
S E
M M
E S

EMS Textbooks in Mathematics

EMS Textbooks in Mathematics is a series of books aimed at students or professional mathematicians seeking an introduction into a particular field. The individual volumes are intended not only to provide relevant techniques, results, and applications, but also to afford insight into the motivations and ideas behind the theory. Suitably designed exercises help to master the subject and prepare the reader for the study of more advanced and specialized literature.

Jørn Justesen and Tom Høholdt, *A Course In Error-Correcting Codes*
Markus Stroppel, *Locally Compact Groups*
Peter Kunkel and Volker Mehrmann, *Differential-Algebraic Equations*
Dorothee D. Haroske and Hans Triebel, *Distributions, Sobolev Spaces, Elliptic Equations*
Thomas Timmermann, *An Invitation to Quantum Groups and Duality*
Oleg Bogopolski, *Introduction to Group Theory*
Marek Jarnicki und Peter Pflug, *First Steps in Several Complex Variables: Reinhardt Domains*
Tammo tom Dieck, *Algebraic Topology*
Mauro C. Beltrametti et al., *Lectures on Curves, Surfaces and Projective Varieties*

Wolfgang Woess

Denumerable Markov Chains

Generating Functions
Boundary Theory
Random Walks on Trees

European Mathematical Society

Author:

Wolfgang Woess
Institut für Mathematische Strukturtheorie
Technische Universität Graz
Steyrergasse 30
8010 Graz
Austria

2000 Mathematical Subject Classification (primary; secondary): 60-01; 60J10, 60J50, 60J80, 60G50, 05C05, 94C05, 15A48

Key words: Markov chain, discrete time, denumerable state space, recurrence, transience, reversible Markov chain, electric network, birth-and-death chains, Galton–Watson process, branching Markov chain, harmonic functions, Martin compactification, Poisson boundary, random walks on trees

ISBN 978-3-03719-071-5

The Swiss National Library lists this publication in The Swiss Book, the Swiss national bibliography, and the detailed bibliographic data are available on the Internet at http://www.helveticat.ch.

Contact address:

European Mathematical Society Publishing House
Seminar for Applied Mathematics
ETH-Zentrum FLI C4
CH-8092 Zürich
Switzerland

Phone: +41 (0)44 632 34 36
Email: info@ems-ph.org
Homepage: www.ems-ph.org

Cover photograph by Susanne Woess-Gallasch, Hakone, 2007

Typeset using the author's TEX files: I. Zimmermann, Freiburg
Printed in Germany

9 8 7 6 5 4 3 2 1

Preface

This book is about time-homogeneous Markov chains that evolve with discrete time steps on a countable state space. This theory was born more than 100 years ago, and its beauty stems from the simplicity of the basic concept of these random processes: "given the present, the future does not depend on the past". While of course a theory that builds upon this axiom cannot explain all the weird problems of life in our complicated world, it is coupled with an ample range of applications as well as the development of a widely ramified and fascinating mathematical theory. Markov chains provide one of the most basic models of stochastic processes that can be understood at a very elementary level, while at the same time there is an amazing amount of ongoing, new and deep research work on that subject.

The present textbook is based on my Italian lecture notes *Catene di Markov e teoria del potenziale nel discreto* from 1996 [W1]. I thank Unione Matematica Italiana for authorizing me to publish such a translation. However, this is not just a one-to-one translation. My view on the subject has widened, part of the old material has been rearranged or completely modified, and a considerable amount of material has been added. Only Chapters 1, 2, 6, 7 and 8 and a smaller portion of Chapter 3 follow closely the original, so that the material has almost doubled.

As one will see from summary (page ix) and table of contents, this is not about applied mathematics but rather tries to develop the "pure" mathematical theory, starting at a very introductory level and then displaying several of the many fascinating features of that theory.

Prerequisites are, besides the standard first year linear algebra and calculus (including power series), an understanding of and – most important – interest in probability theory, possibly including measure theory, even though a good part of the material can be digested even if measure theory is avoided. A small amount of complex function theory, in connection with the study of generating functions, is needed a few times, but only at a very light level: it is useful to know what a singularity is and that for a power series with non-negative coefficients the radius of convergence is a singularity. At some points, some elementary combinatorics is involved. For example, it will be good to know how one solves a linear recursion with constant coefficients. Besides this, very basic Hilbert space theory is needed in §C of Chapter 4, and basic topology is needed when dealing with the Martin boundary in Chapter 7. Here it is, in principle, enough to understand the topology of metric spaces.

One cannot claim that every chapter is on the same level. Some, specifically at the beginning, are more elementary, but the road is mostly uphill. I myself have used different parts of the material that is included here in courses of different levels.

The writing of the Italian lecture notes, seen *a posteriori,* was sort of a "warm up" before my monograph *Random walks on infinite graphs and groups* [W2]. Markov chain basics are treated in a rather condensed way there, and the understanding of a good part of what is expanded here in detail is what I would hope a reader could bring along for digesting that monograph.

I thank Donald I. Cartwright, Rudolf Grübel, Vadim A. Kaimanovich, Adam Kinnison, Steve Lalley, Peter Mörters, Sebastian Müller, Marc Peigné, Ecaterina Sava and Florian Sobieczky very warmly for proofreading, useful hints and some additional material.

Graz, July 2009 Wolfgang Woess

Contents

Introduction

Summary

Chapter 1 starts with elementary examples (§A), the first being the one that is depicted on the cover of the book of KEMENY and SNELL [K-S]. This is followed by an informal description ("What is a Markov chain?", "The graph of a Markov chain") and then (§B) the axiomatic definition as well as the construction of the trajectory space as the standard model for a probability space on which a Markov chain can be defined. This quite immediate first impact of measure theory might be skipped at first reading or when teaching at an elementary level. After that we are back to basic transition probabilities and passage times (§C). In the last section (§D), the first encounter with generating functions takes place, and their basic properties are derived. There is also a short explanation of transition probabilities and the associated generating functions in purely combinatorial terms of paths and their weights.

Chapter 2 contains basic material regarding irreducible classes (§A) and periodicity (§B), interwoven with examples. It ends with a brief section (§C) on the spectral radius, which is the inverse of the radius of convergence of the Green function (the generating function of n-step transition probabilities).

Chapter 3 deals with recurrence vs. transience (§A & §B) and the fundamental convergence theorem for positive recurrent chains (§C & §E). In the study of positive recurrence and existence and uniqueness of stationary probability distributions (§B), a mild use of generating functions and de l'Hospital's rule as the most "difficult" tools turn out to be quite efficient. The convergence theorem for positive recurrent, aperiodic chains appears so important to me that I give two different proofs. The first (§C) applies primarily (but not only) to finite Markov chains and uses Doeblin's condition and the associated contraction coefficient. This is pure matrix analysis which leads to crucial probabilistic interpretations. In this context, one can understand the convergence theorem for finite Markov chains as a special case of the famous Perron–Frobenius theorem for non-negative matrices. Here (§D), I make an additional detour into matrix analysis by reversing this viewpoint: the convergence theorem is considered as a main first step towards the proof of the Perron–Frobenius theorem, which is then deduced. I do not claim that this proof is overall shorter than the typical one that one finds in books such as the one of SENETA [Se]; the main point is that I want to work out how one can proceed by extending the lines of thought of the preceding section. What follows (§E) is another, elegant and much more probabilistic proof of the convergence theorem for general positive recurrent, aperiodic Markov chains. It uses the coupling method,

see LINDVALL [Li]. In the original Italian text, I had instead presented the proof of the convergence theorem that is due to ERDÖS, FELLER and POLLARD [20], a breathtaking piece of "elementary" analysis of sequences; see e.g, [Se, §5.2]. It is certainly not obsolete, but I do not think I should have included a third proof here, too. The second important convergence theorem, namely, the ergodic theorem for Markov chains, is featured in §F. The chapter ends with a short section (§G) about ρ-recurrence.

Chapter 4. The chapter (most of whose material is not contained in [W1]) starts with the network interpretation of a reversible Markov chain (§A). Then (§B) the interplay between the spectrum of the transition matrix and the speed of convergence to equilibrium (= the stationary probability) for finite reversible chains is studied, with some specific emphasis on the special case of symmetric random walks on finite groups. This is followed by a very small introductory glimpse (§C) at the very impressive work on geometric eigenvalue bounds that has been promoted in the last two decades via the work of DIACONIS, SALOFF-COSTE and others; see [SC] and the references therein, in particular, the basic paper by DIACONIS and STROOCK [15] on which the material here is based. Then I consider recurrence and transience criteria for infinite reversible chains, featuring in particular the *flow criterion* (§D). Some very basic knowledge of Hilbert spaces is required here. While being close to [W2, §2.B], the presentation is slightly different and "slower". The last section (§E) is about recurrence and transience of random walks on integer lattices. Those Markov chains are not always reversible, but I figured this was the best place to include that material, since it starts by applying the flow criterion to symmetric random walks. It should be clear that this is just a very small set of examples from the huge world of random walks on lattices, where the classical source is SPITZER's famous book [Sp]; see also (for example) RÉVÉSZ [Ré], LAWLER [La] and FAYOLLE, MALYSHEV and MEN'SHIKOV [F-M-M], as well as of course the basic material in FELLER's books [F1], [F2].

Chapter 5 first deals with two specific classes of examples, starting with birth-and-death chains on the non-negative integers or a finite interval of integers (§A). The Markov chains are nearest neighbour random walks on the underlying graph, which is a half-line or line segment. Amongst other things, the link with analytic continued fractions is explained. Then (§B) the classical analysis of the Galton–Watson process is presented. This serves also as a prelude of the next section (§C), which is devoted to an outline of some basic features of branching Markov chains (BMCs, §C). The latter combine Markov chains with the evolution of a "population" according to a Galton–Watson process. BMCs themselves go beyond the theme of this book, Markov chains. One of their nice properties is that certain probabilistic quantities associated with BMC are expressed in terms of the generating functions of the underlying Markov chain. In particular, ρ-recurrence of the chain has such an interpretation via criticality of an embedded Galton–Watson process. In view of my

insisting on the utility of generating functions, this is a very appealing propaganda instrument regarding their probabilistic nature.

In the sections on the Galton–Watson process and BMC, I pay some extra attention to the rigorous construction of a probability space on which the processes can be defined completely and with all their features; see my remarks about a certain *nonchalance* regarding the existence of the "probabilistic heaven" further below which appear to be particularly appropriate here. (I do not claim that the proposed model probability spaces are the only good ones.)

Of this material, only the part of §A dealing with continued fractions was already present in [W1].

Chapter 6 displays basic notions, terminology and results of potential theory in the discrete context of transient Markov chains. The discrete Laplacian is $P - I$, where P is the transition matrix and I the identity matrix. The starting point (§A) is the finite case, where we declare a part of the state space to be the *boundary* and its complement to be the *interior*. We look for functions that have preassigned value on the boundary and are harmonic in the interior. This discrete *Dirichlet problem* is solved in probabilistic terms.

We then move on to the infinite, transient case and (in §B) consider basic features of harmonic and superharmonic functions and their duals in terms of measures on the state space. Here, functions are thought of as column vectors on which the transition matrix acts from the left, while measures are row vectors on which the matrix acts from the right. In particular, transience is linked with the existence of non-constant positive superharmonic functions. Then (§C) induced Markov chains and their interplay with superharmonic functions and excessive measures are displayed, after which (§D) classical results such as the Riesz decomposition theorem and the approximation theorem for positive superharmonic functions are proved. The chapter ends (§E) with an explanation of "balayage" in terms of first entrance and last exit probabilities, concluding with the domination principle for superharmonic functions.

Chapter 7 is an attempt to give a careful exposition of Martin boundary theory for transient Markov chains. I do not aim at the highest level of sophistication but at the broadest level of comprehensibility. As a mild but natural restriction, only irreducible chains are considered (i.e., all states communicate), but substochastic transition matrices are admitted since this is needed anyway in some of the proofs. The starting point (§A) is the definition and first study of the extreme elements in the convex cone of positive superharmonic functions, in particular, the minimal harmonic functions. The construction/definition of the Martin boundary (§B) is preceded by a preamble on compactifications in general. This section concludes with the statement of one of the two main theorems of that theory, namely convergence to the boundary. Before the proof, martingale theory is needed (§C), and we examine the relation of supermartingales with superharmonic functions and, more subtle and

important here, with excessive measures. Then (§D) we derive the Poisson–Martin integral representation of positive harmonic functions and show that it is unique over the minimal boundary. Finally (§E) we study the integral representation of bounded harmonic functions (the Poisson boundary), its interpretation via terminal random variables, and the probabilistic Fatou convergence theorem. At the end, the alternative approach to the Poisson–Martin integral representation via the approximation theorem is outlined.

Chapter 8 is very short and explains the rather algebraic procedure of finding all minimal harmonic functions for random walks on integer grids.

Chapter 9, on the contrary, is the longest one and dedicated to nearest neighbour random walks on trees (mostly infinite). Here we can harvest in a concrete class of examples from the seed of methods and results of the preceding chapters. First (§A), the fundamental equations for first passage time generating functions on trees are exhibited, and some basic methods for finite trees are outlined. Then we turn to infinite trees and their boundary. The geometric boundary is described via the end compactification (§B), convergence to the boundary of transient random walks is proved directly, and the Martin boundary is shown to coincide with the space of ends (§C). This is also the minimal boundary, and the limit distribution on the boundary is computed. The structural simplicity of trees allows us to provide also an integral representation of *all* harmonic functions, not only positive ones (§D). Next (§E) we examine in detail the Dirichlet problem at infinity and the regular boundary points, as well as a simple variant of the radial Fatou convergence theorem. A good part of these first sections owes much to the seminal long paper by Cartier [Ca], but one of the innovations is that many results do not require local finiteness of the tree. There is a short *intermezzo* (§F) about how a transient random walk on a tree approaches its limiting boundary point. After that, we go back to transience/recurrence and consider a few criteria that are specific to trees, with a special eye on trees with *finitely many cone types* (§G). Finally (§H), we study in some detail two intertwined subjects: rate of escape (i.e., variants of the law of large numbers for the distance to the starting point) and spectral radius. Throughout the chapter, explicit computations are carried out for various examples via different methods.

Examples are present throughout all chapters.

Exercises are not accumulated at the end of each section or chapter but "built in" the text, of which they are considered an integral part. Quite often they are used in the subsequent text and proofs. The imaginary ideal reader is one who solves those exercises in real time while reading.

Solutions of all exercises are given after the last chapter.

The bibliography is subdivided into two parts, the first containing textbooks and other general references, which are recognizable by citations in letters. These are also intended for further reading. The second part consists of research-specific

references, cited by numbers, and I do not pretend that these are complete. I tried to have them reasonably complete as far as material is concerned that is relatively recent, but going back in time, I rely more on the belief that what I'm using has already reached a confirmed status of public knowledge.

Raison d'être

Why another book about Markov chains? As a matter of fact, there is a great number and variety of textbooks on Markov chains on the market, and the older ones have by no means lost their validity just because so many new ones have appeared in the last decade. So rather than just praising in detail my own *opus*, let me display an incomplete subset of the mentioned variety.

For me, the all-time classic is CHUNG's *Markov chains with stationary transition probabilities* [Ch], along with KEMENY and SNELL, *Finite Markov chains* [K-S], whose first editions are both from 1960. My own learning of the subject, years ago, owes most to *Denumerable Markov chains* by KEMENY, SNELL and KNAPP [K-S-K], for which the title of this book is thought as an expression of reverence (without claiming to reach a comparable amplitude). Besides this, I have a very high esteem of SENETA's *Non-negative matrices and Markov chains* [Se] (first edition from 1973), where of course a reader who is looking for stochastic adventures will need previous motivation to appreciate the matrix theory view.

Among the older books, one definitely should not forget FREEDMAN [Fr]; the one of ISAACSON and MADSEN [I-M] has been very useful for preparing some of my lectures (in particular on non time-homogeneous chains, which are not featured here), and REVUZ' [Re] profound French style treatment is an important source permanently present on my shelf.

Coming back to the last 10–12 years, my personal favourites are the monograph by BRÉMAUD [Br] which displays a very broad range of topics with a permanent eye on applications in all areas (this is the book that I suggest to young mathematicians who want to use Markov chains in their future work), and in particular the very nicely written textbook by NORRIS [No], which provides a delightful itinerary into the world of stochastics for a probabilist-to-be. Quite recently, D. STROOCK enriched the selection of introductory texts on Markov processes by [St2], written in his masterly style.

Other recent, maybe more focused texts are due to BEHRENDS [Be] and HÄGGSTRÖM [Hä], as well as the St. Flour lecture notes by SALOFF-COSTE [SC]. All this is complemented by the high level exercise selection of BALDI, MAZLIAK and PRIOURET [B-M-P].

In Italy, my lecture notes (the first in Italian dedicated exclusively to this topic) were followed by the densely written paperback by PINTACUDA [Pi]. In this short

review, I have omitted most of the monographs about Markov chains on non-discrete state spaces, such as NUMMELIN [Nu] or HERNÁNDEZ-LERMA and LASSERRE [H-L] (to name just two besides [Re]) as well as continuous-time processes.

So in view of all this, this text needs indeed some additional reason of being. This lies in the three subtitle topics *generating functions, boundary theory, random walks on trees,* which are featured with some extra emphasis among all the material.

Generating functions. Some decades ago, as an apprentice of mathematics, I learnt from my PhD advisor Peter Gerl at Salzburg how useful it was to use generating functions for analyzing random walks. Already a small amount of basic knowledge about power series with non-negative coefficients, as it is taught in first or second year calculus, can be used efficiently in the basic analysis of Markov chains, such as irreducible classes, transience, null and positive recurrence, existence and uniqueness of stationary measures, and so on. Beyond that, more subtle methods from complex analysis can be used to derive refined asymptotics of transition probabilities and other limit theorems. (See [53] for a partial overview.) However, in most texts on Markov chains, generating functions play a marginal role or no role at all. I have the impression that quite a few of nowadays' probabilists consider this too analytically-combinatorially flavoured. As a matter of fact, the three Italian reviewers of [W1] criticised the use of generating functions as being too heavy to be introduced at such an early stage in those lecture notes. With all my students throughout different courses on Markov chains and random walks, I never noticed any such difficulties.

With humble admiration, I sympathise very much with the vibrant preface of D. STROOCK's masterpiece *Probability theory: an analytic view* [St1]: (quote) "I have never been able to develop sufficient sensitivity to the distinction between a *proof* and a *probabilistic proof*". So, confirming hereby that I'm not a (quote) "dyed-in-the-wool probabilist", I'm stubborn enough to insist that the systematic use of generating functions at an early stage of developing Markov chain basics is very useful. This is one of the specific *raisons d'être* of this book. In any case, their use here is very very mild. My original intention was to include a whole chapter on the application of tools from complex analysis to generating functions associated with Markov chains, but as the material grew under my hands, this had to be abandoned in order to limit the size of the book. The masters of these methods come from analytic combinatorics; see the very comprehensive monograph by FLAJOLET and SEDGEWICK [F-S].

Boundary theory and elements of discrete potential theory. These topics are elaborated at a high level of sophistication by KEMENY, SNELL and KNAPP [K-S-K] and REVUZ [Re], besides the literature from the 1960s and '70s in the spirit of abstract potential theory. While [K-S-K] gives a very complete account, it is not at all easy reading. My aim here is to give an introduction to the language and basics of the potential theory of (transient) denumerable Markov chains, and, in

particular, a rather complete picture of the associated topological boundary theory that may be accessible for good students as well as interested colleagues coming from other fields of mathematics. As a matter of fact, even advanced non-experts have been tending to mix up the concepts of Poisson and Martin boundaries as well as the Dirichlet problem at infinity (whose solution with respect to some geometric boundary does *not* imply that one has identified the Martin boundary, as one finds stated). In the exposition of this material, my most important source was a rather old one, which still is, according to my opinion, the best readable presentation of Martin boundary theory of Markov chains: the expository article by DYNKIN [Dy] from 1969.

Potential and boundary theory is a point of encounter between probability and analysis. While classical potential theory was already well established when its intrinsic connection with Brownian motion was revealed, the probabilistic theory of denumerable Markov chains and the associated potential theory were developed hand in hand by the same protagonists: to their mutual benefit, the two sides were never really separated. This is worth mentioning, because there are not only probabilists but also analysts who distinguish between a *proof* and a *probabilistic proof* – in a different spirit, however, which may suggest that if an analytic result (such as the solution of the Dirichlet problem at infinity) is deduced by probabilistic reasoning, then that result is true only *almost surely* before an *analytic proof* has been found.

What is not included here is the potential and boundary theory of recurrent chains. The former plays a prominent role mainly in relation with random walks on two-dimensional grids, and SPITZER's classic [Sp] is still a prominent source on this; I also like to look up some of those things in LAWLER [La]. Also, not much is included here about the ℓ^2-potential theory associated with reversible Markov chains (networks); the reader can consult the delightful little book by DOYLE and SNELL [D-S] and the lecture notes volume by SOARDI [So].

Nearest neighbour random walk on trees is the third item in the subtitle. Trees provide an excellent playground for working out the potential and boundary theory associated with Markov chains. Although the relation with the classical theory is not touched here, the analogy with potential theory and Brownian motion on the open unit disk, or rather, on the hyperbolic plane, is striking and obvious. The combinatorial structure of trees is simple enough to allow a presentation of a selection of methods and results which are well accessible for a sufficiently ambitious beginner. The resulting, rather long final chapter takes up and elaborates upon various topics from the preceding chapters. It can serve as a link with [W2], where not as much space has been dedicated to this specific theme, and, in particular, the basics are not developed as broadly as here.

In order to avoid the impact of additional structure-theoretic subtleties, I insist on dealing only with nearest neighbour random walks. Also, this chapter is certainly

far from being comprehensive. Nevertheless, I think that a good part of this material appears here in book form for the first time. There are also a few new results and/or proofs.

Additional material can be found in [W2], and also in the ever forthcoming, quite differently flavoured wonderful book by LYONS with PERES [L-P].

At last, I want to say a few words about

the role of measure theory. If one wants to avoid measure theory, and in particular the extension machinery in the construction of the trajectory space of a Markov chain, then one can carry out a good amount of the theory by considering the Markov chain in a finite time interval $\{0, \dots, n\}$. The trajectory space is then countable and the underlying probability measure is atomic. For deriving limit theorems, one may first consider that time interval and then let $n \to \infty$. In this spirit, one can use a rather large part of the initial material in this book for teaching Markov chains at an elementary level, and I have done so on various occasions.

However, it is my opinion that it has been a great achievement that probability has been put on the solid theoretical fundament of measure theory, and that students of mathematics (as well as physics) should be exposed to that theoretical fundament, as opposed to fake attempts to make their curricula more "soft" or "applied" by giving up an important part of the mathematical edifice.

Furthermore, advanced probabilists are quite often – and with very good reason – somewhat *nonchalant* when referring to the spaces on which their random processes are defined. The attitude often becomes one where we are confident that there always is some big probability space somewhere up in the clouds, a kind of *probabilistic heaven,* on which all the random variables and processes that we are working with are defined and comply with all the properties that we postulate, but we do not always care to see what makes it sure that this probabilistic heaven is solid. Apart from the suspicion that this attitude may be one of the causes of the vague distrust of some analysts to which I alluded above, this is fine with me. But I believe this should not be a guideline of the education of master or PhD students; they should first see how to set up the edifice rigorously before passing to nonchalance that is based on firm knowledge.

What is not contained about Markov chains is of course much more than what *is* contained in this book. I could have easily doubled its size, thereby also changing its scope and intentions. I already mentioned recurrent potential and boundary theory, there is a lot more that one could have said about recurrence and transience, one could have included more details about geometric eigenvalue bounds, the Galton–Watson process, and so on. I have not included any hint at continuous-time Markov processes, and there is no random environment, in spite of the fact that this is currently very much *en vogue* and may have a much more *probabilistic* taste than Markov chains that evolve on a deterministic space. (Again, I'm stubborn enough to believe that there is a lot of interesting things to do and to say about the situation

where randomness is restricted to the transition probabilities themselves.) So, as I also said elsewhere, I'm sure that every reader will be able to single out her or his favourite among those topics that are *not* included here. In any case, I do hope that the selected material and presentation may provide some stimulus and usefulness.

Chapter 1
Preliminaries and basic facts

A Preliminaries, examples

The following introductory example is taken from the classical book by KEMENY and SNELL [K-S], where it is called "the weather in the land of OZ".

1.1 Example (The weather in Salzburg). [The author studied and worked in the beautiful city of Salzburg from 1979 to 1981. He hopes that Salzburg tourism authorities won't take offense from the following over-simplified "meteorological" model.] Italian tourists are very fond of Salzburg, the "Rome of the North". Arriving there, they discover rapidly that the weather is not as stable as in the South. There are never two consecutive days of bright weather. If one day is bright, then the next day it rains or snows with equal probability. A rainy or snowy day is followed with equal probability by a day with the same weather or by a change; in case of a change of the weather, it improves only in one half of the cases.

Let us denote the three possible states of the weather in Salzburg by ☼ (bright), 🌧 (rainy) and ❄ (snowy). The following table and figure illustrate the situation.

	☼	🌧	❄
☼	0	1/2	1/2
🌧	1/4	1/2	1/4
❄	1/4	1/4	1/2

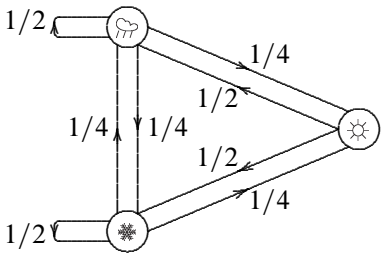

Figure 1

The table (matrix) tells us, for example, in the first row and second column that after a bright day comes a rainy day with probability $1/2$, or in the third row and first column that snowy weather is followed by a bright day with probability $1/4$.

Questions:

a) It rains today. What is the probability that two weeks from now the weather will be bright?

b) How many rainy days do we expect (on the average) during the next month?

c) What is the mean duration of a bad weather period?

1.2 Example (The drunkard's walk [folklore, also under the name of "gambler's ruin"]). A drunkard wants to return home from a pub. The pub is situated on a straight road. On the left, after 100 steps, it ends at a lake, while the drunkard's home is on the right at a distance of 200 steps from the pub, see Figure 2. In each step, the drunkard walks towards his house (with probability $2/3$) or towards the lake (with probability $1/3$). If he reaches the lake, he drowns. If he returns home, he stays and goes to sleep.

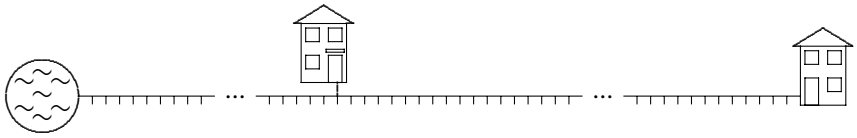

Figure 2

Questions:

a) What is the probability that the drunkard will drown? What is the probability that he will return home?

b) Supposing that he manages to return home, how much time ($\equiv$ how many steps) does it take him on the average?

1.3 Example (P. Gerl). A cat climbs a tree (Figure 3).

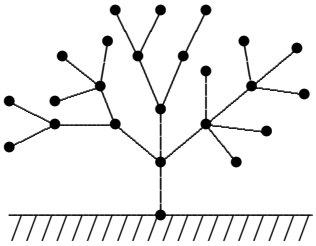

Figure 3

At each ramification point it decides by chance, typically with equal probability among all possibilities, to climb back down to the previous ramification point or to advance to one of the neighbouring higher points. If the cat arrives at the top (at a "leaf") then at the next step it will return to the preceding ramification point and continue as before.

Questions:

a) What is the probability that the cat will ever return to the ground? What is the probability that it will return at the n-th step?

b) How much time does it take on the average to return?

c) What is the probability that the cat will return to the ground before visiting any (or a specific given) leaf of the tree?

d) How often will the cat visit, on the average, a given "leaf" y before returning to the ground?

1.4 Example. Same as Example 1.3, but on another planet, where trees have infinite height (or an infinite number of branchings).

Further examples will be given later on.

What is a Markov chain?

We need the following ingredients.

(1) A *state space* X, finite or countably infinite (with elements u, v, w, x, y, etc., or other notation which is suitable in the respective context). In Example 1.1, $X = \{☼, 🌧, ❄\}$, the three possible states of the weather. In Example 1.2, $X = \{0, 1, 2, \dots, 300\}$, all possible distances (in steps) from the lake. In Examples 1.3 and 1.4, X is the set of all nodes (vertices) of the tree: the root, the ramification points, and the leaves.

(2) A *matrix* (table) of *one-step transition probabilities*

$$P = \big(p(x, y)\big)_{x,y \in X}.$$

Our random process consists of performing steps in the state space from one point to next one, and so on. The steps are random, that is, subject to a probability law. The latter is described by the transition matrix P: if at some instant, we are at some state (point) $x \in X$, the number $p(x, y)$ is the probability that the next step will take us to y, independently of how we arrived at x. Hence, we must have

$$p(x, y) \geq 0 \quad \text{and} \quad \sum_{y \in X} p(x, y) = 1 \quad \text{for all } x \in X.$$

In other words, P is a *stochastic matrix.*

(3) An *initial distribution* ν. This is a probability measure on X, and $\nu(x)$ is the probability that the random process starts at x.

Time is discrete, the steps are labelled by $\mathbb{N}_0$, the set of non-negative integers. At time 0 we start at a point $u \in X$. One after the other, we perform random steps: the position at time n is random and denoted by Z_n. Thus, Z_n, $n = 0, 1, \ldots$, is a sequence of *X-valued random variables*, called a *Markov chain.* Denote by Pr_u the probability of events concerning the Markov chain starting at u. We hence have

$$\mathsf{Pr}_u[Z_{n+1} = y \mid Z_n = x] = p(x, y), \quad n \in \mathbb{N}_0$$

(provided $\mathsf{Pr}_u[Z_n = x] > 0$, since the definition $\mathsf{Pr}(A \mid B) = \mathsf{Pr}(A \cap B)/\mathsf{Pr}(B)$ of conditional probability requires that the condition B has positive probability). In particular, the step which is performed at time n depends only on the current position (state), and not on the past history of how that position was reached, nor on the specific instant n: if $\mathsf{Pr}_u[Z_n = x, Z_{n-1} = x_{n-1}, \ldots, Z_1 = x_1] > 0$ and $\mathsf{Pr}_u[Z_m = x] > 0$ then

$$\begin{aligned} \mathsf{Pr}_u[Z_{n+1} = y \mid Z_n = x, Z_{n-1} = x_{n-1}, \ldots, Z_1 = x_1] \\ = \mathsf{Pr}_u[Z_{n+1} = y \mid Z_n = x] = \mathsf{Pr}_u[Z_{m+1} = y \mid Z_m = x] \\ = p(x, y). \end{aligned} \tag{1.5}$$

The graph of a Markov chain

A graph Γ consists of a denumerable, finite or infinite *vertex set* $V(\Gamma)$ and an edge set $E(\Gamma) \subset V(\Gamma) \times V(\Gamma)$; the edges are oriented: the edge $[x, y]$ goes from x to y. (Graph theorists would use the term "digraph", reserving "graph" to the situation where edges are non-oriented.) We also admit loops, that is, edges of the form $[x, x]$. If $[x, y] \in E(\Gamma)$, we shall also write $x \xrightarrow{1} y$.

1.6 Definition. Let Z_n, $n = 0, 1, \ldots$, be a Markov chain with state space X and transition matrix $P = \big(p(x, y)\big)_{x,y \in X}$. The vertex set of the *graph* $\Gamma = \Gamma(P)$ *of the Markov chain* is $V(\Gamma) = X$, and $[x, y]$ is an edge in $E(\Gamma)$ if and only if $p(x, y) > 0$.

We can also associate weights to the oriented edges: the edge $[x, y]$ is weighted with $p(x, y)$. Seen in this way, a Markov chain is a denumerable, oriented graph with weighted edges. The weights are positive and satisfy

$$\sum_{y : x \xrightarrow{1} y} p(x, y) = 1 \quad \text{for all } x \in V(\Gamma).$$

As a matter of fact, in Examples 1.1–1.3, we have already used those graphs for illustrating the respective Markov chains. We have followed the habit of drawing one non-oriented edge instead of a pair of oppositely oriented edges with the same endpoints.

B Axiomatic definition of a Markov chain

In order to give a precise definition of a Markov chain in the language of probability theory, we start with a probability space [1] $(\Omega^*, \mathcal{A}^*, \mathsf{Pr}^*)$ and a denumerable set X, the state space. With the latter, we implicitly associate the σ-algebra of all subsets of X.

1.7 Definition. A *Markov chain* is a sequence Z_n^*, $n = 0, 1, 2, \ldots$, of random variables (measurable functions) $Z_n^* \colon \Omega^* \to X$ with the following properties.

(i) *Markov property.* For all elements $x_0, x_1, \ldots, x_n$ and $x_{n+1} \in X$ which satisfy $\mathsf{Pr}^*[Z_n^* = x_n,\ Z_{n-1}^* = x_{n-1}, \ldots, Z_0^* = x_0] > 0$, one has

$$\begin{aligned}\mathsf{Pr}^*[Z_{n+1}^* = x_{n+1} \mid Z_n^* = x_n, Z_{n-1}^* = x_{n-1}, \ldots, Z_0^* = x_0] \\ = \mathsf{Pr}^*[Z_{n+1}^* = x_{n+1} \mid Z_n^* = x_n].\end{aligned}$$

(ii) *Time homogeneity.* For all elements $x, y \in X$ and $m, n \in \mathbb{N}_0$ which satisfy $\mathsf{Pr}^*[Z_m^* = x] > 0$ and $\mathsf{Pr}^*[Z_n^* = x] > 0$, one has

$$\mathsf{Pr}^*[Z_{m+1}^* = y \mid Z_m^* = x] = \mathsf{Pr}^*[Z_{n+1}^* = y \mid Z_n^* = x].$$

Here, $[Z_m^* = x]$ is the set ("event") $\{\omega^* \in \Omega^* \mid Z_m^*(\omega^*) = x\} \in \mathcal{A}^*$, and $\mathsf{Pr}^*[\cdot \mid Z_m^* = x]$ is probability conditioned by that event, and so on. In the sequel, the notation for an event [*logic expression*] will always refer to the set of all elements in the underlying probability space for which the logic expression is true.

If we write

$$p(x, y) = \mathsf{Pr}^*[Z_{n+1}^* = y \mid Z_n^* = x]$$

(which is independent of n as long as $\mathsf{Pr}^*[Z_n^* = x] > 0$), we obtain the transition matrix $P = \big(p(x, y)\big)_{x,y \in X}$ of (Z_n^*). The *initial distribution* of (Z_n^*) is the probability measure ν on X defined by

$$\nu(x) = \mathsf{Pr}^*[Z_0^* = x].$$

(We shall always write $\nu(x)$ instead of $\nu(\{x\})$, and $\nu(A) = \sum_{x \in A} \nu(x)$. The initial distribution represents an initial experiment, as in a board game where the initial position of a figure is chosen by throwing a dice. In case $\nu = \delta_u$, the point mass at $u \in X$, we say that the Markov chain *starts at* u.

More generally, one speaks of a Markov chain when just the Markov property (i) holds. Its fundamental significance is the absence of memory: the future (time $n + 1$) depends only on the present (time n) and not on the past (the time instants

[1] That is, a set Ω^* with a σ-algebra $\mathcal{A}^*$ of subsets of Ω^* and a σ-additive probability measure Pr^* on $\mathcal{A}^*$. The * superscript is used because later on, we shall usually reserve the notation $(\Omega, \mathcal{A}, \mathsf{Pr})$ for a specific probability space.

$0, \dots, n-1$). If one does not have property (ii), time homogeneity, this means that $p_{n+1}(x,y) = \mathsf{Pr}^*[Z^*_{n+1} = y \mid Z^*_n = x]$ depends also on the instant n besides x and y. In this case, the Markov chain is governed by a sequence P_n ($n \geq 1$) of stochastic matrices. In the present text, we shall limit ourselves to the study of time-homogeneous chains. We observe, though, that also non-time-homogeneous Markov chains are of considerable interest in various contexts, see e.g. the corresponding chapters in the books by Isaacson and Madsen [I-M], Seneta [Se] and Brémaud [Br].

In the introductory paragraph, we spoke about Markov chains starting directly with the state space X, the transition matrix P and the initial point $u \in X$, and the sequence of "random variables" (Z_n) was introduced heuristically. Thus, we now pose the question whether, with the ingredients X and P and a starting point $u \in X$ (or more generally, an initial distribution ν on X), one can always find a probability space on which the random position after n steps can be described as the n-th random variable of a Markov chain in the sense of the axiomatic Definition 1.7. This is indeed possible: that probability space, called the *trajectory space,* is constructed in the following, natural way.

We set

$$\Omega = X^{\mathbb{N}_0} = \{\omega = (x_0, x_1, x_2, \dots) \mid x_n \in X \text{ for all } n \geq 0\}. \tag{1.8}$$

An element $\omega = (x_0, x_1, x_2, \dots)$ represents a possible evolution (trajectory), that is, a possible sequence of points visited one after the other by the Markov chain. The probability that this single ω will indeed be the actual evolution should be the infinite product $\nu(x_0)p(x_0,x_1)p(x_1,x_2)\dots$, which however will usually converge to 0: as in the case of Lebesgue measure, it is in general not possible to construct the probability measure by assigning values to single elements ("atoms") of Ω.

Let $a_0, a_1, \dots, a_k \in X$. The *cylinder with base* $\boldsymbol{a} = (a_0, a_1, \dots, a_k)$ is the set

$$\begin{aligned} \mathsf{C}(\boldsymbol{a}) &= \mathsf{C}(a_0, a_1, \dots, a_k) \\ &= \{\omega = (x_0, x_1, x_2, \dots) \in \Omega : x_i = a_i, i = 0, \dots, k\}. \end{aligned} \tag{1.9}$$

This is the set of all possible ways how the evolution of the Markov chain may continue after the initial steps through $a_0, a_1, \dots, a_k$. With this set we associate the probability

$$\mathsf{Pr}_\nu\big(\mathsf{C}(a_0, a_1, \dots, a_k)\big) = \nu(a_0)p(a_0,a_1)p(a_1,a_2)\cdots p(a_{k-1},a_k). \tag{1.10}$$

We also consider Ω as a cylinder (with "empty base") with $\mathsf{Pr}_\nu(\Omega) = 1$.

For $n \in \mathbb{N}_0$, let $\mathcal{A}_n$ be the σ-algebra generated by the collection of all cylinders $\mathsf{C}(\boldsymbol{a})$ with $\boldsymbol{a} \in X^{k+1}$, $k \leq n$. Thus, $\mathcal{A}_n$ is in one-to-one correspondence with the collection of all subsets of X^{n+1}. Furthermore, $\mathcal{A}_n \subset \mathcal{A}_{n+1}$, and

$$\mathcal{F} = \bigcup_n \mathcal{A}_n$$

is an algebra of subsets of Ω: it contains Ω and is closed with respect to taking complements and finite unions. We denote by $\mathcal{A}$ the σ-algebra generated by $\mathcal{F}$.

Finally, we define the projections Z_n, $n \geq 0$: for $\omega = (x_0, x_1, x_2, \dots) \in \Omega$,

$$Z_n(\omega) = x_n. \tag{1.11}$$

1.12 Theorem. (a) *The measure* Pr_ν *has a unique extension to a probability measure on* $\mathcal{A}$*, also denoted* Pr_ν.

(b) *On the probability space* $(\Omega, \mathcal{A}, \mathsf{Pr}_\nu)$*, the projections* Z_n*,* $n = 0, 1, 2, \dots$*, define a Markov chain with state space* X*, initial distribution* ν *and transition matrix* P.

Proof. By (1.10), Pr_ν is defined on cylinder sets. Let $A \in \mathcal{A}_n$. We can write A as a finite disjoint union of cylinders of the form $\mathsf{C}(a_0, a_1, \dots, a_n)$. Thus, we can use (1.10) to define a σ-additive probability measure ν_n on $\mathcal{A}_n$ by

$$\nu_n(A) = \sum_{\boldsymbol{a} \in X^{n+1} \,:\, \mathsf{C}(\boldsymbol{a}) \subset A} Pr_\nu\big(\mathsf{C}(\boldsymbol{a})\big).$$

We prove that ν_n coincides with Pr_ν on the cylinder sets, that is,

$$\nu_n\big(\mathsf{C}(a_0, a_1, \dots, a_k)\big) = \mathsf{Pr}_\nu\big(\mathsf{C}(a_0, a_1, \dots, a_k)\big), \quad \text{if } k \leq n. \tag{1.13}$$

We proceed by "reversed" induction on k. By (1.10), the identity (1.13) is valid for $k = n$. Now assume that for some k with $0 < k \leq n$, the identity (1.13) is valid for every cylinder $\mathsf{C}(a_0, a_1, \dots, a_k)$. Consider a cylinder $\mathsf{C}(a_0, a_1, \dots, a_{k-1})$. Then

$$\mathsf{C}(a_0, a_1, \dots, a_{k-1}) = \bigcup_{a_k \in X} \mathsf{C}(a_0, a_1, \dots, a_k),$$

a disjoint union. By the induction hypothesis,

$$\nu_n\big(\mathsf{C}(a_0, a_1, \dots, a_{k-1})\big) = \sum_{a_k \in X} \mathsf{Pr}_\nu\big(\mathsf{C}(a_0, a_1, \dots, a_k)\big).$$

But from the definition of Pr_ν and stochasticity of the matrix P, we get

$$\begin{aligned} \sum_{a_k \in X} \mathsf{Pr}_\nu\big(\mathsf{C}(a_0, a_1, \dots, a_k)\big) &= \sum_{a_k \in X} \mathsf{Pr}_\nu\big(\mathsf{C}(a_0, a_1, \dots, a_{k-1})\big)\, p(a_{k-1}, a_k) \\ &= \mathsf{Pr}_\nu\big(\mathsf{C}(a_0, a_1, \dots, a_{k-1})\big), \end{aligned}$$

which proves (1.13).

(1.13) tells us that for $n \geq k$ and $A \in \mathcal{A}_k \subset \mathcal{A}_n$, we have $\nu_n(A) = \nu_k(A)$. Therefore, we can extend Pr_ν from the collection of all cylinder sets to $\mathcal{F}$ by setting

$$\mathsf{Pr}_\nu(A) = \nu_n(A), \quad \text{if } A \in \mathcal{A}_n.$$

We see that Pr_ν is a finitely additive measure with total mass 1 on the algebra $\mathcal{F}$, and it is σ-additive on each $\mathcal{A}_n$. In order to prove that Pr_ν is σ-additive on $\mathcal{F}$, on the basis of a standard theorem from measure theory (see e.g. HALMOS [Hal]), it is sufficient to verify continuity of Pr_ν at $\emptyset$: if (A_n) is a decreasing sequence of sets in $\mathcal{F}$ with $\bigcap_n A_n = \emptyset$, then $\lim_n \mathsf{Pr}_\nu(A_n) = 0$.

Let us now suppose that (A_n) is a decreasing sequence of sets in $\mathcal{F}$, and that $\lim_n \mathsf{Pr}_\nu(A_n) > 0$. Since the σ-algebras $\mathcal{A}_n$ increase with n, there are numbers $k(n)$ such that $A_n \in \mathcal{A}_{k(n)}$ and $0 \le k(n) < k(n+1)$. We claim that there is a sequence of cylinders $\mathsf{C}_n = \mathsf{C}(\boldsymbol{a}_n)$ with $\boldsymbol{a}_n \in X^{k(n)+1}$ such that for every n

(i) $\mathsf{C}_{n+1} \subset \mathsf{C}_n$,

(ii) $\mathsf{C}_n \subset A_n$, and

(iii) $\lim_m \mathsf{Pr}_\nu(A_m \cap \mathsf{C}_n) > 0$.

Proof. The σ-additivity of Pr_ν on $\mathcal{A}_{k(m)}$ implies

$$\begin{aligned} 0 < \lim_m \mathsf{Pr}_\nu(A_m) &= \lim_m \sum_{\boldsymbol{a} \in X^{k(0)+1}} \mathsf{Pr}_\nu\big(A_m \cap \mathsf{C}(\boldsymbol{a})\big) \\ &= \sum_{\boldsymbol{a} \in X^{k(0)+1}} \lim_m \mathsf{Pr}_\nu\big(A_m \cap \mathsf{C}(\boldsymbol{a})\big). \end{aligned} \tag{1.14}$$

It is legitimate to exchange sum and limit in (1.14). Indeed,

$$\mathsf{Pr}_\nu\big(A_m \cap \mathsf{C}(\boldsymbol{a})\big) \le \mathsf{Pr}_\nu\big(A_0 \cap \mathsf{C}(\boldsymbol{a})\big), \qquad \sum_{\boldsymbol{a} \in X^{k(0)+1}} \mathsf{Pr}_\nu\big(A_0 \cap \mathsf{C}(\boldsymbol{a})\big) = \mathsf{Pr}_\nu(A_0) < \infty,$$

and considering the sum as a discrete integral with respect to the variable $\boldsymbol{a} \in X^{k(0)+1}$, Lebesgue's dominated convergence theorem justifies (1.14).

From (1.14) it follows that there is $\boldsymbol{a}_0 \in X^{k(0)+1}$ such that for $\mathsf{C}_0 = \mathsf{C}(\boldsymbol{a}_0)$ one has

$$\lim_m \mathsf{Pr}_\nu(A_m \cap \mathsf{C}_0) > 0. \tag{1.15}$$

A_0 is a disjoint union of cylinders $\mathsf{C}(\boldsymbol{b})$ with $\boldsymbol{b} \in X^{k(0)+1}$. In particular, either $\mathsf{C}_0 \cap A_0 = \emptyset$ or $\mathsf{C}_0 \subset A_0$. The first case is impossible, since $\mathsf{Pr}_\nu(A_0 \cap \mathsf{C}_0) > 0$. Therefore C_0 satisfies (ii) and (iii).

Now let us suppose to have already constructed $\mathsf{C}_0, \dots, \mathsf{C}_{r-1}$ such that (i), (ii) and (iii) hold for all indices $n < r$. We define a sequence of sets

$$A'_n = A_{n+r} \cap \mathsf{C}_{r-1}, \quad n \ge 0.$$

By our hypotheses, the sequence (A'_n) is decreasing, $\lim_n \mathsf{Pr}_\nu(A'_n) > 0$, and $A'_n \in \mathcal{A}_{k'(n)}$, where $k'(n) = k(n+r)$. Hence, the reasoning that we have applied

above to the sequence (A_n) can now be applied to (A'_n), and we obtain a cylinder $\mathsf{C}_r = \mathsf{C}(\boldsymbol{a}_r)$ with $\boldsymbol{a}_r \in X^{k'(0)+1} = X^{k(n)+1}$, such that

$$\lim_m \mathsf{Pr}_\nu(A'_m \cap \mathsf{C}_r) > 0 \quad \text{and} \quad \mathsf{C}_r \subset A'_0.$$

From the definition of A'_n, we see that also

$$\lim_m \mathsf{Pr}_\nu(A_m \cap \mathsf{C}_r) > 0 \quad \text{and} \quad \mathsf{C}_r \subset A_r \cap \mathsf{C}_{r-1}.$$

Summarizing, (i), (ii) and (iii) are valid for $\mathsf{C}_0 \ldots \mathsf{C}_r$, and the existence of the proposed sequence (C_n) follows by induction.

Since $\mathsf{C}_n = \mathsf{C}(\boldsymbol{a}_n)$, where $\boldsymbol{a}_n \in X^{k(n)+1}$, it follows from (i) that the initial piece of $\boldsymbol{a}_{n+1}$ up to the index $k(n)$ must be $\boldsymbol{a}_n$. Thus, there is a trajectory $\omega = (a_0, a_1, \ldots)$ such that $\boldsymbol{a}_n = (a_0, \ldots, a_{k(n)})$ for each n. Consequently, $\omega \in \mathsf{C}_n$ for each n, and

$$\bigcap_n A_n \supset \bigcap \mathsf{C}_n \neq \emptyset.$$

We have proved that Pr_ν is continuous at $\emptyset$, and consequently σ-additive on $\mathcal{F}$. As mentioned above, now one of the fundamental theorems of measure theory (see e.g. [Hal, §13]) asserts that Pr_ν extends in a unique way to a probability measure on the σ-algebra $\mathcal{A}$ generated by $\mathcal{F}$: statement (a) of the theorem is proved.

For verifying (b), consider $x_0, x_1, \ldots, x = x_n, y = x_{n+1} \in X$ such that $\mathsf{Pr}_\nu[Z_0 = x_0, \ldots, Z_n = x_n] > 0$. Then

$$\begin{aligned} \mathsf{Pr}_\nu[Z_{n+1} = y \mid Z_n = x, Z_i = x_i \text{ for all } i < n] &= \frac{\mathsf{Pr}_\nu\big(\mathsf{C}(x_0, \ldots, x_{n+1})\big)}{\mathsf{Pr}_\nu\big(\mathsf{C}(x_0, \ldots, x_n)\big)} \\ &= p(x, y). \end{aligned} \tag{1.16}$$

On the other hand, consider the events $A = [Z_{n+1} = y]$ and $B = [Z_n = x]$ in $\mathcal{A}$. We can write $B = \bigcup\{\mathsf{C}(\boldsymbol{a}) \mid \boldsymbol{a} \in X^{n+1}, a_n = x\}$. Using (1.16), we get

$$\begin{aligned} \mathsf{Pr}_\nu[Z_{n+1} = y \mid Z_n = x] &= \frac{\mathsf{Pr}_\nu(A \cap B)}{\mathsf{Pr}_\nu(B)} \\ &= \sum_{\substack{\boldsymbol{a} \in X^{n+1} : a_n = x, \\ \mathsf{Pr}_\nu(\mathsf{C}(\boldsymbol{a})) > 0}} \frac{\mathsf{Pr}_\nu\big(A \cap \mathsf{C}(\boldsymbol{a})\big)}{\mathsf{Pr}_\nu(B)} \\ &= \sum_{\substack{\boldsymbol{a} \in X^{n+1} : a_n = x, \\ \mathsf{Pr}_\nu(\mathsf{C}(\boldsymbol{a})) > 0}} \mathsf{Pr}_\nu\big(A \mid \mathsf{C}(\boldsymbol{a})\big) \frac{\mathsf{Pr}_\nu\big(\mathsf{C}(\boldsymbol{a})\big)}{\mathsf{Pr}_\nu(B)} = \end{aligned}$$

$$= \sum_{\substack{\boldsymbol{a}\in X^{n+1}\,:\,a_n=x,\\ \mathsf{Pr}_\nu(\mathsf{C}(\boldsymbol{a}))>0}} p(x,y)\,\frac{\mathsf{Pr}_\nu\big(\mathsf{C}(\boldsymbol{a})\big)}{\mathsf{Pr}_\nu(B)}$$

$$= p(x,y).$$

Thus, (Z_n) is a Markov chain on X with transition matrix P. Finally, for $x \in X$,

$$\mathsf{Pr}_\nu[Z_0 = x] = \mathsf{Pr}_\nu\big(\mathsf{C}(x)\big) = \nu(x),$$

so that the distribution of Z_0 is ν. □

We observe that beginning with the point (1.13) regarding the measures ν_n, the last theorem can be deduced from Kolmogorov's theorem on the construction of probability measures on infinite products of Polish spaces. As a matter of fact, the proof given here is basically the one of Kolmogorov's theorem adapted to our special case. For a general and advanced treatment of that theorem, see e.g. Parthasarathy [Pa, Chapter V].

The following theorem tells us that (1) the "generic" probability space of the axiomatic Definition 1.7 on which the Markov chain (Z_n^*) is defined is always "larger" (in a probabilistic sense) than the associated trajectory space, and that (2) the trajectory space contains already all the information on the individual Markov chain under consideration.

1.17 Theorem. *Let $(\Omega^*, \mathcal{A}^*, \mathsf{Pr}^*)$ be a probability space and (Z_n^*) a Markov chain defined on Ω^*, with state space X, initial distribution ν and transition matrix P. Equip the trajectory space $(\Omega, \mathcal{A})$ with the probability measure Pr_ν defined in (1.10) and Theorem 1.12, and with the sequence of projections (Z_n) of (1.11). Then the function*

$$\tau\colon \Omega^* \to \Omega, \quad \omega^* \mapsto \omega = \big(Z_0^*(\omega^*), Z_1^*(\omega^*), Z_2^*(\omega^*), \dots\big)$$

is measurable, $Z_n^*(\omega^*) = Z_n\big(\tau(\omega^*)\big)$, and

$$\mathsf{Pr}_\nu(A) = \mathsf{Pr}^*\big(\tau^{-1}(A)\big) \quad \textit{for all } A \in \mathcal{A}.$$

Proof. $\mathcal{A}$ is the σ-algebra generated by all cylinders. By virtue of the extension theorems of measure theory, for the proof it is sufficient to verify that $\tau^{-1}(\mathsf{C}) \in \mathcal{A}^*$ and $\mathsf{Pr}_\nu(\mathsf{C}) = \mathsf{Pr}^*\big(\tau^{-1}(\mathsf{C})\big)$ for each cylinder C. Thus, let $\mathsf{C} = \mathsf{C}(a_0, \dots, a_k)$ with $a_i \in X$. Then

$$\tau^{-1}(\mathsf{C}) = \{\omega^* \in \Omega^* \mid Z_i^*(\omega^*) = a_i,\ i = 0, \dots k\}$$

$$= \bigcap_{i=0}^{k} \{\omega^* \in \Omega^* \mid Z_i^*(\omega^*) = a_i\} \in \mathcal{A}^*,$$

and by the Markov property,

$$\begin{aligned}\mathsf{Pr}^*\big(\tau^{-1}(\mathsf{C})\big) &= \mathsf{Pr}^*[Z_k^* = a_k, Z_{k-1}^* = a_{k-1}, \dots Z_0^* = a_0] \\ &= \mathsf{Pr}^*[Z_0^* = a_0] \prod_{i=1}^{k} \mathsf{Pr}^*[Z_i^* = a_i \mid Z_{i-1}^* = a_{i-1}, \dots Z_0^* = a_0] \\ &= \nu(a_0) \prod_{i=1}^{k} p(a_{i-1}, a_i) = \mathsf{Pr}_\nu(\mathsf{C}). \qquad \square\end{aligned}$$

From now on, given the state space X and the transition matrix P, we shall (almost) always consider the trajectory space $(\Omega, \mathcal{A})$ with the family of probability measures Pr_ν, corresponding to all the initial distributions ν on X. Thus, our usual model for the Markov chain on X with initial distribution ν and transition matrix P will always be the sequence (Z_n) of the projections defined on the trajectory space $(\Omega, \mathcal{A}, \mathsf{Pr}_\nu)$. Typically, we shall have $\nu = \delta_u$, a point mass at $u \in X$. In this case, we shall write $\mathsf{Pr}_\nu = \mathsf{Pr}_u$. Sometimes, we shall omit to specify the initial distribution and call *Markov chain* the pair (X, P).

1.18 Remarks. (a) The fact that on the trajectory space one has a whole family of different probability measures that describe the same Markov chain, but with different initial distributions, seems to create every now and then some confusion at an initial level. This can be overcome by choosing and fixing one specific initial distribution ν_0 which is supported by the whole of X. If X is finite, this may be equidistribution. If X is infinite, one can enumerate $X = \{x_k : k \in \mathbb{N}\}$ and set $\nu_0(x_k) = 2^{-k}$. Then we can assign one "global" probability measure on $(\Omega, \mathcal{A})$ by $\mathsf{Pr} = \mathsf{Pr}_{\nu_0}$. With this choice, we get

$$\mathsf{Pr}_u = \mathsf{Pr}[\,\cdot \mid Z_0 = u] \quad \text{and} \quad \mathsf{Pr}_\nu = \sum_{u \in X} \mathsf{Pr}[\,\cdot \mid Z_0 = u]\, \nu(u),$$

if ν is another initial distribution.

(b) As mentioned in the Preface, if one wants to avoid measure theory, and in particular the extension machinery of Theorem 1.12, then one can carry out most of the theory by considering the Markov chain in a finite time interval $[0, n]$. The trajectory space is then X^{n+1} (starting with index 0), the σ-algebra consists of all subsets of X^{n+1} and is in one-to-one correspondence with $\mathcal{A}_n$, and the probability measure $\mathsf{Pr}_\nu = \mathsf{Pr}_\nu^n$ is atomic with

$$\mathsf{Pr}_\nu^n(\boldsymbol{a}) = \nu(a_0) p(a_0, a_1) \cdots p(a_{n-1}, a_n), \quad \text{if } \boldsymbol{a} = (a_0, \dots, a_n).$$

For deriving limit theorems, one may first consider that time interval and then let $n \to \infty$.

C Transition probabilities in *n* steps

Let (X, P) be a Markov chain. What is the probability to be at y at the n-th step, starting from x ($n \geq 0$, $x, y \in X$)? We are interested in the number

$$p^{(n)}(x, y) = \mathsf{Pr}_x[Z_n = y].$$

Intuitively the following is already clear (and will be proved in a moment): if ν is an initial distribution and $k \in \mathbb{N}_0$ is such that $\mathsf{Pr}_\nu[Z_k = x] > 0$, then

$$\mathsf{Pr}_\nu[Z_{k+n} = y \mid Z_k = x] = p^{(n)}(x, y). \tag{1.19}$$

We have

$$p^{(0)}(x, x) = 1, \quad p^{(0)}(x, y) = 0 \text{ if } x \neq y, \quad \text{and} \quad p^{(1)}(x, y) = p(x, y).$$

We can compute $p^{(n+1)}(x, y)$ starting from the values of $p^{(n)}(\cdot, \cdot)$ as follows, by decomposing with respect to the first step: the first step goes from x to some $w \in X$, with probability $p(x, w)$. The remaining n steps have to take us from w to y, with probability $p^{(n)}(w, y)$ not depending on the past history. We have to sum over all possible points w:

$$p^{(n+1)}(x, y) = \sum_{w \in X} p(x, w)\, p^{(n)}(w, y). \tag{1.20}$$

Let us verify the formulas (1.19) and (1.20) rigorously. For $n = 0$ and $n = 1$, they are true. Suppose that they both hold for $n \geq 1$, and that $\mathsf{Pr}_\nu[Z_k = x] > 0$. Recall the following elementary rules for conditional probability. If $A, B, C \in \mathcal{A}$, then

$$\mathsf{Pr}_\nu(A \mid C) = \begin{cases} \mathsf{Pr}_\nu(A \cap C)/\,\mathsf{Pr}_\nu(C), & \text{if } \mathsf{Pr}_\nu(C) > 0, \\ 0, & \text{if } \mathsf{Pr}_\nu(C) = 0; \end{cases}$$

$$\mathsf{Pr}_\nu(A \cap B \mid C) = \mathsf{Pr}_\nu(B \mid C)\ \mathsf{Pr}_\nu(A \mid B \cap C);$$

$$\mathsf{Pr}_\nu(A \cup B \mid C) = \mathsf{Pr}_\nu(A \mid C) + \mathsf{Pr}_\nu(B \mid C), \quad \text{if } A \cap B = \emptyset.$$

Hence we deduce via 1.7 that

$$\begin{aligned} &\mathsf{Pr}_\nu[Z_{k+n+1} = y \mid Z_k = x] \\ &= \mathsf{Pr}_\nu[\exists\, w \in X : (Z_{k+n+1} = y \text{ and } Z_{k+1} = w) \mid Z_k = x] \\ &= \sum_{w \in X} \mathsf{Pr}_\nu[Z_{k+n+1} = y \text{ and } Z_{k+1} = w \mid Z_k = x] \\ &= \sum_{w \in X} \mathsf{Pr}_\nu[Z_{k+1} = w \mid Z_k = x]\ \mathsf{Pr}_\nu[Z_{k+n+1} = y \mid Z_{k+1} = w \text{ and } Z_k = x] \end{aligned}$$

(note that $\mathsf{Pr}_\nu[Z_{k+1} = w \text{ and } Z_k = x] > 0$ if $\mathsf{Pr}_\nu[Z_{k+1} = w \mid Z_k = x] > 0$)

$$= \sum_{w \in X} \mathsf{Pr}_\nu[Z_{k+1} = w \mid Z_k = x]\ \mathsf{Pr}_\nu[Z_{k+n+1} = y \mid Z_{k+1} = w]$$
$$= \sum_{w \in X} p(x, w)\, p^{(n)}(w, y).$$

Since the last sum does not depend on ν or k, it must be equal to

$$\mathsf{Pr}_x[Z_{n+1} = y] = p^{(n+1)}(x, y).$$

Therefore, we have the following.

1.21 Lemma. (a) *The number $p^{(n)}(x, y)$ is the element at position (x, y) in the n-th power P^n of the transition matrix,*

(b) $p^{(m+n)}(x, y) = \sum_{w \in X} p^{(m)}(x, w)\, p^{(n)}(w, y)$,

(c) *P^n is a stochastic matrix.*

Proof. (a) follows directly from the identity (1.20).

(b) follows from the identity $P^{m+n} = P^m P^n$ for the matrix powers of P.

(c) is best understood by the probabilistic interpretation: starting at $x \in X$, it is certain that after n steps one reaches *some* element of X, that is

$$1 = \mathsf{Pr}_x[Z_n \in X] = \sum_{y \in X} \mathsf{Pr}_x[Z_n = y] = \sum_{y \in X} p^{(n)}(x, y). \qquad \square$$

1.22 Exercise. Let (Z_n) be a Markov chain on the state space X, and let $0 \le n_1 < n_2 < \cdots < n_{k+1}$. Show that for $0 < m < n$, $x, y \in X$ and $A \in \mathcal{A}_m$ with $\mathsf{Pr}_\nu(A) > 0$,

$$\text{if } Z_m(\omega) = x \text{ for all } \omega \in A, \text{ then } \mathsf{Pr}_\nu[Z_n = y \,|\, A] = p^{(n-m)}(x, y).$$

Deduce that if $x_1, \ldots, x_{k+1} \in X$ are such that $\mathsf{Pr}_\nu[Z_{n_k} = x_k, \ldots, Z_{n_1} = x_1] > 0$, then

$$\mathsf{Pr}_\nu[Z_{n_{k+1}} = x_{k+1} \,|\, Z_{n_k} = x_k, \ldots, Z_{n_1} = x_1] = \mathsf{Pr}_\nu[Z_{n_{k+1}} = x_{k+1} \,|\, Z_{n_k} = x_k].$$
$\square$

A *real random variable* is a measurable function $\boldsymbol{f} : (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ and $\overline{\mathcal{B}}$ is the σ-algebra of extended Borel sets. If the integral of $\boldsymbol{f}$ with respect to Pr_ν exists, we denote it by

$$\mathsf{E}_\nu(\boldsymbol{f}) = \int_\Omega \boldsymbol{f}\ d\,\mathsf{Pr}_\nu\,.$$

This is the *expectation* or *expected value* of $\boldsymbol{f}$; if $\nu = \delta_x$, we write $\mathsf{E}_x(\boldsymbol{f})$.

An example is the number of visits of the Markov chain to a set $W \subset X$. We define

$$\boldsymbol{v}_n^W(\omega) = \mathbf{1}_W\big(Z_n(\omega)\big), \quad \text{where } \mathbf{1}_W(x) = \begin{cases} 1, & \text{if } x \in W, \\ 0, & \text{otherwise} \end{cases}$$

is the indicator function of the set W. If $W = \{x\}$, we write $\boldsymbol{v}_n^W = \boldsymbol{v}_n^x$.[2] The random variable

$$\boldsymbol{v}_{[k,n]}^W = \boldsymbol{v}_k^W + \boldsymbol{v}_{k+1}^W + \cdots + \boldsymbol{v}_n^W \quad (k \le n)$$

is the number of visits of Z_n in the set W during the time period from step k to step n. It is often called the *local time* spent in W by the Markov chain during that time interval. We write

$$\boldsymbol{v}^W = \sum_{n=0}^{\infty} \boldsymbol{v}_n^W$$

for the total number of visits (total local time) in W. We have

$$\mathsf{E}_\nu(\boldsymbol{v}_{[k,n]}^W) = \sum_{x\in X} \sum_{j=k}^{n} \sum_{y\in W} \nu(x)\, p^{(j)}(x,y). \tag{1.23}$$

1.24 Definition. A *stopping time* is a random variable $\boldsymbol{t}$ taking its values in $\mathbb{N}_0 \cup \{\infty\}$, such that

$$[\boldsymbol{t} \le n] = \{\omega \in \Omega \mid \boldsymbol{t}(\omega) \le n\} \in \mathcal{A}_n \quad \text{for all } n \in \mathbb{N}_0.$$

That is, the property that a trajectory $\omega = (x_0, x_1, \dots)$ satisfies $\boldsymbol{t}(\omega) \le n$ depends only on $(x_0, \dots, x_n)$. In still other words, this means that one can decide whether $\boldsymbol{t} \le n$ or not by observing only $Z_0, \dots, Z_n$.

1.25 Exercise (Strong Markov property). Let $(Z_n)_{n\ge0}$ be a Markov chain with initial distribution ν and transition matrix P on the state space X, and let $\boldsymbol{t}$ be a stopping time with $\mathsf{Pr}_\nu[\boldsymbol{t} < \infty] = 1$. Show that $(Z_{\boldsymbol{t}+n})_{n\ge0}$, defined by

$$Z_{\boldsymbol{t}+n}(\omega) = Z_{\boldsymbol{t}(\omega)+n}(\omega), \quad \omega \in \Omega,$$

is again a Markov chain with transition matrix P and initial distribution

$$\bar\nu(x) = \mathsf{Pr}_\nu[Z_{\boldsymbol{t}} = x] = \sum_{k=0}^{\infty} \mathsf{Pr}_\nu[Z_k = x,\ \boldsymbol{t} = k].$$

[Hint: decompose according to the values of $\boldsymbol{t}$ and apply Exercise 1.22.] □

[2] We shall try to distinguish between the *measure* δ_x and the *function* $\mathbf{1}_x$.

For $W \subset X$, two important examples of – not necessarily a.s. finite – stopping times are the *hitting times,* also called *first passage times*:

$$\boldsymbol{s}^W = \inf\{n \geq 0 : Z_n \in W\} \quad \text{and} \quad \boldsymbol{t}^W = \inf\{n \geq 1 : Z_n \in W\}. \tag{1.26}$$

Observe that, for example, the definition of $\boldsymbol{s}^W$ should be read as $\boldsymbol{s}^W(\omega) = \inf\{n \geq 0 \mid Z_n(\omega) \in W\}$ and the infimum over the empty set is to be taken as $+\infty$. Thus, $\boldsymbol{s}^W$ is the instant of the first visit of Z_n in W, while $\boldsymbol{t}^W$ is the instant of the first visit in W *after starting.* Again, we write $\boldsymbol{s}^x$ and $\boldsymbol{t}^x$ if $W = \{x\}$.

The following quantities play a crucial role in the study of Markov chains.

$$\begin{aligned} G(x,y) &= \mathsf{E}_x(\boldsymbol{v}^y), \\ F(x,y) &= \mathsf{Pr}_x[\boldsymbol{s}^y < \infty], \quad \text{and} \\ U(x,y) &= \mathsf{Pr}_x[\boldsymbol{t}^y < \infty], \quad x, y \in X. \end{aligned} \tag{1.27}$$

$G(x,y)$ is the expected number of visits of (Z_n) in y, starting at x, while $F(x,y)$ is the probability to ever visit y starting at x, and $U(x,y)$ is the probability to visit y *after* starting at x. In particular, $U(x,x)$ is the probability to ever return to x, while $F(x,x) = 1$. Also, $F(x,y) = U(x,y)$ when $y \neq x$, since the two stopping times $\boldsymbol{s}^y$ and $\boldsymbol{t}^y$ coincide when the starting point differs from the target point y. Furthermore, if we write

$$\begin{aligned} f^{(n)}(x,y) &= \mathsf{Pr}_x[\boldsymbol{s}^y = n] = \mathsf{Pr}_x[Z_n = y,\ Z_i \neq y \text{ for } 0 \leq i < n] \quad \text{and} \\ u^{(n)}(x,y) &= \mathsf{Pr}_x[\boldsymbol{t}^y = n] = \mathsf{Pr}_x[Z_n = y,\ Z_i \neq y \text{ for } 1 \leq i < n], \end{aligned} \tag{1.28}$$

then $f^{(0)}(x,x) = 1$, $f^{(0)}(x,y) = 0$ if $y \neq x$, $u^{(0)}(x,y) = 0$ for all x, y, and $f^{(n)}(x,y) = u^{(n)}(x,y)$ for all n, if $y \neq x$. We get

$$G(x,y) = \sum_{n=0}^{\infty} p^{(n)}(x,y), \quad F(x,y) = \sum_{n=0}^{\infty} f^{(n)}(x,y),$$

$$U(x,y) = \sum_{n=0}^{\infty} u^{(n)}(x,y).$$

We can now give first answers to the questions of Example 1.1.

a) If it rains today then the probability to have bright weather two weeks from now is

$$p^{(14)}(\text{🌧}, \text{☼}).$$

This is number obtained by computing the 14-th power of the transition matrix. One of the methods to compute this power in practice is to try to diagonalize the transition matrix via its eigenvalues. In the sequel, we shall also study other

methods of explicit computation and compute the numerical values for the solutions that answer this and the next question.

b) If it rains today, then the expected number of rainy days during the next month (thirty-one days, starting today) is

$$\mathsf{E}_{\text{🌧}}\big(\boldsymbol{v}^{\text{🌧}}_{[0,\,30]}\big) = \sum_{n=0}^{30} p^{(n)}(\text{🌧},\text{🌧}).$$

c) A bad weather period begins after a bright day. It lasts for exactly n days with probability

$$\mathsf{Pr}_{\text{☼}}[Z_1, Z_2, \dots Z_n \in \{\text{🌧},\text{❄}\},\ Z_{n+1} = \text{☼}] = \mathsf{Pr}_{\text{☼}}[\boldsymbol{t}^{\text{☼}} = n+1] = u^{(n+1)}(\text{☼},\text{☼}).$$

We shall see below that the bad weather does not continue forever, that is, $\mathsf{Pr}_{\text{☼}}[\boldsymbol{t}^{\text{☼}} = \infty] = 0$. Hence, the expected duration (in days) of a bad weather period is

$$\sum_{n=1}^{\infty} n\, u^{(n+1)}(\text{☼},\text{☼}).$$

In order to compute this number, we can simplify the Markov chain. We have $p(\text{🌧},\text{☼}) = p(\text{❄},\text{☼}) = 1/4$, so bad weather today (🌧❄) is followed by a bright day tomorrow (☼) with probability $1/4$, and again by bad weather tomorrow (🌧❄) with probability $3/4$, independently of the particular type (🌧 or ❄) of today's bad weather. Bright weather today (☼) is followed with probability 1 by bad weather tomorrow (🌧❄). Thus, we can combine the two states 🌧 and ❄ into a single, new state "bad weather", denoted ☁, and we obtain a new state space $\bar{X} = \{\text{☼}, \text{☁}\}$ with a new transition matrix $\bar{P}$:

$$\begin{aligned} \bar{p}(\text{☼},\text{☼}) &= 0, & \bar{p}(\text{☼},\text{☁}) &= 1,\\ \bar{p}(\text{☁},\text{☼}) &= 1/4, & \bar{p}(\text{☁},\text{☁}) &= 3/4. \end{aligned}$$

For the computation of probabilities of events which do not distinguish between the different types of bad weather, we can use the new Markov chain $(\bar{X}, \bar{P})$, with the associated probability measures $\overline{\mathsf{Pr}}_{\bullet}$, $\bullet \in \bar{X}$. We obtain

$$\begin{aligned} \mathsf{Pr}_{\text{☼}}[\boldsymbol{t}^{\text{☼}} = n+1] &= \overline{\mathsf{Pr}}_{\text{☼}}[\boldsymbol{t}^{\text{☼}} = n+1]\\ &= \bar{p}(\text{☼},\text{☁})\left(\bar{p}(\text{☁},\text{☁})\right)^{n-1}\bar{p}(\text{☁},\text{☼})\\ &= (3/4)^{n-1}\cdot(1/4). \end{aligned}$$

In particular, we verify

$$\mathsf{Pr}_{\text{☼}}[\boldsymbol{t}^{\text{☼}} = \infty] = 1 - \sum_{n=0}^{\infty} \overline{\mathsf{Pr}}_{\text{☼}}[\boldsymbol{t}^{\text{☼}} = n+1] = 0.$$

Using

$$\sum_{n=1}^{\infty} n\, z^{n-1} = \Big(\sum_{n=0}^{\infty} z^n\Big)' = \Big(\frac{1}{1-z}\Big)' = \frac{1}{(1-z)^2}\,,$$

we now find that the mean duration of a bad weather period is

$$\frac{1}{4}\sum_{n=1}^{\infty} n\left(\frac{3}{4}\right)^{n-1} = \frac{1}{4}\cdot\frac{1}{(1-\frac{3}{4})^2} = 4. \qquad \square$$

In the last example, we have applied a useful method for simplifying a Markov chain (X, P), namely *factorization.* Suppose that we have a partition $\bar{X}$ of the state space X with the following property.

$$\text{For all } \bar{x},\ \bar{y} \in \bar{X}, \quad p(x, \bar{y}) = \sum_{y\in\bar{y}} p(x, y) \quad \text{is constant for } x \in \bar{x}. \tag{1.29}$$

If this holds, we can consider $\bar{X}$ as a new state space with transition matrix $\bar{P}$, where

$$\bar{p}(\bar{x}, \bar{y}) = p(x, \bar{y}), \quad \text{with arbitrary } x \in \bar{x}. \tag{1.30}$$

The new Markov chain is the *factor chain* with respect to the given partition.

More precisely, let π be the natural projection $X \to \bar{X}$. Choose an initial distribution ν on X and let $\bar{\nu} = \pi(\nu)$, that is, $\bar{\nu}(\bar{x}) = \sum_{x\in\bar{x}} \nu(x)$. Consider the associated trajectory spaces $(\Omega, \mathcal{A}, \mathsf{Pr}_\nu)$ and $(\bar{\Omega}, \bar{\mathcal{A}}, \overline{\mathsf{Pr}}_{\bar{\nu}})$ and the natural extension $\pi : \Omega \to \bar{\Omega}$, namely $\pi(x_0, x_1, \dots) = \big(\pi(x_0), \pi(x_1), \dots\big)$. Then we obtain for the Markov chains (Z_n) on X and $(\bar{Z}_n)$ on $\bar{X}$

$$\bar{Z}_n = \pi(Z_n) \quad \text{and} \quad \mathsf{Pr}_\nu\big(\pi^{-1}(\bar{A})\big) = \overline{\mathsf{Pr}}_{\bar{\nu}}(\bar{A}) \quad \text{for all } \bar{A} \in \bar{\mathcal{A}}.$$

We leave the verification as an exercise to the reader.

1.31 Exercise. Let (Z_n) be a Markov chain on the state space X with transition matrix P, and let $\bar{X}$ be a partition of X with the natural projection $\pi : X \to \bar{X}$. Show that $\big(\pi(Z_n)\big)$ is (for every starting point in X) a Markov chain on $\bar{X}$ if and only if (1.29) holds. $\square$

D Generating functions of transition probabilities

If (a_n) is a sequence of real or complex numbers then the complex power series $\sum_{n=0}^{\infty} a_n z^n$, $z \in \mathbb{C}$, is called the *generating function* of (a_n). From the study of the generating function one can very often deduce useful information about the sequence. In particular, let (X, P) be a Markov chain. Its *Green function* or *Green kernel* is

$$G(x, y|z) = \sum_{n=0}^{\infty} p^{(n)}(x, y)\, z^n, \quad x, y \in X,\ z \in \mathbb{C}. \tag{1.32}$$

Its radius of convergence is

$$\mathsf{r}(x,y) = 1\Big/ \limsup_n \big(p^{(n)}(x,y)\big)^{1/n} \geq 1. \tag{1.33}$$

If $z = \mathsf{r}(x,y)$, the series may converge or diverge to $+\infty$. In both cases, we write $G\big(x,y|\mathsf{r}(x,y)\big)$ for the corresponding value. Observe that the number $G(x,y)$ defined in (1.27) coincides with $G(x,y|1)$.

Now let $\mathsf{r} = \inf\{\mathsf{r}(x,y) \mid x,y \in X\}$ and $|z| < r$. We can form the matrix

$$\mathcal{G}(z) = \big(G(x,y|z)\big)_{x,y\in X}.$$

We have seen that $p^{(n)}(x,y)$ is the element at position (x,y) of the matrix P^n, and that $P^0 = I$ is the identity matrix over X. We may therefore write

$$\mathcal{G}(z) = \sum_{n=0}^{\infty} z^n P^n,$$

where convergence of matrices is intended pointwise in each pair $(x,y) \in X^2$. The series converges, since $|z| < \mathsf{r}$. We have

$$\mathcal{G}(z) = I + \sum_{n=1}^{\infty} z^n P^n = I + zP \sum_{n=0}^{\infty} z^n P^n = I + zP\,\mathcal{G}(z).$$

In fact, by (1.20)

$$\begin{aligned} G(x,y|z) &= p^{(0)}(x,y) + \sum_{n=1}^{\infty} \sum_{w\in X} p(x,w)\, p^{(n-1)}(w,y)\, z^n \\ &\overset{(*)}{=} p^{(0)}(x,y) + \sum_{w\in X} z\, p(x,w) \sum_{n=1}^{\infty} p^{(n-1)}(w,y)\, z^{n-1} \\ &= p^{(0)}(x,y) + \sum_{w\in X} z\, p(x,w)\, G(w,y|z); \end{aligned}$$

the exchange of the sums in $(*)$ is legitimate because of the absolute convergence. Hence

$$(I - zP)\,\mathcal{G}(z) = I, \tag{1.34}$$

and we may formally write $\mathcal{G}(z) = (I - zP)^{-1}$. If X is finite, the involved matrices are finite dimensional, and this is the usual inverse matrix when $\det(I - zP) \neq 0$. Formal inversion is however not always justified. If X is infinite, one first has to specify on which (normed) linear space these matrices act and then to verify invertibility on that space.

In Example 1.1 we get

$$\mathcal{G}(z) = \begin{pmatrix} 1 & -z/2 & -z/2 \\ -z/4 & 1-z/2 & -z/4 \\ -z/4 & -z/4 & 1-z/2 \end{pmatrix}^{-1}$$
$$= \frac{1}{(1-z)(16-z^2)} \begin{pmatrix} (4-3z)(4-z) & 2z(4-z) & 2z(4-z) \\ z(4-z) & 2(8-4z-z^2) & 2z(2+z) \\ z(4-z) & 2z(2+z) & 2(8-4z-z^2) \end{pmatrix}.$$

We can now complete the answers to questions a) and b): regarding a),

$$G(\text{🌧}, \text{☼}|z) = \frac{z}{(1-z)(4+z)} = \frac{z}{5}\left(\frac{1}{1-z} + \frac{1}{4+z}\right)$$
$$= \sum_{n=1}^{\infty} \frac{1}{5}\left(1 - \frac{(-1)^n}{4^n}\right) z^n.$$

In particular,

$$p^{(14)}(\text{🌧}, \text{☼}) = \frac{1}{5}\left(1 - \frac{(-1)^{14}}{4^{14}}\right).$$

Analogously, to obtain b),

$$G(\text{🌧}, \text{🌧}|z) = \frac{2(8-4z-z^2)}{(1-z)(4-z)(4+z)}$$

has to be expanded in a power series. The sum of the coefficients of z^n, $n = 0, \ldots, 30$, gives the requested expected value.

1.35 Theorem. *If X is finite then $G(x, y|z)$ is a rational function in z.*

Proof. Recall that for a finite, invertible matrix $A = (a_{i,j})$,

$$A^{-1} = \frac{1}{\det(A)}(\hat{a}_{i,j}) \quad \text{with } \hat{a}_{i,j} = (-1)^{i-j} \det(A \mid j, i),$$

where $\det(A \mid j, i)$ is the determinant of the matrix obtained from A by deleting the j-th row and the i-th column. Hence

$$G(x, y|z) = \pm \frac{\det(I - zP \mid y, x)}{\det(I - zP)} \tag{1.36}$$

(with sign to be specified) is the quotient of two polynomials. □

Next, we consider the generating functions

$$F(x,y|z)=\sum_{n=0}^{\infty} f^{(n)}(x,y)\,z^n \quad \text{and} \quad U(x,y|z)=\sum_{n=0}^{\infty} u^{(n)}(x,y)\,z^n. \tag{1.37}$$

For $z=1$, we obtain the probabilities $F(x,y|1)=F(x,y)$ and $U(x,y|1)=U(x,y)$ introduced in the previous paragraph in (1.27). Note that $F(x,x|z)=F(x,x|0)=1$ and $U(x,y|0)=0$, and that $U(x,y|z)=F(x,y|z)$, if $x\neq y$. Indeed, among the U-functions we shall only need $U(x,x|z)$, which is the probability generating function of the first return time to x.

We denote by $\mathsf{s}(x,y)$ the radius of convergence of $U(x,y|z)$. Since $u^{(n)}(x,y)\le p^{(n)}(x,y)\le 1$, we must have

$$\mathsf{s}(x,y)\ge \mathsf{r}(x,y)\ge 1.$$

The following theorem will be useful on many occasions.

1.38 Theorem. (a) $G(x,x|z)=\dfrac{1}{1-U(x,x|z)}$, $|z|<\mathsf{r}(x,x)$.

(b) $G(x,y|z)=F(x,y|z)\,G(y,y|z)$, $|z|<\mathsf{r}(x,y)$.

(c) $U(x,x|z)=\sum_y p(x,y)z\,F(y,x|z)$, $|z|<\mathsf{s}(x,x)$.

(d) *If* $y\neq x$ *then* $F(x,y|z)=\sum_w p(x,w)z\,F(w,y|z)$, $|z|<\mathsf{s}(x,y)$.

Proof. (a) Let $n\ge 1$. If $Z_0=x$ and $Z_n=x$, then there must be an instant $k\in\{1,\dots,n\}$ such that $Z_k=x$, but $Z_j\neq x$ for $j=1,\dots,k-1$, that is, $\boldsymbol{t}^x=k$. The events

$$[\boldsymbol{t}^x=k]=[Z_k=x,\ Z_j\neq x \text{ for } j=1,\dots,k-1],\quad k=1,\dots,n,$$

are pairwise disjoint. Hence, using the Markov property (in its more general form of Exercise 1.22) and (1.19),

$$\begin{aligned}
p^{(n)}(x,x)&=\sum_{k=1}^{n}\mathsf{Pr}_x[Z_n=x,\ \boldsymbol{t}^x=k]\\
&=\sum_{k=1}^{n}\mathsf{Pr}_x[Z_n=x\mid Z_k=x,\ Z_j\neq x\\
&\qquad\qquad \text{for } j=1,\dots,k-1]\ \mathsf{Pr}_x[\boldsymbol{t}^x=k]\\
&\overset{(\times)}{=}\sum_{k=1}^{n}\mathsf{Pr}_x[Z_n=x\mid Z_k=x]\ \mathsf{Pr}_x[\boldsymbol{t}^x=k]\\
&=\sum_{k=1}^{n}p^{(n-k)}(x,x)\,u^{(k)}(x,x).
\end{aligned}$$

In (×), the careful observer may note that one could have $\mathsf{Pr}_x[Z_n = x \mid \boldsymbol{t}^x = k] \neq \mathsf{Pr}_x[Z_n = x \mid Z_k = x]$, but this can happen only when $\mathsf{Pr}_x[\boldsymbol{t}^x = k] = u^{(k)}(x,x) = 0$, so that possibly different first factors in those sums are compensated by multiplication with 0. Since $u^{(0)}(x,x) = 0$, we obtain

$$p^{(n)}(x,x) = \sum_{k=0}^{n} u^{(k)}(x,x)\, p^{(n-k)}(x,x) \quad \text{for } n \geq 1. \tag{1.39}$$

For $n = 0$ we have $p^{(0)}(x,x) = 1$, while $\sum_{k=0}^{n} u^{(k)}(x,x)\, p^{(n-k)}(x,x) = 0$. It follows that

$$\begin{aligned} G(x,x|z) &= \sum_{n=0}^{\infty} p^{(n)}(x,x)\, z^n = 1 + \sum_{n=1}^{\infty} \sum_{k=0}^{n} u^{(k)}(x,x)\, p^{(n-k)}(x,x)\, z^n \\ &= 1 + \sum_{n=0}^{\infty} \sum_{k=0}^{n} u^{(k)}(x,x)\, p^{(n-k)}(x,x)\, z^n \\ &= 1 + U(x,x|z)\, G(x,x|z) \end{aligned}$$

by Cauchy's product formula (Mertens' theorem) for the product of two series, as long as $|z| < \mathsf{r}(x,x)$, in which case both involved power series converge absolutely.

(b) If $x \neq y$, then $p^{(0)}(x,y) = 0$. Recall that $u^{(k)}(x,y) = f^{(k)}(x,y)$ in this case. Exactly as in (a), we obtain

$$p^{(n)}(x,y) = \sum_{k=1}^{n} f^{(k)}(x,y)\, p^{(n-k)}(y,y) = \sum_{k=0}^{n} f^{(k)}(x,y)\, p^{(n-k)}(y,y) \tag{1.40}$$

for *all* $n \geq 0$ (no exception when $n = 0$), and (b) follows again from the product formula for power series.

(d) Recall that $f^{(0)}(x,y) = 0$, since $y \neq x$. If $n \geq 1$ then the events

$$[\boldsymbol{s}^y = n,\ Z_1 = w], \quad w \in X,$$

are pairwise disjoint with union $[\boldsymbol{s}^y = n]$, whence

$$\begin{aligned} f^{(n)}(x,y) &= \sum_{w \in X} \mathsf{Pr}_x[\boldsymbol{s}^y = n,\ Z_1 = w] \\ &= \sum_{w \in X} \mathsf{Pr}_x[Z_1 = w]\ \mathsf{Pr}_x[\boldsymbol{s}^y = n \mid Z_1 = w] \\ &= \sum_{w \in X} \mathsf{Pr}_x[Z_1 = w]\ \mathsf{Pr}_w[\boldsymbol{s}^y = n-1] \\ &= \sum_{w \in X} p(x,w)\, f^{(n-1)}(w,y). \end{aligned}$$

(Note that in the last sum, we may get a contribution from $w = y$ only when $n = 1$, since otherwise $f^{(n-1)}(y, y) = 0$.) It follows that $\mathsf{s}(w, y) \geq \mathsf{s}(x, y)$ whenever $w \neq y$ and $p(x, w) > 0$. Therefore,

$$\begin{aligned} F(x, y|z) &= \sum_{n=1}^{\infty} f^{(n)}(x, y)\, z^n \\ &= \sum_{w \in X} p(x, w)\, z \sum_{n=1}^{\infty} f^{(n-1)}(w, y)\, z^{n-1} \\ &= \sum_{w \in X} p(x, w)\, z\, F(w, y|z). \end{aligned}$$

holds for $|z| < \mathsf{s}(x, y)$. □

1.41 Exercise. Prove formula (c) of Theorem 1.38. □

1.42 Definition. Let Γ be an oriented graph with vertex set X. For $x, y \in X$, a *cut point* between x and y (in this order!) is a vertex $w \in X$ such that every path in Γ from x to y must pass through w.

The following proposition will be useful on several occasions. Recall that $\mathsf{s}(x, y)$ is the radius of convergence of the power series $F(x, y|z)$.

1.43 Proposition. (a) *For all $x, w, y \in X$ and for real z with $0 \leq z \leq \mathsf{s}(x, y)$ one has*

$$F(x, y|z) \geq F(x, w|z)\, F(w, y|z).$$

(b) *Suppose that in the graph $\Gamma(P)$ of the Markov chain (X, P), the state w is a cut point between x and $y \in X$. Then*

$$F(x, y|z) = F(x, w|z)\, F(w, y|z)$$

for all $z \in \mathbb{C}$ with $|z| < \mathsf{s}(x, y)$ and for $z = \mathsf{s}(x, y)$.

Proof. (a) We have

$$\begin{aligned} f^{(n)}(x, y) &= \mathsf{Pr}_x[\boldsymbol{s}^y = n] \\ &\geq \mathsf{Pr}_x[\boldsymbol{s}^y = n,\ \boldsymbol{s}^w \leq n] \\ &= \sum_{k=0}^{n} \mathsf{Pr}_x[\boldsymbol{s}^y = n,\ \boldsymbol{s}^w = k] \\ &= \sum_{k=0}^{n} \mathsf{Pr}_x[\boldsymbol{s}^w = k]\ \mathsf{Pr}_x[\boldsymbol{s}^y = n \mid \boldsymbol{s}^w = k] \\ &= \sum_{k=0}^{n} f^{(k)}(x, w)\, f^{(n-k)}(w, y). \end{aligned}$$

The inequality of statement (a) is true when $F(x, w|\cdot) \equiv 0$ or $F(w, y|\cdot) \equiv 0$. So let us suppose that there is k such that $f^{(k)}(x, w) > 0$. Then $f^{(n-k)}(w, y) \le f^{(n)}(x, y)/f^{(k)}(x, w)$ for all $n \ge k$, whence $\mathsf{s}(w, y) \ge \mathsf{s}(x, y)$. In the same way, we may suppose that $f^{(l)}(w, y) > 0$ for some l, which implies $\mathsf{s}(x, w) \ge \mathsf{s}(x, y)$. Then

$$f^{(n)}(x, y)\, z^n \ge \sum_{k=0}^{n} f^{(k)}(x, w)\, z^k \;\; f^{(n-k)}(w, y)\, z^{n-k}$$

for all real z with $0 \le z \le \mathsf{s}(x, y)$, and the product formula for power series implies the result for all those z with $z < \mathsf{s}(x, y)$. Since we have power series with non-negative coefficients, we can let $z \to \mathsf{s}(x, y)$ from below to see that statement (a) also holds for $z = \mathsf{s}(x, y)$, regardless of whether the series converge or diverge at that point.

(b) If w is a cut point between x and y, then the Markov chain must visit w before it can reach y. That is, $\boldsymbol{s}^w \le \boldsymbol{s}^y$, given that $Z_0 = x$. Therefore the strong Markov property yields

$$\begin{aligned} f^{(n)}(x, y) &= \mathsf{Pr}_x[\boldsymbol{s}^y = n] = \mathsf{Pr}_x[\boldsymbol{s}^y = n,\ \boldsymbol{s}^w \le n] \\ &= \sum_{k=0}^{n} f^{(k)}(x, w)\, f^{(n)}(w, y). \end{aligned}$$

We can now argue precisely as in the proof of (a), and the product formula for power series yields statement (b) for all $z \in \mathbb{C}$ with $|z| < \mathsf{s}(x, y)$ as well as for $z = \mathsf{s}(x, y)$. □

1.44 Exercise. Show that for distinct $x, y \in X$ and for real z with $0 \le z \le \mathsf{s}(x, x)$ one has

$$U(x, x|z) \ge F(x, y|z)\, F(y, x|z). \qquad \square$$

1.45 Exercise. Suppose that w is a cut point between x and y. Show that the expected time to reach y starting from x (given that y is reached) satisfies

$$\mathsf{E}_x(\boldsymbol{s}^y \mid \boldsymbol{s}^y < \infty) = \mathsf{E}_x(\boldsymbol{s}^w \mid \boldsymbol{s}^w < \infty) + \mathsf{E}_w(\boldsymbol{s}^y \mid \boldsymbol{s}^y < \infty).$$

[Hint: check first that $\mathsf{E}_x(\boldsymbol{s}^y \mid \boldsymbol{s}^y < \infty) = F'(x, y|1-)/F(x, y|1)$, and apply Proposition 1.43 (b).] □

We can now answer the questions of Example 1.2. The latter is a specific case of the following type of Markov chain.

1.46 Example. The *random walk with two absorbing barriers* is the Markov chain with state space

$$X = \{0, 1, \ldots, N\}, \quad N \in \mathbb{N}$$

and transition probabilities

$$p(0,0) = p(N,N) = 1,$$
$$p(i,i+1) = p, \quad p(i,i-1) = q \quad \text{for } i = 1,\dots,N-1, \quad \text{and}$$
$$p(i,j) = 0 \quad \text{in all other cases.}$$

Here $0 < p < 1$ and $q = 1 - p$. This example is also well known as the model of *the gambler's ruin.*

The probabilities to reach 0 and N starting from the point j are $F(j,0)$ and $F(j,N)$, respectively. The expected number of steps needed for reaching N from j, under the condition that N is indeed reached, is

$$\mathsf{E}_j(s^N \mid s^N < \infty) = \sum_{n=1}^{\infty} n \ \mathsf{Pr}_j[s^N = n \mid s^N < \infty] \frac{F'(j,N|1-)}{F(j,N|1)}.$$

where $F'(j,N|z)$ denotes the derivative of $F(j,N|z)$ with respect to z, compare with Exercise 1.45.

For the random walk with two absorbing barriers, we have

$$F(0,N|z) = 0, \ F(N,N|z) = 1, \quad \text{and}$$
$$F(j,N|z) = qz\,F(j-1,N|z) + pz\,F(j+1,N|z), \quad j = 1,\dots,N-1.$$

The last identity follows from Theorem 1.38 (d). For determining $F(j,N|z)$ as a function of j, we are thus lead to a linear difference equation of second order with constant coefficients. The associated characteristic polynomial

$$pz\,\lambda^2 - \lambda + qz$$

has roots

$$\lambda_1(z) = \frac{1}{2pz}\left(1 - \sqrt{1-4pqz^2}\right) \quad \text{and} \quad \lambda_2(z) = \frac{1}{2pz}\left(1 + \sqrt{1-4pqz^2}\right).$$

We study the case $|z| < 1/2\sqrt{pq}$: the roots are distinct, and

$$F(j,N|z) = a \cdot \lambda_1(z)^j + b \cdot \lambda_2(z)^j.$$

The constants $a = a(z)$ and $b = b(z)$ are found by inserting the boundary values at $j = 0$ and $j = N$:

$$a + b = 0 \quad \text{and} \quad a \cdot \lambda_1(z)^N + b \cdot \lambda_2(z)^N = 1.$$

The result is

$$F(j,N|z) = \frac{\lambda_2(z)^j - \lambda_1(z)^j}{\lambda_2(z)^N - \lambda_1(z)^N}, \quad |z| < \frac{1}{2\sqrt{pq}}.$$

With some effort, one computes for $|z| < 1/2\sqrt{pq}$

$$\frac{F'(j,N|z)}{F(j,N|z)} = \frac{1}{z\sqrt{1-4pqz^2}}\left(N\,\frac{\alpha(z)^N+1}{\alpha(z)^N-1} - j\,\frac{\alpha(z)^j+1}{\alpha(z)^j-1}\right),$$

where

$$\alpha(z) = \lambda_1(z)/\lambda_2(z).$$

We have $\alpha(1) = \max\{p/q,\ q/p\}$. Hence, if $p \neq 1/2$, then $1 < 1/(2\sqrt{pq})$, and

$$\mathsf{E}_j(\boldsymbol{s}^N \mid \boldsymbol{s}^N < \infty) = \frac{1}{q-p}\left(N\,\frac{(q/p)^N+1}{(q/p)^N-1} - j\,\frac{(q/p)^j+1}{(q/p)^j-1}\right).$$

When $p = 1/2$, one easily verifies that

$$F(j,N|1) = j/N.$$

Instead of $F'(j,N|1)/F(j,N|1)$ one has to compute

$$\lim_{z\to 1-} F'(j,N|z)/F(j,N|z)$$

in this case. We leave the corresponding calculations as an exercise.

In Example 1.2 we have $N = 300$, $p = 2/3$, $q = 1/3$, $\alpha = 1/2$ and $j = 200$. We obtain

$$\mathsf{Pr}_{100}[\boldsymbol{s}^{300} < \infty] = \frac{1-2^{-200}}{1-2^{-300}} \approx 1 \quad \text{and}$$
$$\mathsf{E}_{100}(\boldsymbol{s}^{300} \mid \boldsymbol{s}^{300} < \infty) \approx 600.$$

The probability that the drunkard does not return home is practically 0. For returning home, it takes him 600 steps on the average, that is, 3 times as many as necessary. □

1.47 Exercise. (a) If $p = 1/2$ instead of $p = 2/3$, compute the expected time (number of steps) that the drunkard needs to return home, supposing that he does return.

(b) Modify the model: if the drunkard falls into the lake, he does not drown, but goes back to the previous point on the road at the next step. That is, the lake has become a *reflecting barrier,* and $p(0,0) = 0$, $p(0,1) = 1$, while all other transition probabilities remain unchanged. (In particular, the house remains an absorbing state with $p(N,N) = 1$.) Redo the computations in this case. □

1.48 Exercise. Let (X, P) be an arbitrary Markov chain, and define a new transition matrix $P_a = a \cdot I + (1-a) \cdot P$, where $0 < a < 1$. Let $G(\cdot, \cdot|z)$ and $G_a(\cdot, \cdot|z)$ denote the Green functions of P and P_a, respectively. Show that for all $x, y \in X$

$$G_a(x, y|z) = \frac{1}{1-az}\, G\Big(x, y\Big|\frac{z-az}{1-az}\Big). \qquad \square$$

Further examples that illustrate the use of generating functions will follow later on.

Paths and their weights

We conclude this section with a purely combinatorial description of several of the probabilities and generating functions that we have considered so far. Recall the Definition 1.6 of the (oriented) graph $\Gamma(P)$ of the Markov chain (X, P). A (finite) *path* is a sequence $\pi = [x_0, x_1, \dots, x_n]$ of vertices (states) such that $[x_{i-1}, x_i]$ is an edge, that is, $p(x_{i-1}, x_i) > 0$ for $i = 1, \dots, n$. Here, $n \geq 0$ is the *length* of π, and π is a path *from* x_0 *to* x_n. If $n = 0$ then $\pi = [x_0]$ consists of a single point. The *weight* of π with respect to $z \in \mathbb{C}$ is

$$\mathsf{w}(\pi|z) = \begin{cases} 1, & \text{if } n = 0, \\ p(x_0, x_1)\, p(x_1, x_2) \cdots p(x_{n-1}, x_n)\, z^n, & \text{if } n \geq 1, \end{cases}$$

and

$$\mathsf{w}(\pi) = \mathsf{w}(\pi|1).$$

If Π is a set of paths, then $\Pi^{(n)}$ denotes the set of all $\pi \in \Pi$ with length n. We can consider $\mathsf{w}(\cdot|z)$ as a complex-valued measure on the set of all paths. The weight of Π is

$$\mathsf{w}(\Pi|z) = \sum_{\pi \in \Pi} \mathsf{w}(\pi|z), \quad \text{and} \quad \mathsf{w}(\Pi) = \mathsf{w}(\Pi|1), \tag{1.49}$$

given that the involved sum converges absolutely. Thus,

$$\mathsf{w}(\Pi|z) = \sum_{n=0}^{\infty} \mathsf{w}(\Pi^{(n)})\, z^n.$$

Now let $\Pi(x, y)$ be the set of all paths from x to y. Then

$$p^{(n)}(x, y) = \mathsf{w}\big(\Pi^{(n)}(x, y)\big) \quad \text{and} \quad G(x, y|z) = \mathsf{w}\big(\Pi(x, y)|z\big). \tag{1.50}$$

Next, let $\Pi_\circ(x, y)$ be the set of paths from x to y which contain y only as their final point. Then

$$f^{(n)}(x, y) = \mathsf{w}\big(\Pi_\circ^{(n)}(x, y)\big) \quad \text{and} \quad F(x, y|z) = \mathsf{w}\big(\Pi_\circ(x, y)|z\big). \tag{1.51}$$

Similarly, let $\Pi_\bullet(x,y)$ be the set of all paths $[x = x_0, x_1, \dots, x_n = y]$ for which $n \geq 1$ and $x_i \neq y$ for $i = 1, \dots, n-1$. Thus, $\Pi_\bullet(x,y) = \Pi_\circ(x,y)$ when $x \neq y$, and $\Pi_\bullet(x,x)$ is the set of all paths that return to the initial point x only at the end. We get

$$u^{(n)}(x,y) = \mathsf{w}\big(\Pi_\bullet^{(n)}(x,y)\big) \quad \text{and} \quad U(x,y|z) = \mathsf{w}\big(\Pi_\bullet(x,y)|z\big). \tag{1.52}$$

If $\pi_1 = [x_0, \dots, x_m]$ and $\pi_2 = [y_0, \dots, y_n]$ are two paths with $x_m = y_0$, then we can define their *concatenation* as

$$\pi_1 \circ \pi_2 = [x_0, \dots, x_m = y_0, \dots, y_n].$$

We then have

$$\mathsf{w}(\pi_1 \circ \pi_2|z) = \mathsf{w}(\pi_1|z)\,\mathsf{w}(\pi_2|z). \tag{1.53}$$

If $\Pi_1(x,w)$ is a set of paths from x to w and $\Pi_2(w,y)$ is a set of paths from w to y, then we set

$$\Pi_1(x,w) \circ \Pi_2(w,y) = \{\pi_1 \circ \pi_2 : \pi_1 \in \Pi_1(x,w),\ \pi_2 \in \Pi_2(w,y)\}.$$

Thus

$$\mathsf{w}\big(\Pi_1(x,w) \circ \Pi_2(w,y)\big|z\big) = \mathsf{w}\big(\Pi_1(x,w)\big|z\big)\,\mathsf{w}\big(\Pi_2(w,y)\big|z\big). \tag{1.54}$$

Many of the identities for transition probabilities and generating functions can be derived in terms of weights of paths and their concatenation. For example, the obvious relation

$$\Pi^{(m+n)}(x,y) = \biguplus_w \Pi^{(m)}(x,w) \circ \Pi^{(n)}(w,y)$$

(disjoint union) leads to

$$\begin{aligned} p^{(m+n)}(x,y) &= \mathsf{w}\big(\Pi^{(m+n)}(x,y)\big) \\ &= \sum_w \mathsf{w}\big(\Pi^{(m)}(x,w) \circ \Pi^{(n)}(w,y)\big) = \sum_w p^{(m)}(x,w)\, p^{(n)}(w,y). \end{aligned}$$

1.55 Exercise. Show that

$$\Pi(x,x) = \{[x]\} \uplus \big(\Pi_\bullet(x,x) \circ \Pi(x,x)\big) \quad \text{and} \quad \Pi(x,y) = \Pi_o(x,y) \circ \Pi(y,y),$$

and deduce statements (a) and (b) of Theorem 1.38.

Find analogous proofs in terms of concatenation and weights of paths for statements (c) and (d) of Theorem 1.38. □

Chapter 2
Irreducible classes

A Irreducible and essential classes

In the sequel, (X, P) will be a Markov chain. For $x, y \in X$ we write

a) $x \xrightarrow{n} y$, if $p^{(n)}(x, y) > 0$,

b) $x \to y$, if there is $n \geq 0$ such that $x \xrightarrow{n} y$,

c) $x \not\to y$, if there is no $n \geq 0$ such that $x \xrightarrow{n} y$,

d) $x \leftrightarrow y$, if $x \to y$ and $y \to x$.

These are all properties of the graph $\Gamma(P)$ of the Markov chain that do not depend on the specific values of the weights $p(x, y) > 0$. In the graph, $x \xrightarrow{n} y$ means that there is a path (walk) of length n from x to y. If $x \leftrightarrow y$, we say that the states x and y *communicate.*

The relation $\to$ is reflexive and transitive. In fact $p^{(0)}(x, x) = 1$ by definition, and if $x \xrightarrow{m} w$ and $w \xrightarrow{n} y$ then $x \xrightarrow{m+n} y$. This can be seen by concatenating paths in the graph of the Markov chain, or directly by the inequality $p^{(m+n)}(x, y) \geq p^{(m)}(x, w)p^{(n)}(w, y) > 0$. Therefore, we have the following.

2.1 Lemma. *$\leftrightarrow$ is an equivalence relation on X.*

2.2 Definition. An *irreducible class* is an equivalence class with respect to $\leftrightarrow$.

In graph theoretical terminology, one also speaks of a *strongly connected component.*

In Examples 1.1, 1.3 and 1.4, all elements communicate, and there is a unique irreducible class. In this case, the Markov chain itself is called *irreducible.*

In Example 1.46 (and its specific case 1.2), there are 3 irreducible classes: $\{0\}$, $\{1, \ldots, N-1\}$ and $\{N\}$.

2.3 Example. Let $X = \{1, 2, , \ldots, 13\}$ and the graph $\Gamma(P)$ be as in Figure 4. (The oriented edges correspond to non-zero one-step transition probabilities.) The irreducible classes are $C(1) = \{1, 2\}$, $C(3) = \{3\}$, $C(4) = \{4, 5, 6\}$, $C(7) = \{7, 8, 9, 10\}$, $C(11) = \{11, 12\}$ and $C(13) = \{13\}$.

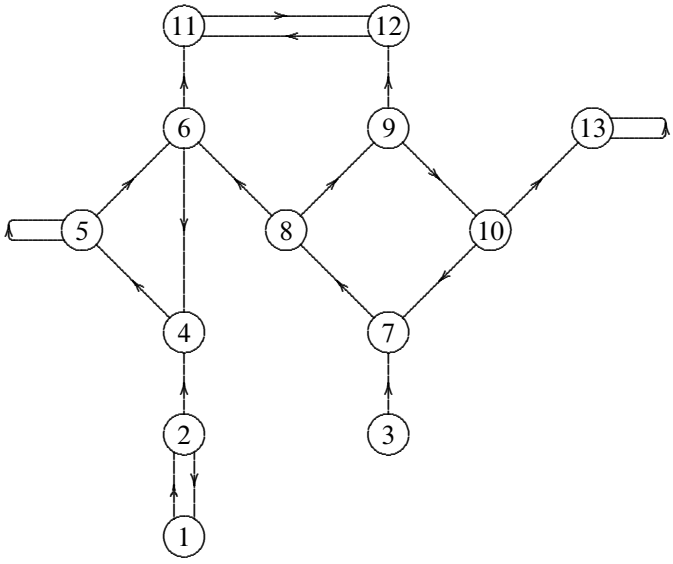

Figure 4

On the irreducible classes, the relation $\to$ becomes a (partial) order: we define

$$C(x) \to C(y) \quad \text{if and only if} \quad x \to y.$$

It is easy to verify that this order is well defined, that is, independent of the specific choice of representatives of the single irreducible classes.

2.4 Lemma. *The relation $\to$ is a partial order on the collection of all irreducible classes of (X, P).*

Proof. Reflexivity: since $x \xrightarrow{0} x$, we have $C(x) \to C(x)$.

Transitivity: if $C(x) \to C(w) \to C(y)$ then $x \to w \to y$. Hence $x \to y$, and $C(x) \to C(y)$.

Anti-symmetry: if $C(x) \to C(y) \to C(x)$ then $x \to y \to x$ and thus $x \leftrightarrow y$, so that $C(x) = C(y)$. □

In Figure 5, we illustrate the partial ordering of the irreducible classes of Example 2.3:

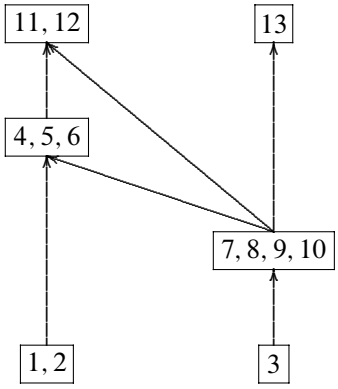

Figure 5

The graph associated with the partial order as in Figure 5 is in general not the graph of a (factor) Markov chain obtained from (X, P). Figure 5 tells us, for example, that from any point in the class $C(7)$ it is possible to reach $C(4)$ with positive probability, but not conversely.

2.5 Definition. The maximal elements (if they exist) of the partial order $\to$ on the collection of the irreducible classes of (X, P) are called *essential classes* (or *absorbing classes*).[1]

A state x is called essential, if $C(x)$ is an essential class.

In Example 2.3, the essential classes are $\{11, 12\}$ and $\{13\}$.

Once it has entered an essential class, the Markov chain (Z_n) cannot exit from it:

2.6 Exercise. Let $C \subset X$ be an irreducible class. Prove that the following statements are equivalent.

(a) C is essential.

(b) If $x \in C$ and $x \to y$ then $y \in C$.

(c) If $x \in C$ and $x \to y$ then $y \to x$. □

If X is finite then there are only finitely many irreducible classes, so that in their partial order, there must be maximal elements. Thus, from each element

[1] In the literature, one sometimes finds the expressions "recurrent" or "ergodic" in the place of "essential", and "transient" in the place of "non-essential". This is justified when the state space is finite. We shall avoid these identifications, since in general, "recurrent" and "transient" have another meaning, see Chapter 3.

it is possible to reach an essential class. If X is infinite, this is no longer true. Figure 6 shows the graph of a Markov chain on $X = \mathbb{N}$ has no essential classes; the irreducible classes are $\{1, 2\}$, $\{3, 4\}$, ...

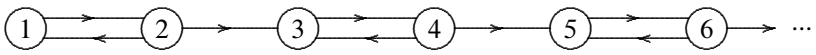

Figure 6

If an essential class contains exactly one point x (which happens if and only if $p(x, x) = 1$), then x is called an *absorbing state.* In Example 1.2, the lake and the drunkard's house are absorbing states.

We call a set $B \subset X$ *convex*, if $x, y \in B$ and $x \to w \to y$ implies $w \in B$. Thus, if $x \in B$ and $w \leftrightarrow x$, then also $w \in B$. In particular, B is a union of irreducible classes.

2.7 Theorem. *Let $B \subset X$ be a finite, convex set that does not contain essential elements. Then there is $\varepsilon > 0$ such that for each $x \in B$ and all but finitely many $n \in \mathbb{N}$,*

$$\sum_{y \in B} p^{(n)}(x, y) \leq (1 - \varepsilon)^n.$$

Proof. By assumption, B is a disjoint union of finite, non-essential irreducible classes $C(x_1), \ldots, C(x_k)$. Let $C(x_1), \ldots, C(x_j)$ be the maximal elements in the partial order $\to$, restricted to $\{C(x_1), \ldots, C(x_k)\}$, and let $i \in \{1, \ldots, j\}$. Since $C(x_i)$ is non-essential, there is $v_i \in X$ such that $x_i \to v_i$ but $v_i \not\to x_i$. By the maximality of $C(x_i)$, we must have $v_i \in X \setminus B$. If $x \in B$ then $x \to x_i$ for some $i \in \{1, \ldots j\}$, and hence also $x \to v_i$, while $v_i \not\to x$. Therefore there is $m_x \in \mathbb{N}$ such that

$$\sum_{y \in B} p^{(m_x)}(x, y) < 1.$$

Let $m = \max\{m_x : x \in B\}$. Choose $x \in B$. Write $m = m_x + \ell_x$, where $\ell_x \in \mathbb{N}_0$. We obtain

$$\begin{aligned}
\sum_{y \in B} p^{(m)}(x, y) &= \sum_{y \in B} \sum_{w \in X} \underbrace{p^{(m_x)}(x, w)\, p^{(\ell_x)}(w, y)}_{\substack{> 0 \text{ only if } w \in B \\ \text{(by convexity of } B)}} \\
&= \sum_{w \in B} p^{(m_x)}(x, w) \underbrace{\sum_{y \in B} p^{(\ell_x)}(w, y)}_{\leq 1} \\
&\leq \sum_{w \in B} p^{(m_x)}(x, w) < 1.
\end{aligned}$$

Since B is finite, there is $\kappa > 0$ such that

$$\sum_{y\in B} p^{(m)}(x,y) \le 1-\kappa \quad \text{for all } x \in B.$$

Let $n \in \mathbb{N}$, $n \ge 2$. Write $n = km + r$ with $0 \le r < m$, and assume that $k \ge 1$ (that is, $n \ge m$). For $x \in B$,

$$\begin{aligned}
\sum_{y\in B} p^{(n)}(x,y) &= \sum_{y\in B}\sum_{w\in X} \underbrace{p^{(km)}(x,w)\,p^{(r)}(w,y)}_{>0 \text{ only if } w\in B} \\
&= \sum_{w\in B} p^{(km)}(x,w) \underbrace{\sum_{y\in B} p^{(r)}(w,y)}_{\le 1} \\
&\le \sum_{w\in B} p^{(km)}(x,w) && \text{(as above)} \\
&= \sum_{y\in B} p^{((k-1)m)}(x,y) \underbrace{\sum_{w\in B} p^{(m)}(y,w)}_{\le 1-\kappa} \\
&\le (1-\kappa)\sum_{y\in B} p^{((k-1)m)}(x,y) && \text{(inductively)} \\
&\le (1-\kappa)^k = \bigl((1-\kappa)^{k/n}\bigr)^n && \text{(since } k/n \ge 1/(2m)\text{)} \\
&\le \bigl((1-\kappa)^{1/2m}\bigr)^n = (1-\varepsilon)^n,
\end{aligned}$$

where $\varepsilon = 1-(1-\kappa)^{1/2m}$. □

In particular, let C be a finite, non-essential irreducible class. The theorem says that for the Markov chain starting at $x \in C$, the probability to remain in C for n steps decreases exponentially as $n \to \infty$. In particular, the expected number of visits in C starting from $x \in C$ is finite. Indeed, this number is computed as

$$\mathsf{E}_x(\boldsymbol{\nu}^C) = \sum_{n=0}^{\infty}\sum_{y\in C} p^{(n)}(x,y) \le 1/\varepsilon,$$

see (1.23). We deduce that $\boldsymbol{\nu}^C < \infty$ almost surely.

2.8 Corollary. *If C is a finite, non-essential irreducible class then*

$$\mathsf{Pr}_x[\exists\, k : Z_n \notin C \text{ for all } n > k] = 1.$$

2.9 Corollary. *If the set of all non-essential states in X is finite, then the Markov chain reaches some essential class with probability one:*

$$\mathsf{Pr}_x[\boldsymbol{s}^{X_{\mathrm{ess}}} < \infty] = 1,$$

where X_{ess} is the union of all essential classes.

Proof. The set $B = X \setminus X_{\mathrm{ess}}$ is a finite union of finite, non-essential classes (indeed, a convex set of non-essential elements). Therefore

$$\mathsf{Pr}_x[\boldsymbol{s}^{X_{\mathrm{ess}}} < \infty] = \mathsf{Pr}_x[\exists\, k : Z_n \notin B \text{ for all } n > k] = 1$$

by the same argument that lead to Corollary 2.8. □

The last corollary does not remain valid when $X \setminus X_{\mathrm{ess}}$ is infinite, as the following example shows.

2.10 Example (Infinite drunkard's walk with one absorbing barrier). The state space is $X = \mathbb{N}_0$. We choose parameters $p, q > 0$, $p + q = 1$, and set

$$p(0,0) = 1, \quad p(k, k+1) = p, \quad p(k, k-1) = q \quad (k > 0),$$

while $p(k, \ell) = 0$ in all other cases.

The state 0 is absorbing, while all other states belong to one non-essential, infinite irreducible class $C = \mathbb{N}$. We observe that

$$F(k+1, k|z) = F(1, 0|z) \quad \text{for each } k \geq 0. \tag{2.11}$$

This can be justified formally by (1.51), since the mapping $n \mapsto n + k$ induces a bijection from $\Pi_\circ(1,0)$ to $\Pi_\circ(k+1,k)$: a path $[1 = x_0, x_1, \ldots, x_l = 0]$ in $\Pi_\circ(1,0)$ is mapped to the path $[k+1 = x_0 + k, x_1 + k, \ldots, x_l + k = k]$. The bijection preserves all single weights (transition probabilities), so that $\mathsf{w}\big(\Pi_\circ(1,0)|z\big) = \mathsf{w}\big(\Pi_\circ(k+1,k)|z\big)$. Note that this is true because the parts of the graph of our Markov chain that are "on the right" of k and "on the right" of 0 are isomorphic as weighted graphs. (Here we cannot apply the same argument to $\Pi(k+1,k)$ in the place of $\Pi_\circ(k+1,k)$!)

2.12 Exercise. Formulate criteria of "isomorphism", resp. "restricted isomorphism" that guarantee $G(x, y|z) = G(x', y'|z)$, resp. $F(x, y|z) = F(x', y'|z)$ for a general Markov chain (X, P) and points $x, y, x', y' \in X$. □

Returning to the drunkard's fate, by Theorem 1.38 (c)

$$F(1, 0|z) = qz + pz\, F(2, 0|z). \tag{2.13}$$

In order to reach state 0 starting from state 2, the random walk must necessarily pass through state 1: the latter is a cut point between 2 and 0, and Proposition 1.43 (b) together with (2.11) yields

$$F(2,0|z) = F(2,1|z)\,F(1,0|z) = F(1,0|z)^2.$$

We substitute this identity in (2.13) and obtain

$$pz\,F(1,0|z)^2 - F(1,0|z) + qz = 0.$$

The two solutions of this equation are

$$\frac{1}{2pz}\left(1 \pm \sqrt{1-4pqz^2}\right).$$

We must have $F(1,0|0) = 0$, and in the interior of the circle of convergence of this power series, the function must be continuous. It follows that of the two solutions, the correct one is

$$F(1,0|z) = \frac{1}{2pz}\left(1 - \sqrt{1-4pqz^2}\right),$$

whence $F(1,0) = \min\{1, q/p\}$. In particular, if $p > q$ then $F(1,0) < 1$: starting at state 1, the probability that (Z_n) never reaches the unique absorbing state 0 is $1 - F(1,0) > 0$.

We can reformulate Theorem 2.7 in another way:

2.14 Definition. For any subset A of X, we denote by P_A the *restriction* of P to A:

$$p_A(x,y) = p(x,y), \quad \text{if } x, y \in A, \text{ and } p_A(x,y) = 0, \quad \text{otherwise.}$$

We consider P_A as a matrix over the whole of X, but the same notation will be used for the truncated matrix over the set A. It is not necessarily stochastic, but always *substochastic:* all row sums are ≤ 1.

The matrix P_A describes the evolution of the Markov chain constrained to staying in A, and the (x,y)-element of the matrix power P_A^n is

$$p_A^{(n)}(x,y) = \mathsf{Pr}_x[Z_n = y,\ Z_k \in A\ (0 \le k \le n)]. \tag{2.15}$$

In particular, $P_A^0 = I_A$, the restriction of the identity matrix to A. For the associated Green function we write

$$G_A(x,y|z) = \sum_{n=0}^{\infty} p_A^{(n)}(x,y)\,z^n, \quad G_A(x,y) = G_A(x,y|1). \tag{2.16}$$

(Caution: this is *not* the restriction of $G(\cdot,\cdot|z)$ to A !) Let $\mathsf{r}_A(x,y)$ be the radius of convergence of this power series, and $\mathsf{r}_A = \inf\{\mathsf{r}_A(x,y) : x,y \in A\}$. If we write $\mathcal{G}_A(z) = \big(G_A(x,y|z)\big)_{x,y\in A}$ then

$$(I_A - zP_A)\mathcal{G}_A(z) = I_A \quad \text{for all } z \in \mathbb{C} \text{ with } |z| < \mathsf{r}_A. \tag{2.17}$$

2.18 Lemma. *Suppose that $A \subset X$ is finite and that for each $x \in A$ there is $w \in X \setminus A$ such that $x \to w$. Then $\mathsf{r}_A > 1$. In particular, $G_A(x,y) < \infty$ for all $x, y \in A$.*

Proof. We introduce a new state † and equip the state space $A \cup \{\dagger\}$ with the transition matrix Q given by

$$q(x,y) = p(x,y), \quad q(x,\dagger) = 1 - p(x,A),$$
$$q(\dagger,\dagger) = 1, \quad \text{and} \quad q(\dagger,x) = 0, \quad \text{if } x,y \in A.$$

Then $Q_A = P_A$, the only essential state of the Markov chain $(A \cup \{\dagger\}, Q)$ is †, and A is convex. We can apply Theorem 2.7 and get $\mathsf{r}_A \geq 1/(1-\varepsilon)$, where $\sum_{y\in A} p_A^{(n)}(x,y) \leq (1-\varepsilon)^n$ for all $x \in A$. □

B The period of an irreducible class

In this section we consider an irreducible class C of a Markov chain (X,P). We exclude the trivial case when C consists of a single point x with $p(x,x) = 0$ (in this case, we say that x is a *ephemeral* state). In order to study the behaviour of (Z_n) inside C, it is sufficient to consider the restriction P_C to the set C of the transition matrix according to Definition 2.14,

$$P_C = \big(p_C(x,y)\big)_{x,y\in C}, \quad \text{where } p_C(x,y) = \begin{cases} p(x,y), & \text{if } x,y \in C, \\ 0, & \text{otherwise.} \end{cases}$$

Indeed, for $x, y \in C$ the probability $p^{(n)}(x,y)$ coincides with the element $p_C^{(n)}(x,y)$ of the n-th matrix power of P_C: this assertion is true for $n = 1$, and inductively

$$p^{(n+1)}(x,y) = \sum_{w\in X} \underbrace{p(x,w)\, p^{(n)}(w,y)}_{=0 \text{ if } w\notin C} = \sum_{w\in C} p(x,w)\, p^{(n)}(w,y)$$
$$= \sum_{w\in C} p_C(x,w)\, p_C^{(n)}(w,y) = p_C^{(n+1)}(x,y).$$

Obviously, P_C is stochastic if and only if the irreducible class C is essential.

By definition, C has the following property.

For each $x, y \in C$ there is $k = k(x, y)$ such that $p^{(k)}(x, y) > 0$.

2.19 Definition. The *period* of C is the number

$$d = d(C) = \gcd(\{n > 0 : p^{(n)}(x, x) > 0\}),$$

where $x \in C$.

Here, gcd is of course the greatest common divisor. We have to check that $d(C)$ is well defined.

2.20 Lemma. *The number $d(C)$ does not depend on the specific choice of $x \in C$.*

Proof. Let $x, y \in C$, $x \neq y$. We write $d(x) = \gcd(\mathbb{N}_x)$, where

$$\mathbb{N}_x = \{n > 0 : p^{(n)}(x, x) > 0\} \tag{2.21}$$

and analogously $\mathbb{N}_y$ and $d(y)$. By irreducibility of C there are $k, \ell > 0$ such that $p^{(k)}(x, y) > 0$ and $p^{(\ell)}(y, x) > 0$. We have $p^{(k+\ell)}(x, x) > 0$, and hence $d(x)$ divides $k + \ell$.

Let $n \in \mathbb{N}_y$. Then $p^{(k+n+\ell)}(x, x) \geq p^{(k)}(x, y)\, p^{(n)}(y, y)\, p^{(\ell)}(y, x) > 0$, whence $d(x)$ divides $k + n + \ell$.

Combining these observations, we see that $d(x)$ divides each $n \in \mathbb{N}_y$. We conclude that $d(x)$ divides $d(y)$. By symmetry of the roles of x and y, also $d(y)$ divides $d(x)$, and thus $d(x) = d(y)$. □

In Example 1.1, $C = X = \{☼, 🌧, ❄\}$ is one irreducible class, $p(🌧, 🌧) > 0$, and hence $d(C) = d(X) = 1$.

In Example 1.46 (random walk with two absorbing barriers), the set $C = \{1, 2, \ldots, N - 1\}$ is an irreducible class with period $d(C) = 2$.

