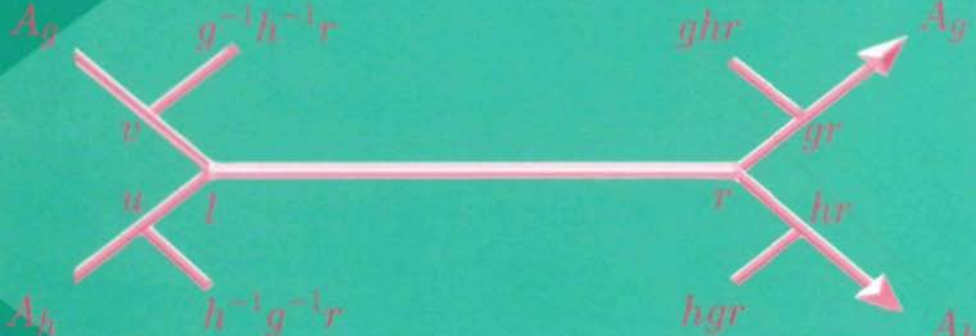
Introduction to Λ-TREES



Ian Chiswell

World Scientific

Introduction to Λ-TREES

Ian Chiswell

School of Mathematical Sciences
Queen Mary and Westfield College
University of London, UK

World Scientific
Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 912805

USA office:

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

INTRODUCTION TO Λ-TREES

ISBN 981-02-4386-3

Printed in Singapore by World Scientific Printers

Preface

A tree (that is, a connected circuit-free graph) defines an integer-valued metric on its set of vertices, called the path metric, just by taking the shortest path between two vertices, and defining their distance apart to be the number of edges in the path. The idea of a Λ-tree (Λ being a totally ordered abelian group) is obtained by considering Λ-valued metrics having certain properties in common with the path metric. Many of the important results concern the case $\Lambda = \mathbb{R}$, the additive group of the real numbers; an $\mathbb{R}$-tree is a generalisation of the geometric realisation of an ordinary tree. However, the view that this is the only interesting case is erroneous. The case $\Lambda = \mathbb{Z} \oplus \Lambda_0$, where Λ_0 is an ordered abelian group and Λ has the lexicographic ordering, has been studied in a paper by Bass [5]. Also, non-standard models of the first-order theory of the non-abelian free groups act freely on Λ-trees, where Λ is a non-standard model of $\mathbb{Z}$, viewed as an ordered abelian group. These non-standard free groups, and their subgroups, can be characterised group-theoretically. This theory is discussed in §5 of Chapter 5.

The theory of Λ-trees has its origins in two papers, by the author [30] in 1976 and by J.Tits [143] in 1977. Tits' paper contains a definition of an $\mathbb{R}$-tree, although the requirement that it should be a complete metric space is not needed for most purposes, and in recent work is assumed only when necessary. The author's paper contains a construction of an $\mathbb{R}$-tree starting from a Lyndon length function on a group, an idea considered earlier by Lyndon [92], who introduced integer-valued length functions to generalise the Nielsen method in free groups and free products. (Real-valued length functions were considered by Harrison [72], whose results, interpreted in terms of Λ-trees, are given in Chapter 5.) However, the author did not show that the space constructed in [30] is an $\mathbb{R}$-tree. This was left to Imrich [79] and (in perhaps a more accessible reference), Alperin and Moss [3]. It should be noted that these authors used various different definitions of an $\mathbb{R}$-tree, the equivalence of which is established in the course of Chapters 2 and 3. In fact, Imrich made use of the "four-point condition", which is equivalent to "0-hyperbolic", using the important notion of hyperbolicity introduced by Gromov [67] (see §2 of Chapter 1).

The theory of $\mathbb{R}$-trees came into prominence with the work of Morgan and

Shalen [106]. They showed that if G is a finitely generated group, then the space of discrete, faithful representations of G as orientation-preserving isometries of hyperbolic space $\mathbb{H}^n$, factored out by conjugacy in the group $\mathrm{SO}(n,1)$, has a compactification, in which the ideal points are obtained from certain actions of G on $\mathbb{R}$-trees. This gave a new viewpoint for work of Thurston on Teichmüller space. Their paper also contains the definition of Λ-tree, and develops the theory of Λ-trees for arbitrary ordered abelian groups Λ. In particular, the construction in the author's paper [30] mentioned above generalises to arbitrary Λ. This construction is given in Theorem 4.4 of Chapter 2, and a version for Λ-valued Lyndon length functions is proved as a corollary in Theorem 4.6. The work of Morgan and Shalen is described in a survey article by Morgan [103]. In their work, they generalise the construction of a tree from a discrete valuation given in [127; Ch. II]. If F is a field with a valuation (not necessarily discrete), this gives a Λ-tree on which $\mathrm{GL}_2(F)$ acts as isometries, where Λ is the value group. The construction is described in §3 of Chapter 4.

The theory was further developed in an important paper by Alperin and Bass [2], which is the standard reference. Another useful paper for the general theory is that by Culler and Morgan [42], which confines attention to $\mathbb{R}$-trees. However, the most interesting results in this paper concern spaces of actions on $\mathbb{R}$-trees, motivated by the fact that the compactification by Morgan and Shalen mentioned previously involves points in this space. They show that, if G is a finitely generated group, then the space $\mathrm{PLF}(G)$ of projectivised non-zero hyperbolic length functions for actions of G on $\mathbb{R}$-trees is itself compact. (The definition of this space is given in §6 of Chapter 4.) They also show that the space $\mathrm{SLF}(G)$ of points in $\mathrm{PLF}(G)$ arising from what are called small actions of G on $\mathbb{R}$-trees is a compact subspace. An action is called *small* if the stabilizer of every segment (as defined in Chapter 1) contains no free subgroup of rank 2. This result is of interest because the points in the Morgan-Shalen compactification belong to $\mathrm{SLF}(G)$. There have been some interesting developments concerning related spaces, which are mentioned at the end of Chapter 4.

Alperin and Bass [2] state a fundamental problem in the theory of Λ-trees. Find the group theoretic information carried by a Λ-tree action, analogous to that presented in [127] for the case $\Lambda = \mathbb{Z}$, due to Bass and Serre. This problem is far from solved. However, some aspects of the theory of group actions on ordinary trees can be generalised to group actions (as isometries) on Λ-trees. These include a classification of single isometries of Λ-trees, and the construction of a Λ-tree arising from a valuation on a field F mentioned above.

Chapter 1 is an introductory chapter, including a discussion of metrics with values in an ordered abelian group. The other topics presented are those needed in the first three chapters, which are not usually taught in the basic

M.Sc. courses in London. This is an attempt to make these chapters accessible to beginning research students. Among the topics assumed to be known are basic facts about free groups, but not free products or HNN-extensions, except for a brief use of HNN-extensions in §4 of Chapter 3. This construction is described, for example, in [95]. Chapters 2 and 3 give the basics of the theory of Λ-trees, based on [2]. Our account differs from that in [2] mostly in detail only. The main differences are in §4 of Chapter 2, and the addition of material specific to $\mathbb{R}$-trees in §2 of Chapter 2. Also, the material on pairs of isometries in §3 of Chapter 3 has been reorganised.

The rest of the book is about group actions by isometries on Λ-trees, and is in places less self-contained. Thus in Chapter 4, §4, knowledge of some basic ideas about manifolds is assumed, while in §5 a knowledge of geometry in the hyperbolic plane is needed. This section involves the theory of non-Euclidean crystallographic groups, and to avoid excessive length, a complete account of this is not given. Details are given only when they are particularly easy, or when it would be tortuous to extract them from the references. (It is unfortunate that many of the standard references deal only with the case of Fuchsian groups.) Nevertheless, throughout the book an effort has been made to maintain an elementary tone, spelling things out as much as possible. This unfortunately leads to some rather long proofs in places, particularly in §§4 and 5 of Chapter 4, and in Chapter 6.

Chapter 4 is mostly a collection of disjoint topics, although §5 depends on §4 and §6 follows on naturally from §5. The main result of §5 is used to prove that with three exceptions, the fundamental group of a compact surface has a free action on an $\mathbb{R}$-tree. The argument for this is sketched in Chapter 5. The results of §3, on the Λ-tree associated with a valuation, are used in §5 of Chapter 5. Otherwise, the results of this chapter are not used later.

The simplest case of the Bass-Serre theory mentioned above is when the action is free, and this suggests looking at free actions on Λ-trees. There has indeed been considerable progress in this area, and this is surveyed, with varying degrees of detail, in Chapter 5. One consequence is that a free product of surface groups (other than the three exceptions) and finitely generated free abelian groups has a free action on an $\mathbb{R}$-tree. The final chapter contains a proof of a converse: a finitely generated group acting freely on an $\mathbb{R}$-tree is a free product of surface groups and free abelian groups. E. Rips outlined a proof of this at a conference at the Isle of Thorns, a conference centre of the University of Sussex, England, in 1991. It was not published, and the proof we give is that of Gaboriau, Levitt and Paulin [53].

The book is based on lectures given by the author at two conferences, at York, England in 1993 and at Banff, Alberta in 1996. These were written up in [34] and [36]. It is not claimed to be a comprehensive account of the subject, and for further reading we recommend the survey articles by Morgan and by Shalen [103], [128], [129]. It is also not completely up-to-date. In particular, there has been recent progress in the theory of even more general kinds of trees, defined without recourse to a metric, which is not covered here. This work is described in two volumes of the AMS Memoirs, by Adeleke and Neumann [1], and by Bowditch [21]. It seems likely that in future, these ideas will be increasingly important in arboreal group theory.

Despite the simplicity of the definition, the basic theory of Λ-trees is long and technical. Readers wishing to learn the basics in the first three chapters are therefore recommended to skip some of the proofs. These could include §2 in Chapter 2, from Lemma 2.4 onwards, and Lemmas 3.1 to 3.5 in §3 of Chapter 3. There are also two theorems (Theorem 3.14 in Chapter 2 and Theorem 1.4 in Chapter 3) where the result for general Λ is much harder to prove than in the case $\Lambda = \mathbb{R}$.

The author thanks D.J. Collins and S.J. O'Rourke for their comments on the first two chapters, and several people for help on points which arose while writing the book, either in conversation or by email. These include M. Bestvina, B. Bowditch, S.R. Bullett, D.B.A. Epstein, G. Levitt and W.P. Thurston.

Theorems, lemmas etc are numbered consecutively in each section. Thus a reference to Theorem 1.4 is to the fourth numbered result in §1 of the same chapter, while Theorem 3.1.4 means Theorem 1.4 in Chapter 3. Similarly, §3.1 means §1 of Chapter 3.

The bibliography contains only items referred to in the text, and is not intended as a survey of the literature on the subject. In particular, it only refers to specific articles in the book Arboreal group theory, (ed. R.C. Alperin. New York: Springer-Verlag 1991). This contains several other interesting papers on $\mathbb{R}$-trees. Also, the text contains several references to W.S. Massey's book [98]; some, but not all of the cited material is reprinted in his later book, A basic course in algebraic topology (New York: Springer-Verlag 1991).

One point should be made about the main index. For the definition of items listed under "$\mathbb{R}$-tree", it may be necessary to look them up under "Λ-tree".

Finally, the subject gives many opportunities for puns. The letter X has been used to denote a Λ-tree, in order to perpetuate the terminology "X-ray" of Alperin and Bass. However, the author could not bring himself to use lower-case γ for a graph, so the term Γ-ray is used. In fact, it is difficult to avoid inadvertent puns, and a colleague once described the theory of Λ-trees, quite seriously, as "a new branch of representation theory".

Contents

CHAPTER 1. Preliminaries

1. Ordered Abelian Groups

A partially ordered abelian group is an abelian group Λ, together with a partial ordering $\leq$ on Λ, such that for all a, b and $c \in \Lambda$, $a \leq b$ implies $a + c \leq b + c$. If the partial ordering is a linear (total) ordering then Λ will be called an ordered abelian group. We shall use additive notation for ordered abelian groups unless otherwise indicated. There will occasionally be situations where multiplicative notation is more natural. Given an ordered abelian group Λ, let $P = \{a \in \Lambda | \ a > 0\}$. Then

(1) Λ is the disjoint union of P, $-P$ and $\{0\}$
(2) $P + P \subseteq P$.

Conversely, given an abelian group Λ and a subset P satisfying (1) and (2), it is easy to see that Λ can be made into an ordered abelian group by defining $a \leq b$ if and only if $b - a \in P \cup \{0\}$. We call P the positive cone of the ordered abelian group Λ.

Examples.

(1) An ordered field is a field K with a linear ordering $\leq$ such that the additive group of K is an ordered abelian group with this ordering, and such that $PP \subseteq P$. Then P is itself an ordered abelian group, using the multiplication of the field K and the ordering $\leq$, suitably restricted. The positive cone is $\{x \in K | \ 1 < x\}$. This is one instance where multiplicative notation is appropriate.
Examples of such fields are the field of real numbers $\mathbb{R}$ and any of its subfields, e.g. $\mathbb{Q}$.

(2) Any subgroup of an ordered abelian group Λ is an ordered abelian group, using the restriction of the ordering on Λ. In particular, the additive subgroups of $\mathbb{R}$, such as $\mathbb{R}$ itself and $\mathbb{Z}$ are ordered abelian groups. We shall say a little more about the additive subgroups of $\mathbb{R}$ shortly.

(3) If Λ_1 and Λ_2 are ordered abelian groups, the direct sum $\Lambda_1 \oplus \Lambda_2$ can be made into an ordered abelian group by defining $(a, b) \leq (c, d)$ if and only if either $a < c$, or $a = c$ and $b \leq d$ (this is the *lexicographic ordering*). In particular, $\mathbb{Z} \oplus \mathbb{Z}$ is an ordered abelian group in this way.

The lexicographic ordering extends inductively to the direct sum of finitely many ordered abelian groups $\Lambda_1 \oplus \cdots \oplus \Lambda_k$.

These will be sufficient examples to illustrate the general theory, although ordered abelian groups can be much more complicated.

Exercise. Show that an ordered abelian group Λ is torsion-free; that is (in additive notation), if $0 \neq a \in \Lambda$ and n is a non-zero integer, then $na \neq 0$. Further, if $n > 0$ and $na > 0$, show that $a > 0$.

If Λ is an ordered abelian group and $a \in \Lambda$, we define the absolute value or modulus $|a|$ by $|a| = \max\{a, -a\}$. This has the following properties (again using additive notation). For all $a, b \in \Lambda$,

(i) $|a| \geq 0$, and $|a| = 0$ if and only if $a = 0$;
(ii) $|a + b| \leq |a| + |b|$.

In the case of an ordered field (Example 1 above), we also have
(iii) $|a||b| = |ab|$.

We can also define intervals in an ordered abelian group. Given $a, b \in \Lambda$ with $a \leq b$, where Λ is an ordered abelian group, we define $[a, b]_\Lambda = \{x \in \Lambda | a \leq x \leq b\}$, and $[b, a]_\Lambda = [a, b]_\Lambda$. Note that $[a, a]_\Lambda = \{a\}$.

Exercise. Show that if a, b and $c \in \Lambda$, then $c \in [a, b]_\Lambda$ if and only if $|a - b| = |a - c| + |c - b|$.

By a morphism of ordered abelian groups we mean a group homomorphism $h : \Lambda \to \Lambda'$, where Λ and Λ' are ordered abelian groups, such that $a \leq b$ implies $h(a) \leq h(b)$. For example, if $\mathbb{Z} \oplus \mathbb{Z}$ has the lexicographic ordering, then projection onto the first factor is a morphism of ordered abelian groups. If the morphism h is one-to-one, we shall call it an embedding of ordered abelian groups.

A subset X of an ordered abelian group Λ is called *convex* if $a, b \in X$ and $a \leq c \leq b$, where $c \in \Lambda$, implies $c \in X$. That is, X is convex if $a, b \in X$ and $c \in [a, b]_\Lambda$ implies $c \in X$. In particular, it makes sense to describe a subgroup of Λ as convex (for example, the trivial subgroup and the whole group are convex).

Lemma 1.1. *The set of convex subgroups of an ordered abelian group is linearly ordered by inclusion.*

Proof. Suppose Λ_1 and Λ_2 are convex subgroups, and it is possible to find $a \in \Lambda_1 \setminus \Lambda_2$ and $b \in \Lambda_2 \setminus \Lambda_1$. Replacing a, b by $-a$, $-b$ as necessary, we can assume that $a > 0$, $b > 0$. Also, either $a \leq b$ or $b \leq a$. If $0 < a \leq b$, then $a \in \Lambda_2$ since Λ_2 is convex, a contradiction. Similarly, if $0 < b \leq a$, we obtain the contradiction $b \in \Lambda_1$. □

If $h : \Lambda \to \Lambda'$ is a morphism of ordered abelian groups, then $\mathrm{Ker}(h)$ is a convex subgroup of Λ. If Λ' is a convex subgroup of Λ, then the quotient

Λ/Λ' can be made into an ordered abelian group by defining $a + \Lambda' > 0$ if and only if $a > b$ for all $b \in \Lambda'$. (This defines the positive cone P, and P satisfies conditions (1) and (2) above, so this specifies the ordering.) Further, the canonical projection $\pi : \Lambda \to \Lambda/\Lambda'$ is a morphism of ordered abelian groups. There is also a correspondence theorem: $M \mapsto \pi^{-1}(M)$ gives a $1-1$ correspondence between the set of convex subgroups of Λ/Λ' and the set of convex subgroups of Λ containing Λ'.

The rank of an ordered abelian group is defined to be the order type of its set of non-trivial convex subgroups. (For the definition of order type of a linearly ordered set see, for example, [101; §15].) In situations where we need to consider the rank, the set of non-trivial convex subgroups will always be finite, with n elements, say, in which case we may take the rank to be n. Thus Λ is of rank 0 if and only if $\Lambda = 0$. The group $\mathbb{Z} \oplus \mathbb{Z}$ with the lexicographic ordering is of rank 2, the proper non-trivial convex subgroup being $0 \oplus \mathbb{Z}$. Rank 1 ordered abelian groups are characterised by the next theorem, but before stating it we need a definition.

Definition. An ordered abelian group Λ is called archimedean if, given a, $b \in \Lambda$ with $b \neq 0$, there exists $n \in \mathbb{Z}$ such that $a < nb$.

The field $\mathbb{R}$ is a complete ordered field. Nowadays this completeness property is usually expressed by saying that a non-empty subset of $\mathbb{R}$ which is bounded above has a least upper bound, equivalently a non-empty subset of $\mathbb{R}$ which is bounded below has a greatest lower bound. It follows easily that $\mathbb{R}$ is archimedean, and also that the field $\mathbb{Q}$ of rational numbers is dense in $\mathbb{R}$. We assume readers are familiar with these facts, which are proved in many elementary texts on real analysis. (We recall that a subspace Y of a topological space X is said to be dense in X if X is the closure of Y. In the case $X = \mathbb{R}$, this means that between any two real numbers there is a point of Y. Equivalently, given $x \in \mathbb{R}$ and a real $\varepsilon > 0$, there exists $y \in Y$ with $|y - x| < \varepsilon$.

Theorem 1.2. *For an ordered abelian group Λ, the following are equivalent.*
(1) *Λ has rank less than or equal to 1;*
(2) *Λ is archimedean;*
(3) *there exists an embedding of ordered abelian groups $\Lambda \to \mathbb{R}$.*

Proof. The proof of (1)$\Longleftrightarrow$(2) is left as an exercise, (3)$\Rightarrow$(2) follows because $\mathbb{R}$ is archimedean, and we prove that (2)$\Rightarrow$(3). Suppose that Λ is archimedean. We can assume $\Lambda \neq 0$, otherwise Λ obviously embeds in $\mathbb{R}$, so we can choose $a \in \Lambda$ with $a > 0$. For $b \in \Lambda$, let L_b be the set of all rational numbers m/n, where m, $n \in \mathbb{Z}$ and $n > 0$, such that $ma < nb$ (note that this condition depends only on the rational number m/n, not on the choice of integers m and $n > 0$ representing it). Also, let R_b be the set of all rational numbers m/n, where m, $n \in \mathbb{Z}$ and $n > 0$, such that $ma \geq nb$. It is easy to see that if $x \in L_b$, $y \in R_b$ then $x < y$, and since Λ is archimedean, L_b and R_b are non-empty. Also, $\mathbb{Q} = L_b \cup R_b$. Thus $\sup(L_b) \leq \inf(R_b)$, and if there is strict inequality, we can find a rational number $q \in (\sup(L_b), \inf(R_b))$, because $\mathbb{Q}$ is dense in $\mathbb{R}$. But then $q \notin L_b \cup R_b$, a contradiction.

Therefore we can define a map $h : \Lambda \to \mathbb{R}$ by $h(b) = \sup(L_b) = \inf(R_b)$. If $b \leq c$, where $c \in \Lambda$, then clearly $L_b \subseteq L_c$, so $h(b) \leq h(c)$. If $x \in L_b$ and $y \in L_c$, then it is easily checked that $x + y \in L_{b+c}$, so $x + y \leq h(b + c)$. It follows that $h(b) + h(c) \leq h(b + c)$ (using the rule $\sup(A + B) = \sup(A) + \sup(B)$, where A, B are non-empty subsets of $\mathbb{R}$, bounded above). By similarly considering R_b and R_c, we find that $h(b) + h(c) \geq h(b + c)$, so h is a homomorphism. If $b > 0$, then since Λ is archimedean, there exists a positive integer n such that $a < nb$. Then $1/n \in L_b$, so $h(b) \geq 1/n > 0$. If $b < 0$, then $h(-b) = -h(b) > 0$. It follows that h has trivial kernel, completing the proof. □

Underlying the proof of Theorem 1.2 is the idea of a Dedekind cut, and the fact that a Dedekind cut in the rationals determines a unique real number. However, the use of Dedekind cuts to construct the real numbers, and to express the completeness property of $\mathbb{R}$, seems to have gone out of fashion (with good reason-there are easier ways to construct the real numbers). Nevertheless, we shall later give a definition of cut in a Λ-tree generalising the notion of Dedekind cut, and these cuts will play a role in the theory. The pair of sets L_b, R_b in the proof above form a cut in $\mathbb{Q}$, and $h(b)$ is the unique real number determined by the cut.

There is one simple fact we can note about the subgroups of $\mathbb{R}$.

Lemma 1.3. *Let Λ be a subgroup of the additive group of $\mathbb{R}$. Then either Λ is cyclic, or it is dense in $\mathbb{R}$.*

Proof. We can assume $\Lambda \neq 0$. Then we can define $a = \inf\{\lambda \in \Lambda |\ \lambda > 0\}$, so $a \in \mathbb{R}$ and $a \geq 0$.

If $a > 0$, then $a \in \Lambda$. For otherwise, there exists $b \in \Lambda$ such that $a < b < 2a$, and there exists $c \in \Lambda$ such that $a < c < b$. Then $0 < b - c < a$, and $b - c \in \Lambda$, a contradiction. Now if $b \in \Lambda$, since Λ is archimedean, there exists an integer n such that $na \leq b < (n+1)a$. Then $0 \leq b - na < a$ and $b - na \in \Lambda$, so $b = na$. Hence Λ is cyclic, generated by a.

If $a = 0$, take $b \in \mathbb{R}$ and choose a real $\epsilon > 0$. There exists $c \in \Lambda$ such that $0 < c < \epsilon$, and again since Λ is archimedean, there is an integer n such that $0 \leq b - nc < c < \epsilon$. Thus $|b - nc| < \epsilon$, and $nc \in \Lambda$, showing that Λ is dense in $\mathbb{R}$. □

Let Λ be an ordered abelian group. Viewing abelian groups as $\mathbb{Z}$-modules, we can form the abelian group $\mathbb{Q} \otimes_{\mathbb{Z}} \Lambda$. It is well-known from commutative algebra that this group can be described as follows. It is the set of equivalence classes of the equivalence relation $\sim$ defined on the set $\Lambda \times (\mathbb{Z} \setminus \{0\})$ by: $(\lambda, m) \sim (\mu, n)$ if and only if $n\lambda = m\mu$. Denoting the equivalence class of (λ, m) by $\frac{\lambda}{m}$, addition is given by

$$\frac{\lambda}{m} + \frac{\mu}{n} = \frac{(n\lambda + m\mu)}{mn}.$$

We can make $\mathbb{Q} \otimes_{\mathbb{Z}} \Lambda$ into an ordered abelian group by defining $\frac{\lambda}{m} > 0$ if and only if $m\lambda > 0$ (it suffices to specify the positive cone P and check that conditions (1) and (2) at the beginning of this section are satisfied). The map $\Lambda \to \mathbb{Q} \otimes_{\mathbb{Z}} \Lambda$ given by $\lambda \mapsto \frac{\lambda}{1}$ is an embedding of ordered abelian groups. In particular, Λ embeds in the subgroup $\frac{1}{2}\Lambda = \left\{\frac{\lambda}{2} | \lambda \in \Lambda\right\}$, which is an ordered abelian group isomorphic to Λ. This group $\frac{1}{2}\Lambda$ will play a role in what follows. Since $\mathbb{Q} \otimes_{\mathbb{Z}} \Lambda$ is divisible, we have shown the following.

Lemma 1.4. *Any ordered abelian group can be embedded in a divisible ordered abelian group.*

□

Exercise. Show that, if Λ is archimedean, then so is $\mathbb{Q} \otimes_{\mathbb{Z}} \Lambda$.

Lemma 1.5. *Suppose Λ is an ordered abelian group which is a direct sum $\Lambda = \Lambda_1 \oplus \Lambda_2$, where Λ_2 is a convex subgroup of Λ. Then the ordering on Λ is the lexicographic ordering.*

Proof. Since the ordering is determined by the positive cone, it suffices to show that, if $a \in \Lambda_1$ and $b \in \Lambda_2$, then $a + b > 0$ if and only if either $a > 0$, or $a = 0$ and $b > 0$. Suppose $a + b > 0$. If $a < 0$, then $0 < -a < b$, which implies $a \in \Lambda_2$, hence $a = 0$, a contradiction. Thus either $a > 0$, or $a = 0$, in which case $b > 0$. The converse is proved similarly. □

Suppose Λ is an ordered abelian group and Λ_1 is a convex subgroup. The quotient Λ/Λ_1 is an ordered abelian group, in particular it is torsion-free, so if Λ is finitely generated, it is free abelian, hence the short exact sequence

$$0 \longrightarrow \Lambda_1 \longrightarrow \Lambda \longrightarrow \Lambda/\Lambda_1 \longrightarrow 0$$

splits, and Λ_1 is a direct summand of Λ. Thus, if $\Lambda_0 \subsetneq \Lambda_1 \subsetneq \ldots \subsetneq \Lambda_k$ is a chain of convex subgroups of Λ, the $\mathbb{Z}$-rank of Λ is at least k, so if Λ is finitely generated, it has finite rank. We can now prove the following lemma, which we believe is due to Zaitseva [150].

Lemma 1.6. *If Λ is a finitely generated ordered abelian group, it has a decomposition $\Lambda = \Lambda_1 \oplus \cdots \oplus \Lambda_k$ for some $k \geq 0$, where the ordering is the lexicographic ordering and $\Lambda_1, \ldots, \Lambda_k$ are of rank 1.*

Proof. This is clear if $\Lambda = 0$. Otherwise, by the above Λ is of finite rank and the convex subgroups are linearly ordered by inclusion, so there is a unique maximal convex subgroup M, and Λ/M is of rank 1 (the correspondence theorem gives a bijection between the convex subgroups of Λ/M and the convex subgroups of Λ containing M). Again by the above, we have $\Lambda = \Lambda_1 \oplus M$ for some Λ_1, and by Lemma 1.5 the ordering is the lexicographic ordering. Also, Λ_1 is isomorphic as an ordered abelian group to Λ/M, so has rank 1. Since the remaining convex subgroups of Λ have to be the convex subgroups of M, the result follows by induction on the rank.

□

2. Metric Spaces

Readers are assumed to be familiar with the basic theory of metric spaces. However, we shall consider metrics with values in an arbitrary ordered abelian group, and to emphasise that the axioms for a metric make sense in this context, we begin with the definition.

Definition. Let X be a non-empty set, Λ an ordered abelian group. A Λ-pseudometric on X is a mapping $d : X \times X \to \Lambda$ such that:

(i) $d(x, y) \geq 0$ for all x, $y \in X$
(ii) $d(x, x) = 0$ for all $x \in X$
(iii) $d(x, y) = d(y, x)$ for all x, $y \in X$
(iv) $d(x, y) \leq d(x, z) + d(y, z)$ for all x, y, $z \in X$.

We call the inequality in (iv) the triangle inequality, as usual. These are the axioms as usually given, but we note that when (iv) is stated this way, (i) and (iii) are redundant-they follow from (ii) and (iv). A Λ-pseudometric d on X is called a Λ-metric if

(ii)′ $d(x, y) = 0$ if and only if $x = y$, for all x, $y \in X$.

We shall be mainly interested in Λ-metrics, but the construction of a metric from a pseudometric, which works for arbitrary Λ just as in the case $\Lambda = \mathbb{R}$, will occasionally be used. If d is a Λ-pseudometric on X, then the relation $\sim$ on X defined by $x \sim y if and only if d(x, y) = 0$ is an equivalence relation on X. If $\overline{X}$ denotes the set of equivalence classes, then we can define a Λ-metric $\bar{d}$ on $\overline{X}$ by putting $\bar{d}(\langle x \rangle, \langle y \rangle) = d(x, y)$, where $\langle x \rangle$ is the equivalence class of x.

A Λ-metric space is a pair (X, d), where X is a non-empty set and d is a Λ-metric on X. If (X, d) and (X', d') are Λ-metric spaces, an isometry from

(X, d) to (X', d') is a mapping $f : X \to X'$ such that $d(x, y) = d'(f(x), f(y))$ for all x, $y \in X$. Such a mapping is necessarily one-to-one, but need not be onto. If it is onto, we call it a metric isomorphism, and (X, d), (X', d') are said to be metrically isomorphic if there exists a metric isomorphism $f : X \to X'$.

We note that, provided $\Lambda \neq \{0\}$, any Λ-metric space (X, d) has a topology with basis the open balls $B(x, r) = \{y \in X \mid d(x, y) < r\}$, where $x \in X$ and $r \in \Lambda$, $r > 0$, just as in the case $\Lambda = \mathbb{R}$.

Exercise. Show that this topology is Hausdorff and normal.

One example of a Λ-metric space is (Λ, d), where $d(a, b) = |a - b|$. We shall need to know about the metric automorphisms of Λ, that is, metric isomorphisms from Λ to Λ. Examples are the *translations* $t_a(x) = a + x$ and the *reflections* $r_a(x) = a - x$, for $a \in \Lambda$. These are in fact the only examples.

Lemma 2.1. (1) *The group G of metric automorphisms of Λ consists of the reflections and translations. The set T of translations is a normal subgroup of G isomorphic to Λ, and G is the semidirect product of T and the cyclic subgroup generated by r_0 ($x \mapsto -x$).*

(2) *If A is non-empty subset of Λ, then any isometry $\alpha : A \to \Lambda$ is the restriction to A of a metric automorphism of Λ, and if A has at least two elements, this metric automorphism is uniquely determined by α.*

Proof. We show first that, if $A \subseteq \Lambda$ and the isometry $\alpha : A \to \Lambda$ fixes two distinct points of A, say a and b, then α is the inclusion map. Let $c \in A$. Then $|c - a| = |\alpha(c) - \alpha(a)| = |\alpha(c) - a|$, and similarly $|c - b| = |\alpha$
$\alpha(c) \neq c$, we have $c - a = a - \alpha(c)$ and $c - b = b - \alpha(c)$, hence $c - 2a = c - 2b$, giving $a = b$, a contradiction. Thus $\alpha(c) = c$, as claimed. This proves the uniqueness statement in (2), on taking $A = \Lambda$.

Now take $a \in A$. Then $\alpha(a) = t(a)$, where $t = t_{\alpha(a)-a}$, so if A has only one element, α is the restriction of t to A. If b is another element of A, then $|t^{-1}\alpha(b) - t^{-1}\alpha(a)| = |b - a|$, whence $t^{-1}\alpha(b) - a = \pm(b - a)$, so either $t^{-1}\alpha(b) = b$, in which case α is the restriction of t to A, since $t^{-1}\alpha$ fixes both a and b, or $t^{-1}\alpha(b) = 2a - b = r_{2a}(b)$. In this case α is the restriction to A of tr_{2a} since both have the same effect on a and b. Using the easily verified formula

$$t_a r_b = r_{a+b} \tag{i}$$

we see that $tr_{2a} = r_c$, where $c = \alpha(a) + a$.

Thus α is the restriction of either a translation or a reflection; this proves (2), and the first part of (1), on taking $A = \Lambda$. Clearly $a \mapsto t_a$ gives an isomorphism of Λ with T, and T is normal in G using the formula

$$r_b t_a r_b^{-1} = t_{-a} \tag{ii}$$

(note that $r_b^2 = 1$). If R is the cyclic subgroup generated by r_0, then $G = TR$ by (1), and clearly $T \cap R = 1$. □

Note that a is called the *amplitude* of the translation t_a.

Definition. A segment in a Λ-metric space (X, d) is the image of an isometry $\alpha : [a, b]_\Lambda \to X$ for some a, $b \in \Lambda$. Note that $a = b$ is allowed. The endpoints of the segment are $\alpha(a)$, $\alpha(b)$. We shall say that the segment joins $\alpha(a)$, $\alpha(b)$.

The endpoints are characterised by the fact that $d(\alpha(a), \alpha(b))$ is the maximum distance apart of any pair of points in the segment. Further, replacing α by $\alpha \circ t_a$ or $\alpha \circ r_a$, as appropriate, we can assume $0 = a \leq b$. If $\beta : [0, b]_\Lambda \to X$ is another isometry with the same image as α, then $\alpha^{-1} \circ \beta$ is a metric automorphism of $[0, b]_\Lambda$, so by Lemma 2.1 is the restriction of a metric automorphism of Λ. This can only be either $t_0 = id$ or r_b, so either $\beta = \alpha$ or $\beta(x) = \alpha(b - x)$. This explains why we usually need only consider the image of α rather than α itself, which is more convenient in the theory which follows.

Note that, if σ is a segment, and u, $v \in \sigma$, then there is a unique segment τ with endpoints u, v and with $\tau \subseteq \sigma$, obtained by restricting an isometry $\alpha : [a, b]_\Lambda \to X$ with image σ to $[\alpha^{-1}(u), \alpha^{-1}(v)]_\Lambda$. For if $\beta : [a', b']_\Lambda \to \sigma$ is an isometry with $\beta(a') = u$, $\beta(b') = v$, then by Lemma 2.1 $\alpha^{-1} \circ \beta$ is the restriction to $[a', b']_\Lambda$ of a translation or reflection, say γ, and it is easily seen that γ maps $[a', b']_\Lambda$ onto $[\gamma(a'), \gamma(b')]_\Lambda = [\alpha^{-1}(u), \alpha^{-1}(v)]_\Lambda$.

In case $\Lambda = \mathbb{Z}$, $[0, n]_\mathbb{Z} = \{0, 1, \ldots, n\}$, and a segment in a $\mathbb{Z}$-metric space (X, d) is the image of an isometry $\alpha : [0, n]_\mathbb{Z} \to X$ for some $n \geq 0$. Putting $v_i = \alpha(i)$, the segment is a finite set $\{v_0, \ldots, v_n\}$ with $d(v_i, v_j) = |i - j|$.

Definition. A Λ-metric space (X, d) is *geodesic* if for all x, $y \in X$, there is a segment in X with endpoints x, y, and (X, d) is *geodesically linear* (or *geodesically linear*) if, for all x, $y \in X$, there is a unique segment in X whose set of endpoints is $\{x, y\}$.

Exercise. Show that if Λ is an ordered abelian group, then (Λ, d) is a geodesically linear Λ-metric space, where $d(a, b) = |a - b|$, and the segment with endpoints a, b is $[a, b]_\Lambda$.

If σ is a segment in a Λ-metric space with endpoints x, y and $z \in \sigma$ then it is the union of two segments, $\sigma_1 = \{u \in \sigma | d(x, u) \leq d(x, z)\}$ with endpoints x, z, and $\sigma_2 = \{u \in \sigma | d(x, u) \geq d(x, z)\}$ with endpoints z, y. Further, $\sigma_1 \cap \sigma_2 = \{z\}$ and if $u \in \sigma_1$, $v \in \sigma_2$ then $d(u, v) = d(u, z) + d(z, v)$. In particular, $d(x, y) = d(x, z) + d(z, y)$. These assertions follow easily from the fact that Λ is an ordered abelian group.

Lemma 2.2. *Let (X, d) be a Λ-metric space.*

(1) *Let σ be a segment in X with endpoints x, z and let τ be a segment in X with endpoints y, z.*
   (a) *Suppose that, for all $u \in \sigma$ and $v \in \tau$, $d(u,v) = d(u,z) + d(z,v)$. Then $\sigma \cup \tau$ is a segment with endpoints x and y;*
   (b) *if $\sigma \cap \tau = \{z\}$ and $\sigma \cup \tau$ is a segment, then its endpoints are x, y.*

(2) *Assume (X, d) is geodesically linear. Let x, y and $z \in X$, and let σ be the segment with endpoints x, y. Then $z \in \sigma$ if and only if $d(x, y) = d(x, z) + d(z, y)$.*

Proof. (1)(a) By using isometries of Λ we can find surjective isometries $\alpha : [0, a]_\Lambda \to \sigma$ and $\beta : [a, b]_\Lambda \to \tau$, where $0 \leq a \leq b$, with $\alpha(a) = \beta(a) = z$. Then $\alpha \cup \beta$ is an isometry with image $\sigma \cup \tau$ and endpoints $\alpha(0) = x$, $\beta(b) = y$.

(b) If $\alpha : [a, b]_\Lambda \to X$ is an isometry with image $\sigma \cup \tau$, then by remarks made above, $\sigma = \alpha([\alpha^{-1}(x), \alpha^{-1}(z)]_\Lambda)$ and $\tau = \alpha([\alpha^{-1}(y), \alpha^{-1}(z)]_\Lambda)$. Hence $[\alpha^{-1}(x), \alpha^{-1}(z)]_\Lambda$, $[\alpha^{-1}(y), \alpha^{-1}(z)]_\Lambda$ intersect only in $\alpha^{-1}(z)$ and their union is $[a, b]_\Lambda$. It follows easily that their union is $[\alpha^{-1}(x), \alpha^{-1}(y)]_\Lambda$ and also that $\{\alpha^{-1}(x), \alpha^{-1}(y)\} = \{a, b\}$, so x, y are the endpoints of $\sigma \cup \tau$.

(2) Assume $d(x, y) = d(x, z) + d(z, y)$, let τ be the segment with endpoints x, z and let ρ be the segment with endpoints z, y. Take $u \in \tau$, $v \in \rho$. Then

$$d(u, z) = d(x, z) - d(x, u) \qquad \text{and} \qquad d(z, v) = d(z, y) - d(y, v).$$

Using this and the triangle inequality,

$$d(u, v) \leq d(u, z) + d(z, v) = d(x, y) - d(x, u) - d(y, v) \leq d(u, v)$$

so there is equality throughout, and by Part (1)(a) $\tau \cup \rho$ is a segment with endpoints x, y. Hence $\sigma = \tau \cup \rho$ and $z \in \sigma$. The converse has already been noted. □

The next lemma appears in [106; Lemma II.1.1], and will be of interest when considering the axioms defining a Λ-tree in Chapter 2.

Lemma 2.3. *Let Λ be an ordered abelian group and let (X, d) be a geodesically linear Λ-metric space. Then the intersection of two segments with a common endpoint u is the image under an isometry of a convex subset of Λ with least element 0. If the intersection is a segment, then x is an endpoint of the intersection.*

If $\Lambda = \mathbb{Z}$ or $\Lambda = \mathbb{R}$ then the intersection is always a segment.

Proof. Let σ and τ be segments with a common endpoint u. Using appropriate reflections and translations, we can find surjective isometries $\alpha : [0,a]_\Lambda \to \sigma$, $\beta : [0,b]_\Lambda \to \tau$ with $\alpha(0) = \beta(0) = u$ and a, $b \geq 0$. Define $I = \{c \in [0, \min\{a,b\}] \mid \alpha(c) = \beta(c)\}$. Clearly $\alpha(I) = \beta(I) = \sigma \cap \tau$. If x, $y \in I$ and $z \in \Lambda$ with $x \leq z \leq y$, then $z \in I$, for otherwise the images of $\alpha|_{[x,y]_\Lambda}$ and $\beta|_{[x,y]_\Lambda}$ are distinct segments in X with the same endpoints, contrary to assumption. Thus I is a convex subset of Λ with least element 0. If $\sigma \cap \tau$ is a segment, it is $\gamma([0,e]_\Lambda)$ for some isometry γ and $e \geq 0$. Then $\alpha^{-1} \circ \gamma : [0,e]_\Lambda \to I$ is a metric isomorphism, so I is a segment in Λ with least element 0, hence $I = [0,e]_\Lambda$, and therefore $x = \alpha(0)$ is an endpoint of $\sigma \cap \tau$.

If $\Lambda = \mathbb{Z}$ then I is finite and equal to $[0,n]_\mathbb{Z}$ for some n, and if $\Lambda = \mathbb{R}$, $\sigma \cap \tau$ is a compact connected subset of X homeomorphic to I, so I is $[0,e]_\mathbb{R}$ for some e. □

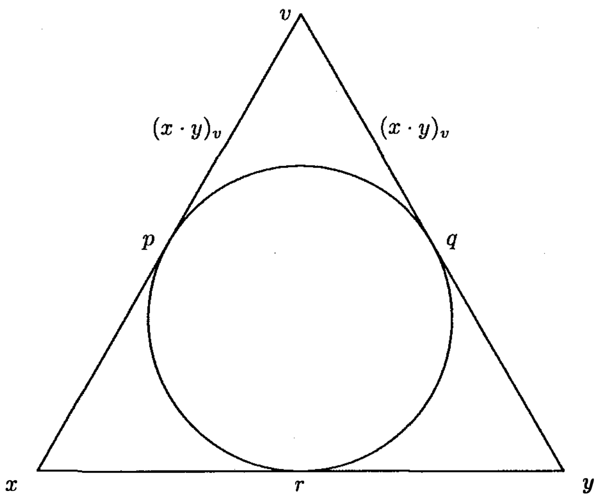
In future, when dealing with a geodesically linear metric space, we shall use the notation $[x,y]$ to denote the segment with endpoints x and y. We shall also occasionally use $[x,y]$ to denote a chosen segment with endpoints x and y, even when the metric space is not geodesically linear.

Remark. It follows from the observations after the definition of segment that, if (X,d) is geodesically linear, and x, $y \in X$, then there is a unique isometry $\alpha : [0,a]_\Lambda \to [x,y]$ such that $\alpha(0) = x$ and $\alpha(a) = y$, where $a = d(x,y)$. Also, if u, $v \in [x,y]$, then $[u,v] \subseteq [x,y]$.

Our final topic in the theory of Λ-metric spaces is the idea of a hyperbolic metric space introduced by Gromov [67]. Let Λ be an ordered abelian group, and suppose (X,d) is a Λ-metric space. Choose a point $v \in X$, and for x, $y \in X$, define

$$(x \cdot y)_v = \tfrac{1}{2}(d(x,v) + d(y,v) - d(x,y)).$$

There is a simple geometric interpretation of $(x \cdot y)_v$ in the case that X is real euclidean space and d is the usual metric. In the following picture, p, q and r are the points of contact of the inscribed circle with the sides of the triangle with vertices x, y and v. Since $d(x,p) = d(x,r)$, $d(y,r) = d(y,q)$ and $d(v,p) = d(v,q)$, it follows easily that $(x \cdot y)_v = d(v,p) = d(v,q)$. There is a similar interpretation in real hyperbolic space. In both cases, it is illustrated by the following picture.



In these examples $(x \cdot y)_v \in \Lambda$, but in general it is an element of $\frac{1}{2}\Lambda$. (For example, take $\Lambda = \mathbb{Z}$, and let X have three points, each at distance 1 from the other two.) Note that, if t is another point of X, we have (by direct calculation) the identity

$$(x \cdot y)_t = d(t, v) + (x \cdot y)_v - (x \cdot t)_v - (y \cdot t)_v \tag{$*$}$$

for all x, $y \in X$. Before proceeding, we record some simple remarks as a lemma, for later use.

Lemma 2.4. (1) *If (X, d) is any Λ-metric space, and v, x, $y \in X$, then* $0 \leq (x \cdot y)_v \leq d(x, v),\ d(y, v)$.

(2) *If (X, d) is a Λ-metric space then the following are equivalent:*

(i) *for some $v \in X$ and all x, $y \in X$, $(x \cdot y)_v \in \Lambda$*

(ii) *for all v, x and $y \in X$, $(x \cdot y)_v \in \Lambda$.*

Proof. (1) follows immediately from the triangle inequality, and (2) follows from equation $(*)$ above. □

We now give Gromov's definition of a hyperbolic space, which is applicable to any Λ-metric space.

Definition. Let $\delta \in \Lambda$ with $\delta \geq 0$. Then (X, d) is δ-hyperbolic with respect to v if, for all x, y, $z \in X$,

$$x \cdot y \geq \min\{x \cdot z, y \cdot z\} - \delta$$

where $x \cdot y$ means $(x \cdot y)_v$, etc.

This means that if we pick out two of $x \cdot y$, $x \cdot z$, $y \cdot z$ which are less than or equal to the third, say a, b, then $|a - b| \leq \delta$. After further elaborating on the definition, we shall comment briefly on its geometrical significance in the case of a geodesic $\mathbb{R}$-metric space.

Lemma 2.5. *If (X, d) is δ-hyperbolic with respect to v, and t is any other point of X, then (X, d) is 2δ-hyperbolic with respect to t.*

Proof. We need to show that $(x \cdot y)_t \geq \min\{(x \cdot z)_t, (y \cdot z)_t\} - 2\delta$, for all x, y, $z \in X$. On adding $(x \cdot t)_v + (y \cdot t)_v + (z \cdot t)_v - d(t, v)$ to both sides and using $(*)$ above, it suffices to show that

$$(x \cdot y)_v + (z \cdot t)_v \geq \min\{(x \cdot z)_v + (y \cdot t)_v, (y \cdot z)_v + (x \cdot t)_v\} - 2\delta$$

for all x, y, z and $t \in X$. Now interchanging x and y if necessary, and also interchanging z and t if appropriate, we can assume that $x \cdot z$ is greater than or equal to $x \cdot t$, $y \cdot z$ and $y \cdot t$, where $x \cdot z$ means $(x \cdot z)_v$, etc. Then by assumption $x \cdot y \geq y \cdot z - \delta$ and $z \cdot t \geq x \cdot t - \delta$. The lemma follows on adding these inequalities. □

Definition. (X, d) is δ-hyperbolic if it is δ-hyperbolic with respect to all points of X, and is hyperbolic if it is δ-hyperbolic for some $\delta \geq 0$.

The definition of δ-hyperbolic can easily be reformulated as follows.

Lemma 2.6. *The Λ-metric space (X, d) is δ-hyperbolic if and only if, for all x, y, z and $t \in X$,*

$$d(x, y) + d(z, t) \leq \max\{d(x, z) + d(y, t), d(y, z) + d(x, t)\} + 2\delta.$$

Proof. Put $a = d(x, y) + d(z, t)$, $b = d(x, z) + d(y, t)$, $c = d(y, z) + d(x, t)$ and $e = d(x, t) + d(y, t) + d(z, t)$. Then $a \leq \max\{b, c\} + 2\delta$ if and only if $e - a \geq \min\{e - b, e - c\} - 2\delta$, and on dividing by 2, this is equivalent to saying that X is δ-hyperbolic with respect to t. □

It is noted in [67] that the notion of hyperbolic metric space can be interpreted geometrically, at least in the case $\Lambda = \mathbb{R}$. If (X, d) is a geodesic $\mathbb{R}$-metric space, then to say that (X, d) is hyperbolic is equivalent to saying that geodesic triangles in X are, in various senses (four, to be exact) "thin". For a full discussion of this see [41; Ch. 1], [61; Ch. 2] or [130; Chs. 1 and 2]. These contain accounts of word hyperbolic (or negatively curved) groups, that is, finitely generated groups whose Cayley diagram is hyperbolic. This is one of the most important class of examples of hyperbolic spaces given in [67]. Another is the class of complete, simply connected Riemannian manifolds, whose sectional curvatures are all at most c for some negative constant c (see Cannon [10; Theorem 11.8]). This applies in particular to real hyperbolic n-space.

We shall see later that Λ-trees are 0-hyperbolic, indeed they can be characterised by this and certain other simple properties. However, we shall not use this as a definition. Instead, we shall arrive at a definition by defining the *path metric* on a connected graph, and studying the special properties of this metric in the case that the graph is a tree. We shall do this in the next section, but first we look in a little more detail at the property of being 0-hyperbolic.

Definition. A triple of elements (a, b, c) of a linearly ordered set is called an isosceles triple if $a \geq \min\{b, c\}$, $b \geq \min\{c, a\}$ and $c \geq \min\{a, b\}$. (This means that at least two of a, b, c are equal, and not greater than the third.)

Note that, by Lemma 2.5, if (X, d) is 0-hyperbolic with respect to one point, then it is 0-hyperbolic with respect to any other point. Also, (X, d) is 0-hyperbolic with respect to v if and only if, for all x, y and $z \in X$, $((x \cdot y)_v$, $(y \cdot z)_v$, $(z \cdot x)_v)$ is an isosceles triple in Λ. Further, by Lemma 2.6, (X,d) is 0-hyperbolic if and only if, for all x, y, z and $t \in X$,

$$d(x,y) + d(z,t) \leq \max\{d(x,z) + d(y,t), d(y,z) + d(x,t)\}.$$

This means that if $a = d(x,y)+d(z,t)$, $b = d(x,z)+d(y,t)$, $c = d(y,z)+d(x,t)$, then at least two of a, b, c are equal and not less than the third, or equivalently $(-a, -b, -c)$ is an isosceles triple, for all x, y, z and $t \in X$. Following [24], we call this the 4-*point condition*. We postpone comment on its geometric significance until Chapter 2.

Remark. If d satisfies the 4-point condition, the triangle inequality follows from this and the other axioms for a metric (put $t = z$). Further, the argument of Lemma 2.5 does not use the triangle inequality, so the triangle inequality follows if we know that for some $v \in X$, $((x \cdot y)_v, (y \cdot z)_v, (z \cdot x)_v)$ is an isosceles triple, for all x, y and $z \in X$.

We close by noting some simple observations on isosceles triples which will be used in Ch. 2.

Lemma 2.7. (1) *If* (a, b, c) *is any triple then* $(\min\{a, b\}, \min\{b, c\}, \min\{c, a\})$ *is an isosceles triple.*

(2) *If* (a, b, c) *and* (d, e, f) *are isosceles triples then so is*

$$(\min\{a, d\},\ \min\{b, e\},\ \min\{c, f\}).$$

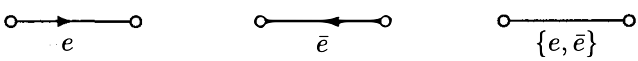
Proof. This is left as an easy exercise (there is a dual version of this in [77]). $\square$

3. Graphs and Simplicial Trees

A Λ-tree is a generalisation of a tree in the usual sense, that is, a connected circuit-free graph. It is a special kind of metric space, where the metric takes values in Λ, which is an ordered abelian group. Any graph induces an integer-valued metric on its vertex set, and to arrive at a definition of Λ-tree we have to look at the special properties of this metric in the case that the graph is a tree. We shall do this later in this introductory section, but we begin by recalling the relevant basic facts about graphs and trees.

There are several different definitions of a graph in the literature, and we start by surveying those we shall use. The definition of a directed graph is standard. It consists of a set V (whose elements are called vertices), a set E (whose elements are called edges), and two mappings $o : E \to V$ and $t : E \to V$. The vertex $o(e)$ is the *origin*, and $t(e)$ is the *terminus* of the edge e. It is required that $V \cap E = \emptyset$.

For us a graph Γ will mean a directed graph with involution, that is, a directed graph with, additionally a mapping $E \to E$, denoted by $e \mapsto \bar{e}$, such that, for all $e \in E$, $\bar{\bar{e}} = e$, $\bar{e} \neq e$, $o(\bar{e}) = t(e)$ and (consequently) $t(\bar{e}) = o(e)$. We write $V(\Gamma)$, $E(\Gamma)$ for V, E. A pair $\{e, \bar{e}\}$ is called an unoriented edge. We call $o(e)$ and $t(e)$ the endpoints of e, and of the corresponding unoriented edge. Vertices are depicted by small circles, and an unoriented edge is depicted by a line segment joining its endpoints. Individual edges e are depicted by the addition of an arrow to the line segment pointing from $o(e)$ to $t(e)$.



Note that a directed graph with vertex set E can be made into a graph in this sense, just by taking a set $\overline{E}$ in one-to one correspondence with E, via a map $E \to \overline{E}$, $e \mapsto \bar{e}$, and with $E \cap \bar{E} = \emptyset$. We then add the elements of $\overline{E}$ to the set of edges, and for $e \in E$, define $o(\bar{e}) = t(e)$, $t(\bar{e}) = o(e)$ and $\bar{\bar{e}} = e$.

Let Γ, Δ be graphs. A morphism of graphs (or graph mapping) from Γ to Δ is a mapping $f : V(\Gamma) \cup E(\Gamma) \to V(\Delta) \cup E(\Delta)$ which maps vertices to vertices and edges to edges, such that, for all edges $e \in V(\Gamma)$, $f(o(e)) = o(f(e))$, $f(t(e)) = t(f(e))$ and $f(\bar{e}) = \overline{f(e)}$. A graph Γ is a subgraph of a graph Δ if $V(\Gamma) \subseteq V(\Delta)$, $E(\Gamma) \subseteq E(\Delta)$ and the inclusion map $V(\Gamma) \cup E(\Gamma) \hookrightarrow V(\Delta) \cup E(\Delta)$ is a graph mapping, that is, if $e \in E(\Gamma)$ then $o(e)$, $t(e)$ and $\bar{e}$ have the same meaning in Δ as they do in Γ. The graph mapping f is called an isomorphism if it is bijective, and an isomorphism from Γ to Γ is called an automorphism of Γ.

Note that there can be more than one unoriented edge with the same endpoints in a graph Γ. If this does not happen we say that Γ has no multiple edges. Also, there is nothing to prevent the occurrence of an edge e with $o(e) = t(e)$. An edge with this property is called a loop.

Graphs in this sense may be viewed as combinatorial descriptions of the class of topological spaces known as 1-dimensional CW-complexes. (See [98; Ch.6], where these spaces are called graphs; we shall show in §2 of Chapter 2 how to associate a 1-dimensional CW-complex with a graph in our sense.) A special case is the class of 1-dimensional polyhedra, that is, geometric realisations of abstract 1-dimensional simplicial complexes. Such a simplicial complex can be described as a set V (the set of vertices) together with a set S of two-element subsets of V (the set of 1-simplices). Given such a simplicial complex, we can construct a graph Γ by defining $V(\Gamma) = V$, $E(\Gamma) = \{(u,v) \in V \times V \mid \{u,v\} \in S\}$ and $o(u,v) = u$, $t(u,v) = v$, $\overline{(u,v)} = (v,u)$. This graph has no loops or multiple edges. Conversely, given a graph Γ with no loops or multiple edges, we may define a 1-dimensional simplicial complex by putting $V = V(\Gamma)$ and

$$S = \{\ \{u,v\} \mid \text{there is an edge in } \Gamma \text{ with endpoints } u,\ v\}.$$

For this reason, we shall call a graph with no loops or multiple edges a simplicial graph.

We are interested mainly in connected graphs, which are the combinatorial analogues of path-connected topological spaces, and are defined as follows. Let L_n be the simplicial graph corresponding to the simplicial complex with vertex set $\{0,1,\ldots,n\}$ and 1-simplices $\{0,1\},\{1,2\},\ldots,\{n-1,n\}$. Here n is a non-negative integer.

L_n: 

We define a path of length n in a graph Γ to be a graph mapping $p : L_n \to X$. We denote $p(0)$ by $o(p)$ and $p(n)$ by $t(p)$ and call these vertices the endpoints of p (they may coincide). We say that p *joins* $o(p)$ to $t(p)$. A graph Γ is called connected if, given x, $y \in V(\Gamma)$, there exists a path in Γ which joins x to y.

This is perhaps the most elegant definition of path since it mimics the definition for topological spaces. However, a path p of length n can be viewed as a sequence $v_0, e_1, v_1, \ldots, e_n, v_n$, where e_i is an edge, $t(e_i) = v_i$ and $o(e_i) = v_{i-1}$ for $1 \le i \le n$. (Formally, the edge e_i is $p(i-1,i)$ and v_i is $p(i)$.) We shall call the v_i the vertices, and the e_i the edges which *belong* to the path. For each vertex v of Γ there is a unique path of length 0 with v as its only endpoint, which is called the trivial path at v, while for a path p of length $n > 0$ we can omit mention of the vertices v_i and think of it as a sequence of edges $e_1, \ldots, e_n$ of Γ with $t(e_i) = o(e_{i+1})$ for $1 \le i \le n-1$, and $o(p) = o(e_1)$, $t(p) = t(e_n)$. This will usually be the most convenient way to think of paths; it should be clear from the notation whether a path is viewed as a sequence of edges, or as a sequence of edges and vertices. Given two paths $p = v_0, e_1, \ldots, e_n, v_n$ and $q = w_0, f_1, \ldots, f_m, w_m$, with $t(p) = o(q)$, the product pq is defined to be $v_0, e_1, \ldots, e_n, v_n, f_1, \ldots, f_m, w_m$, a path from $o(p)$ to $t(q)$. In terms of edge sequences, this can be written as the concatenation $e_1, \ldots, e_n, f_1, \ldots, f_m$, noting that if q is trivial, $pq = p$, and if p is trivial then $pq = q$.

A path $e_1, \ldots, e_n$ is called *reduced* if $e_i \neq \bar{e}_{i+1}$ for $1 \leq i \leq n-1$ and a path is called *closed* if its endpoints coincide (paths of length 0 are regarded as reduced). A *circuit* is a closed reduced path of positive length, $e_1, \ldots, e_n$ such that $o(e_i) \neq o(e_j)$ for $1 \leq i, j \leq n$ with $i \neq j$. The image of such a path is the combinatorial analogue of a homeomorphic image of a circle. We can now define a tree.

Definition. A tree is a connected graph having no circuits.

Note that a graph is simplicial if and only if it has no circuits of length at most 2. Thus trees are simplicial graphs, and we shall sometimes refer to them as simplicial trees to distinguish them from the Λ-trees which will be considered later. We use the term subtree to mean a subgraph which is also a tree.

Remark. If $e_1, \ldots, e_n$ is a path in a graph Γ and $e_{i+1} = \bar{e}_i$ for some i, then removing e_i, e_{i+1} gives a path of shorter length joining the same vertices. Thus paths of shortest length joining two given vertices are reduced.

The only fact about trees which will be needed in the first five chapters is the characterisation in the following lemma.

Lemma 3.1. *A graph Γ is a tree if and only if, given vertices u, v of Γ, there is a unique reduced path which joins u to v.*

Proof. By the remark, Γ is connected if and only if, given vertices u, v, there is a reduced path joining u to v, and it suffices to show that Γ is circuit-free if and only if, given u and v, there is at most one reduced path joining u to v.

If Γ contains a circuit $e_1, \ldots, e_n$, then $e_1, \ldots, e_{n-1}$ and $\bar{e}_n$ are distinct reduced paths from $o(e_1)$ to $o(e_n)$ (if $n = 1$, the first path is the trivial path at $o(e_1)$).

Conversely, suppose $e_1, \ldots, e_n$ and $f_1, \ldots, f_m$ are two distinct reduced paths joining the same pair of vertices, and choose them so that $m + n$ is as small as possible (again if $n = 0$ the first path is the trivial path at $o(f_1)$, and similarly if $m = 0$). Then $e_1, \ldots, e_n, \bar{f}_m, \ldots, \bar{f}_1$ is clearly a non-trivial closed path in Γ, and is reduced. For otherwise $e_n = f_m$, hence $e_1, \ldots, e_{n-1}$ and $f_1, \ldots, f_{m-1}$ are distinct reduced paths joining the same pair of vertices, contradicting the minimality of $m+n$. Thus we can choose a non-trivial closed reduced path in Γ, say $g_1, \ldots, g_k$. Now choose i, j with $1 \leq i \leq j \leq k$ such that $g_i, \ldots, g_j$ is a non-trivial closed path, with $j - i$ as small as possible. It is easily verified that $g_i, \ldots, g_j$ is a circuit in Γ. The lemma follows. □

We introduce some more terminology concerning graphs. The degree of a vertex v in a graph Γ is defined to be the number of edges e of Γ such that $o(e) = v$. A vertex of degree 1 is called a terminal vertex of Γ, and an edge having an endpoint which is a terminal vertex is called a terminal edge of Γ.

Given a collection of subgraphs $\{\Gamma_i \mid i \in I\}$ of a graph Γ, there is clearly a subgraph of Γ, denoted by $\bigcap_{i\in I}\Gamma_i$, with $V(\bigcap_{i\in I}\Gamma_i) = \bigcap_{i\in I}V(\Gamma_i)$ and $E(\bigcap_{i\in I}\Gamma_i) = \bigcap_{i\in I}E(\Gamma_i)$, which we call the intersection of the Γ_i. Similarly we can define the union $\bigcup_{i\in I}\Gamma_i$. If X is a subset of $V(\Gamma)\cup E(\Gamma)$, the subgraph *generated* or *spanned* by X is defined to be the intersection of all subgraphs of Γ containing X.

In Chapter 6 some extra results on trees will be needed, including the following. A *maximal tree* in a graph Γ is a subtree of Γ which is not properly contained in any other subtree of Γ.

Lemma 3.2. *Let Γ be a connected graph. Then a subtree T of Γ is maximal if and only if $V(T) = V(\Gamma)$. Further, Γ has a maximal subtree.*

Proof. Let T be a subtree of Γ with $V(T) \neq V(\Gamma)$, and take $v \in V(\Gamma)\setminus V(T)$. If T is empty, the single vertex v is a subtree of Γ containing T, so T is not maximal. Otherwise, choose $u \in V(T)$ and take a path p in Γ joining v to u, say $p = v_0, e_1, v_1, \ldots, e_n, v_n$, where $v_0 = v$, $v_n = u$. Choose i as large as possible such that $v_i \notin V(T)$, so $v_{i+1} \in V(T)$ (such an I exists since $v_0 \notin V(T)$, and $i < n$ since $v_n \in V(T)$). Now add the edges e_{i+1}, $\bar{e}_{i+1}$ and the vertex v_i to T, to obtain a subgraph T' of Γ containing T. It is easily checked that T' is a tree, using the fact that v_i is a terminal vertex of T', hence T is not maximal.

Conversely, if T is a subtree of Γ with $V(T) = V(\Gamma)$ and Δ is a subgraph of Γ strictly containing T, there is an edge $e \in E(\Delta)\setminus E(T)$. There is a reduced path p in T joining $o(e)$ to $t(e)$, and the path $p, t(e), \bar{e}, o(e)$ is a closed reduced path in Δ, as is the trivial path at $o(e)$, so Δ is not a tree by Lemma 3.1.

To show existence, consider the set Ω of all subtrees of Γ, ordered by inclusion. Given a chain $\{T_i \mid i \in I\}$, where I is linearly ordered and $i \leq j$ implies T_i is a subtree of T_j, let $T = \bigcup_{i\in I} T_i$. We claim T is a subtree of Γ. For if u, $v \in V(T)$, then u, $v \in T_i$ for some $i \in I$, and there is a path in T_i, so in T, from u to v. Hence T is connected. If there is a circuit in T, since it has only finitely many edges it is a circuit in T_i for some i, which is impossible. Hence T is a tree, and is clearly an upper bound in Ω for the chain. By Zorn's Lemma, Ω has a maximal element, and the maximal trees of Γ are precisely the maximal elements of Ω. □

Exercise. Show that if Γ is a connected graph, and finite (i.e. $V(\Gamma)\cup E(\Gamma)$ is finite), then it has a maximal tree, without using Zorn's Lemma. (Hint: if there is a circuit, removing an unoriented edge of the circuit from Γ does not disconnect Γ.)

In a general graph Γ (not necessarily connected), the relation $\sim$ on $V(\Gamma)$ defined by: $u \sim v$ if and only if there is a path joining u to v is an equivalence relation. (If $e_1, \ldots, e_n$ is a path joining u to v, and $f_1, \ldots, f_m$ is a path joining v to w, then $e_1, \ldots, e_n, f_1, \ldots, f_m$ is a path joining u to w, and $\bar{e}_n, \ldots, \bar{e}_1$ is a path joining v to u. Also, the trivial path at u joins u to u.) Let W be an equivalence class. If an edge e has one endpoint in W, both endpoints are in W, since e is a path of length 1 in Γ. Hence W together with all edges having an endpoint in W is a subgraph of Γ, called a component of Γ. Clearly an edge or vertex of Γ belongs to exactly one component, the components are connected, and Γ is connected if and only if it has one component.

A *forest* is by definition a circuit-free graph. This is equivalent to saying that its components are trees. A subforest is a subgraph which is a forest, and a maximal forest is a subforest not properly contained in any other subforest. In a general graph Γ, a maximal subforest consists of a union of maximal trees, one in each component of Γ. Note that a subforest F of Γ is maximal if and only if $V(F) = V(\Gamma)$, by Lemma 3.2.

Exercise. Improve Lemma 3.2, by showing that if Γ is a graph, and F is a subforest, there is a maximal subforest of Γ containing F. In particular, if Γ is connected, there is a maximal subtree of Γ containing F.

We shall need another idea, that of a *core graph*, in Chapter 6.

Definition. Let Γ be a connected graph. The core graph of Γ is the subgraph of Γ generated by the set of all edges and vertices which belong to a circuit of Γ.

Given a connected graph Γ, take a maximal tree T of Γ and $v \in V(\Gamma)$. For each $u \in V(\Gamma)$, let p_u be the reduced path in T joining v to u. For $e \in E(\Gamma)$, let p_e be the closed path $p_{o(e)}, o(e), e, t(e), \bar{p}_{t(e)}$. (If $p = (v_0, e_1, v_1, \ldots, e_n, v_n)$ is a path, $\bar{p}$ means the path $(v_n, \bar{e}_n, v_{n-1}, \ldots, \bar{e}_1, v_0)$.) Let Δ be the union of the vertices and edges occurring in p_e, for every $e \in E(\Gamma) \setminus E(T)$, a subgraph of Γ. Then the core graph of Γ is contained in Δ. For a circuit can be written as $e_1, f_1, e_2, f_2, \ldots, e_n, f_n$, where the e_i are the edges of the circuit not in $E(T)$, and f_i is the reduced path in T joining $t(e_i)$ to $o(e_{i+1})$ (where e_{n+1} means e_1). Now $\bar{p}_{t(e_i)} p_{o(e_{i+1})}$ is a path in T and in Δ, joining $t(e_i)$ to $o(e_{i+1})$, and by successive deletion of consecutive pairs of edges e, $\bar{e}$ from this path, we arrive at the reduced path in T joining these two vertices, i.e. at f_i, so all edges and vertices in f_i are in Δ. Also, each e_i and its endpoints belongs to Δ, so the whole circuit is in Δ. Hence the core graph is contained in Δ, as claimed. It follows that, if $E(\Gamma) \setminus E(T)$ is finite, so is the core graph of Γ.

Again assuming the graph Γ is connected, we can define a metric on its vertex set. For x, $y \in V(\Gamma)$, we define $d(x, y)$ to be the minimal length of a path joining x and y. This is easily seen to be a $\mathbb{Z}$-metric on $V(\Gamma)$ (to see the triangle inequality, if $e_1, \ldots, e_n$ is a path joining x and y, and $f_1, \ldots, f_m$ is a path joining y and z, then $e_1, \ldots, e_n, f_1, \ldots, f_m$ is a path joining x and z). We call d the path metric of Γ. We say that Γ is of finite diameter if there exists an integer n such that, for all x, $y \in V(\Gamma)$, $d(x, y) \leq n$, and the least such n is called the diameter of Γ. Before proceeding, we make a remark that will be useful later, in Ch.2.

Remark. If $e_1, \ldots, e_n$ is a path in a graph Γ and $t(e_i)$ is a terminal vertex for some i with $1 \leq i < n$, then $e_{i+1} = \bar{e}_i$, and the remark preceding Lemma 3.1 applies. Hence, if Γ is connected and Γ' is the graph obtained by removing all terminal vertices and edges from Γ, then Γ' is connected. If Γ is a tree then the shortest paths joining vertices are precisely the reduced paths, by Lemma 3.1. If Γ is a tree of finite diameter ≥ 1, the endpoints of a reduced path of maximal length are terminal vertices. Consequently, Γ' is a subtree of smaller diameter.

We also pause to give some easy exercises; these will be used in Ch.6.

Exercises. Let Γ be a finite graph, let v be the number of vertices and e the number of unoriented edges of Γ.

(1) If Γ is connected, show that $e \geq v - 1$, with equality if and only if Γ is a tree. (First prove $e = v - 1$ when Γ is a tree using the remark and induction on v, then use Lemma 3.2.)

(2) In general, let c be the number of components of Γ. If F is a maximal forest of Γ, show that the number of edges of F is $v - c$, and $e \geq v - c$, with equality if and only if Γ is a forest.

Lemma 3.3. (1) *Let Γ be a connected graph, let $X = V(\Gamma)$ and let d be the path metric of Γ. Then (X,d) is a geodesic $\mathbb{Z}$-metric space. Further, a mapping $\alpha : [0, n]_{\mathbb{Z}} \to X$, $i \mapsto v_i$ (where $n > 0$) is an isometry if and only if there is a path $e_1, \ldots, e_n$ in Γ with $v_0 = o(e_1)$, $v_i = t(e_i)$ for $1 \leq i \leq n$ and $d(v_0, v_n) = n$.*

(2) *Let (X, d) be a geodesic $\mathbb{Z}$-metric space. Then there exists a connected simplicial graph Γ such that $X = V(\Gamma)$ and d is the path metric of Γ.*

Proof. (1). Suppose $e_1, \ldots, e_n$ is a path of shortest length joining two vertices u and v, and put $v_i = t(e_i)$, $v_0 = o(e_1) = u$. Then clearly $e_i, \ldots, e_j$ is a path of shortest length joining v_{i-1} to v_j for $i < j$, and $\alpha : [0, n]_{\mathbb{Z}} \to X$ given by $\alpha(i) = v_i$ is an isometry. Hence (X, d) is geodesic. Conversely, given the isometry α, we have $d(v_{i-1}, v_i) = 1$ for $1 \leq i \leq n$, so by definition of d there is an edge e_i (a path of length 1) with $o(e_i) = v_{i-1}$, $t(e_i) = v_i$. Then $e_1, \ldots, e_n$ is a path with the required properties.

(2). Define Γ to be the graph associated to the 1-dimensional simplicial complex with vertex set X, where $\{x, y\}$ is a 1-simplex if and only if $d(x, y) = 1$.

Let u, $v \in X$, let $n = d(x, y)$ and let $\alpha : [0, n]_{\mathbb{Z}} \to X$ be an isometry with $\alpha(0) = u$, $\alpha(n) = v$. Put $v_i = \alpha(i)$. Then $d(v_{i-1}, v_i) = 1$, so there is an unoriented edge with endpoints v_{i-1}, v_i for $1 \leq i \leq n$, hence a path in Γ joining u to v. In particular Γ is connected. If there is a shorter path joining u, v, say $f_1, \ldots, f_m$, then by definition of Γ, $d(o(f_i), t(f_i)) = 1$, so by the triangle inequality $d(u, v) \leq \sum_{i=1}^{m} d(o(f_i), t(f_i)) = m < n$, a contradiction. Hence d is the path metric. □

If in a graph Γ there are two vertices u, v and more than one unoriented edge with these vertices as endpoints, then replacing these edges by a single edge does not change the path metric. Also, it is not changed by removing loops. Thus we can replace Γ by a simplicial graph without changing the path metric. This explains why the graph Γ in Part (1) of the lemma is not assumed simplicial, but the graph Γ in Part (2) is simplicial. In the next result we consider the path metric in a tree.

Lemma 3.4. *Let Γ be a connected simplicial graph, let $X = V(\Gamma)$ and let d be the path metric of Γ. Then Γ is a tree if and only if the following condition is satisfied.*

If two segments of (X, d) intersect in a single point, which is an endpoint of both, then their union is a segment.

Proof. Suppose that the condition is satisfied, and assume that there is a circuit $e_1, \ldots, e_n$ in Γ. Then $n \geq 3$ since Γ is simplicial. Put $v_i = t(e_i)$ for $1 \leq i \leq n$ and $v_0 = o(e_1) = v_n$. We claim that, for $r \leq n-1$, $\{v_0, \ldots, v_r\}$ is a segment with endpoints v_0, v_r. The proof is by induction on r, and it is clearly true if $r \leq 1$. Suppose it is true for r with $r < n-1$. Then $\{v_r, v_{r+1}\}$ is a segment, and $\{v_0, \ldots, v_r\} \cap \{v_r, v_{r+1}\} = \{v_r\}$, so by assumption $\{v_0, \ldots, v_{r+1}\}$ is a segment. Its endpoints are v_0 and v_{r+1}, because v_i is the unique point of this segment at distance i from v_0, for $i \leq r$, hence $d(v_0, v_{r+1}) = r + 1$. Thus $\{v_0, \ldots, v_{n-1}\}$ is a segment. But so is $\{v_0, v_{n-1}\}$ (because the edge e_n has these vertices as its endpoints), and so $d(v_0, v_{n-1}) = n - 1 = 1$, that is, $n = 2$, a contradiction. It follows that Γ is a tree.

Conversely assume Γ is a tree. Recall from the remark preceding Lemma 3.3 that the shortest paths joining vertices are precisely the reduced paths. Let $\sigma = \{v_0, \ldots, v_n\}$, $\tau = \{w_0, \ldots, w_m\}$ be segments, such that $d(v_i, v_j) = |i - j|$ and $d(w_i, w_j) = |i - j|$ for all relevant values of i and j. Assume they intersect only in the single point $v_n = w_0$. We want to show $\sigma \cup \tau$ is a segment, and we can assume $n > 0$ and $m > 0$. By Lemma 3.3(1) they correspond to reduced paths $e_1, \ldots, e_n$ and $f_1, \ldots, f_m$ with $v_0 = o(e_1)$ and $v_i = t(e_i)$ for $i \geq 1$, with similar equations for the w_i. By assumption $v_{n-1} \neq w_1$, so $e_n \neq \bar{f}_1$. Hence $e_1, \ldots, e_n, f_1, \ldots, f_m$ is a reduced path, and the set of endpoints of the edges in this path is a segment by Lemma 3.3(1). This set of endpoints is $\sigma \cup \tau$. □

Lemma 3.5. *Let (X,d) be a $\mathbb{Z}$-metric space. Then there is a tree Γ, such that $X = V(\Gamma)$ and d is the path metric of Γ, if and only if (a) and (b) below hold.*
(a) *(X,d) is geodesic*
(b) *if two segments of (X,d) intersect in a single point, which is an endpoint of both, then their union is a segment.*

Proof. This is immediate from Lemmas 3.3 and 3.4. □

It follows from Lemma 3.1 and Lemma 3.3(1) that, in the situation of Lemma 3.4, (X,d) is a geodesically linear $\mathbb{Z}$-metric space. This has a generalisation.

Lemma 3.6. *Let Λ be an ordered abelian group and let (X,d) be a geodesic Λ-metric space such that the following condition is satisfied.*

If two segments of (X,d) intersect in a single point, which is an endpoint of both, then their union is a segment.

Then (X,d) is a geodesically linear Λ-metric space.

Proof. Let σ, τ be segments, both with endpoints u, v. Let $w \in \sigma$, and let w' be the point of τ such that $d(u,w) = d(u,w')$ (so that $d(v,w) = d(v,w')$). We have to show $w = w'$. Let ρ be a segment with endpoints w, w'. Now $\sigma = \sigma_1 \cup \sigma_2$, where σ_1 is a segment with endpoints u, w, and σ_2 is a segment with endpoints w, v, by the remarks preceding Lemma 2.2. We claim that either $\sigma_1 \cap \rho = \{w\}$ or $\sigma_2 \cap \rho = \{w\}$. For if $x \in \sigma_1 \cap \rho$ and $y \in \sigma_2 \cap \rho$, then $d(x,y) = d(x,w) + d(w,y)$ (again by the remarks preceding Lemma 2.2), and either $d(w,y) = d(w,x) + d(x,y)$ or $d(w,x) = d(w,y) + d(x,y)$, depending on how x, y are situated in the segment ρ. It follows that either $x = w$ or $y = w$, establishing the claim. If $\sigma_1 \cap \rho = \{w\}$, then by assumption $\sigma_1 \cup \rho$ is a segment, and by Lemma 2.2(1) its endpoints are u, w'. Since $w \in \sigma_1 \cup \rho$, $d(u,w') = d(u,w) + d(w,w')$, so $w = w'$. Similarly if $\sigma_2 \cap \rho = \{w\}$ then $w = w'$. □

If G is a group acting as automorphisms on a connected graph Γ, and d is the path metric of Γ, then G acts as isometries on the $\mathbb{Z}$-metric space (X,d), where $X = V(\Gamma)$, because G preserves shortest paths. Conversely, if G is a group acting as isometries on a geodesic $\mathbb{Z}$-metric space (X,d), then G acts as automorphisms of the corresponding simplicial graph in Lemma 3.3(2), because vertices are adjacent if and only if they are at distance 1 apart. (Two vertices of a graph are *adjacent* if there is an edge having them as its endpoints.)

We end with a caution. A graph such that, for any vertices u, v there is a unique shortest path joining u, v is not necessarily a tree. A counterexample is provided by (the image of) a circuit of odd length.

4. Valuations

The material in this section will not be needed until Ch. 4 §2, and can be omitted on a first reading. We denote the group of units of a ring R by R^*.

Definition. An *absolute value* on a field F is a mapping $F \to K$, $a \mapsto |a|$, where K is an ordered field, such that for all a, $b \in F$,

(i) $|a| \geq 0$, and $|a| = 0$ if and only if $a = 0$;
(ii) $|a + b| \leq |a| + |b|$;
(iii) $|a||b| = |ab|$.

For example, if F is an ordered field, we have noted that $|a| = \max\{a, -a\}$ gives an absolute value on F. The absolute value is called non-archimedean if it satisfies the stronger condition

(ii)′ $|a + b| \leq \max\{|a|, |b|\}$.

This second use of the term archimedean is unfortunate, but is well-established. The reason for it is that an absolute value $|\ |$ is non-archimedean if and only if, for all $n \in \mathbb{Z}$, $|n1_F| \leq 1_K$, where 1_F, 1_K are the multiplicative identity elements of F, K. We leave the proof as an easy exercise. The idea of a valuation is a generalisation of the notion of a non-archimedean absolute value. It is also unfortunate that, in the definition, the inequality used in (ii)′ is reversed, but again this is well-established.

Definition. Let F be a field. A valuation on F is a group homomorphism $v : F^* \to \Lambda$, where Λ is an ordered abelian group, such that, for all a, $b \in F^*$ with $a + b \neq 0$, $v(a + b) \geq \min\{v(a), v(b)\}$.

Thus, given a non-archimedean absolute value as above, we obtain a valuation v by defining $v(a) = |a|^{-1}$. In this case Λ is the positive cone of the additive group of K, and is written multiplicatively. Given a valuation $v : F^* \to \Lambda$, we take an element ∞ not in Λ, with the ordering extended by: $a < \infty$ for all $a \in \Lambda$, and addition extended by: $a + \infty = \infty + a = \infty$ for all $a \in \Lambda \cup \{\infty\}$. We define $v(0) = \infty$; then the inequality $v(a + b) \geq \min\{v(a), v(b)\}$ holds for all a, $b \in F$. The two simplest examples of valuations are the following.

Examples.

(1) Let $F = \mathbb{Q}$, and let $p > 0$ be a prime number in $\mathbb{Z}$. Any non-zero rational number a can be written as $a = \dfrac{m}{n}p^r$, where m, n r are integers, and p does not divide m or n. Putting $v(a) = r$ gives a valuation on $\mathbb{Q}$, which is called the p-adic valuation.

(2) Let F be a field, $F[t]$ the polynomial ring in a single indeterminate t, and let $F(t)$ be its field of quotients. A non-zero element of $F(t)$ can be written as f/g, where f, $g \in F[t]$, and we obtain a valuation by defining $v(f/g) = \deg(g) - \deg(f)$, where deg means the degree of the polynomial in the usual sense.

The next lemma is useful in proving that the mapping v in Example 2 is indeed a valuation.

Lemma 4.1. *Suppose R is an integral domain, with field of fractions F. Let Λ be an ordered abelian group and $v : R \to \Lambda \cup \{\infty\}$ a mapping, such that for all $a,\ b \in R$,*

(i) $v(a) = \infty$ *if and only if* $a = 0$

(ii) $v(a + b) \geq \min\{v(a), v(b)\}$

(iii) $v(ab) = v(a) + v(b)$.

Then there is a unique extension of v to a valuation v' on F.

Proof. If $x \in F$, we can write $x = a/b$, where $a,\ b \in R$, and if v' is a valuation extending v, then $v'(x) = v'(a) - v'(b) = v(a) - v(b)$, showing uniqueness, and we need to show that defining $v'(x) = v(a) - v(b)$ gives a valuation on F. Using the last condition $v(ab) = v(a) + v(b)$, it is easy to see that v' is well-defined, and a homomorphism on F^*. If $x,\ y \in F$, we can write $x = a/c$, $y = b/c$ for some $a,\ b,\ c \in R$. Then

$$\begin{aligned} v'(x + y) = v(a + b) - v(c) &\geq \min\{v(a), v(b)\} - v(c) \\ &= \min\{v(a) - v(c), v(b) - v(c)\} \\ &= \min\{v'(x), v'(y)\}. \end{aligned}$$

This completes the proof. □

Thus applying the lemma with $v(f) = -\deg(f)$, for $f \in F[t]$, gives the valuation in Example 2.

Let v be a valuation on a field F. It is easy to check that that $v(a) = v(-a)$ for all $a \in F$ (by applying v to the equation $a^2 = (-a)^2$). Also, if $a,\ b \in F$, then $(v(a), v(b), v(a+b))$ is an isosceles triple (as defined near the end of §2) in the linearly ordered set $\Lambda \cup \{\infty\}$. In fact, we have the following rather extreme generalisation. We shall not need this result, but it is used in the proof of a result we shall quote in Chapter 5.

Lemma 4.2. *Let R be a ring, Γ a linearly ordered set, $v : R \to \Gamma$ a mapping, such that for all $a,\ b \in R$, $v(a + b) \geq \min\{v(a), v(b)\}$ and $v(a) = v(-a)$.*

Then if $x_1, \ldots, x_n$ are elements of R (where n is a positive integer), we have

$$v(x_1 + \cdots + x_n) \geq \min\{v(x_1), \ldots, v(x_n)\}$$

and if $v(x_1 + \cdots + x_n) > \min\{v(x_1), \ldots, v(x_n)\}$, then $\min\{v(x_1), \ldots, v(x_n)\} = v(x_i) = v(x_j)$ for some $i \neq j$.

Proof. The first assertion follows by induction on n. Suppose there is strict inequality, and the minimum is attained for only one value of i between 1 and

n. Thus $n \geq 2$, and without loss of generality we can assume $i = n$. Then $v(x_1 + \cdots + x_n) > v(x_n)$ and

$$v(x_1 + \cdots + x_{n-1}) \geq \min\{v(x_1), \ldots v(x_{n-1})\} > v(x_n).$$

Let $a = x_1 + \cdots + x_n$, $b = -(x_1 + \cdots + x_{n-1})$. Then

$$v(a + b) \geq \min\{v(a), v(b)\} = \min\{v(a), v(-b)\} > v(a + b),$$

a contradiction. □

Let F be a field, $v : F^* \to \Lambda$ a valuation. Note that $v(F^*)$ is a subgroup of Λ; it is called the value group of v. Define $R = \{a \in F |\ v(a) \geq 0\}$. It is easily checked that R is a subring of F, and it is called the valuation ring of v. Also, if we define $\mathfrak{m} = \{a \in F |\ v(a) > 0\}$, then $\mathfrak{m}$ is an ideal of R. If $a \in F^*$, then since $v(a^{-1}) = -v(a)$, it follows that $a^{-1} \in F \setminus R$ if and only if $a \in \mathfrak{m}$. Thus $R^* = R \setminus \mathfrak{m}$. It follows easily that R is a local ring with $\mathfrak{m}$ as its unique maximal ideal. It also follows that R^* is the kernel of the homomorphism $v : F^* \to \Lambda$. Thus the value group of v is isomorphic to F^*/R^*. Further, if a, $b \in F^*$, then $v(a) \geq v(b)$ if and only if $v(a) - v(b) \geq 0$, that is $v(ab^{-1}) \geq 0$, equivalently $ab^{-1} \in R$. The field $R/\mathfrak{m}$ is called the residue field of v.

Note also that $a \in F \setminus R$ implies $a^{-1} \in R$, and so F is the field of quotients of R. It follows from this and the observations in the previous paragraph that the value group is determined up to isomorphism as an ordered abelian group by the ring R.

A ring is called a valuation ring if it is the valuation ring of some valuation v. The preceding considerations lead to a characterisation of valuation rings. Suppose R is an integral domain, with field of quotients F. For $x \in F^*$, let $\bar{x}$ denote the coset xR^* in F^*/R^*. Then it is easily checked that we can define a partial order on F^*/R^* by $\bar{x} \geq \bar{y}$ if and only if $xy^{-1} \in R$, and that this makes F^*/R^* into a partially ordered abelian group (written multiplicatively).

Lemma 4.3. *In the situation just described the following are equivalent.*
(1) *R is a valuation ring.*
(2) *$a \in F \setminus R$ implies $a^{-1} \in R$.*
(3) *The set of ideals of R is linearly ordered by inclusion.*
(4) *F^*/R^* is linearly ordered by the ordering defined above.*

Proof. We have already observed that (1) implies (2). Assume (2), and let I, J be ideals in R. Suppose $x \in I \setminus J$ and $y \in J \setminus I$. By (2), either $xy^{-1} \in R$ or $yx^{-1} \in R$. If $xy^{-1} \in R$ then $x = (xy^{-1})y \in J$ an contradiction. Similarly if $yx^{-1} \in R$ we obtain the contradiction $y \in I$. Thus (3) holds.

Assume (3), and take x, $y \in F^*$. We can write $x = ab^{-1}$, $y = cd^{-1}$ for some a, b, c, $d \in R$. By (3), either $Rad \subseteq Rbc$ or $Rbc \subseteq Rad$, which implies either $xy^{-1} \in R$ or $yx^{-1} \in R$, so the ordering on F^*/R^* is a linear ordering.

Finally, assume (4), so F^*/R^* is an ordered abelian group. Let $v : F^* \to F^*/R^*$ be the canonical projection, and take x, $y \in F^*$ with $x + y \neq 0$. If $xy^{-1} \in R$ then $v(x) \geq v(y)$ by definition of the ordering. Also, $1 + xy^{-1} \in R$, so $v(1 + xy^{-1}) \geq v(1)$, and we can multiply by $v(y)$ to obtain $v(x + y) \geq v(y) = \min\{v(x), v(y)\}$. Similarly if $yx^{-1} \in R$ we obtain $v(x + y) \geq v(x) = \min\{v(x), v(y)\}$. Thus v is a valuation, and clearly R is the valuation ring of v. □

In the examples we gave above, the value group is infinite cyclic. Valuation rings arising from valuations where the value group is infinite cyclic are called discrete valuation rings, and are of particular interest. There are various ways of characterising them (see [4: Ch. 9] or [81; §9.10]). They arise as localisations of Dedekind domains, and every prime ideal in a Dedekind domain gives a valuation on the field of quotients, generalising the example of the p-adic valuation on $\mathbb{Q}$ (see [81; §10.4]). Discrete valuation rings also occur as local rings at points on a non-singular affine curve ([73; Ch. I, §6]). One characterisation ([81; §9.10]) is that an integral domain is a discrete valuation ring if and only if it is a local principal ideal domain. Thus discrete valuation rings are special kinds of principal ideal domains. There is a well-known structure theory for finitely generated modules over such rings, and we shall now show that there is a similar structure theory for finitely presented modules over any valuation ring.

Let v be a valuation on a field F, let Λ be the value group, and let R be the valuation ring. Let A, B be $m \times n$ matrices over F. We define $A \simeq B$ to mean that there exist $P \in \mathrm{GL}_m(R)$ and $Q \in \mathrm{GL}_n(R)$ such that $PAQ = B$.

By an elementary R-row operation on A we mean one of the following

(i) multiply a row of A by a unit in R;
(ii) add a times one row to another, where $a \in R$.

Note that two rows may be interchanged by a successsion of such operations. Also, an elementary row operation on A leads to the matrix PA, where P is the $m \times m$ matrix obtained by performing the same elementary row operation on I, the $m \times m$ identity matrix, so $P \in \mathrm{GL}_m(R)$.

Elementary R-column operations are defined similarly, and can be performed by postmultiplying by appropriate matrices in $\mathrm{GL}_n(R)$. Thus if B is obtained from A by a succession of elementary row and column operations, then $A \simeq B$.

Lemma 4.4. *Any $m \times n$ matrix A over F (resp. over R) can be transformed by a successsion of elementary R-row and column operations into a diagonal matrix D over F (resp. over R), so $A \simeq D$.*

Proof. Since F is the field of fractions of R, we can write $A = (1/d)A'$, where $d \in R$ and A' has entries in R, so we can assume A has entries in R. Since the principal ideals of R are linearly ordered by inclusion (by Lemma 4.3), there is an entry in A which divides all other entries. Using elementary row and column operations, we can put this entry in the top left-hand corner, then we can make all other entries in the first row and column equal to zero. We can apply induction (on $\max\{m, n\}$) to the matrix B obtained by deleting the first row and column of A to obtain the result. Elementary row and column operations on B correspond to elementary row and column operations on A which do not change the first row or column. $\square$

Corollary 4.5. (1) *suppose L is a finitely generated free R-module, and M is a finitely generated submodule. Then there exist a basis $e_1, \ldots, e_n$ of L and $a_1, \ldots, a_n$ in R such that the elements $a_1e_1, \ldots, a_ne_n$ for which $a_i \neq 0$ form a basis for M.*

(2) *Every finitely presented R-module is isomorphic to $\bigoplus_{i=1}^n R/Ra_i$ for some $n \geq 0$ and $a_i \in R$ $(1 \leq i \leq n)$.*

Proof. (1) Let $x_1, \ldots, x_n$ be a basis for L and let $y_1, \ldots, y_m$ be a finite set of generators for M. Since $y_i \in L$, we can write $y_i = \sum_{j=1}^n a_{ij}x_j$ for $1 \leq i \leq m$, where $a_{ij} \in R$. Let $A = (a_{ij})$ be the $m \times n$ matrix whose i, j entry is a_{ij}. By 4.4, there exist $P \in \mathrm{GL}_m(R)$ and $Q \in \mathrm{GL}_n(R)$ such that $PAQ = D$, where D is a diagonal matrix over R. Write $P = (p_{ij})$, $P^{-1} = (\bar{p}_{ij})$, $Q = (q_{ij})$ and $Q^{-1} = (\bar{q}_{ij})$. Put $e_i = \sum_{j=1}^n \bar{q}_{ij}x_j$ and $f_i = \sum_{j=1}^m p_{ij}y_i$. Then $f_i \in M$, and M is generated by $f_1, \ldots, f_m$, because $\sum_{j=1}^m \bar{p}_{ij}f_j = \sum_{j,k} \bar{p}_{ij}p_{jk}y_k = y_i$, so each y_i is in the submodule generated by $f_1, \ldots, f_m$. Similarly $e_1, \ldots, e_n$ generate L. We can express this by formal matrix equations: writing $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$,

$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$, etc, we have $\mathbf{e} = Q^{-1}\mathbf{x}$, $\mathbf{f} = P\mathbf{y}$, and the calculation just given can be expressed as $\mathbf{y} = P^{-1}\mathbf{f}$. Further, $e_1, \ldots, e_n$ form a basis for L. For suppose $\sum_{i=1}^n r_ie_i = 0$, where $r_i \in R$. Writing $\mathbf{r}$ for the row $(r_1 \cdots r_n)$, we have $\mathbf{re} = 0$. Hence $\mathbf{r}Q^{-1}\mathbf{x} = 0$ (the row with n zeros), and since $x_1, \ldots, x_n$ are linearly independent, $\mathbf{r}Q^{-1} = \mathbf{0}$. On multiplying by Q on the right, we see that $\mathbf{r} = \mathbf{0}$, showing linear independence of $e_1, \ldots, e_n$.

Now $\mathbf{y} = A\mathbf{x}$, so $P\mathbf{y} = (PAQ)Q^{-1}\mathbf{x}$, i.e. $\mathbf{f} = D\mathbf{e}$. Since D is a diagonal matrix, there exist $a_1, \ldots, a_n \in R$ such that $f_i = a_ie_i$ for $1 \leq i \leq m$ (if $n > m$, take $a_i = 0$ for $m + 1 \leq i \leq n$). Then omitting those a_ie_i for which $a_i = 0$ gives a set of generators for M which are linearly independent, because R is an integral domain.

(2) If N is finitely presented then there is an R-epimorphism $\phi : L \to N$, where L is a finitely generated free R-module and $\mathrm{Ker}(\phi)$ is finitely generated. Put

$M = \text{Ker}(\phi)$. By (1), there exist a basis $e_1, \ldots, e_n$ of L and $a_1, \ldots, a_n$ in R such that the $a_1e_1, \ldots, a_ne_n$ with $a_i \neq 0$ form a basis for M. By standard facts on direct products and isomorphism theorems, L/M is isomorphic to $\bigoplus_{i=1}^n R/Ra_i$, and to N. □

Corollary 4.6. *Every finitely generated torsion-free R-module is free.*

Proof. Let L be a finitely generated torsion-free R-module, and let n be the minimum number of generators of L. Thus there is an epimorphism of R-modules $\phi : R^n \to L$. Suppose L is not free, so we can choose $x \in \text{Ker}(\phi)$ with $x \neq 0$. The proof of (1) in Cor. 4.5 shows that there exists a basis $e_1, \ldots, e_n$ of L and $a \in R$ such that ae_1 is a basis for Rx. Thus R^n/Rx is isomorphic to $R^{n-1} \oplus (R/Ra)$, and ϕ induces an R-epimomorphism $\bar{\phi} : R^n/Rx \to L$. Since L is torsion-free, $\bar{\phi}(R/Ra) = 0$, hence L has fewer than n generators, a contradiction. □

This takes care of all the valuation theory we shall need in Chapters 2 to 4. One extra result is needed in Chapter 5, but we shall simply give a reference for this.

This page is intentionally left blank

CHAPTER 2. Λ-trees and their Construction

1. Definition and Elementary Properties

Having investigated $\mathbb{Z}$-metric spaces in §3 of Chapter 1, we are ready to define a Λ-tree. After doing so, we shall consider examples and elementary but important properties of such trees. In what follows, Λ will always denote an ordered abelian group, written additively.

Definition. A Λ-tree is a Λ-metric space (X, d) such that:

(a) (X, d) is geodesic;

(b) if two segments of (X, d) intersect in a single point, which is an endpoint of both, then their union is a segment;

(c) the intersection of two segments with a common endpoint is also a segment.

We can immediately give two examples of Λ-trees.

Example 1. (Λ, d), where d is the usual metric ($d(a, b) = |a - b|$) is a Λ-tree. It is easy to see that a segment in Λ must be of the form $[a, b]_\Lambda$ for some a, $b \in \Lambda$. More generally, any convex subset of Λ is a Λ-tree.

Example 2. A $\mathbb{Z}$-metric space (X, d) is a $\mathbb{Z}$-tree if and only if there is a simplicial tree Γ such that $X = V(\Gamma)$ and d is the path metric of Γ. This follows from Lemmas 1.2.3, 1.3.5 and 1.3.6.

Lemma 1.1. *A Λ-tree is geodesically linear.*

Proof. This follows from Axioms (a) and (b) by Lemma 1.3.6. □

As in Chapter 1, we shall denote the unique segment in a Λ-tree with endpoints x, y by $[x, y]$. As noted in §2 of Chapter 1, the endpoints x, y of a segment are uniquely determined as the pair of points at maximum distance apart in the segment. In Axiom (b), if $[x, y] \cap [x, z] = \{x\}$, then $\sigma = [x, y] \cup [x, z]$ is a segment. By Lemma 1.2.2(1), y, z are the endpoints of σ, that is, $\sigma = [y, z]$. Therefore

(b)$'$ If (X, d) is a Λ-tree, x, y, $z \in X$ and $[x, y] \cap [x, z] = \{x\}$, then $[x, y] \cup [x, z] = [y, z]$.

Also, by Lemma 1.2.3, Axiom (c) for a Λ-tree implies that:

(c)′ If (X, d) is a Λ-tree and $x,\ y,\ z \in X$ then $[x, y] \cap [x, z] = [x, w]$ for some unique $w \in X$.

(Because the intersection has x as an endpoint.) We shall sometimes use the axioms in these generalised forms without comment. Note that, in the cases $\Lambda = \mathbb{R}$ and $\Lambda = \mathbb{Z}$, Axiom (c) follows from (a) and (b), by 1.2.3. On the other hand, it is an easy exercise to show that if a Λ-metric space (X, d) satisfies Axiom (c), then for all $x,\ y \in X$, there is at most one segment in X whose set of endpoints is $\{x, y\}$. The next result is a further elaboration of the axioms; it is essentially the "Y-proposition" of [2], in a slightly generalised form.

Lemma 1.2. *Let x, y and z be points of a Λ-tree (X, d), and let w be the point such that $[x, y] \cap [x, z] = [x, w]$. Then*

(1) $[y, w] \cap [w, z] = \{w\}$, $[y, z] = [y, w] \cup [w, z]$ *and* $[x, y] \cap [w, z] = \{w\}$.

(2) *If y_m is the point on $[x, y]$ at distance m from x, and z_n is the point on $[x, z]$ at distance n from x, where $0 \leq m \leq d(x, y)$ and $0 \leq n \leq d(x, z)$, then $d(y_m, z_n) = m + n - 2\min\{m, n, d(x, w)\}$.*

(3) *w depends only on the set $\{x, y, z\}$, not on the order in which the elements are written.*

Proof. The proof uses Lemma 1.2.2 and the observations preceding it.

(1) Since $y,\ w \in [x, y]$, $[y, w] \subseteq [x, y]$, and similarly $[w, z] \subseteq [x, z]$, so if $u \in [y, w] \cap [w, z]$ then $u \in [x, y] \cap [x, z] = [x, w]$. Hence $u \in [x, w] \cap [y, w] = \{w\}$ (since $w \in [x, y]$). Thus $[y, w] \cap [w, z] = \{w\}$, and $[y, z] = [y, w] \cup [w, z]$ by version (b)′ of Axiom (b). Finally, since $w \in [x, y]$, $[x, y] = [x, w] \cup [w, y]$, so $[x, y] \cap [w, z] = ([x, w] \cap [w, z]) \cup ([y, w] \cap [w, z])$, and both intersections are equal to $\{w\}$ ($w \in [x, z]$).

(2) If $m \leq d(x, w)$ then $y_m,\ z_n \in [x, z]$, and so $d(y_m, z_n) = |m - n|$. Similarly if $n \leq d(x, w)$ then $y_m,\ z_n \in [x, y]$ and $d(y_m, z_n) = |m - n|$. Otherwise, $m > d(x, w)$ and $n > d(x, w)$, so $y_m \in [y, w]$ and $z_n \in [z, w]$, and by (1),

$$\begin{aligned} d(y_m, z_n) &= d(y_m, w) + d(w, z_n) \\ &= (d(x, y_m) - d(x, w)) + (d(x, z_n) - d(x, w)) \\ &= m + n - 2d(x, w). \end{aligned}$$

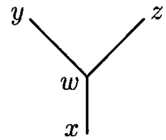
Thus in all cases $d(y_m, z_n) = m + n - 2\min\{m, n, d(x, w)\}$.

(3) We have

$$\begin{aligned} [y, x] \cap [y, z] &= [y, x] \cap ([y, w] \cup [w, z]) \qquad \text{by (1)} \\ &= [y, w] \cup ([y, x] \cap [w, z]) \\ &= [y, w] \cup ([y, w] \cap [w, z]) \cup ([w, x] \cap [w, z]). \end{aligned}$$

Now $[y,w] \cap [w,z] = \{w\}$ by (1) and $[w,x] \cap [w,z] = \{w\}$ since $w \in [x,z]$, hence $[y,x] \cap [y,z] = [y,w]$. Similarly $[z,x] \cap [z,y] = [z,w]$, and (3) follows. □

It follows from (1) and (3) in Lemma 1.2 that $[y,z] \cap [w,x] = \{w\}$ and $[z,x] \cap [w,y] = \{w\}$; also, $[x,y] \cap [y,z] \cap [z,x] = \{w\}$. The segments joining x, y and z are thus illustrated by the following picture, which explains the name "Y-proposition".



We denote the point $w = [x,y] \cap [x,z]$ in Lemma 1.2 by $\mathrm{Y}(x,y,z)$. By 1.2(3), this point is unchanged by permutation of x, y and z. Taking $m = d(x,y)$ and $n = d(x,z)$ in Lemma 1.2(2), we obtain $d(x,w) = (y \cdot z)_x$, in the notation of Ch.1, §2, and by Lemma 1.2(3), $d(y,w) = (x \cdot z)_y$ and $d(z,w) = (x \cdot y)_z$. It follows that $(x \cdot y)_z \in \Lambda$ for all x, y, $z \in X$. Next, we prove a few more elementary but basic facts about segments in Λ-trees. First, a simple consequence of Lemma 1.2.

Corollary 1.3. *If (X,d) is a Λ-tree and $x_0, \ldots, x_n \in X$, then*

$$[x_0, x_n] \subseteq \bigcup_{i=1}^{n} [x_{i-1}, x_i].$$

Proof. If $n = 2$, let $w = \mathrm{Y}(x_0, x_1, x_2)$; then by Lemma 1.2,

$$[x_0, x_2] = [x_0, w] \cup [w, x_2] \subseteq [x_0, x_1] \cup [x_1, x_2].$$

If $n > 2$, then $[x_0, x_n] \subseteq [x_0, x_{n-1}] \cup [x_{n-1}, x_n]$ by the case $n = 2$, and the result follows by induction on n. □

Of course in Corollary 1.3 $d(x_0, x_n) \leq \sum_{i=1}^{n} d(x_{i-1}, x_i)$, by the triangle inequality. We investigate when there is equality.

We write $[x_0, x_n] = [x_0, x_1, \ldots x_n]$ to mean that, if $\alpha : [0, d(x_0, x_n)]_\Lambda \to X$ is the unique isometry with $\alpha(0) = x_0$ and $\alpha(d(x_0, x_n)) = x_n$, then $x_i = \alpha(a_i)$, where $0 = a_0 \leq a_1 \leq \cdots \leq a_n = d(x_0, x_n)$. (We are using the remark after Lemma 1.2.3.) To put it another way, $x_0, x_1, \ldots x_n$ are elements of $[x_0, x_n]$ listed in order of increasing distance from x_0. We shall also say that $[x_0, x_1, \ldots x_n]$ is a piecewise notation for a segment, or simply a piecewise segment, instead of writing $[x_0, x_n] = [x_0, x_1, \ldots x_n]$; indeed, we shall sometimes just write $[x_0, x_1, \ldots x_n]$ alone, meaning that it is a piecewise segment. Note that, if $[x_0, x_1, \ldots x_n]$ is a piecewise segment then $[x_i, x_j] = [x_i, x_{i+1}, \ldots x_j]$

for $0 \le i \le j \le n$ and $[x_0, x_n] = \bigcup_{i=1}^{n}[x_{i-1}, x_i]$ (by induction on n, using the remarks preceding Lemma 1.2.2); also, if $y \in [x_{i-1}, x_i]$, then $[x_0, x_n] = [x_0, \ldots x_{i-1}, y, x_i, \ldots x_n]$. These are observations which will be used frequently without comment.

Remark. (cf [2; Cor.2.12].) For points x, y, z in a Λ-tree, the following are equivalent.

(1) $[x, y] \cap [y, z] = \{y\}$;

(2) $[x, z] = [x, y] \cup [y, z]$;

(3) $[x, z] = [x, y, z]$ i.e. $y \in [x, z]$.

This follows from Axiom (b)′ for a Λ-tree and the remarks preceding Lemma 1.2.2. Clearly, (1) is also equivalent to

(4) $\mathrm{Y}(x, y, z) = y$

and by the observations preceding Cor.1.3, (4) is equivalent to

(5) $(x \cdot z)_y = 0$, i.e. $d(x, z) = d(x, y) + d(y, z)$.

The equivalence of (3) and (5) can be generalised as follows.

Lemma 1.4. *If (X, d) is a Λ-tree and $x_0, \ldots, x_n \in X$, the following are equivalent.*

(1) $[x_0, x_n] = [x_0, x_1, \ldots x_n]$

(2) $d(x_0, x_n) = \sum_{i=1}^{n} d(x_{i-1}, x_i)$.

Proof. (1) implies (2) by an easy induction using the remark preceding the lemma.

We also use induction on n to show (2) implies (1). This is trivial if $n = 1$ and follows from the preceding remark if $n = 2$. If $n > 2$ then by the triangle inequality,

$$d(x_0, x_n) \le d(x_0, x_{n-1}) + d(x_{n-1}, x_n) \le \sum_{i=1}^{n-1} d(x_{i-1}, x_i) + d(x_{n-1}, x_n)$$
$$= d(x_0, x_n)$$

so $d(x_0, x_{n-1}) = \sum_{i=1}^{n-1} d(x_{i-1}, x_i)$, whence $[x_0, x_{n-1}] = [x_0, x_1, \ldots, x_{n-1}]$ by induction. Also, $d(x_0, x_n) = d(x_0, x_{n-1}) + d(x_{n-1}, x_n)$, so by the case $n = 2$,

$$[x_0, x_n] = [x_0, x_{n-1}, x_n] = [x_0, x_1, \ldots, x_{n-1}, x_n].$$

□

If conditions (1) and (2) of Lemma 1.4 are satisfied, then $[x_{i-1}, x_{i+1}] = [x_{i-1}, x_i, x_{i+1}]$, equivalently $[x_{i-1}, x_i] \cap [x_i, x_{i+1}] = \{x_i\}$ for $1 \le i \le n-1$. This alone does not guarantee that $[x_0, x_n] = [x_0, x_1, \ldots x_n]$. For example, take $X = \Lambda = \mathbb{Z}$, $n = 3$, $x_0 = 0$, $x_1 = x_2 = 1$, and $x_3 = 0$. However, the following lemma is very useful.

Lemma 1.5. *Let (X, d) be a Λ-tree. If $x_0, \ldots, x_n \in X$, $x_i \neq x_{i+1}$ for $1 \leq i \leq n-2$ and $[x_{i-1}, x_i] \cap [x_i, x_{i+1}] = \{x_i\}$ (i.e. $[x_{i-1}, x_i, x_{i+1}]$ is a piecewise segment) for $1 \leq i \leq n-1$, then $[x_0, x_n] = [x_0, x_1, \ldots x_n]$.*

Proof. There is nothing to prove if $n \leq 2$. Suppose $n = 3$. We can assume $x_0 \neq x_1$ and $x_2 \neq x_3$, otherwise there is again nothing to prove. Let $w = \mathrm{Y}(x_0, x_2, x_3)$. Now $w \in [x_0, x_2]$ and $x_1 \in [x_0, x_2]$, so $[x_2, w] \cap [x_2, x_1] = [x_2, v]$, where v is either w or x_1, depending on which is closer to x_2. But $[x_2, w] \cap [x_2, x_1] \subseteq [x_2, x_3] \cap [x_2, x_1] = \{x_2\}$, so $v = x_2$. Since $x_1 \neq x_2$, we conclude that $w = x_2$, hence $[x_0, x_2] \cap [x_2, x_3] = \{x_2\}$, which implies $[x_0, x_3] = [x_0, x_2, x_3] = [x_0, x_1, x_2, x_3]$.

Now suppose $n > 3$. By induction $[x_0, x_{n-1}] = [x_0, x_1, \ldots, x_{n-2}, x_{n-1}] = [x_0, x_{n-2}, x_{n-1}]$. By (i),

$$[x_0, x_n] = [x_0, x_{n-2}, x_{n-1}, x_n] = [x_0, x_1, \ldots, x_{n-2}, x_{n-1}, x_n]$$

as required. □

We shall refer to the process of combining piecewise segments via Lemma 1.5 as *splicing* segments together. The case $n = 3$ is particularly useful; given $[x, y, z]$ and $[y, z, w]$ with $y \neq z$, we can splice to get $[x, w] = [x, y, z, w]$. Together, (1.3), (1.4) and (1.5) constitute the "Piecewise Geodesic Proposition" of Alperin and Bass [2; 2.14]. Now we show that Λ-trees are 0-hyperbolic, in the sense of §1.2.

Lemma 1.6. *Let (X,d) be a Λ-tree. Then (X, d) is 0-hyperbolic, and for all x, y and $v \in X$, $(x \cdot y)_v \in \Lambda$.*

Proof. Fix $v \in X$. We have to show $((x{\cdot}y)_v, (y{\cdot}z)_v, (z{\cdot}x)_v)$ is an isosceles triple, for all x, y and z (see §1.2). Let $q = \mathrm{Y}(x, v, y)$, $r = \mathrm{Y}(y, v, z)$, $s = \mathrm{Y}(z, v, x)$. As noted before Cor.1.3, $(x \cdot y)_v = d(v, q)$ by Lemma 1.2(2) (which establishes the last part of the lemma), and similarly $(y{\cdot}z)_v = d(v, s)$, $(z{\cdot}x)_v = d(v, r)$. We may assume without loss of generality that $d(v, q) \leq d(v, r) \leq d(v, s)$, and have to show that $q = r$. Now r, $s \in [v, z]$ by their definition, and $d(v, r) \leq d(v, s)$, so $[v, s] = [v, r, s]$. Also, by definition of s, $[v, x] = [v, s, x] = [v, r, s, x]$. Hence $r \in [v, x] \cap [v, y] = [v, q]$. Since $d(v, q) \leq d(v, r)$, we have $q = r$ as required. □

We look more closely at the proof of Lemma 1.6, where $q = r$ and $d(v, r) \leq d(v, q)$. Put $p = \mathrm{Y}(x, y, z)$. We shall show that $p = s$. First, observe the following:

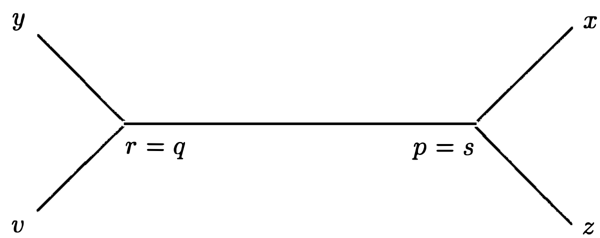
(1) $[v, x] = [v, r, s, x]$	(4) $[y, x] = [y, r, s, x]$
(2) $[v, z] = [v, r, s, z]$	(5) $[y, z] = [y, r, s, z]$
(3) $[v, y] = [v, r, y]$	(6) $[x, z] = [x, s, z]$.

For (1) was proved in the course of proving 1.6, and (3) is immediate from the definition of r. Also, $[v, z] = [v, s, z]$ by definition of s, and $[v, s] = [v, r, s]$

(from (1)) whence (2). Also, (6) is immediate from Lemma 1.2, and by the same lemma, $[y,x] = [y,q,x] = [y,r,x] = [y,r,s,x]$ by (1), and $[y,z] = [y,r,z] = [y,r,s,z]$ by (2).

To prove that $p = s$, we need to show that $[x,y] \cap [x,z] = [x,s]$. In view of (4) and (6), it suffices to show $[y,s] \cap [z,s] = \{s\}$. But this follows from (5).

The situation is illustrated by the diagram below, which will be familiar to those who have worked with simplicial trees, and is intended to represent the union of five segments, some or all of which may be single points.



To see that this diagram accurately reflects the situation, we can prove two further observations. Firstly,

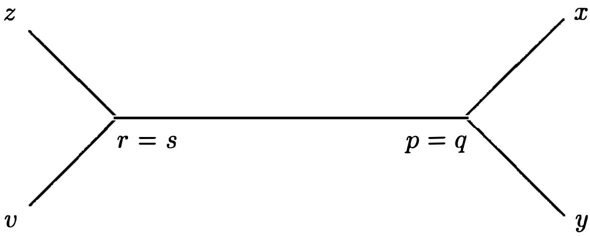
(i) $[v,x] \cap [y,z] = [r,p] = [\mathrm{Y}(v,y,z), \mathrm{Y}(x,y,z)]$.

To prove this, note that $[v,x] \cap [y,r] = [v,x] \cap [y,q] = \{q\} = \{r\}$, by Lemma 1.2 and the definition of q. Also by Lemma 1.2, $[v,x] \cap [z,s] = \{s\}$. The assertion (i) now follows from (1) and (5) above, since $p = s$. We leave it as an exercise to show that $[v,z] \cap [x,y] = [r,p]$. Our second observation is the following.

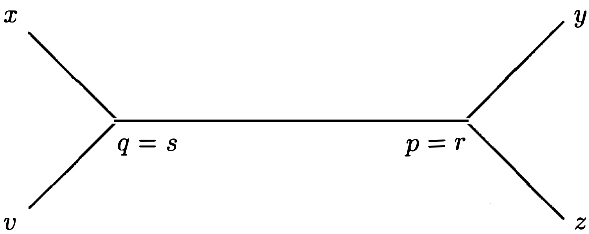
(ii) Either $r \neq s$ and $[v,y] \cap [x,z] = \emptyset$, or $p = q = r = s$ and $[v,y] \cap [x,z] = \{s\}$.

For consider the four intersections $[v,r] \cap [x,s]$, $[v,r] \cap [s,z]$, $[y,r] \cap [x,s]$ and $[y,r] \cap [s,z]$. By (1), (2), (4) and (5) above, these are all empty if $r \neq s$, and all four are equal to $\{s\}$ otherwise. Assertion (ii) now follows from (3) and (6).

In Lemma 1.6, the situation $q = r \leq s$ was obtained by a possible permutation of x, y and z. Given x, y and z, there are therefore two other possible cases (which are not mutually exclusive). The first is $r = s \leq q$, which is obtained from the first case by switching y and z. We conclude that $[v,x] \cap [y,z] = [s,p]$, $p = q$ and $[v,x] \cap [y,z] = \emptyset$ unless $r = q$. The diagram in this case is



The third case, $s = q \leq r$ is obtained by switching x and y in the first case. In this case we obtain $[v, y] \cap [x, z] = [p, q]$, $p = r$ and $[v, x] \cap [y, z] = \emptyset$ unless $p = q$. The diagram for this case is as follows.



The distances between two of x, y, z, v can be read off from these diagrams. If we use them to compute the following:

$$d(x, y) + d(z, v),\ d(x, z) + d(y, v),\ d(x, v) + d(y, z),$$

we find that two of them are equal and not less than the third. (In fact, two are equal and equal to the third plus twice the length of the horizontal segment.) Thus we have a geometric interpretation of the 4-point condition (see the observations before Lemma 1.2.7).

We state a lemma for future use which follows easily from this discussion.

Lemma 1.7. *Let $[x, y]$, $[z, v]$ be segments in a Λ-tree (X, d).*
(1) *If $[x, y] \cap [z, v] \neq \emptyset$, then $[x, y] \cap [z, v] = [p, q]$, where $p = \Upsilon(x, y, z)$ and $q = \Upsilon(x, y, v)$.*
(2) *If $[x, y] \cap [z, v] = \emptyset$, then $[x, z] \cap [y, v] \neq \emptyset$.*

In particular, the intersection of two segments in X is either empty or a segment.

□

Let v be a point of a Λ-tree (X, d). Using Lemma 1.6 we can define the notion of a direction at v, which will be useful in the next section. Define a relation $\equiv$ on the set of segments $\{[v, x] |\ x \in X \setminus \{v\}\ \}$ by $[v, x] \equiv [v, y]$ if and only if $[v, x] \cap [v, y] \neq \{v\}$, that is, $w \neq v$, where $w = \Upsilon(x, v, y)$. Then $\equiv$ is an equivalence relation. It is transitive because $d(v,w) = (x \cdot y)_v$ (as noted before Cor.1.3), and if x, y, $z \in X$, then $((x \cdot y)_v, (y \cdot z)_v, (z \cdot x)_v)$ is an isosceles triple by Lemma 1.6.

Definition. The equivalence classes are called directions (or direction germs) at v. The degree of v, denoted by $\deg(v)$, is defined to be the cardinal number of the set of directions at v. If $\deg(v) \leq 1$, v is called an endpoint of X, if $\deg(v) = 2$ then v is called an edge point of X, and if $\deg(v) \geq 3$ then v is called a branch point of X.

Note that, with this definition, the endpoints of a segment $[x,y]$ are x, y, which is consistent with earlier usage.

Exercise. If Γ is a simplicial tree and (X,d) is the corresponding $\mathbb{Z}$-tree, so $X = V(\Gamma)$ and d is the path metric, show that the degree of $v \in X$ as just defined is the same as the degree of v in the graph Γ, as defined after Lemma 1.3.1.

By a subtree of a Λ-tree (X,d) we mean a convex subset of X, that is, a subset A such that a, $b \in A$ implies $[a,b] \subseteq A$. Then A, with the metric obtained by restriction from d, becomes a Λ-tree. A subtree A is called closed if it is convex-closed, that is, if the intersection of A with any segment of X is either empty or a segment of X.

For example, a segment is a closed subtree by Lemma 1.7, and the subtrees of Λ itself are precisely the convex subsets of Λ, as defined in §1 of Ch.1. An example of a subtree which is not closed is the interval $(0,1)$ in $\mathbb{R}$. The intersection of any collection of subtrees is a subtree, and the intersection of finitely many closed subtrees is a closed subtree. However, the intersection of an arbitrary family of closed subtrees need not be a closed subtree. For example, take $\Lambda = \mathbb{Z} \times \mathbb{Z}$ with the lexicographic ordering, $X = \Lambda$ and let A_i be the segment joining $(0,i)$ and $(1,0)$. Then $\bigcap_{i=0}^{\infty} A_i$ is not a closed subtree (consider its intersection with A_i). Another example is obtained by taking $\Lambda = \mathbb{Q}$, $X = \Lambda$, and A_r to be the segment joining r and 2, where $0 \leq r < \sqrt{2}$, and considering the intersection of $\bigcap_r A_r$ with the segment $[0,2]_{\mathbb{Q}}$. In case $\Lambda = \mathbb{Z}$, $X = V(\Gamma)$ for some simplicial tree Γ, the subtrees of X correspond to subtrees of Γ, and they are all closed.

Definition. If A is a subset of a Λ-tree (X,d), then the subtree of X spanned by A is defined to be the intersection of all subtrees of X containing A. By Lemma 1.2(1), this subtree is equal to $\bigcup_{y \in A}[x,y]$, for any $x \in A$. We say that X is spanned by A if X is equal to the subtree spanned by A.

Lemma 1.8. *If A, B are subtrees of a Λ-tree (X,d) with $A \cap B = \emptyset$, x, $z \in A$ and y, $v \in B$, then $[x,y] \cap [z,v] = [p,q]$, where $p \in A$ and $q \in B$.*

Proof. If $[x,y] \cap [z,v] = \emptyset$ then $[x,z] \cap [y,v] \neq \emptyset$ by Lemma 1.7, which implies

$A \cap B \neq \emptyset$, a contradiction. By Lemma 1.7, $[x,y] \cap [z,v] = [p,q]$, where $p = \mathrm{Y}(x,y,z) \in [x,z]$, so $p \in A$, and $q = \mathrm{Y}(x,y,v) \in [y,v]$, so $q \in B$. □

The main result about closed subtrees is the following lemma

Lemma 1.9. *Let A, B be non-empty subtrees of a Λ-tree (X,d) such that $A \cap B = \emptyset$.*

(a) If A is a closed subtree, there exists a unique element $a \in A$ such that, for all $x \in A$, $y \in B$, $a \in [x,y]$. Further, $[x,y] \cap A = [x,a]$.

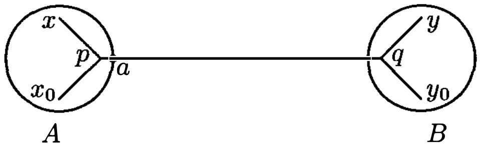
(b) If A and B are closed subtrees, then there exist unique points $a \in A$, $b \in B$ such that $[a,b] \cap A = \{a\}$, $[a,b] \cap B = \{b\}$. Further, if $x \in A$, $y \in B$, then $[a,b] \subseteq [x,y]$.

Proof. Suppose A is a closed subtree, and take $x_0 \in A$, $y_0 \in B$. Since A is closed, it follows easily that $[x_0,y_0] \cap A = [x_0,a]$ for some $a \in A$, and $[a,y_0] \cap A = \{a\}$. Let $x \in A$, $y \in B$. By Lemma 1.8, $[x,y] \cap [x_0,y_0] = [p,q]$, where $p \in A$ and $q \in B$. Now $[x_0,p] \subseteq [x_0,y_0] \cap A = [x_0,a]$, and since $[q,y_0] \subseteq B$, $[q,y_0] \cap A = \emptyset$, hence $[x_0,a] = ([x_0,q] \cup [q,y_0]) \cap A = [x_0,q] \cap A \subseteq [x_0,q]$. Therefore

$$[x_0,p] \subseteq [x_0,a] \subseteq [x_0,q]$$

which implies $[x_0,q] = [x_0,p,a,q]$, so $a \in [p,q] \subseteq [x,y]$. This shows existence of a in (a). If $a' \in A$ also has the property that $a' \in [x,y]$ for all $x \in A$, $y \in B$ then $a' \in [a,y_0] \cap A = \{a\}$, showing uniqueness. Finally, (b) follows easily by applying (a) to both A and B. □

The situation in the proof of Lemma 1.9 is illustrated by the following picture.



Lemmas 1.8 and 1.9 constitute the "Bridge Proposition" in [2]. We call a in 1.9(a) the projection of B onto A, and $[a,b]$ in 1.9(b) is called the bridge between A and B. We also define $d(A,B)$ to be $d(a,b)$, and in case $A \cap B \neq \emptyset$, we put $d(A,B) = 0$. Lemma 1.9 applies in particular when $B = \{b\}$ is a single point and A is a closed subtree, when we call a the projection of b onto A. Also, if A is a closed subtree and $a \in A$, we call $[a,a] = \{a\}$ the bridge between A and a. We shall need the following technical lemma later.

Lemma 1.10. *Suppose A, B and C are closed non-empty subtrees of a Λ-tree (X,d), with $A \cap B = \emptyset$. Let $[a,b]$ be the bridge between A and B, with $a \in A$ and $b \in B$. Then*

$$d(a,C) = \max\{d(A,C), d(B,C) - d(A,B)\}.$$

Proof. Lemma 1.4 will be used several times to compute distances. Let $[a,c]$ be the bridge between a and C ($c = a$ if $a \in C$), and put $w = \mathrm{Y}(a,b,c)$. Since $[a,w] \subseteq [a,b]$, $[a,w] \cap A = \{a\}$, and if $a \neq w$, then $[w,c] \cap A = \emptyset$, since $[a,c] = [a,w,c]$ and $w \notin A$. Also, by definition of c, $[a,c] \cap C = \{c\}$. Hence $[a,c]$ is the bridge between A and C, so $d(a,C) = d(A,C) = d(a,c)$, and $d(B,C) - d(A,B) \leq d(b,c) - d(A,B) = d(b,c) - d(a,b) = d(w,c) - d(w,a) < d(a,c)$.

If $a = w$, and $a \neq c$, then $[b,c] = [b,a,c]$ and $a \notin B \cup C$, hence $[b,c]$ is the bridge between B and C. For $[b,c] \cap B = ([b,a] \cap B) \cup ([a,c]) \cap B) = ([b,a] \cap B) \cup \emptyset = \{b\}$, because if $b' \in [a,c]) \cap B$, then $[b,c] = [b,a,,b',c]$, so $a \in [b,b'] \subseteq B$, a contradiction. Similarly, $[b,c] \cap C = \{c\}$. Thus $d(A,C) \leq d(a,c) = d(a,C) = d(b,c) - d(a,b) = d(B,C) - d(A,B)$.

If $a = c = w$, then $a \in C$, so $d(a,C) = d(A,C) = 0$ and $d(B,C) \leq d(b,a) = d(A,B)$, and the lemma follows. □

As noted in Ch.1, §2, any Λ-metric space (X,d) has a topology with basis the open balls $B(x,r) = \{y \in X \mid d(x,y) < r\}$, where $x \in X$ and $r \in \Lambda$, $r > 0$. Any closed subtree of a Λ-tree is closed in this topology (use Lemma 1.9 to show the complement is open), but the converse is false in general, because of the examples above, where the intersection of a family of closed subtrees is not a closed subtree. However, the converse is true in the case $\Lambda = \mathbb{R}$, for the intersection of a subtree closed in this topology with a segment is a compact connected subset of the segment, hence a segment. We prove one more simple lemma about subtrees of Λ-trees, which shows one way in which they exhibit behaviour like that of subtrees in simplicial trees.

Lemma 1.11. *Suppose $X_1, \ldots, X_n$ are subtrees of a Λ-tree (X,d), such that $X_i \cap X_j \neq \emptyset$ for $1 \leq i,j \leq n$. Then $X_1 \cap \ldots \cap X_n \neq \emptyset$.*

Proof. The proof is by induction on n, the result being obvious for $n \leq 2$. Suppose $n = 3$, and choose $x \in X_1 \cap X_2$, $y \in X_2 \cap X_3$ and $z \in X_3 \cap X_1$. Let $w = \mathrm{Y}(x,y,z)$. Then $w \in [x,y] \cap [y,z] \cap [z,x] \subseteq X_2 \cap X_3 \cap X_1$. If $n > 3$, let $X_3' = X_3 \cap \ldots \cap X_n$. Then $X_1 \cap X_3' \neq \emptyset \neq X_2 \cap X_3'$ by induction, and $X_1 \cap X_2 \cap X_3' \neq \emptyset$ by the case $n = 3$. □

The following exercises will be used later.

Exercises. (1) Given piecewise segments $[x,y,z,w]$ and $[x,y,z',w']$ in a Λ-tree, such that $[y,z] \cap [y,z'] = \{y\}$, show that $[x,w] \cap [x,w'] = [x,y]$.

(2) Let A be a subtree of a Λ-tree (X,d) and suppose $[x,y,z]$ is a piecewise segment in X, where $x \neq y$ and $A \cap [x,y] = \{x\}$. Show that $A \cap [x,z] = \{x\}$.

We shall now prove a series of lemmas giving simple ways in which to construct Λ-trees. They are useful in constructing examples.

Lemma 1.12. *Let (X_1,d_1) and (X_2,d_2) be Λ-trees with $X_1 \cap X_2 = \{x_0\}$, a single point. Let $X = X_1 \cup X_2$ and define $d : X \times X \to \Lambda$ by: $d|_{X_i \times X_i} = d_i$ for $i = 1$, 2, and if $x_i \in X_i$ for $i = 1$, 2 then $d(x_1,x_2) = d(x_2,x_1) = d_1(x_1,x_0) + d_2(x_0,x_2)$. Then (X,d) is a Λ-tree.*

Proof. It is easy to see that d is a metric. We consider only the triangle inequality in detail. Let x, y, $z \in X$; if x, y, z all belong to the same X_i then clearly $d(x,y) \leq d(x,z) + d(z,y)$. If $x \in X_1$, y, $z \in X_2$, then

$$\begin{aligned} d(x,y) &= d_1(x,x_0) + d_2(x_0,y) \\ d(y,z) + d(z,x) &= d_2(y,z) + d_1(x,x_0) + d_2(x_0,z) \end{aligned}$$

and again $d(x,y) \leq d(x,z) + d(z,y)$. The other cases are left to the reader.

Let x, $y \in X$. If they are in the same X_i, then there is a segment with endpoints x, y, otherwise $[x,x_0] \cup [x_0,y]$ (the union of two segments, one in X_1, the other in X_2) is a segment in X with endpoints x, y, by Lemma 1.2.2, Part (1)(a). Thus (X,d) is geodesic. We claim it is geodesically linear. For suppose x, $y \in X_1$, and there is a segment in X joining x, y in X containing a point $z \in X_2 \setminus X_1$. Then

$$\begin{aligned} d_1(x,y) &= d(x,y) \\ &= d(x,z) + d(z,y) \\ &= d_1(x,x_0) + d_2(x_0,z) + d_1(y,x_0) + d_2(x_0,z) \\ &\geq d_1(x,y) + 2d_2(x_0,z) \end{aligned}$$

which implies $d_2(x_0,z) = 0$, so $z = x_0 \in X_1$, a contradiction. Since (X_1,d_1) is geodesically linear, there is only one segment joining x,y. This is similarly true if x, y both belong to X_2. Otherwise, we can assume $x \in X_1$, $y \in X_2$. If a segment joining x, y contains x_0 it must be the segment $[x,x_0] \cup [x_0,y]$ as above since (X_1,d_1) and (X_2,d_2) are geodesically linear. If σ is a segment in X with endpoints x, y, let z be the point of σ such that $d(x,z) = d(x,x_0)$, so that $d(z,y) = d(x_0,y)$. If $z \in X_1$, then by definition of d, $d(z,y) = d(z,x_0) + d(x_0,y)$, while if $z \in X_2$, then $d(z,x) = d(z,x_0) + d(x_0,x)$. In any case $d(z,x_0) = 0$, so $x_0 = z \in \sigma$. We now verify the remaining axioms (b) and (c) (in fact the strengthened forms (b)′ and (c)′) for a Λ-tree.

(b)′ Suppose x, y, $z \in X$ and $[x,y] \cap [x,z] = \{x\}$. We show that $[y,z] = [x,y] \cup [x,z]$. This is clear if all three points belong to the same X_i.

Otherwise we may assume without loss of generality that $x \in X_1$, and by symmetry there are only two cases we need to consider.

Case 1. $y \in X_1$, $z \in X_2$. Then from the above, $\{x\} = [x,y] \cap ([x,x_0] \cup [x_0,z])$ hence $\{x\} = [x,y] \cap [x,x_0]$. Since (X_1,d_1) is a Λ-tree, $[y,x_0] = [y,x] \cup [x,x_0]$, so

$$\begin{aligned}[y,z] &= [y,x_0] \cup [x_0,z] \\ &= [y,x] \cup [x,x_0] \cup [x_0,z] \\ &= [y,x] \cup [x,z].\end{aligned}$$

Case 2. y, $z \in X_2$. Then

$$\begin{aligned}\{x\} &= ([x,x_0] \cup [x_0,y]) \cap ([x,x_0] \cup [x_0,z]) \\ &= [x,x_0] \cup ([x_0,y] \cap [x_0,z])\end{aligned}$$

so $x = x_0$ (since the right-hand side contains x_0) and x, y, z are all in X_2.

(c)′ Suppose x, y, $z \in X$. We have to show that $[x,y] \cap [x,z] = [x,w]$ for some $w \in X$. As before, we can assume $x \in X_1$ and need consider only two cases.

Case 1. $y \in X_1$, $z \in X_2$. Then $[x,z] = [x,x_0] \cup [x_0,z]$ and $[x,z] \cap [x,y] = [x,x_0] \cap [x,y]$ since $[x_0,z] \cap [x,y] \subseteq X_2 \cap X_1 = \{x_0\}$, and $x_0 \in [x,x_0]$. Since (X_1,d_1) satisfies Axiom (c)′, $[x,x_0] \cap [x,y] = [x,w]$ for some $w \in X_1$.

Case 2. y, $z \in X_2$. Then

$$\begin{aligned}[x,y] \cap [x,z] &= [x,x_0] \cup ([x_0,y] \cap [x_0,z]) \\ &= [x,x_0] \cup [x_0,w] = [x,w]\end{aligned}$$

for some $w \in X_2$, since (X_2,d_2) satisfies Axiom (c)′.

This concludes the proof. □

The next lemma gives a generalisation of 1.12.

Lemma 1.13. *Let (X,d) be a Λ-tree, and let $\{(X_j,d_j) \mid j \in J\}$ be a family of Λ-trees, such that, for all $j \in J$, $X_j \cap X = \{x_j\}$, a single point. Also, suppose that $X_j \cap X_k = \{x_j\}$ if $x_j = x_k$, and $X_j \cap X_k = \emptyset$ otherwise. Put $Y = \left(\bigcup_{j\in J} X_j\right) \cup X$, and define $\bar{d} : Y \times Y \to \Lambda$ as follows: $\bar{d}|_{X\times X} = d$, $\bar{d}|_{X_j\times X_j} = d_j$. If $x \in X_j$, $x' \in X_k$ with $j \neq k$, then $\bar{d}(x,x') = d_j(x,x_j) + d(x_j,x_k) + d_k(x_k,x')$. Finally if $x \in X_j$, $x' \in X$, then $\bar{d}(x,x') = \bar{d}(x',x) = d_j(x,x_j) + d(x_j,x')$. Then $(Y,\bar{d})$ is a Λ-tree.*

Proof. If j, k, l are indices in J, then $X_j \cup X_k \cup X_l \cup X$, with the metric obtained from $\bar{d}$ by restriction, can be built up in stages using Lemma 1.12, so is a Λ-tree. Further, any set of at most three points of Y is contained in $X_j \cup X_k \cup X_l \cup X$ for some suitable values of j, k and l. It follows that $(Y,\bar{d})$ is a geodesic Λ-metric space, and the lemma will follow if we can show it is geodesically linear. Suppose x, $y \in Y$. Then x, $y \in X_j \cup X_k \cup X$ for some j, k. Suppose there is a segment in Y joining x, y which contains a point in X_l, with $l \neq j$, k. Then the argument in the previous lemma to show uniqueness of segments (with $X_1 = X_j \cup X_k \cup X$, $X_2 = X_l$) gives a contradiction. Uniqueness of segments in Y follows since $X_j \cup X_k \cup X$ is a Λ-tree. □

The last of this series of lemmas is proved in a similar manner.

