Springer INdAM Series 3

Massimo A. Picardello *Editor*

Trends in Harmonic Analysis

Springer

Springer INdAM Series

Volume 3

For further volumes:
www.springer.com/series/10283

Massimo A. Picardello
Editor

Trends in Harmonic Analysis

Springer

Editor
Massimo A. Picardello
Dipartimento di Matematica
Università di Roma "Tor Vergata"
Rome, Italy

ISSN 2281-518X ISSN 2281-5198 (electronic)
Springer INdAM Series
ISBN 978-88-470-2852-4 ISBN 978-88-470-2853-1 (eBook)
DOI 10.1007/978-88-470-2853-1
Springer Milan Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012948292

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Dedicated to Alessandro Figà-Talamanca on the occasion of his retirement

Preface

This book collects some of the scientific contributions of the participants in the Conference in Harmonic Analysis held at the *Istituto Nazionale di Alta Matematica* from May 30th to June 4th, 2011. It illustrates the wide range of research subjects developed by the Italian research group in harmonic analysis, originally started by Alessandro Figà-Talamanca, to whom the Conference was dedicated on the occasion of his retirement.

In 1978, the mathematicians in this newly formed research group started a cycle of conferences to present and share their research progress. These conferences were held in different places almost every year until this Conference, the thirty-first of the series, and the first whose Proceedings are published.

This book outlines some of the impressive ramifications of the mathematical developments that began when Figà-Talamanca brought the study of harmonic analysis to Italy; the research group that he nurtured has now expanded to cover many areas, and therefore this book is addressed not only to experts in harmonic analysis, summability of Fourier series and singular integrals, but also to experts in potential theory, symmetric spaces, analysis and partial differential equations on Riemannian manifolds, analysis on graphs, trees, buildings and discrete groups, Lie groups and Lie algebras, and even far-reaching applications such as cellular automata and signal processing and connections with mathematical logic.

In the last decades, Alessandro Figà-Talamanca has worked on harmonic analysis on trees, and his influence on ongoing research is also underlined by the fact that several contributions to the present volume, even those dealing with completely different subjects, are related to trees or similar discrete structures.

Rome, Italy Massimo Picardello

Acknowledgements

We acknowledge with pleasure the hospitality of the *Istituto Nazionale di Alta Matematica* and the financial support of the *Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le Applicazioni*.

We are very grateful to the referees for their accurate reading of each contributed paper and for their valuable suggestions.

The Editor is very grateful to all the authors for the patience that they have shown in kindly and promptly replying to his countless requests for revisions.

Contents

The Shifted Wave Equation on Damek–Ricci Spaces and on Homogeneous Trees

Jean-Philippe Anker, Pierre Martinot, Emmanuel Pedon, and Alberto G. Setti

Abstract We solve explicitly the shifted wave equation on Damek–Ricci spaces, using the inverse dual Abel transform and Ásgeirsson's theorem. As an application, we investigate Huygens' principle. A similar analysis is carried out in the discrete setting of homogeneous trees.

Keywords Abel transform · Damek–Ricci space · Homogeneous tree · Huygens' principle · Hyperbolic space · Wave equation · Wave propagation

Mathematics Subject Classification (2010) Primary 35L05 · 43A85 · Secondary 20F67 · 22E30 · 22E35 · 33C80 · 43A80 · 58J45

J.-P. Anker (✉)
Fédération Denis Poisson (FR 2964) and Laboratoire MAPMO (UMR 7349), Bâtiment de Mathématiques, Université d'Orléans and CNRS, B.P. 6759, 45067 Orléans Cedex 2, France
e-mail: anker@univ-orleans.fr

P. Martinot
2 rue de la Maladrie, 54630 Flavigny-sur-Moselle, France
e-mail: pierremartinot@orange.fr

E. Pedon
Fédération ARC (FR 3399) and Laboratoire de Mathématiques (EA 4535), Université de Reims (Champagne-Ardenne), Moulin de la Housse, B.P. 1039, 51687 Reims Cedex 2, France
e-mail: emmanuel.pedon@univ-reims.fr

A.G. Setti
Sezione di Matematica, Dipartimento di Scienza e Alta Tecnologia, Università degli Studi dell'Insubria, via Valleggio 11, 22100 Como, Italy
e-mail: alberto.setti@uninsubria.it

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3,
DOI 10.1007/978-88-470-2853-1_1, © Springer-Verlag Italia 2013

1 Introduction

In the book [17] Helgason uses Ásgeirsson's mean value theorem (see Theorem II.5.28) to solve the wave equation

$$\begin{cases} \partial_t^2 u(x,t) = \Delta_x u(x,t), \\ u(x,0) = f(x), \quad \partial_t|_{t=0}\, u(x,t) = g(x), \end{cases} \tag{1}$$

on Euclidean spaces $\mathbb{R}^d$ (see [17, Exercise II.F.1, p. 342] and its solution at pp. 574–575) and the shifted wave equation

$$\begin{cases} \partial_t^2 u(x,t) = \{\Delta_x + (d-1)^2/4\} u(x,t), \\ u(x,0) = f(x), \quad \partial_t|_{t=0}\, u(x,t) = g(x), \end{cases} \tag{2}$$

on real hyperbolic spaces $H^d(\mathbb{R})$ (see [17, Exercise II.F.2, p. 343] and its solution at pp. 575–577). In this work we extend this approach both to Damek–Ricci spaces and to homogeneous trees. Along the way we clarify the role of the inverse dual Abel transform in solving the shifted wave equation.

Recall that Damek–Ricci spaces are Riemannian manifolds, which contain all hyperbolic spaces $H^d(\mathbb{R})$, $H^d(\mathbb{C})$, $H^d(\mathbb{H})$, $H^2(\mathbb{O})$ as a small subclass and share nevertheless several features with these spaces. Before [17] the shifted wave equation (2) on $H^d(\mathbb{R})$ was solved explicitly in [24, Section 7]. Other hyperbolic spaces were dealt with in [10, 19, 20] and Damek–Ricci spaces in [25]. All these approaches are awkward in our opinion. On one hand, [10, 24] and [19, 20] rely on the method of descent, i.e., on shift operators, which reduce the problem to checking formulae in low dimensions. Moreover [10] involves classical compact dual symmetric spaces and doesn't cover the exceptional case. On the other hand, [25] involves complicated computations and follows two different methods: Helgason's approach for hyperbolic spaces and heat kernel expressions [1] for general Damek–Ricci spaces. In comparison we believe that our presentation is simpler and more conceptual.

Several other works deal with the shifted wave equation (2) without using explicit solutions. Let us mention [7] (see also [18, Section V.5]) for Huygens' principle and the energy equipartition on Riemannian symmetric spaces of the noncompact type. This work was extended to Damek–Ricci spaces in [4], to Chébli-Trimèche hypergroups in [14] and to the trigonometric Dunkl setting in [5, 6]. The non-linear shifted wave equation was studied in [2, 3, 28], first on real hyperbolic spaces and next on Damek–Ricci spaces. These works involve sharp dispersive and Strichartz estimates for the linear equation. Related $L^p \to L^p$ estimates were obtained in [21] on hyperbolic spaces.

Our paper is organized as follows. In Sect. 2, we review Damek–Ricci spaces and spherical analysis thereon. We give in particular explicit expressions for the Abel transform, its dual and the inverse transforms. In Sect. 3 we extend Ásgeirsson's mean value theorem to Damek–Ricci spaces, apply it to solutions to the shifted wave equation and deduce explicit expressions, using the inverse dual Abel transform. As an application, we investigate Huygens' principle. Section 4 deals with the shifted wave equation on homogeneous trees, which are discrete analogs of hyperbolic spaces.

Most of this work was done several years ago. The results on Damek–Ricci spaces were cited in [26] and we take this opportunity to thank François Rouvière for mentioning them and for encouraging us to publish details. We are also grateful to Nalini Anantharaman for pointing out to us the connection between our discrete wave equation (16) on trees and recent works [8, 9] of Brooks and Lindenstrauss.

2 Spherical Analysis on Damek–Ricci Spaces

We shall be content with a brief review about Damek–Ricci spaces and we refer to the lecture notes [26] for more information.

Damek–Ricci spaces are solvable Lie groups $S = N \rtimes A$, which are extensions of Heisenberg type groups N by $A \cong \mathbb{R}$ and which are equipped with a left-invariant Riemannian structure. At the Lie algebra level,

$$\mathfrak{s} \equiv \underbrace{\mathbb{R}^m \oplus \overbrace{\mathbb{R}^k}^{\mathfrak{z}}}_{\mathfrak{n}} \oplus \underbrace{\mathbb{R}}_{\mathfrak{a}},$$

with Lie bracket

$$\big[(X,Y,z),\big(X',Y',z'\big)\big] = \left(\frac{z}{2}X' - \frac{z'}{2}X, zY' - z'Y + \big[X,X'\big], 0\right)$$

and inner product

$$\big\langle (X,Y,z),\big(X',Y',z'\big)\big\rangle = \big\langle X,X'\big\rangle_{\mathbb{R}^m} + \big\langle Y,Y'\big\rangle_{\mathbb{R}^k} + zz'.$$

At the Lie group level,

$$S \equiv \underbrace{\mathbb{R}^m \times \overbrace{\mathbb{R}^k}^{Z}}_{N} \times \underbrace{\mathbb{R}}_{A},$$

with multiplication

$$(X,Y,z)\cdot\big(X',Y',z'\big) = \left(X + e^{z/2}X', Y + e^{z}Y' + \frac{1}{2}e^{z/2}\big[X,X'\big], z + z'\right).$$

So far N could be any simply connected nilpotent Lie group of step $\leqslant 2$. Heisenberg type groups are characterized by conditions involving the Lie bracket and the inner product on $\mathfrak{n}$, that we shall not need explicitly. In particular Z is the center of N and m is even. One denotes by

$$n = m + k + 1$$

the (manifold) dimension of S and by

$$Q = \frac{m}{2} + k$$

the homogeneous dimension of N.

Via the Iwasawa decomposition, all hyperbolic spaces $H^d(\mathbb{R})$, $H^d(\mathbb{C})$, $H^d(\mathbb{H})$, $H^2(\mathbb{O})$ can be realized as Damek–Ricci spaces, real hyperbolic spaces corresponding to the degenerate case where N is abelian. But most Damek–Ricci spaces are not symmetric, although harmonic, and thus provide numerous counterexamples to the Lichnerowicz conjecture [13]. Despite the lack of symmetry, radial analysis on S is similar to the hyperbolic space case and fits into Jacobi function theory [22].

In polar coordinates, the Riemannian volume on S may be written as $\delta(r)\,dr\,d\sigma$, where

$$\begin{aligned}\delta(r) &= \overbrace{2^{m+1}\pi^{n/2}\Gamma\left(\frac{n}{2}\right)^{-1}}^{\text{const}}\left(\sinh\frac{r}{2}\right)^{m}(\sinh r)^{k}\\ &= \underbrace{2^{n}\pi^{n/2}\Gamma\left(\frac{n}{2}\right)^{-1}}_{\text{const}}\left(\cosh\frac{r}{2}\right)^{k}\left(\sinh\frac{r}{2}\right)^{n-1}\end{aligned}$$

is the common surface measure of all spheres of radius r in S and $d\sigma$ denotes the normalized surface measure on the unit sphere in $\mathfrak{s}$. We shall not need the full expression of the Laplace–Beltrami operator Δ on S but only its radial part

$$\operatorname{rad}\Delta = \left(\frac{\partial}{\partial r}\right)^{2} + \underbrace{\left\{\frac{n-1}{2}\coth\frac{r}{2} + \frac{k}{2}\tanh\frac{r}{2}\right\}}_{\frac{\delta'(r)}{\delta(r)}}\frac{\partial}{\partial r}$$

on radial functions and its horocyclic part

$$\Delta f = \left(\frac{\partial}{\partial z}\right)^{2} f - Q\frac{\partial}{\partial z} f \tag{3}$$

on N-invariant functions, i.e., on functions $f = f(X, Y, z)$ depending only on z. The Laplacian Δ commutes both with left translations and with the averaging projector

$$f^{\sharp}(r) = \frac{1}{\delta(r)}\int_{S(e,r)} dx\, f(x),$$

hence with all spherical means

$$f^{\sharp}_{x}(r) = \frac{1}{\delta(r)}\int_{S(x,r)} dy\, f(y).$$

Thus

$$(\Delta f)_x^\sharp = (\operatorname{rad}\Delta)\, f_x^\sharp. \tag{4}$$

Finally Δ has a spectral gap. More precisely, its L^2-spectrum is equal to the half-line $(-\infty, -Q^2/4]$.

Radial Fourier analysis on S may be summarized by the following commutative diagram in the Schwartz space setting [1]:

$$\begin{array}{ccc} & \mathscr{S}(\mathbb{R})_{\text{even}} & \\ \mathscr{H} \nearrow \approx & & \approx \nwarrow \mathscr{F} \\ \mathscr{S}(S)^\sharp & \underset{\mathscr{A}}{\overset{\approx}{\longrightarrow}} & \mathscr{S}(\mathbb{R})_{\text{even}} \end{array}$$

Here

$$\mathscr{H} f(\lambda) = \int_S dx\, \varphi_\lambda(x)\, f(x)$$

denotes the spherical Fourier transform on S,

$$\mathscr{A} f(z) = \mathrm{e}^{-(Q/2)z} \int_{\mathbb{R}^m} dX \int_{\mathbb{R}^k} dY\; f(X, Y, z)$$

the Abel transform,

$$\mathscr{F} f(\lambda) = \int_{\mathbb{R}} dz\, \mathrm{e}^{i\lambda z} f(z)$$

the classical Fourier transform on $\mathbb{R}$ and $\mathscr{S}(S)^\sharp$ the space of smooth radial functions $f(x) = f(|x|)$ on S such that

$$\sup_{r \geqslant 0} (1+r)^M \mathrm{e}^{(Q/2)r} \left| \left(\frac{\partial}{\partial r}\right)^N f(r) \right| < +\infty$$

for every $M, N \in \mathbb{N}$. Recall that the Abel transform and its inverse can be expressed explicitly in terms of Weyl fractional transforms, which are defined by

$$\begin{aligned} \mathscr{W}_\mu^\tau f(r) = \frac{1}{\Gamma(\mu+M)} \int_r^{+\infty} & d(\cosh\tau s)\, (\cosh\tau s - \cosh\tau r)^{\mu+M-1} \\ & \times \left(-\frac{d}{d(\cosh\tau s)}\right)^M f(s) \end{aligned}$$

for $\tau > 0$ and for $\mu \in \mathbb{C}$, $M \in \mathbb{N}$ such that $\operatorname{Re}\mu > -M$. Specifically,

$$\mathscr{A} = c_1 \mathscr{W}_{m/2}^{1/2} \circ \mathscr{W}_{k/2}^{1} \quad \text{and} \quad \mathscr{A}^{-1} = \frac{1}{c_1} \mathscr{W}_{-k/2}^{1} \circ \mathscr{W}_{-m/2}^{1/2},$$

where $c_1 = 2^{(3m+k)/2}\pi^{(m+k)/2}$. More precisely,

$$\mathscr{A}^{-1} f(r) = \frac{1}{c_1}\left(-\frac{d}{d(\cosh r)}\right)^{k/2}\left(-\frac{d}{d(\cosh r/2)}\right)^{m/2} f(r)$$

if n is odd, i.e., k is even, and

$$\mathscr{A}^{-1} f(r) = \frac{1}{c_1\sqrt{\pi}}\int_r^{+\infty}\frac{ds}{\sqrt{\cosh s - \cosh r}} \times \left(-\frac{d}{ds}\right)\left(-\frac{d}{d(\cosh s)}\right)^{(k-1)/2}\left(-\frac{d}{d(\cosh s/2)}\right)^{m/2} f(s)$$

if n is even, i.e., k is odd. Similarly, the dual Abel transform

$$\mathscr{A}^{\star} f(r) = (\tilde{f})^{\sharp}(r), \quad \text{where } \tilde{f}(X,Y,z) = \mathrm{e}^{(Q/2)z} f(z), \tag{5}$$

and its inverse can be expressed explicitly in terms of Riemann-Liouville fractional transforms $\mathscr{R}_{\mu}^{\tau}$, which are defined by

$$\mathscr{R}_{\mu}^{\tau} f(r) = \frac{1}{\Gamma(\mu+M)}\int_0^r d(\cosh\tau s)\,(\cosh\tau r - \cosh\tau s)^{\mu+M-1} \times \left(\frac{d}{d(\cosh\tau s)}\right)^{M} f(s)$$

for $\tau > 0$ and for $\mu \in \mathbb{C}$, $M \in \mathbb{N}$ such that $\operatorname{Re}\mu > -M$.

Theorem 2.1 *The dual Abel transform (5) is a topological isomorphism between $C^{\infty}(\mathbb{R})_{\text{even}}$ and $C^{\infty}(S)^{\sharp} \equiv C^{\infty}(\mathbb{R})_{\text{even}}$. Explicitly,*

$$\mathscr{A}^{\star} f(r) = \frac{c_2}{2}\left(\sinh\frac{r}{2}\right)^{-m}(\sinh r)^{1-k}\mathscr{R}_{k/2}^{1}\left\{\left(\cosh\frac{\cdot}{2}\right)^{-1}\mathscr{R}_{m/2}^{1/2}\left[\left(\sinh\frac{\cdot}{2}\right)^{-1} f\right]\right\}(r)$$

and

$$\left(\mathscr{A}^{\star}\right)^{-1} f(r) = \frac{1}{c_2}\frac{d}{dr}\left(\mathscr{R}_{-m/2}^{1/2}\circ\mathscr{R}_{1-k/2}^{1}\right)\left\{\left(\sinh\frac{\cdot}{2}\right)^{m}(\sinh\cdot)^{k-1} f\right\}(r)$$

where $c_2 = 2^{n/2-1/2}\pi^{-1/2}\Gamma(\frac{n}{2}) = 2^{1/2-n/2}(n-1)!\,\Gamma(\frac{n+1}{2})^{-1}$. More precisely,

$$\left(\mathscr{A}^{\star}\right)^{-1} f(r) = \frac{1}{c_2}\frac{d}{dr}\left(\frac{d}{d(\cosh r/2)}\right)^{m/2}\left(\frac{d}{d(\cosh r)}\right)^{k/2-1} \times \left\{\left(\sinh\frac{r}{2}\right)^{m}(\sinh r)^{k-1} f(r)\right\}$$

if n is odd, i.e., k is even, and

$$\left(\mathscr{A}^{\star}\right)^{-1} f(r)=\frac{1}{c_2\sqrt{\pi}}\,\frac{d}{dr}\left(\frac{d}{d(\cosh r/2)}\right)^{m/2}\left(\frac{d}{d(\cosh r)}\right)^{(k-1)/2} \times\int_0^r \frac{ds}{\sqrt{\cosh r-\cosh s}}\left(\sinh\frac{s}{2}\right)^m (\sinh s)^k f(s)$$

if n is even, i.e., k is odd.

Proof Everything follows from the duality formulae

$$\int_{\mathbb{R}} dr\,\mathscr{A} f(r) g(r)=\int_S dx\, f(x)\mathscr{A}^{\star} g(x),$$

$$\int_0^{+\infty} d(\cosh\tau r)\,\mathscr{W}_{\mu}^{\tau} f(r) g(r)=\int_0^{+\infty} d(\cosh\tau r)\, f(r)\mathscr{R}_{\mu}^{\tau} g(r),$$

and from the properties of the Riemann-Liouville transforms, in particular

$$\mathscr{R}_{1/2}^{\tau}: r^{\ell} C^{\infty}(\mathbb{R})_{\text{even}} \xrightarrow{\approx} r^{\ell+1} C^{\infty}(\mathbb{R})_{\text{even}}$$

for every integer $\ell\geqslant -1$. □

Remark 2.2 In the degenerate case $m=0$, we recover the classical expressions for real hyperbolic spaces $H^n(\mathbb{R})$:

$$\mathscr{A} f(r)=\frac{(2\pi)^{(n-1)/2}}{\Gamma(\frac{n-1}{2})}\int_r^{+\infty} d(\cosh s)\,(\cosh s-\cosh r)^{(n-3)/2} f(s),$$

$$\mathscr{A}^{\star} f(r)=c_3(\sinh r)^{-(n-2)}\int_0^r ds\,(\cosh r-\cosh s)^{(n-3)/2} f(s),$$

where $c_3=\frac{2^{(n-1)/2}\Gamma(\frac{n}{2})}{\sqrt{\pi}\,\Gamma(\frac{n-1}{2})}=\frac{(n-2)!}{2^{(n-3)/2}\Gamma(\frac{n-1}{2})^2}$,

$$\mathscr{A}^{-1} f(r)=(2\pi)^{-(n-1)/2}\left(-\frac{d}{d(\cosh r)}\right)^{(n-1)/2} f(r),$$

$$\left(\mathscr{A}^{\star}\right)^{-1} f(r)=2^{(n-1)/2}\frac{(\frac{n-1}{2})!}{(n-1)!}\,\frac{d}{dr}\left(\frac{d}{d(\cosh r)}\right)^{(n-3)/2}\left\{(\sinh r)^{n-2} f(r)\right\}$$

if n is odd and

$$\mathscr{A}^{-1} f(r)=\frac{1}{2^{(n-1)/2}\pi^{n/2}} \times\int_r^{+\infty}\frac{ds}{\sqrt{\cosh s-\cosh r}}\left(-\frac{d}{ds}\right)\left(-\frac{d}{d(\cosh s)}\right)^{n/2-1} f(s),$$

$$(\mathscr{A}^\star)^{-1} f(r) = \frac{1}{2^{(n-1)/2}(\frac{n}{2}-1)!} \frac{d}{dr} \left(\frac{d}{d(\cosh r)} \right)^{n/2-1} \times \int_0^r \frac{ds}{\sqrt{\cosh r - \cosh s}} (\sinh s)^{n-1} f(s)$$

if n is even.

3 Ásgeirsson's Mean Value Theorem and the Shifted Wave Equation on Damek–Ricci Spaces

Theorem 3.1 *Assume that* $U \in C^\infty(S \times S)$ *satisfies*

$$\Delta_x U(x,y) = \Delta_y U(x,y). \tag{6}$$

Then

$$\int_{S(x,r)} dx' \int_{S(y,s)} dy'\, U(x',y') = \int_{S(x,s)} dx' \int_{S(y,r)} dy'\, U(x',y') \tag{7}$$

for every $x, y \in S$ *and* $r, s > 0$.

Proof The proof is similar to the real hyperbolic space case [17, Sect. II.5.6] once one has introduced the double spherical means

$$U^{\sharp,\sharp}_{x,y}(r,s) = \frac{1}{\delta(r)} \int_{S(x,r)} dx' \frac{1}{\delta(s)} \int_{S(y,s)} dy'\, U(x',y')$$

and transformed (6) into

$$(\mathrm{rad}\,\Delta)_r U^{\sharp,\sharp}_{x,y}(r,s) = (\mathrm{rad}\,\Delta)_s U^{\sharp,\sharp}_{x,y}(r,s). \qquad \square$$

Ásgeirsson's Theorem is the following limit case of Theorem 3.1, which is obtained by dividing (7) by $\delta(s)$ and by letting $s \to 0$.

Corollary 3.2 *Under the same assumptions*,

$$\int_{S(x,r)} dx'\, U(x',y) = \int_{S(y,r)} dy'\, U(x,y').$$

Given a solution $u \in C^\infty(S \times \mathbb{R})$ to the shifted wave equation

$$\partial_t^2 u(x,t) = \big(\Delta_x + Q^2/4\big) u(x,t) \tag{8}$$

on S with initial data $u(x,0) = f(x)$ and $\partial_t|_{t=0}\, u(x,t) = 0$, set

$$U(x,y) = \mathrm{e}^{(Q/2)t} u(x,t), \tag{9}$$

where t is the z coordinate of y. Then (9) satisfies (6), according to (3). By applying Corollary 3.2 to (9) with $y = e$ and $r = |t|$, we deduce that the dual Abel transform of $t \mapsto u(x,t)$, as defined in (5), is equal to the spherical mean $f_x^\sharp(|t|)$ of the initial datum f. Hence

$$u(x,t) = \big(\mathscr{A}^\star\big)^{-1}\big(f_x^\sharp\big)(t).$$

By integrating with respect to time, we obtain the solutions

$$u(x,t) = \int_0^t ds\, \big(\mathscr{A}^\star\big)^{-1}\big(g_x^\sharp\big)(s)$$

to (8) with initial data $u(x,0) = 0$ and $\partial_t|_{t=0}\, u(x,t) = g(x)$. In conclusion, general solutions to the shifted wave equation

$$\begin{cases} \partial_t^2 u(x,t) = (\Delta_x + Q^2/4)u(x,t) \\ u(x,0) = f(x), \quad \partial_t|_{t=0}\, u(x,t) = g(x) \end{cases} \tag{10}$$

on S are given by

$$u(x,t) = \big(\mathscr{A}^\star\big)^{-1}\big(f_x^\sharp\big)(t) + \int_0^t ds\, \big(\mathscr{A}^\star\big)^{-1}\big(g_x^\sharp\big)(s).$$

By using Theorem 2.1, we deduce the following explicit expressions.

Theorem 3.3

(i) *When n is odd, the solution to (10) is given by*

$$u(x,t) = c_4 \frac{\partial}{\partial t}\left(\frac{\partial}{\partial(\cosh t/2)}\right)^{m/2}\left(\frac{\partial}{\partial(\cosh t)}\right)^{k/2-1}\left\{\frac{1}{\sinh t}\int_{S(x,|t|)} dy\, f(y)\right\}$$
$$+ c_4 \left(\frac{\partial}{\partial(\cosh t/2)}\right)^{m/2}\left(\frac{\partial}{\partial(\cosh t)}\right)^{k/2-1}\left\{\frac{1}{\sinh t}\int_{S(x,|t|)} dy\, g(y)\right\},$$

with $c_4 = 2^{-3m/2-k/2-1}\pi^{-(n-1)/2}$.

(ii) *When n is even, the solution to (10) is given by*

$$u(x,t) = c_5 \frac{\partial}{\partial|t|}\left(\frac{\partial}{\partial(\cosh t/2)}\right)^{m/2}\left(\frac{\partial}{\partial(\cosh t)}\right)^{(k-1)/2}$$
$$\times \int_{B(x,|t|)} dy\, \frac{f(y)}{\sqrt{\cosh t - \cosh \mathrm{d}(y,x)}}$$
$$+ c_5\, \mathrm{sign}(t)\left(\frac{\partial}{\partial(\cosh t/2)}\right)^{m/2}\left(\frac{\partial}{\partial(\cosh t)}\right)^{(k-1)/2}$$
$$\times \int_{B(x,|t|)} dy\, \frac{g(y)}{\sqrt{\cosh t - \cosh \mathrm{d}(y,x)}},$$

with $c_5 = 2^{-3m/2-k/2-1}\pi^{-n/2}$.

Remark 3.4 These formulae extend to the degenerate case $m = 0$, which corresponds to real hyperbolic spaces $H^n(\mathbb{R})$:

(i) n odd:

$$u(t,x) = c_6 \frac{\partial}{\partial t}\left(\frac{\partial}{\partial(\cosh t)}\right)^{(n-3)/2}\left\{\frac{1}{\sinh t}\int_{S(x,|t|)} dy\, f(y)\right\}$$
$$+ c_6\left(\frac{\partial}{\partial(\cosh t)}\right)^{(n-3)/2}\left\{\frac{1}{\sinh t}\int_{S(x,|t|)} dy\, g(y)\right\},$$

with $c_6 = 2^{-(n+1)/2}\pi^{-(n-1)/2}$.

(ii) n even:

$$u(t,x) = c_7 \frac{\partial}{\partial |t|}\left(\frac{\partial}{\partial(\cosh t)}\right)^{n/2-1}\int_{B(x,|t|)} dy\, \frac{f(y)}{\sqrt{\cosh t - \cosh \mathrm{d}(y,x)}}$$
$$+ c_7 \,\mathrm{sign}(t)\left(\frac{\partial}{\partial(\cosh t)}\right)^{\frac{n}{2}-1}\int_{B(x,|t|)} dy\, \frac{g(y)}{\sqrt{\cosh t - \cosh \mathrm{d}(y,x)}},$$

with $c_7 = 2^{-(n+1)/2}\pi^{-n/2}$.

As an application, let us investigate the propagation of solutions u to the shifted wave equation (10) with initial data f, g supported in a ball $B(x_0, R)$. The following two statements are immediate consequences of Theorem 3.3. Firstly, waves propagate at unit speed.

Corollary 3.5 *Under the above assumptions,*

$$\mathrm{supp}\, u \subset \left\{(x,t) \in S \mid \mathrm{d}(x,x_0) \leqslant |t| + R\right\}.$$

Secondly, Huygens' principle holds in odd dimension, as in the Euclidean setting. This phenomenon was already observed in [25].

Corollary 3.6 *Assume that n is odd. Then, under the above assumptions,*

$$\mathrm{supp}\, u \subset \left\{(x,t) \in S \mid |t| - R \leqslant \mathrm{d}(x,x_0) \leqslant |t| + R\right\}.$$

In even dimension, $u(x,t)$ may not vanish when $\mathrm{d}(x,x_0) < |t| - R$, but it tends asymptotically to 0. This phenomenon was observed in several settings, for instance on Euclidean spaces in [27], on Riemannian symmetric spaces of the noncompact type [7], on Damek–Ricci spaces [4], for Chébli-Trimèche hypergroups [14], ... Our next result differs from [4, 7, 14] in two ways. On one hand, we use explicit expressions instead of the Fourier transform. On the other hand, we aim at energy estimates as in [27], which are arguably more appropriate than pointwise estimates. Recall indeed that the total energy

$$\mathscr{E}(t) = \mathscr{K}(t) + \mathscr{P}(t) \tag{11}$$

is time independent, where

$$\mathscr{K}(t) = \frac{1}{2} \int_S dx \left| \partial_t u(x,t) \right|^2$$

is the kinetic energy and

$$\begin{aligned} \mathscr{P}(t) &= \frac{1}{2} \int_S dx \left(-\Delta_x - Q^2/4 \right) u(x,t) \overline{u(x,t)} \\ &= \frac{1}{2} \int_S dx \left\{ \left| \nabla_x u(x,t) \right|^2 - Q^2/4 \left| u(x,t) \right|^2 \right\} \end{aligned}$$

the potential energy. By the way, let us mention that the equipartition of (11) into kinetic and potential energies was investigated in [7] and in the subsequent works [4, 5, 14] (see also [18, Sect. V.5.5] and the references cited therein).

Lemma 3.7 *Let u be a solution to* (10) *with smooth initial data f, g supported in a ball $B(x_0, R)$. Then*

$$u(x,t),\ \partial_t u(x,t),\ \nabla_x u(x,t) \quad \text{are} \quad \mathrm{O}\big(\mathrm{e}^{-(Q/2)|t|}\big)$$

for every $x \in S$ and $t \in \mathbb{R}$ such that $\mathrm{d}(x, x_0) \leqslant |t| - R - 1$.

Proof Assume $t > 0$ and consider the second part

$$\begin{aligned} v(x,t) = &\left(\frac{\partial}{\partial(\cosh t/2)} \right)^{m/2} \left(\frac{\partial}{\partial(\cosh t)} \right)^{(k-1)/2} \\ &\times \int_{B(x,t)} dy \, \frac{g(y)}{\sqrt{\cosh t - \cosh \mathrm{d}(y,x)}} \end{aligned} \tag{12}$$

of the solution $u(x,t)$ in part (ii) of Theorem 3.3. The case $t < 0$ and the first part are handled similarly. As $B(x_0, R) \subset B(x,t)$, we have

$$\int_{B(x,t)} dy \, \frac{g(y)}{\sqrt{\cosh t - \cosh \mathrm{d}(y,x)}} = \int_{B(x_0,R)} dy \, \frac{g(y)}{\sqrt{\cosh t - \cosh \mathrm{d}(y,x)}}$$

and thus it remains to apply the differential operator

$$D_t = \left(\frac{\partial}{\partial(\cosh t/2)} \right)^{m/2} \left(\frac{\partial}{\partial(\cosh t)} \right)^{(k-1)/2}$$

to $\{\cosh t - \cosh \mathrm{d}(y,x)\}^{-1/2}$. Firstly

$$\left(\frac{\partial}{\partial(\cosh t)}\right)^{(k-1)/2}\big\{\cosh t - \cosh \mathrm{d}(y,x)\big\}^{-1/2} = \mathrm{const}\big\{\cosh t - \cosh \mathrm{d}(y,x)\big\}^{-k/2}$$

and secondly

$$\begin{aligned}&\left(\frac{\partial}{\partial(\cosh t/2)}\right)^{m/2}\big\{\cosh t - \cosh \mathrm{d}(y,x)\big\}^{-k/2}\\ &\quad= \sum_{0\leqslant j\leqslant \frac{m}{4}} a_j (\cosh t/2)^{m/2-2j}\big\{\cosh t - \cosh \mathrm{d}(y,x)\big\}^{-k/2-m/2+j},\end{aligned}$$

for some constants a_j. As

$$\cosh t - \cosh \mathrm{d}(y,x) = 2\sinh\frac{t+\mathrm{d}(y,x)}{2}\sinh\frac{t-\mathrm{d}(y,x)}{2} \asymp \mathrm{e}^t,$$

we conclude that $D_t\{\cosh t - \cosh \mathrm{d}(y,x)\}^{-1/2}$ and hence $v(x,t)$ are $\mathrm{O}(\mathrm{e}^{-\frac{Q}{2}t})$. The derivatives $\partial_t v(x,t)$ and $\nabla_x v(x,t)$ are estimated similarly. As far as $\nabla_x v(x,t)$ is concerned, we use in addition that

$$\sinh \mathrm{d}(y,x) = \mathrm{O}\big(\mathrm{e}^t\big) \quad \text{and} \quad \big|\nabla_x \mathrm{d}(y,x)\big| \leqslant 1.$$

This concludes the proof of Lemma 3.7. □

Theorem 3.8 *Let u be a solution to* (10) *with initial data $f, g \in C_c^\infty(S)$ and let $R = R(t)$ be a positive function such that*

$$\begin{cases} R(t) \to +\infty \\ R(t) = \mathrm{o}(|t|) \end{cases} \quad \text{as } t \to \pm\infty.$$

Then

$$\int_{\mathrm{d}(x,e)<|t|-R(t)} dx\,\Big\{\big|u(x,t)\big|^2 + \big|\nabla_x u(x,t)\big|^2 + \big|\partial_t u(x,t)\big|^2\Big\}$$

tend to 0 *as $t \to \pm\infty$. In other words, the energy of u concentrates asymptotically inside the spherical shell*

$$\big\{x \in S \mid |t| - R(t) \leqslant \mathrm{d}(x,e) \leqslant |t| + R(t)\big\}.$$

Proof By combining Lemma 3.7 with the volume estimate

$$\operatorname{vol} B\big(e, |t| - R(t)\big) \asymp \mathrm{e}^{Q\{|t|-R(t)\}} \quad \text{as } t \to \pm\infty,$$

we deduce that the three integrals

$$\int_{\mathrm{d}(x,e)<|t|-R(t)} dx\, \bigl|u(x,t)\bigr|^2,$$
$$\int_{\mathrm{d}(x,e)<|t|-R(t)} dx\, \bigl|\nabla_x u(x,t)\bigr|^2,$$
$$\int_{\mathrm{d}(x,e)<|t|-R(t)} dx\, \bigl|\partial_t u(x,t)\bigr|^2$$

are $\mathrm{O}(\mathrm{e}^{-QR(t)})$ and hence tend to 0 as $t \to \pm\infty$. □

4 The Shifted Wave Equation on Homogeneous Trees

This section is devoted to a discrete setting, which is similar to the continuous setting considered so far. A homogeneous tree $\mathbb{T} = \mathbb{T}_q$ of degree $q+1 > 2$ is a connected graph with no loops and with the same number $q+1$ of edges at each vertex. We shall be content with a brief review and we refer to the expository paper [12] for more information (see also the monographs [15, 16]).

For the counting measure, the volume of any sphere $S(x,n)$ in $\mathbb{T}$ is given by

$$\delta(n) = \begin{cases} 1 & \text{if } n = 0, \\ (q+1)q^{n-1} & \text{if } n \in \mathbb{N}^\star. \end{cases}$$

Once we have chosen an origin $0 \in \mathbb{T}$ and a geodesic $\omega : \mathbb{Z} \to \mathbb{T}$ through 0, let us denote by $|x| \in \mathbb{N}$ the distance of a vertex $x \in \mathbb{T}$ to the origin and by $h(x) \in \mathbb{Z}$ its horocyclic height (see Fig. 1).

The combinatorial Laplacian is defined on $\mathbb{Z}$ by

$$\mathscr{L}^{\mathbb{Z}} f(n) = f(n) - \frac{f(n+1) + f(n-1)}{2},$$

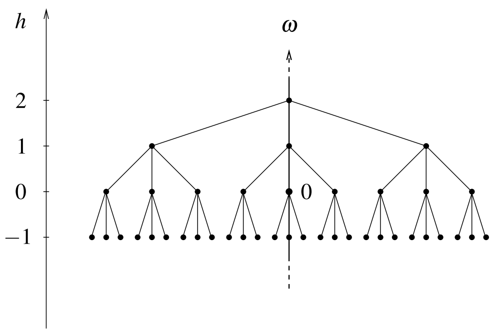


Fig. 1 Upper half space picture of $\mathbb{T}_3$

and similarly on $\mathbb{T}$ by

$$\mathscr{L}^{\mathbb{T}} f(x) = f(x) - \frac{1}{q+1} \sum_{y \in S(x,1)} f(y). \tag{13}$$

The L^2-spectrum of $\mathscr{L}^{\mathbb{T}}$ is equal to the interval $[1-\gamma, 1+\gamma]$, where

$$\gamma = \frac{2}{q^{1/2} + q^{-1/2}} \in (0,1).$$

We have

$$\mathscr{L}^{\mathbb{T}} f(n) = \begin{cases} f(0) - f(1) & \text{if } n = 0, \\ f(n) - \frac{1}{q+1} f(n-1) - \frac{q}{q+1} f(n+1) & \text{if } n \in \mathbb{N}^\star \end{cases} \tag{14}$$

on radial functions and

$$\begin{aligned} \mathscr{L}^{\mathbb{T}} f(h) &= f(h) - \frac{q}{q+1} f(h-1) - \frac{1}{q+1} f(h+1) \\ &= \gamma q^{h/2} \mathscr{L}_h^{\mathbb{Z}} \{ q^{-h/2} f(h) \} + (1-\gamma) f(h) \end{aligned} \tag{15}$$

on horocyclic functions.

Again, radial Fourier analysis on $\mathbb{T}$ may be summarized by the following commutative diagram

$$\begin{array}{ccc} & C^\infty(\mathbb{R}/\tau\mathbb{Z})_{\text{even}} & \\ \mathscr{H} \nearrow \approx & & \approx \nwarrow \mathscr{F} \\ \mathscr{S}(\mathbb{T})^\sharp & \underset{\mathscr{A}}{\overset{\approx}{\longrightarrow}} & \mathscr{S}(\mathbb{Z})_{\text{even}} \end{array}$$

Here

$$\mathscr{H} f(\lambda) = \sum_{x \in \mathbb{T}} \varphi_\lambda(x) f(x) \quad \text{for every } \lambda \in \mathbb{R}$$

denotes the spherical Fourier transform on $\mathbb{T}$,

$$\mathscr{A} f(h) = q^{h/2} \sum_{\substack{x \in \mathbb{T} \\ h(x) = h}} f(|x|) \quad \text{for every } h \in \mathbb{Z}$$

the Abel transform and

$$\mathscr{F} f(\lambda) = \sum_{h \in \mathbb{Z}} q^{i\lambda h} f(h) \quad \text{for every } \lambda \in \mathbb{R}$$

a variant of the classical Fourier transform on $\mathbb{Z}$. Moreover $\tau = \frac{2\pi}{\log q}$, $\mathscr{S}(\mathbb{Z})_{\text{even}}$ denotes the space of even functions on $\mathbb{Z}$ such that

$$\sup_{n\in\mathbb{N}^\star} n^k \big| f(n) \big| < +\infty \quad \text{for every } k \in \mathbb{N},$$

and $\mathscr{S}(\mathbb{T})^\sharp$ the space of radial functions on $\mathbb{T}$ such that

$$\sup_{n\in\mathbb{N}^\star} n^k q^{n/2} \big| f(n) \big| < +\infty \quad \text{for every } k \in \mathbb{N}.$$

Consider finally the dual Abel transform

$$\mathscr{A}^\star f(n) = \frac{1}{\delta(n)} \sum_{\substack{x\in\mathbb{T}\\ |x|=n}} q^{h(x)/2} f\big(h(x)\big) \quad \text{for every } n \in \mathbb{N}.$$

The following expressions are obtained by elementary computations.

Lemma 4.1

(i) *The Abel transform is given by*

$$\mathscr{A} f(h) = q^{|h|/2} f\big(|h|\big) + \frac{q-1}{q} \sum_{k=1}^{+\infty} q^{|h|/2+k} f\big(|h|+2k\big)$$
$$= \sum_{k=0}^{+\infty} q^{|h|/2+k} \big\{ f\big(|h|+2k\big) - f\big(|h|+2k+2\big) \big\} \quad \textit{for every } h \in \mathbb{Z}$$

and the dual Abel transform by

$$\mathscr{A}^\star f(n) = \frac{2q}{q+1} q^{-|n|/2} f(\pm n) + \frac{q-1}{q+1} q^{-|n|/2} \sum_{\substack{-|n|<k<|n| \\ k \text{ has same parity as } n}} f(\pm k)$$

if $n \in \mathbb{Z}^\star$, *resp.* $\mathscr{A}^\star f(0) = f(0)$.

(ii) *The inverse Abel transform is given by*

$$\mathscr{A}^{-1} f(n) = \sum_{k=0}^{+\infty} q^{-n/2-k} \big\{ f(n+2k) - f(n+2k+2) \big\}$$
$$= q^{-n/2} f(n) - (q-1) \sum_{k=1}^{+\infty} q^{-n/2-k} f(n+2k) \quad \textit{for every } n \in \mathbb{N}$$

and the inverse dual Abel transform by

$$\begin{aligned}\big(\mathscr{A}^\star\big)^{-1} f(h) &= \frac{1}{2} q^{h/2} f(h) + \frac{1}{2} q^{-h/2} f(1) \\ &\quad + \frac{1}{2} \sum_{k=1}^{(h-1)/2} q^{h/2-2k+1} \big\{ f(h-2k+2) - f(h-2k) \big\} \\ &= \frac{q^{1/2}+q^{-1/2}}{2} q^{(h-1)/2} f(h) - \frac{q-q^{-1}}{2} q^{-\frac{h}{2}} \sum_{0<k \text{ odd}<h} q^k f(k)\end{aligned}$$

if $h \in \mathbb{N}$ *is odd, respectively*

$$\big(\mathscr{A}^\star\big)^{-1} f(0) = f(0)$$

and

$$\begin{aligned}\big(\mathscr{A}^\star\big)^{-1} f(h) &= \frac{1}{2} q^{h/2} f(h) + \frac{1}{2} q^{-h/2} f(0) \\ &\quad + \frac{1}{2} \sum_{k=1}^{h/2} q^{h/2-2k+1} \big\{ f(h-2k+2) - f(h-2k) \big\} \\ &= \frac{q^{1/2}+q^{-1/2}}{2} q^{(h-1)/2} f(h) - \frac{q^{1/2}-q^{-1/2}}{2} q^{-(h-1)/2} f(0) \\ &\quad - \frac{q-q^{-1}}{2} q^{-h/2} \sum_{0<k \text{ even}<h} q^k f(k)\end{aligned}$$

if $h \in \mathbb{N}^\star$ *is even.*

We are interested in the following shifted wave equation on $\mathbb{T}$:

$$\begin{cases} \gamma \mathscr{L}_n^{\mathbb{Z}} u(x,n) = (\mathscr{L}_x^{\mathbb{T}} - 1 + \gamma) u(x,n), \\ u(x,0) = f(x), \quad \{u(x,1) - u(x,-1)\}/2 = g(x). \end{cases} \tag{16}$$

As was pointed out to us by Nalini Anantharaman, this equation occurs in the recent works [8, 9]. The unshifted wave equation with discrete time was studied in [11] and the shifted wave equation with continuous time in [23].

We will solve (16) by applying the following discrete version of Ásgeirsson's mean value theorem and by using the explicit expression of the inverse dual Abel transform.

Theorem 4.2 *Let* U *be a function on* $\mathbb{T}$ *such that*

$$\mathscr{L}_x^{\mathbb{T}} U(x,y) = \mathscr{L}_y^{\mathbb{T}} U(x,y) \quad \textit{for every } x, y \in \mathbb{T}. \tag{17}$$

Then

$$\sum_{x'\in S(x,m)}\sum_{y'\in S(y,n)} U\big(x',y'\big) = \sum_{x'\in S(x,n)}\sum_{y'\in S(y,m)} U\big(x',y'\big)$$

for every $x, y \in \mathbb{T}$ *and* $m, n \in \mathbb{N}$. *In particular*

$$\sum_{x'\in S(x,n)} U\big(x',y\big) = \sum_{y'\in S(y,n)} U\big(x,y'\big). \tag{18}$$

In order to prove Theorem 4.2, we need the following discrete analog of (4).

Lemma 4.3 *Consider the spherical means*

$$f_x^\sharp(n) = \frac{1}{\delta(n)} \sum_{y\in S(x,n)} f(y) \quad \textit{for every } x \in \mathbb{T}, n \in \mathbb{N}.$$

Then

$$\big(\mathscr{L}^{\mathbb{T}} f\big)_x^\sharp(n) = (\operatorname{rad}\mathscr{L})_n\, f_x^\sharp(n),$$

where $\operatorname{rad}\mathscr{L}$ *denotes the radial part* (14) *of* $\mathscr{L}^{\mathbb{T}}$.

Proof We have

$$\big(\mathscr{L}^{\mathbb{T}} f\big)_x^\sharp(n) = \begin{cases} f(x) - f_x^\sharp(1) & \text{if } n = 0, \\ f_x^\sharp(n) - \frac{1}{q+1} f_x^\sharp(n-1) - \frac{q}{q+1} f_x^\sharp(n+1) & \text{if } n \in \mathbb{N}^\star. \end{cases} \qquad \square$$

Proof of Theorem 4.2 Fix $x, y \in \mathbb{T}$ and consider the double spherical means

$$U_{x,y}^{\sharp,\sharp}(m,n) = \frac{1}{\delta(m)} \sum_{x'\in S(x,m)} \frac{1}{\delta(n)} \sum_{y'\in S(y,n)} U\big(x',y'\big),$$

that we shall denote by $V(m,n)$ for simplicity. According to Lemma 4.3, our assumption (17) may be rewritten as

$$(\operatorname{rad}\mathscr{L})_m V(m,n) = (\operatorname{rad}\mathscr{L})_n V(m,n). \tag{19}$$

Let us prove the symmetry

$$V(m,n) = V(n,m) \quad \text{for every } m, n \in \mathbb{N} \tag{20}$$

by induction on $\ell = m + n$. First of all, (20) is trivial if $\ell = 0$ and (20) with $\ell = 1$ is equivalent to (19) with $m = n = 0$. Assume next that $\ell \geqslant 1$ and that (20) holds for $m + n \leqslant \ell$. On one hand, let $m > n > 0$ with $m + n = \ell + 1$ and let $1 \leqslant k \leqslant m - n$. We deduce from (19) at the point $(m-k, n+k-1)$ that

$$
\begin{aligned}
&V(m-k+1,n+k-1)-V(m-k,n+k)\\
&\quad =q\big\{V(m-k,n+k-2)-V(m-k-1,n+k-1)\big\}.
\end{aligned}
\tag{21}
$$

By adding up (21) over k, we obtain

$$
V(m,n)-V(n,m)=q\big\{V(m-1,n-1)-V(n-1,m-1)\big\},
\tag{22}
$$

which vanishes by induction. On the other hand, we deduce from (19) at the points $(\ell,0)$ and $(0,\ell)$ that

$$
\begin{cases}
V(\ell+1,0)=(q+1)V(\ell,1)-qV(\ell,0),\\
V(0,\ell+1)=(q+1)V(1,\ell)-qV(0,\ell).
\end{cases}
$$

Hence $V(\ell+1,0)=V(0,\ell+1)$ by using (22) and by induction. This concludes the proof of Theorem 4.2. □

Let us now solve explicitly the shifted wave equation (16) on $\mathbb{T}$ as we did in Sect. 3 for the shifted wave equation (10) on Damek–Ricci spaces. Consider first a solution u to (16) with initial data $u(x,0)=f(x)$ and $\{u(x,1)-u(x,-1)\}/2=0$. On one hand, as $(x,n)\mapsto u(x,-n)$ satisfies the same Cauchy problem, we have $u(x,-n)=u(x,n)$ by uniqueness. On the other hand, according to (15), the function

$$
U(x,y)=q^{h(y)/2}u\big(x,h(y)\big)\quad \text{for every } x,y\in\mathbb{T}
$$

satisfies (17). Thus, by applying (18) to U with $y=0$, we deduce that the dual Abel transform of $n\mapsto u(x,n)$ is equal to the spherical mean $f_x^\sharp(n)$ of the initial datum f. Hence

$$
u(x,n)=\big(\mathscr{A}^\star\big)^{-1}\big(f_x^\sharp\big)(n)\quad \text{for every } x\in\mathbb{T}, n\in\mathbb{N}.
$$

Consider next a solution u to (16) with initial data $u(x,0)=0$ and $\{u(x,1)-u(x,-1)\}/2=g(x)$. Then $u(x,n)$ is an odd function of n and

$$
v(x,n)=\frac{u(x,n+1)-u(x,n-1)}{2}
$$

is a solution to (16) with initial data $v(x,0)=g(x)$ and $\{v(x,1)-v(x,-1)\}/2=0$. Hence

$$
u(x,n)=\begin{cases}
2\sum_{0<k\ \text{odd}<n} v(x,k) & \text{if } n\in\mathbb{N}^\star \text{ is even},\\
g(x)+2\sum_{0<k\ \text{even}<n} v(x,k) & \text{if } n\in\mathbb{N}^\star \text{ is odd},
\end{cases}
$$

with $v(x,n)=(\mathscr{A}^\star)^{-1}(g_x^\sharp)(n)$. Using part (ii) of Lemma 4.1, we deduce the following explicit expressions.

Theorem 4.4 *The solution to* (16) *is given by*

$$u(x,n)=\frac{1}{2}q^{-|n|/2}\sum_{\mathrm{d}(y,x)=|n|}f(y)-\frac{q-1}{2}q^{-|n|/2}\sum_{\substack{\mathrm{d}(y,x)<|n|\\|n|-\mathrm{d}(y,x)\ \text{even}}}f(y)$$
$$+\,\mathrm{sign}(n)q^{1/2-|n|/2}\sum_{\substack{\mathrm{d}(y,x)<|n|\\|n|-\mathrm{d}(y,x)\ \text{odd}}}g(y)\quad \textit{for every } x\in\mathbb{T},n\in\mathbb{Z}^{\star}.$$

In other words,

$$u(x,n)=\overbrace{\frac{M_{|n|}-M_{|n|-2}}{2}}^{C_n}f(x)+\overbrace{\mathrm{sign}(n)M_{|n|-1}}^{S_n}g(x), \tag{23}$$

where

$$M_nf(x)=q^{-n/2}\sum_{\substack{\mathrm{d}(y,x)\leqslant n\\n-\mathrm{d}(y,x)\ \text{even}}}f(y) \tag{24}$$

if $n\geqslant 0$ *and* $M_{-1}=0$.

Remark 4.5 Notice that the radial convolution operators C_n and S_n above correspond, via the Fourier transform, to the multipliers

$$\cos_q n\lambda \quad \text{and} \quad \frac{\sin_q n\lambda}{\sin_q\lambda},$$

where $\cos_q\lambda=(q^{i\lambda}+q^{-i\lambda})/2$ and $\sin_q\lambda=(q^{i\lambda}-q^{-i\lambda})/2i$.

As we did in Sect. 3, let us next deduce propagation properties of solutions u to the shifted wave equation (16) with initial data f,g supported in a ball $B(x_0,N)$.

Corollary 4.6 *Under the above assumptions,*

1. $u(x,n)=\mathrm{O}(q^{-|n|/2})$ *for every* $x\in\mathbb{T}$, $n\in\mathbb{Z}$;
2. $\operatorname{supp}u\subset\{(x,n)\in\mathbb{T}\times\mathbb{Z}\mid \mathrm{d}(x,x_0)\leqslant|n|+N\}$.

Obviously Huygens' principle doesn't hold for (16), strictly speaking. Let us show that it holds asymptotically, as for even dimensional Damek–Ricci spaces. For this purpose, define as follows the kinetic energy

$$\mathcal{K}(n)=\frac{1}{2}\sum_{x\in\mathbb{T}}\left|\frac{u(x,n+1)-u(x,n-1)}{2}\right|^2$$

and the potential energy

$$\mathscr{P}(n) = \frac{1}{4q} \sum_{\substack{x,y\in\mathbb{T}\\ \mathrm{d}(x,y)=2}} \left| \frac{u(x,n)-u(y,n)}{2} \right|^2 - \frac{(q-1)^2}{8q} \sum_{x\in\mathbb{T}} \left|u(x,n)\right|^2$$
$$= \frac{q+1}{8} \sum_{x\in\mathbb{T}} (\widetilde{\mathscr{L}}_x - \tilde{\gamma}) u(x,n)\overline{u(x,n)} \tag{25}$$

for solutions u to (16). Here

$$\widetilde{\mathscr{L}} f(x) = f(x) - \frac{1}{q(q+1)} \sum_{y\in S(x,2)} f(y)$$

is the 2-step Laplacian on $\mathbb{T}$ and

$$\tilde{\gamma} = \frac{(q-1)^2}{q(q+1)} \in (0,1).$$

Lemma 4.7

(i) *The L^2-spectrum of $\widetilde{\mathscr{L}}$ is equal to the interval $[\tilde{\gamma}, (q+1)/q]$. Thus the potential energy (25) is nonnegative.*

(ii) *The total energy*

$$\mathscr{E}(n) = \mathscr{K}(n) + \mathscr{P}(n)$$

is independent of $n \in \mathbb{Z}$.

Proof (i) Follows for instance from the relation

$$\widetilde{\mathscr{L}} = \frac{q+1}{q} \mathscr{L}^{\mathbb{T}} \big(2 - \mathscr{L}^{\mathbb{T}}\big)$$

and from the fact that the L^2-spectrum of $\mathscr{L}^{\mathbb{T}}$ is equal to the interval $[1-\gamma, 1+\gamma]$.

(ii) Notice that the shifted wave equation

$$\gamma \mathscr{L}_n^{\mathbb{Z}} u(x,n) = \big(\mathscr{L}_x^{\mathbb{T}} - 1 + \gamma\big) u(x,n)$$

amounts to

$$u(x,n+1) + u(x,n-1) = \frac{1}{\sqrt{q}} \sum_{y\in S(x,1)} u(y,n).$$

As

$$\sum_{x\in\mathbb{T}} \sum_{y,z\in S(x,1)} u(y,n)\overline{u(z,n)} = (q+1) \sum_{x\in\mathbb{T}} \left|u(x,n)\right|^2 + \sum_{\substack{y,z\in\mathbb{T}\\ \mathrm{d}(y,z)=2}} u(y,n)\overline{u(z,n)},$$

we have on one hand

$$\mathscr{K}(n) = \frac{q+1}{8q} \sum_{x\in\mathbb{T}} \big|u(x,n)\big|^2 + \frac{1}{2} \sum_{x\in\mathbb{T}} \big|u(x,n\pm 1)\big|^2 + \frac{1}{8q} \sum_{\substack{x,y\in\mathbb{T}\\ \mathrm{d}(x,y)=2}} u(x,n)\overline{u(y,n)} - \frac{1}{2\sqrt{q}} \sum_{\substack{x,y\in\mathbb{T}\\ \mathrm{d}(x,y)=1}} \mathrm{Re}\big\{u(x,n)\overline{u(y,n\pm 1)}\big\}. \tag{26}$$

On the other hand,

$$\mathscr{P}(n) = \frac{3q-1}{8q} \sum_{x\in\mathbb{T}} \big|u(x,n)\big|^2 - \frac{1}{8q} \sum_{\substack{x,y\in\mathbb{T}\\ \mathrm{d}(x,y)=2}} u(x,n)\overline{u(y,n)}. \tag{27}$$

By adding up (26) and (27), we obtain

$$\mathscr{E}(n) = \frac{1}{2} \sum_{x\in\mathbb{T}} \big|u(x,n)\big|^2 + \frac{1}{2} \sum_{x\in\mathbb{T}} \big|u(x,n\pm 1)\big|^2 - \frac{1}{2\sqrt{q}} \sum_{\substack{x,y\in\mathbb{T}\\ \mathrm{d}(x,y)=1}} \mathrm{Re}\big\{u(x,n)\overline{u(y,n\pm 1)}\big\}$$

and we deduce from this expression that

$$\mathscr{E}(n) = \mathscr{E}(n\pm 1).$$

This concludes the proof of Lemma 4.7. □

Remark 4.8 Alternatively, part (ii) of Lemma 4.7 can be proved by expressing the energies $\mathscr{K}(n)$, $\mathscr{P}(n)$, $\mathscr{E}(n)$ in terms of the initial data f, g and by using spectral calculus. Specifically,

$$\mathscr{K}(n) = \frac{1}{8} \sum_{x\in\mathbb{T}} (C_{n+1} - C_{n-1})^2 f(x)\overline{f(x)} + \frac{1}{8} \sum_{x\in\mathbb{T}} (S_{n+1} - S_{n-1})^2 g(x)\overline{g(x)} + \frac{1}{4} \mathrm{Re} \sum_{x\in\mathbb{T}} (C_{n+1} - C_{n-1})(S_{n+1} - S_{n-1}) f(x)\overline{g(x)}$$

and

$$\begin{aligned}\mathscr{P}(n) &= \frac{1}{4}\sum_{x\in\mathbb{T}}(1-C_2)C_n^2 f(x)\overline{f(x)}\\ &\quad+\frac{1}{4}\sum_{x\in\mathbb{T}}(1-C_2)S_n^2 g(x)\overline{g(x)}\\ &\quad+\frac{1}{2}\operatorname{Re}\sum_{x\in\mathbb{T}}(1-C_2)C_n S_n f(x)\overline{g(x)}.\end{aligned}$$

Here we have used the fact that

$$\frac{q+1}{8}(\widetilde{\mathscr{L}}-\tilde{\gamma}) = \frac{1}{8}(3-M_2) = \frac{1}{4}(1-C_2).$$

Hence

$$\mathscr{E}(n) = \sum_{x\in\mathbb{T}} U_n^+ f(x)\overline{f(x)} + \sum_{x\in\mathbb{T}} V_n^+ g(x)\overline{g(x)} + 2\operatorname{Re}\sum_{x\in\mathbb{T}} W_n^+ f(x)\overline{g(x)},$$

where

$$\begin{aligned}U_n^+ &= \frac{1}{8}(C_{n+1}-C_{n-1})^2 + \frac{1}{4}(1-C_2)C_n^2,\\ V_n^+ &= \frac{1}{8}(S_{n+1}-S_{n-1})^2 + \frac{1}{4}(1-C_2)S_n^2,\\ W_n^+ &= \frac{1}{8}(C_{n+1}-C_{n-1})(S_{n+1}-S_{n-1}) + \frac{1}{4}(1-C_2)C_nS_n.\end{aligned}$$

By considering the corresponding multipliers, we obtain

$$U_n^+ = \frac{1}{4}(1-C_2), \qquad V_n^+ = \frac{1}{2}, \qquad W_n^+ = 0,$$

and conclude that

$$\mathscr{E}(n) = \frac{1}{4}\sum_{x\in\mathbb{T}}(1-C_2)f(x)\overline{f(x)} + \frac{1}{2}\sum_{x\in\mathbb{T}}\left|g(x)\right|^2 = \mathscr{E}(0).$$

Let us turn to the asymptotic equipartition of the total energy $\mathscr{E} = \mathscr{E}(n)$.

Theorem 4.9 *Let u be a solution to (16) with finitely supported initial data f and g. Then the kinetic energy $\mathscr{K}(n)$ and the potential energy $\mathscr{P}(n)$ tend both to $\mathscr{E}/2$ as $n \to \pm\infty$.*

Proof Let us show that the difference $\mathscr{K}(n) - \mathscr{P}(n)$ tends to 0. By resuming the computations in Remark 4.8, we obtain

$$\mathscr{K}(n) - \mathscr{P}(n) = \sum_{x\in\mathbb{T}} U_n^- f(x)\overline{f(x)} + \sum_{x\in\mathbb{T}} V_n^- g(x)\overline{g(x)} + 2\,\mathrm{Re}\sum_{x\in\mathbb{T}} W_n^- f(x)\overline{g(x)},$$

with

$$\begin{aligned} U_n^- &= \frac{1}{8}(C_{n+1} - C_{n-1})^2 - \frac{1}{4}(1 - C_2)C_n^2 = -\frac{1}{4}(1 - C_2)C_{2n}, \\ V_n^- &= \frac{1}{8}(S_{n+1} - S_{n-1})^2 - \frac{1}{4}(1 - C_2)S_n^2 = \frac{1}{2}C_{2n}, \\ W_n^- &= \frac{1}{8}(C_{n+1} - C_{n-1})(S_{n+1} - S_{n-1}) - \frac{1}{4}(1 - C_2)C_n S_n = -\frac{1}{4}(1 - C_2)S_{2n}. \end{aligned}$$

As

$$\|C_{2n} f\|_{\ell^\infty} \leqslant \frac{q-1}{2} q^{-|n|} \|f\|_{\ell^1} \quad \text{and} \quad \big\|(1 - C_2) f\big\|_{\ell^1} \leqslant \left\{\frac{q - q^{-1}}{2} + 2\right\} \|f\|_{\ell^1},$$

the expression

$$\sum_{x\in\mathbb{T}} U_n^- f(x)\overline{f(x)} = -\frac{1}{4}\sum_{x\in\mathbb{T}} C_{2n} f(x)\overline{(1 - C_2) f(x)}$$

tends to 0. The expressions

$$\sum_{x\in\mathbb{T}} V_n^- g(x)\overline{g(x)} = \frac{1}{2}\sum_{x\in\mathbb{T}} C_{2n}\, g(x)\overline{g(x)}$$

and

$$\sum_{x\in\mathbb{T}} W_n^- f(x)\overline{g(x)} = -\frac{1}{4}\sum_{x\in\mathbb{T}} S_{2n} f(x)\overline{(1 - C_2) f(x)}$$

are handled in the same way. This concludes the proof of Theorem 4.9. □

Let us conclude with the asymptotic Huygens principle.

Theorem 4.10 *Let u be a solution to (16) with finitely supported initial data and let $(N_n)_{n\in\mathbb{Z}}$ be a sequence of positive integers such that*

$$\begin{cases} N_n \to +\infty \\ N_n = \mathrm{o}(|n|) \end{cases} \quad \text{as } n \to \pm\infty.$$

Then the expressions

$$\sum_{\substack{x\in\mathbb{T}\\|x|<|n|-N_n}} \big|u(x,n)\big|^2, \qquad \sum_{\substack{x,\,y\,\in\,\mathbb{T}\\|x|,\,|y|\,<\,|n|-N_n\\\mathrm{d}(x,y)=2}} \big|u(x,n)-u(y,n)\big|^2,$$

$$\sum_{\substack{x\in\mathbb{T}\\|x|<|n|-N_n}} \big|u(x,n+1)-u(x,n-1)\big|^2$$

tend to 0 *as* $n \to \pm\infty$. *In other words, the energy of u concentrates asymptotically inside the spherical shell*

$$\big\{x \in \mathbb{T} \mid |n| - N_n \leqslant |x| \leqslant |n| + N_n\big\}.$$

The *proof* is similar to the proof of Theorem 3.8.

Acknowledgements Research partially supported by the European Commission (HCM Network *Fourier Analysis* 1994–1997, TMR Network *Harmonic Analysis* 1998–2002, IHP Network *HARP* 2002–2006) and by a French-Italian research project (PHC Galilée 25970QB *VAMP* 2011–2012).

References

1. Anker, J.-P., Damek, E., Yacoub, C.: Spherical analysis on harmonic AN groups. Ann. Sc. Norm. Super. Pisa, Cl. Sci. **23**, 643–679 (1996)
2. Anker, J.-P., Pierfelice, V., Vallarino, M.: The wave equation on hyperbolic spaces. J. Differ. Equ. **252**, 5613–5661 (2012)
3. Anker, J.-P., Pierfelice, V., Vallarino, M.: The wave equation on Damek–Ricci spaces. arXiv:1012.0689
4. Astengo, F., Di Blasio, B.: Huygens' principle and a Paley-Wiener type theorem on Damek–Ricci spaces. Ann. Math. Blaise Pascal **17**, 327–340 (2010)
5. Ayadi, F.: Equipartition of energy for the wave equation associated to the Dunkl-Cherednik Laplacian. J. Lie Theory **18**, 747–755 (2008)
6. Ben Said, S.: Huygens' principle for the wave equation associated with the trigonometric Dunkl-Cherednik operators. Math. Res. Lett. **13**, 43–58 (2006)
7. Branson, T., Ólafsson, G., Schlichtkrull, H.: Huygens' principle in Riemannian symmetric spaces. Math. Ann. **301**, 445–462 (1995)
8. Brooks, S., Lindenstrauss, E.: Graph eigenfunctions and quantum unique ergodicity. C. R. Acad. Sci. Paris Sér. I (Math.) **348**, 829–834 (2010)
9. Brooks, S., Lindenstrauss, E.: Non-localization of eigenfunctions on large regular graphs. Isr. J. Math. 1–14 (2012). doi:10.1007/s11856-012-0096-y
10. Bunke, U., Olbrich, M.: The wave kernel for the Laplacian on the classical locally symmetric spaces of rank one (theta functions, trace formulas and the Selberg zeta function). Ann. Glob. Anal. Geom. **12**, 357–405 (1994)
11. Cohen, J.M., Pagliacci, M.: Explicit solutions for the wave equation on homogeneous trees. Adv. Appl. Math. **15**, 390–403 (1994)
12. Cowling, M.G., Meda, S., Setti, A.G.: An overview of harmonic analysis on the group of isometries of a homogeneous tree. Expo. Math. **16**, 385–423 (1998)
13. Damek, E., Ricci, F.: A class of nonsymmetric harmonic Riemannian spaces. Bull. Am. Math. Soc. **27**, 139–142 (1992)

14. El Kamel, J., Yacoub, C.: Huygens' principle and equipartition of energy for the modified wave equation associated to a generalized radial Laplacian. Ann. Math. Blaise Pascal **12**, 147–160 (2005)
15. Figà-Talamanca, A., Nebbia, C.: Harmonic Analysis and Representation Theory for Groups Acting on Homogeneous Trees. London Math. Soc. Lect. Note Ser., vol. 162. Cambridge University Press, Cambridge (1991)
16. Figà-Talamanca, A., Picardello, M.A.: Harmonic Analysis on Free Groups. Lect. Notes Pure Appl. Math., vol. 87. Dekker, New York (1983)
17. Helgason, S.: Groups and Geometric Analysis (Integral Geometry, Invariant Differential Operators, and Spherical Functions). Academic Press, Orlando (1984). Am. Math. Soc., Providence (2000)
18. Helgason, S.: Geometric Analysis on Symmetric Spaces. Am. Math. Soc., Providence (1994, 2008)
19. Intissar, A., Vall Ould Moustapha, M.: Solution explicite de l'équation des ondes dans un espace symétrique de type non compact de rang 1. Sémin. Théor. Spectr. Géom. **12**, 25–28 (1994). (Univ. Grenoble 1, 1993–1994)
20. Intissar, A., Val Ould Moustapha, M.: Solution explicite de l'équation des ondes dans l'espace symétrique de type non compact de rang 1. C. R. Acad. Sci. Paris Sér. I (Math.) **321**, 77–80 (1995)
21. Ionescu, A.D.: Fourier integral operators on noncompact symmetric spaces of real rank one. J. Funct. Anal. **174**, 274–300 (2000)
22. Koornwinder, T.H.: Jacobi functions and analysis on noncompact semisimple Lie groups. In: Askey, R.A., Koornwinder, T.H., Schempp, W. (eds.) Special Functions (Group Theoretical Aspects and Applications), pp. 1–85. Reidel, Dordrecht (1984)
23. Medolla, G., Setti, A.G.: The wave equation on homogeneous trees. Ann. Mat. Pura Appl. (4) **176**, 1–27 (1999)
24. Lax, P.D., Phillips, R.S.: The asymptotic distribution of lattice points in euclidean and non-euclidean spaces. J. Funct. Anal. **46**, 280–350 (1982)
25. Noguchi, M.: The solution of the shifted wave equation on Damek–Ricci space. Interdiscip. Inf. Sci. **8**, 101–113 (2002)
26. Rouvière, F.: Espaces de Damek–Ricci (géométrie et analyse). In: Faraut, J., Rouvière, F., Vergne, M. (eds.) Analyse sur les Groupes de Lie et Théorie des Représentations, Kénitra, 1999. Sémin. Congr., vol. 7, pp. 45–100. Soc. Math. France, Paris (2003)
27. Strichartz, R.S.: Asymptotic behavior of waves. J. Funct. Anal. **40**, 341–357 (1981)
28. Tataru, D.: Strichartz estimates in the hyperbolic space and global existence for the semilinear wave equation. Trans. Am. Math. Soc. **353**, 795–807 (2001)

Invariance of Capacity Under Quasisymmetric Maps of the Circle: An Easy Proof

Nicola Arcozzi and Richard Rochberg

Abstract We give a direct, combinatorial proof that the logarithmic capacity is essentially invariant under quasisymmetric maps of the circle.

Keywords Planar quasiconformal maps · Capacity · Quasisymmetric maps of the line

Mathematics Subject Classification (2010) Primary 31A15 · Secondary 31C20

1 Introduction and Statement of the Results

It is a known fact that the logarithmic capacity of closed sets is essentially invariant under quasisymmetric maps of the unit circle. An orientation preserving homeomorphism of the unit disc, identified with an increasing homeomorphism $\varphi : [0, 1] \to [0, 1]$ via the map $e^{2\pi it} \mapsto e^{2\pi i(\varphi(t)+\alpha)}$ for some real α, is *quasisymmetric* if

$$\frac{1}{M} \leqslant \frac{\varphi(x+t) - \varphi(x)}{\varphi(x) - \varphi(x-t)} \leqslant M \tag{QS}$$

for some fixed $M > 1$. Sums are taken modulo integers. The reader should be aware that quasisymmetric maps in [2] form a class more general than (QS): the maps satisfying (QS) are those [5]

(i) having a quasi-conformal extension f mapping the unit disc onto itself;
(ii) such that $f(0) = 0$.

N. Arcozzi (✉)
Dipartimento di Matematica, Università di Bologna, Piazza di Porta S. Donato 5, 40127 Bologna, Italy
e-mail: arcozzi@dm.unibo.it

R. Rochberg
Department of Mathematics, Washington University, St. Louis, MO 63130, USA
e-mail: rr@math.wustl.edu

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3,
DOI 10.1007/978-88-470-2853-1_2, © Springer-Verlag Italia 2013

The logarithmic capacity of a closed subset of the unit circle, identified with a subset E of the unit interval, is comparable with its Bessel (2, 1/2)-capacity $\mathrm{Cap}(E)$. Given a positive Borel measure μ on [0, 1], let

$$\mathscr{E}(\mu)=\int_0^1 K\mu(x)^2\,dx=\int_0^1\left(\int_0^1\frac{d\mu(y)}{|x-y|^{1/2}}\right)^2 dx$$

be its energy. Here, too, differences are taken modulo integers. Then,

$$\mathrm{Cap}(E)=\inf\left\{\frac{\mu(E)^2}{\mathscr{E}(\mu)}:\ \mathrm{supp}(\mu)\subseteq E\right\}.$$

Theorem 1.1 (Corollary of a Theorem of Beurling and Ahlfors [5]). *There is a constant $C(M)$ depending on M only such that the inequalities*

$$\frac{1}{C(M)}\leqslant\frac{\mathrm{Cap}\big(\varphi^{-1}(E)\big)}{\mathrm{Cap}(E)}\leqslant C(M)$$

hold for all closed subsets E of [0, 1].

An indirect proof goes as follows. Let f be a quasiconformal extension of φ such that $f(0)=0$: f leaves essentially invariant the capacities of closed sets in the closed unit disc [2].

We will show that Theorem 1.1 can be rephrased as the *stable* version of Benjamini and Peres' result about the equivalence of classical and (a notion of) discrete capacity. See [1, Chapter 2] for an excellent exposition of potential theory at the level of generality that we need in this note (and wider).

Let T be the usual rooted dyadic tree. We can represent T as a tree with labels (n,k), whose vertices are the dyadic subintervals $J^0_{(n,k)}=[(k-1)/2^n,k/2^n]$ of [0, 1] (with $n\geqslant 0$, $1\leqslant k\leqslant 2^n$), and the edges are given by set inclusion: there is an edge between the vertices (n,k) and $(n+1,k')$ if $J^0_{(n+1,k')}\subset J^0_{(n,k)}$. The *root* of T is $o:(0,1)$. If $\alpha=(n,k)$, we call $\mathrm{d}(\alpha):=$ the *level* of α. Consider a different collection $\mathscr{J}$ of closed intervals $J_\alpha\subseteq[0,1]$ (with $\alpha\in T$) such that

- $\mathscr{J}_o=[0,1]$;
- $\mathscr{J}_{\alpha_+}\cup\mathscr{J}_{\alpha_-}=\mathscr{J}_\alpha$, where $\alpha_\pm$, the *children* of α, are the labels of the two halves of $\mathscr{J}^0_\alpha$, the dyadic interval labeled by α.

Clearly, $\mathscr{J}$ and $\mathscr{J}^0$ have the same combinatorial properties, a fact that will be used below to simplify notation.

A half-infinite geodesic starting at the root is a sequence $\{\zeta_n:\ n\geqslant 0\}$ in T with $\mathrm{d}(\zeta_n)=n$ and $J^0_{\zeta_{n+1}}\subset J^0_{\zeta_n}$. Given $\zeta\neq\xi$ in ∂T, let $\zeta\wedge\xi\in T$ be their *closest common ancestor*, that is, the element of T of highest level T common to both ζ and ξ. The function $\rho(\zeta,\xi):=2^{-\mathrm{d}(\xi\wedge\zeta)}$ defines a distance on ∂T.

A closed subset E of [0, 1] can be identified with a subset of T's boundary ∂T in the natural way, through the map $\Lambda_{\mathscr{J}}:\partial T\to[0,1]$ that maps the geodesic $\zeta=$

$\{\zeta_n\}_{n\geqslant 0} \in \partial T$ to the point $\Lambda_{\mathscr{J}}(\zeta) = \bigcap_{n\geqslant 0} J_{\zeta_n}$. The map $\Lambda_{\mathscr{J}}$ is clearly continuous. In the case of $\mathscr{J} = \mathscr{J}^0$, it is a contraction: $|\Lambda_{\mathscr{J}^0}(\zeta) - \Lambda_{\mathscr{J}^0}(\xi)| \leqslant \rho(\zeta, \xi)$.

Theorem 1.2 *Assume that* $\mathscr{J}$ *satisfies the quasi-symmetry condition*

$$\frac{1}{M} \leqslant \frac{|J_\alpha|}{|J_\beta|} \leqslant M \tag{1}$$

whenever J_α *and* J_β *are intervals in* $\mathscr{J}$ *such that* $\mathrm{d}(\alpha) = \mathrm{d}(\beta)$, *and* J_α *and* J_β *are adjacent* as intervals in $[0, 1]$ (*modulo* 1). *Then* $\mathrm{Cap}_T(\Lambda_{\mathscr{J}}^{-1}(E)) \approx \mathrm{Cap}(E)$.

To request condition (1) when J_α and J_β are adjacent in $[0, 1]$ is more than just requesting the same for $\alpha = \gamma_-$ and $\beta = \gamma_+$: the continuous topology of $[0, 1]$ plays a rôle here. The tree capacity Cap_T is the (linear) one naturally defined on the (unweighted) dyadic tree T. Given $h : T \to [0, \infty)$ and $\zeta = \{\zeta_n : n \geqslant 0\}$ in ∂T, let

$$Ih(\zeta) = \sum_{n=0}^{\infty} h(\zeta_n).$$

For $F \subseteq \partial T$,

$$\mathrm{Cap}_T(F) := \inf\left\{ \|h\|^2_{\ell^2(T)} : h \geqslant 0,\ Ih \geqslant 1 \text{ on } F \right\}.$$

Benjamini and Peres proved that $\mathrm{Cap}(E) \approx \mathrm{Cap}_T(\Lambda^{-1}_{\mathscr{J}^0}(E))$ [4]. The same result, in a more general setting, is proved in [3], whence we take our notation.

2 Proof of the Theorems

We first show that Theorem 1.2 implies Theorem 1.1. Indeed, let φ be quasisymmetric and $J_\alpha = \varphi(J^0_\alpha)$. Then $\mathscr{J}$ satisfies the hypothesis of Theorem 1.2 and $\Lambda_{\mathscr{J}} = \varphi \circ \Lambda_{\mathscr{J}^0}$. If E is a closed subset of $[0, 1]$,

$$\mathrm{Cap}\big(\varphi^{-1}(E)\big) \approx \mathrm{Cap}_T\big(\Lambda^{-1}_{\mathscr{J}^0}\big(\varphi^{-1}(E)\big)\big) = \mathrm{Cap}_T\big(\Lambda^{-1}_{\mathscr{J}}(E)\big) \approx \mathrm{Cap}(E)$$

which is Theorem 1.1.

We now prove Theorem 1.2. Let $\mu \geqslant 0$ be an atomless Borel measure on $[0, 1]$ (atoms make energy infinite), that we may identify with a measure $\mu^\star$ on ∂T by the rule $\mu^\star(\Lambda^{-1}_{\mathscr{J}}(J_\alpha)) := \mu(J_\alpha)$ (indeed, this defines $\mu^\star$ uniquely). Consider the Bessel potential

$$K\mu(x) = \int \frac{d\mu(y)}{|x-y|^{1/2}}.$$

For $x \in [0,1]$, let $P_0(x) = \{\alpha \in T : x \in J_\alpha\}$ and let

$$P_1(x) = \bigcup_{\alpha \in P_0(x)} \{\beta : \mathrm{d}(\alpha) = \mathrm{d}(\beta) \text{ and } d_G(\alpha, \beta) \leqslant 3\}$$

and

$$P_2(x) = \bigcup_{\alpha \in P_0(x)} \{\beta : \mathrm{d}(\alpha) = \mathrm{d}(\beta) \text{ and } 2 \leqslant d_G(\alpha, \beta) \leqslant 3\}.$$

Here, d_G is a *graph distance* which takes into account the adjacency relations of the J_α's in $[0,1]$:

$$\min\bigl\{|z - y| : z \in J_\alpha^0,\ y \in J_\beta^0\bigr\} = \bigl[d_G(\alpha, \beta) - 1\bigr]2^{-\mathrm{d}(\alpha)}$$

if $\mathrm{d}(\alpha) = \mathrm{d}(\beta)$. In other words, moving from J_α to J_β across adjacent intervals at the same level, we have to make $d_G(\alpha, \beta)$ steps. We need the following basic properties of $P_2(x)$:

(i) $\sum_{\alpha \in P_2(x)} \chi_{\mathscr{J}_\alpha} \approx 1$;
(ii) $\min\{|y - x| : y \in J_\alpha,\ \alpha \in P_2(x)\} \approx \max\{|y - x| : y \in J_\alpha,\ \alpha \in P_2(x)\} \approx |J_\alpha|$.

The proofs are easy and are left to the reader. Note that (i) is purely combinatoric, while (ii) relies on the metric hypothesis (1): adjacent intervals have comparable length. Using (i) and then (ii), we have

$$\begin{aligned} K\mu(x) &\approx \sum_{\alpha \in P_2(x)} \int_{J_\alpha} \frac{d\mu(y)}{|x - y|^{1/2}} \\ &\approx \int_0^1 \sum_{\alpha \in P_2(x)} |J_\alpha|^{-1/2} \chi(y \in J_\alpha)\, d\mu(y) \\ &= \sum_{\alpha \in P_2(x)} |J_\alpha|^{-1/2} \mu(J_\alpha). \end{aligned}$$

Replacing P_2 by P_1 makes the argument easier to understand. It is trivial that

$$\sum_{\alpha \in P_2(x)} |J_\alpha|^{-1/2} \mu(J_\alpha) \leqslant \sum_{\alpha \in P_1(x)} |J_\alpha|^{-1/2} \mu(J_\alpha).$$

In the other direction, observe that, if $\alpha \in P_1(x)$, then

$$J_\alpha = \bigcup_{\beta \in P_2(x),\, \mathrm{d}(\beta) \geqslant \mathrm{d}(\alpha)} J_\beta \tag{2}$$

and, for each α in $P_1(x)$ and $y \in J_\alpha$, there are boundedly many β's in $P_2(x)$ such that $\mathrm{d}(\beta) \geqslant \mathrm{d}(\alpha)$ and $y \in J_\beta \subseteq J_\alpha$. The second assertion is obvious. For the first one, given $\alpha \in P_1(\alpha) \setminus P_2(\alpha)$, it is easy to see that J_α can be decomposed as the union of J_β's as in (2).

Then,

$$\sum_{\alpha\in P_1(x)} |J_\alpha|^{-1/2}\mu(J_\alpha) \lesssim \sum_{\alpha\in P_1(x)} |J_\alpha|^{-1/2} \sum_{\beta\in P_2(x),\, \mathrm{d}(\beta)\geqslant \mathrm{d}(\alpha)} \mu(J_\beta)$$

$$= \sum_{\beta\in P_2(x)} \mu(J_\beta) \sum_{\alpha\in P_1(x),\, \mathrm{d}(\alpha)\geqslant \mathrm{d}(\beta)} |J_\alpha|^{-1/2}$$

$$\approx \sum_{\beta\in P_2(x)} |J_\beta|^{-1/2}\mu(J_\beta).$$

The energy becomes

$$\mathcal{E}(\mu) = \int_0^1 K\mu(x)^2\,dx$$

$$\approx \int_0^1 \left(\int_0^1 \sum_{\alpha\in P_1(x)} |J_\alpha|^{-1/2}\chi(y\in J_\alpha)\,d\mu(y)\right)^2 dx$$

$$= \int_0^1 d\mu(y)\int_0^1 d\mu(z)\,H(y,z),$$

where

$$H(y,z) = \sum_{J_\alpha\ni y,\, J_\beta\ni z} |J_\alpha|^{-1/2}|J_\beta|^{-1/2}\int_0^1 \chi\big(\alpha,\beta\in P_1(x)\big)\,dx. \tag{3}$$

The kernel $H(y,z)$ is estimated from above and below by a purely combinatorial quantity. Let $\mathrm{d}(y\tilde{\wedge}z) = n \in \mathbb{N}$ be the greatest integer such that there are elements γ_1,γ_2 at level n with $y\in J_{\gamma_1}$, $z\in J_{\gamma_2}$ and $J_{\gamma_1}\cap J_{\gamma_2}\neq\emptyset$ (γ_1 and γ_2 either coincide or they label adjacent intervals in $\mathcal{J}$). After considering a handful of geometric series (in which hypothesis (1) is crucial), it is easily verified that

$$H(y,z) \approx \mathrm{d}(y\tilde{\wedge}z) + 1. \tag{4}$$

As the quantity $\mathrm{d}(y\tilde{\wedge}z)$ is purely combinatorial, it coincides for $\mathcal{J}$ and $\mathcal{J}^0$, and this remark by itself proves Theorem 1.1, without passing through Theorem 1.2. However, we take a different route. It is proved in [4] (and, in greater generality, in [3]) that

$$\int_0^1 d\mu(y)\int_0^1 d\mu(y)\left[\mathrm{d}(y\tilde{\wedge}z)+1\right] \approx \mathcal{E}_T\big({\Lambda^\star}_{\mathcal{J}}\mu\big),$$

where $\mathcal{E}_T(\cdot)$ is the energy associated with the tree capacity Cap_T:

$$\mathcal{E}_T(\nu) := \int_{\partial T} d\nu(\zeta)\int_{\partial T} d\nu(\xi)\left[\mathrm{d}(\zeta\wedge\xi)+1\right].$$

Equivalence of energies, $\mathcal{E}(\mu)\approx\mathcal{E}_T({\Lambda^\star}_{\mathcal{J}}\mu)$, easily implies the equivalence of capacities in Theorem 1.2.

The proofs of Theorems 1.1 and 1.2 presented here are of combinatorial nature: it is then to be expected that they can be extended to a more general context. We think that a general statement can be proved for quasisymmetric maps between Ahlfors regular spaces, by means of some technical tools contained in [3]. We plan to return to this issue in a forthcoming paper. It would also be interesting to see if there are any relations between our approach to quasisymmetric maps in one dimension and the interesting circle of ideas outlined in [6].

A downside of the approach we take here (which was first considered, we like to stress, in [4]) is that we are able to deal with logarithmic capacity of subsets of the unit circle only. Is there a result like Theorem 1.2, relating logarithmic and tree capacities, that can applied to more general closed subsets of the complex plane? And a last question: can one find estimates for the capacity of *condensers*, rather than *sets*, showing that some features of classical potential theory in the complex plane and of potential theory on trees are essentially equivalent?

References

1. Adams, D.R., Hedberg, L.I.: Function Spaces and Potential Theory. Grundlehren der Mathematischen Wissenschaften, vol. 314. Springer, Berlin (1996)
2. Ahlfors, L.V.: Lectures on Quasiconformal Mappings. Mathematical Studies, vol. 10. Van Nostrand, Princeton (1966)
3. Arcozzi, N., Rochberg, R., Sawyer, E., Wick, B.D.: Potential theory on trees, graphs and Ahlfors-regular metric spaces. arXiv:1010.4788 (2010)
4. Benjamini, I., Peres, Y.: Random walks on a tree and capacity in the interval. Ann. Inst. Henri Poincaré Probab. Stat. **28**, 557–592 (1992)
5. Beurling, A., Ahlfors, L.V.: The boundary correspondence under quasiconformal mappings. Acta Math. **96**, 125–142 (1956)
6. Nag, S., Sullivan, D.: Teichmüller theory and the universal period mapping via quantum calculus and the $H^{1/2}$ space on the circle. Osaka J. Math. **32**, 1–34 (1995)

A Koksma–Hlawka Inequality for Simplices

Luca Brandolini, Leonardo Colzani, Giacomo Gigante, and Giancarlo Travaglini

Abstract We estimate the error in the approximation of the integral of a smooth function over a parallelepiped Ω or a simplex S by Riemann sums with deterministic $\mathbb{Z}^d$-periodic nodes. These estimates are in the spirit of the Koksma–Hlawka inequality, and depend on a quantitative evaluation of the uniform distribution of the sampling points, as well as on the total variation of the function. The sets used to compute the discrepancy of the nodes are parallelepipeds with edges parallel to the edges of Ω or S. Similarly, the total variation depends only on the derivatives of the function along directions parallel to the edges of Ω or S.

Keywords Koksma–Hlawka inequality · Quadrature · Discrepancy · Harmonic analysis

Mathematics Subject Classification (2010) Primary 41A55 · Secondary 11K38 · 42B05

L. Brandolini · G. Gigante (✉)
Dipartimento di Ingegneria, Università di Bergamo, Viale Marconi 5, 24044 Dalmine, Bergamo, Italy
e-mail: giacomo.gigante@unibg.it

L. Brandolini
e-mail: luca.brandolini@unibg.it

L. Colzani
Dipartimento di Matematica e Applicazioni, Edificio U5, Università di Milano Bicocca, Via R. Cozzi 53, 20125 Milan, Italy
e-mail: leonardo.colzani@unimib.it

G. Travaglini
Dipartimento di Statistica e Metodi Quantitativi, Edificio U7, Università di Milano Bicocca, Via Bicocca degli Arcimboldi 8, 20126 Milan, Italy
e-mail: giancarlo.travaglini@unimib.it

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3,
DOI 10.1007/978-88-470-2853-1_3,

1 Introduction

The Koksma–Hlawka inequality gives an estimate of the error in a numerical integration

$$\left| \int_{[0,1]^d} f(x)\,dx - \frac{1}{N} \sum_{j=1}^{N} f(z_j) \right| \leqslant \mathscr{D}(z_j)\mathscr{V}(f).$$

Here $\mathscr{D}(z_j)$ is the discrepancy of the finite set of points $\{z_1, \ldots, z_N\}$ in $[0, 1)^d$, defined by

$$\mathscr{D}(z_j) = \sup_I \left| |I| - \frac{1}{N} \sum_{j=1}^{N} \chi_I(z_j) \right|,$$

where I is an interval of the form $[0, t_1] \times [0, t_2] \times \cdots \times [0, t_d]$ with $0 < t_k < 1$, and $|I| = t_1 t_2 \ldots t_d$ is its measure. The term $\mathscr{V}(f)$ is the so-called Hardy–Krause variation, and when f is smooth (say, $\mathscr{C}^d$) this variation takes the form

$$\mathscr{V}(f) = \sum_{\alpha \in \{0,1\}^d, |\alpha| \neq 0} \int_{Q^\alpha} \left| \left(\frac{\partial}{\partial x} \right)^\alpha f(x) \right| dx.$$

The above sum is over all the non vanishing multiindices $\alpha = (\alpha_1, \ldots, \alpha_d)$ taking only the values 0 and 1, $|\alpha| = \alpha_1 + \cdots + \alpha_d$, $(\partial/\partial x)^\alpha = (\partial/\partial x_1)^{\alpha_1} \ldots (\partial/\partial x_d)^{\alpha_d}$, $Q^\alpha = \{(x_1, \ldots, x_d) \in [0, 1]^d : x_j = 1 \text{ if } \alpha_j = 0\}$ is the $|\alpha|$-dimensional face of $[0, 1]^d$ parallel to $\alpha_1 e_1, \ldots, \alpha_d e_d$ ($\{e_1, \ldots, e_d\}$ is the canonical basis of $\mathbb{R}^d$) containing the vertex $(1, \ldots, 1)$, and dx is the $|\alpha|$-dimensional Lebesgue surface measure (see [5, 2.5], [6, 1.4], [7, 2.2]). There is an extensive literature on this type of estimates, where the contribution to the magnitude of the error given by the irregularity of the point distribution $\{z_1, \ldots, z_N\}$ is isolated from the contribution given by the steepness of the variation of the function f. See e.g. [2–8]. In [1], one such result has been proven, where the cube $[0, 1]^d$ is replaced by a generic bounded Borel subset Ω of $\mathbb{R}^d$. More precisely, let $\{z_1, \ldots, z_N\} \subset [0, 1)^d$ be a distribution of N points in the unit cube, and

$$\mathscr{P} = \left\{ z_j + m : j = 1, \ldots, N,\ m \in \mathbb{Z}^d \right\}$$

its periodic extension to the whole Euclidean space $\mathbb{R}^d$. For any $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and $t = (t_1, \ldots, t_d) \in (0, 1)^d$, let

$$I(x, t) = \bigcup_{m \in \mathbb{Z}^d} \left([0, t_1] \times \cdots \times [0, t_d] + x + m \right)$$

be the periodic extension of the interval with opposite vertices x and $x + t$. Call $\mathscr{I}$ the collection of all such possible periodic intervals $I(x, t)$. Finally, let $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ be the torus. The main result in [1] is the following.

Theorem 1.1 *Let f be a smooth $\mathbb{Z}^d$-periodic function on $\mathbb{R}^d$, Ω a bounded Borel subset of $\mathbb{R}^d$, and $\mathscr{P} = \{z_j + m : j = 1, \dots, N,\ m \in \mathbb{Z}^d\}$ a periodic distribution of points as described above. Then*

$$\left| \int_\Omega f(x)\,dx - \frac{1}{N} \sum_{z \in \mathscr{P} \cap \Omega} f(z) \right| \leqslant \mathscr{D}_{\mathscr{I}}(\Omega, \mathscr{P}) \mathscr{V}_{\mathbb{T}^d}(f),$$

where $\mathscr{D}_{\mathscr{I}}(\Omega, \mathscr{P})$ is the discrepancy

$$\mathscr{D}_{\mathscr{I}}(\Omega, \mathscr{P}) = \sup_{I \in \mathscr{I}} \left| |I \cap \Omega| - \frac{1}{N} \sharp(I \cap \Omega \cap \mathscr{P}) \right|,$$

with $|A|$ and $\sharp(A)$ respectively the Lebesgue measure and cardinality of the set A, and $\mathscr{V}_{\mathbb{T}^d}(f)$ is the total variation

$$\mathscr{V}_{\mathbb{T}^d}(f) = \sum_{\alpha \in \{0,1\}^d} 2^{d-|\alpha|} \int_{\mathbb{T}^d} \left| \left(\frac{\partial}{\partial x} \right)^\alpha f(x) \right| dx,$$

where the sum is over all the multiindices α which take only the values 0 and 1, $|\alpha|$ is the number of 1's, and $(\partial/\partial x)^\alpha = (\partial/\partial x_1)^{\alpha_1} \dots (\partial/\partial x_d)^{\alpha_d}$.

The finite sequence $\{z_1, \dots, z_N\}$ may present repetitions, but in this case the sum $\sum_{z \in \mathscr{P} \cap \Omega} f(z)$ must be replaced by $\sum_{j=1}^N \sum_{m \in \mathbb{Z}^d} f(z_j + m)\chi_\Omega(z_j + m)$, and similarly $\sharp(I \cap \Omega \cap \mathscr{P})$ by $\sum_{j=1}^N \sum_{m \in \mathbb{Z}^d} \chi_{I \cap \Omega}(z_j + m)$.

When Ω is contained in $[0, 1)^d$, the discrepancy $\mathscr{D}_{\mathscr{I}}(\Omega, \mathscr{P})$ is dominated by $2^d \sup ||(B \cap \Omega)| - N^{-1} \sharp(B \cap \Omega \cap \mathscr{P})|$, where the sup is over all the intervals B contained in the unit cube. This reflects the difference between the discrepancy in a torus and the one in a cube, and it is due to the fact that an interval in $\mathbb{T}^d$ can be split into at most 2^d intervals in $[0, 1)^d$.

One of the main features of Theorem 1.1 is the simplicity of its statement, in particular in consideration of the fact that the set Ω is completely arbitrary. On the other hand, observe that the total variation $\mathscr{V}_{\mathbb{T}^d}(f)$ takes into account not only the behavior of f in Ω, but also the behavior outside Ω, which is irrelevant in the estimate of

$$\left| \int_\Omega f(x)\,dx - \frac{1}{N} \sum_{z \in \mathscr{P} \cap \Omega} f(z) \right|.$$

Furthermore, the discrepancy $\mathscr{D}_{\mathscr{I}}(\Omega, \mathscr{P})$ is defined in terms of the family of periodic intervals $\mathscr{I}$, which a priori has no relation with Ω.

The aim of this paper is to show how Theorem 1.1 can be improved in order to overcome the two issues mentioned above: variation of f outside Ω and introduction of arbitrary “directions” in the discrepancy. The improved theorem yields results closer to the original Koksma–Hlawka theorem when Ω is an arbitrary parallelepiped (Theorem 2.2) or a simplex (Theorem 3.2) in $\mathbb{R}^d$, f is a smooth function in $\mathbb{R}^d$ not necessarily periodic, and $\mathscr{P}$ is a $\mathbb{Z}^d$-periodic distribution of points.

For the sake of completeness, we sketch here the proof of Theorem 1.1 [1]. In what follows, we denote by $\hat{f}(n)=\int_{\mathbb{T}^d} f(x)e^{-2\pi i n\cdot x}\,dx$ the Fourier transform and by $g*\mu(x)=\int_{\mathbb{T}^d} g(x-y)\,d\mu(y)$ the convolution. These operators are defined also on distributions.

Lemma 1.2 *Let ϕ be a non vanishing complex sequence on $\mathbb{Z}^d$, and assume that both ϕ and $1/\phi$ have tempered growth in $\mathbb{Z}^d$. Also let f be a smooth function on $\mathbb{T}^d$. Define*

$$g(x)=\sum_{n\in\mathbb{Z}^d}\overline{\phi(n)^{-1}}e^{2\pi i n\cdot x},$$

$$\mathfrak{D}f(x)=\sum_{n\in\mathbb{Z}^d}\phi(n)\hat{f}(n)e^{2\pi i n\cdot x}.$$

Finally, let μ be a finite measure on $\mathbb{T}^d$. Then the following identity holds:

$$\int_{\mathbb{T}^d} f(x)\,\overline{d\mu(x)}=\int_{\mathbb{T}^d}\mathfrak{D}f(x)\overline{g*\mu(x)}\,dx.$$

Lemma 1.3 *Let the function g on $\mathbb{R}^d$ be the superposition of the characteristic functions of all the periodic intervals $I(0,t)$ with $t\in(0,1]^d$,*

$$g(x)=\int_{(0,1]^d}\chi_{I(0,t)}(x)\,dt.$$

Then the function g has Fourier expansion

$$g(x)=\sum_{n\in\mathbb{Z}^d}\left(\prod_{k=1}^{d}\big(2\delta(n_k)+2\pi i n_k\big)^{-1}\right)e^{2\pi i n\cdot x},$$

where $n=(n_1,\ldots,n_d)$, $\delta(n_k)=1$ if $n_k=0$ and $\delta(n_k)=0$ if $n_k\neq 0$.

Lemma 1.4 *If f is a smooth function on $\mathbb{T}^d$, then*

$$\begin{aligned}\mathfrak{D}f(x)&=\sum_{n\in\mathbb{Z}^d}\left(\prod_{k=1}^{d}\big(2\delta(n_k)-2\pi i n_k\big)\right)\hat{f}(n)e^{2\pi i n x}\\&=\sum_{\alpha,\beta\in\{0,1\}^d,\ \alpha+\beta=(1,\ldots,1)}(-1)^{|\alpha|}2^{|\beta|}\int_{[0,1]^{|\beta|}}\left(\frac{\partial}{\partial x}\right)^{\alpha}f\big(x+y^{\beta}\big)\,dy^{\beta},\end{aligned}$$

where $(\partial/\partial x)^{\alpha}=(\partial/\partial x_1)^{\alpha_1}\ldots(\partial/\partial x_d)^{\alpha_d}$ and $y^{\beta}=\sum_{j=1}^{d}\beta_j y_j e_j$ and $\{e_j\}_{j=1}^{d}$ is the canonical basis of $\mathbb{R}^d$, and $dy^{\beta}=dy_1^{\beta_1}\ldots dy_d^{\beta_d}$.

The proofs of the above lemmas are quite straightforward. For details, see [1].

Proof of Theorem 1.1 Write $\mu = dx - N^{-1}\sum_{z\in\mathscr{P}}\delta_z$, where δ_z is the point mass centered at z. Apply Lemma 1.2 to the periodization ν of the measure $\chi_\Omega\mu$ and define g and $\mathfrak{D}f$ as in Lemmas 1.3 and 1.4 respectively. Then, by Hölder inequality,

$$\begin{aligned}\left|\int_\Omega f(x)\,\overline{d\mu(x)}\right| &= \left|\int_{\mathbb{T}^d} f(x)\left(\sum_{n\in\mathbb{Z}^d}\chi_\Omega(x+n)\right)\overline{d\mu(x)}\right| \\ &= \left|\int_{\mathbb{T}^d} f(x)\,\overline{d\nu(x)}\right| \leqslant \|\mathfrak{D}f\|_{L^1(\mathbb{T}^d)}\|g*\nu\|_{L^\infty(\mathbb{T}^d)}.\end{aligned}$$

The estimate for $\|\mathfrak{D}f\|_{L^1(\mathbb{T}^d)}$ follows from Lemma 1.4,

$$\int_{\mathbb{T}^d}\left|\mathfrak{D}f(x)\right|dx \leqslant \sum_{\alpha\in\{0,1\}^d} 2^{d-|\alpha|}\int_{\mathbb{T}^d}\left|\left(\frac{\partial}{\partial x}\right)^\alpha f(x)\right|dx.$$

The estimate for $\|g*\nu\|_{L^\infty(\mathbb{T}^d)}$ follows from Lemma 1.3,

$$\begin{aligned}\left|g*\nu(x)\right| &= \left|\int_{\mathbb{R}^d} g(x-y)\chi_\Omega(y)\,d\mu(y)\right| \\ &= \left|\int_{\mathbb{R}^d}\int_{(0,1]^d}\chi_{I(0,t)}(x-y)\,dt\,\chi_\Omega(y)\,d\mu(y)\right| \\ &\leqslant \int_{(0,1]^d}\left|\int_{\mathbb{R}^d}\chi_{-I(-x,t)}(y)\chi_\Omega(y)\,d\mu(y)\right|dt \\ &\leqslant \sup_{t\in(0,1]^d}\mu\big(\Omega\cap\big(-I(-x,t)\big)\big) \\ &\leqslant \sup_{I\in\mathscr{I}}\left||I\cap\Omega| - \frac{1}{N}\sharp(I\cap\Omega\cap\mathscr{P})\right|.\end{aligned}$$

□

2 Parallelepipeds

So far we have followed [1] almost verbatim. From now on we present some new variants. The next proposition is a first intermediate step: it consists of a version of Theorem 1.1 when Ω is an interval and $\mathscr{V}_{\mathbb{T}^d}(f)$ is replaced by a suitable total variation relative to Ω.

Proposition 2.1 *Let f be a smooth function on $\mathbb{R}^d$, Ω a compact interval in $[0,1)^d$, and $\mathscr{P} = \{z_j + m : j = 1,\ldots,N,\ m\in\mathbb{Z}^d\}$ a periodic distribution of points. Then*

$$\left|\int_\Omega f(x)\,dx - \frac{1}{N}\sum_{z\in\mathscr{P}\cap\Omega}^{*} f(z)\right| \leqslant \mathscr{D}_{\mathscr{I}}(\Omega,\mathscr{P})\mathscr{V}^*_\Omega(f),$$

where $\mathcal{V}^*_\Omega(f)$ *is defined as*

$$\mathcal{V}^*_\Omega(f) = \sum_{\alpha\in\{0,1\}^d} \sum_{\beta\leqslant\alpha} 2^{|\alpha|-|\beta|} \int_{\Omega_\alpha} \left| \left(\frac{\partial}{\partial x}\right)^\beta f(x) \right| dx.$$

The symbol $\sum^*_{z\in\mathcal{P}\cap\Omega} f(z)$ *is defined as follows: if* z *belongs to a* j*-dimensional face of the interval* Ω*, then the term* $f(z)$ *in the sum must be replaced by* $2^{j-d} f(z)$*. A multi-index* β *is less than or equal to another multi-index* α *if* $\beta_j \leqslant \alpha_j$ *for any* $j = 1, \ldots, d$*. Finally,* Ω_α *is the union of all the* $|\alpha|$*-dimensional faces of* Ω *parallel to the directions* $\alpha_1 e_1, \ldots, \alpha_d e_d$ ($\{e_1, \ldots, e_d\}$ *is the canonical basis).*

Proof Since the problem is translation invariant, we may assume that Ω is contained in $(0, 1)^d$. Let φ be a positive radial smooth function supported on the unit ball and with integral 1, and let $\varphi_\varepsilon(x) = \varepsilon^{-d}\varphi(x/\varepsilon)$. Then, for ε small enough, the function $(f\chi_\Omega) * \varphi_\varepsilon$ (here the convolution is intended in $\mathbb{R}^d$) is supported in $(0, 1)^d$ and can therefore be thought of as the image in the unit cube of a smooth periodic function. Now,

$$\begin{aligned}
&\left| \int_\Omega f(x)\,dx - \frac{1}{N} \sum^*_{z\in\mathcal{P}\cap\Omega} f(z) \right| \\
&\quad\leqslant \left| \int_\Omega \big(f(x) - (f\chi_\Omega) * \varphi_\varepsilon(x)\big)\,dx \right| \\
&\qquad + \frac{1}{N} \left| \sum_{z\in\mathcal{P}\cap\Omega} (f\chi_\Omega) * \varphi_\varepsilon(z) - \sum^*_{z\in\mathcal{P}\cap\Omega} f(z) \right| \\
&\qquad + \left| \int_\Omega (f\chi_\Omega) * \varphi_\varepsilon(x)\,dx - \frac{1}{N} \sum_{z\in\mathcal{P}\cap\Omega} (f\chi_\Omega) * \varphi_\varepsilon(z) \right|. \qquad (1)
\end{aligned}$$

It is well known that $(f\chi_\Omega) * \varphi_\varepsilon \to f\chi_\Omega$ as $\varepsilon \to 0$ in the L^1 norm. Hence the first term in the above sum goes to zero. As for the second term, observe that if $z \in \mathcal{P}\cap\Omega$ belongs to a j-dimensional face of Ω, then

$$\lim_{\varepsilon\to 0} \int_{\mathbb{R}^d} \chi_\Omega(z-y)\varphi_\varepsilon(y)\,dy = 2^{j-d}.$$

Similarly,

$$\lim_{\varepsilon\to 0} (f\chi_\Omega) * \varphi_\varepsilon(z) - 2^{j-d} f(z) = 0.$$

Therefore the second term in (1) goes to zero. Now we apply Theorem 1.1 to the smooth function $(f\chi_\Omega) * \varphi_\varepsilon$ and obtain

$$\begin{aligned}
&\left| \int_\Omega (f\chi_\Omega) * \varphi_\varepsilon(x)\,dx - \frac{1}{N} \sum_{z\in\mathcal{P}\cap\Omega} (f\chi_\Omega) * \varphi_\varepsilon(z) \right| \\
&\quad\leqslant \mathcal{D}_\mathcal{I}(\Omega, \mathcal{P}) \mathcal{V}_{\mathbb{T}^d}\big((f\chi_\Omega) * \varphi_\varepsilon\big) \leqslant \mathcal{D}_\mathcal{I}(\Omega, \mathcal{P}) \mathcal{V}_{\mathbb{T}^d}(f\chi_\Omega). \qquad (2)
\end{aligned}$$

Here $\mathscr{V}_{\mathbb{T}^d}(f\chi_\Omega)$ is defined again as

$$\mathscr{V}_{\mathbb{T}^d}(f\chi_\Omega) = \sum_{\alpha\in\{0,1\}^d} 2^{d-|\alpha|} \int_{\mathbb{T}^d} \left| \left(\frac{\partial}{\partial x}\right)^\alpha (f\chi_\Omega)(x) \right| dx,$$

but now the integral $\int_{\mathbb{T}^d} |(\partial/\partial x)^\alpha (f\chi_\Omega)(x)|\, dx$ is intended as the total variation of the finite measure $(\partial/\partial x)^\alpha (f\chi_\Omega)$. That this is a measure follows by applying Leibniz rule,

$$\left(\frac{\partial}{\partial x}\right)^\alpha (f\chi_\Omega)(x) = \sum_{\beta+\gamma=\alpha} \left(\frac{\partial}{\partial x}\right)^\beta f(x) \left(\frac{\partial}{\partial x}\right)^\gamma \chi_\Omega(x),$$

and observing that $(\partial/\partial x)^\gamma \chi_\Omega$ is the (signed) surface measure supported on $\Omega_{(1,\ldots,1)-\gamma}$. Thus, the last inequality in (2) follows from the identity

$$\partial/\partial x_j \big((f\chi_\Omega) * \varphi_\varepsilon\big) = \big(\partial/\partial x_j (f\chi_\Omega)\big) * \varphi_\varepsilon$$

and the inequality

$$\begin{aligned} &\int \left|\big((\partial/\partial x)^\alpha (f\chi_\Omega)\big) * \varphi_\varepsilon(x)\right| dx \\ &\quad \leqslant \left(\int \left|(\partial/\partial x)^\alpha (f\chi_\Omega)(x)\right| dx\right)\left(\int \left|\varphi_\varepsilon(x)\right| dx\right). \end{aligned}$$

Finally, again by Leibniz rule,

$$\begin{aligned} \mathscr{V}_{\mathbb{T}^d}(f\chi_\Omega) &= \sum_{\alpha\in\{0,1\}^d} 2^{d-|\alpha|} \int_{\mathbb{T}^d} \left| \left(\frac{\partial}{\partial x}\right)^\alpha (f\chi_\Omega)(x) \right| dx \\ &= \sum_{\alpha\in\{0,1\}^d} 2^{d-|\alpha|} \int_{\mathbb{R}^d} \left| \sum_{\beta+\gamma=\alpha} \left(\frac{\partial}{\partial x}\right)^\beta f(x) \left(\frac{\partial}{\partial x}\right)^\gamma \chi_\Omega(x) \right| dx \\ &\leqslant \sum_{\alpha\in\{0,1\}^d} 2^{d-|\alpha|} \sum_{\beta+\gamma=\alpha} \int_{\Omega_{(1,\ldots,1)-\gamma}} \left| \left(\frac{\partial}{\partial x}\right)^\beta f(x) \right| dx. \end{aligned}$$

The desired estimate follows by setting $\tilde{\alpha} = (1,\ldots,1) - \alpha + \beta$. □

A homogeneity argument allows to simplify the total variation $\mathscr{V}_\Omega^*(f)$ in the above proposition. We shall present this argument in the more general context of integration over generic parallelepipeds.

Let Ω be any non degenerate compact parallelepiped in $\mathbb{R}^d$, let W be a $d \times d$ non singular real matrix taking the unit cube $[0,1]^d$ to a translated copy of Ω, and let $w_1,\ldots,w_d \in \mathbb{R}^d$ be its columns. For any multi-index $\alpha \in \{0,1\}^d$, define

$$\left(\frac{\partial}{\partial w}\right)^\alpha f(x) = \left(\frac{\partial}{\partial w_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial w_d}\right)^{\alpha_d} f(x),$$

where $\partial/\partial w_j = w_j \cdot \nabla$ are the directional derivatives, and define Ω_α as the union of all the $|\alpha|$-dimensional faces of Ω parallel to the directions $\alpha_1 w_1, \ldots, \alpha_d w_d$.

Theorem 2.2 *Let f be a smooth function on $\mathbb{R}^d$, Ω a compact parallelepiped, and $\mathscr{P} = \{z_j + m : j = 1, \ldots, N,\ m \in \mathbb{Z}^d\}$ a periodic distribution of points. Then*

$$\left| \int_\Omega f(x)\, dx - \frac{1}{N} \sum_{z \in \mathscr{P} \cap \Omega}^{*} f(z) \right| \leqslant \mathscr{D}(\Omega, \mathscr{P}) \mathscr{V}_\Omega(f),$$

where

$$\mathscr{D}(\Omega, \mathscr{P}) = \sup_{I \in \mathscr{I}} \left| \left| W(I) \cap \Omega \right| - \frac{1}{N} \sharp\bigl(W(I) \cap \Omega \cap \mathscr{P}\bigr) \right|,$$

is the discrepancy of $\mathscr{P}$ in Ω with respect to (periodic) parallelepipeds parallel to Ω, and

$$\mathscr{V}_\Omega(f) = \sum_{\alpha \in \{0,1\}^d} \frac{2^{d-|\alpha|}}{|\Omega_\alpha|} \int_{\Omega_\alpha} \left| \left(\frac{\partial}{\partial w} \right)^\alpha f(y) \right| dy$$

is the total variation of f in Ω. As before, the symbol $\sum_{z \in \mathscr{P} \cap \Omega}^{*} f(z)$ means that if z belongs to a j-dimensional face of the parallelepiped Ω, then the term $f(z)$ in the sum must be replaced by $2^{j-d} f(z)$. The integration over Ω_α is intended with respect to the $|\alpha|$-dimensional Lebesgue surface measure.

Observe that, since Ω_α is composed by exactly $2^{d-|\alpha|}$ faces, in the definition of total variation the integral is over all possible faces and it is normalized by dividing by the measure of these faces. Also observe that, while in Proposition 2.1 one integrates over the faces Ω_α all the derivatives of the function f of order $\beta \leqslant \alpha$, in this theorem the integration is only for $\beta = \alpha$.

Here we should emphasize that, via an affine transformation, we can reduce the above problem to an error estimate in a numerical integration over the unit square and then apply the original Koksma–Hlawka inequality. For simplicity, let us assume that $\Omega = W([0, 1]^d)$: then this procedure gives the estimate

$$\begin{aligned} &\left| \int_\Omega f(x)\, dx - \frac{|\Omega|}{n(\mathscr{P}, \Omega)} \sum_{z \in \mathscr{P} \cap W([0,1)^d)} f(z) \right| \\ &\leqslant \sup_I \left| \left| W(I) \right| - \frac{|\Omega|}{n(\mathscr{P}, \Omega)} \sharp\bigl(W(I) \cap \mathscr{P}\bigr) \right| \\ &\times \sum_{\alpha \in \{0,1\}^d, |\alpha| \neq 0} \frac{1}{|\Omega^\alpha|} \int_{\Omega^\alpha} \left| \left(\frac{\partial}{\partial w} \right)^\alpha f(x) \right| dx, \end{aligned} \tag{3}$$

where I is an interval of the form $[0, t_1] \times [0, t_2] \times \cdots \times [0, t_d]$ with $0 < t_k < 1$, $n(\mathscr{P}, \Omega)$ is the number of points of $\mathscr{P}$ contained in $W([0, 1)^d)$, and Ω^α is the

$|\alpha|$-dimensional face of Ω parallel to the directions $\alpha_1 w_1, \ldots, \alpha_d w_d$ containing the vertex $W(1, \ldots, 1)$. The disadvantage of (3) with respect to Theorem 2.2 is in the weight used in the Riemann sums. In Theorem 2.2 this weight is the inverse of N, which is the exact number of points of $\mathscr{P}$ per unit cube. In (3) the weight is the inverse of $n(\mathscr{P}, \Omega)/|\Omega|$, an extrapolation of the number of points per unit cube based on the number of points of $\mathscr{P}$ contained in Ω. As we will see later, in our application to simplices we will need a weight that is independent of the choice of the parallelepiped Ω.

Proof Without loss of generality, assume that $\Omega = W([0, 1]^d)$. For any integer $m \geqslant 2$, define the matrix $V = mW$. Observe that $\widetilde{\Omega} := V^{-1}(\Omega) = [0, m^{-1}]^d$. Also, define the function $\tilde{f}(x) = f(Vx)$. Thus, the restriction to $\widetilde{\Omega}$ of $\tilde{f}$ is an "affine image" of the restriction to Ω of the function f. Finally let $\widetilde{\mathscr{P}}$ be the periodic distribution of points obtained by a periodic extension of the set $(V^{-1}(\mathscr{P})) \cap [0, 1)^d$. Call $n(\widetilde{\mathscr{P}})$ the cardinality of the set $(V^{-1}(\mathscr{P})) \cap [0, 1)^d$. Then

$$\begin{aligned}
&\left|\int_{\Omega} f(x)\,dx - \frac{1}{N} \sum_{z\in\Omega\cap\mathscr{P}}^{*} f(z)\right| \\
&= \left|\int_{\widetilde{\Omega}} f\big(V(y)\big)|\det V|\,dy - \frac{1}{N} \sum_{z\in\widetilde{\Omega}\cap V^{-1}(\mathscr{P})}^{*} f\big(V(z)\big)\right| \\
&= |\det V| \left|\int_{\widetilde{\Omega}} \tilde{f}(y)\,dy - \frac{1}{N|\det V|} \sum_{z\in\widetilde{\Omega}\cap\widetilde{\mathscr{P}}}^{*} \tilde{f}(z)\right|.
\end{aligned}$$

In the last two lines above, the $*$ symbol in the summation signs refers to the faces of the cube $\widetilde{\Omega}$. Observe that we cannot immediately apply Proposition 2.1 to the last line in the above identities, since $N|\det V|$ may be different from $n(\widetilde{\mathscr{P}})$. Anyhow, Proposition 2.1 gives

$$\begin{aligned}
&\left|\int_{\Omega} f(x)\,dx - \frac{1}{N} \sum_{z\in\Omega\cap\mathscr{P}}^{*} f(z)\right| \\
&\leqslant |\det V| \left|\int_{\widetilde{\Omega}} \tilde{f}(y)\,dy - \frac{1}{n(\widetilde{\mathscr{P}})} \sum_{z\in\widetilde{\Omega}\cap\widetilde{\mathscr{P}}}^{*} \tilde{f}(z)\right| \\
&\quad + |\det V| \left|\left(\frac{1}{n(\widetilde{\mathscr{P}})} - \frac{1}{N|\det V|}\right) \sum_{z\in\widetilde{\Omega}\cap\widetilde{\mathscr{P}}}^{*} \tilde{f}(z)\right| \\
&\leqslant |\det V| \mathscr{D}_{\mathscr{I}}(\widetilde{\Omega}, \widetilde{\mathscr{P}}) \mathscr{V}_{\widetilde{\Omega}}^{*}(\tilde{f}) \\
&\quad + |\det V| \left|\left(\frac{1}{n(\widetilde{\mathscr{P}})} - \frac{1}{N|\det V|}\right) \sum_{z\in\widetilde{\Omega}\cap\widetilde{\mathscr{P}}}^{*} \tilde{f}(z)\right|.
\end{aligned}$$

It turns out that the last term is negligible. Indeed,

$$
\begin{aligned}
n(\widetilde{\mathscr{P}}) &= \sharp\big(\big(V^{-1}(\mathscr{P})\big) \cap [0,1)^d\big) = \sum_{j=1}^{N} \sum_{k\in\mathbb{Z}^d} \chi_{[0,1)^d}\big(V^{-1}(z_j+k)\big) \\
&= \sum_{j=1}^{N} \sum_{k\in\mathbb{Z}^d} \chi_{V([0,1)^d)}(z_j+k) = \sharp\big(\mathscr{P} \cap V\big([0,1)^d\big)\big) \\
&= m^d N|\Omega| + \text{error term.}
\end{aligned}
$$

The error term is controlled by N times the number of unit cubes intersecting the boundary of $m\Omega$, thus

$$
\text{error term} = \mathscr{O}\big(Nm^{d-1}\big).
$$

It follows that

$$
\begin{aligned}
&|\det V| \left| \left(\frac{1}{n(\widetilde{\mathscr{P}})} - \frac{1}{N|\det V|} \right) \sum_{z\in\widetilde{\Omega}\cap\widetilde{\mathscr{P}}}^{*} \tilde{f}(z) \right| \\
&\quad \leqslant \left| \left(\frac{m^d|\det W|}{Nm^d|\det W| + \mathscr{O}(Nm^{d-1})} - \frac{1}{N} \right) \right| \left| \sum_{z\in\mathscr{P}\cap\Omega}^{*} f(z) \right| \\
&\quad = \frac{\mathscr{O}(m^{-1})}{N} \left| \sum_{z\in\mathscr{P}\cap\Omega}^{*} f(z) \right|.
\end{aligned}
$$

The right hand side tends to 0 as $m \to +\infty$. Thus we have

$$
\left| \int_\Omega f(x)\,dx - \frac{1}{N} \sum_{z\in\mathscr{P}\cap\Omega}^{*} f(z) \right| \leqslant \lim_{m\to+\infty} |\det V| \mathscr{D}_{\mathscr{I}}(\widetilde{\Omega}, \widetilde{\mathscr{P}}) \mathscr{V}^{*}_{\widetilde{\Omega}}(\tilde{f}).
$$

Consider first the discrepancy factor:

$$
\begin{aligned}
&|\det V| \mathscr{D}_{\mathscr{I}}(\widetilde{\Omega}, \widetilde{\mathscr{P}}) \\
&\quad = |\det V| \sup_{I\subset\mathscr{I}} \left| |I\cap\widetilde{\Omega}| - \frac{1}{n(\widetilde{\mathscr{P}})} \sharp(I\cap\widetilde{\Omega}\cap\widetilde{\mathscr{P}}) \right| \\
&\quad = \sup_{I\subset\mathscr{I}} \left| \big|V(I)\cap\Omega\big| - \frac{|\det V|}{N|\det V| + \mathscr{O}(Nm^{d-1})} \sharp\big(V(I)\cap\Omega\cap\mathscr{P}\big) \right|.
\end{aligned}
$$

Note that, in general, $V(\widetilde{\mathscr{P}})$ does not coincide with $\mathscr{P}$, but the above identity holds because $V(\widetilde{\Omega}\cap\widetilde{\mathscr{P}}) = \Omega\cap\mathscr{P}$. Thus, proceeding as before,

$$|\det V|\mathscr{D}_{\mathscr{I}}(\widetilde{\Omega},\widetilde{\mathscr{P}}) \leqslant \sup_{I\subset\mathscr{I}}\left||V(I)\cap\Omega| - \frac{1}{N}\sharp\big(V(I)\cap\Omega\cap\mathscr{P}\big)\right| + \mathscr{O}\big(m^{-1}\big)\sup_{I\subset\mathscr{I}}\frac{1}{N}\sharp\big(V(I)\cap\Omega\cap\mathscr{P}\big).$$

Since

$$\sup_{I\subset\mathscr{I}}\frac{1}{N}\sharp\big(V(I)\cap\Omega\cap\mathscr{P}\big) \leqslant C|\Omega|,$$

then

$$\lim_{m\to+\infty}|\det V|\mathscr{D}_{\mathscr{I}}(\widetilde{\Omega},\widetilde{\mathscr{P}}) = \lim_{m\to+\infty}\sup_{I\subset\mathscr{I}}\left||V(I)\cap\Omega| - \frac{1}{N}\sharp\big(V(I)\cap\Omega\cap\mathscr{P}\big)\right| = \sup_{I\subset\mathscr{I}}\left||W(I)\cap\Omega| - \frac{1}{N}\sharp\big(W(I)\cap\Omega\cap\mathscr{P}\big)\right|.$$

The last identity follows from the fact that, for every positive integer m, the collection of sets $V(I)\cap\Omega$ coincides with the collection of sets $W(I)\cap\Omega$. Finally, let us study the variation

$$\begin{aligned}\mathscr{V}^*_{\widetilde{\Omega}}(\tilde{f}) &= \sum_{\alpha\in\{0,1\}^d}\sum_{\beta\leqslant\alpha}2^{|\alpha|-|\beta|}\int_{\widetilde{\Omega}_\alpha}\left|\left(\frac{\partial}{\partial x}\right)^\beta(f\circ V)(x)\right|dx\\ &= \sum_{\alpha\in\{0,1\}^d}\sum_{\beta\leqslant\alpha}2^{|\alpha|-|\beta|}\frac{|\widetilde{\Omega}_\alpha|}{|\Omega_\alpha|}\int_{\Omega_\alpha}\left|\left(\frac{\partial}{\partial x}\right)^\beta(f\circ V)\big(V^{-1}(y)\big)\right|dy\\ &= \sum_{\alpha\in\{0,1\}^d}\sum_{\beta\leqslant\alpha}2^{|\alpha|-|\beta|}\frac{2^{d-|\alpha|}m^{-|\alpha|}}{|\Omega_\alpha|}\int_{\Omega_\alpha}\left|\left(\frac{\partial}{\partial v}\right)^\beta f(y)\right|dy\\ &= \sum_{\alpha\in\{0,1\}^d}\sum_{\beta\leqslant\alpha}2^{|\alpha|-|\beta|}\frac{2^{d-|\alpha|}m^{|\beta|-|\alpha|}}{|\Omega_\alpha|}\int_{\Omega_\alpha}\left|\left(\frac{\partial}{\partial w}\right)^\beta f(y)\right|dy.\end{aligned}$$

Finally, when $m\to+\infty$, all the terms in the innermost sum vanish, with the exception of the term with $\beta=\alpha$. Thus,

$$\lim_{m\to+\infty}\mathscr{V}^*_{\widetilde{\Omega}}(\tilde{f}) = \sum_{\alpha\in\{0,1\}^d}\frac{2^{d-|\alpha|}}{|\Omega_\alpha|}\int_{\Omega_\alpha}\left|\left(\frac{\partial}{\partial w}\right)^\alpha f(y)\right|dy.$$

□

3 Simplices

Our last variant of the Koksma–Hlawka inequality refers to simplices. Let now S be a closed simplex in $\mathbb{R}^d$, and let $V_0,\ldots,V_d$ be its vertices. For any $k=0,\ldots,d$, let

$w_1^k, \ldots, w_d^k$, be the vectors joining the vertex V_k with the other vertices, in whatever order. Call W_k the matrix with columns $w_1^k, \ldots, w_d^k$. Let Ω_k be the parallelepiped determined by the vertex V_k and the vectors $w_1^k, \ldots, w_d^k$. Finally, for every multi-index $\alpha \in \{0,1\}^d$, let S_α^k be the (unique) $|\alpha|$-dimensional face of S parallel to the directions $\alpha_1 w_1^k, \ldots, \alpha_d w_d^k$. In order to deduce a Koksma–Hlawka inequality for simplices from the Koksma–Hlawka inequality for parallelepipeds, it suffices to decompose the characteristic function of the simplex S into a weighted sum of characteristic functions of the parallelepipeds Ω_k.

Lemma 3.1 *There exists a constant C_d, depending only on the dimension d, such that for every simplex S there exist smooth functions $\phi_0, \ldots, \phi_d$ satisfying the following conditions*:

(i) *For every $k = 0, \ldots, d$, we have $\phi_k(V_k) = 1$, and the support of ϕ_k is contained in the open half space determined by the facet of S opposite to V_k.*
(ii) *$\sum_{k=0}^d \phi_k(x) = 1$ for every $x \in S$.*
(iii) *For all $k = 0, \ldots, d$, and for all multiindices $\alpha \in \{0,1\}^d$,*

$$\sup_{x \in S} \left| \left(\frac{\partial}{\partial w^k} \right)^\alpha \phi_k(x) \right| \leqslant C_d.$$

Proof When S is the standard simplex, the lemma follows from a simple partition of unit argument. An affine transformation takes the general simplex onto the standard simplex, without changing the norms in point (iii). □

Theorem 3.2 *Let f be a smooth function on $\mathbb{R}^d$, let S be a compact simplex, and let $\mathscr{P} = \{z_j + m : j = 1, \ldots, N,\ m \in \mathbb{Z}^d\}$ be a periodic distribution of points. Then*

$$\left| \int_S f(x)\, dx - \frac{1}{N} \sum_{z \in \mathscr{P} \cap S}^{*} f(z) \right| \leqslant \mathscr{D}(S, \mathscr{P}) \mathscr{V}_S(f),$$

where

$$\mathscr{D}(S, \mathscr{P}) = \max_{k=0,\ldots,d} \mathscr{D}(\Omega_k, \mathscr{P})$$

can be defined as the discrepancy of $\mathscr{P}$ with respect to the $d+1$ parallelepipeds associated with the simplex S, and

$$\mathscr{V}_S(f) = C_d \sum_{k=0}^{d} \sum_{\alpha \in \{0,1\}^d} \sum_{\beta \leqslant \alpha} \frac{1}{|S_\alpha^k|} \int_{S_\alpha^k} \left| \left(\frac{\partial}{\partial w^k} \right)^\beta f(x) \right| dx$$

is the total variation of f in the simplex S. As before, the symbol $\sum_{z \in \mathscr{P} \cap \Omega}^{*} f(z)$ means that if z belongs to a j-dimensional face of the simplex S, then the term $f(z)$ in the sum must be replaced by $2^{j-d} f(z)$. The integration over S_α^k is intended with respect to the $|\alpha|$-dimensional Lebesgue surface measure. Finally, a multi-index β is less than or equal to another multi-index α if $\beta_j \leqslant \alpha_j$ for any $j = 1, \ldots, d$.

Proof By a partition of unit as in the previous lemma, we can write

$$\begin{aligned}
&\left|\int_S f(x)\,dx - \frac{1}{N}\sum_{z\in\mathscr{P}\cap S}^{*} f(z)\right| \\
&= \left|\sum_{k=0}^{d}\left(\int_S f(x)\phi_k(x)\,dx - \frac{1}{N}\sum_{z\in\mathscr{P}\cap S}^{*} f(z)\phi_k(z)\right)\right| \\
&\leqslant \sum_{k=0}^{d}\left|\int_{\Omega_k} f(x)\phi_k(x)\,dx - \frac{1}{N}\sum_{z\in\mathscr{P}\cap\Omega_k}^{*} f(z)\phi_k(z)\right|.
\end{aligned}$$

By Theorem 2.2, each term of the above sum is bounded by

$$\begin{aligned}
&\sup_{I\in\mathscr{I}}\left|\left|W_k(I)\cap\Omega_k\right| - \frac{1}{N}\sharp\big(W_k(I)\cap\Omega_k\cap\mathscr{P}\big)\right| \\
&\quad\times\sum_{\alpha\in\{0,1\}^d}\frac{2^{d-|\alpha|}}{|\Omega_\alpha^k|}\int_{\Omega_\alpha^k}\left|\left(\frac{\partial}{\partial w^k}\right)^{\alpha}(f\phi_k)(x)\right|dx,
\end{aligned}$$

where Ω_α^k is the union of all the $|\alpha|$-dimensional faces of Ω_k parallel to the directions $\alpha_1 w_1^k,\dots,\alpha_d w_d^k$. In the above sum, the term corresponding to $\alpha=(0,\dots,0)$ is just $|f(V_k)|$. When $|\alpha|\neq 0$, by the definition of the functions ϕ_k, the above integrals over the faces of the parallelepipeds can be replaced by the integrals over the faces of the simplex,

$$\frac{2^{d-|\alpha|}}{|\Omega_\alpha^k|}\int_{\Omega_\alpha^k}\left|\left(\frac{\partial}{\partial w^k}\right)^{\alpha}(f\phi_k)(x)\right|dx = \frac{1}{|\alpha||S_\alpha^k|}\int_{S_\alpha^k}\left|\left(\frac{\partial}{\partial w^k}\right)^{\alpha}(f\phi_k)(x)\right|dx.$$

Finally,

$$\left(\frac{\partial}{\partial w^k}\right)^{\alpha}(f\phi_k)(x) = \sum_{\beta+\gamma=\alpha}\left(\frac{\partial}{\partial w^k}\right)^{\beta} f(x)\left(\frac{\partial}{\partial w^k}\right)^{\gamma}\phi_k(x).$$

Hence, by the previous lemma,

$$\begin{aligned}
&\sum_{\alpha\in\{0,1\}^d}\frac{2^{d-|\alpha|}}{|\Omega_\alpha^k|}\int_{\Omega_\alpha^k}\left|\left(\frac{\partial}{\partial w^k}\right)^{\alpha} f(y)\right|dy \\
&\quad\leqslant \left|f(V_k)\right| + C_d\sum_{\substack{\alpha\in\{0,1\}^d\\ |\alpha|\neq 0}}\sum_{\beta\leqslant\alpha}\frac{1}{|\alpha||S_\alpha^k|}\int_{S_\alpha^k}\left|\left(\frac{\partial}{\partial w^k}\right)^{\beta} f(x)\right|dx.
\end{aligned}$$

□

As an example, let us write explicitly the total variation $\mathscr{V}_S(f)$ in the 2-dimensional case. Let S be a triangle with vertices V_1, V_2 and V_3. Call l_k the length of the edge S_k opposite to V_k, and w_k the vector joining the two vertices opposite to V_k. Then the variation is

$$\begin{aligned}\mathscr{V}_S(f) = {} & C_2\left|f(V_1)\right| + C_2\left|f(V_2)\right| + C_2\left|f(V_3)\right| \\ & + C_2\frac{2}{l_1}\int_{S_1}\left(\left|f(x)\right| + \left|\frac{\partial f}{\partial w_1}(x)\right|\right)dx \\ & + C_2\frac{2}{l_2}\int_{S_2}\left(\left|f(x)\right| + \left|\frac{\partial f}{\partial w_2}(x)\right|\right)dx \\ & + C_2\frac{2}{l_3}\int_{S_3}\left(\left|f(x)\right| + \left|\frac{\partial f}{\partial w_3}(x)\right|\right)dx \\ & + C_2\frac{1}{|S|}\int_S\left(3\left|f(x)\right| + 2\left|\frac{\partial f}{\partial w_1}(x)\right| + 2\left|\frac{\partial f}{\partial w_2}(x)\right| + 2\left|\frac{\partial f}{\partial w_3}(x)\right|\right. \\ & \left. + \left|\frac{\partial^2 f}{\partial w_2 \partial w_3}(x)\right| + \left|\frac{\partial^2 f}{\partial w_1 \partial w_3}(x)\right| + \left|\frac{\partial^2 f}{\partial w_1 \partial w_2}(x)\right|\right)dx.\end{aligned}$$

References

1. Brandolini, L., Colzani, L., Gigante, G., Travaglini, G.: On the Koksma–Hlawka inequality. J. Complex, submitted
2. Harman, G.: Variations on the Koksma–Hlawka inequality. Unif. Distrib. Theory **5**, 65–78 (2010)
3. Hickernell, F.J.: Koksma–Hlawka inequality. In: Kotz, S., Read, C.B., Banks, D.L. (eds.): Encyclopedia of Statistical Sciences. Wiley-Interscience, New York (2006)
4. Hlawka, E.: The Theory of Uniform Distribution. A B Academic Publishers, Berkhamsted (1984)
5. Kuipers, L., Niederreiter, H.: Uniform Distribution of Sequences. Dover, New York (2006)
6. Matoušek, J.: Geometric Discrepancy: An Illustrated Guide. Springer, Berlin (2010)
7. Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods. SIAM, Philadelphia (1992)
8. Zaremba, S.K.: Some applications of multidimensional integration by parts. Ann. Pol. Math. **21**, 85–96 (1968)

A Dual Interpretation of the Gromov–Thurston Proof of Mostow Rigidity and Volume Rigidity for Representations of Hyperbolic Lattices

Michelle Bucher, Marc Burger, and Alessandra Iozzi

Abstract We use bounded cohomology to define a notion of volume of an $\mathrm{SO}(n,1)$-valued representation of a lattice $\Gamma < \mathrm{SO}(n,1)$ and, using this tool, we give a complete proof of the volume rigidity theorem of Francaviglia and Klaff (Geom. Dedicata 117, 111–124 (2006)) in this setting. Our approach gives in particular a proof of Thurston's version of Gromov's proof of Mostow Rigidity (also in the non-cocompact case), which is dual to the Gromov–Thurston proof using the simplicial volume invariant.

Keywords Rigidity · Real hyperbolic manifold · Bounded cohomology · Maximal representations

Mathematics Subject Classification (2010) 53C24 · 22E40 · 22E41

1 Introduction

Strong rigidity of lattices was proved in 1965 by Mostow [28] who, while searching for a geometric explanation of the deformation rigidity results obtained by Selberg [32], Calabi–Vesentini [14, 15] and Weil [35, 36], showed the remarkable fact that, under some conditions, topological data of a manifold determine its metric. Namely, he proved that if $M_i = \Gamma_i \backslash \mathbb{H}^n$, $i = 1, 2$ are compact quotients of real hyperbolic n-space and $n \geqslant 3$, then any homotopy equivalence $\varphi : M_1 \to M_2$ is, up to homotopy, induced by an isometry. Shortly thereafter, this was extended to the finite volume case by G. Prasad [29].

M. Bucher
Section de Mathématiques, Université de Genève, 2-4 rue du Lièvre, Case postale 64, 1211 Geneva 4, Switzerland
e-mail: Michelle.Bucher-Karlsson@unige.ch

M. Burger · A. Iozzi (✉)
Department Mathematik, ETH, Rämistrasse 101, 8092 Zurich, Switzerland
e-mail: iozzi@math.ethz.ch

M. Burger
e-mail: burger@math.ethz.ch

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3, DOI 10.1007/978-88-470-2853-1_4,

The methods introduced by Mostow emphasized the role of the quasi-isometries of $\widetilde{M}_i = \mathbb{H}^n$, their quasi-conformal extension to $\partial\mathbb{H}^n$, ergodicity phenomena of the Γ_i-action on $\partial\mathbb{H}^n$, as well as almost everywhere differentiability results *à la* Egorov.

In the 1970's, a new approach for rigidity in the real hyperbolic case was developed by Gromov. In this context he introduced ℓ^1-homology and the simplicial volume: techniques like smearing and straightening became important. This approach was then further developed by Thurston [33, Chap. 6] and one of its consequences is an extension to hyperbolic manifolds of Kneser's theorem for surfaces [25]. To wit, the computation of the simplicial volume $\|M\| = \mathrm{Vol}(M)/v_n$ implies, for a continuous map $f : M_1 \to M_2$ between compact real hyperbolic manifolds, that

$$\deg f \leqslant \frac{\mathrm{Vol}(M_2)}{\mathrm{Vol}(M_1)}.$$

If $\dim M_i \geqslant 3$, Thurston proved that equality holds if and only if f is homotopic to an isometric covering while the topological assertion in the case in which $\dim M_i = 2$ is Kneser's theorem [25].

The next step, in the spirit of Goldman's theorem [20]—what now goes under the theory of *maximal representations*—is to associate an invariant $\mathrm{Vol}(\rho)$ to an arbitrary representation

$$\rho : \pi_1(M) \to \mathrm{Isom}\big(\mathbb{H}^n\big)$$

of the fundamental group of M, satisfying a Milnor–Wood type inequality

$$\mathrm{Vol}(\rho) \leqslant \mathrm{Vol}(i).$$

The equality should be characterized as given by the "unique" lattice embedding i of $\pi_1(M)$, of course provided $\dim M \geqslant 3$. This was carried out in $\dim M = 3$ by Dunfield [17], following Toledo's modification of the Gromov–Thurston approach to rigidity [34].

If M is only of finite volume, a technical difficulty is the definition of the volume $\mathrm{Vol}(\rho)$ of a representation. Dunfield introduced for this purpose the notion of pseudodeveloping map and Francaviglia proved that the definition is independent of the choice of the pseudodeveloping map [18]. Then Francaviglia and Klaff [19] proved a "volume rigidity theorem" for representations

$$\rho : \pi_1(M) \to \mathrm{Isom}\big(\mathbb{H}^k\big),$$

where now k is not necessarily equal to $\dim M$. In their paper, the authors actually succeed in applying the technology developed by Besson–Courtois–Gallot in their seminal work on entropy rigidity [2]. An extension to representations of $\pi_1(M)$ into $\mathrm{Isom}(\mathbb{H}^n)$ for an arbitrary compact manifold M has been given by Besson–Courtois–Gallot [3].

Finally, Bader, Furman and Sauer proved a generalization of Mostow Rigidity for cocycles in the case of real hyperbolic lattices with some integrability condition, using, among others, bounded cohomology techniques, [1].

The aim of this paper is to give a complete proof of volume rigidity from the point of view of bounded cohomology, implementing a strategy first described in [24] and used in the work on maximal representations of surface groups [12, 13], as well as in the proof of Mostow Rigidity in dimension 3 in [11].

Our main contribution consists on the one hand in identifying the top dimensional bounded equivariant cohomology of the full group of isometries $\mathrm{Isom}(\mathbb{H}^n)$, and on the other in giving a new definition of the volume of a representation of $\pi_1(M)$, when M is not compact; this definition, that uses bounded relative cohomology, generalizes the one introduced in [13] for surfaces.

In an attempt to be pedagogical, throughout the paper we try to describe, in varying details, the proof of all results.

Let $\mathrm{Vol}_n(x_0, \dots, x_n)$ denote the signed volume of the convex hull of the points $x_0, \dots, x_n \in \overline{\mathbb{H}^n}$. Then Vol_n is a $G^+ := \mathrm{Isom}^+(\mathbb{H}^n)$-invariant cocycle on $\overline{\mathbb{H}^n}$ and hence defines a top dimensional cohomology class $\omega_n \in H_c^n(G^+, \mathbf{R})$. Let $i : \Gamma \hookrightarrow G^+$ be an embedding of Γ as a lattice in the group of orientation preserving isometries of $\mathbb{H}^n$ and let $\rho : \Gamma \to G^+$ be an arbitrary representation of Γ. Suppose first that Γ is torsion free. Recall that the cohomology of Γ is canonically isomorphic to the cohomology of the n-dimensional quotient manifold $M := i(\Gamma)\backslash\mathbb{H}^n$.

If M is compact, by Poincaré duality the cohomology groups $H^n(\Gamma, \mathbf{R}) \cong H^n(M, \mathbf{R})$ in top dimension are canonically isomorphic to $\mathbf{R}$, with the isomorphism given by the evaluation on the fundamental class $[M]$. We define the volume $\mathrm{Vol}(\rho)$ of ρ by

$$\mathrm{Vol}(\rho) = \big\langle \rho^*(\omega_n), [M] \big\rangle,$$

where $\rho^* : H_c^n(G^+, \mathbf{R}) \to H^n(\Gamma, \mathbf{R})$ denotes the pull-back via ρ. In particular the absolute value of the volume of the lattice embedding i is equal to the volume of the hyperbolic manifold M, $\mathrm{Vol}(M) = \langle i^*(\omega_n), [M] \rangle$.

If M is not compact, the above definition fails since $H^n(\Gamma, \mathbf{R}) \cong H^n(M, \mathbf{R}) = 0$. Thus we propose the following approach: since Vol_n is in fact a bounded cocycle, it defines a bounded class $\omega_n^b \in H_{b,c}^n(G^+, \mathbf{R})$ in the bounded cohomology of G^+ with trivial $\mathbf{R}$-coefficients. Thus associated to a homomorphism $\rho : \Gamma \to G^+$ we obtain $\rho^*(\omega_n^b) \in H_b^n(\Gamma, \mathbf{R})$; since $\widetilde{M} = \mathbb{H}^n$ is contractible, it follows easily that $H_b^n(\Gamma, \mathbf{R})$ is canonically isomorphic to the bounded singular cohomology $H_b^n(M, \mathbf{R})$ of the manifold M (this is true in much greater generality [5, 21], but it will not be used here). To proceed further, let $N \subset M$ be a compact core of M, that is the complement in M of a disjoint union of finitely many horocyclic neighborhoods E_i, $i = 1, \dots, k$, of cusps. Those have amenable fundamental groups and thus the map $(N, \partial N) \to (M, \emptyset)$ induces an isomorphism in cohomology, $H_b^n(N, \partial N, \mathbf{R}) \cong H_b^n(M, \mathbf{R})$, by means of which we can consider $\rho^*(\omega_n^b)$ as a bounded relative class. Finally, the image of $\rho^*(\omega_n^b)$ via the comparison map $c : H_b^n(N, \partial N, \mathbf{R}) \to H^n(N, \partial N, \mathbf{R})$ is an ordinary relative class whose evaluation on the relative fundamental class $[N, \partial N]$ gives the definition of the volume of ρ,

$$\mathrm{Vol}(\rho) := \big\langle \big(c \circ \rho^*\big)\big(\omega_n^b\big), [N, \partial N] \big\rangle,$$

which turns out to be independent of the choice of the compact core N. When M is compact, we recover of course the invariant previously defined. We complete the definition in the case in which Γ has torsion by setting

$$\mathrm{Vol}(\rho) := \frac{\mathrm{Vol}(\rho|_\Lambda)}{[\Gamma : \Lambda]}$$

where $\Lambda < \Gamma$ is a torsion free subgroup of finite index.

Theorem 1.1 *Let $n \geqslant 3$. Let $i : \Gamma \hookrightarrow \mathrm{Isom}^+(\mathbb{H}^n)$ be a lattice embedding and let $\rho : \Gamma \to \mathrm{Isom}^+(\mathbb{H}^n)$ be any representation. Then*

$$\big|\mathrm{Vol}(\rho)\big| \leqslant \big|\mathrm{Vol}(i)\big| = \mathrm{Vol}(M), \tag{1}$$

with equality if and only if ρ is conjugated to i by an isometry.

An analogous theorem, in the more general case of a representation $\rho : \Gamma \to \mathrm{Isom}^+(\mathbb{H}^m)$ with $m \geqslant n$, has been proven by Francaviglia and Klaff [19] with a different definition of volume.

Taking in particular ρ to be another lattice embedding of Γ, we recover Mostow–Prasad Rigidity theorem for hyperbolic lattices:

Corollary 1.2 [28, 29] *Let Γ_1, Γ_2 be two isomorphic lattices in $\mathrm{Isom}^+(\mathbb{H}^n)$. Then there exists an isometry $g \in \mathrm{Isom}(\mathbb{H}^n)$ conjugating Γ_1 to Γ_2.*

As a consequence of Theorem 1.1, we also reprove Thurston's strict version of Gromov's degree inequality for hyperbolic manifolds. Note that this strict version generalizes Mostow Rigidity [33, Theorem 6.4]:

Corollary 1.3 [33, Theorem 6.4] *Let $f : M_1 \to M_2$ be a continuous proper map between two n-dimensional complete finite volume hyperbolic manifolds M_1 and M_2 with $n \geqslant 3$. Then*

$$\deg(f) \leqslant \frac{\mathrm{Vol}(M_2)}{\mathrm{Vol}(M_1)},$$

with equality if and only if f is homotopic to a local isometry.

Our proof of Theorem 1.1 follows closely the steps in the proof of Mostow Rigidity. In particular, the following result is the dual to the use of measure homology and smearing in [33]. We denote by $\varepsilon : G \to \{-1, 1\}$ the homomorphism defined by $\varepsilon(g) = 1$ if g is orientation preserving and $\varepsilon(g) = -1$ if g is orientation reversing.

Theorem 1.4 *Let $M = \Gamma\backslash\mathbb{H}^n$ be a finite volume real hyperbolic manifold. Let $\rho : \Gamma \to \mathrm{Isom}(\mathbb{H}^n)$ be a representation with non-elementary image and let $\varphi : \partial\mathbb{H}^n \to$*

$\partial\mathbb{H}^n$ be the corresponding equivariant measurable map. Then for every $(n+1)$-tuple of points $\xi_0, \ldots, \xi_n \in \partial\mathbb{H}^n$,

$$\int_{\Gamma\backslash \operatorname{Isom}(\mathbb{H}^n)} \varepsilon\big(\dot{g}^{-1}\big) \operatorname{Vol}_n\big(\varphi(\dot{g}\xi_0), \ldots, \varphi(\dot{g}\xi_n)\big)\, d\mu(\dot{g}) = \frac{\operatorname{Vol}(\rho)}{\operatorname{Vol}(M)} \operatorname{Vol}_n(\xi_0, \ldots, \xi_n), \tag{2}$$

where μ is the invariant probability measure on $\Gamma\backslash \operatorname{Isom}(\mathbb{H}^n)$.

This allows us to deduce strong rigidity properties of the boundary map φ from the cohomological information about the boundary that, in turn, are sufficient to show the existence of an element $g \in \operatorname{Isom}^+(\mathbb{H}^n)$ conjugating ρ and i.

To establish the theorem, we first prove the almost everywhere validity of the formula in Theorem 1.4. Ideally, we would need to know that $H^n_{b,c}(G^+, \mathbf{R})$ is 1-dimensional and has no coboundaries in degree n in the appropriate cocomplex. However in general we do not know how to compute $H^n_{b,c}(G^+, \mathbf{R})$, except when $G^+ = \operatorname{Isom}^+(\mathbb{H}^2)$ or $\operatorname{Isom}^+(\mathbb{H}^3)$ and hence there is no direct way to prove the formula in (2). To circumvent this problem, we borrow from [7] (see also [9]) the essential observation that Vol_n is in fact a cocycle equivariant with respect to the full group of isometries $G = \operatorname{Isom}(\mathbb{H}^n)$, that is,

$$\operatorname{Vol}_n(gx_1, \ldots, gx_n) = \varepsilon(g) \operatorname{Vol}_n(x_1, \ldots, x_n).$$

This leads to consider $\mathbf{R}$ as a non-trivial coefficient module $\mathbf{R}_\varepsilon$ for G and in this context we prove that the comparison map

$$H^n_{b,c}(G, \mathbf{R}_\varepsilon) \xrightarrow{\ \cong\ } H^n_c(G, \mathbf{R}_\varepsilon)$$

is an isomorphism. By a slight abuse of notation, we denote again by $\omega^b_n \in H^n_{b,c}(G, \mathbf{R}_\varepsilon)$ and by $\omega_n \in H^n_c(G, \mathbf{R}_\varepsilon)$ the generator defined by Vol_n.

Using this identification and standard tools from the homological algebra approach to bounded cohomology, we obtain the almost everywhere validity of the formula in Theorem 1.4. Additional arguments involving Lusin's theorem are required to establish the formula pointwise. This is essential because one step of the proof (see the beginning of Sect. 4) consists in showing that, if there is the equality in (1), the map φ maps the vertices of almost every positively oriented maximal ideal simplex to vertices of positively (or negatively—one or the other, not both) oriented maximal ideal simplices. Since such vertices form a set of measure zero in the boundary, an almost everywhere statement would not be sufficient.

2 The Continuous Bounded Cohomology of $G = \operatorname{Isom}(\mathbb{H}^n)$

Denote by $G = \operatorname{Isom}(\mathbb{H}^n)$ the full isometry group of hyperbolic n-space, and by $G^+ = \operatorname{Isom}^+(\mathbb{H}^n)$ its subgroup of index 2 consisting of orientation preserving isometries. As remarked in the introduction there are two natural G-module structures on $\mathbf{R}$: the trivial one, which we denote by $\mathbf{R}$, and the one given by multiplication with the homomorphism $\varepsilon : G \to G/G^+ \cong \{+1, -1\}$, which we denote by $\mathbf{R}_\varepsilon$.

Recall that if $q \in \mathbb{N}$, the continuous cohomology groups $H_c^q(G, \mathbf{R})$, respectively $H_c^q(G, \mathbf{R}_\varepsilon)$—or in short $H_c^\bullet(G, \mathbf{R}_{(\varepsilon)})$ for both—of G with coefficient in $\mathbf{R}_{(\varepsilon)}$, is by definition given as the cohomology of the cocomplex

$$C_c\big(G^{q+1}, \mathbf{R}_{(\varepsilon)}\big)^G = \big\{ f : G^{q+1} \to \mathbf{R}_{(\varepsilon)} \,\big|\, f \text{ is continuous and } \varepsilon(g) \cdot f(g_0, \dots, g_q) = f(gg_0, \dots, gg_q) \big\}$$

endowed with its usual homogeneous coboundary operator

$$\delta : C_c\big(G^{q+1}, \mathbf{R}_{(\varepsilon)}\big)^G \to C_c\big(G^{q+2}, \mathbf{R}_{(\varepsilon)}\big)^G$$

defined by

$$\delta f(g_0, \dots, g_{q+1}) := \sum_{j=0}^{q+1} (-1)^j f(g_0, \dots, g_{j-1}, g_{j+1}, \dots, g_{q+1}).$$

This operator clearly restricts to the bounded cochains

$$C_{c,b}\big(G^{q+1}, \mathbf{R}_{(\varepsilon)}\big)^G = \Big\{ f \in C_c\big(G^{q+1}, \mathbf{R}_{(\varepsilon)}\big)^G \,\Big|\, \|f\|_\infty = \sup_{g_0, \dots, g_q \in G} \big|f(g_0, \dots, g_q)\big| < +\infty \Big\}$$

and the continuous bounded cohomology $H_{c,b}^q(G, \mathbf{R}_{(\varepsilon)})$ of G with coefficients in $\mathbf{R}_{(\varepsilon)}$ is the cohomology of this cocomplex. The inclusion

$$C_{c,b}\big(G^{q+1}, \mathbf{R}_{(\varepsilon)}\big)^G \subset C_c\big(G^{q+1}, \mathbf{R}_{(\varepsilon)}\big)^G$$

induces a comparison map

$$c : H_{c,b}^q(G, \mathbf{R}_{(\varepsilon)}) \longrightarrow H_c^q(G, \mathbf{R}_{(\varepsilon)}).$$

We call cochains in $C_{c,(b)}(G^{q+1}, \mathbf{R})^G$ invariant and cochains in $C_{c,(b)}(G^{q+1}, \mathbf{R}_\varepsilon)^G$ equivariant and apply this terminology to the cohomology classes as well. The sup norm on the complex of cochains induces a seminorm in cohomology

$$\|\beta\| = \inf\big\{ \|f\|_\infty \,\big|\, f \in C_{c,(b)}\big(G^{q+1}, \mathbf{R}_{(\varepsilon)}\big)^G, [f] = \beta \big\},$$

for $\beta \in H_{c,(b)}^q(G, \mathbf{R}_{(\varepsilon)})$.

The same definition gives the continuous (bounded) cohomology of any topological group acting either trivially on $\mathbf{R}$ or via a homomorphism into the multiplicative group $\{+1, -1\}$. A continuous representation $\rho : H \to G$ naturally induces pullbacks

$$H_{c,(b)}^\bullet(G, \mathbf{R}) \longrightarrow H_{c,(b)}^\bullet(H, \mathbf{R}) \quad \text{and} \quad H_{c,(b)}^\bullet(G, \mathbf{R}_\varepsilon) \longrightarrow H_{c,(b)}^\bullet(H, \mathbf{R}_\rho),$$

where $\mathbf{R}_\rho$ is the H-module $\mathbb{R}$ with the H-action given by the composition of $\rho : H \to G$ with $\varepsilon : G \to \{+1, -1\}$. Note that $\|\rho^*(\beta)\| \leqslant \|\beta\|$.

Since the restriction to G^+ of the G-action on $\mathbf{R}_{(\varepsilon)}$ is trivial, there is a restriction map in cohomology

$$H^\bullet_{c,(b)}(G, \mathbf{R}_{(\varepsilon)}) \longrightarrow H^\bullet_{c,(b)}\big(G^+, \mathbf{R}\big). \tag{3}$$

In fact, both the continuous and the continuous bounded cohomology groups can be computed isometrically on the hyperbolic n-space $\mathbb{H}^n$, as this space is isomorphic to the quotient of G or G^+ by a maximal compact subgroup. More precisely, set

$$C_{c,(b)}\big(\big(\mathbb{H}^n\big)^{q+1}, \mathbf{R}_{(\varepsilon)}\big)^G = \big\{ f : \big(\mathbb{H}^n\big)^{q+1} \to \mathbf{R} \,\big|\, f \text{ is continuous (and bounded) and } \varepsilon(g) \cdot f(x_0, \dots, x_q) = f(gx_0, \dots, gx_q) \big\}$$

and endow it with its homogeneous coboundary operator. Then the cohomology of this cocomplex is isometrically isomorphic to the corresponding cohomology groups ([22, Chap. III, Prop. 2.3] and [27, Cor. 7.4.10] respectively).

It is now easy to describe the left inverses to the restriction map (3) induced by the inclusion. Indeed, at the cochain level, they are given by maps

$$p : C_{c,(b)}\big(\big(\mathbb{H}^n\big)^{q+1}, \mathbf{R}\big)^{G^+} \longrightarrow C_{c,(b)}\big(\big(\mathbb{H}^n\big)^{q+1}, \mathbf{R}\big)^G$$

and

$$\bar{p} : C_{c,(b)}\big(\big(\mathbb{H}^n\big)^{q+1}, \mathbf{R}\big)^{G^+} \longrightarrow C_{c,(b)}\big(\big(\mathbb{H}^n\big)^{q+1}, \mathbf{R}_\varepsilon\big)^G$$

defined for $x_0, \dots, x_q \in \mathbb{H}^n$ and $f \in C_{c,(b)}((\mathbb{H}^n)^{q+1}, \mathbf{R})^{G^+}$ by

$$p(f)(x_0, \dots, x_q) = \frac{1}{2}\big(f(x_0, \dots, x_q) + f(\tau x_0, \dots, \tau x_q)\big),$$
$$\bar{p}(f)(x_0, \dots, x_q) = \frac{1}{2}\big(f(x_0, \dots, x_q) - f(\tau x_0, \dots, \tau x_q)\big),$$

where $\tau \in G \smallsetminus G^+$ is any orientation reversing isometry. In fact, it easily follows from the G^+-invariance of f that $p(f)$ is invariant, $\bar{p}(f)$ is equivariant, and both $p(f)$ and $\bar{p}(f)$ are independent of τ in $G \smallsetminus G^+$. The following proposition is immediate:

Proposition 2.1 *The cochain map* $(p, \bar{p})$ *induces an isometric isomorphism*

$$H^\bullet_{c,(b)}\big(G^+, \mathbf{R}\big) \cong H^\bullet_{c,(b)}(G, \mathbf{R}) \oplus H^\bullet_{c,(b)}(G, \mathbf{R}_\varepsilon).$$

The continuous cohomology group $H^\bullet_c(G^+, \mathbf{R})$ is well understood since it can, via the van Est isomorphism [22, Corollary 7.2], be identified with the de Rham cohomology of the compact dual to $\mathbb{H}^n$, which is the n-sphere S^n. Thus it is generated by two cohomology classes: the constant class in degree 0, and the volume form in degree n. Recall that the volume form ω_n can be represented by the cocycle $\mathrm{Vol}_n \in C_{c,b}((\mathbb{H}^n)^{n+1}, \mathbf{R}_\varepsilon)^G$ (respectively $\mathrm{Vol}_n \in L^\infty((\partial\mathbb{H}^n)^{n+1}, \mathbf{R}_\varepsilon)^G)$ given by

$$\mathrm{Vol}_n(x_0, \dots, x_n) = \text{signed volume of the convex hull of } x_0, \dots, x_n,$$

for $x_0, \dots, x_n \in \mathbb{H}^n$, respectively $\partial\mathbb{H}^n$. Since the constant class in degree 0 is invariant, and the volume form is equivariant, using Proposition 2.1 we summarize this as follows:

$$H_c^0\big(G^+, \mathbf{R}\big) \cong H_c^0(G, \mathbf{R}) \cong \mathbf{R} \quad \text{and} \quad H_c^n\big(G^+, \mathbf{R}\big) \cong H_c^n(G, \mathbf{R}_\varepsilon) \cong \mathbf{R} \cong \langle \omega_n \rangle.$$

All other continuous cohomology groups are 0. On the bounded side, the cohomology groups are still widely unknown, though they are conjectured to be isomorphic to their unbounded counterparts. The comparison maps for G and G^+ are easily seen to be isomorphisms in degrees 2 and 3 [11]. We show that the comparison map for the *equivariant* cohomology of G is indeed an isometric isomorphism up to degree n, based on the simple Lemma 2.2. Before we prove it, it will be convenient to have yet two more cochain complexes to compute the continuous bounded cohomology groups. If $X = \mathbb{H}^n$ or $X = \partial\mathbb{H}^n$, consider the cochain space $L^\infty(X^{q+1}, \mathbf{R}_{(\varepsilon)})^G$ of G-invariant, resp. G-equivariant, essentially bounded measurable function classes endowed with its homogeneous coboundary operator. It is proven in [27, Cor. 7.5.9] that the cohomology of this cocomplex is isometrically isomorphic to the continuous bounded cohomology groups. Note that the volume cocycle Vol_n represents the same cohomology class viewed as continuous bounded or L^∞-cocycle on $\mathbb{H}^n$, as an L^∞-cocyle on $\partial\mathbb{H}^n$ or, by evaluation on $x \in \mathbb{H}^n$ or $x \in \partial\mathbb{H}^n$, as a continuous bounded or L^∞-cocycle on G.

Lemma 2.2 *For $q < n$ we have*

$$C_c\big((\mathbb{H}^n)^{q+1}, \mathbf{R}_\varepsilon\big)^G = 0,$$
$$L^\infty\big((\mathbb{H}^n)^{q+1}, \mathbf{R}_\varepsilon\big)^G = 0,$$
$$L^\infty\big((\partial\mathbb{H}^n)^{q+1}, \mathbf{R}_\varepsilon\big)^G = 0.$$

Proof Let $f : (\mathbb{H}^n)^{q+1} \to \mathbf{R}_\varepsilon$ or $f : (\partial\mathbb{H}^n)^{q+1} \to \mathbf{R}_\varepsilon$ be G-equivariant. The lemma relies on the simple observation that any $q+1 \leqslant n$ points $x_0, \dots, x_q$ either in $\mathbb{H}^n$ or in $\partial\mathbb{H}^n$ lie either on a hyperplane $P \subset \mathbb{H}^n$ or on the boundary of a hyperplane. Thus there exists an orientation reversing isometry $\tau \in G \setminus G^+$ fixing $(x_0, \dots, x_q)$ pointwise. Using the G-equivariance of f we conclude that

$$f(x_0, \dots, x_q) = -f(\tau x_0, \dots, \tau x_q) = -f(x_0, \dots, x_q),$$

which implies $f \equiv 0$. □

It follows from the lemma that $H_{c,b}^q(G, \mathbf{R}_\varepsilon) \cong H_c^q(G, \mathbf{R}_\varepsilon) = 0$ for $q < n$. Furthermore, we can conclude that the comparison map for the equivariant cohomology of G is injective:

Proposition 2.3 *The comparison map induces an isometric isomorphism*

$$H^n_{c,b}(G, \mathbf{R}_\varepsilon) \xrightarrow{\cong} H^n_c(G, \mathbf{R}_\varepsilon).$$

Proof Since there are no cochains in degree $n-1$, there are no coboundaries in degree n and the cohomology groups $H^n_{c,b}(G, \mathbf{R}_\varepsilon)$ and $H^n_c(G, \mathbf{R}_\varepsilon)$ are equal to the corresponding spaces of cocycles. Thus, we have a commutative diagram

$$\begin{array}{ccc} H^n_{c,b}(G, \mathbf{R}_\varepsilon) & = & \operatorname{Ker}\left\{\delta : C_{c,b}\left(\left(\mathbb{H}^n\right)^{n+1}, \mathbf{R}_\varepsilon\right)^G \to C_{c,b}\left(\left(\mathbb{H}^n\right)^{n+2}, \mathbf{R}_\varepsilon\right)^G\right\} \\ \downarrow & & \downarrow \\ \mathbf{R} \cong H^n_c(G, \mathbf{R}_\varepsilon) & = & \operatorname{Ker}\left\{\delta : C_c\left(\left(\mathbb{H}^n\right)^{n+1}, \mathbf{R}_\varepsilon\right)^G \to C_c\left(\left(\mathbb{H}^n\right)^{n+2}, \mathbf{R}_\varepsilon\right)^G\right\}. \end{array}$$

The proposition follows from the fact that the lower right kernel is generated by the volume form ω_n which is represented by the bounded cocycle Vol_n, hence is in the image of the vertical right inclusion. □

Since there are no coboundaries in degree n in $C_c((\mathbb{H}^n)^{q+1}, \mathbf{R}_\varepsilon)^G$, it follows that the cohomology norm of ω_n is equal to the norm of the unique cocycle representing it. In view of [23], its norm is equal to the volume v_n of an ideal regular simplex in $\mathbb{H}^n$.

Corollary 2.4 *The norm $\|\omega_n\|$ of the volume form $\omega_n \in H^n_c(G^+, \mathbf{R})$ is equal to the volume v_n of a regular ideal simplex in $\mathbb{H}^n$.*

As the cohomology norm $\|\omega_n\|$ is the proportionality constant between simplicial and Riemannian volume for closed hyperbolic manifolds [6, Theorem 2], the corollary gives a simple proof of the proportionality principle $\|M\| = \mathrm{Vol}(M)/v_n$ for closed hyperbolic manifolds, originally due to Gromov and Thurston.

3 Relative Cohomology

3.1 Notation and Definitions

As mentioned in the introduction, we consider a compact core N of the complete hyperbolic manifold M, that is a subset of M whose complement $M \setminus N$ in M is a disjoint union of finitely many geodesically convex cusps of M. If $q \geqslant 0$ and $\sigma : \Delta^q \to M$ denotes a singular simplex, where $\Delta^q = \{(t_0, \ldots, t_q) \in \mathbf{R}^{q+1} : \sum_{j=0}^q t_j = 1,\ t_j \geqslant 0 \text{ for all } j\}$ is a standard q-simplex, we recall that the (singular) cohomology $H^q(M, M \setminus N)$ of M relative to $M \setminus N$ is the cohomology of the cocomplex

$$C^q(M, M \setminus N) = \left\{ f \in C^q(M) \,\middle|\, f(\sigma) = 0 \text{ if } \mathrm{Im}(\sigma) \subset M \setminus N \right\}$$

endowed with its usual coboundary operator. (Here, $C^q(M)$ denotes the space of singular q-cochains on M.) We emphasize that all cohomology groups, singular or relative, are with $\mathbf{R}$ coefficients. The bounded relative cochains $C_b^q(M, M \setminus N)$ are those for which f is further assumed to be bounded, meaning that $\sup\{|f(\sigma)| \mid \sigma : \Delta^q \to M\}$ is finite. The coboundary restricts to bounded cochains and the cohomology of that cocomplex is the bounded cohomology of M relative to $M \setminus N$, which we denote by $H_b^\bullet(M, M \setminus N)$. The inclusion of cocomplexes induces a comparison map $c : H_b^\bullet(M, M \setminus N) \to H^\bullet(M, M \setminus N)$. Similarly, we could define the cohomology of N relative to its boundary ∂N and it is clear, by homotopy invariance, that $H_{(b)}^\bullet(N, \partial N) \cong H_{(b)}^\bullet(M, M \setminus N)$. We can identify the relative cochains on $(M, M \setminus N)$ with the Γ-invariant relative cochains $C^q(\mathbb{H}^n, U)^\Gamma$ on the universal cover $\mathbb{H}^n$ relative to the preimage $U = \pi^{-1}(M \setminus N)$ under the covering map $\pi : \mathbb{H}^n \to M$ of the finite union of horocyclic neighborhoods of cusps. We will identify $H_{(b)}^\bullet(N, \partial N)$ with the latter cohomology group. Note that U is a countable union of disjoint horoballs.

The inclusion $(M, \emptyset) \hookrightarrow (M, M \setminus N)$ induces a long exact sequence on both the unbounded and bounded cohomology groups

$$\cdots \longrightarrow H_{(b)}^{\bullet-1}(M \setminus N) \longrightarrow H_{(b)}^\bullet(M, M \setminus N) \longrightarrow H_{(b)}^\bullet(M) \longrightarrow H_{(b)}^\bullet(M \setminus N) \longrightarrow \cdots$$

Each connected component E_j of $M \setminus N$, $1 \leqslant j \leqslant k$, is a horocyclic neighborhood of a cusp, hence homeomorphic to the product of $\mathbf{R}$ with an $(n-1)$-manifold admitting a Euclidean metric; thus its universal covering is contractible and its fundamental group is virtually abelian (hence amenable). It follows that (see the introduction or [5, 21]) $H_b^\bullet(E_j) \cong H_b^\bullet(\pi_1(E_j)) = 0$ and hence $H_b^\bullet(M \setminus N) = 0$, proving that the inclusion $(M, \emptyset) \hookrightarrow (M, M \setminus N)$ induces an isomorphism on the bounded cohomology groups. Note that based on some techniques developed in [8] we can show that this isomorphism is isometric—a fact that we will not need in this note.

3.2 Transfer Maps

In the following we identify Γ with its image $i(\Gamma) < G^+$ under the lattice embedding $i : \Gamma \to G^+$. There exist natural transfer maps

$$H_b^\bullet(\Gamma) \xrightarrow{\text{trans}_\Gamma} H_{c,b}^\bullet(G, \mathbf{R}_\varepsilon) \quad \text{and} \quad H^\bullet(N, \partial N) \xrightarrow{\tau_{dR}} H_c^\bullet(G, \mathbf{R}_\varepsilon),$$

whose classical constructions we briefly recall here. The aim of this section will then be to establish the commutativity of the diagram (6) in Proposition 3.1. The proof is similar to that in [8], except that we replace the compact support cohomology by the relative cohomology, which leads to some simplifications. In fact, the same proof as in [13] (from where the use of relative bounded cohomology is borrowed) would have worked *verbatim* in this case, but we chose the other (and simpler) approach, to provide a “measure homology-free” proof.

3.2.1 The Transfer Map $\mathrm{trans}_\Gamma : H_b^\bullet(\Gamma) \to H_{c,b}^\bullet(G, \mathbf{R}_\varepsilon)$

We can define the transfer map at the cochain level either as a map

$$\mathrm{trans}_\Gamma : V_q^\Gamma \to V_q^G,$$

where V_q is one of $C_b((\mathbb{H}^n)^{q+1}, \mathbf{R})$, $L^\infty((\mathbb{H}^n)^{q+1}, \mathbf{R})$ or $L^\infty((\partial\mathbb{H}^n)^{q+1}, \mathbf{R})$. The definition is the same in all cases. Let thus c be a Γ-invariant cochain in V_q^Γ. Set

$$\mathrm{trans}_\Gamma(c)(x_0, \dots, x_n) := \int_{\Gamma\backslash G} \varepsilon\big(\dot{g}^{-1}\big) \cdot c(\dot{g}x_0, \dots, \dot{g}x_n)\, d\mu(\dot{g}), \tag{4}$$

where μ is the invariant probability measure on $\Gamma\backslash G$ normalized so that $\mu(\Gamma\backslash G) = 1$. Recall that $\Gamma < G^+$, so that $\varepsilon(\dot{g})$ is well defined. It is easy to check that the resulting cochain $\mathrm{trans}_\Gamma(c)$ is G-equivariant. Furthermore, the transfer map clearly commutes with the coboundary operator, and hence induces a cohomology map

$$H_b^\bullet(\Gamma) \xrightarrow{\mathrm{trans}_\Gamma} H_{c,b}^\bullet(G, \mathbf{R}_\varepsilon).$$

Note that if the cochain c is already G-equivariant, then $\mathrm{trans}_\Gamma(c) = c$, showing that trans_Γ is a left inverse of $i^* : H_{c,b}^\bullet(G, \mathbf{R}_\varepsilon) \to H_b^\bullet(\Gamma)$.

3.2.2 The Transfer Map $\tau_{dR} : H^\bullet(N, \partial N) \to H_c^\bullet(G, \mathbf{R}_\varepsilon)$

The relative de Rham cohomology $H_{dR}^\bullet(M, M \setminus N)$ is the cohomology of the cocomplex of differential forms $\Omega^q(M, M \setminus N)$ which vanish when restricted to $M \setminus N$. Then, as for usual cohomology, there is a de Rham Theorem

$$\Psi : H_{dR}^\bullet(M, M \setminus N) \xrightarrow{\cong} H^\bullet(M, M \setminus N) \cong H^\bullet(N, \partial N)$$

for relative cohomology. The isomorphism is given at the cochain level by integration. In order to integrate, we could either replace the singular cohomology by its smooth variant (i.e. take smooth singular simplices), or we prefer here to integrate the differential form on the straightened simplices. (The geodesic straightening of a continuous simplex is always smooth.) Thus, at the cochain level, the isomorphism is induced by the map

$$\Psi : \Omega^q(M, M \setminus N) \longrightarrow C^q(M, M \setminus N), \tag{5}$$

sending a differential form $\omega \in \Omega^q(M, M \setminus N) \cong \Omega^q(\mathbb{H}^n, U)^\Gamma$ to the singular cochain $\Psi(\omega)$ given by

$$\sigma \mapsto \int_{\pi_* \mathrm{straight}(x_0, \dots, x_q)} \omega,$$

where $\pi : \mathbb{H}^n \to M$ is the canonical projection, the $x_i \in \mathbb{H}^n$ are the vertices of a lift of σ to $\mathbb{H}^n$, and $\mathrm{straight}(x_0, \dots, x_q) : \Delta^q \to \mathbb{H}^n$ is the geodesic straightening. Observe that if σ is in U, then the straightened simplex is as well, since all components of U are geodesically convex.

The transfer map $\mathrm{trans}_{dR} : H^\bullet_{dR}(M, M \smallsetminus N) \to H^\bullet_c(G, \mathbb{R}_\varepsilon)$ is defined through the relative de Rham cohomology and the van Est isomorphism. At the cochain level the transfer

$$\mathrm{trans}_{dR} : \Omega^q\big(\mathbb{H}^n, U\big)^\Gamma \longrightarrow \Omega^q\big(\mathbb{H}^n, \mathbf{R}_\varepsilon\big)^G$$

is defined by sending the differential q-form $\alpha \in \Omega^q(\mathbb{H}^n)^\Gamma$ to the form

$$\mathrm{trans}_{dR}(\alpha) := \int_{\Gamma\backslash G} \varepsilon\big(\dot{g}^{-1}\big) \cdot \big(\dot{g}^*\alpha\big)\, d\mu(\dot{g}),$$

where μ is chosen as in (4). It is easy to check that the resulting differential form $\mathrm{trans}_{dR}(\alpha)$ is G-equivariant. Furthermore, the transfer map clearly commutes with the differential operator, and hence induces a cohomology map

$$\begin{array}{ccccc}
H^\bullet(N, \partial N) & & & & H^\bullet_c(G, \mathbf{R}_\varepsilon) \\
\cong \uparrow \Psi & & & & \uparrow \cong \\
H^\bullet_{dR}(M, M \smallsetminus N) & \xrightarrow{\mathrm{trans}_{dR}} & H^\bullet\big(\Omega^\bullet\big(\mathbb{H}^n, \mathbf{R}_\varepsilon\big)^G\big) & \xrightarrow{=} & \Omega^\bullet\big(\mathbb{H}^n, \mathbf{R}_\varepsilon\big)^G,
\end{array}$$

where the vertical arrow on the right is the van Est isomorphism and the horizontal arrow on the right follows from Cartan's lemma to the extent that any G-invariant differential form on $\mathbb{H}^n$ (or more generally on a symmetric space) is closed.

Let $\omega_{N,\partial N} \in H^n(M, M \smallsetminus N)$ be the unique class with $\langle \omega_{N,\partial N}, [N, \partial N] \rangle = \mathrm{Vol}(M)$. It is easy to check that

$$\mathrm{trans}_{dR}(\omega_{N,\partial N}) = \omega_n \in \Omega^n\big(\mathbb{H}^n, \mathbb{R}_\varepsilon\big)^G \cong H^n_c(G, \mathbf{R}_\varepsilon).$$

3.2.3 Commutativity of the Transfer Maps

Proposition 3.1 *The diagram*

$$\begin{array}{ccc}
H^q_b(\Gamma) & & \\
\cong \uparrow & \searrow^{\mathrm{trans}_\Gamma} & \\
H^q_b(N, \partial N) & & H^q_{c,b}(G, \mathbf{R}_\varepsilon) \\
\downarrow c & & c \downarrow \\
H^q(N, \partial N) & \xrightarrow{\tau_{dR}} & H^q_c(G, \mathbf{R}_\varepsilon)
\end{array} \tag{6}$$

commutes (*here* $\tau_{dR} = \mathrm{trans}_{dR} \circ \Psi^{-1}$).

Proof The idea of the proof is to subdivide the diagram (6) in smaller parts, by defining transfer maps directly on the bounded and unbounded relative singular cohomology of M and show that each of the following subdiagrams commute.

$$\begin{array}{ccc} H_b^q(\varGamma) & & \\ \cong\uparrow & \searrow^{\mathrm{trans}_\varGamma} & \\ H_b^q(N,\partial N) & \overset{\mathrm{trans}_b}{\dashrightarrow} & H_{c,b}^q(G,\mathbf{R}_\varepsilon) \\ \downarrow c & & c\downarrow \\ H^q(N,\partial N) & \overset{\mathrm{trans}}{\dashrightarrow} & H_c^q(G,\mathbf{R}_\varepsilon) \\ \cong\uparrow\varPsi & & \varPhi\uparrow\cong \\ H_{dR}^q(N,\partial N) & \xrightarrow{\mathrm{trans}_{dR}} & \varOmega^q(\mathbb{H}^n,\mathbf{R}_\varepsilon)^G. \end{array} \tag{7}$$

3.2.4 Definition of the Transfer Map for Relative Cohomology

In order to define a transfer map, we need to be able to integrate our cochain on translates of a singular simplex by elements of $\varGamma\backslash G$. This is only possible if the cochain is regular enough.

For $1 \leqslant i \leqslant k$, pick a point $b_i \in E_i$ in each horocyclic neighborhood of a cusp in M and $b_0 \in N$ in the compact core. Let $\beta' : M \to \{b_0, b_1, \dots, b_k\}$ be the measurable map sending N to b_0 and each cusp E_i to b_i. Lift β' to a $\varGamma$-equivariant measurable map

$$\beta : \mathbb{H}^n \longrightarrow \pi^{-1}\big(\{b_0, b_1, \dots, b_k\}\big) \subset \mathbb{H}^n$$

defined as follows. Choose lifts $\tilde{b}_0, \dots, \tilde{b}_k$ of $b_0, \dots, b_k$; for each $j = 1, \dots, k$ choose a Borel fundamental domain $\mathscr{D}_j \ni \tilde{b}_j$ for the $\varGamma$-action on $\pi^{-1}(E_j)$ and choose a fundamental domain $\mathscr{D}_0 \ni \tilde{b}_0$ for the $\varGamma$-action on $\pi^{-1}(N)$. Now define $\beta(\gamma\mathscr{D}_j) := \gamma\tilde{b}_j$. In particular β maps each horoball into itself. Given $c \in C^q(\mathbb{H}^n, U)^\varGamma$, define

$$\beta^*(c) : \big(\mathbb{H}^n\big)^{q+1} \longrightarrow \mathbf{R}$$

by

$$\beta^*(c)(x_0, \dots, x_q) = c\big(\mathrm{straight}\big(\beta(x_0), \dots, \beta(x_q)\big)\big). \tag{8}$$

Remark that $\beta^*(c)$ is $\varGamma$-invariant, vanishes on tuples of points that lie in the same horoball in the disjoint union of horoballs $\pi^{-1}(E_i)$, and is independent of the chosen lift of β' (but not of the points $b_0, \dots, b_k$). Thus, $\beta^*(c)$ is a cochain in $C^q(\mathbb{H}^n, U)^\varGamma$

which is now measurable, so that we can integrate it on translates of a given $(q+1)$-tuple of point. We define

$$\mathrm{trans}_\beta(c) : \left(\mathbb{H}^n\right)^{q+1} \longrightarrow \mathbf{R}$$

by

$$\mathrm{trans}_\beta(c)(x_0, \ldots, x_q) := \int_{\Gamma\backslash G} \varepsilon\left(\dot{g}^{-1}\right) \cdot \left(\beta^*(c)(\dot{g}x_0, \ldots, \dot{g}x_q)\right) d\mu(\dot{g}),$$

where μ is as in (4). It is easy to show that the integral is finite. Indeed, let D be the maximum of the distances between x_0 and x_i, for $i = 1, \ldots, q$. Then for $\dot{g} \in \Gamma\backslash G$ such that $\dot{g}x_0$ lies outside a D-neighborhood of the compact core N, each $\dot{g}x_i$ clearly lies outside N and hence $\beta^*(c)(\dot{g}x_0, \ldots, \dot{g}x_q)$ vanishes for such $\dot{g}$. It follows that the integrand vanishes outside a compact set, within which it takes only finitely many values. Furthermore, it follows from the Γ-invariance of c and $\beta(c)$ that $\mathrm{trans}_\beta(c)$ is G-invariant.