In Example 2.3, the state 3 is ephemeral, and the periods of the other classes are $d(\{1, 2\}) = 2$, $d(\{4, 5, 6\}) = 1$, $d(\{7, 8, 9, 10\}) = 4$, $d(\{11, 12\}) = 2$ and $d(\{13\}) = 1$.

If $d(C) = 1$ then C is called an *aperiodic class*. In particular, C is aperiodic when $p(x, x) > 0$ for some $x \in C$.

2.22 Lemma. *Let C be an irreducible class and $d = d(C)$. For each $x \in C$ there is $m_x \in \mathbb{N}$ such that $p^{(md)}(x, x) > 0$ for all $m \geq m_x$.*

Proof. First observe that the set $\mathbb{N}_x$ of (2.21) has the following property.

$$n_1, n_2 \in \mathbb{N}_x \Longrightarrow n_1 + n_2 \in \mathbb{N}_x. \tag{2.23}$$

(It is a semigroup.) It is well known from elementary number theory (and easy to prove) that the greatest common divisor of a set of positive integers can always be written as a finite linear combination of elements of the set with integer coefficients. Thus, there are $n_1, \dots, n_\ell \in \mathbb{N}_x$ and $a_1, \dots, a_\ell \in \mathbb{Z}$ such that

$$d = \sum_{i=1}^{\ell} a_i\, n_i .$$

Let

$$n^+ = \sum_{i:a_i>0} a_i\, n_i \quad \text{and} \quad n^- = \sum_{i:a_i<0} (-a_i)\, n_i .$$

Now $n^+, n^- \in \mathbb{N}_x$ by (2.23), and $d = n^+ - n^-$. We set $k^+ = n^+/d$ and $k^- = n^-/d$. Then $k^+ - k^- = 1$. We define

$$m_x = k^-(k^- - 1).$$

Let $m \geq m_x$. We can write $m = q\,k^- + r$, where $q \geq k^- - 1$ and $0 \leq r \leq k^- - 1$. Hence, $m = q\,k^- + r(k^+ - k^-) = (q-r)k^- + r\,k^+$ with $(q-r) \geq 0$ and $r \geq 0$, and by (2.23)

$$m\,d = (q-r)m^- + r\,m^+ \in \mathbb{N}_x. \qquad \square$$

Before stating the main result of this section, we observe that the relation $\to$ on the elements of X and the definition of irreducible classes do not depend on the fact that the matrix P is stochastic: more generally, the states can be classified with respect to an arbitrary non-negative matrix P and the associated graph, where an oriented edge is drawn from x to y when $p(x, y) > 0$.

2.24 Theorem. *With respect to the matrix P_C^d, the irreducible class C decomposes into $d = d(C)$ irreducible, aperiodic classes $C_0, C_1, \dots, C_{d-1}$, which are visited in cyclic order by the original Markov chain: if $u \in C_i$, $v \in C$ and $p(u, v) > 0$ then $v \in C_{i+1}$, where $i + 1$ is computed modulo d.*

Schematically,

$$C_0 \xrightarrow{1} C_1 \xrightarrow{1} \cdots \xrightarrow{1} C_{d-1} \xrightarrow{1} C_0, \quad \textit{and}$$

$$x, y \textit{ belong to the same } C_i \iff p^{(md)}(x, y) > 0 \textit{ for some } m \geq 0.$$

Proof. Let $x_0 \in C$. Since $p^{(m_0 d)}(x_0, x_0) > 0$ for some $m_0 > 0$, there are $x_1, \dots, x_{d-1}, x_d \in C$ such that

$$x_0 \xrightarrow{1} x_1 \xrightarrow{1} \cdots \xrightarrow{1} x_{d-1} \xrightarrow{1} x_d \xrightarrow{(m_0-1)d} x_0.$$

Define

$$C_i = \{x \in C : x_i \xrightarrow{md} x \text{ for some } m \geq 0\}, \quad i = 0, 1, \dots, d-1, d.$$

(1) C_i is the irreducible class of x_i with respect to P_C^d:

(a) We have $x_i \in C_i$.

(b) If $x \in C_i$ then $x \in C$ and $x \xrightarrow{n} x_i$ for some $n \geq 0$. Thus $x_i \xrightarrow{md+n} x_i$, and d must divide $md + n$. Consequently, d divides n, and $x \xrightarrow{kd} x_i$ for some $k \geq 0$. It follows that $x \overset{P_C^d}{\longleftrightarrow} x_i$, i.e., x is in the class of x_i with respect to P_C^d.

(c) Conversely, if $x \overset{P_C^d}{\longleftrightarrow} x_i$ then there is $m \geq 0$ such that $x_i \xrightarrow{md} x$, and $x \in C_i$.

(d) By Lemma 2.22, C_i is aperiodic with respect to P_C^d.

(2) $C_d = C_0$: indeed, $x_0 \xrightarrow{d} x_d$ implies $x_d \in C_0$. By (1), $C_d = C_0$.

(3) $C = \bigcup_{i=0}^{d-1} C_i$: if $x \in C$ then $x_0 \xrightarrow{kd+r} x$, where $k \geq 0$ and $0 \leq r < d - 1$. By (1) and (2) there is $\ell \geq 0$ such that $x_r \xrightarrow{d-r} x_d \xrightarrow{\ell d} x_0 \xrightarrow{kd+r} x$, that is, $x_r \xrightarrow{md} x$ with $m = k + \ell + 1$. Therefore $x \in C_r$.

(4) If $x \in C_i$, $y \in C$ and $x \xrightarrow{1} y$ then $y \in C_{i+1}$ $(i = 0, 1, \ldots d - 1)$: there is n with $y \xrightarrow{n} x$. Hence $x \xrightarrow{n+1} x$, and d must divide $n + 1$. On the other hand, $x \xrightarrow{\ell d} x_i$ $(\ell \geq 0)$ by (1), and $x_i \xrightarrow{1} x_{i+1}$. We get $y \xrightarrow{n+\ell d+1} x_{i+1}$. Now let $k \geq 0$ be such that $x_{i+1} \xrightarrow{k} y$. (Such a k exists by the irreducibility of C with respect to P.) We obtain that $y \xrightarrow{k+\ell d+(n+1)} y$, and d divides $k + \ell d + (n + 1)$. But then d divides k, and $k = md$ with $m \geq 0$, so that $x_{i+1} \xrightarrow{md} y$, which implies that $y \in C_{i+1}$. □

2.25 Example. Consider a Markov chain with the following graph.

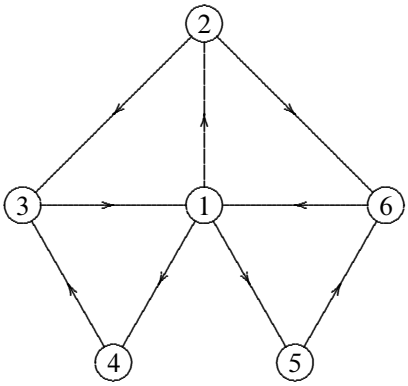

Figure 7

There is a unique irreducible class $C = X = \{1, 2, 3, 4, 5, 6\}$, the period is $d = 3$. Choosing $x_0 = 1$, $x_1 = 2$ and $x_2 = 3$, one obtains $C_0 = \{1\}$, $C_1 = \{2, 4, 5\}$ and $C_2 = \{3, 6\}$, which are visited in cyclic order.

2.26 Exercise. Show that if (X, P) is irreducible and aperiodic, then also (X, P^m) has these properties for each $m \in \mathbb{N}$. □

C The spectral radius of an irreducible class

For $x, y \in X$, consider the number

$$\mathsf{r}(x,y) = 1\Big/ \limsup_n \big(p^{(n)}(x,y)\big)^{1/n},$$

already defined in (1.33) as the radius of convergence of the power series $G(x, y|z)$. Its inverse $1/\mathsf{r}(x, y)$ describes the exponential decay of the sequence $\big(p^{(n)}(x, y)\big)$, as $n \to \infty$.

2.27 Lemma. *If $x \to w \to y$ then $\mathsf{r}(x, y) \le \min\{\mathsf{r}(x, w), \mathsf{r}(w, y)\}$. In particular, $x \to y$ implies $\mathsf{r}(x, y) \le \min\{\mathsf{r}(x, x), \mathsf{r}(y, y)\}$.*

If C is an irreducible class which does not consist of a single, ephemeral state then $\mathsf{r}(x, y) =: \mathsf{r}(C)$ is the same for all $x, y \in C$.

Proof. By assumption, there are numbers $k, \ell \ge 0$ such that $p^{(k)}(x, w) > 0$ and $p^{(\ell)}(w, y) > 0$. The inequality $p^{(n+\ell)}(x, y) \ge p^{(n)}(x, w)\, p^{(\ell)}(w, y)$ yields

$$\big(p^{(n+\ell)}(x,y)^{1/(n+\ell)}\big)^{(n+\ell)/n} \ge p^{(n)}(x,w)^{1/n}\, p^{(\ell)}(w,y)^{1/n}.$$

As $n \to \infty$, the second factor on the right hand side tends to 1, and $1/\mathsf{r}(x, y) \ge 1/\mathsf{r}(x, w)$.

Analogously, from the inequality $p^{(n+k)}(x, y) \ge p^{(k)}(x, w)\, p^{(n)}(w, y)$ one obtains $1/\mathsf{r}(x, y) \ge 1/\mathsf{r}(w, y)$.

If $x \to y$, we can write $x \to x \to y$, and $\mathsf{r}(x, y) \le \mathsf{r}(x, x)$ (setting $w = x$). Analogously, $x \to y \to y$ implies $\mathsf{r}(x, y) \le \mathsf{r}(y, y)$ (setting $w = y$).

To see the last statement, let $x, y, v, w \in C$. Then $v \to x \to y \to w$, which yields $\mathsf{r}(v, w) \le \mathsf{r}(v, y) \le \mathsf{r}(x, y)$. Analogously, $x \to v \to w \to y$ implies $\mathsf{r}(x, y) \le \mathsf{r}(v, w)$. □

Here is a characterization of $\mathsf{r}(C)$ in terms of $U(x, x|z) = \sum_n u^{(n)}(x, x)\, z^n$.

2.28 Proposition. *For any x in the irreducible class C,*

$$\mathsf{r}(C) = \max\{z > 0 : U(x, x|z) \le 1\}.$$

Proof. Write $\mathsf{r} = \mathsf{r}(x, x)$ and $\mathsf{s} = \mathsf{s}(x, x)$ for the respective radii of convergence of the power series $G(x, x|z)$ and $U(x, x|z)$. Both functions are strictly increasing in $z > 0$. We know that $\mathsf{s} \ge \mathsf{r}$. Also, $U(x, x|z) = \infty$ for real $z > \mathsf{s}$. From the equation

$$G(x,x|z) = \frac{1}{1 - U(x,x|z)}$$

we read that $U(x,x|z) < 1$ for all positive $z < \mathsf{r}$. We can let tend $z \to \mathsf{r}$ from below and infer that $U(x,x|\mathsf{r}) \le 1$. The proof will be concluded when we can show that for our power series, $U(x,x|z) > 1$ whenever $z > \mathsf{r}$.

This is clear when $U(x,x|\mathsf{r}) = 1$. So consider the case when $U(x,x|\mathsf{r}) < 1$. Then we claim that $\mathsf{s} = \mathsf{r}$, whence $U(x,x|z) = \infty > 1$ whenever $z > \mathsf{r}$. Suppose by contradiction that $\mathsf{s} > \mathsf{r}$. Then there is a real z_0 with $\mathsf{r} < z_0 < \mathsf{s}$ such that we have $U(x,x|z_0) \le 1$. Set $u_n = u^{(n)}(x,x)\, z_0^n$ and $p_n = p^{(n)}(x,x)\, z_0^n$. Then (1.39) leads to the (*renewal*) recursion[2]

$$p_0 = 1, \quad p_n = \sum_{k=1}^{n} u_k\, p_{n-k} \quad (n \ge 1).$$

Since $\sum_k u_k \le 1$, induction on n yields $p_n \le 1$ for all n. Therefore

$$G(x,x|z) = \sum_{n=0}^{\infty} p_n (z/z_0)^n$$

converges for all $z \in \mathbb{C}$ with $|z| < z_0$. But then $\mathsf{r} \ge z_0$, a contradiction. □

The number

$$\rho(P_C) = 1/\mathsf{r}(C) = \limsup_n \bigl(p^{(n)}(x,y)\bigr)^{1/n}, \quad x,y \in C, \tag{2.29}$$

is called the *spectral radius* of P_C, resp. C. If the Markov chain (X,P) is irreducible, then $\rho(P)$ is the *spectral radius of* P. The terminology is in part justified by the following (which is not essential for the rest of this chapter).

2.30 Proposition. *If the irreducible class C is finite then $\rho(P_C)$ is an eigenvalue of P_C, and every other eigenvalue λ satisfies $|\lambda| \le \rho(P_C)$.*

We won't prove this proposition right now. It is a consequence of the famous *Perron–Frobenius theorem,* which is of utmost importance in Markov chain theory, see Seneta [Se]. We shall give a detailed proof of that theorem in Section 3.D.

Let us also remark that in the case when C is infinite, the name "spectral radius" for $\rho(P_C)$ may be misleading, since it does in general not refer to an action of P_C as a bounded linear operator on a suitable space. On the other hand, later on we shall encounter *reversible* irreducible Markov chains, and in this situation $\rho(P)$ is a "true" spectral radius.

The following is a corollary of Theorem 2.7.

[2] In the literature, what is denoted u_k here is usually called f_k, and what is denoted p_n here is often called u_n. Our choice of the notation is caused by the necessity to distinguish between the generating $F(x,y|z)$ and $U(x,y|z)$ of the stopping time s^y and t^y, respectively, which are both needed at different instances in this book.

2.31 Proposition. *Let C be a finite irreducible class. Then $\rho(P_C) = 1$ if and only if C is essential.*

In other words, if P is a finite, irreducible, substochastic matrix then $\rho(P) = 1$ if and only if P is stochastic.

Proof. If C is non-essential then Theorem 2.7 shows that $\rho(P_C) < 1$. Conversely, if C is essential then all the matrix powers P_C^n are stochastic. Thus, we cannot have $\rho(P_C) < 1$, since in that case we would have $\sum_{y\in C} p^{(n)}(x, y) \to 0$ as $n \to \infty$. □

In the following, the class C need not be finite.

2.32 Theorem. *Let C be an irreducible class which does not consist of a single, ephemeral state, and let $d = d(C)$. If $x, y \in C$ and $p^{(k)}(x, y) > 0$ (such k must exist) then $p^{(m)}(x, y) = 0$ for all $m \not\equiv k \mod d$, and*

$$\lim_{n\to\infty} p^{(nd+k)}(x, y)^{1/(nd+k)} = \rho(P_C) > 0.$$

In particular,

$$\lim_{n\to\infty} p^{(nd)}(x, x)^{1/(nd)} = \rho(P_C) \quad \text{and} \quad p^{(n)}(x, x) \le \rho(P_C)^n \quad \text{for all } n \in \mathbb{N}.$$

For the proof, we need the following.

2.33 Proposition. *Let (a_n) be a sequence of non-negative real numbers such that $a_n > 0$ for all $n \ge n_0$ and $a_m a_n \le a_{m+n}$ for all $m, n \in \mathbb{N}$. Then*

$$\exists \lim_{n\to\infty} a_n^{1/n} = \rho > 0 \quad \text{and} \quad a_n \le \rho^n \ \text{ for all } n \in \mathbb{N}.$$

Proof. Let $n, m \ge n_0$. For the moment, consider n fixed. We can write $m = q_m n + r_m$ with $q_m \ge 0$ and $n_0 \le r_m < n_0 + n$. In particular, $a_{r_m} > 0$, we have $q_m \to \infty$ if and only if $m \to \infty$, and in this case $m/q_m \to n$. Let $\varepsilon = \varepsilon_n = \min\{a_{r_m} : m \in \mathbb{N}\}$. Then by assumption,

$$a_m \ge a_{q_m n} a_{r_m} \ge a_n^{q_m} \varepsilon_n, \quad \text{and} \quad a_m^{1/m} \ge a_n^{q_m/m} \varepsilon_n^{1/m}.$$

As $m \to \infty$, we get $a_n^{q_m/m} \to a_n^{1/n}$ and $\varepsilon_n^{1/m} \to 1$. It follows that

$$\rho := \liminf_{m\to\infty} a_m^{1/m} \ge a_n^{1/n} \quad \text{for each } n \in \mathbb{N}.$$

If now also $n \to \infty$, then this implies

$$\liminf_{m\to\infty} a_m^{1/m} \ge \limsup_{n\to\infty} a_n^{1/n},$$

and $\lim_{n\to\infty} a_n^{1/n}$ exists and is equal to $\liminf_{m\to\infty} a_m^{1/m} = \rho$. □

Proof of Theorem 2.32. By Lemma 2.22, the sequence $(a_n) = \big(p^{(nd)}(x,x)\big)$ fulfills the hypotheses of Proposition 2.33, and $a_n^{1/n} \to \rho(P_C) > 0$.

We have $p^{(nd+k)}(x,y) \geq p^{(nd)}(x,x)\, p^{(k)}(x,y)$, and

$$p^{(nd+k)}(x,y)^{1/(nd+k)} \geq \big(p^{(nd)}(x,x)^{1/nd}\big)^{nd/(nd+k)}\, p^{(k)}(x,y)^{1/(nd+k)}.$$

As $p^{(k)}(x,y) > 0$, the last factor tends to 1 as $n \to \infty$, while by the above $\big(p^{(nd)}(x,x)^{1/(nd)}\big)^{nd/(nd+k)} \to \rho(P_C)$. We conclude that

$$\begin{aligned}\rho(P_C) &\leq \liminf_{n\to\infty} p^{(nd+k)}(x,y)^{1/(nd+k)} \\ &\leq \limsup_{n\to\infty} p^{(nd+k)}(x,y)^{1/(nd+k)} = \limsup_{n\to\infty} p^{(n)}(x,y)^{1/n} = \rho(P_C). \quad \square\end{aligned}$$

2.34 Exercise. Modify Example 1.1 by making the rainy state 🌧 absorbing: set $p(🌧,🌧) = 1$, $p(🌧,☼) = p(🌧,❄) = 0$. The other transition probabilities remain unchanged. Compute the spectral radius of the class $\{☼, ❄\}$. $\square$

Chapter 3

Recurrence and transience, convergence, and the ergodic theorem

A Recurrent classes

The following concept is of central importance in Markov chain theory.

3.1 Definition. Consider a Markov chain (X, P). A state $x \in X$ is called *recurrent*, if

$$U(x, x) = \mathsf{Pr}_x[\exists\, n > 0 : Z_n = x] = 1,$$

and *transient,* otherwise.

In words, x is recurrent if it is certain that the Markov chain starting at x will return to x.

We also define the probabilities

$$H(x, y) = \mathsf{Pr}_x[Z_n = y \text{ for infinitely many } n \in \mathbb{N}], \quad x, y \in X.$$

If it is certain that (Z_n) returns to x at least once, then it will return to x infinitely often with probability 1. If the probability to return at least once is strictly less than 1, then it is unlikely (probability 0) that (Z_n) will return to x infinitely often. In other words, $H(x, x)$ cannot assume any values besides 0 or 1, as we shall prove in the next theorem. Such a statement is called a *zero-one law.*

3.2 Theorem. (a) *The state x is recurrent if and only if $H(x, x) = 1$.*

(b) *The state x is transient if and only if $H(x, x) = 0$.*

(c) *We have $H(x, y) = U(x, y)\, H(y, y)$.*

Proof. We define

$$H^{(m)}(x, y) = \mathsf{Pr}_x[Z_n = y \text{ for at least } m \text{ time instants } n > 0]. \tag{3.3}$$

Then

$$H^{(1)}(x, y) = U(x, y) \quad \text{and} \quad H(x, y) = \lim_{m \to \infty} H^{(m)}(x, y),$$

and

$$H^{(m+1)}(x,y) = \sum_{k=1}^{\infty} \mathsf{Pr}_x[t^y = k \text{ and } Z_n = y \text{ for at least } m \text{ instants } n > k]$$

(using once more the rule $\mathsf{Pr}(A \cap B) = \mathsf{Pr}(A)\ \mathsf{Pr}(B \mid A)$, if $\mathsf{Pr}(A) > 0$)

$$= \sum_{k:u^{(k)}(x,y)>0} f^{(k)}(x,y)\ \mathsf{Pr}_x\left[\begin{array}{c|c} Z_n = y \text{ for at least} & Z_k = y,\ Z_i \neq y \\ m \text{ time instants } n > k & \text{for } i = 1,\dots,k-1 \end{array}\right]$$

(using the Markov property (1.5))

$$= \sum_{k:u^{(k)}(x,y)>0} f^{(k)}(x,y)\ \mathsf{Pr}_x[Z_n = y \text{ for at least } m \text{ instants } n > k \mid Z_k = y]$$

$$= \sum_{k=1}^{\infty} u^{(k)}(x,y)\, H^{(m)}(y,y) = U(x,y)\, H^{(m)}(y,y).$$

Therefore $H^{(m)}(x,x) = U(x,x)^m$. As $m \to \infty$, we get (a) and (b). Furthermore, $H(x,y) = \lim_{m\to\infty} U(x,y)\, H^{(m)}(y,y) = U(x,y)\, H(y,y)$, and we have proved (c). □

We now list a few properties of recurrent states.

3.4 Theorem. (a) *The state x is recurrent if and only if $G(x,x) = \infty$.*

(b) *If x is recurrent and $x \to y$ then $U(y,x) = H(y,x) = 1$, and y is recurrent. In particular, x is essential.*

(c) *If C is a finite essential class then all elements of C are recurrent.*

Proof. (a) Observe that, by monotone convergence[1],

$$U(x,x) = \lim_{z\to 1-} U(x,x|z) \quad \text{and} \quad G(x,x) = \lim_{z\to 1-} G(x,x|z).$$

Therefore Theorem 1.38 implies

$$G(x,x) = \lim_{z\to 1-} \frac{1}{1 - U(x,x|z)} = \begin{cases} \infty, & \text{if } U(x,x) = 1, \\ \dfrac{1}{1-U(x,x)}, & \text{if } U(x,x) < 1. \end{cases}$$

[1] We can interpret a power series $\sum_n a_n z^n$ with $a_n, z \geq 0$ as an integral $\int f_z(n)\, d\sigma(n)$, where σ is the counting measure on $\mathbb{N}_0$ and $f_z(n) = a_n z^n$. Then $f_z(n) \to f_r(n)$ as $z \to r-$ (monotone limit from below), whence $\int f_z(n)\, d\sigma(n) \to \int f_r(n)\, d\sigma(n)$, so that this elementary fact from calculus is indeed a basic variant of the monotone convergence theorem of integration theory.

(b) We proceed by induction: if $x \xrightarrow{n} y$ then $U(y,x) = 1$.

By assumption, $U(x,x) = 1$, and the statement is true for $n = 0$. Suppose that it holds for n and that $x \xrightarrow{n+1} y$. Then there is $w \in X$ such that $x \xrightarrow{n} w \xrightarrow{1} y$. By the induction hypothesis $U(w,x) = 1$. By Theorem 1.38 (c) (with $z = 1$),

$$1 = U(w,x) = p(w,x) + \sum_{v \neq x} p(w,v)\, U(v,x).$$

Stochasticity of P implies

$$0 = \sum_{v \neq x} p(w,v)\big(1 - U(v,x)\big) \geq p(w,y)\big(1 - U(y,x)\big) \geq 0.$$

Since $p(w,y) > 0$, we must have $U(y,x) = 1$.

Hence $H(y,x) = U(y,x)\, H(x,x) = U(y,x) = 1$ for every y with $x \to y$. By Exercise 2.6 (c), x is essential. We now show that also y is recurrent if x is recurrent and $y \leftrightarrow x$: there are $k, \ell \geq 0$ such that $p^{(k)}(x,y) > 0$ and $p^{(\ell)}(y,x) > 0$. Consequently, by (a)

$$G(y,y) \geq \sum_{n=k+\ell}^{\infty} p^{(n)}(y,y) \geq p^{(\ell)}(y,x) \sum_{m=0}^{\infty} p^{(m)}(x,x)\, p^{(k)}(x,y) = \infty.$$

(c) Since C is essential, when starting in $x \in C$, the Markov chain cannot exit from C:

$$\mathsf{Pr}_x[Z_n \in C \text{ for all } n \in \mathbb{N}] = 1.$$

Let $\omega \in \Omega$ be such that $Z_n(\omega) \in C$ for all n. Since C is finite, there is at least one $y \in C$ (depending on ω) such that $Z_n(\omega) = y$ for infinitely many instants n. In other terms,

$$\begin{aligned} &\{\omega \in \Omega \mid Z_n(\omega) \in C \text{ for all } n \in \mathbb{N}\} \\ &\qquad \subset \{\omega \in \Omega \mid \exists\, y \in C : Z_n(\omega) = y \text{ for infinitely many } n\} \end{aligned}$$

and

$$\begin{aligned} 1 &= \mathsf{Pr}_x[\exists\, y \in C : Z_n = y \text{ for infinitely many } n] \\ &\leq \sum_{y \in C} \mathsf{Pr}_x[Z_n = y \text{ for infinitely many } n] = \sum_{y \in C} H(x,y). \end{aligned}$$

In particular there must be $y \in C$ such that $0 < H(x,y) = U(x,y)H(y,y)$, and Theorem 3.2 yields $H(y,y) = 1$: the state y is recurrent, and by (b) every element of C is recurrent. □

Recurrence is thus a property of irreducible classes: if C is irreducible then either all elements of C are recurrent or all are transient. Furthermore a recurrent

irreducible class must always be essential. In a Markov chain with finite state space, all essential classes are recurrent. If X is infinite, this is no longer true, as the next example shows.

3.5 Example (Infinite drunkard's walk). The state space is $X = \mathbb{Z}$, and the transition probabilities are defined in terms of the two parameters p and q ($p, q > 0$, $p + q = 1$), as follows.

$$p(k, k+1) = p, \qquad p(k, k-1) = q, \qquad p(k, \ell) = 0 \quad \text{if } |k - \ell| \neq 1.$$

This Markov chain ("random walk") can also be interpreted as a coin tossing game: if "heads" comes up then we win one Euro, and if "tails" comes up we lose one Euro. The coin is not necessarily fair; "heads" comes up with probability p and "tails" with probability q. The state $k \in \mathbb{Z}$ represents the possible (positive or negative) capital gain in Euros after some repeated coin tosses. If the single tosses are mutually independent, then one passes in a single step (toss) from capital k to $k + 1$ with probability p, and to $k - 1$ with probability q.

Observe that in this example, we have *translation invariance:* $F(k, \ell|z) = F(k - \ell, 0|z) = F(0, \ell - k|z)$, and the same holds for $G(\cdot, \cdot|z)$. We compute $U(0, 0|z)$.

Reasoning precisely as in Example 2.10, we obtain

$$F(1, 0|z) = \frac{1}{2pz}\left(1 - \sqrt{1 - 4pqz^2}\right).$$

By symmetry, exchanging the roles of p and q,

$$F(-1, 0|z) = \frac{1}{2qz}\left(1 - \sqrt{1 - 4pqz^2}\right).$$

Now, by Theorem 1.38 (c)

$$U(0, 0|z) = pz\, F(1, 0|z) + qz\, F(-1, 0|z) = 1 - \sqrt{1 - 4pqz^2}, \tag{3.6}$$

whence

$$U(0, 0) = U(0, 0|1) = 1 - \sqrt{(p - q)^2} = 1 - |p - q|.$$

There is a single irreducible (whence essential) class, but the random walk is recurrent only when $p = q = 1/2$.

3.7 Exercise. Show that if ν is an arbitrary starting distribution and y is a transient state, then

$$\lim_{n \to \infty} \mathsf{Pr}_\nu[Z_n = y] = 0. \qquad \square$$

B Return times, positive recurrence, and stationary probability measures

Let x be a recurrent state of the Markov chain (X, P). Then it is certain that (Z_n), after starting in x, will return to x. That is, the return time $\boldsymbol{t}^x$ is Pr_x-almost surely finite. The expected return time (= number of steps) is

$$\mathsf{E}_x[\boldsymbol{t}^x] = \sum_{n=1}^{\infty} n\, u^{(n)}(x,x) = U'(x,x|1),$$

or more precisely, $\mathsf{E}_x[\boldsymbol{t}^x] = U'(x,x|1-) = \lim_{z\to 1-} U'(x,x|z)$ by monotone convergence. This limit may be infinite.

3.8 Definition. A recurrent state x is called

positive recurrent, if $\mathsf{E}_x[\boldsymbol{t}^x] < \infty$, and

null recurrent, if $\mathsf{E}_x[\boldsymbol{t}^x] = \infty$.

In Example 3.5, the infinite drunkard's random walk is recurrent if and only if $p = q = 1/2$. In this case $U(0,0|z) = 1-\sqrt{1-z^2}$ and $U'(0,0|z) = z\big/\sqrt{1-z^2}$ for $|z| < 1$, so that $U'(0,0|1-) = \infty$: the state 0 is null recurrent.

The next theorem shows that positive and null recurrence are class properties of (recurrent, essential) irreducible classes.

3.9 Theorem. *Suppose that x is positive recurrent and that $y \leftrightarrow x$. Then also y is positive recurrent. Furthermore, $\mathsf{E}_y[\boldsymbol{t}^x] < \infty$.*

Proof. We know from Theorem 3.4 (b) that y is recurrent. In particular, the Green function has convergence radii $\mathsf{r}(x,x) = \mathsf{r}(y,y) = 1$, and by Theorem 1.38 (a)

$$\frac{1-U(x,x|z)}{1-U(y,y|z)} = \frac{G(y,y|z)}{G(x,x|z)} \quad \text{for } 0 < z < 1.$$

As $z \to 1-$, the right hand side becomes an expression of type $\frac{\infty}{\infty}$. Since $0 < U'(y,y|1-) \le \infty$, an application of de l'Hospital's rule yields

$$\frac{\mathsf{E}_x(\boldsymbol{t}^x)}{\mathsf{E}_y(\boldsymbol{t}^y)} = \frac{U'(x,x|1-)}{U'(y,y|1-)} = \lim_{z\to 1-} \frac{G(y,y|z)}{G(x,x|z)}.$$

There are $k, l > 0$ such that $p^{(k)}(x,y) > 0$ and $p^{(l)}(y,x) > 0$. Therefore, if $0 < z < 1$,

$$G(y,y|z) \ge \sum_{n=0}^{k+l-1} p^{(n)}(y,y)\, z^n + p^{(l)}(y,x)\, G(x,x|z)\, p^{(k)}(x,y)\, z^{k+l}.$$

We obtain

$$\lim_{z\to 1-} \frac{G(y,y|z)}{G(x,x|z)} \geq p^{(l)}(y,x)\, p^{(k)}(x,y) > 0.$$

In particular, we must have $\mathsf{E}_y(t^y) < \infty$.

To see that also $\mathsf{E}_y(t^x) = U'(y,x|1-) < \infty$, suppose that $x \xrightarrow{n} y$. We proceed by induction on n as in the proof of Theorem 3.4 (b). If $n = 0$ then $y = x$ and the statement is true. Suppose that it holds for n and that $x \xrightarrow{n+1} y$. Let $w \in X$ be such that $x \xrightarrow{n} w \xrightarrow{1} y$. By Theorem 1.38 (c), for $0 < z \leq 1$,

$$U(w,x|z) = p(w,x)z + \sum_{v\neq x} p(w,v)z\, U(v,x|z).$$

Differentiating on both sides and letting $z \to 1$ from below, we see that finiteness of $U'(w,x|1-)$ (which holds by the induction hypothesis) implies finiteness of $U'(v,x|1-)$ for all v with $p(w,v) > 0$, and in particular for $v = y$. □

An irreducible (essential) class is called positive or null recurrent, if all its elements have the respective property.

3.10 Theorem. *Let C be a finite essential class of (X,P). Then C is positive recurrent.*

Proof. Since the matrix P^n is stochastic, we have for each $x \in X$

$$\sum_{y\in X} G(x,y|z) = \sum_{n=0}^{\infty} \sum_{y\in X} p^{(n)}(x,y)\, z^n = \frac{1}{1-z}, \quad 0 \leq z < 1.$$

Using Theorem 1.38, we can write $G(x,y|z) = F(x,y|z)\big/\big(1-U(y,y|z)\big)$. Thus, we obtain the following important identity (that will be used again later on)

$$\sum_{y\in X} F(x,y|z)\frac{1-z}{1-U(y,y|z)} = 1 \quad \text{for each } x \in X \text{ and } 0 \leq z < 1. \tag{3.11}$$

Now suppose that x belongs to the finite essential class C. Then a non-zero contribution to the sum in (3.11) can come only from elements $y \in C$. We know already (Theorem 3.4) that C is recurrent. Therefore $F(x,y|1-) = U(y,y|1-) = 1$ for all $x, y \in C$. Since C is finite, we can exchange sum and limit and apply de l'Hospital's rule:

$$1 = \lim_{z\to 1-} \sum_{y\in C} F(x,y|z)\frac{1-z}{1-U(y,y|z)} = \sum_{y\in C} \frac{1}{U'(y,y|1-)}. \tag{3.12}$$

Therefore there must be $y \in C$ such that $U'(y,y|1-) < \infty$, so that y, and thus the class C, is positive recurrent. □

3.13 Exercise. Use (3.11) to give a "generating function" proof of Theorem 3.4 (c). □

3.14 Exercise. Let $(\bar{X}, \bar{P})$ be a factor chain of (X, P); see (1.30).

Show that if x is a recurrent state of (X, P), then its projection $\pi(x) = \bar{x}$ is a recurrent state of $(\bar{X}, \bar{P})$. Also show that if x is positive recurrent, then so is $\bar{x}$. □

It is convenient to think of measures ν on X as row vectors $\big(\nu(x)\big)_{x\in X}$, so that νP is the product of the row vector ν with the matrix P,

$$\nu P(y) = \sum_x \nu(x)\, p(x, y). \tag{3.15}$$

Here, we do not necessarily suppose that $\nu(X)$ is finite, so that the last sum might diverge.

In the same way, we consider real or complex functions f on X as column vectors, whence Pf is the function

$$Pf(y) = \sum_y p(x, y) f(y), \tag{3.16}$$

as long as this sum is defined (it might be a divergent series).

3.17 Definition. A (non-negative) measure ν on X is called *invariant* or *stationary,* if $\nu P = \nu$. It is called *excessive,* if $\nu P \le \nu$ pointwise.

3.18 Exercise. Suppose that ν is an excessive measure and that $\nu(X) < \infty$. Show that ν is stationary. □

We say that a set $A \subset X$ *carries* the measure ν on X, if the support of ν, $\mathrm{supp}(\nu) = \{x \in X : \nu(x) \neq 0\}$, is contained in A.

In the next theorem, the class C is not necessarily assumed to be finite.

3.19 Theorem. *Let C be an essential class of (X, P). Then C is positive recurrent if and only if it carries a stationary probability measure m_C. In this case, the latter is unique and given by*

$$\mathsf{m}_C(x) = \begin{cases} 1/\mathsf{E}_x(\boldsymbol{t}^x), & \text{if} \quad x \in C, \\ 0, & \text{otherwise.} \end{cases}$$

Proof. We first show that in the positive recurrent case, m_C is a stationary probability measure. When C is infinite, we cannot apply (3.11) directly to show that $\mathsf{m}_C(C) = 1$, because *a priori* we are not allowed to exchange limit and sum in

(3.12). However, if $A \subset C$ is an arbitrary *finite* subset, then (3.11) implies that for $0 < z < 1$ and $x \in C$,

$$\sum_{y \in A} F(x, y|z) \frac{1 - z}{1 - U(y, y|z)} \leq 1.$$

If we now let $z \to 1$ from below and proceed as in (3.12), we find $\mathsf{m}_C(A) \leq 1$. Therefore the total mass of m_C is $\mathsf{m}_C(X) = \mathsf{m}_C(C) \leq 1$.

Next, recall the identity (1.34). We do not only have $\mathcal{G}(z) = I + zP\,\mathcal{G}(z)$ but also, in the same way,

$$\mathcal{G}(z) = I + \mathcal{G}(z)\, zP.$$

Thus, for $y \in C$,

$$G(y, y|z) = 1 + \sum_{x \in X} G(y, x|z)\, p(x, y) z. \tag{3.20}$$

We use Theorem 1.38 and multiply once more both sides by $1 - z$. Since only elements $x \in C$ contribute to the last sum,

$$\frac{1 - z}{1 - U(y, y|z)} = 1 - z + \sum_{x \in C} F(y, x|z) \frac{1 - z}{1 - U(x, x|z)}\, p(x, y) z. \tag{3.21}$$

Again, by recurrence, $F(y, x|1) = U(x, x|1) = 1$. As above, we restrict the last sum to an arbitrary finite $A \subset C$ and then let $z \to 1-$. De l'Hospital's rule yields

$$\frac{1}{U'(y, y|1-)} \geq \sum_{x \in A} \frac{1}{U'(x, x|1-)}\, p(x, y).$$

Since this inequality holds for every finite $A \subset C$, it also holds with C itself [2] in the place of A. Thus, m_C is an excessive measure, and by Exercise 3.18, it is stationary. By positive recurrence, $\mathrm{supp}(\mathsf{m}_C) = C$, and we can normalize it so that it becomes a stationary probability measure. This proves the "only if"-part.

To show the "if" part, suppose that ν is a stationary probability measure with $\mathrm{supp}(\nu) \subset C$.

Let $y \in C$. Given $\varepsilon > 0$, since $\nu(C) = 1$, there is a finite set $A_\varepsilon \subset C$ such that $\nu(C \setminus A_\varepsilon) < \varepsilon$. Stationarity implies that $\nu P^n = \nu$ for each $n \in \mathbb{N}_0$, and

$$\nu(y) = \sum_{x \in C} \nu(x)\, p^{(n)}(x, y) \leq \sum_{x \in A_\varepsilon} \nu(x)\, p^{(n)}(x, y) + \varepsilon.$$

[2] As a matter of fact, this argument, as well as the one used above to show that $\mathsf{m}_C(C) \leq 1$, amounts to applying Fatou's lemma of integration theory to the sum in (3.21), which once more is interpeted as an integral with respect to the counting measure on C.

Once again, we multiply both sides by $(1-z)z^n$ $(0<z<1)$ and sum over n:

$$\nu(y) \le \sum_{x\in A_\varepsilon} \nu(x)\, F(x,y|z)\frac{1-z}{1-U(y,y|z)} + \varepsilon.$$

Suppose first that C is transient. Then $U(y,y|1-) < 1$ for each $y \in C$ by Theorem 3.4 (b), and when $z \to 1$ from below, the last sum tends to 0. That is, $\nu(y) = 0$ for each $y \in C$, a contradiction. Therefore C is recurrent, and we can apply once more de l'Hospital's rule, when $z \to 1-$:

$$\nu(y) - \varepsilon \le \sum_{x\in A_\varepsilon} \nu(x)\, \mathsf{m}_C(y) = \nu(A_\varepsilon)\, \mathsf{m}_C(y) \le \mathsf{m}_C(y)$$

for each $\varepsilon > 0$. Therefore $\nu(y) \le \mathsf{m}_C(y)$ for each $y \in X$. Thus $\mathsf{m}_C(y) > 0$ for at least one $y \in C$. This means that $\mathsf{E}_y(t^y) = 1/\mathsf{m}_C(y) < \infty$, and C must be positive recurrent.

Finally, since $\nu(X) = 1$ and $\mathsf{m}_C(X) \le 1$, we infer that m_C is indeed a probability measure, and $\nu = \mathsf{m}_C$. In particular, the stationary probability measure m_C carried by C is unique. □

3.22 Exercise. Reformulate the method of the second part of the proof of Theorem 3.19 to show the following.

If (X, P) is an arbitrary Markov chain and ν a stationary probability measure, then $\nu(y) > 0$ implies that y is a positive recurrent state. □

3.23 Corollary. *The Markov chain (X, P) admits stationary probability measures if and only if there are positive recurrent states.*

In this case, let C_i, $i \in I$, be those essential classes which are positive recurrent (with $I = \mathbb{N}$ or $I = \{1,\dots,k\}$, $k \in \mathbb{N}$). For $i \in I$, let $\mathsf{m}_i = \mathsf{m}_{C_i}$ be the stationary probability measure of (C_i, P_{C_i}) according to Theorem 3.19. Consider m_i as a measure on X with $\mathsf{m}_i(x) = 0$ for $x \in X \setminus C_i$.

Then the stationary probability measures of (X, P) are precisely the convex combinations

$$\nu = \sum_{i\in I} c_i \cdot \mathsf{m}_i, \quad \text{where } c_i \ge 0 \text{ and } \sum_{i\in I} c_i = 1.$$

Proof. Let ν be a stationary probability measure. By Exercise 3.22, every x with $\nu(x) > 0$ must be positive recurrent. Therefore there must be positive recurrent essential classes C_i, $i \in I$. By Theorem 3.19, the restriction of ν to any of the C_i must be a non-negative multiple of m_i. Therefore ν must have the proposed form.

Conversely, it is clear that any convex combination of the m_i is a stationary probability measure. □

3.24 Exercise. Let C be a positive recurrent essential class, $d = d(C)$ its period, and $C_0, \dots, C_{d-1}$ its periodic classes (the irreducible classes of P_C^d) according to Theorem 2.24. Determine the stationary probability measure of P_C^d on C_i in terms of the stationary probability m_C of P on d. Compute first $\mathsf{m}_C(C_i)$ for $i = 0, \dots, d-1$. □

C The convergence theorem for finite Markov chains

We shall now study the question of whether the transition probabilities $p^{(n)}(x, y)$ converge to a limit as $n \to \infty$. If y is a transient state then we know from Theorem 3.4 (a) that $G(x, y) = F(x, y)G(y, y) < \infty$, so that $p^{(n)}(x, y) \to 0$. Thus, the question is of interest when y is essential. For the moment, we restrict attention to the case when x and y belong to the same essential class. Since (Z_n) cannot exit from that class, we may assume without loss of generality that this class is the whole of X. That is, we assume to have an irreducible Markov chain (X,P). Before stating results, let us see what we expect in the specific case when X is finite.

Suppose that X is finite and that $p^{(n)}(x, y)$ converges for all x, y as $n \to \infty$. The Markov property ("absence of memory") suggests that on the long run (as $n \to \infty$), the starting point should be "forgotten", that is, the limit should not depend on x. Thus, suppose that $\lim_n p^{(n)}(x, y) = \mathsf{m}(y)$ for all y, where m is a measure on X. Then – since X is finite and P^n is stochastic for each n – we should have $\mathsf{m}(X) = 1$, whence $\mathsf{m}(x) > 0$ for some x. But then $p^{(n)}(x, x) > 0$ for all but finitely many n, so that P should be aperiodic. Also, if we let $n \to \infty$ in the relation

$$p^{(n+1)}(x, y) = \sum_{w \in X} p^{(n)}(x, w)\, p(w, y),$$

we find that m should be the stationary probability measure.

We now know what we are looking for and which hypotheses we need, and can start to work, without supposing right away that X is finite.

Consider the set of all probability distributions on X,

$$M(X) = \{\nu : X \to \mathbb{R} \mid \nu(x) \geq 0 \text{ for all } x \in X \text{ and } \textstyle\sum_{x \in X} \nu(x) = 1\}.$$

We consider $M(X)$ as a subset of $\ell^1(X)$. In particular, $M(X)$ is closed in the metric

$$\|\nu_1 - \nu_2\|_1 = \sum_{x \in X} |\nu_1(x) - \nu_2(x)|.$$

(The number $\|\nu_1 - \nu_2\|_1/2$ is usually called the total variation norm of the signed measure $\nu_1 - \nu_2$.) The transition matrix P acts on $M(X)$, according to (3.15), by $\nu \mapsto \nu P$, which is in $M(X)$ by stochasticity of P. Our goal is to apply the Banach

fixed point theorem to this contraction, if possible. For $y \in X$, we define

$$a(y) = a(y, P) = \inf_{x \in X} p(x, y) \quad \text{and} \quad \tau = \tau(P) = 1 - \sum_{y \in X} a(y).$$

Thus, $a(y)$ is the infimum of the y-column of P, and $0 \le \tau \le 1$. Indeed, if $x \in X$ then $p(x, y) \ge a(y)$ for each y, and

$$1 = \sum_{y \in X} p(x, y) \ge \sum_{y \in X} a(y).$$

We remark that $\tau(P)$ is a so-called *ergodic coefficient,* and that the methods that we are going to use are a particular instance of the theory of those coefficients, see e.g. SENETA [Se, §4.3] or ISAACSON and MADSEN [I-M, Ch. V]. We have $\tau(P) < 1$ if and only if P has a column where all elements are strictly positive. Also, $\tau(P) = 0$ if and only if all rows of P coincide (i.e., they coincide with a single probability measure on X).

3.25 Lemma. *For all* $\nu_1, \nu_2 \in M(X)$,

$$\|\nu_1 P - \nu_2 P\|_1 \le \tau(P)\, \|\nu_1 - \nu_2\|_1.$$

Proof. For each $y \in X$ we have

$$\begin{aligned}
\nu_1 P(y) - \nu_2 P(y) &= \sum_{x \in X} \big(\nu_1(x) - \nu_2(x)\big) p(x, y) \\
&= \sum_{x \in X} |\nu_1(x) - \nu_2(x)| p(x, y) - \sum_{x \in X} \Big(|\nu_1(x) - \nu_2(x)| - \big(\nu_1(x) - \nu_2(x)\big)\Big) p(x, y) \\
&\le \sum_{x \in X} |\nu_1(x) - \nu_2(x)| p(x, y) - \sum_{x \in X} \Big(|\nu_1(x) - \nu_2(x)| - \big(\nu_1(x) - \nu_2(x)\big)\Big) a(y) \\
&= \sum_{x \in X} |\nu_1(x) - \nu_2(x)| \big(p(x, y) - a(y)\big).
\end{aligned}$$

By symmetry,

$$|\nu_1 P(y) - \nu_2 P(y)| \le \sum_{x \in X} |\nu_1(x) - \nu_2(x)| \big(p(x, y) - a(y)\big).$$

Therefore,

$$\begin{aligned}
\|\nu_1 P - \nu_2 P\|_1 &\le \sum_{y \in X} \sum_{x \in X} |\nu_1(x) - \nu_2(x)| \big(p(x, y) - a(y)\big) \\
&= \sum_{x \in X} |\nu_1(x) - \nu_2(x)| \sum_{y \in Y} \big(p(x, y) - a(y)\big) \\
&= \tau(P)\, \|\nu_1 - \nu_2\|_1.
\end{aligned}$$

This proves the lemma. □

The following theorem does not (yet) assume finiteness of X.

3.26 Theorem. *Suppose that (X, P) is irreducible and such that $\tau(P^k) < 1$ for some $k \in \mathbb{N}$. Then P is aperiodic, (positive) recurrent, and there is $\bar\tau < 1$ such that for each $\nu \in M(X)$,*

$$\|\nu P^n - \mathsf{m}\|_1 \le 2\bar\tau^n \quad \text{for all } n \ge k,$$

where m is the unique stationary probability distribution of P.

Proof. The inequality $\tau(P^k) < 1$ implies that $a(y, P^k) > 0$ for some $y \in X$. For this y there is at least one $x \in X$ with $p(y, x) > 0$. We have $p^{(k)}(y, y) > 0$ and $p^{(k)}(x, y) > 0$, and also $p^{(k+1)}(y, y) \ge p(y, x)\, p^{(k)}(x, y) > 0$. For the period d of P we thus have $d \,|\, k$ and $d \,|\, k + 1$. Consequently $d = 1$.

Set $\tau = \tau(P^k)$. By Lemma 3.25, the mapping $\nu \mapsto \nu P^k$ is a contraction of $M(X)$. It follows from Banach's fixed point theorem that there is a unique $\mathsf{m} \in M(X)$ with $\mathsf{m} = \mathsf{m}P^k$, and

$$\|\nu P^{kl} - \mathsf{m}\|_1 \le \tau^l \|\nu - \mathsf{m}\|_1 \le 2\tau^l.$$

In particular, for $\nu = \mathsf{m}P \in M(X)$ we obtain

$$\mathsf{m}P = (\mathsf{m}P^{kl})P = (\mathsf{m}P)P^{kl} \to \mathsf{m}, \quad \text{as } l \to \infty,$$

whence $\mathsf{m}P = \mathsf{m}$. Theorem 3.19 implies positive recurrence, and $\mathsf{m}(x) = 1/\mathsf{E}_x(t^x)$. [Note: it is only for this last conclusion that we use Theorem 3.19 here, while for the rest, the latter theorem and its proof are not essential in the present section.]

If $n \in \mathbb{N}$, write $n = kl + r$, where $r \in \{0, \dots, k-1\}$. Then, for $\nu \in M(X)$,

$$\|\nu P^n - \mathsf{m}\|_1 = \|(\nu P^r)P^{kl} - \mathsf{m}\| \le 2\tau^l \le 2\bar\tau^n,$$

where $\bar\tau = \tau^{1/(k+1)}$. □

We can apply this theorem to Example 1.1. We have

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{pmatrix},$$

and $\tau(P) = 1/2$. Therefore, with $\nu = \delta_x$, $x \in X$,

$$\sum_{y \in X} |p^{(n)}(x, y) - \mathsf{m}(y)| \le 2^{-n+1},$$

where

$$\big(\mathsf{m}(☼), \mathsf{m}(☁), \mathsf{m}(❄)\big) = \big(\tfrac{1}{5}, \tfrac{2}{5}, \tfrac{2}{5}\big).$$

One of the important features of Theorem 3.26 is that it provides an estimate for the speed of convergence of $p^{(n)}(x,\cdot)$ to m, and that convergence is exponentially fast.

3.27 Exercise. Let $X = \mathbb{N}$ and $p(k,k+1) = p(k,1) = 1/2$ for all $k \in \mathbb{N}$, while all other transition probabilities are 0. (Draw the graph of this Markov chain.) Compute the stationary probability measure and estimate the speed of convergence. □

Let us now see how Theorem 3.26 can be applied to Markov chains with finite state space.

3.28 Theorem. *Let (X,P) be an irreducible, aperiodic Markov chain with finite state space. Then there are $k \in \mathbb{N}$ and $\bar\tau < 1$ such that for the stationary probability measure $\mathsf{m}(y) = 1/\mathsf{E}_y(t^y)$ one has*

$$\sum_{y\in X} |p^{(n)}(x,y) - \mathsf{m}(y)| \le 2\bar\tau^n$$

for every $x \in X$ and $n \ge k$.

Proof. We show that $\tau(P^k) < 1$ for some $k \in \mathbb{N}$.

Given $x, y \in X$, by irreducibility we can find $m = m_{x,y} \in \mathbb{N}$ such that $p^{(m)}(x,y) > 0$. Finiteness of X allows us to define

$$k_1 = \max\{m_{x,y} : x, y \in X\}.$$

By Lemma 2.22, for each $x \in X$ there is $\ell = \ell_x$ such that $p^{(q)}(x,x) > 0$ for all $q \ge \ell_x$. Define

$$k_2 = \max\{\ell_x : x \in X\}.$$

Let $k = k_1 + k_2$. If $x, y \in X$ and $n \ge k$ then $n = m + q$, where $m = m_{x,y}$ and $q \ge k_2 \ge \ell_y$. Consequently,

$$p^{(n)}(x,y) \ge p^{(m)}(x,y)\, p^{(q)}(y,y) > 0.$$

We have proved that for each $n \ge k$, all matrix elements of P^n are strictly positive. In particular, $\tau(P^k) < 1$, and Theorem 3.26 applies. □

A more precise estimate of the rate of convergence to 0 of $\|\nu P^n - \mathsf{m}\|_1$, for finite X and irreducible, aperiodic P is provided by the Perron–Frobenius theorem: $\|\nu P^n - \mathsf{m}\|_1$ can be compared with λ_*^n, where $\lambda_* = \{\max|\lambda_i| : \lambda_i < 1\}$ and the λ_i are the eigenvalues of the matrix P. We shall come back to this topic for specific classes of Markov chains in a later chapter, and conclude this section with an important consequence of the preceding results, regarding the spectrum of P.

3.29 Theorem. (1) *If X is finite and P is irreducible, then $\lambda = 1$ is an eigenvalue of P. The left eigenspace consists of all constant multiples of the unique stationary probability measure m, while the right eigenspace consists of all constant functions on X.*

(2) *If $d = d(P)$ is the period of P, then the (complex) eigenvalues λ of P with $|\lambda| = 1$ are precisely the d-th roots of unity $\lambda_j = e^{2\pi \mathrm{i} j/d}$, $j = 0, \dots, d-1$. All other eigenvalues satisfy $|\lambda| < 1$.*

Proof. (1) We start with the right eigenspace and use a tool that we shall re-prove in more generality later on under the name of the *maximum principle*. Let $f : X \to \mathbb{C}$ be a function such that $Pf = f$ in the notation of (3.16) – a *harmonic function*. Since both P and $\lambda = 1$ are real, the real and imaginary parts of f must also be harmonic. That is, we may assume that f is real-valued. We also have $P^n f = f$ for each n.

Since X is finite, there is x such that $f(x) \geq f(y)$ for all $y \in X$. We have

$$\sum_{y \in X} p^{(n)}(x,y) \underbrace{\big(f(x) - f(y)\big)}_{\geq 0} = f(x) - P^n f(x) = 0.$$

Therefore $f(y) = f(x)$ for each y with $p^{(n)}(x,y) > 0$. By irreducibility, we can find such an n for every $y \in X$, so that f must be constant.

Regarding the left eigenspace, we know already from Theorem 3.26 that all *non-negative* left eigenvectors must be constant multiples of the unique stationary probability measure m. In order to show that there may be no other ones, we use a method that we shall also elaborate in more detail in a later chapter, namely *time reversal*. We define the m-reverse $\widehat{P}$ of P by

$$\hat{p}(x,y) = \mathsf{m}(y) p(y,x) / \mathsf{m}(x). \tag{3.30}$$

It is well-defined since $\mathsf{m}(x) > 0$, and stochastic since m is stationary. Also, it inherits irreducibility from P. Indeed, the graph $\Gamma(\widehat{P})$ is obtained from $\Gamma(P)$ by inverting the orientation of each edge. This operation preserves strong connectedness. Thus, we can apply the above to $\widehat{P}$: every right eigenfunction of $\widehat{P}$ is constant. The following is easy.

$$\begin{aligned} &\text{A (complex) measure } \nu \text{ on } X \text{ satisfies } \nu P = \nu \text{ if and only if its} \\ &\text{density } f(x) = \nu(x)/\mathsf{m}(x) \text{ with respect to } \mathsf{m} \text{ satisfies } \widehat{P} f = f. \end{aligned} \tag{3.31}$$

Therefore f must be constant, and $\nu = c \cdot \mathsf{m}$.

(2) First, suppose that (X, P) is aperiodic. Let $\lambda \in \mathbb{C}$ be an eigenvalue of P with $|\lambda| = 1$, and let $f : X \to \mathbb{C}$ be an associated eigenfunction. Then $|f| = |Pf| \leq P|f| \leq P^n|f|$ for each $n \in \mathbb{N}$. Finiteness of X and the convergence

theorem imply that

$$|f(x)| \le \lim_{n\to\infty} P^n|f|(x) = \sum_{y\in X} |f(y)|\,\mathsf{m}(y)$$

for each $x \in X$. If we choose x such that $|f(x)|$ is maximal, then we obtain

$$\sum_{y\in X} \big(|f(x)| - |f(y)|\big)\,\mathsf{m}(y) = 0,$$

and we see that $|f|$ is constant, without loss of generality $|f| \equiv 1$. There is n such $p^{(n)}(x, y) > 0$ for all x, y. Therefore

$$\lambda^n f(x) = \sum_{y\in X} p^{(n)}(x, y)\, f(y)$$

is a strict convex combination of the numbers $f(y) \in \mathbb{C}$, $y \in X$, which all lie on the unit circle. If these numbers did not all coincide then $\lambda^n f(x)$ would have to lie in the interior of the circle, a contradiction. Therefore f is constant, and $\lambda = 1$.

Now let $d = d(P)$ be arbitrary, and let $C_0, \dots, C_{d-1}$ be the periodic classes according to Theorem 2.24. Then we know that the restriction Q_k of P^d to C_k is stochastic, irreducible and aperiodic for each k. If $Pf = \lambda f$ on X, where $|\lambda| = 1$, then the restriction f_k of f to C_k satisfies $Q_k f_k = \lambda^d f_k$. By the above, we conclude that $\lambda^d = 1$. Conversely, if we define f on X by $f \equiv \lambda_j^k$ on C_k, where $\lambda_j = e^{2\pi \mathrm{i} j/d}$, then $Pf = \lambda_j f$, and λ_j is an eigenvalue of P.

Stochasticity implies that all other eigenvalues satisfy $|\lambda| < 1$. □

3.32 Exercise. Prove (3.31). □

D The Perron–Frobenius theorem

We now make a small detour into more general matrix analysis. Theorems 3.28 and 3.29 are often presented as special cases of the Perron–Frobenius theorem, which is a landmark of matrix analysis; see SENETA [Se]. Here we take a reversed viewpoint and regard those theorems as the first part of its proof.

We consider a non-negative, finite square matrix A, which we write $A = \big(a(x, y)\big)_{x,y\in X}$ (instead of $(a_{ij})_{i,j=1,\dots,N}$) in order to stay close to our usual notation. The set X is assumed finite. The n-th matrix power is written $A^n = \big(a^{(n)}(x, y)\big)_{x,y\in X}$. As usual, we think of column vectors as functions and of row vectors as measures on X. The definition of *irreducibility* remains the same as for stochastic matrices.

3.33 Theorem. *$A = \big(a(x,y)\big)_{x,y\in X}$ be a finite, irreducible, non-negative matrix, and let*

$$\rho(A) = \limsup_{n\to\infty} a^{(n)}(x,y)^{1/n}$$

Then $\rho(A) > 0$ is independent of x and y, and

$$\rho(A) = \min\{t > 0 \mid \textit{there is } g\colon X \to (0,\infty) \textit{ with } Ag \le t\cdot g\}.$$

Proof. The property that $\limsup_{n\to\infty} a^{(n)}(x,y)^{1/n}$ is the same for all $x, y \in X$ follows from irreducibility exactly as for Markov chains. Choose x and y in X. By irreducibility, $a^{(k)}(x,y) > 0$ and $a^{(l)}(y,x) > 0$ for some $k, l > 0$. Therefore $\alpha = a^{(k+l)}(x,x) > 0$, and $a^{(n(k+l))}(x,x) \ge \alpha^n$. We deduce that $\rho(A) \ge \alpha^{1/(k+l)} > 0$.

In order to prove the "variational characterization" of $\rho(A)$, let

$$t_0 = \inf\{t > 0 \mid \text{there is } g\colon X \to (0,\infty) \text{ with } Ag \le t\cdot g\}.$$

If $Ag \le t\cdot g$, where g is a positive function on X, then also $A^n g \le t^n \cdot g$. Thus $a^{(n)}(x,y)\, g(y) \le t^n g(x)$, whence

$$a^{(n)}(x,y)^{1/n} \le t\,\big(g(x)/g(y)\big)^{1/n}$$

for each n. This implies $\rho(A) \le t$, and therefore $\rho(A) \le t_0$.

Next, let $G(x,y|z) = \sum_{n=0}^{\infty} a^{(n)}(x,y)\, z^n$. This power series has radius of convergence $1/\rho(A)$, and

$$G(x,y|z) = \delta_x(y) + \sum_w a(x,w)\, z\, G(w,y|z).$$

Now let $t > \rho(A)$, fix $y \in X$ and set

$$g_t(x) = G(x,y|1/t)\big/G(y,y|1/t).$$

Then we see that $g_t(x) > 0$ and $Ag_t(x) \le t\cdot g_t(x)$ for all x. Therefore $t_0 = \rho(A)$.

We still need to show that the infimum is a minimum. Note that $g_t(y) = 1$ for our fixed y. We choose a strictly decreasing sequence $(t_k)_{k\ge1}$ with limit $\rho(A)$ and set $g_k = g_{t_k}$. Then, for each $n \in \mathbb{N}$ and $x \in X$,

$$t_1^n \ge t_k^n = t_k^n \cdot g_k(y) \ge A^n g_k(y) \ge a^{(n)}(y,x) g_k(x).$$

For each x there is n_x such that $a^{(n_x)}(y,x) > 0$. Therefore

$$g_k(x) \le C_x = t_1^{n_x}/a^{(n_x)}(y,x) < \infty \quad \text{for all } k.$$

By the Heine–Borel theorem, there must be a subsequence $\big(g_{k(m)}\big)_{m\ge1}$ that converges pointwise to a limit function h. We have for each $x \in X$

$$\sum_{w\in X} a(x,w)\, g_{k(m)}(w) \le t_{k(m)}\, g_{k(m)}(x).$$

We can pass to the limit as $m \to \infty$ and obtain $Ah \le \rho(A) \cdot h$. Furthermore, $h \ge 0$ and $h(y) = 1$. Therefore $\rho(A)^n h(x) \ge a^{(n)}(x, y)h(y) > 0$ if n is such that $a^{(n)}(x, y) > 0$. Thus, the infimum is indeed a minimum. □

3.34 Definition. The matrix A is called *primitive* if there exists an n_0 such that $a^{(n_0)}(x, y) > 0$ for all $x, y \in X$.

Since X is finite, this amounts to "irreducible & aperiodic" for stochastic matrices (finite Markov chains), compare with the proof of Theorem 3.28.

3.35 Perron–Frobenius theorem. *Let $A = \big(a(x, y)\big)_{x,y\in X}$ be a finite, irreducible, non-negative matrix. Then*

(a) *$\rho(A)$ is an eigenvalue of A, and $|\lambda| \le \rho(A)$ for every eigenvalue λ of A.*

(b) *There is a strictly positive function h on X that spans the right eigenspace of A with respect to the eigenvalue $\rho(A)$. Furthermore, if a function $f : X \to [0, \infty)$ satisfies $Af \le \rho(A) \cdot f$ then $f = c \cdot h$ for some constant c.*

(c) *There is a strictly positive measure ν on X that spans the left eigenspace of A with respect to the eigenvalue $\rho(A)$. Furthermore, if a non-negative measure ν on X satisfies $\mu A \le \rho(A) \cdot \mu$ then $\mu = c \cdot \nu$ for some constant c.*

(d) *If in addition A is primitive, and ν and h are normalized such that $\sum_x h(x)\,\nu(x) = 1$ then $|\lambda| < \rho(A)$ for every $\lambda \in \operatorname{spec}(A) \setminus \{\rho(A)\}$, and*

$$\lim_{n\to\infty} a^{(n)}(x, y)/\rho(A)^n = h(x)\,\nu(y) \quad \text{for all } x, y \in X.$$

Proof. We know from Theorem 3.33 that there is a positive function h on X that satisfies $Ah \le \rho(A) \cdot h$. We define a new matrix P over X by

$$p(x, y) = \frac{a(x, y)\,h(y)}{\rho(A)\,h(x)}. \tag{3.36}$$

This matrix is substochastic. Furthermore, it inherits irreducibility from A. If $D = D_h$ is the diagonal matrix with diagonal entries $h(x)$, $x \in X$, then

$$P = \frac{1}{\rho(A)} D^{-1} A D, \quad \text{whence} \quad P^n = \frac{1}{\rho(A)^n} D^{-1} A^n D \ \text{ for every } n \in \mathbb{N}.$$

That is,

$$p^{(n)}(x, y) = \frac{a^{(n)}(x, y)\,h(y)}{\rho(A)^n\,h(x)}. \tag{3.37}$$

Taking n-th roots and the lim sup as $n \to \infty$, we see that $\rho(P) = 1$. Now Proposition 2.31 tells us that P must be stochastic. But this is equivalent with

$Ah = \rho(A) \cdot h$. Therefore $\rho(A)$ is an eigenvalue of A, and h belongs to the right eigenspace. Now, similarly to (3.31),

$$\begin{aligned} &\text{A (complex) function } g \text{ on } X \text{ satisfies } Ag = \lambda \cdot g \text{ if and only if} \\ &f(x) = g(x)/h(x) \text{ satisfies } Pf = \frac{\lambda}{\rho(A)} \cdot f. \end{aligned} \tag{3.38}$$

Since P is stochastic, $|\lambda/\rho(A)| \le 1$. Thus, we have proved (a).

Furthermore, if $\lambda = \rho(A)$ in (3.38) then $Pf = f$, and Theorem 3.29 implies that f is constant, $f \equiv c$. Therefore $g = c \cdot h$, which shows that the right eigenspace of A with respect to the eigenvalue $\rho(A)$ is spanned by h. Next, let $g \neq 0$ be a non-negative function with $Ag \le \rho(A) \cdot g$. Then $A^n g \le \rho(A)^n \cdot g$ for each n, and irreducibility of A implies that $g > 0$. We can replace h with g in our initial argument, and obtain that $Ag = \rho(A) \cdot g$. But this yields that g is a multiple of h. We have completed the proof of (b).

Statement (c) follows by replacing A with the transposed matrix A^t.

Finally, we prove (d). Since A is primitive, P is irreducible and aperiodic. The fact that $|\lambda| < \rho(A)$ for each eigenvalue $\lambda \neq \rho(A)$ of A follows from Theorem 3.29 via the arguments used to prove (b): namely, λ is an eigenvalue of A if and only if $\lambda/\rho(A)$ is an eigenvalue of P.

Suppose that ν and h are normalized as proposed. Then $\mathsf{m}(x) = \nu(x)h(x)$ is a probability measure on X, and we can write $\mathsf{m} = \nu D$. Therefore

$$\mathsf{m}P = \frac{1}{\rho(A)} \nu D D^{-1} A D = \frac{1}{\rho(A)} \nu A D = \nu D = \mathsf{m}.$$

We see that m is the unique stationary probability measure for P. Theorem 3.28 implies that $p^{(n)}(x, y) \to \mathsf{m}(y) = h(y)\nu(y)$. Combining this with (3.37), we get the proposed asymptotic formula. □

The following is usually also considered as part of the Perron–Frobenius theorem.

3.39 Proposition. *Under the assumptions of Theorem 3.35, $\rho(A)$ is a simple root of the characteristic polynomial of the matrix A.*

Proof. Recall the definition and properties of the *adjunct matrix* $\hat{A}$ of a square matrix (not to be confused with the adjoint matrix $\bar{A}^t$). Its elements have the form

$$\hat{a}(x, y) = (-1)^{\epsilon} \det(A \mid y, x),$$

where $(A \mid y, x)$ is obtained from A by deleting the row of y and the column of x, and $\epsilon = 0$ or $= 1$ according to the parity of the position of (x, y); in particular $\epsilon = 0$ when $y = x$. For $\lambda \in \mathbb{C}$, let $A_\lambda = \lambda I - A$ and $\hat{A}_\lambda$ its adjunct. Then

$$A_\lambda \hat{A}_\lambda = \hat{A}_\lambda A_\lambda = \chi_A(\lambda) I, \tag{3.40}$$

where $\chi_A(\lambda)$ is the characteristic polynomial of A. If we set $\lambda = \rho = \rho(A)$, then we see that each column of $\hat{A}_\rho$ is a right ρ-eigenvector of A, and each row is a left ρ-eigenvector of A. Let h and ν, respectively, be the (positive) right and left eigenvectors of A that we have found in Theorem 3.35, normalized such that $\sum_x h(x)\nu(x) = 1$.

3.41 Exercise. Deduce that $\hat{a}_\rho(x, y) = \alpha \cdot h(x)\, \nu(y)$, where $\alpha \in \mathbb{R}$, whence $\hat{A}_\rho h = \alpha \cdot h$. □

We continue the proof of the proposition by showing that $\alpha > 0$. Consider the matrix A_x over X obtained from A by replacing all elements in the row and the column of x with 0. Then $A_x \leq A$ elementwise, and the two matrices do not coincide. Exercise 3.43 implies that $\rho > |\lambda|$ for every eigenvalue λ of A_x, that is, $\det(\rho\, I - A_x) > 0$. (This is because the leading coefficient of the characteristic polynomial is 1, so that the polynomial is positive for real arguments that are bigger than its biggest real root.) Since $\det(\rho\, I - A_x) = \rho\, \hat{a}_\rho(x, x)$, we see indeed that $\alpha > 0$.

We now differentiate (3.40) with respect to λ and set $\lambda = \rho$: writing $\hat{A}'_\rho$ for that elementwise derivative, since $A'_\rho = I$, the product rule yields

$$\hat{A}'_\rho\, A_\rho + \hat{A}_\rho = \chi'_A(\rho) \cdot I.$$

We get

$$\chi'_A(\rho) \cdot h = \hat{A}'_\rho \underbrace{A_\rho h}_{=0} + \hat{A}_\rho h = \alpha \cdot h.$$

Therefore $\chi'_A(\rho) = \alpha > 0$, and ρ is a simple root of $\chi_A(\cdot)$. □

3.42 Proposition. *Let X be finite and A, B be two non-negative matrices over X. Suppose that A is irreducible and that $b(x, y) \leq a(x, y)$ for all x, y. Then*

$$\max\{|\lambda| : \lambda \in \operatorname{spec}(B)\} \leq \rho(A).$$

Proof. Let $\lambda \in \operatorname{spec}(B)$ and $f : X \to \mathbb{C}$ an associated eigenfunction (right eigenvector). Then

$$|\lambda| \cdot |f| = |Bf| \leq B|f| \leq A\,|f|.$$

Now let ν be as in Theorem 3.35 (c). Then

$$|\lambda| \sum_{x \in X} \nu(x)|f(x)| \leq \nu A\,|f| = \rho(A) \sum_{x \in X} \nu(x)|f(x)|.$$

Therefore $|\lambda| \leq \rho(A)$. □

3.43 Exercise. Show that in the situation of Proposition 3.42, one has that $\max\{|\lambda| : \lambda \in \mathrm{spec}(B)\} = \rho(A)$ if and only if $B = A$.

[Hint: define P as in (3.36), where $Ah = \rho(A) \cdot h$, and let

$$q(x, y) = \frac{b(x, y)h(y)}{\rho(A)h(x)}.$$

Show that $A = B$ if and only if Q is stochastic. Then assume that $B \neq A$ and that B is irreducible, and show that $\rho(B) < \rho(A)$ in this case. Finally, when B is not irreducible, replace B by a slightly bigger matrix that is irreducible and dominated by A.] □

There is a weak form of the Perron–Frobenius theorem for general non-negative matrices.

3.44 Proposition. *Let $A = \big(a(x, y)\big)_{x,y\in X}$ be a finite, non-zero, non-negative matrix. Then A has a positive eigenvalue $\rho = \rho(A)$ with non-negative (non-zero) left and right eigenvectors ν and h, respectively, such that $|\lambda| \leq \rho$ for every eigenvalue of A.*

Furthermore, $\rho = \max_C \rho(A_C)$, where C ranges over all irreducible classes with respect to A, the matrix A_C is the restriction of A to C, and $\rho(A_C)$ is the Perron–Frobenius eigenvalue of the irreducible matrix A_C.

Proof. Let E be the matrix with all entries $= 1$. Let $A_n = A + \frac{1}{n}E$, a non-negative, irreducible matrix. Let $h_n > 0$ and $\nu_n > 0$ be such that $A_n h_n = \rho(A_n) \cdot h_n$ and $\nu_n A_n = \rho(A_n) \cdot \nu_n$. Normalize h_n and ν_n such that $\sum_x h_n(x) = \sum_x \nu_n(x) = 1$. By compactness, there are a subsequence (n') and a function (column vector) h, as well as a measure (row vector) ν such that $h_{n'} \to h$ and $\nu_{n'} \to \nu$. We have $h \geq 0$, $\nu \geq 0$, and $\sum_x h(x) = \sum_x \nu(x) = 1$, so that $h, \nu \neq 0$.

By Proposition 3.42, the sequence $\big(\rho(A_n)\big)$ is decreasing. Let ρ be its limit. Then $Ah = \rho \cdot h$ and $\nu A = \rho \cdot \nu$. Also by Proposition 3.42, every $\lambda \in \mathrm{spec}(A)$ satisfies $|\lambda| \leq \rho(A_n)$. Therefore

$$\max\{|\lambda| : \lambda \in \mathrm{spec}(A)\} \leq \rho,$$

as proposed.

If $h(y) > 0$ and $a(x, y) > 0$ then $h(x) \geq \rho\, a(x, y)\, h(y) > 0$. Therefore $h(x) > 0$ for all $x \in C(y)$, the irreducible class of y with respect to the matrix A. We see that the set $\{x \in X : h(x) > 0\}$ is the union of irreducible classes. Let C_0 be such an irreducible class on which $h > 0$, and which is maximal with this property in the partial order on the collection of irreducible classes. Let h_{C_0} be the restriction of h to C_0. Then $A_{C_0} h_{C_0} = \rho \cdot h_{C_0}$. This implies (why?) that $\rho = \rho(A_{C_0})$.

Finally, if C is any irreducible class, then consider the truncated matrix A_C as a matrix on the whole of X. It is dominated by A, whence by A_n, which implies via Proposition 3.42 that $\rho(A_C) \leq \rho(A_n)$ for each n, and in the limit, $\rho(A_C) \leq \rho$. □

3.45 Exercise. Verify that the "variational characterization" of Theorem 3.33 for $\rho(A)$ also holds when A is an *infinite* non-negative irreducible matrix, as long as $\rho(A)$ is finite. (The latter is true, e.g., when A has bounded row sums or bounded column sums, and in particular, when A is substochastic.) □

E The convergence theorem for positive recurrent Markov chains

Here, we shall (re)prove the analogue of Theorem 3.28 for an arbitrary Markov chain (X, P) that is irreducible, aperiodic and positive recurrent, when X is not necessarily finite, nor $\tau(P^k) < 1$ for some k. The "price" that we have to pay is that we shall not get exponential decay in the approximation of the stationary probability by νP^n. Except for this last fact, we could of course have omitted the extra Section C regarding the finite case and appeal directly to the main theorem that we are going to prove. However, the method used in Section C is interesting on its own right, which is another reason for having chosen this slight redundancy in the presentation of the material.

We shall also determine the convergence behaviour of n-step transition probabilities when (X, P) is null recurrent.

The clue for dealing with the positive recurrent case is the following: we consider two independent versions (Z_n^1) and (Z_n^2) of the Markov chain, one with arbitrary initial distribution ν, and the other with initial distribution m, the stationary probability measure of P. Thus, Z_n^2 will have distribution m for each n. We then consider the stopping time $\boldsymbol{t}^D$ when the two chains first meet. On the event $[\boldsymbol{t}^D \leq n]$, it will be easily seen that Z_n^1 and Z_n^2 have the same distribution. The method also adapts to the null recurrent case.

What we are doing here is to apply the so-called *coupling method:* we construct a larger probability space on which both processes can be defined in such a way that they can be compared in a suitable way. This type of idea has many fruitful applications in probability theory, see the book by Lindvall [Li].

We now elaborate the details of that plan. We consider the new state space $X \times X$ with transition matrix $Q = P \otimes P$ given by

$$q\big((x_1, x_2), (y_1, y_2)\big) = p(x_1, y_1)\, p(x_2, y_2).$$

3.46 Lemma. *If (X, P) is irreducible and aperiodic, then the same holds for the Markov chain $(X \times X, P \otimes P)$. Furthermore, if (X, P) is positive recurrent, then so is $(X \times X, P \otimes P)$.*

Proof. We have

$$q^{(n)}\big((x_1,x_2),(y_1,y_2)\big) = p^{(n)}(x_1,y_1)\,p^{(n)}(x_2,y_2).$$

By irreducibility and aperiodicity, Lemma 2.22 implies that there are indices $n_i = n(x_i,y_i)$ such that $p^{(n)}(x_i,y_i) > 0$ for all $n \geq n_i$, $i = 1,2$. Therefore also Q is irreducible and aperiodic.

If (X,P) is positive recurrent and m is the unique invariant probability measure of P, then $\mathsf{m}\times\mathsf{m}$ is an invariant probability measure for Q. By Theorem 3.19, Q is positive recurrent. □

Let ν_1 and ν_2 be probability measures on X. If we write (Z_n^1, Z_n^2) for the Markov chain with initial distribution $\nu_1\times\nu_2$ and transition matrix Q on $X\times X$, then for $i = 1,2$, (Z_n^i) is the Markov chain on X with initial distribution ν_i and transition matrix P.

Now let $D = \{(x,x) : x \in X\}$ be the diagonal of $X\times X$, and $\boldsymbol{t}^D$ the associated stopping time with respect to Q. This is the time when (Z_n^1) and (Z_n^2) first meet (after starting).

3.47 Lemma. *For every $x \in X$ and $n \in \mathbb{N}$,*

$$\mathsf{Pr}_{\nu_1\times\nu_2}[Z_n^1 = x,\ \boldsymbol{t}^D \leq n] = \mathsf{Pr}_{\nu_1\times\nu_2}[Z_n^2 = x,\ \boldsymbol{t}^D \leq n].$$

Proof. This is a consequence of the Strong Markov Property of Exercise 1.25. In detail, if $k \leq n$, then

$$\begin{aligned}\mathsf{Pr}_{\nu_1\times\nu_2}[Z_n^1 = x,\ \boldsymbol{t}^D = k] &= \sum_{w\in X} \mathsf{Pr}_{\nu_1\times\nu_2}[Z_n^1 = x,\ Z_k^1 = w,\ \boldsymbol{t}^D = k]\\ &= \sum_{w\in X} p^{(n-k)}(w,x)\ \mathsf{Pr}_{\nu_1\times\nu_2}[Z_k^1 = w,\ \boldsymbol{t}^D = k]\\ &= \mathsf{Pr}_{\nu_1\times\nu_2}[Z_n^2 = x,\ \boldsymbol{t}^D = k],\end{aligned}$$

since by definition, $Z_k^1 = Z_k^2$, if $\boldsymbol{t}^D = k$. Summing over all $k \leq n$, we get the proposed identity. □

3.48 Theorem. *Suppose that (X,P) is irreducible and aperiodic.*

(a) *If (X,P) is positive recurrent, then for any initial distribution ν on X,*

$$\lim_{n\to\infty} \|\nu P^n - \mathsf{m}\|_1 = 0,$$

and in particular

$$\lim_{n\to\infty} \mathsf{Pr}_\nu[Z_n = x] = \mathsf{m}(x) \quad \textit{for every } x \in X,$$

where $\mathsf{m}(x) = 1/\mathsf{E}_x(\boldsymbol{t}^x)$ is the stationary probability distribution.

(b) *If (X, P) is null recurrent or transient, then for any initial distribution ν on X,*

$$\lim_{n\to\infty} \mathsf{Pr}_\nu[Z_n = x] = 0 \quad \textit{for every } x \in X.$$

Proof. (a) In the positive recurrent case, we set $\nu_1 = \nu$ and $\nu_2 = \mathsf{m}$ for defining the initial measure of (Z_n^1, Z_n^2) on $X \times X$ in the above construction. Since $\boldsymbol{t}^D \le \boldsymbol{t}^{(x,x)}$, Lemma 3.46 implies that

$$\boldsymbol{t}^D < \infty \quad \mathsf{Pr}_{\nu_1\times\nu_2}\text{-almost surely.} \tag{3.49}$$

(This probability measure lives on the trajectory space associated with Q over $X \times X$!)

We have $\|\nu P^n - \mathsf{m}\|_1 = \|\nu_1 P^n - \nu_2 P^n\|_1$. Abbreviating $\mathsf{Pr}_{\nu_1\times\nu_2} = \mathsf{Pr}$ and using Lemma 3.47, we get

$$\begin{aligned}
\|\nu_1 P^n - \nu_2 P^n\|_1 &= \sum_{y\in X} \big|\mathsf{Pr}[Z_n^1 = y] - \mathsf{Pr}[Z_n^2 = y]\big| \\
&= \sum_{y\in X} \big|\mathsf{Pr}[Z_n^1 = y,\ \boldsymbol{t}^D \le n] + \mathsf{Pr}[Z_n^1 = y,\ \boldsymbol{t}^D > n] \\
&\qquad - \mathsf{Pr}[Z_n^2 = y,\ \boldsymbol{t}^D \le n] - \mathsf{Pr}[Z_n^2 = y,\ \boldsymbol{t}^D > n]\big| \\
&\le \sum_{y\in X} \big(\mathsf{Pr}[Z_n^1 = y,\ \boldsymbol{t}^D > n] + \mathsf{Pr}[Z_n^2 = y,\ \boldsymbol{t}^D > n]\big) \\
&\le 2\ \mathsf{Pr}[\boldsymbol{t}^D > n].
\end{aligned} \tag{3.50}$$

By (3.49), we have that $\mathsf{Pr}[\boldsymbol{t}^D > n] \to \mathsf{Pr}[\boldsymbol{t}^D = \infty] = 0$ as $n \to \infty$.

(b) If (X, P) is transient then statement (b) is immediate, see Exercise 3.7.

So let us suppose that (X, P) is null recurrent. In this case, it may happen that $(X \times X, Q)$ becomes transient. [Example: take for (X, P) the simple random walk on $\mathbb{Z}^2$; see (4.64) in Section 4.E.] Then, setting $\nu_1 = \nu_2 = \nu$, we get that

$$\big(\mathsf{Pr}_\nu[Z_n = x]\big)^2 = \mathsf{Pr}_{\nu_1\times\nu_2}[(Z_n^1, Z_n^2) = (x, x)] \to 0 \quad \text{as } n \to \infty,$$

once more by Exercise 3.7. (Note again that Pr_ν and $\mathsf{Pr}_{\nu_1\times\nu_2}$ live on different trajectory spaces !)

Since (X, P) is a factor chain of $(X \times X, Q)$, the latter cannot be positive recurrent when (X, P) is null recurrent; see Exercise 3.14. Therefore the remaining case that we have to consider is the one where both (X, P) and the product chain $(X \times X, Q)$ are null recurrent. Then we first claim that for any choice of initial probability distributions ν_1 and ν_2 on X,

$$\lim_{n\to\infty} \big(\mathsf{Pr}_{\nu_1}[Z_n = x] - \mathsf{Pr}_{\nu_2}[Z_n = x]\big) = 0 \quad \text{for every } x \in X.$$

Indeed, (3.49) is valid in this case, and we can use once more (3.50) to find that

$$\|\nu_1 P^n - \nu_2 P^n\|_1 \le 2\ \mathsf{Pr}[\boldsymbol{t}^D > n] \to 0.$$

Setting $\nu_1 = \nu$ and $\nu_2 = \nu P^m$ and replacing n with $n-m$, this implies in particular that

$$\lim_{n\to\infty} \big(\mathsf{Pr}_\nu[Z_n = x] - \mathsf{Pr}_\nu[Z_{n-m} = x]\big) = 0 \quad \text{for every } x \in X,\ m \ge 0. \tag{3.51}$$

The remaining arguments involve only the basic trajectory space associated with (X, P). Let $\varepsilon > 0$. We use the "null" of null recurrence, namely, that

$$\mathsf{E}_x(\boldsymbol{t}^x) = \sum_{m=0}^{\infty} \mathsf{Pr}_x[\boldsymbol{t}^x > m] = \infty.$$

Hence there is $M = M_\varepsilon$ such that

$$\sum_{m=0}^{M} \mathsf{Pr}_x[\boldsymbol{t}^x > m] > 1/\varepsilon.$$

For $n \ge M$, the events $A_m = [Z_{n-m} = x,\ Z_{n-m+k} \ne x \text{ for } 1 \le k \le m]$ are pairwise disjoint for $m = 0, \dots, M$. Thus, using the Markov property,

$$\begin{aligned} 1 &\ge \sum_{m=0}^{M} \mathsf{Pr}_\nu(A_m) \\ &= \sum_{m=0}^{M} \mathsf{Pr}_\nu[Z_{n-m+k} \ne x \text{ for } 1 \le k \le m \mid Z_{n-m} = x]\ \mathsf{Pr}_\nu[Z_{n-m} = x] \\ &= \sum_{m=0}^{M} \mathsf{Pr}_x[\boldsymbol{t}^x > m]\ \mathsf{Pr}_\nu[Z_{n-m} = x]. \end{aligned}$$

Therefore, for each $n \ge M$ there must be $m = m(n) \in \{0, \dots, M\}$ such that $\mathsf{Pr}_\nu[Z_{n-m(n)} = x] < \varepsilon$. But $\mathsf{Pr}_\nu[Z_n = x] - \mathsf{Pr}_\nu[Z_{n-m(n)} = x] \to 0$ by (3.51) and boundedness of $m(n)$. Thus

$$\limsup_{n\to\infty} \mathsf{Pr}_\nu[Z_n = x] \le \varepsilon$$

for every $\varepsilon > 0$, proving that $\mathsf{Pr}_\nu[Z_n = x] \to 0$. □

3.52 Exercise. Let (X, P) be a recurrent irreducible Markov chain (X, P), d its period, and $C_0, \dots, C_{d-1}$ its periodic classes according to Theorem 2.24.

(1) Show that (X, P) is positive recurrent if and only if the restriction of P^d to C_i is positive recurrent for some ($\Longleftrightarrow$ all) $i \in \{0, \dots, d-1\}$.

(2) Show that for $x, y \in C_i$,

$$\lim_{n\to\infty} p^{(nd)}(x, y) = d \cdot \mathsf{m}(y) \quad \text{with } \mathsf{m}(y) = 1/\mathsf{E}_y(\boldsymbol{t}^y).$$

(3) Determine the limiting behaviour of $p^{(n)}(x, y)$ for $x \in C_i,\ y \in C_j$. □

In the hope that the reader will have solved this exercise before proceeding, we now consider the general situation. Theorem 3.48 was formulated for an irreducible, positive recurrent Markov chain (X, P). It applies without any change to the restriction of a general Markov chain to any of its essential classes, if the latter is positive recurrent and aperiodic. We now want to find the limiting behaviour of n-step transition probabilities in the case when there may be several essential irreducible classes, as well as non-essential ones.

For $x, y \in X$ and $d = d(y)$, the period of (the irreducible class of) y, we define

$$\begin{aligned} F^r(x, y) &= \mathsf{Pr}_x[\boldsymbol{s}^y < \infty \ \text{ and } \ \boldsymbol{s}^y \equiv r \mod d] \\ &= \sum_{n=0}^{\infty} f^{(nd+r)}(x, y), \quad r = 0, \dots, d-1. \end{aligned} \tag{3.53}$$

We make the following observations.

(i) $F(x, y) = F^0(x, y) + F^1(x, y) + \cdots + F^{d-1}(x, y)$.

(ii) If $x \leftrightarrow y$ then by Theorem 2.24 there is a unique r such that $F(x, y) = F^r(x, y)$, while $F^j(x, y) = 0$ for all other $j \in \{0, 1, \dots, d-1\}$.

(iii) If furthermore y is a recurrent state then Theorem 3.4 (b) implies that $F^r(x, y) = F(x, y) = U(x, y) = 1$ for the index r that we found above in (ii).

(iv) If x and y belong to different classes and $x \to y$, then it may well be that $F^r(x, y) > 0$ for different indices $r \in \{0, 1, \dots, d-1\}$. (Construct examples !)

3.54 Theorem. (a) *Let $y \in X$ be a positive recurrent state and $d = d(y)$ its period. Then for each $x \in X$ and $r \in \{0, 1, \dots, d-1\}$,*

$$\lim_{n\to\infty} p^{(nd+r)}(x, y) = F^r(x, y) \cdot d/\mathsf{E}_y(\boldsymbol{t}^y).$$

(b) *Let $y \in X$ be a transient or null recurrent state. Then*

$$\lim_{n\to\infty} p^{(n)}(x, y) = 0.$$

Proof. (a) By Exercise 3.52, for $\varepsilon > 0$ there is $N_\varepsilon > 0$ such that for every $n \geq N_\varepsilon$, we have $|p^{(nd)}(y,y) - d \cdot \mathsf{m}(y)| < \varepsilon$, where $\mathsf{m}(y) = 1/\mathsf{E}_y(t^y)$. For such n, applying Theorem 1.38 (b) and equation (1.40),

$$\begin{aligned} p^{(nd+r)}(x,y) &= \sum_{\ell=0}^{nd+r} f^{(\ell)}(x,y) \underbrace{p^{(nd+r-\ell)}(y,y)}_{\substack{>0 \text{ only if} \\ \ell - r \equiv 0 \mod d}} \\ &= \sum_{k=0}^{n} f^{(kd+r)}(x,y)\ p^{((n-k)d)}(y,y) \\ &\leq \sum_{k=0}^{n-N_\varepsilon} f^{(kd+r)}(x,y)\left(d \cdot \mathsf{m}(y) + \varepsilon\right) + \sum_{k>n-N_\varepsilon} f^{(kd+r)}(x,y). \end{aligned}$$

Since $\sum_{k>n-N_\varepsilon} f^{(kd+r)}(x,y)$ is a remainder term of a convergent series, it tends to 0, as $n \to \infty$. Hence

$$\limsup_{n\to\infty}\ p^{(nd+r)}(x,y) \leq F^r(x,y)\left(d \cdot \mathsf{m}(y) + \varepsilon\right) \quad \text{for each } \varepsilon > 0.$$

Therefore

$$\limsup_{n\to\infty}\ p^{(nd+r)}(x,y) \leq F^r(x,y) \cdot d \cdot \mathsf{m}(y).$$

Analogously, the inequality

$$p^{(nd+r)}(x,y) \geq \sum_{k=0}^{n-N_\varepsilon} f^{(kd+r)}(x,y)\left(d \cdot \mathsf{m}(y) - \varepsilon\right)$$

yields

$$\liminf_{n\to\infty}\ p^{(nd+r)}(x,y) \geq F^r(x,y) \cdot d \cdot \mathsf{m}(y).$$

This concludes the proof in the positive recurrent case.

(b) If y is transient then $G(x,y) < \infty$, whence $p^{(n)}(x,y) \to 0$. If y is null recurrent then we can apply part (1) of Exercise 3.52 to the restriction of P^d to C_i, the periodic class to which y belongs according to Theorem 2.24. It is again null recurrent, so that $p^{(nd)}(y,y) \to 0$ as $n \to \infty$. The proof now continues precisely as in case (a), replacing the number $\mathsf{m}(y)$ with 0. □

F The ergodic theorem for positive recurrent Markov chains

The purpose of this section is to derive the second important Markov chain limit theorem.

3.55 Ergodic theorem. *Let (X, P) be a positive recurrent, irreducible Markov chain with stationary probability measure $\mathsf{m}(\cdot)$. If $f : X \to \mathbb{R}$ is m-integrable, that is $\int |f| \, d\mathsf{m} = \sum_x |f(x)| \, \mathsf{m}(x) < \infty$, then for any starting distribution,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(Z_n) = \int_X f \, d\mathsf{m} \quad \textit{almost surely.}$$

As a matter of fact, this is a special case of the general ergodic theorem of Birkhoff and von Neumann, see e.g. Petersen [Pe]. Before the proof, we need some preparation. We introduce new probabilities that are "dual" to the $f^{(n)}(x, y)$ which were defined in (1.28):

$$\ell^{(n)}(x, y) = \mathsf{Pr}_x[Z_n = y, \ Z_k \neq x \text{ for } k \in \{1, \dots, n\}] \tag{3.56}$$

is the probability that the Markov chain starting at x is in y at the n-th step before returning to x. In particular, $\ell^{(0)}(x, x) = 1$ and $\ell^{(0)}(x, y) = 0$ if $x \neq y$. We can define the associated generating function

$$L(x, y|z) = \sum_{n=0}^{\infty} \ell^{(n)}(x, y) \, z^n, \quad L(x, y) = L(x, y|1). \tag{3.57}$$

We have $L(x, x|z) = 1$ for all z. Note that while the quantities $f^{(n)}(x, y)$, $n \in \mathbb{N}$, are probabilities of disjoint events, this is not the case for $\ell^{(n)}(x, y)$, $n \in \mathbb{N}$. Therefore, unlike $F(x, y)$, the quantity $L(x, y)$ is not a probability. Indeed, it is the expected number of visits in y before returning to the starting point x.

3.58 Lemma. *If $y \to x$, or if y is a transient state, then $L(x, y) < \infty$.*

Proof. The statement is clear when $x = y$. Since $\ell^{(n)}(x, y) \leq p^{(n)}(x, y)$, it is also obvious when y is a transient state.

We now assume that $y \to x$. When $x \not\to y$, we have $L(x, y) = 0$.

So we consider the case when $x \leftrightarrow y$ and y is recurrent. Let C be the irreducible class of y. It must be essential. Recall the Definition 2.14 of the restriction $P_{C\setminus\{x\}}$ of P to $C \setminus \{x\}$. Factorizing with respect to the first step, we have for $n \geq 1$

$$\ell^{(n)}(x, y) = \sum_{w \in C\setminus\{x\}} p(x, w) \, p^{(n-1)}_{C\setminus\{x\}}(w, y),$$

because the Markov chain starting in x cannot exit from C. Now the Green function of the restriction satisfies

$$G_{C\setminus\{x\}}(w, y) = F_{C\setminus\{x\}}(w, y) \, G_{C\setminus\{x\}}(y, y) \leq G_{C\setminus\{x\}}(y, y) < \infty.$$

Indeed, those quantities can be interpreted in terms of the modified Markov chain where the state x is made absorbing, so that $C \setminus \{x\}$ contains no essential state for

the modified chain. Therefore the associated Green function must be finite. We deduce

$$L(x,y) = \sum_{w \in C\setminus\{x\}} p(x,w)\, G_{C\setminus\{x\}}(w,y) \le G_{C\setminus\{x\}}(y,y) < \infty,$$

as proposed. □

3.59 Exercise. Prove the following in analogy with Theorem 1.38 (b), (c), (d).

$$\begin{aligned} G(x,y|z) &= G(x,x|z)\, L(x,y|z) \quad \text{for all } x,y, \\ U(x,x|z) &= \sum_y L(x,y|z)\, p(y,x)z, \quad \text{and} \\ L(x,y|z) &= \sum_w L(x,w|z)\, p(w,y)z, \quad \text{if } y \neq x \end{aligned}$$

for all z in the common domain of convergence of the power series involved. □

3.60 Exercise. Derive a different proof of Lemma 3.58 in the case when $x \leftrightarrow y$: show that for $|z| < 1$,

$$L(x,y|z)L(y,x|z) = F(x,y|z)F(y,x|z),$$

and let $z \to 1-$.

[Hint: multiply both sides by $G(x,x|z)G(y,y|z)$.] □

3.61 Exercise. Show that when x is a recurrent state,

$$\sum_{y\in X} L(x,y) = \mathsf{E}_x(\boldsymbol{t}^x).$$

[Hint: use again that $\sum_y G(x,y|z) = 1/(1-z)$ and apply the first formula of Exercise 3.59 in the same way as Theorem 1.38 (b) was used in (3.11) and (3.12).] □

Thus, for a positive recurrent chain, $L(x,y)$ has a particularly simple form.

3.62 Corollary. *Let (X,P) be a positive recurrent, irreducible Markov chain with stationary probability measure $\mathsf{m}(\cdot)$. Then for all $x, y \in X$,*

$$L(x,y) = \mathsf{m}(y)/\mathsf{m}(x) = \mathsf{E}_x(\boldsymbol{t}^x)/\mathsf{E}_y(\boldsymbol{t}^y).$$

Proof. We have $L(x,x) = 1 = U(x,x)$ by recurrence. Therefore the second and the third formula of Exercise 3.59 show that for any fixed x, the measure $\nu(y) = L(x,y)$ is stationary. Exercise 3.61 shows that $\nu(X) = \mathsf{E}_x(\boldsymbol{t}^x) = 1/\mathsf{m}(x) < \infty$ in the positive recurrent case. By Theorem 3.19, $\nu = \frac{1}{\mathsf{m}(x)} \cdot \mathsf{m}$. □

After these preparatory exercises, we define a sequence $(\boldsymbol{t}_k^x)_{k\geq 0}$ of stopping times by

$$\boldsymbol{t}_0^x = 0 \quad \text{and} \quad \boldsymbol{t}_k^x = \inf\{n > \boldsymbol{t}_{k-1}^x : Z_n = x\},$$

so that $\boldsymbol{t}_1^x = \boldsymbol{t}^x$, as defined in (1.26). Thus, $\boldsymbol{t}_k^x$ is the random instant of the k-th visit in x after starting.

The following is an immediate consequence of the strong Markov property, see Exercise 1.25. For safety's sake, we outline the proof.

3.63 Lemma. *In the recurrent case, all $\boldsymbol{t}_k^x$ are a.s. finite and are indeed stopping times. Furthermore, the random vectors with random length*

$$(\boldsymbol{t}_k^x - \boldsymbol{t}_{k-1}^x\,;\; Z_i,\; i = \boldsymbol{t}_{k-1}^x, \dots, \boldsymbol{t}_k^x - 1), \quad k \geq 1,$$

are independent and identically distributed.

Proof. We abbreviate $\boldsymbol{t}_k^x = \boldsymbol{t}_k$.

First of all, we can decide whether $[\boldsymbol{t}_k \leq n]$ by looking only at the initial trajectory $(Z_0, Z_1, \dots, Z_n)$. Indeed, we only have to check whether x occurs at least k times in $(Z_1, \dots, Z_n)$. Thus, $\boldsymbol{t}_k$ *is* a stopping time for each k.

By recurrence, $\boldsymbol{t}_1 = \boldsymbol{t}^x$ is a.s. finite. We can now proceed by induction on k. If $\boldsymbol{t}_k$ is a.s. finite, then by the strong Markov property, $(Z_{\boldsymbol{t}_k+n})_{n\geq 0}$ is a Markov chain with the same transition matrix P and starting point x. For this new chain, $\boldsymbol{t}_{k+1} - \boldsymbol{t}_k$ plays the same role as $\boldsymbol{t}^x$ for the original chain. In particular, $\boldsymbol{t}_{k+1} - \boldsymbol{t}_k$ is a.s. finite.

This proves the first statement. For the second statement, we have to show the following: for any choice of $k, l, r, s \in \mathbb{N}_0$ with $k < l$ and all points $x_1, \dots, x_{r-1}$, $y_1, \dots, y_{s-1} \in X$, the events

$$A = [\boldsymbol{t}_k - \boldsymbol{t}_{k-1} = r,\; Z_{\boldsymbol{t}_{k-1}+i} = x_i,\; i = 1, \dots, r-1]$$

and

$$B = [\boldsymbol{t}_l - \boldsymbol{t}_{l-1} = s,\; Z_{\boldsymbol{t}_{l-1}+j} = y_j,\; j = 1, \dots, s-1]$$

are independent. Now, in the time interval $[\boldsymbol{t}_k + 1, \boldsymbol{t}_l]$, the chain (Z_n) visits x precisely $l-k$ times. That is, for the Markov chain $(Z_{\boldsymbol{t}_k+n})_{n\geq 0}$, the stopping time $\boldsymbol{t}_l$ plays the same role as the stopping time $\boldsymbol{t}_{l-k}$ plays for the original chain $(Z_n)_{n\geq 0}$ starting at x.[3] In particular, $\mathsf{Pr}_x(B)$ is the same for each $l \in \mathbb{N}$. Furthermore,

[3] More precisely: in terms of Theorem 1.17, we consider $(\Omega^*, \mathcal{A}^*, \mathsf{Pr}^*) = (\Omega, \mathcal{A}, \mathsf{Pr}_x)$, the trajectory space, but $Z_n^* = Z_{\boldsymbol{t}_k+n}$. Let $\boldsymbol{t}_m^*$ be the stopping time of the m-th return to x of (Z_n^*). Then, under the mapping τ of that theorem,

$$\boldsymbol{t}_m^*(\omega) = \boldsymbol{t}_{k+m}(\tau(\omega)).$$

$A_m = A \cap [\boldsymbol{t}_{k-1} = m] \in \mathcal{A}_{m+r}$ and $Z_{m+r}(\omega) = x$ for all $\omega \in A_m$. Thus,

$$\begin{aligned}\mathsf{Pr}_x(A \cap B) &= \sum_m \mathsf{Pr}_x(B|A_m)\,\mathsf{Pr}_x(A_m)\\ &= \sum_m \mathsf{Pr}_x[\boldsymbol{t}_{l-k} - \boldsymbol{t}_{l-k-1} = s,\ Z_{\boldsymbol{t}_{l-k-1}+j=y_j},\ j = 0,\dots,s-1]\ \mathsf{Pr}_x(A_m)\\ &= \mathsf{Pr}_x(B) \sum_x \mathsf{Pr}_x(A_m) = \mathsf{Pr}_x(B)\ \mathsf{Pr}_x(A),\end{aligned}$$

as proposed. □

Proof of the ergodic theorem. We suppose that the Markov chain starts at x. We write $\boldsymbol{t}_1 = \boldsymbol{t}$ and, as above, $\boldsymbol{t}_k^x = \boldsymbol{t}_k$. Also, we let

$$S_N(f) = \sum_{n=0}^{N-1} f(Z_n).$$

Assume first that $f \geq 0$, and consider the non-negative random variables

$$Y_k = \sum_{n=\boldsymbol{t}_{k-1}}^{n=\boldsymbol{t}_k-1} f(Z_n), \quad k \geq 1.$$

By Lemma 3.63, they are independent and identically distributed. We compute, using Lemma 3.62 in the last step,

$$\begin{aligned}\mathsf{E}_x(Y_k) &= \mathsf{E}_x(Y_1)\\ &= \mathsf{E}_x\Big(\sum_{n=0}^{\boldsymbol{t}-1} f(Z_n)\Big)\\ &= \mathsf{E}_x\Big(\sum_{n=0}^{\infty} f(Z_n)\,\mathbf{1}_{[\boldsymbol{t}>n]}\Big)\\ &= \sum_{n=0}^{\infty}\sum_{y\in X} f(y)\ \mathsf{Pr}_x[Z_n = y,\ \boldsymbol{t} > n]\\ &= \sum_{y\in X} f(y)\,L(x,y) = \frac{1}{\mathsf{m}(x)}\int_X f\,d\mathsf{m},\end{aligned}$$

which is finite by assumption. The strong law of large numbers implies that

$$\frac{1}{k}\sum_{j=1}^{k} Y_j = \frac{1}{k} S_{\boldsymbol{t}_k}(f) \to \frac{1}{\mathsf{m}(x)}\int_X f\,d\mathsf{m} \quad \mathsf{Pr}_x\text{-almost surely, as } k \to \infty.$$

In the same way, setting $f \equiv 1$,

$$\frac{1}{k}\, \boldsymbol{t}_k \to \frac{1}{\mathsf{m}(x)} \quad \mathsf{Pr}_x\text{-almost surely, as } k \to \infty.$$

In particular, $\boldsymbol{t}_{k+1}/\boldsymbol{t}_k \to 1$ almost surely. Now, for $N \in \mathbb{N}$, let $\boldsymbol{k}(N)$ be the (random and a.s. defined) index such that

$$\boldsymbol{t}_{\boldsymbol{k}(N)} \le N < \boldsymbol{t}_{\boldsymbol{k}(N)+1}.$$

As $N \to \infty$, also $\boldsymbol{k}(N) \to \infty$ almost surely. Dividing all terms in this double inequality by $\boldsymbol{t}_{\boldsymbol{k}(N)}$, we find

$$1 \le \frac{N}{\boldsymbol{t}_{\boldsymbol{k}(N)}} \le \frac{\boldsymbol{t}_{\boldsymbol{k}(N)+1}}{\boldsymbol{t}_{\boldsymbol{k}(N)}}.$$

We deduce that $\boldsymbol{t}_{\boldsymbol{k}(N)}/N \to 1$ almost surely. Since $f \ge 0$, we have

$$\frac{1}{N}\, S_{\boldsymbol{t}_{\boldsymbol{k}(N)}}(f) \le \frac{1}{N}\, S_N(f) \le \frac{1}{N}\, S_{\boldsymbol{t}_{\boldsymbol{k}(N)+1}}(f).$$

Now,

$$\frac{1}{N}\, S_{\boldsymbol{t}_{\boldsymbol{k}(N)}}(f) = \underbrace{\frac{\boldsymbol{t}_{\boldsymbol{k}(N)}}{N}}_{\to 1}\, \underbrace{\frac{\boldsymbol{k}(N)}{\boldsymbol{t}_{\boldsymbol{k}(N)}}}_{\to \mathsf{m}(x)}\, \frac{1}{\boldsymbol{k}(N)} S_{\boldsymbol{t}_{\boldsymbol{k}(N)}}(f) \to \int_X f\, d\mathsf{m} \quad \underset{x}{\mathsf{Pr}}\text{-almost surely.}$$

In the same way,

$$\frac{1}{N}\, S_{\boldsymbol{t}_{\boldsymbol{k}(N)+1}}(f) \to \int_X f\, d\mathsf{m} \quad \mathsf{Pr}_x\text{-almost surely.}$$

Thus, $S_N(f)/N \to \int_X f\, d\mathsf{m}$ when $f \ge 0$.

If f is arbitrary, then we can decompose $f = f^+ - f^-$ and see that

$$\frac{1}{N}\, S_N(f) = \frac{1}{N}\, S_N(f^+) - \frac{1}{N}\, S_N(f^-) \to \int_X f^+\, d\mathsf{m} - \int_X f^-\, d\mathsf{m} = \int_X f\, d\mathsf{m}$$

Pr_x-almost surely. □

3.64 Exercise. Above, we have proved the ergodic theorem only for a deterministic starting point. Complete the proof by showing that it is valid for any initial distribution.

[Hint: replace $\boldsymbol{t}_0^x = 0$ with $\boldsymbol{s}^x$, as defined in (1.26), which is almost surely finite.] □

The ergodic theorem for Markov chains has many applications. One of them concerns the statistical estimation of the transition probabilities on the basis of observing the evolution of the chain, which is assumed to be positive recurrent. Recall the random variable $\boldsymbol{v}_n^x = \mathbf{1}_x(Z_n)$. Given the observation of the chain in a time interval $[0,\ N]$, the natural estimate of $p(x, y)$ appears to be the number of times that the chain jumps from x to y relative to the number of visits in x. That is, our estimator for $p(x, y)$ is the statistic

$$T_N = \sum_{n=0}^{N-1} \boldsymbol{v}_n^x \cdot \boldsymbol{v}_{n+1}^y \Big/ \sum_{n=0}^{N-1} \boldsymbol{v}_n^x.$$

Indeed, when $Z_0 = o$, the expected value of the denominator is $\sum_{n=0}^{N-1} p^{(n)}(o, x)$, while the expected value of the denumerator is $\sum_{n=0}^{N-1} p^{(n)}(o, x)\, p(x, y)$. (As a matter of fact, T_N is the maximum likelihood estimator of $p(x, y)$.) By the ergodic theorem, with $f = \mathbf{1}_x$,

$$\frac{1}{N} \sum_{n=0}^{N-1} \boldsymbol{v}_n^x \to \mathsf{m}(x) \quad \mathsf{Pr}_o\text{-almost surely.}$$

3.65 Exercise. Show that

$$\frac{1}{N} \sum_{n=0}^{N-1} \boldsymbol{v}_n^x \boldsymbol{v}_{n+1}^y \to \mathsf{m}(x)\, p(x, y) \quad \mathsf{Pr}_o\text{-almost surely.}$$

[Hint: show that (Z_n, Z_{n+1}) is an irreducible, positive recurrent Markov chain on the state space $\{(x, y) : p(x, y) > 0\}$. Compute its stationary distribution.] □

Combining those facts, we get that

$$T_N \to p(x, y) \quad \text{as } N \to \infty.$$

That is, the estimator is *consistent*.

G ρ-recurrence

In this short section we suppose again that (X, P) is an irreducible Markov chain. From §2.C we know that the radius of convergence $\mathsf{r} = 1/\rho(P)$ of the power series $G(x, y|z)$ does not depend on $x, y \in X$. In fact, more is true:

3.66 Lemma. *One of the following holds for* $\mathsf{r} = 1/\rho(P)$*, where* $\rho(P)$ *is the spectral radius of* (X, P)*.*

(a) $G(x, y|\mathsf{r}) = \infty$ *for all* $x, y \in X$*, or*

(b) $G(x, y|\mathsf{r}) < \infty$ *for all* $x, y \in X$.

In both cases, $F(x, y|\mathsf{r}) < \infty$ *for all* $x, y \in X$.

Proof. Let $x, y, x', y' \in X$. By irreducibility, there are $k, \ell \geq 0$ such that $p^{(k)}(x, x') > 0$ and $p^{(\ell)}(y', y) > 0$. Therefore

$$\sum_{n=k+\ell}^{\infty} p^{(n)}(x, y)\, \mathsf{r}^n \geq p^{(k)}(x, x')\, p^{(\ell)}(y', y)\, \mathsf{r}^{k+\ell} \sum_{n=0}^{\infty} p^{(n)}(x', y')\, \mathsf{r}^n.$$

Thus, if $G(x', y'|\mathsf{r}) = \infty$ then also $G(x, y|\mathsf{r}) = \infty$. Exchanging the roles of x, y and x', y', we also get the converse implication.

Finally, if m is such that $p^{(m)}(y, x) > 0$, then we have in the same way as above that for each $z \in (0, \mathsf{r})$

$$G(y, y|z) \geq p^{(m)}(y, x)\, z^m\, G(x, y|z) = p^{(m)}(y, x)\, z^m\, F(x, y|z)\, G(y, y|z)$$

by Theorem 1.38 (b). Letting $z \to \mathsf{r}$ from below, we find that

$$F(x, y|\mathsf{r}) \leq 1 \big/ \big(p^{(m)}(y, x)\, \mathsf{r}^m\big). \qquad \square$$

3.67 Definition. In case (a) of Lemma 3.66, the Markov chain is called *ρ-recurrent,* in case (b) it is called *ρ-transient.*

This definition and various results are due to Vere-Jones [49], see also Seneta [Se]. This is a formal analogue of usual recurrence, where one studies $G(x, y|1)$ instead of $G(x, y|\mathsf{r})$. While in the previous (1996) Italian version of this chapter, I wrote "there is no analogous probabilistic interpretation of ρ-recurrence", I learnt in the meantime that there is indeed an interpretation in terms of branching Markov chains. This will be explained in Chapter 5. Very often, one finds "r-recurrent" in the place of "ρ-recurrent", where $\rho = 1/\mathsf{r}$. There are good reasons for either terminology.

3.68 Remarks. (a) The Markov chain is ρ-recurrent if and only if $U(x, x|\mathsf{r}) = 1$ for some ($\Longleftrightarrow$ all) $x \in X$. The chain is ρ-transient if and only if $U(x, x|\mathsf{r}) < 1$ for some ($\Longleftrightarrow$ all) $x \in X$. (Recall that the radius of convergence of $U(x, x|\cdot)$ is $\geq \mathsf{r}$, and compare with Proposition 2.28.)

(b) If the Markov chain is recurrent in the usual sense, $G(x, x|1) = \infty$, then $\rho(P) = \mathsf{r} = 1$ and the chain is ρ-recurrent.

(c) If the Markov chain is transient in the usual sense, $G(x, x|1) < \infty$, then each of the following cases can occur. (We shall see Example 5.24 in Chapter 5, Section A.)

- $\mathsf{r} = 1$, and the chain is ρ-transient,

- $\mathsf{r} > 1$, and the chain is ρ-recurrent,
- $\mathsf{r} > 1$, and the chain is ρ-transient.

(d) If $\mathsf{r} > 1$ (i.e., $\rho(P) < 1$) then in any case the Markov chain is transient in the usual sense.

In analogy with Definition 3.8, ρ-recurrence is subdivided in two cases.

3.69 Definition. In the ρ-recurrent case, the Markov chain is called

ρ-positive-recurrent, if $U'(x,x|\mathsf{r}-) = \sum_{n=1}^{\infty} n\, \mathsf{r}^{n-1} u^{(n)}(x,x) < \infty$ for some ($\Longleftrightarrow$ every) $x \in X$, and

ρ-null-recurrent, if $U'(x,x|\mathsf{r}-) = \infty$ for some ($\Longleftrightarrow$ every) $x \in X$.

3.70 Exercise. Prove that in the (irreducible) ρ-recurrent case it is indeed true that when $U'(x,x|\mathsf{r}-) < \infty$ for some x then this holds for all $x \in X$. □

3.71 Exercise. Fix $x \in X$ and let $\mathsf{s} = \mathsf{s}(x,x)$ be the radius of convergence of $U(x,x|z)$. Show that if $U(x,x|\mathsf{s}-) > 1$ then the Markov chain is ρ-positive recurrent. Deduce that when (X,P) is not ρ-positive-recurrent then $\mathsf{s}(x,x) = \mathsf{r}$ for all $x \in X$. □

3.72 Theorem. (a) *If (X,P) is an irreducible, ρ-positive-recurrent Markov chain then for $x, y \in X$*

$$\lim_{n\to\infty} \mathsf{r}^{nd+\ell} p^{(nd+\ell)}(x,y) = d\,\frac{F(x,y|\mathsf{r}-)}{\mathsf{r}\,U'(y,y|\mathsf{r}-)},$$

where d is the period and $\ell \in \{0,\dots,d-1\}$ is such that $x \xrightarrow{kd+\ell} y$ for some $k \geq 0$.

(b) *If (X,P) is ρ-null-recurrent or ρ-transient then for $x, y \in X$*

$$\lim_{n\to\infty} \mathsf{r}^n p^{(n)}(x,y) = 0.$$

Proof. We first consider the case $x = y$. In the ρ-recurrent case we construct the following auxiliary Markov chain on $\mathbb{N}$ with transition matrix $\widetilde{P}$ given by

$$\tilde{p}(1,n) = u^{(nd)}(y,y)\,\mathsf{r}^{nd}, \quad \tilde{p}(n+1,n) = 1,$$

while $p(m,n) = 0$ in all other cases. See Figure 8.

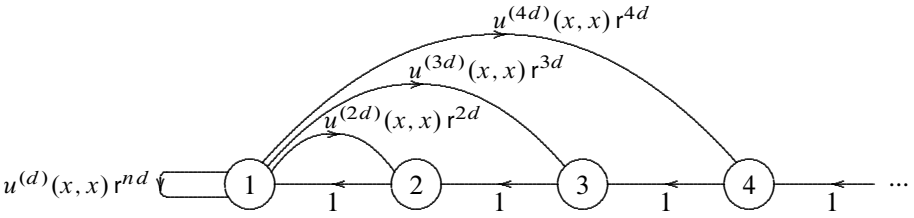

Figure 8

The transition matrix $\widetilde{P}$ is stochastic as (X, P) is ρ-recurrent; see Remark 3.68 (a). The construction is such that the first return probabilities of this chain to the state 1 are $\tilde{u}^{(n)}(1,1) = u^{(nd)}(y,y)\,\mathsf{r}^{nd}$. Applying (1.39) both to (X, P) and to $(\mathbb{N}, \widetilde{P})$, we see that $p^{(nd)}(y,y)\,\mathsf{r}^{nd} = \tilde{p}^{(n)}(1,1)$. Thus, $(\mathbb{N}, \widetilde{P})$ is aperiodic and recurrent, and Theorem 3.48 implies that

$$\lim_{n\to\infty} p^{(nd)}(y,y)\,\mathsf{r}^{nd} = \lim_{n\to\infty} \tilde{p}^{(n)}(1,1) = \frac{1}{\sum_{k=1}^{\infty} k\,\tilde{u}^{(k)}(1,1)} = \frac{d}{\mathsf{r}\,U'(y,y|\mathsf{r}-)}.$$

This also applies to the ρ-null-recurrent case, where the limit is 0.

In the ρ-transient case, it is clear that $p^{(nd)}(y,y)\,\mathsf{r}^{nd} \to 0$.

Now suppose that $x \neq y$. We know from Lemma 3.66 that $F(x,y|\mathsf{r}) < \infty$. Thus, the proof can be completed in the same way as in Theorem 3.54. □

Chapter 4
Reversible Markov chains

A The network model

4.1 Definition. An irreducible Markov chain (X, P) is called *reversible* if there is a positive measure m on X such that

$$\mathsf{m}(x)\, p(x,y) = \mathsf{m}(y)\, p(y,x) \quad \text{for all } x, y \in X.$$

We then call m a *reversing measure* for P.

This symmetry condition allows the development of a rich theory which comprises many important classes of examples and models, such as simple random walk on graphs, nearest neighbour random walks on trees, and symmetric random walks on groups. Reversible Markov chains are well documented in the literature. We refer first of all to the beautiful little book of DOYLE and SNELL [D-S], which lead to a breakthrough of the popularity of random walks. Further valid sources are, among others SALOFF-COSTE [SC], several parts of my monograph [W2], and in particular the (ever forthcoming) perfect book of LYONS with PERES [L-P]. Here we shall only touch a small part of the vast interesting material, and encourage the reader to consult those books.

If (X, P) is reversible, then we call $a(x,y) = \mathsf{m}(x)p(x,y) = a(y,x)$ the *conductance* between x and y, and $\mathsf{m}(x)$ the *total conductance* at x.

Conversely, we can also start with a symmetric function $a\colon X \times X \to [0, \infty)$ such that $0 < \mathsf{m}(x) = \sum_y a(x,y) < \infty$ for every $x \in X$. Then $p(x,y) = a(x,y)/\mathsf{m}(x)$ defines a reversible Markov chain (random walk).

Reversibility implies that X is the union of essential classes that do not communicate among each other. Therefore, it is no restriction that we shall always assume irreducibility of (X, P).

4.2 Lemma. (1) *If (X, P) is reversible then $\mathsf{m}(\cdot)$ is an invariant measure for P with total mass*

$$\mathsf{m}(X) = \sum_{x,y \in X} a(x,y).$$

(2) *In particular, (X, P) is positive recurrent if and only if $\mathsf{m}(X) < \infty$, and in this case,*

$$\mathsf{E}_x(t^x) = \mathsf{m}(X)/\mathsf{m}(x).$$

(3) *Furthermore, also P^n is reversible with respect to m.*

Proof. We have

$$\sum_x \mathsf{m}(x)\, p(x,y) = \sum_x \mathsf{m}(y)\, p(y,x) = \mathsf{m}(y).$$

The statement about positive recurrence follows from Theorem 3.19, since the stationary probability measure is $\frac{1}{\mathsf{m}(X)}\mathsf{m}(\cdot)$ when $\mathsf{m}(X) < \infty$, while otherwise m is an invariant measure with infinite total mass.

Finally, reversibility of P is equivalent with symmetry of the matrix DPD^{-1}, where D is the diagonal matrix over X with diagonal entries $\sqrt{\mathsf{m}(x)}$, $x \in X$. Taking the n-th power, we see that also DP^nD^{-1} is symmetric. □

4.3 Example. Let $\Gamma = (X, E)$ be a *symmetric* (or non-oriented) graph with $V(\Gamma) = X$ and non-empty, symmetric edge set $E = E(\Gamma)$, that is, we have $[x,y] \in E \iff [y,x] \in E$. (Attention: here we distinguish between the two oriented edges $[x,y]$ and $[y,x]$, when $x \neq y$. In classical graph theory, such a pair of edges is usually considered and drawn as one non-oriented edge.)

We assume that Γ is *locally finite*, that is,

$$\deg(x) = |\{y : [x,y] \in E\}| < \infty \quad \text{for all } x \in X,$$

and connected. *Simple random walk* (*SRW*) on Γ is the Markov chain with state space X and transition probabilities

$$p(x,y) = \begin{cases} 1/\deg(x), & \text{if } [x,y] \in E, \\ 0, & \text{otherwise.} \end{cases}$$

Connectedness is equivalent with irreducibility of SRW, and the random walk is reversible with respect to $\mathsf{m}(x) = \deg(x)$. Thus, $a(x,y) = 1$ if $[x,y] \in E$, and $a(x,y) = 0$, otherwise. The resulting matrix $A = \big(a(x,y)\big)_{x,y\in X}$ is called the *adjacency matrix* of the graph.

In particular, SRW on the graph Γ is positive recurrent if and only if Γ is finite, and in this case,

$$\mathsf{E}_x(\boldsymbol{t}^x) = |E|/\deg(x).$$

(Attention: $|E|$ counts all oriented edges. If instead we count undirected edges, then $|E|$ has to be replaced with $2\times$ the number of edges with distinct endpoints plus $1\times$ the number of loops.)

For a general, irreducible Markov chain (X, P) which is reversible with respect to the measure m, we also consider the associated graph $\Gamma(P)$ according to Definition 1.6. By reversibility, it is again non-oriented (its edge set is symmetric). The graph is not necessarily locally finite.

The period of P is $d = 1$ or $d = 2$. The latter holds if and only if the graph $\Gamma(P)$ is *bipartite:* the vertex set has a partition $X = X_1 \cup X_2$ such that each edge

has one endpoint in X_1 and the other in X_2. Equivalently, every closed path in $\Gamma(P)$ has an even number of edges.

For an edge $e = [x, y] \in E = E(P)$, we denote $\check{e} = [y, x]$, and write $e^- = x$ and $e^+ = y$ for its initial and terminal vertex, respectively. Thus, $e \in E$ if and only if $a(x, y) > 0$. We call the number $r(e) = 1/a(x, y) = r(\check{e})$ the *resistance* of e, or rather, the resistance of the non-oriented edge that is given by the pair of oriented edges e and $\check{e}$.

The triple $\mathcal{N} = (X, E, r)$ is called a *network*, where we imagine each edge e as a wire with resistance $r(e)$, and several wires are linked at each node (vertex). Equivalently, we may think of a system of tubes e with cross-section 1 and length $r(e)$, connected at the vertices. If we start with X, E and r then the requirements are that (X, E) is a countable, connected, symmetric graph, and that

$$0 < \mathsf{m}(x) = \sum_{e \in E : e^- = x} 1/r(e) < \infty \quad \text{for each } x \in X,$$

and then $p(x, y) = 1\big/\big(\mathsf{m}(x) r([x, y])\big)$ whenever $r([x, y]) > 0$, as above.

The electric network interpretation leads to nice explanations of various results, see in particular [D-S] and [L-P]. We will come back to part of it at a later stage.

It will be convenient to introduce a potential theoretic setup, and to involve some basic functional analysis, as follows. The (real) Hilbert space $\ell^2(X, \mathsf{m})$ consists of all functions $f : X \to \mathbb{R}$ with $\|f\|^2 = (f, f) < \infty$, where the inner product of two such functions f_1, f_2 is

$$(f_1, f_2) = \sum_{x \in X} f_1(x) f_2(x) \mathsf{m}(x). \tag{4.4}$$

Reversibility is the same as saying that the transition matrix P acts on $\ell^2(X, \mathsf{m})$ as a self-adjoint operator, that is, $(Pf_1, f_2) = (f_1, Pf_2)$ for all $f_1, f_2 \in \ell^2(X, \mathsf{m})$. The action of P is of course given by (3.16), $Pf(x) = \sum_y p(x, y) f(y)$.