Lemma 1.14. *Let (X_i, d_i) $(i \in \mathbb{N})$ be an ascending chain of Λ-trees. (Thus $X_0 \subseteq X_1 \subseteq \cdots$, and for $i < j$, d_i is d_j restricted to $X_i \times X_i$.) Put $X = \bigcup_{i \geq 0} X_i$, $d = \cup_{i \geq 0} d_i$. Then (X,d) is a Λ-tree.*

Proof. Any three points of X are contained in some X_i, so (X,d) is a geodesic Λ-metric space and as in the previous lemma, it suffices to show that it is geodesically linear. Let x, $y \in X$, so x, $y \in X_i$ for some i. Suppose there is a segment in X with endpoints x, y containing a point $z \notin X_i$. Then x, y, $z \in X_j$ for some $j > i$. In X_j, let $w = \mathrm{Y}(x,y,z)$. Then using the remarks before Lemma 1.2.2,

$$\begin{aligned} d(x,y) &= d(x,z) + d(z,y) \\ &= d(x,w) + d(w,z) + d(y,w) + d(w,z) \\ &= d(x,y) + 2d(z,w) \end{aligned}$$

hence $z = w$. The segment joining x and y in X_i must also be the segment $[x,y]$ joining them in X_j since X_j is geodesically linear, and $w \in [x,y]$. It follows that $z \in X_i$, a contradiction. Uniqueness of segments in Y now follows. □

2. Special Properties of $\mathbb{R}$-trees

There are certain results which are special to $\mathbb{R}$-trees and have no analogue for arbitrary Λ. Here the main result will be the construction of the geometric realisation of a simplicial tree, which is an $\mathbb{R}$-tree, and a generalisation. We shall also briefly recall the construction of the metric completion of an $\mathbb{R}$-metric space, and prove the simple result that the completion of a 0-hyperbolic $\mathbb{R}$-metric space is also 0-hyperbolic. This will prepare the ground for the proof that the completion of an $\mathbb{R}$-tree is an $\mathbb{R}$-tree. Before this, we give some alternative ways of characterising $\mathbb{R}$-trees. We begin with two lemmas. The first puts a constraint on the behaviour of paths in $\mathbb{R}$-trees.

Lemma 2.1. *Let (X,d) be an $\mathbb{R}$-tree, and let $\alpha : [a,b]_{\mathbb{R}} \to X$ be a continuous map. If $x = \alpha(a)$ and $y = \alpha(b)$, then $[x,y] \subseteq \mathrm{Im}(\alpha)$.*

Proof. Let $A = \mathrm{Im}(\alpha)$. Since A is a closed subset of X (being compact), it is enough to show that every point of $[x,y]$ is within distance ε of A, for all real $\varepsilon > 0$.

Given $\varepsilon > 0$, the collection $\{\alpha^{-1}(B(x,\varepsilon/2)) \mid x \in A\}$ is an open covering of the compact metric space $[a,b]_{\mathbb{R}}$, so has a Lebesgue number, say $\delta > 0$ (see Lemma 9.11, Chapter I in [22]). Choose a partition of $[a,b]_{\mathbb{R}}$, say $a = t_0 < \cdots < t_n = b$, so that for $1 \le i \le n$, $t_i - t_{i-1} < \delta$, hence $d(\alpha(t_{i-1}),\alpha(t_i)) < \varepsilon$. Then all points of $[\alpha(t_{i-1}),\alpha(t_i)]$ are at distance $< \varepsilon$ from A, for $1 \le i \le n$. Finally, $[x,y] \subseteq \bigcup_{i=1}^{n}[\alpha(t_{i-1}),\alpha(t_i)]$, by Corollary 1.3. □

Recall from §1 that, if (X,d) is an $\mathbb{R}$-tree and $v \in X$, then the set of directions at v is the set of equivalence classes for the equivalence relation $\equiv$ on the set $\{[v,x] \mid x \in X \setminus \{v\}\}$ defined by: $[v,x] \equiv [v,y]$ if and only if $[v,x] \cap [v,y] \neq \{v\}$.

Lemma 2.2. *Let (X,d) be an $\mathbb{R}$-tree, and suppose $v \in X$. Then*

(1) *For x, $y \in X \setminus \{v\}$, $[v,x] \equiv [v,y]$ if and only if x, y are in the same path component of $X \setminus \{v\}$.*
(2) *$X \setminus \{v\}$ is locally path connected, the components of $X \setminus \{v\}$ coincide with its path components, and they are open sets in X.*
(3) *X is contractible.*

Proof. (1) If $v \in [x,y]$, then $[x,v] \cap [v,y] = \{v\}$, so $[v,x] \equiv [v,y]$ implies $[x,y] \subseteq X \setminus \{v\}$, hence x, y are in the same path component of $X \setminus \{v\}$.

Conversely, if $\alpha : [a,b]_{\mathbb{R}} \to X \setminus \{v\}$ is a continuous map, with $x = \alpha(a)$, $y = \alpha(b)$, then $[x,y] \subseteq \mathrm{Im}(\alpha)$ by Lemma 2.1, so $v \notin [x,y]$, and $[v,x] \cap [v,y] \neq \{v\}$ by Axiom (b)$'$ for a Λ-tree, so $[v,x] \equiv [v,y]$.

(2) For $x \in X \setminus \{v\}$, the set $U = \{y \in X \mid d(x,y) < d(x,v)\}$ is an open set in X, $U \subseteq X \setminus \{v\}$, $x \in U$, and U is path connected. For if y, $z \in U$, then $[x,y] \cup [x,z] \subseteq U$, and so $[y,z] \subseteq U$ by Cor. 1.3. Thus $X \setminus \{v\}$ is locally path connected, and it follows that the path components of $X \setminus \{v\}$ are open, so are also closed, and (2) follows easily.

(3) Let $I = [0,1]_{\mathbb{R}}$ be the unit interval, and define $H : X \times I \to X$ by: $H(x,t)$ is the point on $[v,x]$ at distance $td(v,x)$ from v. We take D as metric on $X \times I$, where $D((x,t),(y,s)) = \max\{d(x,y), |t-s|\}$. By Lemma 1.2(2), $d(H(x,t),H(y,s))$ is either $|td(v,x) - sd(v,y)|$ or $td(v,x) + sd(v,y) - 2d(v,w)$, where $w = \mathrm{Y}(v,x,y)$. Now

$$\begin{aligned}|td(v,x) - sd(v,y)| &= |(t-s)d(v,x) + s(d(v,x) - d(v,y))| \\ &\le |t-s|d(v,x) + d(x,y)\end{aligned}$$

using the triangle inequality and the fact that $0 \leq s \leq 1$, while $td(v,x) + sd(v,y) - 2d(v,w) \leq d(v,x) + d(v,y) - 2d(v,w) = d(x,y)$ by Lemma 1.2(2). Given a real number $\varepsilon > 0$, let $\delta = \varepsilon/(d(v,x)+1)$; then $D((x,t),(y,s)) < \delta$ implies $d(H(x,t), H(y,s)) < \varepsilon$. Hence H is continuous at any point (x,t) in $X \times I$, and it is now clear that H is a contracting homotopy, as required. □

Thus the directions at a point v in a $\mathbb{R}$-tree (X,d) are in one-to-one correspondence with the components of $X \setminus \{v\}$.

An arc in an $\mathbb{R}$-metric space is defined to be a homeomorphic image of a compact interval (i.e. a segment) in $\mathbb{R}$ with more than one element, and its endpoints are the images of the endpoints of the interval under the homeomorphism. Thus any segment with more than one point is an arc.

Proposition 2.3. *Let (X,d) be an $\mathbb{R}$-metric space. The following are equivalent.*

(1) *(X,d) is a $\mathbb{R}$-tree.*
(2) *Given two points of X, there is a unique arc having them as endpoints, and it is a segment.*
(3) *(X,d) is geodesic, and X contains no subspace homeomorphic to the circle S^1.*

Proof. (1)$\Longrightarrow$(2). Let α be a homeomorphism from a compact interval $[a,b]_{\mathbb{R}}$ into X, with image A, and let $x = \alpha(a)$, $y = \alpha(b)$. By Lemma 2.1, $[x,y] \subseteq A$. Hence $\alpha^{-1}([x,y])$ is a connected subset of $[a,b]$ containing a and b, so is equal to $[a,b]$. It follows that $A = [x,y]$.

(2)$\Longrightarrow$(3) If X contains a homeomorphic image of a circle, then the images of two semicircles give two distinct arcs joining the same two points of X.

(3)$\Longrightarrow$(1) We have to verify Axioms (b) and (c) for an $\mathbb{R}$-tree. First, (X,d) is geodesically linear. For otherwise, there are isometries α, $\beta : [0,a] \to X$ with $\alpha(0) = \beta(0) = x$, $\alpha(a) = \beta(a) = y$, say, and $\alpha(b) \neq \beta(b)$ for some $b \in [0,a]$. Let l be the largest element of $[0,b]$ on which α, β agree, and let m be the smallest element of $[b,a]$ on which α, β agree. Then $\alpha([l,m])$, $\beta([l,m])$ are two arcs meeting only at their endpoints, so by the "glueing lemma" we can define a continuous bijection from their union onto S^1; since their union is compact and S^1 is Hausdorff, this map is a homeomorphism, contrary to assumption. It now follows from Lemma 1.2.3 that the strengthened version (c)$'$ of Axiom (c) is satisfied, and it remains to verify Axiom (b). Suppose $[x,y] \cap [x,z] = \{x\}$. If $[x,y] \cup [x,z]$ is not a segment, then (using the remarks preceding Lemma 1.2.2) $x \notin [y,z]$, so by (c)$'$,

$$[y,z] \cap [y,x] = [y,w] \qquad \text{with } w \neq x$$
$$[y,z] \cap [z,x] = [z,v] \qquad \text{with } v \neq x.$$

Note that $w \neq v$, otherwise $w \in [x,y] \cap [x,z]$ which implies $w = x$, a contradiction. Now each pair of the segments $[x,w]$, $[w,v]$ and $[v,x]$ meet only in a single endpoint. For example, since $w \in [x,y]$ and w, $v \in [y,z]$, $[x,w] \subseteq [x,y]$ and $[w,v] \subseteq [y,z]$ (remark following Lemma 1.2.3), hence

$$\begin{aligned}[x,w] \cap [w,v] &= [x,w] \cap [x,y] \cap [w,v] \\ &\subseteq [x,w] \cap [x,y] \cap [y,z] \\ &= [x,w] \cap [y,w] = \{w\}\end{aligned}$$

(again using the remarks preceding Lemma 1.2.2 for the last step). It follows as before, using a version of the glueing lemma, that the union of the three segments is homeomorphic to a Euclidean triangle, so to S^1, a contradiction. □

Condition (3) of the last result may be viewed as an analogue of the definition of a simplicial tree, where circuits are replaced by homeomorphic images of a circle, and the equivalence of (2) and (3) is an analogue of Lemma 1.3.1.

We shall construct the geometric realisation real(Γ) of a simplicial tree Γ, in which we view the unoriented edges as part of the tree, each of them being metrically isomorphic to the unit interval $I = [0,1]_{\mathbb{R}}$, with endpoints identified appropriately, and show that it is an $\mathbb{R}$-tree. This formalises the pictorial representation of graphs by drawing lines and points. To begin with, we shall be more general and assume only that Γ is a graph.

Formally, let Γ be a graph, give $V(\Gamma)$ and $E(\Gamma)$ the discrete topology, let $I = [0,1]_{\mathbb{R}}$ be the unit interval with the usual (Euclidean) topology and put $U(\Gamma) = (E(\Gamma) \times I) \cup V(\Gamma)$ (disjoint union), where $E(\Gamma) \times I$ has the product topology and $U(\Gamma)$ has the usual topology on a disjoint union (the open sets in $U(\Gamma)$ are unions of sets which are open in either $E(\Gamma) \times I$ or $V(\Gamma)$). Define a symmetric binary relation $\simeq$ on $U(\Gamma)$ by declaring that

$$\begin{aligned}(e,t) &\simeq (\bar{e}, 1-t) \\ o(e) &\simeq (e,0) \\ (e,0) &\simeq o(e)\end{aligned}$$

for all $e \in E(\Gamma)$ and $t \in I$. Let $\sim$ be the equivalence relation on $U(\Gamma)$ generated by $\simeq$ (note that $t(e) \sim (e,1)$ for all edges e). Define real(Γ) $= U(\Gamma)/\sim$ with the quotient topology. Denote the equivalence class of $u \in U(\Gamma)$ by $\langle u \rangle$, abbreviating $\langle (e,t) \rangle$ to $\langle e,t \rangle$. Let $p : U(\Gamma) \to \text{real}(\Gamma)$, $u \mapsto \langle u \rangle$ be the quotient map. For $e \in E(\Gamma)$, we denote by real(e) the set $\{\langle e,t \rangle | t \in I\} = p(\{e\} \times I)$. The mapping $\phi_e : I \to \text{real}(e)$, $t \mapsto \langle e,t \rangle$ is continuous, being the map $I \to U(\Gamma)$, $t \mapsto (e,t)$ composed with p. Note that real(e) = real($\bar{e}$), and real(Γ) $= \bigcup_{e \in E(\Gamma)} \text{real}(e)$.

Lemma 2.4. *Under these circumstances:*
(1) $p|_{V(\Gamma)}$ *is a homeomorphism onto its image, which is a discrete, closed subspace of* real(Γ)*;*
(2) *for all* $e \in E(\Gamma)$, $p|_{\{e\}\times(0,1)}$ *is a homeomorphism onto its image, which is open in* real(Γ), *and* $p(\{e\} \times (0,1)) \cap p(V(\Gamma)) = \emptyset$. *Further, the closure in* real(Γ) *of* $p(\{e\} \times (0,1))$ *is* real(e)*;*
(3) *a subset* C *of* real(Γ) *is closed if and only if* $C \cap \text{real}(e)$ *is closed in* real(e), *for all* $e \in E(\Gamma)$*;*
(4) real(Γ) *is Hausdorff;*
(5) *if* $e, f \in E(\Gamma)$ *and* $e \neq f, \bar{f}$, *then* $\text{real}(e) \cap \text{real}(f) \subseteq \{\langle o(e)\rangle, \langle t(e)\rangle\}$.

Proof. If $u, v \in U(\Gamma)$ and $u \sim v$, then this means there is a sequence $u = u_1 \simeq u_2 \simeq \cdots \simeq u_n = v$ for some $n \geq 1$. By induction on n, it is easy to simultaneously prove the following four statements.

(a) If e and e_1 are edges, $0 < t < 1$ and $t_1 \in I$, then $\langle e, t\rangle = \langle e_1, t_1\rangle$ implies either $e = e_1$ and $t = t_1$, or $e = \bar{e}_1$ and $t = 1 - t_1$.
(b) If e and e_1 are edges, $t \in \{0, 1\}$ and $t_1 \in I$, then $\langle e, t\rangle = \langle e_1, t_1\rangle$ implies either $o(e) = o(e_1)$ and $t = t_1$, or $o(e) = t(e_1)$ and $t = 1 - t_1$.
(c) If $\langle e, t\rangle = \langle v\rangle$, where v is a vertex, then either $t = 0$ and $v = o(e)$, or $t = 1$ and $v = t(e)$.
(d) If $\langle v\rangle = \langle w\rangle$, where v, w are vertices, then $v = w$.

Part (5) of the lemma follows immediately from (a), and (d) asserts that p restricted to $V(\Gamma)$ is one-to-one. If Z is a subset of I, we define $Z' = \{1-t \mid t \in Z\}$. Let S be a subset of $V(\Gamma)$. By (c), $p^{-1}(p(S)) = S \cup \{(e, 0) \mid o(e) \in S\} \cup \{(e, 1) \mid t(e) \in S\}$, which is the union of three closed subsets of $U(\Gamma)$, so closed. Hence $p(S)$ is closed in real(Γ), and (1) follows.

Suppose Z is an open subset of $(0, 1)$. Then $p^{-1}(p(\{e\} \times Z)) = (\{e\} \times Z) \cup (\{\bar{e}\} \times Z')$, by (a) and (c). This is open in $U(\Gamma)$, so $p(\{e\} \times Z)$ is open in real(Γ). It follows that $p|_{\{e\}\times(0,1)}$ is an open mapping, and it is one-to-one by (a), and the first assertion in (2) follows. The second part of (2) follows immediately from (c). The closure of $\{e\} \times (0, 1)$ in $E(\Gamma) \times I$ is $\{e\} \times I$, so the closure of $p(\{e\} \times (0, 1))$ contains $p(\{e\} \times I) = \text{real}(e)$. By a similar argument to those already used, real(e) is closed in real(Γ), and (2) follows.

Suppose $C \cap \text{real}(e)$ is closed in real(e), for all $e \in E(\Gamma)$. Then $p^{-1}(C \cap \text{real}(e)) \cap \{e\} \times I$ is a closed subset of $\{e\} \times I$, so has the form $\{e\} \times C_e$, where C_e is a closed subset of I. We have

$$p^{-1}(C \cap \text{real}(e)) = (\{e\} \times C_e) \cup (\{\bar{e}\} \times (C_e)') \cup S_e,$$

where S_e is some subset of $V(\Gamma)$. Thus, putting $A_e = C_e \cup (C_{\bar{e}})'$,

$$\begin{aligned} p^{-1}(C) &= p^{-1}\left(\bigcup_{e\in E(\Gamma)} C\cap \mathrm{real}(e)\right) \\ &= \bigcup_{e\in E(\Gamma)} p^{-1}(C\cap \mathrm{real}(e)) \\ &= \bigcup_{e\in E(\Gamma)} (\{e\}\times A_e)\cup S \end{aligned}$$

where S is a subset of $V(\Gamma)$, and $\bigcup_{e\in E(\Gamma)}(\{e\}\times A_e)$ is closed in $E(\Gamma)\times I$, because its complement $\bigcup_{e\in E(\Gamma)}(\{e\}\times(I\setminus A_e))$ is open, since $E(\Gamma)$ has the discrete topology and A_e is closed in I. Hence $p^{-1}(C)$ is closed, so C is closed, and this proves (3), the converse being trivial.

If $v\in V(\Gamma)$ and $0<\varepsilon\leq 1$, then $p\big(\bigcup_{e\in E(\Gamma),o(e)=v}(\{e\}\times[0,\varepsilon))\ \big)$ is an open neighbourhood of $\langle v\rangle$ in real(Γ), using (3) (note that "closed" can be replaced by "open" in (3)). Similarly, if $0<t<1$ and $e\in E(\Gamma)$, then $p(\{e\}\times(t-\varepsilon,t+\varepsilon))$ is an open neighbourhood of $\langle e,t\rangle$ for sufficiently small ε. Using these neighbourhoods, it is easy to see that real(Γ) is Hausdorff. □

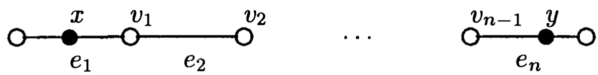
Since I is compact and real(X) is Hausdorff, $\phi_e: I\to \mathrm{real}(e)$ is a closed mapping, so a quotient map. It follows that, if $o(e)\neq t(e)$ then ϕ_e is a homeomorphism, and we can use it to define a metric d_e on real(e), inducing the topology on real(e). Thus $d_e(\langle e,t\rangle,\langle e,s\rangle)=|t-s|$. If $o(e)=t(e)$ then ϕ_e induces a homeomorphism from real(e) to the circle S^1, $\langle e,t\rangle\mapsto\exp(2\pi it)$, and again we have a metric given by $d_e(\langle e,t\rangle,\langle e,s\rangle)=|\exp(2\pi it)-\exp(2\pi is)|$, inducing the topology on real(e). Note that in all cases, $d_e=d_{\bar{e}}$.

It follows from Lemma 2.4 that real(Γ) is a one-dimensional CW-complex, with $p\big(V(\Gamma)$ as 0-skeleton and the ϕ_e as characteristic maps. See [98], Chapters 6 and 7, where "graph" is used to mean a one-dimensional CW-complex. (Strictly, we should choose one map from each pair ϕ_e, $\phi_{\bar{e}}$.) For this reason, we shall call the topology on real(Γ) the CW-topology.

Exercise. Show that any one-dimensional CW-complex is homeomorphic to real(Γ) for some graph Γ, via a homeomorphism preserving cells. (Take $V(\Gamma)$ to be the set of 0-cells, and the unoriented edges of Γ to be the 1-cells of the complex.)

If $f:\Gamma\to\Delta$ is a graph mapping, where Δ is also a graph, there is an obvious induced map $U(f):U(\Gamma)\to U(\Delta)$, which is continuous and clearly preserves $\simeq$, and so $\sim$, hence there is an induced continuous mapping real(f) : real(Γ) → real(Δ). If f is one-to-one, it is easily seen that real(f) is one-to-one. Thus if Γ is a subgraph of Δ, we may view real(Γ) as a subset of real(Δ).

Suppose Γ is connected. We want to define an $\mathbb{R}$-metric d_Γ on real(Γ). The idea is to generalise the definition of the path metric by means of shortest paths.



In the picture, the $v_i \in p(V(\Gamma))$ correspond to the indicated vertices. If x is the point of real(e_1) at distance s from v_1 and y is the point of real(e_1) at distance t from v_{n-1}, then one can travel from x to y by traversing part of real(e_1), then real(e_2), $\ldots$, real(e_{n-1}), then part of real(e_n), covering a total distance $s + (n-2) + t$, assuming $n \geq 2$ (if $n = 1$, the distance is $|s - t|$). We shall define the distance from x to y in such a way that it is the smallest of all such distances. We also view $(x, e_1, v_1, \ldots, v_{n-1}, e_n, y)$ as a generalised path of Γ. To begin with, we allow even more general sequences of edges and points of real(Γ).

Let x, $y \in$ real(Γ) and let $J(x, y)$ denote the set of all sequences

$$\mathbf{v} = (v_0, e_1, v_1, \ldots e_n, v_n),$$

where $e_i \in E(\Gamma)$ for $1 \leq i \leq n$, $x = v_0$, $y = v_n$ and v_{i-1}, $v_i \in$ real(e_i) for $1 \leq i \leq n$. We define $l_{\mathbf{v}} = \sum_{i=1}^{n} d_{e_i}(v_{i-1}, v_i)$. Note that e_i may be replaced by $\bar{e}_i$ without changing $l_{\mathbf{v}}$. Also, $J(x, y) \neq \emptyset$, for there exist e, $f \in E(\Gamma)$ such that $x \in$ real(e), $y \in$ real(f), and there is a path $e_1, \ldots, e_m$ in Γ from an endpoint of e to an endpoint of f with $m \geq 1$. Then $(x, e, o(e_1), e_1, t(e_1), \ldots, t(e_m), f, y) \in J(x, y)$.

Now define $d_\Gamma(x, y) = \inf\{l_{\mathbf{v}} \mid \mathbf{v} \in J(x, y)\}$; we show d_Γ is a metric on real(Γ). If $u = (u_0, e_1, \ldots, e_m, u_m) \in J(x, y)$ and $\mathbf{v} = (v_0, f_1, \ldots, f_n, v_n) \in J(y, z)$, put $w = (u_0, e_1, \ldots u_m, f_1, \ldots, f_n, v_n) \in J(x, z)$; then $l_w = l_u + l_{\mathbf{v}}$, so $d_\Gamma(x, z) \leq l_u + l_{\mathbf{v}}$. It follows that $d_\Gamma(x, z) \leq d_\Gamma(x, y) + d_\Gamma(y, z)$. If $(v_0, e_1, \ldots, e_n, v_n) \in J(x, y)$, then $(v_n, \bar{e}_n, \ldots, \bar{e}_1, v_0) \in J(y, x)$, and it follows easily that $d_\Gamma(x, y) = d_\Gamma(y, x)$. Clearly $d_\Gamma(x, y) \geq 0$, and $d_\Gamma(x, x) = 0$.

Let $J'(x, y)$ be the set of $\mathbf{v} = (v_0, e_1, \ldots, e_n, v_n) \in J(x, y)$ such that

(1) for $1 \leq i \leq n - 1$, real(e_i) $\neq$ real(e_{i+1});
(2) $e_1, \ldots, e_n$ is a reduced path in Γ.

If $\mathbf{v} = (v_0, e_1, \ldots, e_n, v_n) \in J(x, y)$ and real(e_i) $=$ real(e_{i+1}) for some i, then we can omit v_i and e_{i+1} from $\mathbf{v}$ to obtain a new element of $J(x, y)$, without increasing $l_{\mathbf{v}}$ (since d_e satisfies the triangle inequality, for every edge e). Also, if $\mathbf{v} = (v_0, e_1, \ldots, e_n, v_n)$ satisfies (1) then by Lemma 2.4, after replacing each e_i by $\bar{e}_i$ if necessary, $e_1, \ldots, e_n$ is a reduced path and $v_i = \langle o(e_{i+1}) \rangle = \langle t(e_i) \rangle$ for $1 \leq i < n$, and this leaves $l_{\mathbf{v}}$ unaltered. Hence $d_\Gamma(x, y) = \inf\{l_{\mathbf{v}} \mid v \in J'(x, y)\}$.

If $(v_0, e_1, \ldots, e_n, v_n) \in J'(x,y)$, then $d_{e_i}(v_{i-1}, v_i)$ is either 0 or 1, for $2 \leq i \leq n-1$. Also, $v_1 = \langle t(e_1)\rangle$ and $v_{n-1} = \langle t(e_{n-1}\rangle$. Further, e_1 is determined by x and e_n by y. It follows that $d_\Gamma(x,y) = l_{\mathbf{v}}$ for some $\mathbf{v} \in J'(x,y)$. Hence if $x \neq y$, $d_\Gamma(x,y) > 0$, so d_Γ is an $\mathbb{R}$-metric on real(Γ).

Note that, if $(v_0, e_1, \ldots, e_n, v_n) \in J'(x,y)$, none of $v_1, \ldots, v_{n-1}$ is a terminal vertex of Γ (strictly, no $v_i = \langle v\rangle$, where v is a terminal vertex, for $1 \leq i \leq n-1$). Also, it is easy to see that if $x = \langle u\rangle$, $y = \langle v\rangle$, where u, $v \in V(\Gamma)$, then $d_\Gamma(x,y) = d(u,v)$, where d is the path metric of Γ, and if x, $y \in$ real(e), where $e \in E(\Gamma)$, then $d_\Gamma(x,y) = d_e(x,y)$. When it is necessary to specify the graph Γ, we shall write $J_\Gamma(x,y)$ and $J'_\Gamma(x,y)$ for $J(x,y)$ and $J'(x,y)$.

Lemma. *Suppose Γ is a graph with no loops and Δ is a connected subgraph such that, if $e \in E(\Gamma) \setminus E(\Delta)$, then one endpoint of e is in $V(\Delta)$ and the other endpoint is a terminal vertex of Γ (so Γ is connected). If* (real$(\Delta), d_\Delta$) *is an $\mathbb{R}$-tree, so is* (real$(\Gamma), d_\Gamma$)*, and d_Δ is d_Γ restricted to* real$(\Delta) \times$ real(Δ).

Proof. If x, $y \in$ real(Δ), and $\mathbf{v} = (v_0, e_1 \ldots, e_n, v_n) \in J'_\Gamma(x,y)$ is such that $d_\Gamma(x,y) = l_{\mathbf{v}}$, then $\mathbf{v} \in J'_\Delta(x,y)$, because v_i is not a terminal vertex of Γ for $1 \leq i \leq n-1$. Hence $d_\Gamma(x,y) \geq d_\Delta(x,y)$, and $d_\Delta(x,y) \geq d_\Gamma(x,y)$ because $J_\Delta(x,y) \subseteq J_\Gamma(x,y)$. Thus $d_\Gamma(x,y) = d_\Delta(x,y)$.

Suppose $y \in$ real(Δ) and $x \in$ real(e), where $e \in E(\Gamma) \setminus E(\Delta)$ with $t(e)$ a terminal vertex. We claim that $d_\Gamma(x,y) = d_e(x, \langle o(e)\rangle) + d_\Gamma(\langle o(e)\rangle, y)$. This is clear if $x = \langle o(e)\rangle$; if $x \neq \langle o(e)\rangle$, take $\mathbf{v} = (v_0, e_1, \ldots, e_n, v_n) \in J'_\Gamma(x,y)$ with $d_\Gamma(x,y) = l_v$. Then $e_1 = \bar{e}$ and $v_1 = \langle o(e)\rangle$, for otherwise $v_1 = \langle t(e)\rangle$, and $t(e)$ is a terminal vertex, a contradiction. Hence, putting $\mathbf{v}' = (v_1, e_2, \ldots, e_n, v_n)$, we have $\mathbf{v}' \in J_\Gamma(\langle o(e)\rangle, y)$, and

$$\begin{aligned} d_\Gamma(x,y) &= d_e(x, \langle o(e)\rangle) + l_{\mathbf{v}'} = d_\Gamma(x, \langle o(e)\rangle) + l_{\mathbf{v}'} \\ &\geq d_\Gamma(x, \langle o(e)\rangle) + d_\Gamma(\langle o(e)\rangle, y) \end{aligned}$$

and by the triangle inequality $d_\Gamma(x,y) = d_e(x, \langle o(e)\rangle) + d_\Gamma(\langle o(e)\rangle, y)$. Similarly, if $x \in$ real(e), where $e \in E(\Gamma) \setminus E(\Delta)$ with $t(e)$ a terminal vertex, and $y \in$ real(e'), where $e' \in E(\Gamma) \setminus E(\Delta)$ with $t(e')$ a terminal vertex, and $e \neq e'$, then

$$d_\Gamma(x,y) = d_e(x, \langle o(e)\rangle) + d_\Gamma(\langle o(e)\rangle, o(e')) + d_{e'}(\langle o(e')\rangle, y).$$

The result now follows from Lemma 1.13, applied with $X =$ real(Δ) and the X_i the spaces real(e) where $e \in E(\Gamma) \setminus E(\Delta)$ and $t(e)$ is a terminal vertex. □

Proposition 2.5. *If Γ is a simplicial tree, then* (real$(\Gamma), d_\Gamma$) *is an $\mathbb{R}$-tree.*

Proof. If Γ is empty, so is real(Γ). If Γ has finite diameter, it follows by induction on the diameter, using the remark before Lemma 1.3.3 and Lemma

2.4, that $(\text{real}(\Gamma), d_\Gamma)$ is an $\mathbb{R}$-tree. In general, let d be the path metric of Γ, choose a vertex $v \in V(\Gamma)$ and let Γ_n be the subgraph such that $V(\Gamma_n) = \{u \in V(\Gamma) | d(v,u) \leq n\}$, with $E(\Gamma_n)$ the set of edges whose endpoints belong to this set. Then clearly Γ_n is a subtree of Γ of finite diameter (any vertex can be joined to v by a path of length at most n), hence $(\text{real}(\Gamma_n), d_{\Gamma_n})$ is an $\mathbb{R}$-tree for all n. Finally, $\text{real}(\Gamma) = \bigcup_{n \geq 0} \text{real}(\Gamma_n)$, and it follows from Lemma 1.14 that $(\text{real}(\Gamma), d_\Gamma)$ is an $\mathbb{R}$-tree. □

If Γ is a connected graph, the metric d_Γ induces a topology on real(Γ). It would be nice to report that this coincides with the CW-topology, but the situation is described in the next lemma. A graph is said to be locally finite if the degree of every vertex is finite.

Lemma 2.6. *Let Γ be a connected graph. The metric topology on* real(Γ) *is coarser than the CW-topology. The two agree if and only if Γ is locally finite.*

Proof. As noted after Lemma 2.4, d_e, and so d_Γ, induce the relative CW-topology on real(e), for all $e \in E(\Gamma)$, so we can speak of closed subsets of real(e) without ambiguity. Let $C \subseteq \text{real}(\Gamma)$. If C is closed in the metric topology then $C \cap \text{real}(e)$ is closed in real(e), for all $e \in E(\Gamma)$, hence C is closed in the CW-topology.

Suppose Γ is locally finite, and C is closed in the CW-topology. Let (x_n) be a sequence in C and suppose (x_n) converges to x in the metric topology, where $x \in \text{real}(\Gamma)$. If $x \in p(\{e\} \times (0,1))$, then it is easy to see from the definition of the metric that $x_n \in C \cap \text{real}(e)$ for sufficiently large n, and $C \cap \text{real}(e)$ is closed in real(e). Hence $x \in C \cap \text{real}(e) \subseteq C$. If $x = p(v)$ for some $v \in V(\Gamma)$, then for sufficiently large n, there exist $e_n \in E(\Gamma)$ with $o(e_n) = v$ such that $x_n \in \text{real}(e_n)$. By local finiteness there exists $e \in E(\Gamma)$ such that $e_n = e$ for infinitely many n, so there is a subsequence (x_{n_i}) with $x_{n_i} \in C \cap \text{real}(e)$. As before, it follows that $x \in C \cap \text{real}(e)$. These are the only possibilities for x, hence C is closed in the metric topology.

Suppose Γ is not locally finite, so there exist $v \in V(\Gamma)$ and $e_n \in E(\Gamma)$ for $n \geq 1$ such that $e_i \neq e_j$, $\bar{e}_j$ for $i \neq j$ and $o(e_n) = v$ for all n. Let $C_n = p(e_n \times [1/2^n, 1/2])$. Then C_n is compact, so closed in real(e_n). It follows from Lemma 2.4 that the set D defined by $D = \bigcup_{n=1}^\infty C_n$ is closed in the CW-topology. Let $x_n = p(e_n, 1/2^n)$. Then $x_n \in D$, $x_n \to p(v)$ in the metric topology, but $p(v) \notin D$ by Lemma 2.4. □

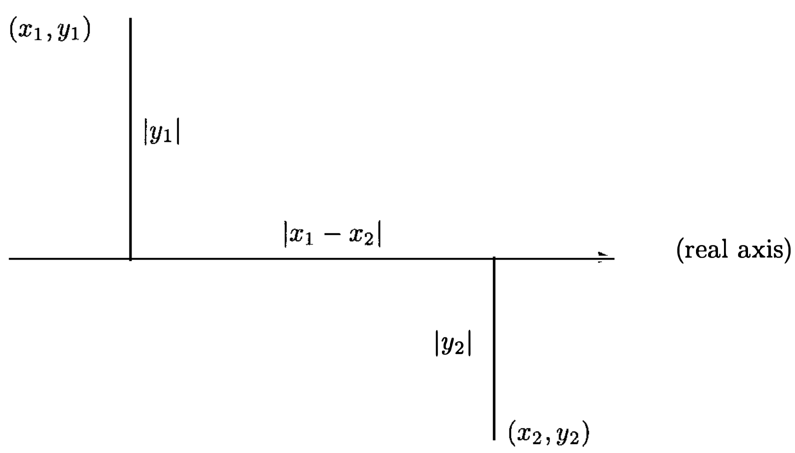
We shall give a well-known example to show that not all $\mathbb{R}$-trees are metrically isomorphic to real(Γ) for some simplicial tree Γ. First, some remarks, indicating that segments in real(Γ) behave as expected, are needed.

Suppose that Γ is a simplicial tree and x, $y \in \text{real}(\Gamma)$. Then $d_\Gamma(x,y) = l_{\mathbf{v}}$ for some $\mathbf{v} = (v_0, e_1, v_1, \ldots, e_n, v_n) \in J'_\Gamma(x,y)$, and $l_{\mathbf{v}} = \sum_{i=1}^n d_{e_i}(v_{i-1}, v_i)$, so $d_\Gamma(x,y) = \sum_{i=1}^n d_\Gamma(v_{i-1}, v_i)$. By Lemma 1.4, $[x,y] = [v_0, v_1, \ldots, v_n]$. As already noted, $v_i = o(e_{i+1}) = t(e_i)$ for $1 \leq i < n$. Also, for any edge e, real(e) is a segment with endpoints $\langle o(e)\rangle$, $\langle t(e)\rangle$, so $[\langle o(e)\rangle, \langle t(e)\rangle] = \text{real}(e)$. (We note in passing that this implies that $p(V(\Gamma))$ spans real(Γ), since every point of real(Γ) is in real(e) for some edge e.) This gives a description of the segment $[x, y]$; $[v_{i-1}, v_i] = \text{real}(e_i)$ for $1 < i < n$, $[v_0, v_1] \subseteq \text{real}(e_1)$ and $[v_{n-1}, v_n] \subseteq \text{real}(e_n)$. In particular, by Lemma 2.4(2), $[x, y] \cap p(V(\Gamma))$ is contained in $\{v_0, \ldots, v_n\}$, so is finite. Further, if $x = \langle e, t\rangle$, where $0 < t < 1$, then $e_1 = e$ by Lemma 2.4 (Parts (2) and (5)), and it follows that there are exactly two directions at x in the $\mathbb{R}$-tree $(\text{real}(\Gamma), d_\Gamma)$, namely the equivalence classes of the segments $[x, \langle o(e)\rangle]$ and $[x, \langle t(e)\rangle]$. Thus the set of points in real(Γ) of degree not equal to 2 is contained in $p(V(\Gamma))$, and this set is closed and discrete in the metric topology, since the restriction of d_Γ is the path metric, which is integer-valued. It also follows using Lemma 2.4(5) that, if $v \in V(\Gamma)$, then the degree of $\langle v\rangle$ is the degree of v in the graph Γ, as defined after Lemma 1.3.1.

Example. Let $Y = \mathbb{R}^2$ be the plane, but with metric d defined by

$$d((x_1, y_1), (x_2, y_2)) = \begin{cases} |y_1| + |y_2| + |x_1 - x_2| & \text{if } x_1 \neq x_2 \\ |y_1 - y_2| & \text{if } x_1 = x_2. \end{cases}$$

Thus, to measure the distance between two points not on the same vertical line, we take their projections onto the horizontal axis, and add their distances to these projections and the distance between the projections (distance in the usual Euclidean sense). If they are on the same vertical line, their distance is the usual Euclidean distance. The metric is illustrated by the following picture.



Now (Y, d) is an $\mathbb{R}$-tree by Lemma 1.13, with X equal to the real axis $\mathbb{R} \times \{0\}$, $J = \mathbb{R}$ and X_j the vertical line $\{j\} \times \mathbb{R}$, for $j \in J$, so that X and all the X_j are isomorphic as $\mathbb{R}$-metric spaces to $\mathbb{R}$. In particular, X is not discrete, and it is easy to see that every point of X has degree 4. Thus Y is not isomorphic to real(Γ) for any simplicial tree Γ.

We can iterate this construction using Lemma 1.13, to make $\mathbb{R}^n$ into an $\mathbb{R}$-tree, and then $\bigcup_{n=1}^{\infty} \mathbb{R}^n$ becomes an $\mathbb{R}$-tree by Lemma 1.14.

Most of the rest of this section is about the idea of a polyhedral $\mathbb{R}$-tree. This will be used only once, in Theorem 5.2.6, and readers may wish to omit the rather technical arguments. However, it ends with a discussion of metric completion, which will be used more extensively. Also, Exercise 1 after Theorem 2.10 will be used in Chapter 6.

If Γ is a simplicial tree, the construction of the $\mathbb{R}$-tree (real(Γ), d_Γ) can be generalised as follows. Let $r : E(\Gamma) \to \mathbb{R}$, $e \mapsto r_e$, be a mapping such that $r_e > 0$ for all edges e, and $r_e = r_{\bar{e}}$ for all $e \in E(\Gamma)$. We call r a real weight function on Γ. In the construction of real(Γ), replace $\{e\} \times [0,1]$ by $\{e\} \times [0, r_e]$, so $U(\Gamma)$ is replaced by $(\bigcup_{e \in E(\Gamma)} \{e\} \times [0, r_e]) \cup V(\Gamma)$, which we denote by U_r. The relation $\simeq$ is modified by replacing $(e,t) \simeq (\bar{e}, 1-t)$ by $(e,t) \simeq (\bar{e}, r_e - t)$, and d_e is defined exactly as before. The construction then works in a similar way to produce an $\mathbb{R}$-tree (X_r, d_r), where now the edge e gives rise to a segment of length r_e, instead of the segment real(e) of length 1. We denote the projection map from U_r to X_r by p_r, so real(e) is replaced by $p_r(e)$ in the definition of $J(x,y)$. There is one important difference; the reason that $d_r(x,y) = l_{\mathbf{v}}$ for some $\mathbf{v} \in J'(x,y)$ is the uniqueness of reduced paths (1.3.1). (The construction no longer works in general if Γ is not a simplicial tree.) Note that (1) and (5) of Lemma 2.4 still hold, as does the second part of (2): $p_r(e \times (0,1)) \cap p_r(V(\Gamma)) = \emptyset$.

The description of segments in real(Γ) preceding the example above still applies (replacing real(e) with $p_r(e)$). In particular, any segment in X_r contains only finitely many points of $p_r(V(\Gamma))$, and $p_r(V(\Gamma))$ spans X_r. This leads to a characterisation of $\mathbb{R}$-trees which arise from this construction.

Lemma 2.7. *Let (X,d) be an $\mathbb{R}$-tree. The following are equivalent.*

(1) *(X,d) is metrically isomorphic to (X_r, d_r) for some simplicial tree Γ and weight function r.*

(2) *there is a subset S of X such that, for all $x, y \in X$, $S \cap [x,y]$ is finite, and S spans X.*

Proof. Given a weight function r on a graph Γ, let $S = p_r(V(\Gamma))$. We have already observed that S has the properties in (2), and these properties are preserved under metric isomorphism, so (1) implies (2).

Assume (2). We define a graph Γ as follows. Let $V(\Gamma) = S$, and let $E(\Gamma)$ be the set of ordered pairs (x,y) of distinct points of X such that x, y are the only points of $[x,y]$ which belong to S. Put $o(x,y) = x$, $t(x,y) = y$, $\overline{(x,y)} =$

(y, x) and define a weight function r by $r_{(x,y)} = d(x, y)$. The assumption that S intersects every segment in finitely many points implies that Γ is connected.

If $e_1, \ldots, e_n$ $(n \geq 1)$ is a reduced path in Γ, we can write $e_i = (v_{i-1}, v_i)$, for $1 \leq i \leq n$, and v_{i-1}, v_i, v_{i+1} are three distinct points of S for $1 \leq i < n$. Let $w_i = \mathrm{Y}(v_{i-1}, v_i, v_{i+1})$. If $w_i = v_{i-1}$ then $v_{i-1} \in [v_i, v_{i+1}] \cap S = \{v_i, v_{i+1}\}$, a contradiction. If $w = v_{i+1}$, we obtain a similar contradiction since then $v_{i+1} \in [v_{i-1}, v_i] \cap S$. It follows that $w_i = v_i$, for otherwise $w_i \in S$, because $[w_i, v_j]$ for $(i-1) \leq j \leq (i+1)$ define three different directions at w_i, by Lemma 1.2. But then $w_i \in S \cap [v_{i-1}, v_i]$, which is again a contradiction. Hence $[v_{i-1}, v_i, v_{i+1}]$ is a piecewise segment (remark preceding Lemma 1.4), so $[v_0, v_1, \ldots v_n]$ is a piecewise segment by Lemma 1.5. It follows that $v_0 \neq v_n$, hence there are no circuits in Γ.

There is a mapping from U_r to X, which is the identity on $S = V(\Gamma)$ and, if $e = (x, y) \in E(\Gamma)$ and $0 \leq t \leq r_e$, sends (e, t) to the point on $[x, y]$ at distance t from x. Further, there is an induced a mapping $f : X_r \to X$, such that $f(p_r(x)) = x$ for $x \in V(\Gamma))$, and which maps $p_r(e)$ isometrically onto $[x, y]$ for every edge $e = (x, y)$. If $\mathbf{v} = (v_0, e_1, \ldots, e_n, v_n) \in J'(x, y)$ is such that $d_r(x, y) = l_{\mathbf{v}}$, then $e_1, \ldots, e_n$ is a reduced path in Γ and as with real(Γ), $v_i = p_r(o(e_{i+1})) = p_r(t(e_i))$ for $1 \leq i < n$. From the above, $[o(e_1), t(e_1), \ldots, t(e_{n-1}), t(e_n)]$ is a piecewise segment, hence so is $[o(e_1), f(x), t(e_1), \ldots, t(e_{n-1}), f(y), t(e_n)]$ $(f(x) \in f(P_r(e_1)) = [o(e_1), t(e_1)]$, etc). Since $f(v_i) = t(e_i)$ for $1 \leq i < n$, $[f(x), f(y)] = [f(v_0), f(v_1), \ldots, f(v_n)]$, and since f restricted to $p_r(e_i)$ is an isometry, it follows from Lemma 1.4 that f is an isometry. Since S spans X, f is an isomorphism. □

The set of all branch points and endpoints of real(Γ), where Γ is a simplicial tree, is a closed, discrete subset of real(Γ). For it is a subset of $p(V(\Gamma))$, which is itself closed and discrete (in the metric topology), as already noted. In the case of an arbitrary weight function r, the set of branch points and endpoints is still closed (it is easy to see that the complement is open), but need no longer be discrete. For example, take a tree with countably many edges e_n, where n is a positive integer, and a vertex v such that $o(e_n) = v$, $v \neq t(e_n)$, for all n, and $t(e_i) \neq t(e_j)$ for all i, j, with $i \neq j$. We can arrange that each $t(e_n)$ is either a branch point or an endpoint. Let $r_{e_n} = 1/n$. Then $t(e_n)$ converges to v.

Definition. An $\mathbb{R}$-tree is called polyhedral if the set of all branch points and endpoints is closed and discrete.

We proceed to characterise polyhedral $\mathbb{R}$-trees. The argument for the next lemma anticipates the discussion of rays and ends in the next section; the sets L_x in the proof are examples of rays in an $\mathbb{R}$-tree.

Lemma 2.8. *Let (X, d) be an $\mathbb{R}$-tree and let C be a closed non-empty discrete* *subset of X containing all branch points and endpoints of X. Then there is a closed discrete subset S of X such that $C \subseteq S$ and S spans X.*

Proof. Let Y be the subtree spanned by C. Suppose $x \in X \setminus Y$ and let y be the nearest point of C to x. There cannot be more than one such point, for if y' is another, let $w = \mathrm{Y}(y, x, y')$. Then $w \neq y$, y' since y, y' are equidistant from x (for example, if $w = y$, then $[y', x] = [y', y, x]$, so $d(y, x) < d(y', x)$). Also, $w \neq x$ since otherwise $x \in [y, y']$, contrary to $x \notin Y$. By Lemma 1.2, w is a branch point of X (see the argument for 2.7), and it is closer to x than y and y', a contradiction. (Much of the rest of the proof is repetition of this idea.) Put

$$L_x = \{z \in X \mid z \in [y, x] \text{ or } x \in [y, z]\}.$$

Note that, if $z \in L_x$, then $[y, z] \subseteq L_x$, and if u, $v \in L_x$ then either $u \in [y, v]$ or $v \in [y, u]$. There are various cases to consider, mostly straightforward, except when $[y, u] = [y, x, u]$ and $[y, v] = [y, x, v]$. In this case let $w = \mathrm{Y}(u, y, v)$; since $x \in [y, u] \cap [y, v] = [y, w]$, $w \neq y$, $w \in L_x$ and $w \notin C$ (otherwise $x \in Y$). Hence w has degree 2, so by Lemma 1.2, either $u = w$ or $v = w$, and $w \in [y, v] \cap [y, u]$.

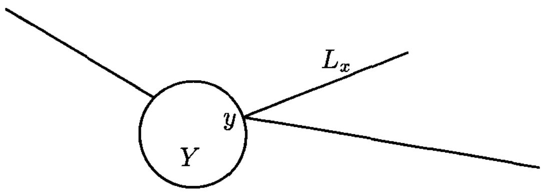
It follows by Lemma 1.4 that there is an isometry $\alpha_x : L_x \to \mathbb{R}$ given by $\alpha_x(z) = d(y, z)$, whose image is an interval $[0, a)$, where $a \in \mathbb{R} \cup \{\infty\}$ and $a > 0$. (If it were an interval of the form $[0, a]$ with $a \in \mathbb{R}$, and $z = \alpha_x^{-1}(a)$, then z would be an endpoint of X, and $x \in [y, z]$, contradicting $x \notin Y$.) Thus L_x is a subtree of X having y as an endpoint.

Further, y is the only element of $L_x \cap Y$. For suppose $z \in L_x \cap Y$. Then $x \notin [y, z]$ since $x \notin Y$, so $z \in [y, x]$. Also, there exist u, $v \in C$ such that $z \in [u, v]$. By Lemma 1.7, $[x, y] \cap [u, v] = [p, q]$, where $p = \mathrm{Y}(x, y, u)$ and $q = \mathrm{Y}(x, y, v)$, and we can assume u, v chosen so that $[y, x] = [y, p, z, q, x]$. If $z \neq y$, then $q \neq y$ and $q \neq x$ since $x \notin Y$, while $q \in Y$ since $q \in [u, v]$. By Lemma 1.2, either $q = v$, or q is a branch point of X, so in any case $q \in C$. This is a contradiction since q is closer to x than y.

Now if $x' \in X \setminus Y$ and y' is the point of C closest to x', and $L_x \cap L_{x'} \neq \emptyset$, then $y = y'$. For let $z \in L_x \cap L_{x'}$ and put $w = \mathrm{Y}(y, z, y')$. Then $w \in [y, z] \cap [y', z] \subseteq L_x \cap L_{x'}$ and $w \in [y, y'] \subseteq Y$, so $w = y = y'$. We claim that, if $u \in L_x$, $u \neq y$ then $L_u = L_x$. For $u \in L_x \cap L_u$, so y is the point of C closest to u. It follows that, if $v \in L_u$, then $v \in L_x$. The only case which needs comment is when $[y, v] = [y, u, v]$ and $[y, x] = [y, u, x]$. Put $w = \mathrm{Y}(y, x, v)$. Then $u \in [y, x] \cap [y, v] = [y, w]$, so $w \neq y$. Therefore w is closer to x than y (since $w \in [x, y]$), so is not a branch point. By Lemma 1.2, either $w = x$ or $w = v$, and in either case $v \in L_x$. Thus $L_u \subseteq L_x$; but $u \in L_x$ implies $x \in L_u$, and so the claim follows.

If $L_x \cap L_{x'} \neq \emptyset$ and $L_x \cap L_{x'} \neq \{y\}$, it follows that $L_x = L_{x'}$. For if $u \in L_x \cap L_{x'}$ and $u \neq y$ then $L_u = L_x = L_{x'}$. Thus two distinct sets L_x, $L_{x'}$

are either disjoint or meet in a single point, which is an endpoint of both. The following diagram illustrates the situation.



Finally, if the isometry α_x has image $[0, a)$, add the points $\alpha_x^{-1}(a(1-1/n))$ $(n \geq 2)$ to C if $a < \infty$, and add the points $\alpha_x^{-1}(n)$ to C if $a = \infty$, for all positive integers n. Doing this for every set L_x gives the required set S. □

Corollary 2.9. *Suppose (X, d) is a polyhedral $\mathbb{R}$-tree. Then (X, d) is metrically isomorphic to (X_r, d_r) for some simplicial tree Γ and bounded weight function r, such that the image of $V(\Gamma)$ in X_r is closed and discrete.*

Proof. If X is empty, take Γ to be empty. Otherwise, take C as in Lemma 2.8 (take C to be the set of branch and endpoints if this set is non-empty, otherwise let C contain a single point of X), and obtain S as in Lemma 2.8. If x, $y \in S$, $x \neq y$ and the segment $[x,y]$ contains no points of S other than x and y, and $d(x,y) > 1$, then we can add finitely many points of $[x,y]$ to S, say $[x, y] = [x_0, \ldots, x_n]$, such that $1/2 \leq d(x_{i-1}, x_i) \leq 1$ for $1 \leq i \leq n$. Do this for each such segment; the resulting set S' is still closed and discrete, hence satisfies the hypotheses of Lemma 2.7 (its intersection with a segment is compact). The construction of that lemma gives the required $\mathbb{R}$-tree (X_r, d_r) metrically isomorphic to X, with r bounded above by 1, and with $V(\Gamma) = S'$. □

Theorem 2.10. *An $\mathbb{R}$-tree (X, d) is polyhedral if and only if it is homeomorphic to real(Γ) (with the metric topology) for some simplicial tree Γ.*

Proof. Suppose (X, d) is polyhedral; by 2.9, we can assume $(X, d) = (X_r, d_r)$ for some bounded weight function r on a simplicial tree Γ, with $p_r(V(\Gamma))$ closed and discrete. There is a mapping $U(\Gamma) \to U_r$ which is the identity mapping on $V(\Gamma)$ and sends (e, t) to $(e, r_e t)$ for $e \in E(\Gamma)$ and $t \in [0, 1]$. This induces a mapping $f : \text{real}(\Gamma) \to X_r$, which is a bijection, and whose restriction to real(e) is a homeomorphism, for $e \in E(\Gamma)$. If $v \in V(\Gamma)$, there exists a real number $R > 0$ such that for all other $w \in V(\Gamma)$, $d_r(p_r(v), p_r(w)) > R$. It follows that $r_e \geq R$ for all edges e of Γ having v as an endpoint. It then follows easily that f is a homeomorphism, using the criterion that, if (X, d) and (Y, d') are metric spaces, and $f : X \to Y$ is a mapping, then f is continuous at $x \in X$ if and only if, for all sequences (x_n) in X which converge to x, $(f(x_n))$ converges to $f(x)$. Details are left to the reader.

Conversely, if $f : \text{real}(\Gamma) \to X$ is a homeomorphism, and x, $y \in \text{real}(\Gamma)$, $x \neq y$, then $f([x, y])$ is an arc in X with endpoints $f(x)$, $f(y)$, so by Prop. 2.3, $f([x, y]) = [f(x), f(y)]$. It follows that f preserves degree of points, and so the set of branch and endpoints. Hence (X, d) is polyhedral. □

Exercises. (1) If an $\mathbb{R}$-tree is spanned by a finite set, show that it has finitely many branch and endpoints, so is polyhedral. Also, it is compact and homeomorphic to real(Γ) for some finite simplicial tree Γ.

(2) If B is the set of branch points and E is the set of endpoints of an $\mathbb{R}$-tree, and B is closed, show that $B \cup E$ is closed.

(3) Give an example of an $\mathbb{R}$-tree where the set of branch points is not closed.

We close by briefly describing the construction of the completion of an $\mathbb{R}$-metric space. This is well-known and details are left to the reader. Let (X, d) be an $\mathbb{R}$-metric space. A *Cauchy sequence* in X is a sequence (x_n) of elements of X such that, for all real $\varepsilon > 0$, there exists an integer N such that m, $n \geq N \Longrightarrow d(x_m, x_n) < \varepsilon$. The metric space (X, d) is called complete if every Cauchy sequence converges to some point of X. It is easy to see that, if (X, d) is an $\mathbb{R}$-metric space and Y is a dense subspace, such that every Cauchy sequence in Y converges in X, then (X, d) is complete.

Let (X, d) be an $\mathbb{R}$-metric space, and let C denote the set of Cauchy sequences in X; for example, if $x \in X$, the constant sequence (x_n) with $x_n = x$ for all n is in C; we denote this constant sequence by $\hat{x}$. Define a binary relation on C by $(x_n) \sim (y_n)$ if and only if $\lim_{n\to\infty} d(x_n, y_n) = 0$. This is an equivalence relation on C, and we put $\hat{X} = C/\sim$, the set of equivalence classes. The equivalence class of (x_n) is denoted by $\langle x_n \rangle$. For (x_n), $(y_n) \in C$, $(d(x_n, y_n))$ is a Cauchy sequence in $\mathbb{R}$, which is complete, hence $\lim_{n\to\infty} d(x_n, y_n)$ exists (and is in $\mathbb{R}$). Further, the limit is unchanged if (x_n) or (y_n) is replaced by a sequence in the same equivalence class. Therefore we can define

$$\hat{d}(\langle x_n \rangle, \langle y_n \rangle) = \lim_{n\to\infty} d(x_n, y_n)$$

and it is easy to see that $(\hat{X}, \hat{d})$ is an $\mathbb{R}$-metric space. The mapping $\phi : X \to \hat{X}$, $x \mapsto \hat{x}$, is an isometry. If $\langle x_n \rangle \in \hat{X}$, then $\widehat{x}_n$ converges to $\langle x_n \rangle$. It follows that $\phi(X)$ is dense in $\hat{X}$, and every Cauchy sequence in $\phi(X)$ converges in $\hat{X}$. Hence $(\hat{X}, \hat{d})$ is complete, and if we identify x with $\phi(x)$, then X is dense in $\hat{X}$. The $\mathbb{R}$-metric space $(\hat{X}, \hat{d})$ is determined up to metric isomorphism by these two properties, and is called the completion of (X, d). We shall show in §4 that, if (X, d) is an $\mathbb{R}$-tree, then so is $(\hat{X}, \hat{d})$. For now, we shall state a simple lemma which will be used in the proof, based on the following remark.

Remark. If (a_n), (b_n), (c_n) are real sequences, converging to a, b, c respectively, and (a_n, b_n, c_n) is an isosceles triple for all n, then (a, b, c) is an isosceles triple.

For if (a, b, c) is not isosceles, then without loss of generality we can assume $a < b$ and $a < c$. Then for sufficiently large n, $a_n < b_n$ and $a_n < c_n$, a contradiction.

Lemma 2.11. *If (X, d) is a 0-hyperbolic $\mathbb{R}$-metric space, then its completion $(\hat{X}, \hat{d})$ is also 0-hyperbolic.*

Proof. Let $x = \langle x_n \rangle$, $y = \langle y_n \rangle$, $z = \langle z_n \rangle$, $v = \langle v_n \rangle$ be points of $\hat{X}$. We have to show that if $a = (x \cdot y)_v$, $b = (y \cdot z)_v$, $c = (x \cdot z)_v$, then (a, b, c) is an isosceles triple. This follows at once from the remark, with $a_n = (x_n \cdot y_n)_{v_n}$, etc. □

If $\alpha : X \to Y$ is an isometry of $\mathbb{R}$-metric spaces, then there is a unique extension of α to an isometry $\hat{\alpha} : \hat{X} \to \hat{Y}$, given by $\hat{\alpha}(\langle x_n \rangle) = \langle \alpha(x_n) \rangle$. We leave it as an exercise to check this. It follows that, if G is a group acting as isometries on X, there is a unique extension to an action by isometries on $\hat{X}$.

3. Linear Subtrees and Ends

The notion of end in a graph is a combinatorial analogue of the idea of an end of a locally compact space introduced by Hopf and Freudenthal, and was used to good effect by Stallings in 1968, who essentially proved that finitely generated groups of cohomological dimension one are free groups. (His argument applied to finitely presented groups, but this was quickly improved to finitely generated groups by several authors.) For more information on ends of graphs, see [38] or [136]. (Note, however, that these do not contain the most up to date version of Stallings' result and its generalisation by Swan to non-finitely generated groups. For this, see [45].)

In the case of a simplicial tree Γ, the ends may be described as follows. A Γ-ray is an infinite sequence $e_1, e_2, \cdots$ of edges such that $t(e_i) = o(e_{i+1})$ and $e_i \neq \bar{e}_{i+1}$ for $i \geq 1$ (a kind of infinite reduced path). Two such rays, $e_1, e_2, \cdots$ and $f_1, f_2, \cdots$ are equivalent if they agree from some point on (after suitable renumbering), that is, there exist integers m and N such that, for all $i \geq N$, $e_i = f_{i+m}$. This is an equivalence relation, and the ends of Γ are the equivalence classes. The idea is that the ends should represent different ways of "going to infinity" in Γ.

In combinatorics, where one considers finite simplicial trees, the terminal vertices (those of degree 1) frequently play a role, and it is convenient to think of these as "closed ends" of the tree, calling the ends just defined open ends. We shall give a definition of ends in an arbitrary Λ-tree which allows for both possibilities. To do so, we introduce the idea of linear subtree with an endpoint.

An example is the set $\{o(e_i) \mid i \geq 1\}$ in the $\mathbb{Z}$-tree corresponding to a simplicial tree Γ, where $e_1, e_2, \cdots$ is a Γ-ray as above.

Ends play a significant role in the theory of Λ-trees, but this will not be apparent until later, in Chapter 3.

Definition. A Λ-tree (X, d) is linear if it is metrically isomorphic to a subtree of Λ. A subset A of a Λ-tree (X, d) is linear if it is contained in a linear subtree of X.

We shall also say "$x_1, \ldots, x_n$ are collinear" to mean that $\{x_1, \ldots, x_n\}$ is linear. Our first observation on linear subtrees is an elaboration of Lemma 1.2.1.

Lemma 3.1. *If (X, d) is a linear Λ-tree, $A \subseteq X$, A has at least two elements, and $\alpha : A \to \Lambda$ is an isometry, then there is a unique extension of α to an isometry $X \to \Lambda$.*

Proof. There is an isometry $\gamma : X \to \Lambda$. Then by Lemma 1.2.1, $\gamma \circ \alpha^{-1} : \alpha(A) \to \Lambda$ is the restriction to $\alpha(A)$ of a metric automorphism of Λ, say q, and $q^{-1} \circ \gamma$ is an extension of α. If β, β' are both extensions of α to X, then $\beta' \circ \beta^{-1} : \beta(X) \to \Lambda$ is the identity on $\alpha(A)$, which has at least two elements, so $\beta = \beta'$ by Lemma 1.2.1. □

Suppose (X, d) is a Λ-tree, $A \subseteq X$ and every three-element subset of A is linear. If x, y, $z \in A$, there is an isometry $\alpha : \{x, y, z\} \to \Lambda$, and after suitable renaming we can arrange that $\alpha(x) \leq \alpha(y) \leq \alpha(z)$, hence $d(x, z) = d(x, y) + d(y, z)$, so $y \in [x, z]$. If S is a finite subset of A, and x, $y \in S$ are chosen with $d(x, y)$ as large as possible, then it follows that $S \subseteq [x, y]$, so every finite subset of A is linear. The next lemma improves on this observation.

Lemma 3.2. *A subset A of a Λ-tree (X, d) is linear if and only if every three-element subset of A is linear, in which case the subtree of X spanned by A is linear.*

Proof. Clearly a subset of a linear set is linear, and a subset with at most one element is linear. Assume every three-element subset of A is linear, and x, y are distinct elements of A. If $z \in A$, there is a segment σ with $\{x, y, z\} \subseteq \sigma$, by the observation preceding the lemma. By Lemma 3.1, there is a unique isometry $\beta : \sigma \to \Lambda$ such that $\beta(x) = 0$ and $\beta(y) = d(x, y)$. If τ is another segment with $\{x, y, z\} \subseteq \tau$, and β' is the corresponding isometry from τ to Λ, then β and β' agree on $\sigma \cap \tau$. (By 3.1, $\beta'|_{\sigma\cap\tau}$ has an extension to σ, and this extension agrees with β on $\{x, y\}$, so is equal to β by 3.1.) Thus, setting $\alpha(z) = \beta(z)$ defines a map $\alpha : A \to \Lambda$. If u, $v \in A$, again by the observation before the lemma, there is a segment σ in X with $\{x, y, u, v\} \subseteq \sigma$, and if $\beta : \sigma \to \Lambda$ is defined as above, then by definition of α, $\alpha(u) = \beta(u)$ and $\alpha(v) = \beta(v)$, hence $d(u, v) = |\beta(u) - \beta(v)| = |\alpha(u) - \alpha(v)|$, showing α is an isometry.

Finally, if A is linear, then $A \subseteq L$ for some linear subtree L, and the subtree spanned by A is contained in L. Clearly a subtree of a linear tree is linear, and the lemma follows. □

Exercises. (1) Show that if L and L' are linear subtrees of a Λ-tree which intersect in a single point, which is an endpoint of both, then $L \cup L'$ is also a linear subtree.

(2) Show that x is an endpoint of X (i.e. has degree at most 1-see the definition after Lemma 1.7) if and only if whenever $x \in [y, z]$, where y, $z \in X$, then either $x = y$ or $x = z$. Also, a Λ-tree is linear if and only if it has no branch points (points of degree greater than 2).

Remark. If x is an endpoint of X, and y, z are points of $X \setminus \{x\}$, then $[x,y]$, $[x,z]$ determine the same direction at x, so $[x,y] \cap [x,z] = [x,u]$ for some $u \neq x$, in particular $[x,y] \cap [x,z]$ has at least two elements. A simple induction shows that, if $y_i \in X \setminus \{x\}$ for $1 \leq i \leq n$, then $[x, y_1] \cap \cdots \cap [x, y_n]$ has at least two elements.

Let (X, d) be a Λ-tree. Take $x \in X$, and let L be a linear subtree of X having x as an endpoint. We call such a subtree of X a linear subtree from x. Define $\alpha : L \to \Lambda$ by $\alpha(y) = d(x, y)$ for $y \in L$. If y, $z \in L$, then by the observations preceding 3.2, one of x, y, z is in the segment determined by the other two. Since x is an endpoint of L, either $y \in [x, z]$ or $z \in [x, y]$, whence $d(y, z) = |d(x,y) - d(x, z)|$ (by Lemma 1.4), showing α is an isometry.

This isometry α induces a linear ordering on L ($x \leq y$ if and only if $\alpha(x) \leq \alpha(y)$). If $y \in L$, then it is easy to see that $L_y = \{z \in L \mid y \leq z\}$ is a linear subtree from y, that $L = [x, y] \cup L_y$ and $[x, y] \cap L_y = \{y\}$. Conversely, if L' is a linear subtree from y, and $x \in X$ is such that $[x, y] \cap L' = \{y\}$, then $[x, y] \cup L'$ is a linear subtree from x. We leave it as an exercise to prove this.

Note that, if $\alpha' : L \to \Lambda$ is any other isometry, then by Lemma 1.2.1, using an argument similar to 3.1, there exists $c \in \Lambda$ such that either $\alpha = \alpha' + c$ or $\alpha = c - \alpha'$. Hence α' is either an increasing or decreasing function on L.

If L is a linear subtree from x, $\alpha : L \to \Lambda$ is the isometry defined above and $u \in L$, $\beta : L_u \to \Lambda$ is defined by $\beta(y) = d(u, y)$, then $\beta(y) = \alpha(y) - d(x, u)$ for $y \in L_u$. Consequently, the ordering on L_u defined by β coincides with the restriction of the ordering on L induced by α. Hence, if $v \in L_u$, $(L_u)_v = L_v$. Also, it follows that if L, L' are linear subtrees from x, x' respectively, and $L_u = L'_u$ for some $u \in L \cap L'$, then $L_v = L'_v$ for all $v \in L_u$.

Lemma 3.3. *Suppose L is a linear subtree from x in a Λ-tree (X, d), and y, $z \in X$. Suppose $L \cap [y, z] \neq \emptyset$. If $L \cap [y, z]$ is bounded above in L, relative to the ordering defined above, then it is a segment, otherwise $L \cap [y, z] = L_v$ for some $v \in L$, and either $L \subsetneq [x, y]$ or $L \subsetneq [x, z]$.*

Proof. Let $v = \text{Y}(x, y, z)$ and choose $u \in L \cap [y, z]$. By Lemma 1.2, $u \in [y, v] \cup [v, z]$, and $v \in [x, y] \cap [x, z]$, so either $[x, v, u, y]$ or $[x, v, u, z]$ is a piecewise segment. In any case, $v \in [x, u]$, so $v \in L$. If w is an upper bound for $L \cap [y, z]$ in L, then $L \cap [y, z] = [x, w] \cap [y, z]$, which is a segment by Lemma 1.7. If $L \cap [y, z]$ is unbounded above in L, then $L_v \subseteq L \cap [y, v, z]$. Since v is an endpoint of L_v either $L_v \subseteq [y, v]$ or $L_v \subseteq [z, v]$. In the first case $L \cap [y, z] = ([x, v] \cup L_v) \cap [y, z] = L_v$, because $[x, v] \cap [y, z] = \{v\}$ (see the remarks after Lemma 2.1). Also, $L_v \underset{\neq}{\subset} [y, v]$ since L_v is unbounded above in L. Hence $L = [x, v] \cup L_v \underset{\neq}{\subset} [x, v] \cup [v, y] = [x, y]$, since $[x, v] \cap [v, y] = \{v\}$. Similarly, if $L_v \subseteq [z, v]$ then $L \underset{\neq}{\subset} [x, z]$. $\square$

Definition. Let (X, d) be a Λ-tree and suppose $x \in X$. A linear subtree from x which is maximal with respect to inclusion is called an X-ray from x. Given X-rays (not necessarily with the same endpoint) L, L', we define $L \equiv L'$ to mean $L \cap L' = L_v = L'_v$ for some v.

Lemma 3.4. (1) *X-rays are closed subtrees of X.*
(2) *The relation $\equiv$ just defined is an equivalence relation on the set of X-rays.*

Proof. (1) If $[y, z]$ is a segment in X, and L is an X-ray from x, then it is impossible that $L \underset{\neq}{\subset} [x, y]$ or $L \underset{\neq}{\subset} [x, z]$, because $[x, y]$ and $[x, z]$ are linear subtrees from x, so by Lemma 3.3, $L \cap [y, z]$ is either empty or a segment.
(2) Clearly $\equiv$ is reflexive and symmetric. Suppose L, L' and L'' are X-rays, from x, x', x'' respectively, and $L \cap L' = L_v = L'_v$, $L' \cap L'' = L'_{v'} = L''_{v'}$ for some v, v'. Let $w = \max\{v, v'\}$ in L', and put $u = \text{Y}(x, w, x'')$. Since $w \in L'_v = L_v$, $L_w = L'_w$ by the observations preceding Lemma 3.3, and similarly $L''_w = L'_w$. By definition of u, $u \in [x, w] \subseteq L$, so $L = [x, u] \cup L_u$, and $w \in L_u$, so $L_u = [u, w] \cup L_w$. Similarly $L'' = [x'', u] \cup L''_u$ and $L''_u = [u, w] \cup L''_w$. It follows that $L_u = L''_u$, and

$$L \cap L'' = ([x, u] \cup L_u) \cap ([x'', u] \cup L''_u) = L_u = L''_u$$

because $[x, u] \cap [x'', u] = \{u\}$ by definition of u, $[x, u] \cap L''_u = [x, u] \cap L_u = \{u\}$, etc. This proves transitivity. $\square$

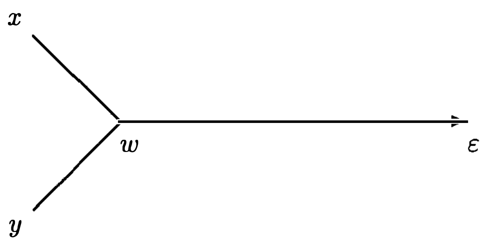
Definition. Let (X, d) be a Λ-tree. The equivalence classes of X-rays for the equivalence relation $\equiv$ defined above are called *ends* of X. If ε is an end of X, and $L \in \varepsilon$, we say that L represents ε.

If y is an endpoint of X then there is an end ε_y whose X-rays are the segments $[x, y]$ for $x \in X$, $x \neq y$. Also, if X has just a single point x, there is an end ε_x whose only ray is $[x, x] = \{x\}$. Ends of the form ε_y for some $y \in X$ are called *closed ends* of X, and ends not of this form are called *open ends* of X.

Lemma 3.5. *Let ε be an open end of a Λ-tree (X,d). For every $x \in X$, there is a unique X-ray from x representing ε. (It is denoted by $[x,\varepsilon\rangle$.)*

Proof. Let L be an X-ray representing ε. Since L is a closed subtree (3.4), there is a bridge $[x,y]$ from x to L (1.9 and the remarks following it). Then L_y is an X-ray from y, since it contains more than one point, because ε is an open end. Hence $L' = [x,y] \cup L_y$ is an X-ray from x (it is linear by Exercise 1 above, and if $L' \subsetneqq L''$, where L'' is a linear subtree from x, then $L_y \subsetneqq L''_y$, a contradiction). Since $L \cap L' = L_y = L'_y$, L' represents ε. If L, L' are two X-rays from x representing ε, then $L \cap L' = L_y = L'_y$ for some y. Since $x \in L \cap L'$, $L = L_y = L'_y = L'$. □

If ε_y is a closed end, we put $[x,\varepsilon_y\rangle = [x,y]$, even when $x = y$. If x and y are points and ε is an end of the Λ-tree (X,d), note that, in all cases, $[x,\varepsilon\rangle \cap [y,\varepsilon\rangle = [w,\varepsilon\rangle$ for some unique $w \in X$. Further, $[x,\varepsilon\rangle = [x,w] \cup [w,\varepsilon\rangle$, $[y,\varepsilon\rangle = [y,w] \cup [w,\varepsilon\rangle$, $[x,w] \cap [w,\varepsilon\rangle = \{w\} = [y,w] \cup [w,\varepsilon\rangle$ and $[x,y] = [x,w,y]$. The situation can be illustrated by the following picture.



As an example, the open interval $(0,1)$ in $\mathbb{R}$ is an $\mathbb{R}$-tree having two open ends, corresponding to 0 and 1. Thus distances on a ray $[x,\varepsilon\rangle$, where ε is an open end, need not be unbounded, in general. This contrasts with the situation for $\mathbb{Z}$-trees.

Lemma 3.6. *Suppose L, L' are closed, linear subtrees in a Λ-tree (X,d), with endpoints x, x' respectively, and $L \cap L' \neq \emptyset$. If $L \cap L'$ is bounded above in L or L', relative to the ordering defined above, then it is a segment. Otherwise, $L \cap L' = L_v = L'_v$ for some $v \in L \cap L'$.*

Proof. Suppose z is an upper bound for $L \cap L'$ in L. Then $L \cap L' = [x,z] \cap L'$ which is a segment since L' is a closed subtree. Suppose $L \cap L'$ is unbounded above in both L and L'. Take $u \in L \cap L'$. Then $L_u \subseteq L \cap L'$, so $L \cap L' = ([x,u] \cap L') \cup L_u$. Since L' is a closed subtree, $[x,u] \cap L' = [v,u]$ for some v. Hence $L \cap L' = [v,u] \cup L_u = L_v$. Similarly, $L \cap L' = L'_{v'}$ for some v'. Since $L \cap L'$ is unbounded in L and L', it cannot have two endpoints, so $v = v'$. □

Lemma 3.7. (1) *Suppose* ε, ε' *are distinct ends of a* Λ*-tree* (X,d). *Then there exists* $x \in X$ *such that* $[x,\varepsilon\rangle \cap [x,\varepsilon'\rangle = \{x\}$. *Further, the set of all* $x \in X$ *with this property is a linear subtree of* X, *denoted by* $\langle\varepsilon,\varepsilon'\rangle$. *For all* $x \in \langle\varepsilon,\varepsilon'\rangle$, $\langle\varepsilon,\varepsilon'\rangle = [x,\varepsilon\rangle \cup [x,\varepsilon'\rangle$. *For all* $y \in X$, $[y,\varepsilon\rangle \cap [y,\varepsilon'\rangle \cap \langle\varepsilon,\varepsilon'\rangle$ *contains a single point of* X.

(2) *If* ε, ε', ε'' *are three distinct ends of* (X,d), *then* $\langle\varepsilon,\varepsilon'\rangle \cap \langle\varepsilon',\varepsilon''\rangle \cap \langle\varepsilon'',\varepsilon\rangle$ *contains a single point of* X.

Proof. (1) Take $y \in X$. By Lemma 3.6, since X-rays are closed subtrees, $[y,\varepsilon\rangle \cap [y,\varepsilon'\rangle$ is a segment, and has y as an endpoint, so is $[y,x]$ for some $x \in X$. Then $[y,\varepsilon\rangle = [y,x] \cup [x,\varepsilon\rangle$ and $[y,x] \cap [x,\varepsilon\rangle = \{x\}$, so x is the only point of $[x,\varepsilon\rangle$ in $[y,\varepsilon\rangle \cap [y,\varepsilon'\rangle$. Similarly it is the only point of $[x,\varepsilon'\rangle$ in $[y,\varepsilon\rangle \cap [y,\varepsilon'\rangle$, hence $[x,\varepsilon\rangle \cap [x,\varepsilon'\rangle = \{x\}$. It follows that $[x,\varepsilon\rangle \cup [x,\varepsilon'\rangle$ is a linear subtree by the exercise after Lemma 3.2. Denote it by L. Note that $[y,\varepsilon\rangle \cap [y,\varepsilon'\rangle \cap L = [y,x] \cap L = \{x\}$.

Suppose x' has the property that $[x',\varepsilon\rangle \cap [x',\varepsilon'\rangle = \{x'\}$, so that $L' = [x',\varepsilon\rangle \cup [x',\varepsilon'\rangle$ is also a linear subtree. Then $[x,\varepsilon\rangle \cap [x',\varepsilon\rangle = [y,\varepsilon\rangle$ for some $y \in X$, and $[x,\varepsilon'\rangle \cap [x',\varepsilon'\rangle = [z,\varepsilon'\rangle$ for some $z \in X$. Since x', y, $z \in L'$ they are collinear, and since $[x',\varepsilon\rangle \cap [x',\varepsilon'\rangle = \{x'\}$, it follows that $[y,z] = [y,x',z]$, hence $x' \in L$. Similarly, $[y,z] = [y,x,z]$ and

$$\begin{aligned} L &= [x,\varepsilon\rangle \cup [x,\varepsilon'\rangle \\ &= ([x,y] \cup [y,\varepsilon\rangle) \cup ([x,z] \cup [z,\varepsilon'\rangle) \\ &= [y,\varepsilon\rangle \cup [y,z] \cup [z,\varepsilon'\rangle \\ &= ([x',y] \cup [y,\varepsilon\rangle) \cup ([x',z] \cup [z,\varepsilon'\rangle) \\ &= [x',\varepsilon\rangle \cup [x',\varepsilon'\rangle. \end{aligned}$$

It remains to show that, if $u \in L$, then $[u,\varepsilon\rangle \cap [u,\varepsilon'\rangle = \{u\}$. If $u \in [x,\varepsilon\rangle$, then $[x,\varepsilon\rangle = [x,u] \cup [u,\varepsilon\rangle$ and $[x,u] \cap [u,\varepsilon\rangle = \{u\}$. Also, $[u,x] \cap [x,\varepsilon'\rangle = \{x\}$ (since $[u,x] \subseteq [x,\varepsilon\rangle$), so $[u,\varepsilon'\rangle = [u,x] \cup [x,\varepsilon'\rangle$. Since $[u,\varepsilon\rangle \subseteq [x,\varepsilon\rangle$, we have $[x,\varepsilon'\rangle \cap [u,\varepsilon\rangle \subseteq \{x\}$, and there is equality if and only if $x = u$ (because $[x,u] \cap [u,\varepsilon\rangle = \{u\}$). Hence $[u,\varepsilon\rangle \cap [u,\varepsilon'\rangle = \{u\}$, and similarly this is true if $u \in [x,\varepsilon'\rangle$.

(2) We can choose $x \in \langle\varepsilon,\varepsilon'\rangle \cap \langle\varepsilon,\varepsilon''\rangle$ and $x' \in \langle\varepsilon,\varepsilon'\rangle \cap \langle\varepsilon',\varepsilon''\rangle$, $x'' \in \langle\varepsilon,\varepsilon''\rangle \cap \langle\varepsilon'',\varepsilon'\rangle$. Let $w = \mathrm{Y}(x,x',x'')$. Then

$$\{w\} = [x,x'] \cap [x',x''] \cap [x'',x] \subseteq \langle\varepsilon,\varepsilon'\rangle \cap \langle\varepsilon',\varepsilon''\rangle \cap \langle\varepsilon'',\varepsilon\rangle.$$

By (1),

$$\langle\varepsilon,\varepsilon'\rangle \cap \langle\varepsilon',\varepsilon''\rangle \cap \langle\varepsilon'',\varepsilon\rangle = ([y,\varepsilon\rangle \cup [y,\varepsilon'\rangle) \cap ([y,\varepsilon'\rangle \cup [y,\varepsilon''\rangle) \cap ([y,\varepsilon''\rangle \cup [y,\varepsilon\rangle).$$

Also by (1), any two of $[y,\varepsilon\rangle$, $[y,\varepsilon'\rangle$, $[y,\varepsilon''\rangle$ have only y in common, and (2) follows. □

We can use the previous result to extend the notation $\mathrm{Y}(x,y,z)$, following §**E1** in [5]. Let (X,d) be a Λ-tree and put $\hat{X} = X \cup Ends(X)$, where $Ends(X)$ denotes the set of ends of X, and an endpoint x of X is identified with the end ε_x it defines. Thus $\hat{X}$ may be viewed as the disjoint union of X and the set of open ends of X. We can define the closed segment $[x,y]$ for x, $y \in \hat{X}$ as follows. If x, $y \in X$, then $[x,y]$ has its usual meaning. If $x \in X$ and y is an open end of X, we put $[x,y] = [x,y\rangle \cup \{y\}$, and if x, y are both open ends of X, $[x,y] = \langle x,y\rangle \cup \{x,y\}$, where $\langle x,y\rangle$ is as defined in 3.7, with the convention that $\langle\varepsilon,\varepsilon\rangle = \emptyset$ if ε is an end of X. We can also define open and half-open segments, for example $[x,y) = [x,y] \setminus \{y\}$.

We now define, for x, y, $z \in \hat{X}$, $\mathrm{Y}(x,y,z)$ to be the unique point in $[x,y]\cap[y,z]\cap[z,x]$. It is easy to see there is exactly one point in this intersection in all cases, using 3.7. Note that $\mathrm{Y}(x,y,z) \in X$ unless two of x, y, z equal some open end ε, in which case $\mathrm{Y}(x,y,z) = \varepsilon$. Further, there is an analogue of Lemma 1.2(1) in all cases. That is, if $w = \mathrm{Y}(x,y,z)$, then $[x,y]\cap[x,z] = [x,w]$ and $[x,y] = [x,w] \cup [w,y]$, and similar statements with x, y, z permuted hold. This is left as an exercise. The next remark will be useful in some several proofs.

Remark. Let L be a linear subtree from a point y in a Λ-tree (X,d). If L is not an X-ray, then $L \subsetneqq L'$ for some linear subtree of X from y, and taking $z \in L' \setminus L$, we have $L \subsetneqq [y,z]$.

If Y is a subtree of a Λ-tree (X,d), we can ask about the connection between the ends of X and those of Y. We say that an end ε of X belongs to Y if $[y,\varepsilon\rangle \subseteq Y$ for some $y \in Y$.

Lemma 3.8. *Suppose Y is a subtree of a Λ-tree (X,d).*

(1) *If ε is an end of X belonging to Y, then $[y,\varepsilon\rangle \subseteq Y$ for all $y \in Y$, and if ε' is another end of X belonging to Y, then $\langle\varepsilon,\varepsilon'\rangle \subseteq Y$.*
(2) *If Y is a closed subtree of X, then every open end of Y uniquely determines an open end of X.*
(3) *If Y is a closed subtree of X, ε is an end of X, and $y \in Y$, then either ε belongs to Y, or $[y,\varepsilon\rangle \cap Y = [y,y_\varepsilon]$ for some $y_\varepsilon \in Y$ which is independent of y.*

Proof. (1) By assumption $[z,\varepsilon\rangle \subseteq Y$ for some $z \in Y$, and $[z,\varepsilon\rangle \cap [y,\varepsilon\rangle = [v,\varepsilon\rangle$ for some v, as noted after Lemma 3.5. Since $[z,\varepsilon\rangle \subseteq Y$, $[v,\varepsilon\rangle \subseteq Y$. Further, $[y,\varepsilon\rangle = [y,v]\cup[v,\varepsilon\rangle$, and $[y,v] \subseteq Y$. To prove the last part, if $y \in Y$, then from the proof of Lemma 3.7 and what has already been proved, $[y,\varepsilon\rangle \cap [y,\varepsilon'\rangle = [x,y] \subseteq Y$ for some x, so $x \in Y$, and $\langle\varepsilon,\varepsilon'\rangle = [x,\varepsilon\rangle \cup [x,\varepsilon'\rangle \subseteq Y$.

(2) Let ε be an open end of Y, and let $y \in Y$. We need to know that the unique Y-ray from y representing ε is an X-ray. Call this Y-ray L. If it is not an X-ray, then by the remark preceding the lemma, $L \subseteq [y, z]$ for some $z \in X$. Since Y is a closed subtree, $Y \cap [y, z] = [y, v]$ for some $v \in Y$. Then $L \subseteq [y, v]$, contradicting the fact that L is a Y-ray representing an open end of Y.

(3) Suppose ε is an end of X not belonging to Y and $y \in Y$. Then by (1) there exists $z \in [y, \varepsilon\rangle$ such that $z \notin Y$. Since $[y, \varepsilon\rangle$ is linear with y as an endpoint, $Y \cap [y, \varepsilon\rangle \subseteq [y, z]$, so $Y \cap [y, \varepsilon\rangle = Y \cap [y, z] = [y, v]$ for some $v \in Y$ (since Y is closed) and $Y \cap [v, \varepsilon\rangle = \{v\}$. If $y' \in Y$, then $[y, \varepsilon\rangle \cap [y', \varepsilon\rangle = [w, \varepsilon\rangle$ for some w, and $w \in Y$ since $[y, y'] = [y, w, y']$. Hence $w \in [y, v]$, and since $[y', \varepsilon\rangle = [y, w] \cup [w, \varepsilon\rangle$ is linear from y, $v \in [w, \varepsilon\rangle \subseteq [y', \varepsilon\rangle$. Hence $[y', \varepsilon\rangle = [y', v] \cup [v, \varepsilon\rangle$, which intersects Y in $[y', v]$. □

If Y is a subtree of X, more generally if $\varphi : Y \to X$ is an isometry of Λ-trees, we can define a map $\hat{\varphi} : \hat{X} \to \hat{Y}$. For notational convenience in making the definition we shall treat φ as an inclusion and suppress it.

Choose $y \in Y$; take $x \in \hat{X}$ and put $L = [y, x] \cap Y$. Then L is a linear subtree of Y with y as an endpoint, and there are two cases.

(1) $L = [y, e]$ for some $e \in Y$. Then $[e, x] \cap Y = \{e\}$. We claim that e depends only on x, not on the choice of y. For if $y_1 \in Y$, let $w = \mathrm{Y}(x, y, y_1)$. Then $w \in [y_1, y] \subseteq Y$, and $w \in [y, x]$, so $w \in [y, e]$, hence $[w, x] = [w, e] \cup [e, x]$, which intersects Y in $[w, e]$. Thus $[y_1, x] \cap Y = ([y_1, w] \cup [w, x]) \cap Y = [y_1, w] \cup ([w, x] \cap Y) = [y_1, w] \cup [w, e] = [y_1, e]$, because $w \in [y_1, x]$, and $[y_1, w] \cap [w, e] \subseteq [y_1, w] \cap [w, x] = \{w\}$. We put $\hat{\varphi}(x) = e$.

(2) $L \neq [y, e]$ for all $e \in Y$. Then L is a ray from y in Y. For otherwise, $L \subseteq [y, z]$ for some $z \in Y$ (remark preceding Lemma 3.8), so $L \subseteq [y, z] \cap [y, x] = [y, e]$ for some $e \in Y$, and $[y, e] \subseteq Y$, hence $L = [y, e]$, contrary to assumption. Let ε be the end of Y defined by L, so $L = [y, \varepsilon\rangle$. Then ε depends only on x, not on the choice of y. For if $y_1 \in Y$, let $w = \mathrm{Y}(x, y, y_1)$. By a similar argument to Case (1), $[w, x]$ meets Y in $[w, \varepsilon\rangle$, and $[y_1, x] \cap Y = ([y_1, w] \cup [w, x]) \cap Y = [y_1, w] \cup ([w, x] \cap Y) = [y_1, w] \cup [w, \varepsilon\rangle = [y_1, \varepsilon\rangle$. We put $\hat{\varphi}(x) = \varepsilon$.

If G is a group acting as isometries on a Λ-tree (X, d) by isometries, then G preserves X-rays and equivalence of rays, so there is an induced action on $Ends(X)$ and on $\hat{X}$.

Proposition 3.9. *Let $\varphi : Y \to X$ be an isometry of Λ-trees. Then*

(1) *$\hat{\varphi}$ is the unique map $\hat{X} \to \hat{Y}$ such that for $x \in \hat{X}$ and $y \in Y$,*

$$[y, x] \cap Y = \begin{cases} [y, \hat{\varphi}(x)] & \text{if } \hat{\varphi}(x) \in Y \\ [y, \hat{\varphi}(x)\rangle & \text{otherwise.} \end{cases}$$

(2) *$\hat{\varphi}$ is surjective.*

(3) *If* $\psi : Z \to Y$ *is also an isometry of* Λ*-trees, then* $\widehat{\varphi \circ \psi} = \hat{\psi} \circ \hat{\varphi}$.
(4) *If* φ *is equivariant for actions of a group* G *on* X *and* Y, *then* $\hat{\varphi}$ *is* G*-equivariant for the induced actions on* $\hat{X}$, $\hat{Y}$.

Proof. (1) is clear from the definition. To prove (2), we again view φ as an inclusion. If $y \in Y$, then $y = \hat{\varphi}(y)$. If ε is an open end of Y, let $y \in Y$ and let $L = [y, \varepsilon\rangle$, a ray from y in Y. If L is a ray from y in X, it defines an end ε' of X, and $\varepsilon = \hat{\varphi}(\varepsilon')$. Otherwise, $L \subseteq [y, x]$ for some $x \in X$, and then $\varepsilon = \hat{\varphi}(x)$. The proof of (3) is left to readers, and (4) follows easily. □

We next consider a generalisation of the idea of a Dedekind cut in the context of Λ-trees.

Definition. A cut in a Λ-tree (X, d) is a pair of disjoint non-empty subtrees (X_1, X_2) of X such that $X = X_1 \cup X_2$.

Lemma 3.10. *Suppose* (X_1, X_2) *is a cut in a* Λ*-tree* (X, d).
(a) *If* X_1 *is not a closed subtree of* X, *then there is a unique end* ε *of* X_1 *such that, for all* $x_1 \in X_1$, $x_2 \in X_2$, $X_1 \cap [x_1, x_2] = [x_1, \varepsilon\rangle$, *and* ε *is an open end of* X_1.
(b) *If* X_1, X_2 *are both closed subtrees of* X, *and* $[z_1, z_2]$ *is the bridge between them, then* $d(z_1, z_2)$ *is the least positive element of* Λ.

Proof. (a) There is a segment $[x, y]$ in X such that $X_1 \cap [x, y]$ is non-empty and is not a segment. Clearly one endpoint, say x, is in X_1 and the other is in X_2. Putting $L = X_1 \cap [x, y]$, it follows that L is a linear subtree in X_1 from x, and is not a segment. We show that L is an X_1-ray from x. For otherwise, $L \subseteq [x, z]$ for some $z \in X_1$ (remark preceding Lemma 3.8). Let $w = \mathrm{Y}(x, y, z)$. Then

$$\begin{aligned} [x, w] &= [x, y] \cap [x, z] \\ &= [x, y] \cap ([x, z] \cap X_1) \quad \text{(since } x, z \in X_1) \\ &= ([x, y] \cap X_1) \cap [x, z] \\ &= L \cap [x, z] = L \end{aligned}$$

contradicting that L is not a segment. Let ε be the end of X_1 defined by L. Since L is not a segment, ε is an open end.

Take $x_i \in X_i$, $i = 1$, 2, and let $L' = X_1 \cap [x_1, x_2]$, a linear subtree of X_1 from x_1. By Lemma 1.8, $[x, y] \cap [x_1, x_2] = [p, q]$ for some $p \in X_1$, $q \in X_2$. Then $L \cap L' = [p, q] \cap X_1$. Since $p \in [x, y]$, $L = ([x, p] \cup [p, y]) \cap X_1 = [x, p] \cup ([p, y] \cap X_1) = [x, p] \cup ([p, q] \cap X_1)$ (because $[p, y] = [p, q, y]$ and $[q, y] \subseteq X_2$), and $[x, p] \cap ([p, q] \cap X_1) = \{p\}$. Hence $[p, q] \cap X_1 = L_p$, and similarly $[p, q] \cap X_1 = L'_p$. Thus $L \cap L' = L_p = L'_p$, so L' is an X_1-ray from x_1 representing ε, and this proves (a).

(b) Since $[z_1, z_2] \cap X_i = \{z_i\}$ for $i = 1$, 2, it follows that $[z_1, z_2] = \{z_1, z_2\}$, so $[0, d(z_1, z_2)]_\Lambda$ has just two points. □

Definition. A cut (X_1, X_2) in a Λ-tree (X, d) is called an open cut if neither X_1 nor X_2 is a closed subtree of X.

In the case of an open cut we have the following, which is an immediate consequence of Lemma 3.10.

Corollary 3.11. *If (X_1, X_2) is an open cut in a Λ-tree (X, d), then there are unique open ends ε_1, ε_2 of X_1, X_2 respectively such that for all $x_i \in X_i$ ($i = 1, 2$), we have $[x_1, x_2] \cap X_i = [x_i, \varepsilon_i\rangle$, and $[x_1, x_2]$ is the disjoint union $[x_1, \varepsilon_1\rangle \cup [x_2, \varepsilon_2\rangle$.*

□

In the situation of Cor. 3.11, we say that the ends ε_1, ε_2 meet at the cut (X_1, X_2).

As examples of cuts, removing an unoriented edge from a simplicial tree results in two disjoint subtrees, and this gives a cut in the corresponding $\mathbb{Z}$-tree, which is an example of Lemma 3.10(b). Writing $\mathbb{R} = (-\infty, 0] \cup (0, \infty)$ gives a cut in the $\mathbb{R}$-tree $\mathbb{R}$, with one subtree closed and the other not. For an example of an open cut, take $\Lambda = \mathbb{Q} = X$, $X_1 = (-\infty, \sqrt{2})_{\mathbb{R}} \cap \mathbb{Q}$ and $X_2 = (\sqrt{2}, \infty)_{\mathbb{R}} \cap \mathbb{Q}$. We leave it as an exercise to give an example of an open cut when $\Lambda = \mathbb{Z} \oplus \mathbb{Z}$ with the lexicographic ordering.

One further significant idea in the theory of ends of Λ-trees is the notion of an end map, which we now describe.

Definition. Let ε be an end of a Λ-tree (X, d). An end map associated to ε is a mapping $\delta : X \to \Lambda$ such that for all $x \in X$, the restriction of δ to $[x, \varepsilon\rangle$ is an isometry.

Lemma 3.12. *Let (X, d) be a Λ-tree, and let ε be an end of X. Then there is an end map associated to ε. In fact, if $x_0 \in X$ is such that $[x_0, \varepsilon\rangle$ has more than one point, and $\delta_0 : [x_0, \varepsilon\rangle \to \Lambda$ is an isometry, then there is a unique end map $\delta : X \to \Lambda$ whose restriction to $[x_0, \varepsilon\rangle$ is δ_0.*

If δ, δ' are two end maps associated to ε then there exists $c \in \Lambda$ such that either $\delta = \delta' + c$ (i.e. $\delta(x) = \delta'(x) + c$ for all $x \in X$) or $\delta = c - \delta'$.

Proof. If X has only a single point, this is clear, otherwise we can take $x_0 \in X$ such that $[x_0, \varepsilon\rangle$ has at least two points, and an isometry $\delta_0 : [x_0, \varepsilon\rangle \to \Lambda$. Let $x \in X$, so $[x_0, \varepsilon\rangle \cap [x, \varepsilon\rangle = [y, \varepsilon\rangle$ for some $y \in X$ by the observations after Lemma 3.5. Further, if $x \notin [x_0, \varepsilon\rangle$, then $[y, \varepsilon\rangle$ has at least two elements. This is clear if ε is an open end, and follows from the remark shortly after Lemma 3.2 if ε corresponds to an endpoint of X. By Lemma 3.1, for any $x \in X$ there is a unique isometry $\delta_x : [x, \varepsilon\rangle \to \Lambda$ agreeing with δ_0 on $[y, \varepsilon\rangle$. If also $z \notin [x_0, \varepsilon\rangle$ then similarly $[x_0, \varepsilon\rangle \cap [x, \varepsilon\rangle \cap [z, \varepsilon\rangle$ has at least two elements, so δ_x and δ_z agree on $[x, \varepsilon\rangle \cap [z, \varepsilon\rangle$, for any x, $z \in X$. Hence, the maps δ_x, for $x \in X$, define an end map δ ($\delta = \bigcup_{x \in X} \delta_x$), and this is the only end map whose restriction to $[x_0, \varepsilon\rangle$ is δ_0.

If δ_0' is another isometry from $[x_0, \varepsilon\rangle$ to Λ, then $\delta_0' \circ (\delta_0)^{-1}$ is an isometry from the image of δ_0 into Λ, so by Lemma 1.2.1 is the restriction of either a translation or reflection. Hence either $\delta_0 = \delta_0' + c$ or $\delta_0 = c - \delta_0'$ for some $c \in \Lambda$. In the first case $\delta = \delta' + c$ since both are end maps with the same restriction to $[x_0, \varepsilon\rangle$. Similarly in the second case $\delta = c - \delta'$. □

Addendum. As we observed earlier (before Lemma 3.3), for all $x \in X$, the mapping δ_x is either increasing or decreasing. If δ_0 is increasing, it follows from the proof that all the δ_x are increasing, so for all $x \in X$, $\delta(x)$ is the least value of δ restricted to $[x, \varepsilon\rangle$. An end map with this property is called an end map towards ε. If δ, δ' are two end maps towards ε, then for some $c \in \Lambda$, $\delta = \delta' + c$. Also, for any end map associated to ε, either δ or $-\delta$ is an end map towards ε.

Corollary 3.13. *If (X_1, X_2) is an open cut in a Λ-tree (X, d), there is a mapping $\alpha : X \to \Lambda$ such that for all $x_i \in X_i$ ($i = 1, 2$), the restriction of α to $[x_1, x_2]$ is an isometry. If α' is another such map, then for some constant $c \in \Lambda$, either $\alpha = \alpha' + c$ or $\alpha = c - \alpha'$.*

Proof. Let ε_i ($i = 1, 2$) be the ends meeting at the cut (with ε_i belonging to X_i). Let $y_1 \in X_1$, $y_2 \in X_2$, and let $\delta_0 : [y_1, y_2] \to \Lambda$ be an isometry. As in the proof of 3.12, we can find end maps $\delta_i : X_i \to \Lambda$ associated to ε_i such that δ_i restricted to $[y_i, \varepsilon_i\rangle$ is δ_0 restricted to $[y_i, \varepsilon_i\rangle$ (recall that $[y_1, y_2] = [y_1, \varepsilon_1\rangle \cup [y_2, \varepsilon_2\rangle$). Set $\alpha = \delta_1 \cup \delta_2$. If $x_i \in X_i$ then $[x_1, x_2] \cap [y_1, y_2] = [z_1, z_2]$ for some $z_i \in X_i$ by Lemma 1.8. Then α is an isometry on $[x_1, \varepsilon_1\rangle$, $[x_2, \varepsilon_2\rangle$ and $[z_1, z_2]$. By Lemma 3.1, the restriction of α to $[z_1, z_2]$ has a unique extension to an isometry $\beta : [x_1, x_2] \to \Lambda$. Now β agrees with α on $[x_1, \varepsilon_1\rangle \cap [z_1, z_2] = [x_1, x_2] \cap X_1 \cap [z_1, z_2] = [z_1, z_2] \cap X_1 = [z_1, \varepsilon_1\rangle$, which has at least two points since ε_1 is an open end, so β and α agree on $[x_1, \varepsilon_1\rangle$ by 3.1. Similarly, they agree on $[x_2, \varepsilon_2\rangle$. Hence α restricted to $[x_1, x_2]$ is β, so is an isometry.

It follows from 3.12 that α is uniquely determined by its restriction to $[y_1, y_2]$, and the last part of the corollary follows by a similar argument to that in Lemma 3.12. □

We finish with a result which will be needed later in the classification of group actions by isometries on Λ-trees. The proof of the second part of the theorem is quite long, and seems to be one of the few places where the theory of Λ-trees for general Λ is noticeably more difficult than the theory of $\mathbb{R}$-trees. The extra work in dealing with general Λ arises mainly from the possibility of cuts occurring in Λ-trees.

Theorem 3.14. *Suppose (X, d) is a Λ-tree and $(A_i)_{i \in I}$ is a family of closed subtrees of X such that, for all i, $j \in I$, $A_i \cap A_j \neq \emptyset$, but $\bigcap_{i \in I} A_i = \emptyset$.*

(i) *At most one end of X belongs to all the A_i. Such an end ε must be open, and for all $i \in I$ and $x \in A_i$, $[x, \varepsilon\rangle \subseteq A_i$.*

(ii) *Suppose there is no end of X common to all A_i. Then there is a unique cut (X_1, X_2) of X such that for all $i \in I$ and $j = 1, 2$, $A_i \cap X_j \neq \emptyset$. The cut is open, and if ε_j $(j = 1, 2)$ are the open ends which meet at the cut, then for all $i \in I$ and $j = 1, 2$, if $x \in A_i \cap X_j$, then $[x, \varepsilon_j\rangle \subseteq A_i \cap X_j$.*

Proof. (i) If ε, ε' are distinct ends belonging to all A_i, then by Lemma 3.8, $\langle\varepsilon, \varepsilon'\rangle \subseteq A_i$ for all $i \in I$, contradicting $\bigcap_{i \in I} A_i = \emptyset$. If ε is a closed end of X belonging to A_i for all $i \in I$, corresponding to an endpoint e of X, then $[x, \varepsilon\rangle = [x, e]$ for $x \in X$, so by Lemma 3.8, $e \in \bigcap_{i \in I} A_i = \emptyset$, contradiction. The last part also follows from 3.8.

(ii) Take $x \in X$. For $i \in I$, let x_i be the projection of x onto A_i, as defined after Lemma 1.9. If F is a finite subset of I and $Y = \bigcap_{i \in F} A_i$, then $Y \neq \emptyset$ by Lemma 1.11, and Y is a closed subtree, being the intersection of finitely many closed subtrees. Let y be the projection of x onto Y. Then for all $i \in F$, $y \in A_i$, so $[x, y] = [x, x_i, y]$ by Lemma 1.9. Thus $x_i \in [x, y]$ for all $i \in F$. It follows that any three-element subset of $\{x\} \cup \{x_i \mid i \in I\}$ is linear, hence the subtree $L(x)$ spanned by this set is linear by Lemma 3.2. Now $L(x) = \bigcup_{i \in I} [x, x_i]$ (see the definition preceding Lemma 1.8), and it also follows that $L(x)$ has x as an endpoint. We pause to state and prove a lemma giving various properties of $L(x)$, which is endowed with the linear ordering described before Lemma 3.3.

Lemma. (1) *For $x \in X$ and $i \in I$, $L(x) \cap A_i = L(x)_{x_i}$. (Recall $L(x)_{x_i} = \{z \in L(x) \mid x_i \leq z\}$.)*
(2) *If $w \in L(x)$ then $L(w) = L(x)_w$.*
(3) *If $x, y \in X$ and $L(x) \cap L(y) \neq \emptyset$ then $L(x) \cap L(y) = L(w)$ for some $w \in X$.*
(4) *If $x, y \in X$ and $L(x) \cap L(y) = \emptyset$, then $[x, y] = L(x) \cup L(y)$.*
(5) *If $x, y, z \in X$, then one of $L(x) \cap L(y)$, $L(y) \cap L(z)$, $L(z) \cap L(x)$ is non-empty.*

Proof. (1) If $z \in L(x)$ and $x_i \leq z$ then $z \in [x, x_j]$ for some j, so taking $Y = A_i \cap A_j$ in the discussion above, we have $[x, y] = [x, x_i, z, x_j, y]$, where y is the projection of x onto Y. Thus $z \in [x_i, y] \subseteq A_i$, and (1) follows since $[x, x_i] \cap A_i = \{x_i\}$ by definition of x_i.

(2) If $x_i \leq w$ then by (1) $w \in A_i$, so $w = w_i$. If $w \leq x_i$, then $w \in [x, x_i]$, so $[w, x_i]$ is the bridge between w and A_i, i.e. $w_i = x_i$. Thus

$$L(x)_w = \bigcup_{i \in I} [x, x_i] \cap L(x)_w = \bigcup_{w \leq x_i} [w, x_i] = \bigcup_{i \in I} [w, w_i] = L(w)$$

as required.

(3) Suppose $L(x) \cap L(y) \neq \emptyset$, let $z \in L(x) \cap L(y)$ and put $w = \mathrm{Y}(x, y, z)$. Then $w \in L(x) \cap L(y)$, (since $w \in [x, z] \cap [y, z]$), and $L(w) = L(x)_w = L(y)_w$ by (2). Hence $L(x) \cap L(y) = ([x, w] \cap [y, w]) \cup L(w) = \{w\} \cup L(w) = L(w)$.

(4) Suppose $L(x) \cap L(y) = \emptyset$. If $i \in I$ then $[x_i, y_i] \subseteq A_i$, and $x_i \neq y_i$. Also, $[x, x_i] \cap A_i = \{x_i\}$ and $[y, y_i] \cap A_i = \{y_i\}$, by definition of x_i, y_i. Hence $[x, y] = [x, x_i, y_i, y]$ by Lemma 1.5, so $[x, x_i] \cup [y, y_i] \subseteq [x, y]$, for all $i \in I$. Consequently, $L(x) \cup L(y) \subseteq [x, y]$. Further, since $[x, x_i] \subseteq L(x)$ and $[y, y_i] \subseteq L(y)$, if $z \in [x, y]$ and $z \notin L(x) \cup L(y)$, then $z \in [x_i, y_i]$ for all i, so $z \in \bigcap_{i \in I} A_i$, a contradiction. This proves (4).

(5) Suppose $L(x) \cap L(y)$, $L(y) \cap L(z)$, $L(z) \cap L(x)$ are all empty. Let $w = \mathrm{Y}(x, y, z)$. Then

$$\begin{aligned} w &\in [x, y] \cap [y, z] \cap [z, x] \\ &= (L(x) \cup L(y)) \cap (L(y) \cup L(z)) \cap (L(z) \cup L(x)) \quad \text{(by (4))} \\ &= \emptyset \end{aligned}$$

a contradiction. □

We return to the proof of Theorem 3.14(ii). Fix $p_1 \in X$, and define

$$X_1 = \{x \in X \mid L(x) \cap L(p_1) \neq \emptyset\}.$$

Suppose $x \in X_1$ and $y \in L(x)$; by (2) and (3), $L(x) \cap L(p_1) = L(x)_z$ for some z, and $L(y) = L(x)_y$. Hence $L(x)_y \cap L(x)_z \subseteq L(y) \cap L(p_1)$. The left-hand intersection is either $L(x)_y$ or $L(x)_z$, so is non-empty, hence $y \in X_1$ and we have shown $L(x) \subseteq X_1$ for all $x \in X_1$. Similarly, if x, $y \in X_1$ then $L(x) \cap L(y) \neq \emptyset$, and if $z \in L(x) \cap L(y)$ then by Cor.1.3, $[x, y] \subseteq [x, z] \cup [y, z] \subseteq L(x) \cup L(y) \subseteq X_1$, so X_1 is a subtree of X, and is non-empty, since $p_1 \in X_1$.

Take $x \in X_1$ and suppose $L(x) \subseteq [x, w]$ for some $w \in X_1$. As noted above, $L(x) \cap L(w) \neq \emptyset$, and we choose $u \in L(x) \cap L(w)$. Choose $i \in I$ such that $u \notin A_i$, so $u \neq u_i$. Then $u \in [x, x_i]$, hence $u_i = x_i$, and similarly $u \in [w, w_i]$ and $u_i = w_i$. Thus $[u, u_i] \subseteq [x, u_i] \cap [w, u_i]$, hence $u_i \notin [x, w]$ by the remark before Lemma 1.4, which contradicts $u_i \in L(x)$. It follows that $L(x)$ is an X_1-ray, and defines an open end of X_1. By (2) and (3), this end is independent of the choice of $x \in X_1$, and we denote it by ε_1. By (1) and (2), ε_1 belongs to $A_i \cap X_1$, for all $i \in I$.

By the assumption in (ii) that there is no end common to all A_i, it follows that $X_1 \neq X$. Choose $p_2 \in X \backslash X_1$, and define $X_2 = \{x \in X \mid L(x) \cap L(p_2) \neq \emptyset\}$. Then again X_2 is a subtree of X, the $L(x)$ for $x \in X_2$ define the same open end ε_2 of X_2 and ε_2 belongs to $A_i \cap X_2$ for all $i \in I$. Suppose $x \in X_1 \cap X_2$; using (2) and (3) we can find $w \in L(x)$ such that $L(x)_w \subseteq L(p_1) \cap L(p_2)$. But by the choice of p_2, $L(p_1) \cap L(p_2) = \emptyset$, a contradiction. Hence $X_1 \cap X_2 = \emptyset$, and by (5)

$X = X_1 \cup X_2$. By (4), if $x \in X_1$ and $y \in X_2$, then $[x,y] \cap X_1 = L(x) = [x, \varepsilon_1\rangle$ and $[x,y] \cap X_2 = L(y) = [x, \varepsilon_2\rangle$, which are not segments. It follows that (X_1, X_2) is an open cut, and that ε_j, $j = 1$, 2 are the open ends meeting at the cut. Since ε_j belongs to $A_i \cap X_j$, the cut has the properties claimed in (ii) of (3.14).

Finally, suppose (X_1', X_2') is another cut of X such that $A_i \cap X_j' \neq \emptyset$ for all $i \in I$ and $j = 1, 2$. Without loss of generality, assume $p_1 \in X_1'$. Let $x \in X_j'$ (where $j = 1$ or 2), and take $y \in A_i \cap X_j'$. Then $x_i \in [y, x] \subseteq X_j'$ by Lemma 1.9, so $[x, x_i] \subseteq X_j'$, for all $i \in I$, hence $L(x) \subseteq X_j'$. In particular, $L(p_1) \subseteq X_1'$, and it follows that $X_1 \subseteq X_1'$. For if $x \in X_1$ and $x \in X \setminus X_1'$, then $x \in X_2'$, so $L(x) \subseteq X_2'$ and therefore $L(x) \cap L(p_1) = \emptyset$, contradicting $x \in X_1$. If $p_2 \in X_1'$ as well, then similarly $X_2 \subseteq X_1'$, hence $X = X_1 \cup X_2 \subseteq X_1'$, a contradiction. Therefore $p_2 \in X_2'$, and similarly $X_2 \subseteq X_2'$. It follows that $X_j = X_j'$ for $j = 1$, 2, completing the proof of 3.14. □

Remark. Alperin and Bass [2:(2.29)] observe that the following are equivalent, for an ordered abelian group Λ.

(i) no closed segment $[x_0, x_1]_\Lambda$ admits an open cut;
(ii) Λ is isomorphic, as an ordered abelian group, to $\{0\}$, $\mathbb{Z}$ or $\mathbb{R}$.

For (i) clearly holds if Λ is isomorphic to $\{0\}$ or $\mathbb{Z}$, and $\mathbb{R}$ satisfies (i) by the completeness property. If (i) is true, Λ is archimedean. For if M is a non-zero, proper convex subgroup of Λ, and $x \notin M$, $x > 0$, then $\big([0,x]_\Lambda \cap M, [0,x]_\Lambda \cap (\Lambda \setminus M)\big)$ is an open cut of $[0,x]_\Lambda$, a contradiction. Now if Λ is not isomorphic to $\{0\}$ or $\mathbb{Z}$, we can assume Λ is a dense subgroup of $\mathbb{R}$, by Lemma 1.1.3. Suppose $y \in \mathbb{R} \setminus \Lambda$. We can find x_0, $x_1 \in \Lambda$ such that $x_0 < y < x_1$, and then $\big([x_0,x_1]_\Lambda \cap (-\infty, y), [x_0,x_1]_\Lambda \cap (y, \infty)\big)$ is a cut of $[x_0,x_1]_\Lambda$, which is open since Λ is dense, a contradiction, so $\Lambda = \mathbb{R}$.

It follows that Case (ii) of Theorem 3.14 cannot occur in case $\Lambda = \mathbb{Z}$ or $\Lambda = \mathbb{R}$. An example of Case (ii) is obtained by taking $\Lambda = X = \mathbb{Z} \oplus \mathbb{Z}$ with the lexicographic ordering, and letting A_i be the segment $[(0,i),(1,i)]$, for $i \in \mathbb{Z}$.

4. Lyndon Length Functions

Our immediate goal is to establish a converse of Lemma 1.6. If (X,d) is a geodesic Λ-metric space, which is 0-hyperbolic, and for all x, y, $v \in X$, $(x \cdot y)_v \in \Lambda$, then (X,d) is a Λ-tree. (We can prove this under slightly weaker hypotheses.) We then give an important construction, which shows that a 0-hyperbolic Λ-metric space can be isometrically embedded in a Λ-tree in a natural way, subject to the condition above that for all x, y, $v \in X$, $(x \cdot y)_v \in \Lambda$. We use this construction to establish an equivalence between actions of a group on Λ-trees and Λ-valued Lyndon length functions defined on the group. We shall not give the definition yet, but the prototype such length function is given by taking a free group G with basis X, and defining $L : G \to \mathbb{Z}$ by $L(g) =$ the length of the reduced word on $X^{\pm 1}$ representing g. This example will be considered in more detail in Ch. 5. We note that interesting results have been obtained by Rhodes [124], using a notion of Lyndon length function in a monoid, and this idea may be capable of further exploitation. We shall also use our construction to define a base-change functor. Given a Λ-tree X and a morphism of ordered abelian groups $h : \Lambda \to \Lambda'$, this produces an induced Λ'-tree which we denote by $\Lambda' \otimes_\Lambda X$.

We begin by proving two lemmas which show that, in a 0-hyperbolic space, segments behave as they do in the Y-proposition (Lemma 1.2) for Λ-trees.

Lemma 4.1. *Assume (X, d) is a 0-hyperbolic Λ-metric space, and let σ, τ be segments in X with endpoints v, x and v, y respectively. Let $x \cdot y$ denote $(x \cdot y)_v$. Then*

(i) *If $x' \in \sigma$, then $x' \in \tau$ if and only if $d(v, x') \leq x \cdot y$.*
(ii) *If $x \cdot y \in \Lambda$, and w is the point of σ at distance $x \cdot y$ from v (which exists by* Lemma 1.2.4(1)*), then $\sigma \cap \tau$ is a segment with endpoints v and w.*

Proof. If $d(x', v) > d(y, v)$ then $x' \notin \tau$, and $d(x', v) > x \cdot y$ by Lemma 1.2.4, so we can assume that $d(x', v) \leq d(y, v)$. Let y' be the point in τ such that $d(v, x') = d(v, y')$. Define $\alpha = x \cdot y$, $\beta = x' \cdot y$, $\gamma = x \cdot x'$, $\alpha' = x' \cdot y'$.

Since $x' \in \sigma$ and $y' \in \tau$, we have $\gamma = d(v, x') = d(v, y') = y \cdot y'$ by the remarks before Lemma 1.2.2, hence (α, β, γ) and (α', β, γ) are isosceles triples, as defined in Ch.1, §2. We have to show that $x' \in \tau$ if and only if $\alpha \geq \gamma$. By Lemma 1.2.4, $x' \cdot y \leq d(v, x')$, that is, $\beta \leq \gamma$. Also,

$$\alpha' = d(v, x') - \tfrac{1}{2} d(x', y') = \gamma - \tfrac{1}{2} d(x', y') \leq \gamma$$

and

$$x' \in \tau \iff x' = y' \iff d(x', y') = 0 \iff \alpha' = \gamma.$$

Now $\alpha' = \gamma$ if and only if $\beta = \gamma$, because (α', β, γ) is an isosceles triple and α', $\beta \leq \gamma$. Since (α, β, γ) is also an isosceles triple, $\beta = \gamma$ is equivalent to $\alpha \geq \gamma$. This proves (i), and Part (ii) of the lemma follows immediately. □

Lemma 4.2. *Once more assume (X, d) is a 0-hyperbolic Λ-metric space, and σ, τ are segments in X with endpoints v, x and v, y respectively. Let $x \cdot y$ denote $(x \cdot y)_v$, and assume that $x \cdot y \in \Lambda$. Let w be the point of σ at distance $x \cdot y$ from v (so w is an endpoint of $\sigma \cap \tau$ by* Lemma 4.1*).*

Suppose $x' \in \sigma$, $y' \in \tau$ and $d(x', v) \geq x \cdot y$, $d(y', v) \geq x \cdot y$. Then

$$d(x', y') = d(x', w) + d(y', w).$$

Proof. The conclusion is clear if $d(x',v) = x \cdot y$ (when $x' = w$) or $d(y',v) = x \cdot y$ (when $y' = w$), so we assume that $d(x',v) > x \cdot y$ and $d(y',v) > x \cdot y$. We again put

$$\alpha = x \cdot y, \ \beta = x' \cdot y, \ \gamma = x \cdot x', \ \alpha' = x' \cdot y'$$

and additionally $\gamma' = y \cdot y'$, so as in the proof of 4.1, $\gamma = d(v,x')$, and similarly $\gamma' = d(v,y')$. Thus $\alpha < \gamma$, hence $\alpha = \beta$ since (α, β, γ) is an isosceles triple. Also, $\alpha < \gamma'$, so $\beta < \gamma'$, hence $\alpha = \alpha' = \beta$ since $(\alpha', \beta, \gamma')$ is an isosceles triple. By definition of α',

$$\begin{aligned} d(x',y') &= d(v,x') + d(v,y') - 2\alpha' \\ &= d(v,x') + d(v,y') - 2\alpha. \end{aligned}$$

Since $w \in \sigma \cap \tau$, $\alpha = d(v,w) < d(v,x')$, $d(v,y')$ and σ, τ are segments, it follows that

$$\begin{aligned} & d(x',w) = d(v,x') - \alpha \\ \text{and} \quad & d(y',w) = d(v,y') - \alpha \end{aligned}$$

and the lemma follows on adding these equations. □

We can now prove our converse of Lemma 1.6.

Lemma 4.3. *Assume (X,d) is a Λ-metric space and v is a point of X with the following properties:*

(i) *for all x, y, $z \in X$, $(x \cdot y)_v \in \Lambda$;*
(ii) *(X,d) is 0-hyperbolic with respect to v;*
(iii) *For every $x \in X$, there is a segment with endpoints v, x.*

Then (X,d) is a Λ-tree.

Proof. Take x, $y \in X$ and let σ, τ be segments with endpoints v, x and v, y respectively. By Lemma 4.1 and the remarks before Lemma 1.2.2, if w is the point of $\sigma \cap \tau$ at distance $(x \cdot y)_v$ from v, then σ is the union $(\sigma \cap \tau) \cup \sigma_1$, where $\sigma_1 = \{u \in \sigma | \ d(v,u) \geq (x \cdot y)_v\}$ is a segment with endpoints w, x. Similarly τ is the union $(\sigma \cap \tau) \cup \tau_1$, where $\tau_1 = \{u \in \tau | \ d(v,u) \geq (x \cdot y)_v\}$ is a segment with endpoints w, y. By Lemma 4.2 and Lemma 1.2.2, $\sigma_1 \cup \tau_1$ is a segment with endpoints x, y. Thus (X,d) is geodesic.

Note that by Lemma 4.1, $\sigma \cap \tau$ is a segment with endpoints v, w. Also by Lemma 4.1, if $\sigma \cap \tau = \{w\}$ then $(x \cdot y)_v = 0$ and $\sigma_1 = \sigma$, $\tau_1 = \tau$, hence $\sigma \cup \tau$ is a segment. Now by Lemma 1.2.4(2) and Lemma 1.2.5, we may replace v in this argument by any other point of X, hence (X,d) satisfies the axioms for a Λ-tree. □

The next result is basic in the theory of Λ-trees.

Theorem 4.4. *Let (X, d) be a Λ-metric space and v a point of X such that*
(i) *for all x, y, $z \in X$, $(x \cdot y)_v \in \Lambda$.*
(ii) *(X, d) is 0-hyperbolic with respect to v*

Then there exist a Λ-tree (X', d') and an isometry $\phi : X \to X'$ such that, if $\psi : X \to Z$ is any isometry of X into a Λ-tree Z, there is a unique isometry $\mu : X' \to Z$ such that $\mu \circ \phi = \psi$, that is, the following diagram is commutative.

$$\begin{array}{ccc} X' & & \\ \phi \uparrow & \searrow^{\mu} & \\ X & \xrightarrow[\psi]{} & Z \end{array}$$

Proof. As usual $x \cdot y$ means $(x \cdot y)_v$ for x, $y \in X$. Let

$$Y = \{(x, m) |\ x \in X,\ m \in \Lambda \text{ and } 0 \leq m \leq d(v, x)\}.$$

Define, for (x,m), $(y,n) \in Y$, $(x,m) \sim (y,n)$ if and only if $x \cdot y \geq m = n$. This is an equivalence relation on Y (it is transitive by assumption (ii)). Let $X' = Y/\sim$, and let $\langle x, m\rangle$ denote the equivalence class of (x, m). We define the metric by $d'(\langle x,m\rangle, \langle y,n\rangle) = m + n - 2\min\{m, n, x \cdot y\}$. It follows from assumption (ii) that d' is well-defined. Note that $d'(\langle x,m\rangle, \langle x,n\rangle) = |m - n|$ and $\langle x, 0\rangle = \langle v, 0\rangle$ for all $x \in X$, so $d'(\langle x,m\rangle, \langle v,0\rangle) = m$.

Clearly d' is symmetric, and it is easy to see that $d'(\langle x,m\rangle, \langle y,n\rangle) = 0$ if and only if $\langle x,m\rangle = \langle y,n\rangle$. Also, in X',

$$(\langle x,m\rangle \cdot \langle y,n\rangle)_{\langle v,0\rangle} = \min\{m, n, x \cdot y\}$$

which is an element of Λ. If $\langle x,m\rangle$, $\langle y,n\rangle$ and $\langle z,p\rangle$ are three points of X', then $(\min\{m,n\}, \min\{n,p\}, \min\{p,m\})$ is an isosceles triple by Lemma 1.2.7(1), hence by Lemma 1.2.7(2) so is $(\min\{m, n, x{\cdot}y\}, \min\{n, p, y{\cdot}z\}, \min\{p, m, z{\cdot}x\})$. It follows that (X', d') is a 0-hyperbolic Λ-metric space (d' satisfies the triangle inequality by the remark just before Lemma 1.2.7).

If $\langle x,m\rangle \in X'$, then the mapping $\alpha : [0, m]_\Lambda \to X'$ given by $\alpha(n) = \langle x, n\rangle$ is an isometry, so $\text{Im}(\alpha)$ is a segment with endpoints $\langle v, 0\rangle$ and $\langle x, m\rangle$. It now follows from Lemma 4.3 that (X', d') is a Λ-tree. Further, the mapping $\phi : X \to X'$ defined by $\phi(x) = \langle x, d(v,x)\rangle$ is easily seen to be an isometry (using Lemma 1.2.4(1)).

Given an isometry $\psi : X \to Z$ and $x \in X$, denote by x_m the point on the segment in Z with endpoints $\psi(v)$ and $\psi(x)$, at distance m from $\psi(v)$. Then the mapping $\mu : X' \to Z$ defined by $\mu(\langle x,m\rangle) = x_m$ is an isometry by Lemma 1.2(2) and is the only one with the required properties, because isometries of Λ-trees preserve segments. □

Remark. In Theorem 4.4, X' is spanned by $\phi(X)$. This can be seen directly from the proof, or by applying the last part of the theorem with Z equal to the subtree of X' spanned by $\phi(X)$, and $\psi = \phi$. The mapping μ obtained is the identity on $\phi(X)$, so on the subtree it spans (because isometries of Λ-trees preserve segments), and is one-to-one, hence the subtree spanned by $\phi(X)$ is X'.

Corollary 4.5. *In the situation of* Theorem 4.4, *suppose* $\theta : X \to X$ *is an isometry. Then there is a unique isometry* $\mu : X' \to X'$ *such that* $\mu \circ \phi = \phi \circ \theta$.

Proof. Apply the last part of Theorem 4.4 to $\psi = \phi \circ \theta$. □

We can now introduce the idea of a Lyndon length function. They were originally introduced to carry out Nielsen-style cancellation arguments, used in free groups and free products, in an abstract setting.

Definition. Let G be a group, Λ an ordered abelian group. A mapping $L : G \to \Lambda$ is called a Lyndon length function if

(1) $L(1) = 0$
(2) For all $g \in G$, $L(g) = L(g^{-1})$
(3) For all $g, h, k \in G$, $c(g,h) \geq \min\{c(g,k), c(h,k)\}$, where $c(g,h)$ is defined to be $\frac{1}{2}(L(g) + L(h) - L(g^{-1}h))$.

It follows from (2) that $c(g,h) = c(h,g)$ for all $g, h \in G$, and it follows from (3) that $(c(g,h), c(h,k), c(k,g))$ is an isosceles triple.

We leave it as an easy exercise to verify that the axioms imply

(4) For all $g \in G$, $L(g) \geq 0$
(5) For all $g, h \in G$, $L(gh) \leq L(g) + L(h)$.

We refer to (5) as the triangle inequality. The following observation is also left as an exercise.

Remark. It follows from (2) and (5) that $0 \leq c(g,h) \leq \min\{L(g), L(h)\}$, for all $g, h \in G$.

As an example, let G be a group acting as isometries on a Λ-tree (X, d) and let $x \in X$. Then defining $L_x(g) = d(x, gx)$ gives a Lyndon length function on G. By applying the isometry g^{-1} to $[x, gx] \cup [x, hx]$, we see that $c(g,h)$ is $d(x,w) = (gx \cdot hx)_x$, where $[x, gx] \cap [x, hx] = [x, w]$ (see the observations after 1.2). It follows by Lemma 1.6 that L_x satisfies Axiom (3) above, and it clearly satisfies (1) and (2). Also, it follows that this example satisfies

(6) For all $g, h \in G$, $c(g,h) \in \Lambda$.

Using Theorem 4.4 we can prove a converse.

Theorem 4.6. *Let G be a group and $L : G \to \Lambda$ a Lyndon length function satisfying* (6). *Then there are a Λ-tree (X', d'), an action of G on X' and a point $x \in X'$ such that $L = L_x$, and with the following property. If (Z, d'') is a Λ-tree on which G acts as isometries, and $w \in Z$ is such that $L = L_w$, then there is a unique G-equivariant isometry $\mu : X' \to Z$ such that $\mu(x) = w$, whose image is the subtree of Z spanned by the orbit Gw of w.*

Proof. The mapping $G \times G \to \Lambda$, $(g, h) \mapsto L(g^{-1}h)$ is easily seen to be a Λ-pseudometric on G, and we let (X, d) be the corresponding Λ-metric space. Thus $X = G/\approx$, where $\approx$ is the equivalence relation on G given by $g \approx h$ if and only if $L(g^{-1}h) = 0$. We denote the equivalence class of g by $\langle g \rangle$, so that d is given by $d(\langle g \rangle, \langle h \rangle) = L(g^{-1}h)$. Also, G acts as isometries on X by $g\langle h \rangle = \langle gh \rangle$. Direct computation shows that

$$(\langle g \rangle \cdot \langle h \rangle)_{\langle 1 \rangle} = c(g, h),$$

so by Axioms (3) and (6), (X, d) satisfies the hypotheses of Theorem 4.4 (with $v = \langle 1 \rangle$), and we obtain a corresponding Λ-tree (X', d') and an isometry $\phi : X \to X'$. We denote the point $\langle \langle g \rangle, m \rangle$ of X' simply by $\langle g, m \rangle$, and take $x = \langle 1, 0 \rangle$ as basepoint. From the construction of X', $\phi(\langle g \rangle) = \langle g, L(g) \rangle$ for $g \in G$, and $x = \phi(v)$. It follows by Corollary 4.5 that the action of G on X induces an action of G on X', such that $g\phi(u) = \phi(gu)$ for all $u \in X$ and $g \in G$. Thus $gx = \langle g, L(g) \rangle$ for $g \in G$, and by definition of d', $d'(x, gx) = L(g)$, as required. Also, it follows that $\phi(X)$ is the orbit Gx.

If Z and w are as described in the theorem, and g $h \in G$ satisfy $g \approx h$, then $L_w(g^{-1}h) = 0$, which means $g^{-1}h$ fixes w, hence $gw = hw$. There is thus a mapping $\psi : X \to Z$ given by $\psi(\langle g \rangle) = gw$, which is clearly G-equivariant. If g, $h \in G$, then $d''(gw, hw) = d''(w, g^{-1}hw) = L(g^{-1}h) = d(\langle g \rangle, \langle h \rangle)$, so ψ is an isometry, and clearly $\psi(X) = Gw$. By 4.4 there is an isometry $\mu : X' \to Z$ with $\mu \circ \phi = \psi$, from which it follows that $\mu(x) = w$, since $x = \phi(\langle 1 \rangle)$.

If $g \in G$, the mapping $\psi_1 : X \to Z$ given by $u \mapsto \psi(gu)$ is also an isometry, and the mappings μ_1, $\mu_2 : X' \to Z$ given by $p \mapsto \mu(gp)$ and $p \mapsto g\mu(p)$ respectively, satisfy $\mu_1 \circ \phi = \mu_2 \circ \phi = \psi_1$, since ϕ and ψ are G-equivariant. By the uniqueness statement in 4.4, $\mu_1 = \mu_2$, hence μ is G-equivariant. Since X' is spanned by $\phi(X)$ (remark after 4.4), the image of μ is spanned by $\mu(\phi(X)) = \psi(X) = Gw$. If $\mu' : X' \to Z$ is another G-equivariant isometry satisfying $\mu'(x) = w$, then μ, μ' agree on the orbit $Gx = \phi(X)$, and by uniqueness in Theorem 4.4, $\mu = \mu'$. □

Our second application of Theorem 4.4 will be to introduce the base-change functor. Given a morphism of ordered abelian groups $h : \Lambda \to \Lambda'$, this associates to each Λ-tree a corresponding Λ'-tree.

Theorem 4.7. *Let $h : \Lambda \to \Lambda'$ be a morphism of ordered abelian groups, and let (X, d) be a Λ-tree. Then there exist a Λ'-tree (X', d') and a mapping $\phi' : X \to X'$ satisfying*

$$d'(\phi'(x), \phi'(y)) = h(d(x, y))$$

for all $x, y \in X$, and such that, if (Z, d'') is a Λ'-tree and $\psi : X \to Z$ is a mapping satisfying

$$d''(\psi(x), \psi(y)) = h(d(x, y))$$

for all $x, y \in X$, then there is a unique Λ'-isometry $\mu : X' \to Z$ such that $\mu \circ \phi' = \psi$. Further, X' is spanned by $\phi'(X)$.

Proof. First, we set aside the case $X = \emptyset$, when we take $X' = \emptyset$ and ϕ' to be the empty mapping. The mapping $X \times X \to \Lambda'$, $(x, y) \mapsto h(d(x, y))$ is a Λ'-pseudometric, giving a Λ'-metric space $((X/\approx), \delta)$ where the equivalence relation $\approx$ is defined by $x \approx y$ if and only if $h(d(x, y)) = 0$. We denote the equivalence class of x by $\langle x \rangle$, so $\delta(\langle x \rangle, \langle y \rangle) = h(d(x, y))$. Since (X, d) is a Λ-tree and h is a morphism, this Λ'-metric space satisfies the hypotheses of Theorem 4.4, on choosing a basepoint v, by Lemma 1.6. Thus Theorem 4.4 gives a Λ'-tree (X', d') and a Λ'-isometry $\phi : (X/\approx) \to X'$. We define $\phi'(x)$ to be $\phi(\langle x \rangle)$; then $d'(\phi'(x), \phi'(y)) = d'(\phi(\langle x \rangle), \phi(\langle y \rangle)) = \delta(\langle x \rangle, \langle y \rangle) = h(d(x, y))$, as required. If ψ is as in the statement of the theorem, then ψ induces a Λ'-isometry $\bar{\psi} : (X/\approx) \to Z$ by $\bar{\psi}(\langle x \rangle) = \psi(x)$. The theorem now follows easily from Theorem 4.4 and the remark following it. □

It follows that the Λ'-tree (X', d') in Theorem 4.7 is determined up to a unique metric isomorphism, by a standard category-theoretic argument (in particular, it does not depend on the choice of basepoint in the proof of 4.7). We denote (X', d') by $\Lambda' \otimes_\Lambda X$, and call the mapping ϕ' in Theorem 4.7 the canonical embedding of X into $\Lambda' \otimes_\Lambda X$.

Corollary 4.8. *Let $h : \Lambda \to \Lambda'$ be a morphism of ordered abelian groups, and let (X, d) be a Λ-tree. Let $\phi' : X \to \Lambda' \otimes_\Lambda X$ be the canonical embedding. If $\theta : X \to X$ is a Λ-isometry, then there is a unique Λ'-isometry $\mu : X' \to X'$ such that $\mu \circ \phi' = \phi' \circ \theta$.*

Proof. Define $\psi = \phi' \circ \theta$. Then

$$d'(\psi(x), \psi(y)) = d'(\phi'(\theta(x)), \phi'(\theta(y))) = h(d(\theta(x), \theta(y)))$$

by Theorem 4.7, and this equals $h(d(x, y))$ since θ is an isometry. The result now follows from Theorem 4.7. □

Corollary 4.9. *With the hypotheses of Cor. 4.8, suppose G is a group acting as Λ-isometries on X. Then there is an action of G on $\Lambda'\otimes_\Lambda X$ as Λ′-isometries such that, for all $x \in X$ and $g \in G$, $\phi'(gx) = g\phi'(x)$. Moreover, the Lyndon length functions are related by $L_{\phi'(x)}(g) = h(L_x(g))$.*

Proof. The existence of the action follows at once from Corollary 4.8, and the formula for the length functions follows easily. □

In particular, if (X, d) is a Λ-tree and g is an isometry of X onto X, there is an induced isometry from $\Lambda' \otimes_\Lambda X$ onto itself, which we denote by $\Lambda' \otimes_\Lambda g$.

To illustrate this construction, we consider $\Lambda'\otimes_\Lambda X$ in some specific cases. Firstly, we consider the inclusion map $h : \mathbb{Z} \to \mathbb{R}$.