Since trans_β commutes with the coboundary operator, it induces a cohomology map

$$\mathrm{trans} : H^q(N, \partial N) \longrightarrow H^q_c(G, \mathbf{R}_\varepsilon).$$

As the transfer map trans_β restricts to a cochain map between the corresponding bounded cocomplexes, it also induces a map on the bounded cohomology groups

$$\mathrm{trans}_b : H^q_b(N, \partial N) \longrightarrow H^q_{c,b}(G, \mathbf{R}_\varepsilon),$$

and the commutativity of the middle diagram in (7) is now obvious.

3.2.5 Commutativity of the Lower Square

Denote by $\Phi : \Omega^q(\mathbb{H}^n, \mathbf{R}_\varepsilon) \longrightarrow L^\infty((\mathbb{H}^n)^{q+1}, \mathbf{R}_\varepsilon)$ the map (analogous to the map Ψ defined in (5)) sending the differential form α to the cochain $\Phi(\alpha)$ mapping a $(q+1)$-tuple of points $(x_0, \ldots, x_q) \in (\mathbb{H}^n)^{q+1}$ to

$$\int_{\mathrm{straight}(x_0, \ldots, x_q)} \alpha.$$

The de Rham isomorphism is realized at the cochain level by precomposing Φ with the map sending a singular simplex in $\mathbb{H}^n$ to its vertices. To check the commutativity of the lower square, observe that

$$\mathrm{trans}_\beta \circ \Phi(\alpha)(x_0, \ldots, x_q) = \int_{\Gamma\backslash G} \varepsilon\left(\dot{g}^{-1}\right) \cdot \left(\int_{\mathrm{straight}(\beta(\dot{g}x_0), \ldots, \beta(\dot{g}x_q))} \alpha\right) d\mu(\dot{g}),$$

while

$$\Phi \circ \mathrm{trans}_{dR}(\alpha)(x_0, \dots, x_q) = \int_{\Gamma\backslash G} \varepsilon\big(\dot{g}^{-1}\big) \cdot \bigg(\int_{\mathrm{straight}(\dot{g}x_0,\dots,\dot{g}x_q)} \alpha\bigg)\, d\mu(\dot{g}).$$

If $d\alpha = 0$, the coboundary of the G-invariant cochain

$$(x_0, \dots, x_{q-1}) \longmapsto \sum_{i=0}^{q-1} (-1)^i \int_{\Gamma\backslash G} \varepsilon\big(\dot{g}^{-1}\big) \cdot \bigg(\int_{\mathrm{straight}(\dot{g}x_0,\dots,\dot{g}x_i,\beta(\dot{g}x_i),\dots,\beta(\dot{g}x_{q-1}))} \alpha\bigg)\, d\mu(\dot{g})$$

is equal to the difference of the two given cocycles.

3.2.6 Commutativity of the Upper Triangle

Observe that the isomorphism $H_b^\bullet(M, M \setminus N) \cong H_b^\bullet(\Gamma)$ can be induced at the cochain level by the map $\beta^* : C_b^q(\mathbb{H}^n, U)^\Gamma \to L^\infty((\mathbb{H}^n)^{q+1}, \mathbf{R})^\Gamma$ defined in (8) (and for which we allow ourselves a slight abuse of notation). It is immediate that we now have commutativity of the upper triangle already at the cochain level,

$$\begin{array}{ccc} L^\infty\big((\mathbb{H}^n)^{q+1}, \mathbf{R}\big)^\Gamma & & \\ \uparrow{\scriptstyle \beta^*} & \searrow^{\mathrm{trans}_\Gamma} & \\ C_b^q\big(\mathbb{H}^n, U\big)^\Gamma & \xrightarrow[\mathrm{trans}_b]{} & L^\infty\big((\mathbb{H}^n)^{q+1}, \mathbf{R}_\varepsilon\big)^G. \end{array}$$

This finishes the proof of the proposition. □

3.3 *Properties of* Vol(ρ)

Lemma 3.2 *Let $i : \Gamma \hookrightarrow G$ be a lattice embedding. Then*

$$\mathrm{Vol}(i) = \mathrm{Vol}(M).$$

Proof Both sides are multiplicative with respect to finite index subgroups. We can hence without loss of generality suppose that Γ is torsion free. By definition, we have

$$\mathrm{Vol}(M) = \big\langle \omega_{N,\partial N}, [N, \partial N] \big\rangle,$$
$$\mathrm{Vol}(i) = \big\langle \big(c \circ i^*\big)\big(\omega_n^b\big), [N, \partial N] \big\rangle.$$

The desired equality would thus clearly follow from $\omega_{N,\partial N} = (c \circ i^*)(\omega_n^b)$. As the transfer map $\tau_{dR} : H^n(N, \partial N) \to H_c^n(G)$ is an isomorphism in top degree and sends $\omega_{N,\partial N}$ to ω_n, this is equivalent to

$$\omega_n = \tau_{dR}(\omega_{N,\partial N}) = \underbrace{\tau_{dR} \circ c}_{c \circ \mathrm{trans}_\Gamma} \circ i^*\big(\omega_n^b\big) = c \circ \mathrm{trans}_\Gamma \circ i^*\big(\omega_n^b\big) = c\big(\omega_n^b\big) = \omega_n,$$

where we have used the commutativity of the diagram (6) in Proposition 3.1 and the fact that $\mathrm{trans}_\Gamma \circ i^* = \mathrm{Id}$. □

Proposition 3.3 *Let $\rho : \Gamma \to G$ be a representation. The composition*

$$\mathbf{R} \cong H^n_{c,b}(G, \mathbf{R}_\varepsilon) \xrightarrow{\ \rho^*\ } H^n_b(\Gamma) \xrightarrow{\ \mathrm{trans}_\Gamma\ } H^n_{c,b}(G, \mathbf{R}_\varepsilon) \cong \mathbf{R}$$

is equal to $\lambda \cdot \mathrm{Id}$, *where*

$$|\lambda| = \frac{|\operatorname{Vol}(\rho)|}{\operatorname{Vol}(M)} \leqslant 1.$$

Proof As the quotient is left invariant by passing to finite index subgroups, we can without loss of generality suppose that Γ is torsion free. Let $\lambda \in \mathbf{R}$ be defined by

$$\mathrm{trans}_\Gamma \circ \rho^*\big(\omega_n^b\big) = \lambda \cdot \omega_n^b. \tag{9}$$

We apply the comparison map c to this equality and obtain

$$c \circ \mathrm{trans}_\Gamma \circ \rho^*\big(\omega_n^b\big) = \lambda \cdot c\big(\omega_n^b\big) = \lambda \cdot \omega_n = \lambda \cdot \tau_{dR}(\omega_{N,\partial N}).$$

The first expression of this line of equalities is equal to $\tau_{dR} \circ c \circ \rho^*(\omega_n^b)$ by the commutativity of the diagram (6). Since τ_{dR} is injective in top degree it follows that $(c \circ \rho^*)(\omega_n^b) = \lambda \cdot \omega_{N,\partial N}$. Evaluating on the fundamental class, we obtain

$$\operatorname{Vol}(\rho) = \big\langle \big(c \circ \rho^*\big)\big(\omega_n^b\big), [N, \partial N]\big\rangle = \lambda \cdot \big\langle \omega_{N,\partial N}, [N, \partial N]\big\rangle = \lambda \cdot \operatorname{Vol}(i) = \lambda \cdot \operatorname{Vol}(M).$$

For the inequality, we take the sup norms on both sides of (9), and get

$$|\lambda| = \frac{\|\mathrm{trans}_\Gamma \circ \rho^*(\omega_n^b)\|}{\|\omega_n^b\|} \leqslant 1,$$

where the inequality follows from the fact that all maps involved do not increase the norm. This finishes the proof of the proposition. □

4 On the Proof of Theorem 1.1

The simple inequality $|\operatorname{Vol}(\rho)| \leqslant |\operatorname{Vol}(i)| = \operatorname{Vol}(M)$ follows from Proposition 3.3 and Lemma 3.2.

The proof is divided into three steps. The first step, which follows essentially Furstenberg's footsteps [37, Chap. 4], consists in exhibiting a ρ-equivariant measurable boundary map $\varphi : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$. In the second step we will establish that φ maps the vertices of almost every positively oriented ideal simplex to vertices of positively (or negatively—one or the other, not both) oriented ideal simplices. In the third and last step we show that φ has to be the extension of an isometry, which will provide the conjugation between ρ and i. The fact that $n \geqslant 3$ will only be used in the third step.

4.1 Step 1: The Equivariant Boundary Map

We need to define a measurable map $\varphi : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ such that

$$\varphi\big(i(\gamma)\cdot\xi\big) = \rho(\gamma)\cdot\varphi(\xi), \tag{10}$$

for every $\xi \in \partial\mathbb{H}^n$ and every $\gamma \in \Gamma$.

The construction of such boundary map is the sore point of many rigidity questions. In the rank one situation in which we are, the construction is well known and much easier, and is recalled here for completeness.

Since $\partial\mathbb{H}^n$ can be identified with $\mathrm{Isom}^+(\mathbb{H}^n)/P$, where $P < \mathrm{Isom}^+(\mathbb{H}^n)$ is a minimal parabolic, the action of Γ on $\partial\mathbb{H}^n$ is amenable. Thus there exists a Γ-equivariant measurable map $\varphi : \partial\mathbb{H}^n \to \mathscr{M}^1(\partial\mathbb{H}^n)$, where $\mathscr{M}^1(\partial\mathbb{H}^n)$ denotes the probability measures on $\partial\mathbb{H}^n$, [37]. We recall the proof here for the sake of the reader familiar with the notion of amenable group but not conversant with that of amenable action, although the result is by now classical.

Lemma 4.1 *Let G be a locally compact group, $\Gamma < G$ a lattice and P an amenable subgroup. Let X be a compact metrizable space with a Γ-action by homeomorphisms. Then there exists a Γ-equivariant boundary map $\varphi : G/P \to \mathscr{M}^1(X)$.*

Proof Let $C(X)$ be the space of continuous functions on X. The space

$$\mathrm{L}^1_\Gamma\big(G, C(X)\big) := \bigg\{ f : G \to C(X) \,\Big|\, f \text{ is measurable, } \Gamma\text{-equivariant and } \int_{\Gamma\backslash G} \big\|f(\dot g)\big\|_\infty \, d\mu(\dot g) < \infty \bigg\},$$

is a separable Banach space whose dual is the space $\mathrm{L}^\infty_\Gamma(G, \mathscr{M}(X))$ of measurable Γ-equivariant essentially bounded maps from G into $\mathscr{M}(X)$, where $\mathscr{M}(X) = C(X)^*$ is the dual of $C(X)$. (Notice that since $C(X)$ is a separable Banach space, the concept of measurability of a function $G \to C(X)^*$ is the same as to whether $C(X)^*$ is endowed with the weak-* or the norm topology.) Then $\mathrm{L}^\infty_\Gamma(G, \mathscr{M}^1(X))$ is a convex compact subset of the unit ball of $\mathrm{L}^\infty_\Gamma(G, \mathscr{M}(X))$ that is right P-invariant. Since P is amenable, there exists a P-fixed point, that is nothing but the map $\varphi : G/P \to \mathscr{M}^1(X)$ we were looking for. □

We are going to associate to every $\mu \in \mathscr{M}^1(\partial\mathbb{H}^n)$ (in the image of φ) a point in $\partial\mathbb{H}^n$.

If the measure μ has only one atom of mass $\geqslant \frac{1}{2}$, then we associate to μ this atom. We will see that all other possibilities result in a contradiction.

If the measure μ has no atoms of mass greater than or equal to $\frac{1}{2}$, we can apply Douady and Earle's barycenter construction [16, Sect. 2] that to such a measure associates equivariantly a point $b_\mu \in \mathbb{H}^n$. By ergodicity of the Γ-action on $\partial\mathbb{H}^n \times \partial\mathbb{H}^n$, the distance $d := d(b_{\varphi(x)}, b_{\varphi(x')})$ between any two of these points is essentially constant. It follows that for a generic $x \in \partial\mathbb{H}^n$, there is a bounded orbit, contradicting the fact that the action is not elementary.

If on the other hand there is more than one atom whose mass is at least $\frac{1}{2}$, then the support of the measure must consist of two points (with an equally distributed measure). Denote by g_x the geodesic between the two points in the support of the measure $\varphi(x) \in \mathscr{M}^1(\partial\mathbb{H}^n)$. By ergodicity of the Γ-action on $\partial\mathbb{H}^n \times \partial\mathbb{H}^n$, the cardinality of the intersection $\operatorname{supp}(\varphi(x)) \cap \operatorname{supp}(\varphi(x'))$ must be almost everywhere constant and hence almost everywhere either equal to 0, 1 or 2.

If $|\operatorname{supp}(\varphi(x)) \cap \operatorname{supp}(\varphi(x'))| = 2$ for almost all $x, x' \in \partial\mathbb{H}^n$, then the geodesic g_x is Γ-invariant and hence the action is elementary.

If $|\operatorname{supp}(\varphi(x)) \cap \operatorname{supp}(\varphi(x'))| = 1$, then we have to distinguish two cases: either for almost every $x \in \partial\mathbb{H}^n$ there is a point $\xi \in \partial\mathbb{H}^n$ such that $\operatorname{supp}(\varphi(x)) \cap \operatorname{supp}(\varphi(x')) = \{\xi\}$ for almost all $x' \in \partial\mathbb{H}^n$, in which case again ξ would be Γ-invariant and the action elementary, or $\operatorname{supp}(\varphi(x)) \cup \operatorname{supp}(\varphi(x')) \cup \operatorname{supp}(\varphi(x''))$ consists of exactly three points for almost every $x', x'' \in \partial\mathbb{H}^n$. In this case the barycenter of the geodesic triangle with vertices in these three points is Γ-invariant and the action is, again, elementary.

Finally, if $|\operatorname{supp}(\varphi(x)) \cap \operatorname{supp}(\varphi(x'))| = 0$, let $D := d(g_x, g_{x'})$. By ergodicity on $\partial\mathbb{H}^n \times \partial\mathbb{H}^n$, d is essentially constant. Let $\gamma \in \rho(\Gamma)$ be a hyperbolic element whose fixed points are not the endpoints of g_x or $g_{x'}$. Then iterates of γ send a geodesic $g_{x'}$ into an arbitrarily small neighborhood of the attractive fixed point of γ, contradicting that g_x is at fixed distance from $g_{x'}$.

4.2 Step 2: Mapping Regular Simplices to Regular Simplices

The next step is to prove Theorem 1.4. Then if $\operatorname{Vol}(\rho) = \operatorname{Vol}(M)$, it will follow that the map φ in Step 1 sends almost all regular simplices to regular simplices.

From Proposition 3.3 we obtain that the composition of the induced map ρ^* and the transfer with respect to the lattice embedding i is equal to $\pm$ the identity on $H^n_{c,b}(G^+, \mathbf{R}_\varepsilon)$. In dimension 3, it follows from [4] that $H^3_{c,b}(\operatorname{Isom}^+(\mathbb{H}^3), \mathbf{R}) \cong \mathbf{R}$ and the proof can be formulated using trivial coefficients; this has been done in [11], which is the starting point of this paper. In higher dimension it is conjectured, but not known, that $H^n_{c,b}(G^+, \mathbf{R}) \cong \mathbf{R}$.

We can without loss of generality suppose that $\operatorname{trans}_\Gamma \circ \rho^*$ is equal to $+\operatorname{Id}$. Indeed, otherwise, we conjugate ρ by an orientation reversing isometry. We will now show that the isomorphism realized at the cochain level leads to the equality (11), which is only an almost everywhere equality. Up to this point, the proof is elementary. The only difficulty in our proof is to show that the almost everywhere equality is a true equality, which we prove in Proposition 4.2. Note however that there are two cases in which Proposition 4.2 is immediate, namely 1) if φ is a homeomorphism, which is the case if Γ is cocompact and ρ is also a lattice embedding (which is the case of the classical Mostow Rigidity Theorem), and 2) if the dimension n equals 3. We give the alternative simple arguments below.

The bounded cohomology groups $H^n_{c,b}(G, \mathbf{R}_\varepsilon)$ and $H^n_b(\Gamma, \mathbf{R})$ can both be computed from the corresponding L^∞ equivariant cochains on $\partial\mathbb{H}^n$. The induced map $\rho^* : H^n_{c,b}(G, \mathbf{R}_\varepsilon) \to H^n_b(\Gamma, \mathbf{R})$ is represented by the pullback by φ, although it should be noted that the pullback in bounded cohomology cannot be implemented with respect to boundary maps in general, unless the class to pull back can be represented by a strict invariant Borel cocycle. This is our case for Vol_n and as a consequence, $\varphi^*(\mathrm{Vol}_n)$ is also a measurable Γ-invariant cocycle. It hence determines a cohomology class in $H^n_b(\Gamma)$ which, by [10, Corollary 3.7], represents $\rho^*(\omega_n)$.

The composition of maps $\mathrm{trans}_\Gamma \circ \rho^*$ is thus realized at the cochain level by

$$\begin{aligned} L^\infty\big((\partial\mathbb{H}^n)^{n+1}, \mathbf{R}_\varepsilon\big)^\Gamma &\longrightarrow L^\infty\big((\partial\mathbb{H}^n)^{n+1}, \mathbf{R}_\varepsilon\big)^G \\ v &\longmapsto \Big\{(\xi_0, \dots, \xi_n) \mapsto \int_{\Gamma\backslash G} \varepsilon\big(\dot g^{-1}\big) v\big(\varphi\big(\dot g\xi_0, \dots, \dot g\xi_n\big)\big)\, d\mu(\dot g)\Big\}. \end{aligned}$$

Since the composition $\mathrm{trans}_\Gamma \circ \rho^*$ is the multiplication by $\frac{\mathrm{Vol}(\rho)}{\mathrm{Vol}(M)}$ at the cohomology level and there are no coboundaries in degree n (Lemma 2.2), the above map sends the cocycle Vol_n to $\frac{\mathrm{Vol}(\rho)}{\mathrm{Vol}(M)} \mathrm{Vol}_n$. Thus, for almost every $\xi_0, \dots, \xi_n \in \partial\mathbb{H}^n$ we have

$$\int_{\Gamma\backslash G} \varepsilon\big(\dot g^{-1}\big) \cdot \mathrm{Vol}_n\big(\varphi(\dot g\xi_0), \dots, \varphi(\dot g\xi_n)\big)\, d\mu(\dot g) = \frac{\mathrm{Vol}(\rho)}{\mathrm{Vol}(M)} \mathrm{Vol}_n(\xi_0, \dots, \xi_n). \quad (11)$$

Let $(\partial\mathbb{H}^n)^{(n+1)}$ be the G-invariant open subset of $(\partial\mathbb{H}^n)^{n+1}$ consisting of $(n+1)$-tuples of points $(\xi_0, \dots, \xi_n)$ such that $\xi_i \neq \xi_j$ for all $i \neq j$. Observe that the volume cocycle Vol_n is continuous when restricted to $(\partial\mathbb{H}^n)^{(n+1)}$ and vanishes on $(\partial\mathbb{H}^n)^{n+1} \smallsetminus (\partial\mathbb{H}^n)^{(n+1)}$. Observe moreover that the volume of ideal simplices is a continuous extension of the volume of simplices with vertices in the interior B^n of the sphere $S^{n-1} = \partial\mathbb{H}^n$.

Proposition 4.2 *Let $i : \Gamma \to G$ be a lattice embedding, $\rho : \Gamma \to G$ a representation and $\varphi : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ a Γ-equivariant measurable map. Identifying Γ with its image $i(\Gamma) < G$ via the lattice embedding, if*

$$\int_{\Gamma\backslash G} \varepsilon\big(\dot g^{-1}\big) \cdot \mathrm{Vol}_n\big(\varphi(\dot g\xi_0), \dots, \varphi(\dot g\xi_n)\big)\, d\mu(\dot g) = \frac{\mathrm{Vol}(\rho)}{\mathrm{Vol}(M)} \mathrm{Vol}_n(\xi_0, \dots, \xi_n)$$

for almost every $(\xi_0, \dots, \xi_n) \in (\partial\mathbb{H}^n)^{n+1}$, then the equality holds everywhere.

Before we proceed with the proof, let us observe that it immediately follows from the proposition that if ρ has maximal volume, then φ maps the vertices of almost every regular simplex to the vertices of a regular simplex of the same orientation, which is the conclusion of Step 2.

Proof if φ is a Homeomorphism Since φ is injective, both sides of the almost everywhere equality are continuous on $(\partial\mathbb{H}^n)^{(n+1)}$. Since they agree on a full measure subset of $(\partial\mathbb{H}^n)^{(n+1)}$, the equality holds on the whole of $(\partial\mathbb{H}^n)^{(n+1)}$. As for its complement, it is clear that if $\xi_i = \xi_j$ for $i \neq j$ then both sides of the equality vanish. □

Proof if $n = 3$ Both sides of the almost equality are defined on the whole of $(\partial\mathbb{H}^3)^4$, are cocycles on the whole of $(\partial\mathbb{H}^3)^4$, vanish on $(\partial\mathbb{H}^3)^4 \setminus (\partial\mathbb{H}^3)^{(4)}$ and are $\mathrm{Isom}^+(\mathbb{H}^3)$-invariant. Let $a, b : (\partial\mathbb{H}^3)^4 \to \mathbf{R}$ be two such functions and suppose that $a = b$ on a set of full measure. This means that for *almost every* $(\xi_0, \dots, \xi_3) \in (\partial\mathbb{H}^3)^4$, we have $a(\xi_0, \dots, \xi_3) = b(\xi_0, \dots, \xi_3)$. Since $\mathrm{Isom}^+(\mathbb{H}^3)$ acts transitively on 3-tuples of distinct points in $\mathbb{H}^3$ and both a and b are $\mathrm{Isom}^+(\mathbb{H}^3)$-invariant, this means that for *every* $(\xi_0, \xi_1, \xi_2) \in (\partial\mathbb{H}^3)^{(3)}$ and almost every $\eta \in \partial\mathbb{H}^3$ the equality

$$a(\xi_0, \xi_1, \xi_2, \eta) = b(\xi_0, \xi_1, \xi_2, \eta)$$

holds. Let $\xi_0, \dots, \xi_3 \in \partial\mathbb{H}^3$ be arbitrary. If $\xi_i = \xi_j$ for $i \neq j$, we have $a(\xi_0, \dots, \xi_3) = b(\xi_0, \dots, \xi_3)$ by assumption. Suppose $\xi_i \neq \xi_j$ whenever $i \neq j$. By the above, for every $i \in 0, \dots, 3$ the equality

$$a(\xi_0, \dots, \widehat{\xi_i}, \dots, \xi_3, \eta) = b(\xi_0, \dots, \widehat{\xi_i}, \dots, \xi_3, \eta)$$

holds for η in a subset of full measure in $\partial\mathbb{H}^3$. Let η be in the (non empty) intersection of these four full measure subsets of $\partial\mathbb{H}^3$. We then have

$$\begin{aligned} a(\xi_0, \dots, \xi_3) &= \sum_{i=0}^{3} (-1)^i a(\xi_0, \dots, \widehat{\xi_i}, \dots, \xi_3, \eta) \\ &= \sum_{i=0}^{3} (-1)^i b(\xi_0, \dots, \widehat{\xi_i}, \dots, \xi_3, \eta) = b(\xi_0, \dots, \xi_3), \end{aligned}$$

where we have used the cocycle relations for a and b in the first and last equality respectively. □

Proof in the General Case Observe first of all that for all $(\xi_0, \dots, \xi_n) \in (\partial\mathbb{H}^n)^{n+1} \setminus (\partial\mathbb{H}^n)^{(n+1)}$ the equality holds trivially.

Using the fact that $\partial\mathbb{H}^n \cong S^{n-1} \subset \mathbf{R}^n$, let us consider the function $\varphi : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ as a function $\varphi : \partial\mathbb{H}^n \to \mathbf{R}^n$ and for $j = 1, \dots, n$ denote by φ_j its coordinates. Since $\partial\mathbb{H}^n \cong G/P$, where P is a minimal parabolic, let ν be the quasi-invariant measure on $\partial\mathbb{H}^n$ obtained from the decomposition of the Haar measure μ_G with respect to the Haar measure μ_P on P, as in (17). According to Lusin's theorem applied to the φ_j for $j = 1, \dots, n$ (see for example [31, Theorem 2.24]), for every $\delta > 0$ there exist a measurable set $B_{\delta,i} \subset \partial\mathbb{H}^n$ with measure $\nu(B_{i,\delta}) \leqslant \delta$ and a continuous function $f'_{j,\delta} : \partial\mathbb{H}^n \to \mathbf{R}$ such that $\varphi_j \equiv f'_{j,\delta}$ on $\partial\mathbb{H}^n \setminus B_{j,\delta}$. Set $f'_\delta := (f_{1,\delta}, \dots, f_{n,\delta}) \to \mathbf{R}^n$ and consider the composition $f_\delta := r \circ f'_\delta$ with the retraction $r : \mathbf{R}^n \to \overline{B^n}$ to the closed unit ball $\overline{B^n} \subset \mathbf{R}^n$. Then, by setting $B_\delta := \bigcup_{j=1}^n B_{j,\delta}$, we see that φ coincides on $\partial\mathbb{H}^n \setminus B_\delta$ with the continuous function $f_\delta : \partial\mathbb{H}^n \to \overline{B^n}$ and $\nu(B_\delta) \leqslant n\delta$.

Let $\mathscr{D} \subset G$ be a fundamental domain for the action of Γ on G. For every measurable subset $E \subset \mathscr{D}$, any measurable map $\psi : \partial\mathbb{H}^n \to \overline{B^n}$ and any point $(\xi_0, \dots, \xi_n) \in (\partial\mathbb{H}^n)^{(n+1)}$, we use the notation

$$\mathscr{I}\big(\psi, E, (\xi_0, \dots, \xi_n)\big) := \int_E \varepsilon\big(g^{-1}\big)\,\mathrm{Vol}_n\big(\psi(g\xi_0), \dots, \psi(g\xi_n)\big)\,d\mu_G(g),$$

so that we need to show that if

$$\mathscr{I}\big(\varphi, \mathscr{D}, (\xi_0, \dots, \xi_n)\big) = \frac{\mathrm{Vol}(\rho)}{\mathrm{Vol}(M)}\,\mathrm{Vol}_n(\xi_0, \dots, \xi_n) \tag{12}$$

for almost every $(\xi_0, \dots, \xi_n) \in (\partial\mathbb{H}^n)^{(n+1)}$, then the equality holds everywhere.

Fix $\epsilon > 0$ and let $K_\epsilon \subset \mathscr{D}$ be a compact set such that $\mu_G(\mathscr{D} \setminus K_\epsilon) < \epsilon$. The proof is broken up in several lemmas, that we state and use here, but whose proof we postpone.

Lemma 4.3 *With the above notations*,

$$\mu_G\big(\{g \in K_\epsilon : g\xi \in B_\delta\}\big) \leqslant \sigma_\epsilon(\delta), \tag{13}$$

where $\sigma_\epsilon(\delta)$ *does not depend on* $\xi \in \partial\mathbb{H}^n$ *and* $\sigma_\epsilon(\delta) \to 0$ *when* $\delta \to 0$.

Replacing φ with f_δ results in the following estimate for the integral.

Lemma 4.4 *With the notation as above, there exists a function* $M_\epsilon(\delta)$ *with the property that* $\lim_{\delta\to 0} M_\epsilon(\delta) = 0$, *such that*

$$\Big|\mathscr{I}\big(\varphi, K_\epsilon, (\xi_0, \dots, \xi_n)\big) - \mathscr{I}\big(f_\delta, K_\epsilon, (\xi_0, \dots, \xi_n)\big)\Big| \leqslant M_\epsilon(\delta),$$

for all $(\xi_0, \dots, \xi_{n+1}) \in (\partial\mathbb{H}^n)^{n+1}$.

Observe that, although

$$\Big|\mathscr{I}\big(\varphi, \mathscr{D}, (\xi_0, \dots, \xi_n)\big) - \mathscr{I}\big(\varphi, K_\epsilon, (\xi_0, \dots, \xi_n)\big)\Big| < \epsilon \|\,\mathrm{Vol}_n\,\|, \tag{14}$$

for all $(\xi_0, \dots, \xi_{n+1}) \in (\partial\mathbb{H}^n)^{(n+1)}$, the estimate

$$\left|\mathscr{I}\big(\varphi, K_\epsilon, (\xi_0, \dots, \xi_n)\big) - \frac{\mathrm{Vol}(\rho)}{\mathrm{Vol}(M)}\,\mathrm{Vol}_n(\xi_0, \dots, \xi_n)\right|$$

$$\leqslant \left|\mathscr{I}\big(\varphi, K_\epsilon, (\xi_0, \ldots, \xi_n)\big) - \mathscr{I}\big(\varphi, \mathscr{D}, (\xi_0, \ldots, \xi_n)\big)\right| + \left|\mathscr{I}\big(\varphi, \mathscr{D}, (\xi_0, \ldots, \xi_n)\big) - \frac{\mathrm{Vol}(\rho)}{\mathrm{Vol}(M)} \mathrm{Vol}_n(\xi_0, \ldots, \xi_n)\right| \leqslant \epsilon \| \mathrm{Vol}_n \|, \tag{15}$$

holds only for almost every $(\xi_0, \ldots, \xi_n) \in (\partial\mathbb{H}^n)^{(n+1)}$, since this is the case for (12).

From (15) and Lemma 4.4, it follows that

$$\begin{aligned} &\left|\mathscr{I}\big(f_\delta, K_\epsilon, (\xi_0, \ldots, \xi_n)\big) - \frac{\mathrm{Vol}(\rho)}{\mathrm{Vol}(M)} \mathrm{Vol}_n(\xi_0, \ldots, \xi_n)\right| \\ &\quad \leqslant \left|\mathscr{I}\big(f_\delta, K_\epsilon, (\xi_0, \ldots, \xi_n)\big) - \mathscr{I}\big(\varphi, K_\epsilon, (\xi_0, \ldots, \xi_n)\big)\right| \\ &\qquad + \left|\mathscr{I}\big(\varphi, K_\epsilon, (\xi_0, \ldots, \xi_n)\big) - \frac{\mathrm{Vol}(\rho)}{\mathrm{Vol}(M)} \mathrm{Vol}_n(\xi_0, \ldots, \xi_n)\right| \\ &\quad < M_\epsilon(\delta) + \epsilon \| \mathrm{Vol}_n \|, \end{aligned} \tag{16}$$

for almost every $(\xi_0, \ldots, \xi_n) \in (\partial\mathbb{H}^n)^{(n+1)}$.

The following lemma uses the continuity of f_δ to deduce that all of the almost everywhere equality that propagated from the use of (12) in (15), can indeed be observed to hold everywhere because of the use of Lusin theorem.

Lemma 4.5 *There exist a function* $L(\epsilon, \delta)$ *such that* $\lim_{\epsilon\to 0} \lim_{\delta\to 0} L(\epsilon, \delta) = 0$ *and*

$$\left|\mathscr{I}\big(f_\delta, K_\epsilon, (\xi_0, \ldots, \xi_n)\big) - \frac{\mathrm{Vol}(\rho)}{\mathrm{Vol}(M)} \mathrm{Vol}_n(\xi_0, \ldots, \xi_n)\right| \leqslant L(\epsilon, \delta)$$

for all $(\xi_0, \ldots, \xi_n) \in (\partial\mathbb{H}^n)^{(n+1)}$.

From this, and from Lemma 4.4, and using once again (14), now all everywhere statements, we conclude that

$$\begin{aligned} &\left|\mathscr{I}\big(\varphi, \mathscr{D}, (\xi_0, \ldots, \xi_n)\big) - \frac{\mathrm{Vol}(\rho)}{\mathrm{Vol}(M)} \mathrm{Vol}_n(\xi_0, \ldots, \xi_n)\right| \\ &\quad \leqslant \left|\mathscr{I}\big(\varphi, \mathscr{D}, (\xi_0, \ldots, \xi_n)\big) - \mathscr{I}\big(\varphi, K_\epsilon, (\xi_0, \ldots, \xi_n)\big)\right| \\ &\qquad + \left|\mathscr{I}\big(\varphi, K_\epsilon, (\xi_0, \ldots, \xi_n)\big) - \mathscr{I}\big(f_\delta, K_\epsilon, (\xi_0, \ldots, \xi_n)\big)\right| \\ &\qquad + \left|\mathscr{I}\big(f_\delta, K_\epsilon, (\xi_0, \ldots, \xi_n)\big) - \frac{\mathrm{Vol}(\rho)}{\mathrm{Vol}(M)} \mathrm{Vol}_n(\xi_0, \ldots, \xi_n)\right| \\ &\quad < M_\epsilon(\delta) + L(\epsilon, \delta) + \epsilon \| \mathrm{Vol}_n \|, \end{aligned}$$

for all $(\xi_0, \ldots, \xi_{n+1}) \in (\partial\mathbb{H}^n)^{n+1}$. This concludes the proof of Proposition 4.2, assuming the unproven lemmas. □

We now proceed to the proof of Lemmas 4.3, 4.4 and 4.5.

Proof of Lemma 4.3 Recall that $\partial\mathbb{H}^n = G/P$, where $P < G$ is a minimal parabolic and let $\eta : G/P \to G$ be a Borel section of the projection $G \to G/P$ such that $F := \eta(G/P)$ is relatively compact [26, Lemma 1.1]. Let $\widetilde{B}_\delta := \eta(B_\delta)$ and, if $\xi \in B_\delta$, set $\tilde{\xi} := \eta(\xi) \in \widetilde{B}_\delta$. On the other hand, if $g \in K_\epsilon$ and $g\xi \in B_\delta$, there exists $p \in P$ such that $g\tilde{\xi}p \in \widetilde{B}_\delta$ and, in fact, the p can be chosen to be in $P \cap F^{-1}(K_\epsilon)^{-1}F =: C_\epsilon$. Thus we have

$$\begin{aligned}\{g \in K_\epsilon : g\xi \in B_\delta\} &= \left\{g \in K_\epsilon : \text{there exists } p \in C_\epsilon \text{ with } g\tilde{\xi}p \in \widetilde{B}_\delta\right\}\\ &= \left\{g \in K_\epsilon \cap \widetilde{B}_\delta p^{-1}\tilde{\xi}^{-1} \text{ for some } p \in C_\epsilon\right\} \subset K_\epsilon \cap \widetilde{B}_\delta C_\epsilon^{-1}\tilde{\xi}^{-1},\end{aligned}$$

and hence

$$\mu_G\big(\{g \in K_\epsilon : g\xi \in B_\delta\}\big) \leqslant \mu_G\big(K_\epsilon\tilde{\xi}^{-1} \cap \widetilde{B}_\delta C_\epsilon^{-1}\big) \leqslant \mu_G\big(\widetilde{B}_\delta C_\epsilon^{-1}\big).$$

To estimate the measure, recall that there is a strictly positive continuous function $q : G \to \mathbf{R}^+$ and a positive measure ν on $\partial\mathbb{H}^n$ such that

$$\int_G f(g)q(g)\,d\mu_G(g) = \int_{\partial\mathbb{H}^n}\left(\int_P f(\dot{g}\xi)\,d\mu_P(\xi)\right)d\nu(\dot{g}), \tag{17}$$

for all continuous functions f on G with compact support, [30, Sect. 8.1].

We may assume that $\mu_G(\widetilde{B}_\delta C_\epsilon^{-1}) \neq 0$ (otherwise we are done). Then, since q is continuous and strictly positive and the integral is on a relatively compact set, there exists a constant $0 < \alpha < \infty$ such that

$$\alpha\mu_G\big(\widetilde{B}_\delta C_\epsilon^{-1}\big) = \int_{\partial\mathbb{H}^n}\left(\int_P \chi_{\widetilde{B}_\delta C_\epsilon^{-1}}(\dot{g}\xi)\,d\mu_P(\xi)\right)d\nu(\dot{g}).$$

But, by construction, if $g \in \widetilde{B}_\delta$, then $g\xi \in \widetilde{B}_\delta C_\epsilon^{-1}$ if and only if $\xi \in C_\epsilon^{-1}$, so that

$$\int_P \chi_{\widetilde{B}_\delta C_\epsilon^{-1}}(\dot{g}\xi)\,d\mu_P(\xi) = \mu_P\big(C_\epsilon^{-1}\big),$$

and hence

$$\alpha\mu_G\big(\widetilde{B}_\delta C_\epsilon^{-1}\big) = \nu(B_\delta)\mu_P\big(C_\epsilon^{-1}\big).$$

Since $\nu(B_\delta) < \delta$, the inequality (13) is proven with $\sigma_\epsilon(\delta) = \frac{1}{\alpha}\mu_P(C_\epsilon^{-1})\delta$. □

Proof of Lemma 4.4 Let us fix $(\xi_0, \dots, \xi_n) \in (\partial\mathbb{H}^n)^{n+1}$. Then we have

$$\begin{aligned}&\big|\mathscr{I}\big(\varphi, K_\epsilon, (\xi_0, \dots, \xi_n)\big) - \mathscr{I}\big(f_\delta, K_\epsilon, (\xi_0, \dots, \xi_n)\big)\big|\\ &\quad\leqslant \big|\mathscr{I}\big(\varphi, K_{\epsilon,0}, (\xi_0, \dots, \xi_n)\big) - \mathscr{I}\big(f_\delta, K_{\epsilon,0}, (\xi_0, \dots, \xi_n)\big)\big|\\ &\qquad+ \big|\mathscr{I}\big(\varphi, K_{\epsilon,1}, (\xi_0, \dots, \xi_n)\big) - \mathscr{I}\big(f_\delta, K_{\epsilon,1}, (\xi_0, \dots, \xi_n)\big)\big|,\end{aligned}$$

where

$$K_{\epsilon,0} := \bigcap_{j=0}^{n} \left\{ g \in K_\epsilon : g\xi_j \in \partial\mathbb{H}^n \setminus B_\delta \right\} \quad \text{and} \quad K_{\epsilon,1} := K_\epsilon \setminus K_{\epsilon,0}.$$

But $\varphi(g) = f_\delta(g)$ for all $g \in K_{\epsilon,0}$, hence the difference of the integrals on $K_{\epsilon,0}$ vanishes. Since

$$\mu_G(K_{\epsilon,1}) = \mu_G\left(K_\epsilon \cap \bigcup_{j=0}^{n} \{g \in K_\epsilon : g\xi_j \in B_\delta\} \right) \leqslant (n+1)\sigma_\epsilon(\delta),$$

we obtain the assertion with $M_\epsilon(\delta) := 2(n+1)\| \operatorname{Vol}_n \| \sigma_\epsilon(\delta)$. □

Proof of Lemma 4.5 If the volume were continuous on $(\partial\mathbb{H}^n)^{n+1}$ or if the function f_δ were injective, the assertion would be obvious.

Observe that φ is almost everywhere injective: in fact, by double ergodicity, the subset of $\partial\mathbb{H}^n \times \partial\mathbb{H}^n$ consisting of pairs (x, y) for which $\varphi(x) = \varphi(y)$ is a set of either zero or full measure and the latter would contradict non-elementarity of the action. Then on a set of full measure in $\partial\mathbb{H}^n \setminus B_\delta$ the function f_δ is injective and hence $\operatorname{Vol}_n(f_\delta(g\xi_0), \dots, f_\delta(g\xi_n))$ is continuous provided the $f_\delta(g\xi_0), \dots, f_\delta(g\xi_n)$ are pairwise distinct.

So, for any $(\xi_0, \dots, \xi_n) \in (\partial\mathbb{H}^n)^{(n+1)}$ we define

$$\mathscr{E}(\xi_0, \dots, \xi_n) := \left\{ g \in K_\epsilon : f_\delta(g\xi_0), \dots, f_\delta(g\xi_n) \text{ are pairwise distinct} \right\}.$$

Let $F \subset (B_\delta^c \times B_\delta^c)^{(2)}$ be the set of distinct pairs $(x, y) \in (B_\delta^c \times B_\delta^c)^{(2)}$ such that $f_\delta(x) = f_\delta(y)$. Then F is of measure zero, and given any $(\xi_0, \xi_1) \in \partial\mathbb{H}^n \times \partial\mathbb{H}^n$ distinct, the set $\{g \in G : g(\xi_0, \xi_1) \in F\}$ is of μ_G-measure zero. This, together with Lemma 4.3, implies that

$$\mu_G\big(K_\epsilon \setminus \mathscr{E}(\xi_0, \dots, \xi_n)\big) \leqslant \mu_G\left(\bigcup_{j=0}^{n} \{g \in K_\epsilon : g\xi_j \in B_\delta\} \right) \leqslant (n+1)\sigma_\epsilon(\delta). \quad (18)$$

Let $\mathscr{S} \subset (\partial\mathbb{H}^n)^{(n+1)}$ be the set of full measure where the inequality (16) holds and let $(\xi_0, \dots, \xi_n) \in (\partial\mathbb{H}^n)^{(n+1)}$. Since $\nu^{n+1}((\partial\mathbb{H}^n)^{(n+1)} \setminus \mathscr{S}) = 0$, there exists a sequence of points $(\xi_0^{(k)}, \dots, \xi_n^{(k)}) \in \mathscr{S}$ with $(\xi_0^{(k)}, \dots, \xi_n^{(k)}) \to (\xi_0, \dots, \xi_n)$. Then for every $g \in \mathscr{E}(\xi)$

$$\lim_{k\to\infty} \operatorname{Vol}_n\big(f_\delta\big(g\xi_0^{(k)}\big), \dots, f_\delta\big(g\xi_n^{(k)}\big)\big) = \operatorname{Vol}_n\big(f_\delta(g\xi_0), \dots, f_\delta(g\xi_n)\big),$$

and, by the dominated convergence theorem applied to the sequence $h_k(g) := \operatorname{Vol}_n(f_\delta(g\xi_0^{(k)}), \dots, f_\delta(g\xi_n^{(k)}))$, we deduce that

$$\lim_{k\to\infty} \mathscr{I}\big(f_\delta, \mathscr{E}(\xi_0, \dots, \xi_n), \big(\xi_0^{(k)}, \dots, \xi_n^{(k)}\big)\big) = \mathscr{I}\big(f_\delta, \mathscr{E}(\xi_0, \dots, \xi_n), (\xi_0, \dots, \xi_n)\big). \quad (19)$$

But then

$$\begin{aligned}
&\left|\mathscr{I}\big(f_\delta, K_\epsilon, (\xi_0,\dots,\xi_n)\big) - \frac{\mathrm{Vol}(\rho)}{\mathrm{Vol}(M)}\mathrm{Vol}_n(\xi_0,\dots,\xi_n)\right| \\
&\quad\leqslant \left|\mathscr{I}\big(f_\delta, K_\epsilon, (\xi_0,\dots,\xi_n)\big) - \mathscr{I}\big(f_\delta, \mathscr{E}(\xi_0,\dots,\xi_n), (\xi_0,\dots,\xi_n)\big)\right| \\
&\qquad + \left|\mathscr{I}\big(f_\delta, \mathscr{E}(\xi_0,\dots,\xi_n), (\xi_0,\dots,\xi_n)\big) - \mathscr{I}\big(f_\delta, \mathscr{E}(\xi_0,\dots,\xi_n), \big(\xi_0^{(k)},\dots,\xi_n^{(k)}\big)\big)\right| \\
&\qquad + \left|\mathscr{I}\big(f_\delta, \mathscr{E}(\xi_0,\dots,\xi_n), \big(\xi_0^{(k)},\dots,\xi_n^{(k)}\big)\big) - \mathscr{I}\big(f_\delta, K_\epsilon, \big(\xi_0^{(k)},\dots,\xi_n^{(k)}\big)\big)\right| \\
&\qquad + \left|\mathscr{I}\big(f_\delta, K_\epsilon, \big(\xi_0^{(k)},\dots,\xi_n^{(k)}\big)\big) - \frac{\mathrm{Vol}(\rho)}{\mathrm{Vol}(M)}\mathrm{Vol}_n\big(\xi_0^{(k)},\dots,\xi_n^{(k)}\big)\right| \\
&\qquad + \left|\frac{\mathrm{Vol}(\rho)}{\mathrm{Vol}(M)}\mathrm{Vol}_n\big(\xi_0^{(k)},\dots,\xi_n^{(k)}\big) - \frac{\mathrm{Vol}(\rho)}{\mathrm{Vol}(M)}\mathrm{Vol}_n(\xi_0,\dots,\xi_n)\right|,
\end{aligned}$$

for all $(\xi_0,\dots,\xi_n) \in (\partial\mathbb{H}^n)^{(n+1)}$.

The first and third lines after the inequality sign are each $\leqslant (n+1)\|\mathrm{Vol}_n\|\sigma_\epsilon(\delta)$ because of (18); the second line after the equality is less than δ if k is large enough because of (19); the fourth line is $\leqslant M_\epsilon(\delta) + \epsilon\|\mathrm{Vol}_n\|$ by (16) since $(\xi_0^{(k)},\dots,\xi_n^{(k)}) \in \mathscr{S}$ and finally the last line is also less than δ if k if large enough. All of the estimate hold for all $(\xi_0,\dots,\xi_n) \in (\partial\mathbb{H}^n)^{(n+1)}$, and hence the assertion is proven with $L(\epsilon,\delta) := 2\delta + 2(n+1)\|\mathrm{Vol}_n\|\sigma_\epsilon(\delta) + M_\epsilon(\delta) + \epsilon\|\mathrm{Vol}_n\|$. ☐

4.3 Step 3: The Boundary Map is an Isometry

Suppose now that the equality $|\mathrm{Vol}(\rho)| = |\mathrm{Vol}(i)|$ holds. Then φ maps enough regular simplices to regular simplices. In this last step of the proof, we want to show that then φ is essentially an isometry, and this isometry will realize the conjugation between ρ and i.

In the case of a cocompact lattice $\Gamma < \mathrm{Isom}(\mathbb{H}^n)$ and a lattice embedding $\rho : \Gamma \to \mathrm{Isom}^+(\mathbb{H}^n)$, the limit map φ is continuous and the proof is very simple based on Lemma 4.6. This is the original setting of Gromov's proof of Mostow rigidity for compact hyperbolic manifolds.

If either the representation ρ is not assumed to be a lattice embedding, or if Γ is not cocompact, then the limit map φ is only measurable and one needs a measurable variant of Lemma 4.6 presented in Proposition 4.7 for $n \geqslant 4$. The case $n = 3$ was first proven by Thurston for his generalization (Corollary 1.3 here) of Gromov's proof of Mostow rigidity. It is largely admitted that the case $n = 3$ easily generalizes to $n \geqslant 4$, although we wish to point out that the proof is very much simpler for $n \geqslant 4$ based on the fact that the reflection group of a regular simplex is dense in the isometry group. For the proof of Proposition 4.7, we will omit the case $n = 3$ which is nicely written down in all necessary details by Dunfield [17, pp. 654–656], following the original [33, two last paragraphs of Sect. 6.4].

Let T denote the set of $(n+1)$-tuples of points in $\partial\mathbb{H}^n$ which are vertices of a regular simplex,

$$T = \left\{\underline{\xi} = (\xi_0, \dots, \xi_n) \in \left(\partial\mathbb{H}^n\right)^{n+1} \mid \underline{\xi} \text{ are the vertices of an ideal regular simplex}\right\}.$$

We shall call an $(n+1)$-tuple in T a regular simplex. Note that the order of the vertices $\xi_0, \dots, \xi_n$ induces an orientation on the simplex $\underline{\xi}$. For $\underline{\xi} \in T$, denote by $\Lambda_{\underline{\xi}} < \mathrm{Isom}(\mathbb{H}^n)$ the reflection group generated by the reflections in the faces of the simplex $\underline{\xi}$.

Lemma 4.6 *Let $n \geqslant 3$. Let $\underline{\xi} = (\xi_0, \dots, \xi_n) \in T$. Suppose that $\varphi : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ is a map such that for every $\gamma \in \Lambda_{\underline{\xi}}$, the simplex with vertices $(\varphi(\gamma\xi_0), \dots, \varphi(\gamma\xi_n))$ is regular and of the same orientation as $(\gamma\xi_0, \dots, \gamma\xi_n) \in T$. Then there exists a unique isometry $h \in \mathrm{Isom}(\mathbb{H}^n)$ such that $h(\xi) = \varphi(\xi)$ for every $\xi \in \bigcup_{i=0}^n \Lambda_{\underline{\xi}}\xi_i$.*

Note that this lemma and its subsequent proposition are the only places in the proof where the assumption $n \geqslant 3$ is needed. The lemma is wrong for $n = 2$ since φ could be any orientation preserving homeomorphism of $\partial\mathbb{H}^2$.

Proof If $\underline{\xi} = (\xi_0, \dots, \xi_n)$ and $(\varphi(\xi_0), \dots, \varphi(\xi_n))$ belong to T, then there exists a unique isometry $h \in \mathrm{Isom}^+(\mathbb{H}^n)$ such that $h\xi_i = \varphi(\xi_i)$ for $i = 0, \dots, n$. It remains to check that

$$h(\gamma\xi_i) = \varphi(\gamma\xi_i) \tag{20}$$

for every $\gamma \in \Lambda_{\underline{\xi}}$. Every $\gamma \in \Lambda_{\underline{\xi}}$ is a product $\gamma = r_k \cdot \ldots \cdot r_1$, where r_j is a reflection in a face of the regular simplex $r_{j-1} \cdot \ldots \cdot r_1(\underline{\xi})$. We prove the equality (20) by induction on k, the case $k = 0$ being true by assumption. Set $\eta_i = r_{k-1} \cdot \ldots \cdot r_1(\xi_i)$. By induction, we know that $h(\eta_i) = \varphi(\eta_i)$. We need to show that $h(r_k\eta_i) = \varphi(r_k\eta_i)$. The simplex $(\eta_0, \dots, \eta_n)$ is regular and r_k is a reflection in one of its faces, say the face containing $\eta_1, \dots, \eta_n$. Since $r_k\eta_i = \eta_i$ for $i = 1, \dots, n$, it just remains to show that $h(r_k\eta_0) = \varphi(r_k\eta_0)$. The simplex $(r_k\eta_0, r_k\eta_1, \dots, r_k\eta_n) = (r_k\eta_0, \eta_1, \dots, \eta_n)$ is regular with opposite orientation to $(\eta_0, \eta_1, \dots, \eta_n)$. This implies on the one hand that the simplex $(h(r_k\eta_0), h(\eta_1), \dots, h(\eta_n))$ is regular with opposite orientation to $(h(\eta_0), h(\eta_1), \dots, h(\eta_n))$, and on the other hand that the simplex $(\varphi(r_k\eta_0), \varphi(\eta_1), \dots, \varphi(\eta_n))$ is regular with opposite orientation to $(\varphi(\eta_0), \dots, \varphi(\eta_n))$. Since $(h(\eta_0), h(\eta_1), \dots, h(\eta_n)) = (\varphi(\eta_0), \dots, \varphi(\eta_n))$ and there is in dimension $n \geqslant 3$ only one regular simplex with face $h(\eta_1), \dots, h(\eta_n)$ and opposite orientation to $(h(\eta_0), h(\eta_1), \dots, h(\eta_n))$ it follows that $h(r_k\eta_0) = \varphi(r_k\eta_0)$. □

If φ were continuous, sending the vertices of all positively (respectively negatively) oriented ideal regular simplices to vertices of positively (resp. neg.) oriented ideal regular simplices, then it would immediately follow from the lemma that φ is equal to an isometry h on the orbits $\bigcup_{i=0}^n \Lambda_{\underline{\xi}}\xi_i$ of the vertices of one regular simplex under its reflection group. Since the set $\bigcup_{i=0}^n \Lambda_{\underline{\xi}}\xi_i$ is dense in $\partial\mathbb{H}^n$, the continuity of φ would imply that φ is equal to the isometry h on the whole $\partial\mathbb{H}^n$.

In the setting of the next proposition, we first need to show that there exist enough regular simplices for which φ maps every simplex of its orbit under reflections to a regular simplex. Second, we apply the lemma to obtain that φ is equal to an isometry on these orbits. Finally, we use ergodicity of the reflection groups to conclude that it is the same isometry for almost all regular simplices. As mentioned earlier, the proposition also holds for $n = 3$ (see [17, pp. 654-656] and [33, two last paragraphs of Sect. 6.4]), but in that case the proof is quite harder, since the reflection group of a regular simplex is discrete in $\mathrm{Isom}(\mathbb{H}^n)$ (indeed, one can tile $\mathbb{H}^3$ by regular ideal simplices) and in particular does not act ergodically on $\mathrm{Isom}(\mathbb{H}^n)$.

Proposition 4.7 *Let $n \geqslant 4$. Let $\varphi : \partial\mathbb{H}^n \to \partial\mathbb{H}^n$ be a measurable map sending the vertices of almost every positively, respectively negatively oriented regular ideal simplex to the vertices of a positively, resp. negatively, oriented regular ideal simplex. Then φ is equal almost everywhere to an isometry.*

Proof Let $T^\varphi \subset T$ denote the following subset of the set T of regular simplices:

$$T^\varphi = \big\{\underline{\xi} = (\xi_0, \dots, \xi_n) \in T \,\big|\, \big(\varphi(\xi_0), \dots, \varphi(\xi_n)\big) \text{ belongs to } T \text{ and has the same orientation as } (\xi_0, \dots, \xi_n)\big\}.$$

By assumption, T^φ has full measure in T. Let $T^\varphi_\Lambda \subset T^\varphi$ be the subset consisting of those regular simplices for which all reflections by the reflection group $\Lambda_{\underline{\xi}}$ are in T^φ,

$$T^\varphi_\Lambda = \big\{\underline{\xi} \in T \,\big|\, \gamma\underline{\xi} \in T^\varphi \ \forall \gamma \in \Lambda_{\underline{\xi}}\big\}.$$

We claim that T^φ_Λ has full measure in T.

To prove the claim, we do the following identification. Since $G = \mathrm{Isom}(\mathbb{H}^n)$ acts simply transitively on the set T of (oriented) regular simplices, given a base point $\underline{\eta} = (\eta_0, \dots, \eta_n) \in T$ we can identify G with T via the evaluation map

$$\begin{aligned} Ev_\eta : G &\longrightarrow T \\ g &\longmapsto g(\underline{\eta}). \end{aligned}$$

The subset T^φ is mapped to a subset $G^\varphi := (Ev_\eta)^{-1}(T^\varphi) \subset G$ via this correspondence. A regular simplex $\underline{\xi} = g(\underline{\eta})$ belongs to T^φ_Λ if and only if, by definition, $\gamma\underline{\xi} = \gamma g\underline{\eta}$ belongs to T^φ for every $\gamma \in \Lambda_{\underline{\xi}}$. Since $\Lambda_{\underline{\xi}} = g\Lambda_{\underline{\eta}}g^{-1}$, the latter condition is equivalent to $g\gamma_0\underline{\eta} \in T^\varphi$ for every $\gamma_0 \in \Lambda_{\underline{\eta}}$, or in other words, $g \in G^\varphi\gamma_0^{-1}$. The subset T^φ_Λ is thus mapped to

$$G^\varphi = Ev_{\underline{\eta}}^{-1}\big(T^\varphi_\Lambda\big) = \bigcap_{\gamma_0 \in \Lambda_{\underline{\eta}}} G^\varphi\gamma_0^{-1} \subset G$$

via the above correspondence. Since a countable intersection of full measure subsets has full measure, the claim is proved.

For every $\underline{\xi} \in T_\Lambda^\varphi$ and hence almost every $\underline{\xi} \in T$ there exists by Lemma 4.6 a unique isometry $h_{\underline{\xi}}$ such that $h_{\underline{\xi}}(\xi) = \varphi(\xi)$ on the orbit points $\xi \in \bigcup_{i=0}^n \Lambda_{\underline{\xi}} \xi_i$. By the uniqueness of the isometry, it is immediate that $h_{\gamma\underline{\xi}} = h_{\underline{\xi}}$ for every $\gamma \in \Lambda_{\underline{\xi}}$. We have thus a map $h : T \to \mathrm{Isom}(\mathbb{H}^n)$ given by $\underline{\xi} \mapsto h_{\underline{\xi}}$ defined on a full measure subset of T. Precomposing h by $Ev_{\underline{\eta}}$, it is straightforward that the left $\Lambda_{\underline{\xi}}$-invariance of h on $\Lambda_{\underline{\xi}}\underline{\xi}$ naturally translates to a global right invariance of $h \circ Ev_{\underline{\eta}}$ on G. Indeed, let $g \in G$ and $\gamma_0 \in \Lambda_{\underline{\eta}}$. We compute

$$h \circ Ev_{\underline{\eta}}(g \cdot \gamma_0) = h_{g\gamma_0\underline{\eta}} = h_{g\gamma_0 g^{-1} g\underline{\eta}} = h_{g\underline{\eta}} = h \circ Ev_{\underline{\eta}}(g),$$

where we have used the left $\Lambda_{g\underline{\eta}}$-invariance of h on the reflections of $g\eta$ in the third equality. (Recall, $g\gamma_0 g^{-1} \in g\Lambda_{\underline{\eta}} g^{-1} = \Lambda_{g\underline{\eta}}$.) Thus, $h \circ Ev_{\underline{\eta}} : G \to G$ is invariant under the right action of $\Lambda_{\underline{\eta}}$. Since the latter group is dense in G, it acts ergodically on G and $h \circ Ev_{\underline{\eta}}$ is essentially constant. This means that also h is essentially constant. Thus, for almost every regular simplex $\underline{\xi} \in T$, the evaluation of φ on any orbit point of the vertices of $\underline{\xi}$ under the reflection group $\Lambda_{\underline{\xi}}$ is equal to h. In particular, for almost every $\underline{\xi} = (\xi_0, \dots, \xi_n) \in T$ and also for almost every $\xi_0 \in \mathbb{H}^n$, we have $\varphi(\xi_0) = h(\xi_0)$, which finishes the proof of the proposition. □

We have now established that φ is essentially equal to the isometry $h \in \mathrm{Isom}(\mathbb{H}^n)$ on $\partial\mathbb{H}^n$. It remains to see that h realizes the conjugation between ρ and i. Indeed, replacing φ by h in (10) we have

$$\big(h \cdot i(\gamma)\big)(\xi) = \big(\rho(\gamma) \cdot h\big)(\xi),$$

for every $\xi \in \partial\mathbb{H}^n$ and $\gamma \in \Gamma$. Since all maps involved (h, $i(\gamma)$ and $\rho(\gamma)$) are isometries of $\mathbb{H}^n$ and two isometries induce the same map on $\partial\mathbb{H}^n$ if and only if they are equal it follows that

$$h \cdot i(\gamma) \cdot h^{-1} = \rho(\gamma)$$

for every $\gamma \in \Gamma$, which finishes the proof of the theorem.

Acknowledgements Michelle Bucher was supported by Swiss National Science Foundation project PP00P2-128309/1, Alessandra Iozzi was partially supported by Swiss National Science Foundation project 2000021-127016/2. The first and third named authors thank the Institute Mittag-Leffler in Djurholm, Sweden, for their warm hospitality during the preparation of this paper. Finally, we thank Beatrice Pozzetti for a thorough reading of the manuscript and numerous insightful remarks.

References

1. Bader, U., Furman, A., Sauer, R.: Integrable measure equivalence and rigidity of hyperbolic lattices. http://arxiv.org/abs/1006.5193v1
2. Besson, G., Courtois, G., Gallot, S.: Entropies et rigidités des espaces localement symétriques de courbure strictement négative. Geom. Funct. Anal. **5**, 731–799 (1995). MR 1354289 (96i:58136)

3. Besson, G., Courtois, G., Gallot, S.: Inégalités de Milnor-Wood géométriques. Comment. Math. Helv. **82**, 753–803 (2007). MR 2341839 (2009e:53055)
4. Bloch, S.J.: Higher Regulators, Algebraic K-theory, and Zeta Functions of Elliptic Curves. CRM Monograph Series, vol. 11. Am. Math. Soc., Providence (2000). MR 1760901 (2001i:11082)
5. Brooks, R.: Some remarks on bounded cohomology. In: Riemann Surfaces and Related Topics: Proceedings of the 1978 Stony Brook Conference, State Univ. New York, Stony Brook, NY, 1978. Ann. of Math. Stud., vol. 97, pp. 53–63. Princeton Univ. Press, Princeton (1981). MR 624804 (83a:57038)
6. Bucher-Karlsson, M.: The proportionality constant for the simplicial volume of locally symmetric spaces. Colloq. Math. **111**, 183–198 (2008). MR 2365796 (2008k:53105)
7. Bucher-Karlsson, M.: The simplicial volume of closed manifolds covered by $\mathbb{H}^2 \times \mathbb{H}^2$. J. Topol. **1**, 584–602 (2008). MR 2417444 (2009i:53025)
8. Bucher, M., Kim, I., Kim, S.: Proportionality principle for the simplicial volume of families of **Q**-rank 1 locally symmetric spaces. arXiv:1209.4697
9. Bucher, M., Monod, N.: The norm of the Euler class. Math. Ann. **353**, 523–544 (2012)
10. Burger, M., Iozzi, A.: Boundary maps in bounded cohomology. Geom. Funct. Anal. **12**(2), 281–292 (2002). MR 1911668 (2003d:53065b). Appendix to: Burger, M., Monod, M.: Geom. Funct. Anal. **12**, 219–280 (2002). MR 1911660 (2003d:53065a)
11. Burger, M., Iozzi, A.: A useful formula from bounded cohomology. In: Géométries à courbure négative ou nulle, groupes discrets et rigidités. Sémin. Congr., vol. 18, pp. 243–292. Soc. Math. France, Paris (2009). MR 2655315 (2011m:22017)
12. Burger, M., Iozzi, A., Wienhard, A.: Surface group representations with maximal Toledo invariant. C. R. Math. Acad. Sci. Paris **336**(5), 387–390 (2003). MR 1979350 (2004e:53076)
13. Burger, M., Iozzi, A., Wienhard, A.: Surface group representations with maximal Toledo invariant. Ann. Math. (2) **172**(1), 517–566 (2010). MR 2680425
14. Calabi, E., Vesentini, E.: Sur les variétés complexes compactes localement symétriques. Bull. Soc. Math. Fr. **87**, 311–317 (1959). MR 0111057 (22 #1922a)
15. Calabi, E., Vesentini, E.: On compact, locally symmetric Kähler manifolds. Ann. Math. **71**, 472–507 (1960). MR 0111058 (22 #1922b)
16. Douady, A., Earle, C.J.: Conformally natural extension of homeomorphisms of the circle. Acta Math. **157**, 23–48 (1986). MR 857678 (87j:30041)
17. Dunfield, N.M.: Cyclic surgery, degrees of maps of character curves, and volume rigidity for hyperbolic manifolds. Invent. Math. **136**, 623–657 (1999). MR 1695208 (2000d:57022)
18. Francaviglia, S.: Hyperbolic volume of representations of fundamental groups of cusped 3-manifolds. Int. Math. Res. Not. **9**, 425–459 (2004). MR 2040346 (2004m:57032)
19. Francaviglia, S., Klaff, B.: Maximal volume representations are Fuchsian. Geom. Dedic. **117**, 111–124 (2006). MR 2231161 (2007d:51019)
20. Goldman, W.M.: Discontinuous groups and the Euler class. ProQuest LLC, Ann Arbor, MI, 1980. Ph. D. Thesis, University of California, Berkeley. MR 2630832
21. Gromov, M.: Volume and bounded cohomology. Inst. Hautes Études Sci. Publ. Math. **56**, 5–99 (1982). MR 686042 (84h:53053)
22. Guichardet, A.: Cohomologie des Groupes Topologiques et des Algèbres de Lie. Textes Mathématiques [Mathematical Texts], vol. 2. CEDIC, Paris (1980). MR 644979 (83f:22004)
23. Haagerup, U., Munkholm, H.J.: Simplices of maximal volume in hyperbolic n-space. Acta Math. **147**, 1–11 (1981). MR 631085 (82j:53116)
24. Iozzi, A.: Bounded cohomology, boundary maps, and rigidity of representations into $\mathrm{Homeo}_+(\mathbf{S}^1)$ and $\mathrm{SU}(1,n)$. In: Rigidity in Dynamics and Geometry, Cambridge, 2000, pp. 237–260. Springer, Berlin (2002). MR 1919404 (2003g:22008)
25. Kneser, H.: Die kleinste Bedeckungszahl innerhalb einer Klasse von Flächenabbildungen. Math. Ann. **103**, 347–358 (1930). MR 1512626
26. Mackey, G.W.: Induced representations of locally compact groups. I. Ann. Math. (2) **55**, 101–139 (1952). MR 0044536 (13,434a)

27. Monod, N.: Continuous Bounded Cohomology of Locally Compact Groups. Lect. Notes in Mathematics, vol. 1758. Springer, Berlin (2001). MR 1840942 (2002h:46121)
28. Mostow, G.D.: Quasi-conformal mappings in n-space and the rigidity of hyperbolic space forms. Inst. Hautes Études Sci. Publ. Math. **34**, 53–104 (1968). MR 0236383 (38 #4679)
29. Prasad, G.: Strong rigidity of **Q**-rank 1 lattices. Invent. Math. **21**, 255–286 (1973). MR 0385005 (52 #5875)
30. Reiter, H., Stegeman, J.D.: Classical Harmonic Analysis and Locally Compact Groups, London Mathematical Society Monographs, New Series, vol. 22, 2nd edn. Clarendon, Oxford University Press, New York (2000). MR 1802924 (2002d:43005)
31. Rudin, W.: Real and Complex Analysis, 3rd edn. McGraw-Hill, New York (1987). MR 924157 (88k:00002)
32. Selberg, A.: On discontinuous groups in higher-dimensional symmetric spaces. In: Contributions to Function Theory, Internat. Colloq. Function Theory, Bombay, 1960, pp. 147–164. Tata Institute of Fundamental Research, Bombay (1960). MR 0130324 (24 #A188)
33. Thurston, W.: Geometry and Topology of 3-Manifolds. Notes from Princeton University, Princeton (1978)
34. Toledo, D.: Representations of surface groups in complex hyperbolic space. J. Differ. Geom. **29**, 125–133 (1989). MR 978081 (90a:57016)
35. Weil, A.: On discrete subgroups of Lie groups. Ann. Math. (2) **72**, 369–384 (1960). MR 0137792 (25 #1241)
36. Weil, A.: On discrete subgroups of Lie groups. II. Ann. Math. (2) **75**, 578–602 (1962). MR 0137793 (25 #1242)
37. Zimmer, R.J.: Ergodic Theory and Semisimple Groups. Monographs in Mathematics, vol. 81. Birkhäuser, Basel (1984). MR 776417 (86j:22014)

The Algebras Generated by the Laplace Operators in a Semi-homogeneous Tree

Enrico Casadio Tarabusi and Massimo A. Picardello

Abstract In a semi-homogeneous tree, the set of edges is a transitive homogeneous space of the group of automorphisms, but the set of vertices is not (unless the tree is homogeneous): in fact, the latter splits into two disjoint homogeneous spaces V_+, V_- according to the homogeneity degree. With the goal of constructing maximal abelian convolution algebras, we consider two different algebras of radial functions on semi-homogeneous trees. The first consists of functions on the vertices of the tree: in this case the group of automorphisms gives rise to a convolution product only on V_+ and V_- separately, and we show that the functions on V_+, V_- that are radial with respect to the natural distance form maximal abelian algebras, generated by the respective Laplace operators. The second algebra consists of functions on the edges of the tree: in this case, by choosing a reference edge, we show that no algebra that contains an element supported on the disc of radius one is radial, not even in a generalized sense that takes orientation into account. In particular, the two Laplace operators on the edges of a semi-homogeneous (non-homogeneous) tree do not generate a radial algebra, and neither does any weighted combination of them. It is also worth observing that the convolution for functions on edges has some unexpected properties: for instance, it does not preserve the parity of the distance, and the two Laplace operators never commute, not even on homogeneous trees.

Keywords Semi-homogeneous trees · Laplace operators · Radial functions

Mathematics Subject Classification (2010) Primary 05C05 · Secondary 43A85

E. Casadio Tarabusi
Dipartimento di Matematica "G. Castelnuovo", Sapienza Università di Roma, Piazzale A. Moro 5, 00185 Rome, Italy
e-mail: casadio@mat.uniroma1.it

M.A. Picardello (✉)
Dipartimento di Matematica, Università di Roma "Tor Vergata", Via della Ricerca Scientifica, 00133 Rome, Italy
e-mail: picard@mat.uniroma2.it

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3,
DOI 10.1007/978-88-470-2853-1_5,

1 Introduction

This short note addresses a question naturally connected with the work of Alessandro Figà-Talamanca in harmonic analysis on homogeneous trees. Indeed, his work revealed a strong analogy between geometry and harmonic analysis on trees and on Riemannian symmetric spaces of rank one [3] (previous progress in this direction, related to the Radon transform, was made in [1]).

A deeper understanding of this analogy can be achieved by regarding trees as p-adic symmetric spaces [5, 7]. In this setting, semi-homogeneous trees are the Bruhat-Tits buildings of rank one. In general, a Bruhat-Tits building of higher rank n is a simplicial complex where a p-adic semisimple group acts naturally. The geometry of this action is used to study harmonic analysis and representation theory of the acting group, and therefore it gives rise to a natural similarity with harmonic analysis for semisimple Lie groups of rank n. In much the same way, the results developed in [3] for some discrete groups acting on trees can be regarded as analogous to harmonic analysis on rank-one semisimple groups.

An important tool for harmonic analysis on a semisimple group G is the maximal abelian algebra R of invariant differential operators. In the rank-one case this algebra has one generator, namely the Laplace operator. The multiplicative functionals on R are the spherical functions of the Gelfand pair (G, K), where K is the maximal compact subgroup of rotations. The algebra R is isomorphic to the algebra of bi-K-invariant functions on G, that is, radial functions on the homogeneous space G/K. Spherical functions are coefficients of unitary or uniformly bounded representations of G. Spherical functions turn out to be eigenfunctions of the Laplace operator; the eigenvalue provides an explicit parametrization and yields a construction of irreducible unitary representations of G. On semisimple groups of rank n there are n independent commuting invariant differential operators; the commutative algebra that they generate provides a parametrization with n complex parameters, and in some cases an explicit construction, of irreducible unitary representations of the group.

In order to develop harmonic analysis and representation theory for free groups and other groups acting on a homogeneous tree T, the tool used by Figà-Talamanca and others (see [1] and further references in [3]) is the (maximal abelian) convolution algebra of summable radial functions on the vertices of T. This algebra, denoted again by R, is defined provided a reference vertex o is chosen as origin, and is generated by the *Laplace operator* of T, that is, by the characteristic function of vertices adjacent to o.

On the other hand, if T is semi-homogeneous but not homogeneous, the set V of its vertices has two orbits V_+, V_- under the action of the group of automorphisms, and each orbit carries a convolution algebra of radial functions (with respect to the respective reference vertex chosen as origin; see [2] for more details). Therefore there are two Laplace operators, each defined by the characteristic function of the set of vertices adjacent to the respective origin; but we shall see that the algebras generated by these two functions have distinct convolution products, and do not embed into a single commutative superalgebra.

We are grateful to Simon Gindikin, who raised the question answered here: whether there exists any non-trivial radial convolution algebra of functions defined on edges of semi-homogeneous trees. We were not able to find similar results in the literature (for related research on buildings, see [6]).

2 Preliminaries and Notation

Let T be a tree, that is, a locally finite, connected graph without loops. Let V be the countable set of vertices of T. The set E of (non-oriented) edges is a proper (except in trivial cases) subset of $V \times V$, such that any pair $(u, v) \in E$ satisfies $u \neq v$ and no non-trivial loops occur. The fact that the edges are unoriented means that we identify (u, v) with (v, u). We say that $u, v \in V$ are *adjacent* if $(u, v) \in E$.

The *degree* (or *homogeneity*) of $v \in V$ is the number of edges that contain v less 1. On T we fix a *parity*, namely a function $\varepsilon : V \to \{+1, -1\}$ such that if $(u, v) \in E$ then $\varepsilon(u) = -\varepsilon(v)$. The group $\mathrm{Aut}(T)$ of *automorphisms* of T, the invertible self-maps of V that preserve adjacency, is transitive on E if and only if T is *semi-homogeneous*, that is, there are non-negative integers q_+, q_- such that the degree of every $v \in V$ is $q_{\varepsilon(v)}$; we shall write $T = T_{q_+,q_-}$ and assume $q_+ \geqslant q_- > 0$ to avoid trivialities. The tree T is *homogeneous* when $q_+ = q_-$, because this is exactly when $\mathrm{Aut}(T)$ acts transitively on V as well. On the other hand, if $q_+ \neq q_-$ then every automorphism must preserve the parity, so the two level sets of ε are the orbits of $\mathrm{Aut}(T)$ on V. This gives rise to an orientation on the edges. On a semi-homogeneous, non-homogeneous tree $\mathrm{Aut}(T)$ preserves this orientation.

Two edges e, e' are *adjacent* if they share exactly one vertex. In general, the *chain* of edges from e to e' is the minimal finite sequence $e = e_0, e_1, \dots, e_n = e'$ in E such that e_{j-1}, e_j are adjacent for every $j = 1, \dots, n$. The (plain) *distance* of e, e' is the non-negative integer n. We refine it to their *signed (or oriented) distance* $\mathrm{dist}(e, e') := \varepsilon(v)n$, where $e_0 \cap e_1 = \{v\}$ if $n > 0$ (note that in general $\mathrm{dist}(e, e') \neq \mathrm{dist}(e', e)$). The *circle* $C_n(e)$ in E of center $e \in E$ and radius a non-negative integer n is the set of edges $e' \in E$ such that $|\mathrm{dist}(e, e')| = n$. In particular, $C_0(e) = \{e\}$. If $n > 0$, the set $C_n(e)$ splits into two disjoint *half-circles* $S_{+n}(e)$, $S_{-n}(e)$ (of *signed radii* $+n$, $-n$ respectively) according to the signed distance from e. Their cardinalities (independent of e) are easily seen to be

$$\bigl|S_j(e)\bigr| = \begin{cases} (q_+q_-)^{|j|/2} & \text{for even } j \neq 0, \\ q_+^{(j+1)/2} q_-^{(j-1)/2} & \text{for odd } j > 0, \\ q_+^{(-j-1)/2} q_-^{(-j+1)/2} & \text{for odd } j < 0, \end{cases} \tag{1}$$

so that $|S_{-j}(e)| = |S_j(e)|$ only for j even or T homogeneous.

Fix $e_0 \in E$. In the specification of a circle or a half-circle, if the center is omitted it will be intended to be e_0. Assume that a function M on E only depends on the signed distance, that is,

$$M(e) = M\bigl(\mathrm{dist}(e_0, e)\bigr) \quad \text{for every } e \in E.$$

The convolution of M with a function f on E is given by

$$M * f(e) = \sum_{e' \in E} M\big(\operatorname{dist}(e, e')\big) f\big(e'\big)$$
$$= \sum_{n=-\infty}^{+\infty} M(n) \sum_{e' \in S_n(e)} f\big(e'\big) \quad \text{for every } e \in E,$$

for instance if M, f are ℓ^2-summable with respect to the counting measure on E. We shall also let M stand for the convolution operator $f \mapsto M * f$.

3 Commutative Algebras of Functions on the Vertices of a Semi-homogeneous Tree

In this section we consider the convolution algebras on the vertices of a semi-homogeneous tree, as well as the geometry of vertices and associated graphs.

In order to construct commutative algebras of functions on V we must restrict attention to proper subsets of V. Indeed, let us consider the two subsets V_+, respectively V_-, of vertices of homogeneity q_+, respectively q_-, and equip each with a graph structure as follows: two vertices in the graph $V_\pm$ are adjacent if both are adjacent in the tree T to the same vertex of $V_\mp$. The set $\Gamma_\pm(v)$ of vertices adjacent in T to a given vertex $v \in V_\mp$ forms a complete subgraph of $V_\pm$ of cardinality $q_\mp + 1$, therefore we shall call V_+, V_- the *locally complete graphs* arising from T. Figures 1 and 2 show the locally complete graphs V_+, respectively V_-, for the tree $T_{3,2}$. These graphs are obtained by taking complete graphs on sets of vertices of cardinality $q_\mp + 1$, that we call *cells*, and connecting them appropriately.

Another interesting graph can be associated to a semi-homogeneous tree as follows. In each cell we select a subgraph isomorphic to a cycle of $q_\mp + 1$ edges, in such a way that each vertex of the cell belongs to the cycle: that is, we choose a path that visits every vertex once and remove the remaining edges. The corresponding graph is a polygonal graph, as defined in [4]: it is obtained by starting with a regular polygon with n sides, joining new regular polygons of m sides at its vertices, and iterating this process with alternating values of m and n. Polygonal graphs from V_+, V_- respectively, arising in this way are called *polygonal graphs associated with the semi-homogeneous tree*. Each of these graphs has homogeneous cells, that is, its cells are polygons of the same number of sides (either three or four in our example). Since a triangle is both a polygon and a complete graph, the polygonal graph associated with vertices of homogeneity 3 coincides with the locally complete graph already shown in Fig. 1; a polygonal graph associated with vertices of homogeneity 4 is shown in Fig. 3.

Remark 3.1 We have just shown how to associate two locally complete graphs V_+ and V_- and two polygonal graphs to a semi-homogeneous tree $T = T_{q_+,q_-}$. Conversely, T can be reconstructed as the *barycentric subdivision* of any of these graphs,

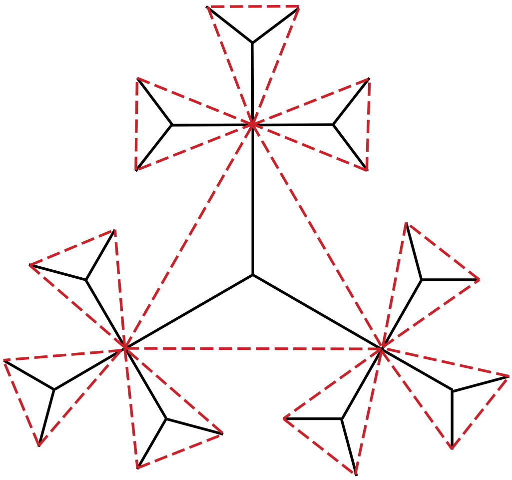


Fig. 1 The semi-homogeneous tree $T_{3,2}$ (*solid edges*) and the locally complete graph V_+ (*dashed edges*). The tree is the barycentric subdivision of the graph

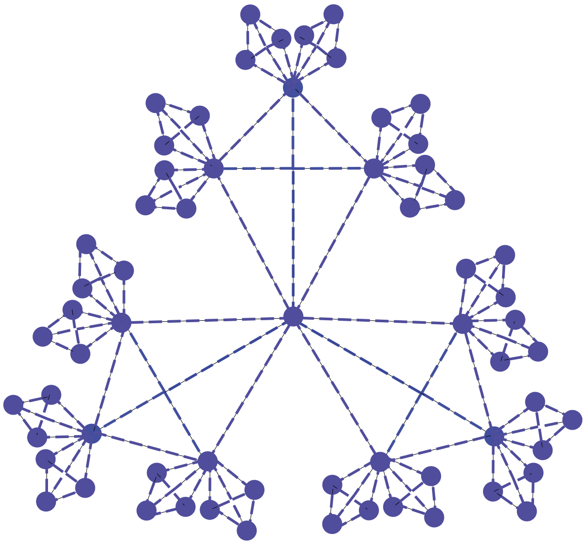


Fig. 2 The locally complete graph V_-. All *dashed edges* in this drawing belong to V_-: its cells are complete subgraphs spanned by polygons of q_+ sides. All its vertices are represented as *dots*: the intersection points of internal edges are not vertices. The edges of the tree $T_{3,2}$, not drawn here, coincide with some of the edges of V_- (see next figure)

namely, the graphs obtained by the following three steps. First we insert into each cell a new vertex, then we join such a new vertex with each old vertex of that cell, and finally we remove all the old edges (but not the old vertices: see Fig. 1).

Instead, each locally complete graph associated with T can be obtained as the *dual graph* of the other, or of the polygonal graph of the opposite sign. Such dual graph is constructed by replacing each cell by a vertex, and considering two such vertices adjacent if their originating cells share an old vertex. Since n cells join at each old vertex (here $n = q_+ + 1$ or $q_- + 1$), this procedure replaces every old vertex by a complete subgraph of n vertices.

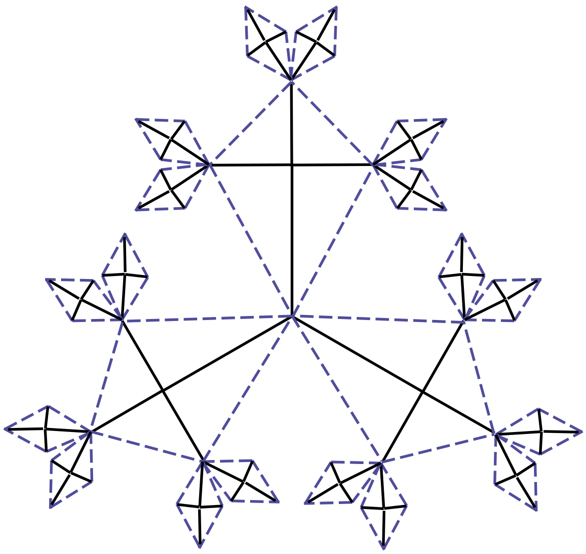


Fig. 3 The semi-homogeneous tree $T_{3,2}$ (*solid edges*) and a polygonal graph extracted from V_+ (*dashed edges*)

By definition, the subsets V_+, V_- are exactly the full sets of vertices of the associated locally complete graphs. Therefore the group of automorphisms of the tree T_{q_+,q_-} is transitive both on the vertices of these two graphs and on their sets of edges.

Remark 3.2 Indeed, more is true: automorphisms of T_{q_+,q_-} are also automorphisms of these two graphs. This is because the edges of the locally complete graphs associated with the tree correspond in a one-to-one way to pairs of vertices of the tree of the same homogeneity and at distance two. Automorphisms of the tree preserve both homogeneity and distance, so they also act on edges of the locally complete graphs. It is also clear that this action is transitive.

Let us denote by G the group of automorphisms of T_{q_+,q_-}, and by $K_\pm$ the isotropy subgroup at a fixed vertex of homogeneity $q_\pm$. As $V_\pm$ is a transitive homogeneous space for G, this group induces a convolution operation on functions on $V_\pm$; in particular, it gives rise to a convolution product on functions on a corresponding polygonal graph associated with the tree. For functions on such a graph there is a natural notion of radiality, introduced in [4] in terms of the *block length distance* from a fixed reference vertex (the block length of every vertex is the number of cells crossed by the minimal path to the reference vertex). If we choose a reference vertex $o_\pm \in V_\pm$, the block length of a vertex v in $V_\pm$ has a natural reinterpretation in terms of the geometry of T_{q_+,q_-}: it is the number of vertices in $V_\mp$ visited by the minimal path from v to $o_\pm$. We shall call this length the *distance* in $V_\pm$. Functions on $V_\pm$ that depend only on the distance are called *radial*. So there are two different spaces of radial functions, one on V_+ and the other on V_-; these two spaces are convolution algebras (under the respective convolution products), indeed they are the algebras generated by the two Laplace operators of the semi-homogeneous tree mentioned at the end of Sect. 1.

It was proved in [4, Corollary 1] that the radial functions on a polygonal graph form a maximal abelian algebra under convolution. Therefore the following statement holds:

Proposition 3.3 *The algebra of radial functions on vertices in* V_+ (*respectively,* V_-) *is maximal abelian.*

4 Commutative Algebras of Functions on the Edges of a Semi-homogeneous Tree

As we have seen, the fact that the group of automorphisms of a semi-homogeneous, non-homogeneous tree is not transitive on vertices does not allow a natural definition of a convolution algebra on all vertices (see [2] for some advances in this direction). So, at first glance, the appropriate way to study convolution algebras on all of the semi-homogeneous tree T (and not just on a subset of it) should be to restrict attention to functions defined on the edges of T. The group of automorphisms acts transitively on $E(T)$, and so it gives rise to a satisfactory convolution product on functions thereon. Since semi-homogeneous trees are (all of) the Bruhat-Tits buildings of rank one, we expect that, once a reference edge e is chosen, the characteristic function of edges adjoining e should be a *Laplace operator* which generates a maximal abelian convolution algebra, whose multiplicative functional should parameterize at least a large class of irreducible unitary, or uniformly bounded, representations of the group of automorphisms.

Here we made use of the natural geometric notion of radiality in terms of distance from e. However, we may want to introduce a more sophisticated definition, inspired by a polarization in $E(T)$. Indeed, the group of automorphisms of T preserves edge orientation, which can be defined positive from its vertex of homogeneity q_- to the vertex of homogeneity q_+. Nevertheless, choosing a reference edge e induces a different orientation on $E(T) \setminus \{e\}$: the outgoing orientation. For some edges the two orientations coincide: in particular, this happens for the edges which join e at its vertex in V_-; for the others, in particular the adjoining edges at the vertex in V_+, the two orientations disagree. Thus, circles around e split in two subsets, which correspond to a *positive* and a *negative* half-circle. Now we see that the characteristic function of edges at distance one from e cannot be expected to generate a radial convolution algebra; perhaps we should instead consider separately the sets of positively and negatively oriented edges at distance one, and define the Laplacian as a suitable linear combination of their characteristic functions. This way we obtain an oriented definition of radiality, such that, in each circle C_n of edges at distance n from e, a radial function may not be constant; instead, it might attain two different values on the two half-circles S_n, S_{-n}. Then the definition of radiality would in general be given in terms of a sequence of appropriate positive weights w_n: a function is radial if the ratio of its values on S_n and S_{-n} equals w_n for all n.

Surprisingly enough, such expectations fail: we show in Theorem 4.2 that not only the characteristic function of C_1 fails to generate a radial algebra in the natural geometric sense, but also that no linear combination of the characteristic functions of S_1 and S_{-1} generates a radial algebra in the polarized sense, unless T is homogeneous. In other words, radial functions with support on C_1 do not generate radial convolution algebras.

This leads to a more general question, that will be addressed in future work: if T is semi-homogeneous but not homogeneous, does there exist any non-trivial radial convolution algebra? Our theorem states that no such algebra may contain a function whose support has radius 1. We conjecture that no radial algebra may contain a function of finite support and give hints for the validity of the general case.

We recall the notation established at the end of Sect. 2. Denoting by χ_S the characteristic function of a subset $S \subset E$, for each relative integer j we consider the function (or convolution kernel)

$$m_j = \begin{cases} \chi_{C_0} = \chi_{\{e_0\}} & \text{if } j = 0, \\ \chi_{S_j} & \text{if } j \neq 0. \end{cases}$$

Then our definition of radiality becomes the following:

Definition 4.1 (Radial algebras of functions on edges) Given a sequence $\{w_n\}$ of positive weights, a function M on the edges of T is *radial* (*in the oriented sense*) *with respect to the weights* $\{w_n\}$ if for each positive n it is equal to the constant $M_{\pm n}$ on the half-circle $S_{\pm n}$, where either $M_n = M_{-n} = 0$, or $M_n/M_{-n} = w_n$ (in particular, M_n and M_{-n} have the same sign). The algebra generated by M is radial with respect to the weights $\{w_n\}$ if, for all positive integers k, h, n, the convolution powers M^h and M^k satisfy the requirement that

$$\frac{(M^h)_n}{(M^h)_{-n}} = \frac{(M^k)_n}{(M^k)_{-n}}$$

(unless $(M^h)_n = (M^h)_{-n} = 0$ or $(M^k)_n = (M^k)_{-n} = 0$), which we can in general rewrite as

$$\big(M^h\big)_n \big(M^k\big)_{-n} = \big(M^k\big)_n \big(M^h\big)_{-n}. \tag{2}$$

Theorem 4.2 *For every $r_{-1}, r_0, r_1 \in \mathbb{R}$ such that r_{-1}, r_1 are not both zero the operator*

$$M = r_{-1} m_{-1} + r_0 m_0 + r_1 m_1$$

does not generate a radial convolution algebra, unless $q_+ = q_-$ and $r_1 = r_{-1}$. In particular this holds for $n = 1$ and $r_{-1} q_- + r_{+1} q_+ = -r_0 = 1$, in which case M is a weighted Laplace operator *on E.*

Proof If one of the coefficients r_{-1} and r_1 is zero, the other must vanish too, because the algebra generated by M would not otherwise be radial. Thus we can assume that they are both non-zero. In order to compute the powers of M, we first check that the following three identities hold:

$$m_{+1}^2 = q_+ m_0 + (q_+ - 1)m_{+1}, \tag{3}$$

$$m_{-1}^2 = q_- m_0 + (q_- - 1)m_{-1}, \tag{4}$$

whereas

$$m_j = \begin{cases} (m_{+1}m_{-1})^{j/2} & \text{for even } j > 0, \\ (m_{-1}m_{+1})^{-j/2} & \text{for even } j < 0, \\ (m_{+1}m_{-1})^{(j-1)/2}m_{+1} & \text{for odd } j > 0, \\ (m_{-1}m_{+1})^{(-j-1)/2}m_{-1} & \text{for odd } j < 0. \end{cases} \tag{5}$$

Indeed, let us compute m_{+1}^2. This is the iteration of two copies of the sum over all forward neighboring (oriented) edges of a given edge in the positive direction. Let us start with any given edge e. Since each neighboring edge e' has the opposite direction, the forward neighbors of e' include e and all the forward neighbors of e except e'; identity (3) follows directly from this remark, and (4) is analogous. To prove the remaining identity, that is (5), let us consider, for example, $m_{-1}m_{+1}$. Now, we first apply the operator m_{-1} of sum over the forward neighbor edges of e, and then the backward sum over all neighbors of such forward neighbors. Because of orientation reversal when passing to neighbors, this is the same as summing over all neighbors of e at distance two in the positive direction. This proves one part of (5), and the other parts follow in the same way.

It is not restrictive to assume that $r_0 = 0$, since m_0 is the identity operator. For the sake of clarity we shall limit the argument to the case $n = 1$, and abbreviate r_{-1}, r_{+1} by r_-, r_+, respectively (the general case is handled in a similar way but computations are more complicated). So r_+ and r_- are both non-zero, therefore, as observed in Definition 4.1, they share the same sign. By the previous identities we find

$$\begin{aligned} M^2 &= \big(r_+^2 q_+ + r_-^2 q_-\big)m_0 \\ &\quad + r_+^2(q_+ - 1)m_{+1} + r_-^2(q_- - 1)m_{-1} \\ &\quad + r_+ r_-(m_{+2} + m_{-2}), \\ M^3 &= \big(r_+^3 q_+(q_+ - 1) + r_-^3 q_-(q_- - 1)\big)m_0 \\ &\quad + \big(r_+^3\big(q_+^2 - q_+ + 1\big) + 2r_+ r_-^2 q_-\big)m_{+1} \\ &\quad + \big(r_-^3\big(q_-^2 - q_- + 1\big) + 2r_- r_+^2 q_+\big)m_{-1} \\ &\quad + \big(r_+^2 r_-(q_+ - 1) + r_+ r_-^2(q_- - 1)\big)(m_{+2} + m_{-2}) \\ &\quad + r_+^2 r_- m_{+3} + r_-^2 r_+ m_{-3}. \end{aligned}$$

If two convolution operators belong to the same radial convolution algebra, then for each positive integer j their coefficients of the terms m_{+j}, m_{-j} must either have the same fixed ratio, say w_j, or must both vanish. Imposing (2) for $n = h = 1$ and $k = 2$, respectively $k = 3$, we obtain

$$r_- r_+^2 (q_+ - 1) = r_+ r_-^2 (q_- - 1), \tag{6}$$

$$r_- \big(r_+^3 \big(q_+^2 - q_+ + 1 \big) + 2 r_+ r_-^2 q_- \big) = r_+ \big(r_-^3 \big(q_-^2 - q_- + 1 \big) + 2 r_- r_+^2 q_+ \big). \tag{7}$$

If $q_+ = q_- = 1$ then (7) simplifies to $r_-^2 = r_+^2$, that implies $r_- = r_+$ (the standard radiality of the homogeneous setting). If $q_- = 1 < q_+$, then (6) cannot hold, since $r_+, r_- \neq 0$. On the other hand, if $q_+, q_- > 1$, it follows from (7) that

$$r_+^2 \big(q_+^2 - 3q_+ + 1 \big) = r_-^2 \big(q_-^2 - 3q_- + 1 \big).$$

From (6) and (7) we have

$$\frac{q_-^2 - 3q_- + 1}{q_+^2 - 3q_+ + 1} = \frac{r_+^2}{r_-^2} = \frac{(q_- - 1)^2}{(q_+ - 1)^2}$$

for every r_+, r_-. By equating the left- and right-hand sides and observing that $(q_-^2 + 1)(q_+ - 1)^2 - (q_+^2 + 1)(q_- - 1)^2 = 2q_-(q_+^2 + 1) - 2q_+(q_-^2 + 1)$, we obtain

$$\frac{q_-}{(q_- - 1)^2} = \frac{q_+}{(q_+ - 1)^2}. \tag{8}$$

Since the function $t \mapsto t/(1+t)^2$ is strictly monotone for $t > 0$, this implies $q_- = q_+$ regardless of the values of r_-, r_+. Finally, by (6), we have $r_+ = r_-$. □

This unexpected behavior of algebras generated by a radial function M is probably not limited to the case where M is supported on edges at distance 1 from the reference edge. Since on any given M of finite support we can impose countably many radiality conditions (2), we expect that such M does not generate a radial algebra unless the tree is homogeneous (i.e., $q_+ = q_-$) and $r_j = r_{-j}$ for all j. Therefore:

Conjecture 4.3 *Let n be a positive integer, and let $(r_j)_{j=-n,\dots,+n}$ be a finite sequence of real numbers such that for each $j = 1, \dots, n$ the terms r_{-j}, r_{+j} are both positive, or both negative, or (if $j < n$) both zero. Then the operator*

$$M = \sum_{j=-n}^{+n} r_j m_j$$

does not generate a radial convolution algebra, unless $q_+ = q_-$ and $r_j = r_{-j}$ for every j.

Next we give evidence for this conjecture in the case the support is contained in the disc of radius 2, that is,

$$M = r_{-2}m_{-2} + r_{-1}m_{-1} + r_0m_0 + r_1m_1 + r_2m_2$$

but again, without loss of generality, from now on we set $r_0 = 0$. First, we rewrite (1) by setting $q_n = |S_n(e)|$:

$$q_n = \begin{cases} q_+^k q_-^k & \text{if } |n| = 2k \geqslant 0, \\ q_+^{k+1} q_-^k & \text{if } n = 2k+1 > 0, \\ q_+^k q_-^{k+1} & \text{if } n = -2k-1 < 0. \end{cases}$$

The multiplication rules between the functions m_n follow implicitly from (1), (3), (4), and (5):

$$m_i m_j = \begin{cases} m_{i+j} & \text{if } i = 0, \text{ or } i > 0 \text{ even and } j \geqslant 0, \text{ or } i < 0 \text{ even and } j \leqslant 0; \\ q_j m_{i+j} + \sum_{k=1}^{-j} (q_{j+k-1} - q_{j+k}) m_{i+j+2k-1} & \text{if } i > 0 \text{ even and } 0 > j \geqslant -i; \\ q_{-i} m_{i+j} + \sum_{k=1}^{i} (q_{-i+k-1} - q_{-i+k}) m_{-i-j+2k-1} & \text{if } i > 0 \text{ even and } j \leqslant -i; \\ q_j m_{i+j} + \sum_{k=1}^{j} (q_{j-k+1} - q_{j-k}) m_{i+j-2k+1} & \text{if } i < 0 \text{ even and } 0 < j \leqslant -i; \\ q_{-i} m_{i+j} + \sum_{k=1}^{-i} (q_{-i-k+1} - q_{-i-k}) m_{-i-j-2k+1} & \text{if } i < 0 \text{ even and } j \geqslant -i; \\ m_{i-j} & \text{if } i > 0 \text{ odd and } j \leqslant 0, \text{ or } i < 0 \text{ odd and } j \geqslant 0; \\ q_j m_{i-j} + \sum_{k=1}^{j} (q_{j-k+1} - q_{j-k}) m_{i-j+2k-1} & \text{if } i > 0 \text{ odd and } 0 < j \leqslant i; \\ q_i m_{i-j} + \sum_{k=1}^{i} (q_{i-k+1} - q_{i-k}) m_{-i+j+2k-1} & \text{if } i > 0 \text{ odd and } j \geqslant i; \\ q_j m_{i-j} + \sum_{k=1}^{-j} (q_{j+k-1} - q_{j+k}) m_{i-j-2k+1} & \text{if } i < 0 \text{ odd and } 0 > j \geqslant i; \\ q_i m_{i-j} + \sum_{k=1}^{-i} (q_{i+k-1} - q_{i+k}) m_{-i+j-2k+1} & \text{if } i < 0 \text{ odd and } j \leqslant i. \end{cases} \tag{9}$$

Now we see that

$$\begin{aligned}\left(M^2\right)_{-2} &= r_{-1}r_1 + r_{-1}r_{-2}(q_- - 1) + r_1 r_{-2}(q_+ - 1),\\ \left(M^2\right)_2 &= r_{-1}r_1 + r_1 r_2(q_+ - 1) + r_{-1}r_2(q_- - 1).\end{aligned} \tag{10}$$

Then the radiality condition (2) for $n = h = 2$ and $k = 1$ gives

$$r_{-2}\big(r_{-1}r_1 + r_1(q_+ - 1) + r_{-1}(q_- - 1)\big) = r_2\big(r_{-1}r_1 + r_1(q_+ - 1) + r_{-1}(q_- - 1)\big),$$

hence $r_{-2} = r_2$ unless $r_{-1} = r_1 = 0$.

Let us deal with the case $r_{-1} = r_1 = 0$, by checking the radiality condition (2) for $n = 1$, $k = 2$, $h = 3$ (the power M^1 vanishes on $C_{\pm 2}(e)$). Now we find

$$\begin{aligned}\left(M^2\right)_{-1} &= r_{-2}r_2(q_- - 1)q_+,\\ \left(M^2\right)_1 &= r_{-2}r_2(q_+ - 1)q_-,\\ \left(M^3\right)_{-1} &= \left(r_{-2}^2 r_2 + r_{-2}r_2^2\right)(q_+ - 1)q_+ q_-,\\ \left(M^3\right)_1 &= \left(r_{-2}^2 r_2 + r_{-2}r_2^2\right)(q_- - 1)q_- q_+.\end{aligned}$$

Thus, (2) yields $q_+(q_- - 1)^2 = q_-(q_+ - 1)^2$, that is, again (8), and it follows again $q_+ = q_-$. Let us write $q_+ = q_- = q$: but then, (10) yields $(M^2)_2 = (M^2)_{-2} = 0$, and by the multiplication rules we see that

$$\begin{aligned}\left(M^3\right)_2 &= r_{-2}r_2^2 q\left(2q^2 - q + 2\right),\\ \left(M^3\right)_{-2} &= r_2 r_{-2}^2 q\left(2q^2 - q + 2\right).\end{aligned}$$

Since the polynomial $2q^2 - q + 2$ has no real roots, it follows that oriented radiality now implies $r_{-2}r_2^2 = r_2 r_{-2}^2$, that is again $r_2 = r_{-2}$, the usual radiality.

Returning to the general environment of non-homogeneous trees, $q_+ \neq q_-$, we now know that we may assume $r_{-2} = r_2$. We should impose conditions (2) to the powers of M. Combinatorics are exceedingly complicated and arithmetic simplifications cannot be carried through by hand, but we checked by symbolic calculus that the radiality conditions for M^2, M^3 and M^4 on edges at distance 1 from the origin are not simultaneously satisfied unless $r_{-1} = 0 = r_1$, the case that we have already dealt with: hence the only radial algebras generated by M with support on the disc of radius 2 arise when the tree is homogeneous and M is radial in the usual sense, as claimed. We verified this by symbolic calculus, using the symbolic manipulation package *Mathematica*. We performed the computations for all $q_+ \neq q_-$ up to 100. Here is the Mathematica code that we wrote for this computation and can be used for larger values of the homogeneity degrees:

```
  (* Cardinalities of half-circles: *)
q[n_]:=If[n ≥ 0,
  If[EvenQ[n], q+^(n/2) q-^(n/2), q+^((n+1)/2) q-^((n-1)/2)],
```

```
  If[EvenQ[n], q+^(-n/2) q-^(-n/2), q+^(-(n+1)/2) q-^(-(n-1)/2)]]

  (* Multiplication rules for m[i]: *)
p[i_, j_]:=
  (* we cannot write m[i]m[j] as the left-hand side, *)
  (* because the factors of product *)
  (* would be automatically reordered *)
  (* even if the product is non-commutative *)
  If[i==0, m[i + j], If[EvenQ[i],
    If[i > 0, If[j ≥ 0, m[i + j],
        If[j ≥ −i, q[j]m[i + j] + Σ_{k=1}^{−j} (q[j + k − 1] − q[j + k])m[i + j + 2k − 1],
          q[−i]m[i + j] + Σ_{k=1}^{i} (q[−i + k − 1] − q[−i + k])m[−i − j + 2k − 1]]],
      If[j ≤ 0, m[i + j],
        If[j ≤ −i, q[j]m[i + j] + Σ_{k=1}^{j} (q[j − k + 1] − q[j − k])m[i + j − 2k + 1],
          q[−i]m[i + j] + Σ_{k=1}^{−i} (q[−i − k + 1] − q[−i − k])m[−i − j − 2k + 1]]]],
    If[i > 0, If[j ≤ 0, m[i − j],
        If[j ≤ i, q[j]m[i − j] + Σ_{k=1}^{j} (q[j − k + 1] − q[j − k])m[i − j + 2k − 1],
          q[i]m[i − j] + Σ_{k=1}^{i} (q[i − k + 1] − q[i − k])m[−i + j + 2k − 1]]],
      If[j ≥ 0, m[i − j],
        If[j ≥ i, q[j]m[i − j] + Σ_{k=1}^{−j} (q[j + k − 1] − q[j + k])m[i − j − 2k + 1],
          q[i]m[i − j] + Σ_{k=1}^{−i} (q[i + k − 1] − q[i + k])m[−i + j − 2k + 1]]]]]]

  (* Function M defined on edges, *)
  (* constant on half-circles: *)
M[n_]:=Sum[r_i mm[i], {i, −n, −1}] + Sum[r_i mm[i], {i, 1, n}]

  (* Convolution powers of function M: *)
powerofM[n_, 0]:=m[0]
powerofM[n_, k_]:=powerofM[n, k − 1]M[n]/.m[i_]mm[j_] → p[i, j]
  (* n is radius of M; k is exponent *)

  (* Component along m[u] *)
  (* of convolution power k of M: *)
coefficientofpowerofM[n_, k_, u_]:=
  Coefficient[powerofM[n, k], m[u]]

  (* Radiality conditions on edges of length u *)
  (* for powers k, h of a function *)
  (* with support on disc of radius n: *)
```

```
equation[n_, k_, h_, u_]:=
  coefficientofpowerofM[n, k, -u]coefficientofpowerofM[n, h, u] -
    coefficientofpowerofM[n, k, u]coefficientofpowerofM[n, h, -u]

  (* For all q- < q+ ≤ 100 *)
  (* this double loop checks radiality conditions *)
  (* for M, its square, and its cube *)
  (* on edges of length 1 *)
qmax:=100
For[q- = 1, q- < qmax, q-++, For[q+ = q-, q+ < qmax + 1, q+++,
  Print[{q-, q+, r-1, r1}/.NSolve[{
    equation[2, 2, 1, 1]==0/.r-2 → 1/.r2 → 1,
    equation[2, 3, 1, 1]==0/.r-2 → 1/.r2 → 1,
    equation[2, 4, 1, 1]==0/.r-2 → 1/.r2 → 1}, {r-1, r1}]]]]
```

Remark 4.4 It follows easily from the multiplication table (9) that the convolution of functions supported on edges of even (respectively, odd) length is not supported on the same sets, that is, it does not respect the parity of the length.

Moreover, the two Laplacians m_1 and m_{-1} of radius 1 never commute, not even on homogeneous trees. So, taken together, they fail to generate a radial algebra and also an abelian algebra. Of course neither of them individually is radial, hence neither generates a radial algebra, not even on a homogeneous tree. We have seen in the proof of Theorem 4.2 that none of their linear combinations generates a radial algebra unless the tree is homogeneous and the linear combination is a multiple of $m_1 + m_{-1}$. A similar statement holds for generators supported on edges of length at most 2, by the argument given above in the proof of Conjecture 4.3 for that case.

References

1. Cartier, P.: Harmonic analysis on trees. In: Harmonic Analysis on Homogeneous Spaces, Williamstown, 1972. Proc. Sympos. Pure Math., vol. 26, pp. 419–424. Am. Math. Soc., Providence (1973)
2. Casadio Tarabusi, E., Cohen, J.M., Korányi, A., Picardello, M.A.: Converse mean value theorems on trees and symmetric spaces. J. Lie Theory **8**, 229–254 (1998)
3. Figà-Talamanca, A., Picardello, M.A.: Harmonic Analysis on Free Groups. Lect. Notes in Pure and Appl. Math., vol. 87. Dekker, New York (1983)
4. Iozzi, A., Picardello, M.A.: Spherical functions on symmetric graphs. In: Harmonic Analysis, Cortona, 1982. Lect. Notes in Mathematics, vol. 992, pp. 344–386. Springer, Berlin (1983)
5. Macdonald, I.G.: Spherical functions on a group of p-adic type. In: Publications of the Ramanujan Institute, vol. 2. Ramanujan Institute, Centre for Advanced Study in Mathematics, University of Madras, Madras (1971)
6. Parkinson, J.: Buildings and Hecke algebras. J. Algebra **297**, 1–49 (2006)
7. Serre, J.-P.: Arbres, Amalgames, SL_2. In: Astérisque, vol. 46. Société Mathématique de France, Paris (1977). English translation: Trees. Springer Monographs in Math., Springer, Berlin (2003)

Surjunctivity and Reversibility of Cellular Automata over Concrete Categories

Tullio Ceccherini-Silberstein and Michel Coornaert

Abstract Following ideas developed by Misha Gromov, we investigate surjunctivity and reversibility properties of cellular automata defined over certain concrete categories.

Keywords Cellular automaton · Concrete category · Closed image property · Surjunctive cellular automaton · Reversible cellular automaton

Mathematics Subject Classification (2010) Primary 37B15 · Secondary 68Q80 · 18B05

1 Introduction

A cellular automaton is an abstract machine which takes as input a configuration of a universe and produces as output another configuration. The universe consists of cells and a configuration is described by the state of each cell of the universe. There is an input and an output state set and these two sets may be distinct. The state sets are also called the sets of colors, the sets of symbols, or the alphabets. The transition rule of a cellular automaton must obey two important properties, namely locality and time-independence. Locality means that the output state of a given cell only depends on the input states of a finite number of its neighboring cells possibly including the cell itself.

In the present paper, we restrict ourselves to cellular automata for which the universe is a group. The cells are the elements of the group. If the input alphabet is denoted by A, the output alphabet by B, and the group by G, this means that a

T. Ceccherini-Silberstein (✉)
Dipartimento di Ingegneria, Università del Sannio, C.so Garibaldi 107, 82100 Benevento, Italy
e-mail: tceccher@mat.uniroma3.it

M. Coornaert
Institut de Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René-Descartes, 67000 Strasbourg, France
e-mail: coornaert@math.unistra.fr

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3, DOI 10.1007/978-88-470-2853-1_6, © Springer-Verlag Italia 2013

cellular automaton is given by a map $\tau\colon A^G \to B^G$, where $A^G := \{x\colon G \to A\}$ is the set of all input configurations and $B^G := \{y\colon G \to B\}$ is the set of all possible output configurations. Besides locality, we will also always require a symmetry condition for τ, namely that it commutes with the shift actions of G on A^G and B^G (see Sect. 2 for precise definitions). In the case when $G = \mathbb{Z}^2$ and $A = B$ is a finite set, such cellular automata were first considered by John von Neumann in the late 1940s to serve as theoretical models for self-reproducing robots. Subsequently, cellular automata over $\mathbb{Z}$, $\mathbb{Z}^2$ or $\mathbb{Z}^3$ were widely used to model complex systems arising in natural or physical sciences. On the other hand, the mathematical study of cellular automata developed as a flourishing branch of theoretical computer science with numerous connections to abstract algebra, topological dynamical systems, ergodic theory, statistics and probability theory.

One of the most famous results in the theory of cellular automata is the Garden of Eden theorem established by Moore [33] and Myhill [34] in the early 1960s. It asserts that a cellular automaton $\tau\colon A^G \to A^G$, with $G = \mathbb{Z}^d$ and A finite, is surjective if and only if it is pre-injective (here *pre-injective* means that two configurations having the same image by τ must be equal if they coincide outside a finite number of cells). The name of this theorem comes from the fact that a configuration that is not in the image of a cellular automaton τ is sometimes called a *Garden of Eden* for τ because in a dynamical picture of the universe, obtained by iterating τ, such a configuration can only appear at time 0. Thus, the surjectivity of τ is equivalent to the absence of Garden of Eden configurations. At the end of the last century, the Garden of Eden theorem was first extended to finitely generated groups of subexponential growth in [30] and then to all amenable groups in [18]. It is now known [5] that the class of amenable groups is precisely the largest class of groups for which the Garden of Eden theorem is valid.

Observe that an immediate consequence of the Garden of Eden theorem is that every injective cellular automaton $\tau\colon A^G \to A^G$, with G amenable and A finite, is surjective and hence bijective. The fact that injectivity implies surjectivity, a property known as *surjunctivity* [21], was extended by Gromov [22] and Weiss [38] to all cellular automata $\tau\colon A^G \to A^G$ with finite alphabet over sofic groups. The class of sofic groups is a class of groups introduced by Gromov [22] containing in particular all amenable groups and all residually finite groups. Actually, there is no known example of a group that is not sofic.

Let us note that in the classical literature on cellular automata, the alphabet sets are most often assumed to be finite. With these hypotheses, topological methods based on properties of compact spaces may be used since A^G is compact for the prodiscrete topology when A is finite (Sect. 2.1). For example, one easily deduces from compactness that every bijective cellular automaton $\tau\colon A^G \to B^G$ is reversible when A is finite (a cellular automaton is called *reversible* if it is bijective and its inverse map is also a cellular automaton, see Sect. 2.6). On the other hand, in the infinite alphabet case, there exist bijective cellular automata that are not reversible.

The aim of the present paper is to investigate cellular automata $\tau\colon A^G \to B^G$ when the alphabets A and B are (possibly infinite) objects in some concrete category $\mathscr{C}$ and the local defining rules of τ are $\mathscr{C}$-morphisms (see Sect. 3 for precise definitions). For example, $\mathscr{C}$ can be the category of (let us say left) modules over some ring R, the category of topological spaces, or some of their subcategories. When $\mathscr{C}$ is the category of vector spaces over a field K, or, more generally, the category of left modules over a ring R, we recover the notion of a *linear cellular automaton* studied in [10–13]. When $\mathscr{C}$ is the category of affine algebraic sets over a field K, this gives the notion of an *algebraic cellular automaton* as in [17].

Following ideas developed by Gromov [22], we shall give sufficient conditions for a concrete category $\mathscr{C}$ that guarantee surjunctivity of all $\mathscr{C}$-cellular automata $\tau\colon A^G \to A^G$ when the group G is residually finite (Sect. 7). We shall also describe conditions on $\mathscr{C}$ implying that all bijective $\mathscr{C}$-automata are reversible (Sect. 8).

The present paper is mostly expository and collects results established in Gromov's seminal article [22] and in a series of papers written by the authors [10–14, 16, 17]. However, our survey contains some new proofs as well as some new results. On the other hand, we hope our concrete categorical approach will help the reader to a better understanding of this fascinating subject connected to so many contemporary branches of mathematics and theoretical computer science.

2 Cellular Automata

In this section, we have gathered background material on cellular automata over groups. The reader is referred to our monograph [15] for a more detailed exposition.

2.1 The Space of Configurations and the Shift Action

Let G be a group and let A be a set (called the *alphabet* or the set of *colors*). The set

$$A^G = \{x\colon G \to A\}$$

is endowed with its *prodiscrete topology*, i.e., the product topology obtained by taking the discrete topology on each factor A of $A^G = \prod_{g \in G} A$. Thus, if $x \in A^G$, a base of open neighborhoods of x is provided by the sets

$$V(x, \Omega) := \left\{ y \in A^G : x|_\Omega = y|_\Omega \right\},$$

where Ω runs over all finite subsets of G and $x|_\Omega \in A^\Omega$ denotes the restriction of $x \in A^G$ to Ω.

The space A^G, which is called the space of *configurations*, is Hausdorff and totally disconnected. It is compact if and only if the alphabet A is finite.

Example 2.1 If G is countably infinite and A is finite of cardinality $|A| \geqslant 2$, then A^G is homeomorphic to the Cantor set. This is the case for $G = \mathbb{Z}$ and $A = \{0, 1\}$, where A^G is the space of bi-infinite sequences of 0's and 1's.

There is a natural continuous left action of G on A^G given by

$$G \times A^G \to A^G,$$
$$(g, x) \mapsto gx$$

where

$$gx(h) = x\big(g^{-1}h\big) \quad \text{for every } h \in G.$$

This action is called the *G-shift* on A^G.

The study of the shift action on A^G is the central theme in *symbolic dynamics*.

2.2 Cellular Automata

Definition 2.2 Let G be a group. Given two alphabets A and B, a map $\tau\colon A^G \to B^G$ is called a *cellular automaton* if there exist a finite subset $M \subset G$ and a map $\mu_M\colon A^M \to B$ such that

$$\big(\tau(x)\big)(g) = \mu_M\big(\big(g^{-1}x\big)\big|_M\big) \quad \text{for every } x \in A^G, g \in G, \tag{1}$$

where $(g^{-1}x)|_M$ denotes the restriction of the configuration $g^{-1}x$ to M. Such a set M is called a *memory set* and the map $\mu_M\colon A^M \to B$ is called the associated *local defining map*.

Example 2.3 The identity map $\mathrm{Id}_{A^G}\colon A^G \to A^G$ is a cellular automaton with memory set $M = \{1_G\}$ and local defining map the identity map $\mu_M = \mathrm{Id}_A\colon A^M = A \to A$.

Example 2.4 If we fix an element $b_0 \in B$, then the constant map $\tau\colon A^G \to B^G$, defined by $\tau(x)(g) = b_0$ for all $x \in A^G$ and $g \in G$, is a cellular automaton with memory set $M = \emptyset$.

Example 2.5 If we fix an element $g_0 \in G$, then the map $\tau\colon A^G \to A^G$, defined by $\tau(x)(g) = x(gg_0)$ for all $x \in A^G$ and $g \in G$, is a cellular automaton with memory set $M = \{g_0\}$.

Example 2.6 (The majority action on $\mathbb{Z}$) Let $G = \mathbb{Z}$, $A = \{0, 1\}$, $M = \{-1, 0, 1\}$, and $\mu_M : A^M \equiv A^3 \to A$ defined by $\mu_M(a_{-1}, a_0, a_1) = 1$ if $a_{-1} + a_0 + a_1 \geqslant 2$ and $\mu_M(a_{-1}, a_0, a_1) = 0$ otherwise. The corresponding cellular automaton $\tau\colon A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ is described in Fig. 1.

Example 2.7 (Hedlund's marker [24]) Let $G = \mathbb{Z}$, $A = \{0, 1\}$, $M = \{-1, 0, 1, 2\}$, and $\mu_M\colon A^M \equiv A^4 \to A$ defined by $\mu_M(a_{-1}, a_0, a_1, a_2) = 1 - a_0$ if $(a_{-1}, a_1, a_2) = (0, 1, 0)$ and $\mu_M(a_{-1}, a_0, a_1, a_2) = a_0$ otherwise. The corresponding cellular automaton $\tau\colon A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ is a non-trivial involution of A^G (see Fig. 2).

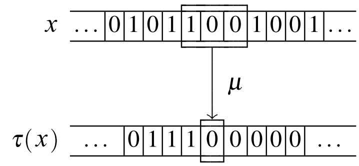


Fig. 1 The cellular automaton defined by the majority action on $\mathbb{Z}$

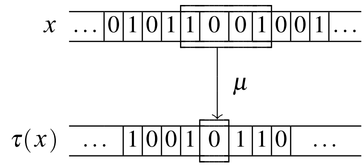


Fig. 2 The cellular automaton defined by the Hedlund marker

Example 2.8 (Conway's Game of Life) Life was introduced by J.H. Conway in the 1970's and popularized by M. Gardner. From a theoretical computer science point of view, it is important because it has the power of a universal Turing machine, that is, anything that can be computed algorithmically can be computed by using the Game of Life.

Let $G = \mathbb{Z}^2$ and $A = \{0, 1\}$. *Life* is the cellular automaton

$$\tau \colon \{0, 1\}^{\mathbb{Z}^2} \to \{0, 1\}^{\mathbb{Z}^2}$$

with memory set $M = \{-1, 0, 1\}^2 \subset \mathbb{Z}^2$ and local defining map $\mu \colon A^M \to A$ given by

$$\mu(y) = \begin{cases} 1 & \text{if } \begin{cases} \sum_{m \in M} y(m) = 3 \\ \text{or } \sum_{m \in M} y(m) = 4 \text{ and } y((0, 0)) = 1 \end{cases} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

for all $y \in A^M$. One thinks of an element g of $G = \mathbb{Z}^2$ as a "cell" and the set gM (we use multiplicative notation) as the set consisting of its eight neighboring cells, namely the North, North-East, East, South-East, South, South-West, West and North-West cells. We interpret state 0 as corresponding to the *absence* of life while state 1 corresponds to the *presence* of life. We thus refer to cells in state 0 as *dead* cells and to cells in state 1 as *live* cells. Finally, if $x \in A^{\mathbb{Z}^2}$ is a configuration at time t, then $\tau(x)$ represents the evolution of the configuration at time $t + 1$. Then the cellular automaton in (2) evolves as follows.

- *Birth:* a cell that is dead at time t becomes alive at time $t + 1$ if and only if three of its neighbors are alive at time t.
- *Survival:* a cell that is alive at time t will remain alive at time $t + 1$ if and only if it has exactly two or three live neighbors at time t.
- *Death by loneliness:* a live cell that has at most one live neighbor at time t will be dead at time $t + 1$.

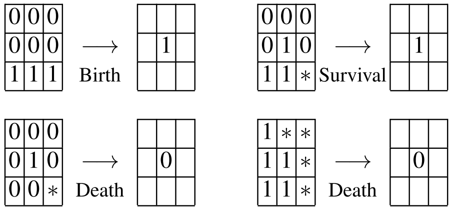


Fig. 3 The evolution of a cell in the Game of Life. The symbol $*$ represents any symbol in $\{0, 1\}$

- *Death by overcrowding:* a cell that is alive at time t and has four or more live neighbors at time t, will be dead at time $t + 1$.

Figure 3 illustrates all these cases.

Observe that if $\tau \colon A^G \to B^G$ is a cellular automaton with memory set M and local defining map μ_M, then μ_M is entirely determined by τ and M since, for all $y \in A^M$, we have

$$\mu_M(y) = \tau(x)(1_G), \tag{3}$$

where $x \in A^G$ is any configuration satisfying $x|_M = y$. Moreover, every finite subset $M' \subset G$ containing M is also a memory set for τ (with associated local defining map $\mu_{M'} \colon A^{M'} \to B$ given by $\mu_{M'}(y) = \mu_M(y|_M)$ for all $y \in A^{M'}$). In fact (see for example [15, Sect. 1.5]), the following holds. Every cellular automaton $\tau \colon A^G \to B^G$ admits a unique memory set $M_0 \subset G$ of minimal cardinality. Moreover, a subset $M \subset G$ is a memory set for τ if and only if $M_0 \subset M$. This memory set M_0 is called the *minimal* memory set of τ.

From the definition, it easily follows that every cellular automaton $\tau \colon A^G \to B^G$ is G-equivariant, i.e.,

$$\tau(gx) = g\tau(x) \quad \text{for every } x \in A^G, g \in G$$

[15, Proposition 1.4.6], and continuous with respect to the prodiscrete topologies on A^G and B^G [15, Proposition 1.4.8].

2.3 Composition of Cellular Automata

The following result is well known. The proof we present here follows the one given in [15, Remark 1.4.10]. An alternative proof will be given in Remark 2.15 below and the result will be strengthened later in Proposition 3.14.

Proposition 2.9 *Let G be a group and let A, B and C be sets. Suppose that $\tau \colon A^G \to B^G$ and $\sigma \colon B^G \to C^G$ are cellular automata. Then the composite map $\sigma \circ \tau \colon A^G \to C^G$ is a cellular automaton.*

Proof Let S (resp. T) be a memory set for σ (resp. τ) and let $\mu\colon B^S \to C$ (resp. $\nu\colon A^T \to B$) be the corresponding local defining map. Then the set $ST = \{st : s \in S, t \in T\} \subset G$ is finite. We have $sT \subset ST$ for every $s \in S$. Consider, for each $s \in S$, the projection map $\pi_s\colon A^{ST} \to A^{sT}$, the bijective map $f_s\colon A^{sT} \to A^T$ defined by $f_s(y)(t) = y(st)$ for all $y \in A^{sT}$ and $t \in T$, and the map $\varphi_s\colon A^{ST} \to B$ given by

$$\varphi_s := \nu \circ f_s \circ \pi_s. \tag{4}$$

Finally, let

$$\Phi := \prod_{s \in S} \varphi_s\colon A^{ST} \to B^S \tag{5}$$

be the product map defined by $\Phi(z)(s) = \varphi_s(z)$ for all $z \in A^{ST}$.

Let us show that $\sigma \circ \tau$ is a cellular automaton with memory set ST and associated local defining map

$$\kappa := \mu \circ \Phi\colon A^{ST} \to C. \tag{6}$$

Let $x \in A^G$. We first observe that, for all $s \in S$ and $t \in T$, we have

$$\begin{aligned}
\left(s^{-1}x\right)(t) &= x(st) \\
&= x|_{ST}(st) \\
&= \left((f_s \circ \pi_s)(x|_{ST})\right)(t),
\end{aligned}$$

so that

$$\left(s^{-1}x\right)\big|_T = (f_s \circ \pi_s)(x|_{ST}).$$

It follows that

$$\tau(x)(s) = \nu\left(\left(s^{-1}x\right)\big|_T\right) = \nu\left(f_s\left(\pi_s(x|_{ST})\right)\right) = \varphi_s(x|_{ST}).$$

Thus, we have

$$\left(\tau(x)\right)\big|_S = \Phi(x|_{ST}). \tag{7}$$

We finally get

$$\begin{aligned}
\left((\sigma \circ \tau)(x)\right)(g) &= \sigma\left(\tau(x)\right)(g) \\
&= \mu\left(\left(g^{-1}\tau(x)\right)\big|_S\right) \\
&= \mu\left(\tau\left(g^{-1}x\right)\big|_S\right) && \text{(by } G\text{-equivariance of } \tau) \\
&= \mu\left(\Phi\left(\left(g^{-1}x\right)\big|_{ST}\right)\right) && \text{(by (7))} \\
&= \kappa\left(\left(g^{-1}x\right)\big|_{ST}\right) && \text{(by (6))}.
\end{aligned}$$

This shows that $\sigma \circ \tau$ is a cellular automaton with memory set ST and associated local defining map κ. □

If we fix a group G, we deduce from Example 2.3 and Proposition 2.9 that the cellular automata $\tau\colon A^G \to B^G$ are the morphisms of a subcategory of the category of sets whose objects are all the sets of the form A^G. We shall denote this subcategory by $\mathrm{CA}(G)$.

2.4 The Curtis-Hedlund Theorem

When the alphabet A is finite, the *Curtis-Hedlund theorem* holds [24] (see also [15, Theorem 1.8.1]):

Theorem 2.10 *Let G be a group, A a finite set and B a set. Let $\tau\colon A^G \to B^G$ be a map. Then the following conditions are equivalent*:

(a) *τ is a cellular automaton*;
(b) *τ is G-equivariant and continuous* (*with respect to the prodiscrete topologies on A^G and B^G*).

As already mentioned in Sect. 2.2, the implication (a) $\Rightarrow$ (b) remains true for A infinite. When the group G is non-periodic (i.e., there exists $g \in G$ of infinite order) and A is infinite, one can always construct a G-equivariant continuous self-mapping of A^G which is not a cellular automaton (see [16] and [15, Example 1.8.2]).

Example 2.11 For $G = A = \mathbb{Z}$, the map $\tau\colon A^G \to A^G$, defined by

$$\tau(x)(n) = x\bigl(x(n) + n\bigr),$$

is G-equivariant and continuous, but τ is not a cellular automaton.

2.5 Uniform Spaces and the Generalized Curtis-Hedlund Theorem

Let X be a set.

We denote by Δ_X the diagonal in $X \times X$, that is the set $\Delta_X = \{(x,x) : x \in X\}$.

The *inverse* $\overset{-1}{U}$ of a subset $U \subset X \times X$ is the subset of $X \times X$ defined by $\overset{-1}{U} = \{(x,y) : (y,x) \in U\}$. We define the *composite* $U \circ V$ of two subsets U and V of $X \times X$ by

$$U \circ V = \bigl\{(x,y) : \text{there exists } z \in X \text{ such that } (x,z) \in U \text{ and } (z,y) \in V\bigr\} \subset X \times X.$$

Definition 2.12 Let X be a set. A *uniform structure* on X is a non-empty set $\mathscr{U}$ of subsets of $X \times X$ satisfying the following conditions:

(UN-1) if $U \in \mathscr{U}$, then $\Delta_X \subset U$;
(UN-2) if $U \in \mathscr{U}$ and $U \subset V \subset X \times X$, then $V \in \mathscr{U}$;
(UN-3) if $U \in \mathscr{U}$ and $V \in \mathscr{U}$, then $U \cap V \in \mathscr{U}$;

(UN-4) if $U \in \mathscr{U}$, then $\overset{-1}{U} \in \mathscr{U}$;
(UN-5) if $U \in \mathscr{U}$, then there exists $V \in \mathscr{U}$ such that $V \circ V \subset U$.

The elements of $\mathscr{U}$ are then called the *entourages* of the uniform structure and the set X is called a *uniform space*.

A subset $\mathscr{B} \subset \mathscr{U}$ is called a *base* of $\mathscr{U}$ if for each $W \in \mathscr{U}$ there exists $V \in \mathscr{B}$ such that $V \subset W$.

Let X and Y be uniform spaces. A map $f: X \to Y$ is called *uniformly continuous* if it satisfies the following condition: for each entourage W of Y, there exists an entourage V of X such that $(f \times f)(V) \subset W$. Here $f \times f$ denotes the map from $X \times X$ into $Y \times Y$ defined by $(f \times f)(x_1, x_2) = (f(x_1), f(x_2))$ for all $(x_1, x_2) \in X \times X$.

If X is a uniform space, there is an induced topology on X characterized by the fact that the neighborhoods of an arbitrary point $x \in X$ consist of the sets $U[x] = \{y \in X : (x, y) \in U\}$, where U runs over all entourages of X. This topology is Hausdorff if and only if the intersection of all the entourages of X is reduced to the diagonal Δ_X. Moreover, every uniformly continuous map $f: X \to Y$ is continuous with respect to the induced topologies on X and Y.

Example 2.13 If (X, d_X) is a metric space, there is a natural uniform structure on X whose entourages are the sets $U \subset X \times X$ satisfying the following condition: there exists a real number $\varepsilon > 0$ such that U contains all pairs $(x, y) \in X \times X$ such that $d_X(x, y) < \varepsilon$. The topology associated with this uniform structure is then the same as the topology induced by the metric. If (Y, d_Y) is another metric space, then a map $f: X \to Y$ is uniformly continuous if and only if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $d_Y(f(x_1), f(x_2)) < \varepsilon$ whenever $d_X(x_1, x_2) < \delta$.

The theory of uniform spaces was developed by A. Weil in [37] (see e.g. [8, 26], [15, Appendix B]).

Let us now return back to configuration spaces and cellular automata. Let G be a group and A be a set. We equip A^G with its *prodiscrete uniform structure*, that is with the uniform structure admitting the sets

$$W(\Omega) := \left\{(x, y) \in A^G \times A^G : x|_\Omega = y|_\Omega\right\},$$

where $\Omega \subset G$ runs over all finites subsets of G, as a base of entourages.

We then have ([16]; see also [15, Theorem 1.9.1]) the following extension of the Curtis-Hedlund theorem:

Theorem 2.14 *Let G be a group and let A and B be sets. Let $\tau : A^G \to B^G$ be a map. Then the following conditions are equivalent*:

(a) *τ is a cellular automaton*;
(b) *τ is G-equivariant and uniformly continuous* (*with respect to the prodiscrete uniform structures on A^G and B^G*).

Remark 2.15 Since the composite of two G-equivariant (resp. uniformly continuous) maps is G-equivariant (resp. uniformly continuous), we immediately deduce from Theorem 2.14 an alternative proof of the fact that the composite of two cellular automata is a cellular automaton (Proposition 2.9).

2.6 Reversible Cellular Automata

Given a group G and two sets A and B, a cellular automaton $\tau \colon A^G \to B^G$ is called *reversible* if τ is bijective and its inverse map $\tau^{-1} \colon B^G \to A^G$ is also a cellular automaton. Observe that the reversible cellular automata $\tau \colon A^G \to B^G$ are precisely the isomorphisms of the category $\mathrm{CA}(G)$ introduced at the end of Sect. 2.3. It immediately follows from Theorem 2.14 that a bijective cellular automaton $\tau \colon A^G \to B^G$ is reversible if and only if the inverse map $\tau^{-1} \colon B^G \to A^G$ is uniformly continuous with respect to the prodiscrete uniform structures on A^G and B^G.

When the alphabet A is finite, every bijective cellular automaton $\tau \colon A^G \to B^G$ is reversible by compactness of A^G. On the other hand, when A is infinite and the group G is non-periodic, there exist bijective cellular automata $\tau \colon A^G \to A^G$ that are not reversible (see [16], [15, Example 1.10.3], and the examples given at the end of the present paper).

2.7 Induction and Restriction of Cellular Automata

Let G be a group, A and B two sets, and H a subgroup of G.

Suppose that a cellular automaton $\tau \colon A^G \to B^G$ admits a memory set M such that $M \subset H$. Let $\mu_M^G \colon A^M \to B$ denote the associated local defining map. Then the map $\tau_H \colon A^H \to B^H$ defined by

$$\tau_H(y)(h) = \mu_M^G\big((h^{-1}y)\big|_M\big) \quad \text{for all } y \in A^H \text{ and } h \in H,$$

is a cellular automaton over the group H with memory set M and local defining map $\mu_M^H := \mu_M^G$. One says that τ_H is the cellular automaton obtained by *restriction* of τ to H.

Conversely, suppose that $\sigma \colon A^H \to B^H$ is a cellular automaton with memory set $N \subset H$ and local defining map $\nu_N^H \colon A^N \to B$. Then the map $\sigma^G \colon A^G \to B^G$ defined by

$$\sigma^G(x)(g) = \nu_N^H\big((g^{-1}x)\big|_N\big) \quad \text{for all } x \in A^G \text{ and } g \in G,$$

is a cellular automaton over the group G with memory set N and local defining map $\nu_N^G := \nu_N^H$. One says that σ^G is the cellular automaton obtained by *induction* of σ to G.

It immediately follows from their definitions that induction and restriction are operations one inverse to the other in the sense that $(\tau_H)^G = \tau$ and $(\sigma^G)_H = \sigma$ for every cellular automaton $\tau\colon A^G \to B^G$ over G admitting a memory set contained in H, and every cellular automaton $\sigma\colon A^H \to B^H$ over H. We shall use the following result, established in [14, Theorem 1.2] (see also [15, Proposition 1.7.4]).

Theorem 2.16 *Let G be a group, A and B two sets, and H a subgroup of G. Suppose that $\tau\colon A^G \to B^G$ is a cellular automaton over G admitting a memory set contained in H and let $\tau_H\colon A^H \to B^H$ denote the cellular automaton over H obtained by restriction. Then the following holds:*

(i) *τ is injective if and only if τ_H is injective;*
(ii) *τ is surjective if and only if τ_H is surjective;*
(iii) *τ is bijective if and only if τ_H is bijective;*
(iv) *$\tau(A^G)$ is closed in B^G for the prodiscrete topology if and only if $\tau_H(A^H)$ is closed in B^H for the prodiscrete topology.*

3 Cellular Automata over Concrete Categories

We assume some familiarity with the basic concepts of category theory (the reader is referred to [31] and [1] for a detailed introduction to category theory). We adopt the following notation:

- Set is the category where objects are sets and morphisms are maps between them;
- Set^f is the full subcategory of Set whose objects are finite sets;
- Grp is the category where objects are groups and morphisms are group homomorphisms between them;
- Rng is the category where objects are rings and morphisms are ring homomorphisms (all rings are assumed to be unital and ring homomorphisms are required to preserve the unity elements);
- Fld is the full category of Rng whose objects are fields (a field is a non-trivial commutative ring in which every nonzero element is invertible);
- R-Mod is the category where objects are left modules over a given ring R and morphisms are R-linear maps between them;
- R-Mod^{f-g} is the full subcategory of R-Mod whose objects are finitely-generated left modules over a given ring R;
- K-Vec $=$ K-Mod is the category where objects are vector spaces over a given field K and morphisms are K-linear maps between them;
- K-Vec^{f-d} $=$ K-Mod^{f-g} is the full subcategory of K-Vec whose objects are finite-dimensional vector spaces over a given field K;
- K-Aal is the category where objects are affine algebraic sets over a given field K and morphisms are regular maps between them. Recall that A is an *affine algebraic set* over a field K if $A \subset K^n$, for some integer $n \geqslant 0$, and A is the set of common zeroes of some set of polynomials $S \subset K[t_1, t_2, \ldots, t_n]$. Recall

also that a map $f\colon A \to B$ from an affine algebraic set $A \subset K^n$ to another affine algebraic set $B \subset K^m$ is called *regular* if f is the restriction of some polynomial map $P\colon K^n \to K^m$ (see for instance [9, 28, 32, 35] for an introduction to affine algebraic geometry);
- Top is the category where objects are topological spaces and morphisms are continuous maps between them;
- CHT is the subcategory of Top whose objects are compact Hausdorff topological spaces;
- Man is the subcategory of Top whose objects are compact topological manifolds (a *topological manifold* is a connected Hausdorff topological space in which every point admits an open neighborhood homeomorphic to $\mathbb{R}^n$ for some integer $n \geqslant 0$).

If $\mathscr{C}$ is a category and A is a $\mathscr{C}$-object, we shall denote by Id_A the identity morphism of A.

3.1 Concrete Categories

A *concrete category* is a pair $(\mathscr{C}, U)$, where $\mathscr{C}$ is a category and $U\colon \mathscr{C} \to \mathrm{Set}$ is a faithful functor from $\mathscr{C}$ to the category Set. The functor U is called the *forgetful functor* of the concrete category $(\mathscr{C}, U)$. We will denote a concrete category $(\mathscr{C}, U)$ simply by $\mathscr{C}$ if its forgetful functor U is clear from the context.

Let $(\mathscr{C}, U)$ be a concrete category. If A is a $\mathscr{C}$-object, the set $U(A)$ is called the *underlying set* of A. Similarly, if $f\colon A \to B$ is a $\mathscr{C}$-morphism, the map $U(f)\colon U(A) \to U(B)$ is called the *underlying map* of f. Note that two distinct $\mathscr{C}$-objects may have the same underlying set. However, the faithfulness of U implies that a $\mathscr{C}$-morphism is entirely determined by its underlying map once its source and target objects are given.

Every subcategory of a concrete category is itself a concrete category. More precisely, if $(\mathscr{C}, U)$ is a concrete category, $\mathscr{D}$ is a subcategory of $\mathscr{C}$, and $U|_{\mathscr{D}}$ denotes the restriction of U to $\mathscr{D}$, then $(\mathscr{D}, U|_{\mathscr{D}})$ is a concrete category.

The categories Set, Grp, Rng, R-Mod, K-Aal and Top, equipped with their obvious forgetful functor to Set, provide basic examples of concrete categories. On the other hand, it can be shown that the homotopy category hTop, where objects are topological spaces and morphisms are homotopy classes of continuous maps between them, is *not concretizable*, in the sense that there exists no faithful functor $U\colon \mathrm{hTop} \to \mathrm{Set}$ (see [20] and [1, Exercise 5J]).

Let $\mathscr{C}$ be a category. Recall that a *product* of a family $(A_i)_{i \in I}$ of $\mathscr{C}$-objects is a pair $(P, (\pi_i)_{i \in I})$, where P is a $\mathscr{C}$-object and $\pi_i\colon P \to A_i$, $i \in I$, are $\mathscr{C}$-morphisms satisfying the following universal property: if B is a $\mathscr{C}$-object equipped with $\mathscr{C}$-morphisms $f_i\colon B \to A_i$, $i \in I$, then there is a unique $\mathscr{C}$-morphism $g\colon B \to P$ such that $f_i = \pi_i \circ g$ for all $i \in I$. When it exists, such a product is essentially unique, in the sense that if $(P, (\pi_i)_{i \in I})$ and $(P', (\pi'_i)_{i \in I})$ are two products of the family

$(A_i)_{i \in I}$ then there exists a unique $\mathscr{C}$-isomorphism $\varphi \colon P \to P'$ such that $\pi_i = \pi'_i \circ \varphi$ for all $i \in I$. By a common abuse, one writes $P = \prod_{i \in I} A_i$ and $g = \prod_{i \in I} f_i$.

One says that a category $\mathscr{C}$ *has products* [1, Definition 10.29.(1)], or that $\mathscr{C}$ satisfies condition (P), if every set-indexed family $(A_i)_{i \in I}$ of $\mathscr{C}$-objects admits a product in $\mathscr{C}$.

One says that a category $\mathscr{C}$ *has finite products* [1, Definition 10.29.(2)], or that $\mathscr{C}$ satisfies (FP), if every finite family $(A_i)_{i \in I}$ of $\mathscr{C}$-objects admits a product in $\mathscr{C}$. By means of an easy induction, one shows that a category $\mathscr{C}$ has finite products if and only if it satisfies the two following conditions:

(a) $\mathscr{C}$ has a terminal object (such an object is then the product of the empty family of $\mathscr{C}$-objects);
(b) any pair of $\mathscr{C}$-objects admits a product [1, Proposition 10.30].

It is clear from these definitions that (P) implies (FP).

The categories Set, Grp, Rng, R-Mod, Top and CHT all satisfy (P). On the other hand, the category Set^f satisfies (FP) but not (P). The category $R\text{-Mod}^{f-g}$ satisfies (FP) but, unless R is a zero ring, does not satisfy (P). The category K-Aal of affine algebraic sets and regular maps over a given field K also satisfies (FP) but not (P). The category Fld does not satisfy (FP) (it does not even admit a terminal object).

Suppose now that $(\mathscr{C}, U)$ is a concrete category. Given a family $(A_i)_{i \in I}$ of $\mathscr{C}$-objects, one says that the pair $(P, (\pi_i)_{i \in I})$ is a *concrete product* of the family $(A_i)_{i \in I}$ if $(P, (\pi_i)_{i \in I})$ is a product of the family $(A_i)_{i \in I}$ in $\mathscr{C}$ and $(U(P), (U(\pi_i))_{i \in I})$ is a product of the family $(U(A_i))_{i \in I}$ in Set [1, Definition 10.52].

One says that $(\mathscr{C}, U)$ has *concrete products*, or that $(\mathscr{C}, U)$ satisfies (CP), if every set-indexed family of $\mathscr{C}$-objects admits a concrete product [1, Definition 10.54].

One says that $(\mathscr{C}, U)$ has *concrete finite products*, or that $(\mathscr{C}, U)$ satisfies (CFP), if every finite family of $\mathscr{C}$-objects admits a concrete product. By using an induction argument, one gets a characterization of (CFP) analogous to the one given above for (FP). More precisely, one can show that a concrete category $(\mathscr{C}, U)$ has concrete finite products if and only if it satisfies the two following conditions:

(a) $\mathscr{C}$ has a terminal object whose underlying set is reduced to a single element (such an object is then the concrete product of the empty family of $\mathscr{C}$-objects);
(b) any pair of $\mathscr{C}$-objects admits a concrete product.

It is clear from these definitions that (CP) implies (CFP).

The concrete categories Set, Grp, Rng, R-Mod, Top and CHT all satisfy (CP). The concrete categories Set^f, $R\text{-Mod}^{f-g}$ (for R a nonzero ring), K-Aal, and Man satisfy (CFP) but not (CP), since they do not satisfy (P). Here is an example of a concrete category that satisfies (P) but not (CFP).

Example 3.1 Fix a set X and consider the category $\mathscr{C}$ defined as follows. The objects of $\mathscr{C}$ are all the pairs (A, α), where A is a set and $\alpha \colon A \to X$ is a map. If (A, α) and (B, β) are $\mathscr{C}$-objects, the morphisms from (A, α) to (B, β) consist of all maps $f \colon A \to B$ such that $\alpha = \beta \circ f$. It is clear that $\mathscr{C}$ is a concrete category for the forgetful functor $U: \mathscr{C} \to \text{Set}$ that associates with each object (A, α) the set A and with each morphism $f: (A, \alpha) \to (B, \beta)$ the map $f: A \to B$. The category $\mathscr{C}$ satisfies (P). Indeed, the product of a set-indexed family $((A_i, \alpha_i))_{i \in I}$ of $\mathscr{C}$-objects is the *fibered product* (F, η), where

$$F := \left\{(a_i)_{i \in I} : \alpha_i(a_i) = \alpha_j(a_j) \text{ for all } i, j \in I\right\} \subset \prod_{i \in I} A_i$$

with the natural projections maps $\pi_i: F \to A_i$. The pair (X, Id_X), where $\text{Id}_X: X \to X$ is the identity map, is clearly a terminal $\mathscr{C}$-object. Since any terminal object in a concrete category satisfying (CFP) must be reduced to a single element, we deduce that the concrete category $(\mathscr{C}, U)$ does not satisfy (CFP) unless X is reduced to a single element (observe that $(\mathscr{C}, U)$ is identical to Set when X is reduced to a single element).

Recall that a morphism $f: A \to B$ in a category $\mathscr{C}$ is called a *retraction* if it is right-invertible, i.e., if there exists a $\mathscr{C}$-morphism $g: B \to A$ such that $f \circ g = \text{Id}_B$. We have the following elementary lemma.

Lemma 3.2 *Let $\mathscr{C}$ be a category. Let A and B be two $\mathscr{C}$-objects admitting a $\mathscr{C}$-product $A \times B$ with first projection $\pi: A \times B \to A$. Then the following conditions are equivalent*:

(a) *π is a retraction*;
(b) *there exists a $\mathscr{C}$-morphism $f: A \to B$.*

Proof Let $\pi': A \times B \to B$ denote the second projection. If $g: A \to A \times B$ is a $\mathscr{C}$-morphism such that $\pi \circ g = \text{Id}_A$, then $f := \pi' \circ g$ is a $\mathscr{C}$-morphism from A to B. This shows that (a) implies (b). Conversely, if (b) is satisfied, then $g := \text{Id}_A \times f: A \to A \times B$ satisfies $\pi \circ g = \text{Id}_A$. □

We say that a concrete category $(\mathscr{C}, U)$ satisfies (CFP+) provided it satisfies (CFP) and the following additional condition:

(C+) Given any $\mathscr{C}$-object A and any $\mathscr{C}$-object B with $U(B) \neq \emptyset$, the first projection morphism $\pi: A \times B \to A$ is a retraction.

Proposition 3.3 *Let $(\mathscr{C}, U)$ be a concrete category satisfying (CFP). Then the following conditions are equivalent*:

(a) *$(\mathscr{C}, U)$ satisfies (CFP+)*;
(b) *for any $\mathscr{C}$-object A and any $\mathscr{C}$-object B with $U(B) \neq \emptyset$, there exists a $\mathscr{C}$-morphism $f: A \to B$*;
(c) *if T is a terminal $\mathscr{C}$-object and B is any $\mathscr{C}$-object with $U(B) \neq \emptyset$, then there exists a $\mathscr{C}$-morphism $g: T \to B$.*

Proof The equivalence of (a) and (b) follows from Lemma 3.2. Condition (b) trivially implies condition (c). To prove that (c) implies (b), it suffices to observe that by composing a $\mathscr{C}$-morphism $h\colon A \to T$ and a $\mathscr{C}$-morphism $g\colon T \to B$, we get a $\mathscr{C}$-morphism $f := g \circ h\colon A \to B$. $\square$

Note that condition (c) in the preceding proposition is satisfied in particular when $\mathscr{C}$ admits a *zero object* 0 (i.e., an object that is both initial and terminal). Hence every concrete category satisfying (CFP) also satisfies (CFP+) if it admits a zero object. This is the case for the categories Grp, Rng, R-Mod, and R-Mod^{f-g}.

On the other hand, the categories Set, Set^f, K-Aal, Top, CHT, and Man also satisfy (CFP+) although they do not admit zero objects. Indeed, in any of these categories, the only initial object is the empty one while the terminal objects are the singletons, and if T is a singleton and B an arbitrary object, then any map from $U(T)$ to $U(B)$ is the underlying map of a morphism from T to B.

3.2 Cellular Automata over Concrete Categories

From now on, in a concrete category, we shall use the same symbol to denote an object (resp. a morphism) and its underlying set (resp. its underlying map).

Proposition 3.4 *Let G be a group and let $\mathscr{C}$ be a concrete category satisfying* (*CFP+*). *Let A and B be two $\mathscr{C}$-objects. Suppose that $\tau\colon A^G \to B^G$ is a cellular automaton. Then the following conditions are equivalent*:

(a) *there exists a memory set M of τ such that the associated local defining map $\mu_M\colon A^M \to B$ is a $\mathscr{C}$-morphism*;
(b) *for any memory set M of τ, the associated local defining map $\mu_M\colon A^M \to B$ is a $\mathscr{C}$-morphism.*

Note that, in the above statement, A^M is a $\mathscr{C}$-object since A is a $\mathscr{C}$-object, M is finite, and $\mathscr{C}$ satisfies (CFP+) and hence (CFP). On the other hand, it may happen that the configuration spaces A^G and B^G are not $\mathscr{C}$-objects (although this is the case if $\mathscr{C}$ satisfies (CP)).

Proof of Proposition 3.4 We can assume $A \neq \emptyset$. Suppose that the local defining map $\mu_M\colon A^M \to B$ is a $\mathscr{C}$-morphism for some memory set M. Let M' be another memory set and let us show that the associated local defining map $\mu_{M'}\colon A^{M'} \to B$ is also a $\mathscr{C}$-morphism. Let M_0 denote the minimal memory set of τ. Recall that we have $M_0 \subset M \cap M'$. After identifying the $\mathscr{C}$-object A^M (resp. $A^{M'}$) with the $\mathscr{C}$-product $A^{M_0} \times A^{M \setminus M_0}$ (resp. $A^{M_0} \times A^{M' \setminus M_0}$), consider the projection map $\pi\colon A^M \to A^{M_0}$ (resp. $\pi'\colon A^{M'} \to A^{M_0}$). We then have

$$\mu_M = \mu_{M_0} \circ \pi \quad \text{and} \quad \mu_{M'} = \mu_{M_0} \circ \pi'. \tag{8}$$

By condition (CFP+), the projection π is a retraction, so that there exists a $\mathscr{C}$-morphism $f\colon A^{M_0} \to A^M$ such that $\pi \circ f = \mathrm{Id}_{A^{M_0}}$. Using (8), we get

$$\mu_{M'} = \mu_{M_0} \circ \pi' = \mu_{M_0} \circ \mathrm{Id}_{A^{M_0}} \circ \pi' = \mu_{M_0} \circ \pi \circ f \circ \pi' = \mu_M \circ f \circ \pi'.$$

Thus, the map $\mu_{M'}$ may be written as the composite of three $\mathscr{C}$-morphisms and therefore is a $\mathscr{C}$-morphism. □

Definition 3.5 Let G be a group and let $\mathscr{C}$ be a concrete category satisfying (CFP+). Let A and B be two $\mathscr{C}$-objects. We say that a cellular automaton $\tau\colon A^G \to B^G$ is a *$\mathscr{C}$-cellular automaton* provided it satisfies one of the equivalent conditions of Proposition 3.4.

Example 3.6 Let G be a group and let $\mathscr{C}$ be a concrete category satisfying (CFP+). Let A be a $\mathscr{C}$-object. Then the identity map $\mathrm{Id}_{A^G}\colon A^G \to A^G$ is the cellular automaton with memory set $M = \{1_G\}$ and local defining map $\mu_M = \mathrm{Id}_A\colon A^M = A \to A$ (see Example 2.3). As Id_A is a $\mathscr{C}$-morphism, we deduce that Id_{A^G} is a $\mathscr{C}$-cellular automaton.

Example 3.7 (The Discrete Laplacian) Let G be a group, $A = \mathbb{R}$, and $S \subset G$ a nonempty finite subset. The map $\Delta_S\colon \mathbb{R}^G \to \mathbb{R}^G$, defined by

$$(\Delta_S x)(g) = x(g) - \frac{1}{|S|}\sum_{s\in S} x(gs)$$

for all $x \in \mathbb{R}^G$ and $g \in G$, is a cellular automaton (with memory set $M = S \cup \{1_G\}$ and associated local defining map $\mu_M\colon \mathbb{R}^M \to \mathbb{R}$ given by $\mu_M(y) = y(1_G) - \frac{1}{|S|}\sum_{s\in S} y(s)$ for all $y \in \mathbb{R}^M$). Since $\mathbb{R}$ is a finite dimensional vector space over itself and μ_M is $\mathbb{R}$-linear, we have that Δ_S is a $\mathscr{C}$-cellular automaton for $\mathscr{C} = \mathbb{R}$-Vec^{f-d}.

Example 3.8 Let $G = \mathbb{Z}$ and let $A = K$ be any field. Then the map $\tau\colon K^{\mathbb{Z}} \to K^{\mathbb{Z}}$, defined by

$$\tau(x)(n) = x(n+1) - x(n)^2$$

for all $x \in K^{\mathbb{Z}}$ and $n \in \mathbb{Z}$, is an $\mathscr{C}$-cellular automaton for $\mathscr{C} = K$-Aal, with memory set $M = \{0, 1\}$ and associated local defining map $\mu_M\colon A^M \to A$ given by $\mu(y) = y(1) - y(0)^2$ for all $y \in A^M$. Observe that τ is not a $\mathscr{C}$-cellular automaton for $\mathscr{C} = K$-Vec unless $K \cong \mathbb{Z}/2\mathbb{Z}$ is the field with two elements.

Example 3.9 Let $G = \mathbb{Z}$. Let also $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$ denote the *unit circle* and, for $n \geqslant 1$, denote by $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n = (\mathbb{S}^1)^n$ the *n-torus*. With each continuous map $f\colon \mathbb{T}^{m+1} \to \mathbb{S}^1$, $m \geqslant 0$, we can associate the cellular automaton $\tau\colon (\mathbb{S}^1)^{\mathbb{Z}} \to (\mathbb{S}^1)^{\mathbb{Z}}$ with memory set $M = \{0, 1, \dots, m\}$ and local defining map $\mu_M = f$. Thus we have

$$\tau(x)(n) = f\big(x(n), x(n+1), \dots, x(n+m)\big)$$

for all $x \in (\mathbb{S}^1)^{\mathbb{Z}}$ and $n \in \mathbb{Z}$. Then τ is a $\mathscr{C}$-cellular automaton for $\mathscr{C} = \mathrm{Man}$.

Example 3.10 (Arnold's cat cellular automaton) Let $G = \mathbb{Z}$. Let also $A = \mathbb{S}^1$ and $B = \mathbb{S}^1 \times \mathbb{S}^1 = \mathbb{T}^2$ and consider the map $\tau \colon A^{\mathbb{Z}} \to B^{\mathbb{Z}}$ defined by

$$\tau(x)(n) = \big(2x(n) + x(n+1), x(n) + x(n+1)\big)$$

for all $x \in A^{\mathbb{Z}}$. Then τ is a $\mathscr{C}$-cellular automaton for $\mathscr{C} = \mathrm{Man}$.

Given sets A and B, a subgroup H of a group G, and a cellular automaton $\tau \colon A^G \to B^G$ admitting a memory set contained in H, we have defined in Sect. 2.7 the cellular automaton $\tau_H \colon A^H \to B^H$ obtained by restriction of τ to H. We also introduced the converse operation, namely induction. It turns out that both restriction and induction of cellular automata preserve the property of being a $\mathscr{C}$-cellular automaton. More precisely, we have the following result.

Proposition 3.11 *Let G be a group and let H be a subgroup of G. Let also $\mathscr{C}$ be a category satisfying* (*CFP+*), *and let A and B be two $\mathscr{C}$-objects. Suppose that $\tau \colon A^G \to B^G$ is a cellular automaton over G admitting a memory set contained in H. Let $\tau_H \colon A^H \to B^H$ denote the cellular automaton over H obtained by restriction. Then τ is a $\mathscr{C}$-cellular automaton if and only if τ_H is a $\mathscr{C}$-cellular automaton.*

Proof If M is a memory set of τ contained in H, then M is also a memory set for τ_H. Moreover, τ and τ_H have the same associated local defining map $\mu_M \colon A^M \to B$ (Sect. 2.7). Therefore, the statement follows immediately from the definition of a $\mathscr{C}$-cellular automaton. □

Proposition 3.12 *Let G be a group and let $\mathscr{C}$ be a concrete category satisfying* (*CP*) *and* (*C+*) (*and hence* (*CFP+*)). *Let A and B be $\mathscr{C}$-objects and let $\tau \colon A^G \to B^G$ be a cellular automaton. Then the following conditions are equivalent*:

(a) *τ is a $\mathscr{C}$-morphism*;
(b) *τ is a $\mathscr{C}$-cellular automaton.*

Note that, in the preceding statement, A^G and B^G are $\mathscr{C}$-objects since $\mathscr{C}$ satisfies (CP).

Proof of Proposition 3.12 Let M be a memory set for τ. We then have

$$\big(\tau(x)\big)(g) = (\mu_M \circ \pi_M)\big(g^{-1}x\big) \quad \text{for every } x \in A^G, g \in G, \tag{9}$$

where $\pi_M \colon A^G \to A^M$ is the projection morphism. By definition, τ is a $\mathscr{C}$-cellular automaton if and only if μ_M is a $\mathscr{C}$-morphism.

Suppose first that μ_M is a $\mathscr{C}$-morphism. For each $g \in G$, the self-map of A^G given by $x \mapsto g^{-1}x$ is a $\mathscr{C}$-morphism since it just permutes coordinates of the $\mathscr{C}$-product A^G. On the other hand, the projection $\pi_M \colon A^G \to A^M$ is also a $\mathscr{C}$-morphism. Therefore, we deduce from (9) that the map from A^G to B given by $x \mapsto \tau(x)(g)$ is a $\mathscr{C}$-morphism for each $g \in G$. It follows that $\tau \colon A^G \to B^G$ is a $\mathscr{C}$-morphism.

Conversely, suppose that τ is a $\mathscr{C}$-morphism. Let us show that τ is a $\mathscr{C}$-cellular automaton. We can assume $A \neq \emptyset$. Denote by $p\colon A^G \to A^{\{1_G\}} = A$ the projection $\mathscr{C}$-morphism $x \mapsto x(1_G)$. Applying (9) with $g = 1_G$, we get

$$(p \circ \tau)(x) = \mu_M(y) \tag{10}$$

for all $x \in A^G$ and $y \in A^M$ with $x|_M = y$. As $\mathscr{C}$ satisfies (C+), the projection $\mathscr{C}$-morphism $\pi\colon A^G = A^M \times A^{G \setminus M} \to A^M$ is a retraction. Therefore there exists a $\mathscr{C}$-morphism $f\colon A^M \to A^G$ such that $\pi \circ f = \mathrm{Id}_{A^M}$. We then deduce from (10) that $\mu_M = p \circ \tau \circ f$. Consequently, μ_M is a $\mathscr{C}$-morphism. This shows that τ is a $\mathscr{C}$-cellular automaton. □

Example 3.13

(i) Take $\mathscr{C} = R$-Mod. Given two left R-modules A and B, a cellular automaton $\tau\colon A^G \to B^G$ is a $\mathscr{C}$-cellular automaton if and only if τ is R-linear with respect to the product R-module structures on A^G and B^G.

(ii) Take $\mathscr{C} = \mathrm{Top}$. Given two topological spaces A and B, a cellular automaton $\tau\colon A^G \to B^G$ is a $\mathscr{C}$-cellular automaton if and only if τ is continuous with respect to the product topologies on A^G and B^G (in general, these topologies are coarser than the prodiscrete topologies).

Proposition 3.14 *Let G be a group. Let $\mathscr{C}$ be a concrete category satisfying (CFP+), and let A, B and C be $\mathscr{C}$-objects. Suppose that $\tau\colon A^G \to B^G$ and $\sigma\colon B^G \to C^G$ are $\mathscr{C}$-cellular automata. Then $\sigma \circ \tau\colon A^G \to C^G$ is a $\mathscr{C}$-cellular automaton.*

Proof We have already seen (Proposition 2.9 and Remark 2.15) that $\sigma \circ \tau$ is a cellular automaton. Thus we are only left to show that $\sigma \circ \tau$ admits a local defining map which is a $\mathscr{C}$-morphism. Let S (resp. T) be a memory set for σ (resp. τ) and let $\mu\colon B^S \to C$ (resp. $\nu\colon A^T \to B$) denote the corresponding local defining map. Then, as we showed in the proof of Proposition 2.9, the set ST is a memory set for $\sigma \circ \tau$ and the map $\kappa\colon A^{ST} \to C$ defined by (6) is the corresponding local defining map. Now, since τ is a $\mathscr{C}$-cellular automaton, we have that ν is a $\mathscr{C}$-morphism. Moreover, the maps $\pi_s\colon A^{ST} \to A^{sT}$ and $f_s\colon A^{sT} \to A^T$ are $\mathscr{C}$-morphisms for every $s \in S$. As $\varphi_s = \nu \circ f_s \circ \pi_s$ (4), it follows that the product map $\Phi = \prod_{s \in S} \varphi_s\colon A^{ST} \to B^S$ (5) is a $\mathscr{C}$-morphism. Since σ is a $\mathscr{C}$-cellular automaton, its local defining map $\mu\colon B^S \to C$ is also a $\mathscr{C}$-morphism and therefore the map $\kappa = \mu \circ \Phi\colon A^{ST} \to C$ is a $\mathscr{C}$-morphism as well. This completes the proof that $\sigma \circ \tau$ is a $\mathscr{C}$-cellular automaton. □

Let G be a group and $\mathscr{C}$ a concrete category satisfying (CFP+). Then it follows from Example 3.6 and Proposition 3.14 that there is a category $\mathrm{CA}(G, \mathscr{C})$ having the same objects as $\mathscr{C}$, in which the identity morphism of an object A is the identity map $\mathrm{Id}_{A^G}\colon A^G \to A^G$, and where the morphisms from an object A to an object B are the $\mathscr{C}$-cellular automata $\tau\colon A^G \to B^G$ (with composition of morphisms given by the usual composition of maps). The category $\mathrm{CA}(G, \mathscr{C})$ is a concrete category when equipped with the functor $U\colon \mathrm{CA}(G,\mathscr{C}) \to \mathrm{Set}$ given by $U(A) = A^G$ and $U(\tau) = \tau$. Observe that the image of the functor U is a subcategory of the category $\mathrm{CA}(G)$ defined at the end of Sect. 2.3.

4 Projective Sequences of Sets

Let us briefly recall some elementary facts about projective sequences of sets and their projective limits.

A *projective sequence of sets* is a sequence $(X_n)_{n\in\mathbb{N}}$ of sets equipped with maps $f_{nm}\colon X_m \to X_n$, defined for all $n, m \in \mathbb{N}$ with $m \geqslant n$, and satisfying the following conditions:

(PS-1) f_{nn} is the identity map on X_n for all $n \in \mathbb{N}$;
(PS-2) $f_{nk} = f_{nm} \circ f_{mk}$ for all $n, m, k \in \mathbb{N}$ such that $k \geqslant m \geqslant n$.

We shall denote such a projective sequence by (X_n, f_{nm}) or simply by (X_n).

Observe that the projective sequence (X_n, f_{nm}) is entirely determined by the maps $g_n = f_{n,n+1}\colon X_{n+1} \to X_n$, $n \in \mathbb{N}$, since

$$f_{nm} = g_n \circ g_{n+1} \circ \cdots \circ g_{m-1} \tag{11}$$

for all $m > n$. Conversely, if we are given a sequence of maps $g_n\colon X_{n+1} \to X_n$, $n \in \mathbb{N}$, then there is a unique projective sequence (X_n, f_{nm}) satisfying (11).

The *projective limit* $X = \varprojlim X_n$ of the projective sequence of sets (X_n, f_{nm}) is the subset $X \subset \prod_{n\in\mathbb{N}} X_n$ consisting of the sequences $x = (x_n)_{n\in\mathbb{N}}$ satisfying $x_n = f_{nm}(x_m)$ for all $n, m \in \mathbb{N}$ such that $m \geqslant n$ (or, equivalently, $x_n = g_n(x_{n+1})$ for all $n \in \mathbb{N}$, where $g_n = f_{n,n+1}$). Note that there is a canonical map $\pi_n\colon X \to X_n$ sending x to x_n and $\pi_n = f_{nm} \circ \pi_m$ for all $m, n \in \mathbb{N}$ with $m \geqslant n$.

The fact that the projective limit $X = \varprojlim X_n$ is not empty clearly implies that all the sets X_n are nonempty. However, it can happen that the projective limit $X = \varprojlim X_n$ is empty even if all the sets X_n are nonempty. The following statement yields a sufficient condition for the projective limit to be nonempty.

Proposition 4.1 *Let (X_n, f_{nm}) be a projective sequence of sets and let $X = \varprojlim X_n$ denote its projective limit. Suppose that all maps $f_{nm}\colon X_m \to X_n$, $m \geqslant n$ are surjective. Then all canonical maps $\pi_n\colon X \to X_n$, $m \in \mathbb{N}$, are surjective. In particular, if in addition $X_{n_0} \neq \emptyset$ for some $n_0 \in \mathbb{N}$, then $X \neq \emptyset$.*

Proof Let $n \in \mathbb{N}$ and $x_n \in X_n$. As the maps $f_{k,k+1}$, $k \geqslant n$, are surjective, we can construct by induction a sequence $(x_k)_{k\geqslant n}$ such that $x_k = f_{k,k+1}(x_{k+1})$ for all $k \geqslant n$. Let us set $x_k = f_{kn}(x_n)$ for $k < n$. Then the sequence $x = (x_k)_{k\in\mathbb{N}}$ is in X and satisfies $x_n = \pi_n(x)$. This shows that π_n is surjective. □

Remark 4.2

(i) For the maps f_{nm}, $m \geqslant n$, to be surjective, it suffices that all the maps $f_{n,n+1}$ are surjective.

(ii) When the maps f_{nm} are all surjective, the following conditions are equivalent:
  (a) there exists $n_0 \in \mathbb{N}$ such that $X_{n_0} \neq \emptyset$;
  (b) $X_n \neq \emptyset$ for all $n \in \mathbb{N}$.

Let (X_n, f_{nm}) be a projective sequence of sets. Property (PS-2) implies that, for each $n \in \mathbb{N}$, the sequence of subsets $f_{nm}(X_m) \subset X_n$, $m \geqslant n$, is non-increasing. Let us set, for each $n \in \mathbb{N}$,

$$X'_n = \bigcap_{m \geqslant n} f_{nm}(X_m) \subset X_n.$$

The set X'_n is called the set of *universal elements* of X_n [23]. Observe that $f_{nm}(X'_m) \subset X'_n$ for all $m \geqslant n$. Thus, the map f_{nm} induces by restriction a map $f'_{nm}\colon X'_m \to X'_n$ for all $m \geqslant n$. Clearly (X'_n, f'_{nm}) is a projective sequence. This projective sequence is called the *universal projective sequence* associated with the projective sequence (X_n, f_{nm}).

Proposition 4.3 *Let (X_n, f_{nm}) be a projective sequence of sets and let (X'_n, f'_{nm}) be the associated universal projective sequence. Then*

$$\varprojlim X_n = \varprojlim X'_n. \tag{12}$$

Proof Let us set $X = \varprojlim X_n$ and $X' = \varprojlim X'_n$. Since $X'_n \subset X_n$ and f'_{nm} is the restriction of f_{nm} to X'_n, for all $n, m \in \mathbb{N}$ with $m \geqslant n$, we clearly have $X' \subset X$. To show the converse inclusion, let $x = (x_n)_{n \in \mathbb{N}} \in X$. We have $x_n = f_{nm}(x_m)$ for all $n, m \in \mathbb{N}$ such that $m \geqslant n$, so that $x_n \in \bigcap_{m \geqslant n} f_{nm}(X_m) = X'_n$. Since $f'_{nm}(x_n) = f_{nm}(x_n)$, we then deduce that $X \subset X'$. This shows (12). □

Corollary 4.4 *Let (X_n, f_{nm}) be a projective sequence of sets. Suppose that the following conditions are satisfied*:

(IP-1) *there exists $n_0 \in \mathbb{N}$ such that $\bigcap_{k \geqslant n_0} f_{n_0 k}(X_k) \neq \emptyset$;*
(IP-2) *for all $n, m \in \mathbb{N}$ with $m \geqslant n$ and all $x'_n \in \bigcap_{i \geqslant n} f_{ni}(X_i)$,*

$$\bigcap_{j \geqslant m} f_{nm}^{-1}\big(x'_n\big) \cap f_{mj}(X_j) \neq \emptyset.$$

Then $\varprojlim X_n \neq \emptyset$.

Proof Consider the universal projective sequence (X'_n, f'_{nm}) associated with the projective sequence (X_n, f_{nm}). Observe that condition (IP-1) says that $X'_{n_0} \neq \emptyset$. On the other hand, condition (IP-2) says that, for all $n, m \in \mathbb{N}$ with $m \geqslant n$, one has $f_{nm}^{-1}(x'_n) \cap X'_m \neq \emptyset$ for all $x'_n \in X'_n$, i.e., that the map f'_{nm} is surjective. Thus, applying Propositions 4.3 and 4.1, we get $\varprojlim X_n = \varprojlim X'_n \neq \emptyset$. □

Remark 4.5 Let (X_n, f_{nm}) be an arbitrary sequence of sets. We claim that, given $m \geqslant n$ and $x'_n \in \bigcap_{i \geqslant n} f_{ni}(X_i)$, one has $f_{nm}^{-1}(x'_n) \cap f_{mj}(X_j) \neq \emptyset$ for every $j \geqslant m$.

Indeed, since $x'_n \in f_{nj}(X_j)$, we can find $y_j \in X_j$ such that $x'_n = f_{nj}(y_j)$. Setting $z_m = f_{mj}(y_j)$, we then have $f_{nm}(z_m) = f_{nm} \circ f_{mj}(y_j) = x'_n$, so that $z_m \in f_{nm}^{-1}(x'_n) \cap f_{mj}(X_j)$. This proves our claim. It follows that the sets $f_{nm}^{-1}(x'_n) \cap f_{mj}(X_j)$ form a non-increasing sequence of nonempty subsets of X'_m. Condition (IP-2) says that the intersections of this sequence is not empty for all $m \geqslant n$ and $x'_n \in X'_n$.

5 Algebraic and Subalgebraic Subsets

5.1 Algebraic Subsets

Definition 5.1 Let $\mathscr{C}$ be a concrete category. Given a $\mathscr{C}$-object A, we say that a subset $X \subset A$ is *$\mathscr{C}$-algebraic* (or simply *algebraic* if there is no ambiguity on the category $\mathscr{C}$) if X is the inverse image of a point by some $\mathscr{C}$-morphism, i.e., if there exist a $\mathscr{C}$-object B, a point $b \in B$, and a $\mathscr{C}$-morphism $f\colon A \to B$ such that $X = f^{-1}(b)$.

Remark 5.2 If $\mathscr{C}$ is a concrete category admitting a terminal object which is reduced to a single element (this is for example the case when $\mathscr{C}$ is a concrete category satisfying (CFP+)), then every $\mathscr{C}$-object A is a $\mathscr{C}$-algebraic subset of itself. Indeed, we then have $A = f^{-1}(t)$, where $f\colon A \to T$ is the unique $\mathscr{C}$-morphism from A to T and t is the unique element of T.

Remark 5.3 Suppose that $\mathscr{C}$ is a concrete category satisfying (CFP) and that A is a $\mathscr{C}$-object. Then the set of $\mathscr{C}$-algebraic subsets of A is closed under finite intersections. Indeed, if $(X_i)_{i \in I}$ is a finite family of $\mathscr{C}$-algebraic subsets of a $\mathscr{C}$-object A, we can find $\mathscr{C}$-morphisms $f_i\colon A \to B_i$ and points $b_i \in B_i$ such that $X_i = f_i^{-1}(b_i)$. Then $\bigcap_{i \in I} X_i = f^{-1}(b)$, where $f = \prod_{i \in I} f_i\colon A \to B$, $B = \prod_{i \in I} B_i$ and $b = (b_i)_{i \in I}$.

Example 5.4

- (i) In the category Set or in its full subcategory Set^f, the algebraic subsets of an object A consist of all the subsets of A.
- (ii) In the category Grp, the algebraic subsets of an object G consist of the empty set $\varnothing$ and all the left-cosets (or, equivalently, all the right-cosets) of the normal subgroups of G, i.e., the subsets of the form gN, where $g \in G$ and N is a normal subgroup of G.
- (iii) In the category Rng, the algebraic subsets of an object R consist of $\varnothing$ and all the translates of the two-sided ideals of R, i.e., the subsets of the form $r + I$, where $r \in R$ and I is a two-sided ideal of R.
- (iv) In the category Fld, the algebraic subsets of an object K are $\varnothing$ and all the singletons $\{k\}$, $k \in K$.
- (v) In the category R-Mod, the algebraic subsets of an object M are $\varnothing$ and all the translates of the submodules of M, i.e., the subsets of the form $m + N$, where $m \in M$ and N is a submodule of M.

(vi) In the category Top, every subset of an object A is algebraic. Indeed, if A is a topological space and X is a nonempty subset of A, then $X = f^{-1}(b_0)$, where B is the quotient space of A obtained by identifying X to a single point b_0 and $f : A \to B$ is the quotient map.

(vii) In the full subcategory of Top whose objects consist of all the normal Hausdorff spaces, the algebraic subsets of an object A are precisely the closed subsets of A.

(viii) In the category K-Aal of affine algebraic sets over a field K, the algebraic subsets of an object A are precisely the subsets of A that are closed in the Zariski topology. If $A \subset K^n$, these subsets are the algebraic subsets of K^n (in the usual sense of algebraic geometry) that are contained in A.

5.2 Subalgebraic Subsets

Definition 5.5 Let $\mathscr{C}$ be a concrete category. Given a $\mathscr{C}$-object B, we say that a subset $Y \subset B$ is *$\mathscr{C}$-subalgebraic* (or simply *subalgebraic* if there is no ambiguity on the category $\mathscr{C}$) if Y is the image of some $\mathscr{C}$-algebraic subset by some $\mathscr{C}$-morphism, i.e., if there exist a $\mathscr{C}$-object A, a $\mathscr{C}$-algebraic subset $X \subset A$, and a $\mathscr{C}$-morphism $f : A \to B$ such that $Y = f(X)$.

Note that, if $\mathscr{C}$ is a concrete category and A is a $\mathscr{C}$-object, then every $\mathscr{C}$-algebraic subset $X \subset A$ is also $\mathscr{C}$-subalgebraic since $X = \mathrm{Id}_A(X)$. Note also that if $g : B \to C$ is a $\mathscr{C}$-morphism and Y is a $\mathscr{C}$-subalgebraic subset of B, then $g(Y)$ is a $\mathscr{C}$-subalgebraic subset of C.

Example 5.6

(i) In the category Set or in the category Set^f, the subalgebraic subsets of an object A coincide with its algebraic subsets, that is, they consist of all the subsets of A.

(ii) In the category Grp, every subalgebraic subset of an object G is either empty or of the form gH, where $g \in G$ and H is a (not necessarily normal) subgroup of G. Observe that every subgroup $H \subset G$ is subalgebraic since it is the image of the inclusion morphism $\iota : H \to G$. This shows in particular that there exist subalgebraic subsets that are not algebraic. On the other hand, a group may contain subgroup cosets which are not subalgebraic. Consider for example the symmetric group $G = S_3$. Then, if we take $g = (12)$ and $H = \langle (13) \rangle = \{1_G, (13)\}$, the coset $gH = \{(12), (123)\}$ is not a subalgebraic subset of G. Otherwise, there would exist a group G', a normal subgroup $H' \subset G'$, an element $g' \in G'$, and a group homomorphism $f : G' \to G$ such that $gH = f(g'H') = f(g')f(H')$. If f was surjective, then $f(H')$ would be normal in G and thus would have 1, 3 or 6 elements, which would contradict the fact that H has order 2. Therefore, the subgroup $f(G')$ must have either 1, 2 or 3 elements. But this is also impossible since the coset $gH = f(g')f(H')$ has two elements and does not contain 1_G.

(iii) In the category Rng, there are subalgebraic subsets that are not algebraic. For example, in the polynomial ring $\mathbb{Z}[t]$, the subring $\mathbb{Z}$ is subalgebraic but not algebraic.
(iv) In the category Fld, the subalgebraic subsets of an object K are its algebraic subsets, i.e., $\emptyset$ and all the singletons $\{k\}$, $k \in K$.
(v) In the category R-Mod, the subalgebraic subsets of an object M coincide with the algebraic subsets of M. Thus the subalgebraic subsets of M consist of $\emptyset$ and the translates of the submodules of M.
(vi) In the category Top, the subalgebraic subsets of an object A coincide with its algebraic subsets, i.e., they consist of all the subsets of A.
(vii) In the full subcategory of Top whose objects are Hausdorff topological spaces, the open interval $]0, 1[\subset \mathbb{R}$ is subalgebraic but not algebraic.
(viii) In the category CHT of compact Hausdorff topological spaces, the subalgebraic subsets of an object A coincide with its algebraic subsets, i.e., are the closed subsets of A.
(ix) In the category K-Aal of affine algebraic sets over an algebraically closed field K, it follows from Chevalley's theorem (see e.g. [7, AG Sect. 10.2] or [32, Theorem 10.2]) that every subalgebraic subsets of an object A is *constructible*, that is, a finite union of subsets of the form $U \cap V$, where $U \subset A$ is open and $V \subset A$ is closed for the Zariski topology.

5.3 The Subalgebraic Intersection Property

The following definition is due to Gromov [22, Sect. 4.C'].

Definition 5.7 Let $\mathscr{C}$ be a concrete category. We say that $\mathscr{C}$ satisfies the *subalgebraic intersection property*, briefly (SAIP), if for every $\mathscr{C}$-object A, every $\mathscr{C}$-algebraic subset $X \subset A$, and every non-increasing sequence $(Y_n)_{n \in \mathbb{N}}$ of $\mathscr{C}$-subalgebraic subsets of A with $X \cap Y_n \neq \emptyset$ for all $n \in \mathbb{N}$, one has $\bigcap_{n \in \mathbb{N}} X \cap Y_n \neq \emptyset$.

Let us introduce one more definition.

Definition 5.8 We say that a concrete category $\mathscr{C}$ is *Artinian* if the subalgebraic subsets of any $\mathscr{C}$-object satisfy the descending chain condition, i.e., if, given any $\mathscr{C}$-object A and any non-increasing sequence $(Y_n)_{n \in \mathbb{N}}$ of $\mathscr{C}$-subalgebraic subsets of A, there exists $n_0 \in \mathbb{N}$ such that $Y_n = Y_{n+1}$ for all $n \geqslant n_0$.