The Hilbert space $\ell^2_\sharp(E, r)$ consists of all functions $\phi : E \to \mathbb{R}$ which are *anti-symmetric*: $\phi(\check{e}) = -\phi(e)$ for each $e \in E$, and such that $\langle \phi, \phi \rangle < \infty$, where the inner product of two such functions ϕ_1, ϕ_2 is

$$\langle \phi_1, \phi_2 \rangle = \frac{1}{2} \sum_{e \in E} \phi_1(e) \phi_2(e)\, r(e). \tag{4.5}$$

We imagine that such a function ϕ represents a "flow", and if $\phi(e) \geq 0$ then this is the amount per time unit that flows from e^- to e^+, while if $\phi(e) < 0$ then $-\phi(e) = \phi(\check{e})$ flows from e^+ to e^-. Note that $\phi(e) = 0$ when e is a loop.

We introduce the difference operator

$$\nabla : \ell^2(X, \mathsf{m}) \to \ell^2_\sharp(E, r), \quad \nabla f(e) = \frac{f(e^+) - f(e^-)}{r(e)}. \tag{4.6}$$

If we interpret f as a potential (voltage) on the set of nodes (vertices) X, then $\nabla\phi(e)$ represents the electric current along the edge e, and the defining equation for ∇ is just *Ohm's law.*

Recall that the adjoint operator $\nabla^*\colon \ell^2_\sharp(E,r) \to \ell^2(X,\mathsf{m})$ is defined by the equation

$$(f, \nabla^*\phi) = \langle \nabla f, \phi\rangle \quad \text{for all } f \in \ell^2(X,\mathsf{m}),\ \phi \in \ell^2_\sharp(E,r).$$

4.7 Exercise. Prove that the operator ∇ has norm $\|\nabla\| \le \sqrt{2}$, that is, $\langle \nabla f, \nabla f\rangle \le 2\,(f, f)$.

Show that ∇^* is given by

$$\nabla^*\phi(x) = \frac{1}{\mathsf{m}(x)} \sum_{e\in E\,:\,e^+=x} \phi(e). \tag{4.8}$$

[Hint: it is sufficient to check the defining equation for the adjoint operator only for finitely supported functions f. Use anti-symmetry of ϕ.] □

In our interpretation in terms of flows,

$$\sum_{e^+=x,\,\phi(e)>0} \phi(e) \quad \text{and} \quad -\sum_{e^+=x,\,\phi(e)<0} \phi(e)$$

are the amounts flowing into node x resp. out of node x, and $\mathsf{m}(x)\nabla^*\phi(x)$ is the difference of those two quantities. Thus, if $\nabla^*\phi(x) = 0$ then this means that the flow has no source or sink at x. This is known as *Kirchhoff's node law.* Later, we shall give a more precise definition of flows. The *Laplacian* is the operator

$$\mathfrak{L} = -\nabla^*\nabla = P - I, \tag{4.9}$$

where I is the identity matrix over X and P is the transition matrix of our random walk, both viewed as operators on functions $X \to \mathbb{R}$.

4.10 Exercise. Verify the equation $\nabla^*\nabla\, f = (I - P)\, f$ for $f \in \ell^2(X,\mathsf{m})$. □

For reversible Markov chains, the name "spectral radius" for the number $\rho(P)$ is justified in the operator theoretic sense by the following.

4.11 Proposition. *If (X, P) is reversible then*

$$\rho(P) = \|P\|,$$

the norm of P as an operator on $\ell^2(X,\mathsf{m})$.

Proof. First of all,

$$p^{(n)}(x,x)\,\mathsf{m}(x) = (P^n\mathbf{1}_x, \mathbf{1}_x) \le \|P\|^n(\mathbf{1}_x,\mathbf{1}_x) = \|P\|^n\,\mathsf{m}(x).$$

Taking n-th roots and letting $n\to\infty$, we see that $\rho(P)\le\|P\|$.

For showing the (more interesting) reversed inequality, we use the fact that the linear space $\ell_0(X)$ of finitely supported real functions on X is dense in $\ell^2(X,\mathsf{m})$. Thus, it is sufficient to show that $(Pf,Pf)\le\rho(P)^2\,(f,f)$ for every non-zero finitely supported function f on X. We first assume that f is non-negative. By self-adjointness of P and a standard use of the Cauchy–Schwarz inequality,

$$(P^{n+1}f, P^{n+1}f)^2 = (P^nf, P^{n+2}f)^2 \le (P^nf, P^nf)(P^{n+2}f, P^{n+2}f).$$

We see that the sequence $\big((P^{n+1}f,P^{n+1}f)/(P^nf,P^nf)\big)_{n\ge0}$ is increasing. A basic lemma of elementary calculus says that when a sequence (a_n) of positive numbers is such that a_{n+1}/a_n converges, then also $a_n^{1/n}$ converges, and the two limits coincide. We claim that

$$\lim_{n\to\infty}(P^nf,P^nf)^{1/n} = \rho(P)^2.$$

Choose $x_0\in\operatorname{supp}(f)$. Then, since $f\ge0$,

$$\begin{aligned}(P^nf,P^nf) &= \sum_{x\in\operatorname{supp}(f)}\mathsf{m}(x)\Big(\sum_{y\in\operatorname{supp}(f)}p^{(n)}(x,y)f(y)\Big)^2\\ &\ge \mathsf{m}(x_0)\,p^{(n)}(x_0,x_0)^2f(x_0)^2.\end{aligned}$$

Therefore the above limit is $\ge\rho(P)^2$. Conversely, given $\varepsilon>0$, there is n_ε such that

$$p^{(n)}(x,y)\le\big(\rho(P)+\varepsilon\big)^n\quad\text{for all } n\ge n_\varepsilon,\ x,y\in\operatorname{supp}(f).$$

We infer that

$$(P^nf,P^nf)\le C\,\big(\rho(P)+\varepsilon\big)^{2n},\quad\text{where } C=\sum_{x\in\operatorname{supp}(f)}\mathsf{m}(x)\Big(\sum_{y\in\operatorname{supp}(f)}f(y)\Big)^2.$$

Taking n-th roots and letting $n\to\infty$, we see that the claim is true. Consequently

$$\frac{(Pf,Pf)}{(f,f)}\le\frac{(P^{n+1}f,P^{n+1}f)}{(P^nf,P^nf)}\le\rho(P)^2$$

for every non-negative function in $\ell_0(X)$. If $f\in\ell_0(X)$ is arbitrary, then

$$(Pf,Pf)\le(P|f|,P|f|)\le\rho(P)^2\,(|f|,|f|)=\rho(P)^2\,(f,f),$$

whence $\|P\|\le\rho(P)$. □

B Speed of convergence of finite reversible Markov chains

In this section, we always assume that X is finite and that (X, P) is irreducible and reversible with respect to the measure m. Since $\mathsf{m}(X) < \infty$, we may assume without loss of generality that m is a probability measure:

$$\sum_{x \in X} \mathsf{m}(x) = 1.$$

If the period of P is $d = 1$, then we know from Theorem 3.28 that the difference $\| p^{(n)}(x, \cdot) - \mathsf{m} \|_1$ tends to 0 exponentially fast.

This fact has an algorithmic use: if X is a very large finite set, and the values $\mathsf{m}(x)$ are very small, then we cannot use a random number generator to simulate the probability measure m. Indeed, such a generator simulates the continuous equidistribution on the interval $[0, 1]$. For simulating m, we should partition $[0, 1]$ into $|X|$ intervals I_x of length $\mathsf{m}(x)$, $x \in X$. If our generator provides a number ξ, then the output of our simulation of m should be the point x, when $\xi \in I_x$. However, if the numbers $\mathsf{m}(x)$ are below the machine's precision, then they will all be rounded to 0, and the simulation cannot work. An alternative is to run a Markov chain whose stationary probability distribution is m. It is chosen such that at each step, there are only relatively few possible transitions which all have relatively large probabilities, so that they can be efficiently simulated by the above method. If we start at x and make n steps then the distribution $p^{(n)}(x, \cdot)$ of Z_n is very close to m. Thus, the "random" element of X that we find after the simulation of n successive steps of the Markov chain is "almost" distributed as m. ("Random" is in quotation marks because the random number generator is based on a clever, but deterministic algorithm whose output is "almost" equidistributed on $[0, 1]$.) The basic question is now: how many steps of the Markov chain should we perform so that the distribution $p^{(n)}(x, \cdot)$ is sufficiently close (for our purposes) to m? That is, given a small $\varepsilon > 0$, we want to know how large we have to choose n such that

$$\| p^{(n)}(x, \cdot) - \mathsf{m} \|_1 < \varepsilon.$$

This is a mathematical analysis that we should best perform before starting our algorithm. The estimate of the speed of convergence, i.e., the parameter $\bar{\tau}$ found in Theorem 3.28, is in general rather crude, and we need better methods of estimation.

Here we shall present only a small glimpse of some basic methods of this type. There is a vast literature, mostly of the last 20–25 years. Good parts of it are documented in the book of Diaconis [Di] and the long and detailed exposition of Saloff-Coste [SC].

The following lemma does not require finiteness of X, but we need positive recurrence, that is, we need that m is a probability measure.

4.12 Lemma. $\|p^{(n)}(x,\cdot) - \mathsf{m}\|_1^2 \le \dfrac{p^{(2n)}(x,x)}{\mathsf{m}(x)} - 1.$

Proof. We use the Cauchy–Schwarz inequality.

$$\begin{aligned}
\|p^{(n)}(x,\cdot) - \mathsf{m}\|_1^2 &= \Big(\sum_{y\in X} \frac{|p^{(n)}(x,y) - \mathsf{m}(y)|}{\sqrt{\mathsf{m}(y)}} \cdot \sqrt{\mathsf{m}(y)} \Big)^2 \\
&\le \sum_{y\in X} \frac{\big(p^{(n)}(x,y) - \mathsf{m}(y)\big)^2}{\mathsf{m}(y)} \cdot \underbrace{\sum_{y\in X} \mathsf{m}(y)}_{=1} \\
&= \sum_{y\in X} \frac{p^{(n)}(x,y)^2}{\mathsf{m}(y)} - 2\sum_{y\in X} p^{(n)}(x,y) + \sum_{y\in X} \mathsf{m}(y) \\
&= \sum_{y\in X} \frac{p^{(n)}(x,y)\, p^{(n)}(y,x)}{\mathsf{m}(x)} - 1 = \frac{1}{\mathsf{m}(x)} p^{(2n)}(x,x) - 1,
\end{aligned}$$

as proposed. □

Since X is finite and P is self-adjoint on $\ell^2(X,\mathsf{m}) \equiv \mathbb{R}^X$, the spectrum of P is real. If (X,P), besides being irreducible, is also aperiodic, then $-1 < \lambda < 1$ for all eigenvalues of P with the exception of $\lambda = 1$; see Theorem 3.29. We define

$$\lambda_* = \lambda_*(P) = \max\{|\lambda| : \lambda \in \operatorname{spec}(P),\ \lambda \ne 1\}. \tag{4.13}$$

4.14 Lemma. $|p^{(n)}(x,x) - \mathsf{m}(x)| \le \big(1 - \mathsf{m}(x)\big)\, \lambda_*^n.$

Proof. This is deduced by diagonalizing the self-adjoint operator P on $\ell^2(X,\mathsf{m})$. We give the details, using only elementary facts from basic linear algebra.

For our notation, it will be convenient to index the eigenvalues with the elements of X, as λ_x, $x \in X$, including possible multiplicities. We also choose a "root" $o \in X$ and let the maximal eigenvalue correspond to o, that is, $\lambda_o = 1$ and $\lambda_x < 1$ for all $x \ne o$. Recall that DPD^{-1} is symmetric, where $D = \operatorname{diag}\big(\sqrt{\mathsf{m}(x)}\big)_{x\in X}$. There is a real matrix $V = \big(v(x,y)\big)_{x,y\in X}$ such that

$$DPD^{-1} = V^{-1}\Lambda V, \quad \text{where } \Lambda = \operatorname{diag}(\lambda_x)_{x\in X},$$

and V is orthogonal, $V^{-1} = V^t$. The row vectors of the matrix VD are left eigenvectors of P. In particular, recall that $\lambda_o = 1$ is a simple eigenvalue, so that each associated left eigenvector is a multiple of the unique stationary probability measure m. Therefore there is $C \ne 0$ such that $v(o,x)\sqrt{\mathsf{m}(x)} = C \cdot \mathsf{m}(x)$ for each x. Since $\sum_x v(o,x)^2 = 1$ by orthogonality of V, we find $C^2 = 1$. Therefore

$$v(o,x)^2 = \mathsf{m}(x).$$

We have $P^n = (D^{-1}V^t)\Lambda^n(VD)$, that is

$$p^{(n)}(x,x') = \sum_{y\in X} \mathsf{m}(x)^{-1/2}\, v(y,x)\ \lambda_y^n\ v(y,x')\, \mathsf{m}(x')^{1/2}.$$

Therefore

$$\begin{aligned}\big|p^{(n)}(x,x) - \mathsf{m}(x)\big| &= \Big|\sum_{y\in X} v(y,x)^2\,\lambda_y^n - v(o,x)^2\Big| = \Big|\sum_{y\neq o} v(y,x)^2\,\lambda_x^n\Big| \\ &\le \Big(\sum_{y\in X} v(y,x)^2 - v(o,x)^2\Big)\lambda_*^n = \big(1-\mathsf{m}(x)\big)\,\lambda_*^n. \qquad \square\end{aligned}$$

As a corollary, we obtain the following important estimate.

4.15 Theorem. *If X is finite, P irreducible and aperiodic, and reversible with respect to the probability measure m on X, then*

$$\|p^{(n)}(x,\cdot) - \mathsf{m}\|_1 \le \sqrt{\big(1-\mathsf{m}(x)\big)/\mathsf{m}(x)}\,\lambda_*^n.$$

We remark that aperiodicity is not needed for the proof of the last theorem, but without it, the result is not useful. The proof of Lemma 4.14 also leads to the better estimate

$$\|p^{(n)}(x,\cdot) - \mathsf{m}\|_1^2 \le \frac{1}{\mathsf{m}(x)} \sum_{y\neq o} v(y,x)^2\,\lambda_x^{2n} \tag{4.16}$$

in the notation of that proof.

If we want to use Theorem 4.15 for bounding the speed of convergence to the stationary distribution, then we need an upper bound for the second largest eigenvalue $\lambda_1 = \max\big(\mathrm{spec}(P)\setminus\{1\}\big)$, and a lower bound for $\lambda_{\min} = \min\mathrm{spec}(P)$ [since $\lambda_* = \max\{\lambda_1, -\lambda_{\min}\}$], unless we can compute these numbers explicitly. In many cases, reasonable lower bounds on $\lambda_{\min}$ are easy to obtain.

4.17 Exercise. Let (X,P) be irreducible, with finite state space X. Suppose that one can write $P = a\cdot I + (1-a)\cdot Q$, where I is the identity matrix, Q is another stochastic matrix, and $0 < a < 1$. Show that $\lambda_{\min}(P) \ge -1 + 2a$, with equality when $\lambda_{\min}(Q) = -1$.

More generally, suppose that there is an *odd* $k \ge 1$ such that $P^k \ge a\cdot I$ elementwise, where $a > 0$. Deduce that $\lambda_{\min}(P) \ge (-1+2a)^{1/k}$. $\square$

Random walks on groups

A simplification arises in the case of random walks on groups. Let $\mathfrak{G}$ be a finite or countable group, in general written multiplicatively (unless the group is Abelian, in which case often "+" is preferred for the group operation). By slightly unusual

notation, we write o for its unit element. (The symbol e is used for edges of graphs here.) Also, let μ be a probability measure on $\mathfrak{G}$. The (*right*) *random walk* on $\mathfrak{G}$ with *law* μ is the Markov chain with state space $\mathfrak{G}$ and transition probabilities

$$p(x, y) = \mu(x^{-1}y), \quad x, y \in \mathfrak{G}. \tag{4.18}$$

The random walk is called *symmetric,* if $p(x, y) = p(y, x)$ for all x, y, or equivalently, $\mu(x^{-1}) = \mu(x)$ for all $x \in \mathfrak{G}$. Then P is reversible with respect to the counting measure (or any multiple thereof).

Instead of the trajectory space, another natural probability space can be used to model the random walk with law μ on $\mathfrak{G}$. We can equip $\Omega^* = \mathfrak{G}^{\mathbb{N}}$ with the product σ-algebra $\mathcal{A}^*$ of the discrete one on $\mathfrak{G}$ (the family of all subsets of $\mathfrak{G}$). As in the case of the trajectory space, it is generated by the family of all "cylinder" sets, which are of the form $\prod_n A_n$, where $A_n \subset \mathfrak{G}$ and $A_n \neq \mathfrak{G}$ for only finitely many n. Then $\mathfrak{G}^{\mathbb{N}}$ is equipped with the product measure $\mathsf{Pr}^* = \mu^{\mathbb{N}}$, which is the unique measure on $\mathcal{A}^*$ that satisfies $\mathsf{Pr}^*\left(\prod_n A_n\right) = \prod_n \mu(A_n)$ for every cylinder set as above. Now let $Y_n \colon \mathfrak{G}^{\mathbb{N}} \to \mathfrak{G}$ be the n-th projection. Then the Y_n are i.i.d. $\mathfrak{G}$-valued random variables with common distribution μ. The random walk (4.18) starting at $x_0 \in \mathfrak{G}$ is then modeled as

$$Z_0^* = x_0, \quad Z_n^* = x_0 Y_1 \cdots Y_n, \quad n \geq 1.$$

Indeed, Y_{n+1} is independent of $Z_0, \ldots, Z_n$, whence

$$\begin{aligned} \mathsf{Pr}^*[Z_{n+1}^* = y \mid Z_n^* = x,\ Z_k^* = x_k\ (k < n)] \\ = \mathsf{Pr}^*[Y_{n+1}^* = x^{-1}y \mid Z_n^* = x,\ Z_k^* = x_k\ (k < n)] \\ = \mathsf{Pr}^*[Y_{n+1}^* = x^{-1}y] = \mu(x^{-1}y) \end{aligned}$$

(as long as the conditioning event has non-zero probability). According to Theorem 1.17, the natural measure preserving mapping τ from the probability space $(\Omega^*, \mathcal{A}^*, \mathsf{Pr}^*)$ to the trajectory space equipped with the measure Pr_{x_0} is given by

$$(y_n)_{n\geq 1} \mapsto (z_n)_{n\geq 0}, \quad \text{where } z_n = x_0 y_1 \cdots y_n.$$

In order to describe the transition probabilities in n steps, we need the definition of the *convolution* $\mu_1 * \mu_2$ of two measures μ_1, μ_2 on the group $\mathfrak{G}$:

$$\mu_1 * \mu_2(x) = \sum_{y\in\mathfrak{G}} \mu_1(y)\mu_2(y^{-1}x). \tag{4.19}$$

4.20 Exercise. Suppose that Y_1 and Y_2 are two independent, $\mathfrak{G}$-valued random variables with respective distributions μ_1 and μ_2. Show that the product $Y_1 Y_2$ has distribution $\mu_1 * \mu_2$. □

We write $\mu^{(n)} = \mu * \cdots * \mu$ (n times) for the n-th convolution power of μ, with $\mu^{(0)} = \delta_o$, the point mass at the group identity. We observe that

$$\operatorname{supp}(\mu^{(n)}) = \{x_1 \cdots x_n \mid x_i \in \operatorname{supp}(\mu)\}.$$

4.21 Lemma. *For the random walk on $\mathfrak{G}$ with law μ, the transition probabilities in n steps are*

$$p^{(n)}(x, y) = \mu^{(n)}(x^{-1}y).$$

The random walk is irreducible if and only if

$$\bigcup_{n=1}^{\infty} \operatorname{supp}(\mu^{(n)}) = \mathfrak{G}.$$

Proof. For $n = 1$, the first assertion coincides with the definition of P. If it is true for n, then

$$\begin{aligned} p^{(n+1)}(x, y) &= \sum_{w \in X} p(x, w)\, p^{(n)}(w, y) \\ &= \sum_{w \in X} \mu(x^{-1}w)\mu^{(n)}(w^{-1}y) \qquad [\text{setting } v = x^{-1}w] \\ &= \sum_{v \in X} \mu(v)\mu^{(n)}(v^{-1}x^{-1}y) \qquad [\text{since } w^{-1} = v^{-1}x^{-1}] \\ &= \mu * \mu^{(n)}(x^{-1}y). \end{aligned}$$

To verify the second assertion, it is sufficient to observe that $o \xrightarrow{n} x$ if and only if $x \in \operatorname{supp}(\mu^{(n)})$, and that $x \xrightarrow{n} y$ if and only if $o \xrightarrow{n} x^{-1}y$. □

Let us now suppose that our group $\mathfrak{G}$ is *finite* and that the random walk on $\mathfrak{G}$ is irreducible, and therefore recurrent. The stationary probability measure is uniform distribution on $\mathfrak{G}$, that is, $\mathfrak{m}(x) = 1/|\mathfrak{G}|$ for every $x \in \mathfrak{G}$. Indeed,

$$\sum_{x \in X} \frac{1}{|\mathfrak{G}|}\, p(x, y) = \frac{1}{|\mathfrak{G}|} \sum_{x \in X} \mu(x^{-1}y) = \frac{1}{|\mathfrak{G}|},$$

the transition matrix is *doubly stochastic* (both row and column sums are equal to 1). If in addition the random walk is also symmetric, then the distinct eigenvalues of P

$$1 = \lambda_0 > \lambda_1 > \cdots > \lambda_q \geq -1$$

are all real, with multiplicities $\operatorname{mult}(\lambda_i)$ and $\sum_{i=0}^{q} \operatorname{mult}(\lambda_i) = |\mathfrak{G}|$. By Theorem 3.29, we have $\operatorname{mult}(\lambda_0) = 1$, while $\lambda_q = -1$ if and only if the random walk has period 2.

4.22 Lemma. *For a symmetric, irreducible random walk on the finite group $\mathfrak{G}$, one has*

$$p^{(n)}(x,x) = \frac{1}{|\mathfrak{G}|} \sum_{i=0}^{q} \operatorname{mult}(\lambda_i)\, \lambda_i^n .$$

Proof. We have $p^{(n)}(x,x) = \mu^{(n)}(o) = p^{(n)}(y,y)$ for all $x, y \in \mathfrak{G}$. Therefore

$$p^{(n)}(x,x) = \frac{1}{|\mathfrak{G}|} \sum_{y \in X} p^{(n)}(y,y) = \frac{1}{|\mathfrak{G}|} \operatorname{tr}(P^n),$$

where $\operatorname{tr}(P^n)$ is the trace (sum of the diagonal elements) of the matrix P^n, which coincides with the sum of the eigenvalues (taking their multiplicities into account). □

In this specific case, the inequalities of Lemmas 4.12 and 4.14 become

$$\begin{aligned} \|p^{(n)}(x,\cdot) - \mathsf{m}\|_1 &\le \sqrt{|\mathfrak{G}|\, p^{(2n)}(x,x) - 1} = \sqrt{\textstyle\sum_{i=1}^{q} \operatorname{mult}(\lambda_i)\, \lambda_i^{2n}} \\ &\le \sqrt{|\mathfrak{G}| - 1}\, \lambda_*^n, \end{aligned} \tag{4.23}$$

where the measure m is equidistribution on $\mathfrak{G}$, that is, $\mathsf{m}(x) = 1/|\mathfrak{G}|$, and $\lambda_* = \max\{\lambda_1, -\lambda_q\}$.

4.24 Example (Random walk on the hypercube). The hypercube is the (additively written) Abelian group $\mathfrak{G} = \mathbb{Z}_2^d$, where $\mathbb{Z}_2 = \{0,1\}$ is the group with two elements and addition modulo 2. We can view it as a (non-oriented) graph with vertex set $\mathbb{Z}_2^d$, and with edges between every pair of points which differ in exactly one component. This graph has the form of a hypercube in d dimensions. Every point has d neighbours. According to Example 4.3, simple random walk is the Markov chain which moves from a point to any of its neighbours with equal probability $1/d$. This is the symmetric random walk on the group $\mathbb{Z}_2^d$ whose law μ is the equidistribution on the points (vectors) $\boldsymbol{e}_i = \big(\delta_i(j)\big)_{j=1,\dots,d}$. The associated transition matrix is irreducible, but its period is 2.

In order to compute the eigenvalues of P, we introduce the set $\mathcal{E}_d = \{-1,1\}^d \subset \mathbb{R}^d$, and define for $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_d) \in \mathcal{E}_d$ the function (column vector) $f_{\boldsymbol{\varepsilon}} \colon \mathbb{Z}_2^d \to \mathbb{R}$ as follows. For $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{Z}_2^d$,

$$f_{\boldsymbol{\varepsilon}}(\boldsymbol{x}) = \varepsilon_1^{x_1} \varepsilon_2^{x_2} \cdots \varepsilon_d^{x_d},$$

where of course $(\pm 1)^0 = 1$ and $(\pm 1)^1 = \pm 1$. It is immediate that

$$f_{\boldsymbol{\varepsilon}}(\boldsymbol{x} + \boldsymbol{y}) = f_{\boldsymbol{\varepsilon}}(\boldsymbol{x})\, f_{\boldsymbol{\varepsilon}}(\boldsymbol{y}) \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}_2^d.$$

Multiplying the matrix P by the vector $f_{\boldsymbol{\varepsilon}}$,

$$Pf_{\boldsymbol{\varepsilon}}(\boldsymbol{x}) = \frac{1}{d}\sum_{i=1}^{d} f_{\boldsymbol{\varepsilon}}(\boldsymbol{x}+\boldsymbol{e}_i) = \Big(\frac{1}{d}\sum_{i=1}^{d} f_{\boldsymbol{\varepsilon}}(\boldsymbol{e}_i)\Big) f_{\boldsymbol{\varepsilon}}(\boldsymbol{x}) = \frac{d-2k(\boldsymbol{\varepsilon})}{d}\, f_{\boldsymbol{\varepsilon}}(\boldsymbol{x}),$$

where $k(\boldsymbol{\varepsilon}) = |\{i : \varepsilon_i = -1\}|$, and thus $\sum_i \varepsilon_i = d - 2k(\boldsymbol{\varepsilon})$.

4.25 Exercise. Show that the functions $f_{\boldsymbol{\varepsilon}}$, where $\boldsymbol{\varepsilon} \in \mathcal{E}_d$, are linearly independent. □

Now, P is a $(2^d \times 2^d)$-matrix, and we have found 2^d linearly independent eigenfunctions (eigenvectors). For each $k \in \{0,\dots,d\}$, there are $\binom{d}{k}$ elements $\boldsymbol{\varepsilon} \in \mathcal{E}_d$ such that $k(\boldsymbol{\varepsilon}) = k$. We conclude that we have found all eigenvalues of P, and they are

$$\lambda_k = 1 - \frac{2k}{d} \quad \text{with multiplicity } \operatorname{mult}(\lambda_k) = \binom{d}{k}, \quad k = 0,\dots,d.$$

By Lemma 4.22,

$$p^{(n)}(x,x) = \frac{1}{2^d}\sum_{k=0}^{d}\binom{d}{k}\Big(1-\frac{2k}{d}\Big)^n.$$

Since P is periodic, we cannot use this random walk for approximating the equidistribution in $\mathbb{Z}_2^d$. We modify P, defining

$$Q = \frac{1}{d+1}\, I + \frac{d}{d+1}\, P.$$

This describes simple random walk on the graph which is obtained from the hypercube by adding to the edge set a loop at each vertex. Now every point has $d+1$ neighbours, including the point itself. The new random walk has law $\tilde{\mu}$ given by $\tilde{\mu}(\mathbf{0}) = \tilde{\mu}(\boldsymbol{e}_i) = 1/(d+1)$, $i = 1,\dots,d$. It is irreducible and aperiodic. We find $Qf_{\boldsymbol{\varepsilon}} = \lambda'_{k(\boldsymbol{\varepsilon})}\, f_{\boldsymbol{\varepsilon}}$, where

$$\lambda'_k = 1 - \frac{2k}{d+1} \quad \text{with multiplicity } \operatorname{mult}(\lambda_k) = \binom{d}{k}.$$

Therefore

$$q^{(n)}(x,x) = \frac{1}{2^d}\sum_{k=0}^{d}\binom{d}{k}\Big(1-\frac{2k}{d+1}\Big)^n. \tag{4.26}$$

Furthermore, $\lambda_1 = -\lambda_d = \frac{d-1}{d+1} = \lambda_*$. Applying (4.23), we obtain

$$\|q^{(n)}(x,\cdot) - \mathsf{m}\|_1 \le \sqrt{2^d - 1}\,\Big(\frac{d-1}{d+1}\Big)^n.$$

We can use this upper bound in order to estimate the necessary number of steps after which the random walk $(\mathbb{Z}^d, Q)$ approximates the uniform distribution m with an error smaller than e^{-C} (where $C > 0$): we have to solve the inequality

$$\sqrt{2^d - 1}\left(\frac{d-1}{d+1}\right)^n \leq e^{-C},$$

and find

$$n \geq \frac{C + \log\sqrt{2^d - 1}}{\log\left(1 + \frac{2}{d-1}\right)},$$

which is asymptotically (as $d \to \infty$) of the order of $\left(\frac{1}{4}\log 2\right)d^2 + \frac{C}{2}d$. We see that for large d, the contribution coming from C is negligible in comparison with the first, quadratic term.

Observe however that the upper bound on $q^{(2n)}(x, x)$ that has lead us to this estimate can be improved by performing an asymptotic evaluation (for $d \to \infty$) of the right hand term in (4.26), a nice combinatorial-analytic exercise.

4.27 Example (The Ehrenfest model). In relation with the discussion of Boltzmann's Theorem H of statistical mechanics, P. and T. EHRENFEST have proposed the following model in 1911. An urn (or box) contains N molecules. Furthermore, the box is separated in two halves (sides) A and B by a "wall" with a small membrane, see Figure 9. In each of the successive time instants, a single molecule chosen randomly among all N molecules crosses the membrane to the other half of the box.

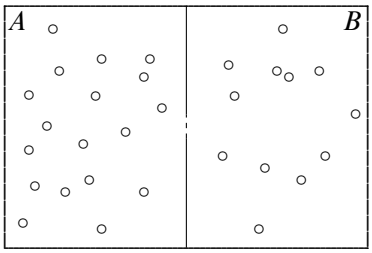

Figure 9

We ask the following two questions. (1) How can one describe the equilibrium, that is, the state of the box after a long time period: what is the approximate probability that side A of the box contains precisely k of the N molecules? (2) If initially side A is empty, how long does one have two wait for reaching the equilibrium, that is, how long does it take until the approximation of the equilibrium is good?

As a Markov chain, the Ehrenfest model is described on the state space $X = \{0, 1, \ldots, N\}$, where the states represent the number of molecules in side A.

The transition matrix, denoted $\bar{P}$ for a reason that will become apparent immediately, is given by

$$\bar{p}(j, j-1) = \frac{j}{N} \qquad (j = 1, \ldots, N)$$

and

$$\bar{p}(j, j+1) = \frac{N-j}{N} \quad (j = 0, \ldots, N-1),$$

where $\bar{p}(j, j-1)$ is the probability, given j particles in side A, that the randomly chosen molecule belongs to side A and moves to side B. Analogously, $\bar{p}(j, j+1)$ corresponds to the passage of a molecule from side B to side A.

This Markov chain cannot be described as a random walk on some group. However, let us reconsider SRW P on the hypercube $\mathbb{Z}_2^N$. We can subdivide the points of the hypercube into the classes

$$C_j = \{\boldsymbol{x} \in \mathbb{Z}_2^N : \boldsymbol{x} \text{ has } N-j \text{ zeros}\}, \quad j = 0, \ldots, N.$$

If $i, j \in \{0, \ldots, N\}$ and $\boldsymbol{x} \in C_i$ then $p(\boldsymbol{x}, C_j) = \bar{p}(i, j)$. This means that our partition satisfies the condition (1.29), so that one can construct the factor chain. The transition matrix of the latter is $\bar{P}$.

Starting with the Ehrenfest model, we can obtain the finer hypercube model by imagining that the molecules have labels 1 through N, and that the state $\boldsymbol{x} = (x_1, \ldots, x_N) \in \mathbb{Z}_2^N$ indicates that the molecules labeled i with $x_i = 0$ are currently on (or in) side A, while the others are on side B. Thus, passing from the hypercube to the Ehrenfest model means that we forget about the labels and just count the number of molecules in A (resp. B).

4.28 Exercise. Suppose that (X, P) is reversible with respect to the measure $\mathsf{m}(\cdot)$, and let $(\bar{X}, \bar{P})$ be a factor chain of (X, P) such that $\mathsf{m}(\bar{x}) < \infty$ for each class $\bar{x}$ in the partition $\bar{X}$ of X. Show that $(\bar{X}, \bar{P})$ is reversible. □

We get that $\bar{P}$ is reversible with respect to the measure on $\{0, \ldots, N\}$ given by

$$\bar{\mathsf{m}}(j) = \mathsf{m}(C_j) = \frac{1}{2^N} \binom{N}{j}.$$

This answers question (1).

The matrix $\bar{P}$ has period 2, and Theorem 3.28 cannot be applied. In this sense, question (2) is not well-posed. As in the case of the hypercube, we can modify the transition probabilities, considering the factor chain of Q, given by $\bar{Q} = \frac{1}{N+1} I + \frac{N}{N+1} \bar{P}$. (Here, I is the identity matrix over $\{0, \ldots, N\}$.) This means that at each step, no molecule crosses the membrane (with probability $\frac{1}{N+1}$), or one random

molecule crosses (with probability $\frac{1}{N+1}$ for each molecule). We obtain

$$\bar{q}(j, j-1) = \frac{j}{N+1} \quad (j = 1, \dots, N),$$
$$\bar{p}(j, j+1) = \frac{N-j}{N+1} \quad (j = 0, \dots, N-1)$$

and

$$\bar{q}(j, j) = \frac{1}{N+1} \quad (j = 0, \dots, N).$$

Then $\bar{Q}$ is again reversible with respect to the probability measure $\bar{\mathsf{m}}$. Since $C_0 = \{\mathbf{0}\}$, we have $\bar{q}^{(n)}(0,0) = q^{(n)}(\mathbf{0}, \mathbf{0})$. Therefore the bound of Lemma 4.12, applied to $\bar{Q}$, becomes

$$\|\bar{q}^{(n)}(0, \cdot) - \bar{\mathsf{m}}\|_1 \le 2^N \, q^{(2n)}(\mathbf{0}, \mathbf{0}) - 1,$$

which leads to the same estimate of the number of steps to reach stationarity as in the case of the hypercube: the approximation error is smaller than e^{-C}, if

$$n \ge \frac{C + \log\sqrt{2^N - 1}}{\log\bigl(1 + \frac{2}{N-1}\bigr)} \sim \Bigl(\frac{1}{4}\log 2\Bigr) N^2 \quad \text{as } N \to \infty.$$

This holds when the starting point is 0 (or N), which means that at the beginning of the process, side A (or side B, respectively) is empty. The reason is that the classes C_0 and C_N consist of single elements. For a general starting point $j \in \{0, \dots, N\}$, we can use the estimate of Theorem 4.15 with $\lambda^* = \frac{N-1}{N+1}$, so that

$$\|\bar{q}^{(n)}(j, \cdot) - \bar{\mathsf{m}}\|_1 \le \sqrt{\frac{2^N}{\binom{N}{j}} - 1} \left(\frac{N-1}{N+1}\right)^n.$$

Thus, the approximation error is smaller than e^{-C}, if

$$n \ge \frac{C + \log\sqrt{\Bigl(2^N / \binom{N}{j}\Bigr) - 1}}{\log\bigl(1 + \frac{2}{N-1}\bigr)} \sim \frac{N}{4} \log \frac{2^N}{\binom{N}{j}} \quad \text{as } N \to \infty. \tag{4.29}$$

4.30 Exercise. Use Stirling's formula to show that when N is even and $j = N/2$, the asymptotic behaviour of the error estimate in (4.29) is

$$\frac{N}{4} \log \frac{2^N}{\binom{N}{N/2}} \sim \frac{N \log N}{8} \quad \text{as } N \to \infty. \qquad \square$$

Thus, a good approximation of the equilibrium is obtained after a number of steps of order $N \log N$, when at the beginning both sides A and B contain the same number of molecules, while our estimate gives a number of steps of order N^2, if at the beginning all molecules stay in one of the two sides.

C The Poincaré inequality

In many cases, λ_1 and $\lambda_{\min}$ cannot be computed explicitly. Then one needs methods for estimating these numbers. The first main task is to find good upper bounds for λ_1. A range of methods uses the geometry of the graph of the Markov chain in order to find such bounds. Here we present only the most basic method of this type, taken from the seminal paper by Diaconis and Stroock [15].

As above, we assume that X is finite and that P is reversible with invariant probability measure m. We can consider the discrete probability space (X, m). For any function $f : X \to \mathbb{R}$, its mean and variance with respect to m are

$$\mathsf{E}_\mathsf{m}(f) = \sum_x f(x)\,\mathsf{m}(x) = (f, \mathbf{1}_X) \quad \text{and}$$
$$\mathsf{Var}_\mathsf{m}(f) = \mathsf{E}_\mathsf{m}\big(f - \mathsf{E}_\mathsf{m}(f)\big)^2 = (f, f) - (f, \mathbf{1}_X)^2,$$

with the inner product $(\cdot, \cdot)$ as in (4.4).

4.31 Exercise. Verify that

$$\mathsf{Var}_\mathsf{m}(f) = \frac{1}{2} \sum_{x,y \in X} \big(f(y) - f(x)\big)^2\, \mathsf{m}(x)\, \mathsf{m}(y). \qquad \square$$

The *Dirichlet norm* or *Dirichlet sum* of a function $f : X \to \mathbb{R}$ is

$$\begin{aligned} D(f) = \langle \nabla f, \nabla f \rangle &= \frac{1}{2} \sum_{e \in E} \frac{\big(f(e^+) - f(e^-)\big)^2}{r(e)} \\ &= \frac{1}{2} \sum_{x,y \in X} \big(f(x) - f(y)\big)^2 \mathsf{m}(x)\, p(x, y). \end{aligned} \tag{4.32}$$

The following is well known matrix analysis.

4.33 Proposition.

$$1 - \lambda_1 = \min \left\{ \frac{D(f)}{\mathsf{Var}_\mathsf{m}(f)} \;\middle|\; f : X \to \mathbb{R} \text{ non-constant} \right\}.$$

Proof. First of all, it is well known that

$$\lambda_1 = \max\left\{\frac{(Pf,f)}{(f,f)} : f \neq 0,\ f\perp \mathbf{1}_X\right\}. \tag{4.34}$$

Indeed, the eigenspace with respect to the largest eigenvalue $\lambda_o = 1$ is spanned by the function (column vector) $\mathbf{1}_X$, so that λ_1 is the largest eigenvector of P acting on the orthogonal complement of $\mathbf{1}_X$: we have an orthonormal basis f_x, $x \in X$, of $\ell^2(X,\mathsf{m})$ consisting of eigenfunctions of P with associated eigenvalues λ_x, such that $f_o = \mathbf{1}_X$, $\lambda_o = 1$, and $\lambda_x < 1$ for all $x \neq o$. If $f\perp\mathbf{1}_X$ then f is a linear combination $f = \sum_{x\neq o}(f,f_x)\cdot f_x$, and

$$(Pf,f) = \sum_{x\neq o}(f,f_x)\cdot(Pf_x,f) = \sum_{x\neq o}\lambda_x\,(f,f_x)^2 \le \lambda_1\sum_{x\neq o}(f,f_x)^2 = (f,f).$$

The maximum in (4.34) is attained for any λ_1-eigenfunction.

Since $D(f) = (\nabla^*\nabla f, f) = (f - Pf, f)$ by Exercise 4.10, we can rewrite (4.34) as

$$1-\lambda_1 = \min\left\{\frac{D(f)}{(f,f)} : f \neq 0,\ f\perp \mathbf{1}_X\right\}.$$

Now, if f is non-constant, then we can write the orthogonal decomposition

$$f = (f,\mathbf{1}_X)\cdot\mathbf{1}_X + g = \mathsf{E}_\mathsf{m}(f)\cdot\mathbf{1}_X + g, \quad \text{where } g\perp\mathbf{1}_X.$$

We then have $D(f) = D(g)$ for the Dirichlet norm, and $\mathsf{Var}_\mathsf{m}(f) = \mathsf{Var}_\mathsf{m}(g)$. Therefore the set of values over which the minimum is taken does not change when we replace the condition "$f\perp\mathbf{1}_X$" with "f non-constant". □

We now consider paths in the oriented graph $\Gamma(P)$ with edge set E. We choose a *length element* $l\colon E \to (0,\infty)$ with $l(\check{e}) = l(e)$ for each $e \in E$. If $\pi = [x_0, x_1, \dots, x_n]$ is such a path and $E(\pi) = \{[x_0,x_1],\dots,[x_{n-1},x_n]\}$ stands for the set of (oriented) edges of X on that path, then we write

$$|\pi|_l = \sum_{e\in E(\pi)} l(e)$$

for its length with respect to $l(\cdot)$. For $l(\cdot)\equiv 1$, we obtain the ordinary length (number of edges) $|\pi|_1 = n$. The other typical choice is $l(e) = r(e)$, in which case we get the *resistance length* $|\pi|_r$ of π.

We select for any ordered pair of points $x, y \in X$, $x \neq y$, a path $\pi_{x,y}$ from x to y. The following definition relies on the choice of all those $\pi_{x,y}$ and the length element $l(\cdot)$.

4.35 Definition. The *Poincaré constant* of the finite network $\mathcal{N} = (X, E, r)$ (with normalized resistances such that $\sum_{e\in E} 1/r(e) = 1$) is

$$\kappa_l = \max_{e\in E} \kappa_l(e), \quad \text{where } \kappa_l(e) = \frac{r(e)}{l(e)} \sum_{x,y\,:\,e\in E(\pi_{x,y})} |\pi_{x,y}|_l \; \mathsf{m}(x)\,\mathsf{m}(y).$$

For simple random walk on a finite graph, we have $\kappa_1 = \kappa_r$.

4.36 Theorem (Poincaré inequality). *The second largest eigenvalue of the reversible Markov chain (X, P), resp. the associated network $\mathcal{N} = (X, E, r)$, satisfies*

$$\lambda_1 \le 1 - \frac{1}{\kappa_l}.$$

with respect to any length element $l(\cdot)$ on E.

Proof. Let $f : X \to \mathbb{R}$ be any function, and let $x \neq y$ and $\pi_{x,y} = [x_0, x_1, \dots, x_n]$. Then by the Cauchy–Schwarz inequality

$$\begin{aligned} \big(f(y) - f(x)\big)^2 &= \Big(\sum_{i=1}^{n} \sqrt{l([x_{i-1}, x_i])} \cdot \frac{f(x_i) - f(x_{i-1})}{\sqrt{l([x_{i-1}, x_i])}}\Big)^2 \\ &\le \sum_{i=1}^{n} l([x_{i-1}, x_i]) \cdot \sum_{i=1}^{n} \frac{\big(f(x_i) - f(x_{i-1})\big)^2}{l([x_{i-1}, x_i])} \qquad (4.37) \\ &= |\pi_{x,y}|_l \sum_{e\in E(\pi_{x,y})} \frac{\big(\nabla f(e)\, r(e)\big)^2}{l(e)}. \end{aligned}$$

Therefore, using the formula of Exercise 4.31,

$$\begin{aligned} \mathsf{Var}_{\mathsf{m}}(f) &\le \frac{1}{2} \sum_{x,y\in X, x\neq y} |\pi_{x,y}|_l \; \mathsf{m}(x)\,\mathsf{m}(y) \sum_{e\in E(\pi_{x,y})} \frac{\big(\nabla f(e)\, r(e)\big)^2}{l(e)} \\ &= \frac{1}{2} \sum_{e\in E} \big(\nabla f(e)\big)^2 r(e) \underbrace{\sum_{\substack{x,y\in X:\\ e\in E(\pi_{x,y})}} \frac{r(e)}{l(e)} \; |\pi_{x,y}|_l \; \mathsf{m}(x)\,\mathsf{m}(y)}_{\kappa_l(e)} \le \kappa_l \cdot D(f). \end{aligned}$$

Together with Proposition 4.33, this proves the inequality. □

The applications of this inequality require a careful choice of the paths $\pi_{x,y}$, $x, y \in X$.

For *simple random walk* on a finite graph $\Gamma = (X, E)$, we have for the stationary probability and the associated resistances

$$\mathsf{m}(x) = \deg(x)/|E| \quad \text{and} \quad r(e) = |E|.$$

(Recall that m and thus also $r(\cdot)$ have to be normalized such that $\mathsf{m}(X) = 1$.) In particular, $\kappa_r = \kappa_1$, and

$$\begin{aligned} \kappa_1(e) &= \frac{1}{|E|} \sum_{x,y\,:\,e\in E(\pi_{x,y})} |\pi_{x,y}|_1 \, \deg(x) \, \deg(y) \\ &\le \kappa^* = \frac{1}{|E|} \max_{x,y\in X} |\pi_{x,y}|_1 \, \big(\max_{x\in X} \deg(x)\big)^2 \max_{e\in E} \gamma(e), \quad \text{where} \\ \gamma(e) &= \big|\{(x,y) \in X^2 : e \in \pi_{x,y}\}\big|, \end{aligned} \tag{4.38}$$

Thus, $\lambda_1 \le 1 - 1/\kappa^*$ for SRW. It is natural to use shortest paths, that is, $|\pi_{x,y}|_1 = d(x,y)$. In this case, $\max_{x,y\in X} |\pi_{x,y}|_1 = \operatorname{diam}(\Gamma)$ is the diameter of the graph Γ. If the graph is *regular,* i.e., $\deg(x) = \deg$ is constant, then $|E| = |X| \cdot \deg$, and

$$\begin{aligned} \kappa_1(e) &= \frac{\deg}{|X|} \sum_{k=1}^{\operatorname{diam}(\Gamma)} k \cdot \gamma_k(e), \quad \text{where} \\ \gamma_k(e) &= \big|\{(x,y) \in X^2 : |\pi_{x,y}|_1 = k, \ e \in \pi_{x,y}\}\big|. \end{aligned} \tag{4.39}$$

4.40 Example (Random walk on the hypercube). We refer to Example 4.24, and start with SRW on $\mathbb{Z}_2^d$, as in that example. We have $X = \mathbb{Z}_2^d$ and

$$E = \{[\boldsymbol{x}, \boldsymbol{x} + \boldsymbol{e}_i] : \boldsymbol{x} \in \mathbb{Z}_2^d, \ i = 1, \dots, d\}.$$

Thus, $|X| = 2^d$, $|E| = d\, 2^d$, $\operatorname{diam}(\Gamma) = d$, and $\deg(\boldsymbol{x}) = d$ for all $\boldsymbol{x}$.

We now describe a natural shortest path $\pi_{\boldsymbol{x},\boldsymbol{y}}$ from $\boldsymbol{x} = (x_1, \dots, x_d)$ to $\boldsymbol{y} = (y_1, \dots, y_d) \ne \boldsymbol{x}$. Let $1 \le i(1) < i(2) < \cdots < i(k) \le d$ be the coordinates where $y_i \ne x_i$, that is, $y_i = x_i + 1$ modulo 2. Then $d(x,y) = k$. We let

$$\begin{aligned} \pi_{\boldsymbol{x},\boldsymbol{y}} &= [\boldsymbol{x} = \boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k = \boldsymbol{y}], \quad \text{where} \\ \boldsymbol{x}_j &= \boldsymbol{x}_{j-1} + \boldsymbol{e}_{i(j)}, \quad j = 1, \dots, k. \end{aligned}$$

We first compute the number $\gamma(e)$ of (4.38) for $e = [\boldsymbol{u}, \boldsymbol{u} + \boldsymbol{e}_i]$ with $\boldsymbol{u} = (u_1, \dots, u_d) \in \mathbb{Z}_2^d$. We have $e \in E(\pi_{\boldsymbol{x},\boldsymbol{y}})$ precisely when $x_j = u_j$ for $j = i, \dots, d$, $y_i = u_i + 1 \mod 2$ and and $y_j = u_j$ for $j = 1, \dots, i-1$. There are 2^{d-i} free choices for the last $d - i$ coordinates of $\boldsymbol{x}$ and 2^{i-1} free choices for the first i coordinates of $\boldsymbol{y}$. Thus, $\gamma(e) = 2^{d-1}$ for every edge e. We get

$$\kappa^* = \frac{1}{d\, 2^d} \, d \cdot d^2 \cdot 2^{d-1} = \frac{d^2}{2}, \quad \text{whence } \lambda_1 \le 1 - \frac{2}{d^2}.$$

Our estimate of the *spectral gap* $1-\lambda_1 \geq 2/d^2$ misses the true value $1-\lambda_1 = 2/d$ by the factor of $1/d$.

Next, we can improve this crude bound by computing $\kappa_1(e)$ precisely. We need the numbers $\gamma_k(e)$ of (4.39). With $e = [\boldsymbol{u}, \boldsymbol{u}+\boldsymbol{e}_i]$, $\boldsymbol{x}$ and $\boldsymbol{y}$ as above such that $e \in E(\pi_{\boldsymbol{x},\boldsymbol{y}})$, we must have (mod 2)

$$\boldsymbol{x} = (x_1, \ldots, x_{i-1}, u_i, u_{i+1}, \ldots, u_d)$$

and

$$\boldsymbol{y} = (u_1, \ldots, u_{i-1}, u_i + 1, y_{i+1}, \ldots, y_d).$$

If $d(\boldsymbol{x}, \boldsymbol{y}) = k$ then $\boldsymbol{x}$ and $\boldsymbol{y}$ differ in precisely k coordinates. One of them is the i-th coordinate. There remain precisely $\binom{d-1}{k-1}$ free choices for the other coordinates where $\boldsymbol{x}$ and $\boldsymbol{y}$ differ. This number of choices is $\gamma_k(e)$. We get

$$\kappa_1 = \kappa_1(e) = \frac{d}{2^d} \sum_{k=1}^{d} k \binom{d-1}{k-1} = \frac{d}{2^d}(d+1)2^{d-2} = \frac{d(d+1)}{4},$$

and the new estimate $1 - \lambda_1 \geq 4/\big(d(d+1)\big)$ is only slightly better, missing the true value by the factor of $2/(d+1)$.

4.41 Exercise. Compute the Poincaré constant κ_1 for the Ehrenfest model of Example 4.27, and compare the resulting estimate with the true value of λ_1, as above.

[Hint: it will turn out after some combinatorial efforts that $\kappa_1(e)$ is constant over all edges.] □

4.42 Example (Card shuffling via random transpositions). What is the purpose of shuffling a deck of N cards? We want to simulate the equidistribution on all $N!$ permutations of the cards. We might imagine an urn containing $N!$ decks of cards, each in a different order, and pick one of those decks at random. This is of course not possible in practice, among other because $N!$ is too big. A reasonable algorithm is to first pick at random one among the N cards (with probability $1/N$ each), then to pick at random one among the remaining $N-1$ cards (with probability $N-1$ each), and so on. We will indeed end up with a random permutation of the cards, such that each permutation occurs with the same probability $1/N!$.

Card shuffling can be formalized as a random walk on the symmetric group $\mathfrak{S}_N$ of all permutations of the set $\{1, \ldots, N\}$ (an enumeration of the cards). The n-th shuffle corresponds to a random permutation Y_n ($n = 1, 2, \ldots$) in $\mathfrak{S}_N$, and a fair shuffler will perform the single shuffles such that they are independent. The random permutation obtained after n shuffles will be the product $Y_1 \ldots Y_n$. Thus, if the law μ of Y_n is such that the resulting random walk is irreducible (compare with Lemma 4.21), then we have a Markov chain whose stationary distribution is equidistribution on $\mathfrak{S}_N$. If, in addition, this random walk is also aperiodic, then the

convergence theorem tells us that for large n, the distribution of $Y_1 \cdots Y_n$ will be a good approximation of that uniform distribution. This explains the true purpose of card shuffling, although one may guess that most card shufflers are not aware of such a justification of their activity.

One of the most common method of shuffling, the *riffle shuffle,* has been analyzed by BAYER and DIACONIS [4] in a piece of work that has become very popular, see also MANN [43] for a nice exposition. Here, we consider another shuffling model, or random walk on $\mathfrak{S}_N$, generated by *random transpositions.*

Throughout this example, we shall write $x, y, \ldots$ for permutations of the set $\{1, \ldots, N\}$, and *id* for the identity. Also, we write the composition of permutations from left to right, that is, $(xy)(i) = y\big(x(i)\big)$. The transposition of the (distinct) elements i, j is denoted $t_{i,j}$. The law μ of our random walk is equidistribution on the set $\mathfrak{T}$ of all transpositions. Thus,

$$p(x, y) = \begin{cases} \dfrac{2}{N(N-1)}, & \text{if } \ y = x\,t \ \text{ for some } \ t \in \mathfrak{T}, \\ 0, & \text{otherwise.} \end{cases}$$

(Note that if $y = x\,t$ then $x = y\,t$.) The stationary probability measure and the resistances of the edges are given by

$$\mathsf{m}(x) = \frac{1}{N!} \quad \text{and} \quad r(e) = \frac{N(N-1)}{2}\,N!, \ \text{ where } e = [x, x\,t], \ x \in \mathfrak{S}_N, \ t \in \mathfrak{T}.$$

For $m < N$, we consider $\mathfrak{S}_m$ as a subgroup of $\mathfrak{S}_N$ via the identification

$$\mathfrak{S}_m = \{x \in \mathfrak{S}_N : x(i) = i \text{ for } i = m+1, \ldots, N\}.$$

Claim 1. Every $x \in \mathfrak{S}_N \setminus \mathfrak{S}_{N-1}$ has a unique decomposition

$$x = y\,t, \quad \text{where } y \in \mathfrak{S}_{N-1} \text{ and } t \in \mathfrak{T}_N = \{t_{j,N} : j = 1, \ldots, N-1\}.$$

Proof. Since $x \notin \mathfrak{S}_{N-1}$, we must have $x(N) = j \in \{1, \ldots, N-1\}$. Set $y = x\,t_{j,N}$. Then $y(N) = t_{j,N}\big(x(N)\big) = N$. Therefore $y \in \mathfrak{S}_{N-1}$ and $x = y\,t_{j,N}$. If there is another decomposition $x = y'\,t_{j',N}$ then $t_{j,N}t_{j',N} = y^{-1}y' \in \mathfrak{S}_{N-1}$, whence $j = j'$ and $y = y'$. □

Thus, every $x \in \mathfrak{S}_N$ has a unique decomposition

$$x = t_{j(1),m(1)} \cdots t_{j(k),m(k)} \tag{4.43}$$

with $0 \le k \le N-1$, $2 \le m(1) < \cdots < m(k) \le N$ and $1 \le j(i) < m(i)$ for all i.

In the graph $\Gamma(P)$, we can consider only those edges $[x, x\,t]$ and $[x\,t, x]$, where $x \in \mathfrak{S}_m$ and $t \in \mathfrak{T}_{m'}$ with $m' > m$. Then we obtain a *spanning tree* of $\Gamma(P)$, that is, a subgraph that contains all vertices but no cycle (A cycle is a

sequence $[x_0, x_1, \ldots x_{k-1}, x_k = x_0]$ such that $k \geq 3$, $x_0, \ldots, x_{k-1}$ are distinct, and $[x_{i-1}, x_i] \in E$ for $i = 1, \ldots, k$.) The tree is *rooted;* the root vertex is *id*. In Figure 10, this tree is shown for $\mathfrak{S}_4$. Since the graph is symmetric, we have drawn non-oriented edges. Each edge $[x, x\,t]$ is labelled with the transposition $t = t_{i,j} = (i, j)$, written in cycle notation. The permutations corresponding to the vertices are obtained by multiplying the transpositions along the edges on the shortest path ("geodesic") from *id* to the respective vertex. Those permutations are again written in cycle notation in Figure 10.

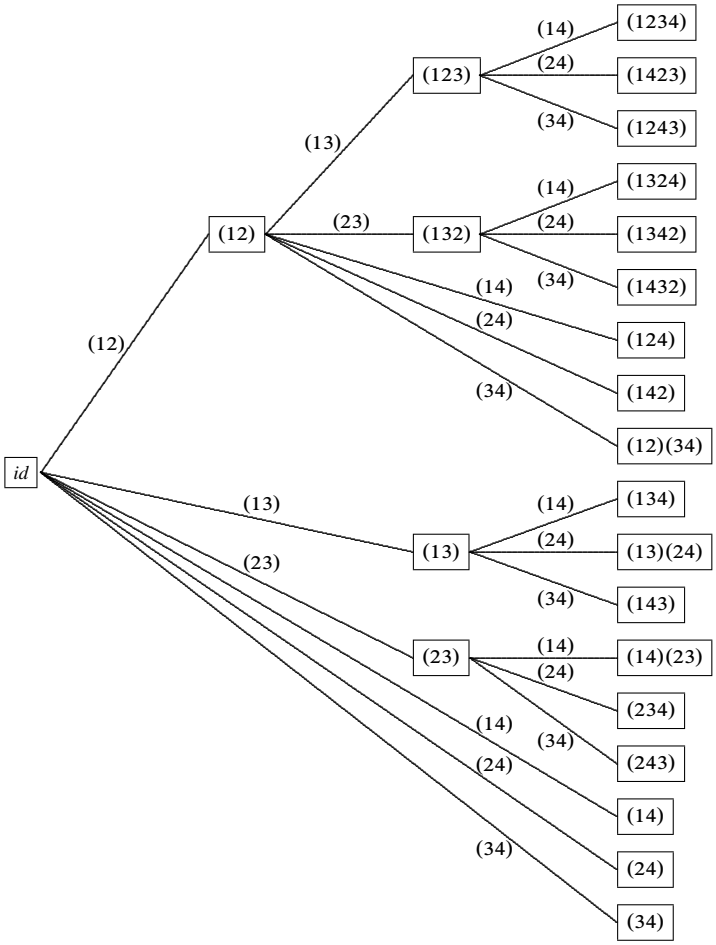

Figure 10. The spanning tree of $\mathfrak{S}_4$.

If $x \in \mathfrak{S}_N \setminus \{id\}$ is decomposed into transpositions as in (4.43), then we choose $\pi_{id,x} = [x_0 = id, x_1, \dots, x_k = x]$ with $x_i = t_{j(1),M(1)} \dots t_{j(i),M(i)}$. This is the shortest path from id to x in the spanning tree. Then we let $\pi_{x,y} = x \cdot \pi_{id,x^{-1}y}$.

For a transposition $t = t_{j,m}$, where $j < m$, we say that an edge e of $\Gamma(P)$ is of *type* t, if it has the form $e = [u, u\,t]$ with $u \in \mathfrak{S}_N$. We want to determine the number κ^* of (4.38).

Claim 2. If $e = [u, u\,t]$ is an edge of type t then

$$\gamma(e) = \big|\big\{x \in \mathfrak{S}_n \setminus \{id\} : \pi_{id,x} \text{ contains an edge of type } t\big\}\big|.$$

Proof. Let $\Gamma(e) = \big\{(x, y) \in \mathfrak{S}_N : x \neq y,\ e \in E(\pi_{x,y})\big\}$ and

$$\Pi(t) = \big\{x \in \mathfrak{S}_N \setminus \{id\} : \pi_{id,x} \text{ contains an edge of type } t\big\}.$$

Then $\gamma(e) = |\Gamma(e)|$ by definition. We show that the mapping $(x, y) \mapsto x^{-1}y$ is a bijection from $\Gamma(e)$ to $\Pi(t)$.

First of all, if $(x, y) \in \Gamma(e)$ then by definition of $\pi_{x,y}$, the edge $[x^{-1}u, x^{-1}ut]$ belongs to $\pi_{id,x^{-1}y}$. Therefore $x^{-1}y \in \Pi(t)$.

Second, if $w \in \Pi_k(t)$ then the decomposition (4.43) of w (in the place of x) contains t. We can write this decomposition as $w = w_1\,t\,w_2$. Then the *unique* edge of type t on $\pi_{id,w}$ is $[w_1, w_1\,t]$. We set $x = u\,w_1^{-1}$ and $y = u\,t\,w_2$. Then $x^{-1}y = w$, so that $\pi_{x,y} = x \cdot \pi_{id,w}$ contains the edge $[x\,w_1, x\,w_1 t] = e$. Thus, the mapping is surjective, and since the edge of type t on $\pi_{id,w}$ is unique, the mapping is also one-to-one (injective). □

So we next have to compute the cardinality of $\Pi(t)$. Let $t = (j, m)$ with $j < m$. By (4.43), every $x \in \Pi(t)$ can be written uniquely as $x = u\,t\,y$, where $u \in \mathfrak{S}_{m-1}$ (and any such u may occur) and y is any element of the form $y = t_{j(1),m(1)} \cdots t_{j(k),m(k)}$ with $m < m(1) < m(k) \le N$ and $j(i) < m(i)$ for all i. (This includes the case $k = 0$, $y = id$.) Thus, we have precisely $(m-1)!$ choices for u. On the other hand, since every element of $\mathfrak{S}_N$ can be written uniquely as $v\,y$ where $v \in \mathfrak{S}_m$ and y has the same form as above, we conclude that the set of all valid elements y forms a set of representatives of the right cosets of $\mathfrak{S}_m$ in $\mathfrak{S}_N$:

$$\mathfrak{S}_N = \biguplus_y \mathfrak{S}_m\,y.$$

The number of those cosets is $N!/m!$, and this is also the number of choices for y in the decomposition $x = u\,t\,y$. Therefore, if e is an edge of type $t = (j, m)$ then

$$\gamma(e) = |\Pi(t)| = (m-1)!\,\frac{N!}{m!} = \frac{N!}{m}.$$

The maximum is obtained when $m = 2$, that is, when $t = (1, 2)$. We compute

$$\kappa^* = \frac{1}{N!} \cdot (N-1) \cdot \frac{N(N-1)}{2} \cdot \frac{N!}{2} = \frac{N(N-1)^2}{4}, \quad \text{and}$$
$$\lambda_1 = \lambda_1(P) \le 1 - \frac{4}{N(N-1)^2}.$$

Again, this random walk has period 2, while we need an aperiodic one in order to approximate the equidistribution on $\mathfrak{S}_N$. Therefore we modify the random transposition of each single shuffle as follows: we select independently and at random (with probability $1/N!$ each) two indices $i, j \in \{1, \dots, N\}$ and exchange the corresponding cards, when $i \neq j$. When $i = j$, no card is moved. The transition probabilities of the resulting random walk of $\mathfrak{S}_N$ are

$$q(x, y) = \begin{cases} 1/N, & \text{if } y = x, \\ 1/N^2, & \text{if } y = x\,t \text{ with } t \in \mathfrak{T}, \\ 0, & \text{in all other cases.} \end{cases}$$

In terms of transition matrices, $Q = \frac{1}{N} I + \frac{N-1}{N} P$. Therefore

$$\lambda_1(Q) = \frac{1}{N} + \frac{N-1}{N} \lambda_1(P) \le 1 - \frac{4}{N^2(N-1)}.$$

By Exercise 4.17, $\lambda_{\min} = -1 + \frac{2}{N}$. Therefore

$$\lambda^*(Q) \le 1 - \frac{4}{N^2(N-1)}.$$

Equation 4.23 leads to

$$\|p^{(n)}(x, \cdot) - \mathsf{m}\|_1 \le \sqrt{N! - 1} \left(1 - \frac{4}{N^2(N-1)}\right)^n.$$

Thus, if we want to be sure that $\|p^{(n)}(x, \cdot) - \mathsf{m}\|_1 < e^{-C}$ then we can choose

$$n \ge \left(C + \tfrac{1}{2} \log(N! - 1)\right) \Big/ \left(-\log\left(1 - \tfrac{4}{N^2(N-1)}\right)\right) \sim \tfrac{N^3}{4} C + \tfrac{N^4}{8} \log N.$$

The above bound for $\lambda^*(Q)$ can be improved by additional combinatorial efforts. However, also in this case the true value is known: $\lambda^*(Q) = 1 - \frac{2}{N}$, see Diaconis and Shahshahani [14].

D Recurrence of infinite networks

In this section, we assume that $\mathcal{N} = (X, E, r)$ is an *infinite* network associated with a reversible, irreducible Markov chain (X, P). We want to establish a set of recurrence (resp. transience) criteria that can actually be applied in a variety of cases. The network is called recurrent (resp. transient), if (X, P) has the respective property.

We shall need very basic properties of Hilbert spaces, namely the Cauchy–Schwarz inequality and the fact that any non-empty closed convex set in a Hilbert space has a unique element with minimal norm.

The *Dirichlet space* $\mathcal{D}(\mathcal{N})$ associated with the network consists of all functions f on X (not necessarily in $\ell^2(X, \mathsf{m})$) such that $\nabla f \in \ell^2_\sharp(E, r)$. The Dirichlet norm $D(f)$ of such a function was defined in (4.32). The kernel of this quasi-norm consists of the constant functions on X. On the space $\mathcal{D}(\mathcal{N})$, we define an inner product with respect to a reference point $o \in X$:

$$(f, g)_D = (f, g)_{D,o} = \langle \nabla f, \nabla g \rangle + f(o)g(o).$$

4.44 Lemma. (a) *$\mathcal{D}(\mathcal{N})$ is a Hilbert space.*

(b) *For each $x \in X$, there is a constant $C_x > 0$ such that*

$$C_x^{-1}\,(f, f)_{D,o} \le (f, f)_{D,x} \le C_x\,(f, f)_{D,o}$$

for all $f \in \mathcal{D}(\mathcal{N})$. That is, changing the reference point o gives rise to an equivalent Hilbert space norm.

(c) *Convergence of a sequence of functions in $\mathcal{D}(\mathcal{N})$ implies their pointwise convergence.*

Proof. Let $x \in X \setminus \{o\}$. By connectedness of the graph $\Gamma(P)$, there is a path $\pi_{o,x} = [o = x_0, x_1, \dots, x_k = x]$ with edges $e_i = [x_{i-1}, x_i] \in E$. Then for $f \in \mathcal{D}(\mathcal{N})$, using the Cauchy–Schwarz inequality as in Theorem 4.36,

$$\big(f(x) - f(o)\big)^2 = \Big(\sum_{i=1}^{k} \frac{f(x_i) - f(x_{i-1})}{\sqrt{r(e_i)}} \sqrt{r(e_i)}\Big)^2 \le c_x\, D(f),$$

where $c_x = \sum_{i=1}^{k} r(e_i) = |\pi_{o,x}|_r$. Therefore

$$f(x)^2 \le 2\big(f(x) - f(o)\big)^2 + 2f(o)^2 \le 2c_x\, D(f) + 2f(o)^2,$$

and

$$(f, f)_{D,x} = D(f) + f(x)^2 \le (2c_x + 1)D(f) + 2f(o)^2 \le C_x (f, f)_{D,o},$$

where $C_x = \max\{2c_x + 1, 2\}$. Exchanging the roles of o and x, we get $(f, f)_{D,o} \le C_x (f, f)_{D,x}$ with the same constant C_x. This proves (b).

We next show that $\mathcal{D}(\mathcal{N})$ is complete. Let (f_n) be a Cauchy sequence in $\mathcal{D}(\mathcal{N})$, and let $x \in X$. Then, by (b),

$$\big(f_n(x) - f_m(x)\big)^2 \le (f_n - f_m, f_n - f_m)_{D,x} \to 0 \quad \text{as } m, n \to \infty.$$

Therefore there is a function f on X such that $f_n \to f$ pointwise. On the other hand, as $m, n \to \infty$,

$$\langle \nabla(f_n - f_m), \nabla(f_n - f_m) \rangle = D(f_n - f_m) \le (f_n - f_m, f_n - f_m)_{D,o} \to 0.$$

Thus (∇f_n) is a Cauchy sequence in the Hilbert space $\ell^2_\sharp(E, r)$. Hence, there is $\phi \in \ell^2_\sharp(E, r)$ such that $\nabla f_n \to \phi$ in that Hilbert space. Convergence of a sequence of functions in the latter implies pointwise convergence. Thus

$$\phi(e) = \lim_{n\to\infty} \frac{f_n(e^+) - f_n(e^-)}{r(e)} = \nabla f(e)$$

for each edge $e \in E$. We obtain $D(f) = \langle \phi, \phi \rangle < \infty$, so that $f \in \mathcal{D}(\mathcal{N})$. To conclude the proof of (a), we must show that $(f_n - f, f_n - f)_{D,o} \to 0$ as $n \to \infty$. But this is true, since both $D(f_n - f) = \langle \nabla f_n - \phi, \nabla f_n - \phi \rangle$ and $\big(f_n(o) - f(o)\big)^2$ tend to 0, as we have just seen.

The proof of (c) is contained in what we have proved above. □

4.45 Exercise. Extension of Exercise 4.10. Show that $\nabla^*(\nabla f) = -\mathfrak{L} f$ for every $f \in \mathcal{D}(\mathcal{N})$, even when the graph $\Gamma(P)$ is not locally finite.

[Hint: use the Cauchy–Schwarz inequality once more to check that for each x, the sum $\sum_{y:[x,y]\in E} |f(x) - f(y)|\, a(x, y)$ is finite.] □

We denote by $\mathcal{D}_0(\mathcal{N})$ the closure in $\mathcal{D}(\mathcal{N})$ of the linear space $\ell_0(X)$ of all finitely supported real functions on X. Since ∇ is a bounded operator, $\ell^2(X, \mathsf{m}) \subset \mathcal{D}_0(\mathcal{N})$ as sets, but in general the Hilbert space norms are not comparable, and equality in the place of "$\subset$" does in general not hold.

Recall the definitions (2.15) and (2.16) of the restriction of P to a subset A of X and the associated Green kernel $G_A(\cdot, \cdot)$, which is finite by Lemma 2.18.

4.46 Lemma. *Suppose that $A \subset X$ is finite, $x \in A$, and let $f \in \ell_0(X)$ be such that* $\operatorname{supp}(f) \subset A$. *Then*

$$\langle \nabla f, \nabla G_A(\cdot, x) \rangle = \mathsf{m}(x) f(x).$$

Proof. The functions f and $G_A(\cdot, x)$ are 0 outside A, and we can use Exercise 4.45 :

$$\begin{aligned}\langle \nabla f, \nabla G_A(\cdot, x)\rangle &= \big(f, (I - P)G_A(\cdot, x)\big)\\ &= \sum_{y\in A} f(y)\Big(G_A(y, x) - \underbrace{\sum_{w\in X} p(y, w)G_A(w, x)}_{=0,\ \text{if } w\notin A}\Big)\\ &= \big(f, (I_A - P_A)G_A(\cdot, x)\big) = (f, \mathbf{1}_x).\end{aligned}$$

In the last step, we have used (2.17) with $z = 1$. □

4.47 Lemma. *If (X, P) is transient, then $G(\cdot, x) \in \mathcal{D}_0(\mathcal{N})$ for every $x \in X$.*

Proof. Let $A \subset B$ be two finite subsets of X containing x. Setting $f = G_A(\cdot, x)$, Lemma 4.46 yields

$$\langle \nabla G_A(\cdot, x), \nabla G_A(\cdot, x)\rangle = \langle \nabla G_A(\cdot, x), \nabla G_B(\cdot, x)\rangle = \mathsf{m}(x)G_A(x, x).$$

Analogously, setting $f = G_B(\cdot, x)$,

$$\langle \nabla G_B(\cdot, x), \nabla G_B(\cdot, x)\rangle = \mathsf{m}(x)G_B(x, x).$$

Therefore

$$\begin{aligned}&D\big(G_B(\cdot, x) - G_A(\cdot, x)\big)\\ &\quad= \langle \nabla G_B(\cdot, x), \nabla G_B(\cdot, x)\rangle\\ &\qquad - 2\langle \nabla G_A(\cdot, x), \nabla G_B(\cdot, x)\rangle + \langle \nabla G_A(\cdot, x), \nabla G_A(\cdot, x)\rangle\\ &\quad= \mathsf{m}(x)\big(G_B(x, x) - G_A(x, x)\big).\end{aligned}$$

Now let $(A_k)_{k\geq 1}$ be an increasing sequence of finite subsets of X with union X. Using (2.15), we see that $p^{(n)}_{A_k}(x, y) \to p^{(n)}(x, y)$ monotonically from below as $k \to \infty$, for each fixed n. Therefore, by monotone convergence, $G_{A_k}(x, x)$ tends to $G(x, x)$ monotonically from below. Hence, by the above, the sequence of functions $\big(G_{A_k}(\cdot, x)\big)_{k\geq 1}$ is a Cauchy sequence in $\mathcal{D}(\mathcal{N})$. By Lemma 4.44, it converges in $\mathcal{D}(\mathcal{N})$ to its pointwise limit, which is $G(\cdot, x)$. Thus, $G(\cdot, x)$ is the limit in $\mathcal{D}(\mathcal{N})$ of a sequence of finitely supported functions. □

4.48 Definition. Let $x \in X$ and $i_0 \in \mathbb{R}$. A *flow with finite power from x to ∞ with input i_0* (also called *unit flow* when $i_0 = 1$) in the network $\mathcal{N}$ is a function $\phi \in \ell^2_\sharp(E, r)$ such that

$$\nabla^*\phi(y) = -\frac{i_0}{\mathsf{m}(x)}\,\mathbf{1}_x(y) \quad \text{for all } y \in X.$$

Its *power* is $\langle \phi, \phi\rangle$.[1]

[1] Many authors call $\langle \phi, \phi\rangle$ the *energy* of ϕ, but in the correct physical interpretation, this should be the power.

The condition means that Kirchhoff's node law is satisfied at every point except x,

$$\sum_{e\in E:e^+=y} \phi(e) = \begin{cases} 0, & y \neq x, \\ -i_0, & y = x. \end{cases}$$

As explained at the beginning of this chapter, we may think of the network as a system of tubes; each edge e is a tube with length $r(e)$ *cm* and cross-section 1 cm^2, and the tubes are connected at their endpoints (vertices) according to the given graph structure. The network is filled with (incompressible) liquid, and at the source x, liquid is injected at a constant rate of i_0 liters per second. Requiring that this be possible with finite power ("effort") $\langle \phi, \phi \rangle$ is absurd if the network is finite (unless $i_0 = 0$). The main purpose of this section is to show that the existence of such flows characterizes transient networks: even though the network is filled, it is so "big at infinity", that the permanently injected liquid can flow off towards infinity at the cost of a finite effort. With this interpretation, recurrent networks correspond more to our intuition of the "real world". An analogous interpretation can of course be given in terms of voltages and electric current.b

4.49 Definition. The *capacity* of a point $x \in X$ is

$$\operatorname{cap}(x) = \inf\{D(f) : f \in \ell_0(X),\ f(x) = 1\}.$$

4.50 Lemma. *It holds that* $\operatorname{cap}(x) = \min\{D(f) : f \in \mathcal{D}_0(\mathcal{N}),\ f(x) = 1\}$, *and the minimum is attained by a unique function in this set.*

Proof. First of all, consider the closure $\mathcal{C}^-$ of $\mathcal{C} = \{f \in \ell_0(X) : f(x) = 1\}$ in $\mathcal{D}(\mathcal{N})$. Every function in $\mathcal{C}^-$ must be in $\mathcal{D}_0(\mathcal{N})$, and (since convergence in $\mathcal{D}(\mathcal{N})$ implies pointwise convergence) has value 1 in x. We see that the inclusion $\mathcal{C}^- \subset \{f \in \mathcal{D}_0(\mathcal{N}) : f(x) = 1\}$ holds. Conversely, let $f \in \mathcal{D}_0(\mathcal{N})$ with $f(x) = 1$. By definition of $\mathcal{D}_0(\mathcal{N})$ there is a sequence of functions f_n in $\ell_0(X)$ such that $f_n \to f$ in $\mathcal{D}(\mathcal{N})$, and in particular $\lambda_n^{-1} = f_n(x) \to f(x) = 1$. But then $\lambda_n \cdot f_n \in \mathcal{C}$ and $\lambda_n \cdot f_n \to f$ in $\mathcal{D}(\mathcal{N})$, since in addition to convergence in the point x we have $D(\lambda_n \cdot f_n - f_n) = (\lambda_n - 1)^2 D(f_n) \to 0 \cdot D(f) = 0$, that is, f_n and $\lambda_n \cdot f_n$ have the same limit in $\mathcal{D}(\mathcal{N})$. We conclude that

$$\mathcal{C}^- = \{f \in \mathcal{D}_0(\mathcal{N}) : f(x) = 1\}.$$

This is a closed convex set in the Hilbert space. It is a basic theorem in Hilbert space theory that such a set possesses a unique element with minimal norm; see e.g. RUDIN [Ru, Theorem 4.10]. Thus, if this element is f_0 then

$$D(f_0) = \min\{D(f) : f \in \mathcal{C}^-\} = \inf\{D(f) : f \in \mathcal{C}\} = \operatorname{cap}(x),$$

since $f \mapsto D(f)$ is continuous. □

The *flow criterion* is now part (b) of the following useful set of necessary and sufficient transience criteria.

4.51 Theorem. *For the network $\mathcal{N}$ associated with a reversible Markov chain (X, P), the following statements are equivalent.*

(a) *The network is transient.*

(b) *For some ($\Longleftrightarrow$ every) $x \in X$, there is a flow from x to ∞ with non-zero input and finite power.*

(c) *For some ($\Longleftrightarrow$ every) $x \in X$, one has* $\operatorname{cap}(x) > 0$.

(d) *The constant function* $\mathbf{1}$ *does not belong to* $\mathcal{D}_0(\mathcal{N})$.

Proof. (a) $\Longrightarrow$ (b). If the network is transient, then $G(\cdot, x) \in \mathcal{D}_0(\mathcal{N})$ by Lemma 4.47. We define $\phi = -\frac{i_0}{\mathsf{m}(x)} \nabla G(\cdot, x)$. Then $\phi \in \ell^2_\sharp(E, r)$, and by Exercise 4.45

$$\nabla^* \phi = \frac{i_0}{\mathsf{m}(x)} \mathfrak{L}\, G(\cdot, x) = -\frac{i_0}{\mathsf{m}(x)} \mathbf{1}_x.$$

Thus, ϕ is a flow from x to ∞ with input i_0 and finite power.

(b) $\Longrightarrow$ (c). Suppose that there is a flow ϕ from x to ∞ with input $i_0 \neq 0$ and finite power. We may normalize ϕ such that $i_0 = -1$. Now let $f \in \ell_0(X)$ with $f(x) = 1$. Then

$$\langle \nabla f, \phi \rangle = (f, \nabla^* \phi) = \left(f, \frac{1}{\mathsf{m}(x)} \mathbf{1}_x \right) = f(x) = 1.$$

Hence, by the Cauchy–Schwarz inequality,

$$1 = |\langle \nabla f, \phi \rangle|^2 \leq \langle \nabla f, \nabla f \rangle \, \langle \phi, \phi \rangle = D(f) \, \langle \phi, \phi \rangle.$$

We obtain $\operatorname{cap}(x) \geq 1/\langle \phi, \phi \rangle > 0$.

(c) $\Longleftrightarrow$ (d). This follows from Lemma 4.50: we have $\operatorname{cap}(x) = 0$ if and only if there is $f \in \mathcal{D}_0(\mathcal{N})$ with $f(x) = 1$ and $D(f) = 0$, that is, $f = \mathbf{1}$ is in $\mathcal{D}_0(\mathcal{N})$. Indeed, connectedness of the network implies that a function with $D(f) = 0$ must be constant.

(c) $\Longrightarrow$ (a). Let $A \subset X$ be finite, with $x \in A$. Set $f = G_A(\cdot, x)/G_A(x, x)$. Then $f \in \ell_0(X)$ and $f(x) = 1$. We use Lemma 4.46 and get

$$\operatorname{cap}(x) \leq D(f) = \frac{1}{G_A(x, x)^2} \langle \nabla G_A(\cdot, x), \nabla G_A(\cdot, x) \rangle = \frac{\mathsf{m}(x)}{G_A(x, x)}.$$

We obtain that $G_A(x, x) \leq \mathsf{m}(x)/\operatorname{cap}(x)$ for every finite $A \subset X$ containing x. Now take an increasing sequence $(A_k)_{k \geq 1}$ of finite sets containing x, whose union is X. Then, by monotone convergence,

$$G(x, x) = \lim_{k \to \infty} G_{A_k}(x, x) \leq \mathsf{m}(x)/\operatorname{cap}(x) < \infty. \qquad \square$$

4.52 Exercise. Prove the following in the transient case. The flow from x to ∞ with input 1 and minimal power is given by $\phi = -\nabla G(\cdot, x)/\mathsf{m}(x)$, its power (the *resistance between x and* ∞) is $G(x,x)/\mathsf{m}(x)$, and $\mathrm{cap}(x) = \mathsf{m}(x)/G(x,x)$. □

The flow criterion became popular through the work of T. Lyons [42]. A few years earlier, Yamasaki [50] had proved the equivalence of the statements of Theorem 4.51 for locally finite networks in a less common terminology of potential theory on networks. The interpretation in terms of Markov chains is not present in [50]; instead, finiteness of the Green kernel $G(x,y)$ is formulated in a non-probabilistic spirit. For non locally finite networks, see Soardi and Yamasaki [48].

If $A \subset X$ then we write E_A for the set of edges with both endpoints in A, and ∂A for all edges e with $e^- \in A$ and $e^+ \in X \setminus A$ (the *edge boundary* of A). Below in Example 4.63 we shall need the following.

4.53 Lemma. *Let ϕ be a flow from x to ∞ with input i_0, and let $A \subset X$ be finite with $x \in A$. Then*

$$\sum_{e \in \partial A} \phi(e) = i_0.$$

Proof. Recall that $\phi(\check{e}) = -\phi(e)$ for each edge. Thus, $\sum_{e \in E\,:\, e^- = y} \phi(e) = i_0 \cdot \mathbf{1}_x(y)$ for each $y \in X$, and

$$\sum_{y \in A} \sum_{e \in E\,:\, e^- = y} \phi(e) = i_0.$$

If $e \in E_A$ then both e and $\check{e}$ appear precisely once in the above sum, and the two of them contribute to the sum by $\phi(\check{e}) + \phi(e) = 0$. Thus, the sum reduces to all those edges e which have only one endpoint in E, that is

$$\sum_{y \in A} \sum_{e \in E\,:\, e^- = y} \phi(e) = \sum_{e \in \partial A} \phi(e). \qquad \square$$

Before using this, we consider a corollary of Theorem 4.51:

4.54 Exercise (Definition). A *subnetwork* $\mathcal{N}' = (X', E', r')$ of $\mathcal{N} = (X, E, r)$ is a connected graph with vertex set X' and symmetric edge set $E' \subset E_{X'}$ such that $r'(e) \geq r(e)$ for each $e \in E'$. (That is, $a'(x,y) \leq a(x,y)$ for all $x, y \in X'$.) We call it an *induced subnetwork*, if $r'(e) = r(e)$ for each $e \in E'$.

Use the flow criterion to show that transience of a subnetwork $\mathcal{N}'$ implies transience of $\mathcal{N}$. □

Thus, if simple random walk on some infinite, connected, locally finite graph is recurrent, then SRW on any subgraph is also recurrent.

The other criteria in Theorem 4.51 are also very useful. The next corollary generalizes Exercise 4.54.

4.55 Corollary. *Let P and Q be the transition matrices of two irreducible, reversible Markov chains on the same state space X, and let D_P and D_Q be the associated Dirichlet norms. Suppose that there is a constant $\varepsilon > 0$ such that*

$$\varepsilon \cdot D_Q(f) \le D_P(f) \quad \text{for each } f \in \ell_0(X).$$

Then transience of (X, Q) implies transience of (X, P).

Proof. The inequality implies that for the capacities associated with P and Q (respectively), one has

$$\mathrm{cap}_P(x) \ge \varepsilon \cdot \mathrm{cap}_Q(x).$$

The statement now follows from criterion (c) of Theorem 4.51. □

We remark here that the notation $D_P(\cdot)$ is slightly ambiguous, since the resistances of the edges in the network depend on the choice of the reversing measure m_P for P: if we multiply the latter by a constant, then the Dirichlet norm divides by that constant. However, none of the properties that we are studying change, and in the above inequality one only has to adjust the value of $\varepsilon > 0$.

Now let $\Gamma = (X, E)$ be an arbitrary connected, locally finite, symmetric graph, and let $d(\cdot, \cdot)$ be the *graph distance,* that is, $d(x, y)$ is the minimal length (number of edges) of a path from x to y. For $k \in \mathbb{N}$, we define the *k-fuzz* $\Gamma^{(k)}$ of Γ as the graph with the same vertex set X, where two points x, y are connected by an edge if and only if $1 \le d(x, y) \le k$.

4.56 Proposition. *Let Γ be a connected graph with uniformly bounded vertex degrees. Then SRW on Γ is recurrent if and only if SRW on the k-fuzz $\Gamma^{(k)}$ is recurrent.*

Proof. Recurrence or transience of SRW on Γ do not depend on the presence of loops (they have no effects on flows). Therefore we may assume that Γ has no loops. Then, as a network, Γ is a subnetwork of $\Gamma^{(k)}$. Thus, recurrence of SRW on $\Gamma^{(k)}$ implies recurrence of SRW on Γ.

Conversely, let $x, y \in X$ with $1 \le d = d(x, y) \le k$. Choose a path $\pi_{x,y} = [x = x_0, x_1, \dots, x_d = y]$ in X. Then for any function f on X,

$$\big(f(y)-f(x)\big)^2 = \Big(\sum_{i=1}^{d} 1\cdot\big(f(x_i)-f(x_{i-1})\big)\Big)^2 \le d\cdot \sum_{e\in E(\pi_{x,y})} 1\cdot\big(f(e^+)-f(e^-)\big)^2$$

by the Cauchy–Schwarz inequality. Here, $E(\pi_{x,y}) = \{[x_0, x_1], \dots, [x_{d-1}, x_d]\}$ stands for the set of edges of X on that path.

We write $\Pi = \{\pi_{x,y} : x, y \in X,\ 1 \le d = d(x, y) \le k\}$. Since the graph X has uniformly bounded vertex degrees, there is a bound $M = M_k$ such that any edge e of X lies on at most M paths in X with length at most k. (Exercise: verify

this fact !) We obtain for $f \in \ell_0(X)$ and the Dirichlet forms associated with SRW on Γ and $\Gamma^{(k)}$ (respectively)

$$\begin{aligned} D_{\Gamma^{(k)}}(f) &= \frac{1}{2} \sum_{(x,y):1\le d(x,y)\le k} \big(f(y)-f(x)\big)^2 \\ &\le \frac{1}{2} \sum_{(x,y):1\le d(x,y)\le k} d(x,y) \sum_{e\in E(\pi_{x,y})} \big(f(e^+)-f(e^-)\big)^2 \\ &= \frac{k}{2} \sum_{\pi\in\Pi} \sum_{e\in E(\pi)} \big(f(e^+)-f(e^-)\big)^2 \\ &= \frac{k}{2} \sum_{e\in E} \sum_{\pi\in\Pi: e\in E(\pi)} \big(f(e^+)-f(e^-)\big)^2 \\ &\le \frac{k}{2} \sum_{e\in E} M\cdot\big(f(e^+)-f(e^-)\big)^2 = kM\cdot D_\Gamma(f). \end{aligned}$$

Thus, we can apply Corollary 4.55 with $\varepsilon = 1/(kM)$, and find that recurrence of SRW on Γ implies recurrence of SRW on $\Gamma^{(k)}$. □

E Random walks on integer lattices

When we speak of $\mathbb{Z}^d$ as a graph, we have in mind the d-dimensional lattice whose vertex set is $\mathbb{Z}^d$, and two points are neighbours if their Euclidean distance is 1, that is, they differ by 1 in precisely one coordinate. In particular, $\mathbb{Z}$ will stand for the graph which is a two-way infinite path. We shall now discuss recurrence of SRW in $\mathbb{Z}^d$.

Simple random walk on $\mathbb{Z}^d$

Since $\mathbb{Z}^d$ is an Abelian group, SRW is the random walk on this group in the sense of (4.18) (written additively) whose law is the equidistribution on the set of integer unit vectors $\{\pm\boldsymbol{e}_1,\dots,\pm\boldsymbol{e}_d\}$.

4.57 Example (Dimension $d = 1$). We first propose different ways for showing that SRW on $\mathbb{Z}$ is recurrent. This is the infinite drunkard's walk of Example 3.5 with $p = q = 1/2$.

4.58 Exercise. (a) Use the flow criterion for proving recurrence of SRW on $\mathbb{Z}$.
(b) Use (1.50) to compute $p^{(2n)}(0,0)$ explicitly: show that

$$p^{(2n)}(0,0) = \frac{1}{2^{2n}}\big|\Pi^{(2n)}(0,0)\big| = \frac{1}{2^{2n}}\binom{2n}{n}.$$

(c) Use generating functions, and in particular (3.6), to verify that $G(0,0|z) = 1/\sqrt{1-z^2}$, and deduce the above formula for $p^{(2n)}(0,0)$ by writing down the power series expansion of that function.

(d) Use Stirling's formula to obtain the asymptotic evaluation

$$p^{(2n)}(0,0) \sim \frac{1}{\sqrt{\pi n}}. \tag{4.59}$$

Deduce recurrence from this formula. □

Before passing to dimension 2, we consider the following variant Q of SRW P on $\mathbb{Z}$:

$$q(k,k\pm 1) = 1/4, \quad q(k,k) = 1/2, \quad \text{and} \quad q(k,l) = 0 \text{ if } |k-l| \geq 2. \tag{4.60}$$

Then $Q = \frac{1}{2}I + \frac{1}{2}P$, and Exercise 1.48 yields

$$G_Q(0,0)(z) = 1/\sqrt{1-z}. \tag{4.61}$$

This can also be computed directly, as in Examples 2.10 and 3.5. Therefore

$$q^{(n)}(0,0) = \frac{1}{4^n}\binom{2n}{n} \sim \frac{1}{\sqrt{\pi n}} \quad \text{as } n \to \infty. \tag{4.62}$$

This return probability can also be determined by combinatorial arguments: of the n steps, a certain number k will go by 1 unit to the left, and then the same number of steps must go by one unit to the left, each of those steps with probability $1/4$. There are $\binom{n}{k} \cdot \binom{n-k}{k}$ distinct possibilities to select those k steps to the right and k steps to the left. The remaining $n-2k$ steps must correspond to loops (where the current position in $\mathbb{Z}$ remains unchanged), each one with probability $1/2 = 2/4$. Thus

$$q^{(n)}(0,0) = \frac{1}{4^n}\sum_{k=0}^{\lfloor n/2\rfloor}\binom{n}{k}\cdot\binom{n-k}{k}\cdot 2^{n-2k}.$$

The resulting identity, namely that the last sum over k equals $\binom{2n}{n}$, is *a priori* not completely obvious.[2]

4.63 Example (Dimension $d = 2$). (a) We first show how one can use the flow criterion for proving recurrence of SRW on $\mathbb{Z}^2$. Recall that all edges have conductance 1. Let ϕ be a flow from $\mathbf{0} = (0,0)$ to ∞ with input $i_0 > 0$. Consider the box $A_n = \{(k,l) \in \mathbb{Z}^2 : |k|, |l| \leq n\}$, where $n \geq 1$. Then Lemma 4.53 and the Cauchy–Schwarz inequality yield

$$i_0 = \sum_{e\in\partial A_n} \phi(e)\cdot 1 \leq |\partial A_n| \sum_{e\in\partial A_n} \phi(e)^2.$$

[2]The author acknowledges an email exchange with Ch. Krattenthaler on this point.

The sets ∂A_n are disjoint subsets of the edge set E. Therefore

$$\langle \phi, \phi \rangle \geq \frac{1}{2} \sum_{n=1}^{\infty} \sum_{e \in \partial A_n} \phi(e)^2 \geq \sum_{n=1}^{\infty} \frac{i_0}{2|\partial A_n|} = \infty,$$

since $|\partial A_n| = 8n + 4$. Thus, every flow from the origin to ∞ with non-zero input has infinite power: the random walk is recurrent.

(b) Another argument is the following. Let two walkers perform the one-dimensional SRW simultaneously and independently, each one with starting point 0. Their joint trajectory, viewed in $\mathbb{Z}^2$, visits only the set of points (k, l) with $k + l$ even. The resulting Markov chain on this state space moves from (k, l) to any of the four points $(k \pm 1, l \pm 1)$ with probability $p_1(k, k \pm 1) \cdot p_1(l, l \pm 1) = 1/4$, where $p_1(\cdot, \cdot)$ now stands for the transition probabilities of SRW on $\mathbb{Z}$. The graph of this "doubled" Markov chain is the one with the dotted edges in Figure 11. It is isomorphic with the lattice $\mathbb{Z}^2$, and the transition probabilities are preserved under this isomorphism. Hence SRW on $\mathbb{Z}^2$ satisfies

$$p^{(2n)}(\mathbf{0}, \mathbf{0}) = \left(\frac{1}{2^{2n}} \binom{2n}{n} \right)^2 \sim \frac{1}{\pi n}. \tag{4.64}$$

Thus $G(\mathbf{0}, \mathbf{0}) = \infty$.

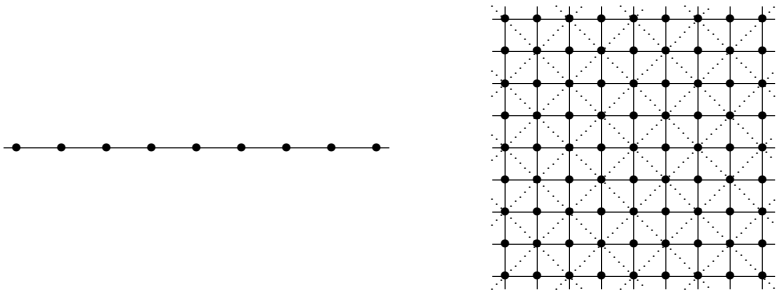

Figure 11. The grids $\mathbb{Z}$ and $\mathbb{Z}^2$.

4.65 Example (Dimension $d = 3$). Consider the random walk on $\mathbb{Z}$ with transition matrix Q given by 4.60. Again, we let three independent walkers start at 0 and move according to Q. Their joint trajectory is a random walk on the Abelian group $\mathbb{Z}^3$ with law μ given by

$$\mu(\pm \boldsymbol{e}_1) = \mu(\pm \boldsymbol{e}_2) = \mu(\pm \boldsymbol{e}_3) = 1/16,$$
$$\mu(\pm \boldsymbol{e}_1 \pm \boldsymbol{e}_2) = \mu(\pm \boldsymbol{e}_1 \pm \boldsymbol{e}_3) = \mu(\pm \boldsymbol{e}_2 \pm \boldsymbol{e}_3) = 1/32,$$
$$\text{and} \quad \mu(\pm \boldsymbol{e}_1 \pm \boldsymbol{e}_2 \pm \boldsymbol{e}_3) = 1/64.$$

We write Q_3 for its transition matrix. It is symmetric (reversible with respect to the counting measure) and transient, since its n-step return probabilities to the origin are

$$q_3^{(n)}(\mathbf{0},\mathbf{0}) = \left(\frac{1}{4^n}\binom{2n}{n}\right)^3 \sim \frac{1}{(\pi\, n)^{3/2}},$$

whence $G(\mathbf{0},\mathbf{0}) < \infty$ for the associated Green function.

If we think of $\mathbb{Z}^3$ made up by cubes then the edges of the associated network plus corresponding resistances are as follows: each side of any cube has resistance 16, each diagonal of each face (square) of any cube has resistance 32, and each diagonal of any cube has resistance 64. In particular, the corresponding network is a subnetwork of the one of SRW on the 3-fuzz of the lattice, if we put resistance 64 on all edges of the latter. Thus, SRW on the 3-fuzz is transient by Exercise 4.54, and Proposition 4.56 implies that also SRW on $\mathbb{Z}^3$ is transient.

4.66 Example (Dimension $d > 3$). Since $\mathbb{Z}^d$ for $d > 3$ contains $\mathbb{Z}^3$ as a subgraph, transience of SRW on $\mathbb{Z}^d$ now follows from Exercise 4.54.

We remark that these are just a few among many different ways for showing recurrence of SRW on $\mathbb{Z}$ and $\mathbb{Z}^2$ and transience of SRW on $\mathbb{Z}^d$ for $d \geq 3$. The most typical and powerful method is to use characteristic functions (Fourier transform), see the historical paper of PÓLYA [47].

General random walks on the group $\mathbb{Z}^d$ are classical sums of i.i.d. $\mathbb{Z}$-valued random variables $Z_n = Y_1 + \cdots + Y_n$, where the distribution of the Y_k is a probability measure μ on $\mathbb{Z}^d$ and the starting point is $\mathbf{0}$. We now want to state criteria for recurrence/transience without necessarily assuming reversibility (symmetry of μ). First, we introduce the (*absolute*) *moment of order* k associated with μ:

$$|\mu|_k = \sum_{\boldsymbol{x}\in\mathbb{Z}^d} |\boldsymbol{x}|^k\, \mu(\boldsymbol{x}).$$

($|\boldsymbol{x}|$ is the Euclidean length of the vector $\boldsymbol{x}$.) If $|\mu|_1$ is finite, the *mean vector*

$$\bar{\mu} = \sum_{\boldsymbol{x}\in\mathbb{Z}^d} \boldsymbol{x}\, \mu(\boldsymbol{x})$$

describes the *average displacement* in a single step of the random walk with law μ.

4.67 Theorem. *In arbitrary dimension d, if $|\mu|_1 < \infty$ and $\bar{\mu} \neq \mathbf{0}$, then the random walk with law μ is transient.*

Proof. We use the *strong law of large numbers* (in the multidimensional version).

Let $(Y_n)_{n\geq 1}$ be a sequence of independent $\mathbb{R}^d$-valued random variables with common distribution μ. If $|\mu|_1 < \infty$ then

$$\lim_{n\to\infty} \tfrac{1}{n}(Y_1 + \cdots + Y_n) = \bar{\mu}$$

with probability 1.

In our particular case, given that $\bar{\mu} \neq \mathbf{0}$, the set of trajectories

$$A = \left\{\omega \in \Omega : \text{there is } n_0 \text{ such that } \left|\tfrac{1}{n} Z_n(\omega) - \bar{\mu}\right| < |\bar{\mu}| \text{ for all } n \geq n_0\right\}$$

contains the event$[\lim_n Z_n/n = \bar{\mu}]$ and has probability 1. (Standard exercise: verify that the latter event as well as A belong to the σ-algebra generated by the cylinder sets.) But for every $\omega \in A$, one may have $Z_n(\omega) = \mathbf{0}$ only for finitely many n. In other words, with probability 1, the random walk (Z_n) returns to the origin no more than finitely many times. Therefore $H(\mathbf{0}, \mathbf{0}) = 0$, and we have transience by Theorem 3.2. □

In the one-dimensional case, the most general recurrence criterion is due to Chung and Ornstein [12].

4.68 Theorem. *Let μ be a probability distribution on $\mathbb{Z}$ with $|\mu|_1 < \infty$ and $\bar{\mu} = 0$. Then every state of the random walk with law μ is recurrent.*

In other words, $\mathbb{Z}$ decomposes into essential classes, on each of which the random walk with law μ is recurrent.

4.69 Exercise. Show that since $\bar{\mu} = 0$, the number of those essential classes is finite except when μ is the point mass in 0. □

For the proof of Theorem 4.68, we need the following auxiliary lemma.

4.70 Lemma. *Let (X, P) be an arbitrary Markov chain. For all $x, y \in X$ and $N \in \mathbb{N}$,*

$$\sum_{n=0}^{N} p^{(n)}(x, y) \leq U(x, y) \sum_{n=0}^{N} p^{(n)}(y, y).$$

Proof.

$$\begin{aligned}\sum_{n=0}^{N} p^{(n)}(x, y) &= \sum_{n=0}^{N} \sum_{k=0}^{n} u^{(n-k)}(x, y)\, p^{(k)}(y, y) \\ &= \sum_{k=0}^{N} p^{(k)}(y, y) \sum_{m=0}^{N-k} u^{(m)}(x, y).\end{aligned}$$

□

Proof of Theorem 4.68. We use once more the law of large numbers, but this time the weak version is sufficient.

Let $(Y_n)_{n\geq 1}$ be a sequence of independent real valued random variables with common distribution μ. If $|\mu|_1 < \infty$ then

$$\lim_{n\to\infty} \mathsf{Pr}\Big[\,\big|\tfrac{1}{n}(Y_1 + \cdots + Y_n) - \bar{\mu}\big| > \varepsilon\Big] = 0$$

for every $\varepsilon > 0$.

Since in our case $\bar{\mu} = 0$, this means in terms of the transition probabilities that

$$\lim_{n\to\infty} \alpha_n = 1, \quad \text{where } \alpha_n = \sum_{k\in\mathbb{Z}:|k|\leq n\varepsilon} p^{(n)}(0,k).$$

Let $M, N \in \mathbb{N}$ and $\varepsilon = 1/M$. (We also assume that $N\varepsilon \in \mathbb{N}$.) Then, by Lemma 4.70,

$$\sum_{n=0}^{N} p^{(n)}(0,k) \leq \sum_{n=0}^{N} p^{(n)}(k,k) = \sum_{n=0}^{N} p^{(n)}(0,0) \quad \text{for all } k \in \mathbb{Z}.$$

Therefore we also have

$$\begin{aligned}\sum_{n=0}^{N} p^{(n)}(0,0) &\geq \frac{1}{2N\varepsilon+1} \sum_{k:|k|\leq N\varepsilon} \sum_{n=0}^{N} p^{(n)}(0,k) \\ &= \frac{1}{2N\varepsilon+1} \sum_{n=0}^{N} \sum_{k:|k|\leq N\varepsilon} p^{(n)}(0,k) \\ &\geq \frac{1}{2N\varepsilon+1} \sum_{n=0}^{N} \alpha_n,\end{aligned}$$

where in the last step we have replaced $N\varepsilon$ with $n\varepsilon$. Since $\alpha_n \to 1$,

$$\lim_{N\to\infty} \frac{1}{2N\varepsilon+1} \sum_{n=0}^{N} \alpha_n = \lim_{N\to\infty} \frac{N}{2N\varepsilon+1} \frac{1}{N} \sum_{n=0}^{N} \alpha_n = \frac{1}{2\varepsilon} = \frac{M}{2}.$$

We infer that $G(0,0) \geq M/2$ for every $M \in \mathbb{N}$. □

Combining the last two theorems, we obtain the following.

4.71 Corollary. *Let μ be a probability measure on $\mathbb{Z}$ with finite first moment and $\mu(0) < 1$. Then the random walk with law μ is recurrent if and only if $\bar{\mu} = 0$.*

Compare with Example 3.5 (infinite drunkard's walk): in this case, $\mu = p \cdot \delta_1 + q \cdot \delta_1$, and $\bar{\mu} = p - q$; as we know, one has recurrence if and only if $p = q$ $(= 1/2)$.

What happens when $|\mu|_1 = \infty$? We present a class of examples.

4.72 Proposition. *Let μ be a* symmetric *probability distribution on $\mathbb{Z}$ such that for some real exponent $\alpha > 0$,*

$$0 < \lim_{k \to \infty} k^\alpha \mu(k) < \infty.$$

Then the random walk with law μ is recurrent if $\alpha \geq 1$, and transient if $\alpha < 1$.

Observe that for $\alpha > 1$ recurrence follows from Theorem 4.68.

In dimension 2, there is an analogue of Corollary 4.71 in presence of finite *second* moment.

4.73 Theorem. *Let μ be a probability distribution on $\mathbb{Z}^2$ with $|\mu|_2 < \infty$. Then the random walk with law μ is recurrent if and only if $\bar{\mu} = \mathbf{0}$.*

(The "only if" follows from Theorem 4.67.)

Finally, the behaviour of the simple random walk on $\mathbb{Z}^3$ generalizes as follows, without any moment condition.

4.74 Theorem. *Let μ be a probability distribution on $\mathbb{Z}^d$ whose support generates a subgroup that is at least 3-dimensional. Then each state of the random walk with law μ is transient.*

The subgroup generated by $\mathrm{supp}(\mu)$ consists of all elements of $\mathbb{Z}^d$ which can be written as a sum of finitely many elements of $-\mathrm{supp}(\mu) \cup \mathrm{supp}(\mu)$. It is known from the structure theory of Abelian groups that any subgroup of $\mathbb{Z}^d$ is isomorphic with $\mathbb{Z}^{d'}$ for some $d' \leq d$. The number d' is the dimension (often called the *rank*) of the subgroup to which the theorem refers. In particular, every irreducible random walk on $\mathbb{Z}^d$, $d \geq 3$, is transient.

We omit the proofs of Proposition 4.72 and of Theorems 4.73 and 4.74. They rely on the use of characteristic functions, that is, harmonic analysis. For a detailed treatment, see the monograph by Spitzer [Sp, §8]. A very nice and accessible account of recurrence of random walks on $\mathbb{Z}^d$ is given by Lesigne [Le].

Chapter 5
Models of population evolution

In this chapter, we shall study three classes of processes that can be interpreted as theoretical models for the random evolution of populations. While in this book we maintain discrete time and discrete state space, the second and the third of those models will go slightly beyond ordinary Markov chains (although they may of course be interpreted as Markov chains on more complicated state spaces).

A Birth-and-death Markov chains

In this section we consider Markov chains whose state space is $X = \{0, 1, \dots, N\}$ with $N \in \mathbb{N}$, or $X = \mathbb{N}_0$ (in which case we write $N = \infty$). We assume that there are non-negative parameters p_k, q_k and r_k ($0 \le k \le N$) with $q_0 = 0$ and (if $N < \infty$) $p_N = 0$ that satisfy $p_k + q_k + r_k = 1$, such that whenever $k-1, k, k+1 \in X$ (respectively) one has

$$p(k, k+1) = p_k, \quad p(k, k-1) = q_k \quad \text{and} \quad p(k, k) = r_k,$$

while $p(k, l) = 0$ if $|k - l| > 1$. See Figure 12. The finite and infinite drunkard's walk and the Ehrenfest urn model are all of this type.

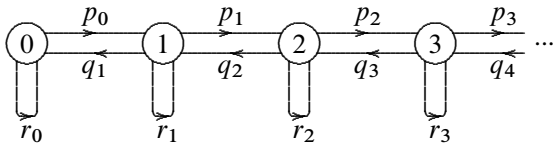

Figure 12

Such a chain is called *random walk on* $\mathbb{N}_0$, or *on* $\{0, 1, \dots, N\}$, respectively. It is also called a *birth-and-death Markov chain.* This latter name comes from the following interpretation. Consider a population which can have any number k of members, where $k \in \mathbb{N}_0$. These numbers are the states of the process, and we consider the evolution of the size of the population in discrete time steps $n = 0, 1, 2, \dots$ (e.g., year by year), so that $Z_n = k$ means that at time n the population has k members. If at some time it has k members then in the next step it can increase by one (with probability p_k – the birth rate), maintain the same size (with probability r_k) or, if $k > 0$, decrease by one individual (with probability q_k – the death rate). The Markov chain describes the random evolution of the population

size. In general, the state space for this model is $\mathbb{N}_0$, but when the population size cannot exceed N, one takes $X = \{0, 1, \ldots, N\}$.

Finite birth-and-death chains

We first consider the case when $N < \infty$, and limit ourselves to the following three cases.

(a) Reflecting boundaries: $p_k > 0$ for each k with $0 \le k < N$ and $q_k > 0$ for every k with $0 < k \le N$.

(b) Absorbing boundaries: $p_k,\ q_k > 0$ for $0 < k < N$, while $p_0 = q_N = 0$ and $r_0 = r_N = 1$.

(c) Mixed case (state 0 is absorbing and state N is reflecting): $p_k,\ q_k > 0$ for $0 < k < N$, and also $q_N > 0$, while $p_0 = 0$ and $r_0 = 1$.

In the reflecting case, we do not necessarily require that $r_0 = r_N = 0$; the main point is that the Markov chain is irreducible. In case (b), the states 0 and N are absorbing, and the irreducible class $\{1, \ldots, N-1\}$ is non-essential. In case (c), only the state 0 is absorbing, while $\{1, \ldots, N\}$ is a non-essential irreducible class.

For the birth-and-death model, the last case is the most natural one: if in some generation the population dies out, then it remains extinct. Starting with $k \ge 1$ individuals, the quantity $F(k, 0)$ is then the probability of extinction, while $1 - F(k, 0)$ is the survival probability.

We now want to compute the generating functions $F(k, m|z)$. We start with the case when $k < m$ and state 0 is reflecting. As in the specific case of Example 1.46, Theorem 1.38 (d) leads to a linear recursion in k:

$$\begin{aligned} F(0, m|z) &= r_0 z\, F(0, m|z) + p_0 z\, F(1, m|z) \quad \text{and} \\ F(k, m|z) &= q_k z\, F(k-1, m|z) + r_k z\, F(k, m|z) + p_k z\, F(k+1, m|z) \end{aligned}$$

for $k = 1, \ldots, m-1$. We write for $z \neq 0$

$$F(k, m|z) = Q_k(1/z)\, F(0, m|z), \quad k = 0, \ldots, m.$$

With the change of variable $t = 1/z$, we find that the $Q_k(t)$ are polynomials with degree k which satisfy the recursion

$$\begin{aligned} Q_0(t) = 1, \quad p_0\, Q_1(t) = t - r_0, \quad \text{and} \\ p_k\, Q_{k+1}(t) = (t - r_k) Q_k(t) - q_k\, Q_{k-1}(t), \quad k \ge 1. \end{aligned} \tag{5.1}$$

It does not matter here whether state N is absorbing or reflecting.

Analogously, if $k > m$ and state N is reflecting then we write for $z \neq 0$

$$F(k, m|z) = Q_k^*(1/z)\, F(N, m|z), \quad k = m, \ldots, N.$$

Again with $t = 1/z$, we obtain the downward recursion

$$\begin{aligned} Q^*_N(t) &= 1, \quad q_N\, Q^*_{N-1}(t) = t - r_N, \quad \text{and} \\ q_k\, Q^*_{k-1}(t) &= (t - r_k) Q^*_k(t) - p_k\, Q^*_{k+1}(t), \quad k \le N-1. \end{aligned} \tag{5.2}$$

The last equation is the same as in (5.1), but with different initial values and working downwards instead of upwards.

Now recall that $F(m, m|z) = 1$. We get the following.

5.3 Lemma. *If the state* 0 *is reflecting and* $0 \le k \le m$, *then*

$$F(k, m|z) = \frac{Q_k(1/z)}{Q_m(1/z)}.$$

If the state N *is reflecting and* $m \le k \le N$, *then*

$$F(k, m|z) = \frac{Q^*_k(1/z)}{Q^*_m(1/z)}.$$

Let us now consider the case when state 0 is absorbing. Again, we want to compute $F(k, m|z)$ for $1 \le k \le m \le N$. Once more, this does not depend on whether the state N is absorbing or reflecting, or even $N = \infty$. This time, we write for $z \ne 0$

$$F(k, m|z) = R_k(1/z)\, F(1, m|z), \quad k = 1, \dots, m.$$

With $t = 1/z$, the polynomials $R_k(z)$ have degree $k - 1$ and satisfy the recursion

$$\begin{aligned} R_1(t) &= 1, \quad p_1\, R_2(t) = t - r_1, \quad \text{and} \\ p_k\, R_{k+1}(t) &= (t - r_k) R_k(t) - q_k\, R_{k-1}(t), \quad k \ge 1, \end{aligned} \tag{5.4}$$

which is basically the same as (5.1) with different initial terms. We get the following.

5.5 Lemma. *If the state* 0 *is absorbing and* $1 \le k \le m$, *then*

$$F(k, m|z) = \frac{R_k(1/z)}{R_m(1/z)}.$$

We omit the analogous case when state N is absorbing and $m \le k < N$. From those formulas, most of the interesting functions and quantities for our Markov chain can be derived.

5.6 Example. We consider simple random walk on $\{0, \dots, N\}$ with state 0 reflecting and state N absorbing. That is, $r_k = 0$ for all $k < N$, $p_0 = r_N = 1$ and $p_k = q_k = 1/2$ for $k = 1, \dots, N-1$. We want to compute $F(k, N|z)$ and $F(k, 0|z)$ for $k < N$.

The recursion (5.1) becomes

$$Q_0(t) = 1, \quad Q_1(t) = t, \quad \text{and} \quad Q_{k+1}(t) = 2t\, Q_k(t) - Q_{k-1}(t) \quad \text{for } k \geq 1.$$

This is the well known formula for the *Chebyshev polynomials of the first kind,* that is, the polynomials that are defined by

$$Q_k(\cos\varphi) = \cos k\varphi.$$

For real $z \geq 1$, we can set $1/z = \cos\varphi$. Thus, $\varphi \mapsto z$ is strictly increasing from $[0,\, \pi/2)$ to $[1,\, \infty)$, and

$$F\left(k, N \middle| \frac{1}{\cos\varphi}\right) = \frac{\cos k\varphi}{\cos N\varphi}.$$

We remark that we can determine the (common) radius of convergence s of the power series $F(k, N|z)$ for $k < N$, which are rational functions. We know that $\mathsf{s} > 1$ and that it is the smallest positive pole of $F(k, N|z)$. Thus, $\mathsf{s} = 1/\cos\frac{\pi}{2N}$.

$F(k, 0|z)$ is the same as the function $F(N-k, N|z)$ in the reversed situation where state 0 is absorbing and state N is reflecting. Therefore, in our example,

$$F(k, 0|z) = \frac{R_{N-k-1}(1/z)}{R_{N-1}(1/z)},$$

where

$$R_0(t) = 1, \quad R_1(t) = 2t, \quad \text{and} \quad R_{k+1}(t) = 2t\, R_k(t) - R_{k-1}(t) \quad \text{for } k \geq 1.$$

We recognize these as the *Chebyshev polynomials of the second kind,* which are defined by

$$R_k(\cos\varphi) = \frac{\sin(k+1)\varphi}{\sin\varphi}.$$

We conclude that for our random walk with 0 reflecting and N absorbing,

$$F\left(k, 0 \middle| \frac{1}{\cos\varphi}\right) = \frac{\sin(N-k)\varphi}{\sin N\varphi},$$

and that the (common) radius of convergence of the power series $F(k, 0|z)$ $(1 \leq k < N)$ is $\mathsf{s}' = 1/\cos\frac{\pi}{N}$. We can also compute $G(0, 0|z) = 1\big/\big(1 - z\, F(1, 0|z)\big)$ in these terms:

$$G\left(0, 0 \middle| \frac{1}{\cos\varphi}\right) = \frac{\tan N\varphi}{\tan\varphi}, \quad \text{and} \quad G(0, 0|z) = \frac{(1/z)R_{N-1}(1/z)}{Q_N(1/z)}.$$

In particular, $G(0, 0|z)$ has radius of convergence $\mathsf{r} = \mathsf{s} = 1/\cos\frac{\pi}{2N}$. The values at $z = 1$ of our functions are obtained by letting $\varphi \to 0$ from the right. For example, the expected number of visits in the starting point 0 before absorption in the state N is $G(0, 0|1) = N$.

We return to general finite birth-and-death chains. If both 0 and N are reflecting, then the chain is irreducible. It is also reversible. Indeed, reversibility with respect to a measure m on $\{0, 1, \dots, N\}$ just means that $\mathsf{m}(k)\, p_k = \mathsf{m}(k+1)\, q_{k+1}$ for all $k < N$. Up to the choice of the value of $\mathsf{m}(0)$, this recursion has just one solution. We set

$$\mathsf{m}(0) = 1 \quad \text{and} \quad \mathsf{m}(k) = \frac{p_0 \cdots p_{k-1}}{q_1 \cdots q_k}, \quad k \geq 1. \tag{5.7}$$

Then the unique stationary probability measure is $\mathsf{m}_0(k) = \mathsf{m}(k) \big/ \sum_{j=0}^{N} \mathsf{m}(j)$. In particular, in view of Theorem 3.19, the expected return time to the origin is

$$\mathsf{E}_0(\boldsymbol{t}^0) = \sum_{j=0}^{N} \mathsf{m}(j).$$

We now want to compute the expected time to reach 0, starting from any state $k > 0$. This is $\mathsf{E}_k(\boldsymbol{s}^0) = F'(k, 0|1)$. However, we shall not use the formula of Lemma 5.3 for this computation. Since $k - 1$ is a cut point between k and 0, we have by Exercise 1.45 that

$$\mathsf{E}_k(\boldsymbol{s}^0) = \mathsf{E}_k(\boldsymbol{s}^{k-1}) + \mathsf{E}_{k-1}(\boldsymbol{s}^0) = \cdots = \sum_{i=0}^{k-1} \mathsf{E}_{i+1}(\boldsymbol{s}^i).$$

Now $U(0, 0|z) = r_0 z + p_0 z\, F(1, 0|z)$. We derive with respect to z and take into account that both $r_0 + p_0 = 1$ and $F(1, 0|1) = 1$:

$$\mathsf{E}_1(\boldsymbol{s}^0) = F'(1, 0|1) = \frac{U'(0, 0|1) - 1}{p_0} = \frac{\mathsf{E}_0(\boldsymbol{t}^0) - 1}{p_0} = \sum_{j=1}^{N} \frac{p_1 \cdots p_{j-1}}{q_1 \cdots q_j}.$$

We note that this number does not depend on p_0. Indeed, the stopping time $\boldsymbol{s}^0$ depends only on what happens until the first visit in state 0 and not on the "outgoing" probabilities at 0. In particular, if we consider the mixed case of birth-and-death chain where the state 0 is absorbing and the state N reflecting, then $F(1, 0|z)$ and $F'(1, 0|1)$ are the same as above. We can use the same argument in order to compute $F'(i+1, i|1)$, by making the state i absorbing and considering our chain on the set $\{i, \dots, N\}$. Therefore

$$\mathsf{E}_{i+1}(\boldsymbol{s}^i) = \sum_{j=i+1}^{N} \frac{p_{i+1} \cdots p_{j-1}}{q_{i+1} \cdots q_j}.$$

We subsume our computations, which in particular give the expected time until extinction for the "true" birth-and-death chain of the mixed model (c).

5.8 Proposition. *Consider a birth-and-death chain on* $\{0, \dots, N\}$*, where* N *is reflecting. Starting at* $k > 0$*, the expected time until first reaching state* 0 *is*

$$\mathsf{E}_k(s^0) = \sum_{i=0}^{k-1} \sum_{j=i+1}^{N} \frac{p_{i+1} \cdots p_{j-1}}{q_{i+1} \cdots q_j}.$$

Infinite birth-and-death chains

We now turn our attention to birth-and-death chains on $\mathbb{N}_0$, again limiting ourselves to two natural cases:

(a) The state 0 is reflecting: $p_k > 0$ for every $k \geq 0$ and $q_k > 0$ for every $k \geq 1$.

(b) The state 0 is absorbing: $p_k,\ q_k > 0$ for all $k \geq 1$, while $p_0 = 0$ and $r_0 = 1$.

Most answers to questions regarding case (b) will be contained in what we shall find out about the irreducible case, so we first concentrate on (a). Then the Markov chain is again reversible with respect to the same measure m as in (5.7), this time defined on the whole of $\mathbb{N}_0$. We first address the question of recurrence/transience.

5.9 Theorem. *Suppose that the random walk on* $\mathbb{N}_0$ *is irreducible* (*state* 0 *is reflecting*)*. Set*

$$S = \sum_{m=1}^{\infty} \frac{q_1 \cdots q_m}{p_1 \cdots p_m} \quad \text{and} \quad T = \sum_{m=1}^{\infty} \frac{p_0 \cdots p_{m-1}}{q_1 \cdots q_m}.$$

Then

(i) *the random walk is transient if* $S < \infty$*,*

(ii) *the random walk is null-recurrent if* $S = \infty$ *and* $T = \infty$*, and*

(iii) *the random walk is positive recurrent if* $S = \infty$ *and* $T < \infty$*.*

Proof. We use the flow criterion of Theorem 4.51. Our network has vertex set $\mathbb{N}_0$ and two oppositely oriented edges between $m-1$ and m for each $m \in \mathbb{N}$, plus possibly additional loops $[m, m]$ at some or all m. The loops play no role for the flow criterion, since every flow ϕ must have value 0 on each of them. With m as in (5.7), the conductance of the edge $[m, m+1]$ is

$$a(m, m+1) = p_0 \frac{p_1 \cdots p_m}{q_1 \cdots q_m}; \quad a(0,1) = p_0.$$

There is only one flow with input $i_0 = 1$ from 0 to ∞, namely the one where $\phi([m-1, m]) = 1$ and $\phi([m, m-1]) = -1$ along the two oppositely oriented

edges between $m-1$ and m. Its power is

$$\langle \phi, \phi \rangle = \sum_{m=1}^{\infty} \frac{1}{a(m-1,m)} = \frac{S+1}{p_0}.$$

Thus, there is a flow from 0 to ∞ with finite power if and only if $S < \infty$: recurrence holds if and only if $S = \infty$. In that case, we have positive recurrence if and only if the total mass of the invariant measure is finite. The latter holds precisely when $T < \infty$. □

5.10 Examples. (a) The simplest example to illustrate the last theorem is the one-sided drunkard's walk, whose absorbing variant has been considered in Example 2.10. Here we consider the reflecting version, where $p_0 = 1$, $p_k = p$, $q_k = q = 1-p$ for $k \geq 1$, and $r_k = 0$ for all k.

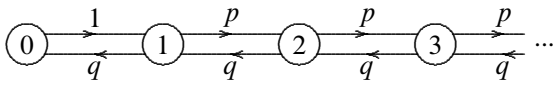

Figure 13

Theorem 5.9 implies that this random walk is transient when $p > 1/2$, null recurrent when $p = 1/2$, and positive recurrent when $p < 1/2$.

More generally, suppose that we have $p_0 = 1$, $r_k = 0$ for all $k \geq 0$, and $p_k \geq 1/2 + \varepsilon$ for all $k \geq 1$, where $\varepsilon > 0$. Then it is again straightforward that the random walk is transient, since $S < \infty$. If on the other hand $p_k \leq 1/2$ for all $k \geq 1$, then the random walk is recurrent, and positive recurrent if in addition $p_k \leq 1/2 - \varepsilon$ for all k.

(b) In view of the last example, we now ask if we can still have recurrence when all "outgoing" probabilities satisfy $p_k > 1/2$. We consider the example where $p_0 = 1$, $r_k = 0$ for all $k \geq 0$, and

$$p_k = \frac{1}{2} + \frac{c}{k}, \quad \text{where } c > 0. \tag{5.11}$$

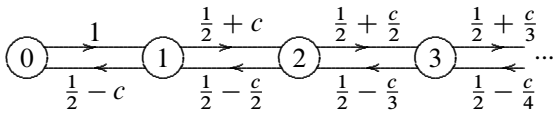

Figure 14

We start by observing that the inequality $\log(1+t) - \log(1-t) \geq 2t$ holds for all real $t \in [0, 1)$. Therefore

$$\log \frac{p_k}{q_k} = \log\left(1 + \frac{2c}{k}\right) - \log\left(1 - \frac{2c}{k}\right) \geq \frac{4c}{k},$$

whence

$$\log \frac{p_1 \cdots p_m}{q_1 \cdots q_m} \geq 4c \sum_{k=1}^{m} \frac{1}{k} \geq 4c \log m .$$

We get

$$S \leq \sum_{m=1}^{\infty} \frac{1}{m^{4c}},$$

and see that $S < \infty$ if $c > 1/4$.