Lemma 4.10. *Let (X, d) be a $\mathbb{Z}$-tree, so $X = V(\Gamma)$ for some simplicial tree Γ. Then $\mathbb{R} \otimes_\mathbb{Z} X$ is metrically isomorphic to* real(Γ).

Proof. From the construction of real(Γ) in §2, there is an isometry $p : X \to$ real(Γ). By 4.7, there is an isometry $\mu : \mathbb{R}\otimes_\mathbb{Z} X \to$ real(Γ) such that $\mu \circ \phi' = p$, where $\phi' : X \to \mathbb{R} \otimes_\mathbb{Z} X$ is the canonical embedding. Thus $p(X)$ is a subset of the image of μ, which is a subtree of real(Γ). If $z \in$ real(Γ), then again from the construction, $z \in$ real(e) for some $e \in E(\Gamma)$, and real(e) $= [p(o(e)), p(t(e))]$ (see the observations after Lemma 2.6), which is contained in the image of μ. Thus z is in this image, so μ is an isomorphism. □

Now we consider a general morphism $h : \Lambda \to \Lambda'$ of ordered abelian groups.

Lemma 4.11. *If (X, d) is a Λ-tree and Y is a subtree of X, then $\Lambda' \otimes_\Lambda Y$ is (metrically isomorphic to) the subtree of $\Lambda' \otimes_\Lambda X$ spanned by $\phi'(Y)$, where $\phi' : X \to \Lambda' \otimes_\Lambda X$ is the canonical embedding.*

Proof. There is also a canonical embedding $\theta : Y \to \Lambda'\otimes_\Lambda Y$ given by Theorem 4.7. Let ψ be the composite of ϕ' and the inclusion map $Y \hookrightarrow X$. By 4.7, there is an isometry $\mu : \Lambda' \otimes_\Lambda Y \to \Lambda' \otimes_\Lambda X$ such that $\mu \circ \theta = \psi$. By 4.7, $\Lambda' \otimes_\Lambda Y$ is spanned by $\theta(Y)$, so $\mu(\Lambda' \otimes_\Lambda Y)$ is spanned by $\psi(Y) = \phi'(Y)$. □

Lemma 4.12. *If $h : \Lambda \to \Lambda'$ is a morphism of ordered abelian groups, then $\Lambda' \otimes_\Lambda \Lambda$ can be identified with the subtree of Λ′ spanned by $h(\Lambda)$, and if a, $b \in \Lambda$, then $\Lambda' \otimes_\Lambda [a, b]_\Lambda = [h(a), h(b)]_{\Lambda'}$.*

Proof. The first part follows by taking $X = \Lambda$, $\psi = h$ and $Z = \Lambda'$ in 4.7. Taking $X = [a, b]_\Lambda$ instead, and noting that $\Lambda' \otimes_\Lambda [a, b]_\Lambda$ is spanned by $\{\phi'(a), \phi'(b)\}$, the last part follows in a similar manner.

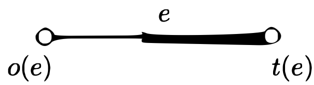
□

One final example is the barycentric subdivision of a simplicial tree. If Γ is any graph, its barycentric subdivision Γ′ can be described geometrically as the graph obtained by adding a vertex in the middle of each (unoriented) edge.

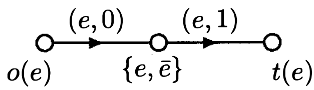
With our combinatorial definition of a graph, this can be formalised as follows. Let $\tilde{E}(\Gamma)$ denote the set of unoriented edges of Γ, and let $V(\Gamma') = V(\Gamma) \cup \tilde{E}(\Gamma)$, $E(\Gamma') = E(\Gamma) \times \{0,1\}$. We set

$$\begin{aligned} o(e,0) &= o(e) \\ t(e,0) &= \{e,\bar{e}\} = o(e,1) \\ t(e,1) &= t(e) \\ \overline{(e,0)} &= (\bar{e},1), \ \overline{(e,1)} = (\bar{e},0). \end{aligned}$$

Thus the edge e illustrated by



is divided into two edges, as shown below.



Assume Γ is connected, let $X = V(\Gamma)$ with the path metric and let $X' = V(\Gamma')$. We view the edges of Γ' as having length $\frac{1}{2}$, that is, we take as metric on X' half the path metric. We leave it to readers to verify that the resulting $\frac{1}{2}\mathbb{Z}$-metric space is isomorphic to the subspace $p(V(\Gamma)) \cup \{\langle e, \frac{1}{2}\rangle \mid e \in E(\Gamma)\}$ of real(Γ). We also leave it as an exercise (using 4.7) to show that, if Γ is a simplicial tree, then X' is isomorphic to $\frac{1}{2}\mathbb{Z} \otimes_{\mathbb{Z}} X$.

For the moment, our final application of Theorem 4.4 is yet another characterisation of $\mathbb{R}$-trees.

Lemma 4.13. *An $\mathbb{R}$-metric space (X,d) is an $\mathbb{R}$-tree if and only if*
(1) *it is connected*
(2) *it is* 0*-hyperbolic.*

Proof. An $\mathbb{R}$-tree is geodesic, so path connected and hence connected, and is 0-hyperbolic by Lemma 1.6. Conversely, assume (1) and (2). By Theorem 4.4 there is an embedding of (X,d) in an $\mathbb{R}$-tree (X',d'). Let x, $y \in X$, suppose $v \in X' \setminus X$ and $v \in [x,y]$. Then $[v,x]$, $[v,y]$ define different directions at v, so by Lemma 2.2, x, y are in different components of $X \setminus \{v\}$. Let C be the component of $X \setminus \{v\}$ containing x. By Lemma 2.2, C is open and closed, so $X \cap C$ is open and closed in X. Since $x \in X \cap C$, $y \notin X \cap C$, this contradicts the connectedness of X. Thus $[x,y] \subseteq X$, and (X,d) is geodesic. It follows that (X,d) is an $\mathbb{R}$-tree by Lemma 4.3. □

Now, as promised in §2, we prove that the completion of an $\mathbb{R}$-tree is an $\mathbb{R}$-tree. The result is due to Imrich [79].

Theorem 4.14. *If* (X, d) *is an* $\mathbb{R}$*-tree, then its metric completion* $(\hat{X}, \hat{d})$ *is also an* $\mathbb{R}$*-tree.*

Proof. By 4.13, (X, d) is connected and 0-hyperbolic, so $(\hat{X}, \hat{d})$ is 0-hyperbolic by Lemma 2.11, and is connected since X is dense in $\hat{X}$. (We are using the fact from elementary point-set topology that the closure of a connected subspace is connected.) By 4.13, $(\hat{X}, \hat{d})$ is an $\mathbb{R}$-tree. □

CHAPTER 3. Isometries of Λ-trees

1. Theory of a Single Isometry

We begin by considering the behaviour of a single isometry of a Λ-tree onto itself. There is a classification of such isometries similar to that for hyperbolic space (and indeed there is such a classification for any proper geodesic hyperbolic $\mathbb{R}$-metric space in the sense of Gromov-see [41] or [61]). There are three possible kinds of isometries in a Λ-tree, elliptic, hyperbolic and inversions. Inversions can be viewed as pathological isometries which do not occur in the case $\Lambda = \mathbb{R}$, or indeed whenever $\Lambda = 2\Lambda$, and in contrast with hyperbolic space, there are no parabolic isometries. From now on, we tacitly assume that all Λ-trees and subtrees under consideration are non-empty, even when not explicitly stated.

Let (X, d) be a Λ-tree, where Λ is an arbitrary ordered abelian group, and let g be an isometry of X onto X. We call g *elliptic* if it has a fixed point. The behaviour of such isometries is described in the next lemma, where $\langle g \rangle$ denotes the cyclic subgroup generated by g (in the group of metric automorphisms of X). If $[x, y]$ is a segment, there is a unique point $p \in [x, y]$ such that $d(x, p) = d(p, y)$ if and only if $d(x, y) \in 2\Lambda$. This point, when it exists, is called the midpoint of $[x, y]$.

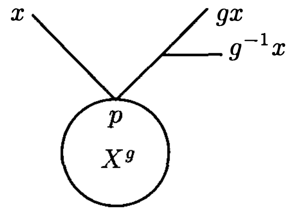
Lemma 1.1. *Let g be an elliptic isometry and let X^g denote the set of fixed points of g. Then X^g is a closed non-empty $\langle g \rangle$-invariant subtree of X. If $x \in X$ and $[x, p]$ is the bridge between x and X^g, then for any $a \in X^g$, $p = \mathrm{Y}(x, a, gx)$, and p is the midpoint of $[x, gx]$. Further, $d(x, gx) = 2d(x, p) = 2d(x, X^g)$, $[x, gx] \cap X^g = \{p\}$ and $[x, p] \subseteq [x, gx] \cap [x, g^{-1}x]$.*

Proof. If g fixes x and y it fixes the segment $[x, y]$, being an isometry, so X^g is a subtree of X, and is plainly non-empty and $\langle g \rangle$-invariant. Let $a \in X^g$ and $x \in X$. If $y \in [a, x]$, then y is the unique point of $[a, x]$ at distance $d(a, y)$ from a, so gy is the unique point of $[a, gx]$ at distance $d(a, y)$ from a, hence $y \in [a, x] \cap [a, gx]$ if and only if $y = gy$, that is, $[a, x] \cap [a, gx] = [a, x] \cap X^g$. By 2.1.2 and the remarks following it, $[a, x] \cap X^g = [a, p]$, where $p = \mathrm{Y}(x, a, gx)$, $[x, gx] = [x, p, gx]$ and $[a, x] = [a, p, x]$. It follows that $[p, x] \cap X^g = \{p\}$ and applying g, $[p, gx] \cap X^g = \{p\}$, hence $[x, gx] \cap X^g = \{p\}$, and since

$d(x,p) = d(gx,gp) = d(gx,p)$, p is the midpoint of $[x,gx]$, so $d(x,gx) = 2d(x,p)$. Replacing g by g^{-1} does not change X^g, so $[x,p] \subseteq [x,gx] \cap [x,g^{-1}x]$. Further, by 2.1.9, $[x,p]$ is the bridge between X^g and x, provided we show that X^g is closed, and this will prove the lemma.

Suppose x, $y \in X$ and $[x,y] \cap X^g \neq \emptyset$. Take $a \in [x,y] \cap X^g$. Let $p = \mathrm{Y}(x,a,gx)$, $q = \mathrm{Y}(y,a,gy)$. By what has already been proved, $[x,y] = [x,p,a,q,y]$, $[x,p] \cap X^g = \{p\}$ and $[y,q] \cap X^g = \{q\}$, hence $[x,y] \cap X^g = [p,q]$. This shows X^g is closed. □

The situation of Lemma 1.1 is illustrated by the following picture.



We next consider the idea of an isometry which is an inversion.

Lemma 1.2. *The following are equivalent:*

(i) *There is a segment of X invariant under g, and the restriction of g to this segment has no fixed points.*
(ii) *There is a segment $[x,y]$ in X such that $gx = y$, $gy = x$ and $d(x,y) \notin 2\Lambda$.*
(iii) *g^2 has a fixed point, but g does not.*
(iv) *g^2 has a fixed point, and for all $x \in X$, $d(x,gx) \notin 2\Lambda$.*

Proof. (i)⇒(ii): suppose $g[x,y] = [x,y]$, where x, $y \in X$ and g has no fixed point in $[x,y]$. Then g must interchange the endpoints x and y, and $[x,y]$ has no midpoint (otherwise it would be fixed by g), so $d(x,y) \notin 2\Lambda$.
(ii)⇒(iii): assuming (ii), it follows from Lemma 1.1 that g has no fixed points (otherwise $d(x,y) = d(x,gx) = 2d(x,X^g)$), but g^2 fixes x and y.
(iii)⇒(iv): Let p be a fixed point of g^2. Then g^2 fixes the endpoints of $[p,gp]$, so fixes this segment pointwise (Lemma 1.2.1). Take $x \in X$, and let $[x,q]$ be the bridge between x and $[p,gp]$. Then $[gx,gq]$ is the bridge between gx and $[p,gp]$, and $gq \neq q$ by assumption. By Lemma 2.1.5, $[x,gx] = [x,q,gq,gx]$, and by Lemma 2.1.4

$$\begin{aligned} d(x,gx) &= d(x,q) + d(q,gq) + d(gq,gx) \\ &= 2d(x,q) + d(q,gq). \end{aligned}$$

Also, g^2 fixes q, and so the segment $[q,gq]$ is invariant under g. Hence $d(q,gq) \notin 2\Lambda$, otherwise g fixes the midpoint of $[q,gq]$, so $d(x,gx) \notin 2\Lambda$.

(iv)$\Rightarrow$(i): let p be a fixed point of g^2. Then $[p, gp]$ is invariant under g, and g has no fixed points at all, by Lemma 1.1. $\square$

Definition. If g satisfies the equivalent conditions of 1.2, it is called an inversion.

Examples of inversions are provided by certain reflections, e.g. take $X = \Lambda = \mathbb{Z}$ and $g : x \mapsto 1 - x$. In fact, if Γ is a simplicial tree and g is a graph automorphism of Γ, then g acts as an isometry of the corresponding $\mathbb{Z}$-tree, and is an inversion if and only if there is an edge e of Γ such that $ge = \bar{e}$. Actions of groups on simplicial trees with inversions cause problems because one cannot then form the quotient graph for the action. However, this problem can be fixed on replacing Γ by its barycentric subdivision (see the discussion after Lemma 2.4.12). This generalises to Λ-trees, in that if a group acts on a Λ-tree X, then the induced action on $\frac{1}{2}\Lambda \otimes_\Lambda X$ given by Cor. 2.4.9 is without inversions. This follows from the next lemma, giving another criterion for g to be an inversion, which unlike those in 1.2, is not intrinsic to (X, d).

Lemma 1.3. *The isometry g is an inversion if and only if: g has no fixed point, and whenever Λ' is an ordered abelian group containing Λ such that $d(x, gx) \in 2\Lambda'$ for some $x \in X$, then $d(x, gx) \in 2\Lambda'$ for all $x \in X$, and the automorphism $\Lambda' \otimes_\Lambda g$ of $\Lambda' \otimes_\Lambda X$ has a unique fixed point.*

Proof. Suppose $\Lambda \subseteq \Lambda'$. Write g' for $\Lambda' \otimes_\Lambda g$, X' for $\Lambda' \otimes_\Lambda X$; we identify $x \in X$ with its image $\phi(x)$ under the canonical embedding $\phi : X \to X'$, so g' is an extension of g to X'. We denote the segment joining u, v in X' by $[u, v]'$, to avoid ambiguity when u, $v \in X$. However, we use d for the metric both in X and X'.

Suppose g is an inversion; then it has no fixed point using (1.2)(iii). Suppose $x \in X$ and $d(x, gx) \in 2\Lambda'$. Let $[x,y]$ be the bridge between x and X^{g^2}, so $g^2 y = y$ and $gy \in X^{g^2}$, hence $[y, gy] \subseteq X^{g^2}$, so $[x, y] \cap [y, gy] = \{y\}$. Applying g gives $[gx, gy] \cap [y, gy] = \{gy\}$, hence $[x, gx] = [x, y, gy, gx]$ by Lemma 2.1.5 ($y \neq gy$ since g has no fixed points). By 2.1.4, $d(x, gx) = 2d(x, y) + d(y, gy)$, hence $d(y, gy) \in 2\Lambda'$. Thus the segment $[y, gy]'$ has a midpoint, say m, which is also the midpoint of $[x, gx]'$; since g interchanges y and gy, $g'm = m$. Thus g' is elliptic, and it follows from Lemma 1.1 that $d(z, g'z) \in 2\Lambda'$ for all $z \in X'$, so $d(z, gz) \in 2\Lambda'$ for all $z \in X$. We need to show that m is independent of the choice of $x \in X$, and by what has already been said, it is enough to show that if y, $z \in X^{g^2}$, then the segments $[y, gy]'$, $[z, gz]'$ have a common midpoint. Now $[z, gz] \cap [y, gy]$ is non-empty, otherwise the bridge between $[z, gz]$ and $[y, gy]$ would be invariant under g (since each of $[y, gy]$, $[z, gz]$ is invariant under g) and so its endpoints would be fixed by g, contrary to assumption. By 2.1.7, $[z, gz] \cap [y, gy] = [p, q]$, where $p = \mathrm{Y}(z, gz, y)$ and $q = \mathrm{Y}(z, gz, gy)$, so $gp = q$ and $gq = p$. It follows by Lemma 1.1 applied to g' that $[p, q]'$ has a midpoint, which coincides with the midpoint of both $[z, gz]'$ and $[y, gy]'$, since $d(y,p) = d(gy,q)$ and $d(z,p) = d(gz,q)$. Thus the point m is independent of the choice of x, as claimed.

Now suppose $r \in (X')^{g'}$. Then $r \in [u,v]'$ for some u, $v \in X$ since X spans X' (Theorem 2.4.7). By Lemma 1.1, $[u,m]'$ is the bridge between u and $(X')^{g'}$, so $m \in [u,r]'$ and similarly $m \in [v,r]'$. Since $r \in [u,v]'$, $[u,r]' \cap [v,r]' = \{r\}$, hence $r = m$. Thus m is the only fixed point of g'.

Conversely, assume that, g has no fixed point, and whenever Λ' is an ordered abelian group containing Λ such that $d(x, gx) \in 2\Lambda'$ for some $x \in X$, then $d(x, gx) \in 2\Lambda'$ for all $x \in X$, and the automorphism $\Lambda' \otimes_\Lambda g$ of $\Lambda' \otimes_\Lambda X$ has a unique fixed point. Apply this with $\Lambda' = \frac{1}{2}\Lambda$. Take $x \in X$, and let $y = Y(x, gx, g^2x)$. By 1.1, the fixed point m of g' belongs to $[x, gx]'$ and to $[x, g^{-1}x]'$, so to $[gx, g^2x]'$. It follows from Lemma 2.1.2 that $m \in [gx, y]'$, hence $[x, gx]' = [x, y, m, gx]'$ and $[g^2x, gx]' = [g^2x, y, m, gx]'$. Thus $y \in [x, m]' \cap [g^2x, m]'$, so y, g^2y both belong to $[m, g^2x]'$. Since $d(m, y) = d(m, g^2y)$, $y = g^2y$. Also, $y \in X$, so by 1.2(iii), g is an inversion. □

Definition. An isometry of a Λ-tree onto itself is *hyperbolic* if it is not an inversion and not elliptic

By 1.2(iii), an isometry g is hyperbolic if and only if g^2 has no fixed point.

Definition. Suppose g is an isometry of a Λ-tree (X, d) onto itself. The characteristic set of g is the subset A_g of X defined by

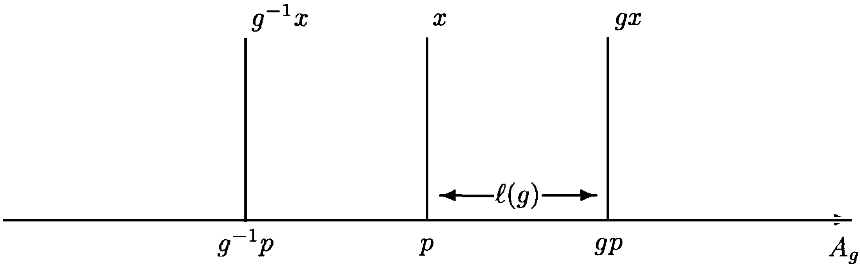
$$A_g = \{p \in X \mid [g^{-1}p, p] \cap [p, gp] = \{p\}\ \}.$$

Thus using the remark preceding Lemma 2.1.4, $A_g = \{p \in X \mid p \in [g^{-1}p, gp]\}$. If g is elliptic, it follows from Lemma 1.1 that $A_g = X^g$. If g is an inversion, then A_g is empty. For suppose $p \in A_g$, g has no fixed point but g^2 does. It follows from Lemma 1.1 applied to g^2 that the midpoint p of $[g^{-1}p, gp]$ is fixed by g^2, equivalently $g^{-1}p = gp$, and since $p \in [g^{-1}p, gp]$, $p = gp$, a contradiction.

If g is hyperbolic, the structure of A_g is described in the next theorem. The proof is lengthy, and is one of the few occasions when the theory of Λ-trees for general Λ is more difficult than for ℝ-trees. The result in this generality is due to Morgan and Shalen [106]. It was proved for $\Lambda = \mathbb{Z}$ by Tits [142] and for $\Lambda = \mathbb{R}$ by Imrich [79]. We shall give the proof in [2].

Theorem 1.4. *Suppose g is hyperbolic. Then A_g is a non-empty closed $\langle g \rangle$-invariant subtree of X. Further, A_g is a linear tree, and g restricted to A_g is equivalent to a translation $a \mapsto a + \ell(g)$ for some $\ell(g) \in \Lambda$ with $\ell(g) > 0$. If $x \in X$ and $[x,p]$ is the bridge between x and A_g, then $p = Y(g^{-1}x, x, gx)$, $[x, gx] \cap A_g = [p, gp]$, $[x, gx] = [x, p, gp, gx]$ and $d(x, gx) = \ell(g) + 2d(x,p)$.*

Proof. The situation we are trying to establish is illustrated by the following picture, where the arrow on A_g indicates g translates from left to right. This illustrates the formula given for $d(x, gx)$.



We first show that A_g is non-empty. Take any $x \in X$, and define $p = \mathrm{Y}(g^{-1}x, x, gx)$. Then $p \in [x, g^{-1}x]$, so $gp \in [x, gx]$, hence either $[x, gx] = [x, p, gp, gx]$ or $[x, gx] = [x, gp, p, gx]$. However, in the second case, $p \in [gp, gx]$ and $gp \in [x, p]$, which implies p, $g^2p \in [gp, gx]$, and $d(gp, p) = d(g^2p, gp)$, hence $g^2p = p$, contradicting that g is hyperbolic. Thus $[x, gx] = [x, p, gp, gx]$, and applying g^{-1}, $[x, g^{-1}x] = [x, p, g^{-1}p, g^{-1}x]$. By 2.1.5, we can splice the piecewise segments $[p, g^{-1}p, g^{-1}x]$ and $[p, g^{-1}p, g^{-1}x]$ to obtain the piecewise segment $[gx, gp, p, g^{-1}p, g^{-1}x]$, whence $[gp, p] \cap [p, g^{-1}p] = \{p\}$, that is, $p \in A_g$.

Now for $p \in A_g$, define $A_p = \bigcup_{n \in \mathbb{Z}} [g^n p, g^{n+1} p]$. If $m > 0$, then by Lemma 2.1.5, $[g^n p, g^{n+m} p] = [g^n p, g^{n+1} p, \ldots, g^{m+n} p]$, hence A_p is a subtree of X, and any three points are collinear, so A_p is a linear subtree by Lemma 2.3.2, and is obviously $\langle g \rangle$-invariant. Also, g is acting as a translation on A_p, of amplitude $d(p, gp)$ (otherwise, g acts as a reflection by Lemma 1.2.1, so g^2 fixes A_p pointwise, contradicting that g is hyperbolic). It follows that every point q of A_p satisfies the condition $[q, g^{-1}q] \cap [q, gq] = \{q\}$, i.e. $A_p \subseteq A_g$.

We pause from the proof to remark that, if $\Lambda = \mathbb{R}$, (or more generally, if Λ is archimedean), then $A_p = A_g$ and the theorem follows easily (see [42], 1.3). However, this is not necessarily the case if Λ is non-archimedean. For example, take $\Lambda = X = \mathbb{Z} \oplus \mathbb{Z}$ with the lexicographic ordering, and let g be the translation $x \mapsto x + (0, 1)$. Then $A_g = X$, but if $p_i = (i, 0)$, then $A_{p_i} = \{i\} \times \mathbb{Z}$, and $A_{p_i} \cap A_{p_j} = \emptyset$ if $i \neq j$. Also, $A_g = \Lambda$ is linearly ordered, and if $i < j$ then all elements of A_{p_i} are less than all elements of A_{p_j}. The strategy of the proof is to show that this reflects the general situation; if p, $q \in A_g$, then A_p, A_q are either disjoint or identical, and one can define a linear ordering on A_g, so that if $A_p \cap A_q = \emptyset$, either all elements of A_p are less than all elements of A_q, or vice-versa.

Returning to the proof, if $p \in A_g$, define $A_p^+ = \bigcup_{n \geq 0} [p, g^n p]$ and $A_p^- = \bigcup_{n \leq 0} [p, g^n p]$, so A_p^+, A_p^- are subtrees of A_p, $A_p = A_p^+ \cup A_p^-$, and $A_p^+ \cap A_p^- = \{p\}$.

Remark. If J is a subtree of A_p containing p, then $gJ \subseteq J$ implies $A_p^+ \subseteq J$, and $J \subseteq gJ$ implies $g^{-1}J \subseteq J$, hence $A_p^- \subseteq J$.

The proof will continue in four steps.

(1) (i) A_p has no proper non-empty g-invariant subtree;
(ii) If p, $q \in A_g$, then $A_p \cap A_q \neq \emptyset$ implies $A_p = A_q$.

Proof. If J is a non-empty $\langle g\rangle$-invariant subtree then $A_p^+ \subseteq J$ and $A_p^- \subseteq J$ by the remark. For if $x \in J$, then $x \in [g^m p, g^{m+1}p]$ for some m, hence $g^{-1}x \in [g^{m-1}p, g^m p]$, which implies $g^m p \in [g^{-1}x, x] \subseteq J$, since A_p is linear. Since J is g-invariant, $p \in J$ and the remark applies. Thus $J = A_p$.

To prove (ii), put $J = A_p \cap A_q$. Then J is $\langle g\rangle$-invariant, since A_p, A_q are, and a subtree of A_p and A_q. If $J \neq \emptyset$, then by (i), $A_p = J = A_q$. □

(2) Let p, $q \in A_g$, with $A_p \cap A_q = \emptyset$. Then either

(i) $[p, gq] = [p, gp, q, gq]$, $p \notin [gp, q]$, $q \in [p, gq]$ and $[p,q] \cap A_p = A_p^+$, $[p,q] \cap A_q = A_q^-$.

or

(ii) $[gp, q] = [gp, p, gq, q]$, $q \notin [gq, p]$, $p \in [gp, q]$ and $[p,q] \cap A_p = A_p^-$, $[p,q] \cap A_q = A_q^+$.

Proof. Let $[p_1, q_1]$ be the bridge between $[p, gp]$ and $[q, gq]$, with $p_1 \in [p, gp]$, $q_1 \in [q, gq]$. We claim that p_1 is either p or gp. For suppose not. Then we claim that $A_p \cap [p_1, q_1] = \{p_1\}$. For if not, we can find r such that $r \in [p_1, q_1] \cap A_p$ and $r \notin [p, gp]$. Since A_p is linear, either p or gp belongs to $[p_1, r]$. Thus either p or gp belongs to $[p_1, q_1] \cap [p, gp] = \{p_1\}$, contrary to assumption. Hence $A_p \cap [p_1, q_1] = \{p_1\}$, and so $A_p \cap [gp_1, gq_1] = \{gp_1\}$. Now $[p_1, q_1] \subseteq [gp_1, gq_1]$ by Lemma 2.1.9, hence $gp_1 = p_1$, contradicting the assumption that g is hyperbolic. This establishes the claim that $p_1 = p$ or gp, and similarly $q_1 = q$ or gq.

Suppose $p_1 = gp$. If $q_1 = gq$ then by 2.1.9, $[p,q] = [p, gp, gq, q]$, which implies $d(p,q) > d(gp, gq)$, a contradiction. Thus $q_1 = q$, and

$$[p, gq] = [p, gp, q, gq] \tag{a}$$

hence $p \notin [gp, q]$ and $q \in [gq, p]$.

Suppose $p_1 = p$. If $q_1 = q$ then by 2.1.9, $[gp, gq] = [gp, p, q, gq]$, which implies $d(p,q) < d(gp, gq)$, a contradiction. Thus $q_1 = gq$, and

$$[gp, q] = [gp, p, gq, q] \tag{b}$$

hence $q \notin [gq, p]$ and $p \in [gp, q]$.

We show that Equation (a) implies the rest of (i); a symmetrical argument shows that Equation (b) implies the rest of (ii). Assume (a); then $[p,q] = J_p \cup J' \cup J_q$, where $J_p = [p,q] \cap A_p$, $J_q = [p,q] \cap A_q$ and $J' = [p,q] \setminus (A_p \cup A_q)$. By (1), $[gp, gq] = J_p' \cup J' \cup J_q'$ (disjoint union), where $J_p' \cup [p, gp] = J_p$ and $J_q' = J_q \cup [q, gq]$. Also, $[gp, gq] = gJ_p \cup gJ' \cup gJ_q$, and $gJ_p \subseteq A_p$, $gJ_q \subseteq A_q$,

$gJ' \cap (A_p \cup A_q) = \emptyset$. It follows that $gJ_p = J'_p$, $gJ_q = J'_q$ and $gJ' = J'$. By the remark preceding Step (1), $A_p^+ \subseteq J_p$, $A_p^- \subseteq J_q$. Since $[p,q] = [p,gp,q]$ and $[p,gp] \subseteq A_p^+$, we have $[p,q] \cap A_p^- = \{p\}$ by Exercise (2) after Lemma 2.1.11, so $J_p \subseteq A_p^+$, hence $J_p = A_p^+$. Similarly, since $[p,gq] = [p,q,gq]$, it follows that $J_q = A_p^-$. □

For p, $q \in A_g$, define $p < q$ to mean that $p \notin [gp,q]$. If $A_p = A_q$, then this is the order on the linear tree A_p induced by the order on Λ for which g translates in the positive direction. Otherwise by (1), $A_p \cap A_q = \emptyset$ and by (2), it follows that either $p < q$ or $q < p$.

(3) If $A_p \cap A_q = \emptyset$ and $p < q$, then $p' < q'$ for all $p' \in A_p$, $q' \in A_q$. Any three points of $A_p \cup A_q$ are collinear.

Proof. By (2), $[p,gq] = [p,gp,q,gq]$ and $[p,q] \cap A_p = A_p^+$. If $p' \in A_p^+$ then it is easily seen that $[p,gp'] = [p,p',gp'] \subseteq A_p^+ \subseteq [p,q]$, hence $[p',q] = [p',gp',q]$ and $p' < q$ since $p' \neq gp'$. If $p' \in A_p^-$, then similarly $[p',gp] = [p',gp',gp]$, and since $[p',p] \subseteq A_p$, $[p'p] \cap [p,q] = [p'p] \cap A_p \cap [p,q] = [p',p] \cap A_p^+ = \{p\}$, since $[p,p'] \subseteq A_p^-$. Thus $[p',q] = [p',p,q] = [p',p,gp,q] = [p',gp,q] = [p',gp',gp,q] = [p',gp',q]$ and again $p' < q$. Similarly, $p < q'$ for $q' \in A_q$, and by (1) $A_q = A_{q'}$, so applying what has been proved with q' in place of q, $p' < q'$ for all $p' \in A_p$, $q' \in A_q$. □

(4) Any three points of A_g are collinear.

Proof. Let p, q, $r \in A_g$. By (3) we need only consider the case that $A_p \cap A_q$, $A_q \cap A_r$, $A_r \cap A_p$ are all empty. Without loss of generality we may assume that $p < q < r$. By (2), $A_q^- \subseteq [p,q]$ and $A_q^+ \subseteq [q,r]$, so $[p,q] = [p,g^{-1}q,q]$ and $[q,r] = [q,gq,r]$. Splicing by Lemma 2.1.5, $[p,r] = [p,g^{-1}q,q,gq,r]$, so p, q, r are collinear. □

We can now finish the proof of Theorem 1.4. If p, $q \in A_g$ and $A_p \cap A_q = \emptyset$, and $J' = [p,q] \setminus (A_p \cup A_q)$, we showed in proving (2) that $gJ' = J'$, so if $r \in J'$, then gr, $g^{-1}r \in J' \subseteq [p,q]$, hence r, gr, $g^{-1}r$ are collinear. Thus $[g^{-1}r,gr] = [g^{-1}r,r,gr]$, for the other possibilities imply $g^2r = r$, since $d(g^2r,gr) = d(gr,r)$, contradicting the assumption that g is hyperbolic. Hence $J' \subseteq A_g$, and it follows that $[p,q] \subseteq A_g$. This shows that A_g is a subtree of X, and it is linear by (4) and Lemma 2.3.2. Further, g acts as a translation on A_p for every $p \in A_g$, so g^2 has no fixed points. By Lemma 1.2.1, g acts as a translation on A_g, of amplitude $\ell(g)$, say, and $\ell(g) = d(p,gp)$ for all $p \in A_g$.

Let $x \in X$ and put $p_x = \Upsilon(g^{-1}x,x,gx)$. Recall from the beginning of the proof that $p_x \in A_g$ and $[x,gx] = [x,p_x,gp_x,gx]$. If $y \in [x,p_x]$, then $[y,gy] = [y,p_x,gp_x,gy]$, so $[y,g^{-1}y] = [y,p_x,g^{-1}p_x,g^{-1}y]$. It follows that $[y,gy] \cap [y,g^{-1}y] = [y,p_x]$, hence $y \in A_g$ if and only if $y = p_x$. Thus $[x,p_x] \cap A_g = \{p_x\}$, hence $[gx,gp_x] \cap A_g = \{gp_x\}$, so $[x,gx] \cap A_g = [p_x,gp_x]$. If $x, y \in X$ and $[x,y] \cap A_g \neq \emptyset$, take $q \in [x,y] \cap A_g$. Then since $[x,p_x] \cap A_g = \{p_x\}$, $[x,q] =$

$[x, p_x, q]$; similarly $[y, q] = [y, p_y, q]$. Thus $[x, y] = [x, q, y] = [x, p_x, q, p_y, y]$, and since also $[y, p_y] \cap A_g = \{p_y\}$, it follows that $[x, y] \cap A_g = [p_x, p_y]$. This shows A_g is a closed subtree, and $[x, p_x]$ is the bridge between x and A_g. Finally, by 2.1.4, $d(x, gx) = d(x, p) + d(p, gp) + d(gp, gx) = 2d(x, p) + \ell(g)$, where $p = p_x$. □

When g is hyperbolic, A_g is called the axis of g (which explains the notation). When g is elliptic or an inversion, we define $\ell(g)$ to be 0.

Corollary 1.5. *If g is an isometry of a Λ-tree onto itself, then g is an inversion if and only if $A_g = \emptyset$. If g is not an inversion, then $\ell(g) = \min\{d(x, gx) |\ x \in X\}$ and $A_g = \{p \in X |\ d(p, gp) = \ell(g)\}$.*

Proof. This follows from Theorem 1.4 and the observations preceding it. □

Corollary 1.5 justifies putting $\ell(g) = 0$ when g is elliptic. When g is an inversion, it is suggested by the fact that the extension of g to an isometry of $\frac{1}{2}\Lambda \otimes_\Lambda X$ has a unique fixed point, by Lemma 1.3. It is justified because it makes the formulas in 1.8 and 2.1 below correct. Note that g is hyperbolic if and only if $\ell(g) > 0$. In all cases, we call $\ell(g)$ the hyperbolic or translation length of g.

Corollary 1.6. *Suppose g is not an inversion. Then A_g meets every $\langle g \rangle$-invariant subtree of X, and A_g is contained in every $\langle g \rangle$-invariant subtree of X with the property that it meets every $\langle g \rangle$-invariant subtree of X.*

Proof. Let Y be a $\langle g \rangle$-invariant subtree of X. By 1.1 and 1.4, if $y \in Y$, and $[y, a]$ is the bridge between y and A_g, then $a \in [y, gy]$, hence $a \in Y \cap A_g$. Suppose Y meets every $\langle g \rangle$-invariant subtree of X. If g is hyperbolic, and $p \in A_g$, then the subtree A_p in the proof of 1.4 is $\langle g \rangle$-invariant, so $Y \cap A_p \neq \emptyset$, and is $\langle g \rangle$-invariant. We showed in Step (1) of the proof of 1.4 that A_p has no proper non-empty invariant subtrees, so $Y \cap A_p = A_p$, i.e. $A_p \subseteq Y$, for all $p \in A_g$. Since $p \in A_p$, $A_g \subseteq Y$. If g is elliptic, then for every $p \in A_g$, $\{p\}$ is a $\langle g \rangle$-invariant subtree of X, so $p \in Y$ and $A_g \subseteq Y$. □

We can now list some simple but important properties of A_g.

Lemma 1.7. *If g, h are both isometries of X onto X, then*
(1) *$A_{hgh^{-1}} = hA_g$ and $\ell(hgh^{-1}) = \ell(g)$;*
(2) *$A_{g^{-1}} = A_g$.*
(3) *If n is an integer then $\ell(g^n) = |n|\ell(g)$, and $A_g \subseteq A_{g^n}$. If $\ell(g) > 0$ and $n \neq 0$ then $A_g = A_{g^n}$.*
(4) *If Y is a $\langle g \rangle$-invariant subtree of X, then $\ell(g) = \ell(g|_Y)$ and $A_{g|_Y} = A_g \cap Y$.*

Proof. If g is an inversion, so is hgh^{-1} and (1) holds, otherwise (1) follows from Cor. 1.5. Also, (2) follows at once from the definition of A_g, and (3) is clear if $\ell(g) = 0$ or $n = 0$. (If g is an inversion, odd powers of g are inversions and even powers are elliptic, by Lemma 1.2(ii).) Assume $\ell(g) > 0$ and $n \neq 0$. Then g^n acts on A_g as a translation of amplitude $|n|\ell(g)$, so by definition of A_{g^n}, $A_g \subseteq A_{g^n}$. Further, by Part (1), A_{g^n} is $\langle g \rangle$-invariant, so since A_{g^n} is linear, g must induce a translation on A_{g^n} (since g^2 has no fixed points). Hence by definition of A_g, $A_{g^n} \subseteq A_g$, and (3) follows.

To prove (4), if g is not an inversion, then $Y \cap A_g \neq \emptyset$ by 1.6, so $\ell(g) = \ell(g|_Y)$ and $A_{g|_Y} = A_g \cap Y$ by 1.5. If g is an inversion, then g has no fixed point in Y, but $Y \cap A_{g^2} \neq \emptyset$ by 1.5, so g^2 has a fixed point in Y. Thus $g|_Y$ is an inversion, and (4) follows. □

Recall that the Lyndon length $L_x(g)$ of g relative to a point $x \in X$ is defined to be $d(x, gx)$. If g is not an inversion, Cor. 1.5 says that $\ell(g) = \min\{L_x(g) |\ x \in X\}$. The next result expresses ℓ in terms of L_x where x is a single point of X.

Lemma 1.8. *For any isometry g from X onto X and any $x \in X$,*

$$\begin{aligned} \ell(g) &= \max\{L_x(g^2) - L_x(g), 0\} \\ &= \max\{d(x, gx) - 2d(x,p), 0\} \end{aligned}$$

where $p = \mathrm{Y}(g^{-1}x, x, gx)$.

Proof. Suppose g is elliptic. From Lemma 1.1 we have $L_x(g) = 2d(x,a)$ where a is a point stabilised by g, so by g^2. By the triangle inequality, $L_x(g^2) \leq d(x,a) + d(a, g^2x) = d(x,a) + d(g^2a, g^2x) = 2d(x,a)$, so $\max\{L_x(g^2) - L_x(g), 0\} = 0 = \ell(g)$.

If g is an inversion, then the proof of (iii)⇒(iv) in Lemma 1.2 gives a point q stabilised by g^2 such that $L_x(g) = 2d(x,q) + d(q, gq)$. Again by the triangle inequality $L_x(g^2) \leq 2d(x,q)$ and so $\max\{L_x(g^2) - L_x(g), 0\} = 0 = \ell(g)$.

Finally, if g is hyperbolic, $[x,p]$ is the bridge between x and A_g, and $L_x(g) = \ell(g) + 2d(p, gp)$, by Theorem 1.4. By Lemma 1.7, $A_{g^2} = A_g$, $\ell(g^2) = 2\ell(g)$ and g^2 is hyperbolic, and applying 1.4 to g^2 gives $L_x(g^2) = 2\ell(g) + 2d(p, gp)$. Therefore $L_x(g^2) - L_x(g) = \ell(g) > 0$.

This establishes the first formula. For the second, note that by Lemma 2.1.2(2),

$$L_x(g^2) = d(g^{-1}x, gx) = d(x, g^{-1}x) + d(x, gx) - 2d(x,p) = 2L_x(g) - 2d(x,p),$$

and the second formula follows. □

If a group acts as isometries on a Λ-tree (X, d), recall that there is an induced action on the set of ends. This action will play a role in what follows.

If g is a hyperbolic isometry of X then there is an isometry $\alpha : A_g \to \Lambda$, and this induces an ordering on A_g which is preserved by g since g is equivalent to a translation. Denoting this ordering by $\leq$, and the corresponding strict ordering by $<$, it follows that A_g has exactly two ends, defined by the A_g-rays $\{x \in A_g \mid p \leq x\}$ and $\{x \in A_g \mid x \leq p\}$, for any $p \in A_g$. These are open ends, since if $x \in A_g$, either $x < gx$ and $g^{-1}x < x$, or $x < g^{-1}x$ and $gx < x$, depending on the choice of α, so these rays are not segments. These two ends are fixed by $\langle g \rangle$.

Lemma 1.9. *Let g be an isometry from X onto X. The open ends of X fixed by g are precisely the open ends belonging to A_g. Consequently, if g fixes an open end of X, it is not an inversion.*

Proof. Firstly g fixes the open ends of A_g; this follows from the preceding observations when g is hyperbolic, and is obvious if g is elliptic ($A_g = X^g$) or an inversion ($A_g = \emptyset$). By Lemma 2.3.8(2), g fixes the open ends of X belonging to A_g.

Suppose ε is an open end of X fixed by g. If g is elliptic, take $x \in A_g$; then g fixes all points of $[x, \varepsilon)$, since it fixes x and preserves distance from x, so $[x, \varepsilon\rangle \subseteq A_g$ and ε belongs to A_g. If g is an inversion, then by criterion (ii) of Lemma 1.2, there is a segment $[x, y]$ whose endpoints are interchanged by g. Now $[x, \varepsilon\rangle \cap [y, \varepsilon\rangle = [z, \varepsilon\rangle$ for some z, hence $[z, \varepsilon\rangle = [gz, \varepsilon\rangle$. Since ε is an open end, $[z, \varepsilon\rangle$ has only one endpoint, so $z = gz$, a contradiction since g has no fixed point. Thus g fixes no open end of X.

Finally, suppose g is hyperbolic, and again take $x \in A_g$. Suppose there exists $z \in [x, \varepsilon\rangle$ such that $z \notin A_g$. Then $[x, \varepsilon\rangle \cap A_g = [x, z] \cap A_g$ since $[x, \varepsilon\rangle$ is a linear subtree of X with x as an endpoint. Since A_g is a closed subtree of X, we obtain $[x, \varepsilon\rangle \cap A_g = [x, y]$ for some $y \in A_g$, and $[y, \varepsilon\rangle \cap A_g = \{y\}$. Hence $[gy, \varepsilon\rangle \cap A_g = \{gy\}$. If $u \in [y, \varepsilon\rangle$, $v \in [gy, \varepsilon\rangle$, then it follows that $[u, y]$ is the bridge between u and A_g, so $[u, gy] = [u, y, gy]$, and similarly $[v, y] = [v, gy, y]$. Splicing these segments (Lemma 2.1.5) gives $[u, v] = [u, y, gy, v]$, so $u \neq v$. Thus $[y, \varepsilon\rangle \cap [gy, \varepsilon\rangle = \emptyset$, a contradiction. There is therefore no such z, so $[x, \varepsilon\rangle \subseteq A_g$ and ε belongs to A_g. □

2. Group Actions as Isometries

We shall establish a classification of group actions on Λ-trees. They can be classified into three basic types, abelian, dihedral and general (or irreducible) type. The abelian actions can be further classified into four types.

Suppose G is a group acting as isometries on a Λ-tree (X, d). For $g \in G$, the characteristic set of the isometry induced on X by g is denoted by A_g. We have a function $\ell : G \to \Lambda$, where $\ell(g)$ is the hyperbolic length of the isometry induced by g, for $g \in G$. This function is known as the hyperbolic or translation length function of the action, also as the unbased length function (to distinguish it from the Lyndon length functions, which are called based length functions since they depend on a choice of basepoint). The action of G is said to be without inversions if no element of G induces an inversion on X, otherwise the action is said to be with inversions.

Lemma 2.1. *Let $h : \Lambda \to \Lambda'$ be a morphism of ordered abelian groups, and let (X,d) be a Λ-tree. Suppose G is a group acting as Λ-isometries on X, so that there is an induced action of G on $\Lambda' \otimes_\Lambda X$ (see Cor 2.4.9). Let ℓ be the hyperbolic length function for the action on X, and ℓ' the hyperbolic length function for the action on $\Lambda' \otimes_\Lambda X$. Then $\ell' = h \circ \ell$.*

Proof. This follows at once from Cor. 2.4.9 and Lemma 1.8. □

In order to obtain our classification, some results on the behaviour of pairs of isometries are needed.

Lemma 2.2. *Suppose (X,d) is a Λ-tree and g, h are isometries of X onto X which are not inversions, with $A_g \cap A_h = \emptyset$. Let $[p,q]$ be the bridge between A_g and A_h, with $p \in A_g$, $q \in A_h$. Then $[p,q] \subseteq A_{gh}$, and*

$$\ell(gh) = \ell(g) + \ell(h) + 2d(A_g, A_h).$$

Further, $A_{gh} \cap A_h = [q, h^{-1}q]$ and $A_{gh} \cap A_g = [p, gp]$, segments of length $\ell(h)$, $\ell(g)$ respectively.

Proof. (Readers may find it helpful to refer to the picture following the lemma; this only shows the case where g and h are both hyperbolic. The argument which follows applies in all cases.) By Lemma 1.1 and Theorem 1.4, applied to g^{-1},

$$[g^{-1}q, q] = [g^{-1}q, g^{-1}p, p, q] \tag{1}$$

and similarly

$$[h^{-1}p, p] = [h^{-1}p, h^{-1}q, q, p]. \tag{2}$$

Applying h^{-1} to Eq. (1) gives

$$[h^{-1}g^{-1}q, h^{-1}q] = [h^{-1}g^{-1}q, h^{-1}g^{-1}p, h^{-1}p, h^{-1}q]. \tag{3}$$

Also $h^{-1}p \neq h^{-1}q$ since $p \neq q$. By Lemma 2.1.5, we may splice the piecewise segments $[h^{-1}p, h^{-1}q, p]$ and $[h^{-1}g^{-1}q, h^{-1}p, h^{-1}q]$, obtained from (2) and (3), giving

$$[h^{-1}g^{-1}q, p] = [h^{-1}g^{-1}q, h^{-1}p, h^{-1}q, p]$$

and from (2) and (3),

$$[h^{-1}g^{-1}q, p] = [h^{-1}g^{-1}q, h^{-1}g^{-1}p, h^{-1}p, h^{-1}q, q, p]. \tag{4}$$

Applying the isometry gh to (4) gives

$$[q, ghp] = [q, p, gp, gq, ghq, ghp] \tag{5}$$

and again using Lemma 2.1.5, we splice the piecewise segments $[h^{-1}g^{-1}q, q, p]$, $[q, p, ghq]$ and use (4) and (5) to obtain

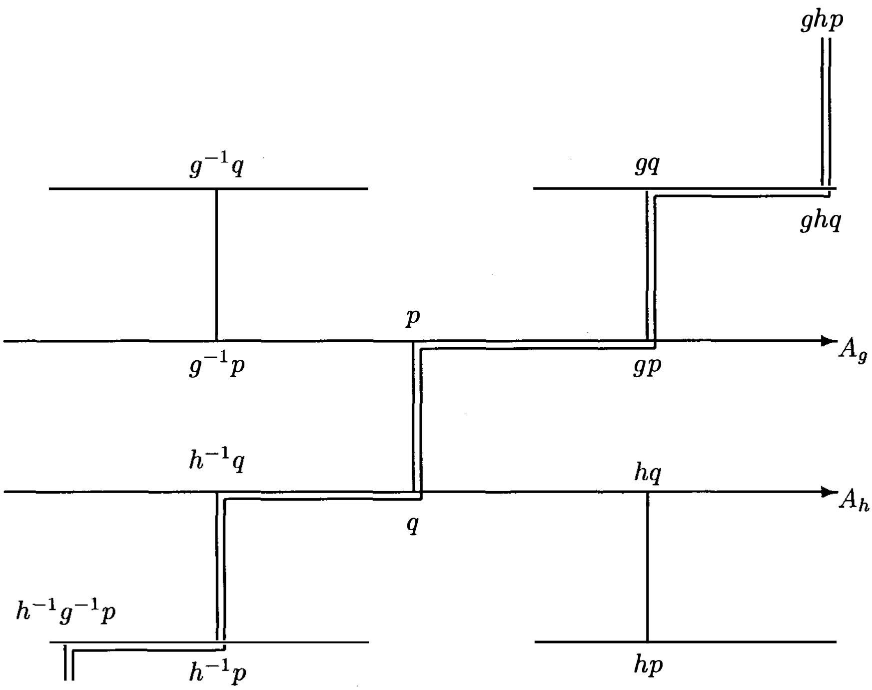
$$[h^{-1}g^{-1}q, ghq] = [h^{-1}g^{-1}q, h^{-1}g^{-1}p, h^{-1}p, h^{-1}q, q, p, gp, gq, ghq]. \tag{6}$$

In particular, $q \in [h^{-1}g^{-1}q, ghq]$, so $q \in A_{gh}$ and $\ell(gh) = d(q, ghq)$ by Cor. 1.5. By Lemma 2.1.4,

$$\begin{aligned} d(q, ghq) &= d(q, p) + d(p, gp) + d(gp, gq) + d(gq, ghq) \\ &= d(p, gp) + d(q, hq) + 2d(p, q) \\ &= \ell(g) + \ell(h) + 2d(A_g, A_h) \end{aligned}$$

since $p \in A_g$, $q \in A_h$ and $[p, q]$ is the bridge between A_g and A_h. Now $q \in A_{gh}$, so $[h^{-1}g^{-1}q, ghq] \subseteq A_{gh}$ and from Eq. (6), $[h^{-1}p, h^{-1}q, q, p] \subseteq A_{gh}$, in particular $[p, q] \subseteq A_{gh}$. Further, $[q, p] \cap A_h = \{q\}$ (by definition of p, q), hence $[h^{-1}p, h^{-1}q] \cap A_h = \{h^{-1}q\}$. Since $d(A_g, A_h) > 0$, $\ell(gh) > 0$, that is, gh is hyperbolic, hence A_{gh} is a linear subtree. It follows easily that $A_{gh} \cap A_h = [q, h^{-1}q]$. The last part of the lemma follows similarly, or by replacing g, h by h^{-1}, g^{-1} and using Lemma 1.7. The segments $[q, h^{-1}q]$ and $[p, gp]$ have the indicated lengths by Cor.1.5. □

As suggested above, the situation is greatly clarified by the following diagram, where A_{gh} is indicated by a double line. It illustrates the case that g and h are both hyperbolic.The picture is built up by starting with A_g, A_h and the bridge $[p, q]$ joining them. Translate by h to obtain the bottom right of the picture, then translate by g to obtain the top right. Similarly, translate by g^{-1} to obtain the top left, then translate by h^{-1} to obtain the bottom left part. Similar diagrams can be drawn when one or both of g and h are elliptic. One has to replace the axis by a circle or other non-descript figure to represent the fixed point set, and if, for example, g is elliptic, then p, $g^{-1}p$ and gp coincide. We leave it to the reader to draw these cases (or to look them up in §8 of [2]).



Lemma 2.3. *If g, h are isometries of a Λ-tree (X,d) and $A_g \cap A_h \neq \emptyset$, then* $\max\{\ell(gh), \ell(gh^{-1})\} \leq \ell(g) + \ell(h)$.

Proof. Choose $p \in A_g \cap A_h$. By Cor. 1.5 and the triangle inequality, $\ell(gh) \leq d(p, ghp) \leq d(p, gp) + d(gp, ghp) = d(p, gp) + d(p, hp) = \ell(g) + \ell(h)$. Since $A_{h^{-1}} = A_h$ by Lemma 1.7, the result follows. □

The following is an immediate consequence of Lemmas 2.2 and 2.3.

Corollary 2.4. *If g, h are isometries of a Λ-tree (X,d) which are not inversions, then $d(A_g, A_h) = \max\{0, \frac{1}{2}(\ell(gh) - \ell(g) - \ell(h))\}$ and A_g, A_h intersect if and only if $\ell(gh) \leq \ell(g) + \ell(h)$.*

□

Definition. An action of a group G on a Λ-tree is called *abelian* if $\ell(gh) \leq \ell(g) + \ell(h)$ for all g, $h \in G$.

Note that an action with inversions is not precluded, although Cor. 2.4 gives an interpretation of the definition in the case of an action without inversions. In this case, the action is abelian if and only if, for all g, $h \in G$, A_g and

A_h intersect. It is in studying abelian actions that the results on ends in §3 of Ch.2 come into play; we begin with a general result applicable to any action without inversions.

Proposition 2.5. *Suppose G is a group acting as isometries, without inversions, on a Λ-tree (X, d), and ε is an end of X. Let δ be an end map towards ε (see the addendum to Lemma 2.3.12) and denote the stabilizer of ε by G_ε. Then there is a homomorphism $\tau : G_\varepsilon \to \Lambda$, depending only on ε, such that, for $g \in G_\varepsilon$, $\ell(g) = |\tau(g)|$, and for all $x \in X$, $\delta(gx) = \delta(x) + \tau(g)$.*

Proof. Let $x \in X$. The mapping $[x, g^{-1}\varepsilon\rangle \to \Lambda$, $y \mapsto \delta(gy)$ is an isometry, with least value $\delta(gx)$, hence $\delta \circ g$ is an end map towards $g^{-1}\varepsilon$.

Suppose $g \in G_\varepsilon$, so $\delta \circ g$ is an end map towards ε. By the addendum after 2.3.12, $\delta \circ g = \delta + \tau(g)$ for some $\tau(g) \in \Lambda$. If also $h \in G_\varepsilon$, then

$$\begin{aligned} \delta(x) + \tau(gh) &= \delta(ghx) \\ &= \delta(hx) + \tau(g) \\ &= \delta(x) + \tau(g) + \tau(h) \end{aligned}$$

showing that τ is a homomorphism.

If $g \in G_\varepsilon$, then ε is an end belonging to A_g by Lemma 1.9, and $A_g = \bigcup_{x \in A_g} [x, \varepsilon\rangle$. If $\ell(g) = 0$, then for $y \in A_g$, $\delta(y) = \delta(gy) = \delta(y) + \tau(g)$, so $\tau(g) = 0$. If $\ell(g) > 0$ then by the discussion preceding Lemma 1.9, if x, $y \in A_g$ then either $[x, \varepsilon\rangle \subseteq [y, \varepsilon\rangle$ or $[y, \varepsilon\rangle \subseteq [x, \varepsilon\rangle$. It follows that $\delta : A_g \to \Lambda$ is an isometry. Thus, if $x \in A_g$, $|\tau(g)| = |\delta(gx) - \delta(x)| = d(gx, x) = \ell(g)$.

It follows from the addendum to 2.3.12 that τ depends only on ε, not on the choice of δ. □

We now show that abelian actions can be further classified into different types. Before stating the result, a simple observation is needed. It is unfortunate that the notation for commutators resembles that for segments, but it should not cause confusion.

Remark. Let G be a group acting on a linear Λ-tree A, with hyperbolic length function ℓ, and let g, $h \in G$. If g acts as a non-trivial translation and h as a reflection (see 1.2.1), then $\ell([g, h]) = 2\ell(g) > 0$, where $[g, h] = ghg^{-1}h^{-1}$; this follows using equations (i) and (ii) in the proof of Lemma 1.2.1. If g, h act as distinct reflections, then gh acts as a non-trivial translation, so $\ell([g, h]) = \ell((gh)^2) = 2\ell(gh) > 0$.

Theorem 2.6. *Let G be a group with an abelian action on a Λ-tree (X, d).*

(1) *If G acts with an inversion, then $\ell = 0$ and $X^G = \emptyset$. If Λ' is an ordered abelian group having Λ as a subgroup, then $(\Lambda' \otimes_\Lambda X)^G = \emptyset$ if $d(x, gx) \notin 2\Lambda'$ for some $x \in X$, $g \in G$; otherwise, $(\Lambda' \otimes_\Lambda X)^G$ consists of a single point.*

(2) *Suppose G acts without inversion. Put $B = \bigcap_{g \in G} A_g$.*
  (i) *Suppose $B \neq \emptyset$. If $\ell = 0$ then $B = X^G$. If $\ell \neq 0$ then $X^G = \emptyset$ and B is a linear G-invariant subtree of X on which G acts as a non-trivial group of translations.*
  (ii) *Suppose $B = \emptyset$ and G fixes an end ε of X. Then ε is the only end fixed by G, it is an open end, and for all $g \in G$ and all $x \in A_g$, $[x, \varepsilon\rangle \subseteq A_g$.*
  (iii) *Suppose $B = \emptyset$ and G fixes no end of X. Then there is a unique open cut (X_1, X_2) of X such that, for all $g \in G$, and $i = 1, 2$, $A_g \cap X_i \neq \emptyset$. Also, G leaves X_1, X_2 invariant, and fixes the ends ε_1, ε_2 which meet at the cut. The action of G on X_1 and X_2 is of type (ii).*

Proof. (1) Assume G contains an inversion; since an inversion has no fixed points (using the criterion 1.2(iii)), $X^G = \emptyset$. Now suppose Λ' is an ordered abelian group containing Λ. Let d' be the metric on $(\Lambda' \otimes_\Lambda X)$, and let U be the set of elements of G which act as inversions on X.

If $d(x, gx) \notin 2\Lambda'$ for some $x \in X$ and $g \in G$ with $\ell(g) = 0$, then by Lemma 1.1, g acts as an inversion on $(\Lambda' \otimes_\Lambda X)$, so again $(\Lambda' \otimes_\Lambda X)^G = \emptyset$ by 1.2(iii).

On the other hand, if $d(x, gx) \in 2\Lambda'$ for all $x \in X$ and $g \in G$ such that $\ell(g) = 0$, then by 1.3 every $g \in U$ has a unique fixed point in $(\Lambda' \otimes_\Lambda X)$, say x_g. If g, $h \in U$, and $x_g \neq x_h$, then $d'(x_g, x_h) = \ell(gh) - \ell(g) - \ell(h) > 0$ by 2.2, contradicting that the action is abelian (note that the action of G on $(\Lambda' \otimes_\Lambda X)$ has the same hyperbolic length function as the action of G on X, by 2.1). Thus $x_g = x_h$, and we denote this point by x. Then for any $g \in G$, gx is also a fixed point of all elements of U (since U is closed under conjugation), so $gx = x$ and $(\Lambda' \otimes_\Lambda X)^G = \{x\}$, hence $\ell(G) = 0$. Since we can find Λ' such that $d(x, gx) \in 2\Lambda'$ for all x and g (e.g. $\Lambda' = \frac{1}{2}\Lambda$), (1) follows.

(2) Assume G acts without inversion. Suppose $B \neq \emptyset$. Note that B is an invariant subtree of X, since $gA_h = A_{ghg^{-1}}$ (by 1.7). If $\ell(G) = 0$, then $B = \bigcap_{g \in G} X^g = X^G$. If $\ell(G) \neq 0$, then since A_g is linear for hyperbolic $g \in G$, B is linear. The action of G on B has the same hyperbolic length function as the action on X by 1.7. Some element g of G acts as a non-trivial translation on B by Theorem 1.4, and if $h \in G$ acts as a reflection on B, $\ell([g, h]) = \ell((ghg^{-1})h^{-1}) \leq \ell(ghg^{-1}) + \ell(h^{-1}) = 2\ell(h) = 0$ (reflections are either elliptic or inversions). This contradicts the remark preceding Theorem 2.6. Hence G acts as a non-trivial group of translations on B.

Suppose $B = \emptyset$, and suppose there is an end ε fixed by G. By Lemma 1.9, ε belongs to A_g, for all $g \in G$. By Theorem 2.3.14, there is only one such end, it is open and for all $x \in A_g$, $[x, \varepsilon\rangle \subseteq A_g$.

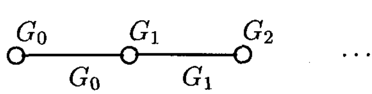
Suppose there is no end ε fixed by G. By Lemma 1.9, there is no end of X belonging to all A_g and by Cor. 2.4 and Theorem 2.3.14, there is a unique cut X_1, X_2 in X such that $A_g \cap X_i \neq \emptyset$ for all $g \in G$ and $i = 1, 2$. Since gX_1,

gX_2 is another cut with this property, for $g \in G$, G preserves the cut. Also, no element of G can interchange X_1 and X_2. For if g is such an element, then so is g^{-1}, and if $x \in X_1 \cap A_g$, then $x \in [g^{-1}x, gx]$ by definition of A_g, so $x \in X_2$, a contradiction. Thus G acts on X_i, and the action is abelian by Lemma 1.7(4). Also by Lemma 1.7(4), if $g \in G$ and g_i denotes its restriction to X_i, then $A_{g_i} = X_i \cap A_g \neq \emptyset$, so the action of G on X_i is without inversions. Further, it follows that $\bigcap_{g\in G} A_{g_i} = B \cap X_i = \emptyset$. By 2.3.14 the cut is open, and the ends ε_i of X_i which meet at the cut are fixed by G, using the characterisation of these ends in Cor. 2.3.11. Thus the action of G on X_i is of type (ii), for $i = 1$, 2. □

Remark. In Case (2) of Theorem 2.6, an action as in (i) is said to be of core type, an action as in (ii) is said to be of end type, and an action as in (iii) is called a cut type action. It follows from the remark after Theorem 2.3.14 that a cut type action cannot occur in case $\Lambda = \mathbb{Z}$ or $\Lambda = \mathbb{R}$.

The example given earlier of an inversion ($X = \Lambda = \mathbb{Z}$ with the reflection $r_1 : x \mapsto 1 - x$) gives an example of Case (1) of (2.6) with $G = \langle r_1 \rangle$ cyclic of order 2. Examples of actions of core type are given by the action of Λ on itself by translations. We shall defer most of the discussion of actions of end and cut type until §4, but for now we shall give examples of actions of end type with the hyperbolic length function ℓ identically zero.

Example. Let G be a group which is a strictly ascending union of a countable collection of subgroups, say $G = \bigcup_{i=1}^{\infty} G_i$, where $G_1 \lneqq G_2 \lneqq \cdots$. Define a graph Γ by putting $V(\Gamma) = \bigcup_{i=1}^{\infty}(G/G_i)$, where $G/G_i = \{gG_i \mid g \in G\}$, and by taking the unoriented edges to be the pairs $\{gG_i, gG_{i+1}\}$ for all $g \in G$, $i \geq 1$. We leave it as an exercise to show that Γ is a simplicial tree. The action of G on G/G_i for $i \geq 1$ gives an action of G on Γ as graph automorphisms, so as isometries of the corresponding $\mathbb{Z}$-tree (whose set of points X is $V(\Gamma)$). (For those familiar with the Bass-Serre theory, this is a special case. For G is the tree product of the tree of groups:



where the monomorphisms from the edge groups to the vertex groups are inclusion and identity maps, and Γ is the tree on which G acts given by the Bass-Serre theory.) Each element of G is in G_i for some i, so fixes the vertex G_i of Γ, hence the action is without inversions, and the hyperbolic length function is identically zero by Cor. 1.5. Further, there is no global fixed point for the action. For if $\text{stab}(gG_n) = G$, then $G_n = \text{stab}(G_n) = g^{-1}Gg = G$, which is impossible as $G_n \lneqq G_{n+1}$. By Theorem 2.6 and the remark after its proof, the action on X is of end type.

There is an induced action on the $\mathbb{R}$-tree $\mathbb{R} \otimes_{\mathbb{Z}} X$ with hyperbolic length function also identically zero, by Lemma 2.1, and any action on an $\mathbb{R}$-tree is without inversions by criterion (ii) of Lemma 1.2. In fact, by Lemma 2.4.10, $\mathbb{R} \otimes_{\mathbb{Z}} X \cong \text{real}(\Gamma)$, and $X = V(\Gamma)$ embeds in $\text{real}(\Gamma)$, so each element of G fixes a point of $\text{real}(\Gamma)$. Again, there is no global fixed point. For if $z \in \text{real}(\Gamma)$ and $gz = z$ for all $g \in G$, then $z \in \text{real}(e)$ for some edge e, by construction of $\text{real}(\Gamma)$, and as observed after the proof of 2.2.6, $\text{real}(e)$ is the segment $[o(e), t(e)]$, so $g\text{real}(e) = \text{real}(ge)$ for all $g \in G$. From what we have already shown, z is not an endpoint of e, so by Lemma 2.2.4(5), $ge{=}e$ or $\bar{e}$ for all $g \in G$. It is clear from the construction of Γ that no element of G can interchange the endpoints of an edge (in fact, this follows since the action on X is without inversions). Hence G fixes the endpoints of e, contradicting what has been proved. It again follows from Theorem 2.6 and the remark after its proof that the action on $\text{real}(X)$ is of end type.

We shall give some alternative characterisations of an abelian action. First, we need another remark.

Remark. If g is an isometry of a Λ-tree (X, d) which is not an inversion, and Y is a subtree of X such that $Y \cap A_g = \emptyset$, then $Y \cap gY = \emptyset$. For if $Y \cap gY \neq \emptyset$, $gy \in Y$ for some $y \in Y$, hence $[y, gy] \subseteq Y$, and this segment contains a point of A_g by Lemma 1.1 and Theorem 1.4.

Proposition 2.7. *Let G be a group acting on a Λ-tree (X, d). The following are equivalent.*

(1) *The action is abelian.*
(2) *The hyperbolic length function is given by $\ell(g) = |\rho(g)|$ for $g \in G$, where $\rho : G \to \Lambda$ is a homomorphism.*
(3) *For all g, $h \in G$, $\ell([g, h]) = 0$.*

Proof. Assume (1). If $\ell = 0$, clearly (2) holds, so we can assume $\ell \neq 0$. If the action is of core type, then $B = \bigcap_{g \in G} A_g$ is a linear invariant subtree on which G acts as translations, say $gx = x + \rho(g)$, for $g \in G$, $x \in B$, where ρ is a homomorphism. Restricting the action to B does not change the hyperbolic length function (by 1.7(4)), so $\ell(g) = |\rho(g)|$, for $g \in G$. If the action is of end type, then (2) follows from Prop. 2.5. If it is of cut type, then restricting the action to one of the subtrees X_i of the cut gives an action of end type, and the hyperbolic length function is unchanged by Lemma 1.7(4), and again (2) holds by Prop. 2.5.

It is obvious that (2) implies (3), so we assume (3) and show the action is abelian. By Lemma 2.1, there is an action of G on the $\frac{1}{2}\Lambda$-tree $X' = \frac{1}{2}\Lambda \otimes_{\Lambda} X$, with the same hyperbolic length function, and this action is without inversions by Lemma 1.3. We can therefore assume the action is without inversions. Let $g, h \in G$; then $\ell([g, h]) \leq \ell(ghg^{-1}) + \ell(h^{-1})$, so by 2.4, $A_{ghg^{-1}} \cap A_{h^{-1}} \neq \emptyset$, that is, using 1.6, $gA_h \cap A_h \neq \emptyset$. By the remark preceding the theorem, $A_g \cap A_h \neq \emptyset$, and by (2.4), the action is abelian. □

We come now to the second type of action.

Definition. Suppose G is a group acting on a Λ-tree (X,d) and let $\ell : G \to \Lambda$ be the hyperbolic length function. The action is called *dihedral* if it is non-abelian but $\ell(gh) \leq \ell(g)+\ell(h)$ whenever g, $h \in G$ act as hyperbolic isometries.

We shall give a characterisation of dihedral actions analogous to that just given in Prop. 2.7 for abelian actions. First we need the following fact.

Lemma 2.8. *Suppose G is a group acting without inversions on a Λ-tree (X,d), and the action is dihedral. Then there is a non-empty linear invariant subtree.*

Proof. Put $A = \bigcap\{A_g \mid g \in G,\ \ell(g) > 0\}$. By Lemma 1.7, A is G-invariant. If $A = \emptyset$, then by Cor. 2.4 and Theorem 2.3.14, G fixes an open end or an open cut of X. Therefore, G has a subgroup H of index at most two which acts on a subtree of X with the same hyperbolic length function and with a fixed end. By Prop. 2.5 and Lemma 1.7, the action of H on X is abelian. If $g \in G$, then $A_g \neq \emptyset$ since the action is without inversions. Arguing as in the proof of Theorem 2.6(2), in the case of an open cut, g cannot interchange the two subtrees of the cut. Thus $G = H$ and the action is abelian, contrary to assumption. Hence A is non-empty and a G-invariant subtree. Since the action is non-abelian, $\ell \neq 0$, so there is a hyperbolic element $g \in G$. Then $A \subseteq A_g$, and A_g is linear, hence so is A. □

In the statement of the next result, $\mathrm{Isom}(\Lambda)$ denotes the group of metric isomorphisms of Λ, a group described in Lemma 1.2.1.

Proposition 2.9. *Let G be a group acting on a Λ-tree (X,d). The following are equivalent*

(1) *The action is dihedral.*
(2) *The hyperbolic length function is given by $\ell(g) = |\rho(g)|$ for $g \in G$, where $\rho : G \to \mathrm{Isom}(\Lambda)$ is a homomorphism whose image contains a reflection and a non-trivial translation, and the absolute value signs denote hyperbolic length for the action of* $\mathrm{Isom}(\Lambda)$.
(3) *For all hyperbolic g, $h \in G$, $\ell([g,h]) = 0$, but there exist g, $h \in G$ such that $\ell([g,h]) \neq 0$.*

Proof. Assume the action is dihedral. By Lemmas 2.1 and 1.3, there is an induced action on $X' = \frac{1}{2}\Lambda \otimes_\Lambda X$ without inversions and with the same hyperbolic length function, so this action is also dihedral. By 2.8, there is a non-empty linear invariant subtree A for the action on X'. By 1.7, the action of G on A has the same hyperbolic length function as the action on X, and by Lemma 1.6, if $\ell(g) = 0$, g has a fixed point in A. Since there is some $g \in G$ with $\ell(g) \neq 0$, A has at least two elements, so by Lemma 1.2.1, the action of G on A extends to an action as isometries on $\frac{1}{2}\Lambda$. If $\ell(g) > 0$, g acts as a translation, either $x \mapsto x+\ell(g)$ or $x \mapsto x-\ell(g)$, and $\ell(g) \in \Lambda$. If $\ell(g) = 0$, then either g acts as the identity on A, or it acts as a reflection with a fixed point, say c, so $g : x \mapsto 2c - x$, and $2c \in \Lambda$. Thus the action of G on $\frac{1}{2}\Lambda$ restricts to an action on Λ, still with the same hyperbolic length function, so $\ell(g) = |\rho(g)|$ for $g \in G$, where $\rho : G \to \text{Isom}(\Lambda)$ is the homomorphism corresponding to the action on Λ. The image of ρ cannot consist entirely of translations, nor can it consist of a single reflection and the identity element, otherwise the action would be abelian by criterion (3) in Prop.2.7. It follows from the remark preceding Theorem 2.6 that (1) implies (2). It also follows from the remark preceding Theorem 2.6 that (2) implies (3).

Assume (3). Then by 2.7 the action is not abelian. Using the induced action on $X' = \frac{1}{2}\Lambda \otimes_\Lambda X$, we can assume the action is without inversions. Arguing as in the proof that (3) implies (1) in 2.7, if g and h are hyperbolic, then $\ell(gh) \leq \ell(g) + \ell(h)$. □

Thus, in the case of either an abelian or dihedral action of a group G on a Λ-tree (X, d), the hyperbolic length function is given by $\ell(g) = |\rho(g)|$ for some homomorphism $\rho : G \to \text{Isom}(\Lambda)$. In the case of an abelian action, the image of ρ is contained in the subgroup of translations, so is abelian. In the case of a dihedral action, the image of ρ is easily seen (as in Lemma 1.2.1) to be the split extension of an infinite abelian group by a cyclic group of order 2, and in case $\Lambda = \mathbb{Z}$, it is isomorphic to the infinite dihedral group. We hope this explains the terminology. In view of Lemma 2.8, we investigate the situation when an action has a linear invariant subtree.

Lemma 2.10. *Let G be a group acting on a Λ-tree (X, d) with hyperbolic length function ℓ, and suppose A is a linear G-invariant subtree.*
(a) *If $g \in G$ and $\ell(g) > 0$ then $A \subseteq A_g$.*
(b) *The action is abelian or dihedral.*
(c) *The action is abelian if and only if either $\ell(G) = 0$ or $A \subseteq A_h$ for all $h \in G$.*

Proof. (a) If $x \in A$ and $\ell(g) > 0$, then $d(x, gx) = d(x, g^{-1}x)$, and $gx \neq g^{-1}x$ since $\ell(g^2) = 2\ell(g) > 0$ by 1.7. Since A is linear, $[x, gx] \cap [x, g^{-1}x] = \{x\}$, so $x \in A_g$.

(b) If g, h are hyperbolic then $A \subseteq A_g \cap A_h$ by (a), so $A_g \cap A_h \neq \emptyset$, and (b) follows by Cor. 2.4.

(c) If $\ell(G) = 0$ the action is clearly abelian. If $A \subseteq A_h$ for all $h \in G$, then $A_g \cap A_h \neq \emptyset$ for all g, $h \in G$, so the action is abelian by Cor. 2.4. Suppose the action is abelian and $\ell(G) \neq 0$. Since $\ell(g) = \ell(g|A)$ for $g \in G$ (by 1.7), some element of G acts as a translation on A, and by 2.7(3) and the remark preceding 2.6, all elements $h \in G$ with $\ell(h) = 0$ fix A pointwise, i.e. $A \subseteq A_h$. By (a), $A \subseteq A_h$ for all $h \in G$. □

Definition. An action which is neither abelian nor dihedral is called irreducible (or of general type).

Thus an action is irreducible if and only if there exist hyperbolic isometries g, h such that $\ell(gh) > \ell(g) + \ell(h)$,where ℓ is hyperbolic length, that is, $A_g \cap A_h = \emptyset$ (by 2.4). Reducible means either abelian or dihedral, in which case the hyperbolic length is given by a homomorphism ρ as above. Once again there is a characterisation of irreducible actions analogous to 2.7 and 2.9. However, it involves more work and is deferred until the next section. For the moment, we shall just give a criterion for irreducibility, or rather reducibility.

Lemma 2.11. *Let G be a group acting on a Λ-tree (X, d). Suppose $k \in G$ acts as a hyperbolic isometry, and for all hyperbolic $g \in G$, $A_g \cap A_k \neq \emptyset$. Then the action is reducible.*

Proof. Suppose g, $h \in G$ act as hyperbolic isometries and $A_g \cap A_h = \emptyset$. Let $[p, q]$ be the bridge between A_g, A_h with $p \in A_g$. Then $[p, q] \subseteq A_k$, since A_k meets A_g and A_h. Also, $[gp, gq]$ is the bridge between A_g and $gA_h = A_{ghg^{-1}}$, and ghg^{-1} is hyperbolic (using Lemma 1.7), so $[gp, gq] \subseteq A_k$, hence $gp \in A_k$. Similarly, $g^{-1}p \in A_k$. But then $[p, gp]$, $[p, g^{-1}p]$ and $[p, q]$ represent three different directions at p in A_k, contradicting the linearity of A_k (see Exercise 1 after Lemma 2.3.2). Thus for all hyperbolic g, $h \in G$, $A_g \cap A_h \neq \emptyset$, and by Cor. 2.4 the action is reducible. (It may illuminate the proof to look at the picture after 2.2.) □

3. Pairs of Isometries

Although we gave a classification of actions in the previous section, using as little information as possible on the behaviour of pairs of isometries, for many purposes more detail is required. Specifically, one often needs information on the hyperbolic length of the product of two isometries g, h. Also one needs to know whether or not A_{gh} intersects A_g and A_h, and if so, the diameter of their intersection. We shall attempt to give that information here. We have already considered the case that $A_g \cap A_h = \emptyset$ in Lemma 2.2. As in that case, the proofs are illuminated by drawing diagrams, which are generated in a similar manner. We shall give these diagrams in some cases, but in all cases we recommend that readers try to do this themselves. The formal proofs consist, in part, of justifying the pictures. This can be quite technical, and readers may wish to skip the details on a first reading. The main ingredients in the proofs are the elementary consequences of the axioms in §1 of Chapter 2, and the results in §1 of this chapter, especially 1.1, 1.4 and 1.7. Now is also the time to remind readers of Exercise 2 after Lemma 2.3.2. A Λ-tree is linear if and only if it has no branch points (points at which there are at least three directions).

The initial problem is in defining the diameter of a set in a Λ-tree, bearing in mind that there is no notion of least upper bound, in general, of a subset of Λ. We resort to a device used by Alperin and Bass ([2]). We partially order the subsets of Λ by inclusion: $A \leq B$ if and only if $A \subseteq B$. If A is a subset of Λ consisting of non-negative elements of Λ and $b \in \Lambda$, $b \geq 0$, then $A \leq b$ means $A \leq [0, b]_\Lambda$, and we similarly define $A < b$, $b < A$, $b \leq A$. If A is a subset of Λ and $b \in \Lambda$, then, in a non-standard use of notation, $A + b$ means $\bigcup_{a \in A}[0, a + b]_\Lambda$, $A - b$ means $A + (-b)$ and $b - A$ means $\bigcup_{a \in A}[0, b - a]_\Lambda$. (In practice this notation will only be used when all the sets involved are sets of non-negative elements of Λ, and when $b \geq 0$.)

If Y is a subset of X, where (X, d) is a Λ-tree, then we define the diameter diam(Y) to be the set $\{d(p,q) \mid p, q \in Y\}$. If Y is a subtree of X, diam(Y) is a subtree of Λ with least element 0. If diam$(Y) = [0, b]_\Lambda$ for some b, then we also call b the diameter of Y. (Identifying b with $[0, b]_\Lambda$ is consistent with the notation introduced in the previous paragraph. Also, if $A \subseteq \Lambda$, $A = b$ means $A = [0, b]_\Lambda$.) If Y is a linear subtree and diam$(Y) = [0, b]_\Lambda$ for some b, then Y is a segment $[p, q]$, where $d(p, q) = b$.

If g, h are isometries of X onto X which are not inversions, and $A_g \cap A_h \neq \emptyset$, then we put $\Delta(g, h) = \text{diam}(A_g \cap A_h)$.

Let A be a linear subtree of a Λ-tree X. By definition, this means that there is a surjective isometry $\alpha : A' \to A$, where A' is a subtree of Λ. The isometry α induces an ordering on A which we denote by $\leq_\alpha$. Thus $a \leq_\alpha b$ if and only if $\alpha^{-1}(a) \leq \alpha^{-1}(b)$, where $\leq$ is the ordering on Λ. If $\beta : B \to A$ is another surjective isometry, then $\beta^{-1}\alpha : A' \to B$ is a surjective isometry and β induces an ordering $\leq_\beta$ on A. By Lemma 1.2.1, $\beta^{-1}\alpha$ is the restriction to A' of a reflection or translation in Λ. Therefore $\beta^{-1}\alpha$ either preserves or reverses the ordering $\leq$ on Λ. It follows that $\leq_\beta$ is either equal to $\leq_\alpha$ or is the dual ordering to $\leq_\alpha$.

Let g be a hyperbolic isometry of X. By Theorem 1.4, the axis A_g is a linear subtree of X, so there is an isometry $\alpha : A' \to A_g$, where A' is a subtree of Λ. Moreover, α can be chosen so that g restricted to A_g corresponds to the translation $a \mapsto a + \ell(g)$. This uniquely determines the ordering $\leq_\alpha$, which is denoted by $\leq_g$, and $<_g$ denotes the corresponding strict ordering (this is precisely the linear ordering constructed in the proof of Theorem 1.4; that is, $p <_g q$ if and only if $p \notin [gp, q]$). Note that $A_g = A_{g^n}$ (1.7), and $\leq_{g^n}$ coincides with $\leq_g$ if n is a positive integer, while $\leq_{g^{-1}}$ is dual to $\leq_g$. Also, g preserves the ordering $\leq_g$.

If g, h are isometries of X onto X, we say that g, h *meet* if $A_g \cap A_h \neq \emptyset$. We say that g, h meet coherently if they meet, and if they are both hyperbolic then the orderings $\leq_g$ and $\leq_h$ agree on $A_g \cap A_h$. (This means that either $A_g \cap A_h$ consists of a single point, or else g and h translate in the same direction along $A_g \cap A_h$.)

Suppose g and h are isometries of X which meet coherently, and at least one, say g, is hyperbolic. Choose $p \in A_g \cap A_h$. Suppose $q \in A_g \setminus A_h$, and $p \leq_g q$. Since A_h is a closed subtree of X, by 1.1 and 1.4, $[p,q] \cap A_h = [p,r]$ for some $r \in A_g \cap A_h$, and since $r \in [p,q]$, $p \leq_g r \leq_g q$. We claim that, if $s \in A_g \cap A_h$, then $s \leq_g r$. Suppose on the contrary that $r <_g s$; if $q \leq_g s$, then $q \in [p,s] \subseteq A_g \cap A_h$, a contradiction, while if $s \leq_g q$,

$$[p,r] \subsetneqq [p,s] \subseteq [p,q] \cap A_h = [p,r],$$

again giving a contradiction, hence $s \leq_g r$. Thus r is an endpoint of the linear subtree $A_g \cap A_h$. We call r the *right-hand endpoint* of g and h (because in pictures the direction of translation will be left to right). If also h is hyperbolic, then $r <_h hr$, and $hr \in A_h \setminus A_g$, for otherwise $hr \in A_g \cap A_h$ and $r <_g hr$ since g and h meet coherently, a contradiction. We can therefore apply the procedure used to find r, with h in place of g, to find $r' \in A_g \cap A_h$ such that $s \leq_h r'$ for all $s \in A_g \cap A_h$. Again since $\leq_h$ coincides with $\leq_g$ on $A_g \cap A_h$, $r = r'$. Thus it makes no difference if we replace g by h. Also, assuming g is hyperbolic, if $A_g \cap A_h$ has no right-hand endpoint, then A_g and A_h have a common end, represented by the ray $\{s \in A_g \mid p \leq_g s\}$, where $p \in A_g \cap A_h$ (and by $\{s \in A_h \mid p \leq_h s\}$, if h is hyperbolic). We call this the *common right end* of g and h when it exists.

Similarly, if g is hyperbolic and there exists $q \in A_g \setminus A_h$, with $q \leq_g p$ then there exists a point $l \in A_g \cap A_h$ such that, if $s \in A_g \cap A_h$, then $l \leq_g s$. Again if h is also hyperbolic, using h instead of g leads to the same point l. We call l the *left-hand endpoint* of g and h. Also if $A_g \cap A_h$ does not have a left-hand endpoint, A_g and A_h have a common end represented by the ray $\{s \in A_g \mid s \leq_g p\}$, called the *common left end* of g and h.

We shall commit an abuse of notation by referring to the right-hand endpoint of $A_g \cap A_h$ instead of the right-hand endpoint of g and h, etc. This is a defective notation, since the right-hand endpoint of g and h is the left-hand endpoint of g^{-1} and h^{-1}, while $A_g \cap A_h = A_{g^{-1}} \cap A_{h^{-1}}$. However, it should not cause confusion, since the subscripts on the characteristic sets will indicate the group elements involved.

If $A_g \cap A_h$ has both a right-hand and a left-hand endpoint, then it follows from the above together with the observation that $A_g \cap A_h$ is a subtree of X that $A_g \cap A_h = [l,r]$. This is precisely the case where $\Delta(g,h)$ is a segment $[0,a]_\Lambda$, and $a = d(l,r)$. For if $\Delta(g,h) = [0,a]_\Lambda$, suppose $A_g \cap A_h$ has no right-hand endpoint. We can assume g is hyperbolic, and choose $p, q \in A_g \cap A_h$ with $d(p,q) = a$ and $p \leq_g q$. Then $gq \in A_g \cap A_h$ and $d(p, gq) > a$, a contradiction. Similarly there is a left-hand endpoint. On the other hand, if g, h are hyperbolic, and $A_g \cap A_h$ has neither a left nor a right-hand endpoint, then $A_g = A_h$.

In the following lemmas, ℓ will, as usual, denote hyperbolic length, and an isometry of a Λ-tree (X, d) will mean an isometry from X onto X.

Lemma 3.1. *Assume g, h are hyperbolic isometries of a Λ-tree (X,d) which meet coherently, and let $\Delta = \Delta(g,h)$. Then gh meets both g and h coherently, $\ell(gh) = \ell(g) + \ell(h)$, $\Delta(gh,h) = \Delta + \ell(h)$ and $\Delta(gh,g) = \Delta + \ell(g)$.*

Proof. Choose $m \in A_g \cap A_h$. If $x \in [g^{-1}m, m] \cap [m, hm]$ then $x \in A_g \cap A_h$ and $x \leq_g m$, $m \leq_h x$ (from the definition of these orderings). Since g, h meet coherently, $x = m$, hence $[g^{-1}m, m] \cap [m, hm] = \{m\}$, so $[g^{-1}m, hm] = [g^{-1}m, m, hm]$. Applying the isometries g and h^{-1} gives

$$[m, ghm] = [m, gm, ghm] \quad \text{and} \quad [h^{-1}g^{-1}m, m] = [h^{-1}g^{-1}m, h^{-1}m, m].$$

Since $m \neq gm$, $h^{-1}m$, we may splice by Lemma 2.1.5 to obtain

$$[h^{-1}g^{-1}m, ghm] = [h^{-1}g^{-1}m, h^{-1}m, m, gm, ghm],$$

whence $m \in A_{gh}$ (which shows $A_g \cap A_h \subseteq A_{gh}$) and by Lemma 2.1.4 and Cor.1.5,

$$\begin{aligned}\ell(gh) = d(m, ghm) &= d(m, gm) + d(gm, ghm) \\ &= d(m, gm) + d(m, hm) = \ell(g) + \ell(h).\end{aligned}$$

Note also that $h^{-1}m$, $gm \in A_{gh}$ and $h^{-1}m \leq_{gh} m \leq_{gh} gm$, so gh meets g, h coherently.

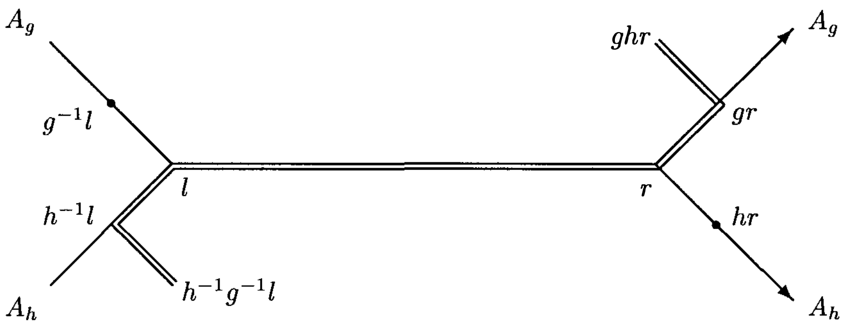
Now if $p, q \in A_g \cap A_h$, with $p \leq_h q$, then $[h^{-1}p, q] = [h^{-1}p, p, q]$ is contained in $A_{gh} \cap A_h$, so $d(h^{-1}p, q) = \ell(h) + d(p,q) \in \Delta(gh, h)$, hence $\Delta + \ell(h) \leq \Delta(gh, h)$, since $\Delta(gh,h)$ is a subtree of Λ containing 0.

If $A_g \cap A_h$ has a right-hand endpoint r, then we claim r is also the right-hand endpoint of $A_{gh} \cap A_h$. For if $s \in A_{gh} \cap A_h$ and $r <_h s$, then $[s, r] \cap [r, h^{-1}r] = \{r\}$, since $h^{-1}r <_h r$. If $x \in [s, r] \cap [r, gr]$, then $x \in A_h \cap A_g$, so $x \leq_g r$ by definition of r, and $r \leq_g x$ since $x \in [r, gr]$, hence $x = r$. Thus $[s, r] \cap [r, gr] = \{r\}$. By the above, the points $h^{-1}r$, r, gr belong to A_{gh}, and $[h^{-1}r, r] \cap [r, gr] = \{r\}$. But then $[r, s]$, $[r, gr]$ and $[r, h^{-1}r]$ define three different directions at r in A_{gh}, contradicting the linearity of A_{gh}. Thus r is the greatest element of $A_{gh} \cap A_h$ with respect to the ordering $\leq_h$, that is, r is the right-hand endpoint of $A_{gh} \cap A_h$, as claimed.

If $A_g \cap A_h$ has a left-hand endpoint l, then we claim $h^{-1}l$ is the left-hand endpoint of $A_{gh} \cap A_h$. For $[g^{-1}l, l] \cap A_h = \{l\}$ by definition of l, so $[h^{-1}g^{-1}l, h^{-1}l] \cap A_h = \{h^{-1}l\}$. By the above,

$$[h^{-1}g^{-1}l, l] = [h^{-1}g^{-1}l, h^{-1}l, l] \subseteq A_{gh},$$

so $[h^{-1}g^{-1}l, l] \cap A_h = [l, h^{-1}l]$. It again follows from the linearity of A_{gh} that A_{gh} contains no points s of A_h with $s <_h h^{-1}l$. For otherwise, there are three different directions at $h^{-1}l$ in A_{gh}, defined by $[h^{-1}l, s]$,$[h^{-1}l, h^{-1}g^{-1}l]$ and $[h^{-1}l, l]$. Thus $h^{-1}l$ is the left-hand endpoint of $A_{gh} \cap A_h$, as claimed. The situation when there is a left and right-hand endpoint is illustrated in the following picture, where A_{gh} is indicated by a double line. If l or r is missing, this gives rise to a common left or right-hand end for A_g, A_h and A_{gh}, because $A_g \cap A_h \subseteq A_{gh}$. If both are missing, the three axes coincide.



Since $d(h^{-1}l, l) = \ell(h)$ when there is a left-hand endpoint, it follows that, in all cases, if $p \in A_{gh} \cap A_h$, then $d(p, A_g \cap A_h) \leq \ell(h)$. If p, $q \in A_{gh} \cap A_h$, it then follows that $d(p,q) \in \Delta(g,h) + \ell(h)$ unless both p, q are not in $A_g \cap A_h$. This can only happen when there is a left-hand endpoint and p, $q \in [h^{-1}l, l]$. But then $d(p,q) \leq d(h^{-1}l, l) = \ell(h)$, so again $d(p,q) \in \Delta(g,h)+\ell(h)$. It follows that $\Delta(gh, h) = \Delta(g,h) + \ell(h)$.

Finally, replacing g, h by h^{-1}, g^{-1} in this argument and using Lemma 1.7 gives

$$\Delta(gh, g) = \Delta(h^{-1}g^{-1}, g^{-1}) = \Delta(h^{-1}, g^{-1}) + \ell(g^{-1}) = \Delta + \ell(g)$$

and the proof is complete. □

Lemma 3.2. *Assume g, h are isometries of a Λ-tree (X,d) which meet coherently, and $\Delta = \Delta(g,h) > 0$. Then gh meets both g and h coherently, $\ell(gh) = \ell(g) + \ell(h)$, $\Delta(gh,h) = \Delta + \ell(h)$ and $\Delta(gh,g) = \Delta + \ell(g)$.*

Proof. In view of the previous lemma, we can assume one of $\ell(g)$, $\ell(h)$ is zero. Suppose first that $\ell(g) \geq \ell(h) = 0$ (so h is elliptic) and put $D = A_g \cap A_h$. If $\ell(g) = 0$, then gh fixes all points of D, so $\ell(gh) = 0$, and $D = A_{gh} \cap A_h = A_{gh} \cap A_g$, and the lemma follows.

Assume $\ell(g) > 0$, and suppose $x \in D$. Then we claim that there exists $y \in D$ such that $x \neq y$ and either $y \in [g^{-1}x, x]$ or $x \in [g^{-1}y, y]$. For $\Delta > 0$, so we can choose $z \in D$ with $z \neq x$. If $z \leq_g x$, we take y to be the maximum of z and $g^{-1}x$ with respect to the linear ordering $\leq_g$, so $g^{-1}x \leq_g y \leq_g x$, hence $y \in [g^{-1}x, x]$. Otherwise, $x \leq_g z$ and we take y to be the minimum of gx, z with respect to $\leq_g$. Then $x \leq_g y \leq_g gx$, hence $g^{-1}y \leq_g x \leq_g y$ (since $\leq_g$ is invariant under g), that is, $x \in [g^{-1}y, y]$ as required.

We next show that $x \in A_{gh}$. In view of the previous paragraph, it suffices to show that, if u, $v \in D$, $u \neq v$ and $v \in [g^{-1}u, u]$, then both u and $v \in A_{gh}$. In this situation, $[g^{-1}u, u] = [g^{-1}u, v, u]$ and applying h^{-1} gives $[h^{-1}g^{-1}u, h^{-1}u] = [h^{-1}g^{-1}u, v, u]$. Therefore,

$$[h^{-1}g^{-1}u, u] \cap [u, ghu] = [h^{-1}g^{-1}u, v, u] \cap [u, gu].$$

Since $[v, u] \cap [u, gu] \subseteq [g^{-1}u, u] \cap [u, gu] = \{u\}$, it follows that $[h^{-1}g^{-1}u, u] \cap [u, ghu] = \{u\}$ (see Exercise (2) after Lemma 2.1.11). Hence, $u \in A_{gh}$ and $v \in [h^{-1}g^{-1}u, u] \subseteq A_{gh}$, as required.

We have shown that if $x \in D$, then $x \in A_{gh}$, so $gx = ghx \in A_{gh}$. This shows that $D \cup gD \subseteq A_{gh}$, and gh meets g, h coherently. Also, choosing $x \in D$, $\ell(gh) = d(x, ghx) = d(x, gx) = \ell(g)$ by Cor.1.5, in particular, gh is hyperbolic, so A_{gh} is linear.

Suppose $x \in A_{gh} \cap A_g$ and let $[p, x]$ be the bridge between A_h and x; we shall show that $p \in D$ and $x \in [p, gp]$. Since D has at least two elements, we can find $q \in D$ with $p \neq q$. By Lemma 1.1, $p = \mathrm{Y}(x, q, hx)$, in particular $p \in [q, x]$ by Lemma 2.1.2, so $p \in A_g$, hence $p \in D \subseteq A_{gh}$. It also follows that

$$[p, hx] \cap A_g = \{p\} \qquad (*)$$

for otherwise $hp = p \neq hx$, so $p \neq x$ and by Lemma 2.1.2, $[p, q]$, $[p, x]$ and $[p, hx] \cap A_g$ define three different directions at p in A_g, contradicting the linearity of A_g. (Since A_g is a closed subtree, $[p, hx] \cap A_g$ is a segment with p as an endpoint.)

If $x <_g p$, then by linearity of A_g, the bridge between $g^{-1}p$ and A_h is $[g^{-1}p, p]$. Applying $(*)$ to $g^{-1}p$ in place of x and h^{-1} in place of h gives $[h^{-1}g^{-1}p, p] \cap A_g = \{p\}$. But then $[h^{-1}g^{-1}p, p]$, $[x, p]$ and $[p, q]$ give three directions in A_{gh} at p, contradicting the linearity of A_{gh} (note that p, $q \in A_{gh}$ since $D \subseteq A_{gh}$). Therefore $p \leq_g x$.

From $(*)$, $[hx, p]$ is the bridge between hx and A_g, so by Theorem 1.4, $[p, ghx] = [p, gp, ghx]$, hence $gp \in A_{gh}$. If $x \notin [p, gp]$, then $gp \in [p, x]$ (since

$p \leq_g x$) and $x \neq gp$. By (*), $[gp, ghx] \cap A_g = \{gp\}$, so $[gp, ghx]$ meets $[gp, p]$ and $[gp, x]$ only in gp. Hence $[gp, p]$, $[gp, x]$ and $[gp, ghx]$ define three different directions at gp in A_{gh}, contradicting the linearity of A_{gh}. Thus $x \in [p, gp]$.

We have shown that, if $x \in A_{gh} \cap A_g$, then $x \in [p, gp]$, where $p \in D$, hence x is in the subtree spanned $D \cup gD$. It follows that $A_{gh} \cap A_g$ is the subtree spanned by $D \cup gD$, a subtree of the linear tree A_g, which is easily seen to have diameter $\Delta + \ell(g)$. Also $D \subseteq A_{gh} \cap A_h$; since h^{-1} is elliptic, gh and h^{-1} meet coherently, and replacing g, h by gh, h^{-1}, we obtain $A_{gh} \cap A_h = D$, which has diameter Δ, as required. Finally, if $\ell(h) \geq \ell(g) = 0$, applying what we have proved with h^{-1}, g^{-1} in place of g, h, and using Lemma 1.7, $\Delta(gh, g) = \Delta(h^{-1}g^{-1}, g^{-1}) = \Delta(h^{-1}, g^{-1}) + \ell(g^{-1}) = \Delta(g, h) + \ell(g)$, similarly $\Delta(gh, h) = \Delta(g, h) + \ell(h)$, and $\ell(gh) = \ell(h^{-1}g^{-1}) = \ell(h^{-1}) + \ell(g^{-1}) = \ell(g) + \ell(h)$. □

Lemma 3.3. *Assume g, h are isometries of a Λ-tree (X, d) which meet coherently, and $\Delta = \Delta(g, h) < \min\{\ell(g), \ell(h)\}$. Then*
(1) *gh meets g, h coherently and $\ell(gh) = \ell(g) + \ell(h)$.*
(2) *gh^{-1} meets g and h^{-1} coherently, and $\ell(gh^{-1}) = \ell(g) + \ell(h) - 2\Delta$.*
(3) *$\Delta(gh, h) = \Delta + \ell(h)$ and $\Delta(gh, g) = \Delta + \ell(g)$.*
(4) *$\Delta(gh^{-1}, h) = \ell(h) - \Delta$ and $\Delta(gh, g) = \ell(g) - \Delta$.*

Proof. In this case g, h are hyperbolic. If $x \in A_g \cap A_h$, then $g^2x \in A_g \setminus A_h$, since $d(x, g^2x) = 2\ell(g) > \Delta$. Hence A_g, A_h do not have a common right end, so $A_g \cap A_h$ has a right-hand endpoint, which we denote by r. Similarly $A_g \cap A_h$ has a left-hand endpoint l. Therefore Δ may be interpreted as $d(l, r)$, and the formulas in the statement of the lemma are to be interpreted accordingly. The situation is illustrated by the diagram following the proof, and the proof consists of justifying this picture. (Note that the double line indicates part of $A_{gh^{-1}}$.)

We have $l <_h r$, $h^{-1}r <_h r$ and $d(l, r) < \ell(h) = d(r, h^{-1}r)$, hence $h^{-1}l <_h h^{-1}r <_h l$, so $[h^{-1}l, l] = [h^{-1}l, h^{-1}r, l]$, and by definition of l, this segment is the bridge between $h^{-1}l$ and A_g. Hence $[gh^{-1}l, gl] = [gh^{-1}l, gh^{-1}r, gl]$ is the bridge between $gh^{-1}l$ and A_g. Since $r \in A_g$,

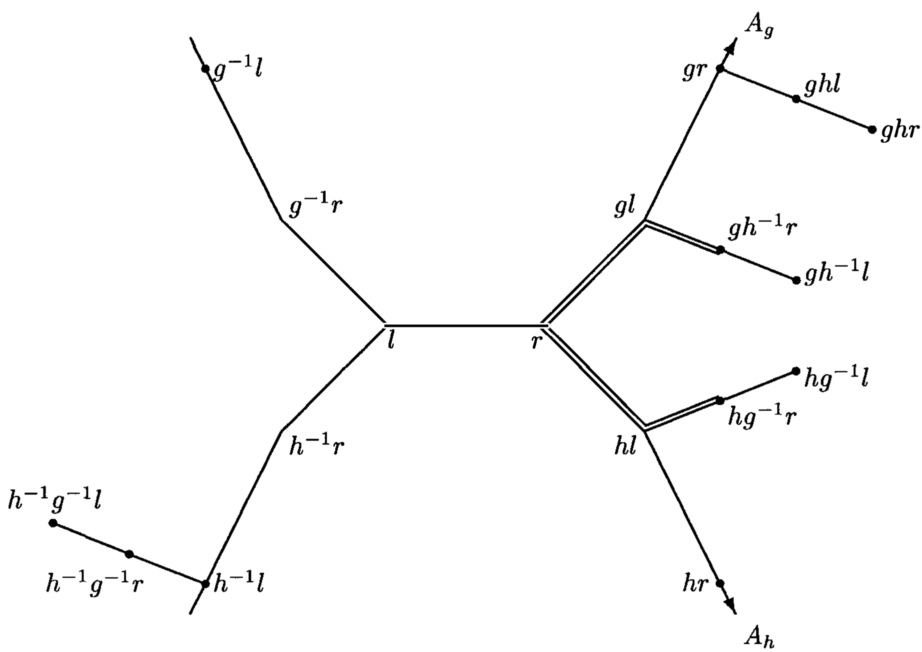
$$[r, gh^{-1}l] = [r, gl, gh^{-1}r, gh^{-1}l]. \tag{1}$$

Similarly, $[g^{-1}l, l] = [g^{-1}l, g^{-1}r, l]$ is the bridge between $g^{-1}l$ and A_h, so the bridge between $hg^{-1}l$ and A_h is $[hg^{-1}l, hl] = [hg^{-1}l, hg^{-1}r, hl]$, hence

$$[r, hg^{-1}l] = [r, hl, hg^{-1}r, hg^{-1}l]. \tag{2}$$

Also, since $d(l, r) < \ell(g) = d(l, gl)$, $r <_g gl$, and similarly $r <_h hl$, so by definition of r, $[r, gl] \cap [r, hl] = \{r\}$. By Lemma 2.1.5, we may splice the segments in (1) and (2), in particular, $r \in [hg^{-1}r, gh^{-1}r]$, that is, $r \in A_{gh^{-1}}$.

Since $gl \in [r, gh^{-1}r]$, $r <_{gh} gl$, hence gh^{-1} meets g coherently. Similarly, since $hl \in [r, hg^{-1}r]$, gh^{-1} meets h^{-1} coherently. Further, $\ell(gh^{-1}) = d(r, gh^{-1}r) = d(r, gl) + d(gl, gh^{-1}r) = (\ell(g) - \Delta) + (\ell(h) - \Delta)$, as required. It also follows easily that $A_{gh^{-1}}$ meets A_g in the segment $[r, gl]$ and meets A_h in the segment $[hl, r]$. This proves (2) and (4). The rest of the lemma can be proved similarly, but it follows from Lemma 3.1. □



Lemma 3.4. *Assume g, h are isometries of a Λ-tree (X, d) which meet coherently, and $\Delta = \Delta(g, h) > \min\{\ell(g), \ell(h)\}$. Then*

(1) *gh meets g, h coherently and $\ell(gh) = \ell(g) + \ell(h)$.*

(2) *Assume $\ell(h) \leq \ell(g)$. Then gh^{-1} meets g and h coherently, and $\ell(gh^{-1}) = \ell(g) - \ell(h)$.*

(3) *$\Delta(gh, h) = \Delta + \ell(h)$ and $\Delta(gh, g) = \Delta + \ell(g)$.*

(4) *Assume $\ell(h) \leq \ell(g)$. Then $\Delta(gh^{-1}, h) = \Delta - \ell(h)$ and $\Delta(gh^{-1}, g) = \Delta + \ell(g) - 2\ell(h)$.*

Proof. (1) and (3) follow from Lemma 3.2. Assume $\ell(h) \leq \ell(g)$. If $\ell(h) = 0$ the result follows by replacing h by h^{-1}, so we assume $\ell(h) > 0$. There is a segment $[p, q] \subseteq A_g \cap A_h$ with $d(p, q) > \ell(h)$, and we assume $p \leq_h q$. Then by

Cor.1.5, $p \leq_h h^{-1}q \leq_h q$, so $h^{-1}q \in [p,q] \subseteq A_g$, and $g^{-1}q \leq_g h^{-1}q <_g q$ (since $\ell(h) \leq \ell(g)$). Thus $[g^{-1}q,q] = [g^{-1}q, h^{-1}q, q]$, hence $[q,gq] = [q, gh^{-1}q, gq]$, and by Lemma 2.1.4,

$$\begin{aligned} d(q,gq) &= d(q,gh^{-1}q) + d(gh^{-1}q,gq) = d(q,gh^{-1}q) + d(hq,q) \\ &= d(q,gh^{-1}q) + \ell(h) \end{aligned}$$

and so $d(q,gh^{-1}q) = \ell(g) - \ell(h)$.

Also, either $[g^{-1}q,q] = [g^{-1}q,p,h^{-1}q,q]$ or $[p,q] = [p,g^{-1}q,h^{-1}q,q]$, and applying h, either $[hg^{-1}q,q] = [hg^{-1}q,hp,q]$ or $[hp,q] = [hp,hg^{-1}q,q]$. By the choice of p and q, $hp <_g q$, hence $[hg^{-1}q,q] \cap [q,gq] = \{q\}$, so $[hg^{-1}q,q] \cap [q,gh^{-1}q] = \{q\}$, because $[q,gq] = [q,gh^{-1}q,gq]$, as already noted. This means $q \in A_{gh^{-1}}$, so $\ell(gh^{-1}) = d(q,gh^{-1}q) = \ell(g) - \ell(h)$. It also follows easily that gh^{-1} meets g, h coherently, and it remains to prove (4).

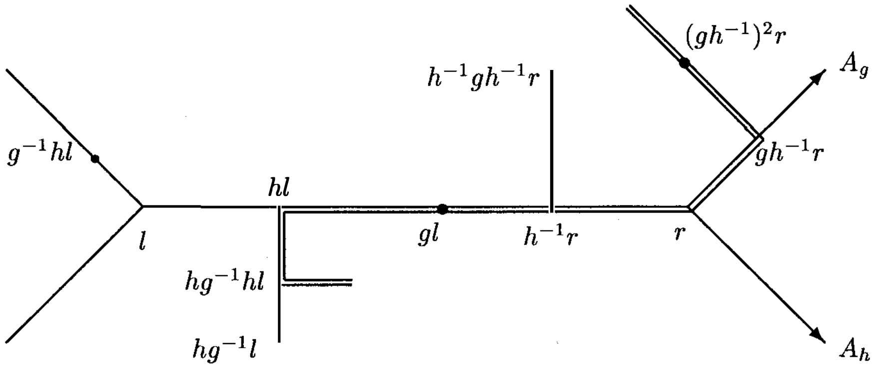
If $A_g \cap A_h$ has a left-hand endpoint l, then by definition of l, $[g^{-1}l,l] \cap A_h = \{l\}$. Also, since $\ell(h) < \delta$, $hl \in A_g \cap A_h$, and since $\ell(h) \leq \ell(g)$, $l \leq_g hl \leq_g gl$, hence $g^{-1}l \leq_g g^{-1}hl \leq_g l$, so $[g^{-1}l,l] = [g^{-1}l, g^{-1}hl, l]$. Therefore $[g^{-1}hl,l] \cap A_h = \{l\}$ and so $[hg^{-1}hl,hl] \cap A_h = \{hl\}$. We claim that $hl \in A_{gh^{-1}}$. Since $gh^{-1}(hl) = gl$, this means that $[hg^{-1}hl,hl] \cap [hl,gl] = \{hl\}$. Suppose not, and put $w = \mathrm{Y}(hg^{-1}hl, hl, gl)$, so $w \neq hl$. Since $w \in [hl,gl]$, $hl <_g w \leq_g gl$, and since $\ell(h) < \Delta$, there exists $q \in A_g \cap A_h$ such that $hl <_g q$. If $w <_g q$, we replace q by w. This gives q such that $q \in A_g \cap A_h$, $q \in [hl,w]$ and $q \neq hl$. But $[hl,w] \subseteq [hl, hg^{-1}hl]$, and we have seen that this segment intersects A_h, and so $A_g \cap A_h$, only in hl, a contradiction, which establishes the claim.

Since $[hg^{-1}hl,hl] \cap A_h = \{hl\}$, $[hg^{-1}hl,hl] \cap [l,hl] = \{hl\}$, and $l <_h hl$, so $l <_g hl$. It follows that hl is the left-hand endpoint of $A_{gh^{-1}} \cap A_h$ and of $A_{gh^{-1}} \cap A_g$.

If $A_g \cap A_h$ has no left-hand endpoint, then $A_g \cap A_h \subseteq A_{gh^{-1}}$. For then if $q \in A_g \cap A_h$, we may take $p = g^{-2}q$ in the argument above to see that $q \in A_{gh^{-1}}$. It follows that A_g, A_h and $A_{gh^{-1}}$ have a common left-hand end.

If $A_g \cap A_h$ has a right-hand endpoint r, then applying the previous argument with g, h replaced by g^{-1}, h^{-1}, we find that $h^{-1}r$ is the left-hand endpoint of $A_{g^{-1}h} \cap A_{h^{-1}}$ and of $A_{g^{-1}h} \cap A_{h^{-1}}$, so is the right-hand endpoint of $A_{h^{-1}g} \cap A_h$ and $A_{h^{-1}g} \cap A_g$. Applying the translations g, h to their axes and using Lemma 1.7, we find r is the right-hand endpoint of $A_{g^{-1}h} \cap A_h$ and $gh^{-1}r$ is the right-hand endpoint of $A_{g^{-1}h} \cap A_g$.

Similarly, if $A_g \cap A_h$ has no right-hand endpoint, A_g, A_h and $A_{gh^{-1}}$ have a common right-hand end. (Alternatively, if p q are as in the first part of the argument, that argument shows that $s \in A_{gh^{-1}} \cap A_h \cap A_g$ for all $s \in A_h$ with $q \leq_h s$.) The lemma now follows easily, and details are left to the reader. □



The situation when $A_g \cap A_h$ has a left and right-hand endpoint and $\Delta > \ell(g) > \ell(h) > 0$ is illustrated in this picture. (The point $h^{-1}r$ could occur in a different position relative to hl and gl.) The axis of gh^{-1} is indicated by the double line.

Lemma 3.5. *Assume g, h are isometries of a Λ-tree (X, d) which meet coherently, and $\Delta = \Delta(g,h) = \min\{\ell(g), \ell(h)\}$. Let r be the right-hand endpoint of $A_g \cap A_h$ (if $\Delta = 0$ then we take r to be the point such that $A_g \cap A_h = \{r\}$). Assume $\ell(h) \leq \ell(g)$, and let $p = \mathrm{Y}(hg^{-1}r, r, gh^{-1}r)$. Then*

(1) *g, h meet gh^{-1} coherently and $\ell(gh^{-1}) = \max\{\ell(g) - \ell(h) - 2d(r,p), 0\}$.*
(2) *If $\ell(gh^{-1}) > 0$, then $A_{gh^{-1}} \cap A_g = [p, gh^{-1}p]$, and $\Delta(gh^{-1}, g) = \ell(gh^{-1})$. Also, $d(A_{gh^{-1}}, A_h) = d(r,p)$, and if $d(r,p) = 0$, then $A_{gh^{-1}} \cap A_h = \{r\}$, so $\Delta(gh^{-1}, h) = 0$.*
(3) *If $\ell(gh^{-1}) = 0$, then either gh^{-1} is an inversion, or $A_{gh^{-1}} \cap A_g = \{q\}$, where q is the midpoint of $[r, gh^{-1}r]$, so $\Delta(gh^{-1}, g) = 0$, and $d(A_{gh^{-1}}, A_h) = d(r, q)$, and $d(r, q) > 0$ unless $\ell(g) = \ell(h)$, in which case $\Delta(gh^{-1}, h) = 0$.*

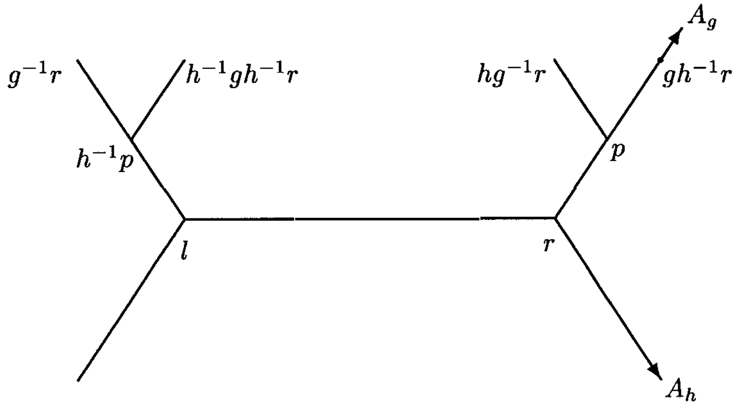
Proof. By Lemma 1.8, $\ell(gh^{-1}) = \max\{d(r, gh^{-1}r) - 2d(r,p), 0\}$. If $\ell(g) = \ell(h) = 0$, then (1) is clearly true. If $\ell(g) > \ell(h) = 0$, then $gh^{-1}r = gr$, so $d(r, gh^{-1}r) = \ell(g)$ and (1) holds. If $\ell(h) > 0$, then $g^{-1}r \leq_g h^{-1}r = l \leq_g r$, so $r \leq_g gh^{-1}r \leq_g gr$, hence $[r, gh^{-1}r, gr]$ is a piecewise segment, and a calculation like that at the start of the proof of the previous lemma, using Lemma 2.1.4, shows that $d(r, gh^{-1}r) = \ell(g) - \ell(h)$.

If $\ell(g) = \ell(h)$, then it follows that $\ell(gh^{-1}) = 0$ and we claim that r is the unique fixed point of gh^{-1} in $A_g \cup A_h$. This is obvious if $\ell(g) = \ell(h) = \Delta = 0$. Otherwise, since $A_{gh^{-1}} = A_{hg^{-1}}$, it suffices by symmetry to check that if $x \in A_g$ and $x \neq r$, then gh^{-1} does not fix x. We leave it to readers to check that in all cases, either $h^{-1}x \notin A_g$, or $h^{-1}x <_g l$, and the second case can occur only when $r <_g x$. (It will help to draw a picture.) It follows that $gh^{-1}x \neq x$, in the first case because A_g is $\langle g \rangle$-invariant, and in the second case because $gh^{-1}x <_g r <_g x$. Thus the assertions in (3) are true in this case.

Assume $\ell(g) > \ell(h) = \Delta$. Let l be the left-hand endpoint of $A_g \cap A_h$ (note that, if $\ell(h) = 0$, then $l = r$), so $r = hl$. Since $\ell(h) < \ell(g)$ and g, h meet coherently, $g^{-1}r <_g l$, so $[g^{-1}r, l]$ meets A_h in l, by definition of l, and applying h, $[hg^{-1}r, r]$ meets A_h in r. Hence $[hg^{-1}r, p] \cap A_h = \emptyset$ unless $p = r$, when $[hg^{-1}r, p] \cap A_h = \{r\}$. Similarly, $r <_g gl = gh^{-1}r$, so by definition of r, $[r, gh^{-1}r] \cap A_h = \{r\}$. Also, since $p \in [r, gh^{-1}r]$, $r \leq_g p \leq_g gh^{-1}r$.

We claim that $[hg^{-1}r, p] \cap A_g = \{p\}$. For suppose $x \in [hg^{-1}r, p] \cap A_g$ and $x \neq p$. Then $[r, hg^{-1}r] = [r, p, x, hg^{-1}r]$, so $r \leq_g p <_g x$, since $r \leq_g p$. If $x \leq_g gh^{-1}r$, then $x \in [p, gh^{-1}r]$, which meets $[p, hg^{-1}r]$ in p (Lemma 2.1.2), so $x = p$, a contradiction. Hence $gh^{-1}r <_g x$, which implies that $d(r, gh^{-1}r) < d(r, x) \leq d(r, hg^{-1}r)$, a contradiction, because $d((r, gh^{-1}r) = d(r, hg^{-1}r)$, since gh^{-1} is an isometry. This establishes the claim.

Since $[r, gh^{-1}r]$ meets A_h in r, applying h^{-1} we find that meets A_h in l. Also, $p \in [r, hg^{-1}r]$ and applying h^{-1}, $h^{-1}p \in [l, g^{-1}r]$, and since $g^{-1}r <_g l$, it follows that $g^{-1}r \leq_g h^{-1}p \leq_g l$. Further, applying g, $gh^{-1}p \in [gh^{-1}r, r]$. A similar argument to that in the last paragraph shows that $[h^{-1}p, h^{-1}gh^{-1}r] \cap A_g = \{h^{-1}p\}$. This is left to readers. Applying g gives $[gh^{-1}p, (gh^{-1})^2 r] \cap A_g = \{gh^{-1}p\}$. The situation in the case that $\ell(h) > 0$ is partly illustrated by the picture below. (When $\ell(h) = 0$, one needs to replace the segment $[l, r]$ by a single point, and remove the segments representing the axis of h.)



Next, we note that

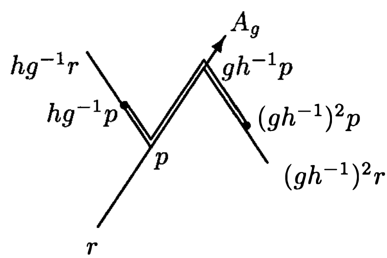
$$d(r, gh^{-1}p) = d(r, hg^{-1}p) = \ell(g) - \ell(h) - d(r, p). \qquad (*)$$

For $d(r, gh^{-1}p) = d(g^{-1}r, h^{-1}p) = d(g^{-1}r, l) - d(h^{-1}p, l) = d(g^{-1}r, r) - d(l, r) - d(p, r) = \ell(g) - \ell(h) - d(r, p)$. Further, $d(hg^{-1}p, r) = d(g^{-1}p, l) = d(g^{-1}p, p) - d(l, r) - d(r, p) = \ell(g) - \ell(h) - d(p, r)$, as required.

Suppose $\ell(gh^{-1}) > 0$; then by $(*)$ and Part (1) of the lemma, $d(r, gh^{-1}p) > d(r,p)$, so $[r, gh^{-1}r] = [r, p, gh^{-1}p, gh^{-1}r]$ hence $p \leq_g gh^{-1}p$. Applying hg^{-1}, $[r, hg^{-1}r] = [r, p, hg^{-1}p, hg^{-1}r]$. Since $p \in [hg^{-1}r, gh^{-1}r]$,

$$[hg^{-1}r, gh^{-1}r] = [hg^{-1}r, hg^{-1}p, p, gh^{-1}p, gh^{-1}r],$$

hence $p \in A_{gh^{-1}}$. Since $gh^{-1}p \in [p, gh^{-1}r]$, which meets A_g in p, we have $[hg^1p, p] \cap A_g = \{p\}$. Further, applying gh^{-1}, $(gh^{-1})^2p \in [gh^{-1}p, (gh^{-1})^2r]$, which meets A_g in $gh^{-1}p$, hence $[gh^{-1}p, (gh^{-1})^2p] \cap A_g = \{gh^{-1}p\}$. By the linearity of $A_{gh^{-1}}$, gh^{-1} meets g coherently in $[p, gh^{-1}p]$, which has length $\ell(gh^{-1})$. It follows from the discussion above that $[p, r]$ is the bridge between A_h and $A_{gh^{-1}}$ and if $p = r$, then $A_h \cap A_{gh^{-1}} = \{r\}$. This case is illustrated by the following elaboration of part of the previous diagram (not to scale). The double line denotes part of $A_{gh^{-1}}$.



Suppose $\ell(gh^{-1}) = 0$. Then $gh^{-1}p$ and $hg^{-1}p$ are both in $[r,p]$, at the same distance from r, hence $gh^{-1}p = hg^{-1}p$, so $(gh^{-1})^2$ has a fixed point. Assuming gh^{-1} is not an inversion, it is elliptic, so fixes the midpoint q of $[r, gh^{-1}r]$, and fixes no other point of $[r, gh^{-1}r]$, by Lemma 1.1. Hence $A_{gh^{-1}} \cap A_g = \{q\}$, by linearity of A_g, and $d(A_{gh^{-1}}, A_h) = d(r,q)$. Also, from the beginning of the proof, $d(r,q) = (1/2)d(r, gh^{-1}r) = (1/2)(\ell(g) - \ell(h)) > 0$. □

Note. To complete the picture in Lemma 3.5, if $\ell(h) = 0$ we can replace h by h^{-1} in the lemma to find $\ell(gh)$, etc. Otherwise, Lemma 3.1 applies.

This concludes the analysis of pairs of isometries. We shall use parts of it to establish the promised characterisation of irreducible actions. This involves the idea of a properly discontinuous action, which is a common one in geometry and topology. However, several different definitions are in use, which are not equivalent for arbitrary spaces. For us, the most appropriate definition is the following.

Definition. Let G be a group acting on a topological space X. The action is called *properly discontinuous* if, for all $x \in X$, there is a neighbourhood N of x such that if $g \in G$ and $gN \cap N \neq \emptyset$, then $gx = x$.

If (X,d) is a Λ-metric space, $\Lambda \neq \langle 0 \rangle$ and X is given the topology induced by d (see §1.2), this is equivalent to requiring that, for all $x \in X$, there exists

$k > 0$ in Λ such that $d(x, gx) < k$ implies $gx = x$, for all $g \in G$. We also have to introduce the idea of a free action, which is standard.

Definition. An action of a group G on a topological space X is *free* if, for all $x \in X$ and $g \in G$, $gx = x$ implies $g = 1$.

Thus a free action is one with all point stabilizers trivial. An action of a group G on a Λ-metric space (X, d) is free and properly discontinuous if and only if, for all $x \in X$, the set $\{d(x, gx) | g \in G, g \neq 1\}$ has a positive lower bound in Λ, again assuming $\Lambda \neq \langle 0 \rangle$.

In dealing with an action of a group G on a Λ-tree (X, d), the assumption that the action is free is often inadequate, because of the possibility of inversions (which have a "phantom" fixed point in $\frac{1}{2}\Lambda \otimes_\Lambda X$). An action is free and without inversions if and only if every non-identity element of the group acts as a hyperbolic isometry. Such actions will be the object of study in Ch.5.

Lemma 3.6. *Suppose G is a group acting on a Λ-tree (X, d), and G is generated by g and h, where g and h are hyperbolic elements such that $A_g \cap A_h$ is a segment of length less than* $\min\{\ell(g), \ell(h)\}$. *Then G is a free group with basis $\{g, h\}$, and there exists $k > 0$, $k \in \Lambda$, such that, for all $a \in G$, we have $\ell(a) \geq k$. Consequently, the action is free, without inversions and properly discontinuous.*

Proof. Replacing h by h^{-1} if necessary, we can assume g, h meet coherently. The situation is illustrated by the picture in the proof of 3.3. The proof is a variant of the "ping-pong" argument used in the Klein-Maskit combination theorems (for algebraic versions of these, see §12, Ch. III in [95]). It will help to refer to the diagram after the proof of Lemma 3.3 during the course of the proof. We define the following subsets of X.

$$\begin{aligned} X_g &= \{x \in X \mid [x, r] = [x, gl, r]\} \\ X_h &= \{x \in X \mid [x, r] = [x, hl, r]\} \\ X_{g^{-1}} &= \{x \in X \mid [x, l] = [x, g^{-1}r, l]\} \\ X_{h^{-1}} &= \{x \in X \mid [x, l] = [x, h^{-1}r, l]\}. \end{aligned}$$

These are non-empty subsets of X (clearly $gl \in X_g$, etc), and are pairwise disjoint. For if x is one of hl, $g^{-1}r$, $h^{-1}r$ then it is easily checked that $[x, gl] = [x, r, gl]$ and $x \neq r \neq gl$. It follows easily that X_g does not intersect the other three sets, and by symmetry the four sets are pairwise disjoint (we can interchange g and h, and replace g, h by g^{-1}, h^{-1}). We claim that

$$\begin{aligned} g(X_g \cup X_h \cup X_{h^{-1}}) &\subseteq X_g \\ h(X_h \cup X_g \cup X_{g^{-1}}) &\subseteq X_h \\ g^{-1}(X_{g^{-1}} \cup X_h \cup X_{h^{-1}}) &\subseteq X_{g^{-1}} \\ h^{-1}(X_{h^{-1}} \cup X_g \cup X_{g^{-1}}) &\subseteq X_{h^{-1}}. \end{aligned}$$

Again by symmetry it suffices to check the first of these. If $x \in X_g$, let $[x,p]$ be the bridge between x and A_g; since $gl \in A_g$, $p \in [x,gl]$ (see Lemma 2.1.9), hence $[x,r] = [x,gl,r] = [x,p,gl,r]$. Since $r <_g gl$ (because $\Delta(g,h) < \ell(g)$), it follows that $gl \leq_g p \leq_g gp$. Also, $[gx,gp]$ is the bridge between gx and A_g and $r \in A_g$, so $[gx,r] = [gx,gp,r] = [gx,gp,p,gl,r]$, hence $gx \in X_g$. Suppose $x \in X_h$. Since $r <_h hl$, (because $\Delta(g,h) < \ell(h)$) it follows easily from the definition of r that $[x,r]$ contains no points of A_g other than r, that is, $[x,r]$ is the bridge between x and A_g, so $[gx,gr]$ is the bridge between gx and A_g, hence $[gx,r] = [gx,gr,r] = [gx,gr,gl,r]$ (because $r <_g gl \leq_g gr$, since $l \leq_g r$), so $gx \in X_g$. Finally if $x \in X_{h^{-1}}$, then similarly $[x,l]$ is the bridge between x and A_g, so $[gx,gl]$ is the bridge between gx and A_g. It follows that $[gx,r] = [gx,gl,r]$, so $gx \in X_g$.

If $w = u_1 \dots u_n \in G$, where $n \geq 1$ and $(u_1,\dots,u_n)$ is a reduced word on $\{g^{\pm 1}, h^{\pm 1}\}$, it follows by induction on n that, if $U = \bigcup_{v \neq u_n^{-1}} X_v$, then $wU \subseteq X_{u_1}$. Therefore, either $wU \subsetneqq U$, or $wU \cap U = \emptyset$. In either case, $w \neq 1$, so G is free on g, h.

Now if $u \in \{g^{\pm 1}, h^{\pm 1}\}$, then $u[l,r] \subseteq X_u$. Hence, $w[l,r] \subseteq X_{u_1}$, so w has no fixed point in $[l,r]$. Further, if $x \in [l,r]$, then $d(x,wx) \geq k$ where $k = \min\{d(r,gl), d(r,hl)\}$, so $k > 0$. If w is cyclically reduced (i.e. $u_1 \neq u_n^{-1}$), then $w^{-1}[l,r] \subseteq X_{u_n^{-1}}$, so $wr \in X_u$, $w^{-1}r \in X_v$ with $u \neq v$. Hence $r \in [w^{-1}r, wr]$ unless $\{u,v\} = \{g^{-1}, h^{-1}\}$, when $[w^{-1}r, r] \cap [r, wr] = [l,r]$. It follows that w is hyperbolic and by Theorem 1.4 either l or $r \in A_w$. For if w is elliptic, it would have to have a fixed point in $[l,r]$, by Lemma 1.1, and similarly w^2 is not elliptic, so w is not an inversion. Thus $\ell(w) \geq k$. Since any element of G is conjugate to a cyclically reduced element and ℓ is invariant under conjugation (Lemma 1.7), the result follows. □

Remark. Suppose G is a group acting on a Λ-tree (X,d) and g, h are hyperbolic elements of G such that $\Delta(g,h) \geq \ell(g) + \ell(h)$, where ℓ denotes hyperbolic length. Then the commutator $[g,h]$ is an elliptic element.

For then $A_g \cap A_h$ contains a segment of length at least $\ell(g) + \ell(h)$, say $[p,q]$ where $p \leq_g q$. If g, h meet coherently, $g^{-1}h^{-1}q \in [p,q]$, at distance $\ell(g) + \ell(h)$ from q, and so $ghg^{-1}h^{-1}q = q$, because g and h act as translations on their axes. If g and h do not meet coherently, then $g^{-1}h^{-1}(gp) \in [p,q]$, at distance $\ell(h)$ from p, so $hg^{-1}h^{-1}(gp) = p$, hence gp is a fixed point of $ghg^{-1}h^{-1}$.

Proposition 3.7. *Let G be a group acting on a Λ-tree (X,d). The following are equivalent.*

(1) *The action is irreducible.*
(2) *There exist hyperbolic elements g, $h \in G$ such that $\ell([g,h]) \neq 0$.*
(3) *There exist hyperbolic elements g, $h \in G$ such that $A_g \cap A_h$ is a segment of length less than $\ell(g) + \ell(h)$.*

(4) *G contains a free subgroup of rank 2 which acts freely, without inversions and properly discontinuously on* X.

Proof. By Propositions 2.7 and 2.9, (1) implies (2). Assume (2) and let g, h be hyperbolic elements of G such that $\ell([g,h]) \neq 0$. Suppose $A_g \cap A_h \neq \emptyset$. By the remark preceding the theorem, $\Delta(g,h) < \ell(g) + \ell(h)$, hence $A_g \cap A_h$ is a segment of length less than $\ell(g) + \ell(h)$. If $A_g \cap A_h = \emptyset$, then by (2.2), $A_{gh} \cap A_h$ is a segment and gh is hyperbolic. Also, $[gh, h] = [g, h]$, so again this segment is of length less than $\ell(gh) + \ell(h)$, hence (3) holds.

Assume (3), and take hyperbolic g, $h \in G$ with $A_g \cap A_h \neq \emptyset$ such that $\Delta = \Delta(g,h) < \ell(g) + \ell(h)$. We may assume g, h meet coherently and $\ell(h) \leq \ell(g)$. By Lemma 3.1, we have $\Delta(gh, g) = \Delta + \ell(g)$ and $\ell(gh) = \ell(g) + \ell(h)$. Thus $\Delta(gh, g) < 2\ell(g) + \ell(h) \leq 3\ell(g) = \ell(g^3)$ and $2\ell(g) + \ell(h) < 2\ell(gh) = \ell((gh)^2)$, using Lemma 1.7. Also by Lemma 1.7, g^3 and $(gh)^2$ have the same axes as g and gh, so by Lemma 3.6, g^3 and $(gh)^2$ freely generate a free group of rank 2 which acts freely, without inversions and properly discontinuously on X, and (3) implies (4).

Finally, if g, h freely generate a free group acting freely, without inversions and properly discontinuously on X, then $[g, h] \neq 1$, so g, h and $[g, h]$ all act as hyperbolic isometries. By Propositions 2.7 and 2.9, the action is irreducible. □

Our next result is an analogue of Lemma 3.6 in the case that $A_g \cap A_h = \emptyset$, which will be needed later. The proof is similar, with the bridge $[p, q]$ between A_g and A_h playing a role similar to the segment $[l, r] = A_g \cap A_h$ in 3.6, and full details will not be given.

Lemma 3.8. *Suppose* G *is a group acting on a* Λ*-tree* (X, d), *and* G *is generated by* g *and* h, *where* g *and* h *are hyperbolic elements such that* $A_g \cap A_h = \emptyset$. *Then* G *is a free group with basis* $\{g, h\}$, *and there exists* $k > 0$, $k \in \Lambda$, *such that, for all* $a \in G$, *we have* $\ell(a) \geq k$. *Consequently, the action is free, without inversions and properly discontinuous.*

Proof. Let $[p, q]$ be the bridge between A_g and A_h, with $p \in A_g$, $q \in A_h$. The situation is illustrated by the picture after the proof of Lemma 2.2. We define the following subsets of X.

$$\begin{aligned} X_g &= \{x \in X \mid [x, p] = [x, gp, p]\} \\ X_h &= \{x \in X \mid [x, q] = [x, hq, q]\} \\ X_{g^{-1}} &= \{x \in X \mid [x, p] = [x, g^{-1}p, p]\} \\ X_{h^{-1}} &= \{x \in X \mid [x, q] = [x, h^{-1}q, q]\}. \end{aligned}$$

These are clearly non-empty subsets of X. We leave it to readers to verify that they are pairwise disjoint, and that the following hold.

$$g(X_g \cup X_h \cup X_{h^{-1}}) \subseteq X_g$$
$$h(X_h \cup X_g \cup X_{g^{-1}}) \subseteq X_h$$
$$g^{-1}(X_{g^{-1}} \cup X_h \cup X_{h^{-1}}) \subseteq X_{g^{-1}}$$
$$h^{-1}(X_{h^{-1}} \cup X_g \cup X_{g^{-1}}) \subseteq X_{h^{-1}}.$$

The proof now proceeds in a similar way to that of 3.6, with l replaced by q, r replaced by p and $k = \min\{d(p, gp), d(q, hq)\}$. Again, all details are left to the reader. □

The next lemma provides an elaboration of the equivalence of (2) and (3) in Proposition 3.7.

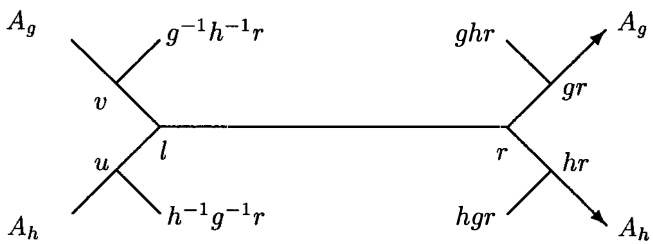
Lemma 3.9. *Let g and h be isometries of a Λ-tree (X, d), and let ℓ denote hyperbolic length.*

(1) *If g and h are hyperbolic isometries which meet coherently and $\Delta(g, h) < \ell(g) + \ell(h)$, then*

$$\ell(ghg^{-1}h^{-1}) = 2(\ell(g) + \ell(h) - \Delta(g, h)).$$

(2) *Suppose g and h commute. If g, h are hyperbolic, then $A_g = A_h$. If g is elliptic and h is hyperbolic, then $A_h \subseteq A_g$.*

Proof. (1) First note that $A_g \cap A_h$ has both a left and right-hand endpoint. For suppose there is no right-hand endpoint, and let $p \in A_g \cap A_h$. Then $hp \in A_h$ and $p \leq_h hp$, so $hp \in A_g \cap A_h$ by the discussion at the start of this section (A_g and A_h have a common right-hand end). Similarly $ghp \in A_g \cap A_h$; but then $d(p, ghp) = \ell(g) + \ell(h) > \Delta(g, h)$ since g and h act as translations in the same direction on $A_g \cap A_h$, which contradicts the definition of Δ. Similarly there is a left-hand endpoint l. The situation is illustrated by the diagram below.