Proposition 5.9 *Every Artinian concrete category satisfies* (*SAIP*).

Proof Let $\mathscr{C}$ be an Artinian concrete category. Let A be a $\mathscr{C}$-object, X an algebraic subset of A, and $(Y_n)_{n \in \mathbb{N}}$ a non-increasing sequence of subalgebraic subsets of A with $X \cap Y_n \neq \emptyset$ for all $n \in \mathbb{N}$. As $\mathscr{C}$ is Artinian, we can find $n_0 \in \mathbb{N}$ such that

$Y_m = Y_{n_0}$ for all $m \geqslant n_0$. We then have

$$\bigcap_{n\in\mathbb{N}} X \cap Y_n = X \cap Y_{n_0} \neq \emptyset.$$

This shows that $\mathscr{C}$ satisfies (SAIP). □

Observe that if a concrete category $\mathscr{C}$ satisfies (SAIP) (resp. is Artinian), then every subcategory of $\mathscr{C}$ satisfies (SAIP) (resp. is Artinian).

Example 5.10

(i) The category Set does not satisfy (SAIP). Indeed, if we take $A = X = \mathbb{N}$ and $Y_n = \{m \in \mathbb{N} : m \geqslant n\}$, $n \in \mathbb{N}$, we have $X \cap Y_n \neq \emptyset$ for all $n \in \mathbb{N}$ but $\bigcap_{n\in\mathbb{N}} X \cap Y_n = \emptyset$.

(ii) The category Set^f is Artinian and hence satisfies (SAIP). Indeed, every non-increasing sequence of subsets of a finite set eventually stabilizes.

(iii) Let R be a nonzero ring and let $\mathscr{C} = R$-Mod. Consider the R-module $M = \bigoplus_{i\in\mathbb{N}} R$ and, for every $n \in \mathbb{N}$, denote by $\pi_n \colon M \to R^{n+1}$ the projection map defined by $\pi_n(m) = (m_0, m_1, \dots, m_n)$ for all $m = (m_i)_{i\in\mathbb{N}} \in M$. Let $y_n := (1, 1, \dots, 1) \in R^{n+1}$ and set $X_n := \pi_n^{-1}(y_n) = \{m = (m_i)_{i\in\mathbb{N}} \in M : m_0 = m_1 = \cdots = m_n = 1\}$. Then X_n is a nonempty $\mathscr{C}$-algebraic subset of M and we have $X_0 \subset X_1 \supset X_2 \supset \cdots$ but $\bigcap_{n\in\mathbb{N}} X_n = \emptyset$. This shows that R-Mod does not satisfy (SAIP) unless R is a zero ring.

(iv) Since $\mathbb{Z}$-Mod is a subcategory of Grp, namely the full subcategory of Grp whose objects are Abelian groups, we deduce that Grp does not satisfy (SAIP) either.

(v) If K is a field, then the category K-Vec $= K$-Mod does not satisfy (SAIP) since any field is a nonzero ring.

(vi) Given an integer $a \geqslant 2$, the subsets $X_n \subset \mathbb{Z}$ defined by

$$X_n = 1 + a + a^2 + \cdots + a^n + a^{n+1}\mathbb{Z},$$

$n \in \mathbb{N}$, are nonempty $\mathscr{C}$-algebraic subsets of $\mathbb{Z}$ for $\mathscr{C} =$ Rng and $\mathscr{C} = \mathbb{Z}$-Mod^{f-g}. As $X_0 \supset X_1 \supset X_2 \supset \cdots$ and $\bigcap_{n\in\mathbb{N}} X_n = \emptyset$ (see the remark following the proof of Proposition 2.2 in [13]), this shows that the categories Rng and $\mathbb{Z}$-Mod^{f-g} do not satisfy (SAIP).

(vii) Let R be a ring. Recall that a left module M is called *Artinian* if the submodules of M satisfy the descending chain condition (see for instance [25, Chap. 8]). It is clear that, in an Artinian module, the translates of the submodules also satisfy the descending chain condition [13, Proposition 2.2]. It follows that the full subcategory R-Mod^{Art} of R-Mod, whose objects consist of all the Artinian left R-modules, is Artinian and hence satisfies (SAIP). Note that R-Mod^{Art} satisfies (CFP+) since the direct sum of two Artinian modules is itself Artinian. If the ring R is *left-Artinian* (i.e., Artinian as a left module over itself), then every finitely generated left module over R is Artinian (Theorem 1.8 in [25, Chap. 8]), so that the category R-Mod^{f-g} is Artinian and hence satisfies (SAIP) [13].

(viii) If K is a field, then the category K-Vec$^{f-d} = K$-Mod^{Art} satisfies (SAIP).

(ix) The category Fld clearly satisfies (SAIP).

(x) The category Top clearly does not satisfy (SAIP). In fact, even the full subcategory of Top whose objects are Hausdorff topological spaces does not satisfy (SAIP) since, in this subcategory, the open intervals $(0, 1/n) \subset \mathbb{R}$ form a non-increasing sequence of subalgebraic subsets of $\mathbb{R}$ with empty intersection.

(xi) The category CHT of compact Hausdorff topological spaces (and therefore its full subcategory Man) satisfies (SAIP). This follows from the fact that, in a compact space, any family of closed subsets with the finite intersection property has a nonempty intersection. Observe that neither CHT nor Man are Artinian since the arcs $X_n := \{e^{i\theta} : 0 \leqslant \theta \leqslant 1/(n+1)\}$, $n \in \mathbb{N}$, form a non-increasing sequence of closed subsets of the unit circle $\mathbb{S}^1 := \{z \in \mathbb{C} : |z| = 1\}$ which does not stabilize.

(xii) Given a field K, the category K-Aal is Artinian if and only if K is finite. To see this, first observe that if K if finite then all objects in K-Aal are finite and hence K-Aal is Artinian. Then suppose that K is infinite and choose a sequence $(a_n)_{n\in\mathbb{N}}$ of distinct elements in K. Now observe that $Y_n = K \setminus \{a_0, a_1, \dots, a_n\}$ is a nonempty K-Aal-subalgebraic subset of K since it is the projection on the x-axis of the affine algebraic curve $X_n \subset K^2$ with equation $(x - a_0)(x - a_1)\cdots(x - a_n)y = 1$. As the sequence Y_n is non-increasing and does not stabilize, this shows that K-Aal is not Artinian.

(xiii) If K is an infinite countable field, then the category K-Aal does not satisfy (SAIP). Indeed, suppose that $K = \{a_n : n \in \mathbb{N}\}$. Then the subsets $Y_n = K \setminus \{a_0, a_1, \dots, a_n\}$ form a non-increasing sequence of nonempty subsets of K with empty intersection. As each Y_n is subalgebraic in K (see the preceding example), this shows that K-Aal does not satisfy (SAIP).

(xiv) If K is an uncountable algebraically closed field, then the category K-Aal of affine algebraic sets over K satisfies (SAIP) [22, Sect. 4.C"], [17, Proposition 4.4].

(xv) The category $\mathscr{C} = \mathbb{R}$-Aal of real affine algebraic sets does not satisfy (SAIP). Indeed, the subsets $X_n = [n, +\infty)$, $n \in \mathbb{N}$, are nonempty $\mathscr{C}$-subalgebraic subsets of $\mathbb{R}$ since $X_n = P_n(\mathbb{R})$ for $P_n(t) = t^2 + n$. On the other hand, we have $X_0 \supset X_1 \supset X_2 \supset \cdots$ but $\bigcap_{n\in\mathbb{N}} X_n = \emptyset$.

5.4 *Projective Sequences of Algebraic Sets*

Let $\mathscr{C}$ be a concrete category satisfying condition (CFP+). We say that a projective sequence (X_n, f_{nm}) of sets is a *projective sequence of $\mathscr{C}$-algebraic sets* provided there is a projective sequence (A_n, F_{nm}), consisting of $\mathscr{C}$-objects A_n and $\mathscr{C}$-morphisms $F_{nm} \colon A_m \to A_n$ for all $n, m \in \mathbb{N}$ such that $m \geqslant n$, satisfying the following conditions:

(PSA-1) X_n is a $\mathscr{C}$-algebraic subset of A_n for every $n \in \mathbb{N}$;
(PSA-2) $F_{nm}(X_m) \subset X_n$ and f_{nm} is the restriction of F_{nm} to X_m for all $m, n \in \mathbb{N}$ such that $m \geqslant n$.

Note that $f_{nm}(X_m) = F_{nm}(X_m)$ in PSA-2 above is a $\mathscr{C}$-subalgebraic subset of A_n for all $n, m \in \mathbb{N}$ such that $m \geqslant n$.

The following result constitutes an essential ingredient in the proof of Theorem 6.1.

Theorem 5.11 *Let $\mathscr{C}$ be a concrete category satisfying (SAIP). Suppose that (X_n, f_{nm}) is a projective sequence of nonempty $\mathscr{C}$-algebraic sets. Then* $\varprojlim X_n \neq \emptyset$.

Proof Let (A_n, F_{nm}) be a projective sequence of $\mathscr{C}$-objects and morphisms satisfying conditions (PSA-1) and (PSA-2) above. Let $n \in \mathbb{N}$. For all $k \geqslant n$, the image set $f_{nk}(X_k) = F_{nk}(X_k)$, being the image of a nonempty $\mathscr{C}$-algebraic subset under a $\mathscr{C}$-morphism, is a nonempty $\mathscr{C}$-subalgebraic subset of A_n. As the sequence $f_{nk}(X_k)$, $k = n, n+1, \ldots,$ is non-increasing, we deduce from (SAIP) that

$$X'_n = \bigcap_{k \geqslant n} f_{nk}(X_k) \neq \emptyset \tag{13}$$

for all $n \in \mathbb{N}$.

Let us fix $m, n \in \mathbb{N}$ with $m \geqslant n$ and $x'_n \in X'_n$. In Remark 4.5, we observed that $f_{nm}^{-1}(x'_n) \cap f_{mj}(X_j) \neq \emptyset$ for all $j \geqslant m$. On the other hand, we have

$$f_{nm}^{-1}\left(x'_n\right) \cap f_{mj}(X_j) = F_{nm}^{-1}\left(x'_n\right) \cap F_{mj}(X_j).$$

As the sets $F_{mj}(X_j)$, for $j = m, m+1, \ldots,$ form a non-increasing sequence of $\mathscr{C}$-subalgebraic subsets of A_m and $F_{nm}^{-1}(x'_n)$ is a $\mathscr{C}$-algebraic subset of A_m, we get

$$\bigcap_{j \geqslant m} f_{nm}^{-1}\left(x'_n\right) \cap f_{mj}(X_j) \neq \emptyset$$

by applying (SAIP) again. This is condition (IP-2) in Corollary 4.4. Since (IP-1) follows from (13), Corollary 4.4 ensures that $\varprojlim X_n \neq \emptyset$. □

6 The Closed Image Property

One says that a map $f : X \to Y$ from a set X into a topological space Y has the *closed image property*, briefly (CIP) [22, Sect. 4.C"], if its image $f(X)$ is closed in Y.

Theorem 6.1 *Let $\mathscr{C}$ be a concrete category satisfying conditions (CFP+) and (SAIP). Let G be a group. Then every $\mathscr{C}$-cellular automaton $\tau : A^G \to B^G$ satisfies (CIP) with respect to the prodiscrete topology on B^G.*

Proof Let $\tau\colon A^G \to B^G$ be a $\mathscr{C}$-cellular automaton. Let $M \subset G$ be a memory set for τ and let $\mu_M\colon A^M \to B$ denote the associated local defining map.

Suppose first that the group G is countable. Then we can find a sequence $(E_n)_{n\in\mathbb{N}}$ of finite subsets of G such that $G = \bigcup_{n\in\mathbb{N}} E_n$, $M \subset E_0$, and $E_n \subset E_{n+1}$ for all $n \in \mathbb{N}$. Consider, for each $n \in \mathbb{N}$, the finite subset $F_n \subset G$ defined by $F_n = \{g \in G : gM \subset E_n\}$. Note that $G = \bigcup_{n\in\mathbb{N}} F_n$, $1_G \in F_0$, and $F_n \subset F_{n+1}$ for all $n \in \mathbb{N}$.

It follows from (1) that if x and x' are elements in A^G such that x and x' coincide on E_n, then the configurations $\tau(x)$ and $\tau(x')$ coincide on F_n. Therefore, we can define a map $\tau_n\colon A^{E_n} \to B^{F_n}$ by setting

$$\tau_n(u) = \bigl(\tau(x)\bigr)\big|_{F_n}$$

for all $u \in A^{E_n}$, where $x \in A^G$ denotes an arbitrary configuration extending u. Observe that both A^{E_n} and B^{F_n} are $\mathscr{C}$-objects as they are finite Cartesian powers of the $\mathscr{C}$-objects A and B respectively.

We claim that τ_n is a $\mathscr{C}$-morphism for every $n \in \mathbb{N}$. Indeed, let $n \in \mathbb{N}$. For every $g \in F_n$, we have $gM \subset E_n$. Denote, for each $g \in F_n$, by $\pi_{n,g}\colon A^{E_n} \to A^{gM}$ the projection $\mathscr{C}$-morphism. Consider also, for all $g \in G$, the $\mathscr{C}$-isomorphism $\phi_g\colon A^{gM} \to A^M$ defined by $(\phi_g(u))(m) = u(gm)$ for all $m \in M$. Then, for each $g \in F_n$, the map $\Phi_g := \mu_M \circ \phi_g \circ \pi_{n,g}\colon A^{E_n} \to B$ is a $\mathscr{C}$-morphism since it is the composite of $\mathscr{C}$-morphisms. Observe that $\Phi_g(x) = \tau(x)(g)$ for all $x \in A^G$. This shows that $\tau_n = \prod_{g\in F_n} \Phi_g$ and the claim follows.

Let now $y \in B^G$ and suppose that y is in the closure of $\tau(A^G)$ for the prodiscrete topology on B^G. Then, for every $n \in \mathbb{N}$, we can find $z_n \in A^G$ such that

$$y|_{F_n} = \bigl(\tau(z_n)\bigr)\big|_{F_n}. \tag{14}$$

Consider, for each $n \in \mathbb{N}$, the $\mathscr{C}$-algebraic subset $X_n \subset A^{E_n}$ defined by $X_n = \tau_n^{-1}(y|_{F_n})$. We have $X_n \neq \emptyset$ for all $n \in \mathbb{N}$ since $z_n|_{E_n} \in X_n$ by (14). Observe that, for $m \geqslant n$, the projection $\mathscr{C}$-morphism $F_{nm}\colon A^{E_m} \to A^{E_n}$ induces by restriction a map $f_{nm}\colon X_m \to X_n$. Conditions (PSA-1) and (PSA-2) are trivially satisfied so that (X_n, f_{nm}) is a projective sequence of nonempty $\mathscr{C}$-algebraic sets. Since $\mathscr{C}$ satisfies (SAIP), we have $\varprojlim X_n \neq \emptyset$ by Theorem 5.11. Choose an element $(x_n)_{n\in\mathbb{N}} \in \varprojlim X_n$. Thus $x_n \in A^{E_n}$ and x_{n+1} coincides with x_n on E_n for all $n \in \mathbb{N}$. As $G = \bigcup_{n\in\mathbb{N}} E_n$, we deduce that there exists a (unique) configuration $x \in A^G$ such that $x|_{E_n} = x_n$ for all $n \in \mathbb{N}$. Moreover, we have $\tau(x)|_{F_n} = \tau_n(x_n) = y_n = y|_{F_n}$ for all n since $x_n \in X_n$. As $G = \bigcup_{n\in\mathbb{N}} F_n$, this shows that $\tau(x) = y$. This completes the proof that τ satisfies condition (CIP) in the case when G is countable.

Let us treat now the case of an arbitrary (possibly uncountable) group G. Let H denote the subgroup of G generated by the memory set M. Observe that H is countable since M is finite. The restriction cellular automaton $\tau_H\colon A^H \to B^H$ is a $\mathscr{C}$-cellular automaton by Proposition 3.11. Thus, by the first part of the proof, τ_H satisfies condition (CIP), that is, $\tau_H(A^H)$ is closed in B^H for the prodiscrete topology on B^H. By applying part (iii) of Theorem 2.16, we deduce that $\tau(A^G)$ is also closed in B^G for the prodiscrete topology on B^G. Thus τ satisfies condition (CIP). □

From Theorem 6.1 and Examples 5.10 we recover results from [10, 14, 16, 17, 22] and [13, Lemma 3.2].

Corollary 6.2 *Let G be a group. Then every $\mathscr{C}$-cellular automaton $\tau\colon A^G \to B^G$ satisfies* (*CIP*) *with respect to the prodiscrete topology on B^G, when $\mathscr{C}$ is one of the following concrete categories*:

- Set^f, *the category of finite sets*;
- K-Vec^{f-d}, *the category of finite-dimensional vector spaces over an arbitrary field K*;
- R-Mod^{Art}, *the category of left Artinian modules over an arbitrary ring R*;
- R-Mod^{f-g}, *the category of finitely generated left modules over an arbitrary left-Artinian ring R*;
- K-Aal, *the category of affine algebraic sets over an arbitrary uncountable algebraically closed field K*;
- CHT, *the category of compact Hausdorff topological spaces*;
- Man, *the category of compact topological manifolds.*

Remark 6.3 When $\mathscr{C}$ is Set^f, we can directly deduce that any $\mathscr{C}$-cellular automaton $\tau\colon A^G \to B^G$ satisfies (CIP) from the compactness of A^G, the continuity of τ, and the fact that B^G is Hausdorff.

Remark 6.4

(i) It is shown in [16] (see also [15, Example 8.8.3]) that if G is a non-periodic group and A is an infinite set, then there exists a cellular automaton $\tau\colon A^G \to A^G$ whose image $\tau(A^G)$ is not closed in A^G for the prodiscrete topology.

(ii) Let K be a field and let $\mathscr{C} = K$-Vec. It is shown in [16] (see also [15, Example 8.8.3]) that is G is a non-periodic group and A is an infinite-dimensional vector space over K, then there exists a $\mathscr{C}$-cellular automaton $\tau\colon A^G \to A^G$ whose image $\tau(A^G)$ is not closed in A^G for the prodiscrete topology.

(iii) In [17, Remark 5.2], it is shown that if G is a non-periodic group, then there exists a $\mathbb{R}$-Aal-cellular automaton $\tau\colon \mathbb{R}^G \to \mathbb{R}^G$ whose image is not closed in $\mathbb{R}^G$ for the prodiscrete topology.

7 Surjunctivity of $\mathscr{C}$-Cellular Automata over Residually Finite Groups

7.1 Injectivity and Surjectivity in Concrete Categories

Let $\mathscr{C}$ be a category and $f\colon A \to B$ a $\mathscr{C}$-morphism. We recall that f is said to be a *$\mathscr{C}$-monomorphism* provided that for any two $\mathscr{C}$-morphisms $g_1, g_2\colon C \to A$ the equality $f \circ g_1 = f \circ g_2$ implies $g_1 = g_2$. We also recall that f is called a *$\mathscr{C}$-epimorphism* if for any two $\mathscr{C}$-morphisms $h_1, h_2\colon B \to C$ the equality $h_1 \circ f =$

$h_2 \circ f$ implies $h_1 = h_2$. In other words, a $\mathscr{C}$-monomorphism (resp. $\mathscr{C}$-epimorphism) is a left-cancellative (resp. right-cancellative) $\mathscr{C}$-morphism.

Suppose now that $\mathscr{C}$ is a concrete category. Then one says that a $\mathscr{C}$-morphism $f\colon A \to B$ is *injective* (resp. *surjective*, resp. *bijective*) if (the underlying map of) f is injective (resp. surjective, resp. bijective) in the set-theoretical sense. It is clear that every injective (resp. surjective) $\mathscr{C}$-morphism is a $\mathscr{C}$-monomorphism (resp. a $\mathscr{C}$-epimorphism). In concrete categories such as Set, Set^f, Grp, R-Mod, R-Mod^{f-g}, Top, CHT, or Man the converse is also true so that the class of injective (resp. surjective) morphisms coincide with the class of monomorphisms (resp. epimorphisms) in these categories (the fact that every epimorphism is surjective in Grp is a non-trivial result, see [29]). However, there exist concrete categories admitting monomorphisms (resp. epimorphisms) that fail to be injective (resp. surjective).

Example 7.1

(i) Let $\mathscr{C}$ be the full subcategory of Grp consisting of all divisible Abelian groups. Recall that an Abelian group G is called *divisible* if for each $g \in G$ and each integer $n \geqslant 1$, there is an element $g' \in G$ such that $ng' = g$. Then the quotient map $f : \mathbb{Q} \to \mathbb{Q}/\mathbb{Z}$ is a non-injective $\mathscr{C}$-monomorphism.
(ii) Let $\mathscr{C} = $ Rng. Then the inclusion map $f : \mathbb{Z} \to \mathbb{Q}$ is a non-surjective $\mathscr{C}$-epimorphism.
(iii) Let $\mathscr{C}$ be the full subcategory of Top whose objects are Hausdorff spaces. Then the inclusion map $f : \mathbb{Q} \to \mathbb{R}$ is a non-surjective $\mathscr{C}$-epimorphism.

7.2 Surjunctive Categories

Definition 7.2 A concrete category $\mathscr{C}$ is said to be *surjunctive* if every injective $\mathscr{C}$-endomorphism $f\colon A \to A$ is surjective (and hence bijective).

Example 7.3

(i) The category Set is not surjunctive but Set^f is. Indeed, a set A is finite if and only if every injective map $f\colon A \to A$ is surjective (Dedekind's characterization of infinite sets).
(ii) The map $f : \mathbb{Z} \to \mathbb{Z}$, defined by $f(n) = 2n$ for all $n \in \mathbb{Z}$, is injective but not surjective. This shows that the categories Grp, $\mathbb{Z}$-Mod and $\mathbb{Z}$-Mod^{f-g} are not surjunctive.
(iii) Let R be a nonzero ring. Then the map $f\colon R[t] \to R[t]$, defined by $P(t) \mapsto P(t^2)$, is injective but not surjective. As f is both a ring and a R-module endomorphism, this shows that the categories Rng and R-Mod are not surjunctive.
(iv) If k is a field and $k(t)$ is the field of rational functions on k, then the map $f : k(t) \to k(t)$, defined by $F(t) \mapsto F(t^2)$, is a field homomorphism which is injective but not surjective. This shows that the category Fld is not surjunctive.

(v) Let K be a field. Then the category K-Vec is not surjunctive but K-Vec^{f-d} is. Indeed, it is well known from basic linear algebra that for a vector space A over K one has $\dim_K(A) < \infty$ if and only if every injective K-linear map $f: A \to A$ is surjective.

(vi) If R is a ring then the category R-Mod^{Art} of Artinian left-modules over R is surjunctive (see e.g. [13, Proposition 2.1]).

(vii) Let R be a left Artinian ring. Then the category R-Mod^{f-g} of finitely-generated left R-modules over R is surjunctive [13, Proposition 2.5].

(viii) In [36], it is shown that, for a commutative ring R, the category R-Mod^{f-g} is surjunctive if and only if all prime ideals in R are maximal (if R is a nonzero ring, this amounts to saying that R has Krull dimension 0). The non-commutative rings R such that R-Mod^{f-g} is surjunctive are characterized in [2].

(ix) Let K be a field. If K is algebraically closed, then the category K-Aal of affine algebraic sets over K is surjunctive: this is a particular case of the *Ax-Grothendieck theorem* [3, Theorem C], [4], and [23, Proposition 10.4.11] (see also [6]).

When the ground field K is not algebraically closed, the category K-Aal may fail to be surjunctive. For instance, the injective polynomial map $f: \mathbb{Q} \to \mathbb{Q}$ defined by $f(x) = x^3$ is not surjective since $2 \notin f(\mathbb{Q})$. This shows that the category $\mathbb{Q}$-Aal is not surjunctive. If k is any field of characteristic $p > 0$ (e.g., $k = \mathbb{Z}/p\mathbb{Z}$) and we denote by $K = k(t)$ the field of rational functions with coefficients in k in one indeterminate t, then, the map $f: k(t) \to k(t)$, defined by $f(R) = R^p$ for all $R \in k(t)$, is injective (it is the *Frobenius endomorphism* of the field $k(t)$) but not surjective since there is no $R \in k(t)$ such that $t = R^p$. Thus, the category $k(t)$-Aal is not surjunctive for any field k of characteristic $p > 0$.

(x) The categories Top and CHT are not surjunctive. Indeed, if we consider the unit interval $[0, 1] \subset \mathbb{R}$, the continuous map $f: [0, 1] \to [0, 1]$, defined by $f(x) = x/2$ for all $x \in [0, 1]$, is injective but not surjective.

(xi) Let M be a compact topological manifold and suppose that $f: M \to M$ is an injective continuous map. Then $f(M)$ is open in M by Brouwer's *invariance of domain* and closed by compactness of M. Since M is connected, we deduce that $f(M) = M$. This shows that the category Man is surjunctive.

7.3 Surjunctive Groups

Definition 7.4 Let $\mathscr{C}$ be a concrete category satisfying condition (CFP+). One says that a group G is *$\mathscr{C}$-surjunctive* if every injective $\mathscr{C}$-cellular automaton $\tau: A^G \to A^G$ is surjective. In other words, G is $\mathscr{C}$-surjunctive if the category CA$(G, \mathscr{C})$ of $\mathscr{C}$-cellular automata over G is surjunctive.

Remark 7.5

(i) The trivial group is $\mathscr{C}$-surjunctive if and only if the category $\mathscr{C}$ is surjunctive.
(ii) There exist no $\mathscr{C}$-surjunctive groups in the case when the category $\mathscr{C}$ is not surjunctive. Indeed, if $f\colon A \to A$ is a $\mathscr{C}$-morphism which is injective but not surjective and G is any group, then the map $\tau\colon A^G \to A^G$, defined by $\tau(x)(g) = f(x(g))$ for all $x \in A^G$ and $g \in G$, is a $\mathscr{C}$-cellular automaton (with memory set $\{1_G\}$) which is injective but not surjective.
(iii) A group G is Set^f-surjunctive if and only if, for any finite alphabet A, every injective cellular automaton $\tau\colon A^G \to A^G$ is surjective. Gottschalk [21] called such a group a *surjunctive group*.

7.4 Residually Finite Groups

Recall that a group G is called *residually finite* if the intersection of its finite index subgroups is reduced to the identity element. This is equivalent to saying that if g_1 and g_2 are distinct elements in G, then we can find a finite group F and a group homomorphism $f\colon G \to F$ such that $f(g_1) \neq f(g_2)$. All finite groups, all free groups, all finitely generated nilpotent groups (and hence all finitely generated Abelian groups, e.g. $\mathbb{Z}^d$ for any $d \in \mathbb{N}$), and all fundamental groups of compact topological manifolds of dimension $\leqslant 3$ are residually finite. All finitely generated linear groups are residually finite by a theorem of Mal'cev. On the other hand, the additive group $\mathbb{Q}$, the group of permutations of $\mathbb{N}$, the Baumslag-Solitar group $BS(2,3) = \langle a, b : a^{-1}b^2a = b^3\rangle$, and all infinite simple groups provide examples of groups which are not residually finite.

The following dynamical characterization of residual finiteness is well known (see e.g. [15, Theorem 2.7.1]):

Theorem 7.6 *Let G be a group. Then the following conditions are equivalent*:

(a) *the group G is residually finite*;
(b) *for every set A, the set of points of A^G which have a finite G-orbit is dense in A^G for the prodiscrete topology.*

7.5 Surjunctivity of Residually Finite Groups

Theorem 7.7 *Let $\mathscr{C}$ be a concrete category satisfying conditions* (*CFP*+) *and* (*SAIP*). *Suppose that $\mathscr{C}$ is surjunctive. Then every residually finite group is $\mathscr{C}$-surjunctive. In other words, every injective $\mathscr{C}$-cellular automaton $\tau\colon A^G \to A^G$ is surjective when G is residually finite* (*e.g.*, $G = \mathbb{Z}^d$).

Before proving Theorem 7.7, let us introduce some additional notation.

Let A, M, and N be sets. Suppose that we are given a map $\rho\colon M \to N$. Then ρ induces a map $\rho^*\colon A^N \to A^M$ defined by $\rho^*(y) = y \circ \rho$ for all $y \in A^N$.

Lemma 7.8 *Let $\mathscr{C}$ be a concrete category satisfying condition (CFP). Let A be a $\mathscr{C}$-object and suppose that we are given a map $\rho\colon M \to N$, where M and N are finite sets. Then the induced map $\rho^*\colon A^N \to A^M$ is a $\mathscr{C}$-morphism.*

Proof We have $\rho^*(y)(m) = y(\rho(m))$ for all $m \in M$ and $y \in A^N$. Thus, if we denote by $\pi_n\colon A^N \to A$, $n \in N$, the $\mathscr{C}$-morphism given by the projection map on the n-factor, we have $\rho^* = \prod_{m\in M} \pi_{\rho(m)}$. Consequently, ρ^* is a $\mathscr{C}$-morphism. □

Proof of Theorem 7.7 Let G be a residually finite group and suppose that $\tau\colon A^G \to A^G$ is an injective $\mathscr{C}$-cellular automaton. For every finite index subgroup H of G we denote by $\mathrm{Fix}(H)$ the subset of A^G consisting of all configurations $x \in A^G$ that are fixed by H, that is, such that $hx = x$ for all $h \in H$. We also denote by $H\backslash G = \{Hg : g \in G\}$ the finite set consisting of all right cosets of H in G and by $\rho_H\colon G \to H\backslash G$ the canonical surjective map sending each $g \in G$ to Hg. Consider the induced map $\rho_H^*\colon A^{H\backslash G} \to A^G$. One immediately checks that $\rho_H^*(A^{H\backslash G}) \subset \mathrm{Fix}(H)$. In fact, the map $\rho_H^*\colon A^{H\backslash G} \to \mathrm{Fix}(H)$ is bijective (see e.g. [15, Proposition 1.3.3]). Observe now that by G-equivariance of τ we have $\tau(\mathrm{Fix}(H)) \subset \mathrm{Fix}(H)$. Denote by $\sigma := \tau|_{\mathrm{Fix}(H)}\colon \mathrm{Fix}(H) \to \mathrm{Fix}(H)$ the map obtained by restricting τ to $\mathrm{Fix}(H)$ and let $\tilde{\sigma}\colon A^{H\backslash G} \to A^{H\backslash G}$ be the conjugate of σ by ρ_H^*, that is, the map given by $\tilde{\sigma} = (\rho_H^*)^{-1} \circ \sigma \circ \rho_H^*$. We claim that $\tilde{\sigma}$ is a $\mathscr{C}$-morphism. To see this, it suffices to prove that, for each $t \in H\backslash G$, the map $\pi_t\colon A^{H\backslash G} \to A$ defined by $\pi_t(y) = \tilde{\sigma}(y)(t)$ is a $\mathscr{C}$-morphism, since then $\tilde{\sigma} = \prod_{t\in T} \pi_t$. Choose a memory set M for τ and let $\mu_M\colon A^M \to A$ denote the associated local defining map. For $t = gH \in T$, consider the map $\psi_t\colon M \to H\backslash G$ defined by $\psi_t(m) = \rho_H(gm)$ for all $m \in M$. It is obvious that ψ_t is well defined (i.e. it does not depend on the particular choice of the representative $g \in G$ of the coset $t = Hg$). If $\psi^*\colon A^{H\backslash G} \to A^M$ is the induced map, we then have $\pi_t = \mu_M \circ \psi_t^*$. But μ_M is a $\mathscr{C}$-morphism since τ is a $\mathscr{C}$-cellular automaton. On the other hand, ψ_t^* is also a $\mathscr{C}$-morphism by Lemma 7.8. It follows that π_t is a $\mathscr{C}$-morphism, proving our claim. Now observe that $\sigma : \mathrm{Fix}(H) \to \mathrm{Fix}(H)$ is injective since it is the restriction of τ. As $\tilde{\sigma}$ is conjugate to σ, we deduce that $\tilde{\sigma}$ is injective as well. Since by our assumptions the category $\mathscr{C}$ is surjunctive, we deduce that $\tilde{\sigma}$ is surjective. Thus, σ is also surjective and hence $\mathrm{Fix}(H) = \sigma(\mathrm{Fix}(H)) = \tau(\mathrm{Fix}(H)) \subset \tau(A^G)$.

Let $E \subset A^G$ denote the set of configurations whose orbit under the G-shift is finite. Then we have

$$E = \bigcup_{H\in\mathscr{F}} \mathrm{Fix}(H) \subset \tau\big(A^G\big),$$

where $\mathscr{F}$ denotes the set of all finite index subgroups of G. On the other hand, the residual finiteness of G implies that E is dense in A^G (Theorem 7.6). As $\tau(A^G)$ is closed in A^G by Theorem 6.1, we conclude that $\tau(A^G) = A^G$. □

From Theorem 7.7, Examples 7.3 and 5.10 we deduce the following:

Corollary 7.9 *All residually finite groups are $\mathscr{C}$-surjunctive when $\mathscr{C}$ is one of the following concrete categories*:

- Set^f, *the category of finite sets*;
- $K\text{-Vec}^{f-d}$, *the category of finite-dimensional vector spaces over an arbitrary field* K;
- $R\text{-Mod}^{Art}$, *the category of left Artinian modules over an arbitrary ring* R;
- $R\text{-Mod}^{f-g}$, *the category of finitely generated left modules over an arbitrary left-Artinian ring* R;
- K-Aal, *the category of affine algebraic sets over an arbitrary uncountable algebraically closed field* K;
- Man, *the category of compact topological manifolds.*

Remark 7.10

(i) Let $\mathscr{C} = \text{Set}^f$. The $\mathscr{C}$-surjunctivity of residually finite groups was established by W. Lawton [21] (see also [15, Theorem 3.3.1]). As mentioned in the Introduction, all amenable groups are $\mathscr{C}$-surjunctive [15, Corollary 5.9.3]. These results were generalized by Gromov [22] and Weiss [38] (see also [15, Theorem 7.8.1]) who proved that all *sofic* groups are $\mathscr{C}$-surjunctive. It is not known whether all groups are $\mathscr{C}$-surjunctive (resp. sofic) or not.

(ii) Let $\mathscr{C} = K\text{-Vec}^{f-d}$, where K is an arbitrary field. In [10] (see also [22]) we proved that residually finite groups and amenable groups are $\mathscr{C}$-surjunctive. More generally, in [11] (see also [22] and [15, Theorem 8.14.4]) we proved that all sofic groups are $\mathscr{C}$-surjunctive. We also proved (see [11] and [15, Corollary 8.15.7]) that a group G is $\mathscr{C}$-surjunctive, if and only if the group ring $K[G]$ is *stably finite*, that is, the following condition holds: if two square matrices a and b with entries in the group ring $K[G]$ satisfy $ab = 1$, then they also satisfy $ba = 1$. We recall that Kaplansky [27] conjectured that all group rings are stably finite. He proved the conjecture when the ground field K has characteristic zero, but for positive characteristic, though proved for all sofic groups by Elek and Szabo [19] (see also [11] and [15, Corollary 8.15.8]), the Kaplansky conjecture remains open. In other words, it is not known whether all groups are $\mathscr{C}$-surjunctive or not when K has positive characteristic.

(iii) In [12, Corollary 1.3], it is shown that if R is a left-Artinian ring and $\mathscr{C} = R\text{-Mod}^{f-g}$, then every sofic group is $\mathscr{C}$-surjunctive.

(iv) Let $\mathscr{C} = K$-Aal, where K is an uncountable algebraically closed field. The fact that all residually finite groups are $\mathscr{C}$-surjunctive was established in [17, Corollary 1.2] (see also [22]). We do not know how to prove that all amenable (resp. all sofic) groups are $\mathscr{C}$-surjunctive.

8 Reversibility of $\mathscr{C}$-Cellular Automata

8.1 The Subdiagonal Intersection Property

Let $\mathscr{C}$ be a concrete category satisfying (CFP) and let $f\colon A \to B$ be a $\mathscr{C}$-morphism. Consider the map $g\colon A \times A \to B \times B$ defined by $g(a_1, a_2) = (f(a_1), f(a_2))$ for all

$(a_1, a_2) \in A \times A$. If $\pi_i^A \colon A \times A \to A$ and $\pi_i^B \colon B \times B \to B$, $i = 1, 2$, denote the projection morphisms, we have $\pi_i^B \circ g = f \circ \pi_i^A$ for $i = 1, 2$. Therefore, the maps $\pi_1^B \circ g$ and $\pi_2^B \circ g$ are $\mathscr{C}$-morphisms. We deduce that g is also a $\mathscr{C}$-morphism. We shall write this morphism $g = f \times f$ and call it the *square* of f (not to be confused with the product map $h \colon A \to B \times B$ defined by $h(a) = (f(a), f(a))$ for all $a \in A$).

Definition 8.1 Let $\mathscr{C}$ be a concrete category satisfying condition (CFP). Let A be a $\mathscr{C}$-object.

A subset $X \subset A \times A$ is called *$\mathscr{C}$-square-algebraic* if it is the inverse image of a point by the square of some $\mathscr{C}$-morphism, i.e., if there exists a $\mathscr{C}$-morphism $f \colon A \to B$ and an element $(b_1, b_2) \in B \times B$ such that

$$X = \{(a_1, a_2) \in A \times A : f(a_1) = b_1 \text{ and } f(a_2) = b_2\}.$$

A subset $X \subset A \times A$ is called *$\mathscr{C}$-prediagonal* if there exists a $\mathscr{C}$-morphism $f \colon A \to B$ such that

$$X = \{(a_1, a_2) \in A \times A : f(a_1) = f(a_2)\}.$$

In other words, a subset $X \subset A \times A$ is $\mathscr{C}$-prediagonal if and only if X is the inverse image of the diagonal $\Delta_B \subset B \times B$ by the square $f \times f \colon A \times A \to B \times B$ of some $\mathscr{C}$-morphism $f \colon A \to B$.

A subset $X \subset A \times A$ is called *$\mathscr{C}$-codiagonal* if there exists a $\mathscr{C}$-morphism $f \colon A \to B$ such that

$$Y = \{(a_1, a_2) \in A \times A : f(a_1) \neq f(a_2)\}.$$

In other words, a subset of $A \times A$ is $\mathscr{C}$-codiagonal if and only if it is the complement of a $\mathscr{C}$-diagonal subset.

A subset $X \subset A \times A$ is called $\mathscr{C}$-subdiagonal if it is the image of a $\mathscr{C}$-prediagonal subset by the square of a $\mathscr{C}$-morphism. In other words, a subset $X \subset A \times A$ is subdiagonal if and only if there exist $\mathscr{C}$-morphisms $f \colon B \to A$ and $g \colon B \to C$ such that

$$X = \{\big(f(b_1), f(b_2)\big) : (b_1, b_2) \in B \times B \text{ and } g(b_1) = g(b_2)\}.$$

Remark 8.2 Suppose that $\mathscr{C}$ is a concrete category satisfying (CFP) and that A is a $\mathscr{C}$-object. Then:

(i) Every $\mathscr{C}$-square-algebraic subset $X \subset A \times A$ is a $\mathscr{C}$-algebraic subset of $A \times A$.
(ii) The set $A \times A$ is a $\mathscr{C}$-square-algebraic subset of itself. Indeed, if $T = \{t\}$ is a terminal $\mathscr{C}$-object, we have $A \times A = \{(a_1, a_2) \in A \times A : f(a_1) = t \text{ and } f(a_2) = t\}$, where $f \colon A \to T$ denotes the unique $\mathscr{C}$-morphism from A to T.
(iii) Every prediagonal subset $X \subset A \times A$ contains the diagonal $\Delta_A \subset A \times A$. Moreover, Δ_A is a $\mathscr{C}$-prediagonal subset of $A \times A$ since $\Delta_A = \{(a_1, a_2) \in A \times A : f(a_1) = f(a_2)\}$ for $f = \mathrm{Id}_A$.

(iv) The set $A \times A$ is a $\mathscr{C}$-prediagonal subset of itself. Indeed if T is a terminal $\mathscr{C}$-object and $f\colon A \to T$ is the unique $\mathscr{C}$-morphism from A to T, then we have $A \times A = \{(a_1, a_2) \in A \times A : f(a_1) = f(a_2)\}$.

(v) The set of $\mathscr{C}$-prediagonal subsets of $A \times A$ is closed under finite intersections. Indeed, if $(X_i)_{i \in I}$ is a finite family of $\mathscr{C}$-prediagonal subsets of $A \times A$, we can find $\mathscr{C}$-morphisms $f_i\colon A \to B_i$ such that $X_i = \{(a_1, a_2) \in A \times A : f_i(a_1) = f_i(a_2)\}$. Then if we set $B = \prod_{i \in I} B_i$ and $f = \prod_{i \in I} f_i\colon A \to B$, we have $\bigcap_{i \in I} X_i = \{(a_1, a_2) \in A \times A : f(a_1) = f(a_2)\}$.

As usual, we shall sometimes omit the letter "$\mathscr{C}$" in the words $\mathscr{C}$-square-algebraic, $\mathscr{C}$-prediagonal, $\mathscr{C}$-codiagonal and $\mathscr{C}$-subdiagonal when the ambient category is clear from the context.

Example 8.3

(i) Let $\mathscr{C} = \text{Set}$. Given a set A, the square-algebraic subsets of $A \times A$ are precisely the subsets of the form $E \times F$, where E and F are arbitrary subsets of A. A subset $X \subset A \times A$ is prediagonal if and only if it is the graph of an equivalence relation on A.

(ii) In the category Grp, given a group G, a subset of $G \times G$ is square-algebraic if and only if it is either empty or of the form $(g_1 N) \times (g_2 N)$, where N is a normal subgroup of G. The prediagonal subsets of $G \times G$ are precisely the subsets of the form $X_N = \{(g_1, g_2) \in G \times G : g_1 N = g_2 N\}$, where N is a normal subgroup of G. In other words, the prediagonal subsets are the graphs of the congruence relations modulo normal subgroups.

(iii) In the category Rng, given a ring R, the prediagonal subsets of $R \times R$ are precisely the subsets of the form $X_I = \{(r_1, r_2) \in R \times R : r_1 + I = r_2 + I\}$, where I is a two-sided ideal of R. In other words, the prediagonal subsets are the graphs of the congruence relations modulo two-sided ideals.

(iv) In the category R-Mod, given a R-module M, the prediagonal subsets of $M \times M$ are precisely the subsets of the form $X_N = \{(m_1, m_2) \in M \times M : m_1 + N = m_2 + N\}$, where N is a submodule of M. In other words, the prediagonal subsets of $M \times M$ are the graphs of congruence relations modulo submodules of M. Observe that every subdiagonal subset of $M \times M$ is a translate of a submodule of $M \times M$.

(v) In the full subcategory of Top whose objects are the Hausdorff topological spaces, given a Hausdorff topological space A, every square-algebraic (resp. prediagonal, resp. codiagonal) subset of $A \times A$ is closed (resp. closed, resp. open) in $A \times A$ for the product topology.

(vi) In CHT, given a compact Hausdorff topological space A, every subdiagonal subset of $A \times A$ is closed in $A \times A$.

(vii) In the category K-Aal of affine algebraic sets over a field K, every square-algebraic (resp. prediagonal, resp. codiagonal) subset of the square $A \times A$ of an object A is closed (resp. closed, resp. open) in the Zariski topology on $A \times A$ (beware that the Zariski topology on $A \times A$ is not, in general, the product of the Zariski topology on A with itself). If K is algebraically closed, it follows from Chevalley's theorem that every subdiagonal subset of $A \times A$ is constructible with respect to the Zariski topology on $A \times A$.

The following definition is introduced by Gromov in [22, Sect. 4.F].

Definition 8.4 Let $\mathscr{C}$ be a concrete category satisfying (CFP). One says that $\mathscr{C}$ has the *subdiagonal intersection property*, briefly (SDIP), provided that the following holds: for every $\mathscr{C}$-object A, any $\mathscr{C}$-square-algebraic subset $X \subset A \times A$, any $\mathscr{C}$-codiagonal subset $Y \subset A \times A$, and any non-increasing sequence $(Z_n)_{n \in \mathbb{N}}$ of $\mathscr{C}$-subdiagonal subsets of $A \times A$ such that $X \cap Y \cap Z_n \neq \emptyset$ for all $n \in \mathbb{N}$,

$$\bigcap_{n \in \mathbb{N}} X \cap Y \cap Z_n \neq \emptyset.$$

Example 8.5

(i) The category $\mathscr{C} = \text{Set}$ does not satisfy (SDIP). To see this, take for example $A = \mathbb{N}$ and consider the maps $f_n \colon A \to A$ defined by

$$f_n(k) = \begin{cases} k & \text{if } k \leqslant n-1, \\ 0 & \text{if } k \geqslant n \end{cases}$$

for all $n, k \in \mathbb{N}$. Then the sets

$$Z_n := \{(a_1, a_2) \in A \times A : f_n(a_1) = f_n(a_2)\}$$

form a non-increasing sequence of $\mathscr{C}$-prediagonal (and therefore $\mathscr{C}$-subdiagonal) subsets of $A \times A$. Take $X = A \times A$ and $Y = A \times A \setminus \Delta_A$. Then $X \subset A \times A$ is $\mathscr{C}$-square-algebraic and $Y \subset A \times A$ is $\mathscr{C}$-codiagonal in $A \times A$. We have $X \cap Y \cap Z_n \neq \emptyset$ for all $n \in \mathbb{N}$ since $(n, n+1) \in X \cap Y \cap Z_n$. On the other hand, we clearly have $\bigcap_{n \in \mathbb{N}} X \cap Y \cap Z_n = \emptyset$. This shows that Set does not satisfy (SDIP).

(ii) The category Set^f satisfies (SDIP) since any non-increasing sequence of finite sets eventually stabilizes.

(iii) Let $\mathscr{C} = \text{Rng}$, Grp, $\mathbb{Z}$-Mod, or $\mathbb{Z}\text{-Mod}^{f-g}$. Take $A = \mathbb{Z}$, $X = A \times A$, $Y = A \times A \setminus \Delta_A$, and, for $n \in \mathbb{N}$,

$$Z_n = \{(a_1, a_2) \in A \times A : a_1 \equiv a_2 \mod 2^n\}.$$

Clearly X is $\mathscr{C}$-square-algebraic, Y is $\mathscr{C}$-codiagonal, and $(Z_n)_{n \in \mathbb{N}}$ is a non-increasing sequence of $\mathscr{C}$-prediagonal (and hence $\mathscr{C}$-subdiagonal) subsets of $A \times A$. We have $X \cap Y \cap Z_n \neq \emptyset$ for all $n \in \mathbb{N}$ but $\bigcap_{n \in \mathbb{N}} X \cap Y \cap Z_n = \emptyset$. This shows that $\mathscr{C}$ does not satisfy (SDIP).

(iv) Given an arbitrary ring R, the category $R\text{-Mod}^{Art}$ satisfies (SDIP). Indeed we have seen that every subalgebraic subset is the translate of some submodule and that, in an Artinian module, every non-increasing sequence consisting of translates of submodules eventually stabilizes.

(v) Let R be a left-Artinian ring. Then the category R-Mod^{f-g} satisfies (SDIP) since it is a subcategory of R-Mod^{Art}.

(vi) Given a field K, the category K-Vec^{f-d} satisfies (SDIP) since K-Vec^{f-d} = K-Mod^{f-g} and every field is a left-Artinian ring.

(vii) The category CHT of compact Hausdorff topological spaces does not satisfy (SDIP). Indeed, let $A = [0, 1]$ denote the unit segment and consider, for every $n \in \mathbb{N}$ the continuous map $f_n \colon [0, 1] \to [0, 1]$ defined by

$$f_n(x) = \begin{cases} x & \text{if } x \leqslant \frac{n}{n+1}, \\ \frac{n}{n+1} & \text{if } \frac{n}{n+1} \leqslant x \end{cases}$$

for all $x \in A$. Take $X = A \times A$, $Y = A \times A \setminus \Delta_A$, and

$$Z_n = \{(a_1, a_2) \in A \times A : f_n(a_1) = f_n(a_2)\}.$$

Then X is square-algebraic, Y is codiagonal, and $(Z_n)_{n \in \mathbb{N}}$ is a non-increasing sequence of prediagonal (and therefore subdiagonal) subsets of $A \times A$. We clearly have $X \cap Y \cap Z_n \neq \emptyset$ for all $n \in \mathbb{N}$ but $\bigcap_{n \in \mathbb{N}} X \cap Y \cap Z_n = \emptyset$. This shows that CHT does not satisfy (SDIP).

(viii) A variant of the previous argument may be used to prove that even Man does not satisfy (SDIP). Indeed, consider the circle $\mathbb{S}^1 := \{z \in \mathbb{C} : |z| = 1\}$ and, for each $n \in \mathbb{N}$, the continuous map $f_n \colon \mathbb{S}^1 \to \mathbb{S}^1$ defined by

$$f_n(z) = \begin{cases} e^{i\frac{n+2}{n+1}\theta} & \text{if } z = e^{i\theta} \text{ with } 0 \leqslant \theta \leqslant \frac{n+1}{n+2}2\pi, \\ 1 & \text{otherwise.} \end{cases}$$

Take $X = A \times A$, $Y = A \times A \setminus \Delta_A$, and

$$Z_n = \{(a_1, a_2) \in A \times A : f_n(a_1) = f_n(a_2)\}.$$

Then X is square-algebraic, Y is codiagonal, and $(Z_n)_{n \in \mathbb{N}}$ is a non-increasing sequence of prediagonal (and therefore subdiagonal) subsets of $A \times A$ in the category Man. We have that $X \cap Y \cap Z_n \neq \emptyset$ for all $n \in \mathbb{N}$, since $(1, e^{i\frac{n+1}{n+2}\pi}) \in Z_n$. On the other hand, we clearly have $\bigcap_{n \in \mathbb{N}} X \cap Y \cap Z_n = \emptyset$. This shows that Man does not satisfy (SDIP).

(ix) Let K be an uncountable algebraically closed field. Then the category K-Aal satisfies (SDIP) [22, 4.F']. This follows from the fact that, when K is an uncountable algebraically closed field, every non-increasing sequence of nonempty constructible subsets of an affine algebraic set over K has a nonempty intersection [17, Proposition 4.4]. Indeed, if X (resp. Y, resp. Z_n) is a $\mathscr{C}$-square-algebraic (resp prediagonal, resp. subdiagonal) subset of $A \times A$, then, as observed above, X (resp. Y, resp. Z_n) is closed (resp. open, resp. constructible) in $A \times A$ and hence $X \cap Y \cap Z_n$ is constructible since any finite intersection of constructible subsets is itself constructible.

8.2 Reversibility of $\mathscr{C}$-Cellular Automata

We shall use the following auxiliary result.

Lemma 8.6 *Let $\mathscr{C}$ be a concrete category satisfying (CFP) and (SDIP). Let (X_n, f_{nm}) be a projective sequence of nonempty sets. Suppose that there is a projective sequence (A_n, F_{nm}), consisting of $\mathscr{C}$-objects A_n and $\mathscr{C}$-morphisms $F_{nm} \colon A_m \to A_n$ for all $n, m \in \mathbb{N}$ such that $m \geqslant n$, satisfying the following conditions*:

(PSD-1) *for all $n \in \mathbb{N}$, one has $X_n = Y_n \cap Z_n$, where $Y_n \subset A_n \times A_n$ is $\mathscr{C}$-codiagonal and $Z_n \subset A_n \times A_n$ is $\mathscr{C}$-prediagonal*;

(PSD-2) *for all $n, m \in \mathbb{N}$ with $m \geqslant n$, setting $S_{nm} = F_{nm} \times F_{nm}$, one has $S_{nm}(X_m) \subset X_n$ and f_{nm} is the restriction of S_{nm} to X_m*;

(PSD-3) *for all $n, m \in \mathbb{N}$ with $m \geqslant n$, one has $S_{nm}(Z_m) \subset Z_n$ and $f_{nm}(X_m) = Y_n \cap S_{nm}(Z_m)$.*

Then $\varprojlim X_n \neq \emptyset$.

Proof Let (X'_n, f'_{nm}) denote the universal projective sequence associated with the projective sequence (X_n, f_{nm}). As $S_{nm}(Z_m)$, $m = n, n+1, \ldots,$ is a non-increasing sequence of $\mathscr{C}$-subdiagonal subsets of $A_n \times A_n$ such that $Y_n \cap S_{nm}(Z_m) = f_{nm}(X_m) \neq \emptyset$ for all $m \geqslant n$, we deduce from (SDIP) that

$$X'_n = \bigcap_{m \geqslant n} f_{nm}(X_m) = \bigcap_{m \geqslant n} Y_n \cap S_{nm}(Z_m) \neq \emptyset. \tag{15}$$

Let now $m, n \in \mathbb{N}$ with $m \geqslant n$ and suppose that $x'_n \in X'_n$. Then we have $f_{nm}^{-1}(x'_n) \cap f_{mk}(X_k) \neq \emptyset$ for all $k \geqslant m$ by Remark 4.5. By applying again (SDIP), we get

$$\bigcap_{k \geqslant m} f_{nm}^{-1}\left(x'_n\right) \cap f_{mk}(X_k) = \bigcap_{k \geqslant m} F_{nm}^{-1}\left(x'_n\right) \cap Y_m \cap S_{mk}(Z_k) \neq \emptyset \tag{16}$$

(observe that $F_{nm}^{-1}(x'_n)$ is $\mathscr{C}$-square-algebraic). From (15) and (16), it follows that conditions (IP-1) and (IP-2) in Corollary 4.4 are satisfied, so that we conclude from this corollary that $\varprojlim X_n \neq \emptyset$. □

Theorem 8.7 *Let $\mathscr{C}$ be a concrete category satisfying (CFP+) and (SDIP), and let G be an arbitrary group. Then every bijective $\mathscr{C}$-cellular automaton $\tau \colon A^G \to B^G$ is reversible.*

Proof Let $\tau \colon A^G \to B^G$ be a bijective $\mathscr{C}$-cellular automaton. We have to show that the inverse map $\tau^{-1} \colon B^G \to A^G$ is a cellular automaton.

As in the proof of Theorem 6.1, we suppose first that the group G is countable. Let us show that the following local property is satisfied by τ^{-1}:

($*$) there exists a finite subset $N \subset G$ such that, for any $y \in B^G$, the element $\tau^{-1}(y)(1_G)$ depends only on the restriction of y to N.

This will show that τ is reversible. Indeed, if $(*)$ holds for some finite subset $N \subset G$, then there exists a (unique) map $\nu\colon B^N \to A$ such that

$$\tau^{-1}(y)(1_G) = \nu(y|_N)$$

for all $y \in B^G$. Now, the G-equivariance of τ implies the G-equivariance of its inverse map τ^{-1}. Consequently, we get

$$\tau^{-1}(y)(g) = g^{-1}\tau^{-1}(y)(1_G) = \tau^{-1}\big(g^{-1}y\big)(1_G) = \nu\big(\big(g^{-1}y\big)\big|_N\big)$$

for all $y \in B^G$ and $g \in G$. This implies that τ^{-1} is the cellular automaton with memory set N and local defining map ν.

Let us assume by contradiction that condition $(*)$ is not satisfied. Let M be a memory set for τ such that $1_G \in M$. Since G is countable, we can find a sequence $(E_n)_{n\in\mathbb{N}}$ of finite subsets of G such that $G = \bigcup_{n\in\mathbb{N}} E_n$, $M \subset E_0$, and $E_n \subset E_{n+1}$ for all $n \in \mathbb{N}$. Consider, for each $n \in \mathbb{N}$, the finite subset $F_n \subset G$ defined by $F_n = \{g \in G : gM \subset E_n\}$. Note that $G = \bigcup_{n\in\mathbb{N}} F_n$, $1_G \in F_0$, and $F_n \subset F_{n+1}$ for all $n \in \mathbb{N}$.

Since $(*)$ is not satisfied, we can find, for each $n \in \mathbb{N}$, two configurations $y'_n, y''_n \in B^G$ such that

$$y'_n\big|_{F_n} = y''_n\big|_{F_n} \quad \text{and} \quad \tau^{-1}\big(y'_n\big)(1_G) \neq \tau^{-1}\big(y''_n\big)(1_G). \tag{17}$$

Recall from the proof of Theorem 6.1, that τ induces, for each $n \in \mathbb{N}$, a $\mathscr{C}$-morphism $\tau_n\colon A^{E_n} \to B^{F_n}$ given by $\tau_n(u) = (\tau(x))|_{F_n}$ for every $u \in A^{E_n}$, where $x \in A^G$ is any configuration extending u.

Consider now, for each $n \in \mathbb{N}$, the subset $X_n \subset A^{E_n} \times A^{E_n}$ consisting of all pairs $(u'_n, u''_n) \in A^{E_n} \times A^{E_n}$ such that $\tau_n(u'_n) = \tau_n(u''_n)$ and $u'_n(1_G) \neq u''_n(1_G)$. We have $X_n = Y_n \cap Z_n$, where

$$Y_n := \big\{\big(u'_n, u''_n\big) \in A^{E_n} \times A^{E_n} : u'_n(1_G) \neq u''_n(1_G)\big\}$$

and

$$Z_n := \big\{\big(u'_n, u''_n\big) \in A^{E_n} \times A^{E_n} : \tau_n\big(u'_n\big) = \tau_n\big(u''_n\big)\big\}.$$

Note that Y_n (resp. Z_n) is a $\mathscr{C}$-codiagonal (resp. $\mathscr{C}$-prediagonal) subset of $A^{E_n} \times A^{E_n}$. Note also that X_n is not empty since

$$\big(\big(\tau^{-1}\big(y'_n\big)\big)\big|_{E_n}, \big(\tau^{-1}\big(y''_n\big)\big)\big|_{E_n}\big) \in X_n$$

by (17).

For $m \geqslant n$, the restriction map $\rho_{nm}\colon A^{E_m} \to A^{E_n}$ gives us a $\mathscr{C}$-square-morphism

$$S_{nm} = \rho_{nm} \times \rho_{nm}\colon A^{E_m} \times A^{E_m} \to A^{E_n} \times A^{E_n}$$

which induces by restriction a map $f_{nm}\colon X_m \to X_n$.

We clearly have $S_{nm}(Z_m) \subset Z_n$ and $S_{nm}(X_m) = Y_n \cap S_{nm}(Z_m)$ for all $n, m \in \mathbb{N}$ such that $m \geqslant n$. Since $\mathscr{C}$ satisfies (SDIP), it follows from Lemma 8.6, that $\varprojlim X_n \neq \emptyset$.

Choose an element $(p_n)_{n\in\mathbb{N}} \in \varprojlim X_n$. Thus $p_n = (u'_n, u''_n) \in A^{E_n} \times A^{E_n}$ and u'_{n+1} (resp. u''_{n+1}) coincides with u'_n (resp. u''_n) on E_n for all $n \in \mathbb{N}$. As $G = \bigcup_{n\in\mathbb{N}} E_n$, we deduce that there exists a (unique) configuration $x' \in A^G$ (resp. $x'' \in A^G$) such that $x'|_{E_n} = u'_n$ (resp. $x''|_{E_n} = u''_n$) for all $n \in \mathbb{N}$. Moreover, we have

$$\left(\tau\left(x'\right)\right)\big|_{F_n} = \tau_n\left(u'_n\right) = \tau_n\left(u''_n\right) = \left(\tau\left(x''\right)\right)\big|_{F_n}$$

for all $n \in \mathbb{N}$. As $G = \bigcup_{n\in\mathbb{N}} F_n$, this shows that $\tau(x') = \tau(x'')$. On the other hand, we have $x'(1_G) = u'_0(1_G) \neq u''_0(1_G) = x''(1_G)$ and hence $x' \neq x''$. This contradicts the injectivity of τ and therefore completes the proof that τ is reversible in the case when G is countable.

We now drop the countability assumption on G and prove the theorem in its full generality. Choose a memory set $M \subset G$ for τ and denote by H the subgroup of G generated by M. Observe that H is countable since M is finite. By Theorem 2.16, the restriction cellular automaton $\tau_H \colon A^H \to A^H$ is a bijective $\mathscr{C}$-cellular automaton. It then follows from the first part of the proof that τ_H is reversible. This implies that τ is reversible as well by part (iii) of Theorem 2.16. □

Corollary 8.8 *If G is an arbitrary group, all bijective $\mathscr{C}$-cellular automata are reversible when $\mathscr{C}$ is one of the following concrete categories*:

- Set^f, *the category of finite sets*;
- K-Vec^{f-d}, *the category of finite-dimensional vector spaces over an arbitrary field K*;
- R-Mod^{Art}, *the category of left Artinian modules over an arbitrary ring R*;
- R-Mod^{f-g}, *the category of finitely generated left modules over an arbitrary left-Artinian ring R*;
- K-Aal, *the category of affine algebraic sets over an arbitrary uncountable algebraically closed field K*.

In the following examples we show that there exist bijective cellular automata which are not reversible. They are modeled after [16] (see also [15, Example 1.10.3]).

Example 8.9

(i) Let p be a prime number and $A = \mathbb{Z}_p = \varprojlim \mathbb{Z}/p^n\mathbb{Z}$ the ring of p-adic integers. Recall that A is a compact Hausdorff topological ring for the topology associated with the p-adic metric. We can regard A as a $\mathscr{C}$-object for $\mathscr{C} = \mathrm{Set}, \mathrm{Grp}$, $\mathbb{Z}$-Mod, and CHT. Consider now the cellular automaton $\tau \colon A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ defined by

$$\tau(x)(n) = x(n) - px(n+1)$$

for all $x \in A^{\mathbb{Z}}$ and $n \in \mathbb{Z}$. It has memory set $M = \{0, 1\} \subset \mathbb{Z}$ and associated local defining map $\mu_M \colon A^M \to A$ given by $\mu_M(y) = y(0) - py(1)$ for all $y \in A^M$. It follows that τ is a $\mathscr{C}$-cellular automaton for $\mathscr{C}$ any of the concrete categories mentioned above. Note that τ is bijective. Indeed the inverse map $\tau^{-1}\colon A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ is given by

$$\tau^{-1}(x)(n) = \sum_{k=0}^{\infty} p^k x(n+k)$$

for all $x \in A^{\mathbb{Z}}$ and $n \in \mathbb{Z}$. However, τ^{-1} is not a cellular automaton. Indeed, let F be a finite subset of $\mathbb{Z}$ and choose an integer $m \geqslant 0$ such that $F \subset (-\infty, m]$. Consider the configurations $y, z \in A^{\mathbb{Z}}$ defined by $y(n) = 0$ if $n \leqslant m$ and $y(n) = 1$ if $n \geqslant m+1$, and $z(n) = 0$ for all $n \in \mathbb{N}$, respectively. Then y and z coincide on F. However, we have

$$\tau^{-1}(y)(0) = \sum_{k=0}^{\infty} p^k y(k) = \sum_{k=m+1}^{\infty} p^k$$

and

$$\tau^{-1}(z)(0) = \sum_{k=0}^{\infty} p^k z(k) = 0.$$

It follows that there is no finite subset $F \subset \mathbb{Z}$ such that $\tau^{-1}(x)(0)$ only depends on the restriction of $x \in A^{\mathbb{Z}}$ to F. Thus, there is no finite subset $F \subset \mathbb{Z}$ which may serve as a memory set for τ^{-1}.

(ii) Let R be a ring and let $A = R[[t]]$ denote the ring of all formal power series in one indeterminate t with coefficients in R. Note that A has a natural structure of a left R-module. Then the cellular automaton $\tau\colon A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ defined by

$$\tau(x)(n) = x(n) - tx(n+1)$$

for all $x \in A^{\mathbb{Z}}$ and $n \in \mathbb{Z}$ is a bijective $\mathscr{C}$-cellular automaton for $\mathscr{C} = R$-Mod. However, this cellular automaton is not reversible unless R is a zero ring. The proof is analogous to the one given in the preceding example (see [15, Example 1.10.3] in the case when R is a field).

Remark 8.10 In Examples 8.9, we can replace the group $\mathbb{Z}$ by any non-periodic group G. Indeed, if G is non-periodic and $H \subset G$ is an infinite cyclic subgroup (thus $H \cong \mathbb{Z}$), the induced cellular automaton $\tau^G\colon A^G \to A^G$ is a bijective $\mathscr{C}$-cellular automaton for $\mathscr{C} = \text{Set}, \text{Grp}, \mathbb{Z}\text{-Mod}$, and CHT (resp. R-Mod with R a nonzero ring) by Theorem 2.16 and Proposition 3.11, which is not reversible.

References

1. Adámek, J., Herrlich, H., Strecker, G.E.: Abstract and concrete categories: The joy of cats. Reprint of the 1990 original. Repr. Theory Appl. Categ. **17** (2006)
2. Armendariz, E.P., Fisher, J.W., Snider, R.L.: On injective and surjective endomorphisms of finitely generated modules. Commun. Algebra **6**, 659–672 (1978)

3. Ax, J.: The elementary theory of finite fields. Ann. Math. (2) **88**, 239–271 (1968)
4. Ax, J.: Injective endomorphisms of varieties and schemes. Pac. J. Math. **31**, 1–7 (1969)
5. Bartholdi, L.: Gardens of Eden and amenability on cellular automata. J. Eur. Math. Soc. **12**, 241–248 (2010)
6. Borel, A.: Injective endomorphisms of algebraic varieties. Arch. Math. (Basel) **20**, 531–537 (1969)
7. Borel, A.: Linear Algebraic Groups, 2nd edn. Graduate Texts in Mathematics, vol. 126. Springer, New York (1991)
8. Bourbaki, N.: Éléments de Mathématique. Topologie Générale. Chapitres 1 à 4. Hermann, Paris (1971)
9. Bump, D.: Algebraic Geometry. World Scientific, River Edge (1998)
10. Ceccherini-Silberstein, T., Coornaert, M.: The Garden of Eden theorem for linear cellular automata. Ergod. Theory Dyn. Syst. **26**, 53–68 (2006)
11. Ceccherini-Silberstein, T., Coornaert, M.: Injective linear cellular automata and sofic groups. Isr. J. Math. **161**, 1–15 (2007)
12. Ceccherini-Silberstein, T., Coornaert, M.: Linear cellular automata over modules of finite length and stable finiteness of group rings. J. Algebra **317**, 743–758 (2007)
13. Ceccherini-Silberstein, T., Coornaert, M.: On the surjunctivity of Artinian linear cellular automata over residually finite groups. In: Geometric Group Theory. Trends in Math., pp. 37–44. Birkhäuser, Basel (2007)
14. Ceccherini-Silberstein, T., Coornaert, M.: Induction and restriction of cellular automata. Ergod. Theory Dyn. Syst. **29**, 371–380 (2009)
15. Ceccherini-Silberstein, T., Coornaert, M.: Cellular Automata and Groups. Springer Monographs in Math. Springer, Berlin (2010)
16. Ceccherini-Silberstein, T., Coornaert, M.: On the reversibility and the closed image property of linear cellular automata. Theor. Comput. Sci. **412**, 300–306 (2011)
17. Ceccherini-Silberstein, T., Coornaert, M.: On algebraic cellular automata. J. Lond. Math. Soc. (2) **84**, 541–558 (2011)
18. Ceccherini-Silberstein, T., Machì, A., Scarabotti, F.: Amenable groups and cellular automata. Ann. Inst. Fourier (Grenoble) **49**, 673–685 (1999)
19. Elek, G., Szabó, E.: Sofic groups and direct finiteness. J. Algebra **280**, 426–434 (2004)
20. Freyd, P.: Homotopy is not concrete. In: The Steenrod Algebra and its Applications, Proc. Conf. to Celebrate N.E. Steenrod's Sixtieth Birthday, Battelle Memorial Inst., Columbus, Ohio, 1970. Lect. Notes in Mathematics, vol. 168, pp. 25–34. Springer, Berlin (1970)
21. Gottschalk, W.: Some general dynamical notions. In: Recent Advances in Topological Dynamics, Proc. Conf. Topological Dynamics, Yale Univ., New Haven, Conn., 1972; in honor of Gustav Arnold Hedlund. Lect. Notes in Mathematics, vol. 318, pp. 120–125. Springer, Berlin (1973)
22. Gromov, M.: Endomorphisms of symbolic algebraic varieties. J. Eur. Math. Soc. **1**, 109–197 (1999)
23. Grothendieck, A.: Éléments de géométrie algébrique. III. Étude cohomologique des faisceaux cohérents. I. Inst. Hautes Études Sci. Publ. Math., 167 (1961)
24. Hedlund, G.A.: Endomorphisms and automorphisms of the shift dynamical system. Math. Syst. Theory **3**, 320–375 (1969)
25. Hungerford, T.W.: Algebra. Graduate Texts in Mathematics, vol. 73. Springer, New York (1980)
26. James, I.M.: Introduction to Uniform Spaces. London Mathematical Society Lecture Note Series, vol. 144. Cambridge University Press, Cambridge (1990)
27. Kaplansky, I.: Fields and Rings. Chicago Lectures in Mathematics. University of Chicago Press, Chicago (1995)
28. Kraft, H.: Basics from algebraic geometry. In: Polynomial Automorphisms and Related Topics, pp. 43–112. Publishing House for Science and Technology, Hanoi (2007)
29. Linderholm, C.E.: A group epimorphism is surjective. Am. Math. Mon. **77**, 176–177 (1970)

30. Machì, A., Mignosi, F.: Garden of Eden configurations for cellular automata on Cayley graphs of groups. SIAM J. Discrete Math. **6**, 44–56 (1993)
31. Mac Lane, S.: Categories for the Working Mathematician. Graduate Texts in Mathematics, vol. 5. Springer, New York (1998)
32. Milne, J.S.: Algebraic Geometry, v5.20 (2009). www.jmilne.org/math/
33. Moore, E.F.: Machine models of self-reproduction. In: Proc. Symp. Appl. Math., vol. 14, pp. 17–34. Am. Math. Soc., Providence (1963)
34. Myhill, J.: The converse of Moore's Garden-of-Eden theorem. Proc. Am. Math. Soc. **14**, 685–686 (1963)
35. Reid, M.: Undergraduate Algebraic Geometry. London Mathematical Society Student Texts, vol. 12. Cambridge University Press, Cambridge (1988)
36. Vasconcelos, W.V.: Injective endomorphisms of finitely generated modules. Proc. Am. Math. Soc. **25**, 900–901 (1970)
37. Weil, A.: Sur les espaces à structure uniforme et sur la topologie générale. Actual. Sci. Ind., vol. 551. Hermann, Paris (1937)
38. Weiss, B.: Sofic groups and dynamical systems. In: Ergodic Theory and Harmonic Analysis, Mumbai, 1999. Sankhyā Ser. A, vol. 62, pp. 350–359 (2000)

Pointwise Convergence of Bochner–Riesz Means in Sobolev Spaces

Leonardo Colzani and Sara Volpi

Abstract The Bochner–Riesz means are defined by the Fourier multiplier operators $(S_R^\alpha * f)\hat{}(\xi) = (1 - |R^{-1}\xi|^2)_+^\alpha \hat{f}(\xi)$. Here we prove that if f has β derivatives in $L^p(\mathbf{R}^d)$, then $S_R^\alpha * f(x)$ converges pointwise to $f(x)$ as $R \to +\infty$ with a possible exception of a set of points with Hausdorff dimension at most $d - \beta p$ if one of the following conditions holds: either $\alpha > (d-1)|1/p - 1/2|$, or $\alpha > d(1/2 - 1/p) - 1/2$ and $\alpha + \beta \geqslant (d-1)/2$. If $\beta > d/p$, then pointwise convergence holds everywhere.

Keywords Bochner–Riesz means · Sobolev space · Hausdorff dimension

Mathematics Subject Classification (2010) 42B08

1 Introduction

The Bochner–Riesz means of order α of functions in $\mathbf{R}^d$ are defined by the Fourier integrals

$$S_R^\alpha * f(x) = \int_{\{|\xi|<R\}} \left(1 - \left|R^{-1}\xi\right|^2\right)^\alpha \hat{f}(\xi) \exp(2\pi i \xi x)\, d\xi.$$

In particular, when $\alpha = 0$ one obtains the spherical partial sums, which are a natural analogue of the partial sums of one-dimensional Fourier series. The almost everywhere convergence of Bochner–Riesz means has been widely studied, however there are still many open problems. See [13] as a general reference. If $\alpha > (d-1)/2$, the critical index, then the Bochner–Riesz maximal operator $S_*^\alpha f = \sup_{R>0} |S_R^\alpha * f|$ is pointwise dominated by the Hardy-Littlewood maximal operator, hence it is of

L. Colzani · S. Volpi (✉)
Dipartimento di Matematica e Applicazioni, Università di Milano Bicocca, Via R. Cozzi 53, 20125 Milan, Italy
e-mail: sara_volpi@yahoo.it

L. Colzani
e-mail: leonardo.colzani@unimib.it

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3, DOI 10.1007/978-88-470-2853-1_7,

weak type $(1,1)$ and of strong type (∞,∞). If $p=2$ and $\alpha>0$, then S_*^α is bounded on $L^2(\mathbf{R}^d)$. Therefore, by complex interpolation, the Bochner–Riesz maximal operator is bounded on $L^p(\mathbf{R}^d)$, with $1<p\leqslant+\infty$ and $\alpha>(d-1)|1/p-1/2|$. From this the almost everywhere convergence follows: if f is in $L^p(\mathbf{R}^d)$, with $1\leqslant p\leqslant+\infty$ and $\alpha>(d-1)|1/p-1/2|$, then $\lim_{R\to+\infty} S_R^\alpha * f(x)=f(x)$ a.e.. This result is not optimal. Indeed, Carbery in [1] established the pointwise convergence of the Bochner–Riesz means when $2\leqslant p<2d/(d-1-2\alpha)$ and $d=2$. The same result for $d\geqslant 3$ has been obtained by Christ in [5], under the extra assumption that $\alpha\geqslant(d-1)/2(d+1)$. For further improvements when $p\geqslant 2$ see also [7]. Finally, in [2] Carbery, Rubio de Francia, and Vega have removed the restriction on α by showing that $S_*^\alpha f$ is bounded on the weighted space $L^2(|x|^{-\lambda}\,dx)$ if $d(1-2/p)\leqslant\lambda<1+2\alpha\leqslant d$ and observing that $L^p\subset L^2+L^2(|x|^{-\lambda}\,dx)$. See also [10]. Moreover, it has been shown by Rubio de Francia that the Bochner–Riesz means of index α are not defined in $L^p(\mathbf{R}^d)$ when $p\geqslant 2d/(d-1-2\alpha)$. Therefore the problem of the almost everywhere convergence of Bochner–Riesz means when $p\geqslant 2$ is essentially solved. On the other hand, as far as we know, sharp results when $p<2$ are not known. For related subjects, see [14].

Here we consider the pointwise convergence of Bochner–Riesz means for more regular functions, in particular functions in Sobolev classes. It has been proved by Carbery and Soria in [3] and Ma in [8] that by putting some smoothness on the function one may decrease the index of almost everywhere summability. Moreover, Carbery and Soria in [4] and Montini in [9] have considered the problem of the capacity and Hausdorff dimension of the divergence set of spherical partial sums of Fourier integrals. Then, in [6] Colzani has shown that the Bochner–Riesz means of functions with β integrable derivatives with $\alpha+\beta>(d-1)/2$ may diverge only in sets of points of Hausdorff dimension at most $d-\beta$. Here we generalise these results to functions with β fractional derivatives in $L^p(\mathbf{R}^d)$, $1<p<\infty$. In particular, we obtain conditions on α, β, p and d that ensure the pointwise convergence up to sets with Hausdorff dimension at most $d-\beta p$. The conditions are the following: either $\alpha>(d-1)|1/p-1/2|$ or $\alpha>d(1/2-1/p)-1/2$ and $\alpha+\beta\geqslant(d-1)/2$. Since the functions we are considering may be infinite precisely on sets of dimension $d-\beta p$, this estimate for the dimension of the divergence sets is the best possible. However, our analysis is not exhaustive and the ranges of the indexes are not optimal.

Before stating our result, we recall some basic definitions.

The *Bochner–Riesz kernel* S_R^α of order α, with $\alpha>0$, is defined by its Fourier transform,

$$\widehat{S_R^\alpha}(\xi)=\left(1-\left|R^{-1}\xi\right|^2\right)_+^\alpha.$$

This kernel can be written explicitly in terms of Bessel functions:

$$S_R^\alpha(x)=\pi^{-\alpha}\Gamma(\alpha+1)R^{d/2-\alpha}|x|^{-\alpha-d/2}J_{\alpha+d/2}\big(2\pi R|x|\big).$$

It follows from the asymptotic formula for Bessel functions that

$$\left|S_R^\alpha(x)\right|\leqslant cR^d\big(1+R|x|\big)^{-\alpha-(d+1)/2}.$$

The *Bessel kernel* G^β, with $\beta > 0$, is defined by its Fourier transform,

$$\widehat{G^\beta}(\xi) = \left(1 + |\xi|^2\right)^{-\beta/2}.$$

Also this kernel can be written explicitly in terms of Bessel functions, however it is more convenient to see it as a superposition of heat kernels,

$$G^\beta(x) = \Gamma(\beta/2)^{-1} \int_0^{+\infty} (4\pi t)^{-d/2} e^{-|x|^2/4t} e^{-t} t^{\beta/2-1}\, dt.$$

It follows from this representation that this kernel is positive and integrable. Moreover, if $0 < \beta < d$ then it is asymptotic to $c|x|^{\beta-d}$ when $x \to 0$ and it has an exponential decay at infinity. If $\beta = d$ then G^β has a logarithmic singularity at the origin, and if $\beta > d$ then it is bounded. See [11]. Finally, the *Riesz kernel* I^β, with $0 < \beta < d$, is given by

$$I^\beta(x) = |x|^{\beta-d}.$$

If $\beta > 0$ and $p > 1$, the *Bessel capacity* of a set $E \subset \mathbf{R}^d$ is defined by

$$B_{\beta,p}(E) = \inf\left\{\|f\|_p^p : G^\beta * f(x) \geqslant 1 \text{ on } E\right\}.$$

The *Riesz capacity* $R_{\beta,p}$ is defined in a similar way, by replacing G^β with I^β. It follows from the definitions that $R_{\beta,p}(E) \leqslant C B_{\beta,p}(E)$. Actually, it is also true that the Bessel and Riesz capacities have the same null sets. See [16, p. 67]. It can be also proved that when the $d - \beta p$ Hausdorff measure of E is finite, then $B_{\beta,p}(E) = 0$. Conversely, if $B_{\beta,p}(E) = 0$, then for every $\varepsilon > 0$ the $d - \beta p + \varepsilon$ Hausdorff measure of E is 0. See [16, Th 2.6.16].

Our main result is the following.

Theorem 1.1 *Assume that* $\alpha > 0$, $\beta > 0$, *and* $f = G^\beta * F$ *with* $F \in L^p(\mathbf{R}^d)$, $1 \leqslant p \leqslant +\infty$. *If* $0 < \beta \leqslant d/p$, *then* $S_R^\alpha * f(x)$ *converges pointwise to* $f(x)$ *as* $R \to +\infty$ *with a possible exception of a set of points with Hausdorff dimension at most* $d - \beta p$, *provided that one of the following conditions holds*:

(i) *either* $\alpha > (d-1)|1/p - 1/2|$;
(ii) *or* $\alpha > d(1/2 - 1/p) - 1/2$ *and* $\alpha + \beta \geqslant (d-1)/2$.

If $\alpha > d(1/2 - 1/p) - 1/2$ *and* $\beta > d/p$, *then the convergence is pointwise everywhere and uniform.*

As mentioned in the Introduction, almost everywhere convergence of Bochner–Riesz means has been considered in [3] and [8], and the case $\alpha = 0$ of our Theorem is contained in [4] and [9]. The cases $p = 1$ and $p = 2$ are already contained in [6], and also the case $p = \infty$ is already known. Indeed, if $p = \infty$ and $\beta > 0$, then $f = G^\beta * F$ is bounded and uniformly continuous, and if $\alpha > (d-1)/2$ then $S_R^\alpha * f$ converges to f uniformly. The assumption $\alpha > d(1/2 - 1/p) - 1/2$ in Theorem 1.1

is necessary in order to define the Bochner–Riesz means in $L^p(\mathbf{R}^d)$. Observe also that when $p < 2d/(d-1)$ this condition reduces to $\alpha + \beta \geqslant (d-1)/2$.

The main point of the proof of the Theorem is an estimate for the maximal Bochner–Riesz operator: for every f with β derivatives in $L^p(\mathbf{R}^d)$ there exists a function H in $L^p(\mathbf{R}^d)$ such that

$$\sup_{R>0} \left| S_R^\alpha * f(x) \right| \leqslant H * |x|^{\beta - d}.$$

Convergence up to a set of Riesz capacity zero follows from this.

2 Proof of Theorem 1.1

We split the proof of Theorem 1.1 into a series of lemmas.

Lemma 2.1 *If* $1 \leqslant p \leqslant +\infty$ *and* $\alpha > d(1/2 - 1/p) - 1/2$, *then for every* $F \in L^p(\mathbf{R}^d)$ *the convolution* $S_R^\alpha * G^\beta * F$ *is well defined and it is commutative and associative*:

$$S_R^\alpha * \left(G^\beta * F\right) = \left(S_R^\alpha * G^\beta\right) * F = G^\beta * \left(S_R^\alpha * F\right).$$

Proof The statement follows from Young's inequality: given $1 \leqslant p, q, r \leqslant \infty$ with $1/p + 1/q = 1 + 1/r$, if $f \in L^p(\mathbf{R}^d)$ and $g \in L^q(\mathbf{R}^d)$, then $f * g \in L^r(\mathbf{R}^d)$. It suffices to observe that the Bessel kernel G^β is an integrable function and that the Bochner–Riesz kernel S_R^α is in $L^q(\mathbf{R}^d)$, with $q > 2d/(2\alpha + d + 1)$. □

Lemma 2.2 *If* $\alpha > (d-1)/2$, *then the maximal operator* $S_*^\alpha f = \sup_{R>0} |S_R^\alpha * f|$ *is of weak type in* $L^1(\mathbf{R}^d)$. *If* $1 < p \leqslant +\infty$ *and* $\alpha > (d-1)|1/p - 1/2|$, *then this maximal operator is bounded on* $L^p(\mathbf{R}^d)$.

Proof This is a classical result of Stein. First one proves the extreme cases $p = 1$ or $p = \infty$ and $\Re(\alpha) > (d-1)/2$, then the case $p = 2$ and $\Re(\alpha) > 0$. The other cases follow by complex interpolation. See [13, Theorem 5.1]. □

Lemma 2.3 *If* $\alpha + \beta \geqslant (d-1)/2$, *then*

$$\sup_{R>0} \left| S_R^\alpha * G^\beta(x) \right| \leqslant C \begin{cases} \min\{|x|^{\beta-d}, |x|^{-\alpha-(d+1)/2}\} & \textit{if } 0 < \beta < d, \\ \min\{\log(1 + 1/|x|), |x|^{-\alpha-(d+1)/2}\} & \textit{if } \beta = d, \\ \min\{1, |x|^{-\alpha-(d+1)/2}\} & \textit{if } \beta > d. \end{cases}$$

Proof First assume that $0 < \beta < d$. By definition

$$S_R^\alpha * G^\beta(x) = \int_{\mathbf{R}^d} \left(1 + |\xi|^2\right)^{-\beta/2} \left(1 - \left|R^{-1}\xi\right|^2\right)_+^\alpha e^{2\pi i \xi \cdot x}\, d\xi.$$

Let $\phi + \psi = 1$ be a partition of unity on $\mathbf{R}_+$, with ϕ and ψ smooth and non-negative, and

$$\phi(\rho) = 1 \quad \text{if } 0 \leqslant \rho \leqslant 1/3 \quad \text{and} \quad \phi(\rho) = 0 \quad \text{if } \rho > 2/3,$$
$$\psi(\rho) = 1 \quad \text{if } 2/3 \leqslant \rho \leqslant 1 \quad \text{and} \quad \psi(\rho) = 0 \quad \text{if } \rho < 1/3.$$

Then

$$S_R^\alpha * G^\beta(x) = \int_{\mathbf{R}^d} \phi\left(\left|R^{-1}\xi\right|\right)\left(1+|\xi|^2\right)^{-\beta/2}\left(1-\left|R^{-1}\xi\right|^2\right)_+^\alpha e^{2\pi i\xi\cdot x}\,d\xi$$
$$+ \int_{\mathbf{R}^d} \psi\left(\left|R^{-1}\xi\right|\right)\left(1+|\xi|^2\right)^{-\beta/2}\left(1-\left|R^{-1}\xi\right|^2\right)_+^\alpha e^{2\pi i\xi\cdot x}\,d\xi.$$

To estimate the first integral set

$$\widehat{K}(\xi) = \phi\left(|\xi|\right)\left(1-|\xi|^2\right)_+^\alpha.$$

Then, if $K_R(x) = R^d K(Rx)$ we can rewrite the first integral as

$$\int_{\mathbf{R}^d} \phi\left(\left|R^{-1}\xi\right|\right)\left(1+|\xi|^2\right)^{-\beta/2}\left(1-\left|R^{-1}\xi\right|^2\right)_+^\alpha e^{2\pi i\xi\cdot x}\,d\xi = K_R * G^\beta(x).$$

Recall that the Hardy-Littlewood maximal function of a locally integrable function is defined by

$$\mathscr{M} f(x) = \sup_{r>0} \frac{1}{|B_r|} \int_{B_r} \left|f(x-y)\right| dy,$$

where B_r denotes the ball of radius r centred at the origin and $|B_r|$ is its volume. Since the multiplier $\widehat{K}$ is smooth with compact support, the kernel $K(x)$ is bounded and rapidly decreasing at infinity. This implies that

$$\sup_{R>0} \left|K_R * G^\beta(x)\right| \leqslant C\mathscr{M} G^\beta(x).$$

Since $G^\beta(x) \leqslant C|x|^{\beta-d}$ and since the Hardy-Littlewood maximal function of a radial homogeneous function is radial homogeneous, it also follows that

$$\mathscr{M} G^\beta(x) \leqslant C\mathscr{M}|x|^{\beta-d} = C|x|^{\beta-d}.$$

We now estimate the second integral. When $|Rx| \leqslant 3$ a crude estimate gives

$$\left|\int_{\mathbf{R}^d} \psi\left(\left|R^{-1}\xi\right|\right)\left(1+|\xi|^2\right)^{-\beta/2}\left(1-\left|R^{-1}\xi\right|^2\right)_+^\alpha e^{2\pi i\xi\cdot x}\,d\xi\right|$$
$$\leqslant \int_{\mathbf{R}^d} \psi\left(\left|R^{-1}\xi\right|\right)\left(1+|\xi|^2\right)^{-\beta/2}\left(1-\left|R^{-1}\xi\right|^2\right)_+^\alpha\,d\xi$$
$$= CR^{d-\beta} \int_0^1 \psi(\rho)\left(R^{-2}+\rho^2\right)^{-\beta/2}\left(1-\rho^2\right)^\alpha \rho^{d-1}\,d\rho$$
$$\leqslant CR^{d-\beta}$$
$$\leqslant C|x|^{\beta-d}.$$

To estimate the integral when $|Rx| > 3$, we introduce another smooth cut-off function $0 \leqslant \chi \leqslant 1$ such that $\chi(\rho) = 1$ if $0 \leqslant \rho \leqslant 1 - 2/|Rx|$ and $\chi(\rho) = 0$ if $\rho \geqslant 1 - 1/|Rx|$. Moreover we require that for all $j = 0, 1, 2, \ldots,$

$$\left| \frac{d^j}{d\rho^j} \chi(\rho) \right| \leqslant C(j)|Rx|^j.$$

Then

$$\begin{aligned}
&\int_{\mathbf{R}^d} \psi\left(\left|R^{-1}\xi\right|\right)\left(1+|\xi|^2\right)^{-\beta/2}\left(1-\left|R^{-1}\xi\right|^2\right)_+^{\alpha} e^{2\pi i \xi\cdot x}\, d\xi \\
&\quad = \int_{\mathbf{R}^d} \left(1-\chi\left(\left|R^{-1}\xi\right|\right)\right)\psi\left(\left|R^{-1}\xi\right|\right)\left(1+|\xi|^2\right)^{-\beta/2}\left(1-\left|R^{-1}\xi\right|^2\right)_+^{\alpha} e^{2\pi i \xi\cdot x}\, d\xi \\
&\qquad + \int_{\mathbf{R}^d} \chi\left(\left|R^{-1}\xi\right|\right)\psi\left(\left|R^{-1}\xi\right|\right)\left(1+|\xi|^2\right)^{-\beta/2}\left(1-\left|R^{-1}\xi\right|^2\right)_+^{\alpha} e^{2\pi i \xi\cdot x}\, d\xi.
\end{aligned}$$

In polar coordinates the first integral becomes

$$\begin{aligned}
&\int_{\mathbf{R}^d} \left(1-\chi\left(\left|R^{-1}\xi\right|\right)\right)\psi\left(\left|R^{-1}\xi\right|\right)\left(1+|\xi|^2\right)^{-\beta/2}\left(1-\left|R^{-1}\xi\right|^2\right)_+^{\alpha} e^{2\pi i \xi\cdot x}\, d\xi \\
&\quad = R^{d-\beta}\int_0^1 \rho^{d-1}\left(1-\chi(\rho)\right)\psi(\rho)\left(R^{-2}+\rho^2\right)^{-\beta/2}\left(1-\rho^2\right)^{\alpha}\int_{|\theta|=1} e^{2\pi i R\rho x\cdot\theta}\, d\theta\, d\rho.
\end{aligned}$$

It is well known that

$$\left| \int_{|\theta|=1} e^{2\pi i x\cdot\theta}\, d\theta \right| \leqslant C|x|^{-(d-1)/2}.$$

This is a standard estimate for oscillatory integrals with non-degenerate critical points, which also follows from the decay of Bessel functions and the explicit formula

$$\int_{|\theta|=1} e^{2\pi i x\cdot\theta}\, d\theta = 2\pi |x|^{(2-d)/2} J_{(d-2)/2}\left(2\pi|x|\right).$$

See for example [12, p. 347]. Therefore, if $\alpha + \beta \geqslant (d-1)/2$ and $|Rx| > 3$,

$$\begin{aligned}
&\left| \int_{\mathbf{R}^d} \left(1-\chi\left(\left|R^{-1}\xi\right|\right)\right)\psi\left(\left|R^{-1}\xi\right|\right)\left(1+|\xi|^2\right)^{-\beta/2}\left(1-\left|R^{-1}\xi\right|^2\right)_+^{\alpha} e^{2\pi i \xi\cdot x}\, d\xi \right| \\
&\quad \leqslant C R^{(d+1)/2-\beta}|x|^{-(d-1)/2}\int_0^1 \left(1-\chi(\rho)\right)(1-\rho)^{\alpha}\, d\rho \\
&\quad \leqslant C R^{(d+1)/2-\beta}|x|^{-(d-1)/2}|Rx|^{-\alpha-1} \\
&\quad = C|Rx|^{(d-1)/2-\alpha-\beta}|x|^{\beta-d} \\
&\quad \leqslant C|x|^{\beta-d}.
\end{aligned}$$

To estimate the second integral, we introduce the Laplacian $\Delta_\xi = -\sum_{j=1}^{d} \partial^2/\partial\xi_j^2$. Since this operator is self-adjoint and $\Delta_\xi^k(e^{2\pi i\xi\cdot x}) = |2\pi x|^{2k} e^{2\pi i\xi\cdot x}$, we obtain

$$\begin{aligned}
&\int_{\mathbf{R}^d} \chi\big(\big|R^{-1}\xi\big|\big)\psi\big(\big|R^{-1}\xi\big|\big)\big(1+|\xi|^2\big)^{-\beta/2}\big(1-\big|R^{-1}\xi\big|^2\big)_+^{\alpha} e^{2\pi i\xi\cdot x}\,d\xi\\
&\quad= \int_{\mathbf{R}^d} \chi\big(\big|R^{-1}\xi\big|\big)\psi\big(\big|R^{-1}\xi\big|\big)\big(1+|\xi|^2\big)^{-\beta/2}\big(1-\big|R^{-1}\xi\big|^2\big)_+^{\alpha} \Delta_\xi^k\left(\frac{e^{2\pi i\xi\cdot x}}{|2\pi x|^{2k}}\right)d\xi\\
&\quad= |2\pi x|^{-2k} R^{-\beta} \int_{\mathbf{R}^d} \Delta_\xi^k\big(g\big(R^{-1}\xi\big)\big)e^{2\pi i\xi\cdot x}\,d\xi,
\end{aligned}$$

where we have set

$$g(\xi) = \chi\big(|\xi|\big)\psi\big(|\xi|\big)\big(R^{-2}+|\xi|^2\big)^{-\beta/2}\big(1-|\xi|^2\big)_+^{\alpha}.$$

Now, denoting by Δ_ρ^k the radial part of the Laplacian and setting $g_0(|\xi|) = g(\xi)$, we get

$$\begin{aligned}
&\int_{\mathbf{R}^d} \chi\big(\big|R^{-1}\xi\big|\big)\psi\big(\big|R^{-1}\xi\big|\big)\big(1+|\xi|^2\big)^{-\beta/2}\big(1-\big|R^{-1}\xi\big|^2\big)_+^{\alpha} e^{2\pi i\xi\cdot x}\,d\xi\\
&\quad= |2\pi x|^{-2k} R^{-\beta-2k} \int_{\mathbf{R}^d} \big(\Delta_\xi^k g\big)\big(R^{-1}\xi\big)e^{2\pi i\xi\cdot x}\,d\xi\\
&\quad= |2\pi x|^{-2k} R^{d-\beta-2k} \int_0^{+\infty} \Delta_\rho^k g_0(\rho)\rho^{d-1} \int_{|\theta|=1} e^{2\pi i R\rho x\cdot\theta}\,d\theta\,d\rho.
\end{aligned}$$

Then, recalling the properties of the cut-off functions χ and ψ, if $k > (\alpha-1)/2$ we finally get

$$\begin{aligned}
&\left|\int_{\mathbf{R}^d} \chi\big(\big|R^{-1}\xi\big|\big)\psi\big(\big|R^{-1}\xi\big|\big)\big(1+|\xi|^2\big)^{-\beta/2}\big(1-\big|R^{-1}\xi\big|^2\big)_+^{\alpha} e^{2\pi i\xi\cdot x}\,d\xi\right|\\
&\quad\leqslant C|x|^{-2k-(d-1)/2} R^{(d+1)/2-\beta-2k} \int_0^{+\infty} \big|\Delta_\rho^k g_0(\rho)\big|\rho^{(d-1)/2}\,d\rho\\
&\quad\leqslant C|x|^{-2k-(d-1)/2} R^{(d+1)/2-\beta-2k} \int_{1/3}^{1-1/R|x|} (1-\rho)^{\alpha-2k}\,d\rho\\
&\quad\leqslant C|x|^{-2k-(d-1)/2} R^{(d+1)/2-\beta-2k} |Rx|^{-\alpha+2k-1}\\
&\quad= C|Rx|^{(d-1)/2-\alpha-\beta}|x|^{\beta-d}\\
&\quad\leqslant C|x|^{\beta-d}.
\end{aligned}$$

We have proved that, if $0<\beta<d$, then for every x

$$\sup_{R>0}\big|S_R^{\alpha} * G^{\beta}(x)\big| \leqslant C|x|^{\beta-d}.$$

This estimate is the best possible if $\alpha+\beta=(d-1)/2$, or if $\alpha+\beta>(d-1)/2$ and $|x|\leqslant 1$. If $(d-1)/2-\beta<\alpha<(d-1)/2$ and $|x|\geqslant 1$, then

$$\begin{aligned}\sup_{R>0}\left|S_R^\alpha * G^\beta(x)\right| &\leqslant G^{\alpha+\beta-(d-1)/2} * \sup_{R>0}\left|S_R^\alpha * G^{(d-1)/2-\alpha}\right|(x)\\ &\leqslant C G^{\alpha+\beta-(d-1)/2} * |x|^{-\alpha-(d+1)/2}\\ &\leqslant C|x|^{-\alpha-(d+1)/2}.\end{aligned}$$

The first inequality follows from the fact that Bessel kernels are positive, and the last inequality follows from the fact that these kernels are integrable with an exponential decay at infinity. If $\alpha\geqslant(d-1)/2$, then

$$\left|S_R^\alpha(x)\right|\leqslant C R^d\big(1+|Rx|\big)^{-\alpha-(d+1)/2}.$$

Again by this estimate it follows that, if $|x|\geqslant 1$,

$$\sup_{R>0}\left|S_R^\alpha * G^\beta(x)\right|\leqslant C|x|^{-\alpha-(d+1)/2}.$$

Hence we have proved that, if $0<\beta<d$,

$$\sup_{R>0}\left|S_R^\alpha * G^\beta(x)\right|\leqslant C\min\big\{|x|^{\beta-d}, |x|^{-\alpha-(d+1)/2}\big\}.$$

The proof of the cases $\beta\geqslant d$ is similar. □

Lemma 2.4 *Let $1<p<+\infty$, $\alpha>0$, $0<\beta<d$, and assume that one of the following properties holds*:

(i) *either $\alpha>(d-1)|1/p-1/2|$;*
(ii) *or $\alpha>d(1/2-1/p)-1/2$ and $\alpha+\beta\geqslant(d-1)/2$.*

Then for every function $F\in L^p(\mathbf{R}^d)$ there exists a function $H\in L^p(\mathbf{R}^d)$ with $\|H\|_p\leqslant C\|F\|_p$ and such that

$$S_*^\alpha\big(G^\beta * F\big)(x)=\sup_{R>0}\left|S_R^\alpha * G^\beta * F(x)\right|\leqslant I^\beta * H(x),$$

where $I^\beta(x)=|x|^{\beta-d}$ is the Riesz kernel.

Proof First assume that $\alpha>(d-1)|1/p-1/2|$. This assumption is stronger than the one in Lemma 2.1, therefore $S_*^\alpha(G^\beta * F)$ is well-defined. Since the Bessel kernel G^β is positive and it is dominated by I^β, we can estimate the maximal function as

$$\sup_{R>0}\left|S_R^\alpha * G^\beta * F(x)\right|\leqslant C I^\beta * \sup_{R>0}\left|S_R^\alpha * F\right|(x).$$

As stated in Lemma 2.2, if $\alpha>(d-1)|1/p-1/2|$ and $F\in L^p(\mathbf{R}^d)$, then also $\sup_{R>0}|S_R^\alpha * F|\in L^p(\mathbf{R}^d)$. Hence, (i) follows with $H=\sup_{R>0}|S_R^\alpha * F|$.

Now assume that $\alpha > d(1/2 - 1/p) - 1/2$ and $\alpha + \beta \geqslant (d-1)/2$. Then, by Lemma 2.3,

$$\sup_{R>0} \left| S_R^\alpha * G^\beta * F(x) \right| \leqslant |F| * \sup_{R>0} \left| S_R^\alpha * G^\beta \right|(x) \leqslant C I^\beta * |F|(x).$$

Hence, (ii) follows with $H = |F|$. □

Proof of Theorem 1.1 As observed before, the case $p = \infty$ is already known. Indeed, if $p = \infty$ and $\beta > 0$, then $f = G^\beta * F$ is bounded and uniformly continuous. Hence, if $\alpha > (d-1)/2$ then $S_R^\alpha * f$ converges to f uniformly. The case $p = 1$ is already contained in [6], however it is also a consequence of the case $p > 1$. Indeed, write $f = G^\beta * F = G^{\beta-\varepsilon} * G^\varepsilon * F$ with $0 < \varepsilon < \beta$. If F is in $L^1(\mathbf{R}^d)$ then, by the Hardy–Littlewood–Sobolev theorem of fractional integration [11, Theorem 5.1], $G^\varepsilon * F$ is in $L^p(\mathbf{R}^d)$ for every $p < d/(d-\varepsilon)$. Hence, assuming that the Theorem holds with $p > 1$, then $S_R^\alpha f$ converges up to a set with Hausdorff dimension at most $d - (\beta - \varepsilon)p$. Finally, letting $\varepsilon \to 0$ and $p \to 1$, one obtains convergence up to a set of dimension at most $d - \beta$. The case $1 < p < \infty$ and $0 < \beta \leqslant d/p$ follows from Lemma 2.4 and the notion of capacity. The proof is standard, anyhow for completeness we include some details. Let $\{F_n\}$ be a sequence of functions in the Schwartz class which converges to F in the metric of $L^p(\mathbf{R}^d)$ and let $f_n = G^\beta * F_n$. Since also f_n is in the Schwartz class, $\lim_{R\to+\infty} S_R^\alpha * f_n = f_n$ pointwise everywhere. Then, for every $t > 0$,

$$\begin{aligned}
&\left\{x : \limsup_{R\to+\infty} \left| S_R^\alpha * f(x) - f(x) \right| > t \right\} \\
&\quad \subseteq \left\{x : \sup_{R>0} \left| S_R^\alpha * f(x) - S_R^\alpha * f_n(x) \right| > t/2 \right\} \cup \left\{x : \left| f_n(x) - f(x) \right| > t/2 \right\}.
\end{aligned}$$

The Bessel capacity of the second term can be estimated by

$$\begin{aligned}
&B_{\beta,p}\left(\left\{x : \left| f_n(x) - f(x) \right| > t/2 \right\}\right) \\
&\quad \leqslant B_{\beta,p}\left(\left\{x : G^\beta * |F_n - F|(x) > t/2 \right\}\right) \\
&\quad \leqslant \left(\frac{t}{2}\right)^{-p} \|F_n - F\|_p^p.
\end{aligned}$$

Hence, this capacity tends to zero as $n \to +\infty$. In case (i) when $\alpha > (d-1)|1/p - 1/2|$, the Bessel capacity of the first term can be estimated by

$$\begin{aligned}
&B_{\beta,p}\left(\left\{x : \limsup_{R\to+\infty} \left| S_R^\alpha * f(x) - S_R^\alpha * f_n(x) \right| > t/2 \right\}\right) \\
&\quad \leqslant B_{\beta,p}\left(\left\{x : G^\beta * \sup_{R>0} \left| S_R^\alpha * (F - F_n) \right|(x) > t/2 \right\}\right)
\end{aligned}$$

$$\leqslant \left(\frac{t}{2}\right)^{-p} \left\| \sup_{R>0} \left| S_R^\alpha * (F - F_n) \right| \right\|_p^p$$
$$\leqslant C \left(\frac{t}{2}\right)^{-p} \|F_n - F\|_p^p.$$

Since $\|F_n - F\|_p \to 0$ as $n \to +\infty$, we obtain

$$B_{\beta,p}\left(\left\{x : \limsup_{R\to+\infty} \left| S_R^\alpha * f(x) - f(x) \right| > t \right\}\right) = 0.$$

In case (ii), when $\alpha \geqslant d(1/2 - 1/p) - 1/2$ and $\alpha + \beta \geqslant (d-1)/2$, the Riesz capacity of the second term can be estimated by

$$R_{\beta,p}\left(\left\{x : \limsup_{R\to+\infty} \left| S_R^\alpha * (f - f_n) \right|(x) > t/2 \right\}\right)$$
$$\leqslant R_{\beta,p}\left(\left\{x : I^\beta * \left|(F - F_n)\right|(x) > Ct/2 \right\}\right)$$
$$\leqslant \left(\frac{Ct}{2}\right)^{-p} \|F_n - F\|_p^p.$$

The last term tends to zero as $n \to +\infty$. Since the Riesz and Bessel capacities have the same null sets, see [16, p. 67], in both cases we get

$$B_{\beta,p}\left(\left\{x : \limsup_{R\to+\infty} \left| S_R^\alpha * f(x) - f(x) \right| > 0 \right\}\right)$$
$$\leqslant \sum_{k=1}^\infty B_{\beta,p}\left(\left\{x : \limsup_{R\to+\infty} \left| S_R^\alpha * f(x) - f(x) \right| > 1/k \right\}\right)$$
$$= 0.$$

Therefore, applying [16, Theorem 2.6.16]), we obtain that the Hausdorff dimension of the set $\{x : \limsup_{R\to+\infty} |S_R^\alpha * f(x) - f(x)| > 0\}$ is at most $d - \beta p$.

If $1 < p \leqslant 2$ and $\beta > d/p$ then, by the Hausdorff-Young inequality, the Fourier transform of f is absolutely integrable. Indeed, if $1/p + 1/q = 1$,

$$\int_{\mathbf{R}^d} \left|\hat{f}(\xi)\right| d\xi = \int_{\mathbf{R}^d} \left(1 + |\xi|^2\right)^{-\beta/2} \left|\widehat{F}(\xi)\right| d\xi$$
$$\leqslant \left(\int_{\mathbf{R}^d} \left(1 + |\xi|^2\right)^{-\beta p/2} d\xi\right)^{1/p} \left(\int_{\mathbf{R}^d} \left|\widehat{F}(\xi)\right|^q d\xi\right)^{1/q}$$
$$\leqslant \left(\int_{\mathbf{R}^d} \left(1 + |\xi|^2\right)^{-\beta p/2} d\xi\right)^{1/p} \left(\int_{\mathbf{R}^d} \left|F(x)\right|^p dx\right)^{1/p}.$$

Hence, by the integrability of the Fourier transform, the inversion formula for the Fourier transform holds everywhere and the convergence of the Bochner–Riesz means is uniform,

$$\begin{aligned}\left|S_R^\alpha * f(x) - f(x)\right| &= \left|\int_{\mathbf{R}^d} \left(\left(1-\left|R^{-1}\xi\right|^2\right)_+^\alpha - 1\right)\hat{f}(\xi)\exp(2\pi i \xi x)\,d\xi\right| \\ &\leqslant \int_{\mathbf{R}^d} \left|\left(1-\left|R^{-1}\xi\right|^2\right)_+^\alpha - 1\right|\left|\hat{f}(\xi)\right|d\xi.\end{aligned}$$

It remains to consider the case $p > 2$, $\alpha > d(1/2 - 1/p) - 1/2$ and $\beta > d/p$. In this case $\alpha + \beta > (d-1)/2$ and Lemma 2.3 applies. Hence, if $1/p + 1/q = 1$, then

$$\begin{aligned}\sup_{R>0}\left|S_R^\alpha * G^\beta * F(x)\right| &\leqslant \left(\sup_{R>0}\left|S_R^\alpha * G^\beta\right|\right) * |F|(x) \\ &\leqslant \left(\int_{\mathbf{R}^d}\left|\sup_{R>0}\left|S_R^\alpha * G^\beta(x)\right|\right|^q dx\right)^{1/q}\left(\int_{\mathbf{R}^d}\left|F(x)\right|^p dx\right)^{1/p}.\end{aligned}$$

Uniform convergence everywhere now follows from the uniform boundedness of this maximal function, via a standard density argument. □

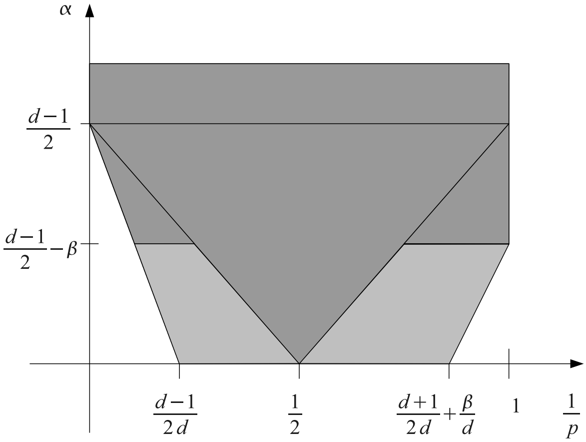
Since functions with β derivatives in $L^p(\mathbf{R}^d)$ may be infinite on sets with Hausdorff dimension $d - \beta p$, the dimension of the divergence set in the statement of the Theorem cannot be decreased. When $p \geqslant 2$, part (i) of Theorem 1.1 can be easily improved by using the bounds for the maximal Bochner–Riesz operator in [1, 5, 7], and [10] for even dimensions. Finally, we want to remark again that our analysis is not exhaustive and the ranges of the indexes are not optimal. However, at least for radial functions, we can prove some definitive results:

Let $\alpha \geqslant 0$, $\beta \geqslant 0$, $1 \leqslant p \leqslant +\infty$, and $2d/(d+1+2\alpha+2\beta) < p < 2d/(d-1-2\alpha)$. Then the Bochner–Riesz means with index α of radial functions with β derivatives in $L^p(\mathbf{R}^d)$ converge pointwise, with the possible exception of a set of points Ω with the following properties:

(i) *if $\beta p \leqslant 1$, then the Hausdorff dimension of Ω is at most $d - \beta p$;*
(ii) *if $1 < \beta p \leqslant d$, then Ω either is empty or it reduces to the origin;*
(iii) *if $\beta p > d$, then Ω is empty.*

For these results, see [15]. The range of the indexes for non-radial functions cannot be larger than for radial function. Figure 1 shows the largest possible region of convergence for our problem. The dark shaded area corresponds to condition (i) and (ii) of Theorem 1.1, whereas we have proved convergence in the light shaded areas only for radial functions. In particular, observe the asymmetry between $p < 2$ and $p > 2$. When $p < 2$ the regularity β lowers the summability index α. On the other hand, when $p > 2$ the indexes of summability α and of regularity β are unrelated.

Fig. 1 The region of convergence



References

1. Carbery, A.: The boundedness of the maximal Bochner–Riesz operator on $L^4(\mathbf{R}^2)$. Duke Math. J. **50**, 409–416 (1983)
2. Carbery, A., Rubio de Francia, J.L., Vega, L.: Almost everywhere summability of Fourier integrals. J. Lond. Math. Soc. **38**, 513–524 (1988)
3. Carbery, A., Soria, F.: Almost-everywhere convergence of Fourier integrals for functions in Sobolev spaces, and an L^2-localisation principle. Rev. Mat. Iberoam. **4**, 319–337 (1988)
4. Carbery, A., Soria, F.: Set of divergence for the localization problem for Fourier integrals. C. R. Acad. Sci. Paris Sér. I Math. **325**, 1283–1286 (1997)
5. Christ, M.: On almost everywhere convergence of Bochner–Riesz means in higher dimensions. Proc. Am. Math. Soc. **95**, 16–20 (1985)
6. Colzani, L.: Fourier expansions of functions with bounded variation of several variables. Trans. Am. Math. Soc. **358**, 5501–5521 (2006)
7. Lee, S.: Improved bounds for Bochner–Riesz and maximal Bochner–Riesz operator. Duke Math. J. **122**, 205–232 (2004)
8. Ma, B.: Almost-everywhere convergence of Bochner–Riesz means in Bessel potential spaces. J. Approx. Theory Appl. **13**, 13–18 (1997)
9. Montini, E.: On the capacity of sets of divergence associated with the spherical partial integral operator. Trans. Am. Math. Soc. **335**, 1415–1441 (2002)
10. Rubio de Francia, J.L.: Transference principles for radial multipliers. Duke Math. J. **58**, 1–19 (1989)
11. Stein, E.M.: Singular Integrals and Differentiability Properties of Functions. Princeton University Press, Princeton (1971)
12. Stein, E.M.: Harmonic Analysis: Real Variable Methods, Orthogonality and Oscillatory Integrals. Princeton University Press, Princeton (1993)
13. Stein, E.M., Weiss, G.: Introduction to Fourier Analysis on Euclidean Spaces. Princeton University Press, Princeton (1971)
14. Tao, T.: On the maximal Bochner–Riesz conjecture in the plane for $p < 2$. Trans. Am. Math. Soc. **354**, 1947–1959 (2002)
15. Volpi, S.: Bochner–Riesz means of eigenfunction expansions and local Hardy spaces on manifolds with bounded geometry. Ph.D. Thesis, Univ. Milano-Bicocca, Milano (2012)
16. Ziemer, W.P.: Weakly Differentiable Functions. Graduate Texts in Math., vol. 120. Springer, New York (1989)

Sub-Finsler Geometry and Finite Propagation Speed

Michael G. Cowling and Alessio Martini

Abstract We prove a number of results on the geometry associated to the solutions of first-order differential operators on manifolds. In particular, we consider distance functions associated to a first-order operator, and discuss the associated geometry, which is sometimes surprisingly different to Riemannian geometry.

Keywords Sub-Finsler geometry · Finite propagation speed

Mathematics Subject Classification (2010) Primary 58J45 · Secondary 58J60 · 53B40

1 Introduction

Suppose that D is a first-order, formally self-adjoint differential operator on a manifold M. Under what circumstances can we define a group of operators e^{itD} (where $t \in \mathbb{R}$) and when can we say that solutions to the corresponding differential equation $(\partial_t - iD)u = 0$ propagate with finite speed? If D is an operator between vector bundles, how do we measure the speed? The aim of this paper is to answer these questions, under some assumptions on D, which are related to the Hörmander condition for families of vector fields. We do this precisely, and while many of the ideas here are in the literature, we have not seen them put together in a coherent way as we do here.

In particular, we establish when formally self-adjoint operators are essentially self-adjoint, and produce sharp estimates for the propagation of solutions, which involve a "sub-Finsler" distance when the operators act between vector bundles. We also give a detailed description of the associated geometry.

M.G. Cowling (✉) · A. Martini
School of Mathematics and Statistics, University of New South Wales, UNSW Sydney
NSW 2052, Australia
e-mail: m.cowling@unsw.edu.au

A. Martini
e-mail: a.martini@unsw.edu.au

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3,
DOI 10.1007/978-88-470-2853-1_8, © Springer-Verlag Italia 2013

Every first-order differential operator D between vector bundles has a symbol $\sigma(D)$, which maps the cotangent space at a point x to the space of linear operators from the fibre of one vector bundle to another. The mapping that sends a cotangent vector ξ to the operator norm of $\sigma(D)(\xi)$ is thus a seminorm P_x on the cotangent space $T_x^* M$. When D is elliptic, the seminorm P_x is a norm at each point x, but when D is not elliptic, the seminorm may well have a non-trivial kernel, and the dimension of this kernel may vary from point to point. Dual to the seminorm on the cotangent space, there is an extended norm P_x^* on the tangent space $T_x M$ (by "extended norm", we mean that some vectors may have infinite norm). The annihilator of the kernel of the seminorm P_x in the tangent space is the space of tangent vectors of finite norm. Thus in general, the geometry that we consider is similar to sub-Riemannian geometry, but we must allow for the possibility that the dimension of the space of vectors of finite norm is not constant. Further, when the bundles are one-dimensional, the seminorm is euclidean (once the kernel is factored out), but when the bundles are higher-dimensional, the norm is more general. Thus we consider "sub-Finsler" geometry, an extension of sub-Riemannian geometry. We define various natural distance functions, and show that under various hypotheses they coincide; but surprisingly, they do not always do so, and we give a number of examples that show that results that are obvious in more restricted circumstances may in fact be false in our more general context. For example, we show that it may not be possible to measure the length of a smooth curve by considering a smooth parametrisation, and that the "right" distance to measure propagation may not be euclidean. Because "obvious" results may be false, we feel that we are justified in giving fairly complete proofs of most results; expert readers may skip over proofs, in the knowledge that they are the proofs that may be expected, but we do suggest looking at the counterexamples later in the paper.

As we commented above, most of the ideas that we consider are not new, but have been considered in less general contexts. For example, our technique for establishing finite propagation speed for first-order operators is well-known in the elliptic context, but less so in general; there are, for instance, a number of proofs in the subelliptic context that consider elliptic approximants to subelliptic operators rather than working directly with subelliptic operators. Some of the analysis of distance functions that we carry out is familiar in the context of "metric spaces", but those who work in the context do not seem usually to consider vector bundles.

We work in the generality of vector bundles, in order to work with self-adjoint operators. Given complex vector bundles $\mathscr{E}$ and $\mathscr{F}$, with hermitean fibre inner products (inner products on each fibre), and a differential operator $D : C^\infty(\mathscr{E}) \to C^\infty(\mathscr{F})$, we define a new differential operator $\not{D} : C^\infty(\mathscr{E} \oplus \mathscr{F}) \to C^\infty(\mathscr{E} \oplus \mathscr{F})$ as the sum of D and its formal adjoint D^+: more precisely, $\not{D}(f, g) = (D^+ g, Df)$. Then $\not{D}$ is formally self-adjoint, and $\not{D}$ induces the same distance function as D. By studying the propagation of solutions to $(\partial_t - \mathrm{i}\not{D})u = 0$, we can say something about the wave equation $(\partial_t^2 - D^+ D)v = 0$. Vector bundles are also a natural context for considering systems of vector fields: to $\{X_1, \dots, X_r\}$, we associate the differential operator sending a function f to the vector-valued function $(X_1 f, \dots, X_r f)$, that is, from a section of a trivial bundle with fibre $\mathbb{C}$ to a section of a trivial bundle with fibre $\mathbb{C}^r$.

1.1 Notation and Background

Throughout, M is an n-dimensional manifold, by which we mean a smooth σ-compact, and hence paracompact, manifold without boundary. Then M admits a countable locally finite atlas $(\varphi_\alpha)_{\alpha\in A}$; here each $U_\alpha \subseteq M$ and each φ_α is a smooth bijection from U_α to $\mathbb{R}^n$ with smooth inverse. By choosing a partition of unity $(\eta_\alpha)_{\alpha\in A}$ subordinate to the cover $(U_\alpha)_{\alpha\in A}$ and then rescaling the φ_α to ensure that $\varphi_\alpha(\operatorname{supp}(\eta_\alpha)) \subseteq B_{\mathbb{R}^n}(0,1)$, where $B_{\mathbb{R}^n}(x,r)$ denotes the open ball in $\mathbb{R}^n$ with centre x and radius r, we may suppose that $\bigcup_{\alpha\in A} V_\alpha = M$, where $V_\alpha = \varphi_\alpha^{-1}(B_{\mathbb{R}^n}(0,1))$. Then $\sum_\alpha \eta_\alpha = 1$ and the η_α are bump functions on M, by which we mean smooth compactly-supported functions taking values in $[0,1]$. We write $\mathfrak{O}(M)$ and $\mathfrak{K}(M)$, or just $\mathfrak{O}$ and $\mathfrak{K}$, for the collections of all open subsets and all compact subsets of M.

We will endow M, and subsets thereof, with various extended distance functions $\varrho : M \times M \to [0,\infty]$; by this, we mean that ϱ satisfies the usual conditions for a distance function, but may take the value ∞. One way to do this is to choose a continuous "fibre seminorm" P on T^*M, that is, P_x is a seminorm on each fibre T_x^*M, and $P : T^*M \to [0,\infty)$ is continuous. Dually, there is an extended fibre norm P^* on the tangent space TM, given by

$$P_x^*(v) = \sup_{\substack{\xi\in T_x^*M \\ P(\xi)\leqslant 1}} \left|\xi(v)\right|.$$

We then say that a curve $\gamma : [a,b] \to M$ is subunit if it is absolutely continuous and $P^*(\gamma') \leqslant 1$ almost everywhere in $[a,b]$. We define the (possibly infinite) distance $\varrho_P(x,y)$ between points x and y in M to be the infimum of the set of lengths of the intervals of definition of subunit curves starting at x and ending at y. We consider both subunit and smooth subunit curves in the text, and show, under suitable hypotheses, that it does not matter which are used, but in general there is a distinction. It is easier to work with P rather than P^*, as describing the continuity requirements on P^* is more complex; further, when P and P^* arise in the analysis of a first-order differential operator, P has a simple description in terms of the symbol of the operator.

In general, the topology induced by ϱ_P may not be equivalent to the original manifold topology of M. It is easier to work with distance functions that do give rise to the original topology, and we give these a special name.

Definition 1.1 An extended distance function is said to be *varietal* if the topology that it induces coincides with the manifold topology.

Given a distance function ϱ on M, a point x in M, and $\varepsilon \in \mathbb{R}^+$, we write

$$B_\varrho(x,\varepsilon) = \left\{y \in M : \varrho(x,y) < \varepsilon\right\} \quad \text{and} \quad B_\varrho^-(x,\varepsilon) = \left\{y \in M : \varrho(x,y) \leqslant \varepsilon\right\};$$

the latter set need not be closed in the manifold topology, and, given a subset X of M, we write X^- for the manifold closure of X. As usual, $\varrho(X,x) = \inf_{y\in X} \varrho(y,x)$. We define $B_\varrho(X,\varepsilon)$ and $B_\varrho^-(X,\varepsilon)$ analogously.

We equip M with a smooth measure that is equivalent to Lebesgue measure in all coordinate charts, and write $\mathrm{d}x$, $\mathrm{d}y$, …, for the measure elements. Take a smooth complex finite-rank fibre-normed vector bundle $\mathscr{E}$ on M. We use "function notation" for spaces of sections of $\mathscr{E}$; for instance, we write $L^p_{\mathrm{loc}}(\mathscr{E})$ for the space of (equivalence classes of) sections f of $\mathscr{E}$ such that $|f|^p$ is locally integrable on M if $p < \infty$, or $|f|$ is essentially bounded if $p = \infty$, and $L^p_{\mathrm{c}}(\mathscr{E})$ for the space of compactly-supported sections in $L^p_{\mathrm{loc}}(\mathscr{E})$. The former space is equipped with a Fréchet structure: $f_m \to f$ in $L^p_{\mathrm{loc}}(\mathscr{E})$ if and only if

$$\int_K \left| f_m(x) - f(x) \right|^p \mathrm{d}x \to 0 \quad \text{as } m \to \infty$$

for all $K \in \mathfrak{K}(M)$ (recall that, in general, a Fréchet space structure involves a countable family of seminorms Q_k such that $f = 0$ if and only if $Q_k(f) = 0$ for all indices k); the latter is an inductive limit of Banach spaces. We write $C(\mathscr{E})$ for the space of continuous sections of $\mathscr{E}$. Then convergence in $C(\mathscr{E})$ means uniform convergence on compacta. If $\mathscr{E}$ has a hermitean fibre inner product $\langle \cdot, \cdot \rangle$, then, for all $f, g \in L^2(\mathscr{E})$, we write $\langle f, g \rangle$ for their pointwise inner product, which is a function on M, and $\langle\langle f, g \rangle\rangle$ for their inner product:

$$\langle\langle f, g \rangle\rangle = \int_M \left\langle f(x), g(x) \right\rangle \mathrm{d}x.$$

We write $\mathscr{T}$ and $\mathscr{T}^r$ for the trivial bundles over M with fibres $\mathbb{C}$ and $\mathbb{C}^r$, and $\mathscr{T}_{\mathbb{R}}$ for the trivial bundle over M with fibre $\mathbb{R}$. Thus $C^\infty_{\mathrm{c}}(\mathscr{T})$ and $C^\infty_{\mathrm{c}}(\mathscr{T}_{\mathbb{R}})$ denote the usual space of smooth compactly-supported complex-valued functions on M, and the subspace thereof of real-valued functions.

Suppose that $\varphi_\alpha \colon U_\alpha \to \mathbb{R}^n$ is a coordinate chart and $\mathscr{E}$ is a vector bundle over M with fibre $\mathbb{C}^r$. On $\mathbb{R}^n$, as on any contractible manifold, all vector bundles are trivialisable [16, Corollary 3.4.8]. Thus, when we consider the restriction $\mathscr{E}|_{U_\alpha}$ of $\mathscr{E}$ to U_α, there are invertible linear maps T_x from $\mathscr{E}_x$, the fibre over x, to $\mathbb{C}^r$, which vary smoothly with x in M, so the map $w \mapsto (\pi(w), T_{\pi(w)} w)$, where π is the projection from $\mathscr{E}$ to M, is a vector bundle isomorphism of $\mathscr{E}|_{U_\alpha}$ with the bundle $U_\alpha \times \mathbb{C}^r$ over U_α. In fact, when $\mathscr{E}$ has a hermitean structure, then the T_x may be chosen to be isometries. Furthermore, the map $\varphi_\alpha \otimes I$ is a vector bundle isomorphism from the bundle $U_\alpha \times \mathbb{C}^r$ over U_α to the bundle $\mathbb{R}^n \times \mathbb{C}^r$ over $\mathbb{R}^n$. This isomorphism in turn induces an identification $\tau_{\mathscr{E},\alpha}$ of the sections of $\mathscr{E}|_{U_\alpha}$ with the sections of the trivial bundle $\mathbb{R}^n \times \mathbb{C}^r$ over $\mathbb{R}^n$, which we identify with the $\mathbb{C}^r$-valued functions on $\mathbb{R}^n$. For instance, $\tau_{\mathscr{E},\alpha} : C^\infty_{\mathrm{c}}(\mathscr{E}|_{U_\alpha}) \to C^\infty_{\mathrm{c}}(\mathbb{R}^n \times \mathbb{C}^r)$ is defined by

$$\tau_{\mathscr{E},\alpha} f(x) = T_{\varphi_\alpha^{-1}(x)} f\left(\varphi_\alpha^{-1}(x)\right) \quad \text{for every } x \in \mathbb{R}^n.$$

At the risk of confusion, we usually just write τ_α rather than $\tau_{\mathscr{E},\alpha}$. We also use τ for the map of other spaces of sections, such as $L^1_{\mathrm{loc}}(\mathscr{E})$. When we write $\tau_\alpha^{-1} f$, where f is a section over $\mathbb{R}^n$, we intend the section of $\mathscr{E}$ that vanishes outside U_α.

We use the letter κ for constants; these may vary from one paragraph to the next. We often highlight the parameters on which these constants depend.

2 Differential Operators and Symbols

We denote by $\mathfrak{D}_k(\mathscr{E}, \mathscr{F})$ the space of smooth linear kth-order differential operators from $C^\infty(\mathscr{E})$ to $C^\infty(\mathscr{F})$, where $\mathscr{E}$ and $\mathscr{F}$ are smooth complex finite-rank vector bundles on M. In local coordinates and trivialisations of the bundles, as described above, each $D \in \mathfrak{D}_k(\mathscr{E}, \mathscr{F})$ may be written as

$$\tau_\alpha(Df)(x) = \sum_{|J| \leqslant k} a_J(x)\partial_J(\tau_\alpha f)(x) \quad \text{for every } x \in \mathbb{R}^n, \tag{1}$$

where the J are multi-indices and the coefficients $a_J(x)$ are matrices that depend smoothly on x in $\mathbb{R}^n$. We also write

$$\tau_\alpha D = \sum_{|J| \leqslant k} a_J \partial_J .$$

Note that $\mathfrak{D}_{k_1}(\mathscr{E}, \mathscr{F}) \subseteq \mathfrak{D}_{k_2}(\mathscr{E}, \mathscr{F})$ if $k_1 \leqslant k_2$.

Each $D \in \mathfrak{D}_k(\mathscr{E}, \mathscr{F})$ has an associated symbol $\sigma_k(D)$, which is a smooth section of $\operatorname{Hom}(S^k(\mathbb{C}T^*M), \operatorname{Hom}(\mathscr{E}, \mathscr{F}))$; that is, the symbol $\sigma_k(D)$ at a point $x \in M$ is a $\operatorname{Hom}(\mathscr{E}_x, \mathscr{F}_x)$-valued symmetric k-linear form on $\mathbb{C}T^*_x M$. In local coordinates and trivialisations, if D is given by (1), then

$$\tau_\alpha\big(\sigma_k(D)\big)(x)\big(\xi^{\odot k}\big) = \sum_{|J|=k} \xi^J a_J(x) \quad \text{for every } x \in \mathbb{R}^n,\ \xi \in \mathbb{C}^n \tag{2}$$

where $\xi^{\odot k}$ denotes the symmetrised version of $\xi \otimes \cdots \otimes \xi$ (with k factors). The mapping $D \mapsto \sigma_k(D)$ is $\mathbb{C}$-linear, and its kernel is $\mathfrak{D}_{k-1}(\mathscr{E}, \mathscr{F})$; furthermore, if $D_1 \in \mathfrak{D}_{k_1}(\mathscr{E}, \mathscr{F})$ and $D_2 \in \mathfrak{D}_{k_2}(\mathscr{F}, \mathscr{G})$, where $\mathscr{G}$ is another vector bundle on M, then $D_2 D_1 \in \mathfrak{D}_{k_1+k_2}(\mathscr{E}, \mathscr{G})$ and

$$\sigma_{k_1+k_2}(D_2 D_1)\big(\xi^{\odot(k_1+k_2)}\big) = \sigma_{k_2}(D_2)\big(\xi^{\odot k_2}\big)\sigma_{k_1}(D_1)\big(\xi^{\odot k_1}\big) \tag{3}$$

for all $\xi \in \mathbb{C}T^*M$.

Recall that M is endowed with a smooth measure that is equivalent to Lebesgue measure in all coordinate charts, and suppose that $\mathscr{E}$ and $\mathscr{F}$ are endowed with hermitean fibre inner products. Then each $D \in \mathfrak{D}_k(\mathscr{E}, \mathscr{F})$ has a formal adjoint $D^+ \in \mathfrak{D}_k(\mathscr{F}, \mathscr{E})$, which is uniquely determined by the identity

$$\langle\!\langle Df, g \rangle\!\rangle = \big\langle\!\big\langle f, D^+ g \big\rangle\!\big\rangle \tag{4}$$

for all $f \in C^\infty_{\mathrm{c}}(\mathscr{E})$ and $g \in C^\infty_{\mathrm{c}}(\mathscr{F})$. This identity extends to sections f and g such that $\operatorname{supp} f \cap \operatorname{supp} g$ is compact, since $\langle\!\langle Df, g \rangle\!\rangle = \langle\!\langle D(\eta f), \eta g \rangle\!\rangle$ for all bump functions η equal to 1 on $\operatorname{supp} f \cap \operatorname{supp} g$. Clearly, the mapping $D \mapsto D^+$ is conjugate-linear and $(D_2 D_1)^+ = D_1^+ D_2^+$. Moreover, for all $\theta \in S^k(\mathbb{C}T^*M)$,

$$\sigma_k\big(D^+\big)(\theta) = (-1)^k \big(\sigma_k(D)(\bar{\theta})\big)^* \tag{5}$$

where the final * denotes the adjoint with respect to the hermitean inner products along the fibres of $\mathscr{E}$ and $\mathscr{F}$; note that the symbol of the formal adjoint does not depend on the choice of measure on M.

2.1 Zeroth-Order Differential Operators

Every $D \in \mathfrak{D}_0(\mathscr{E}, \mathscr{F})$ is a multiplication operator: it is given by multiplication by a smooth section of $\mathrm{Hom}(\mathscr{E}, \mathscr{F})$, namely, the symbol $\sigma_0(D)$. Formal adjunction of D then corresponds to pointwise adjunction of the multiplier:

$$\langle\langle hf, g\rangle\rangle = \big\langle\big\langle f, h^* g\big\rangle\big\rangle \tag{6}$$

for all $h \in C^\infty(\mathrm{Hom}(\mathscr{E}, \mathscr{F}))$, all $f \in C^\infty(\mathscr{E})$, and all $g \in C^\infty(\mathscr{F})$ such that $\mathrm{supp}\, f \cap \mathrm{supp}\, g \cap \mathrm{supp}\, h$ is compact. Here are some special cases of (6).

First, if $\mathscr{E} = \mathscr{F}$ and $h \in C^\infty(\mathscr{T})$, then h corresponds to a scalar section of $\mathrm{Hom}(\mathscr{E}, \mathscr{E})$, whose pointwise adjoint corresponds to the pointwise conjugate $\bar{h}$, so

$$\langle\langle hf, g\rangle\rangle = \langle\langle f, \bar{h} g\rangle\rangle.$$

Next, if $h, g \in C^\infty(\mathscr{E})$ and $f \in C^\infty(\mathscr{T})$, then h corresponds to a smooth section of $\mathrm{Hom}(\mathscr{T}, \mathscr{E})$, whose pointwise adjoint corresponds to the section $h^* = \langle \cdot, h\rangle$ of $\mathscr{E}^*$, which we may identify with $\mathrm{Hom}(\mathscr{E}, \mathscr{T})$; now

$$\langle\langle fh, g\rangle\rangle = \big\langle\big\langle f, h^* g\big\rangle\big\rangle = \big\langle\big\langle f, \langle g, h\rangle\big\rangle\big\rangle.$$

Finally, if $h \in C^\infty(\mathscr{E})$, $f \in C^\infty(\mathrm{Hom}(\mathscr{E}, \mathscr{F}))$ and $g \in C^\infty(\mathscr{F})$, then h corresponds to a smooth section of $\mathrm{Hom}(\mathrm{Hom}(\mathscr{E}, \mathscr{F}), \mathscr{F})$, whose pointwise adjoint, with respect to the Hilbert–Schmidt inner product on $\mathrm{Hom}(\mathscr{E}, \mathscr{F})$, is a section of $\mathrm{Hom}(\mathscr{F}, \mathrm{Hom}(\mathscr{E}, \mathscr{F}))$, given, modulo the identification of $\mathrm{Hom}(\mathscr{E}, \mathscr{F})$ with $\mathscr{E}^* \otimes \mathscr{F}$, by the pointwise tensor product with h^*, and

$$\langle\langle fh, g\rangle\rangle = \big\langle\big\langle f, h^* \otimes g\big\rangle\big\rangle.$$

By the way, by using a partition of unity and local trivialisations, it is easily shown that each smooth compactly-supported section h of $\mathrm{Hom}(\mathscr{E}, \mathscr{F})$ may be written as a finite sum of sections of the form $f^* \otimes g$ for appropriate $f \in C_c^\infty(\mathscr{E})$ and $g \in C_c^\infty(\mathscr{F})$.

2.2 First-Order Differential Operators

Suppose that $D \in \mathfrak{D}_1(\mathscr{E}, \mathscr{F})$. Given any $h \in C^\infty(\mathscr{T})$, denote by $m_{\mathscr{E}}(h)$ and $m_{\mathscr{F}}(h)$ the multiplication operators $f \mapsto hf$ on smooth sections of $\mathscr{E}$ and $\mathscr{F}$, and define

$$\big[D, m(h)\big] = D m_{\mathscr{E}}(h) - m_{\mathscr{F}}(h) D. \tag{7}$$

In local coordinates and trivialisations, if D is given by (1), then

$$\tau_\alpha\big(\big[D, m(h)\big] f\big)(x) = \sum_{j=1}^{n} \partial_j(\tau_\alpha h)(x) a_j(x) \tau_\alpha f(x) \quad \text{for every } x \in \mathbb{R}^n;$$

in other words, the commutator $[D, m(h)]\colon C^\infty(\mathscr{E}) \to C^\infty(\mathscr{F})$ acts by multiplication by $\sigma_1(D)(\mathrm{d}h) \in C^\infty(\mathrm{Hom}(\mathscr{E}, \mathscr{F}))$. Observe that the correspondence $h \mapsto \sigma_1(D)(\mathrm{d}h)$ is a differential operator $D^\sigma \in \mathfrak{D}_1(\mathscr{T}, \mathrm{Hom}(\mathscr{E}, \mathscr{F}))$, given in local coordinates by

$$\tau_\alpha\big(D^\sigma h\big)(x) = \sum_{j=1}^{n} \partial_j(\tau_\alpha h)(x) a_j(x) \quad \text{for every } x \in \mathbb{R}^n.$$

Clearly D^σ is homogeneous, that is, $D^\sigma 1 = 0$, and the map $D \mapsto D^\sigma$ is linear. Moreover (7) may be rewritten as Leibniz' rule for D, that is,

$$D(hf) = \big(D^\sigma h\big) f + h Df$$

for all $f \in C^\infty(\mathscr{E})$ and $h \in C^\infty(\mathscr{T})$. This identity, together with (4) and its zeroth-order instances discussed in Sect. 2.1, easily implies that

$$\big(D^\sigma\big)^+\big(f^* \otimes g\big) = \big\langle D^+ g, f\big\rangle - \langle g, Df\rangle \tag{8}$$

for all $f \in C^\infty(\mathscr{E})$ and $g \in C^\infty(\mathscr{F})$, whereas from (5) it follows that

$$\big(D^+\big)^\sigma h = -\big(D^\sigma \bar{h}\big)^*$$

for all $h \in C^\infty(\mathscr{T})$.

For more on differential operators, see [21, Sect. IV], [4, Sect. 2.1], and [23, Sect. IV.2]; see also [14, Sect. 10] for the first-order case.

3 Distributions and Weak Differentiability

Recall that $C_c^\infty(\mathscr{E})$ denotes the LF-space of compactly-supported smooth sections of $\mathscr{E}$; its conjugate dual $C_c^\infty(\mathscr{E})'$ is the space of $\mathscr{E}$-valued distributions on M. As usual, we identify a locally integrable section $f \in L^1_{\mathrm{loc}}(\mathscr{E})$ with the distribution $\varphi \mapsto \langle\langle f, \varphi\rangle\rangle$ and extend the inner product between sections of $\mathscr{E}$ to denote the duality pairing between $C_c^\infty(\mathscr{E})'$ and $C_c^\infty(\mathscr{E})$.

Every differential operator $D \in \mathfrak{D}_k(\mathscr{E}, \mathscr{F})$ then extends to an operator on distributions: given any $u \in C_c^\infty(\mathscr{E})'$, we define $Du \in C_c^\infty(\mathscr{F})'$ by

$$\langle\langle Du, \varphi\rangle\rangle = \big\langle\big\langle u, D^+\varphi\big\rangle\big\rangle \quad \text{for every } \varphi \in C_c^\infty(\mathscr{F}). \tag{9}$$

Since zeroth-order differential operators are multiplication operators, (9) includes the definition of the "pointwise product" of smooth sections and distributions, in all the variants discussed in Sect. 2.1. Moreover the identity

$$\langle\langle u^*, \varphi\rangle\rangle = \langle\langle u, \varphi^*\rangle\rangle^-$$

allows us to extend pointwise adjunction to $\mathrm{Hom}(\mathscr{E}, \mathscr{F})$-valued distributions.

For a first-order operator D, with these definitions, we may extend the identities of Sect. 2.2 to the realm of distributions. For instance, to show that

$$\big(D^+\big)^\sigma h = -\big(D^\sigma \bar{h}\big)^* \tag{10}$$

for all $h \in C_c^\infty(\mathscr{T})'$, we note that it suffices to test this distributional identity on sections of the form $f^* \otimes g$ where $f \in C_c^\infty(\mathscr{E})$ and $g \in C_c^\infty(\mathscr{F})$; to do this, we apply (8). Similarly it may be proved that

$$D(hf) = \big(D^\sigma h\big) f + h D f \tag{11}$$

when $h \in C_c^\infty(\mathscr{T})'$ and $f \in C^\infty(\mathscr{E})$, or when $h \in C^\infty(\mathscr{T})$ and $f \in C_c^\infty(\mathscr{E})'$.

For $D \in \mathfrak{D}_k(\mathscr{E}, \mathscr{F})$, the definition of the D-derivative of an $\mathscr{E}$-valued distribution is based on that of the formal adjoint D^+ and depends on the choice of measure on M and on the hermitean structures on $\mathscr{E}$ and $\mathscr{F}$; the same holds for the definition of the embedding of $L^1_{\mathrm{loc}}(\mathscr{E})$ in $C_c^\infty(\mathscr{E})'$. However, $L^1_{\mathrm{loc}}(\mathscr{E})$ and $L^1_{\mathrm{loc}}(\mathscr{F})$ do not depend on those structures: if we change the measure or inner products, then we get the same linear spaces, with equivalent families of seminorms and so equivalent Fréchet structures. Moreover, if $f \in L^1_{\mathrm{loc}}(\mathscr{E})$ and $Df \in L^1_{\mathrm{loc}}(\mathscr{E})$, then the section in $L^1_{\mathrm{loc}}(\mathscr{F})$ that corresponds to the distributional derivative Df does not depend on these structures.

We say that $f \in L^1_{\mathrm{loc}}(\mathscr{E})$ is weakly D-differentiable if $Df \in L^1_{\mathrm{loc}}(\mathscr{F})$. Given any $p \in [1, \infty]$ and $D \in \mathfrak{D}_k(\mathscr{E}, \mathscr{F})$, we define the local Sobolev space $W^p_{D,\mathrm{loc}}(\mathscr{E})$ by

$$W^p_{D,\mathrm{loc}}(\mathscr{E}) = \big\{ f \in L^p_{\mathrm{loc}}(\mathscr{E}) : Df \in L^p_{\mathrm{loc}}(\mathscr{F}) \big\},$$

which is given a Fréchet structure by identifying it with a closed subspace of $L^p_{\mathrm{loc}}(\mathscr{E}) \times L^p_{\mathrm{loc}}(\mathscr{F})$ by the map $f \mapsto (f, Df)$. Similarly, we define the Sobolev space $W^p_D(\mathscr{E})$ by

$$W^p_D(\mathscr{E}) = \big\{ f \in L^p(\mathscr{E}) : Df \in L^p(\mathscr{F}) \big\}.$$

The Banach space $W^p_D(\mathscr{E})$ depends on the choice of measure on M and on the hermitean structures on $\mathscr{E}$ and $\mathscr{F}$, while $W^p_{D,\mathrm{loc}}(\mathscr{E})$ does not. Finally, $W^p_{D,0}(\mathscr{E})$ denotes the closure of $C_c^\infty(\mathscr{E})$ in $W^p_D(\mathscr{E})$.

3.1 Mollifiers and Smooth Approximation

Mollifiers, introduced by Friedrichs [10], allow us to approximate distributions, and in particular, locally integrable functions, by smooth functions. We now describe the application of this technique to sections of vector bundles on the manifold M.

For convenience, we first consider the case where M is $\mathbb{R}^n$, equipped with Lebesgue measure and euclidean distance function, and $\mathscr{T}$ is the trivial bundle $\mathbb{R}^n \times \mathbb{C}$ over $\mathbb{R}^n$. Recall that all vector bundles on $\mathbb{R}^n$ are trivialisable, and sections of a trivial bundle over $\mathbb{R}^n$ with fibre $\mathbb{C}^r$ may be identified with functions from $\mathbb{R}^n$ to $\mathbb{C}^r$. Hence it is easy to define mollifiers on $\mathbb{R}^n$ globally, and mollifiers on general manifolds and bundles may then be defined by local trivialisations and partitions of unity.

Choose a bump function $\varphi \in C^\infty_{\mathrm{c}}(\mathscr{T})$ with unit mass and support in the unit ball; for all $\varepsilon \in \,]0, 1]$, define $\varphi_\varepsilon(x) = \varepsilon^{-n}\varphi(\varepsilon^{-1}x)$ for all $x \in \mathbb{R}^n$. For a distributional section $f \in C^\infty_{\mathrm{c}}(\mathscr{T}^r)'$ of a trivial bundle with fibre $\mathbb{C}^r$, we set

$$J_\varepsilon f(x) = \varphi_\varepsilon * f(x) = \sum_{k=1}^{r} \big\langle\!\big\langle f, \varphi_\varepsilon(x - \cdot)e_k \big\rangle\!\big\rangle e_k \quad \text{for every } x \in \mathbb{R}^n, \tag{12}$$

where $\{e_1, \dots, e_r\}$ is the canonical basis of $\mathbb{C}^r$.

Proposition 3.1 *Suppose that $\mathscr{E}$ is the trivial bundle $\mathbb{R}^n \times \mathbb{C}^r$ over $\mathbb{R}^n$, and that $1 \leqslant p \leqslant \infty$. For all $f \in C^\infty_{\mathrm{c}}(\mathscr{E})'$, the formula (12) defines smooth sections $J_\varepsilon f$ of $\mathscr{E}$ that converge to f distributionally as $\varepsilon \to 0$. Moreover, the following hold.*

- (i) (Supports) $\operatorname{supp} J_\varepsilon f \subseteq B^-_{\mathbb{R}^n}(\operatorname{supp} f, \varepsilon)$.
- (ii) (Equicontinuity) *The operators J_ε are bounded on $L^p_{\mathrm{loc}}(\mathscr{E})$, uniformly for ε in $]0, 1]$.*
- (iii) (Approximation) *If $p < \infty$ and $f \in L^p_{\mathrm{loc}}(\mathscr{E})$, then $J_\varepsilon f \to f$ in $L^p_{\mathrm{loc}}(\mathscr{F})$ as $\varepsilon \to 0$; the same holds if $p = \infty$ and $f \in C(\mathscr{E})$.*
- (iv) (Upper bound) *For all continuous fibre seminorms P on $\mathscr{E}$, all $K \in \mathfrak{K}(\mathbb{R}^n)$ and all $f \in L^\infty_{\mathrm{loc}}(\mathscr{E})$,*

$$\limsup_{\varepsilon \to 0} \sup_{x \in K} P(J_\varepsilon f)(x) \leqslant \inf_{\substack{W \in \mathfrak{O} \\ W \supseteq K}} \operatorname*{ess\,sup}_{x \in W} P(f)(x).$$

Proof These are well-known facts about convolution and approximate identities in $\mathbb{R}^n$, and we omit the proofs, except for part (iv).

For all $x \in \mathbb{R}^n$ and $\varepsilon \in \mathbb{R}^+$,

$$J_\varepsilon f(x) = \int_{\mathbb{R}^n} \varphi_\varepsilon(y) f(x - y)\, \mathrm{d}y.$$

Since the functions φ_ε are nonnegative and have unit mass, while $P_x : \mathscr{E}_x \to \mathbb{R}$ is convex, Jensen's inequality implies that

$$P_x\big(J_\varepsilon f(x)\big) \leqslant \int_{\mathbb{R}^n} \varphi_\varepsilon(y) P_x\big(f(x - y)\big)\, \mathrm{d}y,$$

whence

$$P(J_\varepsilon f)(x) \leqslant \int_{\mathbb{R}^n} \varphi_\varepsilon(y) P(f)(x-y)\,\mathrm{d}y \\ + \int_{\mathbb{R}^n} \varphi_\varepsilon(y) \bigl(P_x\bigl(f(x-y)\bigr) - P_{x-y}\bigl(f(x-y)\bigr)\bigr)\,\mathrm{d}y. \qquad (13)$$

Suppose now that $K \subseteq W$, where $K \in \mathfrak{K}(\mathbb{R}^n)$ and $W \in \mathfrak{O}(\mathbb{R}^n)$. Take $\bar{\varepsilon} \in \mathbb{R}^+$ such that $B^-_{\mathbb{R}^n}(K, \bar{\varepsilon}) \subseteq W$. Since $f \in L^\infty_{\mathrm{loc}}(\mathscr{E})$,

$$\operatorname*{ess\,sup}_{x \in B^-_{\mathbb{R}^n}(K,\bar{\varepsilon})} \bigl|f(x)\bigr| < \infty;$$

further, $P\colon \mathscr{E} \to \mathbb{R}$ is continuous, so uniformly continuous when restricted to $B^-_{\mathbb{R}^n}(K, \bar{\varepsilon}) \times \{v \in \mathbb{C}^r : |v| \leqslant R\}$, for all $R \in \mathbb{R}^+$. Thus the second integral on the right-hand side of (13) tends to 0 as $\varepsilon \to 0$, while the first integral is bounded by $\operatorname{ess\,sup}_{x\in W} P(f)(x)$ when $\varepsilon \leqslant \bar{\varepsilon}$. Part (iv) follows. □

The interaction of mollifiers and differentiation is more interesting: for a differential operator $D \in \mathfrak{D}_k(\mathscr{E}, \mathscr{F})$, it is reasonable to ask whether $DJ_\varepsilon f$ converges to Df as $\varepsilon \to 0$. When we are working on $\mathbb{R}^n$ with trivial bundles $\mathscr{E}$ and $\mathscr{F}$, we already know that $J_\varepsilon Df$ approximates Df. In this case, the problem reduces to the study of the commutator operators $[D, J_\varepsilon]$, given by

$$[D, J_\varepsilon]f = DJ_\varepsilon f - J_\varepsilon Df.$$

If D is translation-invariant, then $[D, J_\varepsilon] = 0$. For an arbitrary D, it is clear that $[D, J_\varepsilon]f \to 0$ distributionally as $\varepsilon \to 0$. Stronger forms of convergence to 0 may be proved easily for first-order operators D.

Proposition 3.2 *Suppose that $\mathscr{E}$ and $\mathscr{F}$ are the trivial bundles $\mathbb{R}^n \times \mathbb{C}^r$ and $\mathbb{R}^n \times \mathbb{C}^s$ over $\mathbb{R}^n$, that $D \in \mathfrak{D}_1(\mathscr{E}, \mathscr{F})$, and that $1 \leqslant p \leqslant \infty$. Then the following hold.*

(i) (Equicontinuity) *The operators $[D, J_\varepsilon]$ are bounded from $L^p_{\mathrm{loc}}(\mathscr{E})$ to $L^p_{\mathrm{loc}}(\mathscr{F})$, uniformly for ε in* $]0, 1]$.
(ii) (Vanishing) *If $p < \infty$ and $f \in L^p_{\mathrm{loc}}(\mathscr{E})$, then $[D, J_\varepsilon]f \to 0$ in $L^p_{\mathrm{loc}}(\mathscr{F})$ as $\varepsilon \to 0$; the same holds if $p = \infty$ and $f \in C(\mathscr{E})$.*

Proof Compare with [11, Appendix].

We may suppose that D has the form

$$Df(x) = b(x)f(x) + \sum_{j=1}^{n} a_j(x)\partial_j f(x) = D_0 f(x) + D_1 f(x),$$

say, where the matrix-valued functions a_j and b are smooth.

Since D_0 is a multiplication operator, it is bounded from $L^p_{\mathrm{loc}}(\mathscr{E})$ to $L^p_{\mathrm{loc}}(\mathscr{F})$ for all $p \in [1, \infty]$. Hence, by Proposition 3.1, both $D_0 J_\varepsilon$ and $J_\varepsilon D_0$ are bounded from $L^p_{\mathrm{loc}}(\mathscr{E})$ to $L^p_{\mathrm{loc}}(\mathscr{F})$, uniformly for ε in $]0, 1]$, and

$$D_0 J_\varepsilon f \to D_0 f \quad \text{and} \quad J_\varepsilon D_0 f \to D_0 f$$

in $L^p_{\mathrm{loc}}(\mathscr{F})$ as $\varepsilon \to 0$ if either $p < \infty$ and $f \in L^p_{\mathrm{loc}}$, or $p = \infty$ and $f \in C$. Hence parts (i) and (ii) hold for $[D_0, J_\varepsilon]$, and it suffices to consider D_1.

If $f \in C^\infty_{\mathrm{c}}(\mathscr{E})$, then

$$J_\varepsilon f(x) = \int_{\mathbb{R}^n} f(x-y)\varphi_\varepsilon(y)\,\mathrm{d}y,$$

so

$$\begin{aligned}[D_1, J_\varepsilon] f(x) &= D_1 J_\varepsilon f(x) - J_\varepsilon D_1 f(x) \\ &= \int_{\mathbb{R}^n} \left(\sum_{j=1}^n \big[a_j(x) - a_j(x-y)\big] \partial_j f(x-y) \right) \varphi_\varepsilon(y)\,\mathrm{d}y \\ &= C_\varepsilon f(x),\end{aligned}$$

say.

Define

$$F_\varepsilon(x, y) = \sum_{j=1}^n \Big(\varphi_\varepsilon(y) \partial_j a_j(x-y) + \big[a_j(x) - a_j(x-y)\big] \partial_j \varphi_\varepsilon(y) \Big),$$

and observe that, when $f \in C^\infty_{\mathrm{c}}(\mathscr{E})$ and $\lambda \in \mathbb{C}$,

$$\begin{aligned}C_\varepsilon f(x) &= \int_{\mathbb{R}^n} \left(\sum_{j=1}^n \big[a_j(x) - a_j(y)\big] \partial_j f(y) \right) \varphi_\varepsilon(x-y)\,\mathrm{d}y \\ &= -\int_{\mathbb{R}^n} \left(\sum_{j=1}^n \frac{\partial}{\partial y_j} \big[\big[a_j(x) - a_j(y)\big] \varphi_\varepsilon(x-y)\big] \right) f(y)\,\mathrm{d}y \\ &= -\int_{\mathbb{R}^n} \left(\sum_{j=1}^n \frac{\partial}{\partial y_j} \big[\big[a_j(x) - a_j(y)\big] \varphi_\varepsilon(x-y)\big] \right) \big[f(y) - \lambda f(x)\big]\,\mathrm{d}y \\ &= \int_{\mathbb{R}^n} F_\varepsilon(x, y) \big[f(x-y) - \lambda f(x)\big]\,\mathrm{d}y.\end{aligned}$$

This formula extends to all $f \in C^\infty_{\mathrm{c}}(\mathscr{E})'$ by continuity, since F_ε is smooth and supported in the set $\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : |y| \leqslant \varepsilon\}$.

For $x \in \mathbb{R}^n$, define the quadrilinear form $A(x)$ on $\mathbb{C}^n \times \mathbb{C}^n \times \mathbb{C}^r \times \mathbb{C}^s$ by

$$A(x)\big(u', u, v, w\big) = \sum_{j,k=1}^n u'_j \frac{\partial}{\partial x_j} \big[u_k \big(a_k(x) v, w\big)\big],$$

where (w', w) denotes $\sum_{l=1}^{s} w'_l w_l$, and write $|A(x)|$ for the maximum of the expression $|A(x)(u', u, v, w)|$ as u', u, v and w range over the unit spheres in $\mathbb{C}^n$, $\mathbb{C}^n$, $\mathbb{C}^r$ and $\mathbb{C}^s$. Then for $v \in \mathbb{C}^r$ and $w \in \mathbb{C}^s$,

$$\begin{aligned}\left|\left(\sum_{j=1}^{n}\big[a_j(x)-a_j(x-y)\big]\partial_j\varphi_\varepsilon(y)v,\, w\right)\right| &= \left|\int_0^1 A(x-y+ty)\big(y, \nabla(\varphi_\varepsilon)(y), v, w\big)\right| \\ &\leqslant \sup_{z\in B_{\mathbb{R}^n}(x,|y|)} \big|A(z)\big|\,|y|\,\big|\nabla(\varphi_\varepsilon)(y)\big|\,|v|\,|w| \\ &\leqslant \sup_{z\in B_{\mathbb{R}^n}(x,|y|)} \big|A(z)\big|\,|\nabla\varphi|_\varepsilon(y)|\,|v|\,|w|,\end{aligned}$$

where $|\nabla\varphi|_\varepsilon(y) = \varepsilon^{-n}|\nabla\varphi|(\varepsilon^{-1}y)$, and similarly

$$\begin{aligned}\left|\sum_{j=1}^{n}\varphi_\varepsilon(y)\partial_j a_j(x-y)v\right| &= \big|\varphi_\varepsilon(y)\big|\left|\sum_{j=1}^{n} A(x-y)(e_j, e_j, v, w)\right| \\ &\leqslant n\big|\varphi_\varepsilon(y)\big| \sup_{z\in B_{\mathbb{R}^n}(x,|y|)} \big|A(z)\big|\,|v|\,|w|,\end{aligned}$$

where the e_j are the standard basis vectors in $\mathbb{R}^n$.

Set $\psi = n\varphi + |\nabla\varphi|$ and $\psi_\varepsilon(z) = \varepsilon^{-n}\psi(\varepsilon^{-1}z)$. Then, taking operator norms,

$$\begin{aligned}\big|F_\varepsilon(x,y)\big| &\leqslant \left|\sum_{j=1}^{n}\big[\partial_j a_j(x-y)\big]\varphi_\varepsilon(y)\right| + \left|\sum_{j=1}^{n}\big[\big[a_j(x-y)-a_j(x)\big]\partial_j\varphi_\varepsilon(y)\big]\right| \\ &\leqslant \sup_{z\in B_{\mathbb{R}^n}(x,\varepsilon)} \big|A(z)\big|\psi_\varepsilon(y).\end{aligned}$$

Now ψ is continuous and supported in the unit ball, hence bounded, so

$$\begin{aligned}\big|C_\varepsilon f(x)\big| &\leqslant \left|\int_{\mathbb{R}^n} F_\varepsilon(x,y)\big(f(x-y)-\lambda f(x)\big)\,\mathrm{d}y\right| \\ &\leqslant \sup_{z\in B_{\mathbb{R}^n}(x,\varepsilon)} \big|A(z)\big| \int_{B_{\mathbb{R}^n}(0,\varepsilon)} \psi_\varepsilon(y)\big|f(x-y)-\lambda f(x)\big|\,\mathrm{d}y,\end{aligned}$$

whence, from Minkowski's inequality, for all $K \in \mathfrak{K}(\mathbb{R}^n)$ and $\varepsilon \in]0,1]$,

$$\begin{aligned}&\left(\int_K \big|C_\varepsilon f(x)\big|^p\,\mathrm{d}x\right)^{1/p} \\ &\leqslant \sup_{z\in B_{\mathbb{R}^n}(K,\varepsilon)} \big|A(z)\big| \int_{B_{\mathbb{R}^n}(0,\varepsilon)} \psi_\varepsilon(y)\left(\int_K \big|f(x-y)-\lambda f(x)\big|^p\,\mathrm{d}x\right)^{1/p}\mathrm{d}y \\ &\leqslant \kappa_{n,K,D,\varphi} \sup_{y\in B_{\mathbb{R}^n}(0,1)} \left(\int_K \big|f(x-y)-\lambda f(x)\big|^p\,\mathrm{d}x\right)^{1/p}.\end{aligned}$$

On the one hand, when $\lambda = 1$, the integral on the right-hand side tends to 0 as $y \to 0$, and we obtain part (ii) in the case where $p < \infty$. On the other hand, when $\lambda = 0$, it follows that

$$\left(\int_K \left| C_\varepsilon f(x) \right|^p \, \mathrm{d}x \right)^{1/p} \leqslant \kappa_{n,K,D,\varphi} \left(\int_{B_{\mathbb{R}^n}(K,1)} \left| f(x) \right|^p \, \mathrm{d}x \right)^{1/p}$$

which establishes part (i) in the case where $p < \infty$. To prove parts (i) and (ii) when $p = \infty$, we replace the L^p norm in the argument above by an essential supremum. □

We consider now the general case. Recall that there is a countable locally finite atlas of smooth bijections $\varphi_\alpha : U_\alpha \to \mathbb{R}^n$ on M; here $U_\alpha \subseteq M$ and $\bigcup_{\alpha \in A} V_\alpha = M$, where $V_\alpha = \varphi_\alpha^{-1}(B_{\mathbb{R}^n}(0, 1))$. There is also a partition of unity $(\eta_\alpha)_{\alpha \in A}$ on M for which $\operatorname{supp}(\eta_\alpha) \subseteq V_\alpha$.

For each α, there is a trivialisation τ_α taking sections of $\mathscr{E}$ over U_α to sections of a trivial bundle $\mathscr{T}^r$ on $\mathbb{R}^n$, and similarly for $\mathscr{F}$. Sections of $\mathscr{E}$ with support contained in V_α are then identified with sections of $\mathscr{T}^r$ with support contained in the open unit ball.

Denote by E the set of all sequences $(\varepsilon_\alpha)_{\alpha \in A}$, where each $\varepsilon_\alpha \in \,]0, 1]$, that is, $E = \,]0, 1]^A$. For $\boldsymbol{\varepsilon} \in E$ and $f \in L^1_{\mathrm{loc}}(\mathscr{E})$, define $J_{\boldsymbol{\varepsilon}}^{\mathscr{E},\boldsymbol{\tau}} f$, which we usually write as $J_{\boldsymbol{\varepsilon}}^{\boldsymbol{\tau}} f$, as follows:

$$J_{\boldsymbol{\varepsilon}}^{\boldsymbol{\tau}} f = \sum_\alpha \tau_\alpha^{-1} J_{\varepsilon_\alpha} \tau_\alpha (\eta_\alpha f); \tag{14}$$

since $\operatorname{supp} \tau_\alpha^{-1} J_{\varepsilon_\alpha} \tau_\alpha (\eta_\alpha f) \subseteq U_\alpha$, this sum is locally finite.

Given $\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}' \in E$, we write $\boldsymbol{\varepsilon} \leqslant \boldsymbol{\varepsilon}'$ when $\varepsilon_\alpha \leqslant \varepsilon'_\alpha$ for all $\alpha \in A$. The ordering of E gives a meaning to limit-like expressions along E, such as

$$\limsup_{\boldsymbol{\varepsilon} \to \mathbf{0}} F(\boldsymbol{\varepsilon}) = \inf_{\bar{\boldsymbol{\varepsilon}} \in E} \sup_{\substack{\boldsymbol{\varepsilon} \in E \\ \boldsymbol{\varepsilon} \leqslant \bar{\boldsymbol{\varepsilon}}}} F(\boldsymbol{\varepsilon})$$

for a function $F : E \to [-\infty, \infty]$. We show now that $J_{\boldsymbol{\varepsilon}}^{\boldsymbol{\tau}} f \to f$ as $\boldsymbol{\varepsilon} \to \mathbf{0}$.

Theorem 3.3 *Suppose that $\mathscr{E}$ and $\mathscr{F}$ are vector bundles on a manifold M, that $D \in \mathfrak{D}_1(\mathscr{E}, \mathscr{F})$, and that $1 \leqslant p \leqslant \infty$. Then the linear operators $J_{\boldsymbol{\varepsilon}}^{\boldsymbol{\tau}}$ on $L^1_{\mathrm{loc}}(\mathscr{E})$ defined by* (14) *have the following properties.*

(i) (Smoothing) *If $f \in L^1_{\mathrm{loc}}(\mathscr{E})$, then $J_{\boldsymbol{\varepsilon}}^{\boldsymbol{\tau}} f \in C^\infty(\mathscr{E})$.*
(ii) (Supports) *If C is a closed subset of M and W is an open neighbourhood of C, then there exists $\bar{\boldsymbol{\varepsilon}} \in E$ such that* $\operatorname{supp} J_{\boldsymbol{\varepsilon}}^{\boldsymbol{\tau}} f \subseteq W$ *when* $\operatorname{supp} f \subseteq C$ *and $\boldsymbol{\varepsilon} \leqslant \bar{\boldsymbol{\varepsilon}}$.*
(iii) (Equicontinuity) *If $\zeta \in L^\infty_{\mathrm{loc}}(\mathscr{T})$, then there exists $\xi \in L^\infty_{\mathrm{loc}}(\mathscr{T})$ such that*

$$\left\| \zeta J_{\boldsymbol{\varepsilon}}^{\boldsymbol{\tau}} f \right\|_p \leqslant \| \xi f \|_p$$

for all $f \in L^p_{\mathrm{loc}}(\mathscr{E})$ and $\boldsymbol{\varepsilon} \in E$; if $\zeta \in L^\infty_{\mathrm{c}}$, then we may choose $\xi \in L^\infty_{\mathrm{c}}$.

(iv) (Approximation) *If* $p < \infty$ *and* $f \in L^p_{\mathrm{loc}}(\mathscr{E})$, *then*

$$\limsup_{\boldsymbol{\varepsilon}\to\mathbf{0}} \left\| \zeta \left(J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f - f \right) \right\|_p = 0$$

for all $\zeta \in L^\infty_{\mathrm{loc}}(\mathscr{T})$; *the same holds if* $p = \infty$ *and* $f \in C(\mathscr{E})$.

(v) (Upper bound) *If* $f \in L^\infty_{\mathrm{loc}}(\mathscr{E})$, *then, for each continuous fibre seminorm* P *on* $\mathscr{E}$ *and closed subset* C *of* M,

$$\limsup_{\boldsymbol{\varepsilon}\to\mathbf{0}} \sup_{x\in C} P\left(J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f\right)(x) \leqslant \inf_{\substack{W\in\mathfrak{O}\\ W\supseteq C}} \operatorname{ess\,sup}_{x\in W} P(f)(x). \tag{15}$$

Further, the "commutators" $[D, J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}}]$, *defined by*

$$\left[D, J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}}\right] f = D J^{\mathscr{E},\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f - J^{\mathscr{F},\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} Df,$$

have the following properties.

(vi) (Equicontinuity) *If* $\zeta \in L^\infty_{\mathrm{loc}}(\mathscr{T})$, *then there exists* $\xi \in L^\infty_{\mathrm{loc}}(\mathscr{T})$ *such that*

$$\left\| \zeta \left[D, J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}}\right] f \right\|_p \leqslant \|\xi f\|_p$$

for all $f \in L^p_{\mathrm{loc}}(\mathscr{E})$ *and* $\boldsymbol{\varepsilon} \in E$; *if* $\zeta \in L^\infty_{\mathrm{c}}$, *then we may choose* $\xi \in L^\infty_{\mathrm{c}}$.

(vii) (Vanishing) *If* $f \in L^p_{\mathrm{loc}}(\mathscr{E})$ *and* $p < \infty$, *then*

$$\limsup_{\boldsymbol{\varepsilon}\to\mathbf{0}} \left\| \zeta \left[D, J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}}\right] f \right\|_p = 0$$

for all $\zeta \in L^\infty_{\mathrm{loc}}(\mathscr{T})$; *the same holds if* $p = \infty$ *and* $f \in C(\mathscr{E})$.

As operators on $W^p_{D,\mathrm{loc}}(\mathscr{E})$, *the* $J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}}$ *have the following properties.*

(viii) (Equicontinuity) *If* $\zeta \in L^\infty_{\mathrm{loc}}(\mathscr{T})$, *then there exists* $\xi \in L^\infty_{\mathrm{loc}}(\mathscr{T})$ *such that*

$$\left\| \zeta D J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f \right\|_p \leqslant \|\xi f\|_p + \|\xi Df\|_p$$

for all $f \in W^p_{D,\mathrm{loc}}(\mathscr{E})$ *and* $\boldsymbol{\varepsilon} \in E$; *if* $\zeta \in L^\infty_{\mathrm{c}}$, *then we may choose* $\xi \in L^\infty_{\mathrm{c}}$.

(ix) (Approximation) *If* $f \in W^p_{D,\mathrm{loc}}(\mathscr{E})$ *and* $p < \infty$, *then*

$$\limsup_{\boldsymbol{\varepsilon}\to\mathbf{0}} \left\| \zeta \left(D J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f - Df \right) \right\|_p = 0$$

for all $\zeta \in L^\infty_{\mathrm{loc}}(\mathscr{T})$; *the same holds if* $p = \infty$, $f \in C(\mathscr{E})$, *and* $Df \in C(\mathscr{F})$.

(x) (Upper bound) *If* $f \in C(\mathscr{E})$ *and* $Df \in L^\infty_{\mathrm{loc}}(\mathscr{F})$, *then, for each continuous fibre seminorm* P *on* $\mathscr{F}$ *and closed subset* C *of* M,

$$\limsup_{\boldsymbol{\varepsilon}\to\mathbf{0}} \sup_{x\in C} P\left(D J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f\right)(x) \leqslant \inf_{\substack{W\in\mathfrak{O}\\ W\supseteq C}} \operatorname{ess\,sup}_{x\in W} P(Df)(x).$$

The alert reader will have already noticed that, in contrast to Propositions 3.1 and 3.2, equicontinuity and the limiting properties here refer to a topology on the spaces of sections L^p_{loc} (and C), defined by the "extended seminorms" $f \mapsto \|\zeta f\|_p$, where ζ ranges over $L^\infty_{\mathrm{loc}}(\mathscr{T})$, which is finer than the usual Fréchet topology when M is not compact. Indeed, if M is not compact, then the finer topology, known as the Whitney topology (at least in the case of C [15, Chap. 2]), is not metrisable, nor does it yield a topological vector space structure: the mapping $\lambda \mapsto \lambda f$ is not continuous unless the section f is compactly-supported. However, like the Fréchet topology, the Whitney topology is independent of the measure on M and the hermitean structure of the bundle.

Propositions 3.1 and 3.2 cannot be strengthened by just replacing the Fréchet topology with the Whitney topology; the stronger approximation result of Theorem 3.3 is due to the fact that the approximant $J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f$ depends on the sequence $\boldsymbol{\varepsilon} \in E$ whose components ε_α may be chosen independently.

A propos of limits along E, the following remark will be useful in the course of the proof: if $\{A_\beta\}_{\beta \in B}$ is a collection of subsets of A which is locally finite, in the sense that $\{\beta \in B : \alpha \in A_\beta\}$ is finite for all $\alpha \in A$, then

$$\limsup_{\boldsymbol{\varepsilon}\to\mathbf{0}} \sup_{\beta\in B} F_\beta(\boldsymbol{\varepsilon}|_{A_\beta}) = \sup_{\beta\in B} \limsup_{\boldsymbol{\varepsilon}\to\mathbf{0}} F_\beta(\boldsymbol{\varepsilon}|_{A_\beta}) \tag{16}$$

for all functions $F_\beta :]0,1]^{A_\beta} \to [-\infty, \infty]$.

Proof First, the sum defining $J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}}$ is a locally finite sum of smooth, compactly-supported sections of $\mathscr{E}$, so part (i) clearly holds.

Next, for all closed subsets C of M, open neighbourhoods W of C, and $\alpha \in A$, we may find $\bar{\varepsilon}_\alpha \in]0,1]$ such that

$$B^-_{\mathbb{R}^n}\big(\varphi_\alpha(\operatorname{supp}\eta_\alpha \cap C), \bar{\varepsilon}_\alpha\big) \subseteq \varphi_\alpha(W),$$

and part (ii) follows from part (i) of Proposition 3.1.

Further, if $\zeta \in L^\infty_{\mathrm{loc}}(\mathscr{T})$, then

$$\big|\zeta J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f\big| \leqslant \sum_\alpha \kappa_\alpha \big|J_{\varepsilon_\alpha} \tau_\alpha(\eta_\alpha f)\big|$$

pointwise almost everywhere, where the constants κ_α are independent of f. We deduce from part (ii) of Proposition 3.1 that

$$\big\|J_{\varepsilon_\alpha}\tau_\alpha(\eta_\alpha f)\big\|_p \leqslant \kappa'_\alpha \big\|\tau_\alpha(\eta_\alpha f)\big\|_p$$

and hence

$$\big\|\zeta J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f\big\|_p \leqslant \sum_\alpha \kappa''_\alpha \|\eta_\alpha f\|_p \leqslant \sup_{\alpha\in A} \kappa'''_\alpha \|\eta_\alpha f\|_p$$

for new constants κ'_α, κ''_α and κ'''_α, and we take ξ to be $\sup_{\alpha\in A}\kappa'''_\alpha\eta_\alpha$ to prove part (iii). Analogously, one shows that

$$\left\|\zeta\left(J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f - f\right)\right\|_p \leqslant \sup_{\alpha\in A}\kappa''''_\alpha \left\| J_{\varepsilon_\alpha}\tau_\alpha(\eta_\alpha f) - \tau_\alpha(\eta_\alpha f)\right\|_p$$

for suitable constants κ''''_α, whence

$$\limsup_{\boldsymbol{\varepsilon}\to\mathbf{0}} \left\|\zeta\left(J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f - f\right)\right\|_p \leqslant \sup_{\alpha\in A}\kappa_\alpha \limsup_{t\to 0}\left\| J_t\tau_\alpha(\eta_\alpha f) - \tau_\alpha(\eta_\alpha f)\right\|_p$$

by (16), and part (iv) follows from part (iii) of Proposition 3.1.

Suppose now that $f \in L^\infty_{\mathrm{loc}}(\mathscr{E})$ and P is a continuous fibre seminorm on $\mathscr{E}$. Write P_α for the corresponding seminorm on the fibres of the trivial bundle $\mathbb{R}^n \times \mathbb{C}^r$ over $\mathbb{R}^n$; in other words, $P_\alpha(\tau_\alpha f)(\varphi(x)) = P(f)(x)$ for all $x \in U_\alpha$ for a section f of $\mathscr{E}$ with support in U_α. Given any $\alpha \in A$, write K_α for $\varphi_\alpha^{-1}(B^-_{\mathbb{R}^n}(0,2))$, so $\operatorname{supp}\tau_\alpha^{-1}J_t\tau_\alpha(\eta f) \subseteq K_\alpha$ when $t \leqslant 1$, and given any $\beta \in A$, denote by A_β the finite set of indices α in A such that $K_\alpha \cap K_\beta \neq \emptyset$.

Fix $\delta \in \mathbb{R}^+$. Given any $\beta \in A$, we may find a finite decomposition of K_β as $K_{\beta,1} \cup \dots \cup K_{\beta,k_\beta}$, in which each $K_{\beta,j}$ is compact and the oscillation of η_α on an open neighbourhood $W_{\beta,j}$ of $K_{\beta,j}$ is bounded by $\delta/|A_\beta|$. Take $y_{\beta,j}$ in $K_{\beta,j}$. Then, by part (iv) of Proposition 3.1, for all closed subsets C of M and open neighbourhoods W of C,

$$\begin{aligned}
\limsup_{\boldsymbol{\varepsilon}\to\mathbf{0}} \sup_{x\in K_{\beta,j}\cap C} P\left(J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f\right)(x) &\leqslant \sum_{\alpha\in A_\beta} \limsup_{t\to 0} \sup_{x\in K_{\beta,j}\cap C} P_\alpha\left(J_t\tau_\alpha(\eta_\alpha f)\right)\left(\varphi_\alpha(x)\right) \\
&\leqslant \sum_{\alpha\in A_\beta} \operatorname*{ess\,sup}_{x\in W_{\beta,j}\cap W} P_\alpha\left(\tau_\alpha(\eta_\alpha f)\right)\left(\varphi_\alpha(x)\right) \\
&\leqslant \sum_{\alpha\in A_\beta} \sup_{x\in W_{\beta,j}} \eta_\alpha(x) \operatorname*{ess\,sup}_{z\in W} P(f)(z) \\
&\leqslant \sum_{\alpha\in A_\beta} \left(\eta_\alpha(y_{\beta,j}) + \delta/|A_\beta|\right) \operatorname*{ess\,sup}_{z\in W} P(f)(z) \\
&\leqslant (1+\delta) \operatorname*{ess\,sup}_{z\in W} P(f)(z)
\end{aligned}$$

for each $K_{\beta,j}$. Since the restriction of $P(J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f)$ to $K_{\beta,j}$ depends only on $\boldsymbol{\varepsilon}|_{A_\beta}$, and the set $\{(\beta,j) : \alpha \in A_\beta\} = \bigcup_{\beta\in A_\alpha}\{\beta\}\times\{1,\dots,k_\beta\}$ is finite for all $\alpha \in A$,

$$\begin{aligned}
\limsup_{\boldsymbol{\varepsilon}\to\mathbf{0}} \sup_{x\in C} P\left(J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f\right)(x) &= \sup_{\substack{\beta\in A\\ j=1,\dots,k_\beta}} \limsup_{\boldsymbol{\varepsilon}\to\mathbf{0}} \sup_{x\in K_{\beta,j}\cap C} P\left(J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f\right)(x) \\
&\leqslant (1+\delta) \operatorname*{ess\,sup}_{z\in W} P(f)(z),
\end{aligned}$$

by (16), and part (v) follows from the arbitrariness of δ and W.

We now write $D \in \mathfrak{D}_1(\mathscr{E}, \mathscr{F})$ in local coordinates, and decompose $[D, J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}}]$ as $I^1_{\boldsymbol{\varepsilon}} + I^2_{\boldsymbol{\varepsilon}}$, where

$$I^1_{\boldsymbol{\varepsilon}} f = \sum_{\alpha} \tau_\alpha^{-1} J_{\varepsilon_\alpha} \tau_\alpha \big((D^\sigma \eta_\alpha) f \big)$$

$$I^2_{\boldsymbol{\varepsilon}} f = \sum_{\alpha} \tau_\alpha^{-1} \big[\tau_\alpha(D), J_{\varepsilon_\alpha} \big] \tau_\alpha(\eta_\alpha f).$$

The properties (vi) and (vii) of $[D, J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}}]$ follow from the analogous properties of $I^1_{\boldsymbol{\varepsilon}}$ and $I^2_{\boldsymbol{\varepsilon}}$, which in turn are obtained from Propositions 3.1 and 3.2, by arguing as in the proofs of parts (iii) and (iv) of this theorem and observing that

$$\sum_{\alpha} (D^\sigma \eta_\alpha) f = \Big(D^\sigma \sum_{\alpha} \eta_\alpha \Big) f = 0.$$

Finally, the decomposition

$$D J^{\mathscr{E}, \boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f = \big[D, J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} \big] f + J^{\mathscr{F}, \boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} Df$$

shows that part (viii) follows from parts (iii) and (vi), while part (ix) follows from parts (iv) and (vii). Moreover, given any continuous fibre seminorm P on $\mathscr{F}$,

$$\big| P\big(D J^{\mathscr{E}, \boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f \big) - P\big(J^{\mathscr{F}, \boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} Df \big) \big| \leqslant P\big(\big[D, J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} \big] f \big),$$

and, by part (vii), the right-hand side tends to 0 uniformly as $\boldsymbol{\varepsilon} \to \mathbf{0}$ whenever f is continuous; therefore, under our assumptions,

$$\limsup_{\boldsymbol{\varepsilon} \to \mathbf{0}} \sup_{x \in C} P\big(D J^{\mathscr{E}, \boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f \big)(x) = \limsup_{\boldsymbol{\varepsilon} \to \mathbf{0}} \sup_{x \in C} P\big(J^{\mathscr{F}, \boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} Df \big)$$

for all subsets C of M, and part (x) follows from part (v). □

Not only is the Whitney topology finer than the Fréchet space topology on L^p_{loc}, but also, when restricted to L^p, it is finer than the usual Banach space topology of L^p. Hence the following density result is an immediate consequence of Theorem 3.3.