On the other hand, when $c \leq 1/4$ then $q_k / p_k \geq (2k-1)/(2k+1)$, and

$$\frac{q_1 \cdots q_m}{p_1 \cdots p_m} \geq \frac{1}{2m+1},$$

so that $S = \infty$.

We have shown that the random walk of (5.11) is recurrent if $c \leq 1/4$, and transient if $c > 1/4$. Recurrence must be null recurrence, since clearly $T = \infty$.

Our next computations will lead to another, direct and more elementary proof of Theorem 5.9, that does not involve the flow criterion.

Theorem 1.38 (d) and Proposition 1.43 (b) imply that for $k \geq 1$

$$F(k, k-1|z) = q_k z + r_k z \, F(k, k-1|z) + p_k z \, F(k+1, k-1|z) \quad \text{and}$$
$$F(k+1, k-1|z) = F(k+1, k|z) \, F(k, k-1|z).$$

Therefore

$$F(k, k-1|z) = \frac{q_k z}{1 - r_k z - p_k z \, F(k+1, k|z)}. \tag{5.12}$$

This will allow us to express $F(k, k-1|z)$, as well as $U(0,0|z)$ and $G(0,0|z)$ as a *continued fraction.* First, we define a new stopping time:

$$s_i^k = \min\{n \geq 0 \mid \ Z_n = k, \ |Z_m - k| \leq i \ \text{for all } m \leq n\}.$$

Setting

$$f_i^{(n)}(l,k) = \mathsf{Pr}_l[s_i^k = n] \quad \text{and} \quad F_i(l,k|z) = \sum_{n=0}^{\infty} f_i^{(n)}(l,k) \, z^n,$$

the number $F_i(l,k) = F_i(l,k|1)$ is the probability, starting at l, to reach k before leaving the interval $[k-i, \ k+i]$.

5.13 Lemma. *If $|z| \leq \mathsf{r}$, where $\mathsf{r} = \mathsf{r}(P)$ is the radius of convergence of $G(l,k|z)$, then*

$$\lim_{i \to \infty} F_i(l,k|z) = F(l,k|z).$$

Proof. It is clear that the radius of convergence of the power series $F_i(l,k|z)$ is at least $\mathsf{s}(l,k) \geq \mathsf{r}$. We know from Lemma 3.66 that $F(l,k|\mathsf{r}) < \infty$.

For fixed n, the sequence $\big(f_i^{(n)}(l,k)\big)$ is monotone increasing in i, with limit $f^{(n)}(l,k)$ as $i \to \infty$. In particular, $f_i^{(n)}(l,k)\,|z^n| \leq f^{(n)}(l,k)\,\mathsf{r}^n$. Therefore we can use dominated convergence (the integral being summation over n in our power series) to conclude. □

5.14 Exercise. Prove that for $i \geq 1$,

$$F_i(k+1,k-1|z) = F_{i-1}(k+1,k|z)\,F_i(k,k-1|z). \qquad \square$$

In analogy with (5.12), we can compute the functions $F_i(k,k-1|z)$ recursively.

$$\begin{aligned} &F_0(k,k-1|z) = 0, \quad \text{and} \\ &F_i(k,k-1|z) = \frac{q_k z}{1 - r_k z - p_k z\,F_{i-1}(k+1,k|z)} \quad \text{for } i \geq 1. \end{aligned} \tag{5.15}$$

Note that the denominator of the last fraction cannot have any zero in the domain of convergence of the power series $F_i(k,k-1|z)$. We use (5.15) in order to compute the probability $F(k+1,k)$.

5.16 Theorem. *We have*

$$F(k+1,k) = 1 - \frac{1}{1+S(k)}, \quad \textit{where } S(k) = \sum_{m=1}^{\infty} \frac{q_{k+1}\cdots q_{k+m}}{p_{k+1}\cdots p_{k+m}}.$$

Proof. We prove by induction on i that for all $k, i \in \mathbb{N}_0$,

$$F_i(k+1,k) = 1 - \frac{1}{1+S_i(k)}, \quad \text{where } S_i(k) = \sum_{m=1}^{i} \frac{q_{k+1}\cdots q_{k+m}}{p_{k+1}\cdots p_{k+m}}.$$

Since $S_0(k) = 0$, the statement is true for $i = 0$. Suppose that it is true for $i-1$ (for every $k \geq 0$). Then by (5.15)

$$\begin{aligned} F_i(k+1,k) &= \frac{q_{k+1}}{1 - r_{k+1} - p_{k+1}\,F_{i-1}(k+2,k+1)} \\ &= 1 - \frac{p_{k+1}\big(1 - F_{i-1}(k+2,k+1)\big)}{p_{k+1}\big(1 - F_{i-1}(k+2,k+1)\big) + q_{k+1}} \\ &= 1 - \frac{1}{1 + \dfrac{q_{k+1}}{p_{k+1}}\dfrac{1}{1 - F_{i-1}(k+2,k+1)}} \\ &= 1 - \frac{1}{1 + \dfrac{q_{k+1}}{p_{k+1}}\big(1 + S_{i-1}(k+1)\big)}, \end{aligned}$$

which implies the proposed formula for i (and all k).

Letting $i \to \infty$, the theorem now follows from Lemma 5.13. □

We deduce that

$$F(l,k) = F(l,l-1)\cdots F(k+1,k) = \prod_{j=k}^{l-1} \frac{S(j)}{1+S(j)}, \quad \text{if } l > k, \tag{5.17}$$

while we always have $F(l,k) = 1$ when $l < k$, since the Markov chain must exit the finite set $\{0,\dots,k-1\}$ with probability 1. We can also compute

$$U(k,k) = p_k\, F(k+1,k) + q_k\, F(k-1,k) + r_k = 1 - \frac{p_k}{1+S(k)}.$$

Since $S = S(0)$ for the number defined in Theorem 5.9, we recover the recurrence criterion from above. Furthermore, we obtain formulas for the Green function at $z = 1$:

5.18 Corollary. *When state* 0 *is reflecting, then in the transient case,*

$$G(l,k) = \begin{cases} \dfrac{1+S(k)}{p_k}, & \text{if } l \le k, \\[2ex] \dfrac{S(k)}{p_k} \displaystyle\prod_{j=k+1}^{l-1} \frac{S(j)}{1+S(j)}, & \text{if } l > k. \end{cases}$$

When state 0 is absorbing, that is, for the "true" birth-and-death chain, the formula of (5.17) gives the probability of extinction $F(k,0)$, when the initial population size is $k \ge 1$.

5.19 Corollary. *For the birth-and-death chain on* $\mathbb{N}_0$ *with absorbing state* 0*, extinction occurs almost surely when* $S = \infty$*, while the survival probability is positive when* $S < \infty$*.*

5.20 Exercise. Suppose that $S = \infty$. In analogy with Proposition 5.8, derive a formula for the expected time $\mathsf{E}_k(s^0)$ until extinction, when the initial population size is $k \ge 1$. □

Let us next return briefly to the computation of the generating functions $F(k,k-1|z)$ and $F_i(k,k-1|z)$ via (5.12) and the finite recursion (5.15), respectively. Precisely in the same way, one finds an analogous finite recursion for computing $F(k,k+1|z)$, namely

$$\begin{aligned} F(0,1|z) &= \frac{p_0 z}{1-r_0 z}, \quad \text{and} \\ F(k,k+1|z) &= \frac{p_k z}{1-r_k z - q_k z\, F(k-1,k|z)}. \end{aligned} \tag{5.21}$$

We summarize.

5.22 Proposition. *For $k \geq 1$, the functions $F(k, k+1|z)$ and $F(k, k-1|z)$ can be expressed as finite and infinite continued fractions, respectively. For $z \in \mathbb{C}$ with $|z| \leq \mathsf{r}(P)$,*

$$F(k,k+1|z) = \cfrac{p_k z}{1 - r_k z - \cfrac{q_k p_{k-1} z^2}{1 - r_{k-1} z - \ddots - \cfrac{q_1 p_0 z^2}{1 - r_0 z}}}, \quad \textit{and}$$

$$F(k,k-1|z) = \cfrac{q_k z}{1 - r_k z - \cfrac{p_k q_{k+1} z^2}{1 - r_{k+1} z - \cfrac{p_{k+1} q_{k+2} z^2}{1 - r_{k+1} z - \ddots}}}$$

The last infinite continued fraction is of course intended as the limit of the finite continued fractions $F_i(k, k-1|z)$ – the *approximants* – that are obtained by stopping after the i-th division, i.e., with $1 - r_{k-1+i}\, z$ as the last denominator.

There are well-known recursion formulas for writing the i-th approximant as a quotient of two polynomials. For example,

$$F_i(1,0|z) = \frac{A_i(z)}{B_i(z)},$$

where

$$A_0(z) = 0, \quad B_0(z) = 1, \quad A_1(z) = q_1 z, \quad B_1(z) = 1 - r_1 z,$$

and, for $i \geq 1$,

$$\begin{aligned} A_{i+1}(z) &= (1 - r_{i+1} z) A_i(z) - p_i q_{i+1} z^2 A_{i-1}(z), \\ B_{i+1}(z) &= (1 - r_{i+1} z) B_i(z) - p_i q_{i+1} z^2 B_{i-1}(z). \end{aligned}$$

To get the analogous formulas for $F_i(k, k-1|z)$ and $F(k, k+1|z)$, one just has to adapt the indices accordingly. This opens the door between birth-and-death Markov chains and the classical theory of analytic continued fractions, which is in turn closely linked with orthogonal polynomials. A few references for that theory are the books by Wall [Wa] and Jones and Thron [J-T], and the memoir by Askey and Ismail [2]. Its application to birth-and-death chains appears, for example, in the work of Good [28], Karlin and McGregor [34] (implicitly), Gerl [24] and Woess [52].

More examples

Birth-and-death Markov chains on $\mathbb{N}_0$ provide a wealth of simple examples. In concluding this section, we consider a few of them.

5.23 Example. It may be of interest to compare some of the features of the infinite drunkard's walk on $\mathbb{Z}$ of Example 3.5 and of its reflecting version on $\mathbb{N}_0$ of Example 5.10 (a) with the same parameters p and $q = 1 - p$, see Figure 13. We shall use the indices $\mathbb{Z}$ and $\mathbb{N}$ to distinguish between the two examples.

First of all, we know that the random walk on $\mathbb{Z}$ is transient when $p \neq 1/2$ and null recurrent when $p = 1/2$, while the walk on $\mathbb{N}_0$ is transient when $p > 1/2$, null recurrent when $p = 1/2$, and positive recurrent when $p < 1/2$.

Next, note that for $l > k \geq 0$, then the generating function $F(l, k|z) = F(1, 0|z)^{l-k}$ is the same for both examples, and

$$F(1, 0|z) = \frac{1}{2pz}\left(1 - \sqrt{1 - 4pqz^2}\right).$$

We have already computed $U_{\mathbb{Z}}(0, 0|z)$ in (3.6), and $U_{\mathbb{N}}(0, 0|z) = zF(1, 0|z)$. We obtain

$$G_{\mathbb{Z}}(0, 0|z) = \frac{1}{\sqrt{1 - 4pqz^2}} \quad \text{and} \quad G_{\mathbb{N}}(0, 0|z) = \frac{2p}{2p - 1 + \sqrt{1 - 4pqz^2}}.$$

We compute the asymptotic behaviour of the $2n$-step return probabilities to the origin, as $n \to \infty$. For the random walk on $\mathbb{Z}$, this can be done as in Example 4.58:

$$p_{\mathbb{Z}}^{(2n)}(0, 0) = p^n q^n \binom{2n}{n} \sim \frac{(4pq)^n}{\sqrt{\pi n}}.$$

For the random walk on $\mathbb{N}_0$, we first consider $p < 1/2$ (positive recurrence). The period is $d = 2$, and $\mathsf{E}_0 t^0 = U'_{\mathbb{N}}(0, 0|1) = 2q/(q - p)$. Therefore, using Exercise 3.52,

$$p_{\mathbb{N}}^{(2n)}(0, 0) \to \frac{4q}{q - p}, \quad \text{if } p < 1/2.$$

If $p = 1/2$, then $G_{\mathbb{N}}(0, 0|z) = G_{\mathbb{Z}}(0, 0|z)$. Indeed, if (S_n) is the random walk on $\mathbb{Z}$, then $\big(|S_n|\big)$ is a Markov chain with the same transition probabilities as the random walk on $\mathbb{N}_0$. It is the factor chain as described in (1.29), where the partition of $\mathbb{Z}$ has the blocks $\{k, -k\}$, $k \in \mathbb{N}_0$. Therefore

$$p_{\mathbb{N}}^{(2n)}(0, 0) = p_{\mathbb{Z}}^{(2n)}(0, 0) \sim \frac{1}{\sqrt{\pi n}}, \quad \text{if } p = 1/2.$$

If $p > 1/2$, then we use the fact that $p_{\mathbb{N}}^{(2n)}(0, 0)$ is the coefficient of z^{2n} in the Taylor series expansion at 0 of $G_{\mathbb{N}}(0, 0|z)$, or (since that function depends only

on z^2) equivalently, the coefficient of z^n in the Taylor series expansion at 0 of the function

$$\widetilde{G}(z) = \frac{2p}{2p - 1 + \sqrt{1 - 4pqz}} = \frac{1}{z-1}\Big((p-q) - \sqrt{1-4pqz}\Big).$$

This function is analytic in the open disk $\{z \in \mathbb{C} : |z| < z_0\}$ where $z_0 = 1/(4pq)$. Standard methods in complex analysis (Darboux' method, based on the Riemann–Lebesgue lemma; see Olver [Ol, p. 310]) yield that the Taylor series coefficients behave like those of the function

$$\widetilde{H}(z) = \frac{1}{z_0 - 1}\Big((p-q) - \sqrt{1-4pqz}\Big).$$

The n-th coefficient ($n \geq 1$) of the latter is

$$-\frac{1}{z_0 - 1}(-4pq)^n \binom{1/2}{n} = \frac{2}{z_0 - 1}\,\frac{(4pq)^n}{n\,2^{2n-2}}\binom{2n-2}{n-1}.$$

Therefore, with a use of Stirling's formula,

$$p_{\mathbb{N}}^{(2n)}(0,0) \sim \frac{8pq}{1-4pq}\,\frac{(4pq)^n}{n\sqrt{\pi n}}, \quad \text{if } \ p > 1/2.$$

The spectral radii are

$$\rho(P_{\mathbb{Z}}) = 2\sqrt{pq} \quad \text{and} \quad \rho(P_{\mathbb{N}}) = \begin{cases} 1, & \text{if } p \leq 1/2, \\ 2\sqrt{pq}, & \text{if } p > 1/2. \end{cases}$$

After these calculations, we turn to issues with a less combinatorial-analytic flavour. We can realize our random walks on $\mathbb{Z}$ and on $\mathbb{N}_0$ on one probability space, so that they can be compared (a *coupling*). For this purpose, start with a probability space $(\Omega, \mathcal{A}, \mathsf{Pr})$ on which one can define a sequence of i.i.d. $\{\pm 1\}$-valued random variables $(Y_n)_{n\geq 1}$ with distribution $\mu = p\delta_1 + q\delta_{-1}$. This probability space may be, for example, the product space $\big(\{-1,1\}, \mu\big)^{\mathbb{N}}$, where $\mathcal{A}$ is the product σ-algebra of the discrete one on $\{-1,1\}$. In this case, Y_n is the n-th projection $\Omega \to \{-1,1\}$. Now define

$$S_n = k_0 + Y_1 + \cdots + Y_n = S_{n-1} + Y_n.$$

This is the infinite drunkard's walk on $\mathbb{Z}$ with $p_{\mathbb{Z}}(k, k+1) = p$ and $p_{\mathbb{Z}}(k+1, k) = q$, starting at $k_0 \in \mathbb{Z}$. Analogously, let $k_0 \in \mathbb{N}_0$ and define

$$Z_0 = k_0 \quad \text{and} \quad Z_n = |Z_{n-1} + Y_n| = \begin{cases} 1, & \text{if } Z_{n-1} = 0, \\ Z_{n-1} + Y_n, & \text{if } Z_{n-1} > 0. \end{cases}$$

This is just the reflecting random walk on $\mathbb{N}_0$ of Figure 13. We now suppose that for both walks, the starting point is $k_0 \geq 0$. It is clear that $Z_n \geq S_n$, and we are interested in their difference. We say that *a reflection occurs at time n,* if $Z_{n-1} = 0$ and $Y_n = -1$. At each reflection, the difference increases by 2 and then remains unchanged until the next reflection. Thus, for arbitrary n, we have $Z_n - S_n = 2R_n$, where R_n is the (random) number of reflections that occur up to (and including) time n.

Now suppose that $p > 1/2$. Then the reflecting walk is transient: with probability 1, it visits 0 only finitely many times, so that there can only be finitely many reflections. That is, R_n remains constant from a (random) index onwards, which can also be expressed by saying that $R_\infty = \lim_n R_n$ is almost surely finite, whence $R_n/\sqrt{n} \to 0$ almost surely. By the law of large numbers, $S_n/n \to p - q$ almost surely, and $\big(S_n - n(p-q)\big)\big/2\sqrt{pqn}$ is asymptotically standard normal $N(0,1)$ by the central limit theorem. Therefore

$$\frac{Z_n}{n} \to p - q \text{ almost surely, and } \frac{Z_n - n(p-q)}{2\sqrt{pqn}} \to N(0,1) \text{ in law, if } p > 1/2.$$

In particular, $p - q = 2p - 1$ is the *linear speed* or *rate of escape,* as the random walk tends to $+\infty$.

If $p = 1/2$ then the law of large numbers tells us that $S_n/n \to 0$ almost surely and $S_n/(2\sqrt{n})$ is asymptotically normal. We know that the sequence $\big(|S_n|\big)$ is a model of (Z_n), i.e., it is a Markov chain with the same transition probabilities as (Z_n). Therefore

$$\frac{Z_n}{n} \to 0 \text{ almost surely, and } \frac{Z_n}{2\sqrt{n}} \to |N|(0,1) \text{ in law, if } p = 1/2,$$

where $|N|(0,1)$ is the distribution of the absolute value of a normal random variable; the density is $f(t) = \sqrt{\frac{2}{\pi}}\, e^{-t^2} \mathbf{1}_{[0,\infty)}(t)$.

Finally, if $p < 1/2$ then for each $\alpha > 0$, the function $f(k) = k^{1/\alpha}$ on $\mathbb{N}_0$ is integrable with respect to the stationary probability measure, which is given by $\mathsf{m}(0) = c = (q-p)/(2q)$ and $\mathsf{m}(k) = c\, p^{k-1}/q^k$ for $k \geq 1$. Therefore the Ergodic Theorem 3.55 implies that

$$\frac{1}{n} \sum_{k=0}^{n-1} Z_n^{1/\alpha} \to \sum_{k=1}^{\infty} k^{1/\alpha}\, \mathsf{m}(k) \quad \text{almost surely.}$$

It follows that

$$\frac{Z_n}{n^\alpha} \to 0 \quad \text{almost surely for every } \alpha > 0.$$

Our next example is concerned with ρ-recurrence.

5.24 Example. We let $p > 1/2$ and consider the following slight modification of the reflecting random walk of the last example and Example 5.10 (a): $p_0 = 1$, $p_1 = q_1 = 1/2$, and $p_k = p$, $q_k = q = 1 - p$ for $k \geq 2$, while $r_k = 0$ for all k.

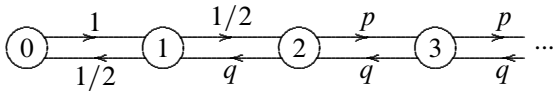

Figure 15

The function $F(2,1|z)$ is the same as in the preceding example,

$$F(2,1|z) = \frac{1}{2pz}\left(1 - \sqrt{1-4pqz^2}\right).$$

By (5.12),

$$F(1,0|z) = \frac{\frac{1}{2}z}{1 - \frac{1}{2}z\,F(2,1|z)},$$

and $U(0,0|z) = z\,F(1,0|z)$. We compute

$$U(0,0|z) = \frac{2pz^2}{4p - 1 + \sqrt{1-4pqz^2}}.$$

We use a basic result from Complex Analysis, *Pringsheim's theorem,* see e.g. Hille [Hi, p. 133]: for a power series with non-negative coefficients, its radius of convergence is a singularity; see e.g. Hille [Hi, p. 133]. Therefore the radius of convergence s of $U(0,0|z)$ is the smallest positive singularity of that function, that is, the zero $\mathsf{s} = 1/\sqrt{4pq}$ of the square root expression. We compute for $p > 1/2$

$$U(0,0|\mathsf{s}) = \frac{1}{(4p-1)(2p-2)} \begin{cases} = 1, & \text{if } p = \frac{3}{4}, \\ < 1, & \text{if } \frac{1}{2} < p < \frac{3}{4}, \\ > 1, & \text{if } \frac{3}{4} < p < 1. \end{cases}$$

Next, we recall Proposition 2.28: the radius of convergence $\mathsf{r} = 1/\rho(P)$ of the Green function is the largest positive real number for which the power series $U(0,0|\mathsf{r})$ converges and has a value ≤ 1. Thus, when $p > \frac{3}{4}$ then r must be the unique solution of the equation $U(0,0|z) = 1$ in the real interval $(0, \mathsf{s})$, which turns out to be $\mathsf{r} = \sqrt{(4p-2)/p}$. We have $U'(0,0|\mathsf{r}) < \infty$ in this case. When $\frac{1}{2} < p \leq \frac{3}{4}$, we conclude that $\mathsf{r} = \mathsf{s}$, and compute $U'(0,0|\mathsf{s}) < \infty$. Therefore we have the following:

- If $\frac{1}{2} < p < \frac{3}{4}$ then $\rho(P) = 2\sqrt{pq}$, and the random walk is ρ-transient.

- If $p = \frac{3}{4}$ then $\rho(P) = \frac{\sqrt{3}}{2}$, and the random walk is ρ-null-recurrent.
- If $\frac{3}{4} < p < 1$ then $\rho(P) = \sqrt{\frac{p}{4p-2}}$, and the random walk is ρ-positive-recurrent.

B The Galton–Watson process

Citing from Norris [No, p. 171], "the original branching process was considered by Galton and Watson in the 1870s while seeking a quantitative explanation for the phenomenon of the disappearance of family names, even in a growing population." Their model is based on the concept that family names are passed on from fathers to sons. (We shall use gender-neutral terminology.)

In general, a Galton–Watson process describes the evolution of successive generations of a "population" under the following assumptions.

- The initial generation number 0 has one member, the *ancestor*.
- The number of children (offspring) of any member of the population (in any generation) is random and follows the *offspring distribution* μ.
- The μ-distributed random variables that represent the number of children of each of the members of the population in all the generations are independent.

Thus, μ is a probability distribution on $\mathbb{N}_0$, the non-negative integers: $\mu(k)$ is the probability to have k children. We exclude the degenerate cases where $\mu = \delta_1$, that is, where every member of the population has precisely one offspring deterministically, or where $\mu = \delta_0$ and there is no offspring at all.

The basic question is: what is the probability that the population will survive forever, and what is the probability of extinction?

To answer this question, we set up a simple Markov chain model: let $N_j^{(n)}$, $j \geq 1$, $n \geq 0$, be a double sequence of independent random variables with identical distribution μ. We write M_n for the random number of members in the n-th generation. The sequence (M_n) is the *Galton–Watson process*. If $M_n = k$ then we can label the members of that generation by $j = 1, \ldots, k$. For each j, the j-th member has $N_j^{(n)}$ children, so that $M_{n+1} = N_1^{(n)} + \cdots + N_k^{(n)}$. We see that

$$M_{n+1} = \sum_{j=1}^{M_n} N_j^{(n)}. \tag{5.25}$$

Since this is a sum of i.i.d. random variables, its distribution depends only on the value of M_n and not on past values. This is the Markov property of $(M_n)_{n\geq 0}$, and

the transition probabilities are

$$p(k,l) = \mathsf{Pr}[M_{n+1} = l \mid M_n = k] = \mathsf{Pr}\Big[\sum_{j=1}^{k} N_j^{(n)} = l\Big] \\ = \mu^{(k)}(l), \quad k,l \in \mathbb{N}_0, \tag{5.26}$$

where $\mu^{(k)}$ is the k-th convolution power of μ, see (4.19) (but note that the group operation is addition here):

$$\mu^{(k)}(l) = \sum_{j=0}^{l} \mu^{(k-1)}(l-j)\,\mu(j), \quad \mu^{(0)}(l) = \delta_0(l).$$

In particular, 0 is an absorbing state for the Markov chain (M_n) on $\mathbb{N}_0$, and the initial state is $M_0 = 1$. We are interested in the probability of absorption, which is nothing but the number

$$F(1,0) = \mathsf{Pr}[\exists\, n \geq 1 : M_n = 0] = \lim_{n\to\infty} \mathsf{Pr}[M_n = 0].$$

The last identity holds because the events $[M_n = 0] = [\exists\, k \leq n : M_k = 0]$ are increasing with limit (union) $[\exists\, n \geq 1 : M_n = 0]$. Let us now consider the probability generating functions of μ and of M_n,

$$f(z) = \sum_{l=0}^{\infty} \mu(l)\, z^l \quad \text{and} \quad g_n(z) = \sum_{k=0}^{\infty} \mathsf{Pr}[M_n = k]\, z^k = \mathsf{E}(z^{M_n}), \tag{5.27}$$

where $0 \leq z \leq 1$. Each of these functions is non-negative and monotone increasing on the interval [0, 1] and has value 1 at $z = 1$. We have $g_n(0) = \mathsf{Pr}[M_n = 0]$. Using (5.26), we now derive a recursion formula for $g_n(z)$. We have $g_0(z) = z$ and $g_1(z) = f(z)$, since the distribution of M_1 is μ. For $n \geq 1$,

$$g_n(z) = \sum_{k,l=0}^{\infty} \mathsf{Pr}[M_n = l \mid M_{n-1} = k]\ \mathsf{Pr}[M_{n-1} = k]\, z^l \\ = \sum_{k=0}^{\infty} \mathsf{Pr}[M_{n-1} = k]\, f_k(z), \quad \text{where } f_k(z) = \sum_{l=0}^{\infty} \mu^{(k)}(l)\, z^l.$$

We have $f_0(z) = 1$ and $f_1(z) = f(z)$. For $k \geq 1$, we can use the product formula for power series and compute

$$f_k(z) = \sum_{l=0}^{\infty} \sum_{j=0}^{l} \big(\mu^{(k-1)}(l-j)\, z^{l-j}\big)\big(\mu(j)\, z^j\big) \\ = \Big(\sum_{m=0}^{\infty} \mu^{(k-1)}(m)\, z^m\Big)\Big(\sum_{j=0}^{\infty} \mu(j)\, z^j\Big) = f_{k-1}(z)\, f(z).$$

Therefore $f_k(z) = f(z)^k$, and we have obtained the following.

5.28 Lemma. *For* $0 \le z \le 1$,

$$g_n(z) = \sum_{k=0}^{\infty} \mathsf{Pr}[M_{n-1} = k]\, f(z)^k$$
$$= g_{n-1}\big(f(z)\big) = \underbrace{f \circ f \circ \ldots \circ f}_{n \text{ times}}(z) = f\big(g_{n-1}(z)\big).$$

In particular, we see that

$$g_1(0) = \mu(0) \quad \text{and} \quad g_n(0) = f\big(g_{n-1}(0)\big). \tag{5.29}$$

Letting $n \to \infty$, continuity of the function $f(z)$ implies that the extinction probability $F(1,0)$ must be a fixed point of f, that is, a point where $f(z) = z$. To understand the location of the fixed point(s) of f in the interval $[0,\ 1]$, note that f is a convex function on that interval with $f(0) = \mu(0)$ and $f(1) = 1$. We do not have $f(z) = z$, since $\mu \ne \delta_1$. Also, unless $f(z) = \mu(0) + \mu(1)z$ with $\mu(1) = 1 - \mu(0)$, we have $f''(z) > 0$ on $(0,\ 1)$. Therefore there are at most two fixed points, and one of them is $z = 1$. When is there a second one? This depends on the slope of the graph of $f(z)$ at $z = 1$, that is, on the left-sided derivative $f'(1-) = \sum_{n=1}^{\infty} n\, \mu(n)$. This is the expected offspring number. (It may be infinite.)

If $f'(1-) \le 1$, then $f(z) > z$ for all $z < 1$, and we must have $F(1,0) = 1$. See Figure 16 a.

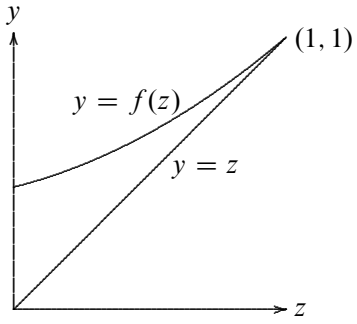

Figure 16 a

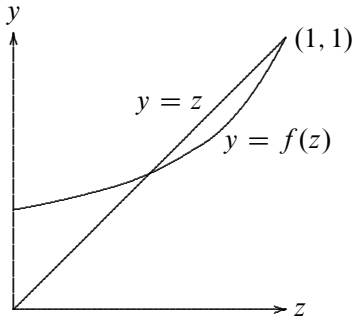

Figure 16 b

If $1 < f'(1-) \le \infty$ then besides $z = 1$, there is a second fixed point $\lambda < 1$, and convexity implies $f'(\lambda) < 1$. See Figure 16 b. Since $f'(z)$ is increasing in z, we get that $f'(z) \le f'(\lambda)$ on $[0, \lambda]$. Therefore λ is an attracting fixed point of f on $[0, \lambda]$, and (5.29) implies that $g_n(0) \to \lambda$.

We subsume the results.

5.30 Theorem. *Let μ be the non-degenerate offspring distribution of a Galton–Watson process and*

$$\bar{\mu} = \sum_{n=1}^{\infty} n\,\mu(n)$$

be the expected offspring number.

If $\bar{\mu} \le 1$, then extinction occurs almost surely.

If $1 < \bar{\mu} \le \infty$ then the extinction probability is the unique non-negative number $\lambda < 1$ such that

$$\sum_{k=0}^{\infty} \mu(k)\,\lambda^k = \lambda,$$

and the probability that the population survives forever is $1 - \lambda > 0$.

If $\bar{\mu} = 1$, the Galton–Watson process is called *critical*, if $\bar{\mu} < 1$, it is called *subcritical*, and if $\bar{\mu} > 1$, the process is called *supercritical*.

5.31 Exercise. Let $\boldsymbol{t}$ be the time until extinction of the Galton–Watson process with offspring distribution μ. Show that $\mathsf{E}(\boldsymbol{t}) < \infty$ if $\bar{\mu} < 1$.

[Hint: use $\mathsf{E}(\boldsymbol{t}) = \sum_n \mathsf{Pr}[\boldsymbol{t} > n]$ and relate $\mathsf{Pr}[\boldsymbol{t} > n]$ with the functions of (5.27).] □

The Galton–Watson process as the basic example of a *branching process* is very well described in the literature. Standard monographs are the ones of Harris [Har] and Athreya and Ney [A-N], but the topic is also presented on different levels in various books on Markov chains and stochastic processes. For example, a nice treatment is given by Lyons with Peres [L-P]. Here, we shall not further develop the detailed study of the behaviour of (M_n).

However, we make one step backwards and have a look beyond counting the number M_n of members in the n-th generation. Instead, we look at the complete information about the *generation tree* of the population. Let $N = \sup\{n : \mu(n) > 0\}$, and write

$$\Sigma = \begin{cases} \{1, \dots, N\}, & \text{if } N < \infty, \\ \mathbb{N}, & \text{if } N = \infty, \end{cases}$$

Any member v of the population will have a certain number k of children, which we denote by $v1, v2, \dots, vk$. If v itself is not the ancestor ϵ of the population, then it is an offspring of some member u of the population, that is, $v = uj$, where $j \in \mathbb{N}$. Thus, we can encode v by a sequence ("word") $j_1 \cdots j_n$ with length $|v| = n \ge 0$ and $j_1, \dots, j_n \in \Sigma$. The set of all such sequences is denoted Σ^*. This includes the *empty* sequence or word ϵ, which stands for the ancestor. If v has the form $v = uj$, then the *predecessor* of v is $v^- = u$. We can draw an edge from u to v in this case. Thus, Σ^* becomes an infinite N-ary tree with root ϵ, where in the case $N = \infty$ this

means that every vertex u has countably many forward neighbours (namely those v with $v^- = u$), while the number of forward neighbours is N when $N < \infty$. For $u, v \in \Sigma^*$, we shall write

$$u \preccurlyeq v, \quad \text{if } v = uj_1 \cdots j_l \text{ with } j_1, \ldots, j_l \in \Sigma \ (l \geq 0).$$

This means that u lies on the shortest path in Σ^* from ϵ to v.

A *full genealogical tree* is a finite or infinite subtree T of Σ^* which contains the root and has the property

$$uk \in \mathsf{T} \Longrightarrow uj \in \mathsf{T}, \ j = 1, \ldots, k-1. \tag{5.32}$$

(We use the word "full" because the tree is thought to describe the generations throughout all times, while later on, we shall consider initial pieces of such trees up to some generation.) Note that in this way, our tree is what is often called a "rooted plane tree" or "planted plane tree". Here, "rooted" refers to the fact that it is equipped with a distinguished root, and "plane" means that isomorphisms of such objects do not only have to preserve the root, but also the drawing of the tree in the plane. For example, the two trees in Figure 17 are isomorphic as usual rooted graphs, but not as planted plane trees.

Figure 17

A full genealogical tree T represents a possible genealogical tree of our population throughout all lifetimes. If it has finite height, then this means that the population dies out, and if it has infinite height, the population survives forever. The n-th generation consists of all vertices of T at distance n from the root, and the *height* of T is the supremum over all n for which the n-th generation is non-empty.

We can now construct a probability space on which one can define a denumerable collection of random variables N_u, $u \in \Sigma^*$, which are i.i.d. with distribution μ. This is just the infinite product space

$$(\Omega^{GW}, \mathcal{A}^{GW}, \mathsf{Pr}^{GW}) = \prod_{u \in \Sigma^*} (S, \mu),$$

where $S = \mathrm{supp}(\mu) = \{k \in \mathbb{N}_0 : \mu(k) > 0\}$ and the σ-algebra on S is of course the family of all subsets of S. Thus, $\mathcal{A}^{GW}$ is the σ-algebra generated by all sets of the form

$$\mathcal{B}_{v,k} = \left\{(k_u)_{u \in \Sigma^*} \in S^{\Sigma^*} : k_v = k\right\}, \quad v \in \Sigma^*, \ k \in S,$$

and if $v_1, \dots, v_r$ are distinct and $k_1, \dots, k_r \in S$ then

$$\mathsf{Pr}^{GW}(\mathcal{B}_{v_1,k_1} \cap \cdots \cap \mathcal{B}_{v_r,k_r}) = \mu(k_1) \cdots \mu(k_r).$$

The random variable N_u is just the projection

$$N_u(\omega) = k_u, \quad \text{if } \omega = (k_u)_{u \in \Sigma^*}.$$

What is then the (random) Galton–Watson tree, i.e., the full genealogical tree $\mathsf{T}(\omega)$ associated with ω? We can build it up recursively. The root ϵ belongs to $\mathsf{T}(\omega)$. If $u \in \Sigma^*$ belongs to $\mathsf{T}(\omega)$, then among its successors, precisely those points uj belong to $\mathsf{T}(\omega)$ for which $1 \le j \le N_u(\omega)$.

Thus, we can also interpret $\mathsf{T}(\omega)$ as the connected component of the root ϵ in a *percolation process.* We start with the tree structure on the whole of Σ^* and decide at random to keep or delete edges: the edge from any $u \in \Sigma^*$ to uj is kept when $j \le N_u(\omega)$, and deleted when $j > N_u(\omega)$. Thus, we obtain several connected components, each of which is a tree; we have a *random forest.* Our Galton–Watson tree is $\mathsf{T}(\omega) = \mathsf{T}_\epsilon(\omega)$, the component of ϵ. Every other connected component in that forest is also a Galton–Watson tree with another root. As a matter of fact, if we only look at forward edges, then every $u \in \Sigma^*$ can thus be considered as the root of a Galton–Watson tree T_u whose distribution is the same for all u. Furthermore, T_u and T_v are independent unless $u \preccurlyeq v$ or $v \preccurlyeq u$.

Coming back to the original Galton–Watson process, $M_n(\omega)$ is now of course the number of elements of $\mathsf{T}(\omega)$ which have length (height) n.

5.33 Exercise. Consider the Galton–Watson process with non-degenerate offspring distribution μ. Assume that $\bar{\mu} \le 1$. Let T be the resulting random tree, and let

$$|\mathsf{T}| = \sum_{n=0}^{\infty} M_n$$

be its size (number of vertices). This is the total number of the population, which is almost surely finite.

(a) Show that the probability generating function

$$g(z) = \sum_{k=1}^{\infty} \mathsf{Pr}[\,|\mathsf{T}| = k\,]\, z^k = \mathsf{E}(z^{|\mathsf{T}|})$$

satisfies the functional equation

$$g(z) = z\, f\big(g(z)\big),$$

where f is the probability generating function of μ as in (5.27).

(b) Compute the distribution of the population size $|\mathsf{T}|$, when the offspring distribution is

$$\mu = q \cdot \delta_0 + p \cdot \delta_2, \quad 0 < p \leq 1/2,\ q = 1 - p.$$

(c) Compute the distribution of the population size $|\mathsf{T}|$, when the offspring distribution is the geometric distribution

$$\mu(k) = q\, p^k, \quad k \in \mathbb{N}_0,\ 0 < p \leq 1/2,\ q = 1 - p. \tag{5.34}$$

One may observe that while the construction of our probability space is simple, it is quite abundant. We start with a very big tree but keep only a part of it as $\mathsf{T}(\omega)$. There are, of course, other possible models which are not such that large parts of the space Ω may remain "unused"; see the literature, and also the next section.

In our model, the offspring distribution μ is a probability measure on $\mathbb{N}_0$, and the number of children of any member of the population is always finite, so that the Galton–Watson tree is always locally finite. We can also admit the case where the number of children may be infinite (countable), that is, $\mu(\infty) > 0$. Then $\Sigma = \mathbb{N}$, and if $u \in \Sigma^*$ is a member of our population which has infinitely many children, then this means that $uj \in \mathsf{T}$ for every $j \in \Sigma$. The construction of the underlying probability space remains the same. In this case, we speak of an *extended* Galton–Watson process in order to distinguish it from the usual case.

5.35 Exercise. Show that when the offspring distribution satisfies $\mu(\infty) > 0$, then the population survives with positive probability. □

5.36 Example. We conclude with an example that will link this section with the one-sided infinite drunkard's walk on $\mathbb{N}_0$ which is reflecting at 0, that is, $p(0, 1) = 1$. See Example 5.10 (a) and Figure 13. We consider the recurrent case, where $p(k, k+1) = p \leq 1/2$ and $p(k, k-1) = q = 1 - p$ for $k \geq 1$. Then $\boldsymbol{t}^0$, the return time to 0, is almost surely finite. We let

$$M_k = \sum_{n=1}^{\boldsymbol{t}^0} \mathbf{1}_{[Z_{n-1}=k,\, Z_n=k+1]}$$

be the number of upward crossings of the edge from k to $k+1$. We shall work out that this is a Galton–Watson process. We have $M_0 = 1$, and a member of the k-th generation is just a single crossing of $[k, k+1]$. Its offspring consists of all crossings of $[k+1, k+2]$ that occur before the next crossing of $[k, k+1]$.

That is, we suppose that $(Z_{n-1}, Z_n) = (k, k+1)$. Since the drunkard will almost surely return to state k with probability 1 after time n, we have to consider the number of times when he makes a step from $k+1$ to $k+2$ before the next return from $k+1$ to k. The steps which the drunkard makes right after those successive

returns to state $k+1$ and before $\boldsymbol{t}^k$ are independent, and will lead to $k+2$ with probability p each time, and at last from $k+1$ to k with probability q. This is just the scheme of subsequent Bernoulli trials until the first "failure" (= step to the left), where the failure probability is q. Therefore we see that the offspring distribution is indeed the same for each crossing of an edge, and it is the geometric distribution of (5.34).

This alone does not yet prove rigorously that (M_k) is a Galton–Watson process with offspring distribution μ. To verify this, we provide an additional explanation of the genealogical tree that corresponds to the above interpretation of offspring of an individual as the upcrossings of $[k+1, k+2]$ in between two subsequent upcrossings of $[k, k+1]$.

Consider any possible *finite* trajectory of the random walk that returns to 0 in the last step and not earlier.

This is a sequence $\boldsymbol{k} = (k_0, k_1, \ldots, k_N)$ in $\mathbb{N}_0$ such that $k_0 = k_N = 0, k_1 = 1$, $k_{n+1} = k_n \pm 1$ and $k_n \neq 0$ for $0 < n < N$. The number N must be even, and $k_{N-1} = 1$.

Which such a sequence, we associate a rooted plane tree $\mathsf{T}(\boldsymbol{k})$ that is constructed in recursive steps $n = 1, \ldots, N-1$ as follows. We start with the root ϵ, which is the current vertex of the tree at step 1. At step n, we suppose to have already drawn the part of the tree that corresponds to $(k_0, \ldots, k_n)$, and we have marked a current vertex, say u, of that current part of the tree. If $n = N-1$, we are done. Otherwise, there are two cases. (1) If $k_{n+1} = k_n - 1$, then the tree remains unchanged, but we mark the predecessor of x as the new current vertex. (2) If $k_{n+1} = k_n + 1$ then we introduce a new vertex, say v, that is connected to u by an edge in the tree and becomes the new current vertex. When the procedure ends, we are back to ϵ as the current vertex.

Another way is to say that the vertex set of $\mathsf{T}(\boldsymbol{k})$ consists of all those initial subsequences $(k_0, \ldots, k_n)$ of $\boldsymbol{k}$, where $0 < n < N$ and $k_n = k_{n-1} + 1$. The predecessor of the vertex corresponding to $(k_0, \ldots, k_n)$ with $n > 1$ is the shortest initial subsequence of $(k_0, \ldots, k_n)$ that ends at k_{n-1}. The root corresponds to $(k_0, k_1) = (0, 1)$.

This is the *walk-to-tree coding* with depth-first traversal. For example, the trajectory $(0, 1, 2, 3, 2, 3, 4, 3, 4, 3, 2, 1, 2, 3, 2, 3, 2, 3, 2, 1, 0)$ induces the first of the two trees in Figure 17. (The initial and final 0 are there "automatically". We might as well consider only sequences in $\mathbb{N}$ that start and end at 1.)

Conversely, when we start with a rooted plane tree, we can read its *contour*, which is a trajectory $\boldsymbol{k}$ as above (with the initial and final 0 added). We leave it to the reader to describe how this trajectory is obtained by a recursive algorithm.

The number of vertices of $\mathsf{T}(\boldsymbol{k})$ is $|\mathsf{T}(\boldsymbol{k})| = N/2$. Any vertex at level (= distance from the root) k corresponds to precisely one step from $k-1$ to k within the trajectory $\boldsymbol{k}$. (It also encodes the next occurring step from k to $k-1$.) Thus, each vertex encodes an upward crossing, and $\mathsf{T}(\boldsymbol{k})$ is the genealogical tree that we have

described at the beginning, of which we want to prove that it is a Galton–Watson tree with offspring distribution μ.

The correspondence between $\boldsymbol{k}$ and $\mathsf{T} = \mathsf{T}(\boldsymbol{k})$ is one-to-one. The probability that our random walk trajectory until the first return to 0 induces T is

$$\mathsf{Pr}^{\mathrm{RW}}(\mathsf{T}) = \mathsf{Pr}_o[Z_0 = k_0, \ldots, Z_N = k_N] = q(pq)^{N/2-1} = q^{|\mathsf{T}|}p^{|\mathsf{T}|-1},$$

where the superscript RW means "random walk". Now, further above a general construction of a probability space was given on which one can realize a Galton–Watson tree with general offspring distribution μ. For our specific example where μ is as in (5.34), we can also use a simpler model.

Namely, it is sufficient to have a sequence $(X_n)_{n\geq 1}$ of i.i.d. ± 1-valued random variables with $\mathsf{Pr}^{\mathrm{GW}}[X_n = 1] = p$ and $\mathsf{Pr}^{\mathrm{GW}}[X_n = -1] = q$. The superscript GW refers to the Galton–Watson tree that we are going to construct. We consider the value $+1$ as "success" and -1 as "failure". With probability one, both values $+1$ and -1 occur infinitely often. With (X_n) we can build up recursively a random genealogical tree T, based on the breadth-first order of any rooted plane tree. This is the linear order where for vertices u, v we have $u < v$ if either the distances to the root satisfy $|u| < |v|$, or $|u| = |v|$ and u is further to the left than v.

At the beginning, the only vertex of the tree is the root ϵ. For each success before the first failure, we draw one offspring of the root. At the first failure, ϵ is declared *processed*. At each subsequent step, we consider the next unprocessed vertex of the current tree in the breadth-first order, and we give it one offspring for each success before the next failure. When that failure occurs, that vertex is processed, and we turn to the next vertex in the list. The process ends when no more unprocessed vertex is available; it continues, otherwise. By construction, the offspring numbers of different vertices are i.i.d. and geometrically distributed. Thus, we obtain a Galton–Watson tree as proposed. Since

$$\bar{\mu} \leq 1 \iff p \leq 1/2,$$

that tree is a.s. finite in our case. The probability that it will be a given finite rooted plane tree T is obtained as follows: each edge of T must correspond to a success, and for each vertex, there must be one failure (when that vertex stops to create offspring). Thus, the first $2|\mathsf{T}| - 1$ members of (X_n) must consist of $|\mathsf{T}| - 1$ successes and $|\mathsf{T}|$ failures. Therefore

$$\mathsf{Pr}^{\mathrm{GW}}(\mathsf{T}) = q^{|\mathsf{T}|}p^{|\mathsf{T}|-1} = \mathsf{Pr}^{\mathrm{RW}}(\mathsf{T}).$$

We have obtained a one-to-one correspondence between random walk trajectories until the first return to 0 and Galton–Watson trees with geometric offspring distribution, and that correspondence preserves the probability measure.

This proves that the random tree created from the upcrossings of the drunkard's walk with $p \leq 1/2$ is indeed the proposed Galton–Watson tree.

We remark that the above correspondence between drunkard's walks and Galton–Watson trees was first described by HARRIS [30].

5.37 Exercise. Show that when $p > 1/2$ in Example 5.36, then (M_k) is *not* a Galton–Watson process.

C Branching Markov chains

We now combine the two models: Markov chain and Galton–Watson process. We imagine that at a given time n, the members of the n-th generation of a finite population occupy various points (sites) of the state space of a Markov chain (X, P). Multiple occupancies are allowed, and the initial generation has only one member (the ancestor). The population evolves according to a Galton–Watson process with offspring distribution μ, and at the same time performs random moves according to the underlying Markov chain. In order to create the members of next generation plus the sites that they occupy, each member of the n-th generation produces its k children, according to the underlying Galton–Watson process, with probability $\mu(k)$ and then dies (or we may say that it splits into k new members of the population). Each of those new members which are thus "born" at a site $x \in X$ then moves instantly to a random new site y with probability $p(x, y)$, independently of all others and independently of the past. In this way, we get the next generation and the positions of its members. Here, we always suppose that the offspring distribution lives on $\mathbb{N}_0$ (that is, $\mu(\infty) = 0$) and is non-degenerate (we do not have $\mu(0) = 1$ or $\mu(1) = 1$). We do allow that $\mu(0) > 0$, in which case the process may die out with positive probability.

The construction of a probability space on which this model may be realized can be elaborated at various levels of rigour. One that comprises all the available information is the following.

A single "trajectory" should consist of a full generation tree T, where to each vertex u of T (= element of the population) we attach an element $x \in X$, which is the position of u. If uj is a successor of u in T, then uj occupies site y with probability $p(x, y)$, given that u occupies x. This has to be independent of all the other members of the population.

Thus (recalling that S is the support of μ), our space is

$$\Omega^{\mathrm{BMC}} = (S \times X)^{\Sigma^*} = \big\{\omega = (k_u, x_u)_{u \in \Sigma^*} : k_u \in S,\ x_u \in X\big\}.$$

It is once more equipped with the product σ-algebra $\mathcal{A}^{\mathrm{BMC}}$ of the discrete one on $S \times X$. For $\omega \in \Omega^{\mathrm{BMC}}$, let $\bar{\omega} = (k_u)_{u \in \Sigma^*}$ be its projection onto Ω^{GW}. The associated Galton–Watson tree is $\mathsf{T}(\omega) = \mathsf{T}(\bar{\omega})$, defined as at the end of the preceding section.

Now let $\boldsymbol{\tau}$ be any finite subtree of Σ^* containing the root, and choose an element $a_u \in X$ for every $u \in \boldsymbol{\tau}$. For every $u \in \boldsymbol{\tau}$, we let

$$\kappa(u) = \kappa_{\boldsymbol{\tau}}(u) = \max\{j \in \Sigma : uj \in \boldsymbol{\tau}\},$$

in particular $\kappa(u) = 0$ if u has no successor in $\boldsymbol{\tau}$. With these data, we associate the basic event (analogous to the cylinder sets of Section 1.B)

$$\mathsf{D}(\boldsymbol{\tau}\,; a_u, u \in \boldsymbol{\tau}) = \big\{ \omega = (k_u, x_u)_{u\in\Sigma^*} : x_u = a_u \text{ and } k_u \geq \kappa(u) \text{ for all } u \in \boldsymbol{\tau} \big\}.$$

Since $k_u = N_u(\omega)$ is the number of children of u in the Galton–Watson process, the condition $k_u \geq \kappa(u)$ means that whenever $uj \in \boldsymbol{\tau}$ then this is one of the children of u. Thus, $\mathsf{D}(\boldsymbol{\tau}; a_u, u \in \boldsymbol{\tau})$ is the event that $\boldsymbol{\tau}$ is part of $\mathsf{T}(\omega)$ and that each member $u \in \boldsymbol{\tau}$ of the population occupies the site a_u of X.

Given a starting point $x \in X$, the probability measure governing the branching Markov chain is now the unique measure on $\mathcal{A}^{\mathrm{BMC}}$ with

$$\begin{aligned} &\mathsf{Pr}_x^{\mathrm{BMC}}\big(\mathsf{D}(\boldsymbol{\tau}; a_u, u \in \boldsymbol{\tau})\big) \\ &\quad = \delta_x(a_\epsilon)\, \mu\big[\kappa(\epsilon), \infty\big) \prod_{u\in\boldsymbol{\tau}\setminus\{\epsilon\}} \mu\big[\kappa(u), \infty\big)\, p(a_{u^-}, a_u), \end{aligned} \tag{5.38}$$

where $\mu[j, \infty) = \mu\big(\{k : k \geq j\}\big)$.

We can now introduce the random variables that describe the branching Markov chain $\mathrm{BMC}(X, P, \mu)$ starting at $x \in X$, where μ is the offspring distribution. If $\omega = (k_u, x_u)_{u\in\Sigma^*}$ then for $u \in \Sigma^*$, we write

$$Z_u(\omega) = x_u$$

for the site occupied by the element u. Of course, we are only interested in $Z_u(\omega)$ when $u \in \mathsf{T}(\omega)$, that is, when u belongs to the population that descends from the ancestor ϵ. The branching Markov chain is then the (random) Galton–Watson tree T together with the family of random variables $(Z_u)_{u\in\mathsf{T}}$ indexed by the Galton–Watson tree T. The Markov property extended to this setting says the following.

5.39 Facts. (1) For $j \in \Sigma \subset \Sigma^*$ and $x, y \in X$,

$$\mathsf{Pr}_x^{\mathrm{BMC}}[j \in \mathsf{T},\ Z_j = y] = \mu[j, \infty)\, p(x, y).$$

(2) If $u, u' \in \Sigma^*$ are such that neither $u \preccurlyeq v$ nor $v \preccurlyeq u$ then the families $\big(\mathsf{T}_u\,; (Z_v)_{v\in\mathsf{T}_u}\big)$ and $\big(\mathsf{T}'_u\,; (Z_{v'})_{v'\in\mathsf{T}_{u'}}\big)$ are independent.

(3) Given that $Z_u = y$, the family $(Z_v)_{v\in\mathsf{T}_u}$ is $\mathrm{BMC}(X, P, \mu)$ starting at y.

(4) In particular, if T contains a *ray* $\{u_n = j_1 \cdots j_n : n \geq 0\}$ (infinite path starting at ϵ) then $(Z_{u_n})_{n\geq 0}$ is a Markov chain on X with transition matrix P.

Property (3) also comprises the generalization of time-homogeneity. The number of members of the n-th generation that occupy the site $y \in X$ is

$$M_n^y(\omega) = \big|\{u \in \mathsf{T}(\omega) : |u| = n,\ Z_u(\omega) = y\}\big|.$$

Thus, the underlying Galton–Watson process (number of members of the n-th generation) is

$$M_n = \sum_{y \in X} M_n^y,$$

while

$$M^y = \sum_{n=0}^{\infty} M_n^y$$

is the total number of occupancies of the site $y \in X$ during the whole lifetime of the BMC. The random variable M^y takes its values in $\mathbb{N} \cup \{\infty\}$. The following may be quite clear.

5.40 Lemma. *Let $u = j_1 \cdots j_n \in \Sigma^*$ and $x, y \in X$. Then*

$$\mathsf{Pr}_x^{\mathrm{BMC}}[u \in \mathsf{T},\ Z_u = y] = p^{(n)}(x,y) \prod_{i=1}^{n} \mu[j_i, \infty).$$

Proof. We use induction on n. If $n = 1$ and $u = j \in \Sigma$ then this is (5.39.1). Now suppose the statement is true for $u = j_1 \cdots j_n$ and let $v = uj_{n+1} \in \Sigma^*$. Then $v \in \mathsf{T}$ implies $u \in \mathsf{T}$. Using this and (5.39.3),

$$\begin{aligned}
&\mathsf{Pr}_x^{\mathrm{BMC}}[v \in \mathsf{T},\ Z_v = y] \\
&= \sum_{w \in X} \mathsf{Pr}_x^{\mathrm{BMC}}[v \in \mathsf{T},\ Z_v = y \mid u \in \mathsf{T},\ Z_u = w]\ \mathsf{Pr}_x^{\mathrm{BMC}}[u \in \mathsf{T}, Z_u = w] \\
&= \sum_{w \in X} \mathsf{Pr}_x^{\mathrm{BMC}}[uj_{n+1} \in \mathsf{T},\ Z_{uj_{n+1}} = y \mid u \in \mathsf{T},\ Z_u = w]\ p^{(n)}(x,w) \prod_{i=1}^{n} \mu[j_i, \infty) \\
&= \sum_{w \in X} \mathsf{Pr}_w^{\mathrm{BMC}}[j_{n+1} \in \mathsf{T},\ Z_{j_{n+1}} = y]\ p^{(n)}(x,w) \prod_{i=1}^{n} \mu[j_i, \infty) \\
&= \sum_{w \in X} p^{(n)}(x,w) p(w,y) \prod_{i=1}^{n+1} \mu[j_i, \infty).
\end{aligned}$$

This leads to the proposed statement. □

5.41 Exercise. (a) Deduce the following from Lemma 5.40. When the initial site is x, then the expected number of members of the n-th generation that occupy the site y is

$$\mathsf{E}_x^{\mathrm{BMC}}(M_n^y) = p^{(n)}(x,y)\, \bar{\mu}^n,$$

where (recall) $\bar{\mu}$ is the expected offspring number.

[Hint: write M_n^y as a sum of indicator functions of events as those in Lemma 5.40.]

(b) Let $x, y \in X$ and $p^{(k)}(x, y) > 0$. Show that $\mathsf{Pr}_x^{\mathrm{BMC}}[M_k^y \geq 1] > 0$. □

The question of recurrence or transience becomes more subtle for BMC than for ordinary Markov chains. We ask for the probability that throughout the lifetime of the process, some site y is occupied by infinitely many members in the successive generations of the population. In analogy with § 3.A, we define the quantities

$$H^{\mathrm{BMC}}(x, y) = \mathsf{Pr}_x^{\mathrm{BMC}}[M^y = \infty], \quad x, y \in X.$$

5.42 Theorem. *One either has* (a) $H^{\mathrm{BMC}}(x, y) = 1$ *for all* $x, y \in X$, *or* (b) $0 < H^{\mathrm{BMC}}(x, y) < 1$ *for all* $x, y \in X$, *or* (c) $H^{\mathrm{BMC}}(x, y) = 0$ *for all* $x, y \in X$.

Before the proof, we need the following.

5.43 Lemma. *For all* $x, y, y' \in X$, *one has* $\mathsf{Pr}_x^{\mathrm{BMC}}[M^y = \infty,\ M^{y'} < \infty] = 0$.

Proof. Let k be such that $p^{(k)}(y, y') > 0$. First of all, using continuity of the probability of increasing sequences,

$$\mathsf{Pr}_x^{\mathrm{BMC}}[M^y = \infty,\ M^{y'} < \infty] = \lim_{m \to \infty} \mathsf{Pr}_x^{\mathrm{BMC}}\big([M^y = \infty] \cap B_m\big),$$

where $B_m = [M_n^{y'} = 0 \text{ for all } n \geq m]$. Now, $[M^y = \infty] \cap B_m$ is the limit (intersection) of the decreasing sequence of the events $A_{m,r}$, where

$$A_{m,r} = \left[\begin{array}{c}\text{There are } n(1), \ldots, n(r) \geq m \text{ with } n(j) > n(j-1) + k \\ \text{such that } M^y_{n(j)} \geq 1 \text{ for all } j\end{array}\right] \cap B_m$$

$$\subset B_{m,r} = \left[\begin{array}{c}\text{There are } n(1), \ldots, n(r) \geq m \text{ with } n(j) > n(j-1) + k \\ \text{such that } M^y_{n(j)} \geq 1 \text{ and } M^{y'}_{n(j)+k} = 0 \text{ for all } j\end{array}\right].$$

If $\omega \in B_{m,r}$ then there are (random) elements $u(j) \in \mathsf{T}(\omega)$ with $|u(j)| = n(j)$ such that $Z_{u(j)}(\omega) = x$. Since $n(j) > n(j-1) + k$, the initial parts with height k of the generation trees rooted at the $u(j)$ are independent. None of the descendants of the $u(j)$ after k generations occupies site y'. Therefore, if we let $\delta = \mathsf{Pr}_y^{\mathrm{BMC}}[M_k^{y'} = 0]$ then $\delta < 1$ by Exercise 5.41 (b), and

$$\mathsf{Pr}_x^{\mathrm{BMC}}(A_{m,r}) \leq \mathsf{Pr}_x^{\mathrm{BMC}}(B_{m,r}) \leq \delta^r.$$

Letting $r \to \infty$, we obtain $\mathsf{Pr}_x^{\mathrm{BMC}}\big([M^y = \infty] \cap B_m\big) = 0$, as required. □

Proof of Theorem 5.42. We first fix y. Let $x, x' \in X$ be such that $p(x', x) > 0$. Choose $k \geq 1$ be such that $\mu(k) > 0$. The probability that the BMC starting at x' produces $\mu(k)$ children, which all move to x, is $\mathsf{Pr}_{x'}^{\mathrm{BMC}}\big[N_\epsilon = k = M_1^x\big] = \mu(k)\, p(x', x)^k$. Therefore

$$\begin{aligned} H^{\mathrm{BMC}}(x', y) &\geq \mathsf{Pr}_{x'}^{\mathrm{BMC}}\big[N_\epsilon = k = M_1^x,\ M^y = \infty\big] \\ &= \mu(k)\, p(x', x)^k\ \mathsf{Pr}_{x'}^{\mathrm{BMC}}\big[M^y = \infty \mid N_\epsilon = k = M_1^x\big]. \end{aligned}$$

The last factor coincides with the probability that BMC with k particles (instead of one) starting at x evolves such that $M^y = \infty$. That is, we have independent replicas $M^{y,1}, \ldots, M^{y,k}$ of M^y descending from each of the k children of ϵ that are now occupying site x, and

$$\begin{aligned} H^{\mathrm{BMC}}(x', y) \big/ \big(\mu(k)\, p(x', x)^k\big) &\geq \mathsf{Pr}_x^{\mathrm{BMC}}[M^{y,1} + \cdots + M^{y,k} = \infty] \\ &= \big(1 - \mathsf{Pr}_x^{\mathrm{BMC}}[M^{y,1} < \infty, \ldots, M^{y,k} < \infty]\big) \\ &= \big(1 - \big(1 - H^{\mathrm{BMC}}(x, y)\big)^k\big) \geq H^{\mathrm{BMC}}(x, y). \end{aligned}$$

Irreducibility now implies that whenever $H^{\mathrm{BMC}}(x, y) > 0$ for some x, then we have $H^{\mathrm{BMC}}(x', y) > 0$ for all $x' \in X$. In the same way,

$$\begin{aligned} 1 - H^{\mathrm{BMC}}(x', y) &= \mathsf{Pr}_{x'}^{\mathrm{BMC}}[M^y < \infty] \\ &\geq \mu(k)\, p(x', x)^k\ \mathsf{Pr}_x^{\mathrm{BMC}}[M^{y,1} + \cdots + M^{y,k} < \infty] \\ &= \mu(k)\, p(x', x)^k \big(1 - H^{\mathrm{BMC}}(x, y)\big)^k. \end{aligned}$$

Again, irreducibility implies that whenever $H^{\mathrm{BMC}}(x, y) < 1$ for some x then $H^{\mathrm{BMC}}(x', y) < 1$ for all $x' \in X$. Now Lemma 5.43 implies that when we have $H^{\mathrm{BMC}}(x, y) = 1$ for some x, y then this holds for all $x, y \in X$, see the next exercise. □

5.44 Exercise. Show that indeed Lemma 5.43 together with the preceding arguments yields the final part of the proof of the theorem. □

The last theorem justifies the following definition.

5.45 Definition. The branching Markov chain (X, P, μ) is called *strongly recurrent* if $H^{\mathrm{BMC}}(x, y) = 1$, *weakly recurrent* if $0 < H^{\mathrm{BMC}}(x, y) < 1$, and *transient* if $H^{\mathrm{BMC}}(x, y) = 0$ for some (equivalently, all) $x, y \in X$.

Contrary to ordinary Markov chains, we do not have a zero-one law here as in Theorem 3.2: we shall see examples for each of the three regimes of Definition 5.45. Before this, we shall undertake some additional efforts in order to establish a general transience criterion.

Embedded process

We introduce a new, embedded Galton–Watson process whose population is just the set of elements of the original Galton–Watson tree T that occupy the starting point $x \in X$ of $\mathrm{BMC}(X, P, \mu)$. We define a sequence $(\mathsf{W}_n^x)_{n\geq 0}$ of subsets of T: we start with $\mathsf{W}_0^x = \{\epsilon\}$, and for $n \geq 1$,

$$\mathsf{W}_n^x = \big\{v \in \mathsf{T} : Z_v = x,\ |\{u \in \mathsf{T} : \epsilon \neq u \preccurlyeq v,\ Z_u = x\}| = n\big\}.$$

Thus, $v \in \mathsf{W}_n$ means that if $\epsilon = u_0, u_1, \dots, u_r = v$ are the successive points on the shortest path in T from ϵ to v, then Z_{u_k} $(k = 0, \dots, r)$ starts at x and returns to x precisely n times. In particular, the definition of W_1 should remind the reader of the "first return" stopping time $\boldsymbol{t}^x$ defined for ordinary Markov chains in (1.26). We set $Y_n^x = |\mathsf{W}_n^x|$.

5.46 Lemma. *The sequence $(Y_n^x)_{n\geq 0}$ is an extended Galton–Watson process with non-degenerate offspring distribution*

$$\nu^x(m) = \mathsf{Pr}_x^{\mathrm{BMC}}[Y_1^x = m], \quad m \in \mathbb{N}_0 \cup \{\infty\}.$$

Its expected offspring number is

$$\overline{\nu^x} = U(x, x|\bar{\mu}) = \sum_{n=1}^{\infty} u^{(n)}(x, x)\, \bar{\mu}^n,$$

where $\bar{\mu}$ is the expected offspring number in $\mathrm{BMC}(X, P, \mu)$, and $U(x, x|\cdot)$ is the generating function of the first return probabilities to x for the Markov chain (X, P), as defined in (1.37).

Proof. We know from (5.39.3) that for distinct elements $u \in \mathsf{W}_n^x$, the families $(Z_v)_{v\in\mathsf{T}_u}$ are independent copies of $\mathrm{BMC}(X, P, \mu)$ starting at x. Therefore the random numbers $|\{v \in \mathsf{W}_{n+1}^x : u \preccurlyeq v\}|$, where $u \in \mathsf{W}_n^x$, are independent and have all the same distribution as Y_1^x. Now,

$$\mathsf{W}_{n+1}^x = \bigcup_{u\in\mathsf{W}_n^x} \{v \in \mathsf{W}_{n+1}^x : u \preccurlyeq v\},$$

a disjoint union. This makes it clear that $(Y_n^x)_{n\geq 0}$ is a (possibly extended) Galton–Watson process with offspring distribution ν^x. The sets W_n^x, $n \in \mathbb{N}$, are its successive generations. In order to describe ν^x, we consider the events $[v \in \mathsf{W}_1^x]$.

5.47 Exercise. Prove that if $v = j_1 \cdots j_n \in \Sigma^+$ then

$$\mathsf{Pr}_x^{\mathrm{BMC}}[v \in \mathsf{W}_1^x] = u^{(n)}(x, x) \prod_{i=1}^{n} \mu[j_i, \infty). \qquad \square$$

We resume the proof of Lemma 5.46 by observing that

$$Y_1^x = \sum_{v\in\Sigma^+} \mathbf{1}_{[v\in \mathsf{W}_1^x]} = \sum_{n=1}^{\infty} \sum_{v\in\Sigma^n} \mathbf{1}_{[v\in \mathsf{W}_1^x]}.$$

Now

$$\begin{aligned} \mathsf{E}_x^{\mathrm{BMC}}\Big(\sum_{v\in\Sigma^n} \mathbf{1}_{[v\in \mathsf{W}_1^x]}\Big) &= \sum_{v=j_1\cdots j_n\in\Sigma^n} \mathsf{Pr}_x^{\mathrm{BMC}}[v\in \mathsf{W}_1^x] \\ &= \sum_{j_1,\dots,j_n\in\mathbb{N}} u^{(n)}(x,x) \prod_{i=1}^{n} \mu[j_i,\infty) \\ &= u^{(n)}(x,x) \prod_{i=1}^{n} \Big(\sum_{j_i\in\mathbb{N}} \mu[j_i,\infty)\Big) \\ &= u^{(n)}(x,x)\, \bar{\mu}^n. \end{aligned} \tag{5.48}$$

This leads to the proposed formula for $\overline{\nu^x} = \mathsf{E}_x^{\mathrm{BMC}}(Y_1^x)$. At last, we show that ν^x is non-degenerate, that is, $\mathsf{Pr}_x^{\mathrm{BMC}}[Y_1^x = 1] < 1$, since clearly $\nu_x(0) < 1$ (e.g., by Exercise 5.47). By assumption, $\mu(1) < 1$. If the population dies out at the first step then also $Y_1^x = 0$. That is, $\mathsf{Pr}_x^{\mathrm{BMC}}[Y_1^x = 0] \geq \mu(0)$, and if $\mu(0) > 0$ then $\mathsf{Pr}_x^{\mathrm{BMC}}[Y_1^x = 1] < 1$. So assume that $\mu(0) = 0$. Then there is $m \geq 2$ such that $\mu(m) > 0$. By irreducibility, there is $n \geq 1$ such that $u^{(n)}(x,x) > 0$. In particular, there are $x_0, x_1, \dots, x_n$ in X with $x_0 = x_n = x$ and $x_k \neq x$ for $1 \leq k \leq n-1$ such that $p(x_{k-1}, x_k) > 0$. We can consider the subtree $\boldsymbol{\tau}$ of Σ^* with height n that consists of all elements $u \in \{1,\dots,m\}^* \subset \Sigma^*$ with $|u| \leq m$. Then (5.38) yields

$$\begin{aligned} \mathsf{Pr}_x^{\mathrm{BMC}}[Y_1^x \geq m^n] &\geq \mathsf{Pr}_x^{\mathrm{BMC}}[\boldsymbol{\tau}\subset \mathsf{T},\ Z_u = x_k \text{ for all } u\in\boldsymbol{\tau} \text{ with } |u| = k] \\ &= \mu[m,\infty)^{1+m+\cdots+m^{n-1}} \prod_{k=1}^{n} p(x_{k-1},x_k)^{m^k} > 0, \end{aligned}$$

and ν^x is non-degenerate. □

5.49 Theorem. *For an irreducible Markov chain (X, P), $\mathrm{BMC}(X, P, \mu)$ is transient if and only if $\bar{\mu} \leq 1/\rho(P)$.*

Proof. For BMC starting at x, the total number of occupancies of x is

$$M^x = \sum_{n=0}^{\infty} Y_n^x.$$

We can apply Theorem 5.30 to the embedded Galton–Watson process $(Y_n^x)_{n\geq 0}$, taking into account Exercise 5.35 in the case when $\mathsf{Pr}_x^{\mathrm{BMC}}[Y_1^x = \infty] > 0$. Namely,

$\mathsf{Pr}_x^{\mathrm{BMC}}[M^x < \infty] = 1$ precisely when the embedded process has average offspring ≤ 1, that is, when $U(x, x|\bar{\mu}) \leq 1$. By Proposition 2.28, the latter holds if and only if $\bar{\mu} \leq 1/\rho(P)$. □

Regarding the last theorem, it was observed by Benjamini and Peres [6] that $\mathrm{BMC}(X, P, \mu)$ is transient when $\bar{\mu} < 1/\rho(P)$ and recurrent when $\bar{\mu} > 1/\rho(P)$. Transience in the critical case $\bar{\mu} = 1/\rho(P)$ was settled by Gantert and Müller [23].[1]

While the last theorem provides a good tool to distinguish between the transient and the recurrent regime, at the moment (2009) no general criterion of comparable simplicity is known to distinguish between weak and strong recurrence. The following is quite obvious.

5.50 Lemma. *Let $x \in X$. $\mathrm{BMC}(X, P, \mu)$ is strongly recurrent if and only if*

$$\mathsf{Pr}_x^{\mathrm{BMC}}[Z_u = x \text{ for some } u \in \mathsf{T} \setminus \{\epsilon\}] = 1,$$

or equivalently (when $|X| \geq 2$), for all $y \in X \setminus \{x\}$,

$$\mathsf{Pr}_y^{\mathrm{BMC}}[Z_u = x \text{ for some } u \in \mathsf{T} \setminus \{\epsilon\}] = 1.$$

For this it is necessary that $\mu(0) = 0$.

Proof. We have strong recurrence if and only if the extinction probability for the embedded Galton–Watson process (Y_n^x) is 0. A quick look at Theorem 5.30 and Figures 16 a, b convinces us that this is true if and only if $\overline{\nu^x} > 1$ and $\nu^x(0) = 0$. Since ν^x is non-degenerate, $\nu^x(0) = 0$ implies $\overline{\nu^x} > 1$. Therefore we have strong recurrence if and only if $\nu^x(0) = \mathsf{Pr}_x^{\mathrm{BMC}}[Y_1^x = 0] = 0$. Since $\mathsf{Pr}_x^{\mathrm{BMC}}[Y_1^x = 0] = 1 - \mathsf{Pr}_x^{\mathrm{BMC}}[Z_u = x \text{ for some } u \in \mathsf{T} \setminus \{\epsilon\}]$, the proposed criterion follows.

If $\mu(0) > 0$ then with positive probability, the underlying Galton–Watson tree T consists only of ϵ, in which case $Y_1^x = 0$. Therefore $\mathsf{Pr}_x^{\mathrm{BMC}}[M^y = \infty] \leq \mathsf{Pr}_x^{\mathrm{BMC}}[Y_1^x > 0] < 1$, and recurrence cannot be strong. This proves the first criterion.

It is clear that the second criterion is necessary for strong recurrence: if infinitely members of the population occupy site x, when the starting point is $y \neq x$, then at least one member distinct from ϵ must occupy x. Conversely, suppose that the criterion is satisfied. Then necessarily $\mu(0) = 0$, and each member of the non-empty first generation moves to some $y \in X$. If one of them stays at x (when $p(x, x) > 0$) then we have an element $u \in \mathsf{T} \setminus \{\epsilon\}$ such that $Z_u = x$. If one of them has moved to $y \neq x$, then our criterion guarantees that one of its descendants will come back to x with probability 1. Therefore the first criterion is satisfied, and we have strong recurrence. □

[1] The above short proof of Theorem 5.49 is the outcome of a discussion between Nina Gantert and the author.

5.51 Exercise. Elaborate the details of the last argument rigorously. □

We remark that in our definition of strong recurrence we mainly have in mind the case when $\mu(0) = 0$: there is always at least one offspring. In the case when $\mu(0) > 0$, we have the obvious inequality $0 \le H^{\mathrm{BMC}}(x, y) \le 1 - \lambda$, where λ is the extinction probability of the underlying Galton–Watson process. One then may strengthen Theorem 5.42 by showing that one of $H^{\mathrm{BMC}}(x, y) = 1 - \lambda$, $0 < H^{\mathrm{BMC}}(x, y) < 1 - \lambda$ or $H^{\mathrm{BMC}}(x, y) = 0$ holds for all $x, y \in X$. Then one may redefine strong recurrence by $H^{\mathrm{BMC}}(x, y) = 1 - \lambda$. We leave the details of the modified proof of Theorem 5.42 to the interested reader.

We note the following consequence of (5.39.4) and Lemma 5.50.

5.52 Exercise. Show that if the irreducible Markov chain (X, P) is recurrent and the offspring distribution satisfies $\mu(0) = 0$ then $\mathrm{BMC}(X, P, \mu)$ is strongly recurrent. □

There is one class of Markov chains where a complete description of strong recurrence is available, namely random walks on groups as considered in (4.18). Here we run into a small conflict of notation, since in the present chapter, μ stands for the offspring distribution of a Galton–Watson process and not for the law of a random walk on a group. Therefore we just write P for the (irreducible) transition matrix of our random walk on $X = \mathfrak{G}$ and recall that it satisfies

$$p^{(n)}(x, y) = p^{(n)}(gx, gy) \quad \text{for all } x, y, g \in \mathfrak{G} \text{ and } n \in \mathbb{N}_0.$$

An obvious, but important consequence is the following.

5.53 Lemma. *If* $(Z_u)_{u \in \mathsf{T}}$ *is* $\mathrm{BMC}(\mathfrak{G}, P, \mu)$ *starting at* $x \in \mathfrak{G}$ *and* $g \in \mathfrak{G}$ *then* $(gZ_u)_{u \in \mathsf{T}}$ *is (a realization of)* $\mathrm{BMC}(\mathfrak{G}, P, \mu)$ *starting at* $y = gx$.

(By "a realization of" we mean the following: we have chosen a concrete construction of a probability space and associated family of random variables that are our model of BMC. The family $(gZ_u)_{u \in \mathsf{T}}$ is not exactly the one of this model, but it has the same distribution.)

5.54 Theorem. *Let* $(\mathfrak{G}, P)$ *be an irreducible random walk on the group* $\mathfrak{G}$. *Then* $\mathrm{BMC}(\mathfrak{G}, P, \mu)$ *is strongly recurrent if and only if the offspring distribution* μ *satisfies* $\mu(0) = 0$ *and* $\bar{\mu} > 1/\rho(P)$.

Proof. We know that transience holds precisely when $\bar{\mu} \le 1/\rho(P)$. So we assume that $\mu(0) = 0$ and $\bar{\mu} > 1/\rho(P)$. Then there is $k \in \mathbb{N}$ such that

$$\alpha = p^{(k)}(x, x)\, \bar{\mu}^k > 1,$$

which is independent of x by group invariance. We fix this k and construct another family of embedded Galton–Watson processes of the BMC. Suppose that $u \in \mathsf{T}$. We define recursively a sequence of subsets V_n^u of T (as long as they are non-empty):

$$\mathsf{V}_0^u = \{u\}, \quad \mathsf{V}_1^u = \{v \in \mathsf{T}_u : |v| = |u| + k, \ Z_v = Z_u\}, \quad \text{and} \quad \mathsf{V}_{n+1}^u = \bigcup_{w \in \mathsf{V}_n^u} \mathsf{V}_1^w.$$

In words, if $|u| = n_0$ then V_1^u consists of all descendants of u in generation number $n_0 + k$ that occupy the same site as u, and V_{n+1}^u consists of all descendants of the elements in V_n^u that belong to generation number $n_0 + (n+1)k$ and occupy again the same site. Then $\big(|\mathsf{V}_n^u|\big)_{n \geq 0}$ is a Galton–Watson process, which we call shortly the (u,k)-process. By Lemma 5.53, all the (u,k)-processes, where $u \in \mathsf{T}$ (and k is fixed), have the same offspring distribution. Suppose that $Z_u = y$. Then, by Exercise 5.41 (a), the average of this offspring distribution is $p^{(k)}(y,y)\,\bar{\mu}^k = \alpha > 1$. By Theorem 5.30, we have $\lambda < 1$ for the extinction probability of the (u,k)-process, and λ does not depend on $y = Z_u$.

Now recall that we assume $\mu(0) = 0$ and that $\mu(1) \neq 1$, so that $\mu[2,\infty) > 0$. We set $u(m) = 0 \cdots 01 \in \Sigma^+$, the word starting with $(m-1)$ letters 0 and the last letter 1, where $m \in \mathbb{N}$. Then the predecessor of $u(m)$ (the word $0 \cdots 0$ with length $m-1$) is in T almost surely, so that $\mathsf{Pr}[u(m) \in \mathsf{T}] = \mu[2,\infty) > 0$. For $m_1 \neq m_2$, none of $u(m_1)$ or $u(m_2)$ is a predecessor of the other. Therefore (5.39.2) implies that all the $\big(u(m),k\big)$-processes are mutually independent, and so are the events

$$B_m = \big[u(m) \in \mathsf{T}, \text{ the } \big(u(m),k\big)\text{-process survives}\big].$$

Since $\mathsf{Pr}(B_m) = (1-\lambda)\,\mu[2,\infty) > 0$ is constant, the complements of the B_m satisfy $\mathsf{Pr}\big(\bigcap_m B_m^c\big) = 0$. On $\bigcup_m B_m$, at least one of the $\big(u(n),k\big)$-processes survives, and all of its members belong to T. Therefore

$$\mathsf{Pr}_x[M^y = \infty \text{ for some } y \in X] = 1.$$

On the other hand, by Lemma 5.43,

$$\mathsf{Pr}_x[M^x < \infty, \ M^y = \infty \text{ for some } y \in X] \leq \sum_{y \in X} \mathsf{Pr}_x[M^x < \infty, \ M^y = \infty] = 0.$$

Thus $\mathsf{Pr}_x[M^x < \infty] = 0$, as proposed. □

We see that in the group invariant case, weak recurrence never occurs. Now we construct examples where one can observe the phase transition from transience via weak recurrence to strong recurrence, as the average offspring number $\bar{\mu}$ increases.

We start with two irreducible Markov chains (X_1, P_1) and (X_2, P_2) and connect the two state spaces at a single "root" o. That is, we assume that $X_1 \cap X_2 = \{o\}$. (Or, in other words, we identify "roots" $o_i \in X_i$, $i = 1,2$, to become one common

point o, while keeping the rest of the X_i disjoint.) Then we choose parameters $\alpha_1, \alpha_2 = 1 - \alpha_1 > 0$ and define a new Markov chain (X, P), where $X = X_1 \cup X_2$ and P is given as follows (where $i = 1, 2$):

$$p(x,y) = \begin{cases} p_i(x,y), & \text{if } x, y \in X_i \text{ and } x \neq o, \\ \alpha_i \, p_i(o,y), & \text{if } x = o \text{ and } y \in X_i \setminus \{o\}, \\ \alpha_1 \, p_1(o,o) + \alpha_2 \, p_2(o,o), & \text{if } x = y = o, \text{ and} \\ 0, & \text{in all other cases.} \end{cases} \tag{5.55}$$

In words, if the Markov chain at some time has its current state in $X_i \setminus \{o\}$, then it evolves in the next step according to P_i, while if the current state is o, then a coin is tossed (whose outcomes are 1 or 2 with probability α_1 and α_2, respectively) in order to decide whether to proceed according to $p_1(o, \cdot)$ or $p_2(o, \cdot)$.

The new Markov chain is irreducible, and o is a cut point between $X_1 \setminus \{o\}$ and $X_2 \setminus \{o\}$ (and vice versa) in the sense of Definition 1.42. It is immediate from Theorem 1.38 that

$$U(o,o|z) = \alpha_1 \, U_1(o,o|z) + \alpha_2 \, U_2(o,o|z).$$

Let s_i and s be the radii of convergence of the power series $U_i(o,o|z)$ $(i = 1, 2)$ and $U(o,o|z)$, respectively. Then (since these power series have non-negative coefficients) $\mathsf{s} = \min\{\mathsf{s}_1, \mathsf{s}_2\}$.

5.56 Lemma. $\min\{\rho(P_1), \rho(P_2)\} \leq \rho(P) \leq \max\{\rho(P_1), \rho(P_2)\}$.

If (X_i, P_i) is not $\rho(P_i)$-positive-recurrent for $i = 1, 2$ then

$$\rho(P) = \max\{\rho(P_1), \rho(P_2)\}.$$

Proof. We use Proposition 2.28. Let $\mathsf{r}_i = 1/\rho(P_i)$ and $\mathsf{r} = 1/\rho(P)$ be the radii of convergence of the respective Green functions.

If $z_0 = \min\{\mathsf{r}_1, \mathsf{r}_2\}$ then $U_i(o,o|z_0) \leq 1$ for $i = 1, 2$, whence $U(o,o|z_0) \leq 1$. Therefore $z_0 \leq \mathsf{r}$.

Conversely, if $z > \max\{\mathsf{r}_1, \mathsf{r}_2\}$ then $U_i(o,o|z) > 1$ for $i = 1, 2$, whence $U(o,o|z) > 1$. Therefore $\mathsf{r} < z$.

Finally, if none (X_i, P_i) is $\rho(P_i)$-positive-recurrent then we know from Exercise 3.71 that $\mathsf{r}_i = \mathsf{s}_i$. With z_0 as above, if $z > z_0 = \min\{\mathsf{s}_1, \mathsf{s}_2\}$, then at least one of the power series $U_1(o,o|z)$ and $U_2(o,o|z)$ diverges, so that certainly $U(o,o|z) > 1$. Again by Proposition 2.28, $\mathsf{r} < z$. Therefore $\mathsf{r} = z_0$, which proves the last statement of the lemma. □

5.57 Proposition. *If P on $X_1 \cup X_2$ with $X_1 \cap X_2 = \{o\}$ is defined as in (5.55), then* $\mathrm{BMC}(X, P, \mu)$ *is strongly recurrent if and only if* $\mathrm{BMC}(X_i, P_i, \mu)$ *is strongly recurrent for $i = 1, 2$.*

Proof. Suppose first that $\mathrm{BMC}(X_i, P_i, \mu)$ is strongly recurrent for $i = 1, 2$. Then $\mathsf{Pr}_y^{\mathrm{BMC}}[Z_u^i = o \text{ for some } u \in \mathsf{T} \setminus \{\epsilon\}] = 1$ for each $y \in X_i$, where $\big(Z_u^i\big)_{u\in\mathsf{T}}$ is of course $\mathrm{BMC}(X_i, P_i, \mu)$. Now, within $X_i \setminus \{o\}$, $\mathrm{BMC}(X, P, \mu)$ and $\mathrm{BMC}(X_i, P_i, \mu)$ evolve in the same way. Therefore

$$\mathsf{Pr}_y^{\mathrm{BMC}}[Z_u^i = o \text{ for some } u \in \mathsf{T} \setminus \{\epsilon\}] = \mathsf{Pr}_y^{\mathrm{BMC}}[Z_u = o \text{ for some } u \in \mathsf{T} \setminus \{\epsilon\}]$$

for all $y \in X_i \setminus \{o\}$, $i = 1, 2$. That is, the second criterion of Lemma 5.50 is satisfied for $\mathrm{BMC}(X, P, \mu)$.

Conversely, suppose that for at least one $i \in \{1, 2\}$, $\mathrm{BMC}(X_i, P_i, \mu)$ is not strongly recurrent. Then, once more by Lemma 5.50, there is $y \in X_i \setminus \{o\}$ such that $\mathsf{Pr}_y^{\mathrm{BMC}}[Z_u^i \neq o \text{ for all } u \in \mathsf{T}] > 0$. Again, this probability coincides with $\mathsf{Pr}_y^{\mathrm{BMC}}[Z_u \neq o \text{ for all } u \in \mathsf{T}]$, and $\mathrm{BMC}(X, P, \mu)$ cannot be strongly recurrent. □

We now can construct a simple example where all three phases can occur.

5.58 Example. Let X_1 and X_2 be two copies of the additive group $\mathbb{Z}$. Let P_1 (on X_1) and P_2 (on X_2) be infinite drunkards' walks as in Example 3.5 with the parameters p_i and $q_i = 1 - p_i$ for $i = 1, 2$. We assume that $\frac{1}{2} < p_1 < p_2$. Thus $\rho(P_i) = \sqrt{4p_iq_i}$, and $\rho(P_1) > \rho(P_2)$. We can use (3.6) to see that these two random walks are ρ-null-recurrent. We now connect the two walks at $o = 0$ as in (5.55). As above, we write (X, P) for the resulting Markov chain. Its graph looks like an infinite cross, that is, four half-lines emanating from o. The outgoing probabilities at o are $\alpha_1 p_1$ in direction East, $\alpha_1 q_1$ in direction West, $\alpha_2 p_2$ in direction North and $\alpha_2 q_2$ in direction South. Along the horizontal line of the cross, all other Eastbound transition probabilities (to the next neighbour) are p_1 and all Westbound probabilities are q_1. Analogously, along the vertical line of the cross, all Northbound transition probabilities (except those at o) are p_2 and all Southbound transition probabilities are q_2. The reader is invited to draw a figure.

By Lemma 5.56, $\rho(P) = \rho(P_1)$. We conclude: in our example, $\mathrm{BMC}(X, P, \mu)$ is

- transient if and only if $\bar{\mu} \leq 1\big/\sqrt{4p_1q_1}$,
- recurrent if and only if $\bar{\mu} > 1\big/\sqrt{4p_1q_1}$,
- strongly recurrent if and only if $\mu(0) = 0$ and $\bar{\mu} > 1\big/\sqrt{4p_2q_2}$.

Further examples can be obtained from arbitrary random walks on countable groups. One can proceed as in the last example, using the important fact that such a random walk can never be ρ-positive recurrent. This goes back to a theorem of Guivarc'h [29], compare with [W2, Theorem 7.8].

The properties of the branching Markov chain (X, P, μ) give us the possibility to give probabilistic interpretations of the Green function of the Markov chain (X, P), as well as of ρ-recurrence and -transience. We subsume.

For real $z > 0$, consider $\mathrm{BMC}(X, P, \mu)$ with initial point x, where the offspring distribution μ has mean $\bar{\mu} = z$. By Exercise 5.41, if $z > 0$, the Green function of the Markov chain (X, P),

$$G(x, y|z) = \sum_{n=0}^{\infty} \mathsf{E}_x^{\mathrm{BMC}}(M_n^y) = \mathsf{E}_x^{\mathrm{BMC}}(M^y),$$

is the expected number of occupancies of the site $y \in X$ during the lifetime of the branching Markov chain. Also, we know from Lemma 5.46 that

$$U(x, x|z) = \mathsf{E}_x^{\mathrm{BMC}}(Y_1^x)$$

is the average offspring number of in the embedded Galton–Watson process $Y_n^x = |\mathsf{W}_n^x|$, where (recall) W_n^x consists of all elements u in T with the property that along the shortest path in the tree T from ϵ to u, the n-th return to site x occurs at u.

Clearly $\mathsf{r} = 1/\rho(P)$ is the maximum value of $z = \bar{\mu}$ for which $\mathrm{BMC}(X, P, \mu)$ is transient, or, equivalently, the process (Y_n^x) dies out almost surely.

Now consider in particular $\mathrm{BMC}(X, P, \mu)$ with $\bar{\mu} = \mathsf{r}$. If the Markov chain (X, P) is ρ-transient, then the embedded process (Y_n^x) is sub-critical: its average offspring number is < 1, and we do not only have $\mathsf{Pr}_x^{\mathrm{BMC}}[M^y < \infty] = 1$, but also $\mathsf{E}_x^{\mathrm{BMC}}(M^y) < \infty$ for all $x, y \in X$. In the ρ-recurrent case, (Y_n^x) is critical: its average offspring number is $= 1$, and while $\mathsf{Pr}_x^{\mathrm{BMC}}[M^y < \infty] = 1$, the expected number of occupancies of y is $\mathsf{E}_x^{\mathrm{BMC}}(M^y) = \infty$ for all $x, y \in X$. We can also consider the expected height in T of an element in the first generation W_1^x of the embedded process. By a straightforward adaptation of (5.48), this is

$$\mathsf{E}_x^{\mathrm{BMC}}\Big(\sum_{u \in \Sigma^*} |u| \, \mathbf{1}_{[u \in \mathsf{W}_1^x]} \Big) = \mathsf{r}\, U'(x, x|\mathsf{r}-).$$

It is finite when (X, P) is ρ-positive recurrent, and infinite when (X, P) is ρ-null-recurrent.

Chapter 6
Elements of the potential theory of transient Markov chains

A Motivation. The finite case

At the centre of classical potential theory stands the Laplace equation

$$\Delta f = 0 \quad \text{on } \mathcal{O} \subset \mathbb{R}^d, \tag{6.1}$$

where Δ is the Laplace operator on $\mathbb{R}^d$ and $\mathcal{O}$ is a relatively compact open domain. A typical problem is to find a function $f \in C(\mathcal{O}^-)$, twice differentiable and solution of (6.1) in $\mathcal{O}$, which satisfies

$$f|_{\partial\mathcal{O}} = g \in C(\partial\mathcal{O}) \tag{6.2}$$

("Dirichlet problem"). The function g represents the boundary data.

Let us consider the simple example where the dimension is $d = 2$ and $\mathcal{O}$ is the interior of the square whose vertices are the points $(-1,-1)$, $(1,-1)$, $(1,1)$ and $(-1,1)$. A typical method for approximating the solution of the problem (6.1)–(6.2) consists in subdividing the square by a partition of its sides in $2n$ pieces of length $\tau = 1/n$; see Figure 18.

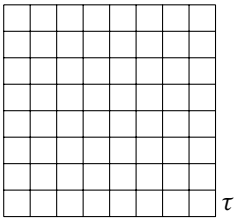

Figure 18

For the second order partial derivatives we then substitute the symmetric differences

$$\frac{\partial^2 f}{\partial x^2}(x,y) \approx \frac{f(x+\tau,y) - 2f(x,y) + f(x-\tau,y)}{\tau^2},$$
$$\frac{\partial^2 f}{\partial y^2}(x,y) \approx \frac{f(x,y+\tau) - 2f(x,y) + f(x,y-\tau)}{\tau^2}.$$

We then write down the difference equation obtained in this way from the Laplace equation in the points

$$(x, y) = (i\tau, j\tau), \quad i, j = -n+1, \dots, 0, \dots, n-1:$$
$$\frac{f\big((i+1)\tau, j\tau\big) - 2f\big(i\tau, j\tau\big) + f\big((i-1)\tau, j\tau\big)}{\tau^2} + \frac{f\big(i\tau, (j+1)\tau\big) - 2f\big(i\tau, j\tau\big) + f\big(i\tau, (j-1)\tau\big)}{\tau^2} = 0.$$

Setting $h(i, j) = f(i\tau, j\tau)$ and dividing by 4, we get

$$\frac{1}{4}\big(h(i+1, j) + h(i-1, j) + h(i, j+1) + h(i, j-1)\big) - h(i, j) = 0, \quad (6.3)$$

that is, $h(i, j)$ must coincide with the arithmetic average of the values of h at the points which are neighbours of (i, j) in the square grid. As i and j vary, this becomes a system of $4(n-1)^2$ linear equations in the unknown variables $h(i, j)$, $i, j = -n+1, \dots, 0, \dots, n-1$; the function g on $\partial\mathcal{O}$ yields the prescribed values

$$h(\pm n, j) = g(\pm 1, j\tau) \quad \text{and} \quad h(i, \pm n) = g(i\tau, \pm 1). \quad (6.4)$$

This system can be resolved by various methods (and of course, the original partial differential equation can be solved by the classical method of separation of the variables). The observation which is of interest for us is that our equations are linked with simple random walk on $\mathbb{Z}^2$, see Example 4.63. Indeed, if P is the transition matrix of the latter, then we can rewrite (6.3) as

$$h(i, j) = Ph(i, j), \quad i, j = -n+1, \dots, 0, \dots, n-1,$$

where the action of P on functions is defined by (3.16). This suggests a strong link with Markov chain theory and, in particular, that the solution of the equations (6.3)–(6.4) can be found as well as interpreted probabilistically.

Harmonic functions and Dirichlet problem for finite Markov chains

Let (X, P) be a finite, irreducible Markov chain (that is, X is finite). We choose and fix a subset $X^o \subset X$, which we call the *interior*, and its complement $\partial X = X \setminus X^o$, the *boundary,* both non-empty. We suppose that X^o is "connected" in the sense that $P_{X^o} = \big(p(x, y)\big)_{x,y\in X^o}$ – the restriction of P to X^o in the sense of Definition 2.14 – is irreducible. (This means that the subgraph of $\Gamma(P)$ induced by X^o is strongly connected; for any pair of points $x, y \in X^0$ there is an oriented path from x to y whose points all lie in X^o.)

We call a function $h\colon X \to \mathbb{R}$ *harmonic on* X^o, if $h(x) = Ph(x)$ for every $x \in X^o$, where (recall) $Ph(x) = \sum_{y\in X} p(x, y)h(y)$. As in Chapter 4, our "Laplace

operator" is $P - I$, acting on functions as the product of a matrix with column vectors. Harmonicity has become a *mean value property*: in each $x \in X^o$, the value $h(x)$ is the weighted mean of the values of h, computed with the weights $p(x, y)$, $y \in X$.

We denote by $\mathcal{H}(X^o) = \mathcal{H}(X^o, P)$ the linear space of all functions on X which are harmonic on X^o. Later on, we shall encounter the following in a more general context. We have already seen it in the proof of Theorem 3.29.

6.5 Lemma (Maximum principle). *Let $h \in \mathcal{H}(X^o)$ and $M = \max_X h(x)$. Then there is $y \in \vartheta X$ such that $h(y) = M$.*

If h is non-constant then $h(x) < M$ for every $x \in X^o$.

Proof. We modify the transition matrix P by setting

$$\begin{aligned} \tilde{p}(x, y) &= p(x, y), && \text{if } x \in X^o,\ y \in X, \\ \tilde{p}(x, x) &= 1, && \text{if } x \in \vartheta X, \\ \tilde{p}(x, y) &= 0, && \text{if } x \in \vartheta X \text{ and } y \neq x. \end{aligned}$$

We obtain a new transition matrix $\widetilde{P}$, with one non-essential class X^o and all points in ϑX as absorbing states. ("We have made all elements of ϑX absorbing.") Indeed, by our assumptions, also with respect to $\widetilde{P}$

$$x \to y \quad \text{for all } x \in X^o,\ y \in X.$$

We observe that $h \in \mathcal{H}(X^o, P)$ if and only if $h(x) = \widetilde{P}h(x)$ for *all* $x \in X$ (not only those in ϑX). In particular, $\widetilde{P}^n h(x) = h(x)$ for every $x \in X$.

Suppose that there is $x \in X^o$ with $h(x) = M$. Take any $x' \in X$. Then $\tilde{p}^{(n)}(x, x') > 0$ for some n. We get

$$\begin{aligned} M = h(x) &= \tilde{p}^{(n)}(x, x')\, h(x') + \sum_{y \neq x'} \tilde{p}^{(n)}(x, y)\, h(y) \\ &\leq \tilde{p}^{(n)}(x, x')\, h(x') + \sum_{y \neq x'} \tilde{p}^{(n)}(x, y)\, M \\ &= \tilde{p}^{(n)}(x, x')\, h(x') + \big(1 - \tilde{p}^{(n)}(x, x')\big) M, \end{aligned}$$

whence $h(x') \geq M$. Since M is the maximum, $h(x') = M$. Thus, h must be constant.

In particular, if h is non-constant, it cannot assume its maximum in X^o. □

Let $\boldsymbol{s} = \boldsymbol{s}^{\vartheta X}$ be the hitting time of ϑX, see (1.26):

$$\boldsymbol{s}^{\vartheta X} = \inf\{n \geq 0 : Z_n \in \vartheta X\},$$

for the Markov chain (Z_n) associated with P, see (1.26). Note that substituting P with $\widetilde{P}$, as defined in the preceding proof, does not change $\boldsymbol{s}$. Corollary 2.9, applied to $\widetilde{P}$, yields that $\mathsf{Pr}_x[\boldsymbol{s}^{\vartheta X} < \infty] = 1$ for every $x \in X$. We set

$$\nu_x(y) = \mathsf{Pr}_x[\boldsymbol{s} < \infty,\ Z_{\boldsymbol{s}} = y], \quad y \in \vartheta X. \tag{6.6}$$

Then, for each $x \in X$, the measure ν_x is a probability distribution on ϑX, called the *hitting distribution* of ϑX.

6.7 Theorem (Solution of the Dirichlet problem). *For every function $g\colon \vartheta X \to \mathbb{R}$ there is a unique function $h \in \mathcal{H}(X^o, P)$ such that $h(y) = g(y)$ for all $y \in \vartheta X$. It is given by*

$$h(x) = \int_{\vartheta X} g \, d\nu_x.$$

Proof. (1) For fixed $y \in \vartheta X$,

$$x \mapsto \nu_x(y) \quad (x \in X)$$

defines a harmonic function. Indeed, if $x \in X^o$, we know that $\boldsymbol{s} \geq 1$, and applying the Markov property,

$$\begin{aligned}
\nu_x(y) &= \sum_{w \in X} \mathsf{Pr}_x[Z_1 = w,\ \boldsymbol{s} < \infty,\ Z_{\boldsymbol{s}} = y] \\
&= \sum_{w \in X} p(x,w)\ \mathsf{Pr}_x[\boldsymbol{s} < \infty,\ Z_{\boldsymbol{s}} = y \mid Z_1 = w] \\
&= \sum_{w \in X} p(x,w)\ \mathsf{Pr}_w[\boldsymbol{s} < \infty,\ Z_{\boldsymbol{s}} = y] \\
&= \sum_{w \in X} p(x,w)\, \nu_w(y).
\end{aligned}$$

Thus,

$$h(x) = \int_{\vartheta X} g \, d\nu_x = \sum_{y \in \vartheta X} g(y)\, \nu_x(y)$$

is a convex combination of harmonic functions. Therefore $h \in \mathcal{H}(X^o, P)$. Furthermore, for $y \in \vartheta X$

$$\nu_y(y) = 1 \quad \text{and} \quad \nu_y(y') = 0 \text{ for all } y' \in \vartheta X,\ y' \neq y.$$

We see that $h(y) = g(y)$ for every $y \in \vartheta X$.

(2) Having found the harmonic extension of g to X^o, we have to show its uniqueness. Let h' be another harmonic function which coincides with g on ϑX. Then $h' - h$ is harmonic, and

$$h(y) - h'(y) = 0 \quad \text{for all } y \in \vartheta X.$$

By the maximum principle, $h - h' \leq 0$. Analogously $h' - h \leq 0$. Therefore h' and h coincide. □

From the last theorem, we see that the potential theoretic task to solve the Dirichlet problem has a probabilistic solution in terms of the hitting distributions. This will be the leading viewpoint in the present chapter, namely, to develop some elements of the potential theory associated with a stochastic transition matrix P and the associated Laplace operator $P - I$ under the viewpoint of its probabilistic interpretation.

We return to the Dirichlet problem for finite Markov chains, i.e., chains with finite state space. Above, we have adopted our specific hypotheses on X^o and ϑX only in order to clarify the analogy with the continuous setting. For a general finite Markov chain, not necessarily irreducible, we define the linear space of harmonic functions on X

$$\mathcal{H} = \mathcal{H}(X, P) = \{h\colon X \to \mathbb{R} \mid h(x) = Ph(x) \text{ for all } x \in X\}.$$

Then we have the following.

6.8 Theorem. *Let (X, P) be a finite Markov chain, and denote its essential classes by C_i, $i \in I = \{1, \dots, m\}$.*

(a) *If h is harmonic on X, then h is constant on each C_i.*

(b) *For each function $g\colon I \to \mathbb{R}$ there is a unique function $h \in \mathcal{H}(X, P)$ such that for all $i \in I$ and $x \in C_i$ one has $h(x) = g(i)$.*

Proof. (a) Let $M_i = \max_{C_i} h$, and let $x \in C_i$ such that $h(x) = M_i$. As in the proof of Lemma 6.5, if $x' \in C_i$, we choose n with $p^{(n)}(x, x') > 0$. Then

$$M_i = h(x) = P^n h(x) \leq \big(1 - p^{(n)}(x, x')\big) M_i + p^{(n)}(x, x') h(x'),$$

and $h(x') \geq M_i$. Hence $h(x') = M_i$.

(b) Let

$$\boldsymbol{s} = \boldsymbol{s}^{X_{\mathrm{ess}}} = \inf\{n \geq 0 : Z_n \in X_{\mathrm{ess}}\},$$

where $X_{\mathrm{ess}} = C_1 \cup \dots \cup C_m$. By Corollary 2.9, we have $\mathsf{Pr}_x[\boldsymbol{s} < \infty] = 1$ for each x. Therefore

$$\nu_x(i) = \mathsf{Pr}_x[\boldsymbol{s} < \infty,\ Z_{\boldsymbol{s}} \in C_i] \tag{6.9}$$

defines a probability distribution on I. As above,

$$h(x) = \sum_{i \in I} g(i) \nu_x(i)$$

defines the unique harmonic function on X with value $g(i)$ on C_i, $i \in I$. We leave the details as an exercise to the reader. □

Note, for the finite case, the analogy with Corollary 3.23 concerning the stationary probability measures.

6.10 Exercise. Elaborate the details from the end of the last proof, namely, that $h(x) = \sum_{i\in I} g(i)\nu_x(i)$ is the unique harmonic function on X with value $g(i)$ on C_i. □

B Harmonic and superharmonic functions. Invariant and excessive measures

In Theorem 6.8 we have described completely the harmonic functions in the finite case. From now on, our focus will be on the infinite case, but most of the results will be valid also when X is finite. However, we shall work under the following restriction.

We always assume that P is irreducible on X.[1]

We do not specify any subset of X as a "boundary": in the infinite case, the boundary will be a set of new points, to be added to X "at infinity". We shall also admit the situation when P is a *substochastic* matrix. In this case, the measures Pr_x ($x \in X$) on the trajectory space, as constructed in Section 1.B are no more probability measures. In order to correct this defect, we can add an absorbing state † to X. We extend the transition probabilities to $X \cup \{\dagger\}$:

$$p(\dagger,\dagger) = 1 \quad \text{and} \quad p(x,\dagger) = 1 - \sum_{y\in X} p(x,y),\ x \in X. \tag{6.11}$$

Now the measures on the trajectory space of $X \cup \{\dagger\}$, which we still denote by Pr_x ($x \in X$), become probability measures. We can think of † as a "tomb": in any state x, the Markov chain (Z_n) may "die" ($\equiv$ be absorbed by †) with probability $p(x,\dagger)$.

We add the state † to X only when the matrix P is strictly substochastic in some x, that is, $\sum_y p(x,y) < 1$. From now on, speaking of the trajectory space $(\Omega, \mathcal{A}, \mathsf{Pr}_x)$, $x \in X$, this will refer to (X, P), when P is stochastic, and to $(X \cup \{\dagger\}, P)$, otherwise.

Harmonic and superharmonic functions

All functions $f : X \to \mathbb{R}$ considered in the sequel are supposed to be *P-integrable:*

$$\sum_{y\in X} p(x,y)\,|f(y)| < \infty \quad \text{for all } x \in X. \tag{6.12}$$

[1] Of course, potential and boundary theory of non-irreducible chains are also of interest. Here, we restrict the exposition to the basic case.

In particular, (6.12) holds for *every* function when P has *finite range*, that is, when $\{y \in X : p(x,y) > 0\}$ is finite for each x.

As previously, we define the *transition operator* $f \mapsto Pf$,

$$Pf(x) = \sum_{y \in X} p(x,y) f(y).$$

We repeat that our discrete analogue of the Laplace operator is $P - I$, where I is the identity operator. We also repeat the definition of harmonic functions.

6.13 Definition. A real function h on X is called *harmonic* if $h(x) = Ph(x)$, and *superharmonic* if $h(x) \geq Ph(x)$ for every $x \in X$.

We denote by

$$\begin{aligned} \mathcal{H} = \mathcal{H}(X,P) &= \{h\colon X \to \mathbb{R} \mid Ph = h\}, \\ \mathcal{H}^+ &= \{h \in \mathcal{H} \mid h(x) \geq 0 \text{ for all } x \in X\} \quad \text{and} \\ \mathcal{H}^\infty &= \{h \in \mathcal{H} \mid h \text{ is bounded on } X\} \end{aligned} \tag{6.14}$$

the linear space of all harmonic functions, the cone of the non-negative harmonic functions and the space of bounded harmonic functions. Analogously, we define $\mathcal{S} = \mathcal{S}(X,P)$, the space of all superharmonic functions, $\mathcal{S}^+$ and $\mathcal{S}^\infty$. (Note that $\mathcal{S}$ is not a linear space.)

The following is analogous to Lemma 6.5. We assume of course irreducibility and that $|X| > 1$.

6.15 Lemma (Maximum principle). *If $h \in \mathcal{H}$ and there is $x \in X$ such that $h(x) = M = \max_X h$, then h is constant. Furthermore, if $M \neq 0$, then P is stochastic.*

Proof. We use irreducibility. If $x' \in X$ and $p^{(n)}(x,x') > 0$, then as in the proof of Lemma 6.5,

$$\begin{aligned} M = h(x) &\leq \sum_{y \neq x'} p^{(n)}(x,y)\, M + p^{(n)}(x,x')\, h(x') \\ &\leq \big(1 - p^{(n)}(x,x')\big)\, M + p^{(n)}(x,x')\, h(x') \end{aligned}$$

where in the second inequality we have used substochasticity. As above it follows that $h(x') = M$. In particular, the constant function $h \equiv M$ is harmonic. Thus, if $M \neq 0$, then the matrix P must be stochastic, since X has more than one element. □

6.16 Exercise. Deduce the following in at least two different ways.
If X is finite and P is irreducible and strictly substochastic in some point, then $\mathcal{H} = \{0\}$. □

We next exhibit two simple properties of superharmonic functions.

6.17 Lemma. (1) *If $h \in \mathcal{S}^+$ then $P^n h \in \mathcal{S}^+$ for each n, and either $h \equiv 0$ or $h(x) > 0$ for every x.*

(2) *If h_i, $i \in I$, is a family of superharmonic functions and $h(x) = \inf_I h_i(x)$ defines a P-integrable function, then also h is superharmonic.*

Proof. (1) Since $0 \le Ph \le h$, the P-integrability of h implies that of Ph and, inductively, also of $P^n h$. Furthermore, the transition operator is monotone: if $f \le g$ then $Pf \le Pg$. In particular, $P^n h \le h$.

Suppose that $h(x) = 0$ for some x. Then for each n,

$$0 = h(x) \ge \sum_y p^{(n)}(x, y) h(y).$$

Since $h \ge 0$, we must have $h(y) = 0$ for every y with $x \xrightarrow{n} y$. Irreducibility implies $h \equiv 0$.

(2) By monotonicity of P, we have $Ph \le Ph_i \le h_i$ for every $i \in I$. Therefore $Ph \le \inf_I h_i = h$. □

6.18 Exercise. Show that in statement (2) of Lemma 6.17, P-integrability of $h = \inf_I h_i$ follows from P-integrability of the h_i, if • the set I is finite, or if • the h_i are uniformly bounded below (e.g., non-negative). □

In the case when P is strictly substochastic in some state x, the elements of X cannot be recurrent. Indeed, if we pass to the stochastic extension of P on $X \cup \{\dagger\}$, we know that the irreducible class X is non-essential, whence non-recurrent by Theorem 3.4 (b).

In general, in the transient (irreducible) case, there is a fundamental family of functions in $\mathcal{S}^+$:

6.19 Lemma. *If (X, P) is transient, then for each $y \in X$, the function $G(\cdot, y)$, defined by $x \mapsto G(x, y)$, is superharmonic and positive. There is at most one $y \in X$ for which $G(\cdot, y)$ is a constant function. If P is stochastic, then $G(\cdot, y)$ is non-constant for every y.*

Proof. We know from (1.34) that

$$PG(\cdot, y) = G(\cdot, y) - \mathbf{1}_y.$$

Therefore $G(\cdot, y) \in \mathcal{S}^+$.

Suppose that there are $y_1, y_2 \in X$, $y_1 \ne y_2$, such that the functions $G(\cdot, y_i)$ are constant. Then, by Theorem 1.38 (b)

$$F(y_1, y_2) = \frac{G(y_1, y_2)}{G(y_2, y_2)} = 1 \quad \text{and} \quad F(y_2, y_1) = \frac{G(y_2, y_1)}{G(y_1, y_1)} = 1.$$

Now Proposition 1.43 (a) implies $F(y_1, y_1) \geq F(y_1, y_2)F(y_2, y_1) = 1$, and y_1 is recurrent, a contradiction.

Finally, if P is stochastic, then every constant function is harmonic, while $G(\cdot, y)$ is strictly subharmonic at y, so that it cannot be constant. □

6.20 Exercise. Show that $G(\cdot, y)$ is constant for the substochastic Markov chain illustrated in Figure 19. □

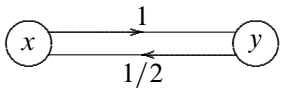

Figure 19

The following is the fundamental result in this section. (Recall our assumption of irreducibility and that $|X| > 1$, while P may be substochastic.)

6.21 Theorem. *(X, P) is recurrent if and only if every non-negative superharmonic function is constant.*

Proof. a) Suppose that (X, P) is recurrent.

First step. We show that $\mathcal{S}^+ = \mathcal{H}^+$:

Let $h \in \mathcal{S}^+$. We set $g(x) = h(x) - Ph(x)$. Then g is non-negative and P-integrable. We have

$$\sum_{k=0}^{n} P^k g(x) = \sum_{k=0}^{n} \big(P^k h(x) - P^{k+1} h(x)\big) = h(x) - P^{n+1} h(x).$$

Suppose that $g(y) > 0$ for some y. Then

$$\sum_{k=0}^{n} p^{(k)}(x, y)\, g(y) \leq \sum_{k=0}^{n} P^k g(x) \leq h(x)$$

for each n, and

$$G(x, y) \leq h(x)/g(y) < \infty,$$

a contradiction. Thus $g \equiv 0$, and h is harmonic. In particular, substochasticity implies that the constant function $\mathbf{1}$ is superharmonic, whence harmonic, and P must be stochastic. (We know this already from the fact that otherwise, X is a non-essential class in $X \cup \{\dagger\}$.)

Second step. Let $h \in \mathcal{S}^+ = \mathcal{H}^+$, and let $x_1, x_2 \in X$. We set $M_i = h(x_i)$, $i = 1, 2$. Then $h_i(x) = \min\{h(x), M_i\}$ is a superharmonic function by Lemma 6.17, hence harmonic by the first step. But h_i assumes its maximum M_i in x_i and must be

constant by the maximum principle (Lemma 6.15): $\min\{h(x), M_i\} = M_i$ for all x. This yields

$$h(x_1) = \min\{h(x_1), h(x_2)\} = h(x_2),$$

and h is constant.

b) Conversely, suppose that $\mathcal{S}^+ = \{\text{constants}\}$. Then, by Lemma 6.19, (X, P) cannot be transient: otherwise there is $y \in X$ for which $G(\cdot, y) \in \mathcal{S}^+$ is non-constant. □

6.22 Exercise. The definition of harmonic and superharmonic functions does of course not require irreducibility.

(a) Show that when P is substochastic, not necessarily irreducible, then the function $F(\cdot, y)$ is superharmonic for each y.

(b) Show that (X, P) is irreducible if and only if every non-negative, non-zero superharmonic function is strictly positive in each point.

(c) Assume in addition that P is stochastic. Show the following. If a superharmonic function attains its minimum in some point x then it has the same value in every y with $x \to y$.

(d) Show for stochastic P that irreducibility is equivalent with the *minimum principle* for superharmonic functions: if a superharmonic function attains a minimum in some point then it is constant. □

Invariant and excessive measures

As above, we suppose that (X, P) is irreducible, $|X| \geq 2$ and P substochastic. We continue, with the proof of Theorem 6.26, the study of invariant measures initiated in Section 3.B. Recall that a measure ν on X is given as a row vector $\big(\nu(x)\big)_{x\in X}$. Here, we consider only non-negative measures. In analogy with P-integrability of functions, we allow only measures which satisfy

$$\nu P(y) = \sum_{x\in X} \nu(x)p(x, y) < \infty \quad \text{for all } y \in X. \tag{6.23}$$

The action of the transition operator is multiplication with P on the right: $\nu \mapsto \nu P$. We recall from Definition 3.17 that a measure ν on X is called *invariant* or *stationary*, if $\nu = \nu P$. Furthermore, ν is called *excessive* or *superinvariant*, if $\nu(y) \geq \nu P(y)$ for every $y \in X$. We denote by $\mathcal{I}^+ = \mathcal{I}^+(X, P)$ and $\mathcal{E}^+ = \mathcal{E}^+(X, P)$ the cones of all invariant and superinvariant measures, respectively.

Theorem 3.19 and Corollary 3.23 describe completely the invariant measures in the case when X is finite and P stochastic, not necessarily irreducible. On the other hand, if X is finite and P irreducible, but strictly substochastic in some point, then the unique invariant measure is $\nu \equiv 0$. In fact, in this case, there are no harmonic

functions $\not\equiv 0$ (see Exercise 6.16). In other words, the matrix P does not have $\lambda = 1$ as an eigenvalue.

The following is analogous to Lemmas 6.17 and 6.19.

6.24 Exercise. Prove the following.

(1) If $\nu \in \mathcal{E}^+$ then $\nu P^n \in \mathcal{E}^+$ for each n, and either $\nu \equiv 0$ or $\nu(x) > 0$ for every x.

(2) If ν_i, $i \in I$, is a family of excessive measures, then also $\nu(x) = \inf_I \nu_i(x)$ is excessive.

(3) If (X, P) is transient, then for each $x \in X$, the measure $G(x, \cdot)$, defined by $y \mapsto G(x, y)$, is excessive. □

Next, we want to know whether there also are excessive measures in the recurrent case. To this purpose, we recall the "last exit" probabilities $\ell^{(n)}(x, y)$ and the associated generating function $L(x, y|z)$ defined in (3.56) and (3.57), respectively. We know from Lemma 3.58 that

$$L(x, y) = \sum_{n=0}^{\infty} \ell^{(n)}(x, y) = L(x, y|1),$$

the expected number of visits in y before returning to x, is finite. Setting $z = 1$ in the second and third identities of Exercise 3.59, we get the following.

6.25 Corollary. *In the recurrent as well as in the transient case, for each $x \in X$, the measure $L(x, \cdot)$, defined by $y \mapsto L(x, y)$, is finite and excessive.*

Indeed, in Section 3.F, we have already used the fact that $L(x, \cdot)$ is invariant in the recurrent case.

6.26 Theorem. *Let (X, P) be substochastic and irreducible. Then (X, P) is recurrent if and only if there is a non-zero invariant measure ν such that each excessive measure is a multiple of ν, that is*

$$\mathcal{E}^+(X, P) = \{c \cdot \nu : c \geq 0\}.$$

In this case, P must be stochastic.

Proof. First, assume that P is recurrent. Then we know e.g. from Theorem 6.21 that P must be stochastic (since constant functions are harmonic). We also know, from Corollary 6.25, that there is an excessive measure ν satisfying $\nu(y) > 0$ for all y. (Take $\nu = L(x, \cdot)$ for some x.) We construct the ν-reversal $\widehat{P}$ of P as in (3.30) by

$$\hat{p}(x, y) = \nu(y) p(y, x)/\nu(x). \tag{6.27}$$

Excessivity of ν yields that $\widehat{P}$ is substochastic. Also, it is straightforward to prove that

$$\hat{p}^{(n)}(x, y) = \nu(y) p^{(n)}(y, x)/\nu(x).$$

Summing over n, we see that also $\widehat{P}$ is recurrent, whence stochastic. Thus, ν must be invariant. If σ is any other excessive measure, and we define the function $h(x) = \sigma(x)/\nu(x)$, then as in (3.31), we find that $\widehat{P}h \le h$. By Theorem 6.21, h must be constant, that is, $\sigma = c \cdot \nu$ for some $c \ge 0$.

To prove the converse implication, we just observe that in the transient case, the measure $\sigma = G(x, \cdot)$ satisfies $\sigma\, P = \sigma - \delta_x$. It is excessive, but not invariant. □

C Induced Markov chains

We now introduce and study an important probabilistic notion for Markov chains, whose relevance for potential theoretic issues will become apparent in the next sections.

Suppose that (X, P) is irreducible and substochastic. Let A be an arbitrary non-empty subset of X. The hitting time $\boldsymbol{t}^A = \inf\{n > 0 : Z_n \in A\}$ defined in (1.26) is not necessarily a.s. finite. We define

$$p^A(x, y) = \mathsf{Pr}_x[\boldsymbol{t}^A < \infty,\ Z_{\boldsymbol{t}^A} = y].$$

If $y \notin A$ then $p^A(x, y) = 0$. If $y \in A$,

$$p^A(x, y) = \sum_{n=1}^{\infty} \sum_{x_1, \dots, x_{n-1} \in X \setminus A} p(x, x_1) p(x_1, x_2) \cdots p(x_{n-1}, y). \tag{6.28}$$

We observe that

$$\sum_{y \in A} p^A(x, y) = \mathsf{Pr}_x[\boldsymbol{t}^A < \infty] \le 1.$$

In other words, the matrix $P^A = \big(p^A(x, y)\big)_{x,y \in A}$ is substochastic. The Markov chain (A, P^A) is called the *Markov chain induced by* (X, P) *on* A.

We observe that irreducibility of (X, P) implies irreducibility of the induced chain: for $x, y \in A$ $(x \ne y)$ there are $n > 0$ and $x_1, \dots, x_{n-1} \in X$ such that $p(x, x_1) p(x_1, x_2) \cdots p(x_{n-1}, y) > 0$. Let $i_1 < \cdots < i_{m-1}$ be the indices for which $x_{i_j} \in A$. Then $p^A(x, x_{i_1}) p^A(x_{i_1}, x_{i_2}) \cdots p^A(x_{i_{m-1}}, y) > 0$, and $x \xrightarrow{m} y$ with respect to P^A.

In general, the matrix P^A is not stochastic. If it is stochastic, that is,

$$\mathsf{Pr}_x[\boldsymbol{t}^A < \infty] = 1 \quad \text{for all } x \in A,$$

then we call the set A *recurrent* for (X, P). If the Markov chain (X, P) is recurrent, then every non-empty subset of X is recurrent for (X, P). Conversely, if there exists a *finite* recurrent subset A of X, then (X, P) must be recurrent.

6.29 Exercise. Prove the last statement as a reminder of the methods of Chapter 3. □

On the other hand, even when (X, P) is transient, one can very well have (infinite) proper subsets of X that are recurrent.

6.30 Example. Consider the random walk on the Abelian group $\mathbb{Z}^2$ whose law μ in the sense of (4.18) (additively written) is given by

$$\mu\big((1,0)\big) = p_1,\ \mu\big((0,1)\big) = p_2,\ \mu\big((-1,0)\big) = p_3,\ \mu\big((0,-1)\big) = p_4,$$

and $\mu\big((k,l)\big) = 0$ in all other cases, where $p_i > 0$ and $p_1 + p_2 + p_3 + p_4 = 1$. Thus

$$\begin{aligned} p\big((k,l),(k+1,l)\big) &= p_1, \quad p\big((k,l),(k,l+1)\big) = p_2,\\ p\big((k,l),(k-1,l)\big) &= p_3, \quad p\big((k,l),(k,l-1)\big) = p_4, \end{aligned}$$

$(k,l) \in \mathbb{Z}^2$.

6.31 Exercise. Show that this random walk is recurrent if and only if $p_1 = p_3$ and $p_2 = p_4$. □

Setting

$$A = \{(k,l) \in \mathbb{Z}^2 : k + \ell \text{ is even}\},$$

one sees immediately that A is a recurrent set for any choice of the p_i. Indeed, $\mathsf{Pr}_{(k,l)}[t^A = 2] = 1$ for every $(k,l) \in A$. The induced chain is given by

$$\begin{aligned} p^A\big((k,l),(k+2,l)\big) &= p_1^2, & p^A\big((k,l),(k+1,l+1)\big) &= 2p_1p_2,\\ p^A\big((k,\ell),(k,\ell+2)\big) &= p_2^2, & p^A\big((k,l),(k-1,l+1)\big) &= 2p_2p_3,\\ p^A\big((k,l),(k-2,l)\big) &= p_3^2, & p^A\big((k,l),(k-1,l-1)\big) &= 2p_3p_4,\\ p^A\big((k,l),(k,l-2)\big) &= p_4^2, & p^A\big((k,l),(k+1,l-1)\big) &= 2p_1p_4,\\ p^A\big((k,l),(k,l)\big) &= 2p_1p_3 + 2p_2p_4. \end{aligned}$$

6.32 Example. Consider the infinite drunkard's walk on $\mathbb{Z}$ (see Example 3.5) with parameters p and $q = 1 - p$. The random walk is recurrent if and only if $p = q = 1/2$.

(1) Set $A = \{0, 1, \dots, N\}$. Being finite, the set A is recurrent if and only the random walk itself is recurrent. The induced chain has the following non-zero transition probabilities.

$$\begin{aligned} p^A(k-1,k) = p, \quad p^A(k,k-1) = q \quad (k = 1,\dots,N), \quad \text{and}\\ p^A(0,0) = p^A(N,N) = \min\{p,q\}. \end{aligned}$$

Indeed, if – starting at state N – the first return to A occurs in N, the first step of (Z_n) has to go from N to $N+1$, after which (Z_n) has to return to N. This means that $p^A(N,N) = p\,F(N+1,N)$, and in Examples 2.10 and 3.5, we have computed $F(N+1,N) = F(1,0) = (1-|p-q|)/2p$, leading to $p^A(N,N) = \min\{p,q\}$. By symmetry, $p^A(0,0)$ has the same value.

In particular, if $p > q$, the induced chain is strictly substochastic only at the point N. Conversely, if $p < q$, the only point of strict substochasticity is 0.