Here $[u, h^{-1}g^{-1}r]$ is the bridge between $h^{-1}g^{-1}r$ and A_h. If $\ell(g) \leq \Delta(g,h)$, $g^{-1}r \in [l,r] \subseteq A_h$, so $h^{-1}g^{-1}r \in A_h$ and u coincides with $h^{-1}g^{-1}r$, so $u <_h l$ since $\Delta(g,h) < \ell(g) + \ell(h)$. If $\ell(g) > \Delta(g,h)$, then $g^{-1}r <_g l$, so $[g^{-1}r, l]$ is the bridge between $g^{-1}r$ and A_h (by definition of l), and applying h^{-1}, we see that $u = h^{-1}l$ and again $u <_h l$. In either case, $u \neq l$, and $[h^{-1}g^{-1}r, l] = [h^{-1}g^{-1}r, u, l]$ since $l \in A_h$. Similarly, $[v, g^{-1}h^{-1}r]$ is the bridge between $g^{-1}h^{-1}r$ and A_g and $v \neq l$. By definition of l, $[u,l] \cap [v,l] = \{l\}$. Further, $r <_h hr$, so $[r, hr]$ is the bridge between hr and A_g (by definition of r), hence $[gr, ghr]$ is the bridge between ghr and A_g, so $[ghr, l] = [ghr, gr, l]$. Since $l \leq_g r <_g gr$, $l \neq gr$ and $r \in [l, gr]$; by Lemma 2.1.5 we may therefore splice the segments $[ghr, gr, l]$ and $[h^{-1}g^{-1}r, u, l]$ to obtain a piecewise segment $[h^{-1}g^{-1}r, u, l, r, gr, ghr]$. Hence $r \in A_{gh}$, this segment is contained in A_{gh} and $\ell(gh) = d(r, ghr) = d(r, gr) + d(gr, ghr) = \ell(g) + \ell(h)$ (although this formula follows from 3.1). Similarly, there is a piecewise segment $[g^{-1}h^{-1}r, v, l, r, hr, hgr]$ in A_{hg} and $\ell(hg) = \ell(g) + \ell(h)$. Hence gh, hg are hyperbolic and A_{gh}, A_{hg} are linear. Also, $[r, gr] \cap [r, hr] = \{r\}$ by definition of r. It follows that gh, gh, g and h all meet coherently and $A_{gh} \cap A_{hg} = [l, r]$. For suppose $w \in A_{gh} \cap A_{hg}$. If $w <_{hg} l$, then $[w, l] \subseteq A_{gh} \cap A_{hg}$, and replacing w by $\max\{w, u\}$ relative to $\leq_{hg}$, we can assume $w \in [u, l]$, so $w <_h l$ and $[w, l] \cap A_g = \{l\}$ by definition of l. Then $[l, w]$, $[l, v]$ and $[l, gr]$ define three different directions at l in A_{gh}, because in A_g, $v <_g l <_g gr$, contradicting the linearity of A_{gh}. If $r <_{hg} w$ we obtain a similar contradiction, hence $w \in [l, r]$. The formula for $\ell(ghg^{-1}h^{-1})$ now follows from Lemma 3.3.

(2) Suppose g, h are not inversions. If $A_g \cap A_h = \emptyset$, then referring to the proof of Lemma 2.2, it is easily seen that $ghp \neq hgp$ (where p is the endpoint of the bridge between A_g and A_h lying in A_g). Thus $A_g \cap A_h \neq \emptyset$, and replacing h by h^{-1} if necessary, we can assume that g and h meet coherently. Suppose g, h are both hyperbolic. If $A_g \cap A_h$ has a right-hand endpoint r, then it is easy to see that $ghr \neq hgr$ (see the proof of (1)), hence $A_g \cap A_h$ has no right-hand endpoint. Similarly, $A_g \cap A_h$ has no left-hand endpoint, so by the remarks at the beginning of this section, $A_g = A_h$.

If g is elliptic and h is hyperbolic, then g fixes some point on A_h, say w. Then $h(gw) = g(hw) = gw$, so g fixes at least two points on A_h. Since A_h is linear, g fixes the whole of A_h pointwise, by Lemma 1.2.1, so $A_h \subseteq A_g$. □

Remark. If g, h are hyperbolic isometries of a Λ-tree with $A_g \cap A_h = \emptyset$, closer inspection of the proof of Lemma 2.2 shows that gh, hg are hyperbolic isometries which meet coherently in $[p, q]$, the bridge between A_g and A_h. It follows from Lemmas 2.2 and 3.3 that $\ell(ghg^{-1}h^{-1}) = \ell(gh) + \ell(hg) - 2d(p, q) = 2\ell(gh) - 2d(A_g, A_h) = 2\ell(g) + 2\ell(h) + 2d(A_g, A_h)$. (Since gh and hg are conjugate, $\ell(gh) = \ell(hg)$ by Lemma 1.7.)

A consequence of Lemma 3.9 is the following, which in the case $\Lambda = \mathbb{R}$ is equivalent to one of Chapman and Wilkens (Proposition 2.6 of [29]), in view of the connection between Lyndon length functions and actions on Λ-trees (Theorem 2.4.6), and Lemma 1.8. (According to Lemma 1.8, if g is an element of a group acting on a Λ-tree, and x is a point of the Λ-tree, then g is hyperbolic if and only if $L_x(g^2) > L_x(g)$.)

Corollary 3.10. *Suppose G is a group acting on a Λ-tree, and g, h, k are hyperbolic elements, such that g, h commute, and h, k commute. Then $[g,k]$ is an elliptic element.*

Proof. By Lemma 3.9, $A_g = A_h = A_k$, and so $A_g \cap A_k$ contains a segment of length at least $\ell(g) + \ell(k)$, where ℓ denotes hyperbolic length (e.g. $[p, gkp]$, where p is any point of A_k, if g, k meet coherently, or $[p, g^{-1}kp]$ if not). By the remark preceding Prop. 3.7, $[g,k]$ is elliptic. □

4. Minimal Actions

An action of a group G on a Λ-tree (X,d) is called *minimal* if there is no proper (and non-empty) G-invariant subtree. A G-invariant subtree Y of X is called minimal if the action of G restricted to Y is minimal. We shall study the following questions. Given an action of G on (X,d), is there a minimal invariant subtree, and is it unique (absolutely, or up to isomorphism)? Given a minimal action of G, to what extent is the action determined (up to G-equivariant isomorphism) by the hyperbolic length function $\ell : G \to \Lambda$? This will be significant when we consider spaces of actions of groups on $\mathbb{R}$-trees in §4.6. The exposition which follows is based on [37].

We begin by summarising the classification of actions given in §2. The initial classification is into abelian, dihedral and irreducible actions. Given an action of a group G on a Λ-tree, put $N = \{g \in G | \ \ell(g) = 0\}$, the set of elements of G which do not act as hyperbolic isometries. The abelian actions can be further classified into five types. Put $B = \bigcap_{g\in G} A_g$. An abelian action is of exactly one of the following types.

(1) Some element of G acts as an inversion. In this case $N = G$.
(2) $B \neq \emptyset$ and $N = G$. Then B is the set of fixed points of G.
(3) $B \neq \emptyset$ and $N \neq G$. Then B is a G-invariant linear subtree on which G acts as translations. Also, no point of X is fixed by G.
(4) End type.
(5) Cut type.

Types (2) and (3) are the actions of core type. They are the abelian actions for which $B \neq \emptyset$. In type (4) there is a unique end of X fixed by G, and in type (5) X is the disjoint union of two G-invariant subtrees, and the action of G on both subtrees is of end type. An action is called *trivial* if $N = G$, that is, there are no hyperbolic isometries. Such an action is necessarily abelian. In

[42], an action is called semisimple if it is non-abelian or abelian of core type, and actions of type (3) are called shifts.

The following observation will be useful in some later proofs.

Remark If a group G acts on a Λ-tree (X,d) and $x \in X$, then the subtree spanned by the orbit Gx is $\bigcup_{g\in G}[x,gx]$, as noted in the definition preceding Lemma 2.1.8, and it is clearly a G-invariant subtree.

The first result is due to Alperin and Bass ([2; 7.13]). For an action of a group G on a Λ-tree (X,d) we denote the hyperbolic length function by ℓ_X.

Theorem 4.1. *Let G be a group with a non-abelian action without inversions on a Λ-tree (X,d). Then there is a unique minimal G-invariant subtree $X_{\min}$ of X, and $\ell_{X_{\min}} = \ell_X$. If Y is another Λ-tree on which G acts without inversions, with $\ell_X = \ell_Y$, then there is a unique G-isomorphism $X_{\min} \to Y_{\min}$.*

Proof. By Corollary 2.4, we can choose g, $h \in G$ such that $A_g \cap A_h = \emptyset$. Let $[x_g, x_h]$ be the bridge between A_g and A_h and define $X_{\min}$ to be the subtree spanned by the orbit Gx_g. This is plainly G-invariant. By Corollary 1.6, A_g and A_h meet every G-invariant subtree of X, so every such subtree contains the bridge between them, hence contains $X_{\min}$. Thus $X_{\min}$ is the unique minimal invariant subtree of X, and $\ell_X = \ell_{X_{\min}}$ by Lemma 1.7.

By Lemma 1.1 and Theorem 1.4, the Lyndon length function L_{x_g} is given by $L_{x_g}(k) = \ell_X(k) + 2d(x_g, A_k)$, and by Lemma 2.1.10,

$$d(x_g, A_k) = \max\{d(A_k, A_g), d(A_k, A_h) - d(A_g, A_h)\}.$$

Using Corollary 2.4, this can be expressed in terms of ℓ_X applied to g, h, k and certain of their products, so ℓ_X determines L_{x_g}. Let (X', d') be the Λ-tree corresponding to L_{x_g} given by Theorem 2.4.6, so G acts on X' and there are a point $x \in X'$ such that $L_x = L_{x_g}$ and a G-equivariant isometry $X' \to X_{\min}$, which is an isomorphism by minimality of $X_{\min}$. Thus if G acts on Y and $\ell_Y = \ell_X$, then $Y_{\min}$ and $X_{\min}$ are G-isomorphic, both being isomorphic to X' by Theorem 2.4.6. (Note that the axes of g and h in Y do not intersect, by Corollary 2.4, so if $[y_g, y_h]$ is the bridge between them, then $Y_{\min}$ is spanned by Gy_g, and $L_{y_g} = L_{x_g}$ since both can be expressed by the same formula in terms of $\ell_Y = \ell_X$.)

To show uniqueness of the isomorphism, two isomorphisms differ by a G-automorphism ϕ of $X_{\min}$, and ϕ preserves $A_k \cap X_{\min}$, for each $k \in G$ (from the definition of A_k or by Cor.1.5), so ϕ fixes x_g. Thus the fixed point set $X_{\min}^{\phi}$ is non-empty, so is a subtree, and it is fixed by G, so is all of $X_{\min}$ by minimality, i.e. $\phi = 1$. □

We refer to (7.14) in [2], for an example of a group acting on two Λ-trees with the the same non-abelian hyperbolic length function, where one action is with inversions and the other is not.

The last part of the next result is essentially due to Culler and Morgan ([42; 3.1]), who stated it only for $\mathbb{R}$-trees. The proof gives an alternative description of the minimal invariant subtree for a non-trivial action when Λ is archimedean.

Theorem 4.2. *Let G be a group acting non-trivially on a Λ-tree (X, d).*
(a) *If the action is abelian of core type there is a minimal invariant subtree.*
(b) *If the action is abelian of core type and minimal, and G acts on a Λ-tree (X', d') with minimal abelian action of core type, with the same hyperbolic length function as that for the action on X, then X and X' are G-equivariantly isomorphic.*
(c) *If Λ is archimedean then there is a unique minimal invariant subtree $X_{\min}$, and $\ell_X = \ell_{X_{\min}}$.*

Proof. (a) Let $B = \bigcap_{g \in G} A_g$, so B is a linear G-invariant subtree on which G acts as a non-trivial group of translations. Let $y \in B$, and let $Y = \bigcup_{g \in G}[y, gy]$, the subtree spanned by the orbit Gy. Clearly Y is G-invariant. Suppose Z is a G-invariant subtree of X and $Z \subseteq Y$, and take $z \in Z$. Then $z \in [y, gy]$ for some $g \in G$. Thus $y \leq_g z \leq_g gy$ in A_g, so $g^{-1}z \leq_g y \leq_g z$, hence $y \in [z, g^{-1}z] \subseteq Z$. It follows that $Gy \subseteq Z$ and so $Y \subseteq Z$, i.e. $Z = Y$, and Y is a minimal G-invariant subtree.

(b) It follows from the proof of (a) that $X = A_g$ for every $g \in G$. Consequently, if $x \in X$ then $\ell_X = L_x$, the Lyndon length function based at x. As in the proof of 4.1, X is G-equivariantly isomorphic to Y, where Y is the tree given by Theorem 2.4.6 using the Lyndon length function L_x. Similarly X' is G-equivariantly isomorphic to Y, so to X.

(c) We define $X_{\min}$ to be $\bigcup\{A_g \mid g \in G \text{ and } \ell(g) \neq 0\}$. If $A_g \cap A_h = \emptyset$, then by Lemma 2.2, $\ell(gh) > 0$ and A_{gh} intersects A_g and A_h, and it follows that $X_{\min}$ is a subtree of X, which is G-invariant by Lemma 1.7. If Z is an invariant subtree, and $g \in G$ with $\ell(g) \neq 0$, then by Corollary 1.6, A_g meets Z, say $p \in A_g \cap Z$. If $q \in A_g$, $d(p, q) \leq n\ell(g)$ for some positive integer n, since Λ is archimedean, so either $q \in [p, g^n p]$ or $q \in [g^{-n}p, p]$ since A_g is linear and g acts as a translation of amplitude $\ell(g)$ on it. In any case $q \in Z$, so $A_g \subseteq Z$. Hence $X_{\min} \subseteq Z$, and $X_{\min}$ is the unique minimal invariant subtree. Again by Lemma 1.7, $\ell_X = \ell_{X_{\min}}$. □

In the situation of Theorem 4.2(a), there may be more than one minimal G-invariant subtree. A simple example is $\Lambda = \mathbb{Z} \oplus \mathbb{Z}$ with the lexicographic ordering, $X = \Lambda$, with G the infinite cyclic group generated by the translation $x \mapsto a + x$, where $a = (0, 1)$. For all $n \in \mathbb{Z}$, $\{n\} \times \mathbb{Z}$ is a minimal G-invariant subtree.

In Theorem 4.2(b), the G-isomorphism $X \to X'$ need not be unique. For if $\alpha : X \to X'$ is a G-isomorphism, so is $x \mapsto g\alpha(x)$, where $g \in G$ acts as a non-trivial translation on X'.

Remark. In the situation of Theorem 4.2(c), if $g \in G$, $\ell(g) \neq 0$ and $p \in A_g$, we saw in the proof of 1.4 that $A_p \subseteq A_g$, where $A_p = \bigcup_{n\in\mathbb{Z}}[g^n p, g^{n+1}p]$. Since A_g embeds in Λ and $d(p, g^n p) = |n|\ell(g)$, it follows since Λ is archimedean that $A_p = A_g$ is metrically isomorphic to Λ (cf. [42: 1.3]).

In the case of a trivial action on a Λ-tree, if there is a fixed point, such a point will obviously be a minimal invariant subtree, but will not necessarily be unique. For abelian actions with inversions there is the following result.

Proposition 4.3. *Let G be a group with an abelian action with inversions on a Λ-tree (X, d), and suppose there is a minimal G-invariant subtree. Then Λ has a least positive element denoted by* 1, *and the minimal invariant subtree is unique and consists of two points at distance* 1 *apart.*

Proof. By Theorem 2.6(1) there is a unique fixed point p for the induced action on the $\frac{1}{2}\Lambda$-tree $X' = \frac{1}{2}\Lambda \otimes_\Lambda X$, whose metric we denote by d'. We use the notation $[x, y]'$ to denote the segment joining x, y in X', to distinguish from segments in X. Let Y be a minimal invariant subtree of X, and choose $x \in Y$. Then $Y = \bigcup_{g\in G}[x, gx]$, since the right-hand union is a G-invariant subtree by the remark preceding Theorem 4.1.

By Lemma 1.1, p is the midpoint of $[x, gx]'$, for all $g \in G$. Let $g \in G$, and suppose $y \in [x, gx]$. Then $y \in Y$, so also $Y = \bigcup_{h\in G}[y, hy]$, and p is the midpoint of $[y, hy]'$, for all $h \in G$. If z, $w \in Y$, say $z \in [y, hy]$, $w \in [y, h'y]$ (h, $h' \in G$), then $d(z, w) = d'(z, w) \leq d'(z, p) + d'(w, p) \leq 2d'(y, p)$. Since x, $gx \in Y$, $d(x, gx) = 2d'(x, p) \leq 2d'(y, p)$, that is, $d'(x, p) \leq d'(y, p)$, and since $y \in [x, gx]$, $d'(y, p) = d'(x, p) = d'(gx, p)$. It follows that $[x, gx] = \{x, gx\}$. Choosing g so that $gx \neq x$ (e.g. an inversion), it follows that $d(x, gx)$ is the least positive element 1 of Λ. Now for all $h \in G$, $Y(x, gx, hx) \in [x, gx] \cap [x, hx] \cap [gx, hx] = \{x, gx\} \cap \{x, hx\} \cap \{gx, hx\}$ by Lemma 2.1.2(3). It follows that either $hx = x$ or $hx = gx$. Thus the G-orbit of x is $\{x, gx\}$, which is the subtree $[x, gx]$ of Y. Hence $Y = \{x, gx\}$.

Suppose Z is a minimal G-invariant subtree of X. The argument just given shows that $Z = \{z, hz\}$ for some $z \in Z$, $h \in G$, with $d(z, hz) = 1$, and p is the midpoint of $[z, hz]'$. Then $Y(x, z, hz) \in [x, z] \cap [x, hz] \cap \{z, hz\}$, so either $z \in [x, hz]$ or $hz \in [x, z]$. Further, $d(x, z) \leq d'(x, p) + d'(z, p) = 1$, and similarly $d(x, hz) \leq 1$. Thus either $z = x$ or $hz = x$, and it follows that $Z = Y$. □

The simplest example of the situation in Proposition 4.3 is when $\Lambda = \mathbb{Z}$, and X has two points x, y with $d(x, y) = 1$. The cyclic group of order 2 acts by interchanging x and y, and the action is minimal. If F is a free group of rank 2 with basis a, b, we obtain two abelian actions of F on X with inversions by letting a interchange x and y and letting b fix both points, then by letting b

interchange x, y while a fixes both of them. Note that there is no F-equivariant isomorphism from X to X for these two actions.

Another example is obtained by letting $\Lambda = \mathbb{Z} \oplus \mathbb{Z}$ with the lexicographic ordering, and taking $X = \Lambda$. The cyclic group of order 2 acts by means of the reflection $x \mapsto a - x$, where $a = (1, 0)$. This action is abelian with inversions, and there is no minimal invariant subtree. For if Y is an invariant subtree, let $y \in Y$. Since $a - y \in Y$, we can find (m, n), $(1 - m, -n) \in Y$ with $m \leq 0$. Then $[(m,n),(1-m,-n)]_\Lambda \subseteq Y$, and so $[(m,n+1),(1-m,-n-1)]_\Lambda \subsetneqq Y$. But $[(m,n+1),(1-m,-n-1)]_\Lambda$ is an invariant subtree, so Y is not minimal. Note that Λ has a least positive element, ruling out a possible converse to Proposition 4.3. There remains the question of what happens in the case of actions of end and of cut type. We shall concentrate on actions of end type, then use the results to study actions of cut type. Our first result on actions of end type is a generalisation of an example in [42: 3.10].

Proposition 4.4. *Let G be a group with an action on a Λ-tree (X, d) which is trivial of end type. Then there is no minimal invariant subtree.*

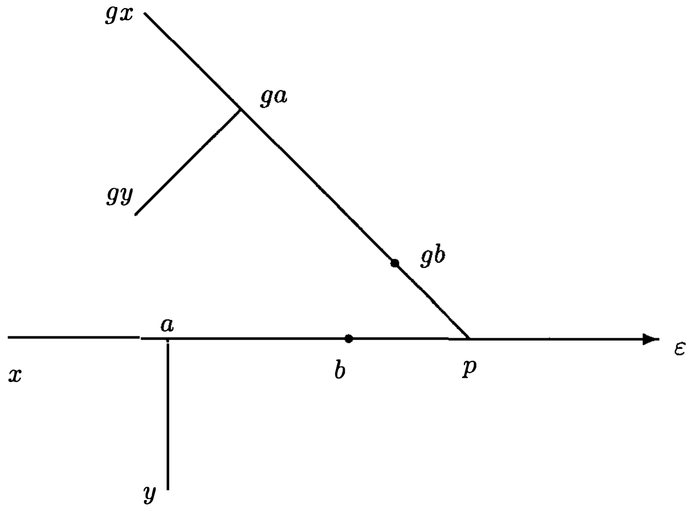
Proof. By Lemma 1.9, all elements of G are elliptic. Choose $x \in X$, and let ε be the fixed end. For $z \in [x, \varepsilon\rangle$, put $E_z = \bigcup_{g \in G} g[z, \varepsilon\rangle$. For g, $h \in G$, $g[z,\varepsilon\rangle \cap h[z,\varepsilon\rangle = [gz,\varepsilon\rangle \cap [hz,\varepsilon\rangle = [w,\varepsilon\rangle$ for some w, by the definition of ends. If $u \in [gz,\varepsilon\rangle$, $v \in [hz,\varepsilon\rangle$ then $[u,v] \subseteq [u,w] \cup [v,w] \subseteq E_z$, by Cor. 2.1.3, so E_z is a subtree of X, and it is plainly G-invariant. Further, if $u \in [x,\varepsilon\rangle$ and $z < u$ (using the ordering in §3 of Chapter 2 on $[x,\varepsilon\rangle$), then E_u is a proper subtree of E_z. For $[u,\varepsilon\rangle \subseteq [z,\varepsilon\rangle$, so E_u is a subtree of E_z. We show it is a proper subtree by showing that $z \notin E_u$. If $g \in G$, $[u,\varepsilon\rangle \cap [gu,\varepsilon\rangle = [p,\varepsilon\rangle$ for some p. By Lemma 1.9, ε belongs to A_g and we choose $r \in A_g \cap [p,\varepsilon\rangle$, so $p \in [u,r] \cap [gu,r]$ and $gr = r$. Thus $d(gu,p) = d(gu,r) - d(r,p) = d(u,r) - d(r,p) = d(u,p)$, and $p \in [u,gu]$ since $[u,p] \cap [gu,p] = \{p\}$, so p is the midpoint of $[u,gu]$. It follows by Lemma 1.1 that $p \in A_g$ and $[u,p] \cap A_g = \{p\}$. Since ε belongs to A_g, $A_g \cap [u,\varepsilon\rangle = [p,\varepsilon\rangle$.

Now if $q \in [gu,\varepsilon\rangle$, either $q \in [p,\varepsilon\rangle$, when $z < u \leq p \leq q$, or $q \in [p,gu]$, when $d(p,q) \leq d(p,gu) = d(p,u) < d(p,z)$. Thus $z \notin [gu,\varepsilon\rangle$, and this is true for any $g \in G$, so $z \notin E_u$.

Suppose Y is an invariant subtree, take $y \in Y$, and let $[y,a]$ be the bridge between y and $[x,\varepsilon\rangle$. We claim that $[a,\varepsilon\rangle \subseteq Y$. Assume $b \in [a,\varepsilon\rangle$. Since $\bigcap_{g \in G} A_g = \emptyset$, there exists $g \in G$ such that $gb \neq b$. As in the previous part of the argument, $[x,\varepsilon\rangle \cap [gx,\varepsilon\rangle = A_g \cap [x,\varepsilon\rangle = [p,\varepsilon\rangle$ for some p, and it follows that $A_g \cap [x,p] = \{p\}$. Now x, a, b, p occur in order in the linear tree $[x,\varepsilon\rangle$, so $[x,p] = [x,a,b,p]$, hence $A_g \cap [b,p] = \{p\}$, and by Lemma 1.1, $[b,gb] = [b,p,gb]$. Also, $[y,p] = [y,a,p] = [y,a,b,p] = [y,b,p]$ by definition of a, and applying g, $[gy,p] = [gy,gb,p]$. It follows by Lemma 2.1.5 that $[y,gy] = [y,b,p,gb,gy]$, so $b \in [y,gy] \subseteq Y$.

Since Y is G-invariant, E_a is a subtree of Y. Since ε is an open end, there exists $b \in [x, \varepsilon\rangle$ with $a < b$. Then E_b is a proper G-invariant subtree of Y, so Y is not minimal. □

The situation in the last part of the proof is illustrated in the following picture.



Before stating the next result we need to introduce some notation. Let Λ be an ordered abelian group, and let $x \in \Lambda$, $x \geq 0$. Let C be the union of all the convex subgroups of Λ which do not contain x. Then C is a convex subgroup of Λ, since the convex subgroups of Λ are linearly ordered by inclusion. Let D be the intersection of all convex subgroups of Λ containing x, so D is a convex subgroup of Λ, and again because the the convex subgroups are linearly ordered by inclusion, $C \leq D$. The pair (C, D) is called the *convex jump* determined by x. There are no convex subgroups strictly between C and D, so the quotient ordered abelian group D/C is archimedean. If $y \in \Lambda$, $y \geq 0$, we define $y << x$ to mean $y \in C$. It is easy to see that this is equivalent to the requirement that $ny < x$ for all $n \in \mathbb{Z}$. Note that if x, $y << z$ then $x + y << z$, and if $0 \leq x \leq y << z$ then $x << z$.

Theorem 4.5. *Let G be a finitely generated group acting on a Λ-tree (X, d) with action of end type. Then the action is non-trivial, and there is a unique minimal invariant subtree.*

Proof. Let ε be the fixed end, and let G be generated by the finite set $\{g_1, \ldots, g_n\}$, which we assume is closed under taking inverses. If $u_i \in A_{g_i}$ then $[u_i, \varepsilon\rangle \subseteq A_{g_i}$ by Lemma 1.9, so $\bigcap_{i=1}^{n} [u_i, \varepsilon\rangle \subseteq \bigcap_{i=1}^{n} A_{g_i}$, and $\bigcap_{i=1}^{n} [u_i, \varepsilon\rangle = [u, \varepsilon\rangle$ for some $u \in X$, by the definition of ends. Let $v \in [u, \varepsilon\rangle$.

Suppose (for a contradiction) that $d(u, g_i^n u) < d(u, v)$ for all $n \in \mathbb{Z}$ and $1 \le i \le n$. Then $\ell(g_i) << d(u, v)$ for $1 \le i \le n$ (since $d(u, g_i^n u) = |n|\ell(g_i)$). Suppose $g \in G$ and write $g = g_{i_1} \cdots g_{i_k}$, where $1 \le i_j \le n$. Since the action is abelian, $\ell(g) \le k \max_{1 \le i \le n}\{\ell(g_i)\} << d(u, v)$. That is, for all $g \in G$, $\ell(g) << d(u, v)$. We claim that, for all $g \in G$, there is a point $p_g \in [u, \varepsilon\rangle \cap A_g$ with $d(u, p_g) << d(u, v)$. We use induction on k, where g is written as a product of the generators as above. For $k = 0$ or 1 we may take $p_g = u$. To finish the induction, it suffices to show that, if it is true for g then it is true for gg_i, where $1 \le i \le n$. Suppose then, that $p_g \in [u, \varepsilon\rangle \cap A_g$ and $d(u, p_g) << d(u, v)$. Let $[u, w]$ be the bridge between u and A_g; then $A_g \cap [u, \varepsilon\rangle = [w, \varepsilon\rangle$ by Lemma 1.9, and since $[u, p_g] = [u, w, p_g]$, $d(u, w) \le d(u, p_g) << d(u, v)$. Now

$$\begin{aligned} d(w, gg_i w) &\le d(w, gw) + d(gw, gg_i w) \\ &= \ell(g) + d(w, g_i w) \\ &= \ell(g) + \ell(g_i) << d(u, v) \end{aligned}$$

since $w \in [u, \varepsilon\rangle \subseteq A_{g_i}$.

By Lemma 1.1 and Theorem 1.4, $[w, gg_i w]$ contains a point of A_{gg_i}, say q, and again by Lemma 1.9, $[q, \varepsilon\rangle \subseteq A_{gg_i}$. Further, $[q, \varepsilon\rangle \cap [u, \varepsilon\rangle = [p, \varepsilon\rangle$ for some p, by definition of ends, and $p \in [u, q]$ since $[u, p] \cap [q, p] = \{p\}$. Therefore,

$$\begin{aligned} d(p, u) \le d(u, q) &\le d(u, w) + d(w, q) \\ &\le d(u, w) + d(w, gg_i w) << d(u, v). \end{aligned}$$

We may therefore take $p_{gg_i} = p$, completing the induction. It follows that $v \in [p_g, \varepsilon\rangle$ for all $g \in G$, and invoking Lemma 1.9 yet again, $[p_g, \varepsilon\rangle \subseteq A_g$. But then $v \in \bigcap_{g \in G} A_g$ and this intersection is empty since the action is of end type, a contradiction.

This shows that, for $v \in [u, \varepsilon\rangle$, there exists an integer n and an integer i with $1 \le i \le n$ such that $d(u, v) \le d(u, g_i^n u)$. Since u, $v \in A_{g_i}$ it follows that $u \in [g_i^{\pm n} v, v]$. (This is obvious if $v = u$, otherwise $d(u, g_i^n u) = |n|\ell(g_i) > 0$, so g_i acts as a translation of amplitude $\ell(g_i)$ on the linear subtree A_{g_i}.) Since ε is an open end, it follows that there exists $v \in [u, \varepsilon\rangle$ with $v \ne u$, hence the action of G is non-trivial.

Now define $Y = \bigcup_{g \in G} g[u, \varepsilon\rangle = \bigcup_{g \in G}[gu, \varepsilon\rangle$. Then (as in the proof of Prop. 4.4) Y is a G-invariant subtree of X. Suppose Z is a G-invariant subtree of X. Choose $x \in Z$; then $[u, \varepsilon\rangle \cap [x, \varepsilon\rangle = [s, \varepsilon\rangle$ for some $s \in X$. Since the action is of end type, there exists $g \in G$ such that $s \notin A_g$. Since ε belongs to A_g, $[x, \varepsilon\rangle \cap A_g = [z, \varepsilon\rangle$, where $[x, z]$ is the bridge between x and A_g, and $[x, \varepsilon\rangle = [x, z] \cup [z, \varepsilon\rangle$. Thus $s \in [x, z]$ and $z \in [s, \varepsilon\rangle \subseteq [u, \varepsilon\rangle$. By Lemma 1.1 and Theorem 1.4, $z \in [x, gx]$, so $z \in Z \cap [u, \varepsilon\rangle$.

By the above, $u \in [g_i^m z, z]$ for some integer m and some i with $1 \le i \le n$, so $u \in Z$. If $v \in [u, \varepsilon\rangle$, again by the above there exist n and i such that

$d(u,v) \leq d(u, g_i^n u)$. It follows as above that $v \in [u, g_i^{\pm n} u]$, hence $v \in Z$. Thus $[u, \varepsilon\rangle \subseteq Z$, hence $Y \subseteq Z$ and Y is the unique minimal G-invariant subtree of X. □

One can also ask, given two non-trivial minimal actions of end type of a group G on Λ-trees with the same hyperbolic length function, are the two trees necessarily isomorphic (G-equivariantly or otherwise). We shall give an example to show that the answer is no. This will involve the idea of an HNN-extension of groups; readers unfamiliar with this are referred to Chapter 1 in [39] or Chapter IV in [95]. If a group G acts on a Λ-tree (X, d), with action of end type and fixed end ε, and $\delta : X \to \Lambda$ is an end map towards ε, then by Prop. 2.5 there is a homomorphism $\tau : G \to \Lambda$, not depending on δ, such that for all $x \in X$ and $g \in G$, $\delta(gx) = \delta(x) + \tau(g)$. Further, the hyperbolic length function ℓ for the action is given by $\ell(g) = |\tau(g)|$. We call τ the homomorphism associated to the action.

Now if G is a properly ascending HNN-extension, that is, has a presentation $G = \langle\ t,\ A \mid tAt^{-1} = B\ \rangle$, where B is a proper subgroup of A, then G has an action of end type on a $\mathbb{Z}$-tree with associated homomorphism the canonical map $G \to \mathbb{Z}$ sending t to 1 and all elements of A to 0 (see Example 1, §1 in [23] and also 3.9 in [42]). The tree is the $\mathbb{Z}$-tree associated to the usual simplicial tree given by the Bass-Serre theory ([127; 5.3]), viewing an HNN-extension as the fundamental group of a graph of groups where the graph has one vertex and one unoriented edge which is a loop (see Example 3 before Prop. 20 in [127; 5.1]). Recall that the points of the $\mathbb{Z}$-tree associated to a simplicial tree are the vertices of the simplicial tree, and the metric is the usual path metric on the tree. Also, the action is minimal since it is transitive (as a G-set, the $\mathbb{Z}$-tree is isomorphic to the coset space $G\backslash A$).

In particular, this applies when $A = \langle a \rangle$ is infinite cyclic and

$$G_n = \langle\ t,\ a \mid tat^{-1} = a^n\ \rangle,$$

where $n > 1$. In the corresponding simplicial tree, all vertices have degree $n+1$. For there is an (unoriented) edge for each coset $g\langle a \rangle$, where $g \in G$, joining the vertices $g\langle a \rangle$ and $gt\langle a \rangle$ (identifying the vertices with $G\backslash A$). It follows easily (using the normal form theorem for HNN-extensions) that the only vertices adjacent to $\langle a \rangle$ are $a^i t\langle a \rangle$ for $0 \leq i < n$ and $t^{-1}\langle a \rangle$, and these are distinct. Since G acts transitively, all vertices have degree $n+1$ as claimed.

Let F be a free group of rank 2 with basis $\{x, y\}$. The homomorphism $F \to G_n$ sending x to t, y to a gives a minimal action of end type of F on the $\mathbb{Z}$-tree (X_n, d_n) on which G_n acts. The associated homomorphism $\tau : F \to \mathbb{Z}$ sends x to 1 and y to 0, so τ, and therefore ℓ, are independent of n. However the $\mathbb{Z}$-trees (X_n, d_n) are pairwise non-isomorphic since the vertices of the corresponding simplicial trees have different degrees (any isomorphism as

$\mathbb{Z}$-trees gives a graph isomorphism of the corresponding simplicial trees). This gives the required counterexamples.

By contrast with Theorem 4.5, there are examples of an infinitely generated group having a non-trivial action of end type with no minimal invariant subtree. We refer to [37] for an example. The construction used there can be easily modified to give examples of actions of cut type. Such examples can also be constructed using Theorem 4.7 below. Before proceeding to this result, we state some observations on actions of cut type which follow easily from the discussion of actions of end type.

Theorem 4.6. *Let G be a group with an action of cut type on a Λ-tree (X, d).*
(a) *If the action is trivial, there is no minimal invariant subtree.*
(b) *If G is finitely generated then the action of G is non-trivial, and there are exactly two minimal invariant subtrees.*

Proof. Let the fixed cut consist of the two subtrees X_0, X_1 of X. By Theorem 2.6, the action of G restricted to X_0 and to X_1 is of end type. Moreover, any minimal G-invariant subtree must be entirely contained in one of X_0, X_1 since these two subtrees are disjoint and G-invariant. Thus (a) follows at once from Prop. 4.4, and (b) follows from Theorem 4.5. □

To give an example of a non-trivial action of cut type with no minimal invariant subtree, we shall use the next result. The proof will use the notation $\mathrm{Y}(x, y, z)$ when x, y, z may be ends or points of a Λ-tree, which was defined after Lemma 2.3.7.

The proof will also use the idea of an end of full Λ-type ([5; §1]). Given an ordered abelian group, let $[0, \infty)_\Lambda = \{a \in \Lambda \mid a \geq 0\}$. An end ε of a Λ-tree (X, d) is said to be of full Λ-type if for some $x \in X$, $[x, \varepsilon)$ is metrically isomorphic to $[0, \infty)_\Lambda$. (We leave it as an exercise to show that if this is true for some $x \in X$, then it is true for all $x \in X$.)

Theorem 4.7. *Let Γ be an ordered abelian group and put $\Lambda = \mathbb{Z} \oplus \Gamma$ with the lexicographic ordering. Suppose a group G has two actions of end type on Γ-trees (X_0, d_0) and (X_1, d_1) with associated homomorphisms τ, $-\tau$ respectively. Assume further that the fixed ends of X_0, X_1 are of full Γ-type. Then G has an action of cut type on a Λ-tree with the same hyperbolic length function as that for the actions on X_0 and X_1.*

Proof. For $i = 0, 1$, let ε_i be the fixed end of X_i for the action of G, and let δ_i be an end map towards ε_i (as defined in the addendum to 2.3.12). We can assume that $X_0 \cap X_1 = \emptyset$, and we put $X = X_0 \cup X_1$. We define $d : X \times X \to \Lambda$ by: $d|_{X_i \times X_i} = d_i$ for $i = 0, 1$, and for $x \in X_0$, $y \in X_1$, $d(x, y) = d(y, x) = (1, -\delta_0(x) - \delta_1(y))$ (Γ is viewed as a subgroup of Λ via the embedding $a \mapsto (0, a)$).

Clearly d satisfies all the axioms for a metric other than the triangle inequality. To verify the triangle inequality, take x, y, $z \in X$. If all three points are in one of X_0, X_1 the triangle inequality follows since d_i is a metric. Suppose x, $y \in X_0$ and $z \in X_1$. Then clearly $d(x,y) < d(x,z) + d(y,z)$. Let $p = Y(x,y,\varepsilon_0)$. Then $d(x,y) = d(x,p) + d(y,p) = 2\delta_0(p) - \delta_0(x) - \delta_0(y)$, by definition of end map. Therefore, $d(x,y)+d(y,z) = (1, 2\delta_0(p) - \delta_0(x) - 2\delta_0(y) - \delta_1(z))$, while $d(x,z) = (1, -\delta_0(x) - \delta_1(z))$. Thus $d(x,z) \leq d(x,y)+d(y,z)$ since $\delta_0(y) \leq \delta_0(p)$ (because $p \in [y, \varepsilon_0\rangle$). Similarly, $d(y,z) \leq d(x,z) + d(x,y)$, and a similar argument establishes the triangle inequality when two of the points are in X_1 and one is in X_0.

Now we show (X,d) is a Λ-tree. The first thing is to show it is geodesic, that is, if x, $y \in X$, there is a segment $[x,y]$ with endpoints x, y. If both points are in one of X_0, X_1 we take $[x,y]$ to be the unique segment with endpoints x, y in the Γ-tree X_0 or X_1, as appropriate, which is a segment in the Λ-tree X. Thus we can assume $x \in X_0$, $y \in X_1$. We claim that $[x,y] = [x,\varepsilon_0\rangle \cup [y,\varepsilon_1\rangle$ is the required segment. For define $\alpha : [x,y] \to \Lambda$ by $\alpha(z) = d(x,z)$. To show α is an isometry, we need to show that, if z, $w \in [x,y]$, then $d(z,w) = |d(x,z) - d(x,w)|$. If z, $w \in X_0$ this is clear since $[x,\varepsilon_0\rangle$ is a linear subtree of X_0 with endpoint x. Suppose $z \in X_0$, $w \in X_1$. Then $|d(x,z) - d(x,w)| = |(0, d_0(x,z)) - (1, -\delta_0(x) - \delta_1(w))| = (1, \delta_0(x) - \delta_1(w) - d_0(x,z))$, and $d(z,w) = (1, -\delta_0(z) - \delta_1(w))$, so we need $d_0(x,z) = \delta_0(z) - \delta_0(x)$, which is true because δ_0 restricted to $[x,\varepsilon_0\rangle$ is an isometry. Finally, if z, $w \in X_1$ then

$$\begin{aligned} |d(x,z) - d(x,w)| &= |(1, -\delta_0(x) - \delta_1(z)) - (1, -\delta_0(x) - \delta_1(w))| \\ &= |(0, \delta_1(w) - \delta_1(z))| = (0, d_1(w,z)) = d(w,z), \end{aligned}$$

because δ_1 restricted to $[y,\varepsilon_1\rangle$ is an isometry. It follows that α is an isometry, and since ε_i ($i = 0$, 1) are of full Γ-type, the image of α is easily seen to be $[0, d(x,y)]_\Lambda$, hence $[x,y]$ is a segment.

Suppose x, y and $u \in X$, and $u \notin [x,y]$. We show that $d(x,u) + d(u,y) > d(x,y)$, so u does not lie on any segment in X with endpoints x, y. This will show uniqueness of segments in X.

First, assume x, $y \in X_0$. If $u \in X_0$, let $w = \mathrm{Y}(x,y,u)$. As noted before Cor. 2.1.3, $d(x,u) + d(u,y) = d(x,y) + 2d(u,w) > d(x,y)$ (because $w \in [x,y]$, so $w \neq u$).If $u \in X_1$, then $d(x,u) + d(u,y) > d(x,y)$ by definition of d.

Next, assume $x \in X_0$, $y \in X_1$. If $u \in X_0$, let $p = Y(u,x,\varepsilon_0)$. Then it is easily checked that $d(u,y) = d(u,p) + d(p,y)$ and $d(u,x) = d(u,p) + d(p,x)$, so $d(u,x) + d(u,y) = d(x,y) + 2d(u,p) > d(x,y)$ (because $p \in [x,y]$, so $p \neq u$). Similarly (interchanging x and y in this argument) if $u \in X_1$, $d(u,x)+d(u,y) > d(x,y)$. By symmetry this covers all cases and we have shown uniqueness of segments.

Now suppose x, y, $z \in X$. We need to show that $[x,y] \cap [x,z]$ is a segment. Suppose $x \in X_0$. It is easily checked that the following holds

$$[x,y] \cap [x,z] = \begin{cases} [x, Y(x,y,z)] & \text{if } y,\ z \in X_0 \\ [x, Y(x,y,\varepsilon_0)] & \text{if } y \in X_0,\ z \in X_1 \\ [x, Y(z,y,\varepsilon_1)] & \text{if } y,\ z \in X_1. \end{cases}$$

A similar argument works if $x \in X_1$.

It remains to show that, if $[x,y] \cap [x,z] = \{x\}$, then $[x,y] \cup [x,z]$ is a segment. Again by symmetry we need only consider the case $x \in X_0$. If y, $z \in X_0$, this follows since X_0 is a Γ-tree. Suppose $y \in X_0$, $z \in X_1$. Then $[x,z] = [x,\varepsilon_0\rangle \cup [z,\varepsilon_1\rangle$, so $[x,y] \cap [x,\varepsilon_0\rangle = \{x\}$, hence $[y,\varepsilon_0\rangle = [x,y] \cup [x,\varepsilon_0\rangle$ (which follows from the discussion after Lemma 2.3.2). Thus $[x,y] \cup [x,z] = [x,y] \cup [x,\varepsilon_0\rangle \cup [z,\varepsilon_1\rangle = [y,\varepsilon_0\rangle \cup [z,\varepsilon_1\rangle = [y,z]$, as required. If y, $z \in X_1$, then $[x,\varepsilon_0\rangle \subseteq [x,y] \cap [x,z]$, and $[x,\varepsilon_0\rangle \neq \{x\}$ since ε_0 is an open end (Theorem 2.6), so this case is impossible. This covers all cases, and we have shown that (X,d) is a Λ-tree.

Next, the actions of G on X_i give an action on X, and we need to show this is an action by isometries, that is, if x, $y \in X$ and $g \in G$, then $d(x,y) = d(gx,gy)$. This is clear if both points belong to one of X_0, X_1, so we can assume $x \in X_0$ and $y \in X_1$. By Prop. 2.5, $\delta_0(gx) = \delta_0(x) + \tau(g)$ and $\delta_1(gy) = \delta_1(y) - \tau(g)$. Thus $d(gx,gy) = (1, -\delta_0(gx) - \delta_1(gy)) = (1, -\delta_0(x) - \delta_1(y)) = d(x,y)$.

The action of G on X has the same hyperbolic length function ℓ as that for the actions on X_0, X_1 by Lemma 1.7, consequently the action is abelian. If some element g of G acts as an inversion, that is, g has no fixed point but g^2 does, then g acts as an inversion on either X_0 or X_1, which is impossible since the actions on X_i are of end type. Hence the action on X is without inversions. Thus for $g \in G$, the characteristic set A_g for the action on X is equal to $\{\, x \in X \mid d(x,gx) = \ell(g)\}$, so $A_g \cap X_i$ is the characteristic set for the action on X_i. Therefore $\bigcap_{g\in G} A_g = \bigcap_{g\in G}(A_g \cap X_0) \cup \bigcap_{g \in G}(A_g \cap X_1) = \emptyset$, so the action on X is not of core type.

If ε is a fixed end for the action on X, take $x \in X_0$. If $y \in [x,\varepsilon\rangle$ and $y \in X_1$, then $[y,\varepsilon\rangle \subseteq X_1$, for if $z \in [y,\varepsilon\rangle \cap X_0$ then $y \in [x,z]$, which implies $y \in X_0 \cap X_1$, an impossibility. Thus for some i and $w \in X_i$, $[w,\varepsilon\rangle \subseteq X_i$, i.e. ε belongs to X_i. By Lemma 2.3.8(1), $[w,\varepsilon\rangle$ is a ray in X_i, and the end of X_i it defines is invariant under G. Since the action on X_i is of end type, this end must be ε_i. But this is impossible since the X_i-ray $[w,\varepsilon_i\rangle$ is not an X-ray, being contained in $[w,u]$ for any $u \in X_{1-i}$. It follows that the action on X is of cut type, and the fixed cut is the pair of subtrees X_0, X_1 (and the ε_i are the ends meeting at the cut). □

As noted before Theorem 4.6, we can find a non-trivial action of a group G on a Γ-tree of end type with no minimal invariant subtree, and in fact, there are examples with $\Gamma = \mathbb{Z} \oplus \mathbb{Z}$ with the lexicographic ordering, and with G countable. Further, in these examples the fixed end is of full Γ-type (see [37]). Let F be a free group of countable rank with basis $\{a_i \mid i \in \mathbb{N}\}$, and let $\phi : F \to G$ be a surjective homomorphism. This gives a non-trivial action of end type of F on X with no minimal invariant subtree, with associated homomorphism τ, say. Now let $\psi : F \to G$ be the homomorphism which sends a_i to $\phi(a_i)^{-1}$. Then ψ gives an action of F on X, also non-trivial of end type and with no minimal invariant subtree, but with associated homomorphism $-\tau$. We can therefore find two disjoint copies X_0, X_1 of X with F having non-trivial action of end type and no minimal invariant subtree, but with associated homomorphisms τ, $-\tau$. Theorem 4.7 then gives a non-trivial action of cut type of F on a Λ-tree, where $\Lambda = \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$ with the lexicographic ordering. As noted in the proof of Theorem 4.6, any invariant subtree lies in one of X_0, X_1, so there is no minimal invariant subtree for this action.

Theorem 4.6(b) raises the question of whether the two minimal invariant subtrees must be isomorphic. We shall use Theorem 4.7 to show that this need not be the case. Recall the example after Theorem 4.5 of the HNN-extension

$$G_n = \langle\ t,\ a \mid tat^{-1} = a^n\ \rangle$$

which acts on a $\mathbb{Z}$-tree with action of end type, and the associated homomorphism is the canonical map $G_n \to \mathbb{Z}$. The action is minimal, and for different values of n, the $\mathbb{Z}$-trees are non-isomorphic (G_n-equivariantly or otherwise). Again let F be a free group of rank 2 with basis x, y and consider the homomorphisms $F \to G_2$ ($x \mapsto t$, $y \mapsto a$) and $F \to G_3$ ($x \mapsto t^{-1}$, $y \mapsto a$). This gives two non-trivial actions of F of end type on $\mathbb{Z}$-trees with associated homomorphisms which can be denoted by τ and $-\tau$. In a $\mathbb{Z}$-tree, an open end is necessarily of full $\mathbb{Z}$-type, so Theorem 4.7 gives a non-trivial action of F on a $\mathbb{Z} \oplus \mathbb{Z}$-tree of cut type. The construction gives an example where the two minimal subtrees in Theorem 4.6(b) are the subtrees of the cut, and are non-isomorphic.

CHAPTER 4. Aspects of Group Actions on Λ-trees

1. Introduction

Not only this chapter, but the rest of these notes will be concerned with actions of groups as isometries on Λ-trees. The main question, which was posed in [2] is: find the group-theoretic information carried by a Λ-tree action, analogous to that presented in Serre's book [127] for the case $\Lambda = \mathbb{Z}$ (simplicial trees). However it is in general an intractable problem. One way to proceed is to put restrictions on the action, a point of view which will be adopted in Chapter 5. Note that every group has a trivial action on a Λ-tree, that is, one with no hyperbolic isometries (e.g. take the tree to have just a single point). Thus when considering this question it is reasonable to confine attention to non-trivial actions. (However, the reader should not conclude that trivial actions are uninteresting. There is a forceful reminder of this in Shalen's survey [129; 1.4], where there is a discussion of work of Gupta and Sidki [70].) In §2 we shall consider a different question. If we put group-theoretic restrictions on a group G, what restrictions does this place on the type of action G can have on a Λ-tree?

One reason for the importance of Λ-trees is the work of Morgan and Shalen ([102], [106], [107], [108]). They showed, among other things, the following. Let G be a finitely generated group, let $\mathrm{Isom}^+(\mathbb{H}^n)$ denote the group $\mathrm{SO}_0^+(n, 1)$ of orientation-preserving isometries of real n-dimensional hyperbolic space $\mathbb{H}^n$. Give G the discrete topology, and $\mathrm{Hom}(G, \mathrm{Isom}^+(\mathbb{H}^n))$ the compact-open topology. Let $\mathcal{F}_n(G)$ be the subspace of this consisting of the discrete, faithful representations, and let $\mathcal{H}_n(G)$ be the quotient space $\mathcal{F}_n(G)/\sim$, where, for α, $\beta \in \mathcal{F}_n(G)$, $\alpha \sim \beta$ if and only if there exists $\gamma \in \mathrm{Isom}^+(\mathbb{H}^n)$ such that $\alpha(g) = \gamma\beta(g)\gamma^{-1}$ for all $g \in G$. Suppose G is not abelian by finite. Then $\mathcal{H}_n(G)$ has a compactification $\mathcal{H}_n(G) \cup \mathcal{T}$, where the points of $\mathcal{T}$ arise from non-trivial actions on $\mathbb{R}$-trees in which the stabiliser of any segment with more than one point is abelian by finite. This work will not be presented here.

In the case that G is the fundamental group of a closed surface with negative Euler characteristic, $\mathcal{H}_n(G)$ is one description of the Teichmüller space of the surface. A compactification of this space was achieved by Thurston using the idea of a measured geodesic lamination, and the compactification of

Morgan and Shalen is a generalisation of this. Morgan and Shalen [109] showed how to construct an $\mathbb{R}$-tree from a measured lamination. We shall present this result in §4, and in §5, following the account of Morgan and Shalen, prove Thurston's result that the space of "projectivised measured geodesic laminations" for a closed surface with negative Euler characteristic is compact. (Thurston showed that this space is a sphere of dimension $-3\chi - 1$, where χ is the Euler characteristic of the surface. See [139], p.8.60 and [140], Theorem 5.2.) This compactness result will be of use in the next chapter, where we summarise the argument ([109]) showing that most surface groups act freely on an $\mathbb{R}$-tree.

The work of Morgan and Shalen gives another way in which Λ-trees arise in nature. This is a means of constructing a Λ-tree on which the group $SL_2(F)$ acts without inversions, where F is a field equipped with a valuation, and Λ is the value group. This will be discussed in §3.

The space $\mathcal{T}$ may be viewed as a subspace of the space of "projectivised length functions" of the finitely generated group G. This will be defined in §6. In view of the results mentioned above, it is natural to ask if this space is itself compact. This was shown to be so by Culler and Morgan [42]. We shall prove this, and consider compactness results for various subspaces in §6. We shall also briefly mention, with no details, results on the contractibility of certain subspaces in the case that G is a free group, and applications to automorphisms of free groups.

One interesting topic which we shall not discuss in detail is the connection between actions on $\mathbb{R}$-trees and the Bieri-Neumann-Strebel Invariant, which is briefly mentioned at the end of §2.

We close this section by showing that, if a finitely generated group has a non-trivial action on a Λ-tree for some ordered abelian group Λ, then it has a non-trivial action on an $\mathbb{R}$-tree. First we need a lemma.

Lemma 1.1. *Let G be a group acting without inversions on a Λ-tree X. Suppose G is generated by $g_1, \ldots, g_n$. Then there exist generators $h_1, \ldots, h_n$ for G, obtained from $g_1, \ldots, g_n$ by Nielsen transformations, such that $A_{h_1} \cap \ldots \cap A_{h_n} \neq \emptyset$.*

Proof. We use induction on n. Since the action is without inversions we have $A_{g_1} \neq \emptyset$, so if $n = 1$ we take $h_1 = g_1$. Suppose $n > 1$ and the lemma is true for $n-1$, so $\langle g_1, \ldots, g_{n-1} \rangle = \langle h_1, \ldots, h_{n-1} \rangle$, where $A_{h_1} \cap \ldots \cap A_{h_{n-1}} \neq \emptyset$. Put $A = A_{h_1} \cap \ldots \cap A_{h_{n-1}}$, a closed subtree of X, being the intersection of finitely many closed subtrees. If $A \cap A_{g_n} \neq \emptyset$, we can take $h_n = g_n$, so suppose $A \cap A_{g_n} = \emptyset$. Let $[u, v]$ be the bridge between A and A_{g_n} with $u \in A_{g_n}$ and $v \in A$. Since A_{h_i} is a closed subtree and $v \in A_{h_i}$, $A_{h_i} \cap [u, v] = [v_i, v]$ for some $v_i \in X$ (where $1 \leq i \leq n-1$). Re-order the h_i so that the v_i occur in order in $[u, v]$, that is, $[u, v] = [u, v_1, \ldots, v_{n-1}, v]$. Then $v_{n-1} \in [v_i, v] \subseteq A_{h_i}$

for $1 \leq i \leq n-1$, so $v_{n-1} \in A$, hence $[v_{n-1}, v] \subseteq A \cap [u,v] = \{v\}$, that is, $v_{n-1} = v$.

Thus $A_{h_{n-1}} \cap [u,v] = \{v\}$ and $A_{g_n} \cap [u,v] = \{u\}$. It follows that $A_{h_{n-1}} \cap A_{g_n} = \emptyset$. For if $w \in A_{h_{n-1}} \cap A_{g_n}$, then $[u,w] \cap [v,w] = [w,w']$ for some $w' \in [u,v]$; but $[u,w] \subseteq A_{g_n}$ and $[v,w] \subseteq A_{h_{n-1}}$, so $w' \in A_{h_{n-1}} \cap A_{g_n} \cap [u,v]$, which is impossible as this intersection is empty. It follows from Lemma 2.1.9 that $[u,v]$ is the bridge between $A_{h_{n-1}}$ and A_{g_n}. Now let $h_n = g_n h_{n-1}$, so $G = \langle h_1, \ldots, h_n \rangle$, and inductively $h_1, \ldots, h_n$ is obtained from $g_1, \ldots, g_n$ by Nielsen transformations. By Lemma 3.2.2, $[u,v] \subseteq A_{h_n}$; in particular, $v \in A \cap A_{h_n}$, so this intersection is non-empty, as required. □

The following result first appeared in [31]. We give a simplified proof which has also been noted in [80].

Theorem 1.2. *If a finitely generated group G has a non-trivial action on a Λ-tree for some ordered abelian group Λ, then it has a non-trivial action on some $\mathbb{R}$-tree.*

Proof. Suppose G acts non-trivially on a Λ-tree (X,d). Recall from Chapter 1, §1 that there is an embedding of Λ into $\frac{1}{2}\Lambda$. By Lemma 3.1.3, the induced action of G on $\frac{1}{2}\Lambda \otimes_\Lambda X$ is without inversions, and is non-trivial by Lemma 3.2.1. Thus we may assume the action is without inversions. By Lemma 1.1, we can choose generators h_i $(1 \leq i \leq n)$ for G such that $A_{h_1} \cap \ldots \cap A_{h_n} \neq \emptyset$. Choose a point $u \in A_{h_1} \cap \ldots \cap A_{h_n}$, and let L_u be the corresponding Lyndon length function. Reorder the generators so that $L_u(h_1) \leq \ldots \leq L_u(h_n)$. Let H be the intersection of all convex subgroups of Λ containing $L_u(h_n)$. Since L_u satisfies the triangle inequality $L_u(gh) \leq L_u(g) + L_u(h)$ and $L_u(g^{-1}) = L_u(g)$, it follows that $L_u(g) \in H$ for all $g \in G$, by writing g as a product of the generators and their inverses. Also $L_u(h_n) \neq 0$, otherwise L_u would be identically zero and the action would be trivial by Lemma 3.2.1.

Let K be the union of all convex subgroups of Λ not containing $L_u(h_n)$. Since the convex subgroups are linearly ordered by inclusion, K is a convex subgroup contained in H, $L_u(h_n) \notin K$, and there are no convex subgroups strictly between K and H. Thus H/K is a rank 1 ordered abelian group. Also, since $u \in A_{h_n}$, $L_u(h_n) = \ell(h_n)$, the hyperbolic length of h_n, by Cor. 3.1.5, so $\ell(h_n) \notin K$.

Let $Y = \bigcup_{g \in G} [u, gu]$, the subtree of X spanned by the orbit Gu. By Lemma 2.1.2, the metric on Y obtained from d by restriction takes values in H, since H is a convex subgroup of Λ. Thus Y is an H-tree on which G acts by restriction. Also, L_u is the Lyndon length function corresponding to the point u and the action of G on Y, and passing from X to Y does not change the hyperbolic length by Lemma 3.2.1. Put $Z = (H/K) \otimes_H Y$, formed using the canonical projection $H \to H/K$. There is an induced action of G on Z, and by Lemma 3.2.1 h_n has non-zero hyperbolic length for this action, which is therefore non-trivial.

By Theorem 1.1.2, H/K embeds as an ordered abelian group in $\mathbb{R}$, and there is an induced action of G on the $\mathbb{R}$-tree $\mathbb{R}\otimes_{H/K} Z$, which again by Lemma 3.2.1 is non-trivial, as required. □

Thus if we want to investigate the class of finitely generated groups which have a non-trivial action on a Λ-tree, it suffices to consider the case $\Lambda = \mathbb{R}$. This is no longer true, however, if we put further restrictions on the kind of action allowed, as we shall see in Chapter 5.

2. Actions of Special Classes of Groups

We now consider certain group-theoretic properties, and the restrictions on the type of action on a Λ-tree which a group with one of these properties can have. The best results in this direction are those of Khan and Wilkens [85], and it is mostly these we shall present. The arguments, which mainly apply only when $\Lambda = \mathbb{R}$, use machinery developed in earlier papers of Wilkens, including [147], [148] and [149]. Some of the results are couched in terms of Lyndon length functions, but we shall usually restate them in terms of group actions on trees, and use the geometry of Λ-trees in the proofs. The starting point is actually a discussion of trivial actions. The argument for the next lemma and its corollary is given in [106] (see also [128; 1.14]), and generalises a result of Serre [127] for simplicial trees.

Lemma 2.1. *Let G be generated by $g_1, \ldots, g_n$ and assume that G acts on a Λ-tree. Suppose that g_i , for $1 \leq i \leq n$, and $g_i g_j$, for $j < i$ all have a fixed point. Then there is point of the tree fixed by all of G.*

Proof. We use induction on n, the result being clear if $n = 1$. Let H be the subgroup generated by $g_1, \ldots, g_{n-1}$, and assume inductively that there is a point $v \in X$ fixed by all of H. By Lemma 1.1, the midpoint u of $[v, g_n v]$ is fixed by g_n, and indeed by $g_n g_i$ for $1 \leq i \leq n-1$, because $g_n g_i v = g_n v$. It follows that u is fixed by $g_1, \ldots, g_n$, hence by G. □

If G is finitely generated, the next result says that a trivial action (which is clearly abelian) must be of type (1) or (2) at the beginning of §3.4, giving an easier proof of part of Theorem 3.4.5. (In general, a trivial action can also be of type (4) or (5); see §3.10 in [42] for an example.)

Corollary 2.2. *Let G be a finitely generated group acting trivially and without inversions on a Λ-tree (X, d). Then there is a point of X which is fixed by all elements of G.*

Proof. If there are no inversions, every element of G acts as an elliptic isometry, and the result is immediate by Lemma 2.1. □

Next we shall give another result asserting that under certain circumstances, an action has a fixed point. This needs two technical lemmas. Recall that, if a group G acts on a Λ-tree (X,d) and $x \in X$, the Lyndon length function $L_x : G \to \Lambda$ is defined by $L_x(g) = d(x, gx)$.

Lemma 2.3. *Let G be a group acting on a Λ-tree (X,d), and let $x \in X$. If a, $b \in G$ and a, b, ab are all elliptic isometries, then two of $L_x(a)$, $L_x(b)$, $L_x(ab)$ are equal, with the third no greater. (Thus $(-L_x(a), -L_x(b), -L_x(ab))$ is an isosceles triple.) Moreover, if u is the midpoint of $[x, cx]$, where $c \in \{a,b\}$ is such that $L_x(c) = \max\{L_x(a), L_x(b)\}$, then u is fixed by a and b.*

Proof. By Lemma 3.2.2, $A_a \cap A_b$ is non-empty, and the intersection of any two of A_a, A_b, A_{ab} coincides with the intersection of all three. Let $[x, p_y]$ denote the bridge between x and A_y, for y not an inversion, and choose $q \in A_a \cap A_b \cap A_{ab}$. Then p_a, p_b, p_{ab} all lie in $[x,q]$. For example, suppose $[x,q] = [x, p_a, p_b, p_{ab}, q]$, then $p_b \in [p_a, q]$, hence $p_b \in A_a \cap A_b \subseteq A_{ab}$. It follows that $p_b = p_{ab}$, so by Lemma 3.1.1, $d(x, bx) = d(x, abx) = 2d(x, p_b) \geq 2d(x, p_a) = d(x, ax)$. Further, $u = p_b$ is fixed by a and b. The other cases are similar. □

Lemma 2.4. *Let $L : G \to \Lambda$ be a Lyndon length function, and let $g \in G$.*
(i) *If $L(g^2) \leq L(g)$, then $L(g^n) \leq L(g)$ for all integers n.*
(ii) *If $L(g^2) > L(g)$, then $L(g^n) = L(g) + (n-1)(L(g^2) - L(g))$ for $n \geq 1$.*

Proof. We may replace L by $2L$ and so assume $L = L_x$ for some Λ-tree (X,d) on which G acts without inversions, and some $x \in X$, by Theorem 2.4.6 and the characterisation of inversions in Lemma 3.1.2(iv). By Lemma 3.1.8, if $L(g^2) \leq L(g)$ then g has a fixed point, and $L_x(g) = 2d(x, A_g)$. Since $A_g \subseteq A_{g^n}$, $L_x(g^n) = 2d(x, A_{g^n}) \leq 2d(x, A_g)$, whence (i). If $L(g^2) > L(g)$, then g is hyperbolic by Lemma 3.1.8, and $A_g = A_{g^n}$ for $n > 0$ by Lemma 3.1.7, so

$$L_x(g^n) = 2d(x, A_g) + \ell(g^n) = 2d(x, A_g) + n\ell(g) = L_x(g) + (n-1)\ell(g)$$

using Theorem 3.1.4. □

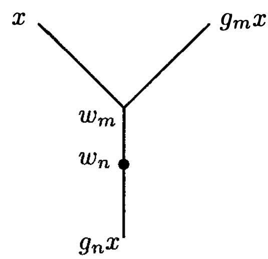
Suppose G acts on a Λ-tree (X,d) and x, $y \in X$. Then for $g \in G$, $L_x(g) = d(x, gx) \leq d(x,y) + d(y, gy) + d(gy, gx) = 2d(x,y) + L_y(g)$. Thus if the Lyndon length function L_y is bounded, in the sense that there is $r \in \Lambda$ such that $L_y(g) \leq r$ for all $g \in G$, then L_x is bounded for all $x \in X$. In this case we say that the action of G is bounded.

It follows from Lemma 2.4(ii) that if Λ is archimedean, a bounded action is trivial. This, and the fact that the next result uses the notion of a complete $\mathbb{R}$-tree, are the main reasons why we confine attention to $\mathbb{R}$-trees in the results which follow. The next result is due to Wilkens [148].

Lemma 2.5. *Suppose G has a bounded action on a complete $\mathbb{R}$-tree (X, d). Then there is a fixed point for the action.*

Proof. As we have noted, the action is trivial by Lemma 2.4. Take $x \in X$, and let $r = \sup\{L_x(g) \mid g \in G\}$. If there is an element $g \in G$ such that $L_x(g) = r$ then by Lemma 2.3 the midpoint of $[x, gx]$ is fixed by G.

Suppose no such g exists. For each positive integer n, choose $g_n \in G$ such that $L_x(g_n) > r - (1/n)$ and $L_x(g_n) > L_x(g_m)$ if $n > m$. Let w_n be the midpoint of $[x, g_n x]$. If $n > m$ then by Lemma 2.3, $L_x(g_n^{-1} g_m) = L_x(g_n) > L_x(g_m)$. Thus $d(x, g_n x) = d(x, g_n^{-1} g_m x) = d(g_n x, g_m x)$, and so by Lemma 2.1.2, $[x, g_n x] \cap [x, g_m x] = [x, w_m]$. Also, $w_n \in [w_m, g_n x]$. The situation is illustrated by the following picture.



It follows that

$$d(w_n, w_m) = (1/2)(L_x(g_n) - L_x(g_m)) \leq (1/2)(r - (r - (1/m))) = 1/2m,$$

hence (w_n) is a Cauchy sequence. By completeness, it converges; let $w = \lim_{n \to \infty} w_n$.

We claim w is a fixed point for G. For suppose $gw \neq w$ for some $g \in G$. Put $\epsilon = d(w, gw)$. Choose n such that $L_x(g_n) > L_x(g)$ and $d(w, w_n) < \epsilon/2$. By Lemma 2.3, w_n is fixed by g, so $d(w_n, gw) = d(gw_n, gw) = d(w_n, w)$. But then $d(w, gw) \leq d(w, w_n) + d(w_n, gw) < \epsilon$, a contradiction, which establishes the lemma. □

We continue to investigate bounded actions. Suppose a group G acts on a Λ-tree (X, d), $x \in X$ and L_x is the corresponding Lyndon length function. We define $H_x = \{a \in G \mid L_x(ag) = L_x(g) \text{ for all } g \in G \setminus N\}$.

Lemma 2.6. *In these circumstances, $H_x \subseteq N$ and H_x is a subgroup of G. Further, for all $a \in H_x$, $g \in G \setminus N$, $L_x(a) < L_x(g)$, so if $N \neq G$, H_x has bounded action on X.*

Proof. Write H for H_x. If $a \in H$ and $a \notin N$, then taking $g = a$ in the definition of H, $L(a^2) = L(a)$, so $a \in N$, a contradiction. Hence $H \subseteq N$.

Next, we show that if $a \in H$, $g \in G \setminus N$, then $L_x(a) < L_x(g)$. Firstly, $A_a \cap [x, gx] \neq \emptyset$. For suppose this intersection is empty, and let $[p, q]$ be the bridge between $[x, gx]$ and A_a, with $p \in [x, gx]$. Then $[x, q] = [x, p, q]$ and $[gx, q] = [gx, p, q]$, so $[agx, q] = [agx, ap, q]$ since $aq = q$. By Lemma 3.1.1, $[p, q] \cap [ap, q] = \{q\}$, so $[x, agx] = [x, p, q, ap, agx]$, and by Lemma 2.1.4, $d(x, agx) = d(x, p) + d(p, q) + d(q, ap) + d(ap, agx) = d(x, p) + 2d(p, q) + d(p, gx) = d(x, gx) + 2d(p, q)$, and $d(p, q) > 0$, so $L_x(ag) > L_x(g)$, contradicting $a \in H$. Since $g^{-1} \in G \setminus N$, it follows that $A_a \cap [x, gx]$, $A_a \cap [x, g^{-1}x]$ are non-empty, so by Lemma 2.1.11, $A_a \cap [x, gx] \cap [x, g^{-1}x]$ is non-empty. Take r in their intersection, so $r \in [x, w]$, where $w = \text{Y}(x, gx, g^{-1}x)$. Then by Lemma 3.1.1,

$$L_x(a) = 2d(x, A_a) \leq 2d(x, r) \leq 2d(x, w) < 2d(x, w) + \ell(g) = L_x(g)$$

by Theorem 3.1.4 (where ℓ is hyperbolic length).

If $a \in H$, we need to show that $a^{-1} \in H$, equivalently that $L_x(ga) = L_x(g)$ for all $g \in G \setminus N$. Since N is closed under conjugation (by Lemma 3.1.7), $a^{-1}ga$, $a^{-1}g^{-1}a \notin N$. Also, $L_x(a^{-1}ga) = L_x(a^{-1}g^{-1}a)$, hence

$$\begin{aligned} L_x(ga) &= L_x(aa^{-1}ga) = L_x(a^{-1}ga) \\ &= L_x(a^{-1}g^{-1}a) = L_x(aa^{-1}g^{-1}a) = L_x(g^{-1}a). \end{aligned}$$

If $ga \notin N$, then $L_x(ga) = L_x(a^{-1}g^{-1}) = L_x(aa^{-1}g^{-1}) = L_x(g^{-1}) = L_x(g)$. If $g^{-1}a \notin N$, then $L_x(ga) = L_x(g^{-1}a) = L_x(a^{-1}g) = L_x(aa^{-1}g) = L_x(g)$. Otherwise, ga, $g^{-1}a \in N$, hence ag^{-1}, $a^2 \in N$ by Lemma 3.1.7 (ag^{-1} being a conjugate of $g^{-1}a$). By Lemma 2.3, two of $L_x(ag^{-1})$, $L_x(ga)$ and $L_x(a^2)$ are equal, with the third no greater. By what has already been proved, $L_x(ag^{-1}) = L_x(g^{-1}) = L_x(g) > L_x(a) \geq L_x(a^2)$ (using Lemma 3.1.8 for the last inequality). Hence $L_x(ga) = L_x(ag^{-1}) = L_x(g)$.

Now suppose a, $b \in H$ and $g \in G \setminus N$. If $bg \notin N$ then $L(abg) = L(bg) = L(g)$, while if $abg \notin N$, then since $a^{-1} \in H$, $L(abg) = L(a^{-1}(abg)) = L(bg) = L(g)$. Otherwise by Lemma 2.3, two of $L(a)$, $L(bg)$, $L(abg)$ are equal, with the third no greater. But by what is already proved, $L(bg) = L(g) > L(a)$, so $L(abg) = L(bg) = L(g)$. It follows that $ab \in H$. Since $1 \in H$, it follows that H is a subgroup of G. □

Exercise. For any ordered abelian group Λ, if A is a proper subset of Λ, then Λ is not metrically isomorphic to A. (Use Lemma 1.2.1.)

Lemma 2.7. *Let G act on an $\mathbb{R}$-tree (X, d) and let $L = L_x$, where $x \in X$. Suppose K is a normal subgroup of G. If K has bounded action then K fixes A_g pointwise for all $g \notin N$. If $N \neq G$, then K has bounded action if and only if $K \leq H_x$.*

Proof. The action of G extends to an action on the metric completion $\hat{X}$, as noted at the end of §2 in Ch.2, and $\hat{X}$ is an $\mathbb{R}$-tree by Theorem 2.4.14. This action has the same Lyndon length function L_x based at x. It follows that it has the same hyperbolic length function ℓ as the action on X, by Lemma 3.1.8. If $g \in G$ acts as a hyperbolic isometry, it follows that A_g (the axis of g in X) is contained in $\hat{A}_g$ (the axis of g in $\hat{X}$), by Cor. 3.1.5. By the remark after Theorem 3.4.2, A_g and $\hat{A}_g$ are both metrically isomorphic to $\mathbb{R}$, hence $A_g = \hat{A}_g$, using the exercise preceding the lemma.

If K has bounded action (on X, so on $\hat{X}$) then there is a fixed point for the action on $\hat{X}$, say y, by Lemma 2.5. If $g \in G \setminus N$, K fixes $g^n y$ for every integer n, since K is normal in G. Hence K fixes $[g^{n-1}y, g^n y]$ pointwise, for every integer n. Let $[y, q]$ be the bridge between y and A_g; by Theorem 3.1.4, $[q, gq] \subseteq [y, gy]$, so $[g^{n-1}q, g^n q] \subseteq [g^{n-1}y, g^n y]$ for all integers n. By the remark after Theorem 3.4.2, $A_g \subseteq \bigcup_{n\in\mathbb{Z}}[g^{n-1}y, g^n y]$, hence K fixes A_g pointwise.

Now if $a \in K$, and $[x, p]$ is the bridge between x and A_g, then $[x, gx] = [x, p, gp, gx]$ by Theorem 3.1.4. Also, $[agp, agx] = [gp, agx]$ is the bridge between A_g and agx, so $[x, agx] = [x, p, gp, agx]$, by Lemma 2.1.5. Further, $d(gp, agx) = d(agp, agx) = d(gp, gx)$, hence $L_x(g) = L_x(ag)$ by Lemma 2.1.4. We conclude that $a \in H_x$, so $K \leq H_x$. Conversely, if $K \subseteq H_x$ then K has bounded action on X by Lemma 2.6. □

Thus if G acts on an $\mathbb{R}$-tree (X, d) and $L = L_x$ for some $x \in X$, and the action is non-trivial, there is a maximal normal subgroup of G having bounded action on X, namely the core of H in G, where $H = H_x$, that is, $\bigcap_{g\in G} gHg^{-1}$, which we denote by $\mathrm{core}_G(H)$.

In order to state our results it is convenient to divide group actions on $\mathbb{R}$-trees into the following types. There are no actions with inversions on an $\mathbb{R}$-tree since $\mathbb{R} = 2\mathbb{R}$ (see Lemma 3.1.2), and no actions of cut type (by the remark after Theorem 3.2.6), so from §3.2 an action of a group G on an $\mathbb{R}$-tree is of exactly one of the following types, where N denotes the set of non-hyperbolic elements of G.

(I) $N = G$ and there is a fixed point for the action (Type (2) at the beginning of §3.4).

(II) $N = G$ and there is no fixed point for the action (end type (4) with $N = G$).

(III) (a) A dihedral action.

(b) A shift (i.e. Type (3) at the beginning of §3.4).

(IV) End type (4) with $N \neq G$.

(V) Irreducible action.

In the results which follow, type will refer to one of types (I)-(V).

Remark. For an action of a group G on an $\mathbb{R}$-tree, the following are equivalent.

(a) The action is of type III.

(b) There is exactly one axis for the hyperbolic elements of G.

(c) The action is non-trivial and there is a (non-empty) linear G-invariant subtree.

Proof. Let $A = \bigcap_{g \in G \setminus N} A_g$ and $B = \bigcap_{g \in G} A_g$.

If the action is dihedral, it was shown in the proof of 3.2.8 that A is a (non-empty) G-invariant linear subtree. If the action is a shift, $\emptyset \neq B \subseteq A$ and by 3.1.7 A is a linear invariant subtree. If $p \in A$, then $A_p \subseteq A$, where $A_p = \bigcup_{n \in \mathbb{Z}} [g^n p, g^{n+1} p]$. By the remark after Theorem 3.4.2, $A = A_g$ for all $g \in G \setminus N$. Thus (a) implies (b), and trivially (b) implies (c).

Assume the action is non-trivial and C is a linear invariant subtree. Then the action is dihedral or abelian by Lemma 3.2.10. In the abelian case, $C \subseteq B$ is non-empty by 3.2.10(c), so the action is a shift. Thus (c) implies (a), completing the proof.

[In fact, if C is a linear invariant subtree and g is hyperbolic, there is a point $p \in A_g \cap C$ by 3.1.6. By the remark following 3.4.2, $A_g \subseteq C$ and A_g is metrically isomorphic to $\mathbb{R}$. By the exercise preceding Lemma 2.7, $C = A_g$ for all hyperbolic g. Thus an action is of type III if and only if there is a unique invariant line (where "line" means an isometric copy of $\mathbb{R}$).] □

Theorem 2.8. *Suppose a group G acts on an $\mathbb{R}$-tree (X, d), and H is a subgroup of finite index in G. Then the action of H on X by restriction is of the same type as that of G. If the action of G is minimal, so is that of H.*

Proof. If the action is of type I, any point fixed by G is fixed by H, so the action of H is of type I. Suppose the action of G is of type II. Let $C = \text{core}_G(H)$, so C is normal and of finite index in G. If H has a fixed point, so does its subgroup C, and the set X^C of points fixed by C is a subtree of X. Further, G/C acts by isometries on X^C, and since G/C is finite, the action is bounded, hence trivial, so has a fixed point by Corollary 2.2. This point is then a fixed point for the action of G, a contradiction, so the action of H is of type II.

In the remaining cases, $N \neq G$ (as usual, N is the set of non-hyperbolic elements of G). If $g \in G \setminus N$, then $g^r \in G \setminus N$ and $A_{g^r} = A_g$ for $r \neq 0$, by 3.1.7, and we can find $r > 0$ such that $g^r \in H$. Thus the axis of every hyperbolic element of G is also the axis of some hyperbolic element of H.

Thus if there is a single axis for the action of G, the same is true for the action of H, so if G has action of type III, so does H, by the remark preceding the theorem.

In Case IV, $B = \emptyset$, where $B = \bigcap_{g \in G} A_g$, and it follows that $A = \bigcap_{g \in G \setminus N} A_g = \emptyset$. Otherwise, A is a linear G-invariant subtree, and by Lemma 3.2.10 $A \subseteq A_h$ for all $h \in G$, so $A \subseteq B$ and $B \neq \emptyset$, a contradiction. From what we have proved, $A = \bigcap_{g \in H \setminus N} A_g$, so $\bigcap_{g \in H} A_g \subseteq A = \emptyset$. Since H fixes the end fixed by G, the action of H is of type IV by Theorem 3.2.6.

By Corollary 3.2.4, an action on an $\mathbb{R}$-tree is irreducible if and only if $A_g \cap A_h = \emptyset$ for some hyperbolic elements g, h. Since G and H have the same axes for hyperbolic elements, if the action of G is of type V, so is that of H.

If the action of G is minimal, it cannot be of type II by Prop. 3.4.4, and if it is of type I, X consists of a single point and the action of H is certainly minimal. Otherwise, by the proof of Lemma 3.4.2, $X = \bigcup_{g \in G \setminus N} A_g$, so by what has been proved, $X = \bigcup_{g \in H \setminus N} A_g$, and again by the proof of Lemma 3.4.2, the action of H is minimal. □

The next result investigates the inherited action of a normal subgroup by restriction on an $\mathbb{R}$-tree, and its connection with the original action.

Lemma 2.9. *Suppose a group G acts on an $\mathbb{R}$-tree (X, d) and H is a normal subgroup of G.*

(1) *If the action of G is of type III or V and $H \subseteq N$, then H fixes some point of X.*
(2) *If $N \neq G$ and H is not contained in N, then the action of H on X is of the same type as that of G.*
(3) *If $N \neq G$, $H \subseteq N$ and H fixes no point of X, then the action of G on X is of type IV.*

Proof. (1) Let $g \in G$ and $h \in H$. Since H is normal, h, ghg^{-1} and $ghg^{-1}h$ are in H, so are in N. By Corollary 3.2.4, $A_h \cap A_{ghg^{-1}} \neq \emptyset$, that is, $A_h \cap gA_h \neq \emptyset$. By the remark preceding Prop. 3.2.7, $A_h \cap A_g \neq \emptyset$. Thus h fixes some point of A_g.

If the action of G is non-abelian (Type III(a) or V), then by Corollary 3.2.4, there exist g_1, $g_2 \in G$ such that $A_{g_1} \cap A_{g_2} = \emptyset$. If $h \in H$, then h fixes some point of A_{g_1} and some point of A_{g_2}, so fixes pointwise the segment having these two points as endpoints. This segment contains the bridge between A_{g_1} and A_{g_2}. Thus H fixes every point of this bridge.

If the action of G is of type III(b), then $B = \bigcap_{g \in G} A_g \neq \emptyset$ is a linear G-invariant subtree, and by Lemma 2.7, H fixes B pointwise.

(2) In this case the action of H is of type III, IV or V. If the action is of type III there is a single axis $A = A_h$ for all $h \in H \setminus N$, (remark preceding 2.8). If $g \in G$, then $A = A_h = A_{ghg^{-1}} = gA_h$, by Lemma 3.1.7, hence A is a linear G-invariant subtree. By the remark preceding 2.8, the action of G is of type III.

If the action of H is of type IV then H, and so G, has more than one axis, hence the action of G is of type IV or V. Let $h \in H \setminus N$, and let $g \in G \setminus N$. Since the action of H is of type IV, A_h and $A_{ghg^{-1}}$ have a common end by Lemma 3.1.9, in particular $A_h \cap A_{ghg^{-1}} \neq \emptyset$, that is, $A_h \cap gA_h \neq \emptyset$. By the remark preceding Prop. 3.2.7, $A_g \cap A_h \neq \emptyset$. By Lemma 3.2.11, the action of G is not of type V, so it is of type IV.

If the action of H is of Type V, then by Corollary 3.2.4 there are a pair of disjoint axes of hyperbolic elements for the action of H, so for the action of

G, hence the action of G is of type V.

(3) In this case the action of G is not of type I or II since $N \neq G$, and is not of type III or V by (1). □

If G acts on an $\mathbb{R}$-tree and K is a normal subgroup of G, then in general the quotient G/K will not act on X. However, one can construct a kind of quotient space for the action of K, and, following [148], we shall show that, if K has bounded action, this quotient space is an $\mathbb{R}$-tree.

If (X, d) is any $\mathbb{R}$-metric space, and a group G acts as isometries, the quotient space X/G (in the usual sense) has a pseudometric $\bar{d}$ given by

$$\bar{d}(\langle x \rangle, \langle y \rangle) = \inf_{g,h \in G} d(gx, hy) = \inf_{g \in G} d(x, gy),$$

where $\langle x \rangle$ denotes the orbit Gx. (To check this is an easy exercise.) Let (Z, d') be the corresponding metric space. We refer to Z as X/G *made Hausdorff*, and denote Z by $\widehat{X/G}$. We denote the equivalence class of $\langle x \rangle$ in $\widehat{X/G}$ by $[x]$. Thus $[x] = [y]$ if and only if $\inf_{g \in G} d(x, gy) = 0$, and $d'([x], [y]) = \inf_{g \in G} d(x, gy)$. We have a canonical projection map $p : X \to \widehat{X/G}$, $x \mapsto [x]$, and p is continuous; one reason for this is that $d'([x], [y]) \leq d(x, y)$. If Y is a G-invariant subspace of X, then denoting the canonical projection by p_Y, there is an isometric embedding $\widehat{Y/G} \to \widehat{X/G}$ given by $p_Y(x) \mapsto p_X(x)$. This is again easily checked. We use this to identify $\widehat{Y/G}$ with a subspace of $\widehat{X/G}$. Thus if $y \in Y$, we can use the notation $[y]$ unambiguously for $p_X(y)$ and $p_Y(y)$, and we can denote the metric in both $\widehat{Y/G}$ and $\widehat{X/G}$ by d'.

Remark. It can be shown that if (X, d) is complete then so is $(\widehat{X/G}, d')$. See [148; Prop. 1.2].

Theorem 2.10. *Suppose a group G has a bounded action on an $\mathbb{R}$-tree (X, d). Then $(\widehat{X/G}, d')$ is an $\mathbb{R}$-tree.*

Proof. As noted in the proof of 2.7, the action of G extends to an action on the metric completion $\bar{X}$, which is an $\mathbb{R}$-tree, and passing from X to $\bar{X}$ does not change the Lyndon length function L_x, for any $x \in X$. Thus the action on $\bar{X}$ is bounded. By Lemma 2.5, there is a fixed point, say x_0, for the action of G on $\bar{X}$. For simplicity write d for the metric in $\bar{X}$ and d' for the metric in $\widehat{\bar{X}/G}$.

We show that $\widehat{\bar{X}/G}$ is 0-hyperbolic with respect to $t = [x_0]$. This means that, if $x = [u]$, $y = [v]$, $z = [w] \in \widehat{\bar{X}/G}$, then $x \cdot y$, $x \cdot z$, $y \cdot z$ is an isosceles triple, where $x \cdot y$ means $(x \cdot y)_t$, etc. (see §1.2). Suppose $x \cdot y \geq m$, $x \cdot z \geq m$, where $m \in \mathbb{R}$; we show that $y \cdot z \geq m$. The first inequality, $x \cdot y \geq m$, implies that $d(u, x_0) + d(v, x_0) \geq d'(x, y) + 2m$. Since $d(u, x_0) = d(gu, gx_0) = d(gu, x_0)$ for all $g \in G$, it follows that $d'(x, t) = d(u, x_0) = d(gu, x_0)$ for every $g \in G$, and similarly $d'(y, t) = d(gv, x_0)$ for all $g \in G$. Let ϵ be a positive real number. We can find $g \in G$ such that $d(u, gv) < d'(x, y) + \epsilon$, and since $d(v, x_0) = d(gv, x_0)$, we conclude that $(u \cdot gv)_{x_0} \geq m - (\epsilon/2)$. Similarly, we can find $h \in G$ such that $(u \cdot hw)_{x_0} \geq m - (\epsilon/2)$. Since $\hat{X}$ is 0-hyperbolic, it follows that $(hw \cdot gv)_{x_0} \geq m - (\epsilon/2)$. This means that

$$d'(t, y) + d'(t, z) \geq d(gv, hw) + 2m - \epsilon \geq d'(y, z) + 2m - \epsilon.$$

This is true for all $\epsilon > 0$, so $d'(t, y) + d'(t, z) \geq d'(y, z) + 2m$, that is, $y \cdot z \geq m$. It follows that $y \cdot z \geq \min\{x \cdot y, x \cdot z\}$, and on permuting x, y, z in this argument, that $x \cdot y$, $x \cdot z$, $y \cdot z$ is an isosceles triple, as required. Thus $\widehat{\bar{X}/G}$ is 0-hyperbolic with respect to t, so with respect to any point.

To finish the proof, X, being an $\mathbb{R}$-tree, is connected, so $\widehat{X/G} = p_X(X)$ is connected since p_X is continuous. As a subspace of $\widehat{\bar{X}/G}$, $\widehat{X/G}$ inherits the property of being 0-hyperbolic. By Lemma 2.4.13, $(\widehat{X/G}, d')$ is an $\mathbb{R}$-tree. □

If (X, d) is an $\mathbb{R}$-tree and G is a countable group acting on (X, d), a general criterion for $(\widehat{X/G}, d')$ to be an $\mathbb{R}$-tree is given in Theorem 1 of [90].

Lemma 2.11. *Let G act on an $\mathbb{R}$-tree (X, d), and let $x \in X$. Let K be a normal subgroup of G with bounded action. Then G/K acts as isometries on the $\mathbb{R}$-tree $(\widehat{X/K}, d')$, and $L_{[x]}(gK) = \inf_{k \in K} L_x(kg)$ for $g \in G$, $x \in X$.*

Proof. For $g \in G$, define an action of G by $g[x] = [gx]$. This is clearly an action provided we show it is well-defined. If $[x] = [y]$ then

$$\inf_{k \in K} d(gx, kgy) = \inf_{k \in K} d(x, g^{-1}kgy) = \inf_{k \in K} d(x, ky) = 0$$

so $[gx] = [gy]$, and the action is well-defined. Clearly $k[x] = [x]$ for $k \in K$, so there is an induced action of G/K, given by $(gK)[x] = [gx]$. This is an action by isometries, for

$$\begin{aligned} d'((gK)[x], (gK)[y]) &= d'([gx], [gy]) = \inf_{k \in K} d(gx, kgy) \\ &= \inf_{k \in K} d(x, g^{-1}kgy) = \inf_{k \in K} d(x, ky) = d'([x], [y]). \end{aligned}$$

Finally,

$$L_{[x]}(gK) = d'([x], [gx]) = \inf_{k \in K} d(x, kgx) = \inf_{k \in K} L_x(kg).$$

□

Lemma 2.12. *Let G act on an $\mathbb{R}$-tree (X,d) with $N \neq G$ and let $x \in X$. Let $H_x = \{a \in G \mid L_x(ag) = L_x(g)$ for all $g \in G \setminus N\}$, and let $K = \mathrm{core}_G(H_x)$, the maximal normal subgroup of G having bounded action on X (by Lemma 2.7). Then*

(1) *there is no non-trivial normal subgroup of G/K with bounded action on $\widehat{X/K}$;*

(2) *the action of G/K on $\widehat{X/K}$ is the same type as that of G on X. Further, if the action of G is of type III(a), so is that of G/K.*

Proof. (1) Suppose there is a non-trivial normal subgroup of G/K having bounded action on $\widehat{X/K}$, say M/K, where M is normal in G. Let $x \in X$. There are real constants r, s such that $L_x(k) \leq r$ for all $k \in K$, and $L_{[x]}(gK) \leq s$ for all $g \in M$. Suppose $g \in M$, let $l = L_{[x]}(gK)$ and let ϵ be a positive real number. By Lemma 2.11, there exists $k \in K$ such that $L_x(kg) < l + \epsilon$. Then $L_x(g) \leq L_x(k^{-1}) + L_x(kg) \leq r + l + \epsilon \leq r + s + \epsilon$. This is true for all $\epsilon > 0$, so $L_x(g) \leq r + s$ for all $g \in M$. Therefore M is a normal subgroup of G having bounded action on X, and $K \leq M$, so $K = M$ by the maximality of K. This means $M/K = 1$, a contradiction.

(2) Note that $K = \mathrm{core}_G(H_y)$ for every $y \in X$, so $K \subseteq H_y$, and by definition of H_y and Lemma 2.11, $L_{[y]}(gK) = L_y(g)$ for all $g \in G$. By Cor. 3.1.5, the hyperbolic length of gK for the action on $\widehat{X/K}$ is equal to the hyperbolic length of g for the action on X. In particular, if g is hyperbolic, so is gK. It follows that the action of G is of type I (resp. type II) if and only if the action of G/K is of type I (resp. type II). (A point $x \in X$ is a fixed point for G if and only if $L_x(g) = 0$ for all $g \in G$, which is equivalent to saying that $[x] \in \widehat{X/K}$ is a fixed point for G/K.) Also, since dihedral actions are characterised by the hyperbolic length function, if the action of G is dihedral (III(a)), so is that of G/K.

The set of axes of hyperbolic elements for the action of G on X is mapped isomorphically to $\widehat{X/K}$ by the projection p_X, by Lemma 2.7. If g is a hyperbolic element of G, then $(gK)p_X(A_g) \subseteq p_X(A_g)$, so $p_X(A_g)$ is invariant under the cyclic subgroup generated by gK, and is metrically isomorphic to $\mathbb{R}$. By 3.2.10, $p_X(A_g) \subseteq A_{gK}$, and by the exercise preceding Lemma 2.7, $p_X(A_g) = A_{gK}$. It follows that the actions of G and G/K are of the same type. For type III actions are characterised by the existence of a unique axis for hyperbolic elements, by the remark preceding Theorem 2.8; Type IV actions are characterised by the fact that the action is non-trivial, abelian and $\bigcap_{g \in G \setminus N} A_g = \emptyset$ (using 3.2.10 as in Theorem 2.8, Case IV), and Type V is characterised by the existence of two hyperbolic elements whose axes do not intersect, by Corollary 3.2.4. □

After this lengthy buildup, it is time to introduce the classes of groups which are of interest.

Definition. A group is hypercentral if every non-trivial quotient group has non-trivial centre. A group is hypercyclic (resp. hyperabelian) if every non-trivial quotient group has a non-trivial normal subgroup which is cyclic (resp. abelian).

Clearly hypercentral implies hypercyclic implies hyperabelian. There are equivalent formulations of the definitions. By an ascending series in a group G we mean a family $\{H_\alpha | \alpha \leq \beta\}$ of subgroups of G indexed by all ordinals less than or equal to some ordinal β, satisfying:

(a) $\alpha_1 \leq \alpha_2 \leq \beta$ implies H_{α_1} is a subgroup of H_{α_2};
(b) $H_0 = \langle 1 \rangle$ and $H_\beta = G$;
(c) H_α is a normal subgroup of $H_{\alpha+1}$ for $\alpha < \beta$;
(d) if α is a limit ordinal with $\alpha \leq \beta$, then $H_\alpha = \bigcup_{\gamma<\alpha} H_\gamma$.

The factors of the series are the quotients $H_{\alpha+1}/H_\alpha$ for $\alpha < \beta$. It is called a normal series if H_α is normal in G for all $\alpha \leq \beta$. It is called a central series if $H_{\alpha+1}/H_\alpha \leq Z(G/H_\alpha)$ for all $\alpha < \beta$, where $Z(K)$ denotes the centre of the group K. We leave it as an exercise to show that a group G is hyperabelian (resp. hypercyclic) if and only if it has an ascending normal series with all factors abelian (resp. cyclic). Also, G is hypercentral if and only if it has an ascending central series. Thus soluble groups are hyperabelian, supersoluble groups are hypercyclic and nilpotent groups are hypercentral.

We have finally arrived at the main result. Recall that, if $\mathfrak{X}$ is a class of groups, a group G is virtually in $\mathfrak{X}$ if it has a subgroup of finite index in $\mathfrak{X}$.

Theorem 2.13. *Let G be a group acting on an $\mathbb{R}$-tree (X, d).*
(1) *If G is hypercentral, the action is of type I, II or III(b).*
(2) *If G is virtually hypercyclic, the action is of type I, II or III.*
(3) *If G is virtually hyperabelian, the action is of type I, II, III or IV.*

Proof. By Theorem 2.8, we can assume G is hypercyclic in Case (2) and hyperabelian in Case (3). Suppose $N \neq G$ (i.e. the action is not of type I or II). By Lemma 2.7, G has a unique maximal normal subgroup with bounded action on X, which we denote by K. By Lemma 2.6, $K \leq N$, so G/K is non-trivial, hence has a non-trivial normal subgroup, say H/K, which is central in Case (1), cyclic in Case (2) and abelian in Case (3). We consider the action of G/K on $\widehat{X/K}$, which has the same type as that of G on X by Lemma 2.12(2). Denote the set of non-hyperbolic elements of this action by N', so $N' \neq G/K$.

If H/K acts non-trivially on $\widehat{X/K}$, then the action of H/K is of the same type as the action of G/K on $\widehat{X/K}$, by Lemma 2.9(2), so is of the same type as the action of G on X. In all three cases, H/K is an abelian group, so by Lemma 3.3.9(2) there is a single axis A for the action of H/K, which is therefore of Type III. Also, A is G-invariant by Lemma 3.1.7(i). If H/K is central in G/K, then there is a hyperbolic isometry in H/K, say hK, which is in the centre of G/K. Again by Lemma 3.3.9(2), $A = A_{hK}$ is contained in A_{gK} for every element $gK \in G/K$, so $\bigcup_{gK \in G/K} A_{gK} \neq \emptyset$. It follows by Lemma 3.2.10 that the action of G/K is of type III(b), hence the action of G on X is of type III(b) by Lemma 2.12(2).

Suppose $H/K \subseteq N'$ and H/K fixes no point of $\widehat{X/K}$. By Lemma 2.9(3), the action is of type IV. If H/K is central, it fixes each axis of $\widehat{X/K}$ pointwise by Lemma 3.3.9(2), so this possibility cannot occur in this case. Also, this possibility cannot occur if H/K is cyclic, since then H/K would fix a point of $\widehat{X/K}$ by Corollary 2.2.

Finally, if H/K fixes a point of $\widehat{X/K}$, the action of H/K is bounded (the Lyndon length function based at a fixed point is identically zero), which is impossible by Lemma 2.12(1). □

Concerning Part (1) of 2.13, we can only conclude that an action of a virtually hypercentral group on an $\mathbb{R}$-tree must be of type I, II or III. It seems reasonable to prove the three parts of Theorem 2.13 by a uniform method, but we should note an alternative proof of Part (3). The class of hyperabelian groups is subgroup closed; this can be seen from the alternative characterisation of hyperabelian groups by ascending series, or directly as follows. Suppose G is hyperabelian and $M \trianglelefteq H \leq G$ with $M \neq H$. The set $\Omega = \{K \trianglelefteq G | \ K \cap H \leq M\}$ is inductively ordered by inclusion, and non-empty ($\langle 1 \rangle \in \Omega$), so has a maximal element, say K_0, by Zorn's Lemma. Since $M \neq H$, $K_0 \neq G$, so we can find a non-trivial abelian normal subgroup of G/K_0, say L/K_0, where $K_0 \leq L \trianglelefteq G$. It follows that $(L \cap H)M/M$ is a non-trivial abelian normal subgroup of H/M, hence H is hyperabelian.

The free group of rank 2 is not hyperabelian, for an abelian subgroup is cyclic, and it is easy to see that no non-trivial cyclic subgroup is normal. Thus a hyperabelian group has no free subgroup of rank 2, so cannot have an action of Type V by Prop. 3.3.7. Invoking 2.8, we obtain Part (3) of Theorem 2.13.

It follows as a special case that an action of a virtually soluble group on an $\mathbb{R}$-tree must be of type I, II, III or IV. This was first proved in [143]. Also, as a special case of 2.13(2), an action of a virtually supersoluble group must be of type I, II or III. The next result, which is Prop. 3.8 in [42], improves on this.

Proposition 2.14. *An action of a virtually polycyclic group on an $\mathbb{R}$-tree must be of type I or III.*

Proof. It suffices by 2.8 to show that an action of a polycyclic group on an $\mathbb{R}$-tree is of type I or III. Such a group G has an ascending series $\langle 1 \rangle = G_0 \trianglelefteq G_1 \trianglelefteq \ldots G_{n-1} \trianglelefteq G_n = G$ for some natural number n, with G_i/G_{i-1} cyclic for

$1 \le i \le n$. It suffices by induction on n to show that, if G is a group, $H \trianglelefteq G$, G/H is cyclic, and any action of H on an $\mathbb{R}$-tree is of type I or III, then any action of G on an $\mathbb{R}$-tree must also be of type I or III. Suppose G acts on an $\mathbb{R}$-tree (X, d). If the action of H by restriction is of type III, the action of G is of type III by 2.9(2). Suppose the action of H is of type I; then the set X^H of fixed points of H is a subtree of X, and G acts on X^H since $H \trianglelefteq G$, with the same hyperbolic length function as that for the action on X (3.1.7(4)). The action of G induces an action of G/H on X^H. If the action of G/H is non-trivial, then since G/H is cyclic, there is a unique axis for the hyperbolic elements of G/H, by Lemma 3.1.7(3). This is a linear subtree of X, invariant under the action of G on X, which is non-trivial, so this action is of type III by the remark preceding 2.8. If the action of G/H on X^H is trivial, there is a fixed point for the action by Corollary 2.2. This point is then a fixed point for the action of G on X, which is therefore of type I. □

One can also say something about inherited actions of finitely generated groups.

Lemma 2.15. *If G acts on an $\mathbb{R}$-tree (X, d) with $N \neq G$, then there is a subgroup H with at most two generators whose action on X by restriction is of the same type as that of G. If the action of G is of type III(a), such a subgroup H can be found with action also of type III(a).*

Proof. If the action is of type III(a) then by Prop. 3.2.9 there is a homomorphism $\rho : G \to \mathrm{Isom}(\mathbb{R})$ whose image contains a non-trivial translation and a reflection. Choose two elements of g, one mapping to the translation and the other to the reflection, and let H be the subgroup they generate. Then the action of H is of type III(a) by Prop. 3.2.9.

If the action is of type III(b), choose an element $g \in G \setminus N$. Then g acts as a non-trivial translation on $B = \bigcap_{g \in G} A_g$, by Theorem 3.2.6. Let H be the infinite cyclic group generated by g. Then H has action of type III(b) by Theorem 3.2.6.

If the action is of type IV, there exist g, $h \in G \setminus N$ such that $A_g \neq A_h$, by Theorem 3.2.6. Take H to be the subgroup they generate. If the action of H were of type III(b), then all axes of hyperbolic elements in H would coincide (remark preceding 2.8), so the action of H is of type IV. (The action of G, so of any subgroup, is abelian.)

If the action of G is of type V, there exist g, $h \in G \setminus N$ such that $A_g \cap A_h = \emptyset$ by Corollary 3.2.4. Let H be the subgroup generated by g and h. Then the action of H is of type V. □

Recall that, if $\mathfrak{X}$ is a class of groups, the class $\mathrm{L}\mathfrak{X}$ of groups which are locally $\mathfrak{X}$ is the class of all groups G such that every finite subset of G is contained in a subgroup of G which belongs to $\mathfrak{X}$. If $\mathfrak{X}$ is subgroup closed, then $G \in \mathrm{L}\mathfrak{X}$ if and only if every finitely generated subgroup of G is in $\mathfrak{X}$.

Theorem 2.16.

(1) *An action of a locally nilpotent group on an $\mathbb{R}$-tree must be of type I, II or III(b).*

(2) *An action of a locally polycyclic group on an $\mathbb{R}$-tree must be of type I or III.*

(3) *An action of a locally soluble group on an $\mathbb{R}$-tree must be of type I, II, III or IV.*

Proof. Immediate from 2.13, 2.14 and 2.15. □

In Theorems 2.13 and 2.16, we cannot omit type II from the statement of parts (1) and (3), for there are actions of abelian groups of type II, using the example after Theorem 3.2.6. Any abelian group G which can be written as the union of a strictly ascending chain of subgroups, $G = \bigcup_{i=1}^{\infty} G_i$, where $G_1 \lneqq G_2 \lneqq \dots$, has such an action.

Also, in connection with Part (3) in 2.13 and 2.16, there are finitely generated soluble groups having an action of type IV on an $\mathbb{R}$-tree. For we saw after the proof of Theorem 3.4.5 that the group $G_n = \langle\ t,\ a \mid tat^{-1} = a^n\ \rangle$ has a non-trivial action of end type on a $\mathbb{Z}$-tree, say X_n. Arguing as in the example after Theorem 3.2.6, the extension to an action on $\mathbb{R} \otimes_{\mathbb{Z}} X_n$ is of type IV. Further, G_n is soluble, in fact it is metabelian. For using the relations $ta = a^n t$ and $at^{-1} = t^{-1}a^n$, we can write any element of G_n in the form $t^{-m}a^k t^l$, where k is an integer and m, l are non-negative integers. This is in the kernel of the natural homomorphism $\phi : G_n \to \mathbb{Z}$, sending t to 1 and a to 0, if and only if $m = l$. It follows that $\mathrm{Ker}(\phi) = \bigcup_{m\in\mathbb{Z}} t^m \langle a \rangle t^{-m}$. Since $t^m\langle a\rangle t^{-m} \subseteq t^{m-1}\langle a \rangle t^{-m+1}$ for all $m \in \mathbb{Z}$, it follows that $\mathrm{Ker}(\phi)$ is locally cyclic, so abelian. Hence G_n is an extension of the abelian group $\mathrm{Ker}(\phi)$ by an infinite cyclic group. (In fact, G_n can be described as a split extension of the additive group of the ring $\mathbb{Z}[1/n]$ by an infinite cyclic group, acting as multiplication by n; we leave this as an exercise.)

In §3.4, to see that G_n has a non-trivial action of end type, we quoted results in a paper by K.S. Brown [23]. In fact, actions on an $\mathbb{R}$-tree of type IV are of interest because of a connection, discovered by Brown and described in [23], with a group theoretic invariant called the Bieri-Neumann-Strebel Invariant (see [17]). Let G be a finitely generated group. The set $\mathrm{Hom}(G, \mathbb{R})$ is a vector space over $\mathbb{R}$ of dimension n, where n is the number of infinite cyclic factors in the decomposition of $G/[G, G]$ as a direct sum of cyclic groups. The multiplicative group $\mathbb{R}^+$ of positive real numbers acts on $\mathrm{Hom}(G, \mathbb{R}) \setminus \{0\}$. If $\phi : G \to \mathbb{R}$ is a non-zero homomorphism, and r is a positive real number, $r\phi$ sends $g \in G$ to $r\phi(g)$. Under the identification of $\mathrm{Hom}(G, \mathbb{R})$ with $\mathbb{R}^n$, this corresponds to scalar multiplication (or "homothety"), so the quotient space $(\mathrm{Hom}(G,\mathbb{R}) \setminus \{0\})/\mathbb{R}^+$ is a sphere of dimension $(n-1)$ which we denote by $S(G)$.

Now consider an action G on an $\mathbb{R}$-tree of type IV (end type with $N \neq G$). By Prop. 3.2.5 there is an associated homomorphism $\tau : G \to \mathbb{R}$ such that $\ell(g) = |\tau(g)|$ for all $g \in G$, where ℓ denotes hyperbolic length. It is given by $\tau(g) = \delta(gx) - \delta(x)$, for every $x \in X$, where δ is an end map towards the fixed end, and this is independent of the choice of δ. The homomorphism $h = -\tau$ is an element of $\mathrm{Hom}(G,\mathbb{R}) \setminus \{0\}$, since $N \neq G$, also uniquely determined by the action. Such a homomorphism h will be called an exceptional homomorphism. If $E(G)$ denotes the image of the set of exceptional homomorphisms in $S(G)$, the complement $S(G) \setminus E(G)$ is the Bieri-Neumann-Strebel Invariant of G, denoted by $\Sigma(G)$. It is shown in Brown's paper that this coincides with the set $\Sigma(G)$ as defined by Bieri, Neumann and Strebel. These authors make good use of this invariant. For example, they show that If G is a finitely presented group whose abelianisation has at least two infinite cyclic factors, then either G contains a free group of rank 2, or G has a finitely generated normal subgroup N with G/N infinite cyclic. It would be interesting to have a proof of this using entirely arboreal methods. See §1.6 of Shalen's article [129] for further discussion. As Brown points out, this gives one generalisation of the Bieri-Neumann-Strebel Invariant, by considering non-trivial actions of end type on Λ-trees for a fixed ordered abelian group Λ. One still has a notion of exceptional homomorphism in $\mathrm{Hom}(G,\Lambda) \setminus \{0\}$.

Remark. If ℓ is the hyperbolic length function for an action of a group G on a Λ-tree and $h : G \to \Lambda$ is a group homomorphism, such that $\ell(g) = |h(g)|$ for all $g \in G$, where ℓ is hyperbolic length, then the only other homomorphism from G to Λ with this property is $-h$. We shall not prove this, but note that it is a consequence of (1.4) in [2].

3. The Action of the Special Linear Group

Here we describe an interesting method, due to Serre [127; Ch.II §1] in the case of simplicial trees ($\Lambda = \mathbb{Z}$) and to Morgan and Shalen [106] in general, of constructing an action of $\mathrm{SL}_2(F)$ on a Λ-tree, where F is a field with a valuation. The method was obtained in the case $\Lambda = \mathbb{Z}$ by studying Bruhat-Tits buildings, and the method in general is a generalisation of this (although having obtained the set of points of the tree, we have to construct a metric rather than the set of edges of a simplicial tree, and verify that we have a Λ-tree). In fact, we shall construct an action of $\mathrm{GL}_2(F)$ whose restriction to $\mathrm{SL}_2(F)$ is without inversions. In order to understand this section, a knowledge of the basics of module theory and of $1.4 is needed.

Let F be a field and let v be a valuation on F with valuation ring R and value group Λ. Let V be the vector space F^2. An R-lattice in V is a finitely generated R-module $L \subseteq V$ which spans V as an F-module. An R-lattice is a free R-module by Corollary 1.4.6. Since F is the field of fractions of R, a subset of L is R-linearly independent if and only if it is F-linearly independent. Consequently, L is a free R-module of rank 2.

Remark. If L, L' are lattices, we can find a basis e_1, e_2 for L and a, $b \in F^*$ (the multiplicative group of F) such that ae_1, be_2 span L'. For since L spans V, L' is finitely generated and F is the field of fractions of R, there exists $\gamma \in R \setminus \{0\}$ such that $\gamma L' \subseteq L$. By Corollary 1.4.5, there is a basis e_1, e_2 for L and α, $\beta \in R$ such that αe_1, βe_2 span $\gamma L'$ as R-module. Put $a = \alpha/\gamma$, $b = \beta/\gamma$. Then ae_1, be_2 span L' as R-module, so span V as F-module, hence a, $b \in F^*$.

Given lattices L, L', we define $L \sim L'$ to mean that $L' = aL$ for some $a \in F^*$. This is clearly an equivalence relation on the set of R-lattices in V and we denote the set of equivalence classes by X_v. We shall define a Λ-metric on X_v and show that this gives a Λ-tree.

If L, L' are lattices, we write $L' \leq L$ to mean that $L' \subseteq L$ and L/L' is a cyclic R-module. If $L' \leq L$, by Corollary 1.4.5 we can find bases e_1, e_2 for L and ae_1, be_2 for L', where a, $b \in R$, and $L/L' \cong (R/Ra) \oplus (R/Rb)$. If a is not a unit in R, then $a \in \mathfrak{m}$, where $\mathfrak{m}$ is the maximal ideal of R, so $Ra \subseteq \mathfrak{m}$. Now applying the functor $(R/\mathfrak{m}) \otimes_R -$ to the short exact sequence $0 \to Ra \to R \to R/Ra \to 0$, and using the usual isomorphism $(R/\mathfrak{m}) \otimes_R R \cong R/\mathfrak{m}$, we see that $(R/\mathfrak{m}) \otimes_R (R/Ra) \cong R/\mathfrak{m}$. Similar remarks apply to b, so if neither a nor b is a unit, then $(R/\mathfrak{m}) \otimes_R (L/L')$ is a vector space of dimension 2 over $R/\mathfrak{m}$, contradicting that L/L' is cyclic. Thus one of a, b is a unit in R, say b, and $L/L' \cong R/Ra$. Since L' spans V as F-module, $a \neq 0$, so $v(a) \in \Lambda$ and $v(a) \geq 0$. Further, a is determined up to a unit in R, so $v(a)$ depends only on L/L' and will be denoted by $|L/L'|$.

Lemma 3.1. *Given two lattices L and L', there is a unique lattice L_0 such that $L_0 \sim L'$ and $L_0 \leq L$.*

Proof. By the remark above, we can find a basis e_1, e_2 for L and a, $b \in F^*$ such that ae_1, be_2 span L' as R-module. We may assume by renaming that $v(a) \leq v(b)$. Put $L_0 = a^{-1}L'$, so $L_0 \sim L'$. Then L_0 is spanned by e_1 and ce_2, where $c = a^{-1}b$, and $v(c) = v(b) - c(a) \geq 0$, so $c \in R$. Hence $L_0 \subseteq L$ and $L/L_0 \cong R/Rc$, so $L_0 \leq L$, showing existence.

If L_1 is a lattice with $L_1 \sim L'$, then $L_1 \sim L_0$, so there exists $f \in F^*$ such that $L_1 = fL_0$. If $v(f) < 0$, then $fe_1 \notin L$ and $L_1 \not\subseteq L$. If $v(f) > 0$ then $L/L_1 \cong R/fR \oplus R/fcR$, and f, cf belong to the maximal ideal $\mathfrak{m}$ of R. From the observations preceding the lemma, L/L_1 is not a cyclic R-module. Thus if

$L_1 \leq L$, then $v(f) = 0$, i.e. f is a unit in R, so $L_1 = L_0$. This concludes the proof. □

It follows that, in Lemma 3.1, $|L/L_0|$ depends only on the pair L, L'; and it will be denoted by $d(L, L')$.

Lemma 3.2. *Let x, $y \in X_v$. Then there exist lattices L, L', representing x, y respectively, such that $L' \leq L$. Further, $d(L, L')$ depends only on the pair (x, y), not on the choice of L and L'.*

Proof. The first assertion follows at once from Lemma 3.1. Let $L/L' \cong R/Rb$, where $b \in R$ and suppose $L_1 \sim L$, $L_1' \sim L'$, $L_1' \leq L_1$. Then $L_1 = aL$ for some $a \in F^*$. The mapping $L \to aL$, $x \mapsto ax$ is an R-isomorphism, so $L/L' \cong aL/aL' = L_1/aL'$, hence $aL' \leq L_1$, so by the uniqueness statement in Lemma 3.1, $aL' = L_1'$. Thus $L_1/L_1' \cong R/Rb$, so $|L_1/L_1'| = v(b) = |L/L'|$. □

If x, $y \in X_v$, we can therefore define $d(x, y) = d(L, L')$, where L, L' are representative lattices for x, y respectively, such that $L' \leq L$. After some preliminaries, we shall prove that (X_v, d) is a Λ-tree.

Lemma 3.3. *Let L_0, L_1 and L be R-lattices in V such that $L_i \leq L$ for $i = 1$, 2 and $L_0 + L_1 = L$. Then $d(L_0, L_1) = d(L, L_0) + d(L, L_1)$.*

Proof. This is clear if $L = L_0$ or $L = L_1$, so assume this is not the case. We can write $L/L_0 \cong R/Ra$, $L/L_1 \cong R/Rb$ for some non-units a, $b \in R$, that is, a, $b \in \mathfrak{m}$, where $\mathfrak{m}$ is the maximal ideal of R. Now the R-homomorphism $f : L \to (L/L_0) \oplus (L/L_1)$ given by $z \mapsto (z + L_0, z + L_1)$ is surjective by the Chinese Remainder Theorem. (Explicitly, since $L = L_0 + L_1$, given x, $y \in L$, we can find $x' \in L_0$, $y' \in L_1$ so that $x - x' \in L_1$ and $y - y' \in L_0$. Then, putting $z = x - x' + y - y'$, $(x + L_0, y + L_1) = (z + L_0, z + L_1)$.) The kernel of f is $L_0 \cap L_1$. Also, $\mathfrak{m}(R/Ra) = \mathfrak{m}/Ra$ and $\mathfrak{m}(R/Rb) = \mathfrak{m}/Rb$, hence $f(L)/\mathfrak{m}f(L) \cong (R/\mathfrak{m}) \oplus (R/\mathfrak{m})$. Composing f with the projection $f(L) \to f(L)/\mathfrak{m}f(L)$ gives a surjection $L \to f(L)/\mathfrak{m}f(L)$ with kernel $(L_0 \cap L_1) + \mathfrak{m}L$, which we denote by K. This gives a short exact sequence

$$K/\mathfrak{m}L \rightarrowtail L/\mathfrak{m}L \overset{\bar{f}}{\twoheadrightarrow} f(L)/\mathfrak{m}f(L).$$

But $L/\mathfrak{m}L$ and $f(L)/\mathfrak{m}f(L)$ are both vector spaces of dimension 2 over $R/\mathfrak{m}$, so $\bar{f}$ is an isomorphism. Hence $K = \mathfrak{m}L$, that is, $L_0 \cap L_1 \subseteq \mathfrak{m}L$.

Choose $e_1 \in L_0$, $e_0 \in L_1$ such that $L = L_0 + Re_0 = L_1 + Re_1$. Then e_0, e_1 generate L mod $L_0 \cap L_1$, so generate L (corollary after Nakayama's Lemma in [81], after Proposition 7.18). Since L is a lattice, e_0, e_1 is a basis for L. Since $e_1 \in L_0$ and Ra is the annihilator of (L/L_0), L_0 is generated by ae_0 and e_1. Similarly, L_1 is generated by e_0 and be_1. Hence aL_1 is generated by ae_0 and abe_1, so $aL_1 \subseteq L_0$ and $L_0/aL_1 \cong R/Rab$. It follows that $d(L_0, L_1) = v(ab) = v(a) + v(b) = |L/L_0| + |L/L_1| = d(L, L_0) + d(L, L_1)$, as claimed. □

Remark. If L, L' are lattices then it follows at once from the definition that $L + L'$ is a lattice. Also, if L, L' are lattices with $L' \leq L$ and L'' is a lattice with $L' \subseteq L'' \subseteq L$, then $L' \leq L'' \leq L$. For L/L'' is cyclic, being a homomorphic image of L/L', and from the observations before Lemma 3.1, we can write $L/L' \cong R/Ra$, $L/L'' \cong R/Rb$ for some non-zero a, $b \in R$, and R/Ra is a homomorphic image of R/Rb. Hence Ra annihilates R/Rb, so $Ra \subseteq Rb$ and we can write $a = bc$ for some $c \in R$. Any epimorphism $R/Ra \to R/Rb$ is the canonical projection composed with multiplication by some element of R relatively prime to b, so has kernel $Rb/Ra \cong R/Rc$. Thus the short exact sequence $L''/L' \rightarrowtail L/L' \twoheadrightarrow L/L''$ is isomorphic to one of the form $R/Rc \rightarrowtail R/Ra \twoheadrightarrow R/Rb$ and $L''/L' \cong R/Rc$, so $L' \leq L''$. Further, $|L/L'| = v(a) = v(b) + v(c) = |L/L''| + |L''/L'|$. In particular, $|L/L'| \geq |L/L''|$.

Theorem 3.4. *The pair* (X_v, d) *is a* Λ*-tree.*

Proof. Clearly $d(x,x) = 0$ for all $x \in X_v$, and it is clear from the definition that $d(x,y) \geq 0$. If $d(x,y) = 0$ then, in Lemma 3.2, $L/L' \cong R/Rc$ where c is a unit in R, so $L = L'$, hence $x = y$.

Let x, $y \in X_v$ and choose lattices L, L' representing x, y respectively with $L' \leq L$. As we saw in the proof of Lemma 3.1, there is a basis e_1, e_2 for L and $c \in R$ such that L' is spanned by e_1 and $e_2' = ce_2$, and $v(c) = d(x,y)$. Then $cL \sim L$ and cL is spanned by ce_1 and e_2', so $cL \leq L'$ and $L'/cL \cong R/Rc$. Hence $v(c) = d(y,x)$, showing d is symmetric.

Now fix a lattice L_0 and let x_0 be the point of X_v it represents. For $x \in X_v$, let L_x denote the unique lattice representing x such that $L_x \leq L_0$. If x, $y \in X_v$, we claim that $(x \cdot y)_{x_0} = |L_0/(L_x + L_y)|$, in particular $(x \cdot y)_{x_0} \in \Lambda$. For, using Lemma 3.3 and the remark preceding the theorem:

$$\begin{aligned} 2(x \cdot y)_{x_0} &= d(x, x_0) + d(y, x_0) - d(x,y) \\ &= |L_0/L_x| + |L_0/L_y| - (|(L_x + L_y)/L_x| + |(L_x + L_y)/L_y|) \\ &= 2|L_0/(L_x + L_y)|. \end{aligned}$$

Next, we claim that, if x, y, $z \in X_v$, then two of $L_x + L_y$, $L_x + L_z$, $L_y + L_z$ are equal and contain the third. The submodules of the cyclic module L_0/L_y are linearly ordered by the correspondence theorem and Lemma 1.4.3. Hence one of $L_x + L_y$, $L_z + L_y$ contains the other. Suppose $L_z + L_y \subseteq L_x + L_y$ (the other case is similar); then $L_x + L_y = L_x + L_y + L_z$. Similarly one of $L_x + L_z$, $L_z + L_y$ equals $L_x + L_y + L_z$, and the claim follows.

We conclude from the remark preceding the theorem that $(x \cdot y)_{x_0}$, $(x \cdot z)_{x_0}$, $(y \cdot z)_{x_0}$ is an isosceles triple, hence (X_v, d) is a 0-hyperbolic Λ-metric space by the remark preceding Lemma 1.2.7.

To finish the proof, it suffices by Lemma 2.4.3 to show that, if $x \in X_v$, there is a segment in X_v with endpoints x_0 and x. As in the proof of Lemma

3.1, there is a basis e, f of L_0 and $a \in R$ such that e, af is a basis for L_x. For $\lambda \in [0, d(x_0, x)]_\Lambda$, choose $a_\lambda \in R$ such that $v(a_\lambda) = \lambda$, and let L_λ be the lattice generated by e, $a_\lambda f$, and let x_λ be the element of X_v represented by L_λ. Define $\alpha : [0, d(x_0, x)]_\Lambda \to X_v$ by $\alpha(\lambda) = x_\lambda$. If $0 \leq \lambda \leq \mu \leq d(x_0, x)$, then $Ra_\mu \subseteq R\lambda$, $L_\mu \leq L_\lambda$ and $L_\lambda / L_\mu \cong Ra_\lambda / Ra_\mu \cong R/R(a_\mu/a_\lambda)$, so $d(L_\lambda, L_\mu) = v(a_\mu/a_\lambda) = \mu - \lambda$, hence α is an isometry, and its image is the desired segment. □

The natural action of $\mathrm{GL}(V) = \mathrm{GL}_2(F)$ on V induces an action on the set of R-lattices in V, which clearly preserves the relation $\sim$ and is transitive, so $\mathrm{GL}(V)$ acts transitively on X_v. If L, L' are lattices and $L' \leq L$, then as in the proof of Lemma 3.1 we can find a basis e_1, e_2 for L and $a \in R$ such that e_1, ae_2 spans L'. If $g \in \mathrm{GL}(V)$ then ge_1, ge_2 is a basis for gL and gL' is spanned by ge_1, age_2, so $gL' \leq gL$. It follows that the action on X_v is by isometries.

Put $L_0 = R^2$ and let x_0 be the element of X_v it represents. Note that if $g \in \mathrm{GL}_2(R)$ then $gL_0 = L_0$, so $gx_0 = x_0$. If $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(F)$, define

$$v(g) = \min\{v(a), v(b), v(c), v(d)\}.$$

If h is obtained from g by an elementary R-row or column operation (see §4 in Ch.1), it is easy to see using the axioms for a valuation that $v(h) \geq v(g)$. Since the elementary operation is reversible, it follows that $v(g) = v(h)$. The next result gives the Lyndon length function based at x_0 for the action of $\mathrm{GL}(V)$ on X_v.

Lemma 3.5. *For $g \in \mathrm{GL}_2(F)$, $d(gx_0, x_0) = v(\det(g)) - 2v(g)$.*

Proof. Let a be an entry of g with $v(a)$ as small as possible, so $v(a) = v(g)$. Then $a \in F^*$ and we can write $g = ah$, where $h \in \mathrm{GL}_2(F)$ and $v(h) = 0$. By Lemma 1.4.4, h can be transformed by a sequence of elementary R-row and column operations to a diagonal matrix d, and consequently there exist p, $q \in \mathrm{GL}_2(R)$ such that $h = pdq$. As has been observed before the lemma, $v(d) = v(h) = 0$, so by further elementary operations (to interchange the rows and columns, and multiply a row by a unit of R, if necessary), we can assume $d = \begin{pmatrix} b & 0 \\ 0 & 1 \end{pmatrix}$ for some $b \in R$. Then $L_0/dL_0 \cong R/Rb$, so $dL_0 \leq L_0$. Also, $gL_0 = ahL_0 \sim hL_0$, so $gx_0 = hx_0$. Hence

$$\begin{aligned} d(x_0, gx_0) &= d(x_0, pdqx_0) = d(p^{-1}x_0, dqx_0) \\ &= d(x_0, dx_0) = |L_0/dL_0| = v(b) \\ &= v(\det(d)) \\ &= v(\det(h)) \qquad \text{(since } \det(p), \det(q) \text{ are units in } R) \\ &= v(a^{-2}\det(g)) \\ &= v(\det(g)) - 2v(a) = v(\det(g)) - 2v(g) \end{aligned}$$

and the lemma is proved. □

We shall also need to know the hyperbolic length function ℓ for this action.

Lemma 3.6. *For* $g \in \mathrm{GL}_2(F)$, $\ell(g) = \max\{v(\det(g)) - 2v(\mathrm{tr}(g)), 0\}$.

Proof. First, if $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(F)$, then $v(g) \leq \min\{v(a), v(d)\} \leq v(a + d) = v(\mathrm{tr}(g))$. If g is an inversion, then g^2 has a fixed point, and since the action of $\mathrm{GL}_2(F)$ on X_v is transitive, there is a conjugate of g, say h, such that h^2 fixes x_0. Then by Lemma 3.5, $v(\det(h^2)) = 2v(h^2) \leq 2v(\mathrm{tr}(h^2))$, so $v(\det(h)) \leq v(\mathrm{tr}(h^2))$; since trace and determinant are unchanged by conjugation, $v(\det(g)) \leq v(\mathrm{tr}(g^2))$, which means $\mathrm{tr}(g^2)/\det(g) \in R$. By Cayley-Hamilton, $g^2 = (\mathrm{tr}(g))g - (\det(g))1$, where 1 denotes the 2×2 identity matrix, hence $\mathrm{tr}(g^2) = (\mathrm{tr}(g))^2 - 2\det(g)$, so $(\mathrm{tr}(g))^2/\det(g) \in R$. It follows that $2v(\mathrm{tr}(g)) \geq v(\det(g))$. The formula for $\ell(g)$ is therefore true in this case, and we assume g is not an inversion.

Since the action is transitive, if $x \in X_v$ then $x = hx_0$ for some $h \in \mathrm{GL}_2(F)$. Then

$$\begin{aligned} d(x, gx) &= d(hx_0, hgx_0) = d(x_0, h^{-1}ghx_0) \\ &= v(\det(h^{-1}gh)) - 2v(h^{-1}gh) \qquad \text{(Lemma 3.5)} \\ &\geq v(\det(h^{-1}gh)) - 2v(\mathrm{tr}(h^{-1}gh)) \\ &= v(\det(g)) - 2v(\mathrm{tr}(g)). \end{aligned}$$

It follows from Cor. 3.1.5 that $\ell(g) \geq \max\{v(\det(g)) - 2v(\mathrm{tr}(g)), 0\}$. We have to show the reverse inequality.

Put $\delta = \det(g)$, $\tau = \mathrm{tr}(g)$. From the Cayley-Hamilton formula $g^2 = \tau g - \delta 1$, we obtain

$$\begin{aligned} v(g^2) &= \min\{v(\tau a - \delta), v(\tau b), v(\tau c), v(\tau d - \delta)\} \\ &\geq \min\{v(g) + v(\tau), v(\delta)\}. \end{aligned}$$

From the first formula for ℓ in Lemma 3.1.8 and Lemma 3.5 we obtain

$$\begin{aligned} \ell(g) &= \max\{v(\delta) - 2(v(g^2) - v(g)), 0\} \\ &\leq \max\{v(\delta) - 2v(\tau), 2v(g) - v(\delta), 0\} \\ &= \max\{v(\delta) - 2v(\tau), -d(x_0, gx_0), 0\} \\ &= \max\{v(\delta) - 2v(\tau), 0\} \end{aligned}$$

as required. □

For $g \in \mathrm{SL}_2(F)$, Lemma 3.5 gives $d(gx_0, x_0) = -2v(g) \in 2\Lambda$, so by criterion (iv) of Lemma 3.1.2 the action of $\mathrm{SL}_2(F)$ on X_v is without inversions.

Hence $g \in \mathrm{SL}_2(F)$ has a fixed point if and only if $\ell(g) = 0$, which by Lemma 3.6 happens if and only if $v(\mathrm{tr}(g)) \geq 0$. Thus g acts as a hyperbolic isometry if and only if $v(\mathrm{tr}(g)) < 0$. A similar construction to that given here applies to the group $\mathrm{SO}(n, 1)$ over the field F, given a valuation on F whose residue field is formally real, to give a Λ-tree on which $\mathrm{SO}(n, 1)$ acts (see [102]).

4. Measured Laminations

In this section it is assumed the reader is familiar with the notion of a manifold. For us, a topological n-manifold means a Hausdorff topological space such that every point has an open neighbourhood homeomorphic to an open subspace of $\mathbb{R}^n$. If M is such a manifold, a chart for M is a pair (U, ϕ), where U is an open subset of M and ϕ is a homeomorphism from U to some open subspace of $\mathbb{R}^n$. An *atlas* is a collection of charts $\{(U_i, \phi_i) \mid i \in I\}$ such that $M = \bigcup_{i \in I} U_i$. Given two charts (U_i, ϕ_i), (U_j, ϕ_j), we have a corresponding *transition map*

$$\phi_{ij} = \phi_j \circ \phi_i^{-1} : \phi_i(U_i \cap U_j) \to \phi_j(U_i \cap U_j).$$

We obtain extra structure on the manifold by taking an atlas and imposing restrictions on the transition maps. An elegant way of doing this is to introduce the notion of a pseudogroup.

Definition. Let X be a topological space. A pseudogroup on X is a collection P of homeomorphisms between open subsets of X satisfying the following.

(1) The domains of the elements of P cover X.
(2) If $\phi : U \to V$ is in P and W is an open subset of U, then $\phi|_W$ is in P.
(3) If ϕ, $\psi \in P$ then $\psi \circ \phi \in P$ provided it is defined.
(4) If $\phi \in P$ then $\phi^{-1} \in P$.
(5) If $\phi : U \to V$ is a homeomorphism between open sets in X, and $\{U_\alpha\}$ is an open covering of U, such that $\phi|_{U_\alpha}$ is in P for all α, then $\phi \in P$.

If P is a pseudogroup on $\mathbb{R}^n$, two charts for an n-manifold M are said to be P-compatible if the corresponding transition map is in P. An atlas for an n-manifold M is called a P-atlas on M if the charts are pairwise P-compatible. For example, if P consists of all homeomorphisms between open subsets of $\mathbb{R}^n$ which have continuous rth order partial derivatives on their domain, then a P-atlas on M is called a C^r-structure on M. If P consists of the homeomorphisms having continuous partial derivatives of all orders, we refer to a C^∞ structure on M, and call M with this structure a *smooth* manifold. A manifold with a C^r structure (where $r \leq \infty$) is called a differentiable manifold.

Given a P-atlas $\mathcal{A}$ on a manifold M, there is a unique maximal P-atlas containing it, obtained by adjoining to $\mathcal{A}$ all charts which are P-compatible with every chart in $\mathcal{A}$. This is left as an exercise, with the comment that it is here that condition (5) in the definition of pseudogroup is important. It is frequently convenient to work with a maximal atlas in this sense.

Another exercise is to show that, if P_0 is a collection of homeomorphisms between open subsets of M, there is a smallest pseudogroup P on M containing P_0, which we call the pseudogroup generated by P_0.

We assume the reader is familiar with the idea of a covering map, which is discussed in many books on topology (e.g. [22; Chapter III, §3]). Suppose M is a manifold with a P-atlas and $p : \tilde{M} \to M$ is a covering map. Let $x \in M$, and let (U, ϕ) be a chart of the P-atlas with $x \in U$. Let V be an evenly covered neighbourhood of x (see Definition 3.1, Chapter III in [22]) and put $W = U \cap V$, so W is also an evenly covered neighbourhood of x. Thus $p^{-1}(W)$ is a disjoint union of open sets mapped homeomorphically onto W by p. Let $\tilde{W}$ be one of these open sets and put $\psi = \phi \circ (p|_{\tilde{W}})$, so that $(\tilde{W}, \psi)$ is a chart for $\tilde{M}$. Taking all such charts as x and $\tilde{W}$ vary gives an atlas for $\tilde{M}$. If $(\tilde{W}_1, \psi_1)$ and $(\tilde{W}_2, \psi_2)$ are two of the charts for this atlas, arising from charts (U_1, ϕ_1) and (U_2, ϕ_2) for M, then $\psi_1 \circ \psi_2^{-1}$ is $\phi_1 \circ \phi_2^{-1}$ restricted to $\phi_2 p(\tilde{W}_2)$, so is in P. Thus we have defined a P-atlas on $\tilde{M}$, which we call a *lift* of the P-atlas on M.

An important idea in the theory of manifolds is that of a foliation. We view $\mathbb{R}^n = \mathbb{R}^{n-k} \times \mathbb{R}^k$ as the union $\bigcup_{y \in \mathbb{R}^k} \mathbb{R}^{n-k} \times \{y\}$. In the case $n = 3$, $k = 1$, this is a union of horizontal planes, which we picture as infinite "leaves". We let P_0 be the set of all homeomorphisms ϕ between open subsets of $\mathbb{R}^n$ which preserve this decomposition of $\mathbb{R}^n$, i.e. which have the form $\phi(x, y) = (f(x, y), g(y))$, where x, $f(x, y) \in \mathbb{R}^{n-k}$ and y, $g(y) \in \mathbb{R}^k$. Let P be the pseudogroup generated by P_0; a foliation on a manifold M is defined to be a P-atlas on M, with the restriction that for any chart (U, ϕ), $\phi(U) = V \times W$ for some open sets $V \subseteq \mathbb{R}^{n-k}$ and $W \subseteq \mathbb{R}^k$. The integer k is called the codimension of the foliation. By minor modifications of the observations above, any foliation is contained in a unique maximal foliation, and a foliation can be lifted to a foliation of a covering space.

We are interested in a generalisation of a foliation called a lamination, in which the decomposition of $\mathbb{R}^n$ above is preserved only on some closed subset of M.

Definition. Let M be an n-manifold, and let C be a closed subspace of M. A lamination on M of codimension k with support C is a covering of C by charts such that:

(1) for each chart (U, ϕ), $\phi(U) = V \times W$ for some open sets $V \subseteq \mathbb{R}^{n-k}$ and $W \subseteq \mathbb{R}^k$ and $\phi(U \cap C) = V \times X$ for some closed subset X of W;

(2) if (U_i, ϕ_i), (U_j, ϕ_j) are two charts, then the transition map ϕ_{ij} has the property that any point $(x, y) \in \phi_i(U_i \cap U_j)$ has a neighbourhood N such that $\phi_{ij}(x, y) = (f_{ij}(x, y), g_{ij}(y))$ for $(x, y) \in N \cap \phi_i(C)$. (Here, x,

$f_{ij}(x, y) \in \mathbb{R}^{n-k}$ and y, $g_{ij}(y) \in \mathbb{R}^k$.)