Corollary 3.4 *Suppose that $D \in \mathfrak{D}_1(\mathscr{E}, \mathscr{F})$ and $1 \leqslant p < \infty$. Then $C^\infty_{\mathrm{c}}(\mathscr{E})$ is dense in $W^p_{D,\mathrm{loc}}(\mathscr{E})$, and $W^p_D \cap C^\infty(\mathscr{E})$ is dense in $W^p_D(\mathscr{E})$.*

An analogous result when $p = \infty$ may be obtained by restricting to continuous sections with continuous D-derivatives. The following weaker result, however, does not require the continuity of the D-derivatives.

Corollary 3.5 *Suppose that $D \in \mathfrak{D}_1(\mathscr{E}, \mathscr{F})$, $f \in C(\mathscr{E})$, $Df \in L^\infty_{\mathrm{loc}}(\mathscr{F})$, and, for some continuous fibre seminorm P on $\mathscr{F}$, $\|P(Df)\|_\infty \leqslant 1$. Then there exists a sequence of $C^\infty(\mathscr{E})$-sections f_m that converges to f uniformly on compacta, such* *that* $\|P(Df_m)\|_\infty \leqslant 1$ *for all* m *and* $\operatorname{supp} f_m \subseteq W$ *for all open neighbourhoods* W *of* $\operatorname{supp} f$ *once* m *is large enough. Moreover, if* $\mathscr{E} = \mathscr{T}$ *and* f *is real-valued, then the* f_m *may be chosen to be real-valued.*

Proof By parts (iv) and (x) of Theorem 3.3, $\limsup_{\boldsymbol{\varepsilon}\to\mathbf{0}} \|J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f - f\|_\infty = 0$ and

$$\limsup_{\boldsymbol{\varepsilon}\to\mathbf{0}} \left\| P\big(DJ^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f\big)\right\|_\infty \leqslant \left\|P(Df)\right\|_\infty \leqslant 1.$$

We fix a decreasing countable base $\{W_m\}_{m\in\mathbb{N}}$ of open neighbourhoods of $\operatorname{supp} f$, and then choose, for all $m \in \mathbb{N}$, a sequence $\boldsymbol{\varepsilon}$ in E such that the section $g_m = J^{\boldsymbol{\tau}}_{\boldsymbol{\varepsilon}} f$ satisfies $\|g_m - f\|_\infty \leqslant 2^{-m}$ and $\|P(Dg_m)\|_\infty \leqslant 1 + 2^{-m}$. We may also assume that $\operatorname{supp} g_m \subseteq W_m$ by part (ii) of Theorem 3.3. The conclusion then follows by taking $f_m = (1 + 2^{-m})^{-1} g_m$. □

3.2 Integration and Differentiation

Approximation using mollifiers allows us to extend results such as integration by parts, Leibniz' rule, and the chain rule to the realm of weakly differentiable sections (see, for instance, [13, Chap. 7]). In what follows, p' denotes the index conjugate to p, that is, $1/p' + 1/p = 1$.

Proposition 3.6 (Integration by parts) *Suppose that* $D \in \mathfrak{D}_1(\mathscr{E}, \mathscr{F})$ *and that* $1 \leqslant p \leqslant \infty$*. If* $f \in W^p_{D,\mathrm{loc}}(\mathscr{E})$ *and* $g \in W^{p'}_{D^+,\mathrm{loc}}(\mathscr{F})$*, and moreover* $\operatorname{supp}(f \otimes g)$ *is compact, then*

$$\langle\!\langle Df, g\rangle\!\rangle = \big\langle\!\big\langle f, D^+ g\big\rangle\!\big\rangle.$$

Proof By exchanging f with g and D with D^+ if necessary, we may suppose that $p < \infty$.

Take a bump function η equal to 1 on $\operatorname{supp}(f \otimes g)$. By Corollary 3.4, there exists a sequence of $C^\infty_{\mathrm{c}}(\mathscr{E})$-sections f_m such that $(f_m, Df_m) \to (f, Df)$ in $L^p_{\mathrm{loc}}(\mathscr{E} \oplus \mathscr{F})$, and then $(\eta f_m, D(\eta f_m)) \to (\eta f, D(\eta f))$ in $L^p_{\mathrm{c}}(\mathscr{E} \oplus \mathscr{F})$. Now

$$D(\eta f) = \big(D^\sigma \eta\big) f + \eta Df$$

by Leibniz' rule (11) for a smooth function η and a distribution f, and moreover $D^\sigma \eta$ vanishes on $\operatorname{supp}(f \otimes g)$, so $\langle\!\langle (D^\sigma \eta) f, g\rangle\!\rangle = 0$. Hence (9) implies that

$$\begin{aligned}\langle\!\langle Df, g\rangle\!\rangle &= \big\langle\!\big\langle D(\eta f), g\big\rangle\!\big\rangle = \lim_m \big\langle\!\big\langle D(\eta f_m), g\big\rangle\!\big\rangle \\ &= \lim_m \big\langle\!\big\langle \eta f_m, D^+ g\big\rangle\!\big\rangle = \big\langle\!\big\langle \eta f, D^+ g\big\rangle\!\big\rangle \\ &= \big\langle\!\big\langle f, D^+ g\big\rangle\!\big\rangle,\end{aligned}$$

as required. □

Proposition 3.7 (Leibniz' rule) *Suppose that $D \in \mathfrak{D}_1(\mathscr{E}, \mathscr{F})$ and $1 \leqslant p \leqslant \infty$. Suppose also that $h \in W^p_{D^\sigma,\mathrm{loc}}(\mathscr{T})$ and $f \in W^{p'}_{D,\mathrm{loc}}(\mathscr{E})$. Then $hf \in W^1_{D,\mathrm{loc}}(\mathscr{E})$ and*

$$D(hf) = \big(D^\sigma h\big) f + hDf.$$

Proof By Hölder's inequality, $hf \in L^1_{\mathrm{loc}}(\mathscr{E})$ and $(D^\sigma h) f + hDf \in L^1_{\mathrm{loc}}(\mathscr{F})$. We must show that the $\mathscr{F}$-valued distributions $D(hf)$ and $(D^\sigma h) f + hDf$ coincide. In fact, for all $\varphi \in C^\infty_{\mathrm{c}}(\mathscr{F})$,

$$\langle\!\langle D(hf), \varphi\rangle\!\rangle = \langle\!\langle hf, D^+\varphi\rangle\!\rangle = \langle\!\langle f, \bar{h} D^+\varphi\rangle\!\rangle.$$

Leibniz' rule (11) for a smooth section φ and a distribution $\bar{h}$, together with (10), leads to the distributional equality

$$D^+(\bar{h}\varphi) = -\big(D^\sigma h\big)^*\varphi + \bar{h} D^+\varphi.$$

By the hypotheses, each summand in the right-hand side lies in $L^p_{\mathrm{c}}(\mathscr{E})$, therefore $D^+(\bar{h}\varphi) \in L^p_{\mathrm{c}}(\mathscr{E})$ too, and

$$\langle\!\langle D(hf), \varphi\rangle\!\rangle = \langle\!\langle f, D^+(\bar{h}\varphi) + \big(D^\sigma h\big)^*\varphi\rangle\!\rangle = \langle\!\langle hDf + \big(D^\sigma h\big) f, \varphi\rangle\!\rangle$$

by Proposition 3.6, since $\bar{h}\varphi$ is compactly-supported and in $L^p_{\mathrm{loc}}(\mathscr{F})$. □

Proposition 3.8 (Chain rule) *Suppose that $D \in \mathfrak{D}_1(\mathscr{T}, \mathscr{E})$ is homogeneous. If $1 \leqslant p < \infty$ and $f \in W^p_{D,\mathrm{loc}}(\mathscr{T})$ is real-valued, and the function $g \colon \mathbb{R} \to \mathbb{C}$ is continuously differentiable and g' is bounded, then $g \circ f \in W^p_{D,\mathrm{loc}}(\mathscr{T})$ and*

$$D(g \circ f) = \big(g' \circ f\big) Df.$$

Proof By Corollary 3.4, there exists a sequence of $C^\infty_{\mathrm{c}}(\mathscr{T}_{\mathbb{R}})$-sections f_m such that $(f_m, Df_m) \to (f, Df)$ in $L^p_{\mathrm{loc}}(\mathscr{T} \oplus \mathscr{E})$; by extracting a subsequence, we may suppose that the convergence is also pointwise almost everywhere. Now g is Lipschitz and $f_m \to f$ in $L^p_{\mathrm{loc}}(\mathscr{T})$, and so $g \circ f_m \to g \circ f$ in $L^p_{\mathrm{loc}}(\mathscr{T})$. Moreover, by the chain rule for C^1-functions,

$$D(g \circ f_m) = \big(g' \circ f_m\big) Df_m = \big(g' \circ f_m\big)(Df_m - Df) + \big(g' \circ f_m\big) Df.$$

Since $\|g' \circ f_m\|_\infty \leqslant \|g'\|_\infty < \infty$, the first summand converges to 0 in $L^p_{\mathrm{loc}}(\mathscr{E})$; moreover, since g' is continuous, $g' \circ f_m \to g' \circ f$ pointwise almost everywhere, and therefore the second summand converges to $(g' \circ f) Df$ in $L^p_{\mathrm{loc}}(\mathscr{E})$ by the dominated convergence theorem. Thus $g \circ f_m \to g \circ f$ in $W^p_{D,\mathrm{loc}}(\mathscr{T})$, and the conclusion follows. □

4 Reversible Sub-Finsler Geometry

Suppose that P is a continuous fibre seminorm on T^*M, and P^* is the dual extended fibre norm on TM. Thus

$$P_x(\xi) = \sup_{\substack{v \in T_x M \\ P^*(v) \leqslant 1}} \left|\xi(v)\right| \quad \text{and} \quad P_x^*(v) = \sup_{\substack{\omega \in T_x^* M \\ P(\omega) \leqslant 1}} \left|\xi(v)\right|$$

by the finite-dimensional Hahn–Banach theorem.

As a function on the tangent bundle, P^* need not be continuous. However, it may be approximated by continuous fibre norms on TM, as we are about to show.

Lemma 4.1 *There exists a countable family $\mathfrak{G}_P$ of Riemannian metrics on M such that*

$$P^*(v) = \sup_{g \in \mathfrak{G}_P} |v|_g \quad \textit{for every } v \in TM \tag{17}$$

or, equivalently,

$$P(\xi) = \inf_{g \in \mathfrak{G}_P} |\xi|_g \quad \textit{for every } \xi \in T^*M. \tag{18}$$

Proof Recall that a Riemannian metric on M is given by a smooth fibre inner product on TM, or, by duality, by a smooth fibre inner product on T^*M. From a geometric point of view, proving (17) amounts to realising the closed unit ball of P^* at a point $x \in M$ (which is convex but may have no interior) as the intersection of the closed unit balls of the metrics g in $\mathfrak{G}_P$, which are ellipsoids, and proving (18) amounts to realising the open unit ball of P at a point $x \in M$ (which is convex, but may be unbounded) as the union of the open unit balls of the metrics g, which are also ellipsoids. In general, this may require an infinite number of ellipsoids, as we may see by considering the problem of realising a square as an intersection or union of g balls. We consider the cotangent space problem only.

It is easy to show that a Riemannian metric that satisfies

$$P(\xi) \leqslant |\xi|_g \quad \text{for every } \xi \in T^*M \tag{19}$$

exists. Indeed, if g is a Riemannian metric on M, then the function

$$x \mapsto \sup_{\substack{\xi \in T_x^* M \\ |\xi|_g \leqslant 1}} P(\xi)$$

is locally finite, therefore it is majorised by a strictly positive function $\psi \in C^\infty(\mathscr{T})$, and one simply needs to rescale g by ψ^2.

Take a Riemannian metric g on M satisfying (19) and the countable atlas $(\varphi_\alpha)_{\alpha \in A}$. Recall that each φ_α maps U_α to $\mathbb{R}^n$, that $V_\alpha = \varphi_\alpha^{-1}(B_{\mathbb{R}^n}(0, 1))$, and that $M = \bigcup_{\alpha \in A} V_\alpha$. Each subbundle T^*U_α of T^*M is trivialisable.

Fix $\alpha \in A$, and choose a bump function ζ_α with compact support in U_α that is equal to 1 on V_α and a countable set $\mathscr{Y}_\alpha$ of smooth sections of T^*U_α such that

$$\{\omega(x) : \omega \in \mathscr{Y}_\alpha\}^- = \{\xi \in T_x^*M : |\xi|_g = 1\} \tag{20}$$

for all $x \in U_\alpha$. To do this, it is sufficient to consider constant sections taking values in a countable dense subset of the unit sphere with respect to a trivialisation of T^*U_α given by a g-orthonormal frame.

Next, fix $\omega \in \mathscr{Y}_\alpha$. Since $P(\omega)$ is a continuous nonnegative function on U_α, there is a sequence of smooth functions $\psi_{\omega,k} : U_\alpha \to \mathbb{R}$ such that

$$P(\omega) + 2^{-k} \leqslant \psi_{\omega,k} \leqslant P(\omega) + 2^{1-k}.$$

We now define, for all $k \in \mathbb{N}$, a smooth inner product $(\cdot,\cdot)_{\alpha,\omega,k}$ and associated norm $|\cdot|_{\alpha,\omega,k}$ along the fibres of T^*U_α by

$$\begin{aligned}(\xi_1, \xi_2)_{\alpha,\omega,k} &= \psi_{\omega,k}(x)^2 \big(\langle \pi(\xi_1), \pi(\xi_2)\rangle_g + 2^{2k}\langle \xi_1 - \pi(\xi_1), \xi_2 - \pi(\xi_2)\rangle_g\big) \\ &= \psi_{\omega,k}(x)^2 \big(\langle \xi_1, \omega(x)\rangle_g \langle \xi_2, \omega(x)\rangle_g \\ &\quad + 2^{2k}\big(\langle \xi_1, \xi_2\rangle_g - \langle \xi_1, \omega(x)\rangle_g \langle \xi_2, \omega(x)\rangle_g\big)\big)\end{aligned}$$

for all $\xi_1, \xi_2 \in T_x^*M$ and $x \in U_\alpha$, where $\pi(\xi)$ is the projection of ξ in T_xM onto $\mathbb{R}\omega(x)$, that is, $\pi(\xi) = \langle \xi, \omega(x)\rangle_g \omega(x)$. Now

$$\big|\pi(\xi)\big|_g \leqslant \frac{|\xi|_{\alpha,\omega,k}}{\psi_{\omega,k}(x)} \quad \text{and} \quad \big|\xi - \pi(\xi)\big|_g \leqslant \frac{|\xi|_{\alpha,\omega,k}}{2^k \psi_{\omega,k}(x)}$$

for all $\xi \in T_x^*M$ and $x \in U_\alpha$, and so, from (19),

$$\begin{aligned}P(\xi) &\leqslant P\big(\pi(\xi)\big) + P\big(\xi - \pi(\xi)\big) \\ &\leqslant \big|\langle \xi, \omega(x)\rangle_g\big| P\big(\omega(x)\big) + \big|\xi - \pi(\xi)\big|_g \\ &\leqslant \frac{|\xi|_{\alpha,\omega,k}}{\psi_{\omega,k}(x)} \big(P\big(\omega(x)\big) + 2^{-k}\big) \\ &\leqslant |\xi|_{\alpha,\omega,k}.\end{aligned}$$

Hence

$$P(\xi) \leqslant \inf_{\substack{k\in\mathbb{N} \\ \omega \in \mathscr{Y}_\alpha}} |\xi|_{\alpha,\omega,k}. \tag{21}$$

Moreover, for all $x \in U_\alpha$, from the definitions of $\psi_{\omega,k}$ and $(\cdot,\cdot)_{\alpha,\omega,k}$,

$$P\big(\omega(x)\big) + 2^{1-k} \geqslant \psi_{\omega,k}(x) = \big|\omega(x)\big|_{\alpha,\omega,k}.$$

More generally, for all $\xi \in T_x^*U_\alpha$ such that $|\xi|_g = 1$ and all $k \in \mathbb{N}$, we may choose $\omega \in \mathscr{Y}_\alpha$ such that $|\xi - \omega(x)|_g \leqslant 2^{-2k}$, by (20). Then by construction,

$|\xi - \omega(x)|_{\alpha,\omega,k} \leqslant 3 \times 2^{-2k}$ and $P(\xi - \omega(x)) \leqslant 2^{-2k}$, hence

$$P(\xi) \geqslant |\xi|_{\alpha,\omega,k} - 2^{1-k} - 3 \times 2^{-k} - 2^{-2k}$$

and the reverse of inequality (21) follows. Putting everything together, we deduce that

$$P(\xi) = \inf_{\substack{k\in\mathbb{N}\\ \omega\in\mathscr{Y}_\alpha}} |\xi|_{\alpha,\omega,k}$$

for all $x \in U_\alpha$ and $\xi \in T_x^* M$.

For each $\alpha \in A$, choose a bump function ζ_α with compact support in U_α that is equal to 1 on V_α, and define a Riemannian metric $g_{\alpha,\omega,k}$ on M by setting

$$\langle \xi, \xi \rangle_{g_{\alpha,\omega,k}} = \zeta_\alpha(x)(\xi, \xi)_{\alpha,\omega,k} + \bigl(1 - \zeta_\alpha(x)\bigr)\langle \xi, \xi \rangle_g$$

for all $x \in M$ and $\xi \in T_x^* M$; the first summand is defined to vanish whenever $x \notin U_\alpha$. Then clearly $P(\xi) \leqslant |\xi|_{g_{\alpha,\omega,k}}$ for all $\xi \in T^* M$. Moreover, if $x \in M$, then $x \in V_\alpha$ for some α, therefore $|\xi|_{g_{\alpha,\omega,k}} = |\xi|_{\alpha,\omega,k}$ for all $\xi \in T_x^* M$. We now set

$$\mathfrak{G}_P = \{g_{\alpha,\omega,k} : \omega \in \mathscr{Y}_\alpha, k \in \mathbb{N}, \alpha \in A\},$$

and the desired conclusion follows. □

Define the finite subspace of P^* in TM and the zero subspace of P in T^*M by

$$F\bigl(P_x^*\bigr) = \bigl\{v \in T_x M : P^*(v) < \infty\bigr\} \quad \text{and} \quad Z(P_x) = \bigl\{\xi \in T_x^* P : P(\xi) = 0\bigr\}.$$

Then $F(P_x^*)$ is the annihilator of $Z(P_x)$, so $\dim F(P_x^*) = \operatorname{codim} Z(P_x)$, and the function $x \mapsto \dim F(P_x^*)$ is lower-semicontinuous. When this function is continuous, that is, when it is locally constant, P^* has additional continuity properties. We define $F(P^*) = \bigcup_{x\in M} F(P_x^*)$.

Proposition 4.2 *Suppose that $x \mapsto \dim F(P_x^*)$ is continuous. Then $F(P^*)$ is closed in TM, and P^* restricted to $F(P^*)$ is continuous.*

Proof Without loss of generality, we suppose that M is connected, so the function $x \mapsto \dim F(P_x^*)$ is constant, that is, $\operatorname{codim} F(P_x^*) = k$ for some k and all $x \in M$.

Write G for the kth grassmannian bundle over T^*M, so G_x is the set of k-dimensional subspaces of T_x^*M, and define $X = \{S \in G : P|_S = 0\}$. Then X is closed in G, because P is continuous, and $X \cap G_x = \{Z(P_x)\}$ for all $x \in M$. Since G has compact fibres, X is the image of a continuous section of G, and this section may be lifted locally to a continuous section of the frame bundle of T^*M. Thus there is a continuous local frame $\{\omega_1, \dots, \omega_n\}$ for T^*M in the neighbourhood of each point of M such that $Z(P_x) = \operatorname{span}\{\omega_1|_x, \dots, \omega_k|_x\}$, whence

$$F\bigl(P_x^*\bigr) = \ker \omega_1|_x \cap \dots \cap \ker \omega_k|_x = \operatorname{span}\bigl\{\omega_{k+1}^*\big|_x, \dots, \omega_n^*\big|_x\bigr\},$$

where $\{\omega_1^*, \dots, \omega_n^*\}$ is the dual local frame for TM. This proves that $F(P^*)$ is closed in TM, and determines a continuous subbundle E of TM.

Denote by $\iota^* \colon T^*M \to E^*$ the pointwise transpose of the inclusion map $\iota \colon E \to TM$. Then $Z(P_x) = \ker \iota^*|_{T_x^*M}$, hence P induces a continuous fibre norm Q on E^* such that $Q \circ \iota^* = P$. It is then easily checked that the restriction of P^* to E is the dual norm of Q pointwise.

By the use of local trivialisations of E^*, we may find, for all $x \in M$, a neighbourhood U of x and linear isomorphisms $t_y \colon E_x^* \to E_y^*$ for all $y \in U$ such that the mapping $(y, \xi) \mapsto t_y(\xi)$ is continuous from $U \times E_x^*$ to E^*. The continuity of Q and the compactness of the unit sphere of Q_x in E_x^* then imply that, for all positive ε, there is a neighbourhood V of x in M such that, for all $y \in V$,

$$(1+\varepsilon)^{-1} Q_y \circ t_y \leqslant Q_x \leqslant (1+\varepsilon) Q_y \circ t_y$$

and correspondingly

$$(1+\varepsilon)^{-1} P_y^*|_{F(P_y^*)} \leqslant P_x^*|_{F(P_x^*)} \circ t_y^* \leqslant (1+\varepsilon) P_y^*|_{F(P_y^*)}.$$

This proves the continuity of $P^*|_{F(P^*)}$. □

Definition 4.3 A tangent vector $v \in T_xM$ is said to be P-subunit if $P^*(v) \leqslant 1$.

Definition 4.4 We write $\Gamma^k([a,b])$ for the set of all curves $\gamma \colon [a,b] \to M$ of class C^k; here k may be ∞. A curve $\gamma \colon [a,b] \to M$ is said to be P-subunit if it is absolutely continuous and $\gamma'(t)$ is P-subunit for almost all $t \in [a,b]$. We write $\Gamma_P^{\mathrm{unit}}(I)$ for the set of all P-subunit curves defined on the interval I, and Γ_P^{unit} for the set of all P-subunit curves when the interval of definition may vary. We write ϱ_P for the distance function induced by P, that is, $\varrho_P(x,y)$ is the infimum of the set of all $T \in \mathbb{R}^+$ for which there exists $\gamma \in \Gamma_P^{\mathrm{unit}}([0,T])$ such that $\gamma(0) = x$ and $\gamma(T) = y$.

The infimum need not be attained: for instance, in $\mathbb{R}^2 \setminus \{(0,0)\}$ with the euclidean metric, there is no minimising curve joining $(-1,0)$ and $(1,0)$.

Absolute continuity may be defined in various equivalent ways: here is one.

Definition 4.5 The curve $\gamma \colon [a,b] \to M$ is said to be *absolutely continuous* if $\varphi_\alpha \circ \gamma$ is locally absolutely continuous for each φ_α in the atlas A. We write $AC([a,b])$ for the set of all absolutely curves on the interval $[a,b]$.

Suppose that P is a norm induced by a Riemannian metric g on M and ϱ_g is the distance function induced by g. If $\gamma \in \Gamma_P^{\mathrm{unit}}([a,b])$, then

$$\varrho_g\big(\gamma(s), \gamma(t)\big) \leqslant |s-t| \quad \text{for every } s,t \in [a,b]. \tag{22}$$

Conversely, a curve $\gamma \colon [a,b] \to M$ that satisfies (22) is P-subunit: the derivative $\gamma'(t)$ may be computed in exponential coordinates centred at $\gamma(t)$, and the difference quotient is controlled by the Lipschitz constant.

Thanks to Lemma 4.1, a similar result may be proved for an arbitrary fibre seminorm P. We use the family $\mathfrak{G}_P$ of Riemannian metrics defined in Lemma 4.1.

Proposition 4.6 *The function ϱ_P is an extended distance function on M,*

$$\varrho_g(x, y) \leqslant \varrho_P(x, y) \tag{23}$$

for all $x, y \in M$ and $g \in \mathfrak{G}_P$, and the topology induced by ϱ_P is at least as fine as the manifold topology. Further, for a function $\gamma : [a, b] \to M$, the following conditions are equivalent:

(i) *γ is a P-subunit curve*;
(ii) *$\varrho_P(\gamma(s), \gamma(t)) \leqslant |s - t|$ for all $s, t \in [a, b]$*;
(iii) *$\varrho_g(\gamma(s), \gamma(t)) \leqslant |s - t|$ for all $s, t \in [a, b]$ and all $g \in \mathfrak{G}_P$.*

Proof For a Riemannian metric $g \in \mathfrak{G}_P$, inequality (23) follows easily from the fact that, for each $\gamma \in \Gamma_P^{\mathrm{unit}}([0, T])$ joining x to y, the g-norm of $\gamma'(t)$ is at most 1 for almost all t, so

$$\varrho_g(x, y) \leqslant \int_0^T \left|\gamma'(t)\right|_g \mathrm{d}t \leqslant T.$$

From (23), if $\varrho_P(x, y) = 0$, then $\varrho_g(x, y) = 0$ and hence $x = y$; it follows immediately that ϱ_P satisfies the other axioms for an (extended) distance function. Moreover, again by (23), the topology induced by ϱ_P is no coarser than the topology induced by ϱ_g, that is, the original topology of M.

Further, for a function $\gamma : [a, b] \to M$, condition (i) implies condition (ii) by the definition of ϱ_P, while condition (ii) implies condition (iii) by (23). Finally, if condition (iii) holds, then γ is absolutely continuous and, for all $g \in \mathfrak{G}_P$, $|\gamma'(t)|_g \leqslant 1$ for almost all $t \in [a, b]$. As $\mathfrak{G}_P$ is countable, we may reverse the order of the quantifiers on g and t, and deduce from Lemma 4.1 that $P^*(\gamma'(t)) \leqslant 1$ for almost all $t \in [a, b]$, which is condition (i). □

4.1 Topologies on M

By Proposition 4.6, the topology induced by ϱ_P is no coarser than the original manifold topology of M; recall (from Definition 1.1) that ϱ_P is varietal if the two topologies are equivalent. In general, the topology induced by ϱ_P may be finer than the original manifold topology of M. Unless otherwise specified, we do not assume that ϱ_P is varietal, and topological concepts such as compactness and convergence refer to the original topology of M.

Lemma 4.7 *Suppose that a sequence of curves $\gamma_m \in \Gamma_P^{\mathrm{unit}}([a, b])$ converges pointwise to a curve $\gamma : [a, b] \to M$. Then $\gamma \in \Gamma_P^{\mathrm{unit}}([a, b])$.*

Proof The characterisation of P-subunit curves in Proposition 4.6(iii) is preserved by pointwise convergence. □

Recall that $B_P^-(K,R)$ denotes $\{x \in M : \varrho_P(K,x) \leqslant R\}$.

Proposition 4.8 *Suppose that K is a compact subset of M and $R \in \mathbb{R}^+$. If $B_P^-(K,R)$ has compact closure, then it is compact and coincides with the set of all $x \in M$ for which there exists $\gamma \in \Gamma_P^{\mathrm{unit}}([0,R])$ such that $\gamma(0) \in K$ and $\gamma(R) = x$. Moreover $B_P^-(K,R')$ is also compact for some R' greater than R.*

Proof Take a Riemannian metric $g \in \mathfrak{G}_P$. If $B_P^-(K,R)$ is relatively compact, then $B_g^-(B_P^-(K,R), R\varepsilon)$ is compact for sufficiently small positive ε.

Take $x \in B_P^-(K,R)$. We may then find $\gamma_m \in \Gamma_P^{\mathrm{unit}}([0, R(1+\varepsilon_m)])$ such that $\gamma_m(0) \in K$ and $\gamma_m((1+\varepsilon_m)R) = x$, where $0 \leqslant \varepsilon_m \leqslant \varepsilon$ and $\varepsilon_m \downarrow 0$. Hence the images of the γ_m are all contained in $B_g^-(B_P^-(K,R), R\varepsilon)$. If we rescale these curves so that they are all defined on $[0,R]$, we obtain a sequence of curves $\tilde{\gamma}_m : [0,R] \to M$ which are $(1+\varepsilon)$-Lipschitz with respect to ϱ_g, and whose images are all contained in $B_g^-(B_P^-(K,R), R\varepsilon)$. By the Arzelà–Ascoli theorem, after taking a subsequence, the $\tilde{\gamma}_m$ converge uniformly to a curve $\gamma : [0,R] \to M$ for which $\gamma(0) \in K$ and $\gamma(R) = x$.

Now, for all positive δ, the rescalings of the curves $\tilde{\gamma}_m$ on $[0,(1+\delta)R]$ are eventually P-subunit (since $\varepsilon_m \to 0$), therefore their limit, that is, the rescaling of γ on $[0,(1+\delta)R]$, is also P-subunit, by Lemma 4.7. In other words, $P^*(\gamma'(t)) \leqslant 1+\delta$ for all positive δ and almost all $t \in [0,R]$. It follows by exchanging quantifiers that γ is P-subunit.

Finally, take x in the closure of $B_P^-(K,R)$. Then there is a sequence of P-subunit curves $\gamma_m : [0,R] \to M$ such that $\gamma_m(0) \in K$ and $\gamma_m(R) \to x$. As before, we may extract a subsequence that converges uniformly to a P-subunit curve $\gamma : [0,R] \to M$ such that $\gamma(0) \in K$ and $\gamma(R) = x$, and therefore $x \in B_P^-(K,R)$. This shows that $B_P^-(K,R)$ is closed, hence compact. The same argument proves that $B_P^-(K, R(1+\varepsilon))$ is compact too, since it is contained in the compact set $B_g^-(B_P^-(K,R), R\varepsilon)$. □

If ϱ_P is varietal, then the proof of Proposition 4.8 may be simplified.

Definition 4.9 For a compact subset K of M, we define

$$R_P(K) = \sup\left\{R \in \mathbb{R}^+ : B_P^-(K,R) \in \mathfrak{K}(M)\right\}. \tag{24}$$

For a point x in M, we write $R_P(x)$ instead of $R_P(\{x\})$.

By Proposition 4.8, the supremum is never a maximum and is always strictly positive.

4.2 Distance, Rectifiability and Length

The previous characterisation of P-subunit curves shows that (M, ϱ_P) is an (extended) length space, in the sense of Gromov (see, for instance, [22]).

Definition 4.10 Suppose that $\gamma : [a, b] \to M$ is a continuous curve. The P-length of γ, written $\ell_P(\gamma)$, is defined to be

$$\sup\left\{\sum_{j=1}^{m} \varrho_P\big(\gamma(t_{j-1}), \gamma(t_j)\big) : m \in \mathbb{N},\ a = t_0 \leqslant \cdots \leqslant t_m = b\right\}. \tag{25}$$

To help state the next results, we define $\Gamma_P([a, b])$ and $\Gamma([a, b])$ to be the sets of all ϱ_P-continuous and all continuous curves $\gamma : [a, b] \to M$.

Proposition 4.11 *For all $x, y \in M$, the distance $\varrho_P(x, y)$ is equal to*

$$\inf\big\{\ell_P(\gamma) : \gamma \in \Gamma_P\big([a, b]\big),\ \gamma(a) = x,\ \gamma(b) = y\big\}. \tag{26}$$

Proof Write $\tilde{\varrho}_P(x, y)$ for the expression (26). On the one hand, $\tilde{\varrho}_P$ is an extended distance function and $\tilde{\varrho}_P \geqslant \varrho_P$ since $\ell_P(\gamma) \geqslant \varrho_P(\gamma(a), \gamma(b))$ for all $\gamma \in \Gamma_P([a, b])$. On the other hand, if $\gamma \in \Gamma_P^{\text{unit}}([0, T])$, then $\gamma \in \Gamma_P([a, b])$ and $\ell_P(\gamma) \leqslant T$, by Proposition 4.6, and the reverse inequality $\tilde{\varrho}_P \leqslant \varrho_P$ follows. □

Next we show that the expression (26) does not change if we require only that the curves γ are continuous with respect to the manifold topology.

Proposition 4.12 *If $\gamma \in \Gamma([a, b])$ and $\ell_P(\gamma) < \infty$, then $\gamma \in \Gamma_P([a, b])$, and the topology on $\gamma([a, b])$ induced by ϱ_P and the relative topology coincide.*

Proof Since γ is continuous, $\gamma([a, b])$ is compact, hence $R_P(\gamma([a, b])) > 0$ by Proposition 4.8.

Fix now $\bar{t} \in [a, b[$. First, since $\ell_P(\gamma) < \infty$,

$$\inf_{\varepsilon > 0} \sup_{t, t' \in]\bar{t}, \bar{t} + \varepsilon[} \varrho_P\big(\gamma(t), \gamma\big(t'\big)\big) = 0;$$

in fact, if the infimum η were positive, then we could find a decreasing sequence $(t_m)_{m \in \mathbb{N}}$ tending to $\bar{t}$ such that $\varrho_P(\gamma(t_{2k+1}), \gamma(t_{2k})) \geqslant \eta/2$, and deduce that

$$\ell_P(\gamma) \geqslant \sum_{k=0}^{j-1} \varrho_P\big(\gamma(t_{2k+1}), \gamma(t_{2k})\big) \geqslant j\frac{\eta}{2}$$

for all $j \in \mathbb{N}$, which is a contradiction.

Therefore, for every positive δ, there is a positive ε such that

$$\gamma\big(]\bar{t}, \bar{t} + \varepsilon[\big) \subseteq B_P^-\big(\gamma(t), \delta\big)$$

for all $t \in]\bar{t}, \bar{t}+\varepsilon[$. If $\delta < R_P(\gamma([a,b]))$, then $B_P^-(\gamma(t),\delta)$ is closed, hence

$$\gamma(\bar{t}) \in \gamma\big(]\bar{t}, \bar{t}+\varepsilon[\big)^- \subseteq B_P^-\big(\gamma(t),\delta\big)$$

by the continuity of γ, which means that $\varrho_P(\gamma(\bar{t}),\gamma(t)) \leqslant \delta$.

This proves that $\lim_{t\to\bar{t}+} \varrho_P(\gamma(t),\gamma(\bar{t})) = 0$. The proof when $\bar{t} \in]a,b]$ and $t \to \bar{t}-$ is similar. To conclude, recall that a continuous map from a compact space to a Hausdorff space is closed; hence every topology on M that makes γ continuous induces the quotient topology induced by γ on $\gamma([a,b])$. □

Corollary 4.13 *For all $x, y \in M$, the distance $\varrho_P(x,y)$ is equal to*

$$\inf\big\{\ell_P(\gamma) : \gamma \in \Gamma\big([a,b]\big),\ \gamma(a)=x,\ \gamma(b)=y\big\}.$$

Proof This follows immediately from Propositions 4.11 and 4.12. □

We may express the length of an absolutely continuous curve as an integral.

Proposition 4.14 *Suppose that $\gamma \in AC([a,b])$. Then*

$$\ell_P(\gamma) = \int_a^b P^*\big(\gamma'(t)\big)\,\mathrm{d}t. \tag{27}$$

If $\ell_P(\gamma) < \infty$, then γ is also ϱ_P-absolutely continuous and

$$P^*\big(\gamma'(t)\big) = \lim_{s\to t} \frac{\varrho_P(\gamma(s),\gamma(t))}{|s-t|} \tag{28}$$

for almost all $t \in [a,b]$.

We remark that, in the general theory of absolutely continuous curves in metric spaces (see, for example, [3, Sect. 4.1] or [2, Sect. 1.1]), the right-hand side of (28) is known as the *metric derivative* of γ.

Proof Note first that the corresponding statement for a Riemannian metric g on M is easily proved. To do so, define ℓ_g like ℓ_P in Definition 4.10, but with ϱ_P replaced by ϱ_g. By using exponential coordinates centred at $\gamma(t)$, one sees that

$$\big|\gamma'(t)\big|_g = \lim_{s\to t} \frac{\varrho_g(\gamma(s),\gamma(t))}{|s-t|} \tag{29}$$

for all points t in $[a,b]$ at which γ is differentiable, and it follows from the theory of absolutely continuous curves in metric spaces that

$$\ell_g(\gamma|_{[t_1,t_2]}) = \int_{t_1}^{t_2} \big|\gamma'(\tau)\big|_g\,\mathrm{d}\tau \tag{30}$$

whenever $a \leqslant t_1 \leqslant t_2 \leqslant b$. From (17), (23) and (29), we deduce that

$$P^*\big(\gamma'(t)\big) \leqslant \liminf_{s \to t} \frac{\varrho_P(\gamma(s), \gamma(t))}{|s-t|} \tag{31}$$

for all $t \in [a,b]$ where γ is differentiable.

Suppose now that $\ell_P(\gamma) < \infty$. Then the function $r\colon [a,b] \to \mathbb{R}$, defined by $r(t) = \ell_P(\gamma|_{[a,t]})$, is non-decreasing, so differentiable almost everywhere, and

$$\int_a^b r'(\tau)\,\mathrm{d}\tau \leqslant r(b) - r(a) = \ell_P(\gamma).$$

Now

$$\varrho_P\big(\gamma(t_1), \gamma(t_2)\big) \leqslant \ell_P(\gamma|_{[t_1,t_2]}) = r(t_2) - r(t_1)$$

whenever $a \leqslant t_1 \leqslant t_2 \leqslant b$, so $P^*(\gamma'(t)) \leqslant r'(t)$ for almost all $t \in [a,b]$ from (31), and *a fortiori*

$$\int_a^b P^*\big(\gamma'(\tau)\big)\,\mathrm{d}\tau \leqslant \ell_P(\gamma).$$

The same inequality holds trivially when the right-hand side is infinite.

Conversely, if $T = \int_a^b P^*(\gamma'(\tau))\,\mathrm{d}\tau < \infty$, then define $\tilde{r}\colon [a,b] \to [0,T]$ by

$$\tilde{r}(t) = \int_a^t P^*\big(\gamma'(\tau)\big)\,\mathrm{d}\tau.$$

The function $\tilde{r}$ is non-decreasing and surjective. Further, if $a \leqslant t_1 \leqslant t_2 \leqslant b$ and $g \in \mathfrak{G}_P$, then

$$\varrho_g\big(\gamma(t_1), \gamma(t_2)\big) \leqslant \int_{t_1}^{t_2} \big|\gamma'(\tau)\big|_g\,\mathrm{d}\tau \leqslant \int_{t_1}^{t_2} P^*\big(\gamma'(\tau)\big)\,\mathrm{d}\tau = \tilde{r}(t_2) - \tilde{r}(t_1),$$

by (17) and (30). In particular, if $\tilde{r}(t_1) = \tilde{r}(t_2)$ then $\gamma(t_1) = \gamma(t_2)$, hence we may define a function $\tilde{\gamma}\colon [0,T] \to M$ by $\gamma = \tilde{\gamma} \circ \tilde{r}$, and $\tilde{\gamma}$ is 1-Lipschitz with respect to ϱ_g for every $g \in \mathfrak{G}_P$. By Proposition 4.6, this implies that $\tilde{\gamma}\colon [0,T] \to M$ is 1-Lipschitz with respect to ϱ_P, hence $\tilde{\gamma} \in \Gamma_P^{\mathrm{unit}}([0,T])$ and

$$\ell_P(\gamma) = \ell_P(\tilde{\gamma}) \leqslant T = \int_a^b P^*\big(\gamma'(\tau)\big)\,\mathrm{d}\tau.$$

Again, this inequality holds trivially when the right-hand side is infinite, and we have proved (27).

If $\ell_P(\gamma) < \infty$, then $P^*(\gamma'(\cdot))$ is integrable on $[a,b]$ and

$$\ell_P(\gamma|_{[t_1,t_2]}) = \int_{t_1}^{t_2} P^*\big(\gamma'(\tau)\big)\,\mathrm{d}\tau$$

whenever $a \leqslant t_1 \leqslant t_2 \leqslant b$; now (28) follows from the theory of absolutely continuous curves in metric spaces. □

Corollary 4.15 *For all* $x, y \in M$,

$$\varrho_P(x, y) = \inf\left\{ \int_a^b P^*\big(\gamma'(t)\big)\,\mathrm{d}t : \gamma \in AC\big([a, b]\big),\ \gamma(a) = x,\ \gamma(b) = y \right\}.$$

We conclude our discussion of curves and lengths by pointing out that any curve of finite P-length may be reparametrised using arc-length, and then becomes a subunit curve, from part (iii) of Proposition 4.6.

4.3 Completeness

We say that the fibre seminorm P is *complete* if the set $B_P^-(K, R)$ is relatively compact for all compact subsets K of M and all positive R. By Proposition 4.8, P is complete if and only if $R_P(K) = \infty$ for all compact subsets K of M.

Proposition 4.16 *If P is complete, then the metric space (M, ϱ_P) is complete. The converse holds if ϱ_P is varietal.*

Proof Suppose that P is complete, and take a ϱ_P-Cauchy sequence $(x_m)_{m\in\mathbb{N}}$ in M. The set $\{x_m\}_{m\in\mathbb{N}}$ is ϱ_P-bounded, hence it is relatively compact, thus we may find a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ that converges to a point $x \in M$ in the manifold topology. By completeness and Proposition 4.8, the function $\varrho_P(x_m, \cdot)$ is lower-semicontinuous, whence

$$\varrho_P(x_m, x) \leqslant \liminf_{k\to\infty} \varrho_P(x_m, x_{n_k}),$$

and the right-hand side tends to 0 as m tends to ∞ since $(x_m)_{m\in\mathbb{N}}$ is ϱ_P-Cauchy, hence $\varrho_P(x_m, x)$ tends to 0.

Conversely, suppose that (M, ϱ_P) is complete and ϱ_P is varietal. Then each compact subset K of M is ϱ_P-bounded, so $B_P^-(K, R)$ is also bounded for all positive R, and it is closed because ϱ_P is continuous. Since (M, ϱ_P) is a complete locally compact length space, closed ϱ_P-bounded subsets of M are compact [22, Theorem 1.5], and we are done. □

Definition 4.17 The closed set $\{x \in M : P_x \neq 0\}^-$ is said to be the *support* of the fibre seminorm P.

Proposition 4.18 *If P is compactly-supported, then it is complete.*

Proof If $x \in M \setminus \mathrm{supp}(P)$, then the only P-subunit vector in $T_x M$ is the null vector. Hence all P-subunit curves passing through $M \setminus \mathrm{supp}(P)$ are constant, and every point of $M \setminus \text{supp}(P)$ has infinite ϱ_P-distance to every other point of M. Hence, for all compact subsets K of M and all positive R,

$$B_P^-(K, R) = K \cup B_P^-\big(K \cap \text{supp}(P), R\big),$$

and necessarily $B_P^-(K \cap \text{supp}(P), R) \subseteq \text{supp}(P)$. Thus $B_P^-(K \cap \text{supp}(P), R)$ is compact, by Proposition 4.8, and consequently $B_P^-(K, R)$ is compact. □

4.4 Subunit Vector Fields and Hörmander's Condition

We begin with a definition.

Definition 4.19 A smooth section of TM that is P-subunit everywhere in M is said to be a P-subunit vector field.[1] We write $\mathfrak{X}_P$ for the set of all P-subunit vector fields and $L(\mathfrak{X}_P)$ for the Lie algebra of vector fields generated by $\mathfrak{X}_P$.

Various extended distance functions may be defined as in Definition 4.4, by restricting γ to a subclass of Γ_P^{unit}. For example, we might restrict out attention to smooth P-subunit curves, or piecewise smooth P-subunit curves, or flow curves along P-subunit vector fields. More precisely, the flow curves of a P-subunit vector field are smooth P-subunit curves, and any concatenation of flow curves of P-subunit fields is a P-subunit curve; such a concatenation will be called a P-subunit piecewise flow curve.

Definition 4.20 We write Γ_P^∞ for the set of smooth P-subunit curves, Γ_P^{flow} for the set of P-subunit piecewise flow curves, ϱ_P^∞ for the extended distance corresponding to the class Γ_P^∞ and ϱ_P^{flow} for the extended distance function corresponding to the class Γ_P^{flow}.

It is not obvious that these three distance functions are the same, however

$$\varrho_P \leqslant \varrho_P^\infty \leqslant \varrho_P^{\text{flow}}. \tag{32}$$

The second inequality is justified because we may obtain the distance function ϱ_P^∞ either by taking the class of smooth P-subunit curves or by taking the class of piecewise smooth P-subunit curves, that is, the P-subunit curves $\gamma : [a, b] \to M$ for which a finite subdivision $\{t_0, \ldots, t_k\}$ of $[a, b]$ such that $\gamma|_{[t_{j-1},t_j]}$ is smooth when $j = 1, \ldots, k$. Indeed, for every positive ε, there is a smooth increasing bijection $\eta\colon [a, b+\varepsilon] \to [a, b]$ such that $\eta' \leqslant 1$ and $\eta^{(h)}(\eta^{-1}(t_j)) = 0$ when $j = 0, \ldots, k$ and $h \geqslant 1$, so the reparametrisation $\gamma \circ \eta\colon [a, b + \varepsilon] \to M$ is P-subunit and smooth.

Example 4.21 Suppose that $\varphi\colon \mathbb{R} \to \mathbb{R}$ is continuous but not differentiable anywhere, and that $\varphi(0) = 0$. Given $(p, q) \in \mathbb{R}^2$, define the seminorm $P_{(p,q)}\colon \mathbb{R}^2 \to$

[1] We consider all vector fields to be smooth.

$[0, \infty[$ by

$$P_{(p,q)}(\xi, \eta) = \left|\xi + \varphi(q)\eta\right| / \left(1 + \varphi(q)^2\right)^{1/2}.$$

Then

$$P^*_{(p,q)}(u, v) = \begin{cases} |(u, v)| & \text{if } (u, v) \in \mathbb{R}(1, \varphi(q)) \\ \infty & \text{otherwise.} \end{cases}$$

It is easy to check that the vector field $\partial/\partial x$ along the x axis does not extend to a P-subunit vector field, because we require vector fields to be smooth. In fact, there are no nonnull P-subunit vector fields. It follows that $\varrho_P^{\text{flow}}((0, 0), (1, 0)) = \infty$, while $\varrho_P^{\infty}((0, 0), (1, 0)) = 1$.

Definition 4.22 The fibre seminorm P is said to satisfy Hörmander's condition if $\{X|_x : X \in L(\mathfrak{X}_P)\} = T_x M$ for every $x \in M$.

Proposition 4.23 *If P satisfies Hörmander's condition, then ϱ_P^{flow} is varietal and* a fortiori *ϱ_P and ϱ_P^{∞} are varietal too.*

Proof Recall that $L(\mathfrak{X}_P)$ is the linear span of the iterated Lie brackets of elements of $\mathfrak{X}_P$. Hence, for every fixed $x \in M$, there is a finite subset $\mathfrak{X}$ of $\mathfrak{X}_P$ such that the iterated commutators of elements of $\mathfrak{X}$ up to some order, m say, evaluated at x, span $T_x M$. Denote by $\varrho_{\mathfrak{X}}$ the extended distance function corresponding to the class of P-subunit curves that are concatenations of flow curves of vector fields in $\mathfrak{X}$. Then clearly $\varrho_P^{\text{flow}} \leqslant \varrho_{\mathfrak{X}}$, so, by Chow's theorem (see, for example, [20, Chap. 2]), for some $g \in \mathfrak{G}_P$ and constant κ,

$$B_g(x, r) \subseteq B_{\varrho_{\mathfrak{X}}}\left(x, \kappa r^{1/m}\right) \subseteq B_{\varrho_P^{\text{flow}}}\left(x, \kappa r^{1/m}\right)$$

for all sufficiently small positive r. Hence the ϱ_P^{flow}-balls centred in x are neighbourhoods of x and, by the arbitrariness of x, the ϱ_P^{flow}-open sets are open. The conclusion follows from the inequalities (23) and (32). □

We remark that when the dimension of the spaces of finite vectors varies from point to point, the Hörmander condition depends on the seminorm P as well as on the vector space of finite vectors. For example, take a smooth function $\varphi\colon \mathbb{R} \to \mathbb{R}$, and define the seminorm $P_{(p,q)}$ on $\mathbb{R}^2$ by

$$P_{(p,q)}(u, v) = \left(u^2 + \varphi^2(p)v^2\right)^{1/2}.$$

Then P satisfies Hörmander's condition if $\varphi(p) = p^k$ where $k \in \mathbb{N}$, but not if φ is the smooth extension of $p \mapsto \mathrm{e}^{-1/p^2}$ to $\mathbb{R}$ (see the discussion in Sect. 8.7). However, the spaces of finite vectors coincide everywhere for these two examples.

Definition 4.24 The fibre seminorm P is said to satisfy the Lipschitz seminorm condition if, for every $\alpha \in A$, there is a countable family $\mathfrak{X}$ of P-subunit vector fields on U_α and a constant L, which may depend on α, such that

(i) $|\tau_\alpha X(x) - \tau_\alpha X(y)| \leqslant L|x-y|$ for all $x, y \in B_{\mathbb{R}^n}(0,1)$ and $X \in \mathfrak{X}$, and
(ii) $\{X|_x : X \in \mathfrak{X}\}^- = \{v \in T_x M : P^*(v) \leqslant 1\}$ for all $x \in V_\alpha$.

Since the V_α are relatively compact in M and form a locally finite cover of M, the Lipschitz seminorm condition for P does not depend on the choice of the atlas $\{\varphi_\alpha\}_{\alpha \in A}$.

Theorem 4.25 *Suppose that P satisfies the Lipschitz seminorm condition, and that $\gamma \in \Gamma_P^{\mathrm{unit}}([0,T])$. For all neighbourhoods W of $\gamma(T)$ there exists a neighbourhood U of $\gamma(0)$ such that, for all $x \in U$, there exists $\delta \in \Gamma_P^{\mathrm{flow}}([0,T])$ for which $\delta(0) = x$ and $\delta(T) \in W$.*

Proof Fix a Riemannian metric $g \in \mathfrak{G}_P$.

With a view to a contradiction, suppose that $\gamma \in \Gamma_P^{\mathrm{unit}}([0,T])$ is "bad", that is, the conclusion does not hold for γ. Clearly $\gamma|_{[0,T/2]}$ or $\gamma(\cdot + T/2)|_{[0,T/2]}$ is bad too. Iteration of this bisection procedure, together with a compactness argument, shows that we may suppose that $\gamma([0,T]) \subseteq B_g^-(z,r)$ for some $z \in M$ and $r \in \mathbb{R}^+$ such that $B_g^-(z,3r) \subseteq V_\alpha$ for some $\alpha \in A$; further iteration allows us to suppose that $T < r$, so

$$\gamma\big([0,T]\big) \subseteq B_P^-\big(B_g^-\big(\gamma(0),T\big),T\big) \subseteq V_\alpha. \tag{33}$$

Now take the countable family $\mathfrak{X}$ of P-subunit vector fields X_k and the Lipschitz constant L corresponding to α as in Definition 4.24. There is a constant κ such that

$$\big|\tau_\alpha(v)\big| \leqslant \kappa P^*(v) \quad \text{for every } v \in T V_\alpha, \tag{34}$$

where $|\cdot|$ denotes the euclidean norm on $\mathbb{R}^n$.

Take $x \in B_g^-(\gamma(0),T)$. We aim to construct $\delta\colon [0,T] \to M$ that is a piecewise flow curve of fields in $\mathfrak{X}$, such that $\delta(0) = x$, and $\delta(T)$ is arbitrarily near $\gamma(T)$ whenever x is sufficiently near $\gamma(0)$. The image of any such δ is contained in V_α by (33), therefore from now on we work in the coordinates φ_α. For simplicity, we continue to write γ rather than $\varphi_\alpha \circ \gamma$. Hence

$$\gamma(t) = \gamma(0) + \int_0^t \gamma'(\tau)\,\mathrm{d}\tau. \tag{35}$$

By altering γ' on a negligible subset of $[0,T]$, we may suppose that γ' is a Borel function, $\gamma'(t)$ is P-subunit for all $t \in [0,T]$, and (35) still holds.

Fix $\varepsilon \in \mathbb{R}^+$. By the density and smoothness properties of the family $\mathfrak{X}$, the function $\nu_0\colon [0,T] \to \mathbb{N}$, given by

$$\nu_0(t) = \min\big\{k \in \mathbb{N} : \big|X_k\big(\gamma(t)\big) - \gamma'(t)\big| \leqslant \varepsilon\big\},$$

is well-defined and Borel. Fix $N \in \mathbb{Z}^+$ and set $d = T/N$ and $b(t) = \lfloor t/d \rfloor d$. Then the function $\nu_1\colon [0,T] \to \mathbb{N}$, given by

$$\nu_1(t) = \min\big\{k \in \mathbb{N} : \big|X_k\big(\gamma\big(b(t)\big)\big) - X_{\nu_0(t)}\big(\gamma\big(b(t)\big)\big)\big| \leqslant \varepsilon\big\},$$

is also well-defined and Borel; furthermore, since b takes its values in the finite set $\{0, d, 2d, \dots, Nd\}$ and the unit P^*-ball at $\gamma(jd)$ is compact when $j = 0, \dots, N$, the function ν_1 takes a finite number of values too.

Set $I_j = [jd, (j+1)d[$, where $j = 0, \dots, N-1$. We define $\nu_2\colon [0, T] \to \mathbb{N}$ to be the increasing rearrangement of ν_1 on each of the intervals I_j, that is, $\nu_2(t) = n$ when $|I_{\lfloor t/d \rfloor} \cap \{\nu_1 \leqslant n-1\}| \leqslant t - b(t) < |I_{\lfloor t/d \rfloor} \cap \{\nu_1 \leqslant n\}|$, and set $\nu_2(T) = \nu_1(T)$. Hence ν_2 takes a finite number of values and is piecewise constant. Concatenating flow curves along fields in $\mathfrak{X}$, we define $\delta\colon [0, T] \to M$ by

$$\delta(t) = x + \int_0^t X_{\nu_2(\tau)}\big(\delta(\tau)\big)\,\mathrm{d}\tau. \tag{36}$$

We want now to estimate $|\gamma(T) - \delta(T)|$.

Set $D_j = |\delta(jd) - \gamma(jd)|$ when $j = 0, \dots, N$. Clearly

$$D_{j+1} \leqslant D_j + \left| \int_{I_j} \big(\gamma'(\tau) - X_{\nu_2(\tau)}\big(\delta(\tau)\big)\big)\,\mathrm{d}\tau \right|.$$

Decompose the integrand as

$$\begin{aligned} \gamma'(\tau) - X_{\nu_2(\tau)}\big(\delta(\tau)\big) &= \gamma'(\tau) - X_{\nu_0(\tau)}\big(\gamma(\tau)\big) \\ &\quad + X_{\nu_0(\tau)}\big(\gamma(\tau)\big) - X_{\nu_0(\tau)}\big(\gamma(jd)\big) \\ &\quad + X_{\nu_0(\tau)}\big(\gamma(jd)\big) - X_{\nu_1(\tau)}\big(\gamma(jd)\big) \\ &\quad + X_{\nu_1(\tau)}\big(\gamma(jd)\big) - X_{\nu_2(\tau)}\big(\gamma(jd)\big) \\ &\quad + X_{\nu_2(\tau)}\big(\gamma(jd)\big) - X_{\nu_2(\tau)}\big(\delta(\tau)\big). \end{aligned}$$

The norms of the first and third pieces are at most ε, by definition of ν_0 and ν_1. The second and the fifth pieces are controlled by the Lipschitz seminorm condition, together with inequalities

$$\big|\gamma(\tau) - \gamma(jd)\big| \leqslant \kappa d \quad \text{and} \quad \big|\delta(\tau) - \gamma(jd)\big| \leqslant D_j + \kappa d$$

for all $\tau \in I_j$, by (34), (35) and (36). The fourth piece vanishes after integration over I_j, because it is the difference of two simple functions, one of which is a rearrangement of the other. Putting everything together,

$$\begin{aligned} D_{j+1} &\leqslant D_j + \varepsilon d + L\kappa d^2 + \varepsilon d + L(D_j + \kappa d)d \\ &= (1 + Ld)D_j + 2\varepsilon d + 2L\kappa d^2, \end{aligned}$$

and by induction,

$$D_j \leqslant (1 + Ld)^j D_0 + 2(\kappa d + \varepsilon/L)\big((1 + Ld)^j - 1\big).$$

Since $d = T/N$ and $(1 + LT/N)^N \leqslant \mathrm{e}^{LT}$,

$$\big|\delta(T) - \gamma(T)\big| \leqslant \mathrm{e}^{LT}\big|x - \gamma(0)\big| + 2(\kappa T/N + \varepsilon/L)\big(\mathrm{e}^{LT} - 1\big).$$

Note now that T, κ and L do not depend on the parameters x, ε and N of the construction. Hence by taking N sufficiently large, ε sufficiently small, and x sufficiently near $\gamma(0)$, we may construct a subunit piecewise flow curve δ for which $|\delta(T) - \gamma(T)|$ is arbitrarily small. This contradicts the badness of γ and proves the desired result. □

Corollary 4.26 *Suppose that P satisfies the Lipschitz seminorm condition. For all $x, y \in M$ such that $x \neq y$,*

$$\varrho_P(x, y) \geqslant \liminf_{z \to y} \varrho_P^{\mathrm{flow}}(x, z). \tag{37}$$

If $\varrho_P^{\mathrm{flow}}$ is varietal, then $\varrho_P = \varrho_P^{\mathrm{flow}}$. More generally, if ϱ_P^{∞} is varietal, then $\varrho_P(x, y) = \varrho_P^{\infty}(x, y)$, and both are equal to

$$\inf\bigl\{\ell_P(\gamma) : \gamma \in \Gamma^{\infty}\bigl([a, b]\bigr),\ \gamma(a) = x,\ \gamma(b) = y\bigr\} \tag{38}$$

for all $x, y \in M$.

Proof The inequality (37) is trivially satisfied when $\varrho_P(x, y) = \infty$. Otherwise, by Theorem 4.25, for all open neighbourhoods W of y that do not contain x and all T greater than $\varrho_P(x, y)$, we may find $\delta \in \Gamma_P^{\mathrm{flow}}([0, T])$ such that $\delta(0) = x$ and $\delta(T) \in W$. Since $\delta([0, T])$ is connected and δ is not constant, $\delta([0, T]) \cap W \neq \{y\}$; hence we may find $z \in \delta([0, T]) \cap W \setminus \{y\}$ such that $\varrho_P^{\mathrm{flow}}(x, z) \leqslant T$.

In particular, if $\varrho_P^{\mathrm{flow}}$ is varietal, then it is continuous, hence

$$\varrho_P^{\mathrm{flow}}(x, y) \geqslant \varrho_P(x, y) \geqslant \liminf_{z \to y} \varrho_P^{\mathrm{flow}}(x, z) = \varrho_P^{\mathrm{flow}}(x, y)$$

by (32) and (37), and the equality $\varrho_P = \varrho_P^{\mathrm{flow}}$ follows. In fact, (32) and (37) also imply that $\varrho_P(x, y) \geqslant \liminf_{z \to y} \varrho_P^{\infty}(x, z)$ when $x \neq y$, therefore the same argument proves that $\varrho_P = \varrho_P^{\infty}$ whenever ϱ_P^{∞} is varietal. It remains to note that the infimum (38) is at least $\varrho_P(x, y)$ by Corollary 4.13, and at most $\varrho_P^{\infty}(x, y)$ because $\ell_P(\gamma) \leqslant T$ for all $\gamma \in \Gamma_P^{\infty}([0, T])$. □

It is interesting to compare the expressions for the distance as the infimum of the lengths of curves in the preceding corollary, Proposition 4.11, and Corollary 4.13. When the function $x \mapsto \dim Z(P_x)$ is continuous, the equality of $\varrho_P^{\infty}(x, y)$ and (38) may be obtained without the hypotheses of Corollary 4.26, thanks to the following result.

Proposition 4.27 *Suppose that $x \mapsto \dim Z(P_x)$ is continuous and that $\gamma \in \Gamma^1([a, b])$. Then $\ell_P(\gamma)$ is the infimum of the set of all $T \in \mathbb{R}^+$ for which there is a smooth diffeomorphism $r : [0, T] \to [a, b]$ such that $\gamma \circ r$ is P-subunit.*

Proof By Proposition 4.6, $\ell_P(\gamma) = \ell_P(\gamma \circ r) \leqslant T$ when $\gamma \circ r \in \Gamma_P^{\mathrm{unit}}([0, T])$. Hence we are done if $\ell_P(\gamma) = \infty$. Otherwise, by Propositions 4.2 and 4.14, the set $\{t \in [a,b] : P^*(\gamma'(t)) < \infty\}$ is closed in $[a,b]$ and has full measure, so is all of $[a,b]$. This means that $P^*(\gamma'(t)) < \infty$ for all $t \in [a,b]$, and hence $t \mapsto P^*(\gamma'(t))$ is continuous by Proposition 4.2 again.

Take now $\varepsilon \in \mathbb{R}^+$. We may find a smooth function $h_\varepsilon \colon [a,b] \to \mathbb{R}$ such that

$$P^*\big(\gamma'(t)\big) + \varepsilon/2 \leqslant h_\varepsilon(t) \leqslant P^*\big(\gamma'(t)\big) + \varepsilon \quad \text{for every } t \in [a,b],$$

and then define the smooth diffeomorphism $s_\varepsilon \colon [a,b] \to [0, s_\varepsilon(b)]$ by

$$s_\varepsilon(t) = \int_a^t h_\varepsilon(\tau)\,\mathrm{d}\tau.$$

Denote by $r_\varepsilon \colon [0, s_\varepsilon(b)] \to [a,b]$ the inverse of s_ε. Then $\gamma \circ r_\varepsilon$ is P-subunit since $(\gamma \circ r_\varepsilon)' = (\gamma' \circ r_\varepsilon)/(h_\varepsilon \circ r_\varepsilon)$. The conclusion now follows because

$$s_\varepsilon(b) = \int_a^b h_\varepsilon(\tau)\,\mathrm{d}\tau \to \ell_P(\gamma)$$

as $\varepsilon \to 0$ by Proposition 4.14. □

In Sect. 8.7, we show that Proposition 4.27 need not hold if $x \mapsto \dim Z(P_x)$ is not continuous.

5 The Control Distance for a Differential Operator

Take $D \in \mathfrak{D}_1(\mathscr{E}, \mathscr{F})$. We define a continuous fibre seminorm P_D on T^*M by

$$P_D(\xi) = \big|\sigma_1(D)(\xi)\big|_{\mathrm{op}}.$$

All the notions introduced in Sect. 4 in connection with the seminorm P_D may be applied to D: we will speak, for example, of D-subunit vectors and D-subunit curves, and the distance function ϱ_{P_D} will be called the *control distance function* associated to D and written ϱ_D. These notions depend only on the seminorm P_D, so by (5), they do not change if we replace D with D^+, or with the operator $\mathit{Đ} \in \mathfrak{D}_1(\mathscr{E} \oplus \mathscr{F}, \mathscr{E} \oplus \mathscr{F})$ given by

$$\mathit{Đ}(f,g) = \big(D^+ g, Df\big), \tag{39}$$

which satisfies $\mathit{Đ} = \mathit{Đ}^+$ and

$$\sigma_1(\mathit{Đ})(\xi) = \begin{pmatrix} 0 & \sigma_1(D^+)(\xi) \\ \sigma_1(D)(\xi) & 0 \end{pmatrix},$$

so $P_{\mathit{Đ}} = P_D = P_{D^+}$.

We will also say that D is complete if P_D is complete. In particular, by Proposition 4.18, D is complete if its symbol $\sigma_1(D)$ is a compactly-supported section of $\mathrm{Hom}(\mathbb{C}T^*M, \mathrm{Hom}(\mathscr{E}, \mathscr{F}))$.

5.1 The Weak Differentiability of Lipschitz Functions

Recall from Sect. 4.4 that a D-subunit vector field is a smooth section X of TM that is D-subunit at each point of M, and that $\varrho_D^{\text{flow}} \geqslant \varrho_D$ because any concatenation of flow curves of D-subunit fields is a D-subunit curve.

Lemma 5.1 *Suppose that a representative of $f \in L^1_{\text{loc}}(\mathscr{T})$ satisfies*

$$\left| f(x) - f(y) \right| \leqslant L \varrho_D^{\text{flow}}(x, y) \quad \textit{for every } x, y \in M,$$

where $L \in \mathbb{R}^+$. Then, for all D-subunit vector fields X, the distributional derivative Xf is in $L^\infty(\mathscr{T})$ and $\|Xf\|_\infty \leqslant L$.

Proof Compare with [12, Theorem 1.3]. Without loss of generality, we suppose that $L = 1$.

Take $x \in M$, and a smooth bump function η that is equal to 1 in a neighbourhood of x. It is enough to show that $\eta X f \in L^1_{\text{loc}}(\mathscr{T})$ and $\|\eta X f\|_\infty \leqslant 1$. We may therefore suppose that $\operatorname{supp} X$ is compact and contained in a coordinate chart. Moreover, since weak derivatives are independent of the measure on M, we may suppose that the measure coincides with Lebesgue measure in coordinates. It will then be sufficient to show that

$$\left| \langle\!\langle Xf, \bar{\varphi} \rangle\!\rangle \right| \leqslant \|\varphi\|_1$$

for all $\varphi \in C_c^\infty(\mathscr{T})$ with support contained in the coordinate chart.

Now

$$\langle\!\langle Xf, \bar{\varphi} \rangle\!\rangle = \left\langle\!\left\langle f, X^+ \bar{\varphi} \right\rangle\!\right\rangle = -\int f(x) X\varphi(x)\,\mathrm{d}x - \int f(x)(\operatorname{div} X)(x)\varphi(x)\,\mathrm{d}x.$$

Denote by $(t, x) \mapsto F_t(x)$ the flow of X, so

$$\begin{aligned} -\int f(x) X\varphi(x)\,\mathrm{d}x &= \lim_{t \to 0} \int f(x) \frac{\varphi(F_{-t}(x)) - \varphi(x)}{t}\,\mathrm{d}x \\ &= \lim_{t \to 0} \int \frac{f(F_t(x)) - f(x)}{t} \varphi(x) \det \mathrm{d}F_t(x)\,\mathrm{d}x \\ &\quad + \lim_{t \to 0} \int f(x)\varphi(x) \frac{\det \mathrm{d}F_t(x) - 1}{t}\,\mathrm{d}x. \end{aligned}$$

Note that the last limit exists, since $(\mathrm{d} \det \mathrm{d}F_t(x)/\mathrm{d}x)|_{t=0} = \operatorname{div} X(x)$, hence

$$\langle\!\langle Xf, \bar{\varphi} \rangle\!\rangle = \lim_{t \to 0} \int \frac{f(F_t(x)) - f(x)}{t} \varphi(x) \det \mathrm{d}F_t(x)\,\mathrm{d}x.$$

Further, $t \mapsto F_t(x)$ is a flow curve of the D-subunit field X for all x. Therefore $\varrho_D^{\text{flow}}(F_t(x), x) \leqslant |t|$ and so $|f(F_t(x)) - f(x)| \leqslant |t|$. Moreover, $\det \mathrm{d}F_0$ is identically equal to 1, and the desired conclusion follows. □

Proposition 5.2 *Suppose that a representative of $f \in L^1_{\mathrm{loc}}(\mathscr{T}_{\mathbb{R}})$ satisfies*

$$\left|f(x) - f(y)\right| \leqslant L\varrho_D^{\mathrm{flow}}(x, y) \quad \textit{for every } x, y \in M,$$

where $L \in \mathbb{R}^+$. Then f is weakly D^σ-differentiable and $\|\,|D^\sigma f|_{\mathrm{op}}\|_\infty \leqslant L$.

Proof Again, we suppose that $L = 1$.

Consider first the case where $\mathscr{E} = \mathscr{F}$ and $D = D^+$. Given any section $V \in C^\infty(\mathscr{E})$, define the differential operator $X_V \in \mathfrak{D}_1(\mathscr{T}, \mathscr{T})$ by

$$X_V h = \mathrm{i}V^*\big(D^\sigma h\big)V = \big\langle \mathrm{i}\sigma_1(D)(\mathrm{d}h)V, V\big\rangle;$$

this is the differential operator D^σ composed with the multiplication operator $g \mapsto \mathrm{i}V^* g V$. This operator is homogeneous (that is, it annihilates constants) and preserves real-valued functions by (10), because $D = D^+$; therefore X_V corresponds to a smooth vector field on M. Moreover, if $\|V\|_\infty \leqslant 1$, then

$$\left|\mathrm{d}h(X_V)\right| = \left|\big\langle \mathrm{i}\sigma_1(D)(\mathrm{d}h)V, V\big\rangle\right| \leqslant \left|\sigma_1(D)(\mathrm{d}h)\right|_{\mathrm{op}},$$

from which it follows that X_V is a D-subunit vector field; in this case, therefore, $X_V f \in L^\infty(\mathscr{T})$ and $\|X_V f\|_\infty \leqslant 1$ by Lemma 5.1.

More generally, we may define the operators $X_{V,W} h = \mathrm{i}W^*(D^\sigma h)V$, and (10) implies that $(X_{V,W} h)^- = X_{W,V}\bar{h}$. Since f is real-valued,

$$X_{V,W} f = \frac{1}{2}(X_{V+W} f - X_V f - X_W f) + \frac{\mathrm{i}}{2}(X_{V+\mathrm{i}W} f - X_V f - X_{\mathrm{i}W} f),$$

so $\|X_{V,W} f\|_\infty \leqslant 3$ when $\max\{\|V\|_\infty, \|W\|_\infty\} \leqslant 1$.

To prove that $D^\sigma f \in L^\infty$, it will be sufficient to show that there is a constant κ such that

$$\left|\langle\!\langle D^\sigma f, \varphi\rangle\!\rangle\right| \leqslant \kappa \|\varphi\|_1$$

for each $\varphi \in C^\infty_{\mathrm{c}}(\mathrm{Hom}(\mathscr{E}, \mathscr{E}))$ supported in an open subset U of M in which $\mathscr{E}$ is trivialisable. We may write φ as

$$\varphi = \sum_{j,k} \varphi_{j,k} V_k^* \otimes V_j$$

for a suitable choice of orthonormal frame $\{V_1, \dots, V_r\}$ of $\mathscr{E}|_U$ and sections $\varphi_{j,k} \in C^\infty_{\mathrm{c}}(\mathscr{T}|_U)$, so

$$\langle\!\langle D^\sigma f, \varphi\rangle\!\rangle = \sum_{j,k} \langle\!\langle V_j^*\big(D^\sigma f\big)V_k, \varphi_{j,k}\rangle\!\rangle = -\mathrm{i}\sum_{j,k} \langle\!\langle X_{V_k,V_j} f, \varphi_{j,k}\rangle\!\rangle,$$

hence

$$\left|\langle\!\langle D^\sigma f, \varphi\rangle\!\rangle\right| \leqslant 3\sum_{j,k} \|\varphi_{j,k}\|_1 \leqslant 3r^2\|\varphi\|_1.$$

Thus $D^\sigma f$ is an L^∞-section of $\mathrm{Hom}(\mathscr{E}, \mathscr{E})$ that satisfies $(D^\sigma f)^* = -D^\sigma f$ pointwise almost everywhere. By using local trivialisations, it is easy to construct a countable family of sections $V_m \in C^\infty_c(\mathscr{E})$ such that $\|V_m\|_\infty \leqslant 1$ and the set of the $V_m(x)$ of unit norm is dense in the unit sphere of $\mathscr{E}_x$ for all $x \in M$. Thus

$$\left|D^\sigma f\right|_{\mathrm{op}} = \sup_{m \in \mathbb{N}} \left|\left\langle \left(D^\sigma f\right) V_m, V_m \right\rangle\right| = \sup_{m \in \mathbb{N}} |X_{V_m} f| \leqslant 1$$

pointwise almost everywhere.

In the general case, if $\not{D}$ is defined as in (39), then $\not{D}^+ = \not{D}$ and $\varrho_D = \varrho_{\not{D}}$, therefore our last result implies that $\not{D}^\sigma f \in L^\infty$ and $\| \, |\not{D}^\sigma f|_{\mathrm{op}} \|_\infty \leqslant 1$. However,

$$\not{D}^\sigma h = \begin{pmatrix} 0 & (D^+)^\sigma h \\ D^\sigma h & 0 \end{pmatrix},$$

so weak $\not{D}^\sigma$-differentiability implies weak D^σ-differentiability, and $|\not{D}^\sigma f|_{\mathrm{op}} = |D^\sigma f|_{\mathrm{op}}$ pointwise almost everywhere. The conclusion follows. □

Proposition 5.2 extends to complex-valued functions f, by decomposing f in its real and imaginary parts; however in this way one obtains the weaker estimate $|D^\sigma f|_{\mathrm{op}} \leqslant 2$. The example where $M = \mathbb{R}^2$, $\mathscr{E} = \mathscr{T}$, $D = \partial_1 - \mathrm{i}\partial_2$, and $f(x_1, x_2) = x_1 + \mathrm{i}x_2$ shows that this estimate cannot be improved; since $|\not{D}^\sigma f|_{\mathrm{op}} \geqslant |D^\sigma f|_{\mathrm{op}}$, the assumption that $D = D^+$ does not help.

A partial converse of Proposition 5.2 is easily established under additional regularity assumptions on f.

Proposition 5.3 *Suppose that $f \in C^1(\mathscr{T}_\mathbb{R})$ and $\| \, |D^\sigma f|_{\mathrm{op}} \|_\infty \leqslant L$, where $L \in \mathbb{R}^+$. Then*

$$\left|f(x) - f(y)\right| \leqslant L\varrho_D(x, y) \quad \textit{for every } x, y \in M.$$

Proof Again, we suppose that $L = 1$.

Take a D-subunit curve $\gamma \colon [0, T] \to M$ from x to y. Then $f \circ \gamma \colon [0, T] \to \mathbb{R}$ is absolutely continuous and

$$\left|(f \circ \gamma)'(t)\right| = \left|\mathrm{d}f|_{\gamma(t)}\big(\gamma'(t)\big)\right| \leqslant \left|\sigma_1(D)(\mathrm{d}f|_{\gamma(t)})\right|_{\mathrm{op}} = \left|D^\sigma f\big(\gamma(t)\big)\right|_{\mathrm{op}} \leqslant 1$$

for almost all $t \in [0, T]$, since $\gamma'(t)$ is D-subunit. Hence

$$\left|f(x) - f(y)\right| \leqslant \int_0^T \left|(f \circ \gamma)'(t)\right| \mathrm{d}t \leqslant T.$$

The conclusion follows from the arbitrariness of γ. □

To remove the regularity assumptions on f from the previous statement, we need an extra hypothesis on the ϱ_D-topology.

Proposition 5.4 *Suppose that* $f \in W^{\infty}_{D^\sigma,\mathrm{loc}}(\mathscr{T}_\mathbb{R})$ *and* ϱ_D *is varietal. Then* f *has a continuous representative. If also* $\| |D^\sigma f|_{\mathrm{op}} \|_\infty \leqslant L$*, where* $L \in \mathbb{R}^+$*, then this continuous representative satisfies*

$$\left| f(x) - f(y) \right| \leqslant L \varrho_D(x, y) \quad \textit{for every } x, y \in M.$$

Proof Since f is real-valued, the smooth approximants $J_{\boldsymbol{\varepsilon}} f$ given by Theorem 3.3 are real-valued too, and form a bounded set in $W^{\infty}_{D^\sigma,\mathrm{loc}}(\mathscr{T})$.

Given any $x \in M$, take R_x in $]0, R_D(\{x\})[$ and write K_x for the closed ball $B^{-}_{D}(x, R_x)$. By Proposition 4.8, for all $y \in K_x$ there exists a D-subunit curve $\gamma \colon [0, T] \to M$ joining x to y such that $T = \varrho_D(x, y) \leqslant R_x$, hence the points of γ lie in K_x. Now, arguing as in the proof of Proposition 5.3,

$$\left| J_{\boldsymbol{\varepsilon}} f(x) - J_{\boldsymbol{\varepsilon}} f(y) \right| \leqslant \varrho_D(x, y) \sup_{z \in K_x} \left| D^\sigma J_{\boldsymbol{\varepsilon}} f(z) \right|_{\mathrm{op}}.$$

Since ϱ_D is varietal, the boundedness of the set $\{J_{\boldsymbol{\varepsilon}} f\}_{\boldsymbol{\varepsilon} \in E}$ in $W^{\infty}_{D^\sigma,\mathrm{loc}}(\mathscr{T})$ implies the local equiboundedness and equicontinuity of the $J_{\boldsymbol{\varepsilon}} f$, so, by the Arzelà–Ascoli theorem, there exists a subsequence of the net $(J_{\boldsymbol{\varepsilon}} f)_{\boldsymbol{\varepsilon} \in E}$ that converges uniformly on compacta to a continuous function $g \colon M \to \mathbb{R}$. Further, $f = g$ pointwise almost everywhere since $J_{\boldsymbol{\varepsilon}} f \to f$ in $L^1_{\mathrm{loc}}(\mathscr{E})$, and by replacing f with g we may suppose that f is continuous on M.

If moreover $\| |D^\sigma f|_{\mathrm{op}} \|_\infty \leqslant L$ almost everywhere, then Corollary 3.5 yields a sequence of smooth real-valued functions f_m that converges locally uniformly to f, for which $\| |D^\sigma f_m|_{\mathrm{op}} \|_\infty \leqslant L$. Since these f_m are L-Lipschitz with respect to ϱ_D by Proposition 5.3, their limit f is L-Lipschitz too. □

In general, Propositions 5.3 and 5.4 do not extend to complex-valued functions f. Indeed, suppose that $M = \mathbb{C} = \mathbb{R}^2$, $\mathscr{E} = \mathscr{T}$, and $D = \partial_1 + i\partial_2$. Then a holomorphic function $f \colon \mathbb{C} \to \mathbb{C}$ satisfies $Df = 0$ but may not be globally Lipschitz. However, when $D = D^+$, the propositions do extend, since, by (10),

$$\left| D^\sigma \operatorname{Re} f \right|_{\mathrm{op}} \leqslant \left(\left| D^\sigma f \right|_{\mathrm{op}} + \left| D^\sigma \bar{f} \right|_{\mathrm{op}} \right)/2 = \left| D^\sigma f \right|_{\mathrm{op}}.$$

We present now a consequence of Proposition 5.2 that does not require ϱ_D to be varietal.

Proposition 5.5 *Suppose that* $1 \leqslant p < \infty$*. Then* $L^p_c \cap W^p_D(\mathscr{E}) \subseteq W^p_{D,0}(\mathscr{E})$*. If* D *is complete, then* $W^p_D(\mathscr{E}) = W^p_{D,0}(\mathscr{E})$*.*

Proof If $f \in W^p_D(\mathscr{E})$ is compactly-supported, then it may be approximated in W^p_D by compactly-supported smooth sections of $\mathscr{E}$, by parts (i), (ii), (iv), and (ix) of Theorem 3.3, hence $f \in W^p_{D,0}(\mathscr{E})$.

Suppose now that D is complete. Take a sequence of compact sets K_m in M such that K_m is contained in the interior of K_{m+1} and $\bigcup_m K_m = M$, and define

$$g_m(x) = \left(1 - n^{-1} \varrho_D(K_m, x)\right)_+$$

for all positive n. By Proposition 4.8, the g_m are upper-semicontinuous and compactly-supported; moreover $0 \leqslant g_m \leqslant 1$, $g_m \uparrow 1$ pointwise and

$$\left|g_m(x) - g_m(y)\right| \leqslant n^{-1} \varrho_D(x, y),$$

so g_m is weakly D^σ-differentiable and $\| |D^\sigma g_m|_{\mathrm{op}} \|_\infty \leqslant n^{-1}$, by Proposition 5.2. If $u \in W^p_D(\mathscr{E})$, then, by Proposition 3.7,

$$D(g_m u) = \big(D^\sigma g_m\big) u + g_m Du,$$

therefore $g_m u, D(g_m u) \in L^p$ and $(g_m u, D(g_m u)) \to (u, Du)$ in L^p by the dominated convergence theorem. Since the $g_m u$ are compactly-supported, they belong to $W^p_{D,0}(\mathscr{E})$, and we conclude that $u \in W^p_{D,0}(\mathscr{E})$ too. □

5.2 Equivalent Definitions of the Control Distance

Since the control distance function ϱ_D is the distance function associated with the fibre seminorm P_D, various equivalent characterisations of ϱ_D are contained in Sect. 4. In particular, the results of Sect. 4.4 apply because of the following property.

Proposition 5.6 *The fibre seminorm P_D satisfies the Lipschitz seminorm condition.*

Proof Since $P_D = P_{\underline{D}}$, it is not restrictive to suppose that $D = D^+$. We may suppose moreover that the local trivialisations of $\mathscr{E}$ are isometric. Take $\alpha \in A$, and choose a bump function η on $\mathbb{R}^n$ which is equal to 1 on $B_{\mathbb{R}^n}(0, 1)$, and a countable dense set $\mathscr{W}$ of the unit sphere in $\mathbb{C}^r$. For every $w \in \mathscr{W}$, define $V_w \in C^\infty_{\mathrm{c}}(\mathscr{E})$ by requiring that V_w is supported in U_α and

$$\tau_\alpha(V_w)(x) = \eta(x) w \quad \text{for every } x \in \mathbb{R}^n,$$

and then define the D-subunit field X_w by

$$X_w h = \mathrm{i} V_w^* \big(D^\sigma h\big) V_w = \big\langle \mathrm{i}\sigma_1(D)(\mathrm{d}h) V_w, V_w \big\rangle.$$

If D is expressed in coordinates, as in (1), then (2) implies that

$$\tau_\alpha X_w(x) = \big(\mathrm{i}\eta(x)^2 \langle a_j(x) w, w\rangle\big)_j \quad \text{for every } x \in \mathbb{R}^n,$$

from which it is clear that the family $\{\tau_\alpha X_w\}_{w \in \mathscr{W}}$ is equi-Lipschitz, with a Lipschitz constant depending on the derivatives of the smooth coefficients a_j of D. Moreover, for all $x \in V_\alpha$, the set $\{V_w|_x\}_{w \in \mathscr{W}}$ is dense in the unit sphere of $\mathscr{E}_x$, so

$$P_D(\xi) = \big|\sigma_1(D)(\xi)\big|_{\mathrm{op}} = \sup_{w \in \mathscr{W}} \big|\xi(X_w|_x)\big|$$

for all $x \in V_\alpha$ and $\xi \in T_x^* M$. The bipolar theorem [19, Sect. 20.8] implies that, for all $x \in V_\alpha$, the set $\{v \in T_x M : P_D^*(v) \leqslant 1\}$ is the closed convex envelope of $\{\pm X_w|_x : w \in \mathscr{W}\}$. Hence the set $\mathfrak{X}$ of convex combinations with rational coefficients of elements of $\{\pm X_w : w \in \mathscr{W}\}$ is a countable family of compactly-supported D-subunit fields, such that $\{\tau_\alpha X\}_{X \in \mathfrak{X}}$ is equi-Lipschitz, and $\{X|_x\}_{X \in \mathfrak{X}}$ is dense in $\{v \in T_x M : P_D^*(v) \leqslant 1\}$ for all $x \in V_\alpha$. □

The following result is an immediate consequence of Proposition 4.23 and Corollary 4.26.

Corollary 5.7 *If P_D satisfies Hörmander's condition, then ϱ_D^{flow} is varietal, and, $\varrho_D(x, y) = \varrho_D^\infty(x, y) = \varrho_D^{\text{flow}}(x, y)$ for all $x, y \in M$; further, all are equal to*

$$\inf\big\{\ell_D(\gamma) : \gamma \in \Gamma^\infty\big([a, b]\big),\ \gamma(a) = x,\ \gamma(b) = y\big\}.$$

Another characterisation of the control distance may be given in terms of smooth functions with "bounded gradient".

Proposition 5.8 *Suppose that ϱ_D is varietal. Then*

$$\varrho_D(x, y) = \sup\big\{\big|\xi(x) - \xi(y)\big| : \xi \in C^\infty(\mathscr{T}_\mathbb{R}),\ \big\|\big|D^\sigma \xi\big|_{\text{op}}\big\|_\infty \leqslant 1\big\}. \tag{40}$$

If D is complete, then the supremum may be restricted to ξ in $C_c^\infty(\mathscr{T}_\mathbb{R})$.

Proof The left-hand side of (40) is greater than or equal to the right-hand side, without any assumptions on D, from Proposition 5.3. For the reverse inequality, take $x, y \in M$ and $\lambda \in]0, \varrho_D(x, y)[$, and define

$$f(z) = \big(\lambda - \varrho_D(x, z)\big)_+ \quad \text{for every } z \in M.$$

Then f is finite and continuous; moreover, by Proposition 5.2, f is weakly D^σ-differentiable and $\|\,|D^\sigma f|_{\text{op}}\|_\infty \leqslant 1$. Therefore, by Corollary 3.5, there is a sequence of real-valued smooth functions f_m such that $|D^\sigma f_m|_{\text{op}} \leqslant 1$ and f_m converges locally uniformly to f; thus

$$\big|f_m(x) - f_m(y)\big| \to \big|f(x) - f(y)\big| = \lambda,$$

and the first part of the conclusion follows by the arbitrariness of λ. If D is complete, then the function f is compactly-supported, and by Corollary 3.5 the smooth approximants f_m may also be chosen compactly-supported. □

Now we are going to show that the characterisation of ϱ_D given by Proposition 5.8 may hold even when ϱ_D is not varietal.

Fix a Riemannian metric g on M. This induces a fibre inner product on $\mathbb{C}T^*M$. For all $m \in \mathbb{N}$, define $D_m \in \mathfrak{D}_1(\mathscr{E} \oplus \mathscr{T}, \mathscr{F} \oplus \mathbb{C}T^*M)$ by

$$D_m(f, g) = \big(Df, 2^{-m}\, dg\big). \tag{41}$$

Then

$$(D_m)^\sigma h = \begin{pmatrix} D^\sigma h & 0 \\ 0 & 2^{-m}\,\mathrm{d}h \end{pmatrix},$$

so

$$P_{D_m}(\xi) = \max\{P_D(\xi), 2^{-m}|\xi|_g\} \quad \text{for every } \xi \in T^*M.$$

In particular, a vector $v \in TM$ is D-subunit if and only if it is D_m-subunit for all $m \in \mathbb{N}$. Moreover, $\varrho_{D_m} \leqslant 2^m \varrho_g$, so ϱ_{D_m} is varietal for all $m \in \mathbb{N}$.

Proposition 5.9 *Suppose that* D_0, *given by* (41), *is complete. Then*

$$\varrho_D(x, y) = \sup_{m\in\mathbb{N}} \varrho_{D_m}(x, y) \quad \textit{for every } x, y \in M.$$

Proof Recall from Proposition 4.6 that a curve is P-subunit if and only if it is 1-Lipschitz with respect to ϱ_P. From the definition of D_0, it is clear that $\varrho_D \geqslant \varrho_{D_m}$ for all $m \in \mathbb{N}$.

Fix $x, y \in M$ such that $\sup_{m\in\mathbb{N}} \varrho_{D_m}(x, y)$ is finite, and take a finite T greater than the supremum. For all $m \in \mathbb{N}$, we choose a D_m-subunit curve $\gamma_m : [0, T] \to M$ such that $\gamma_m(0) = x$ and $\gamma_m(T) = y$. If $m \leqslant m'$, then $\varrho_{D_m} \leqslant \varrho_{D_{m'}}$ by definition, hence $\gamma_{m'}$ is also D_m-subunit.

Consequently, all the curves γ_m are 1-Lipschitz with respect to ϱ_{D_0}, and all take their values in $B^-_{D_0}(x, T)$, which is compact because D_0 is complete. By the Arzelà–Ascoli theorem, there is a subsequence of $(\gamma_m)_{m\in\mathbb{N}}$ that converges uniformly to a continuous curve $\gamma : [0, T] \to M$. Then $\gamma(0) = x$ and $\gamma(T) = y$; further, γ is 1-Lipschitz with respect to all the ϱ_{D_m}, that is, γ is D_m-subunit for all $m \in \mathbb{N}$. Thus $\gamma'(t)$ is D_m-subunit for almost every $t \in [0, T]$ and all $m \in \mathbb{N}$, which implies that $\gamma'(t)$ is D-subunit. Hence γ is D-subunit, and $\varrho_D(x, y) \leqslant T$. □

Corollary 5.10 *Suppose that* D_0, *given by* (41), *is complete. Then, for all* $x, y \in M$,

$$\varrho_D(x, y) = \sup\{|\xi(x) - \xi(y)| : \xi \in C^\infty_c(\mathscr{T}_\mathbb{R}),\ \||D^\sigma\xi|_{\mathrm{op}}\|_\infty \leqslant 1\}.$$

Proof Fix $x, y \in M$, and take λ less than $\varrho_D(x, y)$. By Proposition 5.9, there exists $m \in \mathbb{N}$ such that $\lambda < \varrho_{D_m}(x, y)$. Since ϱ_{D_m} is varietal and D_m is complete, there exists $\xi \in C^\infty_c(\mathscr{T}_\mathbb{R})$ such that $|\xi(x) - \xi(y)| > \lambda$ and $\||D^\sigma_m\xi|_{\mathrm{op}}\|_\infty \leqslant 1$ by Proposition 5.8. However,

$$|D^\sigma_m\xi|_{\mathrm{op}} = \max\{|D^\sigma\xi|_{\mathrm{op}}, 2^{-m}|\mathrm{d}\xi|_g\} \geqslant |D^\sigma\xi|_{\mathrm{op}},$$

so $\||D^\sigma\xi|_{\mathrm{op}}\|_\infty \leqslant 1$. □

In general, the completeness of D_0 depends on the choice of the Riemannian metric g in (41). However, if M is compact, then any Riemannian metric g gives a complete D_0. Moreover, if D_0 is complete for some g, then D is complete too.

The following examples show that the completeness hypothesis is stronger than necessary to ensure (40), but that it does not always hold.

Example 5.11 Suppose that $M =]0,1[^2$ and $D = \partial/\partial x_1$. Then

$$\varrho_D(x,y) = \begin{cases} |x_1 - y_1| & \text{if } x_2 = y_2, \\ \infty & \text{otherwise,} \end{cases}$$

and (40) holds without ϱ_D being varietal or D being complete.

Example 5.12 Suppose that $M = \mathbb{R}^2 \setminus \{(0,0)\}$ and $D = \partial/\partial x_1$. Then

$$\varrho_D\big((-1,0),(1,0)\big) = \infty.$$

However, for all smooth $\xi : M \to \mathbb{R}$, the condition that $\|D\xi\|_\infty \leqslant 1$ implies that $|\xi(-1,t) - \xi(1,t)| \leqslant 2$ when $t \neq 0$ by the mean value theorem, and hence $|\xi(-1,0) - \xi(1,0)| \leqslant 2$ by continuity.

6 The L^2 Theory: Formal and Essential Self-adjointness

Given a differential operator $D \in \mathfrak{D}_k(\mathscr{E},\mathscr{F})$, we denote for the moment by D_s its restriction to compactly-supported smooth sections and by D_d its extension to distributions. Then D_s may be thought of as a densely defined operator $L^2(\mathscr{E}) \dashrightarrow L^2(\mathscr{F})$, and we may consider its Hilbert space adjoint $(D_s)^* : L^2(\mathscr{F}) \dashrightarrow L^2(\mathscr{E})$. It is easily checked that the domain of $(D_s)^*$ is the space $W^2_{D^+}(\mathscr{F})$, that is,

$$\big\{ f \in L^2(\mathscr{F}) : D_d^+ f \in L^2(\mathscr{E}) \big\},$$

and $(D_s)^*$ is the restriction of D_d^+ to this domain. In particular, $(D_s)^* \supseteq D_s^+$, so $(D_s)^*$ is densely defined, D_s is closable and

$$(D_s)^- = (D_s)^{**} \subseteq \big(D_s^+\big)^*.$$

The domains of $(D_s)^-$ and $(D_s^+)^*$ will be called the minimal and maximal domains of D respectively; note that the maximal domain of D is $W^2_D(\mathscr{E})$, whereas the minimal domain of D is $W^2_{D,0}(\mathscr{E})$. If D is formally self-adjoint, that is, $\mathscr{E} = \mathscr{F}$ and $D = D^+$, then

$$(D_s)^- = (D_s)^{**} \subseteq (D_s)^*,$$

and clearly $(D_s)^-$ and $(D_s)^*$ are the minimal and maximal closed symmetric extensions of D_s respectively; the essential self-adjointness of D_s is thus equivalent to the equality of the minimal and maximal domains of D.

Henceforth, we will write D, D^* and D^- instead of D_s, $(D_s)^*$ and $(D_s)^-$.

We now rephrase the content of Proposition 5.5 when $p = 2$.

Proposition 6.1 *Suppose that $D \in \mathfrak{D}_1(\mathscr{E}, \mathscr{F})$. A compactly-supported section $f \in L^2(\mathscr{E})$ belongs to the maximal domain of D if and only if it belongs to its minimal domain. If moreover D is complete, then the minimal and maximal domains of D coincide.*

Corollary 6.2 *Suppose that $D \in \mathfrak{D}_1(\mathscr{E}, \mathscr{E})$ is complete and formally self-adjoint. Then D is essentially self-adjoint.*

7 Finite Propagation Speed

Take a formally self-adjoint element $D \in \mathfrak{D}_1(\mathscr{E}, \mathscr{E})$. We say that u_t is a *solution* of

$$\frac{\mathrm{d}}{\mathrm{d}t} u_t = \mathrm{i} D^* u_t \tag{42}$$

if $t \mapsto u_t$ is an $L^2(\mathscr{E})$-valued map defined on a subinterval I of $\mathbb{R}$, which is continuously differentiable on I (as an $L^2(\mathscr{E})$-valued map), takes its values in the domain of D^* and satisfies (42) for all $t \in I$; we say moreover that u_t is *energy-preserving* if $t \mapsto \|u_t\|_2$ is constant.

If D admits a self-adjoint extension $\widetilde{D}$, so

$$D^- \subseteq \widetilde{D} = (\widetilde{D})^* \subseteq D^*,$$

then $u_t = \mathrm{e}^{\mathrm{i}t\widetilde{D}} f$ is an energy-preserving solution of (42) for all f in the domain of $\widetilde{D}$, since $\mathrm{e}^{\mathrm{i}t\widetilde{D}}$ is unitary. In fact, $u_t = \mathrm{e}^{\mathrm{i}t\widetilde{D}} f$ is defined for an arbitrary $f \in L^2(\mathscr{E})$, but need not be differentiable in t, and satisfies an integral version of the (42), that is,

$$u_t = u_0 + \mathrm{i} D^* \int_0^t u_s \,\mathrm{d}s$$

[8, Lemma II.1.3]; however such a "mild solution" of (42) may be approximated by "classical solutions" because the domain of $\widetilde{D}$ is dense in $L^2(\mathscr{E})$.

In any case, for an arbitrary D, compactly-supported solutions automatically preserve energy.

Proposition 7.1 *If u_t is a solution of (42), defined on an interval I, and $\operatorname{supp} u_t$ is compact for all $t \in I$, then u_t is energy-preserving.*

Proof The function $t \mapsto \|u_t\|_2^2$ is differentiable, with derivative

$$\mathrm{i}\langle\!\langle D^* u_t, u_t \rangle\!\rangle - \mathrm{i}\langle\!\langle u_t, D^* u_t \rangle\!\rangle.$$

Since $\operatorname{supp} u_t$ is compact, u_t is in the domain of D^- by Proposition 6.1, therefore

$$\langle\!\langle u_t, D^* u_t \rangle\!\rangle = \langle\!\langle D^- u_t, u_t \rangle\!\rangle = \langle\!\langle D^* u_t, u_t \rangle\!\rangle,$$

and hence the derivative of $t \mapsto \|u_t\|_2^2$ is identically null. □

The relationship between preservation of energy and compactness of support may be partially reversed.

Theorem 7.2 *Suppose that $K \subseteq W$, where $K \in \mathfrak{K}(M)$ and $W \in \mathfrak{O}(M)$. There exists ε, depending on K and W, such that, for all energy-preserving solutions u_t of* (42) *defined on an interval I containing 0, if*

$$\operatorname{supp} u_0 \subseteq K$$

then

$$\operatorname{supp} u_t \subseteq W$$

for all $t \in I \cap \,]-\varepsilon, \varepsilon[$.

Proof See [14, Proposition 10.3.1].

Choose a bump function g that is equal to 1 on K and supported in W, and take ε less than $\| \, |D^\sigma g|_{\mathrm{op}} \|_\infty^{-1}$. Choose also a smooth non-decreasing function $\varphi \colon \mathbb{R} \to [0, 1]$ such that $\varphi(t) < 1$ when $t < 1$ and $\varphi(t) = 1$ when $t \geqslant 1$.

Define functions $h_t \colon M \to [0, 1]$ for all $t \in [0, \infty[$ by

$$h_t(x) = \varphi\big(g(x) + \varepsilon^{-1} t\big).$$

Now h_t depends smoothly on t, with derivative (relative to t) given by

$$\dot{h}_t(x) = \varepsilon^{-1} \varphi'\big(g(x) + \varepsilon^{-1} t\big) \geqslant 0,$$

since φ is smooth and non-decreasing; further, by the chain rule,

$$D^\sigma h_t(x) = \varphi'\big(g(x) + \varepsilon^{-1} t\big)\, D^\sigma g(x) = \varepsilon\, \dot{h}_t(x)\, D^\sigma g(x).$$

Consequently,

$$\dot{h}_t - \mathrm{i} D^\sigma h_t = \dot{h}_t \big(1 - \mathrm{i}\varepsilon D^\sigma g\big) \geqslant 0 \tag{43}$$

pointwise as a section of $\mathrm{Hom}(\mathscr{E}, \mathscr{E})$, since $\mathrm{i}\varepsilon D^\sigma g$ is pointwise self-adjoint by (10), and $\| \, |\mathrm{i}\varepsilon D^\sigma g|_{\mathrm{op}} \|_\infty \leqslant 1$.

Take u_t as in the hypotheses. Then, by Leibniz' rule,

$$D(h_t u_t) = \big(D^\sigma h_t\big) u_t + h_t D u_t.$$

Now $h_t u_t \in L^2$ and $D(h_t u_t) \in L^2$ since $u_t \in L^2$, $D u_t \in L^2$, $h_t \in L^\infty$, and $D^\sigma h_t \in L^\infty$. Moreover $h_t u_t$ is compactly-supported, so belongs to the domain of D^- by Proposition 6.1, and

$$\big\langle\!\big\langle h_t u_t, D^* u_t \big\rangle\!\big\rangle = \big\langle\!\big\langle D^-(h_t u_t), u_t \big\rangle\!\big\rangle = \big\langle\!\big\langle D^*(h_t u_t), u_t \big\rangle\!\big\rangle,$$

hence

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\langle h_t u_t, u_t\rangle\rangle = \langle\langle \dot{h}_t u_t, u_t\rangle\rangle + \mathrm{i}\big\langle\big\langle h_t D^* u_t, u_t\big\rangle\big\rangle - \mathrm{i}\big\langle\big\langle h_t u_t, D^* u_t\big\rangle\big\rangle$$
$$= \langle\langle \dot{h}_t u_t, u_t\rangle\rangle - \mathrm{i}\big\langle\big\langle \big(D^\sigma h_t\big) u_t, u_t\big\rangle\big\rangle \geqslant 0$$

by Leibniz' rule and (43). Therefore, for all positive t,

$$\langle\langle h_t u_t, u_t\rangle\rangle \geqslant \langle\langle h_0 u_0, u_0\rangle\rangle = \langle\langle u_0, u_0\rangle\rangle = \langle\langle u_t, u_t\rangle\rangle,$$

since u_t is energy-preserving, $\operatorname{supp} u_0 \subseteq K$ and h_0 is equal to 1 on K. But then $h_t u_t = u_t$ almost everywhere, since $0 \leqslant h_t \leqslant 1$. Note that $h_t = \varphi(\varepsilon^{-1} t) < 1$ on the open set $M \setminus \operatorname{supp} g$ when $t < \varepsilon$; hence $\operatorname{supp} u_t \subseteq \operatorname{supp} g \subseteq W$.

The case where $t < 0$ may be treated by replacing D with $-D$ and u_t with u_{-t}. □

As a consequence, we establish the uniqueness of energy-preserving solutions of (42) for small times and compactly-supported initial datum.

Corollary 7.3 *Suppose that K is a compact subset of M. There exists $\varepsilon \in \mathbb{R}^+$, depending on K, such that, for all $f \in L^2(\mathscr{E})$ for which* $\operatorname{supp} f \subseteq K$, *two energy-preserving solutions u_t and v_t of* (42) *that satisfy $u_0 = v_0 = f$ coincide when $|t| < \varepsilon$. In particular, when $|t| < \varepsilon$, the value of $\mathrm{e}^{\mathrm{i}t\widetilde{D}} f$ does not depend on the self-adjoint extension $\widetilde{D}$ of D.*

Proof Take a relatively compact open neighbourhood W of K in M, and take ε, depending on K and W, as in Theorem 7.2.

Write w_t for $u_t - v_t$. The w_t is a solution of (42), and $\operatorname{supp} w_t \subseteq W$ when $|t| < \varepsilon$ by Theorem 7.2, therefore w_t is energy-preserving when $|t| < \varepsilon$ by Proposition 7.1, and the conclusion follows since $\|w_0\|_2 = 0$. □

7.1 Propagation and the Control Distance

By using the control distance function ϱ_D associated to $D = D^+ \in \mathfrak{D}_1(\mathscr{E}, \mathscr{E})$, we establish a quantitative version of Theorem 7.2. Recall Definition 4.9 of $R_D(K)$.

Theorem 7.4 *Suppose that K is a compact subset of M. If U is an energy-preserving solution of* (42) *defined on an interval I containing* 0 *and*

$$\operatorname{supp} u_0 \subseteq K,$$

then

$$\operatorname{supp} u_t \subseteq B^-_D\big(K, |t|\big)$$

for all $t \in I \cap \,]{-R_D(K)}, R_D(K)[$.

Proof It suffices to prove that $\operatorname{supp} u_t \subseteq B_D^-(K,\varepsilon)$ when $|t| < \varepsilon$, since

$$B_D^-\big(K,|t|\big) = \bigcap_{\delta>|t|} B_D^-(K,\delta).$$

Take any ε such that $|t| < \varepsilon < R_D(K)$. We now follow the proof of Theorem 7.2, with one modification: we define g, which is no longer smooth, by

$$g(x) = \big(1 - \varepsilon^{-1}\varrho_D(K,x)\big)_+ \quad \text{for every } x \in M.$$

Again by Proposition 4.8, $B_D^-(K,r)$ is compact when $r \leqslant \varepsilon$, and hence g is upper-semicontinuous; moreover it is clear that $0 \leqslant g \leqslant 1$, that $g = 1$ on K, that $\operatorname{supp} g \subseteq B_D^-(K,\varepsilon)$ and that

$$\big|g(x) - g(y)\big| \leqslant \varepsilon^{-1}\varrho_D(x,y),$$

so g is weakly D^σ-differentiable and $\| \, |D^\sigma g|_{\mathrm{op}}\|_\infty \leqslant \varepsilon^{-1}$ by Proposition 5.2. The steps of the proof of Theorem 7.2 may now be repeated, interpreting D^σ-derivatives in the weak sense, and using Propositions 3.7 and 3.8 whenever Leibniz' rule and the chain rule are invoked. □

A quantitative version of Corollary 7.3 on uniqueness of energy-preserving solutions may be derived as before. In fact, with a little more effort, we also establish an existence result. To avoid boundary value problems, we restrict attention to the interval $]-R_D(K), R_D(K)[$.

Theorem 7.5 *Suppose that $K \in \mathfrak{K}(M)$ and that $f \in W_D^2(\mathscr{E})$ is supported in K. Then there exists an energy-preserving solution u_t of (42) on the interval $]-R_D(K)$, $R_D(K)[$ such that $u_0 = f$; moreover, any other energy-preserving solution of (42) with initial datum f coincides with u_t on the intersection of their domains.*

Proof Every energy-preserving solution u_t such that $\operatorname{supp} u_0 \subseteq K$ remains compactly-supported when $|t| < R_D(K)$ by Theorem 7.4, so uniqueness of the solution on $]-R_D(K), R_D(K)[$ is proved as in Corollary 7.3. It remains to show the existence of an energy-preserving solution on all intervals $[-R, R]$ where $R < R_D(K)$.

Note that, if D is complete, then D^* is self-adjoint by Corollary 6.2, and $u_t = \mathrm{e}^{\mathrm{i}tD^*} f$ is the required solution. In the general case, take a bump function η that is equal to 1 on a neighbourhood of $B_D^-(K,R)$, and define $D_0 \in \mathfrak{D}_1(\mathscr{E},\mathscr{E})$ by

$$D_0 f = \frac{1}{2}\big(\eta Df + D(\eta f)\big) = \eta Df + \frac{1}{2}\big(D^\sigma \eta\big) f.$$

Then it is easily checked that D_0 is formally self-adjoint and $\sigma_1(D_0) = \eta\sigma_1(D)$ is compactly-supported, therefore D_0 is complete by Proposition 4.16, and we may take $u_t = \mathrm{e}^{\mathrm{i}tD_0^*} f$, which is an energy-preserving solution of

$$\frac{\mathrm{d}}{\mathrm{d}t} u_t = \mathrm{i} D_0^* u_t$$

for all $t \in \mathbb{R}$. Since $\sigma_1(D_0) = \eta\sigma_1(D)$ and $|\eta| \leqslant 1$, all D_0-subunit vectors are D-subunit, hence $\varrho_D \leqslant \varrho_{D_0}$ and consequently, by Theorem 7.4,

$$\operatorname{supp} u_t \subseteq B^-_{D_0}(K, R) \subseteq B^-_D(K, R)$$

for all $|t| \leqslant R$. Moreover, D and D_0 coincide on a neighbourhood of $B^-_D(K, R)$ by construction, therefore u_t is a solution of (42) when $|t| \leqslant R$. □

7.2 Second-Order Operators

Consider now the second-order equation

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^2 u_t = -L^* u_t, \tag{44}$$

for some positive $L \in \mathfrak{D}_2(\mathscr{E}, \mathscr{E})$. Suppose that $\widetilde{L}$ is a positive self-adjoint extension of L, denote the continuous extension of $\lambda \mapsto \lambda^{-1} \sin \lambda$ to $\mathbb{R}$ by sinc, and define u_t by

$$u_t = \cos\big(t\widetilde{L}^{1/2}\big) f + t \operatorname{sinc}\big(t\widetilde{L}^{1/2}\big) g. \tag{45}$$

It is well-known that u_t satisfies (44) together with the initial conditions $u_0 = f$ and $\dot{u}_0 = g$, at least when f is in the domain of $\widetilde{L}$ and g is in the domain of $\widetilde{L}^{1/2}$.

Suppose that L factorises as $L = D^+ D$ for some $D \in \mathfrak{D}_1(\mathscr{E}, \mathscr{F})$, and recall that $\not{D} = D \oplus D^+$. If $\widetilde{\not{D}}$ is any self-adjoint extension of $\not{D}$, then $\widetilde{\not{D}}^2$ preserves the decomposition of $L^2(\mathscr{E} \oplus \mathscr{F})$ as $L^2(\mathscr{E}) \oplus L^2(\mathscr{F})$, and $\widetilde{\not{D}}^2(f, 0) = (\widetilde{L} f, 0)$ for some positive self-adjoint extension $\widetilde{L}$ of L. In particular,

$$\begin{aligned}\big(\cos\big(t\widetilde{L}^{1/2}\big) f, 0\big) &= \cos\big(t\big(\widetilde{\not{D}}^2\big)^{1/2}\big)(f, 0) = \cos(t\widetilde{\not{D}})(f, 0) \\ &= \frac{\mathrm{e}^{\mathrm{i}t\widetilde{\not{D}}} + \mathrm{e}^{-\mathrm{i}t\widetilde{\not{D}}}}{2}(f, 0),\end{aligned}$$

because the cosine function is even, and moreover

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(t \operatorname{sinc}\big(t\widetilde{L}^{1/2}\big) g\big) = \cos\big(t\widetilde{L}^{1/2}\big) g.$$

Therefore if $\operatorname{supp} f \cup \operatorname{supp} g$ is compact and u_t is defined by (45), then from Theorem 7.4 we deduce that

$$\operatorname{supp} u_t \subseteq B^-_D\big(\operatorname{supp} f \cup \operatorname{supp} g, |t|\big)$$

whenever $|t| < R_D(\operatorname{supp} f \cup \operatorname{supp} g)$.

Note that

$$P_D(\xi) = \big|\sigma_1(D)(\xi)\big|_{\mathrm{op}} = \big(\big|\sigma_2(L)\big(\xi^{\odot 2}\big)\big|_{\mathrm{op}}\big)^{1/2}$$

by (3) and (5), so the fibre seminorm P_D and the associated distance function may be expressed directly in terms of the second-order symbol of L.

When P_D is complete, $\not{D}$ is essentially self-adjoint by Corollary 6.2. In fact the smoothness of solutions of symmetric hyperbolic systems with smooth coefficients (see [1, Sect. 7.6] for an elementary proof), together with the finite propagation speed, implies that the operators $e^{it\not{D}^*}$ preserve $C_c^\infty(\mathscr{E} \oplus \mathscr{F})$, and an argument of Chernoff [6, Lemma 2.1] proves that $\not{D}^2$ is essentially self-adjoint too. In particular, if $L = D^+D$, then L is essentially self-adjoint, and $W_L^2(\mathscr{E}) \subseteq W_D^2(\mathscr{E})$ with continuous inclusion, because

$$\langle\langle Df, Df\rangle\rangle = \langle\langle f, Lf\rangle\rangle \leqslant \|f\|_2 \|Lf\|_2$$

for every f in the maximal domain of L. It is then not difficult to deduce that, for all maps $t \mapsto u_t$ in $C^2(I; L^2(\mathscr{E}))$ that satisfy (44), the equality

$$v_t = (\dot{u}_t, \mathrm{i}Du_t)$$

defines a mild solution of

$$\frac{\mathrm{d}}{\mathrm{d}t} v_t = \mathrm{i}\not{D}^* v_t,$$

and consequently $v_t = e^{it\not{D}^*} v_0$ (see [8, Propositions VI.3.2 and II.6.4]). This implies that (44) has a unique solution for given initial data $u_0 = f$ and $\dot{u}_0 = g$, that is,

$$u_t = \cos\bigl(t\bigl(L^*\bigr)^{1/2}\bigr) f + t \operatorname{sinc}\bigl(t\bigl(L^*\bigr)^{1/2}\bigr) g.$$

8 Examples

This section contains examples that illustrate our theory. We begin with a discussion of multilinear algebra, then pass to the examples. Most of these are concerned with applications to differential operators, but the final example shows that smooth subunit parametrisations of smooth curves may not enable us to compute length.

8.1 Preliminaries on Multilinear Algebra

Suppose that V is an n-dimensional vector space over $\mathbb{C}$. As usual, if $\{v_1, \dots, v_n\}$ is a basis of V and $J = \{j_1, \dots, j_k\}$, where $1 \leqslant j_1 < \cdots < j_k \leqslant n$, then we define the element v_J of the exterior algebra $\bigoplus_{k=0}^n \Lambda^k V$ by

$$v_J = v_{j_1} \wedge \cdots \wedge v_{j_k}.$$

When V is endowed with a hermitean inner product $\langle \cdot, \cdot \rangle$, there exists a unique hermitean inner product $\langle \cdot, \cdot \rangle$ on the exterior algebra ΛV such that $\Lambda^k V \perp \Lambda^{k'} V$

when $k \neq k'$ and, for every orthonormal basis $\{v_1, \ldots, v_n\}$ of V, the multivectors v_J, where J varies over the k-element subsets of $\{1 \ldots, n\}$, form an orthonormal basis of $\Lambda^k V$ when $k = 0, \ldots, n$.

Given any $\alpha, \beta \in \Lambda V$, we define $\alpha \vee \beta \in \Lambda V$ by requiring that

$$\langle \alpha \vee \beta, \gamma \rangle = \langle \beta, \alpha \wedge \gamma \rangle \quad \text{for every } \gamma \in \Lambda V.$$

The map $(\alpha, \beta) \mapsto \alpha \vee \beta$ is sesquilinear (conjugate-linear in the first variable), and moreover

$$\langle \alpha \wedge \beta, \alpha \vee \gamma \rangle = \langle \alpha \wedge \alpha \wedge \beta, \gamma \rangle = 0,$$

so

$$|\alpha \wedge \beta + \alpha \vee \gamma|^2 = |\alpha \wedge \beta|^2 + |\alpha \vee \gamma|^2.$$

Suppose now $\alpha \in \Lambda^1 V = V$. Then we set $\alpha = |\alpha| v_1$ and extend v_1 to an orthonormal basis $\{v_1, \ldots, v_n\}$ of V. If $\beta = \sum_J b_J v_J$ for some $b_J \in \mathbb{C}$, where J ranges over the subsets of $\{1, \ldots, n\}$, then

$$\alpha \wedge \beta = |\alpha| \sum_{J \not\ni 1} b_J v_{J \cup \{1\}} \quad \text{and} \quad \alpha \vee \beta = |\alpha| \sum_{J \not\ni 1} b_{J \cup \{1\}} v_J,$$

hence

$$|\alpha \wedge \beta - \alpha \vee \beta|^2 = |\alpha \wedge \beta|^2 + |\alpha \vee \beta|^2 = |\alpha|^2 |\beta|^2. \tag{46}$$

8.2 Riemannian Manifolds

Let M be an n-dimensional manifold. The exterior algebra ΛM over the complexified cotangent bundle $\mathbb{C}T^*M$ is the bundle $\bigoplus_{k=0}^m \Lambda^k M$; its sections are known as differential forms. In particular, $\Lambda^0 M = \mathscr{T}$, $\Lambda^1 M = \mathbb{C}T^*M$, and the differential $\mathrm{d} \in \mathfrak{D}_1(\Lambda^0 M, \Lambda^1 M)$ extends to the exterior derivative $\mathrm{d} \in \mathfrak{D}_1(\Lambda M, \Lambda M)$, which satisfies $\mathrm{d}^2 = 0$ and

$$\mathrm{d}\alpha \in C^\infty\big(\Lambda^{k+1} M\big) \quad \text{and} \quad \mathrm{d}(\alpha \wedge \beta) = \mathrm{d}\alpha \wedge \beta + (-1)^k \alpha \wedge \mathrm{d}\beta$$

for all $\alpha \in C^\infty(\Lambda^k M)$ and all $\beta \in C^\infty(\Lambda M)$. Hence

$$\big[\mathrm{d}, m(h)\big]\alpha = \mathrm{d}(h\alpha) - h\mathrm{d}\alpha = \mathrm{d}h \wedge \alpha,$$

for all $h \in C^\infty(\mathscr{T})$ and $\alpha \in C^\infty(\Lambda M)$, that is,

$$\sigma_1(\mathrm{d})(\xi)\beta = \xi \wedge \beta,$$

when $x \in M$, $\xi \in \mathbb{C}T_x^* M$, and $\beta \in \Lambda_x M$.

Suppose now that M is endowed with a Riemannian metric g. This defines a hermitean fibre inner product on $\mathbb{C}T^*M$, which in turn extends to a hermitean fibre inner product $\langle\cdot,\cdot\rangle$ on ΛM. The formal adjoint d^+ of the exterior derivative d is then defined, and satisfies

$$\sigma_1\big(\mathrm{d}^+\big)(\xi)\beta = -\bar{\xi} \vee \beta$$

where $x \in M$, $\xi \in \mathbb{C}T_x^*M$ and $\beta \in \Lambda_x M$, by (5). We set $D = \mathrm{d} + \mathrm{d}^+$. Then D is formally self-adjoint and

$$\sigma_1(D)(\xi)\beta = \xi \wedge \beta - \bar{\xi} \vee \beta,$$

so, when $\xi = \bar{\xi} \in T_x^*M$ is real,

$$\big|\sigma_1(D)(\xi)\beta\big| = |\xi||\beta|,$$

by (46), and

$$P_D(\xi) = \big|\sigma_1(D)(\xi)\big|_{\mathrm{op}} = |\xi|.$$

Thus the control distance function ϱ_D associated to D is just the Riemannian distance function ϱ_g on M.

Define $\Delta = D^2 = \mathrm{d}\mathrm{d}^+ + \mathrm{d}^+\mathrm{d}$. This is the Laplace operator on forms induced by the Riemannian structure. Hence, according to Sect. 7.2, when (M, g) is complete, the Riemannian distance also describes the propagation of the solution u_t of the second-order equation $\ddot{u}_t = -\Delta u_t$ given by

$$u_t = \cos\big(t\Delta^{1/2}\big)u_0 + t\,\mathrm{sinc}\big(t\Delta^{1/2}\big)\dot{u}_0.$$

Since Δ preserves the degree of forms, such a solution u_t is a k-form for all $t \in \mathbb{R}$ whenever the initial data u_0 and $\dot{u}_0$ are both k-forms.

8.3 Hermitean Complex Manifolds

Suppose now that M is a complex manifold of real dimension $2n$. The decomposition

$$\mathbb{C}T^*M = \Lambda^{1,0}M \oplus \Lambda^{0,1}M$$

given by the complex structure in turn induces a decomposition of $\Lambda^k M$, namely,

$$\Lambda^k M = \bigoplus_{p+q=k} \Lambda^{p,q}M;$$

then $\bigoplus \Lambda^{p,q}M$ is an algebra bigrading of ΛM. Let $\pi_{p,q} \in \mathrm{Hom}(\Lambda M, \Lambda M)$ denote the projection onto $\Lambda^{p,q}M$. The exterior derivative d decomposes as $\partial + \bar{\partial}$, where

$$\partial\alpha = \pi_{p+1,q}\,\mathrm{d}\alpha \quad \text{and} \quad \bar{\partial}\alpha = \pi_{p,q+1}\,\mathrm{d}\alpha \quad \text{for every } \alpha \in C^\infty\big(\Lambda^{p,q}M\big);$$

then $\partial^2 = \partial\bar\partial + \bar\partial\partial = \bar\partial^2 = 0$ and

$$\partial(\alpha\wedge\beta) = \partial\alpha\wedge\beta + (-1)^k \alpha\wedge\partial\beta \quad \text{and} \quad \bar\partial(\alpha\wedge\beta) = \bar\partial\alpha\wedge\beta + (-1)^k \alpha\wedge\bar\partial\beta$$

for all $\alpha \in C^\infty(\Lambda^k M)$, $\beta \in C^\infty(\Lambda M)$. As before,

$$\big[\partial, m(h)\big]\alpha = \partial h\wedge\alpha \quad \text{and} \quad \big[\bar\partial, m(h)\big]\alpha = \bar\partial h\wedge\alpha,$$

so

$$\sigma_1(\partial)(\xi)\beta = \pi_{1,0}\xi\wedge\beta \quad \text{and} \quad \sigma_1(\bar\partial)(\xi)\beta = \pi_{0,1}\xi\wedge\beta.$$

For any choice of Riemannian metric g on M,

$$\sigma_1\big(\partial^+\big)(\xi)\beta = -\pi_{1,0}\bar\xi\vee\beta \quad \text{and} \quad \sigma_1\big(\bar\partial^+\big)(\xi)\beta = -\pi_{0,1}\bar\xi\vee\beta,$$

by (5), hence also, when $\xi = \bar\xi$, that is, ξ is real,

$$\big|\sigma_1\big(\partial+\partial^+\big)(\xi)\beta\big| = |\pi_{1,0}\xi||\beta| \quad \text{and} \quad \big|\sigma_1\big(\bar\partial+\bar\partial^+\big)(\xi)\beta\big| = |\pi_{0,1}\xi||\beta|$$

by (46). In particular, if g is compatible with the complex structure (that is, the complex structure $J\colon T_xM \to T_xM$ is an isometry for every $x\in M$), then for real ξ,

$$|\pi_{1,0}\xi|^2 = |\pi_{0,1}\xi|^2 = |\xi|^2/2,$$

so the distance functions associated to $\partial+\partial^+$ and $\bar\partial+\bar\partial^+$ coincide with the Riemannian distance function on M multiplied by $\sqrt{2}$ (that is, the propagation speed with respect to the Riemannian distance is at most $1/\sqrt{2}$). The complex Laplacian $\Box$ on forms is given by

$$\Box = \big(\bar\partial+\bar\partial^+\big)^2 = \bar\partial\bar\partial^+ + \bar\partial^+\bar\partial;$$

when M is a Kähler manifold, $\Delta = 2\Box$, which is consistent with the result already obtained for Δ.

See [9, 23] for more on the material in this subsection.

8.4 CR Manifolds

Let M be an n-dimensional manifold with a CR structure of codimension $n-2k$, that is, an involutive complex subbundle $\mathscr{L}$ of $\mathbb{C}TM$ of rank k such that $\mathscr{L}_x \cap \overline{\mathscr{L}}_x = \{0\}$ for all x in M. The exterior algebra $\Lambda^{0,\bullet}M = \Lambda(\overline{\mathscr{L}}^*)$ over the dual of $\overline{\mathscr{L}}$ may be identified with the quotient of ΛM by a suitable graded fibre ideal $\mathscr{I}$. Correspondingly $C^\infty(\Lambda^{0,\bullet}M)$ may be identified with $C^\infty(\Lambda M)/C^\infty(\mathscr{I})$. The exterior derivative d passes to the quotient bundle, giving a differential operator $\bar\partial_b \in \mathfrak{D}_1(\Lambda^{0,\bullet}M, \Lambda^{0,\bullet}M)$ that satisfies

$$\bar\partial_b^2 = 0,$$

$$\bar{\partial}_b\alpha \in C^\infty\big(\Lambda^{0,q+1}M\big),$$

$$\bar{\partial}_b(\alpha \wedge \beta) = \bar{\partial}_b\alpha \wedge \beta + (-1)^q \alpha \wedge \bar{\partial}_b\beta$$

for all $\alpha \in C^\infty(\Lambda^{0,q}M)$ and all $\beta \in C^\infty(\Lambda^{0,\bullet}M)$.

Note that $\Lambda^{0,0}M = \Lambda^0 M = \mathscr{T}$, so $\bar{\partial}_b f = \pi\, \mathrm{d}f$, where $\pi : \mathbb{C}T^*M \to \overline{\mathscr{L}}^*$ is the restriction morphism. Thus

$$\big[\bar{\partial}_b, m(h)\big]\alpha = \bar{\partial}_b h \wedge \alpha \quad \text{and} \quad \sigma_1(\bar{\partial}_b)(\xi)\beta = \pi\xi \wedge \beta.$$

Any choice of hermitean fibre inner product on $\overline{\mathscr{L}}$ induces a hermitean inner product along the fibres of $\Lambda^{0,\bullet}M$, and

$$\sigma_1\big(\bar{\partial}_b^+\big)(\xi)\beta = -\pi\bar{\xi} \vee \beta,$$

so again, for all real ξ,

$$\big|\sigma_1\big(\bar{\partial}_b + \bar{\partial}_b^+\big)(\xi)\beta\big| = |\pi\xi||\beta|,$$

that is, if $D = \bar{\partial}_b + \bar{\partial}_b^+$, then

$$P_D(\xi) = |\pi\xi|.$$

If $\xi \in T^*M$, then $\pi\xi = 0$ if and only if ξ vanishes on $TM \cap (\mathscr{L} \oplus \overline{\mathscr{L}})$; in other words, the Levi distribution $TM \cap (\mathscr{L} \oplus \overline{\mathscr{L}})$ is the subbundle spanned by the D-subunit vectors. In particular, if M is a non-degenerate CR manifold and $n = 2k + 1$, then P_D satisfies Hörmander's condition; for a discussion of the higher-codimensional case, see, for example, [5, Sect. 12.1].

Note moreover that the Kohn Laplacian $\Box_b$ on the tangential Cauchy–Riemann complex is given by

$$\Box_b = D^2 = \bar{\partial}_b\bar{\partial}_b^+ + \bar{\partial}_b^+\bar{\partial}_b.$$

For more information on CR manifolds, see, for example, [5, 7].

8.5 Sub-Riemannian Structures

Let E be a real vector bundle on M, endowed with a fibre inner product and a smooth bundle homomorphism $\mu\colon E \to TM$. Consider the adjoint morphism $\mu^*\colon T^*M \to E^*$, and its complexification $\mu^*\colon \mathbb{C}T^*M \to \mathbb{C}E^*$. Define the differential operator $D \in \mathfrak{D}_1(\mathscr{T}, \mathbb{C}E^*)$ by $Df = \mu^*(\mathrm{d}f)$. Then $D^\sigma = D$, modulo the identification $\mathrm{Hom}(\mathscr{T}, \mathbb{C}E^*) = \mathbb{C}E^*$; further $P_D(\xi) = |\mu^*(\xi)|$, $P_D^*(v) = \inf\{|w| : v = \mu(w)\}$, and the D-subunit vectors are the images under μ of the $w \in E$ such that $|w| \leqslant 1$.

A commonly considered case is when E is a subbundle of TM and μ is the inclusion map. Then E is called the horizontal distribution [20, Sect. 1.4], and is the set of the tangent vectors v for which $P_D^*(v) < \infty$.

Another commonly considered case [12, 17] is when E is the trivial bundle $\mathscr{T}^r$ with the standard inner product. In this case, there are (subunit) vector fields $X_j = \mu(Y_j)$, where the Y_j are the constant sections of E corresponding to the standard basis of $\mathbb{R}^r$. Hence

$$P_D(\xi)^2 = \sum_j \left|\mu^*(\xi)(Y_j)\right|^2 = \sum_j \left|\xi(X_j)\right|^2,$$

so

$$\left|D^\sigma f\right|^2_{\mathrm{op}} = |Df|^2 = \sum_j |X_j f|^2,$$

and

$$P_D^*(v)^2 = \inf\left\{\sum_j c_j^2 : v = \sum_j c_j X_j|_x\right\} \quad \text{for every } v \in T_x M.$$

8.6 Non-Riemannian Propagation

The fibre seminorm P_D on T^*M associated to $D \in \mathfrak{D}_1(\mathscr{E}, \mathscr{F})$ is defined to be the pullback of an operator norm along the fibres of $\mathrm{Hom}(\mathscr{E}, \mathscr{F})$. In the previous examples, however, P_D is actually induced by some (possibly degenerate) inner product on T^*M. We present now a simple example showing that this is not always the case.

Let M be $\mathbb{R}^n$, take $\mathscr{E} = \mathscr{F} = \mathscr{T}^n$, and define D by

$$D(f_1, \dots, f_n) = (\mathrm{i}\partial_1 f_1, \dots, \mathrm{i}\partial_n f_n),$$

where $\partial_1, \dots, \partial_n$ are the partial derivatives on $\mathbb{R}^n$. Then

$$\sigma_1(D)(\xi) = \begin{pmatrix} \mathrm{i}\xi_1 & & \\ & \ddots & \\ & & \mathrm{i}\xi_n \end{pmatrix},$$

so $P_D(\xi) = |\xi|_\infty$ and $P_D^*(v) = |v|_1$; here, as usual, $|\xi|_\infty = \max_j |\xi_j|$ and $|v|_1 = \sum_j |v_j|$. Consequently, $\varrho_D(x, y) = |x - y|_1$, hence ϱ_D is varietal and D is complete, therefore D is essentially self-adjoint, and

$$\mathrm{e}^{\mathrm{i}tD}(f_1, \dots, f_n)(x) = \big(f_1(x_1 - t, x_2, \dots, x_n), \dots, f_n(x_1, \dots, x_{n-1}, x_n - t)\big).$$

Hence the condition $\mathrm{supp}(\mathrm{e}^{\mathrm{i}tD} f) \subseteq B_D^-(\mathrm{supp}\, f, |t|)$ given by Theorem 7.4 is optimal. This shows that the natural distance describing the propagation of solutions of (42) need not be Riemannian or even sub-Riemannian.

8.7 Nonsmooth Arc-Length Reparametrisation

Construction of the Sub-Riemannian Structure Take $M = \mathbb{R}^2$ with Lebesgue measure.

Fix a smooth $u\colon \mathbb{R}^2 \to \mathbb{R}$. Define the smooth vector fields X, Y on $\mathbb{R}^2$ by

$$X|_p = \frac{2}{\sqrt{4+3u(p)^2}}\left(\frac{\partial}{\partial x} + u(p)\frac{\sqrt{3}}{2}\frac{\partial}{\partial y}\right),$$

$$Y|_p = \frac{u(p)}{2\sqrt{1+u(p)^2}}\frac{2}{\sqrt{4+3u(p)^2}}\left(-u(p)\frac{\sqrt{3}}{2}\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right),$$

where $\{\partial/\partial x, \partial/\partial y\}$ denotes the standard basis of $\mathbb{R}^2$. With respect to the standard Riemannian (that is, euclidean) structure of $\mathbb{R}^2$,

$$\langle X, Y\rangle = 0, \qquad |X| = 1, \quad \text{and} \quad |Y| = \frac{|u|}{2\sqrt{1+u^2}}$$

at every point of $\mathbb{R}^2$. Indeed, if we define the "matrix field" M by

$$M|_p = \frac{1}{\sqrt{4+3u(p)^2}}\begin{pmatrix} 2 & -u(p)\sqrt{3} \\ u(p)\sqrt{3} & 2\end{pmatrix},$$

then M is pointwise orthogonal and

$$X = M\frac{\partial}{\partial x}, \qquad Y = \frac{u}{2\sqrt{1+u^2}}M\frac{\partial}{\partial y}$$

pointwise.

Define the differential operator $D \in \mathfrak{D}_1(\mathscr{T}, \mathscr{T}^2)$ by $Df = (Xf, Yf)$. Then

$$\sigma_1(D)|_p(\xi) = \big(\langle X|_p, \xi\rangle, \langle Y|_p, \xi\rangle\big),$$

hence the associated fibre seminorm on the cotangent bundle $T^*\mathbb{R}^2$ is given by

$$\begin{aligned} P_D|_p(\xi)^2 &= \big|\sigma_1(D)|_p(\xi)\big|_{\mathrm{op}}^2 = \langle X|_p, \xi\rangle^2 + \langle Y|_p, \xi\rangle^2 \\ &= \langle \xi, H|_p\,\xi\rangle, \end{aligned}$$

where

$$H = \frac{1}{4(1+u^2)}\begin{pmatrix} 4+u^2 & 2\sqrt{3}u \\ 2\sqrt{3}u & 4u^2\end{pmatrix}.$$

On the one hand, at points p where $u(p) \neq 0$, the matrix $H|_p$ is non-degenerate; in this case, the norm P_D^* on the tangent bundle is given by

$$P_D|_p(v)^2 = \big\langle v, H|_p^{-1}v\big\rangle,$$

where

$$H^{-1} = \frac{1}{u^2}\begin{pmatrix} 4u^2 & -2\sqrt{3}u \\ -2\sqrt{3}u & 4+u^2 \end{pmatrix},$$

and $\{X|_p, Y|_p\}$ is an orthonormal basis for the corresponding inner product on $T_p\mathbb{R}^2$.

On the other hand, at points p where $u(p) = 0$,

$$P_D|_p(\xi) = \left|\left\langle \frac{\partial}{\partial x}, \xi \right\rangle\right|,$$

hence P_D^* is the extended norm

$$P_D^*|_p(v) = \begin{cases} |\langle \partial/\partial x, v\rangle| & \text{if } \langle \partial/\partial y, v\rangle = 0, \\ \infty & \text{otherwise,} \end{cases}$$

and $X|_p = \partial/\partial x$ and $Y|_p = 0$. In particular,

$$P_D^*|_p\left(\frac{\partial}{\partial x}\right) = \begin{cases} 2 & \text{if } u(p) \neq 0, \\ 1 & \text{if } u(p) = 0. \end{cases}$$

A Choice of *u* Let $\mathbb{Q} = \{q_m\}_{m\in\mathbb{N}}$ be an enumeration of the rational numbers, and set

$$A = \bigcup_{m\in\mathbb{N}}]q_m - 2^{-m-3}, q_m + 2^{-m-3}[.$$

Then A is a dense open subset of $\mathbb{R}$ whose measure $|A|$ is at most $\sum_{m=0}^{\infty} 2^{-m-2}$, that is, $1/2$.

Since $\mathbb{R} \setminus A$ is closed in $\mathbb{R}$, there exists a smooth function $v\colon \mathbb{R} \to [0, 1]$ such that $v^{-1}(0) = \mathbb{R} \setminus A$ [18, Theorem 1.5]. In fact, after composing v with a smooth function from $\mathbb{R}$ to [0, 1] that vanishes exactly on $]-\infty, 0]$, we may suppose that v vanishes to infinite order at all points of $\mathbb{R} \setminus A$. Set then $u(x, y) = v(x)$.

Hörmander's Condition Let Z be a D-subunit field. Then $Z = \varphi X + \psi Y$ for some real-valued functions φ, ψ with $\varphi^2 + \psi^2 = 1$. Since $\langle Z, X\rangle = \varphi$, we see that φ is smooth, so φX and ψY are smooth too. Moreover, since $|\psi| \leqslant 1$, the smooth field ψY vanishes at least to the same order as Y, at every point of $\mathbb{R}^2$, and hence ψY vanishes to infinite order at every point of $(\mathbb{R} \setminus A) \times \mathbb{R}$.

Take now a system $Z_1, \dots, Z_r$ of D-subunit vector fields, and decompose Z_j as $\varphi_j X + \psi_j Y$. Then any iterated Lie bracket of $Z_1, \dots, Z_r$ is the sum of an iterated Lie bracket of $\varphi_1 X, \dots, \varphi_r X$ and of iterated Lie brackets where some of the $\psi_j Y$ occur. The first summand is then a smooth multiple of X, whereas the other summands vanish to infinite order at every point of $(\mathbb{R} \setminus A) \times A$ (indeed, the set of smooth vector fields vanishing to infinite order at some $p \in M$ is an ideal of the Lie algebra of smooth vector fields). We conclude that the iterated Lie bracket of $Z_1, \dots, Z_r$, evaluated at any point of $(\mathbb{R} \setminus A) \times \mathbb{R}$, is a multiple of X.

Hence Hörmander's condition for P_D fails at all points of $(\mathbb{R} \setminus A) \times \mathbb{R}$.

Topologies Define $Z = 2^{-1}\partial/\partial x$ and $W = u(\sqrt{4+u^2})^{-1}\partial/\partial y$, and then set $\mathfrak{X} = \{Z, W\}$. Then $\mathfrak{X}$ is a system of smooth D-subunit vector fields on $\mathbb{R}^2$. Write $\varrho_{\mathfrak{X}}$ for the distance function corresponding to the class of D-subunit curves that are piecewise flow curves of Z or W. Clearly

$$\varrho_D \leqslant \varrho_D^{\infty} \leqslant \varrho_D^{\text{flow}} \leqslant \varrho_{\mathfrak{X}}.$$

We now show that $\varrho_{\mathfrak{X}}$ is varietal, so all the other distance functions above are.

Take $(x, y) \in \mathbb{R}^2$ and $r \in \mathbb{R}^+$. We want to prove that $B^-_{\mathfrak{X}}((x, y), r)$ is a neighbourhood of (x, y). Since A is dense in $\mathbb{R}$, there is $x' \in A$ such that $|x - x'| < r/8$ and $v(x') \neq 0$. We claim that every point $(\tilde{x}, \tilde{y}) \in \mathbb{R}^2$ such that

$$\left|(\tilde{x}, \tilde{y}) - (x, y)\right|_{\infty} < \min\left\{\frac{r}{8}, \frac{r}{4}\frac{|v(x')|}{\sqrt{4+v(x')^2}}\right\}$$

belongs to $B^-_{\mathfrak{X}}((x, y), r)$. The idea is to go from (x, y) to (x', y) along the flow of Z, then from (x', y) to $(x', \tilde{y})$ along the flow of W, and finally from $(x', \tilde{y})$ to $(\tilde{x}, \tilde{y})$ along the flow of Z. Such a curve is defined on an interval of length

$$2\left|x - x'\right| + \frac{\sqrt{4+v(x')^2}}{|v(x')|}|y - \tilde{y}| + 2\left|x' - \tilde{x}\right| < \frac{r}{4} + \frac{r}{4} + \frac{r}{2} = r,$$

hence its final point $(\tilde{x}, \tilde{y})$ belongs to the ball $B^-_{\mathfrak{X}}((x, y), r)$.

A Smooth Curve with Nonsmooth Arc-Length Let $\varphi\colon [0, T] \to \mathbb{R}$ be absolutely continuous, and define $\gamma(t) = (\varphi(t), 0)$. Then

$$P_D^*\big|_{\gamma(t)}\big(\gamma'(t)\big) = \begin{cases} 2|\varphi'(t)| & \text{if } \varphi(t) \in A, \\ |\varphi'(t)| & \text{if } \varphi(t) \notin A, \end{cases}$$

at every point t where φ is differentiable.

The set $\widetilde{A} = \varphi^{-1}(A)$ is open in $[0, T]$, but need not be dense. However, any connected subset of $[0, T] \setminus \widetilde{A}$ is mapped by φ onto a connected subset of $\mathbb{R} \setminus A$, which has at most one element because A is dense. Hence φ is locally constant on the interior of $[0, T] \setminus \widetilde{A}$, so $\varphi'(t) = 0$ for every interior point t of $[0, T] \setminus \widetilde{A}$. The remaining points of $[0, T] \setminus \widetilde{A}$, that is, the boundary points, also belong to the closure of $\widetilde{A}$.

Suppose now that γ is D-subunit and C^1. Then $|\varphi'(t)| \leqslant 1/2$ for all $t \in \widetilde{A}$. Since φ' is continuous, $|\varphi'| \leqslant 1/2$ on the closure of $\widetilde{A}$, hence on all $[0, T]$. This means that

$$\left|\varphi(T) - \varphi(0)\right| \leqslant \int_0^T \left|\varphi'(t)\right| \mathrm{d}t \leqslant T/2,$$

that is, $T \geqslant 2|\varphi(T) - \varphi(0)|$.

Suppose further that φ is non-decreasing, $\varphi(0) = 0$ and $\varphi(T) = 1$, so $T \geqslant 2$. The length $\ell_D(\gamma)$ does not depend on the parametrisation. Therefore, if we define $\tilde{\gamma}(t) = (t, 0)$ for all $t \in [0, 1]$, then

$$\begin{aligned}\ell_D(\gamma) = \ell_D(\tilde{\gamma}) &= \int_0^1 P_D^*\big|_{\tilde{\gamma}(t)}\big(\tilde{\gamma}'(t)\big)\,\mathrm{d}t \\ &= 2\big|[0, 1] \cap A\big| + \big|[0, 1] \setminus A\big| \\ &= 1 + \big|[0, 1] \cap A\big| \leqslant 3/2 < 2.\end{aligned}$$

In summary, every D-subunit, C^1 reparametrisation of γ is defined on an interval of width at least 2. By contrast, the arc-length reparametrisation of γ is D-subunit and defined on an interval of width at most $3/2$.

Acknowledgements This research was supported under the Australian Research Council's Discovery Projects funding scheme (project number DP-110102488).

References

1. Alinhac, S.: Hyperbolic Partial Differential Equations. Universitext. Springer, Dordrecht (2009)
2. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures, 2nd edn. Lectures in Mathematics ETH Zürich. Birkhäuser, Basel (2008)
3. Ambrosio, L., Tilli, P.: Topics on Analysis in Metric Spaces. Oxford Lecture Series in Mathematics and Its Applications, vol. 25. Oxford University Press, Oxford (2004)
4. Berline, N., Getzler, E., Vergne, M.: Heat Kernels and Dirac Operators. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 298. Springer, Berlin (1992)
5. Boggess, A.: CR Manifolds and the Tangential Cauchy–Riemann Complex. Studies in Advanced Mathematics. CRC Press, Boca Raton (1991)
6. Chernoff, P.R.: Essential self-adjointness of powers of generators of hyperbolic equations. J. Funct. Anal. **12**, 401–414 (1973)
7. Dragomir, S., Tomassini, G.: Differential Geometry and Analysis on CR Manifolds. Progress in Mathematics, vol. 246. Birkhäuser Boston, Boston (2006)
8. Engel, K.J., Nagel, R.: One-Parameter Semigroups for Linear Evolution Equations. Graduate Texts in Mathematics, vol. 194. Springer, New York (2000). With contributions by S. Brendle, M. Campiti, T. Hahn, G. Metafune, G. Nickel, D. Pallara, C. Perazzoli, A. Rhandi, S. Romanelli and R. Schnaubelt
9. Folland, G.B., Kohn, J.J.: The Neumann Problem for the Cauchy–Riemann Complex. Annals of Mathematics Studies, vol. 75. Princeton University Press, Princeton (1972)
10. Friedrichs, K.O.: The identity of weak and strong extensions of differential operators. Trans. Am. Math. Soc. **55**, 132–151 (1944)
11. Garofalo, N., Nhieu, D.M.: Isoperimetric and Sobolev inequalities for Carnot-Carathéodory spaces and the existence of minimal surfaces. Commun. Pure Appl. Math. **49**, 1081–1144 (1996)
12. Garofalo, N., Nhieu, D.M.: Lipschitz continuity, global smooth approximations and extension theorems for Sobolev functions in Carnot-Carathéodory spaces. J. Anal. Math. **74**, 67–97 (1998)

13. Gilbarg, D., Trudinger, N.S.: Elliptic Partial Differential Equations of Second Order. Classics in Mathematics. Springer, Berlin (2001). Reprint of the 1998 edn.
14. Higson, N., Roe, J.: Analytic K-Homology. Oxford Mathematical Monographs. Oxford University Press, Oxford (2000)
15. Hirsch, M.W.: Differential Topology. Graduate Texts in Mathematics, vol. 33. Springer, New York (1994). Corrected reprint of the 1976 original
16. Husemoller, D.: Fibre Bundles, 3rd edn. Graduate Texts in Mathematics, vol. 20. Springer, New York (1994)
17. Jerison, D., Sánchez-Calle, A.: Subelliptic, second order differential operators. In: Complex Analysis, III, College Park, MD, 1985–1986. Lect. Notes in Mathematics, vol. 1277, pp. 46–77. Springer, Berlin (1987)
18. Kahn, D.W.: Introduction to Global Analysis. Pure and Applied Mathematics, vol. 91. Academic Press [Harcourt Brace Jovanovich], New York (1980)
19. Köthe, G.: Topological Vector Spaces. I. Grundlehren der Mathematischen Wissenschaften, vol. 159. Springer, New York (1969). Translated from the German by D.J.H. Garling
20. Montgomery, R.: A Tour of Subriemannian Geometries, Their Geodesics and Applications. Mathematical Surveys and Monographs, vol. 91. Am. Math. Soc., Providence (2002)
21. Palais, R.S.: Seminar on the Atiyah-Singer Index Theorem. With Contributions by M.F. Atiyah, A. Borel, E.E. Floyd, R.T. Seeley, W. Shih and R. Solovay. Annals of Mathematics Studies, vol. 57. Princeton University Press, Princeton (1965)
22. Roe, J.: Lectures on Coarse Geometry. University Lecture Series, vol. 31. Am. Math. Soc., Providence (2003)
23. Wells, R.O. Jr.: Differential Analysis on Complex Manifolds, 3rd edn. Graduate Texts in Mathematics, vol. 65. Springer, New York (2008). With a new appendix by Oscar Garcia-Prada

On the Boundary Behavior of Holomorphic and Harmonic Functions

Fausto Di Biase

Abstract We present old and new problems in the study of the boundary behavior of holomorphic and harmonic functions. The key issue is the *sharpness* of the natural approach regions that appear in qualitative results (such as the Fatou-type theorems on the a.e. convergence) as well as in quantitative results (such as the Hardy–Littlewood inequality). Roughly speaking, we address the following question: Are there *larger* approach regions for which these results hold? The answer turns out to depend on the meaning we assign to the notion of being larger.

Keywords Harmonic functions · Holomorphic functions · Approach regions · Boundary behavior · Independence results · NTA domains · Hardy spaces

Mathematics Subject Classification (2010) Primary 32A40 · Secondary 31A20

1 Starting from the Unit Disc

The unit disc $\mathbb{D}$ lies at the confluence of two subjects, or viewpoints, just as the city of St. Louis sits at the confluence of two rivers. The first viewpoint is complex analysis, where we look at $\mathbb{D}$ as the set

$$\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}.$$

Here $\mathbb{D}$ is a ground for holomorphic functions. The second viewpoint is potential theory (and real-variables methods), where we look at $\mathbb{D}$ as the set

$$\mathbb{D} = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 < 1\}.$$

Here $\mathbb{D}$ is a ground for harmonic functions. Indeed, harmonic functions and holomorphic functions do exhibit a different behavior, but this is true in $\mathbb{C}$ much less than in $\mathbb{C}^n$. Moreover, the unit disc also serves as spring for the natural stream of classical Fourier analysis.

F. Di Biase (✉)
Dipartimento di Economia, Università “G. D’Annunzio”, Viale Pindaro 42, 65127 Pescara, Italy
e-mail: fdibiase@unich.it

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3,
DOI 10.1007/978-88-470-2853-1_9,

1.1 The First Viewpoint

We start with a basic theorem of Fatou [17], where $\mathbb{D}$ is seen from the first viewpoint. In order to state this result, we need some terminology.

$\mathbb{T}$	The topological boundary of $\mathbb{D}$ is denoted by $\mathbb{T}$. Thus $\mathbb{T} = \{z \in \mathbb{C} : \lvert z\rvert = 1\}$.
$\mathscr{P}(\mathbb{D})$	The set of all subsets of $\mathbb{D}$ is denoted by $\mathscr{P}(\mathbb{D})$.
$\mathscr{O}(\mathbb{D})$	The set of holomorphic functions on $\mathbb{D}$ is denoted by $\mathscr{O}(\mathbb{D})$.

Definition 1.1 An *approach region in* $\mathbb{D}$ *at* $w \in \mathbb{T}$ is an element of $\mathscr{P}(\mathbb{D})$ whose closure (in the ambient space $\mathbb{C}$) contains w. Let $\mathbb{D}_w$ be the set of all approach regions in $\mathbb{D}$ at w.

According to this definition, an approach region need not be open. If $A \in \mathbb{D}_w$, and $\varphi : \mathbb{D} \to \mathbb{C}$, then the notion of the limit value $\lim_{A \ni z \to w} \varphi(z)$ can be defined in the familiar fashion. Here the relevant part of A that matters is the one that lies close to w: We call it a *tail* of A.

Definition 1.2 If $A \in \mathbb{D}_w$, a *tail* of A (at w) is a subset of A of the form

$$\left\{z \in \mathbb{D} : z \in A,\ |z - w| < r\right\}$$

for some $r > 0$.

Definition 1.3 Any open Euclidean triangle ∇ of vertex $w \in \mathbb{T}$, contained in $\mathbb{D}$, is called a *Stolz approach region at* w *in* $\mathbb{D}$. We denote by STOLZ_w the collection of all the Stolz approach regions at w in $\mathbb{D}$. Hence $\mathrm{STOLZ}_w \subset \mathbb{D}_w$.

If $z \in \mathbb{D}$ converges to $w \in \mathbb{T}$, and $z \in \nabla$, where $\nabla \in \mathrm{STOLZ}_w$, then z fails to approach the boundary tangentially. In other words, the quantity[1]

$$d_w(z) = \frac{\text{distance}(z, \mathbb{T})}{\text{distance}(z, w)} = \frac{1 - |z|}{|z - w|} \tag{1}$$

stays bounded away from 0. Observe that $0 < d_w(z) \leqslant 1$ for each $z \in \mathbb{D}$ and $w \in \mathbb{T}$.

1.1.1 Qualitative Results for Holomorphic Functions

Definition 1.4 Let φ be a complex-valued function defined on $\mathbb{D}$, and let $w \in \mathbb{T}$. We say that *the angular limit of* φ *exists at* w, or that $\varphi_\flat(w)$ *exists*, if the limit value $\lim_{\nabla \ni z \to w} \varphi(z)$ exists for each $\nabla \in \mathrm{STOLZ}_w$. Since the various triangles in STOLZ_w are not all mutually disjoint, this limit value (when it exists) is independent of ∇ and is denoted by $\varphi_\flat(w)$.

[1] This is the *normalized distance to the boundary*, where the distance from z to w serves as normalizing term.

Theorem 1.5 (Fatou, 1906) *If* $F \in \mathscr{O}(\mathbb{D})$ *and* $\sup_{z\in\mathbb{D}} |F(z)| < \infty$ *then there is a subset* $\mathscr{E}_F$ *of* $\mathbb{T}$ *such that*

(i) *the set* $\mathscr{E}_F$ *has zero linear* (*Lebesgue*) *measure*;
(ii) *for each point* $w \in \mathbb{T} \setminus \mathscr{E}_F$, *the angular limit* $F_\flat(w)$ *exists.*

See [17]. The set of functions that satisfy the hypothesis of Theorem 1.5 is denoted by $H^\infty(\mathbb{D})$. Theorem 1.5 says that, if $F \in H^\infty(\mathbb{D})$, then, for a.e. boundary point $w \in \mathbb{T}$ (with respect to linear measure), $F(z)$ has a limit value, when $z \to w$ and z stays within any open triangle contained in $\mathbb{D}$, with vertex at w.

In 1923, F. Riesz obtained a result stronger than Theorem 1.5; see [54]. He proved that the same conclusion of Theorem 1.5 holds under weaker hypothesis, as follows.

Theorem 1.6 (F. Riesz, 1923) *Let* $0 < p < \infty$. *If* F *is a holomorphic function on* $\mathbb{D}$, *and*

$$\sup_{0<r<1} \int_0^{2\pi} \left|F\left(re^{i\theta}\right)\right|^p d\theta < \infty$$

then, for a.e. $w \in \mathbb{T}$ (*with respect to linear measure*), $F_\flat(w)$ *exists.*

The set of functions that satisfy the hypothesis of Theorem 1.6 is denoted by $H^p(\mathbb{D})$. The function spaces $H^p(\mathbb{D})$ are the *Hardy spaces* of holomorphic functions in $\mathbb{D}$.

The conclusion reached in Theorems 1.5 and 1.6 (on the existence a.e. of angular limits) is of a qualitative nature. Before stating a quantitative version of these basic result, we state another result, also of a qualitative nature, that may appear paradoxical. First, we need one more notion. Let $n > 1$ and $w \in \mathbb{T}$, and d_w be as in (1), and define

$$\Gamma_n(w) = \left\{ z \in \mathbb{D} : d_w(z) > \frac{1}{n} \right\}.$$

Observe that $\Gamma_n(w) \in \mathbb{D}_w$, and that $n \mapsto \Gamma_n(w)$ is a strictly increasing (set-valued) function.

The approach regions $\Gamma_n(w)$ and the Stolz approach regions at w are equivalent, from the point of view of limit values. Indeed, there is a tail of $\Gamma_n(w)$ that is contained in a Stolz approach region at w (the larger is n, the wider will be the angle of the Stolz approach), and any Stolz approach region at w has a tail that is contained in $\Gamma_n(w)$, for n large enough.

The precise value of the parameter n is not important, if we are interested in the behavior of f for a.e. w, rather than in the behavior at a single $w \in \mathbb{T}$. The following result may appear paradoxical but it clarifies this claim.

Theorem 1.7 *Let* $\varphi : \mathbb{D} \to \mathbb{C}$ *be any function. Let* $n_0 > 1$ *be fixed. Assume that, for a.e.* $w \in \mathbb{T}$, *the limit*

$$\lim_{\Gamma_{n_0}(w) \ni z \to w} \varphi(z)$$

exists. Then, for a.e. $w \in \mathbb{T}$, *the angular limit of* φ *at* w *exists.*

A proof (of a more general version) of Theorem 1.7 can be found in [11].

Theorem 1.7 is one of several examples of phenomena that hold *almost everywhere* but do *not* hold at a single fixed point.

Observe that the function φ in Theorem 1.7 is not assumed to be continuous or even measurable. Indeed, *a priori*, it is not necessary, in order for the conclusion of this theorem to make sense, to assume that φ is continuous or measurable.

Definition 1.8 A *regularity condition*, in the statement of a theorem, is a hypothesis that is not *a priori* needed in order for the conclusion of the theorem to make sense.

Osgood (1916) had the following to say about this issue; see [47].

> It is unsatisfactory, in stating an important theorem, not to know whether a given hypothesis is needed merely for convenience of proof, or whether the theorem would be false if it were omitted. The situation is still more annoying when it is conceivable that the theorem could be proven with about the same ease without the hypothesis, if one were only able to see more clearly.

Here are three examples that should convince the reader that it is not possible to predict *a priori* whether a regularity condition in a theorem can be safely omitted.

- A theorem of Osgood (1899) says that a bounded function of several complex variables, that is holomorphic in each variable separately, is holomorphic; see [46]. In this statement, the boundedness hypothesis is a regularity condition: It is not necessary to give a meaning to the conclusion of the theorem. In 1906, Hartogs proved that this regularity condition may be omitted [23]. In other words, the conclusion holds even if this hypothesis is omitted.
- In [7, Theorem 2 (INT IV.64)], the hypothesis that "the set A is countable" is a regularity condition.
- In Fatou's theorem, the hypothesis that the function is bounded is a regularity condition. If we omit this regularity hypothesis than the conclusion fails, in general. However, Theorem 1.12 (stated below), due to Plessner, shows that if we omit this hypothesis than a *weaker* conclusion follows. The curious fact is that the new theorem thus obtained actually *implies* Fatou's Theorem.

In Theorem 1.7, there are no hypothesis on the function f, but the width n is fixed and independent of w. What happens if we allow the width to depend on w?

Definition 1.9 We say that a function $F : \mathbb{D} \to \mathbb{C}$ *is non-tangentially bounded at a point* $w \in \mathbb{T}$ if there exists $n_0 > 1$ and there is a tail T of $\Gamma_{n_0}(w)$ on which $|F|$ is bounded, i.e., such that $\sup_{z\in T} |F(z)| < \infty$.

Theorem 1.10 (Privalov, 1923) *Let* $F \in \mathscr{O}(\mathbb{D})$. *Assume that for each* $w \in \mathbb{T}$, *the function* F *is non-tangentially bounded at* w. *Then, for a.e.* $w \in \mathbb{T}$, *the angular limit of* F *at* w *exists and is finite.*

Observe the lack of uniformity: The value of n_0 and the bound $\sup_{z\in T} |F(z)|$, in the definition of non-tangential boundedness, depend on w. Theorem 1.10 is due to

Table 1 Outline of the qualitative results of this section for $\mathscr{O}(\mathbb{D})$

P. Fatou (1906)	Boundedness
F. Riesz (1923)	Hardy-type growth condition
Privalov (1923)	Non-tangential boundedness
Plessner (1927)	No growth condition required

Privalov [51]. Actually, in this work he proved a more general, local version, known as the *local Fatou theorem*. His proof was based on the Riemann mapping theorem.

Theorem 1.11 (Privalov, 1923) *Let $F \in \mathscr{O}(\mathbb{D})$. Let $E \subset \mathbb{T}$. Assume that for each $w \in E$, the function F is non-tangentially bounded at w. Then, for a.e. $w \in E$, the angular limit of F at w exists and is finite.*

Privalov's result actually *implies* Fatou's theorem.

The following and last result of this section, due to Plessner [49], says that the a.e. angular boundary behavior of holomorphic functions on $\mathbb{D}$ is either "good" or "bad".

Theorem 1.12 (Plessner, 1927) *Let $F : \mathbb{D} \to \mathbb{C}$ be holomorphic.*[2] *Then there exists a set $\mathscr{E}_F \subset \mathbb{T}$ such that*

- *the linear measure of $\mathscr{E}_F$ is zero*;
- *for each $w \in \mathbb{T} \setminus \mathscr{E}_F$, either one of the following properties occurs*:

(Plessner points) *For each $\nabla \in \text{STOLZ}_w$, the set $\{F(z) : z \in \nabla\}$ is dense in the one-point compactification of $\mathbb{C}$.*

(Fatou points) *The angular limit of F exists and is finite.*

Observe that Plessner's result actually *implies* Fatou's theorem.

In this section we have presented four basic results, of a qualitative nature, on the boundary behavior of holomorphic functions on the unit disc (possibly subject to certain growth conditions of Hardy-type). Table 1 contains an outline of these results, with an indication of the boundedness conditions under which the result holds.

1.1.2 Quantitative Results for Holomorphic Functions

The following result, due to Hardy and Littlewood [22], is rather subtle, and gives a quantitative version of Fatou's theorem on the a.e. existence of angular limits.

[2] As a matter of fact, the conclusion also holds for functions that are merely meromorphic on $\mathbb{D}$.

Theorem 1.13 (Hardy–Littlewood, 1930) *Let* $0 < p < \infty$, *and fix* $n > 1$. *Then there exists a constant* $c = c(p, n)$, *with* $0 < c < \infty$, *such that, if* $F \in H^p(\mathbb{D})$, *then*

$$\int_0^{2\pi} \sup_{z \in \Gamma_n(e^{i\theta})} \left|F(z)\right|^p d\theta \leqslant c \int_0^{2\pi} \left|F_\flat\left(e^{i\theta}\right)\right|^p d\theta.$$

1.1.3 The Optimality Issue

J.E. Littlewood posed the following problem: Is it possible to obtain a result stronger than Theorem 1.5, by way of proving *a stronger conclusion* with the same hypothesis? More precisely, his question may be rephrased as follows.

Littlewood's Question Are there approach regions that are *larger* than the Stolz approach and that, moreover, also work for the a.e. convergence of functions in $H^p(\mathbb{D})$?

The catch here is: *larger in what sense?* A simple example of an approach region that is set-theoretically *larger* than the Stolz approach regions is given by a curve that is *tangential* to the boundary at the given point. Of course, we need to fix a curve for every single boundary point, in order for the problem to be meaningful.

Definition 1.14 A *family of approach regions in* $\mathbb{D}$ is a function

$$\gamma : \mathbb{T} \to \mathscr{P}(\mathbb{D}), \quad \text{such that} \quad \gamma(w) \in \mathbb{D}_w, \quad \text{for every } w \in \mathbb{T}.$$

Definition 1.15 Let γ be a family of approach regions in $\mathbb{D}$.

(c) We say that *each* $\gamma(w)$ *is a curve in* $\mathbb{D}$ *ending at* w if the following condition holds: For each $w \in \mathbb{T}$ there exists a continuous function $[0, 1) \to \mathbb{D}$, which, with slight abuse of language, we call γ_w, such that

$$\lim_{s \to 1} \gamma_w(s) = w, \qquad \gamma(w) = \left\{\gamma_w(s) : s \in [0, 1)\right\}.$$

(tg) We say that *the curves* $\gamma(w)$ *are tangential to the boundary* if[3]

- each $\gamma(w)$ is a curve in $\mathbb{D}$ ending at w;
- for every $w \in \mathbb{T}$, $\lim_{s \uparrow 1} d_w(\gamma_w(s)) = 0$.

(aecv) We say that *the functions in* $H^\infty(\mathbb{D})$ *converge a.e. along* γ *to their angular limits* if the following conditions holds: For every $f \in H^\infty(\mathbb{D})$, the set

[3]Recall that $d_w(z)$ is the normalized distance to the boundary. See (1).

$$C(f,\gamma) := \left\{ w \in \mathbb{T} : \text{ both } f_\flat(w) \text{ and } \lim_{\gamma(w)\ni z \to w} f(z) \text{ exist and are equal} \right\}$$

has full linear measure in $\mathbb{T}$.

The STRONG SHARPNESS STATEMENT is the following claim.

Strong Sharpness Statement There is no family of approach regions in $\mathbb{D}$ for which (c), (tg), and (aecv) for H^∞ hold.

This claim is consistent with a principle—implicit in Fatou (1906)—whose first precise rendition is due to Littlewood (1927). He proved the claim under an additional regularity condition.

Definition 1.16 A family γ of approach regions in $\mathbb{D}$ is *rotationally-invariant* if

$$\text{for each } w \in \mathbb{T}, \quad \gamma(w) = \left\{ wz : z \in \gamma(1) \right\}.$$

Here is the precise statement of Littlewood's result. See [37].

Theorem 1.17 (Littlewood, 1927) *Let γ be a family of approach regions in $\mathbb{D}$. Assume that*

- (c) *each $\gamma(w)$ is a curve in $\mathbb{D}$ ending at w*;
- (tg) *the curves $\gamma(w)$ are tangential to the boundary*;
- *γ is rotationally-invariant.*

Then (aecv) *for $H^\infty(\mathbb{D})$ does not hold for γ. Indeed, there exists $f \in H^\infty(\mathbb{D})$ such that, for a.e. $w \in \mathbb{T}$, the limit value*

$$\lim_{\gamma(w)\ni z \to w} f(z)$$

does not exist.

Littlewood's result is called a *negative* theorem, since it bars certain families of approach regions, *larger* than the Stolz approach regions, from begin conducive to a.e. convergence for functions in $H^\infty(\mathbb{D})$. The hypothesis of rotational-invariance in Theorem 1.17 is a regularity condition, in the sense of Definition 1.8. The shape, and, in particular, the order of tangency of $\gamma(w)$ to the boundary, is fixed and independent of w.

Question (Littlewood, 1927) May we allow the curves to *change their shape from point to point* in Theorem 1.17?

This question goes back to Littlewood (1927). It amounts to ask whether it is possible to omit the regularity condition of rotational-invariance from Theorem 1.17, and therefore it brings us back to the general question we posed at the beginning of Sect. 1.1.3 and to the issue of the truth-value of the STRONG SHARPNESS STATEMENT. Roughly speaking, we may imagine two radically different scenarios.

- If the hypothesis of rotational-invariance *may* be omitted from the hypothesis of Theorem 1.17 *without* invalidating its conclusion, then the STRONG SHARPNESS STATEMENT would be true, and this would mean that Fatou's theorem would be fairly sharp in its conclusion, from the viewpoint of the set-theoretical largeness of the approach regions.
- Assume, instead, that when you omit the hypothesis of rotational-invariance from Theorem 1.17, the conclusion may turn out to be false. Then the STRONG SHARPNESS STATEMENT would be false, and this would mean that the Stolz approach is not as optimal as it could possibly be.

The following result is stronger than Theorem 1.17, since it has a stronger conclusion. It was proved in 1957 by Lohwater and Piranian [38].

Theorem 1.18 (Lohwater and Piranian, 1957) *Assume that γ is a family of approach regions in $\mathbb{D}$. Assume that*

- (c) *each $\gamma(w)$ is a curve in $\mathbb{D}$ ending at w*;
- (tg) *the curves $\gamma(w)$ are tangential to the boundary*;
- *γ is rotationally-invariant.*

Then it is not true that (aecv) *holds for γ for $H^\infty(\mathbb{D})$: There exists $f \in H^\infty(\mathbb{D})$ such that, for every $w \in \mathbb{T}$, the limit value*

$$\lim_{\gamma(w)\ni z\to w} f(z)$$

does not exist.

We will now take a brief detour, where we look at the problem of proving a theorem of Littlewood type with *weaker* hypothesis. This detour is motivated by our desire to understand whether it is possible to prove a theorem of Littlewood type where the regularity condition of rotational-invariance is *omitted*.

Aikawa proved a theorem of Littlewood type where the condition that γ is rotationally-invariant is weakened to a hypothesis of a *quantitative* nature. He obtained the following result; see [1] and [2].

Theorem 1.19 (Aikawa, 1990–1991) *Assume that γ is a family of approach regions in $\mathbb{D}$. Assume that*

- (c) *each $\gamma(w)$ is a curve in $\mathbb{D}$ ending at w*;
- (tg) *the curves $\gamma(w)$ are tangential to the boundary*;
- *the curves $\gamma(w)$ are uniformly bi-Lipschitz equivalent.*

Then (aecv) *for* $H^{\infty}(\mathbb{D})$ *does not hold for* γ. *Indeed*, *there exists* $f \in H^{\infty}(\mathbb{D})$ *such that*, *for a.e.* $w \in \mathbb{T}$, *the limit value* $\lim_{\gamma(w)\ni z\to w} f(z)$ *does not exist*.

In Theorem 1.19 the rotational invariance in Theorem 1.17 is replaced by the condition that the curves in the family γ are uniformly bi-Lipschitz equivalent.

Di Biase et al. (2006) proved a theorem of Littlewood type where the condition that γ is rotationally-invariant is weakened to a hypothesis of a *qualitative* nature, and obtained the following result [14].

Theorem 1.20 (Di Biase, Stokolos, Svensson, and Weiss, 2006) *Assume that* γ *is a family of approach regions in* $\mathbb{D}$. *Assume that*

- (c) *each* $\gamma(w)$ *is a curve in* $\mathbb{D}$ *ending at* w;
- (tg) *the curves* $\gamma(w)$ *are tangential to the boundary*;
- *for each open* $U \subset \mathbb{D}$, *the set*

$$\{w \in \mathbb{T} : U \cap \gamma(w) \neq \emptyset\}$$

  is a measurable subset of $\mathbb{T}$.

Then (aecv) *for* $H^{\infty}(\mathbb{D})$ *does not hold for* γ. *Indeed*, *there exists* $f \in H^{\infty}(\mathbb{D})$ *such that*, *for a.e.* $w \in \mathbb{T}$, *the limit value* $\lim_{\gamma(w)\ni z\to w} f(z)$ *does not exist*.

Definition 1.21 Let γ be a family of approach regions in $\mathbb{D}$. We say that γ is *regular* if, for each open set $U \subset \mathbb{D}$, the set $\{w \in \mathbb{T} : U \cap \gamma(w) \neq \emptyset\}$ is measurable.

Definition 1.22 If $U \subset \mathbb{D}$, the set $\{w \in \mathbb{T} : U \cap \gamma(w) \neq \emptyset\}$ is called the *shadow that* U *projects on the boundary via* γ, or, briefly, the γ-shadow of U.

Hence in Theorem 1.20 the rotational-invariance (in Theorem 1.17) is replaced by the regularity condition that the γ-shadow of each open subset of the unit disc is measurable.

Observe that the regularity conditions appearing in Theorems 1.19 and 1.20 are independent of each other.

We now look at the issue of whether it is possible to weaken the other hypothesis that appear in Littlewood's theorem (Theorem 1.17). A fairly precise answer, obtained by Di Biase et al. [14] in 2006, is in Theorem 1.25.

Definition 1.23 A family γ of approach regions in $\mathbb{D}$ is $*$-*connected* if

$$\text{For each } w \in \mathbb{T} \text{ the set } \{w\} \cup \gamma(w) \text{ is connected.}$$

Observe that, if each $\gamma(w)$ is a curve in $\mathbb{D}$ ending at w, then γ is $*$-connected, but the converse statement does not hold.

An example of a family γ of approach regions in $\mathbb{D}$ that is *not* $*$-connected is given as follows: For each $w \in \mathbb{T}$, $\gamma(w)$ is a countable set whose closure contains w.

Definition 1.24 A family γ of approach regions in $\mathbb{D}$ *lies eventually outside of the angular approach* if the following condition holds:

$$\text{For all } w \in \mathbb{T},\ \text{ for all } \nabla \in \text{STOLZ}_w,\ \text{ the set } \gamma(w) \text{ has a tail that is disjoint from } \nabla.$$

Observe that if the curves $\gamma(w)$ are tangential to the boundary, then γ lies eventually outside of the angular approach, but the converse statement does not hold.

Theorem 1.25 (Di Biase, Stokolos, Svensson, and Weiss, 2006) *Assume that γ is a family of approach regions in $\mathbb{D}$. Assume that*

- *γ is $*$-connected, in the sense of Definition* 1.23;
- *γ lies eventually outside of the angular approach*;
- *γ is regular.*

Then (aecv) *for $H^\infty(\mathbb{D})$ does not hold for γ. Indeed, there exists $f \in H^\infty(\mathbb{D})$ such that, for a.e. $w \in \mathbb{T}$, the limit value* $\lim_{\gamma(w)\ni z\to w} f(z)$ *does not exist.*

Observe that the hypothesis that γ is $*$-connected is significantly weaker than the hypothesis that each $\gamma(w)$ is a curve in $\mathbb{D}$ ending at w: The former hypothesis captures the property of curves that is relevant to a theorem of Littlewood type. Moreover, the hypothesis that γ lies eventually outside of the angular approach does not depend on the preliminary assumption that each $\gamma(w)$ is a curve in $\mathbb{D}$ ending at w, and does not depend on any smoothness assumptions.

Is it possible to omit from Theorem 1.25 the hypothesis that γ is $*$-connected? In 1979, Rudin [56], the author constructed a highly oscillating inner function in $\mathbb{D}$ that led to a result that must have been considered surprising at that time. In order to state Rudin's result, we need the following notion.

Definition 1.26 Let γ be a family of approach regions in $\mathbb{D}$. We say that γ *lies frequently outside of the angular approach* if the following condition holds:

$$\text{For each } w \in \mathbb{T},\ \text{ for each } \nabla \in \text{STOLZ}_w,\ \text{ no tail of } \gamma(w) \text{ is entirely contained in } \nabla.$$

Here is Rudin's surprising result.

Theorem 1.27 (Rudin, 1979) *There exists a family γ of approach regions in $\mathbb{D}$ that has the following properties*:

- *each $\gamma(w)$ is a curve in $\mathbb{D}$ ending at w*;
- *γ lies frequently outside of the angular approach*;
- *γ is regular*;
- (aecv) for $H^\infty(\mathbb{D})$ *holds for γ.*

Observe that Rudin's (positive) result is compatible with the previous (negative) results, since the condition that γ lies frequently outside of the angular approach is strictly weaker than the condition that it lies eventually outside of the angular approach.

Rudin's theorem (Theorem 1.27) implies that in Theorem 1.25 one cannot weaken the condition the γ lies eventually outside the angular approach to the condition that γ lies frequently outside of the angular approach.

The family of approach regions γ constructed by Rudin in Theorem 1.27 is *not* rotationally-invariant. At the time, this point must have appeared the crucial one, at least to the author. Indeed, Rudin asked the following question.

Question (Rudin, 1979) Is there a family γ of approach regions in $\mathbb{D}$ with the following properties?

- for every $w \in \mathbb{T}$, the set $\gamma(w)$ is a sequence $\{z_w(n)\}_{n\in\mathbb{N}}$ of points ending at w;
- for every $w \in \mathbb{T}$, $\lim_{n\to\infty} d_w(z_w(n)) = 0$;
- γ is rotationally-invariant;
- if $f : \mathbb{T} \to \mathbb{R}$ is Lebesgue integrable, then

$$f(w) = \lim_{n\to\infty} \frac{1}{|\{e^{i\delta} \in \mathbb{T} : z_w(n) \in \Gamma_1(e^{i\delta})\}|} \int_{\{e^{i\delta}\in\mathbb{T}:z_w(n)\in\Gamma_1(e^{i\delta})\}} f\left(e^{i\theta}\right) d\theta$$

  for a.e. $w \in \mathbb{T}$ (where vertical bars denote Lebesgue linear measure).

Rudin conjectured that the answer would be in the *negative*. Observe that the set that appears in the integrals written above is precisely the Γ_1-shadow of $z_w(n)$. The issue raised by Rudin is therefore related to the possibility of having a Lebesgue differentiation theorem where the sets over which we evaluate the mean values are located in an eccentric manner with respect to the "center" w. There is an intimate tie that binds the Lebesgue Differentiation Theorem to the boundary behavior of holomorphic (and harmonic) functions (a hint of this tie is glimpsed in Theorem 1.38 below). As a consequence of this tie, Rudin's conjecture is equivalent to the following.

Claim *There is no family γ of approach regions in $\mathbb{D}$ that has the following properties*:

- γ *lies eventually outside of the angular approach*;
- γ *is rotationally-invariant*;
- (aecv) *for $H^\infty(\mathbb{D})$ holds for γ.*

Observe that if this conjecture were true, it would be a *negative* result. *Caveat:* In this statement it is *not* required that each $\gamma(w)$ be a curve in $\mathbb{D}$ ending at w, and indeed such a requirement would be *inconsistent* with the theorems of Littlewood type that we have seen so far. Here the set $\gamma(w)$ is assumed to be a *sequence* of points in $\mathbb{D}$, ending at w, and tangential to the boundary. In other words, since a rotationally-invariant family of approach regions is regular, Rudin's conjecture is related to the question of whether it is possible to omit condition *(c)* from Theorem 1.17. It is also related to the question of whether it is possible to omit the hypothesis that γ is $*$-connected from Theorem 1.25. Indeed, if a family γ of approach regions exists, with the properties listed above, then Theorem 1.25 implies that γ *cannot* be $*$-connected.

Rudin's conjecture was disproved by Nagel and Stein (1984), who proved the following surprising results; see [43].

Theorem 1.28 (Nagel and Stein, 1984) *There is a family γ of approach regions in $\mathbb{D}$ with the following properties*:

- *for every $w \in \mathbb{T}$, the set $\gamma(w)$ is a sequence $\{z_w(n)\}_{n\in\mathbb{N}}$ of points ending at w*;
- *for every $w \in \mathbb{T}$, $\lim_{n\to\infty} d_w(z_w(n)) = 0$*;
- *γ is rotationally-invariant*;
- *if $f : \mathbb{T} \to \mathbb{R}$ is Lebesgue integrable, then, for a.e. $w \in \mathbb{T}$,*

$$f(w) = \lim_{n\to\infty} \frac{1}{|\{e^{i\delta} \in \mathbb{T} : z_w(n) \in \Gamma_1(e^{i\delta})\}|} \int_{\{e^{i\delta}\in\mathbb{T}:z_w(n)\in\Gamma_1(e^{i\delta})\}} f\left(e^{i\theta}\right) d\theta.$$

Theorem 1.29 (Nagel and Stein, 1984) *There exists a family γ of approach regions in $\mathbb{D}$ such that*

- *each $\gamma(w)$ is a curve in $\mathbb{D}$ ending at w*;
- *γ lies frequently outside of the angular approach*;
- *γ is rotationally-invariant*;
- (aecv) for $H^\infty(\mathbb{D})$ *holds for γ.*

Theorem 1.29 is stronger than Rudin's Theorem, since it shows that γ may be chosen to be rotationally-invariant.

Theorem 1.30 (Nagel and Stein, 1984) *There exists a family γ of approach regions in $\mathbb{D}$ such that*

- *γ lies eventually outside of the angular approach*;
- *γ is rotationally-invariant*;
- (aecv) for $H^\infty(\mathbb{D})$ *holds for γ.*

These results disprove Rudin's conjecture. They also imply that the family γ of approach regions constructed in Theorem 1.30 cannot be $*$-connected. Indeed, for each $w \in \mathbb{T}$, $\gamma(w)$ consists of a sequence of points in $\mathbb{D}$ converging to w and tangential to the boundary of the unit disc. The results of Nagel and Stein imply that *not* every such sequence may be used for such a construction.

Theorems 1.29 and 1.30 have the following consequences.

- In Theorem 1.25, it is not possible to omit the condition that γ is *-connected.
- In Theorem 1.25, it is not possible to relax the condition that γ lies *eventually* outside of the angular approach to the condition that it lies *frequently* outside of the angular approach.
- *Within the collection of all families of regular, $*$-connected approach regions*, the angular approach is optimal (in the sense that "nothing larger will do"), provided we interpret "larger" in the set-theoretical sense of "lying eventually outside [of the angular approach]".
- As soon as we allow families of approach regions that consist of *sequences*, then "something larger will do". This is the so-called *Nagel & Stein phenomenon*.

We are now ready to go back to the issue of the role played by the regularity hypothesis in Theorem 1.25. In Rudin (1979), the author also asked the following question.

Question (Rudin, 1979) Is there a family of approach regions in $\mathbb{D}$ for which (c), (tg), and (aecv) for $H^{\infty}(\mathbb{D})$ hold?

This question leads us back again to the issue of the truth-value of the STRONG SHARPNESS STATEMENT. Rudin conjectured that the answer would be in the negative. The answer turned out to be outflanking [14].

Theorem 1.31 (Di Biase, Stokolos, Svensson, and Weiss, 2006) *Within the axiomatic framework that is currently taken as a foundation of mathematics*,[4] *it is neither possible to prove nor to disprove the* STRONG SHARPNESS STATEMENT. *In other words*, *the* STRONG SHARPNESS STATEMENT *is independent of ZFC*.

For the details, see [14]. The proof relies on a combination of techniques from mathematical analysis and modern logic, based upon an insight about the nature of the link that makes the combination possible. Littlewood could not have possibly stated this result, since it is based on ideas that appeared after 1927.

Gödel believed that if a meaningful statement can be neither proved nor disproved within a certain axiomatic framework, then one should look for other axioms that are better suited to capture the mathematical reality which is addressed by the statement.

> Only someone who [...] denies that the concepts and axioms of classical set theory have any meaning (or any well-defined meaning) could be satisfied with such a solution, not someone who believes them to describe some well-determined reality. For in this reality Cantor's conjecture must be either true or false, and its undecidability from the axioms as knows today can only mean that these axioms do not contain a complete description of this reality; and such a belief is by no means chimerical, since it is possible to points out ways in

[4] To wit: ZFC, i.e., the Zermelo–Fraenkel axioms together with the Axiom of Choice.

Table 2 A (poetic) analogy

Axiomatic framework	Representations of the axioms
Euclid's axioms	Models of Euclidean geometry
The group axioms	Groups
A Midsummer night's dream	A representation of this play
The axioms in ZFC	

> which a decision of the question, even if it is undecidable from the axioms in their present form, might nevertheless be obtained.[5]

See [20]. It seems to us that Gödel's remarks apply equally well to the Strong Sharpness Statement, if one shares the Platonist viewpoint of Gödel, i.e., if one believes that holomorphic functions and approach regions do exists somewhere in the realm of mathematical objects.