(2) Set $A = \mathbb{N}_0$. The transition probabilities of the induced chain coincide with those of the original random walk in each point $k > 0$. Reasoning as above in (1), we find

$$p^{\mathbb{N}_0}(0,1) = p \quad \text{and} \quad p^{\mathbb{N}_0}(0,0) = \min\{p,q\}.$$

We see that the set $\mathbb{N}_0$ is recurrent if and only if $p \geq q$. Otherwise, the only point of strict substochasticity is 0.

6.33 Lemma. *If the set A is recurrent for (X,P) then*

$$\mathsf{Pr}_x[\boldsymbol{t}^A < \infty] = 1 \quad \textit{for all } x \in X.$$

Proof. Factoring with respect to the first step, one has – even when the set A is not recurrent –

$$\mathsf{Pr}_x[\boldsymbol{t}^A < \infty] = \sum_{y\in A} p(x,y) + \sum_{y\in X\setminus A} p(x,y)\ \mathsf{Pr}_y[\boldsymbol{t}^A < \infty] \quad \text{for all } x \in X.$$

(Observe that in case $y = Z_1 \in A$, one has $\boldsymbol{t}^A = 1$.) In particular, if we have $\mathsf{Pr}_y[\boldsymbol{t}^A < \infty] = 1$ for every $y \in A$, then the function $h(x) = \mathsf{Pr}_x[\boldsymbol{t}^A < \infty]$ is harmonic and assumes its maximum value 1. By the maximum principle (Lemma 6.15), $h \equiv 1$. □

Observe that the last lemma generalizes Theorem 3.4 (b) in the irreducible case. Indeed, one may as well introduce the basic potential theoretic setup – in particular, the maximum principle – at the initial stage of developing Markov chain theory and thereby simplify a few of the proofs in the first chapters.

The following is intuitively obvious, but laborious to formalize.

6.34 Theorem. *If $A \subset B \subset X$ then $(P^B)^A = P^A$.*

Proof. Let (Z_n^B) be the Markov chain relative to (B, P^B). It is a random subsequence of the original Markov chain (Z_n), which can be realized on the trajectory space associated with (X,P) (which includes the "tomb" state †). We use the random variable $\boldsymbol{v}^B$ introduced in 1.C (number of visits in B), and define $\boldsymbol{w}_n^B(\omega) = k$, if $n \leq \boldsymbol{v}^B(\omega)$ and k is the instant of the n-th return visit to B. Then

$$Z_n^B = \begin{cases} Z_{\boldsymbol{w}_n^B}, & \text{if } n \leq \boldsymbol{v}^B, \\ \dagger, & \text{otherwise.} \end{cases} \tag{6.35}$$

Let t_B^A be the stopping time of the first visit of (Z_n^B) in A. Since $A \subset B$, we have for every trajectory $\omega \in \Omega$ that $t^A(\omega) = \infty$ if and only if $t_B^A(\omega) = \infty$. Furthermore, $t^A(\omega) \geq t^B(\omega)$. Hence, if $t^A(\omega) < \infty$, then (6.35) implies

$$Z^B_{t_B^A(\omega)}(\omega) = Z_{t^A(\omega)}(\omega),$$

that is, the first return visits in A of (Z_n^B) and of (Z_n) take place at the same point. Consequently, for $x, y \in A$,

$$\begin{aligned}(p^B)^A(x,y) &= \mathsf{Pr}_x[t_B^A < \infty,\ Z^B_{t_B^A} = y] \\ &= \mathsf{Pr}_x[t^A < \infty,\ Z_{t^A} = y] = p^A(x,y). \qquad \square\end{aligned}$$

If A and B are two arbitrary non-empty subsets of X (not necessarily such that one is contained in the other), we define the *restriction* of P to $A \times B$ by

$$P_{A,B} = \big(p(x,y)\big)_{x\in A, y\in B}. \tag{6.36}$$

In particular, if $A = B$, we have $P_{A,A} = P_A$, as defined in Definition 2.14. Recall (2.15) and the associated Green function $G_A(x,y)$ for $x, y \in A$, which is finite by Lemma 2.18.

6.37 Lemma. $P^A = P_A + P_{A,X\setminus A}\, G_{X\setminus A}\, P_{X\setminus A,A}$.

Proof. We use formula (6.28), factorizing with respect to the first step. If $x, y \in A$, then the induced chain starting at x and going to y either moves to y immediately, or else exits A and re-enters into A only at the last step, and the re-entrance must occur at y:

$$p^A(x,y) = p(x,y) + \sum_{v\in X\setminus A} p(x,v)\ \mathsf{Pr}_v[t^A < \infty,\ Z_{t^A} = y]. \tag{6.38}$$

We now factorize with respect to the last step, using the Markov property:

$$\begin{aligned}\mathsf{Pr}_v[t^A < \infty,\, Z_{t^A} = y] &= \sum_{w\in X\setminus A} \mathsf{Pr}_v[t^A < \infty,\ Z_{t^A-1} = w,\ Z_{t^A} = y] \\ &= \sum_{w\in X\setminus A} \sum_{n=1}^{\infty} \mathsf{Pr}_v[t^A = n,\ Z_{n-1} = w,\ Z_n = y] \\ &= \sum_{w\in X\setminus A} \sum_{n=1}^{\infty} \mathsf{Pr}_v[Z_n = y,\ Z_{n-1} = w,\ Z_i \notin A \text{ for all } i < n] \\ &= \sum_{w\in X\setminus A} \sum_{n=0}^{\infty} \mathsf{Pr}_v[Z_n = w,\ Z_i \notin A \text{ for all } i \leq n]\, p(w,y) \\ &= \sum_{w\in X\setminus A} G_{X\setminus A}(v,w)\, p(w,y).\end{aligned}$$

[In the last step we have used (2.15).] Thus

$$p^A(x,y) = p(x,y) + \sum_{v\in X\setminus A}\sum_{w\in X\setminus A} p(x,v)\, G_{X\setminus A}(v,w)\, p(w,y). \qquad \square$$

6.39 Theorem. *Let $\nu \in \mathcal{E}^+(X,P)$, $A \subset X$ and ν_A the restriction of ν to A. Then*

$$\nu_A \in \mathcal{E}^+(A, P^A).$$

Proof. Let $x \in A$. Then

$$\nu_A(x) = \nu(x) \geq \nu P(x) = \nu_A\, P_A(x) + \nu_{X\setminus A}\, P_{X\setminus A,A}(x).$$

Hence

$$\nu_A \geq \nu_A\, P_A + \nu_{X\setminus A}\, P_{X\setminus A,A},$$

and by symmetry

$$\nu_{X\setminus A} \geq \nu_{X\setminus A}\, P_{X\setminus A} + \nu_A\, P_{A,X\setminus A}.$$

Applying $\sum_{k=0}^{n-1} P_{X\setminus A}^k$ from the right, the last relation yields

$$\nu_{X\setminus A} \geq \nu_{X\setminus A}\, P_{X\setminus A}^n + \nu_A\, P_{A,X\setminus A}\Big(\sum_{k=0}^{n-1} P_{X\setminus A}^k\Big) \geq \nu_A\, P_{A,X\setminus A}\Big(\sum_{k=0}^{n-1} P_{X\setminus A}^k\Big)$$

for every $n \geq 1$. By monotone convergence,

$$\nu_A\, P_{A,X\setminus A}\Big(\sum_{k=0}^{n-1} P_{X\setminus A}^k\Big) \;\to\; \nu_A\, P_{A,X\setminus A}\, G_{X\setminus A}$$

pointwise, as $n \to \infty$. Therefore

$$\nu_{X\setminus A} \geq \nu_A\, P_{A,X\setminus A}\, G_{X\setminus A}.$$

Combining the inequalities and applying Lemma 6.37,

$$\nu_A \geq \nu_A\, P_A + \nu_A\, P_{A,X\setminus A}\, G_{X\setminus A}\, P_{X\setminus A,A} = \nu_A\, P^A,$$

as proposed $\square$

6.40 Exercise. Prove the "dual" to the above result for superharmonic functions: if $h \in \mathcal{S}^+(X,P)$ and $A \subset X$, then the restriction of h to the set A satisfies $h_A \in \mathcal{S}^+(A, P^A)$. $\square$

D Potentials, Riesz decomposition, approximation

With Theorems 6.21 and 6.26, we have completed the description of all positive superharmonic functions and excessive measures in the recurrent case. Therefore, in the rest of this chapter,

we assume that (X, P) is irreducible and transient.

This means that

$$0 < G(x, y) < \infty \quad \text{for all } x, y \in X.$$

6.41 Definition. A *G-integrable* function $f : X \to \mathbb{R}$ is one that satisfies

$$\sum_y G(x, y)\, |f(y)| < \infty$$

for each $x \in X$. In this case,

$$g(x) = Gf(x) = \sum_{y \in X} G(x, y)\, f(y)$$

is called the *potential* of f, while f is called the *charge* of g.

If we set $f^+(x) = \max\{f(x), 0\}$ and $f^-(x) = \max\{-f(x), 0\}$ then f is G-integrable if and only f^+ and f^- have this property, and $Gf = Gf^+ - Gf^-$. In the sequel, when studying potentials Gf, we shall always assume tacitly G-integrability of f. The *support of* f is, as usual, the set $\operatorname{supp}(f) = \{x \in X : f(x) \neq 0\}$.

6.42 Lemma. (a) *If g is the potential of f, then $f = (I - P)g$. Furthermore, $P^n g \to 0$ pointwise.*

(b) *If f is non-negative, then $g = Gf \in \mathcal{S}^+$, and g is harmonic on $X \setminus \operatorname{supp}(f)$, that is, $Pg(x) = g(x)$ for every $x \in X \setminus \operatorname{supp}(f)$.*

Proof. We may suppose that $f \geq 0$. (Otherwise, decomposing $f = f^+ - f^-$, the extension to the general case is immediate.)

Since all terms are non-negative, convergence of the involved series is absolute, and

$$P\,Gf = G\,Pf = \sum_{n=1}^{\infty} P^n f = Gf - f.$$

This implies the first part of (a) as well as (b). Furthermore

$$P^n g(x) = GP^n f(x) = \sum_{k=n}^{\infty} P^k f(x)$$

is the n-th rest of a convergent series, so that it tends to 0. □

Formally, one has $G = \sum_{n=0}^{\infty} P^n = (I - G)^{-1}$ (geometric series), but – as already mentioned in Section 1.D – one has to pay attention on which space of functions (or measures) one considers G to act as an operator. For the G-integrable functions we have seen that $(I - P)Gf = G(I - P)f = f$. But in general, it is not true that $G(I - P)f = f$, even when $(I - P)f$ is a G-integrable function. For example, if P is stochastic and $f(x) = c > 0$, then $(I - P)f = 0$ and $G(I - P)f = 0 \neq f$.

6.43 Riesz decomposition theorem. *If $u \in \mathcal{S}^+$ then there are a potential $g = Gf$ and a function $h \in \mathcal{H}^+$ such that*

$$u = Gf + h.$$

The decomposition is unique.

Proof. Since $u \geq 0$ and $u \geq Pu$, non-negativity of P implies that for every $x \in X$ and every $n \geq 0$,

$$P^n u(x) \geq P^{n+1} u(x) \geq 0.$$

Therefore, there is the limit function

$$h(x) = \lim_{n \to \infty} P^n u(x).$$

Since $0 \leq h \leq u$ and u is P-integrable, Lebesgue's theorem on dominated convergence implies

$$Ph(x) = P\big(\lim_{n \to \infty} P^n u\big)(x) = \lim_{n \to \infty} P^n (Pu)(x) = \lim_{n \to \infty} P^{n+1} u(x) = h(x),$$

and h is harmonic. We set

$$f = u - Pu.$$

This function is non-negative, and P^k-integrable along with u and Pu. In particular, $P^k f = P^k u - P^{k+1} u$ for every $k \geq 0$:

$$u - P^{n+1} u = \sum_{k=0}^{n} (P^k u - P^{k+1} u) = \sum_{k=0}^{n} P^k f.$$

Letting $n \to \infty$ we obtain

$$u - h = \sum_{k=0}^{\infty} P^k f = Gf = g.$$

This proves existence of the decomposition. Suppose now that $u = g_1 + h_1$ is another decomposition. We have $P^n u = P^n g_1 + h_1$ for every n. By Lemma 6.42 (a), $P^n g_1 \to 0$ pointwise. Hence $P^n u \to h_1$, so that $h_1 = h$. Therefore also $g_1 = g$ and, again by Lemma 6.42 (a), $f_1 = (I - P)g_1 = (I - P)g = f$. □

6.44 Corollary. (1) *If g is a non-negative potential then the only function $h \in \mathcal{H}^+$ with $g \geq h$ is $h \equiv 0$.*

(2) *If $u \in \mathcal{S}^+$ and there is a potential $g = Gf$ with $g \geq u$, then u is the potential of a non-negative function.*

Proof. (1) By Lemma 6.42 (a), we have $h = P^n h \leq P^n g \to 0$ pointwise, as $n \to \infty$.

(2) We write $u = Gf_1 + h_1$ with $h_1 \in \mathcal{H}^+$ and $f_1 \geq 0$ (Riesz decomposition). Then $h_1 \leq g$, and $h_1 \equiv 0$ by (1). □

We now illustrate what happens in the case when the state space is finite.

The finite case

(I) If X is finite and P is irreducible but strictly substochastic in some point, then $\mathcal{H} = \{0\}$, see Exercise 6.16. Consequently every positive superharmonic function is a potential Gf, where $f \geq 0$. In particular, the constant function 1 is superharmonic, and there is a function $\varphi \geq 0$ such that $G\varphi \equiv 1$. Let u be a superharmonic function that assumes negative values. Setting $M = -\min_X u(x) > 0$, the function $x \mapsto u(x) + M$ becomes a non-negative superharmonic function. Therefore every superharmonic function (not necessarily positive) can be written in the form

$$u = G(f - M \cdot \varphi),$$

where $f \geq 0$.

(II) Assume that X is finite and P stochastic. If P is irreducible then (X, P) is recurrent, and all superharmonic functions are constant: indeed, by Theorem 6.21, this is true for non-negative superharmonic functions. On the other hand, the constant functions are harmonic, and every superharmonic function can be written as $u - M$, where M is constant and $u \in \mathcal{S}^+$.

(III) Let us now assume that (X, P) is finite, stochastic, but not irreducible, with the essential classes C_i, $i \in I = \{1, \ldots, m\}$. Consider X_{ess}, the union of the essential classes, and the probability distributions ν_x on I as in (6.9). The set

$$X^o = X \setminus X_{\text{ess}}$$

is assumed to be non-empty. (Otherwise, (X, P) decomposes into a finite number of irreducible Markov chains – the restrictions to the essential classes – which do not communicate among each other, and to each of them one can apply what has been said in (II).)

Let $u \in \mathcal{S}(X, P)$. Then the restriction of u to C_i is superharmonic for P_{C_i}. The Markov chain (C_i, P_{C_i}) is recurrent by Theorem 3.4 (c). If $g(i) = \min_{C_i} u$,

then $u|_{C_i} - g(i) \in \mathcal{S}^+(C_i, P_{C_i})$. Hence $u|_{C_i}$ is constant by Lemma 6.17 (1) or Theorem 6.21,

$$u|_{C_i} \equiv g(i).$$

We set

$$h(x) = \int_I g \, d\nu_x = \sum_{i \in I} g(i)\, \nu_x(i),$$

see Theorem 6.8 (b). Then h is the unique harmonic function on X which satisfies $h(x) = g(i)$ for each $i \in I$, $x \in C_i$, and so $u - h \in \mathcal{S}(X, P)$ and $u(y) - h(y) = 0$ for each $y \in X_{\text{ess}}$. Exercise 6.22 (c) implies that $u - h \geq 0$ on the whole of X. (Indeed, if the minimum of $u - h$ is attained at x then there is $y \in X_{\text{ess}}$ such that $x \to y$, and the minimum is also attained in y.) We infer that $v \geq 0$, and $(u - h)|_{X^o} \in \mathcal{S}^+(X^o, P_{X^o})$.

6.45 Exercise. Deduce that there is a unique function f on X^o such that

$$(u - h)|_{X^o} = G_{X^o} f = Gf,$$

and $f \geq 0$ with $\text{supp}(f) \subset X^o$. (Note here that (X^o, P_{X^o}) is substochastic, but not necessarily irreducible.) □

We continue by observing that $G(x, y) = 0$ for every $x \in X_{\text{ess}}$ and $y \in X^o$, so that we also have $u - h = Gf$ on the whole of X. We conclude that every function $u \in \mathcal{S}^+(X, P)$ can be uniquely represented as

$$u(x) = Gf(x) + \int_I g \, d\nu_x,$$

where f and g are functions on X^0 and I, respectively. □

Let us return to the study of positive superharmonic functions in the case where (X, P) is irreducible, transient, not necessarily stochastic. The following theorem will be of basic importance when X is infinite.

6.46 Approximation theorem. *If $h \in \mathcal{S}^+(X, P)$ then there is a sequence of potentials $g_n = Gf_n$, $f_n \geq 0$, such that $g_n(x) \leq g_{n+1}(x)$ for all x and n, and*

$$\lim_{n \to \infty} g_n(x) = h(x).$$

Proof. Let A be a finite subset of X. We define the *reduced function* of h on A: for $x \in X$,

$$R^A[h](x) = \inf \{u(x) : u \in \mathcal{S}^+,\ u(y) \geq h(y) \text{ for all } y \in A\}.$$

The reduced function is also defined when A is infinite, and – since h is superharmonic –

$$R^A[h] \leq h.$$

In particular, we have

$$R^A[h](x) = h(x) \quad \text{for all } x \in A.$$

Furthermore, $R^A[h] \in \mathcal{S}^+$ by Lemma 6.17(2). Let $f_0(x) = h(x)$, if $x \in A$, and $f_0(x) = 0$, otherwise. f_0 is non-negative and finitely supported. (It is here that we use the assumption of finiteness of A for the first time.) In particular, the potential Gf_0 exists and is finite on X. Also, $Gf_0 \geq f_0$. Thus Gf_0 is a positive superharmonic function that satisfies $Gf_0(y) \geq h(y)$ for all $y \in A$. By definition of the reduced function, we get $R^A[h] \leq Gf_0$. Now we see that $R^A[h]$ is a positive superharmonic function majorized by a potential, and Corollary 6.44(2) implies that there is a function $f = f_{h,A} \geq 0$ such that

$$R^A[h] = Gf.$$

Let B be another finite subset of X, containing A. Then $R^B[h]$ is a positive superharmonic function that majorizes h on the set A. Hence

$$R^B[h] \geq R^A[h], \quad \text{if } B \supset A.$$

Now we can conclude the proof of the theorem. Let (A_n) be an increasing sequence of finite subsets of X such that $X = \bigcup_n A_n$, and let

$$g_n = R^{A_n}[h].$$

Then each g_n is the potential of a non-negative function f_n, we know that $g_n \leq g_{n+1} \leq h$, and g_n coincides with h on A_n. □

The approximation theorem applies in particular to positive harmonic functions. In the Riesz decomposition of such a function h, the potential is 0. Nevertheless, h can not only be approximated from below by potentials, but the latter can be chosen such as to coincide with h on arbitrarily large finite subsets of X.

E "Balayage" and domination principle

For $A \subset X$ and $x, y \in X$ we define

$$\begin{aligned} F^A(x,y) &= \sum_{n=0}^{\infty} \mathsf{Pr}_x[Z_n = y,\ Z_j \notin A \text{ for } 0 \leq j < n] \cdot \mathbf{1}_A(y), \\ L^A(x,y) &= \sum_{n=0}^{\infty} \mathsf{Pr}_x[Z_n = y,\ Z_j \notin A \text{ for } 0 < j \leq n] \cdot \mathbf{1}_A(x). \end{aligned} \tag{6.47}$$

Thus, $F^A(x,y) = \mathsf{Pr}_x[Z_{s^A} = y]$ is the probability that the first visit in the set A of the Markov chain starting at x occurs at y. On the other hand, $L^A(x,y)$ is the expected number of visits in the point y before re-entering A, where $Z_0 = x \in A$.

$F^{\{y\}}(x,y) = F(x,y) = G(x,y)/G(y,y)$ coincides with the probability to reach y starting from x, defined in (1.27). In the same way $L^{\{x\}}(x,y) = L(x,y) = G(x,y)/G(x,x)$ is the quantity defined in (3.57). In particular, Lemma 3.58 implies that $L^A(x,y) \le L(x,y)$ is finite even when the Markov chain is recurrent.

Paths and their weights have been considered at the end of Chapter 1. In that notation,

$$F^A(x,y) = \mathsf{w}\big(\{\pi \in \Pi(x,y) : \pi \text{ meets} A \text{ only in the terminal point}\}\big), \quad \text{and}$$
$$L^A(x,y) = \mathsf{w}\big(\{\pi \in \Pi(x,y) : \pi \text{ meets} A \text{ only in the initial point}\}\big).$$

6.48 Exercise. Prove the following duality between F^A and L^A: let $\widehat{P}$ be the reversal of P with respect to some excessive (positive) measure ν, as defined in (6.27), then

$$\widehat{L}^A(x,y) = \frac{\nu(y)F^A(y,x)}{\nu(x)} \quad \text{and} \quad \widehat{F}^A(x,y) = \frac{\nu(y)L^A(y,x)}{\nu(x)},$$

where $\widehat{F}^A(x,y)$ and $\widehat{L}^A(x,y)$ are the quantities of (6.47) relative to $\widehat{P}$. □

The following two identities are obvious.

$$x \in A \Longrightarrow F^A(x,\cdot) = \delta_x, \quad \text{and} \quad y \in A \Longrightarrow L^A(\cdot,y) = \mathbf{1}_y. \tag{6.49}$$

(It is always useful to think of functions as column vectors and of measures as row vectors, whence the distinction between $\mathbf{1}_y$ and δ_x.) We recall for the following that we consider the restriction $P_{X\setminus A}$ of P to $X \setminus A$ and the associated Green function $G_{X\setminus A}$ on the whole of X, taking values 0 if $x \in A$ or $y \in A$.

6.50 Lemma. (a) $G = G_{X\setminus A} + F^A\, G$, (b) $G = G_{X\setminus A} + G\, L^A$.

Proof. We show only (a); statement (b) follows from (a) by duality (6.48). For $x, y \in X$,

$$\begin{aligned} p^{(n)}(x,y) &= \mathsf{Pr}_x[Z_n = y,\ s^A > n] + \mathsf{Pr}_x[Z_n = y,\ s^A \le n] \\ &= p^{(n)}_{X\setminus A}(x,y) + \sum_{v\in A} \mathsf{Pr}_x[Z_n = y,\ s^A \le n,\ Z_{s^A} = v] \\ &= p^{(n)}_{X\setminus A}(x,y) + \sum_{v\in A}\sum_{k=0}^{n} \mathsf{Pr}_x[Z_n = y,\ s^A = k,\ Z_k = v] \end{aligned}$$

$$= p^{(n)}_{X\setminus A}(x,y) + \sum_{v\in A}\sum_{k=0}^{n} \mathsf{Pr}_x[\boldsymbol{s}^A = k,\ Z_k = v]\ \mathsf{Pr}_x[Z_n = y \mid Z_k = v]$$

$$= p^{(n)}_{X\setminus A}(x,y) + \sum_{v\in A}\sum_{k=0}^{n} \mathsf{Pr}_x[\boldsymbol{s}^A = k,\ Z_k = v]\ p^{(n-k)}(v,y).$$

Summing over all n and applying (as so often) the Cauchy formula for the product of two absolutely convergent series,

$$\begin{aligned} G(x,y) &= G_{X\setminus A}(x,y) + \sum_{v\in A}\Big(\sum_{k=0}^{\infty} \mathsf{Pr}_x[\boldsymbol{s}^A = k,\ Z_k = v]\Big)\Big(\sum_{n=0}^{\infty} p^{(n)}(v,y)\Big) \\ &= G_{X\setminus A}(x,y) + \sum_{v\in A} \mathsf{Pr}_x[\boldsymbol{s}^A < \infty,\ Z_{\boldsymbol{s}^A} = v]\, G(v,y) \\ &= G_{X\setminus A}(x,y) + \sum_{v\in X} F^A(x,v)\, G(v,y), \end{aligned}$$

as proposed. □

The interpretation of statement (a) in terms of weights of paths is as follows. Recall that $G(x,y)$ is the weight of the set of all paths from x to y. It can be decomposed as follows: we have those paths that remain completely in the complement of A – their contribution to $G(x,y)$ is $G_{X\setminus A}(x,y)$ – and every other path must posses a first entrance time into A, and factorizing with respect to that time one obtains that the overall weight of the latter set of paths is $\sum_{v\in A} F^A(x,v)\, G(v,y)$. The interpretation of statement (b) is analogous, decomposing with respect to the last visit in A.

6.51 Corollary. $F^A\, G = G\, L^A$.

There is also a link with the induced Markov chain, as follows.

6.52 Lemma. *The matrix P^A over $A \times A$ satisfies*

$$P^A = P_{A,X}\, F^A = L^A\, P_{X,A}.$$

Proof. We can rewrite (6.38) with $\boldsymbol{s}^A$ in the place of $\boldsymbol{t}^A$, since these two stopping times coincide when the initial point is not in A:

$$\begin{aligned} p^A(x,y) &= p(x,y) + \sum_{v\in X\setminus A} p(x,v)\ \mathsf{Pr}_v[\boldsymbol{s}^A < \infty,\ Z_{\boldsymbol{s}^A} = y] \\ &= \sum_{v\in A} p(x,v)\,\delta_v(y) + \sum_{v\in X\setminus A} p(x,v)\, F^A(v,y) \\ &= \sum_{v\in X} p(x,v)\, F^A(v,y). \end{aligned}$$

Observing that (6.28) implies $\hat{p}^A(x,y) = \nu(y)\, p^A(y,x)/\nu(x)$ (where ν is an excessive measure for P), the second identity follows by duality. □

6.53 Lemma. (1) *If $h \in \mathcal{S}^+(X,P)$, then $F^A h(x) = \sum_{y\in A} F^A(x,y)\, h(y)$ is finite and*

$$F^A h(x) \le h(x) \quad \textit{for all } x \in X.$$

(2) *If $\nu \in \mathcal{E}^+(X,P)$, then $\nu L^A(y) = \sum_{x\in A} \nu(x)\, L^A(x,y)$ is finite and*

$$\nu L^A(y) \le \nu(y) \quad \textit{for all } y \in X.$$

Proof. As usual, (2) follows from (1) by duality. We prove (1). By the approximation theorem (Theorem 6.46) we can find a sequence of potentials $g_n = Gf_n$ with $f_n \ge 0$ and $g_n \le g_{n+1}$, such that $\lim_n g_n = h$ pointwise on X. The f_n can be chosen to have finite support. Lemma 6.50 implies

$$F^A g_n = F^A G\, f_n = Gf_n - G_{X\setminus A}\, f_n \le g_n \le h.$$

By the monotone convergence theorem,

$$F^A h = F^A\big(\lim_{n\to\infty} g_n\big) = \lim_{n\to\infty}\big(F^A g_n\big) \le h,$$

which proves the claims. □

Recall the definition of the *reduced function on A* of a positive superharmonic function h: for $x \in X$,

$$R^A[h](x) = \inf\{u(x) : u \in \mathcal{S}^+,\ u(y) \ge h(y) \text{ for all } y \in A\}.$$

Analogously one defines the *reduced measure on A* of an excessive measure ν:

$$R^A[\nu](x) = \inf\{\mu(x) : \mu \in \mathcal{E}^+,\ \mu(y) \ge \nu(y) \text{ for all } y \in A\}.$$

We are now able to describe the reduced functions and measures in terms of matrix operators.

6.54 Theorem. (i) *If $h \in \mathcal{S}^+$ then $R^A[h] = F^A h$. In particular, $R^A[h]$ is harmonic in every point of $X \setminus A$, while $R^A[h] \equiv h$ on A.*

(ii) *If $\nu \in \mathcal{E}^+$ then $R^A[\nu] = \nu L^A$. In particular, $R^A[\nu]$ is invariant in every point of $X \setminus A$, while $R^A[\nu] \equiv \nu$ on A.*

Proof. Always by duality it is sufficient to prove (a).

1.) If $x \in X \setminus A$ and $y \in A$, we factorize with respect to the first step: by (6.49)

$$F^A(x,y) = p(x,y) + \sum_{v\in X\setminus A} p(x,v)\, F^A(v,y) = \sum_{v\in X} p(x,v)\, F^A(v,y).$$

In particular,

$$F^A h(x) = P(F^A h)(x), \quad x \in X \setminus A.$$

2.) If $x \in A$ then by Lemma 6.52 and the "dual" of Theorem 6.39 (Exercise 6.40),

$$P\,(F^A h)(x) = \sum_{y \in A} P\,F^A(x, y)\, h(y) = P^A h(x) \le h(x).$$

3.) Again by (6.49),

$$F^A h(x) = h(x) \quad \text{for all } x \in A.$$

Combining 1.), 2.) and 3.), we see that

$$F^A h \in \{u \in \mathcal{S}^+ : u(y) \ge h(y) \text{ for all } y \in A\}.$$

Therefore $R^A[h] \le F^A h$.

4.) Now let $u \in \mathcal{S}^+$ and $u(y) \ge h(y)$ for every $y \in A$. By Lemma 6.53, for very $x \in X$

$$u(x) \ge \sum_{y \in A} F^A(x, y)\, u(y) \ge \sum_{y \in A} F^A(x, y)\, h(y) = F^A h(x).$$

Therefore $R^A[h] \ge F^A h$. □

In particular, let f be a non-negative G-integrable function and $g = Gf$ its potential. By Corollary 6.44(2), $R^A[g]$ must be a potential. Indeed, by Corollary 6.51,

$$R^A[g] = F^A\, G\, f = G\, L^A\, f$$

is the potential of $L^A f$.

Analogously, if μ is a non-negative, G-integrable measure (that is, $\mu G(y) = \sum_x \mu(x)\, G(x, y) < \infty$ for all y), then its *potential* is the excessive measure $\nu = \mu G$. In this case,

$$R^A[\nu] = \mu\, F^A\, G$$

is the potential of the measure μF^A.

6.55 Definition. (1) If f is a non-negative G-integrable function on X, then the *balayée of* f is the function $f^A = L^A f$.

(2) If μ is a non-negative, G-integrable measure on X, then the *balayée of* μ is the measure $\mu^A = \mu F^A$.

The meaning of "balayée" (French, balayer ≡ sweep out) is the following: if one considers the potential $g = Gf$ only on the set A, the function f (the charge) contains "superfluous information". The latter can be eliminated by passing to the charge $L^A f$ which has the same potential on the set A, while on the complement of A that potential is as small as possible.

An important application of the preceding results is the following.

6.56 Theorem (Domination principle). *Let f be a non-negative, G-integrable function on X, with support A. If $h \in \mathcal{S}^+$ is such that $h(x) \geq Gf(x)$ for every $x \in A$, then $h \geq Gf$ on the whole of X.*

Proof. By (6.49), $f^A = f$. Lemma 6.53 and Corollary 6.51 imply

$$\begin{aligned} h(x) \geq F^A h(x) &= \sum_{y \in A} F^A(x,y) h(y) \\ &\geq \sum_{y \in A} F^A(x,y)\, Gf(y) = F^A\, Gf(x) = Gf^A(x) = Gf(x) \end{aligned}$$

for every $x \in X$. □

6.57 Exercise. Give direct proofs of all statements of the last section, concerning excessive and invariant measures, where we just relied on duality.

In particular, formulate and prove directly the dual domination principle for excessive measures. □

6.58 Exercise. Use the domination principle to show that

$$G(x,y) \geq F(x,w) G(w,y).$$

Dividing by $G(y,y)$, this leads to Proposition 1.43 (a) for $z = 1$. □

Chapter 7
The Martin boundary of transient Markov chains

A Minimal harmonic functions

As in the preceding chapter, *we always suppose that* (X, P) *is irreducible and* P *substochastic.* We want to undertake a more detailed study of harmonic and superharmonic functions.

We know that $\mathcal{S}^+ = \mathcal{S}^+(X, P)$, besides the $\mathbf{0}$ function, contains all non-negative constant functions. As we have already stated, $\mathcal{S}^+$ is a *cone* with vertex $\mathbf{0}$: if $u \in \mathcal{S}^+ \setminus \{\mathbf{0}\}$, then the ray (half-line) $\{a \cdot u : a \geq 0\}$ starting at $\mathbf{0}$ and passing through u is entirely contained in $\mathcal{S}^+$. Furthermore, the cone $\mathcal{S}^+$ is *convex:* if u_1, u_2 are non-negative superharmonic functions and $a_1, a_2 \geq 0$ then $a_1 \cdot u_1 + a_2 \cdot u_2 \in \mathcal{S}^+$. (Since we have a cone with vertex $\mathbf{0}$, it is superfluous to require that $a_1 + a_2 = 1$.)

A *base* of a cone with vertex $\bar{v}$ is a subset $\mathcal{B}$ such that each element of the cone different from $\bar{v}$ can be uniquely written as $\bar{v} + a \cdot (u - \bar{v})$ with $a > 0$ and $u \in \mathcal{B}$. Let us fix a reference point ("origin") $o \in X$. Then the set

$$\mathcal{B} = \{u \in \mathcal{S}^+ : u(o) = 1\} \tag{7.1}$$

is a base of the cone $\mathcal{S}^+$. Indeed, if $v \in \mathcal{S}^+$ and $v \neq \mathbf{0}$, then $v(x) > 0$ for each $x \in X$ by Lemma 6.17. Hence $u = \frac{1}{v(o)} v \in \mathcal{B}$ and we can write $v = a \cdot u$ with $a = v(o)$.

Finally, we observe that $\mathcal{S}^+$, as a subset of the space of all functions $X \to \mathbb{R}$, carries the *topology of pointwise convergence:* a sequence of functions $f_n : X \to \mathbb{R}$ converges to the function f if and only if $f_n(x) \to f(x)$ for every $x \in X$. This is the product topology on $\mathbb{R}^X$.

We shall say that our Markov chain has *finite range,* if for every $x \in X$ there is only a finite number of $y \in X$ with $p(x, y) > 0$.

7.2 Theorem. (a) $\mathcal{S}^+$ *is closed and* $\mathcal{B}$ *is compact in the topology of pointwise convergence.*

(b) *If* P *has finite range then* $\mathcal{H}^+$ *is closed.*

Proof. (a) In order to verify compactness of $\mathcal{B}$, we observe first of all that $\mathcal{B}$ is closed. Let (u_n) be a sequence of functions in $\mathcal{S}^+$ that converges pointwise to the function $u : X \to \mathbb{R}$. Then, by Fatou's lemma (where the action of P represents the integral),

$$Pu = P\,(\liminf u_n) \leq \liminf Pu_n \leq \liminf u_n = u,$$

and u is superharmonic. Let $x \in X$. By irreducibility we can choose $k = k_x$ such that $p^{(k)}(o,x) > 0$. Then for every $u \in \mathcal{B}$

$$1 = u(o) \geq P^k u(o) \geq p^{(k)}(o,x)u(x),$$

and

$$u(x) \leq C_x = 1/p^{(k_x)}(o,x). \tag{7.3}$$

Thus, $\mathcal{B}$ is contained in the compact set $\prod_{x\in X}[0,\ C_x]$. Being closed, $\mathcal{B}$ is compact.

(b) If (h_n) is a sequence of non-negative harmonic functions that converges pointwise to the function h, then $h \in \mathcal{S}^+$ by (a). Furthermore, for each $x \in X$, the summation in $Ph_n(x) = \sum_y p(x,y)h_n(y)$ is finite. Therefore we may exchange summation and limit,

$$Ph = P\,(\lim h_n) = \lim Ph_n = \lim h_n = h,$$

and h is harmonic. □

We see that $\mathcal{S}^+$ is a convex cone with compact base $\mathcal{B}$ which contains $\mathcal{H}^+$ as a convex sub-cone. When P does not have finite range, that sub-cone is not necessarily closed. It should be intuitively clear that in order to know $\mathcal{S}^+$ (and consequently also $\mathcal{H}^+$) it will be sufficient to understand $\mathrm{d}\mathcal{B}$, the set of extremal points of the convex set $\mathcal{B}$. Recall that an element u of a convex set is called *extremal* if it cannot be written as a convex combination $a \cdot u_1 + (1-a)\cdot u_2$ $(0 < a < 1)$ of distinct elements u_1, u_2 of the same set. Our next aim is to determine the elements of $\mathrm{d}\mathcal{B}$.

In the transient case we know from Lemma 6.19 that for each $y \in X$, the function $x \mapsto G(x,y)$ belongs to $\mathcal{S}^+$ and is strictly superharmonic in the point y. However, it does not belong to $\mathcal{B}$. Hence, we normalize by dividing by its value in o, which is non-zero by irreducibility.

7.4 Definition. (i) The *Martin kernel* is

$$K(x,y) = \frac{F(x,y)}{F(o,y)}, \quad x, y \in X.$$

(ii) A function $h \in \mathcal{H}^+$ is called *minimal,* if

- $h(o) = 1$, and
- if $h_1 \in \mathcal{H}^+$ and $h \geq h_1$ in each point, then h_1/h is constant.

Note that in the recurrent case, the Martin kernel is also defined and is constant $= 1$, and $\mathcal{S}^+ = \mathcal{H}^+ = \{\text{non-negative constant functions}\}$ by Theorem 6.21. Thus, we may limit our attention to the transient case, in which

$$K(x,y) = \frac{G(x,y)}{G(o,y)}. \tag{7.5}$$

7.6 Theorem. *If (X, P) is transient, then the extremal elements of $\mathcal{B}$ are the Martin kernels and the minimal harmonic functions:*

$$\mathrm{d}\mathcal{B} = \{K(\cdot, y) : y \in X\} \cup \{h \in \mathcal{H}^+ : h \text{ is minimal}\}.$$

Proof. Let u be an extremal element of $\mathcal{B}$. Write its Riesz decomposition (Theorem 6.43): $u = Gf + h$ with $f \geq 0$ and $h \in \mathcal{H}^+$.

Suppose that both Gf and h are non-zero. By Lemma 6.17, the values of these functions in o are (strictly) positive, and we can define

$$u_1 = \frac{1}{Gf(o)}\, Gf, \quad u_2 = \frac{1}{h(o)}\, h \in \mathcal{B}.$$

Since $Gf(o) + h(o) = u(o) = 1$, we can write u as a convex combination $u = a \cdot u_1 + (1-a) \cdot u_2$ with $0 < a = Gf(o) < 1$. But u_1 is strictly superharmonic in at least one point, while u_2 is harmonic, so that we must have $u_1 \neq u_2$. This contradicts extremality of u. Therefore u is a potential or a harmonic function.

Case 1. $u = Gf$, where $f \geq 0$. Let $A = \operatorname{supp}(f)$. This set must have at least one element y. Suppose that A has more than one element. Consider the restrictions f_1 and f_2 of f to $\{y\}$ and $A \setminus \{y\}$, respectively. Then $u = Gf_1 + Gf_2$, and as above, setting $a = Gf_1(o)$, we can rewrite this identity as a convex combination,

$$u = a \cdot Gg_1 + (1-a) \cdot Gg_2, \quad \text{where } g_1 = \tfrac{1}{a} f_1 \text{ and } g_2 = \tfrac{1}{1-a} f_2.$$

By assumption, u is extremal. Hence we must have $Gg_1 = Gg_2$ and thus (by Lemma 6.42) also $g_1 = g_2$, a contradiction.

Consequently $A = \{y\}$ and $f = a \cdot \mathbf{1}_y$ with $a > 0$. Since $a \cdot G(o, y) = Gf(o) = u(o) = 1$, we find

$$u = K(\cdot, y),$$

as proposed.

Case 2. $u = h \in \mathcal{H}^+$. We have to prove that h is minimal. By hypothesis, $h(o) = 1$. Suppose that $h \geq h_1$ for a function $h_1 \in \mathcal{H}^+$. If $h_1 = \mathbf{0}$ or $h_1 = h$, then h_1/h is constant. Otherwise, setting $h_2 = h - h_1$, both h_1 and h_2 are strictly positive harmonic functions (Lemma 6.17). As above, we obtain a convex combination

$$h = h_1(o) \cdot \frac{h_1}{h_1(o)} + h_2(o) \cdot \frac{h_2}{h_2(o)}.$$

But then we must have $\frac{1}{h_i(o)} h_i = h$. In particular, h_1/h is constant.

Conversely, we must verify that the functions $K(\cdot, y)$ and the minimal harmonic functions are elements of $\mathrm{d}\mathcal{B}$.

Consider first the function $K(\cdot, y)$ with $y \in X$. It can be written as the potential Gf, where $f = \frac{1}{G(o,y)}\mathbf{1}_y$. Suppose that

$$K(\cdot, y) = a \cdot u_1 + (1-a) \cdot u_2$$

with $0 < a < 1$ and $u_i \in \mathcal{B}$. Then $u_1 \leq G\left(\frac{1}{a} f\right)$, and u_1 is dominated by a potential. By Corollary 6.44, it must itself be a potential $u_1 = Gf_1$ with $f_1 \geq 0$. If $\operatorname{supp}(f_1)$ contained some $w \in X$ different from y, then u_1, and therefore also $K(\cdot, y)$, would be strictly superharmonic in w, a contradiction. We conclude that $\operatorname{supp}(f_1) = \{y\}$, and $f_1 = c \cdot G(\cdot, y)$ for some constant $c > 0$. Since $u_1(o) = 1$, we must have $c = 1/G(o, y)$, and $u_1 = K(\cdot, y)$. It follows that also $u_2 = K(\cdot, y)$. This proves that $K(\cdot, y) \in \mathrm{d}\mathcal{B}$.

Now let $h \in \mathcal{H}^+$ be a minimal harmonic function. Suppose that

$$h = a \cdot u_1 + (1-a) \cdot u_2$$

with $0 < a < 1$ and $u_i \in \mathcal{B}$. None of the functions u_1 and u_2 can be strictly subharmonic in some point (since otherwise also h would have this property). We obtain $a \cdot u_1 \in \mathcal{H}^+$ and $h \geq a \cdot u_1$. By minimality of h, the function u_1/h is constant. Since $u_1(o) = 1 = h(o)$, we must have $u_1 = h$ and thus also $u_2 = h$. This proves that $h \in \mathrm{d}\mathcal{B}$. □

We shall now exhibit two general criteria that are useful for recognizing the minimal harmonic functions. Let h be an arbitrary positive, non-zero superharmonic function. We use h to define a new transition matrix $P_h = \big(p_h(x, y)\big)_{x,y \in X}$:

$$p_h(x, y) = \frac{p(x, y)h(y)}{h(x)}. \tag{7.7}$$

The Markov chain with these transition probabilities is called the *h-process,* or also Doob's h-process, see his fundamental paper [17].

We observe that in the notation that we have introduced in §1.B, the random variables of this chain remain always Z_n, the projections of the trajectory space onto X. What changes is the probability measure on Ω (and consequently also the distributions of the Z_n). We shall write Pr_x^h for the family of probability measures on $(\Omega, \mathcal{A})$ with starting point $x \in X$ which govern the h-process. If h is strictly subharmonic in some point, then recall that we have to add the "tomb" state † as in (6.11).

The construction of P_h is similar to that of the reversed chain with respect to an excessive measure as in (6.27). In particular,

$$p_h^{(n)}(x, y) = \frac{p^{(n)}(x, y)h(y)}{h(x)}, \quad \text{and} \quad G_h(x, y) = \frac{G(x, y)h(y)}{h(x)}, \tag{7.8}$$

where G_h denotes the Green function associated with P_h in the transient case.

7.9 Exercise. Prove the following simple facts.

(1) The matrix P_h is stochastic if and only if $h \in \mathcal{H}^+$.

(2) One has $u \in \mathcal{S}(X, P)$ if and only if $\bar{u} = u/h \in \mathcal{S}(X, P_h)$. Furthermore, u is harmonic with respect to P if and only if $\bar{u}$ is harmonic with respect to P_h.

(3) A function $u \in \mathcal{H}^+(X, P)$ with $u(o) = 1$ is minimal harmonic with respect to P if and only if $h(o) \cdot \bar{u} \in \mathcal{H}^+(X, P_h)$ is minimal harmonic with respect to P_h. □

We recall that $\mathcal{H}^\infty = \mathcal{H}^\infty(X, P)$ denotes the linear space of all bounded harmonic functions.

7.10 Lemma. *Let (X, P) be an irreducible Markov chain with stochastic transition matrix P. Then $\mathcal{H}^\infty = \{\text{constants}\}$ if and only if the constant harmonic function $\mathbf{1}$ is minimal.*

Proof. Suppose that $\mathcal{H}^\infty = \{\text{constants}\}$. Let h_1 be a positive harmonic function with $\mathbf{1} \geq h_1$. Then h_1 is bounded, whence constant by the assumption. Therefore $h_1/\mathbf{1}$ is constant, and $\mathbf{1}$ is a minimal harmonic function.

Conversely, suppose that the harmonic function $\mathbf{1}$ is minimal. If $h \in \mathcal{H}^\infty$ then there is a constant M such that $h_1 = h + M$ is a positive function. Since P is stochastic, h_1 is harmonic. But h_1 is also bounded, so that $\mathbf{1} \geq c \cdot h_1$ for some $c > 0$. By minimality of $\mathbf{1}$, the ratio $h_1/\mathbf{1}$ must be constant. Therefore also h is constant. □

Setting $u = h$ in Exercise 7.9(3), we obtain the following corollary (valid also when P is substochastic, since what matters is stochasticity of P_h).

7.11 Corollary. *A function $h \in \mathcal{H}^+(X, P)$ is minimal if and only if one has $\mathcal{H}^\infty(X, P_h) = \{\text{constants}\}$.*

If C is a compact, convex set in the Euclidean space $\mathbb{R}^d$, and if the set $\mathrm{d}C$ of its extremal points is finite, then it is known that every element $\boldsymbol{x} \in C$ can be written as a weighted average $\boldsymbol{x} = \sum_{\boldsymbol{c} \in \mathrm{d}C} \nu(\boldsymbol{c}) \cdot \boldsymbol{c}$ of the elements of $\mathrm{d}C$. The numbers $\nu(\boldsymbol{c})$ make up a probability measure on $\mathrm{d}C$. (Note that in general, $\mathrm{d}C$ is not the topological boundary of C.) If $\mathrm{d}C$ is infinite, the weighted sum has to be replaced by an integral $\boldsymbol{x} = \int_{\mathrm{d}C} \boldsymbol{c}\, d\nu(\boldsymbol{c})$, where ν is a probability measure on $\mathrm{d}C$. In general, this integral representation is not unique: for example, the interior points of a rectangle (or a disk) can be written in different ways as weighted averages of the four vertices (or the points on the boundary circle of the disk, respectively). However, the representation does become unique if the set C is a *simplex:* a triangle in dimension 2, a tetrahedron in dimension 3, etc.; in dimension d, a simplex has $d + 1$ extremal points.

We use these observations as a motivation for the study of the compact convex set $\mathcal{B}$, base of the cone $\mathcal{S}^+$. We could appeal to CHOQUET's representation theory of convex cones in topological linear spaces, see for example PHELPS [Ph]. However, in our direct approach regarding the (X, P), we shall obtain a more detailed specific understanding. In the next sections, we shall prove (among other) the following results:

- The base $\mathcal{B}$ is a simplex (usually infinite dimensional) in the sense that every element of $\mathcal{B}$ can be written uniquely as the integral of the elements of $\mathrm{d}\mathcal{B}$ with respect to a suitable probability measure.
- Every minimal harmonic function can be approximated by a sequence of functions $K(\cdot, y_n)$, where $y_n \in X$.

We shall obtain these results via the construction of the *Martin compactification,* a compactification of the state space X defined by the Martin kernel.

B The Martin compactification

Preamble on compactifications

Given the countably infinite set X, by a *compactification* of X we mean a compact topological Hausdorff space $\widehat{X}$ containing X such that

- the set X is dense in $\widehat{X}$, and
- in the induced topology, $X \subset \widehat{X}$ is discrete.

The set $\widehat{X} \setminus X$ is called the *boundary* or *ideal boundary* of X in $\widehat{X}$. We consider two compactifications of X as "equal", that is, *equivalent*, if the identity function $X \to X$ extends to a homeomorphism between the two. Also, we consider one compactification bigger than a second one, if the identity $X \to X$ extends to a continuous surjection from the first onto the second. The following topological exercise does not require Cantor–Bernstein or other deep theorems.

7.12 Exercise. Two compactifications of X are equivalent if and only if each of them is bigger than the other one.

[Hint: use sequences in X.] □

Given a family $\mathcal{F}$ of real valued functions on X, there is a standard way to associate with $\mathcal{F}$ a compactification (in general not necessarily Hausdorff) of X. In the following theorem, we limit ourselves to those hypotheses that will be used in the sequel.

7.13 Theorem. *Let $\mathcal{F}$ be a denumerable family of bounded functions on X. Then there exists a unique (up to equivalence) compactification $\widehat{X} = \widehat{X}_{\mathcal{F}}$ of X such that*

(a) *every function* $f \in \mathcal{F}$ *extends to a continuous function on* $\widehat{X}$ (*which we still denote by* f), *and*

(b) *the family* $\mathcal{F}$ *separates the boundary points: if* $\xi, \eta \in \widehat{X} \setminus X$ *are distinct, then there is* $f \in \mathcal{F}$ *with* $f(\xi) \neq f(\eta)$.

Proof. 1.) *Existence* (*construction*).

For $x \in X$, we write $\mathbf{1}_x$ for the indicator function of the point x. We add all those indicator functions to $\mathcal{F}$, setting

$$\mathcal{F}^* = \mathcal{F} \cup \{\mathbf{1}_x : x \in X\}. \tag{7.14}$$

For each $f \in \mathcal{F}^*$, there is a constant C_f such that $|f(x)| \leq C_f$ for all $x \in X$. Consider the topological product space

$$\Pi_{\mathcal{F}} = \prod_{f \in \mathcal{F}^*} [-C_f, C_f] = \{\phi \colon \mathcal{F}^* \to \mathbb{R} \mid \phi(f) \in [-C_f, C_f] \text{ for all } f \in \mathcal{F}^*\}.$$

The topology on $\Pi_{\mathcal{F}}$ is the one of pointwise convergence: $\phi_n \to \phi$ if and only if $\phi_n(f) \to \phi(f)$ for every $f \in \mathcal{F}^*$. A neighbourhood base at $\phi \in \Pi_{\mathcal{F}}$ is given by the finite intersections of sets of the form $\{\psi \in \Pi_{\mathcal{F}} : |\psi(f) - \phi(f)| < \varepsilon\}$, as $f \in \mathcal{F}^*$ and $\varepsilon > 0$ vary.

We can embed X into $\Pi_{\mathcal{F}}$ via the map

$$\iota : X \hookrightarrow \Pi_{\mathcal{F}}, \quad \iota(x) = \phi_x, \quad \text{where } \phi_x(f) = f(x) \text{ for } f \in \mathcal{F}^*.$$

If x, y are two distinct elements of X then $\phi_x(\mathbf{1}_x) = 1 \neq 0 = \phi_y(\mathbf{1}_x)$. Therefore ι is injective. Furthermore, the neighbourhood $\{\psi \in \Pi_{\mathcal{F}} : |\psi(\mathbf{1}_x) - \phi_x(\mathbf{1}_x)| < 1\}$ of $\iota(x) = \phi_x$ contains none of the functions ϕ_y with $y \in X \setminus \{x\}$. This means that $\iota(X)$, with the induced topology, is a discrete subset of $\Pi_{\mathcal{F}}$. Thus we can identify X with $\iota(X)$. [Observe how the enlargement of $\mathcal{F}$ by the indicator functions has been crucial for this reasoning.]

Now $\widehat{X} = \widehat{X}_{\mathcal{F}}$ is defined as the closure of X in $\Pi_{\mathcal{F}}$. It is clear that this is a compactification of X in our sense. Each $\xi \in \widehat{X} \setminus X$ is a function $\mathcal{F}^* \to \mathbb{R}$ with $|\xi(f)| \leq C_f$. By the construction of $\widehat{X}$, there must be a sequence (x_n) of distinct points in X that converges to ξ, that is, $f(x_n) = \phi_{x_n}(f) \to \xi(f)$ for every $f \in \mathcal{F}^*$. We prefer to think of ξ as a limit point of X in a more "geometrical" way, and *define* $f(\xi) = \xi(f)$ for $f \in \mathcal{F}$. Observe that since $\phi_{x_n}(\mathbf{1}_x) = 0$ when $x_n \neq x$, we have $\mathbf{1}_x(\xi) = \xi(\mathbf{1}_x) = 0$ for every $x \in X$, as it should be.

If (x_n) is an arbitrary sequence in X which converges to ξ in the topology of $\widehat{X}$, then for each $f \in \mathcal{F}$ one has

$$f(x_n) = \phi_{x_n}(f) \to \xi(f) = f(\xi).$$

Thus, f has become a continuous functions on $\widehat{X}$. Finally, $\mathcal{F}$ separates the points of $\widehat{X} \setminus X$: if ξ, η are two distinct boundary points, then they are also distinct in

their original definition as functions on $\mathcal{F}^*$. Hence there is $f \in \mathcal{F}^*$ such that $\xi(f) \neq \eta(f)$. Since $\xi(\mathbf{1}_x) = \eta(\mathbf{1}_x) = 0$ for every $x \in X$, we must have $f \in \mathcal{F}$. With the "reversed" notation that we have introduced above, $f(\xi) \neq f(\eta)$.

2.) *Uniqueness.* To show uniqueness of $\widehat{X}$ up to homeomorphism, suppose that $\widetilde{X}$ is another compactification of X with properties (a) and (b). We only use those defining properties in the proof, so that the roles of $\widehat{X}$ and $\widetilde{X}$ can be exchanged ("symmetry"). In order to distinguish the continuous extension of a function $f \in \mathcal{F}$ to $\widehat{X}$ from the one to $\widetilde{X}$, in this proof we shall write $\hat{f}$ for the former and $\tilde{f}$ for the latter.

We construct a function $\tau\colon \widetilde{X} \to \widehat{X}$: if $x \in X \subset \widetilde{X}$ then we set $\tau(x) = x$ (the latter seen as an element of $\widehat{X}$).

If $\tilde{\xi} \in \widetilde{X} \setminus X$, then there must be a sequence (x_n) in X such that in the topology of $\widetilde{X}$, one has $x_n \to \tilde{\xi}$. We show that $x_n\ (= \tau(x_n))$ has a limit in $\widehat{X} \setminus X$: let $\hat{\xi} \in \widehat{X}$ be an accumulation point of (x_n) in the latter compact space. If $\hat{\xi} \in X$ then $x_n = \hat{\xi}$ for infinitely many n, which contradicts the fact that $x_n \to \tilde{\xi}$ in $\widetilde{X}$. Therefore every accumulation point of (x_n) in $\widehat{X}$ lies in the boundary $\widehat{X} \setminus X$. Suppose there is another accumulation point $\hat{\eta} \in \widehat{X} \setminus X$. Then there is $f \in \mathcal{F}$ with $\hat{f}(\hat{\xi}) \neq \hat{f}(\hat{\eta})$. As $\hat{f}$ is continuous, we find that the real sequence $\big(f(x_n)\big)$ possesses the two distinct real accumulation points $\hat{f}(\xi)$ and $\hat{f}(\eta)$. But this is impossible, since with respect to the compactification $\widetilde{X}$, we have that $f(x_n) \to \tilde{f}(\tilde{\xi})$.

Therefore there is $\hat{\xi} \in \widehat{X} \setminus X$ such that $x_n \to \hat{\xi}$ in the topology of $\widehat{X}$. We define $\tau(\tilde{\xi}) = \hat{\xi}$. This mapping is well defined: if (y_n) is another sequence that tends to $\tilde{\xi}$ in $\widetilde{X}$, then the union of the two sequences (x_n) and (y_n) also tends to $\tilde{\xi}$ in $\widetilde{X}$, so that the argument used a few lines above shows that the union of those two sequences must also converge to $\hat{\xi}$ in the topology of $\widehat{X}$.

By construction, τ is continuous. [Exercise: in case of doubts, prove this.] Since X is dense in both compactifications, τ is surjective.

In the same way (by symmetry), we can construct a continuous surjection $\widehat{X} \to \widetilde{X}$ which extends the identity mapping $X \to X$. It must be the inverse of τ (by continuity, since this is true on X). We conclude that τ is a homeomorphism. $\square$

We indicate two further, equivalent ways to construct the compactification $\widehat{X}$.

1.) Let us say that $x_n \to \infty$ for a sequence in X, if for every finite subset A of X, there are only finitely many n with $x_n \in A$. Consider the set

$$\mathcal{X}_\infty = \{(x_n) \in X^{\mathbb{N}} : x_n \to \infty \text{ and } \big(f(x_n)\big) \text{ converges for every } f \in \mathcal{F}\}.$$

On $\mathcal{X}_\infty$, we consider the following equivalence relation:

$$(x_n) \sim (y_n) \iff \lim f(x_n) = \lim f(y_n) \quad \text{for every } f \in \mathcal{F}.$$

The boundary of our compactification is $\mathcal{X}_\infty/\sim$, that is, $\widehat{X} = X \cup (\mathcal{X}_\infty/\sim)$. The topology is defined via convergence: on X, it is discrete; a sequence (x_n) in X converges to a boundary point ξ if $(x_n) \in \mathcal{X}_\infty$ and (x_n) belongs to ξ as an equivalence class under $\sim$.

2.) Consider the countable family of functions $\mathcal{F}^*$ as in (7.14). Each $f \in \mathcal{F}^*$ is bounded by a constant C_f. We choose weights $w_f > 0$ such that $\sum_{\mathcal{F}^*} w_f\, C_f < \infty$. Then we define a metric on X:

$$\theta(x, y) = \sum_{f \in \mathcal{F}^*} w_f \; |f(x) - f(y)|. \tag{7.15}$$

In this metric, X is discrete, while a sequence (x_n) which tends to ∞ in the above sense is a Cauchy sequence if and only if $\big(f(x_n)\big)$ converges in $\mathbb{R}$ for every $f \in \mathcal{F}$. The completion of (X, θ) is (homeomorphic with) $\widehat{X}$.

Observation: if $\mathcal{F}$ contains only constant functions, then $\widehat{X}$ is the one-point compactification: $\widehat{X} = X \cup \{\infty\}$, and convergence to ∞ is defined as in 1.) above.

7.16 Exercise. Elaborate the details regarding the constructions in 1.) and 2.) and show that the compactifications obtained in this way are (equivalent with) $\widehat{X}_{\mathcal{F}}$. □

After this preamble, we can now give the definition of the Martin compactification.

7.17 Definition. Let (X, P) be an irreducible, (sub)stochastic Markov chain. The *Martin compactification* of X with respect to P is defined as $\widehat{X}(P) = \widehat{X}_{\mathcal{F}}$, the compactification in the sense of Theorem 7.13 with respect to the family of functions $\mathcal{F} = \{K(x, \cdot) : x \in X\}$. The *Martin boundary* $\mathcal{M} = \mathcal{M}(P) = \widehat{X}(P) \setminus X$ is the ideal boundary of X in this compactification.

Note that all the functions $K(x, \cdot)$ of Definition 7.4 are bounded. Indeed, Proposition 1.43 (a) implies that

$$K(x, y) = \frac{F(x, y)}{F(o, y)} \le \frac{1}{F(o, x)} = C_x$$

for every $y \in X$. By (7.15), the topology of $\widehat{X}(P)$ is induced by a metric (as it has to be, since $\widehat{X}(P)$ is a compact separable Hausdorff space). If $\xi \in \mathcal{M}$, then we write of course $K(x, \xi)$ for the value of the extended function $K(x, \cdot)$ at ξ.

If (X, P) is recurrent, $\mathcal{F} = \{\mathbf{1}\}$, and the Martin boundary consists of one element only. We note at this point that for recurrent Markov chains, another notion of Martin compactification has been introduced, see Kemeny and Snell [35]. In the transient case, we have the following, where (attention) we now consider $K(\cdot, \xi)$ as a function on X for every $\xi \in \widehat{X}(P)$. In our notation, when applied from the left to the Martin kernel, the transition operator P acts on the first variable of $K(\cdot, \cdot)$.

7.18 Lemma. *If (X, P) is transient and $\xi \in \mathcal{M}$ then $K(\cdot, \xi)$ is a positive superharmonic function. If P has finite range at $x \in X$ (that is, for the given x, the set $\{y \in X : p(x, y) > 0\}$ is finite), then the function $K(\cdot, \xi)$ is harmonic in x.*

Proof. By construction of $\mathcal{M}$, there is a sequence (y_n) in X, tending to ∞, such that $K(\cdot, y_n) \to K(\cdot, \xi)$ pointwise on X. Thus $K(\cdot, \xi)$ is the pointwise limit of the superharmonic functions $K(\cdot, y_n)$ and consequently a superharmonic function.

By (1.34) and (7.5),

$$PK(x, y_n) = \sum_{y : p(x,y)>0} p(x, y)\, K(y, y_n) = K(x, y_n) - \frac{\delta_x(y_n)}{K(o, y_n)}.$$

If the summation is finite, it can be exchanged with the limit as $n \to \infty$. Since $y_n \neq x$ for all but (at most) finitely many n, we have that $\delta_x(y_n) \to 0$. Therefore $PK(x, \xi) = K(x, \xi)$. □

In particular, if the Markov chain has finite range (at every point), then for every $\xi \in \mathcal{M}$, the function $K(\cdot, \xi)$ is positive harmonic with value 1 in o.

Another construction in case of finite range

The last observation allows us to describe a fourth, more specific construction of the Martin compactification in the case when (X, P) is *transient* and has *finite range*.

Let $\mathcal{B} = \{u \in \mathcal{S}^+ : u(o) = 1\}$ be the base of the cone $\mathcal{S}^+$, defined in (7.1), with the topology of pointwise convergence. We can embed X into $\mathcal{B}$ via the map $y \mapsto K(\cdot, y)$. Indeed, this map is injective (one-to-one): suppose that $K(\cdot, y_1) = K(\cdot, y_2)$ for two distinct elements $y_1, y_2 \in X$. Then

$$\frac{1}{F(o, y_1)} = K(y_1, y_1) = K(y_1, y_2) = \frac{F(y_1, y_2)}{F(o, y_2)} \quad \text{and}$$
$$\frac{1}{F(o, y_2)} = K(y_2, y_2) = K(y_2, y_1) = \frac{F(y_2, y_1)}{F(o, y_1)}.$$

We deduce $F(y_1, y_2)F(y_2, y_1) = 1$ which implies $U(y_1, y_1) = 1$, see Exercise 1.44.

Now we identify X with its image in $\mathcal{B}$. The Martin compactification is then the closure of X in $\mathcal{B}$. Indeed, among the properties which characterize $\widehat{X}(P)$ according to Theorem 7.13, the only one which is not immediate is that $X \equiv \{K(\cdot, y) : y \in X\}$ is discrete in $\mathcal{B}$: let us suppose that (y_n) is a sequence of distinct elements of X such that $K(\cdot, y_n)$ converges pointwise. By finite range, the limit function is harmonic. In particular, it cannot be one of the functions $K(\cdot, y)$, where $y \in X$, as the latter is strictly superharmonic at y. In other words, no element of X can be an accumulation point of (y_n).

We remark at this point that in the original article of Doob [17] and in the book of Kemeny, Snell and Knapp [K-S-K] it is not required that X be discrete in the Martin compactification. In their setting, the compactification can always be described as the closure of (the embedding of) X in $\mathcal{B}$. However, in the case when P does not have finite range, the compact space thus obtained may be smaller than in our construction, which follows the one of Hunt [32]. In fact, it can happen that there are $y \in X$ and $\xi \in \mathcal{M}$ such that $K(\cdot, y) = K(\cdot, \xi)$: in our construction, they are considered distinct in any case (since ξ is a limit point of a sequence that tends to ∞), while in the construction of [17] and [K-S-K], ξ and y would be identified.

A discussion, in the context of random walks of trees with infinite vertex degrees, can be found in Section 9.E after Example 9.47. In few words, one can say that for most probabilistic purposes the smaller compactification is sufficient, while for more analytically flavoured issues it is necessary to maintain the original discrete topology on X.

We now want to state a first fundamental theorem regarding the Martin compactification. Let us first remark that $\widehat{X}(P)$, as a compact metric space, carries a natural σ-algebra, namely the *Borel σ-algebra,* which is generated by the collection of all open sets. Speaking of a "random variable with values in $\widehat{X}(P)$", we intend a function from the trajectory space $(\Omega, \mathcal{A})$ to $\widehat{X}(P)$ which is measurable with respect to that σ-algebra.

7.19 Theorem (Convergence to the boundary). *If (X, P) is stochastic and transient then there is a random variable Z_∞ taking its values in $\mathcal{M}$ such that for each $x \in X$,*

$$\lim_{n\to\infty} Z_n = Z_\infty \quad \mathsf{Pr}_x\text{-almost surely}$$

in the topology of $\widehat{X}(P)$.

In terms of the trajectory space, the meaning of this statement is the following. Let

$$\Omega_\infty = \left\{\omega = (x_n) \in \Omega : \begin{array}{c}\text{there is } x_\infty \in \mathcal{M} \text{ such that} \\ x_n \to x_\infty \text{ in the topology of } \widehat{X}(P)\end{array}\right\}. \tag{7.20}$$

Then

$$\Omega_\infty \in \mathcal{A} \quad \text{and} \quad \mathsf{Pr}_x(\Omega_\infty) = 1 \quad \text{for every } x \in X.$$

Furthermore,

$$Z_\infty(\omega) = x_\infty \quad (\omega \in \Omega_\infty)$$

defines a random variable which is measurable with respect to the Borel σ-algebra on $\widehat{X}(P)$.

When P is strictly substochastic in some point, we have to modify the statement of the theorem. As we have seen in Section 7.B, in this case one introduces the

absorbing ("tomb") state † and extends P to $X \cup \{\dagger\}$. The construction of the Martin boundary remains unchanged, and does not involve the additional point †. However, the trajectory space $(\Omega, \mathcal{A})$ with the probability measures Pr_x, $x \in X$ now refers to $X \cup \{\dagger\}$. In reality, in order to construct it, we do not need *all* sequences in $X \cup \{\dagger\}$: it is sufficient to consider

$$\Omega = X^{\mathbb{N}_0} \cup \Omega_\dagger, \quad \text{where}$$
$$\Omega_\dagger = \left\{ \omega = (x_n) : \text{ there is } k \geq 1 \text{ with } \left\{ \begin{array}{ll} x_n \in X & \text{for all } n \leq k, \\ x_n = \dagger & \text{for all } n > k \end{array} \right\} \right. . \tag{7.21}$$

Indeed, once the Markov chain has reached †, it has to stay there forever. We write $\boldsymbol{\epsilon}(\omega) = k$ for $\omega \in \Omega_\dagger$, with k as in the definition of $\Omega_\dagger$, and $\boldsymbol{\epsilon}(\omega) = \infty$ for $\omega \in \Omega \setminus \Omega_\dagger$. Thus,

$$\boldsymbol{\epsilon} = \boldsymbol{t}^\dagger - 1$$

is a stopping time, the *exit time* from X – the last instant when the Markov chain is in X. With $\Omega_\dagger$ and Ω_∞ as in (7.21) and (7.20), respectively, we now define

$$\Omega_{\boldsymbol{\epsilon}} = \Omega_\dagger \cup \Omega_\infty, \quad \text{and}$$
$$Z_{\boldsymbol{\epsilon}} : \Omega_{\boldsymbol{\epsilon}} \to \widehat{X}(P), \quad Z_{\boldsymbol{\epsilon}}(\omega) = \begin{cases} Z_{\boldsymbol{\epsilon}(\omega)}(\omega), & \omega \in \Omega_\dagger, \\ Z_\infty(\omega), & \omega \in \Omega_\infty. \end{cases}$$

In this setting, Theorem 7.19 reads as follows.

7.22 Theorem. *If (X, P) is transient then for each $x \in X$,*

$$\lim_{n \to \boldsymbol{\epsilon}} Z_n = Z_{\boldsymbol{\epsilon}} \quad \mathsf{Pr}_x\text{-almost surely}$$

in the topology of $\widehat{X}(P)$.

Theorem 7.19 arises as a special case. Note that Theorem 7.22 comprises the following statements.

(a) $\Omega_{\boldsymbol{\epsilon}}$ belongs to the σ-algebra $\mathcal{A}$;

(b) $\mathsf{Pr}_x(\Omega_{\boldsymbol{\epsilon}}) = 1$ for every $x \in X$;

(c) $Z_{\boldsymbol{\epsilon}} : \Omega_{\boldsymbol{\epsilon}} \to \widehat{X}(P)$ is measurable with respect to the Borel σ-algebra of $\widehat{X}(P)$.

The most difficult part is the proof of (b), which will be elaborated in the next section. Thereafter, we shall deduce from Theorem 7.19 resp. 7.22 that

- every minimal harmonic function is a Martin kernel $K(\cdot, \xi)$, with $\xi \in \mathcal{M}$;
- every positive harmonic function h has an integral representation $h(x) = \int_{\mathcal{M}} K(x, \cdot)\, d\nu^h$, where ν^h is a Borel measure on $\mathcal{M}$.

The construction of the Martin compactification is an abstract one. In the study of specific classes of Markov chains, typically the state space carries an algebraic or geometric structure, and the transition probabilities are in some sense adapted to this structure; compare with the examples in earlier chapters. In this context, one is searching for a concrete description of the Martin compactification in terms of that underlying structure. In Chapters 8 and 9, we shall explain some examples; various classes of examples are treated in detail in the book of Woess [W2]. We observe at this point that in all cases where the Martin boundary is known explicitly in this sense, one also knows a simpler and more direct (structure-specific) method than that of the proof of Theorem 7.19 for showing almost sure convergence of the Markov chain to the "geometric" boundary.

C Supermartingales, superharmonic functions, and excessive measures

This section follows the exposition by Dynkin [Dy], which gives the clearest and best readable account of Martin boundary theory for denumerable Markov chains so far available in the literature (old and good, and certainly not obsolete). The method for proving Theorem 7.22 presented here, which combines the study of non-negative supermartingales with *time reversal,* goes back to the paper by Hunt [32].

Readers who are already familiar with martingale theory can skip the first part. Also, since our state space X is countable, we can limit ourselves to a very elementary approach to this theory.

I. Non-negative supermartingales

Let Ω be as in (7.21), with the associated σ-algebra $\mathcal{A}$ and the probability measure Pr (one of the measures Pr_x, $x \in X$). Even when P is stochastic, we shall need the additional absorbing state $\dagger$.

Let $Y_0, Y_1, \dots, Y_N$ be a finite sequence of random variables $\Omega \to X \cup \{\dagger\}$, and let $W_0, W_1, \dots, W_N$ be a sequence of real valued, composed random variables of the form

$$W_n = f_n(Y_0, \dots, Y_n), \quad \text{with } f_n \colon (X \cup \{\dagger\})^{n+1} \to [0, \infty).$$

7.23 Definition. The sequence $W_0, \dots, W_N$ is called a *supermartingale with respect to* $Y_0, \dots, Y_N$, if for each $n \in \{1, \dots, N\}$ one has

$$\mathsf{E}(W_n \mid Y_0, \dots, Y_{n-1}) \le W_{n-1} \quad \text{almost surely.}$$

Here, $\mathsf{E}(\cdot \mid Y_0, \dots, Y_{n-1})$ denotes conditional expectation with respect to the σ-algebra generated by $Y_0, \dots, Y_{n-1}$. On the practical level, the inequality means

that for all $x_0, \dots, x_{n-1} \in X \cup \{\dagger\}$ one has

$$\sum_{y \in X \cup \{\dagger\}} f_n(x_0, \dots, x_{n-1}, y)\ \mathsf{Pr}[Y_0 = x_0, \dots, Y_{n-1} = x_{n-1}, Y_n = y] \\ \leq f_{n-1}(x_0, \dots, x_{n-1})\ \mathsf{Pr}[Y_0 = x_0, \dots, Y_{n-1} = x_{n-1}], \tag{7.24}$$

or, equivalently,

$$\mathsf{E}\big(W_n\, g(Y_0, \dots, Y_{n-1})\big) \leq \mathsf{E}\big(W_{n-1}\, g(Y_0, \dots, Y_{n-1})\big) \tag{7.25}$$

for every function $g\colon (X \cup \{\dagger\})^n \to [0, \infty)$.

7.26 Exercise. Verify the equivalence between (7.24) and (7.25). Refresh your knowledge about conditional expectation by elaborating the equivalence of those two conditions with the supermartingale property. □

Clearly, Definition 7.23 also makes sense when the sequences Y_n and W_n are infinite ($N = \infty$). Setting $g = \mathbf{1}$, it follows from (7.25) that the expectations $\mathsf{E}(W_n)$ form a decreasing sequence. In particular, if W_0 is integrable, then so are all W_n.

Extending, or specifying, the definition given in Section 1.B, a random variable $\boldsymbol{t}$ with values in $\mathbb{N}_0 \cup \{\infty\}$ is called a *stopping time with respect to* $Y_0, \dots, Y_N$ (or with respect to the infinite sequence $Y_0, Y_1, \dots$), if for each integer $n \leq N$,

$$[\boldsymbol{t} \leq n] \in \mathcal{A}(Y_0, \dots, Y_n),$$

the σ-algebra generated by $Y_0, \dots, Y_n$. As usual, $W_{\boldsymbol{t}}$ denotes the random variable defined on the set $\{\omega \in \Omega : \boldsymbol{t}(\omega) < \infty\}$ by $W_{\boldsymbol{t}}(\lambda) = W_{\boldsymbol{t}(\omega)}(\omega)$. If $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ are two stopping times with respect to the Y_n, then so are $\boldsymbol{t}_1 \wedge \boldsymbol{t}_2 = \inf\{\boldsymbol{t}_1, \boldsymbol{t}_2\}$ and $\boldsymbol{t}_1 \vee \boldsymbol{t}_2 = \sup\{\boldsymbol{t}_1, \boldsymbol{t}_2\}$.

7.27 Lemma. *Let $\boldsymbol{s}$ and $\boldsymbol{t}$ be two stopping times with respect to the sequence (Y_n) such that $\boldsymbol{s} \leq \boldsymbol{t}$. Then*

$$\mathsf{E}(W_{\boldsymbol{s}}) \geq \mathsf{E}(W_{\boldsymbol{t}}).$$

Proof. Suppose first that W_0 is integrable and N is finite, so that $\boldsymbol{s} \leq \boldsymbol{t} \leq N$. As usual, we write $\mathbf{1}_A$ for the indicator function of an event $A \in \mathcal{A}$. We decompose

$$W_{\boldsymbol{s}} = \sum_{n=0}^{N} W_n\, \mathbf{1}_{[\boldsymbol{s}=n]} = W_n\, \mathbf{1}_{[\boldsymbol{s}\leq 0]} + \sum_{n=1}^{N} W_n\, (\mathbf{1}_{[\boldsymbol{s}\leq n]} - \mathbf{1}_{[\boldsymbol{s}\leq n-1]}) \\ = \sum_{n=0}^{N} W_n\, \mathbf{1}_{[\boldsymbol{s}\leq n]} - \sum_{n=0}^{N-1} W_{n+1}\, \mathbf{1}_{[\boldsymbol{s}\leq n]}.$$

We infer that $W_{\boldsymbol{s}}$ is integrable, since the sum is finite and the W_n are integrable. We decompose $W_{\boldsymbol{t}}$ in the same way and observe that $[\boldsymbol{s} \le N] = [\boldsymbol{t} \le N] = \Omega$, so that $\mathbf{1}_{[\boldsymbol{s}\le N]} = \mathbf{1}_{[\boldsymbol{t}\le N]}$. We obtain

$$W_{\boldsymbol{s}} - W_{\boldsymbol{t}} = \sum_{n=0}^{N-1} W_n(\mathbf{1}_{[\boldsymbol{s}\le n]} - \mathbf{1}_{[\boldsymbol{t}\le n]}) - \sum_{n=0}^{N-1} W_{n+1}(\mathbf{1}_{[\boldsymbol{s}\le n]} - \mathbf{1}_{[\boldsymbol{t}\le n]}).$$

For each n, the random variable $\mathbf{1}_{[\boldsymbol{s}\le n]} - \mathbf{1}_{[\boldsymbol{t}\le n]}$ is non-negative and measurable with respect to $\mathcal{A}(Y_0, \dots, Y_n)$. The latter σ-algebra is generated by the disjoint events (atoms) $[Y_0 = x_0, \dots, Y_n = x_n]$ $(x_0, \dots, x_n \in X)$, and every $\mathcal{A}(Y_0, \dots, Y_n)$-measurable function must be constant on each of those sets. (This fact also stands behind the equivalence between (7.24) and (7.25).) Therefore we can write

$$\mathbf{1}_{[\boldsymbol{s}\le n]} - \mathbf{1}_{[\boldsymbol{t}\le n]} = g_n(Y_0, \dots, Y_n),$$

where g_n is a non-negative function on $(X \cup \{\dagger\})^{n+1}$. Now (7.25) implies

$$\mathsf{E}\big(W_{n+1}(\mathbf{1}_{[\boldsymbol{s}\le n]} - \mathbf{1}_{[\boldsymbol{t}\le n]})\big) \le \mathsf{E}\big(W_n(\mathbf{1}_{[\boldsymbol{s}\le n]} - \mathbf{1}_{[\boldsymbol{t}\le n]})\big)$$

for every n, whence $\mathsf{E}(W_{\boldsymbol{s}} - W_{\boldsymbol{t}}) \ge 0$.

If W_0 does not have finite expectation, we can apply the preceding inequality to the supermartingale $(W_n \wedge c)_{n=0,\dots,N}$. By monotone convergence,

$$\mathsf{E}(W_{\boldsymbol{s}}) = \lim_{c\to\infty} \mathsf{E}\big((W \wedge c)_{\boldsymbol{s}}\big) \le \lim_{c\to\infty} \mathsf{E}\big((W \wedge c)_{\boldsymbol{t}}\big) = \mathsf{E}(W_{\boldsymbol{t}}).$$

Finally, if the sequence is infinite, we may apply the inequality to the stopping times $(\boldsymbol{s} \wedge N)$ and $(\boldsymbol{t} \wedge N)$ and use again monotone convergence, this time for $N \to \infty$. □

Let (r_n) be a finite or infinite sequence of real numbers, and let $[a, b]$ be an interval. Then the number of *downward crossings* of the interval by the sequence is

$$D_{\downarrow}\big((r_n) \,\big|\, [a, b]\big) = \sup\left\{ k \ge 0 : \begin{array}{c} \text{there are } n_1 \le n_2 \le \dots \le n_{2k} \\ \text{with } r_{n_i} \ge b \text{ for } i = 1, 3, \dots, 2k-1 \\ \text{and } r_{n_j} \le a \text{ for } j = 2, 4, \dots, 2k \end{array} \right\}.$$

In case the sequence is finite and terminates with r_N, one must require that $n_{2k} \le N$ in this definition, and the supremum is a maximum. For an infinite sequence,

$$\begin{gathered} \lim_{n\to\infty} r_n \in [-\infty, \infty] \quad \text{exists} \iff \\ D_{\downarrow}\big((r_n) \,\big|\, [a, b]\big) < \infty \quad \text{for every interval } [a, b]. \end{gathered} \tag{7.28}$$

Indeed, if $\liminf r_n < a < b < \limsup r_n$ then $D_\downarrow\big((r_n) \,\big|\, [a, b]\big) = \infty$. Observe that in this reasoning, it is sufficient to consider only the – countably many – intervals with rational endpoints.

Analogously, one defines the number of *upward crossings* of an interval by a sequence (notation: $D^\uparrow$).

7.29 Lemma. *Let (W_n) be a non-negative supermartingale with respect to the sequence (Y_n). Then for every interval $[a, b] \subset \mathbb{R}^+$*

$$\mathsf{E}\big(D_\downarrow\big((W_n) \,\big|\, [a, b]\big)\big) \le \frac{1}{b-a}\mathsf{E}(W_0).$$

Proof. We suppose first to have a finite supermartingale $W_0, \dots, W_N$ ($N < \infty$). We define a sequence of stopping times relative to $Y_0, \dots, Y_N$, starting with $\boldsymbol{t}_0 = 0$. If n is odd,

$$\boldsymbol{t}_n = \begin{cases} \min\{i \ge \boldsymbol{t}_{n-1} : W_i \ge b\}, & \text{if such } i \text{ exists;} \\ N, & \text{otherwise.} \end{cases}$$

If $n > 0$ is even,

$$\boldsymbol{t}_n = \begin{cases} \min\{j \ge \boldsymbol{t}_{n-1} : W_j \le a\}, & \text{if such } j \text{ exists;} \\ N, & \text{otherwise.} \end{cases}$$

Setting $\boldsymbol{d} = D_\downarrow\big(W_0, \dots, W_N \,\big|\, [a, b]\big)$, we get $\boldsymbol{t}_n = N$ for $n \ge 2\boldsymbol{d} + 2$. Furthermore, $\boldsymbol{t}_n = N$ also for $n \ge N$. We choose an integer $m \ge N/2$ and consider

$$\overline{W} = \underbrace{W_{\boldsymbol{t}_1} + \sum_{j=1}^{m} (W_{\boldsymbol{t}_{2j+1}} - W_{\boldsymbol{t}_{2j}})}_{(1)} = \underbrace{\sum_{i=1}^{\boldsymbol{d}} (W_{\boldsymbol{t}_{2i-1}} - W_{\boldsymbol{t}_{2i}}) + W_{\boldsymbol{t}_{2\boldsymbol{d}+1}}}_{(2)}.$$

(We have used the fact that $W_{\boldsymbol{t}_{2\boldsymbol{d}+2}} = W_{\boldsymbol{t}_{2\boldsymbol{d}+3}} = \cdots = W_{\boldsymbol{t}_{2m+1}} = W_N$.) Applying Lemma 7.27 to term (1) gives

$$\mathsf{E}(\overline{W}) = \mathsf{E}(W_{\boldsymbol{t}_1}) + \sum_{j=1}^{m} \big(\mathsf{E}(W_{\boldsymbol{t}_{2j+1}}) - \mathsf{E}(W_{\boldsymbol{t}_{2j}})\big) \le \mathsf{E}(W_{\boldsymbol{t}_1}) \le \mathsf{E}(W_0).$$

The expression (2) leads to $\overline{W} \ge (b-a)\,\boldsymbol{d}$ and thus also to

$$\mathsf{E}(\overline{W}) \ge (b-a)\mathsf{E}(\boldsymbol{d}).$$

Combining these inequalities, we find that

$$\mathsf{E}\big(D_\downarrow(W_0, \dots, W_N \,\big|\, [a, b])\big) \le \frac{1}{b-a}\mathsf{E}(W_0).$$

If the sequences (Y_n) and (W_n) are infinite, we can let N tend to ∞ in the latter inequality, and (always by monotone convergence) the proposed statement follows. □

II. Supermartingales and superharmonic functions

As an application of Lemma 7.29, we obtain the limit theorem for non-negative supermartingales.

7.30 Theorem. *Let $(W_n)_{n\geq 0}$ be a non-negative supermartingale with respect to $(Y_n)_{n\geq 0}$ such that $\mathsf{E}(W_0) < \infty$. Then there is an integrable (whence almost surely finite) random variable W_∞ such that*

$$\lim_{n\to\infty} W_n = W_\infty \quad \textit{almost surely.}$$

Proof. Let $[a_i, b_i]$, $i \in \mathbb{N}$, be an enumeration of all intervals with non-negative rational endpoints. For each i, let

$$\boldsymbol{d}_i = D_\downarrow\big((W_n) \mid [a_i, b_i]\big).$$

By Lemma 7.29, each $\boldsymbol{d}_i$ is integrable and thus almost surely finite. Let

$$\overline{\Omega} = \bigcap_{i\in\mathbb{N}} [\boldsymbol{d}_i < \infty].$$

Then $\mathsf{Pr}(\overline{\Omega}) = 1$, and for every $\omega \in \overline{\Omega}$, each interval $[a_i, b_i]$ is crossed downwards only finitely many times by $\big(W_n(\omega)\big)$. By (7.28), there is $W_\infty = \lim_n W_n \in [0, \infty]$ almost surely. By Fatou's lemma, $\mathsf{E}(W_\infty) \leq \lim_n \mathsf{E}(W_n) \leq \mathsf{E}(W_0) < \infty$. Consequently, W_∞ is a.s. finite. □

This theorem applies to positive superharmonic functions. Consider the Markov chain $(Z_n)_{n\geq 0}$. Recall that $\mathsf{Pr} = \mathsf{Pr}_x$ some $x \in X$. Let $f : X \to \mathbb{R}$ be a non-negative function. We extend f to $X \cup \{\dagger\}$ by setting $f(\dagger) = 0$. By (7.24), the sequence of real-valued random variables $\big(f(Z_n)\big)_{n\geq 0}$ is a supermartingale with respect to $(Z_n)_{n\geq 0}$ if and only if for every n and all $x_0, \ldots, x_{n-1}$

$$\sum_{y\in X\cup\{\dagger\}} \delta_x(x_0)\, p(x_0, x_1) \cdots p(x_{n-2}, x_{n-1})\, p(x_{n-1}, y)\, f(y)$$
$$\leq \delta_x(x_0)\, p(x_0, x_1) \cdots p(x_{n-2}, x_{n-1})\, f(x_{n-1}),$$

that is, if and only if f is superharmonic.

7.31 Corollary. *If $f \in \mathcal{S}^+(X, P)$ then $\lim_{n\to\infty} f(Z_n)$ exists and is Pr_x-almost surely finite for every $x \in X$.*

If P is strictly substochastic in some point, then the probability that Z_n "dies" (becomes absorbed by $\dagger$) is positive, and on the corresponding set $\Omega_\dagger$ of trajectories, $f(Z_n)$ tends to 0. What is interesting for us is that in any case, the set of trajectories in $X^{\mathbb{N}_0}$ along which $f(Z_n)$ does not converge has measure 0.

7.32 Exercise. Prove that for all $x, y \in X$,

$$\lim_{n\to\infty} G(Z_n, y) = 0 \quad \mathsf{Pr}_x\text{-almost surely.}$$ □

III. Supermartingales and excessive measures

A specific example of a positive superharmonic function is $K(\cdot, y)$, where $y \in X$. Therefore, $K(Z_n, y)$ converges almost surely, the limit is 0 by Exercise 7.32. However, in order to prove Theorem 7.22, we must verify instead that $\mathsf{Pr}_x(\Omega_{\boldsymbol{\epsilon}}) = 1$, or equivalently, that $\lim_{n\to\boldsymbol{\epsilon}} K(y, Z_n)$ exists Pr_x-almost surely for all $x, y \in X$. We shall first prove this with respect to Pr_o (where the starting point is the same "origin" o as in the definition of the Martin kernel).

Recall that we are thinking of functions on X as column vectors and of measures as row vectors. In particular, $G(x, \cdot)$ is a measure on X. Now let μ be an arbitrary probability measure on X. We observe that

$$\mu\, G(y) = \sum_x \mu(x)\, G(x, y) \le \sum_x \mu(x)\, G(y, y) = G(y, y)$$

is finite for every y. Then $\nu = \mu\, G$ is an excessive measure by the dual of Lemma 6.42 (b). Furthermore, we can write

$$\mu\, G(y) = f_\mu(y)\, G(o, y), \quad \text{where } f_\mu(y) = \sum_x \mu(x)\, K(x, y). \tag{7.33}$$

We may consider the function f_μ as the density of the excessive measure $\mu\, G$ with respect to the measure $G(o, \cdot)$. We extend f_μ to $X \cup \{\dagger\}$ by setting $f_\mu(\dagger) = 0$.

We choose a finite subset $V \subset X$ that contains the origin o. As above, we define the exit time of V:

$$\boldsymbol{\epsilon}_V = \sup\{n : Z_n \in V\}. \tag{7.34}$$

Contrary to $\boldsymbol{\epsilon} = \boldsymbol{t}^\dagger - 1$, this is not a stopping time, as the property that $\boldsymbol{\epsilon}_V = k$ requires that $Z_n \notin V$ for all $n > k$. Since our Markov chain is transient and $o \in V$,

$$\mathsf{Pr}_o[0 \le \boldsymbol{\epsilon}_V < \infty] = 1,$$

that is, $\mathsf{Pr}_o(\Omega_V) = 1$, where $\Omega_V = \{\omega \in \Omega : 0 \le \boldsymbol{\epsilon}_V(\omega) < \infty\}$. Observe that for $x \notin V$, it can occur with positive probability that (Z_n) never enters V, in which case $\boldsymbol{\epsilon}_V = \infty$, while $\boldsymbol{\epsilon}_V$ is finite only for those trajectories starting from x that visit V.

Given an arbitrary interval $[a, b]$, we want to control the number of its upward crossings by the sequence $f_\mu(Z_0), \ldots, f_\mu(Z_{\boldsymbol{\epsilon}_V})$, where f_μ is as in (7.33). Note that this sequence has a random length that is almost surely finite. For $n \ge 0$ and $\omega \in \Omega_V$ we set

$$Z_{\boldsymbol{\epsilon}_V - n}(\omega) = \begin{cases} Z_{\boldsymbol{\epsilon}_V(\omega) - n}(\omega), & n \le \boldsymbol{\epsilon}(\omega); \\ \dagger, & n > \boldsymbol{\epsilon}(\omega). \end{cases}$$

Then with probability 1 (that is, on the event $[\boldsymbol{\epsilon}_V < \infty]$)

$$\begin{aligned} &D^{\uparrow}\Big(f_\mu(Z_0), f_\mu(Z_1), \dots, f_\mu(Z_{\boldsymbol{\epsilon}_V}) \,\Big|\, [a, b]\Big) \\ &\quad = \lim_{N\to\infty} D_{\downarrow}\Big(f_\mu(Z_{\boldsymbol{\epsilon}_V}), f_\mu(Z_{\boldsymbol{\epsilon}_V - 1}), \dots, f_\mu(Z_{\boldsymbol{\epsilon}_V - N}) \,\Big|\, [a, b]\Big). \end{aligned} \tag{7.35}$$

With these ingredients we obtain the following

7.36 Proposition. (1) $\mathsf{E}_o\big(f_\mu(Z_{\boldsymbol{\epsilon}_V})\big) \le 1$.

(2) *$f_\mu(Z_{\boldsymbol{\epsilon}_V})$, $f_\mu(Z_{\boldsymbol{\epsilon}_V - 1}), \dots, f_\mu(Z_{\boldsymbol{\epsilon}_V - N})$ is a supermartingale with respect to $Z_{\boldsymbol{\epsilon}_V}, Z_{\boldsymbol{\epsilon}_V - 1}, \dots, Z_{\boldsymbol{\epsilon}_V - N}$ and the measure Pr_o on the trajectory space.*

Proof. First of all, we compute for $x, y \in V$

$$\mathsf{Pr}_x[Z_{\boldsymbol{\epsilon}_V} = y] = \sum_{n=0}^{\infty} \mathsf{Pr}_x[\boldsymbol{\epsilon}_V = n,\ Z_n = y].$$

(Note that more precisely, we should write $\mathsf{Pr}_x[0 \le \boldsymbol{\epsilon}_V < \infty,\ Z_{\boldsymbol{\epsilon}_V} = y]$ in the place of $\mathsf{Pr}_x[Z_{\boldsymbol{\epsilon}_V} = y]$.) The event $[\boldsymbol{\epsilon}_V = n]$ depends only on those Z_k with $k \ge n$ (the future) , and not on $Z_0, \dots, Z_{n-1}$ (the past). Therefore $\mathsf{Pr}_x[\boldsymbol{\epsilon}_V = n,\ Z_n = y] = p^{(n)}(x, y)\ \mathsf{Pr}_y[\boldsymbol{\epsilon}_V = 0]$, and

$$\mathsf{Pr}_x[Z_{\boldsymbol{\epsilon}_V} = y] = G(x, y)\ \mathsf{Pr}_y[\boldsymbol{\epsilon}_V = 0]. \tag{7.37}$$

Taking into account that $0 \le \boldsymbol{\epsilon}_V < \infty$ almost surely, we now get

$$\begin{aligned} \mathsf{E}_o\big(f_\mu(Z_{\boldsymbol{\epsilon}_V})\big) &= \sum_{y\in V} f_\mu(y)\ \mathsf{Pr}_o[Z_{\boldsymbol{\epsilon}_V} = y] \\ &= \sum_{y\in V} f_\mu(y)\, G(o, y)\ \mathsf{Pr}_y[\boldsymbol{\epsilon}_V = 0] \\ &= \sum_{y\in V} \mu\, G(y)\ \mathsf{Pr}_y[\boldsymbol{\epsilon}_V = 0] \\ &= \sum_{x\in X} \mu(x) \sum_{y\in V} G(x, y)\ \mathsf{Pr}_y[\boldsymbol{\epsilon}_V = 0] \\ &= \sum_{x\in X} \mu(x) \sum_{y\in V} \mathsf{Pr}_x[Z_{\boldsymbol{\epsilon}_V} = y] \le 1. \end{aligned}$$

This proves (1). To verify (2), we first compute for $x_0 \in V$, $x_1 \ldots, x_n \in X$

$$\begin{aligned}
&\mathsf{Pr}_o[Z_{\boldsymbol{\epsilon}_V} = x_0,\ Z_{\boldsymbol{\epsilon}_V - 1} = x_1, \ldots, Z_{\boldsymbol{\epsilon}_V - n} = x_n] \\
&\qquad \text{(since we have } \boldsymbol{\epsilon}_V \geq n, \text{ if } Z_{\boldsymbol{\epsilon}_V - n} \neq \dagger) \\
&\qquad = \sum_{k=n}^{\infty} \mathsf{Pr}_o[\boldsymbol{\epsilon}_V = k,\ Z_k = x_0,\ Z_{k-1} = x_1, \ldots, Z_{k-n} = x_n] \\
&\qquad = \sum_{k=n}^{\infty} p^{(k-n)}(o, x_n)\, p(x_n, x_{n-1}) \cdots p(x_1, x_0)\ \mathsf{Pr}_{x_0}[\boldsymbol{\epsilon}_V = 0] \\
&\qquad = G(o, x_n)\, p(x_n, x_{n-1}) \cdots p(x_1, x_0)\ \mathsf{Pr}_{x_0}[\boldsymbol{\epsilon}_V = 0].
\end{aligned}$$

We now check (7.24) with $f_n(x_0, \ldots, x_n) = f_\mu(x_n)$. Since $f_\mu(\dagger) = 0$, we only have to sum over elements $x \in X$. Furthermore, if $Y_n = Z_{\boldsymbol{\epsilon}_V - n} \in X$ then we must have $\boldsymbol{\epsilon}_V \geq n$ and $Z_{\boldsymbol{\epsilon}_V - k} \in X$ for $k = 0, \ldots, n$. Hence it is sufficient to consider only the case when $x_0, \ldots, x_{n-1} \in X$:

$$\begin{aligned}
&\sum_{x \in X} f_\mu(x)\ \mathsf{Pr}_0[Z_{\boldsymbol{\epsilon}_V} = x_0, \ldots, Z_{\boldsymbol{\epsilon}_V - n + 1} = x_{n-1},\ Z_{\boldsymbol{\epsilon}_V - n} = x] \\
&\qquad = \sum_{x \in X} f_\mu(x)\, G(o, x)\, p(x, x_{n-1}) \cdots p(x_1, x_0)\ \mathsf{Pr}_{x_0}[\boldsymbol{\epsilon}_V = 0] \\
&\qquad = \Big(\sum_{x \in X} \mu\, G(x)\, p(x, x_{n-1}) \Big) p(x_{n-1}, x_{n-2}) \cdots p(x_1, x_0)\ \mathsf{Pr}_{x_0}[\boldsymbol{\epsilon}_V = 0] \\
&\qquad \leq \big(\mu\, G(x_{n-1}) \big) p(x_{n-1}, x_{n-2}) \cdots p(x_1, x_0)\ \mathsf{Pr}_{x_0}[\boldsymbol{\epsilon}_V = 0] \\
&\qquad = f_\mu(x_{n-1})\ \mathsf{Pr}_o[Z_{\boldsymbol{\epsilon}_V} = x_0, \ldots, Z_{\boldsymbol{\epsilon}_V - n + 1} = x_{n-1}].
\end{aligned}$$

In the inequality we have used that $\mu\, G$ is an excessive measure. □

This proposition provides the main tool for the proof of Theorem 7.22.

7.38 Corollary.

$$\mathsf{E}_o\Big(D^{\uparrow}\big((f_\mu(Z_n))_{n \leq \boldsymbol{\epsilon}} \,\big|\, [a, b] \big) \Big) \leq \frac{1}{b - a}.$$

Proof. If $V \subset X$ is finite, then (7.35), Lemma 7.29 and Proposition 7.36 imply, by virtue of the monotone convergence theorem, that

$$\begin{aligned}
\mathsf{E}_o\big(D^{\uparrow}\big(f_\mu(Z_0),\ f_\mu(Z_1),\ \ldots,\ f_\mu(Z_{\boldsymbol{\epsilon}_V}) \,\big|\, [a, b] \big) \big) &\leq \frac{1}{b - a} \mathsf{E}_o\big(f_\mu(Z_{\boldsymbol{\epsilon}_V}) \big) \\
&\leq \frac{1}{b - a}.
\end{aligned}$$

We choose a sequence of finite subsets V_k of X containing o, such that $V_k \subset V_{k+1}$ and $\bigcup_k V_k = X$. Then $\lim_{k\to\infty} \boldsymbol{\epsilon}_{V_k} = \boldsymbol{\epsilon}$, and

$$\lim_{k\to\infty} D^{\uparrow}\Big(\big(f_\mu(Z_n)\big)_{n\le \boldsymbol{\epsilon}_{V_k}} \,\Big|\, [a,\, b]\Big) = D^{\uparrow}\Big(\big(f_\mu(Z_n)\big)_{n\le \boldsymbol{\epsilon}} \,\Big|\, [a,\, b]\Big).$$

Using monotone convergence once more, the result follows. □

IV. Proof of the boundary convergence theorem

Now we can finally prove Theorem 7.22. We have to prove the statements (a), (b), (c) listed after the theorem. We start with (a), $\Omega_{\boldsymbol{\epsilon}} \in \mathcal{A}$, whose proof is a standard exercise in measure theory. (In fact, we have previously omitted such detailed considerations on several occasions, but it is good to go through this explicitly on at least one occasion.)

It is clear that $\Omega_\dagger \in \mathcal{A}$, since it is a countable union of basic cylinder sets. On the other hand, we now prove that Ω_∞ can be obtained by countably many intersections and unions, starting with cylinder sets in $\mathcal{A}$. First of all $\Omega_\infty = \bigcap_x \Omega_x$, where

$$\Omega_x = \big\{\omega = (x_n) \in X^{\mathbb{N}_0} : \lim_{n\to\infty} K(x, x_n) \text{ exists in } \mathbb{R}\big\}.$$

Now $\big(K(x, x_n)\big)$ is a bounded sequence, whence by (7.28)

$$\Omega_x = \bigcap_{[a,b] \text{ rational}} A_x([a,\, b]), \quad \text{where}$$
$$A_x([a,\, b]) = \Big\{(x_n) : D_{\downarrow}\big(\big(K(x, x_n)\big) \,\big|\, [a,\, b]\big) < \infty\Big\}.$$

We show that $\Omega_\infty \setminus A_x([a,\, b]) \in \mathcal{A}$ for any fixed interval $[a,\, b]$: this is

$$\Big\{(x_n) : D_{\downarrow}\big(\big(K(x, x_n)\big) \,\big|\, [a,\, b]\big) = \infty\Big\}$$
$$= \bigcap_k \bigcup_{l,m\ge k} \big(\{(x_n) : K(x, x_l) \ge b\} \cap \{(x_n) : K(x, x_m) \le a\}\big).$$

Each set $\{(x_n) : K(x, x_l) \ge b\}$ depends only on the value $K(x, x_l)$ and is the union of all cylinder sets of the form $\mathsf{C}(y_0, \dots, y_l)$ with $K(x, y_l) \ge b$. Analogously, $\{(x_n) : K(x, x_m) \le a\}$ is the union of all cylinder sets $\mathsf{C}(y_0, \dots, y_m)$ with $K(x, y_m) \le a$.

(b) We set $\mu = \delta_x$ and apply Corollary 7.38: $f_\mu(y) = K(x, y)$, whence

$$\lim_{n\to\boldsymbol{\epsilon}} K(x, Z_n) \quad \text{exists } \mathsf{Pr}_o\text{-almost surely for each } x.$$

Therefore, (Z_n) converges Pr_o-almost surely in the topology of $\widehat{X}(P)$. In other terms, $\mathsf{Pr}_o(\Omega_\epsilon) = 1$. In order to see that the initial point o can be replaced with any $x_0 \in X$, we use irreducibility. There are $k \geq 0$ and $y_1, \dots, y_{k-1} \in X$ such that

$$p(o, y_1)p(y_1, y_2)\cdots p(y_{k-1}, x_0) > 0.$$

Therefore

$$\begin{aligned} &p(o, y_1)\, p(y_1, y_2)\cdots\, p(y_{k-1}, x_0)\ \mathsf{Pr}_{x_0}(\Omega \setminus \Omega_\epsilon) \\ &\quad = p(o, y_1)\, p(y_1, y_2)\cdots p(y_{k-1}, x_0)\ \mathsf{Pr}_{x_0}\big(\mathsf{C}(x_0) \cap (\Omega \setminus \Omega_\epsilon)\big) \\ &\quad = \mathsf{Pr}_o\big(\mathsf{C}(o, y_1, \dots, y_{k-1}, x_0) \cap (\Omega \setminus \Omega_\epsilon)\big) = 0. \end{aligned}$$

We infer that $\mathsf{Pr}_{x_0}(\Omega \setminus \Omega_\epsilon) = 0$.

(c) Since X is discrete in the topology of $\widehat{X}(P)$, it is clear that the restriction of Z_ϵ to $\Omega_\dagger$ is measurable. We prove that also the restriction to Ω_∞ is measurable with respect to the Borel σ-algebra on $\mathcal{M}$. In view of the construction of the Martin boundary (see in particular the approach using the completion of the metric (7.15) on X), a base of the topology is given by the collection of all finite intersections of sets of the form

$$B_{x,\xi,\varepsilon} = \big\{\eta \in \mathcal{M} : |K(x, \eta) - K(x, \xi)| < \varepsilon\big\},$$

where $x \in X$, $\xi \in \mathcal{M}$ and $\varepsilon > 0$ vary. We prove that $[Z_\epsilon \in B_{x,\xi,\varepsilon}] \in \mathcal{A}$ for each of those sets. Write $c = K(x, \xi)$. Then

$$\begin{aligned} [Z_\epsilon \in B_{x,\xi,\varepsilon}] &= \big\{\omega = (x_n) \in \Omega_\infty : |K(x, x_\infty) - c| < \varepsilon\big\} \\ &= \big\{\omega = (x_n) \in \Omega_\infty : \big|\lim_{n\to\infty} K(x, x_n) - c\big| < \varepsilon\big\}. \end{aligned}$$

7.39 Exercise. Prove in analogy with (a) that the latter set belongs to the σ-algebra $\mathcal{A}$. □

This concludes the proof of Theorem 7.22. □

Theorem 7.19 follows immediately from Theorem 7.22. Indeed, if P is stochastic then $\mathsf{Pr}_x(\Omega_\dagger) = 0$ for every $x \in X$, and $\epsilon = \infty$. In this case, adding $\dagger$ to the state space is not needed and is just convenient for the technical details of the proofs.

D The Poisson–Martin integral representation theorem

Theorems 7.19 and 7.22 show that the Martin compactification $\widehat{X} = \widehat{X}(P)$ provides a "geometric" model for the limit points of the Markov chain (always considering the transient case). With respect to each starting point $x \in X$, we consider the distribution ν_x of the random variable Z_ϵ: for a Borel set $B \subset \widehat{X}$,

$$\nu_x(B) = \mathsf{Pr}_x[Z_\epsilon \in B].$$

In particular, if $y \in X$, then (7.37) (with $V = X$) yields

$$\nu_x(y) = G(x,y)\Big(1 - \sum_{w\in X} p(y,w)\Big) = G(x,y)\, p(y,\dagger).^{1} \tag{7.40}$$

If $f : \widehat{X} \to \mathbb{R}$ is ν_x-integrable then

$$\mathsf{E}_x\big(f(Z_{\boldsymbol{\epsilon}})\big) = \int_{\widehat{X}} f\, d\nu_x. \tag{7.41}$$

7.42 Theorem. *The measure ν_x is absolutely continuous with respect to ν_o, and (a realization of) its Radon–Nikodym density is given by $\dfrac{d\nu_x}{d\nu_o} = K(x,\cdot)$. Namely, if $B \subset \widehat{X}$ is a Borel set then*

$$\nu_x(B) = \int_B K(x,\cdot)\, d\nu_o.$$

Proof. As above, let V be a finite subset of X and $\boldsymbol{\epsilon}_V$ the exit time from V. We assume that $o, x \in V$. Applying formula (7.37) once with starting point x and once with starting point o, we find

$$\mathsf{Pr}_x[Z_{\boldsymbol{\epsilon}_V} = y] = K(x,y)\ \mathsf{Pr}_o[Z_{\boldsymbol{\epsilon}_V} = y]$$

for every $y \in V$. Let $f : \widehat{X} \to \mathbb{R}$ be a continuous function. Then

$$\begin{aligned}\mathsf{E}_x\big(f(Z_{\boldsymbol{\epsilon}_V})\big) &= \sum_{y\in V} f(y)\ \mathsf{Pr}_x[Z_{\boldsymbol{\epsilon}_V} = y] \\ &= \sum_{y\in V} f(y)\, K(x,y)\ \mathsf{Pr}_o[Z_{\boldsymbol{\epsilon}_V} = y] = \mathsf{E}_o\big(f(Z_{\boldsymbol{\epsilon}_V})\, K(x, Z_{\boldsymbol{\epsilon}_V})\big).\end{aligned}$$

We now take, as above, an increasing sequence of finite sets V_k with limit (union) X. Then $\lim_k Z_{\boldsymbol{\epsilon}_{V_k}} = Z_{\boldsymbol{\epsilon}}$ almost surely with respect to Pr_x and Pr_o. Since f and $K(x,\cdot)$ are continuous functions on the compact set $\widehat{X}$, Lebesgue's dominated convergence theorem implies that one can exchange limit and expectation. Thus

$$\mathsf{E}_x\big(f(Z_{\boldsymbol{\epsilon}})\big) = \mathsf{E}_o\big(f(Z_{\boldsymbol{\epsilon}})\, K(x, Z_{\boldsymbol{\epsilon}})\big),$$

that is,

$$\int_{\widehat{X}} f(\xi)\, d\nu_x(\xi) = \int_{\widehat{X}} f(\xi)\, K(x,\xi)\, d\nu_o(\xi)$$

for every continuous function $f : \widehat{X} \to \mathbb{R}$. Since the indicator functions of open sets in $\widehat{X}$ can be approximated by continuous functions, it follows that

$$\nu_x(B) = \int_B K(x,\cdot)\, d\nu_o$$

[1] For any measure ν, we always write $\nu(w) = \nu(\{w\})$ for the mass of a singleton.

for every open set B. But the open sets generate the Borel σ-algebra, and the result follows.

(We can also use the following reasoning: two Borel measures on a compact metric space coincide if and only if the integrals of all continuous functions coincide. For more details regarding Borel measures on metric spaces, see for example the book by PARTHASARATHY [Pa].) □

We observe that by irreducibility, also ν_o is absolutely continuous with respect to ν_x, with Radon–Nikodym density $1/K(x,\cdot)$. Thus, all the limit measures ν_x, $x \in X$, are mutually absolutely continuous. We add another useful proposition involving the measures ν_x.

7.43 Proposition. *If $f : \widehat{X} \to \mathbb{R}$ is a continuous function then*

$$\begin{aligned}\mathsf{E}_x\big(f(Z_\epsilon)\big) &= \sum_{y\in X} f(y)\,\nu_x(y) + \lim_{n\to\infty} P^n f(x)\\ &= \sum_{y\in X} f(y)\,G(x,y)\,p(y,\dagger) + \lim_{n\to\infty} P^n f(x).\end{aligned}$$

Proof. We decompose

$$\mathsf{E}_x\big(f(Z_\epsilon)\big) = \mathsf{E}_x\big(f(Z_\epsilon)\,\mathbf{1}_{\Omega_\dagger}\big) + \mathsf{E}_x\big(f(Z_\epsilon)\,\mathbf{1}_{\Omega_\infty}\big).$$

The first term can be rewritten as

$$\mathsf{E}_x\big(f(Z_\epsilon)\,\mathbf{1}_{[\epsilon<\infty]}\big) = \sum_{y\in X} f(y)\ \mathsf{Pr}_x[\epsilon<\infty,\ Z_\epsilon = y] = \sum_{y\in X} f(y)\,\nu_x(y).$$

Using continuity of f and dominated convergence, the second term can be written as

$$\lim_{n\to\infty} \mathsf{E}_x\big(f(Z_n)\,\mathbf{1}_{[\epsilon\ge n]}\big).$$

On the set $[\epsilon \ge n]$ we have $Z_k \in X$ for each $k \le n$. Hence

$$\mathsf{E}_x\big(f(Z_n)\,\mathbf{1}_{[\epsilon\ge n]}\big) = \sum_{y\in X} f(y)\ \mathsf{Pr}_x[Z_n = y] = P^n f(x).$$

Combining these relations, we obtain the first of the proposed identities. The second one follows from (7.40). □

The *support* of a (non-negative) Borel measure ν is the set

$$\operatorname{supp}(\nu) = \{\xi : \nu(V) > 0 \text{ for every neighbourhood } V \text{ of } \xi\}.$$

If we set $h = \int_{\hat{X}} K(\cdot, \xi)\, d\nu(\xi)$ then

$$Ph = \int_{\hat{X}} PK(\cdot, \xi)\, d\nu(\xi). \tag{7.44}$$

Indeed, if P has finite range at $x \in X$ then $Ph(x) = \sum_y p(x, y)h(y)$ is a finite sum which can be exchanged with the integral. Otherwise, we choose an enumeration y_k, $k = 1, 2, \ldots$, of the y with $p(x, y) > 0$. Then

$$\sum_{k=1}^{n} p(x, y_k)h(y_k) = \int_{\hat{X}} \sum_{k=1}^{n} p(x, y_k)\, K(y_k, \xi)\, d\nu(\xi)$$

for every n. Using monotone convergence as $n \to \infty$, we get (7.44). In particular, h is a superharmonic function.

We have now arrived at the point where we can prove the second main theorem of Martin boundary theory, after the one concerning convergence to the boundary.

7.45 Theorem (Poisson–Martin integral representation). *Let (X, P) be substochastic, irreducible and transient, with Martin compactification $\hat{X}$ and Martin boundary $\mathcal{M}$. Then for every function $h \in \mathcal{S}^+(X, P)$ there is a Borel measure ν^h on $\hat{X}$ such that*

$$h(x) = \int_{\hat{X}} K(x, \cdot)\, d\nu^h \quad \textit{for every } x \in X.$$

If h is harmonic then $\operatorname{supp}(\nu^h) \subset \mathcal{M}$.

Proof. We exclude the trivial case $h \equiv 0$. Then we know (from the minimum principle) that $h(x) > 0$ for every x, and we can consider the h-process (7.7). By (7.8), the Martin kernel associated with P_h is $K_h(x, y) = K(x, y)h(o)/h(x)$. In view of the properties that characterize the Martin compactification, we see that $\hat{X}(P_h) = \hat{X}(P)$, and that for every $x \in X$

$$K_h(x, \cdot) = K(x, \cdot)\, \frac{h(o)}{h(x)} \quad \text{on } \hat{X}. \tag{7.46}$$

Let $\tilde{\nu}_x$ be the distribution of Z_ϵ with respect to the h-process with starting point x, that is, $\tilde{\nu}_x(B) = \mathsf{Pr}_x^h[Z_\epsilon \in B]$. [At this point, we recall once more that when working with the trajectory space, the mappings Z_n and Z_ϵ defined on the latter do not change when we consider a modified process; what changes is the probability measure on the trajectory space.] We apply Theorem 7.42 to the h-process, setting $B = \hat{X}$, and use the fact that $K_h(x, \cdot) = d\tilde{\nu}_x / d\tilde{\nu}_o$:

$$1 = \tilde{\nu}_x(\hat{X}) = \int_{\hat{X}} K_h(x, \cdot)\, d\tilde{\nu}_o.$$

We set $\nu^h = h(o)\,\tilde{\nu}_o$ and multiply by $h(x)$. Then (7.46) implies the proposed integral representation $h(x) = \int_{\widehat{X}} K(x,\cdot)\,d\nu^h$.

Let now h be a harmonic function. Suppose that $y \in \operatorname{supp}(\nu^h)$ for some $y \in X$. Since X is discrete in $\widehat{X}$, we must have $\nu_h(y) > 0$. We can decompose $\nu^h = a\cdot\delta_y + \nu'$, where $\operatorname{supp}(\nu') \subset \widehat{X}\setminus\{y\}$. Then we get $h(x) = a\,K(x,y) + h'(x)$, where $h'(x) = \int_{\widehat{X}} K(x,\cdot)\,d\nu'$. But $K(\cdot,y)$ and h' are superharmonic functions, and the first of the two is strictly superharmonic in y. Therefore also h must be strictly superharmonic in y, a contradiction. □

The proof has provided us with a natural choice for the measure ν^h in the integral representation: for a Borel set $B \subset \widehat{X}$

$$\nu^h(B) = h(o)\ \mathsf{Pr}_o^h[Z_\epsilon \in B]. \tag{7.47}$$

7.48 Lemma. *Let h_1, h_2 be two strictly positive superharmonic functions, let $a_1, a_2 > 0$ and $h = a_1\cdot h_1 + a_2\cdot h_2$. Then $\nu^h = a_1\cdot\nu^{h_1} + a_2\cdot\nu^{h_2}$.*

Proof. Let $o = x_0, x_1, \dots, x_k \in X \cup \{\dagger\}$. Then, by construction of the h-process,

$$\begin{aligned}
&h(o)\ \mathsf{Pr}_o^h[Z_0 = x_0, \dots, Z_k = x_k]\\
&= h(o)\,\frac{p(x_0,x_1)h(x_1)}{h(x_0)} \cdots \frac{p(x_{k-1},x_k)h(x_k)}{h(x_{k-1})}\\
&= p(x_0,x_1)\cdots p(x_{k-1},x_k)\,h(x_k)\\
&= a_1\,p(x_0,x_1)\cdots p(x_{k-1},x_k)\,h_1(x_k) + a_2\,p(x_0,x_1)\cdots p(x_{k-1},x_k)\,h_2(x_k)\\
&= a_1h_1(o)\ \mathsf{Pr}_o^{h_1}[Z_0 = x_0, \dots, Z_k = x_k]\\
&\qquad + a_2h_2(o)\ \mathsf{Pr}_o^{h_2}[Z_0 = x_o, \dots, Z_k = x_k].
\end{aligned}$$

We see that the identity between measures

$$h(o)\ \mathsf{Pr}_o^h = a_1\,h_1(o)\cdot\mathsf{Pr}_o^{h_1} + a_2\,h_2(o)\cdot\mathsf{Pr}_o^{h_2}$$

is valid on all cylinder sets, and therefore on the whole σ-algebra $\mathcal{A}$. If B is a Borel set in $\widehat{X}$ then we get

$$\begin{aligned}
\nu^h(B) &= h(o)\ \mathsf{Pr}_o^h[Z_\epsilon \in B]\\
&= a_1h_1(o)\ \mathsf{Pr}_o^{h_1}[Z_\epsilon \in B] + a_2h_2(o)\ \mathsf{Pr}_o^{h_2}[Z_\epsilon \in B]\\
&= a_1\,\nu^{h_1}(B) + a_2\,\nu^{h_2}(B),
\end{aligned}$$

as proposed. □

7.49 Exercise. Let $h \in \mathcal{S}^+(X,P)$. Use (7.8), (7.40) and (7.47) to show that

$$\nu^h(y) = G(o,y)\big(h(y) - Ph(y)\big).$$

In particular, let $y \in X$ and $h = K(\cdot,y)$. Show that $\nu^h = \delta_y$. □

In general, the measure in the integral representation of a positive (super)harmonic function is not necessarily unique. We still have to face the question under which additional properties it does become unique. Prior to that, we show that every minimal harmonic function is of the form $K(\cdot, \xi)$ with $\xi \in \mathcal{M}$ – always under the hypotheses of (sub)stochasticity, irreducibility and transience.

7.50 Theorem. *Let h be a minimal harmonic function. Then there is a point $\xi \in \mathcal{M}$ such that the unique measure ν on $\widehat{X}$ which gives rise to an integral representation $h = \int_{\widehat{X}} K(\cdot, \eta)\, d\nu(\eta)$ is the point mass $\nu = \delta_\xi$. In particular,*

$$h = K(\cdot, \xi).$$

Proof. Suppose that we have

$$h = \int_{\widehat{X}} K(\cdot, \eta)\, d\nu(\eta).$$

By Theorem 7.45, such an integral representation does exist. We have $\nu(\widehat{X}) = 1$ because $h(o) = K(o, \eta) = 1$ for all $\eta \in \widehat{X}$. Suppose that $B \subset \widehat{X}$ is a Borel set with $0 < \nu(B) < 1$. Set

$$h_B(x) = \frac{1}{\nu(B)} \int_B K(x, \eta)\, d\nu(\eta)$$

and

$$h_{\widehat{X}\setminus B}(x) = \frac{1}{\nu(\widehat{X} \setminus B)} \int_{\widehat{X}\setminus B} K(x, \eta)\, d\nu(\eta).$$

Then h_B and $h_{\widehat{X}\setminus B}$ are positive superharmonic with value 1 at o, and

$$h = \nu(B) \cdot h_B + \big((1 - \nu(B)\big) \cdot h_{\widehat{X}\setminus B}$$

is a convex combination of two functions in the base $\mathcal{B}$ of the cone $\mathcal{S}^+$. Therefore we must have $h = h_B = h_{\widehat{X}\setminus B}$. In particular,

$$\int_B h(x)\, d\nu(\eta) = \nu(B)\, h(x) = \int_B K(x, \eta)\, d\nu(\eta)$$

for every $x \in X$ and every Borel set $B \subset X$ (if $\nu(B) = 0$ or $\nu(B) = 1$, this is trivially true). It follows that for each $x \in X$, one has $K(x, \eta) = h(x)$ for ν-almost every η. Since X is countable, we also have $\nu(A) = 1$, where

$$A = \{\eta \in \widehat{X} : K(x, \eta) = h(x) \text{ for all } x \in X\}.$$

This set must be non-empty. Therefore there must be $\xi \in A$ such that $h = K(\cdot, \xi)$. If $\eta \neq \xi$ then $K(\cdot, \eta) \neq K(\cdot, \xi)$ by the construction of the Martin compactification. In other words, A cannot contain more than the point ξ, and $\nu = \delta_\xi$. Since h is harmonic, we must have $\xi \in \mathcal{M}$. □

We define the *minimal Martin boundary* $\mathcal{M}_{\min}$ as the set of all $\xi \in \mathcal{M}$ such that $K(\cdot, \xi)$ is a minimal harmonic function. By now, we know that every minimal harmonic function arises in this way.

7.51 Corollary. *For a point $\xi \in \mathcal{M}$ one has $\xi \in \mathcal{M}_{\min}$ if and only if the limit distribution of the associated h-process with $h = K(\cdot, \xi)$ is $\nu^{K(\cdot,\xi)} = \delta_\xi$.*

Proof. The "only if" is contained in Theorem 7.50.

Conversely, let $\nu^{K(\cdot,\xi)} = \delta_\xi$. Suppose that $K(\cdot, \xi) = a_1 \cdot h_1 + a_2 \cdot h_2$ for two positive superharmonic functions h_1, h_2 with $h_i(o) = 1$ and constants $a_1, a_2 > 0$. Since $h_i(o) = 1$, the ν^{h_i} are probability measures. By Lemma 7.48

$$a_1 \cdot \nu^{h_1} + a_2 \cdot \nu^{h_2} = \delta_\xi.$$

This implies $\nu^{h_1} = \nu^{h_2} = \delta_\xi$, and $K(\cdot, \xi) \in \mathrm{d}\mathcal{B}$.

Now suppose that $K(\cdot, \xi) = K(\cdot, y)$ for some $y \in X$. (This can occur only when P does not have finite range, since finite range implies that $K(\cdot, \xi)$ is harmonic, while $K(\cdot, y)$ is not.) But then we know from Exercise 7.49 that $\nu^h = \delta_y \neq \delta_\xi$ in contradiction with the initial assumption. By Theorem 7.6, it only rests that h is a minimal harmonic function. □

Note a small subtlety in the last lines, where the proof relies on the fact that we distinguish $\xi \in \mathcal{M}$ from $y \in X$ even when $K(\cdot, \xi) = K(\cdot, y)$. Recall that this is because we wanted X to be discrete in the Martin compactification, following the approach of Hunt [32]. It may be instructive to reflect about the necessary modifications in the original approach of Doob [17], where ξ would not be distinguished from y.

We can combine the last characterization of $\mathcal{M}_{\min}$ with Proposition 7.43 to obtain the following.

7.52 Lemma. *$\mathcal{M}_{\min}$ is a Borel set in $\widehat{X}$.*

Proof. As we have seen in (7.15), the topology of $\widehat{X}$ is induced by a metric $\theta(\cdot, \cdot)$. Let $\xi \in \mathcal{M}$. For a function $h \in \mathcal{S}^+$ with $h(o) = 1$, we consider the h-process and apply Proposition 7.43 to the starting point o and the continuous function $f_m = e^{-m\,\theta(\cdot,\xi)}$:

$$\int_{\widehat{X}} f_m \, d\nu^h = \mathsf{E}_o^h\big(f_m(Z_\epsilon)\big) = \sum_{x \in X} f_m(x)\, \nu^h(x) + \lim_{n\to\infty} \sum_{y \in X} p^{(n)}(o, y)\, h(y)\, f_m(y).$$

If $m \to \infty$ then $f_m \to \mathbf{1}_{\{\xi\}}$, and by dominated convergence $\sum_{x \in X} f_m(x)\, \nu^h(x)$ tends to 0, while the integral on the left hand side tends to $\nu^h(\xi)$. Therefore

$$\lim_{m\to\infty} \lim_{n\to\infty} \sum_{y \in X} p^{(n)}(o, y)\, h(y)\, f_m(y) = \nu^h(\xi).$$

Setting $h = K(\cdot, \xi)$ and applying Corollary 7.51, we see that

$$\mathcal{M}_{\min} = \big\{\xi \in \mathcal{M} : \lim_{m\to\infty} \lim_{n\to\infty} \textstyle\sum_{y\in X} p^{(n)}(o,y)\, K(y,\xi)\, e^{-m\,\theta(y,\xi)} = 1\big\}.$$

Thus we have characterized $\mathcal{M}_{\min}$ as the set of points $\xi \in \mathcal{M}$ in which the triple limit (the third being summation over y) of a certain sequence of continuous functions on the compact set $\mathcal{M}$ is equal to 1. Therefore, $\mathcal{M}_{\min}$ is a Borel set by standard measure theory on metric spaces. □

We remark here that in general, $\mathcal{M}_{\min}$ can very well be a proper subset of $\mathcal{M}$. Examples where this occurs arise, among other, in the setting of Cartesian products of Markov chains, see Picardello and Woess [46] and [W2, §28.B]. We can now deduce the following result on uniqueness of the integral representation.