Remark. In (2) of the definition, suppose that $\phi_i(U) = V_i \times W_i$ for some open sets $V_i \subseteq \mathbb{R}^{n-k}$ and $W_i \subseteq \mathbb{R}^k$, and that $\phi_i(U_i \cap C) = V_i \times X_i$. Interchanging i and j, there is a neighbourhood N' of $\phi_{ij}(x, y)$ such that $\phi_{ji}(x', y') = (f_{ji}(x', y'), g_{ji}(y'))$ for $(x', y') \in N' \cap \phi_j(C)$. Put $Z = N \cap \phi_{ji}(N')$. Then for $(x, y) \in Z$ such that $y \in X_i$, ϕ_{ij} maps $(V_i \times \{y\}) \cap Z$ to $(V_j \times \{y'\}) \cap \phi_{ij}(Z)$, where $y' = g_{ij}(y)$, and $(V_j \times \{y'\}) \cap \phi_{ij}(Z) \subseteq N'$. It follows that ϕ_{ij} maps $(V_i \times \{y\}) \cap Z$ homeomorphically onto $(V_j \times \{y'\}) \cap \phi_{ij}(Z)$, the inverse map being the restriction of ϕ_{ji}.

The support of a lamination $\mathcal{L}$ will be denoted by $|\mathcal{L}|$. As with foliations, any lamination can be uniquely extended to a maximal one, with the same support, and any lamination $\mathcal{L}$ can be lifted to a lamination on a covering space, whose support is the inverse image under the covering map of $|\mathcal{L}|$.

Remark. If (U, ϕ) is a chart for a lamination $\mathcal{L}$, and $\phi(U) = V \times W$ as in the definition, suppose V_1 is an open subset of V and W_1 is an open subset of W. Put $U_1 = \phi^{-1}(V_1 \times W_1)$. Then $(U_1, \phi|_{U_1})$ is a chart for the maximal lamination corresponding to $\mathcal{L}$.

We shall be exclusively interested in codimension 1 laminations, and it will be assumed from now on that all laminations are codimension 1. Given such a lamination, by restricting the mappings ϕ in the charts (U, ϕ), we obtain a covering of C by *flow boxes*, that is, charts (U, ϕ), where $\phi(U)$ has the form $V \times (a, b)$, with V an open connected subset of $\mathbb{R}^{n-1}$ and (a, b) a bounded open interval in $\mathbb{R}$. By the remark, these flow boxes belong to the corresponding maximal lamination. Further, we can cover C by flow boxes with V an open ball, so that $\phi(U)$ is convex in $\mathbb{R}^n$.

If (U, ϕ) is a flow box, with $\phi(C \cap U) = V \times X$, we call X the local leaf space of the flow box, and the sets $\phi^{-1}(V \times \{x\})$ for $x \in X$ are called *local leaves* of (U, ϕ).

It follows from the remark that a lamination on M with support C determines a topology $\mathcal{T}$ on C, with basis the set of local leaves. Its components are called the *leaves* of the lamination. Further, a local leaf in a flow box (U, ϕ) is homeomorphic to an open path connected subset of $\mathbb{R}^{n-1}$ via ϕ, and the topology induced by ϕ and the relative topology induced by $\mathcal{T}$ on a local leaf coincide. Hence any two points on the same local leaf lie in the same leaf of the lamination. Since a connected, locally path connected space is path connected, the leaves are path connected in the topology $\mathcal{T}$ (and are $(n-1)$-dimensional manifolds). Suppose L is a leaf and x, $y \in L$. We can find a path $\alpha : I = [0, 1]_{\mathbb{R}} \to L$ joining x to y. Each point in $\alpha(I)$ lies in a local leaf contained in L, giving an open cover for $\alpha(I)$ in L. Choose a Lebesgue number δ for the inverse image under α of this covering, and a partition $0 = t_0 < t_1 \ldots < t_k = 1$ such that

$t_i - t_{i-1} < \delta$ for $1 \leq i \leq k$, and let $x_i = \alpha(t_i)$. This gives finitely many points $x_0, \ldots, x_k$ in $\alpha(I)$ such that $x_0 = x$, $x_k = y$ and x_{i-1}, x_i lie in the same local leaf for $1 \leq i \leq k$. Conversely, given a finite collection of points x_i $(0 \leq i \leq k)$ satisfying $x_0 = x$, $x_k = y$ and with x_{i-1}, x_i in the same local leaf for $1 \leq i \leq k$, x and y lie in the same leaf.

Let $\alpha : J \to I$ be a continuous map of intervals in $\mathbb{R}$, and let X be a closed subset of I.

Definition. The map α is monotone over X if $\alpha|_{\alpha^{-1}(X)} : \alpha^{-1}(X) \to X$ is strictly monotone.

Suppose $\alpha : J \to M$ is a continuous map, where J is an interval and M is a manifold, with $\alpha(J) \subseteq U$ for some flow box (U, ϕ) of a lamination $\mathfrak{L}$, with $\phi(U) = V \times I$ and $\phi(|\mathfrak{L}| \cap U) = V \times X$, and let $p : V \times I \to I$ be the projection map.

Definition. The map α is transverse to $\mathfrak{L}$ in U if

(1) every point of J has an interval neighbourhood on which $p \circ \phi \circ \alpha$ is monotone over X;

(2) if t is an interior point of J and $\alpha(t) \in |\mathfrak{L}|$, then for every interval neighbourhood N of t in J, there exist $t_1, t_2 \in N$ such that $p\phi\alpha(t_1) < p\phi\alpha(t) < p\phi\alpha(t_2)$.

Definition. We say that an arbitrary map $\alpha : J \to M$ is transverse to $\mathfrak{L}$ if every point of J has a neighbourhood which α maps into a flow box, in which the restriction of α is transverse to $\mathfrak{L}$ in the flow box, as just defined.

To be a reasonable definition, this should be independent of the choice of flow box. This follows from the next lemma.

Lemma 4.1. *If (U_1, ϕ_1), (U_2, ϕ_2) are flow boxes for $\mathfrak{L}$ and $\alpha : J \to U_1 \cap U_2$ is a continuous map, where J is an interval, then α is transverse to U_1 in $\mathfrak{L}$ if and only if it is transverse to U_2 in $\mathfrak{L}$.*

Proof. We first show that α satisfies (1) in the definition above with respect to (U_1, ϕ_1) if and only if it satisfies it with respect to (U_2, ϕ_2).

For $i = 1, 2$, let $\phi_i(U_i) = V_i \times (a_i, b_i)$ where V_i is open in $\mathbb{R}^{n-1}$ and (a_i, b_i) is a bounded real interval, and let $\phi_i(U_i) \cap |\mathfrak{L}| = V_i \times X_i$. Also, let $p_i : V_i \times (a_i, b_i) \to (a_i, b_i)$ be the projection map. Let $t_0 \in J$, let $\phi_i\alpha(t_0) = (x_0, y_0)$ and suppose $y_0 \in X_i$. Let ϕ_{ij} be the transition map, where $j = 1, 2$ and $j \neq i$. By the remark after the definition of lamination, there is a neighbourhood Z_i of (x_0, y_0) in $\phi_i(U_1 \cap U_2)$ such that $\phi_{ij}(Z_i) = Z_j$ and for $(x, y) \in Z_i$ with $y \in X_i$, $\phi_{ij}(x, y) = (f_{ij}(x, y), g_{ij}(y))$, and for $y \in p_i(Z_i) \cap X_i$, ϕ_{ij} maps $(U_i \times \{y\}) \cap Z_i$ homeomorphically to $(U_j \times \{y'\}) \cap Z_j$, where $y' = g_{ij}(y) \in p_j(Z_j) \cap X_j$. We can assume Z_i is path connected. Put $Y_i = p_i(Z_i) \cap X_i$.

Assume α satisfies (1) with respect to (U_i, ϕ_i). By continuity of α, there is an interval neighbourhood I of t_0 in J such that $\phi_i\alpha(t) \in Z_i$ for $t \in I$. Now $\phi_j\alpha = \phi_{ij}\phi_i\alpha$, so for $t \in I$ such that $p_i\phi_i\alpha(t) \in Y_i$,

$$p_j\phi_j\alpha(t) = p_j\phi_{ij}\phi_i\alpha(t) = g_{ij}p_i\phi_i\alpha(t),$$

so to obtain the desired conclusion (that α satisfies (1) with respect to (U_j, ϕ_j), where $j \neq i$) it suffices to show that $g_{ij} : Y_i \to Y_j$ is strictly monotone.

Suppose not. Then there exist $y_i \in Y_i$ for $1 \leq i \leq 3$ such that $y_1 < y_2 < y_3$ and either $g_{ij}(y_1) > g_{ij}(y_2) < g_{ij}(y_3)$ or $g_{ij}(y_1) < g_{ij}(y_2) > g_{ij}(y_3)$. We assume $g_{ij}(y_2) < g_{ij}(y_1) \leq g_{ij}(y_3)$. (The other cases are similar and left to the reader.) Choose x_2 such that $(x_2, y_2) \in Z_i$ and x_3 such that $(x_3, y_3) \in Z_i$.

Now Z_i is path connected, so there is a path $\beta : [0,1] \to Z_i$ with $\beta(0) = (x_2, y_2)$ and $\beta(1) = (x_3, y_3)$. Put $c = \sup\{t \in [0,1] \mid p_i\beta(t) = y_2\}$, and then put $d = \inf\{t \in [c,1] \mid p_i\beta(t) = y_3\}$. It is an easy consequence of the Intermediate Value Theorem that $p_i\beta$ maps $[c,d]$ onto $[y_2, y_3]$. The mapping $p_j\phi_{ij}\beta : [c,d] \to p_j(Z_j)$ is continuous, $p_j\phi_{ij}\beta(c) = g_{ij}p_i\beta(c) = g_{ij}(y_2)$ and $p_j\phi_{ij}\beta(d) = g_{ij}(y_3)$. Again by the Intermediate Value Theorem, there exists $t \in [c,d]$ such that $p_j\phi_{ij}\beta(t) = g_{ij}(y_1)$. By the way Z_i was chosen, it follows that $p_i\beta(t) = y_1$, hence $y_1 \in [y_2, y_3]$, a contradiction.

We now need to show α satisfies (2) in the definition with respect to (U_1, ϕ_1) if and only if it satisfies (2) with respect to (U_2, ϕ_2). In the preceding setup, fixing i, we can assume Z_i is convex. If $(x, y) \in Z_i$ with $y \in X_i$, then $Z_i \backslash p_i^{-1}(y)$ has two path components, namely $p_i^{-1}(\{z \mid z > y\})$ and $p_i^{-1}(\{z \mid z < y\})$, hence $Z_j \setminus p_j^{-1}(y')$, where $y' = g_{ij}(y)$, has two path components, which are the images of these two path components under the homeomorphism ϕ_{ij}. These two path components must be $p_j^{-1}(\{z \mid z > y'\})$ and $p_j^{-1}(\{z \mid z < y'\})$, since any path component of $Z_j \setminus p_j^{-1}(y')$ is contained in one of these two sets, by continuity of p_j. If t is an interior point of J and $\phi_i\alpha(t) = (x, y)$, and N is an interval neighbourhood of t in J, then N has a smaller interval neighbourhood mapped into Z_i by $\phi_i \circ \alpha$. The lemma follows. □

We shall need some information on how transverse paths behave with respect to homotopy.

Lemma 4.2. *Let $\mathfrak{L}$ be a (codimension 1) lamination on a manifold M, such that all leaves are closed in M. Let $I = [0,1]_{\mathbb{R}}$ and suppose $H : I \times I \to M$ is a continuous map. Put $h_t(s) = H(s,t)$ and assume h_t is transverse to $\mathfrak{L}$ for all $t \in I$. Further, assume that for all $t \in I$, $H(0,t)$ lies in some fixed leaf or some fixed component of $M \setminus |\mathfrak{L}|$, and for all $t \in I$, $H(1,t)$ satisfies the same condition. Then $h_0(I)$ and $h_1(I)$ meet exactly the same set of leaves of $\mathfrak{L}$.*

Proof. Let L be a leaf meeting $h_0(I)$ but not $h_1(I)$. Let Z be the image of $H^{-1}(L)$ under the projection map from $I \times I$ onto the second coordinate. Since $H^{-1}(L)$ is closed, it is compact, so Z is compact, hence closed in I.

We claim Z is open in I. For suppose $t \in Z$, so we can choose $s \in I$ such that $H(s,t) \in L$. We need to find $\varepsilon > 0$ such that $(t-\varepsilon, t+\varepsilon) \cap I \subseteq Z$. If $s = 0$ or $s = 1$ this is true by assumption on H. Suppose $0 < s < 1$. Choose a flow box (U, ϕ) containing $H(s,t)$, with $\phi(U) = V \times J$, where V is an open ball in $\mathbb{R}^{n-1}$ and J is an open real interval. Let $p : V \times J \to J$ be the projection map. Since $H^{-1}(U)$ is open, it contains a set of the form $(s-\delta, s+\delta) \times \big((t-\delta, t+\delta) \cap I\big)$, for some $\delta > 0$. Since h_t is transverse to $\mathfrak{L}$, we can find s_1, $s_2 \in (s-\delta, s+\delta)$ such that $p\phi h_t(s_1) < p\phi h_t(s) < p\phi h_t(s_2)$. By continuity of H, there is a neighbourhood $N = (t-\varepsilon, t+\varepsilon) \cap I$ (with $0 < \varepsilon \leq \delta$) such that for $t' \in N$, $p\phi H(s_1, t') < p\phi h_t(s) < p\phi H(s_2, t')$. For such t', there exists $s' \in (s_1, s_2)$ such that $p\phi H(s', t') = p\phi h_t(s)$, which implies $H(s', t') \in L$. Hence $N \subseteq Z$, as required.

Now $0 \in Z$ and $1 \notin Z$, and this contradicts the connectedness of I. Similarly, (replacing H by the map $(s,t) \mapsto H(s, 1-t)$), there is no leaf meeting $h_1(I)$ but not $h_0(I)$, proving the lemma. □

Before introducing the idea of a measured lamination, we recall some basic definitions and results from measure theory. Let $\mathcal{E}$ be a non-empty collection of sets which form a ring, in the sense that if A, $B \in \mathcal{E}$ then $A \cup B \in \mathcal{E}$ and $A \setminus B \in \mathcal{E}$. (It follows easily that if A, $B \in \mathcal{E}$ then $A \cap B \in \mathcal{E}$, and that $\emptyset \in \mathcal{E}$.)

Definition. A mapping $\mu : \mathcal{E} \to \mathbb{R} \cup \{\pm\infty\}$ is called a measure if
(a) For all $A \in \mathcal{E}$, $\mu(A) \geq 0$.
(b) If $\{A_n \mid n \in \mathbb{N}\}$ is a collection of pairwise disjoint subsets of $\mathcal{E}$, and $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{E}$, then $\mu\big(\bigcup_{n\in\mathbb{N}} A_n\big) = \sum_{n\in\mathbb{N}} \mu(A_n)$.
(c) $\mu(\emptyset) = 0$.

It follows that, if A, $B \in \mathcal{E}$, then $\mu(B) = \mu(A \cap B) + \mu(B \setminus A)$. Hence, if $A \subseteq B$ then $\mu(A) \leq \mu(B)$. Also, $\mu(A \cup B) = \mu(A) + \mu(B \setminus A)$, hence $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$, so $\mu(A \cup B) \leq \mu(A) + \mu(B)$. It follows by induction and taking limits that, if $\{A_n \mid n \in \mathbb{N}\}$ is a collection of subsets of $\mathcal{E}$, and $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{E}$, then $\mu\big(\bigcup_{n\in\mathbb{N}} A_n\big) \leq \sum_{n\in\mathbb{N}} \mu(A_n)$.

The sets in $\mathcal{E}$ are called measurable sets with respect to μ. A *point mass* is a singleton $\{x\}$ which is measurable and such that $\mu(\{x\}) > 0$. A ring of sets is called a σ-ring if it is closed under countable unions.

Let X be a locally compact Hausdorff space, let $\mathcal{C}$ be the set of all compact subsets of X, let $\mathcal{B}$ be the σ-ring generated by $\mathcal{C}$ and let $\mathcal{U}$ be the set of all open subsets of X which belong to $\mathcal{B}$. Members of $\mathcal{B}$ are called Borel subsets of X. We define a Borel measure on X to be a measure μ on $\mathcal{B}$, such that $\mu(C) < \infty$ for all $C \in \mathcal{C}$.

Remark. If μ is a Borel measure on X with $\mu(X) < \infty$, then the set D of point masses of μ is countable. For suppose not. Let $A_n = \{x \in D \mid \mu(x) \geq 1/n\}$. If all A_n are finite, then $D = \bigcup_{n=1}^{\infty} A_n$ is countable, so some A_n, say A_N, is infinite. We can then choose an integer $k > N\mu(X)$, and k points $x_1, \ldots, x_k \in A_N$. Then $\mu(\{x_1, \ldots, x_k\}) = \mu(x_1) + \ldots + \mu(x_k) > \mu(X)$, a contradiction.

Definition. A Borel measure μ is *regular* if, for $A \in \mathcal{B}$, $\mu(A) = \inf\{\mu(U) \mid U \in \mathcal{U} \text{ and } A \subseteq U\}$ and $\mu(A) = \sup\{\mu(C) \mid C \in \mathcal{C} \text{ and } C \subseteq A\}$.

In order to construct Borel measures, we introduce the notion of a content on the locally compact space X.

Definition. A content on X is a mapping $\nu : \mathcal{C} \to \mathbb{R} \cup \{\pm\infty\}$ satisfying the following.

(i) For all $C \in \mathcal{C}$, $0 \leq \nu(C) < \infty$.
(ii) If C, $D \in \mathcal{C}$ and $C \subseteq D$, then $\nu(C) \leq \nu(D)$.
(iii) If C, $D \in \mathcal{C}$ and $C \cap D = \emptyset$, then $\nu(C \cup D) = \nu(C) + \nu(D)$.
(iv) If C, $D \in \mathcal{C}$ then $\nu(C \cup D) \leq \nu(C) + \nu(D)$.

A content ν is called *regular* if, for all $C \in \mathcal{C}$, $\nu(C) = \inf\{\nu(D) \mid C \subseteq D^\circ \subseteq D \in \mathcal{C}\}$. Note that a regular Borel measure defines a regular content by restriction to compact sets.

Proposition 4.3. *A regular content ν on X has a unique extension to a regular Borel measure on X.*

Proof. See [12], §§63-64 or [71], Chapter X. □

Let $C(X)$ be the space of continuous real-valued functions on X with compact support (the support supp(f) of a function $f : X \to \mathbb{R}$ is the closure of the set $\{x \in X \mid f(x) \neq 0\}$). This is a Banach space, with the norm $||f|| = \sup_{x \in X} |f(x)|$. Let $L(X) = \mathrm{Hom}_{\mathbb{R}}(C(X), \mathbb{R})$, the space of all linear functionals on $C(X)$. We topologise $L(X)$ with the weak topology. This is the topology with subbasis the collection $U(f, V)$ where f runs through $C(X)$ and V through open subsets of $\mathbb{R}$, and $U(f, V) = \{\phi \in L(X) \mid \phi(f) \in V\}$. (This is also called the weak-$*$ topology.) This topology coincides with the subspace topology when $L(X)$ is viewed as a subset of $\mathbb{R}^{C(X)}$, which is given the product topology.

For f, $g \in C(X)$, it is easy to see that $\{\phi \in L(X) \mid \phi(f) > \phi(g)\}$ coincides with $\bigcup_{a \in \mathbb{R}} U(f, (a, \infty)) \cap U(g, (-\infty, a))$, so is open, and the same is true with f and g interchanged. It follows that $\{\phi \in L(X) \mid \phi(f) = \phi(g)\}$ is closed. Similarly, for $a \in \mathbb{R}$, the sets $\{\phi \in L(X) \mid \phi(f) = a\}$ and $\{\phi \in L(X) \mid \phi(f) \geq a\}$ are closed, and $\{\phi \in L(X) \mid \phi(f) > a\}$ is open.

An element $\phi \in L(X)$ is called positive if, for all $f \in C(X)$ such that $f(x) \geq 0$ for all $x \in X$, we have $\phi(f) \geq 0$. Denote the set of positive elements of $L(X)$ by $L^+(X)$. Note that $L^+(X)$ is a closed subspace of $L(X)$, because by the remarks just made, it is an intersection of closed sets, one for each $f \in C(X)$ such that $f(x) \geq 0$ for all $x \in X$. Let $M(X)$ denote the set of regular Borel measures on X. If $\mu \in M(X)$, we obtain $\phi \in L^+(X)$, by $\phi(f) = \int_X f d\mu$ (we refer to [12] or [71] for the definition of the integral). There is a converse to this.

Proposition 4.4. (Riesz-Markov Representation Theorem.) *If ϕ is a positive linear functional on $C(X)$, there is a unique regular Borel measure μ on X such that for all $f \in C(X)$, $\phi(f) = \int_X f d\mu$.*

Proof. See [12], §69 or [71]. □

Thus there is a one-to-one correspondence between $M(X)$ and $L^+(X)$, and we topologise $M(X)$ so that this bijection is a homeomorphism. The following remarks will be needed later, in §5.

Remarks. **(1)** If U is an open subset of X and $a \in \mathbb{R}$, then the set

$$\{\mu \in M(X) \mid \mu(U) > a\}$$

is open in $M(X)$. To see this, let C be a compact subset of U. We can find an open set V and a compact set D such that $C \subseteq V \subseteq D \subseteq U$ (this is well-known and easy to see). By the version of Urysohn's Lemma for locally compact spaces (Chapter X, §50, Theorem B in [71]), there is a continuous function $f_C : X \to [0,1]$ such that $f_C(x) = 1$ for $x \in C$ and $f_C(x) = 0$ for $x \notin V$, so $\operatorname{supp}(f_C) \subseteq D$ is compact. Then we claim that

$$\{\mu \in M(X) \mid \mu(U) > a\} = \bigcup_C \{\mu \in M(X) \mid \textstyle\int f_C d\mu > a\}$$

(union over all compact subsets C of U) which is a union of open sets, so open. For $\chi_C \leq f_C \leq \chi_U$, where χ denotes characteristic function, so $\mu(C) \leq \int f_C d\mu \leq \mu(U)$, hence the union is contained in $\{\mu \in M(X) \mid \mu(U) > a\}$. If $\mu(U) > a$, let $\varepsilon = \mu(U) - a$. Then by regularity we can choose C such that $\mu(U) < \mu(C) + \varepsilon$. It follows that $\int f_C d\mu > a$, establishing the claim.

(2) If $g : X \to Y$ is a homeomorphism, there is an induced homeomorphism $\bar{g} : L^+(X) \to L^+(Y)$, given by $(\bar{g}\phi)(f) = \phi(f \circ g)$ for $f \in C(Y)$. This induces a homeomorphism $\bar{g} : M(X) \to M(Y)$, which is given by $\bar{g}\mu(A) = \mu(g^{-1}(A))$, for Borel sets A of Y. To prove this, first assume A is compact. Using regularity, given $\varepsilon > 0$, it is easy to find an open set U of Y such that $A \subseteq U$, $\bar{g}\mu(U) < \bar{g}\mu(A) + \varepsilon$ and $\mu(g^{-1}(U)) < \mu(g^{-1}(A)) + \varepsilon$. Note that, by definition of $\bar{g}\mu$, $\int f d(\bar{g}\mu) = \int (f \circ g^{-1}) d\mu$ for $f \in C(Y)$. We can take f to be a function which is 1 on A and 0 outside U, as in Remark 1. It follows, using inequalities similar to those in Remark 1, that $\bar{g}\mu(A) = \mu(g^{-1}A)$. Details are left to the reader. It now follows that $\bar{g}\mu(A) = \mu(g^{-1}A)$ for any Borel set A by uniqueness in Proposition 4.3.

(3) Suppose X is compact. For $\phi \in L^+(X)$, associated to the measure $\mu \in M(X)$, we define $\|\phi\| = \sup_{f \in C(X)}(|\int_X f d\mu| / \|f\|)$ ($\|\phi\| = 0$ if ϕ is identically zero). Then $||\phi|| \leq \mu(X) < \infty$ (this follows easily from Theorem 6, Ch.4 in [12]), so $L^+(X)$ is contained in the space $C(X)^*$ of bounded linear functionals on $C(X)$. Hence the set $\{\mu \in M(X) \mid \mu(X) = 1\}$ corresponds to a subset of the closed unit ball in $C(X)^*$, which is compact (see, for example, Theorem 2, §5.4 in [65]). Since $\mu(X) = \phi(\chi_X)$, where χ_X is the constant function 1 on X, it follows from the observations preceding Proposition 4.4 that $\{\mu \in M(X) \mid \mu(X) = 1\}$ is closed in $M(X)$, and therefore compact.

If X is a locally compact space, μ is a Borel measure on X and Y is a subspace of X, there is an induced Borel measure $\mu|_Y$, since any Borel set in Y is a Borel set of X. If Y is a Borel set and μ is regular, then $\mu|_Y$ is regular. If Y is open, the mapping $M(X) \to M(Y)$, $\mu \mapsto \mu|_Y$ is continuous, because any $f \in C(Y)$ can be extended to a map in $C(X)$ by defining f to be zero outside Y (continuity follows by the glueing lemma applied to $X \setminus \mathrm{supp}(f)$ and Y).

Proposition 4.5. *Let $\{Y_i\}_{i\in I}$ be an open covering of the locally compact Hausdorff space X, and let μ_i be a regular Borel measure on Y_i, for each $i \in I$, such that the restrictions of μ_i and μ_j to $Y_i \cap Y_j$ agree, for every $i, j \in I$. Then there is a unique regular Borel measure μ on X whose restriction to Y_i is μ_i, for every $i \in I$.*

Proof. In view of Proposition 4.4, this follows from the argument for Prop.1 in Ch.III, §2 of [20]. □

Corollary 4.6. *Let $\{C_i\}_{i\in I}$ be a collection of compact subsets of the locally compact Hausdorff space X, whose interiors cover X, and let μ_i be a regular Borel measure on C_i, for each $i \in I$, such that the restrictions of μ_i and μ_j to $C_i \cap C_j$ agree, for every $i, j \in I$. Then there is a unique regular Borel measure μ on X whose restriction to C_i is μ_i, for every $i \in I$.*

Proof. Let $Y_i = \mathring{C}_i$, the interior of C_i, and let $\nu_i = \mu_i|_{Y_i}$. The boundary ∂C of C in X is a closed subspace of C, so compact, hence $Y_i = C \setminus \partial C$ is a Borel set in X and ν_i is regular. By Proposition 4.5 there is a unique regular Borel measure μ on X whose restriction to Y_i is ν_i for every $i \in I$, and it suffices to show that $\mu|_{C_j} = \mu_j$ for $j \in I$. Fix $j \in I$, and for $i \in I$, let $\lambda_i = \mu_i|_{Y_i \cap C_j}$. For a Borel set A in $Y_i \cap C_j$, $\lambda_i(A) = \mu_i(A) = \nu_i(A) = \mu(A)$ and $\mu_i(A) = \mu_j(A)$ since $A \subseteq C_i \cap C_j$, hence $\mu(A) = \mu_j(A) = \lambda_i(A)$. Applying uniqueness in Proposition 4.5 to the collection λ_i, $\mu|_{C_j} = \mu_j$. □

A regular Borel measure μ on a locally compact space X is said to be *supported* on a measurable set A if $\mu(X \setminus A) = 0$. Assume that all open sets in X are Borel sets. Let $\mathcal{O}$ be the set of all open sets U in X such that $\mu(U) = 0$ and put $O = \bigcup_{U\in\mathcal{O}} U$. Applying uniqueness in Prop. 4.5 to O and the collection $\mathcal{O}$, where each $U \in \mathcal{O}$ has the measure which is identically zero, we conclude that $\mu(O) = 0$. The complement $X \setminus O$ is called the *support* of μ, denoted

$\text{supp}(\mu)$, a closed set. Note that, for any Borel set A, $\mu(A) = \mu(A \cap \text{supp}(\mu))$. For $\mu(A) = \mu(A \cap \text{supp}(\mu)) + \mu(A \setminus \text{supp}(\mu))$, and $\mu(A \setminus \text{supp}(\mu)) = 0$ since $A \setminus \text{supp}(\mu) \subseteq O$.

Remark. Under the assumption that all open subsets of X are Borel sets, $\mathcal{B}$ is the σ-ring generated by the collection of open subsets of X, because compact sets are closed. If X is second countable then all open sets are Borel sets. For if U is open, each $x \in U$ has a compact neighbourhood C_x contained in U (see Chapter I, Theorem 11.2 in [22]). Since U is second countable, it is Lindelöf, so U is the union of countably many of the sets C_x, hence $U \in \mathcal{B}$. In applications, X will be a homeomorph of a real interval, so second countable. Note that, in this case, all subintervals of X are Borel sets.

We are now ready to introduce the idea of a measured lamination. Let (U, ϕ) be a flow box for a lamination $\mathcal{L}$ with $\phi(U) = V \times I$ and $\phi(U \cap |\mathcal{L}|) = V \times X$.

Definition. A transverse measure for $\mathcal{L}$ in (U, ϕ) is a regular Borel measure μ on I which is supported on X. It is of full support if the support is X.

A local leaf $\phi^{-1}(V \times \{x\})$ is said to support a point mass of μ if x is a point mass of μ.

Let μ be a transverse measure for $\mathcal{L}$ in (U, ϕ). Let $\alpha : J \to U$ be continuous, where J is a compact interval. Let $p : V \times I \to I$ be the projection map (where $\phi(U) = V \times I$). Suppose projection of α onto I, that is, $p\phi\alpha$, is monotone over X (the local leaf space). We define

$$\int_\alpha \mu = \mu(p\phi\alpha(J)).$$

If $\alpha : J \to U$ is transverse to $\mathcal{L}$ in U, but not necessarily monotone over X, each point $t \in J$ has an open neighbourhood N_t in which α is monotone over X; let δ be a Lebesgue number for the covering $\{N_t \mid t \in J\}$ of J and take a partition $t_0 < t_1 < \ldots < t_r$ of J such that $t_i - t_{i-1} < \delta$ for $1 \leq i \leq r$. Then $\alpha|_{J_i}$ is monotone over X, where $J_i = [t_{i-1}, t_i]$, and by slightly perturbing the t_i if necessary, using a previous remark, we can assume $\alpha(t_i)$ does not lie on a local leaf supporting a point mass of μ for $1 \leq i < r$. We then put

$$\int_\alpha \mu = \sum_i \int_{\alpha_i} \mu$$

where α_i is the restriction of α to $[t_{i-1}, t_i]$. Adding an extra point t to the partition, where $\alpha(t)$ does not lie on a local leaf supporting a point mass of μ, does not alter this sum. Since two partitions have a common refinement (namely their union), $\int_\alpha \mu$ is independent of the choice of partition.

If (U_1, ϕ_1), (U_2, ϕ_2) are flow boxes for $\mathfrak{L}$ and μ_1, μ_2 are transverse measures for $\mathfrak{L}$ in (U_1, ϕ_1), (U_2, ϕ_2) respectively, then μ_1, μ_2 are said to be compatible if they assign the same measure to all transverse paths to $\mathfrak{L}$ in $U_1 \cap U_2$.

Definition. A transverse measure for a lamination $\mathfrak{L}$ is a family of compatible measures for $\mathfrak{L}$ in the flow boxes of $\mathfrak{L}$. It is of full support if the measure on each flow box is of full support. A measured lamination is a lamination together with a transverse measure for it.

As a matter of notation, we shall use a single letter, usually μ, to denote a transverse measure, meaning the family of measures $\{\mu_{(U,\phi)}\}$, where (U, ϕ) runs through the flow boxes of the lamination and $\mu_{(U,\phi)}$ is the transverse measure in (U, ϕ). We leave it to the reader to verify that a measured lamination on a manifold lifts to a measured lamination on any covering space of the manifold.

If μ is a transverse measure, and $\alpha : J \to M$ is continuous, where J is a compact interval, $\int_\alpha \mu$ is defined by partitioning J into subintervals, each of which maps into a flow box. (Again, choose the interior points of the partition not to map to leaves supporting point masses of the measures.) The compatibility condition ensures this is independent of the choice of partition. We shall abbreviate $\int_\alpha \mu$ to $\mu(\alpha)$. Note that μ has an additivity property. If α is the product of two transverse paths α_1 and α_2 such that $\alpha_1(1) = \alpha_2(0)$ does not lie on a local leaf supporting a point mass, then $\mu(\alpha) = \mu(\alpha_1) + \mu(\alpha_2)$.

Remark. Suppose $\alpha : J \to M$ is continuous, where J is a real interval, and μ is a transverse measure of full support for a lamination on the manifold M. If $\alpha(t)$ lies on a leaf, where t is not an endpoint of J, then $\mu(\alpha) > 0$.

For we can assume α maps into a flow box (U, ϕ), with $\phi(U) = V \times (a, b)$ and $p : V \times (a, b) \to (a, b)$ the projection map, and that α is monotone over the local leaf space. Suppose $\alpha(t)$ lies on a leaf. Choose an open interval J' in J containing t, and let $I = p\phi\alpha(J')$. Then $p\phi\alpha(t)$ is in the interior $\mathring{I}$ of the interval I by the assumption that α is monotone.

Thus $\mu(\alpha) \geq \mu(\alpha|_{J'}) = \mu_{(U,\phi)}(I) \geq \mu_{(U,\phi)}(\mathring{I}) > 0$. For if $\mu_{(U,\phi)}(\mathring{I}) = 0$, then $(a, b) \setminus \mathring{I} \supseteq \mathrm{supp}(\mu_{(U,\phi)})$, hence $p\phi\alpha(t) \in (a, b) \setminus \mathring{I}$, a contradiction.

Given a measured lamination $\mathfrak{L}$ on M, we shall now use it to construct an $\mathbb{R}$-tree, provided certain restrictions are imposed. Let $\tilde{\mathfrak{L}}$ be the induced measured lamination on the universal covering $\tilde{M}$ and let $\tilde{C} = |\tilde{\mathfrak{L}}|$. Suppose

(a) $\mathfrak{L}$ is of full support and $C = |\mathfrak{L}|$ is nowhere dense in M.
(b) Each leaf of $\tilde{\mathfrak{L}}$ is a closed subset of $\tilde{M}$.
(c) M is a connected, paracompact C^∞ manifold, and the leaves are C^∞ submanifolds of M.
(d) There is a covering of $\tilde{M}$ by flow boxes (U, ϕ) such that, if p and q are points of $U \setminus \tilde{C}$, then there is a smooth path $\omega : [0, 1] \to U$ with $\omega(0) = p$ and

$\omega(1) = q$, which is transverse (in the usual sense of differential topology) to each leaf of $\tilde{\mathfrak{L}}$ and which crosses each leaf at most once.

For a manifold, paracompactness is equivalent to certain other topological properties (see [135; Appendix A, Theorem 1]). These equivalent conditions are also equivalent to saying that the manifold admits a Riemannian structure. See for example [18], Chapter V, Theorem 4.5, where it is shown that a second countable manifold has a Riemannian structure, and Chapter V, Theorem 3.1, where it is shown that a connected Riemannian manifold is metrisable.

Note that (a) implies $\tilde{C}$ is nowhere dense in $\tilde{M}$.

Lemma 4.7. *Assume conditions* (a), (b), (c) *and* (d) *are satisfied. Let x, $y \in X$, where X is the set of components of $\tilde{M} \setminus \tilde{C}$. Suppose ω and ω' are transverse paths in $\tilde{M}$ from $p \in x$ to $q \in y$ and from $p' \in x$ to $q' \in y$ respectively, with ω crossing each leaf at most once. Then any leaf of $\tilde{\mathfrak{L}}$ which meets ω also meets ω'. Any leaf which meets ω' but not ω crosses ω' at least twice.*

Proof. Let δ be a Lebesgue number for the covering $\omega^{-1}(U)$, of $[0,1]$, where (U, ϕ) runs through all flow boxes as in (d), and let $0 = t_0 < \ldots < t_k = 1$ be a partition of $[0,1]$ such that $t_i - t_{i-1} < \delta$ for $1 \leq i \leq k$, so $\omega([t_{i-1}, t_i])$ lies in a flow box of $\tilde{\mathfrak{L}}$, say (U_i, ϕ_i), satisfying the conditions in (d). Let $\phi_i(U_i) = V_i \times J_i$ where J_i is an open real interval. Since $\tilde{C}$ is nowhere dense, we can assume $t_i \notin \tilde{C}$ for all i. There is a smooth path λ_i in U_i from $\omega(t_{i-1})$ to $\omega(t_i)$, transverse to $\tilde{\mathfrak{L}}$ and crossing each leaf at most once. Then $\lambda = \lambda_1 \ldots \lambda_n$ meets the same set of leaves as ω, and crosses each such leaf just once. (Apply the Intermediate Value Theorem to the projections of $\phi_i\omega|_{[t_{i-1},t_i]}$ and $\phi_i\lambda_i$ onto J_i.) Then λ is piecewise smooth, and since $\omega(t_i)$ has a neighbourhood not meeting $\tilde{C}$, we can make λ smooth. Thus we can assume that ω is smooth, and similarly that ω' is smooth.

Let α, β be smooth paths in $\tilde{M} \setminus \tilde{C}$ from $\omega(0)$ to $\omega'(0)$ and from $\omega(1)$ to $\omega'(1)$ respectively. Again we can modify the paths so that $\omega\beta\omega'^{-1}\alpha^{-1}$ is smooth (viewed as a map from the circle S^1 to $\tilde{M}$).

Since $\tilde{M}$ is simply connected, there is a homotopy $H : I \times I \to \tilde{M}$, where $I = [0,1]$, with $H|_{[0,1]\times\{0\}} = \omega$, $H|_{[0,1]\times\{1\}} = \omega'$, $H(\{0\} \times [0,1]) = \alpha$ and $H(\{1\} \times [0,1]) = \beta$. By Theorem 11.8 and the remarks at the end of §11 in [22], we can assume H is smooth. Let L be a leaf of $\tilde{\mathfrak{L}}$. By §3, Chapter 2 of [68], we can assume H is transverse to L, in the usual sense of differential topology. Then by §1, Chapter 2 in [68], $H^{-1}(L)$ is a submanifold of $I \times I$ of dimension 1, whose boundary is the intersection of $H^{-1}(L)$ with the boundary of $I \times I$.

By Condition (b), $H^{-1}(L)$ is compact. Hence, each component of $H^{-1}(L)$ is either a circle or an arc with endpoints in $[0,1] \times \{0,1\}$. No such component can have both endpoints in $[0,1] \times \{0\}$, because ω crosses each leaf of $\tilde{\mathfrak{L}}$ at most once. Hence the components which meet the boundary of $[0,1] \times [0,1]$ are either arcs from $[0,1] \times \{0\}$ to $[0,1] \times \{1\}$ or arcs with both endpoints in $[0,1] \times \{1\}$. The lemma follows. □

Lemma 4.8. *Under the hypotheses of Lemma 4.7, let $\tilde{\mu}$ be the lift of the transverse measure for $\mathfrak{L}$ to a transverse measure for $\tilde{\mathfrak{L}}$. Then*

$$\tilde{\mu}(\omega) \le \tilde{\mu}(\omega')$$

with equality if and only if ω' crosses each leaf of $\tilde{\mathfrak{L}}$ at most once.

Proof. The proof is a version of the construction of the leaf holonomy map for a foliation. Let $x \in \mathrm{Im}(\omega)$, $x' \in \mathrm{Im}(\omega')$ be points on the same leaf L of $\tilde{\mathfrak{L}}$. Choose a path $\gamma : [0,1] \to L$ with $\gamma(0) = x$, $\gamma(1) = x'$. Let $0 = t_0 < \ldots < t_n = 1$ be a partition of $[0,1]$ such that, for $1 \le i \le n$, $\gamma([t_{i-1}, t_i])$ lies in U_i, where (U_i, ϕ_i) is a flow box, with $\phi_i(U_i) = V_i \times I_i$, where I_i is an interval. For each i, $1 \le i < n$, let ω_i be a path transverse to $\tilde{C}$ in $U_i \cap U_{i+1}$ passing through $\gamma(t_i)$ (e.g. a linear path in $V_i \times I_i$ through $\phi_i\gamma_i(t_i)$ orthogonal to the sets $V_i \times \{a\}$ for $a \in I_i$, lifted to U_i via ϕ_i^{-1} and restricted so it lies in U_{i+1}). Let $\omega_i(s_i) = \gamma(t_i)$ $(1 \le i < n)$, $\omega(s_0) = \gamma(0)$ and $\omega'(s_n) = \gamma(1)$. Note that s_i $(1 \le i < n)$ is intended to be an interior point of the domain of ω_i. Let $p_i : V_i \times I_i \to I_i$ be the projection map and let $x_i = p_i\phi_i\gamma(t_i)$.

There are interval neighbourhoods N_0, N_1 of s_0, s_1 respectively such that $p_1\phi_1\omega|_{N_0}$ and $p_1\phi_1\omega_1|_{N_1}$ are monotone over $p_1\phi_1(U_1 \cap \tilde{C})$. Thus $p_1\phi_1\omega(N_0)$ and $p_1\phi_1\omega_1(N_1)$ are non-degenerate intervals containing x_1. (A non-degenerate interval is one containing more than one point.) Hence we can find a, b, both in $p_1\phi_1\omega(N_0)$ and $p_1\phi_1\omega_1(N_1)$, such that $a < x_1 < b$. Choose $u_0, v_0 \in N_0$ such that $p_1\phi_1\omega(u_0) = a$, $p_1\phi_1\omega(v_0) = b$, and choose $u_1, v_1 \in N_1$ such that $p_1\phi_1\omega_1(u_1) = a$, $p_1\phi_1\omega_1(v_1) = b$. Then $\omega([u_0, v_0])$ and $\omega_1([u_1, v_1])$ meet exactly the same set of leaves of $\tilde{\mathfrak{L}}$, namely those having a local leaf of the form $\phi_1^{-1}(V_i \times \{y\})$ for some $y \in [a,b]$ (this is left as an exercise). Further, $\tilde{\mu}(\omega|_{[u_0,v_0]}) = \tilde{\mu}(\omega_1|_{[u_1,v_1]})$, both being equal to $\tilde{\mu}_{(U_1,\phi_1)}([a,b])$.

Similarly, we can find non-degenerate subintervals $[u_1', v_1']$ and $[u_2, v_2]$ of $[0,1]$ such that $\omega_1([u_1', v_1'])$ and $\omega_2([u_2, v_2])$ meet exactly the same leaves of $\tilde{\mathfrak{L}}$, $s_1 \in (u_1', v_1')$, $s_2 \in (u_2, v_2)$ and $\tilde{\mu}(\omega_1|_{[u_1',v_1']}) = \tilde{\mu}(\omega_2|_{[u_2,v_2]})$. Replacing $[u_1, v_1]$ by $[u_1, v_1] \cap [u_1', v_1']$ and making appropriate adjustments to $[u_0, v_0]$ and $[u_2, v_2]$, we obtain that $\omega([u_0, v_0])$, $\omega_1([u_1, v_1])$ and $\omega_2([u_2, v_2])$ meet exactly the same set of leaves of $\tilde{\mathfrak{L}}$ and $\tilde{\mu}(\omega|_{[u_0,v_0]}) = \tilde{\mu}(\omega_1|_{[u_1,v_1]}) = \tilde{\mu}(\omega_2|_{[u_2,v_2]})$. Proceeding inductively, we eventually obtain non-degenerate intervals $[u, v]$ and $[u', v']$ containing s_0, s_n respectively as interior points, such that $\omega([u, v])$ and $\omega'([u', v'])$ meet exactly the same set of leaves of $\tilde{\mathfrak{L}}$, cross each such leaf exactly once, and such that $\tilde{\mu}(\omega|_{[u,v]}) = \tilde{\mu}(\omega'|_{[u',v']})$. It is clear from the construction that $[u, v]$ can be replaced by a non-degenerate subinterval, and $[u', v']$ by a corresponding subinterval, so that these properties are still satisfied.

By Lemma 4.7, for each $s \in [0,1]$ such that $\omega(s) \in \tilde{C}$, we can find $x' \in \tilde{C} \cap \mathrm{Im}(\omega')$ and in the same leaf of $\tilde{\mathcal{L}}$. We then have intervals $[u,v]$, $[u',v']$ as in the last paragraph, containing $\omega(s)$ and x', respectively. Denote these intervals by W_s, W_s' respectively. If $s \in [0,1]$ and $\omega(s) \notin \tilde{C}$, take W_s to be any compact interval containing s in its interior in $[0,1]$, such that $\omega(W_s)$ does not meet $\tilde{C}$, and W_s' any compact interval such that $\omega'(W_s')$ does not meet $\tilde{C}$.

The interiors of the W_s, for $s \in [0,1]$, cover $[0,1]$. Let δ be a Lebesgue number for this covering. Choose a partition $0 = p_0 < \ldots < p_k = 1$ such that $p_j - p_{j-1} < \delta$ for $i \leq j \leq k$ and no p_j supports a point mass of $\tilde{\mu}$. Then each interval $[p_{j-1}, p_j]$ is contained in one of the W_s, so we can find a compact interval Z_j such that $\omega([p_{j-1}, p_j])$ and $\omega'(Z_j)$ meet exactly the same leaves of $\tilde{\mathcal{L}}$ and $\tilde{\mu}(\omega|_{[p_{j-1},p_j]}) = \tilde{\mu}(\omega'|_{Z_j})$. Thus

$$\tilde{\mu}(\omega) = \sum_{j=1}^{k} \tilde{\mu}(\omega|_{[p_{j-1},p_j]}) = \sum_{j=1}^{k} \tilde{\mu}(\omega'|_{Z_j}).$$

Also, for $j \neq l$, $\omega'(Z_j) \cap \omega'(Z_l)$ meets at most one leaf of $\tilde{\mathcal{L}}$, which does not support a point mass. Hence $\tilde{\mu}(\omega'|_{Z_j \cap Z_l}) = 0$. Now $\bigcup_{j=1}^{k} Z_j$ is a union of finitely many disjoint compact intervals, say $K_1, \ldots, K_r$. It follows easily that

$$\sum_{j=1}^{k} \tilde{\mu}(\omega'|_{Z_j}) = \sum_{i=1}^{r} \tilde{\mu}(\omega'|_{K_i}) \leq \tilde{\mu}(\omega').$$

Further, if ω' crosses some leaf twice, then this leaf meets $\omega'(U)$, where U is one of the open intervals in the complement of $\bigcup_{j=1}^{k} Z_j$ and

$$\sum_{j=1}^{k} \tilde{\mu}(\omega'|_{Z_j}) < \sum_{i=1}^{r} \tilde{\mu}(\omega'|_{K_i}) + \tilde{\mu}(\omega'|_U) \leq \tilde{\mu}(\omega')$$

using the remark after the definition of transverse measure (recall that the endpoints of ω' do not lie in $\tilde{C}$).

Finally if ω' crosses each leaf at most once, then we may interchange ω, ω' in the argument to obtain $\tilde{\mu}(\omega') \leq \tilde{\mu}(\omega)$, so there is equality. □

In addition to assumptions (a)-(d) in the previous two lemmas, assume also:

(e) for any p, $q \in \tilde{M} \setminus \tilde{C}$, there is a path ω in $\tilde{M}$ from p to q which is transverse to $\tilde{\mathcal{L}}$ and crosses each leaf at most once.

We define an $\mathbb{R}$-metric d on X, the set of components of $\tilde{M} \setminus \tilde{C}$. Suppose x, $y \in X$, choose $p \in x$, $q \in y$ and let $\omega : [0,1] \to \tilde{M}$ be a path with $\omega(0) = p$ and

$\omega(1) = q$, which is transverse to $\tilde{\mathfrak{L}}$ and which crosses each leaf at most once. We put $d(x,y) = \tilde{\mu}(\omega)$. It follows from Lemma 4.8 that d is well-defined and is easily seen to be a metric. Now $\pi_1(M)$ is the group of covering transformations of $\tilde{M}$ and $\tilde{\mathfrak{L}}$ is $\pi_1(M)$-invariant, hence so is X. If $g \in \pi_1(M)$ and ω is a path with endpoints in $\tilde{M} \setminus \tilde{C}$ which is transverse to $\tilde{\mathfrak{L}}$ and which crosses each leaf at most once, then it is left as an exercise to show $\tilde{\mu}(\omega) = \tilde{\mu}(g\omega)$. Hence d is $\pi_1(M)$-invariant. The main result of this section is that (X,d) can be embedded in an $\mathbb{R}$-tree, and the action of $\pi_1(M)$ extends to the $\mathbb{R}$-tree.

Before proceeding to the main theorem, we note a consequence of Lemma 4.7. In the situation of that lemma, and additionally assuming Condition (e), if L is a leaf of $\tilde{\mathfrak{L}}$, $\tilde{M} \setminus L$ has two components. To see this, for each $p \in \tilde{M} \setminus L$, choose a path connected neighbourhood N_p of p in $\tilde{M} \setminus L$, and choose $p' \in N_p$ such that $p' \in \tilde{M} \setminus \tilde{C}$ (recall that $\tilde{C}$ is nowhere dense). Now for p, $q \in \tilde{M} \setminus L$, define $p \sim q$ to mean there is a path from p' to q' transverse to $\tilde{\mathfrak{L}}$ crossing each leaf at most once which does not cross L.

To see this depends only on p and q, it suffices to show it does not depend on the choice of p'. Suppose $p'' \in N_p$ and $p'' \notin \tilde{C}$. Take paths α from p' to q', and β from p'' to q', transverse to $\tilde{\mathfrak{L}}$ and crossing each leaf at most once. Let $\gamma : [0,1] \to N_p$ be a path from p' to p''. For each $t \in [0,1]$, there is a flow box (U_t, ϕ_t) such that $\gamma(t) \in U_t \subseteq N_p$ and $\phi(U_t)$ is convex. By the usual Lebesgue number argument, we can find a partition $0 = t_0 < t_1 < \ldots < t_k = 1$ and flow boxes (U_i, ϕ_i) for $1 \le i \le k$ such that $\gamma([t_{i-1}, t_i]) \subseteq U_i \subseteq N_p$ and $\phi_i(U_i)$ is convex. Since $\tilde{C}$ is nowhere dense, we can assume $\gamma(t_i) \notin \tilde{C}$ for $0 \le i \le k$. Let λ_i be the linear path in $\phi_i(U_i)$ from $\phi_i\gamma(t_{i-1})$ to $\phi_i\gamma(t_i)$ and let $\gamma_i = \phi_i^{-1} \circ \lambda_i$. Then $\gamma_1 \ldots \gamma_k$ is a path in N_p from p' to p'' transverse to $\tilde{\mathfrak{L}}$. Thus we can assume γ is transverse to $\tilde{\mathfrak{L}}$. Now if α crosses L, then so does $\gamma\beta$, by Lemma 4.7. Since γ does not cross L, β crosses L. By symmetry, α crosses L if and only if β does.

If $p \not\sim q$, then we claim that p, q are in different components of $\tilde{M} \setminus L$. For suppose there is a path from p to q not meeting L. Then there is a path α from p' to q' not meeting L. Arguing as in the previous paragraph, we can replace α by β, transverse to $\tilde{\mathfrak{L}}$, such that if β crosses a leaf, so does α. Thus β does not cross L. But then by Lemma 4.7, $p \sim q$, a contradiction.

If $p \not\sim q$ and $p \not\sim r$, we claim that $q \sim r$. For take paths α from p' to q', β from q' to r' and γ from r' to p', transverse to $\tilde{\mathfrak{L}}$ and crossing each leaf at most once. Let σ be a 2-simplex; the paths α, β and γ define a continuous map $\partial\sigma \to \tilde{M}$, which extends to a map $H : \sigma \to \tilde{M}$, since $\tilde{M}$ is simply connected. Arguing as in Lemma 4.7, we can modify H so that $H^{-1}(L)$ has components which are circles or arcs joining two points of $\partial\sigma$, and so that the modified versions of α, β, γ cross the same leaves as the original paths. It follows that there is only one arc, joining a point of α to a point of γ. Hence β (modified or not) does not cross L, so $q \sim r$.

Note that there is a transverse path crossing each leaf at most once which crosses L (e.g. a suitable linear path in a flow box containing points of L). Because $\tilde{C}$ is nowhere dense, we can choose such a path with endpoints outside $\tilde{C}$, and by Lemma 4.7 the endpoints of this path are in different components of $\tilde{M} \setminus L$.

Thus $\tilde{M} \setminus L$ has two components, called *sides* of L; fixing $p \in \tilde{M} \setminus L$, they are $\{q \in \tilde{M} \setminus L \mid q \sim p\}$ and $\{q \in \tilde{M} \setminus L \mid q \not\sim p\}$. Hence, if a transverse path joining two complementary regions of $\tilde{\mathfrak{L}}$ crosses each leaf at most once, its endpoints are on different sides of the leaves crossed by the path, and on the same side of the other leaves.

We can now prove the main theorem. Recall (definition after Lemma 2.1.7) that an edge point of a Λ-tree is one at which there are two directions.

Theorem 4.9. *Let $\mathfrak{L}$ be a measured lamination with a transverse measure of full support on a manifold M satisfying* (a)-(e) *above, and let X be the set of components of $\tilde{M} \setminus \tilde{C}$. Then there exist an $\mathbb{R}$-tree T and an isometry $\phi : X \to T$ such that:*

(1) *T is spanned by $\phi(X)$;*
(2) *any point of $T \setminus \phi(X)$ is an edge point of T;*
(3) *The action of $\pi_1(M)$ on X extends uniquely to an action of $\pi_1(M)$ as isometries on T.*

Between any two such $\mathbb{R}$-trees there is an equivariant isometry commuting with the embeddings of X.

Proof. By Theorem 2.4.4 and the remark after its proof, in order to prove (1) and (3), it suffices to show (X,d) is 0-hyperbolic. Let x, $y \in X$, and let ω be a transverse path in $\tilde{M}$ from $p \in x$ to $q \in y$ which crosses each leaf at most once. If $v \in X$ and ω meets v, say $\omega(t) = r \in v$, then ω can be written as the product of two paths, ω_1 from p to r and ω_2 from r to q, which are transverse to $\mathfrak{L}$ and meet each leaf at most once. Then, denoting the transverse measure for $\tilde{\mathfrak{L}}$ by $\tilde{\mu}$,

$$d(x,y) = \tilde{\mu}(\omega) = \tilde{\mu}(\omega_1) + \tilde{\mu}(\omega_2) = d(x,v) + d(v,y)$$

using additivity of $\tilde{\mu}$.

Conversely, if ω_1 is a path from p to r and ω_2 a path from r to q, both transverse to $\mathfrak{L}$ and crossing each leaf at most once, and ω is their product, then by Lemma 4.8,

$$d(x,y) \leq \tilde{\mu}(\omega) = \tilde{\mu}(\omega_1) + \tilde{\mu}(\omega_2) = d(x,v) + d(v,y)$$

and if $d(x,y) = d(x,v) + d(v,y)$ then ω is a transverse path from x to y crossing each leaf at most once. Now put $[x,y] = \{v \in X \mid (x,y) = d(x,v) + d(v,y)\}$, for x, $y \in X$, so if ω is a transverse path joining x and y crossing each leaf at most once, then $[x,y]$ is the set of complementary regions of $\tilde{C}$ that ω meets. The next stage of the proof is to establish the following lemma.

Lemma.

(i) *the mapping* $[x,y] \to [0, d(x,y)]_{\mathbb{R}}$, $v \mapsto d(x,v)$ *is an isometry, for all* x, $y \in X$;

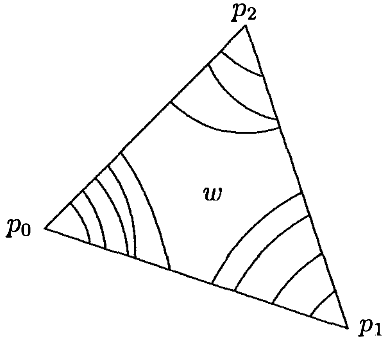
(ii) *for all* x, y, $z \in X$, $[x,y] \cap [x,z] = [x,w]$ *for some* $w \in X$;

(iii) *for all* x, y, $z \in X$, *if* $[x,y] \cap [x,z] = \{x\}$, *then* $[y,z] = [x,y] \cup [x,z]$.

(In the terminology of [63], (X,d) *is an* $\mathbb{R}$*-pretree.)*

Proof. It is easily seen from the description of $[x,y]$ just given that (i) and (iii) hold. To prove (ii), choose points $p_0 \in x$, $p_1 \in y$ and $p_2 \in z$. Let ω_{ij} be a path transverse to $\tilde{\mathcal{L}}$, crossing each leaf at most once, beginning at p_i and ending at p_j, with ω_{ji} the opposite path to ω_{ij}. Let σ be a 2-simplex. The paths ω_{ij} define a map $\partial\sigma \to \tilde{M}$, sending the vertices of σ to p_0, p_1 and p_2, which can be extended to a map $\Omega : \sigma \to \tilde{M}$, since $\tilde{M}$ is simply connected. Arguing as in Lemma 4.7, we can modify the ω_{ij} to smooth paths ω'_{ij} so that ω'_{ij} crosses the same leaves as ω_{ij} exactly once, and in the same order. Further, we can modify Ω so that the inverse images of any finite number of leaves are either circles or arcs joining two boundary points of σ, and there can be at most one arc. By Lemma 4.7, any leaf crossed by ω_{01} is crossed by ω_{02} or ω_{12}, and since we can make its inverse image an arc, it is crossed by exactly one of them.

Also, if $\omega_{01}(t)$ is on a leaf meeting ω_{02} and $\omega_{01}(t')$ is on a leaf meeting ω_{12}, then $t < t'$. For otherwise, we can modify Ω so the inverse images of the two leaves in σ each contain a single arc and these arcs cross, which is impossible since distinct leaves are disjoint. By symmetry, the situation is as illustrated in the following diagram (where Ω^{-1} has been omitted on the labels, and w will be defined later).



Let C_k be the set of leaves meeting both ω_{ki} and ω_{kj}, where i, j and k are distinct. Then $\omega_{ki}^{-1}(C_k)$ and $\omega_{kj}^{-1}(C_k)$ are order isomorphic in the ordering induced from $[0,1]_{\mathbb{R}}$ (the domain of ω_{ij}). For if the points in these sets corresponding to two leaves in C_k occur in opposite order in the two sets, we can modify Ω so the inverse images of these two leaves each contain an arc, and these arcs cross in σ, which is impossible.

Suppose $\omega_{01}^{-1}(C_0)$ has no largest element. Let $l = \sup \omega_{01}^{-1}(C_0)$. Then $l \in \omega_{01}^{-1}(\tilde{C})$ since $\tilde{C}$ is closed. Therefore $\omega_{01}(l)$ lies on a leaf of C_1, which meets ω_{12}, in the point $\omega_{12}(u)$, say. Denote this leaf by L. Similarly, if $l' = \sup \omega_{02}^{-1}(C_0)$, $\omega_{02}(l')$ lies on a leaf L' which meets ω_{12} in a point $\omega_{12}(v)$, say. Note that L and L' are distinct leaves, since a leaf meeting ω_{12} cannot meet both ω_{01} and ω_{02}, hence $u \neq v$. Further, the interval (u, v) is mapped by ω_{12} to a complementary region of $\tilde{C}$, all points of $\omega_{12}^{-1}(C_1)$ are in $[y, u]_{\mathbb{R}}$ and all points of $\omega_{12}^{-1}(C_2)$ are in $[z, v]_{\mathbb{R}}$. Otherwise, we again obtain the contradiction that two arcs belonging to different leaves cross. Choose $p \in (u, v)$.

Suppose $t \in [0,1]_{\mathbb{R}}$, $l < t$ and $\omega_{01}(t) \in \tilde{C}$ (so $\omega_{01}(t) \in C_1$). Let s be the point of $[0,1]_{\mathbb{R}}$ such that $\omega_{12}(s)$ is in the same leaf as $\omega_{01}(t)$. Since $\tilde{C}$ is nowhere dense and has a product structure in each flow box, the interval (l, t) is not contained in $\omega_{01}^{-1}(\tilde{C})$, and we choose $t' \in (l, t)$ such that $\omega_{01}(t') \notin \tilde{C}$. The complementary region R of $\tilde{C}$ containing $\omega_{01}(t')$ is determined by the leaves which ω_{10} crosses in going from y to $\omega_{01}(t')$. In view of the order isomorphism between $\omega_{10}^{-1}(C_1)$ and $\omega_{12}^{-1}(C_1)$, we can find $t'' \in [0,1]$ such that $\omega_{12}(t'')$ lies in R, and $t'' \in (u, s) \subseteq [u, y]_{\mathbb{R}}$.

Now the path from x to $\omega_{12}(p)$ obtained by traversing ω_{01} from x to $\omega_{01}(t')$, then a path in R from $\omega_{01}(t')$ to $\omega_{12}(t'')$, then ω_{12} from $\omega_{12}(t'')$ to $\omega_{12}(p)$ is transverse, crosses only leaves in C_0 and C_1 and does not cross the leaf containing $\omega_{01}(t)$. It follows from Lemma 4.7 that, if α is a transverse path from x to $\omega_{12}(p)$ crossing each leaf at most once, then α meets only leaves in C_0 and possibly L.

However, by a symmetrical argument with ω_{02} instead of ω_{01}, α meets only leaves in C_0 and possibly L'. Since L and L' are distinct leaves, α crosses only leaves in C_0. Further, α crosses all leaves of C_0. For traversing α, followed by ω_{21} from $\omega_{12}(p)$ to y gives a transverse path from x to y crossing each leaf at most once, so by Lemma 4.7 it crosses the same leaves as ω_{01}.

It follows as before that $\omega_{01}^{-1}(C_0)$ and $\alpha^{-1}(C_0) = \alpha^{-1}(\tilde{C})$ are order isomorphic. But $\alpha^{-1}(\tilde{C})$ is closed, so has a maximum element, hence so does $\omega_{01}^{-1}(C_0)$, a contradiction. Thus in fact $\omega_{01}^{-1}(C_0)$ has a maximum element l, and similarly $\omega_{02}^{-1}(C_0)$ has a maximum element l'. By symmetry $\omega_{01}^{-1}(C_1)$ has a minimum element m and $\omega_{02}^{-1}(C_2)$ has a minimum element m', say. Let w be the complementary region of $\tilde{C}$ lying on the opposite side to p_0 of all leaves in C_0 and on the same side of p_0 as all other leaves. Then $l < m$, $l' < m'$ and $\omega_{01}(l, m)$ and $\omega_{02}(l', m')$ belong to w. Further, ω_{01} and ω_{02} cross the same set of complementary regions up to w, after which they cross no region in common. It follows that $[x, y] \cap [x, z] = [x, w]$. □

Returning to the proof of Theorem 4.9, it follows from the lemma that, if $w \in [x,y]$, then $[x,w] \cap [w,y] = \{w\}$. For if $v \in [x,w] \cap [w,y]$, then $d(x,w) = d(x,v) + d(v,w)$ and $d(y,w) = d(y,v) + d(v.w)$, and adding gives $d(x,y) = d(x,v) + d(y,v) + 2d(v,w)$, so by the triangle inequality $d(v,w) = 0$, i.e. $v = w$. It follows from (iii) that $[x,y] = [x,w] \cup [w,y]$.

The argument of Lemma 2.1.2 now shows that, if x, y and $z \in X$, and $[x,y] \cap [x,z] = [x,w]$, then the conclusions of parts (1) and (3) of that lemma are valid in this situation, and the calculation in the proof of part (2) of Lemma 2.1.2 (with $y_m = y$, $z_n = z$) shows that $(y \cdot z)_x = d(x,w)$.

Further, if r, $s \in [x,y]$ and $d(x,r) \leq d(x,s)$, then $r \in [x,s]$. For otherwise, $r \in [s,y]$ and $r \neq s$, so $d(y,r) = d(y,s) - d(r,s) < d(y,s)$, hence $d(x,r) = d(x,y) - d(y,r) > d(x,y) - d(y,s) = d(x,s)$, a contradiction. The argument of Lemma 2.1.6 now shows (X,d) is 0-hyperbolic, as required to show the existence of the $\mathbb{R}$-tree T. Also, if x, y and $z \in X$, and $w \in X$ is the point in (ii) of the lemma, then $\phi(w)$ is the point $\mathrm{Y}(\phi(x), \phi(y), \phi(z))$ in T, since both points lie in the segment of T with endpoints $\phi(x)$ and $\phi(y)$, at distance $(y \cdot z)_x$ from $\phi(x)$.

To prove Part (2) of the theorem, let $v \in T \setminus X$. Then $v \in [\phi(x), \phi(y)]$ for some x, $y \in X$, since $\phi(X)$ spans T, where now square brackets denote segments in T. Thus there are at least two directions at v, given by $[v, \phi(x)]$ and $[v, \phi(y)]$. Suppose there is a third, given by $[v,u]$. Then $[\phi(x), v] \cap [v,u] = \{v\}$, so $[\phi(x), u] = [\phi(x), v, u]$. Since $T = \bigcup_{z \in X} [\phi(x), \phi(z)]$ (see the definition preceding Lemma 2.1.8), $u \in [\phi(x), \phi(z)]$ for some $z \in X$. Thus $[\phi(x), \phi(z)] = [\phi(x), u, \phi(z)] = [\phi(x), v, u, \phi(z)]$, hence $[v, \phi(z)]$ defines the same direction at v as $[v,u]$. It follows that $v = \mathrm{Y}(\phi(x), \phi(y), \phi(z)) \in \phi(X)$, a contradiction. Therefore there are exactly two directions at v in T. □

If $\mathfrak{L}$ satisfies Conditions (a) to (e) above, the tree T in Theorem 4.9 will be called the dual $\mathbb{R}$-tree to $\mathfrak{L}$.

Lemma 4.10. *Assume the measured lamination $\mathfrak{L}$ satisfies Conditions (a)-(e) and let T be the dual $\mathbb{R}$-tree. For any $x \in X$, the stabilizer in $\pi_1(M)$ of $\phi(x)$ for the action on T is equal to the stabilizer of x. For any $y \in T \setminus \phi(X)$, there is a leaf of $\tilde{\mathfrak{L}}$ whose stabilizer contains that of y.*

Proof. The first part of the lemma is clear. Suppose $y \in T \setminus \phi(X)$. Since $\phi(X)$ spans T, and $T \setminus \{y\}$ has two components, by Lemma 2.2.2 and Theorem 4.9(2), we can choose points $\phi(c)$, $\phi(c') \in \phi(X)$ in different components of $T \setminus \{y\}$. Take a path α from c to c' transverse to $\tilde{\mathfrak{L}}$, crossing each leaf at most once. For $t \in [0,1] \setminus \alpha^{-1}(\tilde{C})$, let c_t be the complementary region of $\tilde{C}$ containing $\alpha(t)$. Let U be the set of points t in $[0,1] \setminus \alpha^{-1}(\tilde{C})$ such that $\phi(c_t)$ belongs to the same component of $T \setminus \{y\}$ as $\phi(c)$, and let V be the set of points t in $[0,1] \setminus \alpha^{-1}(\tilde{C})$ such that $\phi(c_t)$ belongs to the same component of $T \setminus \{y\}$ as $\phi(c')$. From the description of the sets $[a,b]$ (a, $b \in X$) in the lemma above, all points of U occur before all points of V in $[0,1]$. Hence $\sup(U) = \inf(V) = l$, say, otherwise the interval between them maps to a complementary region of $\tilde{C}$, so is in either U or V, which is impossible. Further, l cannot belong to U or V, so $\alpha(l)$ is in a leaf of $\tilde{\mathfrak{L}}$ which separates the two sets of complementary regions containing points of $\alpha(U)$ and of $\alpha(V)$. It follows easily from Lemma 2.1.7 that L is independent of the choice of c and c', and we denote it by L_y. Clearly, if $g \in \pi_1(M)$ then $gL_y = L_{gy}$, and it follows that the stabilizer of y is contained in the stabilizer of L_y. □

It will be convenient to have a description of a transverse measure independent of the flow boxes. If κ is an arc in a manifold M, given by a homeomorphism α from a compact interval $[a,b]_{\mathbb{R}}$ onto κ, the corresponding *open arc* is $\alpha\big((a,b)\big)$, denoted by κ_o.

Lemma 4.11. *Let $\mathfrak{L}$ be a measured lamination with a transverse measure of full support on a manifold M, all of whose leaves are closed in M. Then there is an induced regular Borel measure μ_κ on each arc κ homeomorphic to* $[0,1]$ *embedded transversely to $\mathfrak{L}$, satisfying:*

(i) *For all such arcs κ, the support of μ_κ restricted to κ_o is $|\mathfrak{L}| \cap \kappa_o$.*

(ii) *If κ, κ' are two such arcs, homotopic by means of arcs embedded transversely to $\mathfrak{L}$, by a homotopy whose extremities remain in the same leaf of $\mathfrak{L}$ or in the same component of $M \setminus |\mathfrak{L}|$, then $\mu_\kappa(\kappa) = \mu_{\kappa'}(\kappa')$.*

(iii) *If κ, κ' are two such arcs and A is a Borel subset of κ and of κ', then $\mu_\kappa(A) = \mu_{\kappa'}(A)$.*

Conversely, given a lamination and a regular Borel measure μ_κ on each arc κ homeomorphic to $[0,1]$ *embedded transversely to $\mathfrak{L}$, satisfying* (i), (ii) *and* (iii), *there is a transverse measure of full support on the lamination inducing the Borel measures μ_κ.*

Proof. Given a transverse measure μ for $\mathfrak{L}$ of full support, and arc κ embedded transversely to $\mathfrak{L}$, κ is the image of some homeomorphism α from a compact real interval J. The hypothesis that κ is embedded transversely means that α is transverse to $\mathfrak{L}$. We can now define the measure of any compact subset $K = \alpha(C)$, where C is a closed subset of J, by slightly modifying the definition of $\mu(\alpha)$ as given above. Thus, let (U, ϕ) be a flow box, with $\phi(U) = V \times I$ where I is an interval, let $p : V \times I \to I$ be the projection map, assume α maps into U and $p\phi\alpha$ is monotone over the local leaf space X. Then $p\phi$ restricted to $\kappa \cap |\mathfrak{L}|$ is one-to-one, since $\phi(|\mathfrak{L}|) = U \times X$, and $\kappa \cap |\mathfrak{L}|$ is compact, I is Hausdorff, so $p\phi$ maps $\kappa \cap |\mathfrak{L}|$ homeomorphically onto its image. We can therefore define a Borel measure μ_κ on κ by $\mu_\kappa(A) = \mu(p\phi(A \cap |\mathfrak{L}|)$ for Borel subsets A of κ, and it is left as an exercise to show that μ_κ is regular. Note that $\mu_\kappa(A) = \mu(p\phi(A) \cap X)$, since $\phi(|\mathfrak{L}|) = U \times X$.

Let U be an open subset of κ_o. If $U \cap |\mathfrak{L}| = \emptyset$ then clearly $\mu_\kappa(U) = 0$. If $U \cap |\mathfrak{L}| \neq \emptyset$, take $x \in U \cap |\mathfrak{L}|$. Then $p\phi(x) \in K \subseteq p\phi(U)$ for some open subinterval K of I, by transversality of κ. Since $K \setminus X$ is an open subset of I and $X = \text{supp}(\mu)$, $\mu(K \setminus X) = 0$, so if $\mu(K \cap X) = 0$ then $\mu(K) = 0$, and since K is open in I, $K \subseteq I \setminus X$, contradicting $p\phi(x) \in X$. Thus $\mu(p\phi(U) \cap X) \geq \mu(K \cap X) > 0$, that is, $\mu_\kappa(U) > 0$. It follows that μ_κ satisfies condition (i).

In general $\mu_\kappa(K)$ is defined by taking partitions of J, with subintervals J_i as previously, avoiding point masses and defining $\mu_\kappa(K)$ to be $\sum_i \mu_{\kappa_i}(K_i)$, where $K_i = \alpha(C \cap J_i)$ and κ_i is $\alpha(J_i)$, first assuming α maps into a flow box, then in general. One checks that μ_κ is a regular content, so by Proposition 4.3, μ_κ extends to a Borel measure on κ, which satisfies (i) by an argument similar to that already given. To prove (ii), one can argue as in Lemma 4.8, using Lemma 4.2 instead of Lemma 4.7. Details are left to the reader, as is the proof of (iii), which is not difficult.

Suppose conversely that the measures μ_κ are given satisfying (i), (ii) and (iii). Let (U, ϕ) be a flow box with $\phi(U) = V \times I$, where I is a bounded open real interval. Choose $v \in V$, and if C is a compact subset of I, a compact interval $J \subseteq I$ with $C \subseteq J$. Then put $\kappa = \phi^{-1}(\{v\} \times J)$, an arc in M, and define $\mu(C) = \mu_\kappa(\phi^{-1}(\{v\} \times C))$. If $v' \in V$, take a path $\alpha : [0,1] \to V \times I$ with $\alpha(0) = v$, $\alpha(1) = v'$. Define a homotopy $H : J \times [0,1] \to V \times I$ by $H(t,s) = (\alpha(s), t)$. Then applying (ii) to $\phi^{-1} \circ H$, we see that the definition does not depend on the choice of v, and it is independent of the choice of J by (iii). By Proposition 4.3, μ extends to a transverse measure in (U, ϕ). Before proceeding we make a remark.

Remark. Suppose κ is an arc embedded transversely to $\mathfrak{L}$. Then
(a) if $x \in \kappa \setminus |\mathfrak{L}|$ then $\mu_\kappa(\{x\}) = 0$;
(b) if A is a Borel subset of κ, then $\mu_\kappa(A) = \mu_\kappa(A \cap |\mathfrak{L}|)$.

For we can find an arc κ' with $x \in \kappa'_o$ and $\kappa' \cap |\mathfrak{L}| = \emptyset$ (since M is a manifold and $M \setminus |\mathfrak{L}|$ is open). Then $\mu_\kappa(\{x\}) = \mu_{\kappa'}(\{x\})$ by (iii), and $\mu_{\kappa'}(\{x\}) = 0$ by (i). If p, q are the endpoints of κ and A is a Borel set, then

$$\begin{aligned}\mu_\kappa(A) &= \mu_\kappa(A \cap \kappa_o) + \mu_\kappa(A \cap \{p\}) + \mu_\kappa(A \cap \{q\}) \\ &= \mu_\kappa(A \cap |\mathfrak{L}| \cap \kappa_o) + \mu_\kappa(A \cap |\mathfrak{L}| \cap \{p\}) + \mu_\kappa(A \cap |\mathfrak{L}| \cap \{q\}) \\ &\quad \text{(by (i) and part (a))} \\ &= \mu_\kappa(A \cap |\mathfrak{L}|).\end{aligned}$$

If J is an interval in I such that $J \cap X = \emptyset$, where X is the local leaf space, then $\mu(J) = 0$. For we can write $J = \bigcup_{n=1}^\infty J_n$, where each J_n is a compact interval, so $\mu(J) \leq \sum_{n=1}^\infty \mu(J_n)$, and from definition of μ and the remark, $\mu(J_n) = 0$ for all n. If U is an open subset of I such that $U \cap X = \emptyset$, then U is a countable union of open intervals, so $\mu(U) = 0$. On the other hand, if U is an open subset of I meeting X, say $x \in U \cap X$, then $\mu(U) > 0$. For we can find an open interval J with $x \in J \subseteq \overline{J} \subseteq U$, and putting $\kappa = \phi^{-1}(\{v\} \times \overline{J})$,

$$\mu(\overline{J}) = \mu_\kappa(\kappa) \geq \mu_\kappa(\phi^{-1}(\{v\} \times J)) > 0$$

by (i), since $\phi^{-1}(\{v\} \times J)$ is an open subset of κ_0 meeting $|\mathcal{L}|$. It follows that μ is of full support. It remains to show that the measures $\mu_{(U,\phi)}$ we have defined in the flow boxes are compatible, so define a measured lamination, and this induces the original measures μ_κ. The proof is rather lengthy, and we first prove a technical lemma.

Lemma 4.12. *Let (U, ϕ) be a flow box with $\phi(U) = V \times I$, where I is a bounded open interval, and let X be the local leaf space (a closed subset of I). Let J be a compact real interval and $\alpha : J \to U$ be a continuous map which is transverse to $|\mathcal{L}|$ in U and such that $p\phi\alpha$ is strictly monotone over X. Put $p\phi\alpha(J) = [c, d]$. Then:*

1. *There is a continuous map $\beta : J' \to U$, where J' is a compact subinterval of J, such that $\beta(t) = \alpha(t)$ for t such that $p\phi\alpha(t) \in X$ (i.e. $\alpha(t) \in |\mathcal{L}|$), $p\phi\beta$ is strictly monotone over I, $\beta(J') \cap |\mathcal{L}| = \alpha(J) \cap |\mathcal{L}|$ and $p\phi\beta(J') = [c, d]$;*
2. *if $p\phi\alpha$ is strictly monotone over I, so α is an arc, say κ, then for every Borel set A in κ, $\mu_\kappa(A) = \mu_{(U,\phi)}(p\phi(A))$.*

Proof. (1) Choose t_0 such that $\alpha(t_0) = c$ and t_1 such that $\alpha(t_1) = d$ and let $J' = [t_0, t_1]$. Define $\beta(t_0) = c$ and $\beta(t_1) = d$. For t such that $p\phi\alpha(t) \in X$, $t \in J'$ since α is strictly monotone over X, and we define $\beta(t) = \alpha(t)$. The set of points $t \in J'$ for which $\beta(t)$ has not been defined is open, so a disjoint union of intervals of the form (a, b), where $a < b$. Note that $\beta(a)$ and $\beta(b)$ are already defined and equal to $\alpha(a)$, $\beta(b)$ respectively. Let γ be $\phi \circ \alpha$ composed with the projection map $V \times I \to V$. Define β on (a, b) by

$$\phi\beta(t) = \left(\gamma(t), \left(\frac{b-t}{b-a}\right) p\phi\beta(a) + \left(\frac{t-a}{b-a}\right) p\phi\beta(b)\right).$$

(This defines β since ϕ is a homeomorphism.) It follows from the continuity of α that β is continuous (note that, if $c \notin X$ then there is an interval $[c, e)$ not intersecting X with $e > c$, and similarly for b). It is easy to see that β has the required properties.

(2) For convenience reparametrise α so that $J = [0, 1]$. Choose $v \in V$ and let κ' be the arc $\phi^{-1}(\{v\} \times [c, d])$. Let α' be $\phi \circ \alpha$ composed with the projection map $V \times I \to V$, and let $\gamma : [0, 1] \to V$ be a path with $\gamma(0) = \alpha'(0)$ and $\gamma(1) = v$. Define a homotopy $H : [0, 1] \times [0, 1] \to V \times I$ by

$$H(t, s) = \begin{cases} (\alpha'((1-2s)t), p\phi\alpha(t)) & \text{if } 0 \leq s \leq \frac{1}{2}; \\ (\gamma(2s-1), p\phi\alpha(t)) & \text{if } \frac{1}{2} \leq s \leq 1. \end{cases}$$

The map $\phi^{-1}H(t,0) \mapsto \phi^{-1}H(t,1)$, that is $\alpha(t) \mapsto \phi^{-1}(v, p\phi\alpha(t))$ is a homeomorphism, so defines a bijection from subsets of κ to subsets of κ', $A \mapsto A'$, sending compact sets to compact sets, and so Borel sets to Borel sets. If A is a compact interval, $\mu_\kappa(A) = \mu_{\kappa'}(A')$ by (ii) (using the homotopy $\phi^{-1} \circ H$ restricted to A). Also, we can define a regular Borel measure μ' on κ by $\mu'(A) = \mu_{\kappa'}(A')$ for Borel sets A in κ. Thus $\mu'(A) = \mu_\kappa(A)$ for compact intervals A, so by uniqueness in Corollary 4.6, $\mu' = \mu_\kappa$, that is, $\mu_\kappa(A) = \mu_{\kappa'}(A')$ for all Borel sets A. Now $A' = \phi^{-1}(\{v\} \times p\phi(A))$, and so $\mu_{\kappa'}(A') = \mu_{(U,\phi)}(p\phi(A))$. For this is true if A is compact, by definition of $\mu_{(U,\phi)}$, and so for all Borel subsets A by an argument like that just given, using Corollary 4.6. □

Returning to the proof of Lemma 4.11, in the situation of Part (1) of Lemma 4.12, we can compute $\mu(\alpha)$ using the measure $\mu_{(U,\phi)}$. Letting $\kappa = \beta(J')$, we obtain

$$\begin{aligned}\mu(\alpha) &= \mu_{(U,\phi)}(p\phi\alpha(J)) \\ &= \mu_{(U,\phi)}(p\phi\alpha(J) \cap X) = \mu_{(U,\phi)}(p\phi(\alpha(J) \cap |\mathcal{L}|)) \\ &= \mu_{(U,\phi)}(p\phi(\kappa \cap |\mathcal{L}|)) = \mu_\kappa(\kappa \cap |\mathcal{L}|)\end{aligned}$$

by Part (2) of Lemma 4.12. If we replace the flow box by another, say (U_1, ϕ_1), with $\alpha(J) \subseteq U$, so the hypotheses of Part (1) of Lemma 4.12 are satisfied, then we obtain an analogous arc κ_1 such that, measuring $\mu(\alpha)$ in this flow box, $\mu(\alpha) = \mu_{\kappa_1}(\kappa_1 \cap |\mathcal{L}|)$ and $\kappa_1 \cap |\mathcal{L}| = \alpha(J) \cap |\mathcal{L}|$. By (iii), this gives the same answer for $\mu(\alpha)$. It follows easily from the definition of $\mu(\alpha)$ that the measures in the flow boxes are compatible, and we have defined a transverse measure for $\mathcal{L}$.

Now in Part (1) of Lemma 4.12, let α be a parametrisation of an arc κ, and let κ' be the arc parametrised by β. For $A \subseteq \kappa$, $A \cap |\mathcal{L}| \subseteq \kappa'$, and the collection of Borel sets A of κ such that $A \cap |\mathcal{L}|$ is a Borel subset of κ' is easily seen to be a σ-ring, and contains all compact subsets of κ. Hence for all Borel sets A in κ, $A \cap |\mathcal{L}|$ is a Borel subset of κ'. By the remark, hypothesis (iii) and (2) of Lemma 4.12, for any Borel set A,

$$\begin{aligned}\mu_\kappa(A) &= \mu_\kappa(A \cap |\mathcal{L}|) \\ &= \mu_{\kappa'}(A \cap |\mathcal{L}|) \\ &= \mu_{(U,\phi)}(p\phi(A \cap |\mathcal{L}|)) = \mu_{(U,\phi)}(p\phi(A) \cap X) \\ &= \mu_{(U,\phi)}(p\phi(A)) \quad (\text{since } X = \operatorname{supp}(\mu_{(U,\phi)})).\end{aligned}$$

It follows easily (and is left as an exercise) that the transverse measure induces the original measures μ_κ. □

Lemma 4.11 enables us to construct examples of measured laminations. Let $\mathcal{L}$ be a lamination on a manifold M with finitely many leaves, $\{L_i\}_{i\in I}$ (it is left to the reader to construct examples of such laminations). Associate to each L_i a positive real number μ_i. If κ is a transversely embedded arc and A is a Borel subset of κ, put $\mu_\kappa(A) = \sum_{i \in I} \mu_i |A \cap L_i|$, where $|A \cap L_i|$ means the cardinality of $A \cap L_i$. This is finite since κ can be covered by finitely many flow boxes, in each of which it crosses every leaf at most once.

In the next section, we shall make brief use of the tangent bundle of a smooth manifold M. There are various ways of defining the tangent space $\mathrm{T}_x(M)$ at a point $x \in M$. The most natural seems as follows. Let (U, ϕ) be a chart with $x \in U$. If α, β are smooth curves in M (smooth curve meaning a C^∞ map from a real interval into M) defined at 0 and such that $\alpha(0) = \beta(0) = x$. Define $\alpha \sim \beta$ to mean $(\phi \circ \alpha)'(0) = (\phi \circ \beta)'(0)$, where the prime denotes differentiation. This is easily seen to be independent of the choice of chart and an equivalence relation. We define $\mathrm{T}_x(M)$ to be the set of equivalence classes, and denote the equivalence class of α by $[\alpha]$. The mapping $\mathrm{T}_x(M) \to \mathbb{R}^n$, $[\alpha] \mapsto (\phi \circ \alpha)'(0)$, is a bijection (where M is an n-manifold). This induces a topology on $\mathrm{T}_x(M)$ from the Euclidean topology, as well as a vector space structure, both of which are independent of the choice of chart.

The tangent bundle $\mathrm{T}(M)$ is defined to be $\coprod_{x \in M} \mathrm{T}_x(M)$, the disjoint union of the tangent spaces, and we view elements of $\mathrm{T}(M)$ as ordered pairs $(x, [\alpha])$, where $x \in M$ and α is a smooth curve with $\alpha(0) = x$. We topologise $\mathrm{T}(M)$ as follows. Let $\pi : \mathrm{T}(M) \to M$ be the projection map: $\pi(x, [\alpha]) = x$. If (U, ϕ) is a chart, there is a bijective map $t : \pi^{-1}(U) \to U \times \mathbb{R}^n$, and we can take the sets $t^{-1}(A)$, where A is an open subset of $U \times \mathbb{R}^n$ and (U, ϕ) runs through the charts in a maximal C^∞ structure, as a basis for the topology on $\mathrm{T}(M)$. This makes $\mathrm{T}(M)$ into a $2n$-manifold. Further details can be found in any text on differential geometry, and there is a comprehensive account of different ways of defining $\mathrm{T}(M)$ in [135], Chapter 3.

5. Hyperbolic Surfaces

In the last section we developed the theory of measured laminations without giving many examples. We shall now consider the examples which mainly concern us, namely measured geodesic laminations on hyperbolic surfaces. Before we can give the definition of such a lamination, we shall need to develop some of the theory of hyperbolic surfaces.

We assume some familiarity with hyperbolic geometry. This is discussed in two of our main references, [9] and [82; Ch.5], as well as [83]. Our other main source is [27]; this also contains information on hyperbolic geometry in Ch.1. We denote by $\mathbb{H}^2$ the hyperbolic plane, and although we shall have occasion to use other models, we shall always view this as the Poincaré disk model, unless stated otherwise. Thus the underlying set is the interior of the unit disk in $\mathbb{R}^2$, where the geodesics (straight lines) are the arcs of circles inside the disk which intersect the boundary at right angles. Note that $\mathbb{H}^2$ is a geodesic $\mathbb{R}$-metric space, whose segments are closed segments of hyperbolic lines. In the disk model, the hyperbolic lines intersect the boundary circle S^1 in two points, which are called endpoints or ideal points of the line. We shall occasionally pass between the Poincaré disk model and the upper half-plane model, and the reader is expected to be able to do this without difficulty. It can be accomplished by the transformation $z \mapsto \dfrac{z-i}{z+i}$ from the upper half-plane to the disk.

We recall ([9; §7.4]) that, using the upper half-plane model, isometries of $\mathbb{H}^2$ are of two kinds:

$$z \mapsto \frac{az+b}{cz+d} \tag{1}$$

$$z \mapsto \frac{a\bar{z}+b}{c\bar{z}+d} \tag{2}$$

where a, b, c, $d \in \mathbb{R}$, $ad-bc > 0$ in Case (1) and $ad-bc < 0$ in Case (2). Isometries of type (1) are called orientation-preserving isometries, those of type (2) orientation-reversing. The set of orientation-preserving isometries forms a subgroup of $\mathrm{Isom}(\mathbb{H}^2)$ of index 2, denoted $\mathrm{Isom}^+(\mathbb{H}^2)$. There is a surjective homomorphism from $\mathrm{GL}_2(\mathbb{R})$ to $\mathrm{Isom}(\mathbb{H}^2)$, sending the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ to the transformation (1) if its determinant is positive, and to the transformation (2) if the determinant is negative. The kernel is $\{\alpha I \mid \alpha \in \mathbb{R},\ \alpha \neq 0\}$, where I is the identity matrix, so $\mathrm{Isom}(\mathbb{H}^2)$ is isomorphic to $\mathrm{PGL}_2(\mathbb{R})$. Given an isometry of type (1), we can assume (dividing through by $\sqrt{ad-bc}$) that $ad-bc = 1$. Thus, restricted to $\mathrm{SL}_2(\mathbb{R})$, the homomorphism maps onto $\mathrm{Isom}^+(\mathbb{H}^2)$, so $\mathrm{Isom}^+(\mathbb{H}^2)$ is isomorphic to $\mathrm{PSL}_2(\mathbb{R})$.

By an easy calculation of fixed points ([27; p.14]), non-identity elements of $\mathrm{Isom}^+(\mathbb{H}^2)$ fall into three subtypes: *elliptic* ($|a+d| > 2$), *parabolic* ($|a+d| = 2$) and *hyperbolic* ($|a+d| < 2$). An elliptic isometry has a unique fixed point in $\mathbb{H}^2$, a parabolic isometry has exactly one fixed point on $\partial\mathbb{H}^2$, and a hyperbolic isometry has exactly two fixed points on $\partial\mathbb{H}^2$. Note that in the disk model, the Euclidean boundary $\partial\mathbb{H}^2$ is S^1, while in the upper half-plane model $\partial\mathbb{H}^2 = \mathbb{R} \cup \{\infty\}$.

Given two hyperbolic lines, there is an orientation-preserving isometry sending one to the other (see Exercise 1.1 in [83]), so sending the endpoints of one to the endpoints of the other. Thus, a hyperbolic isometry is conjugate in $\mathrm{Isom}^+(\mathbb{H}^2)$ to one fixing 0 and ∞ (in the upper half-plane model), which is of the form $z \mapsto a^2 z$, where $a \in \mathbb{R}$, $a \neq 0, 1$. Similarly, a parabolic isometry is conjugate to one fixing ∞, which is of the form $z \mapsto z+b$, $b \in \mathbb{R}$, $b \neq 0$.

A similar calculation shows that an isometry of type (2) either fixes a hyperbolic line pointwise ($a+d = 0$) or two points on $\partial\mathbb{H}^2$ ($a+d \neq 0$). These are conjugate in $\text{Isom}(\mathbb{H}^2)$ to an isometry $z \mapsto -a^2\bar{z}$, where $a \neq 0$, and $a = 1$ in the first case, $a \neq 1$ in the second. In the first case, the isometry is hyperbolic reflection in the fixed line (exercise-see [9; §3.1] and [27; p.4]). In the second case, the isometry is called a (hyperbolic) *glide reflection* and is the product of a reflection and a hyperbolic isometry, both fixing the same hyperbolic line, because $z \mapsto -a^2\bar{z}$ is the product of $z \mapsto -\bar{z}$ and $z \mapsto a^2 z$.

Thus an element of finite order in $\text{Isom}(\mathbb{H}^2)$ is either elliptic or a reflection. Reflections have order 2, and in the disk model, elliptic elements are conjugate to elements fixing the origin, which are Euclidean rotations $z \mapsto e^{i\theta}z$ (see [9; §7.33]). The elliptic elements and reflections are precisely the elements of $\text{Isom}(\mathbb{H}^2)$ having a fixed point in $\mathbb{H}^2$. Hence the only element of $\text{Isom}(\mathbb{H}^2)$ which fixes an open set is the identity, so if two isometries agree on an open set, they are equal.

Exercise. Let W be an open connected subset of $\mathbb{H}^2$, and let $\phi : W \to \mathbb{H}^2$ be a mapping, such that every point of W has a neighbourhood such that ϕ restricted to this neighbourhood equals the restriction of an isometry of $\mathbb{H}^2$. Show that ϕ is the restriction to W of a unique element of $\text{Isom}(\mathbb{H}^2)$.

Definition. Let S be a surface (i.e. a connected 2-manifold). A hyperbolic structure on S is a P-atlas on S, where P is the pseudogroup generated by all homeomorphisms between open subsets of $\mathbb{H}^2$ which are restrictions of isometries of $\mathbb{H}^2$.

This means that, if (U_i, ϕ_i) and (U_j, ϕ_j) are two charts, every point $x \in U_i \cap U_j$ has an open neighbourhood V such that the transition map ϕ_{ij}, restricted to $\phi_i(V)$, is the restriction of an isometry of $\mathbb{H}^2$. Such isometries are C^∞ maps on $\mathbb{H}^2$, so any hyperbolic structure is also a C^∞ structure. Also, it follows that if W is any connected open subset of $U_i \cap U_j$, then the restriction of ϕ_{ij} to $\phi_i(W)$ is the restriction of a unique isometry, by the exercise above. The most obvious example of a hyperbolic structure is $S = \mathbb{H}^2$, with a single chart $(\mathbb{H}^2, id_{\mathbb{H}^2})$.

If P_0 is the pseudogroup generated by all homeomorphisms between open subsets of $\mathbb{H}^2$ which are restrictions of orientation-preserving isometries of $\mathbb{H}^2$, then a P_0-atlas is a hyperbolic structure, and makes S into an oriented manifold, in the usual sense of differential geometry ([22], Definition 2.6, Ch.2).

As with any P-atlas, any hyperbolic structure extends uniquely to a maximal hyperbolic structure, and when we refer to a chart we mean a chart in this maximal structure. Also, any hyperbolic structure lifts to a hyperbolic structure on a covering space. We call a surface with a hyperbolic structure a hyperbolic surface.

Let S be a hyperbolic surface. We can make S into a metric space, by first defining a notion of length of paths, as follows. We do not want to confine attention to paths defined on the unit interval, so in this section "path" will mean a continuous map $\alpha : [a,b]_{\mathbb{R}} \to S$ where a, $b \in \mathbb{R}$ and $a \leq b$. First suppose that the image of the path α is contained in a chart (U, ϕ), such that the hyperbolic length of $\phi \circ \alpha$ is defined ([9; §7.2]). We define the length of α to be the hyperbolic length of $\phi \circ \alpha$. This is independent of the choice of chart, for if (U', ϕ') is another chart containing the image of α, then $\phi' \circ \alpha = (\phi' \circ \phi^{-1}) \circ (\phi \circ \alpha)$, and the restriction of $\phi' \circ \phi^{-1}$ to the component of $\phi(U \cap U')$ containing the image of $\phi \circ \alpha$ is the restriction of an isometry, and these preserve hyperbolic length in $\mathbb{H}^2$. (For orientation-preserving isometries see Theorem 5.3.1 in [82], and it is easily seen that $z \mapsto -\bar{z}$ preserves hyperbolic length. Together, these generate Isom($\mathbb{H}^2$).) We denote the length of α by $l(\alpha)$. Note that length is additive, in the sense that if $c \in [a,b]$ and $\alpha_1 = \alpha|_{[a,c]}$, $\alpha_2 = \alpha|_{[c,b]}$, then $l(\alpha)$ is defined if and only if $l(\alpha_1)$ and $l(\alpha_2)$ are defined, in which case $l(\alpha) = l(\alpha_1) + l(\alpha_2)$. This follows from the additivity of the integral used to define hyperbolic length.

Next, suppose there is a partition $P: \ t_0 < t_1 < \ldots < t_n$ of $[a,b]_{\mathbb{R}}$ and charts (U_i, ϕ_i) such that $\alpha(t) \in U_i$ for $t_{i-1} \leq t \leq t_i$ and $1 \leq i \leq n$. Call such a partition a *suitable* partition for α. If $l(\alpha|[t_{i-1}, t_i])$ is defined for all i, we define $l(\alpha) = \sum_{i=1}^{n} l(\alpha|_{[t_{i-1},t_i]})$. From the previous paragraph, this sum depends only on the partition, not on the choice of the U_i. Refining the partition gives another suitable partition, and does not change the sum, by additivity. Since two partitions have a common refinement, the sum is independent of the choice of partition. For similar reasons, this sum is defined for any suitable partition. Further, with this extended definition, l remains additive in the sense above.

Exercise. If $\alpha : [a,b] \to S$ is continuous, and the opposite path $\beta : [a,b] \to S$ is defined by $\beta(t) = \alpha(a + b - t)$, show that $l(\alpha)$ is defined if and only if $l(\beta)$ is, in which case $l(\alpha) = l(\beta)$.

Now consider a reparametrisation of α, that is, let $\beta(t) = \alpha(f(t))$, where $f : [c,d] \to [a,b]$ is a piecewise continuously differentiable, strictly monotone function. Assume f is strictly increasing. Then if $t_0 < t_1 < \ldots < t_n$ is a suitable partition for α, put $s_i = f^{-1}(t_i)$ to get a suitable partition for β. Since arclength in $\mathbb{H}^2$ is invariant under reparametrisation (an easy exercise), $l(\beta|[s_{i-1}, s_i]) = l(\alpha|[t_{i-1}, t_i])$, so $l(\beta) = l(\alpha)$. If f is strictly decreasing, the same argument works, except that the s_i need renumbering (or one can deduce it from the increasing case using the exercise).

We can now define the metric. If x, $y \in S$, let α be a path in S from x to y. For each point t in the domain of α we can find a chart (U_t, ϕ_t) such that $\alpha(t) \in U_t$ and $\phi_t(U_t)$ is a hyperbolic disk. Let δ be a Lebesgue number for the covering $\{\alpha^{-1}(U_t)\}$ of the domain, and let $P: \ t_0 < t_1 < \ldots < t_n$ be a partition of the domain of α with $t_i - t_{i-1} < \delta$ for $1 \leq i \leq n$. Then there are charts (U_i, ϕ_i) such that $\alpha(t) \in U_i$ for $t_{i-1} \leq t \leq t_i$ and $\phi_i(U_i)$ is a disk in $\mathbb{H}^2$. Let β_i be a path in U_i from $\alpha(t_{i-1})$ to $\alpha(t_i)$ such that $l(\beta_i)$ is defined (e.g.

$\phi_i^{-1} \circ \gamma_i$, where γ_i is a suitable parametrisation of a hyperbolic line segment). If α' is the concatenation $\beta_1 \dots \beta_n$, then $l(\alpha')$ is defined. We can now define, for x, $y \in S$, $d(x,y) = \inf_\alpha l(\alpha)$, the infimum being over all paths α from x to y such that $l(\alpha)$ is defined. It is easily checked that d is an $\mathbb{R}$-pseudometric. We claim (S,d) is an $\mathbb{R}$-metric space.

Suppose x, $y \in S$ with $x \neq y$. Choose a chart (U,ϕ) such that $x \in U$ and $\phi(U)$ is a disk, and let V be an open neighbourhood of x contained in U such that $\phi(V)$ is an open hyperbolic disk in the interior of $\phi(U)$, centre $\phi(x)$ and radius r, such that $y \notin V$. Suppose $\alpha : [0,1] \to S$ is a path from x to y. Let $s = \inf\{t \in [0,1] \mid \alpha(t) \notin V\}$. It is easy to see that $\alpha(s)$ is a boundary point of V, hence $\rho(\phi(\alpha(s)), \phi(x)) = r$, where ρ is the hyperbolic metric ([9; §7.2]), and $\alpha(s) \in U$. Suppose $l(\alpha)$ is defined and choose a suitable partition $0 = t_0 < t_1 < \dots t_n = 1$ of $[0,1]$, so $l(\alpha) = \sum_{i=1}^n l(\alpha|_{[t_{i-1},t_i]})$. We can assume $t_1 = s$ (add s to the partition if necessary, then remove all t_i with $0 < t_i < s$; the partition remains suitable). Then $l(\alpha) \geq l(\alpha|_{[0,s]}) \geq \rho(\phi(\alpha(s)), \phi(x)) = r$, the second inequality being by definition of ρ. Hence $d(x,y) \geq r > 0$, so d is a metric as claimed.

Exercise. If x, y are points in a chart (U,ϕ), show that $d(x,y) = \rho(\phi(x), \phi(y))$.

If $p : \tilde{S} \to S$ is a covering map, then $\tilde{S}$ inherits a hyperbolic structure from S, so there is a corresponding metric $\tilde{d}$ on $\tilde{S}$. If (U,ψ) is a chart for the hyperbolic structure on $\tilde{S}$, and γ is a covering (deck) transformation, then $(\gamma U, \psi\gamma^{-1})$ is also a chart. It follows that, if α is a path in $\tilde{S}$, then $l(\alpha) = l(\gamma\alpha)$, hence γ preserves length. It follows that γ is an isometry of the $\mathbb{R}$-metric space $(\tilde{S}, \tilde{d})$.

We shall also need the idea of a geodesic curve in a hyperbolic surface. Our definition is based on that used in Riemannian geometry. Let l be a hyperbolic line. We can parametrise l by arclength measured from some fixed point. Any parametrisation of l obtained from this by a linear change of variable $s \mapsto as+b$, where a, b are real constants and $a \neq 0$, is called a *geodesic parametrisation* of l.

Suppose S is a hyperbolic surface. A *curve* on S is a continuous map $\alpha : J \to S$ for some real interval J (not necessarily compact). Such a curve α is called a *geodesic curve* if J is non-degenerate and, for all $t \in J$, there are a neighbourhood I of t in J and a chart (U,ϕ) such that $\alpha(I) \subseteq U$ and $\phi \circ \alpha$ is the restriction to I of a geodesic parametrisation of a hyperbolic line. This condition does not depend on the choice of chart (exercise-see [82; Theorem 5.3.1]).

Given such a geodesic curve α, choose $t_0 \in J$. We define arclength $s(t)$ by putting $s(t) = l(\alpha|_{[t_0,t]})$ if $t \geq t_0$ and $s(t) = -l(\alpha|_{[t,t_0]})$ if $t \leq t_0$. If $t \in J$, and $\alpha(t) \in U$, where (U,ϕ) is a chart, then s is a linear function of t in a neighbourhood of t, hence s is a linear function of t on J. Note that

s is a strictly increasing function of t, so we can parametrise the curve by arclength. Because hyperbolic arclength is unchanged by reparametrisation, if β is a reparametrisation of α, and β is a geodesic curve, then $\beta(t) = \alpha(at + b)$ for some real constants a, b.

Remarks.(1) If $\alpha : J \to S$ is a geodesic curve and an endpoint b of J belongs to J, then α can be extended to a geodesic parametrisation defined on an interval $J \cup (b - \varepsilon, b + \varepsilon)$ for some $\varepsilon > 0$, and the extension is unique.

(2) If α, $\beta : J \to S$ are geodesic curves which agree on a non-degenerate subinterval I of J, then $\alpha = \beta$.

To see (1), let (U, ϕ) be a chart with $\alpha(b) \in U$. Then there is an interval $[b, b \pm \varepsilon)$, where $\varepsilon > 0$, on which $\phi \circ \alpha$ is the restriction of a geodesic parametrisation of a hyperbolic line. Call this parametrisation of the line β. Then defining $\alpha(t) = \phi^{-1}\beta(t)$ for all $t \in (b - \varepsilon, b + \varepsilon)$ gives the required extension. If two geodesic parametrisations of a hyperbolic line agree on a non-degenerate interval, they are the same, hence the extension is unique.

To see (2), consider subintervals I' of J containing I such that α and β agree on I'. Let I_0 be the union of all such subintervals. Then I_0 is a maximal subinterval of J on which α, β agree. If I_0 is not equal to J, one of its endpoints is not in I_0, by (1), but the set of points at which α, β agree is closed, a contradiction. Hence $I_0 = J$, as required.

Suppose α is a geodesic curve defined on a non-degenerate interval J, and β, γ are two geodesic curves defined on intervals J', J'' respectively, both containing J, and that β and γ both extend α. Then by Remark (2) $\beta(t) = \gamma(t)$ on $J' \cap J''$. Therefore, there is a unique maximal interval I containing J such that α can be extended to a geodesic curve on I, namely the union of all intervals I' containing J such that α can be extended to I', and the extension to I is unique.

Definition. A geodesic curve on a hyperbolic surface S is called complete if it is defined on $\mathbb{R}$. The surface S is called complete if it is complete as a metric space using the metric defined above.

Lemma 5.1. *If a hyperbolic surface S is complete, then every geodesic curve can be extended to a complete geodesic curve.*

Proof. Denote the metric on S by d. Let $\alpha : J \to S$ be a geodesic curve which cannot be extended to a larger interval. By Remark (1) above, $J = (a, b)$ for some a, $b \in \mathbb{R} \cup \{\pm\infty\}$, and we assume $a < b$. Suppose $b \in \mathbb{R}$. Let (t_n) be a strictly increasing sequence of points in J converging to b. Then (t_n) is a Cauchy sequence. Now arclength is a linear function of the parameter, so the length of α restricted to the closed interval with endpoints t_m, t_n, is $k|t_m - t_n|$ for some real constant k. By definition of d, $d(\alpha(t_m), \alpha(t_n)) \leq k|t_m - t_n|$, hence $(\alpha(t_n))$ is a Cauchy sequence, so converges to some point p in S. Put $p_n = \alpha(t_n)$, and choose a chart (U, ϕ) with $p \in U$. Then p_n is in U for all $n \geq N$, for some suitably chosen N. Now $\phi \circ \alpha$ restricted to $[t_N, t_n]$ is the restriction of a geodesic parametrisation of a hyperbolic line, for $n > N$. By remark (2) above, this parametrisation is the same for all n. Call this parametrisation β. Define $\alpha(b) = p$. Then $\alpha = \phi^{-1} \circ \beta$ in a neighbourhood of b in $(a, b]$, so this extends α to a geodesic curve defined on $(a, b]$, a contradiction, hence $b = \infty$. A similar argument shows $a = -\infty$. □

We shall need a converse of Lemma 5.1, and the next two lemmas are intended for this purpose.

Lemma 5.2. *Let S be a hyperbolic surface. If a curve α from x to y has $l(\alpha) = d(x, y)$, then it is a geodesic curve when parametrised by arclength.*

Proof. Suppose $\alpha : [a, b] \to S$ with $\alpha(a) = x$, $\alpha(b) = y$. Let $a \leq c \leq e \leq b$ and let $z = \alpha(c)$, $w = \alpha(e)$. It follows easily from the triangle inequality and additivity of length that $l(\alpha|_{[c,e]}) = d(z, w)$. Choose $t \in (a, b)$ and let (U, ϕ) be a chart with $\alpha(t) \in U$. If $t_1 \leq t \leq t_2$ and $\alpha(t_1)$, $\alpha(t_2) \in U$, then by the observation just made and an exercise above, $\rho(\phi\alpha(t_1), \phi\alpha(t_2)) = d(\alpha(t_1), \alpha(t_2)) = l(\alpha|_{[t_1,t_2]}) = l(\phi \circ \alpha|_{[t_1,t_2]})$. It follows that $\phi \circ \alpha$ restricted to a suitable interval around t is a parametrisation of a hyperbolic line segment (see Theorem 7.3.1 in [9]). If α is a parametrisation by arclength of the curve, measured from some point, then $\phi \circ \alpha$ is a parametrisation of the hyperbolic line segment by arclength, which is a geodesic parametrisation. Hence α is a geodesic curve. □

Lemma 5.3. *Let S be a hyperbolic surface with metric d. Suppose every geodesic curve in S can be extended to a complete geodesic curve. Then every two points x, y in S can be joined by a geodesic curve whose length is $d(x, y)$.*

Proof. Choose a chart (U, ϕ) such that $x \in U$, and choose a subset V of U such that $\phi(V)$ is a hyperbolic circle centred at $\phi(x)$, with radius r, say, such that $\phi(y) \notin \phi(V)$. Then V is compact, so there exists $z_0 \in V$ such that $d(z_0, y) = \inf_{z \in V} d(z, y)$. Note that $d(x, z_0) = r$ since ϕ is an isometry by an earlier exercise. It follows that $d(x, y) = d(x, z_0) + d(z_0, y)$. For otherwise, there is a curve α joining x to y with $l(\alpha) < d(x, z_0) + d(z_0, y) = r + d(z_0, y)$. It must intersect V at some point z'. Suppose $\alpha(t_0) = x$, $\alpha(t_1) = z'$ and $\alpha(t_2) = y$. Then

$$l(\alpha) = l(\alpha|_{[t_0,t_1]}) + l(\alpha|_{[t_1,t_2]}) \geq d(x, z') + d(z', y) = r + d(z', y),$$

hence $d(z', y) < d(z_0, y)$, contradicting the choice of z_0. Now there is a geodesic curve joining x and z_0, of length $d(x, z_0) = r$, obtained by taking the inverse image under ϕ of a hyperbolic line segment, suitably parametrised. Let β be the extension of this geodesic curve to a complete geodesic curve, parametrised by arclength, in such a way that $\beta(0) = x$ and $y = \beta(s)$ for some $s > 0$. Put $b = d(x,y)$. Consider real values s of arclength satisfying $0 \leq s \leq b$, $d(x,\beta(s)) = l(\beta|_{[0,s]}) = s$ and $d(x,\beta(s)) + d(\beta(s),y) = d(x,y)$. Note that $\beta(r) = z_0$ and $s = r$ is one such value. It is left as an exercise to show that the set of all such s forms a closed interval $[0,c]$ with $c \leq b$. It suffices to show that $c = b$, for then $d(x,\beta(b)) = b = d(x,y)$, so $d(\beta(b),y) = 0$, i.e. $\beta(b) = y$, as required. Since $r > 0$ and $r \in [0,c]$, the interval $[0,c]$ is non-degenerate.

Suppose on the contrary that $c < b$, and let $z_1 = \beta(c)$. Repeating the argument above, with z_1 in place of x, we obtain $w \in S$, with $w \in U'$ for some chart (U',ϕ') containing z_1, such that $\phi'(w)$ lies on a hyperbolic circle, contained in $\phi'(U')$ and centred at $\phi'(z_1)$, and such that $d(z_1,w) + d(w,y) = d(z_1,y)$. Then

$$d(x,y) = d(x,z_1) + d(z_1,y) = d(x,z_1) + d(z_1,w) + d(w,y)$$

and it follows by use of the triangle inequality that $d(x,z_1) + d(z_1,w) = d(x,w)$. Again, there is a geodesic curve β' joining z_1 and w, of length $d(z_1,w)$. Concatenating this with β restricted to $[0,c]$ gives a curve of length $d(x,z_1) + d(z_1,w) = d(x,w)$. By the previous lemma, this curve, when parametrised by arclength, is a geodesic curve. It extends $\beta|_{[0,c]}$, so by Remark (2) above, must be $\beta|_{[0,d]}$, where $d = d(x,w)$ is the length of $\beta|_{[0,d]}$, and $d > c$. This contradicts the choice of c, proving the lemma. □

Before proving the converse of Lemma 5.1, we need to introduce the exponential map. We assume all geodesics on S extend to complete geodesics. Choose $x \in S$ and a chart (U,ϕ) with $\phi(x) = 0$ (in the disk model). Parametrise hyperbolic line segments from 0 by arclength. If $\alpha : [0,\infty) \to \mathbb{H}^2$ is such a parametrisation, then $\phi^{-1} \circ \alpha$ is defined on some interval $[0,a]$ with $a > 0$ and is a geodesic curve. It extends uniquely to a complete geodesic curve β defined on $\mathbb{R}$, which is parametrised by arclength measured from x. We then define $e_x(\alpha(s)) = \beta(s)$ for $0 \leq s < \infty$. This defines a mapping $e_x : \mathbb{H}^2 \to S$, called the exponential map (the reason for this is too long a story to give here). The notation suppresses dependence on the choice of (U,ϕ), but this will not cause problems. We shall show e_x is continuous.

With α, β as above and $s \geq 0$, let $y = \beta(s)$. As noted in the definition of length of paths, we can choose a partition $0 = s_0 < s_1 < \ldots < s_n = s$ of $[0,s]$ and charts (U_i,ϕ_i) for $1 \leq i \leq n$ such that $\beta([s_{i-1},s_i]) \subseteq U_i$ and $\phi_i(U_i)$ is a hyperbolic disk. We can also assume $(U_1,\phi_1) = (U,\phi)$. Let V_i be the component of $U_i \cap U_{i+1}$ containing $\beta(s_i)$, for $1 \leq i < n$. Then $\phi_i = g_i \circ \phi_{i+1}$ on V_i for a unique isometry g_i of $\mathbb{H}^2$, for $1 \leq i < n$. Now replace ϕ_2 by $g_1 \circ \phi_2$. Successively, we can assume $\phi_i = \phi_{i+1}$ on V_i, for $1 \leq i < n$. It then follows by induction on i that on $[s_{i-1},s_i]$, $\beta = \phi_i^{-1} \circ \alpha$, using the uniqueness of extensions of geodesic curves noted above.

Let $\tilde{y} = \phi_n(y) = \alpha(s)$, and let $\tilde{z} \in \mathbb{H}^2$ be sufficiently close to $\tilde{y}$ to ensure that $\phi_n^{-1}(\tilde{z}) \in U_n$ and $\rho(\tilde{z}, 0) \geq \rho(\alpha(s_{n-1}), 0) = s_{n-1}$. Let $\alpha' : [0, \infty) \to \mathbb{H}^2$ be the parametrisation by arclength of the hyperbolic line segment from 0 passing through $\tilde{z}$, and let $\tilde{z} = \alpha'(\tilde{s})$. Let $\varepsilon > 0$. By making $\rho(\tilde{y}, \tilde{z})$ sufficiently small, we can make the angle between the segments parametrised by α and α' as small as we like, and can therefore make $\rho(\alpha(s_i), \alpha'(s_i)) < \varepsilon$ for $0 \leq i \leq n-1$. This follows from the hyperbolic cosine rule (version I in §7.12 of [9]) and is left as an exercise. Choose ε so that the open disk of radius ε in S, centred at $\beta(s_i) = \phi_i^{-1}\alpha(s_i)$ is contained in V_i for $1 \leq i < n$. Then $\phi_i^{-1}\alpha'(s_i) \in V_i$ (recall from an earlier exercise that ϕ_i is an isometry). Since $\phi_i(U_i)$ is a disk, so a convex hyperbolic set, it follows that $\phi_i^{-1}\alpha'([s_{i-1}, s_i]) \subseteq U_i$ for $1 \leq i < n$, and $\phi_n^{-1}\alpha'([s_{n-1}, \tilde{s}]) \subseteq U_n$.

Let β' be the extension of $\phi^{-1} \circ \alpha'$ to a complete geodesic curve. It again follows by induction that $\beta' = \phi_i^{-1} \circ \alpha'$ on $[s_{i-1}, s_i]$ for $1 \leq i < n$, and $\beta' = \phi_n^{-1} \circ \alpha'$ on $[s_{n-1}, \tilde{s}]$. Thus

$$e_x(\tilde{z}) = \beta'(\tilde{s}) = \phi_n^{-1}\alpha'(\tilde{s}) = \phi_n^{-1}(\tilde{z}).$$

This is true for all $\tilde{z}$ in a neighbourhood of $\tilde{y}$, and the continuity of e_x follows.

Note that, by definition of e_x, if β is any geodesic curve with $\beta(0) = x$, parametrised by arclength, then for $0 \leq s < \infty$ such that $\beta(s)$ is defined, $\beta(s) = e_x(z)$ for some $z \in \mathbb{H}^2$ at hyperbolic distance s from 0.

We now come to the promised converse of Lemma 5.1. It is a version in the present context of a theorem of Riemannian geometry, the Hopf-Rinow Theorem.

Theorem 5.4. *A hyperbolic surface S is complete if and only if every geodesic curve can be extended to a complete geodesic curve.*

Proof. Assume all geodesic curves can be extended to complete geodesic curves. Let C be a closed bounded set in S. We claim C is compact. Choose $x \in C$ and put $a = \sup_{y \in C} d(x, y)$, so $a \in \mathbb{R}$. For any $y \in C$, there is a geodesic curve from x to y of length $d(x, y)$, by Lemma 5.3. Parametrise this curve by arclength, so it is defined on $[0, d(x, y)]$. Then from an observation preceding the theorem, $y = e_x(z)$ for some $z \in \mathbb{H}^2$ at hyperbolic distance $d(x, y)$ from 0. Thus, if $B = \{z \in \mathbb{H}^2 \mid \rho(0, z) \leq a\}$, where ρ is the hyperbolic metric, then $C \subseteq e_x(B)$. Now B is compact, hence $e_x(B)$ is compact, and C is a closed subset, so C is compact, as claimed.

The proof can now be finished with arguments from elementary analysis. Let (x_n) be a Cauchy sequence in S. Then it is bounded, so there is some closed disk C containing all x_n. Now C is compact, so sequentially compact, hence (x_n) has a convergent subsequence, converging to x, say. Since the sequence is Cauchy, the entire sequence converges to x. Hence S is complete. The converse has already been established. □

This has a useful consequence. Let $p : \tilde{S} \to S$ be a covering map. If α is a curve in $\tilde{S}$, then it is easy to see that α is a geodesic curve in $\tilde{S}$ if and only if $p \circ \alpha$ is a geodesic curve in S. Clearly, α is complete if and only if $p \circ \alpha$ is complete. On the other hand, if β is a curve $J \to S$, since J is simply connected, from the standard theory of covering spaces there is a curve $\alpha : J \to \tilde{S}$ such that $\beta = p \circ \alpha$. This gives an immediate corollary to the previous theorem.

Corollary 5.5. *Let S be a hyperbolic surface and let $p : \tilde{S} \to S$ be a covering map. Then S is complete if and only if $\tilde{S}$ is complete.*

□

The next result is again a version, in the present context, of a theorem of Riemannian geometry, the Cartan-Hadamard Theorem.

Theorem 5.6. *A complete, simply connected hyperbolic surface S is metrically isomorphic to $\mathbb{H}^2$.*

Proof. Choose $x \in S$ and a chart (U, ϕ) such that $\phi(x) = 0$ and $\phi(U)$ is a disk. We define a map $D : S \to \mathbb{H}^2$ (a version of the "developing map"). Let $y \in S$. Using previous arguments, we can choose a path $\alpha : [a, b] \to S$ joining x to y, a suitable partition $t_0 < t_1 \ldots < t_n$ of $[a, b]$ and charts (U_i, ϕ_i) for $1 \leq i \leq n$ such that $\alpha([t_{i-1}, t_i]) \subseteq U_i$ and $\phi_i(U_i)$ is a hyperbolic disk. We can also assume that $(U_1, \phi_1) = (U, \phi)$ and $\phi_i = \phi_{i+1}$ on V_i where V_i is the component of $U_i \cap U_{i+1}$ containing $\alpha(t_i)$, for $1 \leq i < n$.

We put $D(y) = \phi_n(y)$. As with the definition of length of curves, by taking refinements of partitions we see that this depends only on α and (U, ϕ), not on the choice of partition or charts (U_i, ϕ_i) for $i \geq 2$. To show D is well-defined, and depends only on x and (U, ϕ), we need to show $D(y)$ is independent of the choice of α.

If β is another path from x to y which is homotopic to α, via a homotopy F (so $F(t, 0) = \alpha(t)$, $F(t, 1) = \beta(t)$, $F(0, s) = x$, $F(1, s) = y$ for all s, $t \in [0, 1]$), then the sets $F^{-1}(U')$, where (U', ϕ') is a chart such that $\phi'(U')$ is a disk, form an open covering of $[0, 1] \times [0, 1]$. Take a Lebesgue number δ for this covering, and take partitions $t_0 < t_1 \ldots < t_m$ and $s_0 < s_1 \ldots < s_k$ of $[0, 1]$ such that the rectangles $R_{ij} = [t_{i-1}, t_i] \times [s_{j-1}, s_j]$ all have diameter less than δ, so that $F(R_{ij}) \subseteq U_{ij}$ for some chart (U_{ij}, ϕ_{ij}) such that $\phi_{ij}(U_{ij})$ is a disk. Let α_j be the path $t \mapsto F(t, s_j)$.

Since F is uniformly continuous, by refining the partitions we can assume $F(R_{1j}) \subseteq U$ for $1 \leq j \leq k$, and so we can assume $(U_{1j}, \phi_{1j}) = (U, \phi)$ for $1 \leq j \leq k$. Then, computing $D(y)$ using the charts (U, ϕ), $(U_{2j}, \phi_{2j}), \ldots, (U_{mj}, \phi_{mj})$, after modifying the chart maps to satisfy the appropriate conditions, we see that replacing $\alpha = \alpha_0$ by α_1 does not change the value of $D(y)$, and successively, $D(y)$ is unchanged using $\alpha_2, \ldots, \alpha_k = \beta$ instead of α. (Note that $\alpha_j(t_i)$ and $\alpha_{j-1}(t_i)$ are in the same component of $U_{ij} \cap U_{i+1,j}$, because $s \mapsto F(t_i, s)$

for $s_{j-1} \leq s \leq s_j$ provides a path joining them, where $1 \leq i < m$, $1 \leq j \leq k$.) Since S is simply connected, it follows that D is well-defined.

Note that, in the definition of D, $D(z) = \phi_n(z)$ for all $z \in U_n$, since U_n is path connected, hence D is continuous. Further, D restricted to U_n is an isometry, by an earlier exercise. Thus, if γ is a path in S, we can find a partition $r_0 < \ldots < r_p$ of its domain and charts (W_i, ψ_i) such that $\gamma([r_{i-1}, r_i]) \subseteq W_i$ and D restricted to W_i is an isometry, for $1 \leq i \leq p$. It follows that $l(\gamma)$ is defined if and only if $l(D \circ \gamma)$ is defined, in which case $l(D \circ \gamma) = l(\gamma)$.

If we define the exponential map e_x using the chart (U, ϕ), then from the proof that e_x is continuous, $D \circ e_x = id_{\mathbb{H}^2}$. It follows that $e_x \circ D$ is a retraction of S onto the subset $e_x(\mathbb{H}^2)$ of S. Since S is Hausdorff, $e_x(\mathbb{H}^2)$ is closed (an exercise in point-set topology). But $e_x(\mathbb{H}^2)$ is open by Invariance of Domain (see Corollary 19.9, Chapter IV, in [22]), and S is connected, so D and e_x are inverse homeomorphisms. Since D preserves length of paths, it is an isometry. $\square$

Thus if S is a complete hyperbolic surface, we can identify the universal covering space with $\mathbb{H}^2$. As already noted, the group G of covering transformations acts as isometries on $\mathbb{H}^2$. From the general theory of covering spaces, the action of G is free and properly discontinuous (see the definitions preceding Lemma 3.3.6). In particular, the action is faithful, so G can be viewed as a subgroup of $\mathrm{PGL}_2(\mathbb{R}) = \mathrm{Isom}(\mathbb{H}^2)$. A subgroup of $\mathrm{PGL}_2(\mathbb{R})$ which acts properly discontinuously on $\mathbb{H}^2$ is called a *non-Euclidean crystallographic group*, abbreviated to NEC group. An NEC group consisting entirely of orientation-preserving isometries is called a *Fuchsian group*. Any NEC group G has a subgroup of index at most 2 which is a Fuchsian group, namely $G \cap \mathrm{Isom}^+(\mathbb{H}^2)$.

A subgroup of $\mathrm{PGL}_2(\mathbb{R})$ is an NEC group if and only if it is a discrete subgroup. To see this, we make some observations.

(1) If K is a compact subset of $\mathbb{H}^2$ and $z \in \mathbb{H}^2$, the set

$$E = \{g \in \mathrm{PGL}_2(\mathbb{R}) \mid gz \in K\}$$

is compact.

To prove (1), let $q : \mathrm{GL}_2(\mathbb{R}) \to \mathrm{PGL}_2(\mathbb{R})$ be the canonical projection, and let $E_z = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}) \;\middle|\; \frac{az+b}{cz+d} \in K \right\}$. Then E_z is compact ([82; Lemma 5.6.1]), and the mapping $\phi : \mathrm{GL}_2(\mathbb{R}) \to \mathrm{GL}_2(\mathbb{R})$, $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} -a & b \\ -c & d \end{pmatrix}$ is a homeomorphism. Since $E = q(E_z \cup \phi(E_{-\bar{z}}))$, E is compact.

(2) If G is an NEC group, then for any $z \in \mathbb{H}^2$, the stabiliser $\mathrm{stab}_G(z)$ is finite.

For G contains a Fuchsian subgroup H of index at most 2, which is discrete by [82; Theorem 5.6.3], and $\text{stab}_H(z)$ is finite cyclic (see the proof of Theorem 5.7.2 in [82]).

(3) If G is an NEC group, the set of points in $\mathbb{H}^2$ fixed by no element of $G \setminus \{1\}$ is open and dense in $\mathbb{H}^2$.

To see this, let $z \in \mathbb{H}^2$ and choose an open disk neighbourhood U of z such that for $g \in G$, $U \cap gU \neq \emptyset$ implies $gz = z$. If $g \in G$ fixes a point in U, then $g \in \text{stab}_G(z)$, and there are finitely many such g by (2). For each such g other than 1, the set of fixed points of g in U is either the single point z, or the intersection of U and a hyperbolic line passing through z, and (3) follows.

Using (1) and (3), the proof of Theorem 5.6.3(i) in [82] can now be easily modified to show that a subgroup of $\text{PGL}_2(\mathbb{R})$ is an NEC group if and only if it is discrete. In particular, the following observation is a consequence of (1).

(4) If G is an NEC group, K is a compact subset of $\mathbb{H}^2$ and $z \in \mathbb{H}^2$, the set $\{g \in G \mid gz \in K\}$ is finite.

Further, the proof of Corollary 5.6.4 in [82] works to show that a subgroup G of $\text{PGL}_2(\mathbb{R})$ is an NEC group if and only if, for all $z \in \mathbb{H}^2$, the orbit Gz is a discrete subspace of $\mathbb{H}^2$.

By Theorem 5.7.3 in [82], an elliptic element in an NEC group has finite order. Since the orientation-reversing isometries of $\mathbb{H}^2$ having a fixed point are precisely those having finite order (the hyperbolic reflections), it follows that an NEC group acts freely on $\mathbb{H}^2$ if and only if it is torsion-free.

From the general theory of covering spaces, if S is a complete hyperbolic surface, then $S = \mathbb{H}^2/G$, where G is the group of covering transformations (a torsion-free NEC group) and G is isomorphic to $\pi_1(S)$.

Conversely, given a torsion-free NEC group G, G acts freely and properly discontinuously on $\mathbb{H}^2$, so the projection map $p : \mathbb{H}^2 \to \mathbb{H}^2/G$ is a covering map. Moreover, $\mathbb{H}^2/G$ can be made into a hyperbolic surface as follows. For each $z \in \mathbb{H}^2$, we can choose an open disk N_z such that $N_z \cap gN_z = \emptyset$ for $1 \neq g \in G$. Let $p : \mathbb{H}^2 \to \mathbb{H}^2/G$ be the projection map. We take as charts all pairs $(p(N_z), p_z^{-1})$ as z ranges over $\mathbb{H}^2$, where p_z is the restriction of p to N_z. Since p is an open map and injective when restricted to N_z, p_z^{-1} is a homeomorphism onto N_z. Since p is surjective these charts cover S, hence S is a surface. Consider a transition map $p_u^{-1} \circ p_v : p_v^{-1}(p(N_u) \cap p(N_v)) \to p_u^{-1}(p(N_u) \cap p(N_v))$. If $z \in p_v^{-1}(p(N_u) \cap p(N_v))$, $p(z) = p(w)$ for some unique $w \in N_u$, and so $w = gz$ for some $g \in G$. The maps p_v and $p_u \circ g$ agree on $N_v \cap g^{-1}N_u$ since they both agree with p on this set, hence $p_u^{-1} \circ p_v$ agrees with g on $N_v \cap g^{-1}N_u$, which is an open neighbourhood of z in $p_v^{-1}(p(N_u) \cap p(N_v))$. Thus $p_u^{-1} \circ p_v$ restricted to $N_v \cap g^{-1}N_u$ is the restriction of a hyperbolic isometry. It follows that the transition map $p_u^{-1} \circ p_v$ belongs to the pseudogroup P in the definition of hyperbolic structure, and we have defined a hyperbolic structure on S. With this structure it is complete, by Corollary 5.5.

If G is a Fuchsian group, the transition maps $p_u^{-1} \circ p_v$ are restrictions of orientation-preserving isometries, so as noted earlier, the hyperbolic structure makes S an oriented surface.

Note that an NEC group G is countable, by (4) above, because if we fix $z \in \mathbb{H}^2$, the closed hyperbolic disk of radius n centred at z contains gz for only finitely many $g \in G$, for $n = 1, 2, \ldots$, and these disks cover $\mathbb{H}^2$.

If G is an NEC group, a *fundamental domain* for G is defined to be a closed subset F of $\mathbb{H}^2$ such that $GF = \mathbb{H}^2$ and $g\mathring{F} \cap \mathring{F} = \emptyset$ for $g \in G$, $g \neq 1$, where $\mathring{F}$ means the interior of F. This is an important idea in the theory of NEC groups. To construct one, choose $x \in \mathbb{H}^2$ such that x is not fixed by any element of $G \setminus \{1\}$; this is always possible by observation (3) above. Put $H_g = \{y \in \mathbb{H}^2 \mid \rho(x,y) \leq \rho(x,gy)\}$, where ρ is the hyperbolic metric and $g \in G$. Define $F = \bigcap_{g \in G} H_g$. Then F is a convex fundamental domain for G (see [82; Theorem 5.8.3] or [83; Theorem 3.2.2]; the argument works for any NEC group). We call F the *Dirichlet region* for G corresponding to x (because the analogue for the Euclidean plane was considered by Dirichlet).

We investigate the boundary of the Dirichlet region F. For $g \in G \setminus \{1\}$, let $L_g = \{y \in \mathbb{H}^2 \mid \rho(x,y) = \rho(x,gy)\}$, the bisector of x and $g^{-1}x$. Then L_g is a hyperbolic line, and $\rho(x, L_g) = \frac{1}{2}\rho(x,gx)$ (see Theorem 7.2.1, (7.20.4) and §7.21 in [9]). Suppose $z \in F$ and let D be a closed hyperbolic disk centred at z, of radius r, say. Then D meets only finitely many of the lines L_g, for $g \in G \setminus \{1\}$. For otherwise, $\rho(z,gx) \leq \rho(z,x) + \rho(x,gx) \leq \rho(z,x) + 2r$ for infinitely many $g \in G$. That is, gx is in the closed disk of radius $\rho(z,x) + 2r$ centred at z for infinitely many g, contradicting observation (4) above. Hence either $z \in L_g$ for some $g \in G \setminus \{1\}$, or there is a hyperbolic disk $D' \subseteq D$, centre z, such that D' does not meet any L_g. In the second case, suppose $g \in G \setminus \{1\}$. Since D' is connected, the image of the function $D' \to \mathbb{R}$, $y \mapsto \rho(x,y) - \rho(x,gy)$ is an interval, and does not contain 0. Since $z \in H_g$, $D' \subseteq H_g$. It follows that $D' \subseteq F$ and z is an interior point of F. Thus if $z \in \partial F$ (the boundary of F in $\mathbb{H}^2$), then $z \in L_g$ for some $g \in G \setminus \{1\}$, so $\partial F \subseteq \bigcup_{g \in G \setminus \{1\}} L_g$, a countable union of hyperbolic lines. It follows that the boundary of F has zero hyperbolic area.

Lemma 5.7. *If F_1, F_2 are fundamental domains for the NEC group G and the boundaries of F_1, F_2 have zero hyperbolic area, then $\mu(F_1) = \mu(F_2)$, where μ denotes hyperbolic area.*

Proof. Note that μ is a measure in the sense of §4, provided the Lebesgue integral is used to define it ([9;§7.2]). Also, for all $g \in \mathrm{Isom}(\mathbb{H}^2)$ and measurable sets E, $\mu(gE) = \mu(E)$. This is proved for orientation-preserving g in [82; Theorem 5.3.1], and it suffices to prove it when g is $z \mapsto -\bar{z}$, since $\mathrm{Isom}(\mathbb{H}^2)$

is clearly generated by this isometry and the orientation-preserving isometries. This is left as a very easy exercise.

Let G_i be the interior of F_i. Since F_i is closed, it is measurable with respect to μ, and $\mu(F_i) = \mu(G_i)$, for $i = 1, 2$. Now $\bigcup_{g\in G}(F_1 \cap gG_2) \subseteq F_1$, and since $gG_2 \cap hG_2 = \emptyset$ for $g \neq h$, g, $h \in G$, this union is a disjoint union. Using this and the fact that μ is G-invariant, we have

$$\begin{aligned}\mu(F_1) \geq \mu\Big(\bigcup_{g\in G}(F_1 \cap gG_2)\Big) &= \sum_{g\in G}\mu(F_1 \cap gG_2)\\ &= \sum_{g\in G}\mu(g^{-1}F_1 \cap G_2)\\ &= \sum_{g\in G}\mu(gF_1 \cap G_2)\\ &\geq \mu\Big(\bigcup_{g\in G}(gF_1 \cap G_2)\Big) = \mu(G_2) = \mu(F_2)\end{aligned}$$

because as g^{-1} ranges through G, so does g, and $GF_1 = \mathbb{H}^2$. Thus $\mu(F_1) \geq \mu(F_2)$, and a symmetrical argument shows $\mu(F_2) \geq \mu(F_1)$, proving the lemma. □

Let S be a complete hyperbolic surface, so $S = \mathbb{H}^2/G$, where G is an NEC group acting freely on $\mathbb{H}^2$. By Lemma 5.7, we can define the area of S to be $\mu(F)$, where F is any fundamental domain for G whose boundary has zero hyperbolic area. For example, we could take F to be a Dirichlet region for G corresponding to some point x not fixed by any element of $G \setminus \{1\}$. The area can be $+\infty$. For example, take the infinite cyclic group generated by the transformation $z \mapsto z+1$ of the upper half-plane. The Dirichlet region corresponding to i is the vertical strip $\{z \in \mathbb{H}^2 \mid -\frac{1}{2} \leq \Re(z) \leq \frac{1}{2}\}$, and has infinite area. This is left as an exercise.

We shall be interested in complete hyperbolic surfaces of finite area, and to study these, we need to look more closely at the boundary of Dirichlet regions, and consider Euler characteristics.

Let G be an NEC group and F be a Dirichlet region for G corresponding to a point x not fixed by any element of $G \setminus \{1\}$. One reason for this assumption on x is that it ensures F is *locally finite*, that is, every compact subset of $\mathbb{H}^2$ meets the set gF for only finitely many $g \in G$. See [82; Theorem 5.8.5]. The argument is valid for any NEC group, in view of observation (4) above. Another consequence of this assumption is that x is not on any of the lines L_g, for $g \neq 1$, so x is in the interior of F, hence F has non-empty interior. Also, $L_g = L_h$ implies $g = h$.