How can a theorem such as Theorem 1.31 be proved? Modern mathematical logic gives us the tools to show that certain statements can be neither proved nor disproved. The basic idea is familiar: If different models ("representations") of some axioms exhibit different properties, then those properties do not follow from the axioms. Suppose our statement is the *parallel postulate*. How do we prove that we cannot prove/disprove it?

DON'TS: Spend time trying to deduce this postulate from the other axioms of Euclidean geometry.
DOS: Find models of Euclidean geometry where this postulate holds/does not hold.

Similarly, the existence of a non-Abelian group shows that the statement of commutativity cannot be derived from the group axioms.

In order to carry out this scheme, one has to adopt an axiomatic framework. The currently adopted axioms for mathematics are due to Zermelo and Fraenkel, together with the axiom of choice. These axioms are known by the acronym of ZFC. To prove a theorem amounts to deduce the statement of the theorem from ZFC. We may understand these matters by using the (poetic) analogy in Table 2.

The empty space in Table 1 is filled up by the notion of *model of* ZFC. Roughly speaking, a model of ZFC stands to ZFC as, say, a group stands to the group axioms.

If ZFC is consistent, then it has several, *different* models. Gödel showed, in his *completeness theorem*, that a statement can be deduced from ZFC *if and only if* it holds in every model of ZFC. In other words, in order to show that a statement cannot be proved or disproved within ZFC, one has to find one model of ZFC where the statement does not hold, and another model where the statement holds.

[5]See [20]. Here Gödel is concerned with Cantor's conjecture (also known as the Continuum Hypothesis). The "solution" which he refers to was an hypothetical proof of his conjecture that (a generalized version of) Cantor's conjecture was actually independent of ZFC. In 1940, Gödel proved part of this conjecture. The conjecture was proved in full by Cohen in 1963.

The notion of *a model of* ZFC can be rather disturbing for working mathematicians, who are used to think (rightly so, in my opinion) that the objects they have been studying for several or many years do have some well-determined reality and do exist somewhere in the realm of mathematical objects. As Gödel observed in the passage quoted above, those who (perhaps unconsciously) nurture this belief should consider Theorem 1.31 as an indication of the fact that ZFC *does not* contain a complete description of this reality. One is thus led to hope that new axioms may do a better job.[6]

We should perhaps reassure the working mathematicians, since they usually meet the notion of "a model of ZFC" with the suspicion that these are tricky objects where weird things happen: In any model of ZFC all our familiar theorems do quietly hold, no more and no less than in Rudin's books. See [9, 16, 26] and [34].

Instead, the boundary behavior of functions in $H^\infty(\mathbb{D})$ is radically different in different models of ZFC. The following three results will substantiate this claim, and also prove Theorem 1.31.

Theorem 1.32 (Di Biase, Stokolos, Svensson, and Weiss, 2006) *There is a model of ZFC where there exists a family γ of approach regions such that*

- *each $\gamma(w)$ is a curve in $\mathbb{D}$ ending at w*;
- *the curves are tangential to the boundary*;
- (aecv) *for $H^\infty(\mathbb{D})$ holds for γ.*

Theorem 1.33 (Di Biase, Stokolos, Svensson, and Weiss, 2006) *There is a model of ZFC with the following property*: *If γ is any family of approach regions in $\mathbb{D}$ such that*

- *γ is $*$-connected*;
- *γ lies eventually outside of the angular approach*

then there is a function $f \in H^\infty(\mathbb{D})$ such that the set

$$\left\{ w \in \mathbb{T} : \text{ the limit } \lim_{\gamma(w) \ni z \to w} f(z) \text{ does not exist } \right\}$$

has full outer linear measure.

The following result can be proved in ZFC, hence it holds in every model of ZFC.

Theorem 1.34 (Di Biase, Stokolos, Svensson, and Weiss, 2006) *There exists a family γ of approach regions in $\mathbb{D}$ such that*

- *each $\gamma(w)$ is a curve in $\mathbb{D}$ ending at w*;
- *the curves $\gamma(w)$ are tangential to the boundary*;
- *for each $f \in H^\infty(\mathbb{D})$ the set $C(f, \gamma)$, introduced in Definition 1.15, has full outer linear measure in $\mathbb{T}$.*

[6] Gödel himself nurtured this belief.

We close this section with the following quotation.

> Mathematics as (the narrative of) an adventure; [...] a mathematical problem is like a lantern that can help us to shed light on our path. This viewpoint (more emotional than rational) leads us to believe that, when a problem turns out to be undecidable, the lantern burns brighter and enables us to look at the path (and upon our adventure) from above, and see the farthest regions of the landscape, which perhaps are the axioms (of) ZFC.[7]

We hope of course to get closer to the "farthest regions of the landscape" and see what lies beyond, in the new horizon that will then open its view to the working mathematician.

1.2 The Second Viewpoint

Consider the basic results of the previous section, on the boundary behavior of holomorphic functions (in the Hardy spaces $H^p(\mathbb{D})$), and observe that the holomorphic Hardy spaces H^p have a natural counterpart in the Hardy spaces h^p of harmonic functions.

Question h (vague version)
Is it possible to obtain corresponding results in the study of the boundary behavior of harmonic functions (in the Hardy spaces h^p)?

A precise definition of the Hardy spaces h^p of harmonic functions is needed.

Definition 1.35 (The Hardy spaces of harmonic functions in the unit disc)

- $h(\mathbb{D})$ is the set of real-valued harmonic functions on $\mathbb{D}$.
- $h^\infty(\mathbb{D})$ is the set $\{u \in h(\mathbb{D}) : \sup_{z\in\mathbb{D}} |u(z)| < \infty\}$.
- $h^p(\mathbb{D})$ is the set $\{u \in h(\mathbb{D}) : \sup_{0<r<1} \int_0^{2\pi} |u(re^{i\theta})|^p \, d\theta < \infty\}$.

Observe that

$$H^p(\mathbb{D}) = \mathscr{O} \cap h^p(\mathbb{D}).$$

If $D \subset \mathbb{R}^n$ is a bounded domain,[8] then the definition of the Hardy spaces $h^p(D)$ of harmonic functions in D may be given in terms of the Green function of D. Recall that a real-valued function f defined on D is called harmonic if

---

[7]"La matematica come racconto di un'avventura. [...] un problema matematico da risolvere è come una lanterna che si può usare per illuminare la strada. Allora, da questo punto di vista (e sto parlando di emozioni non di pensiero razionale) se il problema si rivela non decidibile, la lanterna finisce per brillare ancora di più, anzi permette di guardare la strada, o l'avventura, dall'alto, in modo che si possano contemplare i lontani confini del paesaggio, che forse sono gli assiomi di ZFC." [19].

[8]A *domain* in a topological space is an open and connected set.

$$\sum_{k=1}^{n} \frac{\partial^2 f}{\partial {x_k}^2} = 0,$$

where $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. Denote by $G(x, y)$ the Green function of D, and fix $x_0 \in D$. If $r > 0$ then the superlevel set $\{z \in D : G(x_0, z) > r\}$ is denoted by D_r. The boundary of D_r is the level set $bD_r = \{z \in D : G(x_0, z) = r\}$.

Definition 1.36 (The Hardy spaces of harmonic functions on bounded domains) Let $D \subset \mathbb{R}^n$ be a bounded domain.

- $h(D)$ is the set of real-valued harmonic functions on D.
- $h^\infty(D)$ is the set $\{u \in h(D) : \sup_{z \in D} |u(z)| < \infty\}$.
- $h^p(D)$ is the set of all functions $u \in h(D)$ such that

$$\sup_{0<r} \int_{bD_r} \left|u(w)\right|^p \omega_r(dw) < \infty$$

here $\omega_r(dw)$ denotes harmonic measure for D_r with pole at x_0 (see *infra*).

For example, if $D = B_n \subset \mathbb{R}^n$ is the ball $B_n = \{x \in \mathbb{R}^n : x_1^2 + x_2^2 + \cdots + x_n^2 < 1\}$, then the Hardy spaces $h^p(B_n)$ of harmonic functions on B_n are defined by the following growth condition:

$$\sup_{0<r<1} \int_{S_n} \left|u(rw)\right|^p \omega(dw) < \infty,$$

where S_n is the boundary of B_n and ω is "surface" measure on S_n, i.e., the normalized Hausdorff $(n-1)$-dimensional measure on S_n.

It is now possible to give a more precise version of Question h.

Question h in the unit disc Consider the boundary behavior of *harmonic* functions in the unit disc (belonging to the Hardy spaces $h^p(\mathbb{D})$): Are there results corresponding to those obtained for the holomorphic Hardy spaces $H^p(\mathbb{D})$?

This problem is natural enough and, indeed, it has been posed at an early stage in the development of the subject. For example, the work of Hardy and Littlewood (where the quantitative estimate in Theorem 1.13 is proved) is already set in the real-variable, potential-theoretic framework of harmonic functions.[9]

Question h inherently leads to a more general version of itself.

[9] The maximal operator, a fundamental real-variable construct, has been introduced and exploited precisely in this paper. See also [59].

Question h in higher dimensions Consider the boundary behavior of *harmonic* functions defined on *smoothly bounded* domains $D \subset \mathbb{R}^n$ (belonging to the Hardy spaces $h^p(D)$). Are there results corresponding to those obtained for the holomorphic Hardy spaces $H^p(\mathbb{D})$?

We are thus led to the following more general problem.

Question h in higher dimensions with rough boundary Consider the boundary behavior of *harmonic* functions defined on domains $D \subset \mathbb{R}^n$ (belonging to the Hardy spaces $h^p(D)$). Are there results corresponding to those obtained for the holomorphic Hardy spaces $H^p(\mathbb{D})$, even for domains with rough boundary? Relax the smoothness assumptions on the boundary of the domain D as much as possible.

The first lucid statement of this problem is due to E.M. Stein (1962).

> [. . .] it would be desirable to extend these results by considering non-tangential behavior for sets lying on more general hyper-surfaces. Presumably this could be done without too much difficulty if the bounding hyper-surface were smooth enough. It would be of definite interest, however, to allow the most general bounding hyper-surface for which non-tangential behavior is meaningful. Hence, extension of these results to the case when the bounding hyper-surface is of class C^1 would have genuine merit. Whether this can be done is an open problem.

It is important to point out that the domains considered in these studies are not necessarily endowed with a group structure (suitably acting on the domain itself), and have *rank one*. Here *rank one* means, roughly speaking, that essentially only one "radius" can be issued from a boundary point inside the domain.

1.2.1 Motivations

Before we enter a little deeper in these matters, and see the answer that have been given to these questions, let us look at some of the original motivations that have prompted mathematical research in this subject.

There are at least three closely allied motivations for Question h in higher dimensions.

Zygmund's Vision Firstly, we have to recall Zygmund's efforts to extend the central results of harmonic analysis to higher dimensions. Indeed, the extension to harmonic functions in several variables of the results obtained for $H^p(\mathbb{D})$ is due, to an essential extent, to the development of *Zygmund's vision*, by the Chicago school of analysis. The difficulty here is that some of the results on the boundary behavior of holomorphic functions (in the Hardy spaces) have been proved using complex analytic tools (such as the Riemann mapping theorem) that are not available in higher dimensions. Therefore, the first task has been that of recapturing these results without using complex analysis. For example, the result of Privalov, given in Theorem 1.11, has actually been formulated and proved not only for the unit disc, but also for planar domains with rectifiable boundary. However, the proof given by Privalov used the Riemann mapping theorem (of which there is no higher-dimensional version), and therefore could not be adapted to harmonic functions in higher-dimensional domains.

The Hilbert Transform and Fourier Series Classical one-dimensional Fourier analysis lives in the boundary of the unit disc and has been another motivation for adopting a real-variable viewpoint in phenomena of boundary behavior. It is useful to see how the need for a real-variable approach arose. Here a major player has been the Moscow school of function theory, led by Egorov and Luzin.[10] Recall that, if f is an integrable function on $\mathbb{T}$, then the *Hilbert transform of* f is the singular integral operator

$$H(f)\big(e^{ix}\big) = -\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f(e^{i(x+u)})}{\tan(\frac{u}{2})}\,du$$

where the integral (whenever it exists) is interpreted as a principal value.[11] Luzin (1913, 1915) proved that this limit exists almost everywhere, provided f is square-summable, and that it is equal (almost everywhere) to the boundary values of the harmonic function that is conjugate to Pf. Privalov (1918) employed Theorem 1.5 (the basic result on the boundary behavior of holomorphic functions) in order to extend Luzin's result to functions f that are merely integrable on $\mathbb{T}$. Now, the partial sums $S_n(f)$ of the Fourier series of f are linked to the Hilbert transform by the formula

$$S_n(f) = \sin_n H(f\cos_n) - \cos_n H(f\sin_n) + o(1),$$

where $\sin_n(x) = \sin(nx)$, $\cos_n(x) = \cos(nx)$, and $o(1) \to 0$ as $n \to \infty$ uniformly in $\mathbb{T}$ (see [39] and [40]). Luzin suspected that

> [the convergence of the singular integral defining the Hilbert transform in $\mathbb{T}$, due to] a certain interference of positive and negative values [...], should be considered as an actual cause for the a.e. convergence of Fourier–Lebesgue series [of L^2 functions].

Luzin believed that, in order to prove his conjecture, one had to understand the properties of the Hilbert transform *without entering inside the unit disc*, i.e. without using complex analysis. Here lie one of the earliest motivations for the search of the real-variable roots of central results in harmonic analysis.

The (Classical) Dirichlet Problem In the classical Dirichlet's problem one has to determine a function that is harmonic in a given domain, and has prescribed

---

[10] It is not a coincidence that Privalov also proved some of the first results on the boundary behavior of (sub)harmonic functions in domains $D \subset \mathbb{R}^3$. See for example [31].

[11] This means that it is interpreted as the limit value $\lim_{\varepsilon\downarrow 0}(\int_{-\pi}^{-\varepsilon} + \int_{\varepsilon}^{\pi})$.

values at the boundary. This boundary value problem is inherently apt to be posed for higher-dimensional domains.

1.2.2 Two Further Questions

Two further questions contributed to shape the subject.

Question h and probability theory Is it possible to understand (at least in part) the boundary behavior of harmonic functions using probabilistic notions?

The answer turned out to be in the positive. The tool here is an appropriate diffusion process (Brownian motion), although one may also adopt, as a discrete counterpart, an appropriate Markov process (see *infra*).

Question h with intrinsic methods Study the boundary behavior of harmonic functions defined on domains D that are defined by intrinsic properties, rather than by an embedding in Euclidean spaces.

A suggestive answer to this problem has been given by Ancona in 1987. See [3].

1.2.3 The Dirichlet Problem for the Unit Disc

In order to see the answers that have been given to these questions, it is useful to look in some detail at the *classical Dirichlet's problem*.

The *classical Dirichlet's problem for* $\mathbb{D}$ is the problem of determining a function that is harmonic in the unit disc, and has prescribed values at the boundary. More precisely, if f is the function, defined on $\mathbb{T}$, that encodes the prescribed boundary values, then we seek an *unrestricted harmonic extension of* f *to* $\mathbb{D}$, i.e., a function Pf, defined on $\mathbb{D}$ and *harmonic* therein, such that

$$\lim_{\mathbb{D}\ni z\to w} Pf(z) = f(w), \quad \forall w \in \mathbb{T}. \tag{2}$$

The maximum principle implies that the classical Dirichlet problem has at most one solution. If the classical Dirichlet problem for $\mathbb{D}$ has a solution, then the boundary datum f has to be a continuous function on the boundary (because here we have unrestricted convergence). The following result shows that, *in the unit disc*, the continuity of the boundary function is also sufficient.

Theorem 1.37 (Prym 1871, Schwarz 1872) *The classical Dirichlet problem for the unit disc has a solution for each continuous boundary function. The solution is given by the* Poisson integral of f

$$Pf(z) = \frac{1}{2\pi}\int_0^{2\pi} \frac{1-|z|^2}{|e^{i\theta}-z|^2} f\left(e^{i\theta}\right) d\theta \quad (z \in \mathbb{D}). \tag{3}$$

Hence the boundary limit (2) recaptures a continuous boundary function from its Poisson integral. The boundary limit in (2) is called *unrestricted* because no restriction on z is required, apart from the necessary requirement that it belongs to $\mathbb{D}$.

Observe that the Poisson integral Pf in (3) makes also sense when the boundary function f is assumed to be merely Lebesgue integrable, rather than continuous. In 1906, just a few years after the introduction of Lebesgue measure, Fatou published a seminal memoir [17], where he treated the following natural question.

An Inversion Problem for $f \mapsto Pf$ Is it possible to recapture f from Pf when f is assumed to be merely Lebesgue integrable, rather than continuous?

If the answer were positive, then the Poisson integral Pf could be seen as a solution, *in a wide sense*, of the Dirichlet problem. Significantly, Fatou was also interested in the problem of recapturing a (Lebesgue integrable) function from the sequence of its Fourier coefficients. The link between these two inversion problems is not merely formal. Indeed, several results on Fourier series were proved for the first time by "entering inside the unit disc" and using results on the boundary behavior of harmonic, or holomorphic, functions.[12]

Since a Lebesgue integrable function is only defined almost everywhere, we can only hope to recapture f from Pf at almost every point. According to a result of Lebesgue, for almost every point $w \in \mathbb{T}$ the value $f(w)$ is equal to the limit value of the average of f on small boundary intervals containing w (as the intervals "shrink" to the point). These points are called the *Lebesgue points* of f. Fatou showed that the angular limit of Pf exists, and equals f, at every Lebesgue point of f.

Theorem 1.38 (Fatou 1906) *If f is a Lebesgue integrable on $\mathbb{T}$, and $w \in \mathbb{T}$, then*

$$\text{if } w \text{ is a Lebesgue point of } f \text{ then } (Pf)_\flat(w) = f(w).$$

In particular, $(Pf)_\flat(w) = f(w)$ almost everywhere.

A Lebesgue integrable function on $\mathbb{T}$ may thus be recaptured from its Poisson integral, which may be seen as a solution, in a wide sense, of the Dirichlet problem. Theorem 1.38 hints at the following general principle. See [59].

[12] The study of the boundary behavior of holomorphic and harmonic functions has remained dear to the heart of many harmonic analysts, even in cases where the boundary exhibits no group structure. Indeed, the study of these cases have led to an understanding of boundary behavior from the viewpoint of real-analysis. It has also led to the construction of (real-variable) tools that have proved to be most efficient in these studies. See [18, 59].

Fatou's Principle The "differentiability" properties of a boundary function f at a point $w \in \mathbb{T}$ positively affect the boundary behavior of Pf at w: higher regularity of f implies better boundary behavior of Pf.

The following result follows from Theorem 1.38 and is virtually due to Fatou (1906). It shows that a Fatou-type theorem for $h^\infty(\mathbb{D})$ does indeed hold.

Theorem 1.39 *If $u \in h^\infty(\mathbb{D})$ then the angular limit $u_\flat(w)$ exists for a.e. $w \in \mathbb{T}$.*

An answer to the other questions has been found by delving deeper into the Dirichlet problem and by looking to higher dimensions.

1.2.4 The Dirichlet Problem for Bounded Domains in $\mathbb{R}^n$

The classical Dirichlet problem for the unit disc is a special case of a more general problem. If $D \subset \mathbb{R}^2$ is a bounded domain, the *classical Dirichlet problem* for D is defined just as in the case of the unit disc, *mutatis mutandis*. Indeed, it is enough to replace $\mathbb{D}$ with D, and $\mathbb{T}$ with bD, the geometric boundary of D.

The maximum principle for harmonic functions implies that the classical Dirichlet problem for a bounded domain D has at most one solution. Moreover, it implies that, if the classical Dirichlet problem has a solution, then the boundary datum f has to be a continuous function on the boundary. Is continuity a sufficient condition for a solution to exist, in this generality?

Motivations There are at least three reasons that explain why the Dirichlet problem has kept mathematicians busy for several decades.

- Firstly, it all started with Riemann. The existence of a solution to the Dirichlet problem for a planar domain D was invoked by Riemann (1851) in order to prove that if $D \subset \mathbb{R}^2$ is bounded simply connected domain, then it is conformally equivalent to the unit disc. However, in 1870 Weierstrass pointed out that the variational argument, used by Riemann to establish the existence of a solution, appeared to be incomplete. As we have seen, the first proofs that the classical Dirichlet problem *for the unit disc* can be solved were given by Prym (1871) and Schwarz (1872), after mistaken attempts made by Neumann (1865) and Hankel (1870). See [52, 57, 58].
- There is no higher-dimensional version of the Riemann mapping theorem, but there is, as we have seen, a natural higher-dimensional version of the Dirichlet problem (set in bounded domains in $\mathbb{R}^n$). This more general problem has exerted a fascinating influence and has pushed mathematical research for several decades, also because of its connections with various fields.
- The Dirichlet problem (for planar domains as well as for domains in $\mathbb{R}^3$ and $\mathbb{R}^n$) plays an important role in many problems of mathematical physics. In 1890, Poincaré published an influential work on these aspects [50].

The Classical Solution Operator Let $D \subset \mathbb{R}^n$ be a bounded domain. The following notation will be used.

$C(bD)$ is the Banach space of real-valued continuous functions defined on $\mathbb{T}$.
$C_h(bD)$ is the set of $f \in C(bD)$ for which the classical Dirichlet problem has a solution.

Definition 1.40 If $D \subset \mathbb{R}^n$ is a bounded domain, then (the boundary of) D is called *regular* (for the Dirichlet problem) if the classical Dirichlet problem for D has a solution for each continuous real-valued boundary function. The *classical solution operator*

$$P : C_h(bD) \to h(D)$$

maps $f \in C_h(bD)$ to the solution Pf of the classical Dirichlet problem with boundary values f.

The set $C_h(bD)$ is nonempty, since it contains constant functions and the restrictions of harmonic polynomials. The maximum principle for harmonic functions implies that

- the set $C_h(bD)$ is a closed subspace of $C(bD)$;
- the operator $P : C_h(bD) \to h(D)$ is positive and linear.

We have already observed that if the classical Dirichlet problem has a solution, then the boundary datum f has to be a continuous function on the boundary. However, the *continuity of f may not be enough to ensure the existence of a solution in the classical sense*. This fact was shown by examples given in 1911 by Zaremba [66] and in 1912 by Lebesgue [36]. The following regularity criterion is useful.

Theorem 1.41 *If $D \subset \mathbb{R}^n$ is a bounded domain, then the following conditions are equivalent*:

- *D is regular for the Dirichlet problem, i.e. $C_h(bD) = C(bD)$*;
- *$C_h(bD)$ is dense in $C(bD)$ with respect to the uniform norm*;
- *if f and g belong to $C_h(bD)$ then fg belongs to $C_h(bD)$.*

A version of Theorem 1.41 is contained in Riemann's *Habilitationsschrift*, written in 1854, devoted to the problem of recapturing a periodic function from the sequence of its Fourier coefficients. Riemann had understood that the following two statements are *equiveridical*:

(i) the unit disc is regular for the Dirichlet problem;
(ii) trigonometric polynomials[13] are dense in $C(bD)$.

[13] Indeed, a trigonometric polynomial $f(\theta) = \frac{1}{2}a_0 + \sum_{k=1}^{N}(a_k \cos k\theta + b_k \sin k\theta)$ belongs to $C_h(\mathbb{T})$, since the function, whose value at $z = re^{i\theta} \in \mathbb{D}$ equals to $\frac{1}{2}a_0 + \sum_{k=1}^{N}(a_k \cos k\theta + b_k \sin k\theta)r^k$, is the unrestricted harmonic extension of f.

In 1897, Volterra observed that the second statement was virtually within the reach of the knowledge available in 1854.[14]

The Generalized Solution Operator The inclusion $C_h(bD) \subset C(bD)$ is proper precisely when D is *not* regular. In these cases, if we try and extend P to an operator $\overline{P}$ defined on all of $C(bD)$, as in the following diagram,

$$\begin{array}{ccc} C_h(bD) & \xrightarrow{\text{inclusion}} & C(bD) \\ & \searrow^{P} & \vdots\ \overline{P}=? \\ & & h(D) \end{array} \tag{4}$$

then it is *not* possible to obtain such an extension that would *also* preserve (2). In other words, if D is not regular (for the Dirichlet problem), then it is *not* possible to define $\overline{P}$ in such a way that the diagram in (4) commutes, and, moreover, ensure that

$$\lim_{\mathbb{D}\ni z\to w} \overline{P}f(z) = f(w) \tag{5}$$

for each $f \in C(bD)$ and each $w \in bD$. The following theorem is the result of various decades of efforts. It is a remarkable result.[15] Indeed, it is plausible that an extension of P to $C(bD)$ should exist, but no special meaning could *a priori* be attached to it, if it were not unique.[16]

Theorem 1.42 (Wiener 1924, Perron 1923, Remak 1924, Keldych 1945) *Let $D \subset \mathbb{R}^n$ be a bounded domain. There exists one and only one operator*

$$\overline{P} : C(bD) \to h(D)$$

with the following two properties:

(i) $\overline{P}f = Pf$ *for each* $f \in C_h(bD)$;
(ii) $\overline{P}$ *is linear and positive.*

---

[14]Indeed, since continuous functions on $\mathbb{T}$ are uniformly continuous, they may be approximated in the uniform norm by piecewise linear functions, which are piecewise monotonic and to which, therefore, a theorem of Dirichlet (1829) on Fourier series applies. See [64]. Riemann's insight, as well as many others, remained hidden for some time in his work. Weierstrass gave the first published proof of the second statement, but his proof was not the one indicated by Volterra.

[15]The uniqueness part of this theorem means that there is indeed an operator $\overline{P}$ that extends P, and which is not merely the result of some (arbitrary) construction, since it has an intrinsic meaning. The first constructions of $\overline{P}$ appear in [48, 53, 65]. The uniqueness is due to Keldych, a work that is curiously ignored by most treatises on potential theory and harmonic measure. See [28, 29].

[16]Some authors apply the Hahn-Banach theorem, but this method does not yield uniqueness.

Probabilistic Counterpart The following proof of the existence of the operator $\overline{P}$ is due to Kakutani, and has the virtue that it brings to light in a natural way the probabilistic counterpart of $\overline{P}$. Suppose, for notational simplicity, that $D \subset \mathbb{R}^2$, denote by $\overline{D}$ the closure of D, and define the operator

$$K : C(\overline{D}) \to C(\overline{D}) \tag{6}$$

as follows:

$$K\varphi(z) = \begin{cases} \frac{1}{2\pi}\int_0^{2\pi} \varphi(z + \frac{\text{dist}(z)}{2} e^{i\theta})\, d\theta & \text{if } z \in D, \\ \varphi(z) & \text{if } z \in bD \end{cases}$$

where $\varphi \in C(\overline{D})$ and $\text{dist}(z) = \text{distance}(z, bD) = \min\{|z - w| : w \in bD\}$. Denote by K^n the composition of K with itself n times. If $f \in C(bD)$, let φ be *any* continuous extension of f to $\overline{D}$. Define $\overline{P} f$ as $\lim_{n\to\infty} K^n \varphi$. Indeed, one can prove that this limit exists in the topology of uniform convergence in $C(\overline{D})$, is harmonic on D, and only depends on f.

The operator K defined in (6) is linked to a Markov process that is a discrete version of Brownian motion. Indeed, $f \in C(bD)$ can be recaptured from $\overline{P} f$, even if D is not regular, by way of Brownian motion, a diffusion process strongly linked to harmonic functions, as the following result shows.

Theorem 1.43 *Let $D \subset \mathbb{R}^n$ be a bounded domain and let $f \in C(bD)$. Then almost every Brownian path from a point of D tends to a point $w \in D$ depending on the path, and $\overline{P} f$ evaluated along the path has limit $f(w)$.*

The domain in Theorem 1.43 is *not* assumed to be regular for the Dirichlet problem. Theorem 1.43 shows that we may recapture a boundary function f by way of evaluating $\overline{P} f$ along *Brownian paths* in D.

We now go back to the unrestricted approach to the boundary.

Regular Points for the Dirichlet Problem Let $D \subset \mathbb{R}^n$ be a bounded domain. A point $w \in bD$ is called *regular (for the Dirichlet problem)* if (5) holds for each $f \in C(bD)$.

Observe that D is regular if and only if each point of bD is regular.

Theorem 1.44 *Let $D \subset \mathbb{R}^n$ be a bounded domain, and let $w \in bD$. The following conditions are equivalent:*

- *w is regular;*
- *there is a strictly positive superharmonic function u on D such that*

$$\lim_{z\to w} u(z) = 0;$$

- $\lim_{z\to w} G_D(x, z) = 0$, *for some (equivalently every) point $x \in D$, where G_D is the Green function of D with pole x.*

See [15, Chapter 1.VIII]. The following geometrical conditions ensure regularity.

Theorem 1.45 *Let $D \subset \mathbb{R}^n$ be a bounded domain. Let $w \in bD$, and assume that either one of the following conditions holds*:

- *there is a closed ball which meets the closure of D at w but at no other point*;
- *$n = 2$ and w is the endpoint of a simple continuous arc which is contained in $\mathbb{R}^2 \setminus D$ except for w.*

Then w is regular.

The following result shows that the *irregular* points are *not too many*.

Theorem 1.46 (Kellogg 1928, Evans 1933) *If $D \subset \mathbb{R}^n$ is a bounded domain, then the set of irregular points has capacity zero.*

A bridge between the geometry of the domain and the operator $\overline{P} : C(bD) \to h(D)$ is given by the notion of *harmonic measure*.

Harmonic Measure The following notation will be used, where $D \subset \mathbb{R}^n$ is a bounded domain.

$\mathscr{B}(bD)$ The σ-algebra of Borel subsets of bD is denoted by $\mathscr{B}(bD)$.

$C^*(bD)$ Let $C^*(bD)$ denote the dual of $C(bD)$.

$C^*_+(bD)$ We denote by $C^*_+(bD) \subset C^*(bD)$ the cone in $C^*(bD)$ consisting of the *positive* linear functionals on bD. An element of $C^*_+(bD)$ may and will be identified to a positive measure defined on $\mathscr{B}(bD)$. The value of $\mu \in C^*_+(bD)$ at $f \in C^*(bD)$ is denoted by $\langle \mu, f \rangle$ if μ is seen as a linear functional, and by $\int_{bD} f(w)\mu(dw)$ if μ is seen as a positive Borel measure.

$C^*_1(bD)$ We denote by $C^*_1(bD)$ the convex subset of $C^*_+(bD)$ consisting of *probability* measures on bD.

Finally, the *Dirac embedding*

$$\varepsilon : bD \to C^*_1(bD) \tag{7}$$

is the function that maps $w \in bD$ to the Dirac measure ε_w concentrated at w. The set $\{\varepsilon_w : w \in bD\}$ is the set of the extreme points of $C^*_1(bD)$.

Definition 1.47 Let $D \subset \mathbb{R}^n$ be a bounded domain. The *harmonic measure in D* is the harmonic function[17]

$$\omega : D \to C^*_1(bD) \tag{8}$$

[17] Observe that the function $\omega : D \to C^*(bD)$ is *harmonic*, in the sense that it is continuous and it has the *mean-value property*, i.e., $\omega(z) = \frac{1}{2\pi} \int_0^{2\pi} \omega(z + re^{i\theta})\, d\theta$ for all $z \in D$, with $r > 0$ small enough, where the integral is interpreted in the natural way.

such that

$$\overline{P}f(z) = \langle \omega_z, f \rangle = \int_{bD} f(w)\omega_z(dw)$$

for each $f \in C(bD)$ and each $z \in D$ (where we write ω_z for $\omega(z)$).

The Borel probability measure ω_z on bD is called the *harmonic measure for D with pole z*. If we need to emphasize the dependence of ω_z on the domain D, we write

$$\omega_z(A|D)$$

instead of $\omega_z(A)$, where $A \in \mathscr{B}(bD)$.

The harmonic function $\omega : D \to C_1^*(bD)$ relates the operator $\overline{P} : C(bD) \to h(D)$ to the geometry of the domain. For example, Theorem 1.46 implies that[18]

$$\lim_{D \ni z \to w} \omega_z = \varepsilon_w \tag{9}$$

for quasi every points $w \in bD$. Indeed, $w \in bD$ is regular if and only if (9) holds.

Harmonic Null Boundary Sets The maximum principle for harmonic functions implies that if $A \in \mathscr{B}(bD)$, and $\omega_{z_0}(A) = 0$ for some $z_0 \in D$, then $\omega_z(A) = 0$ for each $z \in D$. See [15, Chapter 1.VIII]. It follows that the notion of "boundary Borel subset of harmonic measure zero" is well-defined, in the sense that it does not depend on the choice of the pole.

ω-Measurable Boundary Subsets The preceding remark implies that the measure-theoretic completion of the σ-algebra $\mathscr{B}(bD)$, with respect to harmonic measure, does not depend on the choice of the pole: This σ-algebra will be denoted by $\mathscr{R}(bD)$. Its elements are the *resolutive* boundary subsets. Hence a boundary subset A is resolutive if and only if there exist boundary Borel subsets B_1 and B_2 such that $B_1 \subset A \subset B_2$ and $B_2 \setminus B_1$ has harmonic measure zero.

The σ-algebra $\mathscr{B}(bD)$ is strictly contained in $\mathscr{R}(bD)$, and the measure ω_z extends in a unique way to $\mathscr{R}(bD)$. The sets in $\mathscr{R}(bD)$ are also called *ω-measurable*, or *measurable with respect to harmonic measure*.

The following result helps us cope with measurability issues of boundary subsets that are not Borel.

Analytic Boundary Subsets are ω-measurable Recall that a subset of bD is called *analytic* if it is the continuous image of the Baire space.[19] Every Borel set

[18]Loosely speaking, ω is a *harmonic* extension to D of the Dirac embedding given in (7), and ω embeds D in the convex (and weak* compact) set $C_1^*(bD)$.

[19]The Baire space is the set of all (infinite) sequences of natural numbers, endowed with the product topology, where $\mathbb{N}$ is taken discrete.

is analytic, but there are analytic sets that are not Borel.[20] A theorem of Luzin and Sierpinski implies the following useful result.

Theorem 1.48 *If $D \subset \mathbb{R}^n$ is a bounded domain, then every analytic set in bD is measurable with respect to harmonic measure.*

Theorem 1.48 is useful to deal with boundary subsets that are analytic but not Borel.

The following result (Theorem 1.49, due to Beurling, and curiously omitted from major treatises on harmonic measure) illustrates the fact that harmonic measure relates the operator $\overline{P}$ with the geometry of the domain. See [4].

Beurling's Estimate The quantity

$$d_A(z) = \frac{\text{distance}(z, bD)}{\text{distance}(z, A)} = \frac{\min\{|z-w| : w \in bD\}}{\inf\{|z-a| : a \in A\}}$$

(where $z \in D$ and $A \subset bD$), is a normalized distance to the boundary, with the distance from z to A serving as normalizing term. Observe that $0 < d_A(z) \leqslant 1$.

Theorem 1.49 *Let $D \subset \mathbb{R}^2$ be a bounded domain in the plane, and let $A \subset bD$ be measurable for harmonic measure. Then*

$$\omega_z(A) \leqslant \frac{4}{\pi} \arctan \sqrt{d_A(z)}.$$

Beurling proved Theorem 1.49 by using deep tools of complex-analysis.[21]

1.2.5 Stolz Approach Regions for Harmonic Functions of Several Variables

We have seen that, in its original formulation, the classical Dirichlet problem *cannot* be solved for every bounded domain $D \subset \mathbb{R}^n$ and every continuous boundary function. However, it is useful to look at a weaker form of the Dirichlet problem.

A Weak Form of the Dirichlet Problem Determine a function $\overline{P}f$ that is harmonic in a given bounded domain $D \subset \mathbb{R}^n$, and has prescribed continuous boundary values $f \in C(bD)$, in the sense that *f can be recaptured from $\overline{P}f$ by taking boundary values by some appropriate method*.

This weaker problem can be solved for every bounded domain $D \subset \mathbb{R}^n$ and every $f \in C(bD)$. The solution to this weaker problem is the map $\overline{P} : C(bD) \to h(D)$

[20] Lebesgue [35, pp.191–192] claimed, by a simple, short, but false argument, that the projection on the x-axis of a Borel subset of $\mathbb{R}^2$ is *obviously* Borel. Suslin showed that this claim is false [61, Theorem V].

[21] Is it possible to prove Theorem 1.49 (or a weaker form) by only using potential-theoretic tools?

described in Theorem 1.42. So far we have seen three "appropriate methods" that can be used to recover f from $\overline{P}f$:

- Brownian motion (or its discrete version, given by Kakutani's Markov chain), where one evaluates $\overline{P}f$ along Brownian paths (ending at boundary points);
- unrestricted boundary values (to be used at regular points);
- angular limits, to be used at Lebesgue points in the unit disc.

We will now dwell on the third method listed above. So far we have seen in action the third method listed above in the unit disc, where it is strongly linked to Fatou's Principle, enunciated in Sect. 1.2.3. It is possible to apply it for more general bounded domains in $\mathbb{R}^n$? In order to do so, we need a notion of "non-tangential approach region" for a bounded domain D in $\mathbb{R}^n$. The following natural questions are also strongly linked to this one.

For which bounded domains $D \subset \mathbb{R}^n$ it is possible to define the notion of "non-tangential" approach region that is relevant in the boundary behavior of harmonic functions (in the Hardy spaces)?

As in Definition 1.1, we denote by $\mathscr{P}(D)$ the collection of all subsets of D. An *approach region in D at $w \in bD$* is an element of $\mathscr{P}(D)$ whose closure (in the ambient space $\mathbb{R}^n$) contains w. Let D_w be the set of all approach regions in D at w. Let us first look at the ball $B_n = \{x \in \mathbb{R}^n : x_1^2 + x_2^2 + \cdots + x_n^2 < 1\}$. In this case, the notion of "Stolz approach region" may be formulated as follows.

Definition 1.50 If D is the unit ball in $\mathbb{R}^n$ then a *Stolz approach region at $w \in bD$ in D* is the intersection of the following three sets:

- the domain D itself;
- an open cone of revolution with vertex w, axis of rotation the inner normal to the boundary at w and half-angle less than $\pi/2$;
- a ball in the ambient space $\mathbb{R}^n$ centered at w and sufficiently small radius.

We denote by $\textsc{Stolz}_w$ the collection of all the Stolz approach regions at w in D. Hence $\textsc{Stolz}_w \subset D_w$. If $z \in D$ converges to w through a Stolz approach region at w, then z fails to be tangential to the boundary, in the sense that $d_w(z)$ stays bounded away from zero.[22] If φ be a complex-valued function defined on D, and $w \in bD$, then we say, as in Definition 1.4, that *the angular limit of φ exists at w*, or that $\varphi_\flat(w)$ *exists*, if the limit value $\lim_{\nabla \ni z \to w} \varphi(z)$ exists for each $\nabla \in \textsc{Stolz}_w$. This limit value (when it exists) is independent of ∇ and is denoted by $\varphi_\flat(w)$.

Similar definition can be given when the boundary of D is smooth, or even merely Lipschitz.

[22]Here $d_w(z) = \frac{\text{distance}(z,bD)}{\text{distance}(z,w)}$ is the normalized distance to the boundary, analogous to (1).

Let f be a real-valued function defined on the boundary of the unit ball in $\mathbb{R}^n$, and assume that f is integrable with respect to "surface" measure ω.[23] The notion of "Lebesgue point" of f is defined with respect to the "balls" in bD of the natural isotropic metric induced on bD by restriction of the Euclidean metric of the ambient space. The following result was obtained at an early stage. See [62].

Theorem 1.51 (Tsuji) *Let f be a real-valued function defined on the boundary of the unit ball in $\mathbb{R}^n$. If f is integrable with respect to surface measure and $w \in bD$ is a Lebesgue point of f, then*

$$(\overline{P} f)_\flat(w) = f(w),$$

i.e., the angular limit of $\overline{P} f$ exists and equals $f(w)$ through any Stolz approach region in D at w.

Theorem 1.51 extends and includes Theorem 1.38. The following result follows from Theorem 1.51 just as Theorem 1.39 follows from Theorem 1.38.

Theorem 1.52 *Let $u \in h^\infty(B_n)$, where B_n is the unit ball in $\mathbb{R}^n$. Then $u_\flat(w)$ exists for almost every point in the boundary of the unit ball* (*with respect to surface measure*).

Tsuji also obtained the quantitative estimate of Hardy–Littlewood in the unit ball. See [63].

In 1982, Jerison and Kenig introduced the class of NTA domains in $\mathbb{R}^n$. See [27]. We will not give the precise definition here. The following facts explain the importance of the class of NTA domains.

- Every Lipschitz domain is an NTA domain.
- There are NTA domains that are not Lipschitz. For example, the von Koch snowflake domain is an NTA domain, but it is not Lipschitz.
- The boundary of an NTA domain need not admit a tangent hyperplane. For example, the von Koch snowflake is an NTA domain but it is not rectifiable. In particular, any attempt to define the "non-tangential" approach regions using open cones of revolution (as in the case of the unit ball) will fail, in this generality. For example, almost every point in the boundary of the von Koch snowflake (with respect to harmonic measure) is a twist point, and therefore it is not sectorially accessible.[24]
- If $D \subset \mathbb{R}^n$ is an NTA domain and $w \in bD$, then the approach regions that are relevant in the study of the boundary behavior of harmonic functions in D are defined as follows, for $n > 1$:

$$\Gamma_n(w) = \left\{ z \in D : d_w(z) > \frac{1}{n} \right\}.$$

[23] The "surface measure" ω is the $(n-1)$-dimensional Hausdorff measure restricted to bD.

[24] This means that it is not the vertex of any open triangle contained in the domain.

- Jerison and Kenig proved a Fatou-type theorem for functions in $h^p(D)$, with $1 \leqslant p \leqslant \infty$, where $D \subset \mathbb{R}^n$ is an NTA domain. Here the almost everywhere result is expressed in terms of harmonic measure.
- Jerison and Kenig also proved a result of Privalov-type for NTA domains (in the spirit of Theorem 1.11). A result of Plessner-type for NTA domains (in the spirit of Theorem 1.12) follows from the former, if one pays some attention to certain measurability issues.
- Jerison and Kenig also proved a Hardy–Littlewood estimate (similar to the one given by Theorem 1.13) in the context of NTA domain.
- A Littlewood-type result (in the spirit of Theorem 1.25) also holds for NTA domains. See [14].

The unit disc is an NTA domain, and these results, once we set them in the unit disc, give a fairly precise answer to Question h in higher dimensions (with rough boundary), posed at the beginning of Sect. 1.2.

2 Return to Holomorphic Functions

Now the natural question is whether these results also hold in the study of the boundary behavior of holomorphic functions (in the Hardy spaces) for bounded domains $D \subset \mathbb{C}^n$. Here, as before, we set[25]

$$H^p(D) = \mathscr{O}(D) \cap h^p(D),$$

where $\mathscr{O}(D)$ is the space of all holomorphic functions defined on D.

Korányi's Results The first surprising result is due to Korányi in 1969. See [30]. Let

$$\mathbb{B} = \left\{ z \in \mathbb{C}^n : |z_1|^2 + |z_2|^2 + \cdots + |z_n|^2 < 1 \right\}$$

and, for $w \in b\mathbb{B}$ and $n > 1$, define[26]

$$\mathscr{K}_n(w) = \left\{ z \in \mathbb{B} : \frac{\text{distance}(z, b\mathbb{B})}{\text{distance}(z, \tau_w)} > \frac{1}{n} \right\},$$

where τ_w is the complex tangent space to $b\mathbb{B}$ at w. What is the shape of $\mathscr{K}_n(w)$? Since τ_w is the maximal complex subspace of the tangent hyperplane to $b\mathbb{B}$ at w, if $z \to w$ and $z \in \mathscr{K}_n(w)$, then $d_w(z)$ may actually converge to 0, provided $z - w$ is orthogonal to w *in the sense of the Hermitian inner product of* $\mathbb{C}^n$. Therefore $\mathscr{K}_n(w)$ is actually tangential to the boundary in the complex tangential directions. Korányi

[25] A different, somehow unyielding, definition of Hardy spaces of holomorphic functions for domains in $\mathbb{C}^n$ is due to Lumer. See [55] and references therein.

[26] This definition is due to Stein. See also [25].

proved that every function $F \in H^p(\mathbb{B})$ converges a.e. through $\mathscr{K}_n$ to $F_\flat$. This is a result of Fatou-type. This result was surprising because the approach regions $\mathscr{K}_n$ are actually tangential to the boundary (in the complex tangential directions).

Stein's Results Stein showed that it is possible to define approach regions in the spirit of $\mathscr{K}_n$ *for any bounded domain* D *in* $\mathbb{C}^n$ *with smooth boundary*. We shall also denote these approach regions by $\mathscr{K}_n$; see [60]. Moreover, he showed that every $F \in H^p(D)$ converges a.e. through $\mathscr{K}_n$ to $F_\flat$, and also proved a Hardy–Littlewood type inequality for $\mathscr{K}_n$.

The Optimality Issue for the Fatou-Type Theorem Here, as before, one may ask whether the approach regions $\mathscr{K}_n$ are optimal. In other words, whether a result of Littlewood-type (in the spirit of Theorem 1.17) holds for them. A more precise version of this question will be given momentarily.

Notation We denote by $\mathscr{K}_\bullet(w)$ the sequence $\{\mathscr{K}_n\}_n(w)$.

Definition 2.1 Let $D \subset \mathbb{C}^n$ be a bounded domain with smooth boundary. Let $w \in bD$. Let $A \subset D$, and assume that w belongs to the closure of A. We say that A lies eventually outside of $\mathscr{K}_\bullet(w)$ if the following condition holds:

For every $n > 1$, the set A has a tail[27] at w which is disjoint from $\mathscr{K}_n(w)$.

D'Angelo's Domains We denote by **DA** the class of all bounded domains D in $\mathbb{C}^n$ that are pseudoconvex, have smooth boundary, and are of finite type in the sense of D'Angelo. See [10].

Conjecture 2.2 *Let* $D \in$ ***DA****. Let* γ *be a family of approach regions in* D*. Assume that* γ *is* $*$*-connected,*[28] *that* γ *is regular,*[29] *and that the set*

$$E := \left\{w \in bD : \gamma(w) \text{ lies eventually outside } \mathscr{K}_\bullet\right\}$$

has positive outer measure. Then there is a bounded holomorphic functions on D *which, for a.e.* $w \in E$*, fails to have limits through* γ*.*

As a matter of fact, this conjecture is compatible with all results obtained so far in the area. In particular, it is compatible with results of Hakim and Sibony (1983); see [21] and [24], which are set in the unit ball $\mathbb{B}$.

Caveat Let us assume for a moment that Conjecture 2.2 turned out to be true. Then this would not be the end of the story! In order to make this point precise, we have to look at a subclass of **DA**.

27 A *tail of* A *at* w is a subset of D of the form $\{z \in D : z \in A, |z - w| < r\}$, for some $r > 0$.

28 As in Definition 1.23.

29 Recall that this condition means that the γ-shadow of any open subset of D is a measurable subset of the boundary. See Definition 1.21.

Convex D'Angelo's Domains Let us denote by **CDA** the class of all bounded domains D in $\mathbb{C}^n$ that are convex, have smooth boundary, and are of finite type in the sense of D'Angelo. See [5] and [42]. Observe that

$$\mathbf{CDA} \subset \mathbf{DA}.$$

The Adapted Approach Regions for CDA Let $D \in \mathbf{CDA}$. Then there exist in D still other approach regions that enter in the picture in a somewhat subtle manner. See [12]. We shall call these approach regions *adapted*,[30] and denote them by $\mathscr{C}_n(w)$, where $n > 1$ and $w \in bD$. The following facts might at first appear confusing to the reader but should eventually clarify the picture.

(i) A Fatou-type convergence result holds for $\mathscr{C}_\bullet$, i.e., every $F \in H^p(D)$ converges to $F_\flat$ a.e. through $\mathscr{C}_n$.
(ii) If $w \in bD$ is only weakly pseudoconvex,[31] then, for each $n > 1$, the set $\mathscr{C}_n(w)$ is *essentially larger than* $\mathscr{K}(w)$, i.e., $\mathscr{C}_n(w)$ contains a curve ending at w that lies eventually outside of $\mathscr{K}_\bullet(w)$.
(iii) a.e. $w \in bD$ is strongly pseudoconvex.
(iv) If $w \in bD$ is strongly pseudoconvex, then $\mathscr{C}_\bullet(w)$ is equivalent to $\mathscr{K}_\bullet(w)$, in the sense that

- for each $n > 1$ there exists $j > 1$ such that $\mathscr{C}_n(w)$ has a tail at w which is contained in $\mathscr{K}_j(w)$;
- for each $n > 1$ there exists $j > 1$ such that $\mathscr{K}_n(w)$ has a tail at w which is contained in $\mathscr{C}_j(w)$.

(v) The Hardy–Littlewood inequality holds for $\mathscr{C}_\bullet$, i.e., for each $p \in (0, \infty)$ and each $n > 1$, there exists $c = c(p, n)$ such that, if $F \in H^p(D)$, then

$$\int_{bD} \sup_{z \in \mathscr{C}_n(w)} \left|F(z)\right|^p \omega(dw) \leqslant c \int_{bD} \left|F_\flat(w)\right|^p \omega(dw),$$

where ω is the normalized surface measure on bD and $F_\flat(w)$ is the angular limit of F at w.
(vi) $\mathscr{C}_\bullet$ is distributionally larger than $\mathscr{K}_\bullet$, i.e., for each $n > 1$, there is no constant $c > 0$ such that the inequality

$$\omega\Big(\Big\{w \in bD : \sup_{z \in \mathscr{C}_n(w)} \varphi(z) > \lambda\Big\}\Big) \leqslant c\omega\Big(\Big\{w \in bD : \sup_{z \in \mathscr{K}_n(w)} \varphi(z) > \lambda\Big\}\Big)$$

holds for each function $\varphi : D \to [0, +\infty)$ and for each $\lambda > 0$. In other words, the distribution function of $\sup_{\mathscr{C}_n(\cdot)} |f|$ is not bounded in terms of the distribution function of $\sup_{\mathscr{K}_n(\cdot)} |f|$.

[30] We prefer to avoid the term *admissible*, since the latter has been used during the years to denote different objects.

[31] See [32].

One may ask: What is the point of introducing these "larger" approach regions $\mathscr{C}_\bullet$? The point is that a Hardy–Littlewood inequality holds for $\mathscr{C}_\bullet$, as we will now explain.

A Few Words of Explanation

- Observe that (i) follows immediately from (iii) and (iv), once we take into account Stein's results about $\mathscr{K}_\bullet$. In other words, the Fatou-type theorem for $\mathscr{C}_\bullet$ follows immediately from the Fatou-type theorem that Stein has proved for $\mathscr{K}_\bullet$.
- However, in view of (vi), (v) *does not* follow from Stein's results about $\mathscr{K}_\bullet$.

The reason that the adapted regions $\mathscr{C}_\bullet$ are interesting *does not* reside in the (qualitative) Fatou-type result given in (i), since (iii) and (iv) imply that the latter result follows immediately from the Fatou-type result that Stein proved for $\mathscr{K}_\bullet$. The reason that the adapted regions $\mathscr{C}_\bullet$ are interesting is that the (quantitative) Hardy–Littlewood maximal estimate given in (v) holds for these families of approach regions, and this result, in view of (vi), *does not* follow from the corresponding one that Stein proved for $\mathscr{K}_\bullet$.

Therefore, in several complex variables, the most salient result in the boundary behavior is the Hardy–Littlewood maximal inequality, that can be considered as a quantitative version of the a.e. convergence result.

So far, the "adapted" approach regions have been defined for convex domains of finite type only.[32]

Conjecture 2.3 *The adapted approach regions $\mathscr{C}_\bullet$ for domains in **CDA** are optimal for the Hardy–Littlewood maximal inequality, in the following sense. Let $D \in$ **CDA**. Assume that, for each $n > 1$ and $w \in bD$, an approach region $\mathscr{E}_n(w)$ at w in D is given. Assume that*

- *for each $n > 1$, $\mathscr{E}_n$ is a regular and $*$-connected family of approach regions in D;*
- *for each $n > 1$ and $j > 1$, there is a constant c such that*

$$\omega\Big(\Big\{w \in bD : \sup_{z\in\mathscr{E}_n(w)} \varphi(z) > \lambda\Big\}\Big) \leqslant c\omega\Big(\Big\{w \in bD : \sup_{z\in\mathscr{E}_j(w)} \varphi(z) > \lambda\Big\}\Big)$$

for each positive-valued function φ on D and each $\lambda > 0$;
- *if $0 < p \leqslant \infty$ and $F \in H^p(D)$ then F converges to $F_\flat$ a.e. through $\mathscr{E}_n$;*
- *the Hardy–Littlewood inequality holds for $\mathscr{E}_\bullet$, i.e., for each $p \in (0, \infty)$ there exists $c = c(p)$ such that, if $F \in H^p(D)$, then*

$$\int_{bD} \sup_{z\in\mathscr{E}_1(w)} \big|F(z)\big|^p \omega(dw) \leqslant c \int_{bD} \big|F_\flat(w)\big|^p \omega(dw).$$

[32] In the specific case of the "egg domains", such as, for example, $\{z \in \mathbb{C}^2 : |z_1|^2 + |z_2|^4 < 1\}$ the "adapted" approach regions made their first appearance in [6].

Then there exists a constant c such that

$$\omega\Big(\Big\{w \in bD : \sup_{z \in \mathscr{E}_1(w)} \varphi(z) > \lambda\Big\}\Big) \leqslant c\omega\Big(\Big\{w \in bD : \sup_{z \in \mathscr{C}_1(w)} \varphi(z) > \lambda\Big\}\Big)$$

for each positive-valued function φ on D and $\lambda > 0$.

In other words, Conjecture 2.3 says that $\mathscr{C}_\bullet$ is optimal for the Hardy–Littlewood maximal inequality, in the sense that *nothing larger will make the Hardy–Littlewood inequality work*, where *larger* is defined in terms of the distribution function.

The following observation will clarify Conjecture 2.3.

Remark 2.4 Assume that $D \subset \mathbb{C}^3$ belongs to **CDA**, and, to fix ideas, suppose that

$$D = \big\{z \in \mathbb{C}^3 : |z_1|^2 + |z_2|^4 + |z_3|^6 < 1\big\}.$$

Then the construction given in [44] applies and yields a certain one-parameter family $\mathscr{A}_n(w)$ of approach regions in D (called *admissible* in [44]), with the following properties. See also [45].

(i) Each functions $F \in H^p(D)$ converges a.e. to $F_\flat$ through $\mathscr{A}_n$.
(ii) For a.e. $w \in bD$, $\mathscr{A}_\bullet(w)$ is equivalent to $\mathscr{K}_\bullet(w)$.
(iii) The Hardy–Littlewood estimate holds for $\mathscr{A}_n$.
(iv) $\mathscr{A}_\bullet$ is distributionally larger than $\mathscr{K}_\bullet$.

Observe that here as well, in light of (ii), the a.e. convergence result (i) is an immediate consequence of the corresponding one for $\mathscr{K}_\bullet$, but, in light of (iv), the Hardy–Littlewood estimate for $\mathscr{A}_\bullet$ *does not* follow from the corresponding one that is known to hold for $\mathscr{K}_\bullet$.

However, if we look at how the approach regions $\mathscr{A}_n(w)$ compare with $\mathscr{C}_n(w)$, we find out that, on the one hand, $\mathscr{C}_\bullet$ is distributionally larger than $\mathscr{A}_\bullet$, and, on the other hand, $\mathscr{A}_\bullet$ is subordinate to $\mathscr{C}_\bullet$, in the sense that, for each $n > 1$, there is a constant c such that

$$\omega\Big(\Big\{w \in bD : \sup_{z \in \mathscr{A}_n(w)} \varphi(z) > \lambda\Big\}\Big) \leqslant c\omega\Big(\Big\{w \in bD : \sup_{z \in \mathscr{C}_n(w)} \varphi(z) > \lambda\Big\}\Big)$$

for all φ and $\lambda > 0$, as above. It follows that the Hardy–Littlewood maximal estimate for $\mathscr{A}_\bullet$ given in (iii) is an immediate consequence of the corresponding one that holds for $\mathscr{C}_\bullet$, but the converse is false. In other words, the admissible approach regions $\mathscr{A}_\bullet$ are *not* optimal for the Hardy–Littlewood maximal estimate.

We like to stress that the a.e. convergence results for $\mathscr{A}_\bullet$ and for $\mathscr{C}_\bullet$ are both an immediate consequence of the a.e. convergence result that is known to hold for $\mathscr{K}_\bullet$, given the fact that these approach regions are a.e. equivalent to $\mathscr{K}_\bullet$.

Other Open Problems

(i) So far, the "adapted" approach regions have been defined for convex domains of finite type only. Is it possible to extend this definition to all domains in **DA**, and obtain an optimality result similar to the one described in Conjecture 2.3?
(ii) It would be interesting to relax the smoothness assumption on the boundary as much as possible, just as in the case of the NTA domains for potential theory.[33]
(iii) It would be interesting to find out whether there is a diffusion process that plays a role in this context, just as Brownian motion plays a role in potential theory.[34]
(iv) It would be interesting to develop this theory in an intrinsic setting, along the lines of Ancona's work.[35] Cf. [3].

Čirka, in the early seventies, wrote that

> [...] in the still young theory of boundary properties of holomorphic functions of several complex variables there is an entire field of white nothingness with widely spaced isolated results

(see [8]). It seems that now at least we see the contours of the field.

Acknowledgements I am deeply grateful to Steven G. Krantz for sharing with me his insight during many years, first as a mentor and then as a friend and colleague.

References

1. Aikawa, H.: Harmonic functions having no tangential limits. Proc. Am. Math. Soc. **108**, 457–464 (1990)
2. Aikawa, H.: Harmonic functions and Green potentials having no tangential limits. J. Lond. Math. Soc. (2) **43**, 125–136 (1991)
3. Ancona, A.: Negatively curved manifolds, elliptic operators, and the martin boundary. Ann. of Math. (2) **125**, 495–536 (1987)
4. Beurling, A.: Etudes sur un problème de majoration. Almqvist & Wiksell, Uppsala (1933)
5. Boas, H.P., Straube, E.J.: On equality of line type and variety type of real hypersurfaces in $\mathbb{C}^n$. J. Geom. Anal. **2**, 95–98 (1992)
6. Bonami, A., Lohoué, N.: À propos du théorème de Fatou sur certains domaines faiblement pseudo-convexes. In: Seminar on Harmonic Analysis (1979–1980). Publ. Math. Orsay 80, vol. 7, pp. 34–39. Univ. Paris XI, Orsay (1980)
7. Bourbaki, N.: Elements of Mathematics. Integration I. Chapters 1–6. Springer, Berlin (2004)
8. Čirka, E.M.: The Lindelöf and Fatou theorems in C^n. Mat. Sb. (N.S.) **92**(134), 622–644 (1973). English translation in: Math. USSR Sb. **21**, 619–639 (1973)
9. Cohen, P.J.: Set Theory and the Continuum Hypothesis. Benjamins, Amsterdam (1966)
10. D'Angelo, J.P.: Several Complex Variables and the Geometry of Real Hypersurfaces. CRC Press, Boca Raton (1993)

---

[33]The notion of finite type can be actually be redefined so as to avoid smoothness assumptions on the boundary.

[34]P. Malliavin proved some results that give hints in this direction. See [41].

[35]Krantz proved some results in this direction. See [33]. See also [13].

11. Di Biase, F.: Fatou Type Theorems: Maximal Functions and Approach Regions. Birkhäuser, Boston (1998)
12. Di Biase, F., Fischer, B.: Boundary behavior of H^p functions on convex domains of finite type in $\mathbb{C}^n$. Pac. J. Math. **183**, 25–38 (1998)
13. Di Biase, F., Krantz, S.G.: Boundary Behavior of Holomorphic Functions. Birkhäuser Mathematics (2013, to appear)
14. Di Biase, F., Stokolos, A., Svensson, O., Weiss, T.: On the sharpness of the Stolz approach. Ann. Acad. Sci. Fenn. Math. **31**, 47–59 (2006)
15. Doob, J.L.: Classical Potential Theory and Its Probabilistic Counterpart. Springer, New York (1984)
16. Drake, F.R.: Set Theory. North-Holland, Amsterdam (1974)
17. Fatou, P.: Séries trigonométriques et séries de Taylor. Acta Math. **30**, 335–400 (1906)
18. Fefferman, C., Stein, E.M.: Some maximal inequalities. Am. J. Math. **93**, 107–115 (1971)
19. Figà-Talamanca, A.: Personal communication. September 14, 2009
20. Gödel, K.: What is Cantor's continuum problem? Am. Math. Mon. **54**, 515–525 (1983)
21. Hakim, M., Sibony, N.: Fonctions holomorphes bornées et limites tangentielles. Duke Math. J. **50**, 133–141 (1983)
22. Hardy, G.H., Littlewood, J.E.: A maximal inequality with function-theoretic applications. Acta Math. **54**, 81–116 (1930)
23. Hartogs, F.: Eine folgerungen aus der Cauchyschen integral formel bei functionen mehrer Veranderlichen. Sitzungsber. Math.-Phys. Kl. Bayer. Akad. Wiss. Münch. **36**, 223–242 (1906)
24. Hirata, K.: Sharpness of the Korányi approach region. Proc. Am. Math. Soc. **133**, 2309–2317 (2005)
25. Hörmander, L.: L^p estimates for (pluri-) subharmonic functions. Math. Scand. **20**, 65–78 (1967)
26. Jech, T.: Set Theory. Academic Press, New York (1978)
27. Jerison, D., Kenig, C.: Boundary behavior of harmonic functions in non-tangentially accessible domains. Adv. Math. **46**, 80–147 (1982)
28. Keldych, M.V.: On the solubility and stability of the Dirichlet problem. Usp. Mat. Nauk **8**, 171–231 (1941)
29. Keldych, M.V.: Sur le problème de Dirichlet. C. R. (Dokl.) Acad. Sci. URSS **32**, 308–309 (1941)
30. Korányi, A.: Harmonic functions on Hermitian hyperbolic space. Trans. Am. Math. Soc. **135**, 507–516 (1969)
31. Kouznetzoff, P., Privalov, I.I.: Sur les problèmes limites et les classes différentes de fonctions harmoniques et subharmoniques définies dans un domaine arbitraire. Rec. Math. [Mat. Sb.] New Ser. **6** (Old Series **48**), 345–376 (1939)
32. Krantz, S.G.: Function Theory of Several Complex Variables. Wadsworth & Brooks/Cole, Pacific Grove (1992)
33. Krantz, S.G.: Invariant metrics and the boundary behavior of holomorphic functions on domains in $\mathbb{C}^n$. J. Geom. Anal. **1**, 71–97 (1991)
34. Kunen, K.: Set Theory. An Introduction to Independence Proofs. North-Holland, Amsterdam (1980)
35. Lebesgue, H.: Sur les fonctions représentables analytiquement. J. Math. Pures Appl. (6) **1**, 139–216 (1905)
36. Lebesgue, H.: Sur des cas d'impossibilité du problème de Dirichlet ordinaire. Bull. Soc. Math. Fr. **41** (supplément spécial: Vie de la Société), 17 (1912)
37. Littlewood, J.E.: Mathematical notes (4): On a theorem of Fatou. J. Lond. Math. Soc. **2**, 172–176 (1927)
38. Lohwater, A.J., Piranian, G.: The Boundary Behavior of Functions Analytic in a Disc. Ann. Acad. Sci. Fenn. Ser. A. 1 Math., vol. 239, pp. 17 (1957)
39. Luzin, N.N.: Sur la convergence des séries trigonométriques de Fourier. C. R. Math. Acad. Sci. Paris **156**, 1655–1658 (1913)

40. Luzin, N.N.: Integral i trigonometriceskii ryad [Integral and Trigonometric Series]. Mat. Sb. **30**, 1–242 (1916)
41. Malliavin, P.: Équation de la chaleur associée à une fonction plurisousharmonique d'exhaustion et comportement frontière. Ann. Inst. Fourier **25**, 447–464 (1975)
42. McNeal, J.D.: Convex domains of finite type. J. Funct. Anal. **108**, 61–373 (1992)
43. Nagel, A., Stein, E.M.: On certain maximal functions and approach regions. Adv. Math. **54**, 83–106 (1984)
44. Nagel, A., Stein, E.M., Wainger, S.: Boundary behavior of functions holomorphic in domains of finite type. Proc. Natl. Acad. Sci. USA **78**, 6596–6599 (1981)
45. Nagel, A., Stein, E.M., Wainger, S.: Balls and metrics defined by vector fields I: Basic properties. Acta Math. **155**, 103–147 (1985)
46. Osgood, W.: Note über analytische Functionen mehrerer Veränderlichen. Math. Ann. **52**, 462–464 (1899)
47. Osgood, W.: On functions of several complex variables. Trans. Am. Math. Soc. **17**, 1–8 (1916)
48. Perron, O.: Eine neue Behandlung der ersten Randwertaufgabe fuer $\Delta u = 0$. Math. Z. **18**, 42–54 (1923)
49. Plessner, A.: Über das verhalten analytischer Funktionen am Rande ihres Definitionsbereics. J. Reine Angew. Math. **158**, 219–227 (1927)
50. Poincaré, H.: Sur les équations aux dérivées partielles de la physique mathématique. Am. J. Math. **12**, 211–294 (1890)
51. Privalov, I.I.: Sur une généralisation du théorème de Fatou. Mat. Sb. **31**, 232–235 (1923)
52. Prym, F.E.: Zur integration der differentialgleichung $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$. J. Reine Angew. Math. **73**, 340–364 (1871)
53. Remak, R.: Über potentialconvexe Funktionen. Math. Z. **20**, 126–130 (1924)
54. Riesz, F.: Über die Randwerte einer analytischen Funktion. Math. Z. **18**, 87–95 (1923)
55. Rudin, W.: Lumer's Hardy spaces. Mich. Math. J. **24**, 1–5 (1977)
56. Rudin, W.: Inner function images of radii. Math. Proc. Camb. Philos. Soc. **85**, 357–360 (1979)
57. Schwarz, H.A.: Zur Integration der partiellen Differentialgleichung $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$ für die Fläche eines Kreises. Vierteljahrsschr. Nat.forsch. Ges. Zür. **15**, 113–128 (1870)
58. Schwarz, H.A.: Zur Integration der partiellen Differentialgleichung $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$. J. Reine Angew. Math. **74**, 218–253 (1872)
59. Stein, E.M.: Singular Integrals and Differentiability Properties of Functions. Princeton University Press, Princeton (1970)
60. Stein, E.M.: Bounday Behaviour of Holomorphic Functions of Several Complex Variables. Princeton University Press, Princeton (1972)
61. Suslin, M.: Sur une définition des ensembles mesurables B sans nombres transfinis. C. R. Math. Acad. Sci. Paris **164**, 88–91 (1917)
62. Tsuji, M.: On the boundary value of a harmonic function in space. Jpn. J. Math. **19**, 111–137 (1944)
63. Tsuji, M.: On Hardy–Littlewood's maximal theorem. Comment. Math. Univ. St. Pauli **6**, 33–42 (1957)
64. Volterra, V.: Sul principio di Dirichlet. Rend. Circ. Mat. Palermo **XI**, 83–96 (1897)
65. Wiener, N.: Certain notions in potential theory. J. Math. Phys. Mass. Inst. Tech. **3**, 24–51 (1924)
66. Zaremba, S.: Sur le principe de Dirichlet. Acta Math. **34**, 293–316 (1911)

Constructing Laplacians on Limit Spaces of Self-similar Groups

Alfredo Donno

Abstract This paper is basically an explanatory survey focusing on a new and stimulating connection between Analysis on fractals and self-similar group Theory. It is shown how to construct Dirichlet forms on fractals which are obtained as limit spaces of contracting self-similar groups acting on regular rooted trees. The key idea is to give a prefractal approximation of the limit space by using the sequence of finite Schreier graphs of the action of the group: under certain conditions of compatibility, a Dirichlet form on the limit space is obtained as limit of the sequence of finite Dirichlet forms associated with the classical discrete Laplacian on them. Some known and new examples are described.

Keywords Dirichlet form · Laplacian · Post-critically finite self-similar fractal · Limit space · Julia set · Schreier graph · Self-similar group

Mathematics Subject Classification (2010) 28A80 · 31C25 · 20E08 · 37F50

1 Introduction

The problem of constructing a Laplacian on a fractal, in order to describe physical phenomena like percolation, diffusion across highly conductive layers, diffusion through porous media, is one of the most important target of modern mathematics and this is, in particular, what Analysis on fractals is about. Motivated by studies in Physics, Kusuoka [20] and Goldstein [14] made the first step in this direction, by constructing a Brownian motion on the Sierpiński gasket. This method is now called the "probabilistic approach". They considered a sequence of random walks on graphs approximating the Sierpiński gasket (see Fig. 1) and proved that, by taking a certain scaling factor, those random walks converged to a diffusion process on the Sierpiński gasket. In this probabilistic approach, a Laplacian is the infinitesimal generator of the semigroup associated with the diffusion process. In [18], a direct

A. Donno (✉)
Dipartimento di Scienze di Base e Applicate per l'Ingegneria, Sezione di Matematica, Sapienza Università di Roma, Via A. Scarpa 16, 00161 Rome, Italy
e-mail: alfredo.donno@sbai.uniroma1.it

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3,
DOI 10.1007/978-88-470-2853-1_10,

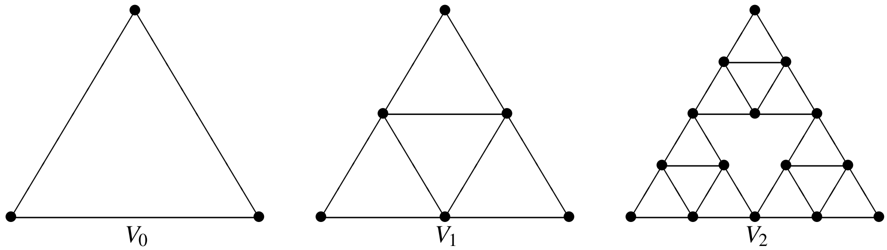


Fig. 1 Approximation of the Sierpiński gasket by graphs V_m

definition of a Laplacian on the Sierpiński gasket was given, following the so-called "analytical approach". Instead of a sequence of random walks, Kigami considered a sequence of discrete Laplacians on finite graphs approximating the Sierpiński gasket, proving that, by choosing a proper scaling, those discrete Laplacians converged to a good operator, called the "standard Laplacian" on the Sierpiński gasket. He described the structure of harmonic functions, Green's functions, Dirichlet forms: these constructions were later extended to post-critically finite self-similar sets [19], which essentially represent the only fractals on which such analysis has been developed. It is worth mentioning the recent paper [7], where a definition of differential 1-forms on the Sierpiński gasket and their integral on paths is given. In that setting, the Dirichlet form studied by Kigami provides the class of finite energy functions, which plays the role of a Sobolev space on the fractal, but it also allows to develop a canonical first order differential calculus, as described in [8].

In [27] V. Nekrashevych and A. Teplyaev investigate the relation between Analysis on fractals and self-similar group Theory. More precisely, with each contracting self-similar group one can associate a limit space defined as a quotient of the space $X^{-\infty}$ of the left-infinite words over a finite alphabet X, modulo an asymptotic equivalence relation that can be described using the Moore diagram of the automaton generating the group. This limit space is approximated by a sequence of finite graphs, namely the finite Schreier graphs associated with the action of the group on the regular rooted tree of degree $|X|$ (Theorem 3.2).

Under certain conditions, the limit space is homeomorphic to a post-critically finite self-similar fractal, and so one can obtain a Laplacian on it by using the methods of Kigami [19]: a sequence of finite Dirichlet forms and the associated Laplacians can be recursively constructed on the finite Schreier graphs and, if this sequence is compatible in a certain sense, then a Laplacian on the limit space is obtained as limit of the Laplacians on the finite Schreier graphs, and it is associated with a local regular Dirichlet form on the limit space. Having a Laplacian, one can investigate the spectral problems, solutions to wave and heat equations and so on. The existence of Dirichlet forms on a post-critically finite self-similar fractal was proven only under special symmetry conditions, e.g., for *nested fractals* (Lindstrøm [21]) or *strongly symmetric fractals* (Kigami [19, Theorem 3.8.10]).

In the latest years, several efforts have been made in order to enlarge the class of fractals on which harmonic analysis can be done, i.e., to define an energy form (and hence a Laplacian) also for non self-similar fractals. We mention, for instance, the papers [23, 24] of Mosco, where a Lagrangian approach is developed, i.e., the energy form is obtained by integrating a local energy measure, which is called Lagrangian, or the paper [13] by Freiberg and Lancia, where energy forms on conformal C^1-diffeomorphic images of the Sierpiński gasket are constructed, by mixing a Lagrangian approach and a pull-back argument.

The interaction between Analysis on fractals and the theory of self-similar groups seems very stimulating in this direction and provides a new class of examples. First, one can ask whether there is a natural non-trivial Laplacian on the limit space of any contracting group and how it inherits the property of self-similarity of the group. Moreover, one can ask whether discrete Laplacians on the finite Schreier graphs of a contracting group converge, with some appropriate normalization, to a non-trivial Laplacian on the limit space.

Schreier graphs arise naturally from the action of a group on a set. In this paper, we consider groups acting by automorphisms on rooted trees. Let T be a regular rooted tree and $G < \mathrm{Aut}(T)$ be a finitely generated group of automorphisms of T. By fixing a finite set S of generators of G, we naturally get a sequence $\{\Gamma_n\}_{n\geqslant 1}$ of finite Schreier graphs of the action of G on T. The vertex set of Γ_n coincides with the set L_n of vertices of the nth level of T, and two vertices $v, v' \in L_n$ are connected by an edge in Γ_n if there exists $s \in S$ such that $s(v) = v'$. If G acts transitively on each level, then it follows that the graph Γ_n is connected, for each $n \geqslant 1$.

Similarly, the action of G on the boundary ∂T of the tree gives rise to an uncountable family of infinite orbital Schreier graphs $\{\Gamma_\xi\}_{\xi\in\partial T}$, with $V(\Gamma_\xi) = G \cdot \xi$. It turns out that, for each $\xi \in \partial T$, the orbital Schreier graph (Γ_ξ, ξ), viewed as a graph rooted at the vertex ξ, is exactly the limit, in the pointed Gromov-Hausdorff topology, of the sequence $\{(\Gamma_n, \xi_n)\}_{n\geqslant 1}$ of the finite Schreier graphs rooted at ξ_n, where ξ_n denotes the prefix of length n of ξ (see Sect. 2.2).

The study of Schreier graphs of some examples of groups acting on rooted trees was initiated by L. Bartholdi and R.I. Grigorchuk [1] in connection with the problem of determining the spectrum of the associated Laplace operator. Particularly interesting examples come from the class of self-similar groups, which are connected to self-similar sets via the notion of limit space, which is a compact space that can be associated with any contracting self-similar group [26]. In the case of Iterated Monodromy Groups of post-critically finite rational functions, the limit space associated with the monodromy action on the tree is homeomorphic to the Julia set of the rational map and it has a fractal nature. Finite Schreier graphs form an approximating sequence for this limit space. Recently, several models coming from physics and statistical mechanics have been studied on Schreier graphs of self-similar groups [10, 11, 22]. See also [6, 12], where the partition function of the Ising model on those graphs is obtained as a special evaluation of the Tutte polynomial of them.

The paper is organized as follows. In Sect. 2, we recall some basic facts concerning self-similar groups and Schreier graphs. In Sect. 3, we present the notion of limit space of a contracting self-similar group, focusing our attention on iterated monodromy groups. The definition of post-critically finite self-similar fractal is given in Sect. 3.3. In Sect. 4, we illustrate the Kigami methods [19] for the construction of self-similar Dirichlet forms on post-critically finite self-similar fractals.

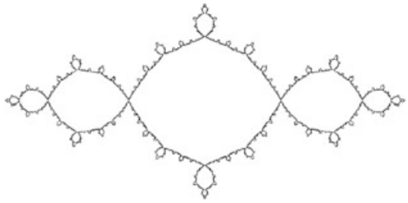


Fig. 2 Julia set of the polynomial $z^2 - 1$ defining the Basilica group

Finally, in Sect. 5, we explicitly consider some known and new examples of limit spaces of iterated monodromy groups, and we use the theory from Sect. 4 in order to obtain Dirichlet forms on them.

2 Self-similar Groups and Schreier Graphs

In this section we recall some basic definitions and properties about self-similar groups and Schreier graphs associated with their action on the regular rooted tree. Finite Schreier graphs will play a crucial role in the next sections, by providing a sequence of graphs approximating the limit space associated with any contracting self-similar group (see Theorem 3.2).

2.1 Groups Acting on Rooted Trees

Let T be the infinite regular rooted tree of degree q, i.e., the rooted tree in which each vertex has q offsprings. Given a finite alphabet $X = \{0, 1, \dots, q-1\}$ of q elements, we denote by X^n the set of words of length n over X and put $X^* = \bigcup_{n \geqslant 0} X^n$, where the set X^0 consists of the empty word. Moreover, we denote X^∞ the set of (right-) infinite words in X. In this way, each vertex of the nth level L_n of the tree can be regarded as an element of X^n and the set X^∞ can be identified with the set ∂T of infinite geodesic rays emanating from the root of T. The set X^∞ can be equipped with the direct product topology. The basis of open sets is the collection of all cylindrical sets $\{wX^\infty, w \in X^*\}$. The space X^∞ is totally disconnected and homeomorphic to the Cantor set. The cylindrical sets generate a σ-algebra of Borel subsets of the space X^∞.

We denote by $\mathrm{Aut}(T)$ the group of all automorphisms of T, i.e., the group of all bijections of the set of vertices of T preserving the incidence relation. Clearly, the root and hence the levels of the tree are preserved by any automorphism of T. A group $G \leqslant \mathrm{Aut}(T)$ is said *spherically transitive* if it acts transitively on each level of the tree.

The stabilizer of a vertex v in T is the subgroup of G defined as $Stab_G(v) = \{g \in G : g(v) = v\}$; the stabilizer of the level L_n is $Stab_G(L_n) = \bigcap_{v \in L_n} Stab_G(v)$; finally, the stabilizer of a boundary point $\xi \in X^\infty$ is $Stab_G(\xi) = \{g \in G : g(\xi) = \xi\}$.

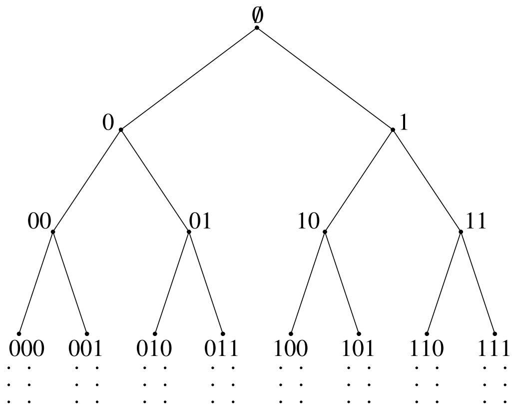


Fig. 3 The regular rooted tree of degree 2

If $g \in \mathrm{Aut}(T)$ and $v \in X^*$, define $g|_v \in \mathrm{Aut}(T)$, called the *restriction* of the action of g to the subtree rooted at v, by

$$g(vw) = g(v)g|_v(w),$$

for all $v, w \in X^*$. Every subtree of T rooted at a vertex is isomorphic to T. Therefore, every automorphism $g \in \mathrm{Aut}(T)$ induces a permutation of the vertices of the first level of the tree and q restrictions, $g|_0, \ldots, g|_{q-1}$, to the subtrees rooted at the vertices of the first level. Hence, it can be written as $g = \tau_g(g|_0, \ldots, g|_{q-1})$, where $\tau_g \in \mathrm{Sym}(q)$ describes the action of g on L_1.

Definition 2.1 A group G acting by automorphisms on a regular rooted tree T of degree q is *self-similar* if $g|_v \in G$ for every $v \in X^*$ and $g \in G$.

It follows that, if G is self-similar, an automorphism $g \in G$ can be represented as $g = \tau_g(g|_0, \ldots, g|_{q-1})$, where $\tau_g \in \mathrm{Sym}(q)$ describes the action of g on L_1, and $g|_i \in G$ is the restriction of the action of g on the subtree T_i rooted at the ith vertex of the first level. So, for every $x \in X$ and $w \in X^*$, one has:

$$g(xw) = \tau_g(x)g|_x(w).$$

Definition 2.2 A self-similar group G is *fractal* (or self-replicating) if it acts transitively on the first level of the tree and, for all $x \in X$, the map $g \mapsto g|_x$ from $Stab_G(x)$ to G is surjective.

2.2 Schreier Graphs

Let $G < \mathrm{Aut}(T)$ be a group acting on T by automorphisms, generated by a finite symmetric set $S \subset G$. Throughout the paper we will assume that the action of G is spherically transitive (note that any action by automorphisms is level-preserving).

Definition 2.3 The nth *Schreier graph* Γ_n of the action of G on T, with respect to the generating set S, is a (labeled) graph with $V(\Gamma_n) = X^n$, and edges (u, v) between vertices u and v such that u is moved to v by the action of some generator $s \in S$. The edge (u, v) is then labeled by s.

For an infinite ray $\xi \in \partial T$, the *orbital Schreier graph* Γ_ξ has vertex set $G \cdot \xi$ and the edge set determined by the action of generators on this orbit, as above.

It is not difficult to see that the orbital Schreier graphs are infinite and that the finite Schreier graphs $\{\Gamma_n\}_{n=1}^{\infty}$ form a sequence of graph coverings. More precisely, for each $n \geqslant 1$, the projection $\pi_{n+1} : V(\Gamma_{n+1}) \longrightarrow V(\Gamma_n)$ defined by $\pi_{n+1}(x_1 \dots x_n x_{n+1}) = x_1 \dots x_n$ induces a surjective morphism between Γ_{n+1} and Γ_n, which is a graph covering of degree q.

Finite Schreier graphs converge to infinite Schreier graphs in the space of rooted (labeled) graphs with local convergence (*rooted Gromov-Hausdorff convergence* [17, Chap. 3]). More precisely, for an infinite ray $\xi \in X^\infty$ denote by ξ_n the nth prefix of the word ξ. Then the sequence of rooted graphs $\{(\Gamma_n, \xi_n)\}_{n=1}^{\infty}$ converges to the infinite rooted graph (Γ_ξ, ξ) in the space $\mathscr{X}$ of (rooted isomorphism classes of) rooted graphs endowed with the following metric: the distance between two rooted graphs (Y_1, v_1) and (Y_2, v_2) is

$$\mathrm{Dist}\big((Y_1, v_1), (Y_2, v_2)\big) := \inf\left\{\frac{1}{r+1} : B_{Y_1}(v_1, r) \text{ is isomorphic to } B_{Y_2}(v_2, r)\right\}$$

where $B_Y(v, r)$ is the ball of radius r in Y centered in v.

2.3 Self-similar Groups and Automata

An *automaton* is a quadruple $\mathscr{A} = (\mathscr{S}, X, \mu, \nu)$, where $\mathscr{S}$ is the set of states; X is an alphabet; $\mu : \mathscr{S} \times X \to \mathscr{S}$ is the transition map; and $\nu : \mathscr{S} \times X \to X$ is the output map. The automaton $\mathscr{A}$ is finite if $\mathscr{S}$ is finite and it is invertible if, for all $s \in \mathscr{S}$, the transformation $\nu(s, \cdot) : X \to X$ is a permutation of X. An automaton $\mathscr{A}$ can be represented by its *Moore diagram*. This is a directed labeled graph whose vertices are identified with the states of $\mathscr{A}$. For every state $s \in \mathscr{S}$ and every letter $x \in X$, the diagram has an arrow from s to $\mu(s, x)$ labeled by $x|\nu(s, x)$. A natural action on the words over X is induced, so that the maps μ and ν can be extended to $\mathscr{S} \times X^*$:

$$\begin{aligned} \mu(s, xw) &= \mu\big(\mu(s, x), w\big) \\ \nu(s, xw) &= \nu(s, x)\nu\big(\mu(s, x), w\big), \end{aligned} \tag{1}$$

where we set $\mu(s, \emptyset) = s$ and $\nu(s, \emptyset) = \emptyset$, for all $s \in \mathscr{S}, x \in X$ and $w \in X^*$. Moreover, (1) defines uniquely a map $\nu : \mathscr{S} \times X^\infty \to X^\infty$.

If we fix an initial state s in an automaton $\mathscr{A}$, then the transformation $\nu(s, \cdot)$ on the set $X^* \cup X^\infty$ is defined by (1); it is denoted by $\mathscr{A}_s$. The image of a word $x_1 x_2 \dots$ under $\mathscr{A}_s$ can be easily found using the Moore diagram.

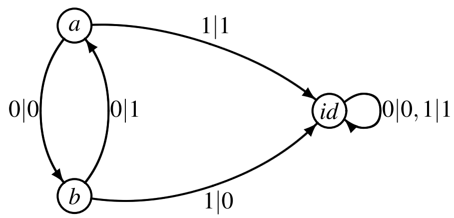


Fig. 4 The automaton generating the Basilica group

More precisely, consider the directed path starting at the state s with consecutive labels $x_1|y_1, x_2|y_2, \ldots$; the image of the word $x_1x_2\ldots$ under the transformation $\mathscr{A}_s$ is then $y_1y_2\ldots$. More generally, given an invertible automaton $\mathscr{A} = (\mathscr{S}, X, \mu, \nu)$, one can consider the group generated by the transformations $\mathscr{A}_s$, for $s \in \mathscr{S}$; this group is called the *automaton group* generated by $\mathscr{A}$ and is denoted by $G(\mathscr{A})$. See, for instance, the automaton in Fig. 4 generating the *Basilica* group, which is an automorphisms group of the rooted binary tree, whose generators have the following self-similar presentation:

$$a = (b, id), \qquad b = (01)(a, id). \tag{2}$$

A basic theorem [26] states that the action of a group G on $X^* \cup X^\infty$ is self-similar if and only if G is generated by an invertible automaton.

Let $\mathscr{A}$ be a finite automaton with the set of states $\mathscr{S}$ and alphabet X and let us denote $\alpha(k, s)$, for $k \in \mathbb{N}$ and $s \in \mathscr{S}$, the number of words $w \in X^k$ such that $s|_w \neq id$. Sidki suggested to call $\mathscr{A}$ *bounded*, if the sequence $\alpha(k, s)$ is bounded as a function of k for each state $s \in \mathscr{S}$. He showed in [30] that a finite invertible automaton is bounded if and only if any two non-trivial cycles in the Moore diagram of the automaton are disjoint and not connected by a directed path. More generally, the automaton $\mathscr{A}$ (or, equivalently, the group $G(\mathscr{A})$), is said *polynomial* of degree n if the sequence $\alpha(k, s)$ is bounded by a polynomial of the smallest degree n for each $s \in \mathscr{S}$ (see definition in [30] or [3, Chap. IV]). In [4], finite and infinite Schreier graphs associated with a group generated by a polynomial but non bounded automaton are studied, in connection with isomorphism problem, growth, amenability.

It can be shown that any group generated by a bounded automaton is *contracting* [5], i.e., there exists a finite set $\mathscr{N} \subset G$ such that for every $g \in G$ there exists $k \in \mathbb{N}$ such that $g|_v \in \mathscr{N}$, for all words v of length greater or equal to k. The smallest set $\mathscr{N}$ satisfying this property is called the *nucleus* of the contracting group. For a finitely generated self-similar group, the contracting property means that the length of the group elements contracts under taking restrictions.

2.4 Matrix Recursion

We know that if G is a self-similar group, then every automorphism $g \in G$ induces a permutation of the vertices of the first level of the tree and q restrictions, $g|_0, \ldots, g|_{q-1}$, to the subtrees rooted at the vertices of the first level, with $g|_i \in G$.

Therefore, it can be written as $g = \tau_g(g|_0, \dots, g|_{q-1})$, where $\tau_g \in \mathrm{Sym}(q)$ describes the action of g on L_1.

This property can be encoded by associating a matrix $\psi(g) = (a_{ij})_{i,j=0,\dots,q-1}$ with each $g \in G$, defined as follows:

$$a_{ij} = \begin{cases} g|_i & \text{if } g(i) = j, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, $\psi(g)$ is obtained from a permutational matrix by replacing its non-zero entries by elements of G. The map ψ, called *matrix recursion*, can be extended, by linearity, to a homomorphism from $\mathbb{C}G$ to $\mathscr{M}_{q\times q}(\mathbb{C}G)$. This recursion can be iterated, so that we get a map $\psi_n : \mathbb{C}G \to \mathscr{M}_{q^n\times q^n}(\mathbb{C}G)$, defined by recurrence by replacing each entry a_{ij} of $\psi_{n-1}(g)$ by $\psi(a_{ij})$, and putting $\psi_1 = \psi$.

Remark 2.4 It follows from the definition of matrix recursion and of Schreier graph that, if S is a symmetric generating set of G, then the matrix $M = \sum_{s\in S}\psi_n(s)$ is the *adjacency matrix* of the Schreier graph Γ_n, for each $n \geqslant 1$. In order to get this interpretation, the rows and columns of M have to be indexed by the vertices of Γ_n (words of length n over the alphabet $\{0, 1, \dots, q-1\}$), ordered lexicographically.

3 Limit Spaces and Post-Critically Finite Self-similar Fractals

In this section, the definition of limit space associated with the self-similar action of a contracting group is presented, with a special attention to the case of iterated monodromy groups of a rational function. We also recall the classical definition of self-similar structure (see, for instance, [19]).

3.1 Limit Spaces of Contracting Self-similar Groups

Let G be a contracting self-similar group with nucleus $\mathscr{N}$ and denote by $X^{-\infty} = \{\dots x_2x_1 : x_i \in X\}$ the set of left-infinite words over the alphabet X.

Definition 3.1 Two sequences $\dots x_2x_1$ and $\dots y_2y_1$ in $X^{-\infty}$ are said asymptotically equivalent if there exists a finite set $K \subset G$ and a sequence $\{g_n\}_{n\geqslant 1}$, with $g_n \in K$, such that

$$g_n(x_n \dots x_1) = y_n \dots y_1, \quad \text{for every } n \geqslant 1.$$

The quotient $\mathscr{J}_G$ of $X^{-\infty}$ modulo this equivalence relation is the *limit space* of G.

Since the shift map s defined by

$$s(\dots x_3x_2x_1) = \dots x_3x_2, \quad \text{for all } \dots x_3x_2x_1 \in X^{-\infty},$$

preserves the asymptotic equivalence relation, a dynamical system $(\mathcal{J}_G, s)$ is defined and it is called the *limit dynamical system* of G. One can show that two sequences $\ldots x_2x_1$ and $\ldots y_2y_1$ are asymptotically equivalent if and only if there exists an oriented left-infinite path $\ldots e_2e_1$ in the Moore diagram of the nucleus $\mathcal{N}$ of G such that the arrow e_n is labeled by (x_n, y_n), for each $n \geqslant 1$. The limit space $\mathcal{J}_G$ is compact, metrizable and has finite topological dimension. If G is finitely generated and spherically transitive, then $\mathcal{J}_G$ is connected; moreover, if G is fractal, then $\mathcal{J}_G$ is locally connected and path connected [26, Theorem 3.6.3].

The sequence of finite Schreier graphs $\{\Gamma_n\}_{n\geqslant 1}$ of a contracting group constitutes an approximation of the limit space $\mathcal{J}_G$, via the notion of Gromov hyperbolic space, as formalized in the following theorem (see, for instance, [25]).

Theorem 3.2 *Let G be a contracting self-similar group generated by a finite set S. Consider the self-similarity graph Γ, with vertex set X^* and where two vertices are connected by an edge if*

- *they are of the form v and xv, with $v \in X^*$ and $x \in X$ (vertical edges);*
- *they are of the form v and $s(v)$, with $v \in X^*$ and $s \in S$ (horizontal edges).*

Then Γ is Gromov hyperbolic and its hyperbolic boundary is homeomorphic to $\mathcal{J}_G$.

Observe that horizontal edges constitute a graph which is isomorphic to the disjoint union of the finite Schreier graphs $\{\Gamma_n\}_{n\geqslant 1}$. As an example, we draw in Fig. 5 the (unlabeled) Schreier graphs Γ_n of the Basilica group, for $n = 1, \ldots, 6$, approximating the limit space of the group, which is homeomorphic to the Julia set of the complex polynomial $z^2 - 1$ (see Sect. 3.2), which is drawn in Fig. 2.

The limit space has a "*cellular structure*", as formalized by the definition of tiles.

Let $\pi : X^{-\infty} \to \mathcal{J}_G$ be the projection defining the limit space. Then, for each $v \in X^*$, the *tile* $\mathcal{T}_v$ is defined as

$$\mathcal{T}_v = \pi\left(X^{-\infty}v\right).$$

If $v \in X^n$, we say that $\mathcal{T}_v$ is a tile of level n of $\mathcal{J}_G$. Each tile is a compact subset of the limit space $\mathcal{J}_G$. One has:

$$\mathcal{T}_\emptyset = \mathcal{J}_G, \qquad \mathcal{T}_v = \bigcup_{x\in X} \mathcal{T}_{xv}.$$

Moreover,

$$\mathcal{J}_G = \bigcup_{|v|=n} \mathcal{T}_v \quad \text{for each } n \geqslant 1, \qquad s(\mathcal{T}_{vx}) = \mathcal{T}_v$$

and $\bigcap_{n\geqslant 1} \mathcal{T}_{x_n\ldots x_1}$ is given by only one point, for each $\ldots x_2x_1 \in X^{-\infty}$.

We say that a contracting self-similar group G satisfies the open set condition if for every $g \in G$ there exists $v \in X^*$ such that $g|_v = id$. If this condition is verified, then one has:

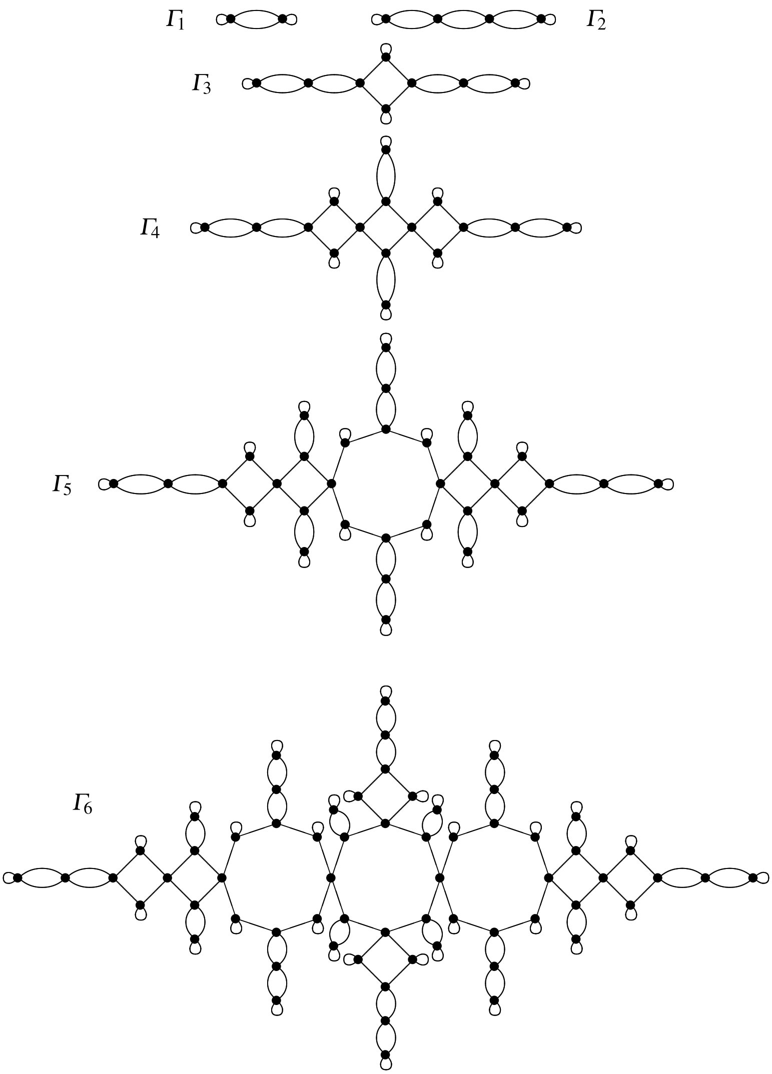


Fig. 5 Schreier graphs Γ_n of the Basilica group, for $n = 1, \ldots, 6$

$$\partial \mathscr{T}_v = \mathscr{T}_v \bigcap_{|u|=|v|, u \neq v} \mathscr{T}_u.$$

In [5], it is shown that a contracting self-similar group is bounded if and only if it satisfies the open set condition and the boundaries of the tiles $\mathscr{T}_v$ are finite, for all $v \in X^*$. Finally, one has:

$$\mathscr{T}_u \cap \mathscr{T}_v \neq \emptyset \quad \Leftrightarrow \quad \exists\, g \in \mathscr{N} \text{ s.t.} \quad g(u) = v.$$

In other words, the adjacency of the tiles of level n of $\mathscr{J}_G$ reflects the adjacency of the vertices of the Schreier graph Γ_n of the group G, with respect to the generating set $\mathscr{N}$ of G.

The *critical set* of G is defined as

$$\mathscr{C} = \pi^{-1}\left(\bigcup_{x \neq y} (\mathscr{T}_x \cap \mathscr{T}_y)\right) \subset X^{-\infty}.$$

The *post-critical set* of G is given by

$$\mathscr{P} = \bigcup_{n \geqslant 1} s^n(\mathscr{C}) \subset X^{-\infty}.$$

One can show that $\mathscr{P}$ contains exactly all the left-infinite words $\ldots x_2 x_1 \in X^{-\infty}$ such that there exists a path $\ldots e_2 e_1$ in the Moore diagram of the automaton defining G, which ends in a non-trivial state and is labeled by $\ldots x_2 x_1 | *$ or $* | \ldots x_2 x_1$.

We say that $\mathscr{J}_G$ is *post-critically finite* if $\mathscr{P}$ is finite. One can show that this is equivalent to say that two distinct tiles of the same level have finite intersection, i.e, $\mathscr{J}_G$ is *finitely ramified*. Moreover, it is known that this is the case if and only if the group G is generated by a bounded automaton [3]. The set $\pi(\mathscr{P})$ is called the *boundary* of $\mathscr{J}_G$.

3.2 Iterated Monodromy Groups

An important class of self-similar groups is given by Iterated Monodromy groups. We recall here some basic definitions.

A *partial self-covering* is a covering map $f : N \to M$, i.e., a map f such that every point $x \in M$ has a neighborhood U_x such that $f^{-1}(U_x)$ is the disjoint union of subsets U_i such that $f : U_i \to U$ is a homeomorphism. The cardinality $|f^{-1}(x)|$ is the *degree* of the covering. Here, we suppose that M is a path connected and locally path connected topological space and $N \subset M$. Given such a covering f and a basepoint $t \in M$, one can consider the tree of preimages of t

$$T_t = \bigcup_{n \geqslant 0} f^{-n}(t),$$

i.e., the rooted tree whose vertex set is given by the formal disjoint union of the iterated inverse images of t under the action of f, where the vertex $v \in f^{-n}(t)$ is connected by an edge to the vertex $f(v) \in f^{-(n-1)}(t)$. The fundamental group $\pi_1(M, t)$ acts on T_t as follows: given a loop $\gamma \in \pi_1(M, t)$, a point $v \in f^{-n}(t)$ is mapped by γ to the endpoint of the unique lift of γ by f^n starting at v. This action is called *iterated monodromy action*.

The Iterated Monodromy group $IMG(f)$ of the partial self-covering f is defined as the quotient of $\pi_1(M, t)$ modulo the kernel of the iterated monodromy action:

this is a self-similar automorphisms group of the rooted tree whose degree equals the degree of the covering, and it does not depend on the choice of the basepoint.

In particular, if M is a complete Riemannian manifold and f is an expanding self-covering, then $IMG(f)$ is a contracting group.

As a fundamental example, consider a rational function $f(z) = \frac{p(z)}{q(z)} \in \mathbb{C}(z)$. It defines a branched self-covering of $\widehat{\mathbb{C}}$. Hence, considering

$$\mathscr{C}_f = \{z : f'(z) = 0\}, \qquad \mathscr{P}_f = \bigcup_{n \geqslant 0} f^n(\mathscr{C}_f),$$

i.e., the *critical* and the *post-critical* set of f, respectively, one has that

$$f : \widehat{\mathbb{C}} \setminus f^{-1}(\overline{\mathscr{P}_f}) \longrightarrow \widehat{\mathbb{C}} \setminus \overline{\mathscr{P}_f}$$

is a partial self-covering of $M = \widehat{\mathbb{C}} \setminus \overline{\mathscr{P}_f}$ of degree $d = \max\{\deg p, \deg q\}$. The map f is said *post-critically finite* if $|\mathscr{P}_f| < \infty$. If this is the case, the covering is expanding [27], so that $IMG(f)$ is a contracting self-similar group, and its limit space $\mathscr{J}_G$ is homeomorphic to the *Julia set* $J(f)$ of the function f, i.e., the set of the accumulation points of the backward orbit $\bigcup_{n \geqslant 0} f^{-n}(t)$. More precisely, there exists a homeomorphism φ conjugating the dynamical systems $(\mathscr{J}_G, s)$ and $(J(f), f)$, such that the following diagram commutes.

$$\begin{array}{ccc} \mathscr{J}_G & \xrightarrow{\ s\ } & \mathscr{J}_G \\ {\scriptstyle\varphi}\downarrow & & \downarrow{\scriptstyle\varphi} \\ J(f) & \xrightarrow{\ f\ } & J(f) \end{array}$$

In particular, it is known that, if f is a post-critically finite polynomial, then the limit space $\mathscr{J}_G$ is finitely ramified and the group $IMG(f)$ is amenable, since it is generated by a bounded automaton.

We have already given the example of the Basilica group, which is isomorphic to the iterated monodromy group $IMG(z^2 - 1)$ and which is generated by the automorphisms of the rooted binary tree given in (2). Another example is given by the Hanoi Towers group $H^{(3)}$, so called since it models the classical Hanoi Towers game on 3 pegs, and which is isomorphic to the group $IMG(z^2 - \frac{16}{27z})$. The group $H^{(3)}$ is an automorphisms group of the rooted tree of degree 3 and it is generated by the following automorphisms [15]:

$$a = (01)(id, id, a), \qquad b = (02)(id, b, id), \qquad c = (12)(c, id, id).$$

In Fig. 6, we draw its Schreier graphs Γ_n, for $n = 1, 2, 3$ (compare with the finite approximations of the Sierpiński gasket in Fig. 1). Its limit space is homeomorphic to the Sierpiński gasket (Fig. 7).

Other examples of limit spaces of iterated monodromy groups of rational functions are represented in Figs. 8, 9, 10 and 11.

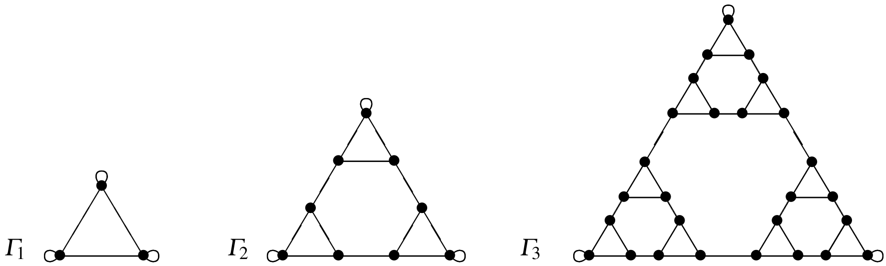


Fig. 6 Schreier graphs Γ_n of the Hanoi Towers group, for $n = 1, 2, 3$

Fig. 7 Sierpiński gasket

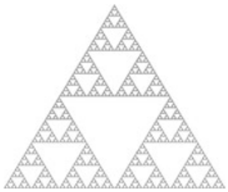


Fig. 8 $J(z^2 + i) =$ "Dendrite"

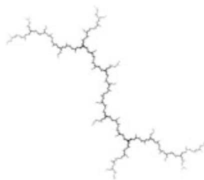


Fig. 9 $J(z^2 + \frac{2}{27z}) =$ "Apollonian gasket"

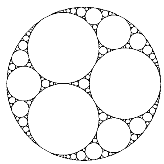


Fig. 10 $J(\frac{(z-1)^2}{(z+1)^2}) =$ "Pillow"

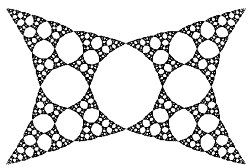
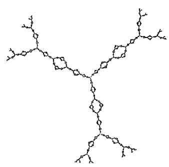


Fig. 11 $J(1 - \frac{1}{z^2})$



3.3 Post-Critically Finite Self-similar Fractals

Contracting self-similar groups are connected to fractal geometry through the notion of limit space. In this subsection, we recall some basic definitions concerning self-similar structures. Self-similar sets are in fact the simplest and basic structures in fractal theory. Moreover, post-critically finite self-similar structures are essentially the only class of fractals on which analysis has been developed. In the next section, we will briefly expose the analytic approach of Kigami [19] for constructing Dirichlet forms on post-critically finite self-similar structures, based on their approximation by an increasing sequence of finite sets $\{V_m\}_{m\geqslant 0}$.

Let K be a compact connected metrizable space and $X = \{0, 1, \ldots, q-1\}$ be a finite alphabet. Then K is a *self-similar structure* if there exist continuous injections $F_i : K \longrightarrow K$, with $i \in X$, and a continuous surjection $\pi : X^{\infty} \longrightarrow K$ such that the following diagram is commutative

$$\begin{array}{ccc} X^{\infty} & \xrightarrow{\sigma_i} & X^{\infty} \\ \pi\downarrow & & \downarrow\pi \\ K & \xrightarrow{F_i} & K \end{array}$$

where σ_i is defined by $\sigma_i(w_1w_2\ldots) = iw_1w_2\ldots$, for every $i \in X$ and $w_1w_2\ldots \in X^{\infty}$. Note that

$$K = \bigcup_{i\in X} F_i(K).$$

We set $F_{w_1\ldots w_m} = F_{w_1} \circ \cdots \circ F_{w_m}$. Then one has

$$\pi(w) = \bigcap_{m\geqslant 1} F_{w_1\ldots w_m}(K), \qquad \text{for each } w = w_1w_2\ldots \in X^{\infty}.$$

We call *m-cell* of K each set of the form $F_w(K)$, with $w \in X^m$. Let us define the shift map σ as $\sigma(w_1w_2\ldots) = w_2w_3\ldots$.

The *critical set* of K is defined by $\mathscr{C} = \pi^{-1}(\bigcup_{i\neq j}(F_i(K)\cap F_j(K))) \subset X^{\infty}$. The *post-critical set* is given by $\mathscr{P} = \bigcup_{n\geqslant 1}\sigma^n(\mathscr{C}) \subset X^{\infty}$.

Finally, K is said *post-critically finite* if $\mathscr{P}$ is finite. The set $V_0 = \pi(\mathscr{P})$ is called the *boundary* of K. By putting $V_m = \bigcup_{w\in X^m} F_w(V_0)$, one obtains an increasing (i.e. $V_m \subseteq V_{m+1}$) sequence of sets $\{V_m\}_{m\geqslant 0}$ such that $V_{m+1} = \bigcup_{i\in X} F_i(V_m)$. Finally, we set $V_* = \bigcup_{m\geqslant 0} V_m$. Hence, one has $\overline{V_*} = K$ if $V_0 \neq \emptyset$.

Remark 3.3 Note the analogy of the "cellular" structures of the limit space $\mathscr{J}_G$ and of the self-similar set K. In particular the tiles $\mathscr{T}_v$ in $\mathscr{J}_G$, with $v \in X^n$, play the same role as the cells $F_w(K)$ in K, with $w \in X^n$.

4 Construction of a Laplacian on Post-Critically Finite Self-similar Fractals

In this section, we present some basic elements of the theory of Dirichlet forms (or equivalently Laplacians) on a finite set, then we discuss limits of discrete Laplacians on an increasing sequence of finite sets, giving an approximation of a post-critically finite self-similar structure.

4.1 Dirichlet Forms and Laplacians on Finite Sets

Let V be a finite set and put $\ell(V) = \{f : V \to \mathbb{R}\}$. The set $\ell(V)$ can be endowed with the inner product defined by

$$(u, v) = \sum_{p \in V} u(p)v(p),$$

for all $u, v \in \ell(V)$.

Definition 4.1 A *Dirichlet form* on $\ell(V)$ is a symmetric bilinear form $\mathscr{E}$ on $\ell(V)$ satisfying the following properties:

(i) $\mathscr{E}(u, u) \geqslant 0$, for each $u \in \ell(V)$;
(ii) $\mathscr{E}(u, u) = 0$ if and only if u is constant on V;
(iii) (Markov property) $\mathscr{E}(u, u) \geqslant \mathscr{E}(\bar{u}, \bar{u})$, for every $u \in \ell(V)$, where

$$\bar{u}(p) = \begin{cases} 1 & \text{if } u(p) \geqslant 1, \\ u(p) & \text{if } 0 < u(p) < 1, \\ 0 & \textit{if } u(p) \leqslant 0. \end{cases}$$

For $U \subseteq V$, the characteristic function of U is defined as

$$\chi_U^V(p) = \begin{cases} 1 & \text{if } p \in U, \\ 0 & \text{otherwise.} \end{cases}$$

If $U = \{q\} \subset V$ is a singleton, we put $\chi_{\{q\}}^V = \chi_q$.

Now let $H : \ell(V) \to \ell(V)$ be a linear map and put $H_{pq} = (H\chi_q)(p)$, for all $p, q \in V$, so that one has, by linear algebra,

$$(Hf)(p) = \sum_{q \in V} H_{pq} f(q),$$

for every $f \in \ell(V)$.

Definition 4.2 A *Laplacian* on V is a symmetric linear operator $H : \ell(V) \to \ell(V)$ satisfying the following properties:

(i) H is non-positive definite;
(ii) $Hu = 0$ if and only if u is constant on V;
(iii) $H_{pq} \geqslant 0$, for all $p, q \in V$, $p \neq q$.

There exists a natural bijective correspondence between Dirichlet forms on $\ell(V)$ and Laplacians on V. More precisely, given a Laplacian H on V, a Dirichlet form $\mathscr{E}_H$ on $\ell(V)$ can be defined as

$$\mathscr{E}_H(u, v) = -(u, Hv),$$

for all $u, v \in \ell(V)$, and it is not difficult to check that this correspondence is bijective.

Now let H be a Laplacian on V and let $U \subset V$. Define $T_U : \ell(U) \to \ell(U)$, $J_U : \ell(U) \to \ell(V \setminus U)$ and $X_U : \ell(V \setminus U) \to \ell(V \setminus U)$ by

$$H = \begin{pmatrix} T_U & J_U^t \\ J_U & X_U \end{pmatrix}.$$

Theorem 4.3 [19, Theorem 2.1.6] *For any $u \in \ell(U)$, define $h(u) \in \ell(V)$ as*

$$h(u)|_U = u \quad \textit{and} \quad h(u)|_{V \setminus U} = -X_U^{-1} J_U u.$$

Then $h(u)$ is the unique element that attains $\min_{v \in \ell(V), v|_U = u} \mathscr{E}_H(v, v)$. Also define $P_{V,U}(H) = T_U - J_U^t X_U^{-1} J_U$. Then $P_{V,U}(H)$ is a Laplacian on U and

$$\mathscr{E}_{P_{V,U}(H)}(u, u) = \mathscr{E}_H\big(h(u), h(u)\big) = \min_{v \in \ell(V), v|_U = u} \mathscr{E}_H(v, v).$$

The linear operator $P_{V,U}(H)$ represents the restriction of H to U. Moreover, $h(u)$ is the unique solution of the problem

$$\begin{cases} (Hv)|_{V \setminus U} = 0, \\ v|_U = u \end{cases}$$

and is called the *harmonic function* with boundary value $u \in \ell(U)$.

Definition 4.4 Put

$$R_H(p, q) = \big(\min\big\{\mathscr{E}_H(u, u) : u \in \ell(V), u(p) = 1, u(q) = 0\big\}\big)^{-1}, \quad \text{for } p \neq q$$

and $R_H(p, p) = 0$. $R_H(p, q)$ is called the *effective resistance* between p and q.

We write

$$(V_1, H_1) \leqslant (V_2, H_2)$$

if $V_1 \subseteq V_2$, H_1 (resp. H_2) is a Laplacian on V_1 (resp. V_2) and $P_{V_2, V_1}(H_2) = H_1$.

It is possible to show that $(V_1, H_1) \leqslant (V_2, H_2)$ if and only if $R_{H_1}(p,q) = R_{H_2}(p,q)$, for all $p, q \in V_1$. Moreover, the following theorem holds.

Theorem 4.5 [19, Theorem 2.1.14] *If H is a Laplacian on a finite set V, then R_H is a metric on V.*

R_H is called the *effective resistance metric* on V associated with H. The following is an equivalent expression of $R_H(p,q)$, with $p, q \in V$:

$$R_H(p,q) = \max\left\{ \frac{|u(p) - u(q)|^2}{\mathscr{E}_H(u,u)} : u \in \ell(V), \mathscr{E}_H(u,u) \neq 0 \right\}. \tag{3}$$

It follows from (3) that, for every $p, q \in V$ and $u \in \ell(V)$, one has

$$\left|u(p) - u(q)\right|^2 \leqslant R_H(p,q)\mathscr{E}_H(u,u).$$

Observe that $\sqrt{R_H(\cdot,\cdot)}$ is also a metric on V.

Definition 4.6 The sequence $\{(V_m, H_m)\}_{m\geqslant 0}$ is called a *compatible sequence* if H_m is a Laplacian on V_m and $(V_m, H_m) \leqslant (V_{m+1}, H_{m+1})$, for each $m \geqslant 0$.

If $\mathscr{S} = \{(V_m, H_m)\}_{m\geqslant 0}$ is such a sequence, we put $V_* = \bigcup_{m\geqslant 0} V_m$ and

$$\mathscr{F}(\mathscr{S}) = \left\{ u : u \in \ell(V_*), \lim_{m\to\infty} \mathscr{E}_{H_m}(u|_{V_m}, u|_{V_m}) < +\infty \right\}$$

$$\mathscr{E}_{\mathscr{S}}(u,v) = \lim_{m\to\infty} \mathscr{E}_{H_m}(u|_{V_m}, v|_{V_m}),$$

for all $u, v \in \mathscr{F}(\mathscr{S})$. Then, for $p, q \in V_*$, the effective resistance $R_{\mathscr{S}}$ can be defined by putting

$$R_{\mathscr{S}}(p,q) = R_{H_m}(p,q),$$

with m such that $p, q \in V_m$. The resistance $R_{\mathscr{S}}$ is well-defined since the sequence $\{(V_m, H_m)\}_{m\geqslant 0}$ is compatible. Roughly speaking, the word "compatible" means that the Dirichlet forms associated with the Laplacians appearing in the sequence induce the same effective resistance on the increasing union of finite sets. One can easily show that both $R_{\mathscr{S}}$ and $R_{\mathscr{S}}^{1/2}$ are metrics on V_*.

Similarly as in the finite case, there exists a linear map $h_m : \ell(V_m) \to \mathscr{F}(\mathscr{S})$ satisfying $h_m(u)|_{V_m} = u$, for every $u \in \ell(V_m)$, and

$$\mathscr{E}_{H_m}(u,u) = \mathscr{E}_{\mathscr{S}}\big(h_m(u), h_m(u)\big) = \min_{v\in\mathscr{F}(\mathscr{S}), v|_{V_m}=u} \mathscr{E}_{\mathscr{S}}(v,v).$$

Moreover if $v \in \mathscr{F}(\mathscr{S})$, with $v|_{V_m} = u$, attains the minimum above, then $v = h_m(u)$. Finally, $h_m(u)$ is the unique solution of the problem

$$\begin{cases} (H_n v|_{V_n})|_{V_n\setminus V_m} = 0 & \text{for all } n > m, \\ v|_{V_m} = u, \end{cases}$$

where $v \in \ell(V_*)$. Then, $h_m(u)$ is called the *harmonic function* with boundary value $u \in \ell(V_m)$. It is easy to prove that the following equalities hold:

$$R_{\mathscr{S}}(p,q) = \left(\min\left\{\mathscr{E}_{\mathscr{S}}(u,u) : u \in \mathscr{F}(\mathscr{S}), u(p)=1, u(q)=0\right\}\right)^{-1}$$
$$= \max\left\{\frac{|u(p)-u(q)|^2}{\mathscr{E}_{\mathscr{S}}(u,u)} : u \in \mathscr{F}(\mathscr{S}), \mathscr{E}_{\mathscr{S}}(u,u) > 0\right\}.$$

Definition 4.7 Let X be a set. $(\mathscr{E}, \mathscr{F})$ is a *resistance form* on X if the following conditions are satisfied.

(i) $\mathscr{F}$ is a linear subspace of $\ell(X)$ containing constants and $\mathscr{E}$ is a non-negative symmetric quadratic form on $\mathscr{F}$. Moreover, $\mathscr{E}(u,u) = 0$ if and only if u is constant on X.
(ii) Define $u \sim v$, with $u, v \in \mathscr{F}$, if $u - v$ is constant on X. Then $(\mathscr{F}/\sim, \mathscr{E})$ is a Hilbert space.
(iii) For any finite subset $V \subset X$ and $v \in \ell(V)$, there exists $u \in \mathscr{F}$ such that $u|_V = v$.
(iv) For any $p, q \in X$,

$$\sup\left\{\frac{|u(p)-u(q)|^2}{\mathscr{E}(u,u)} : u \in \mathscr{F}, \mathscr{E}(u,u) > 0\right\} < \infty.$$

(v) (*Markov property*) If $u \in \mathscr{F}$, then $\bar{u} \in \mathscr{F}$ and $\mathscr{E}(u,u) \geqslant \mathscr{E}(\bar{u},\bar{u})$, where $\bar{u}$ is as in Definition 4.1.

One can show that $(\mathscr{E}_{\mathscr{S}}, \mathscr{F}(\mathscr{S}))$ is a *resistance form* on V_* and $R_{\mathscr{S}}$ is called the associated *resistance metric* on V_*. Since the space V_* is just a countable set, we have to consider the completion of V_* with respect to $R_{\mathscr{S}}$. More precisely, the following result holds.

Theorem 4.8 [19, Theorem 2.3.10] *Let $(\mathscr{E}, \mathscr{F})$ be a resistance form on X and let R be the associated resistance metric on X. If Ω is the completion of X with respect to R, then $(\mathscr{E}, \mathscr{F})$ is a resistance form on Ω. Moreover, the resistance metric associated with $(\mathscr{E}, \mathscr{F})$ on Ω is the natural extension of the resistance metric R associated with $(\mathscr{E}, \mathscr{F})$ on X.*

4.2 Dirichlet Forms on Locally Compact Metric Spaces

Let $\mathscr{H}$ be a separable Hilbert space with inner product $(\cdot,\cdot)$.

Definition 4.9 A non-negative quadratic form $\mathscr{E}$ on $\mathscr{H}$ with dense domain $\mathrm{Dom}(\mathscr{E})$ is said a *closed form* if it satisfies one of the following equivalent conditions:

(i) $\mathrm{Dom}(\mathscr{E}) = \mathrm{Dom}(H^{1/2})$ and $\mathscr{E}(u,v) = (H^{1/2}u, H^{1/2}v)$ for all $u, v \in \mathrm{Dom}(H^{1/2})$, for some non-negative self-adjoint operator H on $\mathscr{H}$;

(ii) Define $\mathscr{E}_*(u, v) = \mathscr{E}(u, v) + (u, v)$, for every $u, v \in \mathrm{Dom}(\mathscr{E})$. Then $(\mathrm{Dom}(\mathscr{E}), \mathscr{E}_*)$ is a Hilbert space.

Now let X be a locally compact metric space and let μ be a σ-finite Borel measure on X satisfying $\mu(A) < \infty$ for any compact A and $\mu(O) > 0$ for every nonempty open set O. Put

$$C_0(X) = \big\{f : X \to \mathbb{R},\ f \text{ continuous and } supp(f) \text{ is compact}\big\}.$$

Definition 4.10

(i) Let $\mathscr{E}$ be a closed form on $L^2(X, \mu)$ and let $\mathscr{F} = \mathrm{Dom}(\mathscr{E})$. Then $\mathscr{E}$ is a *Dirichlet form* on $L^2(X, \mu)$ if $\bar{u} \in \mathscr{F}$ and $\mathscr{E}(u, u) \geqslant \mathscr{E}(\bar{u}, \bar{u})$ for any $u \in \mathscr{F}$, where $\bar{u}$ is as in Definition 4.1 (Markov property).

(ii) A Dirichlet form is called *regular* if there exists $\mathscr{C} \subseteq \mathscr{F} \cap C_0(X)$ which is dense in $\mathscr{F}$ with respect to the $\mathscr{E}_*$-norm and in $C_0(X)$ with respect to the uniform norm.

(iii) A Dirichlet form is called *local* if $\mathscr{E}(u, v) = 0$ whenever $u, v \in \mathscr{F}$ and $supp(u)$ and $supp(v)$ are compact and disjoint.

Remark 4.11 There exists a one-to-one correspondence between Dirichlet forms on $L^2(K, \mu)$ and strongly continuous semigroups $\{T_t\}_{t>0}$ on $L^2(K, \mu)$ with the Markov property, i.e., if $u \in L^2(K, \mu)$ and $0 \leqslant u(x) \leqslant 1$ for μ-a.e. $x \in K$, then $0 \leqslant (T_t u)(x) \leqslant 1$ for μ-a.e. $x \in K$ and any $t > 0$.

4.3 Harmonic Structures

Let K and V_m be as defined in Sect. 3.3. We aim at constructing a self-similar compatible sequence $\{(V_m, H_m)\}_{m\geqslant 0}$ of Laplacians on the sets V_m and taking its limit in order to get a Dirichlet form on K.

Let D be a Laplacian on V_0 and $\mathbf{r} = (r_0, \ldots, r_{q-1})$, with $r_i > 0$. Moreover, let $\mathscr{E}_D$ be the Dirichlet form on V_0 associated with D. For each $m \geqslant 1$, we define the Dirichlet form on $\ell(V_m)$ as

$$\mathscr{E}^{(m)}(u, v) = \sum_{w\in X^m} \frac{1}{r_w} \mathscr{E}_D(u \circ F_w, v \circ F_w), \tag{4}$$

for every $u, v \in \ell(V_m)$, with $r_w = \prod_{i=1}^{m} r_{w_i}$, for $w = w_1 \ldots w_m \in X^m$. One can easily verify that one has

$$\mathscr{E}^{(m+1)}(u, v) = \sum_{i=0}^{q-1} \frac{1}{r_i} \mathscr{E}^{(m)}(u \circ F_i, v \circ F_i),$$

for all $u, v \in \ell(V_{m+1})$. Let H_m be the Laplacian on V_m associated with $\mathscr{E}^{(m)}$.

Definition 4.12 $(D, \mathbf{r})$ is called a harmonic structure if $\{(V_m, H_m)\}_{m \geqslant 1}$ is a compatible sequence. Also, a harmonic structure $(D, \mathbf{r})$ is said regular if $0 < r_i < 1$, for each $i = 0, \dots, q-1$.

The crucial problem is to establish if there exists a harmonic structure on a post-critically finite self-similar fractal. The following proposition simplifies the problem, reducing the question to verify the compatibility condition between V_0 and V_1.

Proposition 4.13 [19, Proposition 3.1.3] $(D, \mathbf{r})$ *is a harmonic structure if and only if* $(V_0, D) \leqslant (V_1, H_1)$.

Given $\mathbf{r} = (r_0, \dots, r_{q-1})$, define the *renormalization operator* $\mathscr{R}_{\mathbf{r}}$ as

$$\mathscr{R}_{\mathbf{r}}(D) = P_{V_1, V_0}(H_1),$$

where H_1 is the Laplacian on V_1 associated with the form

$$\mathscr{E}^{(1)}(u, v) = \sum_{i=0}^{q-1} \frac{1}{r_i} \mathscr{E}_D(u \circ F_i, v \circ F_i), \quad \text{for } u, v \in \ell(V_1),$$

defined in (4). We already know that $P_{V_1, V_0}(H_1)$ is a Laplacian on V_0 (see Theorem 4.3). Hence, by Proposition 4.13, $(D, \mathbf{r})$ is a harmonic structure if and only if D is a fixed point of $\mathscr{R}_{\mathbf{r}}$. Now let $a, \lambda \in \mathbb{R}$. Then it is easy to verify that

$$\mathscr{R}_{\lambda \mathbf{r}}(aD) = \frac{a}{\lambda} \mathscr{R}_{\mathbf{r}}(D).$$

Therefore, if D is an eigenvector for $\mathscr{R}_{\mathbf{r}}$ of eigenvalue λ, then D is a fixed point for $\mathscr{R}_{\lambda \mathbf{r}}$. As a consequence, the problem of existence of harmonic structures on K is reduced to a fixed point problem, or an eigenvalue problem, for the non-linear operator $\mathscr{R}_{\mathbf{r}}$. Although the investigation of this problem is very hard and it was not solved in the general case, the existence of a harmonic structure was established in the case of *nested fractals* by Lindstrøm [21] and of post-critically finite self-similar fractals with three boundary points [28] and in some other situations (see, for instance, [27, Theorem 6.1]).

Example 4.14 The interval $I = [0, 1]$ is a post-critically finite self-similar structure with respect to the map $F_0(x) = \frac{x}{2}$ and $F_1(x) = \frac{x}{2} + \frac{1}{2}$. One has $V_m = \{2^{-m} i\}_{i=0,1,\dots,2^m}$, for each $m \geqslant 0$. Consider the Laplacian D on V_0 associated with the matrix $\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$ and the corresponding Dirichlet form $\mathscr{E}_D$. Then it is defined on $\ell(V_1)$ the form

$$\mathscr{E}^{(1)}(u, v) = \frac{1}{r_0} \mathscr{E}_D(u \circ F_0, v \circ F_0) + \frac{1}{r_1} \mathscr{E}_D(u \circ F_1, v \circ F_1),$$

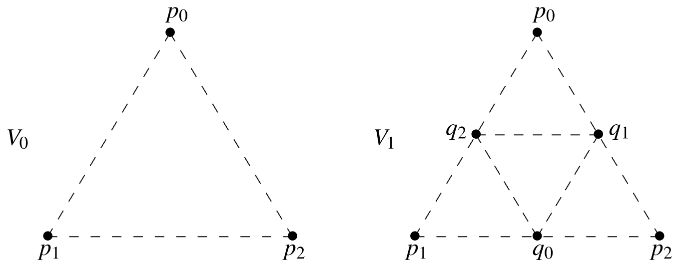


Fig. 12 The sets V_0 and V_1 approximating the Sierpiński gasket

whose associated Laplacian is

$$H_1 = \begin{pmatrix} -\frac{1}{r_0} & 0 & \frac{1}{r_0} \\ 0 & -\frac{1}{r_1} & \frac{1}{r_1} \\ \frac{1}{r_0} & \frac{1}{r_1} & -\frac{1}{r_0} - \frac{1}{r_1} \end{pmatrix}.$$

Here, the rows and columns of the matrix are indexed by the elements $0, 1, \frac{1}{2}$ of V_1. Then, the compatibility condition $P_{V_1,V_0}(H_1) = D$ gives

$$\begin{pmatrix} -\frac{1}{r_0} + \frac{r_1}{r_0(r_0+r_1)} & \frac{1}{r_0+r_1} \\ \frac{1}{r_0+r_1} & -\frac{1}{r_1} + \frac{r_0}{r_1(r_0+r_1)} \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

One easily gets that $(D, \mathbf{r})$ is a harmonic structure if and only if $r_0 + r_1 = 1$, and $0 < r_i < 1$ for $i = 0, 1$.

Example 4.15 The Sierpiński gasket (see Fig. 7 in Sect. 3.2) is a post-critically finite self-similar structure with respect to the maps, defined on $\mathbb{C}$, given by

$$F_i(z) = \frac{z - p_i}{2} + p_i,$$

for $i = 0, 1, 2$, where p_0, p_1, p_2 are the vertices of an equilateral triangle. In this case one has $X = \{0, 1, 2\}$ and $p_i = \pi(i^\infty)$. If we put

$$q_0 = \pi\left(12^\infty\right) = \pi\left(21^\infty\right), \qquad q_1 = \pi\left(02^\infty\right) = \pi\left(20^\infty\right),$$
$$q_2 = \pi\left(01^\infty\right) = \pi\left(10^\infty\right),$$

then one has $V_0 = \{p_0, p_1, p_2\}$ and $V_1 = \{p_0, p_1, p_2, q_0, q_1, q_2\}$ (see Fig. 12). Consider the Laplacian on V_0 associated with the matrix $\begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}$. Then the structure $(D, \mathbf{r})$ obtained by choosing $r_0 = r_1 = r_2 = \frac{3}{5}$ is harmonic and it is called *standard harmonic structure* on the Sierpiński gasket.

The fundamental idea is to obtain a Laplacian (or, equivalently, a Dirichlet form) on K, by constructing a "self-similar" compatible sequence of Laplacians on the sets $\{V_m\}_{m\geqslant 0}$. Firstly, observe that the resistance form $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S}))$ on V_* obtained as limit of a harmonic structure $(D, \mathbf{r})$ is self-similar, i.e.,

$$\mathcal{E}_{\mathcal{S}}(u, v) = \sum_{i=0}^{q-1} \frac{1}{r_i} \mathcal{E}_{\mathcal{S}}(u \circ F_i, v \circ F_i),$$

as follows from the definition of $\mathcal{E}^{(m)}$ in (4).

If the completion Ω of V_* with respect to $\mathcal{R}_{\mathcal{S}}$ coincides with K, it is true [19, Chap. 2] that $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S}))$ is a regular local Dirichlet form on $L^2(K, \mu)$, for every Borel regular probability measure μ on K. Then, it follows from the classical potential theory that there exists a unique self-adjoint, non-positive operator H on $L^2(K, \mu)$, with domain $\mathrm{Dom}(H) \subset \mathcal{F}(\mathcal{S})$ dense in $L^2(K, \mu)$, satisfying

$$\mathcal{E}_{\mathcal{S}}(u, v) = -\int_K (Hu)v \, d\mu \quad \text{for every } u \in \mathrm{Dom}(H) \text{ and } v \in \mathcal{F}(\mathcal{S}).$$

This operator is a Laplacian on the fractal K. If Ω does not coincide with K, it turns out that, however, there exists a continuous injective map from Ω to K, whose restriction to V_* is the identity. The following theorem holds.

Theorem 4.16 [19, Theorem 3.3.4] *The following conditions are equivalent*:

(i) $\Omega = K$;
(ii) $(\Omega, R_{\mathcal{S}})$ *is compact*;
(iii) $(\Omega, R_{\mathcal{S}})$ *is bounded*;
(iv) $(D, \mathbf{r})$ *is regular.*

If this is the case, $R_{\mathcal{S}}$ *is a metric on* K *which is compatible with the original topology of* K.

Remark 4.17 If $(D, \mathbf{r})$ is not regular, then Ω is a proper subset of K. However, $\mathcal{F}(\mathcal{S})$ can be embedded in $L^2(K, \mu)$ for a certain measure μ such that $(\mathcal{E}_{\mathcal{S}}, \mathcal{F}(\mathcal{S}))$ is a local regular Dirichlet form on $L^2(K, \mu)$ (see [19] for details).

5 Examples

We explicitly describe here some examples of contracting self-similar group (the Adding Machine, the Basilica group and the group $IMG(z^2 + i)$), for which the approximation by finite Schreier graphs allows to construct a Laplacian on the corresponding limit space.

5.1 An Approach by Self-similar Random Walks

Let G be a contracting self-similar group and define a Markov operator on G as $M = \sum_{g\in G} \mu(g)g$, where μ is a probability measure on G. We assume that μ is symmetric, i.e., $\mu(g) = \mu(g^{-1})$ and non-degenerate, i.e., its support generates G. The operator M defines a random walk on G, where $\mu(g)$ represents the probability of transition from h to hg, for every $h \in G$. The associated Laplacian is $\Delta = M - 1$, where 1 denotes the identity operator. If μ is finitely supported by a set S, then Δ is the discrete Laplacian on the Cayley graph $Cay(G, S)$ of G with respect to S, i.e., the graph whose vertex set is G and where (g, h) is an edge if there exists $s \in S$ such that $gs = h$. Moreover, $\psi_n(\Delta)$ (see Sect. 2.4) is the matrix associated with the discrete Laplacian on the Schreier graph Γ_n of G.

One can consider the Dirichlet forms $\mathscr{E}_n$ on $\ell(X^n)$ given by the matrices $\Delta_n = \psi_n(1 - M) = -\psi_n(\Delta)$. Choose a letter $x_0 \in X$, and let $w = x_0^{-\infty} \in X^{-\infty}$ and $V_n \subset \mathscr{J}_G$ be the set of points of the limit space $\mathscr{J}_G$ represented by the sequences of the form wv, for $v \in X^n$. We consider $\mathscr{E}_n$ to be forms on V_n, identifying $v \in X^n$ with the corresponding point $wv \in V_n$. Then, the restriction of the form $\mathscr{E}_{n+1}$ on V_n, that we call the *trace* and denote by $Tr_{V_n}\mathscr{E}_{n+1}$, is given by the Schur complement

$$-\psi_n\big(\Delta_{x_0,x_0} - \Delta_{x_0,\overline{x_0}}(\Delta_{\overline{x_0},\overline{x_0}})^{-1}\Delta_{\overline{x_0},x_0}\big),$$

where Δ has the following block decomposition:

$$\Delta = \begin{pmatrix} \Delta_{x_0,x_0} & \Delta_{x_0,\overline{x_0}} \\ \Delta_{\overline{x_0},x_0} & \Delta_{\overline{x_0},\overline{x_0}} \end{pmatrix}.$$

Here we are assuming, without loss of generality, that x_0 is the first letter in the ordering of X. The Markov operator M is said *self-similar* if

$$\Delta_{x_0,x_0} - \Delta_{x_0,\overline{x_0}}(\Delta_{\overline{x_0},\overline{x_0}})^{-1}\Delta_{\overline{x_0},x_0} = \lambda\Delta,$$

for some $\lambda < 1$. In our setting, if M is a self-similar Markov operator, then the sequence of Dirichlet forms $\widetilde{\mathscr{E}_n}$ given by the matrices $\lambda^{-n}\psi_n(1 - M)$ is compatible (in the sense of Definition 4.6) and so can be used to construct a Laplacian on the limit space of the group.

5.2 The Circle $\mathbb{R}/\mathbb{Z}$ via the Adding Machine

Consider the automorphisms group G of the rooted binary tree generated by the automorphism a whose self-similar presentation is

$$a = (01)(id, a).$$

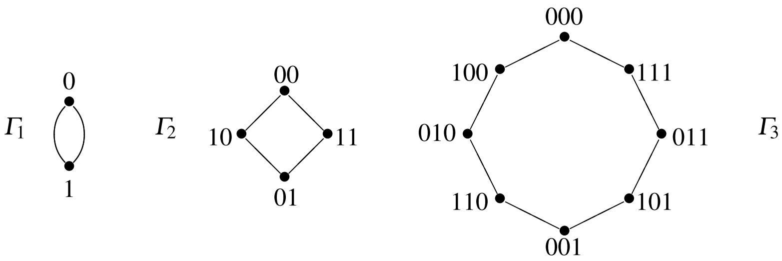


Fig. 13 Schreier graphs Γ_n of the Adding Machine, for $n = 1, 2, 3$

Observe that $G = \langle a \rangle \cong \mathbb{Z}$. Moreover, G can also be obtained as the iterated monodromy group $IMG(z^2)$. For each $x_1x_2 \ldots x_n \in \{0,1\}^n$, one has

$$a(x_1x_2\ldots x_n) = y_1y_2\ldots y_n \iff \sum_{i=1}^{n} x_i 2^{i-1} + 1 = \sum_{i=1}^{n} y_i 2^{i-1} (\text{mod } 2^n).$$

For this reason, the group G is called the "Adding Machine". Its Schreier graph Γ_n is a cycle of length 2^n (see Fig. 13), since $Stab_G(L_n) =< a^{2^n} >$ and so G acts on the nth level of the tree as the cyclic group $G/Stab_G(L_n) \cong \mathbb{Z}/2^n\mathbb{Z}$.

From the description of the generator of the group it is easy to deduce that the asymptotic equivalence relations between left-infinite words defining $\mathscr{J}_G$ are the following:

$$0^{-\infty}1u \sim 1^{-\infty}0u, \quad \text{for every } u \in \{0,1\}^* \quad \text{and} \quad 0^{-\infty} \sim 1^{-\infty}$$

and so $\ldots x_2x_1 \sim \ldots y_2y_1 \iff \sum_{n\geqslant 1} x_n 2^{-n} = \sum_{n\geqslant 1} y_n 2^{-n} (\text{mod } 1)$. Since the limit space is given by $\mathscr{J}_G = \{0,1\}^{-\infty}/\sim$, one gets $\mathscr{J}_G = \mathbb{R}/\mathbb{Z}$, which is the circle.

We use here the techniques developed in Sect. 5.1, in order to get a Dirichlet form on $\mathscr{J}_G$. Let $L_0 : \{0,1\}^n \to \{0,1\}^{n+1}$ such that $L_0(v) = 0v$ and let $V_n = \pi(0^{-\infty}X^n) \subset \mathscr{J}_G$. One has $V_n = \{\frac{k}{2^n} : k = 0, 1, \ldots, 2^n - 1\} \subset \mathbb{R}/\mathbb{Z}$. Then it is clear that $V_n \subset V_{n+1}$, where the embedding is induced by the map L_0.

The matrix recursion for the generator a of the group and its inverse gives

$$\psi(a) = \begin{pmatrix} 0 & id \\ a & 0 \end{pmatrix}, \qquad \psi(a^{-1}) = \begin{pmatrix} 0 & a^{-1} \\ id & 0 \end{pmatrix}.$$

Consider now the Dirichlet form $\mathscr{E}_n$ on $\ell(V_n)$ associated with the matrix

$$D_n = \psi_n\big(2id - (a + a^{-1})\big) = \psi_{n-1} \begin{pmatrix} 2id & -(id + a^{-1}) \\ -(id + a) & 2id \end{pmatrix},$$

corresponding to the discrete Laplacian on Γ_n. Using the Schur complement technique, one gets $Tr_{V_n}\mathscr{E}_{n+1} = \frac{1}{2}\mathscr{E}_n$ and so $\{\widetilde{\mathscr{E}_n} = 2^n\mathscr{E}_n\}_{n\geqslant 1}$ is a compatible sequence of Dirichlet forms. Therefore, one can pass to the limit on V_* by putting

$$\mathscr{E}(f, f) = \lim_{n\to\infty} \widetilde{\mathscr{E}_n}(f|_{V_n}, f|_{V_n}),$$

for each $f \in \ell(V_*)$. Then a Laplacian Δ_n can be associated with $\widetilde{\mathscr{E}}_n$ by putting:

$$\widetilde{\mathscr{E}}_n(f, f) = -\langle f, \Delta_n f \rangle.$$

One gets $\Delta_n = -2^{2n}\psi_n(2id - (a + a^{-1}))$. In particular, if $f \in \ell(V_n)$ and $t = \frac{k}{2^n} \in V_n$, we find:

$$(\Delta_n f)(t) = -\frac{2f(t) - (f(t + \frac{1}{2^n}) + f(t - \frac{1}{2^n}))}{1/2^{2n}} \to f''(t), \quad \text{as } n \to \infty,$$

obtaining the classical Laplacian on the circle $\mathbb{R}/\mathbb{Z}$.

Remark 5.1 Compare with the Laplacian on the unit interval [0, 1] obtained in [19, Example 3.7.2], which coincides with the standard Laplacian on $C^2([0, 1])$. (Note that [0, 1] is homeomorphic to the circle, when the boundary points 0 and 1 are identified.) One has in this case:

$$V_m = \left\{ \frac{i}{2^m} : i = 0, 1, \dots, 2^m \right\}, \qquad V_m \subset V_{m+1}, \quad \text{for each } m \geqslant 0.$$

More precisely, let $D = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$ be the Laplacian on $V_0 = \{0, 1\}$ defined in Example 4.14 and consider the harmonic structure $(D, \mathbf{r})$, obtained by choosing $r_0 = r_1 = \frac{1}{2}$. One has, for each $m \geqslant 1$,

$$(H_m u)(p) = 2^m \begin{cases} u(p + \frac{1}{2^m}) + u(p - \frac{1}{2^m}) - 2u(p) & p = \frac{i}{2^m} \neq 0, 1, \\ u(\frac{1}{2^m}) - u(0) & p = 0, \\ u(1 - \frac{1}{2^m}) - u(1) & p = 1. \end{cases}$$

As $m \to \infty$, one gets

$$(\Delta u)(x) = u''(x), \quad \text{for every } u \in C^2\big([0, 1]\big).$$

5.3 *The Basilica Fractal via the Group $IMG(z^2 - 1)$*

The so-called "Basilica group" is the automorphisms group of the rooted binary tree generated by the automorphisms whose self-similar presentation is

$$a = (b, id), \qquad b = (01)(a, id).$$

It can also be obtained as the iterated monodromy group $IMG(z^2 - 1)$ (see, e.g., [26]). It was introduced by R.I. Grigorchuk and A. Żuk in [16], where they show that it does not belong to the closure of the set of groups of subexponential growth under the operations of group extension and direct limit. L. Bartholdi and B. Virág further proved its amenability, making Basilica the first example of an amenable but not subexponentially amenable group [2]. The Schreier graphs Γ_n of the group, for $n = 1, \dots, 6$, have been drawn in Fig. 5 (see [9] for a detailed study of both finite and infinite Schreier graphs of the Basilica group).

From the description of the generators of the group it is easy to deduce that the asymptotic equivalence relation is generated by the following identifications:

$$(00)^{-\infty}1u \sim (10)^{-\infty}0u, \qquad (00)^{-\infty}1u \sim (01)^{-\infty}1u \quad \text{for every } u \in \{0,1\}^*,$$
$$(00)^{-\infty} \sim (01)^{-\infty}, \qquad (00)^{-\infty} \sim (10)^{-\infty}.$$

The limit space is homeomorphic to the Julia set of the complex polynomial $z^2 - 1$ (see Fig. 2).

Although the limit space $\mathcal{J}_G$ is not a post-critically finite self-similar set in the classical sense of Sect. 3.3, we can construct a Dirichlet form on it by using the technique of the self-similar random walk introduced in Sect. 5.1.

Let M be the operator associated with the matrix $\alpha(a+a^{-1}) + \beta(b+b^{-1})$, with $\alpha, \beta > 0$ and $\alpha + \beta = \frac{1}{2}$. Observe that one has:

$$\psi(a) = \begin{pmatrix} b & 0 \\ 0 & id \end{pmatrix}, \qquad \psi(b) = \begin{pmatrix} 0 & a \\ id & 0 \end{pmatrix},$$
$$\psi(a^{-1}) = \begin{pmatrix} b^{-1} & 0 \\ 0 & id \end{pmatrix}, \qquad \psi(b^{-1}) = \begin{pmatrix} 0 & id \\ a^{-1} & 0 \end{pmatrix}$$

since $a^{-1} = (b^{-1}, id)$ and $b^{-1} = (01)(id, a^{-1})$. As usual, the set X^n can be identified with the subset V_n of the limit space, given by $V_n = \pi(\{0^{-\infty}v : v \in X^n\}) \subset \mathcal{J}_G$.

Let $\mathcal{E}_n$ be the Dirichlet form on $\ell(X^n)$ (or, equivalently, on $\ell(V_n)$) given by:

$$D_n = \psi_n\big(id - \alpha\big(a + a^{-1}\big) - \beta\big(b + b^{-1}\big)\big) = \psi_{n-1}\begin{pmatrix} id - \alpha(b+b^{-1}) & -\beta(id+a) \\ -\beta(id + a^{-1}) & id - 2\alpha \end{pmatrix}$$
$$= \begin{pmatrix} D_{00} & D_{0\overline{0}} \\ D_{\overline{0}0} & D_{\overline{00}} \end{pmatrix}.$$

We want to find $\lambda < 1$, and $\alpha, \beta > 0$ satisfying $\alpha + \beta = \frac{1}{2}$, such that

$$D_{00} - D_{0\overline{0}} D_{\overline{00}}^{-1} D_{\overline{0}0} = \lambda D_n.$$

Using the Schur complement technique, we get

$$\lambda = \frac{1}{\sqrt{2}}, \qquad \alpha = \frac{\sqrt{2}-1}{2}, \qquad \beta = \frac{2-\sqrt{2}}{2}.$$

As a consequence, the form

$$\widetilde{\mathcal{E}}_n = 2^{n/2}\psi_n\left(id - \frac{\sqrt{2}-1}{2}\big(a + a^{-1}\big) - \frac{2-\sqrt{2}}{2}\big(b + b^{-1}\big)\right)$$

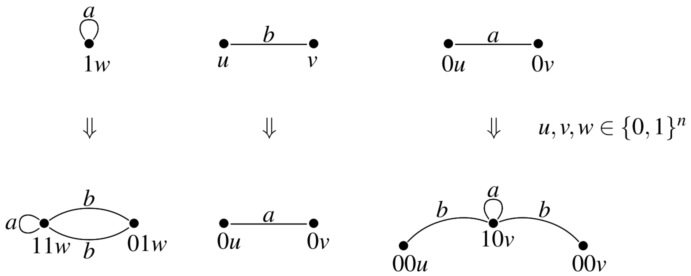


Fig. 14 Substitutional rules for the construction of Γ_{n+1} starting from Γ_n

is such that the sequence $\{\widetilde{\mathscr{E}}_n\}_{n\geqslant 1}$ is compatible. Therefore, we can pass to the limit by putting

$$\mathscr{E}(f, f) = \lim_{n\to\infty} \widetilde{\mathscr{E}}_n(f|_{V_n}, f|_{V_n}), \quad \text{for every } f \in \ell(V_*).$$

Remark 5.2 In [29], the Julia set of the complex polynomial $p(z) = z^2 - 1$ is endowed with a different finitely ramified cellular structure. Using the methods of Kigami for the construction of resistance forms and Dirichlet forms, the authors describe all possible Dirichlet forms on it, for which the topology given by the effective resistance coincides with the usual topology. Between them, there exists a unique (up to a constant) form that has a self-similar scaling under the action of $p(z)$ and it is found by using the substitutional rules [9] allowing to recursively construct the Schreier graphs of the group (Fig. 14). This last form coincides with the one that we have found by using matrices $\{D_n\}_{n\geqslant 1}$.

5.4 *The Dendrite Fractal via the Group IMG($z^2 + i$)*

The limit space of the iterated monodromy group of the complex polynomial $z^2 + i$ is homeomorphic to the Dendrite fractal (see Fig. 8). The group $IMG(z^2 + i)$ is the automorphisms group of the rooted binary tree whose generators have the following self-similar form:

$$a = (01)(id, id), \qquad b = (a, c), \qquad c = (b, id).$$

The Schreier graphs Γ_n, for $n = 1, \dots, 4$ are shown in Fig. 15, where loops are omitted. The sequence of the Schreier graphs $\{\Gamma_n\}_{n\geqslant 1}$ can be drawn in the complex plane in such a way that they converge, in the Hausdorff metric, to the Julia set of the polynomial $z^2 + i$.

This group is an example of a *strictly post-critically finite group* (strictly p.c.f. for short), i.e., it is generated by a bounded automaton and every element of its nucleus changes at most one letter in every word $v \in \{0, 1\}^*$. Therefore, we can apply the

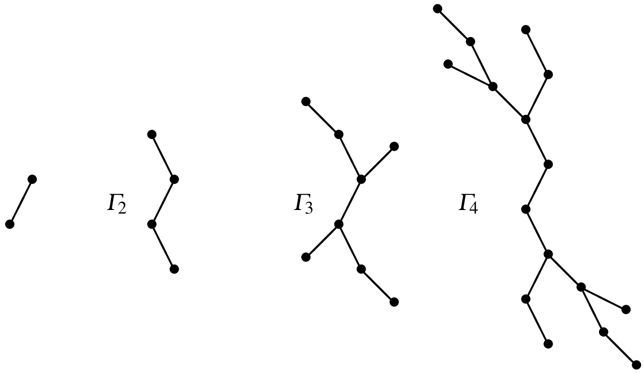


Fig. 15 Schreier graphs Γ_n of the group $IMG(z^2+i)$, for $n = 1, \ldots, 4$

standard techniques of analysis on finitely ramified fractals. More precisely, let G be a strictly p.c.f. finitely generated fractal group, generated by its nucleus. Let $\mathscr{P}$ be the post-critical set of G, defined in Sect. 3.1, i.e., the set of sequences $w = \ldots x_2x_1 \in X^{-\infty}$ such that there exists an oriented path $\ldots e_2e_1$ in the Moore diagram of the nucleus ending in a non-trivial element and such that the arrow e_n is labeled by (x_n, y_n), for some y_n. Then, if G is strictly p.c.f., it must be $y_n = x_n$. Denote by $g_w \in \mathbb{C}G$ the average of the elements of G corresponding to the ends of such paths. Take now the symmetric matrix $D(\vec{x}, y) = (D_{w_1,w_2}) \in M_{|\mathscr{P}|\times|\mathscr{P}|}(\mathbb{C}G)$, where $\vec{x} = (x_{\{w_1,w_2\}})$ is a real valued vector whose components are indexed by the 2-subsets of $\mathscr{P}$, and $y \in \mathbb{R}$, whose elements are defined as follows: $D_{w_1,w_2} = -x_{\{w_1,w_2\}}$ if $w_1 \neq w_2$, and $D_{w,w} = y(1 - g_w) + \sum_{w_2\neq w} x_{\{w,w_2\}}$.

Set $\mathscr{P}X^n = \{wv : w \in \mathscr{P}, v \in X^n\}$ and let $\ell(V_n)$ be the subspace of $\ell(\mathscr{P}X^n)$ consisting of functions which are constant on the asymptotic equivalence classes. It is possible to prove that the trace of the quadratic form associated with the matrix $D(\vec{x}, y)$ on the subspace $\ell(\mathscr{P})$ of $\ell(\mathscr{P}X)$ is a quadratic form with matrix $D(R(\vec{x}, y))$, for some rational function R. Moreover, the values of the quadratic form with matrix $\psi_n(D(\vec{x}, y))$ on the subspace $\ell(V_n)$ of $\ell(\mathscr{P}X^n)$ do not depend on y, so that one can pass to the limit as $y \to \infty$. In this way, the problem of constructing a self-similar Dirichlet form on the limit space is reduced to a non-linear finite-dimensional eigenvector problem. This strategy is used in [27] in order to get a Laplacian on the Sierpiński gasket, regarded as the limit space of the Hanoi Towers group, as well as for the "Pillow fractal" case (see Fig. 10).

Let us return now to the group $IMG(z^2+i)$, which is the group generated by the automaton in Fig. 16.

One can easily check that the asymptotic equivalence relations are given by

$$(10)^{-\infty}00v \sim (10)^{-\infty}01v, \quad \text{for every } v \in \{0,1\}^*.$$

We have

$$\mathscr{P} = \{w_1 = (10)^{-\infty}, w_2 = (01)^{-\infty}, w_3 = (10)^{-\infty}0\},$$

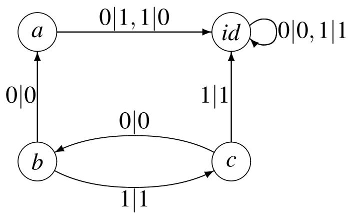


Fig. 16 The automaton generating the group $IMG(z^2+i)$

so that

$$w_1 1 = w_2, \qquad w_1 0 = w_3, \qquad w_2 0 = w_1.$$

Moreover, one has

$$g_{w_1} = b, \qquad g_{w_2} = c, \qquad g_{w_3} = a.$$

Put $\vec{x} = (x_{12}, x_{13}, x_{23}) = (r, s, t)$, with $r, s, t \in \mathbb{R}$. Let $\mathscr{E}_{n,\vec{x},y}$ be the Dirichlet form on $\ell(\mathscr{P}X^n)$ associated with the matrix $D_n = \psi_n(D)$, with

$$D = D(r,s,t,y) = \begin{pmatrix} y(1-b)+r+s & -r & -s \\ -r & y(1-c)+r+t & -t \\ -s & -t & y(1-a)+s+t \end{pmatrix}.$$

A long computation, via the Schur complement, shows that

$$Tr_{\mathscr{P}X^n}\mathscr{E}_{n+1,\vec{x},y} = \psi_n\big(D(R_r, R_s, R_t, y)\big),$$

with

$$R_r = R_r(r,s,t,y) = \frac{ty(rs+rt+st)}{2rys+2ryt+rs^2+2rst+rt^2+2syt+t^2y+s^2t+t^2s},$$

$$R_s = R_s(r,s,t,y) = \frac{(rs+rt+st)(rs+2ry+rt+st+yt)}{2rys+2ryt+rs^2+2rst+rt^2+2syt+t^2y+s^2t+t^2s},$$

$$R_t = R_t(r,s,t,y) = \frac{sy(rs+rt+st)}{2rys+2ryt+rs^2+2rst+rt^2+2syt+t^2y+s^2t+t^2s}.$$

Let $\ell(V_n)$ be the subspace of $\ell(\mathscr{P}X^n)$ consisting of the functions which are constant on the asymptotic equivalence classes. It follows from the description of the equivalence classes that the values of $\mathscr{E}_{n,\vec{x},y}$ on $\ell(V_n)$ do not depend on y. This allows us to simplify our computations, so that we can assume:

$$Tr_{\mathscr{P}X^n}\mathscr{E}_{n+1,\vec{x},y} = \frac{rs+rt+st}{2rs+2rt+2st+t^2} D_n(t, 2r+t, s, y).$$

Now let $\mathscr{E}^V_{n,\vec{x}}$ be the restriction of $\mathscr{E}_{n,\vec{x},y}$ to $\ell(V_n)$. One can show that $Tr_{V_n}\mathscr{E}^V_{n+1,\vec{x}}$ is equal to the restriction of $\lim_{y\to\infty} Tr_{\mathscr{P}X^n}\mathscr{E}_{n+1,\vec{x},y}$ onto $\ell(V_n)$. Observe that letting y go to infinity corresponds to gluing together equivalent sequences of $\mathscr{P}X^n$, which have to be identified in $\mathscr{J}_G$ (this means that conductances between them become infinite).

Let $\omega \approx 1.5213\ldots$ be the real root of the polynomial $p(x) = x^3 - x - 2$. Then, if we choose

$$r = 1, \qquad s = \omega^2, \qquad t = \omega,$$

we get

$$Tr_{V_n} \mathscr{E}^V_{n+1,1,\omega^2,\omega} = \gamma \mathscr{E}^V_{n,1,\omega^2,\omega},$$

with $\gamma = (2\omega^2 + 3\omega + 2)/(3\omega^2 + 4\omega + 4)$. Therefore, we have a Dirichlet form on the limit space if we consider the limit of the form

$$\widetilde{\mathscr{E}}_n = \gamma^{-n} \mathscr{E}^V_{n,1,\omega^2,\omega}.$$

Remark 5.3 It is interesting to observe that I. Bondarenko showed in [3] that the diameters of the Schreier graphs $\{\Gamma_n\}_{n\geqslant 1}$ of the group $IMG(z^2+i)$ have growth ω^n.

References

1. Bartholdi, L., Grigorchuk, R.I.: On the spectrum of Hecke type operators related to some fractal groups. In: Din. Sist., Avtom. i Beskon. Gruppy, Collected papers. Trudy Mat. Inst. Steklova. vol. 231, pp. 5–45 (2000). English translation in: Proc. Steklov Inst. Math., vol. 231, pp. 1–41 (2000)
2. Bartholdi, L., Virág, B.: Amenability via random walks. Duke Math. J. **130**, 39–56 (2005)
3. Bondarenko, I.: Groups generated by bounded automata and their Schreier graphs. Ph.D. Thesis, Texas A&M University, College Station, TX (2007). http://repository.tamu.edu/bitstream/handle/1969.1/85845/Bondarenko.pdf
4. Bondarenko, I., Ceccherini-Silberstein, T., Donno, A., Nekrashevych, V.: On a family of Schreier graphs of intermediate growth associated with a self-similar group. Eur. J. Comb. **33**, 1408–1421 (2012)
5. Bondarenko, I., Nekrashevych, V.: Post-critically finite self-similar groups. Algebra Discrete Math. **4**, 21–32 (2003)
6. Ceccherini-Silberstein, T., Donno, A., Iacono, D.: The Tutte polynomial of the Schreier graphs of the Grigorchuk group and the Basilica group. In: Bianchi, M., Longobardi, P., Maj, M., Scoppola, C.M. (eds.) Ischia Group Theory 2010: Proceedings of the Conference, pp. 45–68. World Scientific, Singapore (2011)
7. Cipriani, F., Guido, D., Isola, T., Sauvageot, J.L.: Differential 1-forms, their integrals and potential theory on the Sierpiński gasket. http://arxiv.org/abs/1105.1995
8. Cipriani, F., Sauvageot, J.L.: Derivations as square roots of Dirichlet forms. J. Funct. Anal. **201**, 78–120 (2003)
9. D'Angeli, D., Donno, A., Matter, M., Nagnibeda, T.: Schreier graphs of the Basilica group. J. Mod. Dyn. **4**, 167–205 (2010)
10. D'Angeli, D., Donno, A., Nagnibeda, T.: Counting dimer coverings on self-similar Schreier graphs. Eur. J. Comb. **33**, 1484–1513 (2012)
11. D'Angeli, D., Donno, A., Nagnibeda, T.: Partition functions of the Ising model on some self-similar Schreier graphs. In: Lenz, D., Sobieczky, F., Woess, W. (eds.) Progress in Probability: Random Walks, Boundaries and Spectra, vol. 64, pp. 277–304. Springer, Basel (2011)

12. Donno, A., Iacono, D.: The Tutte polynomial of the Sierpiński and Hanoi graphs. Adv. Geom., in press
13. Freiberg, U.R., Lancia, M.R.: Energy forms on conformal $\mathscr{C}^1$-diffeomorphic images of the Sierpiński gasket. Math. Nachr. **281**, 337–349 (2008)
14. Goldstein, S.: Random walks and diffusions on fractals. In: Kesten, H. (ed.) Percolation Theory and Ergodic Theory of Infinite Particle Systems. I.M.A. Math. Appl., vol. 8, pp. 121–129. Springer, New York (1987)
15. Grigorchuk, R.I., Šunić, Z.: Self-similarity and branching in group theory. In: Groups St. Andrews 2005, I. London Math. Soc. Lecture Note Ser., vol. 339, pp. 36–95. Cambridge University Press, Cambridge (2007)
16. Grigorchuk, R.I., Żuk, A.: On a torsion-free weakly branch group defined by a three-state automaton. Int. J. Algebra Comput. **12**, 223–246 (2002)
17. Gromov, M.: Structures métriques pour les variétés riemanniennes. Lafontaine, J., Pansu, P. (eds.) Textes Mathématiques, vol. 1. CEDIC, Paris (1981). ISBN 2-7124-0714-8
18. Kigami, J.: A harmonic calculus on the Sierpiński spaces. Jpn. J. Appl. Math. **6**, 259–290 (1989)
19. Kigami, J.: Analysis on Fractals. Cambridge Tracts in Mathematics, vol. 143. Cambridge University Press, Cambridge (2001)
20. Kusuoka, S.: A diffusion process on a fractal. In: Ito, K., Ikeda, N. (eds.) Probabilistic Methods on Mathematical Physics, Proc. Taniguchi International Symp., Katata & Kyoto, 1985, pp. 251–274. Academic Press, Boston (1987)
21. Lindstrøm, T.: Brownian Motion on Nested Fractals. Mem. Amer. Math. Soc., vol. 420. Am. Math. Soc., Providence (1990)
22. Matter, M., Nagnibeda, T.: Abelian sandpile model on randomly rooted graphs and self-similar groups. http://arxiv.org/abs/1105.4036
23. Mosco, U.: Dirichlet forms and self-similarity. In: New Directions in Dirichlet Forms. A.M.S./IP Studies in Advanced Mathematics, vol. 8, pp. 117–155. Am. Math. Soc., Providence (1998)
24. Mosco, U.: Lagrangian metrics on fractals. In: Spigler, R., Venakides, S. (eds.) Proceedings of Symposia in Applied Mathematics, vol. 54, pp. 301–323. Am. Math. Soc., Providence (1998)
25. Nekrashevych, V.: Hyperbolic spaces from self-similar group actions. Algebra Discrete Math. **1**, 77–86 (2003)
26. Nekrashevych, V.: Self-similar Groups. Mathematical Surveys and Monographs, vol. 117. Am. Math. Soc., Providence (2005). ISBN 0-8218-3831-8
27. Nekrashevych, V., Teplyaev, A.: Groups and analysis on fractals. In: Analysis on Graphs and Its Applications. Proc. Symp. Pure Math. Amer. Math. Soc., vol. 77, pp. 143–180. Am. Math. Soc., Providence (2008)
28. Peirone, R.: Convergence of Dirichlet forms on fractals. In: Topics on Concentration Phenomena and Problems with Multiple Scales. Lect. Notes Unione Mat. Ital., vol. 2, pp. 139–188. Springer, Berlin (2006)
29. Rogers, L., Teplyaev, A.: Laplacians on the Basilica Julia set. Commun. Pure Appl. Anal. **9**, 211–231 (2010)
30. Sidki, S.: Automorphisms of one-rooted trees: Growth, circuit structure and acyclicity. J. Math. Sci. (N.Y.) **100**, 1925–1943 (2000)

Some Remarks on Generalized Gaussian Noise

Saverio Giulini

Abstract The aim of this paper is to obtain sharp estimates about the behavior (local and at infinity) of the convolution of n copies of Generalized Gaussian Densities both in the symmetric and asymmetric case. Moreover, we obtain some improvement of the estimate of the parameters of these densities from data samples.

Keywords Asymptotic approximations · Generalized Gaussian processes · Noisy environment · Edgeworth expansion

Mathematics Subject Classification (2010) Primary 41A60 · 60F05 · Secondary 60G50 · 62P30

1 Introduction

Signal processing concerns itself with the treatment of signals in additive white noise, which is often assumed to be Gaussian. This assumption is justified by the Central Limit Theorem, since environmental noise is supposed to result from the combination of a large number of independent sources.

More realistically, the distribution of source energies is characterized by a small number of very strong sources and a huge number of very weak sources [22]. So the Lindberg condition, a key requirement of the Central Limit Theorem, could be violated.

Actually in many real applications (terrestrial noise disturbances in low frequency radio-communication, ship-traffic-radiated noise in sonar and underwater communication systems, etc.) background noise appears to deviate from Gaussianity both in terms of asymmetry and sharpness.

Since the performances of signal processing algorithms, optimized in presence of Gaussian noise, may decay significantly in non-Gaussian environments, it is crucial to provide realistic (and simple) modeling of a wide set of noise probability density functions (pdf's) in order to optimize signal detection.

S. Giulini (✉)
D.S.A., Facoltà di Architettura, Università di Genova, Str. S. Agostino 37, 16123 Genova, Italy
e-mail: giulini@dima.unige.it

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3,
DOI 10.1007/978-88-470-2853-1_11,

The family of Generalized Gaussian pdf's

$$p_{gG}(x) = \frac{\gamma^{\frac{1}{c}} c}{2\Gamma(\frac{1}{c})} e^{-\gamma|x|^c}, \quad \gamma, c \in (0, +\infty)$$

was introduced in 1972 by J.H. Miller and J.B. Thomas [16] to illustrate the structure and performance of a class of non-linear detectors for discrete-time signals. This family of noise densities was commonly used in communication problems to model certain atmospheric impulsive noise (see e.g. [12, 14]). More recently, the so called hyperbolic assumption [22] (i.e., one supposes that the product of the strength and the position of any source is approximately constant, when the sources of the background noise are arranged by decreasing strength) justifies the use of the Champernowne model

$$p_C(x) = \frac{1}{\alpha} \frac{\sinh(\frac{\alpha\pi}{\beta})}{\cosh(\frac{\pi(\alpha+x)}{2\beta}) \cosh(\frac{\pi(x-\alpha)}{2\beta})}.$$

Because of their attractive properties (mainly stability under convolution), also symmetric alpha stable models

$$\widehat{p_\alpha}(\xi) = e^{-\gamma|\xi|^\alpha}, \quad \gamma > 0,\ 0 < \alpha \leqslant 2$$

($\hat{\cdot}$ denotes the Fourier transform on the line) were applied to communications, sonar, radar, etc. [18].

Finally, experimental results on the detection of known deterministic signals corrupted by real underwater acoustic noise has been used to compare different classes of pdf models [20, 21]: more general results were obtained in the case of Asymmetric Generalized Gaussian densities

$$p_{\alpha gG}(x) = \frac{c}{\Gamma(\frac{1}{c})} \frac{(\gamma_l \gamma_r)^{1/c}}{\gamma_l^{1/c} + \gamma_r^{1/c}} \begin{cases} e^{-\gamma_l(-x)^c} & \text{if } x < 0, \\ e^{-\gamma_r x^c} & \text{if } x \geqslant 0, \end{cases} \quad \gamma_l, \gamma_r > 0,\ c \in (0, +\infty)$$

introduced in [21].

Our goal is to obtain sharp estimates about the behavior (local and at infinity) of the convolution of n copies of Generalized Gaussian densities both in the symmetric and asymmetric cases. We shall also obtain some improvement of the estimate of the parameters of the densities p_{gG} from data samples.

2 Analysis at Infinity

We denote by $p_{\gamma_l,\gamma_r,c}$ the Asymmetric Generalized Gaussian density corresponding to the parameters γ_l, γ_r, c (in the symmetric case we put $p_{\gamma_l,c} = p_{\gamma,\gamma,c}$). We can rewrite

$$p_{\gamma_l,\gamma_r,c}(x) = \frac{c}{\Gamma(\frac{1}{c})} \frac{(\gamma_l\gamma_r)^{1/c}}{\gamma_l^{1/c} + \gamma_r^{1/c}} e^{-(\gamma_l+(\gamma_r-\gamma_l)H(x))|x|^c}$$

where H is the Heavyside function $H(x) = \begin{cases} 0 & \text{if } x<0, \\ 1 & \text{if } x\geqslant 0. \end{cases}$

Our goal is to obtain a sharp estimate, for large values of x, of the convolution power $p^{*(n+1)}_{\gamma_l,\gamma_r,c}$. When $\gamma_l = \gamma_r$ explicit computation can be done only in the "central" case $c = 2$ and in the extreme cases $c = 1$ and $c = \infty$ (where $p_{\cdot,\infty} = \frac{1}{2}\chi_{[-1,1]}$).

Indeed, it is well known that

$$\left(\sqrt{\frac{\gamma}{\pi}} e^{-\gamma(\cdot)^2}\right)^{*(n+1)}(x) = \sqrt{\frac{\gamma}{(n+1)\pi}} e^{-\frac{\gamma}{n+1}x^2},$$

while, with a long but straightforward computation, one can show that

$$\left(\frac{\gamma}{2} e^{-\gamma|\cdot|}\right)^{*(n+1)}(x) = \frac{\gamma}{2^{2n+1}} \sum_{j=0}^{n} \binom{2n-j}{n} \frac{\gamma^j}{j!} |x|^j e^{-\gamma|x|}, \tag{1}$$

$$\left(\frac{1}{2}\chi_{[-1,1]}\right)^{*(n+1)}(x) = \frac{1}{n!2^{n+1}} \sum_{j=0}^{[\frac{n}{2}]} (-1)^j \binom{n+1}{j} \big(n+1-2j-|x|\big)_+^n, \tag{2}$$

where $[y]$ denotes the integral part of y and $(y)_+ = \begin{cases} y & \text{if } y>0, \\ 0 & \text{if } y\leqslant 0. \end{cases}$

We need to estimate

$$p^{*(n+1)}_{\gamma_l,\gamma_r,c}(x) = \int_{\mathbb{R}^n} p_{\gamma_l,\gamma_r,c}(x-x_1) \prod_{j=2}^{n} p_{\gamma_l,\gamma_r,c}(x_{j-1}-x_j) p_{\gamma_l,\gamma_r,c}(x_n)\, dx_1\, dx_2 \cdots dx_n.$$

Along the guidelines of (1) and (2), we shall prove an "asymptotic almost stability" property for convolution powers, in the sense that

$$p^{*(n+1)}_{\gamma_l,\gamma_r,c}(x) \sim a_{n,c} \left(\frac{x}{n+1}\right)^{b_{n,c}} p_{\gamma_l(n+1),\gamma_r(n+1),c}\left(\frac{x}{n+1}\right)$$

for suitable constants $a_{n,c}, b_{n,c}$.

We put

$$f(y_1, y_2, \ldots, y_{n-1}, y_n) = \sum_{j=1}^{n+1} \big(\gamma_l + (\gamma_r - \gamma_l)H(y_j)\big)|y_j|^c \tag{3}$$

and we perform the change of variables $\frac{x_j}{x} = t_j$, for every $j = 1, 2, \ldots, n$. Then

$$p^{*(n+1)}_{\gamma_l,\gamma_r,c}(x)$$
$$= \left(\frac{c}{\Gamma(\frac{1}{c})} \frac{(\gamma_l\gamma_r)^{1/c}}{\gamma_l^{1/c} + \gamma_r^{1/c}}\right)^{n+1}$$

$$\times\ |x|^n \int_{\mathbb{R}^n} e^{-|x|^c f((1-t_1)\,\mathrm{sign}\,x,(t_1-t_2)\,\mathrm{sign}\,x,\ldots,(t_{n-1}-t_n)\,\mathrm{sign}\,x,t_n\,\mathrm{sign}\,x)} dt_1\,dt_2\cdots dt_n. \tag{4}$$

The Laplace method for multiple integrals can be used to estimate (4). We state in advance a compactification Lemma [4, Lemma 38].

Lemma 2.1 *Let f and $h\colon \mathbb{R}^n \to \mathbb{R}$ be continuous functions such that*

$$\int_{\mathbb{R}^n} \left|h(t)\right| e^{-f(t)}\,dt < +\infty.$$

Suppose that the function f attains its minimum value in a unique point $t_{\min}$. Moreover suppose there exists a neighborhood of $t_{\min}$ where the function h is different from 0. Then for any neighborhood V of $t_{\min}$

$$\int_{\mathbb{R}^n} \left|h(t)\right| e^{-\lambda f(t)}\,dt \sim \int_V \left|h(t)\right| e^{-\lambda f(t)}\,dt$$

as $\lambda \longrightarrow +\infty$.

Our main tool will be the Hsu's extension of the Laplace method to the case of multiple integrals [11] (see also [4, Thm. 41]).

Lemma 2.2 *Let U and V be two subsets of $\mathbb{R}^n$ such that V is compact, U is open and $V \subset U$. Suppose $f\colon U \to \mathbb{R}$ is a twice differentiable function and $h\colon V \to \mathbb{R}$ is a continuous function such that $\int_V |h(t)|e^{-\lambda f(t)}\,dt$ is finite, if λ is large enough. If $t_{\min}$ is the unique point of minimum for f inside V and the Hessian matrix of f therein, $H(t_{\min})$, is positive definite, then*

$$\int_V h(t)e^{-\lambda f(t)}\,dt \longrightarrow \left(\frac{2\pi}{\lambda}\right)^{n/2} \frac{h(t_{\min})}{\sqrt{|\det H(t_{\min})|}} e^{-\lambda f(t_{\min})}$$

as $\lambda \longrightarrow +\infty$.

Now we are ready to prove our main result.

Theorem 2.3 *Suppose $c > 1$. Then, as $x \longrightarrow \pm\infty$, the following estimate holds*

$$p^{*(n+1)}_{\gamma_l,\gamma_r,c}(x) \sim \left(\frac{2c\pi}{\Gamma^2(1/c)(c-1)\gamma_\pm(\gamma_l^{-1/c}+\gamma_r^{-1/c})^2}\right)^{n/2} \times (n+1)^{\frac{1}{2}-\frac{1}{c}} \left(\frac{|x|}{n+1}\right)^{n(1-\frac{c}{2})} p_{\gamma_l(n+1),\gamma_r(n+1),c}\left(\frac{x}{n+1}\right)$$

where $\gamma_\pm = \begin{cases} \gamma_r & \text{if } x \to -\infty, \\ \gamma_l & \text{if } x \to +\infty. \end{cases}$

Proof Let us put $t_0 = 1$, $t_{n+1} = 0$. Then the functions

$$f_{\pm}^{(n)}(t_1, t_2, \dots, t_n) = \sum_{j=1}^{n+1} \big(\gamma_l + (\gamma_r - \gamma_l) H\big(\pm(t_{j-1} - t_j)\big)\big) |t_{j-1} - t_j|^c$$

are continuous in $\mathbb{R}^n$ if $c > 0$ and $\mathscr{C}^\infty$ with the exception of the points lying in the hyperplanes $t_{j-1} = t_j$, $j = 1, 2, \dots, n+1$. Such hyperplanes split $\mathbb{R}^n$ into $2^{n+1} - 1$ open convex region S_k, $k = 0, 1, \dots, 2(2^n - 1)$. Only one of them, say S_0, is bounded:

$$S_0 = \big\{(t_1, t_2, \dots, t_n) \in \mathbb{R}^n : 0 < t_n < t_{n-1} < \dots < 1\big\}.$$

Since

$$\begin{aligned}\frac{\partial}{\partial t_k} & f_{\pm}^{(n)}(t_1, t_2, \dots, t_n) \\ &= c \sum_{j=0}^{1} (-1)^j \big(\gamma_l + (\gamma_r - \gamma_l) H\big(\pm(t_{j-1} - t_j)\big)\big) \frac{|t_{k-j} - t_{k-j+1}|^c}{t_{k-j} - t_{k-j+1}},\end{aligned}$$

the gradient of $f_{\pm}^{(n)}$ does not vanish in S_j if $j \neq 0$. On the other hand, in S_0,

$$\frac{\partial}{\partial t_k} f_{\pm}^{(n)}(t_1, t_2, \dots, t_n) = c\big(\gamma_l + (\gamma_r - \gamma_l) H(\pm 1)\big)\big((t_k - t_{k+1})^{c-1)} - (t_{k-1} - t_k)^{c-1}\big)$$

and therefore $\nabla f_{\pm}^{(n)} = 0$ (in S_0) if and only if

$$t_1 = \frac{n}{n+1}, \quad t_2 = \frac{n-1}{n+1}, \quad \dots, \quad t_n = \frac{1}{n+1}.$$

We put $\mathbf{t}_{\min} = (\frac{n}{n+1}, \frac{n-1}{n+1}, \dots, \frac{1}{n+1})$ and $\gamma_{\pm} = \gamma_l + (\gamma_r - \gamma_l) H(\pm 1)$; in such a point

$$f_{\pm}^{(n)}(\mathbf{t}_{\min}) = \gamma_{\pm}(n+1)^{1-c}.$$

Now the restriction of $f_{\pm}^{(n)}$ to the intersection of m hyperplanes, say $t_{j_r - 1} = t_{j_r}$ ($r = 1, 2, \dots, m$) coincides with $f_{\pm}^{(n-m)}(\mathbf{t}')$ where $(\mathbf{t}') \in \mathbb{R}^{n-m}$. Therefore the infimum of such a restriction of $f_{\pm}^{(n)}$ is $\gamma_{\pm}(n - m + 1)^{1-c}$.

It follows that

$$\min_{\mathbf{t} \in \mathbb{R}^n} = f_{\pm}^{(n)}(\mathbf{t}) = \gamma_{\pm}(n+1)^{1-c}.$$

On the other hand, if $c < 1$, the function $f_{\pm}^{(n)}$ attains its minimum value $\gamma_{\pm}$ in the point $t_1 = t_2 = \dots = t_n = 1$.

The Hessian matrix $\mathbb{H}_n$ of $f_{\pm}^{(n)}$ at the critical point $\mathbf{t}_{\min}$ is

$$\mathbb{H}_n(\mathbf{t}_{\min}) = c(c-1)\gamma_{\pm}(n+1)^{2-c} \begin{pmatrix} 2 & -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}. \quad (5)$$

If A_n denotes the $n \times n$ matrix on the right side of (5), a simple induction argument proves that its determinant satisfies the identity $\det A_n = n+1$. Therefore

$$\det \mathbb{H}_n(\mathbf{t}_{\min}) = \big(c(c-1)\gamma_{\pm}\big)^n (n+1)^{1+(2-c)n}.$$

The thesis follows at once from (4), Lemmas 2.1 and 2.2, with $\lambda = |x|^c$. □

3 Cumulants and Edgeworth Expansion

A large part of this section has no claim of originality: it is a simple remark about the correct use of some expansions related to the Central Limit Theorem.