7.53 Theorem (Uniqueness of the representation). *If $h \in \mathcal{S}^+$ then the unique measure ν on $\widehat{X}$ such that*

$$\nu(\mathcal{M} \setminus \mathcal{M}_{\min}) = 0$$

and

$$h(x) = \int_{\widehat{X}} K(x,\cdot)\, d\nu \quad \text{for all } x \in X$$

is given by $\nu = \nu^h$, defined in (7.47).

Proof. 1.) Let us first verify that $\nu^h(\mathcal{M} \setminus \mathcal{M}_{\min}) = 0$. We may suppose that $h(o) = 1$. Let $f, g\colon \widehat{X} \to \mathbb{R}$ be two continuous functions. Then

$$\begin{aligned}\mathsf{E}_o^h\big(f(Z_n)\, g(Z_{n+m})\, \mathbf{1}_{[\epsilon\geq n+m]}\big) &= \sum_{x,y\in X} p_h^{(n)}(o,x)\, f(x)\, p_h^{(m)}(x,y)\, g(y)\\ &= \sum_{x\in X} p^{(n)}(o,x)\, h(x)\, f(x)\, \mathsf{E}_x^h\big(g(Z_m)\, \mathbf{1}_{[\epsilon\geq m]}\big).\end{aligned}$$

Letting $m \to \infty$, by dominated convergence

$$\mathsf{E}_o^h\big(f(Z_n)\, g(Z_\infty)\, \mathbf{1}_{\Omega_\infty}\big) = \sum_{x\in X} p^{(n)}(o,x)\, h(x)\, f(x)\, \mathsf{E}_x^h\big(g(Z_\infty)\, \mathbf{1}_{\Omega_\infty}\big), \quad (7.54)$$

since on Ω_∞ we have $\epsilon = \infty$ and $Z_\epsilon = Z_\infty$. Now, on Ω_∞ we also have $Z_\infty \in \mathcal{M}$. Considering the restriction $g|_{\mathcal{M}}$ of g to $\mathcal{M}$, and applying (7.41) and Theorem 7.42 to the h-process, we see that

$$\mathsf{E}_x^h\big(g(Z_\infty)\, \mathbf{1}_{\Omega_\infty}\big) = \mathsf{E}_x^h\big(g|_{\mathcal{M}}(Z_\epsilon)\big) = \int_{\mathcal{M}} K^h(x,\xi)\, g(\xi)\, d\nu^h(\xi). \quad (7.55)$$

Hence we can rewrite the right hand side of (7.54) as

$$\begin{aligned}
&\int_{\mathcal{M}} \sum_{x\in X} p^{(n)}(o,x)\, h(x)\, f(x)\, K^h(x,\xi)\, g(\xi)\, d\nu^h(\xi) \\
&\quad = \int_{\mathcal{M}} \sum_{x\in X} p^{(n)}(o,x)\, f(x)\, K(x,\xi)\, g(\xi)\, d\nu^h(\xi) \\
&\quad = \int_{\mathcal{M}} \sum_{x\in X} p^{(n)}_{K(\cdot,\xi)}(o,x)\, f(x)\, g(\xi)\; d\nu^h(\xi) \\
&\quad = \int_{\mathcal{M}} \mathsf{E}_o^{K(\cdot,\xi)}\left(f(Z_n)\,\mathbf{1}_{[\epsilon\geq n]}\right) g(\xi)\, d\nu^h(\xi).
\end{aligned}$$

If $n\to\infty$ then – as in (7.55), with $x=o$ and with $K(\cdot,\xi)$ in the place of h –

$$\mathsf{E}_o^{K(\cdot,\xi)}\left(f(Z_n)\,\mathbf{1}_{[\epsilon\geq n]}\right) \to \mathsf{E}_o^{K(\cdot,\xi)}\left(f\big|_{\mathcal{M}}(Z_\epsilon)\right) = \int_{\mathcal{M}} f(\eta)\, d\nu^{K(\cdot,\xi)}(\eta).$$

In the same way,

$$\mathsf{E}_o^h\big(f(Z_n)\, g(Z_\infty)\,\mathbf{1}_{\Omega_\infty}\big) \to \mathsf{E}_o^h\big(f(Z_\infty)\, g(Z_\infty)\,\mathbf{1}_{\Omega_\infty}\big) = \int_{\mathcal{M}} f(\xi)\, g(\xi)\, d\nu^h(\xi).$$

Combining these equations, we get

$$\int_{\mathcal{M}} f(\xi) g(\xi)\, d\nu^h(\xi) = \int_{\mathcal{M}} \left(\int_{\mathcal{M}} f(\eta)\, d\nu^{K(\cdot,\xi)}(\eta)\right) g(\xi)\, d\nu^h(\xi).$$

This is true for any choice of the continuous function g on $\widehat{X}$. One deduces that

$$f(\xi) = \int_{\mathcal{M}} f(\eta)\, d\nu^{K(\cdot,\xi)}(\eta) \quad \text{for } \nu^h\text{-almost every } \xi\in\mathcal{M}. \tag{7.56}$$

This is valid for every continuous function f on $\widehat{X}$. The boundary $\mathcal{M}$ is a compact space with the metric $\theta(\cdot,\cdot)$. It has a denumerable dense subset $\{\xi_k : k\in\mathbb{N}\}$. We consider the countable family of continuous functions $f_{k,m} = e^{-m\,\theta(\cdot,\xi_k)}$ on $\widehat{X}$. Then $\nu^h(B_{k,m})=0$, where $B_{k,m}$ is the set of all ξ which do not satisfy (7.56) with $f=f_{k,m}$. Then also $\nu^h(B)=0$, where $B=\bigcup_{k,m} B_{k,m}$. If $\xi\in\mathcal{M}\setminus B$ then for every m and k

$$e^{-m\,\theta(\xi,\xi_k)} = \int_{\mathcal{M}} e^{-m\,\theta(\eta,\xi_k)}\, d\nu^{K(\cdot,\xi)}(\eta).$$

There is a subsequence of (ξ_k) which tends to ξ. Passing to the limit,

$$1 = \int_{\mathcal{M}} e^{-m\,\theta(\eta,\xi)}\, d\nu^{K(\cdot,\xi)}(\eta).$$

If now $m \to \infty$, then the last right hand term tends to $\nu^{K(\cdot,\xi)}(\xi)$. Therefore $\nu^{K(\cdot,\xi)} = \delta_\xi$, and via Corollary 7.51 we deduce that $\xi \in \mathcal{M}_{\min}$. Consequently

$$\nu^h(\mathcal{M} \setminus \mathcal{M}_{\min}) \le \nu^h(B) = 0.$$

2.) We show uniqueness. Suppose that we have a measure ν with the stated properties. We can again suppose without loss of generality that $h(o) = 1$. Then ν and ν^h are probability measures. Let $f : \widehat{X} \to \mathbb{R}$ be continuous. Applying (7.41), Proposition 7.43 and (7.40) to the h-process,

$$\begin{aligned}
&\int_{\mathcal{M}_{\min}} f(\xi)\, d\nu^h(\xi) \\
&= \sum_{y\in X} f(y)\, G_h(o,y) \Big(1 - \sum_{w\in X} p_h(y,w)\Big) + \lim_{n\to\infty} P_h^n f(o) \\
&= \sum_{y\in X} f(y)\, G(o,y) \Big(h(y) - \sum_{w\in X} p(y,w)\, h(w)\Big) \\
&\quad + \lim_{n\to\infty} \sum_{x\in X} p^{(n)}(o,x)\, f(x)\, h(x).
\end{aligned} \tag{7.57}$$

Choose $\eta \in \mathcal{M}_{\min}$. Substitute h with $K(\cdot,\eta)$ in the last identity. Corollary 7.51 gives

$$\begin{aligned}
f(\eta) &= \int_{\mathcal{M}_{\min}} f(\xi)\, d\nu^{K(\cdot,\eta)}(\xi) \\
&= \sum_{y\in X} f(y)\, G(o,y) \Big(K(y,\eta) - \sum_{w\in X} p(y,w)\, K(w,\eta)\Big) \\
&\quad + \lim_{n\to\infty} \sum_{x\in X} p^{(n)}(o,x)\, f(x)\, K(x,\eta).
\end{aligned}$$

Integrating the last expression with respect to ν over $\mathcal{M}_{\min}$, the sums and the limit can exchanged with the integral (dominated convergence), and we obtain precisely the last line of (7.57). Thus

$$\int_{\mathcal{M}_{\min}} f(\eta)\, d\nu(\eta) = \int_{\mathcal{M}_{\min}} f(\xi)\, d\nu^h(\xi)$$

for every continuous function f on $\widehat{X}$: the measures ν and ν^h coincide. □

E Poisson boundary. Alternative approach to the integral representation

If ν is a Borel measure on $\mathcal{M}$ then

$$h = \int_{\mathcal{M}_{\min}} K(\cdot,\xi)\, d\nu(\xi)$$

defines a non-negative harmonic function. Indeed, by monotone convergence (applied to the summation occurring in Ph), one can exchange the integral and the application of P, and each of the functions $K(\cdot,\xi)$ with $\xi \in \mathcal{M}_{\min}$ is harmonic. If $u \in \mathcal{S}^+$ then by Theorem 7.45

$$u(x) = \sum_{y\in X} K(x,y)\,\nu^u(y) + \int_{\mathcal{M}_{\min}} K(x,\xi)\,d\nu^u(\xi).$$

Set $g(x) = \sum_{y\in X} K(x,y)\,\nu^u(y)$ and $h(x) = \int_{\mathcal{M}_{\min}} K(x,\xi)\,d\nu^u(\xi)$. Then, as we just observed, h is harmonic, and $\nu^h = \nu^u|_{\mathcal{M}}$ by Theorem 7.53.

In view of Exercise 7.49, we find that $g = Gf$, where $f = u - Pu$. In this way we have re-derived the Riesz decomposition $u = Gf + h$ of the superharmonic function u, with more detailed information regarding the harmonic part.

The constant function $\mathbf{1} = \mathbf{1}_X$ is harmonic precisely when P is stochastic, and superharmonic in general. If we set $B = \widehat{X}$ in Theorem 7.42, then we see that the measure on $\widehat{X}$, which gives rise to the integral representation of $\mathbf{1}_X$ in the sense of Theorem 7.53, is the measure ν_o. That is, for any Borel set $B \subset \widehat{X}$,

$$\nu^{\mathbf{1}}(B) = \nu_o(B) = \mathsf{Pr}_o[Z_{\boldsymbol{\epsilon}} \in B].$$

If P is stochastic then $\boldsymbol{\epsilon} = \infty$ and ν_o as well as all the other measures ν_x, $x \in X$, are probability measures on $\mathcal{M}_{\min}$. If P is strictly substochastic in some point, then $\mathsf{Pr}_x[\boldsymbol{\epsilon} < \infty] > 0$ for every x.

7.58 Exercise. Construct examples where $\mathsf{Pr}_x[\boldsymbol{\epsilon} < \infty] = 1$ for every x, that is, the Markov chain does not escape to infinity, but vanishes (is absorbed by †) almost surely.

[Hint: modify an arbitrary recurrent Markov chain suitably.] □

In general, we can write the Riesz decomposition of the constant function $\mathbf{1}$ on X:

$$\mathbf{1} = h_0 + Gf_0 \quad \text{with } h_0(x) = \nu_x(\mathcal{M}) \text{ for every } x \in X. \tag{7.59}$$

Thus, $h_0 \equiv 0 \iff \boldsymbol{\epsilon} < \infty$ almost surely $\iff \mathbf{1}_X$ is a potential.

In the sequel, when we speak of a function φ on $\mathcal{M}$, then we tacitly assume that φ is extended to the whole of $\widehat{X}$ by setting $\varphi = 0$ on X. Thus, a ν_o-integrable function φ on $\mathcal{M}$ is intended to be one that is integrable with respect to $\nu_o|_{\mathcal{M}}$. (Again, these subtleties do not have to be considered when P is stochastic.) The *Poisson integral* of φ is the function

$$h(x) = \int_{\mathcal{M}} K(x,\cdot)\,\varphi\,d\nu_o = \int_{\mathcal{M}} \varphi\,d\nu_x = \mathsf{E}_x\big(\varphi(Z_\infty)\,\mathbf{1}_{\Omega_\infty}\big), \quad x \in X. \tag{7.60}$$

It defines a harmonic function (not necessarily positive). Indeed, we can decompose $\varphi = \varphi_+ - \varphi_-$ and consider the non-negative measures $d\nu_\pm(\xi) = \varphi_\pm(\xi)\,d\nu_o(\xi)$.

Since $\nu_o(\mathcal{M} \setminus \mathcal{M}_{\min}) = 0$ (Theorem 7.53), the functions $h_\pm = \int_{\mathcal{M}} K(\cdot, \xi)\, d\nu_\pm(\xi)$ are harmonic, and $h = h_+ - h_-$. If φ is a bounded function then also h is bounded. Conversely, the following holds.

7.61 Theorem. *Every bounded harmonic function is the Poisson integral of a bounded measurable function on $\mathcal{M}$.*

Proof. Claim. For every bounded harmonic function h on X there are constants $a, b \in \mathbb{R}$ such that $a \cdot h_0 \le h \le b \cdot h_0$.

To see this, we start with $b \ge 0$ such that $h(x) \le b$ for every x. That is, $u = b \cdot \mathbf{1}_X - h$ is a non-negative superharmonic function. Then $u = \bar{h} + b \cdot Gf_0$, where $\bar{h} = b \cdot h_0 - h$ is harmonic. Therefore $0 \le P^n u = \bar{h} + b \cdot P^n G f_0 \to \bar{h}$ as $n \to \infty$, whence $\bar{h} \ge 0$.

For the lower bound, we apply this reasoning to $-h$.

The claim being verified, we now set $c = b - a$ and write $c \cdot h_0 = h_1 + h_2$, where $h_1 = b \cdot h_0 - h$ and $h_2 = h - a \cdot h_0$. The h_i are non-negative harmonic functions. By Lemma 7.48,

$$\nu^{h_1} + \nu^{h_2} = c \cdot \nu^{h_0} = c \cdot \mathbf{1}_{\mathcal{M}}\, \nu_o,$$

where $\mathbf{1}_{\mathcal{M}}\, \nu_o$ is the restriction of ν_o to $\mathcal{M}$. In particular, both ν^{h_i} are absolutely continuous with respect to $\mathbf{1}_{\mathcal{M}}\, \nu_o$ and have non-negative Radon–Nikodym densities φ_i supported on $\mathcal{M}$ with respect to ν_o. Thus, for $i = 1, 2$,

$$h_i = \int_{\mathcal{M}} K(\cdot, \xi)\, \varphi_i(\xi)\, d\nu_o(\xi).$$

Adding the two integrals, we see that

$$\varphi_1 + \varphi_2 = c \cdot \mathbf{1}_{\mathcal{M}}$$

ν_o-almost everywhere, whence the φ_i are ν_o-almost everywhere bounded. We now get

$$h = h_2 + a \cdot h_0 = \int_{\mathcal{M}} K(\cdot, \xi)\, \varphi(\xi)\, d\nu_o(\xi), \quad \text{where } \varphi = \varphi_2 + a \cdot \mathbf{1}_{\mathcal{M}}.$$

This is the proposed Poisson integral. □

The last proof becomes simpler when P is stochastic.

We underline that by the uniqueness theorem (Theorem 7.42), the bounded harmonic function h determines the function φ uniquely ν_o-almost everywhere.

For the following measure-theoretic considerations, we continue to use the essential part of the trajectory space, namely $\Omega = X^{\mathbb{N}_0} \cup \Omega_\dagger$ as in (7.21). We write

(by slight abuse of notation) $\mathcal{A}$ for the σ-algebra restricted to that set, on which all our probability measures Pr_x live. Let $\boldsymbol{\sigma}$ be the shift operator on Ω, namely

$$\boldsymbol{\sigma}(x_0, x_1, x_2, \dots) = (x_1, x_2, x_3, \dots).$$

Its action extends to any extended real random variable W defined on Ω by $\boldsymbol{\sigma} W(\omega) = W(\boldsymbol{\sigma}\omega)$

Following once more the lucid presentation of DYNKIN [Dy], we say that such a random variable is *terminal* or *final*, if $\boldsymbol{\sigma} W = W$ and, in addition, $W \equiv 0$ on $\Omega_\dagger$. Analogously, an event $A \in \mathcal{A}$ is called terminal or final, if its indicator function $\mathbf{1}_A$ has this property. The terminal events form a σ-algebra of subsets of the "ordinary" trajectory space $X^{\mathbb{N}_0}$. Every non-negative terminal random variable can be approximated in the standard way by non-negative simple terminal random variables (i.e., linear combinations of indicator functions of terminal events).

"Terminal" means that the value of $W(\omega)$ does not depend on the deletion, insertion or modification of an initial (finite) piece of a trajectory within $X^{\mathbb{N}_0}$. The basic example of a terminal random variable is as follows: let $\varphi\colon \mathcal{M}_{\min} \to [-\infty, +\infty]$ be measurable, and define

$$W(\omega) = \begin{cases} \varphi\big(Z_\infty(\omega)\big), & \text{if } \omega \in \Omega_\infty, \\ 0, & \text{otherwise.} \end{cases}$$

There is a direct relation between terminal random variables and harmonic functions, which will lead us to the conclusion that every terminal random variable has the above form.

7.62 Proposition. *Let $W \geq 0$ be a terminal random variable satisfying $0 < \mathsf{E}_o(W) < \infty$. Then $h(x) = \mathsf{E}_x(W) < \infty$, and h is harmonic on X.*

The probability measures with respect to the h-process are given by

$$\mathsf{Pr}^h_x(A) = \frac{1}{h(x)} \mathsf{E}_x(\mathbf{1}_A\, W), \quad A \in \mathcal{A},\ x \in X.$$

Proof. (Note that in the last formula, expectation E_x always refers to the "ordinary" probability measure Pr_x on the trajectory space.)

If W is an arbitrary (not necessarily terminal) random variable on Ω, we can write $W(\omega) = W\big(Z_0(\omega), Z_1(\omega), \dots\big)$. Let $y_1, \dots, y_k \in X$ ($k \geq 1$) and denote, as usual, by $\mathsf{E}_x(\cdot \mid Z_1 = y_1, \dots, Z_k = y_k)$ expectation with respect to the probability measure $\mathsf{Pr}_x(\cdot \mid Z_1 = y_1, \dots, Z_k = y_k)$. By the Markov property and

time-homogeneity,

$$\begin{aligned}
&\mathsf{E}_x\big(\mathbf{1}_{[Z_1=y_1,\dots,Z_k=y_k]}W(Z_k,Z_{k+1},\dots)\big)\\
&\quad=\mathsf{Pr}_x[Z_1=y_1,\dots,Z_k=y_k]\cdot\\
&\qquad\cdot\mathsf{E}_x\big(W(Z_k,Z_{k+1},\dots)\mid Z_1=y_1,\dots,Z_k=y_k\big)\\
&\quad=p(x,y_1)\cdots p(y_{k-1},y_k)\,\mathsf{E}_x\big(W(Z_k,Z_{k+1},\dots)\mid Z_k=y_k\big)\\
&\quad=p(x,y_1)\cdots p(y_{k-1},y_k)\,\mathsf{E}_y\big(W(Z_0,Z_1,\dots)\big)
\end{aligned}\tag{7.63}$$

In particular, if $W\ge 0$ is terminal, that is, $W(Z_0,Z_1,Z_2,\dots)=W(Z_1,Z_2,\dots)$, then for arbitrary y,

$$\mathsf{E}_x(\mathbf{1}_{[Z_1=y]}W)=p(x,y)\,\mathsf{E}_y(W).$$

Thus, if $\mathsf{E}_x(W)<\infty$ and $p(x,y)>0$ then $\mathsf{E}_y(W)<\infty$. Irreducibility now implies that when $\mathsf{E}_o(W)<\infty$ then $\mathsf{E}_x(W)<\infty$ for all x. In this case, using the fact that $W\equiv 0$ on $\Omega_\dagger$,

$$\mathsf{E}_x(W)=\mathsf{E}_x\Big(\sum_{y\in X\cup\{\dagger\}}\mathbf{1}_{[Z_1=y]}W\Big)=\sum_{y\in X}\mathsf{E}_x(\mathbf{1}_{[Z_1=y]}W)=\sum_{y\in X}p(x,y)\,\mathsf{E}_y(W);$$

the function h is harmonic.

In order to prove the proposed formula for Pr_x^h, we only need to verify it for an arbitrary cylinder set $A=\mathsf{C}(a_0,a_1,\dots,a_k)$, where $a_0,\dots,a_k\in X$. (We do not need to consider $a_j=\dagger$, since the h-process does not visit $\dagger$ when it starts in X.) For our cylinder set,

$$\begin{aligned}
\mathsf{Pr}_x^h(A)&=\delta_x(a_0)p_h(a_0,a_1)\cdots p_h(a_{k-1},a_k)\\
&=\frac{h(a_k)}{h(x)}\delta_x(a_0)p(a_0,a_1)\cdots p(a_{k-1},a_k).
\end{aligned}$$

On the other hand, we apply (7.63) and get, using that W is terminal,

$$\frac{1}{h(x)}\mathsf{E}_x(\mathbf{1}_A\,W)=\frac{1}{h(x)}\delta_x(a_0)p(a_0,a_1)\cdots p(a_{k-1},a_k)\,\mathsf{E}_{a_k}(W),$$

which coincides with $\mathsf{Pr}_x^h(A)$ as claimed. □

The first part of the last proposition remains of course valid for any final random variable W that is E_o-integrable: then it is E_x-integrable for every $x\in X$, and $h(x)=\mathsf{E}_x(W)$ is harmonic. If W is (essentially) bounded then h is a bounded harmonic function.

7.64 Theorem. *Let W be a bounded, terminal random variable, and let φ be the bounded measurable function on $\mathcal{M}$ such that $h(x) = \mathsf{E}_x(W)$ satisfies*

$$h(x) = \int_{\mathcal{M}} \varphi \, d\nu_x.$$

Then

$$W = \varphi(Z_\infty) \cdot \mathbf{1}_{\Omega_\infty} \quad \mathsf{Pr}_x\text{-almost surely for every } x \in X.$$

Proof. As proposed, we let φ be the function on $\mathcal{M}$ that appears in the Poisson integral representation of h. Then $W' = \varphi(Z_\infty) \cdot \mathbf{1}_{\Omega_\infty}$ is a bounded terminal random variable that satisfies

$$\mathsf{E}_x(W') = h(x) = \mathsf{E}_x(W) \quad \text{for all } x \in X.$$

Therefore the second part of Proposition 7.62 implies that

$$\int_A W \, d\,\mathsf{Pr}_x = \int_A W' \, d\,\mathsf{Pr}_x \quad \text{for every } A \in \mathcal{A}.$$

The result follows. □

7.65 Corollary. *Every terminal random variable W is of the form*

$$W = \varphi(Z_\infty) \cdot \mathbf{1}_{\Omega_\infty} \quad \mathsf{Pr}_x\text{-almost surely for every } x \in X,$$

where φ is a measurable function on $\mathcal{M}$.

Proof. For bounded W, this follows from Theorems 7.61 and 7.64. If W is non-negative, choose $n \in \mathbb{N}$ and set $W_n = \min\{n, W\}$ (pointwise). This is a bounded, non-negative terminal random variable, whence there is a bounded, non-negative function φ_n on $\mathcal{M}$ such that

$$W_n = \varphi_n(Z_\infty) \cdot \mathbf{1}_{\Omega_\infty} \quad \mathsf{Pr}_x\text{-almost surely for every } x \in X.$$

If we set $\varphi = \limsup_n \varphi_n$ (pointwise on $\mathcal{M}$) then $W = \varphi(Z_\infty)\,\mathbf{1}_{\Omega_\infty}$.

Finally, if W is arbitrary, then we decompose $W = W_+ - W_-$. We get $W_\pm = \varphi_\pm(Z_\infty) \cdot \mathbf{1}_{\Omega_\infty}$. Then we can set $\varphi = \varphi_+ - \varphi_-$ (choosing the value to be 0 whenever this is of the indefinite form $\infty - \infty$; note that $W_+ = 0$ where $W_- > 0$ and vice versa). □

For the following, we write $\nu = \mathbf{1}_{\mathcal{M}}\,\nu_o$ for the restriction of ν_o to $\mathcal{M}$.

The pair $(\mathcal{M}, \nu)$, as a measure space with the Borel σ-algebra, is called the *Poisson boundary* of (X, P). It is a probability space if and only if the matrix P is stochastic, in which case $\nu = \nu_o$. Since $\nu(\mathcal{M} \setminus \mathcal{M}_{\min}) = 0$, we can identify $(\mathcal{M}, \nu)$ with $(\mathcal{M}_{\min}, \nu)$. Besides being large enough for providing a unique integral

representation of all bounded harmonic functions, the Poisson boundary is also the "right" model for the distinguishable limit points at infinity which the Markov chain (Z_n) can attain. Note that for describing those limit points, we do not only need the topological model (the Martin compactification), but also the distribution of the limit random variable. Theorem 7.64 and Corollary 7.65 show that the Poisson boundary is the finest model for distinguishing the behaviour of (Z_n) at infinity.

On the other hand, in many cases the Poisson boundary can be "smaller" than $\mathcal{M}_{\min}$ in the sense that the support of ν does not contain all points of $\mathcal{M}_{\min}$. In particular, we shall say that the Poisson boundary is *trivial,* if $\operatorname{supp}(\nu)$ consists of a single point. In formulating this, we primarily have in mind the case when P is stochastic. In the stochastic case, triviality of the Poisson boundary amounts to the (weak) *Liouville property:* all bounded harmonic functions are constant.

When P is strictly substochastic in some point, recall that it may also happen that $\boldsymbol{\epsilon} < \infty$ almost surely, in which case $\nu_x(\mathcal{M}) = 0$ for all x, and there are no non-zero bounded harmonic functions. In this case, the Poisson boundary is *empty* (to be distinguished from "trivial").

7.66 Exercise. As in (7.59), let h_0 be the harmonic part in the Riesz decomposition of the superharmonic function $\mathbf{1}$ on X. Show that the following statements are equivalent.

(a) The Poisson boundary of (X, P) is trivial.

(b) The function h_0 of (7.59) is non-zero, and every bounded harmonic function is a constant multiple of h_0.

(c) One has $h_0(o) \neq 0$, and $\frac{1}{h_0(o)} h_0$ is a minimal harmonic function. □

We next deduce the following theorem of convergence to the boundary.

7.67 Theorem (Probabilistic Fatou theorem). *If φ is a ν_o-integrable function on $\mathcal{M}$ and h its Poisson integral, then*

$$\lim_{n\to\infty} h(Z_n) = \varphi(Z_\infty) \quad \nu_o\text{-almost surely on } \Omega_\infty.$$

Proof. Suppose first that φ is bounded. By Corollary 7.31, $W = \lim_{n\to\infty} h(Z_n)$ exists Pr_x-almost surely. This W is a terminal random variable. (From the lines preceding the corollary, we see that $W \equiv 0$ on $\Omega_\dagger$.) Since h is bounded, we can use Lebesgue's theorem (dominated convergence) to obtain

$$\mathsf{E}_x(W) = \lim_{n\to\infty} \mathsf{E}_x\big(h(Z_n)\big) = \lim_{n\to\infty} P^n h(x) = h(x)$$

for every $x \in X$. Now Theorem 7.64 implies that $W = \varphi(Z_\infty) \cdot \mathbf{1}_{\Omega_\infty}$, as claimed.

Next, suppose that φ is non-negative and ν_o-integrable. Let $N \in \mathbb{N}$ and define $\varphi_N = \varphi \cdot \mathbf{1}_{[\varphi \le N]}$ and $\psi_N = \varphi \cdot \mathbf{1}_{[\varphi > N]}$. Then $\varphi_N + \psi_N = \varphi$. Let g_N and h_N be the

Poisson integrals of φ_N and ψ_N, respectively. These two functions are harmonic, and $g_N(x) + h_N(x) = h(x)$.

We can write

$$h(x) = \mathsf{E}_x(Y), \quad \text{where } Y = \varphi(Z_\infty) \cdot \mathbf{1}_{\Omega_\infty},$$
$$g_N(x) = \mathsf{E}_x(V_N), \quad \text{where } V_N = \varphi_N(Z_\infty) \cdot \mathbf{1}_{\Omega_\infty},$$

and

$$h_N(x) = \mathsf{E}_x(Y_N), \quad \text{where } Y_N = \psi_N(Z_\infty) \cdot \mathbf{1}_{\Omega_\infty}.$$

Since φ_N is bounded, we know from the first part of the proof that

$$\lim_{n\to\infty} g_N(Z_n) = V_N \quad \nu_o\text{-almost surely on } \Omega_\infty.$$

Furthermore, we know from Corollary 7.31 that $W = \lim_{n\to\infty} h(Z_n)$ and $W_N = \lim_{n\to\infty} h_N(Z_n)$ exist and are terminal random variables.

We have $W = V_N + W_N$ and $Y = V_N + Y_N$ and need to show that $W = Y$ almost surely. We cannot apply the dominated convergence theorem to $h_N(Z_n)$, as $n \to \infty$, but by Fatou's lemma,

$$\mathsf{E}_x(W_N) \le \lim_{n\to\infty} \mathsf{E}_x\big(h_N(Z_n)\big) = h_N(x).$$

Therefore

$$\mathsf{E}_x\big(|W - Y|\big) = \mathsf{E}_x\big(|W_N - Y_N|\big) \le \mathsf{E}_x(W_N) + \mathsf{E}_x(Y_N) \le 2h_N(x).$$

Now by irreducibility, there is $C_x > 0$ such that $h_N(x) \le C_x\, h_N(o)$, see (7.3). Since φ is ν_o-integrable, $h_N(o) = \nu_o[\varphi > N] \to 0$ as $N \to \infty$. Therefore $W - Y = 0$ Pr_x-almost surely, as proposed.

Finally, in general we can decompose $\varphi = \varphi_+ - \varphi_-$ and apply what we just proved to the positive and negative parts. □

Besides the Riesz decomposition (see above), also the approximation theorem (Theorem 6.46) can be easily deduced by the methods developed in this section: if $h \in \mathcal{S}^+$ and $V \subset X$ is finite, then by transience of the h-process one has

$$\sum_{y\in V} \mathsf{Pr}_x^h[Z_{\epsilon_V} = y] \begin{cases} = 1, & \text{if } x \in V, \\ \le 1, & \text{otherwise.} \end{cases}$$

Applying (7.37) to the h-process, this relation can be rewritten as

$$\sum_{y\in V} G(x,y)\, h(y)\ \mathsf{Pr}_y^h[\epsilon_V = 0] \begin{cases} = h(x), & \text{if } x \in V, \\ \le h(x), & \text{otherwise.} \end{cases}$$

If we choose an increasing sequence of finite sets V_n with union X, we obtain h as the pointwise limit from below of a sequence of potentials.

Alternative approach to the Poisson–Martin integral representation

The above way of deducing the approximation theorem, involving Martin boundary theory, is of course far more complicated than the one in Section 7.D. Conversely, the integral representation can also be deduced directly from the approximation theorem without prior use of the theorem on convergence to the boundary.

Alternative proof of the integral representation theorem. Let $h \in \mathcal{S}^+$. By Theorem 6.46, there is a sequence of non-negative functions f_n on X such that $g_n(x) = Gf_n(x) \to h(x)$ pointwise. We rewrite

$$Gf_n(x) = Gf_n(o) \sum_{y \in X} K(x, y)\, \nu_n(y),$$

where

$$\nu_n(y) = \frac{G(o, y)\, f_n(y)}{Gf_n(o)}.$$

Then ν_n is a probability distribution on the set X, which is discrete in the topology of $\widehat{X}$. We can consider ν_n as a Borel measure on $\widehat{X}$ and rewrite

$$Gf_n(x) = Gf_n(o) \int_{\widehat{X}} K(x, \cdot)\, d\nu_n.$$

Now, the set of all Borel probability measures on a compact metric space is compact in the topology of convergence in law (weak convergence), see for example [Pa]. This implies that there are a subsequence ν_{n_k} and a probability measure ν on $\widehat{X}$ such that

$$\int_{\widehat{X}} f\, d\nu_{n_k} \to \int_{\widehat{X}} f\, d\nu$$

for every continuous function f on $\widehat{X}$. But the functions $K(x, \cdot)$ are continuous, and in the limit we obtain

$$h(x) = h(o) \int_{\widehat{X}} K(x, \cdot)\, d\nu.$$

This provides an integral representation of h with respect to the Borel measure $h(o) \cdot \nu$. □

This proof appears simpler than the road we have taken in order to achieve the integral representation. Indeed, this is the approach chosen in the original paper by Doob [17]. It uses a fundamental and rather profound theorem, namely the one on compactness of the set of Borel probability measures on a compact metric space. (This can be seen as a general version of Helly's principle, or as a special case of Alaoglu's theorem of functional analysis: the dual of the Banach space of

all continuous functions on $\widehat{X}$ is the space of all signed Borel measures; in the weak topology, the set of all probability measures is a closed subset of the unit ball and thus compact.) After this proof of the integral representation, one still needs to *deduce* from the latter the other theorems regarding convergence to the boundary, minimal boundary, uniqueness of the representation. In conclusion, the probabilistic approach presented here, which is due to HUNT [32], has the advantage to be based only on a relatively elementary version of martingale theory.

Nevertheless, in lectures where time is short, it may be advantageous to deduce the integral representation directly from the approximation theorem as above, and then prove that all minimal harmonic functions are Martin kernels, precisely as in Theorem 7.50. After this, one may state the theorems on convergence to the boundary and uniqueness of the representation without proof.

This approach is supported by the observations at the end of Section B: in all classes of specific examples of Markov chains where one is able to elaborate a concrete description of the Martin compactification, one also has at hand a direct proof of convergence to the boundary that relies on the specific features of the respective example, but is usually much simpler than the general proof of the convergence theorem.

What we mean by "concrete description" is the following. Imagine to have a class of Markov chains on some state space X which carries a certain geometric, algebraic or combinatorial structure (e.g., a hyperbolic graph or group, an integer lattice, an infinite tree, or a hyperbolic graph or group). Suppose also that the transition probabilities of the Markov chain are adapted to that structure. Then we are looking for a "natural" compactification of that structure, *a priori* defined in the respective geometric, algebraic or combinatorial terms, maybe without thinking yet about the probabilistic model (the Markov chain and its transition matrix). Then we want to know if this model may also serve as a concrete description of the Martin compactification.

In Chapter 9, we shall carry out this program in the class of examples which is simplest for this purpose, namely for random walks on trees. Before that, we insert a brief chapter on random walks on lattices.

Chapter 8
Minimal harmonic functions on Euclidean lattices

In this chapter, we consider irreducible random walks on the Abelian group $\mathbb{Z}^d$ in the sense of (4.18), but written additively. Thus,

$$p^{(n)}(\boldsymbol{k}, \boldsymbol{l}) = \mu^{(n)}(\boldsymbol{l} - \boldsymbol{k}) \quad \text{for } \boldsymbol{k},\ \boldsymbol{l} \in \mathbb{Z}^d,$$

where μ is a probability measure on $\mathbb{Z}^d$ and $\mu^{(n)}$ its n-th convolution power given by $\mu^{(1)} = \mu$ and

$$\mu^{(n)}(\boldsymbol{k}) = \sum_{\boldsymbol{k}_1 + \cdots + \boldsymbol{k}_n = \boldsymbol{k}} \mu(\boldsymbol{k}_1) \cdots \mu(\boldsymbol{k}_n) = \sum_{\boldsymbol{l} \in \mathbb{Z}^d} \mu^{(n-1)}(\boldsymbol{l})\, \mu(\boldsymbol{k} - \boldsymbol{l}).$$

(Compare with simple random walk on $\mathbb{Z}^d$, where μ is equidistribution on the set of integer unit vectors.) Recall from Lemma 4.21 that irreducibility means $\bigcup_n \operatorname{supp}(\mu^{(n)}) = \mathbb{Z}^d$, with $\operatorname{supp}(\mu^{(n)}) = \{\boldsymbol{k}_1 + \cdots + \boldsymbol{k}_n : \boldsymbol{k}_i \in \operatorname{supp}(\mu)\}$. The action of the transition matrix on functions $f : \mathbb{Z}^d \to \mathbb{R}$ is

$$Pf(\boldsymbol{k}) = \sum_{\boldsymbol{l} \in \mathbb{Z}^d} f(\boldsymbol{l})\, \mu(\boldsymbol{l} - \boldsymbol{k}) = \sum_{\boldsymbol{m} \in \mathbb{Z}^d} f(\boldsymbol{k} + \boldsymbol{m})\, \mu(\boldsymbol{m}). \tag{8.1}$$

We first study the bounded harmonic functions, that is, the Poisson boundary. The following theorem was first proved by Blackwell [8], while the proof given here goes back to a paper by Dynkin and Malyutov [19], who attribute it to A. M. Leonotovič.

8.2 Theorem. *All bounded harmonic functions with respect to μ are constant.*

Proof. The proof is based on the following.

Let $h \in \mathcal{H}^\infty$ and $\boldsymbol{l} \in \operatorname{supp}(\mu)$. Then we claim that

$$h(\boldsymbol{k} + \boldsymbol{l}) \le h(\boldsymbol{k}) \quad \text{for every } \boldsymbol{k} \in \mathbb{Z}^d. \tag{8.3}$$

Proof of the claim. Let $c > 0$ be a constant such that $|h(\boldsymbol{k})| \le c$ for all $\boldsymbol{k}$. We set $g(\boldsymbol{k}) = h(\boldsymbol{k} + \boldsymbol{l}) - h(\boldsymbol{k})$. We have to verify that $g \le 0$. First of all,

$$Pg(\boldsymbol{k}) = \sum_{\boldsymbol{m} \in \mathbb{Z}^d} \big(h(\boldsymbol{k} + \boldsymbol{l} + \boldsymbol{m}) - h(\boldsymbol{k} + \boldsymbol{m})\big)\, \mu(\boldsymbol{m}) = g(\boldsymbol{k}).$$

(Note that we have used in a crucial way the fact that $\mathbb{Z}^d$ is an Abelian group.) Therefore $g \in \mathcal{H}^\infty$, and $|g(\boldsymbol{k})| \le 2c$ for every $\boldsymbol{k}$.

Suppose by contradiction that

$$b = \sup_{\boldsymbol{k}\in\mathbb{Z}^d} g(\boldsymbol{k}) > 0.$$

We observe that for each $N \in \mathbb{N}$ and every $\boldsymbol{k} \in \mathbb{Z}^d$,

$$\sum_{n=0}^{N-1} g(\boldsymbol{k} + n\boldsymbol{l}) = h(\boldsymbol{k} + N\boldsymbol{l}) - h(\boldsymbol{k}) \le 2c.$$

We choose N sufficiently large so that $N\frac{b}{2} > 2c$. We shall now find $\boldsymbol{k} \in \mathbb{Z}^d$ such that $g(\boldsymbol{k} + n\boldsymbol{l}) > \frac{b}{2}$ for all $n < N$, thus contradicting the assumption that $b > 0$.

For arbitrary $\boldsymbol{k} \in \mathbb{Z}^d$ and $n < N$,

$$p^{(n)}(\boldsymbol{k}, \boldsymbol{k} + n\boldsymbol{l}) = \mu^{(n)}(n\boldsymbol{l}) \ge \big(\mu(\boldsymbol{l})\big)^n \ge \big(\mu(\boldsymbol{l})\big)^{N-1} = a,$$

where $0 < a < 1$. Since $b(1 - \frac{a}{2}) < b$, there must be $\boldsymbol{k} \in \mathbb{Z}^d$ with

$$g(\boldsymbol{k}) > b\left(1 - \frac{a}{2}\right).$$

Observing that $P^n g = g$, we obtain for this $\boldsymbol{k}$ and for $0 \le n < N$

$$\begin{aligned} b\left(1 - \frac{\mu^{(n)}(n\boldsymbol{l})}{2}\right) \le b\left(1 - \frac{a}{2}\right) &< g(\boldsymbol{k}) \\ &= g(\boldsymbol{k} + n\boldsymbol{l})\,\mu^{(n)}(n\boldsymbol{l}) + \sum_{\boldsymbol{m}\neq n\boldsymbol{l}} g(\boldsymbol{k} + \boldsymbol{m})\,\mu^{(n)}(\boldsymbol{m}) \\ &\le g(\boldsymbol{k} + n\boldsymbol{l})\,\mu^{(n)}(n\boldsymbol{l}) + \sum_{\boldsymbol{m}\neq n\boldsymbol{l}} b\,\mu^{(n)}(\boldsymbol{m}) \\ &= g(\boldsymbol{k} + n\boldsymbol{l})\,\mu^{(n)}(n\boldsymbol{l}) + b\big(1 - \mu^{(n)}(n\boldsymbol{l})\big). \end{aligned}$$

Simplifying, we get $g(\boldsymbol{k} + n\boldsymbol{l}) > \frac{b}{2}$ for every $n < N$, as proposed. This completes the proof of (8.3).

If $h \in \mathcal{H}^\infty$ then we can apply (8.3) both to h and to $-h$ and obtain

$$h(\boldsymbol{k} + \boldsymbol{l}) = h(\boldsymbol{k}) \quad \text{for every } \boldsymbol{k} \in \mathbb{Z}^d \text{ and every } \boldsymbol{l} \in \operatorname{supp}(\mu).$$

Now let $\boldsymbol{k} \in \mathbb{Z}^d$ be arbitrary. By irreducibility, we can find $n > 0$ and elements $\boldsymbol{l}_1, \dots, \boldsymbol{l}_n \in \operatorname{supp}(\mu)$ such that $\boldsymbol{k} = \boldsymbol{l}_1 + \cdots + \boldsymbol{l}_n$. Then by the above

$$h(\boldsymbol{0}) = h(\boldsymbol{l}_1) = h(\boldsymbol{l}_1 + \boldsymbol{l}_2) = \cdots = h(\boldsymbol{l}_1 + \cdots + \boldsymbol{l}_{n-1}) = h(\boldsymbol{k}),$$

and h is constant. □

Besides the constant functions, it is easy to spot another class of functions on $\mathbb{Z}^d$ that are harmonic for the transition operator (8.1). For $\boldsymbol{c} \in \mathbb{R}^d$, let

$$f_{\boldsymbol{c}}(\boldsymbol{k}) = e^{\boldsymbol{c}\cdot\boldsymbol{k}}, \quad \boldsymbol{k} \in \mathbb{Z}^d. \tag{8.4}$$

(Here, $\boldsymbol{c}\cdot\boldsymbol{k}$ denotes the standard scalar product in $\mathbb{R}^d$.) Then $f_{\boldsymbol{c}}(\boldsymbol{k})$ is P-integrable if and only if

$$\varphi(\boldsymbol{c}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} e^{\boldsymbol{c}\cdot\boldsymbol{k}}\, \mu(\boldsymbol{k}) \tag{8.5}$$

is finite, and in this case,

$$P f_{\boldsymbol{c}}(\boldsymbol{k}) = \sum_{\boldsymbol{l}\in\mathbb{Z}^d} e^{\boldsymbol{c}\cdot\boldsymbol{k}}\, e^{\boldsymbol{c}\cdot(\boldsymbol{l}-\boldsymbol{k})}\, \mu(\boldsymbol{l}-\boldsymbol{k}) = \varphi(\boldsymbol{c})\, f_{\boldsymbol{c}}(\boldsymbol{k}). \tag{8.6}$$

Note that $\varphi(\boldsymbol{c}) = P f_{\boldsymbol{c}}(\boldsymbol{0})$. If $\varphi(\boldsymbol{c}) = 1$ then $f_{\boldsymbol{c}}$ is a positive harmonic function. For the reference point o, our natural choice is the origin $\boldsymbol{0}$ of $\mathbb{Z}^d$, so that $f_{\boldsymbol{c}}(o) = 1$.

8.7 Theorem. *The minimal harmonic functions for the transition operator* (8.1) *are precisely the functions* $f_{\boldsymbol{c}}$ *with* $\varphi(\boldsymbol{c}) = 1$.

Proof. A. Let h be a minimal harmonic function. For $\boldsymbol{l} \in \mathbb{Z}^d$, set $h_{\boldsymbol{l}}(\boldsymbol{k}) = h(\boldsymbol{k}+\boldsymbol{l})/h(\boldsymbol{l})$. Then, as in the proof of Theorem 8.2,

$$P h_{\boldsymbol{l}}(\boldsymbol{k}) = \frac{1}{h(\boldsymbol{l})} \sum_{\boldsymbol{m}\in\mathbb{Z}^d} h(\boldsymbol{k}+\boldsymbol{m}+\boldsymbol{l})\, \mu(\boldsymbol{m}) = h_{\boldsymbol{l}}(\boldsymbol{k}).$$

Hence $h_{\boldsymbol{l}} \in \mathcal{B}$. If $n \geq 1$, by iterating (8.1), we can write the identity $P^n h = h$ as

$$h(\boldsymbol{k}) = \sum_{\boldsymbol{l}\in\mathbb{Z}^d} h_{\boldsymbol{l}}(\boldsymbol{k})\, h(\boldsymbol{l})\, \mu^{(n)}(\boldsymbol{l}).$$

That is, $h = \sum_{\boldsymbol{l}} a_{\boldsymbol{l}} \cdot h_{\boldsymbol{l}}$ with $a_{\boldsymbol{l}} = h(\boldsymbol{l})\, \mu^{(n)}(\boldsymbol{l})$, and h is a convex combination of the functions $h_{\boldsymbol{l}}$ with $a_{\boldsymbol{l}} > 0$ (which happens precisely when $\boldsymbol{l} \in \operatorname{supp}(\mu^{(n)})$). By minimality of h we must have $h_{\boldsymbol{l}} = h$ for each $\boldsymbol{l} \in \operatorname{supp}(\mu^{(n)})$. This is true for every n, and $h_{\boldsymbol{l}} = h$ for every $\boldsymbol{l} \in \mathbb{Z}^d$, that is,

$$h(\boldsymbol{k}+\boldsymbol{l}) = h(\boldsymbol{k})\, h(\boldsymbol{l}) \quad \text{for all } \boldsymbol{k}, \boldsymbol{l} \in \mathbb{Z}^d.$$

Now let $\boldsymbol{e}_i$ be i-th unit vector in $\mathbb{Z}^d$ ($i = 1, \dots, d$) and $\boldsymbol{c} \in \mathbb{R}^d$ the vector whose i-th coordinate is $c_i = \log h(\boldsymbol{e}_i)$. Then

$$h(\boldsymbol{k}) = e^{\boldsymbol{c}\cdot\boldsymbol{k}},$$

and harmonicity of h implies $\varphi(\boldsymbol{c}) = 1$.

B. Conversely, let $\boldsymbol{c} \in \mathbb{R}^d$ with $\varphi(\boldsymbol{c}) = 1$. Consider the $f_{\boldsymbol{c}}$-process:

$$p_{f_{\boldsymbol{c}}}(\boldsymbol{k}, \boldsymbol{l}) = \frac{p(\boldsymbol{k}, \boldsymbol{l})\, f_{\boldsymbol{c}}(\boldsymbol{l})}{f_{\boldsymbol{c}}(\boldsymbol{k})} = e^{\boldsymbol{c}\cdot(\boldsymbol{l}-\boldsymbol{k})}\, \mu(\boldsymbol{l}-\boldsymbol{k}).$$

These are the transition probabilities of the random walk on $\mathbb{Z}^d$ whose law is the probability distribution $\mu_{\boldsymbol{c}}$, where

$$\mu_{\boldsymbol{c}}(\boldsymbol{k}) = e^{\boldsymbol{c}\cdot\boldsymbol{k}}\, \mu(\boldsymbol{k}), \quad \boldsymbol{k} \in \mathbb{Z}^d.$$

We can apply Theorem 8.2 to $\mu_{\boldsymbol{c}}$ in the place of μ, and infer that all bounded harmonic functions with respect to $P_{f_{\boldsymbol{c}}}$ are constant. Corollary 7.11 yields that $f_{\boldsymbol{c}}$ is minimal with respect to P. □

8.8 Exercise. Check carefully all steps of the proofs to show that the last two theorems are also valid if instead of irreducibility, one only assumes that $\operatorname{supp}(\mu)$ generates $\mathbb{Z}^d$ as a *group*. □

Since we have developed Martin boundary theory only in the irreducible case, we return to this assumption. We set

$$\boldsymbol{C} = \{\boldsymbol{c} \in \mathbb{R}^d : \varphi(\boldsymbol{c}) = 1\}. \tag{8.9}$$

This set is non-empty, since it contains $\mathbf{0}$. By Theorem 8.7, the minimal Martin boundary is parametrised by $\boldsymbol{C}$. For a sequence $(\boldsymbol{c}_n)$ we have $\boldsymbol{c}_n \to \boldsymbol{c}$ if and only if $f_{\boldsymbol{c}_n} \to f_{\boldsymbol{c}}$ pointwise on $\mathbb{Z}^d$. Now recall that the topology on the Martin boundary is that of pointwise convergence of the Martin kernels $K(\cdot, \xi)$, $\xi \in \mathcal{M}$. Thus, the bijection $\boldsymbol{C} \to \mathcal{M}_{\min}$ induced by $\boldsymbol{c} \mapsto f_{\boldsymbol{c}}$ is a homeomorphism. Theorem 7.53 implies the following.

8.10 Corollary. *For every positive function h on $\mathbb{Z}^d$ which is harmonic with respect to μ there is a unique Borel measure ν on $\boldsymbol{C}$ such that*

$$h(\boldsymbol{k}) = \int_{\boldsymbol{C}} e^{\boldsymbol{c}\cdot\boldsymbol{k}}\, d\nu(\boldsymbol{c}) \quad \textit{for all } \boldsymbol{k} \in \mathbb{Z}^d.$$

8.11 Exercise (Alternative proof of Theorems 8.2 and 8.7). A shorter proof of the two theorems can be obtained as follows.

• Start with part A of the proof of Theorem 8.7, showing that every minimal harmonic function has the form $f_{\boldsymbol{c}}$ with $\boldsymbol{c} \in \boldsymbol{C}$.

• Arguing as before Corollary 8.10, infer that there is a subset $\boldsymbol{C}'$ of $\boldsymbol{C}$ such that the mapping $\boldsymbol{c} \mapsto f_{\boldsymbol{c}}$ ($\boldsymbol{c} \in \boldsymbol{C}'$) induces a homeomorphism from $\boldsymbol{C}'$ to $\mathcal{M}_{\min}$. It follows that every positive harmonic function h has a unique integral representation as in Corollary 8.10 with $\nu(\boldsymbol{C} \setminus \boldsymbol{C}') = 0$.

Now consider a *bounded* harmonic function h. Show that the representing measure must be $\nu = c \cdot \delta_{\mathbf{0}}$, a multiple of the point mass at $\mathbf{0}$.

[Hint: if $\operatorname{supp} \nu \neq \{\mathbf{0}\}$, show that the function h cannot be bounded.]

Theorem 8.2 follows.

• Now conclude with part B of the proof of Theorem 8.7 without any change, showing *a posteriori* that $\boldsymbol{C}' = \boldsymbol{C}$. □

We remark that the proof outlined in the last exercises is shorter than the road taken above only because it uses the highly non-elementary integral representation theorem. Thus, our completely elementary approach to the proofs of Theorems 8.2 and 8.7 is in reality more economic.

For the rest of this chapter, we assume in addition to irreducibility that P has finite range, that is, $\operatorname{supp}(\mu)$ is finite. We analyze the properties of the function φ.

$$\varphi(\boldsymbol{c}) = \sum_{\boldsymbol{k} \in \operatorname{supp}(\mu)} e^{\boldsymbol{c}\cdot\boldsymbol{k}}\, \mu(\boldsymbol{k})$$

is a finite sum of exponentials, hence a convex function, defined and differentiable on the whole of $\mathbb{R}^d$. (It is of course also convex on the set where it is finite when μ does not have finite support).

8.12 Lemma. $\lim_{|\boldsymbol{c}| \to \infty} \varphi(\boldsymbol{c}) = \infty$.

Proof. (8.6) implies that $P^n f_{\boldsymbol{c}} = \varphi(\boldsymbol{c})^n f_{\boldsymbol{c}}$. Hence, applying (8.5) and (8.6) to the probability measure $\mu^{(n)}$,

$$\varphi(\boldsymbol{c})^n = \sum_{\boldsymbol{k} \in \operatorname{supp}(\mu^{(n)})} e^{\boldsymbol{c}\cdot\boldsymbol{k}}\, \mu^{(n)}(\boldsymbol{k}).$$

By irreducibility, we can find $n \in \mathbb{N}$ and $\alpha > 0$ such that

$$\sum_{k=1}^{n} \mu^{(k)}(\pm \boldsymbol{e}_i) \geq \alpha$$

for $i = 1, \ldots, d$. Therefore

$$\sum_{k=1}^{n} \varphi(\boldsymbol{c})^k \geq \sum_{k=1}^{n} \sum_{i=1}^{d} \left(e^{\boldsymbol{c}\cdot\boldsymbol{e}_i}\, \mu^{(k)}(\boldsymbol{e}_i) + e^{-\boldsymbol{c}\cdot\boldsymbol{e}_i}\, \mu^{(k)}(-\boldsymbol{e}_i) \right) \geq \alpha \sum_{i=1}^{d} (e^{c_i} + e^{-c_i}),$$

from which the lemma follows. □

8.13 Exercise. Deduce the following. The set $\{\boldsymbol{c} \in \mathbb{R}^d : \varphi(\boldsymbol{c}) \leq 1\}$ is compact and convex, and its topological boundary is the set $\boldsymbol{C}$ of (8.9). Furthermore, the

function φ assumes its absolute minimum in the unique point $\boldsymbol{c}_{\min}$ which is the solution of the equation

$$\sum_{\boldsymbol{k}\in\operatorname{supp}(\mu)} e^{\boldsymbol{c}\cdot\boldsymbol{k}}\,\mu(\boldsymbol{k})\,\boldsymbol{k} = \boldsymbol{0}. \qquad \square$$

In particular,

$$\boldsymbol{c}_{\min} = \boldsymbol{0} \iff \bar{\mu} = \boldsymbol{0}, \quad \text{where } \bar{\mu} = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \mu(\boldsymbol{k})\,\boldsymbol{k}.$$

(The vector $\bar{\mu}$ is the average displacement of the random walk in one step.) Thus

$$\boldsymbol{C} = \{\boldsymbol{0}\} \iff \bar{\mu} = \boldsymbol{0}.$$

Otherwise, $\boldsymbol{c}_{\min}$ belongs to the interior of the set $\{\varphi \le 1\}$, and the latter is homeomorphic to the closed unit ball in $\mathbb{R}^d$, while the boundary $\boldsymbol{C}$ is homeomorphic to the unit sphere $\mathbb{S}_{d-1} = \{\boldsymbol{u} \in \mathbb{R}^d : |\boldsymbol{u}| = 1\}$ in $\mathbb{R}^d$.

8.14 Corollary. *Let μ be a probability measure on $\mathbb{Z}^d$ that gives rise to an irreducible random walk.*

(1) *If $\bar{\mu} = \boldsymbol{0}$ then all positive μ-harmonic functions are constant, and the minimal Martin boundary consists of a single point.*

(2) *Otherwise, the minimal Martin boundary $\mathcal{M}_{\min}$ is homeomorphic with the set $\boldsymbol{C}$ and with the unit sphere $\mathbb{S}_{d-1}$ in $\mathbb{R}^d$.*

So far, we have determined the minimal Martin boundary $\mathcal{M}_{\min}$ and its topology, and we know how to describe all positive harmonic functions with respect to μ. These results are due to Doob, Snell and Williamson [18], Choquet and Deny [11] and Hennequin [31]. However, they do not provide the complete knowledge of the full Martin compactification. We still have the following questions. (a) Do there exist non-minimal elements in the Martin boundary? (b) What is the topology of the full Martin compactification of $\mathbb{Z}^d$ with respect to μ? This includes, in particular, the problem of determining the "directions of convergence" in $\mathbb{Z}^d$ along which the functions $f_{\boldsymbol{c}}$ ($\boldsymbol{c} \in \boldsymbol{C}$) arise as pointwise limits of Martin kernels. The answers to these questions require very serious work. They are due to Ney and Spitzer [45]. Here, we display the results without proof.

8.15 Theorem. *Let μ be a probability measure with finite support on $\mathbb{Z}^d$ which gives rise to an irreducible random walk which is transient ($\iff \bar{\mu} \ne \boldsymbol{0}$ or $d \ge 3$).*

(a) *If $\bar{\mu} = \boldsymbol{0}$ then the Martin compactification coincides with the one-point-compactification of $\mathbb{Z}^d$.*

(b) *If $\bar{\mu} \neq \mathbf{0}$ then the Martin boundary is homeomorphic with the unit sphere $\mathbb{S}_{d-1}$. The topology of the Martin compactification is obtained as the closure of the immersion of $\mathbb{Z}^d$ into the unit ball via the mapping*

$$\boldsymbol{k} \mapsto \frac{\boldsymbol{k}}{1 + |\boldsymbol{k}|}.$$

If $(\boldsymbol{k}_n)$ is a sequence in $\mathbb{Z}^d$ such that $\boldsymbol{k}_n/(1 + |\boldsymbol{k}_n|) \to \boldsymbol{u} \in \mathbb{S}_{d-1}$ then $K(\cdot, \boldsymbol{k}_n) \to f_{\boldsymbol{c}}$, where $\boldsymbol{c}$ is the unique vector in in $\mathbb{R}^d$ such that $\varphi(\boldsymbol{c}) = 1$ and the gradient $\nabla\varphi(\boldsymbol{c})$ is collinear with $\boldsymbol{u}$.

The proof of this theorem in the original paper of NEY and SPITZER [45] requires a large amount of subtle use of characteristic function theory. A shorter proof, also quite subtle (communicated to the author by M. BABILLOT), is presented in §25.B of the monograph [W2].

In particular, we see from Theorem 8.15 that $\mathcal{M} = \mathcal{M}_{\min}$. From Theorem 8.2 we know that the Poisson boundary is always trivial. In the case $\bar{\mu} \neq \mathbf{0}$ it is easy to find the boundary point to which (Z_n) converges almost surely in the topology of the Martin compactification. Indeed, by the law of large numbers, $\frac{1}{n} Z_n \to \bar{\mu}$ almost surely. Therefore,

$$\frac{Z_n}{1 + |Z_n|} \to \frac{\bar{\mu}}{|\bar{\mu}|} \quad \text{almost surely.}$$

In the following figure, we illustrate the Martin compactification in the case $d = 2$ and $\bar{\mu} \neq 0$. It is a fish-eye's view of the world, seen from the origin.

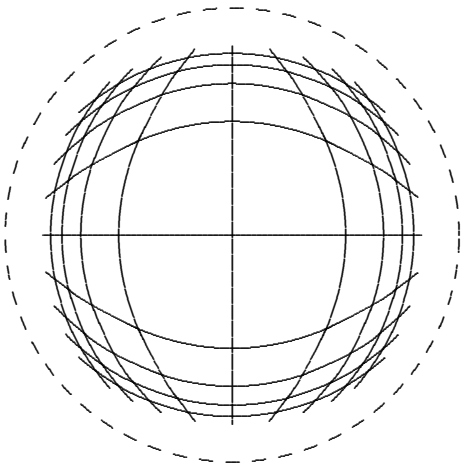

Figure 20

Chapter 9
Nearest neighbour random walks on trees

In this chapter we shall study a large class of Markov chains for which many computations are accessible: we assume that the transition matrix P is adapted to a specific graph structure of the underlying state space X. Namely, the graph $\Gamma(P)$ of Definition 1.6 is supposed to be a *tree*, so that we speak of a random walk on that tree. In general, the use of the term *random walk* refers to a Markov chain whose transition probabilities are adapted in some way to a graph or group structure that the state space carries. Compare with random walks on groups (4.18), where adaptedness is expressed in terms of the group operation. If we start with a locally finite graph, then simple random walk is the standard example of a Markov chain adapted to the graph structure, see Example 4.3. More generally, we can consider *nearest neighbour* random walks, such that

$$p(x, y) > 0 \quad \text{if and only if} \quad x \sim y, \tag{9.1}$$

where $x \sim y$ means that the vertices x and y are neighbours.

Trees lend themselves particularly well to computations with generating functions. At the basis stands Proposition 1.43 (b), concerning cut points. We shall be mainly interested in infinite trees, but will also come back to finite ones.

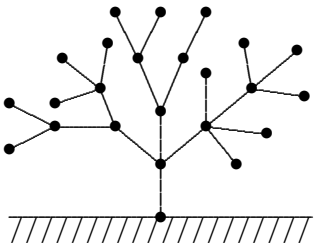

Figure 21

A Basic facts and computations

Recall that a tree T is a finite or infinite, symmetric graph that is connected and contains no cycle. (A cycle in a graph is a sequence $[x_0, x_1, \ldots, x_{k-1}, x_k = x_0]$ such that $k \geq 3$, $x_0, \ldots, x_{k-1}$ are distinct, and $x_{i-1} \sim x_i$ for $i = 1, \ldots, k$.) All our trees have to be finite or denumerable.

We choose and fix a reference point (root) $o \in T$. (Later, o will also serve as the reference point in the construction of the Martin kernel). We write $|x| = d(x, o)$ for the *length* $x \in T$. $|x| \geq 1$ then we define the *predecessor* x^- of x to be the unique neighbour of x which is closer to the root: $|x^-| = |x| - 1$.

In a tree, a *geodesic arc* or *path* is a finite sequence $\pi = [x_0, x_1, \ldots, x_{k-1}, x_k]$ of distinct vertices such that $x_{j-1} \sim x_j$ for $j = 1, \ldots, k$. A *geodesic ray* or just *ray* is a one-sided infinite sequence $\pi = [x_0, x_1, x_2, \ldots]$ of distinct vertices such that $x_{j-1} \sim x_j$ for all $j \in \mathbb{N}$. A 2-*sided infinite geodesic* or just *geodesic* is a sequence $\pi = [\ldots, x_{-2}, x_{-1}, x_0, x_1, x_2, \ldots]$ of distinct vertices such that $x_{j-1} \sim x_j$ for all $j \in \mathbb{Z}$. In each of those cases, we have $d(x_i, x_j) = |j - i|$ for the graph distance in T. An infinite tree that is not locally finite may not possess any ray. (For example, it can be an infinite star, where a root has infinitely many neighbours, all of which have degree 1.)

A crucial property of a tree is that for every pair of vertices x, y, there is a unique geodesic arc $\pi(x, y)$ starting at x and ending at y.

If $x, y \in T$ are distinct then we define the associated *cone* of T as

$$T_{x,y} = \{w \in T : y \in \pi(x, w)\}.$$

(That is, w lies behind y when seen from x.) This is (or spans) a subtree of T. See Figure 22. (The boundary $\partial T_{x,y}$ appearing in that figure will be explained in the next section.)

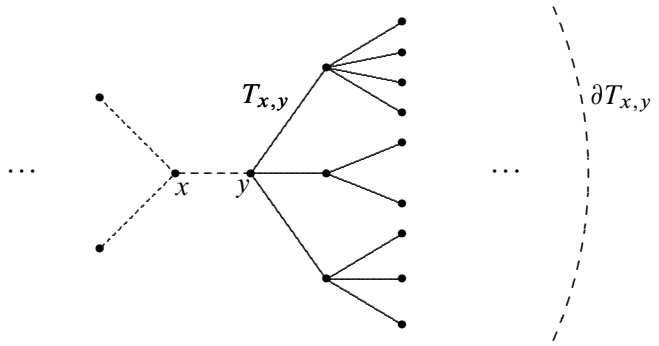

Figure 22

On some occasions, it will be (technically) convenient to consider the following variants of the cones. Let $[x, y]$ be an (oriented) edge of T. We augment $T_{x,y}$ by x and write $B_{[x,y]}$ for the resulting subtree, which we call a *branch* of T. Given P on T, we define $P_{[x,y]}$ on $B_{[x,y]}$ by

$$p_{[x,y]}(v, w) = \begin{cases} p(v, w), & \text{if } v, w \in B_{[x,y]},\ v \neq x, \\ 1, & \text{if } v = x,\ w = y. \end{cases} \tag{9.2}$$

Then $F(y,x|z)$ is the same for P on T and for $P_{[x,y]}$ on $B_{[x,y]}$. Indeed, if the random walk starts at y it cannot leave $B_{[x,y]}$ before arriving at x. (Compare with Exercise 2.12.)

The following is *the* basic ingredient for dealing with nearest neighbour random walks on trees.

9.3 Proposition. (a) *If $x, y \in T$ and w lies on the geodetic arc $\pi(x,y)$ then*

$$F(x,y|z) = F(x,w|z)F(w,y|z).$$

(b) *If $x \sim y$ then*

$$F(y,x|z) = p(y,x)\,z + \sum_{\substack{w:w\sim y\\ w\neq x}} p(y,w)\,z\,F(w,y|z)\,F(y,x|z)$$

$$= \frac{p(y,x)\,z}{1 - \sum_{\substack{w:w\sim y\\ w\neq x}} p(y,w)\,z\,F(w,y|z)}.$$

Proof. Statement (a) follows from Proposition 1.43 (b), because w is a cut point between x and y. Statement (b) follows from (a) and Theorem 1.38:

$$F(y,x|z) = p(y,x)\,z + \sum_{\substack{w:w\sim y\\ w\neq x}} p(y,w)\,z\,F(w,x|z),$$

and $F(w,x|z) = F(w,y|z)F(y,x|z)$ by (a). □

9.4 Exercise. Deduce from Proposition 9.3 (b) that

$$F'(y,x|z) = \frac{F(y,x|z)^2}{p(y,x)\,z^2} + \sum_{\substack{w:w\sim y\\ w\neq x}} F(y,x|z)^2\,\frac{p(y,w)}{p(y,x)}F'(w,y|z).$$

(This will be used in Section H). □

On the basis of these formulas, there is a simple algorithm to compute all the generating functions on a *finite* tree T. They are determined by the functions $F(y,x|z)$, where x and y run through all ordered pairs of neighbours in T.

If y is a *leaf* of T, that is, a vertex which has a unique neighbour x in T, then $F(y,x|z) = p(y,x)\,z$. (In the stochastic case, we have $p(y,x) = 1$, but below we shall also refer to the case when P is substochastic.)

Otherwise, Proposition 9.3 (b) says that $F(y,x|z)$ is computed in terms of the functions $F(w,y|z)$, where w varies among the neighbours of y that are distinct from x. For the latter, one has to consider the sub-branches $B_{[y,w]}$ of $B_{[x,y]}$. Their sizes are all smaller than that of $B_{[x,y]}$. In this way, the size of the branches reduces step by step until one arrives at the leaves. We describe the algorithm, in which every oriented edge $[x,y]$ is recursively labeled by $F(x,y|z)$.

(1) For each leaf y and its unique neighbour x, label the oriented edge $[y, x]$ with $F(y, x|z) = p(y, x)\, z$ ($= z$ when P is stochastic).

(2) Take any edge $[y, x]$ which has not yet been labeled, but such that all the edges $[w, y]$ (where $w \neq x$) already carry their label $F(w, y|z)$. Label $[y, x]$ with the rational function

$$F(y, x|z) = \frac{p(y, x)\, z}{1 - \sum\limits_{\substack{w: w\sim y \\ w\neq x}} p(y, w)\, z\, F(w, y|z)}.$$

Since the tree is finite, the algorithm terminates after $|E(T)|$ steps, where $E(T)$ is the set of oriented edges (two oppositely oriented edges between any pair of neighbours). After that, one can use Lemma 9.3 (a) to compute $F(x, y|z)$ for arbitrary $x, y \in T$: if $\pi(x, y) = [x = x_0, x_1, \ldots, x_k = y]$ then

$$F(x, y|z) = F(x_0, x_1|z) F(x_1, x_2|z) \cdots F(x_{k-1}, x_k|z)$$

is the product of the labels along the edges of $\pi(x, y)$. Next, recall Theorem 1.38:

$$U(x, x|z) = \sum_{y\sim x} p(x, y)\, z\, F(y, x|z)$$

is obtained from the labels of the ingoing edges at $x \in T$, and finally

$$G(x, y|z) = \frac{F(x, y|z)}{1 - U(y, y|z)}.$$

The reader is invited to carry out these computations for her/his favorite examples of random walks on finite trees. The method can also be extended to certain classes of infinite trees & random walks, see below. In a similar way, one can compute the expected hitting times $\mathsf{E}_y(\boldsymbol{t}^x)$, where $x, y \in T$. If $x \neq y$, this is $F'(y, x|1-)$. However, here it is better to proceed differently, by first computing the stationary measure. The following is true for any tree, finite or not.

9.5 Fact. *P is reversible.*

Indeed, for our "root" vertex $o \in T$, we define $\mathsf{m}(o) = 1$. Then we can construct the reversing measure m recursively:

$$\mathsf{m}(x) = \mathsf{m}(x^-) \frac{p(x^-, x)}{p(x, x^-)}.$$

That is, if $\pi(o, x) = [o = x_0, x_1, \ldots, x_{k-1}, x_k = x]$ then

$$\mathsf{m}(x) = \frac{p(x_0, x_1)\, p(x_1, x_2) \cdots p(x_{k-1}, x_k)}{p(x_1, x_0)\, p(x_2, x_1) \cdots p(x_k, x_{k-1})}. \tag{9.6}$$

9.7 Exercise. Change of base point: write m^o for the measure of (9.6) with respect to the root o. Verify that when we choose a different point y as the root, then

$$\mathsf{m}^o(x) = \mathsf{m}^o(y)\,\mathsf{m}^y(x). \qquad \square$$

The measure m of (9.6) is not a probability measure. We remark that one need not always use (9.6) in order to compute m. Also, for reversibility, we are usually only interested in that measure up to multiplication with a constant. For simple random walk on a locally finite tree, we can always use $\mathsf{m}(x) = \deg(x)$, and for a symmetric random walk, we can also take m to be the counting measure.

9.8 Proposition. *The random walk is positive recurrent if and only if the measure* m *of* (9.6) *satisfies* $\mathsf{m}(T) = \sum_{x\in T} \mathsf{m}(x) < \infty$. *In this case,*

$$\mathsf{E}_x(\boldsymbol{t}^x) = \frac{\mathsf{m}(T)}{\mathsf{m}(x)}.$$

When $y \sim x$ *then*

$$\mathsf{E}_y(\boldsymbol{t}^x) = \frac{\mathsf{m}(T_{x,y})}{\mathsf{m}(y)\,p(y,x)}.$$

Proof. The criterion for positive recurrence and the formula for $\mathsf{E}_x(\boldsymbol{t}^x)$ are those of Lemma 4.2.

For the last statement, we assume first that $y = o$ and $x \sim o$. We mentioned already that $\mathsf{E}_o(\boldsymbol{t}^x) = \mathsf{E}_o(\boldsymbol{t}^x_{[x,o]})$, where $\boldsymbol{t}^x_{[x,o]}$ is the first passage time to the point x for the random walk on the branch $B_{[x,o]}$ with transition probabilities given by (9.2). If the latter walk starts at x, then its first step goes to o. Therefore

$$\mathsf{E}_x(\boldsymbol{t}^x_{[x,o]}) = 1 + \mathsf{E}_o(\boldsymbol{t}^x).$$

Now the measure $\mathsf{m}_{[x,o]}$ with respect to base point o that makes the random walk on the branch $B_{[x,o]}$ reversible is given by

$$\mathsf{m}_{[x,o]}(w) = \begin{cases} \mathsf{m}(w), & \text{if } w \in B_{[x,o]} \setminus \{x\}, \\ p(o,x), & \text{if } w = x. \end{cases}$$

Applying to the branch what we stated above for the random walk on the whole tree, we have

$$\mathsf{E}_x(\boldsymbol{t}^x_{[x,o]}) = \mathsf{m}_{[x,o]}(B_{[x,o]})/\mathsf{m}_{[x,o]}(x) = \big(\mathsf{m}(T_{x,o}) + p(o,x)\big)/p(o,x).$$

Therefore

$$\mathsf{E}_o(\boldsymbol{t}^x) = \mathsf{m}(T_{x,o})/p(o,x).$$

Finally, if y is arbitrary and $x \sim y$ then we can use Exercise 9.7 to see how the formula has to be adapted when we change the base point from o to y. $\square$

Let $x, y \in T$ be distinct and $\pi(y,x) = [y = y_0, y_1, \dots, y_k = x]$. By Exercise 1.45

$$\mathsf{E}_y(t^x) = \sum_{j=1}^{k} \mathsf{E}_{y_{j-1}}(t^{y_j}).$$

Thus, we have nice explicit formulas for all the expected first passage times in the positive recurrent case, and in particular, when the tree is finite.

Let us now answer the questions of Example 1.3 (the cat), which concerns the simple random walk on a finite tree.

a) The probability that the cat will ever return to the root vertex o is 1, since the random walk is recurrent (as T is finite).

The probability that it will return at the n-th step (and not before) is $u^{(n)}(o,o)$. The explicit computation of that number may be tedious, depending on the structure of the tree. In principle, one can proceed by computing the rational function $U(o,o|z)$ via the algorithm described above: start at the leaves and compute recursively all functions $F(x,x^-|z)$. Since $\deg(o) = 1$ in our example, we have $U(o,o|z) = z\,F(v,o|z)$ where v is the unique vertex with $v^- = o$. Then one expands that rational function as a power series and reads off the n-th coefficient.

b) The average time that the cat needs to return to o is $\mathsf{E}_o(t^o)$. Since $\mathsf{m}(x) = \deg(x)$ and $\deg(o) = 1$, we get $\mathsf{E}_o(t^o) = \sum_{x\in T} \deg(x) = 2(|T| - 1)$. Indeed, the sum of the vertex degrees is the number of oriented edges, which is twice the number of non-oriented edges. In a tree T, that last number is $|T| - 1$.

c) The probability that the cat returns to the root before visiting a certain subset Y of the set of leaves of the tree can be computed as follows: cut off the leaves of that set, and consider the resulting truncated Markov chain, which is substochastic at the truncation points. For that new random walk on the truncated tree $T^\times$ we have to compute the functions $F^\times(x,x^-) = F^\times(x,x^-|1)$ following the same algorithm as above, but now starting with the leaves of $T^\times$. (The superscript refers of course to the truncated random walk.) Then the probability that we are looking for is $U^\times(o,o) = F^\times(v,o)$, where v is the unique neighbour of o.

A different approach to the same question is as follows: the probability to return to o before visiting the chosen set Y of leaves is the same as the probability to reach o from v before visiting Y. The latter is related with the Dirichlet problem for finite Markov chains, see Theorem 6.7. We have to set $\partial T = \{o\} \cup L$ and $T^o = T \setminus \partial T$. Then we look for the unique harmonic function h in $\mathcal{H}(T^o, P)$ that satisfies $h(o) = 1$ and $h(y) = 0$ for all $y \in Y$. The value $h(v)$ is the required probability. We remark that the same problem also has a nice interpretation in terms of electric currents and voltages, see [D-S].

d) We ask for the expected number of visits to a given leaf y before returning to o. This is $L(o,y) = L(o,y|1)$, as defined in (3.57). Always because $p(o,v) = 1$, this number is the same as the expected number of visits to y before visiting o,

when the random walk starts at v. This time, we "chop off" the root o and consider the resulting truncated random walk on the new tree $T^\bullet = T \setminus \{o\}$, which is strictly substochastic at v. Then the number we are looking for is $G^\bullet(v, y) = G^\bullet(v, y|1)$, where $G^\bullet$ is the Green function of the truncated walk. This can again be computed via the algorithm described above. Note that $U^\bullet(y, y) = F^\bullet(y^-, y)$, since y^- is the only neighbour of the leaf y. Therefore $G^\bullet(v, y) = F^\bullet(v, y)\big/\big(1-F^\bullet(y^-, y)\big)$.

On can again relate this to the Dirichlet problem. This time, we have to set $\partial T = \{v, y\}$. We look for the unique function h that is harmonic on $T^o = T \setminus \partial T$ which satisfies $h(o) = 0$ and $h(y) = 1$. Then $h(x) = F^\bullet(x, y)$ for every x, so that $G^\bullet(v, y) = h(v)\big/\big(1 - h(y^-)\big)$.

B The geometric boundary of an infinite tree

After this prelude we shall now concentrate on infinite trees and transient nearest neighbour random walks. We want to see how the boundary theory developed in Chapter 7 can be implemented in this situation. For this purpose, we first describe the natural geometric compactification of an infinite tree.

9.9 Exercise. Show that a locally finite, infinite tree possesses at least one ray. □

As mentioned above, an infinite tree that is not locally finite might not possess any ray. In the sequel we assume to have a tree that does possess rays. We do not assume local finiteness, but it may be good to keep in mind that case. A ray describes a path from its starting point x_0 to infinity. Thus, it is natural to use rays in order to distinguish different directions of going to infinity. We have to clarify when two rays define the same point at infinity.

9.10 Definition. Two rays $\pi = [x_0, x_1, x_2, \dots]$ and $\pi' = [y_0, y_1, y_2, \dots]$ in the tree T are called *equivalent,* if their symmetric difference is finite, or equivalently, there are $i, j \in \mathbb{N}_0$ such that $x_{i+n} = y_{j+n}$ for all $n \geq 0$.

An *end* of T is an equivalence class of rays.

9.11 Exercise. • Prove that equivalence of rays is indeed an equivalence relation.

• Show that for every point $x \in T$ and end ξ of T, there is a unique geodesic ray $\pi(x, \xi)$ starting at x that is a representative of ξ (as an equivalence class).

• Show that for every pair of distinct ends ξ, η of T, there is a unique geodesic $\pi(\xi, \eta) = [\dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots]$ such that $\pi(x_0, \xi) = [x_0, x_{-1}, x_{-2}, \dots]$ and $\pi(x_0, \eta) = [x_0, x_1, x_2, \dots]$. □

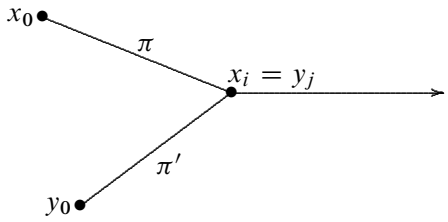

Figure 23. Two equivalent rays.

We write ∂T for the set of all ends of T. Analogously, when $x, y \in T$ are distinct, we define

$$\partial T_{x,y} = \{\xi \in \partial T : y \in \pi(x, \xi)\},$$

compare with Figure 22. This is the set of ends of T which have a representative ray that lies in $T_{x,y}$. Equivalently, we may consider it as the set of ends of the tree $T_{x,y}$. (Attention: work out why this identification is legitimate!)

When $x = o$, the "root", then we just write $T_y = T_{o,y} = T_{y^-,y}$ and $\partial T_y = \partial T_{o,y} = \partial T_{y^-,y}$.

If $\xi, \eta \in T \cup \partial T$ are distinct, then their *confluent* $\xi \wedge \eta$ is the vertex with maximal length on $\pi(o, \xi) \cap \pi(o, \eta)$, see Figure 24. If on the other hand $\xi = \eta$, then we set $\xi \wedge \eta = \eta$. The only case in which $\xi \wedge \eta$ is not a vertex of T is when $\xi = \eta \in \partial T$. For vertices, we have

$$|x \wedge y| = \frac{1}{2}\big(|x| + |y| - d(x, y)\big), \quad x, y \in T.$$

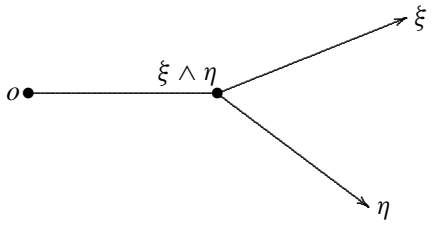

Figure 24

9.12 Lemma. *For all $\xi, \eta, \zeta \in T \cup \partial T$,*

$$|\xi \wedge \eta| \geq \min\{|\xi \wedge \zeta|,\ |\zeta \wedge \eta|\}.$$

Proof. We assume without loss of generality that ξ, η, ζ are distinct and set $x = \xi \wedge \eta$ and $y = \xi \wedge \zeta$. Then we either have $y \in \pi(o, x)$ or $y \in \pi(x, \xi) \setminus \{x\}$.

If $y \in \pi(o, x)$ then

$$|\xi \wedge \eta| = |x| \geq |y| = \min\{\underbrace{|\xi \wedge \zeta|}_{= |y|},\ \underbrace{|\zeta \wedge \eta|}_{\geq |y|}\}.$$

If $y \in \pi(x,\xi) \setminus \{x\}$ then $|\zeta \wedge \eta| = |x|$, and the statement also holds. (The reader is invited to visualize the two cases by figures.) □

We can use the confluents in order to define a new metric on $T \cup \partial T$:

$$\theta(\xi,\eta) = \begin{cases} e^{-|\xi\wedge\eta|}, & \text{if } \xi \neq \eta, \\ 0, & \text{if } \xi = \eta. \end{cases}$$

9.13 Proposition. *θ is an* ultrametric *on $T \cup \partial T$, that is,*

$$\theta(\xi,\eta) \leq \max\{\theta(\xi,\zeta),\ \theta(\zeta,\eta)\} \quad \textit{for all } \xi,\eta,\zeta \in T \cup \partial T.$$

Convergence of a sequence (x_n) of vertices or ends in that metric is as follows.

- *If $x \in T$ then $x_n \to x$ if and only if $x_n = x$ for all $n \geq n_0$.*
- *If $\xi \in \partial T$ then $x_n \to \xi$ if and only if $|x_n \wedge \xi| \to \infty$.*

T is a discrete, dense subset of this metric space. The space is totally disconnected: *every point has a neighbourhood base consisting of open and closed sets. If the tree T is locally finite, then $T \cup \partial T$ is compact.*

Proof. The ultrametric inequality follows from Lemma 9.12, so that θ is a metric.

Let $x \in T$ and $|x| = k$. Then $|x \wedge v| \leq k$ and consequently $\theta(x,v) \geq e^{-k}$ for every $v \in T \cup \partial T$. Therefore the open ball with radius e^{-k} centred at x contains only x. This means that T is discrete.

The two statements about convergence are now immediate, and the second implies that T is dense.

A neighbourhood base of $\xi \in \partial T$ is given by the family of all sets

$$\begin{aligned} \{\eta \in T \cup \partial T : \theta(\xi,\eta) < e^{-k+1}\} &= \{\eta \in T \cup \partial T : |\xi \wedge \eta| > k-1\} \\ &= \{\eta \in T \cup \partial T : |\xi \wedge \eta| \geq k\} \\ &= \{\eta \in T \cup \partial T : \theta(\xi,\eta) \leq e^{-k}\} \\ &= T_{x_k} \cup \partial T_{x_k}, \end{aligned}$$

where $k \in \mathbb{N}$ and x_k is the point on $\pi(o,\xi)$ with $|x_k| = k$. These sets are open as well as closed balls.

Finally, assume that T is locally finite. Since T is dense, it is sufficient to show that every sequence in T has a convergent subsequence in $T \cup \partial T$. Let (x_n) be such a sequence. If there is $x \in T$ such that $x_n = x$ for infinitely many n, then we have a subsequence converging to x. We now exclude this case.

There are only finitely many cones $T_y = T_{o,y}$ with $y \sim o$. Since $x_n \neq o$ for all but finitely many n, there must be $y_1 \sim o$ such that T_{y_1} contains infinitely many of the x_n. Again, there are only finitely many cones T_v with $v^- = y_1$, so that there

must be y_2 with $y_2^- = y_1$ such that T_{y_2} contains infinitely many of the x_n. We now proceed inductively and construct a sequence (y_k) such that $y_{k+1}^- = y_k$ and each cone T_{y_k} contains infinitely many of the x_n. Then $\pi = [o, y_1, y_2, \dots]$ is a ray. If ξ is the end represented by π, then it is clearly an accumulation point of (x_n). □

If T is locally finite, then we set

$$\widehat{T} = T \cup \partial T,$$

and this is a compactification of T in the sense of Section 7.B. The ideal boundary of T is ∂T.

However, when T is not locally finite, $T \cup \partial T$ is not compact. Indeed, if y is a vertex which has infinitely many neighbours x_n, $n \in \mathbb{N}$, then (x_n) has no convergent subsequence in $T \cup \partial T$. This defect can be repaired as follows. Let

$$T^\infty = \{y \in T : \deg(y) = \infty\}$$

be the set of all vertices with infinite degree. We introduce a new set

$$T^* = \{y^* : y \in T^\infty\},$$

disjoint from $T \cup \partial T$, such that the mapping $y \mapsto y^*$ is one-to-one on T^∞. We call the elements of T^* the *improper vertices.* Then we define

$$\widehat{T} = T \cup \partial^* T, \quad \text{where } \partial^* T = T^* \cup \partial T. \tag{9.14}$$

For distinct points $x, y \in T$, we define $\widehat{T}_{x,y}$ accordingly: it consists of all vertices in $T_{x,y}$, of all improper vertices v^* with $v \in T_{x,y}$ and of all ends $\xi \in \partial T$ which have a representative ray that lies in that cone. The topology on $\widehat{T}$ is such that each singleton $\{x\} \subset T$ is open (so that T is discrete). A neighbourhood base of $\xi \in \partial T$ is given by all $\widehat{T}_{x,y}$ that contain ξ, and it is sufficient to take only the sets $\widehat{T}_y = \widehat{T}_{o,y}$, where $y \in \pi(o, \xi) \setminus \{o\}$. A neighbourhood base of $y^* \in T^*$ is given by the family of all sets that are finite intersections of sets $\widehat{T}_{x,v}$ that contain y, and it is sufficient to take just all the finite intersections of sets of the form $\widehat{T}_{x,y}$, where $x \sim y$.

We explain what convergence of a sequence (w_n) of elements of $T \cup \partial T$ means in this topology.

- We have $w_n \to x \in T$ precisely when $w_n = x$ for all but finitely many n.
- We have $w_n \to \xi \in \partial T$ precisely when $|w_n \wedge \xi| \to \infty$.
- We have $w_n \to y^* \in T^*$ precisely when for every finite set A of neighbours of y, one has $\pi(y, w_n) \cap A = \emptyset$ for all but finitely many n. (That is, w_n "rotates" around y.)

Convergence of a sequence (x_n^*) of improper vertices is as follows.

- We have $x_n^* \to x^* \in T^*$ or $x_n^* \to \xi \in \partial T$ precisely when in the above sense, $x_n \to x^*$ or $x_n \to \xi$, respectively.

9.15 Exercise. Verify in detail that this is indeed the convergence of sequences induced by the topology introduced above. Prove that $\widehat{T}$ is compact. □

The following is a useful criterion for convergence of a sequence of vertices to an *end.*

9.16 Lemma. *A sequence (x_n) of vertices of T converges to an end if and only if*

$$|x_n \wedge x_{n+1}| \to \infty.$$

Proof. The "only if" is straightforward and left to the reader.

For sufficiency of the criterion, we first observe that confluents can also be defined when improper vertices are involved: for $x^*, w^* \in T^*$, $y \in T$, and $\xi \in \partial T$,

$$x^* \wedge y = x \wedge y, \quad x^* \wedge w^* = x \wedge w, \quad \text{and} \quad x^* \wedge \xi = x \wedge \xi.$$

Let (x_n) satisfy $|x_n \wedge x_{n+1}| \to \infty$. By Lemma 9.12,

$$|x_m \wedge x_n| \geq \min\{|x_k \wedge x_{k+1}| : i = m, \dots, n-1\}$$

which tends to ∞ as $m \to \infty$ and $n > m$. Suppose that (x_n) has two distinct accumulation points η, ξ in $\widehat{T}$. Let $v = \eta \wedge \xi$, a vertex of T. Then there must be infinitely many m and $n > m$ such that $x_m \wedge x_n = v$, a contradiction. Therefore (x_n) must have a limit in $\widehat{T}$. This cannot be a vertex. If it were an improper vertex x^* then again $x_m \wedge x_n = x$ for infinitely many m and $n > m$, another contradiction. Therefore the limit must be an end. □

The reader may note that the difficulty that has lead us to introducing the improper vertices arises because we insist that T has to be discrete in our compactification. The following will also be needed later on for the study of random walks, both in the case when the tree is locally finite or not.

A function $f : T \to \mathbb{R}$ is called *locally constant,* if the set of edges

$$\{[x, y] \in E(T) : f(x) \neq f(y)\}$$

is finite.

9.17 Exercise. Show that the locally constant functions constitute a linear space $\mathcal{L}$ which is spanned by the set

$$\mathcal{L}_0 = \{\mathbf{1}_{T_{x,y}} : [x, y] \in E(T)\}$$

of the indicator functions of all branches of T. (Once more, edges are oriented here!) □

We can now apply Theorem 7.13: there is a unique compactification of T which contains T as a discrete, dense subset with the following properties: (a) every function in $\mathcal{L}_0$, and therefore also every function in $\mathcal{L}$, extends continuously, and (b) for every pair of points ξ, η in the associated ideal boundary, there is an extended function that assumes different values at ξ and η.

It is now obvious that this is just the compactification described above, $\widehat{T} = T \cup T^* \cup \partial T$. Indeed, the continuous extension of $\mathbf{1}_{T_{x,y}}$ is just $\mathbf{1}_{\widehat{T}_{x,y}}$, and it is clear that those functions separate the points of $\partial^* T$.

C Convergence to ends and identification of the Martin boundary

As one may expect in view of the efforts undertaken in the last section, we shall show that for a transient nearest neighbour random walk on a tree T, the Martin compactification coincides with the geometric compactification. At the end of Section 7.B, we pointed out that in most known concrete examples where one is able to achieve that goal, one can also give a direct proof of boundary convergence. In the present case, this is particularly simple.

9.18 Theorem. *Let (Z_n) be a transient nearest neighbour random walk with stochastic transition matrix on the infinite tree T. Then Z_n converges almost surely to a random end of T: there is a ∂T-valued random variable Z_∞ such that in the topology of the geometric compactification $\widehat{T}$,*

$$\lim_{n\to\infty} Z_n = Z_\infty \quad \mathsf{Pr}_x\text{-almost surely for every starting point } x \in T.$$

Proof. Consider the following subset of the trajectory space $\Omega = T^{\mathbb{N}_0}$:

$$\Omega_0 = \left\{ \omega = (x_n)_{n\geq 0} \in \Omega : \begin{array}{l} x_n \sim x_{n-1}, \\ |\{n \in \mathbb{N}_0 : x_n = y\}| < \infty \text{ for every } y \in T \end{array} \right\}.$$

Then $\mathsf{Pr}_x(\Omega_0) = 1$ for every starting point x, and of course $Z_n(\omega) = x_n$. Now let $\omega = (x_n) \in \Omega_0$. We define recursively a subsequence $(x_{\boldsymbol{\epsilon}_k})$, starting with

$$\boldsymbol{\epsilon}_0 = \max\{n : x_n = x_0\}, \quad \text{and} \quad \boldsymbol{\epsilon}_{k+1} = \max\{n : x_n = x_{\boldsymbol{\epsilon}_k+1}\}. \tag{9.19}$$

By construction of Ω_0, $\boldsymbol{\epsilon}_k$ is well defined for each k. The points $x_{\boldsymbol{\epsilon}_k}$, $k \in \mathbb{N}_0$, are all distinct, and $x_{\boldsymbol{\epsilon}_{k+1}} \sim x_{\boldsymbol{\epsilon}_k}$. Thus $[x_{\boldsymbol{\epsilon}_0}, x_{\boldsymbol{\epsilon}_1}, x_{\boldsymbol{\epsilon}_2}, \dots]$ is a ray and defines an end ξ of T. Also by construction, $x_n \in T_{x_0, x_{\boldsymbol{\epsilon}_k}}$ for all $k \geq 1$ and all $n \geq \boldsymbol{\epsilon}_k$. Therefore $x_n \to \xi$. □

In probabilistic terminology, the *exit times* $\boldsymbol{\epsilon}_k$ are non-negative random variables (but not stopping times). If we recall the notation of (7.34), then $\boldsymbol{\epsilon}_k = \boldsymbol{\epsilon}_{B(x_0,k)}$,

the exit time from the ball $B(x_0, k) = \{y \in T : d(y, x_0) \le k\}$ in the graph metric of T. We see that by transience, those exit times are almost surely finite even in the case when T is not locally finite and the balls may be infinite. The following is of course quite obvious from the beginning.

9.20 Corollary. *If the tree T contains no geodesic ray then every nearest neighbour random walk on T is recurrent.*

Note that when T is not locally finite, then it may well happen that its diameter $\sup\{d(x, y) : x, y \in T\}$ is infinite, while T possesses no ray.

Let us now study the Martin compactification. Setting $z = 1$ in Proposition 9.3 (a), we obtain for the Martin kernel with respect to the root o,

$$K(x, y) = \frac{F(x, y)}{F(o, y)} = \frac{F(x, x \wedge y)\, F(x \wedge y, y)}{F(o, x \wedge y)\, F(x \wedge y, y)} = \frac{F(x, x \wedge y)}{F(o, x \wedge y)} = K(x, x \wedge y).$$

If we write $\pi(o, x) = [o = v_0, v_1, \dots, v_k = x]$ then

$$K(x, y) = \begin{cases} K(x, o) = F(x, o) & \text{for } y \in T_{v_1, o}, \\ K(x, v_j) & \text{for } y \in T_{v_j} \setminus T_{v_{j+1}} \quad (j = 1, \dots, k-1), \\ K(x, x) = 1/F(o, x) & \text{for } y \in T_x. \end{cases}$$

This can be rewritten as

$$K(x, y) = K(x, o) + \sum_{j=1}^{k} \big(K(x, v_j) - K(x, v_{j-1})\big)\, \mathbf{1}_{T_{v_j}}(y). \tag{9.21}$$

We see that $K(x, \cdot)$ is a locally constant function, which leads us to the following.

9.22 Theorem. *Let P be the stochastic transition matrix of a transient nearest neighbour random walk on the infinite tree T. Then the Martin compactification of (T, P) coincides with the geometric compactification $\widehat{T}$. The continuous extension of the Martin kernel to the Martin boundary $\partial^* T = T^* \cup \partial T$ is given by*

$$K(x, y^*) = K(x, y) \ \text{ for } y^* \in T^*, \quad \text{and} \quad K(x, \xi) = K(x, x \wedge \xi) \ \text{ for } \xi \in \partial T.$$

Each function $K(\cdot, \xi)$, where $\xi \in \partial T$, is harmonic. The minimal Martin boundary is the space of ends ∂T of T.

Proof. We know from the preceding computation that for each $x \in T$, the kernel $K(x, \cdot)$ is locally constant on T. Therefore it has a continuous extension to $\widehat{T}$. This extension is the one stated in the theorem: when $y_n \to y^*$ then $x \wedge y_n = x \wedge y$ for all but finitely many n, and $K(x, y_n) = K(x, x \wedge y) = K(x, y)$ for those n. Analogously, if $y_n \to \xi$ then $x \wedge y_n = x \wedge \xi$ and thus $K(x, y_n) = K(x, x \wedge \xi)$ for all but finitely many n.

We next show that $K(\cdot, \xi)$ is harmonic when $\xi \in \partial T$. (When T is locally finite, this is clear, because all extended kernels are harmonic when P has finite range.) Let $x \in T$ and let v be the point on $\pi(o, \xi)$ with $v^- = x \wedge \xi$. Then $K(y, \xi) = K(y, v)$ for all $y \sim x$, and $K(x, \xi) = K(x, v)$. Now the function $K(\cdot, v) = G(\cdot, v)/G(o, v)$ is harmonic in every point except v. In particular, it is harmonic in x. Therefore $K(\cdot, \xi)$ is harmonic in every x.

The extended kernel separates the boundary points:

1.) If $w^*, y^* \in T^*$ are distinct, then the functions $K(\cdot, w^*) = K(\cdot, w)$ and $K(\cdot, y^*) = K(\cdot, y)$ are distinct, since the first is harmonic everywhere except at w, while the second is not harmonic at y.

2.) If $y^* \in T^*$ and $\xi \in \partial T$ then $K(\cdot, \xi)$ is harmonic, while $K(\cdot, y^*)$ is strictly superharmonic at y. Therefore the two functions do not coincide.

3.) The interesting case is the one where $\xi, \eta \in \partial T$ are distinct. Let $x = \xi \wedge \eta$ and let y be the neighbour of x on $\pi(x, \xi)$, see Figure 25.

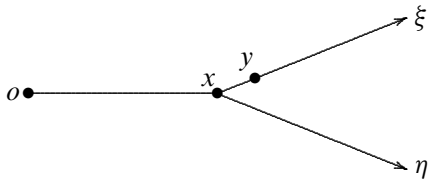

Figure 25

Then, since $F(o, y) = F(o, x)F(x, y)$,

$$\frac{K(y, \eta)}{K(y, \xi)} = \frac{K(y, x)}{K(y, y)} = \frac{F(o, y)F(y, x)}{F(o, x)} = F(x, y)F(y, x) \leq U(x, x)$$

by Exercise 1.44. Now $U(x, x) < 1$ by transience, so that $K(\cdot, \eta) \neq K(\cdot, \xi)$.

We see that the Martin compactification is $\widehat{T}$. The last step is to show that $K(\cdot, \xi)$ is a minimal harmonic function for every end $\xi \in \partial T$. Suppose that

$$K(\cdot, \xi) = a \cdot h_1 + (1 - a) \cdot h_2 \quad (0 < a < 1)$$

for two positive harmonic functions with $h_i(o) = 1$.

Let $x \in T$, and let y be an arbitrary point on the ray $\pi(x, \xi)$. Lemma 6.53, applied to $A = \{y\}$, implies that $h(x) \geq F(x, y)h(y)$ for every positive harmonic function. [This can be seen more directly: $G_h(x, y) = G(x, y)h(y)/h(x)$ is the Green kernel associated with the h-process, and $G_h(x, y) = F_h(x, y)G_h(y, y) \leq G_h(y, y)$. Dividing by $G_h(y, y)$ and multiplying by $h(x)$, one gets the inequality.] On the other hand, our choice of y implies – by Lemma 9.3 (a), as so often – that

$K(x,\xi) = F(x,y)K(y,\xi)$. Therefore

$$\begin{aligned} K(x,\xi) &= a\cdot h_1(x) + (1-a)\cdot h_2(x) \\ &\geq a\cdot F(x,y)h_1(y) + (1-a)\cdot F(x,y)h_2(y) \\ &= F(x,y)K(y,\xi) = K(x,\xi). \end{aligned}$$

The inequality in the middle cannot be strict, and we deduce that

$$F(x,y)h_i(y) = h_i(x) \quad (i=1,2)$$

for all $x \in T$ and all $y \in \pi(x,\xi)$. In particular, choose $y = x\wedge\xi$. Then, applying the last formula to the points x and o (in the place of x),

$$h_i(x) = \frac{h_i(x)}{h_i(o)} = \frac{F(x,y)h_i(y)}{F(o,y)h_i(y)} = K(x,y) = K(x,\xi).$$

We conclude that $K(\cdot,\xi)$ is minimal. □

We now study the family of limit distributions $(\nu_x)_{x\in T}$ on the space of ends ∂T, given by

$$\nu_x(B) = \mathsf{Pr}_x[Z_\infty \in B], \quad B \text{ a Borel set in } \partial T.$$

(We can also consider ν_x as a measure on the compact set $\partial^* T = T^* \cup \partial T$ that does not charge T^*.) For any fixed x, the Borel σ-algebra of ∂T is generated by the family of sets $\partial T_{x,y}$, where y varies in $T\setminus\{x\}$. Therefore ν_x is determined by the measures of those sets.

9.23 Proposition. *Let $x, y \in T$ be distinct, and let w be the neighbour of y on the arc $\pi(x,y)$. Then*

$$\nu_x(\partial T_{x,y}) = F(x,y)\frac{1-F(y,w)}{1-F(w,y)F(y,w)}.$$

Proof. Note that $\partial T_{x,y} = \partial T_{w,y}$. If $Z_0 = x$ and $Z_\infty \in \partial T_{x,y}$ then $\boldsymbol{s}^y < \infty$, the random walk has to pass through y, since y is a cut point between x and $T_{x,y}$. Therefore, via the Markov property,

$$\begin{aligned} \nu_x(\partial T_{x,y}) &= \mathsf{Pr}_x[Z_\infty \in \partial T_{x,y},\ \boldsymbol{s}^y < \infty] \\ &= \sum_{n=1}^{\infty} \mathsf{Pr}_x[Z_\infty \in \partial T_{x,y} \mid \boldsymbol{s}^y = n]\ \mathsf{Pr}_x[\boldsymbol{s}^y = n] \\ &= \sum_{n=1}^{\infty} \mathsf{Pr}_y[Z_\infty \in \partial T_{x,y}]\ \mathsf{Pr}_x[\boldsymbol{s}^y = n] \\ &= F(x,y)\,\nu_y(\partial T_{w,y}) = F(x,y)\big(1-\nu_y(\partial T_{y,w})\big), \end{aligned}$$

since $\partial T_{w,y} = \partial T \setminus \partial T_{y,w}$. In particular,

$$\nu_w(\partial T_{w,y}) = F(w,y)\big(1 - \nu_y(\partial T_{y,w})\big) \quad \text{and}$$
$$\nu_y(\partial T_{y,w}) = F(y,w)\big(1 - \nu_w(\partial T_{w,y})\big).$$

From these two equations, we compute

$$\nu_w(\partial T_{w,y}) = F(w,y)\frac{1 - F(y,w)}{1 - F(w,y)F(y,w)}.$$

Finally, we have $\nu_x(\partial T_{x,y}) = F(x,w)\,\nu_w(\partial T_{w,y})$, since the random walk starting at x must pass through w when $Z_\infty \in \partial T_{x,y}$. Recalling that $F(x,y) = F(x,w)F(w,y)$, this leads to the proposed formula. □

The next formula follows from the fact that $\nu_x(\partial T) = 1$.

$$\sum_{y\,:\,y\sim x} F(x,y)\frac{1 - F(y,x)}{1 - F(x,y)F(y,x)} = 1 \quad \text{for every } x \in T. \tag{9.24}$$

We are primarily interested in $\nu_o = \nu^{\mathbf{1}}$, the probability measure on ∂T that appears in the integral representation of the constant harmonic function $\mathbf{1}$. As mentioned already, the sets $\partial T_x = \partial T_{o,x}$, where $x \neq o$, are a base of the topology on ∂T. By the above,

$$\nu_o(\partial T_x) = F(o,x)\frac{1 - F(x,x^-)}{1 - F(x^-,x)F(x,x^-)}, \tag{9.25}$$

where (recall) x^- is the predecessor of x on $\pi(o,x)$. We see that the support of ν_o is

$$\begin{aligned}\operatorname{supp}(\nu_o) &= \{\xi \in \partial T : \nu_o(\partial T_x) > 0 \text{ for all } x \in \pi(o,\xi),\ x \neq o\}\\ &= \{\xi \in \partial T : F(x,x^-) < 1 \text{ for all } x \in \pi(o,\xi),\ x \neq o\}.\end{aligned}$$

We call an end ξ of T *transient* if $F(x,x^-) < 1$ for all $x \in \pi(o,\xi)\setminus\{o\}$, and we call it *recurrent* otherwise. This terminology is justified by the following observations: if $x \sim y$ then the function $z\ F(y,x|z)$ is the generating function associated with the first return time $\boldsymbol{t}^x_{[x,y]}$ to x for $P_{[x,y]}$. Thus, $F(y,x) = 1$ if and only if $P_{[x,y]}$ on the branch $B_{[x,y]}$ is recurrent. In this case we shall also say more sloppily that the cone $T_{x,y}$ is recurrent, and call it transient, otherwise. We see that an end ξ is transient if and only if each cone T_x is transient, where $x \in \pi(o,\xi) \setminus \{o\}$.

9.26 Exercise. Show that when $x \sim y$ and $F(y,x) = 1$ then $F(w,v) = 1$ for all $v, w \in T_{x,y}$ with $v \sim w$ and $d(v,y) < d(w,y)$. Thus, when $T_{x,y}$ is recurrent, then all ends in $\partial T_{x,y}$ are recurrent.

Conclude that transience of an end does not depend on the choice of the root o. □

We see that for a transient random walk on T, there must be at least one transient end, and that all the measures ν_x, $x \in T$, have the same support, which consists precisely of all transient ends. This describes the Poisson boundary. Furthermore, the transient ends together with the origin o span a subtree T_{tr} of T, the *transient skeleton* of (T, P). Namely, by the above,

$$T_{\mathrm{tr}} = \{o\} \cup \{x \neq o : F(x, x^-) < 1\} \tag{9.27}$$

is such that when $x \in T_{\mathrm{tr}} \setminus \{o\}$ then $x^- \in T_{\mathrm{tr}}$, and x must have at least one forward neighbour. Then, by construction ∂T_{tr} consists precisely of all transient ends. Therefore $Z_\infty \in \partial T_{\mathrm{tr}}$ Pr_x-almost surely for every x. On its way to the limit at infinity, the random walk (Z_n) can of course make substantial "detours" into $T \setminus T_{\mathrm{tr}}$.

Note that T_{tr} depends on the choice of the root, $T_{\mathrm{tr}} = T_{\mathrm{tr}}^o$. For $x \sim o$, we have three cases.

(a) If $F(x, o) < 1$ and $F(o, x) < 1$, then $T_{\mathrm{tr}}^x = T_{\mathrm{tr}}^o$.

(b) If $F(x, o) < 1$ but $F(o, x) = 1$, then $T_{\mathrm{tr}}^x = T_{\mathrm{tr}}^o \setminus \{o\}$.

(c) If $F(x, o) = 1$ but $F(o, x) < 1$, then $T_{\mathrm{tr}}^x = T_{\mathrm{tr}}^o \cup \{x\}$.

Proceeding by induction on the distance $d(o, x)$, we see that T_{tr}^x and T_{tr}^o only differ by at most finitely many vertices from the geodesic segment $\pi(o, x)$.

It may also be instructive to spend some thoughts on $T_{\mathrm{tr}} = T_{\mathrm{tr}}^o$ as an induced subnetwork of T in the sense of Exercise 4.54;

$$a(x, y) = \mathsf{m}(x)p(x, y), \quad \text{if } x, y \in T_{\mathrm{tr}},\ x \sim y.$$

With respect to those conductances, we get new transition probabilities that are reversible with respect to a new measure m_{tr}, namely

$$\begin{aligned} \mathsf{m}_{\mathrm{tr}}(x) &= \sum_{y \sim x\,:\, y \in T_{\mathrm{tr}}} a(x, y) = \mathsf{m}(x)p(x, T_{\mathrm{tr}}) \quad \text{and} \\ p_{\mathrm{tr}}(x, y) &= p(x, y)/p(x, T_{\mathrm{tr}}), \quad \text{if } x, y \in T_{\mathrm{tr}},\ x \sim y. \end{aligned} \tag{9.28}$$

The resulting random walk is (Z_n), conditioned to stay in T_{tr}.

Let us now look at some examples.

9.29 Example (The homogeneous tree). This is the tree $T = \mathbb{T}_q$ where every vertex has degree $q + 1$, with $q \geq 2$. (When $q = 1$ this is $\mathbb{Z}$, visualized as the two-way-infinite path.)

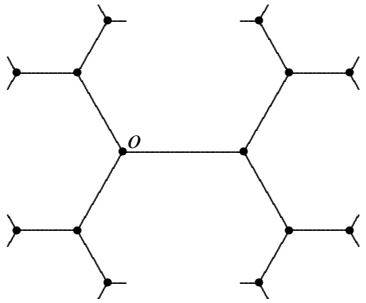

Figure 26

We consider the simple random walk. There are many different ways to see that it is transient. For example, one can use the flow criterion of Theorem 4.51. We choose a root o and define the flow ϕ by $\phi(e) = 1\big/\big((q+1)q^{n-1}\big) = -\phi(\check{e})$ if $e = [x^-, x]$ with $|x| = n$. Then it is straightforward that ϕ is a flow from o to ∞ with input 1 and with finite power.

The hitting probability $F(x, y)$ must be the same for every pair of neighbours x, y. Thus, formula (9.24) becomes $(q+1)F(x,y)\big/\big(1+F(x,y)\big) = 1$, whence $F(x, y) = 1/q$. For an arbitrary pair of vertices (not necessarily neighbours), we get

$$F(x, y) = q^{-d(x,y)}, \quad x, y \in T.$$

We infer that the distribution of Z_∞, given that $Z_0 = o$, is equidistribution on ∂T,

$$\nu_o(\partial T_x) = \frac{1}{(q+1)q^{n-1}}, \quad \text{if } |x| = n \geq 1.$$

We call it "equidistribution" because it is invariant under "rotations" of the tree around o. (In graph theoretical terminology, it is invariant under the group of all self-isometries of the tree that fix o.) Every end is transient, that is, the Poisson boundary (as a set) is $\mathrm{supp}(\nu_o) = \partial T$.

The Martin kernel at $\xi \in \partial T$ is given by

$$K(x, \xi) = q^{-\mathrm{hor}(x,\xi)}, \quad \text{where } \mathrm{hor}(x, \xi) = d(x, x \wedge \xi) - d(o, x \wedge \xi).$$

Below, we shall immediately come back to the function hor. The Poisson–Martin integral representation theorem now says that for every positive harmonic function h on $T = \mathbb{T}_q$, there is a unique Borel measure ν^h on ∂T such that

$$h(x) = \int_{\partial T} q^{-\mathrm{hor}(x,\xi)}\, d\nu^h(\xi).$$

Let us now give a geometric meaning to the function hor. In an arbitrary tree, we can define for $x \in T$ and $\xi \in T \cup \partial T$ the *Busemann function* or *horocycle index* of x with respect to ξ by

$$\mathrm{hor}(x, \xi) = d(x, x \wedge \xi) - d(o, x \wedge \xi). \tag{9.30}$$

9.31 Exercise. Show that for $x \in X$ and $\xi \in \partial T$,

$$\mathrm{hor}(x, \xi) = \lim_{y \to \xi} \mathrm{hor}(x, y) = \lim_{y \to \xi} d(x, y) - d(o, y). \qquad \square$$

The Busemann function should be seen in analogy with classical hyperbolic geometry, where one starts with a geodesic ray $\pi = \big(\pi(t)\big)_{t \geq 0}$, that is, an isometric embedding of the interval $[0, \infty)$ into the hyperbolic plane (or another suitable metric space), and considers the Busemann function

$$\mathrm{hor}(x, \pi) = \lim_{t \to \infty} \big(d\big(x, \pi(t)\big) - d\big(o, \pi(t)\big)\big).$$

A *horocycle* in a tree with respect to an end ξ is a level set of the Busemann function $\mathrm{hor}(\cdot, \xi)$:

$$\mathrm{Hor}_k = \mathrm{Hor}_k(\xi) = \{x \in T : \mathrm{hor}(x, \xi) = k\}, \quad k \in \mathbb{Z}. \tag{9.32}$$

This is the analogue of a horocycle in hyperbolic plane: in the Poincaré disk model of the latter, a horocycle at a boundary point ξ on the unit circle (the boundary of the hyperbolic disk) is a circle inside the disk that is tangent to the unit circle at ξ.

Let us consider another example.

9.33 Example. We now construct a new tree, here denoted again T, by attaching a ray ("hair") at each vertex of the homogeneous tree $\mathbb{T}_q$, see Figure 27. Again, we consider simple random walk on this tree. It is transient by Exercise 4.54, since the $\mathbb{T}_q$ is a subgraph on which SRW is transient.

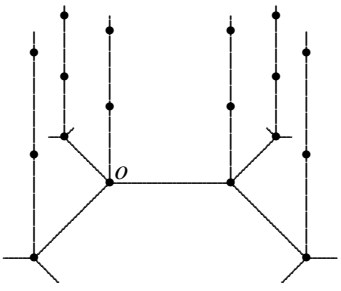

Figure 27

The hair attached at $x \in \mathbb{T}_q$ has one end, for which we write η_x. Since simple random walk on a single ray (a standard birth-and-death chain with forward and

backward probabilities equal to $1/2$) is recurrent, each of the ends η_x, $x \in \mathbb{T}_q$, is recurrent.

Let $\bar{x}$ be the neighbour of x on the hair at x. Then $F(\bar{x}, x) = 1$. Again, by homogeneity of the structure, $F(x, y)$ is the same for all pairs x, y of neighbours in T that belong both to $\mathbb{T}_q$. We infer that $T_{\mathrm{tr}} = \mathbb{T}_q$, and formula (9.24) at x becomes

$$(q+1)\frac{F(x,y)}{1+F(x,y)} + \underbrace{F(x,\bar{x})\frac{1-F(\bar{x},x)}{1-F(x,\bar{x})F(\bar{x},x)}}_{=0} = 1.$$

Again, $F(x, y) = 1/q$ for neighbours $x, y \in \mathbb{T}_q \subset T$. The limit distribution ν_o on ∂T is again uniform distribution on $\partial \mathbb{T}_q \subset \partial T$. What we mean here is that we know already that $\nu_o(\{\eta_x : x \in \mathbb{T}_q\}) = 0$, so that we can think of ν_o as a Borel probability measure on the compact subset $\partial \mathbb{T}_q$ of ∂T, and ν_o is equidistributed on that set in the above sense. (Of course, o is chosen to be in $\mathbb{T}_q$.)

It may be noteworthy that in this example, the set $\{\eta_x : x \in \mathbb{T}_q\}$ of recurrent ends is dense in ∂T in the topology of the geometric compactification of T. On the other hand, each η_x is isolated, in that the singleton $\{\eta_x\}$ is open and closed.

Note also that of course each of the recurrent as well as of the transient ends ξ defines a minimal harmonic function $K(\cdot, \xi)$. However, since $\nu_x(\eta_y) = 0$ for all x, y, every *bounded* harmonic function is constant on each hair. (This can also be easily verified directly via the linear recursion that a harmonic function satisfies on each hair.) That is, it arises from a bounded harmonic function h for SRW on $\mathbb{T}_q$ such that its value is $h(x)$ along the hair attached at x.

For computing the (extended) Martin kernel, we shall need $F(x, \bar{x})$. We know that $F(y, x) = 1/q$ for all $y \sim x$, $y \in \mathbb{T}_q$. By Proposition 9.3 (b),

$$F(x,\bar{x}) = \frac{p(x,\bar{x})}{1 - \sum_{y\sim x, y\in\mathbb{T}_q} p(x,y)\,F(y,x)} = \frac{\frac{1}{q+2}}{1-\frac{q+1}{(q+2)q}} = \frac{q}{q^2+q-1}.$$

Now let $x, y \in \mathbb{T}_q$ be distinct (not necessarily neighbours), $w \in \pi(x, \eta_x)$ (a generic point on the hair at x), and $\xi \in \partial \mathbb{T}_q \subset \partial T$. We need to compute $K(w, \xi)$, $K(w, \eta_y)$ and $K(w, \eta_x)$ in order to cover all possible cases.

(i) We have $F(w, x) = 1$ by recurrence of the end η_x. Therefore

$$K(w,\xi) = F(w,x)\,K(x,\xi) = K(x,\xi) = q^{-\mathrm{hor}(x,\xi)}.$$

(ii) We have $w \wedge \eta_y = w \wedge y$. Therefore

$$K(w,\eta_y) = K(w, w\wedge y) = K(w,y) = F(w,x)\,K(x,y) = K(x,y) = q^{-\mathrm{hor}(x,y)}.$$

(iii) We have $w \wedge \eta_x = w$ for every $w \in \pi(x, \eta_x)$, so that $K(w, \eta_x) = K(w, w) = 1/F(o, w)$. In order to compute this explicitly, we use harmonicity of

the function $h_x = K(\cdot, \eta_x)$. Write $\pi(x, \eta_x) = [x = w_0, w_1 = \bar{x}, w_2, \dots]$. Then

$$h_x(w_0) = \frac{1}{F(o,x)} = q^{|x|} \quad \text{and} \quad h_x(w_1) = \frac{1}{F(o,x)F(x,\bar{x})} = \frac{q^2+q-1}{q}\, q^{|x|},$$

and $h_x(w_n) = \frac{1}{2}\big(h_x(w_{n-1}) + h_x(w_{n+1})\big)$ for $n \geq 1$. This can be rewritten as

$$h_x(w_{n+1}) - h_x(w_n) = h_x(w_n) - h_x(w_{n-1}) = h_x(w_1) - h_x(w_0) = \frac{q^2-1}{q}\, q^{|x|}.$$

We conclude that $h_x(w_n) = h_x(w_0) + n\big(h_x(w_1) - h_x(w_0)\big)$, and find $K(w, \eta_x)$.

We also write the general formula for the kernel at η_y (since w varies and η_y is fixed, we have to exchange the roles of x and y with respect to the above !):

$$K(w, \eta_y) = \begin{cases} q^{|y|}\left(1 + d(w,y)\frac{q^2-1}{q}\right), & \text{if } w \in \pi(y, \eta_y), \\ q^{-\mathrm{hor}(x,y)}, & \text{if } w \in \pi(x, \eta_x),\ x \in \mathbb{T}_q,\ x \neq y. \end{cases}$$

Thus, for every positive harmonic function h on T there is a Borel measure ν^h on ∂T such that

$$\begin{aligned} h(w) &= h_1(w) + h_2(w), \quad \text{where for } w \in \pi(x, \eta_x),\ x \in \mathbb{T}_q, \\ h_1(w) &= \int_{\partial \mathbb{T}_q} q^{-\mathrm{hor}(x,\xi)}\, d\nu^h(\xi) \quad \text{and} \\ h_2(w) &= q^{|x|}\Big(1 + d(w,x)\tfrac{q^2-1}{q}\Big)\, \nu^h(\eta_x) + \sum_{y \in \mathbb{T}_q,\ y \neq x} q^{-\mathrm{hor}(x,y)}\, \nu^h(\eta_y). \end{aligned}$$

The function h_1 in this decomposition is constant on each hair.

D The integral representation of all harmonic functions

Before considering further examples, let us return to the general integral representation of harmonic functions. If h is positive harmonic for a transient nearest neighbour random walk on a tree T, then the measure ν^h on ∂T in the Poisson–Martin integral representation of h is $h(o) \times$ (the limit distribution of the h-process), see (7.47). The Green function of the h-process is $G_h(x,y) = G(x,y)h(y)/h(x)$, and $G_h(x,x) = G(x,x)$. Furthermore, the associated hitting probabilities are $F_h(x,y) = F(x,y)h(y)/h(x)$. We can apply formula (9.25) to the h-process, replacing F with F_h:

$$\nu^h(\partial T_x) = F(o,x)\frac{h(x) - F(x,x^-)h(x^-)}{1 - F(x^-,x)F(x,x^-)}.$$

Conversely, if we start with ν on ∂T, then the associated harmonic function is easily computed on the basis of (9.21). If $x \in T$ and the geodesic arc from o to x is $\pi(o,x) = [o = v_0, v_1, \dots, v_k = x]$ then

$$\begin{aligned} h(x) &= \int_{\partial T} K(x,\cdot)\, d\nu \\ &= K(x,o)\,\nu(\partial T) + \sum_{j=1}^{k} \big(K(x,v_j) - K(x,v_{j-1})\big)\,\nu(\partial T_{v_j}). \end{aligned} \tag{9.34}$$

Note that $K(x,v_j) - K(x,v_{j-1}) = K(x,v_j)\big(1 - F(v_j,v_{j-1})F(v_{j-1},v_j)\big)$.

We see that the integral in (9.34) takes a particularly simple form due to the fact that the integrand is the extension to the boundary of a locally constant function, and we do not need the full strength of Lebesgue's integration theory here. This will allow us to extend the Poisson–Martin representation to get an integral representation over the boundary of *all* (not necessarily positive) harmonic functions.

Before that, we need two further identities for generating functions that are specific to trees.

9.35 Lemma. *For a transient nearest neighbour random walk on a tree T,*

$$G(x,x|z)\,p(x,y)z = \frac{F(x,y|z)}{1 - F(x,y|z)F(y,x|z)} \quad \text{if } y \sim x, \quad \text{and}$$
$$G(x,x|z) = 1 + \sum_{y:y\sim x} \frac{F(x,y|z)F(y,x|z)}{1 - F(x,y|z)F(y,x|z)}$$

for all z with $|z| < \mathsf{r}(P)$, and also for $z = \mathsf{r}(P)$.

Proof. We use Proposition 9.3 (b) (exchanging x and y). It can be rewritten as

$$F(x,y|z) = p(x,y)z + \big(U(x,x|z) - p(x,y)z\,F(y,x|z)\big)F(x,y|z).$$

Regrouping,

$$p(x,y)z\big(1 - F(x,y|z)F(y,x|z)\big) = F(x,y|z)\big(1 - U(x,x|z)\big). \tag{9.36}$$

Since $G(x,x|z) = 1\big/\big(1 - U(x,x|z)\big)$, the first identity follows. For the second identity, we multiply the first one by $F(y,x|z)$ and sum over $y \sim x$ to get

$$\sum_{y:y\sim x} \frac{F(x,y|z)F(y,x|z)}{1 - F(x,y|z)F(y,x|z)} = \sum_{y:y\sim x} p(x,y)z\,G(y,x|z) = G(x,x|z) - 1$$

by (1.34). □

For the following, recall that $T_x = T_{o,x}$ for $x \neq o$. For convenience we write $T_o = T$.

A *signed measure* ν on the collection of all sets

$$\mathcal{F}_o = \{\partial T_x : x \in T\}$$

is a set function $\nu\colon \mathcal{F}_o \to \mathbb{R}$ such that for every x

$$\nu(\partial T_x) = \sum_{y : y^- = x} \nu(\partial T_y).$$

When $\deg(x) = \infty$, the last series has to converge absolutely. Then we use formula (9.34) in order to *define* $\int_{\partial T} K(x,\cdot)\, d\nu$. The resulting function of x is called the *Poisson transform* of the measure ν.

9.37 Theorem. *Suppose that P defines a transient nearest neighbour random walk on the tree T.*

A function $h\colon T \to \mathbb{R}$ is harmonic with respect to P if and only if it is of the form

$$h(x) = \int_{\partial T} K(x,\cdot)\, d\nu,$$

where ν is a signed measure on $\mathcal{F}_o$. The measure ν is determined by h, that is, $\nu = \nu^h$, where

$$\nu^h(\partial T) = h(o) \quad \textit{and} \quad \nu^h(\partial T_x) = F(o,x)\frac{h(x) - F(x,x^-)h(x^-)}{1 - F(x^-,x)F(x,x^-)}, \quad x \neq o.$$

Proof. We have to verify two principal facts.

First, we start with h and have to show that ν^h, as defined in the theorem, is a signed measure on $\mathcal{F}_o$, and that h is the Poisson transform of ν^h.

Second, we start with ν and define h by (9.34). We have to show that h is harmonic, and that $\nu = \nu^h$.