Let ∂F be the boundary of F in $\mathbb{H}^2$. Since F is a fundamental domain, any set $F \cap gF$ with $g \in G$, $g \neq 1$ is contained in ∂F (and in $\partial(gF)$), and

$F \cap gF$ is convex, being the intersection of two convex sets. Further, it cannot contain three non-collinear points, otherwise it would contain a non-degenerate triangle, contradicting the fact that the boundary of F has zero area. Thus $F \cap gF$ could be empty or a single point, but is otherwise a segment (not necesssarily compact) of a hyperbolic line of positive length; in this case we call $F \cap gF$ a *side* of F.

In fact, ∂F is the union of the sides of F. For suppose $z \in \partial F$; we need to find a side containing z. Now $z \in F$, and by taking disks of decreasing radius centred at z, we can find a sequence of points (z_n) in $\mathbb{H}^2 \setminus F$ converging to z. Then since the sets gF, for $g \in G$, cover $\mathbb{H}^2$, $z_n \in g_nF$ for some $g_n \neq 1$, and since F is locally finite, there exists $g \in G \setminus \{1\}$ such that $z_n \in gF$ for infinitely many n. Since gF is closed, $z \in gF$. Now let K be a disk centred at z. Then since $z \in F$ and F has more than one point, (it has non-empty interior), there is a hyperbolic line segment from z to another point of F, which lies in F since F is convex. Hence K contains a point of F and a point not in F, neither of which is z. Thus, there is a path α in $K \setminus \{z\}$, from a point in F to a point not in F and this path must contain a point of ∂F ($\alpha^{-1}(F)$ is compact, so has a largest element t, and $\alpha(t) \in \partial F$). Hence, by taking disks centred at z of decreasing radius, we can find a sequence (w_n) of points in ∂F, so in F, not equal to z, converging to z. As with z, we can find $g_n \neq 1$ such that $w_n \in g_nF$, and again by local finiteness, there is a $g \in G \setminus \{1\}$ such that, for infinitely many n, $w_n \in F \cap gF$, a closed set. It follows that $F \cap gF$ is a side and $z \in F \cap gF$, as required. It is easy to see that, for any $g \in G \setminus \{1\}$, $F \cap gF \subseteq L_g$, so the sides are segments of the lines L_g.

Suppose $F \cap gF$ is a side and $h \in G$, $h \neq 1$, $h \neq g$. Then no interior point of the segment $F \cap gF$ belongs to hF. For if x is such a point, choose points y, $z \in F \cap gF$ so that $y \neq x \neq z$ and x is in the line segment joining them, and choose $u \in \mathring{F}$, $v \in g\mathring{F}$, $w \in h\mathring{F}$. Then the non-degenerate triangles uyz, vyz are contained, respectively, in F, gF, and the line segment joining w and x lies in hF and intersects the interior of either uyz or vyz, contradicting that F is a fundamental domain for G. It follows that $F \cap gF \cap hF$ is either empty or consists of a single point. For it is a convex subset of $F \cap gF$, so if it contains more than one point, it is a hyperbolic line segment, hence contains an interior point of $F \cap gF$, a contradiction. If $F \cap gF \cap hF$ consists of a single point, that point is called a *vertex* of F; it is in ∂F, so belongs to at least one side. Further, if two sides $F \cap gF$, $F \cap hF$ meet, they do so in a single point, which is a vertex and an endpoint of both sides.

If a side $F \cap gF$ has an endpoint, say v, in $\mathbb{H}^2$, it is an endpoint of at least two sides (so is a vertex). For let K be a closed disk centred at v and put $K' = K \setminus (F \cap gF)$. Since $F \cap gF$ is a line segment and v is an endpoint, K' is path connected, and arguing as above, we can find a point of ∂F in K', hence a sequence of points z_n in ∂F, not in $F \cap gF$, converging to v, and $h \in G$ such that $z_n \in F \cap hF$ for infinitely many n. Then $F \cap hF$ is a side containing v and not equal to $F \cap gF$.

Suppose a vertex lies on at least three sides. Choose a point on each side so that the three chosen points are not collinear. Then the triangle obtained by joining these three points is contained in F. The triangle contains points of one of the sides in its interior, contradicting that the sides are contained in ∂F. Hence each vertex is an endpoint of exactly two sides. Similarly, a point on the circle at infinity S^1 cannot be an endpoint of more than two sides of F.

Note that, since G is countable, there are only countably many vertices and sides.

Let G^* be the set of $g \in G$ such that $F \cap gF$ is a side of F. Let Σ be the set of sides of F. The map $G^* \to \Sigma$, $g \mapsto F \cap gF$ is clearly surjective, and in fact is bijective. For if $F \cap gF$, $F \cap hF$ are sides and $F \cap gF = F \cap hF$, then $F \cap gF \cap hF$ has more than one point, so by the discussion above $h = g$.

Further, if $s = F \cap gF$ is a side, so is $g^{-1}s = F \cap g^{-1}F$, and $s \mapsto g^{-1}s$ is a pairing of the sides, that is, a mapping $\Sigma \to \Sigma$ whose square is the identity map. It can happen that $s = g^{-1}s$, but then, by considering the effect of g on the hyperbolic line containing s, using the fact that $\mathring{F} \cap g\mathring{F} = \emptyset$, g is of order 2 and has a fixed point on s (this is left as an exercise). Moreover, if s is a side and hs is also a side, where $h \in G$, then either $h = 1$, or $s = F \cap h^{-1}F$, so that s and hs are paired by this mapping. This again follows easily from the discussion above.

Now we need to consider Euler characteristics. For the definition of Euler characteristic of a space of "bounded finite type" see, for example, Definition 13.2, Chapter IV in [22]. Let S be a surface whose Euler characteristic $\chi(S)$ is defined, and let $x \in S$. It is well known that

$$\mathrm{H}_n(S, S \setminus \{x\}) = \begin{cases} 0 & \text{if } n \neq 2 \\ \mathbb{Z} & \text{if } n = 2 \end{cases}$$

(see the observations after Corollary 6 in §2, Ch.6 of [134]). Using the long exact homology sequence of the pair $(S, S \setminus \{x\})$, we see that the Euler characteristic of $S \setminus \{x\}$ is defined, and equal to $\chi(S) - 1$, using Theorem 3.14, Chapter 4 in [134] (an exact sequence is a chain complex with zero homology). Also, $S \setminus \{x\}$ is clearly a surface.

Suppose S is a compact surface. Then S has a triangulation (see Theorem 3, Chapter 8 in [100]), which is necessarily finite (see Theorem 8.2, Chapter IV in [22]) and of dimension 2 (see 3.4.4 in [99]). In fact, any decomposition of S as a CW-complex has finitely many cells (see Theorem 8.2, Chapter IV in [22]), and is of dimension 2, by applying invariance of dimension (Corollary 19.10, Chapter IV in [22]) to an open cell of highest dimension, which is an open submanifold. Given a decomposition of S as a CW-complex, let v be the number of vertices (0-cells), e the number of edges (1-cells) and f the number of faces (2-cells). Then $\chi(S)$ is defined and equal to $v - e + f$ (see Theorem 13.3, Chapter IV in [22]). If S' is a surface obtained from S by removing a finite number, n, of points, we see by induction on n that $\chi(S')$ is defined and equal to $v - e + f - n$.

We can now give a result on the structure of complete hyperbolic surfaces of finite area. In the proof we follow Lemma 2.9 in [27].

Lemma 5.8. *A complete hyperbolic surface S of finite area is homeomorphic to a compact surface with a finite number of points removed, and the area $\mu(S)$ is $-2\pi\chi(S)$, where χ denotes Euler characteristic. In particular, $\chi(S) < 0$.*

Proof. Write $S = \mathbb{H}^2/G$, where G is an NEC group acting freely on $\mathbb{H}^2$, let F be a Dirichlet region for G corresponding to a point x which is fixed by no element of $G \setminus \{1\}$, and let $\overline{F}$ be its closure in $\mathbb{H}^2 \cup S^1$. Let $n > 2$ be an integer. If we can find n points on S^1 belonging to $\overline{F}$, then joining them in cyclic order around S^1 with hyperbolic line segments gives an ideal hyperbolic polygon with n vertices, whose area is $(n-2)\pi$ (see [9; Theorem 7.15.1]). This polygon is contained in $\overline{F}$ (in the Klein model [9; §7.1], F is convex as a subset of the Euclidean plane, being an intersection of Euclidean half-planes with the open unit disk (corresponding to the sets H_g in the definition of F), hence $\overline{F}$ is convex, and the polygon is the Euclidean convex hull of its vertices). Hence all but the vertices of the polygon lie in F, so $(n-2)\pi \leq \mu(F)$, where μ is hyperbolic area, and it follows that $\overline{F}$ has only finitely many points in S^1.

We now show that F has only finitely many vertices in $\mathbb{H}^2$. Suppose $v_1, \ldots, v_n$ are vertices. The line segment joining x and v_i meets ∂F only in v_i, otherwise it crosses one of the lines L_g, so contains points not in the half-plane H_g, hence not in F, a contradiction since F is convex. Hence, after renumbering, the triangles xv_iv_{i+1} (indices mod n) fit together to form a polygon P in F with vertices $v_1, \ldots, v_n$.

If v is a vertex of F, let θ_v be the internal angle at v in F (the angle between the sides meeting at v_i), and let α_i be the internal angle at v_i in P, so $\alpha_i \leq \theta_{v_i}$. By [9; Theorem 7.15.1], $\mu(P) = (n-2)\pi - \sum_{i=1}^n \alpha_i$. Hence

$$\sum_{i=1}^{n}(\pi - \theta_{v_i}) \leq \sum_{i=1}^{n}(\pi - \alpha_i) = \mu(P) + 2\pi \leq \mu(F) + 2\pi$$

so the series $\sum_v(\pi - \theta_v)$ (sum over all vertices of F) converges with sum at most $\mu(F) + 2\pi$. Divide the vertices into two subsets:

$$A = \{v \mid \theta_v \leq \tfrac{2\pi}{3}\}, \quad B = \{v \mid \theta_v > \tfrac{2\pi}{3}\}.$$

Since $\sum_v(\pi - \theta_v)$ converges, A is finite.

We call two vertices of F congruent if they are in the same G-orbit. This is an equivalence relation, whose equivalence classes are called cycles. Let C

be a cycle; then $\sum_{v \in C} \theta_v = 2\pi$, by [82; Theorem 5.10.2], since G acts freely (the argument works for any NEC group). Hence C contains no more than two vertices from B. Since $0 < \theta_v < \pi$ (v is the intersection of two distinct sides of F), C contains at least one vertex from A. It follows that there are finitely many cycles, each containing finitely many vertices, so there are finitely many vertices.

From the observations preceding the lemma, there are only finitely many sides, and these are identified in pairs by G (since G acts freely, no side is paired with itself). Hence the number of sides is $2E$ for some integer E. From the discussion in Chapter I, 4D in [89] (which works for NEC groups), since $\overline{F} \cap S^1$ is finite, there are no "free sides" as defined there, so if z is a point of $\overline{F} \cap S^1$, z is an endpoint of exactly two sides.

Thus $\overline{F}$ is homeomorphic to a Euclidean polygon, whose vertices correspond to those of F and the points of $\overline{F} \cap S^1$, which are called proper infinite vertices of F, in the terminology of [9]. We extend the notation and call two vertices (in $\mathbb{H}^2$ or infinite) congruent if they are in the same G-orbit, and call the equivalence classes cycles. Let V be the total number of cycles.

Now S is homeomorphic to F/G by [82; Theorem 5.9.6] (the argument works for any NEC group), and F/G is homeomorphic to $(\overline{F}/G) \setminus I$, where I is the finite set consisting of the cycles of vertices lying on S^1 (it is left as an exercise to check this). Also, from the discussion above, $\overline{F}/G$ is obtained by identifying sides of the polygon $\overline{F}$ in pairs. Hence $\overline{F}/G$ is a compact surface (see Theorem 3.4.6 et seq., Examples 3.4.8 and Theorem 3.4.11 in [99]). Further, this identification gives a decomposition of $\overline{F}/G$ as a CW-complex with V vertices, E edges and 1 face (by a simple generalisation of Example 7.3.3(b) in [99]). Let N be the cardinality of I. Then from the discussion preceding the lemma, $\chi(S) = V - E + 1 - N$. By [9; Theorem 7.15.1], $\mu(F) = (2E-2)\pi - \sum_v \theta_v$, the sum being over all vertices of F. If $v \in S^1$, then $\theta_v = 0$, while if C is a cycle in $\mathbb{H}^2$, then $\sum_{v \in C} \theta_v = 2\pi$, as already noted. Therefore,

$$\begin{aligned} \mu(F) &= (2E-2)\pi - 2\pi(V-N) \\ &= 2\pi(E-1-V+N) = -2\pi\chi(S) \end{aligned}$$

as required. Finally, $\mu(F) > 0$ since F has non-empty interior. □

We use the term *geodesic* to mean the image of a geodesic curve, and *complete geodesic* means the image of a complete geodesic curve. We now come to the main object of study in this section.

Definition. Let S be a complete hyperbolic surface of finite area. A geodesic lamination on S is a codimension 1 lamination all of whose leaves are complete geodesics.

Since the leaves are 1-manifolds, every leaf of a geodesic lamination is homeomorphic to the real line or to a circle. We call such a geodesic a simple geodesic. Let $\mathrm{L}(S)$ denote the set of all closed subsets C of S such that C is a disjoint union of simple geodesics. Thus if $C = |\mathcal{L}|$ is the support of a geodesic lamination, then $C \in \mathrm{L}(S)$. Write $S = \mathbb{H}^2/G$, where G is the group of covering transformations. For any $C \in \mathrm{L}(S)$, the inverse image $\tilde{C}$ of C in $\mathbb{H}^2$ has the following properties.

(i) $\tilde{C}$ is closed in $\mathbb{H}^2$.

(ii) $\tilde{C}$ is a disjoint union of hyperbolic lines, say $\tilde{C} = \bigcup_{i \in I} l_i$, where $l_i \cap l_j = \emptyset$ for $i \neq j$, and the collection $\{l_i \mid i \in I\}$ is invariant under G.

To see (ii), from the comments preceding Corollary 5.5, the lift of a complete geodesic is a union of hyperbolic lines, which is clearly invariant under G. It therefore suffices to show that two distinct hyperbolic lines which project to the same simple geodesic σ in S are disjoint. Suppose on the contrary that two distinct lines l_1, l_2 project onto σ and intersect in a point x. The inverse image $\tilde{\sigma}$ of σ in $\mathbb{H}^2$ is a 1-manifold, so we can find a disk D centred at x such that $D \cap \tilde{\sigma}$ is homeomorphic to a real interval. Further, $D \cap (l_1 \cup l_2)$ is a connected subset (shaped like an "X"), so is itself homeomorphic to an interval. But this is impossible, since removing x from $D \cap (l_1 \cup l_2)$ leaves a space with four components (in fact, $D \cap (l_1 \cup l_2)$ is an $\mathbb{R}$-tree with four directions at x). We now show that every element of $\mathrm{L}(S)$ is the support of a geodesic lamination on S.

Lemma 5.9. *Let S be a complete hyperbolic surface of finite area. If $C \in \mathrm{L}(S)$, then there is a geodesic lamination on S with support C, whose leaves are the geodesics whose disjoint union is C.*

Proof. Let $\tilde{C} = p^{-1}(C)$, where $p : \mathbb{H}^2 \to S$ is the universal covering map. Then $\tilde{C}$ is closed and a disjoint union of hyperbolic lines, no two of which intersect. Let $z \in \tilde{C}$. We first construct a flow box (U, ϕ) for $\tilde{C}$ containing z. To do so it is convenient to use the Klein model of $\mathbb{H}^2$, in which the underlying space is again the interior of the unit disk in $\mathbb{R}^2$, but the lines are the intersections of Euclidean lines with this set. (To see how to get from the Poincaré disk model to this model, see [9; §7.1].) Let l be the line of $\tilde{C}$ which contains z, and let D be an open Euclidean disk centred at z contained in $\mathbb{H}^2$.

If possible, take U to be the interior of a rectangle centred at z with two sides parallel to l and intersecting no other lines of $\tilde{C}$, and let $\phi : U \to I \times I$, where $I = [0, 1]$, be an affine homeomorphism such that $\phi(l \cap U)$ is parallel to the horizontal axis.

Otherwise choose a line l_1 of $\tilde{C}$, disjoint from l, which intersects D. Then choose a line l_2 which intersects D, disjoint from l and on the opposite side of l from l_1, and choose l_2 to be a line of $\tilde{C}$ if possible. Then z does not lie on l_1 or l_2. Let $l_1 \cap \partial D = \{a, b\}$, $l_2 \cap \partial D = \{\mathrm{c}, d\}$, with the notation such that the convex hull of $\{a, b, \mathrm{c}, d\}$ is the quadrilateral $abcd$, whose interior we denote by U. Note that $z \in U$. Let α be the affine homeomorphism from the line segment

da to I sending d to 0 and a to 1, and let β be the affine homeomorphism from the line segment cb to I sending c to 0 and b to 1. There is a homeomorphism $\theta : I \times I \to abcd$, given by $\theta(s,t) = (1-s)\alpha^{-1}(t) + s\beta^{-1}(t)$. It is surjective because $abcd$ is the convex hull of $\{a,b,c,d\}$, and to see it is injective is an exercise in Euclidean geometry.

Suppose m is a line of $\tilde{C}$ intersecting U. Then m meets da in a single point u and cb in a single point v. Now the assignment $u \mapsto v$ defines a map $f : X \to I$, where X is a subset of I, by $f(\alpha(u)) = \beta(v)$. Note that f is strictly increasing, otherwise two distinct lines in $\tilde{C}$ would intersect at a point in U. Also, $X \subseteq (0,1)$ and we put $f(0) = 0$, $f(1) = 1$. Then f remains strictly increasing and is continuous, otherwise two distinct lines in $\tilde{C}$ would intersect somewhere in $\mathbb{H}^2$. Since $\tilde{C}$ is closed, $X \cup \{0,1\} = \alpha(da \cap \tilde{C}) \cup \{0,1\}$ is closed in I, hence its complement is a disjoint union of open intervals. If (q,r) is such an open interval, then q, $r \in X \cup \{0,1\}$, and we extend f linearly to $[q,r]$. This extends f to a continuous strictly increasing bijection from I to I.

We can now define a homeomorphism $\psi : I \times I \to I \times I$ by: $\psi(s,t) = (s, (1-s)t + sf(t))$. Putting $\phi = \psi^{-1} \circ \theta^{-1}$ gives the desired flow box (U, ϕ). Take such a flow box for each $z \in \tilde{C}$ and corresponding disk D.

To define a lamination in S, take $x \in C$, and choose z so that $p(z) = x$. Let V be an evenly covered neighbourhood of x, so there is a neighbourhood W of z such that $p|_W : W \to V$ is a homeomorphism. Choose a disk D centred at z with $D \subseteq W$, and let (U, ϕ) be the flow box constructed above. Then $(p(U), p^{-1} \circ \phi)$ is a flow box for C. Doing this for each $x \in C$ gives a collection $\mathcal{K}$ of flow boxes which cover C, and it is left as an exercise to show this collection defines a lamination on S.

Finally it is clear from the construction that if x, y belong to the same flow box in $\mathcal{K}$, then x and y are in the same local leaf of the flow box if and only if they are on the same geodesic of C. As noted in the discussion of leaves at the start of §4, two points x, $y \in C$ are in the same leaf if and only if there exist finitely many points $x_1, \ldots, x_n$ with $x_1 = x$, $x_n = y$ and x_i, x_{i+1} in the same local leaf of a flow box in $\mathcal{K}$ for $1 \leq i \leq n-1$. The lemma follows. □

Exercise. Let $\mathcal{L}$ be a geodesic lamination on a complete hyperbolic surface of finite area, and let $\tilde{\mathcal{L}}$ be the lift of the lamination to $\mathbb{H}^2$. Show that, if l is a hyperbolic line which is not a leaf of $\tilde{\mathcal{L}}$, with a suitable parametrisation (for example, a geodesic parametrisation), then l is transverse to $\tilde{\mathcal{L}}$, in the sense of §4.

We want eventually to concentrate on $\mathrm{L}(S)$ rather than geodesic laminations, that is, we want to forget about flow boxes, and also give other interpretations of $\mathrm{L}(S)$. The series of lemmas which follow are for this purpose.

The next lemma involves the use of the projective tangent bundle $\mathrm{PT}(S)$ of S. To define this, we first define the unit tangent bundle of S. Choose a chart (U, ϕ) with $x \in U$. Recall that, for $x \in S$, $\mathrm{T}_x(S)$ is in one-to-one correspondence with $\mathbb{R}^2$, which we identify with $\mathbb{C}$, via $[\alpha] \mapsto (\phi \circ \alpha)'(0)$, and we use this to endow $\mathrm{T}_x(S)$ with an inner product, $(u, v) \mapsto u \cdot v$, using the usual Euclidean inner product on $\mathbb{R}^2$. The usual Riemannian structure on $\mathbb{H}^2$ induces a Riemannian structure on S. Explicitly, (see [83; §1.3]) the Riemannian inner product $\langle\ ,\ \rangle$ on $\mathrm{T}_x(S)$ is given by:

$$\langle u, v\rangle = \frac{1}{\Im(\phi(x))^2}\,(u \cdot v)$$

where $\Im$ means imaginary part. We define $\mathrm{UT}_x(S)$ to be the set of all elements of $\mathrm{T}_x(S)$ having norm 1 with respect to this inner product. Thus $\mathrm{UT}_x(S)$ consists of all elements $[\alpha]$ of $\mathrm{T}_x(S)$ satisfying $|(\phi \circ \alpha)'(0)| = \Im(\phi(x))$. One can check that this condition is independent of the choice of chart (using equation (1.1.3) in [83]). We define the unit tangent bundle $\mathrm{UT}(S)$ to be $\coprod_{x \in S} \mathrm{UT}_x(S)$, a subspace of the tangent bundle $\mathrm{T}(S)$.

The mapping $\mathrm{UT}(S) \to \mathrm{UT}(S)$, $(x, [\alpha]) \mapsto (x, [\bar{\alpha}])$, where $\bar{\alpha}(t) = \alpha(-t)$, gives an action of the cyclic group of order 2 on $\mathrm{UT}(S)$, and we define $\mathrm{PT}(S)$, the projective tangent bundle of S, to be the quotient space for this action. If $p : \mathrm{UT}(S) \to \mathrm{PT}(S)$ is the projection map onto the quotient, we can view $p(x, [\alpha])$ as $(x, \langle\alpha\rangle)$, where $\langle\alpha\rangle = [\alpha] \cup [\bar{\alpha}]$.

Since two geodesic parametrisations of a hyperbolic line differ by a linear change of variable, if $x \in S$ and σ is a geodesic on which x lies, there are exactly two parametrisations α of σ such that $(x, [\alpha]) \in \mathrm{UT}(S)$, and they are of the form $(x, [\alpha])$, $(x, [\bar{\alpha}])$. Consequently, elements of $\mathrm{PT}(S)$ may be viewed, when convenient, as pairs (x, σ), where $x \in S$ and σ is a complete geodesic on which x lies.

If S is a complete hyperbolic surface, which is not orientable, it has a 2-sheeted covering space which is orientable. This can be seen without invoking general manifold theory. We have seen that if the group G of covering transformations is Fuchsian, then S can be oriented. Therefore, G is an NEC group having a Fuchsian subgroup H of index 2, and so $\mathbb{H}^2/H$ is the required covering space, by the theory of covering spaces. (Orientability of manifolds is discussed in [22]. A less formal discussion for surfaces can be found in Chapter 1 of [98].)

Lemma 5.10. *Let S be a complete hyperbolic surface of finite area, let $C \in \mathrm{L}(S)$, and let $\tilde{C}$ be the inverse image of C in $\mathbb{H}^2$. Then $\tilde{C} = \bigcup_{l \in \mathcal{C}} l$ for a unique collection $\mathcal{C}$ of pairwise disjoint geodesics in $\mathbb{H}^2$. Further, C is nowhere dense in S.*

Proof. The first step in the proof is to show that $C \neq S$. To prove this in general would require the development of more theory of hyperbolic manifolds, and we give an argument valid only when C is compact. This will be the case in applications. For the general case we refer to Proposition 1.6.3 in [119] or Theorem 4.9 in [27], where it is shown that C in Lemma 5.10 has zero hyperbolic area, by showing that S and $S \setminus C$ have the same area. (It follows immediately, of course, that C is nowhere dense.)

Suppose $C = S$. There is a map $d : S \to \mathrm{PT}(S)$ given by $d(x) = (x, \sigma)$, where σ is the geodesic in $\mathfrak{L}$ on which x lies. If x_i is a sequence of points in $|\mathfrak{L}|$, converging to $x \in C$, let $d(x_i) = (x_i, \sigma_i)$ and $d(x) = (x, \sigma)$. Let (U, ϕ) be a chart with $x \in U$, so for sufficiently large i, $x_i \in U$. Then as $i \to \infty$, the angle between $\phi \circ \sigma$ and $\phi \circ \sigma_i$ approaches zero (an exercise in hyperbolic geometry). It follows that $(x_i, \sigma_i) \to (x, \sigma)$ as $i \to \infty$, hence d is continuous.

Let $\tilde{S}$ be the double covering which orients d. This is obtained from a pullback diagram:

$$\begin{array}{ccc} \tilde{S} & \xrightarrow{r} & \mathrm{UT}(S) \\ {\scriptstyle q}\downarrow & & \downarrow{\scriptstyle p} \\ S & \xrightarrow[d]{} & \mathrm{PT}(S) \end{array}$$

Thus $\tilde{S} = \{(x, v) \in S \times \mathrm{UT}(S) \mid d(x) = p(v)\}$, q and r are the restrictions of the usual projection maps and p is the projection map in the definition of $\mathrm{PT}(S)$ above. If $v = (y, [\alpha]) \in \mathrm{UT}(S)$ and $(x, v) \in \tilde{S}$, then $d(x) = (y, \langle\alpha\rangle)$, hence $x = y$, so we can view elements of $\tilde{S}$ as pairs $(x, [\alpha]) \in \mathrm{UT}(S)$, where α is a geodesic parametrisation of the leaf on which x lies. If $q(x, [\alpha]) = q(y, [\beta])$ then $x = y$ and $p([\alpha]) = p([\beta])$, so either $[\beta] = [\alpha]$ or $[\beta] = [-\alpha]$.

Let $x \in S$; using Lemma 5.9, let (U, ϕ) be a flow box with $x \in U$, and $\phi(U)$ the interior of a rectangle. Choose a geodesic parametrisation α of the local leaf on which x lies. The local leaves map to parallel line segments under ϕ, so for each $y \in U$, we can define α_y to be the geodesic curve such that $(y, [\alpha_y]) \in \mathrm{UT}(S)$ and such that $\phi \circ \alpha_y$, $\phi \circ \alpha$ traverse parallel lines in the same direction in $\phi(U)$. The mapping $s : U \to \tilde{S}$, $y \mapsto (y, [\alpha_y])$ is continuous and an inverse map to $q|_{s(U)}$. Replacing α by $\bar{\alpha}$ replaces s by $\bar{s}$, where $\bar{s}(y) = (y, [\bar{\alpha}_y])$ and $q^{-1}(U) = s(U) \cup \bar{s}(U)$, so U is an evenly covered neighbourhood of S. For $v = (x, [\alpha]) \in \tilde{S}$, we let s_v be whichever of s, $\bar{s}$ takes x to v.

Now $\tilde{S}$ inherits a hyperbolic structure from S (any covering of a manifold with a P-atlas inherits a P-atlas) and every component of $\tilde{S}$ is a hyperbolic surface. Further, q^{-1} restricted to an evenly covered neighbourhood U respects the Riemannian structure, and so induces a mapping $\mathrm{UT}_x(S) \to \mathrm{UT}_{q^{-1}(x)}(\tilde{S})$ for every $x \in U$. Consequently, there is a mapping $\tilde{r} : \tilde{S} \to \mathrm{UT}(\tilde{S})$, where if $v = (x, [\alpha])$, $\tilde{r}(v) = (v, [s_v \circ \alpha])$, which is clearly continuous at v, for all v, so continuous.

Let $\hat{S}$ be a component of $\tilde{S}$ and $\hat{r}$ the restriction of $\tilde{r}$ to $\hat{S}$, a nowhere vanishing tangent vector field on $\hat{S}$. Then $\hat{S}$ is a covering space of S with at most two sheets. Since $S = C$ is a compact surface, as we observed before Lemma 5.8, it has a decomposition as a finite CW-complex. Therefore, $\hat{S}$ has such a decomposition (see [22; Ch.IV, Theorem 8.10]) and the Euler characteristic $\chi(\hat{S}) = \chi(S)$ or $2\chi(S)$ ([22; Ch.IV, Prop.13.5]). If S is orientable, so is $\hat{S}$ (an easy exercise), and by [22] Chapter VII, Corollary 14.5, $\chi(\hat{S}) = 0$, hence $\chi(S) = 0$, contradicting Lemma 5.8. If S is non-orientable, by the discussion preceding the lemma, it has an orientable 2-sheeted covering space S', which is of finite area by [83; Theorem 3.1.2] (which works for NEC groups), and we can lift the lamination to S', again giving a contradiction. Thus $C \neq S$.

The rest of the proof is valid whether or not C is compact, and involves considering $\tilde{C}$, the lift of C to $\mathbb{H}^2$ via the universal covering map $\mathbb{H}^2 \to S$. For $x \in \tilde{C}$, let l_x be the hyperbolic line in $\tilde{C}$ passing through x, so $\tilde{C} = \bigcup_{x \in \tilde{C}} l_x$, and for x, $y \in \tilde{C}$, l_x and l_y are either disjoint or identical. Suppose l is an arc of a hyperbolic line contained in $\tilde{C}$ but not contained in any l_x. We show that this leads to a contradiction. Let $R = \bigcup_{y \in l} l_y$. Choose a direction along l, so there is an induced orientation on each l_y, for $y \in l$ (so that, say, looking along l in the positive direction, l_y crosses l from left to right). Define a mapping $f : l \times \mathbb{R} \to \mathbb{H}^2$ by letting $f(y,t)$ be the point on l_y at distance t from y in the positive direction. Note that for y, $y' \in L$ with $y \neq y'$, $l_y \neq l_{y'}$. It follows that f is continuous (exercise), and obviously its image is in R. Further, f extends to a continuous map from the compact set $l \times (\mathbb{R} \cup \{\pm\infty\})$ into the closed unit disk in $\mathbb{R}^2$, which is a homeomorphism onto the Euclidean closure of R.

Let G be the group of covering transformations of the universal covering $\mathbb{H}^2 \to S$, so G is an NEC group. Let F be a fundamental domain for G of finite hyperbolic area. Choose z to be one of the endpoints $f(y, \pm\infty)$ of l_y, where y is an interior point of the arc l. For sufficiently small $\varepsilon > 0$, the Euclidean disk D with centre z and radius ε has the property that $D \cap \mathbb{H}^2 \subseteq R$. Let $D' \subseteq D$ be the Euclidean disk with centre z and radius $\varepsilon/2$. Then $D' \cap \mathbb{H}^2$ has infinite hyperbolic area (this is left as an exercise), and is covered by the translates of F under G. Hence it meets infinitely many of the sets gF, where $g \in G$ (if $D' \cap \mathbb{H}^2 \subseteq g_1 F \cup \ldots \cup g_n F$, then $\mu(D' \cap \mathbb{H}^2) \leq n\mu(F) < \infty$, where μ denotes hyperbolic area, because as noted in the proof of Lemma 5.7, G preserves μ). Enumerate G, say $G = \{g_1, g_2, \ldots\}$; by Lemma 4.5.3 in [83] or Property (7) in §9.3 of [9], the Euclidean diameter of $g_k F \to 0$ as $k \to \infty$ (the argument works for any NEC group). Hence $D' \cap \mathbb{H}^2$ meets a translate gF of diameter less than $\varepsilon/2$. But then $gF \subseteq D \cap \mathbb{H}^2 \subseteq R \subseteq \tilde{C}$. Since the translates of F cover $\mathbb{H}^2$ and $\tilde{C}$ is G-invariant, we conclude that $\tilde{C} = \mathbb{H}^2$, hence $C = S$, a contradiction.

Thus no such arc l exists. The uniqueness of $\mathcal{C}$ follows, and it also follows that $\tilde{C}$, and so C, has empty interior. Since C is closed, it is nowhere dense. □

Let $\mathcal{G}$ denote the set of hyperbolic lines in $\mathbb{H}^2$. For $l \in \mathcal{G}$, let $e(l)$ be the unordered pair $\{a, b\}$, where a, b are the ideal points of l in $\partial\mathbb{H}^2$, which in the Poincaré disk model is S^1. Then e is a bijection from $\mathcal{G}$ to the set of unordered pairs of distinct elements of $\partial\mathbb{H}^2$, which we may view as the quotient of $((S^1 \times S^1) \setminus \Delta)$, where Δ is the diagonal, by a cyclic group $\langle\tau\rangle$ of order two, where $\tau(a,b) = (b,a)$. This gives $e(\mathcal{G})$ a natural topology, and we topologise $\mathcal{G}$ so that e becomes a homeomorphism. The space $e(\mathcal{G})$ can be described as follows. Let $T = \{(s,t) \in I \times I \mid 0 \leq s < t \leq 1,\ (s,t) \neq (0,1)\}$. Then $e(\mathcal{G})$ is the quotient of T obtained by identifying the pairs of points $(0,t)$ and $(t,1)$ for $0 < t < 1$ (the quotient map $I \times I \to S^1 \times S^1$, $(s,t) \mapsto (e^{2\pi is}, e^{2\pi it})$, restricted to T, is injective, and T is open in $I \times I$). It is easy to see that this quotient space, and so $\mathcal{G}$, is homeomorphic to the open Möbius band. Thus $\mathcal{G}$ is locally compact Hausdorff and second countable.

We make an observation for future use. Let S be a complete hyperbolic surface of finite area and let G be the group of covering transformations for the universal covering map $\mathbb{H}^2 \to S$. Then there is a compact set F in $e(\mathcal{G})$ such that the translates under G of the interior of F cover $e(\mathcal{G})$. Since G has a Fuchsian subgroup H of index at most 2, and $\mathbb{H}^2/H$ has finite area by [83; Theorem 3.1.2] (the argument works for any NEC group), it suffices to prove this when G is Fuchsian.

To do this, choose real a, b with $0 < a < b < 1$, and let F be the image in $e(\mathcal{G})$ of the set $\{(s,t) \in T \mid a \leq t - s \leq b\}$, a compact set. Let x, y be the two points on $S^1 = \partial\mathbb{H}^2$ corresponding to $(t,s) \in T$. Join x, y to the origin by Euclidean line segments. Choose c such that $a < c < b$. Note that G is of the first kind since S has finite area ([83; Theorem 4.5.2]). By Theorem 2.2.2 in [114], we can choose x to be a line transitive point (see [114] for the definition), and then choose y so that $t - s = c$. Since G acts as hyperbolic isometries, it follows from Theorem 2.2.1 and the remarks preceding it in [114] that, if l is a hyperbolic line, we can find $g \in G$ such that the endpoints of gl are as close as we like to x, y (in the Euclidean metric). We can therefore ensure $ge(l) = e(gl)$ lies in the interior of F. Thus F is the required compact set.

Lemma 5.11. *If $\mathcal{C}$ is a collection of pairwise disjoint hyperbolic lines in $\mathbb{H}^2$, then $\bigcup_{l\in\mathcal{C}} l$ is closed in $\mathbb{H}^2$ if and only if $\mathcal{C}$ is closed in $\mathcal{G}$.*

Proof. Suppose that $\mathcal{C}$ is closed in $\mathcal{G}$, and (x_i) is a sequence of points in $\bigcup_{l\in\mathcal{C}} l$ which converge in $\mathbb{H}^2$ to x. We need to show $x \in \bigcup_{l\in\mathcal{C}} l$. Since hyperbolic lines are closed in $\mathbb{H}^2$, we may assume, on passing to a subsequence, that each x_i lies on a hyperbolic line l_i of $\mathcal{C}$, where $l_i \neq l_j$ for $i \neq j$. Let $e(l_i) = \{a_i, b_i\}$. Since $S^1 \times S^1$ is (sequentially) compact, some subsequence of the ordered pairs (a_i, b_i) converges to a point $(a,b) \in S^1 \times S^1$. Then $a \neq b$, otherwise x would lie on the boundary S^1 of $\mathbb{H}^2$. Since $\mathcal{C}$ is closed and e is a homeomorphism, $e(\mathcal{C})$ is closed, so the hyperbolic line l with ideal points a, b is in $\mathcal{C}$. If $x \notin l$, the hyperbolic half-plane determined by l in which x lies contains a neighbourhood of x which is disjoint from all but finitely many l_i (which can be easily seen using the upper half-plane model, taking l to be a vertical half-line). This contradicts the fact that x_i converges to x, so $x \in l$ and $l \in \mathcal{C}$, as required.

Conversely, suppose that $\bigcup_{l\in\mathcal{C}} l$ is closed, and l_i converges to l in $\mathcal{G}$, where $l_i \in \mathcal{C}$. Then $e(l_i)$ converges to $e(l)$; put $e(l_i) = \{a_i, b_i\}$ and $e(l) = \{a, b\}$. Choose a hyperbolic line l' such that the ideal points of l' separate a from b in $\partial\mathbb{H}^2$. Since $\{a_i, b_i\}$ converges to $\{a, b\}$, it follows that, for sufficiently large i, $l_i \cap l' = \{x_i\}$ for some $x_i \in \mathbb{H}^2$, so we obtain a sequence (x_i) of points in $\bigcup_{l\in\mathcal{C}} l$. Some subsequence of (x_i) converges to a point in $\mathbb{H}^2 \cup S^1$, say x, and in fact $x \in \mathbb{H}^2$. Otherwise, after passing to subsequences, (a_i) and (b_i) both converge to x (again this is easy to see using the upper half-plane model), which is impossible as $\{a_i, b_i\}$ converges to $\{a, b\}$. Since $\bigcup_{l\in\mathcal{C}} l$ is closed, $x \in m$ for some $m \in \mathcal{C}$. In fact, $e(m) = \{a, b\}$, otherwise $l_i \cap m$ consists of a single point for sufficiently large i (once more using the upper half-plane model, with m a vertical half-line). This contradicts the assumption that the hyperbolic lines in $\mathcal{C}$ are pairwise disjoint. Hence $l = m \in \mathcal{C}$. □

Let S be a complete hyperbolic surface of finite area. If $C \in \mathrm{L}(S)$, and $\tilde{C}$ is the inverse image of C in $\mathbb{H}^2$, then as we observed above, $\tilde{C} = \bigcup_{l\in\mathcal{C}} l$ for some $\mathcal{C}$ as in Lemma 5.10. The assignment $C \mapsto \mathcal{C}$ is a mapping from $\mathrm{L}(S)$ to the collection of subsets of $\mathcal{G}$. It is well-defined by Lemma 5.10 and clearly injective.

Corollary 5.12. *The mapping just defined is a one-to-one correspondence from $\mathrm{L}(S)$ to the collection of all closed G-invariant subsets $\mathcal{C}$ of $\mathcal{G}$, where G is the group of covering transformations (so G is isomorphic to $\pi_1(S)$), with the property that the hyperbolic lines in $\mathcal{C}$ are pairwise disjoint.*

Proof. If $C \mapsto \mathcal{C}$ under this mapping, we have already observed that $\tilde{C}$ is closed, G-invariant and the lines in $\mathcal{C}$ are pairwise disjoint. Hence $\mathcal{C}$ is G-invariant by Lemma 5.10 and is closed by Lemma 5.11.

Conversely, if $\mathcal{C}$ is a closed, G-invariant subset of $\mathcal{G}$ whose elements are pairwise disjoint, let $\tilde{C} = \bigcup_{l\in\mathcal{C}} l$, let $p : \mathbb{H}^2 \to \mathbb{H}^2/G = S$ be the projection map and let $C = p(\tilde{C})$. Since $\mathcal{C}$ is G-invariant, so is $\tilde{C}$, hence $\tilde{C} = p^{-1}(C)$, by the theory of covering spaces, and it suffices to show $C \in \mathrm{L}(S)$. By Lemma 5.11, $\tilde{C}$ is closed in $\mathbb{H}^2$. Since p is a covering map, C is closed in S, and if $l \in \mathcal{C}$ then its image in S under projection is a 1-manifold, so a simple geodesic. Finally if l_1, l_2 are two lines in $\mathcal{C}$ such that $p(l_1)$ and $p(l_2)$ intersect, then for some $g \in G$, l_1 and gl_2 intersect, hence $l_1 = gl_2$, so $p(l_1) = p(l_2)$. Thus $C \in \mathrm{L}(S)$. □

In order to be able to construct $\mathbb{R}$-trees using geodesic laminations, we need to consider transverse measures on them.

Definition. A measured geodesic lamination is a geodesic lamination with a transverse measure of full support.

Let S be a complete hyperbolic surface. We denote by $\mathrm{ML}(S)$ the set of all pairs (C, μ), where $C \in \mathrm{L}(S)$, so that C defines a lamination $\mathcal{L}$ on S

by Lemma 5.9, and μ assigns to every arc κ transverse to $\mathcal{L}$ a regular Borel measure μ_κ such that (i)–(iii) in Lemma 4.11 hold.

Lemma 5.13. *An element* (C, μ) *of* ML(S) *determines a measured geodesic lamination on* S *satisfying Conditions* (a)-(e) *in Theorem* 4.9, *so has a dual* $\mathbb{R}$*-tree.*

Proof. It determines a geodesic lamination $\mathcal{L}$ by Lemma 5.9. The leaves of $\tilde{\mathcal{L}}$ (the lift of $\mathcal{L}$ to $\mathbb{H}^2$) are hyperbolic lines, so are closed in $\mathbb{H}^2$ and Condition (b) is satisfied. Further, the leaves of C are closed, being the images of hyperbolic lines under a covering map. By Lemma 4.11, (C, μ) determines a measured geodesic lamination, and Condition (a) is satisfied. We have seen that S is metrisable, so paracompact, and the hyperbolic structure gives a C^∞ structure on S. Since the leaves are projections of hyperbolic lines in $\mathbb{H}^2$, they are C^∞, hence (c) holds.

Now $\mathbb{H}^2$ has a covering by flow boxes which are convex hyperbolic quadrilaterals (see the proof of Lemma 5.9). If p, q are points in such a flow box, not in the support of the lamination, then the hyperbolic line segment pq between them meets every other hyperbolic line in at most one point, and the tangents to the two lines at the point of intersection differ. Hence pq lies in the flow box, crosses each leaf at most once and is smooth, when suitably parametrised to give a path between p and q, whence (d). Finally, (e) holds for similar reasons, using the exercise after Lemma 5.9. □

We can now give an alternative description of elements of ML(S), using that we gave for L(S) in Corollary 5.12.

Lemma 5.14. *Let* S *be a complete hyperbolic surface of finite area and let* $G = \pi_1(S)$. *Elements of* ML(S) *are in one-to-one correspondence with the set of regular Borel measures* μ *on* $\mathcal{G}$ *satisfying*

(a) μ *is invariant under* G;

(b) *If* l_1, l_2 *are distinct hyperbolic lines in the support of* μ *then* $l_1 \cap l_2 = \emptyset$.

Proof. Take $(C, \mu) \in$ ML(S), and let $\mathcal{C}$ be the corresponding closed subset of $\mathcal{G}$ given by Corollary 5.12. We define a Borel measure on $\mathcal{G}$ with support $\mathcal{C}$. Recall that the hyperbolic lines in $\mathcal{C}$ are pairwise disjoint, so (b) is satisfied. Note that (C, μ) determines a measured lamination on S by Lemma 5.13, which lifts to a measured lamination on $\mathbb{H}^2$ of full support, so there is a regular Borel measure μ_κ on each arc κ in $\mathbb{H}^2$ transverse to this lamination, by Lemma 4.11, and the support of μ_κ restricted to the open arc κ_o is $\kappa_o \cap \tilde{C}$, where $\tilde{C} = \bigcup_{l \in \mathcal{C}} l$, the lift of C to $\mathbb{H}^2$.

Let I, J be two disjoint closed intervals on $S^1 = \partial\mathbb{H}^2$. The set $\mathcal{I}$ of hyperbolic lines which have one end in I and the other in J contains two extremal lines l_1, l_2 such that all lines in $\mathcal{I}$ lie between l_1 and l_2, in what is hoped is an obvious sense. Note that $\mathcal{I}$ is compact, since $e(\mathcal{I})$ is the image of

$I \times J$ under the projection onto $((S^1 \times S^1) \setminus \Delta)/\langle\tau\rangle$ (see the observations after Lemma 5.10). Let κ be a geodesic segment joining a point of l_1 to a point of l_2. Then the mapping $f : \mathcal{I} \to \kappa$ given by $f(l) =$ the point of intersection of l and κ is continuous and κ is Hausdorff, hence $f|_{\mathcal{I}\cap\mathcal{C}}$ is a homeomorphism. If $\mathcal{A}$ is a Borel subset of $\mathcal{I}$, $\mathcal{A} \cap \mathcal{C}$ is a Borel subset of $\mathcal{C}$, so $f(\mathcal{A} \cap \mathcal{C})$ is a Borel subset of κ, and we define $\mu_{\mathcal{I}}(\mathcal{A}) = \mu_\kappa(f(\mathcal{A} \cap \mathcal{C}))$. It is easily seen that $\mu_{\mathcal{I}}$ is a regular Borel measure on $\mathcal{I}$ with support $\mathcal{I} \cap \mathcal{C}$. If I or J is replaced by a subinterval, to obtain $\mathcal{J}$, then $\mu_{\mathcal{I}}(\mathcal{A}) = \mu_{\mathcal{J}}(\mathcal{A})$ for Borel subsets $\mathcal{A}$ in $\mathcal{J}$, using uniqueness in Corollary 4.6. Now any two arcs joining a point of l_1 to a point in l_2 are homotopic by a homotopy as described in (ii) of Lemma 4.11 (exercise). An argument similar to that in Part (2) of Lemma 4.12, together with Remark (b) in the proof of Lemma 4.11, shows that $\mu_{\mathcal{I}}$ is independent of the choice of κ. Details are left to the determined reader.

Hence, if $\mathcal{I}'$ is obtained from intervals I', J' in the same way as $\mathcal{I}$ is obtained from I, J, then $\mu_{\mathcal{I}}$ and $\mu_{\mathcal{I}'}$ agree on Borel subsets of $\mathcal{I} \cap \mathcal{I}'$, so by Corollary 4.6 the measures $\mu_{\mathcal{I}}$ have a unique extension to a regular Borel measure ν on $\mathcal{G}$. Now for $g \in G$ and Borel subsets A of κ, $\mu_\kappa(A) = \mu_{g\kappa}(gA)$ (exercise), and it follows that, for $g \in G$, and Borel subsets $\mathcal{A}$ of $\mathcal{I}$, $\mu_{g\mathcal{I}}(g\mathcal{A}) = \mu_{\mathcal{I}}(\mathcal{A})$. Defining $\nu'(\mathcal{A}) = \nu(g\mathcal{A})$, for Borel subsets $\mathcal{A}$ of $\mathcal{G}$, gives a regular Borel measure ν' on $\mathcal{G}$, and if $\mathcal{A} \subseteq \mathcal{I}$, $\nu'(\mathcal{A}) = \nu(g\mathcal{A}) = \mu_{g\mathcal{I}}(g\mathcal{A}) = \mu_{\mathcal{I}}(\mathcal{A})$, so by uniqueness $\nu' = \nu$, that is, ν is G-invariant, so Condition (a) holds.

Since $\mathcal{G}$ is second countable, if $\mathcal{U}$ is an open subset of $\mathcal{G}$, there is a countable collection of sets $\{\mathcal{I}_j\}_{j=1}^{\infty}$ as above such that $\mathcal{U} \subseteq \bigcup_{j=1}^{\infty} \mathring{\mathcal{I}}_j$, where $\mathring{\mathcal{I}}_j$ means $\mathcal{I}_j$ with its two extremal lines removed. For each j, there are arcs κ_j and maps $f_j : \mathcal{I}_j \to \kappa_j$ as above. If $\mathcal{U} \cap \mathcal{C} = \emptyset$, then $\nu(\mathcal{U}) \leq \sum_{j=1}^{\infty} \mu_{\mathcal{I}_j}(\mathcal{I}_j \cap \mathcal{U}) = 0$, since $\mathcal{I}_j \cap \mathcal{U} \cap \mathcal{C} = \emptyset$. If $\mathcal{U} \cap \mathcal{C} \neq \emptyset$, choose j such that $\mathring{\mathcal{I}}_j \cap \mathcal{U} \cap \mathcal{C} \neq \emptyset$. Then $\nu(\mathcal{U}) \geq \nu(\mathring{\mathcal{I}}_j \cap \mathcal{U} \cap \mathcal{C}) = \mu_{\kappa_j}(f_j(\mathring{\mathcal{I}}_j \cap \mathcal{U} \cap \mathcal{C})) > 0$, since $f_j(\mathring{\mathcal{I}}_j \cap \mathcal{U} \cap \mathcal{C})$ is an open subset of $\lambda \cap \tilde{C}$, where $\lambda = (\kappa_j)_o$, and $\lambda \cap \tilde{C}$ is the support of μ_{κ_j} restricted to λ. Hence $\mathcal{C}$ is the support of ν.

Conversely, suppose given a regular Borel measure ν on $\mathcal{G}$ with support $\mathcal{C}$ satisfying (b) and invariant under $G = \pi_1(S)$, so $\mathcal{C}$ is G-invariant. Recall (§4) that the support of a regular Borel measure is closed, so $\mathcal{C}$ determines an element C of $\mathrm{L}(S)$ by Corollary 5.12, giving a geodesic lamination $\mathcal{L}$ on S by Lemma 5.9. Let $\tilde{\mathcal{L}}$ be the lift of $\mathcal{L}$ to a lamination on $\mathbb{H}^2$, so $|\mathcal{L}| = \bigcup_{l\in\mathcal{C}} l = \tilde{C}$, the lift of C.

If κ is an arc in S transverse to $\mathcal{L}$, choose a lifting to an arc $\tilde{\kappa}$ in $\mathbb{H}^2$. Now κ has an open covering by flow boxes lying in an evenly covered neighbourhood in which κ intersects each leaf in at most one point. By taking a Lebesgue number, and using the fact that C is nowhere dense (Lemma 5.10), we can divide κ into finitely many subarcs, each of which intersects every leaf of $\mathcal{L}$ in at most one point, in such a way that the endpoints of the arcs do not lie in C. Hence there are points in $\tilde{\kappa} \setminus \tilde{C}$ which cut $\tilde{\kappa}$ into finitely many arcs κ_i which intersect each leaf of $\tilde{\mathcal{L}}$ at most once. Since $\mathbb{H}^2 \setminus \tilde{C}$ is open, we can extend κ_i to a slightly larger subarc, without changing $\kappa_i \cap \tilde{C}$, so that the interiors of the κ_i now cover $\tilde{\kappa}$. Let $\mathcal{K}_i = \{l \in \mathcal{C} \mid l \cap \kappa_i \neq \emptyset\}$. The mapping $g : \kappa_i \cap \tilde{C} \to \mathcal{K}_i$ given by $g(x) =$ the leaf of $\tilde{\mathcal{L}}$ on which x lies is continuous (exercise in hyperbolic geometry), bijective, from a compact to a Hausdorff space, so a homeomorphism. Hence there is a regular Borel measure μ_i on κ_i given by $\mu_i(A) = \nu(g(A))$, which extend to a regular Borel measure $\mu_{\tilde{\kappa}}$ on $\tilde{\kappa}$ by Corollary 4.6. By projection we obtain a regular Borel measure μ_κ on κ. This is independent of the choice of lifting $\tilde{\kappa}$ because ν is G-invariant. It is again left to the reader to check that this gives an element of ML(S), and the required bijection from ML(S) to the set of regular Borel measures on $\mathcal{G}$ satisfying (a) and (b). □

Thus, if S is a complete hyperbolic surface of finite area, ML(S) can be identified with a subset of the space $M(\mathcal{G})$ of all regular Borel measures on $\mathcal{G}$. We topologise $M(\mathcal{G})$ with the weak topology as in §4, and we give ML(S) the subspace (relative) topology.

Lemma 5.15. *The subspace* ML(S) *is closed in* $M(\mathcal{G})$.

Proof. The action of $G = \pi_1(S)$ on $\mathcal{G}$ induces an action on the space $M(\mathcal{G})$ (see Remark (2) after Proposition 4.4), and $M(\mathcal{G})$ is Hausdorff (for example, because it is a subspace of $\mathbb{R}^{C(\mathcal{G})}$–see §4). For $g \in G$, the set of fixed points of g in $M(\mathcal{G})$ is therefore closed, and so the set of G-invariant measures in $M(\mathcal{G})$ is closed, being an intersection of closed sets, one for each $g \in G$.

It suffices to show that the set V of measures in ML(S) which fail to satisfy (b) in Lemma 5.14 is open. Let $\mu \in V$. Then there are hyperbolic lines l_1, l_2 in the support of μ which cross. By definition of the topology on $\mathcal{G}$, there are open neighbourhoods U_i of l_i ($i = 1, 2$) in $\mathcal{G}$ such that, for all $l_1' \in U_1$, $l_2' \in U_2$, we have $l_1' \cap l_2' \neq \emptyset$. The measures that are positive on both U_1 and U_2 form an open neighbourhood of μ in the weak topology, by Remark 1 after Proposition 4.4, and this neighbourhood is contained in V, so V is open. (For any open set U of $\mathcal{G}$ and $\nu \in M(\mathcal{G})$, $\nu(U) > 0$ if and only if U intersects supp(ν).) □

The multiplicative group $\mathbb{R}^+$ of positive real numbers acts as homeomorphisms on $M(\mathcal{G}) \setminus \{0\}$ by "homothety": $(r\mu)(A) = r\mu(A)$ for $r \in \mathbb{R}^+$, $\mu \in M(\mathcal{G}) \setminus \{0\}$ and A a Borel set in $\mathcal{G}$. Let P($\mathcal{G}$) be the quotient $(M(\mathcal{G}) \setminus \{0\})/\mathbb{R}^+$, with the quotient topology, so P($\mathcal{G}$) is a Hausdorff space. The set ML(S) $\setminus \{0\}$ is invariant under this action, and its image in P($\mathcal{G}$) will be called ML(S) *projectivised*, denoted by PML(S) (by analogy with real n-dimensional projective space, which is the quotient of $\mathbb{R}^{n+1} \setminus \{0\}$ by the action of $\mathbb{R}^+$ by multiplication). Inspection of the proof of Theorem 5.14 shows that two elements (C, μ), (C', μ') of ML(S) with μ, μ' not identically zero represent the same element of PML(S) if and only if $C = C'$ and $\mu = r\mu'$ for some positive real number r.

Theorem 5.16. *Let S be a complete hyperbolic surface of finite area. Then the space* PML(S) *is compact and Hausdorff.*

Proof. Since P($\mathcal{G}$) is Hausdorff, so is PML(S). By the remarks preceding Lemma 5.11, we can find a compact set F in $\mathcal{G}$ such that the translates under $G = \pi_1(S)$ of $\mathring{F}$ (the interior of F) cover $\mathcal{G}$. Let E be the set of all G-invariant measures $\mu \in M(\mathcal{G})$ such that $\mu(F) = 1$. The projection $\rho : \mathrm{ML}(S) \cap E \to \mathrm{PML}(S)$ is bijective, so it is enough to show $\mathrm{ML}(S) \cap E$ is compact. By Lemma 5.15, it suffices to show that E is compact. Let E_0 be the set of all measures $\nu \in M(F)$ such that $\nu(F) = 1$ and

(∗) for all $g \in G$ and every Borel set $A \subseteq F \cap g^{-1}F$, it is the case that $\nu(A) = \nu(gA)$.

It follows from condition (∗) that, if $\nu \in E_0$, we can define, for each $g \in G$, a regular Borel measure ν_g on gF by $\nu_g(A) = \nu(g^{-1}A)$, and that if A is a Borel set in $g_1F \cap g_2F$, then $\nu_{g_1}(A) = \nu_{g_2}(A)$. By Corollary 4.6, the measures ν_g have a unique extension to a regular Borel measure μ on $\mathcal{G}$. For $g \in G$, we can define μ_g by $\mu_g(A) = \mu(gA)$, to get another extension of the ν_g. By uniqueness, $\mu_g = \mu$ for all $g \in G$, that is, μ is G-invariant. This defines a bijection $\Phi : E_0 \to E$. To see Φ is continuous, if $f \in C(\mathcal{G})$ (the space of continuous real valued functions on $\mathcal{G}$ with compact support), by the argument for Prop.1 in Ch.III, §2 of [20], we can find finitely many elements of G, say $g_1, \ldots, g_n$, and $f_i \in C(\mathcal{G})$ for $1 \leq i \leq n$ such that $\mathrm{supp}(f) \subseteq g_1\mathring{F} \cup \ldots \cup g_n\mathring{F}$, $\mathrm{supp}(f_i) \subseteq g_i\mathring{F}_i$ for $1 \leq i \leq n$ and $f = \sum_{i=1}^{n} f_i$. Define $h_i \in C(\mathcal{G})$ by $h_i(x) = f_i(g_i x)$, so that h_i can be viewed as an element of $C(F)$ by restriction. Then for $\nu \in E_0$ and $\mu = \Phi(\nu)$,

$$\int f d\mu = \sum_{i=1}^{n} \int f_i d\mu = \sum_{i=1}^{n} \int h_i d\mu = \sum_{i=1}^{n} \int h_i d\nu$$

and it follows easily that Φ is continuous. Thus it suffices to show E_0 is compact.

By Remark 3 after Proposition 4.4, the set of measures $\nu \in M(F)$ such that $\nu(F) = 1$ is compact, so it suffices to show that the set of measures $\nu \in M(F)$ satisfying (∗) is closed. First note that this set coincides with the set of $\nu \in M(F)$ satisfying:

(∗∗) for all $g \in G$ and every Borel set $A \subseteq \mathring{F} \cap g^{-1}\mathring{F}$, $\nu(A) = \nu(gA)$.

For repeating the argument above, if ν satisfies (∗∗), there is a unique Borel measure on $\mathcal{G}$ extending the measures ν_g restricted to $g\mathring{F}$, by Proposition 4.5, and this measure must be $\Phi(\nu)$, which is G-invariant, hence ν satisfies (∗). Fix $g \in G$, let $f : M(F) \to M(\mathring{F} \cap g^{-1}\mathring{F})$, $f' : M(F) \to M(\mathring{F} \cap g\mathring{F})$ be the restriction maps and let $\bar{g} : M(\mathring{F} \cap g^{-1}\mathring{F}) \to M(\mathring{F} \cap g\mathring{F})$ be the homeomorphism induced by g (Remark 2 after Proposition 4.4). Note that, as observed before Proposition 4.5, f and f' are continuous. Then the set

$$\{\nu \in M(F) \mid \text{for all Borel sets } A \subseteq \mathring{F} \cap g^{-1}\mathring{F},\ \nu(A) = \nu(gA)\}$$

is closed, being the equaliser of $\bar{g} \circ f$ and f'. Hence the set of $\nu \in M(F)$ satisfying $(**)$ is closed, being an intersection of closed sets. □

We shall show that, with a few exceptions, a closed surface, that is, a compact connected 2-manifold without boundary, can be made into a complete hyperbolic surface of finite area. There is a well-known classification of closed surfaces, as follows. Let S^n denote the n-sphere. The orientable surface of genus $g \geq 0$ is defined to be the connected sum $S^2 \# T \# \ldots \# T$, where $T = S^1 \times S^1$ is the torus, and there are g factors T. The non-orientable surface of genus $g \geq 1$ is the connected sum $\mathbb{R}P^2 \# \ldots \# \mathbb{R}P^2$ of g copies of the real projective plane $\mathbb{R}P^2$. (The notion of connected sum is discussed in [98], Ch.1, §4.) Let S be a closed surface; then S is homeomorphic to exactly one of these surfaces (see [99], Theorem 3.4.7 and Corollary 3.4.10, or [98], Ch.1, §10). Ignoring the sphere S^2, they can all be obtained by appropriately identifying the sides of a regular polygon, with $4g$ sides in the orientable case and $2g$ sides in the non-orientable case. This gives corresponding presentations of their fundamental group (see [99], Theorem 3.4.9). The group $\pi_1(S)$ has either a presentation

$$\langle a_1, b_1, \ldots, a_g, b_g \mid [a_1, b_1] \ldots [a_g, b_g] = 1 \rangle$$

(for the orientable surface of genus g), or

$$\langle x_1, \ldots, x_g \mid x_1^2 \ldots x_g^2 = 1 \rangle$$

(for the non-orientable surface of genus g). This identification of sides leads to a decomposition of the surface as a CW-complex with one face and one vertex (all the vertices of the polygon are identified), and $2g$ edges in the orientable case, giving the Euler characteristic $\chi(S) = 2 - 2g$, and g edges in the non-orientable case, giving $\chi(S) = 2 - g$. The first formula also works for the sphere (orientable with $g = 0$) as is well-known and easy to see. (The sphere decomposes as a CW-complex with one point as 0-skeleton and a great circle passing through it as 1-skeleton, giving two hemispheres as faces.)

If $g \geq 2$ in the orientable case, or $g \geq 3$ in the non-orientable case, the polygon has n sides, where $n > 4$, so can be realised as a regular polygon in the hyperbolic plane in which the sum of the interior angles is 2π. (See §7.16 in [9].) Further, the necessary identifications of pairs of sides can be performed by hyperbolic isometries, in such a way that a theorem of Poincaré, as stated in §2 of [97], applies. This is nicely illustrated for the orientable genus 2 case in [141], Figs. 1.12 and 1.13, and in general is left to the reader. The theorem of Poincaré guarantees that the group G generated by the side-pairing isometries is an NEC group, and that the regular polygon together with its interior is a fundamental domain for the action of G on $\mathbb{H}^2$. Hence S is homeomorphic to $\mathbb{H}^2/G$, by [82; Theorem 5.9.6], which works for any NEC group. Poincaré's theorem also implies G is isomorphic to $\pi_1(S)$, which is torsion-free. This follows from the theory of one-relator groups (see [95; Ch.IV, Theorem 5.2]), or just by writing $\pi_1(S)$ as a free product with amalgamation of two free groups. Therefore G acts freely on $\mathbb{H}^2$. The polygon has zero area and the polygon together with its interior is compact, so has finite area. Thus S has been made into a complete hyperbolic surface of finite area.

For the other compact surfaces, the Euler characteristic is greater than or equal to 0, so they cannot be made into complete hyperbolic surfaces. Otherwise, a corresponding Dirichlet region would be compact ([82; Theorem 5.9.8], which works for NEC groups), so the area of the surface would be finite, hence the Euler characteristic would be negative (Lemma 5.8), a contradiction.

Finally, if S is a complete hyperbolic surface of finite area which is not compact, then $\pi_1(S)$ is free, of rank $1 - \chi(S)$. (See the exercises at the end of Ch.6, §4 in [98]; if S has n punctures, its deformation retracts onto a compact bordered surface with n holes.)

6. Spaces of Actions on $\mathbb{R}$-trees

In §5 it was shown that the space PML(S) of projectivised measured geodesic laminations on a hyperbolic surface S of finite area is compact, and by Lemma 5.13, any element of ML(S) gives rise to an action of $\pi_1(S)$ on an $\mathbb{R}$-tree. This leads to the question of whether or not the space of actions of a given group on $\mathbb{R}$-trees, suitably projectivised, is compact. The initial problem is to make a topological space out of the actions of the group on $\mathbb{R}$-trees. Apart from any other problems in doing this, there is the set theoretic one that the class of all such actions is not a set, so we cannot simply consider all these actions as points of our space. The way to proceed is to note that, by 3.4.2(c), if an action is non-trivial, there is a minimal invariant subtree, and this subtree is uniquely determined by the hyperbolic length function for the action. We therefore consider the set of all hyperbolic length functions for actions of a given group G on $\mathbb{R}$-trees. This set is a subset of $\mathbb{R}^G$ which we denote by LF(G). Giving $\mathbb{R}^G$ the product (Tychonoff) topology and LF(G) the subspace (relative) topology, we make a topological space out of the actions of G on $\mathbb{R}$-trees. The space LF(G) is a subspace of $[0,\infty)^G$, and the multiplicative group $\mathbb{R}^+$ of positive real numbers acts as homeomorphisms on $[0,\infty)^G \setminus \{0\}$ by "homothety": $(rf)(g) = rf(g)$ for $r \in \mathbb{R}^+$, $f \in [0,\infty)^G \setminus \{0\}$ and $g \in G$.

The quotient space $([0,\infty)^G \setminus \{0\})/\mathbb{R}^+$, with the quotient topology, is denoted by $\mathrm{P}(G)$ (if G is finite with n elements, then $\mathrm{P}(G)$ is a subspace of $(n-1)$-dimensional real projective space). The image of $\mathrm{LF}(G)$ in $\mathrm{P}(G)$ under the canonical projection map $p : [0,\infty)^G \setminus \{0\} \to \mathrm{P}(G)$ is denoted by $\mathrm{PLF}(G)$, and is given the subspace topology. (Note that, if ℓ is a hyperbolic length function of G, $r\ell$ (where $r > 0$) is the hyperbolic length function for the action on the $\mathbb{R}$-tree obtained by scaling the metric by r. Also, in taking a quotient of $[0,\infty)^G \setminus \{0\}$, we exclude all trivial actions, which are those for which the hyperbolic length function is identically zero.)

We shall show that, if G is finitely generated, then $\mathrm{PLF}(G)$ is compact, a result of Culler and Morgan [42]. If G is finitely generated, it is countable, so $[0,\infty)^G \setminus \{0\}$ is a second-countable space. Since $p^{-1}p(U) = \bigcup_{r\in\mathbb{R}^+} rU$ for $U \subseteq [0,\infty)^G$, and since $\mathbb{R}^+$ acts as homeomorphisms, p is an open map, hence $\mathrm{P}(G)$ is second-countable, and therefore so is any subspace. Thus for subspaces of $\mathrm{P}(G)$, compactness is equivalent to sequential compactness.

One problem is to explicitly describe elements of $\mathrm{PLF}(G)$, or to put it another way, to characterise hyperbolic length functions for actions on $\mathbb{R}$-trees. In fact, one can give axioms for hyperbolic length functions, which work for any group and any ordered abelian group Λ. We shall only give the first step in doing this, which is the following lemma.

Lemma 6.1. *Let $\ell : G \to \Lambda$, where G is a group and Λ is an ordered abelian group, be the hyperbolic length function for the action of G on some Λ-tree. Then for all g, $h \in G$:*

(0) *if $\ell(g) > 0$ and $\ell(h) > 0$ then* $\max\{0, \ell(gh) - \ell(g) - \ell(h)\} \in 2\Lambda$;
(I) $\ell(ghg^{-1}) = \ell(h)$;
(II) *either* $\ell(gh) = \ell(gh^{-1})$ *or* $\max\{\ell(gh), \ell(gh^{-1})\} \leq \ell(g) + \ell(h)$;
(III) *if* $\ell(g) > 0$ *and* $\ell(h) > 0$ *then either* $\ell(gh) = \ell(gh^{-1}) > \ell(g) + \ell(h)$ *or* $\max\{\ell(gh), \ell(gh^{-1})\} = \ell(g) + \ell(h)$.

Proof. The first property (0) follows from Corollary 3.2.4, and II follows from Lemmas 3.2.2 and 3.2.3, bearing in mind that $\ell(h) = \ell(h^{-1})$ and $A_h = A_{h^{-1}}$ by Lemma 3.1.7, and I also follows from Lemma 3.1.7. Finally III follows from Lemma 3.2.2 and Lemmas 3.3.3, 3.3.4 and 3.3.5 (in the last case also invoking Lemma 3.3.1). □

We shall call a function $\ell : G \to \Lambda$ such that $\ell(g) \geq 0$ for all $g \in G$ and which satisfies 0, I, II and III in Lemma 6.1 a *pseudo-length function*. It has been shown by Parry [115] that every pseudo-length function is the hyperbolic length function for the action of G on some Λ-tree. Thus hyperbolic length functions are characterised axiomatically by 0, I, II and III. However, the proof of Parry's Theorem is long and is omitted. Note that, in case $\Lambda = \mathbb{R}$, Axiom (0) is redundant. Also, it is an easy exercise to deduce from these axioms that

$\ell(1) = 0$, $\ell(h) \geq 0$ and $\ell(h) = \ell(h^{-1})$ for $h \in G$. The axioms are based on those proposed in [42].

The image of the set of $\mathbb{R}$-valued pseudo-length functions in $\mathrm{P}(G)$ will be denoted by $\Psi\mathrm{LF}(G)$. Following Culler and Morgan [42], we shall show that $\Psi\mathrm{LF}(G)$ is compact. It follows from Parry's result that $\mathrm{PLF}(G) = \Psi\mathrm{LF}(G)$, so $\mathrm{PLF}(G)$ is compact. However, Culler and Morgan were able to show that $\mathrm{PLF}(G)$ is a closed subspace of $\Psi\mathrm{LF}(G)$, and so deduce the compactness of $\mathrm{PLF}(G)$, without recourse to Parry's result, which was not available at the time. We shall indicate how they showed $\mathrm{PLF}(G)$ is closed, but some of the details will be omitted.

Suppose G is a finitely generated group and fix a set of generators for G, say $g_1, \ldots, g_n$. Let D be the set of all elements of G of the form $g_{i_1}^{\pm 1} \ldots g_{i_k}^{\pm 1}$, where $i_1, \ldots, i_k$ are distinct elements from $\{1, \ldots, n\}$, so D is finite. Further, let $|g|$ be the length of a shortest word in the generators and their inverses representing g.

Lemma 6.2. *In these circumstances, if $\ell : G \to \mathbb{R}$ is a pseudo-length function, then $\ell(g) \leq M|g|$, where $M = \max_{g \in D}(\ell(g))$.*

Proof. The proof is by induction on $|g|$, and we first note that the result is clearly true if $g \in D$. Assume that $\ell(h) \leq M|g|$ when $|h| < n$. Let $g \in G$ have $|g| = n$, and let w be a word in the generators and their inverses of length n which represents g. Replacing w by a cyclic conjugate does not increase $|g|$, and $\ell(g)$ depends only on the conjugacy class of g, so it suffices to consider the case that w is cyclically reduced.

We can also assume $g \notin D$, so w may be assumed to be of one of the following two forms:

(a) $w = ag_ibg_i$
(b) $w = ag_ibg_i^{-1}$

for some i, where a, b are subwords of w and in case (b) are non-empty. By Axiom III in Lemma 6.1 and induction, in case (b) either $\ell(g) = \ell(ag_i^2b^{-1})$, or

$$\ell(g) \leq \ell(ag_i) + \ell(g_ib^{-1}) \leq M(|ag_i| + |g_ib^{-1}|) = M\,|g|.$$

(The right-hand equality is because w is a word of minimal length representing g.) Thus we can assume $\ell(g) = \ell(ag_i^2b^{-1}) = \ell(b^{-1}ag_i^2)$, and either $|b^{-1}ag_i^2| < |g|$, when the result follows by induction, or $b^{-1}ag_i^2$ falls in Case (a). Thus it suffices to consider Case (a); in this case, again using Axiom III and the induction hypothesis, either

$$\ell(g) = \ell((ag_i)(bg_i)) = \ell(ab^{-1}) \leq M(|g| - 2) < M|g|$$

or

$$\ell(g) \leq \ell(ag_i) + \ell(bg_i) \leq M(|ag_i| + |bg_i|) = M\,|g|.$$

This completes the inductive proof. □

Theorem 6.3. *If G is a finitely generated group, the space $\Psi\mathrm{LF}(G)$ is compact.*

Proof. Let $J = \prod_{g\in G}[0, |g|]$, a compact subspace of $[0,\infty)^G$ by Tychonoff's Theorem. Let K be the subspace of J consisting of all $x \in J$ such that $x(g) = 1$ for some $g \in D$. Denoting the projection onto the gth coordinate by π_g (so $\pi_g(x) = x(g)$), $K = \bigcup_{g\in D}(J \cap \pi_g^{-1}(1))$ is a finite union of closed sets in J, so is closed and hence compact.

Let L be the set of elements of K which are pseudo-length functions. If $\alpha \in \Psi\mathrm{LF}(G)$, and $\ell \in [0,\infty)^G \setminus \{0\}$ is a pseudo-length function with $p(\ell) = \alpha$, then $\ell(g) > 0$ for some $g \in D$, and multiplying ℓ by a positive real number, we can assume $M = \max\{\ell(g)|\ g \in D\} = 1$. By Lemma 6.2, $\ell(g) \in [0, |g|]$ for all $g \in G$, and $\ell(g) = 1$ for some $g \in D$. Thus $\ell \in L$, so $\alpha \in p(L)$ and $\Psi\mathrm{LF}(G) = p(L)$. It therefore suffices to show L is compact, and so it suffices to show it is closed in K. But if ℓ_n is a sequence in L converging to ℓ in K, then $\ell \in L$, that is, ℓ satisfies I, II and III in Lemma 6.2. This follows because of the continuity of the projection maps π_g and elementary properties of real sequences. Since K is second-countable, it follows that L is closed. □

To deduce that $\mathrm{PLF}(G)$ is compact, we shall deviate slightly from [42] and use some results in [2]. The starting point is a simple lemma.

Lemma 6.4. *Let $\ell : G \to \Lambda$ be a pseudo-length function. Then for all $g \in G$ and integers n, $\ell(g^n) = |n|\ell(g)$.*

Proof. Since $\ell(g) = \ell(g^{-1})$, we can assume $n \geq 0$, and the result is clearly true when $n = 0$, 1. We use induction on n, so we suppose $n \geq 2$ and the result is true for all non-negative integers less than n.

If $\ell(g) = 0$, then apply Axiom II to g^{n-1} and g, to see that either $\ell(g^n) = \ell(g^{n-2}) = 0$, or

$$\max\{\ell(g^n), \ell(g^{n-2})\} \leq \ell(g^{n-1}) + \ell(g) = 0,$$

so $\ell(g^n) = 0$.

If $\ell(g) > 0$, then apply Axiom III to g^{n-1} and g. By induction, $\ell(g^{n-2}) = \ell(g^{n-1}g^{-1}) < \ell(g^{n-1}) + \ell(g)$, so $\max\{\ell(g^n), \ell(g^{n-2})\} = \ell(g^{n-1}) + \ell(g) = |n|\ell(g)$, so $\ell(g^n) = |n|\ell(g)$. □

In Chapter 3 we classified actions as abelian, dihedral or irreducible, and we can do the same for pseudo-length functions.

Definition. A pseudo-length function $\ell : G \to \Lambda$ is *abelian* if $\ell(gh) \leq \ell(g) + \ell(h)$ for all g, $h \in G$, *dihedral* if it is non-abelian but $\ell(gh) \leq \ell(g) + \ell(h)$ for all g, $h \in G$ with $\ell(g)$, $\ell(h) > 0$, and *irreducible* if it is neither abelian nor dihedral.

We shall need a limited version of Parry's Theorem for abelian pseudo-length functions.

Lemma 6.5. *If $\ell : G \to \Lambda$ is an abelian pseudo-length function, then ℓ is the hyperbolic length function for the action of G on some Λ-tree.*

Proof. We shall deduce this from a result in [2], and to do this we note that the following are true, for all g, $h \in G$:

(1) $0 = \ell(1) \leq \ell(g) = \ell(g^{-1})$;
(2) $\ell(ghg^{-1}) = \ell(h)$;
(3) $\{\beta(g,h), \beta(g,h^{-1})\} = \{0, 2m\}$, where $\beta(g,h) = \ell(gh) - \ell(g) - \ell(h)$ and $m = \min\{\ell(g), \ell(h)\}$.

The only assertion which requires comment is (3). To prove this, we can assume without loss of generality (using (1) and (2)) that $\ell(g) \geq \ell(h)$. If $\ell(h) = 0$, then $\ell(gh) = \ell(g) = \ell(gh^{-1})$, so both sides of the equation in (3) are 0. For $\ell(gh) \leq \ell(g) + \ell(h) = \ell(g)$, and $\ell(g) = \ell(gh.h^{-1}) \leq \ell(gh) + \ell(h^{-1}) = \ell(gh)$, hence $\ell(gh) = \ell(g)$, and we can replace h by h^{-1} to see that $\ell(gh^{-1}) = \ell(g)$.

Suppose $\ell(h) > 0$. Then by Axiom III and the fact that ℓ is abelian,

$$\max\{\beta(g,h), \beta(g,h^{-1})\} = 0.$$

It therefore suffices to show that, if $\ell(gh^{-1}) \leq \ell(gh)$, then $\ell(gh^{-1}) = \ell(g) - \ell(h)$. Assuming this, $\beta(g,h) = 0$, that is, $\ell(gh) = \ell(g) + \ell(h)$. Now $\ell(gh^{-1}) \geq \ell(g) - \ell(h)$ since ℓ is abelian. Suppose $\ell(gh^{-1}) > \ell(g) - \ell(h)$. Then $\ell(gh)$, $\ell(gh^{-1}) > 0$, so by Axiom III and assumption that ℓ is abelian,

$$\begin{aligned}\max\{\ell(ghgh^{-1}), \ell(gh^2g^{-1})\} &= \ell(gh) + \ell(gh^{-1}) \\ &> (\ell(g) + \ell(h)) + (\ell(g) - \ell(h)) \\ &= 2\ell(g).\end{aligned}$$

However, by Lemma 6.4 and property (2), $\ell(gh^2g^{-1}) = 2\ell(h) \leq 2\ell(g)$, and since ℓ is abelian, $\ell(ghgh^{-1}) \leq \ell(g) + \ell(hgh^{-1}) = 2\ell(g)$. This contradiction establishes (3).

By 1.4 in [2], there is a homomorphism $h : G \to \Lambda$ with $\ell(g) = |h(g)|$ for all $g \in G$. This gives an action of G on Λ as translations, with ℓ as the hyperbolic length function. □

We can now prove one of the major results in [42].

Theorem 6.6. *If G is a finitely generated group, the space* PLF(G) *is compact.*

Proof. Take a sequence of points (α_n) in PLF(G). By 6.3, there is a subsequence converging to a point α in ΨLF(G). Let $p : [0,\infty)^G \setminus \{0\} \to \mathrm{P}(G)$ be the projection map, and choose a pseudo-length function ℓ such that $p(\ell) = \alpha$. Since p is open and there is a countable basis for the neighbourhoods of ℓ in $[0, \infty)^G \setminus \{0\}$, we can obtain from the (α_n) a sequence ℓ_m of hyperbolic length functions converging to ℓ. Let (X_m, d_m) be an $\mathbb{R}$-tree on which G acts with hyperbolic length function ℓ_m. We show ℓ is also a hyperbolic length function. By Lemma 6.5, we may assume ℓ is non-abelian. This means that there are elements g, $h \in G$ with $\ell(gh) > \ell(g) + \ell(h)$. For sufficiently large m, it follows that $\ell_m(gh) > \ell_m(g) + \ell_m(h)$. By Corollary 3.2.4, the characteristic sets of g, h for the action of G on X_m do not intersect. Let $[p_m, q_m]$ be the bridge joining these two characteristic sets, with p_m in the characteristic set of g. Arguing as in Theorem 3.4.1, the Lyndon length function L_{p_m} can be expressed in terms of ℓ_m, using linear combinations and max, and the expression is the same for each m. It follows that the sequence (L_{p_m}) converges in $\mathbb{R}^G$ to a point L, so L is a function from G to $\mathbb{R}$. Further, it is easy to see by inspection of the axioms that the set of real-valued Lyndon length functions on G is a closed subset of $\mathbb{R}^G$, so L is a Lyndon length function.

By Lemma 3.4.6, there are an $\mathbb{R}$-tree (X, d) on which G acts, and a point $p \in X$ such that $L = L_p$. By Lemma 3.1.8, $\ell_m(g) = \max\{L_{p_m}(g^2) - L_{p_m}(g), 0\}$, and taking limits gives $\ell(g) = \max\{L_p(g^2) - L_p(g), 0\}$. Again by Lemma 3.4.6, ℓ is the hyperbolic length function for the action of G on X, as required. □

The compactification achieved by Morgan and Shalen of the space in the introduction was by means of elements of PLF(G) arising from what are called small actions.

Definition. A non-trivial action of a group G on a Λ-tree is called small if the stabilizer of every segment $[x, y]$ with $x \neq y$ contains no free subgroup of rank 2. The space SLF(G) is the subspace of PLF(G) consisting of the images of hyperbolic length functions of small actions of G on $\mathbb{R}$-trees.

It is therefore of some interest to know that SLF(G) is also compact. Following [42] we shall prove this in two stages. The first stage requires some preliminary remarks. We call a segment $[x, y]$ with $x \neq y$ a *non-degenerate* segment.

Remark 1. Let G have a non-trivial action without inversions on a Λ-tree (X, d) and let K be a normal subgroup of G whose action by restriction on (X, d) is trivial. Then any finitely generated subgroup of K fixes a non-degenerate segment in X.

For suppose H is a finitely generated subgroup of K. Choose $g \in G$ with g hyperbolic. Then $\langle H, gHg^{-1}\rangle$ is also a finitely generated subgroup of K, so has a fixed point, say p, by Lemma 2.1. Then H fixes $[p, g^{-1}p]$, and $p \neq g^{-1}p$ since g is hyperbolic.

Remark 2. The group Isom(ℝ) of metric automorphisms of ℝ contains no free subgroup of rank 2.

For Isom(ℝ) is a semidirect product $\mathbb{R} \rtimes C_2$, where C_2 is cyclic of order 2, by Lemma 1.2.1, so is soluble.

Lemma 6.7. *Let G be a finitely generated group. If there is a reducible action of G with hyperbolic length function ℓ, such that the projective class of ℓ is in the closure of SLF(G), then G contains no free subgroup of rank 2. Consequently, SLF(G)=PLF(G).*

Proof. Suppose firstly that G has a small, reducible action on an ℝ-tree, with hyperbolic length function ℓ. By Props. 3.2.7 and 3.2.9, ℓ is induced from a homomorphism ρ from G to Isom(ℝ). The kernel, K, of ρ has trivial action, so any finitely generated subgroup of K fixes a non-degenerate segment by Remark 1 above. Hence K has no free subgroup of rank 2. Suppose F is a free subgroup of G of rank 2. By Remark 2 above, ρ restricted to F has a non-trivial kernel $F \cap K$, which is therefore non-cyclic, being normal in F. But then $F \cap K$, and so K, contains a free subgroup of rank 2, a contradiction.

Now suppose G has no small, reducible action on an ℝ-tree. Since PLF(G) is compact, the closure of SLF(G) is contained in PLF(G). The proof is completed by showing every hyperbolic length function ℓ whose projective class is a limit point of SLF(G) is the length function of an irreducible action. Given such a length function ℓ, we can find (as in Theorem 6.6) a sequence (ℓ_m) of hyperbolic length functions converging to ℓ in $\mathbb{R}^G$, where each ℓ_m corresponds to a small, irreducible action on an ℝ-tree (X_m, d_m). We denote the characteristic set of $g \in G$ for the action on X_m by A_g^m.

Choose $g \in G$ with $\ell(g) > 0$, and choose M with $\ell_m(g) > 0$ for all $m \geq M$. By Lemma 3.2.11, there is a hyperbolic element h of G such that $A_g^M \cap A_h^M = \emptyset$. Put $k = hgh^{-1}$, so $\ell_m(g) = \ell_m(k) > 0$ for all $m \geq M$ by Lemma 3.1.7. Now if $[p, q]$ is the bridge between A_g^M and A_h^M, with $p \in A_g$, then $[hp, hq]$ is the bridge between A_h^M and $A_k^M = hA_g^M$, and $hp \neq p$. It follows that if $a \in A_g$, $b \in A_h$, then $[a, b] = [a, p, q, hq, hp, b]$ by 2.1.5. Thus $A_g^M \cap A_k^M = \emptyset$, and the bridge joining them is $[p, hp]$. By Lemma 3.3.8, g, k freely generate a free group in G.

It follows that, for $m \geq M$, either $A_g^m \cap A_k^m = \emptyset$, or $A_g^m \cap A_k^m$ is a segment of length Δ_m, say, with $\Delta_m < 4\ell_m(g) = 4\ell_m(k)$. For otherwise, suppose $[x, y]$ is a segment of length $4\ell_m(g)$ in $A_g^m \cap A_k^m$, with $x \leq_g y$. If g, k meet coherently then the commutators $[g, k]$ and $[g^2, k]$ fix the segment $[x, z]$, where $z \in [x, y]$ is the point with $d_m(x, z) = \ell(g) > 0$. But $[g, k]$, $[g^2, k]$ freely generate a subgroup of the free group $\langle g, k\rangle$, contradicting the assumption that the action on X_m is small. Similarly, if g, k meet incoherently, $\langle [g, k^{-1}], [g^2, k^{-1}]\rangle$ is a free group fixing a non-degenerate segment, a contradiction.

On replacing X_m by a subsequence, and replacing k by k^{-1} if necessary, we can assume that either $A_g^m \cap A_k^m = \emptyset$ for all $m \geq M$, in which case we put $\Delta_m = 0$, or g, k meet coherently in a segment of length $\Delta_m < 4\ell_m(g)$ for $m \geq M$. Since $A_{g^4}^m = A_g^m$ and $A_{k^4}^m = A_k^m$, it follows from Lemma 3.3.9 and the remark following it that, for $m \geq M$, $\ell_m(g^4k^4g^{-4}k^{-4}) \geq 2\ell_m(g^4) + 2\ell_m(k^4) - 2\Delta_m \geq 8\ell_m(g)$. On taking limits, $\ell(g^4k^4g^{-4}k^{-4}) \geq 8\ell(g) > 0$. By Proposition 3.3.7, the action of G on an $\mathbb{R}$-tree corresponding to ℓ is irreducible, completing the proof. □

Before proceeding to the second and final stage in proving PSL(G) is compact, we need a simple technical lemma.

Lemma 6.8. *Let x_1, x_2 be points of a Λ-tree (X, d), and let g_1, g_2 be isometries of X, not inversions, such that $d(x_i, g_jx_i) < \epsilon$, for $1 \leq i, j \leq 2$, where $\epsilon \in \Lambda$. If $d(p_1, p_2) > 2\epsilon$, then $A_{g_1} \cap A_{g_2}$ contains a segment σ of length at least $d(p_1, p_2) - 2\epsilon$, with $\sigma \subseteq [x_1, x_2]$.*

Proof. Let $B_i = \{x \in X \mid d(x_i, x) \leq \epsilon\}$, for $i = 1, 2$. We leave it as an exercise to verify that these are closed subtrees of X, and they are clearly disjoint. Let σ be the bridge between them. By 3.1.1 and 3.1.4, $2d(x_i, A_{g_j}) = d(x_i, g_jx_i) - \ell(g_j) < \epsilon$, so $B_i \cap A_{g_j} \neq \emptyset$, i.e. A_{g_j} contains a point of B_1 and of B_2, so contains σ, for $j = 1, 2$. Thus $\sigma \subseteq A_{g_1} \cap A_{g_2}$, and the length of σ is $d(p_1, p_2) - 2\epsilon$, since $\sigma \subseteq [x_1, x_2]$. □

Theorem 6.9. *If G is a finitely generated group, then SLF(G) is a compact subspace of $P(G)$.*

Proof. By Theorem 6.6, it suffices to show SLF(G) is a closed subspace of PLF(G). Let ℓ be a hyperbolic length function whose projective class is a limit point of SLF(G). As in Lemma 6.7, there is a sequence (ℓ_m) of hyperbolic length functions converging to ℓ in $\mathbb{R}^G$, where each ℓ_m corresponds to a small action on an $\mathbb{R}$-tree (X_m, d_m) and ℓ corresponds to an action on an $\mathbb{R}$-tree (X, d). By Lemma 6.7, we can assume the action on X is irreducible. Further, by Theorem 3.4.2(c), we can assume the action on X and every X_m is minimal. We assume there is a non-degenerate segment σ in X which is stabilised by a free subgroup of rank 2 in G, and obtain a contradiction.

A subgroup of index at most two in G fixes the endpoints of σ, so there is a rank 2 free subgroup of G which fixes the endpoints of σ, with basis a, b say. Since the action on X is irreducible, there exist hyperbolic isometries s, u in G such that $\ell(su) > \ell(s) + \ell(u)$. For sufficiently large m, it follows that $\ell_m(su) > \ell_m(s) + \ell_m(u)$. Arguing as in the proof of Theorem 6.6, we find points $p_m \in X_m$ and $p \in X$ such that the Lyndon length functions L_{p_m} converge in $\mathbb{R}^G$ to L_p (p is the endpoint of the bridge joining A_s and A_u in X which lies in A_s; L_p can be expressed in terms of ℓ in the same way that L_{p_m} can be expressed in terms of ℓ_m). Now $X_m = \bigcup_{g\in G}[p_m, gp_m]$, because the right-hand side is a G-invariant subtree of X_m (see the remark preceding Theorem 3.4.1), and the action is minimal. Similarly, $X = \bigcup_{g\in G}[p, gp]$. For $g \in G$ and $t \in [0,1]_{\mathbb{R}}$, let $\langle g,t\rangle_m$ denote the point in $[p_m, gp_m]$ at distance $td_m(p_m, gp_m)$ from p_m, and $\langle g,t\rangle$ the point in $[p, gp]$ at distance $td(p, gp)$ from p. By 2.1.2(2), $d(\langle g,t\rangle, \langle h,r\rangle) = tL_p(g) + rL_p(h) - 2\min\{tL_p(g), rL_p(h), d(p,w)\}$, where $w = \mathrm{Y}(p, gp, hp)$. Further, as noted after Lemma 2.1.2, $d(p,w) = (gp\cdot hp)_p = \frac{1}{2}(L_p(g) + L_p(h) - L_p(g^{-1}h))$, since $d(gp, hp) = d(p, g^{-1}hp)$. Thus $d(\langle g,t\rangle, \langle h,r\rangle)$ can be expressed in terms of L_p, and there is a similar expression for $d_m(\langle g,t\rangle_m, \langle h,r\rangle_m)$ in terms of L_{p_m}. It follows that

$$\lim_{m\to\infty} d_m(\langle g,t\rangle_m, \langle h,r\rangle_m) = d(\langle g,t\rangle, \langle h,r\rangle). \tag{1}$$

Let e, f be the endpoints of σ, and write $e = \langle g_1, t_1\rangle$, $f = \langle g_2, t_2\rangle$. Let $e_m = \langle g_1, t_1\rangle_m$, $f_m = \langle g_2, t_2\rangle_m$ and put $\sigma_m = [e_m, f_m]$. By (1), $\lim\limits_{m\to\infty} \mathrm{length}(\sigma_m) = \mathrm{length}(\sigma) > 0$. Also, $\lim\limits_{m\to\infty} \ell_m(a) = \ell(a) = 0$, since a fixes e and f, and similarly $\lim\limits_{m\to\infty} \ell_m(b) = \ell(b) = 0$.

Now by considering the subtree of X spanned by p, gp and $h^{-1}p$, and applying the isometry h to it, we find that $h\langle g,t\rangle = \langle h, 1-t\rangle$ if $tL_p(g) \le d(p,v)$, where $v = \mathrm{Y}(p, gp, h^{-1}p)$, and $h\langle g,t\rangle = \langle hg, 1-(1-t)(L_p(g)/L_p(hg))\rangle$ if $tL_p(g) > d(p,v)$. (The last condition ensures that $hgp \ne p$.) Further, $d(p,v)$ can be expressed in terms of L_p in a similar manner to $d(p,w)$ and, putting $v_m = \mathrm{Y}(p_m, gp_m, h^{-1}p_m)$, $d_m(p_m, v_m)$ can be similarly expressed in terms of L_{p_m}. Therefore, $\lim\limits_{m\to\infty} d_m(p_m, v_m) = d(p,v)$. Also, $\lim\limits_{m\to\infty} tL_{p_m}(g) = tL_p(g)$, so if $tL_p(g) < d(p,v)$, then $tL_{p_m}(g) < d_m(p_m, v_m)$ for sufficiently large m and, if $tL_p(g) > d(p,v)$, then $tL_{p_m}(g) > d_m(p_m, v_m)$ for sufficiently large m. If $tL_p(g) = d(p,v)$ then $\langle h, 1-t\rangle = \langle hg, 1-(1-t)(L_p(g)/L_p(hg))\rangle = hv$. We conclude from (1) that

$$\lim_{m\to\infty} d_m(\langle g,t\rangle_m, h\langle g,t\rangle_m) = d(\langle g,t\rangle, h\langle g,t\rangle). \tag{2}$$

In particular, $\lim_{m\to\infty} d_m(e_m, ae_m) = 0 = \lim_{m\to\infty} d_m(f_m, af_m)$, and the same is true with a replaced by b. By Lemma 6.8, for sufficiently large m, $A_g \cap A_h$ contains the middle third of σ_m. Arguing as in the proof of 6.7, it follows that, for sufficiently large m, one of $\langle [a,b], [a^2,b]\rangle$, $\langle [a,b^{-1}], [a^2,b^{-1}]\rangle$ fixes a non-degenerate interval contained in σ_m. But these are free groups of rank 2, and the action on X_m is small, giving the desired contradiction. □

We denote by $\mathrm{CV}(G)$ the subspace of $\mathrm{PLF}(G)$ arising from free actions of G on ℝ-trees which are polyhedral (see Chapter 2, §2). This is clearly a subspace of $\mathrm{SLF}(G)$. The space $\mathrm{CV}(F_n)$, where F_n denotes the free group of rank n, is called Outer Space of rank n. This space was studied by Culler and Vogtmann [43], who showed that outer space is triangulable, contractible and the outer automorphism group $\mathrm{Out}(F_n)$ acts properly discontinuously on it, which explains the terminology. (They gave a different but equivalent definition of outer space in terms of marked graphs.) This leads to important information on $\mathrm{Out}(F_n)$, including its virtual cohomological dimension (it is $2n - 3$, for $n \geq 2$).

There have been some interesting recent developments concerning spaces of actions, which we shall not discuss in detail. Notably, there is now a description of the closure $\overline{\mathrm{CV}(F_n)}$ in $\mathrm{P}(F_n)$. This involves the notion of a very small action.

Definition. An action of a group G on a Λ-tree is called very small if it is small and for all $1 \neq g \in G$, X^g is linear and $X^g = X^{g^n}$ for all integers n such that $g^n \neq 1$; VSLF(G) is the subspace of PLF(G) arising from very small actions.

It has been shown by Bestvina and Feighn [14], building on work of Cohen and Lustig [40], that the closure of Outer Space in $\mathrm{SLF}(F_n)$ is $\mathrm{VSLF}(F_n)$. They showed that for $n \geq 2$, this closure has dimension $3n - 4$, and it was shown by Gaboriau and Levitt [51] that the boundary of outer space has dimension $3n - 5$. In the case $n = 2$, the boundary of Outer Space has been explicitly described in [44]. Very recently, Bestvina and Feighn [15] have constructed an analogue, for Outer Space, of the Borel-Serre bordification of a symmetric space. They make use of this to show that, for $n \geq 2$, $\mathrm{Out}(F_n)$ is $(2n - 5)$-connected at infinity (we refer to their paper for the definition) and is a virtual duality group of dimension $(2n - 3)$.

There has also been some interesting work on the contractibility of spaces related to outer space, including its closure. See [88], [133], [138] and [146]. In particular, Skora's paper [133] contains a uniform method for showing $\mathrm{CV}(F_n)$, $\overline{\mathrm{CV}(F_n)} = \mathrm{VSLF}(F_n)$, $\mathrm{SLF}(F_n)$ and $\mathrm{PLF}(F_n)$ are contractible.

Outer space was invented by Culler and Vogtmann to study the outer automorphism group of a free group. There have been several recent papers which use $\mathbb{R}$-trees to study automorphisms of a free group, for example, [16], [52] and [55]. In addition there has been some work on actions of word hyperbolic (negatively curved) groups on $\mathbb{R}$-trees, including [117] and [126].

We close the chapter by mentioning some questions raised by Shalen in [128].

Question 1. Let G be a finitely generated group. Does SLF(G) have a dense subspace consisting of projectivised integer-valued hyperbolic length functions?

As motivation for this question, we sketch a proof that the answer is yes when $G = \pi_1(S)$, where S is a compact hyperbolic surface. Call an action of G on an $\mathbb{R}$-tree *geometric* if it is the action on an $\mathbb{R}$-tree dual to an element of

ML(S) (Lemma 5.13). Skora ([131], [132]) has shown that an action of G on an $\mathbb{R}$-tree is small if and only if it is geometric, using a result proved by Hatcher [74] and by Morgan and Otal [105]. (Skora also notes that the small subgroups of $\pi_1(S)$ are precisely the cyclic subgroups.) An element (C, μ) of ML(S) is called *discrete* if μ is obtained by the procedure described after the proof of Lemma 4.11, using a finite set $\{\mu_i \mid i \in I\}$ of positive real numbers. Skora [131] also observes that the discrete elements of ML(S) are dense in ML(S), by work of Thurston [139], and this implies that the set of projectivised length functions arising from geometric actions corresponding to discrete elements of ML(S) is dense in SLF(G).

Let (C, μ) be a discrete element of ML(S) corresponding to the finite set $\{\mu_i \mid i \in I\}$ and let ε be a positive real number. Choose positive rational numbers r_i such that $|\mu_i - r_i| < \varepsilon$, and let (C, μ') be the discrete element of ML(S) obtained by replacing each μ_i by r_i. Let the dual $\mathbb{R}$-trees for (C, μ) and (C, μ') be (T, d) and (T', d') respectively, with respective hyperbolic length functions ℓ and ℓ'. These are obtained from metric spaces (X, d), (X, d') by the construction of Theorem 2.4.4, where X is the set of components of $\mathbb{H}^2 \setminus \tilde{C}$ and $\tilde{C}$ is the lift of C to $\mathbb{H}^2$. Let x, $y \in X$; one can check that there are integers n_i such that $d(x, y) = \sum_{i \in I} n_i \mu_i$ and $d'(x, y) = \sum_{i \in I} n_i r_i$. In particular, fixing x, if $g \in G$ this is true for $y = gx$. It follows by Lemma 3.1.8 that, if N is a neighbourhood of ℓ in $[0, \infty)^G$, then by making ε sufficiently small, we can ensure that $\ell' \in N$.

Hence it suffices to show that the equivalence class of ℓ' in SLF(G) contains an integer-valued length function. But from the expression for $d'(x, y)$ just given, (X, d') is a $\frac{1}{m}\mathbb{Z}$-metric space, where m is the product over $i \in I$ of the denominators of the r_i, hence (T, d') is a $\frac{1}{m}\mathbb{Z}$-tree. Therefore $m\ell'$ is the required integer-valued length function.

It should be noted that other notions of geometric action have been used in the literature, based on different notions of dual $\mathbb{R}$-tree. In particular, Levitt and Paulin [91] use a definition based on the dual tree to a measured foliation on a complex, as constructed in [63]. They prove several results on a geometric action in this sense, and give a criterion for an action to be geometric.

We can also ask the analogue of Question 1 for PLF(G).

Question 2. Let G be a finitely generated group. Does PLF(G) have a dense subspace consisting of projectivised integer-valued hyperbolic length functions?

Progress has been made on Questions 1 and 2 in [64], [51] and more recently in [69], where the main result in [14] is generalised. In [129], 2.5.5, Shalen conjectures that the answer to Question 1 is yes when the small subgroups of G are finitely generated, and gives a stronger conjecture in 5.5.6.

Note that, if ℓ is a hyperbolic length function representing an element of PLF(G) which is integer-valued, then 2ℓ is the hyperbolic length function for an action of G on a $\mathbb{Z}$-tree. For ℓ is the hyperbolic length function for an action of G on an $\mathbb{R}$-tree, say (X,d). If $x \in X$, the Lyndon length function L_x is integer-valued (see the argument for Theorem 3.4.1), so by Theorem 2.4.6 there is a $\mathbb{Z}$-tree (Y,d') on which G acts, whose Lyndon length function with respect to some point is $2L_x$. It follows by Lemma 3.1.8 that the hyperbolic length function for this action is 2ℓ.

Theorem 2.4.6 also gives an $\mathbb{R}$-tree on which G acts, whose Lyndon length function with respect to some point is $2L_x$, and it is clear from the construction that this $\mathbb{R}$-tree contains Y. Arguing as in Theorem 3.4.1, this $\mathbb{R}$-tree is isomorphic to the minimal invariant subtree of $(X,2d)$ (which exists by Theorem 3.4.2). It follows easily, using Lemma 1.2.2(2), that if the action of G on (X,d) is small, so is its action on (Y,d').

Thus the following are weakened versions of Questions 1 and 2.

Question 1′. Let G be a finitely generated group having a non-trivial small action on an $\mathbb{R}$-tree. Does G have a non-trivial small action on a $\mathbb{Z}$-tree?

Question 2′. Let G be a finitely generated group having a non-trivial action on an $\mathbb{R}$-tree. Does G have a non-trivial action on a $\mathbb{Z}$-tree?

A positive answer to Question 1 (resp. 2) clearly implies a positive answer to Question 1′ (resp. 2′). In view of the Bass-Serre theory ([39],[127]), to say that G has a non-trivial action on a $\mathbb{Z}$-tree is equivalent to saying that it decomposes either as a non-trivial free product with amalgamation or as an HNN-extension. To say that G has a non-trivial small action on a $\mathbb{Z}$-tree is equivalent to saying that it decomposes either as a non-trivial free product with amalgamation with small amalgamated subgroup, or as an HNN-extension with the associated pair of subgroups small. (For an explanation of these terms see [39; Chapter 1].) Results on the existence of such decompositions are contained in [13], [118] and [126], based on the argument we shall give in Chapter 6 concerning free actions on $\mathbb{R}$-trees.

CHAPTER 5. Free Actions

1. Introduction

We return to the question: what group-theoretic information can be obtained from an action of a group on a Λ-tree, analogous to the Bass-Serre theory ([127; Ch.I], [39; Ch.8]) in the case $\Lambda = \mathbb{Z}$? In fact, given the apparent difficulty of this problem, one should probably ask the abbreviated question: what information on the group can be obtained from an action on a Λ-tree? The simplest case of the Bass-Serre theory is when the action is free (and without inversions), and this suggests that the study of free actions on Λ-trees may be simpler than the study of actions on Λ-trees in general.

Recall that an action of a group G on a Λ-tree is free and without inversions if, for all $1 \neq g \in G$, g acts as a hyperbolic isometry (see the observations preceding Lemma 3.3.6). Equivalently, all non-identity elements of G have positive hyperbolic length. This is the opposite extreme to a trivial action. If N denotes the set of non-hyperbolic elements of G, then trivial means $N = G$, while free and without inversions means $N = 1$. If Λ is an ordered abelian group, we call a group Λ-free if it has a free action without inversions on some Λ-tree, and a group is called tree-free if it is Λ-free for some Λ. Note that any subgroup of a Λ-free group is Λ-free. Also, if a group G acts freely without inversions on a Λ-tree (X, d) and Λ embeds in an ordered abelian group Λ', then G acts freely without inversions on the Λ'-tree $\Lambda' \otimes_\Lambda X$ by Lemma 3.2.1. One other useful observation is that tree-free groups are torsion-free, by Lemma 3.1.7(3).

Remark. Any ordered abelian group Λ is Λ-free, because Λ acts on itself as translations. It follows that any torsion-free abelian group is tree-free, for such a group can be made into an ordered abelian group (see [58; Lemma 2]). This applies to free abelian groups. Free abelian groups of rank at most $2^{\aleph_0}$ are $\mathbb{R}$-free since they are isomorphic to subgroups of $\mathbb{R}$.

It is a consequence of the Bass-Serre theory that a group is $\mathbb{Z}$-free if and only if it is a free group (see [127; Ch.I, §3]). In fact it is easy to see, without recourse to the general theory, that a free group acts freely on its Cayley graph with respect to a basis, which is a simplicial tree, giving a free action without inversions on the corresponding $\mathbb{Z}$-tree. (See §8.1, particularly Lemma 1, and also Theorem 5, Chapter 7 in [39].) It is also a consequence of Prop. 1.1 below that free groups, being free products of infinite cyclic groups, are $\mathbb{Z}$-free.

We shall give further examples in the course of this chapter. We shall summarise a proof that the fundamental groups of compact surfaces, with three exceptions, are $\mathbb{R}$-free. We shall also show that non-standard free groups, constructed as ultrapowers of free groups, are tree-free, hence so are their subgroups, and we shall give a group-theoretic characterisation of such subgroups. Before this, however, we shall prove some general results on tree-free groups. The most important of these is Harrison's Theorem, characterising tree-free groups with at most two generators. We begin with two simple general results. The first depends on another argument of Harrison [72; Proposition 1.2] on Lyndon length functions, together with the connection between Lyndon length functions and actions on Λ-trees established in §2.4. (The argument in [72] is given for $\Lambda = \mathbb{R}$ and finitely many free factors, but works in general.)

Proposition 1.1. *If $\{G_i \mid i \in I\}$ is a collection of Λ-free groups then the free product $*_{i\in I} G_i$ is Λ-free.*

Proof. Let $G = *_{i\in I} G_i$. For $i \in I$, let (X_i, d_i) be a Λ-tree on which G_i acts freely without inversions. Choose a basepoint x_i in each X_i, and let $L_i = L_{x_i}$ be the corresponding Lyndon length function (see §2.4). We define a mapping $L : G \to \Lambda$ by letting $L(1) = 0$, and if $1 \neq g \in G$, let $g = g_1 \ldots g_n$ be the expression of g in normal form (so $g_j \in G_{i_j}$ for some $i_j \in I$, and $i_j \neq i_{j+1}$ for $1 \leq j \leq n-1$), and set $L(g) = \sum_{j=1}^{n} L_{i_j}(g_j)$. We shall show that L is a Lyndon length function. Axiom (1) for a Lyndon length function is satisfied by definition, and if g is expressed as above in normal form, then the normal form of g^{-1} is $g^{-1} = g_n^{-1} \ldots g_1^{-1}$, hence Axiom (2) ($L(g) = L(g^{-1})$ for all $g \in G$) is satisfied.

Note that L restricted to G_i is L_i, so we can omit the subscript on L_i. We next observe that L satisfies the triangle inequality, that is, $L(gh) \leq L(g)+L(h)$ for all g, $h \in G$. This is clear if one of g, h is 1, otherwise we have expressions in normal form: $g = g_n \ldots g_1$, $h = h_1 \ldots h_m$. Let p be the least integer $j \geq 1$ such that $g_j \neq h_j^{-1}$. Then we can write $gh = g_n \ldots g_p h_p \ldots h_m$ (note that $p = n+1$ is allowed, when $gh = h_{n+1} \ldots h_m$, as is $p = m+1$, and both can happen, in which case $gh = 1$). There are two cases.

Case 1. Either g_p, h_p are in different free factors, or $p = n+1$, or $p = m+1$. In this case, $L(gh) = \sum_{j=p}^{n} L(g_j) + \sum_{j=p}^{m} L(h_j) \leq L(g) + L(h)$. (If $p > n$ then $\sum_{j=p}^{n} L(g_j)$ is to be interpreted as 0; similar conventions are used throughout the proof.)

Case 2. Otherwise, g_p, h_p are in the same free factor G_i, but $g_p \neq h_p^{-1}$. In this case, $L(gh) = \sum_{j=p+1}^{n} L(g_j) + \sum_{j=p+1}^{m} L(h_j) + L(g_p h_p)$, and $L(g_p h_p) \leq L(g_p) + L(h_p)$ since $L|_{G_i} = L_i$ satisfies the triangle inequality (Property (5) of a Lyndon length function). Thus we again obtain $L(gh) \leq L(g) + L(h)$.

We have to verify Axiom (3) for a Lyndon length function, that is, for all g, h and $k \in G$,

$$c(g,h) \geq \min\{c(g,k), c(h,k)\}$$

where $c(g,h) = \frac{1}{2}(L(g) + L(h) - L(g^{-1}h))$. (Note that, by Axiom (2), c is symmetric, that is, $c(g,h) = c(h,g)$.) This inequality is clear if one of g, h, k is 1, otherwise we can write normal forms

$$g = g_1 \dots g_m$$
$$h = h_1 \dots h_n$$
$$k = k_1 \dots k_p.$$

Let q be the least integer $j \geq 1$ such that $k_j \neq g_j$, and let r be the least integer $j \geq 1$ such that $k_j \neq h_j$. We can assume without loss of generality that $q \leq r$.

Assume first that k^{-1}, g fall in Case 1 above. Then it is easily seen that $c(k,g) = \sum_{j=1}^{q-1} L(k_j)$. Now $k_j = g_j$ for $j < q$, and $k_j = h_j$ for $j < r$, so $g_j = h_j$ for $j < q$. Hence, using the properties of L already established,

$$L(g^{-1}h) = L(g_m^{-1} \dots g_q^{-1}h_q \dots h_n) \leq \sum_{j=q}^{m} L(g_j) + \sum_{j=q}^{n} L(h_j)$$

from which we obtain $c(g,h) \geq \frac{1}{2}(\sum_{j=1}^{q-1} L(g_j) + \sum_{j=1}^{q-1} L(h_j)) = \sum_{j=1}^{q-1} L(k_j) = c(k,g)$.

Now assume k^{-1}, g fall in Case 2 above. Then

$$L(k^{-1}g) = \sum_{j=q+1}^{p} L(k_j) + \sum_{j=q+1}^{m} L(g_j) + L(k_q^{-1}g_q)$$

from which we obtain $c(k,g) = \sum_{j=1}^{q-1} L(k_j) + c(k_q, g_q)$.

If $q < r$ then $k_q = h_q$ but $k_q \neq g_q$, so g_q, h_q are in the same free factor but $g_q \neq h_q$, hence

$$c(g,h) = \sum_{j=1}^{q-1} L(g_j) + c(g_q, h_q) = \sum_{j=1}^{q-1} L(k_j) + c(g_q, k_q) = c(g,k).$$

Assume $q = r$. If k^{-1} and h fall in Case 1 above, then so do g^{-1} and h, hence $c(g,h) = \sum_{j=1}^{q-1} L(g_j) = \sum_{j=1}^{q-1} L(k_j) = c(h,k)$. If k^{-1}, h fall in Case 2 above, then $c(g,k) = c(h,k) = \sum_{j=1}^{q-1} L(k_j) + c(k_q, h_q)$. Also,

$$L(g^{-1}h) = L(g_m^{-1} \dots g_q^{-1}h_q \dots h_n) \leq \sum_{j=q-1}^{m} L(g_j) + \sum_{j=q-1}^{n} L(h_j) + L(g_q^{-1}h_q).$$

Hence

$$c(g,h) \geq \frac{1}{2}(\sum_{j=1}^{q} L(g_j) + \sum_{j=1}^{q} L(h_j) - L(g_q^{-1}h_q)) = \sum_{j=1}^{q-1} L(k_j) + c(g_q, h_q).$$

Now since g_q, h_q, k_q are all in the same free factor, and L restricted to this factor is a Lyndon length function, $c(g_q, h_q) \geq \min\{c(g_q, k_q), c(h_q, k_q)\}$, hence again $c(g,h) \geq \min\{c(g,k), c(h,k)\}$, and L is a Lyndon length function.

Since the action on each X_i is free and without inversions, $L(g^2) > L(g)$ and $L(g) \neq 0$, for all $1 \neq g \in G_i$ and all $i \in I$, by Lemma 3.1.8 and because x_i has trivial stabiliser. If $1 \neq g \in G$, we can write g in reduced form

$$g = h_1 \dots h_m(g_1 \dots g_n)h_m^{-1} \dots h_1^{-1},$$

where $n \geq 1$, and either $n = 1$ or $g_1 \neq g_n^{-1}$. It follows easily that $L(g^2) > L(g)$. Replacing L by $2L$ gives a Lyndon length function on G, satisfying $c(g,h) \in \Lambda$ for all g, $h \in G$, and still satisfying $L(g^2) > L(g)$ for all $1 \neq g \in G$. By Theorem 2.4.6, there is an action of G on a Λ-tree (X,d), such that $L = L_v$ for some $v \in X$. By Lemma 3.1.8, the action of G on X is free and without inversions. □

Lyndon called elements g satisfying $L(g^2) > L(g)$ archimedean, but having already used this term in Chapter 1 with two different meanings, we prefer to call them hyperbolic. Before stating our second general result, some definitions are needed.

Definition. A subgroup H of a group G is called malnormal (or conjugately separated) if $H \cap gHg^{-1} \neq \{1\}$ implies $g \in H$, for all $g \in G$. A group is called a CSA group if all its maximal abelian subgroups are malnormal.

The definition of CSA appears in [112] and [62].

Definition. A group G is commutative transitive if the following equivalent conditions are satisfied.

(1) Given a, b, $c \in G$ such that $[a,b] = 1$, $[b,c] = 1$ and $b \neq 1$, then $[a,c] = 1$.
(2) Centralisers of non-identity elements of G are abelian.
(3) If M_1, M_2 are maximal abelian subgroups of G and $M_1 \neq M_2$, then $M_1 \cap M_2 = \{1\}$.

We leave it as a simple exercise to show that the three conditions in the definition of commutative transitive are indeed equivalent, and that a CSA group is commutative transitive. The next result is part of 1.8 and 1.9 in [5].

Lemma 1.2. *A tree-free group is CSA and commutative transitive.*

Proof. Commutative transitivity follows at once from Corollary 3.3.10. Let G be a group acting freely without inversions on a Λ-tree (X,d). If g, $h \in$

$G \setminus \{1\}$, then g, h commute if and only if $A_g = A_h$. For if $A_g = A_h$, the commutator $[g,h]$ fixes this common axis pointwise, so $[g,h] = 1$ since the action is free. The converse follows from Lemma 3.3.9. Hence if M is a maximal abelian subgroup of G, all elements of $M \setminus \{1\}$ have the same axis, say A, and $M \setminus \{1\} = \{ g \in G \setminus \{1\} \mid A_g = A\}$. Suppose $M \cap gMg^{-1} \neq \{1\}$. We have to show $g \in M$, and we can assume $g \neq 1$. Let $h \in M \cap gMg^{-1}$, $h \neq 1$. Then $A_h = A = A_{g^{-1}hg} = g^{-1}A_h$ using Lemma 3.1.7, that is, $gA_h = A_h$. Thus g acts as an isometry on the linear tree A_h, and g^2 has no fixed points in A_h since g^2 acts as a hyperbolic isometry of X, again by Lemma 3.1.7. By Lemma 1.2.1, g acts as a translation on A_h, as does h. Again it follows that $[g,h]$ fixes A_h pointwise and so $[g,h] = 1$. By the previous observations, $A_g = A_h$ and $g \in M$, as required. □

In the proof, we did not need the exercise (CSA implies commutative transitive). The fact that Λ-free groups are commutative transitive was proved, in terms of Lyndon length functions and for $\Lambda = \mathbb{R}$, in 5.13 of [72].

2. Harrison's Theorem

We shall prove the following theorem, which was proved for the case $\Lambda = \mathbb{R}$ in terms of Lyndon length functions in [72]. The case of arbitrary Λ was treated in [32] and [144], and we shall draw on arguments from both. The terminology in the theorem is taken from §3.3, so that $\Delta(g,h)$ is the diameter of $A_g \cap A_h$, assuming this intersection is non-empty.

Theorem 2.1. *Let G be a group acting freely and without inversions on a Λ-tree X, and suppose $g, h \in G \setminus \{1\}$. Then $\langle g,h\rangle$ is either free of rank two or abelian. Moreover, $\langle g,h\rangle$ is abelian if and only if $A_g \cap A_h \neq \emptyset$ and $\Delta(g,h) \geq \ell(g) + \ell(h)$, if and only if $A_g = A_h$. (Here ℓ denotes, as usual, hyperbolic length.)*

Some of the work needed to prove this has already been done in 3.3.6 and 3.3.8, but more lemmas are needed to take care of the case that Λ is non-archimedean, asserting that under certain circumstances, two isometries g, h generate a free group. As might be expected from the statement of the theorem, they will all involve the assumption that $\Delta(g,h) < \ell(g) + \ell(h)$ and g, h act as hyperbolic isometries. Then (as at the beginning of §3.3) $A_g \cap A_h$ has a left and right-hand endpoint, denoted l, r respectively, and $A_g \cap A_h = [l,r]$. Further, we may take $\Delta(g,h)$ to be $d(l,r)$. As in §3.3, the arguments are made more transparent by drawing suitable pictures, indeed it is probably essential to do so, but mostly this will be left to the reader. The main lemma is 2.2 below. The idea is that if we take a reduced word in $g^{\pm 1}$, $h^{\pm 1}$ and successively apply the letters in this word to l, we obtain a sequence of points, and provided this sequence avoids l and r, once the sequence leaves $[l, r]$ it never returns to it. There is one case (2.5) where the sequence need not avoid l and r, which needs a separate argument.

Suppose, then, that a group G acts on a Λ-tree (X, d), that g, $h \in G$ both act as hyperbolic isometries, and $\Delta < \ell(g) + \ell(h)$, where $\Delta = \Delta(g, h)$. We also assume that g, h meet coherently, as defined in §3.3. Let W_g be the set of all elements of G represented by a reduced word in $g^{\pm 1}$, $h^{\pm 1}$ which begin with $g^{\pm 1}$, and similarly define W_h (we shall not distinguish such a word w from the element of G it represents, to simplify notation; it should be clear from the context what is meant). Now define $X_g(l)$ to be the set of all points wl in X, where $w \in W_g$, such that for all terminal segments v of w, $vl \in [l, r]$. In a similar manner, we can define $X_g(r)$, $X_h(l)$, $X_h(r)$. Further, define $X_g(l, r) = X_g(l) \cup X_g(r) \cup \{l, r\}$, and similarly define $X_h(l, r)$.

Lemma 2.2. *Under the assumptions of the previous paragraph, suppose that, for every non-zero integer n,*

$$h^n(X_g(l, r)) \cap \{l, r\} = g^n(X_h(l, r)) \cap \{l, r\} = \emptyset.$$

Then $\langle g, h \rangle$ is a free group of rank 2.

Proof. We begin by establishing the following observation.

(1) If $[x, y] \subseteq X$ is a segment with $y \in X_h(l, r)$ and $[x, y] \cap A_g = \{y\}$, then for every integer $k \neq 0$ there is a point $z \in X_g(l, r)$ with $[g^k x, z] \cap A_h = \{z\}$ and $d(g^k x, z) \geq d(x, y)$. In particular, if $x \neq y$, then $g^k x \notin [l, r]$

For if $g^k y \in [l, r]$ we may take $z = g^k y$. Clearly $d(g^k x, z) = d(x, y)$, and by hypothesis $g^k y \neq l$, r, so $[g^k y, g^k x] \cap A_h = \{g^k y\}$. If $g^k y \notin [l, r]$, let $[g^k y, z]$ be the bridge between $g^k y$ and $[l, r]$, so $z \in \{l, r\} \subseteq X_g(l, r)$. It is easy to see that $[g^k x, z] \cap A_h = \{z\}$, and

$$d(g^k x, z) = d(g^k x, g^k y) + d(g^k y, z) = d(x, y) + d(g^k y, z) > d(x, y).$$

Interchanging the roles of g, h we obtain

(2) If $[x, y] \subseteq X$ is a segment with $y \in X_g(l, r)$ and $[x, y] \cap A_h = \{y\}$, then for every integer $k \neq 0$ there is a point $z \in X_h(l, r)$ with $[h^k x, z] \cap A_g = \{z\}$ and $d(h^k x, z) \geq d(x, y)$. In particular, if $x \neq y$, then $h^k x \notin [l, r]$

To prove Lemma 2.2, suppose $w \in W_g \cup W_h$; it suffices to show that $wl \neq l$. If $wl \in X_g(l) \cup X_h(l)$ this follows by hypothesis. Otherwise, we can find a longest terminal subword w_1 of w such that for all terminal segments w' of w_1, $w'l \in [l, r]$. We can then write $w = ug^n w_1$ or $w = uh^n w_1$ for some non-zero integer n, with n as large as possible. We assume $w = uh^n w_1$ (the other case follows on interchanging g and h) and put $v = h^n w_1$, so $w = uv$, and the last letter of u is different from the first letter of v. It is easy to see that $vl \in A_h \setminus [l,r]$, and either $[vl,r] \cap A_g = \{r\}$ and $d(vl,r) > 0$, or $[vl,l] \cap A_g = \{l\}$ and $d(vl,l) > 0$. Writing u as a product of powers of g and h, and successively applying (1) and (2), it follows that $wl \notin [l,r]$. □

Corollary 2.3. *Let G act on a Λ-tree (X,d), let g, $h \in G$ act as hyperbolic isometries which meet coherently. Let $\Delta = \Delta(g,h)$ and assume $\Delta < \ell(g) + \ell(h)$. Suppose also that $\ell(g)$, $\ell(h)$ are $\mathbb{Z}$-linearly independent, and that $\Delta \notin \langle \ell(g), \ell(h) \rangle$. Then $\langle g,h \rangle$ is a free group of rank 2.*

Proof. Let $w \in W_g$ be such that $vl \in [l,r]$ for all terminal segments v of w, so that $wl \in X_g(l)$. Because of the condition $\Delta < \ell(g) + \ell(h)$, it follows that all the powers of g which occur in w have the same sign, and the powers of h which occur have the opposite sign. Hence $d(l,wl) = \alpha\ell(g) + \beta\ell(h)$, where α is the exponent sum of g in w, and β is the exponent sum of h in w, and $\alpha \neq 0$. The hypotheses imply $h^n wl \neq l$ and $h^n wl \neq r$, for all integers n. Similar considerations with $X_g(r)$, $X_h(l)$, $X_h(r)$ show that the hypotheses of Lemma 2.2 are satisfied, and the result follows. □

Corollary 2.4. *Let G act on a Λ-tree (X,d), let g, $h \in G$ act as hyperbolic isometries which meet coherently. Let $\Delta = \Delta(g,h)$ and assume $\ell(h^r) < \ell(g)$ and $\ell(h^r) < \Delta < \ell(g) + \ell(h)$, for all integers r. Assume also that $\Delta \neq \ell(g)$. Then $\langle g,h \rangle$ is a free group of rank 2.*

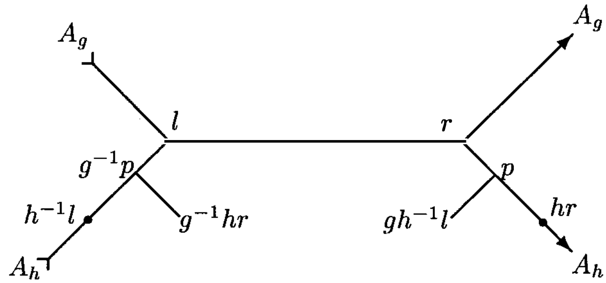
Proof. The only possible element of $W_g(l)$ is gl, and the only possible elements of $W_h(l)$ are $h^m l$ and $h^{-m} gl$, where m is positive, and similar comments apply to $W_g(r)$, $W_h(r)$. It follows that the hypotheses of Lemma 2.2 are satisfied, whence the result. □

We shall also need an analogue of Corollary 2.4 when $\ell(g) = \Delta$, which will be proved by an ad hoc argument.

Lemma 2.5. *Let G act on a Λ-tree (X,d), let g, $h \in G$ act as hyperbolic isometries which meet coherently. Let $\Delta = \Delta(g,h)$ and assume $\ell(h^r) < \ell(g)$ and $\ell(h^r) < \Delta < \ell(g) + \ell(h)$, for all integers r. Assume also that $\Delta = \ell(g)$ and $ghg^{-1}h$ acts as a hyperbolic isometry of X. Then $\langle g,h \rangle$ is a free group of rank 2.*

Proof. Let the bridge between $gh^{-1}l$ and A_h be $[p, gh^{-1}l]$.

Since $[l, h^{-1}l] \cap A_g = \{l\}$, we have $[r, gh^{-1}l] \cap A_g = \{r\}$, so $r \leq_h p$. Also, $d(gh^{-1}l, r) = d(h^{-1}l, l) = \ell(h) = d(r, hr)$. Hence $d(p,r) \leq \ell(h)$, in fact $p <_h hr$ and $p \neq gh^{-1}l$, otherwise $hr = gh^{-1}l = gh^{-1}g^{-1}r$ which implies $ghg^{-1}h$ is not hyperbolic, contrary to assumption. Since $[l, g^{-1}p, h^{-1}l] \subseteq A_h$, $[g^{-1}p, g^{-1}hr]$ is the bridge between $g^{-1}hr$ and A_h, and $g^{-1}hr \neq g^{-1}p$. The situation is illustrated below ($p = r$ is allowed, but as we have observed, $[gh^{-1}l, r]$ has no point in common with A_g except r).



For $x \in X$, let the bridge between x and A_g be $[x, x_g]$. We define the following subsets of X (the notation $<_g$ was introduced in §3.3):

$U = \{x \in X \mid r <_g x_g\}$
$V = \{x \in X \mid x_g <_g l\}$
$W = \{x \in X \mid [x, r] = [x, gh^{-1}l, p, r]\}$
$Z = \{x \in X \mid [x, l] = [x, g^{-1}hr, g^{-1}p, l]\}$.

Note that none of these sets contains r. If m, n are non-zero integers, we leave it to the reader to check that

$g^m h^n U \subseteq U \cup V \cup Z$
$g^m h^n V \subseteq U \cup V \cup W$
$g^m h^n W \subseteq U \cup V \cup Z$
$g^m h^n Z \subseteq U \cup V \cup W$.

The details are in effect contained in the proof of Lemma 6 in [32]. Further, $g^m h^n r \in U \cup V \cup Z$. It follows that if $w = g^{m_1} h^{n_1} \dots g^{m_k} h^{n_k}$, where the m_i and n_i are integers, none of which is zero, and $k \geq 1$, then $wr \neq r$. Hence $\langle g, h \rangle$ is the free product of the cyclic subgroups generated by g and h, which are infinite cyclic since g, h act as hyperbolic isometries. Thus $\langle g, h \rangle$ is free with basis g, h. □

Proof of Theorem 2.1. We consider several cases.

Case 1. $A_g \cap A_h = \emptyset$. By Lemma 3.3.8, $\langle g, h \rangle$ is a free group of rank 2.

In the other cases, we assume $A_g \cap A_h \neq \emptyset$ and put $\Delta = \Delta(g, h)$ (see §3.3).

Case 2. $\Delta < \min\{\ell(g), \ell(h)\}$. Then similarly using Lemma 3.3.6, g and h generate a free group of rank 2.

In the remaining cases, we can assume without loss of generality that $\ell(h) \leq \ell(g)$, and replacing h by h^{-1} if necessary, that g and h meet coherently.

Case 3. $\Delta = \ell(h) \leq \ell(g)$. Then $\Delta > 0$ (otherwise $h = 1$ since the action is free without inversions, contrary to assumption). Also, $\ell(ghg^{-1}h^{-1}) =$

$2\ell(g) > 0$ by Lemma 3.3.9, hence $\ell(gh^{-1}) \neq 0$ (otherwise $g = h$, which implies $\ell(ghg^{-1}h^{-1}) = \ell(1) = 0$, a contradiction). By Lemma 3.3.5, either $A_{gh^{-1}} \cap A_h = \emptyset$, in which case gh^{-1}, h generate a free group of rank 2 by Case 1, or $A_{gh^{-1}} \cap A_h$ is a single point, in which case gh^{-1}, h generate a free group of rank 2 by Case 2. Hence g, h generate a free group of rank 2 since gh^{-1}, h is obtained by a Nielsen transformation from g, h.

Case 4. $\ell(h) < \Delta, \ell(h) \leq \ell(g)$ and $\Delta < \ell(g) + \ell(h)$. Then by Lemma 3.3.4, $\ell(gh^{-1}) = \ell(g) - \ell(h)$, $\Delta(gh^{-1}, h) = \Delta - \ell(h)$, and gh^{-1}, h meet coherently. If $\ell(gh^{-1}) \geq \ell(h)$, put $h_1 = h$ and $g_1 = gh^{-1}$, while if $\ell(gh^{-1}) < \ell(h)$, put $h_1 = gh^{-1}$ and $g_1 = h$. Also, let $\Delta_1 = \Delta(g_1, h_1) = \Delta - \ell(h)$. Thus $\Delta_1 < \ell(g_1) + \ell(h_1)$, $\ell(h_1) \leq \ell(h)$, $\ell(g_1) \leq \ell(g)$ and $\ell(h_1) \leq \ell(g_1)$. Also, denoting the commutator $ghg^{-1}h^{-1}$ by $[g, h]$, we have $\ell([g, h]) = 2(\ell(g) + \ell(h) - \Delta) = 2(\ell(g_1) + \ell(h_1) - \Delta_1) = \ell([g_1, h_1]) > 0$ by Lemma 3.3.9 (in fact $[g_1, h_1] = [g, h]^{\pm 1}$), so $h_1 \neq 1$ and $\ell(h_1) > 0$. Therefore, either g_1, h_1 satisfy the hypotheses of a previous case, or we remain in Case 4, and we can then repeat this procedure on g_1, h_1 to obtain g_2, h_2, and if we remain in Case 4 we can continue to obtain g_3, h_3 etc. If this procedure stops after finitely many steps, so that, for some n, g_n and h_n generate a free group by a previous case, then so do g and h since g_n and h_n are obtained from g and h by a succession of Nielsen transformations.

Suppose the procedure does not stop, and let M be the subgroup of Λ generated by $\ell(g)$ and $\ell(h)$. Inductively, M is generated by $\ell(g_n)$ and $\ell(h_n)$, and $(\ell(h_n))_{n\geq 1}$, $(\ell(g_n))_{n\geq 1}$ are decreasing sequences of elements of M and $0 < \ell(h_n) \leq \ell(g_n)$ for all n. We put $\Delta_n = \Delta(g_n, h_n)$, so inductively $\Delta_{n+1} = \Delta_n - \ell(h_n)$, and $(\Delta_n)_{n\geq 1}$ is a decreasing sequence of elements of Λ with $\Delta_n \geq 0$ for all n. Note that M is not cyclic, otherwise $(\ell(g_n) + \ell(h_n))_{n\geq 1}$ may be viewed as a strictly decreasing sequence of positive integers, which is impossible. In fact, we shall show that M is non-archimedean.

For assume that M is archimedean, so may be viewed as a subgroup of the additive group of $\mathbb{R}$. Then we have three decreasing sequences of real numbers, $(\ell(h_n))_{n\geq 1}$, $(\ell(g_n))_{n\geq 1}$ and $(\Delta_n)_{n\geq 1}$. They are bounded below by 0, so we may define $\alpha = \lim_{n\to\infty} \ell(h_n)$, $\beta = \lim_{n\to\infty} \ell(g_n)$ and $\gamma = \lim_{n\to\infty} \Delta_n$. Thus $\gamma \geq 0$, and since $0 < \ell(h_n) \leq \ell(g_n)$ for all n, $0 \leq \alpha \leq \beta$. Also, for all n, $\Delta_{n+1} = \Delta_n - \ell(h_n)$, so taking limits gives $\gamma = \gamma - \alpha$, i.e. $\alpha = 0$.

If $\beta > 0$, we can therefore choose n such that $0 < \ell(h_n) < \beta$. Then g_{n+1} cannot be h_n since $\ell(g_{n+1}) \geq \beta$, so $g_{n+1} = g_n h_n^{-1}$ and $h_{n+1} = h_n$. Inductively $h_{n+r} = h_n$ for all $r \geq 0$, so $\ell(h_n) = \alpha = 0$, a contradiction. Hence $\alpha = \beta = 0$.

Inductively, $\ell([g, h]) = 2(\ell(g_n) + \ell(h_n) - \Delta_n) > 0$ for all n, and taking limits gives $\ell([g, h]) = -2\gamma \leq 0$, a contradiction, so M is non-archimedean, as claimed.

Assume that $\Delta \notin M$. Since M is non-cyclic, $\ell(g)$, $\ell(h)$ are $\mathbb{Z}$-linearly independent. By Corollary 2.3, $\langle g,h\rangle$ is a free group of rank 2.

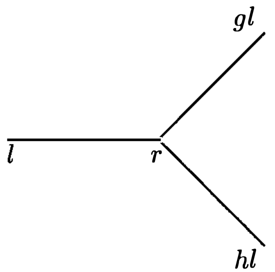
Suppose $\Delta \in M$, so $\Delta_n \in M$ for all n. By Lemma 1.1.6, M is isomorphic to $\mathbb{Z} \oplus \mathbb{Z}$ with the lexicographic ordering, so there is a non-trivial order-preserving homomorphism ϕ from M to $\mathbb{Z}$ whose kernel is infinite cyclic. For some n, we have $\phi\ell(h_n) = 0$, otherwise for some n, $\phi(\Delta_n)$ would be less than zero, which is impossible. Also, $\phi\ell(g_n) > 0$, since M is non-cyclic. It follows that $h_{n+r} = h_n$, and so $\Delta_{n+r} = \Delta_n - r\ell(h_n)$ for $r \geq 0$, hence $\Delta_n > r\ell(h_n)$ for all $r \geq 0$. Replacing g, h by g_n, h_n therefore puts us in the situation of either Corollary 2.4 or Lemma 2.5. Note that $ghg^{-1}h$ is hyperbolic, otherwise since the action is free without inversions, $h^{-1} = ghg^{-1}$. However, this is impossible as $A_{h^{-1}} = A_h$ and $A_{ghg^{-1}} = gA_h$ by Lemma 3.1.7, but $A_h \neq gA_h$ (for example, in the notation of Lemma 2.2, $r \in A_h$ but $gr \notin gA_h$). It follows from these lemmas that g and h generate a free group of rank 2.