Many people used the Hermite polynomial expansion (Gram–Charlier series; see (9) below) to approximate the convolution powers of a suitable density f. Despite some good results (see e.g. [1, 23]), this expansion has no probabilistic meaning and does not supply any indication about the speed of convergence. In fact H. Cramer noticed, in a slightly different setting, that only a rearrangement of (9), the so called Edgeworth series (see (8) below) gives an asymptotic expansion of the convolution power $f^{*(n)}$ ([6, Ch. VII]; see also [5]). Actually the Gram–Charlier series seems to be more user-friendly than the Edgeworth expansion because of the explicit presence in the latter of cumulants, whose general definition may appear slightly intricate (see [17] for an introduction to the use of cumulants in signal processing; see also [8–10]). I hope that formula (7) will convince the reader that cumulants are easy to compute.

It is well known that the Central Limit Theorem asserts that, under suitable conditions, sums of independent random variables are asymptotically normally distributed. Of course this Theorem does not supply any information on the behavior at infinity of the density of such sums. For many purposes, however, the local information appears inadequate as well seems. When higher order moments exist, we make use of the so called Edgeworth expansion.

Let F be a probability distribution with density f and characteristic function $\hat{f}(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)\,dx$. Let μ_k denote the moment of order k

$$\mu_k = \int_{-\infty}^{+\infty} x^k f(x)\,dx.$$

We suppose $\mu_1 = 0$, $\mu_2 < +\infty$. and we put, as usual, $\mu_2 = \sigma^2$.

If the absolute moment $\int_{-\infty}^{+\infty} x^s f(x)\,dx$ is finite for some positive integer s, then the function $\log \hat{f}$ (here log denotes the principal branch of the logarithm) is s times differentiable at the origin and we put

$$\chi_k = (-1)^k \frac{d^k}{dt^k}(\log \hat{f})(0), \quad k = 1, 2, \ldots, s. \tag{6}$$

The number χ_k is called the k-cumulant (or semi-invariant) of f.

The following general formula for cumulants can be easily deduced from their definition.

$$\chi_k = k! \sum_{s=1}^{k} (-1)^{s-1}(s-1)! \sum^{*} \prod_{j=1}^{k} \frac{\mu_j^{h_j}}{h_j!(j!)^{h_j}} \tag{7}$$

where $\sum^*$ denotes the sum over all the k-ple of non negative integers $h_1, h_2, \ldots, h_k$ such that $\sum_{j=1}^{k} j h_j = k$, $\sum_{j=1}^{k} h_j = s$.

It follows from definition (6) that cumulants of order $k > 2$ vanish if the signal is Gaussian and thus non zero cumulants provide a measure of the deviation from Gaussianity. Moreover the cumulant of two statistically independent random processes equals the sum of the cumulants of the individual random process (in analytical terms, the cumulants of a convolution are the sum of cumulants), whereas this is false for higher order moments.

Denote by $\mathfrak{n}_\sigma$ the Gaussian density (with variance σ^2)

$$\mathfrak{n}_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

($\mathfrak{n} = \mathfrak{n}_1$) and define the polynomial $P_{k,\sigma}$ ($k \geqslant 3$) in the following way

$$\begin{aligned}
&\mathfrak{n}_\sigma(x) P_{k,\sigma}(x) \\
&= \sum_{m=1}^{k-2} \frac{1}{m!} \left(\sum_{j_1+j_2+\cdots+j_m=k-2} \frac{\chi_{j_1+2}\chi_{j_2+2}\cdots\chi_{j_m+2}}{(j_1+2)!(j_2+2)!\cdots(j_m+2)!} \left(-\frac{d}{dx}\right)^{k+2m-2} \mathfrak{n}_\sigma(x) \right) \\
&= \sum_{m=1}^{k-2} \frac{1}{m!} \left(\sum_{j_1+j_2+\cdots+j_m=k-2} \frac{\chi_{j_1+2}\chi_{j_2+2}\cdots\chi_{j_m+2}}{(j_1+2)!(j_2+2)!\cdots(j_m+2)!\sigma^{k+2m-2}} H_{k+2m-2}\left(\frac{x}{\sigma}\right) \right)
\end{aligned}$$

where H_j denotes the Hermite polynomial

$$\mathfrak{n}(x) H_j(x) = (-1)^j \left(\frac{d}{dx}\right)^j \mathfrak{n}(x).$$

The following theorem holds (see e.g. [7, XVI.2] or [2, Th. 19.2] for a slightly more general version):

Theorem 3.1 *Suppose that the moments $\mu_3, \dots, \mu_r$ exist and that $\hat{f} \in L^p(\mathbb{R})$ for some $p \geqslant 1$. Then*

$$f^{*(n)}(x) - \frac{1}{\sqrt{n}} \mathfrak{n}_\sigma\left(\frac{y}{\sqrt{n}}\right)\left(1 + \sum_{k=3}^{r} n^{1-\frac{k}{2}} P_{k,\sigma}\left(\frac{y}{\sqrt{n}}\right)\right) = o\left(n^{\frac{1-r}{2}}\right). \tag{8}$$

The next lemma allows us to apply Theorem 3.1 to the density $p_{\gamma,c}$.

Lemma 3.2 $\widehat{p_{\gamma,c}} \in L^1$ *for any* $c > 0$.

Proof The proof is very easy if $c > 1$. A little additional effort allows us to prove the result for general positive c.

We define

$$L^2_\alpha(\mathbb{R}) = \left\{ f \in L^2(\mathbb{R}) : \int_{\mathbb{R}} \left(1 + |\xi|\right)^\alpha \left|\hat{f}(\xi)\right|^2 d\xi < +\infty \right\}.$$

Of course $\{f \in L^2_\alpha(\mathbb{R})$ for some $\alpha > \frac{1}{2}\} \Longrightarrow \hat{f} \in L^1(\mathbb{R})$. If $c > \frac{1}{2}$, then the derivative

$$\frac{d}{dx} e^{-\gamma |x|^c} = -\gamma c \operatorname{sign} x |x|^{c-1} e^{-\gamma |x|^c}$$

belongs to $L^2(\mathbb{R})$, therefore $e^{-\gamma|\cdot|^c} \in L^2_1(\mathbb{R})$.

If $0 < c < \frac{1}{2}$, we want to prove that some fractional derivative of order $\alpha \in (\frac{1}{2}, 1)$ of $e^{-\gamma|x|^c}$ belongs to $L^2(\mathbb{R})$.

Thanks to symmetry, it is enough to consider $x > 0$. Then we may define

$$D^\alpha e^{-\gamma|\cdot|^c}(x) = \frac{1}{\Gamma(1-\alpha)} \int_x^{+\infty} (y - x)^{-\alpha} \frac{d}{dy} e^{-\gamma|y|^c} \, dy.$$

It is obvious that the previous integral decays exponentially as $x \longrightarrow +\infty$. With the intent of studying the behavior as $x \longrightarrow 0$, we perform the change of variable $y = tx$ and obtain

$$D^\alpha e^{-\gamma|\cdot|^c}(x) = \frac{c\gamma}{\Gamma(1-\alpha)} \int_1^{+\infty} (t-1)^{-\alpha} t^{c-1} e^{-\gamma|tx|^c} \, dt.$$

It follows from Lebesgue dominated convergence theorem that

$$D^\alpha e^{-\gamma|\cdot|^c}(x) \sim C_{\gamma,\alpha,c} x^{c-\alpha} \quad \text{as } x \to 0.$$

Therefore $D^\alpha e^{-\gamma|\cdot|^c}$ is in $L^2(\mathbb{R})$, provided that $\alpha < c + \frac{1}{2}$. □

Let $c_{n,k} = \frac{1}{k!}\int_{-\infty}^{+\infty} f^{*(n)}(y) H_k(\frac{y}{\sigma\sqrt{n}})\,dy$ (of course $c_{1,k} = c_{2,k} = 0$ if $\mu_1 = 0$). As remarked before, the so called Gram–Charlier series

$$f^{*(n)}(x) = \frac{1}{\sqrt{n}}\mathfrak{n}_\sigma\left(\frac{y}{\sqrt{n}}\right)\left(1 + \sum_{k=3}^{+\infty} c_{n,k} H_k\left(\frac{y}{\sigma\sqrt{n}}\right)\right) \tag{9}$$

does not supply any information about the rate of convergence (in the case of generalized Gaussian densities the Gram–Charlier series is convergent). Moreover there is no practical reason to prefer the expansion (9): indeed the coefficient therein are not easier to compute than cumulants. Unfortunately it is very difficult to obtain good *numerical* estimates of the error in both series. These estimates should involve the cumulants, and the (almost obvious) inequality

$$|\chi_k| \leqslant k^k \mu_k$$

is far from being sharp. For example, in the Gaussian case, the cumulants of order larger than 2 are zero, while $\mu_k = \frac{(2k)!}{2^k k!}$.

4 Parameters Estimates

A general problem in designing a signal detector in non-Gaussian environments is the extraction, from real data samples, of the parameters the models of noise pdf depend on. When dealing with generalized Gaussian functions, the parameter c is difficult to estimate from the data; however it is linked with the sharpness of the pdf. The Higher-Order-Statistics parameter which provide the best description of sharpness variability is the fourth order kurtosis

$$\beta_2(c) = \frac{\mu_4}{\sigma_4} = \frac{\Gamma(\frac{5}{c})\Gamma(\frac{1}{c})}{\Gamma^2(\frac{3}{c})}.$$

To express the pdf p_{gG} in terms of the kurtosis β_2, we need to invert the above expression; of course an explicit inversion is not possible. A largely used analytical approximation of this inverse function is

$$c = c_1(\beta_2) = \sqrt{\frac{5}{\beta_2 - 1.865}} - 0.12, \tag{10}$$

in the range $1.87 \leqslant \beta_2 \leqslant 30$. For example, ocean noise distributions have been reported to have kurtosis in the range [2.3, 3.67] ([15, Table 5]; see also [22]). It is quite easy to do better.

The quantities

$$\beta_k(c) = \frac{\mu_{2k}}{\sigma_{2k}} = \frac{\Gamma(\frac{2k+1}{c})\Gamma^{k-1}(\frac{1}{c})}{\Gamma^k(\frac{3}{c})}$$

are decreasing functions of the variable c. Indeed

$$\beta_k'(c) = \frac{1}{c^2} \frac{\Gamma(\frac{2k+1}{c})\Gamma(\frac{3}{c})\Gamma^{k-1}(\frac{1}{c})}{\Gamma^{k+1}(\frac{3}{c})} \times \left(-(2k+1)\psi\left(\frac{2k+1}{c}\right) - (k-1)\psi\left(\frac{1}{c}\right) + 3k\psi\left(\frac{3}{c}\right)\right), \tag{11}$$

where $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ is the logarithmic derivative of the function Γ. Since

$$\psi(x) = -\gamma + \sum_{n=0}^{+\infty}\left(\frac{1}{n+1} - \frac{1}{n+x}\right)$$

and the series converges uniformly on every compact subset of $(0, +\infty)$, the second factor in (11) equals

$$4k(1-k)\sum_{n=0}^{+\infty} \frac{n}{(\frac{1}{c}+n)(\frac{3}{c}+n)(\frac{2k+1}{c}+n)} < 0.$$

From the definition of Γ function and Stirling's formula it follows that

$$\beta_k(c) = \begin{cases} \frac{3^k}{2k+1}(1 + \frac{\pi^2}{c^2}k(k-1) + \mathscr{O}(\frac{1}{c^4})) & \text{as } c \longrightarrow +\infty, \\ \frac{(2k+1)^{\frac{2k+1}{c}-\frac{1}{2}}}{3^{k(\frac{3}{c}-\frac{1}{2})}}(1 + \mathscr{O}(c)) & \text{as } c \longrightarrow +0. \end{cases} \tag{12}$$

Although the forthcoming steps are valid for general k, we limit ourself to the most interesting case, $k = 2$. If we are concerned with small values of the parameter c, formula (12) suggests the following approximation

$$\beta_2(c) \approx \frac{5^{\frac{5}{c}} - \frac{1}{2}}{3^{\frac{6}{c}} - 1} + \frac{2}{9}$$

(note that the two functions coincide if $c = 2$). Therefore an approximate inverse of β_2 is

$$c_2(\beta_2) = \frac{5\ln 5 - 6\ln 3}{\ln(\frac{\sqrt{5}}{3}(\beta_2 - \frac{2}{9}))}. \tag{13}$$

Formula (12) provides better estimates than (10) in the range $\beta_2 \geqslant 2.2$ (that is, $0 < c \leqslant 3.9$). On the other hand, for large values of c, still better results can be obtained by interpolating the function β_2 with a polynomial of degree 2 in the variable $1/c$. Then we have

$$\beta_2(c) \approx \frac{9}{5}\left(1 + \frac{2}{15c} - \frac{4}{5c^2}\right).$$

We force the functions to have the same value if $c = 1$ ($\beta_2 = 6$) and $c = 2$ ($\beta_2 = 3$) and we get the approximate inverse function

$$c_3(\beta_2) = -\frac{3(1+\sqrt{40\beta_2 - 71})}{2(9-5\beta_2)}.$$

The function c_3 works very well in the wide range $c \leqslant 9$ ($\beta_2 \geqslant 1.9$) and it provides the best approximation when $c > 3$ (that is, $\beta_2 < 2.41$; compare with [Appendix, Table 5]).

Acknowledgements The author is grateful to Carlo S. Regazzoni and Claudio Sacchi for introducing him to the study of non-Gaussian noise. He would also like to thank Alessandro Veneruso for helpful comments and suggestions.

Appendix

In this appendix we collect some computations about the coefficients of the expansions related to the Central Limit Theorem. Table 5 compares different approximate inverse functions of the kurtosis.

Tables 1 and 2 contain the first values of the cumulants of even order (less than 10; formula (7)) and the corresponding coefficients of the Gram–Charlier series, both in the case of symmetric pdf's. It is worthwhile to remark that the computation of the Gram–Charlier coefficients can be efficiently performed by means of the formula

$$\int_{\mathbb{R}} f^{*(n)}(x)x^k\,dx = \sum_{\substack{k_1+k_2+\cdots+k_n=k\\ k_1,k_2,\ldots,k_n\geqslant 0}} \frac{k!}{k_1!k_2!\cdots k_n!}\mu_{k_1}\mu_{k_2}\cdots\mu_{k_n}.$$

In Table 3, for $k = 4, 6, 8, 10$, we compare the first terms of the Edgeworth and Gram–Charlier expansions (see (8) and (9)); we put $\tilde{\chi}_k = \chi_k/k!\mu_2^{\frac{k}{2}}$. Of course the argument of the Hermite polynomials is $y/\sqrt{n\mu_2}$.

It is apparent that the computational cost of the expansion (9) is approximately the same as Edgeworth series (8); however the latter expansion provides a better approximation, as Table 4 shows. Anyway, as noticed in [3], "for situations where

Table 1 Cumulants

$\chi_2 = \mu_2$
$\chi_4 = \mu_4 - 3\mu_2^2$
$\chi_6 = \mu_6 - 15\mu_2\mu_4 + 30\mu_2^3$
$\chi_8 = \mu_8 - 28\mu_2\mu_6 - 35\mu_4^2 + 420\mu_2^2\mu_4 - 630\mu_2^4$
$\chi_{10} = \mu_{10} - 45\mu_2\mu_8 - 210\mu_6(\mu_4 - 6\mu_2^2) + 3150\mu_2\mu_4^2 - 18900\mu_2^3\mu_4 + 22680\mu_2^5$

Table 2 Coefficients of Gram–Charlier series

$c_{n,2} = 0$	
$c_{n,4} = \frac{1}{4!n}\frac{\chi_4}{\mu_2^2}$	$c_{n,6} = \frac{1}{6!n^2}\frac{\chi_6}{\mu_2^3}$
$c_{n,8} = \frac{1}{8!}(\frac{1}{n^3}\frac{\chi_8}{\mu_2^4} + \frac{35}{n^2}(\frac{\chi_4}{\mu_2^2})^2)$	$c_{n,10} = \frac{1}{10!}(\frac{1}{n^4}\frac{\chi_{10}}{\mu_2^5} + \frac{210}{n^3}\frac{\chi_4}{\mu_2^2}\frac{\chi_6}{\mu_2^3})$

Table 3 Expansion first terms

Edgeworth	Gram–Charlier
$\frac{1}{n}\tilde{\chi}_4 H_4$	$\frac{1}{n}\tilde{\chi}_4 H_4$
$\frac{1}{n^2}(\tilde{\chi}_6 H_6 + \frac{1}{2}\tilde{\chi}_4^2 H_8)$	$\frac{1}{n^2}\tilde{\chi}_6 H_6$
$\frac{1}{n^3}(\tilde{\chi}_8 H_8 + \tilde{\chi}_4\tilde{\chi}_6 H_8 + \frac{1}{6}\tilde{\chi}_4^3 H_{12})$	$(\frac{1}{n^3}\tilde{\chi}_8 + \frac{1}{2n^2}\tilde{\chi}_4^2) H_8$
$\frac{1}{n^4}(\tilde{\chi}_{10} H_{10} + (\tilde{\chi}_4\tilde{\chi}_8 + \frac{1}{2}\tilde{\chi}_6^2) H_{12} + \frac{1}{2}\tilde{\chi}_4^2\tilde{\chi}_6 H_{14} + \frac{1}{4!}\tilde{\chi}_4^4 H_{16})$	$(\frac{1}{n^4}\tilde{\chi}_{10} + \frac{1}{n^3}\tilde{\chi}_4\tilde{\chi}_6) H_{10}$

Table 4 Maximum error

c	$*$	Edg.	Max. Err.	G-C	Max. Err.
1.5	6	6	2.27×10^{-5}	6	2.26×10^{-4}
1.5	10	6	6.44×10^{-6}	6	8.41×10^{-5}
1.5	10	8	2.66×10^{-6}	8	7.46×10^{-6}
4	6	6	2.33×10^{-5}	6	4.68×10^{-4}
4	10	6	3.61×10^{-6}	6	1.30×10^{-4}
4	10	8	3.84×10^{-7}	8	2.16×10^{-5}

Table 5 Approx. inverse functions of the kurtosis

β	$c(\beta)$	$c_1(\beta)$	$c_2(\beta)$	$c_3(\beta)$
1959.3	0.2	−0.0694592	0.19976	0.0043038
51.95	0.4	0.196	0.3985	0.0274
15.5788	0.6	0.483818	0.597098	0.533383
8.56514	0.8	0.743859	0.796443	0.77517
6	1	0.979632	0.996843	1
4.74348	1.2	1.19796	1.19806	1.21252
4.01786	1.4	1.40397	1.39959	1.41647
3.5527	1.6	1.60123	1.60084	1.61452
3.23235	1.8	1.79225	1.80118	1.80857
3	2	1.97888	2	2
2.82473	2.2	2.1625	2.19673	2.18982
2.68841	2.4	2.3442	2.39087	2.37881
2.57977	2.6	2.52485	2.58194	2.56756
2.49143	2.8	2.70519	2.76957	2.75656
2.4184	3	2.88584	2.95341	2.94619
2.18844	4	3.811778	3.80808	3.91307
2.0701	5	4.81746	4.54645	4.92674
2	6	5.96581	5.17111	6
1.95463	7	7.34883	5.69406	7.14057
1.92341	8	9.13209	6.1302	8.35377
1.90093	9	11.6771	6.49049	9.64364
1.88416	10	16.0347	6.7986	11.0134
1.86121	12	–	7.27084	14.0026

the estimate of a deviation of a pdf from a Gaussian one is needed, the asymptotic Edgeworth expansion is indispensable, and for high order moments the form of this expansion found by Petrov is necessary" (see also [13, 19]).

The first column of Table 4 lists the values of the parameter c, the second the number of convolutions, the third and fifth the number of terms that we consider in Edgeworth and Gram–Charlier series, respectively, and the fourth and sixth contain the absolute values of the corresponding maximum error in the interval $[-5, 5]$.

References

1. Bartucca, F., Biglieri, E.: DS-SSMA capacity for a mobile satellite system. Telecommun. Syst. **7**, 157–164 (1996)
2. Bhattacharya, R.N., Ranga Rao, R.: Normal Approximation and Asymptotic Expansions. Wiley, New York (1976). MR 04 #36272, **55**:9219
3. Blinnikov, S., Moessner, R.: Expansions for nearly Gaussian distributions. Astron. Astrophys. Suppl. Ser. **130**(1), 193–205 (1998)
4. Breitung, K.W.: Asymptotic Approximations for Probability Integrals. Springer, Berlin (1994)
5. Cramer, H.: On the composition of elementary errors. Skand. Aktuarietidskr. **11**, 13–74 and 141–180 (1928).
6. Cramer, H.: Random Variables and Probability Distributions. Cambridge University Press, Cambridge (1970)
7. Feller, W.: An Introduction to Probability Theory and Its Applications, II. Wiley, New York (1966). MR 00 #88081, **19**:466a
8. Giulini, S., Regazzoni, C.S., Sacchi, C., Teschioni, A.: Higher order statistics based sharpness evaluation of a generalized Gaussian pdf model in impulsive noisy environment. In: 9th IEEE SP Workshop on Statistical Signal and Array Processing, Portland, pp. 411–414 (1998)
9. Giulini, S., Oberti, F., Regazzoni, C.S., Sacchi, C.: Use of higher-order moments and related distribution series expansions for noise characterisation and performance evaluation in few-user asynchronous DS/CDMA systems. In: IEEE Global Conference on Telecommunications, GLOBECOM 99, Rio de Janeiro, pp. 472–476 (1999)
10. Giulini, S., Regazzoni, C.S., Sacchi, C., Vernazza, G.: Reliable parameter estimation for generalised Gaussian pdf models: Application to signal detection in non-Gaussian noisy environment. In: NSIP01, IEEE-EURASIP, Workshop on Nonlinear Signal and Image Processing, Baltimora, pp. 1–5 (2001)
11. Hsu, L.C.: On the asymptotic evaluation of a class of multiple integrals involving a parameter. Duke Math. J. **15**, 625–634 (1948)
12. Kassam, S.A.: Signal Detection in Non-Gaussian Noise. Springer, Berlin (1988)
13. Kolassa, J.E.: Series Approximation Methods in Statistics, 3rd edn. Lect. Notes in Statistics, vol. 88. Springer, New York (2006). MR 12 #95242, **95j**:62011
14. Lu, N.H., Eisenstein, B.A.: Detection of weak signals in non-Gaussian noise. IEEE Trans. Inf. Theory **IT-27**(7), 755–771 (1981)
15. Machell, E.J., Penrod, C.S., Ellis, G.E.: Statistical characteristics of ocean acoustic noise. In: Wegman, E.J., Schwartz, S.C., Thomas, J.B. (eds.) Topics in Non-Gaussian Signal Processing. Springer, New York (1984)
16. Miller, J.H., Thomas, J.B.: Detectors for discrete-time signals in non-Gaussian noise. IEEE Trans. Inf. Theory **IT-18**, 241–250 (1972)
17. Nikias, C.L., Mendel, J.: Signal processing with higher-order spectra. IEEE Signal Process. Mag. **10**, 10–27 (1993)
18. Nikias, C.L., Shao, M.: Signal Processing with Alpha-Stable Distributions and Applications. Wiley, New York (1995)

19. Petrov, V.V.: Limit Theorems of Probability Theory: Sequences of Independent Random Variables. Oxford Studies in Probability, vol. 4. Clarendon Press, New York (1995). MR 13 #53441, **96h**:60048
20. Tesei, A., Regazzoni, C.S.: Use of the fourth-order statistics for non-Gaussian noise modelling: The generalized Gaussian pdf in terms of kurtosis. In: VIII European Signal Processing Conference, EUSIPCO-96, Trieste, pp. 671–674 (1996)
21. Tesei, A., Regazzoni, C.S.: The asymmetric generalized Gaussian function: A new HOS-based model for generic noise pdfs. In: 8th IEEE SP Workshop on Statistical Signal and Array Processing, Corfu, pp. 210–213 (1996)
22. Webster, R.J.: Ambient noise statistics. IEEE Trans. Signal Process. **41**(7), 2249–2253 (1993)
23. Wu, K.T.: Bandwidth utilization of direct sequence spread-spectrum system. In: Proceedings International Symposium on Communication, Taiwan, pp. 307–311 (1991)

Eigenvalues of the Vertex Set Hecke Algebra of an Affine Building

Anna Maria Mantero and Anna Zappa

Abstract We study the eigenvalues of the vertex set Hecke algebra of an affine building, and prove, by a direct approach, the Weyl group invariance of any eigenvalue associated to a character. Moreover, we construct the Satake isomorphism of the Hecke algebra and we prove, by this isomorphism, that every eigenvalue arises from a character.

Keywords Buildings · Hecke algebras · Eigenvalues · Poisson kernel

Mathematics Subject Classification (2010) Primary 51E24 · 20C08 · Secondary 22E50

1 Introduction

The aim of this paper is to discuss the eigenvalues of the vertex set Hecke algebra $\mathscr{H}(\Delta)$ of any affine building Δ, using only its geometric properties. We avoid making use of the structure of any group acting on Δ.

To every multiplicative function χ on the fundamental apartment $\mathbb{A}$ of the building we associate an eigenvalue Λ_χ that can be expressed in terms of the Poisson kernel relative to the character χ. We prove the invariance of the eigenvalue Λ_χ with respect to the action of the finite Weyl group $\mathbf{W}$ on the characters. Moreover we prove that every eigenvalue arises from a character. Following the method used by Macdonald in his paper [7], our basic tool to obtain this characterization is the definition of the Satake isomorphism between the algebra $\mathscr{H}(\Delta)$ and the Hecke algebra of all $\mathbf{W}$-invariant operators on the fundamental apartment $\mathbb{A}$.

Our approach strongly depends on the definition of an α-boundary Ω_α, for every simple root α. Indeed, to every point of Ω we associate a tree, called *tree at infinity*,

A.M. Mantero (✉)
D.S.A., Università di Genova, Str. S. Agostino 37, 16123 Genova, Italy
e-mail: mantero@dima.unige.it

A. Zappa
D.I.M.A., Università di Genova, Via Dodecaneso 35, 16146 Genova, Italy
e-mail: zappa@dima.unige.it

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3,
DOI 10.1007/978-88-470-2853-1_12,

and define the α-boundary Ω_α as the collection of all such isomorphic trees. Then we show that the maximal boundary splits as the product of Ω_α and the boundary ∂T of the tree at infinity, and so any probability measure on Ω decomposes as the product of a probability measure on Ω_α and the standard measure on ∂T.

Our goal is a proof where the geometry of the building is foremost. Since we intend to address a non-specialized audience, we make use of a language that reduces to a minimum the algebraic prerequisites on affine buildings. So we make the paper as self-contained as possible, and for this goal we give the main results about buildings and their maximal boundary Ω, without claims of originality except possibly in the presentation. Our results are used in [11] to construct Macdonald's formula for the spherical functions on a building.

For buildings of type $\widetilde{A}_2$, $\widetilde{B}_2$ and $\widetilde{G}_2$ the eigenvalues of the algebra $\mathscr{H}(\Delta)$ are described in detail in [8–10] respectively.

We point out that an exhaustive exposition of the features of an affine building and its maximal boundary can be also found in the articles [12, 13] of J. Parkinson.

2 Affine Buildings

This section collects the fundamental definitions and properties concerning buildings and establishes notation that will be used in the rest of the paper. Our presentation is based on [3, 14, 15]; we refer the reader to these books for more details. For a similar exposition on buildings, see [12].

2.1 Labeled Chamber Complexes

A *simplicial complex* (with vertex set $\mathscr{V}$) is a collection Δ of finite subsets of $\mathscr{V}$ (called *simplices*) such that every singleton $\{v\}$ is a simplex and every subset of a simplex A is a simplex (called a *face* of A). The cardinality r of A is called the *rank* of A, and $r-1$ is called the *dimension* of A. Moreover a simplicial complex is said to be a *chamber complex* if all maximal simplices have the same dimension d and any two can be connected by a *gallery*, that is, by a sequence of maximal simplices in which any two consecutive ones are adjacent, that is, have a common codimension 1 face. The maximal simplices will then be called *chambers* and the rank $d+1$ (respectively the dimension d) of any chamber is called the *rank* (respectively the *dimension*) of Δ. The chamber complex is said to be *thin* (respectively *thick*) if every codimension 1 simplex is a face of exactly two chambers (respectively at least three chambers).

A *labeling* of the chamber complex Δ by a set I is a function τ which assigns to each vertex an element of I (the *type* of the vertex), in such a way that the vertices of every chamber are mapped bijectively onto I. The number of labels or types used is required to be the rank of Δ (that is the number of vertices of any chamber), and joinable vertices are required to have different types. When a chamber complex Δ is endowed by a labeling τ, we say that Δ is a *labeled chamber complex*. For every $A \in \Delta$, we will call $\tau(A)$ the type of A, that is, the subset of I consisting of the types of the vertices of A; moreover we call $I \setminus \tau(A)$ the co-type of A.

A *chamber system* over a set I is a set $\mathscr{C}$ such that each $i \in I$ determines a partition of $\mathscr{C}$, two elements in the same class of this partition being called *i-adjacent*. The elements of $\mathscr{C}$ are called chambers and we write $c \sim_i d$ to mean that the chambers c and d are i-adjacent. Then a labeled chamber complex is a chamber system over the set I of the types and two chambers are i-adjacent if they share a face of co-type i.

2.2 Coxeter Systems

Let W be a group (possibly infinite) and S be a set of generators of W of order 2. Then W is called a *Coxeter group* and the pair (W, S) is called a *Coxeter system*, if W admits the presentation

$$\left\langle S; (st)^{m(s,t)} = 1 \right\rangle,$$

where $m(s,t)$ is the order of st and there is one relation for each pair s, t, with $m(s,t) \leqslant \infty$. We shall assume that S is finite, and denote by N the cardinality of S. Then, if I is an arbitrary index set with $|I| = N$, we can write $S = (s_i)_{i \in I}$ and

$$W = \left\langle (s_i)_{i \in I}; (s_i s_j)^{m_{ij}} = 1 \right\rangle,$$

where $m(s_i s_j) = m_{ij}$. When $w \in W$ is written as $w = s_{i_1} s_{i_2} \cdots s_{i_k}$, with $i_j \in I$ and k minimal, we say that the expression is reduced and we call *length* $|w|$ of w the number k. The matrix $M = (m_{ij})_{i,j \in I}$, with entries $m_{ij} \in \mathbb{Z} \cup \{\infty\}$, is called the *Coxeter matrix* of W. We shall represent M by its diagram D: the nodes of D are the elements of I (or of S) and between two nodes there is a bond if $m_{ij} \geqslant 3$, with the label m_{ij} over the bond if $m_{ij} \geqslant 4$. We call D the *Coxeter diagram* or the *Coxeter graph* of W. We often say that W *has type* M if M is its Coxeter matrix.

2.3 Coxeter Complexes

Let (W, S) be a Coxeter system, with $S = (s_i)_{i \in I}$ finite. Define a *special coset* to be a coset $w\langle S' \rangle$, with $w \in W$ and $S' \subset S$, and define $\Sigma = \Sigma(W, S)$ to be the set of special cosets, partially ordered by the opposite of the inclusion relation: $B \leqslant A$ in Σ if and only if $B \supseteq A$ as subsets of W, in which case we say that B is a *face* of A. The set Σ is a simplicial complex; moreover it is a thin, chamber complex of rank $N =$ card S that admits a labeling, and the W-action on Σ is type-preserving. We remark that the chambers of Σ are the elements of W and, for each $i \in I$, $w \sim_i w'$ means that $w' = ws_i$ or $w' = w$. Following Tits, we shall call Σ the *Coxeter complex* associated to (W, S), or the *Coxeter complex of type M,* if M is the Coxeter matrix of W.

2.4 Buildings

Let (W, S) be a Coxeter system, and let $M = (m_{ij})_{i,j\in I}$ its Coxeter matrix. A *building of type M* (Tits, [15]) is a simplicial complex Δ, which can be expressed as the union of subcomplexes $\mathscr{A}$ (called *apartments*) satisfying the following axioms:

(B_0) each apartment $\mathscr{A}$ is isomorphic to the Coxeter complex $\Sigma(W, S)$ of type M of W;
(B_1) for any two simplices $A, B \in \Delta$, there is an apartment $\mathscr{A}(A, B)$ that contains both of them;
(B_2) if $\mathscr{A}$ and $\mathscr{A}'$ are two apartments containing A and B, there is an isomorphism $\mathscr{A} \to \mathscr{A}'$ fixing A and B pointwise.

Hence each apartment of Δ is a thin, labeled chamber complex over I of rank $N = |I|$. It can be proved that a building of type M is a chamber system over the set I with the properties:

(i) for each chamber $c \in \Delta$ and $i \in I$, there is a chamber $d \neq c$ in Δ such that $d \sim_i c$;
(ii) there exists a W-distance function

$$\delta : \Delta \times \Delta \to W$$

such that, if $f = i_1 \cdots i_k$ is a reduced word in the free monoid on I and $w_f = s_{i_1} \cdots s_{i_k} \in W$, then

$$\delta(c, d) = w_f$$

when c and d can be joined by a gallery of type f. We write $d = c \cdot \delta(c, d)$.

Actually it can be proved that each chamber system over a set I satisfying these properties is a building.

To ensure that the labeling of Δ and $\Sigma(W, S)$ are compatible, we assume that the isomorphisms in (B_0) and (B_2) are *type-preserving*; this implies that the isomorphism in (B_2) is unique. We write $\mathscr{C}(\Delta)$ for the chamber set of Δ. We call *rank of* Δ the cardinality N of the index set I.

We always assume that Δ is irreducible, that is, the associated Coxeter group W is irreducible (that is, its Coxeter graph is connected). Moreover we confine ourselves to consider *thick* buildings. In any building we always consider the complete system of apartments.

2.5 Regularity and Parameter System

Let Δ be a (irreducible) building of type M, with associated Coxeter group W over index set I, with $|I| = N$. We say that Δ is *locally finite* if

$$\left|\left\{d \in \mathscr{C}(\Delta), c \sim_i d\right\}\right| < \infty, \quad \text{for every } i \in I, \text{for every } c \in \mathscr{C}(\Delta).$$

Moreover we say that Δ is *regular* if this number does not depend on the chamber c. We shall assume that Δ is locally finite and regular. Since, for every $i \in I$, the set

$$\mathscr{C}_i(c) = \{d \in \mathscr{C}(\Delta), c \sim_i d\}$$

has a cardinality which does not depend on the choice of the chamber c, we put

$$q_i = |\mathscr{C}_i(c)|, \quad \text{for every } c \in \mathscr{C}(\Delta).$$

The set $\{q_i\}_{i\in I}$ is called the *parameter system* of the building. Notice that the parameter system has the following properties (see for instance [12] for the proof):

(i) $q_i = q_j$, whenever $m_{ij} < \infty$ is odd;
(ii) if $s_j = w s_i w^{-1}$ for some $w \in W$, then $q_i = q_j$.

Property (ii) implies that, for $w \in W$, the monomial $q_{i_1} \cdots q_{i_k}$ is independent of the particular reduced decomposition $w = s_{i_1} \cdots s_{i_k}$ of w [2]. So we define, for every $w \in W$,

$$q_w = q_{i_1} \cdots q_{i_k}$$

if $s_{i_1} \cdots s_{i_k}$ is any reduced expression for w. If we set, for every $c \in \mathscr{C}(\Delta)$ and every $w \in W$,

$$\mathscr{C}_w(c) = \{d \in \mathscr{C}(\Delta), \delta(c,d) = w\},$$

it can be proved that

$$|\mathscr{C}_w(c)| = q_w = q_{i_1} \cdots q_{i_k},$$

whenever $w = s_{i_1} \cdots s_{i_k}$ is a reduced expression for w. Hence the cardinality of the set $\mathscr{C}_w(c)$ does not depend on the choice of the chamber c. Obviously, $q_w = q_{w^{-1}}$.

If U is any finite subset of W, we define

$$U(q) = \sum_{w\in U} q_w$$

and we call it the *Poincaré polynomial* of U.

2.6 Affine Buildings

According to [2], W is called an *affine reflection group* if W is a group of affine isometries of a Euclidean space $\mathbb{V}$ (of dimension $n \geqslant 1$) generated by reflections s_H, where H ranges over a set locally finite $\mathscr{H}$ of affine hyperplanes of $\mathbb{V}$ which is W-invariant. We also assume W infinite. It is known that an affine reflection group is in fact a Coxeter group, because it has a finite set S of $n+1$ generators and admits the presentation

$$\langle S; (st)^{m(s,t)} = 1\rangle.$$

A building Δ (of type M) is said *affine* if the associated Coxeter group W is an affine reflection group. It is well known that each affine reflection group can be seen as the affine Weyl group of a root system. So we can define an affine building as a building associated to the affine Weyl group of a root system.

For the purpose of establishing notation, we shall give a brief discussion of root systems and its affine Weyl group, and we shall describe the geometric realization of the Coxeter complex associated to this group. We suggest [2] as an exhaustive reference on this subject.

2.7 Root Systems

Let $\mathbb{V}$ be a vector space over $\mathbb{R}$, of dimension $n \geqslant 1$, with inner product $\langle\cdot,\cdot\rangle$. For every $v \in \mathbb{V} \setminus \{0\}$ we define

$$v^\vee = \frac{2v}{\langle v, v\rangle}.$$

Let R be an irreducible, but not necessarily reduced, *root system* in $\mathbb{V}$ (see [2]). The elements of R are called *roots* and the rank of R is n.

Let $B = \{\alpha_i, i \in I_0\}$ be a basis of R, where $I_0 = \{1, \dots, n\}$. Thus B is a subset of R such that

(i) B is a vector space basis of $\mathbb{V}$;
(ii) each root in R can be written as a linear combination

$$\sum_{i\in I_0} k_i\alpha_i,$$

with integer coefficients k_i which are either all nonnegative or all nonpositive.

The roots in B are called *simple*. We say that $\alpha \in R$ is *positive* (respectively *negative*) if $k_i \geqslant 0$, for every $i \in I_0$ (respectively $k_i \leqslant 0$, for every $i \in I_0$). We denote by R^+ (respectively R^-) the set of all positive (respectively negative) roots. Thus $R^- = -R^+$ and $R = R^+ \cup R^-$ (as disjoint union). Define the *height* (with respect to B) of $\alpha = \sum_{i\in I_0} k_i\alpha_i$ by

$$ht(\alpha) = \sum_{i\in I_0} k_i.$$

There exists a unique root $\alpha_0 \in R$ whose height is maximal, and if we wright $\alpha_0 = \sum_{i\in I_0} m_i\alpha_i$, then $m_i \geqslant k_i$ for every root $\alpha = \sum_{i\in I_0} k_i\alpha_i$; in particular $m_i > 0$, for every $i \in I_0$ [2].

The set $R^\vee = \{\alpha^\vee, \alpha \in R\}$ is an irreducible root system, which is reduced if and only if R is. We call $R^\vee$ the *dual* (or *inverse*) of R and we call co-roots its elements.

For each $\alpha \in R$, denote by H_α the linear hyperplane of $\mathbb{V}$ defined by $\langle v, \alpha\rangle = 0$ and by $\mathscr{H}_0$ the family of all linear hyperplanes H_α. For every $\alpha \in R$, let s_α be the reflection with reflecting hyperplane H_α and $\mathbf{W}$ the subgroup of $GL(\mathbb{V})$ generated by $\{s_\alpha, \alpha \in R\}$. $\mathbf{W}$ permutes the set R and is a finite group, called the *Weyl group* of R. Note that $\mathbf{W}(R) = \mathbf{W}(R^\vee)$.

The hyperplanes in $\mathscr{H}_0$ split up $\mathbb{V}$ into finitely many regions; the connected components of $\mathbb{V} \setminus \bigcup_\alpha H_\alpha$ are (open) sectors based at 0, called the (open) *Weyl chambers* of $\mathbb{V}$ (with respect to R). The so called *fundamental Weyl chamber* or *fundamental sector* based at 0 (with respect to B) is the Weyl chamber

$$\mathbb{Q}_0 = \left\{ v \in \mathbb{V} : \langle v, \alpha_i \rangle > 0, i \in I_0 \right\}.$$

It is known that

(i) $\mathbf{W}$ is generated by $S_0 = \{s_i = s_{\alpha_i}, i \in I_0\}$ and hence $(\mathbf{W}, S_0)$ is a finite Coxeter system;
(ii) $\mathbf{W}$ acts simply transitively on Weyl chambers;
(iii) $\overline{\mathbb{Q}_0}$ is a fundamental domain for the action of $\mathbf{W}$ on $\mathbb{V}$.

Moreover, for every $\mathbf{w} \in \mathbf{W}$, we have $|\mathbf{w}| = n(\mathbf{w})$, where $n(\mathbf{w})$ is the number of positive roots α for which $\mathbf{w}(\alpha) < 0$. Recall that at most two root lengths occur in R if R is reduced, and all roots of a given length are conjugate under $\mathbf{W}$. When there are in R two distinct root lengths, we speak of *long* and *short* roots. In this case, the highest root α_0 is long.

The root system (or the associated Weyl group) can be characterized by the Dynkin diagram, which is the usual Coxeter graph D_0 of the group $\mathbf{W}$, where we add an arrow pointing to the shorter of the two roots. We refer to [2] for the classification of (irreducible) root systems. Observe that, for every $n \geqslant 1$, there is exactly one irreducible non-reduced root system (up to isomorphism) of rank n, denoted by BC_n. If we take $\mathbb{V} = \mathbb{R}^n$, with the usual inner product, the root system BC_n is the following:

$$R = \{\pm e_k, \pm 2e_k, 1 \leqslant k \leqslant n\} \cup \{\pm e_i \pm e_j, 1 \leqslant i < j \leqslant n\}.$$

Hence we can choose $B = \{\alpha_i, 1 \leqslant i \leqslant n\}$, if $\alpha_i = e_i - e_{i+1}$, $1 \leqslant i \leqslant n-1$ and $\alpha_n = e_n$. Moreover

$$R^+ = \{e_k, 2e_k, 1 \leqslant k \leqslant n\} \cup \{e_i \pm e_j, 1 \leqslant i < j \leqslant n\}$$

and $\alpha_0 = 2e_1$. In this case $R^\vee = R$ and $\mathbf{W}(BC_n) = \mathbf{W}(C_n) = \mathbf{W}(B_n)$.

We now decompose $R = R_1 \cup R_2 \cup R_0$, as disjoint union, by defining

$$R_1 = \{\alpha \in R : \ \alpha/2 \in R, 2\alpha \notin R\},$$

$$R_2 = \{\alpha \in R : \ \alpha/2 \notin R, 2\alpha \in R\},$$

$$R_0 = \{\alpha \in R : \ \alpha/2, 2\alpha \notin R\}.$$

Then $\alpha_0 \in R_1$, $\alpha_n \in R_2$, and $\alpha_i \in R_0$, for every $i = 1, \ldots, n-1$, and $\mathbf{W}$ stabilizes each R_j.

The $\mathbb{Z}$-span $L(R)$ of the root system R is called the *root lattice* of $\mathbb{V}$ and $L(R^\vee)$ is called the *co-root lattice* of $\mathbb{V}$ associated to R. Notice that $L(BC_n) = L(C_n) =$

$L(B_n^\vee)$. We simply denote by L the *co-root lattice* of $\mathbb{V}$ associated to R. Moreover we set

$$L^+ = \left\{ \sum_{\alpha\in R^+} n_\alpha \alpha^\vee, n_\alpha \in \mathbb{N} \right\}.$$

2.8 Affine Weyl Group of a Root System

Let R be an irreducible root system, not necessarily reduced. For every $\alpha \in R$ and $k \in \mathbb{Z}$, define an affine hyperplane

$$H_\alpha^k = \{ v \in \mathbb{V} : \langle v, \alpha \rangle = k \}.$$

We remark that $H_\alpha^k = H_{-\alpha}^{-k}$ and $H_\alpha^0 = H_\alpha$; moreover H_α^k can be obtained by translating H_α^0 by $\frac{k}{2}\alpha^\vee$.

When R is reduced we define $\mathscr{H} = \bigcup_{\alpha\in R^+} \mathscr{H}(\alpha)$, where, for every $\alpha \in R^+$,

$$\mathscr{H}(\alpha) = \{ H_\alpha^k \text{ for all } k \in \mathbb{Z} \}.$$

When R is not reduced, we note that, for every $\alpha \in R_2$, $H_\alpha^k = H_{2\alpha}^{2k}$; then we define

$$\mathscr{H}_1 = \{ H_\alpha^k : \alpha \in R_1, k \in 2\mathbb{Z}+1 \},$$
$$\mathscr{H}_2 = \{ H_\alpha^k : \alpha \in R_2, k \in \mathbb{Z} \},$$
$$\mathscr{H}_0 = \{ H_\alpha^k : \alpha \in R_0, k \in \mathbb{Z} \},$$

and $\mathscr{H} = \mathscr{H}_1 \cup \mathscr{H}_2 \cup \mathscr{H}_0$, as disjoint union. Since $\mathscr{H}_1 \cup \mathscr{H}_2 = \{H_\alpha^k, \alpha \in R_1, k \in \mathbb{Z}\}$, we can write

$$\mathscr{H} = \bigcup_{\alpha\in R_1\cup R_0} \mathscr{H}(\alpha),$$

by setting, for every $\alpha \in R_0$ or $\alpha \in R_1$, $\mathscr{H}(\alpha) = \{H_\alpha^k, \text{ for all } k \in \mathbb{Z}\}$, as in the reduced case. Actually, $R_1 \cup R_0$ is the root system of type C_n and the hyperplanes described before are these associated with this reduced root system.

Given an affine hyperplane $H_\alpha^k \in \mathscr{H}$, the affine reflection with respect to H_α^k is the map s_α^k defined by

$$s_\alpha^k(v) = v - \big(\langle v, \alpha \rangle - k\big)\alpha^\vee, \quad \text{for every } v \in \mathbb{V}.$$

The reflection s_α^k fixes H_α^k and sends 0 to $k\alpha^\vee$; in particular $s_\alpha^0 = s_\alpha$, for every $\alpha \in R$. Denote by $\mathscr{S}$ the set of all affine reflections, and let Aff($\mathbb{V}$) be the set of maps $v \mapsto Tv + \lambda$, for all $T \in GL(\mathbb{V})$ and $\lambda \in \mathbb{V}$. Then the *affine Weyl group W* of R is the subgroup of Aff($\mathbb{V}$) generated by all affine reflections s_α^k, $\alpha \in R$, $k \in \mathbb{Z}$.

Let $s_0 = s_{\alpha_0}^1$ and $I = I_0 \cup \{0\}$: then it can be proved that W is a Coxeter group over I, generated by the set $S = \{s_i, i \in I\}$. Writing t_λ for the translation $v \mapsto v + \lambda$, we can consider $\mathbb{V}$ as a subgroup of Aff($\mathbb{V}$), by identifying λ and t_λ. In this sense $\mathrm{Aff}(\mathbb{V}) = GL(\mathbb{V}) \ltimes \mathbb{V}$. In the same sense, if we consider the affine Weyl group W, the co-root lattice L can be seen as a subgroup of W, since t_λ, $\lambda \in L$, are the only translations of $\mathbb{V}$ belonging to W, and we have

$$W = \mathbf{W} \ltimes L.$$

We point out that $W(BC_n) = W(C_n)$, whereas $W(BC_n) \supset W(B_n)$. Hence each $w \in W$ can be uniquely written as $w = \mathbf{w}t_\lambda$ for some $\mathbf{w} \in \mathbf{W}$ and $\lambda \in L$; moreover, if $w_1 = \mathbf{w}_1 t_{\lambda_1}$ and $w_2 = \mathbf{w}_2 t_{\lambda_2}$, then $w_2^{-1} w_1 \in L$ if and only if $\mathbf{w}_1 = \mathbf{w}_2$. This implies that there is a bijection between the quotient W/L and $\mathbf{W}$: each coset wL determines a unique $\mathbf{w} \in \mathbf{W}$. So we denote by $\mathbf{w}$ the coset whose representative in W is w, and we write $w \in \mathbf{w}$ to intend that $w = \mathbf{w}t_\lambda$ for some $\lambda \in L$.

It is not difficult to construct, for each irreducible root system R, the Coxeter graph D of the affine Weyl group W; one just needs to work out the order of $s_i s_0$ for each $i \in I_0$, to see what new bonds and labels occur when the new node is adjoined to the Coxeter graph D_0 of $\mathbf{W}$, that is, of R. We refer to [5] for the classification of all affine Weyl groups.

2.9 Co-weight Lattice

Following standard notation, we call *weight lattice* of $\mathbb{V}$ associated to the root system R the $\mathbb{Z}$-span $\widehat{L}(R)$ of the vectors $\{\lambda_i^\vee, i \in I_0\}$, defined by $\langle \lambda_i^\vee, \alpha_j^\vee \rangle = \delta_{ij}$. We call $\widehat{L}(R^\vee)$ the *co-weight lattice* of $\mathbb{V}$ associated to the root system R. We simply set $\widehat{L} = \widehat{L}(R^\vee)$. Then $\widehat{L}$ is the $\mathbb{Z}$-span of the vectors $\{\lambda_i, i \in I_0\}$, defined by

$$\langle \lambda_i, \alpha_j \rangle = \delta_{ij}, \quad \text{for every } i, j \in I_0.$$

It is easy to see that, when R is reduced, $\widehat{L}$ contains L as a subgroup of finite index $\mathbf{f}$, the so called *index of connection*, with $1 \leqslant \mathbf{f} \leqslant n + 1$. Instead, when R is non-reduced, that is, when R has type BC_n, then $\widehat{L}(BC_n) = L(BC_n)$; thus, in this case

$$L(C_n) = L(BC_n) = \widehat{L}(BC_n) \not\supseteq \widehat{L}(C_n).$$

A co-weight λ is said *dominant* (respectively *strongly dominant*) if $\langle \lambda, \alpha_i \rangle \geqslant 0$ (respectively $\langle \lambda, \alpha_i \rangle > 0$) for every $i \in I_0$. We denote by $\widehat{L}^+$ (resp. $\widehat{L}^{++}$) the set of all dominant (respectively strongly dominant) co-weights. Thus $\lambda \in \widehat{L}^+$ if and only if $\lambda \in \overline{\mathbb{Q}_0}$ and $\lambda \in \widehat{L}^{++}$ if and only if $\lambda \in \mathbb{Q}_0$. Observe that L^+ does not lie in $\widehat{L}^+$ and $L^+ \cap \widehat{L}^+$ consists of all dominant coweights of type 0.

2.10 Geometric Realization of an Affine Coxeter Complex

Let W be the affine Weyl group of a root system R; let $\mathscr{H}$ be the collection of the affine hyperplanes associated to W. The open connected components of $\mathbb{V} \setminus \bigcup_{\alpha,k} H_\alpha^k$

are called *chambers*. Since R is irreducible, each chamber is an open (geometric) simplex of rank $n+1$ and dimension n. The extreme points of the closure of any chamber are called *vertices* and the 1 codimension faces of any chamber are called *panels*.

Denote by $\mathbb{A} = \mathbb{A}(R)$ the vector space $\mathbb{V}$ equipped with chambers, vertices, panels as defined above. Hence $\mathbb{A}$ is a geometric simplicial complex of rank $n+1$ and dimension n, realized as a tessellation of the vector space $\mathbb{V}$ in which all chambers are isomorphic.

It is convenient to single out one chamber, called *fundamental chamber* of $\mathbb{A}$, in the following way:

$$C_0 = \left\{v \in \mathbb{V} : 0 < \langle v, \alpha\rangle < 1 \; \forall \alpha \in R^+\right\} = \left\{v \in \mathbb{V} : \langle v, \alpha_i\rangle > 0 \; \forall i \in I_0, \langle v, \alpha_0\rangle < 1\right\}.$$

Define *walls* of C_0 the hyperplanes H_{α_i}, $i \in I_0$ and $H^1_{\alpha_0}$: then the group W is generated by the set of the reflections with respect to the walls of the fundamental chamber C_0.

We denote by $\mathscr{C}(\mathbb{A})$ the set of chambers and by $\mathscr{V}(\mathbb{A})$ the set of vertices of $\mathbb{A}$. It can be proved that W acts simply transitively on $\mathscr{C}(\mathbb{A})$ and $\overline{C_0}$ is a fundamental domain for the action of W on $\mathbb{V}$. Moreover $\mathbf{W}$ acts simply transitively on the set $\mathscr{C}(0)$ of all chambers C such that $0 \in \overline{C}$. Hence, there are well defined walls for each chamber $C \in \mathscr{C}(\mathbb{A})$: the walls of C are the images of the walls of C_0 under w, if $C = wC_0$. If we declare $wC_0 \sim_i wC_0$ and $wC_0 \sim_i ws_iC_0$, for each $w \in W$ and each $i \in I$, then the map

$$w \mapsto wC_0$$

is an isomorphism of the labeled chamber complex of W onto $\mathscr{C}(\mathbb{A})$. For every $w \in W$, we set $C_w = wC_0$. The vertices of C_0 are $X^0_0, X^0_1, \ldots, X^0_n$, where $X^0_0 = 0$ and $X^0_i = \lambda_i/m_i$, $i \in I_0$.

We declare $\tau(0) = 0$ and $\tau(\lambda_i/m_i) = i$, for $i \in I_0$; more generally we declare that a vertex X of $\mathbb{A}$ has type i, $i \in I$ if $X = w(X^0_i)$ for some $w \in W$. This define a unique labeling

$$\tau : \mathscr{V}(\mathbb{A}) \to I,$$

and the action of W on $\mathbb{A}$ is type-preserving. For any $i \in I$, we say that a panel of C_0 has *co-type* i if i is the type of the vertex of C_0 not lying in the panel; this extends to a unique labeling on the panels of $\mathbb{A}$.

Hence the Coxeter complex $\Sigma(W, S)$ associated to the affine Weyl group W admits a unique type-preserving isomorphism onto $\mathbb{A}$. Thus $\mathbb{A}$ may be regarded as the geometric realization of Σ. Up to this isomorphism, the co-root lattice L consists of all type 0 vertices of $\mathbb{A}$ and W acts on L. We point out that, for every $w \in W$, the chamber C_w can be joined to C_0 by a gallery $\gamma(C_0, C_w)$ of type $f = i_1 \cdots i_k$, if $w = s_{i_1} \cdots s_{i_k}$; so, recalling the definition of the W-distance function given in Sect. 2.4, we have $w = \delta(C_0, C_w)$. This suggests to denote the chamber C_w by $C_0 \cdot w$.

As in [2], a vertex X is a *special vertex* of $\mathbb{A}$ if, for every $\alpha \in R^+$, there exists $k \in \mathbb{Z}$ such that $X \in H^k_\alpha$. In particular the vertex 0 is special and hence every vertex of type 0 is special, but in general not all vertices of $\mathbb{A}$ are special. We shall denote by $\mathscr{V}_{sp}(\mathbb{A})$ the set of all special vertices of $\mathbb{A}$. We point out that, when R is reduced, $\mathscr{V}_{sp}(\mathbb{A}) = \widehat{L}$. More precisely, if R has type A_n, all $n+1$ types are special. Furthermore, the special types that occur in root systems R of type D_n, E_6 and G_2, are respectively four, three and only one. In all other cases there are two special types. In particular, if R has type B_n or C_n, the special vertices have type 0 or n. We refer the reader to [5] for more details.

When R has type C_n and $\alpha = \alpha_n$, then all vertices of type 0 lie in hyperplanes H_α^{2k}, for $k \in \mathbb{Z}$, whereas all vertices of type n lie in hyperplanes H_α^{2k+1}, for $k \in \mathbb{Z}$. Actually, the reflection s_{α_0} fixes each hyperplane H_α^h and the panel of co-type n containing 0, of the hyperplane $H_{\alpha_0}^0$. On the other hand, for every j, the reflection with respect to $H_{\alpha_0}^j$ fixes its panel and each hyperplane H_α^h. The same is true for every long root. If R has type B_n the previous property holds for each simple root $\alpha = \alpha_i, i = 1, \dots, n-1$, and then for every long root.

When R is non-reduced, the Coxeter complex $\Sigma(W, S)$ associated to the root system of type BC_n has the same geometric realization as the one associated to the root system of type C_n. Then the special types are type 0 and type n, and are arranged as above. Since $\widehat{L}(BC_n) = L(BC_n)$, the lattice $\widehat{L}(BC_n)$ is a proper subset of $\mathscr{V}_{sp}(\mathbb{A})$ and it consists of all type 0 vertices, lying in the hyperplanes H_i^{2k}, for $k \in \mathbb{Z}$ and $i = 0, n$.

In general we denote by $\widehat{\mathscr{V}}(\mathbb{A})$ the set of all special vertices of $\mathbb{A}$ belonging to $\widehat{L}$; so $\widehat{\mathscr{V}}(\mathbb{A})$ inherits the group structure of $\widehat{L}$. If we define $\hat{I} := \{\tau(\lambda) : \lambda \in \widehat{L}\}$, then $\widehat{\mathscr{V}}(\mathbb{A})$ is the set of all special vertices of $\mathbb{A}$ whose type belongs to $\hat{I}$. We remark that $\hat{I} = \{i \in I : m_i = 1\}$. See [12] for a proof of this property.

For every $\lambda \in \widehat{L}^+$, we define

$$\mathbf{W}_\lambda = \{\mathbf{w} \in \mathbf{W} : \mathbf{w}\lambda = \lambda\}.$$

If X_λ is the special vertex of $\mathbb{A}$ associated with λ and C_λ is the chamber containing X_λ in the minimal gallery connecting C_0 to X_λ, that is, the chamber of $\mathbb{Q}_0$ containing X_λ and nearest to C_0, then the set $\mathbf{W}_\lambda$ is the stabilizer of X_λ in $\mathbf{W}$. Moreover denote by w_λ the unique element of W such that $C_\lambda = w_\lambda(C_0)$.

Finally, for each $i \in \hat{I}$, denote by $\mathbf{W}_i$ the stabilizer in W of the vertex X_i^0 of type i lying in the fundamental chamber C_0, that is, the Weyl group associated with $I_i = I \setminus \{i\}$. Obviously $\mathbf{W}_0 = \mathbf{W}$.

2.11 Extended Affine Weyl Group of R

Let us consider in $\mathrm{Aff}(\mathbb{V})$ the translation group corresponding to $\widehat{L}$; since this group is also normalized by $\mathbf{W}$, we can form the semi-direct product

$$\widehat{W} = \mathbf{W} \ltimes \widehat{L},$$

called the *extended affine Weyl group* of R. Observe that $\widehat{W}/W$ is isomorphic to $\widehat{L}/L$, hence $\widehat{W}$ contains W as a normal subgroup of finite index $\mathbf{f}$. In particular when R is non-reduced, then $\widehat{W}(BC_n) = W(BC_n)$, as in this case $\widehat{L}(BC_n) = L(BC_n)$; moreover $\widehat{W}(BC_n)$ is not isomorphic to $\widehat{W}(C_n)$, since $\widehat{W}(C_n)$ is larger than $W(C_n)$. Notice that $\widehat{W}$ permutes the hyperplanes in $\mathscr{H}$ and acts transitively, but not simply transitively, on $\mathscr{C}(\mathbb{A})$.

Given any two special vertices X, Y of $\mathbb{A}$, there exists a unique $\hat{w} \in \widehat{W}$ such that $\hat{w}(X) = 0$ and $\hat{w}(Y)$ belongs to $\overline{\mathbb{Q}}_0$. Call *shape* of Y with respect to X the element $\lambda = \hat{w}(Y)$ of $\widehat{L}^+$, and denote it by $\sigma(X, Y)$. For every $\lambda \in \widehat{L}^+$, we set

$$\mathscr{V}_\lambda(X) = \big\{Y \in \mathscr{V}(\mathbb{A}) : \sigma(X, Y) = \lambda\big\}.$$

As for W/L, there is a bijection between the quotient $\widehat{W}/\widehat{L}$ and $\mathbf{W}$: each coset $\hat{w}\widehat{L}$ determines a unique $\mathbf{w} \in \mathbf{W}$. Denote by $\mathbf{w}$ the coset whose representative in $\mathbf{W}$ is $\mathbf{w}$. Hence $\hat{w} \in \mathbf{w}$ means that $\hat{w} = \mathbf{w}t_\lambda$ for some $\lambda \in \widehat{L}$.

For every $\hat{w} \in \widehat{W}$, let

$$\mathscr{L}(\hat{w}) = \big|\big\{H \in \mathscr{H} : H \text{ separates } C_0 \text{ and } \hat{w}(C_0)\big\}\big|.$$

If $w \in W$, then $\mathscr{L}(w) = |w|$. The subgroup $G = \{g \in \widehat{W} : \mathscr{L}(g) = 0\}$ is the stabilizer of C_0 in $\widehat{W}$ and

$$\widehat{W} \cong G \ltimes W.$$

Hence $G \cong \widehat{L}/L$ and is a finite abelian group. If R is reduced, it can be proved that $G = \{g_i, i \in \hat{I}\}$ if $g_0 = 1$, and, for every $i \in I_0$, $g_i = t_{\lambda_i}\mathbf{w}^0_{\lambda_i}\mathbf{w}_0$, where $\mathbf{w}_0$ and $\mathbf{w}^0_{\lambda_i}$ are the longest elements of $\mathbf{W}$ and $\mathbf{W}_{\lambda_i} = \{\mathbf{w} \in \mathbf{W} : \mathbf{w}\lambda_i = \lambda_i\}$ respectively. A proof of this property can be found in [12]. Obviously, if R is non-reduced, then G is trivial.

We extend to $\widehat{W}$ the definition of q_w given in Sect. 2.5 for $w \in W$, by setting

$$q_{\hat{w}} = q_w \quad \text{if } \hat{w} = wg,$$

where $w \in W$ and $g \in G$. In particular, for each $\lambda \in \widehat{L}$, $q_{t_\lambda} = q_{u_\lambda}$ if $t_\lambda = u_\lambda g$.

2.12 *Automorphisms of* $\mathbb{A}$ *and D*

As usual, an automorphism of $\mathbb{A}$ is a bijection φ on $\mathbb{V}$ that maps chambers to chambers, with the property that $\varphi(C)$ and $\varphi(C')$ are adjacent if and only if C and C' are adjacent. If D denotes the Coxeter graph of W, then an automorphism of D is a permutation σ of I such that $m_{\sigma(i),\sigma(j)} = m_{ij}$, for every $i, j \in I$. Aut$(\mathbb{A})$ and Aut(D) denote the automorphism groups of $\mathbb{A}$ and D, respectively. It can be proved (see for instance [12]) that, for every $\varphi \in$ Aut$(\mathbb{A})$, there exists $\sigma \in$ Aut(D) such that, for every $X \in \mathscr{V}(\mathbb{A})$,

$$\tau \circ \varphi(X) = \sigma \circ \tau(X),$$

and $\varphi(C) \sim_{\sigma(i)} \varphi(C')$ if $C \sim_i C'$.

Obviously $\mathbf{W}$, W and $\widehat{W}$ can be seen as subgroups of Aut($\mathbb{A}$) such that $\mathbf{W} \leqslant W \leqslant \widehat{W} \leqslant \mathrm{Aut}(\mathbb{A})$ (in some cases $\widehat{W}$ is a proper subgroup). Consider in particular the finite abelian group G and, for every $i \in \hat{I}$, denote by σ_i the automorphism of D such that $\tau \circ g_i = \sigma_i \circ \tau$. Then $\sigma_i(0) = i$ for every $i \in \hat{I}$. Hence every σ_i, $i \in \hat{I}$ is called *type-rotating*, and we set

$$\mathrm{Aut}_{tr}(D) = \{\sigma_i, i \in \hat{I}\}.$$

In particular $\sigma_0 = 1$. Note that $\mathrm{Aut}(D) = \mathrm{Aut}(D_0) \ltimes \mathrm{Aut}_{tr}(D)$, where D_0 is the Coxeter graph of $\mathbf{W}$, and $\mathrm{Aut}_{tr}(D)$ acts simply transitively on $\hat{I}$. Since each $w \in W$ is type-preserving, it corresponds to the element $\sigma_0 = 1$ of $\mathrm{Aut}_{tr}(D)$; actually W is the subgroup of all type-preserving automorphisms of $\mathbb{A}$. Keeping in mind that $\widehat{W} \cong G \ltimes W$, we call *type-rotating* automorphism of $\mathbb{A}$ any element of $\widehat{W}$.

The group $\mathrm{Aut}_{tr}(D)$ acts on W as follows: for every $\sigma \in \mathrm{Aut}_{tr}(D)$ and $w = s_{i_1} \cdots s_{i_k} \in W$,

$$\sigma(w) = s_{\sigma(i_1)} \cdots s_{\sigma(i_k)}.$$

In particular, for every $i \in \hat{I}$, we have $\mathbf{W}_i = \sigma_i(\mathbf{W})$.

Consider now the map

$$\iota(\mu) = -\mathbf{w}_0(\mu), \quad \text{for every } \mu \in \mathbb{A}.$$

Since the map $\mu \mapsto -\mu$ is an automorphism of $\mathbb{A}$, then $\iota \in \mathrm{Aut}(\mathbb{A})$; moreover $\iota^2 = 1$ and $\iota(\mathbb{Q}_0) = \mathbb{Q}_0$. Therefore either ι is the identity or it permutes the walls of the sector $\mathbb{Q}_0$. Since the identity is the unique element of $\mathbf{W}$ which fixes the sector $\mathbb{Q}_0$, by virtue of the simple transitivity of $\mathbf{W}$ on the sectors based at 0, it follows that ι belongs to $\mathbf{W}$ only when is the identity. This happens when the map $\mu \mapsto -\mu$ belongs to $\mathbf{W}$, that is, when $\mathbf{w}_0 = -1$. Hence, if we consider the automorphism $\sigma_\star$ of D induced by ι, then in general $\sigma_\star$ is not an element of $\mathrm{Aut}_{tr}(D)$, but $\sigma_\star \in \mathrm{Aut}_{tr}(D)$ if and only if $\sigma_\star = 1$. Moreover, when $\sigma_\star \neq 1$, then it belongs to $\mathrm{Aut}(D_0)$. On the other hand, $\mathrm{Aut}(D_0)$ is non-trivial only for a root system of type A_l ($l \geqslant 2$), D_l ($l \geqslant 4$) and E_6. Hence, apart these three cases, ι is always the identity, or equivalently, the map $\mu \mapsto -\mu$ belongs to $\mathbf{W}$.

Simple computations allow to determine if ι is trivial or not for a Dynkin diagram D_0 of type A_l ($l \geqslant 2$), D_l ($l \geqslant 4$) and E_6. The results are listed in the following proposition.

Proposition 2.1 *Let R be an irreducible root system*:

(i) *if R has type A_l ($l \geqslant 2$), then ι induces the unique non-trivial automorphism of the diagram D_0*;
(ii) *if R has type D_l ($l \geqslant 4$), then ι is the identity for n even and it induces the unique non-trivial automorphism of the diagram D_0 for n odd*;
(iii) *if R has type E_6, then ι induces the unique non-trivial automorphism of the diagram D_0.*

For every $\mu \in \mathscr{V}_{sp}(\mathbb{A})$, let $\mu^\star = \iota(\mu)$: then $\mu^\star \in \overline{\mathbb{Q}_0}$ for each $\mu \in \overline{\mathbb{Q}_0}$.

2.13 Affine Buildings of Type $\widetilde{X}_n$

Let Δ be an affine building; we assume Δ is irreducible, locally finite, regular and we denote by $\{q_i\}_{i\in I}$ its parameter system. By definition, there is a Coxeter group W canonically associated to Δ and W is an affine reflection group, which can be interpreted as the affine Weyl group of a (irreducible) root system R. Hence there is a root system R canonically associated to each (irreducible, locally finite, regular) affine building. In most cases the choice of R is straightforward, since in general different root systems have different affine Weyl group.

The only exceptions to this rule are the root systems of type C_n and BC_n, which have the same affine Weyl group. So, when the group W associated to the building is the affine Weyl group of the root systems of type C_n and BC_n, we have to choose the root system. We assume to operate this choice according to the parameter system of the building. Actually, we choose R to ensure that in each case the group $\mathrm{Aut}_{tr}(D)$ preserves the parameter system of the building, i.e., so as to have, for each $\sigma \in \mathrm{Aut}_{tr}(D)$, the identity $q_{\sigma(i)} = q_i$ for all $i \in I$. In the cases $R = C_n$ or BC_n, one has $q_1 = q_2 = \cdots = q_{n-1}$, but in general $q_0 \neq q_1 \neq q_n$. On the other hand, if $R = C_n$, then $\mathrm{Aut}_{tr}(D) = \{1, \sigma\}$, while, if $R = BC_n$, then $\mathrm{Aut}_{tr}(D) = \{1\}$. Thus, if $R = C_n$, the condition $q_{\sigma(0)} = q_0$ implies $q_n = q_0$, while, if $R = BC_n$, q_0 and q_n can have different values.

Keeping in mind the above choice and the classification of root systems, we shall say that

(i) Δ is an affine building of type $\widetilde{X}_n$, if R has type X_n, in the following cases:

$$X_n = A_n\ (n \geqslant 2),\quad B_n\ (n \geqslant 3),\quad D_n\ (n \geqslant 4),\quad E_n\ (n = 6, 7, 8),\quad F_4,\quad G_2;$$

(ii) Δ is an affine building of type
  a. $\widetilde{A}_1$, associated to a root system of type A_1, if $q_0 = q_1$ (homogeneous tree);
  b. $\widetilde{BC}_1$, associated to a root system of type BC_1, if $q_0 \neq q_1$ (semi-homogeneous tree);

(iii) Δ is an affine building of type
  a. $\widetilde{C}_n$, $n \geqslant 2$, associated to a root system of type C_n, if $q_0 = q_n$;
  b. $\widetilde{BC}_n$, $n \geqslant 2$, associated to a root system of type BC_n, if $q_0 \neq q_n$.

We refer to Appendix of [12] for the classification of all irreducible, locally finite, regular affine buildings, in terms of diagram and parameter system.

In each case $\mathrm{Aut}_{tr}(D)$ preserves the parameter system of the building. Actually, if we define

$$\mathrm{Aut}_q(D) = \left\{\sigma \in \mathrm{Aut}(D) : q_{\sigma(i)} = q_i, i \in I\right\},$$

then in each case $\mathrm{Aut}_{tr}(D) \cup \{\sigma_\star\} \subset \mathrm{Aut}_q(D)$.

2.14 Subgroups of G

We wish to determine the subsets of the set $\hat{I}$ of special types corresponding to sublattices of $\widehat{L}$. In order to solve this problem we have to determine all the subgroups of the finite group $G = \widehat{L}/L$ of order **f**. We only consider buildings of type $\widetilde{A}_n$, $\widetilde{D}_n$ and $\widetilde{E}_6$, as only in these cases **f** is greater than 2 and so there can be proper subgroups of $\widehat{L}/L$. Since the order of a proper subgroup of a finite group must be a divisor of the order of the group, then in the cases $\widetilde{E}_6$ and $\widetilde{A}_n$, n a prime number, there is no proper subgroup of $\widehat{L}/L$. So we have to consider only the cases of $\widetilde{A}_n$, if n is a composite number, and of $\widetilde{D}_n$. Direct computations lead to The following results.

Proposition 2.2 *Let Δ be a building of type $\widetilde{D}_n$. Then*

(i) *if n is even, G has three subgroups of order two, namely*

$$G_{0,1} = \langle g_0, g_1\rangle, \qquad G_{0,n-1} = \langle g_0, g_{n-1}\rangle \quad \textit{and} \quad G_{0,n} = \langle g_0, g_n\rangle,$$

that correspond to types $\{0, 1\}$, $\{0, n-1\}$ *and* $\{0, n\}$ *respectively*;

(ii) *if n is odd, then $G_{0,1} = \langle g_0, g_1\rangle$ is the unique subgroup of order two of G, and it corresponds to the types* $\{0, 1\}$.

Proposition 2.3 *Let Δ be a building of type $\widetilde{A}_n$; assume $n = lm$ for some $l, m \in \mathbb{Z}$, $1 < l, m < n$. Then $\{g_0, g_l, g_{2l}, \dots, g_{(m-1)l}\}$ generate the unique subgroup of order m in G.*

Proposition 2.2 implies that, for a building of type $\widetilde{D}_n$, the vertices of $\mathbb{A}$ of types 0 and 1 form a sublattice of $\widehat{L}$, for every n; moreover, when n is even, also the vertices of types $\{0, n-1\}$ and the vertices of type $\{0, n\}$ form a sublattice of $\widehat{L}$. Besides, the types $\{n-1, n\}$ do not correspond to a subgroup of order two in $\widehat{L}/L$, but to its complement; this means that the vertices of $\mathbb{A}$ of types $n-1$ and n form an affine lattice which does not contain the origin 0. The same is true, when n is even, for the types $\{1, n-1\}$ and $\{1, n\}$.

Proposition 2.3 implies that the vertices of $\mathbb{A}$ of types $\{0, l, 2l, \dots, (m-1)l\}$ form a sublattice of $\widehat{L}$, whereas, for $0 < j < l$, the types $\{j, j+l, j+2l, \dots, j+(m-1)l\}$, do not correspond to any subgroup of order m in $\widehat{L}/L$, but to a lateral of this subgroup. This means that, for $0 < j < l$, the vertices of $\mathbb{A}$ of types $\{j, j+l, j+2l, \dots, j+(m-1)l\}$ form an affine lattice which does not contain the origin 0.

2.15 Geometric Realization of an Affine Building

Let Δ be any affine building of type $\widetilde{X}_n$. The affine Coxeter complex $\mathbb{A}$ associated to W is called the *fundamental apartment* of the building. By definition, each apartment $\mathscr{A}$ of Δ is isomorphic to $\mathbb{A}$ and hence it can be regarded as a Euclidean space, tessellated by a family of affine hyperplanes isomorphic to the family $\mathscr{H}$. Moreover every such isomorphism is type-preserving or type-rotating. If $\psi : \mathscr{A} \to \mathbb{A}$ is any type-preserving isomorphism, then, for each $\hat{w} \in \widehat{W}$, $\psi' = \hat{w}\psi$ is a type-rotating isomorphism and, for every vertex x of type i, the type of $\psi'(x)$ is $\sigma_j(i)$ if $\hat{w} = wg_j$. Each type-rotating isomorphism $\psi' : \mathscr{A} \to \mathbb{A}$ is obtained in this way.

For any apartment $\mathscr{A}$, denote by $\mathscr{H}(\mathscr{A})$ the family of all hyperplanes h of $\mathscr{A}$. If $\psi : \mathscr{A} \to \mathbb{A}$ is any type-rotating isomorphism, we set $h = h_\alpha^k$ if $\psi(h) = H_\alpha^k$. Obviously k and α depend on ψ.

Denote by $\mathscr{V}(\Delta)$ the set of all vertices of the building and by $\mathscr{V}_i(\Delta)$ the set of all type i vertices of Δ, for each $i \in I$.

There is a natural way to extend to Δ the definition of special vertices given in $\mathbb{A}$; we call special each vertex x of Δ such that its image on $\mathbb{A}$, under any type-preserving isomorphism between any apartment containing x and the fundamental apartment, is a special vertex of $\mathbb{A}$. Recall that all types are special for a building of type $\widetilde{A}_n$. Furthermore four, three and only one special type occur for a building of type $\widetilde{D}_n$, $\widetilde{E}_6$ and $\widetilde{G}_2$ respectively. In all other cases the special types are two. We denote by $\mathscr{V}_{sp}(\Delta)$ the set of all special vertices of Δ.

Finally, let $\widehat{\mathscr{V}}(\Delta)$ be the set of all vertices of type $i \in \hat{I}$, that is, the set of all vertices x whose image on $\mathbb{A}$, under any isomorphism type-preserving between any apartment containing x and the fundamental apartment, belongs to $\widehat{L}$. It is obvious that $\widehat{\mathscr{V}}(\Delta) = \mathscr{V}_{sp}(\Delta)$ if Δ is reduced, while $\widehat{\mathscr{V}}(\Delta) = \mathscr{V}_0(\Delta)$ if Δ is not reduced. We always refer to vertices of $\widehat{\mathscr{V}}(\Delta)$.

Recall that, for every pair of chambers $c, d \in \mathscr{C}(\Delta)$, there exists a minimal gallery $\gamma(c,d)$ from c to d, lying in any apartment containing both chambers. The type of $\gamma(c,d)$ is $f = i_1 \cdots i_k$ if $\delta(c,d) = w_f$. If $\{q_i\}_{i \in I}$ is the parameter system of the building, for every $c \in \mathscr{C}(\Delta)$ and $w \in W$, we have $|\mathscr{C}_w(c)| = q_w$, if $\mathscr{C}_w(c) = \{d \in \mathscr{C}(\Delta) : \delta(c,d) = w\}$.

Analogously, given a vertex $x \in \widehat{\mathscr{V}}(\Delta)$ and a chamber d, there exists a minimal gallery $\gamma(x,d)$ from x to d, lying in any apartment containing x and d. If c is the chambers of $\gamma(x,d)$ containing x, then the type of this gallery is $f = i_1 \cdots i_k$ if $\delta(c,d) = w_f$. Let $\delta(x,d) = \delta(c,d)$, and, for every $x \in \widehat{\mathscr{V}}(\Delta)$ and $w \in W$,

$$\mathscr{C}_w(x) = \{d \in \mathscr{C}(\Delta) : \delta(x,d) = w\}.$$

For every $x \in \widehat{\mathscr{V}}(\Delta)$, denote by $\mathscr{C}(x)$ the set of all chambers containing x. Then $\mathscr{C}_w(x) = \bigcup_{c \in \mathscr{C}(x)} \mathscr{C}_w(c)$, a disjoint union. Observe that, for every x of type $i \in \hat{I}$ and any fixed chamber c containing x,

$$\mathscr{C}(x) = \{c' \in \mathscr{C}(\Delta) : \delta(c,c') = w \in \mathbf{W}_i\},$$

where $\mathbf{W}_i = \sigma_i(\mathbf{W})$ is the stabilizer of the type i vertex of C_0. Hence the cardinality of the set $\mathscr{C}(x)$ is the Poincaré polynomial $\mathbf{W}_i(q)$ of $\mathbf{W}_i$. On the other hand, $\mathbf{W}_i(q) = \mathbf{W}_{\sigma_i(0)}(q) = \mathbf{W}(q)$. Thus, in each case,

$$|\mathscr{C}(x)| = \mathbf{W}(q).$$

Therefore, for every $x \in \mathscr{V}_{sp}(\Delta)$ and $w \in W$, the cardinality of the set $\mathscr{C}_w(x)$ does not depend on x and

$$|\mathscr{C}_w(x)| = \mathbf{W}(q)q_w.$$

For any pair of facets $\mathscr{F}_1, \mathscr{F}_2$ of the building Δ, there exists an apartment $\mathscr{A}(\mathscr{F}_1, \mathscr{F}_2)$ containing them. We call *convex hull* of $\{\mathscr{F}_1, \mathscr{F}_2\}$ the minimal convex region $[\mathscr{F}_1, \mathscr{F}_2]$ delimited by hyperplanes of $\mathscr{A}(\mathscr{F}_1, \mathscr{F}_2)$ that contains $\{\mathscr{F}_1, \mathscr{F}_2\}$.

Given two special vertices x, y, there exists a minimal gallery $\gamma(x, y)$ from x to y, lying in any apartment $\mathscr{A}(x, y)$ that contains x and y. If c and d are the chambers of $\gamma(x, y)$ containing x and y respectively, and $\delta(c, d) = w_f$, then the type of this gallery is $f = i_1 \cdots i_k$. Moreover, denoting by φ any type-preserving isomorphism from $\mathscr{A}(x, y)$ onto $\mathbb{A}$, we define the *shape* of y with respect to x as

$$\sigma(x, y) = \sigma(X, Y) \quad \text{if } X = \varphi(x), Y = \varphi(y).$$

Hence, by definition of $\sigma(X, Y)$, the shape $\sigma(x, y)$ is an element of $\widehat{L}^+$ and, if $\sigma(x, y) = \lambda$, there exists a type-rotating isomorphism $\psi : \mathscr{A}(x, y) \to \mathbb{A}$ such that $\psi(x) = 0$ and $\psi(y) = \lambda$.

For every vertex $x \in \widehat{\mathscr{V}}(\Delta)$ and every $\lambda \in \widehat{L}^+$, we define

$$\mathscr{V}_\lambda(x) = \left\{ y \in \widehat{\mathscr{V}}(\Delta) : \sigma(x, y) = \lambda \right\}.$$

It is easy to prove that, for every $x \in \widehat{\mathscr{V}}(\Delta)$, we have $\widehat{\mathscr{V}}(\Delta) = \bigcup_{\lambda \in \widehat{L}^+} \mathscr{V}_\lambda(x)$, a disjoint union.

The following proposition provides a formula for the cardinality of the set $\mathscr{V}_\lambda(x)$.

Proposition 2.4 *Let $x \in \widehat{\mathscr{V}}(\Delta)$ and $\lambda \in \widehat{L}^+$. If $\tau(x) = i$, $\tau(X_\lambda) = l$ and $j = \sigma_i(l)$, then*

$$\left| \mathscr{V}_\lambda(x) \right| = \frac{1}{\mathbf{W}(q)} \sum_{w \in \mathbf{W} w_\lambda \mathbf{W}_j} q_w = \frac{\mathbf{W}(q)}{\mathbf{W}_\lambda(q)} q_{w_\lambda}.$$

In particular $|\mathscr{V}_\lambda(x)| = \mathbf{W}(q) q_{w_\lambda}$ for $\lambda \in L^{++}$.

Proof For every chamber c of Δ and for every $i \in I$, denote by $v_i(c)$ the vertex of type i of c. Then

$$\mathscr{V}_\lambda(x) = \left\{ y = v_j(d), d \in \mathscr{C}(\Delta) : \delta(x, d) = \sigma_i(w_\lambda) \right\}.$$

If we define

$$\mathscr{C}_\lambda(x) = \left\{ d \in \mathscr{C}(\Delta) : v_j(d) \in \mathscr{V}_\lambda(x) \right\},$$

then it is immediate to note that, for each $y \in \mathscr{V}_\lambda(x)$, there are $\mathbf{W}(q)$ chambers in $\mathscr{C}_\lambda(x)$ which contain y. Therefore $|\mathscr{C}_\lambda(x)| = \mathbf{W}(q) |\mathscr{V}_\lambda(x)|$. On the other hand, for any chamber c in the set $\mathscr{C}(x)$, it can be proved that, as disjoint union,

$$\mathscr{C}_\lambda(x) = \bigcup_{w \in \mathbf{W}_i \sigma_i(w_\lambda) \mathbf{W}_j} \mathscr{C}_w(c).$$

This implies that $|\mathscr{C}_\lambda(x)| = \sum_{w\in\mathbf{W}_i\sigma_i(w_\lambda)\mathbf{W}_j} |\mathscr{C}_w(c)|$. Moreover, since $\mathbf{W}_i\sigma_i(w_\lambda) \times \mathbf{W}_j = \sigma_i(\mathbf{W}w_\lambda\mathbf{W}_j)$ and $q_{\sigma_i(w)} = q_w$, it follows that

$$\left|\mathscr{C}_\lambda(x)\right| = \sum_{w\in\mathbf{W}w_\lambda\mathbf{W}_j} q_w.$$

This proves the first formula.

For $y \in \mathscr{V}_\lambda(x)$, a minimal gallery γ from x to y starts with a chamber $c \in \mathscr{C}(x)$ and is equal to $\gamma(c, y)$. This gallery has type $\sigma_i(f_\lambda)$ if f_λ is the type of the gallery $\gamma(C_0, C_\lambda)$. Therefore there are $|\mathscr{C}(x)| = \mathbf{W}(q)$ choices for c and, for each c, q_{w_λ} choices for $\gamma(c, y)$. On the other hand, c is determined by y up to $\mathbf{W}_\lambda(q)$. This yields the last formula. □

Proposition 2.4 shows that $|\mathscr{V}_\lambda(x)|$ does not depend on x. Therefore we can set, for every vertex $x \in \widehat{\mathscr{V}}(\Delta)$,

$$N_\lambda = \left|\mathscr{V}_\lambda(x)\right|.$$

Set $\lambda^\star = \iota(\lambda)$. Then $y \in \mathscr{V}_\lambda(x)$ if and only if $x \in \mathscr{V}_{\lambda^\star}(y)$. Hence $N_\lambda = N_{\lambda^\star}$.

We provide an alternative formula for N_λ, in terms of q_{t_λ}.

Proposition 2.5 *Let* $\lambda \in \widehat{L}^+$. *Then*

$$N_\lambda = \frac{\mathbf{W}(q^{-1})}{\mathbf{W}_\lambda(q^{-1})} q_{t_\lambda}.$$

In particular, for $\lambda \in L^{++}$,

$$N_\lambda = \mathbf{W}\bigl(q^{-1}\bigr) q_{t_\lambda}.$$

Proof For any $x \in \widehat{\mathscr{V}}(\Delta)$ and $y \in \mathscr{V}_\lambda(x)$, denote by c_x and c_y the chambers containing x and y respectively in any minimal gallery connecting x to y. Let

$$\mathscr{C}_{t_\lambda}(x, y) = \bigl\{d \in \mathscr{C}(\Delta) : y \in d, \delta(x, d) = t_\lambda\bigr\}.$$

Then it is easy to check that

$$\mathscr{C}_{t_\lambda}(x, y) = \bigl\{d \in \mathscr{C}(\Delta) : \delta(c_y, d) = w_j^0 w_{j,\lambda}^0\bigr\},$$

if w_j^0 and $w_{j,\lambda}^0$ are the longest elements of $\mathbf{W}_j$ and $\mathbf{W}_{j,\lambda} = \{w \in \mathbf{W}_j, : w\lambda = \lambda\}$ respectively. Therefore,

$$\left|\mathscr{C}_{t_\lambda}(x, y)\right| = q_{w_j^0 w_{j,\lambda}^0} = q_{w_j^0} q_{w_{j,\lambda}^0}^{-1} = q_{\mathbf{w}_0} q_{\mathbf{w}_\lambda^0}^{-1}$$

and

$$q_{t_\lambda} = q_{w_\lambda} q_{\mathbf{w}_0} q_{\mathbf{w}_\lambda^0}^{-1}.$$

Thus

$$N_\lambda = \frac{\mathbf{W}(q)}{\mathbf{W}_\lambda(q)} q_{\mathbf{w}_0}^{-1} q_{\mathbf{w}_\lambda^0} q_{t_\lambda}.$$

Since $\mathbf{W}(q) = q_{\mathbf{w}_0}\mathbf{W}(q^{-1})$ and $\mathbf{W}_\lambda(q) = q_{\mathbf{w}_\lambda^0}\mathbf{W}_\lambda(q^{-1})$ [12], we conclude that

$$N_\lambda = \frac{\mathbf{W}(q^{-1})}{\mathbf{W}_\lambda(q^{-1})} q_{t_\lambda}.$$

In particular, for $\lambda \in L^{++}$,

$$N_\lambda = \mathbf{W}\big(q^{-1}\big) q_{t_\lambda}. \qquad \square$$

2.16 Parameter System of *R*

Let Δ be a building of type $\widetilde{X}_n$ and let $\{q_i\}_{i\in I}$ the parameter system of Δ. As we said in Sect. 2.13, $q_{\sigma(i)} = q_i$ for every $i \in I$ and every $\sigma \in \mathrm{Aut}_{tr}(D)$. Moreover we notice that $q_i = q_j$ if there exists an hyperplane h on any apartment of the building which contains two panels π_i and π_j of co-type i and j respectively. Hence for every hyperplane h of the building we may define $q_h = q_i$ if there is a panel of co-type i lying in h. Notice that if h and h' are two hyperplanes of the building, lying in $\mathscr{A}$ and $\mathscr{A}'$ respectively, and there exists a type-rotating isomorphism $\psi : \mathscr{A} \to \mathscr{A}'$ such that $h' = \psi(h)$, then $q_{h'} = q_h$; actually, if π_i is a panel lying in h, then h' contains a panel of co-type $\sigma(i)$ for some $\sigma \in \mathrm{Aut}_{tr}(D)$.

Consider any apartment $\mathscr{A}$ of Δ and the set $\mathscr{H}(\mathscr{A})$ of all the hyperplanes of $\mathscr{A}$. Let $\psi : \mathscr{A} \to \mathbb{A}$ any type-rotating isomorphism. Following notation of Sect. 2.15, we set $h = h_\alpha^k$ if $\psi(h) = H_\alpha^k$, for any positive root α and any $k \in \mathbb{Z}$. In this case we define

$$q_{\alpha,k} = q_h.$$

This definition is independent of the particular choice of $\mathscr{A}$ and ψ. Actually, if $\psi' : \mathscr{A}' \to \mathbb{A}$ is another type-rotating isomorphism and $\psi(h) = \psi'(h') = H_\alpha^k$, then $q_{h'} = q_h$, since $\psi'^{-1}\psi$ is a type-rotating automorphism mapping h onto h'.

If R is reduced, it is easy to check that $q_{\alpha,k} = q_{\alpha',k'}$ if $H_{\alpha'}^{k'} = \hat{w}(H_\alpha^k)$, for some $\hat{w} \in \widehat{W}$. Indeed $q_{h'} = q_h$ if $\psi(h) = H_\alpha^k$ and $\psi(h') = H_{\alpha'}^{k'}$ for any $\psi : \mathscr{A} \to \mathbb{A}$. In particular $q_{\alpha,0} = q_{\alpha',0}$ if $\alpha' = \mathbf{w}(\alpha)$ for some $\mathbf{w} \in \mathbf{W}$ and, for every $\alpha \in R^+$, $q_{\alpha,k} = q_{\alpha,0}$ for every $k \in \mathbb{Z}$. Moreover $q_{\alpha_i,0} = q_i$, $i = 1, \ldots, n$ and $q_{\alpha_0,1} = q_0$. These properties suggest to define, for every $\alpha \in R^+$,

$$q_\alpha = q_{\alpha,k}, \qquad \text{for every } k \in \mathbb{Z}.$$

Then $q_{\alpha_i} = q_i$ for every $i \in I$; moreover, for every $\alpha \in R^+$, $q_\alpha = q_{\alpha_i}$ if $\alpha = \mathbf{w}\alpha_i$ for some $\mathbf{w} \in \mathbf{W}$. Hence $q_\alpha = q_{\alpha_i}$ if $|\alpha| = |\alpha_i|$. It turns out that, if all roots have the same length (as in the case of R of type A_n), then $q_i = q$ for every $i \in I$ and $q_\alpha = q$

for every $\alpha \in R$. Moreover, if R contains long and short roots, then $q_i = q$ if α_i is long, and $q_i = p$ if α is short; so $q_\alpha = q$, for all long α, and $q_\beta = p$, for all short β.

Consider now the case of a non-reduced root system of type BC_n. Since $\widehat{L} = L$ and $\widehat{W} = W$, then every isomorphism of an apartment $\mathscr{A}$ onto $\mathbb{A}$ is type-preserving and $q_{\alpha,k} = q_{\alpha',k'}$ if $H_{\alpha'}^{k'} = w(H_\alpha^k)$ for some $w \in W$. Hence it is easy to prove that, for all $k \in \mathbb{Z}$,

$$q_{\alpha,2k+1} = q_{\alpha,1} = q_{\alpha_0,1} \quad \text{for every } \alpha \in R_1,$$

$$q_{\alpha,k} = q_{\alpha,0} = q_{\alpha_n,0} \quad \text{for every } \alpha \in R_2,$$

$$q_{\alpha,k} = q_{\alpha,0} = q_{\alpha_i,0}, \quad i = 1, \dots, n-1, \text{ if } \alpha \in R_0 \text{ and } \alpha = \mathbf{w}\alpha_i \text{ for some } \mathbf{w} \in \mathbf{W}.$$

Moreover

$$q_{\alpha_0,1} = q_1, \qquad q_{\alpha_i,0} = q_0, \quad \text{for every } i = 1, \dots, n-1 \text{ and } q_{\alpha_n,0} = q_n.$$

Hence, setting

$$q_\alpha = \begin{cases} q_{\alpha,2k+1} & \text{for every } \alpha \in R_1, \text{for every } k \in \mathbb{Z}, \\ q_{\alpha,k}, & \text{for every } \alpha \in R_2 \cup R_0, \text{for every } k \in \mathbb{Z}, \end{cases}$$

we have

$$q_\alpha = \begin{cases} q_1 & \text{for every } \alpha \in R_1, \\ q_0 & \text{for every } \alpha \in R_0, \\ q_n & \text{for every } \alpha \in R_2. \end{cases}$$

For the sake of simplicity, set $q_1 = p$, $q_0 = q$, $q_n = r$. In each case it is convenient to extend the definition of q_α by setting $q_\alpha = 1$ if $\alpha \notin R$. Thus, $q_\alpha = p$, $q_{\alpha/2} = r$ if $\alpha \in R_1$, $q_\alpha = q$, $q_{\alpha/2} = 1$ if $\alpha \in R_0$, and $q_\alpha = r$, $q_{\alpha/2} = 1$ if $\alpha \in R_2$.

Let us give the following useful alternative characterization of q_{t_λ}, for every $\lambda \in \widehat{L}^+$.

Proposition 2.6 *For every $\lambda \in \widehat{L}^+$, then*

$$q_{t_\lambda} = \prod_{\alpha \in R^+} q_\alpha^{\langle \lambda, \alpha \rangle} q_{2\alpha}^{-\langle \lambda, \alpha \rangle}.$$

Proof Recall that q_{u_λ} denotes the number of chambers c' connected to any chamber c by a gallery of type u_λ. Moreover $q_{t_\lambda} = q_{u_\lambda} = q_{i_1} \cdots q_{i_r}$ if $t_\lambda = u_\lambda g_l$ and $u_\lambda = s_{i_1} \cdots s_{i_r}$.

In the building Δ, fix two chambers c, c' such that $\delta(c, c') = u_\lambda$. Denote by $\mathscr{A}$ any apartment that contains c, c' (and hence the gallery $\gamma(c, c')$ of type u_λ), and consider the isomorphism $\psi : \mathscr{A} \to \mathbb{A}$ such that $\psi(c) = C_0$. Through this isomorphism, the chamber c' maps to the chamber $u_\lambda(C_0)$, lying in $\mathbb{Q}_0$. For every $i_1, \dots, i_r$, the panel π_{i_j} of the gallery belongs to a hyperplane h of $\mathscr{A}$ such that $\psi(h) = H_\alpha^j$ for some $\alpha \in R^+$ and $j \in \mathbb{Z}$. Therefore it follows that

$$q_{t_\lambda} = \prod_{\alpha \in R^+} q_\alpha^{k_\alpha},$$

where, for each $\alpha \in R^+$, k_α denotes the number of hyperplanes in $\mathscr{H}(\alpha)$ separating C_0 and $u_\lambda(C_0)$. Since $v_l(u_\lambda(C_0)) = \lambda$, we notice that $k_\alpha = \langle \lambda, \alpha \rangle$ when $\alpha/2 \notin R$, and $k_\alpha = \langle \lambda, \alpha/2 \rangle$ otherwise. This proves the statement. $\square$

Corollary 2.7 *Let $\lambda \in \widehat{L}^+$. Then*

$$N_\lambda = \frac{\mathbf{W}(q^{-1})}{\mathbf{W}_\lambda(q^{-1})} \prod_{\alpha \in R^+} q_\alpha^{\langle \lambda, \alpha \rangle} q_{2\alpha}^{-\langle \lambda, \alpha \rangle}.$$

In particular, if $\lambda \in \widehat{L}^{++}$, we have

$$N_\lambda = \mathbf{W}\big(q^{-1}\big) \prod_{\alpha \in R^+} q_\alpha^{\langle \lambda, \alpha \rangle} q_{2\alpha}^{-\langle \lambda, \alpha \rangle}.$$

2.17 The Algebra $\mathscr{H}(\mathscr{C})$

Denote by $\mathscr{L}(\mathscr{C})$ the space of all finitely supported functions on $\mathscr{C} = \mathscr{C}(\Delta)$. Each function $f \in \mathscr{L}(\mathscr{C})$ can be written uniquely as $f = \sum_c f(c)\mathbb{1}_c$, where, for each chamber $c \in \mathscr{C}(\Delta)$,

$$\mathbb{1}_c\big(c'\big) = \begin{cases} 1, & c' = c, \\ 0, & c' \neq c. \end{cases}$$

For each $w \in W$, we define

$$T_w \mathbb{1}_c = \sum_{\delta(c',c)=w} \mathbb{1}_{c'}.$$

The operator T_w may be extended by linearity to the space $\mathscr{L}(\mathscr{C})$, by setting $T_w f = \sum_c f(c) T_w \mathbb{1}_c$ if $f = \sum_c f(c)\mathbb{1}_c$. It is easy to prove that, for every c,

$$T_w f(c) = \sum_{\delta(c,c')=w} f\big(c'\big).$$

Indeed,

$$T_w f(c) = \langle T_w f, \mathbb{1}_c \rangle = \sum_{c'} f\big(c'\big) \sum_{\delta(c'',c')=w} \langle \mathbb{1}_{c''}, \mathbb{1}_c \rangle = \sum_{\delta(c,c')=w} f\big(c'\big),$$

since we can choose $c'' = c$ in the sum only in the case $\delta(c, c') = w$ and $\langle \mathbb{1}_{c''}, \mathbb{1}_c \rangle = 0$ for $c'' \neq c$.

We denote by $\mathscr{H}(\mathscr{C})$ the linear span of $\{T_w, w \in W\}$. We shall prove that $\mathscr{H}(\mathscr{C})$ is in fact an algebra.

Lemma 2.8 *Let S be the finite set of generators of W. For every $s \in S$, let $q_s = q_\alpha$ when $s = s_\alpha$. Then*

$$T_s^2 = q_s I + (q_s - 1)T_s.$$

Proof Fix $s \in S$. Then, for every chamber c,

$$T_s^2 \mathbb{1}_c = \sum_{\delta(c',c)=s} T_s \mathbb{1}_{c'} = \sum_{\delta(c',c)=s} \sum_{\delta(c'',c')=s} \mathbb{1}_{c''} = \sum_{\delta(c',c)=s} \left(\mathbb{1}_c + \sum_{\delta(c'',c')=s,\, c''\neq c} \mathbb{1}_{c''}, \right).$$

Since q_s is the number of chambers c' such that $\delta(c, c') = \delta(c', c) = s$, we conclude that

$$T_s^2 = q_s \mathbb{1}_c + (q_s - 1) \sum_{\delta(c',c)=s} \mathbb{1}_{c'} = q_s I + (q_s - 1)T_s. \qquad \square$$

Proposition 2.9 *For every $w \in W$ and $s \in S$,*

$$T_w T_s = \begin{cases} T_{ws} & \text{if } |ws| = |w| + 1, \\ q_s T_{ws} + (q_s - 1)T_w & \text{if } |ws| = |w| - 1. \end{cases}$$

Proof For each function $f \in \mathscr{L}(\mathscr{C})$, and each chamber c, we have by definition

$$(T_w T_s) f(c) = \sum_{\delta(c,c')=w} \sum_{\delta(c',c'')=s} f\big(c''\big) \quad \text{and} \quad T_{ws} f(c) = \sum_{\delta(c,\tilde{c})=ws} f(\tilde{c}).$$

If $|ws| = |w| + 1$, then, for every $\tilde{c}$, there exists c' such that $\delta(c, c') = w$ and $\delta(c', \tilde{c}) = s$; hence $\mathscr{C}_{ws}(c) = \{\tilde{c} : \delta(c, \tilde{c}) = ws\} = \bigcup_{\delta(c,c')=w} \{c'' : \delta(c', c'') = s\}$. Therefore $(T_w T_s) f(c) = T_{ws} f(c)$.

Assume now $|ws| = |w| - 1$ and define $w_1 = ws$. In this case $w = w_1 s$, with $|w_1 s| = |w_1| + 1$. Therefore $T_w = T_{w_1 s} = T_{w_1} T_s$ and, by Lemma 2.8,

$$\begin{aligned} T_w T_s = T_{w_1} T_s^2 &= q_s T_{w_1} + (q_s - 1) T_{w_1} T_s = q_s T_{w_1} + (q_s - 1) T_{w_1 s} \\ &= q_s T_{ws} + (q_s - 1) T_w. \end{aligned} \qquad \square$$

Theorem 2.10 *Let $w_1, w_2 \in W$ for every $w, w_1, w_2 \in W$ there exists $N_w(w_1, w_2)$ such that*

$$T_{w_1} T_{w_2} = \sum_{w \in W} N_w(w_1, w_2) T_w.$$

Moreover, the set $\{w \in W : N_w(w_1, w_2) \neq 0\}$ is finite, for all $w_1, w_2 \in W$.

Proof We proceed by induction on $|w_2|$. If $|w_2| = 1$, then $w_2 = s$ for some $s \in S$ and the identity follows from Proposition 2.9. If $|w_2| = n$ for $n > 1$, write $w_2 = w's$ for some s and w' such that $|w'| = n - 1$. Hence $T_{w_1}T_{w_2} = T_{w_1}T_{w'}T_s$. Now assume that the identity is true for each $k < n$. Then

$$T_{w_1}T_{w_2} = (T_{w_1}T_{w'})T_s = \left(\sum_{w\in W} N_w\big(w_1, w'\big)T_w\right)T_s = \sum_{w\in W} N_w\big(w_1, w'\big)(T_w T_s).$$

Therefore the identity follows from Proposition 2.9. □

Corollary 2.11 *Let $w_1, w_2 \in W$. If $|w_1w_2| = |w_1| + |w_2|$, then $T_{w_1}T_{w_2} = T_{w_1w_2}$.*

Proof If $|w_2| = 1$, the identity follows from Proposition 2.9. If $|w_2| = n$ for $n > 1$, and $w_2 = w's$ for some s and w' such that $|w'| = n - 1$, then $|w_1w'| = |w_1| + |w'|$, and $|w_1w_2| = |w_1w'| + |s|$. Thus, assuming the identity true for each $k < n$, by Proposition 2.9 we have

$$T_{w_1}T_{w_2} = T_{w_1}T_{w'}T_s = T_{w_1w'}T_s = T_{w_1w's} = T_{w_1w_2}.$$ □

Theorem 2.10 shows that $\mathscr{H}(\mathscr{C})$ is an associative algebra, generated by $\{T_s, s \in S\}$. We refer to the numbers $N_w(w_1, w')$ as the *structure constants* of the algebra $\mathscr{H}(\mathscr{C})$. Observe that $\mathscr{H}(\mathscr{C})$ is (up to an isomorphism) the Hecke algebra $\mathscr{H}(q_s, q_s - 1)$ associated to W and S [5, Chap. 7].

We shall need some particular operators of the algebra $\mathscr{H}(\mathscr{C})$. For every $i \in \hat{I}$ and for any chamber c, we set

$$T_i\mathbb{1}_c = \sum_{v_i(c')=v_i(c)} \mathbb{1}_{c'},$$

if, as usual, $v_i(c)$ denotes the vertex of type i lying in c. We extend T_i to the space $\mathscr{L}(\mathscr{C})$ by linearity.

Proposition 2.12 *For every $i \in \hat{I}$, the operator T_i belongs to the algebra $\mathscr{H}(\mathscr{C})$. Moreover $T_i^\star = T_i$.*

Proof We observe that $T_i \in \mathscr{H}(\mathscr{C})$ for every $i \in \hat{I}$, because $T_i = \sum_{w\in W_i} T_w$; actually

$$\big\{c' : v_i\big(c'\big) = v_i(c)\big\} = \bigcup_{w\in W_i} \big\{c' : \delta\big(c, c'\big) = w\big\}.$$

Let us prove that T_i is selfadjoint. For all c_1, c_2,

$$\langle T_i\mathbb{1}_{c_1}, \mathbb{1}_{c_2}\rangle = \sum_{v_i(c')=v_i(c_1)} \langle \mathbb{1}_{c'}, \mathbb{1}_{c_2}\rangle \quad\text{and}\quad \langle \mathbb{1}_{c_1}, T_i\mathbb{1}_{c_2}\rangle = \sum_{v_i(c'')=v_i(c_2)} \langle \mathbb{1}_{c_1}, \mathbb{1}_{c''}\rangle.$$

Notice that $\langle \mathbb{1}_{c'}, \mathbb{1}_{c_2}\rangle \neq 0$ only for $c' = c_2$: we can choose $c' = c_2$ in the set $\{c' : v_i(c') = v_i(c_1)\}$ only if $v_i(c_1) = v_i(c_2)$. Analogously, $\langle \mathbb{1}_{c_1}, \mathbb{1}_{c''}\rangle \neq 0$ only for $c''=c_1$ and we can choose $c''=c_1$ in the set $\{c'' : v_i(c'')=v_i(c_2)\}$ only if $v_i(c_1)=v_i(c_2)$. Therefore

$$\langle T_i \mathbb{1}_{c_1}, \mathbb{1}_{c_2}\rangle = \langle \mathbb{1}_{c_1}, T_i \mathbb{1}_{c_2}\rangle = \begin{cases} 1 & \text{if } v_i(c_1)=v_i(c_2), \\ 0 & \text{if } v_i(c_1)\neq v_i(c_2). \end{cases}$$ □

2.18 Chamber and Vertex Regularity of the Building

For every $w_0, w_1, w_2 \in W$ and every pair of chambers c_1, c_2 such that $\delta(c_1,c_2)=w_0$, consider the set

$$\{c' \in \mathscr{C}(\Delta) : \delta(c_1,c')=w_1, \delta(c_2,c')=w_2\}.$$

We say that the building Δ is *chamber regular* if the cardinality of this set does not depend on the choice of the chambers, but only depends on w_0, w_1, w_2.

Proposition 2.13 *The building Δ is chamber regular.*

Proof Fix $w_0, w_1, w_2 \in W$ and a pair of chambers c_1, c_2 such that $\delta(c_1,c_2)=w_0$. Consider the operator $T_{w_1}T_{w_2^{-1}}$. For any chamber c,

$$(T_{w_1}T_{w_2^{-1}})\mathbb{1}_c = \sum_{\delta(c',c)=w_2^{-1}} \sum_{\delta(c'',c')=w_1} \mathbb{1}_{c''} = \sum_{\delta(c,c')=w_2} \sum_{\delta(c'',c')=w_1} \mathbb{1}_{c''}.$$

Let $c_1, c_2 \in \mathscr{C}(\Delta)$ and assume that $\delta(c_1,c_2)=w_0$. Then

$$\begin{aligned}\big\langle (T_{w_1}T_{w_2^{-1}})\mathbb{1}_{c_2}, \mathbb{1}_{c_1}\big\rangle &= \sum_{\delta(c_2,c')=w_2} \sum_{\delta(c'',c')=w_1} \langle \mathbb{1}_{c''}, \mathbb{1}_{c_1}\rangle \\ &= \big|\{c' : \delta(c_1,c')=w_1, \delta(c_2,c')=w_2\}\big|,\end{aligned}$$

since $\langle \mathbb{1}_{c''}, \mathbb{1}_{c_1}\rangle = 1$ if $c''=c_1$ and $\langle \mathbb{1}_{c''}, \mathbb{1}_{c_1}\rangle = 0$ otherwise. On the other hand, as we have proved in Sect. 2.17, there exist constants $N_w(w_1, w_2^{-1})$, $w \in W$ such that $T_{w_1}T_{w_2^{-1}} = \sum_{w\in W} N_w(w_1, w_2^{-1}) T_w$. Therefore

$$\begin{aligned}\big\langle (T_{w_1}T_{w_2^{-1}})\mathbb{1}_{c_2}, \mathbb{1}_{c_1}\big\rangle &= \sum_{w\in W} N_w\big(w_1, w_2^{-1}\big) \langle T_w \mathbb{1}_{c_2}, \mathbb{1}_{c_1}\rangle \\ &= \sum_{w\in W} N_w\big(w_1, w_2^{-1}\big) \sum_{\delta(d,c_2)=w} \langle \mathbb{1}_d, \mathbb{1}_{c_1}\rangle = N_{w_0}\big(w_1, w_2^{-1}\big),\end{aligned}$$

since $\langle \mathbb{1}_d, \mathbb{1}_{c_1}\rangle \neq 0$ only if $d=c_1$ and this equality is possible only in the case $w=w_0$, as we assumed $\delta(c_1,c_2)=w_0$. Therefore

$$\big|\{c' : \delta(c_1,c')=w_1, \delta(c_2,c')=w_2\}\big| = N_{w_0}\big(w_1, w_2^{-1}\big).$$

This proves the statement. □

By means of the operators T_i defined in Sect. 2.17, we extend the previous result to every set

$$\{c' \in \mathscr{C}(\Delta) : \delta(x, c') = w_1, \delta(c, c') = w_2\}.$$

Proposition 2.14 *Let $w_0, w_1, w_2 \in W$. If $x \in \mathscr{V}_{sp}(\Delta)$ and $c \in \mathscr{C}(\Delta)$ satisfy $\delta(x, c) = w_0$, then*

$$\left|\{c' \in \mathscr{C}(\Delta) : \delta(x, c') = w_1, \delta(c, c') = w_2\}\right|$$

does not depend on x and c, but only on w_0, w_1, w_2 and $i = \tau(x)$.

Proof Let x be a special vertex and let c a chamber; assume $\delta(x, c) = w_0$. This means that $\delta(c_x, c) = w_0$, where c_x is the chamber that contains x in a minimal gallery $\gamma(x, c)$. If $\tau(x) = i$, we have

$$\begin{aligned}
\left\langle (T_{w_1} T_{w_2^{-1}}) \mathbb{1}_c, T_i \mathbb{1}_{c_x} \right\rangle &= \sum_{c'_x : x \in c'_x} \left\langle (T_{w_1} T_{w_2^{-1}}) \mathbb{1}_c, \mathbb{1}_{c'_x} \right\rangle \\
&= \sum_{c'_x : x \in c'_x} \left|\{c' : \delta(c'_x, c') = w_1, \delta(c, c') = w_2\}\right| \\
&= \left|\{c' : \delta(x, c') = w_1, \delta(c, c') = w_2\}\right|.
\end{aligned}$$

On the other hand T_i is a selfadjoint operator in the algebra generated by $\{T_w, w \in W\}$. Hence

$$\left\langle (T_{w_1} T_{w_2^{-1}}) \mathbb{1}_c, T_i \mathbb{1}_{c_x} \right\rangle = \left\langle (T_i T_{w_1} T_{w_2^{-1}}) \mathbb{1}_c, \mathbb{1}_{c_x} \right\rangle$$

and there exist constants $n^i_w(w_1, w_2^{-1})$ such that $T_i T_{w_1} T_{w_2^{-1}} = \sum_{w \in W} n^i_w(w_1, w_2^{-1}) T_w$. Therefore, by the same argument of Proposition 2.13,

$$\left\langle (T_{w_1} T_{w_2^{-1}}) \mathbb{1}_c, T_i \mathbb{1}_{c_x} \right\rangle = \sum_{w \in W} n^i_w(w_1, w_2^{-1}) \langle T_w \mathbb{1}_c, \mathbb{1}_{c_x} \rangle = n^i_{w_0}(w_1, w_2^{-1}).$$

This proves the statement, as

$$\left|\{c' : \delta(x, c') = w_1, \delta(c, c') = w_2\}\right| = n^i_{w_0}(w_1, w_2^{-1}). \qquad \square$$

Corollary 2.15 *Let $\lambda \in \widehat{L}^+$ and $w_1, w_2 \in W$. If $x, y \in \widehat{\mathscr{V}}(\Delta)$, and $\sigma(x, y) = \lambda$, then*

$$\left|\{c' \in \mathscr{C}(\Delta) : \delta(x, c') = w_1, \delta(y, c') = w_2\}\right|$$

does not depend on x and y, but only on λ, w_1, w_2.

For every triple $\lambda, \mu, \nu \in \widehat{L}$ and every pair $x, y \in \widehat{\mathscr{V}}(\Delta)$ such that $\sigma(x, y) = \lambda$, consider the set

$$\{z \in \widehat{\mathscr{V}}(\Delta) : \sigma(x, z) = \mu, \sigma(y, z) = \nu\}.$$

We say that the building Δ is *vertex regular* if the cardinality of this set does not depend on the choice of the vertices, but only depends on λ, μ, ν.

Proposition 2.16 *The building is vertex regular. Moreover*

$$\begin{aligned} &\left|\left\{z \in \widehat{\mathscr{V}}(\Delta) : \sigma(x,z) = \mu, \sigma(y,z) = \nu\right\}\right| \\ &\quad = \left|\left\{z \in \widehat{\mathscr{V}}(\Delta) : \sigma(x,z) = \nu^\star, \sigma(y,z) = \mu^\star\right\}\right|. \end{aligned} \tag{1}$$

Proof Let $\lambda \in \widehat{L}^+$ and $\sigma(x,y) = \lambda$. Consider in W the elements $\sigma_i(w_\mu), \sigma_j(w_\nu)$, where $i = \tau(x)$, $j = \tau(y)$. By Corollary 2.15, the cardinality of the set

$$A = \left\{c' \in \mathscr{C}(\Delta) : \delta\left(x, c'\right) = \sigma_i(w_\mu), \delta\left(y, c'\right) = \sigma_j(w_\nu)\right\}$$

does not depend on x and y. On the other hand $\sigma(x,z) = \mu$, $\sigma(y,z) = \nu$ if and only if $z = v_l(c')$ for some $c' \in A$ and some $l \in \hat{I}$. This proves that the set $\{z \in \widehat{\mathscr{V}}(\Delta) : \sigma(x,z) = \mu, \sigma(y,z) = \nu\}$ has a cardinality independent of x and y. Moreover, if $\sigma(x,y) = \lambda$, then $\sigma(y,x) = \lambda^\star$. Therefore

$$\begin{aligned} &\left|\left\{z \in \widehat{\mathscr{V}}(\Delta) : \sigma(x,z) = \mu, \sigma(y,z) = \nu\right\}\right| \\ &\quad = \left|\left\{z' \in \widehat{\mathscr{V}}(\Delta) : \sigma\left(y,z'\right) = \mu^\star, \sigma\left(x,z'\right) = \nu^\star\right\}\right| \end{aligned} \tag{2}$$

(see also [12, Theorem 5.21]). This completes the proof. □

Set

$$N(\lambda,\mu,\nu) := \left|\left\{z \in \widehat{\mathscr{V}}(\Delta) : \sigma(x,z) = \mu, \sigma(y,z) = \nu\right\}\right| = N\left(\lambda, \nu^\star, \mu^\star\right), \tag{3}$$

if $\sigma(x,y) = \lambda$.

2.19 Partial Ordering on $\mathbb{A}$

Define a partial order on $\widehat{L}$, by setting

$$\mu \preceq \lambda, \quad \text{if } \lambda - \mu \in L^+.$$

Since $\widehat{\mathscr{V}}(\mathbb{A})$ may be identified with the co-weight lattice $\widehat{L}$, the partial ordering defined on $\widehat{L}$ applies to $\widehat{\mathscr{V}}(\mathbb{A})$. For every $\lambda \in \widehat{L}^+$, we define

$$\Pi_\lambda = \left\{\mathbf{w}\mu : \mu \in \widehat{L}^+, \mu \preceq \lambda, \mathbf{w} \in \mathbf{W}\right\}.$$

This set is *saturated*: for every $\eta \in \Pi_\lambda$ and every $\alpha \in R$, then $\eta - j\alpha^\vee \in \Pi_\lambda$ for every $0 \leqslant j \leqslant \langle \eta, \alpha \rangle$. Hence it is stable under $\mathbf{W}$. Moreover λ is the highest co-weight of Π_λ. It is easy to prove that $\Pi_\lambda + \Pi_\mu \subset \Pi_{\lambda+\mu}$ for every $\lambda, \mu \in \widehat{L}^+$. Recall that W is equipped with the Bruhat ordering, defined as follows [6]. We declare $w_1 < w_2$

if there exists a sequence $w_1 = u_0 \to u_1, \dots, u_{k-1} \to u_k = w_2$, where $u_j \to u_{j+1}$ means that $u_{j+1} = u_j t$ for some conjugate t of an $s \in S$, and $|u_j| < |u_{j+1}|$. This defines a partial order on W that can be extended to $\widehat{W}$, by setting $\hat{w}_1 \leqslant \hat{w}_2$ if $\hat{w}_1 = w_1 g_1$ and $\hat{w}_2 = w_2 g_2$ with $w_1 < w_2$. We remark that $w_1 \leqslant w_2$ if and only if w_1 can be obtained as a sub-expression $s_{i_{k_1}} \cdots s_{i_{k_m}}$ of any reduced expression $s_{i_1} \cdots s_{i_r}$ for w_2. Observe that, for every $\lambda \in \widehat{L}^+$ and $\hat{w}' \leqslant \hat{w}$, if $\hat{w}(0) \in \Pi_\lambda$ then $\hat{w}'(0) \in \Pi_\lambda$.

Define also a partial ordering on $\mathscr{C}(\mathbb{A})$, in the following way. Given two chambers C_1, C_2, consider all the hyperplanes H_α^k separating C_1 and C_2. We declare $C_1 \prec C_2$ if C_2 belongs to the positive half-spaces determined by each of these hyperplanes. It is clear that the resulting relation $C_1 \preceq C_2$ is a partial ordering of $\mathscr{C}(\mathbb{A})$. Observe that, by definition of $\mathbb{Q}_0$, we have $C_0 \prec C$ if and only if $C \subset \mathbb{Q}_0$. Moreover, if C is any chamber and $s = s_\alpha^k$ is the affine reflection with respect to the hyperplane that contains a panel of C, then $C \prec s(C)$ or $s(C) \prec C$, since C and $s(C)$ are adjacent. Since $\mathscr{C}(\mathbb{A})$ may be identified with W, the previous definition induces a partial ordering on W. We point out that this ordering is different from the Bruhat order. Moreover, on $\mathbf{W}$, we have

$$\mathbf{w}_1(C_0) \prec \mathbf{w}_2(C_0) \quad \text{if and only if} \quad \mathbf{w}_1 > \mathbf{w}_2.$$

Proposition 2.17 *Let C be a chamber of $\mathbb{A}$, $s = s_\alpha^k$ the affine reflection with respect to the hyperplane H_α^k containing a panel of C and $\mathbf{s} = s_\alpha^0$. Assume that $C \prec s(C)$. Let $w \in W$. If $w = \mathbf{w}t_\lambda$ for some $\mathbf{w} \in \mathbf{W}$ and $\lambda \in L$, then*

(i) *if $w(C) \prec ws(C)$ then $\mathbf{w} < \mathbf{ws}$;*
(ii) *if $ws(C) \prec w(C)$ then $\mathbf{ws} < \mathbf{w}$.*

Proof Since α is positive and $C \prec s(C)$, then C and $s(C)$ belong respectively to the negative and the positive half-space determined by the affine hyperplane H_α^k. That is, for every vertex v lying in C,

$$\langle v, \alpha \rangle \leqslant k, \qquad \big\langle s(v), \alpha \big\rangle \geqslant k.$$

The adjacent chambers $w(C)$ and $ws(C)$ share a panel that belongs to the hyperplane $w(H_\alpha^k) = H_{\mathbf{w}(\alpha)}^{k'}$; moreover, for every $v \in C$,

$$\big\langle w(v), \mathbf{w}(\alpha) \big\rangle \leqslant k' \quad \text{and} \quad \big\langle ws(v), \mathbf{w}(\alpha) \big\rangle \geqslant k'.$$

Indeed, set $k' = k + \langle \lambda, \alpha \rangle$. Then

$$\big\langle w(v), \mathbf{w}(\alpha) \big\rangle = \big\langle \mathbf{w}t_\lambda(v), \mathbf{w}(\alpha) \big\rangle = \big\langle t_\lambda(v), \alpha \big\rangle = \langle v, \alpha \rangle + \langle \lambda, \alpha \rangle \leqslant k',$$
$$\big\langle ws(v), \mathbf{w}(\alpha) \big\rangle = \big\langle \mathbf{w}t_\lambda s(v), \mathbf{w}(\alpha) \big\rangle = \big\langle t_\lambda s(v), \alpha \big\rangle = \big\langle s(v), \alpha \big\rangle + \langle \lambda, \alpha \rangle \geqslant k'.$$

This implies that $\mathbf{w}(\alpha)$ is positive in case (i) and negative in case (ii). If $\mathbf{w}(\alpha) > 0$, then, for every $v \in \mathbb{Q}_0$,

$$\big\langle \mathbf{w}^{-1}v, \alpha \big\rangle = \big\langle v, \mathbf{w}(\alpha) \big\rangle > 0, \qquad \big\langle (\mathbf{ws})^{-1}v, \alpha \big\rangle = \big\langle v, \mathbf{ws}(\alpha) \big\rangle = -\big\langle v, \mathbf{w}(\alpha) \big\rangle < 0,$$

since $\langle v, \mathbf{s}(\alpha)\rangle = -\langle v, \alpha\rangle$. Therefore $\mathbb{Q}_0$ and $\mathbf{w}^{-1}(\mathbb{Q}_0)$ belong to the same half-space determined by H_α, while H_α separates $(\mathbf{ws})^{-1}(\mathbb{Q}_0)$ and $\mathbb{Q}_0$. So the number of hyperplanes separating $\mathbb{Q}_0$ and $(\mathbf{ws})^{-1}(\mathbb{Q}_0)$ is larger than the number of hyperplanes separating $\mathbb{Q}_0$ and $\mathbf{w}^{-1}(\mathbb{Q}_0)$, and we conclude that $\mathbf{w} < \mathbf{ws}$.

On the contrary, if $\mathbf{w}(\alpha) < 0$, then, for every $v \in \mathbb{Q}_0$,

$$\left\langle \mathbf{w}^{-1} v, \alpha \right\rangle < 0, \qquad \left\langle (\mathbf{ws})^{-1} v, \alpha \right\rangle > 0,$$

and therefore $\mathbf{w} > \mathbf{ws}$. □

2.20 Retraction ρ_x

Let x be any special vertex of Δ (say $\tau(x) = i$). For every $c \in \mathscr{C}(\Delta)$, denote by $\mathrm{proj}_x(c)$ the chamber containing x in any minimal gallery $\gamma(x, c)$. In particular we write $\mathrm{proj}_0(c)$ when x is the fundamental vertex e. We note that $\mathrm{proj}_x(c)$ does not depend on the minimal gallery considered.

In the fundamental apartment $\mathbb{A}$, let $\mathbb{Q}_0^- = \mathbf{w}_0(\mathbb{Q}_0)$ and let C_0^- be the base chamber of $\mathbb{Q}_0^-$.

Definition 2.18 For every $c \in \mathscr{C}(\Delta)$, the *retraction* of c *with respect to* x is defined as

$$\rho_x(c) = C_0^- \cdot \delta_i\big(\mathrm{proj}_x(c), c\big),$$

where, for every pair c, d of chambers, $\delta_i(c, d) := w_{\sigma_i^{-1}(f)}$ when $\delta(c, d) = w_f$. In particular, if $\tau(x) = 0$,

$$\rho_x(c) = C_0^- \cdot \delta\big(\mathrm{proj}_x(c), c\big).$$

Obviously, $\rho_x(c)$ belongs to $\mathbb{Q}_0^-$ for every c. We extend the previous definition to all special vertices. For every $y \in \mathscr{V}_{sp}(\Delta)$, say $\tau(y) = j$, we set

$$\rho_x(y) = v_l\big(\rho_x(c)\big)$$

if c is any chamber that contains y and $l = \sigma_i^{-1}(j)$. This definition is well posed: it does not depend on the choice of the chamber containing the vertices y, as $v_l(c_1) = v_l(c_2)$ implies $v_l(\rho_x(c_1)) = v_l(\rho_x(c_2))$. In particular, denote by ρ_0 the retraction with respect to the fundamental vertex e. We shall use the fact that, if $\lambda \in \widehat{L}^+$, and $t_\lambda = u_\lambda g_l$, then, for every c such that $\delta(\mathrm{proj}_0(c), c) = u_\lambda$, one has $\rho_0(c) = \mathbf{w}_0 u_\lambda(C_0)$. Therefore, if $\sigma(e, x) = \lambda$, then $\rho_0(x) = \mathbf{w}_0\lambda$.

2.21 Extended Chambers

Recall that the action of $\widehat{W}$ on the set $\mathscr{C}(\mathbb{A})$ is transitive but not simply transitive; indeed, if $\hat{w}_i = w g_i$, then $\hat{w}_i(C_0) = w(C_0)$, for every $w \in W$ and for every $i \in \hat{I}$.

Nevertheless, the action of the elements $\hat{w}_i$ on the special vertices $v_j(C_0)$ of C_0 depends on i, because

$$\hat{w}_i\big(v_j(C_0)\big) = v_{\sigma_i(j)}\big(w(C_0)\big).$$

This suggest to enlarge the set $\mathscr{C}(\mathbb{A})$ in the following way. We call extended chambers of $\mathbb{A}$ the pairs $\widehat{C} = (C, \sigma)$, where $C \in \mathscr{C}(\mathbb{A})$ and $\sigma \in \mathrm{Aut}_{tr}(D)$. Denote by $\widehat{\mathscr{C}}(\mathbb{A})$ the set of all extended chambers. A straightforward consequence of this definition is that $\widehat{W}$ acts simply transitively on $\widehat{\mathscr{C}}(\mathbb{A})$: for every couple of extended chambers $\widehat{C}_1 = (C_1, \sigma_{i_1})$ and $\widehat{C}_2 = (C_2, \sigma_{i_2})$, there exists a unique element $\hat{w} \in \widehat{W}$ such that $\widehat{C}_2 = \hat{w}(\widehat{C}_1)$. Indeed, if $C_2 = w(C_1)$, $g = g_{i_2} g_{i_1}^{-1}$ and σ is the automorphism of D corresponding to g, then $\hat{w} = wg = g\sigma(w)$. In particular, for every $\widehat{C} = (C, \sigma_i)$, then $\hat{w} = wg_i = g_i\sigma_i(w)$ is the unique element of $\widehat{W}$ such that $\hat{w}(C_0) = \widehat{C}$ if $C = w(C_0)$.

In the same way we enlarge the set $\mathscr{C}(\Delta)$ and define

$$\widehat{\mathscr{C}}(\Delta) = \big\{\hat{c} = (c, \sigma_i), c \in \mathscr{C}(\Delta), i \in \hat{I}\big\}.$$

Notice that, for every $c \in \mathscr{C}(\Delta)$ and $i \in \hat{I}$, there exists a unique $\hat{c}$ such that $v_i(c) = v_0(\hat{c})$; indeed, this element is $\hat{c} = (c, \sigma_i)$. The W-distance on $\mathscr{C}(\Delta)$ can be extended to a $\widehat{W}$-distance on $\widehat{\mathscr{C}}(\Delta)$ in the following way: for every pair of extended chambers $\hat{c}_1 = (c_1, \sigma_{i_1})$ and $\hat{c}_2 = (c_2, \sigma_{i_2})$, set

$$\hat{\delta}(\hat{c}_1, \hat{c}_2) = \delta(c_1, c_2) g_{i_2} g_{i_1}^{-1}.$$

For every $\lambda \in \widehat{L}^+$, with $\tau(\lambda) = l$, consider the translation $t_\lambda = u_\lambda g_l$. Then $t_\lambda(C_0) = (u_\lambda(C_0), g_l)$ and $v_0(t_\lambda(C_0)) = v_l(u_\lambda(C_0))$.

3 Maximal Boundary

This section describes the maximal boundary of an affine building.

3.1 Sectors of $\mathbb{A}$

Let R be a root system and let $\mathbb{A} = \mathbb{A}(R)$. In Sect. 2.7 we defined a sector of $\mathbb{A}$, based at 0, as any connected component of $\mathbb{V} \setminus \bigcup_\alpha H_\alpha$; in particular $\mathbb{Q}_0 = \{v \in \mathbb{V} : \langle v, \alpha\rangle > 0, i \in I_0\}$ is the fundamental sector based at 0. For every chamber C that contains 0, denote by $Q_0(C)$ the sector based at 0 with base chamber C; in particular, C_0 is the base chamber of $\mathbb{Q}_0$. Observe that $Q_0(C) = \mathbf{w}\mathbb{Q}_0$ for some $\mathbf{w} \in \mathbf{W}$.

More generally, for each special vertex X of $\mathbb{A}$, in particular for every $X \in \widehat{\mathscr{V}}(\mathbb{A})$, we call *sector* of $\mathbb{A}$, based at X, any connected component of $\mathbb{V} \setminus \bigcup_{H_\alpha^k \in \mathscr{H}_X} H_\alpha^k$, where $\mathscr{H}_X$ denotes the collection of all hyperplanes of $\mathscr{H}$ sharing X. For every chamber C that contains X, denote by $Q_X(C)$ the sector based at X with base chamber C. Observe that, for every $X \in \widehat{\mathscr{V}}(\mathbb{A})$ and every C containing X, there exists a unique $\hat{w} \in \widehat{W}$ such that $Q_X(C) = \hat{w}\mathbb{Q}_0$.

3.2 Maximal Boundary

We extend to any irreducible regular affine building Δ the definition of sector given on its fundamental apartment $\mathbb{A} = \mathbb{A}(R)$, declaring that, for any $x \in \mathscr{V}_{sp}(\Delta)$, a *sector* of Δ with base vertex x is a subcomplex Q_x of any apartment $\mathscr{A}$ of the building such that $\psi_{tp}(Q_x) = Q_X$, where X is any special vertex such that $\tau(X) = \tau(x)$, and $\psi_{tp} : \mathscr{A} \to \mathbb{A}$ is a type-preserving isomorphism mapping x to X. We note that, given any apartment $\mathscr{A}$ of the building, for every sector $Q_x \subset \mathscr{A}$ there exists a unique type-rotating isomorphism $\psi_{tr} : \mathscr{A} \to \mathbb{A}$ mapping Q_x to $\mathbb{Q}_0$.

We say that a sector Q_y is a *subsector* of a sector Q_x if $Q_y \subset Q_x$. Two sectors Q_x and Q_y are said to be *equivalent* if they share a subsector Q_z. Each equivalence class of sectors is called a *boundary point* of the building and is denoted by ω; the set of all equivalence classes of sectors is called the *maximal boundary* of the building and it is denoted by Ω. As immediate consequence of the definition, for every special vertex x and $\omega \in \Omega$ there is one and only one sector in the class ω based at x, denoted by $Q_x(\omega)$.

For every special vertex $x \in \mathscr{V}_{sp}(\Delta)$ and every $\omega \in \Omega$, there exists an apartment $\mathscr{A}(x, \omega)$ containing x and ω (in fact containing $Q_x(\omega)$). Analogously, for every chamber c and every $\omega \in \Omega$, there exists an apartment $\mathscr{A}(c, \omega)$ that contains c and ω, that is, c and a sector in the class ω. On this apartment we denote by $Q_c(\omega)$ the intersection of all sectors in the class ω containing c.

For every $x \in \mathscr{V}_{sp}(\Delta)$ and every chamber $c \in \mathscr{C}(\Delta)$, on the maximal boundary Ω we define the set

$$\Omega(x, c) = \left\{\omega \in \Omega : Q_x(\omega) \supset c\right\}.$$

Analogously, for every pair of special vertices x, y, we can define the subset $\Omega(x, y)$ of Ω given by

$$\Omega(x, y) = \left\{\omega \in \Omega : y \in Q_x(\omega)\right\}.$$

We note that, for every x,

$$\Omega\left(x, c'\right), \Omega(x, z) \supset \Omega(x, c) \quad \text{for every } c', z \text{ in the convex hull of } \{x, c\},$$

$$\Omega\left(x, c'\right), \Omega(x, z) \supset \Omega(x, y) \quad \text{for every } c', z \text{ in the convex hull of } \{x, y\}.$$

From now on we shall limit attention to sectors based at a vertex of $\widehat{\mathscr{V}}(\Delta)$.

3.3 Retraction ρ_ω^x

Let $\omega \in \Omega$ and $x \in \widehat{\mathscr{V}}(\Delta)$. For every apartment $\mathscr{A} = \mathscr{A}(x,\omega)$ that contains ω and x, there exists a unique type-rotating isomorphism $\psi_{tr} : \mathscr{A} \to \mathbb{A}$ such that $\psi_{tr}(Q_x(\omega)) = \mathbb{Q}_0$. On the other hand, if $\mathscr{A}'$ contains a subsector $Q_y(\omega)$ of $Q_x(\omega)$, but not x, then there exists a type-preserving isomorphism $\phi : \mathscr{A}' \to \mathscr{A}(x,\omega)$ fixing $Q_y(\omega)$; hence the type-rotating isomorphism $\psi'_{tr} = \psi_{tr}\phi : \mathscr{A}' \to \mathbb{A}$ is well defined. Since every facet $\mathscr{F}$ of the building lies in an apartment $\mathscr{A}'$ that contains a subsector $Q_y(\omega)$ of $Q_x(\omega)$ (possibly $Q_x(\omega)$), then, according to previous notation, $\mathscr{F}$ maps uniquely onto the facet $\mathbf{F} = \psi'_{tr}(\mathscr{F})$ of $\mathbb{A}$.

Definition 3.1 We call *retraction of Δ on $\mathbb{A}$ with respect to ω and center x* the map

$$\rho_\omega^x : \Delta \to \mathbb{A},$$

such that, for every apartment $\mathscr{A}'$ and for every facet $\mathscr{F} \in \mathscr{A}'$ $\rho_\omega^x(\mathscr{F}) = \mathbf{F} = \psi'_{tr}(\mathscr{F})$.

In particular we remark that $\rho_\omega^x(x) = 0$, and, if c_ω^x is the base chamber of $Q_x(\omega)$, then $\rho_\omega^x(c_\omega^x) = C_0$. Moreover, for every chamber $c \in Q_x(\omega)$ and for every special vertex $y \in Q_x(\omega)$,

$$\rho_\omega^x(c) = C_0 \cdot \delta\big(c_\omega^x, c\big) \quad \text{and} \quad \rho_\omega^x(y) = X_\mu$$

if X_μ is the special vertex associated with $\mu = \sigma(x,y)$. For the sake of simplicity, we simply write $\rho_\omega^x(z) = \mu$ to mean that $\rho_\omega^x(y) = X_\mu$. In the case $x = e$, set $\rho_\omega = \rho_\omega^e$.

Proposition 3.2 *Let $x \in \widehat{\mathscr{V}}(\Delta)$, $c \in \mathscr{C}(\Delta)$ and $\omega \in \Omega$. If $d \subset Q_x(\omega) \cap Q_c(\omega)$, then $\delta(x,d)\delta(d,c)$ is independent of d. Moreover*

$$\rho_\omega^x(c) = C_0 \cdot \delta(x,d)\delta(d,c).$$

Proof Fix $d \in Q_x(\omega) \cap Q_c(\omega)$. For every $d' \in Q_d(\omega)$,

$$\delta\big(x,d'\big) = \delta\big(c_\omega^x, d'\big) = \delta\big(c_\omega^x, d\big)\delta\big(d,d'\big) \quad \text{and} \quad \delta\big(c,d'\big) = \delta(c,d)\delta\big(d,d'\big),$$

where c_ω^x is the base chamber of the sector $Q_x(\omega)$. Hence $\delta(c_\omega^x, d')\delta(c,d')^{-1} = \delta(c_\omega^x, d)\delta(c,d)^{-1}$. Then, given d_1 and d_2 in $Q_x(\omega) \cap Q_c(\omega)$, for $d' \in Q_{d_1}(\omega) \cap Q_{d_2}(\omega)$,

$$\delta\big(c_\omega^x, d_1\big)\delta(c,d_1)^{-1} = \delta\big(c_\omega^x, d'\big)\delta\big(c,d'\big)^{-1} = \delta\big(c_\omega^x, d_2\big)\delta(c,d_2)^{-1}.$$

By definition of ρ_ω^x,

$$\rho_\omega^x(d) = \rho_\omega^x\big(c_\omega^x\big) \cdot \delta\big(c_\omega^x, d\big) = C_0 \cdot \delta\big(c_\omega^x, d\big) \quad \text{and} \quad \rho_\omega^x(d) = \rho_\omega^x(c) \cdot \delta(c,d).$$

Indeed, since $d \subset Q_x(\omega) \cap Q_c(\omega)$, the retraction of a gallery $\gamma(c^x_\omega, d)$ is a gallery $\Gamma(\rho^x_\omega(c^x_\omega), \rho^x_\omega(d))$ of the same type as $\gamma(c^x_\omega, d)$ and the retraction of a gallery $\gamma(c, d)$ is a gallery $\Gamma(\rho^x_\omega(c), \rho^x_\omega(d))$ of the same type as $\gamma(c, d)$. Therefore

$$\rho^x_\omega(c) = \rho^x_\omega(d) \cdot \delta(c, d)^{-1} = \rho^x_\omega(d) \cdot \delta(d, c) = C_0 \cdot \delta\big(c^x_\omega, d\big)\delta(d, c). \qquad \square$$

An analogous of Proposition 3.2 holds for the retraction ρ^x_ω of special vertices of the building.

Proposition 3.3 *Let $x, y \in \widehat{\mathscr{V}}(\Delta)$ and $\omega \in \Omega$. For every $z \in Q_x(\omega) \cap Q_y(\omega)$, $\sigma(x, z) - \sigma(y, z)$ is independent of z. Moreover*

$$\rho^x_\omega(y) = \sigma(x, z) - \sigma(y, z).$$

Proof Fix $z \in Q_x(\omega) \cap Q_y(\omega)$ and assume that $\sigma(x, z) = \mu$ and $\sigma(y, z) = \nu$. For every $z' \in Q_z(\omega)$, we have $\sigma(x, z') = \mu + \lambda'$, $\sigma(y, z') = \nu + \lambda'$ if $\sigma(z, z') = \lambda'$; hence $\sigma(x, z') - \sigma(y, z') = \mu - \nu$. Given z_1 and z_2 in $Q_x(\omega) \cap Q_y(\omega)$, for every $z' \in Q_{z_1}(\omega) \cap Q_{z_2}(\omega)$ we conclude that

$$\sigma(x, z_1) - \sigma(y, z_1) = \sigma\big(x, z'\big) - \sigma\big(y, z'\big) = \sigma(x, z_2) - \sigma(y, z_2).$$

This proves that $\sigma(x, z) - \sigma(y, z)$ does not depend on the choice of z in $Q_x(\omega) \cap Q_y(\omega)$.

In order to prove that $\rho^x_\omega(y) = \sigma(x, z) - \sigma(y, z)$ for every $z \in Q_x(\omega) \cap Q_y(\omega)$ we fix any apartment $\mathscr{A}(x, \omega)$ that contains $Q_x(\omega)$. If $y \in \mathscr{A}(x, \omega)$, and $z \in Q_x(\omega) \cap Q_y(\omega)$, then $\rho^x_\omega(x) = 0$, $\rho^x_\omega(z) = \mu$; moreover, if we set $\rho^x_\omega(y) = \eta$, then $\tau_{-\eta}(Q_\eta) = \mathbb{Q}_0$, and in particular $\mu - \eta = \tau_{-\eta}(\rho^x_\omega(z)) = \nu$. If, instead, $y \notin \mathscr{A}(x, \omega)$, then there is $y' \in \mathscr{A}(x, \omega)$ such that $\rho^x_\omega(y) = \rho^x_\omega(y')$ and $\sigma(y, z) = \sigma(y', z) = \mu - \nu$. Hence, as before, $\mu - \eta = \tau_{-\eta}(\rho^x_\omega(z)) = \nu$. $\square$

Corollary 3.4 *For all x, y, z in $\widehat{\mathscr{V}}(\Delta)$ and for each $\omega \in \Omega$,*

$$\rho^y_\omega(z) = \rho^x_\omega(z) - \rho^x_\omega(y).$$

Proof Let $z' \in Q_x(\omega) \cap Q_y(\omega) \cap Q_z(\omega)$. Then $\rho^x_\omega(y) = \sigma(x, z') - \sigma(y, z')$, $\rho^x_\omega(z) = \sigma(x, z') - \sigma(z, z')$ and $\rho^y_\omega(z) = \sigma(y, z') - \sigma(z, z')$. Therefore

$$\rho^x_\omega(z) - \rho^x_\omega(y) = \sigma\big(y, z'\big) - \sigma\big(z, z'\big) = \rho^y_\omega(z). \qquad \square$$

Notice that, if $z = x$, then $\rho^y_\omega(x) = -\rho^x_\omega(y)$. In particular, for all x, y special and for each $\omega \in \Omega$,

$$\rho^x_\omega(y) = \rho_\omega(y) - \rho_\omega(x).$$

In fact this formula is independent of the choice of the fundamental vertex e.

We shall prove that, for every $\lambda \in \widehat{L}^+$, it is possible to choose μ large enough with respect to λ such that Proposition 3.3 holds for every $y \in V_\lambda(x)$ and every $\omega \in \Omega$.

For every chamber c denote by $\ell(x,c)$ the length of the element $w=\delta(x,c)$, that is, the number of hyperplanes separating x and c. On the fundamental apartment $\mathbb{A}$ define, for every $v \in \mathbb{Q}_0$,

$$\partial(v,\partial\mathbb{Q}) = \min\bigl\{\langle v,\alpha_i\rangle, i \in I_0\bigr\}.$$

Extend this definition to all special vertices of $Q_x(\omega)$, for all x and ω, in the following way: for each special vertex $y \in Q_x(\omega)$,

$$\partial\bigl(y,\partial Q_x(\omega)\bigr) = \partial\bigl(\rho^x_\omega(y),\partial\mathbb{Q}_0\bigr).$$

We define, for $k \in \mathbb{N}$,

$$Q^k_x(\omega) = \bigl\{y \in Q_x(\omega) : \partial\bigl(y,\partial Q_x(\omega)\bigr) \geqslant k\bigr\}.$$

Lemma 3.5 *Let $x \in \widehat{\mathscr{V}}(\Delta)$, $\omega \in \Omega$ and $k > 0$. Then*

$$Q^k_x(\omega) \subset Q_c(\omega) \tag{4}$$

for every $c \in \mathscr{C}(\Delta)$ such that $\ell(x,c) \leqslant k$.

Proof If $k=0$, then $x \in c$, hence $Q_x(\omega) \subset Q_c(\omega)$. On the other hand, the set $\{y \in Q_x(\omega) : \partial(y,\partial Q_x(\omega)) \geqslant 0\}$ coincides with $Q_x(\omega)$, hence (4). By induction, assume now that (4) holds for every c such that $\ell(x,c) \leqslant k$. Let c_1 such that $\ell(x,c_1)=k+1$. If $\gamma(x,c_1)$ is a gallery joining x to c_1, denote by d_1 the chamber of this gallery adjacent to c_1. Then $\ell(x,d_1)=k$, therefore

$$\bigl\{y \in Q_x(\omega) : \partial\bigl(y,\partial Q_x(\omega)\bigr) \geqslant k\bigr\} \subset Q_{d_1}(\omega).$$

Thus the result follows readily if $Q_{c_1}(\omega) \supset Q_{d_1}(\omega)$. Otherwise, $Q_{c_1}(\omega) \subset Q_{d_1}(\omega)$ and, for every $y \in (Q_{d_1}(\omega) \setminus Q_{c_1}(\omega)) \cap Q_x(\omega)$, one has $\langle\rho^x_\omega(y),\alpha\rangle = k$ for some $\alpha \in R^+$, and $\langle\rho^x_\omega(y),\alpha'\rangle = k \geqslant k$ for $\alpha' \neq \alpha$. On the other hand,

$$\begin{aligned}&\bigl\{y \in Q_x(\omega) : \partial\bigl(y,\partial Q_x(\omega)\bigr) \geqslant k+1\bigr\}\\&\quad= \bigl\{y \in Q_x(\omega) : \partial\bigl(y,\partial Q_x(\omega)\bigr) \geqslant k\bigr\} \setminus \bigl\{y \in Q_x(\omega) : \partial\bigl(y,\partial Q_x(\omega)\bigr) = k\bigr\}\end{aligned}$$

and $\{y \in Q_x(\omega) : \partial(y,\partial Q_x(\omega)) = k\}$ is the set of all $y \in Q_x(\omega)$ such that $\langle\rho^x_\omega(y),\alpha\rangle = k$ for some $\alpha \in R^+$, and $\langle\rho^x_\omega(y),\alpha'\rangle = k' \geqslant k$ for $\alpha' \neq \alpha$. Thus (4) is true also in this case. □

Let $x \in \widehat{\mathscr{V}}(\Delta)$ and $\omega \in \Omega$. For every $w \in W$, denote by $Q_w(\omega)$ the intersection of all sectors in the class ω containing the chamber d_w such that $\delta(c_x(\omega),d_w)=w$.

Proposition 3.6 *For every $w_1 \in W$ there exists $w_0 \in W$ such that, for every x and c such that $\delta(x,c)=w_1$ and for every $\omega \in \Omega$,*

$$Q_{w_0}(\omega) \subset Q_x(\omega) \cap Q_c(\omega).$$

Moreover, for every chamber d of $Q_{w_0}(\omega)$,

$$\rho_\omega^x(c) = C_0 \cdot \delta\big(c_x(\omega), d\big)\delta(d, c).$$

Proof Let $k > 0$ and $Q_k = \{v \in \mathbb{Q}_0 : \langle v, \alpha_i \rangle \geqslant k,\ \text{for every } i \in I_0\}$. Choose a chamber $D \subset Q_k$ and let w_k be the element of W such that $D = C_0 \cdot w_k$. For every ω, consider the chamber d_{w_k} such that $\delta(c_x(\omega), d_{w_k}) = w_k$ and the sector $Q_{w_k}(\omega)$. If k is larger than the length of w_1, that is, $\ell(x, c) \leqslant k$, then Lemma 3.5 implies that for every ω the sector $Q_{w_k}(\omega)$ lies in $Q_x(\omega) \cap Q_c(\omega)$. Therefore $w_0 = w_k$ is the required element of W. Moreover, Proposition 3.2 implies that, for every chamber d of $Q_{w_0}(\omega)$,

$$\rho_\omega^x(c) = C_0 \cdot \delta\big(c_x(\omega), d\big)\delta(d, c).$$ □

Given x and ω, for every $\lambda \in \widehat{L}^+$ denote by z_λ the unique vertex of $Q_x(\omega)$ such that $\sigma(x, z_\lambda) = \lambda$ and by $Q_\lambda(\omega)$ the subsector of $Q_x(\omega)$ of base vertex z_λ. Moreover, denote by k_λ the number of hyperplanes separating 0 and λ.

Proposition 3.7 *Let $\lambda \in \widehat{L}^+$; there exists $\mu \in \widehat{L}^+$ (large enough with respect to λ) such that, for every pair $x, y \in \widehat{\mathscr{V}}(\Delta)$ and for every $\omega \in \Omega$,*

$$Q_\mu(\omega) \subset Q_x(\omega) \cap Q_y(\omega).$$

Moreover, for every ν such that $\nu - \mu \in \widehat{L}^+$,

$$\rho_\omega^x(y) = \mu - \sigma(y, z_\mu) = \nu - \sigma(y, z_\nu).$$

Proof Let $\lambda \in \widehat{L}^+$ and $Q_{k_\lambda} = \{v \in \mathbb{Q}_0 : \langle v, \alpha_i \rangle > k_\lambda,\ \text{for every } i \in I_0\}$. Choose a special vertex $\mu \in Q_{k_\lambda}$. For every ω consider the special vertex z_μ of $Q_x(\omega)$ such that $\sigma(x, z_\mu) = \mu$, and the sector $Q_\mu(\omega)$ based at z_μ. By Proposition 3.6, for every ω, the sector $Q_\mu(\omega)$ lies in $Q_x(\omega) \cap Q_c(\omega)$. Therefore, by Proposition 3.3, $\rho_\omega^x(y) = \mu - \sigma(y, z_\mu)$. The same is true for every ν such that $\nu - \mu \in \widehat{L}^+$; indeed, if $\nu - \mu \in \widehat{L}^+$, one has $z_\nu \in Q_\mu(\omega)$. □

Observe that Proposition 3.7 holds if $\langle \mu, \alpha_i \rangle \geqslant k_\lambda$, for every $i \in I_0$.

As a consequence of Proposition 3.7 the following result holds.

Theorem 3.8 *Let $y \in \mathscr{V}_\lambda(x)$ and $z \in \mathscr{V}_\mu(x)$. If μ is large enough with respect to λ, then $\Omega(x, z) \subset \Omega(y, z)$. Moreover, for all $\omega \in \Omega(x, z)$, $\rho_\omega^x(y) = \mu - \nu$ if $\sigma(y, z) = \nu$.*

Proof If $\omega \in \Omega(x, z)$, then $z \in Q_x(\omega)$ and therefore, if μ is large enough, $z \in Q_y(\omega)$ by Proposition 3.7: that is, $\omega \in \Omega(y, z)$. The second part of the theorem follows readily from Proposition 3.3. □

Corollary 3.9 *Let $y \in \mathscr{V}_\lambda(x)$ and $z \in \mathscr{V}_\mu(x) \cap \mathscr{V}_\nu(y)$. If μ is large enough with respect to λ and ν is large enough with respect to $\lambda^\star$, then $\Omega(x, z) = \Omega(y, z)$.*

Let $y \in \mathscr{V}_\lambda(x)$ and $\omega \in \Omega$. We know that $\rho_\omega^x(y) = \lambda$ if $y \in Q_x(\omega)$. The following proposition describes the retraction of the vertices of the set $\mathscr{V}_\lambda(x)$.

Proposition 3.10 *Let $\omega \in \Omega$, x be special and $\lambda \in \widehat{L}^+$. Then $\rho_\omega^x(y) \in \Pi_\lambda$ for every $y \in \mathscr{V}_\lambda(x)$.*

Proof Let f_λ be the type of a minimal gallery connecting 0 to λ. Then each vertex $y \in \mathscr{V}_\lambda(x)$ is connected to x by a minimal gallery $\gamma(x, y)$ of type $\sigma_i(f_\lambda)$ (Sect. 2.12). This implies that $\rho_\omega^x(\gamma(x, y))$ is a gallery of type f_λ (possibly not reduced) on $\mathbb{A}$ that joins 0 to $\mu = \rho_\omega^x(y)$. Thus there is a reduced gallery from 0 to μ, of type, say, f'. Let $\lambda' = s_{f'} g_l(0)$. Since $\lambda = w_\lambda g_l(0)$ and $s_{f'} \leqslant w_\lambda$, then $\lambda' \in \Pi_\lambda$. On the other hand, if c and d are the chambers of $\gamma(x, y)$ containing x and y respectively, there exists $\mathbf{w} \in \mathbf{W}$ such that $\rho_\omega^x(c) = \mathbf{w}(C_0)$, hence $\rho_\omega^x(d) = \mathbf{w}(s_{f'}(C_0))$. This implies that $\mu = \mathbf{w}(\lambda')$ belongs to Π_λ. □

Let us determine how many vertices of $\mathscr{V}_\lambda(x)$ are mapped by ρ_ω^x onto an element of Π_λ. By means of Proposition 2.14 of Sect. 2.18, we now prove that this number is actually independent of x and ω.

Theorem 3.11 *Let $x \in \widehat{\mathscr{V}}(\Delta)$ and $\omega \in \Omega$. For $w, w_1 \in W$, then*

$$\left|\left\{c \in \mathscr{C}(\Delta) : \delta(x, c) = w_1, \rho_\omega^x(c) = C_0 \cdot w\right\}\right|$$

is independent of x and ω.

Proof Fix $w_1 \in W$; by Proposition 3.6, there exists $w_0 \in W$ such that, for every chamber c such that $\delta(x, c) = w_1$ and for every $\omega \in \Omega$, the set $Q_x(\omega) \cap Q_c(\omega)$ contains a chamber c' such that $\delta(x, c') = w_0$. Moreover, by Proposition 3.2, $\rho_\omega^x(c) = C_0 \cdot \delta(c_\omega^x, c')\delta(c', c) = C_0 \cdot w_0\delta(c', c)$. Hence, for any $w \in W$,

$$\begin{aligned}&\left\{c : \delta(x, c) = w_1, \rho_\omega^x(c) = C_0 \cdot w\right\}\\ &= \left\{c : \delta(x, c) = w_1, w_0\delta\left(c', c\right) = w\right\} = \left\{c : \delta(x, c) = w_1, \delta\left(c', c\right) = w_0^{-1}w\right\}.\end{aligned}$$

On the other hand, Proposition 2.14 of Sect. 2.18 implies that $|\{c : \delta(x, c) = w_1, \delta(c', c) = w_0^{-1}w\}|$ only depends on $\tau(x)$ and w_0, w_1, $w_0^{-1}w$. The statement follows. □

Finally, one has

Theorem 3.12 *Let $x \in \widehat{\mathscr{V}}(\Delta)$ and $\omega \in \Omega$. For every $\lambda \in \widehat{L}^+$ and $\mu \in \Pi_\lambda$,*

$$\left|\left\{y \in \mathscr{V}_\lambda(x) : \rho_\omega^x(y) = \mu\right\}\right|$$

is independent of x and ω.

Proof Let $\lambda \in \widehat{L}^+$, $\mu \in \Pi_\lambda$ and $\omega \in \Omega$. Consider the set

$$A = \{y : \sigma(x, y) = \lambda, \rho_\omega^x(y) = \mu\}.$$

For any $y \in \mathscr{V}_\lambda(x)$, denote by c_λ the chamber that contains y in a minimal gallery $\gamma(x, y)$. Then $y = v_j(c_\lambda)$ if $\tau(y) = j$, and $\delta(x, c_\lambda) = w_\lambda$. Thus

$$A = \{v_j(c), \delta(x, c) = w_\lambda, v_j(\rho_\omega^x(c)) = \mu\}.$$

Let W_μ be the stabilizer of μ in W. For every $w \in W_\mu$, consider the set of chambers

$$B_w = \{c : \delta(x, c) = w_\lambda, \rho_\omega^x(c) = C_0 \cdot w\}$$

and $B = \bigcup_{w \in W_\mu} B_w$. Notice that, if $v_j(\rho_\omega^x(c)) = \mu$, then $\rho_\omega^x(c) = C_0 \cdot w$ for some $w \in W_\mu$. Therefore $A = \{v_j(c), c \in B\}$, and then $|A| = |B| = \sum_{w \in W_\mu} |B_w|$. As Theorem 3.11 implies that $|B_w|$ is independent of x and ω, the same is true for $|A|$. □

This theorem allows us to set, for every $x \in \mathscr{V}_\lambda(x)$ and $\omega \in \Omega$,

$$N(\lambda, \mu) = \left|\{y \in \mathscr{V}_\lambda(x) : \rho_\omega^x(y) = \mu\}\right|. \tag{5}$$

For every $x \in \widehat{\mathscr{V}}(\Delta)$ and $\omega \in \Omega$, we now compare the retraction ρ_ω^x with the retraction ρ_x with respect to x, defined in Sect. 2.20.

Lemma 3.13 *Let c be any chamber and let y be any special vertex of $\widehat{\mathscr{V}}(\Delta)$.*

(i) *If c (respectively y) lies in the sector $Q_x^-(\omega)$ opposite to the sector $Q_x(\omega)$ in any apartment $\mathscr{A}(x, \omega)$, then*

$$\rho_\omega^x(c) = \rho_x(c) \quad \big(\text{respectively } \rho_\omega^x(y) = \rho_x(y)\big).$$

(ii) *If c (respectively y) belongs to the sector $(Q_x^\alpha)^-(\omega)$ α-adjacent to $Q_x^-(\omega)$ in a suitable apartment that contains c and $Q_x(\omega)$, then*

$$\rho_\omega^x(c) = s_\alpha \rho_x(c) \quad \big(\text{respectively } \rho_\omega^x(y) = s_\alpha \rho_x(y)\big).$$

Proof Assume first $\tau(x) = 0$.

(i) Let us show that $\rho_\omega^x(c) = \rho_x(c)$ for every chamber c of $Q_x^-(\omega)$. Since c lies in the sector $Q_x^-(\omega)$, then $Q_c(\omega) \supset Q_x(\omega)$. Therefore c_ω^x belongs to $Q_c(\omega)$. This implies that $\rho_\omega^x(c) = C_0 \cdot \delta(c^x(\omega), c)$. On the other hand

$$\delta(c^x(\omega), c) = \delta(c^x(\omega), \mathrm{proj}_x(c))\delta(\mathrm{proj}_x(c), c) = \mathbf{w}_0\delta(\mathrm{proj}_x(c), c),$$

therefore

$$\rho_\omega^x(c) = C_0 \cdot \mathbf{w}_0\delta(\mathrm{proj}_x(c), c) = C_0^- \cdot \delta(\mathrm{proj}_x(c), c) = \rho^x(c).$$

If $y \in Q_x^-(\omega)$, we may choose $\gamma(x, y)$ in $Q_x^-(\omega)$. Therefore, if c is the chamber of $\gamma(x, y)$ that contains y, we have $\rho_\omega^x(c) = \rho_x(c)$ and so $\rho_\omega^x(y) = \rho_x(y)$.

(ii) Let us now show that $\rho^x_\omega(c) = s_\alpha \rho_x(c)$ for every chamber c of $(Q^\alpha_x)^-(\omega)$. Since c lies in the sector $(Q^\alpha_x)^-(\omega)$, then $\mathrm{proj}_x(c)$ is the base chamber of the sector $(Q^\alpha_x)^-(\omega)$, that is, the chamber opposite to the base chamber $c^\alpha_x(\omega)$ of the sector $(Q^\alpha_x)(\omega)$ which is α-adjacent to $(Q_x)^-(\omega)$. This implies that

$$\delta\big(c^x(\omega), \mathrm{proj}_x(c)\big) = s_\alpha \delta\big(c^\alpha_x(\omega), \mathrm{proj}_x(c)\big) = s_\alpha \mathbf{w}_0.$$

From this equality it follows that

$$\delta\big(c^x(\omega), c\big) = \delta\big(c^x(\omega), \mathrm{proj}_x(c)\big)\delta\big(\mathrm{proj}_x(c), c\big) = s_\alpha \mathbf{w}_0 \delta\big(\mathrm{proj}_x(c), c\big).$$

Therefore

$$\rho^x_\omega(c) = C_0 \cdot s_\alpha \mathbf{w}_0 \delta\big(\mathrm{proj}_x(c), c\big) = s_\alpha (C_0 \cdot \mathbf{w}_0 \delta\big(\mathrm{proj}_x(c), c\big) = s_\alpha \rho^x(c).$$

If $y \in (Q^\alpha_x)^-_x(\omega)$, we may choose $\gamma(x, y)$ in $(Q^\alpha_x)^-_x(\omega)$. Therefore, if c is the chamber of $\gamma(x, y)$ that contains y, it follows $\rho^x_\omega(c) = s_\alpha \rho_x(c)$ and so $\rho^x_\omega(y) = s_\alpha \rho_x(y)$.

Finally, if $\tau(x) = i \neq 0$, swap δ with δ_i and use the same argument. □

3.4 Topologies on the Maximal Boundary

The maximal boundary Ω may be equipped with a totally disconnected compact Hausdorff topology in the following way. Fix a special vertex $x \in \widehat{\mathscr{V}}(\Delta)$, say of type $i = \tau(x)$, and consider the family

$$\mathscr{B}_x = \big\{\Omega(x, c), c \in \mathscr{C}(\Delta)\big\}.$$

Then $\mathscr{B}_x$ generates a totally disconnected compact Hausdorff topology on Ω. For every $\omega \in \Omega$, a local base at ω is given by

$$\mathscr{B}_{x,\omega} = \big\{\Omega(x, c), c \subset Q_x(\omega)\big\}.$$

We observe that it suffices to consider, as a local base at ω, only the chambers c lying in $Q_x(\omega)$ such that, for some $\lambda \in \widehat{L}^+$, $\delta(c_x(\omega), c) = \sigma_i(t_\lambda)$, where $c_x(\omega)$ is the base chamber of the sector $Q_x(\omega)$ and $i = \tau(x)$.

Remark 3.14 For every special vertex $y \in \widehat{\mathscr{V}}(\Delta)$, let $\lambda = \sigma(x, y)$ and denote by $\mathscr{C}_y$ the set of all chambers containing y such that $\delta(x, c) = \sigma_i(t_\lambda)$. This is the set of all chambers containing y and opposite to the chamber that contains y in a minimal gallery connecting x and y. It is easy to check that

$$\Omega(x, y) = \bigcup_{c \in \mathscr{C}_y} \Omega(x, c).$$

Moreover, for every chamber c choose $\bar{y} \in \widehat{\mathscr{V}}(\Delta)$ such that c lies in $[x, \bar{y}]$ and let $\lambda = \sigma(x, \bar{y})$. Then

$$\Omega(x, c) = \bigcup_{y \in V_\lambda(x),\, [x,y] \supset c} \Omega(x, y).$$

Hence the family $\widetilde{\mathscr{B}}_x = \{\Omega(x, y), y \in \widehat{\mathscr{V}}(\Delta)\}$ generates on Ω the same topology as $\mathscr{B}_x$ and, for every $\omega \in \Omega$, a local base at ω is given by $\widetilde{\mathscr{B}}_{x,\omega} = \{\Omega(x, y), y \subset Q_x(\omega)\}$.

Proposition 3.15 *The topology on Ω does not depend on the choice of $x \in \widehat{\mathscr{V}}(\Delta)$.*

Proof Let $x, y \in \widehat{\mathscr{V}}(\Delta)$, $\lambda = \sigma(x, y)$ and $\omega_0 \in \Omega$. We prove that, for every neighborhood $\Omega(y, z)$ of ω_0, there exists a neighborhood $\Omega(x, z')$ of ω_0 such that $\Omega(x, z') \subset \Omega(y, z)$. Indeed, if z' is a vertex of $Q_x(\omega_0) \cap Q_y(\omega_0)$ such that $z \in [y, z']$, then $\omega_0 \in \Omega(y, z') \cap \Omega(x, z')$ and $\Omega(y, z') \subset \Omega(y, z)$. On the other hand, if $\sigma(x, z') = \mu$, then Theorem 3.8 of the previous section. yields μ large enough with respect to λ so that $\Omega(x, z') \subset \Omega(y, z')$. $\square$

3.5 Probability Measures on the Maximal Boundary

For each vertex x of $\widehat{\mathscr{V}}(\Delta)$, denote by ν_x the regular Borel probability measure on Ω such that, for every $y \in \widehat{\mathscr{V}}(\Delta)$,

$$\nu_x\big(\Omega(x, y)\big) = N_\lambda^{-1} = \frac{\mathbf{W}_\lambda(q^{-1})}{\mathbf{W}(q^{-1})} \prod_{\alpha \in R^+} q_\alpha^{-\langle \lambda, \alpha \rangle} q_{2\alpha}^{\langle \lambda, \alpha \rangle} \quad \text{if } y \in V_\lambda(x).$$

In fact there exists a unique regular Borel probability measure on Ω, satisfying this property; indeed ν_x is the measure such that, for every $f \in \mathscr{C}(\Omega)$,

$$J(f) = \int_\Omega f(\omega)\, d\nu_x(\omega),$$

where J denotes the linear functional on $\mathscr{C}(\Omega)$ that extends the linear functional on the space of all locally constant functions on Ω defined by

$$J(f) = N_\lambda^{-1} \sum_{\sigma(x,y)=\lambda} f_y,$$

where, for each $y \in V_\lambda(x)$, we set $f_y = f(\omega)$, for every $\omega \in \Omega(x, y)$.

The following property of the measure ν_x is a consequence of Proposition 3.6 and Theorem 3.11 of Sect. 3.3.

Theorem 3.16 *Let $x \in \widehat{\mathscr{V}}(\Delta)$ and $w, w_0 \in W$. For each $c \in \mathscr{C}(\Delta)$ such that $\delta(x,c) = w_0$,*

$$\nu_x\bigl(\bigl\{\omega \in \Omega : \rho^x_\omega(c) = C_0 \cdot w\bigr\}\bigr)$$

is independent of x and c.

Proof Fix $w_0 \in W$ and a chamber c such that $\delta(x,c) = w_0$. By Proposition 3.6 of Sect. 3.3, there exists $w_1 \in W$ such that, for every ω, $Q_{w_1}(\omega) \subset Q_x(\omega) \cap Q_c(\omega)$. Moreover $\rho^x_\omega(c) = C_0 \cdot \delta(x,d)\delta(d,c)$ if d is any chamber of $Q_{w_1}(\omega)$. In particular,

$$\rho^x_\omega(c) = C_0 \cdot w_1 \delta\bigl(d_{w_1}(\omega), c\bigr),$$

where $d_{w_1}(\omega)$ denotes the chamber of $Q_{w_1}(\omega)$ such that $\delta(x, d_{w_1}(\omega)) = w_1$. Therefore, for any $w \in W$, we have $\rho^x_\omega(c) = C_0 \cdot w$ if and only if $w = w_1\delta(d_{w_1}(\omega), c)$, that is, if and only if $\delta(c, d_{w_1}(\omega)) = w^{-1}w_1$. Therefore, by setting $w^{-1}w_1 = w_2$ and $\mathscr{C}(w_1, w_2) = \{c' : \delta(x,c') = w_1, \delta(c,c') = w_2\}$, one has

$$\bigl\{\omega \in \Omega : \rho^x_\omega(c) = C_0 \cdot w\bigr\} = \bigcup_{c' \in \mathscr{C}(w_1,w_2)} \Omega\bigl(x, c'\bigr).$$

This implies that

$$\nu_x\bigl(\bigl\{\omega \in \Omega : \rho^x_\omega(c) = C_0 \cdot w\bigr\}\bigr) = \sum_{c' \in \mathscr{C}(w_1,w_2)} \nu_x\bigl(\Omega\bigl(x, c'\bigr)\bigr).$$

On the other hand, $\nu_x(\Omega(x,c'))$ has the same value for each chamber c' such that $\delta(x,c') = w_1$. Therefore, by fixing any chamber c' such that $\delta(x,c') = w_1$,

$$\begin{aligned}&\nu_x\bigl(\bigl\{\omega \in \Omega : \rho^x_\omega(c) = C_0 \cdot w\bigr\}\bigr)\\ &\quad = \nu_x\bigl(\Omega\bigl(x,c'\bigr)\bigr)\bigl|\bigl\{c' \in \mathscr{C}(\Delta) : \delta\bigl(x,c'\bigr) = w_1, \delta\bigl(c,c'\bigr) = w_2\bigr\}\bigr|.\end{aligned}$$

Thus Theorem 3.11 of Sect. 3.3 implies that $\nu_x(\{\omega \in \Omega : \rho^x_\omega(c) = C_0 \cdot w\})$ is independent of the choice of x and c, but depends only on w, w_0. □

A variant of this theorem holds for the set of vertices.

Theorem 3.17 *Let x be a special vertex of $\widehat{\mathscr{V}}(\Delta)$, $\lambda \in \widehat{L}^+$ and $\mu \in \Pi_\lambda$. For each $y \in \widehat{\mathscr{V}}(\Delta)$ such that $\sigma(x,y) = \lambda$, $\nu_x(\{\omega \in \Omega : \rho^x_\omega(y) = \mu\})$ is independent of x and y.*

Proof Fix $y \in \widehat{\mathscr{V}}(\Delta)$ such that $\sigma(x,y) = \lambda$, and for every $\mu \in \Pi_\lambda$ consider the set

$$\Omega_\mu = \bigl\{\omega \in \Omega : \rho^x_\omega(y) = \mu\bigr\}.$$

If $\tau(x) = i$, $\tau(y) = j$, then $\tau(X_\lambda) = l = \sigma_i^{-1}(j)$. Therefore

$$\Omega_\mu = \bigl\{\omega \in \Omega : v_l\bigl(\rho^x_\omega(c_\lambda)\bigr) = \mu\bigr\},$$

where c_λ denotes, as usual, the chamber that contains the vertex y in a minimal gallery connecting x and y. Therefore, if W_μ is the stabilizer of μ in W, then

$$\Omega_\mu = \{\omega \in \Omega : \rho^x_\omega(y) = C_0 \cdot w, w \in W_\mu\} = \bigcup_{w \in W_\mu} \{\omega \in \Omega : \rho^x_\omega(y) = C_0 \cdot w\}.$$

Now the proof follows from Theorem 3.16. □

4 The α-Boundary Ω_α

For every simple root α, we define in this section the *α-boundary* of the building, that consists of all equivalence classes of boundary points.

4.1 Walls

Let Δ be an affine building and let R be its root system. In the fundamental apartment $\mathbb{A} = \mathbb{A}(R)$ consider the fundamental sector $\mathbb{Q}_0 = Q_0(C_0)$. It would be natural to call walls of $\mathbb{Q}_0$ the walls of C_0 that contain 0 (Sect. 2.10). Actually, we slightly change this definition by calling *wall* of $\mathbb{Q}_0$ the intersection with $\overline{\mathbb{Q}_0}$ of any hyperplane $H_i = H_{\alpha_i}$, $i \in I_0$. Moreover, for each $i \in I_0$, we say that a wall of $\mathbb{Q}_0$ is the *wall of type* i, or i*-type wall* of $\mathbb{Q}_0$ if it lies in H_i. This is the case if and only if the wall contains the co-type i panel of C_0. For every $i \in I_0$, denote by $H_{0,i}$ the i-type wall of $\mathbb{Q}_0$.

We extend this definition to each sector of $\mathbb{A}$ by declaring that, for every special vertex X_λ in $\mathbb{A}$, and for every chamber C sharing X_λ, the walls of the sector $Q_\lambda(C)$ based at X_λ are the intersection with $\overline{Q_\lambda(C)}$ of any affine hyperplane H^k_α, $\alpha \in R^+$, $k \in \mathbb{Z}$, which is a wall of the chamber C. Moreover we say that a wall of $Q_\lambda(C)$ has type i for some $i \in I_0$ if there is a type-preserving isomorphism on $\mathbb{A}$ mapping the wall in an affine hyperplane $H^k_i = H^k_{\alpha_i}$ for some $i \in I_0$ and $k \in \mathbb{Z}$.

The definition of wall can be extended to each sector of the building; indeed, if $Q_x(c)$ is any sector of Δ, and $\mathscr{A}$ is any apartment of the building containing $Q_x(c)$, then the *walls* of $Q_x(c)$ are the inverse images of the walls of the sector $Q_\lambda(C) = \psi_{tp}(Q_x(c))$, under a type-preserving isomorphism $\psi_{tp} : \mathscr{A} \to \mathbb{A}$. Moreover, for every $i \in I_0$, a wall of $Q_x(c)$ has type i, if its image in $\mathbb{A}$ has type i. The previous definition does not depend on the choice of the apartment $\mathscr{A}$ containing the sector and of the type-preserving isomorphism $\psi_{tp} : \mathscr{A} \to \mathbb{A}$. For every sector $Q_x(c)$ and for every $i \in I_0$, denote by $h_{x,i}(c) = h_{x,i}(Q_x(c))$ the wall of type i of the sector. If ω is any element of the maximal boundary Ω, then, for every $x \in \mathscr{V}_{sp}(\Delta)$ and $i \in I_0$, we simply write $h_{x,i}(\omega)$ for the wall of type i of the sector $Q_x(\omega)$. If α is a simple root, that is, $\alpha = \alpha_i$ for some $i \in I_0$, then for every special vertex x of Δ and for every $\omega \in \Omega$ we denote by $h_{x,\alpha}(\omega)$ the wall of $Q_x(\omega)$ of type i, and refer to it simply as the α-wall of $Q_x(\omega)$. In general, for every simple root α, we denote by $h_{x,\alpha}$ the α-wall of any sector based at x.

Definition 4.1 Let $x, y \in \mathscr{V}_{sp}(\Delta)$, $x \neq y$. Let $h_{x,\alpha}$ and $h_{y,\alpha}$ be α-walls, based at x and y respectively.

(i) The walls $h_{x,\alpha}$ and $h_{y,\alpha}$ are said to be *equivalent* if they eventually coincide, i.e., if there is $h_{z,\alpha}$ such that $h_{z,\alpha} \subset h_{x,\alpha} \cap h_{y,\alpha}$.
(ii) The walls $h_{x,\alpha}$ and $h_{y,\alpha}$ are said to be *parallel* if they are not equivalent, but are both contained in the same apartment and, through any type-preserving isomorphism ψ_{tp} of this apartment onto $\mathbb{A}$, they correspond to walls of $\mathbb{A}$ that lie in parallel affine α-hyperplanes H_α^k, H_α^j for some $k, j \in \mathbb{Z}$.
(iii) The walls $h_{x,\alpha}$ and $h_{y,\alpha}$ are said to be *eventually parallel* if there exist $h_{x',\alpha} \subset h_{x,\alpha}$ and $h_{y',\alpha} \subset h_{y,\alpha}$ which are parallel. If $h_{x,\alpha}$ and $h_{y,\alpha}$ are eventually parallel, we call *distance* between the two walls the usual distance between the two hyperplanes of $\mathbb{A}$ that contain the images of their parallel sub-walls, that is, the positive integer $|j-k|$, if j and k are such that $\psi_{tp}(h_{x,\alpha}) = H_\alpha^k$ and $\psi_{tr}(h_{y,\alpha}) = H_\alpha^j$.

Proposition 4.2 *For every $\omega \in \Omega$ and for every pair of special vertices $x, y \in \mathscr{V}_{sp}(\Delta)$, the walls $h_{x,\alpha}(\omega)$ and $h_{y,\alpha}(\omega)$ are equivalent or eventually parallel.*

Proof Fix $\omega \in \Omega$, $x \neq y$ in $\mathscr{V}_{sp}(\Delta)$ and consider the α-walls $h_{x,\alpha}(\omega)$ and $h_{y,\alpha}(\omega)$. We assume that $h_{x,\alpha}(\omega)$ and $h_{y,\alpha}(\omega)$ are not equivalent and prove that they are eventually parallel. Observe that, if there exists an apartment $\mathscr{A}$ that contains $h_{x,\alpha}(\omega)$ and $h_{y,\alpha}(\omega)$, then the two walls are parallel. Actually, if ω' denotes a boundary point lying into the apartment $\mathscr{A}$ such that $h_{x,\alpha}(\omega) = h_{x,\alpha}(\omega')$, then $\rho_{\omega'}^x$ is a type-rotating isomorphism from $\mathscr{A}$ onto $\mathbb{A}$ such that $\rho_{\omega'}^x(h_{x,\alpha}(\omega))$ lies in H_α and $\rho_{\omega'}^x(h_{y,\alpha}(\omega))$ lies in H_α^k for some $k \in \mathbb{Z}$. Hence, in order to prove that $h_{x,\alpha}(\omega)$ and $h_{y,\alpha}(\omega)$ are eventually parallel, we only have to show that there exists an apartment $\mathscr{A}$ that contains sub-walls $h_{x',\alpha}(\omega) \subset h_{x,\alpha}(\omega)$ and $h_{y',\alpha}(\omega) \subset h_{y,\alpha}(\omega)$. We prove this by induction with respect to the distance between x and y.

We consider at first the case when $\mathscr{V}_{sp}(\Delta)$ contains vertices of different types. This happens for every building of type different from $\widetilde{G}_2$. If $\text{dist}(x, y) = 1$, the vertices x and y are adjacent. Then there exists a chamber c such that $x, y \in c$. If $\mathscr{A}$ is an apartment that contains ω and c, we have $Q_x(\omega), Q_y(\omega) \subset \mathscr{A}$. Thus $h_{x,\alpha}(\omega)$, $h_{y,\alpha}(\omega)$ lie in $\mathscr{A}$. Moreover the distance between $h_{x,\alpha}(\omega)$ and $h_{y,\alpha}(\omega)$ is zero or one.

By induction, assume that, when $\text{dist}(x, y) \leqslant n$, the walls $h_{x,\alpha}(\omega)$ and $h_{y,\alpha}(\omega)$ have sub-walls $h_{x',\alpha}(\omega)$ and $h_{y',\alpha}(\omega)$ lying in the same apartment. Then $h_{x',\alpha}(\omega)$ and $h_{y',\alpha}(\omega)$ are parallel and their distance is less than or equal to n. Actually we may assume, without loss of generality, that $\text{dist}(x', y') \leqslant n$. Let now $\text{dist}(x, y) = n + 1$ and choose z such that $\text{dist}(y, z) = 1$ and $\text{dist}(x, z) = n$. By the inductive hypothesis, there exist x', z' with $\text{dist}(x', z') = n$ such that the sub-walls $h_{x',\alpha}(\omega) \subset h_{x,\alpha}(\omega)$ and $h_{z',\alpha}(\omega) \subset h_{z,\alpha}(\omega)$ lie in the same apartment $\mathscr{A}_1$ and are parallel, at distance less than or equal to n. Without loss of generality, we may assume, for the sake of simplicity, that $x' = x$ and $z' = z$. Moreover, if c is a chamber such that $y, z \in c$, then there exists an apartment $\mathscr{A}_2$ that contains $h_{y,\alpha}(\omega)$, $h_{z,\alpha}(\omega)$ and c. We prove that there exists an apartment $\mathscr{A}$ that contains $h_{x,\alpha}(\omega)$,

$h_{z,\alpha}(\omega)$ and $h_{y,\alpha}(\omega)$. If $h_{y,\alpha}(\omega)$ lies in $\mathscr{A}_1$, then $\mathscr{A}_2 = \mathscr{A}_1$, the required apartment is $\mathscr{A}_1$ and, in this apartment, the distance between the parallel hyperplanes $h_{x,\alpha}(\omega)$, $h_{y,\alpha}(\omega)$ is less than or equal to n. If, on the contrary, $h_{y,\alpha}(\omega)$ does not lie in $\mathscr{A}_1$, we consider two isomorphisms $\psi_1 : \mathscr{A}_1 \to \mathbb{A}$ and $\psi_2 : \mathscr{A}_2 \to \mathbb{A}$ such that $\psi_1(h_{z,\alpha}(\omega)) = \psi_2(h_{z,\alpha}(\omega)) = H_{0,\alpha}$. Then,

$$\psi_1\big(h_{x,\alpha}(\omega)\big) = H_{h,\alpha}, \qquad \psi_2\big(h_{y,\alpha}(\omega)\big) = H_{k,\alpha},$$

for some $h, k \in \mathbb{Z}$. When $hk < 0$, then $H_{h,\alpha}$ and $H_{k,\alpha}$ lie in distinct half-apartments $\mathbb{A}^+_{0,\alpha}$, $\mathbb{A}^-_{0,\alpha}$, say $H_{h,\alpha} \subset \mathbb{A}^+_{0,\alpha}$ and $H_{k,\alpha} \subset \mathbb{A}^-_{0,\alpha}$: in this case consider the apartment $\mathscr{A} = \psi^{-1}(\mathbb{A})$, if $\psi = \psi_1$ on $\mathbb{A}^+_{0,\alpha}$ and $\psi = \psi_2$ on $\mathbb{A}^-_{0,\alpha}$. On the contrary, when $hk > 0$, then $H_{h,\alpha}$ and $H_{k,\alpha}$ lie in a same half-apartment $\mathbb{A}^+_{0,\alpha}$ or $\mathbb{A}^-_{0,\alpha}$, say $H_{h,\alpha}$, $H_{k,\alpha} \subset \mathbb{A}^+_{0,\alpha}$: in this case consider the apartment $\mathscr{A} = \psi^{-1}(\mathbb{A})$, if $\psi = \psi_1$ on $\mathbb{A}^+_{0,\alpha}$ and $\psi = \psi_2 s_\alpha$ on $\mathbb{A}^-_{0,\alpha}$. In both cases $\mathscr{A}$ is the required apartment that contains $h_{x,\alpha}(\omega)$, $h_{z,\alpha}(\omega)$ and $h_{y,\alpha}(\omega)$.