1.) Given the harmonic function h, we claim that for any $x \in T$,

$$h(x) = \sum_{y\,:\,y\sim x} F(x,y)\frac{h(y) - F(y,x)h(x)}{1 - F(x,y)F(y,x)}. \tag{9.38}$$

We can regroup the terms and see that this is equivalent with

$$\Bigl(1 + \sum_{y\,:\,y\sim x} \frac{F(x,y)F(y,x)}{1 - F(x,y|z)F(y,x|z)}\Bigr)h(x) = \sum_{y\,:\,y\sim x} \frac{F(x,y)}{1 - F(x,y)F(y,x)}\, h(y).$$

By Lemma 9.35, this reduces to

$$G(x,x)\, h(x) = \sum_{y\sim x} G(x,x)\, p(x,y)\, h(y),$$

which is true by harmonicity of h.

If we set $x = o$, then (9.38) says that $\nu^h(\partial T) = \sum_{y\sim o} \nu^h(\partial T_y)$. Suppose that $x \neq o$. Then by (9.38),

$$\begin{aligned}\sum_{y:y^-=x} \nu^h(\partial T_y) &= F(o,x) \sum_{y\,:\,y^-=x} F(x,y)\frac{h(y) - F(y,x)h(x)}{1 - F(x,y)F(y,x)} \\ &= F(o,x)\Big(h(x) - F(x,x^-)\frac{h(x^-) - F(x^-,x)h(x)}{1 - F(x,x^-)F(x^-,x)}\Big) \\ &= F(o,x)\frac{h(x) - F(x,x^-)h(x^-)}{1 - F(x,x^-)F(x^-,x)} = \nu^h(\partial T_x).\end{aligned}$$

We have shown that ν^h is indeed a signed measure on $\mathcal{F}_o$. Now we check that $\int_{\partial T} K(x,\cdot)\, d\nu^h = h(x)$. For $x = o$ this is true by definition. So let $x \neq o$. Using the same notation as in (9.34), we simplify

$$\big(K(x,v_j) - K(x,v_{j-1})\big)\, \nu^h(\partial T_{v_j}) = F(x,v_j)\, \big(h(v_j) - F(v_j,v_{j-1})h(v_{j-1})\big),$$

whence we obtain a "telescope sum"

$$\begin{aligned}\int_{\partial T} K(x,\cdot)\, d\nu^h &= K(x,o)h(o) + \sum_{j=1}^{k}\big(F(x,v_j)h(v_j) - F(x,v_{j-1})h(v_{j-1})\big) \\ &= F(x,x)h(x) = h(x).\end{aligned}$$

2.) Given ν, let $h(x) = \int_{\partial T} K(x,\cdot)\, d\nu$.

9.39 Exercise. Show that h is harmonic at o. □

Resuming the proof of Theorem 9.37, we suppose again that $x \neq o$ and use the notation of (9.34). With the fixed index $k = |x|$, we consider the function

$$g(w) = K(w,o)\, \nu(\partial T) + \sum_{j=1}^{k}\big(K(w,v_j) - K(w,v_{j-1})\big)\, \nu(\partial T_{v_j}), \quad w \in T.$$

Since for $i < j$ one has $K(v_i,v_j) = K(v_i,v_i)$, we have $h(v_j) = g(v_j)$ for all $j \le k$. In particular, $h(x) = g(x)$ and $h(x^-) = g(x^-)$. Also,

$$h(y) = g(y) + \big(K(y,y) - K(y,x)\big)\, \nu(\partial T_y), \quad \text{when } y^- = x.$$

Recalling that $P\, K(x,v) = K(x,v) - \frac{1}{G(o,v)}\mathbf{1}_v(x)$, we first compute

$$Pg(x) = g(x) - \frac{1}{G(o,x)}\nu(\partial T_x).$$

Now, using Lemma 9.35,

$$\begin{aligned} Ph(x) &= Pg(x) + \sum_{y:y^-=x} p(x,y)\frac{1-F(x,y)F(y,x)}{F(o,x)F(x,y)}\,\nu(\partial T_y) \\ &= h(x) - \frac{1}{G(o,x)}\nu(\partial T_x) + \frac{1}{F(o,x)}\sum_{y:y^-=x}\frac{1}{G(x,x)}\,\nu(\partial T_y) \\ &= h(x) - \frac{1}{G(o,x)}\nu(\partial T_x) + \frac{1}{G(o,x)}\sum_{y:y^-=x}\nu(\partial T_y) = h(x). \end{aligned}$$

Finally, to show that $\nu^h(\partial T_x) = \nu(\partial T_x)$ for all $x \in T$, we use induction on $k = |x|$. The statement is trivially true for $x = o$. So let once more $|x| = k \geq 1$ and $\pi(o,x) = [v_0 = o, v_1, \dots, v_k = x]$, and assume that $\nu^h(\partial T_{v_j}) = \nu(\partial T_{v_j})$ for all $j < k$. Since we know that $\int_{\partial T} K(x,\cdot)\,d\nu^h = h(x) = \int_{\partial T} K(x,\cdot)\,d\nu$, the induction hypothesis and (9.34) yield that

$$\big(K(x,x) - K(x,v_{j-1})\big)\,\nu^h(\partial T_x) = \big(K(x,x) - K(x,v_{j-1})\big)\,\nu(\partial T_x).$$

Therefore $\nu^h(\partial T_x) = \nu(\partial T_x)$. □

In the case when h is a positive harmonic function, the measure ν^h is a non-negative measure on $\mathcal{F}_o$. Since the sets in $\mathcal{F}_o$ generate the Borel σ-algebra of ∂T, one can justify with a little additional effort that ν^h extends to a Borel measure on ∂T. In this way, one can deduce the Poisson–Martin integral theorem directly, without having to go through the whole machinery of Chapter 7.

Some references are due here. The results of this and the preceding section are basically all contained in the seminal article of Cartier [Ca]. Previous results, regarding the case of random walks on free groups (≡ homogeneous trees) can be found in the note of Dynkin and Malyutov [19]. The part of the proof of Theorem 9.22 regarding minimality of the extended kernels $K(\cdot,\xi)$, $\xi \in \partial T$, is based on an argument of Derriennic [13], once more in the context of free groups. The integral representation of arbitrary harmonic functions is also contained in [Ca], but was also proved more or less independently by various different methods in the paper of Koranyi, Picardello and Taibleson [38] and by Steger, see [21], as well as in some other work.

All those references concern only the locally finite case. The extension to arbitrary countable trees goes back to an idea of Soardi, see [10].

E Limits of harmonic functions at the boundary

The Dirichlet problem at infinity

We now consider another potential theoretic problem. In Chapter 6 we have studied and solved the Dirichlet problem for finite Markov chains with respect to a subset – the "boundary" – of the state space. Now let once more T be an infinite tree and P the stochastic transition matrix of a nearest neighbour random walk on T. The *Dirichlet problem at infinity* is the following. Given a *continuous* function φ on $\partial^* T = \widehat{T} \setminus T$, is there a continuous extension of φ to $\widetilde{T}$ that is harmonic on T?

This problem is related with the limit distributions ν_x, $x \in T$, of the random walk. We have been slightly ambiguous when speaking about the (common) support of these probability measures: since we know that $Z_\infty \in \partial T$ almost surely, so far we have considered them as measures on ∂T. In the spirit of the construction of the geometric compactification (which coincides with the Martin compactification here) and the general theorem of convergence to the boundary, the measures should a priori be considered to live on the compact set $\partial^* T = \widehat{T} \setminus T = \partial T \cup T^*$. This is the viewpoint that we adopt in the present section, which is more topology-oriented. If T is locally finite, then of course there is no difference, and in general, none of the two interpretations regarding where the limit distributions live is incorrect.

In any case, now $\operatorname{supp}(\nu_o)$ is the compact subset of $\partial^* T$ that consists of all points $\xi \in \partial^* T$ with the property that $\nu_x(V) > 0$ for every neighbourhood V of ξ in $\widehat{T}$. We point out that $\operatorname{supp}(\nu_x) = \operatorname{supp}(\nu_o)$ for every $x \in T$, and that ν_x is absolutely continuous with respect to ν_o with Radon–Nikodym density

$$\frac{d\nu_x}{d\nu_o} = K(x, \cdot),$$

see Theorem 7.42.

9.40 Proposition. *If the Dirichlet problem at infinity admits a solution for every continuous function φ on the boundary, then the solution is unique and given by the harmonic function*

$$h(x) = \int_{\partial^* T} \varphi \, d\nu_x = \int_{\partial^* T} K(x, \cdot)\, \varphi \, d\nu_o.$$

Proof. If h is the solution with boundary data given by φ, then h is a bounded harmonic function. By Theorem 7.61, there is a bounded measurable function ψ on the Martin boundary $\mathcal{M} = \widehat{T} \setminus T$ such that

$$h(x) = \int_{\partial^* T} \psi \, d\nu_x.$$

By the probabilistic Fatou theorem 7.67,

$$h(Z_n) \to \psi(Z_\infty) \quad \mathsf{Pr}_o\text{-almost surely.}$$

Since h provides the continuous harmonic extension of φ, and since $Z_n \to Z_\infty$ Pr_o-almost surely,

$$h(Z_n) \to \varphi(Z_\infty) \quad \mathsf{Pr}_o\text{-almost surely.}$$

We conclude that ψ and φ coincide ν_o-almost surely. This proves that the solution of the Dirichlet problem is as stated, whence unique. □

Let us remark here that one can state the Dirichlet problem for an arbitrary Markov chain and with respect to the ideal boundary in an arbitrary compactification of the infinite state space. In particular, it can always be stated with respect to the Martin boundary. In the latter context, Proposition 9.40 is correct in full generality.

There are some degenerate cases: if the random walk is recurrent and $|\partial^* T| \geq 2$ then the Dirichlet problem does not admit solution, because there is some non-constant continuous function on the boundary, while all bounded harmonic functions are constant. On the other hand, when $|\partial^* T| = 1$ then the constant functions provide the trivial solutions to the Dirichlet problem.

The last proposition leads us to the following local version of the Dirichlet problem.

9.41 Definition. (a) A point $\xi \in \partial^* T$ is called *regular for the Dirichlet problem* if for every continuous function φ on $\partial^* T$, its Poisson integral $h(x) = \int_{\partial^* T} \varphi \, d\nu_x$ satisfies

$$\lim_{x \to \xi} h(x) = \varphi(\xi).$$

(b) We say that *the Green kernel vanishes at* ξ, if

$$\lim_{y \to \xi} G(y, o) = 0.$$

We remark that $\lim_{y \to \xi} G(y, o) = 0$ if and only if $\lim_{y \to \xi} G(y, x) = 0$ for some ($\Longleftrightarrow$ every) $x \in T$. Indeed, let $k = d(x, o)$. Then $p^{(k)}(x, o) > 0$, and $G(y, x) p^{(k)}(x, o) \leq G(x, o)$.

Also, if $x^* \in T^*$ and $\{y_k : k \in \mathbb{N}\}$ is an enumeration of the neighbours of x in T, then the Green kernel vanishes at x^* if and only if

$$\lim_{k \to \infty} G(y_k, x) = 0.$$

This holds because if (w_n) is an arbitrary sequence in T that converges to x^*, then $k(n) \to \infty$, where $k(n)$ is the unique index such that $w_n \in T_{x, y_{k(n)}}$: then $G(w_n, x) = F(w_n, y_{k(n)}) G(y_{k(n)}, x) \to 0$.

Regularity of a point $\xi \in \partial^* T$ means that

$$\lim_{y \to \xi} \nu_y = \delta_\xi \quad \text{weakly,}$$

where weak convergence of a sequence of finite measures means that the integrals of any continuous function converge to its integral with respect to the limit measure. The following is quite standard.

9.42 Lemma. *A point $\xi \in \partial^* T$ is regular for the Dirichlet problem if and only if for every set $\partial^* T_{v,w} = \widehat{T}_{v,w} \setminus T_{v,w}$ that contains ξ ($v, w \in T$, $w \neq v$),*

$$\lim_{y \to \xi} \nu_y(\partial^* T_{v,w}) = 1.$$

Proof. The indicator function $\mathbf{1}_{\partial^* T_{v,w}}$ is continuous. Therefore regularity implies that the above limit is 1.

Conversely, assume that the limit is 1. Let φ be a continuous function on the compact set $\partial^* T$, and let $M = \max |\varphi|$. Write h for the Poisson integral of φ. We have $\varphi(\xi) = \int_{\partial^* T} \varphi(\xi)\, d\nu_y(\eta)$. Given $\varepsilon > 0$, we first choose $\widehat{T}_{v,w}$ such that $|\varphi(\xi) - \varphi(\eta)| < \varepsilon$ for all $\eta \in \partial^* T_{v,w}$. Then, if $y \in T_{v,w}$ is close enough to ξ, we have $\nu_y(\partial^* T_{w,v}) < \varepsilon$. For such y,

$$\begin{aligned} |h(y) - \varphi(\xi)| &\leq \int_{\partial^* T_{v,w}} |\varphi(\xi) - \varphi(\eta)|\, d\nu_y(\eta) + \int_{\partial^* T_{w,v}} |\varphi(\xi) - \varphi(\eta)|\, d\nu_y(\eta) \\ &\leq \varepsilon\, \nu_y(\partial^* T_{v,w}) + 2M\, \nu_y(\partial^* T_{w,v}) < (1 + 2M)\varepsilon. \end{aligned}$$

This concludes the proof. □

9.43 Theorem. *Consider a transient nearest neighbour random walk on the tree T.*

(a) *A point $\xi \in \partial^* T$ is regular for the Dirichlet problem if and only if the Green kernel vanishes at ξ, and in this case, $\xi \in \operatorname{supp}(\nu_o) \subset \partial^* T$.*

(b) *The regular points form a Borel set that has ν_x-measure* 1.

Proof. Consider the set

$$B = \big\{\xi \in \partial^* T : \lim_{y \to \xi} G(y, x) = 0\big\}.$$

Since $G(w, o) \leq G(y, o)$ for all $w \in T_y$, we can write

$$B = \bigcap_{n \in \mathbb{N}} \bigcup_{y : G(y,o) < 1/n} \partial^* T_y,$$

so that B is a Borel set. We know that it is independent of the choice of x. Our first claim is that $\nu_x(B) = 1$ for some and hence every $x \in T$. We take $x = o$. Consider the event

$$\overline{\Omega} = \left\{\omega \in \Omega : \begin{array}{c} Z_0(\omega) = o,\ Z_{n+1}(\omega) \sim Z_n(\omega) \text{ for all } n, \\ G\big(Z_n(\omega), o\big) \to 0,\ Z_n(\omega) \to Z_\infty(\omega) \in \partial T \end{array} \right\}.$$

Then $\mathsf{Pr}_o(\overline{\Omega}) = 1$, see Exercise 7.32. For $\xi \in \partial T$, let us write $v_k(\xi)$ for the point on $\pi(o,\xi)$ at distance k from o. For $\omega \in \overline{\Omega}$, the sequence $\big(Z_n(\omega)\big)$ visits every point on the ray from o to $Z_\infty(\omega)$. Recall the sequence of exit times $\boldsymbol{\epsilon}_k(\omega) = \max\big\{n : Z_n(\omega) = v_k\big(Z_\infty(\omega)\big)\big\}$. Then $\boldsymbol{\epsilon}_k(\omega) \to \infty$ and

$$G\big(v_k\big(Z_\infty(\omega)\big), o\big) = G\big(Z_{\boldsymbol{\epsilon}_k}(\omega), o\big) \to 0.$$

We obtain

$$\begin{aligned}\nu_o\big(\{\xi \in \partial T : G\big(v_k(\xi), o\big) \to 0\}\big)\\ = \mathsf{Pr}_o\big[Z_\infty \in \{\xi \in \partial T : G\big(v_k(\xi), o\big) \to 0\}\big] = 1.\end{aligned}$$

If $y \to \xi$ and $G\big(v_k(\xi), o\big) \to 0$ then $y \wedge \xi = v_k(\xi)$ with $k = k(y) \to \infty$, and

$$G(y,o) = F(y, y\wedge\xi)\, G\big(v_k(\xi), o\big) \le G\big(v_k(\xi), o\big) \to 0.$$

This shows that $\nu_o(B) = 1$.

We now prove statement (a), and then statement (b) follows from the above.

Suppose that the Green kernel vanishes at $\xi \in \partial^* T$. Let $\widehat{T}_{v,w}$ be a neighbourhood of ξ, where $v \sim w$. Its complement in $\widehat{T}$ is $\widehat{T}_{w,v}$. Let $y \in T_{v,w}$. Then

$$\nu_y(\widehat{T}_{w,v}) = F(y,w)\, \nu_w(\widehat{T}_{w,v}) \le G(y,w) \to 0, \quad \text{as } y \to \xi.$$

Therefore any finite intersection $U = \bigcap_{j=1}^m \widehat{T}_{v_j,w_j}$ of such neighbourhoods of ξ satisfies $\nu_y(U) \to 1$ as $y \to \xi$. Lemma 9.42 yields that ξ is regular, and we also conclude that $\nu_o(U) > 0$ for every basic neighbourhood of ξ, whence $\xi \in \mathrm{supp}(\nu_o)$.

Conversely, suppose that ξ is regular. If $\nu_o = \delta_\xi$, then ξ must be an end, and since $\nu_x(B) = 1$, the Green kernel vanishes at ξ. So suppose that $\mathrm{supp}(\nu_o)$ has at least two elements. One of them must be ξ. There must be a basic neighbourhood $\widehat{T}_{v,w}$ $(v \sim w)$ of ξ such that its complement contains some element of $\mathrm{supp}(\nu_o)$. Let $\varphi = \mathbf{1}_{\partial^* T_{w,v}}$ and h its Poisson integral. By assumption,

$$0 = \lim_{y\to\xi} h(y) = \lim_{y\to\xi} F(y,w)\, \underbrace{\nu_w(\partial^* T_{w,v})}_{>0}.$$

Therefore $G(y,w) = F(y,w)G(w,w) \to 0$ as $y \to \xi$. □

The proof of the following corollary is left as an exercise.

9.44 Corollary (Exercise). *The Dirichlet problem at infinity admits solution if and only if the Green kernel vanishes at infinity, that is, for every $\varepsilon > 0$ there is a finite subset $A_\varepsilon \subset T$ such that*

$$G(x,o) < \varepsilon \quad \textit{for all } \ x \in T \setminus A_\varepsilon.$$

In this case, $\mathrm{supp}(\nu_o) = \partial^* T$, *the full boundary.*

Let us next consider some examples.

9.45 Examples. (a) For simple random walk on the homogeneous tree of Example 9.29 and Figure 26, we have $F(x, y) = q^{-d(x,y)}$. We see that the Green kernel vanishes at infinity, and the Dirichlet problem is solvable.

(b) For simple random walk on the tree of Example 9.33 and Figure 27, all the "hairs" give rise to recurrent ends η_x, $x \in \mathbb{T}_q$. If $w \in \pi(x, \eta_x)$ then $F(w, x) = 1$, so that η_x is not regular for the Dirichlet problem. On the other hand, the Green kernel of simple random walk on T clearly vanishes at infinity on the subtree $\mathbb{T}_q$, and all ends in $\partial\mathbb{T}_q \subset \partial T$ are regular for the Dirichlet problem. The set $\partial\mathbb{T}_q$ is a closed subset of the boundary, and every continuous function φ on $\partial\mathbb{T}_q$ has a continuous extension to $\widehat{T}$ which is harmonic on T. For the extended function, the values at the ends η_x, $x \in \mathbb{T}_q$, are forced by φ, since the harmonic function is constant on each hair.

As $\partial\mathbb{T}_q$ is dense in ∂T, in the spirit of this section we should consider the support of the limit distribution to be $\mathrm{supp}(\nu_o) = \partial T$, although the random walk does not converge to one of the recurrent ends.

In this example, as well as in (a), the measure ν_o is *continuous:* $\nu_o(\xi) = 0$ for every $\xi \in \partial T$.

We also note that each non-regular (recurrent) end is itself an isolated point in ∂T.

We now consider an example with a non-regular point that is not isolated.

9.46 Example. We construct a tree starting with the half-line $\mathbb{N}_0$, whose end is $\varpi = +\infty$. At each point we attach a finite path of length $f(k)$ (a finite "hair"). At the end of each of those "hairs", we attach a copy of the binary tree by its root. (The binary tree is the tree where the root, as well as any other vertex x, has precisely two forward neighbours: the root has degree 2, while all other points have degree 3.) See Figure 28. We write T for the resulting tree.

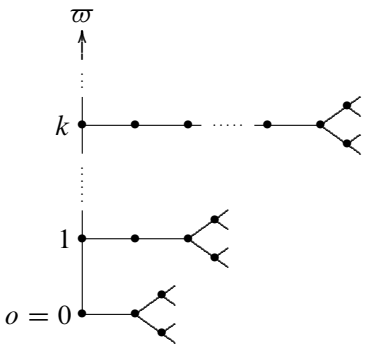

Figure 28

Our root vertex is $o = 0$ on the "backbone" $\mathbb{N}_0$. We consider once more simple random walk. If w is one of the vertices on one of the attached copies of the binary tree, then $F(w, w^-)$ is the same as on the homogeneous tree with degree 3. (Compare with Exercise 2.12 and the considerations following (9.2).) From Example 9.29, we know that $F(w, w^-) = 1/2$. Exercise 9.26 implies that $F(x, x^-) < 1$ for every $x \neq o$. All ends are transient, and $\mathrm{supp}(\nu_o) = \partial T$, the full boundary. The Green kernel vanishes at every end of each of the attached binary trees. That is, every end in $\partial T \setminus \{\varpi\}$ is regular for the Dirichlet problem. We now show that with a suitable choice of the lengths $f(k)$ of the "hairs", the end ϖ is not regular. We suppose that $f(k) \to \infty$ as $k \to \infty$. Then it is easy to understand that $F(k, k-1) \to 1$.

Indeed, consider first a tree $\widetilde{T}$ similar to ours, but with $f(k) = \infty$ for each k, that is, each of the "hairs" is an infinite ray (no binary tree). SRW on this tree is recurrent (why?) so that $\widetilde{F}(1, 0) = 1$ for the associated hitting probability. Then let $\widetilde{F}_n(1, 0)$ be the probability that SRW on $\widetilde{T}$ starting at 1 reaches 0 before leaving the ball $\widetilde{B}_n$ with radius n around 0 in the graph metric of $\widetilde{T}$. Then $\lim_{n\to\infty} \widetilde{F}_n(1, 0) = \widetilde{F}(1, 0) = 1$ by monotone convergence. On the other hand, in our tree T, consider the cone $T_k = T_{0,k}$. If $n(k) = \inf\{f(m) : m \geq k\}$ then the ball of radius $n(k)$ centred at the vertex k in T_k is isomorphic with the ball $\widetilde{B}_{n(k)}$ in $\widetilde{T}$. Therefore $F(k, k-1) \geq \widetilde{F}_{n(k)}(1, 0) \to 1$ as $k \to \infty$.

Next, we write k' for the neighbour of the vertex $k \in \mathbb{N}$ that lies on the "hair" attached at k. We have $F(k', k) \geq \big(f(k) - 1\big)\big/ f(k)$, since the latter quotient is the probability that SRW on T starting at k' reaches k before the root of the binary tree attached at the k-th "hair". (To see this, we only need to consider the drunkard's walk of Example 1.46 on $\{0, 1, \ldots, f(k)\}$ with $p = q = 1/2$.)

Now Proposition 9.3 (b) yields for SRW on T

$$\begin{aligned} 1 - F(k, k-1) &= 1 - \frac{1}{3 - F(k+1, k) - F(k', k)} \\ &= \frac{\big(1 - F(k+1, k)\big) + \big(1 - F(k', k)\big)}{1 + \big(1 - F(k+1, k)\big) + \big(1 - F(k', k)\big)} \\ &\leq \big(1 - F(k+1, k)\big) + \big(1 - F(k', k)\big). \end{aligned}$$

Recursively, we deduce that for each $r \geq k$,

$$1 - F(k, k-1) \leq \big(1 - F(r+1, r)\big) + \sum_{m=k}^{r} \big(1 - F(\bar{m}, m)\big).$$

We let $r \to \infty$ and get

$$1 - F(k, k-1) \leq \sum_{m=k}^{\infty} \big(1 - F(\bar{m}, m)\big).$$

Therefore

$$\sum_{k=1}^{\infty}\big(1-F(k,k-1)\big)\le\sum_{k=1}^{\infty}k\,\big(1-F(k',k)\big)\le\sum_{k=1}^{\infty}\frac{k}{f(k)}.$$

If we choose $f(k)>k$ such that the last series converges (for example, $f(k)=k^3$), then

$$F(n,0)=\prod_{k=1}^{n}F(k,k-1)\to\prod_{k=1}^{\infty}F(k,k-1)>0$$

(because for a sequence of numbers $a_k\in(0,\,1)$, the infinite product $\prod_k a_k$ is >0 if and only if $\sum_k(1-a_k)<\infty$), and the end $\varpi=+\infty$ is non-regular.

Next, we give an example of a non-simple random walk, also involving vertices with infinite degree.

9.47 Example. Let again $T=\mathbb{T}_{s-1}$ be the homogeneous tree with degree $s\ge 3$, but this time we also allow that $s=\infty$. We colour the non-oriented edges of T by the numbers ("colours") in the set $\mathcal{I}$, where $\mathcal{I}=\{1,\dots,s\}$ when $s<\infty$, and $\mathcal{I}=\mathbb{N}$ when $s=\infty$. This coloring is such that every vertex is incident with precisely one edge of each colour. We choose probabilities $p_i>0$ ($i\in\mathcal{I}$) such that $\sum_{i\in\mathcal{I}}p_i=1$ and define the following symmetric nearest neighbour random walk.

$$p(x,y)=p(y,x)=p_i,\quad\text{if the edge }[x,y]\text{ has colour }i.$$

See Figure 29, where $s=3$.

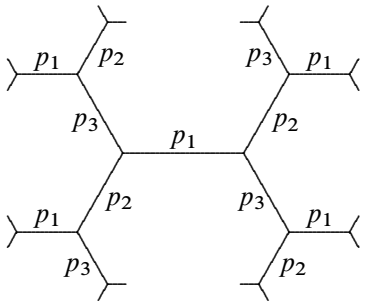

Figure 29

We remark here that this is a random walk on the group with the presentation

$$\mathfrak{G}=\langle a_i,\,i\in\mathcal{I}\mid a_i^2=o\ \text{ for all }i\in\mathcal{I}\rangle,$$

where (recall) we denote by o the unit element of $\mathfrak{G}$. For readers who are not familiar with group presentations, it is easy to describe $\mathfrak{G}$ directly. It consists of all *words*

$$x=a_{i_1}a_{i_2}\cdots a_{i_n},\quad\text{where }n\ge 0,\ i_j\in\mathcal{I},\ i_{j+1}\ne i_j\text{ for all }j.$$

The number n is the length of the word, $|x| = n$. When $n = 0$, we obtain the *empty word* o. The group operation is concatenation of words followed by cancellations. Namely, if the last letter of x coincides with the first letter of y, then both are cancelled in the product xy because of the relation $a_i^2 = o$. If in the remaining word (after concatenation and cancellation) one still has an a_j^2 in the middle, this also has to be cancelled, and so on until one gets a square-free word. The latter is the product of x and y in $\mathfrak{G}$. In particular, for x as above, the inverse is $x^{-1} = a_{i_n} \cdots a_{i_2} a_{i_1}$. In the terminology of combinatorial group theory, $\mathfrak{G}$ is the free product over all $i \in \mathcal{I}$ of the 2-element groups $\{o, a_i\}$ with $a_i^2 = o$.

The tree is the *Cayley graph* of $\mathfrak{G}$ with respect to the symmetric set of generators $S = \{a_i : i \in \mathcal{I}\}$: the set of vertices of the graph is $\mathfrak{G}$, and two elements x, y are neighbours in the graph if $y = xa_i$ for some $a_i \in S$. Our random walk is a random walk on that group. Its law μ is supported by the set of generators, and $\mu(a_i) = p_i$ for each $i \in \mathcal{I}$.

We now want to compute the Green function and the functions $F(x, y|z)$. We observe that $F(y, x|z) = F_i(z)$ is the same for all edges $[x, y]$ with colour i. Furthermore, the function $G(x, x|z) = G(z)$ is the same for all $x \in T$. We know from Theorem 1.38 and Proposition 9.3 that

$$\begin{aligned} G(z) &= \frac{1}{1 - \sum_i p_i z\, F_i(z)} \quad \text{and} \\ F_i(z) &= \frac{p_i z}{1 - \sum_{j \neq i} p_j z\, F_j(z)} = \frac{p_i z}{\dfrac{1}{G(z)} + p_i z\, F_i(z)}. \end{aligned} \tag{9.48}$$

We obtain a second order equation for the function $F_i(z)$, whose two solutions are $\Big(1 \pm \sqrt{1 + 4p_i^2 z^2 G(z)^2} - 1\Big) \big/ \big(2 p_i z\, G(z)\big)$. Since $F_i(0) = 0$, while $G(0) = 1$, the right solution is

$$F_i(z) = \frac{\sqrt{1 + 4p_i^2 z^2 G(z)^2} - 1}{2 p_i z\, G(z)}. \tag{9.49}$$

Combining (9.48) and (9.49), we obtain an implicit equation for the function $G(z)$:

$$G(z) = \Phi\big(zG(z)\big), \quad \text{where } \Phi(t) = 1 + \frac{1}{2} \sum_{i \in \mathcal{I}} \Big(\sqrt{1 + 4p_i^2 t^2} - 1 \Big). \tag{9.50}$$

We can analyze this formula by use of some basic facts from the theory of complex functions. The function $G(z)$ is defined by a power series and is analytic in the disk of convergence around the origin. (It can be extended analytically beyond that disk.) Furthermore, the coefficients are non-negative, and by Pringsheim's theorem (which we have used already in Example 5.24), its radius of convergence must be

a singularity. Thus, $\mathsf{r} = \mathsf{r}(P)$ is the smallest positive singularity of $G(z)$. We are lead to studying the equation (9.50) for $z,\ t \geq 0$.

For $t \geq 0$, the function $t \mapsto \Phi(t)$ is monotone increasing, convex, with $\Phi(0) = 1$ and $\Phi'(0) = 0$. For $t \to \infty$, it approaches the asymptote with equation $y = t - \frac{s-2}{2}$ in the (t, y)-plane. For $0 < z < \mathsf{r}(P)$, by (9.50), $G(z)$ is the y-coordinate of a point where the curve $y = \Phi(t)$ intersects the line $y = \frac{1}{z}t$, see Figures 30 and 31.

Case 1. $s = 2$. The asymptote is $y = t$, there is a unique intersection point of $y = \Phi(t)$ and $y = \frac{1}{z}t$ for each fixed $z \in (0, 1)$, and the angle of intersection is non-zero.

By the implicit function theorem, there is a unique analytic solution of the equation (9.50) for $G(z)$ in some neighbourhood of z. This gives $G(z)$, which is analytic in each real $z \in (0, 1)$. On the other hand, if $z > 1$, the curve $y = \Phi(t)$ and the line $y = \frac{1}{z}t$ do not intersect in any point with real coordinates: there is no real solution of (9.50). It follows that $\mathsf{r} = 1$. Furthermore, we see from Figure 30 that $G(z) \to \infty$ when $z \to 1$ from the left. Therefore the random walk is recurrent. Our random walk is symmetric, whence reversible with respect to the counting measure, which has infinite mass. We conclude that recurrence is null recurrence.

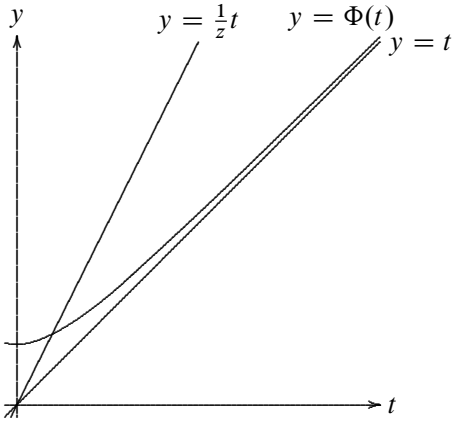

Figure 30

Case 2. $3 \leq s \leq \infty$. The asymptote intersects the y-axis at $-\frac{s-2}{2} < 0$. By the convexity of Φ, there is a unique tangent line to the curve $y = \Phi(t)$ $(t \geq 0)$ that emanates from the origin, as one can see from Figure 31.

Let ρ be the slope of that tangent line. Its slope is smaller than that of the asymptote, that is, $\rho < 1$. For $0 < z \leq 1$, the line $y = \frac{1}{z}t$ intersects the curve $y = \Phi(t)$ in a unique point with positive real coordinates, and the same argument

as in Case 1 shows that $G(\cdot)$ is analytic at z. For $1 < z < 1/\rho$, there are two intersection points with non-zero angles of intersection. Therefore both give rise to an analytic solution of the equation (9.50) for $G(z)$ in a complex neighbourhood of z. Continuity of $G(z)$ and analytic continuation imply that $G(z) = \sum_n p^{(n)}(x,x)\, z^n$ is finite and coincides with the y-coordinate of the intersection point that is more to the left.

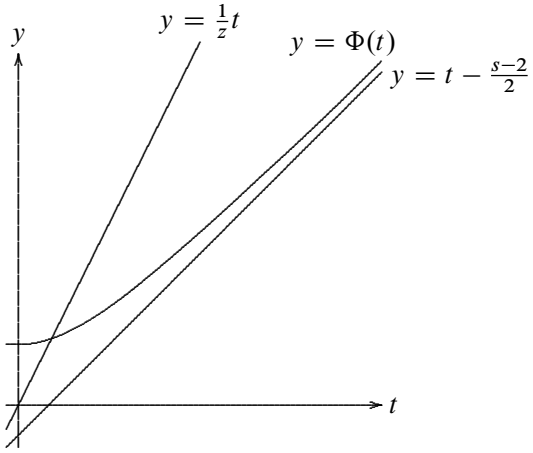

Figure 31

On the other hand, for $z > 1/\rho$ there is no real solution of (9.50). We conclude that the radius of convergence of $G(z)$ is $\mathsf{r} = 1/\rho$. We also find a formula for $\rho = \rho(P)$. Namely, $\rho = \Phi(t_0)$, where t_0 is the unique positive solution of the equation $\Phi'(t) = \Phi(t)/t$, which can be written as

$$\frac{1}{2}\sum_{i\in\mathcal{I}}\left(1 - \frac{1}{\sqrt{1+4p_i^2t^2}}\right) = 1.$$

We also have $\rho(P) = \min\{\Phi(t)/t : t > 0\}$.

In particular, $G(1)$ is finite, and the random walk is transient. (This can of course be shown in various different ways, including the flow criterion.) Also, $G(\mathsf{r}) = \lim_{z\to\mathsf{r}-} G(z) = \Phi(t_0)$ is finite, so that the random walk is also ρ-transient.

Let us now consider the Dirichlet problem at infinity in Case 2. By equation (9.49), we clearly have $F_i = F_i(1) < 1$ for each i. If $s = \infty$, then we also observe that by (9.48)

$$F_i \le p_i G(1) \to 0, \quad \text{as } i \to \infty.$$

Thus we always have $\bar{F} = \max_{i\in\mathcal{I}} F_i < 1$. For arbitrary y,

$$G(y,o) = \bar{F}(y,o)G(o,o) \le \bar{F}^{|y|}\, G(o,o).$$

Thus,

$$\lim_{y\to\xi} G(y,o) = 0 \quad \text{for every end } \xi \in \partial T.$$

When $3 \le s < \infty$, we see the Dirichlet problem admits a solution for every continuous function on ∂T.

Now consider the case when $s = \infty$. Every vertex has infinite degree, and T^* is in one-to-one correspondence with T. We see from the above that the Green kernel vanishes at every end. We still have to show that it vanishes at every improper vertex; compare with the remarks after Definition 9.41. By the spatial homogeneity of our random walk (i.e., because it is a random walk on a group), it is sufficient to show that the Green kernel vanishes at o^*. The neighbours of o are the points a_i, $i \in I$, where the colour of the edge $[o, a_i]$ is i. We have $G(a_i, o) = F_i\, G(o, o)$, which tends to 0 when $i \to \infty$. Now $y \to o^*$ means that $i(y) \to \infty$, where $i = i(y)$ is such that $y \in T_{o,a_i}$.

We have shown that the Green kernel vanishes at every boundary point, so that the Dirichlet problem is solvable also when $s = \infty$.

At this point we can make some comments on why it has been preferable to introduce the improper vertices. This is because we want for a general irreducible, transient Markov chain (X, P) that the state space X remains *discrete* in the Martin compactification. Recall what has been said in Section 7.B (before Theorem 7.19): in the original article of Doob [17] it is not required that X be discrete in the Martin compactification. In that setting, the compactification is the closure of (the embedding of) X in $\mathcal{B}$, the base of the cone of all positive superharmonic functions. In the situation of the last example, this means that the compactification is $T \cup \partial T$, but a sequence that converges to an improper vertex x^* in our setting will then converge to the "original" vertex x.

Now, for this smaller compactification of the tree with $s = \infty$ in the last example, the Dirichlet problem cannot admit solution. Indeed, there every vertex is the limit of a sequence of ends, so that a continuous function on ∂T forces the values of the continuous extension (if it exists) at each vertex before we can even start to consider the Poisson integral. For example, if $x \sim y$ then the continuous extension of the function $\mathbf{1}_{\partial T_{x,y}}$ to $T \cup \partial T$ in the "wrong" compactification is $\mathbf{1}_{T_{x,y} \cup \partial T_{x,y}}$. It is not harmonic at x and at y.

We see that for the Dirichlet problem it is really relevant to have a compactification in which the state space remains discrete.

We remark here that Dirichlet regularity for points of the Martin boundary of a Markov chain was first studied by Knapp [37]. Theorem 9.43 (a) and Corollary 9.44 are due to Benjamini and Peres [5] and, for trees that are not necessarily locally finite, Cartwright, Soardi and Woess [10]. Example 9.46, with a more general and more complicated proof, is due to Amghibech [1]. The paper [5] also contains an example with an uncountable set of non-regular points that is dense in the boundary.

A radial Fatou theorem

We conclude this section with some considerations on the Fatou theorem. Its classical versions are non-probabilistic and concern convergence of the Poisson integral $h(x)$ of an integrable function φ on the boundary. Here, φ is not assumed to be continuous, and the Dirichlet problem at infinity may not admit solution. We look for a more restricted variant of convergence of h when approaching the boundary. Typically, $x \to \xi$ (a boundary point) in a specific way ("non-tangentially", "radially", etc.), and we want to know whether $h(x) \to \varphi(\xi)$. Since h does not change when φ is modified on a set of ν_o-measure 0, such a result can in general only hold ν_o-almost surely. This is of course similar to (but not identical with) the probabilistic Fatou theorem 7.67, which states convergence along almost every trajectory of the random walk. We prove the following Fatou theorem on radial convergence.

9.51 Theorem. *Consider a transient nearest neighbour random walk on the countable tree T. Let φ be a ν_o-integrable function on the space of ends ∂T, and $h(x)$ its Poisson integral.*

Then for ν_o-almost every $\xi \in \partial T$,

$$\lim_{k\to\infty} h\big(v_k(\xi)\big) = \varphi(\xi),$$

where $v_k(\xi)$ is the vertex on $\pi(o,\xi)$ at distance k from o.

Proof. Similarly to the proof of Theorem 9.43, we define the event

$$\Omega' = \left\{\omega \in \Omega : \begin{array}{c} Z_0(\omega) = o,\ Z_{n+1}(\omega) \sim Z_n(\omega) \text{ for all } n, \\ Z_n(\omega) \to Z_\infty(\omega) \in \partial T,\ h\big(Z_n(\omega)\big) \to \varphi\big(Z_\infty(\omega)\big) \end{array}\right\}.$$

Then $\mathsf{Pr}_o(\Omega') = 1$. Precisely as in the proof of Theorem 9.43, we consider the exit times $\boldsymbol{\epsilon}_k = \boldsymbol{\epsilon}_k(\omega) = \max\big\{n : Z_n(\omega) = v_k\big(Z_\infty(\omega)\big)\big\}$ for $\omega \in \Omega'$. Then $\boldsymbol{\epsilon}_k \to \infty$, whence

$$h\big(v_k(Z_\infty)\big) = h(Z_{\boldsymbol{\epsilon}_k}) \to \varphi(Z_\infty) \quad \text{on } \Omega'.$$

Therefore, setting

$$B = \{\xi \in \partial T : h\big(v_k(\xi)\big) \to \varphi(\xi)\},$$

we have $\nu_o(B) = \mathsf{Pr}_o[Z_\infty \in B] = 1$, as proposed. □

9.52 Exercise. Verify that the set B defined at the end of the last proof is a Borel subset of ∂T. □

Note that the improper vertices make no appearance in Theorem 9.51, since $\nu_o(T^*) = 0$.

Theorem 9.51 is due to Cartier [Ca]. Much more work has been done regarding Fatou theorems on trees, see e.g. the monograph by Di Biase [16].

F The boundary process, and the deviation from the limit geodesic

We know that a transient nearest neighbour random walk on a tree T converges almost surely to a random end. We next want to study how and when the initial pieces with lengths $k \in \mathbb{N}_0$ of $\pi(Z_0, Z_n)$ stabilize. Recall the exit times $\boldsymbol{\epsilon}_k$ that were introduced in the proof of Theorem 9.18: $\boldsymbol{\epsilon}_k = n$ means that n is the last instant when $d(Z_0, Z_n) = k$. In this case, $Z_n = v_k(Z_\infty)$, where (as above) for an end ξ, we write $v_k(\xi)$ for the k-th point on $\pi(o, \xi)$.

9.53 Definition. For a transient nearest neighbour random walk on a tree T, the *boundary process* is $W_k = v_k(Z_\infty) = Z_{\boldsymbol{\epsilon}_k}$, and the *extended boundary process* is $\big(W_k, \boldsymbol{\epsilon}_k\big)$, $k \geq 0$.

When studying the boundary process, we shall always assume that $Z_0 = o$. Then $|W_k| = k$, and for $x \in T$ with $|x| = k$, we have $\mathsf{Pr}_o[W_k = x] = \nu_o(\partial T_x)$.

9.54 Exercise. Setting $z = 1$, deduce the following from Proposition 9.3 (b). For $x \in T \setminus \{o\}$,

$$\mathsf{Pr}_x[Z_n \in T_x \setminus \{x\} \text{ for all } n \geq 1] = p(x, x^-)\frac{1 - F(x, x^-)}{F(x, x^-)}.$$

[Hint: decompose the probability on the left hand side into terms that correspond to first moving from x to some forward neighbour y and then never going back to x.] □

9.55 Proposition. *The extended boundary process* $\big(W_k, \boldsymbol{\epsilon}_k\big)_{k \geq 1}$ *is a (non-irreducible) Markov chain with state space* $T_{\mathrm{tr}} \times \mathbb{N}_0$. *Its transition probabilities are*

$$\begin{aligned}\mathsf{Pr}_o[W_k = y, \boldsymbol{\epsilon}_k = n \mid W_{k-1} = x,\ \boldsymbol{\epsilon}_{k-1} = m] \\ = \frac{F(x, x^-)}{F(y, x)} \frac{1 - F(y, x)}{1 - F(x, x^-)} \frac{p(x, y)}{p(x, x^-)} f^{(n-m)}(y, x),\end{aligned}$$

where $y \in T_{\mathrm{tr}}$ *with* $|y| = k$, $x = y^-$, *and* $n - m \in \mathbb{N}$ *is odd.*

Proof. For the purpose of this proof, the quantity computed in Exercise 9.54 is denoted

$$g(x) = \mathsf{Pr}_x[Z_i \in T_x \setminus \{x\} \text{ for all } i \geq 1].$$

Let $[o = x_0, x_1, \ldots, x_k]$ be any geodesic arc in T_{tr} that starts at o, and let $m_1 < m_2 < \cdots < m_k$ be positive integers such that $m_j - m_{j-1} \in \mathbb{N}_{\mathrm{odd}}$ for all j. (Of course, $\mathbb{N}_{\mathrm{odd}}$ denotes the odd positive integers.) We consider the following events

in the trajectory space.

$$
\begin{aligned}
A_k &= [W_k = x_k,\ \boldsymbol{\epsilon}_k = m_k],\\
B_k &= [W_{k-1} = x_{k-1}, \boldsymbol{\epsilon}_{k-1} = m_{k-1},\ W_k = x_k, \boldsymbol{\epsilon}_k = m_k] = A_{k-1} \cap A_k,\\
C_k &= [W_1 = x_1, \boldsymbol{\epsilon}_1 = m_1, \dots, W_k = x_k, \boldsymbol{\epsilon}_k = m_k], \quad \text{and}\\
D_k &= [\,|Z_n| > k \ \text{ for all } n > m_k].
\end{aligned}
$$

We have to show that $\mathsf{Pr}_o[A_k \mid C_{k-1}] = \mathsf{Pr}_o[A_k \mid A_{k-1}]$, and we want to compute this number. A difficulty arises because the two conditioning events depend on all future times after $\boldsymbol{\epsilon}_{k-1} = m_{k-1}$. Therefore we also consider the events

$$
\begin{aligned}
A_k^* &= [Z_{m_k} = x_k],\\
B_k^* &= [Z_{m_{k-1}} = x_{k-1},\ Z_{m_k} = x_k,\ |Z_i| \geq k \ \text{ for } i = m_{k-1}+1, \dots, m_k],
\end{aligned}
$$

and

$$
C_k^* = \left[\begin{array}{c} Z_{m_j} = x_j \ \text{ for } j = 1, \dots, k, \\ |Z_i| \geq j \ \text{ for } i = m_{j-1}+1, \dots, m_j,\ j = 2, \dots, k \end{array} \right].
$$

Each of them depends only on $Z_0, \dots, Z_{m_k}$. Now we can apply the Markov property as follows.

$$
\mathsf{Pr}_o(A_k) = \mathsf{Pr}_o(D_k \cap A_k^*) = \mathsf{Pr}_o(D_k \mid A_k^*)\ \mathsf{Pr}_o(A_k^*) = g(x_k)\ \mathsf{Pr}_o(A_k^*),
$$

and analogously

$$
\mathsf{Pr}_o(B_k) = g(x_k)\ \mathsf{Pr}_o(B_k^*) \quad \text{and} \quad \mathsf{Pr}_o(C_k) = g(x_k)\ \mathsf{Pr}_o(C_k^*).
$$

Thus, noting that $C_k^* = B_k^* \cap C_{k-1}^*$,

$$
\begin{aligned}
\mathsf{Pr}_o(A_k \mid C_{k-1}) &= \frac{\mathsf{Pr}_o(C_k)}{\mathsf{Pr}_o(C_{k-1})}\\
&= \frac{g(x_k)}{g(x_{k-1})}\, \frac{\mathsf{Pr}_o(C_k^*)}{\mathsf{Pr}_o(C_{k-1}^*)}\\
&= \frac{g(x_k)}{g(x_{k-1})}\ \mathsf{Pr}_o(B_k^* \mid C_{k-1}^*)\\
&= \frac{g(x_k)}{g(x_{k-1})}\ \mathsf{Pr}_o(B_k^* \mid A_{k-1}^*)\\
&= \frac{g(x_k)}{g(x_{k-1})}\, \frac{\mathsf{Pr}_o(B_k^*)}{\mathsf{Pr}_o(A_{k-1}^*)}\\
&= \frac{\mathsf{Pr}_o(B_k)}{\mathsf{Pr}_o(A_{k-1})}\\
&= \mathsf{Pr}_o(A_k \mid A_{k-1}),
\end{aligned}
$$

as required. Now that we know that the process is Markovian, we assume that $x_{k-1} = x$, $x_k = y$, $m_{k-1} = m$ and $m_k = n$, and easily compute

$$\begin{aligned}\mathsf{Pr}_o(B_k^*) &= \mathsf{Pr}_o(A_{k-1}^*)\ \mathsf{Pr}_x[Z_{n-m} = y,\ Z_i \neq x \ \text{ for } i = 1, \dots, n-m] \\ &= \mathsf{Pr}_o(A_{k-1}^*)\, \ell^{(n-m)}(x, y),\end{aligned}$$

where $\ell^{(n-m)}(x, y)$ is the "last exit" probability of (3.56). By reversibility (9.5) and Exercise 3.59, we have for the generating function $L(x, y|z)$ of the $\ell^{(n)}(x, y)$ that

$$\mathsf{m}(x)\, L(x, y|z) = \frac{\mathsf{m}(x)\, G(x, y|z)}{G(x, x|z)} = \frac{\mathsf{m}(y)\, G(y, x|z)}{G(x, x|z)} = \mathsf{m}(y)\, F(y, x|z).$$

Therefore, $\ell^{(n-m)}(x, y) = \mathsf{m}(y)\, f^{(n-m)}(y, x)/\mathsf{m}(x)$, and of course we also have $\mathsf{m}(y)\, p(y, x)/\mathsf{m}(x) = p(x, y)$. Putting things together, the transition probability from (x, m) to (y, n) is

$$\mathsf{Pr}_o(A_k \mid A_{k-1}) = \frac{g(y)}{g(x)}\, \ell^{(n-m)}(x, y),$$

which reduces to the stated formula. □

Note that the transition probabilities in Proposition 9.55 depend only on x, y and the increment $\boldsymbol{\delta}_{k+1} = \boldsymbol{\epsilon}_{k+1} - \boldsymbol{\epsilon}_k$. Thus, also the process $\big(W_k, \boldsymbol{\delta}_k\big)_{k\geq 1}$ is a Markov chain, whose transition probabilities are

$$\begin{aligned}&\mathsf{Pr}_o[W_{k+1} = y,\ \boldsymbol{\delta}_{k+1} = n \mid W_k = x,\ \boldsymbol{\delta}_k = m] \\ &\qquad = \frac{F(x, x^-)}{F(y, x)}\, \frac{1 - F(y, x)}{1 - F(x, x^-)}\, \frac{p(x, y)}{p(x, x^-)}\, f^{(n)}(y, x).\end{aligned} \tag{9.56}$$

Here, n has to be odd, so that the state space is $T_{\mathrm{tr}} \times \mathbb{N}_{\mathrm{odd}}$. Also, the process factorizes with respect to the projection $(x, n) \mapsto x$.

9.57 Corollary. *The boundary process $(W_k)_{k\geq 1}$ is also a Markov chain. If $x, y \in T_{\mathrm{tr}}$ with $|x| = k$ and $y^- = x$ then*

$$\mathsf{Pr}_o[W_{k+1} = y \mid W_k = x] = \frac{\nu_o(\partial T_y)}{\nu_o(\partial T_x)} = F(x, x^-)\, \frac{1 - F(y, x)}{1 - F(x, x^-)}\, \frac{p(x, y)}{p(x, x^-)}.$$

We shall use these observations in the next sections.

9.58 Exercise. Compute the transition probabilities of the extended boundary process in Example 9.29.

[Hint: note that the probabilities $f^{(n)}(x, x^-)$ can be obtained explicitly from the closed formula for $F(x, x^-|z)$, which we know from previous computations.] □

Now that we have some idea how the boundary process evolves, we want to know how far (Z_n) can deviate from the limit geodesic $\pi(o, Z_\infty)$ in between the exit times. That is, we want to understand how $d\big(Z_n, \pi(o, Z_\infty)\big)$ behaves, where (Z_n) is a transient nearest neighbour random walk on an infinite tree T. Here, we just consider one basic result of this type that involves the limit distributions on ∂T.

9.59 Theorem. *Suppose that there is a decreasing function $\phi\colon \mathbb{R}^+ \to \mathbb{R}^+$ with $\lim_{t\to\infty} \phi(t) = 0$ such that*

$$\nu_x(\partial T_{x,y}) \le \phi\big(d(x,y)\big) \quad \textit{for all } x, y \in T \ (x \ne y).$$

Then, whenever (r_n) is an increasing sequence in $\mathbb{N}$ such that $\sum_n \phi(r_n) < \infty$, one has

$$\limsup_{n\to\infty} \frac{d\big(Z_n, \pi(x, Z_\infty)\big)}{r_n} \le 1 \quad \mathsf{Pr}_x\textit{-almost surely}$$

for every starting point $x \in T$.

In particular, if

$$\sup\big\{F(x,y) : x, y \in T,\ x \sim y\big\} = \lambda < 1$$

then

$$\limsup_{n\to\infty} \frac{d\big(Z_n, \pi(x, Z_\infty)\big)}{\log n} \le \log(1/\lambda) \quad \mathsf{Pr}_x\textit{-almost surely for every } x.$$

Proof. The assumptions imply that each ν_x is a continuous measure, that is, $\nu_x(\xi) = 0$ for every end ξ of T. Therefore ∂T must be uncountable. We choose and fix an end $\xi \in \partial T$, and assume without loss of generality that the starting point is $Z_0 = o$.

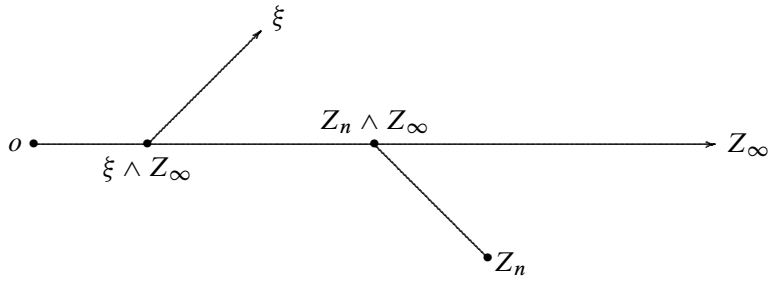

Figure 32

From Figure 32 one sees the following: if $|Z_n \wedge Z_\infty| > |\xi \wedge Z_\infty|$ then one has $d\big(Z_n, \pi(o, Z_\infty)\big) = d\big(Z_n, \pi(\xi, Z_\infty)\big)$. (Note that $\pi(\xi, Z_\infty)$ is a bi-infinite

geodesic.) Therefore, if $r > 0$, then

$$\begin{aligned}
&\mathsf{Pr}_o\big[d\big(Z_n,\, \pi(o, Z_\infty)\big) \geq r,\ |Z_n \wedge Z_\infty| > |\xi \wedge Z_\infty|\big] \\
&\quad \leq \mathsf{Pr}_o\big[d\big(Z_n, \pi(\xi, Z_\infty)\big) \geq r\big] \\
&\quad = \sum_{x \in T} \mathsf{Pr}_o[Z_n = x]\ \mathsf{Pr}_x\big[d\big(x, \pi(\xi, Z_\infty)\big) \geq r\big] \\
&\quad \leq \sum_{x \in T} \mathsf{Pr}_o[Z_n = x]\, \phi(r) = \phi(r),
\end{aligned}$$

since $d\big(x, \pi(\xi, Z_\infty)\big) \geq r$ implies that $Z_\infty \in \partial T_{x,y}$, where y is the element on the ray $\pi(x, \xi)$ at distance r from x. Now consider the sequence of events

$$A_n = \big[d\big(Z_n, \pi(o, Z_\infty)\big) \geq r_n,\ |Z_n \wedge Z_\infty| > |\xi \wedge Z_\infty|\,\big]$$

in the trajectory space. Then by the above, $\sum_n \mathsf{Pr}_o(A_n) < \infty$ and, by the Borel–Cantelli lemma, $\mathsf{Pr}_o(\limsup_n A_n) = 0$. We know that $\mathsf{Pr}_o[Z_\infty \neq \xi] = 1$, since ν_o is continuous. That is, $|\xi \wedge Z_\infty| < \infty$ almost surely. On the other hand, $|Z_n \wedge Z_\infty| \to \infty$. We see that

$$\mathsf{Pr}_o\big(\liminf_n \big[\,|Z_n \wedge Z_\infty| > |\xi \wedge Z_\infty|\,\big]\big) \geq \mathsf{Pr}_o[Z_\infty \neq \xi] = 1.$$

Therefore $\mathsf{Pr}_o\big(\limsup_n \big[d\big(Z_n, \pi(o, Z_\infty)\big) \geq r_n\big]\big) = 0$, which implies the proposed general result.

In the specific case when $\lambda = \sup\big\{F(x, y) : x, y \in T,\ x \sim y\big\} < 1$, we can set $\phi(t) = \lambda^t$. If we choose $r_n = \big\lceil (1+\alpha) \log n \big/ \log(1/\lambda) \big\rceil$, where $\alpha > 0$, then we see that $\phi(r_n) \leq 1/n^{1+\alpha}$, whence

$$\limsup_{n \to \infty} \frac{d\big(Z_n, \pi(x, Z_\infty)\big)}{\log n} \leq \frac{\log(1/\lambda)}{1+\alpha} \qquad \mathsf{Pr}_x\text{-almost surely,}$$

and this holds for every $\alpha > 0$. □

The $(\log n)$-estimate in the second part of the theorem was first proved for random walks on free groups by Ledrappier [40]. A simplified generalization is presented in [44], but it contains a trivial error at the end (the exponential function does not vary regularly at infinity), which was observed by Gilch [26].

The boundary process was used by Lalley [39] in the context of finite range random walks on free groups.

G Some recurrence/transience criteria

So far, we have always assumed transience. We now want to present a few (of the many) criteria for transience and recurrence of a nearest neighbour random walk with stochastic transition matrix P on a tree T.

In view of reversibility (9.5), we already have a criterion for positive recurrence, see Proposition 9.8. We can use the flow criterion of Theorem 4.51 to study transience. If $x \neq o$ and $\pi(o,x) = [o = x_0, x_1, \dots, x_{k-1}, x_k = x]$ then the resistance of the edge $e = [x^-, x] = [x_{k-1}, x_k]$ is

$$r(e) = \frac{1}{p(x_0,x_1)} \prod_{i=1}^{k-1} \frac{p(x_i,x_{i-1})}{p(x_i,x_{i+1})}.$$

There are various simple choices of unit flows from o to ∞ which we can use to test for transience.

9.60 Example. The *simple flow* on a locally finite tree T with root o and $\deg(x) \geq 2$ for all $x \neq o$ is defined recursively as

$$\phi(o,x) = \frac{1}{\deg(x)}, \text{ if } x^- = o, \quad \text{and} \quad \phi(x,y) = \frac{\phi(x^-,x)}{\deg(x)-1}, \text{ if } y^- = x \neq o.$$

Of course $\phi(x,x^-) = -\phi(x^-,x)$. With respect to simple random walk, its power is just

$$\langle \phi, \phi \rangle = \sum_{x \in T \setminus \{o\}} \phi(x^-,x)^2.$$

This sum can be nicely interpreted as the total area of a square tiling that is filled into the strip $[0,\ 1] \times [0,\ \infty)$ in the plane. Above the base segment we draw $\deg(o)$ squares with side length $1/\deg(o)$. Each of them corresponds to one of the edges $[o,x]$ with $x^- = o$. If we already have drawn the square corresponding to an edge $[x^-,x]$, then we subdivide its top side into $\deg(x) - 1$ segments of equal length. Each of them becomes the base segment of a new square that corresponds to one of the edges $[x,y]$ with $y^- = x$. See Figure 33. If the total area of the tiling is finite then the random walk is transient.

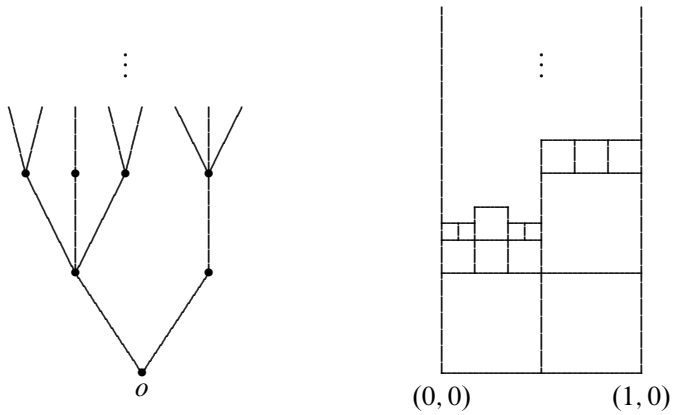

Figure 33. A tree and the associated square tiling.

This square tiling associated with SRW on a locally finite tree was first considered by GERL [25]. More generally, square tilings associated with random walks on planar graphs appear in the work of BENJAMINI and SCHRAMM [7].

Another choice is to take an end ξ and send a unit flow ϕ_ξ from o to ξ: if $\pi(o,\xi) = [o = x_0, x_1, x_2, \dots]$ then this flow is given by $\phi(e) = 1$ and $\phi(\check{e}) = -1$ for the oriented edge e from x_{i-1} to x_i, while $\phi(e) = 0$ for all edges that do not lie on the ray $\pi(o,\xi)$. If ϕ has finite power then the random walk is transient. We obtain the following criterion.

9.61 Corollary. *If T has an end ξ such that for the geodesic ray $\pi(o,\xi) = [o = x_0, x_1, x_2, \dots]$,*

$$\sum_{k=1}^{\infty} \prod_{i=1}^{k} \frac{p(x_i, x_{i-1})}{p(x_i, x_{i+1})} < \infty,$$

then the random walk is transient.

9.62 Exercise. Is it true that any end ξ that satisfies the criterion of Corollary 9.61 is a transient end? □

When T itself is a half-line (ray), then the condition of Corollary 9.61 is the necessary and sufficient criterion of Theorem 5.9 for birth-and-death chains. In general, the criterion is far from being necessary, as shows for example simple random walk on $\mathbb{T}_q$.

In any case, the criterion says that the ingoing transition probabilities (towards the root) are strongly dominated by the outgoing ones. We want to formulate a more general result in the same spirit. Let f be any strictly positive function on T. We define $\lambda = \lambda^f : T \setminus \{o\} \to (0,\infty)$ and $g = g^f : T \to [0,\infty)$ by

$$\begin{aligned} &\lambda(x) = \frac{1}{p(x,x^-)f(x)} \sum_{y\,:\,y^-=x} p(x,y) f(y), \\ &g(o) = 0, \text{ and if } x \neq o,\ \pi(o,x) = [o = x_0, x_1, \dots, x_m = x], \text{ then} \\ &g(x) = f(x_1) + \sum_{k=1}^{m-1} \frac{f(x_{k+1})}{\lambda(x_1)\cdots\lambda(x_k)}. \end{aligned} \tag{9.63}$$

Admitting the value $+\infty$, we can extend $g = g^f$ to ∂T by setting

$$g(\xi) = f(x_1) + \sum_{k=1}^{\infty} \frac{f(x_{k+1})}{\lambda(x_1)\cdots\lambda(x_k)}, \quad \text{if } \pi(o,\xi) = [o = x_0, x_1, \dots]. \tag{9.64}$$

9.65 Lemma. *Suppose that T is such that $\deg(x) \geq 2$ for all $x \in T \setminus \{o\}$. Then the function $g = g^f$ of (9.63) satisfies*

$$Pg(x) = g(x) \text{ for all } x \neq o, \quad \text{and} \quad Pg(o) = Pf(o) > 0 = g(o);$$

*it is **subharmonic**: $Pg \geq g$.*

Proof. Since $\deg(x) \geq 2$, we have $\lambda(x) > 0$ for all $x \in T \setminus \{o\}$. Let us define $a(o) = 1$ and recursively for $x \neq o$

$$a(x) = a(x^-)/\lambda(x).$$

Then $g(x)$ is also defined recursively by

$$g(o) = 0 \quad \text{and} \quad g(x) = g(x^-) + a(x^-)f(x), \ x \neq o.$$

Therefore, if $x \neq o$ then

$$\begin{aligned} Pg(x) &= p(x,x^-)g(x^-) + \sum_{y:y^-=x} p(x,y)\big(g(x) + a(x)f(y)\big) \\ &= p(x,x^-)\big(g(x) - a(x^-)f(x)\big) \\ &\qquad + \big(1 - p(x,x^-)\big)g(x) + \underbrace{a(x)\lambda(x)}_{a(x^-)} p(x,x^-)f(x) \\ &= g(x). \end{aligned}$$

It is clear that $Pg(o) = Pf(o) = \sum_x p(o,x)f(x)$. □

9.66 Corollary. *If* $\deg(x) \geq 2$ *for all* $x \in T \setminus \{o\}$ *and* $g = g^f$ *is bounded then the random walk is transient.*

Proof. If $g(x) < M$ for all x, then $M - g(x)$ is a non-constant, positive superharmonic function. Theorem 6.21 yields transience. □

We have defined the function $g^f = g_o^f$ with respect to the root o, and we can also define an analogous function g_v with respect to another reference vertex v. (Predecessors have then to be considered with respect to v instead of o.) Then, with $\lambda(x)$ defined with respect to o as above, we have the following.

$$\text{If } v \sim o \text{ then } g_o^f(x) = f(v) + \frac{1}{\lambda(v)} g_v^f(x) \text{ for all } x \in T_{o,v} \setminus \{v\}.$$

9.67 Corollary. *If there is a cone* $T_{v,w}$ $(v \sim w)$ *of* T *such that* $\deg(x) \geq 2$ *for all* $x \in T_{v,w}$*, and the function* $g = g^f$ *of* (9.63) *is bounded on that cone, then the random walk on* T *is transient, and every element of* $\partial T_{v,w}$ *is a transient end.*

Proof. By induction on $d(o,v)$, we infer from the above formula that $g = g_o$ is bounded on $T_{v,w}$ if and only if g_v is bounded on $T_{v,w}$. Equivalently, it is bounded on the branch $B_{[v,w]} = T_{v,w} \cup \{v\}$. On $B_{[v,w]}$, we have the random walk with transition matrix $P_{[v,w]}$ which coincides with the restriction of P along each oriented edge of that branch, with the only exception that $p_{[v,w]}(v,w) = 1$. See (9.2). We can now apply Corollary 9.66 to $P_{[v,w]}$ on $B_{[v,w]}$, with g_v in the place of g. It follows that

$P_{[v,w]}$ is transient on $B_{[v,w]}$. We know that this holds if and only if $F(w,v) < 1$, which implies transience of P on T.

Furthermore, If g is bounded on $T_{v,w}$, then it is bounded on every sub-cone $T_{x,y}$, where $x \sim y$ and $x \in \pi(w,y)$. Therefore $F(y,x) < 1$ for all those edges $[x,y]$. This says that all ends of $T_{v,w}$ are transient. □

Another by-product of Corollaries 9.66 and 9.67 is the following criterion.

9.68 Corollary. *Suppose that there are a bounded, strictly positive function f on $T \setminus \{o\}$ and a number $\lambda > 0$ such that*

$$\sum_{y\,:\,y^- = x} p(x,y) f(y) \geq \lambda\, p(x,x^-) f(x) \quad \text{for all } x \in T \setminus \{o\}.$$

If $\lambda > 1$ then every end of T is transient.

Proof. We can choose $f(o) > 0$ arbitrarily. Since f is bounded and $\lambda(x) \geq \lambda > 1$ for all $x \neq o$, we have $\sup g^f \leq \sup f \cdot \sum_n \lambda^{-n} < \infty$. □

We remark that the criteria of the last corollaries can also be rewritten in terms of the conductances $a(x,y) = \mathsf{m}(x)p(x,y)$. Therefore, if we find an arbitrary subtree of T to which one of them applies with respect to a suitable root vertex, then we get transience of that subtree as a subnetwork, and thus also of the random walk on T itself. See Exercise 4.54.

We now consider the extension (9.64) of $g = g^f$ to the space of ends of T.

9.69 Theorem. *Suppose that T is such that* $\deg(x) \geq 2$ *for all $x \in T \setminus \{o\}$.*

(a) *If $g(\xi) < \infty$ for all $\xi \in \partial T$, then the random walk is transient.*

(b) *If the random walk is transient, then $g(\xi) < \infty$ for ν_o-almost every $\xi \in T$.*

Proof. (a) Suppose that the random walk is recurrent. By Corollary 9.66, g cannot be bounded. There is a vertex $w(1) \neq o$ such that $g\big(w(1)\big) \geq 1$. Let $v(1) = w(1)^-$. Then $F\big(w(1), v(1)\big) = 1$, the cone $T_{v(1),w(1)}$ is recurrent. By Corollary 9.67, g is unbounded on $T_{v(1),w(1)}$, and we find $w(2) \neq w(1)$ in $T_{v(1),w(1)}$ such that $g\big(w(2)\big) \geq 2$.

We now proceed inductively. Given $w(n) \in T_{v(n-1),w(n-1)} \setminus \{w(n-1)\}$ with $g\big(w(n)\big) \geq n$, we set $v(n) = w(n)^-$. Recurrence of $T_{v(n),w(n)}$ implies via Corollary 9.67 that g is unbounded on that cone, and there must be $w(n+1) \in T_{v(n),w(n)} \setminus \{w(n)\}$ with $g\big(w(n+1)\big) \geq n+1$.

The points $w(n)$ constructed in this way lie on an infinite ray. If ξ is the corresponding end, then $g(\xi) \geq g\big(w(n)\big) \to \infty$. This contradicts the assumption of finiteness of g on ∂T.

(b) Suppose transience. Recall that $P\,G(x,o) = G(x,o)$, when $x \neq o$, while $P\,G(o,o) = G(o,o)-1$. Taking Lemma 9.65 into account, we see that the function $h(x) = g(x)+c\,G(x,o)$ is positive harmonic, where $c = Pf(o)$ is chosen in order to compensate the strict subharmonicity of g at o. By Corollary 7.31 from the section about martingales, we know that $\lim h(Z_n)$ exists and is Pr_o-almost surely finite. Also, by Exercise 7.32, $\lim G(Z_n,o) = 0$ almost surely. Therefore

$$\lim_{n\to\infty} g(Z_n) \text{ exists and is } \mathsf{Pr}_o\text{-almost surely finite.}$$

Proceeding once more as in the proof of Theorem 9.43, we have $\epsilon_k \to \infty$ Pr_o-almost surely, where $\epsilon_k = \max\{n : Z_n = v_k(Z_\infty)\}$. Then

$$\lim_{k\to\infty} g\big(v_k(Z_\infty)\big) < \infty \quad \mathsf{Pr}_o\text{-almost surely.}$$

Therefore

$$\nu_o\big(\{\xi \in \partial T : g(\xi) < \infty\}\big) = \mathsf{Pr}_o\big(Z_\infty \in \{\xi \in \partial T : \lim_{k\to\infty} g\big(v_k(\xi)\big) < \infty\}\big) = 1,$$

as proposed. □

9.70 Exercise. Prove the following strengthening of Theorem 9.69 (a).

Suppose that there is a cone $T_{v,w}$ ($v \sim w$) of T such that $\deg(x) \geq 2$ *for all $x \in T_{v,w}$, and the extension* (9.64) *of $g = g^f$ to the boundary satisfies $g(\xi) < \infty$ for every $\xi \in \partial T_{v,w}$. Then the random walk is transient. Furthermore, every end $\xi \in \partial T_{v,w}$ is transient.* □

The simplest choice for f is $f \equiv 1$. In this case, the function $g = g^1$ has the following form.

$$\begin{aligned} &g(o) = 0, \quad g(v) = 1 \text{ for } v \sim o, \quad \text{and} \\ &g(x) = 1 + \sum_{k=1}^{m-1} \prod_{i=1}^{k} \frac{p(x_i, x_{i-1})}{1 - p(x_i, x_{i-1})}, \end{aligned} \tag{9.71}$$

if $\pi(o,x) = [o = x_0, x_1, \dots, x_m = x]$ with $m \geq 2$.

9.72 Examples. In the following examples, we always choose $f \equiv 1$, so that $g = g^1$ is as in (9.71).

(a) For simple random walk on the homogeneous tree with degree $q+1$ of Example 9.29 and Figure 26, we have for the function $g = g^1$

$$g(x) = 1 + q^{-1} + \cdots + q^{-|x|+1} \quad \text{for } x \neq o.$$

The function g is bounded, and SRW is transient, as we know.

(b) Consider SRW on the tree of Example 9.33 and Figure 27. There, the recurrent ends are dense in ∂T, and $g(\eta_x) = \infty$ for each of them. Nevertheless, the random walk is transient.

(c) Next, consider SRW on the tree T of Figure 28 in Example 9.46.

For the vertex $k \geq 2$ on the backbone, we have $g(k) = 2 - 2^{-k+2}$.

For its neighbour k' on the finite hair with length $f(k)$ attached at k, the value is $g(k') = 2 - 2^{-k+1}$.

The function increases linearly along that hair, and for the root o_k of the binary tree attached at the endpoint of that hair, $g(o_k) = g(k) + 2^{-k+1} f(k)$.

Finally, if x is a vertex on that binary tree and $d(x, o_k) = m$ then $g(x) = g(o_k) + 2^{-k+1}(1 - 2^{-m})$.

We obtain the following values of g on ∂T.

$$g(\varpi) = 2 \quad \text{and} \quad g(\xi) = 2 - 2^{-k+2} + 2^{-k+1}\big(f(k) + 1\big), \quad \text{if } \xi \in \partial T_{k'}.$$

If we choose $f(k)$ big enough, e.g. $f(k) = k\, 2^k$, then the function g is everywhere finite, but unbounded on ∂T.

(d) Finally, we give an example which shows that for the transience criterion of Theorem 9.69 it is essential to have no vertices with degree 1 (at least in some cone of the tree). Consider the half line $\mathbb{N}$ with root $o = 1$, and attach a "dangling edge" at each $k \in \mathbb{N}$. See Figure 34. SRW on this tree is recurrent, but the function g is bounded.

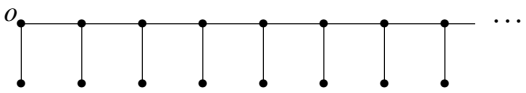

Figure 34

9.73 Exercise. Show that for the random walk of Example 9.47, when $3 \leq s \leq \infty$, the function g is bounded.

[Hint: when $\max_i p_i < 1/2$ this is straightforward. For the general case show that $\max\big\{\frac{p_i}{1-p_i}\frac{p_j}{1-p_j} : i, j \in \mathcal{I}, i \neq j\big\} < 1$ and use this fact appropriately.] □

With respect to $f \equiv 1$, the function $g = g^1$ of (9.71) was introduced by Bajunaid, Cohen, Colonna and Singman [3] (it is denoted H there). The proofs of the related results have been generalized and simplified here. Corollary 9.68 is part of a criterion of Gilch and Müller [27], who used a different method for the proof.

Trees with finitely many cone types

We now introduce a class of infinite, locally finite trees & random walks which comprise a variety of interesting examples and allow many computations. This includes a practicable recurrence criterion that is both necessary and sufficient.

We start with (T, P), where P is a nearest neighbour transition matrix on the locally finite tree T. As above, we fix an "origin" $o \in T$. For $x \in T \setminus \{o\}$, we consider the cone $T_x = T_{x^-,x}$ of x as labelled tree with root x. The labels are the probabilities $p(v, w)$, $v, w \in T_x$ $(v \sim w)$.

9.74 Definition. Two cones are *isomorphic,* if there is a root-preserving bijection between the two that also preserves neighbourhood as well as the labels of the edges. A *cone type* is an isomorphism class of cones T_x, $x \neq o$.

The pair (T, P) is called a *tree with finitely many cone types* if the number of distinct cone types is finite.

We write $\mathcal{I}$ for the finite set of cone types. The type (in $\mathcal{I}$) of $x \in T \setminus \{o\}$ is the cone type of T_x and will be denoted by $\iota(x)$. Suppose that $\iota(x) = i$. Let $\mathsf{d}(i, j)$ be the number of neighbours of x in T_x that are of type j. We denote

$$\mathsf{p}(i, j) = \sum_{y\,:\, y^- = x, \iota(y) = j} p(x, y), \quad \text{and} \quad \mathsf{p}(i-) = p(x, x^-) = 1 - \sum_{j \in \mathcal{I}} \mathsf{p}(i, j).$$

As the notation indicates, those numbers depend only on i and j, resp. (for the backward probability) only on i. In particular, we must have $\sum_j \mathsf{p}(i, j) < 1$. We also admit leaves, that have no forward neighbour, in which case $\mathsf{p}(i-) = 1$.

We can encode this information in a labelled oriented graph with multiple edges over the vertex set $\mathcal{I}$. For $i, j \in \mathcal{I}$, there are $\mathsf{d}(i, j)$ edges from i to j, which carry the labels $p(x, y)$, where x is any vertex with type i and y runs through all forward neighbours with type j of x. This does not depend on the specific vertex x with $\iota(x) = i$.

Next, we augment the graph of cone types $\mathcal{I}$ by the vertex o (the root of T). In this new graph $\mathcal{I}_o$, we draw $\mathsf{d}(o, i)$ edges from o to i, where $\mathsf{d}(o, i)$ is the number of neighbours x of o in T with $\iota(x) = i$. Each of those edges carries one of the labels $p(o, x)$, where $x \sim o$ and $\iota(x) = i$. As above, we write

$$\mathsf{p}(o, i) = \sum_{x\,:\, x \sim o, \iota(x) = i} p(o, x).$$

Then $\sum_i \mathsf{p}(o, i) = 1$. The original tree with its transition probabilities can be recovered from the graph $\mathcal{I}_o$ as the *directed cover*. It consists of all oriented paths in $\mathcal{I}_o$ that start at o. Since we have multiple edges, such a path is not described fully by its sequence of vertices; a path is a finite sequence of oriented edges of $\mathcal{I}$ with the property that the terminal vertex of one edge has to coincide with the initial vertex of the next edge in the path. This includes the empty path without any edge that starts and ends at o. (This path is denoted o.) Given two such paths in $\mathcal{I}_o$, here denoted x, y, we have $y^- = x$ as vertices of the tree if y extends the path x by one edge at the end. Then $p(x, y)$ is the label of that final edge from $\mathcal{I}_o$. The backward probability is then $p(y, x) = \mathsf{p}(j-)$, if the path y terminates at $j \in \mathcal{I}$.

9.75 Examples. (a) Consider simple random walk on the homogeneous tree with degree $q+1$ of Example 9.29 and Figure 26. There is one cone type, $\mathcal{I} = \{1\}$, we have $\mathsf{d}(1,1) = q$, and each of the q loops at vertex ($\equiv$ type) 1 carries the label (probability) $1/(q+1)$. Furthermore, $\mathsf{d}(o,1) = q+1$, and each of the $q+1$ edges from o to 1 carries the label $1/(q+1)$.

(b) Consider simple random walk on the tree of Example 9.33 and Figure 27. There are two cone types, $\mathcal{I} = \{1,2\}$, where 1 is the type of any vertex of the homogeneous tree, and 2 is the type of any vertex on one of the hairs. We have $\mathsf{d}(1,1) = q$, and each of the q loops at vertex ($\equiv$ type) 1 carries the label $1/(q+2)$, while $\mathsf{d}(1,2) = \mathsf{d}(2,2) = 1$ with labels $1/(q+2)$ and $1/2$, respectively.

Furthermore, $\mathsf{d}(o,1) = q+1$, and each of the $q+1$ edges from o to 1 carries the label (probability) $1/(q+2)$, while $\mathsf{d}(o,2) = 1$ and $p(o,1) = 1/(q+2)$.

Figure 35 refers to those two examples with $q = 2$.

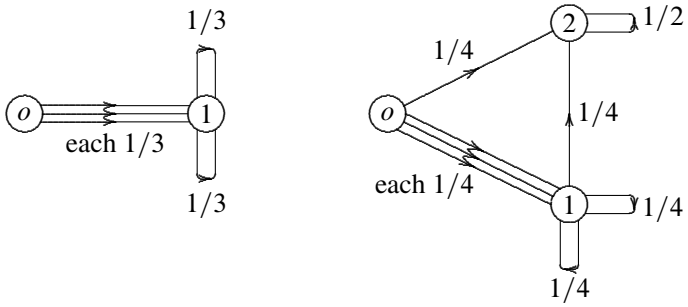

Figure 35. The graphs $\mathcal{I}_o$ for SRW on $\mathbb{T}_2$ and on $\mathbb{T}_2$ with hairs.

(c) Consider the random walk of Example 9.47 in the case when $s < \infty$. Again, we have finitely many cone types. As a set, $\mathcal{I} = \{1, \dots, s\}$. The graph structure is that of a complete graph: there is an oriented edge $[i,j]$ for every pair of distinct elements $i, j \in \mathcal{I}$. We have $\mathsf{d}(i,j) = 1$ and $\mathsf{p}(i,j) = \mathsf{p}(j-) = p_j$. (As a matter of fact, when some of the p_j coincide, we have a smaller number of distinct cone types and higher multiplicities $\mathsf{d}(i,j)$, but we can as well maintain the same model.)

In addition, in the augmented graph $\mathcal{I}_o$, there is an edge $[o,i]$ with $\mathsf{d}(o,i) = 1$ and $\mathsf{p}(o,i) = p_i$ for each $i \in \mathcal{I}$.

The reader is invited to draw a figure.

(d) Another interesting example is the following. Let $0 < \alpha < 1$ and $\mathcal{I} = \{1,-1\}$. We let $\mathsf{d}(1,1) = 2$, $\mathsf{d}(-1,1) = \mathsf{d}(-1,-1) = 1$ and $\mathsf{d}(1,-1) = 0$. The probability labels are $\alpha/2$ at each of the loops at vertex ($\equiv$ type) 1 as well as at the edge from -1 to 1. Furthermore, the label at the loop at vertex -1 is $1-\alpha$.

For the augmented graph $\mathcal{I}_o$, we let $\mathsf{d}(o,-1) = 1$ with $\mathsf{p}(o,-1) = 1-\alpha$, and $\mathsf{d}(o,1) = 2$ with each of the resulting two edges having label $\alpha/2$.

The graph $\mathcal{I}_o$ is shown in Figure 36. The reader is solicited to draw the resulting tree and transition probabilities. We shall reconsider that example later on.

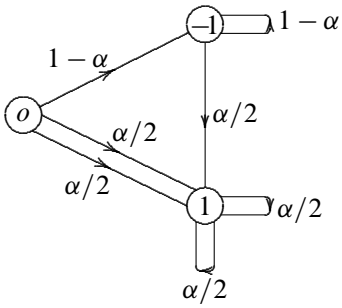

Figure 36. Random walk on $\mathbb{T}_2$ in horocyclic layers.

We return to general trees with finitely many cone types. If $\iota(x) = i$ then $F(x, x^-|z) = F_i(z)$ depends only on the type i of x. Proposition 9.3 (b) leads to a finite system of algebraic equations of degree ≤ 2,

$$F_i(z) = \mathsf{p}(i-)z + \sum_{j \in \mathcal{I}} \mathsf{p}(i, j)\, z\, F_j(z) F_i(z). \tag{9.76}$$

(If x is a terminal vertex then $F_i(z) = z$.) In Example 9.47, we have already worked with these equations, and we shall return to them later on.

We now define the non-negative matrix

$$A = \big(\mathsf{a}(i, j)\big)_{i,j \in \mathcal{I}} \quad \text{with } \mathsf{a}(i, j) = \mathsf{p}(i, j)/\mathsf{p}(i-). \tag{9.77}$$

Let $\rho(A)$ be its largest non-negative eigenvalue; compare with Proposition 3.44. This number can be seen as an overall average or balance of quotients of forward and backward probabilities.

9.78 Theorem. *Let (T, P) be a random walk on a tree with finitely many cone types, and let A be the associated matrix according to (9.77). Then the random walk is*

- *positive recurrent if and only if $\rho(A) < 1$,*
- *null recurrent if and only if $\rho(A) = 1$, and*
- *transient if and only if $\rho(A) > 1$.*

Proof. First, we consider positive recurrence. We have to show that $\mathsf{m}(T) < \infty$ for the measure of (9.6) if and only if $\rho(A) < 1$.

Let $T_i = T_x$ be a cone in T with type $\iota(x) = i$. We define a measure m_i on T_i by

$$\mathsf{m}_i(x) = 1 \quad \text{and} \quad \mathsf{m}_i(y) = \frac{p(y^-, y)}{p(y, y^-)} \mathsf{m}_i(y^-) \quad \text{for } y \in T_x \setminus \{x\}.$$

Then

$$\mathsf{m}(T) = 1 + \sum_{i \in \mathcal{I}} \frac{\mathsf{p}(o, i)}{\mathsf{p}(i-)} \mathsf{m}_i(T_i).$$

We need to show that $\mathsf{m}_i(T_i) < \infty$ for all i. For $n \geq 0$, let $T_i^n = T_x^n$ be the ball of radius n centred at x in the cone T_x (all vertices at graph distance $\leq n$ from x). Then

$$\mathsf{m}_i(T_i^0) = 1 \quad \text{and} \quad \mathsf{m}(T_i^n) = 1 + \sum_{j \in \mathcal{I}} \mathsf{a}(i, j) \mathsf{m}_j(T_j^{n-1}), \quad n \geq 1$$

for each $i \in \mathcal{I}$. Consider the column vectors $\mathbf{m} = \big(\mathsf{m}(T_i)\big)_{i \in \mathcal{I}}$ and $\mathbf{m}^{(n)} = \big(\mathsf{m}(T_i^n)\big)_{i \in \mathcal{I}}$, $n \geq 1$, as well as the vector $\mathbf{1}$ over $\mathcal{I}$ with all entries equal to 1. Then we see that

$$\mathbf{m}^{(n)} = \mathbf{1} + A\, \mathbf{m}^{(n-1)} = \mathbf{1} + A\, \mathbf{1} + A^2\, \mathbf{1} + \cdots + A^n\, \mathbf{1},$$

and $\mathbf{m}^{(n)} \to \mathbf{m}$ as $n \to \infty$. By Proposition 3.44, this limit is finite if and only if $\rho(A) < 1$.

Next, we show that transience holds if and only if $\rho(A) > 1$.

We start with the "if" part. Assume that $\rho = \rho(A) > 1$. Once more by Proposition 3.44, there is an irreducible class $\mathcal{J} \subset \mathcal{I}$ of the matrix A such that $\rho(A) = \rho(A_{\mathcal{J}})$, where $A_{\mathcal{J}}$ is the restriction of A to $\mathcal{J}$. By the Perron–Frobenius theorem, there is a column vector $\mathbf{h} = \big(h(i)\big)_{i \in \mathcal{J}}$ with strictly positive entries such that $A_{\mathcal{J}} \mathbf{h} = \rho(A) \cdot \mathbf{h}$.

We fix a vertex x_0 of T with $\iota(x_0) \in \mathcal{J}$ and consider the subtree $T_{\mathcal{J}}$ of T_{x_0} which is spanned by all vertices $y \in T_{x_0}$ which have the property that $\iota(w) \in \mathcal{J}$ for all $w \in \pi(x_0, y)$. Note that $T_{\mathcal{J}}$ is infinite. Indeed, every vertex in $T_{\mathcal{J}}$ has at least one successor in $T_{\mathcal{J}}$, because the matrix $A_{\mathcal{J}}$ is non-zero and irreducible.

We define $f(x) = h\big(\iota(x)\big)$ for $x \in T_{\mathcal{J}}$. This is a bounded, strictly positive function, and

$$\sum_{y \in T_{\mathcal{J}},\, y^- = x} p(x, y) f(y) = \rho \cdot p(x, x^-) f(x) \quad \text{for all } x \in T_{\mathcal{J}}.$$

Therefore Corollary 9.68 applies with $\lambda = \rho(A) > 1$ to $T_{\mathcal{J}}$ as a subnetwork of T, and transience follows; compare with the remarks after the proof of that corollary.

Finally, we assume transience and have to show that $\rho(A) > 1$.

Consider the transient skeleton T_{tr} as a subnetwork, and the associated transition probabilities of (9.28). Then $(T_{\mathrm{tr}}, P_{\mathrm{tr}})$ is also a tree with finitely many cone types: if $x_1, x_2 \in T_{\mathrm{tr}}$ have the same cone type in (T, P), then they also have the same cone type in $(T_{\mathrm{tr}}, P_{\mathrm{tr}})$. These transient cone types are just

$$\mathcal{I}_{\mathrm{tr}} = \{i \in \mathcal{I} : F_i(1) < 1\}.$$

Furthermore, it follows from Exercise 9.26 that when $i \in \mathcal{I}_{\mathrm{tr}}$, $j \in \mathcal{I}$ and $j \to i$ with respect to the non-negative matrix A (i.e., $\mathsf{a}^{(n)}(j,i) > 0$ for some n), then $j \in \mathcal{I}_{\mathrm{tr}}$. Therefore, $\mathcal{I}_{\mathrm{tr}}$ is a union of irreducible classes with respect to A. From (9.28), we also see that the matrix A_{tr} associated with P_{tr} according to (9.77) is just the restriction of A to $\mathcal{I}_{\mathrm{tr}}$. Proposition 3.44 implies $\rho(A) \geq \rho(A_{\mathrm{tr}})$. The proof will be completed if we show that $\rho(A_{\mathrm{tr}}) > 1$.

For this purpose, we may assume without loss of generality that $T = T_{\mathrm{tr}}$, $P = P_{\mathrm{tr}}$ and $A = A_{\mathrm{tr}}$. Consider the diagonal matrix

$$D(z) = \operatorname{diag}\big(F_i(z)\big)_{i \in \mathcal{I}},$$

and let I be the identity matrix over $\mathcal{I}$. For $z = 1$, we can rewrite (9.76) as

$$1 - F_i(1) = \sum_j F_i(1)\, \mathsf{a}(i,j)\, \big(1 - F_j(1)\big).$$

Equivalently, the non-negative, irreducible matrix

$$Q = \big(I - D(1)\big)^{-1} D(1)\, A \big(I - D(1)\big) \tag{9.79}$$

is stochastic. Thus, $\rho(Q) = 1$, and therefore also $\rho\big(D(1)\, A\big) = 1$. The (i,j)-element of the last matrix is $F_i(1)\mathsf{a}(i,j)$. It is 0 precisely when $\mathsf{a}(i,j) = 0$, while $F_i(1)\mathsf{a}(i,j) < \mathsf{a}(i,j)$ strictly, when $\mathsf{a}(i,j) > 0$. Therefore Proposition 3.44 in combination with Exercise 3.43 (applying the latter to the irreducible classes of A) yields $\rho(A) > \rho\big(D(1)\, A\big) = 1$. □

The above recurrence criterion extends the one for "homesick" random walk of Lyons [41]; it was proved by Nagnibeda and Woess [44], and a similar result in completely different terminology had been obtained by Gairat, Malyshev, Men'shikov and Pelikh [22]. (In [44], the proof that $\lambda(A) = 1$ implies null recurrence is somewhat sloppy.)

9.80 Exercise. Compute the largest eigenvalue $\rho(A)$ of A for each of the random walks of Example 9.75. □

H Rate of escape and spectral radius

We now plan to give a small glimpse at some results concerning the asymptotic behaviour of the graph distances $d(Z_n, o)$, which will turn out to be related with the spectral radius $\rho(P)$.

A sum $S_n = X_1 + \cdots + X_n$ of independent, integer (or real) random variables defines a random walk on the additive group of integer (or real) numbers, compare with (4.18). If the X_k are integrable then the law of large numbers implies that

$$\frac{1}{n} d(S_n, 0) = \frac{1}{n} |S_n| \to \ell \quad \text{almost surely, where } \ell = |\mathsf{E}(X_1)|.$$

We can ask whether analogous results hold for an arbitrary Markov chain with respect to some metric on the underlying state space. A natural choice is of course the graph metric, when the graph of the Markov chain is symmetric. Here, we shall address this in the context of trees, but we mention that there is a wealth of results for different types of Markov chains, and we shall only scratch the surface of this topic.

Recall the definition (2.29) of the spectral radius of an irreducible Markov chain (resp., an irreducible class). The following is true for arbitrary graphs in the place of trees.

9.81 Theorem. *Let X be a connected, symmetric graph and P the transition matrix of a nearest neighbour random walk on X. Suppose that there is $\varepsilon_0 > 0$ such that $p(x, y) \geq \varepsilon_0$ whenever $x \sim y$, and that $\rho(P) < 1$. Then there is a constant $\underline{\ell} > 0$ such that*

$$\liminf_{n\to\infty} \frac{1}{n} d(Z_n, Z_0) \geq \underline{\ell} \quad \mathsf{Pr}_x \text{-almost surely for every } x \in X.$$

Proof. We set $\rho = \rho(P)$ and claim that for all $x, y \in X$ and $n \geq 0$,

$$p^{(n)}(x, y) \leq (\rho/\varepsilon_0)^{d(x,y)}\, \rho^n \quad \text{for all } x, y \in X \text{ and } n \geq 0.$$

This is true when $x = y$ by Theorem 2.32. In general, let $d = d(x, y)$. We have by assumption $p^{(d)}(y, x) \geq \varepsilon_0^d$, and therefore

$$p^{(n)}(x, y)\, \varepsilon_0^d \leq p^{(n+d)}(x, x) \leq \rho^{n+d}.$$

The claimed inequality follows by dividing by ε_0^d.

Note that we must have $\deg(x) \leq M = \lfloor 1/\varepsilon_0 \rfloor$ for every $x \in X$, where $M \geq 2$. This implies that for each $x \in X$ and $k \geq 1$,

$$|\{y \in X : d(x, y) \leq k\}| \leq M(M-1)^{k-1}.$$

Also, if $x \sim y$ then $\rho^2 \geq p^{(2)}(x,x) \geq p(x,y)p(y,x) \geq \varepsilon_0^2$, so that $\varepsilon_0 \leq \rho$. Since $\rho < 1$, we can now choose a real number $\underline{\ell} > 0$ such that $(\rho/\varepsilon_0^2)^{\underline{\ell}} < 1/\rho$. Consider the sets

$$A_n = [d(Z_n, Z_0) < \underline{\ell}\, n]$$

in the trajectory space. We have

$$\begin{aligned}
\mathsf{Pr}_x(A_n) &= \sum_{y: d(x,y) < \underline{\ell} n} p^{(n)}(x,y) \\
&\leq \sum_{y: d(x,y) < \underline{\ell} n} (\rho/\varepsilon_0)^{d(x,y)} \rho^n \\
&\leq \rho^n \Big(1 + \sum_{k=1}^{\lfloor \underline{\ell} n \rfloor} \sum_{y: d(x,y)=k} (\rho/\varepsilon_0)^k \Big) \\
&= \rho^n \Big(1 + \sum_{k=1}^{\lfloor \underline{\ell} n \rfloor} M(M-1)^{k-1} (\rho/\varepsilon_0)^k \Big) \\
&= \rho^n \left(1 + M(\rho/\varepsilon_0) \frac{\big((M-1)\rho/\varepsilon_0\big)^{\lfloor \underline{\ell} n \rfloor} - 1}{\big((M-1)\rho/\varepsilon_0\big) - 1} \right) \\
&\leq C \left((\rho/\varepsilon_0^2)^{\underline{\ell}} \rho \right)^n,
\end{aligned}$$

where $C = \frac{M\rho}{(M-1)\rho - \varepsilon_0} > 0$. Therefore $\sum_n \mathsf{Pr}_x(A_n) < \infty$, and by the Borel–Cantelli lemma, $\mathsf{Pr}_x(\limsup A_n) = 0$, or equivalently,

$$\mathsf{Pr}_x \Big[\bigcup_{k \geq 1} \bigcap_{n \geq k} A_n^c \Big] = 1$$

for the complements of the A_n. But this says that Pr_x-almost surely, one has $d(Z_n, Z_0) \geq \underline{\ell}\, n$ for all but finitely many n. □

We see that a "reasonable" random walk with $\rho(P) < 1$ moves away from the starting point at linear speed. ("Reasonable" means that $p(x,y) \geq \varepsilon_0$ along each edge). So we next ask when it is true that $\rho(P) < 1$. We start with simple random walk. When Γ is an arbitrary symmetric, locally finite graph, then we write $\rho(\Gamma) = \rho(P)$ for the spectral radius of the transition matrix P of SRW on Γ.

9.82 Exercise. Show that for simple random walk on the homogeneous tree $\mathbb{T}_q$ with degree $q+1 \geq 2$, the spectral radius is

$$\rho(\mathbb{T}_q) = \frac{2\sqrt{q}}{q+1}.$$

[Hints. Variant 1: use the computations of Example 9.47. Variant 2: consider the factor chain $(\overline{Z}_n)$ on $\mathbb{N}_0$ where $\overline{Z}_n = |Z_n| = d(Z_n, o)$ for the simple random walk (Z_n) on $\mathbb{T}_q$ starting at o. Then determine $\rho(\overline{P})$ from the computations in Example 3.5.] □

For the following, T need not be locally finite.

9.83 Theorem. *Let T be a tree with root o and P a nearest neighbour transition matrix with the property that $p(x, x^-) \le 1 - \alpha$ for each $x \in T \setminus \{o\}$, where $1/2 < \alpha < 1$. Then*

$$\rho(P) \le 2\sqrt{\alpha(1-\alpha)}$$

and

$$\liminf_{n\to\infty} \frac{1}{n} d(Z_n, Z_0) \ge 2\alpha - 1 \quad \mathsf{Pr}_x\text{-almost surely for every } x \in T.$$

Furthermore,

$$F(x, x^-) \le (1-\alpha)/\alpha \quad \text{for every } x \in T \setminus \{o\}.$$

In particular, if T is locally finite then the Green kernel vanishes at infinity.

Proof. We define the function $g = g_\alpha$ on $\mathbb{N}_0$ by

$$g_\alpha(0) = 1 \quad \text{and} \quad g_\alpha(n) = \big(1 + (2\alpha - 1)n\big)\left(\tfrac{1-\alpha}{\alpha}\right)^{n/2} \quad \text{for } n \ge 1.$$

Then $g(1) = 2\sqrt{\alpha(1-\alpha)}$, and a straightforward computation shows that

$$(1-\alpha)\, g(n-1) + \alpha\, g(n+1) = 2\sqrt{\alpha(1-\alpha)}\, g(n) \quad \text{for } n \ge 1.$$

Also, g is decreasing. On our tree T, we define the function f by $f(x) = g_\alpha(|x|)$, where (recall) $|x| = d(x, o)$. Then we have for the transition matrix P of our random walk

$$Pf(o) = g_\alpha(1) = 2\sqrt{\alpha(1-\alpha)}\, g_\alpha(0) = 2\sqrt{\alpha(1-\alpha)}\, f(o),$$

and for $x \ne o$

$$Pf(x) - 2\sqrt{\alpha(1-\alpha)}\, f(x) = \big(p(x, x^-) - (1-\alpha)\big) \underbrace{\big(g_\alpha(|x|-1) - g_\alpha(|x|+1)\big)}_{>0}.$$

We see that when $p(x, x^-) \le 1 - \alpha$ for all $x \ne o$ then $Pf \le 2\sqrt{\alpha(1-\alpha)}\, f$ and

$$p^{(n)}(o, o) \le P^n f(o) \le \Big(2\sqrt{\alpha(1-\alpha)}\Big)^n f(o) = \Big(2\sqrt{\alpha(1-\alpha)}\Big)^n$$

for all n. The proposed inequality for $\rho(P)$ follows.

Next, we consider the rate of escape. For this purpose, we construct a *coupling* of our random walk (Z_n) on T and the random walk (sum of i.i.d. random variables) (S_n) on $\mathbb{Z}$ with transition probabilities $\bar{p}(k,k+1)=\alpha$ and $\bar{p}(k,k-1)=1-\alpha$ for $k\in\mathbb{Z}$. Namely, on the state space $X=T\times\mathbb{Z}$, we consider the Markov chain with transition matrix Q given by

$$\begin{aligned}
q\big((o,k),(y,k+1)\big) &= p(o,y)\,\alpha \quad \text{and}\\
q\big((o,k),(y,k-1)\big) &= p(o,y)\,(1-\alpha), \quad \text{if } y^-=o,\\
q\big((x,k),(x^-,k-1)\big) &= p(x,x^-), \quad \text{if } x\neq o,\\
q\big((x,k),(y,k+1)\big) &= p(x,y)\frac{\alpha}{1-p(x,x^-)} \quad \text{and}\\
q\big((x,k),(y,k-1)\big) &= p(x,y)\frac{1-\alpha-p(x,x^-)}{1-p(x,x^-)}, \quad \text{if } x\neq o,\ y^-=x.
\end{aligned}$$

All other transition probabilities are $=0$. The two projections of X onto T and onto $\mathbb{Z}$ are compatible with these transition probabilities, that is, we can form the two corresponding factor chains in the sense of (1.30). We get that the Markov chain on X with transition matrix Q is $(Z_n,S_n)_{n\geq 0}$. Now, our construction is such that when Z_n walks backwards then S_n moves to the left: if $|Z_{n+1}|=|Z_n|-1$ then $S_{n+1}=S_n-1$. In the same way, when $S_{n+1}=S_n+1$ then $|Z_{n+1}|=|Z_n|+1$. We conclude that

$$|Z_n|\geq S_n \text{ for all } n, \text{ provided that } |Z_0|\geq S_0.$$

In particular, we can start the coupled Markov chain at $(x,|x|)$. Then S_n can be written as $S_n=|x|+X_1+\cdots+X_n$ where the X_n are i.i.d. integer random variables with $\mathsf{Pr}[X_k=1]=\alpha$ and $\mathsf{Pr}[X_k=-1]=1-\alpha$. By the law of large numbers, $S_n/n\to 2\alpha-1$ almost surely. This leads to the lower estimate for the velocity of escape of (Z_n) on T.

With the same starting point $(x,|x|)$, we see that (Z_n) cannot reach x^- before (S_n) reaches $|x|-1$ for the first time. That is, in our coupling we have

$$\boldsymbol{t}_{\mathbb{Z}}^{|x|-1}\leq \boldsymbol{t}_T^{x^-} \qquad \mathsf{Pr}_{(x,|x|)}\text{-almost surely,}$$

where the two first passage times refer to the random walks on $\mathbb{Z}$ and T, respectively. Therefore,

$$F_T(x,x^-)=\mathsf{Pr}_{(x,|x|)}\big[\boldsymbol{t}_T^{x^-}<\infty\big]\leq \mathsf{Pr}_{(x,|x|)}\big[\boldsymbol{t}_{\mathbb{Z}}^{|x|-1}<\infty\big]=F_{\mathbb{Z}}(|x|,|x|-1).$$

We know from Examples 3.5 and 2.10 that $F_{\mathbb{Z}}(|x|,|x|-1)=F_{\mathbb{Z}}(1,0)=(1-\alpha)/\alpha$. This concludes the proof. □

9.84 Exercise. Show that when $p(x,x^-)=1-\alpha$ for each $x\in T\setminus\{o\}$, where $1/2<\alpha<1$, then the three inequalities of Theorem 9.83 become equalities.

[Hint: verify that $(|Z_n|)_{n\geq 0}$ is a Markov chain.] □

9.85 Corollary. *Let T be a locally finite tree with $\deg(x) \geq q+1$ for all $x \in T$. Then the spectral radius of simple random walk on T satisfies $\rho(T) \leq \rho(\mathbb{T}_q)$.*

Indeed, in that case, Theorem 9.83 applies with $\alpha = q/(q+1)$.

Thus, when $\deg(x) \geq 3$ for all x then $\rho(T) < 1$. We next ask what happens with the spectral radius of SRW when there are vertices with degree 2.

We say that an *unbranched path* of length N in a symmetric graph Γ is a path $[x_0, x_1, \ldots, x_N]$ of distinct vertices such that $\deg(x_k) = 2$ for $k = 1, \ldots, N-1$. The following is true in any graph (not necessarily a tree).

9.86 Lemma. *Let Γ be a locally finite, connected symmetric graph. If Γ contains unbranched paths of arbitrary length, then $\rho(\Gamma) = 1$.*

Proof. Write $p_{\mathbb{Z}}^{(n)}(0,0)$ for the transition probabilities of SRW on $\mathbb{Z}$ that we have computed in Exercise 4.58 (b). We know that $\rho(\mathbb{Z}) = 1$ in that example.

Given any $n \in \mathbb{N}$, there is an unbranched path in Γ with length $2n+2$. If x is its midpoint, then we have for SRW on Γ that

$$p^{(2n)}(x,x) = p_{\mathbb{Z}}^{(2n)}(0,0),$$

since within the first $2n$ steps, our random walk cannot leave that unbranched path, where it evolves like SRW on $\mathbb{Z}$. We know from Theorem 2.32 that $p^{(2n)}(x,x) \leq \rho(\Gamma)^{2n}$. Therefore

$$\rho(\Gamma) \geq p_{\mathbb{Z}}^{(2n)}(0,0)^{1/(2n)} \to 1 \quad \text{as } n \to \infty.$$

Thus, $\rho(\Gamma) = 1$. □

Now we want to know what happens when there is an upper bound on the lengths of the unbranched paths. Again, our considerations are valid for an arbitrary symmetric graph $\Gamma = (X, E)$. We can construct a new graph $\widetilde{\Gamma}$ by replacing each non-oriented edge of Γ (= pair of oppositely oriented edges with the same endpoints) with an unbranched path of length k (depending on the edge). The vertex set X of Γ is a subset of the vertex set of $\widetilde{\Gamma}$. We call $\widetilde{\Gamma}$ a *subdivision* of Γ, and the maximum of those numbers k is the *maximal subdivision length* of $\widetilde{\Gamma}$ with respect to Γ.

In particular, we write $\Gamma_{(N)}$ for the subdivision of Γ where each non-oriented edge is replaced by a path of the same length N. Let (Z_n) be SRW on $\Gamma_{(N)}$. Since the vertex set $X_{(N)}$ of $\Gamma_{(N)}$ contains X as a subset, we can define the following sequence of stopping times.

$$\boldsymbol{t}_0 = 0, \quad \boldsymbol{t}_j = \inf\{n > \boldsymbol{t}_{j-1} : Z_n \in X,\ Z_n \neq Z_{\boldsymbol{t}_{j-1}}\} \quad \text{for } j \geq 1.$$

The set of those n is non-empty with probability 1, in which case the infimum is a minimum.

9.87 Lemma. (a) *The increments $\boldsymbol{t}_j - \boldsymbol{t}_{j-1}$ ($j \geq 1$) are independent and identically distributed with probability generating function*

$$\mathsf{E}_{x_0}(z^{\boldsymbol{t}_j - \boldsymbol{t}_{j-1}}) = \sum_{n=1}^{\infty} \mathsf{Pr}_x[\boldsymbol{t}_1 = n]\, z^n = \phi(z) = 1/Q_N(1/z), \quad x_0, x \in X,$$

where $Q_N(t)$ is the N-th Chebyshev polynomial of the first kind; see Example 5.6.

(b) *If y is a neighbour of x in Γ then*

$$\sum_{n=1}^{\infty} \mathsf{Pr}_x[\boldsymbol{t}_1 = n,\ Z_{\boldsymbol{t}_1} = y]\, z^n = \frac{1}{\deg(x)}\phi(z).$$

(c) *For any $x \in X$,*

$$\sum_{n=0}^{\infty} \mathsf{Pr}_x[Z_n = x,\ \boldsymbol{t}_1 > n]\, z^n = \psi(z) = \frac{(1/z)R_{N-1}(z)}{Q_N(1/z)},$$

where $R_{N-1}(t)$ is the $(N-1)$-st Chebyshev polynomial of the second kind.

Proof. (a) Let $Z_{\boldsymbol{t}_{j-1}} = x \in X \subset X_{(N)}$, and let $S(x)$ be the set consisting of x and the neighbours of x in the original graph Γ. In $\Gamma_{(N)}$, this set becomes a star-shaped graph $S_{(N)}(x)$ with centre x, where for each terminal vertex $y \in N(x) \setminus \{x\}$ there is a path from x to y with length N. Up to the stopping time $\boldsymbol{t}_j$, (Z_n) is SRW on $S_{(N)}(x)$. But for the latter simple random walk, we can construct the factor chain $\overline{Z}_n = d(Z_n, x)$, with $d(\cdot,\cdot)$ being the graph distance in $S_{(N)}(x)$. This is the birth-and-death chain on $\{0, \dots, N\}$ which we have considered in Example 5.6. Its transition probabilities are $\bar{p}(0,1) = 1$ and $\bar{p}(k, k\pm 1) = 1/2$ for $k = 1, \dots, N-1$. Then $\boldsymbol{t}_j$ is the first instant n after time $\boldsymbol{t}_{j-1}$ when $\overline{Z}_n = N$. But this just says that

$$\mathsf{Pr}_{x_0}[\boldsymbol{t}_j - \boldsymbol{t}_{j-1} = k] = f^{(k)}(0, N),$$

the probability that $(\overline{Z}_n)$ starting at 0 will first hit the state N at time k. The associated generating function $F(0, N|z) = 1/Q_N(1/z)$ was computed in Example 5.6.

Since this distribution does not depend on the specific starting point x_0 nor on $x = Z_{\boldsymbol{t}_{j-1}}$, the increments must be independent. The precise argument is left as an exercise.

(b) In $S_{(N)}(x)$ every terminal vertex y occurs as $Z_{\boldsymbol{t}_j}$ with the same probability $1/\deg(x)$. This proves statement (b).

(c) $\mathsf{Pr}_x[Z_n = x,\ \boldsymbol{t}_1 > n]$ is the probability that SRW on $S_{(N)}(x)$ returns to x at time n before visiting any of the endpoints of that star. With the same factor chain argument as above, this is the probability to return to 0 for the simple birth-and-death Markov chain on $\{0, 1, \dots, N\}$ with state 0 reflecting and state N absorbing. Therefore $\psi(z)$ is the Green function at 0 for that chain, which was computed at the end of Example 5.6. □

9.88 Exercise. (1) Complete the proof of the fact that the increments $\boldsymbol{t}_j - \boldsymbol{t}_{j-1}$ are independent. Does this remain true for an arbitrary subdivision of Γ in the place of $\Gamma_{(N)}$?

(2) Use Lemma 9.87 (b) and induction on k to show that for all $x, y \in X \subset X_{(N)}$,

$$\sum_{n=0}^{\infty} \mathsf{Pr}_x\big[\boldsymbol{t}_k = n,\ Z_n = y\big]\, z^n = p^{(k)}(x,y)\, \phi(z)^k,$$

where $p^{(k)}(x,y)$ refers to SRW on Γ.

[Hint: Lemma 9.87 (b) is that statement for $k = 1$.] □

9.89 Theorem. *Let $\Gamma = (X, E)$ be a connected, locally finite symmetric graph.*

(a) *The spectral radii of SRW on Γ and its subdivision $\Gamma_{(N)}$ are related by*

$$\rho(\Gamma_{(N)}) = \cos \frac{\arccos \rho(\Gamma)}{N}.$$

(b) *If $\widetilde{\Gamma}$ is an arbitrary subdivision of Γ with maximal subdivision length N, then*

$$\rho(\Gamma) \le \rho(\widetilde{\Gamma}) \le \rho(\Gamma_{(N)}).$$

Proof. (a) We write $G(x, y|z)$ and $G_{(N)}(x, y|z)$ for the Green functions of SRW on $\Gamma_{(N)}$ and Γ, respectively. We claim that for $x, y \in X \subset X_{(N)}$,

$$G_{(N)}(x, y|z) = G\big(x, y|\phi(z)\big)\, \psi(z), \tag{9.90}$$

where $\phi(z)$ and $\psi(z)$ are as in Lemma 9.87. Before proving this, we explain how it implies the formula for $\rho(\Gamma_{(N)})$.

All involved functions in (9.90) arise as power series with non-negative coefficients that are ≤ 1. Their radii of convergence are the respective smallest positive singularities (by Pringsheim's theorem, already used several times). Let $\mathsf{r}(\Gamma_{(N)})$ be the (common) radius of convergence of all the functions $G_{(N)}(x, y|z)$. By (9.90), it is the minimum of the radii of convergence of $G\big(x, y|\phi(z)\big)$ and $\psi(z)$.

We know from Example 5.6 that the radii of convergence of $\phi(z)$ and $\psi(z)$ (which are the functions $F(0, N|z)$ and $G(0, 0|z)$ of that example, respectively) coincide and are equal to $\mathsf{s} = 1/\cos\frac{\pi}{2N}$. Therefore the function $G\big(x, y|\phi(z)\big)$ is finite and analytic for each $z \in (0, \mathsf{s})$ that satisfies $\phi(z) < \mathsf{r}(\Gamma) = 1/\rho(\Gamma)$, while the function has a singularity at the point where $\phi(z) = \mathsf{r}(\Gamma)$, that is, $Q_N(1/z) = \rho(\Gamma)$. The unique solution of this last equation for $z \in (0, \mathsf{s})$ is $z = 1/\cos\frac{\arccos \rho(\Gamma)}{N}$. This must be $\mathsf{r}(\Gamma_{(N)})$, which yields the stated formula for $\rho(\Gamma_{(N)})$.

So we now prove (9.90).

We consider the random variables $\boldsymbol{j}_n = \max\{j : \boldsymbol{t}_j \le n\}$ and decompose, for $x, y \in X \subset X_{(N)}$,

$$G_{(N)}(x, y|z) = \underbrace{\sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \mathsf{Pr}_x[Z_n = y,\ \boldsymbol{j}_n = k]\, z^n}_{W_k(x, y|z)}.$$

When $Z_n = y$ and $\boldsymbol{j}_n = k$ then we cannot have $Z_{\boldsymbol{t}_k} = y' \neq y$, since the random walk cannot visit any new point in X (i.e., other than y') in the time interval $[\boldsymbol{t}_k, n]$. Therefore $Z_{\boldsymbol{t}_k} = y$ and $Z_i \notin X \setminus \{y\}$ when $\boldsymbol{t}_k < i < n$. Using Lemma 9.87 and Exercise 9.88(2),

$$\begin{aligned}
&W_k(x, y|z) \\
&\quad = \sum_{m=0}^{\infty} \sum_{n=m}^{\infty} \mathsf{Pr}_x\big[\boldsymbol{t}_k = m,\ Z_m = Z_n = y,\ Z_i \notin X \setminus \{y\}\ (m < i < n)\big]\, z^n \\
&\quad = \sum_{m=0}^{\infty} \mathsf{Pr}_x\big[\boldsymbol{t}_k = m,\ Z_m = y\big]\, z^m \\
&\qquad \times \sum_{n=m}^{\infty} \mathsf{Pr}_x\big[Z_n = y,\ Z_i \notin X \setminus \{y\}\ (m < i < n) \mid Z_m = y\big]\, z^{n-m} \\
&\quad = p^{(k)}(x, y)\, \phi(z)^k \sum_{n=0}^{\infty} \mathsf{Pr}_y\big[Z_n = y,\ Z_i \notin X \setminus \{y\}\ (0 < i < n)\big]\, z^n \\
&\quad = p^{(k)}(x, y)\, \phi(z)^k\, \psi(z),
\end{aligned}$$

where $p^{(k)}(x, y)$ refers to SRW on Γ. We conclude that

$$G_{(N)}(x, y|z) = \sum_{k=0}^{\infty} p^{(k)}(x, y)\, \phi(z)^k\, \psi(z),$$

which is (9.90).

(b) The proof of the inequality uses the network setting of Section 4.A, and in particular, Proposition 4.11. We write X and $\widetilde{X}$ for the vertex sets, and P and $\widetilde{P}$ for the transition operators (matrices) of SRW on Γ and $\widetilde{\Gamma}$, respectively. We distinguish the inner products on the associated ℓ^2-spaces as well as the Dirichlet norms of (4.32) by a Γ, resp. $\widetilde{\Gamma}$ in the index. Since P is self-adjoint, we have

$$\|P\| = \sup\left\{\frac{(Pf, f)_\Gamma}{(f, f)_\Gamma} : f \in \ell_0(X),\ f \neq 0\right\}.$$

Let $\widetilde{X}$ be the vertex set of $\widetilde{\Gamma}$. We define a map $g\colon \widetilde{X} \to X$ as follows. For $x \in X \subset \widetilde{X}$, we set $g(x) = x$. If $\tilde{x} \in \widetilde{X} \setminus X$ is one of the "new" vertices on

one of the inserted paths of the subdivision, then $g(\tilde{x})$ is the closer one among the two endpoints in X of that path. When $\tilde{x}$ is the midpoint of such an inserted path, we have to choose one of the two endpoints as $g(x)$. Given $f \in \ell_0(X)$, we let $\tilde{f} = f \circ g \in \ell_0(\widetilde{X})$. The following is simple.

9.91 Exercise. Show that

$$(\tilde{f}, \tilde{f})_{\widetilde{\Gamma}} \geq (f, f)_\Gamma \quad \text{and} \quad D_{\widetilde{\Gamma}}(\tilde{f}) = D_\Gamma(f). \qquad \square$$

Having done this exercise, we can resume the proof of the theorem. For arbitrary $f \in \ell_0(X)$,

$$\begin{aligned}(f, f)_\Gamma - (Pf, f)_\Gamma = D_\Gamma(f) = D_{\widetilde{\Gamma}}(\tilde{f}) &= (\tilde{f}, \tilde{f})_{\widetilde{\Gamma}} - (\widetilde{P}\,\tilde{f}, \tilde{f})_{\widetilde{\Gamma}} \\ &\geq (1 - \|\widetilde{P}\|)(\tilde{f}, \tilde{f})_{\widetilde{\Gamma}} \geq (1 - \|\widetilde{P}\|)(f, f)_\Gamma.\end{aligned}$$

That is, for arbitrary $f \in \ell_0(X)$,

$$(Pf, f)_\Gamma \leq \|\widetilde{P}\|(f, f)_\Gamma.$$

We infer that $\rho(\Gamma) = \|P\| \leq \|\widetilde{P}\| = \rho(\widetilde{\Gamma})$.

Since $\Gamma_{(N)}$ is in turn a subdivision of $\widetilde{\Gamma}$, this also yields $\rho(\widetilde{\Gamma}) \leq \rho(\Gamma_{(N)})$. $\square$

Combining the last theorem with Lemma 9.86 and Corollary 9.85 we get the following.

9.92 Theorem. *Let T be a locally finite tree without vertices of degree* 1. *Then SRW on T satisfies $\rho(T) < 1$ if and only if there is a finite upper bound on the lengths of all unbranched paths in T.*

Another typical proof of the last theorem is via comparison of Dirichlet forms and quasi-isometries (rough isometries), which is more elegant but also a bit less elementary. Compare with Theorem 10.9 in [W2]. Here, we also get a numerical bound. When $\deg(x) \geq q + 1$ for all vertices x with $\deg(x) > 2$ and the upper bound on the lengths of unbranched paths is N, then

$$\rho(T) \leq \cos \frac{\arccos \rho(\mathbb{T}_q)}{N}.$$

The following exercise takes us again into the use of the Dirichlet sum of a reversible Markov chain, as at the end of the proof of Theorem 9.89. It provides a tool for estimating $\rho(P)$ for non-simple random walks.

9.93 Exercise. Let T be as in Corollary 9.92. Let P be the transition matrix of a nearest neighbour random walk on the locally finite tree T with associated measure m according to (9.6). Suppose the following holds.

(i) There is $\varepsilon > 0$ such that $\mathsf{m}(x)p(x,y) \geq \varepsilon$ for all $x, y \in T$ with $x \sim y$.

(ii) There is $M < \infty$ such that $\mathsf{m}(x) \leq M \deg(x)$ for every $x \in T$.

Show that the spectral radii of P and of SRW on T are related by

$$1 - \rho(P) \geq \frac{\varepsilon}{M}\big(1 - \rho(T)\big).$$

In particular, $\rho(P) < 1$ when T is as in Theorem 9.92.

[Hint: establish inequalities between the ℓ^2 and Dirichlet norms of finitely supported functions with respect to the two reversible Markov chains. Relate $D(f)/(f,f)$ with the spectral radius.] □

We next want to relate the rate of escape with natural projections of a tree onto the integers. Recall the definition (9.30) of the horocycle function $\mathrm{hor}(x,\xi)$ of a vertex x with respect to the end ξ of a tree T with a chosen root o. As above, we write $v_n(\xi)$ for the n-th vertex on the geodesic ray $\pi(o,\xi)$. Once more, we do *not* require T to be locally finite.

9.94 Proposition. *Let (x_n) be a sequence in a tree T with $x_{n-1} \sim x_n$ for all n. Then the following statements are equivalent.*

(i) *There is a constant $a \geq 0$ (the **rate of escape** of the sequence) such that*

$$\lim_{n\to\infty} \frac{1}{n}|x_n| = a.$$

(ii) *There is an end $\xi \in \partial T$ and a constant $b \geq 0$ such that*

$$\lim_{n\to\infty} \frac{1}{n} d\big(x_n, v_{\lfloor bn \rfloor}(\xi)\big) = 0.$$

(iii) *For some ($\Longleftrightarrow$ every) end $\xi \in \partial T$, there is a constant $a_\xi \in \mathbb{R}$ such that*

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{hor}(x_n, \xi) = a_\xi.$$

Furthermore, we have the following.

(1) *The numbers a, b and a_ξ of the three statements are related by $a = b = |a_\xi|$.*

(2) *If $b > 0$ in statement* (ii), *then $x_n \to \xi$.*

(3) *If $a_\xi < 0$ in statement* (iii), *then $x_n \to \xi$, while if $a_\xi > 0$ then one has $\lim x_n \in \partial T \setminus \{\xi\}$.*

Proof. (i) $\Longrightarrow$ (ii). If $a = 0$ then we set $b = 0$ and see that (ii) holds for arbitrary $\xi \in \partial T$. So suppose $a > 0$. Then

$$|x_n \wedge x_{n+1}| = \frac{1}{2}\big(|x_n| + |x_{n+1}| - d(x_n, x_{n+1})\big) = na + \mathfrak{o}(n) \to \infty.$$

By Lemma 9.16, there is $\xi \in \partial T$ such that $x_n \to \xi$. Thus,

$$|x_n \wedge \xi| = \lim_{m\to\infty} |x_n \wedge x_m| \geq \lim_{m\to\infty} \min\{|x_i \wedge x_{i+1}| : i = n, \dots, m-1\} = na + \mathfrak{o}(n).$$

Since $|x_n \wedge \xi| \leq |x_n| = na + \mathfrak{o}(n)$, we see that $|x_n \wedge \xi| = na + \mathfrak{o}(n)$. Therefore, on one hand

$$d(x_n, x_n \wedge \xi) = |x_n| - |x_n \wedge \xi| = \mathfrak{o}(n),$$

while on the other hand $x_n \wedge \xi$ and $v_{\lfloor an \rfloor}(\xi)$ lie both on $\pi(o, \xi)$, so that

$$d\big(x_n \wedge \xi, v_{\lfloor an \rfloor}(\xi)\big) = \mathfrak{o}(n).$$

We conclude that $d\big(x_n, v_{\lfloor an \rfloor}(\xi)\big) = \mathfrak{o}(n)$, as proposed. (The reader is invited to visualize these arguments by a figure.)

(ii) $\Longrightarrow$ (iii). Let ξ be the end specified in (ii). We have

$$d(x_n, x_n \wedge \xi) = d\big(x_n, \pi(o, \xi)\big) \leq d\big(x_n, v_{\lfloor bn \rfloor}(\xi)\big) = \mathfrak{o}(n).$$

Also,

$$d\big(x_n \wedge \xi, v_{\lfloor bn \rfloor}(\xi)\big) \leq d\big(x_n, v_{\lfloor bn \rfloor}(\xi)\big) = \mathfrak{o}(n).$$

Therefore

$$\begin{aligned} |x_n \wedge \xi| &= |v_{\lfloor bn \rfloor}(\xi)| + \mathfrak{o}(n) = bn + \mathfrak{o}(n), \quad \text{and} \\ \operatorname{hor}(x_n, \xi) &= d(x_n, x_n \wedge \xi) - |x_n \wedge \xi| = -bn + \mathfrak{o}(n). \end{aligned}$$

Thus, statement (iii) holds with respect to ξ with $a_\xi = -b$. It is clear that $x_n \to \xi$ when $b > 0$, and that we do not have to specify ξ when $b = 0$.