Case 5. $\Delta \geq \ell(g) + \ell(h)$. By the remark preceding Prop. 3.3.7, $[g,h]$ has a fixed point, so $[g,h] = 1$ since the action is free, i.e. g and h commute, and $A_g = A_h$ by Lemma 3.3.9. This completes the proof. □

Note that, in Theorem 2.1, if $\langle g,h\rangle$ is abelian it will be free abelian since G is torsion-free. The proof in the case $\Lambda = \mathbb{R}$, which is essentially contained in [44], shows that, if $\langle g,h\rangle$ is non-abelian, then by Nielsen transformations, we obtain a free basis a, b such that either $A_a \cap A_b = \emptyset$ or $\Delta < \min\{\ell(a), \ell(b)\}$. For in this case, the procedure in Case 4 stops after finitely many steps, because the group M is archimedean. Therefore after a finite succession of Nielsen transformations, we end in either Case 1 or Case 2 of the proof. This leads to information on free actions of the free group of rank 2 on $\mathbb{R}$-trees (recall that there are no inversions for actions on $\mathbb{R}$-trees, by Lemma 3.1.2). The main part of the argument for the next theorem is a simplified special case of Lemma 3.1 in [44]. Before stating it, we recall from §2.2 that a polyhedral $\mathbb{R}$-tree is one whose set of branch and endpoints (i.e. the set of points of degree $\neq 2$) is closed and discrete. By 2.2.10, this is equivalent to requiring the $\mathbb{R}$-tree to be homeomorphic to real(Γ) for some simplicial tree Γ.

Theorem 2.6. *Suppose F_2 acts freely on an $\mathbb{R}$-tree (X,d). Then there is a basis $\{g,h\}$ for F_2 such that either $\Delta(g,h) < \max\{\ell(g),\ell(h)\}$ or $A_g \cap A_h = \emptyset$. If the action is minimal then X is a polyhedral $\mathbb{R}$-tree.*

Proof. The first part follows from the remarks preceding the theorem. Assume the action is minimal. Consider the case $\Delta < \max\{\ell(g),\ell(h)\}$, where $\Delta = \Delta(g,h)$. As usual, we let l, r denote the left and right-hand endpoints of $A_g \cap A_h$, and let H be the subtree spanned by l, gl and hl, so $H = [l,gl] \cup [l,hl]$, and $[l,r] \subseteq H$. Thus H is illustrated by the following, which is part of the picture after the proof of Lemma 3.3.3.



We now have to slightly modify the sets X_u, for $u \in \{g^{\pm 1}, h^{\pm 1}\}$ used in the proof of Lemma 3.3.6. Put

$$\begin{aligned} Y_g &= \{x \in X \mid [x,r] = [x,gl,r]\} \\ Y_h &= \{x \in X \mid [x,r] = [x,hl,r]\} \\ Y_{g^{-1}} &= \{x \in X \mid [x,l] = [x,g^{-1}r,l]\} \cup [l,g^{-1}l] \\ Y_{h^{-1}} &= \{x \in X \mid [x,l] = [x,h^{-1}r,l]\} \cup [l,h^{-1}l]. \end{aligned}$$

Then one of the main assertions in the proof of Lemma 3.3.6 remains true, namely

$$u(Y_u \cup Y_v \cup Y_{v^{-1}}) \subseteq Y_u \qquad (*)$$

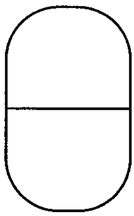
for u, $v \in \{g^{\pm 1}, h^{\pm 1}\}$ with $v \neq u$, u^{-1}. Further, it is easily verified that $uH \cap H$ contains a single point which is one of l, gl, hl, so $uH \cap H \neq \emptyset$, and that $uH \subseteq Y_u$.

We next show that F_2H is a subtree of X. Since it is clearly F_2-invariant, it will follow that $X = F_2H$. Let X_n be the union of all subtrees gH, where g is an element of F_2 of word length at most n in the basis g, h. We show that X_n is a subtree of X by induction on n. For $n = 0$, $X_0 = H$ is a subtree, and X_1 is a subtree by the observation above that $uH \cap H \neq \emptyset$. For $n > 1$, $X_{n+1} = gX_n \cup g^{-1}X_n \cup hX_n \cup h^{-1}X_n$, a union of four subtrees by induction, each of which contains H, hence X_{n+1} is a subtree. Since F_2H is the ascending union of the X_n, it is a subtree of X.

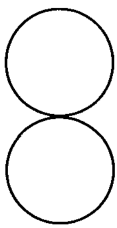
Now we show that, for all $1 \neq w \in F_2$, $wH \subseteq Y_u$, where u is the first letter in the expression of w as a reduced word relative to the basis g, h. The proof is by induction on the length n of this word. The case $n = 1$ has been noted above. For $n > 1$, write $w = uv$, where v has length $n - 1$, and v has first letter z not equal to u^{-1}. By induction, $vH \subseteq Y_z$, so $wH \subseteq Y_u$ by (*).

Since $X = F_2H$, it follows that the only branch points in X are the elements of $F_2l \cup F_2r$, and there are no endpoints of X. Further, the distance from r to the nearest branch point other than r (which is one of l, gl, hl) is

$\min\{\Delta, \ell(g) - \Delta, \ell(h) - \Delta\}$, or $\min\{\ell(g) - \Delta, \ell(h) - \Delta\}$ if $\Delta = 0$ (i.e $l = r$). By symmetry (applying this to g^{-1}, h^{-1}), this is the distance from l to the nearest branch point other than l. Thus the set of branch points is closed and discrete, as required. Incidentally, it also follows that the only points of H which are identified in the quotient X/F_2 are l, gl and hl, and all points of X have finite degree. Hence X/F_2 is a "theta graph", homeomorphic to the quotient of H obtained by identifying l, gl and hl.



The formal proof of this is left as an exercise (see the proof of Theorem 9.2.4 in [9] for inspiration). In the case $\Delta = 0$, the theta graph becomes a "rose", although a rather unimpressive one with only two petals.



In the case $A_g \cap A_h = \emptyset$, we take H to be the subtree $[p, q] \cup [p, gp] \cup [g, hq]$, where p, q are as in Lemma 3.2.2 (see the diagram after the proof of 3.2.2). We modify the sets X_u in the proof of Lemma 3.3.8 as follows:

$$\begin{aligned}
Y_g &= \{x \in X \mid [x,p] = [x, gp, p]\} \\
Y_h &= \{x \in X \mid [x,q] = [x, hq, q]\} \\
Y_{g^{-1}} &= \{x \in X \mid [x,p] = [x, g^{-1}p, p]\} \cup [g^{-1}p, p] \\
Y_{h^{-1}} &= \{x \in X \mid [x,q] = [x, h^{-1}q, q]\} \cup [h^{-1}q, q].
\end{aligned}$$

The proof now proceeds on similar lines to the first case, and details are left to the reader. The quotient X/F_2 in this case is a "barbell", as illustrated below, obtained from H by identifying the pairs of points p, gp and q, hq.

□

By contrast with Theorem 2.6, there are free, minimal actions of F_n on non-polyhedral $\mathbb{R}$-trees for $n > 2$. Examples have been given by Bestvina and Feighn (see §2.2 in [129]), by Levitt [90] and by Holton and Zamboni [78]. By Lemmas 3.3.6 and 3.3.8, in Theorem 2.6 the action is properly discontinuous, so the quotient map $X \to X/F_2$ is a covering map, and since $\mathbb{R}$-trees are contractible (by 2.2.2(4)), it is a universal covering map, so by the general theory of covering spaces, F_2 is isomorphic to $\pi_1(X/F_2)$. One can read off a basis for $\pi_1(X/F_2)$, and so for F_2, using the proof of [95; Ch.III, Prop.2.1]. (The relation between the topological and combinatorial fundamental group is discussed at the start of Chapter 6.) Morgan and Shalen [109] make use of this in proving the next result, which is a technical addendum to Theorem 2.6 which we shall need later. We shall give a proof which avoids the theory of covering spaces. Nevertheless, those familiar with this theory may find the argument of Morgan and Shalen easier to follow.

Lemma 2.7. *Suppose F_2 acts freely on an $\mathbb{R}$-tree (X, d), and g, h form a basis for F_2. Let A denote the characteristic set of the commutator $[g, h]$. Then there exist a point $p \in A$ and $u \in F_2$ such that $up \in A$ and $d(p, up) = \frac{1}{2}\ell(g)$.*

Proof. From the observations preceding Theorem 2.6 we obtain, by a finite succession of Nielsen transformations, a free basis a, b such that either $\Delta(a, b) < \min\{\ell(a), \ell(b)\}$ or $A_a \cap A_b = \emptyset$. These Nielsen transformations replace $[g, h]$ by a conjugate of $[g, h]$ or its inverse, say c, so by Lemma 3.1.7, $A_c = vA$ for some $v \in F_2$, and $\ell(c) = \ell([g, h])$. Thus we can assume that either $\Delta(g, h) < \min\{\ell(g), \ell(h)\}$ or $A_g \cap A_h = \emptyset$. We can also assume in the first case that g and h meet coherently.

Suppose $\Delta(g, h) < \min\{\ell(g), \ell(h)\}$. By further elaborating the diagram following the proof of Lemma 3.3.3, one can see that

$$[r, [g, h]r] = [r, gr, ghl, ghg^{-1}l, [g, h]r] \qquad (*)$$

where l, r are the left and right-hand endpoints of $A_g \cap A_h$. One can prove this using Lemma 2.1.5 and the information contained in the diagram after 3.3.3. For example, $[g^{-1}r, r] \cap [r, hl] = \{r\}$ by definition of r, so $[r, gr] \cap [gr, ghl] = \{gr\}$. Also $gr \neq ghl$ since $r \neq hl$ ($r <_h hl$ since $\ell(h) > \Delta(g, h)$). Continuing in this way we see that 2.1.5 applies.

Similarly, interchanging the roles of g, h, we have

$$[r, [g, h]^{-1}r] = [r, hr, hgl, hgh^{-1}l, [h, g]r].$$

Since $[r, gr] \cap [r, hr] = \{r\}$ and $gr \neq r \neq hr$, it follows that $[r, [g, h]r] \cap [r, [g, h]^{-1}r] = \{r\}$. Hence $r \in A$, and $[r, [g, h]r] \subseteq A$. Now let p be the midpoint of $[gr, ghl]$, so $p \in A$, and let $u = [g, h]g^{-1}$. Then up is the midpoint of $[ghg^{-1}l, [g, h]r]$, so $up \in A$. From $(*)$, using Lemma 2.1.4, we calculate that

$$\begin{aligned} d(p, up) &= (1/2)d(gr, ghl) + d(ghl, ghg^{-1}l) + (1/2)d(ghg^{-1}l, [g, h]r) \\ &= d(r, hl) + d(l, g^{-1}l) \\ &= (\ell(h) - \Delta(g, h)) + \ell(g) = (1/2)\ell([g, h]). \end{aligned}$$

The final equation comes from Lemma 3.3.9, or by using $(*)$ to directly calculate $d(r, [g, h]r)$.

Suppose $A_g \cap A_h = \emptyset$. Referring to Lemma 3.2.2 and the diagram for its proof, we find that

$$[p, [g, h]p] = [p, gp, gq, ghq, ghp, ghg^{-1}p, ghg^{-1}q, [g, h]q, [g, h]p].$$

This can be proved using Lemma 2.1.5. Applying $[h, g] = [g, h]^{-1}$ to this, we see that $[p, [g, h]p] \cap [p, [g, h]^{-1}p] = \{p\}$, so $p \in A$ and $ghp \in A$. Using Lemma 2.1.4, we compute that $\ell([g, h]) = d(p, [g, h]p) = 2\ell(g) + 2\ell(h) + 4d(p, q)$, and $d(p, ghp) = \ell(g) + \ell(h) + 2d(p, q)$, so we may take $u = gh$ to prove the lemma in this case. □

One can also say something about actions of two-generator groups on Λ-trees, even when the action is not free, by a generalisation of the argument for Theorem 2.1. We shall give no details of the proof, but refer to [33], because the lemmas leading up to it are considerably more complicated, and in dire need of simplification.

Theorem 2.8. *Let $G = \langle g, h \rangle$ act on a Λ-tree (X, d). Then one of the following possibilities occurs*

(i) *G acts freely and properly discontinuously on X, and g, h freely generate G.*
(ii) *By a finite succession of elementary Nielsen transformations, g and h can be transformed into new generators u and v such that at least one of $\ell(u)$, $\ell(v)$ is zero.*
(iii) *$\ell(g) > 0$, $\ell(h) > 0$ but $ghg^{-1}h^{-1}$ has a fixed point.*
(iv) *By a finite succession of elementary Nielsen transformations, g and h can be transformed into new generators u and v such that $\ell(u) > 0$, $\ell(v) > 0$,*

u and v meet coherently, $A_u \cap A_v$ is a segment of length $\ell(u)$, $\ell(v^n) < \ell(u)$ for all integers n, and $uvu^{-1}v$ has a fixed point.

If Λ is archimedean then one of (i), (ii), (iii) occurs.

□

In Case (i), one can use Lemmas 3.3.6 and 3.3.8 to see the action is without inversions.

3. Some Examples

The rest of this chapter will be devoted to examples of tree-free groups. In this section we shall mention some examples of tree-free groups which will not be discussed in detail. Our other examples, surface groups and non-standard free groups, are discussed in the remaining two sections. In this section and the next we shall mainly be looking at $\mathbb{R}$-free groups. Firstly, however, we note that the structure of Λ-free groups for $\Lambda = \mathbb{Z} \oplus \Lambda_0$, where Λ_0 is an ordered abelian group and Λ has the lexicographic ordering, has been determined in Theorem 4.7 of [5]. We shall not reproduce the result, but as a special case, if G and H are free groups, the free product with amalgamation $G \underset{a=b}{*} H$, amalgamating cyclic subgroups generated by a, b, is $(\mathbb{Z} \oplus \mathbb{Z})$-free provided a is not a proper power in G and b is not a proper power in H. Also, the HNN-extension $G = \langle F, t \mid tat^{-1} = b \rangle$, where F is a free group, a, b are not proper powers, a is not conjugate in F to b^{-1} and F acts freely without inversions on some Λ_0-tree in such a way that $\ell(a) = \ell(b)$, is $(\mathbb{Z} \oplus \Lambda_0)$-free. In proving these results, the theory of ends of Λ-trees is further developed, and the structure of Λ-trees for non-archimedean Λ is examined. We mention in passing that the results on the structure of Λ-trees have led Zamboni [151] to introduce the notion of a tree fibration.

We now consider $\mathbb{R}$-free groups, which are groups having a free action on some $\mathbb{R}$-tree, since the action must be without inversions by Criterion (iv) of Lemma 3.1.2. By Theorem 2.4.6 and Lemma 3.1.8, a free action of a group G on an $\mathbb{R}$-tree is given by a real-valued Lyndon length function L satisfying $L(g^2) > L(g)$ for all $g \in G \setminus \{1\}$. By Proposition 3.2.7, a free action of $\mathbb{R}$ itself on an $\mathbb{R}$-tree is given by an embedding $\rho : \mathbb{R} \to \mathbb{R}$, with hyperbolic length function $\ell(g) = |\rho(g)|$, which is the hyperbolic length function for an action of $\mathbb{R}$ on $\mathbb{R}$ by translations: $gx = x + \rho(g)$, for g, $x \in \mathbb{R}$. It is also the Lyndon length function relative to any point of $\mathbb{R}$ for this action.

By Proposition 1.1, if G is a free product of copies of $\mathbb{R}$, it is $\mathbb{R}$-free. We use the length function $L(a) = |a|$ on each free factor in the decomposition of G. The Lyndon length function on G constructed in the proof of Proposition 1.1 is then given by: if $g \in G$ then $L(g) = \sum_{i=1}^{n} |g_i|$, where $g = g_1 \dots g_n$ is the expression of g in normal form. (Since $\mathbb{R} = 2\mathbb{R}$, there is no need to double L.)

Lyndon (after Proposition 9.2, Ch.I in [95]) conjectured that, given any Lyndon length function $L : G \to \mathbb{R}$ satisfying $L(g^2) > L(g)$ for all $g \in G \setminus \{1\}$, G embeds in a free product of copies of $\mathbb{R}$, with L the restriction of the length function described in the previous paragraph. (A stronger version of this conjecture is attributed to Lyndon in [30], but the author can no longer remember how or when this conjecture was made.) Counterexamples to Lyndon's conjecture were given by Alperin and Moss [3] and by Promislow [121]. Interesting though their constructions are, we shall not attempt to describe them here.

These examples left open the question of whether or not an $\mathbb{R}$-free group is a free product of subgroups of $\mathbb{R}$. Had Lyndon's conjecture been true, this would have followed by the Kurosh Subgroup Theorem. This was resolved by Morgan and Shalen [109], who showed that, with three exceptions, the fundamental groups of closed surfaces are $\mathbb{R}$-free. (These groups were described at the end of §4.5.) The work of Morgan and Shalen will be briefly described in §4; it depends on the results of §4 and §5 in Chapter 4.

The fundamental groups of compact surfaces are not free products of subgroups of $\mathbb{R}$, apart from the fundamental groups of the sphere and torus, which are respectively trivial and free abelian of rank 2. For these groups are freely indecomposable, by a result of Shenitzer (see [95], Chapter II, Propositions 5.12-5.14), and the only ones which are abelian are the fundamental groups of the sphere, torus and projective plane. (These are, respectively, the orientable surfaces of genera 0 and 1, and the non-orientable surface of genus 1, and the fundamental group of the projective plane is cyclic of order 2.) The other groups are non-abelian because they can be written as non-trivial free products with amalgamation. For example, $\langle x_1, \ldots, x_g \mid x_1^2 \ldots x_g^2 = 1 \rangle$ $(g \geq 2)$ is a free product of the free group on $x_1, \ldots, x_{g-1}$ and the infinite cyclic group generated by x_g, amalgamating the infinite cyclic groups generated by $x_1^2 \ldots x_{g-1}^2$ and x_g^{-2}.

We therefore have examples of $\mathbb{R}$-free groups which are not free products of subgroups of $\mathbb{R}$.

Thus free groups, subgroups of $\mathbb{R}$ and most surface groups are $\mathbb{R}$-free, and so are free products of these groups by Proposition 1.1. (No extra examples are obtained by taking free products of subgroups of these groups, for subgroups of surface groups are either surface groups or free groups. See [75] and [76].) One of the best results on tree-free groups which has been proved so far is the following, which asserts that for finitely generated groups, these are the only examples of $\mathbb{R}$-free groups.

Rips' Theorem. *A finitely generated $\mathbb{R}$-free group G can be written as a free product $G = G_1 * \cdots * G_n$ for some integer $n \geq 1$, where each G_i is either a finitely generated free abelian group or the fundamental group of a closed surface.*

(It is unnecessary to allow the G_i to be free groups, since free groups are free products of infinite cyclic groups.) Rips outlined a proof of this at a conference at the Isle of Thorns, a conference centre of the University of Sussex, England, in 1991. It was not published, but versions of it have now appeared in [13] and [53]. See also [125]. The proof in [53] will be given in Chapter 6.

This still leaves the question of whether or not any $\mathbb{R}$-free group is a free product of subgroups of $\mathbb{R}$ and surface groups. Counterexamples have been given by Dunwoody [46] and by Zastrow [153]. We shall describe Zastrow's example.

Let F_n be a free group of rank n with basis $x_1, \ldots, x_n$, and let $p_n : F_{n+1} \to F_n$ be the homomorphism sending x_i to x_i for $1 \leq i \leq n$ and x_{n+1} to 1. Let F be the inverse limit of the system $F_1 \xleftarrow{p_1} F_2 \xleftarrow{p_2} F_3 \ldots$, so that elements of F are infinite sequences $(g_1, g_2, \ldots)$ where $g_i \in F_i$ and $g_i = p_i(g_{i+1})$ for all $i \geq 1$, with coordinate-wise multiplication. Let w_i be the reduced word in $x_1^{\pm 1}, \ldots, x_i^{\pm 1}$ representing g_i, and let n_{ij} be the number of letters $x_j^{\pm 1}$ occurring in w_i. Then as i increases and j is fixed, either n_{ij} is eventually constant or $n_{ij} \to \infty$. Let H be the subset of F consisting of those elements for which n_{ij} is eventually constant for all $j \geq 1$. Then H is a subgroup of F, in fact it is the fundamental group of the Hawaiian earring (see [104]).

For $g = (g_1, g_2, \ldots) \in H$, let w_{ij} be the word on $x_1^{\pm 1}, \ldots, x_m^{\pm 1}$, where $m = \min\{i, j\}$, obtained from w_i by deleting all occurrences of $x_k^{\pm 1}$ with $k > j$ from w_i. Because $g \in H$, when j is fixed and i is increased, w_{ij} is eventually constant; denote this eventually constant word by γ_j. The sequence $\gamma = (\gamma_1, \gamma_2, \ldots)$ is called the word sequence associated with g. Put $L_i(g) =$ the number of occurrences of $x_i^{\pm 1}$ in γ_i, and

$$L(g) = \sum_{i=1}^{\infty} \frac{1}{i} L_i(g).$$

Now let $G = \{g \in H \mid L(g) < \infty\}$. It is shown in [153] that G is a subgroup of H and L is a Lyndon length function on G, satisfying $L(g^2) > L(g)$ for all $g \in G$, $g \neq 1$. We shall not reproduce the details. It follows that G is $\mathbb{R}$-free, by Theorem 2.4.6 and Lemma 3.1.8. To show G is not a free product of surface groups and subgroups of $\mathbb{R}$, some lemmas are needed. We shall use without comment the Nielsen-Schreier Theorem (subgroups of free groups are free-see Theorem 1, Chapter 7 in [39]).

Lemma 3.1. (1) *For $1 \neq g \in H$, there is an integer N such that if $g = h^n$, where $h \in H$ and $n \in \mathbb{N}$, then $|n| \leq N$.*

(2) *Any finitely generated subgroup of H is free.*

Proof. (1) If $g = (g_1, g_2, \ldots) \neq 1$, fix i with $g_i \neq 1$. If $g = h^n$, then applying the canonical projection homomorphism $\pi_i : H \to F_i$, which sends an element to its ith coordinate, $g_i = \pi_i(h)^n$, and there is a bound N, depending on g_i, for $|n|$, by well-known properties of free groups.

(2) Let K be a subgroup of H whose minimal number of generators is $d < \infty$. The group $\pi_i(K)$ is then free (being a subgroup of F_i) of rank at most d (being an epimorphic image of K). Also, the map $p_i : F_{i+1} \to F_i$ used to define F maps $\pi_{i+1}(K)$ onto $\pi_i(K)$, so the rank of $\pi_i(K)$ increases with i. Therefore, there is an integer k such that $\pi_i(K)$ has constant rank, say r, for $i \geq k$. Since free groups are Hopfian (see the corollary to Prop. 5, Chapter 1 in [39]), $p_i : \pi_{i+1}(K) \to \pi_i(K)$ is an isomorphism for $i \geq k$. Hence the inverse limit of $\pi_1(K) \xleftarrow{p_1} \pi_2(K) \xleftarrow{p_2} \pi_3(K) \ldots$ is the free group of rank r, and K is a subgroup of this inverse limit, so K is free. □

Note that, if $g = (g_1, g_2, \ldots) \in H$ and $\gamma = (\gamma_1, \gamma_2, \ldots)$ is the corresponding word sequence, the mapping $g \mapsto \gamma$ is one-to-one, because freely reducing the word γ_i gives the reduced word representing g_i. It is not hard to see that the image of this mapping is the set of sequences $(\gamma_1, \gamma_2, \ldots)$, where γ_i is a word on $x_1^{\pm 1}, \ldots, x_i^{\pm 1}$, satisfying, for all i:

(i) γ_i is obtained by deleting all occurrences of $x_{i+1}^{\pm 1}$ in γ_{i+1};
(ii) if γ_i contains two adjacent letters which are inverses, then there exists $k > 0$ such that the subword of γ_{i+k} between the two corresponding letters of γ_{i+k} cannot be freely reduced to the empty word.

Lemma 3.2. *Let* $G_1 = \left\{g \in H \;\middle|\; \sum_{i=1}^{\infty} \frac{1}{4^i} L_i(g) < \infty\right\}$. *Then the abelianisation* $G_1/[G_1, G_1]$ *is not free abelian.*

Proof. The details of this in [153] are rather complicated to write out, and are omitted. It depends on a proof that the group $\mathbb{Z}^{\mathbb{N}}$ is not free abelian. There is a homomorphism $H \to \mathbb{Z}^{\mathbb{N}}$, $(\gamma_1, \gamma_2, \ldots) \mapsto (n_1, n_2, \ldots)$, where n_i is the exponent sum of x_i in γ_i, and $(\gamma_1, \gamma_2, \ldots)$ is a word sequence representing an element of H (it is convenient in this proof to view elements of H as word sequences). This map factors through $H/[H, H]$, and the computations to show $\mathbb{Z}^{\mathbb{N}}$ is not free abelian can be mimicked in $H/[H, H]$ to show this group is not free abelian. Finally, one observes that the argument applies also to $G_1/[G_1, G_1]$. □

Lemma 3.3. *The group G is not free.*

Proof. We construct an embedding $\phi : G_1 \to G$. If $g = (g_1, g_2, \ldots) \in H$ with corresponding word sequence $(\gamma_1, \gamma_2, \ldots)$, let γ' be the result of replacing all occurrences of x_i^{ε} in γ_i by $x_{4^i}^{\varepsilon}$, for all i, where $\varepsilon = \pm 1$, and put $\gamma_0' = 1$ (the empty word). Define the sequence $\delta = (\delta_1, \delta_2, \ldots)$ by $\delta_k = \gamma_{i-1}'$ for $4^{i-1} \leq k < 4^i$, $i \geq 1$. Then δ is a word sequence representing an element of H, which we denote by $\phi(g)$, so defining a mapping $\phi : H \to H$ which is clearly a homomorphism. Also, for $g \in H$, $\sum_{i=1}^{\infty} \frac{1}{4^i} L_i(g) = \sum_{i=1}^{\infty} \frac{1}{i} L_i(\phi(g))$, hence ϕ

restricted to G_1 maps G_1 to G and is an embedding. By Lemma 3.2 G_1 is not free, so neither is G. □

Theorem 3.4. *The group G is not a free product of fundamental groups of closed surfaces and subgroups of $\mathbb{R}$.*

Proof. By Lemma 3.1(2) no surface group can occur as a free factor in a free product decomposition of G, and if a finitely generated abelian group occurs, it must be infinite cyclic. Suppose A is an infinitely generated abelian subgroup of G. Then given $a \in A$, with $a \neq 1$, we can find an infinite sequence $a_1, a_2, \ldots$ of elements of A with $a_1 = a$ and $a_i \notin \langle a_1, \ldots, a_{i-1} \rangle$ for $i > 1$. Again by Lemma 3.1(2), $\langle a_1, \ldots, a_i \rangle$ is cyclic, generated by b_i, say, and we may take $b_1 = a$. Then for $i > 1$, $b_{i-1} = b_i^{n_i}$ for some integer n_i with $|n_i| > 1$, so $a = b_1 = b_2^{n_2} = b_3^{n_2 n_3} = \ldots$, contradicting Lemma 3.1(1). Thus if there is a decomposition of G as in the theorem, it can have only infinite cyclic factors, which implies G is a free group, contradicting Lemma 3.3. □

Dunwoody's example [46] is a group originally considered by Kurosh:

$$\langle a_1, b_1, a_2, b_2, \ldots \mid b_1 = [a_2, b_2], b_2 = [a_3, b_3], \ldots \rangle.$$

The proof that it is $\mathbb{R}$-free uses the theory of protrees, as developed in [46], and will not be reproduced here. Dunwoody points out that the method also shows that $\langle a_1, b_1, a_2, b_2, \ldots \mid b_1 = a_2^2 b_2^2, b_2 = a_3^2 b_3^2, \ldots \rangle$ is $\mathbb{R}$-free. Zastrow [153] notes that this group embeds in the group G in Theorem 3.4. In connection with Zastrow's example, we mention a recent paper on the structure of the Hawaiian earring group by Cannon and Conner [25].

4. Free Actions of Surface Groups

By a surface group we mean the fundamental group of a closed surface, that is, of a compact connected 2-manifold without boundary. As noted at the end of §4.5, there is a well-known classification of such surfaces, and corresponding presentations of their fundamental group. Such a group has either a presentation

$$\langle a_1, b_1, \ldots, a_g, b_g \mid [a_1, b_1] \ldots [a_g, b_g] = 1 \rangle$$

(the orientable surface of genus g), or

$$\langle x_1, \ldots, x_g \mid x_1^2 \ldots x_g^2 = 1 \rangle$$

(the non-orientable surface of genus g).

We shall exclude the sphere S^2 (the orientable surface of genus 0) from consideration. Its fundamental group is trivial, so Λ-free for any Λ. The following result is due to Morgan and Shalen [109].

Theorem 4.1. *The fundamental group of a closed surface is $\mathbb{R}$-free, except for the non-orientable surfaces of genus 1, 2 and 3 (the connected sum of 1, 2 or 3 real projective planes).*

We call the non-orientable surfaces of genus 1, 2 and 3 *exceptional* surfaces, and their fundamental groups *exceptional surface groups*. The orientable surface of genus 1 (the torus) has fundamental group $\mathbb{Z} \oplus \mathbb{Z}$, which is embeddable in $\mathbb{R}$, so is $\mathbb{R}$-free, as noted in §1. In the remaining cases, the surface has a hyperbolic structure, and has finite area, as noted at the end of §4.5. Thus we assume S is a closed non-exceptional hyperbolic surface of finite area, and need to prove $\pi_1(S)$ is $\mathbb{R}$-free. The proof of Theorem 4.1 uses the results in §4.5, that the space PML(S) of projectivised measured geodesic laminations in S is compact, and elements of ML(S) have dual $\mathbb{R}$-trees. It would involve too great a digression to give the argument in detail, but we shall give a summary of the steps involved.

First, if S is orientable, let Z be the set of simple closed geodesics in S which are non-separating in S. If S is non-orientable, let Z be the set of simple closed geodesics which are non-separating, orientation-preserving and which have non-orientable complement. Each element α of Z determines an element of L(S), (see §4.5) and we can obtain an element of ML(S) by the procedure given after Lemma 4.4.11, and in this case there is only one positive real number μ_i involved, hence α determines a unique element of PML(S) (see the remarks before Theorem 4.5.16). Thus we have an injective map $Z \to$ PML(S). Let W be the closure of the image of Z under this map. Since PML(S) is a compact Hausdorff space, by Theorem 4.5.16, so is W.

From the remarks at the end of §4.5, S can be realised as $\mathbb{H}^2/G$, where $G = \pi_1(S)$ is an NEC group with a certain convex hyperbolic polygon and its interior as a fundamental domain. In the orientable case, the image under the projection map $\mathbb{H}^2 \to \mathbb{H}^2/G$ of a side of the polygon gives an element of Z (all vertices are identified to a single point). More care is needed in the non-orientable case, and it is left to the reader to find an element of Z. Thus W is non-empty.

If (C, μ) is an element of ML(S), then C is uniquely determined by the image of (C, μ) in PML(S) under projection (see the observations preceding Theorem 4.5.16). Hence an element of PML(S) determines a geodesic lamination on S, by Lemma 4.5.9. By a leaf of an element in PML(S), we mean a leaf of the corresponding geodesic lamination.

If α is a simple closed geodesic in S, Let U_α be the subset of PML(S) consisting of all elements of PML(S) which have at least one leaf crossing α. We shall now state two lemmas without proof. It is in the first that essential use is made of the fact that S is non-exceptional.

Lemma 4.2. *Suppose $\gamma \in Z$. Let α be a simple closed geodesic which does* *not cross γ. Then there an orientation-preserving simple closed geodesic β which crosses α, and which crosses γ in a single point.*

□

The next lemma is the main step in showing S is $\mathbb{R}$-free. It uses Lemma 4.2 and a result whose proof uses hyperbolic geometry, which is given in an appendix to [109].

Lemma 4.3. *The subset $W \cap U_\alpha$ is open and dense in W.*

□

The final step is the following.

Lemma 4.4. *There exists a measured geodesic lamination on S such that the leaves and components of the complement are simply connected.*

Proof. Since W is a non-empty compact Hausdorff space, it follows from Lemma 4.3 and the Baire Category Theorem (Corollary 17.3, Chapter I in [22]) that $W \cap \bigcap_\alpha U_\alpha \neq \emptyset$. (This is an intersection of countably many sets, by Lemme 1 and Lemme 4, exposé 3 in [47]; for Lemme 4, see also Lemma 2.3 in [27].) By Lemma 4.5.13, there is a measured geodesic lamination $\mathfrak{L}$ in S such that for every simple closed geodesic α there is a leaf of $\mathfrak{L}$ crossing α. It follows that $\mathfrak{L}$ contains no leaf which is a closed geodesic, so all leaves are homeomorphic to $\mathbb{R}$, and there is no simple closed geodesic which fails to meet the support of $\mathfrak{L}$.

It follows that the complementary regions of $\mathfrak{L}$ are simply connected. For let U be a complementary region and let $\tilde{U}$ be a component of $p^{-1}(U)$, where $p : \mathbb{H}^2 \to S$ is the universal covering map. We claim that $p : \tilde{U} \to U$ is a homeomorphism. By the general theory of covering spaces, it is a covering map (see Lemma 2.1, Chapter 5 in [98]), so it suffices to show it is injective. Suppose not, and take $x,\ y \in \tilde{U}$ such that $x \neq y$ and $p(x) = p(y)$. Let $\alpha : [a,b] \to \mathbb{H}^2$ be a simple parametrisation of the hyperbolic line segment joining x and y. Then $\alpha([a,b]) \subseteq \tilde{U}$, since $\tilde{U}$ is hyperbolically convex (see the proof of Lemma 4.1 in [27]). Each point $\alpha(t)$ has an open neighbourhood U_t such that $p|_{U_t}$ is a homeomorphism, and we let δ be a Lebesgue number for the covering $\{\alpha^{-1}(U_t) \mid t \in [a,b]\}$ of $[a,b]$. Since $p\alpha(a) = p\alpha(b)$, $\delta \leq |a-b|$. The set of points $(s,t) \in [a,b] \times [a,b]$ such that $\alpha(s) = \alpha(t)$ and $|s-t| \geq \delta$ is therefore non-empty and clearly compact, so there is an element (c,d) in this set with $|c-d|$ as small as possible. Then $p \circ \alpha$ restricted to $[c,d]$ has image a simple closed geodesic lying in U, a contradiction. Thus $p : \tilde{U} \to U$ is a homeomorphism, and since $\tilde{U}$ is hyperbolically convex, U is simply connected.

□

We can now show that S has a free action on an $\mathbb{R}$-tree. Let $\mathfrak{L}$ be a lamination as in Lemma 4.4. By Lemma 4.5.13 there is a dual $\mathbb{R}$-tree on which $\pi_1(S)$ acts, spanned by the complementary regions of the leaves of $\tilde{\mathcal{L}}$, the lift of $\mathcal{L}$ to the universal covering $\mathbb{H}$ of S. Let $p : \mathbb{H} \to S$ be the covering projection. The restriction of p to a leaf or complementary region of $\tilde{\mathcal{L}}$ is a covering map (Lemma 2.1, Chapter 5 in [98]). If an element of $\pi_1(S)$ stabilises a leaf or complementary region, it acts as a covering transformation for the restriction of p, and since the leaves and complementary regions of $\mathcal{L}$ are simply connected, it fixes the leaf or complementary region pointwise (see Theorem 6.8, Chapter III in [22]). But the action of $\pi_1(S)$ by covering transformations is free (see Ch. III in [22], before Proposition 6.2), so the stabiliser of each leaf and complementary region is trivial. The action of $\pi_1(S)$ on the dual $\mathbb{R}$-tree is free by Lemma 4.4.10.

A much simpler construction of a free action in the orientable case is given in [137], using measured foliations rather than measured geodesic laminations. A version of the theory in §4.4 for measured foliations is given in [63], using a generalisation of a triangulated manifold, and allowing singularities in the foliation.

To complete the proof of Theorem 4.1, it remains to show that the fundamental group of the non-orientable surface group of genus n is not $\mathbb{R}$-free for $1 \leq n \leq 3$. This involves only the theory of $\mathbb{R}$-trees and can be given in detail. For $n = 3$, the fundamental group is the group $\langle x, y, z \mid x^2y^2z^2 = 1\rangle$. Putting $g = z^{-1}y^{-2}x^{-1}$, $h = xy$ and $v = z^{-1}y^{-1}x^{-1}$, so $x = v^{-1}gh$, $y = h^{-1}g^{-1}vh$ and $z = h^{-1}v^{-1}$, we obtain by Tietze transformations the presentation $\langle g, h, v \mid [g, h] = v^2\rangle$. (Tietze transformations are discussed in [39; Ch.1, §1.2].) This reflects the fact that the connected sum of three projective planes is homeomorphic to the connected sum of a torus and a projective plane (see [98; Ch.1, Lemma 7.1]).

Theorem 4.5. *The fundamental group of the non-orientable surface group of genus n is not $\mathbb{R}$-free for $1 \leq n \leq 3$.*

Proof. For $n = 1$ the group is cyclic of order 2, so is not tree-free since tree-free groups are torsion-free. For $n = 2$, the group has presentation $\langle x, y \mid x^2 = y^2\rangle$, which is the free product of two infinite cyclic groups amalgamating their subgroups of index 2, so is non-abelian, and it is not free (this follows from §3, but is most easily seen by abelianising). By Theorem 2.1, this group is not tree-free. For $n = 3$, we have noted that the group, which we denote by G, has the presentation $\langle g, h, v \mid [g, h] = v^2\rangle$, which is the free product with amalgamation of the free group F of rank 2 with basis g, h, and the infinite cyclic group generated by v, amalgamating the cyclic subgroups generated by $[g, h]$ and v^2. Suppose G acts freely on an $\mathbb{R}$-tree (X, d), so F acts freely by restriction. By Lemma 2.7, there exist a point $p \in A_{[g,h]}$ and $u \in F$ such that $up \in A_{[g,h]}$ and $d(p, up) = (1/2)\ell([g, h])$. Since $v^2 = [g, h]$, $A_v = A_{[g,h]}$ and $\ell(v) = (1/2)\ell([g, h])$ by Lemma 3.1.7, and v acts as a translation on A_v. Hence

$v^{\pm 1}up = p$, so $u = v^{\pm 1}$, a contradiction since $v \notin F$. □

This completes the proof of Theorem 4.1. We note that, although the fundamental group G of the non-orientable surface of genus 3 is not $\mathbb{R}$-free, it is tree-free, in fact it is Λ-free, where $\Lambda = \mathbb{Z} \oplus \mathbb{Z}$ with the lexicographic ordering. This has been observed after Theorem 3 in [60], and can be seen as follows. In the presentation $G = \langle g, h, v \mid [g, h] = v^2 \rangle$, put $w = hv$, $t = v^{-2}g$ to obtain $G = \langle t, v, w \mid t(wv^{-1})t^{-1} = wv \rangle$, an HNN-extension with base the free group F generated by v, w. Let $\langle a, b \rangle$ be a free group of rank 2, acting on its Cayley graph with respect to these generators, viewed as a $\mathbb{Z}$-tree which we denote by X. There is an embedding of F into $\langle a, b \rangle$ given by $w \mapsto a$, $v \mapsto bab^{-1}$, giving a free action of F on X. For this action, it is easily checked that $A_v \cap A_w = \emptyset$, so by Lemma 3.2.2, $\ell(wv^{-1}) = \ell(wv)$ for this action (both being equal to $\ell(w) + \ell(v) + 2d(A_v, A_w) = 4$). Further, wv^{-1} and wv are not proper powers in F and, being cyclically reduced, wv and vw^{-1} are not conjugate in F. By 4.17 in [5], G is $(\mathbb{Z} \oplus \mathbb{Z})$-free, as claimed.

It follows from Theorem 4.1, the remark near the beginning of §1 and Proposition 1.1 that a free product of finitely generated free abelian groups and non-exceptional surface groups is $\mathbb{R}$-free. Together with Rips' Theorem (which was stated in the previous section and will be proved in Chapter 6), this gives the following characterisation.

Theorem 4.6. *A finitely generated group G is $\mathbb{R}$-free if and only if it can be written as a free product $G = G_1 * \cdots * G_n$ for some integer $n \geq 1$, where each G_i is either a finitely generated free abelian group or a non-exceptional surface group.*

□

(No exceptional surface groups can occur in the decomposition by Theorem 4.5.)

5. Non-standard Free Groups

A non-standard free group is a non-free model of the first order theory of some free group, in the first-order language of groups, which has function symbols $\cdot$ (binary) and $^{-1}$ (unary) and constant symbol 1. Actually, we shall only consider models which are ultrapowers of free groups, and keep the model theory to a minimum. We shall show that all such ultrapowers are tree-free, hence so are all their subgroups, and this includes all non-standard free groups. This follows from general model-theoretic results (for example, see [11; Ch. 9, Lemma 3.8]). We shall also give a group-theoretic characterisation of subgroups of ultrapowers of free groups. First, we describe the construction of ultraproducts and ultrapowers in general.

Let I be a set and let $\mathcal{P}(I)$ be the Boolean algebra of all subsets of I. A filter in $\mathcal{P}(I)$ is a subset $\mathcal{F}$ of $\mathcal{P}(I)$ such that

(i) $A \in \mathcal{F}$ and $A \subseteq B \subseteq I$ implies $B \in \mathcal{F}$
(ii) A, $B \in \mathcal{F}$ implies $A \cap B \in \mathcal{F}$
(iii) $\emptyset \notin \mathcal{F}$.

Conditions (i) and (ii) are equivalent to the requirement that the set $\{A \in \mathcal{P}(I) \mid I \setminus A \in \mathcal{F}\}$ is an ideal in the Boolean ring $\mathcal{P}(I)$ (where addition is symmetric difference, and multiplication is intersection). Given (i) and (ii), (iii) is equivalent to requiring that this set is a proper ideal. A filter $\mathcal{F}$ is called an ultrafilter if it satisfies the following stronger version of (iii).

(iii)′ for all $A \in \mathcal{P}(I)$, exactly one of A, $I \setminus A$ belongs to $\mathcal{D}$.

Equivalently, a filter $\mathcal{D}$ is an ultrafilter if and only if $\{A \in \mathcal{P}(I) \mid I \setminus A \in \mathcal{D}\}$ is a maximal ideal in the Boolean ring $\mathcal{P}(I)$ (exercise). Before proceeding, we prove a result on the existence of ultrafilters.

Definition. Let I be a set. A subset $\mathcal{B}$ of $\mathcal{P}(I)$ is called a filter base if
(i) $\emptyset \notin \mathcal{B}$
(ii) for all B, $B' \in \mathcal{B}$, there exists $B'' \in \mathcal{B}$ such that $B'' \subseteq B \cap B'$.

Lemma 5.1. *If $\mathcal{B}$ is a filter base in $\mathcal{P}(I)$, then there is an ultrafilter $\mathcal{D}$ on I such that $\mathcal{B} \subseteq \mathcal{D}$.*

Proof. Let $\mathcal{F} = \{X \in \mathcal{P}(I) \mid \exists B \in \mathcal{B}$ such that $B \subseteq X\}$. It is easily checked that $\mathcal{F}$ is a filter in $\mathcal{P}(I)$. Now apply Zorn's Lemma to see that the set

$$\{\mathcal{F}' \mid \mathcal{F}' \text{ is a filter in } \mathcal{P}(I) \text{ and } \mathcal{F} \subseteq \mathcal{F}'\},$$

ordered by inclusion, has a maximal element $\mathcal{D}$. Then $\mathcal{D}$ is the required ultrafilter. □

Let $\{X_i \mid i \in I\}$ be a family of sets indexed by I, and let $\mathcal{D}$ be an ultrafilter in $\mathcal{P}(I)$. The ultraproduct $\prod_{i\in I} X_i/\mathcal{D}$ is defined to be the quotient set $\prod_{i\in I} X_i/\sim$, where $(x_i)_{i\in I} \sim (y_i)_{i\in I}$ if and only if $\{i \in I \mid x_i = y_i\} \in \mathcal{D}$. This is easily checked to be an equivalence relation. We shall denote the equivalence class of an element $(x_i)_{i\in I} \in \prod_{i\in I} X_i$ by $\langle x_i\rangle_{i\in I}$, and usually abbreviate this to $\langle x_i\rangle$. If all $X_i = X$, a single set, then the ultraproduct is $X^I/\mathcal{D}$, which is called an ultrapower of X.

Denoting the ultrapower $X^I/\mathcal{D}$ by ${}^*X$, given a function $f : X^n \to Y$ of n variables, we obtain an extension to a function ${}^*f : ({}^*X)^n \to {}^*Y$ by defining ${}^*f(x^{(1)}, \ldots, x^{(n)}) = \langle f(x_i^{(1)}, \ldots, x_i^{(n)})\rangle$, where $x^1 = \langle x_i^1\rangle$, etc. Similarly any finitary relations on X can be extended to ${}^*X$. Thus if X is a group, then we can define a multiplication in ${}^*X$ by $\langle x_i\rangle\langle y_i\rangle = \langle x_iy_i\rangle$, and this makes ${}^*X$ into a group. If X is abelian, so is $X^I/\mathcal{D}$, and if X is an ordered abelian group, then $X^I/\mathcal{D}$ can be made into an ordered abelian group by defining $\langle x_i\rangle \leq \langle y_i\rangle$ if and only if $\{i \in I \mid x_i \leq y_i\} \in \mathcal{D}$. Similarly if X is a ring or a field, so is $X^I/\mathcal{D}$. These assertions can be proved directly, but all our claims follow from Łoš's Theorem (see Corollary 2.2 or Lemma 2.3, Ch.5 in [11], or for those not familiar with first order logic, [145]).

Given a set X, an index set I and an ultrafilter $\mathcal{D}$ on I, there is a canonical embedding $\phi : X \to X^I/\mathcal{D}$, given by $x \mapsto \langle x_i\rangle$, where $x_i = x$ for all $i \in I$. If X is a group, this is an embedding of groups, and similarly if X is an ordered abelian group, ring, etc.

An ultrafilter $\mathcal{D}$ on I is called *principal* if some finite subset of I belongs to $\mathcal{D}$. In this case, the embedding $\phi : X \to X^I/\mathcal{D}$ is surjective, so identifies $X^I/\mathcal{D}$ with X. To see this, we first note that, if $A \in \mathcal{D}$ and A is finite, then $\{i\} \in \mathcal{D}$ for some $i \in A$. For otherwise, $I \setminus \{i\} \in \mathcal{D}$ for all $i \in A$ (by Axiom (iii)$'$ for an ultrafilter), hence $\bigcap_{i\in A} I\setminus\{i\} \in \mathcal{D}$ (Axiom (ii)), that is, $I\setminus A \in \mathcal{D}$, contradicting Axiom (iii)$'$. Thus if we fix $j \in I$ such that $\{j\} \in \mathcal{D}$, and an element $\langle x_i\rangle$ of $X^I/\mathcal{D}$, then $\langle x_i\rangle = \phi(x_j)$. For this reason, it is often assumed that $\mathcal{D}$ is non-principal, but this is unnecessary in what follows and we shall not do so.

If F is a free group, we noted in §1 that F acts freely on its Cayley graph, relative to some basis X. Then using the vertex 1 as basepoint, it is easy to see that the corresponding Lyndon length function L is given by: $L(u)$ is the length of the reduced word on $X^{\pm 1}$ representing u, and $c(u, v)$ is the length of the longest common terminal segment of the reduced words representing u and v. Since the action is free and without inversions, $L(u^2) > L(u)$ for all $u \in F \setminus \{1\}$ by Lemma 3.1.8. Let I be an index set, let $\mathcal{D}$ be an ultrafilter in $\mathcal{P}(I)$ and denote the ultrapower $X^I/\mathcal{D}$ by ${}^*X$. Then as indicated above we obtain a function ${}^*L : {}^*F \to {}^*\mathbb{Z}$, given by ${}^*L(\langle u_i\rangle) = \langle L(u_i)\rangle$, and ${}^*L$ is a Lyndon length function. This can be checked directly, but follows from the version of Łoš's Theorem given in 3.18 of [145]. Further, ${}^*L(u^2) > {}^*L(u)$ for all $u \in {}^*F\setminus\{1\}$. Therefore there is an action of ${}^*F$ on a ${}^*\mathbb{Z}$-tree by Theorem 2.4.6, and this action is free without inversions by Lemma 3.1.8. Thus ${}^*F$, and all its subgroups, are tree-free.

This example is due to Gaglione and Spellman [58], [59] (see also Remeslennikov [123]). These authors were led to consider ultrapowers of free groups by model-theoretic considerations. We shall give a generalisation of some of the results of Gaglione and Spellman in Theorem 5.9 below. This follows easily from a group-theoretic characterisation of subgroups of ultrapowers of free groups (Theorem 5.8), which depends on further work of Remeslennikov [122].

Let $\mathfrak{X}$ be a class of groups (required only to be non-empty). A group G is called n-residually $\mathfrak{X}$, where n is a positive integer, if given a collection $g_1, \ldots, g_n$ of elements of $G \setminus \{1\}$, there is a surjective homomorphism $\phi : G \to$

H, where $H \in \mathfrak{X}$, such that $\phi(g_i) \neq 1$ for $1 \leq i \leq n$. A group is called fully residually $\mathfrak{X}$ if it is n-residually $\mathfrak{X}$ for all $n > 0$. Also, "residually $\mathfrak{X}$", without qualification, means 1-residually $\mathfrak{X}$. If $\mathfrak{X}$ is subgroup closed, it is unnecessary in the definition to require ϕ to be surjective.

Remark. It is well-known that for any class of groups $\mathfrak{X}$, a residually $\mathfrak{X}$ group embeds in a direct product of groups in $\mathfrak{X}$. For suppose G is residually $\mathfrak{X}$. Obviously if G is the trivial group, it embeds in a direct product of groups in $\mathfrak{X}$, so we put $X = G \setminus \{1\}$ and assume $X \neq \emptyset$. For $x \in X$, let $\phi_x : G \to H_x$ be a homomorphism with $\phi_x(x) \neq 0$ and $H_x \in \mathfrak{X}$. Then $\phi = \prod_{x \in X} \phi_x : G \to \prod_{x \in X} H_x$ is an embedding.

Conversely, if $H = \prod_{i \in I} H_i$, where $H_i \in \mathfrak{X}$ for all i, then we see that H is residually $\mathfrak{X}$, by use of the projection maps $H \to H_i$. If $\mathfrak{X}$ is subgroup closed, it is easily seen that the class of residually $\mathfrak{X}$ groups is also subgroup closed. Thus in this case, a group G is residually $\mathfrak{X}$ if and only if it embeds in a direct product of groups in $\mathfrak{X}$.

These definitions apply to the class of all free groups. We shall show that a group G embeds in some ultrapower of a free group if and only if it is locally fully residually free ("locally" was defined just before Theorem 4.2.16). First, we shall establish some results on residually free groups; not all of these will be used, but it is of interest that they involve properties of tree-free groups established in §§1 and 2. Further information on residually and fully residually free groups, including examples, is contained in [6] and [7]. We begin with abelian groups, following [6]. Since abelian subgroups of free groups are cyclic, an abelian group is residually free (resp. fully residually free) if and only if it is residually infinite cyclic (resp. fully residually infinite cyclic).

Proposition 5.2. *For an abelian group G, the following are equivalent.*
(1) *G is residually free;*
(2) *G embeds in a direct product of infinite cyclic groups;*
(3) *G is fully residually free.*

Consequently, free abelian groups are fully residually free.

Proof. It follows from the remark and the observations immediately preceding the proposition that (1) implies (2).

Since the class of fully residually free groups is clearly subgroup closed, to prove (2) implies (3) it suffices to show that a direct product $P = \prod_{i \in I} \mathbb{Z}$ is fully residually free (viewing $\mathbb{Z}$ as an infinite cyclic group). Given finitely many non-identity elements of P, there is a finite subset J of I such that the projection map $P \to \prod_{i \in J} \mathbb{Z}$ maps these elements to non-identity elements of $\prod_{i \in J} \mathbb{Z}$. We may therefore assume I is finite. By induction on $|I|$, it suffices to show that, if A, B are abelian fully residually free groups, then $A \times B$ is fully residually free. Let $(a_1, b_1), \ldots, (a_n, b_n)$ be non-identity elements of $A \times B$.

Since A, B are abelian, there are homomorphisms $\phi : A \to \mathbb{Z}$, $\psi : B \to \mathbb{Z}$ such that

$$\begin{aligned} \phi(a_i) &\neq 0 && \text{for all } i \text{ such that } a_i \neq 1 \\ \psi(b_i) &\neq 0 && \text{for all } i \text{ such that } b_i \neq 1. \end{aligned}$$

Then $\theta = \phi \times \psi : A \times B \to \mathbb{Z} \times \mathbb{Z}$ is a homomorphism with $\theta(a_i, b_i) \neq 0$ for $1 \leq i \leq n$. It therefore suffices to show that $\mathbb{Z} \times \mathbb{Z}$ is fully residually free. Let $(c_1, d_1), \ldots, (c_n, d_n)$ be non-identity elements of $\mathbb{Z} \times \mathbb{Z}$. Let x be a prime not occurring in the factorisation of any non-zero c_i or d_i. The homomorphism $f : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ given by $f(1,0) = x$, $f(0,1) = 1$ then satisfies $f(c_i, d_i) \neq 0$ for $1 \leq i \leq n$, as required.

Obviously (3) implies (1), and the final assertion follows from the implication (2) $\Rightarrow$ (3). □

In the next result, the fact that (1) and (3) are equivalent appears in [6], and the equivalence of (1) and (2) is noted in [122]. For definitions of the terms used in (3) and (4), see §1.

Proposition 5.3. *For a residually free group G, the following are equivalent.*
(1) *G is fully residually free;*
(2) *G is 2-residually free;*
(3) *G is commutative transitive;*
(4) *G is CSA.*

Proof. Obviously (1) implies (2). Assume (2). Suppose x, y, $z \in G$, $[x, y] = [y, z] = 1$ and $y \neq 1$. We need to show $[x, z] = 1$. If not, there is a homomorphism $\phi : G \to F$, where F is a free group, with $\phi(y) \neq 1$ and $\phi([x, z]) = [\phi(x), \phi(z)] \neq 1$. But then $[\phi(x), \phi(y)] = [\phi(y), \phi(z)] = 1$, contradicting that F is commutative transitive. Thus (2) implies (3).

Assume (3). To prove (4), we assume M is a maximal abelian subgroup of G, $g \in G \setminus M$ and $M \cap gMg^{-1} \neq 1$, and obtain a contradiction. For some $m \in M$, with $m \neq 1$, $gmg^{-1} \in M$. By commutative transitivity, $[g, m] \neq 1$ (otherwise $\langle g, M \rangle$ would be abelian). Hence there is a free group F and a homomorphism $\phi : G \to F$ such that $\phi([g, m]) \neq 1$, so $\phi(m)$, $\phi(g) \neq 1$. But then $1 \neq \phi(gmg^{-1}) \in \phi(M) \cap \phi(g)\phi(M)\phi(g)^{-1}$, hence $\phi(g)$ is in the maximal abelian subgroup of F containing $\phi(m)$, since F is CSA. Hence $[\phi(g), \phi(m)] = 1$, contradicting the choice of ϕ.

It remains to show (4) implies (1). Assume (4). It was left as an exercise in §1 to show that G is then commutative transitive. If G is abelian, then (1) follows from Proposition 5.2, so we assume G is non-abelian. Then if x, $y \in G \setminus \{1\}$, there exists $g \in G$ such that $[gxg^{-1}, y] \neq 1$. This is obvious if $[x, y] \neq 1$, so assume $[x, y] = 1$. Then x, y belong to the same maximal abelian subgroup, say M, of G, and $M = C_G(y)$ since G is commutative transitive. Since G is non-abelian, we can choose $g \in G \setminus M$. Then $M \cap gMg^{-1} = \{1\}$, so $gxg^{-1} \notin M$, hence $[gxg^{-1}, y] \neq 1$.

Now let $x_1, \ldots, x_n$ be non-identity elements of G. Put $y_1 = x_1$, let y_2 be a conjugate of y_1 such that $[y_2, y_1] \neq 1$, let y_3 be a conjugate of x_3 such that $[y_3, [y_2, y_1]] \neq 1$, and so on. We obtain $y_1, \ldots, y_n$ such that y_i is conjugate to x_i and the commutator $[y_n, \ldots, y_1] \neq 1$. There is a homomorphism $\phi : G \to F$, where F is free, such that $\phi([y_n, \ldots, y_1]) \neq 1$, hence $\phi(y_i) \neq 1$, and so $\phi(x_i) \neq 1$ for $1 \leq i \leq n$. This proves (1). □

Thus if F is a non-abelian free group, $F \times F$ is residually free by the remark preceding Prop.5.2, but not fully residually free by Prop.5.2, since it is not commutative transitive.

Our final property of residually free groups is an observation in [7].

Lemma 5.4. *If G is a residually free group, every two-generator subgroup is either free or free abelian.*

Proof. A residually free group is easily seen to be torsion free, so an abelian 2-generator subgroup is free abelian. Suppose $x, y \in G$ do not commute. Then there is a homomorphism $\phi : G \to F$, where F is free, such that $\phi([x, y]) \neq 1$. The subgroup of F generated by $\phi(x)$, $\phi(y)$ is therefore non-abelian, so freely generated by them, since $\phi(x)$, $\phi(y)$ can be transformed by Nielsen transformations to a basis of the subgroup they generate (see 2.1-2.5 in [95]). Hence x, y freely generate a subgroup of G (either by the normal form theorem or the universal property of a free group). □

We proceed to show that a group embeds in some ultrapower of a free group if and only if it is locally fully residually free. For the moment, it is convenient to confine attention to F_2, the free group of rank 2.

Lemma 5.5. *Let G be a locally fully residually free group. Then G embeds in some ultrapower of F_2.*

Proof. Let I be the set of all finite subsets of G. For $E \in I$, let $a_E = \{E' \in I \mid E \subseteq E'\}$. Then $a_E \neq \emptyset$ since $E \in a_E$, and if E, $E' \in I$, then $a_{E \cup E'} \subseteq a_E \cap a_{E'}$. By 5.1, there is an ultrafilter $\mathcal{D}$ on I such that $a_E \in \mathcal{D}$ for all $E \in I$.

For $E \in I$, let G_E denote the subgroup of G generated by E. Since all countable free groups embed in F_2, there is a group homomorphism $\phi_E : G_E \to F_2$ such that, for all $x \in E \setminus \{1\}$, $\phi_E(x) \neq 1$. Extend ϕ_E to G by putting $\phi_E(x) = 1$ for $x \notin G_E$. Now define $\phi : G \to F_2^I/\mathcal{D}$ by $\phi(x) = \langle \phi_E(x) \rangle_{E \in I}$, where $\langle \phi_E(x) \rangle_{E \in I}$ means the equivalence class of $(\phi_E(x))_{E \in I}$ in $F_2^I/\mathcal{D}$. If x, $y \in G$ then ϕ_E restricted to G_E is a group homomorphism for all $E \in a_{\{x,y\}}$ and $a_{\{x,y\}} \in \mathcal{D}$. It follows that ϕ is a group homomorphism.

Suppose $\phi(x) = 1$; then $\phi_E(x) = 1$ for almost all E, that is, for all $E \in A$, where A is some element of $\mathcal{D}$. Then $a_{\{x\}} \cap A \in \mathcal{D}$, so is non-empty. Take $E \in a_{\{x\}} \cap A$; then $\phi_E(x) = 1$ and $x \in E$, so $x = 1$ by the definition of ϕ_E, hence ϕ is an embedding. □

To prove the converse of Lemma 5.5, we need to show that finitely generated subgroups of ultrapowers of F_2 are fully residually free. This is due to Remeslennikov [122], and we shall give his argument. The only difference is that in [122], use was made of the fact that it suffices to show that such subgroups are 2-residually free (Prop. 5.3), which slightly simplifies the notation. Fix a set I and an ultrafilter $\mathcal{D}$ on I, and denote the ultrapower $X^I/\mathcal{D}$ by ${}^*X$. We shall view ${}^*\mathbb{Z}$ as an extension of the ring $\mathbb{Z}$, and identify $\mathrm{SL}_2({}^*\mathbb{Z})$ with ${}^*\mathrm{SL}_2(\mathbb{Z})$, using the map

$$\begin{pmatrix} \langle a_i \rangle & \langle b_i \rangle \\ \langle c_i \rangle & \langle d_i \rangle \end{pmatrix} \mapsto \left\langle \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix} \right\rangle_{i \in I}.$$

It is left to the reader to check this is a well-defined isomorphism of groups. The next lemma refers to the first order language of rings. By this we mean the first order language with equality as its only predicate symbol, binary function symbols $+$ and $\cdot$, unary function symbol $-$, and constant symbols 0, 1. (The proof could be altered to avoid the use of first-order logic, but the result would not be very elegant.)

Lemma 5.6. *Let R be a finitely generated subring of ${}^*\mathbb{Z}$. Then as a ring, R is fully residually $\mathbb{Z}$. That is, given $n \geq 1$ and $x_1, \ldots, x_n \in R \setminus \{0\}$, there is a ring homomorphism $\phi : R \to \mathbb{Z}$ such that $\phi(x_i) \neq 0$ for $1 \leq i \leq n$.*

Proof. There is a short exact sequence

$$J \rightarrowtail \mathbb{Z}[y_1, \ldots, y_k] \overset{\theta}{\twoheadrightarrow} R$$

where the y_i are commuting indeterminates, and J is a finitely generated ideal. Choose generators $f_1, \ldots, f_m$ for J and $g_j \in \mathbb{Z}[y_1, \ldots, y_k]$ such that $\theta(g_j) = x_j$ for $1 \leq j \leq n$. Denoting $z_1, \ldots, z_k$ by $\mathbf{z}$, the sentence

$$\exists z_1 \cdots \exists z_k (f_1(\mathbf{z}) = 0 \wedge \ldots \wedge f_m(\mathbf{z}) = 0 \wedge \neg(g_1(\mathbf{z}) = 0) \wedge \ldots \wedge \neg(g_n(\mathbf{z}) = 0))$$

in the first-order language of rings is then valid in ${}^*\mathbb{Z}$ (assign $\theta(y_j)$ to the variable z_j), hence is valid in $\mathbb{Z}$, by Łoš's Theorem ([11; Ch. 5, Lemma 2.3]). Thus there are integers $u_1, \ldots, u_k$ in $\mathbb{Z}$ such that $f_1(\mathbf{u}) = \cdots = f_m(\mathbf{u}) = 0$ and $g_j(\mathbf{u}) \neq 0$ for $1 \leq j \leq n$. The mapping $\mathbb{Z}[y_1, \ldots, y_k] \to \mathbb{Z}$ which sends y_j to u_j and is the identity on $\mathbb{Z}$ induces a ring homomorphism $\phi : R \to \mathbb{Z}$, with $\phi(x_j) = g_j(\mathbf{u}) \neq 0$, as required. □

Let p be an odd positive prime and let $K = \mathrm{Ker}(\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z}))$. It is well-known that K is a free group. One way to see this is to note that $\mathrm{SL}_2(\mathbb{Z})$ is a free product of two cyclic groups of orders 4 and 6 amalgamating their subgroups of order 2 (see [127; Ch.I, 4.3]). One can obtain explicit generators for the cyclic free factors from the proof, and it is easily checked that K intersects these free factors, and so all their conjugates, trivially. Hence K is free by the subgroup theorem for amalgamated free products (see Ch. I, §4.4, Prop. 18 in [127]). It is easy to see that K is non-abelian, so there is an embedding of F_2 into K, and this has an extension to an embedding of ${}^*F_2$ into ${}^*K$.

Lemma 5.7. *Let G be a finitely generated subgroup of ${}^*F_2$. Then G is fully residually free.*

Proof. By the above we may assume $G \subseteq {}^*K \subseteq \mathrm{SL}_2({}^*\mathbb{Z})$. Let $\{g_1, \ldots, g_n\}$ be a set of generators for G, and for simplicity assume this set is closed under taking inverses. We can write

$$g_j = \begin{pmatrix} 1+pa_j & pb_j \\ pc_j & 1+pd_j \end{pmatrix}$$

where a_j, b_j, c_j and d_j are all in ${}^*\mathbb{Z}$. Let R be the subring of ${}^*\mathbb{Z}$ generated by $\{a_j, b_j, c_j, d_j \,|1 \leq j \leq n\}$. If $g \in G$, an easy induction on the length of a word in the g_j representing g shows that we can write

$$g = \begin{pmatrix} 1+pa & pb \\ pc & 1+pd \end{pmatrix}$$

where $a, b, c, d \in R$. In particular, $G \subseteq \mathrm{SL}_2(R)$.

Suppose $h_1, \ldots, h_k$ are non-identity elements of G. If

$$h_j = \begin{pmatrix} a^j_{11} & a^j_{12} \\ a^j_{21} & a^j_{22} \end{pmatrix}$$

then by Lemma 4.3 there is a ring homomorphism $\phi : R \to \mathbb{Z}$ such that $\phi(a^j_{kl}) \neq 0$ whenever $a^j_{kl} \neq 0$ and $\phi(a^j_{kl}) \neq 1$ whenever $a^j_{kl} \neq 1$. Let $\psi : \mathrm{SL}_2(R) \to \mathrm{SL}_2(\mathbb{Z})$ be the group homomorphism induced by ϕ. Then $\psi(h_j) \neq 1$ for $1 \leq j \leq k$ and for $g \in G$ written as above, we have

$$\psi(g) = \psi \begin{pmatrix} 1+pa & pb \\ pc & 1+pd \end{pmatrix} = \begin{pmatrix} 1+p\phi(a) & p\phi(b) \\ p\phi(c) & 1+p\phi(d) \end{pmatrix} \in K.$$

Thus $\psi|_G$ maps G into K which is free. □

We can now give our characterisation of subgroups of ultrapowers of free groups.

Theorem 5.8. *Let G be any group. The following are equivalent.*
(1) *G is locally fully residually free.*
(2) *G embeds in some ultrapower $^*F_2$ of the free group of rank 2.*
(3) *G embeds in some ultrapower of a free group.*

Proof. (1) and (2) are equivalent by Lemmas 5.6 and 5.7, and obviously (2) implies (3). If F is a free group, then F embeds in some ultrapower of F_2, say $F_2^I/\mathcal{D}_1$, because (1) implies (2). Hence an ultrapower of F, say $F^J/\mathcal{D}_2$, embeds in $(F_2^I/\mathcal{D}_1)^J/\mathcal{D}_2$. But this iterated ultrapower is isomorphic to ${F_2}^K/\mathcal{D}$ where $K = I \times J$ and $\mathcal{D}$ is the ultrafilter $\mathcal{D}_1 \otimes \mathcal{D}_2$. (See Exercises 2.6 and 2.7 in Ch.6 of [11].) Thus (3) implies (2). □

Note that $^*F_2 = F_2^I/\mathcal{D}$ itself is not residually free if $\mathcal{D}$ is non-principal and I is countable, say $I = \mathbb{N}$. For there are many embeddings of the infinite cyclic group $\mathbb{Z}$ into F_2, which extend to embeddings of $^*\mathbb{Z}$ into $^*F_2$, and $^*\mathbb{Z}$ is not residually free (i.e. not residually infinite cyclic, since it is abelian). To see this, note that the image of the embedding $\phi : \mathbb{Z} \to {}^*\mathbb{Z}$ discussed after Lemma 5.1 is not surjective and we can find a positive non-standard integer N, i.e. a positive element of $^*\mathbb{Z} \setminus \phi(\mathbb{Z})$, for example $N = \langle n_i \rangle$, where $n_i = i$ for all $i \in I$. The mapping $n \mapsto n!$ extends as explained earlier to a map $^*\mathbb{Z} \to {}^*\mathbb{Z}$ (putting $n! = 0$ for $n < 0$), and it is an exercise to show $N!$ is divisible by every non-zero standard integer (i.e. by every $n \in \mathbb{Z}$ with $n \neq 0$, identifying n with $\phi(n)$). The additive group of $\mathbb{Q}$ can therefore be embedded in $^*\mathbb{Z}$ by $p/q \mapsto N!(p/q)$, where p, q are (standard) integers with $q \neq 0$. Since $\mathbb{Q}$ is not residually infinite cyclic, neither is $^*\mathbb{Z}$.

There is a model-theoretic consequence of Theorem 5.8. The term "universal theory" is defined, for example, in [59]. For every cardinal r let F_r denote the free group of rank r. It is well-known that, for any cardinals r, s greater than 1, F_r and F_s have the same universal theory in the first-order language of groups. This is because the free group of countably infinite rank embeds in the free group of rank 2, so if r, $s \leq \omega$, F_r embeds in F_s, while if $\omega \leq r \leq s$, then F_r is an elementary substructure of F_s, by a theorem of Vaught (see Theorem 4, §38 in [66]). It follows that F_r, F_s have the same universal theory for all r, $s > 1$ using Lemma 3.7 in [11]. We denote by Φ the set of all universal and existential sentences true in the non-abelian free groups, so that a group G has the same universal theory as the non-abelian free groups if and only if $G \models \Phi$. It is easy, using 5.8, to give a group-theoretic characterisation of such groups.

Theorem 5.9. *Let G be any group. Then the following are equivalent.*
(1) *G is non-abelian and locally fully residually free.*
(2) *G is a model of the set of sentences Φ.*

Proof. Assume (1). To prove (2) we use an observation in the introduction of [58]. By Lemma 5.4, G contains a subgroup isomorphic to F_2. Thus by Theorem 5.8 $F_2 \subseteq G \subseteq {F_2}^J/\mathcal{D}$ for some ultrapower ${F_2}^J/\mathcal{D}$. It follows from Lemma 3.7 in [11] that an existential sentence true in F_2 is true in G, and any existential sentence true in G is true in ${F_2}^J/\mathcal{D}$. Since negations of existential sentences are equivalent to universal sentences, any universal sentence true in ${F_2}^J/\mathcal{D}$ is true in G. By Łoš's Theorem ([11; Ch. 5, Lemma 2.3]), ${F_2}^J/\mathcal{D}$ is elementarily equivalent to F_2, in particular ${F_2}^J/\mathcal{D}$ is a model of Φ. Hence G is a model of Φ.

Conversely if (2) holds, then by [11; Ch. 9, Lemma 3.8], G embeds in some ultrapower of F_2, so by Theorem 5.8 G is locally fully residually free. Since F_2 is non-abelian, the existential sentence $\exists x \exists y(\neg(xy = yx))$ belongs to Φ, so G is non-abelian. □

Theorems 5.8 and 5.9 appear in [34]. Theorem 5.9 generalises the results that non-abelian fully residually free groups are models of Φ [57], and that non-abelian locally free groups are models of Φ ([56], after Question 3).

Exercises. (1) Show that, for an abelian group G, the following are equivalent.

(a) G is locally fully residually free;
(b) G embeds in some ultrapower of the infinite cyclic group;
(c) G is torsion-free.

(To show (a) implies (b), adapt the argument of Lemma 5.5.)

(2) Show that a group has the same universal theory as the infinite cyclic group if and only if it is non-trivial, abelian and locally fully residually free. (Modify the argument of Theorem 5.9.)

We present another argument by Remeslennikov [123], which is an interesting application of the action of $\mathrm{SL}_2(F)$ on a Λ-tree, where F is a field with a valuation, and Λ is the value group.

Theorem 5.10. *If G is a finitely generated fully residually free group, then G is Λ-free for some finitely generated ordered abelian group Λ.*

Proof. We shall not give all the details of the proof. Let $P = \mathbb{Q}(x_1, x_2, y_1, y_2)$ be a field extension of $\mathbb{Q}$ by four algebraically independent elements. There is a valuation v of P given by $v(f/g) = \deg(g) - \deg(f)$, where f, g are polynomials in x_1, x_2, y_1, y_2 and deg means total degree. Define

$$A = \begin{pmatrix} x_1 & x_1x_2 - 1 \\ 1 & x_2 \end{pmatrix} \quad B = \begin{pmatrix} y_1 & y_1y_2 - 1 \\ 1 & y_2 \end{pmatrix}$$

so that A, $B \in \mathrm{SL}_2(P)$. Let H be the subgroup generated by A and B. If w is a reduced word on $\{A^{\pm 1}, B^{\pm 1}\}$, let $\|w\|$ denote the length of the cyclically reduced word corresponding to w. Remeslennikov shows that, if h is the element of H represented by w, then $v(\mathrm{tr}(h)) = -\|w\|$. It follows that H is freely generated by A and B, and that $v(\mathrm{tr}(h)) < 0$ for all $1 \neq h \in H$.

Now let G be a finitely generated fully residually free group, so by Theorem 5.9 G embeds in some ultrapower ${}^*F_2$, which is isomorphic to ${}^*H$, a subgroup of $\mathrm{SL}_2({}^*P)$. By a similar argument to that used in Lemma 5.7, G embeds in ${}^*H \cap \mathrm{SL}_2(L)$, where L is some subfield of ${}^*P$ which is finitely generated over $\mathbb{Q}$. The valuation v extends to a valuation ${}^*v$ of ${}^*P$, and by restriction we obtain a valuation v_1 of L. Now v_1 inherits from v the property that $v_1(\mathrm{tr}(g)) < 0$ for all $1 \neq g \in G$, and it follows from Lemma 4.3.6 and the remarks following it that G acts freely without inversions on the Λ-tree (X_{v_1}, d), where Λ is the value group of v_1. Also, v_1 is identically zero on the multiplicative group of $\mathbb{Q}$, and so Λ is finitely generated (see Ch. VII, §10.3, Corollaire 1 in [19]). □

However, recent progress has improved on Theorem 5.10. This involves the idea of a group admitting exponents from an arbitrary ring, an idea of Lyndon ([93]).

Definition. Let R be a ring. An R-group is a group G together with a mapping $G \times R \to G$, $(g, \alpha) \mapsto g^\alpha$, satisfying the following, for all g, $h \in G$ and all α, $\beta \in R$.

(1) $g^1 = g$
(2) $g^{\alpha+\beta} = g^\alpha g^\beta$.
(3) $g^{\alpha\beta} = (g^\alpha)^\beta$.
(4) $h^{-1}g^\alpha h = (h^{-1}gh)^\alpha$.

Clearly the R-groups, for a given R, form a category. It follows from general results in universal algebra that given a set X, there is a free object in this category on the set X, which we denote by $F(X)^R$. Lyndon established a normal form theorem for free R-groups, where $R = \mathbb{Z}[t_1, \ldots t_n]$, the polynomial ring in finitely many commuting indeterminates. Using this, Pfander [120] has shown that $F_2^{\mathbb{Z}[t]}$, the free $\mathbb{Z}[t]$-group on two generators, embeds in $F_2^I/\mathfrak{D}$, where I is a countable set and $\mathfrak{D}$ is a non-principal ultrafilter on I. He further shows that if G is a finitely generated subgroup of $F_2^{\mathbb{Z}[t]}$, then the Lyndon length function ${}^*L$ on ${}^*F_2$ constructed above, when restricted to G, takes values in a group isomorphic to $\mathbb{Z} \oplus \ldots \oplus \mathbb{Z}$, the direct sum of finitely many copies of $\mathbb{Z}$ with the lexicographic ordering. Using this and the results of Bass mentioned in Example 5 above, he shows that G is finitely presented.

Myasnikov and Kharlampovich ([86], [87]) have shown that a finitely generated group is fully residually free if and only if it embeds in $F_2^{\mathbb{Z}[t]}$. They also have an alternative method of showing that finitely generated subgroups of $F_2^{\mathbb{Z}[t]}$ are finitely presented, and it follows that finitely generated fully residually free groups are finitely presented. A proof of this last result has also been announced by Z. Sela. This answers affirmatively a question recorded in [60].

The motivation for this question is that, in the case of a finitely presented subgroup of an ultraproduct ${}^*F_2$, there is a simpler proof of Lemma 5.7, similar to the proof of Lemma 5.6. (See the proof of Theorem 6 in [56].) The structure of fully residually free groups with at most three generators has been described in [48].

We note that Myasnikov and Kharlampovich use an extra axiom in their definition of an R-group: if g, h commute, then $(gh)^\alpha = g^\alpha h^\alpha$, for all $g, h \in G$ and $\alpha \in R$. All free R-groups in the sense above satisfy this axiom.

Recently, Myasnikov and Kharlampovich [110], [111], have solved two old and well-known problems of Tarski, by showing that the elementary theories of all non-abelian free groups coincide, and that the elementary theory of a free group is decidable. Underlying the work of these authors is an interesting new technique in combinatorial group theory, known as "algebraic geometry over groups". For an account of this, see [8].

Making use of Rips' Theorem, stated in §3, we shall show that $\mathbb{R}$-free groups are locally fully residually free.

Lemma 5.11. *The non-exceptional surface groups are fully residually free.*

Proof. The surface group with presentation $\langle x_1, \ldots, x_g \mid x_1^2 \ldots x_g^2 = 1\rangle$ $(g \geq 4)$ is fully residually free by §7 in [6]. This group has the surface group

$$\langle a_1, b_1, \ldots, a_{g-1}, b_{g-1} \mid [a_1, b_1] \ldots [a_{g-1}, b_{g-1}] = 1\rangle$$

as a subgroup of index 2 (see §2.2 in [154]). Since the class of fully residually free groups is clearly closed under taking subgroups, the group

$$\langle a_1, b_1, \ldots, a_g, b_g \mid [a_1, b_1] \ldots [a_g, b_g] = 1\rangle \quad (g \geq 3)$$

is fully residually free. For $g = 2$, the group $\langle a_1, b_1, a_2, b_2 \mid [a_1, b_1][a_2, b_2] = 1\rangle$ is a free product with amalgamation $F \underset{u=\bar{u}}{*} \bar{F}$, where F is the free group with basis $\{a_1, b_1\}$, $\bar{F}$ the free group on $\{a_2, b_2\}$, isomorphic to F via the isomorphism sending a_1 to b_2 and b_1 to a_2, $u = [a_1, b_1]$ and $\bar{u}$ is the image of u under this isomorphism. Further u is not a proper power in F. It follows from Theorem 8 in [6], using the observations in §1.3 of [7], that this surface group is fully residually free. For $g = 1$, $\langle a_1, b_1 \mid [a_1, b_1] = 1\rangle$ is free abelian of rank 2, so fully residually free by Prop.5.2. Finally, for $g = 0$ (the 2-sphere), the trivial group is free (of rank 0). □

The following well-known result, and the remark following the proof, will be useful in the next chapter.

Corollary 5.12. *Free products of free abelian groups and surface groups are residually finite.*

Proof. Free abelian groups and all the orientable surface groups are residually free by 5.2 and 5.11, so are residually finite since free groups are residually finite (see Prop.5, Ch.1 in [39]). By §2.2 in [154], the non-orientable surface groups of genus at least 2 contain an orientable surface group as a subgroup of index 2, so are residually finite. (We are using the elementary fact that the class of residually finite groups is closed under finite extensions; see 26.12 in [113].) The non-orientable surface group of genus 1 is cyclic of order 2, so is finite. Finally, free products of residually finite groups are residually finite (see the corollary to Prop.22, Chapter 1 in [39]). □

Remark. Abelianising $\langle a_1, b_1, \ldots, a_g, b_g \mid [a_1, b_1] \ldots [a_g, b_g] = 1 \rangle$ gives a free abelian group of rank $2g$, hence the minimum number of generators of this surface group is $2g$, while abelianising $\langle x_1, \ldots, x_g \mid x_1^2 \ldots x_g^2 = 1 \rangle$ gives the direct product of a free abelian group of rank $g-1$ and a cyclic group of order 2, so the minimum number of generators of this group is g.

Assuming Rips' Theorem (§3), it is now easy to deduce the following result.

Proposition 5.13. *An $\mathbb{R}$-free group is locally fully residually free.*

Proof. We need to show that finitely generated $\mathbb{R}$-free groups are fully residually free. By the argument in the proof of Theorem 6 in [6], the free product of two fully residually free groups is fully residually free. By 5.2 and 5.11, finitely generated free abelian groups and the non-exceptional surface groups are fully residually free. The result follows by induction and Rips' Theorem. □

By Theorem 5.8, locally fully residually free groups embed in ultrapowers of F_2, so are tree-free. We therefore have inclusions of classes of groups:

$\mathbb{R}$-free groups $\subset$ locally fully residually free groups $\subset$ tree-free groups.

Moreover the inclusions are strict. For the right-hand inclusion, we noted at the end of §4 that the non-orientable surface group

$$G = \langle x_1, x_2, x_3 \mid x_1^2 x_2^2 x_3^2 = 1 \rangle$$

is tree-free. It is not residually free because in a free group, the equation $a^2b^2c^2 = 1$ implies that a, b and c commute (see [94]). Thus if F is a free group and $\phi : G \to F$ is any homomorphism, $\phi([x_1, x_2]) = 1$, but $[x_1, x_2] \neq 1$ in G. This can be seen by viewing G as the free product of a free group generated by x_1 and x_2 and the cyclic group generated by x_3, amalgamating the cyclic subgroups generated by $x_1^2x_2^2$ and x_3^{-2}. Another example, given in [123], is $\langle t, x, y \mid tx^2y^2t^{-1} = xyx^{-1}y^{-1}\rangle$.

To see that the left-hand inclusion is strict, note that by the remark after Corollary 5.12 and Rips' Theorem, abelianising a finitely generated $\mathbb{R}$-free group results in a finite direct product of infinite cyclic groups and cyclic groups of order 2. Now $G = \langle x_1, x_2, x_3, x_4 \mid x_1^3x_2^3x_3^3x_4^3 = 1\rangle$ is fully residually free by §7 in [6], but is not $\mathbb{R}$-free since its abelianisation contains 3-torsion.

These inclusions can be extended to the right. By Lemma 1.2 and the remarks preceding it, we have further inclusions:

$$\text{tree-free groups} \subset \text{CSA groups} \subset \text{ commutative transitive groups.}$$

The right-hand inclusion is strict, as noted in Remark 5 of [112] (a simple example is the infinite dihedral group). The left-hand inclusion is also strict. In [35] there are examples of word hyperbolic groups of cohomological dimension 2 (so torsion-free) which are not tree-free. These are CSA by Prop.12 in [112].

One other point is the following, noted in [60]. If G is a finitely generated fully residually free group, then the maximal abelian subgroups of G are finitely generated. For G is Λ-free for some finitely generated Λ, by 5.10. If A is an abelian subgroup, then obviously $\ell([g,h]) = 0$ for all g, $h \in A$, where as usual ℓ means hyperbolic length. By 3.2.7, the action of A is abelian and there is a homomorphism $\rho : A \to \Lambda$ such that $\ell(g) = |\rho(g)|$ for all $g \in A$. Since the action is free and without inversions, $\ker(\rho)$ is trivial, hence A embeds in Λ, so is finitely generated. This raises the following questions.

Question 1. If G is a finitely generated tree-free group, is G Λ-free for some finitely generated ordered abelian group Λ?

Question 2. If G is a finitely generated tree-free group, are its abelian subgroups finitely generated?

Before posing a third question, another definition is needed. A *right-ordered group* is a group G with a linear ordering $\leq$ on G such that, if g, h and $k \in G$ and $g \leq h$, then $gk \leq hk$. If, in addition, $g \leq h$ implies $kg \leq kh$ for all g, h and $k \in G$, then G is called an *ordered group*. This is of course a generalisation of the idea of an ordered abelian group. A group is called right-orderable (resp. orderable) if there exists a linear ordering on it making it into a right-ordered (resp. ordered) group.

Now free groups, in particular F_2, are orderable (see [116], Ch.13, Cor. 2.8), and the ordering on F_2 can be extended to an ordering on any ultraproduct of F_2 as indicated in the discussion of ultraproducts above. It is easy to see that this makes any such ultraproduct into an ordered group. Clearly subgroups of ordered groups are ordered. It follows from Theorem 5.8 that locally fully residually free groups are orderable. This suggests the following.

Question 3. Is every tree-free group orderable, or at least right-orderable?

CHAPTER 6. Rips' Theorem

1. Systems of Isometries

We recall the statement of Rips' Theorem:

A finitely generated $\mathbb{R}$-free group G can be written as a free product $G = G_1 * \cdots * G_n$ for some integer $n \geq 1$, where each G_i is either a finitely generated free abelian group or the fundamental group of a closed surface.

We also recall that it is unnecessary to include the possibility that G_i is a free group, because free groups are free products of infinite cyclic groups. The basic idea, due to Rips, is to associate what is known as a system of isometries with an action on an $\mathbb{R}$-tree. To such a system one can associate a group G, and by taking a direct limit, it suffices to show that G has the form claimed in the theorem. Thus the argument involves mainly the study of systems of isometries rather than $\mathbb{R}$-trees, but unfortunately this is just as technical as the study of $\mathbb{R}$-trees. The main prerequisite for understanding this chapter not covered earlier is an understanding of elementary homotopy theory.

The proof to be given is that in [53], with the modification that the treatment of "independent generators" in §3 follows [50]. This has the disadvantage of slightly lengthening the already technical argument, but gives a better result than the one in §5 of [53], given the results at the end of this section. It does not need any ideas from differential geometry, and ultimately (the proof of Theorem 3.1) it is an elegant argument. Use has also been made of [118], which describes the argument in [53] in greater generality, allowing non-free actions. We begin with definitions of the basic ideas.

A *finite tree* is an $\mathbb{R}$-tree which is spanned by a finite set of points. Such a tree is compact and polyhedral (see §2.2), and has only finitely many branch points. A *system of isometries* $S = (D, \{\varphi_i\}_{i=1,\ldots,k})$ consists of a disjoint union D of finitely many finite trees, called the *domain*, and a finite set of isometries $\varphi_i : A_i \to B_i$ (mapping A_i onto B_i), where A_i, B_i are finite subtrees of D. The trees A_i, B_i are called the *bases* of the system, and the φ_i are called the *generators* of S.

Associated to S is a finite 2-complex $\Sigma(S)$ as follows. Take the disjoint union of D and $A_i \times [0,1]$, and identify every $(x, 0) \in A_i \times \{0\}$ with $x \in D$, and $(x, 1) \in A_i \times \{1\}$ with $\varphi_i(x) \in D$. We identify D with its image in $\Sigma(S)$.

Let $\sim$ be the equivalence relation on Σ generated by: $x \sim y$ if for some i with $1 \leq i \leq k$, x, y both belong to the image of $\{a\} \times [0,1]$, where $a \in A_i$. The equivalence classes will be called leaves, and the set $\mathcal{F}$ of leaves will be called the foliation of $\Sigma(S)$. (We shall not consider any other foliations, and will not attempt a general definition of foliation which encompasses both this and the foliations of manifolds mentioned in §4.3.)

We call S connected if $\Sigma(S)$ is connected; since $\Sigma(S)$ is clearly locally path connected, this is equivalent to saying that $\Sigma(S)$ is path connected. We shall show that $\Sigma(S)$ has the homotopy type of a finite 1-complex, but first we consider the fundamental group of a 1-complex. Recall (exercise after Lemma 2.2.4) that such a 1-complex is real(Γ) for some graph Γ. Note that all graphs we shall encounter will be locally finite, so the metric and CW topologies on their realisations coincide (Lemma 2.2.6).

Let Γ be a connected graph and let $v \in V(\Gamma)$. There is an associated group $\pi_1(\Gamma, v)$ defined as follows. Let Π be the set of all closed paths in Γ joining v to v (viewing paths as sequences of edges). For p, $q \in \Pi$, define $p \simeq q$ if p is obtained from q by inserting or deleting a consecutive pair of edges of the form $e, \bar{e}$ for some edge e. Let $\sim$ be the equivalence relation generated by $\simeq$, so $p \sim q$ if and only if there is a finite sequence of paths $p = p_1, \ldots, p_n = q$ such that $p_i \simeq p_{i+1}$ for $1 \leq i \leq n-1$. Let $[p]$ denote the equivalence class of p, and let $\pi_1(\Gamma, v)$ be the set of equivalence classes. One can define a multiplication on $\pi_1(\Gamma, v)$ by $[p][q] = [pq]$, where pq is as defined in §1.3. This makes $\pi_1(\Gamma, v)$ into a group, with identity element the class of the trivial path at v. This is a simple generalisation of one construction of free groups [39; §1.1], to which it reduces when Γ has one vertex.

Let $X = \text{real}(\Gamma)$. It is a consequence of the Simplicial Approximation Theorem that $\pi_1(\Gamma, v)$ is isomorphic to $\pi_1(X, v)$; see [99; Theorem 3.3.9]. To use this argument, one may need to barycentrically subdivide twice to make Γ simplicial (see §1.3), a process described after Lemma 2.4.12. Subdividing once removes edges which are loops and existing multiple edges, and repeating removes multiple edges arising from any original loops. It is left as an exercise to show this does not change $\pi_1(\Gamma, v)$ (up to isomorphism) or X (up to homeomorphism). Also, note that in a simplicial graph, a path can be described as a sequence of vertices, where each consecutive pair is a 1-simplex, and it follows easily that, if Γ is simplicial, then $\pi_1(\Gamma, v)$ is isomorphic to the fundamental group of the corresponding simplicial complex, as defined after Proposition 3.3.7 in [99]. For each edge e, choose a path p_e in real(e) from $o(e)$ to $t(e)$ in such a way that $p_{\bar{e}}$ is the opposite path to p_e. In addition, if $o(e) = t(e)$, we require p_e to traverse real(e) just once (real(e) is a circle, so has infinite cyclic fundamental group). The isomorphism is induced by the map sending a path $e_1, \ldots, e_n$ in Γ to the path $p_1 \ldots p_n$ in X, where $p_i = p_{e_i}$. This is again left as an exercise.

Thus any element of $\pi_1(X, v)$ is represented by a closed path $p_1 \dots p_n$ from v to v, where $e_1, \dots, e_n$ is a path in Γ and p_i is a path from $o(e_i)$ to $t(e_i)$ in real(e_i) for $1 \leq i \leq n$, and if $o(e_i) = t(e_i)$, then p_i goes around real(e_i) once. (The path corresponding to the identity element of $\pi_1(X, v)$ is the constant path at v.) If $o(e_i) \neq t(e_i)$, real(e_i) is a copy of $[0, 1]$, so any two paths from $o(e_i)$ to $t(e_i)$ are homotopic relative to their endpoints. It follows from the isomorphism with $\pi_1(\Gamma, v)$ that two such paths are homotopic rel v if and only if one can be obtained from the other by a sequence of the following moves:

(1) insert or delete a path $p\bar{p}$, where p is a path in real(e) from $o(e)$ to $t(e)$ for some edge e and $\bar{p}$ is the opposite path to p in the topological sense;

(2) replace some p_i by a path in real(e_i), homotopic rel $o(e_i)$, $t(e_i)$ to p_i.

(In (1), insertions must be made so that the result is a path. Also, if the path is just $p\bar{p}$, deletion results in the constant path at v.)

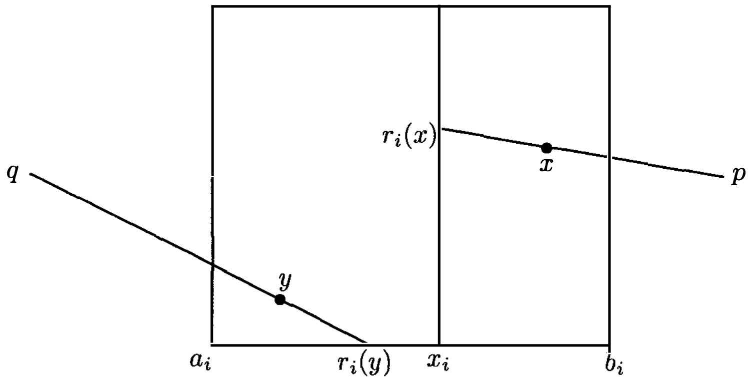
If we drop the requirement that p_i traverses real(e_i) once when $o(e_i) = t(e_i)$, we have to add a third move:

(3) replace p_i, where $o(e_i) = t(e_i)$, by p^n (product of n copies of p) for some integer $n > 0$ and path p from $o(e_i)$ to $t(e_i)$, where p^n is homotopic rel $o(e_i)$ to p_i, or the opposite of this (replace a string of n consecutive paths, all equal to p, by a single path homotopic to p^n).

Assume Γ is connected and let T be a maximal tree in Γ. Then $\pi_1(\Gamma, v)$ has a presentation with $E(\Gamma)$ as the set of generators, and relations $e\bar{e} = 1$ for all $e \in E(\Gamma)$ and $e = 1$ for $e \in E(T)$. The isomorphism is induced by the map sending a path $e_1 \dots e_n$ from v to v in Γ to the product $e_1 \dots e_n$ in the group with this presentation. See Proposition 2.1, Chapter III in [95]. Thus $\pi_1(\Gamma, v)$ is a free group, with a basis obtained by choosing one edge from each pair $\{e, \bar{e}\}$ in $E(\Gamma) \setminus E(T)$.

Lemma 1.1. *The complex $\Sigma(S)$ has the homotopy type of a finite 1-complex, and if $x \in \Sigma(S)$, $\pi_1(\Sigma(S), x)$ is a free group.*

Proof. Choose a point $x_i \in A_i$ for $1 \leq i \leq k$ and let $H_i = (A_i \times \{0, 1\}) \cup (\{x_i\} \times [0, 1])$. This space is a strong deformation retract of $A_i \times [0, 1]$. To see this, first suppose A_i is a segment, which we identify with an interval $[a_i, b_i]$ in $\mathbb{R}$. Choose a point p in $\mathbb{R}^2$ with coordinates $(u, \frac{1}{2})$ where $u > b_i$ and a point q with coordinates $(v, \frac{1}{2})$, where $v < a_i$. The required retraction r_i is given by projection from p and q, as illustrated in the picture which follows.



The homotopy F_i from the identity map to r_i is given simply by $F_i(x,t) = (1-t)x + tr_i(x)$. In general, A_i is a union of finitely many segments (it is homeomorphic to the geometric realisation of a finite simplicial tree), and we obtain r_i and F_i by a succession of retractions similar to that just described. We start with a segment $[a,b]$, not containing x_i, where a is an endpoint of A_i, and retract $[a,b] \times [0,1]$ onto $([a,b] \times \{0,1\}) \cup (\{b\} \times [0,1])$, leaving the rest of $A_i \times [0,1]$ fixed. It is left to the reader to write out the details.

Let U be the disjoint union of the spaces $A_i \times [0,1]$ and D, and let $p : U \to \Sigma(S)$ be the projection map. The maps F_i, together with the projection $D \times [0,1] \to D$ give a homotopy $F : U \times [0,1] \to U$, and $F(-,1)$ is a strong deformation retraction of U onto $E = D \cup H_1 \cup \ldots \cup H_k$. This is because $F(-,0)$ is the identity map on U and $F(y,t) = y$ for all $y \in E$ and $t \in [0,1]$. In turn, F induces a homotopy $G : \Sigma(S) \times [0,1] \to \Sigma(S)$, because $p \times id : U \times [0,1] \to \Sigma(S) \times [0,1]$ is a quotient map (Prop.13.9, Chapter 1 in [22]) and $p(y) = p(z)$ implies $y = z$ or y, $z \in E$, so $pF(y,t) = pF(z,t)$ for $t \in [0,1]$. Further, $r = G(-,1)$ is a strong deformation retraction of $\Sigma(S)$ onto $p(E)$, so $\Sigma(S)$ is homotopy equivalent to $p(E)$. Since D is compact, p maps D homeomorphically onto $p(D)$, and maps the compact space $D \cup \bigcup_{i=1}^{k} \{x_i\} \times [0,1]$ onto $p(E)$. Thus $p(E)$ is obtained from D by attaching the endpoints of the segments $\{x_i\} \times [0,1]$ to D. Hence $p(E)$ is a finite 1-complex, whose 1-cells are the images under p of the segments $\{x_i\} \times [0,1]$ and certain segments contained in D. By the observations preceding the lemma, $\pi_1(\Sigma(S), x)$ is free. □

Remarks. (1) Let X be the space $p(E)$ in the proof just given. Then X is real(Γ) for some finite graph Γ, and from the observations preceding Lemma 1.1, any element of $\pi_1(\Sigma(S), x)$ is represented by a path which is a product of paths α_i contained in a leaf in the image of some $A_j \times [0,1]$, and paths β_i contained in D. Also, D is real(F) for some subforest F of Γ. If S is connected, so is X, and we can choose a maximal tree T in Γ containing F (an exercise after Lemma 1.3.2). Again from the remarks preceding the lemma, this gives a presentation of $\pi_1(\Sigma(S))$ with generators $\overline{\varphi}_1, \ldots, \overline{\varphi}_k$ in one-to-one correspondence with the generators of S, and relations $\overline{\varphi}_i = 1$ for certain values of i (those for which the edge of X corresponding to $A_i \times [0,1]$ belongs to real(T)).

(2) In fact, using the Simplicial Approximation Theorem as in the discussion before Lemma 1.1, any path in the space X is homotopic (relative to its endpoints) to a path which is a product of paths α_i and β_i as in Remark (1). Since X is a deformation retract of $\Sigma(S)$, $\Sigma(S)$ is connected if and only if X is path connected. Clearly X is path connected if and only if any two points of D can be joined by a path in X. It follows that $\Sigma(S)$ is connected if and only if any two components of D are equivalent under the equivalence relation $\simeq$ generated by: $A \simeq B$ if and only if there is a leaf which meets A and B. This equivalence relation is also generated by: $A \simeq B$ if and only if there exists i with $1 \leq i \leq k$ such that $A_i \subseteq A$ and $B_i \subseteq B$.

The Cayley graph of S is the graph whose vertices are the points of D, with a directed edge joining a to $\varphi_i(a)$, for all $a \in A_i$ and all i with $1 \leq i \leq k$. As noted in §1.3, this can be made into a graph (directed graph with involution) by adjoining an opposite edge for each directed edge, and we assume this done. As with Cayley graphs for groups, we assign to each edge a "label": the directed edge joining a to $\varphi_i(a)$ has label φ_i and its opposite edge has label φ_i^{-1}.

The *orbits* of S are the equivalence classes of vertices, where two vertices are equivalent if they are in the same component of this graph. To obtain an alternative description, we define an S-word to be a word in $\varphi_i^{\pm 1}$. Such a word defines a partial isometry of D whose domain is either a finite subtree of D or is empty. We shall commit the usual abuse of notation by failing to distinguish the word and the partial isometry it defines. Two points x, $y \in D$ are in the same orbit if and only if there is an S-word sending x to y. We denote the orbit containing x by $S(x)$, so that $S(x)$ is the set of vertices of the component of the Cayley graph containing x. Note that two points of D are in the same S-orbit if and only if they are in the same leaf of $\mathcal{F}$. In fact, if $x \in D$ and L is the leaf containing x, there is a continuous bijection $f : \text{real}(\Gamma(x)) \to L$, where $\Gamma(x)$ is the component of the Cayley graph containing x. This is left as an exercise-if e is an edge from u to v in $\Gamma(x)$, with label φ_i^{ε}, real(e) is mapped to the image in $\Sigma(S)$ of the segment $\{u\} \times [0,1]$ or $\{v\} \times [0,1]$ in $A_i \times [0,1]$, according as $\varepsilon = 1$ or -1. This bijection need not be a homeomorphism-we shall be dealing, for example, with situations where the orbit of x is dense in D, when this map will not be proper (i.e. the inverse image of a compact set need not be compact). However, if Δ is a finite subgraph of $\Gamma(x)$, then real(Δ) is compact, and $\Sigma(S)$ is Hausdorff (being a 2-complex), so real(Δ) is homeomorphic to its image in L. Further, every path in L has the form $f \circ \beta$ for some unique path β in real($\Gamma(x)$). We give a proof of this.

Removing $A_i \times [1/3, 2/3]$ from $\Sigma(S)$, for each i, leaves an open set U, which retracts onto D, via a retraction r which preserves leaves. Let α be a path in a leaf L corresponding to the component $\Gamma(x)$ of the Cayley graph. Suppose $t \in [0,1]$ and $\alpha(t) \in D$. There exists an interval neighbourhood N_t of t in $[0,1]$ such that $\alpha(N_t) \subseteq U$. Then $(r \circ \alpha)(N_t)$ is path connected and contained in $D \cap L$, so is an $\mathbb{R}$-tree (Lemma 2.4.13). But $D \cap L$ is an orbit, so is countable (there are only countably many S-words). It follows that $(r \circ \alpha)(N_t)$ is a single point, so $\alpha(N_t)$ is contained in the image under f of the vertex $\alpha(t)$ and the edges incident with it in $\Gamma(x)$.

If $t \in [0,1]$ and $\alpha(t) \in A_i \times (0,1)$ for some i, then there is an interval neighbourhood N_t of t such that $\alpha(N_t) \subseteq A_i \times (0,1)$. If $p : A_i \times (0,1) \to D$ is projection onto the first coordinate, then similarly $(p \circ \alpha)(N_t)$ consists of a single point, so $\alpha(N_t)$ is contained in $f(\text{real}(e))$ for some edge e of $\Gamma(x)$.

Since $\Gamma(x)$ is locally finite, for each $t \in [0,1]$, there is an open neighbourhood N_t of t such that $\alpha(N_t) \subseteq f(\text{real}(\Delta_t))$, for some finite subgraph Δ_t of $\Gamma(x)$, hence $f^{-1} \circ \alpha : N_t \to \text{real}(\Gamma(x))$ is continuous. Thus $f^{-1} \circ \alpha : [0,1] \to \text{real}(\Gamma(x))$ is continuous at every point of $[0,1]$, so is continuous. Hence every path in L has the form $f \circ \beta$ for some unique path β in real($\Gamma(x)$), as claimed. Note that any such path β is contained in real(Δ) for some finite subgraph Δ of $\Gamma(x)$. (This is a standard fact about CW-complexes-see, for example, Theorem 8.2, Ch.IV in [22].)

Denote by $\mathcal{C}$ the set of all closed paths in $\Sigma(S)$ which are contained in a single leaf of the foliation $\mathcal{F}$.

Abbreviating $\Sigma(S)$ to Σ, and assuming S connected, choose a basepoint y in Σ. We shall suppress the basepoint and write $\pi_1(\Sigma)$ for $\pi_1(\Sigma, y)$. If l is a closed path in Σ, join it by a path α to y. Then $\bar{\alpha} l \alpha$ represents an element of $\pi_1(\Sigma)$. Replacing α by a different path replaces $\bar{\alpha} l \alpha$ by a conjugate in $\pi_1(\Sigma)$. Thus it makes sense to define $N(S)$ to be the normal subgroup of $\pi_1(\Sigma)$ generated by the set $\mathcal{C}$. If two closed paths are freely homotopic, they represent conjugate elements of $\pi_1(\Sigma)$, so $N(S)$ can be described as the normal subgroup of $\pi_1(\Sigma)$ generated by free homotopy classes of closed paths in $\mathcal{C}$. We define $G(S)$ to be the quotient group $\pi_1(\Sigma)/N(S)$, which is independent of the choice of basepoint y. It is this group which will be the main object of study, and we shall see why shortly. First, we give an algebraic description of $G(S)$, by finding a presentation for it.

Let $l \in \mathcal{C}$, so l is the image under the above map f of a path in real(Δ), for some finite subgraph Δ of a component $\Gamma(x)$ of the Cayley graph. Also, f restricted to real(Δ) is a homeomorphism, and will be suppressed. Choose as basepoint a vertex v of $\Gamma(x)$. Then we can find a path q such that $\bar{q} l q$ represents an element of $\pi_1(\text{real}(\Delta), v)$. By the observations before Lemma

1.1, $\bar{q}lq$ is homotopic to a path p of the form $p = p_1 \dots p_n$, where p_i is a path in real(e_i) from $o(e_i)$ to $t(e_i)$ for some $e_i \in E(\Delta)$. Let $\varphi_{j_i}^{\varepsilon_i}$ be the label on e_i. In the proof of Lemma 1.1, a retraction $r : \Sigma(S) \to \text{real}(\Gamma)$ was defined, where Γ is a finite graph. The path $r \circ p_i$ has the form $\alpha_i p_i' \beta_i$, where α_i, β_i are paths in D, and p_i' is a path in real(e_i') from $o(e_i')$ to $t(e_i')$ for some $e_i' \in E(\Gamma)$. All the paths p_i, α_i, β_i and p_i' lie in the image of $A_{j_i} \times [0,1]$ in $\Sigma(S)$, and p_i, p_i' lie in leaves. Thus

$$r \circ p = \alpha_1 p_1' (\beta_1 \alpha_2) p_2' \dots (\beta_{n-1} \alpha_n) p_n' \beta_n \qquad (*)$$

so $\bar{\alpha}_1 (r \circ p) \alpha_1 = p_1' \gamma_1 p_2' \dots p_n' \gamma_n$, where $\gamma_i = \beta_i \alpha_{i+1}$, with indices taken mod n. Note that γ_i is a path in D and joins two vertices of Γ. Since r and the inclusion map real$(\Gamma) \to \Sigma(S)$ induce inverse isomorphisms on fundamental groups, $N(S)$ is generated as a normal subgroup by the homotopy classes of all closed paths obtained in this way. Further, we obtain such a path for each $x \in D$ and word $\varphi_{j_1}^{\varepsilon_1} \dots \varphi_{j_n}^{\varepsilon_n}$ having x as a fixed point (corresponding to closed paths in the Cayley graph).

Choosing a maximal tree T as in Remark 1 after Lemma 1.1, to obtain a presentation of $\pi_1(\Sigma)$, the conjugacy class of $\bar{\alpha}_1 (r \circ p) \alpha_1$ is that of the group element represented by the word $\overline{\varphi}_{j_1}^{\varepsilon_1} \dots \overline{\varphi}_{j_n}^{\varepsilon_n}$. It follows that $N(S)$ is generated as a normal subgroup by the words $\overline{\varphi}_{j_1}^{\varepsilon_1} \dots \overline{\varphi}_{j_n}^{\varepsilon_n}$ such that $\varphi_{j_1}^{\varepsilon_1} \dots \varphi_{j_n}^{\varepsilon_n}$ has a fixed point in D.

Thus a presentation of $G(S)$ is obtained by taking generators $\overline{\varphi}_1, \dots, \overline{\varphi}_k$ in one-to-one correspondence with the generators of S, with certain relations $\overline{\varphi}_i = 1$, together with $\overline{\varphi}_{j_1}^{\varepsilon_1} \dots \overline{\varphi}_{j_n}^{\varepsilon_n} = 1$ for all words $\varphi_{j_1}^{\varepsilon_1} \dots \varphi_{j_n}^{\varepsilon_n}$ which have a fixed point in D. (In what follows, we shall also need to remember the explicit description of which $\overline{\varphi}_i$ are set equal to 1.)

A generator φ_i is called a *singleton* if A_i consists of a single point. A real interval is called *non-degenerate* if it has more than one point. The system S is called *non-degenerate* if every base is non-degenerate, that is, there are no singletons. We next consider certain operations which can be performed on a connected system of isometries S.

Splitting a base. Let x be a point of a base A_i which is not an endpoint of A_i. Let $\varphi_{i1}, \dots, \varphi_{is}$ be the restrictions of φ_i to the closures $A_{i1}, \dots, A_{is}$ of the connected components of $A_i \setminus \{x\}$. This gives a new system $S' = (D, \{\varphi_{il}, \varphi_j\}_{1 \le l \le s,\ j \ne i})$. The effect on the Cayley graph is to replace the edge from x to $\varphi_i(x)$ with label φ_i by s edges from x to $\varphi_i(x)$ having labels $\varphi_{i1}, \dots, \varphi_{is}$, and to replace all other labels $\varphi_i^{\pm 1}$ with one of $\varphi_{i1}^{\pm 1}, \dots, \varphi_{is}^{\pm 1}$. Hence, in the presentation of $G(S')$ above, $\overline{\varphi}_{i1} = \overline{\varphi}_{il}$ for $1 \le l \le s$. Further, choosing maximal trees suitably, there is an isomorphism from $G(S')$ to $G(S)$, induced by sending all $\overline{\varphi}_{ij}$ to $\overline{\varphi}_i$, and $\overline{\varphi}_j$ to $\overline{\varphi}_j$ for $j \ne i$. (The inverse map is induced by sending $\overline{\varphi}_i$ to $\overline{\varphi}_{i1}$.) Thus $G(S)$ is unchanged by this procedure. Similarly we can split a base B_i.

Splitting the domain. Let x be a point of a component X of D which is not an endpoint of X. Split every base which contains x and for which x is not an endpoint, to obtain a system $S_1 = (D, \{\psi_j : A'_j \to B'_j\}_{1\leq j\leq p})$. Let $D_1, \dots, D_s$ be the closures of the connected components of $D \setminus \{x\}$, let D' be the disjoint union of the D_i and let x_i be the point of D_i corresponding to x, for $1 \leq i \leq s$. Each base A'_j, B'_j is contained in one of the D_i (uniquely determined unless the base is just $\{x\}$), so making arbitrary choices where necessary, each ψ_j determines a partial isometry of D', also denoted by ψ_j. Let Φ be the set of singletons $\{x_1 \mapsto x_2, \dots, x_1 \mapsto x_s\}$. Put $S' = (D', \Phi \cup \{\psi_j\}_{1\leq j\leq p})$. The segments corresponding to the singletons in Φ form a subcomplex of $\Sigma(S')$ which is an $\mathbb{R}$-tree (so contractible). Since the inclusion map of a subcomplex into a CW-complex is a cofibration (Corollary 1.4, Chapter VII in [22]), contracting this $\mathbb{R}$-tree to a point gives a homotopy equivalence of $\Sigma(S')$ with $\Sigma(S_1)$, which induces an isomorphism of $G(S_1)$ and $G(S')$. In terms of the presentations, we can choose a maximal tree so that $\overline{\varphi} = 1$ in $\pi_1(\Sigma(S'))$, for all $\varphi \in \Phi$. These generators can be deleted from the relations of the presentation for $G(S')$, giving the presentation for $G(S_1)$ relative to a suitable maximal tree. Thus $G(S)$ and $G(S')$ are isomorphic.

Removing a singleton. Let S' be obtained by removing a singleton $\varphi : \{a\} \to \{b\}$. The effect of this on $\Sigma(S)$ is to remove a segment with endpoints a and b. If S' is connected, then $\pi_1(\Sigma(S)) = \pi_1(\Sigma(S')) * \mathbb{Z}$, since adding the segment adds a new generator to the presentation of the fundamental group. If $b \in S'(a)$, it is easy to see that the new generator belongs to $N(S)$ and $G(S) = G(S')$. If $b \notin S'(a)$ the new generator does not occur in any non-trivial relation of $G(S')$ and $G(S) = G(S') * \mathbb{Z}$. Otherwise, $\Sigma(S')$ has two components, U_1, U_2; let S_i have domain $D \cap U_i$, and generators those φ_j whose bases belong to U_i. Then $\Sigma(S)$ is obtained by joining $\Sigma(S_1)$ and $\Sigma(S_2)$ by a segment with endpoints a and b. It follows that, in the presentation of $\pi_1(\Sigma(S))$, with any choice of maximal tree, $\overline{\varphi} = 1$ and that $\pi_1(\Sigma(S)) = \pi_1(\Sigma(S_1)) * \pi_1(\Sigma(S_2))$. Further, $G(S) = G(S_1) * G(S_2)$, because a relation for $G(S)$ can be written as a product of relations for $G(S_1)$ and $G(S_2)$, on omitting occurrences of $\overline{\varphi}$.

Now suppose D consists of a single finite tree, and let S' be the system obtained by splitting S at its set of branch points. The domain D' of S' is then a disjoint union of finitely many compact real intervals, which we call a multi-interval. We always assume that a multi-interval is embedded in $\mathbb{R}$ in such a way that the embedding restricted to each component of the domain is an isometry (but impose no further conditions on the embedding).

Suppose G is a group acting on an $\mathbb{R}$-tree (X, d), and K is a finite subtree of X. Suppose further that $\{g_1, \dots, g_k\}$ is a finite set of generators for G. Each g_i defines a partial isometry $\varphi_i : K \cap g_i^{-1}(K) \to K \cap g_i(K)$ by restriction, giving a system $S_0 = (K, \{\varphi_i\}_{1\leq i\leq k})$. We let $S(K)$ be the system obtained by splitting this system at the branch points of K, so that the domain of $S(K)$ is a multi-interval.

Lemma 1.2. *Assume $K \cap g_i(K) \neq \emptyset$ for $1 \leq i \leq k$. Then the group $G(S(K))$ has a presentation with generators $g_1, \ldots, g_k$ and relators the words $g_{i_1}^{e_1} \ldots g_{i_n}^{e_n}$ with $e_i = \pm 1$ such that there exists $x \in K$ with $g_{i_1}^{e_1} \ldots g_{i_n}^{e_n}(x) = x$ and $g_{i_m}^{e_m} \ldots g_{i_n}^{e_n}(x) \in K$ for $1 \leq m \leq n$.*

Proof. The presentation of $G(S_0)$ described earlier is clearly that in the lemma (in Remark 1 after Lemma 1.1, we may choose T so that $K = \text{real}(T)$, so no $\overline{\varphi}_i$ are put equal to 1). By the discussion above, $G(S_0)$ is isomorphic to $G(S(K))$. □

Let $x \in X$, and take an enumeration of G, $G = \{g_1, g_2, \ldots\}$. Let K_n be the subtree spanned by $\{g_1 x, \ldots, g_n x\}$. Replacing X by the subtree spanned by the G-orbit of x (a G-invariant subtree) if necessary, we can assume that X is the union of the increasing sequence $(K_n)_{n \geq 1}$ of finite subtrees. By Lemma 1.2, for sufficiently large n there are natural epimorphisms $\sigma_n : G(S(K_n)) \to G(S(K_{n+1}))$, and the direct limit of the σ_n is the quotient G/H, where H is the subgroup of G generated by the set of elliptic elements. For if F is the free group on $g_1, \ldots g_n$ then the natural maps $F \to G(S(K_n))$ given by Lemma 1.2 induce epimorphisms $G \to G(S(K_n))$ which commute with the σ_n, giving an epimorphism onto the direct limit. The kernel is clearly H.

Now if we can show that each G_n is a free product of surface groups and free abelian groups, it will follow that G/H is isomorphic to G_n for some n, and Rips' Theorem follows. For the minimal number of generators of G_n is bounded, so only finitely many isomorphism classes of groups occur among the G_n. Further, the G_n are finitely generated and by Corollary 5.5.12 are residually finite, so they are Hopfian by a very well-known theorem of Mal'cev (see the corollary to Prop.5, Ch.1 in [39]). Hence σ_n is an isomorphism for n sufficiently large. Thus to establish Rips' Theorem, it suffices to prove the following.

Theorem 1.3. *If S is a connected system of isometries whose domain is a multi-interval, then $G(S)$ is a free product of surface groups and free abelian groups.*

However, the proof of this is a lengthy case-by-case reduction, and occupies the rest of the chapter. Before proceeding we shall introduce some more of the terminology and ideas that are needed, and prove a result needed in §4.

Let S be a system of isometries with domain a multi-interval. We can associate to every generator φ_i which is not a singleton a sign ± 1. By Lemma 1.2.1, φ_i is the restriction to its domain of a unique isometry σ from $\mathbb{R}$ onto $\mathbb{R}$; we put $sgn(\varphi_i) = 1$ if σ is a translation (so preserves the ordering on $\mathbb{R}$) and

$sgn(\varphi_i) = -1$ if σ is a reflection (so reverses the ordering on $\mathbb{R}$). For an S-word $w = \varphi_{i_1}^{e_1} \dots \varphi_{i_l}^{e_l}$, we define $sgn(w) = \prod_{j=1}^{l} sgn(\varphi_{i_j})$ whenever the product on the right is defined. Note that this depends on the embedding of the domain in $\mathbb{R}$.

Suppose $S = (D, \{\varphi_i\}_{i=1,\dots,k})$ (where D is a multi-interval). A restriction of S is a system $\tilde{S} = (D, \{\tilde{\varphi}_i\}_{i=1,\dots,k})$, where $\tilde{\varphi}_i$ is the restriction of φ_i to a subinterval of its domain, for $1 \le i \le k$. The subintervals are not required to be closed, so $\tilde{S}$ need not be a system of isometries in the sense we have defined. However, the notion of Cayley graph and orbit as defined above still make sense, and the component of the Cayley graph containing $x \in D$ is denoted by $\tilde{S}(x)$. There is an obvious notion of $\tilde{S}$-word. If $w = \psi_1 \dots \psi_p$, where each ψ_i is a generator or inverse of a generator of S, the $\tilde{S}$-word $\tilde{\psi}_1 \dots \tilde{\psi}_p$ will be denoted by $\tilde{w}$.

A particular case which will interest us is when each φ_i is the restriction of φ_i to the interior of its domain, which is an open interval. In this case we denote $\tilde{\varphi}_i$ by $\mathring{\varphi}_i$, $\tilde{S}$ by $\mathring{S}$ and $\tilde{S}(x)$ by $\mathring{S}(x)$, and of course, speak of $\mathring{S}$-words. Again we can give a sign ± 1 to every $\mathring{S}$-word $\mathring{\varphi}$ with non-empty domain. We denote by $\mathring{S}^+(x)$ the orbit of x under isometries of sign 1. Thus $y \in \mathring{S}^+(x)$ if and only if $y = w(x)$ for some $\mathring{S}$-word w with $sgn(w) = 1$, that is, w is the restriction of a translation. If w is an S-word, we shall use $\mathring{w}$ for the corresponding $\mathring{S}$-word.

Another important special case is S_{-t}, which is defined to be the system $(D, \{(\varphi_i)_{-t}\}_{i=1,\dots,k})$, where $(\varphi_i)_{-t}$ is the restriction of φ_i to $[a_i + t, b_i - t]$, and $[a_i, b_i] = A_i$ (the domain of φ_i). This is defined for $t \ge 0$ small enough, and is a system of isometries. The orbit of x in this case is denoted by $S_{-t}(x)$. We can obtain $\mathring{S}_{-t}$ from S_{-t} in the same way as $\mathring{S}$ is obtained from S, and replacing S by S_{-t} in the notation $\mathring{S}^+(x)$ above leads to $\mathring{S}^+_{-t}(x)$, the orbit of x under $\mathring{S}_{-t}$-words of sign 1. Readers will be relieved to know that this notation will be used on only a few occasions. If w is an S-word, there is a corresponding S_{-t}-word, denoted by w_{-t}.

We shall now give some technical results from [54] about systems of isometries, for the purpose of showing that for a given domain D which is a multi-interval, $G(S)$ depends only on the orbits of S. We begin by introducing another graph Γ_S associated with the system $S = (D, \{\varphi_i\}_{i=1,\dots,k})$, where D is a multi-interval. Suppose $\varphi_i : [a_i, a_i'] \to [b_i, b_i']$, where $\varphi_i(a_i) = b_i$. The vertices of Γ_S are the $\mathring{S}$-orbits which contain a point of the set $\Omega = \{a_i, a_i', b_i, b_i' \mid 1 \le i \le k\}$. There are $2k$ unoriented edges of Γ_S, one starting at the $\mathring{S}$-orbit of a_i and ending at the $\mathring{S}$-orbit of b_i and one starting at the $\mathring{S}$-orbit of a_i' and ending at the $\mathring{S}$-orbit of b_i' for $1 \le i \le k$. We define D_S to be the number of edges in a maximal forest of Γ_S, that is, the number of vertices minus the number of components of Γ_S (see the exercise preceding Lemma 1.3.3).

Definition. A point $x \in \mathbb{R}$ will be called *bad* for S if there is no $\mathring{S}$-word $\mathring{\varphi}$ with $sgn(\mathring{\varphi}) = -1$ such that $\mathring{\varphi}(x) = x$, but there is an $\mathring{S}$-word of sign -1 sending the interval $(x - u, x)$ to $(x, x + u)$ for some $u > 0$. (Thus x is an endpoint of the domain of such a word.) The system S is called *good* if it has no bad point.

If x is bad for S, so is every point in its $\mathring{S}$-orbit. Also, endpoints of domains and codomains of $\mathring{S}$-words are in $\mathring{S}$-orbits containing a point of the set Ω (this is easy to see by induction on the length of the word). Hence, if S has a bad point, some element of Ω is bad.

Lemma 1.4. *If S is a system of isometries with domain a multi-interval, there is a good system of isometries S' with the same domain, such that S and S' have the same orbits, as do $\mathring{S}$ and $\mathring{S}'$, and $D_S = D_{S'}$.*

Proof. Suppose $x \in \Omega$ is bad for S, and w is an $\mathring{S}$-word of sign -1 sending $(x - u, x)$ to $(x, x + u)$ for some $u > 0$. Take d with $0 < d < u$ and add a generator of sign -1 fixing x, with domain (and codomain) $[x - d, x + d]$, to obtain a system S'. The orbits of S and S' are the same, as are the orbits of $\mathring{S}$ and $\mathring{S}'$. Now the union of the $\mathring{S}$-orbits of the points in Ω is a countable set, so we may choose d so that $x \pm d$ do not belong to the $\mathring{S}$-orbit of a point in Ω. Then $x \pm d$ are not bad for S', so no new bad points are added to Ω in passing from S to S'. The graph $\Gamma_{S'}$ is obtained from Γ_S by adding a new component with one vertex and two edges, so $D_S = D_{S'}$. Repetition of this procedure reduces the number of bad points in Ω, so eventually gives the desired system. □

Lemma 1.5. *Let S be a good system of isometries. Then there exists $\delta > 0$ such that, if p, $q : [0, a] \to \mathbb{R}$ are isometric embeddings with $0 < a < \delta$, and $p(t)$, $q(t)$ are in the same $\mathring{S}$-orbit for all $t \in (0, a)$, then there is an $\mathring{S}$-word sending $p(t)$ to $q(t)$ for all $t \in (0, a)$.*

Proof. For each u, v in the set Ω defined before Lemma 1.4, such that there is an $\mathring{S}$-word sending u to v with sign μ, fix such a word $w_{u,v,\mu}$. Since Ω is finite, there is a $\delta > 0$ such that, for all u, every $w_{u,v,\mu}$ is defined on the interval $(u - \delta, u + \delta)$. Suppose $0 < a < \delta$.

If $q \circ p^{-1}$ is the restriction of a translation of $\mathbb{R}$, define $\lambda = +1$, otherwise put $\lambda = -1$. If w is an $\mathring{S}$-word sending $p(t)$ to $q(t)$, where $t \in (0, a)$, and $sgn(w) = -\lambda$, then w does not map $p(t')$ to $q(t')$ for any other $t' \in (0, a)$. Since there are only countably many $\mathring{S}$-words, it follows that for all but countably many $t \in (0, a)$, there is an $\mathring{S}$-word w sending $p(t)$ to $q(t)$, with $sgn(w) = \lambda$.

Let $w = \mathring{\varphi}_{i_1}^{e_1} \dots \mathring{\varphi}_{i_n}^{e_n}$ be an $\mathring{S}$-word sending $p(t)$ to $q(t)$ for some $t \in (0, a)$, with $sgn(w) = \lambda$. Define $x_0(t) = q(t)$, and inductively define $x_j(t)$ by $\mathring{\varphi}_{i_j}^{e_j}(x_j(t)) = x_{j-1}(t)$, so $x_n(t) = p(t)$. For all $t' \in [0, a]$, put $x_0(t') = q(t')$ and $x_j(t') = x_j(t) + sgn(\mathring{\varphi}_{i_j}^{e_j})(x_{j-1}(t') - x_{j-1}(t))$. Then if $\mathring{\varphi}_{i_j}^{e_j}(x_j(t'))$ is defined, it is equal to $x_{j-1}(t')$. Also, since $sgn(w) = \lambda$, if $w(p(t'))$ is defined, it equals $q(t')$. Let $c(w)$ be the number of $j \in \{1, \dots, n\}$ such that $\mathring{\varphi}_{i_j}^{e_j}$ is not defined on all of the interval $(x_j(0), x_j(a))$.

Now take w with $c(w)$ as small as possible. If $c(w) = 0$, w is the required $\mathring{S}$-word, so suppose $c(w) > 0$. The set of t' such that $w(p(t')) = q(t')$ is $p^{-1}(I)$, where I is the domain of w, so is an open interval, say (t_0, t_1), where $t_0 < t < t_1$, and w is not defined at t_0, nor at t_1. If $t_0 = 0$ and $t_1 = a$ then $c(w) = 0$, contrary to assumption.

Suppose $t_0 > 0$. Let s be the smallest value of j such that $\mathring{\varphi}_{i_{j+1}}^{-e_{j+1}}$ is not defined at $x_j(t_0)$, and let r be the largest value of j such that $\mathring{\varphi}_{i_j}^{e_j}$ is not defined at $x_j(t_0)$. Since w is not defined at t_0, $0 \leq s < r \leq n$. Put $u = x_r(t_0)$, $v = x_s(t_0)$. Also, let $w_1 = \mathring{\varphi}_{i_{s+1}}^{e_{s+1}} \dots \mathring{\varphi}_{i_r}^{e_r}$, a subword of w sending $x_r(t')$ to $x_s(t')$ for $t' > t_0$ in some neighbourhood of t_0, and let $\mu = sgn(w_1)$.

By the choice of r and s, v and $q(t_0)$ are in the same $\mathring{S}$-orbit, as are u and $p(t_0)$. Since $p(t_0)$ and $q(t_0)$ are in the same $\mathring{S}$-orbit, there is an $\mathring{S}$-word w' with $w'(u) = v$. If $sgn(w') = -\mu$, then $(w')^{-1}w_1$ is the restriction of a reflection with centre u, defined on $(u, u+\varepsilon)$ for some $\varepsilon > 0$, so its inverse maps $(u-\varepsilon, u)$ onto $(u, u+\varepsilon)$. Since S is good, there is an $\mathring{S}$-word w'' of sign -1 fixing u, and replacing w' by $w'w''$, we may assume $sgn(w') = \mu$. Also, u, $v \in \Omega$ by the choice of r and s, so there is a previously chosen word $w_{u,v,\mu}$. We can replace the subword w_1 of w by $w_{u,v,\mu}$, to obtain a word w_1 with $c(w_1) < c(w)$, because $a < \delta$, so $w_{u,v,\mu}$ is defined on $(x_r(0), x_r(a))$. This contradicts the choice of w.

Similarly if $t_1 < a$ we obtain a contradiction. This shows that $c(w) = 0$, completing the proof. □

Lemma 1.6. *Let S be a system of isometries. If p, $q : [0, a] \to \mathbb{R}$ are isometric embeddings with $0 < a$, and $p(t)$, $q(t)$ are in the same S-orbit for all but countably many values of t, then $p(t)$, $q(t)$ are in the same S-orbit for all t. Moreover, there exist finitely many S-words $w_1, \dots, w_n$ such that for all $t \in [0, a]$, there is some i with $1 \leq i \leq n$ such that $w_i(p(t)) = q(t)$.*

Proof. We begin by proving the first part. Define λ as in the start of the proof of Lemma 1.5. Using Lemma 1.4, we can assume S is good. Let U be the set of all $t \in [0, a]$ such that there is an $\mathring{S}$-word with sign λ sending $p(t)$ to $q(t)$, and let $F = [0, a] \setminus U$. Given such an $\mathring{S}$-word, it is defined at $p(t')$ for all $t' \in (t-u, t+u) \cap [0, a]$ for some $u > 0$, and for all t' in this interval, it sends $p(t')$ to $q(t')$. Hence U is open in $[0, a]$, so F is closed in $\mathbb{R}$.

Now $q(t)$ is in the same $\mathring{S}$-orbit as $p(t)$ for all but countably many t, since $\mathring{S}(x) = S(x)$ for all but countably many points x (the points in orbits of endpoints of the domain or a base). Arguing as in Lemma 1.5, for all but countably many $t \in [0, a]$, there is an $\mathring{S}$-word w sending $p(t)$ to $q(t)$ with $sgn(w) = \lambda$, that is, F is countable.

If F is finite, there is a partition $0 = t_0 < t_1 < \ldots < t_n = a$ including all points of F, with $t_i - t_{i-1} < \delta$ for $1 \leq i \leq n$, where δ is chosen as in Lemma 1.5. By Lemma 1.5, there exist S-words $w_1, \ldots, w_n$ such that $\mathring{w}_i$ sends $p(t)$ to $q(t)$ for all $t \in (t_{i-1}, t_i)$ and $1 \leq i \leq n$. Then w_i sends $p(t)$ to $q(t)$ for all $t \in [t_{i-1}, t_i]$.

If F is infinite, it still has an isolated point in $(0, a)$, being closed and countable (exercise). Take such a point t_0. Suppose $p(t_0)$ and $q(t_0)$ are in the same $\mathring{S}$-orbit, and take an $\mathring{S}$-word w sending $p(t_0)$ to $q(t_0)$. Since $t_0 \in F$, $sgn(w) = -\lambda$, and w is defined on some interval $(p(t_0 - u), p(t_0 + u))$ for some $u > 0$, so that $w(p(t_0 + t)) = q(t_0 - t)$ for $-u < t < u$. By Lemma 1.5, we can choose u so that there is an $\mathring{S}$-word w' such that $w'(q(t_0 + t)) = p(t_0 + t)$ for $0 < t < u$. Then $w'w$ is the restriction of a reflection with centre t_0, mapping $(p(t_0) - u, p(t_0))$ to $(p(t_0), p(t_0) + u)$. Since S is good, there is an $\mathring{S}$-word w'' fixing $p(t_0)$. Then $sgn(ww'') = \lambda$ and $ww''(p(t_0)) = q(t_0)$, contradicting $t_0 \in F$. Thus $p(t_0)$, $q(t_0)$ are in different $\mathring{S}$-orbits.

Let S' be the system of isometries obtained from S by adding a new generator, with domain $[p(t_0 - v), p(t_0 + v)]$, where $v > 0$, sending $p(t)$ to $q(t)$ for $t \in [t_0 - v, t_0 + v]$. Since t_0 is an isolated point of F, we can choose v so that $p(t)$, $q(t)$ are in the same $\mathring{S}$-orbit for $t \in [t_0 - v, t_0 + v]$ with $t \neq t_0$. By Lemma 1.5, we can also choose v sufficiently small so that there is an $\mathring{S}$-word w sending $(p(t_0), p(t_0 + v))$ to $(q(t_0), q(t_0 + v))$, and the corresponding S-word sends $p(t_0)$ to $q(t_0)$. Thus $p(t)$ and $q(t)$ are in the same S-orbit for all $t \in [t_0 - v, t_0 + v]$, hence S and S' have the same orbits. Write the S-word corresponding to w as $\psi_n \ldots \psi_1$, where ψ_i is a generator or inverse of a generator and put $x_i = \psi_i \ldots \psi_1(p(t_0))$ for $0 \leq 1 \leq n$. For $i < n$, if x_i is an interior point of the domain of ψ_{i+1} then x_i and x_{i+1} are in the same $\mathring{S}$-orbit, otherwise x_i, x_{i+1} are in Ω and their $\mathring{S}$-orbits are joined by an edge in Γ_S. The second possibility occurs for at least one value of i, since $p(t_0)$ and $q(t_0) = x_n$ are in different $\mathring{S}$-orbits, hence $p(t_0)$ and $q(t_0)$ are in distinct $\mathring{S}$-orbits of points in Ω which are joined by a path in Γ_S. Also, the union of the $\mathring{S}$-orbits of points in Ω is countable, hence we can choose v so that $p(t_0 - v)$ and $q(t_0 + v)$ do not belong to this union.

It follows that $\Gamma_{S'}$ is obtained from Γ_S by identifying two vertices in the same component, and adding either two components, each with one vertex and two edges, or a single component with one vertex and two edges, according as $p(t_0 - v)$ and $q(t_0 + v)$ are in the same $\mathring{S}$-orbit or not. Hence $D_{S'} = D_S - 1$.

Now iterate, first using Lemma 1.4 to make S' good, then using the procedure above. This decreases D_S, so terminates eventually (with F finite). It also does not change the S-orbits, so this proves the first part of the lemma.

To prove the second part, first note that in the argument above, the S-word corresponding to w maps $[p(t_0), p(t_0+v)]$ to $[q(t_0), q(t_0+v)]$. By a similar argument there is an S-word mapping $[p(t_0), p(t_0 - v)]$ to $[q(t_0), q(t_0 - v)]$

(and sending $p(t_0)$ to $q(t_0)$). Inspecting the proof of Lemma 1.4, we see that whenever a new generator φ is introduced, either using Lemma 1.4 or the procedure above, φ has domain a non-degenerate interval I which is the union of two closed subintervals, on each of which φ agrees with some S-word (where S is the system to which φ is added). It follows by induction on the number of new generators introduced, that the domain of each new generator φ is a finite union of non-degenerate closed intervals, on each of which φ agrees with some S-word (S being the original system).

When the procedure stops, the words w_i obtained above in the case F is finite, defined on $[p(t_{i-1}), p(t_i)]$, may involve the new generators. Suppose $w_i = w'\varphi w$, where φ is a new generator or its inverse. Then there is a partition $s_0 < s_1 < \ldots < s_m$ of $w([p(t_{i-1}), p(t_i)])$ such that on $[s_{j-1}, s_j]$, φ agrees with some S-word u_j. Now let $I_j = [w^{-1}(s_{j-1}), w^{-1}(s_j)]$ (giving a partition of $[p(t_{i-1}), p(t_i)]$ into m subintervals) and replace w_i by $w'u_jw$ on I_j. Repetition of this removes all occurrences of new generators and gives the required finite set of S-words. □

Proposition 1.7. *Let S, S' be connected systems of isometries whose domains are multi-intervals, with S connected and suppose S' is obtained from S by adding a new generator $\varphi : A \to B$, such that x and $\varphi(x)$ are in the same S-orbit for all $x \in A$. Then S' is connected and $G(S)$ is isomorphic to $G(S')$.*

Proof. Recall that $G(S)$ has a presentation $\langle X \mid R\rangle$, where $X = \{\overline{\varphi}_1, \ldots \overline{\varphi}_k\}$ is a set in one-to-one correspondence with the set of generators $\{\varphi_1, \ldots \varphi_k\}$ of S. If w is an S'-word, let $\overline{w}$ denote the word in the generators obtained by replacing all occurrences of the symbol φ by $\overline{\varphi}$. Then R consists of relators $\overline{\varphi}_i$ for certain values of i, together with $\overline{w}$ for all S-words w having a fixed point. Clearly S' is connected, and from the description of which $\overline{\varphi}_i$ are set equal to 1, there is a corresponding presentation for $G(S')$, namely $\langle X, \overline{\varphi} \mid R \cup R'\rangle$, where R' consists of relators $\overline{w}$, where w is an S' word having a fixed point and at least one occurrence of $\overline{\varphi}^{\pm 1}$. We shall convert this to the presentation $\langle X \mid R\rangle$ by Tietze transformations, showing $G(S') \cong G(S)$.