Assume now that Δ has type $\widetilde{G_2}$. In this case, all special vertices have type 0 and we cannot choose x, y adjacent. However, if we choose as x and y the vertices of type 0 of two adjacent chambers c, c', it is a consequence of the geometry of the building that the walls $h_{x,\alpha}(\omega)$, $h_{y,\alpha}(\omega)$ are eventually parallel and have distance 0 or 1. Therefore the result follows from the same inductive argument as before. □

We point out that, if Δ has type $\widetilde{C}_n$ or $\widetilde{BC}_n$, then a wall of type n of any sector of the building contains special vertices of only one type, that is,only of type 0 or only of type n (the same is true for a wall of type i, $i < n$, of a building of type $\widetilde{B}_n$).

From now on we shall limit attention to walls based at special vertices of the set $\widehat{\mathscr{V}}(\Delta)$.

4.2 The α-Boundary Ω_α

Let α be a simple root, that is, $\alpha = \alpha_i$ for some $i \in I_0$. For every special vertex x of $\widehat{\mathscr{V}}(\Delta)$ and for every $\omega \in \Omega$, we consider the α-wall $h_{x,\alpha}(\omega)$ of $Q_x(\omega)$.

Lemma 4.3 *Let $\omega_1, \omega_2 \in \Omega$. If there exists a vertex $x \in \widehat{\mathscr{V}}(\Delta)$ such that $h_{x,\alpha}(\omega_1) = h_{x,\alpha}(\omega_2)$, then $h_{y,\alpha}(\omega_1) = h_{y,\alpha}(\omega_2)$ for every $y \in \widehat{\mathscr{V}}(\Delta)$.*

Proof We split the proof in two parts.

(i) At first assume that there exists an apartment $\mathscr{A}$ that contains $Q_x(\omega_1)$ and $Q_x(\omega_2)$. Since $h_{x,\alpha}(\omega_1) = h_{x,\alpha}(\omega_2)$, there exists a type-rotating isomorphism $\psi_{tr} : \mathscr{A} \to \mathbb{A}$, mapping $Q_x(\omega_1)$ onto $\mathbb{Q}_0$ and $Q_x(\omega_2)$ onto $s_\alpha\mathbb{Q}_0$. Hence the same property holds for each $y \in \mathscr{A}$. This proves that $h_{y,\alpha}(\omega_1) = h_{y,\alpha}(\omega_2)$ for every $y \in \mathscr{A}$. On the other hand, if $y \notin \mathscr{A}$, the sectors $Q_y(\omega_1)$ and $Q_y(\omega_2)$ do not lie in $\mathscr{A}$, but there exists $z \in \mathscr{A}$ such that $Q_z(\omega_1) \subset Q_y(\omega_1)$, $Q_z(\omega_2) \subset Q_y(\omega_2)$ and $h_{z,\alpha}(\omega_1) = h_{z,\alpha}(\omega_2)$. Therefore $Q_y(\omega_1) \cap Q_y(\omega_2)$ contains $h_{z,\alpha}(\omega_1) = h_{z,\alpha}(\omega_2)$, besides y. This implies that $Q_y(\omega_1) \cap Q_y(\omega_2)$ contains the convex hull of y and

$h_{z,\alpha}(\omega_1) = h_{z,\alpha}(\omega_2)$, which includes the wall of type α of the two sectors; thus $h_{y,\alpha}(\omega_1) = h_{y,\alpha}(\omega_2)$.

(ii) If no apartment contains both $Q_x(\omega_1)$ and $Q_x(\omega_2)$, then there exists a vertex z such that $Q_z(\omega_1) \subset Q_x(\omega_1)$ and $Q_z(\omega_2) \subset Q_x(\omega_2)$, and $Q_z(\omega_1)$ and $Q_z(\omega_2)$ lie in some apartment $\mathscr{A}$; moreover $h_{z,\alpha}(\omega_1) = h_{z,\alpha}(\omega_2)$. The proof now follows from the same argument as in (i). □

Definition 4.4 Let $\omega, \omega' \in \Omega$. We say that ω is α*-equivalent* to ω', and we write $\omega \sim_\alpha \omega'$, if, for some x, $h_{x,\alpha}(\omega) = h_{x,\alpha}(\omega')$.

Lemma 4.3 implies that the definition of α-equivalence does not depend on the vertex x such that $h_{x,\alpha}(\omega) = h_{x,\alpha}(\omega')$. Moreover, if ω is α-equivalent to ω', and $\mathscr{A} = \mathscr{A}(\omega, \omega')$ denotes any apartment having ω and ω' as boundary points, then for every $x \in \mathscr{A}$, the sectors $Q_x(\omega)$ and $Q_x(\omega')$ are α-adjacent, that is, there exists a type rotating isomorphism $\psi_{tr} : \mathscr{A} \to \mathbb{A}$, mapping $Q_x(\omega)$ onto $\mathbb{Q}$ and $Q_x(\omega')$ onto $s_\alpha \mathbb{Q}_0$. On the contrary, if x does not lie in any $\mathscr{A}(\omega, \omega')$, then the intersection of $Q_x(\omega)$ and $Q_x(\omega')$ contains properly their common α-wall.

Definition 4.5 We call α*-boundary* of the building Δ the set $\Omega_\alpha = \Omega/{\sim_\alpha}$, consisting of all equivalence classes $[\omega]_\alpha$ of boundary points and we denote by η_α any element of Ω_α. Hence $\eta_\alpha = [\omega]_\alpha$, if ω belongs to the equivalence class η_α.

Fix $\omega \in \Omega$ and consider the set $\mathscr{H}_\alpha(\omega) = \{h_{x,\alpha}(\omega), x \in \widehat{\mathscr{V}}(\Delta)\}$. If $\omega' \sim_\alpha \omega$ then, for every x, $h_{x,\alpha}(\omega') = h_{x,\alpha}(\omega)$ and hence $\mathscr{H}_\alpha(\omega) = \mathscr{H}_\alpha(\omega')$. Therefore the set $\mathscr{H}_\alpha(\omega)$ only depends on the equivalence class $\eta_\alpha = [\omega]_\alpha$ represented by ω and we shall denote $\mathscr{H}(\eta_\alpha) = \mathscr{H}_\alpha(\omega)$, if $\omega \in \eta_\alpha$. Moreover, if $\omega \not\sim_\alpha \omega'$, then, for every $x \in \widehat{\mathscr{V}}(\Delta)$, $h_{x,\alpha}(\omega) \neq h_{x,\alpha}(\omega')$ and so $\mathscr{H}_\alpha(\omega) \cap \mathscr{H}_\alpha(\omega') = \emptyset$. This implies that the map

$$\eta_\alpha \to \mathscr{H}(\eta_\alpha)$$

is a bijection between the α-boundary Ω_α and the set $\{\mathscr{H}(\eta_\alpha)\}$. In particular, for every $x \in \widehat{\mathscr{V}}(\Delta)$, each element η_α of Ω_α determines one α-wall based at x: we denote this wall by $h_x(\eta_\alpha)$. Of course, $h_x(\eta_\alpha) = h_{x,\alpha}(\omega)$ for every $\omega \in \eta_\alpha$.

4.3 Trees at Infinity

Let us consider the α-boundary Ω_α that corresponds to a simple root α of the building. We claim that it is possible to construct a graph associated to each element η_α of Ω_α, and this graph is in fact a tree, whose boundary can be canonically identified with the set of all ω belonging to the class η_α. For this purpose we look in detail at the set $\mathscr{H}(\eta_\alpha)$ for any class η_α, and we show how this set gives rise to a tree. Proposition 4.2 of Sect. 4.1 implies the following corollary.

Corollary 4.6 *For every $\eta_\alpha \in \Omega_\alpha$, the set $\mathscr{H}(\eta_\alpha)$ consists of walls equivalent or eventually parallel.*

Let η_α be a fixed element of Ω_α. For every $x \in \widehat{\mathscr{V}}(\Delta)$, consider the wall $h_x(\eta_\alpha)$ of $\mathscr{H}(\eta_\alpha)$ and the class of all walls $h_{x'}(\eta_\alpha)$ equivalent to $h_x(\eta_\alpha)$, according to part (i) of Definition 4.1 of Sect. 4.1. We simply write $\mathbf{x}$ for this equivalence class represented by the wall $h_x(\eta_\alpha)$. Obviously, $\mathbf{x} = \mathbf{y}$ if and only if $h_x(\eta_\alpha)$ and $h_y(\eta_\alpha)$ are equivalent.

Remark 4.7 Consider, in the fundamental apartment $\mathbb{A}$, the α-wall of any sector Q_X equivalent to $\mathbb{Q}_0$. Each of these walls lies in an affine hyperplane H_α^k for some $k \in \mathbb{Z}$. For every $k \in \mathbb{Z}$, we write in short as $\mathbf{X}_k$ the class of all walls lying in H_α^k, and set

$$\Gamma_0 = \{\mathbf{X}_k, k \in \mathbb{Z}\}.$$

For every apartment $\mathscr{A}$ of the building and for any η_α, we consider the walls of $\mathscr{H}(\eta_\alpha)$ which lie in $\mathscr{A}$, and the equivalence classes $\mathbf{x}$ represented by these walls. By a type-preserving isomorphism $\psi_{tp} : \mathscr{A} \to \mathbb{A}$, each $\mathbf{x}$ maps to an element $\mathbf{X}_k$ of Γ_0 for some $k \in \mathbb{Z}$.

Recall that, if the root system R has type C_n or BC_n and $\alpha = \alpha_n$, then, for every $j \in \mathbb{Z}$, H_α^{2j} only contains special vertices of type 0 and H_α^{2j+1} only contains special vertices of type n (the same is true if R has type B_n and $\alpha = \alpha_i$, $i < n$). Hence in this case it is natural to endow the set Γ_0 with a labeling in the following way: we say that $\mathbf{X}_k$ has type 0 if $k = 2j$ and has type 1 if $k = 2j + 1$, for $j \in \mathbb{Z}$. This labeling can be extended as follows to all equivalence classes $\mathbf{x}$ represented by walls of $\mathscr{H}(\eta_\alpha)$ lying in any apartment $\mathscr{A}$, and hence to all walls of the building: we say that $\mathbf{x}$ has type 0 if it maps to some $\mathbf{X}_{2j}$ through any type-preserving isomorphism, and that it has type 1 if it maps to some $\mathbf{X}_{2j+1}$.

Definition 4.8 Let $\eta_\alpha \in \Omega_\alpha$. We denote by $T(\eta_\alpha)$ the graph having as *vertices* the classes $\mathbf{x}$ of equivalent walls associated to η_α, and as *edges* the pairs $[\mathbf{x}, \mathbf{y}]$ of equivalence classes represented by (eventually parallel) walls $h_x(\eta_\alpha)$ and $h_y(\eta_\alpha)$ at distance one.

Then, for every $\omega \in \eta_\alpha$ we associate to ω the graph $T_\alpha(\omega) = T(\eta_\alpha)$ and for every $\omega \in \Omega$ we associate to ω the graph of the element η_α of the α-boundary represented by ω.

Recall that, according to notation introduced in Sect. 2.16, the simple root α belongs to R_2 if and only if R is not reduced and $\alpha = \alpha_n = e_n$. In this particular case, $H_\alpha^k = H_{2\alpha}^{2k}$ for every $k \in \mathbb{Z}$: therefore the parallel hyperplanes of $\mathbb{A}$ orthogonal to α are the hyperplanes $H_{2\alpha}^h$ for all $h \in \mathbb{Z}$. Moreover, for every $k \in \mathbb{Z}$,

$$q_{2\alpha,2k} = q_{\alpha,k} = q_\alpha = r, \qquad q_{2\alpha,2k+1} = q_{2\alpha} = p.$$

In all other cases, that is, for all simple root of a reduced building or for all simple root α_i, $i \neq n$, of a building of type $\widetilde{BC}_n$, we always have $\alpha \in R_0$, hence

$$q_{\alpha,k} = q_\alpha \quad \text{for every } k \in \mathbb{Z}.$$

Proposition 4.9 *For every simple root α and for every $\eta_\alpha \in \Omega_\alpha$, the graph $T(\eta_\alpha)$ is a tree.*

(i) *If $\alpha \in R_0$, the tree is homogeneous, with homogeneity q_α.*
(ii) *If $\alpha \in R_2$, the tree is labeled and semi-homogeneous; each vertex of type 0 shares $q_{2\alpha} = p$ edges and each vertex of type 1 shares $q_\alpha = r$ edges.*

Proof We have to prove that $T(\eta_\alpha)$ is connected and has no loops.

Let $\mathbf{x}$, $\mathbf{y}$ be two vertices of the graph, let $\omega \in \eta_\alpha$ and $h_{x,\alpha}(\omega)$, $h_{y,\alpha}(\omega)$ are representatives of $\mathbf{x}$ and $\mathbf{y}$ respectively. We may assume, without loss of generality, that the two walls are parallel, hence that they lie in an apartment $\mathscr{A}$. Let n be the distance between the two walls on this apartment. We can choose $x_0, x_1, \dots, x_n$ on $\mathscr{A}$ such that $x_0 \in h_{x,\alpha}(\omega)$, $x_n \in h_{y,\alpha}(\omega)$ and $\mathrm{dist}(x_{i-1}, x_i) = 1$ for every $i = 1, \dots, n$. The walls $h_{x_0,\alpha}(\omega), h_{x_1,\alpha}(\omega), \dots, h_{x_n,\alpha}(\omega)$ are pairwise adjacent on $\mathscr{A}$ and

$$h_{x_0,\alpha}(\omega) \sim h_{x,\alpha}(\omega), \qquad h_{x_n,\alpha}(\omega) \sim h_{y,\alpha}(\omega).$$

Therefore, if $\mathbf{x}_i$ is the vertex of the graph represented by $h_{x_i,\alpha}(\omega)$, for $i = 0, \dots, n$, then $\mathrm{dist}(\mathbf{x}_{i-1}, \mathbf{x}_i) = 1$ for $i = 0, \dots, n$, and $\mathbf{x} = \mathbf{x}_0$, $\mathbf{y} = \mathbf{x}_n$. This proves that $\mathbf{x}$, $\mathbf{y}$ are connected by a path of length n.

For every $n \geqslant 2$, consider a path $\mathbf{x}_0, \dots, \mathbf{x}_n$ in the graph such that $\mathbf{x}_{i-1} \neq \mathbf{x}_i, \mathbf{x}_{i+1}$, for $i = 1, \dots, n-1$. We prove by induction that $\mathbf{x}_0 \neq \mathbf{x}_n$. If $n = 2$, the property holds by definition; let us assume the property is true for $n - 1$ and show that it is true also for n. Indeed, if $h_{x_0,\alpha}(\omega), \dots, h_{x_{n-1},\alpha}(\omega), h_{x_n,\alpha}(\omega)$ are representatives of the vertices $\mathbf{x}_0, \dots, \mathbf{x}_{n-1}, \mathbf{x}_n$ respectively, we know that there exists an apartment $\mathscr{A}$ that contains all the walls $h_{x_0,\alpha}(\omega), \dots, h_{x_{n-1},\alpha}(\omega)$ and in this apartment the distance between $h_{x_0,\alpha}(\omega)$ and $h_{x_{n-1},\alpha}(\omega)$ is $n-1$. On the other hand, the apartment $\mathscr{A}$ can be chosen in such a way that also the wall $h_{x_n,\alpha}(\omega)$ lies in it. In this apartment, $\mathrm{dist}(h_{x_0,\alpha}(\omega), h_{x_n,\alpha}(\omega)) = n$, as $h_{x_n,\alpha}(\omega) \neq h_{x_{n-2},\alpha}(\omega)$. This proves that $\mathbf{x}_0 \neq \mathbf{x}_n$.

Finally, if R is not reduced and $\alpha = \alpha_n = e_n$, the parallel hyperplanes of $\mathbb{A}$, orthogonal to α, are the hyperplanes $H^k_{2\alpha}$, for all $k \in \mathbb{Z}$. Moreover, for every $j \in \mathbb{Z}$,

$$q_{2\alpha,2j} = q_{\alpha,k} = q_\alpha = r, \qquad q_{2\alpha,2j+1} = q_{2\alpha} = p.$$

Hence, in this case the number of edges sharing any vertex $\mathbf{x}$ of type 0 is r, while the number of edges sharing the vertex $\mathbf{y}$ is p.

In all other cases, that is, for all simple roots of a reduced building or for all simple roots α_i, $i \neq n$, for a building of type $\widetilde{BC}_n$, we always have $\alpha \in R_0$, and so

$$q_{\alpha,k} = q_\alpha \quad \text{for every } k \in \mathbb{Z}.$$

Therefore, each wall $h_{x,\alpha}(\omega)$ is adjacent to q_α walls $h_{y,\alpha}(\omega)$. Hence each vertex $\mathbf{x}$ belongs to q_α edges. □

Remark 4.10 For every apartment $\mathscr{A}$, the walls $h_{x,\alpha}(\omega)$ of $\mathscr{H}(\eta_\alpha)$ which lie in $\mathscr{A}$, determine a geodesic $\gamma(\eta_\alpha)$ of the tree $T(\eta_\alpha)$, that consists of all vertices $\mathbf{x}$ associated to these walls and of all edges connecting each pair of adjacent vertices $\mathbf{x}$, $\mathbf{y}$.

The set Γ_0 can be seen as the *fundamental geodesic* of the tree, since each geodesic $\gamma(\eta_\alpha)$ of the building is isomorphic to Γ_0 through any type-preserving isomorphism $\psi_{tp} : \mathscr{A} \to \mathbb{A}$, if $\mathscr{A}$ denotes any apartment that contains $\gamma(\eta_\alpha)$.

The tree $T(\eta_\alpha)$ is labeled and semi-homogeneous only when R is not reduced and $\alpha = \alpha_n = e_n$, i.e., only when the building has type $\widetilde{BC}_n$: in this case $\widehat{\mathscr{V}}(\Delta)$ consists only of vertices of type 0. Therefore for such a tree it is natural to restrict attention to its vertices of type 0. Now, if $\mathbf{x}, \mathbf{y}$ are vertices of type 0, then the geodesic $[\mathbf{x}, \mathbf{y}]$ has length $2n$, for some $n \in \mathbb{N}$. Moreover in the fundamental geodesic Γ_0 we consider only the vertices X_{2n}, for $n \in \mathbb{N}$.

Proposition 4.9 shows that, for every element $\eta_\alpha \in \Omega_\alpha$, we may identify the set $\mathscr{H}(\eta_\alpha)$ with a tree $T(\eta_\alpha)$. Moreover trees $T(\eta_{\alpha,1})$, $T(\eta_{\alpha,2})$ associated to any two $\eta_{\alpha,1}$, $\eta_{\alpha,2}$ in Ω_α are isomorphic. For every $x \in \widehat{\mathscr{V}}(\Delta)$, the vertex $\mathbf{x}$ can be seen as the projection of x onto the tree $T(\eta_\alpha)$. In this sense we can refer to $T(\eta_\alpha)$ as to *the tree at infinity* associated to the element η_α of the α-boundary.

Proposition 4.11 *For every $\eta_\alpha \in \Omega_\alpha$, the set*

$$\{\omega \in \Omega : \omega \in \eta_\alpha\}$$

can be identified with the boundary $\partial T(\eta_\alpha)$ of the tree $T(\eta_\alpha)$.

Proof We fix $x \in \widehat{\mathscr{V}}(\Delta)$. For every ω in the class $\eta_\alpha = [\omega]_\alpha$, we consider the sector $Q_x(\omega)$ based at x and its wall $h_{x,\alpha}(\omega)$. Let us denote by $h_{x_j,\alpha}(\omega)$, $j \geqslant 0$, a sequence of walls lying in $Q_x(\omega)$ such that

$$h_{x_0,\alpha}(\omega) = h_{x,\alpha}(\omega) \quad \text{and} \quad \operatorname{dist}\bigl(h_{x_j,\alpha}(\omega), h_{x_{j+1},\alpha}(\omega)\bigr) = 1, \qquad j \geqslant 0.$$

The sequence $\mathbf{x}_j$, $j \geqslant 0$, is a geodesic of the tree $T(\eta_\alpha)$ starting at $\mathbf{x}_0 = \mathbf{x}$ and hence it determines, as usual, a boundary point $\bar{\omega}$ of the tree. The map $\omega \to \bar{\omega}$ is a bijection of $\eta_\alpha = [\omega]_\alpha$ onto $\partial T(\eta_\alpha)$, since each boundary point of the tree can be obtained from a suitable ω in the class η_α with the procedure described before, and $\bar{\omega}_1 \neq \bar{\omega}_2$ if $\omega_1 \neq \omega_2$ are two elements of the same class η_α. □

Since the trees $T(\eta_{\alpha,1})$, $T(\eta_{\alpha,2})$ associated to any two $\eta_{\alpha,1}$, $\eta_{\alpha,2}$ in Ω_α are isomorphic, the same is true for their boundaries $\partial T(\eta_{\alpha,1})$, $\partial T(\eta_{\alpha,2})$. We denote by T_α an abstract tree such that

$$T(\eta_\alpha) \sim T_\alpha, \quad \text{for every } \eta_\alpha \in \Omega_\alpha.$$

Moreover, we denote by $\mathbf{t}$ any element of T_α and by $\mathbf{b}$ any element of its boundary ∂T_α.

As a consequence of Proposition 4.11, the maximal boundary Ω of the building can be decomposed as a disjoint union of boundaries of trees, one for each equivalence class $\eta_\alpha = [\omega]_\alpha$:

$$\Omega = \bigcup_{\eta_\alpha \in \Omega_\alpha} \partial T(\eta_\alpha).$$

The previous decomposition implies that each boundary point ω of the building can be seen as a pair $(\eta_\alpha, \mathbf{b}) \in \Omega_\alpha \times \partial T_\alpha$, where η_α is the equivalence class $[\omega]_\alpha$ that contains ω and $\mathbf{b}$ is the boundary point of T_α that corresponds to $\bar{\omega}$ on $\partial T(\eta_\alpha)$. In this sense we may write, up to isomorphism,

$$\Omega = \Omega_\alpha \times \partial T_\alpha.$$

4.4 Orthogonal Decomposition with Respect to a Root α

Definition 4.12 Let s_α be the reflection with respect to the linear hyperplane H_α of $\mathbb{A}$. For every vector v of the Euclidean space supporting $\mathbb{A}$, set

$$P_\alpha(v) = \frac{v - s_\alpha v}{2}, \qquad Q_\alpha(v) = \frac{v + s_\alpha v}{2}.$$

By definition, $P_\alpha(v) + Q_\alpha(v) = v$ and $Q_\alpha(v) - P_\alpha(v) = s_\alpha v$. Moreover

$$P_\alpha(s_\alpha v) = -P_\alpha(v) \quad \text{and} \quad Q_\alpha(s_\alpha v) = Q_\alpha(v).$$

We observe that, for every v, $Q_\alpha(v)$ lies in H_α and $P_\alpha(v)$ is the component of the vector v along the direction orthogonal to the hyperplane H_α, that is, in the direction of the vector α.

Proposition 4.13 *Let ω_1, ω_2 be α-equivalent. Then, for every $x, y \in \widehat{\mathscr{V}}(\Delta)$,*

$$Q_\alpha\big(\rho_{\omega_2}(y) - \rho_{\omega_2}(x)\big) = Q_\alpha\big(\rho_{\omega_1}(y) - \rho_{\omega_1}(x)\big).$$

If x, y belong to an apartment that contains both the boundary points ω_1, ω_2, then

$$P_\alpha\big(\rho_{\omega_2}(y) - \rho_{\omega_2}(x)\big) = -P_\alpha\big(\rho_{\omega_1}(y) - \rho_{\omega_1}(x)\big).$$

Proof Let $x, y \in \widehat{\mathscr{V}}(\Delta)$ and $\eta_\alpha = [\omega]_\alpha$, for $\omega \in \Omega$. Consider the tree $T(\eta_\alpha)$ and let $\mathbf{x}$ and $\mathbf{y}$ be the vertices of this tree, associated to x and y respectively.

If $\mathbf{x} = \mathbf{y}$, the walls $h_{x,\alpha}(\omega)$ and $h_{y,\alpha}(\omega)$ are equivalent, hence they intersect in a wall $h_{z,\alpha}(\omega)$. In this case, $Q_\alpha(\rho_\omega(y) - \rho_\omega(x))$ is given by the difference between $\sigma(y, z)$ and $\sigma(x, z)$.

Assume now $\mathbf{x} \neq \mathbf{y}$. If $\mathbf{b}$ is the boundary point of the tree that corresponds to ω, we consider the geodesics $[\mathbf{x}, \mathbf{b}]$, $[\mathbf{y}, \mathbf{b}]$ from $\mathbf{x}$ and from $\mathbf{y}$ to $\mathbf{b}$ respectively. We denote by $\mathbf{z}$ the vertex of the tree such that $[\mathbf{z}, \mathbf{b}] = [\mathbf{x}, \mathbf{b}] \cap [\mathbf{y}, \mathbf{b}]$, and by z a vertex of the building corresponding to $\mathbf{z}$ such that $Q_z(\omega) \subset Q_x(\omega) \cap Q_y(\omega)$. In the case when $[\mathbf{y}, \mathbf{b}] \subset [\mathbf{x}, \mathbf{b}]$, then $\mathbf{z} = \mathbf{y}$, and hence $h_{z,\alpha}(\omega) \subset h_{y,\alpha}(\omega)$. Otherwise, $h_{z,\alpha}(\omega)$ and $h_{x,\alpha}(\omega)$ are eventually parallel; if $h_{x',\alpha}(\omega)$ is the sub-wall of $h_{x,\alpha}(\omega)$ parallel to $h_{z,\alpha}(\omega)$, it is easy to check that $Q_\alpha(\rho_\omega(y) - \rho_\omega(x))$ is given by the difference between $\sigma(y, z)$ and $\sigma(x, x')$. In the case when $[\mathbf{x}, \mathbf{b}] \subset [\mathbf{y}, \mathbf{b}]$, a similar argument shows that $Q_\alpha(\rho_\omega(y) - \rho_\omega(x))$ is given by the difference between $\sigma(y, y')$ and $\sigma(x, z)$, where y' is the vertex such that $h_{y',\alpha}(\omega)$ is the sub-wall of $h_{y,\alpha}(\omega)$ parallel to $h_{z,\alpha}(\omega)$. Finally, if $\mathbf{z} \neq \mathbf{x}$ and $\mathbf{z} \neq \mathbf{y}$, then both the walls $h_{x,\alpha}(\omega)$ and $h_{y,\alpha}(\omega)$ are eventually parallel to $h_{z,\alpha}(\omega)$. If we denote by $h_{x',\alpha}(\omega)$ and by $h_{y',\alpha}(\omega)$ the sub-wall of $h_{x,\alpha}(\omega)$ and of $h_{y,\alpha}(\omega)$ respectively, which are parallel to $h_{z,\alpha}(\omega)$, then $Q_\alpha(\rho_\omega(y) - \rho_\omega(x))$ is given by the difference between $\sigma(y, y')$ and $\sigma(x, x')$. In every case $Q_\alpha(\rho_\omega(y) - \rho_\omega(x))$ is a vector lying in the hyperplane H_α and it is the same for all boundary points α-equivalent to ω. Assume now that there exists an apartment that contains x, y and both the boundary points ω_1, ω_2. In this particular case, $\rho_{\omega_2}(y) - \rho_{\omega_2}(x) = s_\alpha(\rho_{\omega_1}(y) - \rho_{\omega_1}(x))$. Therefore in this case

$$P_\alpha\big(\rho_{\omega_2}(y) - \rho_{\omega_2}(x)\big) = -P_\alpha\big(\rho_{\omega_1}(y) - \rho_{\omega_1}(x)\big). \qquad \square$$

4.5 Topologies on Ω_α

As well as the maximal boundary, also each α-boundary Ω_α may be equipped with a totally disconnected compact Hausdorff topology. Let x, y be special vertices in $\widehat{\mathscr{V}}(\Delta)$; consider the set $\Omega(x, y)$, defined in Sect. 3.2. Define a subset of Ω_α in the following way:

$$\Omega_\alpha(x, y) = \big\{\eta_\alpha = [\omega]_\alpha, \omega \in \Omega(x, y)\big\}.$$

Let $x \in \widehat{\mathscr{V}}(\Delta)$; the family

$$\widetilde{\mathscr{B}}^x_\alpha = \Big\{\Omega_\alpha(x, y), y \in \widehat{\mathscr{V}}(\Delta), y \in \bigcup h_{x,\alpha}\Big\}$$

generates a (totally disconnected compact Hausdorff) topology on Ω_α. For every $\eta_\alpha \in \Omega_\alpha$, say $\eta_\alpha = [\omega]_\alpha$, a local base at η_α is given by

$$\widetilde{\mathscr{B}}_{x,\eta_\alpha} = \big\{\Omega_\alpha(x, y), y \in Q_x(\omega)\big\}.$$

Observe that there exists an α-wall based at x containing y if and only if $y \in \mathscr{V}_\lambda(x)$, with $\lambda \in H_{0,\alpha}$. Then, for every pair of vertices $x, y \in \widehat{\mathscr{V}}(\Delta)$ such that $y \in \mathscr{V}_\lambda(x)$, for any $\lambda \in H_{0,\alpha}$, we have

$$\Omega_\alpha(x, y) = \big\{\eta_\alpha \in \Omega_\alpha : y \in h_{x,\alpha}(\eta_\alpha)\big\}.$$

Moreover the family

$$\mathscr{B}^x_\alpha = \big\{\Omega_\alpha(x, y), y \in \mathscr{V}_\lambda(x), \lambda \in H_{0,\alpha}\big\}$$

generates the same topology on Ω_α as before; hence, for every $\eta_\alpha \in \Omega_\alpha$, a local base at η_α is given by

$$\mathscr{B}_{x,\eta_\alpha} = \big\{\Omega_\alpha(x, y), y \in h_{x,\alpha}(\eta_\alpha)\big\}.$$

It follows from the same argument used for the maximal boundary that the topology on Ω_α does not depend on the particular $x \in \widehat{\mathscr{V}}(\Delta)$.

4.6 Probability Measures on the α-Boundary

For every x of $\widehat{\mathscr{V}}(\Delta)$, we define a regular Borel measure ν_x^α on Ω_α, in the following way. For every $y \in \widehat{\mathscr{V}}(\Delta)$, let $\lambda = \sigma(x, y)$. Then $\sigma(\mathbf{x}, \mathbf{y}) = P_\alpha \lambda$ if $\mathbf{x}$ and $\mathbf{y}$ are the projections of x and y on the tree at infinity associated with any $\omega \in \Omega(x, y)$. Thus, define

$$\nu_x^\alpha\big(\Omega_\alpha(x, y)\big) = \frac{N^\alpha_{P_\alpha \lambda}}{N_\lambda},$$

if $N^\alpha_{P_\alpha\lambda} = |\{\mathbf{z} : \sigma(\mathbf{x}, \mathbf{z}) = P_\alpha \lambda\}|$. In fact the same argument used for the maximal boundary shows that there exists a unique regular Borel probability measure ν_x^α on Ω satisfying this property. Notice that, if $\lambda \in H_{0,\alpha}$, then $\mathbf{y} = \mathbf{x}$ and then $P_\alpha \lambda = 0$. Therefore in this case

$$\nu_x^\alpha\big(\Omega_\alpha(x, y)\big) = \nu_x\big(\Omega(x, y)\big).$$

Define

$$R_\alpha^+ = \big\{\beta \in R^+, \beta \neq \alpha, 2\alpha\big\}.$$

Then, recalling the formula for N_λ given in Corollary 2.7 of Sect. 2.16, we have

$$\nu_x^\alpha\big(\Omega_\alpha(x, y)\big) = \begin{cases} \frac{\mathbf{W}_\lambda(q^{-1})}{\mathbf{W}(q^{-1})} \prod_{\beta \in R_\alpha^+} q_\beta^{-\langle\lambda,\beta\rangle} q_{2\beta}^{\langle\lambda,\beta\rangle} & \text{if } \lambda \in H_{0,\alpha}, \\ \frac{\mathbf{W}_\lambda(q^{-1})(1+q_\alpha^{-1})}{\mathbf{W}(q^{-1})} \prod_{\beta \in R_\alpha^+} q_\beta^{-\langle\lambda,\beta\rangle} q_{2\beta}^{\langle\lambda,\beta\rangle} & \text{otherwise.} \end{cases}$$

4.7 Topologies and Probability Measures on the Trees at Infinity

Let T_α be the abstract tree isomorphic to each tree at infinity $T(\eta_\alpha)$ and let ∂T_α be its boundary. As usual, denote by $\widehat{\mathscr{V}}(T_\alpha)$ the set of all vertices of T_α when the tree is homogeneous, or the set of all vertices of type 0 when the tree is semi-homogeneous. For every $\mathbf{t} \in \widehat{\mathscr{V}}(T_\alpha)$ and every $\mathbf{b} \in \partial T_\alpha$, denote by $\gamma(\mathbf{t}, \mathbf{b})$ the geodesic from $\mathbf{t}$ to $\mathbf{b}$. For every $\mathbf{t}, \mathbf{t}' \in \widehat{\mathscr{V}}(T_\alpha)$, let $B(\mathbf{t}, \mathbf{t}') = \{\mathbf{b} \in \partial T_\alpha : \mathbf{t}' \in \gamma(\mathbf{t}, \mathbf{b})\}$. It is well known that, for every $\mathbf{t} \in \widehat{\mathscr{V}}(T_\alpha)$, the family

$$\mathscr{B}_\mathbf{t} = \big\{B\big(\mathbf{t}, \mathbf{t}'\big), \mathbf{t}' \in \widehat{\mathscr{V}}(T_\alpha)\big\},$$

generates a totally disconnected compact Hausdorff topology on ∂T_α. Moreover, for every element $\mathbf{b}$, a local base at $\mathbf{b}$ is given by

$$\mathscr{B}_{\mathbf{t},\mathbf{b}} = \big\{B\big(\mathbf{t}, \mathbf{t}'\big), \mathbf{t}' \in \gamma_\mathbf{t}(\mathbf{b})\big\}.$$

Denote by $\mu_\mathbf{t}$ the usual probability measure on ∂T_α associated with the isotropic random walk on T_α starting at the vertex $\mathbf{t}$. We refer the reader to [1, 4] for the definition of this measure. Recall that, in the homogeneous case, with homogeneity q_α, for every vertex $\mathbf{t}'$ one has

$$\mu_{\mathbf{t}}\big(B\big(\mathbf{t}, \mathbf{t}'\big)\big) = \frac{1}{q_\alpha + 1} q_\alpha^{1-n},$$

where n is the length of the finite geodesic $[\mathbf{t}, \mathbf{t}']$. On the other hand, in the semi-homogeneous case, with homogeneities p, r, for every vertex $\mathbf{t}'$ at distance $2n$ from $\mathbf{t}$, one has

$$\mu_{\mathbf{t}}\big(B\big(\mathbf{t}, \mathbf{t}'\big)\big) = \frac{1}{p(1+r)} (pr)^{1-n}.$$

Since, for every element $\eta_\alpha \in \Omega_\alpha$, the tree $T(\eta_\alpha)$ is isomorphic to the abstract tree T_α, all previous arguments apply to $\partial T(\eta_\alpha)$ if $\mathbf{t}$ is replaced by the projection $\mathbf{x}$ on $T(\eta_\alpha)$ of some $x \in \widehat{\mathscr{V}}(\Delta)$, and in particular $\mathbf{e}$ is the projection on $T(\eta_\alpha)$ of the fundamental vertex e of the building. We point out that, for every $x \in \widehat{\mathscr{V}}$, the measure $\mu_{\mathbf{x}}$ on $\partial T(\eta_\alpha)$ defined before can be seen as a measure on Ω supported on $[\omega]_\alpha$ if $\eta_\alpha = [\omega]_\alpha$. Actually, it is easy to check that, if $\eta_\alpha = [\omega]_\alpha$, then, through the identification of $\partial T(\eta_\alpha)$ with the subset $[\omega]_\alpha$ of the maximal boundary, the measure $\mu_{\mathbf{x}}$ determines a probability measure on $[\omega]_\alpha$.

4.8 Decomposition of the Measure ν_x

Let $x \in \widehat{\mathscr{V}}(\Delta)$, let $\mathbf{x}$ be its projection on the tree $T(\eta_\alpha)$ associated with an assigned $\omega \in \Omega$ and $\mathbf{t}$ the element of the abstract tree T_α which corresponds to the vertex $\mathbf{x}$. For the sake of simplicity, from now on we identify $\mathbf{t}$ with $\mathbf{x}$. We also identify the maximal boundary Ω with $\Omega_\alpha \times \partial T_\alpha$ according to Sect. 4.3, and thereby we claim that each probability measure ν_x splits as product of the probability measure ν_x^α on the α-boundary Ω_α and the canonical probability measure $\mu_{\mathbf{x}}$ on the boundary of the tree T_α. Le us show this. For $x, y \in \widehat{\mathscr{V}}(\Delta)$, consider the set $\Omega(x, y)$. If $\omega \in \Omega(x, y)$ and $\omega = (\eta_\alpha, \mathbf{b})$, then $\eta_\alpha \in \Omega_\alpha(x, y)$ and $\mathbf{b} \in B(\mathbf{x}, \mathbf{y})$. Hence

$$\Omega(x, y) = \Omega_\alpha(x, y) \times B(\mathbf{x}, \mathbf{y}).$$

Proposition 4.14 $\nu_x = \nu_x^\alpha \times \mu_{\mathbf{x}}$, *for every* $x \in \widehat{\mathscr{V}}(\Delta)$.

Proof Let $x, y \in \widehat{\mathscr{V}}(\Delta)$ and $y \in \mathscr{V}_\lambda(x)$. Let $\mathbf{x}$ and $\mathbf{y}$ be the projection of x and y on the tree at infinity associated with any $\omega \in \Omega(x, y)$. We prove that

$$\nu_x\big(\Omega(x, y)\big) = \nu_x^\alpha\big(\Omega_\alpha(x, y)\big)\mu_{\mathbf{x}}\big(B(\mathbf{x}, \mathbf{y})\big).$$

In the case $\lambda \in H_{0,\alpha}$, we already proved that $\nu_x(\Omega(x, y)) = \nu_x^\alpha(\Omega_\alpha(x, y))$. On the other hand, in this case $\mathbf{y} = \mathbf{x}$, therefore $B(\mathbf{x}, \mathbf{y}) = \partial T_\alpha$. Hence $\mu_{\mathbf{x}}(B(\mathbf{x}, \mathbf{y})) = 1$ and the statement follows. Assume now $\lambda \notin H_{0,\alpha}$; in this case $\mu_{\mathbf{x}}(B(\mathbf{x}, \mathbf{y})) = (N^\alpha_{P_\alpha \lambda})^{-1}$. Then the desired formula is a direct consequence of the definition of $\nu_x^\alpha(\Omega_\alpha(x, y))$. □

5 Characters and Poisson Kernels

In this section we state the basic features of the characters on the fundamental apartment of the building.

5.1 *Characters of* $\mathbb{A}$

We call *character* of $\mathbb{A}$ any multiplicative complex-valued function χ defined on $\widehat{L}$:

$$\chi(\lambda_1 + \lambda_2) = \chi(\lambda_1)\chi(\lambda_2), \quad \text{for } \lambda_1, \lambda_2 \in \widehat{L}.$$

We assume, without loss of generality, that a character of $\mathbb{A}$ is the restriction to $\widehat{L}$ of a multiplicative complex-valued function on $\mathbb{V}$. Denote by $\mathbb{X}(\widehat{L})$ the group of all characters of $\mathbb{A}$. If $n = \dim \mathbb{V}$, then $\mathbb{X}(\widehat{L}) \cong (\mathbb{C}^\times)^n$, and $\mathbb{X}(\widehat{L})$ can be equipped with the weak topology and the usual measure of $\mathbb{C}^n$.

The Weyl group $\mathbf{W}$ acts on $\mathbb{X}(\widehat{L})$ in the following way: for every $\mathbf{w} \in \mathbf{W}$ and for every $\chi \in \mathbb{X}(\widehat{L})$,

$$(\mathbf{w}\chi)(\lambda) = \chi\big(\mathbf{w}^{-1}(\lambda)\big), \quad \text{for all } \lambda \in \widehat{L}.$$

It is immediately seen that $\mathbf{w}\chi$ is a character: denote it simply by $\chi^{\mathbf{w}}$.

5.2 *The Fundamental Character* χ_0 *of* $\mathbb{A}$

We are interested in a particular character of $\mathbb{A}$.

Definition 5.1 We denote by χ_0 the following function on $\widehat{L}$:

$$\chi_0(\lambda) = \prod_{\alpha \in R^+} q_\alpha^{\langle \lambda, \alpha \rangle} q_{2\alpha}^{-\langle \lambda, \alpha \rangle}, \quad \text{for every } \lambda \in \widehat{L}.$$

Being α a linear functional on the vector space $\mathbb{V}$ supporting $\mathbb{A}$, the function χ_0 is a character of $\mathbb{A}$, called the *fundamental* character of $\mathbb{A}$. Observe that $\chi_0(\lambda) > 1$, for all $\lambda \in \widehat{L}^+$.

If R is reduced, then $2\alpha \notin R$ and therefore $q_{2\alpha} = 1$ for every $\alpha \in R$. Hence

$$\chi_0(\lambda) = \prod_{\alpha \in R^+} q_\alpha^{\langle \lambda, \alpha \rangle}.$$

In particular if R is reduced and all roots have the same length, that is, for buildings of type $\widetilde{A}_n$, $\widetilde{D}_n$, $\widetilde{E}_6$, $\widetilde{E}_7$ and $\widetilde{E}_8$, then $q_\alpha = q$ for every $\alpha \in R^+$ and

$$\chi_0(\lambda) = q^{\sum_{\alpha \in R^+} \langle \lambda, \alpha \rangle} = q^{2\langle \lambda, \delta \rangle},$$

where $\delta = \frac{1}{2}(\sum_{\alpha\in R^+}\alpha)$. Instead, if R is reduced but it contains long and short roots, then, denoting by α any long root and by β any short root and setting $\delta_l = \frac{1}{2}(\sum\alpha)$, $\delta_s = \frac{1}{2}(\sum\beta)$, we see that

$$\chi_0(\lambda) = q^{2\langle\lambda,\delta_l\rangle} p^{2\langle\lambda,\delta_s\rangle}.$$

This happens for buildings of type $\widetilde{B}_n$, $\widetilde{C}_n$, $\widetilde{F}_4$ and $\widetilde{G}_2$.

Assume now that R is not reduced, that is, the building has type $\widetilde{(BC)}_n$. In this case $R = R_0 \cup R_1 \cup R_2$. We denote by α, β and γ any root of R_0, R_1 and R_2 respectively. Then, keeping in mind that $R_2 = \{\beta/2, \beta \in R_1\}$, we now see that

$$\begin{aligned}\chi_0(\lambda) &= \prod_{\alpha\in R_0^+} q_\alpha^{\langle\lambda,\alpha\rangle} \prod_{\beta\in R_1^+} q_\beta^{\langle\lambda,\beta\rangle} \prod_{\gamma\in R_2^+} q_\gamma^{\langle\lambda,\gamma\rangle} q_{2\gamma}^{-\langle\lambda,\gamma\rangle}\\ &= \prod_{\alpha\in R_0^+} q_\alpha^{\langle\lambda,\alpha\rangle} \prod_{\beta\in R_1^+} q_\beta^{\langle\lambda,\beta\rangle} \prod_{\beta\in R_1^+} q_{\beta/2}^{\langle\lambda,\beta/2\rangle} q_\beta^{-\langle\lambda,\beta/2\rangle}\\ &= \prod_{\alpha\in R_0^+} q_\alpha^{\langle\lambda,\alpha\rangle} \prod_{\beta\in R_1^+} (q_{\beta/2}q_\beta)^{\langle\lambda,\beta/2\rangle} = q^{2\langle\lambda,\delta_0\rangle}(pr)^{\langle\lambda,\delta_1\rangle} \qquad (6)\end{aligned}$$

where $\delta_0 = \frac{1}{2}\sum\alpha$, $\delta_1 = \frac{1}{2}\sum\beta$.

Notice that, by Proposition 2.6 of Sect. 2.16,

$$\chi_0(\lambda) = q_{t_\lambda}, \quad \text{for every } \lambda \in \widehat{L}^+.$$

More generally, if λ is any element of $\widehat{L}$, and $t_\lambda = u_\lambda g_l$, with $u_\lambda = s_{i_1}\cdots s_{i_r}$, then the same argument used in Proposition 2.6 of Sect. 2.16 shows that,

$$\chi_0(\lambda) = \prod_{j\in J^+} q_{i_j} \cdot \prod_{j\in J^-} q_{i_j}^{-1},$$

where

$$J^+ = \{j : s_{i_1}\cdots s_{i_{j-1}}(C_0) \prec s_{i_1}\cdots s_{i_j}(C_0)\},$$
$$J^- = \{j : s_{i_1}\cdots s_{i_j}(C_0) \prec s_{i_1}\cdots s_{i_{j-1}}(C_0)\}.$$

In fact, notice that, when λ is dominant, then $J^- = \emptyset$ and thus $J^+ = \{1,\dots,n\}$. Therefore we obtain the previous formula for $\chi_0(\lambda)$.

We can easily compute the fundamental character in each simple co-root $\alpha^\vee$. We consider separately the reduced and non-reduced case.

Proposition 5.2 *Let R be a reduced root system*; *for every simple root α, then*

$$\chi_0(\alpha^\vee) = q_\alpha^2.$$

Proof Observe that, for every simple α, we have $\langle\alpha^\vee, \delta\rangle = 1$. This is a consequence of (13.3) in [5]. □

Proposition 5.3 *Let R be a non-reduced root system*; *then*

(i) $\chi_0(\alpha^\vee) = q^2$, *for every* $\alpha = e_i - e_{i+1}$, $i = 1, \dots, n-1$;
(ii) $\chi_0(\beta^\vee) = pr$, *for* $\beta = 2e_n$.

Proof We compute $\chi_0(\alpha^\vee)$ and $\chi_0(\beta^\vee)$ by using the equality (6) for $\chi_0(\lambda)$ given above.

(i) If $\alpha = \alpha_i = e_i - e_{i+1}$ for some $i = 1, \dots, n-1$, then $\alpha_i^\vee = \alpha_i$, and, by definition,

$$\chi_0(\alpha_i^\vee) = \chi_0(\alpha_i) = \Big(\prod_{\alpha\in R_0^+} q^{\langle\alpha_i,\alpha\rangle}\Big)\Big(\prod_{\beta\in R_1^+} p^{\langle\alpha_i,\beta\rangle}\Big(\frac{r}{p}\Big)^{\langle\alpha_i,\beta/2\rangle}\Big)$$
$$= q^{\langle\alpha_i,\sum_{\alpha\in R_0^+}\alpha\rangle} p^{\langle\alpha_i,\sum_{\beta\in R_1^+}\beta\rangle}\Big(\frac{r}{p}\Big)^{\langle\alpha_i,\sum_{\beta\in R_1^+}\beta/2\rangle}.$$

Notice that

$$\sum_{\alpha\in R_0^+}\alpha = 2\big[(n-1)e_1 + (n-2)e_2 + \cdots + e_{n-1}\big] \quad\text{and}\quad \sum_{\beta\in R_1^+}\beta = 2\sum_{k=1}^n e_k.$$

Hence, for every $i = 1, \dots, n-1$,

$$\Big\langle\alpha_i, \sum_{\alpha\in R_0^+}\alpha\Big\rangle = 2\big[(n-i)-(n-i-1)\big] = 2 \quad\text{and}\quad \Big\langle\alpha_i, \sum_{\beta\in R_1^+}\beta\Big\rangle = 0,$$

since $\langle e_i - e_{i+1}, 2e_k\rangle = 2, -2, 0$, if $k = i$, $k = i+1$ or $k \neq i$, $i+1$ respectively. Therefore

$$\prod_{\alpha\in R_0^+} q^{\langle\alpha_i,\alpha\rangle} = q^2 \quad\text{and}\quad \prod_{\beta\in R_1^+} p^{\langle\alpha_i,\beta\rangle} = \prod_{\beta\in R_1^+}\Big(\frac{r}{p}\Big)^{\langle\alpha_i,\beta/2\rangle} = 1$$

and we conclude that $\chi_0(\alpha_i^\vee) = q^2$ for every i.

(ii) If $\beta = \beta_n = 2e_n$, then $\beta^\vee = e_n$. Therefore

$$\chi_0(\beta_n^\vee) = \Big(\prod_{\alpha\in R_0^+} q^{\langle\beta_n^\vee,\alpha\rangle}\Big)\Big(\prod_{\beta\in R_1^+} p^{\langle\beta_n^\vee,\beta\rangle}\Big(\frac{r}{p}\Big)^{\langle\beta_n^\vee,\beta/2\rangle}\Big)$$
$$= q^{\langle\beta_n^\vee,\sum_{\alpha\in R_0^+}\alpha\rangle} p^{\langle\beta_n^\vee,\sum_{\beta\in R_1^+}\beta\rangle}\Big(\frac{r}{p}\Big)^{\langle\beta_n^\vee,\sum_{\beta\in R_1^+}\beta/2\rangle}.$$

On the other hand

$$\Big\langle\beta_n^\vee, \sum_{\alpha\in R_0^+}\alpha\Big\rangle = 0 \quad\text{and}\quad \Big\langle\beta_n^\vee, \sum_{\beta\in R_1^+}\beta\Big\rangle = 2,$$

since $\langle \beta_n^\vee, e_k\rangle = \langle e_n, 2e_k\rangle = 2$ or 0, depending on whether $k = n$ or $k \neq n$. Therefore

$$\prod_{\alpha\in R_0^+} q^{\langle \beta_n^\vee,\alpha\rangle} = 1, \qquad \prod_{\beta\in R_1^+} p^{\langle \beta_n^\vee,\beta\rangle} = p^2, \qquad \prod_{\beta\in R_1^+} \left(\frac{r}{p}\right)^{\langle \beta_n^\vee,\frac{\beta}{2}\rangle} = \frac{r}{p}$$

and we conclude that $\chi_0(\beta^\vee) = pr$. □

For every simple root α we define, for every $\lambda \in \widehat{L}$,

$$\chi_0^\alpha(\lambda) = \prod_{\beta\in R_\alpha^+} q_\beta^{\langle\lambda,\beta\rangle} q_{2\beta}^{-\langle\lambda,\beta\rangle}.$$

Obviously χ_0^α is a character on $\mathbb{A}$; moreover it is easy to check that, if $\lambda \in H_{0,\alpha}$, then

$$\chi_0^\alpha(\lambda) = \chi_0(\lambda),$$

since, for every $\lambda \in H_{0,\alpha}$, we have $\langle\lambda,\alpha\rangle = \langle\lambda, 2\alpha\rangle = 0$ and therefore

$$\prod_{\beta\in R_\alpha^+} q_\beta^{\langle\lambda,\beta\rangle} q_{2\beta}^{-\langle\lambda,\beta\rangle} = \prod_{\beta\in R^+} q_\beta^{\langle\lambda,\beta\rangle} q_{2\beta}^{-\langle\lambda,\beta\rangle} = \chi_0(\lambda).$$

Let T_α be the abstract tree isomorphic to each tree at infinity $T(\eta_\alpha)$. We denote by Γ_0 the fundamental geodesic of the tree and by Γ_0^+ the fundamental geodesic based at 0. Define a character $\bar{\chi}_0$ on Γ_0 in the following way:

$\bar{\chi}_0(X_n) = q_\alpha^n$, *where* X_n *is the vertex of* Γ_0^+ *at distance n from 0, in the homogeneous case;*
$\bar{\chi}_0(X_{2n}) = (pr)^n$, *where* X_{2n} *is the vertex of* Γ_0^+ *at distance* $2n$ *from* 0, *otherwise.*

The characters χ_0, χ_0^α and $\bar{\chi}_0$ are related through the operators P_α and Q_α defined in Sect. 4.4, as the following lemma shows.

Lemma 5.4 *Let* $\lambda \in \widehat{L}$ *such that* $\lambda \in H_{n,\alpha}$ *if* $\alpha \in R_0$, *and* $\lambda \in H_{2n,\alpha}$ *if* $\alpha \in R_2$. *Then*

(i) $\chi_0(Q_\alpha(\lambda)) = \chi_0^\alpha(Q_\alpha(\lambda)) = \chi_0^\alpha(\lambda)$;

(ii) $\chi_0(P_\alpha(\lambda)) = \begin{cases} \bar{\chi}_0(\mathbf{X}_n) = q_\alpha^n & \text{if } \alpha \in R_0, \\ \bar{\chi}_0(\mathbf{X}_{2n}) = (pr)^n & \text{if } \alpha \in R_2. \end{cases}$

Proof (i) Notice at first that $\langle Q_\alpha(\lambda), \alpha\rangle = 0$ for every α. Hence

$$\chi_0^\alpha\big(Q_\alpha(\lambda)\big) = \prod_{\beta\in R_\alpha^+} q_\beta^{\langle Q_\alpha(\lambda),\beta\rangle} q_{2\beta}^{-\langle Q_\alpha(\lambda),\beta\rangle} = \chi_0\big(Q_\alpha(\lambda)\big).$$

Moreover, it is easy to prove that

$$\prod_{\beta\in R_\alpha^+} q_\beta^{\langle P_\alpha(\lambda),\beta\rangle} q_{2\beta}^{-\langle P_\alpha(\lambda),\beta\rangle} = 1.$$

Indeed, for every $\beta \in R_\alpha^+$, the root $s_\alpha\beta$ belongs to R_α^+, and $\langle P_\alpha(\lambda), \beta\rangle = -\langle P_\alpha(\lambda), \sigma_\alpha\beta\rangle$. Therefore,

$$\begin{aligned}\chi_0^\alpha(\lambda) &= \prod_{\beta\in R_\alpha^+} q_\beta^{\langle\lambda,\beta\rangle} q_{2\beta}^{-\langle\lambda,\beta\rangle}\\ &= \prod_{\beta\in R_\alpha^+} q_\beta^{\langle Q_\alpha(\lambda),\beta\rangle} q_{2\beta}^{-\langle Q_\alpha(\lambda),\beta\rangle} \prod_{\beta\in R_\alpha^+} q_\beta^{\langle P_\alpha(\lambda),\beta\rangle} q_{2\beta}^{-\langle P_\alpha(\lambda),\beta\rangle} = \chi_0^\alpha\big(Q_\alpha(\lambda)\big).\end{aligned}$$

(ii) By the same argument of part (i),

$$\begin{aligned}\chi_0\big(P_\alpha(\lambda)\big) &= q_\alpha^{\langle P_\alpha(\lambda),\alpha\rangle} q_{2\alpha}^{-\langle P_\alpha(\lambda),\alpha\rangle} \prod_{\beta\in R_\alpha^+} q_\beta^{\langle P_\alpha(\lambda),\beta\rangle} q_{2\beta}^{-\langle P_\alpha(\lambda),\beta\rangle} = q_\alpha^{\langle P_\alpha(\lambda),\alpha\rangle} q_{2\alpha}^{-\langle P_\alpha(\lambda),\alpha\rangle}\\ &= q_\alpha^{\langle\lambda,\alpha\rangle} q_{2\alpha}^{-\langle\lambda,\alpha\rangle}.\end{aligned}$$

Then part (ii) follows, because

$$q_\alpha^{\langle\lambda,\alpha\rangle} q_{2\alpha}^{-\langle\lambda,\alpha\rangle} = \begin{cases}\bar\chi_0(\mathbf{X}_n) & \text{if } \alpha\in R_0,\\ \bar\chi_0(\mathbf{X}_{2n}) & \text{if } \alpha\in R_2.\end{cases}$$

□

Corollary 5.5 *For every $\lambda \in \widehat{L}$, $\chi_0(\lambda) = \chi_0^\alpha(Q_\alpha(\lambda))\bar\chi_0(\mathbf{X}_\lambda)$, if $\mathbf{X}_\lambda$ is the vertex of Γ_0 that corresponds to $P_\alpha(\lambda)$.*

Let $\rho_{\mathbf{b}}$ be the retraction of the tree on Γ_0, with respect to the boundary point $\mathbf{b}$ such that $\rho_{\mathbf{b}}(\gamma(\mathbf{e},\mathbf{b})) = \Gamma_0^+$ (here $\mathbf{e}$ is the fundamental vertex of the tree).

Proposition 5.6 *Let $x, y \in \widehat{\mathscr{V}}(\Delta)$ and $\omega \in \Omega$. Let $\mathbf{x}$ and $\mathbf{y}$ be the projections of x and y on the tree at infinity $T(\eta_\alpha)$ associated with ω. Then*

(i) $\chi_0(Q_\alpha(\rho_\omega(y) - \rho_\omega(x))) = \chi_0^\alpha(\rho_\omega(y) - \rho_\omega(x))$;
(ii) $\chi_0(P_\alpha(\rho_\omega(y) - \rho_\omega(x))) = \bar\chi_0(\rho_{\mathbf{b}}(\mathbf{y}) - \rho_{\mathbf{b}}(\mathbf{x}))$.

Proof Let $x, y \in \widehat{\mathscr{V}}(\Delta)$ and $\omega \in \Omega$. If $\lambda = \rho_\omega(y) - \rho_\omega(x)$, (i) follows from part (i) of Lemma 5.4.

Let $\eta_\alpha = [\omega]_\alpha$, and consider the vertices $\mathbf{x}, \mathbf{y}$ of the tree $T(\eta_\alpha)$ which correspond to x, y. If $\mathbf{b}$ is the boundary point of this tree that corresponds to ω, then $\mathbf{b} \in B(\mathbf{x}, \mathbf{y})$. This implies that $\rho_{\mathbf{b}}(\mathbf{y}) - \rho_{\mathbf{b}}(\mathbf{x}) = n$ if $\langle\lambda, \alpha\rangle = n$. Hence (ii) follows from part (iii) of Lemma 5.4. □

5.3 Probability Measures on the Boundaries

The measure ν_x, defined on the maximal boundary Ω for any $x \in \widehat{\mathscr{V}}(\Delta)$, may be described in terms of the character χ_0.

Proposition 5.7 *Let x and y be vertices of $\widehat{\mathscr{V}}(\Delta)$; then, for every $\omega \in \Omega(x, y)$,*

$$\nu_x\big(\Omega(x, y)\big) = \frac{\mathbf{W}_\lambda(q^{-1})}{\mathbf{W}(q^{-1})}\chi_0^{-1}\big(\rho_\omega^x(y)\big) = \frac{\mathbf{W}_\lambda(q^{-1})}{\mathbf{W}(q^{-1})}\chi_0^{-1}\big(\rho_\omega(y) - \rho_\omega(x)\big).$$

Proof Since $\chi_0(\lambda) = q_{t_\lambda}$ for every $\lambda \in \widehat{L}^+$, then, by definition of ν_x,

$$\nu_x\big(\Omega(x, y)\big) = \frac{\mathbf{W}_\lambda(q^{-1})}{\mathbf{W}(q^{-1})}\chi_0^{-1}(\lambda), \quad \text{for every } y \in \mathscr{V}_\lambda(x).$$

On the other hand, as we proved in Sect. 3.3, if $y \in Q_x(\omega)$, then $\rho_\omega^x(y) = \sigma(x, y)$ and $\rho_\omega^x(y) = \rho_\omega(y) - \rho_\omega(x)$. The statement follows. □

Let α be any simple root of the root system R associated with Δ. The measure ν_x^α defined in Sect. 4.6 on the α-boundary can be characterized in terms of the character χ_0^α.

Proposition 5.8 *Let $\lambda \in \widehat{L}^+$ and $y \in \mathscr{V}_\lambda(x)$; then, for every $\eta_\alpha \in \Omega_\alpha(x, y)$ and every $\omega \in \eta_\alpha$,*

$$\nu_x^\alpha\big(\Omega_\alpha(x, y)\big) = \begin{cases} \frac{\mathbf{W}_\lambda(q^{-1})}{\mathbf{W}(q^{-1})}(\chi_0^\alpha)^{-1}(\rho_\omega(y) - \rho_\omega(x)), & \textit{if } \lambda \in H_{0,\alpha}, \\ \frac{\mathbf{W}_\lambda(q^{-1})(1+q_\alpha^{-1})}{\mathbf{W}(q^{-1})}(\chi_0^\alpha)^{-1}(\rho_\omega(y) - \rho_\omega(x)), & \textit{otherwise.} \end{cases}$$

Proof Recalling the definition of the character χ_0^α, we have

$$\nu_x^\alpha\big(\Omega_\alpha(x, y)\big) = \begin{cases} \frac{\mathbf{W}_\lambda(q^{-1})}{\mathbf{W}(q^{-1})}(\chi_0^\alpha)^{-1}(\lambda), & \text{if } \lambda \in H_{0,\alpha}, \\ \frac{\mathbf{W}_\lambda(q^{-1})(1+q_\alpha^{-1})}{\mathbf{W}(q^{-1})}(\chi_0^\alpha)^{-1}(\lambda), & \text{otherwise.} \end{cases}$$

On the other hand, for every $\eta_\alpha \in \Omega_\alpha(x, y)$ and every ω in the class η_α,

$$\rho_\omega(y) - \rho_\omega(x) = \lambda, \quad \text{if } \sigma(x, y) = \lambda.$$

In particular, if we assume $y \in V_\lambda(x)$, with $\lambda \in H_{0,\alpha}$, then the vector $\rho_\omega(y) - \rho_\omega(x)$ belongs to $H_{0,\alpha}$. □

Taking in account Proposition 5.6 of the previous section, we can express the measures ν_x^α and $\mu_{\mathbf{x}}$ in terms of the character χ_0 and the operators P_α and Q_α.

Corollary 5.9 *Let $x, y \in \widehat{\mathscr{V}}(\Delta)$ and $y \in \mathscr{V}_\lambda(x)$. Let $\mathbf{x}$ and $\mathbf{y}$ be the projections of x and y on the tree at infinity $T(\eta_\alpha)$ associated with any $\omega \in \Omega(x, y)$. Then*

$$\nu_x^\alpha\big(\Omega_\alpha(x, y)\big) = \begin{cases} \frac{\mathbf{W}_\lambda(q^{-1})}{\mathbf{W}(q^{-1})} \chi_0^{-1}(\rho_\omega(y) - \rho_\omega(x)), & \lambda \in H_{0,\alpha}, \\ \frac{\mathbf{W}_\lambda(q^{-1})(1+q_\alpha^{-1})}{\mathbf{W}(q^{-1})} \chi_0^{-1}(Q_\alpha(\rho_\omega(y) - \rho_\omega(x))), & \textit{otherwise}. \end{cases}$$

Moreover

$$\mu_{\mathbf{x}}\big(B(\mathbf{x}, \mathbf{y})\big) = \begin{cases} 1, & \textit{if } \lambda \in H_{0,\alpha}, \\ \frac{q_\alpha}{1+q_\alpha} \chi_0^{-1}(P_\alpha(\rho_\omega(y) - \rho_\omega(x))), & \textit{otherwise}. \end{cases}$$

In view of Corollary 5.9, the decomposition of the measure ν_x for the maximal boundary, stated in Sect. 4.8, is a direct consequence of the orthogonal decomposition $\chi_0(\lambda) = \chi_0(P_\alpha(\lambda))\chi_0(Q_\alpha(\lambda))$.

5.4 Poisson Kernel and Poisson Transform

Proposition 5.10 *For $x, y \in \widehat{\mathscr{V}}(\Delta)$ the measures ν_x, ν_y are mutually absolutely continuous and the Radon–Nikodym derivative of ν_y with respect to ν_x is given by*

$$\frac{d\nu_y}{d\nu_x}(\omega) = \chi_0\big(\rho_\omega^x(y)\big) = \chi_0\big(\rho_\omega(y) - \rho_\omega(x)\big), \quad \textit{for } \omega \in \Omega.$$

Proof We fix x, y and ω; by Corollary 3.9 of Sect. 3.3, we can choose a special vertex z lying into $Q_y(\omega) \cap Q_x(\omega)$, so that $\Omega(x, z) = \Omega(y, z)$. We set $\Omega_z = \Omega(x, z) = \Omega(y, z)$. Of course ω belongs to Ω_z. By Proposition 5.7 of the previous section,

$$\nu_x(\Omega_z) = \nu_x\big(\Omega(x, z)\big) = \frac{\mathbf{W}_\lambda(q^{-1})}{\mathbf{W}(q^{-1})} \chi_0^{-1}\big(\rho_\omega(z) - \rho_\omega(x)\big),$$

$$\nu_y(\Omega_z) = \nu_y\big(\Omega(y, z)\big) = \frac{\mathbf{W}_\lambda(q^{-1})}{\mathbf{W}(q^{-1})} \chi_0^{-1}\big(\rho_\omega(z) - \rho_\omega(y)\big).$$

So we conclude that

$$\frac{\nu_y(\Omega_z)}{\nu_x(\Omega_z)} = \frac{\chi_0^{-1}(\rho_\omega(z) - \rho_\omega(y))}{\chi_0^{-1}(\rho_\omega(z) - \rho_\omega(x))} = \chi_0\big(\rho_\omega(y) - \rho_\omega(x)\big).$$

This proves that ν_y is absolutely continuous with respect to ν_x and shows the required formula for the Radon–Nikodym derivative of ν_y with respect to ν_x. □

Definition 5.11 We call *Poisson kernel* of the building Δ the function

$$P(x, y, \omega) = \chi_0\big(\rho_\omega(y) - \rho_\omega(x)\big) = \chi_0\big(\rho_\omega^x(y)\big), \quad \text{for } x, y \in \widehat{\mathscr{V}}(\Delta) \text{ and } \omega \in \Omega.$$

This definition does not depend on the choice of the special vertex e. By Proposition 5.10, for every choice of x, y in $\widehat{\mathscr{V}}(\Delta)$, the function $P(x, y, \cdot)$ is the Radon–Nikodym derivative of ν_y with respect to ν_x:

$$\frac{d\nu_y}{d\nu_x}(\omega) = P(x, y, \omega), \quad \text{for } \omega \in \Omega.$$

The same argument of Proposition 5.10 proves also the following proposition.

Proposition 5.12 *For $x, y \in \widehat{\mathscr{V}}(\Delta)$, the measures ν_x^α, ν_y^α are mutually absolutely continuous and*

$$\frac{d\nu_y^\alpha}{d\nu_x^\alpha}(\eta_\alpha) = \chi_0^\alpha\big(\rho_\omega(y) - \rho_\omega(x)\big), \quad \textit{for } \omega \in \eta_\alpha, \eta_\alpha \in \Omega_\alpha.$$

For every $x, y \in \widehat{\mathscr{V}}(\Delta)$ and for every $\eta_\alpha \in \Omega_\alpha$, let

$$P^\alpha(x, y, \eta_\alpha) = \frac{d\nu_y^\alpha}{d\nu_x^\alpha}(\eta_\alpha) = \chi_0^\alpha\big(\rho_\omega(y) - \rho_\omega(x)\big), \quad \text{for } \omega \in \eta_\alpha.$$

It is known that, for every pair of vertices $\mathbf{t}, \mathbf{t}'$ in $\widehat{\mathscr{V}}(T_\alpha)$, the measure $\mu_{\mathbf{t}'}$ is absolutely continuous with respect to $\mu_{\mathbf{t}}$. The Radon–Nikodym derivative $d\mu_{\mathbf{t}'}/d\mu_{\mathbf{t}}(\mathbf{b})$ is the Poisson kernel $P(\mathbf{t}, \mathbf{t}', \mathbf{b})$, where

$P(\mathbf{t}, \mathbf{t}', \mathbf{b}) = q_\alpha^{n-1}$ if $\mathrm{dist}(\mathbf{t}, \mathbf{t}') = n$, in the homogeneous case;
$P(\mathbf{t}, \mathbf{t}', \mathbf{b}) = (pr)^{n-1}$ if $\mathrm{dist}(\mathbf{t}, \mathbf{t}') = 2n$, in the semi-homogeneous case.

In both cases, as a straightforward consequence of the definition,

$$P\big(\mathbf{t}, \mathbf{t}', \mathbf{b}\big) = \bar{\chi}_0\big(\rho_{\mathbf{b}}\big(\mathbf{t}'\big) - \rho_{\mathbf{b}}(\mathbf{t})\big), \quad \text{for } \mathbf{b} \in \partial T_\alpha.$$

Since, for every pair of vertices $x, y \in \widehat{\mathscr{V}}(\Delta)$, the measure ν_y on Ω is absolutely continuous with respect to ν_x, the measure ν_y^α on Ω_α is absolutely continuous with respect to ν_x^α and the measure $\mu_{\mathbf{y}}$ on ∂T_α is absolutely continuous with respect to $\mu_{\mathbf{x}}$. Moreover,

Corollary 5.13 *Let $x, y \in \widehat{\mathscr{V}}(\Delta)$, and $\omega \in \Omega$. If $\omega = (\eta_\alpha, \mathbf{b})$, and $\mathbf{x}$ and $\mathbf{y}$ are the projections of x and y on the tree at infinity $T(\eta_\alpha)$, then*

$$P(x, y, \omega) = P^\alpha(x, y, \eta_\alpha) P(\mathbf{x}, \mathbf{y}, \mathbf{b}).$$

Proof By Proposition 5.6 of Sect. 5.2, for every $x, y \in \widehat{\mathscr{V}}(\Delta)$ and $\omega \in \Omega$,

$$P^\alpha(x, y, \eta_\alpha) = \chi_0(Q_\alpha\big(\rho_\omega(y) - \rho_\omega(x)\big) \quad \text{and} \quad P(\mathbf{x}, \mathbf{y}, \mathbf{b}) = \chi_0(P_\alpha\big(\rho_\omega(y) - \rho_\omega(x)\big).$$

Therefore, the decomposition of the Poisson kernel $P(x, y, \omega)$ is a direct consequence of the orthogonal decomposition $\chi_0(\lambda) = \chi_0(P_\alpha(\lambda))\chi_0(Q_\alpha(\lambda))$. □

Definition 5.11 can be generalized, if the character χ_0 is replaced by any character χ.

Definition 5.14 We call *generalized Poisson kernel* of the building Δ associated with the character χ the function

$$P^{\chi}(x, y, \omega) = \chi\big(\rho_\omega(y) - \rho_\omega(x)\big), \quad \text{for } x, y \in \widehat{\mathscr{V}}(\Delta) \text{ and } \omega \in \Omega.$$

It is obvious that also this definition does not depend on the choice of the vertex e. According to this definition, $P(x, y, \omega) = P^{\chi_0}(x, y, \omega)$.

The following proposition describes the properties of the functions $P^{\chi}(x, y, \omega)$.

Proposition 5.15 *Let χ be a character on $\mathbb{A}$. Then,*

(i) $P^{\chi}(x, x, \omega) = 1$, *for every x and every ω; moreover, for every x, y and every ω,*

$$P^{\chi}(y, x, \omega) = \big(P^{\chi}(x, y, \omega)\big)^{-1} = P^{\chi^{-1}}(x, y, \omega);$$

(ii) *for every x and every ω, the function $P^{\chi}(x, \cdot, \omega)$ is constant on the set*

$$\big\{y \in \widehat{\mathscr{V}}(\Delta) : \sigma(x, y) = \lambda, \rho^x_\omega(y) = \mu\big\},$$

for any $\lambda \in \widehat{L}^+$ and $\mu \in \Pi_\lambda$.

(iii) *for every x, y, the function $P^{\chi}(x, y, \cdot)$ is locally constant on Ω, and, if $\sigma(x, y) = \lambda$, then $P^{\chi}(x, y, \omega) = \chi(\lambda)$, for all $\omega \in \Omega(x, y)$.*

Proof (i) and (ii) follow directly from the definition. Moreover (iii) is a consequence of the properties of the retraction ρ^x_ω, proved in Sect. 3.3. Indeed, if $\sigma(x, y) = \lambda$, and we choose μ big enough with respect to λ, then $\Omega = \bigcup_{z \in V_\mu(x)} \Omega(x, z)$ and $\rho^x_\omega(y)$ does not depend on the choice of ω in each set $\Omega(x, z)$. In particular, $\rho^x_\omega(y) = \lambda$, for all $\omega \in \Omega(x, y)$. □

Definition 5.16 Let $x_0 \in \widehat{\mathscr{V}}(\Delta)$ and let χ be a character on $\mathbb{A}$. For any complex valued function f on Ω, we call *generalized Poisson transform of f* with base point x_0, associated with the character χ, the function on $\widehat{\mathscr{V}}(\Delta)$ defined by

$$\mathscr{P}^{\chi}_{x_0} f(x) = \int_\Omega P^{\chi}(x_0, x, \omega) f(\omega) d\nu_{x_0}(\omega) = \int_\Omega \chi\big(\rho_\omega(x) - \rho_\omega(x_0)\big) f(\omega) d\nu_{x_0}(\omega),$$

for every $x \in \widehat{\mathscr{V}}(\Delta)$ and $f \in L^1(\Omega, \nu_{x_0})$.

In particular, set $\mathscr{P}_{x_0} = \mathscr{P}^{\chi_0}_{x_0}$ and $\mathscr{P} = \mathscr{P}_e$.

6 The Algebra $\mathscr{H}(\Delta)$ and Its Eigenvalues

6.1 The Algebra $\mathscr{H}(\Delta)$

For every $\lambda \in \widehat{L}^+$, define an operator A_λ acting on the space of complex valued functions f on $\widehat{\mathscr{V}}(\Delta)$ by

$$(A_\lambda f)(x) = \sum_{y \in V_\lambda(x)} f(y) = \sum_{y \in \widehat{\mathscr{V}}(\Delta)} \mathbb{1}_{V_\lambda(x)}(y) f(y), \quad \text{for } x \in \widehat{\mathscr{V}}(\Delta).$$

The operators A_λ are linear; moreover, for each y, the coefficient $\mathbb{1}_{V_\lambda(x)}(y)$ depends only on λ. Observe that the operators A_λ, $\lambda \in \widehat{L}^+$ are linearly independent. Actually, if assume $\sum_{\lambda \in \widehat{L}^+} a_\lambda A_\lambda = 0$, then

$$\sum_{\lambda \in \widehat{L}^+} a_\lambda (A_\lambda \delta_y)(x) = 0, \quad \text{for every } x, y \in \widehat{\mathscr{V}}(\Delta).$$

On the other hand $\sum_{\lambda \in \widehat{L}^+} a_\lambda (A_\lambda \delta_y)(x) = a_\mu$ if $\sigma(x, y) = \mu$. Hence we get $a_\mu = 0$ for every $\mu \in \widehat{L}^+$.

Denote by $\mathscr{H}(\Delta)$ the linear span of $\{A_\lambda, \lambda \in \widehat{L}^+\}$ over $\mathbb{C}$.

Proposition 6.1 *The space $\mathscr{H}(\Delta)$ is a commutative $\mathbb{C}$-algebra.*

Proof Let us prove that, for every λ, μ the operator $A_\lambda \circ A_\mu$ is a finite linear combination of operators A_ν, for suitable ν. Indeed, recalling (3), for every function f and for every $x \in \widehat{\mathscr{V}}(\Delta)$ one has

$$\begin{aligned}
A_\lambda \circ A_\mu f(x) &= \sum_{y \in \widehat{\mathscr{V}}(\Delta)} \mathbb{1}_{V_\lambda(x)}(y) A_\mu f(y) = \sum_{y \in \widehat{\mathscr{V}}(\Delta)} \mathbb{1}_{V_\lambda(x)}(y) \sum_{z \in \widehat{\mathscr{V}}(\Delta)} \mathbb{1}_{V_\mu(y)}(z) f(z) \\
&= \sum_{z \in \widehat{\mathscr{V}}(\Delta)} \Bigg(\sum_{y \in \widehat{\mathscr{V}}(\Delta)} \mathbb{1}_{V_\lambda(x)}(y) \mathbb{1}_{V_\mu(y)}(z) \Bigg) f(z) \\
&= \sum_{z \in \widehat{\mathscr{V}}(\Delta)} \big| \big\{ y \in \widehat{\mathscr{V}}(\Delta) : \sigma(x, y) = \lambda, \sigma(y, z) = \mu \big\} \big| f(z) \\
&= \sum_{\nu \in \widehat{L}^+} \sum_{z \in V_\nu(x)} N\big(\nu, \lambda, \mu^\star\big) f(z) = \sum_{\nu \in \widehat{L}^+} N\big(\nu, \lambda, \mu^\star\big) (A_\nu f)(x)
\end{aligned}$$

and $N(\nu, \lambda, \mu^\star)$ is different from zero only for finitely many ν. Moreover

$$A_\mu \circ A_\lambda f = \sum_{\nu \in \widehat{L}^+} N\big(\nu, \mu, \lambda^\star\big)(A_\nu f) = \sum_{\nu \in \widehat{L}^+} N\big(\nu, \lambda, \mu^\star\big)(A_\nu f) = A_\lambda \circ A_\mu f.$$

This completes the proof. □

We refer to the numbers $N(\nu, \lambda, \mu^\star)$ in Proposition 6.1 as the *structure constants* of $\mathscr{H}(\Delta)$.

6.2 Eigenvalue of the Algebra $\mathscr{H}(\Delta)$ Associated with a Character χ

Let χ be a character on $\mathbb{A}$, and consider the generalized Poisson kernel $P^\chi(x, y, \omega)$ associated with χ.

Lemma 6.2 *Let $z \in \widehat{\mathscr{V}}(\Delta)$ and $\omega \in \Omega$. For every $\lambda \in \widehat{L}^+$,*

$$\sum_{y \in \mathscr{V}_\lambda(z)} \chi\big(\rho_\omega(y) - \rho_\omega(z)\big) = \sum_{\mu \in \Pi_\lambda} N(\lambda, \mu)\chi(\mu).$$

Proof For every $z \in \widehat{\mathscr{V}}(\Delta)$, $\omega \in \Omega$ and $\lambda \in \widehat{L}^+$,

$$\sum_{y \in \mathscr{V}_\lambda(z)} \chi\big(\rho_\omega(y) - \rho_\omega(z)\big) = \sum_{\mu \in \Pi_\lambda} \big|\big\{y \in \mathscr{V}_\lambda(z) : \rho_\omega(y) - \rho_\omega(z) = \mu\big\}\big|\chi(\mu).$$

By Theorem 3.12 of Sect. 3.3 and (5), for every $\mu \in \Pi_\lambda$,

$$\big|\big\{y \in \mathscr{V}_\lambda(z) : \rho_\omega(y) - \rho_\omega(z) = \mu\big\}\big| = \big|\big\{y \in \mathscr{V}_\lambda(e) : \rho_\omega(y) = \mu\big\}\big| = N(\lambda, \mu).$$

This proves the statement. □

For every $\lambda \in \widehat{L}^+$, define

$$\Lambda^\chi(\lambda) = \sum_{\mu \in \Pi_\lambda} N(\lambda, \mu)\chi(\mu). \tag{7}$$

Proposition 6.3 *For every $\lambda \in \widehat{L}^+$, $\Lambda^\chi(\lambda)$ is an eigenvalue of the operator A_λ and, for every $x \in \widehat{\mathscr{V}}(\Delta)$ and $\omega \in \Omega$, the function $P^\chi(x, \cdot, \omega)$ is an eigenfunction of A_λ associated to the eigenvalue $\Lambda^\chi(\lambda)$:*

$$A_\lambda P^\chi(x, \cdot, \omega) = \Lambda^\chi(\lambda) P^\chi(x, \cdot, \omega).$$

Proof For every $z \in \widehat{\mathscr{V}}(\Delta)$, we can write

$$\begin{aligned} A_\lambda P^\chi(x, \cdot, \omega)(z) &= \sum_{y \in \mathscr{V}_\lambda(z)} P^\chi(x, y, \omega) = \sum_{y \in \mathscr{V}_\lambda(z)} \chi\big(\rho_\omega(y) - \rho_\omega(x)\big) \\ &= \sum_{y \in \mathscr{V}_\lambda(z)} \chi\big(\rho_\omega(y)\big)\chi\big(-\rho_\omega(x)\big) \\ &= \chi\big(\rho_\omega(z) - \rho_\omega(x)\big) \sum_{y \in \mathscr{V}_\lambda(z)} \chi\big(\rho_\omega(y) - \rho_\omega(z)\big) \\ &= P^\chi(x, z, \omega) \sum_{y \in \mathscr{V}_\lambda(z)} \chi\big(\rho_\omega(y) - \rho_\omega(z)\big). \end{aligned}$$

Hence, by Lemma 6.2, we conclude that $A_\lambda P^\chi(x, \cdot, \omega) = \Lambda^\chi(\lambda) P^\chi(x, \cdot, \omega)$. □

Since $\{A_\lambda, \lambda \in \widehat{L}^+\}$ generates $\mathscr{H}(\Delta)$, then $\{\Lambda^\chi(\lambda), \lambda \in \widehat{L}^+\}$ generates an algebra homomorphism Λ^χ from $\mathscr{H}(\Delta)$ to $\mathbb{C}$ such that $\Lambda^\chi(A_\lambda) = \Lambda^\chi(\lambda)$ for every $\lambda \in \widehat{L}^+$. Moreover, for every $x \in \widehat{\mathscr{V}}(\Delta)$ and $\omega \in \Omega$, the function $P^\chi(x, \cdot, \omega)$ is an eigenfunction of $\mathscr{H}(\Delta)$ associated with the eigenvalue Λ^χ.

In the particular case when $\chi = \chi_0$, then, for every $x \in \widehat{\mathscr{V}}(\Delta)$ and every $\omega \in \Omega$, the Poisson kernel $P(x, \cdot, \omega)$ is an eigenfunction of all operators A_λ, with associated eigenvalue $\Lambda^{\chi_0}(\lambda)$. Since $P(x, y, \omega)$ is the Radon–Nikodym derivative of the measure ν_y with respect to the measure ν_x, this implies that

$$\sum_{y \in \mathscr{V}_\lambda(x)} \nu_y = \Lambda^{\chi_0}(\lambda)\nu_x.$$

On the other hand

$$\sum_{y \in \mathscr{V}_\lambda(x)} \nu_y = \left|\mathscr{V}_\lambda(x)\right|\nu_x = N_\lambda \nu_x,$$

since ν_y and ν_x are probability measures on Ω. Therefore $\Lambda^{\chi_0}(\lambda) = N_\lambda$ and hence

$$\sum_{\mu \in \Pi_\lambda} N(\lambda, \mu)\chi_0(\mu) = N_\lambda.$$

Corollary 6.4 *For every character χ, the Poisson transform $\mathscr{P}_x^\chi(f)$ (with base point x) of every $f \in L^1(\Omega, \nu_x)$ is an eigenfunction of the algebra $\mathscr{H}(\Delta)$ associated with the eigenvalue Λ^χ.*

Proof For every $\lambda \in \widehat{L}^+$,

$$\begin{aligned} A_\lambda \mathscr{P}_x^\chi(f)(z) &= \sum_{y \in \mathscr{V}_\lambda(z)} \mathscr{P}_x^\chi(f)(y) = \sum_{y \in \mathscr{V}_\lambda(z)} \int_\Omega P^\chi(x, y, \omega) f(\omega)\, d\nu_x(\omega) \\ &= \int_\Omega \Bigg(\sum_{y \in \mathscr{V}_\lambda(z)} P^\chi(x, y, \omega) \Bigg) f(\omega)\, d\nu_x(\omega) \\ &= \int_\Omega \Lambda^\chi(\lambda) P^\chi(x, z, \omega) f(\omega)\, d\nu_x(\omega) \\ &= \Lambda^\chi(\lambda) \mathscr{P}_x^\chi(f)(z). \end{aligned}$$

□

The Weyl group $\mathbf{W}$ acts on the eigenvalues Λ^χ of the algebra $\mathscr{H}(\Delta)$. We shall prove that in fact these eigenvalues are invariant with respect to the action of $\mathbf{W}$, in the sense that, for every character χ,

$$\Lambda^{\chi \chi_0^{1/2}} = \Lambda^{\chi^{\mathbf{w}} \chi_0^{1/2}} \quad \text{for every } \mathbf{w} \in \mathbf{W}.$$

6.3 Preliminary Results

Let χ be a fixed character on $\mathbb{A}$ and let α be a fixed simple root.

Definition 6.5 Let $x \in \widehat{\mathscr{V}}(\Delta)$. For each pair ω_1, ω_2 in a class $\eta_\alpha \in \Omega_\alpha$, we fix a vertex of $\widehat{\mathscr{V}}(\Delta)$, say $e = e_{\omega_1,\omega_2}$, in any apartment $\mathscr{A}(\omega_1, \omega_2)$ containing both boundary points. We set

$$j^\alpha_{x,\chi}(\omega_1, \omega_2) = \chi\chi_0^{1/2}\big(P_\alpha\big(\rho_{\omega_1}(e) + \rho_{\omega_2}(e) - \rho_{\omega_1}(x) - \rho_{\omega_2}(x)\big)\big).$$

Remark 6.6 The function $j^\alpha_{x,\chi}(\omega_1, \omega_2)$ does not depend on the choice of the vertex e_{ω_1,ω_2} in any apartment $\mathscr{A}(\omega_1, \omega_2)$. Indeed, if e and e' are two vertices in this apartment, then, for every $x \in \widehat{\mathscr{V}}(\Delta)$,

$$\begin{aligned} &P_\alpha\big(\rho_{\omega_1}(x) - \rho_{\omega_1}(e) + \rho_{\omega_2}(x) - \rho_{\omega_2}(e)\big) - P_\alpha\big(\rho_{\omega_1}(x) - \rho_{\omega_1}(e') + \rho_{\omega_2}(x) - \rho_{\omega_2}(e')\big) \\ &\quad = P_\alpha\big(\big(\rho_{\omega_1}(e') - \rho_{\omega_1}(e)\big) + \big(\rho_{\omega_2}(e') - \rho_{\omega_2}(e)\big)\big) \\ &\quad = P_\alpha\big(\big(\rho_{\omega_1}(e') - \rho_{\omega_1}(e)\big)\big) + P_\alpha\big(\big(\rho_{\omega_2}(e') - \rho_{\omega_2}(e)\big)\big) = 0, \end{aligned}$$

since $P_\alpha((\rho_{\omega_1}(e') - \rho_{\omega_1}(e))) = -P_\alpha((\rho_{\omega_2}(e') - \rho_{\omega_2}(e)))$, as we proved in Proposition 4.13 of Sect. 4.4.

For every $\omega \in \Omega$, let η_α be the element of the α-boundary Ω_α such that $\omega \in \eta_\alpha$. We denote by $\nu^\alpha_{x,\omega}$ the restriction of the measure ν_x to the set $\{\omega' \in \Omega : \omega' \in \eta_\alpha\}$. Since this set can be identified with the boundary of the tree $T(\eta_\alpha)$, then $\nu^\alpha_{x,\omega}$ can be seen as the usual measure $\mu_{\mathbf{x}}$ on $\partial T(\eta_\alpha)$.

Definition 6.7 Let $x \in \widehat{\mathscr{V}}(\Delta)$. We denote by $J^\alpha_{x,\chi}$ the following operator acting on the complex valued functions f defined on Ω:

$$J^\alpha_{x,\chi}(f)(\omega_0) = \int_\Omega j^\alpha_{x,\chi}(\omega_0, \omega) f(\omega)\, d\nu^\alpha_{x,\omega_0}(\omega), \quad \text{for } \omega_0 \in \Omega.$$

Theorem 6.8 *Assume that* $|\chi(\alpha^\vee)| < 1$; *then*

(i) $J^\alpha_{x,\chi}\mathbf{1} = c(\chi)\mathbf{1}$, *where*

$$c(\chi) = \begin{cases} \frac{q_\alpha}{q_\alpha+1} + \frac{q_\alpha-1}{q_\alpha+1}\sum_{j\geqslant 1}(\bar{\chi}(\mathbf{X}_1))^{2j} & \text{if } \alpha \in R_0, \\ \frac{r}{(r+1)} + [\frac{r-1}{r+1} + \frac{r(p-1)}{r+1}\frac{1}{\sqrt{pr}}\bar{\chi}(-\mathbf{X}_2)]\sum_{j\geqslant 1}(\bar{\chi}(\mathbf{X}_2))^{2j} & \text{if } \alpha \in R_2; \end{cases}$$

(ii) $J^\alpha_{x,\chi} : L^\infty(\Omega) \to L^\infty(\Omega)$ *is a bounded operator.*

Proof (i) Fix ω_0 in Ω and let $\eta_\alpha = [\omega_0]_\alpha$. By Definitions 6.5 and 6.7,

$$\begin{aligned} J^\alpha_{x,\chi}\mathbf{1}(\omega_0) &= \int_\Omega j^\alpha_{x,\chi}(\omega_0,\omega)\,d\nu^\alpha_{x,\omega_0}(\omega) \\ &= \int_{[\omega_0]_\alpha} \chi\chi_0^{1/2}\big(P_\alpha\big(\rho_{\omega_0}(e)+\rho_\omega(e)-\rho_{\omega_0}(x)-\rho_\omega(x)\big)\big)\,d\nu^\alpha_{x,\omega_0}(\omega), \end{aligned}$$

where e is a vertex in any apartment that contains ω_0 and ω.

Consider the tree $T(\eta_\alpha)$ and its boundary $\partial T(\eta_\alpha)$. In line with notation of Sect. 5.2, we simply denote by $\bar\chi$ the character on the fundamental geodesic Γ_0 of the tree, such that, for every $n \in \mathbb{Z}$,

$$\begin{aligned} \bar\chi(\mathbf{X}_n) &= \chi\big(P_\alpha(\lambda)\big) \quad \text{if } \alpha \in R_0, \\ \bar\chi(\mathbf{X}_{2n}) &= \chi\big(P_\alpha(\lambda)\big) \quad \text{if } \alpha \in R_2, \end{aligned}$$

where $\lambda \in \widehat{L}$ satisfies $\langle\lambda,\alpha\rangle = n$. Since the set $[\omega_0]_\alpha$ can be identified with the boundary of the tree $T(\eta_\alpha)$ and the measure ν^α_{x,ω_0} can be seen as the usual measure $\mu_{\mathbf{x}}$ on $\partial T(\eta_\alpha)$, we can write

$$J^\alpha_{x,\chi}\mathbf{1}(\omega_0) = \int_{\partial T(\eta_\alpha)} \bar\chi\bar\chi_0^{1/2}\big(\rho_{\mathbf{b}_0}(\mathbf{e})+\rho_{\mathbf{b}}(\mathbf{e})-\rho_{\mathbf{b}_0}(\mathbf{x})-\rho_{\mathbf{b}}(\mathbf{x})\big)\,d\mu_{\mathbf{x}}(\mathbf{b}),$$

where $\mathbf{b}_0$ is the boundary point of the tree that corresponds to ω_0, $\mathbf{b}$ is the boundary point of the tree that corresponds to ω for every $\omega \in [\omega_0]_\alpha$, and $\mathbf{e}$ is the vertex of the geodesic $\gamma(\mathbf{b}_0,\mathbf{b})$ obtained as projection of e on the tree $T(\eta_\alpha)$. For every $x \in \widehat{\mathscr{V}}(\Delta)$, let $\mathbf{x}$ be the vertex of the tree that corresponds to x and denote by $N_{\mathbf{x}}(\mathbf{b}_0,\mathbf{b})$ the distance of $\mathbf{x}$ from the geodesic $[\mathbf{b}_0,\mathbf{b}]$, that is, the minimal distance of $\mathbf{x}$ from the set $\{\mathbf{y} \in \mathscr{V}(T(\eta_\alpha)) : \mathbf{y} \in [\mathbf{b}_0,\mathbf{b}]\}$. For every $j \geqslant 0$, set

$$B_j(\mathbf{x},\mathbf{b}_0) = \big\{\mathbf{b} \in \partial T(\eta_\alpha) : N_{\mathbf{x}}(\mathbf{b}_0,\mathbf{b}) = j\big\}.$$

Then $\partial T(\eta_\alpha)$ can be decomposed as a disjoint union in the following way:

$$\partial T(\eta_\alpha) = \bigcup_j B_j(\mathbf{x},\mathbf{b}_0).$$

We can easily compute $\mu_{\mathbf{x}}(B_j(\mathbf{x},\mathbf{b}_0))$, for $j \geqslant 0$. If $\alpha \in R_0$, the tree $T(\eta_\alpha)$ is homogeneous and

$$\mu_{\mathbf{x}}\big(B_0(\mathbf{x},\mathbf{b}_0)\big) = \frac{q_\alpha}{q_\alpha+1} \quad \text{and} \quad \mu_{\mathbf{x}}\big(B_j(\mathbf{x},\mathbf{b}_0)\big) = \frac{q_\alpha-1}{q_\alpha+1}q_\alpha^{-j} \quad \text{for all } j>0.$$

Otherwise, if $\alpha \in R_2$, the tree $T(\eta_\alpha)$ is semi-homogeneous and

$$\begin{aligned} \mu_{\mathbf{x}}\big(B_0(\mathbf{x},\mathbf{b}_0)\big) &= \frac{r}{r+1}, \\ \mu_{\mathbf{x}}\big(B_{2j}(\mathbf{x},\mathbf{b}_0)\big) &= \frac{r-1}{r+1}(pr)^{-j}, \quad \text{for } j>0, \end{aligned}$$

$$\mu_{\mathbf{x}}\big(B_{2j+1}(\mathbf{x},\mathbf{b}_0)\big) = \frac{p-1}{p(r+1)}(pr)^{-j}, \quad \text{for } j \geqslant 0.$$

It is easy to see that, for every $j \geqslant 0$,

$$\rho_{\mathbf{b}_0}(\mathbf{e}) + \rho_{\mathbf{b}}(\mathbf{e}) - \rho_{\mathbf{b}_0}(\mathbf{x}) - \rho_{\mathbf{b}}(\mathbf{x}) = X_{2j} \quad \text{for all } \mathbf{b} \in B_j(\mathbf{x},\mathbf{b}_0).$$

Thus

$$J^{\alpha}_{x,\chi}\mathbf{1}(\omega_0) = \sum_{j=0}^{\infty} \mu_{\mathbf{x}}\big(B_j(\mathbf{x},\mathbf{b}_0)\big)\bar{\chi}\bar{\chi}_0^{1/2}(\mathbf{X}_{2j}).$$

Therefore, if $\alpha \in R_0$,

$$\begin{aligned} J^{\alpha}_{x,\chi}\mathbf{1}(\omega_0) &= \frac{q_\alpha}{q_\alpha+1}\bar{\chi}\bar{\chi}_0^{1/2}(0) + \sum_{j\geqslant 1}\frac{q_\alpha-1}{q_\alpha+1}q_\alpha^{-j}\bar{\chi}\bar{\chi}_0^{1/2}(\mathbf{X}_{2j}) \\ &= \frac{q_\alpha}{q_\alpha+1} + \frac{q_\alpha-1}{q_\alpha+1}\sum_{j\geqslant 1} q_\alpha^{-j}q_\alpha^{j}\bar{\chi}(2j\mathbf{X}_1) \\ &= \frac{q_\alpha}{q_\alpha+1} + \frac{q_\alpha-1}{q_\alpha+1}\sum_{j\geqslant 1}\big(\bar{\chi}(\mathbf{X}_1)\big)^{2j}. \end{aligned}$$

Analogously, if $\alpha \in R_2$, then

$$\begin{aligned} J^{\alpha}_{x,\chi}\mathbf{1}(\omega_0) &= \frac{r}{r+1}\bar{\chi}\bar{\chi}_0^{1/2}(\mathbf{X}_0) + \sum_{j\geqslant 1}\frac{r-1}{r+1}(pr)^{-j}\bar{\chi}\bar{\chi}_0^{1/2}(\mathbf{X}_{4j}) \\ &\quad + \sum_{j\geqslant 1}\frac{r(p-1)}{r+1}(pr)^{-j}\bar{\chi}\bar{\chi}_0^{1/2}(\mathbf{X}_{4j-2}) \\ &= \frac{r}{(r+1)} + \frac{r-1}{r+1}\sum_{j\geqslant 1}(pr)^{-j}(pr)^{j}\bar{\chi}(2j\mathbf{X}_2) \\ &\quad + \frac{r(p-1)}{r+1}\frac{1}{\sqrt{pr}}\sum_{j\geqslant 1}(pr)^{-j}(pr)^{j}\bar{\chi}\big((2j-1)\mathbf{X}_2\big) \\ &= \frac{r}{(r+1)} + \left[\frac{r-1}{r+1} + \frac{r(p-1)}{r+1}\frac{1}{\sqrt{pr}}\bar{\chi}(-\mathbf{X}_2)\right]\sum_{j\geqslant 1}\big(\bar{\chi}(\mathbf{X}_2)\big)^{2j}. \end{aligned}$$

Since $\bar{\chi}(\mathbf{X}_2) = \chi(\alpha^\vee)$ and $\bar{\chi}(\mathbf{X}_1) = \chi^{1/2}(\alpha^\vee)$, by assuming $|\chi(\alpha^\vee)| < 1$ it follows that $|\bar{\chi}(\mathbf{X}_1)| < 1$ if $\alpha \in R_0$, and that $|\bar{\chi}(\mathbf{X}_2)| < 1$ if $\alpha \in R_2$; hence the geometric series $\sum_{j\geqslant 1}(\bar{\chi}(\mathbf{X}_1))^{2j}$ and $\sum_{j\geqslant 1}(\bar{\chi}(\mathbf{X}_2))^{2j}$ converge. Since the sum of these series does not depend on the choice of x and ω_0, (i) is proved.

(ii) The same argument, applied to the real character $|\chi|$, shows that

$$\int_{\Omega}\big|j^{\alpha}_{x,\chi}(\omega_0,\omega)\big|\,d\nu^{\alpha}_{x,\omega_0}(\omega) = k(\chi),$$

being $k(\chi)$ a real positive number. Hence, for any $f \in L^\infty(\Omega)$, and for every $\omega_0 \in \Omega$,

$$\left|J^\alpha_{x,\chi} f(\omega_0)\right| \leqslant \|f\|_\infty \int_\Omega \left|j^\alpha_{x,\chi}(\omega_0,\omega)\right| d\nu^\alpha_{x,\omega_0}(\omega) = k(\chi)\|f\|_\infty.$$

This proves that $J^\alpha_{x,\chi} f$ belongs to $L^\infty(\Omega)$ and that $J^\alpha_{x,\chi}$ is a bounded operator on $L^\infty(\Omega)$. □

Remark 6.9 The constant $c(\chi)$ is equal to 1 for $\chi = \chi_0^{-1}$.

Definition 6.10 Let $x, y \in \widehat{\mathscr{V}}(\Delta)$. Denote by $T^\chi_{x,y}$ the operator

$$T^\chi_{x,y}(f)(\omega) = P^{\chi\chi_0^{-1}}(x, y, \omega) f(\omega), \quad \text{for } \omega \in \Omega.$$

For every $x, y \in \widehat{\mathscr{V}}(\Delta)$, $T^\chi_{x,y}$ is bounded on $L^\infty(\Omega)$, because $P^{\chi\chi_0^{-1}}(x, y, \cdot)$ is locally constant on Ω, hence bounded by compactness.

Proposition 6.11 *Assume* $|\chi(\alpha^\vee)| < 1$. *For every pair of vertices* $x, y \in \widehat{\mathscr{V}}(\Delta)$,

$$J^\alpha_{y,\chi} \circ T^{\chi\chi_0^{1/2}}_{x,y} = T^{\chi^{s_\alpha}\chi_0^{1/2}}_{x,y} \circ J^\alpha_{x,\chi}.$$

Proof By Theorem 6.8, the assumption $|\chi(\alpha^\vee)| < 1$ assures that, for every pair $x, y \in \widehat{\mathscr{V}}(\Delta)$, the operators $J^\alpha_{x,\chi}$, $J^\alpha_{y,\chi}$ are bounded on the space $L^\infty(\Omega)$. By Definitions 6.5, 6.7 and 6.10, for every function f and every boundary point ω_0,

$$\begin{aligned}
&\left(T^{\chi^{s_\alpha}\chi_0^{1/2}}_{x,y} \circ J^\alpha_{x,\chi}\right) f(\omega_0)\\
&\quad = P^{\chi^{s_\alpha}\chi_0^{-1/2}}(x, y, \omega_0) \int_\Omega j^\alpha_{x,\chi}(\omega_0,\omega) f(\omega)\, d\nu^\alpha_{x,\omega_0}(\omega)\\
&\quad = \int_\Omega j^\alpha_{x,\chi}(\omega_0,\omega) P^{\chi^{s_\alpha}\chi_0^{-1/2}}(x, y, \omega_0) f(\omega)\, d\nu^\alpha_{x,\omega_0}(\omega)\\
&\quad = \int_\Omega \frac{j^\alpha_{x,\chi}(\omega_0,\omega)}{j^\alpha_{y,\chi}(\omega_0,\omega)} j^\alpha_{y,\chi}(\omega_0,\omega) P^{\chi^{s_\alpha}\chi_0^{-1/2}}(x, y, \omega_0) f(\omega) \frac{d\nu^\alpha_{x,\omega_0}(\omega)}{d\nu^\alpha_{y,\omega_0}(\omega)}\, d\nu^\alpha_{y,\omega_0}(\omega).
\end{aligned}$$

Definition 6.5 implies that, for any e lying in any apartment that contains ω_0 and ω,

$$\begin{aligned}
\frac{j^\alpha_{x,\chi}(\omega_0,\omega)}{j^\alpha_{y,\chi}(\omega_0,\omega)} &= \frac{\chi\chi_0^{1/2}(P_\alpha(\rho_{\omega_0}(e) + \rho_\omega(e) - \rho_{\omega_0}(x) - \rho_\omega(x))}{\chi\chi_0^{1/2}(P_\alpha(\rho_{\omega_0}(e) + \rho_\omega(e) - \rho_{\omega_0}(y) - \rho_\omega(y))}\\
&= \frac{\chi\chi_0^{1/2}(P_\alpha(-\rho_{\omega_0}(x) - \rho_\omega(x))}{\chi\chi_0^{1/2}(P_\alpha(-\rho_{\omega_0}(y) - \rho_\omega(y))}.
\end{aligned}$$

Moreover, according to the definition of the measure $\nu^{\alpha}_{x,\omega_0}$,

$$\frac{d\nu^{\alpha}_{x,\omega_0}(\omega)}{d\nu^{\alpha}_{y,\omega_0}(\omega)} = \chi_0(P_\alpha\big(\rho_\omega(x) - \rho_\omega(y)\big).$$

Therefore

$$\begin{aligned}
\frac{j^{\alpha}_{x,\chi}(\omega_0,\omega)}{j^{\alpha}_{y,\chi}(\omega_0,\omega)}\,\frac{d\nu^{\alpha}_{x,\omega_0}(\omega)}{d\nu^{\alpha}_{y,\omega_0}(\omega)} &= \frac{\chi\chi_0^{1/2}(P_\alpha(-\rho_{\omega_0}(x) - \rho_\omega(x)))}{\chi\chi_0^{1/2}(P_\alpha(-\rho_{\omega_0}(y) - \rho_\omega(y)))}\chi_0\big(P_\alpha\big(\rho_\omega(x) - \rho_\omega(y)\big)\big)\\
&= \frac{\chi(P_\alpha(\rho_\omega(y) - \rho_\omega(x)))}{\chi(P_\alpha(\rho_{\omega_0}(x) - \rho_{\omega_0}(y)))}\chi_0^{1/2}\big(P_\alpha\big(\rho_{\omega_0}(y) - \rho_{\omega_0}(x)\big)\big)\\
&\quad \times \chi_0^{-1/2}\big(P_\alpha\big(\rho_\omega(y) - \rho_\omega(x)\big)\big)\\
&= \frac{\chi\chi_0^{-1/2}(P_\alpha(\rho_\omega(y) - \rho_\omega(x)))}{\chi^{s_\alpha}\chi_0^{-1/2}(P_\alpha(\rho_{\omega_0}(y) - \rho_{\omega_0}(x)))}.
\end{aligned}$$

Recall also that $Q_\alpha(\rho_{\omega_0}(y) - \rho_{\omega_0}(x)) = Q_\alpha(\rho_\omega(y) - \rho_\omega(x))$ (Proposition 4.13 of Sect. 4.4). Therefore

$$\frac{j^{\alpha}_{x,\chi}(\omega_0,\omega)}{j^{\alpha}_{y,\chi}(\omega_0,\omega)}\,\frac{d\nu^{\alpha}_{x,\omega_0}(\omega)}{d\nu^{\alpha}_{y,\omega_0}(\omega)} = \frac{\chi\chi_0^{-1/2}(\rho_\omega(y) - \rho_\omega(x))}{\chi^{s_\alpha}\chi_0^{-1/2}(\rho_{\omega_0}(y) - \rho_{\omega_0}(x))} = \frac{P^{\chi\chi_0^{-1/2}}(x,y,\omega)}{P^{\chi^{s_\alpha}\chi_0^{-1/2}}(x,y,\omega_0)}.$$

Therefore

$$\begin{aligned}
&\big(T^{\chi^{s_\alpha}\chi_0^{1/2}}_{x,y} \circ J^{\alpha}_{x,\chi}\big) f(\omega_0)\\
&\quad = \int_\Omega j^{\alpha}_{y,\chi}(\omega_0,\omega)\frac{P^{\chi\chi_0^{-1/2}}(x,y,\omega)}{P^{\chi^{s_\alpha}\chi_0^{-1/2}}(x,y,\omega_0)}P^{\chi^{s_\alpha}\chi_0^{-1/2}}(x,y,\omega_0) f(\omega)\,d\nu^{\alpha}_{y,\omega_0}(\omega)\\
&\quad = \int_\Omega j^{\alpha}_{y,\chi}(\omega_0,\omega)P^{\chi\chi_0^{-1/2}}(x,y,\omega) f(\omega)\,d\nu^{\alpha}_{y,\omega_0}(\omega)\\
&\quad = \int_\Omega j^{\alpha}_{y,\chi}(\omega_0,\omega)T^{\chi\chi_0^{1/2}}_{x,y}(f)(\omega)\,d\nu^{\alpha}_{y,\omega_0}(\omega) = \big(J^{\alpha}_{y,\chi} \circ T^{\chi\chi_0^{1/2}}_{x,y}\big) f(\omega_0). \qquad \square
\end{aligned}$$

6.4 W-*Invariance of the Eigenvalues*

Theorem 6.12 *For every character* χ *and for every simple root* α,

$$\Lambda^{\chi\chi_0^{1/2}} = \Lambda^{\chi^{s_\alpha}\chi_0^{1/2}}. \tag{8}$$

Proof Let χ be a character. We split the proof into three parts.

(i) At first, assume $|\chi(\alpha^\vee)| > 1$. Then $|\chi^{-1}(\alpha^\vee)| < 1$, and Theorem 6.8 of the previous section implies that, for every $x, y \in \widehat{\mathscr{V}}(\Delta)$, $J^\alpha_{x,\chi^{-1}}$ and $J^\alpha_{y,\chi^{-1}}$ are bounded operators on $L^\infty(\Omega)$. Therefore it follows from Proposition 6.11 of the previous section that, for every $x, y \in \widehat{\mathscr{V}}(\Delta)$,

$$J^\alpha_{y,\chi^{-1}} \circ T^{\chi^{-1}\chi_0^{1/2}}_{x,y} \mathbf{1}(\omega) = T^{(\chi^{s_\alpha})^{-1}\chi_0^{1/2}}_{x,y} \circ J^\alpha_{x,\chi^{-1}} \mathbf{1}(\omega) \quad \text{for every } \omega \in \Omega,$$

since $(\chi^{s_\alpha})^{-1} = (\chi^{-1})^{s_\alpha}$. Now fix $y \in \widehat{\mathscr{V}}(\Delta)$ and sum on all $x \in \mathscr{V}_\lambda(y)$ for $\lambda \in \widehat{L}^+$. By linearity,

$$\begin{aligned}\sum_{x\in\mathscr{V}_\lambda(y)} J^\alpha_{y,\chi^{-1}} \circ T^{\chi^{-1}\chi_0^{1/2}}_{x,y} \mathbf{1} &= J^\alpha_{y,\chi^{-1}} \circ \sum_{x\in\mathscr{V}_\lambda(y)} T^{\chi^{-1}\chi_0^{1/2}}_{x,y} \mathbf{1} \\ &= J^\alpha_{y,\chi^{-1}} \Bigg(\sum_{x\in\mathscr{V}_\lambda(y)} P^{\chi^{-1}\chi_0^{-1/2}}(x, y, \cdot) \Bigg).\end{aligned}$$

Moreover $\sum_{x\in\mathscr{V}_\lambda(y)} P^{\chi^{-1}\chi_0^{-1/2}}(x, y, \omega) = \sum_{x\in\mathscr{V}_\lambda(y)} P^{\chi\chi_0^{1/2}}(y, x, \omega) = \Lambda^{\chi\chi_0^{1/2}}(\lambda)$. Therefore, for every $\omega \in \Omega$,

$$\begin{aligned}\sum_{x\in\mathscr{V}_\lambda(y)} J^\alpha_{y,\chi^{-1}} \circ T^{\chi^{-1}\chi_0^{1/2}}_{x,y} \mathbf{1}(\omega) &= J^\alpha_{y,\chi^{-1}} \big(\Lambda^{\chi\chi_0^{1/2}}(\lambda)\mathbf{1}\big)(\omega) \\ &= \Lambda^{\chi\chi_0^{1/2}}(\lambda) J^\alpha_{y,\chi^{-1}} \mathbf{1}(\omega) = \Lambda^{\chi\chi_0^{1/2}}(\lambda) c\big(\chi^{-1}\big).\end{aligned}$$

On the other hand,

$$\begin{aligned}&\sum_{x\in\mathscr{V}_\lambda(y)} T^{(\chi^{s_\alpha})^{-1}\chi_0^{1/2}}_{x,y} \circ J^\alpha_{x,\chi^{-1}} \mathbf{1}(\omega) \\ &= \sum_{x\in\mathscr{V}_\lambda(y)} T^{(\chi^{s_\alpha})^{-1}\chi_0^{1/2}}_{x,y} \big(c\big(\chi^{-1}\big)\mathbf{1}\big)(\omega) = c\big(\chi^{-1}\big) \sum_{x\in\mathscr{V}_\lambda(y)} T^{(\chi^{s_\alpha})^{-1}\chi_0^{1/2}}_{x,y} \mathbf{1}(\omega) \\ &= c\big(\chi^{-1}\big) \sum_{x\in\mathscr{V}_\lambda(y)} P^{(\chi^{s_\alpha})^{-1}\chi_0^{-1/2}}(x, y, \omega) \\ &= c\big(\chi^{-1}\big) \sum_{x\in\mathscr{V}_\lambda(y)} P^{\chi^{s_\alpha}\chi_0^{1/2}}(y, x, \omega) = c\big(\chi^{-1}\big) \Lambda^{\chi^{s_\alpha}\chi_0^{1/2}}(\lambda).\end{aligned}$$

If $c(\chi^{-1})$ is a real number different from zero, the identity

$$c\big(\chi^{-1}\big)\Lambda^{\chi\chi_0^{1/2}}(\lambda) = c\big(\chi^{-1}\big)\Lambda^{\chi^{s_\alpha}\chi_0^{1/2}}(\lambda)$$

implies $\Lambda^{\chi\chi_0^{1/2}}(\lambda) = \Lambda^{\chi^{s_\alpha}\chi_0^{1/2}}(\lambda)$ for every $\lambda \in \widehat{L}$.

On the other hand, $c(\chi^{-1})$ can be zero for at most two values of $\chi(\alpha^\vee)$. So the previous identity follows also in this case from a standard argument of continuity.

(ii) If $|\chi(\alpha^\vee)| < 1$, then $|\chi^{s_\alpha}(\alpha^\vee)| > 1$, and therefore, by part (i) of this proof,

$$\Lambda^{\chi^{s_\alpha}\chi_0^{1/2}} = \Lambda^{\chi^{s_\alpha^2}\chi_0^{1/2}} = \Lambda^{\chi\chi_0^{1/2}}.$$

(iii) Finally, if $|\chi(\alpha^\vee)| = 1$, the required identity can be proved by a standard argument of continuity, as the eigenvalue associated to a character χ depends continuously on χ, with respect to the weak topology on the space $\mathrm{Hom}(\widehat{L}, \mathbb{C})$. Indeed, there exists a character χ', with $|\chi'(\alpha^\vee)| < 1$, arbitrarily close to χ. □

Corollary 6.13 *For every character χ and for every $\mathbf{w} \in \mathbf{W}$,*

$$\Lambda^{\chi\chi_0^{1/2}} = \Lambda^{\chi^{\mathbf{w}}\chi_0^{1/2}}.$$

6.5 Technical Results About the Poisson Transform

Let us study the relationship between the Poisson transform and the operators defined in Sect. 6.3.

Proposition 6.14 *For every pair $x, y \in \widehat{\mathscr{V}}(\Delta)$, and for every $f \in L^\infty(\Omega)$,*

$$\mathscr{P}_y^\chi\big(T_{x,y}^\chi f\big) = \mathscr{P}_x^\chi(f).$$

Proof For every vertex $z \in \widehat{\mathscr{V}}(\Delta)$,

$$\begin{aligned}
&\mathscr{P}_y^\chi\big(T_{x,y}^\chi f\big)(z)\\
&\quad= \int_\Omega P^\chi(y,z,\omega) P^{\chi\chi_0^{-1}}(x,y,\omega) f(\omega)\, d\nu_y(\omega)\\
&\quad= \int_\Omega \chi\big(\rho_\omega(z) - \rho_\omega(y)\big)\chi\big(\rho_\omega(y) - \rho_\omega(x)\big) f(\omega)\chi_0\big(\rho_\omega(x) - \rho_\omega(y)\big)\, d\nu_y(\omega)\\
&\quad= \int_\Omega \chi\big(\rho_\omega(z) - \rho_\omega(x)\big) f(\omega)\frac{d\nu_x(\omega)}{d\nu_y(\omega)}\, d\nu_y(\omega) = \int_\Omega P^\chi(x,z,\omega) f(\omega)\, d\nu_x(\omega)\\
&\quad= \mathscr{P}_x^\chi f(z).
\end{aligned}$$

□

By Corollary 6.4 of Sect. 6.2, for every $f \in L^\infty(\Omega)$, $\mathscr{P}_x^{\chi\chi_0^{1/2}}(f)$ and $\mathscr{P}_x^{\chi^{s_\alpha}\chi_0^{1/2}}(f)$ are eigenfunctions of the algebra $\mathscr{H}(\Delta)$, associated to eigenvalues $\Lambda^{\chi\chi_0^{1/2}}$ and $\Lambda^{\chi^{s_\alpha}\chi_0^{1/2}}$ respectively. Since $\Lambda^{\chi\chi_0^{1/2}} = \Lambda^{\chi^{s_\alpha}\chi_0^{1/2}}$ by Theorem 6.12 of the previous section, then, for every $f \in L^\infty(\Omega)$, $\mathscr{P}_x^{\chi\chi_0^{1/2}}(f)$ and $\mathscr{P}_x^{\chi^{s_\alpha}\chi_0^{1/2}}(f)$ are eigenfunctions associated to the same eigenvalue. When $|\chi(\alpha^\vee)| < 1$, the following theorem, for every $f \in L^\infty(\Omega)$, constructs a function $g \in L^\infty(\Omega)$ such that

$$\mathscr{P}_x^{\chi^{s_\alpha}\chi_0^{1/2}}(g) = c(\chi)\mathscr{P}_x^{\chi\chi_0^{1/2}}(f).$$

Theorem 6.15 *Assume that* $|\chi(\alpha^\vee)| < 1$. *Then, for every* $x \in \widehat{\mathscr{V}}(\Delta)$ *and for every* $f \in L^\infty(\Omega)$,

$$\mathscr{P}_x^{\chi^{s_\alpha}\chi_0^{1/2}}\big(J_{x,\chi}^\alpha f\big) = c(\chi)\mathscr{P}_x^{\chi\chi_0^{1/2}}(f). \tag{9}$$

Proof Let $f \in L^\infty(\Omega)$. We split the proof into two parts.

(i) For any given x, we first prove (9) at the point $y = x$ (that is, when the variable of both sides has value x).

Since $P^{\chi^{s_\alpha}\chi_0^{1/2}}(x, x, \omega) = 1$, then

$$\mathscr{P}_x^{\chi^{s_\alpha}\chi_0^{1/2}}\big(J_{x,\chi}^\alpha f\big)(x) = \int_\Omega J_{x,\chi}^\alpha f(\omega_0)\, d\nu_x(\omega_0).$$

Thus, by Definition 6.7 of Sect. 6.3,

$$\mathscr{P}_x^{\chi^{s_\alpha}\chi_0^{1/2}}\big(J_{x,\chi}^\alpha f\big)(x) = \int_\Omega \left(\int_\Omega j_{x,\chi}^\alpha(\omega_0, \omega) f(\omega)\, d\nu_{x,\omega_0}^\alpha(\omega)\right) d\nu_x(\omega_0).$$

By taking into account that, for every ω_0,

$$\nu_x(\omega_0) = \nu_x^\alpha(\omega_0) \times \nu_{x,\omega_0}^\alpha(\omega_0),$$

the measure $\nu_{x,\omega}^\alpha$ can be regarded as the restriction of the measure ν_x to the subset $\{\omega' \in \Omega : \omega' \in [\omega]_\alpha\}$. Then, setting $j_{x,\chi}^\alpha(\omega_0, \omega) = 0$ for $\omega \notin [\omega_0]_\alpha$, we obtain

$$\mathscr{P}_x^{\chi^{s_\alpha}\chi_0^{1/2}}\big(J_{x,\chi}^\alpha f\big)(x) = \int_\Omega \left(\int_\Omega j_{x,\chi}^\alpha(\omega_0, \omega) f(\omega)\, d\nu_x(\omega)\right) d\nu_x(\omega_0).$$

On the other hand,

$$\begin{aligned} &\int_\Omega \left(\int_\Omega j_{x,\chi}^\alpha(\omega_0, \omega) f(\omega)\, d\nu_x(\omega)\right) d\nu_x(\omega_0) \\ &\quad = \int_\Omega \left(\int_\Omega j_{x,\chi}^\alpha(\omega_0, \omega)\, d\nu_x(\omega_0)\right) f(\omega)\, d\nu_x(\omega), \end{aligned}$$

since the integral is absolutely convergent. Therefore

$$\begin{aligned} \mathscr{P}_x^{\chi^{s_\alpha}\chi_0^{1/2}}\big(J_{x,\chi}^\alpha f\big)(x) &= \int_\Omega \left(\int_\Omega j_{x,\chi}^\alpha(\omega_0, \omega)\, d\nu_x(\omega_0)\right) f(\omega)\, d\nu_x(\omega) \\ &= \int_\Omega \left(\int_\Omega j_{x,\chi}^\alpha(\omega, \omega_0)\, d\nu_x(\omega_0)\right) f(\omega)\, d\nu_x(\omega) \end{aligned}$$

$$= \int_{\Omega} J^{\alpha}_{x,\chi} \mathbf{1}(\omega) f(\omega)\, d\nu_x(\omega)$$
$$= c(\chi) \int_{\Omega} f(\omega)\, d\nu_x(\omega) = c(\chi) \mathscr{P}_x^{\chi \chi_0^{1/2}}(f)(x).$$

(ii) Choose now the variable y to be any other vertex: that is, let $x, y \in \widehat{\mathscr{V}}(\Delta)$ with $y \neq x$. By Proposition 6.14, $\mathscr{P}_x^{\chi} f(y) = \mathscr{P}_y^{\chi}(T^{\chi}_{x,y} f)(y)$. Hence, if we apply (i), replacing x with y and f with $T^{\chi}_{x,y} f$, we obtain

$$\mathscr{P}_y^{\chi^{s_\alpha} \chi_0^{1/2}} \big(J^{\alpha}_{y,\chi}\big(T^{\chi\chi_0^{1/2}}_{x,y} f\big)\big)(y) = c(\chi) \mathscr{P}_y^{\chi\chi_0^{1/2}} \big(T^{\chi\chi_0^{1/2}}_{x,y} f\big)(y) = c(\chi) \mathscr{P}_x^{\chi\chi_0^{1/2}} f(y).$$

On the other hand, by Proposition 6.11 of Sect. 6.3,

$$\mathscr{P}_y^{\chi^{s_\alpha} \chi_0^{1/2}} \big(J^{\alpha}_{y,\chi}\big(T^{\chi\chi_0^{1/2}}_{x,y} f\big)\big)(y) = \mathscr{P}_y^{\chi^{s_\alpha} \chi_0^{1/2}} \big(T^{\chi^{s_\alpha} \chi_0^{1/2}}_{x,y} \big(J^{\alpha}_{x,\chi} f\big)\big)(y).$$

Again by Proposition 6.14 we conclude that

$$\mathscr{P}_x^{\chi^{s_\alpha} \chi_0^{1/2}} \big(J^{\alpha}_{x,\chi} f\big)(y) = c(\chi) \mathscr{P}_x^{\chi\chi_0^{1/2}} f(y). \qquad \square$$

Remark 6.16 Theorem 6.15 provides a different proof of the identity $\Lambda^{\chi^{s_\alpha} \chi_0^{1/2}} = \Lambda^{\chi\chi_0^{1/2}}$, when $|\chi(\alpha^\vee)| < 1$. Indeed, for every $f \in L^\infty(\Omega)$, the function $\mathscr{P}_x^{\chi^{s_\alpha} \chi_0^{1/2}}(f)$ is an eigenfunction of the algebra $\mathscr{H}(\Delta)$ associated with the eigenvalue $\Lambda^{\chi^{s_\alpha} \chi_0^{1/2}}$ and, when $|\chi(\alpha^\vee)| < 1$, $J^{\alpha}_{x,\chi} f$ belongs to $L^\infty(\Omega)$. Then

$$A_\lambda\big(\mathscr{P}_x^{\chi^{s_\alpha} \chi_0^{1/2}} \big(J^{\alpha}_{x,\chi} f\big)\big) = \Lambda^{\chi^{s_\alpha} \chi_0^{1/2}} \mathscr{P}_x^{\chi^{s_\alpha} \chi_0^{1/2}} \big(J^{\alpha}_{x,\chi} f\big), \quad \text{for every } \lambda \in \widehat{L}.$$

On the other hand, for every $f \in L^\infty(\Omega)$, $\mathscr{P}_x^{\chi\chi_0^{1/2}}(f)$ is an eigenfunction of the algebra $\mathscr{H}(\Delta)$ associated with the eigenvalue $\Lambda^{\chi\chi_0^{1/2}}$. Therefore

$$A_\lambda\big(c(\chi) \mathscr{P}_x^{\chi\chi_0^{1/2}}(f)\big) = \Lambda^{\chi\chi_0^{1/2}} c(\chi) \mathscr{P}_x^{\chi\chi_0^{1/2}}(f), \quad \text{for every } \lambda \in \widehat{L}.$$

Now, by Theorem 6.15,

$$A_\lambda\big(\mathscr{P}_x^{\chi^{s_\alpha} \chi_0^{1/2}} \big(J^{\alpha}_{x,\chi} f\big)\big) = \Lambda^{\chi\chi_0^{1/2}} \mathscr{P}_x^{\chi^{s_\alpha} \chi_0^{1/2}} \big(J^{\alpha}_{x,\chi} f\big), \quad \text{for every } \lambda \in \widehat{L}.$$

So we have proved that, if $|\chi(\alpha^\vee)| < 1$, then, for every $f \in L^\infty(\Omega)$, $\mathscr{P}_x^{\chi^{s_\alpha} \chi_0^{1/2}}(J^{\alpha}_{x,\chi} f)$ belongs to the eigenspaces associated to both the eigenvalues $\Lambda^{\chi^{s_\alpha} \chi_0^{1/2}}$ and $\Lambda^{\chi\chi_0^{1/2}}$. This implies that $\Lambda^{\chi^{s_\alpha} \chi_0^{1/2}} = \Lambda^{\chi\chi_0^{1/2}}$.

7 Satake Isomorphism

7.1 Convolution Operators on $\mathbb{A}$

The set $\widehat{\mathscr{V}}(\mathbb{A}) = \widehat{L}$ can be identified with $\mathbb{Z}^n$, by identifying each $\lambda \in \widehat{L}$ with $(m_1, \ldots, m_n)$ of $\mathbb{Z}^n$ if $\lambda = \sum_{i=1}^n m_i \lambda_i$. Hence $\widehat{L}$ inherits the structure of finitely generated free abelian group of $\mathbb{Z}^n$. We denote by $\mathscr{L}(\widehat{L})$ the $\mathbb{C}$-algebra of all complex-valued functions on $\widehat{L}$, with finite support. Each function h in $\mathscr{L}(\widehat{L})$ determines a convolution operator on the functions defined on $\widehat{L}$

$$\tau_h(F) = h * F.$$

Proposition 7.1 *Every character χ on $\mathbb{A}$ is an eigenfunction of all operators τ_h, $h \in \mathscr{L}(\widehat{L})$, with associated eigenvalue $\Theta^\chi(h) = \sum_{\mu \in \widehat{L}} h(\mu)\chi(\mu)$.*

Proof For every $\lambda \in \widehat{L}$,

$$(\tau_h \chi)(\lambda) = \sum_{\mu \in \widehat{L}} h(\mu)\chi(\lambda + \mu) = \Big(\sum_{\mu \in \widehat{L}} h(\mu)\chi(\mu)\Big)\chi(\lambda).$$

□

Proposition 7.2 *Let $h \in \mathscr{L}(\widehat{L})$. Then*

$$h = 0 \quad \Longleftrightarrow \quad \Theta^\chi(h) = 0 \quad \text{for all } \chi \in \operatorname{Hom}\big(\widehat{L}, \mathbb{C}^\times\big).$$

Proof There is a natural identification of $\widehat{L}$ with the group T of all translations t_λ, $\lambda \in \widehat{L}$. Hence $\mathscr{L}(\widehat{L})$ is the algebra $\mathscr{L}(T)$ defined by [7, (1.1)]. Using this identification and following notation of [7], the mapping

$$h \mapsto \sum_{\lambda \in \widehat{L}} h(\lambda)\lambda,$$

is a $\mathbb{C}$-algebra isomorphism of $\mathscr{L}(\widehat{L})$ onto the group algebra $\mathbb{C}[\widehat{L}]$ of $\widehat{L}$ over $\mathbb{C}$. Since $\widehat{L}$ is a free abelian group generated by the finite set $\{\lambda_1, \ldots, \lambda_n\}$, it follows that $\mathbb{C}[\widehat{L}] = \mathbb{C}[\pm\lambda_i, i = 1, \ldots, n]$, hence it is a commutative integral domain. Consequently $\mathbb{C}[\widehat{L}]$ is the coordinate ring of an affine algebraic variety, say S, whose points are the $\mathbb{C}$-algebra homomorphisms $s : \mathbb{C}[\widehat{L}] \to \mathbb{C}$. The restriction of these homomorphisms to $\widehat{L}$ gives a bijection of S onto $\mathbb{X}(\widehat{L}) = \operatorname{Hom}(\widehat{L}, \mathbb{C}^\times)$ and we identify $\mathbb{X}(\widehat{L})$ with S in this way. The elements of $\mathbb{C}[\widehat{L}]$ can therefore be regarded as functions on $\mathbb{X}(\widehat{L})$. Hence, by the Nullstellensatz, if $\eta \in \mathbb{C}[\widehat{L}]$,

$$\eta = 0 \quad \Longleftrightarrow \quad \chi(\eta) = 0 \quad \text{for all } \chi \in \mathbb{X}(\widehat{L}).$$

Keeping in mind the $\mathbb{C}$-algebra isomorphism of $\mathscr{L}(\widehat{L})$ onto $\mathbb{C}[\widehat{L}]$, each χ defines a homomorphism $\mathscr{L}(\widehat{L}) \to \mathbb{C}$, namely

$$\chi(h) = \sum_{\lambda \in \widehat{L}} h(\lambda)\chi(\lambda),$$

and $h=0$ if and only if $\chi(h)=0$ for all $\chi\in\mathbb{X}(\widehat{L})$. On the other hand, for every h in $\mathscr{L}(\widehat{L})$, $\chi(h)=\Theta^{\chi}(h)$, according to Proposition 7.1. Hence

$$h=0 \quad\Longleftrightarrow\quad \Theta^{\chi}(h)=0, \quad \text{for all } \chi\in\mathbb{X}(\widehat{L}). \qquad \square$$

7.2 *The Hecke Algebra on* $\mathbb{A}$

The group $\mathbf{W}$ acts on $\mathscr{L}(\widehat{L})$ in the following way: for every $h\in\mathscr{L}(\widehat{L})$,

$$h^{\mathbf{w}}(\lambda)=(\mathbf{w}h)(\lambda)=h\big(\mathbf{w}^{-1}(\lambda)\big) \quad \text{for every } \lambda\in\widehat{L}.$$

We denote by $\mathscr{L}(\widehat{L})^{\mathbf{W}}$ the subring of $\mathscr{L}(\widehat{L})$ that consists of all $\mathbf{W}$-invariant functions in $\mathscr{L}(\widehat{L})$, that is, the functions h in $\mathscr{L}(\widehat{L})$ such that $h^{\mathbf{w}}=h$ for every $\mathbf{w}\in\mathbf{W}$.

Proposition 7.3 *For every h in $\mathscr{L}(\widehat{L})^{\mathbf{W}}$, the operator τ_h is $\mathbf{W}$-invariant.*

Proof Fix $\mathbf{w}\in\mathbf{W}$. For every function F and every λ, it follows from the $\mathbf{W}$-invariance of h that

$$(\tau_h F)\big(\mathbf{w}^{-1}(\lambda)\big)=\sum_{\mu\in\widehat{L}} h(\mu)F\big(\mathbf{w}^{-1}(\lambda)+\mu\big)=\sum_{\mu\in\widehat{L}} h\big(\mathbf{w}(\mu)\big)F\big(\mathbf{w}^{-1}(\lambda)+\mu\big).$$

By setting $\mathbf{w}(\mu)=\mu'$, this becomes

$$\begin{aligned}(\tau_h F)\big(\mathbf{w}^{-1}(\lambda)\big)&=\sum_{\mu'\in\widehat{L}} h\big(\mu'\big)F\big(\mathbf{w}^{-1}(\lambda)+\mathbf{w}^{-1}\big(\mu'\big)\big)=\sum_{\mu'\in\widehat{L}} h\big(\mu'\big)F\big(\mathbf{w}^{-1}\big(\lambda+\mu'\big)\big)\\&=\sum_{\mu'\in\widehat{L}} h\big(\mu'\big)F^{\mathbf{w}}\big(\lambda+\mu'\big)=\big(\tau_h F^{\mathbf{w}}\big)(\lambda).\end{aligned}$$

This proves the invariance property in the statement. $\square$

Let

$$\mathscr{H}(\mathbb{A})=\big\{\tau_h, h\in\mathscr{L}(\widehat{L})^{\mathbf{W}}\big\}.$$

Obviously, $\mathscr{H}(\mathbb{A})$ is a $\mathbb{C}$-algebra; following Humphreys [5], we call $\mathscr{H}(\mathbb{A})$ the *Hecke algebra on* $\mathbb{A}$.

Proposition 7.1 of the previous section implies that every character χ on $\widehat{L}$ is an eigenfunction of the whole algebra $\mathscr{H}(\mathbb{A})$. We denote by Θ^{χ} the associated eigenvalue, that is, the homomorphism from the algebra $\mathscr{H}(\mathbb{A})$ to $\mathbb{C}^{\times}$ such that, for every $h\in\mathscr{L}(\widehat{L})^{\mathbf{W}}$,

$$\Theta^{\chi}(\tau_h)=\Theta^{\chi}(h)=\sum_{\mu\in\widehat{L}} h(\mu)\chi(\mu).$$

Note that, since $\mathbf{W}$ is finite, $\mathscr{L}(\widehat{L})$ is integral over $\mathscr{L}(\widehat{L})^{\mathbf{W}}$. Therefore every eigenvalue of $\mathscr{L}(\widehat{L})^{\mathbf{W}}$ may be extended as an eigenvalue of $\mathscr{L}(\widehat{L})$. Keeping in mind that the restriction to $\widehat{L}$ of Θ^{χ} is the character χ, we easily obtain the following proposition.

Proposition 7.4 *For every eigenvalue Θ of the Hecke algebra of $\mathbb{A}$ there exists a character χ on $\widehat{L}$ such that*

$$\Theta = \Theta^{\chi}.$$

Proof For every $\lambda \in \widehat{L}$, let δ_λ be the function on $\widehat{L}$ such that $\delta_\lambda(\lambda) = 1$ and $\delta_\lambda(\mu) = 0$ for every $\mu \neq \lambda$. Then each $h \in \mathscr{L}(\widehat{L})^{\mathbf{W}}$ can be written as $h = \sum_\lambda h(\lambda)\delta_\lambda$. Let Θ be any eigenvalue of $\mathscr{H}(\mathbb{A})$ and χ its restriction to $\widehat{L}$

$$\chi(\lambda) = \Theta(\delta_\lambda), \quad \text{for every } \lambda \in \widehat{L}.$$

It is immediate to observe that χ belongs to $\mathbb{X}(\widehat{L})$ and, for every $h \in \mathscr{L}(\widehat{L})^{\mathbf{W}}$,

$$\Theta(h) = \sum_\lambda h(\lambda)\Theta(\delta_\lambda) = \sum_\lambda h(\lambda)\chi(\lambda) = \Theta^{\chi}(h).$$

□

7.3 Operators $\widetilde{A}_\lambda$

Fix $\omega \in \Omega$. For every $\lambda \in \widehat{L}^+$ and every vertex $\mu \in \widehat{L}$, the number $N(\lambda, \mu)$ defined in (5) does not depend on the choice of ω. For every $\lambda \in \widehat{L}^+$, let h_λ be the following function on $\widehat{L}$:

$$h_\lambda(\mu) = \chi_0^{1/2}(\mu)N(\lambda, \mu), \quad \text{for every } \mu \in \widehat{L}.$$

As $N(\lambda, \mu) = 0$ except for finitely many $\mu \in \widehat{L}$, we see that $h_\lambda \in \mathscr{L}(\widehat{L})$.

Definition 7.5 For every $\lambda \in \widehat{L}^+$, denote by $\widetilde{A}_\lambda$ the convolution operator associated with the function h_λ: for every function F on $\widehat{L}$,

$$\widetilde{A}_\lambda F(\mu) = h_\lambda * F(\mu) = \sum_{\mu' \in \widehat{L}} N(\lambda, \mu')\chi_0^{1/2}(\mu')F(\mu + \mu'), \quad \text{for } \mu \in \widehat{L}.$$

By Proposition 7.1 of Sect. 7.1, every character χ on $\widehat{L}$ is an eigenfunction of $\widetilde{A}_\lambda$, $\lambda \in \widehat{L}$, with associated eigenvalue

$$\Theta^{\chi}(\lambda) = \Theta^{\chi}(h_\lambda) = \sum_{\mu \in \widehat{L}} h_\lambda(\mu)\chi(\mu) = \sum_{\mu \in \widehat{L}} N(\lambda, \mu)\chi_0^{1/2}(\mu)\chi(\mu).$$

Then (7) implies that

$$\Theta^{\chi}(\lambda) = \Lambda^{\chi\chi_0^{1/2}}(\lambda). \tag{10}$$

Proposition 7.6 *For every* $\mathbf{w} \in \mathbf{W}$, $h_\lambda = h_\lambda^{\mathbf{w}}$.

Proof Since the Weyl group $\mathbf{W}$ is generated by reflections s_α, $\alpha \in B$, we only need to prove that $h_\lambda = h_\lambda^{s_\alpha}$ for every simple root α. Fix s_α and consider

$$h_\lambda^{s_\alpha}(\mu) = \chi_0^{1/2}\big(s_\alpha(\mu)\big) N\big(\lambda, s_\alpha(\mu)\big), \quad \text{for } \mu \in \widehat{L}.$$

For every character χ and every $\mu \in \widehat{L}$,

$$h_\lambda * \chi(\mu) = \Theta^\chi(h_\lambda)\chi(\mu), \qquad h_\lambda^{s_\alpha} * \chi(\mu) = \Theta^\chi\big(h_\lambda^{s_\alpha}\big)\chi(\mu).$$

On the other hand, as remarked before,

$$\Theta^\chi(h_\lambda) = \sum_{\mu \in \widehat{L}} N(\lambda, \mu)\chi_0^{1/2}(\mu)\chi(\mu) = \Lambda^{\chi \chi_0^{1/2}}(\lambda),$$

and, with $\mu' = s_\alpha(\mu)$,

$$\begin{aligned}\Theta^\chi\big(h_\lambda^{s_\alpha}\big) &= \sum_{\mu \in \widehat{L}} N\big(\lambda, s_\alpha(\mu)\big)\chi_0^{1/2}\big(s_\alpha(\mu)\big)\chi(\mu) = \sum_{\mu' \in \widehat{L}} N\big(\lambda, \mu'\big)\chi_0^{1/2}\big(\mu'\big)\chi^{s_\alpha}\big(\mu'\big) \\ &= \Lambda^{\chi^{s_\alpha} \chi_0^{1/2}}(\lambda).\end{aligned}$$

Thus, Theorem 6.12 of Sect. 6.4 implies $\Theta^\chi(h_\lambda^{s_\alpha}) = \Theta^\chi(h_\lambda)$ for every χ. So $h_\lambda = h_\lambda^{s_\alpha}$, by Proposition 7.2 of Sect. 7.1. □

Corollary 7.7 *For every* $\lambda \in \widehat{L}^+$, *the operator* $\widetilde{A}_\lambda$ *belongs to the Hecke algebra* $\mathscr{H}(\mathbb{A})$.

Proposition 7.8 *The operators* $\widetilde{A}_\lambda$, $\lambda \in \widehat{L}^+$, *form a* $\mathbb{C}$-*basis of* $\mathscr{H}(\mathbb{A})$.

Proof We only need to show that the functions h_λ, $\lambda \in \widehat{L}^+$, form a $\mathbb{C}$-basis of $\mathscr{L}(\widehat{L})^{\mathbf{W}}$. For each $\lambda \in \widehat{L}^+$, let ξ_λ be the characteristic function of the $\mathbf{W}$-orbit of λ. The functions ξ_λ, as λ runs through $\widehat{L}^+$, form a $\mathbb{C}$-basis of $\mathscr{L}(\widehat{L})^{\mathbf{W}}$. Hence, by summing on all λ' in $\widehat{L}^+$,

$$h_\lambda = \sum_{\lambda' \in \widehat{L}^+} h_\lambda\big(\lambda'\big)\xi_{\lambda'}.$$

Since $N(\lambda, \lambda) = 1$, then $h_\lambda(\lambda) = \chi_0^{1/2}(\lambda)$. Consequently the previous sum takes the form

$$h_\lambda = \chi_0^{1/2}(\lambda)\xi_\lambda + \sum_{\lambda' \neq \lambda} h_\lambda\big(\lambda'\big)\xi_{\lambda'}.$$

In this sum $h_\lambda(\lambda') = 0$, but for $\lambda' \in \Pi_\lambda$. Since $\chi_0^{1/2}(\lambda) \neq 0$, we conclude that the h_λ form a $\mathbb{C}$-basis of $\mathscr{L}(\widehat{L})^{\mathbf{W}}$. □

Definition 7.9 For every $\lambda \in \widehat{L}^+$, let g_λ be the function defined as $g_\lambda(\mu) = N(\lambda, \mu)$, for $\mu \in \widehat{L}$. We denote by B_λ the following operator acting on the complex-valued functions F on $\widehat{L}$:

$$B_\lambda F(\mu) = g_\lambda \star F(\mu) = \sum_{\mu' \in \widehat{L}} N(\lambda, \mu') F(\mu + \mu'), \quad \text{for } \mu \in \widehat{L}.$$

Observe that the operator B_λ is linear and invariant with respect to any translation in $\mathbb{A}$, as their coefficients $N(\lambda, \mu')$ do not depend on μ. However, B_λ is not $\mathbf{W}$-invariant, because g_λ does not belong to $\mathscr{L}(\widehat{L})^{\mathbf{W}}$, as $N(\lambda, \mu) \neq N(\lambda, \mathbf{w}^{-1}\mu)$ for $\mathbf{w} \in \mathbf{W}$ different from the identity.

Proposition 7.10 *For every function F on $\widehat{L}$, let $f(x) = F(\rho_\omega(x))$, for $x \in \widehat{\mathscr{V}}(\Delta)$. Then, for every $\lambda \in \widehat{L}^+$,*

$$A_\lambda f(x) = B_\lambda F(\mu) \quad \textit{if } \mu = \rho_\omega(x).$$

Proof By definition of A_λ, for every function f,

$$A_\lambda(f)(x) = \sum_{y \in \mathscr{V}_\lambda(x)} f(y) = \sum_{\nu \in \widehat{L}} \left(\sum_{\{y : \sigma(x,y) = \lambda,\, \rho_\omega(y) = \nu\}} f(y) \right).$$

In the case $f(x) = F(\rho_\omega(x))$, for every $\nu \in \widehat{L}$ one has $f(y) = F(\nu)$ for all y such that $\rho_\omega(y) = \nu$. Hence, by setting $\mu = \rho_\omega(x)$ and $\mu + \mu' = \nu$,

$$A_\lambda(f)(x) = \sum_{\mu' \in \widehat{L}} N(\lambda, \mu') F(\mu + \mu') = B_\lambda F(\mu).$$

□

Proposition 7.11 *For every $\lambda \in \widehat{L}^+$ and every function F,*

$$\widetilde{A}_\lambda F = \chi_0^{-1/2} B_\lambda\left(\chi_0^{1/2} F\right), \qquad B_\lambda F = \chi_0^{1/2} \widetilde{A}_\lambda\left(\chi_0^{-1/2} F\right).$$

Proof For every $\mu \in \widehat{L}$, by Definitions 7.5 and 7.9,

$$\begin{aligned}
(\widetilde{A}_\lambda F)(\mu) &= \sum_{\mu' \in \widehat{L}} N(\lambda, \mu') \chi_0^{1/2}(\mu') F(\mu + \mu') \\
&= \chi_0^{-1/2}(\mu) \sum_{\mu' \in \widehat{L}} N(\lambda, \mu') \chi_0^{1/2}(\mu + \mu') F(\mu + \mu') \\
&= \chi_0^{-1/2}(\mu) B_\lambda\left(\chi_0^{1/2} F\right)(\mu).
\end{aligned}$$

Moreover

$$(B_\lambda F)(\mu) = \sum_{\mu' \in \widehat{L}} N(\lambda, \mu') F(\mu + \mu')$$

$$= \chi_0^{1/2}(\mu) \sum_{\mu' \in \widehat{L}} N(\lambda, \mu') \chi_0^{1/2}(\mu') \chi_0^{-1/2}(\mu + \mu') F(\mu + \mu')$$

$$= \chi_0^{1/2}(\mu) \sum_{\mu' \in \widehat{L}} N(\lambda, \mu') \chi_0^{1/2}(\mu') (\chi_0^{-1/2} F)(\mu + \mu')$$

$$= \chi_0^{1/2}(\mu) \widetilde{A}_\lambda (\chi_0^{-1/2} F)(\mu). \qquad \square$$

7.4 Satake Isomorphism

Consider the mapping

$$i : A_\lambda \to \widetilde{A}_\lambda, \quad \text{for } \lambda \in \widehat{L}^+.$$

Since $\{A_\lambda, \lambda \in \widehat{L}^+\}$ is a basis for the algebra $\mathscr{H}(\Delta)$, we extend this map to the whole Hecke algebra $\mathscr{H}(\Delta)$ by linearity.

Theorem 7.12 *The mapping $i : A_\lambda \to \widetilde{A}_\lambda$ is a $\mathbb{C}$-algebra isomorphism of $\mathscr{H}(\Delta)$ onto $\mathscr{H}(\mathbb{A})$.*

Proof First of all, we prove that i is a $\mathbb{C}$-algebra homomorphism from $\mathscr{H}(\Delta)$ to $\mathscr{H}(\mathbb{A})$. By definition, if $A = \sum_{j=1}^k c_j A_{\lambda_j}$, then

$$i(A) = \sum_{j=1}^k c_j i(A_{\lambda_j}) = \sum_{j=1}^k c_j \widetilde{A}_{\lambda_j}.$$

Now, for any pair $\lambda, \lambda' \in \widehat{L}^+$, we consider the operator $A_\lambda \circ A_{\lambda'}$ and prove that

$$i(A_\lambda \circ A_{\lambda'}) = i(A_\lambda) \circ i(A_{\lambda'}).$$

We know that $A_\lambda \circ A_{\lambda'}$ is a linear combination of operators $A_{\lambda_1}, \ldots, A_{\lambda_k}$, for suitable $\lambda_1, \ldots, \lambda_k$:

$$(A_\lambda \circ A_{\lambda'}) f = \sum_{j=1}^k c_j A_{\lambda_j} f.$$

Hence, $i(A_\lambda \circ A_{\lambda'}) = \tau_{h_{\lambda,\lambda'}}$, where $h_{\lambda,\lambda'}$ is the $\mathbf{W}$-invariant function on $\widehat{L}$ defined by $h_{\lambda,\lambda'} := \sum_{j=1}^k c_j h_{\lambda_j}$. This proves that $i(A_\lambda \circ A_{\lambda'})$ belongs to the algebra $\mathscr{H}(\mathbb{A})$.

Now we prove that, for every pair λ, λ',

$$i(A_\lambda \circ A_{\lambda'}) = i(A_\lambda) \circ i(A_{\lambda'}).$$

For this goal, for every character χ consider the eigenvalue $\Theta^\chi(h_{\lambda,\lambda'})$. For the sake of simplicity, set $\Theta^\chi(\lambda,\lambda') = \Theta^\chi(h_{\lambda,\lambda'})$. Since $\tau_{h_{\lambda,\lambda'}} = \sum_{j=1}^k c_j \tau_{h_{\lambda_j}}$,

$$\Theta^\chi\bigl(\lambda,\lambda'\bigr) = \sum_{j=1}^k c_j \Theta^\chi(\lambda_j).$$

Therefore, by keeping in mind (10),

$$\begin{aligned}\Theta^\chi\bigl(\lambda,\lambda'\bigr) &= \sum_{j=1}^k c_j \Lambda^{\chi\chi_0^{1/2}}(\lambda_j) = \Lambda^{\chi\chi_0^{1/2}}(A_\lambda \circ A_{\lambda'}) = \Lambda^{\chi\chi_0^{1/2}}(\lambda)\Lambda^{\chi\chi_0^{1/2}}\bigl(\lambda'\bigr)\\ &= \Theta^\chi(\lambda)\Theta^\chi\bigl(\lambda'\bigr).\end{aligned}$$

Therefore, for every χ,

$$\Theta^\chi\bigl(i(A_\lambda \circ A_{\lambda'})\bigr) = \Theta^\chi\bigl(i(A_\lambda)\bigr)\Theta^\chi\bigl(i(A_{\lambda'})\bigr) = \Theta^\chi\bigl(i(A_\lambda)\circ i(A_{\lambda'})\bigr).$$

Thus Proposition 7.2 of Sect. 7.1 implies that $i(A_\lambda \circ A_{\lambda'}) = i(A_\lambda)\circ i(A_{\lambda'})$. This proves that i is a $\mathbb{C}$-algebra homomorphism from $\mathscr{H}(\Delta)$ to $\mathscr{H}(\mathbb{A})$.

Since the operators A_λ form a $\mathbb{C}$-basis of $\mathscr{H}(\Delta)$ and, according to Proposition 7.8 of the previous section, the operators $\widetilde{A}_\lambda = i(A_\lambda)$ form a $\mathbb{C}$-basis of $\mathscr{H}(\mathbb{A})$, it follows readily that the operator i is a bijection from the algebra $\mathscr{H}(\Delta)$ onto the algebra $\mathscr{H}(\mathbb{A})$. □

The operator i is called the *Satake isomorphism* between $\mathscr{H}(\Delta)$ and $\mathscr{H}(\mathbb{A})$.

7.5 Characterization of the Eigenvalues of the Algebra $\mathscr{H}(\Delta)$

We proved in Sect. 7.1 that, for every eigenvalue Θ of the algebra $\mathscr{H}(\mathbb{A})$, there exists a character χ such that $\Theta = \Theta^\chi$. The Satake isomorphism between $\mathscr{H}(\Delta)$ and $\mathscr{H}(\mathbb{A})$ allows us to extend this characterization to the eigenvalues of the algebra $\mathscr{H}(\Delta)$.

Corollary 7.13 *For every eigenvalue Λ of the algebra $\mathscr{H}(\Delta)$ there exists a character χ on $\widehat{L}$ such that $\Lambda = \Lambda^{\chi\chi_0^{1/2}}$.*

Proof Let Λ be an eigenvalue of the algebra $\mathscr{H}(\Delta)$. By Theorem 7.12 of the previous section, there exists a unique eigenvalue $\Theta \in \mathrm{Hom}(\mathscr{H}(\mathbb{A}),\mathbb{C})$ such that

$$\Theta(\lambda) = \Lambda(\lambda), \quad \text{for every } \lambda \in \widehat{L}^+.$$

Since, by Proposition 7.4 of Sect. 7.2, there exists a character χ such that $\Theta = \Theta^\chi$, and taking in account the identity (10), we conclude that $\Lambda = \Lambda^{\chi\chi_0^{1/2}}$. □

Acknowledgements We would like to acknowledge several useful suggestions of T. Steger during the preparation of this paper.

References

1. Betori, W., Pagliacci, M.: Harmonic analysis for groups acting on trees. Boll. Unione Mat. Ital., B (6) **3**, 333–349 (1984)
2. Bourbaki, N.: Lie Groups and Lie Algebras. Chapters 4–6. Elements of Mathematics. Springer, Berlin (2002)
3. Brown, K.: Buildings. Springer, New York (1989)
4. Figà-Talamanca, A., Picardello, M.A.: Harmonic Analysis on Free Groups. Lectures Notes in Pure and Applied Math., vol. 87. Dekker, New York (1983)
5. Humphreys, J.E.: Introduction to Lie Algebras and Representation Theory. Graduate Texts in Mathematics, vol. 9. Springer, New York (1978)
6. Humphreys, J.E.: Reflection Groups and Coxeter Groups. Cambridge Studies in Advanced Mathematics, vol. 29. Cambridge University Press, Cambridge (1990)
7. Macdonald, I.G.: Spherical Functions on a Group of p-Adic Type. Publications of the Ramanujan Institute, vol. 2. Ramanujan Institute, Centre for Advanced Study in Mathematics, University of Madras, Madras (1971)
8. Mantero, A.M., Zappa, A.: Eigenfunctions of the Laplace operators for a building of type $\widetilde{A_2}$. J. Geom. Anal. **10**, 339–363 (2000)
9. Mantero, A.M., Zappa, A.: Eigenfunctions of the Laplace operators for a building of type $\widetilde{B_2}$. Boll. Unione Mat. Ital., B (8) **5**, 163–195 (2002)
10. Mantero, A.M., Zappa, A.: Eigenfunctions of the Laplace operators for a building of type $\widetilde{G_2}$. Boll. Unione Mat. Ital. (9) **2**, 483–508 (2009)
11. Mantero, A.M., Zappa, A.: Macdonald formula for spherical functions on affine buildings. Ann. Fac. Sci. Toulouse (4) **20**, 669–758 (2011)
12. Parkinson, J.: Buildings and Hecke algebras. J. Algebra **297**, 1–49 (2006)
13. Parkinson, J.: Spherical harmonic analysis on affine buildings. Math. Z. **253**, 571–606 (2006)
14. Ronan, M.: Lectures on Buildings. Perspect. in Mathematics. Academic Press, London (1989)
15. Tits, J.: Reductive groups over local fields. Automorphic forms, representations and L-functions. In: Proc. Sympos. Pure Math., vol. XXXIII, pp. 29–69. Am. Math. Soc., Providence (1979)

A Liouville Type Theorem for Carnot Groups: A Case Study

Alessandro Ottazzi and Ben Warhurst

Abstract Capogna and Cowling have shown that, if ϕ is 1-quasiconformal on an open subset of a Carnot group G, then precomposing with ϕ preserves Q-harmonic functions, where Q denotes the homogeneous dimension of G. They combine this with a regularity theorem for Q-harmonic functions to show that ϕ is in fact C^∞. As an application, they observe that for some Carnot groups of step 2 the space of 1-quasiconformal mappings forms a finite dimensional space. This is a generalization of a classical theorem of Liouville from 1850.

In this chapter we show, using the Engel group as an example, that a Liouville type theorem can be proved for every Carnot group. Indeed, the smoothness of 1-quasiconformal maps allows us to obtain a proof of the theorem using Tanaka prolongation theory.

Keywords Conformal map · Quasiconformal map · Engel group

Mathematics Subject Classification (2010) Primary 30C65 · Secondary 58D05 · 22E25 · 53C17

1 Introduction

The Liouville Theorem states that every C^4-conformal map between domains of $\mathbb{R}^3$ is the restriction of the action of some element of the group $O(1,4)$. The same result holds in $\mathbb{R}^n$ when $n > 3$ (see, e.g., Nevanlinna [5]). A major advance in the theory was the passage from smoothness assumptions to metric assumptions (Gehring [4] and Reshetnyak [12]): the conclusion of Liouville's theorem holds

A. Ottazzi (✉)
CIRM Fondazione Bruno Kessler, Via Sommarive 14, 38123 Trento, Italy
e-mail: ottazzi@fbk.eu

B. Warhurst
Dipartimento di Matematica e Applicazioni, Università di Milano Bicocca, Via Cozzi 53, 20125 Milano, Italy
e-mail: benwarhurst68@gmail.com

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3,
DOI 10.1007/978-88-470-2853-1_13,

for the 1-quasiconformal maps. When the ambient space is not Riemannian there are similar theorems. Capogna and Cowling proved in [1] that 1-quasiconformal maps defined on open subsets of a Carnot group G are smooth and they applied this to show some Liouville type results. In particular, it is now known that 1-quasiconformal maps between open subsets of H-type groups, whose Lie algebra has dimension larger than 2, form a finite dimensional space. This follows by combining the smoothness result in [1] and the work of Reimann [11], who established the corresponding infinitesimal result. Moreover, if G is a Carnot group of step two, such that the strata preserving automorphisms are all dilations, then 1-quasiconformal maps are translations composed with dilations [1].

In this chapter we combine the smoothness result in [1] with the Tanaka prolongation theory, to show that, when G is the Engel group (step three), 1-quasiconformal maps form a finite-dimensional space. A characterization of conformal maps for the Engel group can also be found in [2]. There, the Engel group is viewed as the nilradical of the group $\mathrm{Sp}(2, \mathbb{R})$, and the theory of semisimple Lie groups plays a central role. The main point of interest in our approach is that it extends to all Carnot groups. The general case together with an exposition on the Tanaka prolongation method is found in [9].

The next section is devoted to the study of 1-quasiconformal maps on the Engel group. First we define the basic formalism: we introduce the contact structure, the sub-Riemannian metric, and we give the definition of quasiconformal maps. In particular, we interpret 1-quasiconformal maps as conformal transformations. Next, we restrict the attention to the infinitesimal level introducing conformal vector fields. This leads to a system of differential equations. We then proceed by constructing a prolongation of these differential equations, formalized in terms of Tanaka prolongation. The latter is a graded Lie algebra that is isomorphic to the Lie algebra of conformal vector fields. Finally, we use a standard argument to show that any conformal map can be interpreted as the restriction of the action of some element in the automorphism group of this prolonged algebra.

2 A Case Study: The Engel Group

2.1 Notation

Let $\mathfrak{g}$ be the real Lie algebra generated by vectors X_1, X_2, Y, Z with nonzero brackets $[X_1, X_2] = Y$ and $[X_1, Y] = Z$. This is a stratified nilpotent Lie algebra of step three. Namely $\mathfrak{g} = \mathfrak{g}_{-1} + \mathfrak{g}_{-2} + \mathfrak{g}_{-3}$, where $\mathfrak{g}_{-1} = \mathrm{span}\{X_1, X_2\}$, $\mathfrak{g}_{-2} = \mathbb{R}Y$ and $\mathfrak{g}_{-3} = \mathbb{R}Z$; here the symbol $+$ denotes the direct sum of vector spaces. We write G for the connected and simply connected Lie group whose Lie algebra is $\mathfrak{g}$. We choose exponential coordinates $(x_1, x_2, y, z) = \exp(x_2X_2 + yY + zZ)\exp(x_1X_1)$. We identify the Lie algebra with the tangent space T_eG to G at the identity e, and for every X in $\mathfrak{g}$ we write $\widetilde{X}$ for the left-invariant vector field that agrees with X

at the identity e. The left-invariant vector fields corresponding to the basis vectors are

$$\widetilde{X}_1 = \frac{\partial}{\partial x_1}, \qquad \widetilde{X}_2 = \frac{\partial}{\partial x_2} + x_1 \frac{\partial}{\partial y} + \frac{x_1^2}{2}\frac{\partial}{\partial z},$$
$$\widetilde{Y} = \frac{\partial}{\partial y} + x_1 \frac{\partial}{\partial z},$$
$$\widetilde{Z} = \frac{\partial}{\partial z}.$$

At each point of $p \in G$, the span of these vector fields determines a subspace of the tangent space that we call the *horizontal space* and denote by $\tilde{\mathfrak{g}}_{-1,p}$. These subspaces vary smoothly and give rise to the horizontal tangent bundle $\tilde{\mathfrak{g}}_{-1}$. Since iterated brackets of X_1 and X_2 generate $\mathfrak{g}$, also iterated brackets of the sections $\widetilde{X}_1$ and $\widetilde{X}_2$ generate T_pG at every point p. We define an inner product $\langle\, ,\rangle$ on $\mathfrak{g}$ for which the different layers are orthogonal and we carry the inner product to the tangent space at each point using left translation. We denote by $\langle\, ,\rangle_p$ the inner product in T_pG. This allows us to define a left-invariant Carnot–Carathéodory metric d on G as follows. A smooth curve is said to be *horizontal* if its tangent vectors at every point p are in $\tilde{\mathfrak{g}}_{-1,p}$. The length of a horizontal curve is the integral of the lengths of its tangent vectors. The distance between two points is then the infimum of the lengths of the horizontal curves joining them. A nilpotent stratified Lie group with such a metric is called a Carnot group and the example we are considering is known as the Engel group.

A diffeomorphism ϕ between open subsets of G is called a *contact mapping* if its differential ϕ_* maps the horizontal space at p to the horizontal space at $\phi(p)$. Given a 1-parameter group of contact mappings, we denote by V the corresponding infinitesimal generator and call it a *contact vector field*. Contact vector fields are characterized by differential equations arising from the condition that for every horizontal vector field $\widetilde{X}$ one has $[V, \widetilde{X}] = f\widetilde{X}_1 + g\widetilde{X}_2$ for some smooth functions f and g. It is well known [6, 15] that for the Engel group the space of contact vector fields is infinite dimensional. Indeed, it is straightforward to prove that $V = f\widetilde{Z} + \widetilde{X}_1 f\widetilde{Y} + \widetilde{X}_1^2 f\widetilde{X}_2$ is a contact vector field for every smooth function $f = f(x_1)$.

Let $\mathscr{U}, \mathscr{V} \subset G$ be open domains and $\phi : \mathscr{U} \to \mathscr{V}$ a homeomorphism. For $p \in \mathscr{U}$, and for small $t \in \mathbb{R}$, we define the distortion as

$$H_\phi(p,t) = \frac{\max\{d(\phi(p),\phi(q)) \mid d(p,q) = t\}}{\min\{d(\phi(p),\phi(q)) \mid d(p,q) = t\}}.$$

We say that ϕ is *quasiconformal* if there exists a constant λ such that

$$\limsup_{t\to 0} H_\phi(p,t) \leqslant \lambda$$

for all $p \in \mathscr{U}$. Furthermore, ϕ is locally quasiconformal if it is quasiconformal in a neighborhood of each point and 1-quasiconformal when $\lambda = 1$.

In [1], the authors prove that 1-quasiconformal mappings are smooth and characterized by the conditions of being locally quasiconformal with Pansu differential $D\phi(p)$ equal to a similarity at every point p (that is, the product of a dilation and an isometry). In fact, ϕ is contact [10]. We remind the reader that

$$D\phi(p)\, q = \lim_{t \to 0} \delta_t^{-1} \circ l_{\phi(p)}^{-1} \circ \phi \circ l_p \circ \delta_t(q),$$

where $p, q \in G$, l_p denotes the left multiplication by p and for every $t \in \mathbb{R}_+$, δ_t denotes the automorphic dilation that in our case study is

$$\delta_t(x_1, x_2, y, z) = \left(e^t x_1, e^t x_2, e^{2t} y, e^{3t} z\right).$$

The Pansu differential is an automorphism of G and its Lie derivative defines an automorphism of the Lie algebra $\mathfrak{g}$ that we denote by $d\phi(p)$. The condition that $D\phi(p)$ coincides with a similarity implies that $d\phi(p)$ is a similarity when restricted to $\mathfrak{g}_{-1}$ [1, Lemma 5.2]. If we define

$$M_\phi(p) = \left(l_{\phi(p)}^{-1} \circ \phi \circ l_p\right)_{*e},$$

it turns out that $M_\phi(p)$ and $d\phi(p)$ coincide when restricted to the horizontal space. Therefore, the condition of 1-quasiconformality is equivalent to $M_\phi(p)$ being a similarity of the horizontal space at every point. More explicitly, for all horizontal vectors $X, X' \in \mathfrak{g}_{-1}$ we have

$$\begin{aligned} \left\langle M_\phi(p)^{tr} M_\phi(p) X, X'\right\rangle &= \left\langle M_\phi(p) X, M_\phi(p) X'\right\rangle \\ &= k\left\langle X, X'\right\rangle, \end{aligned}$$

whence for every $p \in \mathscr{U}$, the linear map $M_\phi(p)$ restricted to the horizontal space lies in the two dimensional conformal group

$$CO(2) = \left\{A \in GL(2, \mathbb{R}) \mid AA^{tr} = kI,\ k \in \mathbb{R}_+\right\}.$$

In view of this fact we shall also refer to 1-quasiconformal maps as *conformal* maps.

2.2 Conformal Vector Fields

Let ϕ_t be a 1-parameter group of conformal maps defined on an open set $\mathscr{U}$ and let V be the vector field whose local flow is ϕ_t. Using the basis of left-invariant vector fields, we can write $V = f_1\widetilde{X}_1 + f_2\widetilde{X}_2 + g\widetilde{Y} + h\widetilde{Z}$ where the coefficients are smooth functions. Then a direct calculation shows that

$$\frac{d}{dt}\big(M_{\phi_s}(p)\big|_{\mathfrak{g}_{-1}}\big)\big|_{t=0} = \begin{bmatrix} \widetilde{X}_1 f_1(p) & \widetilde{X}_2 f_1(p) \\ \widetilde{X}_1 f_2(p) & \widetilde{X}_2 f_2(p) \end{bmatrix}. \tag{1}$$

The contact conditions for V are of the form

$$[V, \widetilde{X}_1] = a\widetilde{X}_1 + b\widetilde{X}_2,$$
$$[V, \widetilde{X}_2] = c\widetilde{X}_1 + d\widetilde{X}_2,$$

for some functions a, b, c, d. They yield the following system of differential equations:

$$\begin{cases} \widetilde{X}_1 g = -f_2, \\ \widetilde{X}_1 h = -g, \end{cases} \qquad \begin{cases} \widetilde{X}_2 g = f_1, \\ \widetilde{X}_2 h = 0, \end{cases} \qquad \widetilde{Y} h = f_1. \tag{2}$$

Furthermore, since $M_{\phi_s}(p)|_{\mathfrak{g}_{-1}} \in CO(2)$, it follows that $\frac{d}{dt}(M_{\phi_s}(p)|_{\mathfrak{g}_{-1}})|_{t=0}$ is in $\mathfrak{co}(2) = \{A \in \mathfrak{gl}(2, \mathbb{R}) \mid A + A^{tr} = kI\}$. Therefore, by (1),

$$\begin{cases} \widetilde{X}_1 f_1 = \widetilde{X}_2 f_2, \\ \widetilde{X}_2 f_1 = -\widetilde{X}_1 f_2. \end{cases} \tag{3}$$

A vector field that satisfies (2) and (3) on its domain of definition is said to be a *conformal vector field*.

2.3 Prolongation of the Differential Equations

In order to gather information on the space of functions that solve (2) and (3), we consider higher order derivatives: roughly speaking, if the derivatives in all directions of the coefficients of V vanish at a certain order, we may conclude that V has polynomial coefficients and therefore varies in a finite dimensional space. This idea was formalized by the prolongation of Singer and Sternberg [13] in their study of infinitesimal automorphisms of G-structures and later refined by Tanaka [14]. For the study of contact mappings, Tanaka prolongation theory was used in [17] and more recently by the authors of this chapter in different collaborations [3, 7–9, 16].

Instead of giving the general argument of Tanaka prolongation theory, we choose here to illustrate the method in our case study. In order to do that, we first fix the following notation. For every $p \in \mathscr{U}$ we define

$$A_V^{-1}(p) = \big(f_1(p), f_2(p), 0, 0\big),$$
$$A_V^{-2}(p) = \big(0, 0, g(p), 0\big),$$
$$A_V^{-3}(p) = \big(0, 0, 0, h(p)\big),$$

where we interpret the vectors on the right hand side as elements of $\mathfrak{g}$ according to the fixed basis. It follows that the equations in (2) are equivalent to

$$\big[A_V^j(p), X\big] = \widetilde{X}\big(A_V^{j-1}\big)(p),$$

where $X \in \mathfrak{g}_j$ and $j = -1, -2$.

Next, for every $p \in G$, we define a strata preserving linear map $A_V^0(p) : \mathfrak{g} \to \mathfrak{g}$ by setting

$$A_V^0(p)(X) = \widetilde{X}\big(A_V^j\big)(p),$$

where $X \in \mathfrak{g}_j$, $j = -1, -2, -3$. Namely

$$\begin{aligned}
A_V^0(p)(X_1) &= \widetilde{X}_1\big(A_V^{-1}\big)(p) = \big(\widetilde{X}_1 f_1(p), \widetilde{X}_1 f_2(p), 0, 0\big),\\
A_V^0(p)(X_2) &= \widetilde{X}_2\big(A_V^{-1}\big)(p) = \big(\widetilde{X}_2 f_1(p), \widetilde{X}_2 f_2(p), 0, 0\big),\\
A_V^0(p)(Y) &= \widetilde{Y}\big(A_V^{-2}\big)(p) = \big(0, 0, \widetilde{Y} g(p), 0\big),\\
A_V^0(p)(Z) &= \widetilde{Z}\big(A_V^{-3}\big)(p) = \big(0, 0, 0, \widetilde{Z} h(p)\big),
\end{aligned}$$

or equivalently

$$A_V^0(p) = \begin{bmatrix} \widetilde{X}_1 f_1(p) & \widetilde{X}_2 f_1(p) & 0 & 0 \\ \widetilde{X}_1 f_2(p) & \widetilde{X}_2 f_2(p) & 0 & 0 \\ 0 & 0 & \widetilde{Y} g(p) & 0 \\ 0 & 0 & 0 & \widetilde{Z} h(p) \end{bmatrix}. \tag{4}$$

If we set $[A_V^0(p), X] := A_V^0(p)(X)$, it follows by direct computation that the Jacobi identity

$$\big[A_V^0(p), [S, T]\big] = \big[A_V^0(p)(S), T\big] - \big[A_V^0(p)(T), S\big]$$

holds for every $S \in \mathfrak{g}_s$ and $T \in \mathfrak{g}_l$, with $s, l = -3, -2, -1$. This implies that $A_V^0(p) \in \mathrm{Der}_0(\mathfrak{g})$, the space of strata preserving derivations of $\mathfrak{g}$. Notice that $A_V^0(p)$ coincides with $\frac{d}{dt}(M_{\phi_s}(p)|_{\mathfrak{g}_{-1}})|_{t=0}$ when restricted to the horizontal space, so that conditions (3) can be expressed in terms of $A_V^0(p)$. This implies that $A_V^0(p)$ must lie in

$$\mathfrak{g}_0 = \big\{D \in \mathrm{Der}_0(\mathfrak{g}) \mid D|_{\mathfrak{g}_{-1}} \in \mathfrak{co}(2)\big\}.$$

The fact that $A_V^0(p) \in \mathfrak{g}_0$ imposes conditions on higher order derivatives of the coefficients of V. Indeed, for each $p \in G$, we define the linear map

$$A_V^1(p) : \mathfrak{g} \to \mathfrak{g} + \mathfrak{g}_0$$

by setting $A_V^1(p)(X) = \widetilde{X}(A_V^{j+1})(p)$ for every $X \in \mathfrak{g}_j$.

In particular,

$$A_V^1(p)(X_1) = \begin{bmatrix} \widetilde{X}_1^2 f_1(p) & \widetilde{X}_1\widetilde{X}_2 f_1(p) & 0 & 0 \\ \widetilde{X}_1^2 f_2(p) & \widetilde{X}_1\widetilde{X}_2 f_2(p) & 0 & 0 \\ 0 & 0 & \widetilde{X}_1\widetilde{Y} g(p) & 0 \\ 0 & 0 & 0 & \widetilde{X}_1\widetilde{Z} h(p) \end{bmatrix},$$

$$A_V^1(p)(X_2) = \begin{bmatrix} \widetilde{X}_2\widetilde{X}_1 f_1(p) & \widetilde{X}_2^2 f_1(p) & 0 & 0 \\ \widetilde{X}_2\widetilde{X}_1 f_2(p) & \widetilde{X}_2^2 f_2(p) & 0 & 0 \\ 0 & 0 & \widetilde{X}_2\widetilde{Y} g(p) & 0 \\ 0 & 0 & 0 & \widetilde{X}_2\widetilde{Z} h(p) \end{bmatrix}, \tag{5}$$

$$A_V^1(p)(Y) = \big(\widetilde{Y} f_1(p), \widetilde{Y} f_2(p), 0, 0\big),$$

$$A_V^1(p)(Z) = \big(0, 0, \widetilde{Z} g(p), 0\big).$$

Writing $[A_V^1(p), X] := A_V^1(p)(X)$ we obtain the Jacobi identity:

$$\big[A_V^1(p), [S, T]\big] = \big[A_V^1(p)(S), T\big] - \big[A_V^1(p)(T), S\big],$$

for every $S \in \mathfrak{g}_s$ and $T \in \mathfrak{g}_l$, with $s, l = -3, -2, -1$. This implies that $A_V^1(p)$ lies in the space

$$\mathfrak{g}_1 := \big\{u : \mathfrak{g} \to \mathfrak{g} + \mathfrak{g}_0 \,\big|\, u(\mathfrak{g}_j) \subset \mathfrak{g}_{j+1}, \\ u[S, T] = \big[u(S), T\big] - \big[u(T), S\big],\ \forall S \in \mathfrak{g}_s,\ \forall T \in \mathfrak{g}_l,\ s, l = -3, -2, -1\big\}.$$

In general, we continue this procedure inductively and introduce the linear maps

$$A_V^i(p) : \mathfrak{g} \to \mathfrak{g} + \mathfrak{g}_0 + \mathfrak{g}_1 + \cdots + \mathfrak{g}_{i-1},$$

that can vary in the space

$$\mathfrak{g}_i := \big\{u : \mathfrak{g} \to \mathfrak{g} + \mathfrak{g}_0 + \mathfrak{g}_1 + \cdots + \mathfrak{g}_{i-1} \,\big|\, u(\mathfrak{g}_j) \subset \mathfrak{g}_{j+i}, \\ u[S, T] = \big[u(S), T\big] - \big[u(T), S\big],\ \forall S \in \mathfrak{g}_s,\ \forall T \in \mathfrak{g}_l,\ s, l = -3, -2, -1\big\}.$$

We say that $\mathfrak{g}_i$ is the ith *prolongation space* of $\mathfrak{g}$ through $\mathfrak{g}_0$. The fact that $A_V^i(p) \in \mathfrak{g}_i$ provides information on the $i+1$ order derivatives of f_1 and f_2, the $i+2$ order derivatives of g, and $i+3$ order derivatives of h. If the process is finite, then one ends up with a graded Lie algebra $\mathfrak{g} + \mathfrak{g}_0 + \sum_{k \geqslant 1} \mathfrak{g}_k$, where $[u, X] := u(X)$, for every $X \in \mathfrak{g}$ and $u \in \sum_{j \geqslant 0} \mathfrak{g}_k$. We shall see that this algebra is isomorphic to the space of vector fields.

2.4 The Lie Algebra of Conformal Vector Fields and the Group of Conformal Maps

We proceed by computing the prolongation spaces. First, we compute $\mathfrak{g}_0$. Write elements in $\mathrm{Der}_0\mathfrak{g}$ as $D = d_{ij} \in \mathfrak{gl}(4, \mathbb{R})$. Since D is a strata preserving derivation it follows that

$$DY = D[X_1, X_2] = \left[\sum_{i=1}^{2} d_{i1} X_i, X_2\right] + \left[X_1, \sum_{j=1}^{2} d_{j2} X_j\right] = (d_{11} + d_{22})Y,$$

$$DZ = D[X_1, Y] = \left[\sum_{i=1}^{2} d_{i1} X_i, Y\right] + \left[X_1, (d_{11} + d_{22})Y\right] = (2d_{11} + d_{22})Z.$$

Since $[X_2, Y] = 0$,

$$0 = D[X_2, Y] = \left[\sum_{j=1}^{2} d_{j2} X_j, Y\right] = d_{12} Z,$$

whence $d_{12} = 0$. The condition $D|_{\mathfrak{g}_{-1}} \in \mathfrak{co}(2)$ implies $d_{11} = d_{22}$ and $d_{21} = -d_{12} = 0$. In conclusion, $\mathfrak{g}_0 = \mathbb{R}D$, where $D = \mathrm{diag}\{1, 1, 2, 3\}$.

The calculation of $\mathfrak{g}_1$ goes as follows. If $u \in \mathfrak{g}_1$, then we set $u(X_1) = aD$ and $u(X_2) = bD$ with $a, b \in \mathbb{R}$. By the Jacobi identity we obtain $u(Y) = aX_2 - bX_1$ and $u(Z) = 3aY$. Therefore

$$0 = u[X_1, Z] = [aD, Z] - [3aY, X_1] = 6aZ,$$

whence $a = 0$, and

$$0 = u[X_2, Z] = [bD, Z] - 0 = 3bZ,$$

whence $b = 0$. Thus $u = 0$ and so $\mathfrak{g}_1 = \{0\}$.