If $\overline{w} \in R'$ has more than one occurrence of $\overline{\varphi}^{\pm 1}$, consider one such occurrence and write $\overline{w} = \overline{w}_1\overline{\varphi}^e\overline{w}_2$, where $e = \pm 1$. Let a be a fixed point for w, let $b = w_2(a)$ and let $c = \varphi^e(b)$. By assumption, $c = w_3(b)$ for some S-word w_3, and $\overline{\varphi}^{-e}\overline{w}_3 \in R'$, so we can replace $\overline{\varphi}^e$ by $\overline{w}_3$ in $\overline{w}$. Repetition replaces $\overline{w}$ by a word having only one occurrence of $\overline{\varphi}^{\pm 1}$, and we can do this for each such w. On removing cyclic conjugates and inverses, we obtain a presentation $\langle X, \overline{\varphi} \mid R \cup R''\rangle$, where R'' consists of all words in R' of the form $\overline{w}^{-1}\overline{\varphi}$, where w is an S-word and $w^{-1}\varphi$ has a fixed point.

By Lemma 1.6, We can find a partition $t_0 < t_1 < \ldots < t_n$ of A and S-words $w_1, \ldots, w_n$ such that w_i agrees with φ on $[t_{i-1}, t_i]$. If $\overline{w}^{-1}\overline{\varphi} \in R''$, $w^{-1}\varphi$ has a fixed point a, so $\varphi(a) = w(a) = w_i(a)$ for some i, hence $\overline{w}^{-1}\overline{w}_i \in R$.

Therefore the presentation can be transformed into $\langle X, \overline{\varphi} \mid R, \overline{w}_i^{-1}\overline{\varphi}\ (1 \leq i \leq n)\ \rangle$.

Now for $1 \leq i \leq n-1$, $w_i(t_i) = \varphi(t_i) = w_{i+1}(t_i)$, hence $\overline{w}_i^{-1}\overline{w}_{i+1} \in R$, so we can transform the presentation into $\langle X, \overline{\varphi} \mid R, \overline{w}_1^{-1}\overline{\varphi}\rangle$. A final Tietze transformation gives $\langle X \mid R\rangle$, as required. □

Corollary 1.8. *If $S = (D, \{\varphi_i\}_{i=1,\ldots,k})$, $S' = (D, \{\psi_i\}_{i=1,\ldots,l})$ are systems of isometries having the same multi-interval D as domain, and having the same orbits, then S is connected if and only if S' is, in which case $G(S)$ and $G(S')$ are isomorphic.*

Proof. Since the orbits of a system of isometries are the intersections of D with the leaves, the first part follows from Remark (2) after Lemma 1.1. Assume S and S' are connected. Let $S'' = (D, \{\varphi_i\}_{i=1,\ldots,k} \cup \{\psi_i\}_{i=1,\ldots,l})$. Then S'' can be obtained from S by successively adding the generators ψ_i, and by Proposition 1.7, $G(S) \cong G(S'')$. By symmetry, $G(S') \cong G(S'')$. □

2. Minimal Components

From now on all systems of isometries will be assumed to have domain a multi-interval. Let S be a non-degenerate system of isometries with domain D. Recall that $\mathring{S}$ is obtained by replacing each generator of S with its restriction to the interior of its domain. An $\mathring{S}$-orbit is called *singular* if it consists of an endpoint of D, or contains an endpoint of a base. Otherwise it is called *regular*. Denote by E the union of the finite singular $\mathring{S}$-orbits and the finite $\mathring{S}$-orbits containing the unique fixed point of some reflection (i.e. a word in the generators of $\mathring{S}$, whose domain is a non-degenerate interval, having a fixed point and of sign -1). Our first theorem is a major result, taken from [53], where it is noted that it is a version of a theorem of Imanishi. First, a simple remark is needed.

Remark. Let x be an element of D with infinite $\mathring{S}$-orbit and let l be a positive real number. Then we can choose y, z in the orbit $\mathring{S}(x)$ such that $z \in \mathring{S}^+(y)$ and $0 < |y - z| < l$.

To see this, since $\mathring{S}(x)$ is infinite, it has an accumulation point in the compact set D. Hence we can find an open interval J of length less than l containing infinitely many points of $\mathring{S}(x)$. Choose three points u, v and $w \in J \cap \mathring{S}(x)$, and $\mathring{S}$-words g, h such that $v = gu$ and $w = hv$. Then at least one of g, h, gh has sign 1, so we can take for y and z two of the points u, v and w.

Theorem 2.1. *In this situation, E is finite and $D \setminus E$ is a disjoint union of open $\mathring{S}$-invariant sets $U_1, \ldots, U_r$, such that, for each i, exactly one of the following holds.*

(a) *U_i is a family of finite orbits, that is:*
  (i) *U_i consists of intervals of equal length;*
  (ii) *each interval contains only points belonging to finite $\mathring{S}$-orbits;*
  (iii) *each interval contains exactly one point of each $\mathring{S}$-orbit contained in U_i.*

(b) *U_i is a minimal component, that is, every $\mathring{S}$-orbit contained in U_i is dense in U_i.*

Proof. The union E' of the finite singular $\mathring{S}$-orbits is clearly finite since there are only finitely many endpoints of D and the bases. Hence $D \setminus E'$ is a finite union of intervals open in D.

Let x be a point of a finite regular $\mathring{S}$-orbit, and let d be the minimum distance of a point of $\mathring{S}(x)$ to an endpoint of D or a base, so $d > 0$. Let $y \in (x - d, x + d)$ and suppose w is an $\mathring{S}$-word. Then $w(x)$ is defined if and only if $w(y)$ is defined, in which case $|w(x) - w(y)| = |x - y| < d$. This follows by induction on the length of w. It follows that $\mathring{S}(y)$ is finite, equal to either $|\mathring{S}(x)|$ or $2|\mathring{S}(x)|$, depending on whether or not x is the centre of a reflection and $|\mathring{S}(y)|$ is a regular orbit.

There is therefore a maximal open interval I containing x such that, for $y \in I$ and all $\mathring{S}$-words w, $w(y)$ is defined if and only if $w(x)$ is defined, and such that all points of I are in finite regular $\mathring{S}$-orbits (namely, the union of all such intervals). Further, for all $y \in I$, $|\mathring{S}(y)| \leq 2|\mathring{S}(x)|$. Let $I = (a, b)$, where $a < b$). Then we claim the $\mathring{S}$-orbit of a is finite.

For suppose not, and put $n = 2|\mathring{S}(x)|$. There are $\mathring{S}$-words, say $w_1, \ldots, w_m$ such that $w_1(a), \ldots, w_m(a)$ are pairwise distinct and $m > n$. Choose $e > 0$ such that w_i is defined on $(a - 2e, a + 2e)$ for $1 \leq i \leq m$, $a + e \in I_x$ and $w_1(a + e), \ldots, w_m(a + e)$ are distinct. Then $|\mathring{S}(a + e)| > n$, a contradiction since $a + e \in I$. Now a is not in a regular $\mathring{S}$-orbit, otherwise if d' is the minimal distance from a point of $\mathring{S}(a)$ to an endpoint of D or of a base, the argument already given shows that $I \cup (a - d', a + d')$ contradicts the maximality of I. Hence $a \in E'$, and similarly $b \in E'$. Thus I is the component of $D \setminus E'$ containing x.

If $x \in D \setminus E'$ is the centre of a reflection, then x is the midpoint of I. For otherwise, if w is a reflection fixing x, it is defined on I, and will send some point of I to an endpoint of I, which is impossible since points of I are in regular orbits. Hence E is finite, and since it is $\mathring{S}$-invariant, $D \setminus E$ decomposes into finitely many disjoint $\mathring{S}$-invariant sets, each of which is a finite union of intervals open in D.

For $x \in D \setminus E$, let I_x be the component of $D \setminus E$ containing x. Then I_x is either the component of $D \setminus E'$ containing x, or a half-interval of this component, and consists of points in finite regular $\mathring{S}$-orbits. Further, if w is an $\mathring{S}$-word and $w(x)$ is defined, then $w(y)$ is defined for all $y \in I_x$. It follows that

$wI_x = I_{wx}$. (Since wI_x is an interval and does not meet $E = wE$, $wI_x \subseteq I_{wx}$, and we have equality since w^{-1} is defined on I_{wx}, being defined at $w(x)$.) Hence there are only finitely many intervals of the form wI_x. Further, if $I_x \cap wI_x \neq \emptyset$, then $I_x = wI_x$. Otherwise, I_x contains an endpoint of wI_x, which is in E, a contradiction. If $I_x = wI_x$, then if w is the restriction of a reflection to I_x, it must fix the midpoint of I_x, while if w is the restriction of a translation, it must be the identity map. Since the mid-point of I_x is not in E and has finite $\mathring{S}$-orbit, w must be the identity map. It follows that each interval in the collection $\{wI_x\}$, where w runs through $\mathring{S}$-words with wx defined, contains exactly one point of each $\mathring{S}$-orbit which meets $\bigcup_w wI_x$.

To prove the theorem, it suffices to show that, if an infinite $\mathring{S}$-orbit L meets a component J of $D \setminus E$, then L is dense in J. We first show that, if $D \setminus \overline{L}$ is non-empty, there exists a real $\delta > 0$ such that every component of $D \setminus \overline{L}$ has length at least δ.

To see this, choose $\delta > 0$ such that

(i) any two points of E are at distance greater than δ apart.

(ii) if a is an endpoint of a base A_i, then the distance between a and $\mathring{A}_i \cap L$ is either 0 or greater than δ.

Suppose J is a component of $D \setminus \overline{L}$ of length $l < \delta$. Then we claim that at least one endpoint of J has an infinite $\mathring{S}$-orbit. For if an endpoint x has finite $\mathring{S}$-orbit, it cannot be in $D \setminus E$. Otherwise, by the above, there is an open interval I_x containing x consisting of points with finite $\mathring{S}$-orbits, hence containing no points of L, and so no points of $\overline{L}$. Thus $I_x \cup J$ is an interval contained in $D \setminus \overline{L}$ and strictly containing J, a contradiction. Thus if both endpoints of J have finite orbits, they are in E, at distance $l < \delta$ apart, contradicting the choice of δ.

Let x be an endpoint of J with infinite $\mathring{S}$-orbit. By the remark preceding the theorem, we can choose two points y, z in the orbit $\mathring{S}(x)$ such that $z \in \mathring{S}^+(y)$ and $0 < |y - z| < l$. Interchanging y and z if necessary, we can find an $\mathring{S}$-word $w = \mathring{\varphi}_{j_q}^{e_q} \dots \mathring{\varphi}_{j_1}^{e_1}$ with $w(x) = y$ and which sends some subinterval of J with endpoint x to a subinterval of $[y, z]$ with endpoint y. Let x' be the point of J at distance $|y - z|$ from x.

Now x is a boundary point of $\overline{L}$, which is closed, so $x \in \overline{L}$. It follows that $z \in \overline{L}$ and that $\mathring{S}(z) \subseteq \overline{L}$. Hence $x' \notin \mathring{S}(z)$ since $J \subseteq D \setminus \overline{L}$. Therefore, there exists i such that $\mathring{\varphi}_{j_i}^{e_i} \dots \mathring{\varphi}_{j_1}^{e_1}$ is not defined at x'. Take the smallest possible value of i, and let x'' be the point of $(x, x']$ closest to x such that $\varphi_{j_i}^{e_i} \dots \varphi_{j_1}^{e_1}$ is not defined at x''. Then $\mathring{\varphi}_{j_{i-1}}^{e_{i-1}} \dots \mathring{\varphi}_{j_1}^{e_1}(x'')$ (meaning x'' if $i = 1$) is defined ($\mathring{\varphi}_{j_{i-1}}^{e_{i-1}} \dots \mathring{\varphi}_{j_1}^{e_1}$ is defined at x and x' and its domain is an interval), is not in $\overline{L}$ (since $x'' \in J$), is an endpoint of the domain of $\mathring{\varphi}_{j_i}^{e_i}$, and is at distance less than δ from $\mathring{\varphi}_{j_{i-1}}^{e_{i-1}} \dots \mathring{\varphi}_{j_1}^{e_1}(x) \in \overline{L}$. This contradicts the choice of δ, and establishes the existence of δ with the desired property, that every component of $D \setminus \overline{L}$

has length greater than δ.

To finish the proof, let L be an infinite $\mathring{S}$-orbit meeting a component J of $D \setminus E$. Suppose $L \cap J$ is not dense in J. Then the boundary of $\overline{L} \cap J$ in J is non-empty since J is connected. Let L' be an $\mathring{S}$-orbit meeting the boundary of $\overline{L} \cap J$, and let x be a point in the intersection of L' with this boundary. Thus $x \in J \subseteq D \setminus E$, so if L' is finite, then by the above x has a neighbourhood in which all points belong to finite orbits. hence L' is infinite. We can therefore choose $\delta > 0$ such that every component of $D \setminus \overline{L'}$ has length at least δ. Further, if w is a word defined at x, then it is defined in a neighbourhood of x, hence every neighbourhood of $w(x)$ contains points of L, so $L' \subseteq \overline{L}$, which implies $\overline{L'} \subseteq \overline{L}$. Moreover, every neighbourhood of x contains points not in $\overline{L}$, so the same is true of every point y in L'.

By the remark preceding the theorem, we can find $y, z \in L'$ such that $0 < |y - z| < \delta$ and $z \in \mathring{S}^+(y)$. Then by the choice of δ, we have $[y, z] \subseteq \overline{L'}$, hence $[y, z] \subseteq \overline{L}$. Since y, z are in the same $\mathring{S}^+$-orbit (i.e. $y = wz$ for some $\mathring{S}$-word w which is the restriction of a translation), there is an open interval containing y contained in $\overline{L}$, contradicting the previous paragraph. This completes the proof. □

The following remark will be useful later, in §4.

Remark. If S is non-degenerate and K is an interval of positive length contained in a minimal component U, then there exists an integer N such that every $x \in \overline{U}$ can be mapped into K by an S-word of length at most N.

Proof of remark. Let a be a point in the boundary of U, so $a \in E$ (the set in Theorem 2.1). Let J be a non-degenerate interval with a as a boundary point and $J \setminus \{a\} \subseteq U$. Let w be an S-word defined on a non-degenerate interval $[a, b] \subseteq J$. Then $w(a, b) = \mathring{w}(a, b) \subseteq U$, so either $w(a) \in U$ or $w(a)$ is in the boundary of U, hence in E. We claim that, for at least one such word w, $w(a) \in U$.

Suppose not, so all such images are in the finite set E. We show this leads to a contradiction. For $x \in E$, let D_x be the minimum distance from x to a base not containing x ($D_x = \infty$ if no such base exists), and put $D = \min_{x \in E} D_x$. Choose a real δ such that $0 < |\delta| < D$ and $(a, a+\delta) \subseteq J$. Suppose $y \in (a, a+\delta)$ and u is an S-word; then if $u(y)$ is defined, it follows easily by induction on the length of u that $u(a)$ is defined, $u(a) \in E$ and of course $|u(y) - u(a)| = |y - a|$. Since there are only finitely many choices for $u(a)$, there are only finitely many for $u(y)$, hence the S-orbit, and so the $\mathring{S}$-orbit of y is finite, contradicting that $\mathring{S}$-orbits in U are dense.

Now let $I = \mathring{K}$. Let u be an S-word defined on some non-degenerate interval $[a, b] \subseteq J$ with $u(a) \in U$. Since $\mathring{S}$-orbits in U are dense, there is an S-word u' defined on some open interval containing $u(a)$ sending $u(a)$ into I.

It follows that there is some non-degenerate interval $(a, b') \subseteq U$ such that $u'u$ is defined on $[a, b')$ and maps it into I.

Suppose V is a component of U, with endpoints a_1, a_2. By what we have proved, there are non-degenerate intervals $[a_i, b_i)$ with $(a_i, b_i) \subseteq V$ and S-words v_i, for $i = 1, 2$, such that $v_i[a_i, b_i) \subseteq I$. Then $[b_1, b_2] \subseteq V \subseteq U$, so for every $x \in [b_1, b_2]$, there is an S-word w such that $\mathring{w}(x) \in I$ (again because $\mathring{S}$-orbits of U are dense). Thus $[b_1, b_2] \subseteq \bigcup_w \mathring{w}^{-1}I$. Since $[b_1, b_2]$ is compact and $\mathring{w}^{-1}I$ is open, we can find finitely many S-words $w_1, \ldots, w_n$ such that $[b_1, b_2] \subseteq \mathring{w}_1^{-1}I \cup \ldots \cup \mathring{w}_n^{-1}I$. It follows that

$$\overline{V} = [a_1, a_2] \subseteq v_1^{-1}I \cup v_2^{-1}I \cup w_1^{-1}I \cup \ldots \cup w_n^{-1}I.$$

Since U has only finitely many components, there is a finite set W of S-words such that $\overline{U} \subseteq \bigcup_{w \in W} w^{-1}I$. Thus for every $x \in \overline{U}$, there exists $w \in W$ such that $w(x) \in I \subseteq K$. We can therefore take N to be the maximum length of a word in W. □

Using Theorem 2.1 we can make two definitions which will be significant in what follows. The notation is taken from Theorem 2.1.

Definition. A non-degenerate system of isometries S is called *minimal* if $D \setminus E$ consists of a single minimal component.

Definition. A system of isometries S is *pure* if it is connected, non-degenerate, and E consists only of endpoints of D.

The sets U_i occurring in Theorem 2.1 are unions of the components of $D \setminus E$. It follows that if S is pure, then either $D \setminus E$ is a family of finite orbits or S is minimal.

Theorem 2.2. *If S is connected, then there exist pure systems of isometries $S_1, \ldots, S_r$ and a free group F such that $G(S) \cong G(S_1) * \ldots * G(S_r) * F$.*

Proof. We use the operations on systems of isometries described in §1. First, remove all singletons from S to get a non-degenerate system S'. Define E as before, and then split the domain at each point of $E \cap \mathring{D}$, but without adding the singletons used in this operation. We get pure systems $S_1, \ldots, S_r$ corresponding to the sets $U_1, \ldots, U_r$ associated to S' by Theorem 2.1. The result follows by the description of the effect of these operations on $G(S)$ given in §1 (viewing the second set of operations as splitting the domain at the points in question, which does not change $G(S)$, followed by further removal of singletons). □

Using this we can deal with one case in the proof of Theorem 1.3.

Proposition 2.3. *If S is a connected system of isometries and all orbits of S are finite, then $G(S)$ is free.*

Proof. The systems $S_1, \ldots, S_r$ in Theorem 2.2 also have all orbits finite, so we can assume S is pure, in which case we show that $G(S) = 1$. If S is pure, then the components of $D \setminus E$ (in the notation of Theorem 2.1) are the components of $\mathring{D}$, and every base is the closure of a component of $\mathring{D}$, so every base has the same length, say e, by Theorem 2.1. Thus $\Sigma(S)$ is obtained by glueing copies of $[0, e] \times [0, 1]$ to D, and by Theorem 2.1, there is a leaf L passing through the midpoint of every component of $\mathring{D}$, consisting of these points and the images of the copies of $\{\frac{1}{2}e\} \times [0, 1]$. There is an obvious strong deformation retraction of $[0, e] \times [0, 1]$ onto $\{\frac{1}{2}e\} \times [0, 1]$ (given by the homotopy $((x, t), s) \mapsto ((1 - s)x + \frac{1}{2}se, t)$, for $x \in [0, e]$, s, $t \in [0, 1]$), and it is easy to see this induces a strong deformation retraction of $\Sigma(S)$ onto L (because E consists only of endpoints of D). Hence any closed path in $\Sigma(S)$ is homotopic to a closed path in L, so $G(S) = 1$. □

To prove Theorem 1.3, it suffices to prove the following.

Theorem 2.4. *If S is a pure system of isometries, then $G(S)$ is either a free group, a free abelian group, or a surface group.*

Further, in view of Prop. 2.3, we need only prove this for pure, minimal systems. The next part of the proof deals with a special case, the homogeneous case, where, roughly speaking, the orbits coincide with orbits of a dense subgroup of $\mathrm{Isom}(\mathbb{R})$. In this context, a subgroup P of $\mathrm{Isom}(\mathbb{R})$ is *dense* if $P \cap T$ is dense in $\mathrm{Isom}(\mathbb{R}) \cap T$, where T is the subgroup of translations in $\mathrm{Isom}(\mathbb{R})$. (Recall from Lemma 1.2.1 that T is isomorphic to $\mathbb{R}$, and it is topologised accordingly.)

In a non-degenerate system S, every generator φ_i is the restriction of a unique element of $\mathrm{Isom}(\mathbb{R})$, by Lemma 1.2.1, and we let Q be the subgroup of $\mathrm{Isom}(\mathbb{R})$ generated by these elements. If x, $y \in D$, where D is the domain of S, then obviously if x, y are in the same S-orbit, they are in the same Q-orbit, but the converse is false in general; it is left as an exercise to give counterexamples.

Definition. Let S be a non-degenerate system of isometries. Let U be a minimal component of S (see Theorem 2.1). Then U is called *homogeneous* if there exists an interval $J \subseteq U$ of positive length, and a dense subgroup P of $\mathrm{Isom}(\mathbb{R})$, such that for x, $y \in J$, x and y are in the same $\mathring{S}$-orbit if and only if they are in the same P-orbit.

To investigate the subgroup P in this definition, we first make the following observation. Let G be a dense subgroup of $\mathrm{Isom}(\mathbb{R})$. Take an open interval $I \subseteq \mathbb{R}$ and $\varepsilon > 0$. Then G is generated by its translations of amplitude less than ε and its reflections with centre in I. This is left as an exercise. Also, using Lemma 1.2.1, it is easy to see that, if G is countable, and α is an element of Isom($\mathbb{R}$) such that x and $\alpha(x)$ are in the same G-orbit for uncountably many x, then $\alpha \in G$.

Now let P, J be as in the definition and let Q be as defined above. Choose $x \in \mathring{J}$ and $\varepsilon > 0$ such that $(x-\varepsilon, x+\varepsilon) \subseteq J$. If $p \in P$ is a translation of amplitude at most $\varepsilon/2$, then y and py are in the same $\mathring{S}$-orbit for all $y \in (x-\varepsilon/2, x+\varepsilon/2)$, so in the same Q-orbit. Hence $p \in Q$ (Q is finitely generated, so countable). Similarly, if $p \in P$ is a reflection with centre $x \in \mathring{J}$, and ε is as above, y and py are in the same Q-orbit, for all $y \in (x-\varepsilon, x+\varepsilon)$, so $p \in Q$. Hence $P \leq Q$. Now Isom($\mathbb{R}$) is a split extension of $\mathbb{R}$ by a cyclic group of order 2. It follows that any finitely generated subgroup of Isom($\mathbb{R}$) has all its subgroups finitely generated (exercise). Hence P is finitely generated, in particular, P is countable.

Call a non-degenerate interval J' contained in U *good* for P if, for all x, $y \in J'$, x and y are in the same $\mathring{S}$-orbit if and only if they are in the same P-orbit. An argument similar to that just given shows that if J is good for P and for P', where P' is a dense subgroup of Isom($\mathbb{R}$), then $P \leq P'$, so by symmetry, $P = P'$.

Suppose J_1 and J_2 are subintervals of U which are good for P such that $J_1 \cap J_2$ is a non-degenerate interval. Since $\mathring{S}$-orbits in U are dense and P is a dense subgroup of Isom($\mathbb{R}$), it follows easily that $J_1 \cup J_2$ is good for P.

Let J be a good interval for P, and let I be the union of all subintervals J' of U containing J which are good for P. Clearly I is good for P. We claim that I is a component of U.

For otherwise, some endpoint of I, say a, is in U, and since $\mathring{S}$-orbits in U are dense, there is an $\mathring{S}$-word w such that $w(a) \in \mathring{I}$. We can find $\varepsilon \neq 0$ such that $(a-\varepsilon, a+\varepsilon) \subseteq U$, $(a-\varepsilon, a) \subseteq I$, w is defined on $(a-\varepsilon, a+\varepsilon)$ and $w(a-\varepsilon, a+\varepsilon) \subseteq I$. Now w is the restriction of an element $p \in$ Isom($\mathbb{R}$), and for all $x \in (a-\varepsilon, a) \subseteq I$, x and $px = w(x)$ are in the same P-orbit, hence $p \in P$. Thus, for x, $y \in (a-\varepsilon, a+\varepsilon)$,

$$\begin{aligned} x, y \in \text{ the same } \mathring{S}-\text{orbit} &\iff w(x), w(y) \in \text{ the same } \mathring{S}-\text{orbit} \\ &\iff w(x), w(y) \in \text{ the same } P-\text{orbit} \\ &\iff p(x), p(y) \in \text{ the same } P-\text{orbit} \\ &\iff x, y \in \text{ the same } P-\text{orbit}. \end{aligned}$$

We have shown $(a-\varepsilon, a+\varepsilon)$ is good for P, hence so is $I \cup (a-\varepsilon, a+\varepsilon)$, which strictly contains I, which is impossible by definition of I. This establishes the claim, and shows that any interval in the same component of U as J is good for P.

Now let J be good for P and Let J' be any non-degenerate subinterval of U. Choose an interior point x of J'. Since $\mathring{S}$-orbits of U are dense, there is an

$\mathring{S}$-word w such that $w(x) \in J'$. There is a non-degenerate subinterval I of J such that w is defined on I and sends I $w(I)$ to a non-degenerate interval I' contained in J'. Also, w is the restriction of an element s in Isom($\mathbb{R}$). Now if y, $z \in I'$, by an argument similar to that above, y, z are in the same $\mathring{S}$-orbit if and only if they are in the same sPs^{-1}-orbit. Thus I' is good for sPs^{-1}, and by the above, so is J'.

We have shown that the subgroup P in the definition of homogeneous is uniquely determined by J, and is independent of J, up to conjugation in Isom($\mathbb{R}$). If P is a group of translations, then it is normal in Isom($\mathbb{R}$) (see equation (ii) in the proof of Lemma 1.2.1), so is independent of J. We now take care of the proof of the main theorem in the case where there is a homogeneous component.

Theorem 2.5. *Suppose S is pure, minimal and homogeneous. Then $G(S)$ is free abelian.*

Proof. The proof begins with some observations. If D is the domain of S, $\mathring{D}$ is the homogeneous component. Let $[a, a+c]$, $[b, b+c]$ be intervals in $\mathring{D}$, where $c > 0$, and suppose $w(a) = b$, where w is an $\mathring{S}$-word with $sgn(w) = 1$. Let s be the extension of w to an isometry of $\mathbb{R}$ (i.e. translation by $b - a$). According to the previous discussion, $[a, a+c]$ is good for some subgroup P of Isom($\mathbb{R}$), and $[b, b+c]$ is good for sPs^{-1}. Also, w is defined on some interval $[a, a+\varepsilon)$, where $0 < \varepsilon \leq c$. If $a + t \in [a, a+c]$, then since $\mathring{S}$-orbits are dense, there is some $p \in P$ such that $p(a+t) \in [a, a+\varepsilon)$, and $p(a+t)$ is in the same $\mathring{S}$-orbit as $a+t$, as is $w(p(a+t)) = sp(a+t) = sps^{-1}(s(a+t)) = sps^{-1}(b+t)$. But this is in the same $\mathring{S}$-orbit as $b+t$, so $a+t$, $b+t$ are in the same $\mathring{S}$-orbit.

Hence, there is a path γ_t joining $a+t$ to $b+t$ lying in a leaf of $\mathcal{F}$, the foliation of $\Sigma(S)$. Let λ_t be the closed path which is the product of the paths γ_0, l, $\overline{\gamma}_t$, l', where l is a linear parametrisation of the segment $[b, b+t]$ in D (i.e. $l(s) = (1-s)b + s(b+t)$) and l' is a linear parametrisation of $[a+t, a]$. The element $[\lambda_t]$ of $G(S)$ it represents is independent of the choice of γ_0 and γ_t, by definition of $G(S)$. Further, it is locally constant. For we can find an $\mathring{S}$-word w such that $w(a+t) = b+t$, and w is defined in a neighbourhood of $a+t$. The path γ_u, for all u in this neighbourhood, can be chosen using w; if $w = \mathring{\varphi}_{i_p}^{e_p} \ldots \mathring{\varphi}_{i_1}^{e_1}$, and C_j is the domain of $\mathring{\varphi}_{i_j}^{e_j}$, take $\eta_{1,u}$ to be the linear parametrisation of $[a+u, \mathring{\varphi}_{i_1}^{e_1}(a+u)]$ in the image of $C_1 \times [0,1]$ in $\Sigma(S)$, $\eta_{2,u}$ to be the linear parametrisation of $[\mathring{\varphi}_{i_1}^{e_1}(a+u), \mathring{\varphi}_{i_2}^{e_2}\mathring{\varphi}_{i_1}^{e_1}(a+u)]$ in (the image of) $C_2 \times [0,1]$, etc, and let $\gamma_u = \eta_{1,u} \ldots \eta_{p,u}$. Let α_0 be a linear parametrisation of the segment $[a+u, a+t]$, and let α_i be a linear parametrisation of the segment in D from the endpoint of $\eta_{1,u}$ to the endpoint of $\eta_{1,t}$. Arguing as in Lemma 1.1, $\eta_{1,u}$ is homotopic to $\alpha_{i-1}\eta_{1,t}\overline{\alpha}_i$ for $1 \leq i \leq p$, hence γ_u is homotopic to $\alpha_0\gamma_t\overline{\alpha}_p$ (homotopy relative to endpoints), and it follows easily that $[\lambda_u] = [\lambda_t]$. Thus the map $[0, c] \to G(S)$, $t \mapsto [\lambda_t]$, viewing $G(S)$ as a discrete space, is continuous, so constant, hence $[\lambda_t] = [\lambda_0] = 1$, for all $t \in [0, c]$.

Similarly, if $[a, a+c]$ and $[b, b-c]$ are intervals in $\mathring{D}$ and there is a word w sending a to b with $sgn(w) = -1$, we can find γ_t joining $a+t$ to $b-t$ for $0 \le t \le c$, and analogously define λ_t, such that $[\lambda_t] = 1$ for all $t \in [0, c]$.

Call a path γ in $\Sigma(S)$ *admissible* if it can be written as $\gamma = \beta_1\alpha_1 \ldots \alpha_{n-1}\beta_n$ where each β_i is a path in D and each α_i is a path in a leaf whose image is the image in $\Sigma(S)$ of a segment $\{x\} \times [0,1]$ of $A_j \times [0,1]$ for some i, so α_i joins x and $\varphi_j(x)$. Note that we can replace β_i by β_i' for any path β_i' in D with the same endpoints, without changing the homotopy class of γ relative to its endpoints (the components of D are convex, so β_i and β_i' are homotopic by a linear homotopy). If β_i is closed, we can therefore replace it by a trivial path, which can be omitted in the expression for γ. The proof itself now falls into two cases.

(1) Assume that the group P in the definition of homogeneous consists only of translations, so may be viewed as a subgroup of $\mathbb{R}$. We first claim that we can change the embedding of D in $\mathbb{R}$ so that $sgn(\varphi) = +1$ for all generators φ. To prove this, we begin with a remark.

Remark. Let $w = \varphi_1 \ldots \varphi_n$ be an S-word (where φ_i is a generator or the inverse of a generator), such that $\mathrm{Im}(\varphi_i)$ and $\mathrm{Dom}(\varphi_{i+1})$ are in the same component of D for $1 \le i < n$. Then there are intervals $[a, a+c] \subseteq \mathrm{Dom}(\varphi_1)$ and $[b, b+sgn(w)c] \subseteq \mathrm{Im}(\varphi_n)$, where $c > 0$, such that, for $0 \le t \le c$, $a+t$ and $b + sgn(w)t$ are in the same $\mathring{S}$-orbit.

This is proved by induction on n. When $n = 1$ it follows easily since the system is non-degenerate. Assume it is true for n, and let w, a, b and c be as in the statement of the remark. Suppose φ_{n+1} is a generator or the inverse of a generator such that $\mathrm{Im}(\varphi_n)$ and $\mathrm{Dom}(\varphi_{n+1})$ are in the same component of D. Since P is dense, there exists $p \in P$ and c' such that $0 < c' \le c$ and $p([b, b+sgn(w)c']) \subseteq \mathrm{Dom}(\mathring{\varphi}_{n+1})$. Since the system is pure and minimal, it follows that $[b, b+sgn(w)c']$ is good for P, so $b+sgn(w)t$ and $p(b+sgn(w)t) = p(b) + sgn(w)t$ are in the same $\mathring{S}$-orbit for $0 \le t \le c'$, hence so are $b+sgn(w)t$ and $\mathring{\varphi}_{n+1}(p(b) + sgn(w)t) = \varphi_{n+1}(p(b) + sgn(w)t)$. That is, $b + sgn(w)t$ and $\varphi_{n+1}(p(b)) + sgn(\varphi_{n+1})sgn(w)t$ are in the same $\mathring{S}$-orbit. Put $b' = \varphi_{n+1}(p(b))$, so $a+t$, $b' + sgn(w\varphi_{n+1})t$ are in the same $\mathring{S}$-orbit for $0 \le t \le c'$, completing the induction.

To establish our claim, form a graph Γ with vertices the components of D, and an edge for each generator φ, from the component of D containing $\mathrm{Dom}(\varphi)$ to the component containing $\mathrm{Im}(\varphi)$, with label φ, and as usual adjoin an opposite edge for each such edge, with label φ^{-1}. To this pair of opposite edges we associate a sign, namely $sgn(\varphi) = sgn(\varphi^{-1})$. Since the system is connected, so is Γ, by Remark 2 after Lemma 1.1, and we choose a maximal tree T in Γ. Fix a vertex v. Each vertex at distance n from v, in the path metric on T, is adjacent in T to a unique vertex at distance $n-1$ from v, for $n > 0$, by uniqueness of reduced paths. By induction on n, reflecting components of D around their midpoints as necessary, we can assume that all edges of T have sign $+1$. We shall show that all edges of Γ now have sign $+1$, proving the claim.

For otherwise, there is an edge e with sign -1; let p be a path in T from v to $o(e)$ and q a path in T from $t(e)$ to v. Then peq is a closed non-trivial path from v to v. If $\varphi_1, \dots \varphi_n$ are the labels on the successive edges of this path, put $w = \varphi_1 \dots \varphi_n$. Then w is a word satisfying the hypotheses of the remark, with $sgn(w) = -1$ and such that $\mathrm{Dom}(\varphi_1)$ and $\mathrm{Im}(\varphi_n)$ are in the same component of D (namely v). We shall show this is impossible.

We can choose a, b and c as in the remark, such that $[a, a+c]$ and $[b, b+sgn(w)c] = [b, b-c]$ are contained in the interior of the component of D to which they both belong, which we denote by J. Since the system is pure and minimal, J is good for P by the discussion preceding the theorem. Thus $a+t$ and $b-t$ are in the same P-orbit, for $0 \leq t \leq c$. But then P contains the translation $x \mapsto x+b-a-2t$, for all $t \in [0, c]$, contradicting that P is countable.

After this preamble, we assume $sgn(\varphi) = 1$ for all generators φ. We shall define a homomorphism from $G(S)$ to $\mathbb{R}$. First, if $\gamma = \beta_1\alpha_1 \dots \alpha_{n-1}\beta_n$ is an admissible path as above, define $\rho(\gamma)$ to be $\sum_i(b_i - a_i)$, where β_i starts at a_i and ends at b_i (it is easy to see this depends only on γ). Let Γ be a finite graph and $r : \Sigma(S) \to \mathrm{real}(\Gamma)$ a retraction as in Lemma 1.1. Then $r \circ \gamma$ is admissible, of the form $\beta'_1\alpha'_1 \dots \alpha'_{n-1}\beta'_n$, and $\rho(\gamma) = \rho(r \circ \gamma)$. Choose a basepoint v which is a vertex of $\mathrm{real}(\Gamma)$, and suppose γ is a closed path from v to v. Then the paths α'_i, β'_i all join vertices of $\mathrm{real}(\Gamma)$, so using the Simplicial Approximation Theorem as in the remarks before Lemma 1.1 (see the proof of Theorem 3.3.9 in [99]), can be replaced by homotopic paths which are products of paths which lie in $\mathrm{real}(e)$ for some edge e of Γ, and join the endpoints of e. Therefore γ is homotopic rel v to a path γ' which is also a product of such paths, and therefore admissible, and $\rho(\gamma') = \rho(r \circ \gamma) = \rho(\gamma)$. Note that any element of $\pi_1(\Sigma(S), v)$ is represented by such an admissible path, by the remarks preceding Lemma 1.1. Now if η is another closed, admissible path from v to v homotopic rel v to γ, then γ' and η' are homotopic rel v, so one can be obtained from the other by a sequence of moves (1), (2) and (3) as described before Lemma 1.1. These moves do not change ρ, hence $\rho(\gamma) = \rho(\eta)$. We can therefore define $\bar{\rho} : \pi_1(\Sigma(S), v) \to \mathbb{R}$ by $\bar{\rho}(g) = \rho(\gamma)$, where γ is any admissible path in the homotopy class g. Clearly this is a group homomorphism (products of admissible paths are admissible). Also, $\bar{\rho}(N(S)) = 0$. For $N(S)$ is generated as a normal subgroup by closed paths arising from paths in $\mathrm{real}(\Gamma(x))$, where $x \in D$, which (again using the Simplicial Approximation Theorem) are homotopic to admissible paths of the form $\alpha_1 \dots \alpha_n$, where the α_i are paths all in the same leaf, and $\rho(\alpha_1 \dots \alpha_n) = 0$.

Thus we have a homomorphism $\tau : G(S) \to \mathbb{R}$ induced by $\bar{\rho}$.

We claim that τ maps $G(S)$ to P. To see this, take any admissible path $\gamma = \beta_1\alpha_1 \dots \alpha_{n-1}\beta_n$, where β_i runs from a_i to b_i. Since $\mathring{S}$-orbits are dense and there are only finitely many singular $\mathring{S}$-orbits, we can choose the graph Γ to consist of segments in D and segments in regular leaves. Then replacing ρ by $r \circ \rho$, we can assume, without changing $\rho(\gamma)$, that all α_i are paths in regular leaves. We shall show that $\rho(\gamma) \equiv b_n - c \bmod P$, for some c in the same component of $\mathring{D}$ as b_n and in the same $\mathring{S}$-orbit as a_1.

The proof is by induction on n. When $n = 1$ we may take $c = a_1$, so assume $n > 1$ and $\rho(\beta_1\alpha_1 \dots \alpha_{n-2}\beta_{n-1}) = b_{n-1} - c'$, where c' is in the same component of $\mathring{D}$ as b_{n-1} and the same orbit as a_1. Now b_{n-1}, a_n are in the same $\mathring{S}$-orbit, being on the same regular leaf, so $b_{n-1} = \mathring{\varphi}(a_n)$ for some word $\mathring{\varphi}$, which is defined on some non-degenerate interval $I \subseteq [a_{n-1}, b_{n-1}]$. Since $\mathring{S}$-orbits are dense, there is a point $d \in I$ in the same orbit as a_1 and c', so d and c' are in the same P-orbit. Put $c = \mathring{\varphi}(d)$. Then

$$\begin{aligned}
\rho(\gamma) &\equiv (b_{n-1} - c') + (b_n - a_n) \bmod P \\
&\equiv (b_{n-1} - d) + (b_n - a_n) \bmod P \\
&= (a_n - c) + (b_n - a_n) \quad (\text{since } sgn(\mathring{\varphi}) = 1) \\
&= b_n - c
\end{aligned}$$

and c has the required properties. Hence, if γ is a closed path, $\rho(\gamma) \equiv b_n - c = a_1 - c \equiv 0 \bmod P$, since a_1, c are in the same component of $\mathring{D}$ and the same $\mathring{S}$-orbit.

Thus τ is a homomorphism from $G(S)$ to P, and we shall show it is a monomorphism. For let $\gamma = \beta_1\alpha_1 \dots \alpha_{n-1}\beta_n$ be a closed admissible path with $\rho(\gamma) = 0$. If γ is not contained in a leaf, then two consecutive terms in the sum defining $\rho(\gamma)$ have opposite signs, say $b_{i-1} - a_{i-1}$ and $b_i - a_i$. Suppose $|b_{i-1} - a_{i-1}| \leq |b_i - a_i|$. By the observations at the start of the proof, $b_{i-1} + a_i - b_i$ is in $[a_{i-1}, b_{i-1}]$, and is in the same $\mathring{S}$-orbit as b_i. Again using the observations at the start of the proof, we can replace β_{i-1} by a linear path from a_{i-1} to $b_{i-1} + a_i - b_i$ in D, α_{i-1} by a path α in a leaf from $b_{i-1} + a_i - b_i$ to b_i, and omit β_i, to obtain a path γ' representing the same element of $G(S)$ as γ. The path $\alpha\alpha_i$ lies in a leaf and joins two points of D, so is homotopic relative to its endpoints to an admissible path in this leaf, so we can modify γ' to make it admissible. If $|b_{i-1} - a_{i-1}| \geq |b_i - a_i|$, we can make similar modifications to γ which involve omitting β_{i-1}. Thus we can decrease the number of segments β_i in γ without changing the element of $G(S)$ it represents. By repeating this we replace γ with a closed path lying in a leaf, which represents 1 in $G(S)$, hence so does γ. Since P is a finitely generated subgroup of $\mathbb{R}$, it is free abelian, and therefore so is $G(S)$. It is left as an exercise to show τ is onto, so $G(S)$ is isomorphic to P.

(2) Assume P contains a reflection. An $\mathring{S}$-orbit O is called *one-sided* if for some $x \in O$, x is the fixed point of an $\mathring{S}$-word $\mathring{w}$ with $sgn(\mathring{w}) = -1$. If this is true for one $x \in O$, it is true for all $x \in O$. (If $y = \mathring{v}(x)$, for some $\mathring{S}$-word $\mathring{v}$, then y is the fixed point of $\mathring{v}\mathring{w}\mathring{w}^{-1}$.)

Since P contains a reflection, and is dense, infinitely many points $x \in \mathring{D}$ are centres of reflections in P. For such x, $x - t$ and $x + t$ are in the same $\mathring{S}$-orbit for all sufficiently small t. Since there are countably many $\mathring{S}$-words, there are countably many which are restrictions of translations, hence x is the fixed point of some $\mathring{S}$-word $\mathring{w}$. Since there are infinitely many choices for x, we can assume x is in a regular S-orbit, which then coincides with the $\mathring{S}$-orbit of x. Let L be the leaf of $\mathcal{F}$ through x.

Suppose $\gamma = \beta_1\alpha_1 \ldots \alpha_{n-1}\beta_n$ is an admissible, closed path. Since L is dense, we may take the graph Γ as in the previous case to consist of segments in D and segments in L. Replacing γ by $r \circ \gamma$, we can assume, without changing the homotopy class of γ, that all α_i are in L. Then the endpoints a_i, b_i of β_i and x are all in the same $\mathring{S}$-orbit, so there are $\mathring{S}$-words w sending a_i to b_i and v sending x to a_i, as well as an $\mathring{S}$-word u fixing x. Replacing w by $w(vuv^{-1})$ if necessary, we can assume $sgn(w) = -1$. By the remarks at the beginning of the proof, $a_i + t$ and $b_i - t$ are in the same $\mathring{S}$-orbit for $0 \leq t \leq b_i - a_i$, or $b_i - a_i \leq t \leq 0$, as appropriate. That is, if $c_i = (a_i + b_i)/2$, $c_i - t$ and $c_i + t$ are in the same $\mathring{S}$-orbit, for $0 \leq t \leq |b_i - a_i|/2$. Again by the remarks at the beginning, if μ is a path in L from a_i to b_i, λ is the trivial path from c_i to c_i, l is a linear parametrisation of $[c_i, a_i]$ and l' a linear parametrisation of $[b_i, c_i]$, then the path $l\mu l'\lambda$ represents 1 in $G(S)$. Hence $\mu l''$, where l'' is a linear parametrisation of $[b_i, a_i]$, and so $\mu\beta_i$, represents 1 in $G(S)$. We can then replace β_i by $\bar{\mu}$ in γ, without changing the element of $G(S)$ it represents. Doing this for each i results in a closed path in L, which represents 1 in $G(S)$. Hence $G(S)$ is trivial. □

Underlying the proof of Case (1) in Theorem 2.5 is the idea that Lebesgue measure on $\mathbb{R}$ gives a "transverse measure" to the foliation $\mathcal{F}$. This lifts to a measured foliation on the covering of $\Sigma(S)$ corresponding to $N(S)$, and so to an $\mathbb{R}$-tree on which $G(S) = \pi_1(S)/N(S)$ acts, by a construction analogous to that for measured laminations (Chapter 4, §4). For details, we refer to [90] and [54].

We finish this section with a result on homogeneous components (or rather the lack of them) for later use.

Theorem 2.6. *Assume no minimal component of the non-degenerate system of isometries S is homogeneous. Then for $t > 0$, all orbits of S_{-t} are finite.*

Proof. First note that, if an $\mathring{S}_{-t}$-word is defined at a point x, then the corresponding $\mathring{S}$-word is defined on $[x-k, x+k]$, for any $k < t$, by an easy induction on the length of the word.

We assume S_{-t} has a minimal component, say V, and we prove that S has a homogeneous component. By an easy adaptation of the proof of the remark preceding Theorem 2.1, we can find an interval $J = [a, a+l]$ contained in V, where $0 < l < t$ and $a + l \in \mathring{S}^{+}_{-t}(a)$, and $\mathring{S}^{+}_{-t}(a) \cap J$ is dense in J. Thus there is an isometry τ of $\mathring{S}$ which is restriction of translation by l to its domain (by Lemma 1.2.1), and τ is defined on $[a-l, a+l]$, while τ^{-1} is defined on $[a, a+2l]$, since $l < t$.

If x_0, $y_0 \in J$ and $y_0 \in \mathring{S}_{-t}(x_0)$, then $y_0 = w(x_0)$ for some $\mathring{S}_{-t}$-word w, which is the restriction of a unique isometry p of $\mathbb{R}$ by Lemma 1.2.1 (its domain is an open interval, and non-empty, so is not a single point). Let X be the set of all isometries p obtained in this way by varying x_0, y_0 and w, and let P_0 be the subgroup of Isom($\mathbb{R}$) generated by X. Note that $X^{-1} = X$. Further, P_0 contains a set of translations taking a to a dense subset of J, so the subgroup of translations of P_0 is not cyclic, hence by Lemma 1.1.3, P_0 is dense. We claim that, if $y - px \in l\mathbb{Z}$, where $p \in P_0$, $x \in J$, and n is an integer with $y + nl \in J$, then $y + nl \in \mathring{S}(x)$. The proof is by induction on k, where $p = p_1 \dots p_k$ and each $p_i \in X$.

Assume for the moment this is true when $k = 1$ and consider the general case. Put $x' = p_2 \dots p_k x$, and choose an integer m such that $x' + ml \in J$. Then $x' + ml \in \mathring{S}(x)$ by induction. Further, $y + rl = p_1 x' = p_1(x' + ml) \pm ml$, for some $r \in \mathbb{Z}$, so $y + nl = p_1(x' + ml) + (nl \pm ml - rl) \in J$. By the case $k = 1$, $y + nl \in \mathring{S}(x' + ml)$, hence $y + nl \in \mathring{S}(x)$.

We now have to consider the situation $y + rl = px$, $x \in J$, $y + nl \in J$, $p \in X$, $r \in \mathbb{Z}$. Thus there are $x_0, y_0 \in J$ with $y_0 = w(x_0)$ for some $\mathring{S}_{-t}$-word w, and w is the restriction of p. Further, the corresponding $\mathring{S}$-word v is defined on J and is obtained by restriction from p. If $p = t_{y_0 - x_0}$ (translation by $y_0 - x_0$), then $y + rl = v(x) = x + (y_0 - x_0)$, so $y + rl \in \mathring{S}(x)$ and $y + rl \in [a-l, a+2l]$. It follows that $n = r$, $r+1$ or $r-1$, and if $n = r$ we are finished. If $n = r+1$, then $y + rl \in [a-l, a]$ and $\tau(y + rl) = y + nl$, so $y + nl \in \mathring{S}(y)$, hence $y + nl \in \mathring{S}(x)$. If $n = r - 1$, similarly $y + nl = \tau^{-1}y$ and $y + nl \in \mathring{S}(x)$.

Otherwise, p is the reflection $r_{x_0+y_0}$, whose fixed point $b = (x_0 + y_0)/2$ is in J. Then $|(y+rl) - b| = |x - b| \leq l$, since $x \in J$, so again $y + rl \in [a-l, a+2l]$, and as in the previous case $y + nl \in \mathring{S}(x)$, completing the inductive proof.

Careful inspection of the inductive proof shows that not only is $y + nl \in \mathring{S}(x)$, but $y + nl = w(x)$ for some $\mathring{S}$-word w with $sgn(w) = sgn(p)$ ($sgn(p) = 1$ if p is a translation, $sgn(p) = -1$ otherwise). If x, $y \in J$ are in the same P_0-orbit, then taking $n = 0$ in the claim, $y = w(x)$ for some $\mathring{S}$-word w with $sgn(w) = sgn(p)$, in particular $y \in \mathring{S}(x)$.

Now suppose x_0, $y_0 \in J$ and $y_0 = w(x_0)$ for some $\mathring{S}$-word w. Then w is the restriction of a unique element $q \in \mathrm{Isom}(\mathbb{R})$. Suppose $y = qx$, where x, $y \in J$. We claim that $y = u(x)$ for some $\mathring{S}$-word u with $sgn(u) = sgn(q)$, so u is obtained from q by restriction.

If x_0, y_0 are the endpoints of J and q is translation by $\pm l$, then x, y are also the endpoints of J and we may take $u = w$. Otherwise, since P_0 is dense, we can find a translation $p \in P_0$ such that px is in the domain of w, $px \in J$ and $w(px) \in J$. From what we have proved, $px = v(x)$ for some $\mathring{S}$-word v with $sgn(v) = sgn(p) = 1$. Then

$$wv(x) = w(px) = qpx = (qpq^{-1})y = p^{\pm 1}y$$

(see equation (ii) in the proof of Lemma 1.2.1), so $y = p^{\mp 1}(wv(x))$, and again by what we have proved, $y = v'wv(x)$ for some $\mathring{S}$-word v' with $sgn(v') = sgn(p^{\mp 1}) = 1$. Thus $sgn(v'wv) == sgn(w) = sgn(q)$, and we can take u to be $v'wv$.

Now suppose x_0, $y_0 \in J$ and $y_0 = w(x_0)$ for some $\mathring{S}$-word w, so w is the restriction of a unique isometry p of $\mathbb{R}$. Let P be the subgroup of $\mathrm{Isom}(\mathbb{R})$ generated by all isometries p obtained in this way by varying x_0, y_0 and w. Clearly $P_0 \leq P$, so P is dense. By minor modifications of the inductive proof above, we can show that if x, $y \in J$ are in the same P-orbit, then $y \in \mathring{S}(x)$. It is clear from the definition of P that if x, $y \in J$ and $y \in \mathring{S}(x)$, then x, y are in the same P-orbit. It follows from Theorem 2.1 that J is contained in a minimal component of $\mathring{S}$, which is then a homogeneous component of $\mathring{S}$, as required, since J is good for P. □

3. Independent Generators

We now have to consider systems of isometries with a non-homogeneous minimal component. One key idea in dealing with this is that of an independent set of generators. Let $S = (D, \{\varphi_i\}_{i=1,\dots,k})$ be a system of isometries (with D a multi-interval). An S-word $w = \varphi_{i_1}^{e_1} \dots \varphi_{i_l}^{e_l}$ is called *reduced* if there is no value of j $(1 \leq j \leq l-1)$ such that $i_j = i_{j+1}$ and $e_j = -e_{j+1}$. It is called *cyclically reduced* if in addition, it is not the case that $i_p = i_1$ and $e_p = -e_1$. If w is cyclically reduced, its *cyclic conjugates* are the reduced words $\varphi_{i_j}^{e_j} \dots \varphi_{i_l}^{e_l}\varphi_{i_1}^{e_1} \dots \varphi_{i_{j-1}}^{e_{j-1}}$ for $1 \leq j \leq l$. Obviously, there are similar definitions for $\mathring{S}$-words. By the trivial word we mean the empty string having no letters.

Definition. The system S has independent generators if, for every non-trivial reduced $\mathring{S}$-word, the corresponding partial isometry has at most one fixed point.

Equivalently, no non-trivial reduced word in the $\mathring{\varphi}_i^{\pm 1}$ with non-empty domain is a restriction of the identity map on $\mathbb{R}$. This is also equivalent to: no non-trivial reduced word in the $\varphi_i^{\pm 1}$ with non-degenerate domain is a restriction of the identity map on $\mathbb{R}$. Note that adding singletons to S, or removing them, does not affect whether or not S has independent generators. We shall prove the following result, due to Gaboriau [50].

Theorem 3.1. *Let $S = (D, \{\varphi_i\}_{i=1,\dots,k})$ be a system of isometries without a minimal homogeneous component. Then there is a system of isometries $S' = (D, \{\varphi'_i\}_{i=1,\dots,k})$ with independent generators, having the same orbits as S, where each φ'_i is the restriction of φ_i to a closed subinterval of its domain. If S is non-degenerate, we can remove all singletons from S' without changing the orbits.*

An immediate corollary of this and Corollary 1.8 is the following.

Corollary 3.2. *Let $S = (D, \{\varphi_i\}_{i=1,\dots,k})$ be a connected system of isometries without a minimal homogeneous component. Then there is a connected system of isometries $S' = (D, \{\varphi'_i\}_{i=1,\dots,k})$ with independent generators such that $G(S)$ and $G(S')$ are isomorphic. If S is non-degenerate, there is such a system S' which is non-degenerate.*

□

To begin, we have a simple criterion for independent generators. Given a system of isometries S, and x in the domain of S, let $\mathring{\Gamma}(x)$ denote the component of x in the Cayley graph of $\mathring{S}$.

If $w = \psi_1 \dots \psi_p$ is an S-word (each ψ_i being a generator or its inverse) and $u \in D$ (the domain of S), there is a path p_w in the Cayley graph from u to $w(u)$, such that writing down the labels on the edges of p_w in succession gives the word $\psi_p \dots \psi_1$.

Lemma 3.3. *The system S has independent generators if and only if, for all $x \in D$, the graph $\mathring{\Gamma}(x)$ is a tree.*

Proof. If $e_1, \dots, e_r$ is a circuit in $\mathring{\Gamma}(x)$, with label $\mathring{\varphi}_{i_j}^{\varepsilon_j}$ on e_j, the partial isometry $(\mathring{\varphi}_{i_r}^{\varepsilon_r} \circ \dots \circ \mathring{\varphi}_{i_1}^{\varepsilon_1})^2$ fixes x and has sign $+1$, so is the restriction of the identity map on $\mathbb{R}$, and has open, non-empty domain, hence fixes more than one point.

Conversely, if there is a reduced S-word w such that the corresponding isometry fixes two points, it fixes their midpoint, say z, and is defined on an open interval containing z, hence defines a closed, non-trivial reduced path p_w in $\mathring{\Gamma}(z)$. Arguing as in the proof of 1.3.1, it follows that $\mathring{\Gamma}(z)$ contains a circuit.

□

Our first goal is to develop another criterion for an independent set of generators, using a different graph, which is a variant of the graph Γ_S used in §1. Let $S = (D, \{\varphi_i\}_{i=1,\dots,k})$ be a system of isometries. Let φ_i have domain $[a_i, a'_i]$ and codomain $[b_i, b'_i]$, where $a_i \leq a'_i$, $b_i = \varphi_i(a_i)$ and $b'_i = \varphi_i(a'_i)$.

Choose a set V with $4k$ elements denoted by c_i, c_i', d_i, d_i' for $1 \leq i \leq k$. The mapping $c_i \mapsto a_i$, $c_i' \mapsto a_i'$, $d_i \mapsto b_i$, $d_i' \mapsto b_i'$ will be denoted by $v \mapsto v(S)$. Note that there is a base associated to each $v \in V$ ($[a_i, a_i']$ to c_i and c_i', and $[b_i, b_i']$ to d_i and d_i').

Form a directed graph Δ with vertex set V, and for each i, two edges with label φ_i, one from c_i to d_i, the other from c_i' to d_i'. Now define an equivalence relation $\mathcal{R}(S)$ on V as follows. For v_1, $v_2 \in V$, $v_1\mathcal{R}(S)v_2$ if and only if there is an S-word w, with corresponding $\mathring{S}$-word $\mathring{w}$, such that

(1) $\mathring{w}v_1(S) = v_2(S)$;

(2) $sgn(w) = \varepsilon(v_1)\varepsilon(v_2)$;

where $\varepsilon : V \to \{\pm 1\}$ is defined by $\varepsilon(c_i) = -1$, $\varepsilon(c_i') = 1$, $\varepsilon(d_i) = -sgn(\varphi_i)$ and $\varepsilon(d_i') = sgn(\varphi_i)$. Thus $\varepsilon(v)$ represents the position of the endpoint $v(S)$ of the base associated to v. It is the left-hand endpoint if $\varepsilon(v) = -1$ and the right-hand endpoint if $\varepsilon(v) = 1$. (Note that we cannot give a sign ± 1 to the endpoints themselves; the left-hand endpoint of one base might coincide with the right-hand endpoint of another base, and S might contain singletons, when the left and right-hand endpoints of the corresponding bases coincide.) Thus if S is non-degenerate, then for sufficiently small $t > 0$, the interval $[v(S), v(S) - \varepsilon(v)t]$ is contained in the base associated with v.

We make Δ into a graph (in the sense of §1.3) by adding an opposite edge for each original edge e, and if e has label φ, then the opposite edge is given label φ^{-1}. The graph obtained from Δ by identifying vertices equivalent under $\mathcal{R}(S)$, with the same edges (and labels) as Δ will be denoted by $\Delta(S)$. There is a projection map $\pi : \Delta \to \Delta(S)$ which is a graph mapping and a bijection on edges.

If v_1, $v_2 \in V$ are $\mathcal{R}(S)$-equivalent, we fix an S-word of minimal length satisfying (1) and (2) above, and denote it by $w_{[v_2,v_1]}$. We do this in such a way that $w_{[v_1,v_2]} = w_{[v_2,v_1]}^{-1}$ and $w_{[v_1,v_1]} = 1$.

Remark. For non-degenerate S, Condition (1) in the definition above implies that $w_{[v_2,v_1]}$ is defined on $[v_1(S), v_1(S) - \varepsilon(v_1)t]$ for sufficiently small t (the domain of $\mathring{w}$ is an open interval). Condition (2) implies that $w_{[v_2,v_1]}$ maps $[v_1(S), v_1(S) - \varepsilon(v_1)t]$ onto $[v_2(S), v_2(S) - \varepsilon(v_2)t]$. Further, $v_1(S) - \varepsilon(v_1)t = v_1(S_{-t})$ and $v_2(S) - \varepsilon(v_2)t = v_2(S_{-t})$, so $w_{[v_2,v_1]}$ sends $v_1(S_{-t})$ to $v_2(S_{-t})$.

Lemma 3.4. *If S is a non-degenerate system of isometries, there exists a real number $t_0 > 0$ such that for all $t \in [0, t_0]$, S_{-t} is non-degenerate and the equivalence relations $\mathcal{R}(S)$ and $\mathcal{R}(S_{-t})$ are the same, so $\Delta(S) = \Delta(S_{-t})$.*

Proof. Choose $t_0 > 0$ such that, for every pair (v_1, v_2) of elements of V which are $\mathcal{R}$-equivalent, the domain of $w_{[v_2,v_1]}$ contains the closed interval centred at $v_1(S)$ of length $6t_0$, and such that $2t_0 < \min\{a_i' - a_i \mid 1 \leq i \leq k\}$ (to ensure S_{-t} is non-degenerate). Then for $t \leq t_0$, and such a pair (v_1, v_2), let

w_t denote $(w_{[v_2,v_1]})_{-t}$. By an easy induction on the length of w, the domain of w_t contains the closed interval centred at $v_1(S)$ of length $4t_0$. The point $v_1(S_{-t})$ is at distance $t \leq t_0$ from $v_1(S)$, so is in the interior of the domain of w_t. From the remark preceding the lemma, w_t sends $v_1(S_{-t})$ to $v_2(S_{-t})$. Thus Condition (1) in the definition of $\mathcal{R}(S_{-t})$ is satisfied, using this word w_t, and Condition (2) is clearly satisfied. Hence $v_1\mathcal{R}(S_{-t})v_2$. If z is an S-word and z_{-t} is defined in a neighbourhood of $v_1(S_{-t})$, then z is defined in a neighbourhood of $v_1(S)$ (induction on the length of z). Hence if $v_1\mathcal{R}(S_{-t})v_2$ then $v_1\mathcal{R}(S)v_2$. □

Before stating the next result, some terminology is needed. Let $\tilde{S}$ be a restriction of S which is a system of isometries. Thus $\tilde{S}$ is obtained by replacing the base $[a_i, a'_i]$ by $[a_i + t_i, a'_i - t'_i]$ for some t_i, $t'_i \geq 0$. Let e_i be the unoriented edge of Δ with endpoints c_i and d_i and let e'_i be the unoriented edge with endpoints c'_i and d'_i. We say that $\tilde{S}$ is obtained from S by shortening S by t_i on the side of e_i, for $i \in I$ and by t'_i on the side of e'_i for $i \in J$, where $I = \{i \mid 1 \leq i \leq k \text{ and } t_i > 0\}$ and $J = \{i \mid 1 \leq i \leq k \text{ and } t'_i > 0\}$. We also say that S is obtained from $\tilde{S}$ by lengthening $\tilde{S}$ by t_i on the side of e_i, for $i \in I$ and by t'_i on the side of e'_i for $i \in J$.

Proposition 3.5. *Let S be a non-degenerate system of isometries. Choose a maximal forest F in $\Delta(S)$ (see the second exercise after Lemma 1.3.2) and let F' be its preimage under π in Δ. Then there exists $t_0 > 0$ such that, for all $t \in [0, t_0]$, the system $\tilde{S}$ obtained by shortening S by t on the side of all edges in $\Delta \setminus F'$ is non-degenerate and has the same orbits as S.*

Proof. Take t_0 to satisfy the conditions in the proof of Lemma 3.4. Before giving the proof, a remark is needed.

Remark. Let e be an edge of F with label ψ and put $v_0 = o(e)$, $v_1 = t(e)$. Let $v \in V$ be $\mathcal{R}(S)$-equivalent to v_1, and let $w = w_{[v,v_1]}\psi$. Then the isometry $\tilde{w}$ of $\tilde{S}$ corresponding to w (for $0 \leq t \leq t_0$) is defined on $I = [v_0(S), v_0(S) - 2\varepsilon(v_0)t_0]$, and maps it to $[v(S), v(S) - 2\varepsilon(v)t_0]$.

For $\tilde{\psi} = \psi$, since $e \in F$, so $\tilde{\psi}$ is defined on I, and maps it to $K = [v_1(S), v_1(S) - 2\varepsilon(v_1)t_0]$. Now (see the proof of Lemma 3.4) $w_{[v,v_1]}(S)$ is defined on the interval $[v_1 - 3t_0, v_1 + 3t_0]$, so $\tilde{w}_{[v,v_1]}$ is defined on $[v_1 - 2t_0, v_1 + 2t_0]$. It therefore sends K to $[v(S), v(S) - 2\varepsilon(v_1)sgn(w_{[v,v_1]})t_0]$. But $\varepsilon(v_1)sgn(w_{[v,v_1]}) = \varepsilon(v)$, by (2) in the definition of $\mathcal{R}(S)$. This proves the remark.

We can now prove the proposition. Let e be an edge of $\Delta(S) \setminus F$, with label φ. In Δ, let $u = o(e)$, $v = t(e)$, so $\pi(u)$, $\pi(v)$ are in the same component of $\Delta(S)$. Let $e_2 \dots e_p$ be the successive edges in the reduced path in F from $\pi(u)$ to $\pi(v)$, and let $\varphi_{i_j}^{\epsilon_j}$ be the label on e_j. Put $u_j = o(e_j)$, $v_j = t(e_j)$ in Δ and let $u_{p+1} = v$, $v_1 = u$. Now for $1 \leq j \leq p$, $\pi(v_j) = \pi(u_{j+1})$, i.e. v_j

and u_{j+1} are $\mathcal{R}(S)$-equivalent. Let $w_j = w_{[u_{j+1},v_j]}\varphi_{i_j}^{\epsilon_j}$ for $2 \leq j \leq p$ and let $w = w_p \dots w_2 w_{[u_2,v_1]}$. Now $w_{[u_2,v_1]} = w_{[u_2,u]}$ is defined on $J = [u(S), u(S) - 2\varepsilon(u)t_0]$, and maps it to $[u_2(S), u_2(S) - 2\varepsilon(u_2)t_0]$, as noted in the proof of the remark. Successively applying the remark to $w_2, \dots, w_p$, we see that $\tilde{w}$ is defined on $J = [u(S), u(S) - 2\varepsilon(u)t_0]$ and maps it to $J' = [v(S), v(S) - 2\varepsilon(v)t_0]$. Hence $sgn(w) = \varepsilon(u)\varepsilon(v) = sgn(\varphi)$, so $\imath\tilde{w}(x) = \varphi(x)$ for all $x \in J$. Thus $\varphi(x)$ is in the same $\tilde{S}$-orbit as x, and it follows that S and $\tilde{S}$ have the same orbits. □

An edge e of the Cayley graph is called singular if $o(e)$ is an endpoint of the domain of ψ, where ψ is the label on e, and regular otherwise. If S is non-degenerate and e is singular, there is a unique edge e' of Δ with the same label as e, such that if $o(e') = v$ then $o(e) = v(S)$, giving a bijection between singular edges and edges of Δ.

The next result explains the significance of the graph $\Delta(S)$.

Proposition 3.6. *A non-degenerate system S has independent generators if and only if $\Delta(S)$ is a forest.*

Proof. Suppose S does not have independent generators, so there is a non-trivial reduced S-word w fixing every point of its domain, which is a non-degenerate compact interval. Choose such a word w of minimal length and let its domain be $[x, x + \delta]$, where $\delta > 0$. Thus w determines a closed path p_w from x to x in the Cayley graph. Note that, if $w = uvu^{-1}$, where u, v are S-words, then v fixes $u[x, x + \delta]$ pointwise. Since w has minimal length, it is therefore cyclically reduced, and replacing it by a cyclic conjugate if necessary, we may suppose $p_w = e_1\gamma_1 e_2\gamma_2 \dots e_p\gamma_p$, where γ_j is a sequence of regular edges and e_j is a singular edge. (The γ_j may be empty, but there is at least one e_j since otherwise the domain of w would strictly contain $[x, x + \delta]$.) Let w_j be the word obtained by writing down from right to left, in succession, the labels on the edges of γ_j, let e'_j be the edge of Δ associated with e_j and let $u_j = o(e'_j)$, $v_j = t(e'_j)$ in Δ. Also, let ψ_j be the label on e_j. Thus $o(e_j) = u_j(S)$, $t(e_j) = v_j(S)$, $\psi_j(u_j(S)) = v_j(S)$ and $w_j(v_j(S)) = u_{j+1}(S)$ for $1 \leq j \leq p$, with indices taken mod p. Also, $w = w_p\psi_p \dots w_1\psi_1$. We claim that v_j and u_{j+1} are $\mathcal{R}(S)$-equivalent for $1 \leq j \leq p$, where $u_{p+1} = u_1$. Since $w_j(v_j(S)) = u_{j+1}(S)$, and w_j is defined in a neighbourhood of $v_j(S)$ (being obtained from a sequence of regular edges) it suffices to show that condition (2) in the definition of $\mathcal{R}(S)$ is satisfied, using the word w_j. Now w has domain an interval of length δ, hence the domain of ψ_j contains the interval $[u_j(S), u_j(S) - \varepsilon(u_j)\delta]$, which it maps to $[v_j(S), v_j(S) - \varepsilon(v_j)\delta]$. Further, each w_j is defined on $[v_j(S), v_j(S) - \varepsilon(v_j)\delta]$, which it maps to $[u_{j+1}(S), u_{j+1}(S) - \varepsilon(u_{j+1})\delta]$. But w_j maps $[v_j(S), v_j(S) - \varepsilon(v_j)\delta]$ to $[u_{j+1}(S), u_{j+1}(S) - \varepsilon(u_j)sgn(w_j)\delta]$. Condition (2) follows.

It follows that $e'_1 \dots e'_p$ is the sequence of edges in a closed path in $\Delta(S)$. If this path is not reduced, then $e'_{i+1} = \bar{e}'_i$ for some i, so $e_{i+1} = \bar{e}_i$ and we can remove the subword of w corresponding to $e_i\gamma_i e_{i+1}$ (i.e. $\psi_{i+1}w_i\psi_i = \psi_i^{-1}w_i\psi_i$), which is not equal to w since w is cyclically reduced, to obtain a word of shorter length than w whose restriction to an interval of length δ is the identity map, a contradiction. It follows that $\Delta(S)$ is not a forest (see the proof of Lemma 3.1).

Conversely, if $\Delta(S)$ is not a forest, choose an edge e outside a maximal forest. Then we can construct w and φ as in the proof of Proposition 3.5, and by deletions we can replace w by a reduced word w' such that w is a restriction of w'. Since $\tilde{w}$, and so $\tilde{w}'$, is defined on the interval J in Proposition 3.5, while $\tilde{\varphi}$ is not, $w \neq \varphi$. After a further deletion if necessary, the word $\varphi^{-1}w$ gives a non-empty reduced word fixing the interval J of positive length, so S does not have independent generators. □

The strategy of the proof of Theorem 3.1 is to first prove it when S has finite orbits, then to deduce the general case using Lemma 3.4, Proposition 3.5 and Theorem 2.6. To prove the finite orbit case, we use Proposition 3.6 and the following lemma. We need an elaboration of notation used in §2. Denote by $E(S)$ the union of the finite singular $\mathring{S}$-orbits and the finite orbits containing the unique fixed point of some reflection (i.e. a word in the generators of S, whose domain is a non-degenerate interval, having a fixed point and of sign -1). Also, $E'(S)$ denotes the union of the finite singular $\mathring{S}$-orbits. By Theorem 2.1, $E(S)$ is finite.

Lemma 3.7. *Let S be a non-degenerate system of isometries whose orbits are all finite. Let T be the least distance between two points of $E(S)$. Let F' be the lift of a maximal forest F of $\Delta(S)$ to Δ, and let e be an edge in $\Delta \setminus F'$. Then there exists $t \geq T$ such that the system $\tilde{S}$ obtained by shortening by t on the side of e has the same orbits as S, and such that:*
(1) *$E(\tilde{S}) \subseteq E(S)$;*
(2) *There is a maximal forest of $\Delta(\tilde{S})$ whose lift $\tilde{F}'$ to Δ contains F'.*

Proof. Suppose e has label φ, $o(e) = u$ and $t(e) = v$ in Δ. Let I be the connected component of $D \setminus E(S)$ which contains the point $u(S) - \varepsilon(u)\delta$ for all sufficiently small δ. Take t to be the length of the interval I, so $I = (u(S), u(S) - \varepsilon(u)t)$. Let $K = [u(S), u(S) - \varepsilon(u)t)$, the interval removed from the domain of φ to obtain $\tilde{S}$.

We use the notation in the proof of Proposition 3.5. Let $e_2 \dots e_p$ be the successive edges in the reduced path in F from $\pi(u)$ to $\pi(v)$, let $\varphi_{i_j}^{\epsilon_j}$ be the label on e_j, and let $u_j = o(e_j)$, $v_j = t(e_j)$ in Δ. For $1 \leq j \leq p$, putting $u_{p+1} = v$ and $v_1 = u$, let $w_j = w_{[u_{j+1},v_j]}\varphi_{i_j}^{\epsilon_j}$ for $2 \leq j \leq p$ and let $w = w_p \dots w_2 w_{[u_2,v_1]}$. From the proof of Proposition 3.5, $\tilde{w}$ is defined on $J = [u(S), u(S) - 2\varepsilon(u)t_0]$ for some $t_0 > 0$.

Since I contains no point of a singular orbit, w is defined on all of K. It follows easily from Theorem 2.1 that, for all $x \in K$, $S(x) \cap K = \{x\}$ (recall

$S(x)$ is the S-orbit of x). For $x \in K$, $w(x) = \varphi(x)$ (since w and φ agree on a non-degenerate subinterval of K; see the proof of Proposition 3.5). Let p_w be the sequence of successive edges in the path in the Cayley graph from x to $\varphi(x)$ determined by w. We claim that p_w contains no edge whose origin y is in K and whose label is φ.

For otherwise, w can be written $w = w''\varphi w'$ where $w'(x) = y \in K$, so $y \in K \cap S(x)$, hence $y = x$. If w' is the restriction of a reflection with fixed point x, then $x \in E$, so $x \notin I$, hence $x = u(S)$. But since w is defined on K, $w'(K)$ is contained in the domain of φ, so w cannot be a reflection in $u(S)$, a contradiction. Hence w' is the restriction of a translation, and fixes x, so w' restricted to K is the identity map, in particular, $w'(u(S)) = u(S)$. But the occurrence of φ in w cannot be in one of the $w_{[u_{j+1},v_j]}$, because then w' can be written as the product of two subwords z_1z_2, where $z_2(u(S)) = v_j$, $z_1(v_j) = u(S)$, and φz_1 is a subword of $w_{[u_{j+1},v_j]}$. This contradicts the requirement that $w_{[u_{j+1},v_j]}$ must be defined in an open interval around v_j (condition (1) in the definition of $\mathcal{R}(S)$), because φ is not defined in an open neighbourhood of $u(S)$, which is an endpoint of its domain. Thus this occurrence of φ is one of the $\varphi_{i_j}^{\epsilon_j}$, and $u = v_j$. Hence $e_j = e$, so $e \in F$, a contradiction. This establishes the claim.

It follows that the isometry $\tilde{w}$ is defined on K. The Cayley graph of $\tilde{S}$ is obtained from that of S by removing the edge labelled φ with origin x, for each $x \in K$. This does not disconnect the component of x in the Cayley graph of S, whose vertex set is $S(x)$ (from the above, $S(x) \cap S(y) = \emptyset$ for $x, y \in K$ with $x \neq y$, so only one edge is removed from the component of x, and its endpoints are joined by the path p_w in the Cayley graph of $\tilde{S}$). It follows that S and $\tilde{S}$ have the same orbits. Further, by definition of t, $u(S) - \varepsilon(u)t \in E(S)$, and (1) follows.

Finally, if $v_1, v_2 \in V \setminus \{u, v\}$ (recall that V is the vertex set of Δ), and v_1, v_2 are $\mathcal{R}(\tilde{S})$-equivalent, then they are $\mathcal{R}(S)$-equivalent. This follows since $v_i(S) = v_i(\tilde{S})$ and $\tilde{S}$ is a restriction of S. Statement (2) of the lemma now follows easily. □

Remark. If, in the proof of Lemma 3.7, $\tilde{\varphi}$ is a singleton, its domain is $\{u(\tilde{S})\}$, and $u(\tilde{S}) = u(S) - \varepsilon(u)t$. The domain of $\tilde{w}$ is a compact interval and contains K, so contains its endpoint $u(\tilde{S})$, and is non-degenerate, hence φ does not occur in $\tilde{w}$. Further, $w(u(\tilde{S})) = \varphi(u(\tilde{S}))$ since w and φ agree on K. Therefore, removing the edge with label $\tilde{\varphi}$ and origin $u(\tilde{S})$ from the Cayley graph of $\tilde{S}$ does not change its components. that is, removing the singleton $\tilde{\varphi}$ from $\tilde{S}$ does not change the orbits.

We can now prove Theorem 3.1 in the case of finite orbits.

Theorem 3.8. *Let S be a non-degenerate system of isometries, with all orbits finite, and let F be a maximal forest in $\Delta(S)$. Let F' be the lift of F to Δ.*

Then there exist edges $e_1, \ldots, e_p$ of $\Delta \setminus F'$ and real numbers $t_1, \ldots, t_p \geq 0$ such that the system $\tilde{S}$ obtained by shortening S by t_i on the side of e_i for $1 \leq i \leq p$ has the same orbits as S and independent generators. The system obtained by removing all singletons from $\tilde{S}$ also has the same orbits as S and independent generators.

Proof. If $\Delta(S)$ is not a forest, we can repeatedly apply Lemma 3.7. (At each stage we remove a singleton if necessary before proceeding to the next stage, which does not change the orbits, by the remark after Lemma 3.6.) This terminates eventually since the lengths by which one shortens are bounded below by T, which is a uniform bound for these lengths by (1) in Lemma 3.7. Thus it terminates with a system S' with $\Delta(S')$ a forest. In view of (2) in Lemma 3.7, one only shortens on the side of edges not in F. The system S' has independent generators by Proposition 3.6. Adding the removed singletons does not change the orbits or independence of the generators, and results in the required system $\tilde{S}$. □

Before proving the general case, we need another criterion for independent generators, in terms of certain numerical invariants associated with a system of isometries S.

Definition. If U is a family of finite orbits, we define the *width* of U to be the common length of the disjoint intervals whose union is U. For a system of isometries S, we put

$$m(S) = \text{the total length of } D;$$
$$l(S) = \text{the sum of the lengths of the domains } A_i \text{ of the generators};$$
$$e(S) = \text{the sum of the widths of the finite orbits.}$$

Our criterion for independent generators involves these three numbers, but to prove it, another numerical invariant is involved.

Definition. For a system of isometries S, $c(S)$ is the number of edges in a maximal forest of $\Delta(S)$. Equivalently, it is the number of vertices minus the number of components of $\Delta(S)$ (see the exercise preceding Lemma 1.3.3).

Lemma 3.9. *Suppose the system S is non-degenerate and without minimal homogeneous component. The function $t \mapsto e(S_{-t})$ is continuous, piecewise linear and has right-hand derivative $c(S)$ at 0.*

Proof. We first show that $t \mapsto e(S_{-t})$ is continuous. In passing from S_{-t} to S_{-t+s}, at most the orbits which pass through $4k$ intervals of length $|s|$ are affected, so $|e(S_{-t}) - e(S_{-t+s})| \leq 4k|s|$, showing continuity.

To show the function is piecewise linear, we first assume S has finite orbits, so that S_{-t} does. It is sufficient to show that it is linear on some interval

$[0, t_0]$, for then we may apply this result to S_{-t} and use a simple compactness argument to show the function is piecewise linear. Choose t_0 such that the distance between two distinct points of $E(S)$ is greater than $4t_0$, and so that $\mathcal{R}(S)$ and $\mathcal{R}(S_{-t})$ are the same for $t \in [0, t_0]$ (using Lemma 3.4). Take $t \in [0, t_0]$. Note that the smallest distance between two points of $E(S_{-t})$ is greater than $2t$.

For $v \in V$ (recall V is the vertex set of Δ), put $I_v = (v(S), v(S) - \varepsilon(v)t)$. Also, let U_v be the union of all S-orbits containing a point of I_v, and let U_v^t be the union of all S_{-t}-orbits containing a point of I_v. Now by the choice of t, I_v contains no point of a singular orbit (for S or for S_{-t}), so if an S-word w is defined at $x \in I_v$, then w is defined on $\overline{I}_v$, and if w_{-t} is defined at x, then w_{-t} is defined on $\overline{I}_v$. It follows that, if v, $v' \in V$, then either $U_v = U_{v'}$ or $U_v \cap U_{v'} = \emptyset$, and either $U_v^t = U_{v'}^t$ or $U_v^t \cap U_{v'}^t = \emptyset$. We investigate when $U_v = U_{v'}$ and when $U_v^t = U_{v'}^t$.

We claim that $U_v = U_{v'}$ if and only if $\pi(v)$, $\pi(v')$ are in the same component of $\Delta(S)$. (Recall $\pi : \Delta \to \Delta(S)$ is the projection map.) Suppose $U_v = U_{v'}$; there is an S-word w and $x \in I_v$ such that $w(x) \in I_{v'}$. As noted above, $w(\overline{I}_v) = \overline{I}_{v'}$, and by the choice of t, $w(v(S)) = v'(S)$. The path p_w is a path from $v(S)$ to $v'(S)$ in the Cayley graph of S. We can write $p_w = \gamma_0 e_1 \gamma_1 e_2 \gamma_2 \ldots e_p \gamma_p$, where γ_j is a sequence of regular edges and e_j is a singular edge. Let e_j' be the edge of Δ associated with e_j and let $u_j = o(e_j')$, $v_j = t(e_j')$ in Δ. As in the proof of Proposition 3.6, we can write $w = w_p \psi_p \ldots w_1 \psi_1 w_0$. Now w_0 is defined in a neighbourhood of $v(S)$ and maps $v(S)$ to $u_1(S)$. Also, it maps $\bar{I}_v$ to a subset of the domain of ψ_1, which is therefore $[u_1(S), u_1(S) - \varepsilon(u_1)t]$. But w_0 sends $\bar{I}_v$ to $[u_1(S), u_1(S) - \varepsilon(v) sgn(w_0)t]$. Hence $v\mathcal{R}(S)u_1$ using the word w_0. Arguing as in the proof of Prop. 3.6, $v_j\, \mathcal{R}(S) u_{j+1}$ for $1 \le j \le (p-1)$, and $\psi_p \ldots w_1 \psi_1 w_0$ maps $\bar{I}_v$ to $[v_p(S), v_p(S) - \varepsilon(v_p)t]$, and w_p maps this interval to $\bar{I}_{v'}$, since $w(\overline{I}_v) = \overline{I}_{v'}$. Hence $v_p \mathcal{R}(S) v'$ using the word w_p. Thus $e_1' \ldots e_p'$ is a path in $\Delta(S)$ from $\pi(v)$ to $\pi(v')$, so $\pi(v)$, $\pi(v')$ are in the same component of $\Delta(S)$.

For the converse, suppose first that $v\mathcal{R}(S)v'$. This means there is an S-word w such that $\mathring{w}(v(S)) = v'(S)$ and $sgn(w) = \varepsilon(v)\varepsilon(v')$. Then w maps some subinterval of I_v with endpoint $v(S)$ onto a subinterval of $I_{v'}$ with endpoint $v'(S)$, hence $U_v \cap U_{v'} \neq \emptyset$, so $U_v = U_{v'}$.

Next, suppose there is an edge in Δ joining v and v'. Then there is a generator or its inverse φ such that $\varphi(v(S)) = v'(S)$, and I_v is contained in the domain of φ, so $\varphi(I_v) = I_{v'}$. Hence $U_v \cap U_{v'} \neq \emptyset$, so $U_v = U_{v'}$.

It follows that, if $\pi(v)$, $\pi(v')$ are connected by an edge in $\Delta(S)$, then $U_v = U_{v'}$. Hence if $\pi(v)$, $\pi(v')$ are in the same component of $\Delta(S)$ we have $U_v = U_{v'}$, as required.

We claim that $U_v^t = U_{v'}^t$ if and only if $v\mathcal{R}(S_{-t})v'$. Suppose $U_v^t = U_{v'}^t$. There is an S-word w and $x \in I_v$ such that $w_{-t}(x) \in I_{v'}$. Then $w_{-t}(\overline{I}_v) = \overline{I}_{v'}$, and by the choice of t, w_{-t} is defined in a neighbourhood of $v(S)$ and so the same is true of w. Also, $w(\overline{I}_v) = \overline{I}_{v'}$ and by choice of t, w sends $v(S)$ to $v'(S)$. It follows that $sgn(w) = \varepsilon(v)\varepsilon(v')$. Hence $v\mathcal{R}(S)v'$. Again by the choice of t, $v\mathcal{R}(S_{-t})v'$.

Conversely, if $v\mathcal{R}(S_{-t})v'$, then a similar argument to that given when $v\mathcal{R}(S)v'$ shows that $U_v^t = U_{v'}^t$.

In passing from S to S_{-t}, orbits not in $E(S)$ and outside $\bigcup_v U_v$ are not affected, while for those orbits in $\bigcup_v U_v$, certain edges of the Cayley graph are removed, and the endpoints of these edges are in $\bigcup_v I_v$. Hence U_v is the union of certain of the sets $U_{v'}^t$. Now $I_v \subseteq D \setminus E(S)$, where D is the domain of S, and from what we have proved and Theorem 2.1, an S-orbit $S(x)$ with $x \in I_v$ splits into n_v S_{-t}-orbits, where n_v is the number of vertices of the component of $\pi(v)$ in $\Delta(S)$. Further, the orbit of I_v contributes t to $e(S)$, and this splits into orbits which contribute $n_v t$ to $e(S_{-t})$, an increase of $(n_v - 1)t$. Summing over all the components of $\Delta(S)$ gives $e(S_{-t}) = e(S) + c(S)t$, as required.

In general, when S has no homogeneous minimal component, by Theorem 2.5 and what has been proved, $e(S_{-t-s}) = e(S_{-s}) + c(S_{-s})t$, for sufficiently small s, t, since $S_{-t-s} = (S_{-s})_{-t}$. Also, by Lemma 3.4, $c(S_{-s}) = c(S)$ for sufficiently small s. Taking limits as $s \to 0$, by continuity of e we have $e(S_{-t}) = e(S) + c(S)t$, concluding the proof. □

We can now state our criterion for independent generators.

Proposition 3.10. *Let S be non-degenerate without homogeneous component. Then $e(S) + l(S) \geq m(S)$, with equality if and only if S has independent generators.*

Proof. We begin by showing that $e(S) + l(S) \geq m(S)$ assuming all orbits are finite. Suppose $U_1, \ldots, U_p$ are the families of finite orbits. For $x \in U_i$, the component $\Gamma(x)$ of the Cayley graph of S containing x is independent, up to isomorphism, of the choice of x in U_i, and we denote this labelled graph by Γ_i. Let v_i be the number of vertices of Γ_i, let r_i be the number of unoriented edges of Γ_i, and let e_i be the width of U_i. Note that r_i is the number of oriented edges of Γ_i with label a generator (rather than the inverse of a generator). Also, v_i is the number of intervals of U_i. We claim that

$$e(S) = \sum_{i=1}^{p} e_i, \quad l(S) = \sum_{i=1}^{p} r_i e_i \quad \text{and} \quad m(S) = \sum_{i=1}^{p} v_i e_i.$$

These are clear except for the middle equality. To see this, note that $A_j \setminus E(S)$ is a disjoint union of intervals, each of which is an interval of one of the U_i, and $|A_j|$ is the sum of the lengths of the intervals of $A_j \setminus E(S)$. Further, there is a one-to-one correspondence between intervals of $A_j \setminus E(S)$ contained in U_i

and edges labelled φ_j in Γ_i, hence

$$|A_j| = \sum_{i=1}^{p}(\text{the number of intervals of } A_j \setminus E(S) \text{ contained in } U_i)e_i$$
$$= \sum_{i=1}^{p}(\text{the number of directed edges in } \Gamma_i \text{ with label } \varphi_j)e_i.$$

Summing over j,

$$l(S) = \sum_{j=1}^{k}|A_j| = \sum_{j=1}^{k}\sum_{i=1}^{p}(\text{the number of directed edges in } \Gamma_i \text{ with label } \varphi_j)e_i$$
$$= \sum_{i=1}^{p}\sum_{j=1}^{k}(\text{the number of directed edges in } \Gamma_i \text{ with label } \varphi_j)e_i$$
$$= \sum_{i=1}^{p} r_i e_i$$

as claimed.

Now $1 + r_i \geq v_i$, with equality if and only if Γ_i is a tree (see the exercise preceding Lemma 1.3.3) and the required inequality follows, on multiplying by e_i and summing over i. Further, if S has independent generators, then Γ_i is a tree for $1 \leq i \leq p$, by Lemma 3.3, hence $e(S) + l(S) = m(S)$.

In the general case, note that the function $t \mapsto l(S_{-t})$ is continuous for sufficiently small t, in fact, it is clearly linear: $l(S_{-t}) = l(S) - 2kt$. By Lemma 3.9, $e(S) + l(S) = e(S_{-t}) + l((S_{-t}) + (2k - c(S))t$ for sufficiently small t, and since $\Delta(S)$ has $2k$ edges, $c(S) \leq 2k$, with equality if and only if $\Delta(S)$ is a forest. By Theorem 2.6 and what we have proved, $e(S) + l(S) \geq m(S)$, and if $e(S) + l(S) = m(S)$, $\Delta(S)$ is a forest, so S has independent generators by Proposition 3.6.

Conversely, if S has independent generators, so does S_{-t}, and again by Theorem 2.6 and what we have proved $e(S_{-t}) + l(S_{-t}) = m(S_{-t}) = m(S)$ for small $t > 0$. It follows on taking limits that $e(S) + l(S) = m(S)$. □

We can now prove Theorem 3.1, whose statement we recall:

Let $S = (D, \{\varphi_i\}_{i=1,\ldots,k})$ be a system of isometries without a minimal homogeneous component. Then there is a system $S' = (D, \{\varphi_i'\}_{i=1,\ldots,k})$ with independent generators, having the same orbits as S, where each φ_i' is the restriction of φ_i to a closed subinterval of its domain. If S is non-degenerate, we can remove all singletons from S' without changing the orbits.

Proof. First assume S is non-degenerate. The proof is in 4 steps. Choose a maximal forest F in $\Delta(S)$, and let F' be its lift to Δ. Also, choose $t > 0$ small enough that $\Delta(S) = \Delta(S_{-t})$ (using Lemma 3.4) and S_{-t} has finite orbits (Theorem 2.6), and such that S_{-t} is non-degenerate. In view of Lemma 3.10, we can also assume that $e(S_{-t}) = e(S) + c(S)t$. In all these steps the domain D does not change, and we denote the total length of D by m.

First, we replace S by S_1, obtained by shortening S by t on the sides of the edges in $\Delta \setminus F'$. By Proposition 3.5, S_1 has the same orbits as S. Next, we replace S_1 by S_2 by shortening S_1 on the sides of the edges in F'. Thus $S_2 = S_{-t}$, so by the choice of t, $\Delta(S_2) = \Delta(S)$, and F is a maximal forest of $\Delta(S_2)$. Note that $e(S_2) = e(S) + c(S)t$, and S_2 has finite orbits, by choice of t.

We now apply Theorem 3.8 to the system S_2 to obtain a non-degenerate system S_3 with the same orbits as S_2 and independent generators. It follows that $e(S_3) = e(S_2)$, consequently

$$e(S_3) = e(S) + c(S)t. \tag{1}$$

By Proposition 3.10,

$$e(S_3) + l(S_3) = m. \tag{2}$$

Note that in this step one only shortens on the side of edges not in F', so no edges of F' are deleted as a result of removing singletons. The final step is to lengthen by t on the side of all edges of F', to obtain a non-degenerate system S_4. This can be done in view of Step 2. Since F' has $c(S)$ edges, we have

$$l(S_4) = l(S_3) + c(S)t. \tag{3}$$

We claim that S_4 is the desired system S'. First we have to show that S_4 has the same orbits as S. Let $\Gamma(x)$ be the component of the Cayley graph of S containing the point $x \in D$. In passing from S to S_1, certain edges of Γ are removed, resulting in a graph $\Gamma_1(x)$, which is connected since S and S_1 have the same orbits. Now in passing from S_1 to S_2, further edges are removed from $\Gamma_1(x)$ to obtain a graph $\Gamma_2(x)$ which is the union of a family of finite connected subgraphs, say $\Gamma_2^i(x)$, for i in some index set I, since S_2 has finite orbits. In the next step, from S_2 to S_3, further edges are deleted from $\Gamma_2(x)$ to obtain a graph $\Gamma_3(x)$ which is the union of a family $\{\Gamma_3^i(x) \mid i \in I\}$, obtained from the $\Gamma_2^i(x)$ by deleting edges. However, the $\Gamma_3^i(x)$ remain connected since S_2 and S_3 have the same orbits. Finally, a graph $\Gamma_4(x)$ is obtained from $\Gamma_3(x)$ by replacing all the edges removed in passing from $\Gamma_1(x)$ to $\Gamma_2(x)$. This reconnects all the subgraphs $\Gamma_3^i(x)$, into a single component $\Gamma_4(x)$. It follows that S_4 and S have the same orbits (the vertex set of every $\Gamma_i(x)$ is the orbit $S(x)$).

It remains to see that S_4 has independent generators. By what we have just proved, $e(S) = e(S_4)$. Therefore:

$$\begin{aligned} e(S_4) + l(S_4) &= e(S) + (l(S_3) + c(S)t) && \text{by (3)} \\ &= (e(S_3) - c(S)t) + (l(S_3) + c(S)t) && \text{by (1)} \\ &= e(S_3) + l(S_3) = m && \text{by (2).} \end{aligned}$$

It follows from Proposition 3.10 that S_4 has independent generators.

Finally, if S is degenerate, first remove all singletons from S and apply the argument given to the new, non-degenerate system to obtain a system with independent generators. Adding the singletons originally removed from S to this system gives the desired system S'. □

In the next section we shall use Theorem 3.1, or rather its corollary, 3.2, to finish the proof of Rips' Theorem.

4. Interval Exchanges and Conclusion

The final steps in the proof of Theorem 2.4 are to first consider another special case, that of an *interval exchange*, and then give a procedure for modifying a pure, minimal system S with independent generators which either stops after a finite number of steps with an interval exchange, when $G(S)$ is a surface group, or continues indefinitely. In this last case, we show that $G(S)$ is a free group.

Definition. A system of isometries S with domain D is called an interval exchange if it is connected, non-degenerate and for all but finitely many $x \in D$, x belongs to exactly two bases.

Note that the number of bases to which a point x belongs means the sum of the number of i for which $x_i \in A_i$ and the number of i for which $x \in B_i$, where $\varphi_i : A_i \to B_i$ are the generators. Otherwise put, it is the degree of x in the Cayley graph of the system. The proof of our first lemma uses the following remark. We denote the length of an interval J in $\mathbb{R}$ by $\mu(J)$.

Remark. Let I be an interval in $\mathbb{R}$, and let $I_1, \ldots, I_n$ be subintervals of I, such that every point of I is in I_j for at least two values of j. Then $\mu(I) \leq (1/2)\sum_{j=1}^{n} \mu(I_j)$, with strict inequality if there are distinct indices j, k and l such that $I_j \cap I_k \cap I_l$ is infinite (i.e. a non-degenerate interval).

The case $I = \emptyset$ is clear, otherwise we can renumber so that the left-hand endpoint a of I is the left-hand endpoint of I_1 and I_2, and if $a \in I$, then $a \in I_1$ and $a \in I_2$, and also $\mu(I_1) \leq \mu(I_2)$. Let $I' = I \setminus I_1$. Then every point of I' is in at least two sets in the sequence $I_2 \cap I', \ldots, I_n \cap I'$. The remark follows by induction on n; details are left to the reader.

Lemma 4.1. *Let S be connected, non-degenerate with independent generators, and suppose S has only finitely many finite orbits, and every $x \in D$ belongs to at least two bases. Then S is an interval exchange.*

Proof. Since there are only finitely many finite orbits, in Proposition 3.10 the number $e(S) = 0$, so $l(S) = m(S)$. By the remark, summing over all components I of the domain D, the intersection of three distinct indexed bases is finite, so the union of all these triple intersections is finite, that is, finitely many points of D belong to more than two bases, hence S is an interval exchange. □

Note that, by the remark after Theorem 2.1, if S is minimal, all S-orbits are dense, in particular, they are infinite.

Proposition 4.2. *If S is a pure, minimal interval exchange, then $G(S)$ is a surface group.*

Proof. Recall from §1 that, if L is a leaf of the foliation of S, there is a continuous bijection $f : \text{real}(\Gamma(x)) \to L$, where x is a point of $D \cap L$, D is the domain of S and $\Gamma(x)$ is the component of the Cayley graph containing x. Also, restricted to $\text{real}(\Delta)$, where Δ is a finite subgraph of $\Gamma(x)$, f is a homeomorphism. Further, a path in L is the image under f of a unique path in $\text{real}(\Gamma(x))$, and this path is contained in $\text{real}(\Delta)$, for some finite subgraph Δ of $\Gamma(x)$. Certain edges of the Cayley graph may correspond to segments of the leaf of the form $\{a\} \times [0, 1]$, where a is an endpoint of a base. We call such edges boundary edges.

Note that a point of D is in at least two bases. It must be in at least one since its S-orbit is infinite, and if it is in exactly one, all nearby points would be in at most one base, contrary to assumption.

Let O be an S-orbit in which each point belongs to exactly two bases. Note that all S-orbits are infinite, and all points of O have degree 2 in the Cayley graph. Since O is infinite, the corresponding component of the Cayley graph is a copy of L, the graph with vertex set $\mathbb{Z}$ and an edge joining n to $n + 1$ for each n. In particular, it is a tree, so any path in the corresponding leaf is nullhomotopic. (It is the image of a path in $\text{real}(\Delta)$, where Δ is a finite tree, so $\text{real}(\Delta)$ is an $\mathbb{R}$-tree, hence contractible, by Lemma 2.2.2, Proposition 2.2.5 and Lemma 2.2.6.)

Now consider $S(x)$, where x is a point in more than two bases. First, x cannot belong to the interior of more than 1 base, otherwise there would be a non-degenerate interval $[x, y]$, all of whose points belong to at least 3 bases. Thus x is an endpoint of at least two bases.

Suppose x is an endpoint of at least three bases. Take three such bases; they must be of the form $[x, a]$, $[x, b]$, $[x, c]$, where $[x, c]$ meets $[x, a]$ and $[x, b]$ only in x. Otherwise there is again a non-degenerate interval $[x, y]$ with all points in at least three bases. Then x cannot be an endpoint of D, so by purity the $\mathring{S}$-orbit of x is infinite. In particular, x belongs to the interior of a base, say B. But then $[x, a] \cap [x, b] \cap B$ is a non-degenerate interval with all points in at least three bases, a contradiction.

It follows that x is an interior point of one base, and an endpoint of a further two, which must be of the form $[x, a]$ and $[x, b]$, where $a < x < b$, otherwise there is a non-degenerate interval of the form $[x, y]$ with all points in at least three bases. Thus x has degree three in the Cayley graph, and two of the edges incident with x are boundary edges. Also, it is easy to see that x has an open neighbourhood in $\Sigma(S)$ homeomorphic to a closed half plane. It also follows that x is not an endpoint of D, so by purity the $\mathring{S}$-orbit of x is infinite.

If y has degree 2 in the Cayley graph, it cannot be an endpoint of one base and an interior point of another, otherwise there will be a non-degenerate interval of points belonging to only one base. Thus either both edges are boundary edges, in which case y has an open neighbourhood in $\Sigma(S)$ homeomorphic to a closed half plane, or neither edge is a boundary edge, in which case y has an open neighbourhood homeomorphic to the plane. Clearly a point in the interior of a band $A \times [0, 1]$, where A is a base, has an open neighbourhood homeomorphic to the plane (an open disk), and a point of $\{a\} \times (0, 1)$, where a is an endpoint of A, has an open neighbourhood homeomorphic to a closed half plane. We have now shown that $\Sigma(S)$ is a 2-manifold with boundary.

If x is a vertex of degree 3 in the Cayley graph, two of the incident edges are removed to form the Cayley graph of $\mathring{S}$. If y is a vertex of degree 2, it is of degree 0 or 2 in the Cayley graph of $\mathring{S}$. Since the $\mathring{S}$-orbit of x is infinite, it follows that the component of the Cayley graph of $\mathring{S}$ containing x is isomorphic to L^+, the graph with vertex set $\mathbb{N}$, and edges joining n and $n+1$ for all $n \in \mathbb{N}$, the isomorphism sending x to 0. Note that there are only finitely many points of D incident with a boundary edge, since such a point is an endpoint of a base.

Therefore, the component of x in the Cayley graph of S consists of a circuit, whose edges are boundary edges, together with copies of L^+ attached to some vertices of the circuit, and there are only finitely many such components. Considering the leaves corresponding to these components, we conclude that the boundary of Σ consists of finitely many circles, say $C_1, \dots, C_p$, corresponding to the circuits in these components. It also follows that any closed path in the leaf corresponding to such a component is homotopic to a power of a closed path which goes round one of the C_i exactly once. Choose such a closed path $\alpha_i : [0, 1] \to C_i$ for $1 \leq i \leq p$. Then $N(S)$ is generated as a normal subgroup of $\pi_1(\Sigma(S))$ by the homotopy classes of $\alpha_1, \dots, \alpha_p$.

Each α_i induces a map $\beta_i : S^1 \to C_i$, and we can attach p disks to $\Sigma(S)$ using these as attaching maps. The result is a compact surface, whose fundamental group is isomorphic to $\pi_1(\Sigma(S))/N(S)$, i.e. to $G(S)$. This is a consequence of van Kampen's Theorem (see [39; Prop.4.11] or [98; Ch.7, Theorem 2.1]). Thus $G(S)$ is a surface group. □

We now give the procedure mentioned above, due to Rips. Let $S_0 = (D_0, \{\varphi_i\})$ be a minimal, pure system of isometries. Let $L(S_0)$ be the set of points of D_0 belonging to only one base; it is open in D_0 and a union of finitely many intervals. If it is not empty, define S_1 to be the system obtained from S_0 by replacing D_0 by $D_1 = D_0 \setminus L(S_0)$, which is clearly a multi-interval, and replacing every φ_i by its restrictions to the components of $D_1 \setminus \varphi_i^{-1}(L(S_0))$, which are finite in number and are compact intervals. Any singletons created by this procedure (when a component of $L(S_0)$ coincides with a component of D_0) are removed. Now every point of $\varphi_i^{-1}(L(S_0))$ is contained in a base other than the domain of φ_i. Otherwise, there would be a point x in $L(S_0)$ such that $\Gamma(x)$ has vertices x and $\varphi_i^{-1}(x)$, with an edge joining them. Also, X is not an endpoint of D. Since the orbit of x is finite, $x \notin D_0 \setminus E(S_0)$ by minimality, that is, $x \in E(S_0)$, contradicting purity of S_0 (the set $E(S_0)$ is defined just before Lemma 3.7). Hence $\varphi_i^{-1}(L(S_0))$ does not meet $L(S_0)$, so the closure of $\varphi_i^{-1}(L(S_0))$ does not meet $L(S_0)$. It follows that $\Sigma(S_1)$ is a strong deformation retract of $\Sigma(S_0)$. If C is a component of $(L(S_0))$, the band $C \times [0, 1]$ can be retracted onto the three sides belonging to $\Sigma(S_1)$, by a procedure similar to that in Lemma 1.1. It is easy to see that S_1 is still pure. The Cayley graph of $\mathring{S}_1$ is obtained from that of $\mathring{S}_0$ by deleting terminal vertices and the corresponding edges, and it follows easily that S_1 is minimal. It also follows from Remark 1 after Lemma 1.1 that a path in a leaf of S_0 is homotopic to one in S_1, and it follows that $G(S_1)$ is isomorphic to $G(S_0)$. Clearly, if the generators of S_0 are independent, then so are those of S_1.

We say that S_1 is obtained from S_0 by erasing intervals. If D_1 has points belonging to only one base, then we can perform this erasing procedure on S_1 to obtain S_2. Iteration leads to a sequence S_n of systems which either stops at $n = N$ if every point of D_N belongs to at least two bases, or continues indefinitely. Note that, if D_n is domain of S_n, then $D_0 \supset D_1 \supset D_2 \ldots$.

Lemma 4.3. *Suppose S_0 is pure, minimal and has independent generators. If the erasing process just described continues indefinitely, Then $\bigcap_{n=0}^{\infty} D_n$ is nowhere dense in D_0.*

Proof. Suppose on the contrary that there is a non-degenerate interval $I \subseteq \bigcap_{n=0}^{\infty} D_n$. If U is the minimal component of S_0, then U contains some non-degenerate subinterval of I. By the remark after Theorem 2.1, there is an integer N such that every $x \in D_0$ can be mapped into I by an S_0-word of length at most N. Every orbit $S_n(x)$ of S_n is the set of vertices of a component of the Cayley graph of S_n. By the distance between two points of $S_n(x)$, we mean the distance in the path metric. The Cayley graph of S_{n+1} is obtained from that of S_n by deleting all the terminal vertices and the corresponding edges, hence for $x \in S_n$, $\Gamma_n(x) \subseteq \Gamma_0(x)$, where $\Gamma_n(x)$ is the component of x in the Cayley graph of S_n. Since the erasing process can be carried out $N+1$ times, there is an S_{N+1}-orbit V and a point y such that, in the Cayley graph of S_0, y is a terminal vertex, $V \subseteq S_0(y)$ and the distance from y to V is $N+1$. Now $I \subseteq D_{N+1}$, so any points of I in $S_0(y)$ belong to V, hence are at distance at least $N+1$ from y in $S_0(y)$. But then no S_0-word of length at most N can send y to a point of I, contradicting the choice of N. □

We can now finish the proof of Theorem 2.4. We have already observed that, by Proposition 2.3, we can suppose S is pure and minimal. By Theorem 2.5 we can assume S is not homogeneous, and by Corollary 3.2 we can replace S by a system S' having independent generators. We apply Theorem 2.2 to S'. It is easily checked that the systems $S_1, \dots, S_r$ obtained also have independent generators. It therefore suffices to prove Theorem 2.4 for a system S_0 which is pure, minimal and has independent generators. We can also assume S_0 is not an interval exchange, by Proposition 4.2. Now apply the erasing process to $S = S_0$. If it stops after finitely many steps, with X_n, say, then X_n is an interval exchange by Lemma 4.1, and we are finished by Proposition 4.2, since $G(S_n) \cong G(S_0)$.

Suppose the erasing process can be continued indefinitely, and let Σ_n be the 2-complex $\Sigma(S_n)$. Let Γ be the Cayley graph of S_0. Again recall (§1) that, if $x \in D_0$, there is a continuous bijection f from real($\Gamma(x)$) onto the leaf L passing through x, where $\Gamma(x)$ is the component of Γ containing x. Also, a closed path in L is the image under f of a path in real(Δ), for some finite subgraph Δ of $\Gamma(x)$, and f restricted to real(Δ) is a homeomorphism onto its image, for any finite subgraph Δ. The Cayley graph F of $\mathring{S}_0$ is obtained from Γ by removing the finitely many edges with label φ (a generator or its inverse) whose origin is an endpoint of the domain of φ. By Lemma 3.3, F is a forest. It follows that the core graph C_x of each component $\Gamma(x)$ is finite (see the discussion after Lemma 1.3.2). Now let L be a leaf, with corresponding component $\Gamma(x)$. If $S_0(x)$ contains no endpoint of a base then $\Gamma(x)$ is a tree, since no edges of it were removed to obtain F, and $\Gamma(x)$ is locally finite, so from §2.2, real($\Gamma(x)$) is an $\mathbb{R}$-tree, hence contractible. It follows that any closed path in L is nullhomotopic. If x is an endpoint of a base, then given a path in L we may take the graph Δ above to be connected and to contain C_x. It is left as an exercise to show that real(C_x) is a deformation retract of real(Δ). Hence any closed path in L is homotopic to a closed path in f(real(C_x)). Thus we can find a finite 1-complex K in $\Sigma(S_0)$, homeomorphic to the disjoint union of the complexes real(C_x), where x runs through the finite set of endpoints of bases, such that $N(S_0)$ is generated as a normal subgroup of $\pi_1(\Sigma(S_0))$ by free homotopy classes of closed paths in K. Further, the 0-skeleton of K is $K \cap D_0$ and the edges (1-cells) of K have the form $\{a\} \times [0,1]$ for some a in a base.

Suppose first that there is at most one such edge in each band $A_i \times (0,1)$. Choose a maximal forest F' in K. Take a segment $\{a\} \times [0,1]$ in each band, choosing the segment in K whenever possible. Let X be the 1-complex which is the union of all these segments and D. Then X is connected, and $K \subseteq X$, so we can choose a maximal tree T in X containing F' (see the second exercise after Lemma 1.3.2). Also, X is a deformation retract of $\Sigma(S_0)$ (see Lemma 1.1), and from §1, $\pi_1(\Sigma(S_0))$ has a presentation with generators the edges e of X, and relations $e = 1$ for $e \in T$. To get $G(S)$, we add the relations $e_1 \dots e_r = 1$, where $e_1, \dots, e_r$ is a sequence of edges in K corresponding to a closed path in K. If $e \in K \setminus F'$, the endpoints of e are in the same component of K, so there is a path in F' joining them. Hence there is a relation $e_1 \dots e_r = 1$ with $e_1 = e$ and $e_2, \dots, e_r \in F'$, hence $e = 1$ in $G(S)$. We can use this to remove each such e from the generators by Tietze transformations. The relations are now of the form $e = 1$ for $e \in T$, and further relations $e_1 \dots e_r = 1$, where the e_i are all in F', which can be omitted. Hence $G(S_0)$ is a free group.

In general, by Lemma 4.3, we can perform the erasing process until all points of the finite set $K \cap D_0$ are in different components of D_n. (By the way K was defined, no points of $K \cap D_0$ are lost in the erasing process, since core graphs have no terminal vertices.) The argument just given then shows that $G(S_n)$ is free, and we have noted that $G(S_0) \cong G(S_n)$. This finishes the proof, and indeed, this book.

References

[1] S.A. Adeleke and P.M. Neumann, *Relations related to betweenness: their structure and automorphisms.* Memoirs Amer. Math. Soc. **623**. Providence, RI: Amer. Math. Soc. 1998.

[2] R.C. Alperin and H. Bass, "Length functions of group actions on Λ-trees". In: Combinatorial group theory and topology (ed. S.M.Gersten and J.R.Stallings), Annals of Math. Studies **111**, pp. 265-378. Princeton: University Press 1987.

[3] R.C. Alperin and K.N. Moss, "Complete trees for groups with a real-valued length function", *J. London Math. Soc.* **31** (1985), 55-68.

[4] M.F. Atiyah and I.G. Macdonald, *Introduction to commutative algebra.* Reading, Mass.: Addison-Wesley 1969.

[5] H. Bass, "Group actions on non-archimedean trees". In: Arboreal group theory, (ed. R.C. Alperin), MSRI Publications **19**, pp. 69-131. New York: Springer-Verlag 1991.

[6] B. Baumslag, "Residually free groups", *Proc. London Math. Soc.* (3) **17** (1967), 402–418.

[7] G. Baumslag, "On generalised free products", *Math. Z.* **78** (1962), 423–438.

[8] G. Baumslag, A. Myasnikov and V.N. Remeslennikov, "Algebraic geometry over groups I. Algebraic sets and ideal theory", *J. Algebra* **219** (1999), 16-79.

[9] A.F. Beardon, *The geometry of discrete groups.* New York, Heidelberg, Berlin: Springer 1983.

[10] T. Bedford, M. Keane and C. Series, *Ergodic theory, symbolic dynamics and hyperbolic spaces.* Oxford: University Press 1991.

[11] J.L. Bell and A.B. Slomson, *Models and ultraproducts.* Amsterdam: North-Holland 1971.

[12] S.K. Berberian, *Measure and integration.* New York: Macmillan 1965.

[13] M. Bestvina and M. Feighn, "Stable actions of groups on real trees", *Invent. Math.* **121** (1995), 287-321.

[14] M. Bestvina and M. Feighn, "Outer limits", preprint.

[15] M. Bestvina and M. Feighn, "The topology at infinity of $Out(F_n)$", *Invent. Math.* **140** (2000), 651-692.

[16] M. Bestvina, M. Feighn and M. Handel, "Laminations, trees, and irreducible automorphisms of free groups", *Geom. Funct. Anal.* **7** (1997), 215-244; erratum ibid. **7**, 1143 (1997).

[17] R. Bieri, W.D. Neumann and R. Strebel, "A geometric invariant of discrete groups", *Invent. Math.* **90** (1987), 451-477.

[18] W.M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry.* San Diego: Academic Press 1986.

[19] N. Bourbaki, *Éléments de mathématique X, algèbre commutative.* Paris: Hermann 1964.

[20] N. Bourbaki, *Éléments de mathématique XIII, intégration.* Paris: Hermann 1965.

[21] B.H. Bowditch, *Treelike structures arising from continua and convergence groups.* Memoirs Amer. Math. Soc. **662**. Providence, RI: Amer. Math. Soc. 1999.

[22] G.E. Bredon, *Topology and geometry.* New York, Berlin, etc.: Springer 1993.

[23] K.S. Brown, "Trees, valuations and the Bieri-Neumann-Strebel invariant", *Invent. Math.* **90** (1987), 479-504.

[24] P. Buneman, "A note on the metric property of trees", *J. Combin. Theory Ser. B* **17**, (1974), 48-50.

[25] J.W. Cannon and G.R. Conner, "The combinatorial structure of the Hawaiian earring group", to appear in *Topology and its Applications.*

[26] J.W.S. Cassels, *Local fields*, London Math. Soc. Student Texts **3**. Cambridge: University Press 1986.

[27] A.J. Casson and S.A. Bleiler, *Automorphisms of surfaces after Nielsen and Thurston*, London Math. Soc. Student Texts **9**. Cambridge: University Press 1988.

[28] C.C. Chang and H.J. Keisler, *Model Theory.* Amsterdam: North-Holland 1973.

[29] G.R. Chapman and D.L. Wilkens, "On length functions, trivializable subgroups and centres of groups", *Michigan Math. J.* **30** (1983), 165-173.

[30] I.M. Chiswell, "Abstract length functions in groups", *Math. Proc. Cambridge Philos. Soc.* **80** (1976), 451-463.

[31] I.M. Chiswell, "Non-trivial group actions on Λ-trees", *Bull. London Math. Soc.* **24** (1992), 277-280.

[32] I.M. Chiswell, "Harrison's theorem for Λ-trees", *Quart. J. Math Oxford* (2) **45** (1994), 1-12.

[33] I.M. Chiswell, "Properly discontinuous actions on Λ-trees", *Proc. Edinburgh Math. Soc.* **37** (1994), 423-444.

[34] I.M. Chiswell, "Introduction to Λ-trees". In: Semigroups, formal languages and groups (ed. J. Fountain), pp. 255-293. Dordrecht, Boston, London: Kluwer 1995.

[35] I.M. Chiswell, "Some examples of groups having no non-trivial action on a Λ-tree", *Mathematika* **42** (1995), 214-219.

[36] I.M. Chiswell, "Length functions and Λ-trees". In: Summer School in group theory in Banff, 1996, CRM Proceedings and Lecture Notes **17**. Providence, RI: Amer. Math. Soc. 1999.

[37] I.M. Chiswell, "Minimal group actions on Λ-trees", *Proc. Roy. Soc. Edinburgh Sect. A* **128** (1998), 23-36.

[38] D.E. Cohen, *Groups of cohomological dimension 1*, Lecture Notes in Mathematics **245**. Berlin, Heidelberg, New York: Springer 1972.

[39] D.E. Cohen, *Combinatorial group theory: a topological approach*, London Math. Soc. Student Texts **14**. Cambridge: University Press 1989.

[40] M.M. Cohen and M. Lustig, "Very small group actions on $\mathbb{R}$-trees and Dehn twist automorphisms", *Topology* **34** (1995), 575-617.

[41] M. Coornaert, T. Delzant and A. Papadopoulos, *Géométrie et théorie des groupes*, Lecture Notes in Mathematics **1441**. Berlin, Heidelberg: Springer 1990.

[42] M. Culler and J. Morgan, "Group actions on $\mathbb{R}$-trees", *Proc. London Math. Soc.(3)* **55** (1987), 571-604.

[43] M. Culler and K. Vogtmann, "Moduli of graphs and automorphisms of free groups", *Invent. Math.* **84** (1986), 91-119.

[44] M. Culler and K. Vogtmann, "The boundary of outer space in rank two". In: Arboreal group theory, (ed. R.C. Alperin), MSRI Publications **19**, pp. 189-230. New York: Springer-Verlag 1991.

[45] W. Dicks and M.J. Dunwoody, *Groups acting on graphs*. Cambridge: University Press 1989.

[46] M.J. Dunwoody, "Groups acting on protrees", *J. London Math. Soc.* **56** (1997), 125-136.

[47] A. Fathi, F. Laudenbach and V. Poénaru, *Travaux de Thurston sur les surfaces, (Astérisque* **66-67**). Paris: Société Mathématique de France 1991.

[48] B. Fine, A.M. Gaglione, A. Myasnikov, G. Rosenberger and D. Spellman, "A classification of fully residually free groups of rank three or less", *J. Algebra* **200** (1998), 571-605.

[49] B. Fine, A.M. Gaglione, G. Rosenberger and D. Spellman, "n-free groups and questions about universally free groups". In: 93 Galway/St. Andrews. Proceedings of the international conference, held in Galway, Ireland, August 1-14, 1993, Vol. 1 (eds. C.M. Campbell et al.), London Math. Soc. Lecture Notes **211**, pp. 191-204. Cambridge: University Press 1995.

[50] D. Gaboriau, "Générateurs indépendants pour les systèmes d'isométries de dimension un", *Ann. Inst. Fourier* **47** (1997), 101-122.

[51] D. Gaboriau and G. Levitt, "The rank of actions on $\mathbb{R}$-trees", *Ann. Sci. École Norm. Sup. (4)* **28** (1995), 549-570.

[52] D. Gaboriau, G. Levitt and M. Lustig, "A dendrological proof of the Scott conjecture for automorphisms of free groups", *Proc. Edinburgh Math. Soc. (2)* **41** (1998), 325-332.

[53] D. Gaboriau, G. Levitt and F. Paulin, "Pseudogroups of isometries of $\mathbb{R}$ and Rips' Theorem on free actions on $\mathbb{R}$-trees", *Israel. J. Math.* **87** (1994), 403-428.

[54] D. Gaboriau, G. Levitt and F. Paulin, "Pseudogroups of isometries of $\mathbb{R}$ and reconstruction of free actions on $\mathbb{R}$-trees", *Ergodic Theory Dynam. Systems* **15** (1995), 633-652.

[55] D. Gaboriau, G. Levitt, A. Jaeger and M. Lustig, "An index for counting fixed points of automorphisms of free groups", *Duke Math. J.* **93** (1998), 425-452.

[56] A.M. Gaglione and D. Spellman, "More model theory of free groups", *Houston J. Math.* **21** (1995), 225-245.

[57] A.M. Gaglione and D. Spellman, "Even more model theory of free groups". In: Infinite groups and group rings. Proceedings of the AMS special session, Tuscaloosa, AL, USA, March 13-14, 1992 (eds. J.M. Corson et al.), pp. 37-40. Singapore: World Scientific 1993.

[58] A.M. Gaglione and D. Spellman, "Does Lyndon's length function imply the universal theory of free groups". In: The mathematical legacy of Wilhelm Magnus: groups, geometry and special functions (Brooklyn, NY, 1992), Contemp. Math. **169**, pp. 277–281. Providence, RI: Amer. Math. Soc. 1994.

[59] A.M. Gaglione and D. Spellman, "Every 'universally free' group is tree free". In: Group theory (Granville, Ohio 1992), pp. 149-154. River Edge, NJ: World Scientific 1993.

[60] A.M. Gaglione and D. Spellman, "Generalisations of some free groups: some questions", *Commun. Algebra* **22** (1994), 3159-3169.

[61] E. Ghys and P. de la Harpe, *Sur les groupes hyperboliques d'après Mikhael Gromov*. Boston: Birkhäuser 1990.

[62] D. Gildenhuys, O. Kharlampovich and A.G. Myasnikov, " CSA groups and separated free constructions", *Bull. Austral. Math. Soc.* **52** (1995), 63-84.

[63] H. Gillet and P.B. Shalen, "Dendrology of groups in low $\mathbb{Q}$-ranks", *J. Differential Geom.* **32** (1990), 605-712.

[64] H. Gillet, P.B. Shalen and R. Skora, "Simplicial approximation and low-rank trees", *Comment. Math. Helv.* **66** (1991), 521-540.

[65] C. Goffman and G. Pedrich, *First course in functional analysis*. Englewood Cliffs, NJ: Prentice Hall 1965.

[66] G. Grätzer, *Universal algebra*. Princeton: van Nostrand 1968.

[67] M. Gromov, "Hyperbolic groups". In: Essays in group theory, (ed. S.M. Gersten), MSRI Publications **8**, pp. 75-263. New York: Springer-Verlag 1987.

[68] V. Guillemin and A. Pollack, *Differential topology*. Englewood Cliffs: Prentice-Hall 1974.

[69] V. Guirardel, "Approximations of stable actions on ℝ-trees", *Comment. Math. Helv.* **73** (1998), 89-121.

[70] N. Gupta and S. Sidki, "On the Burnside problem for periodic groups", *Math. Z.* **182** (1983), 385-388.

[71] P.R. Halmos, *Measure theory.* New York: van Nostrand 1950.

[72] N. Harrison, "Real length functions in groups", *Trans. Amer. Math. Soc.* **174** (1972), 77-106.

[73] R. Hartshorne, *Algebraic geometry.* New York: Springer 1977.

[74] A.E. Hatcher, "Measured lamination spaces for surfaces, from the topological viewpoint", *Topology Appl.* **30** (1988), 63-88.

[75] A.H.M. Hoare, A. Karrass and D. Solitar, "Subgroups of finite index in Fuchsian groups", *Math. Z.* **120** (1971), 289-298.

[76] A.H.M. Hoare, A. Karrass and D. Solitar, "Subgroups of infinite index in Fuchsian groups", *Math. Z.* **125** (1972), 59-69.

[77] A.H.M. Hoare and D.L. Wilkens, "On groups with unbounded non-archimedean elements". In: Groups-St. Andrews 1981 (eds. C.M. Campbell and E. Robertson), London Math. Soc. Lecture Notes **71**, pp. 228-236. Cambridge: University Press 1982.

[78] C. Holton and L.Q. Zamboni, "Exotic actions on trees", *Bull. London Math. Soc.* **29** (1972), 309-313.

[79] W.Imrich, "On metric properties of tree-like spaces". In: Beiträge zur Graphentheorie und deren Anwendungen, (ed. Sektion MARÖK der Technischen Hochschule Ilmenau), pp. 129-156. Oberhof, DDR 1977.

[80] S. Jackson and L.Q. Zamboni, "A note on a theorem of Chiswell", *Proc. Amer. Math. Soc.* **123** (1995), 2629-2631.

[81] N. Jacobson, *Basic algebra II.* San Francisco: W.H. Freeman 1980.

[82] G.A. Jones and D. Singerman, *Complex functions.* Cambridge: University Press 1987.

[83] S. Katok, *Fuchsian groups.* Chicago: University Press 1992.

[84] O.H. Kegel and B.A.F. Wehrfritz, *Locally finite groups.* Amsterdam, London: North-Holland 1973.

[85] Z. Khan and D.L. Wilkens, "Inherited group actions on ℝ-trees", *Mathematika* **42** (1995), 206-213.

[86] O. Kharlampovich and A. Myasnikov, "Irreducible affine varieties over a free group. I: Irreducibility of quadratic equations and Nullstellensatz", *J. Algebra* **200** (1998), 472-516.

[87] O. Kharlampovich and A. Myasnikov, "Irreducible affine varieties over a free group. II: Systems in triangular quasi-quadratic form and description of residually free groups", *J. Algebra* **200** (1998), 517-570.

[88] S. Krstic and K. Vogtmann, "Equivariant outer space and automorphisms of free-by-finite groups", *Comment. Math. Helv.* **68** (1993), 216-262.

[89] J. Lehner, *A short course on automorphic functions*. New York: Holt, Rinehart and Winston 1966.

[90] G. Levitt, "Constructing free actions on $\mathbb{R}$-trees", *Duke Math. J.* **69** (1993), 615-633.

[91] G. Levitt and F. Paulin,"Geometric group actions on trees", *Amer. J. Math.* **119** (1997), 83-102.

[92] R.C. Lyndon, "Length functions in groups", *Math. Scand.* **12** (1963), 209-234.

[93] R.C. Lyndon, "Groups with parametric exponents", *Trans. Amer. Math. Soc.* **96** (1960), 518-533.

[94] R.C. Lyndon, "The equation $a^2b^2 = c^2$ in free groups", *Michigan Math. J.* **6** (1959), 155-164.

[95] R.C. Lyndon and P.E. Schupp, *Combinatorial group theory*, Ergebnisse der Math. **89**. Berlin, Heidelberg, New York: Springer 1977.

[96] W. Magnus, *Noneuclidean tesselations and their groups*. New York, London: Academic Press 1974.

[97] B. Maskit, "On Poincaré's theorem on fundamental regions", *Adv. Math.* **7** (1971), 219-230.

[98] W.S. Massey, *Algebraic topology: an introduction*. New York: Harcourt Brace and World 1967.

[99] C.R.F. Maunder, *Algebraic topology*. Cambridge: University Press 1980.

[100] E.E. Moise, *Geometric topology in dimensions* 2 *and* 3. New York, Heidelberg, Berlin: Springer 1977.

[101] J.D. Monk, *Introduction to set theory*. New York: McGraw-Hill 1969.

[102] J.W. Morgan, "Group actions on trees and the compactification of the space of classes of $\mathrm{SO}(n,1)$-representations", *Topology* **25** (1986), 1-33.

[103] J.W. Morgan, "Λ-trees and their applications", *Bull. Amer. Math. Soc.* **26** (1992), 87-112.

[104] J.W. Morgan and I.Morrison,"A van Kampen Theorem for weak joins", *Proc. London Math. Soc.* **53** (1986), 562-576.

[105] J.W. Morgan and J.-P. Otal,"Relative growth rates of closed geodesics on a surface under varying hyperbolic structures", *Comment. Math. Helv.* **68** (1993), 171-208.

[106] J.W. Morgan and P.B. Shalen, "Valuations, trees and degenerations of hyperbolic structures: I", *Annals of Math.* (2) **122** (1985), 401-476.

[107] J.W. Morgan and P.B. Shalen, "Degenerations of hyperbolic structures: II: measured laminations in 3-manifolds", *Annals of Math.* (2) **127** (1988), 403-456.

[108] J.W. Morgan and P.B. Shalen, "Degenerations of hyperbolic structures: III: actions of 3-manifold groups on trees and Thurston's compactness theorem", *Annals of Math.* (2) **127** (1988), 457-519.

[109] J.W. Morgan and P.B. Shalen, "Free actions of surface groups on $\mathbb{R}$-trees", *Topology* **30** (1991), 143-154.

[110] A.G. Myasnikov and O. Kharlampovich, "Tarski's problem about the elementary theory of free groups has a positive solution", *Electron. Res. Announc. Amer. Math. Soc.* 4 (1998), 101-108.

[111] A.G. Myasnikov and O. Kharlampovich, "Elementary theory of free nonabelian groups", preprint.

[112] A.G. Myasnikov and V.N. Remeslennikov, "Exponential groups. II: extensions of centralizers and tensor completion of CSA-groups", *Internat. J. Algebra Comput.* **6** (1996), 687-711.

[113] H. Neumann, *Varieties of groups*. Berlin, Heidelberg, New York: Springer 1967.

[114] P.J. Nicholls, *The ergodic theory of discrete groups*, London Math. Soc. Lecture Notes **14**. Cambridge: University Press 1989.

[115] W. Parry, "Axioms for translation length functions". In: Arboreal group theory, (ed. R.C. Alperin), MSRI Publications **19**, pp. 295-330. New York: Springer-Verlag 1991.

[116] D.S. Passman, *The algebraic structure of group rings*. New York, London etc.: John Wiley 1977.

[117] F. Paulin, "Outer automorphisms of hyperbolic groups and small actions on $\mathbb{R}$-trees". In: Arboreal group theory, (ed. R.C. Alperin), MSRI Publications **19**, pp. 331-343. New York: Springer-Verlag 1991.

[118] F. Paulin, "Actions de groupes sur les arbres", *Astérisque* **241** (1997), 97-137.

[119] R.C. Penner and J.L. Harer, *Combinatorics of train tracks*. Annals of Math. Studies **125**. Princeton: University Press 1992.

[120] P.H. Pfander, "Finitely generated subgroups of the free $\mathbb{Z}[t]$ group on two generators". In: Model theory of groups and automorphism groups (ed. D.M. Evans), London Math. Soc. Lecture Notes **244**, pp. 166-187. Cambridge: University Press 1997.

[121] D. Promislow, "Equivalence classes of length functions on groups", *Proc. London Math. Soc.* (3) **51** (1985), 449-477.

[122] V.N. Remeslennikov, "$\exists$-free groups", *Siberian Math. J.* **30** (1989), 998-1001; translation (1990) from *Sibirsk. Mat. Zh.* **30** (1989), 193-197.

[123] V.N. Remeslennikov, "$\exists$-free groups as groups with length function", *Ukrainian Math. J.* 44 (1992), 733-738; translation from *Ukrain. Mat. Zh.* 44 (1992), 813-818.

[124] J. Rhodes, "Monoids acting on trees: elliptic and wreath products and the holonomy theorem for arbitrary monoids with applications to infinite groups", *Internat. J. Algebra Comput.* **1** (1991), 253-279.

[125] F.S. Rimlinger, "$\mathbb{R}$-trees and normalisation of pseudogroups", *Experiment. Math.* **1** (1992), 95-114.

[126] E. Rips and Z. Sela, "Structure and rigidity in hyperbolic groups. I" *Geom. Funct. Anal.* 4 (1994), 337-371.

[127] J.-P. Serre, *Trees.* New York: Springer 1980.

[128] P.B. Shalen, "Dendrology of groups: an introduction". In: Essays in group theory, (ed. S.M. Gersten), MSRI Publications **8**, pp. 265-319. New York: Springer-Verlag 1987.

[129] P.B. Shalen, "Dendrology and its applications". In: Group theory from a geometrical viewpoint (eds. E. Ghys, A. Haefliger and A. Verjovsky). Singapore, New Jersey, London, Hong Kong: World Scientific 1991.

[130] H. Short (ed.), "Notes on word hyperbolic groups". In: Group theory from a geometrical viewpoint (eds. E. Ghys, A. Haefliger and A. Verjovsky). Singapore, New Jersey, London, Hong Kong: World Scientific 1991.

[131] R. Skora, "Splittings of surfaces", *Bull. Amer. Math. Soc.* **23** (1990), 85-90.

[132] R. Skora, "Splittings of surfaces", *J. Amer. Math. Soc.* **9** (1996), 605-616.

[133] R. Skora, "Deformations of length functions in groups", preprint.

[134] E. Spanier, *Algebraic topology.* New York, etc.: McGraw-Hill 1966.

[135] M. Spivak, *A comprehensive introduction to differential geometry*, Volume I. Houston: Publish or Perish 1971.

[136] J.R. Stallings, *Group Theory and Three-dimensional manifolds.* New Haven and London: Yale University Press 1971.

[137] J.R. Stallings, "Adian groups and pregroups". In: Essays in group theory, (ed. S.M. Gersten), MSRI Publications **8**, pp. 321-342. New York: Springer-Verlag 1987.

[138] M. Steiner, "Gluing data and group actions on Λ-trees", preprint.

[139] W.P. Thurston, *Geometry and topology of 3-manifolds.* Unpublished. At the time of writing, available from the Web site of the Mathematical Sciences Research Institute, Berkeley, Ca., URL: `http://www.msri.org/publications/books/gt3m/index.html`

[140] W.P. Thurston, "Three dimensional manifolds, Kleinian groups and hyperbolic geometry", *Bull. Amer. Math. Soc.* **6** (1982), 357-379.

[141] W.P. Thurston (ed. S. Levy), *Three-dimensional geometry and topology.* Princeton: University Press 1997.

[142] J. Tits, "Sur le groupe des automorphismes d'un arbre". In: Essays on topology and related topics, pp. 188-211. New York: Springer 1970.

[143] J. Tits, "A 'theorem of Lie-Kolchin' for trees". In: Contributions to algebra: a collection of papers dedicated to Ellis Kolchin. New York: Academic Press 1977.

[144] M. Urbański and L.Q. Zamboni, "On free actions on Λ-trees", *Math. Proc. Cambridge. Philos. Soc.* **113** (1993), 535-542.

[145] L. van den Dries and A. Wilkie, "Gromov's Theorem on groups of polynomial growth and elementary logic", *J. Algebra* 8**9** (1984), 349-374.

[146] T. White, "Fixed points of finite groups of free group automorphisms", *Proc. Amer. Math. Soc.* **118** (1993), 681-688.

[147] D.L. Wilkens, "Length functions and normal subgroups", *J. London Math. Soc.* (2) **22** (1980), 439-448.

[148] D.L. Wilkens, "Group actions on trees and length functions", *Michigan Math. J.* **35** (1988), 141-150.

[149] D.L. Wilkens, "Bounded group actions on trees and hyperbolic and Lyndon length functions", *Michigan Math. J.* **36** (1989), 303-308.

[150] M.I. Zaitseva, "On the set of orderings of an abelian group", (Russian) *Uspekhi Mat. Nauk* **8(53)** (1953), 135-137.

[151] L.Q. Zamboni, "Tree fibrations", *Houston J. Math.* **21** (1995), 53-74.

[152] O. Zariski and P. Samuel, *Commutative algebra*, Vol. II. Princeton, New York: van Nostrand 1960.

[153] A. Zastrow, "Construction of an infinitely generated group that is not a free product of surface groups and abelian groups, but which acts freely on an $\mathbb{R}$-tree", *Proc. Royal Soc. Edinburgh(A)* **128** (1998), 433-445.

[154] H. Zieschang, E. Vogt and H.-D. Coldewey, *Surfaces and planar discontinuous groups*, Lecture Notes in Mathematics **835**. Berlin, Heidelberg, New York: Springer 1980.

Index of Notation

Index

Introduction to Λ-TREES

The theory of Λ-trees has its origin in the work of Lyndon on length functions in groups. The first definition of an ℝ-tree was given by Tits in 1977. The importance of Λ-trees was established by Morgan and Shalen, who showed how to compactify a generalisation of Teichmüller space for a finitely generated group using ℝ-trees. In that work they were led to define the idea of a Λ-tree, where Λ is an arbitrary ordered abelian group. Since then there has been much progress in understanding the structure of groups acting on ℝ-trees, notably Rips' theorem on free actions. There has also been some progress for certain other ordered abelian groups Λ, including some interesting connections with model theory.

Introduction to Λ-Trees will prove to be useful for mathematicians and research students in algebra and topology.

ISBN 981-02-4386-3

9 789810 243869

www.worldscientific.com

4495 hc