The contact equations (2), the formula (4) and the fact that $A_V^0(p) \in \mathfrak{g}_0$, lead to the differential equations

$$\begin{aligned} \widetilde{X}_1 f_1 &= \widetilde{X}_2 f_2, \qquad \widetilde{X}_2 f_1 = -\widetilde{X}_1 f_2 = 0, \\ \widetilde{Y} g &= 2\widetilde{X}_1 f_1, \\ \widetilde{Z} h &= 3\widetilde{X}_1 f_1. \end{aligned} \tag{6}$$

Since $A_V^1(p) \in \mathfrak{g}_1 = \{0\}$, (5) and (6) yield

$$\widetilde{X}_1^2 f_1 = 0, \qquad \widetilde{X}_2^2 f_2 = 0.$$

The differential equations above and the contact equations lead to a system of differential equations for all the coefficients of V. Moreover, (2) implies that the coefficients of V are determined by h. So we can restrict our attention to the equations involving h:

$$\widetilde{X}_1^3 h = 0, \qquad \widetilde{X}_2 h = 0, \qquad \widetilde{Y}^2 h = 0, \qquad \widetilde{Z}^2 h = 0.$$

It is easy to verify that h then varies in a space of polynomials of dimension 5, so that the space of conformal vector fields, say $\mathscr{C}(G)$, has dimension 5. In fact, the Tanaka prolongation of $\mathfrak{g}$ through $\mathfrak{g}_0$, namely $\mathfrak{s} = \mathfrak{g} + \mathfrak{g}_0$, is isomorphic as Lie algebra to $\mathscr{C}(G)$. The isomorphism $\tau : \mathfrak{s} \to \mathscr{C}(G)$ is

$$\tau(X) f(p) = \frac{d}{dt} f(\exp tX \cdot p)|_{t=0},$$

where $\cdot$ denotes the action of $\exp \mathfrak{s}$ on $\mathfrak{g}$. More precisely, the symbol $\cdot$ represents the product on G if $X \in \mathfrak{g}$ and the action of the automorphism $\exp tX$ if $X \in \mathfrak{g}_0$. By abuse of notation, a basis of $\mathfrak{s}$ is $\{X_1, X_2, Y, Z, D\}$. A direct calculation shows that

$$\begin{aligned}
\tau(D) &= \left(3z - 2x_1 y + \frac{x_1^2 x_2}{2}\right)\widetilde{Z} + (2y - x_1 x_2)\widetilde{Y} + x_1 \widetilde{X}_1 + x_2 \widetilde{X}_2, \\
\tau(X_1) &= (y - x_1 x_2)\widetilde{Z} + x_2 \widetilde{Y} + \widetilde{X}_1, \\
\tau(X_2) &= \frac{x_1^2}{2}\widetilde{Z} - x_1 \widetilde{Y} + \widetilde{X}_2, \\
\tau(Y) &= -x_1 \widetilde{Z} + \widetilde{Y}, \\
\tau(Z) &= \widetilde{Z}.
\end{aligned}$$

We showed that the space of vector fields whose local flow is given by conformal mappings is finite dimensional and it coincides with $\mathfrak{s}$. Now we prove that, if the map $\phi : \mathscr{U} \to \mathscr{V}$ is conformal, then ϕ is the restriction to $\mathscr{U}$ of the action of some element in $\mathrm{Aut}(\mathfrak{s})$, the automorphism group of $\mathfrak{s}$. The conclusion will be that the space of conformal maps is contained in $\mathrm{Aut}(\mathfrak{s})$ and contains $\exp \mathfrak{s}$. By composing with left translations, we may assume that $e \in \mathscr{U} \cap \mathscr{V}$ and it is enough to show that any conformal map which preserves the identity is such a restriction.

We show that ϕ induces an automorphism of $\mathfrak{s}$. If $V \in \tau(\mathfrak{s})$ and ψ_t is the corresponding flow, then $\phi_* V$ is the infinitesimal generator of the 1-parameter group of conformal maps $\phi \psi_t \phi^{-1}$. Therefore $\phi_* V$ is conformal, whence $\phi_* V \in \tau(\mathfrak{s})$, and $\tau^{-1}\phi_* \tau \in \mathrm{Aut}(\mathfrak{s})$. Since the domain of ϕ is connected and contains e, it follows that ϕ_* determines the map ϕ uniquely. In fact, consider a conformal map φ such that $\varphi(e) = e$ and $\phi_* V = \varphi_* V$. Then we have $\psi_t \circ \varphi^{-1} \circ \phi = \varphi^{-1} \circ \phi \circ \psi_t$. In particular, if V is the infinitesimal generator of right translations, we conclude that $\varphi^{-1} \circ \phi$ commutes with the right translations and therefore it is a left translation, whence $\varphi^{-1} \circ \phi$ is the identity.

It is worth noticing that not all automorphisms of $\mathfrak{s}$ define conformal maps. For example, let α be the automorphism of $\mathfrak{s}$ defined by $\alpha(X_1) = X_1$ and $\alpha(X_2) = 2X_2$. This automorphism cannot arise as $\tau^{-1}\phi_* \tau$ for some conformal map ϕ.

References

1. Capogna, L., Cowling, M.G.: Conformality and Q-harmonicity in Carnot groups. Duke Math. J. **135**, 455–479 (2006)
2. Cowling, M.G., De Mari, F., Korányi, A., Reimann, H.M.: Contact and conformal maps on Iwasawa *N* groups. Rend. Mat. Accad. Lincei (9) **13**, 219–232 (2002)
3. De Mari, F., Ottazzi, A.: Rigidity of Carnot groups relative to multicontact structures. Proc. Am. Math. Soc. **138**, 1889–1895 (2010)
4. Gehring, F.W.: Rings and quasiconformal mappings in space. Trans. Am. Math. Soc. **103**, 353–393 (1962)
5. Nevanlinna, R.: On differentiable mappings. In: Analytic Functions, pp. 3–9. Princeton Univ. Press, Princeton (1960)
6. Ottazzi, A.: A sufficient condition for nonrigidity of Carnot groups. Math. Z. **259**, 617–629 (2008)
7. Ottazzi, A., Warhurst, B.: Rigidity of Iwasawa nilpotent Lie groups via Tanaka's theory. Note Mat. **30**, 141–146 (2010)
8. Ottazzi, A., Warhurst, B.: Algebraic prolongation, rigidity of Carnot groups. Monatshefte Math. **162**, 179–195 (2011)
9. Ottazzi, A., Warhurst, B.: Contact and 1-quasiconformal maps on Carnot groups. J. Lie Theory **21**, 787–811 (2011)
10. Pansu, P.: Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un. Ann. Math. (2) **129**, 1–60 (1989)
11. Reimann, H.M.: Rigidity of H-type groups. Math. Z. **237**, 697–725 (2001)
12. Rešetnjak, Yu.G.: Liouville's conformal mapping theorem under minimal regularity hypotheses. Sib. Mat. Zh. **8**, 835–840 (1967) (in Russian). English translation in: Sib. Math. J. **8**, 631–634 (1967)
13. Singer, I.M., Sternberg, S.: The infinite groups of Lie and Cartan. I. The transitive groups. J. Anal. Math. **15**, 1–114 (1965)
14. Tanaka, N.: On differential systems, graded Lie algebras and pseudogroups. J. Math. Kyoto Univ. **10**, 1–82 (1970)
15. Warhurst, B.: Jet spaces as nonrigid Carnot groups. J. Lie Theory **15**, 341–356 (2005)
16. Warhurst, B.: Tanaka prolongation of free Lie algebras. Geom. Dedic. **130**, 59–69 (2007)
17. Yamaguchi, K.: Differential systems associated with simple graded Lie algebras. In: Progress in Differential Geometry. Adv. Studies Pure Math., vol. 22, pp. 413–494. Math. Soc. Japan, Tokyo (1995)

Stochastic Properties of Riemannian Manifolds and Applications to PDE's

Gregorio Pacelli Bessa, Stefano Pigola, and Alberto G. Setti

Abstract The aim of this note is to describe geometric conditions under which a Riemannian manifold enjoys the Feller property and to show how the validity of the Feller property in combination with stochastic completeness provides a new viewpoint to study qualitative properties of solutions of semilinear elliptic PDE's defined outside a compact set.

Keywords Feller property · Stochastic completeness · Comparison results

Mathematics Subject Classification (2010) Primary 58J05 · 31B35 · Secondary 58J65

1 Introduction

The asymptotic behavior of the heat kernel of a Riemannian manifold gives rise to the classical concepts of parabolicity, stochastic completeness (or conservative property) and Feller property (or C^0-diffusion property). Both parabolicity and stochastic completeness have been subject to a systematic study which led to the discovery not only of sharp geometric conditions for their validity but also of an incredibly rich family of tools, techniques and equivalent concepts ranging from maximum principles at infinity, function theoretic tests (Khas'minskii criterion), comparison techniques and so on. The purpose of this note is twofold. First we describe geometric conditions that ensure that a manifold enjoys the Feller property, for short,

G. Pacelli Bessa
Departamento de Matemática, Universidade Federal do Ceará-UFC, 60455-760 Fortaleza, CE, Brazil
e-mail: bessa@mat.ufc.br

S. Pigola · A.G. Setti (✉)
Sezione di Matematica, Dipartimento di Scienza e Alta Tecnologia, Università dell'Insubria, via Valleggio 11, 22100, Como, Italy
e-mail: alberto.setti@uninsubria.it

S. Pigola
e-mail: stefano.pigola@uninsubria.it

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3,
DOI 10.1007/978-88-470-2853-1_14,

is Feller. We will see that, although there are similarities with the case of stochastic completeness, their situation is indeed quite different.

Our second goal is to describe the consequences of the Feller property on the behavior of solutions of PDE's involving the Laplacian. It is well understood that stochastic properties of a Riemannian manifold $(M, \langle\, ,\rangle)$, like parabolicity and stochastic completeness have important counterparts on the behavior of solutions of PDE's defined on the manifold. Indeed, M is parabolic, respectively stochastically complete, if every subharmonic function bounded from above is constant, respectively if every non-negative bounded solution of the differential inequality $\Delta u \geqslant \lambda u$ for $\lambda > 0$ is constant, and therefore vanishes identically. It is apparent by the very definition of these stochastic properties that global solutions must be considered. The introduction of the Feller property, that is the property that the heat semigroup maps the space of continuous functions vanishing at infinity into itself, in combination with stochastic completeness, will enable us to get important information even in the case of solutions at infinity.

In fact, using a suitable comparison theory, we are going to show that manifolds which are both stochastically complete and Feller do represent a natural environment where solutions of PDE's at infinity can be studied.

Sections 2–4 contain foundational material and the results recently obtained in [20]. In Sect. 5 we present application of the Feller property to geometry and PDE's taken from [3].

2 Stochastic Completeness vs. the Feller Property

In what follows, $(M, \langle\, ,\rangle)$, often abbreviated by M, denotes a connected complete Riemannian manifold of dimension m, and $d\,\mathrm{vol}$, ∇, Δ are the Riemannian measure, the gradient and the Laplace operator of M. We denote by $B(x, r)$ and $\partial B(x, r)$ the geodesic ball of radius r centered at x and its boundary. Let g_{ij} be the components of the metric $\langle\, ,\rangle$ in local coordinates x^i, g^{ij} the components of the inverse matrix, and $g = \det g_{ij}$. Recall that

$$d\,\mathrm{vol} = \sqrt{g}\,dx, \qquad \nabla f = g^{ij}\frac{\partial f}{\partial x^i}\frac{\partial}{\partial x^j}, \qquad \Delta f = \frac{1}{\sqrt{g}}\frac{\partial}{\partial x^i}\left(\sqrt{g}g^{ij}\frac{\partial f}{\partial x^j}\right).$$

The heat kernel $p_t(x, y)$ of M is the minimal positive solution of the problem

$$\begin{cases} \Delta p_t = \frac{\partial p_t}{\partial t}, \\ p_{0+}(x, y) = \delta_y(x), \end{cases} \tag{1}$$

and can be obtained as limit of the Dirichlet heat kernels $p_t^{\Omega_n}(x, y)$ of any smooth, relatively compact exhaustion $\Omega_n \nearrow M$ (see details in [10]). We recall the following properties:

(i) $p_t(x, y) > 0$ is a symmetric function of x and y.
(ii) $\int_M p_t(x, z)p_s(z, y)\,d\,\mathrm{vol}(z) = p_{t+s}(x, y)$ for every $t, s > 0$ and $x, y \in M$.

(iii) $\int_M p_t(x,y)\,d\,\mathrm{vol}(y) \leqslant 1$, for every $t > 0$ and $x \in M$.
(iv) For every bounded continuous function u on M, if we set

$$P_t u(x) = \int_M p_t(x,y)u(y)\,d\,\mathrm{vol}(y),$$

then $P_t u(x)$ satisfies the heat equation on $M \times (0,+\infty)$. Moreover, by (ii) and (iii), P_t extends to a contraction semigroup on every L^p, called the heat semigroup of M.

From the probabilistic viewpoint, the heat kernel $p_t(x,y)$ represents the transition probability density of the Brownian motion $t \to X_t$ of M. In this respect, property (iii) stated above means that $t \to X_t$ is, in general, sub-Markovian.

Definition 2.1 We say that M is stochastically complete if heat is conserved, i.e., if for all $t > 0$ and some (and therefore all) $x \in M$

$$\int_M p_t(x,y)\,d\,\mathrm{vol}(y) = 1.$$

Stochastic completeness has a number of equivalent formulations. For instance,

- solutions of the heat equation with bounded initial data are unique;
- for some (and therefore all) $\lambda > 0$, bounded nonnegative solutions on M of the differential inequality $\Delta u \geqslant \lambda u$, vanish identically. See [13].

For the purposes of this note the most useful equivalent formulation is in terms of the weak maximum principle at infinity:

Definition 2.2 We say that the weak maximum principle at infinity holds on M if, for every $u \in C^2(M)$ with $\sup_M u = u^* < +\infty$, there exists a sequence $\{x_k\}$ along which

$$\text{(i)} \quad u(x_k) > u^\star - \frac{1}{k}, \qquad \text{(ii)} \quad \Delta u(x_k) < \frac{1}{k}.$$

It was proved in [17] (see also [18]) that

- M is stochastically complete if and only if the weak maximum principle at infinity holds on M.

The geometric conditions which imply stochastic completeness are subsumed either by a lower bound on the Ricci curvature Ricc of the underlying manifold or by an upper bound on the volume growth of geodesic balls.

Theorem 2.3 *Let M be a complete Riemannian manifold and $r(x) = \mathrm{dist}(o,x)$ denote the geodesic distance function from a reference point o. Then M is stochastically complete provided one of the following conditions hold*:

(i) *the Ricci curvature satisfies* $\mathrm{Ricc}(x) \geqslant -G^2(r(x))$, *where* G *is a positive, continuous increasing function satisfying* $\int^{+\infty} \frac{1}{G(r)}\,dr = +\infty$ [14, 22];

(ii) $\int^{+\infty} \frac{r}{\log(\mathrm{vol}\, B(o,r))}\,dr = +\infty$ [11].

These two conditions are essentially sharp, and although Ricci curvature lower bounds imply volume upper bounds, by the Bishop–Gromov volume comparison theorem, the conditions are related but independent. Note that heuristically, the obstruction to stochastic completeness is the fact that the manifold grows too fast at infinity.

We recall for comparison that M is parabolic if positive superharmonic functions are necessarily constant. This is equivalent to the non-existence of a positive minimal Green's kernel, and to the recurrence of Brownian motion. We also recall that a geodesically complete manifold is parabolic provided it has at most quadratic volume growth. Indeed, a sufficient condition for parabolicity is that

$$\int^{+\infty} \frac{1}{\mathrm{vol}\, \partial B(o,r)}\,dr = +\infty.$$

Let us now turn to the Feller condition.

Definition 2.4 We say that M satisfies the Feller condition, (for short, that M is Feller), if the heat semigroup P_t maps $C_0(M)$ into itself, that is, if

$$P_t u(x) = \int_M p_t(x,y)u(y)\,d\,\mathrm{vol}(y) \to 0, \quad \text{as } x \to +\infty \tag{2}$$

for every $u \in C_0(M) = \{u\colon M \to \mathbb{R} \text{ continuous}\colon u(x) \to 0 \text{ as } x \to \infty\}$.

Since $p_t(x,\cdot)$ is uniformly integrable, using a cut-off argument, one can easily prove the following

Lemma 2.5 *Assume that* M *is geodesically complete. Then* M *is Feller if and only if it satisfies one of the following equivalent conditions*:

(i) *the limit in* (2) *holds for every non-negative function* $u \in C_c(M)$;

(ii) *for some* (*and therefore all*) $p \in M$ *and for every* $R > 0$,

$$\int_{B(p,R)} p_t(x,y)\,d\,\mathrm{vol}(y) \to 0 \quad \text{as } x \to +\infty.$$

According to R. Azencott [1], the Feller property can be characterized in terms of asymptotic properties of solutions of exterior boundary value problems. We write

$$\Omega \Subset M \quad \text{if } \overline{\Omega} \text{ is compact and contained in } M.$$

Given a smooth open set $\Omega \Subset M$ and $\lambda > 0$, the problem

$$\begin{cases} \Delta h = \lambda h & \text{on } M \setminus \overline{\Omega}, \\ h = 1 & \text{on } \partial\Omega, \\ h > 0 & \text{on } M \setminus \Omega \end{cases} \tag{3}$$

has a (unique) minimal smooth solution $h : M \setminus \overline{\Omega} \to \mathbb{R}$. By the maximum principle $0 < h \leqslant 1$ and h is obtained as the limit $h(x) = \lim_{n\to+\infty} h_n(x)$, where Ω_n is a smooth exhaustion of M and h_n solves $\Delta h_n = \lambda h_n$ on $\Omega_n \setminus \overline{\Omega}$ and has boundary values $h_n = 1$ on $\partial\Omega$ and $h = 0$ on $\partial\Omega_n$.

Theorem 2.6 [1] *M is Feller if and only if for some (hence any) open set $\Omega \Subset M$ with smooth boundary and for some (hence any) constant $\lambda > 0$, the minimal solution $h : M \setminus \Omega \to \mathbb{R}$ of problem (3) satisfies*

$$h(x) \to 0 \quad \text{as } x \to \infty. \tag{4}$$

After the pioneering work of Azencott, the investigation has focused on finding optimal geometric conditions ensuring that a manifold is Feller [8, 10, 14–16, 23], and with the only exception of [8], the geometric conditions are always expressed in terms of Ricci curvature lower bounds. The methods range from estimates of solutions of parabolic equations [10, 16, 23] to estimates of the probability that the Brownian motion on M be found in certain regions before a fixed time [14]. The best known result in this direction is due to E. Hsu [14]. It uses a probabilistic approach and relies on a result by Azencott (see also [15]) according to which *M is Feller if and only if, for every compact set K and for every $t_0 > 0$, the probability that Brownian motion X_t issuing from x_0 enters K before the time t_0 tends to zero as $x_0 \to \infty$.*

Theorem 2.7 [14] *Let M be a complete, non compact Riemannian manifold of dimension* $\dim M = m$. *Assume that*

$$\operatorname{Ricc} \geqslant -(m-1)G^2\big(r(x)\big), \tag{5}$$

where $r(x) = \operatorname{dist}(x, o)$ is the distance function from a fixed reference point $o \in M$ and G is a positive, increasing function on $[0, +\infty)$ satisfying

$$\frac{1}{G} \notin L^1(+\infty). \tag{6}$$

Then M is Feller.

It is remarkable that, to the best to our knowledge, there is no analytic proof of this result. Note also that (6) is precisely the condition on the Ricci curvature that ensures the stochastic completeness of M. So one may be led to believe that as in the case of stochastic completeness "big volumes" are an obstruction to the Feller property. In fact this is not the case: in some sense, the obstruction is given by "small volumes". Indeed we have the following:

Theorem 2.8 [1] *If M is a Cartan–Hadamard manifold (complete, simply connected with nonpositive sectional curvature), then M is Feller.*

3 Model Manifolds and Comparison Results

Model manifolds also shed light on the relationship between the Feller property and the geometry of the manifold. Recall that a model manifold M_f^m is $\mathbb{R}^m = [0, \infty) \times \mathbb{S}^{m-1}$ with the metric given in polar coordinates by $\langle\, , \rangle = dr^2 + f(r)^2 d\theta^2$, where f is odd and $f'(0) = 1$. For instance if $f(r) = r$ then $M^m = \mathbb{R}^m$, if $f(r) = \sin r$ then $M^m = \mathbb{S}^m$ and if $f(r) = \sinh r$ then $M^m = \mathbb{H}^m$.

The following result holds.

Theorem 3.1 [1, 20] *An m-dimensional model manifold M_f^m with warping function f is Feller if and only if either*

$$\frac{1}{f^{m-1}(r)} \in L^1(+\infty) \tag{7}$$

or

$$\text{(i)} \quad \frac{1}{f^{m-1}(r)} \notin L^1(+\infty) \quad \textit{and} \quad \text{(ii)} \quad \frac{\int_r^{+\infty} f^{m-1}(t)\,dt}{f^{m-1}(r)} \notin L^1(+\infty). \tag{8}$$

In (8), condition (ii) is considered automatically satisfied if $f^{m-1} \notin L^1(+\infty)$.

Proof The proof is of elliptic nature. One easily observes that, on a model manifold, the minimal solution of the problem

$$\begin{cases} \Delta h = \lambda h & \text{on } M_f^m \setminus \overline{B(0,1)}, \\ h = 1 & \text{on } \partial B(0,1), \\ h > 0 & \text{on } M_f^m \setminus B(0,1) \end{cases}$$

is necessarily radial.

Next, one shows that the minimal solution $h(r)$ of the radialized 1-dimensional problem

$$\begin{cases} \left(f^{m-1} h'\right)' = \lambda f^{m-1} h & \text{on } (1, +\infty), \\ h(1) = 1 \end{cases}$$

tends to zero as $r \to +\infty$ if and only either condition (7) or condition (8) holds. □

Since $\operatorname{vol} \partial B(0,r) = c_m f^{m-1}(r)$, conditions (7) and (8) can be restated in more geometrical terms by saying that M is Feller if either

$$\frac{1}{\operatorname{vol}(\partial B_r)} \in L^1(+\infty) \tag{9}$$

or

$$\text{(i)} \quad \frac{1}{\mathrm{vol}(\partial B_r)} \notin L^1(+\infty) \quad \text{and} \quad \text{(ii)} \quad \frac{\mathrm{vol}(M_g)}{\mathrm{vol}(\partial B_r)} - \frac{\mathrm{vol}(B_r)}{\mathrm{vol}(\partial B_r)} \notin L^1(+\infty). \quad (10)$$

In particular, a model manifold with infinite volume is always Feller.

Note also that $1/f^{m-1}(r) \in L^1(+\infty)$ is the necessary and sufficient condition for a model manifold M_f to be non-parabolic. Indeed,

$$G(x,0) := \int_{r(x)}^{+\infty} \frac{dt}{f^{m-1}(t)}$$

is the Green kernel with pole at 0 of the Laplace–Beltrami operator of M_f^m. Since parabolicity implies stochastic completeness, stochastically incomplete models are always Feller.

For comparison, it may be interesting to notice that a model manifold M_f^m is stochastically complete if

$$\int^{+\infty} \frac{\int_0^r f^{m-1}(s)\,ds}{f^{m-1}(r)}\, dr = +\infty\,,$$

that is if and only if

$$r \to \frac{\mathrm{vol}\, B(0,r)}{\mathrm{vol}\, \partial B(0,r)} \notin L^1(+\infty).$$

Indeed, the function

$$u(x) = \int_0^{r(x)} \frac{\int_0^r f^{m-1}(s)\,ds}{f^{m-1}(r)}\, dr$$

satisfies $\Delta u = 1$. Therefore, if it is bounded, it violates the weak maximum principle at infinity and M_f^m is not stochastically complete. The other implication follows from a comparison argument.

However, neither parabolicity nor, a fortiori, stochastic completeness imply the Feller property. Indeed, fix $\beta > 2$ and $\alpha > 0$, and let $f(t) : \mathbb{R} \to \mathbb{R}$ be any smooth, positive, odd function satisfying $f'(0) = 1$ and $f(r) = \exp(-\alpha r^\beta)$ for $r \geqslant 10$. Then,

$$\frac{1}{f(r)} = \exp\bigl(\alpha r^\beta\bigr) \notin L^1(+\infty) \quad \text{and} \quad \frac{\int_r^{+\infty} f(t)\,dt}{f(r)} \asymp r^{1-\beta} \in L^1(+\infty)$$

and the 2-dimensional model M_f^2 is parabolic and therefore stochastically complete, but it is not Feller. Actually we shall see in Sect. 4 below that, using a gluing technique, one can construct Feller manifolds which are neither parabolic nor stochastically complete.

Parabolicity and stochastic completeness of a general manifold can be deduced from those of a model manifold via curvature comparisons (see, e.g., [13]). We are going to describe how this technique may be extended to the Feller property.

To this goal, recall that the minimal solution h of the exterior problem (3) is the limit of solutions h_n which vanish on the boundary $\partial\Omega_n$ of an exhaustion of M. Therefore standard comparison results show that, if u is a supersolution of (3), that is

$$\begin{cases} \Delta u \leqslant \lambda u & \text{on } M \setminus \overline{\Omega}, \\ u \geqslant 1 & \text{on } \partial\Omega, \end{cases}$$

then

$$h \leqslant u, \quad \text{on } M \setminus \Omega.$$

In particular, if $u(x) \to 0$ as $x \to \infty$, then M is Feller.

In the case where the manifold M is stochastically complete we obtain a somewhat complementary result.

Theorem 3.2 *Let M be stochastically complete, and let u be a bounded solution $u > 0$ of*

$$\Delta u \geqslant \lambda u$$

outside a smooth domain $\Omega \Subset M$. If $h > 0$ is the minimal solution of

$$\begin{cases} \Delta h = \lambda h & \text{on } M \setminus \overline{\Omega}, \\ h = 1 & \text{on } \partial\Omega, \end{cases}$$

then there is a constant $c > 0$ such that

$$u(x) \leqslant \mathrm{ch}(x) \quad \text{on } M \setminus \Omega.$$

Proof Let $c = \sup_{\partial\Omega} u$. Then, for every $\varepsilon > 0$, $\Delta(u - ch - \varepsilon) \geqslant \lambda(u - ch) \geqslant \lambda(u - ch - \varepsilon)$ on $M \setminus \overline{\Omega}$ and $u - ch - \varepsilon \leqslant -\varepsilon$ on $\partial\Omega$. Therefore the function $v_\varepsilon = \max\{0, u - ch - \varepsilon\}$ is bounded, non-negative and satisfies $\Delta v_\varepsilon \geqslant \lambda v_\varepsilon$. Since M is stochastically complete $v_\varepsilon \equiv 0$, that is, $u \leqslant ch + \varepsilon$. The conclusion follows letting $\varepsilon \to 0$. □

In particular, if $h(x) \to 0$ as $x \to \infty$, we can deduce that the same holds for the original function u. This leads to the following

Corollary 3.3 [20] *Let M be stochastically complete. If M is Feller, then every bounded solution $v > 0$ of*

$$\Delta v \geqslant \lambda v \quad \text{on } M \setminus \overline{\Omega}$$

satisfies

$$v(x) \to 0 \quad \text{as } x \to \infty.$$

To state the announced result of comparison with models, given a smooth even function $G : \mathbb{R} \to \mathbb{R}$, we let $f : [0, +\infty) \to [0 + \infty)$ be the unique solution of the Cauchy problem

$$\begin{cases} f'' + Gf = 0, \\ f(0) = 0, \quad f'(0) = 1. \end{cases} \tag{11}$$

Then we have:

Theorem 3.4 [20] *Let M be a complete Riemannian m-manifold.*

(i) *Assume that M has a pole at o and that the radial sectional curvature with respect to o satisfies*

$$^{M}\,\mathrm{Sec}_{\mathrm{rad}} \leqslant G\big(r(x)\big) \quad \text{on } M \tag{12}$$

for some smooth even function $G : \mathbb{R} \to \mathbb{R}$. If the m-dimensional model M^m_f is Feller then M is Feller.

(ii) *Assume that the radial Ricci curvature of M satisfies*

$$^{M}\,\mathrm{Ricc}(\nabla r, \nabla r) \geqslant (m-1)G\big(r(x)\big),$$

where $r(x) = \mathrm{dist}(x, o)$.

If the m-dimensional model M^m_f is not Feller (thus, it has finite volume) then also M is not Feller.

Proof We give only an outline of the proof. In case (i), one shows that the minimal radial solution α of the exterior problem on $M_f \setminus B^{M_f}(0,1)$ is decreasing and since M^m_f is Feller it tends to zero at infinity.

Let $u(x) = \alpha(r(x))$. Then the curvature condition and the Laplacian Comparison Theorem imply that $\Delta r \geqslant (m-1) f'/f$, hence

$$\Delta u = \alpha''\big(r(x)\big) + \alpha'\big(r(x)\big)\Delta r \leqslant \alpha''\big(r(x)\big) + (m-1)\frac{f'}{f}\alpha'\big(r(x)\big) = \lambda u.$$

By the comparison result, the minimal solution of the exterior Dirichlet problem h satisfies $h \leqslant u$. Since $u(r(x)) \to 0$ as $r(x) \to \infty$, M is Feller.

To prove (ii), let us note that since M_f is not Feller, by Theorem 3.1, $f^{m-1} \in L^1(+\infty)$, $1/f^{m-1} \notin L^1(+\infty)$ and

$$\frac{\int_r^{+\infty} f^{m-1}(t)\,dt}{f^{m-1}(r)} \in L^1(+\infty).$$

Define

$$\alpha(r) = \int_r^{+\infty} \frac{\int_s^{+\infty} f^{m-1}(t)\,dt}{f^{m-1}(s)}\,ds.$$

A direct computation shows that

$$^{M_f}\Delta\alpha = 1.$$

Now consider

$$v(x) = \alpha\big(r(x)\big) + 1 \quad \text{on } M \setminus B_1.$$

Clearly, v is a positive bounded function, and since $\alpha' \leqslant 0$, by Laplacian comparison we have

$$\Delta v \geqslant 1 \geqslant \lambda v,$$

where $\lambda = 1/\sup v$. Since $v(x) \to 1$ as $x \to \infty$, by Corollary 3.3 M is not Feller. □

Note that in (i) above the (sectional) curvature is bounded from *above*. This is the opposite of the inequality assumed in Hsu's result, and it shows that, in contrast with what happens for stochastic completeness, Hsu's result is a genuine estimation result, and does not follow from a comparison argument.

4 Ends and Further Geometric Conditions for the Feller Property

It is clear that the Feller property is affected only by the properties of M outside a compact set Ω. The set $M \setminus \Omega$ has a finite number of unbounded connected components E_i, called the ends of M with respect to Ω. Thus, the minimal solution h of

$$\begin{cases} \Delta h = \lambda h & \text{on } M \setminus \overline{\Omega}, \\ h = 1 & \text{on } \partial\Omega, \\ h > 0 & \text{on } M \setminus \Omega, \end{cases}$$

restricts to the minimal solution h_j of the same Dirichlet problem on E_j with respect to the compact boundary ∂E_j. Furthermore, h tends to zero at infinity in M if and only if each function $h_j(x)$ tends to 0 as $E_j \ni x \to \infty$.

This suggests that the property of being Feller may be localized at the ends of M.

We say that an end E is Feller if, for some $\lambda > 0$, the minimal solution $g : E \to (0, 1]$ of the Dirichlet problem

$$\begin{cases} \Delta g = \lambda g & \text{on } \mathrm{int}(E), \\ g = 1 & \text{on } \partial E \end{cases}$$

satisfies $g(x) \to 0$ as $x \to \infty$. The usual exhausting procedure shows that g actually exists. The following statement holds:

Proposition 4.1 *Let $(M, \langle\,,\rangle)$ be a complete Riemannian manifold and let $E_1, \dots, E_k$ be the ends of M with respect to the smooth compact domain Ω. Then, the following are equivalent:*

(i) *M is Feller;*
(ii) *each end E_j has the Feller property;*
(iii) *the double $\mathcal{D}(E_j)$ of each end has the Feller property.*

Using this observation, one can easily construct new Feller or non-Feller manifolds from old ones by adding suitable ends. For instance, consider complete Riemannian manifolds M and N of the same dimension m and form their connected sum $M\#N$. This latter is Feller if and only if both M and N has the Feller property.

Since the same property also holds for parabolicity and stochastic completeness (for the latter see [2]), one may then construct examples of manifolds which show that are no obvious implications between stochastic completeness and the Feller property. By way of example, consider the warped product $M = \mathbb{R} \times_f \mathbb{S}^{m-1}$ with warping function $f(t)$ such that

$$f(t) = \begin{cases} e^{t^4} & \text{if } t \geqslant 1, \\ e^{-t^4} & \text{if } t \leqslant -1. \end{cases}$$

Then the "positive" end of M is Feller and stochastically incomplete, while the "negative" end of M is parabolic and non-Feller, so that M is both non-Feller and stochastically incomplete.

We conclude this summary of the geometric properties leading to the Feller property with two last results.

Using heat kernel estimates in the presence of an isoperimetric inequality of A. Grigor'yan [12] and a result of G. Carron [7] we obtain the following result.

Theorem 4.2 [20] *Assume that M supports an L^2-Sobolev inequality of the form*

$$\|\nabla u\|_{L^2} \geqslant S_{2,p} \|u\|_{L^{\frac{2p}{p-2}}}, \quad \textit{for every } u \in C^1_c(M).$$

Then M is Feller.

Note that according to a result of Carron [6], if the L^2-isoperimetric inequality holds off a compact set then it holds everywhere and M is Feller.

Corollary 4.3 [20] *Let M be isometrically immersed into a Cartan–Hadamard manifold. If its mean curvature vector field H satisfies*

$$\|H\|_{L^m(M)} < +\infty,$$

then M is Feller. In particular,

(i) *every Cartan–Hadamard manifold is Feller;*
(ii) *every complete, minimal submanifold in a Cartan–Hadamard manifold is Feller.*

The above result has been completed in the very recent paper [5], where it is shown that bounded mean curvature hypersurfaces properly immersed in Cartan–Hadamard manifold are Feller. The proof relies upon the comparison principle (Theorem 3.2), by means of a suitable test function $u(x)$.

Finally, we address the following

Problem 4.4 Suppose we are given a Riemannian covering

$$\pi : \big(\widehat{M}, \widehat{\langle\, ,\rangle}\big) \to \big(M, \langle\, ,\rangle\big).$$

What are the relationships between the validity of the Feller property on the covering space $\widehat{M}$ and on the base manifold M?

By comparison, recall that M is stochastically complete if and only if so is $\widehat{M}$ (see, e.g., [9]). As for parabolicity, the situation is quite different. Using subharmonic functions it is easy to see that if $\widehat{M}$ is parabolic then the base manifold M is also parabolic. In general, the converse is not true, as shown e.g. by the twice punctured complex plane, which is a parabolic manifold, as can be seen by using the well know Khas'minskii test [13], and which is universally covered by the (non-parabolic) Poincaré disk.

Let us now consider the Feller property. To begin with, consider the easiest case of coverings with a finite number of sheets.

Observe that in a finite covering one can pass from functions on M to functions on $\widehat{M}$, and vice-versa, and that a sequence of points in $\widehat{M}$ goes to infinity if and only if their projections tend to infinity in M. So one has

Proposition 4.5 [20] *Let $\pi : (\widehat{M}, \widehat{\langle\, ,\rangle}) \to (M, \langle\, ,\rangle)$ be a k-fold Riemannian covering, with $k < +\infty$. Then $\widehat{M}$ is Feller if and only if M is Feller.*

In general

$$\widehat{M} \text{ Feller} \not\Longrightarrow M \text{ Feller}.$$

Consider the 2-dimensional warped product $M = \mathbb{R} \times_f \mathbb{S}^1$ where $f(t) = e^{t^3}$. Combining the necessary and sufficient condition for a model to be Feller and the results on the Feller property for manifolds with ends, we see that M is not Feller. But since the Gaussian curvature of M is given by

$$K(t, \theta) = -\frac{f''(t)}{f(t)} \leqslant 0,$$

the universal covering $\widehat{M}$ is Cartan–Hadamard, and hence Feller by Azencott's result.

However the reverse implication

$$M \text{ is Feller} \Longrightarrow \widehat{M} \text{ is Feller}$$

always holds. Indeed, by means of results by M. Bordoni [4] on the relationship between the heat kernel of M and that of its covering $\widehat{M}$, we obtain

Theorem 4.6 [20] *If M is Feller then so is $\widehat{M}$.*

5 Applications to Geometry and PDE's

The weak maximum principle at infinity is a powerful tool to deduce qualitative information on the solutions of differential inequalities of the form

$$\Delta u \geqslant \Lambda(u). \tag{13}$$

Indeed, it implies that every solution u of (13) on the whole manifold M such that $u^* = \sup_M u < +\infty$ satisfies

$$\Lambda\big(u^*\big) \leqslant 0.$$

This fact has many applications in geometric analysis. Our aim is to apply the Feller property to investigate qualitative properties of solutions of (13) which are defined only in a neighborhood of infinity. This section is based on [3].

Recall that, according to Corollary 3.3, if M is stochastically complete and Feller, then every bounded solution $v > 0$ of $\Delta v \geqslant \lambda v$ on $M \setminus \overline{\Omega}$ satisfies $v(x) \to 0$ as $x \to \infty$. On the basis of these remarks, we prove the following:

Theorem 5.1 *Let M be a stochastically complete and Feller manifold. Consider the differential inequality*

$$\Delta u \geqslant \Lambda(u) \quad \text{on } M \setminus \Omega, \tag{14}$$

where $\Omega \Subset M$ and $\Lambda : [0, +\infty) \to [0, +\infty)$ is either continuous or it is a non-decreasing function which satisfies the following conditions:

$$\text{(i)} \quad \Lambda(0) = 0; \qquad \text{(ii)} \quad \Lambda(t) > 0 \quad \text{for every } t > 0; \qquad \text{(iii)} \quad \liminf_{t\to 0+} \frac{\Lambda(t)}{t^{\xi}} > 0$$

for some $0 \leqslant \xi \leqslant 1$. Then every bounded solution $u > 0$ of (14) must satisfy

$$\lim_{x\to\infty} u(x) = 0.$$

Proof Suppose Λ is non-decreasing. By assumption, there exists $0 < \varepsilon < 1/2$ and $c > 0$ such that

$$\Lambda(t) \geqslant ct^{\xi} \quad \text{on } (0, 2\varepsilon).$$

As $t^{\xi} \geqslant t$ on $(0, 1]$, and Λ is non-decreasing, then

$$\Lambda\big(u(x)\big) \geqslant \Lambda_{\varepsilon}\big(u(x)\big) = \begin{cases} cu & \text{if } u(x) < \varepsilon, \\ c\varepsilon & \text{if } u(x) \geqslant \varepsilon. \end{cases}$$

Since $u > 0$ is bounded, if we set $u^{\star} = \sup_{M\setminus\Omega} u$, then

$$c\varepsilon \geqslant \frac{c\varepsilon}{u^{\star}} u^{\star} \geqslant \frac{c\varepsilon}{u^{\star}} u.$$

It follows that

$$\Delta u \geqslant \Lambda_\varepsilon(u) \geqslant \lambda u,$$

where

$$\lambda = c \min\left\{1, \frac{\varepsilon}{u^\star}\right\} > 0.$$

Using the Feller property we now conclude that $u(x) \to 0$ as $x \to \infty$. □

As shown in Theorem 5.1, using the Feller property on a stochastically complete manifold enables one to extend the investigation of qualitative properties of solution of PDEs to the case where these are defined only in a neighborhood of infinity.

We are going to exemplify the use of this viewpoint in various geometric and analytic settings. We stress that the needed stochastic assumptions are enjoyed by a very rich family of examples. For instance, as seen in Sect. 2, we have the class of complete manifolds such that $\mathrm{Ricc} \geqslant -G^2(r)$, where $G(r) > 0$ is an increasing function satisfying $1/G \notin L^1(+\infty)$. Another admissible category is given by Cartan–Hadamard manifolds, or minimal submanifolds of Cartan–Hadamard manifolds (which are Feller by Corollary 4.3) with at most quadratic exponential volume growth (to guarantee stochastic completeness).

5.1 Isometric Immersions

An application of the weak maximum principle shows that if a Riemannian manifold $(M, \langle\, ,\rangle)$ is stochastically complete, then the mean curvature $\mathbf{H}$ of a bounded isometric immersion $f : M \to \mathbb{B}(O, R) \subset \mathbb{R}^n$ must satisfy

$$\sup_M |\mathbf{H}| R \geqslant 1.$$

In particular, a stochastically complete minimal submanifold in Euclidean space is necessarily unbounded.

The next result extends this to the case where the complement of a compact domain in M admits a bounded isometric immersion into $\mathbb{R}^n$.

Theorem 5.2 *Let the Riemannian manifold M be stochastically complete and Feller. Assume that, outside a compact set $\Omega \subset M$, there exists a bounded isometric immersion $f : M \setminus \Omega \to \mathbb{B}(O, R) \subset \mathbb{R}^n$. Then*

$$\sup_{M\setminus\Omega} |\mathbf{H}| R \geqslant 1.$$

Proof Assume by contradiction that

$$\sup_{M\setminus\Omega} |\mathbf{H}| R < 1, \tag{15}$$

and let $u(x) = |f(x) - O|^2 \geqslant 0$. Then

$$\Delta u \geqslant c \quad \text{on } M \setminus \overline{\Omega},$$

where we have set $c = 2m(1 - \sup_{M \setminus \Omega} |\mathbf{H}|R) > 0$. By Theorem 5.1, $u \to 0$ and therefore $f(x) \to O$ as $x \to \infty$.

Now, as strict inequality holds in (15), for $R' > R$ sufficiently close to R we have $\sup_{M \setminus \Omega} |\mathbf{H}|R' < 1$, and clearly $f(M \setminus \Omega) \subset \mathbb{B}(O', R')$ provided $|O' - O| < R' - R$. Thus we can repeat the argument with $u'(x) = |f(x) - O'|^2$ for which again we have

$$\Delta u' \geqslant c$$

with the same value c, and then $u'(x) \to 0$, i.e., $f(x) \to O' \neq O$, as $x \to \infty$. This yields the required contradiction and proves the theorem. □

We note that a modification of the above argument allows to consider just one of the ends of M with respect to Ω is isometrically immersed in a ball.

5.2 Conformal Deformations

Given a Riemannian manifold $(M, \langle\, ,\rangle)$ of dimension $m \geqslant 3$ consider the conformally related metric $\overline{\langle\, ,\rangle} = v^{\frac{4}{m-2}} \langle\, ,\rangle$ where $v > 0$ is a smooth function. Thus, the conformality factor v obeys the Yamabe equation

$$c_m^{-1} \Delta v - S v = -\overline{S} v^{\frac{m+2}{m-2}},$$

where S and $\overline{S}$ denote the scalar curvatures of $\langle\, ,\rangle$ and $\overline{\langle\, ,\rangle}$, respectively. Assume that M is stochastically complete and that

$$\sup_M S(x) \leqslant S^\star, \qquad \inf_M \overline{S}(x) \geqslant \overline{S}_\star,$$

for some constants $S^\star \geqslant 0$ and $\overline{S}_\star > 0$. An application of the weak minimum principle at infinity to the Yamabe equation shows that

$$\left(\frac{S^\star}{\overline{S}_\star}\right)^{\frac{m-2}{4}} \geqslant v_\star = \inf_M v.$$

In particular, if $S(x) \leqslant 0$ on M, then $v_\star = 0$. Actually, since the infimum of v cannot be attained, for every $\Omega \Subset M$

$$\inf_{M \setminus \Omega} v = 0.$$

Clearly, to reach these conclusions the scalar curvature bound must hold on M. As a consequence of Theorem 5.1, we obtain the following non-existence result.

Note that this applies e.g. to an expanding, gradient Ricci soliton M. Indeed, in this case, the scalar curvature assumption is compatible with the restriction $\inf_M S \leqslant 0$ imposed by the soliton structure.

Theorem 5.3 *Let $(M, \langle\, ,\rangle)$ be a stochastically complete and Feller manifold of dimension $m \geqslant 6$ such that, for some relatively compact domain Ω*

$$\sup_{M\setminus\Omega} S(x) \leqslant 0.$$

On M, one cannot perform a conformal change $\overline{\langle\, ,\rangle} = v^{\frac{4}{m-2}}\langle\, ,\rangle$ in such a way that

$$0 < v_\star \leqslant v(x) \leqslant v^\star < +\infty$$

and

$$\liminf_{x\to\infty} \overline{S}(x) = \overline{S}_\star > 0.$$

Proof Simply observe that the positive, bounded function $u(x) = v(x)^{-1}$ satisfies

$$c_m^{-1}\Delta u \geqslant -Su + \overline{S}u^{\frac{m-6}{m-2}} \geqslant \overline{S}u^{\frac{m-6}{m-2}}.$$

Since

$$0 \leqslant \frac{m-6}{m-2} < 1,$$

Theorem 5.1 yields

$$u(x) \to 0 \quad \text{as } x \to \infty. \qquad \square$$

One may wonder if the assumption that S be nonnegative at infinity implies that it can be made nonnegative everywhere on M with a conformal change of metric. However this in general would require a somewhat implicit control on the positive part of S in the set Ω (see, e.g., Proposition 1.2 in [21]).

5.3 Compact Support Property of Bounded Solutions of PDEs

We say that a certain PDE satisfies the compact support principle if all solutions in the exterior of a compact set which are non-negative and decay at infinity must have compact support. We are going to analyze some situations where the decay assumption can be relaxed. This has applications to the Yamabe problem.

Theorem 5.4 *Let M be a geodesically complete and stochastically complete, Cartan–Hadamard manifold. Let $u \geqslant 0$ be a bounded solution of*

$$\Delta u \geqslant \Lambda(u) \quad \text{on } M \setminus \Omega \tag{16}$$

for some domain $\Omega \Subset M$ *and for some non-decreasing function* $\Lambda : [0, +\infty) \to [0, +\infty)$ *satisfying the following conditions*:

$$\text{(i)}\quad \Lambda(0) = 0; \qquad \text{(ii)}\quad \Lambda(t) > 0 \quad \textit{for every } t > 0; \qquad \text{(iii)}\quad \liminf_{t \to 0+} \frac{\Lambda(t)}{t^{\xi}} > 0 \tag{17}$$

for some $0 \leqslant \xi < 1$. *Then* u *has compact support.*

Proof Recall that a Cartan–Hadamard manifold is Feller. By Theorem 5.1 we know that $u(x) \to 0$ as $x \to \infty$. The conclusion now follows from the compact support principle, that is valid under the stated assumptions on M and Λ [19, Theorem 1.1]. □

Of course for the conclusion of Theorem 5.4 to hold it suffices that M be stochastically complete, Feller and that the compact support principle hold for solutions of (16).

The above theorems can be applied to obtain nonexistence results. For instance, combining Theorems 5.4 and 5.3 we obtain

Corollary 5.5 *Let* $(M, \langle\,,\rangle)$ *be a stochastically complete Cartan–Hadamard manifold of dimension* $m \geqslant 6$. *Then the metric of* M *cannot be conformally deformed to a new metric* $\overline{\langle\,,\rangle} = v^2\langle\,,\rangle$ *with* $v_* > 0$ *and scalar curvature* $\overline{S}$ *satisfying* $\liminf_{x\to\infty} \overline{S} > 0$.

References

1. Azencott, R.: Behavior of diffusion semi-groups at infinity. Bull. Soc. Math. Fr. **102**, 193–240 (1974)
2. Bär, C., Bessa, G.P.: Stochastic completeness and volume growth. Proc. Am. Math. Soc. **138**, 2629–2640 (2010)
3. Bessa, G.P., Pigola, S., Setti, A.G.: Spectral and stochastic properties of the f-Laplacian, solutions of PDE's at infinity, and geometric applications. Rev. Mat. Iberoam., to appear. arXiv:1107.1172v1
4. Bordoni, M.: Comparing heat operators through local isometries and fibrations. Bull. Soc. Math. Fr. **128**, 151–178 (2000)
5. Brandão, C.M., Oliveira, J.: Stochastic properties of the Laplacian on Riemannian submersions. Geom. Dedic., to appear. arXiv:1109.3380v1. doi:10.1007/s10711-012-9732-2
6. Carron, G.: Une suite exacte en L^2-cohomologie. Duke Math. J. **95**, 343–372 (1998)
7. Carron, G.: Inégalités isopérimétriques et inégalités de Faber–Krahn. Sémin. Théor. Spectrale Géom. **13**, 63–66 (Année 1994–1995)
8. Davies, E.B.: Heat kernel bounds, conservation of probability and the Feller property. In: Festschrift on the occasion of the 70th birthday of Shmuel Agmon. J. Anal. Math. **58**, 99–119 (1992)
9. Elworthy, K.D.: Stochastic Differential Equations on Manifolds. London Math. Soc. Lect. Note Series, vol. 70. Cambridge University Press, Cambridge (1982)
10. Dodziuk, J.: Maximum principle for parabolic inequalities and the heat flow on open manifolds. Indiana Univ. Math. J. **32**, 703–716 (1983)

11. Grigor'yan, A.: On stochastically complete manifolds. Dokl. Akad. Nauk SSSR **290**, 534–537 (1986) (in Russian). English translation: Sov. Math. Dokl. **34**, 310–313 (1987)
12. Grigor'yan, A.: Heat kernel upper bounds on a complete non-compact manifold. Rev. Mat. Iberoam. **10**, 395–452 (1994)
13. Grigor'yan, A.: Analytic and geometric background of recurrence and non-explosion of the Brownian motion on Riemannian manifolds. Bull. Am. Math. Soc. **36**, 135–249 (1999)
14. Hsu, E.-P.: Heat semigroup on a complete Riemannian manifold. Ann. Probab. **17**, 1248–1254 (1989)
15. Hsu, E.-P.: Stochastic Analysis on Manifolds. Grad. Studies Math., vol. 38. Am. Math. Soc., Providence (2002)
16. Li, P., Karp, L.: The heat equation on complete Riemannian manifolds. Unpublished preprint. http://www.math.uci.edu/~pli/
17. Pigola, S., Rigoli, M., Setti, A.G.: A remark on the maximum principle and stochastic completeness. Proc. Am. Math. Soc. **131**, 1283–1288 (2002)
18. Pigola, S., Rigoli, M., Setti, A.G.: Maximum principles on Riemannian manifolds and applications. Mem. Am. Math. Soc. **174** (2005)
19. Pucci, P., Rigoli, R., Serrin, J.: Qualitative properties for solutions of singular elliptic inequalities on complete manifolds. J. Differ. Equ. **234**, 507–543 (2007)
20. Pigola, S., Setti, A.G.: The Feller property on Riemannian manifolds. J. Funct. Anal. **262**, 2481–2515 (2012)
21. Ratto, A., Rigoli, M., Veron, L.: Scalar curvature and conformal deformations of noncompact Riemannian manifolds. Math. Z. **25**, 395–426 (1997)
22. Varopoulos, N.Th.: Potential theory and diffusion on Riemannian manifolds. In: Conference on Harmonic Analysis in Honor of Antoni Zygmund, Chicago, Ill., 1981. Wadsworth Math. Series, vols. I, II, pp. 821–837. Wadsworth, Belmont (1983)
23. Yau, S.T.: Some function-theoretic properties of complete Riemannian manifold and their applications to geometry. Indiana Univ. Math. J. **25**, 659–670 (1976)

Characterization of Carleson Measures for Besov Spaces on Homogeneous Trees

Maria Rosaria Tupputi

Abstract We give a characterization of Carleson measures for diagonal Besov spaces of martingales on homogeneous trees by means of testing type conditions. The corresponding problem for spaces harmonic functions in the unit ball was solved by means of a capacitary condition in (Guliyev, Wu in Further Progress in Analysis, pp. 132–141 (1999)). For a given measure, however, capacitary conditions are more difficult to verify than testing conditions. The problem of characterizing Carleson measures for Besov spaces of harmonic functions by means of testing conditions is still open.

Keywords Weighted inequalities · Analysis on homogeneous trees

Mathematics Subject Classification (2010) 31C20

1 Introduction

In this note we characterize the Carleson measures for some Besov spaces on the homogeneous tree. Let T be a k-homogeneous tree with root o (see below for these and other definitions). Given a martingale $\varphi : T \to \mathbb{R}$, consider the *martingale difference function*

$$D\varphi(\alpha) = \begin{cases} \varphi(\alpha) - \varphi(\alpha^-) & \text{if } \alpha \in T \setminus \{o\}, \\ \varphi(o) & \text{if } \alpha = o. \end{cases}$$

Here α^- is the predecessor of α in T with respect to o. On the level of metaphor, we think of the tree as a model for the unit disc in the complex plane, of martingales as harmonic functions and of martingale differences as gradients of harmonic functions. This viewpoint is well known. It has its roots in the seminal work of Cartier (with harmonic functions instead of martingales) [4] and in the influential article [7].

M.R. Tupputi (✉)
Dipartimento di Matematica, Università di Bologna, Piazza di Porta S. Donato 5, 40126 Bologna, Italy
e-mail: mariarosaria.tupput2@unibo.it

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3,
DOI 10.1007/978-88-470-2853-1_15,

Our characterization is based on *testing conditions* similar to those obtained by Kerman and Saywer for weighted trace inequalities [6]. Carleson measures could also be characterized in terms of suitable capacity conditions, by a successful idea of Maz'ya [8].

We now state our main result more precisely.

We consider a k-homogeneous tree T, and observe that the nth generation of vertices consists of k^n vertices. Therefore we can uniquely parameterize $T := \{(n, j) \in \mathbb{N} \times \mathbb{N} : 1 \leqslant j \leqslant k^n\}$. Now we give an equivalent realization where the vertices are represented by subintervals of [0, 1]. Let us associate to each index pair $\alpha = (n, j) \in T$ the interval $I(\alpha) := [k^{-n}(j-1), k^{-n} j]$, and denote by $|I(\alpha)|$ its length; set $o := [0, 1]$ and consider it to be the *root* of the tree. We can define on T a partial ordering by saying that $\alpha < \beta$ if $I(\beta) \subset I(\alpha)$. If $\alpha < \beta$ or $\alpha > \beta$ and $|I(\alpha)| \cdot |I(\beta)|^{-1} \in \{k, 1/k\}$ we say that there is an *edge* between α and β. We write $\text{dist}(\alpha, \beta)$ to indicate the minimum number of edges between α and β, in particular we set $\text{dist}(\alpha) := \text{dist}(\alpha, o)$, the distance in terms of edges from a generic element α to the root.

We call *predecessor of* α the element α^- such that $\alpha > \alpha^-$ and $\text{dist}(\alpha^-) = \text{dist}(\alpha) - 1$, and *successor of* α any element β such that $\alpha < \beta$ and $\text{dist}(\beta) = \text{dist}(\alpha) + 1$. We know that, if T is a k-homogeneous tree, then every element α we has k successors $\alpha_1, \alpha_2, \ldots, \alpha_k$. A function $\varphi : T \to \mathbb{R}$ is a *martingale* if, for all $\alpha \in T$,

$$\varphi(\alpha) = \frac{1}{k} \sum_{j=1}^{k} \varphi(\alpha_j). \tag{1}$$

To each function $\varphi : T \to \mathbb{R}$ we can associate

- the discrete derivative $D\varphi : T \to \mathbb{R}$, defined by $D\varphi(\alpha) := \varphi(\alpha) - \varphi(\alpha^-)$ if $\alpha \neq o$ and by $D\varphi(o) := \varphi(o)$ at the root;
- the discrete primitive $I\varphi : T \to \mathbb{R}$ defined by $I\varphi(\alpha) := \sum_{o \leqslant \beta \leqslant \alpha} \varphi(\beta)$.

It is straightforward to verify that $D \circ I = I \circ D = Id$.

We introduce Besov spaces of martingales on T, denoted by $\mathscr{B}_p$, with $1 < p < +\infty$. A martingale $\varphi : T \to \mathbb{R}$ belongs to $\mathscr{B}_p$ if

$$\|\varphi\|_{\mathscr{B}_p}^p = \sum_{\alpha \in T} \left|D\varphi(\alpha)\right|^p < +\infty.$$

For $p = 2$ this is the Dirichlet space $\mathscr{B}_2$, which is an Hilbert space with inner product

$$\langle \varphi, \psi \rangle_{\mathscr{B}_2} := \sum_{\alpha \in T} D\varphi(\alpha) \overline{D\psi(\alpha)}.$$

Definition 1.1 A positive measure μ on T is a Carleson measure for $\mathscr{B}_p$, $1 < p < +\infty$, if there exists a constant $C(\mu) > 0$ such that for each $\varphi \in \mathscr{B}_p$ the following inequality holds:

$$\sum_{\alpha \in T} \left|\varphi(\alpha)\right|^p \mu(\alpha) \leqslant C(\mu) \sum_{\alpha \in T} \left|D\varphi(\alpha)\right|^p. \tag{2}$$

Theorem 1.2 *Given $p > 1$, let p' the conjugate of p. A positive and bounded measure μ on T is a Carleson measure for $\mathscr{B}_p$ if and only if, for all $\alpha \in T$,*

$$\sum_{x \geqslant \alpha} \mu\big(S(x)\big)^{p'} \leqslant C(\mu)\mu\big(S(\alpha)\big), \tag{3}$$

where, for each $\alpha \in T$, we have set $S(\alpha) := \{\beta \mid \beta \geqslant \alpha\}$, the box of vertex α.

Carleson measures are characterized in Theorem 1.2 below by testing inequality (2) on characteristic functions of special sets. The result is similar to those obtained in a different context by Kerman and Sawyer. We now make some comments about the proof. Inequality (2) means that the operator $Id : \mathscr{B}_p \to L^p(\mu)$ is bounded. This is equivalent to the boundedness of the adjoint operator $\Theta : L^{p'}(\mu) \to B_{p'}$. Now, Θ is an integral operator with a kernel having many cancellations. It was proved in [2] that the hypothesis of Theorem 1.2 is equivalent to the boundedness of the operator $|\Theta|$ which is obtained from Θ by replacing the signed kernel with its absolute value. Then we are left with the task of proving that boundedness of Θ, the signed kernel, implies inequality (3). The same problem arises in the dyadic case, which was considered in [1]. To get around the problem of cancellations we use a stopping time argument. In the continuous case, a result analogous to Theorem 1.2 for the corresponding Besov spaces of harmonic functions on $\mathbb{R}^n_+$ has not yet been proved. The cancellations in the kernel are rather similar to those of the discrete case. The way they are related with the hyperbolic geometry of the upper-half space, however, is more involved than in the discrete case. Moreover, the fact that *Carleson boxes* have trivial boundary in the tree in but not the continuous case provides an additional difficulty. We believe that the continuous version of Theorem 1.2 holds true, but we are not yet able to give a proof. It is possible that a version of the classical Calderòn and Zygmund decomposition argument might help here, although that argument is usually employed to show that cancellations *do* matter, while our goal is to prove that, for our kernels, they do not. At present, we have not yet been able to use the Calderòn–Zygmund techniques efficiently.

2 Proofs of the Results

First, we find the reproducing kernel for $\mathscr{B}_2$.

Definition 2.1 $K : T \times T \to \mathbb{R}$ is a reproducing kernel for $\mathscr{B}_2$ if for all martingales $\varphi \in \mathscr{B}_2$ and for all $\alpha \in T$:

$$\varphi(\alpha) = \langle K_\alpha, \varphi \rangle_{\mathscr{B}_2} = \sum_{\beta \in T} DK_\alpha(\beta)\overline{D\varphi(\beta)}. \tag{4}$$

Proposition 2.2 *The reproducing kernel for $\mathscr{B}_2$ is*

$$K_\alpha(\beta) = \begin{cases} \frac{k-1}{k}\operatorname{dist}(\alpha) & \text{if } \alpha \leqslant \beta, \\ \frac{k-1}{k}(\operatorname{dist}(\beta)+1) & \text{if } \alpha > \beta. \end{cases}$$

Combining property (4) with Proposition 2.2 we obtain the following identities for $\varphi \in L^{p'}(\mu)$:

$$\Theta\varphi(x) = \langle K_x, \Theta\varphi\rangle_{\mathscr{B}_2} = \langle K_x, \varphi\rangle_{L^2(\mu)} = \sum_{y\in T} K_x(y)\varphi(y)\mu(y).$$

Hence

$$D_x\Theta\varphi(x) = \sum_{y\in T} D_x K_x(y)\varphi(y)\mu(y).$$

Therefore

$$\|\Theta\varphi\|_{\mathscr{B}_{p'}}^{p'} = \sum_{x\in T}\left|\frac{k-1}{k}\sum_{y\in S(x)}\varphi(y)\mu(y) - \frac{1}{k}\sum_{y\approx x}\sum_{z\in S(y)}\varphi(z)\mu(z)\right|^{p'} \tag{5}$$

where we denote by $y \approx x$ the successors of the predecessor x^- of x such that $y \neq x$ (if one regards the tree as a genealogic tree, then $y \approx x$ means that y and x are brothers).

The proof of Theorem 1.2 shows that the cancellations in (5) do not add up to zero: in particular this shows that the $\mathscr{B}_{p'}$-norm of Θ applied to the characteristic functions is equivalent to the term at the left side of (3).

Proof of Proposition 2.2 If $\{\varphi_r\}_r$ is an orthonormal basis of $\mathscr{B}_2$, then it is known that, for all $\alpha, \beta \in T$:

$$K_\alpha(\beta) = \sum_r \varphi_r(\alpha)\overline{\varphi_r(\beta)}. \tag{6}$$

We denote by D_α and D_β the corresponding derivatives of K with respect to the variables α and β. From (6),

$$\begin{aligned} D_\beta K_\alpha(\beta) &= K_\alpha(\beta) - K_\alpha\left(\beta^-\right) \\ &= \sum_r \varphi_r(\alpha)\overline{\varphi_r(\beta)} - \sum_r \varphi_r(\alpha)\overline{\varphi_r\left(\beta^-\right)} = \sum_r \varphi_r(\alpha)\overline{D\varphi_r(\beta)}. \end{aligned}$$

Therefore,

$$\begin{aligned} D_\alpha D_\beta K_\alpha(\beta) &= D_\beta K_\alpha(\beta) - D_\beta K_{\alpha^-}(\beta) \\ &= \sum_r \varphi_r(\alpha)\overline{D\varphi_r(\beta)} - \sum_r \varphi_r\left(\alpha^-\right)\overline{D\varphi_r(\beta)} = \sum_r D\varphi_r(\alpha)\overline{D\varphi_r(\beta)}. \end{aligned} \tag{7}$$

Let $\mathscr{D}\mathscr{B}_2$ be the space of functions ψ for which there is a martingale $\varphi \in \mathscr{B}_2$ such that $\psi = D\varphi$. Then (1) is equivalent to

$$\sum_{j=1}^{k} \psi(\alpha_j) = 0, \tag{8}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_k$ are the successors of an element $\alpha \in T$. The ℓ^2-norm on $\mathscr{D}\mathscr{B}_2$ satisfies the identity

$$\|\psi\|^2_{\ell^2(T)} = \sum_{\alpha \in T} \left|\psi(\alpha)\right|^2 = \sum_{\alpha \in T} \left|D(I\psi)(\alpha)\right|^2 = \|I\psi\|^2_{\mathscr{B}_2} = \|\varphi\|^2_{\mathscr{B}_2}.$$

Therefore it is obvious that $\{\varphi_r\}$ is an orthonormal basis in $\mathscr{B}_2$ if and only if $\{\psi_r\}$ is an orthonormal basis in $\mathscr{D}\mathscr{B}_2$. We can choose as orthonormal basis of $\mathscr{D}\mathscr{B}_2$ the normalized kth roots of unity:

$$\psi_{\alpha,r}(\alpha_l) := \frac{1}{\sqrt{k}} e^{2\pi i r l/k}, \quad l \in \{1, 2, \ldots, k\}$$

where $r \in \{0, 1, \ldots, k-1\}$, α is an element of T and $\alpha_1, \alpha_2, \ldots, \alpha_k$ are its successors. Each function $\psi_{\alpha,r}$ is zero on all other elements of T different from the successors of α. Condition (8) is verified for all $\alpha \in T$ and $r \in \{1, 2, \ldots, k-1\}$. Clearly $\{\psi_{\alpha,j} \,|\, \alpha \in T, 1 \leqslant j \leqslant k-1\}$ is an orthonormal basis of $\mathscr{D}\mathscr{B}_2$. By (7),

$$D_\alpha D_\beta K_\alpha(\beta) = \sum_{\delta \in T} \sum_{r=1}^{k-1} \psi_{\delta,r}(\alpha) \overline{\psi_{\delta,r}(\beta)}.$$

The last expression is non-zero only if there is $\delta \in T$ such that $\alpha, \beta \in \{\delta_1, \delta_2, \ldots, \delta_k\}$. If so, there are $m, n \in \{1, 2, \ldots, k\}$ such that $\alpha = \delta_m$ and $\beta = \delta_n$, hence

$$\begin{aligned} D_\alpha D_\beta K_\alpha(\beta) &= \sum_{r=1}^{k-1} \psi_{\delta,r}(\delta_m) \overline{\psi_{\delta,r}(\delta_n)} \\ &= \frac{1}{k} \sum_{r=1}^{k-1} e^{2\pi i \frac{r(m-n)}{k}} = \begin{cases} \frac{k-1}{k} & \text{if } m = n, \\ -\frac{1}{k} & \text{if } m \neq n. \end{cases} \end{aligned}$$

Integrating over the variable β, we obtain

$$\begin{aligned} D_\alpha K_\alpha(\beta) &= I_\beta\big(D_\alpha D_\beta K_\alpha(\beta)\big) \\ &= \sum_{\delta=o}^{\beta} D_\alpha D_\beta K_\alpha(\delta) = \begin{cases} \frac{k-1}{k} & \text{if } \alpha \in [o, \beta], \\ -\frac{1}{k} & \text{if } \alpha \notin [o, \beta],\ \operatorname{dist}(\alpha, \alpha \wedge \beta) = 1, \end{cases} \end{aligned}$$

where $\alpha \in [o, \beta]$ means that $o \leqslant \alpha \leqslant \beta$ and we denote by $\alpha \wedge \beta$ *the confluent of* α *and* β, i.e. the only intersection point of the two geodesics starting at o and containing α and β respectively. We recall that a geodesic of T is a sequence $\{z_n\}_{n \geqslant 0} \subseteq T$

such that $z_n > z_{n-1}$ and $\operatorname{dist}(z_n, z_{n-1}) = 1$ for each $n \geqslant 1$. Integrating now over the variable α, we have:

$$\begin{aligned} K_\alpha(\beta) &= I_\alpha\big(D_\alpha K_\alpha(\beta)\big) \\ &= \sum_{\delta=0}^{\alpha} D_\alpha K_\delta(\beta) = \begin{cases} \frac{k-1}{k}\operatorname{dist}(\alpha) & \text{if } \alpha \leqslant \beta, \\ \frac{k-1}{k}(\operatorname{dist}(\beta)+1) & \text{if } \alpha > \beta. \end{cases} \end{aligned}$$ □

Proof of Theorem 1.2 For $x \in T$ we set $M_x := \mu(S(x))$. If $\varphi = \chi_{S(\alpha)}$, it follows from (5) that

$$\|\Theta\varphi\|^{p'}_{\mathscr{B}_{p'}} = \sum_{x>\alpha}\left|\frac{k-1}{k}M_x - \frac{1}{k}\sum_{y\approx x} M_y\right|^{p'} + \sum_{x\leqslant\alpha}\left|\frac{k-1}{k}M_\alpha\right|^{p'}. \tag{9}$$

To prove Theorem 1.2 it is enough to show that

$$\sum_{x>\alpha}\left|\frac{k-1}{k}M_x - \frac{1}{k}\sum_{y\approx x} M_y\right|^{p'} + \left|\frac{k-1}{k}M_\alpha\right|^{p'} \approx \sum_{x\geqslant\alpha} M_x^{p'}. \tag{10}$$

Indeed, denote by $I^*_\mu : L^{p'}(\mu) \to L^{p'}$ the adjoint of $I : L^p \to L^p(\mu)$. Then, for $\varphi := \chi_{S(\alpha)}$ and $x \in T$,

$$I^*_\mu(\varphi)(x) = \sum_{y\in S(x)} \varphi(y)\mu(y) = \begin{cases} M_x & \text{if } x > \alpha, \\ M_\alpha & \text{if } x \leqslant \alpha. \end{cases}$$

Thus

$$\big\|I^*_\mu(\varphi)\big\|^{p'}_{L^{p'}} = \sum_{x\in T}\big|I^*_\mu(\varphi)(x)\big|^{p'} = \sum_{x\geqslant\alpha} M_x^{p'} + \sum_{x<\alpha} M_\alpha^{p'}. \tag{11}$$

If (10) is true, it follows from (9) and (11) that

$$\|\Theta\varphi\|^{p'}_{\mathscr{B}_{p'}} \approx \big\|I^*_\mu(\varphi)\big\|^{p'}_{L^{p'}}. \tag{12}$$

It has been proved in [2] that I^*_μ is bounded if and only if it is bounded only on characteristic functions $\chi_{S(\alpha)}$. By (12) it is sufficient to test the boundedness of the operator Θ only on this type of functions.

It is immediately seen that

$$\begin{aligned} \sum_{x>\alpha}\left|\frac{k-1}{k}M_x - \frac{1}{k}\sum_{y\approx x} M_y\right|^{p'} &\leqslant \sum_{x>\alpha}\left|\frac{k-1}{k}M_x + \frac{1}{k}\sum_{y\approx x} M_y\right|^{p'} \\ &\leqslant \sum_{x>\alpha}\left|\frac{k-1}{k}M_{x^-} + \frac{k-1}{k}M_{x^-}\right|^{p'} \leqslant C\sum_{x\geqslant\alpha} M_x^{p'}. \end{aligned}$$

The opposite inequality

$$\sum_{x \geqslant \alpha} M_x^{p'} \leqslant C \left| \frac{k-1}{k} M_\alpha \right|^{p'} + C \sum_{x > \alpha} \left| \frac{k-1}{k} M_x - \frac{1}{k} \sum_{y \approx x} M_y \right|^{p'} \tag{13}$$

is not trivial.

To prove (13) we need some preliminary definitions. Let $\varepsilon > 0$ to be specified later. Let $\alpha \in T$ and $\xi := \{z_n\}_{n \geqslant 0}$ a geodesic such that $z_0 = \alpha$. We define *m-stopping times* recursively as follows:

$$t_m := t_m(\xi) := \inf \left\{ t > t_{m-1} : M_{z_t} > \left(\frac{1+\varepsilon}{k} \right) M_{z_t^-} \right\},$$

where $m > 0$ and $t_0 := 0$. A point $b \in T$ such that $b = z_{t_m}$ on some geodesic starting at α is called an *m-stopping point*. Let $A(b)$ be the set of the $(m+1)$-stopping points u such that $u > b$. If $SP(\alpha)$ denotes the set of the stopping points, we prove that

$$\sum_{x \geqslant \alpha} M_x^{p'} \leqslant C \sum_{x \geqslant \alpha, x \in SP(\alpha)} M_x^{p'}.$$

To prove this last inequality is enough to show that, for all $m \geqslant 0$ and for each m-stopping point b,

$$\sum_{b \leqslant c < A(b)} M_c^{p'} \leqslant C M_b^{p'}, \tag{14}$$

where $c < A(b)$ means that no $u \in A(b)$ verifies the inequality $u \leqslant c$. Let n be a positive integer and $b \leqslant c < A(b)$ such that $\operatorname{dist}(b, c) = n$. Let $c_1, c_2, \ldots, c_k$ be the successors of c^-. Then there is $j \in \{1, 2, \ldots, k\}$ such that $c = c_j$. If there exists $l \in \{1, 2, \ldots, k-1\}$ such that for all $1 \leqslant s \leqslant l$ we have $b < c_s < A(b)$, then

$$M_c^{p'} + \sum_{s=1}^{l} M_{c_s}^{p'} \leqslant (l+1) \left(\frac{1+\varepsilon}{k} \right)^{p'} M_{c^-}^{p'} \leqslant k^{1-p'} (1+\varepsilon)^{p'} M_{c^-}^{p'}. \tag{15}$$

If $c_1, \ldots, c_{j-1}, c_{j+1}, \ldots, c_k$ are not between b and $A(b)$, we consider:

$$M_c^{p'} \leqslant \left(\frac{1+\varepsilon}{k} \right)^{p'} M_{c^-}^{p'} \leqslant k^{1-p'} (1+\varepsilon)^{p'} M_{c^-}^{p'}. \tag{16}$$

We choose ε such that $k^{1-p'}(1+\varepsilon)^{p'} = 1 - \delta < 1$, where $0 < \delta < 1$. If either (15) or (16) holds, we have, by iteration $\operatorname{dist}(b, c)$:

$$\begin{aligned} \sum_{b < c < A(b),\, \operatorname{dist}(b,c) = n} M_c^{p'} &\leqslant (1-\delta) \sum_{b < c < A(b),\, \operatorname{dist}(b,c) = n-1} M_c^{p'} \\ &\leqslant (1-\delta)^2 \sum_{b < c < A(b),\, \operatorname{dist}(b,c) = n-2} M_c^{p'} \\ &\leqslant \cdots \leqslant (1-\delta)^n M_b^{p'}. \end{aligned} \tag{17}$$

Inequality (14) now follows from summing over $n > 0$ in (17). Indeed

$$\sum_{b \leqslant c < A(b)} M_c^{p'} \leqslant M_b^{p'} \sum_{n=0}^{+\infty} (1-\delta)^n \leqslant C M_b^{p'}.$$

Now let $b > \alpha$ be a stopping point. For $y \approx b$ the boxes $S(y) \subset S(b^-)$ are all mutually disjoint. Therefore

$$\sum_{y \approx b} M_y \leqslant M_{b^-} - M_b.$$

Then

$$\begin{aligned} \frac{1}{k} \sum_{y \approx b} M_y &\leqslant \frac{1}{k}(M_{b^-} - M_b) \leqslant \frac{1}{k}\left(\frac{k}{1+\varepsilon} - 1\right) M_b \\ &\leqslant \left(\frac{1}{1+\varepsilon} - 1\right) M_b + \frac{k-1}{k} M_b. \end{aligned}$$

Hence

$$\frac{k-1}{k} M_b - \frac{1}{k} \sum_{y \approx b} M_y \geqslant (\varepsilon/1+\varepsilon) M_b.$$

Putting $C := (1+\varepsilon/\varepsilon)^{-p'}$, we obtain:

$$M_b^{p'} \leqslant C \left| \frac{k-1}{k} M_b - \frac{1}{k} \sum_{y \approx b} M_y \right|^{p'}.$$

Finally, summing over b, we have

$$\begin{aligned} \sum_{b > \alpha,\, b \in SP(\alpha)} M_b^{p'} &\leqslant C \sum_{b > \alpha,\, b \in SP(\alpha)} \left| \frac{k-1}{k} M_b - \frac{1}{k} \sum_{y \approx b} M_y \right|^{p'} \\ &\leqslant C \left| \frac{k-1}{k} M_\alpha \right|^{p'} + C \sum_{b > \alpha} \left| \frac{k-1}{k} M_b - \frac{1}{k} \sum_{y \approx b} M_y \right|^{p'}. \end{aligned}$$

By the definition of stopping point,

$$\sum_{b \geqslant \alpha} M_b^{p'} \leqslant \sum_{b > \alpha,\, b \in SP(\alpha)} M_b^{p'}.$$

Hence (13) is fully proved. □

Acknowledgements We thank the referee for her/his interesting comments and suggestions which helped, we hope, to improve our presentation.

References

1. Arcozzi, N., Rochberg, R.: Topics in dyadic Dirichlet spaces. N.Y. J. Math. **10**, 45–67 (2004)
2. Arcozzi, N., Rochberg, R., Sawyer, E.: Carleson measures for analytic Besov spaces. Rev. Mat. Iberoam. **18**, 443–510 (2002)
3. Aronszajn, N.: Theory of reproducing kernels. Trans. Am. Math. Soc. **68**, 337–404 (1950)
4. Cartier, P.: Fonctions harmoniques sur un arbre. Symp. Math. **9**, 203–270 (1972)
5. Guliyev, V.S., Wu, Z.: Strong-type estimate and Carleson measures for weighted Besov-Lipschitz spaces. In: Further Progress in Analysis, pp. 132–141 (1999)
6. Kerman, R., Sawyer, E.: The trace inequality and eigenvalue estimates for Schrödinger operators. Ann. Inst. Fourier **36**(4), 207–228 (1985)
7. Korányi, A., Picardello, M.A., Taibleson, M.H.: Hardy spaces on non-homogeneous trees. Symp. Math. **29**, 205–265 (1984)
8. Maz'ya, V.: Sobolev Spaces. Springer, Berlin (1985)

Atomic and Maximal Hardy Spaces on a Lie Group of Exponential Growth

Maria Vallarino

Abstract The main part of this paper is a shorter version of a joint work with P. Sjögren. Let G be the Lie group $\mathbb{R}^2 \rtimes \mathbb{R}^+$ endowed with the Riemannian symmetric space structure. Let X_0, X_1, X_2 be a distinguished basis of left-invariant vector fields of the Lie algebra of G and define the Laplacian $\Delta = -(X_0^2 + X_1^2 + X_2^2)$. We recall the definition and the main properties of the atomic Hardy space H^1_{at} introduced on the group G in a previous paper of the author. Then we introduce a maximal Hardy space $H^1_{\mathrm{max,h}}$ on G defined in terms of the maximal function associated with the heat kernel of the Laplacian Δ. We show that the atomic Hardy space is strictly included in the heat maximal Hardy space. In the last part of the paper, which is new, we consider the maximal Hardy space $H^1_{\mathrm{max,p}}$ defined in terms of the Poisson kernel of the Laplacian Δ and show that it strictly contains the atomic Hardy space H^1_{at}.

Keywords Heat kernel · Maximal function · Hardy space · Lie groups · Exponential growth

Mathematics Subject Classification (2010) Primary 22E30 · 42B30 · Secondary 35K08 · 42B25

1 Introduction

The theory of Hardy spaces on the Euclidean space $\mathbb{R}^n$ plays an important role in harmonic analysis and has been systematically developed (see, for instance, [15, 16]). Different equivalent definitions of the classical Hardy space are available in the literature. In particular, we shall recall here the *atomic* and the *maximal* definitions. The atomic Hardy space $H^1_{\mathrm{at}}(\mathbb{R}^n)$ is defined as the set of integrable functions f which admit an atomic decomposition of the form $f = \sum_j c_j a_j$, where $\sum_j |c_j| < \infty$ and each a_j is an atom, i.e., a function supported in a ball, with vanishing integral

M. Vallarino (✉)
Dipartimento di Scienze Matematiche, Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Torino, Italy
e-mail: maria.vallarino@polito.it

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3,
DOI 10.1007/978-88-470-2853-1_16,

and satisfying a suitable size condition. The maximal Hardy space $H^1_{\max}(\mathbb{R}^n)$ is the set of functions $f \in L^1(\mathbb{R}^n)$ such that $\mathscr{M}_\phi f \in L^1(\mathbb{R}^n)$, where

$$\mathscr{M}_\phi f(x) = \sup_{t>0} \left| f * \phi_t(x) \right| \quad \text{for every } x \in \mathbb{R}^n,$$

ϕ is any Schwartz function with $\int \phi(x)\,dx \neq 0$ and $\phi_t(x) = t^{-n}\phi(t^{-1}x)$. It is well known that the maximal Hardy space does not depend on the test function ϕ chosen to define the maximal function and that the atomic and maximal Hardy spaces coincide.

In the last forty years many efforts have been made to develop a theory of Hardy spaces in settings which are different from the Euclidean one.

On a space $\mathbb{X}$ of homogeneous type R.R. Coifman and G. Weiss [2, 3] introduced an atomic Hardy space $H^1_{\mathrm{at}}(\mathbb{X})$. Under some additional assumption, different maximal characterizations of the Hardy space $H^1_{\mathrm{at}}(\mathbb{X})$ were obtained by R.A. Macías and C. Segovia [12], W.M. Li [10], L. Grafakos, L. Liu and D. Yang [7, 8].

On the Euclidean space $\mathbb{R}^n$ endowed with a nondoubling Radon measure μ of polynomial growth, X. Tolsa [17] defined an atomic Hardy space $H^1_{\mathrm{at}}(\mu)$ and proved that it can be characterized by a grand maximal operator as in the classical setting.

In this paper we shall work on a Lie group of exponential growth: in this setting we define a suitable atomic Hardy space and then discuss a maximal characterization of such a space.

Let G be the Lie group $\mathbb{R}^2 \rtimes \mathbb{R}^+$ with the following product rule:

$$(x_1, x_2, a) \cdot \left(x_1', x_2', a'\right) = \left(x_1 + a x_1', x_2 + a x_2', a a'\right)$$

for $(x_1, x_2, a), (x_1', x_2', a') \in G$. We shall denote by x the point (x_1, x_2, a). The group G is not unimodular; its right and left Haar measures are given by

$$d\rho(x) = a^{-1}\,dx_1\,dx_2\,da \quad \text{and} \quad d\lambda(x) = a^{-3}\,dx_1\,dx_2\,da,$$

respectively. The modular function is thus $\delta(x) = a^{-2}$. Throughout this paper, unless explicitly stated, we use the right measure ρ on G and denote by L^p, $\|\cdot\|_p$ and $\langle\cdot,\cdot\rangle$ the L^p-space, the L^p-norm and the L^2-scalar product with respect to the measure ρ.

The group G has a Riemannian symmetric space structure, and the corresponding metric, which we denote by d, is that of the three-dimensional hyperbolic half-space. The metric d is invariant under left translation and given by

$$\cosh r(x) = \frac{a + a^{-1} + a^{-1}|x|^2}{2}, \tag{1}$$

where $r(x) = \mathrm{dist}(x, e)$ denotes the distance of the point x from the identity $e = (0, 0, 1)$ of G and $|x|$ denotes the Euclidean norm of the two-dimensional vector (x_1, x_2): $|x| = \sqrt{x_1^2 + x_2^2}$. The measure of a hyperbolic ball B_r, centered at the identity and of radius r, behaves like

$$\lambda(B_r) = \rho(B_r) \sim \begin{cases} r^3 & \text{if } r < 1, \\ e^{2r} & \text{if } r \geqslant 1. \end{cases}$$

Thus G is a group of *exponential growth*. In this context, the classical Calderón–Zygmund theory and the classical definitions of the atomic Hardy space [2, 3, 15] do not apply. But W. Hebisch and T. Steger [9] have constructed a Calderón–Zygmund theory which applies to some spaces of exponential growth, in particular to the space (G, d, ρ) defined above. The main idea is to replace the family of balls which is used in the classical Calderón–Zygmund theory by a suitable family of parallelepipeds which we call *Calderón–Zygmund sets*. The definition appears in [9] and implicitly in [6], and reads as follows.

Definition 1.1 A Calderón–Zygmund set is a parallelepiped $R = [x_1 - L/2, x_1 + L/2] \times [x_2 - L/2, x_2 + L/2] \times [ae^{-r}, ae^r]$, where $L > 0$, $r > 0$ and $(x_1, x_2, a) \in G$ are related by

$$e^2 ar \leqslant L < e^8 ar \quad \text{if } r < 1,$$
$$ae^{2r} \leqslant L < ae^{8r} \quad \text{if } r \geqslant 1.$$

The point (x_1, x_2, a) is called the center of R.

We let $\mathscr{R}$ denote the family of all Calderón–Zygmund sets, and observe that $\mathscr{R}$ is invariant under left translation. Given $R \in \mathscr{R}$, we define its dilated set as $R^* = \{x \in G : \operatorname{dist}(x, R) < r\}$. There exists an absolute constant $C_0 > 0$ such that $\rho(R^*) \leqslant C_0 \rho(R)$ and

$$R \subset B\big((x_1, x_2, a), C_0 r\big). \tag{2}$$

By using the Calderón–Zygmund sets, it is natural to introduce an atomic Hardy space H^1_{at} on the group G, as follows (see [18] for details). We define an *atom* as a function A in L^1 such that

(i) A is supported in a Calderón–Zygmund set R;
(ii) $\|A\|_\infty \leqslant \rho(R)^{-1}$;
(iii) $\int A \, d\rho = 0$.

The atomic Hardy space is now defined in the standard way:

Definition 1.2 The atomic Hardy space H^1_{at} is the space of all functions f in L^1 which can be written as $f = \sum_j c_j A_j$, where A_j are atoms and c_j are complex numbers such that $\sum_j |c_j| < \infty$. We denote by $\|f\|_{H^1_{\mathrm{at}}}$ the infimum of $\sum_j |c_j|$ over such decompositions.

The dual of the atomic Hardy space has been identified in [18] and is defined as follows.

Definition 1.3 The space $\mathscr{BMO}$ is the space of all functions in L^1_{loc} such that

$$\sup_{R\in\mathscr{R}} \frac{1}{\rho(R)} \int_R |g - g_R| \, d\rho < \infty,$$

where g_R denotes the mean value of g in the set R. The space BMO is $\mathscr{BMO}$ modulo the subspace of constant functions. It is a Banach space endowed with the norm

$$\|g\|_{\mathrm{BMO}} = \sup\left\{ \frac{1}{\rho(R)} \int_R |g - g_R| \, d\rho : R \in \mathscr{R} \right\}.$$

For any g in BMO the functional ℓ defined on any atom A by

$$\ell(A) = \int g A \, d\rho,$$

extends to a bounded functional on H^1_{at}. Furthermore, any functional in the dual of H^1_{at} is of this type, and $\|\ell\|_{(H^1_{\mathrm{at}})^*} \sim \|g\|_{\mathrm{BMO}}$. Given functions g in BMO and f in H^1_{at} we shall denote by $\langle g, f\rangle$ the action of g on f in this duality.

The atomic Hardy space has the following interesting interpolation property [11]:

$$\left(H^1_{\mathrm{at}}, L^2\right)_{[\theta]} = L^p \quad \text{for every } \theta \in (0,1),\ \frac{1}{p} = 1 - \frac{\theta}{2},$$

where $(A, B)_{[\theta]}$ denotes the interpolation space between two Banach spaces A and B of index θ via the complex interpolation method. It is thus useful to study the H^1_{at}–L^1 boundedness of linear operators which are bounded on L^2, because it will imply their boundedness on L^p, for every $p \in (1,2)$. In particular, the atomic Hardy space has turned out useful for studying the boundedness of singular integral operators related to a distinguished Laplacian on G, which we now define.

Let X_0, X_1, X_2 denote the left-invariant vector fields

$$X_0 = a\partial_a, \qquad X_1 = a\partial_{x_1}, \qquad X_2 = a\partial_{x_2},$$

which span the Lie algebra of G. The Laplacian $\Delta = -(X_0^2 + X_1^2 + X_2^2)$ is a left-invariant operator which is essentially selfadjoint on $L^2(\rho)$. A suitable class of Mihlin–Hörmander multipliers of Δ are bounded from the atomic Hardy space to L^1 [18]. It is proved in [13] that the first order Riesz transforms $X_i\Delta^{-1/2}$, $i = 0, 1, 2$, are bounded from H^1_{at} to L^1, while the Riesz transforms $\Delta^{-1/2}X_i$, $i = 0, 1, 2$, are unbounded from H^1_{at} to L^1.

In this paper we aim to investigate further properties of the atomic Hardy space. In particular, we shall study a maximal characterization of this space via the heat kernel of the Laplacian Δ. It is well known that the heat semigroup $(e^{-t\Delta})_{t>0}$ is given by a kernel h_t, in the sense that $e^{-t\Delta} f = f * h_t$ for suitable functions f. Let $\mathscr{M}_h$ denote the corresponding maximal operator, defined by

$$\mathscr{M}_h f(x) = \sup_{t>0} \left| f * h_t(x) \right| \quad \text{for every } x \in G. \tag{3}$$

M. Cowling, G. Gaudry, S. Giulini and G. Mauceri [4] proved that $\mathscr{M}_h$ is of weak type $(1, 1)$ and bounded on L^p for all $p > 1$. We shall now define a maximal Hardy space on G by means of the heat maximal function.

Definition 1.4 The maximal Hardy space $H^1_{\mathrm{max,h}}$ is the space of all functions f in L^1 such that $\mathscr{M}_h f$ is in L^1. The norm on this space is defined by

$$\|f\|_{H^1_{\mathrm{max,h}}} = \|\mathscr{M}_h f\|_1 \quad \text{for every } f \in H^1_{\mathrm{max,h}}.$$

We are interested in the relationship between the atomic and the maximal Hardy spaces. In Sect. 2 we shall prove that the heat maximal function is bounded from the atomic Hardy space to L^1. This implies that $H^1_{\mathrm{at}} \subseteq H^1_{\mathrm{max,h}}$. In Sect. 3 we shall prove that the previous inclusion is strict. Thus in our setting there is no characterization of the atomic Hardy space by means of the heat maximal operator. This is in contrast with the classical theory and other settings, as we mentioned at the beginning of the Introduction.

It is natural to ask whether we are able to find a maximal characterization of the atomic Hardy space, in terms of a different maximal function. In the last section of the paper we consider the Poisson kernel p_t associated with the Laplacian Δ, i.e., the convolution kernel of the semigroup $(e^{-t\sqrt{\Delta}})_{t>0}$, and the corresponding Poisson maximal function $\mathscr{M}_p$ defined as

$$\mathscr{M}_p f(x) = \sup_{t>0} \big| f * p_t(x) \big| \quad \text{for every } x \in G.$$

We then define a maximal Hardy space on G by means of the Poisson maximal function.

Definition 1.5 The Poisson maximal Hardy space $H^1_{\mathrm{max,p}}$ is the space of all functions f in L^1 such that $\mathscr{M}_p f$ is in L^1. The norm on this space is defined by

$$\|f\|_{H^1_{\mathrm{max,p}}} = \|\mathscr{M}_p f\|_1 \quad \text{for every } f \in H^1_{\mathrm{max,p}}.$$

We shall show in Sect. 4 that also the Poisson maximal Hardy space strictly contains H^1_{at}. Thus there is no characterization of the atomic Hardy space by means of the Poisson maximal operator.

The content of this paper was the subject of a talk given by the author at the XXXI Conference in Harmonic Analysis dedicated to Alessandro Figà-Talamanca (Roma, May 30–June 4 2011). We should mention that all the details of the proofs of the results announced in Sects. 2 and 3 are contained in the paper [14], while the content of Sect. 4 is new.

2 The Inclusion $H^1_{\mathrm{at}} \subseteq H^1_{\mathrm{max,h}}$

Let us first introduce some notation and give some technical preliminary results.

We shall use the following integration formula [5, Lemma 1.3]: for any radial function f such that $\delta^{1/2} f$ is integrable,

$$\int_G \delta^{1/2} f \, d\rho = \int_0^\infty f(r) r \sinh r \, dr. \tag{4}$$

The convolution in G is defined by

$$f * g(x) = \int f\big(xy^{-1}\big) g(y) \, d\rho(y). \tag{5}$$

Let h_t denote the heat kernel associated with Δ. It is well known [4, Theorem 5.3, Proposition 5.4], [1, formula (5.7)] that

$$h_t(x) = \frac{1}{8\pi^{3/2}} \delta^{1/2}(x) \frac{r(x)}{\sinh r(x)} t^{-3/2} e^{-r^2(x)/4t} \quad \text{for every } x \in G. \tag{6}$$

Note that $h_t = \delta^{1/2} q_t$, where q_t is the heat kernel associated with the operator $\mathscr{L} - I$ and $\mathscr{L}$ is the Laplace–Beltrami operator on the three-dimensional hyperbolic space. The kernel q_t and its gradient were studied in [1, 5]. Since $r(x^{-1}) = r(x)$ we have that

$$h_t(x) = \delta(x) h_t\big(x^{-1}\big) \quad \text{for every } x \in G. \tag{7}$$

Via a change of variables in (5), this implies

$$f * h_t(x) = \int f\big(z^{-1}\big) h_t(zx) \, d\rho(z) = \delta(x) \int f(y) h_t\big(x^{-1} y\big) \, d\rho(y). \tag{8}$$

We shall need the observation that for any $r > 0$

$$\sup_{t>0} t^{-3/2} e^{-\frac{r^2}{4t}} \sim r^{-3}. \tag{9}$$

We now give some properties of the heat kernel which follow easily by formulae (6) and (9) and direct computations.

Lemma 2.1 *For any point $x \in G$ the derivatives of the heat kernel h_t along the vector fields X_i, $i = 0, 1, 2$, are the following, where r denotes $r(x)$:*

(i) $X_i h_t(x) = h_t(x) \dfrac{x_i}{\sinh r} \left(\dfrac{1}{r} - \dfrac{\cosh r}{\sinh r} - \dfrac{r}{2t} \right)$ *for $i = 1, 2$;*

(ii) $X_0 h_t(x) = h_t(x) \left(\dfrac{a}{r \sinh r} - \dfrac{\cosh r}{r \sinh r} - \dfrac{a \cosh r}{\sinh^2 r} + \dfrac{1}{\sinh^2 r} + \dfrac{r}{2t} \dfrac{\cosh r - a}{\sinh r} \right)$;

(iii) $\sup_{t>0} |X_i h_t(x)| \lesssim r^{-4}$ *for $i = 0, 1, 2$ and for all $x \in B_1$;*

(iv) $\sup_{t>0} |X_i h_t(x)| \lesssim \dfrac{|x|}{a r^2 \cosh^2 r}$ *for $i = 1, 2$ and for all $x \in B_1^c$;*

(v) $\sup_{t>0} |X_0 h_t(x)| \lesssim \dfrac{1}{a r^3 \cosh r} + \dfrac{1}{r^2 \cosh^2 r}$ *for all $x \in B_1^c$.*

For suitable functions F, we denote by $\|\nabla F\|$ the Euclidean norm of the vector $\nabla F = (X_0 F, X_1 F, X_2 F)$.

By applying the previous lemma, it is easy to estimate the integral over the complement of a small ball B_{r_1} of the function

$$G_{r_1}(x) = \sup_{B(x, r_1/2)} \sup_{t>0} \|\nabla h_t\|.$$

Lemma 2.2 *For any* $r_1 \in (0, 2C_0)$

$$\int_{B_{r_1}^c} G_{r_1}\, d\rho \leqslant \frac{C}{r_1},$$

where C_0 *is the constant which appears in* (2).

The result of the previous lemma is not surprising, since it is similar to the analogous result in the Euclidean setting. We do not give the details of the proof, which is based on Lemma 2.1 and formula (1).

We now prove the main result of this section.

Theorem 2.3 *The heat maximal function* $\mathscr{M}_h$ *is bounded from* H^1_{at} *to* L^1.

Proof To prove the theorem it suffices to show that there exists a positive constant κ such that, for any atom A supported in a Calderón–Zygmund set centered at the identity,

$$\|\mathscr{M}_h A\|_1 \leqslant \kappa. \tag{10}$$

Let us thus suppose that A is an atom supported in the Calderón–Zygmund set $R = Q \times [e^{-r_0}, e^{r_0}] = [-L/2, L/2] \times [-L/2, L/2] \times [e^{-r_0}, e^{r_0}]$. We study the cases when $r_0 < 1$ and $r_0 \geqslant 1$.

Case $r_0 < 1$. By (2) $R \subset B_{C_0 r_0}$. Since $r_0 < 1$, the set $\widetilde{R} = B_{2C_0 r_0}$ has measure comparable with $\rho(R)$, and we get

$$\|\mathscr{M}_h A\|_{L^1(\widetilde{R})} \leqslant \rho(\widetilde{R})^{1/2} \|\mathscr{M}_h\|_{L^2 \to L^2} \|A\|_2 \lesssim \rho(R)^{1/2} \rho(R)^{-1/2} \lesssim 1. \tag{11}$$

It remains to consider $\mathscr{M}_h A$ on the complement of $\widetilde{R}$. For any point x in $\widetilde{R}^c$, we obtain from (8) and the cancellation condition of the atom

$$A * h_t(x) = \delta(x) \int_R A(y) \big[h_t\big(x^{-1} y\big) - h_t\big(x^{-1}\big) \big]\, d\rho(y). \tag{12}$$

From the Mean Value Theorem in $\mathbb{R}^3$, the left invariance of ∇ and the estimate $\|A\|_1 \leq 1$, we obtain that

$$\big| A * h_t(x) \big| \lesssim \delta(x) r_0 \sup_{B(x^{-1}, C_0 r_0)} \|\nabla h_t\| \int_R A(y)\, d\rho(y) \lesssim r_0 \delta(x) \sup_{B(x^{-1}, C_0 r_0)} \|\nabla h_t\|,$$

and so

$$\left|\mathscr{M}_h A(x)\right| \lesssim r_0 \delta(x) \sup_{B(x^{-1}, C_0 r_0)} \sup_{t>0} \|\nabla h_t\|.$$

We now integrate over $x \in \widetilde{R}^c$, inverting the variable x. Then Lemma 2.2 with $r_1 = 2C_0 r_0$ leads to

$$\int_{\widetilde{R}^c} \left|\mathscr{M}_h A(x)\right| d\rho(x) \lesssim 1.$$

This ends the case $r_0 < 1$.

Case $r_0 \geqslant 1$. In this case define $R^{**} = \{x : |x| < \gamma L, \frac{1}{L} < a < \gamma L\}$, where $\gamma > 1$ is a suitable constant. Then $\rho(R^{**}) \sim \rho(R)$ and there exists a constant C such that

$$\|\mathscr{M}_h A\|_{L^1(R^{**})} \leqslant \rho\big(R^{**}\big)^{1/2} \|\mathscr{M}_h\|_{L^2 \to L^2} \|A\|_2 \leqslant C \rho(R)^{1/2} \rho(R)^{-1/2} \leqslant C. \quad (13)$$

It remains to consider $\mathscr{M}_h A$ on the complementary set of R^{**}. We write $(R^{**})^c = \Omega_1 \cup \Omega_2 \cup \Omega_3$, where

$$\Omega_1 = \left\{x \in G : a < \frac{1}{L}, |x| < \gamma L\right\},$$
$$\Omega_2 = \big\{x \in G : a > \gamma L, |x| < a\big\},$$
$$\Omega_3 = \big\{x \in G : |x| > \max(a, \gamma L)\big\}.$$

In (12) we write $y = (y_1, y_2, b) \in R$ and $x = (x_1, x_2, a)$ so that $x^{-1} y = (a^{-1}(y_1 - x_1), a^{-1}(y_2 - x_2), a^{-1} b)$. We now study $\mathscr{M}_h A$ on $\Omega_1, \Omega_2, \Omega_3$ separately.

For $x \in \Omega_1$ and $y \in R$, (1) implies

$$\cosh r\big(x^{-1} y\big) \geqslant \frac{a^{-1} b + a^{-1} b^{-1} |x - y|^2}{2} \geqslant \frac{a^{-1} b}{2} \geqslant \frac{a^{-1/2}}{2} \geqslant \frac{e}{2}.$$

Therefore, by (9)

$$\begin{aligned}
\sup_{t>0} h_t\big(x^{-1} y\big) &\lesssim \delta^{1/2}\big(x^{-1} y\big) r\big(x^{-1} y\big)^{-3} \frac{r(x^{-1} y)}{\sinh r(x^{-1} y)} \\
&\lesssim \big(a^{-1} b\big)^{-1} \frac{(\log a^{-1})^{-2}}{a^{-1} b^{-1} (b^2 + |x - y|^2)} \\
&= \frac{a^2}{(\log a^{-1})^2 (b^2 + |x - y|^2)}.
\end{aligned}$$

The previous estimate and an integration over the set R implies that for all x in Ω_1,

$$\sup_{t>0} \left|A * h_t(x)\right| = \delta(x) \sup_{t>0} \left| \int_R A(y) h_t\big(x^{-1} y\big) \, d\rho(y) \right|$$

$$\lesssim \delta(x)\|A\|_\infty \int_R \sup_{t>0} h_t\big(x^{-1}y\big)\, d\rho(y)$$
$$\lesssim \frac{r_0}{L^2(\log a^{-1})^2}.$$

By taking the integral over the region Ω_1, we obtain that

$$\int_{\Omega_1} \mathscr{M}_h A(x)\, d\rho(x) \lesssim \frac{r_0}{L^2} \int_0^{1/L} \frac{da}{a(\log a^{-1})^2} \int_{|x|<\gamma L} dx_1\, dx_2$$
$$\lesssim \frac{r_0}{\log L} \sim 1. \tag{14}$$

In order to study the integral of $\mathscr{M}_h A$ on the remaining regions, we use the cancellation condition of the atom and write

$$A * h_t(x) = \delta(x) \int_R A(y)\big[h_t\big(x^{-1}y\big) - h_t\big(x^{-1}\tilde{y}\big) + h_t\big(x^{-1}\tilde{y}\big) - h_t\big(x^{-1}\big)\big]\, d\rho(y),$$

where for $y=(y_1, y_2, b)\in R$ we write $\tilde{y}=(y_1, y_2, 1)$ and thus $y=\tilde{y}\exp(\log b X_0)$. Setting $q_1(x)=\sup_{t>0, y\in R} |h_t(x^{-1}y)-h_t(x^{-1}\tilde{y})|$ and $q_2(x)=\sup_{t>0, y\in R} |h_t(x^{-1}\tilde{y}) - h_t(x^{-1})|$, we conclude that

$$\mathscr{M}_h A(x) \leqslant \delta(x)\big(q_1(x) + q_2(x)\big) \int_R \big|A(y)\big|\, d\rho(y) \leqslant \delta(x)\big(q_1(x) + q_2(x)\big).$$

To estimate q_2 observe that

$$h_t\big(x^{-1}\tilde{y}\big) - h_t\big(x^{-1}\big) = \int_0^1 (y_1X_1 + y_2X_2) h_t\big(x^{-1}\exp(sy_1X_1 + sy_2X_2)\big)\, ds. \tag{15}$$

By applying (1) and part (iv) of Lemma 2.1 to (15) we see that

$$\int_{\Omega_2} \delta(x)q_2(x)\, d\rho(x) \lesssim \frac{1}{(\log L)^2} \quad \text{and} \quad \int_{\Omega_3} \delta(x)q_2(x)\, d\rho(x) \lesssim \frac{1}{\log L}. \tag{16}$$

To estimate q_1, we proceed in a similar way, using the derivative of the heat kernel along the vector field X_0 instead of X_i, $i = 1, 2$, and getting

$$\int_{\Omega_2} \delta(x)q_1(x)\, d\rho(x) \lesssim \frac{1}{r_0} \quad \text{and} \quad \int_{\Omega_3} \delta(x)q_1(x)\, d\rho(x) \lesssim 1. \tag{17}$$

Combining (14), (16), (17), we obtain the estimate $\int \mathscr{M}_h A\, d\rho \lesssim 1$ in the case $r_0 \geqslant 1$, which completes the proof. □

As a consequence of the previous theorem we have that there exists a positive constant C such that, for all f in H^1_{at},

$$\|f\|_{H^1_{\text{max,h}}} = \|\mathscr{M}_h f\|_1 \leqslant C\|f\|_{H^1_{\text{at}}}.$$

Thus $H^1_{\text{at}} \subseteq H^1_{\text{max,h}}$.

3 The Converse Inclusion

In this section we shall construct a family of functions $\{f_L\}_{L>2}$ in H^1_{at} such that

$$\lim_{L\to+\infty}\frac{\|f_L\|_{H^1_{\mathrm{at}}}}{\|\mathscr{M}_h f_L\|_1}=+\infty.$$

Fix $L>2$ and consider the rectangles $R_0=[-1,1]\times[-1,1]\times[\frac{1}{e},e]$ and $R_L=(L,0,1)\cdot R_0=[L-1,L+1]\times[-1,1]\times[\frac{1}{e},e]$. We then define $f_L=\chi_{R_L}-\chi_{R_0}$. Obviously f_L is a multiple of an atom, so it lies in the Hardy space H^1_{at}. We shall estimate the H^1_{at}-norm of the function f_L from below and the $H^1_{\mathrm{max,h}}$-norm of f_L from above.

Lemma 3.1 *There exists a positive constant C such that*

(i) $\|f_L\|_{H^1_{\mathrm{at}}}\geqslant C\log L$ *for every* $L>2$;
(ii) $\|f_L\|_{H^1_{\mathrm{max,h}}}\leqslant C\log\log L$ *for every* $L>2$.

Proof We first prove (i). Let us consider the function h in $\mathrm{BMO}(\mathbb{R})$ given by $h(s)=\log|s|$ for all s in $\mathbb{R}$ and define $g(x_1,x_2,a)=h(x_1)=\log|x_1|$. It is easy to see that g is in BMO and $\|g\|_{\mathrm{BMO}}=\|h\|_{\mathrm{BMO}(\mathbb{R})}$.

Since f_L is a multiple of an atom, $\langle g,f_L\rangle=\int gf_L\,d\rho$, and it is easy to verify that $|\int gf_L\,d\rho|\geqslant C\log L$. On the other hand, $|\langle g,f_L\rangle|\lesssim\|g\|_{\mathrm{BMO}}\|f_L\|_{H^1_{\mathrm{at}}}$. Estimate (i) follows.

The proof of (ii) is very technical, and we sketch it here, skipping many details.

Denote by $2R_0$ the rectangle $[-2,2]\times[-2,2]\times[\frac{1}{e^2},e^2]$ and by $2R_L$ the rectangle $(L,0,1)\cdot(2R_0)$. We shall estimate the L^1-norm of $\mathscr{M}_h f_L$ by integrating it over different regions of the space.

Step 1. The operator $\mathscr{M}_h$ is a contraction on L^∞, so that $\mathscr{M}_h f_L\leqslant 1$, and since $\rho(2R_0)=\rho(2R_L)\sim 1$,

$$\int_{2R_0\cup 2R_L}\mathscr{M}_h f_L\,d\rho\lesssim 1. \tag{18}$$

Step 2. Choose a ball $B=B(e,r_B)$ with $r_B=(\log L)^\alpha$, where $\alpha>2$ is a constant. Then (1) implies that $B\supset 2R_0\cup 2R_L$ if L is large enough. We shall estimate the maximal function on $B\setminus(2R_L\cup 2R_0)$. From (8) we see that, for any x in G,

$$\mathscr{M}_h\chi_{R_0}(x)=\sup_t\int\chi_{R_0}(y)h_t\big(y^{-1}x\big)\,d\lambda(y)\lesssim\sup_{y\in R_0}\sup_t h_t\big(y^{-1}x\big). \tag{19}$$

If $x\in(2R_0)^c$ and $y\in R_0$, then $\delta^{1/2}(y^{-1}x)\sim\delta^{1/2}(x)$ and $|r(y^{-1}x)-r(x)|\leqslant C$. Applying (6) and (9), we have that, for any $x\in(2R_0)^c$,

$$\mathscr{M}_h\chi_{R_0}(x)\lesssim\sup_{y\in R_0}\sup_t h_t\big(y^{-1}x\big)\lesssim\delta^{1/2}(x)\frac{1}{r(x)^2\sinh r(x)}. \tag{20}$$

By applying (4) and (20), we have

$$\int_{B\setminus 2R_0} \mathcal{M}_h \chi_{R_0}\, d\rho \lesssim \int_1^{r_B} r^{-2} \frac{1}{\sinh r} r \sinh r\, dr \lesssim \log\log L. \tag{21}$$

By an easy translation argument, we also obtain

$$\int_{B\setminus 2R_L} \mathcal{M}_h \chi_{R_L}(x)\, d\rho(x) \lesssim \log\log L. \tag{22}$$

We now split the complement of B into the following three regions:

$$\begin{aligned} \Gamma_1 &= \big\{x=(x_1,x_2,a)\in B^c : a<a^*,\, |x|<f(a)\big\},\\ \Gamma_2 &= \big\{x=(x_1,x_2,a)\in B^c : a\geqslant a^*\big\},\\ \Gamma_3 &= \big\{x=(x_1,x_2,a)\in B^c : a<a^*,\, |x|\geqslant f(a)\big\}, \end{aligned} \tag{23}$$

where $a^* = e^{-r_B/8}$ and $f(a) = e^{\sqrt{\log a^{-1}}}$.

Step 3. To estimate the integral of $\mathcal{M}_h f_L$ over the region Γ_1 we use the simple fact that $\mathcal{M}_h f_L \leqslant \mathcal{M}_h \chi_{R_L} + \mathcal{M}_h \chi_{R_0}$. For any point x in Γ_1 we have $\log a^{-1} \lesssim r(x)$. From (20) we obtain

$$\int_{\Gamma_1} \mathcal{M}_h \chi_{R_0}\, d\rho \lesssim \int_0^{a^*} \frac{da}{a} \int_{|x|<f(a)} a^{-1}\big[\log a^{-1}\big]^{-2} \frac{dx}{a^{-1}(1+|x|^2)} \lesssim 1. \tag{24}$$

By a translation argument, we see also that

$$\int_{\Gamma_1} \mathcal{M}_h \chi_{R_L}\, d\rho \lesssim 1. \tag{25}$$

Step 4. In order to estimate the integrals over Γ_2 and Γ_3, we first write the convolution $f_L * h_t(x)$ at a point $x \in B^c$ as follows:

$$\begin{aligned} f_L * h_t(x) &= \int_{R_L} h_t\big(y^{-1}x\big)\, d\lambda(y) - \int_{R_0} h_t\big(y^{-1}x\big)\, d\lambda(y)\\ &= \int_{R_0} \big[h_t\big(y^{-1}(-L,0,1)x\big) - h_t\big(y^{-1}x\big)\big]\, d\lambda(y). \end{aligned}$$

Let now $y^{-1} = (y_1, y_2, b)$ be any point in $(R_0)^{-1}$ and $x = (x_1, x_2, a)$ any point in B^c. Then $y^{-1}(-L,0,1)x = (-bL,0,1)y^{-1}x$, and the Mean Value Theorem implies

$$h_t\big(y^{-1}(-L,0,1)x\big) - h_t\big(y^{-1}x\big) = -bL\partial_1 h_t\big((s,0,1)y^{-1}x\big),$$

for some $s \in (-bL, 0)$. By the fact that $X_1 = a\partial_1$ and the explicit expression for the derivative $X_1 h_t$ given by part (i) of Lemma 2.1, and by some computations,

$$\sup_t \left| h_t\big(y^{-1}(-L,0,1)x\big) - h_t\big(y^{-1}x\big)\right| \lesssim \frac{L^3}{a^2} r(x)^{-2} \frac{1}{\sinh^2 r(x)} \big(L + |x|\big),$$

which allows to conclude that

$$\sup_t \left| f_L * h_t(x)\right| \lesssim \frac{L^3}{a^2} r(x)^{-2} \frac{1}{\sinh^2 r(x)} \big(L + |x|\big). \tag{26}$$

Integrating (26) on $\Gamma_2 \cup \Gamma_3$ we obtain that

$$\int_{\Gamma_2 \cup \Gamma_3} \mathscr{M}_h f_L \, d\rho \lesssim 1. \tag{27}$$

Part (ii) of the lemma is proved by summing up the estimates obtained in Steps 1–4. □

By Lemma 3.1 it follows that

$$\lim_{L \to +\infty} \frac{\| f_L \|_{H^1_{\mathrm{at}}}}{\| f_L \|_{H^1_{\mathrm{max,h}}}} = +\infty.$$

This implies that there does not exist a positive constant C such that $\| f_L \|_{H^1_{\mathrm{at}}} \leqslant C \| f_L \|_{H^1_{\mathrm{max,h}}}$. Thus the atomic Hardy space is strictly contained in the heat maximal Hardy space.

4 The Poisson Maximal Hardy Space

Let p_t denote the Poisson kernel associated with the Laplacian Δ, i.e., the convolution kernel of the semigroup $(e^{-t\sqrt{\Delta}})_{t>0}$, which is given by the subordination formula

$$p_t(x) = \int_0^\infty \frac{t}{2\sqrt{\pi}} s^{-3/2} e^{-t^2/4s} h_s(x)\, ds = \int_0^\infty w_t(s) h_s(x)\, ds \quad \text{for every } x \in G, \tag{28}$$

where the weight function w_t is positive and has integral 1. The explicit formula for the Poisson kernel is then

$$p_t(x) = \frac{1}{\pi^2} \delta^{1/2}(x) \frac{r(x)}{\sinh r(x)} \frac{t}{(t^2 + r^2(x))^2} \quad \text{for every } x \in G. \tag{29}$$

From Theorem 2.3 we deduce the boundedness of the Poisson maximal function from the atomic Hardy space to L^1.

Theorem 4.1 *The Poisson maximal function $\mathscr{M}_p$ is bounded from H^1_{at} to L^1.*

Proof Suppose that A is an atom supported in the Calderón–Zygmund set R. We then have

$$\begin{aligned}\mathscr{M}_p A(x) &= \sup_{t>0}\big|A * p_t(x)\big| \\ &\leqslant \sup_{t>0}\left|\int_0^\infty w_t(s) A * h_s(x)\,ds\right| \\ &\leqslant \sup_{t>0}\int_0^\infty w_t(s)\sup_{s>0}\big|A * h_s(x)\big|\,ds \\ &\leqslant \sup_{t>0}\int_0^\infty w_t(s)\mathscr{M}_h A(x)\,ds \\ &\leqslant \mathscr{M}_h A(x).\end{aligned}$$

This implies that

$$\|\mathscr{M}_p A\|_1 \leqslant \|\mathscr{M}_h A\|_1 \leqslant \kappa,$$

where κ is the constant which appears in (10). This completes the proof. □

As a consequence of the previous theorem we have that there exists a positive constant C such that for all f in H^1_{at}

$$\|f\|_{H^1_{\mathrm{max,p}}} = \|\mathscr{M}_p f\|_1 \leqslant C\|f\|_{H^1_{\mathrm{at}}}.$$

Thus $H^1_{\mathrm{at}} \subseteq H^1_{\mathrm{max,p}}$. In order to discuss the other inclusion, we shall need the observation that for any $r > 0$

$$\sup_{t>0}\frac{t}{(t^2+r^2)^2} \sim r^{-3}, \tag{30}$$

and the following analog of Lemma 2.1.

Lemma 4.2 *For any point $x \in G$ the derivatives of the Poisson kernel p_t along the vector fields X_i, $i = 0, 1, 2$, are the following, where r denotes $r(x)$:*

(i) $X_i p_t(x) = p_t(x)\dfrac{x_i}{\sinh r}\left(\dfrac{1}{r} - \dfrac{\cosh r}{\sinh r} - \dfrac{4r}{t^2+r^2}\right)$ *for $i = 1, 2$;*

(ii) $X_0 p_t(x)$

$$= p_t(x)\left(\frac{a}{r\sinh r} - \frac{\cosh r}{r\sinh r} - \frac{a\cosh r}{\sinh^2 r} + \frac{1}{\sinh^2 r} + \frac{4r}{t^2+r^2}\,\frac{\cosh r - a}{\sinh r}\right);$$

(iii) $\sup_{t>0}\big|X_i p_t(x)\big| \lesssim \dfrac{|x|}{ar^2\cosh^2 r}$ *for $i = 1, 2$ and for all $x \in B_1^c$;*

(iv) $\sup_{t>0}\big|X_0 p_t(x)\big| \lesssim \dfrac{1}{ar^3\cosh r} + \dfrac{1}{r^2\cosh^2 r}$ *for all $x \in B_1^c$.*

Let now $f_L = \chi_{R_L} - \chi_{R_0}$ be the functions defined at the beginning of Sect. 3.

Lemma 4.3 *There exists a positive constant C such that*

$$\|f_L\|_{H^1_{\max,p}} \leqslant C \log\log L \quad \text{for every } L > 2.$$

Proof The proof follows the same outline of the proof of part (ii) of Lemma 3.1. More precisely, the local estimates of $\mathscr{M}_p f_L$ on the sets $2R_0 \cup 2R_L$ and $B = B(e, r_B)$, with $r_B = (\log L)^\alpha$, follow exactly as for the heat maximal function. Indeed, it is enough to apply (29) and (30) to repeat the arguments used for h_t and deduce that

$$\int_B \mathscr{M}_p f_L \, d\rho \lesssim \log\log L.$$

We then split the complement of B into the three regions Γ_j, $j = 1, 2, 3$ defined in (23). The estimate

$$\int_{\Gamma_1} \mathscr{M}_p f_L \, d\rho \lesssim 1$$

follows as in Step 3 of the proof of Lemma 3.1. In order to estimate the integrals over Γ_2 and Γ_3, we first write the convolution $f_L * p_t(x)$ at a point $x \in B^c$ as follows:

$$\begin{aligned} f_L * p_t(x) &= \int_{R_L} p_t\big(y^{-1}x\big)\, d\lambda(y) - \int_{R_0} p_t\big(y^{-1}x\big)\, d\lambda(y) \\ &= \int_{R_0} \big[p_t\big(y^{-1}(-L,0,1)x\big) - p_t\big(y^{-1}x\big)\big]\, d\lambda(y). \end{aligned}$$

Let now $y^{-1} = (y_1, y_2, b)$ be any point in $(R_0)^{-1}$ and $x = (x_1, x_2, a)$ any point in B^c. Then $y^{-1}(-L, 0, 1)x = (-bL, 0, 1)y^{-1}x$, and the Mean Value Theorem implies

$$p_t\big(y^{-1}(-L,0,1)x\big) - p_t\big(y^{-1}x\big) = -bL\partial_1 p_t\big((s,0,1)y^{-1}x\big),$$

for some $s \in (-bL, 0)$. By the fact that $X_1 = a\partial_1$ and the explicit expression for the derivative $X_1 p_t$ given in part (ii) of Lemma 4.2, and by some computations, one obtains

$$\begin{aligned} &\sup_t \big|p_t\big(y^{-1}(-L,0,1)x\big) - p_t\big(y^{-1}x\big)\big| \\ &\quad \lesssim \frac{L}{a^2 b} \frac{1}{r((s,0,1)y^{-1}x)^2 \sinh^2 r((s,0,1)y^{-1}x)} \big(L + |x|\big). \end{aligned}$$

Using the fact that $y \in (R_0)^{-1}$, $x \in B^c$ and $s \in (-bL, 0)$ we deduce that

$$\sup_t \big|p_t\big(y^{-1}(-L,0,1)x\big) - p_t\big(y^{-1}x\big)\big| \lesssim \frac{L^3}{a^2} r(x)^{-2} \frac{1}{\sinh^2 r(x)} \big(L + |x|\big),$$

which allows to conclude that

$$\sup_t \left| f_L * p_t(x) \right| \lesssim \frac{L^3}{a^2} r(x)^{-2} \frac{1}{\sinh^2 r(x)} \big(L + |x|\big). \tag{31}$$

By integrating (31) on $\Gamma_2 \cup \Gamma_3$ we obtain that

$$\int_{\Gamma_2 \cup \Gamma_3} \mathscr{M}_p f_L \, d\rho \lesssim 1.$$

This completes the proof of the lemma. □

By Lemma 4.3 it follows that

$$\lim_{L \to +\infty} \frac{\| f_L \|_{H^1_{\mathrm{at}}}}{\| f_L \|_{H^1_{\mathrm{max,p}}}} = +\infty.$$

This implies that there does not exist a positive constant C such that $\| f_L \|_{H^1_{\mathrm{at}}} \leqslant C \| f_L \|_{H^1_{\mathrm{max,p}}}$. Thus the atomic Hardy space is strictly contained in the Poisson maximal Hardy space.

Remark 4.4 There are some final remarks and open questions that we would like to address:

- We showed that $H^1_{\mathrm{at}} \subsetneq H^1_{\mathrm{max,h}} \subsetneq L^1$ and $H^1_{\mathrm{at}} \subsetneq H^1_{\mathrm{max,p}} \subsetneq L^1$. This implies that both the heat and the Poisson maximal Hardy spaces inherit all nice interpolation properties satisfied by H^1_{at} and L^1.
- How do the Poisson and the heat maximal Hardy spaces relate? Do they coincide?
- In the Euclidean setting a characterization of the Hardy space in terms of Riesz transforms is available. Is there any characterization of the space H^1_{at} in terms of the first order Riesz transforms in our setting?

Acknowledgements Work partially supported by Progetto GNAMPA 2011 "Analisi armonica su gruppi di Lie e varietà".

References

1. Anker, J.-P., Damek, E., Yacoub, C.: Spherical analysis on harmonic *AN* groups. Ann. Sc. Norm. Super. Pisa, Cl. Sci. **23**, 643–679 (1996)
2. Coifman, R.R., Weiss, G.: Analyse harmonique non commutative sur certains espaces homogenes. Lect. Notes in Mathematics, vol. 242. Springer, Berlin (1971)
3. Coifman, R.R., Weiss, G.: Extensions of Hardy spaces and their use in analysis. Bull. Am. Math. Soc. **83**, 569–645 (1977)
4. Cowling, M.G., Gaudry, G., Giulini, S., Mauceri, G.: Weak type (1, 1) estimates for heat kernel maximal functions on Lie groups. Trans. Am. Math. Soc. **323**, 637–649 (1991)
5. Cowling, M.G., Giulini, S., Hulanicki, A., Mauceri, G.: Spectral multipliers for a distinguished Laplacian on certain groups of exponential growth. Stud. Math. **111**, 103–121 (1994)

6. Giulini, S., Sjögren, P.: A note on maximal functions on a solvable Lie group. Arch. Math. **55**, 156–160 (1990)
7. Grafakos, L., Liu, L., Yang, D.: Maximal function characterizations of Hardy spaces on RD-spaces and their applications. Sci. China Ser. A **51**, 2253–2284 (2008)
8. Grafakos, L., Liu, L., Yang, D.: Radial maximal function characterizations for Hardy spaces on RD-spaces. Bull. Soc. Math. Fr. **13**, 225–251 (2009)
9. Hebisch, W., Steger, T.: Multipliers and singular integrals on exponential growth groups. Math. Z. **245**, 37–61 (2003)
10. Li, W.M.: A maximal function characterization of Hardy spaces on spaces of homogeneous type. J. Approx. Theory Appl. **14**, 12–27 (1998)
11. Liu, L., Vallarino, M., Yang, D.: Dyadic sets, maximal functions and applications on ax + b-groups. Math. Z. **270**, 515–529 (2012)
12. Macías, R.A., Segovia, C.: A decomposition into atoms of distributions on spaces of homogeneous type. Adv. Math. **33**, 271–309 (1979)
13. Sjögren, P., Vallarino, M.: Boundedness from H^1 to L^1 of Riesz transforms on a Lie group of exponential growth. Ann. Inst. Fourier **58**, 1117–1151 (2008)
14. Sjögren, P., Vallarino, M.: Heat maximal function on a Lie group of exponential growth. Ann. Acad. Sci. Fenn. **37**, 491–507 (2012)
15. Stein, E.M.: Harmonic Analysis. Princeton University Press, Princeton (1993)
16. Stein, E.M., Weiss, G.: On the theory of harmonic functions of several variables. I. The theory of Hp-spaces. Acta Math. **103**, 25–62 (1960)
17. Tolsa, X.: The space H^1 for nondoubling measures in terms of a grand maximal operator. Trans. Am. Math. Soc. **355**, 315–348 (2003)
18. Vallarino, M.: Spaces H^1 and BMO on exponential growth groups. Collect. Math. **60**, 277–295 (2009)

The Maximal Singular Integral: Estimates in Terms of the Singular Integral

Joan Verdera

Abstract This paper considers estimates of the maximal singular integral $T^\star f$ in terms of the singular integral Tf only. The most basic instance of the estimates we look for is the $L^2(\mathbb{R}^n)$ inequality $\|T^\star f\|_2 \leqslant C\|Tf\|_2$. We present the complete characterization, recently obtained by Mateu, Orobitg, Pérez and the author, of the smooth homogeneous convolution Calderón–Zygmund operators for which such inequality holds. We focus attention on special cases of the general statement to convey the main ideas of the proofs in a transparent way, as free as possible of the technical complications inherent to the general case. Particular attention is devoted to higher Riesz transforms.

Keywords Maximal singular integrals · Calderón–Zygmund operators · Fourier multipliers

Mathematics Subject Classification (2010) Primary 42B20 · Secondary 42B25

1 Introduction

In this expository paper we consider the problem of estimating the Maximal Singular Integral $T^\star f$ only in terms of the Singular Integral Tf. In other words, the function f should appear in the estimates only through Tf. The context is that of classical Calderón–Zygmund theory: we deal with smooth homogeneous convolution singular integral operators of the type

$$Tf(x) = \text{p.v.}\int f(x-y)K(y)\,dy \equiv \lim_{\varepsilon\to 0} T^\varepsilon f(x), \tag{1}$$

where

$$T^\varepsilon f(x) = \int_{|y-x|>\varepsilon} f(x-y)K(y)\,dy$$

J. Verdera (✉)
Departament de Matemàtiques, Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain
e-mail: jvm@mat.uab.cat

M.A. Picardello (ed.), *Trends in Harmonic Analysis*, Springer INdAM Series 3, DOI 10.1007/978-88-470-2853-1_17,

is the truncated integral at level ε. The kernel K is

$$K(x)=\frac{\Omega(x)}{|x|^n}, \quad x\in\mathbb{R}^n\setminus\{0\}, \tag{2}$$

where Ω is a (real valued) homogeneous function of degree 0 whose restriction to the unit sphere S^{n-1} is of class $C^\infty(S^{n-1})$ and satisfies the cancellation property

$$\int_{|x|=1}\Omega(x)\,d\sigma(x)=0, \tag{3}$$

σ being the normalized surface measure on S^{n-1}. The maximal singular integral is

$$T^\star f(x)=\sup_{\varepsilon>0}\left|T^\varepsilon f(x)\right|, \quad x\in\mathbb{R}^n.$$

As we said before, the problem we are envisaging consists in estimating $T^\star f$ in terms of Tf only. Write the maximal Hardy–Littlewood operator as M and consider the well known Cotlar's inequality

$$T^\star f(x)\leqslant C\big(M(Tf)(x)+Mf(x)\big), \quad x\in\mathbb{R}^n. \tag{4}$$

Here this inequality is of no use because it contains the term f besides Tf. The most basic form of the estimate we are looking for is the L^2 inequality

$$\left\|T^\star f\right\|_2\leqslant C\|Tf\|_2, \quad f\in L^2\big(\mathbb{R}^n\big). \tag{5}$$

This problem arose when the author was working at the David–Semmes problem ([2, p. 139, first paragraph]). It was soon discovered [7] that the parity of the kernel plays an essential role. Some years after, a complete characterization of the even operators for which (5) holds was presented in [5] and afterwards the case of odd kernels was solved in [6]. Unfortunately there does not seem to be a way of adapting the techniques of those papers to the Ahlfors regular context in which the David–Semmes problem was formulated.

The proof of the main result in [5] and [6] is long and technically involved. The purpose of this paper is to describe the main steps of the argument in the most transparent possible way. We give complete proofs of particular instances of the main results of the above mentioned papers, so that the reader may grasp, in a simple situation, the idea behind the proof of the general cases. Thus, in a sense, the present paper could serve as an introduction to [5] and [6].

Notice that (5) is true whenever T is a continuous isomorphism of $L^2(\mathbb{R}^n)$ onto itself. Indeed a classical estimate, which follows from Cotlar's inequality, states that

$$\left\|T^\star f\right\|_2\leqslant C\|f\|_2, \quad f\in L^2\big(\mathbb{R}^n\big), \tag{6}$$

which, combined with the assumption that T is an isomorphism, gives (5). Thus (5) is true for the Hilbert Transform and for the Beurling Transform. The first non-trivial case is a scalar Riesz transform in dimension 2 or higher. Denote by $|x|$ the

Euclidean norm of $x \in \mathbb{R}^n$. Recall that the jth Riesz transform is the Calderón–Zygmund operator with kernel

$$\frac{x_j}{|x|^{n+1}}, \quad x \in \mathbb{R}^n \setminus \{0\}, \; 1 \leqslant j \leqslant n.$$

The first non-trivial case for even operators is any second order Riesz transform. For example, the second order Riesz transform with kernel

$$\frac{x_1 x_2}{|x|^{n+2}}, \quad x \in \mathbb{R}^n \setminus \{0\}.$$

We prove the L^2 estimate (5) for the second order Riesz transform above in Sect. 2 and for the jth Riesz transform in Sect. 4. Indeed, in both cases we prove a stronger pointwise estimate which works for all higher Riesz transforms. Recall that a higher Riesz transform is a smooth homogeneous convolution singular integral operator with kernel of the type

$$\frac{P(x)}{|x|^{n+d}}, \quad x \in \mathbb{R}^n \setminus \{0\},$$

where P is a harmonic homogeneous polynomial of degree $d \geqslant 1$. The mean value property of harmonic functions combined with homogeneity yields the cancellation property (3). The following theorem holds [5]:

Theorem 1.1 *If T is an even higher Riesz transform, then*

$$T^\star f(x) \leqslant C M(Tf)(x), \quad x \in \mathbb{R}^n, f \in L^2(\mathbb{R}^n), \tag{7}$$

where M is the maximal Hardy–Littlewood operator.

Indeed, for a second order Riesz transform S one has that the truncation at level ε is a mean of $S(f)$ on a ball. More precisely one has

$$S^\varepsilon(f)(x) = \frac{1}{|B(x,\varepsilon)|} \int_{B(x,\varepsilon)} S(f)(y)\, dy. \tag{8}$$

A weighted variant of the preceding identity works for a general even higher Riesz transform. Of course, (5) for even higher Riesz transforms follows immediately from (7). As we explain in Sect. 3, it turns out that inequality (7) does not hold for odd Riesz transforms, not even for the Hilbert transform. But we can prove the following substitute result [6], which obviously takes care of inequality (5) for odd higher Riesz transforms.

Theorem 1.2 *If T is an odd higher Riesz transform, then*

$$T^\star f(x) \leqslant C M^2(Tf)(x), \quad x \in \mathbb{R}^n, f \in L^2(\mathbb{R}^n), \tag{9}$$

where $M^2 = M \circ M$ is the iteration of the maximal Hardy–Littlewood operator.

Without any harmonicity assumption the L^2 estimate (5) does not hold. The simplest example involves the Beurling transform B, which is the singular integral operator in the plane with complex valued kernel

$$-\frac{1}{\pi}\frac{1}{z^2} = -\frac{1}{\pi}\frac{\bar{z}^2}{|z|^4} = -\frac{1}{\pi}\frac{x^2-y^2}{|z|^4} + \mathrm{i}\frac{1}{\pi}\frac{2xy}{|z|^4}.$$

The Fourier transform of the tempered distribution p. v. $(-\frac{1}{\pi}\frac{1}{z^2})$ is the function $\bar{\xi}/\xi$, hence B is an isometry of $L^2(\mathbb{R}^2)$ onto itself. It turns out that the singular integral

$$T = B + B^2 = B(I + B)$$

does not satisfy the L^2 control (5). The reason for that, as we shall see later on in this section, is that the operator $I + B$ is not invertible in $L^2(\mathbb{R}^2)$.

One way to explain the difference between the even and odd cases is as follows. Theorem 1.1 concerns an even higher Riesz transform determined by a harmonic homogeneous polynomial of degree, say, d. In its proof one is lead to consider the operator $(-\triangle)^{d/2}$, which is a differential operator. Instead, in Theorem 1.2, d is odd and thus $(-\triangle)^{d/2}$ is only a pseudo-differential operator. The effect of this is that in the odd case certain functions are not compactly supported and are not bounded. Nevertheless, they still satisfy a BMO condition, which is the key fact in obtaining the second iteration of the maximal operator.

The search for a description of those singular integrals T of a given parity for which (5) holds begun just after [7] was published. The final answer was given in [5] and [6]. To state the result denote by A the Calderón–Zygmund algebra consisting of the operators of the form $\lambda I + T$, where T is a smooth homogeneous convolution singular integral operator and λ a real number.

Theorem 1.3 *Let T be an even smooth homogeneous convolution singular integral operator with kernel $\Omega(x)/|x|^n$. Then the following are equivalent.*

(i)

$$T^\star f(x) \leqslant C M(Tf)(x), \quad x \in \mathbb{R}^n,\ f \in L^2(\mathbb{R}^n),$$

where M is the Hardy–Littlewood maximal operator.

(ii)

$$\int |T^\star f|^2 \leqslant C \int |Tf|^2, \quad f \in L^2(\mathbb{R}^n).$$

(iii) *If the spherical harmonics expansion of Ω is*

$$\Omega(x) = P_2(x) + P_4(x) + \cdots, \quad |x| = 1,$$

then there exist an even harmonic homogeneous polynomial P of degree d, such that P divides P_{2j} (in the ring of all polynomials in n variables with real coefficients) for all j, $T = R_P \circ U$, where R_P is the higher Riesz transform with *kernel* $P(x)/|x|^{n+d}$*, and* U *is an invertible operator in the Calderón–Zygmund algebra* A.

Several remarks are in order. First, it is surprising that the L^2 control we are looking for, that is, condition (ii) above, is equivalent to the apparently much stronger pointwise inequality (i). We do not know any proof of this fact which does not go through the structural condition (iii). Second, condition (iii) on the spherical harmonics expansion of Ω is purely algebraic and easy to check in practice on the Fourier transform side. Observe that if condition (iii) is satisfied, then the polynomial P must be a scalar multiple of the first non-zero spherical harmonic P_{2j} in the expansion of Ω. We illustrate this with an example.

Example 1.4 Let $P(x,y) = -\frac{1}{\pi}xy$ and denote by R_P the second order Riesz transform in the plane associated with the harmonic homogeneous polynomial P. Its kernel is

$$-\frac{1}{\pi}\frac{xy}{|z|^4}, \quad z = x+iy \in \mathbb{C}\setminus\{0\}. \tag{10}$$

According to a well known formula [9, p. 73] the Fourier transform of the principal value distribution associated with this kernel is

$$\frac{uv}{|\xi|^2}, \quad \xi = u+iv \in \mathbb{C}\setminus\{0\}.$$

This is also the symbol (or Fourier multiplier) of R_P, in the sense that

$$\widehat{R_P(f)}(\xi) = \frac{uv}{|\xi|^2}\hat{f}(\xi), \quad \xi \neq 0,\ f \in L^2(\mathbb{R}^n).$$

Similarly, the Fourier multiplier of the fourth order Riesz transform with kernel

$$\frac{2}{\pi}\frac{x^3y - xy^3}{|z|^6}, \quad z \neq 0,$$

is

$$\frac{u^3v - uv^3}{|\xi|^4}, \quad \xi \neq 0.$$

Given a real number λ let T be the singular integral with kernel

$$-\frac{1}{\pi}\frac{2xy}{|z|^4} + \lambda\frac{2}{\pi}\frac{x^3y - xy^3}{|z|^6}.$$

Its symbol is

$$\frac{uv}{|\xi|^2}\left(1 + \lambda\frac{u^2 - v^2}{|\xi|^2}\right).$$

We clearly have

$$T = R_P \circ U,$$

U being the bounded operator on $L^2(\mathbb{R}^n)$ with symbol $1 + \lambda \frac{u^2 - v^2}{|\xi|^2}$. Notice that the multiplier $1 + \lambda \frac{u^2 - v^2}{|\xi|^2}$ vanishes at some point of the unit sphere if and only if $|\lambda| \geqslant 1$. Therefore condition (iii) of Theorem 1.3 is satisfied if and only if $|\lambda| < 1$. For instance, taking $\lambda = 1$ one gets an operator for which neither the L^2 estimate (ii) nor the pointwise inequality (i) hold.

To grasp the subtlety of the division condition in (iii) it is instructive to consider the special case of the plane. The function Ω, which is real, has a Fourier series expansion

$$\begin{aligned}\Omega\left(e^{i\theta}\right) = \sum_{n=-\infty}^{\infty} c_n\, e^{in\theta} &= \sum_{n=1}^{\infty} c_n\, e^{in\theta} + \overline{c_n}\, e^{-in\theta} \\ &= \sum_{n=1}^{\infty} 2\,\mathrm{Re}\left(c_n\, e^{in\theta}\right).\end{aligned}$$

The expression $2\,\mathrm{Re}(c_n\, e^{in\theta})$ is the general form of the restriction to the unit circle of a harmonic homogeneous polynomial of degree n on the plane. There are exactly $2n$ zeroes of $2\,\mathrm{Re}(c_n\, e^{in\theta})$ on the circle, which are uniformly distributed. They are the $2n$th roots of unity if and only if c_n is purely imaginary.

Since Ω is even, only the Fourier coefficients with even index may be non-zero and so

$$\Omega\left(e^{i\theta}\right) = \sum_{n=1}^{\infty} 2\,\mathrm{Re}\left(c_{2n}\, e^{i2n\theta}\right).$$

Replacing θ by $\theta + \alpha$ we obtain

$$\Omega\left(e^{i(\theta+\alpha)}\right) = \sum_{n=N}^{\infty} 2\,\mathrm{Re}\left(c_{2n}\, e^{i2n\alpha}\, e^{i2n\theta}\right),$$

where $c_{2N} \neq 0$. Take α so that $c_{2N}\, e^{i2N\alpha}$ is purely imaginary. Set $\gamma_{2n} = c_{2n}\, e^{i2n\alpha}$. Then

$$\Omega\left(e^{i(\theta+\alpha)}\right) = \sum_{n=N}^{\infty} 2\,\mathrm{Re}\left(\gamma_{2n}\, e^{i2n\theta}\right).$$

If $\mathrm{Re}(\gamma_{2N}\, e^{i2N\theta})$ divides $\mathrm{Re}(\gamma_{2n}\, e^{i2n\theta})$, then, for some positive integer k,

$$k\frac{\pi}{4n} = \frac{\pi}{4N},$$

or $n = kN$. This means that only the Fourier coefficients whose index is a multiple of $2N$ may be non-zero:

$$\Omega\left(e^{i(\theta+\alpha)}\right) = \sum_{p=1}^{\infty} 2\,\mathrm{Re}\left(\gamma_{2Np}\, e^{i2Np\theta}\right).$$

Moreover γ_{2Np} must be purely imaginary, that is, $\gamma_{2Np} = r_{2Np} i$, with r_{2Np} real. Replacing $\theta + \alpha$ by θ we get

$$\begin{aligned}\Omega\left(e^{i\theta}\right) &= \sum_{p=1}^{\infty} 2\,\mathrm{Re}\left(r_{2Np} i\, e^{-i2Np\alpha}\, e^{i2Np\theta}\right),\\ &= \sum_{p=1}^{\infty} r_{2Np} i\, e^{-i2Np\alpha}\, e^{i2Np\theta} - r_{2Np} i\, e^{i2Np\alpha}\, e^{-i2Np\theta}.\end{aligned}$$

As well known, the sequence of the r_{2Np} $(p = 1, 2, \dots)$ is rapidly decreasing, because $\Omega(e^{i\theta})$ is infinitely differentiable. Therefore the division property in condition (iii) of Theorem 1.1 can be reformulated as a statement about the arguments and the support of the Fourier coefficients of $\Omega(e^{i\theta})$.

For odd operators the statement of Theorem 1.3 must be slightly modified [6].

Theorem 1.5 *Let T be an odd smooth homogeneous convolution singular integral operator with kernel $\Omega(x)/|x|^n$. Then the following are equivalent.*

(i)

$$T^\star f(x) \leqslant C M^2(Tf)(x), \quad x \in \mathbb{R}^n,\ f \in L^2\left(\mathbb{R}^n\right),$$

$M^2 = M \circ M$ being the iterated Hardy–Littlewood maximal operator.

(ii)

$$\int \left|T^\star f\right|^2 \leqslant C \int |Tf|^2, \quad f \in L^2\left(\mathbb{R}^n\right).$$

(iii) *If the spherical harmonics expansion of Ω is*

$$\Omega(x) = P_1(x) + P_3(x) + \cdots, \quad |x| = 1,$$

then there exist an odd harmonic homogeneous polynomial P of degree d, such that P divides P_{2j+1} (in the ring of all polynomials in n variables with real coefficients) for all j, $T = R_P \circ U$, where R_P is the higher Riesz transform with kernel $P(x)/|x|^{n+d}$, and U is an invertible operator in the Calderón–Zygmund algebra A.

Sections 2 and 4 contain, respectively, the proofs of Theorems 1.1 and 1.2 for the most simple kernels. In Sect. 3 we show that the Hilbert transform does not satisfy the pointwise inequality (7). In Sect. 5 we prove that condition (iii) in Theorem 1.3 is necessary and in Sect. 6 that it is sufficient, in both cases in particularly simple situations. Section 7 contains brief comments on the proof of the general case and some open problems.

2 Proof of Theorem 1.1 for Second Order Riesz Transforms

For the sake of clarity we work only with the second order Riesz transform T with kernel

$$\frac{x_1 x_2}{|x|^{n+2}}.$$

The inequality to be proven, namely (7), is invariant by translations and by dilations, so that we only need to show that

$$\left|T^1 f(0)\right| \leqslant C M(T f)(0), \tag{11}$$

where

$$T^1 f(0) = \int_{\mathbb{R}^n \setminus B} \frac{x_1 x_2}{|x|^{n+2}} f(x)\, dx$$

is the truncation at level 1 at the origin. Here B is the unit (closed) ball centered at the origin. A natural way to prove (11) is to find a function b such that

$$\chi_{\mathbb{R}^n \setminus B}(x) \frac{x_1 x_2}{|x|^{n+2}} = T(b).$$

One should keep in mind that T is injective but not onto. Then there is no reason whatsoever for such a b to exist. If such a b exists then

$$T^1 f(0) = \int T b(x) f(x)\, dx = \int b(x) T(f)(x)\, dx. \tag{12}$$

Moreover, if b is in $L^\infty(\mathbb{R}^n)$ and is supported on B, it follows that

$$\left|T_1 f(0)\right| \leqslant \|b\|_\infty |B| \frac{1}{|B|} \int_B \left|T(f)(x)\right| dx \leqslant C M\big(T(f)\big)(0).$$

Thus everything has been reduced to the following lemma.

Lemma 2.1 *There exists a bounded measurable function b supported on B such that*

$$\chi_{\mathbb{R}^n \setminus B}(x) \frac{x_1 x_2}{|x|^{n+2}} = T(b)(x), \quad \textit{for almost all } x \in \mathbb{R}^n.$$

Proof Let E be the standard fundamental solution of the Laplacian in $\mathbb{R}^n$. Then, for some dimensional constant c_n (that is, depending only on n), in the distribution sense we have

$$\partial_1 \partial_2 E = c_n \,\text{p. v.}\, \frac{x_1 x_2}{|x|^{n+2}}. \tag{13}$$

Let us define a function φ by

$$\varphi(x) = \begin{cases} E(x) & \text{on } \mathbb{R}^n \setminus B, \\ A_0 + A_1|x|^2 & \text{on } B \end{cases} \tag{14}$$

where the constants A_0 and A_1 are chosen so that φ and $\nabla\varphi$ are continuous on $\mathbb{R}^n$. This is possible because, for each i,

$$\partial_i \varphi(x) = \begin{cases} c_n x_i/|x|^n, & x \in \mathbb{R}^n \setminus B, \\ 2A_1 x_i, & x \in B \end{cases}$$

and so, for an appropriate choice of A_1, the above two expressions coincide on ∂B for all i, or, equivalently, $\nabla\varphi$ is continuous. The continuity of φ is now just a matter of choosing A_0 so that $E(x) = A_0 + A_1|x|^2$ on ∂B, which is possible because E is radial.

The continuity of φ and $\nabla\varphi$ guarantees that we can compute a second order derivative of φ in the distribution sense by just computing it pointwise on B and on $\mathbb{R}^n \setminus B$. The reason is that no boundary terms will appear when applying the Green–Stokes theorem to compute the action of the second order derivative of φ on a test function. Therefore

$$\triangle\varphi = 2nA_1\chi_B \equiv b,$$

where the last identity is the definition of b. Since $\varphi = E * \Delta\varphi$ we obtain, for some dimensional constant c_n,

$$\partial_1\partial_2\varphi = \partial_1\partial_2 E * \Delta\varphi = c_n \, \text{p. v.} \, \frac{x_1 x_2}{|x|^{n+2}} * \Delta\varphi = c_n T(b).$$

On the other hand, by (14) and noticing that $\partial_1\partial_2|x|^2 = 0$, we have

$$\partial_1\partial_2\varphi = \chi_{\mathbb{R}^n\setminus B}(x) c_n \frac{x_1 x_2}{|x|^{n+2}},$$

and the proof of Lemma 2.1 is complete. □

Notice that (12) together with the special form of the function b found in the proof of Lemma 2.1 yields formula (8), namely, that a truncation at level ε at a point x of $S(f)$, S being a second order Riesz transform, is the mean of $S(f)$ on the ball $B(x, \varepsilon)$.

3 The Pointwise Control of $T^\star$ by $M \circ T$ Fails for the Hilbert Transform

We now show that, if H is the Hilbert transform, the following inequality *fails*:

$$H^\star f(x) \leqslant CM(Hf)(x), \quad x \in \mathbb{R}, \, f \in L^2(\mathbb{R}). \tag{15}$$

Replacing f by $H(f)$ in (15) and recalling that $H(Hf)=-f$, $f\in L^2(\mathbb{R})$, we see that (15) is equivalent to

$$H^\star\big(H(f)\big)(x)\leqslant CM(f)(x),\quad x\in\mathbb{R},\, f\in L^2(\mathbb{R}).$$

It turns out that the operator $H^\star\circ H$ is not of weak type (1, 1).

Let us prove that, if $f=\chi_{(0,1)}$, then there are positive constants m and C such that, whenever $x>m$,

$$H^\star(Hf)(x)\geqslant C\frac{\log x}{x}. \tag{16}$$

This shows that $H^\star\circ H$ is not of weak type (1, 1). Indeed, choosing $m>e$ if necessary, we have

$$\begin{aligned}\sup_{\lambda>0}\lambda\big|\{x\in\mathbb{R}:H^\star(Hf)(x)>\lambda\}\big| &\geqslant \sup_{\lambda>0}\lambda\left|\left\{x>m:\frac{\log x}{x}>C^{-1}\lambda\right\}\right|\\ &= C\sup_{\lambda>0}\lambda\left|\left\{x>m:\frac{\log x}{x}>\lambda\right\}\right|\\ &\geqslant C\sup_{\lambda>0}\lambda\big(\varphi^{-1}(\lambda)-e\big),\end{aligned}$$

where the decreasing function $\varphi:(e,\infty)\to(0,e^{-1})$ is given by $\varphi(x)=\frac{\log x}{x}$. To conclude, observe that the right hand side of the estimate is unbounded as $\lambda\to 0$:

$$\lim_{\lambda\to 0}\lambda\varphi^{-1}(\lambda)=\lim_{\lambda\to\infty}\varphi(\lambda)\lambda=\infty.$$

To prove (16) we recall that, for $f=\chi_{(0,1)}$,

$$Hf(y)=\log\frac{|y|}{|y-1|}.$$

Let $m>1$ big enough to be chosen later on. Take $x>m$. By definition of $H^\star$

$$H^\star(Hf)(x)\geqslant\left|\int_{|y-x|>m+x}\frac{1}{y-x}\log\frac{|y|}{|y-1|}\,dy\right|$$

and the integral splits in the obvious way

$$\begin{aligned}&\int_{-\infty}^{-m}\frac{1}{y-x}\log\frac{-y}{-y+1}\,dy+\int_{2x+m}^{\infty}\frac{1}{y-x}\log\frac{y}{y-1}\,dy\\ &\quad=\int_m^{\infty}\frac{1}{x+y}\log\frac{y+1}{y}\,dy+\int_{2x+m}^{\infty}\frac{1}{y-x}\log\frac{y}{y-1}\,dy=A(x)+B(x),\end{aligned}$$

where both $A(x)$ and $B(x)$ are positive. Hence

$$H^\star(Hf)(x)\geqslant A(x).$$

Since

$$\log\left(1+\frac{1}{y}\right) \approx \frac{1}{y} \quad \text{as } y \to \infty,$$

there is a constant $m > 1$ such that, whenever $y > m$,

$$\frac{1}{2} < \frac{\log(1+\frac{1}{y})}{\frac{1}{y}} < \frac{3}{2}.$$

Hence, for this constant m we have

$$A(x) = \int_m^\infty \frac{1}{x+y} \log\left(1+\frac{1}{y}\right) dy \approx \int_m^\infty \frac{1}{x+y} \frac{dy}{y} = \frac{1}{x} \log \frac{y}{x+y}\Big|_m^\infty \approx \frac{\log x}{x},$$

which proves (16).

Notice that the term $B(x)$ is better behaved:

$$B(x) \leqslant \int_{2x+m}^\infty \frac{1}{y-x} \log \frac{y}{y-1}\, dy \leqslant \int_{2x+m}^\infty \frac{2}{y} \frac{dy}{y} \leqslant \frac{1}{x}.$$

4 Proof of Theorem 1.2 for First Order Riesz Transforms

In this section we prove that

$$R_j^\star(f)(x) \leqslant C M^2\big(R_j(f)\big), \quad x \in \mathbb{R}^n, \tag{17}$$

where R_j is the jth Riesz transform, namely, the Calderón–Zygmund operator with kernel

$$\frac{x_j}{|x|^{n+1}}, \quad x \in \mathbb{R}^n \setminus \{0\}, 1 \leqslant j \leqslant n.$$

Recall that $M^2 = M \circ M$ and notice that for $n = 1$ we are dealing with the Hilbert transform. Inequality (17) for the Hilbert transform is, as far as we know, new. To have a glimpse at the difficulties we will encounter in proving (17) we start by discussing the case of the Hilbert transform.

As in the even case we want to find a function b such that

$$\frac{1}{x} \chi_{\mathbb{R}\setminus(-1,1)}(x) = H(b).$$

Since $H(-H) = I$,

$$\begin{aligned} b(x) &= -H\left(\frac{1}{y}\chi_{\mathbb{R}\setminus(-1,1)}(y)\right)(x) \\ &= \frac{1}{\pi}\int_{|y|>1} \frac{1}{y-x}\frac{1}{y}\,dy \\ &= \frac{1}{\pi x}\log\frac{|1+x|}{|1-x|}. \end{aligned}$$

We conclude that, unlike the even case, the function b is unbounded and is not supported in the unit interval $(-1, 1)$. On the positive side, we see that b is a function in BMO = BMO($\mathbb{R}$), the space of function of bounded mean oscillation on the line. Since b decays at infinity as $1/x^2$, b is integrable on the whole line. However, the minimal decreasing majorant of the absolute value of b is not integrable, due to the poles at ± 1. This prevents a pointwise estimate of $H^\star f$ by a constant times $M(Hf)$. We can now proceed with the proof of (17) keeping in mind the kind of difficulties we shall have to overcome.

We start with the analog of Lemma 2.1. We denote by BMO the space of functions of bounded mean oscillation on $\mathbb{R}^n$.

Lemma 4.1 *There exists a function $b \in$ BMO such that*

$$\chi_{\mathbb{R}^n\setminus B}(x)\frac{x_j}{|x|^{n+1}} = R_j(b)(x), \quad \textit{for almost all } x \in \mathbb{R}^n, 1 \leqslant j \leqslant n. \tag{18}$$

Proof For an appropriate constant c_n the function

$$E(x) = c_n\frac{1}{|x|^{n-1}}, \quad 0 \neq x \in \mathbb{R}^n$$

satisfies

$$\widehat{E}(\xi) = \frac{1}{|\xi|}, \quad 0 \neq \xi \in \mathbb{R}^n.$$

Since the pseudo-differential operator $(-\Delta)^{1/2}$ is defined on the Fourier transform side as

$$\widehat{(-\Delta)^{1/2}\psi}(\xi) = |\xi|\hat{\psi}(\xi),$$

E may be understood as a fundamental solution of $(-\Delta)^{1/2}$. This will allow to structure our proof in complete analogy to that of Lemma 2.1 until new facts emerge. Consider the function φ that takes the value c_n on B and $E(x)$ on $\mathbb{R}^n \setminus B$. We have that $\varphi = E * (-\Delta)^{1/2}\varphi$ and we define b as $(-\Delta)^{1/2}\varphi$.

As it is well known,

$$\partial_j E = -(n-1)c_n\,\text{p.v.}\,\frac{x_j}{|x|^{n+1}},$$

in the distribution sense and, since φ is continuous on the boundary of B,

$$\partial_j \varphi = -(n-1)c_n \chi_{\mathbb{R}^n \setminus B}(x) \frac{x_j}{|x|^{n+1}} \tag{19}$$

also in the distribution sense. Then

$$\begin{aligned} -(n-1)c_n \chi_{\mathbb{R}^n \setminus B}(x) \frac{x_j}{|x|^{n+1}} &= \partial_j \varphi \\ &= \partial_j E * b \\ &= -(n-1)c_n \,\text{p. v.}\, \frac{x_j}{|x|^{n+1}} * b, \end{aligned}$$

which is (18). It remains to show that $b \in \mathrm{BMO}$.

Checking on the Fourier transform side we easily see that

$$b = (-\Delta)^{1/2} \varphi = \gamma_n \sum_{k=1}^{n} R_k(\partial_k \varphi), \tag{20}$$

for some dimensional constant γ_n. Since $\partial_k \varphi$ is a bounded function by (19) and R_k maps L^∞ into BMO, b is in BMO and the proof is complete. □

Unfortunately b is not bounded and is not supported on $\mathbb{R}^n \setminus B$. Moreover one can easily check that b blows up at the boundary of B as the function $\log(1/|1-|x||)$. This entails that the minimal decreasing majorant of the absolute value of b is not integrable, as in the one dimensional case.

We now take up the proof of (17). By translation and dilation invariance we only have to estimate the truncation of $R_j f$ at the point $x = 0$ and at level $\varepsilon = 1$. By Lemma 4.1

$$\begin{aligned} R_j^1 f(0) &= -\int \chi_{\mathbb{R}^n \setminus B}(x) \frac{x_j}{|x|^{n+1}} f(x)\,dx = -\int R_j b(x) f(x)\,dx \\ &= \int b(x) R_j f(x)\,dx. \end{aligned}$$

Let b_{2B} denote the mean of b on the ball $2B$. We split the last integral above into three pieces

$$\begin{aligned} R_j^1 f(0) &= \int_{2B} \big(b(x) - b_{2B}\big) R_j f(x)\,dx + b_{2B} \int_{2B} R_j f(x)\,dx \\ &\quad + \int_{\mathbb{R}^n \setminus 2B} b(x) R_j f(x)\,dx \\ &= I_1 + I_2 + I_3. \end{aligned}$$

Since b_{2B} is a dimensional constant the term I_2 can be immediately estimated by $CM(R_j f)(0)$. The term I_3 can easily be estimated if we first prove that

$$\left|b(x)\right| \leqslant C \frac{1}{|x|^{n+1}}, \quad |x| \geqslant 2. \tag{21}$$

Indeed, this decay estimate yields

$$|I_3| \leqslant C \int_{\mathbb{R}^n \setminus 2B} |R_j f(x)| \frac{1}{|x|^{n+1}}\, dx \leqslant CM(R_j f)(0).$$

To prove (21), let us express b by means of (20):

$$\begin{aligned}
\frac{b}{\gamma_n} &= \sum_{k=1}^{n} R_k * \chi_{\mathbb{R}^n \setminus B}(x) \frac{x_k}{|x|^{n+1}} = \sum_{k=1}^{n} R_k * R_k - \sum_{k=1}^{n} R_k * \chi_B(x) \frac{x_k}{|x|^{n+1}} \\
&= \gamma_n' \delta_0 - \sum_{k=1}^{n} R_k \left(\chi_B(x) \frac{x_k}{|x|^{n+1}} \right),
\end{aligned}$$

where γ_n' is a dimensional constant and δ_0 the Dirac delta at the origin. The preceding formula for b looks magical and one may even think that some terms make no sense. For instance, the term $R_k * R_k$ should not be thought as the action of the kth Riesz transform of the distribution p. v. $x_k/|x|^{n+1}$. It is more convenient to look at it on the Fourier transform side, where you see immediately that it is $\gamma_n' \delta_0$. The term $R_k * (\chi_B(x) x_k/|x|^{n+1})$ should be thought as a distribution, which acts on a test function as one would expect via principal values (see below).

If $|x| > 1$ we have

$$\begin{aligned}
R_k \left(\chi_B(x) \frac{x_k}{|x|^{n+1}} \right)(x) &= \lim_{\varepsilon \to 0} \int_{\varepsilon < |y| < 1} \frac{x_k - y_k}{|x - y|^{n+1}} \frac{y_k}{|y|^{n+1}}\, dy \\
&= \lim_{\varepsilon \to 0} \int_{\varepsilon < |y| < 1} \left(\frac{x_k - y_k}{|x - y|^{n+1}} - \frac{x_k}{|x|^{n+1}} \right) \frac{y_k}{|y|^{n+1}}\, dy.
\end{aligned}$$

Since

$$\left| \frac{x_k - y_k}{|x - y|^{n+1}} - \frac{x_k}{|x|^{n+1}} \right| \leqslant C \frac{|y|}{|x|^{n+1}} \quad \text{for every } |x| \geqslant 2,\ |y| \leqslant 1,$$

we obtain, for $|x| \geqslant 2$,

$$\left| R_k \left(\chi_B(x) \frac{x_k}{|x|^{n+1}} \right)(x) \right| \leqslant C \int_{|y|<1} \frac{1}{|x|^{n+1}} \frac{1}{|y|^{n-1}}\, dy = \frac{C}{|x|^{n+1}},$$

which gives (21).

We are left with the term I_1. Since b is in BMO, it is exponentially integrable by John–Nirenberg's theorem. We estimate I_1 by Hölder's inequality associated with the "dual" Young functions $e^t - 1$ and $t + t\log^+ t$ [4, p. 165]. We obtain

$$|I_1| \leqslant C\|b\|_{\mathrm{BMO}}\|R_j f\|_{L\log L(2B)},$$

where, for an integrable function g on $2B$,

$$\|g\|_{L\log L(2B)} = \inf\left\{\lambda > 0 : \frac{1}{|2B|}\int_{2B}\left(\frac{|g(x)|}{\lambda} + \frac{|g(x)|}{\lambda}\log^+\left(\frac{|g(x)|}{\lambda}\right)\right)dx \leqslant 1\right\}.$$

Consider the maximal operator associated with $L\log L$, that is,

$$M_{L(\log L)}g(x) = \sup_{Q\ni x}\|f\|_{L(\log L),Q},$$

the supremum being taking over all balls Q. It is a nice fact (see [8] or [4, p. 159]) that M satisfies

$$M_{L(\log L)}f(x) \approx M^2 f(x), \quad x \in \mathbb{R}^n. \tag{22}$$

Thus

$$|I_1| \leqslant CM^2(R_j f)(0)$$

and the proof of (17) is complete.

5 Necessary Conditions for the L^2 Estimate of $T^\star f$ by Tf

In this section we find necessary conditions for the L^2 estimate

$$\left\|T^\star f\right\|_2 \leqslant C\|Tf\|_2, \quad f \in L^2\left(\mathbb{R}^n\right) \tag{23}$$

which are stated in part (iii) of Theorem 1.3 for the case of even kernels. In particular, this will supply many even kernels for which this estimate fails (and thus the pointwise estimate in part (i) of Theorem 1.3 also fails).

We shall look at the simplest possible case. The kernel of our operator T in the plane is of the form

$$K(z) = -\frac{1}{\pi}\frac{xy}{|z|^4} + \frac{2}{\pi}\frac{P_4(z)}{|z|^6}, \tag{24}$$

where $z = x + iy$ is the complex variable in the plane $\mathbb{C}$ and P_4 is a harmonic homogeneous polynomial of degree 4. The constants in front of the two terms are set so that the expression of the Fourier multiplier is the simplest. Indeed, by [9, p. 73], the Fourier transform of the principal value tempered distribution associated with K is

$$\widehat{\mathrm{p.v.}\ K}(\xi) = \frac{uv}{|\xi|^2} + \frac{P_4(\xi)}{|\xi|^4}, \quad 0 \neq \xi = u + iv \in \mathbb{C}.$$

Our purpose is to find necessary conditions on P_4 so that the L^2 estimate (23) holds. Notice that the kernel K is not harmonic, except in the case $P_4 = 0$ which we ignore. The spherical harmonics expansion of K is reduced to the sum of the two terms in (24).

Let E be the standard fundamental solution of the bi-Laplacian Δ^2 in the plane. Thus

$$E(z) = \frac{1}{8\pi}|z|^2 \log|z|, \quad 0 \neq z \in \mathbb{C},$$

and $\widehat{E}(\xi) = |\xi|^{-4}$, $0 \neq \xi \in \mathbb{C}$. We have

$$\big(\partial_1\partial_2\triangle + P_4(\partial_1,\partial_2)\big)(E) = \text{p. v. } K$$

as one easily checks on the Fourier transform side. Here we adopt the usual convention of denoting by $P_4(\partial_1,\partial_2)$ the differential operator obtained by replacing the variables x and y of P_4 by ∂_1 and ∂_2 respectively.

Define a function φ by

$$\varphi(z) = \begin{cases} E(z) & \text{on } \mathbb{C}\setminus B \\ A_0 + A_1|z|^2 + A_2|z|^4 + A_3|z|^6 & \text{on } B \end{cases}$$

where B is the ball centered at the origin of radius 1. The constants A_j, $0 \leqslant j \leqslant 3$, are chosen so that all derivatives of φ of order not greater than 3 are continuous. This can be done because E is radial. With this choice, in order to compute a fourth order derivative of φ in the distribution sense we only need to compute the corresponding pointwise derivative of φ in B and on its complement. Set $b = \Delta^2\varphi$, so that

$$\varphi = E * \Delta^2\varphi = E * b.$$

A straightforward computation yields

$$b = \Delta^2\varphi = \chi_B(z)\big(\alpha + \beta|z|^2\big),$$

for some constants α and β. Then, as in the proof of the L^2 estimate (23) for even second order Riesz transforms presented in Sect. 2, b is supported on the ball B and is bounded. Set

$$L = \partial_1\partial_2\triangle + P_4(\partial_1,\partial_2),$$

so that

$$L(\varphi) = L(E) * b = \text{p. v. } K * b = T(b).$$

On the other hand, by the definition of φ,

$$L(\varphi) = \chi_{\mathbb{C}\setminus B}(z)K(z) + L\big(A_0 + A_1|z|^2 + A_2|z|^4 + A_3|z|^6\big)\chi_B(z).$$

The term $L(A_0 + A_1|z|^2 + A_2|z|^4 + A_3|z|^6)$ does not vanish. Indeed, one can see that for some constant c

$$L\big(A_0 + A_1|z|^2 + A_2|z|^4 + A_3|z|^6\big) = cxy.$$

The result follows from the following three facts:

$$(\partial_1\partial_2\triangle)\big(|z|^4\big) = 0,$$
$$(\partial_1\partial_2\triangle)\big(|z|^6\big) = cxy$$

and

$$P_4\big(|z|^4\big) = P_4\big(|z|^6\big) = 0.$$

The last identity is due to the fact that P_4 is a homogeneous harmonic polynomial of degree 4. Notice that a priori $P_4(|z|^4)$ is a constant and $P_4(|z|^6)$ is a homogeneous polynomial of degree 2. The reader can verify that they are both zero just by taking the Fourier transform and then checking their action on a test function.

The conclusion is that

$$T(b) = \chi_{\mathbb{C}\setminus B}(z)K(z) + cxy\chi_B(z). \tag{25}$$

The novelty with respect to the argument of Sect. 2 involving second order Riesz transforms is the second term in the right hand side of the preceding formula. Convolving (25) with a function f in $L^2(\mathbb{C})$ one gets

$$cxy\chi_B(z) * f = T(f) * b - T^1(f),$$

where $T^1 f$ is the truncation at level 1. Now, if (23) holds then, since $b \in L^1(\mathbb{C})$,

$$\big\|cxy\chi_B(z) * f\big\|_2 \leqslant C\big\|T(f)\big\|_2, \quad f \in L^2(\mathbb{C}),$$

hence, passing to the multipliers,

$$\big|\widehat{cxy\chi_B(z)}(\xi)\big| \leqslant C\frac{|uv|\xi|^2 + P_4(\xi)|}{|\xi|^4}, \quad \xi \neq 0. \tag{26}$$

Our next task is to understand the left hand side of the above inequality to obtain useful relations between the zero sets of the various polynomials at hand. We should recall that the Fourier transform of the characteristic function of the unit ball in $\mathbb{R}^2$ is $J_1(\xi)/|\xi|$, where $J_1(\xi)$ is the Bessel function of order 1. Write $G_m(\xi) = J_m(\xi)/|\xi|^m$. The functions G_m are radial and so we can view them as depending on a non-negative real variable r. We have [3, p. 425] the useful identity

$$\frac{1}{r}\frac{dG_m}{dr}(r) = -G_{m+1}(r), \quad 0 \leqslant r.$$

From this it is easy to obtain the formula

$$\begin{aligned}\widehat{xy\chi_B(z)}(\xi) &= -\partial_1\partial_2\big(G_1\big(|\xi|\big)\big)\\ &= -uvG_3\big(|\xi|\big),\end{aligned}$$

which transforms (26) into

$$\big|uvG_3\big(|\xi|\big)\big| \leqslant C\frac{|uv|\xi|^2 + P_4(\xi)|}{|\xi|^4}, \quad \xi \neq 0. \tag{27}$$

Set

$$Q(\xi) = uv|\xi|^2 + P_4(\xi), \quad \xi \in \mathbb{C}.$$

Then on the unit circle (27) becomes

$$|uv| \leqslant C\big|Q(\xi)\big| \quad \text{for every } |\xi| = 1. \tag{28}$$

The above inequality encodes valuable information on the zero set of P_4. Recall that our goal is to show that uv divides P_4.

Observe that Q is a real polynomial with zero integral on the unit circle, as sum of two non- constant homogeneous harmonic polynomials. Thus Q vanishes at some point $\xi = u + iv$ on the unit circle. Then $uv = 0$ by (28) and so $P_4(\xi) = 0$, due to the definition of Q. We need now a precise expression for P_4. The general harmonic homogeneous polynomial of degree 4 is

$$\operatorname{Re}\big(\lambda\xi^4\big) = \alpha\big(u^3v - v^3u\big) + \beta\big(u^4 + v^4 - 6u^2v^2\big), \tag{29}$$

where λ is a complex number and α and β are real. Assume that P_4 is as above. We know that $u^2 + v^2 = 1$, $P_4(u, v) = 0$ and that $uv = 0$. If $u = 0$, then $\beta v^4 = 0$, which yields $\beta = 0$. If $v = 0$, then $\beta u^4 = 0$ and we conclude again that $\beta = 0$. Therefore

$$P_4(u, v) = \alpha\big(u^3v - v^3u\big) \tag{30}$$

and uv divides $P_4(u, v)$. We immediately conclude that the operator T with kernel

$$K(z) = \frac{xy}{|z|^4} + \frac{x^4 + y^4 - 6x^2y^2}{|z|^6}, \quad 0 \neq z \in \mathbb{C},$$

is an example in which the L^2 inequality (23) fails. Before going on we remark that a key step in proving the division property has been that Q has at least one zero on the circle. This is also a central fact in the proof of the general case.

We can easily deduce now another necessary condition for (23). Substituting (30) in (28) and simplifying the common factor uv we get

$$0 < \big|G_3(1)\big| \leqslant C\big(1 + \alpha\big(u^2 - v^2\big)\big), \quad |\xi| = 1,$$

which means that the right hand side cannot vanish on the unit circle, namely, $|\alpha| < 1$. Therefore we get the structural condition

$$T = R_P \circ U,$$

where R_P is the Riesz transform associated with the polynomial $P(x, y) = -(1/\pi)xy$ and U is an invertible operator in the Calderón–Zygmund algebra A.

Taking $\alpha = 1$ we get an operator T for which (23) fails but whose kernel

$$K(z) = -\frac{1}{\pi}\frac{xy}{|z|^4} + \frac{2}{\pi}\frac{x^3y - x, y^3}{|z|^6}, \quad 0 \neq z \in \mathbb{C},$$

satisfies the division property of part (iii) of Theorem 1.3.

6 Sufficient Conditions for the L^2 Estimate of $T^\star f$ by Tf

In this section we show how condition (iii) in Theorem 1.3 yields the pointwise inequality

$$T^\star f(z) \leqslant CM(Tf)(z), \quad z \in \mathbb{C}. \tag{31}$$

As in the previous section, we work in the particularly simple case in which the spherical harmonics expansion of the kernel is reduced to two terms. The first is a harmonic homogeneous polynomial of degree 2, which for definiteness is taken to be

$$P(z) = -\frac{1}{\pi}xy.$$

The second term is a fourth degree harmonic homogeneous polynomial. The division assumption in part (iii) of Theorem 1.3 is that P divides this second term. In view of the general form of a fourth degree harmonic homogeneous polynomial (29) we conclude that our kernel must be of the form

$$K(z) = -\frac{1}{\pi}\frac{xy}{|z|^4} + \frac{2}{\pi}\alpha\frac{x^3y - xy^3}{|z|^6}, \quad 0 \neq z \in \mathbb{C}, \alpha \in \mathbb{R}.$$

The second assumption in part (iii) of Theorem 1.3 is that T is of the form $T = R_P \circ U$, where R_P is the second order Riesz transform determined by P and U is an invertible operator in the Calderón–Zygmund algebra A. This is equivalent to $|\alpha| < 1$, as one can easily check looking at multipliers in the Fourier transform side.

In the simple context we have just set the two assumptions of condition (iii) of Theorem 1.3 are not independent. The reader can easily check that the structural condition $T = R_P \circ U$ implies the division property, that is, that P divides the fourth degree term. We will point out later on where this simplifies the argument.

We start now the proof of (31). Recall that, as we showed in the preceding section, there exists a bounded measurable function b supported on the unit ball B and a constant c such that

$$T(b) = \chi_{\mathbb{C}\setminus B}(z)K(z) + cxy\chi_B(z). \tag{32}$$

Our goal is to express the second term in the right hand side above as

$$cxy\chi_B(z) = T(\beta)(z) \quad \text{for almost all } z \in \mathbb{C}, \tag{33}$$

where β is a bounded measurable function such that

$$\left|\beta(z)\right| \leqslant \frac{C}{|z|^3}, \quad |z| \geqslant 2. \tag{34}$$

We first show that this is enough for (31). The only difficulty is that β is not supported in B, but the decay inequality (34) is an excellent substitute. Set $\gamma = b - \beta$. Then (dA being planar Lebesgue measure)

$$\begin{aligned} T^1 f(0) &= \int \chi_{\mathbb{C}\setminus B}(z)K(z)f(z)\,dA(z) \\ &= \int T(\gamma)(z)f(z)\,dA(z) \\ &= \int \gamma(z)Tf(z)\,dA(z) \\ &= \int_{2B} \gamma(z)Tf(z)\,dz + \int_{\mathbb{C}\setminus 2B} \gamma(z)Tf(z)\,dA(z). \end{aligned}$$

The first term is clearly less that a constant times $M(TF)(0)$, because γ is bounded, and the second too, because of (34) with β replaced by γ.

The proof of (33) is divided into two steps. The first step consists in showing that there exists a function β_0 such that

$$cxy\chi_B(z) = R(\beta_0)(z) \quad \text{for almost all } z \in \mathbb{C},$$

where $R = R_P$. To find β_0 let us look for a function ψ such that

$$P(\partial)\psi = cxy\chi_B(z). \tag{35}$$

Assume that we have found ψ and that it is regular enough that

$$\psi = E * \Delta\psi,$$

where E is the standard fundamental solution of the Laplacian. Then

$$\begin{aligned} cxy\chi_B(z) &= P(\partial)\psi = P(\partial)E * \Delta\psi \\ &= c\,\text{p.v.}\,\frac{P(x)}{|z|^4} * \Delta\psi = R(\beta_0), \end{aligned}$$

where $\beta_0 = c\Delta\psi$.

Taking the Fourier transform in (35) gives

$$P(\xi)\hat{\psi}(\xi) = c\partial_1\partial_2\widehat{\chi_B}(\xi) = cuvG_3\big(|\xi|\big).$$

For the definition of G_3 see the paragraph below (26). Hence

$$\hat{\psi}(\xi) = cG_3(\xi),$$

where c is some constant. It is a well known fact in the elementary theory of Bessel functions [3, p. 429] that

$$cG_3(\xi) = \Big(\big(1-|z|^2\big)^2\chi_B(z)\Big)\widehat{\ }(\xi).$$

In other words,

$$\psi(z) = c\big(1-|z|^2\big)^2\chi_B(z).$$

Clearly ψ and its first order derivatives are continuous functions supported on the closed unit ball B. The second order derivatives of ψ are supported on B and on B they are polynomials. In particular, we get that $\beta_0 = c\Delta\psi$ is a function supported on B, which satisfies a Lipschitz condition on B and satisfies the cancellation property $\int\beta_0 = c\int\Delta\psi = 0$.

It is worth remarking that in the general case, where the spherical harmonic expansion of the kernel contains many terms, one has to resort to the division assumption of part (iii) of Theorem 1.3 to complete the proof of the first step.

We proceed now with the second step. Since $T = R\circ U$ we have

$$cxy\chi_B(z) = R(\beta_0)(z) = T\big(U^{-1}(\beta_0)\big)(z).$$

Set $\beta = U^{-1}(\beta_0)$, so that (33) is satisfied. We are left with the task of showing that β is bounded and satisfies the decay estimate (34).

The inverse of U is an operator in the Calderón–Zygmund algebra A. Thus

$$\beta = U^{-1}(\beta_0) = (\lambda I + V)(\beta_0) = \lambda\beta_0 + V(\beta_0),$$

where λ is a real number and V an even convolution smooth homogeneous Calderón–Zygmund operator. The desired decay estimate for β now follows readily, because β_0 is supported in the closed ball B and has zero integral. It remains to show that $V(\beta_0)$ is bounded. At first glance this is quite unlikely because V is a general even convolution smooth homogeneous Calderón–Zygmund operator and β_0 has no global smoothness properties in the plane. Indeed, although β_0 is Lipschitz on B, it has a jump at the boundary of B. Assume for a moment that $\beta_0 = \chi_B$. It is then known that $V(\chi_B)$ is a bounded function because V is an even Calderón–Zygmund operator and the boundary of B is smooth. Here the fact that the operator is even is crucial as one can see by considering the action of the Hilbert transform on the interval $(-1,1)$. We are not going to present the nice argument for the proof that $V(\beta_0)$ is bounded [5]. Let us only mention that this result for the Beurling transform and smoothly bounded domains was known to experts in Fluid Dynamics. It plays a basic role in the regularity theory of certain solutions of the Euler equation in the plane [1].

7 The Proof in the General Case and Final Comments

The proof of Theorems 1.3 and 1.5 in the general case proceeds in two stages. First one proves the Theorems in the case in which the spherical harmonics expansion of the kernel contains finitely many non-zero terms. Then one has to truncate the expansion of the kernel and see that some of the estimates obtained in the first step do not depend on the number of terms. This is a delicate issue at some moments, but necessary to perform a final compactness argument. In both steps there are difficulties of various types to be overcome and a major computational issue, lengthly and involved, which might very likely be significantly simplified by a more clever argument.

A final word on the proof for the necessity of the division condition. To show that a polynomial with complex coefficients divides another, one often resorts to Hilbert's Nullstellensatz, the zero set theorem of Hilbert, which states that if P is a prime polynomial with complex coefficients and finitely many variables, then P divides another such polynomial Q if Q vanishes on the zeros of P. This fails for real polynomials, as simple examples show. Now, since we are working with real polynomials, we cannot straightforwardly apply Hilbert's theorem. What saves us is that our real polynomials have a fairly substantial amount of zeros, just because they have zero integral on the unit sphere. We can then jump to the complex case and come back to the real by checking that the Hausdorff dimension of the zero set of certain polynomials is big enough.

There are several questions about Theorems 1.3 and 1.5 that deserve further study. The first is a potential application to the David–Semmes problem mentioned in the introduction, which was the source of the question. Another is the smoothness of the kernels. It is not known how to prove the analogs of Theorems 1.3 and 1.5 for kernels of moderate smoothness, say of class C^m for some positive integer m. Finally it is has recently been shown by Bosch, Mateu and Orobitg that

$$\left\| T^{\star} f \right\|_p \leqslant C \| T f \|_p, \quad f \in L^p\left(\mathbb{R}^n\right), 1 < p < \infty,$$

implies any of the three equivalent conditions in Theorems 1.3 and 1.5.

Acknowledgements The author is grateful to the organizers of the XXXI Conference in Harmonic Analysis, held in Rome at the beginning of June 2011 in honor of Professor A.Figà-Talamanca, for the kind invitation to participate and for their efficient work. The author has been partially supported by the grants MTM2010-15657 and 2009SGR420.

References

1. Chemin, J.Y.: Fluides parfaits incompressibles. Astérisque, vol. 230. Soc. Math. France, Paris (1995)

2. David, G., Semmes, S.: Singular integrals and rectifiable sets in $\mathbb{R}^n$: Au-delàs des graphes lipschitziens. Astérisque, vol. 193. Soc. Math. France, Paris (1991)
3. Grafakos, L.: Classical Fourier Analysis, 2nd edn. Graduate Texts in Mathematics, vol. 249. Springer, Berlin (2008)
4. Grafakos, L.: Modern Fourier Analysis, 2nd edn. Graduate Texts in Mathematics, vol. 250. Springer, Berlin (2008)
5. Mateu, J., Orobitg, J., Verdera, J.: Estimates for the maximal singular integral in terms of the singular integral: The case of even kernels. Ann. Math. **174**, 1429–1483 (2011)
6. Mateu, J., Orobitg, J., Pérez, C., Verdera, J.: New estimates for the maximal singular integral. Int. Math. Res. Not. **2010**, 3658–3722 (2010)
7. Mateu, J., Verdera, J.: L^p and weak L^1 estimates for the maximal Riesz transform and the maximal Beurling transform. Math. Res. Lett. **13**, 957–966 (2006)
8. Pérez, C.: Endpoint estimates for commutators of singular integral operators. J. Funct. Anal. **128**, 163–185 (1995)
9. Stein, E.M.: Singular Integrals and Differentiability Properties of Functions. Princeton University Press, Princeton (1970)