To complete this step of the proof, let $\eta \neq \xi$ be another end of T, and assume $b > 0$. Let $y = \xi \wedge \eta$. Then $x_n \wedge \xi \in \pi(y, \xi)$ and $x_n \wedge \eta = y$ for n sufficiently large. For such n,

$$\begin{aligned} \operatorname{hor}(x_n, \eta) &= d(x_n, x_n \wedge \eta) - |x_n \wedge \eta| = d(x_n, y) - |y| \\ &= d(x_n, x_n \wedge \xi) + |x_n \wedge \xi| - 2|y| = bn + \mathfrak{o}(n). \end{aligned}$$

(The reader is again invited to draw a figure.) Thus, statement (iii) holds with respect to η with $a_\eta = b$.

(iii) $\Longrightarrow$ (i). We suppose to have one end ξ such that $h(x_n, \xi)/n \to a_\xi$. Without loss of generality, we may assume that $x_0 = o$. Since for any x,

$$|x| = d(x, x \wedge \xi) + |x \wedge \xi| \quad \text{and} \quad \operatorname{hor}(x, \xi) = d(x, x \wedge \xi) - |x \wedge \xi|,$$

we have

$$\liminf_{n\to\infty} \frac{|x_n|}{n} \geq \lim_{n\to\infty} \frac{|\operatorname{hor}(x_n,\xi)|}{n} = |a_\xi|.$$

Next, note that every point on $\pi(o, x_n)$ is some x_k with $k \leq n$. In particular, $x_n \wedge \xi = x_{k(n)}$ with $k(n) \leq n$.

Consider first the case when $a_\xi < 0$. Then

$$|x_{k(n)}| = d(x_n, x_{k(n)}) - \operatorname{hor}(x_n,\xi) \geq -\operatorname{hor}(x_n,\xi) \to \infty,$$

so that $k(n) \to \infty$,

$$\frac{|x_n|}{n} = \frac{\operatorname{hor}(x_n,\xi) + 2|x_n \wedge \xi|}{n} = \frac{\operatorname{hor}(x_n,\xi)}{n} + \frac{2|\operatorname{hor}(x_{k(n)},\xi)|}{k(n)} \underbrace{\frac{k(n)}{n}}_{\leq 1},$$

which implies

$$\limsup_{n\to\infty} \frac{|x_n|}{n} \leq a_\xi + 2|a_\xi| = |a_\xi|.$$

The same argument also applies when $a_\xi = 0$, regardless of whether $k(n) \to \infty$ or not.

Finally, consider the case when $a_\xi > 0$. We claim that $k(n)$ is bounded. Indeed, if $k(n)$ had a subsequence $k(n')$ tending to ∞, then along that subsequence $\operatorname{hor}(x_{k(n')},\xi) = a_\xi\, k(n') + \mathfrak{o}\big(k(n')\big) \to \infty$, while we must have $\operatorname{hor}(x_{k(n')},\xi) \leq 0$ for all n. Therefore

$$\frac{|x_n|}{n} = \frac{\operatorname{hor}(x_n,\xi)}{n} + \underbrace{\frac{2|x_{k(n)}|}{n}}_{\to 0} \to a_\xi$$

as $n \to \infty$. □

Proposition 9.94 contains information about the rate of escape of a nearest neighbour random walk on a tree T. Namely, $\lim |Z_n|/n$ exists a.s. if and only if $\lim \operatorname{hor}(Z_n,\xi)/n$ exists a.s., and in that case, the former limit is the absolute value of the latter. The process $S_n = \operatorname{hor}(Z_n,\xi)$ has the integer line $\mathbb{Z}$ as its state space, but in general, it is not a Markov chain. We next consider a class of simple examples where this "horocyclic projection" *is* Markovian.

9.95 Example. Choose and fix an end ϖ of the homogeneous tree $\mathbb{T}_q$, and let $\operatorname{hor}(x) = \operatorname{hor}(x,\varpi)$ be the Busemann function with respect to ϖ. Recall the definition (9.32) of the horocycles Hor_k, $k \in \mathbb{Z}$, with respect to that end. Instead of the "radial" picture of $\mathbb{T}_q$ of Figure 26, we can look at it in horocyclic layers with each Hor_k on a horizontal line.

This is the "upper half plane" drawing of the tree. We can think of $\mathbb{T}_q$ as an infinite genealogical tree, where the horocycles are the successive, infinite generations and ϖ is the "mythical ancestor" from which all elements of the population descend. For each k, every element in the k-th generation Hor_k has precisely one predecessor $x^\sim \in \mathrm{Hor}_{k-1}$ and q successors in Hor_{k+1} (vertices y with $y^\sim = x$). Analogously to the notation for x^- as the neighbour of x on $\pi[x, o]$, the predecessor $x^\sim$ is the neighbour of x on $\pi[x, \varpi]$).

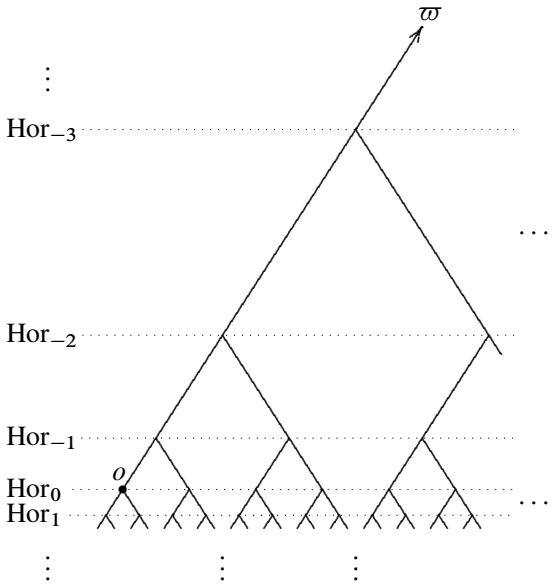

Figure 37

We look at the simplest type of nearest neighbour transition probabilities on $\mathbb{T}_q$ that are compatible with this generation structure. For a fixed parameter $\alpha \in (0, 1)$, define

$$p(x^\sim, x) = \frac{\alpha}{q} \quad \text{and} \quad p(x, x^\sim) = 1 - \alpha, \quad x \in \mathbb{T}_q.$$

For the resulting random walk (Z_n) on the tree, $X_n = \mathrm{hor}(Z_n)$ defines clearly a factor chain in the sense of (1.29). Its state space is $\mathbb{Z}$, and its (nearest neighbour) transition probabilities are

$$\bar{p}(k-1, k) = \alpha \quad \text{and} \quad \bar{p}(k, k-1) = 1 - \alpha, \quad x \in \mathbb{T}_q.$$

This is the infinite drunkard's walk on $\mathbb{Z}$ of Example 3.5 with $p = \alpha$. (Attention: our q here is not $1 - p$, but the branching number of $\mathbb{T}_q$!) If we have the starting point $Z_0 = x$ and $k = \mathrm{hor}(x)$, then we can represent S_n as a sum

$$S_n = k + X_1 + \cdots + X_n,$$

where the X_j are independent with $\mathsf{Pr}[X_j = 1] = \alpha$ and $\mathsf{Pr}[X_j = -1] = 1 - \alpha$. The classical law of large numbers tells us that $S_n/n \to 2\alpha - 1$ almost surely. Proposition 9.94 implies immediately that on the tree,

$$\lim_{n\to\infty} \frac{|Z_n|}{n} = |2\alpha - 1| \quad \mathsf{Pr}_x\text{-almost surely for every } x.$$

We next claim that Pr_x-almost surely for every starting point $x \in \mathbb{T}_q$,

$$\begin{aligned} &Z_n \to Z_\infty \in \partial T \setminus \{\varpi\}, && \text{when } \alpha > 1/2, \quad \text{and} \\ &Z_n \to \varpi, && \text{when } \alpha \le 1/2. \end{aligned}$$

In the first case, Z_∞ is a "true" random variable, while in the second case, the limiting end is deterministic. In fact, when $\alpha \neq 1/2$, this limiting behaviour follows again from Proposition 9.94. The "drift-free" case $\alpha = 1/2$ requires some additional reasoning.

9.96 Exercise. Show that for every $\alpha \in (0, 1)$, the random walk is transient (even for $\alpha = 1/2$).

[Hint: there are various possibilities to verify this. One is to compute the Green function, another is to use the fact the we have finitely many cone types; see Figure 35.] □

Now we can show that $Z_n \to \varpi$ when $\alpha = 1/2$. In this case, $S_n = \mathrm{hor}(Z_n)$ is recurrent on $\mathbb{Z}$. In terms of the tree, this means that Z_n visits Hor_0 infinitely often; there is a random subsequence $(\boldsymbol{n}_k)$ such that $Z_{\boldsymbol{n}_k} \in \mathrm{Hor}_0$. We know from Theorem 9.18 that (Z_n) converges almost surely to a random end. This is also true for the subsequence $(Z_{\boldsymbol{n}_k})$. Since $\mathrm{hor}(Z_{\boldsymbol{n}_k}) = 0$ for all k, this random end cannot be distinct from ϖ.

9.97 Exercise. Compute the Green and Martin kernels for the random walk of Example 9.95. Show that the Dirichlet problem at infinity is solvable if and only if $\alpha > 1/2$. □

We mention that Example 9.95 can be interpreted in terms of products of random affine transformations over a non-archimedean local field (such as the p-adic numbers), compare with Cartwright, Kaimanovich and Woess [9]. In that paper, a more general version of Proposition 9.94 is proved. It goes back to previous work of Kaimanovich [33].

Rate of escape on trees with finitely many cone types

We now consider (T, P) with finitely many cone types, as in Definition 9.74. Consider the associated matrix A over the set of cone types $\mathcal{I}$, as in (9.77).

Here, we assume to have *irreducible cone types*, that is, the matrix A is irreducible. In terms of the tree T, this means that for all $i, j \in \mathcal{I}$, every cone with type i contains a sub-cone with type j. Recall the definition of the functions $F_i(z)$, $i \in \mathcal{I}$, and the associated system (9.76) of algebraic equations.

9.98 Lemma. *If (T, P) has finitely many, irreducible cone types and the random walk is transient, then for every $i \in \mathcal{I}$,*

$$F_i(1) < 1 \quad \textit{and} \quad F_i'(1) < \infty.$$

Proof. We always have $F_i(1) \le 1$ for every $i \in \mathcal{I}$. If $F_j(1) < 1$ for some j then (9.76) yields $F_i(1) < 1$ for every i with $\mathsf{a}(i, j) > 0$. Now irreducibility yields that $F_i(1) < 1$ for all i.

Besides the diagonal matrix $D(z) = \operatorname{diag}\big(F_i(z)\big)_{i\in\mathcal{I}}$ that we introduced in the proof of Theorem 9.78, we consider

$$B = \operatorname{diag}\big(p(i-)\big)_{i\in\mathcal{I}}.$$

Then we can write the formula of Exercise 9.4 as

$$D'(z) = \frac{1}{z^2} D(z)^2 B^{-1} + D(z)^2 A D'(z),$$

where $D'(z)$ refers to the elementwise derivative. We also know that $\rho(Q) = 1$ for the matrix of (9.79), and also that $\rho\big(D(1)\,A\big) = 1$. Proposition 3.42 and Exercise 3.43 imply that $\rho\big(D(z)^2\,A\big) \le \rho\big(D(1)^2\,A\big) < 1$ for each $z \in [0, 1]$. Therefore the inverse matrix $\big(I - D(z)^2\,A\big)^{-1}$ exists, is non-negative, and depends continuously on $z \in [0, 1]$. We deduce that

$$D'(z) = \frac{1}{z^2}\big(I - D(z)^2\,A\big)^{-1} D(z)^2 B^{-1}$$

is finite in each (diagonal) entry for every $z \in [0, 1]$. □

The last lemma implies that the Dirichlet problem at infinity admits solution (Corollary 9.44) and that $\limsup_n d\big(Z_n, \pi(o, Z_\infty)\big)\big/ \log n \le C < \infty$ almost surely (Theorem 9.59).

Now recall the boundary process $(W_k)_{k\ge1}$ of Definition 9.53, the increments $\boldsymbol{\delta}_{k+1} = \boldsymbol{\epsilon}_{k+1} - \boldsymbol{\epsilon}_k$ of the exit times, and the associated Markov chain $\big(W_k, \boldsymbol{\delta}_k\big)_{k\ge1}$. The formula of Corollary 9.57 shows that the transition probabilities of (W_k) depend only on the types of the points. That is, we can build the $\mathcal{I}$-valued factor chain $\big(\iota(W_k)\big)_{k\ge1}$, whose transition matrix is just the matrix $Q = \big(\mathsf{q}(i, j)\big)_{i,j\in\mathcal{I}}$ of (9.79). It is irreducible and finite, so that it admits a unique stationary probability measure

σ on $\mathcal{I}$. Note that $\sigma(i)$ is the asymptotic frequency of cone type i in the boundary process: by the ergodic theorem (Theorem 3.55),

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_i\big(\iota(W_k)\big) = \sigma(i) \quad \mathsf{Pr}_o\text{-almost surely.}$$

The sum appearing on the left hand side is the number of all vertices that have cone type i among the first n points (after o) on the ray $\pi(o, Z_\infty)$. For the following, we write

$$\mathsf{f}^{(n)}(j-) = f^{(n)}(y, y^-), \quad \text{where } y \in T \setminus \{o\},\ \iota(y) = j.$$

9.99 Lemma. *If (T, P) has finitely many, irreducible cone types and the random walk is transient, then the sequence $\big(\iota(W_k), \delta_k\big)_{k\geq 1}$ is a positive recurrent Markov chain with state space $\mathcal{I} \times \mathbb{N}_{\text{odd}}$. Its transition probabilities are*

$$\tilde{\mathsf{q}}\big((i,m),(j,n)\big) = \mathsf{q}(i,j)\mathsf{f}^{(n)}(j-)/F_j(1).$$

Its stationary probability measure is given by

$$\tilde{\sigma}(j,n) = \sum_{i\in\mathcal{I}} \sigma(i)\, \tilde{\mathsf{q}}\big((i,m),(j,n)\big)$$

(*independent of m*).

Proof. We have that $\mathsf{f}^{(n)}(j-) > 0$ for every odd n: if x has type j and y is a forward neighbour of x, then $f^{(2m+1)}(x, x^-) \geq \big(p(x,y)p(y,x)\big)^m p(x, x^-)$. We see that irreducibility of Q implies irreducibility of $\widetilde{Q}$. It is a straightforward exercise to compute that $\tilde{\sigma}$ is a probability measure and that $\tilde{\sigma}\,\widetilde{Q} = \tilde{\sigma}$. □

9.100 Theorem. *If (T, P) has finitely many, irreducible cone types and the random walk is transient, then*

$$\lim_{n\to\infty} \frac{|Z_n|}{n} = \ell \quad \mathsf{Pr}_o\text{-almost surely, where } \ell = 1\Big/ \sum_{i\in\mathcal{I}} \sigma(i) \frac{F_i'(1)}{F_i(1)}.$$

Proof. Consider the projection $g : \mathcal{I} \times \mathbb{N}_{\text{odd}} \to \mathbb{N}$, $(i,n) \mapsto n$. Then

$$\begin{aligned}
\int_{\mathcal{I}\times\mathbb{N}_{\text{odd}}} g\, d\tilde{\sigma} &= \sum_{j\in\mathcal{I}} \sum_{n\in\mathbb{N}_{\text{odd}}} n\, \tilde{\sigma}(j,n) \\
&= \sum_{i,j\in\mathcal{I}} \sum_{n\in\mathbb{N}} \sigma(i)\, \mathsf{q}(i,j)\, n\, \mathsf{f}^{(n)}(j-)/F_j(1) \\
&= \sum_{j\in\mathcal{I}} \sigma(j)\, F_j'(1)/F_j(1) = 1/\ell,
\end{aligned}$$

where ℓ is defined by the first of the two formulas in the statement of the theorem. The ergodic theorem implies that

$$\lim_{k\to\infty} \frac{1}{k} \sum_{m=1}^{k} g\big(\iota(W_m), \boldsymbol{\delta}_m\big) = 1/\ell \quad \mathsf{Pr}_o\text{-almost surely.}$$

The sum on the left hand side is $\boldsymbol{\epsilon}_k - \boldsymbol{\epsilon}_0$, and we get that $k/\boldsymbol{\epsilon}_k \to \ell$ Pr_o-almost surely.

We now define the integer valued random variables

$$\boldsymbol{k}(n) = \max\{k : \boldsymbol{\epsilon}_k \le n\}.$$

Then $\boldsymbol{k}(n) \to \infty$ Pr_o-almost surely, $Z_{\boldsymbol{\epsilon}_{\boldsymbol{k}(n)}} = W_{\boldsymbol{k}(n)}$, and since $Z_n \in T_{W_{\boldsymbol{k}(n)}}$, the cone rooted at the random vertex $W_{\boldsymbol{k}(n)}$,

$$\begin{aligned} |Z_n| - \boldsymbol{k}(n) = |Z_n| - |W_{\boldsymbol{k}(n)}| &= d(Z_n, W_{\boldsymbol{k}(n)}) \\ &\le n - \boldsymbol{\epsilon}_{\boldsymbol{k}(n)} < \boldsymbol{\epsilon}_{\boldsymbol{k}(n)+1} - \boldsymbol{\epsilon}_{\boldsymbol{k}(n)} = \boldsymbol{\delta}_{\boldsymbol{k}(n)+1}. \end{aligned}$$

(The middle "$\le$" follows from the nearest neighbour property.) Now

$$0 < \frac{\boldsymbol{\epsilon}_{\boldsymbol{k}(n)+1} - n}{n} \le \frac{\boldsymbol{\epsilon}_{\boldsymbol{k}(n)+1} - \boldsymbol{\epsilon}_{\boldsymbol{k}(n)}}{n} \le \frac{\boldsymbol{\epsilon}_{\boldsymbol{k}(n)+1} - \boldsymbol{\epsilon}_{\boldsymbol{k}(n)}}{\boldsymbol{\epsilon}_{\boldsymbol{k}(n)}} \to 0 \quad \mathsf{Pr}_o\text{-almost surely,}$$

as $n \to \infty$, because $\boldsymbol{\epsilon}_{k+1}/\boldsymbol{\epsilon}_k \to 1$ as $k \to \infty$. Consequently,

$$\frac{|Z_n| - \boldsymbol{k}(n)}{n} \to 0 \quad \text{and} \quad \frac{\boldsymbol{\epsilon}_{\boldsymbol{k}(n)}}{n} \to 1.$$

We conclude that

$$\frac{|Z_n|}{n} = \frac{|Z_n| - \boldsymbol{k}(n)}{n} + \frac{\boldsymbol{k}(n)}{\boldsymbol{\epsilon}_{\boldsymbol{k}(n)}} \frac{\boldsymbol{\epsilon}_{\boldsymbol{k}(n)}}{n} \to \ell \quad \mathsf{Pr}_o\text{-almost surely,}$$

as $n \to \infty$. □

The last theorem is taken from [44].

9.101 Exercise. Show that the formula for the rate of escape in Theorem 9.100 can be rewritten as

$$\ell = 1 \Big/ \sum_{i\in\mathcal{I}} \sigma(i) \frac{F_i(1)}{\mathsf{p}(i-)\big(1 - F_i(1)\big)}.$$ □

9.102 Example. As an application, let us compute the rate of escape for the random walk of Example 9.47. When $\mathcal{I}$ is finite, there are finitely many cone types,

$\mathsf{d}(i,j) = 1$ and $\mathsf{p}(i,j) = p_j$ when $j \neq i$, and $\mathsf{p}(i-) = p_i$. The numbers $F_i(1)$ are determined by (9.49), and

$$\mathsf{q}(i,j) = F_i(1)\frac{p_j}{p_i}\frac{1-F_j(1)}{1-F_i(1)}.$$

We compute the stationary probability measure σ for Q. Introducing the auxiliary term

$$H = \sum_{j\in\mathcal{I}} \frac{\sigma(j)}{p_j}\frac{F_j(1)}{1-F_j(1)},$$

the equation $\sigma\, Q = \sigma$ becomes

$$\left(H - \frac{\sigma(i)}{p_i}\frac{F_i(1)}{1-F_i(1)}\right) p_i\big(1-F_i(1)\big) = \sigma(i), \quad i \in \mathcal{I}.$$

Therefore

$$\sigma(i) = H\, p_i \frac{1-F_i(1)}{1+F_i(1)} = \frac{H}{G(1)}\frac{F_i(1)}{\big(1+F_i(1)\big)^2}.$$

The last identity holds because $p_i\big(1-F_i(1)^2\big) = F_i(1)/G(1)$ by (9.48), which also implies that

$$\sum_{i\in\mathcal{I}} \frac{F_i(1)}{1+F_i(1)} = G(1) - \sum_{i\in\mathcal{I}} p_i\, F_i(1)\, G(1) = 1$$

by (1.34). Now, since σ is a probability measure,

$$\sigma(i) = \frac{F_i(1)}{\big(1+F_i(1)\big)^2} \Big/ \sum_{j\in\mathcal{I}} \frac{F_j(1)}{\big(1+F_j(1)\big)^2}.$$

Combining all those identities with the formula of Exercise 9.101, we compute with some final effort that the rate of escape is

$$\ell = \frac{1}{G(1)} \sum_{i\in\mathcal{I}} \frac{F_i(1)}{\big(1+F_i(1)\big)^2}.$$

I first learnt this specific formula from T. Steger in the 1980s.

Solutions of all exercises

Exercises of Chapter 1

Exercise 1.22. This is straightforward, but formalizing it correctly needs some work. Working with the trajectory space, we first assume that A is a cylinder in $\mathcal{A}_m$, that is, $A = \mathsf{C}(a_0, \dots, a_{m-1}, x)$ with $a_0, \dots, a_{m-1} \in X$. Then, by the Markov property as stated in Definition 1.7, combined with (1.19),

$$\begin{aligned}
&\mathsf{Pr}_\nu[Z_n = y \mid A] \\
&\quad = \sum_{x_{m+1},\dots,x_{n-1} \in X} \mathsf{Pr}_\nu \left[\begin{array}{c|c} Z_n = y, Z_j = x_j & Z_m = x, Z_i = a_i \\ (j = m+1, \dots, n-1) & (i = 0, \dots, m-1) \end{array} \right] \\
&\quad = \sum_{x_{m+1},\dots,x_{n-1} \in X} \mathsf{Pr}_x[Z_{n-m} = y, Z_j = x_j \ (j = 1, \dots, n-m-1)] \\
&\quad = \mathsf{Pr}_x[Z_{n-m} = y] = p^{(n-m)}(x, y).
\end{aligned}$$

Next, a general set $A \in \mathcal{A}_m$ with the stated properties must be a finite or countable disjoint union of cylinders $\mathsf{C}_i \in \mathcal{A}_m$, $i \in \mathcal{I}$, each of the form $\mathsf{C}(a_0, \dots, a_{m-1}, x)$ with certain $a_0, \dots, a_{m-1} \in X$. Therefore

$$\mathsf{Pr}_\nu[Z_n = y \mid \mathsf{C}_i] = p^{(n-m)}(x, y) \quad \text{whenever } \mathsf{Pr}_\nu(\mathsf{C}_i) > 0.$$

So by the rules of conditional probability,

$$\begin{aligned}
\mathsf{Pr}_\nu[Z_n = y \mid A] &= \frac{1}{\mathsf{Pr}_\nu(A)} \sum_i \mathsf{Pr}_\nu([Z_n = y] \cap \mathsf{C}_i) \\
&= \frac{1}{\mathsf{Pr}_\nu(A)} \sum_i \mathsf{Pr}_\nu[Z_n = y \mid \mathsf{C}_i] \, \mathsf{Pr}_\nu(\mathsf{C}_i) \\
&= \frac{1}{\mathsf{Pr}_\nu(A)} \sum_i p^{(n-m)}(x, y) \, \mathsf{Pr}_\nu(\mathsf{C}_i) = p^{(n-m)}(x, y).
\end{aligned}$$

The last statement of the exercise is a special case of the first one. □

Exercise 1.25. Fix $n \in \mathbb{N}_0$ and let $x_0, \dots, x_{n+1} \in X$. For any $k \in \mathbb{N}_0$, the event $A_k = [\boldsymbol{t} = k, Z_{k+j} = x_j \ (j = 0, \dots, n)]$ is in $\mathcal{A}_{k+n}$. Therefore we can apply

Exercise 1.22:

$$\begin{aligned}
&\mathsf{Pr}_\nu[Z_{\boldsymbol{t}+n+1} = x_{n+1},\ Z_{\boldsymbol{t}+j} = x_j\ (j = 0,\dots,n)] \\
&\quad = \sum_{k=0}^{\infty} \mathsf{Pr}_\nu[Z_{k+n+1} = x_{n+1},\ Z_{k+j} = x_j\ (j = 0,\dots,n),\ \boldsymbol{t} = k] \\
&\quad = \sum_{k=0}^{\infty} \mathsf{Pr}_\nu[Z_{k+n+1} = x_{n+1} \mid A_k]\ \mathsf{Pr}_\nu(A_k) \\
&\quad = \sum_{k=0}^{\infty} p(x_n, x_{n+1})\ \mathsf{Pr}_\nu(A_k) \\
&\quad = p(x_n, x_{n+1})\ \mathsf{Pr}_\nu[Z_{\boldsymbol{t}+j} = x_j\ (j = 0,\dots,n)].
\end{aligned}$$

Dividing by $\mathsf{Pr}_\nu[Z_{\boldsymbol{t}+j} = x_j\ (j = 0,\dots,n)]$, we see that the statements of the exercise are true. (The computation of the initial distribution is straightforward.) □

Exercise 1.31. We start with the "if" part, which has already been outlined. We *define* $\overline{\mathsf{Pr}}$ by $\overline{\mathsf{Pr}}(\bar{A}) = \mathsf{Pr}_\nu\big(\pi^{-1}(\bar{A})\big)$ for $\bar{A} \in \bar{\mathcal{A}}$ and have to show that under this probability measure, the sequence of n-th projections $\overline{Z}_n \colon \overline{\Omega} \to \overline{X}$ is a Markov chain with the proposed transition probabilities $\bar{p}(\bar{x}, \bar{y})$ and initial distribution $\bar{\nu}$. For this, we only need to show that $\overline{\mathsf{Pr}} = \overline{\mathsf{Pr}}_{\bar{\nu}}$, and equality has only to be checked for cylinder sets because of the uniqueness of the extended measure. Thus, let $\bar{A} = \mathsf{C}(\bar{x}_0, \dots, \bar{x}_n) \in \bar{\mathcal{A}}$. Then

$$\pi^{-1}(\bar{A}) = \biguplus_{x_0 \in \bar{x}_0, \dots, x_n \in \bar{x}_n} \mathsf{C}(x_0, \dots, x_n),$$

and (inductively)

$$\begin{aligned}
\overline{\mathsf{Pr}}(\bar{A}) &= \sum_{x_0 \in \bar{x}_0, \dots, x_n \in \bar{x}_n} \nu(x_0)\, p(x_0, x_1) \cdots p(x_{n-1}, x_n) \\
&= \sum_{x_0 \in \bar{x}_0} \nu(x_0) \sum_{x_1 \in \bar{x}_1} p(x_0, x_1) \cdots \underbrace{\sum_{x_n \in \bar{x}_n} p(x_{n-1}, x_n)}_{\bar{p}(\bar{x}_{n-1}, \bar{x}_n)} \\
&= \bar{\nu}(\bar{x}_0)\, \bar{p}(\bar{x}_0, \bar{x}_1) \cdots \bar{p}(\bar{x}_{n-1}, \bar{x}_n) = \overline{\mathsf{Pr}}_{\bar{\nu}}(\bar{A}).
\end{aligned}$$

For the "only if", suppose that $\big(\pi(Z_n)\big)$ is (for every starting point $x \in X$) a Markov chain on $\overline{X}$ with transition probabilities $\bar{p}(\bar{x}, \bar{y})$. Then, given two classes $\bar{x}, \bar{y} \in \overline{X}$, for every $x_0 \in \bar{x}$ we have

$$\bar{p}(\bar{x}, \bar{y}) = \mathsf{Pr}_{x_0}[\pi(Z_1) = \bar{y} | \pi(Z_0) = \bar{x}] = \mathsf{Pr}_{x_0}[Z_1 \in \bar{y}] = \sum_{y \in \bar{y}} p(x_0, y),$$

as required. □

Exercise 1.41. We decompose with respect to the first step. For $n \geq 1$

$$u^{(n)}(x,x) = \sum_{y\in X} \mathsf{Pr}_x[\boldsymbol{t}^x = n,\ Z_1 = y] = \sum_{y\in X} p(x,y)\ \mathsf{Pr}_x[\boldsymbol{t}^x = n \mid Z_1 = y]$$
$$= \sum_{y\in X} p(x,y)\ \mathsf{Pr}_y[\boldsymbol{s}^x = n-1] = \sum_{y\in X} p(x,y)\ f^{(n-1)}(y,x).$$

Multiplying by z^n and summing over all n, we get the formula of Theorem 1.38 (c).

Exercise 1.44. This works precisely as in the proof of Proposition 1.43.

$$\begin{aligned} u^{(n)}(x,x) &= \mathsf{Pr}_x[\boldsymbol{t}^x = n] \\ &\geq \mathsf{Pr}_x[\boldsymbol{t}^x = n,\ \boldsymbol{s}^y \leq n] \\ &= \sum_{k=0}^{n} \mathsf{Pr}_x[\boldsymbol{s}^y = k]\ \mathsf{Pr}_x[\boldsymbol{t}^x = n \mid \boldsymbol{s}^y = k] \\ &= \sum_{k=0}^{n} f^{(k)}(x,y)\ f^{(n-k)}(y,x), \end{aligned}$$

since $u^{(n-k)}(y,x) = f^{(n-k)}(y,x)$ when $x \neq y$. □

Exercise 1.45. We have

$$\begin{aligned} \mathsf{E}_x(\boldsymbol{s}^y \mid \boldsymbol{s}^y < \infty) &= \sum_{n=1}^{\infty} n\ \mathsf{Pr}_x[\boldsymbol{s}^y = n \mid \boldsymbol{s}^y < \infty] \\ &= \sum_{n=1}^{\infty} n\ \frac{\mathsf{Pr}_x[\boldsymbol{s}^y = n]}{Pr_x[\boldsymbol{s}^y < \infty]} \\ &= \sum_{n=1}^{\infty} \frac{n\ f^{(n)}(x,y)}{F(x,y|1)} = \frac{F'(x,y|1)}{F(x,y|1)}. \end{aligned}$$

More precisely, in the case when $z = 1$ is on the boundary of the disk of convergence of $F(x,y|z)$, that is, when $\mathsf{s}(x,y) = 1$, we can apply the theorem of Abel: $F'(j,N|1)$ has to be replaced with $F'(j,N|1-)$, the left limit along the real line. (Actually, we just use the monotone convergence theorem, interpreting $\sum_n n\ f^{(n)}(x,y)z^n$ as an integral with respect to the counting measure and letting $z \to 1$ from below.)

If w is a cut point between x and y, then Proposition 1.43 (b) implies

$$\frac{F'(x,y|z)}{F(x,y|z)} = \frac{F'(x,w|z)}{F(x,w|z)} + \frac{F'(w,y|z)}{F(x,y|z)},$$

and the formula follows by setting $z = 1$. □

Exercise 1.47. (a) If $p = q = 1/2$ then we can rewrite the formula for

$$F'(j, N|z)/F(j, N|z)$$

of Example 1.46 as

$$\frac{F'(j, N|z)}{F(j, N|z)} = \frac{2}{z^2\lambda_1(z)} \frac{1}{\alpha(z) - 1} \left(N \frac{\alpha(z)^N + 1}{\alpha(z)^N - 1} - j \frac{\alpha(z)^j + 1}{\alpha(z)^j - 1} \right).$$

We have $\lambda_1(1) = \alpha(1) = 1$. Letting $z \to 1-$, we see that

$$\mathsf{E}_j(\boldsymbol{s}^N \mid \boldsymbol{s}^N < \infty) = \lim_{\alpha \to 1} \frac{2}{\alpha - 1} \left(N \frac{\alpha^N + 1}{\alpha^N - 1} - j \frac{\alpha^j + 1}{\alpha^j - 1} \right).$$

This can be calculated in different ways. For example, using that

$$(\alpha^N - 1)(\alpha^j - 1) \sim (\alpha - 1)^2 \, jN \quad \text{as} \quad \alpha \to 1,$$

we compute

$$\begin{aligned}
&\frac{2}{\alpha - 1} \left(N \frac{\alpha^N + 1}{\alpha^N - 1} - j \frac{\alpha^j + 1}{\alpha^j - 1} \right) \\
&\quad \sim \frac{2}{(\alpha - 1)^3 \, jN} \Big((N - j)(\alpha^{N+j} - 1) - (N + j)(\alpha^N - \alpha^j) \Big) \\
&\quad = \frac{2}{(\alpha - 1)^2 \, jN} \left((N - j) \sum_{k=0}^{N+j-1} \alpha^k - (N + j) \sum_{k=j}^{N-1} \alpha^k \right) \\
&\quad = \frac{2}{(\alpha - 1)^2 \, jN} \left((N - j) \sum_{k=1}^{N+j-1} (\alpha^k - 1) - (N + j) \sum_{k=j}^{N-1} (\alpha^k - 1) \right) \\
&\quad = \frac{2}{(\alpha - 1) \, jN} \left((N - j) \sum_{k=1}^{N+j-1} \sum_{m=0}^{k-1} \alpha^m - (N + j) \sum_{k=j}^{N-1} \sum_{m=0}^{k-1} \alpha^m \right) \\
&\quad = \frac{2}{(\alpha - 1) \, jN} \left((N - j) \sum_{k=1}^{N+j-1} \sum_{m=1}^{k-1} (\alpha^m - 1) - (N + j) \sum_{k=j}^{N-1} \sum_{m=1}^{k-1} (\alpha^m - 1) \right) \\
&\quad \sim \frac{2}{jN} \left((N - j) \sum_{k=1}^{N+j-1} \sum_{m=1}^{k-1} m - (N + j) \sum_{k=j}^{N-1} \sum_{m=1}^{k-1} m \right) \\
&\quad = \frac{N^2 - j^2}{3}.
\end{aligned}$$

(b) If the state 0 is absorbing, then the linear recursion

$$F(j, N|z) = qz \, F(j - 1, N|z) + pz \, F(j + 1, N|z), \quad j = 1, \ldots, N - 1,$$

remains the same, as well as the boundary value $F(N, N|z) = 1$. The boundary value at 0 has to be replaced with the equation $F(0, N|z) = z\, F(1, N|z)$. Again, for $|z| < 1/2\sqrt{pq}$, the solution has the form

$$F(j, N|z) = a \cdot \lambda_1(z)^j + b \cdot \lambda_2(z)^j,$$

but now

$$a + b = a \cdot z\, \lambda_1(z) + b \cdot z\, \lambda_2(z) \quad \text{and} \quad a \cdot \lambda_1(z)^N + b \cdot \lambda_2(z)^N = 1.$$

Solving in a and b yields

$$F(j, N|z) = \frac{\big(1 - z\lambda_1(z)\big)\lambda_2(z)^j - \big(1 - z\lambda_2(z)\big)\lambda_1(z)^j}{\big(1 - z\lambda_1(z)\big)\lambda_2(z)^N - \big(1 - z\lambda_2(z)\big)\lambda_1(z)^N}.$$

In particular, $\mathsf{Pr}_j[s^N < \infty] = F(j, N|1) = 1$, as it must be (why ?), and $\mathsf{E}_j(s^N) = F'(j, N|1-)$ is computed similarly as before. We omit those final details. □

Exercise 1.48 Formally, we have $\mathcal{G}(z) = (I - zP)^{-1}$, and this is completely justified when X is finite. Thus,

$$\mathcal{G}_a(z) = \big(I - z(aI + (1-a)P)\big)^{-1} = \tfrac{1}{1-az}\big(I - \tfrac{z-az}{1-az}P\big)^{-1} = \tfrac{1}{1-az}\,\mathcal{G}\big(\tfrac{z-az}{1-az}\big).$$

For general X, we can argue by approximation with respect to finite subsets, or as follows:

$$\begin{aligned} G_a(x, y|z) &= \sum_{n=0}^{\infty} z^n \sum_{k=0}^{n} \binom{n}{k} a^{n-k}\,(1-a)^k\, p^{(k)}(x, y) \\ &= \sum_{k=0}^{\infty} \left(\frac{1-a}{a}\right)^k p^{(k)}(x, y) \sum_{n=k}^{\infty} \binom{n}{k} (az)^n. \end{aligned}$$

Since

$$\sum_{n=k}^{\infty} \binom{n}{k} (az)^n = \frac{1}{1-az}\Big(\frac{az}{1-az}\Big)^k \quad \text{for } |az| < 1,$$

the formula follows. □

Exercise 1.55. Let $\pi = [x = x_0, x_1, \ldots, x_n = x]$ be a path in $\Pi(x, x) \setminus \{[x]\}$. Since $x_n = x$, we have $1 \le k \le n$, where $k = \min\{j \ge 1 : x_j = x\}$. Then

$$\pi_1 = [x_0, x_1, \ldots, x_k] \in \Pi_\bullet(x, x), \quad \pi_2 = [x_k, x_{k+1}, \ldots, x_n] \in \Pi(x, x),$$
$$\text{and} \quad \pi = \pi_1 \circ \pi_2.$$

This decomposition is unique, which proves the first formula. We deduce

$$\begin{aligned} G(x,x|z) &= \mathsf{w}\big(\Pi(x,x)|z\big) \\ &= \mathsf{w}\big([x]|z\big) + \mathsf{w}\big(\Pi_\bullet(x,x)\circ\Pi(x,x)|z\big) \\ &= 1 + \mathsf{w}\big(\Pi_\bullet(x,x)|z\big)\,\mathsf{w}\big(\Pi(x,x)|z\big) \\ &= 1 + U(x,x|z)G(x,x|z). \end{aligned}$$

Analogously, if $\pi = [x = x_0, x_1, \dots, x_n = y]$ is a path in $\Pi(x,y)$ then we let $m = \min\{j \ge 0 : x_j = y\}$. We get

$$\pi_1 = [x_0, x_1, \dots, x_m] \in \Pi_\circ(x,y), \quad \pi_2 = [x_m, x_{k+1}, \dots, x_n] \in \Pi(y,y), \\ \text{and} \quad \pi = \pi_1 \circ \pi_2.$$

Once more, the decomposition is unique, and

$$\begin{aligned} G(x,y|z) &= \mathsf{w}\big(\Pi(x,y)|z\big) \\ &= \mathsf{w}\big(\Pi_\circ(x,y)\circ\Pi(y,y)|z\big) \\ &= F(x,y|z)G(y,y|z). \end{aligned}$$

Regarding Theorem 1.38 (c), every $\pi \in \Pi_\bullet(x,x)$ has a unique decomposition $\pi = [x,y]\circ\pi_2$, where $[x,y] \in E\big(\Gamma(P)\big)$ and $\pi_2 \in \Pi_\circ(y,x)$. That is,

$$\Pi_\bullet(x,x) = \biguplus_{y:[x,y]\in E\left(\Gamma(P)\right)} [x,y]\circ\Pi_\circ(y,x),$$

which yields the formula for $U(x,x|z)$. In the same way, let $y \ne x$ and $\pi \in \Pi_\circ(x,y)$. Then either $\pi = [x,y]$, which is possible only when $[x,y] \in E\big(\Gamma(P)\big)$, or else $\pi = [x,w]\circ\pi_2$, where $[x,w] \in E\big(\Gamma(P)\big)$ and $\pi_2 \in \Pi_\circ(w,y)$. Thus, noting that $\Pi_\circ(y,y) = \{[y]\}$,

$$\Pi_\circ(x,y) = \biguplus_{w:[x,w]\in E(\Gamma(P))} [x,w]\circ\Pi_\circ(w,y), \quad y \ne x,$$

which yields Theorem 1.38 (d) in terms of weights of paths. □

Exercises of Chapter 2

Exercise 2.6. (a) $\Longrightarrow$ (b). If C is essential and $x \to y$ then $C = C(x) \to C(y)$ in the partial order of irreducible classes. Since C is maximal in that order, $C(y) = C$, whence $y \in C$.

(b) $\Longrightarrow$ (c). Let $x \in C$ and $x \to y$. By assumption, $y \in C$, that is, $x \leftrightarrow y$.

(c) $\Longrightarrow$ (a). Let $C(y)$ be any irreducible class such that $C \to C(y)$ in the partial order of irreducible classes. Choose $x \in C$. Then $x \to y$. By assumption, also $y \to x$, whence $C(y) = C(x) = C$. Thus, C is maximal in the partial order. □

Exercise 2.12. Let (X_1, P_1) and (X_2, Y_2) be Markov chains.

An *isomorphism* between (X_1, P_1) and (X_2, P_2) is a bijection $\varphi\colon X_1 \to X_2$ such that

$$p_2(\varphi x_1, \varphi y_1) = p_1(x_1, y_1) \quad \text{for all } x_1, y_1 \in X_1.$$

Note that this definition does not require that the matrix P is stochastic. An *automorphism* of (X, P) is an isomorphism of (X, P) onto itself. Then we have the following obvious fact.

If there is an automorphism φ of (X, P) such that $\varphi x = x'$ and $\varphi y = y'$ then $G(x, y|z) = G(x', y'|z)$.

Next, let $y \neq x$ and define the *branch*

$$B_{y,x} = \{w \in X : x \to w \to y\}.$$

We let $P_{y,x}$ be the restriction of the matrix P to that branch.

We say that the branches $B_{y,x}$ and $B_{y',x'}$ are isomorphic, if there is an isomorphism φ of $(B_{y,x}, P_{y,x})$ onto $(B_{y',x'}, P_{y',x'})$ such that $\varphi x = x'$ and $\varphi y = y'$. Again, after formulating this definition, the following fact is obvious.

If $B_{y,x}$ and $B_{y',x'}$ are isomorphic, then $F(x, y|z) = F(x', y'|z)$.

Indeed, before reaching y for the first time, the Markov chain starting at x can never leave $B_{y,x}$. □

Exercise 2.26. If (X, P) is irreducible and aperiodic and $x, y \in X$, then there is $k = k_{x,y}$ such that $p^{(k)}(x, y) > 0$. Also, By Lemma 2.22, there is m_x such that $p^{(m)}(x, x) > 0$ for all $m \geq m_x$. Therefore $p^{(qn)}(x, y) > 0$ for all q with $qn - k_{x,y} \geq m_x$. □

Exercise 2.34. For $C = \{☼, ❋\}$, the truncated transition matrix is

$$P_C = \begin{pmatrix} 0 & 1/2 \\ 1/4 & 1/2 \end{pmatrix}.$$

Therefore, using (1.36),

$$G_C(❋, ❋|z) = 1/\det(I - zP_C) = 8/(8 - 4z - z^2).$$

Its radius of convergence is the root of the denominator with smallest absolute value. The spectral radius is the inverse of that root, that is, the largest eigenvalue. We get $\rho(P_C) = (1 + \sqrt{3})/4$. □

Exercises of Chapter 3

Exercise 3.7. If y is transient, then

$$\sum_{n=0}^{\infty} \mathsf{Pr}_\nu[Z_n = y] = \sum_x \nu(x) G(x,y)$$
$$= \sum_x \nu(x) \underbrace{F(x,y)}_{\le 1} G(y,y) \le G(y,y) < \infty.$$

In particular, $\mathsf{Pr}_\nu[Z_n = y] \to 0$. □

Exercise 3.13. If C is a finite, essential class and $x \in C$, then by (3.11),

$$\sum_{y \in C} F(x,y|z) \frac{1-z}{1-U(y,y|z)} = 1.$$

Suppose that some and hence every element of C is transient. Then $U(y,y) < 1$ for all $y \in C$, so that finiteness of C yields

$$\lim_{z \to 1-} \sum_{y \in C} F(x,y|z) \frac{1-z}{1-U(y,y|z)} = 0,$$

a contradiction. □

Exercise 3.14. We always think of $\bar{X}$ as a partition of X. We realize the factor chain $(\bar{Z}_n)$ on the trajectory space of (X,P) (instead of its own trajectory space), which is legitimate by Exercise 1.31: $\bar{Z}_n = \pi(Z_n)$. Since $x \in \bar{x}$, the first visit of Z_n in the set $\bar{x}$ cannot occur after the first visit in the point $x \in \bar{x}$. That is, $\boldsymbol{t}^{\bar{x}} \le \boldsymbol{t}^x$.

Therefore, if x is recurrent, $\mathsf{Pr}_x[\boldsymbol{t}^x < \infty] = 1$, then also $\mathsf{Pr}_{\bar{x}}[\boldsymbol{t}^{\bar{x}} < \infty] = \mathsf{Pr}_x[\boldsymbol{t}^{\bar{x}} < \infty] = 1$. In the same way, if x is positive recurrent, then also

$$\mathsf{E}_{\bar{x}}(\boldsymbol{t}^{\bar{x}}) = \mathsf{E}_x(\boldsymbol{t}^{\bar{x}}) \le \mathsf{E}_x(\boldsymbol{t}^x) < \infty.$$ □

Exercise 3.18. We have

$$\nu(X) \ge \sum_{y \in X} \nu P(y) = \sum_{x \in X} \sum_{y \in X} \nu(x) p(x,y) = \nu(X).$$

Thus, we cannot have $\nu P(y) < \nu(y)$ for any $y \in X$. □

Exercise 3.22. Let $\varepsilon = \nu(y)/2$. We can find a finite subset A_ε of X such that $\nu(X \setminus A_\varepsilon) < \varepsilon$. As in the proof of Theorem 3.19, for $0 < z < 1$,

$$2\varepsilon = \nu(y) \le \sum_{x \in A_\varepsilon} \nu(x)\, F(x,y|z) \frac{1-z}{1-U(y,y|z)} + \varepsilon.$$

Therefore

$$\sum_{x\in A_\varepsilon} \nu(x)\, F(x,y|z)\frac{1-z}{1-U(y,y|z)} \geq \varepsilon.$$

Suppose first that $U(y,y|1) < 1$. Then the left hand side in the last inequality tends to 0 as $z \to 1-$, a contradiction. Therefore y must be recurrent, and we can apply de l'Hospital's rule. We find

$$\sum_{x\in A_\varepsilon} \nu(x)\, F(x,y)\frac{1}{U'(y,y|1-)} \geq \varepsilon.$$

Thus, $U'(y,y|1-) < \infty$, and y is positive recurrent. □

Exercise 3.24. We write $C_d = C_0$. Let $i \in \{1,\dots,d\}$. If $y \in C_i$ and $p(x,y) > 0$ then $x \in C_{i-1}$. Therefore

$$\mathsf{m}_C(C_i) = \sum_{y\in C_i}\sum_{x\in C_{i-1}} \mathsf{m}_C(x)p(x,y) = \sum_{x\in C_{i-1}} \mathsf{m}_C(x) \underbrace{\sum_{y\in C_i} p(x,y)}_{=1} = \mathsf{m}(C_{i-1}).$$

Thus, $\mathsf{m}_C(C_i) = 1/d$. Now write m_i for the stationary probability measure of P_C^d on C_i. We claim that

$$\mathsf{m}_i(x) = \begin{cases} d\cdot \mathsf{m}_C(x), & \text{if } x \in C_i,\\ 0, & \text{otherwise.}\end{cases}$$

By the above, this is a probability measure on C_i. If $y \in C_i$, then

$$\mathsf{m}_i(y) = d\sum_{x\in X} \mathsf{m}_C(x) \underbrace{p^{(d)}(x,y)}_{>0 \text{ only if } x\in C_i} = \mathsf{m}_i P^d(y),$$

as claimed. The same result can also be deduced by observing that

$$\boldsymbol{t}^x_{P^d} = \boldsymbol{t}^x_P/d, \quad \text{if } Z_0 = x,$$

where the indices P^d and P refer to the respective Markov chains. □

Exercise 3.27. (We omit the figure.) Since $p(k,1) = 1/2$ for all k, while every other column of the transition matrix contains a 0, we have $\tau(P) = 1/2$. Theorem 3.26 implies that the Markov chain is positive recurrent. The stationary probability measure must satisfy

$$\mathsf{m}(1) = \sum_{k\in\mathbb{N}} \mathsf{m}(k)\, p(k,1) = 1/2 \quad \text{and} \quad \mathsf{m}(k+1) = \mathsf{m}(k)\, p(k,k+1) = \mathsf{m}(k)/2.$$

Thus, $\mathsf{m}(k) = 2^{-k}$, and for every $j \in \mathbb{N}$,

$$\sum_{k\in\mathbb{N}} |p^{(n)}(j,k) - 2^{-k}| \le 2^{-n+1}. \qquad \square$$

Exercise 3.32. Set $f(x) = \nu(x)/\mathsf{m}(x)$. Then

$$\hat{P} f(x) = \sum_y \frac{\mathsf{m}(y)p(y,x)}{\mathsf{m}(x)} \frac{\nu(y)}{\mathsf{m}(y)} = \frac{\nu P(x)}{\mathsf{m}(x)}.$$

Thus $\hat{P} f = f$ if and only if $\nu P = \nu$. $\square$

Exercise 3.41. In addition to the normalization $\sum_x h(x)\nu(x) = 1$ of the right and left ρ-eigenvectors of A, we can also normalize such that $\sum_x \nu(x) = 1$. With those conditions, the eigenvectors are unique.

Consider first the y-column $\hat{a}_\rho(\cdot, y)$ of $\hat{A}_\rho$. Since it is a right ρ-eigenvector of A, there must be a constant $c(y)$ depending on y such that $\hat{a}_\rho(x, y) = c(y)h(x)$.

On the other hand, the x-row $\hat{a}_\rho(x, \cdot) = c(\cdot)h(x)$ is a left ρ-eigenvector of A, and since $h(x) > 0$, also $c(\cdot)$ is a left ρ-eigenvector. Therefore there is a constant α such that $c(y) = \alpha \cdot \nu(y)$ for all y. $\square$

Exercise 3.43. With

$$p(x,y) = \frac{a(x,y)h(y)}{\rho(A)h(x)} \quad \text{and} \quad q(x,y) = \frac{b(x,y)h(y)}{\rho(A)h(x)}$$

we have that P is stochastic and $q(x, y) \le p(x, y)$ for all $x, y \in X$. If Q is also stochastic, then

$$\sum_y \underbrace{p(x,y) - q(x,y)}_{\ge 0} = 0$$

for every x, whence $P = Q$.

Now suppose that B is irreducible and $B \ne A$. Then Q is irreducible and strictly substochastic in at least one row. By Proposition 2.31, $\rho(Q) < 1$. But $\rho(B) = \rho(Q)/\rho(A)$, so that $\rho(B) < \rho(A)$. Finally, if B is not irreducible and dominated by A, let $C = \frac{1}{2}(A + B)$. Then C is irreducible, dominates B and is dominated by A. Furthermore, there must be $x, y \in X$ such that $a(x, y) > 0$ and $b(x, y) = 0$, so that $c(x, y) < a(x, y)$. By Proposition 3.42 we have that $\max\{|\lambda| : \lambda \in \operatorname{spec}(B)\} \le \rho(C)$, and by the above, $\rho(C) < \rho(A)$ strictly. $\square$

Exercise 3.45. We just have to check that every step of the proof that

$$\rho(A) = \min\{t > 0 \mid \text{there is } g\colon X \to (0, \infty) \text{ with } Ag \le t \cdot g\}$$

remains valid even when X is an infinite (countable) set. This is indeed the case, with some care where the Heine–Borel theorem is used. Here one can use the classical diagonal method for extracting a subsequence $(g_{k(m)})$ that converges pointwise. $\square$

Exercise 3.52. (1) Let $(Z_n)_{n\geq 0}$ be the Markov chain on X with transition matrix P. We know from Theorem 2.24 that Z_n can return to the starting point only when n is a multiple of d. Thus,

$$\boldsymbol{t}_{P^d}^x = \boldsymbol{t}_P^x/d, \quad \text{if } Z_0 = x,$$

where the indices P^d and P refer to the respective Markov chains, as mentioned above. Therefore $\mathsf{E}_x(\boldsymbol{t}_P^x) < \infty$ if and only if $\mathsf{E}_x(\boldsymbol{t}_{P^d}^x) < \infty$.

(2) This follows by applying Theorem 3.48 to the irreducible, aperiodic Markov chain $(C_i, P_{C_i}^d)$.

(3) Let $r = j - i$ if $j \geq i$ and $r = j - i + d$ if $j < i$. Then we know from Theorem 2.24 that for $w \in C_j$ we have $p^{(m)}(x, w) > 0$ only if $d \,|\, m - r$. We can write

$$p^{(nd+r)}(x, y) = \sum_{w \in C_j} p^{(r)}(x, w)\, p^{(nd)}(w, y).$$

By (2), $p^{(nd)}(w, y) \to d \cdot \mathsf{m}(y)$ for all $w \in C_j$. Since $\sum_{w\in C_j} p^{(r)}(x, w) = 1$, we get (by dominated convergence) that for $x \in C_i$ and $y \in C_j$,

$$\lim_{n\to\infty} p^{(nd+j-i)}(x, y) = d \cdot \mathsf{m}(y). \qquad \square$$

Exercise 3.59. The first identity is clear when $x = y$, since $L(x, x|z) = 1$. Suppose that $x \neq y$. Then

$$\begin{aligned} p^{(n)}(x, y) &= \sum_{k=0}^{n-1} \mathsf{Pr}_x[Z_k = x,\ Z_j \neq x\ (j = k+1, \dots, n),\ Z_n = y] \\ &= \sum_{k=0}^{n-1} p^{(k)}(x, x)\, \ell^{(n-k)}(x, y) = \sum_{k=0}^{n} p^{(k)}(x, x)\, \ell^{(n-k)}(x, y), \end{aligned}$$

since $\ell^{(0)}(x, y) = 0$. The formula now follows once more from the product rule for power series.

For the second formula, we decompose with respect to the last step: for $n \geq 2$,

$$\begin{aligned} \mathsf{Pr}_x[\boldsymbol{t}^x = n] &= \sum_{y\neq x} \mathsf{Pr}_x[Z_j \neq x\ (j = 1, \dots, n-1),\ Z_{n-1} = y,\ Z_n = x] \\ &= \sum_{y\neq x} \ell^{(n-1)}(x, y)\, p(y, x) = \sum_{y} \ell^{(n-1)}(x, y)\, p(y, x), \end{aligned}$$

since $\ell^{(n-1)}(x, x) = 0$. The identity $\mathsf{Pr}_x[\boldsymbol{t}^x = n] = \sum_y \ell^{(n-1)}(x, y)\, p(y, x)$ also remains valid when $n = 1$. Multiplying by z^n and summing over all n, we get the proposed formula.

The proof of the third formula is completely analogous to the previous one. □

Exercise 3.60. By Theorem 1.38 and Exercise 3.59,

$$G(x,y|z)G(y,x|z) = \begin{cases} F(x,y|z)G(y,y|z)F(y,x|z)G(x,x|z) \\ G(x,x|z)L(x,y|z)G(y,y|z)L(y,x|z). \end{cases}$$

For real $z \in (0,1)$, we can divide by $G(x,x|z)G(y,y|z)$ and get the proposed identity

$$L(x,y|z)L(y,x|z) = F(x,y|z)F(y,x|z).$$

Since $x \leftrightarrow y$, we have $L(x,y|1-) > 0$ and $L(y,x|1-) > 0$. But

$$L(x,y|1-)L(y,x|1-) = F(x,y)F(y,x) \le 1,$$

so that we must have $L(x,y|1-) < \infty$ and $L(y,x|1-) < \infty$. □

Exercise 3.61. If x is a recurrent state and $L(x,y|z) > 0$ for $z > 0$ then $x \leftrightarrow y$ (since x is essential). Using the suggestion and Exercise 3.59,

$$\sum_{y\in X} G(x,x|z)\, L(x,y|z) = \frac{1}{1-z}, \quad \text{or equivalently,}$$

$$\sum_{y\in X} L(x,y|z) = \frac{1-U(x,x|z)}{1-z}.$$

Letting $z \to 1-$, the formula follows. □

Exercise 3.64. Following the suggestion, we consider an arbitrary initial distribution ν and choose a state $x \in X$. Then we define $\boldsymbol{t}_0 = \boldsymbol{s}^x$ and, as before, $\boldsymbol{t}_k = \inf\{n > \boldsymbol{t}_{k-1} : Z_n = x\}$ for $k \ge 1$. Then we let $Y_0 = \sum_{n=0}^{\boldsymbol{s}^x} f(Z_n)$, while Y_k for $k \ge 1$ remains as before. The Y_k, $k \ge 0$, are independent, and for $k \ge 1$, they all have the same distribution. In particular,

$$\frac{1}{k} S_{\boldsymbol{t}_k}(f) = \underbrace{\frac{1}{k} Y_0}_{\to 0} + \frac{1}{k} \sum_{j=1}^{k} Y_j \to \frac{1}{\mathsf{m}(x)} \int_X f \, d\mathsf{m}.$$

The proof now proceeds precisely as before. □

Exercise 3.65. Let $\widetilde{Z}_n = (Z_n, Z_{n+1})$. It is immediate that this is a Markov chain. For $(x,y),(v,w) \in \widetilde{X}$, write $\tilde{p}\big((x,y),(v,w)\big)$ for its transition probabilities. Then

$$\tilde{p}\big((x,y),(v,w)\big) = \begin{cases} p(y,w), & \text{if } v = y, \\ 0, & \text{otherwise.} \end{cases}$$

By induction (prove this in detail!),

$$\tilde{p}^{(n)}\big((x,y),(v,w)\big) = p^{(n-1)}(y,v)\,p(v,w).$$

From this formula, it becomes apparent that the new Markov chain inherits irreducibility and aperiodicity from the old one. The limit theorem for (X, P) implies that

$$\tilde{p}^{(n)}\big((x,y),(v,w)\big) \to \mathsf{m}(v)\,p(v,w), \quad as\ n \to \infty.$$

Thus, the stationary probability distribution for the new chain is given by $\tilde{\mathsf{m}}(x,y) = \mathsf{m}(x)\,p(x,y)$. We can apply the ergodic theorem with $f = \mathbf{1}_{(x,y)}$ to $(\widetilde{Z}_n)$ and get

$$\frac{1}{N}\sum_{n=0}^{N-1} \boldsymbol{v}_n^x \boldsymbol{v}_{n+1}^y = \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{1}_{(x,y)}(\widetilde{Z}_n) \to \tilde{\mathsf{m}}(x,y) = \mathsf{m}(x)\,p(x,y)$$

almost surely. □

Exercise 3.70. This is proved exactly as Theorem 3.9. For $0 < z < \mathsf{r}$,

$$\frac{1-U(x,x|z)}{1-U(y,y|z)} = \frac{G(y,y|z)}{G(x,x|z)} \geq p^{(l)}(y,x)\,p^{(k)}(x,y)\,z^{k+l},$$

where k and l are chosen such that $p^{(k)}(x,y) > 0$ and $p^{(l)}(y,x) > 0$. Therefore, in the ρ-recurrent case, once again via de l'Hospital's rule,

$$\frac{U'(x,x|\mathsf{r}-)}{U(y,y|\mathsf{r}-)} = \lim_{z\to 1-} \frac{G(y,y|z)}{G(x,x|z)} \geq p^{(l)}(y,x)\,p^{(k)}(x,y)\,\mathsf{r}^{k+l} > 0. \qquad \square$$

Exercise 3.71. The function $z \mapsto U(x,x|z)$ is monotone increasing and differentiable for $z \in (0, \mathsf{s})$. It follows from Proposition 2.28 that r must be the unique solution in $(0, \mathsf{s})$ of the equation $U(x,x|z) = 1$. In particular, $U'(x,x|\mathsf{r}) < \infty$. □

Exercises of Chapter 4

Exercise 4.7. For the norm of ∇,

$$\begin{aligned}
\langle \nabla f, \nabla f\rangle &= \frac{1}{2}\sum_{e\in E} \frac{1}{r(e)}\big(f(e^+) - f(e^-)\big)^2 \\
&\leq \sum_{e\in E} \frac{1}{r(e)}\big(f(e^+)^2 + f(e^-)^2\big) \\
&= \sum_{x\in X}\Big(\sum_{e\,:\,e^+=x} \frac{1}{r(e)} + \sum_{e\,:\,e^-=x} \frac{1}{r(e)}\Big) f(x)^2 \\
&= \sum_{x\in X} 2\,\mathsf{m}(x)\,f(x)^2 = 2\,(f,f).
\end{aligned}$$

Regarding the adjoint operator, we have for every finitely supported $f : X \to \mathbb{R}$

$$\begin{aligned}\langle \nabla f, \phi \rangle &= \frac{1}{2} \sum_{e \in E} \big(f(e^+) - f(e^-)\big)\phi(e) \\ &= \frac{1}{2} \sum_{x \in X} \Big(\sum_{e\,:\,e^+ = x} f(x)\phi(e) - \sum_{e\,:\,e^- = x} f(x)\phi(e) \Big) \\ &= \sum_{x \in X} f(x) \sum_{e\,:\,e^+ = x} \phi(e) \qquad \big(\text{since } -\phi(e) = \phi(\check{e})\big) \\ &= (f, g), \qquad \text{where } g(x) = \frac{1}{\mathsf{m}(x)} \sum_{e\,:\,e^+ = x} \phi(e).\end{aligned}$$

Therefore $\nabla^* \phi = g$. □

Exercise 4.10. It is again sufficient to verify this for finitely supported f. Since $\frac{1}{\mathsf{m}(x) r(e)} = p(x, y)$ for $e = [x, y]$, we have

$$\begin{aligned}\nabla^*(\nabla f)(x) &= \frac{1}{\mathsf{m}(x)} \sum_{e\,:\,e^+ = x} \frac{f(e^+) - f(e^-)}{r(e)} \\ &= \sum_{y\,:\,[x,y] \in E} \big(f(x) - f(y)\big) p(x, y) = f(x) - Pf(x),\end{aligned}$$

as proposed. □

Exercise 4.17. If $Qf = \lambda \cdot f$ then $Pf = \big(a + (1-a)\lambda\big) \cdot f$. Therefore

$$\lambda_{\min}(P) = a + (1-a)\lambda_{\min}(Q) \geq a - (1-a),$$

with equality precisely when $\lambda_{\min}(Q) = -1$.

If $P^k \geq a \cdot I$ then we can write $P^k = a \cdot I + (1-a) \cdot Q$, where Q is stochastic. By the first part, $\lambda_{\min}(P^k) \geq -1 + 2a$. When k is odd, $\lambda_{\min}(P^k) = \big(\lambda_{\min}(P)\big)^k$. □

Exercise 4.20. We have

$$\begin{aligned}\mathsf{Pr}[Y_1 Y_2 = x] &= \sum_y \mathsf{Pr}[Y_1 = y,\ Y_1 Y_2 = x] = \sum_y \mathsf{Pr}[Y_1 = y,\ Y_2 = y^{-1}x] \\ &= \sum_y \mathsf{Pr}[Y_1 = y]\ \mathsf{Pr}[Y_2 = y^{-1}x] = \sum_y \mu_1(y)\, \mu_2(y^{-1}x).\end{aligned}$$

This is $\mu_1 * \mu_2(x)$. □

Exercise 4.25. We use induction on d. Thus, we need to indicate the dimension in the notation $f_{\boldsymbol{\varepsilon}} = f_{\boldsymbol{\varepsilon}}^{(d)}$.

For $d = 1$, linear independence is immediate. Suppose linear independence holds for $d-1$. For $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_d) \in \mathcal{E}_d$ we let $\boldsymbol{\varepsilon}' = (\varepsilon_1, \dots, \varepsilon_{d-1}) \in \mathcal{E}_{d-1}$, so that $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}', \varepsilon_d)$. Analogously, if $\boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{Z}_2^d$, then we write $\boldsymbol{x} = (\boldsymbol{x}', x_d)$. Now suppose that for a set of real coefficients $c_{\boldsymbol{\varepsilon}}$, we have

$$\sum_{\boldsymbol{\varepsilon} \in \mathcal{E}_d} c_{\boldsymbol{\varepsilon}} \cdot f_{\boldsymbol{\varepsilon}}^{(d)}(\boldsymbol{x}) = 0 \quad \text{for all } \boldsymbol{x} \in \mathbb{Z}_2^d.$$

Since $f_{\boldsymbol{\varepsilon}}^{(d)}(\boldsymbol{x}) = \boldsymbol{\varepsilon}_d^{x_d} f_{\boldsymbol{\varepsilon}'}^{(d-1)}(\boldsymbol{x}')$, we can rewrite this as

$$\sum_{\boldsymbol{\varepsilon}' \in \mathcal{E}_{d-1}} \big(c_{(\boldsymbol{\varepsilon}',1)} + (-1)^{x_d} c_{(\boldsymbol{\varepsilon}',-1)}\big) \cdot f_{\boldsymbol{\varepsilon}'}^{(d-1)}(\boldsymbol{x}') = 0 \quad \text{for all } \boldsymbol{x}' \in \mathbb{Z}_2^{d-1},\ x_d \in \mathbb{Z}_2.$$

By the induction hypothesis, with $x_d = 0$ and $= 1$, respectively,

$$c_{(\boldsymbol{\varepsilon}',1)} + c_{(\boldsymbol{\varepsilon}',-1)} = 0 \quad \text{and} \quad c_{(\boldsymbol{\varepsilon}',1)} - c_{(\boldsymbol{\varepsilon}',-1)} = 0 \quad \text{for all } \boldsymbol{\varepsilon}' \in \mathcal{E}_{d-1},\ x_d \in \mathbb{Z}_2.$$

Therefore $c_{(\boldsymbol{\varepsilon}',1)} = c_{(\boldsymbol{\varepsilon}',-1)} = 0$, that is $c_{\boldsymbol{\varepsilon}} = 0$ for all $\boldsymbol{\varepsilon} \in \mathcal{E}_d$, completing the induction argument. □

Exercise 4.28. We define the measure $\bar{\mathsf{m}}$ on $\bar{X}$ by $\bar{\mathsf{m}}(\bar{x}) = \mathsf{m}(\bar{x}) = \sum_{x \in \bar{x}} \mathsf{m}(x)$. Then

$$\bar{\mathsf{m}}(\bar{x})\, \bar{p}(\bar{x}, \bar{y}) = \sum_{x \in \bar{x}} \mathsf{m}(x) \sum_{y \in \bar{y}} p(x, y) = \sum_{y \in \bar{y}} \sum_{x \in \bar{x}} \mathsf{m}(y)\, p(y, x) = \bar{\mathsf{m}}(\bar{y})\, \bar{p}(\bar{y}, \bar{x}).$$

□

Exercise 4.30. Stirling's formula says that

$$N! \sim (N/e)^N \sqrt{2N\pi}, \quad \text{as } N \to \infty.$$

Therefore

$$\binom{N}{N/2} \sim \frac{(N/e)^N \sqrt{2N\pi}}{\big(N/(2e)\big)^N N\pi} = 2^{N+\frac{1}{2}} \Big/ \sqrt{N\pi}$$

and

$$\frac{N}{4} \log \frac{2^N}{\binom{N}{N/2}} \sim \frac{N}{4} \log \sqrt{\frac{N\pi}{2}} = \frac{N}{8} \big(\log N + \log(\pi/2)\big) \sim \frac{N \log N}{8},$$

as $N \to \infty$. □

Exercise 4.41. In the Ehrenfest model, the state space is $X = \{0, \dots, N\}$, the edge set is $E = \{[j-1, j],\ [j, j-1] : j = 1, \dots, N\}$,

$$\mathsf{m}(j) = \frac{\binom{N}{j}}{2^N}, \quad \text{and} \quad r([j-1, j]) = \frac{2^N}{\binom{N-1}{j-1}}.$$

For $i < k$, we have the obvious choices $\pi_{i,k} = [i, i+1, \dots, k]$ and $\pi_{k,i} = [k, k-1, \dots, i]$. We get

$$\kappa_1([j, j-1]) = \kappa_1([j-1, j]) = \frac{1}{2^N \binom{N-1}{j-1}} \underbrace{\sum_{i=0}^{j-1} \sum_{k=j}^{N} (k-i) \binom{N}{i} \binom{N}{k}}_{= S_1 - S_2}.$$

We compute

$$S_1 = \sum_{i=0}^{j-1} \binom{N}{i} \sum_{k=j}^{N} k \binom{N}{k} = N \sum_{i=0}^{j-1} \binom{N}{i} \sum_{k=j}^{N} \binom{N-1}{k-1}$$
$$= N \sum_{i=0}^{j-1} \binom{N}{i} \left(2^{N-1} - \sum_{k=0}^{j-2} \binom{N-1}{k} \right)$$

and

$$S_2 = \sum_{i=0}^{j-1} i \binom{N}{i} \sum_{k=j}^{N} \binom{N}{k} = N \sum_{i=1}^{j-1} \binom{N-1}{i-1} \sum_{k=j}^{N} \binom{N}{k}$$
$$= N \sum_{i=0}^{j-2} \binom{N-1}{i} \left(2^N - \sum_{k=0}^{j-1} \binom{N}{k} \right).$$

Therefore, using $\binom{N}{i} - \binom{N-1}{i} = \binom{N-1}{i-1}$,

$$\begin{aligned} S_1 - S_2 &= N\, 2^{N-1} \sum_{i=0}^{j-1} \binom{N}{i} - N\, 2^N \sum_{i=0}^{j-2} \binom{N-1}{i} \\ &= N\, 2^{N-1} \sum_{i=0}^{j-1} \binom{N}{i} - N\, 2^{N-1} \sum_{i=0}^{j-1} \binom{N-1}{i} \\ &\quad - N\, 2^{N-1} \sum_{i=0}^{j-2} \binom{N-1}{i} + N\, 2^{N-1} \binom{N-1}{j-1} \\ &= N\, 2^{N-1} \sum_{i=1}^{j-1} \binom{N-1}{i-1} - N\, 2^{N-1} \sum_{i=0}^{j-2} \binom{N-1}{i} \\ &\quad + N\, 2^{N-1} \binom{N-1}{j-1} \\ &= N\, 2^{N-1} \binom{N-1}{j-1}. \end{aligned}$$

We conclude that $\kappa_1(e) = N/2$ for each edge. Thus, our estimate for the spectral gap becomes $1 - \lambda_1 \geq 2/N$, which misses the true value $2/(N+1)$ by the (asymptotically as $N \to \infty$) negligible factor $N/(N+1)$.[1] □

Exercise 4.45. We have to take care of the case when $\deg(x) = \infty$. Following the suggestion,

$$\begin{aligned}\sum_{y:[x,y]\in E} |f(x) - f(y)|\, a(x,y) &= \sum_{y:[x,y]\in E} \big(|f(x) - f(y)|\,\sqrt{a(x,y)}\,\big)\sqrt{a(x,y)} \\ &\leq \sum_{y:[x,y]\in E} \big(f(x) - f(y)\big)^2\, a(x,y) \sum_{y:[x,y]\in E} a(x,y) \\ &\leq D(f)\,\mathsf{m}(x),\end{aligned}$$

which is finite, since $f \in \mathcal{D}(\mathcal{N})$. Now both sums

$$\sum_{y:[x,y]\in E} \big(f(x) - f(y)\big)\, a(x,y) = \mathsf{m}(x)\,\nabla^*(\nabla f)(x) \quad \text{and}$$

$$\sum_{y:[x,y]\in E} f(x)\, a(x,y) = \mathsf{m}(x)\, f(x)$$

are absolutely convergent. Therefore we may separate the differences:

$$\begin{aligned}\mathsf{m}(x)\,\nabla^*(\nabla f)(x) &= \sum_{y:[x,y]\in E} f(x)\, a(x,y) - \sum_{y:[x,y]\in E} f(y)\, a(x,y) \\ &= \mathsf{m}(x)\big(f(x) - Pf(x)\big).\end{aligned}$$

This is the proposed formula. □

Exercise 4.52. Let $\phi = -\nabla G(\cdot, x)/\mathsf{m}(x)$. Since $G(\cdot, x) \in \mathcal{D}(\mathcal{N})$, we can apply Exercise 4.45 to compute the power of ϕ:

$$\begin{aligned}\langle \phi, \phi \rangle &= \frac{1}{\mathsf{m}(x)^2}\big(G(\cdot,x), \nabla^*\nabla G(\cdot,x)\big) \\ &= \frac{1}{\mathsf{m}(x)^2}\big(G(\cdot,x), \underbrace{(I-P)G(\cdot,x)}_{=\mathbf{1}_x}\big) = \frac{G(x,x)}{\mathsf{m}(x)}.\end{aligned}$$

From the proof of Theorem 4.51, we see that for finite $A \subset X$,

$$\frac{\mathsf{m}(x)}{G_A(x,x)} \geq \mathrm{cap}(x) \geq \frac{1}{\langle \phi, \phi \rangle} = \frac{\mathsf{m}(x)}{G(x,x)}.$$

If we let $A \nearrow X$, we get the proposed formula for $\mathrm{cap}(x)$.

[1] The author acknowledges input from Theresia Eisenkölbl regarding the computation of $S_1 - S_2$.

Finally, we justify writing "*The* flow from x to ∞ with...": the set of all flows from x to ∞ with input 1 is closed and convex in $\ell^2_\sharp(E,r)$, so that it has indeed a *unique* element with minimal norm. □

Exercise 4.54. Suppose that ϕ' is a flow from $x \in X'$ to ∞ with input 1 and finite power in the subnetwork $\mathcal{N}' = (X', E', r')$ of $\mathcal{N} = (X, E, r)$. We can extend ϕ' to a function on E by setting

$$\phi'(e) = \begin{cases} \phi'(e), & \text{if } e \in E', \\ 0, & \text{if } e \in E \setminus E'. \end{cases}$$

Then ϕ is a flow from $x \in X$ to ∞ with input 1. Its power in $\ell^2_\sharp(E,r)$ reduces to

$$\sum_{e \in E'} \phi'(e)^2 \, r(e) \le \sum_{e \in E'} \phi'(e)^2 \, r'(e),$$

which is finite by assumption. Thus, also $\mathcal{N}$ is transient. □

Exercise 4.58. (a) The edge set is $\{[k, k \pm 1] : k \in \mathbb{Z}\}$. All resistances are $= 1$. Suppose that ϕ is a flow with input 1 from state 0 to ∞. Let $\phi([0,1]) = \alpha$. Then $\phi([0,-1]) = 1 - \alpha$. Furthermore, we must have $\phi([k, k+1]) = \alpha$ and $\phi([-k, -k-1]) = 1 - \alpha$ for all $k \ge 0$. Therefore

$$\langle \phi, \phi \rangle = \left(\alpha^2 + (1-\alpha)^2\right) \cdot \infty.$$

Thus, every flow from 1 to ∞ with input 1 has infinite power, so that SRW is recurrent.

(b) We use the classical formula "number of favourable cases divided by the number of possible cases", which is justified when all cases are equally likely. Here, a single case is a trajectory (path) of length $2n$ that starts at state 0. There are 2^{2n} such trajectories, each of which has probability $1/2^{2n}$.

If the walker has to be back at state 0 at the $2n$-th step, then of those $2n$ steps, n must go right and the other n must go left. Thus, we have to select from the time interval $\{1, \dots, 2n\}$ the subset of those n instants when the walker goes right. We conclude that the number of favourable cases is $\binom{2n}{n}$. This yields the proposed formula for $p^{(2n)}(0,0)$.

(c) We use (3.6) with $p = q = 1/2$ and get $U(0,0|z) = 1 - \sqrt{1-z^2}$. By binomial expansion,

$$G(0,0|z) = \frac{1}{\sqrt{1-z^2}} = (1-z^2)^{-1/2} = \sum_{n=0}^{\infty} (-1)^n \binom{-1/2}{n} z^{2n}.$$

The coefficient of z^{2n} is

$$p^{(2n)}(0,0) = (-1)^n \binom{-1/2}{n} = \frac{1}{4^n} \binom{2n}{n}.$$

(d) The asymptotic evaluation has already been carried out in Exercise 4.30. We see that $\sum_n p^{(n)}(0,0) = \infty$. □

Exercise 4.69. The irreducible class of 0 is the additive semigroup S generated by $\mathrm{supp}(\mu)$. Since it is essential, $k \to 0$ for every $k \in S$. But this means that $-k \in S$, so that S is a subgroup of $\mathbb{Z}$. Under the stated assumptions, $S \neq \{0\}$. But then there must be $k_0 \in \mathbb{N}$ such that $S = \{k_0\, n : n \in \mathbb{Z}\}$. In particular, S must have finite index in $\mathbb{Z}$. The irreducible (whence essential) classes are the cosets of S in $\mathbb{Z}$, so that there are only finitely many classes. □

Exercises of Chapter 5

Exercise 5.14. We modify the Markov chain by restricting the state space to $\{0, 1, \ldots, k-1+i\}$. We make the point $k-1+i$ absorbing, while all other transition probabilities within that set remain the same. Then the generating functions $F_i(k+1, k-1|z)$, $F_i(k, k-1|z)$ and $F_{i-1}(k+1, k|z)$ with respect to the original chain coincide with the functions $F(k+1, k-1|z)$, $F(k, k-1|z)$ and $F(k+1, k|z)$ of the modified chain. Since k is a cut point between $k+1$ and $k-1$, the formula now follows from Proposition1.43 (b), applied to the modified chain. □

Exercise 5.20. In the null-recurrent case ($S = \infty$), we know that $\mathsf{E}_k(\boldsymbol{s}^0) = \infty$. Let us therefore suppose that our birth-and-death chain is positive recurrent. Then $\mathsf{E}_0(\boldsymbol{t}^0) = \sum_{j=0}^{\infty} \mathsf{m}(j)$. We can the proceed precisely as in the computations that led to Proposition 5.8 to find that

$$\mathsf{E}_k(\boldsymbol{s}^0) = \sum_{i=0}^{k-1} \sum_{j=i+1}^{\infty} \frac{p_{i+1} \cdots p_{j-1}}{q_{i+1} \cdots q_j}.$$ □

Exercise 5.31. We have for $n \geq 1$

$$\mathsf{Pr}[\boldsymbol{t} > n] = 1 - \mathsf{Pr}[M_n = 0] = 1 - g_n(0).$$

Using the mean value theorem of differential calculus and monotonicity of $f'(z)$,

$$1 - g_n(0) = f(1) - f\big(g_{n-1}(0)\big) = \big(1 - g_{n-1}(0)\big)\, f'(\xi) \leq \big(1 - g_{n-1}(0)\big)\, \bar{\mu},$$

where $g_{n-1}(0) < \xi < 1$. Inductively,

$$1 - g_n(0) \leq \big(1 - \mu(0)\big)\, \bar{\mu}^{n-1}.$$

Therefore

$$\mathsf{E}(\boldsymbol{t}) = \sum_{n=0}^{\infty} \mathsf{Pr}[\boldsymbol{t} > n] \leq 1 + \big(1 - \mu(0)\big) \sum_{n=1}^{\infty} \bar{\mu}^{n-1} = 1 + \frac{1 - \mu(0)}{1 - \bar{\mu}}.$$ □

Exercise 5.35. With probability $\mu(\infty)$, the first generation is infinite, that is, $\Sigma \subset \mathsf{T}$. But then

$$\mathsf{Pr}[N_j < \infty \text{ for all } j \in \Sigma \mid \Sigma \subset \mathsf{T}] = 0;$$

at least one of the members of the first generation has infinitely many children. Repeating the argument, at least one of the latter must have infinitely many children, and so on (inductively). We conclude that conditionally upon the event $[\Sigma \subset \mathsf{T}]$, all generations are infinite:

$$\mathsf{Pr}[M_n = \infty \text{ for all } n \geq 1 \mid M_1 = \infty] = 1. \qquad \square$$

Exercise 5.33. (a) If the ancestor ϵ has no offspring, $M_1 = 0$, then $|\mathsf{T}| = 1$. Otherwise,

$$|\mathsf{T}| = 1 + |\mathsf{T}_1| + |\mathsf{T}_2| + \cdots + |\mathsf{T}_{M_1}|,$$

where T_j is the subtree of T rooted at the j-th offspring of the ancestor, $j \in \Sigma$. Since these trees are i.i.d., we get

$$\mathsf{E}(z^{|\mathsf{T}|}) = \mu(0)\, z + \sum_{k=1}^{\infty} \mathsf{Pr}[M_1 = k]\, \mathsf{E}\big(z^{1+|\mathsf{T}_1|+|\mathsf{T}_2|+\cdots+|\mathsf{T}_k|}\big) = \sum_{k=1}^{\infty} \mu(k)\, z\, g(z)^k.$$

(b) We have $f(z) = q + p\, z^2$. We get a quadratic equation for $g(z)$. Among the two solutions, the right one must be monotone increasing for $z > 0$ near 0. Therefore

$$g(z) = \frac{1}{2pz}\Big(1 - \sqrt{1 - 4pqz^2}\Big).$$

It is now a straightforward task to use the binomial theorem to expand $g(z)$ into a power series. The series' coefficients are the desired probabilities.

(c) In this case, $f(z) = q/(1 - zp)$, and

$$g(z) = \frac{1}{2p}\Big(1 - \sqrt{1 - 4pqz}\Big).$$

The computation of the probabilities $\mathsf{Pr}[\,|\mathsf{T}| = k]$ is almost the same as in (b) and also left to the reader. $\square$

Exercise 5.37. When $p > 1/2$, the drunkard's walk is transient. Therefore $Z_n \to \infty$ almost surely. We infer that

$$\mathsf{Pr}_0[\boldsymbol{t}^0 = \infty,\ Z_n \to \infty] = \mathsf{Pr}_0[\boldsymbol{t}^0 = \infty] = 1 - F(1, 0) > 0.$$

On the event $[\boldsymbol{t}^0 = \infty,\ Z_n \to \infty]$, every edge is crossed. Therefore

$$\mathsf{Pr}[M_k > 0 \text{ for all } k] > 0.$$

If (M_k) were a Galton–Watson process then it would be supercritical. But the average number of upcrossings of $[k+1, k+2]$ that take place after an upcrossing of $[k, k+1]$ and before another upcrossing of $[k, k+1]$ may occur (i.e., before the drunkard returns from state $k+1$ to state k) coincides with

$$\begin{aligned}
\mathsf{E}_0(M_1) &= \sum_{n=1}^{\infty} \mathsf{Pr}_0[Z_{n-1} = 1,\ Z_n = 2,\ n < \boldsymbol{t}^0] \\
&= \sum_{n=1}^{\infty} \mathsf{Pr}_1[Z_{n-1} = 1,\ Z_n = 2,\ Z_i \neq 0\ (i < n)] \\
&= \sum_{n=1}^{\infty} \frac{p}{q}\ \mathsf{Pr}_1[Z_{n-1} = 1,\ Z_n = 0,\ Z_i \neq 0\ (i < n)] \\
&= \frac{p}{q}\, F(1,0) = 1.
\end{aligned}$$

But if the average offspring number is 1 (and the offspring distribution is non-degenerate, which must be the case here) then the Galton–Watson process dies out almost surely, a contradiction. □

Exercise 5.41. (a) Following the suggestion, we decompose

$$M_n^y = \sum_{u \in \Sigma^n} \mathbf{1}_{[u \in \mathsf{T},\ Z_u = y]}.$$

Therefore, using Lemma 5.40,

$$\begin{aligned}
\mathsf{E}_x^{\mathrm{BMC}}(M_n^y) &= \sum_{u \in \Sigma^n} \mathsf{Pr}_x^{\mathrm{BMC}}[u \in \mathsf{T},\ Z_u = y] \\
&= \sum_{j_1, \dots, j_n \in \Sigma} p^{(n)}(x, y) \prod_{i=1}^{n} \mu[j_i, \infty) = p^{(n)}(x, y) \prod_{i=1}^{n} \sum_{j_i \in \Sigma} \mu[j_i, \infty).
\end{aligned}$$

Since $\sum_{j \in \Sigma} \mu[j, \infty) = \bar{\mu}$, the formula follows.

(b) By part (a), we have $\mathsf{E}_x^{\mathrm{BMC}}(M_k^y) > 0$. Therefore $\mathsf{Pr}_x^{\mathrm{BMC}}[M_k^y \neq 0] > 0$. □

Exercise 5.44. To conclude, we have to show that when there are x, y such that $H^{\mathrm{BMC}}(x, y) = 1$ then $H^{\mathrm{BMC}}(x, y') = 1$ for every $y' \in X$. If $\mathsf{Pr}_x^{\mathrm{BMC}}[M^y = \infty] = 1$, then

$$\mathsf{Pr}_x^{\mathrm{BMC}}[M^{y'} < \infty] = \mathsf{Pr}_x^{\mathrm{BMC}}[M^y = \infty,\ M^{y'} < \infty],$$

which is $= 0$ by Lemma 5.43. □

Exercise 5.47. Let $u(0) = \epsilon, u(i) = j_1 \cdots j_i, i = 1, \ldots, n-1$ be the predecessors of $u(n) = v$ in Σ^*, and let $\boldsymbol{\tau}$ be the subtree of Σ^* spanned by the $u(i)$. Let $y_1, \ldots, y_{n-1} \in X$ and set $y_n = x$. Then we can consider the event $\mathsf{D}(\boldsymbol{\tau}; a_u, u \in \boldsymbol{\tau})$, where $a_{u(j)} = y_j$, and (5.38) becomes

$$\mathsf{Pr}_x^{\mathrm{BMC}}[v \in \mathsf{T},\ Z_v = x,\ Z_{u_i} = y_k\ (i = 1, \ldots, n-1)] = \prod_{k=1}^{n} \mu[j_i, \infty)\, p\big(y_{i-1}, y_i\big).$$

Now we can compute

$$\begin{aligned}\mathsf{Pr}_x^{\mathrm{BMC}}[v \in \mathsf{W}_1^x] &= \sum_{y_1,\ldots,y_{n-1} \in X\setminus\{x\}} \mathsf{Pr}_x^{\mathrm{BMC}}[v \in \mathsf{T},\ Z_{u_i} = y_k\ (i = 1, \ldots, n)] \\ &= \prod_{k=1}^{n} \mu[j_i, \infty) \sum_{y_1,\ldots,y_{n-1} \in X\setminus\{x\}} p(x, y_1)p(y_1, y_2)\cdots p(y_{n-1}, x).\end{aligned}$$

The last sum is $u^{(n)}(x, x)$. □

Exercise 5.51. It is definitely left to the reader to formulate this with complete rigour in terms of events in the underlying probability space and their probabilities. □

Exercise 5.52. If $\mu(0) = 0$ then each of the elements $u_n = 1^n = 1 \cdots 1$ ($n \geq 0$ times) belongs to T with probability 1. We know from (5.39.4) that $(Z_{u_n})_{n\geq 0}$ is a Markov chain on X with transition matrix P. If (X, P) is recurrent then this Markov chain returns to the starting point x infinitely often with probability 1. That is, $H^{\mathrm{BMC}}(x, x) = 1$. □

Exercises of Chapter 6

Exercise 6.10. The function $x \mapsto \nu_x(i)$ has the following properties: if $x \in C_i$ then $\nu_x(i) = \sum_y p(x, y)\nu_y(i) = 1$, since $y \in C_i$ when $p(x, y) > 0$. If $x \in C_j$ with $j \neq i$ then $\nu_x(i) = 0$. If $x \notin C_i$ then we can decompose with respect to the first step and obtain also $\nu_x(i) = \sum_y p(x, y)\nu_y(i)$. Therefore $h(x) = \sum_i g(i)\, \nu_x(i)$ is harmonic, and has value $g(i)$ on C_i for each i.

If h' is another harmonic function with value $g(i)$ on C_i for every i, then $g = h' - h$ is harmonic with value 0 on X_{ess}. By a straightforward adaptation of the maximum principle (Lemma 6.5), g must assume its maximum on X_{ess}, and the same holds for $-g$. Therefore $g \equiv 0$. □

Exercise 6.16. Variant 1. If the matrix P is irreducible and strictly substochastic in some row, then we know that it cannot have 1 as an eigenvalue.

Variant 2. Let $h \in \mathcal{H}$. Let x_0 be such that $h(x_0) = \max h$. Then, repeating the argument from the proof of the maximum principle, $h(y) = h(x_0)$ for every y with $p(x_0, y) > 0$. By irreducibility, $h \equiv c$ is constant. If now v_0 is such that $\sum_y p(v_0, y) < 1$ then $c = \sum_y p(v_0, y)\, c$, which is possible only when $c = 0$. □

Exercise 6.18. If I is finite then $|\inf_I h_i| \le \sum_I |h_i|$ is P-integrable.

If $h_i(x) \ge C$ for all $x \in X$ and all $i \in I$ then $C \le \inf_I h_i \le h_1$, so that $|\inf_I h_i| \le |C| + |h_1|$, a P-integrable upper bound. □

Exercise 6.20. We have

$$P = \begin{pmatrix} 0 & 1 \\ 1/2 & 0 \end{pmatrix} \quad \text{and} \quad \mathcal{G} = (I - P)^{-1} = \begin{pmatrix} 2 & 2 \\ 1 & 2 \end{pmatrix},$$

and $G(\cdot, y) \equiv 2$. □

Exercise 6.40. This can be done by a straightforward adaptation of the proof of Theorem 6.39 and is left entirely to the reader.

Exercise 6.22. (a) By Theorem 1.38 (d), $F(\cdot, y)$ is harmonic at every point except y. At y, we have

$$\sum_w p(y, w)\, F(w, y) = U(y, y) \le 1 = F(y, y).$$

(b) The "only if" is part of Lemma 6.17. Conversely, suppose that every non-negative, non-zero superharmonic function is strictly positive. For any $y \in X$, the function $F(\cdot, y)$ is superharmonic and has value 1 at y. The assumption yields that $F(x, y) > 0$ for all x, y, which is the same as irreducibility.

(c) Suppose that u is superharmonic and that x is such that $u(x) \le u(y)$ for all $y \in X$. Since P is stochastic,

$$\sum_y p(x, y) \underbrace{\big(u(y) - u(x)\big)}_{\ge 0} \ge 0.$$

Thus, $u(y) = u(x)$ whenever $x \overset{1}{\to} y$, and consequently whenever $x \to y$.

(d) Assume first that (X, P) is irreducible. Let u be superharmonic, and let x_0 be such that $u(x_0) = \min_X u(x)$. Since P is stochastic, the function $u - u(x_0)$ is again superharmonic, non-negative, and assumes the value 0. By Lemma 6.17(1), $u - u(x_0) \equiv 0$.

Conversely, assume that the minimum principle for superharmonic functions is valid. If $F(x, y) = 0$ for some x, y then the superharmonic function $F(\cdot, y)$ assumes 0 as its minimum. But this function cannot be constant, since $F(y, y) = 1$. Therefore we cannot have $F(x, y) = 0$ for any x, y. □

Exercise 6.24. (1) If $0 \le \nu P \le \nu$ then inductively

$$\nu P^n \le \nu P^{n-1} \le \cdots \le \nu P \le \nu.$$

(2) For every $i \in I$, we have $\nu P \le \nu_i P \le \nu_i$. Thus, $\nu P \le \inf_I \nu_i = \nu$.

(3) This is immediate from (3.20). □

Exercise 6.29. If $\mathsf{Pr}_x[t^A < \infty] = 1$ for every $x \in A$, then the Markov chain must return to A infinitely often with probability 1. Since A is finite, there must be at least one $y \in A$ that is visited infinitely often:

$$1 = \mathsf{Pr}_x[\exists y \in A : Z_n = y \text{ for infinitely many } n] \le \sum_{y \in A} H(x, y).$$

Thus there must be $y \in A$ such that $H(x, y) = U(x, y)H(y, y) > 0$. We see that $H(y, y) > 0$, which is equivalent with recurrence of the state y (Theorem 3.2). □

Exercise 6.31. We can compute the mean vector of the law of this random walk: $\bar{\mu} = \binom{p_1 - p_3}{p_2 - p_4}$, and appeal to Theorem 4.68.

Since that theorem was not proved here, we can also give a direct proof.

We first prove recurrence when $p_1 = p_3$ and $p_2 = p_4$. Then we have a symmetric Markov chain with $\mathsf{m}(\boldsymbol{x}) = 1$ and resistances p_1 on the horizontal edges and p_2 on the vertical edges of the square grid. Let $\mathcal{N}'$ be the resulting network, and let $\mathcal{N}$ be the network with the same underlying graph but all resistances equal to $\max\{p_1, p_2\}$. The Markov chain associated with $\mathcal{N}$ is simple random walk on $\mathbb{Z}^2$, which we know to be recurrent from Example 4.63. By Exercise 4.54, also $\mathcal{N}'$ is recurrent.

If we do not have $p_1 = p_3$ and $p_2 = p_4$, then let $\boldsymbol{c} = (c_1, c_2) \in \mathbb{R}_2$ and define the function $f_{\boldsymbol{c}}(\boldsymbol{x}) = e^{c_1 x_1 + c_2 x_2}$ for $\boldsymbol{x} = (x_1, x_2) \in \mathbb{Z}^2$. The one verifies easily that

$$P f_{\boldsymbol{c}} = \lambda_{\boldsymbol{c}} \cdot f_{\boldsymbol{c}}.$$

Then also $P^n f_{\boldsymbol{c}} = \lambda_{\boldsymbol{c}}^n \cdot f_{\boldsymbol{c}}$, so that $\rho(P) \le \lambda_{\boldsymbol{c}}$ for every $\boldsymbol{c}$. By elementary calculus,

$$\min\{\lambda_{\boldsymbol{c}} : \boldsymbol{c} \in \mathbb{R}^2\} = 2\sqrt{p_1 p_3} + 2\sqrt{p_2 p_4} < 1.$$

Therefore $\rho(P) < 1$, and we have transience. □

Exercise 6.45. We know that $g = u - h \ge 0$ and $g_{X_{\text{ess}}} \equiv 0$. We can proceed as in the proof of the Riesz decomposition: set $f = g - Pg$, which is ≥ 0. Then $\text{supp}(f) \subset X^o$, and

$$g - P^{n+1} g = \sum_{k=0}^{n} P^k f.$$

Now $p^{(n+1)}(x,y) \to 0$ as $n \to \infty$, whenever $y \in X^o$. Therefore $P^{n+1}g \to 0$, and $g = Gf$. Uniqueness is immediate, as in the last line of the proof of Theorem 6.43. □

Exercise 6.48. The notation $\widehat{\Pi}$ will refer to sets of paths in $\Gamma(\widehat{P})$, and $\widehat{\mathsf{w}}(\cdot)$ to their weights with respect to $\widehat{P}$. Then it is clear that the mapping

$$\pi = [x_0, x_1, \dots, x_n] \mapsto \hat{\pi} = [x_n, \dots, x_1, x_0]$$

is a bijection between $\{\pi \in \Pi(x,y) : \pi \text{ meets } A \text{ only in the terminal point}\}$ and $\{\hat{\pi} \in \widehat{\Pi}(x,y) : \hat{\pi} \text{ meets } A \text{ only in the initial point}\}$. It is also a bijection between $\{\pi \in \Pi(x,y) : \pi \text{ meets } A \text{ only in the initial point}\}$ and $\{\hat{\pi} \in \widehat{\Pi}(x,y) : \hat{\pi} \text{ meets } A \text{ only in the terminal point}\}$.

By a straightforward computation, $\widehat{\mathsf{w}}(\hat{\pi}) = \nu(x_0)\mathsf{w}(\pi)/\nu(x_n)$ for any path $\pi = [x_0, x_1, \dots, x_n]$. Summing over all paths in the respective sets, the formulas follow. □

Exercise 6.57. This is left entirely to the reader.

Exercise 6.58. We fix w and y. Set $h = G(\cdot, y)$ and $f = \frac{G(w,y)}{G(w,w)}\mathbf{1}_w$. Then $Gf(x) = F(x,w)G(w,y)$. We have $h(w) = Gf(w)$. By the domination principle, $h(x) \geq Gf(x)$ for all x. □

Exercises of Chapter 7

Exercise 7.9. (1) We have $\sum_y p_h(x,y) = \frac{1}{h(x)} Ph(x)$, which is $= 1$ if and only if $Ph(x) = h(x)$.

(2) In the same way,

$$P_h\bar{u}(x) = \sum_y \frac{p(x,y)h(y)}{h(x)} \frac{u(y)}{h(y)} = \frac{1}{h(x)} Pu(x),$$

which is $\leq \bar{u}(x)$ if and only if $Pu(x) \leq u(x)$.

(3) Suppose that u is minimal harmonic with respect to P. We know from (2) that $h(o) \cdot \bar{u}$ is harmonic with respect to P_h. Suppose that $h(o) \cdot \bar{u} \geq \bar{h}_1$, where $\bar{u} = u/h$ as above, and $\bar{h}_1 \in \mathcal{H}^+(X, P_h)$. By (2), $h_1 = h\bar{h}_1 \in \mathcal{H}^+(X,P)$. On the other hand, $u \geq \frac{1}{h(o)} h_1$. Since u is minimal, u/h_1 is constant. But then also $h(o) \cdot \bar{u}/\bar{h}_1$ is constant. Thus, $h(o) \cdot \bar{u}$ is minimal harmonic with respect to P_h.

The converse is proved in the same way and is left entirely to the reader.

Exercise 7.12. We suppose to have two compactifications $\widehat{X}$ and $\widetilde{X}$ of X and continuous surjections $\tau : \widehat{X} \to \widetilde{X}$ and $\sigma : \widetilde{X} \to \widehat{X}$ such that $\tau(x) = \sigma(x) = x$ for all $x \in X$.

Let $\xi \in \widetilde{X}$. There is a sequence (x_n) in X such that $x_n \to \xi$ in the topology of $\widetilde{X}$. By continuity of τ, $x_n = \tau(x_n) \to \tau(\xi)$ in the topology of $\widehat{X}$. But then by continuity of σ, $x_n = \sigma(x_n) \to \sigma\big(\tau(\xi)\big)$ in the topology of $\widetilde{X}$. Therefore $\sigma\big(\tau(\xi)\big) = \xi$. Therefore τ is injective, whence bijective, and σ is the inverse mapping. □

Exercise 7.16. We start with the second model, and will also see that it is equivalent with the first one.

First of all, X is discrete because when $x \in X$ and $y \neq x$ then $\theta(x, y) \geq w_{\mathbf{1}_x} |\mathbf{1}_x(x) - \mathbf{1}_x(y)| = w_{\mathbf{1}_x}$, so that the open ball with centre x and radius $w_{\mathbf{1}_x}$ consists only of x. Next, let (x_n) be a Cauchy sequence in the metric θ. Then $\big(f(x_n)\big)$ must be a Cauchy sequence in $\mathbb{R}$ and $\lim_n f(x_n)$ exists for every $f \in \mathcal{F}^*$.

Conversely, suppose that $\big(f(x_n)\big)$ is a Cauchy sequence in $\mathbb{R}$ for every $f \in \mathcal{F}^*$. Given $\varepsilon > 0$, there is a finite subset $\mathcal{F}_\varepsilon \subset \mathcal{F}^*$ such that $\sum_{\mathcal{F}^* \setminus \mathcal{F}_\varepsilon} w_f\, C_f < \varepsilon/4$. Now let N_ε be such that $|f(x_n) - f(x_m)| < C_f\, \varepsilon/(2S)$ for all $n, m \geq N_\varepsilon$ and all $f \in \mathcal{F}_\varepsilon$, where $S = \sum_{\mathcal{F}^*} w_f\, C_f\big)$. Then for such n, m

$$\theta(x_n, x_m) < \sum_{\mathcal{F}^* \setminus \mathcal{F}_\varepsilon} w_f \left(|f(x_n)| + |f(x_m)|\right) + \sum_{\mathcal{F}_\varepsilon} w_f\, C_f \varepsilon/(2S) \leq \varepsilon.$$

Thus, (x_n) is a Cauchy sequence in the metric $\theta(\cdot, \cdot)$.

Let (x_n) be such a sequence. If there is $x \in X$ such that $x_{n_k} = x$ for an infinite subsequence (n_k), then $\mathbf{1}_x(x_{n_k}) = 1$ for all k. Since $\big(\mathbf{1}_x(x_n)\big)$ converges, the limit is 1. We see that every Cauchy sequence (x_n) is such that either $x_n = x$ for some $x \in X$ and all but finitely many n, or else (x_n) tends to ∞ and $\big(f(x_n)\big)$ is a convergent sequence in $\mathbb{R}$ for every $f \in \mathcal{F}$.

At this point, recall the general construction of the completion of the metric space (X, θ). The completion consists of all equivalence classes of Cauchy sequences, where $(x_n) \sim (y_n)$ if $\theta(x_n, y_n) \to 0$. In our case, this means that either $x_n = y_n = x$ for some $x \in X$ and all but finitely many n, or – as above – that both sequences tend to ∞ and $\lim f(x_n) = \lim f(y_n)$ for every $f \in \mathcal{F}$. The embedding of X in this space of equivalence classes is via the equivalence classes of constant sequences in X, and the extended metric is $\theta(\xi, \eta) = \lim_n \theta(x_n, y_n)$, where (x_n) and (y_n) are arbitrary representatives of the equivalence classes ξ and η, respectively.

At this point, we see that the completion X coincides with $\mathcal{X}_\infty$ of the first model, including the notion of convergence in that model.

We now write $\widetilde{X}$ for this completion. We show compactness. Since X is dense in $\widetilde{X}$, it is sufficient to show that every sequence (x_n) in X has a convergent subsequence. Now, the sequence $\big(f(x_n)\big)$ is bounded for every $f \in \mathcal{F}^*$. Since $\mathcal{F}^*$ is countable, we can use the well-known diagonal argument to extract a subsequence (x_{n_k}) such that $\big(f(x_{n_k})\big)$ converges for every $f \in \mathcal{F}^*$. We know that this is a Cauchy sequence, whence it converges to some element of $\widetilde{X}$.

By construction, every $f \in \mathcal{F}$ extends to a continuous function on $\widetilde{X}$. Finally, if $\xi, \eta \in \widetilde{X} \setminus X$ are distinct, then let the sequences (x_n) and (y_n) be representatives of those two respective equivalence classes. They are not equivalent, so that there must be $f \in \mathcal{F}$ such that $\lim_n f(x_n) \neq \lim_n f(y_n)$. That is, $f(\xi) \neq f(\eta)$. We conclude that $\mathcal{F}$ separates the boundary points.

By Theorem 7.13, $\widetilde{X}$ is (equivalent with) $\widetilde{X}_{\mathcal{F}}$. □

Exercise 7.26. Let us refresh our knowledge about conditional expectation: we have the probability space $(\Omega, \mathcal{A}, \mathsf{Pr})$, and a sub-$\sigma$-algebra $\mathcal{A}_{n-1}$, which in our case is the one generated by $(Y_0, \dots, Y_{n-1})$. We take a non-negative, real, $\mathcal{A}$-measurable random variable W, in our case $W = W_n$. We can define a measure Q_W on $\mathcal{A}$ by $Q_W(A) = \int_A W \, d\,\mathsf{Pr}$. Both Pr and Q_W are now considered as measures on $\mathcal{A}_{n-1}$, and Q_W is absolutely continuous with respect to Pr. Therefore, it has an $\mathcal{A}_{n-1}$-measurable Radon–Nikodym density, which is Pr-almost surely unique. By definition, this is $\mathsf{E}(W|\mathcal{A}_{n-1})$. If W itself is $\mathcal{A}_{n-1}$-measurable then it is itself that density. In general, Q_W can also be defined on $\mathcal{A}_{n-1}$ via integration: for every non-negative $\mathcal{A}_{n-1}$-measurable function (random variable) V on Ω,

$$\int_\Omega V \, dQ_A = \int_\Omega V \cdot W \, d\,\mathsf{Pr}\,.$$

Thus, by the construction of the conditional expectation, the latter is characterized as the a.s. unique $\mathcal{A}_{n-1}$-measurable random variable $\mathsf{E}(W|\mathcal{A}_{n-1})$ that satisfies

$$\int_\Omega V \cdot \mathsf{E}(W|\mathcal{A}_{n-1}) \, d\,\mathsf{Pr} = \int_\Omega V \cdot W \, d\,\mathsf{Pr},$$

that is,

$$\mathsf{E}\big(V \cdot \mathsf{E}(W|\mathcal{A}_{n-1})\big) = \mathsf{E}(V \cdot W)$$

for every V as above.

In our case, $\mathcal{A}_{n-1}$ is generated by the atoms $[Y_0 = x_0, \dots, Y_{n-1} = x_{n-1}]$, where $x_0, \dots, x_{n-1} \in X \cup \{\dagger\}$. Every $\mathcal{A}_{n-1}$-measurable function is constant on each of those atoms. In other words, every such function has the form $V = g(Y_0, \dots, Y_{n-1})$. We can reformulate: $\mathsf{E}(W|Y_0, \dots, Y_n)$ is the a.s. unique $\mathcal{A}_{n-1}$-measurable random variable that satisfies

$$\mathsf{E}\big(g(Y_0, \dots, Y_{n-1}) \cdot \mathsf{E}(W|\mathcal{A}_{n-1})\big) = \mathsf{E}\big(g(Y_0, \dots, Y_{n-1}) \cdot W\big)$$

for every function $g\colon (X \cup \{\dagger\})^n \to [0, \infty)$.

If we now set $W = W_n = f_n(Y_0, \dots, Y_n)$, then the supermartingale property $\mathsf{E}(W_n \mid Y_0, \dots, Y_{n-1}) \leq W_{n-1}$ implies

$$\begin{aligned}\mathsf{E}(g(Y_0, \dots, Y_{n-1}) \cdot W_n) &= \mathsf{E}\big(g(Y_0, \dots, Y_{n-1}) \cdot \mathsf{E}(W_n|\mathcal{A}_{n-1})\big) \\ &\leq \mathsf{E}\big(g(Y_0, \dots, Y_{n-1}) \cdot W_{n-1}\big)\end{aligned}$$

for every g as above. Conversely, if the inequality between the second and the third term holds for every such g, then the measures Q_{W_n} and $Q_{W_{n-1}}$ on $\mathcal{A}_{n-1}$ satisfy $Q_{W_n} \le Q_{W_{n-1}}$. But then their Radon–Nikodym derivatives with respect to Pr_x also satisfy

$$\mathsf{E}(W_n \mid Y_0, \dots, Y_{n-1}) = \frac{dQ_{W_n}}{d\,\mathsf{Pr}} \le \frac{dQ_{W_{n-1}}}{d\,\mathsf{Pr}} = \mathsf{E}(W_{n-1} \mid Y_0, \dots, Y_{n-1}) = W_{n-1}$$

almost surely. (The last identity holds because W_{n-1} is $\mathcal{A}_{n-1}$-measurable.)

So now we see that (7.25) is equivalent with the supermartingale property. If we take $g = \mathbf{1}_{(x_0,\dots,x_{n-1})}$, where $(x_0, \dots, x_{n-1}) \in (X \cup \{\dagger\})^n$, then (7.25) specializes to (7.24). Conversely, every function $g\colon (X \cup \{\dagger\})^n \to [0, \infty)$ is a finite, non-negative linear combination of such indicator functions. Therefore (7.24) implies (7.25). □

Exercise 7.32. The function $G(\cdot, y)$ is superharmonic and bounded by $G(y, y)$. Thus, $\big(G(Z_n, y)\big)$ is a bounded, positive supermartingale, and must converge almost surely. Let W be the limit random variable. By dominated convergence, $\mathsf{E}_x(W) = \lim_n \mathsf{E}_x\big(G(Z_n, y)\big)$. But

$$\mathsf{E}_x\big(G(Z_n, y)\big) = \sum_{v \in X} p^{(n)}(x, v)\, G(v, y) = \sum_{k=n}^{\infty} p^{(k)}(x, y),$$

which tends to 0 as $n \to \infty$. Therefore $\mathsf{E}_x(W) = 0$, so that (being non-negative) $W = 0$ Pr_x-almost surely. □

Exercise 7.39. Let $k, l, r \in \mathbb{N}$ and consider the set

$$A_{k,l,r} = \big\{\omega = (x_n) \in X^{\mathbb{N}_0} : K(x, x_r) < c + \varepsilon - \tfrac{1}{k} + \tfrac{1}{l}\big\}.$$

This set is a union of basic cylinder sets, since it depends only on the r-th projection x_r of ω. We invite the reader to check that

$$\bigcup_{k=1}^{\infty} \bigcap_{l=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{r=m}^{\infty} A_{k,l,r} = A_{\limsup} = \big\{\omega = (x_n) \in X^{\mathbb{N}_0} : \limsup_{n\to\infty} K(x, x_n) < c + \varepsilon\big\}.$$

Therefore the latter set is in $\mathcal{A}$.

We leave it entirely to the reader to work out that analogously,

$$A_{\liminf} = \big\{\omega = (x_n) \in X^{\mathbb{N}_0} : \liminf_{n\to\infty} K(x, x_n) > c - \varepsilon\big\} \in \mathcal{A}.$$

Then our set is

$$\Omega_\infty \cap A_{\limsup} \cap A_{\liminf} \in \mathcal{A}.$$ □

Exercise 7.49. By (7.8), the Green function of the h-process is

$$G_h(x,y) = G(x,y)h(y)/h(x).$$

Therefore (7.47) and (7.40) yield

$$\begin{aligned}\nu^h(v) &= h(o)\,G_h(o,v)\Big(1-\sum_{w\in X} p_h(v,w)\Big)\\ &= G(o,v)\,h(v)\Big(1-\sum_{w\in X} p(v,w)\frac{h(w)}{h(v)}\Big)\\ &= G(o,v)\Big(h(v)-\sum_{w\in X} p(v,w)h(w)\Big),\end{aligned}$$

as proposed. Setting $h = K(\cdot, y)$, we get for $v \in X$

$$\begin{aligned}\nu^h(v) &= G(o,v)\Big(K(v,y)-\sum_{w\in X} p(v,w)K(w,y)\Big)\\ &= G(v,y) - PG(v,y) = \delta_y(v).\end{aligned}$$

□

Exercise 7.58. Let (X, Q) be irreducible and recurrent. In particular, Q is stochastic. Choose $w \in X$ and define a transition matrix P by

$$p(x,y) = \begin{cases} q(x,y)/2, & \text{if } x = w,\\ q(x,y), & \text{if } x \neq w.\end{cases}$$

Thus, $p(w,\dagger) = 1/2$ and $p(x,\dagger) = 0$ when $x \neq w$. Then $F(x,w)$ is the same for P and Q, because $F(x,w)$ does not depend on the outgoing probabilities at w; compare with Exercise 2.12. Thus $F(x,w) = 1$ for all x. For the chain extended to $X \cup \{\dagger\}$, the state w is a cut point between any $x \in X \setminus \{w\}$ and $\dagger$. Via Theorem 1.38 and Proposition 1.43,

$$F(w,\dagger) = \frac{1}{2} + \sum_{x\in X} p(w,x)F(x,w)F(w,\dagger) = \frac{1}{2} + \frac{1}{2}F(w,\dagger).$$

We see that $F(w,\dagger) = 1$, and for $x \in X \setminus \{w\}$, we also have $F(x,\dagger) = F(x,w)F(w,\dagger) = 1$. Thus, the Markov chain with transition matrix P is absorbed by $\dagger$ almost surely for every starting point. □

Exercise 7.66. (a) $\Longrightarrow$ (b). Let $\operatorname{supp}(\nu) = \{\xi\}$. Then $h_0(x) = \nu_x(\xi) = K(x,\xi)\,\nu(\xi)$ is non-zero. If h is a bounded harmonic function then there is a bounded measurable function φ on $\mathcal{M}$ such that

$$h(x) = \int_{\mathcal{M}} K(x,\cdot)\,\varphi\,d\nu = K(x,\xi)\,\varphi(\xi)\,\nu(\xi) = \varphi(\xi)\,h_0(x).$$

(b) $\Longrightarrow$ (c). Let h_1 be a non-negative harmonic function such that $h_0 \geq h_1$. Then h_1 is bounded, and by the hypothesis, h_1/h_0 is constant. That is, $\frac{1}{h_0(o)}h_0$ is minimal.

(c) $\Longrightarrow$ (a). If $\frac{1}{h_0(o)}h_0$ is minimal then it must be a Martin kernel $K(\cdot,\xi)$. Thus

$$h_0(x) = h_0(o)\int_{\mathcal{M}_{\min}} K(x,\cdot)\,d\delta_\xi \quad = \int_{\mathcal{M}_{\min}} K(x,\cdot)\,d\nu.$$

By the uniqueness of the integral representation, $\nu = h_0(o)\cdot\delta_\xi$. □

Exercises of Chapter 8

Exercise 8.8. In the proof of Theorem 8.2, irreducibility is only used in the last 4 lines. Without any change, we have $h(\boldsymbol{k}+\boldsymbol{l}) = h(\boldsymbol{k})$ for every $\boldsymbol{k} \in \mathbb{Z}^d$ and every $\boldsymbol{l} \in \operatorname{supp}(\mu)$. But then also $h(\boldsymbol{k}-\boldsymbol{l}) = h(\boldsymbol{k})$. Suppose that $\operatorname{supp}(\mu)$ generates $\mathbb{Z}^d$ as a group, and let $\boldsymbol{k} \in \mathbb{Z}^d$. Then we can find $n > 0$ and elements $\boldsymbol{l}_1, \ldots, \boldsymbol{l}_n \in \operatorname{supp}(\mu)$ such that $\boldsymbol{k} = \pm\boldsymbol{l}_1 \pm \cdots \pm \boldsymbol{l}_n$. Again, we get $h(\boldsymbol{0}) = h(\boldsymbol{k})$.

In part A of the proof of Theorem 8.7, irreducibility is not used up to the point where we obtained that $h_{\boldsymbol{l}} = h$ for every $\boldsymbol{l} \in \operatorname{supp}(\mu^{(n)}$ and each $n \geq 0$. That is, fur such $\boldsymbol{l}$ and every $\boldsymbol{k} \in \mathbb{Z}^d$, $h(\boldsymbol{k}+\boldsymbol{l}) = h(\boldsymbol{k})h(\boldsymbol{l})$. But then also

$$h(\boldsymbol{k}) = h(\boldsymbol{k}-\boldsymbol{l}+\boldsymbol{l}) = h(\boldsymbol{k}-\boldsymbol{l})h(\boldsymbol{l}).$$

With $\boldsymbol{k} = \boldsymbol{0}$, we find $h(-\boldsymbol{l}) = 1/h(\boldsymbol{l})$, and then in general

$$h(\boldsymbol{k}-\boldsymbol{l}) = h(\boldsymbol{k})/h(\boldsymbol{l}) = h(\boldsymbol{k})h(-\boldsymbol{l}).$$

Now, as above, if $\operatorname{supp}(\mu)$ generates $\mathbb{Z}^d$ as a group, then every $\boldsymbol{l} \in \mathbb{Z}^d$ has the form $\boldsymbol{l} = \boldsymbol{l}_1 - \boldsymbol{l}_2$, where $\boldsymbol{l}_i \in \operatorname{supp}(\mu^{(n_i)})$ for some $n_i \in \mathbb{N}_0$ $(i = 1, 2)$. But then

$$h(\boldsymbol{k}+\boldsymbol{l}) = h(\boldsymbol{k}+\boldsymbol{l}_1-\boldsymbol{l}_2) = h(\boldsymbol{k})h(\boldsymbol{l}_1)h(-\boldsymbol{l}_2) = h(\boldsymbol{k})h(\boldsymbol{l}),$$

and this is true for all $\boldsymbol{k}, \boldsymbol{l} \in \mathbb{Z}^d$.

In part B of the proof of Theorem 8.7, irreducibility is not used directly. We should go back a little bit and see whether irreducibility is needed for Corollary 7.11, or for Exercise 7.9 and Lemma 7.10 which lead to that corollary. There, the crucial point is that we are allowed to define the h-process for $h = f_{\boldsymbol{c}}$, since that function does not vanish at any point. □

Exercise 8.11. Part A of the proof of Theorem 8.7 remains unchanged: every minimal harmonic function has the form $f_{\boldsymbol{c}}$ with $\boldsymbol{c} \in \boldsymbol{C}$. Pointwise convergence within that set of functions is the same as usual convergence in the set $\boldsymbol{C}$. The topology on $\mathcal{M}_{\min}$ is the one of pointwise convergence. This yields that there is a

subset $\boldsymbol{C}'$ of $\boldsymbol{C}$ such that the mapping $\boldsymbol{c} \mapsto f_{\boldsymbol{c}}$ ($\boldsymbol{c} \in \boldsymbol{C}'$) is a homeomorphism from $\boldsymbol{C}'$ to $\mathcal{M}_{\min}$. Therefore every positive harmonic function h has a unique integral representation

$$h(\boldsymbol{k}) = \int_{\boldsymbol{C}'} e^{\boldsymbol{c}\cdot\boldsymbol{k}}\, d\nu(\boldsymbol{c}) \quad \text{for all } \boldsymbol{k} \in \mathbb{Z}^d.$$

Suppose that there is $\boldsymbol{c} \neq \boldsymbol{0}$ in $\operatorname{supp}(\nu)$. Let B be the open ball in $\mathbb{R}^d$ with centre $\boldsymbol{c}$ and radius $|\boldsymbol{c}|/2$. The cone $\{t \cdot \boldsymbol{x} : t \geq 0,\ \boldsymbol{x} \in B\}$ opens with the angle $\pi/3$ at its vertex $\boldsymbol{0}$. Therefore, if $\boldsymbol{k} \in \mathbb{Z}^d \setminus \{\boldsymbol{0}\}$ is in that cone, then

$$|\boldsymbol{x} \cdot \boldsymbol{k}| \geq \cos(\pi/3)\, |\boldsymbol{x}|\, |\boldsymbol{k}| \geq |\boldsymbol{k}|\, |\boldsymbol{c}|/4 \quad \text{for all } \boldsymbol{x} \in B.$$

For such $\boldsymbol{k}$, we get $h(\boldsymbol{k}) \geq |\boldsymbol{k}|\, \nu(B)\, |\boldsymbol{c}|/4$, which is unbounded. We see that when h is a bounded harmonic function, then $\operatorname{supp}(\nu)$ contains no non-zero element. That is, ν is a multiple of the point mass at $\boldsymbol{0}$, and h is constant.

Theorem 8.2 follows.

As suggested, we now conclude our reasoning with part B of the proof of Theorem 8.7 without any change. This shows also that $\boldsymbol{C}'$ cannot be a proper subset of $\boldsymbol{C}$. □

Exercise 8.13. Since the function φ is convex, the set $\{\boldsymbol{c} \in \mathbb{R}^d : \varphi(\boldsymbol{c}) \leq 1\}$ is convex. Since φ is continuous, that set is closed, and by Lemma 8.12, it is bounded, whence compact. Its interior is $\{\boldsymbol{c} \in \mathbb{R}^d : \varphi(\boldsymbol{c}) < 1\}$, so that its topological boundary is $\boldsymbol{C}$. As φ is strictly convex, it has a unique minimum, which is the point where the gradient of φ is $\boldsymbol{0}$. This leads to the proposed equation for the point where the minimum is attained. □

Exercises of Chapter 9

Exercise 9.4. This is straightforward by the quotient rule. □

Exercise 9.7. We first show this when $y \sim o$. Then $\mathsf{m}^o(y) = p(o,y)/p(y,o) = 1/\mathsf{m}^y(o)$. We have to distinguish two cases.
(1) If $x \in T_{o,y}$ then $y = x_1$ in the formula (9.6) for $\mathsf{m}^o(x)$, and

$$\mathsf{m}^o(x) = \frac{p(o,y)}{p(y,o)} \frac{p(y,x_2)\cdots p(x_{k-1},x_k)}{p(x_2,y)\cdots p(x_k,x_{k-1})} = \mathsf{m}^o(y)\mathsf{m}^y(x).$$

(2) If $x \in T_{y,o}$ then we can exchange the role of o and y in the first case and get $\mathsf{m}^y(x) = \mathsf{m}^y(o)\mathsf{m}^o(x) = \mathsf{m}^o(x)/\mathsf{m}^o(y)$.

The rest of the proof is by induction on the distance $d(y,o)$ in T. We suppose the statement is true for y^-, that is, $\mathsf{m}^o(x) = \mathsf{m}^o(y^-)\mathsf{m}^{y^-}(x)$ for all x. Applying the initial argument to y^- in the place of o, we get $\mathsf{m}^{y^-}(x) = \mathsf{m}^{y^-}(y)\mathsf{m}^y(x)$. Since $\mathsf{m}^o(y^-)\mathsf{m}^{y^-}(y) = \mathsf{m}^o(y)$, the argument is complete. □

Exercise 9.9. Among the finitely many cones T_x, where $x \sim o$, at least one must be infinite. Let this be T_{x_1}. We now proceed by induction. If we have already found a geodesic arc $[o, x_1, \dots, x_n]$ such that T_{x_n} is infinite, then among the finitely many T_y with $y^- = x_n$, at least one must be infinite. Let this be $T_{x_{n+1}}$.

In this way, we get a ray $[o, x_1, x_2, \dots]$. □

Exercise 9.11. Reflexivity and symmetry of the relation are clear. For transitivity, let $\pi = [x_0, x_1, \dots]$, $\pi' = [y_0, y_1, \dots]$ and $\pi'' = [z_0, z_1, \dots]$ be rays such that π and π' as well as π' and π'' are equivalent. Then there are i, j and k, l such that $x_{i+n} = y_{j+n}$ and $y_{k+n} = z_{l+n}$ for all n. Then $x_{(i+k)+n} = y_{j+k+n} = z_{(j+l)+n}$ for all n, so that π and π'' are also equivalent.

For the second statement, let $\pi = [y_0, y_1, \dots]$ be a geodesic ray that represents the end ξ. For $x \in X$, consider $\pi(x, y_0) = [x = x_0, x_1, \dots, x_m = y_0]$. Let j be minimal in $\{0, \dots, m\}$ such that $x_j \in \pi$. That is, $x_j = y_k$ for some k. Then

$$\pi' = [x = x_0, x_1, \dots, x_j = y_k, y_{k+1}, \dots]$$

is a geodesic ray equivalent with π that starts at x. Uniqueness follows from the fact that a tree has no cycles.

For two ends ξ, η, let $\pi = \pi(o, \xi) = [o = w_0, w_1, \dots]$ and $\pi' = \pi(o, \eta) = [o = y_0, y_1, \dots]$. These rays are not equivalent. Let k be minimal such that $w_{k+1} \neq y_{k+1}$. Then $k \geq 1$, and the rays $[w_{k+1}, w_{k+2}, \dots]$ and $[y_{k+1}, y_{k+2}, \dots]$ must be disjoint, since otherwise there would be a cycle in T. We can set $x_0 = y_k = w_k$, $x_n = y_{k+n}$ and $x_{-n} = w_{k+n}$ for $n > 0$. Then $[\dots, -x_2, -x_1, x_0, x_1, x_2, \dots]$ is a geodesic with the required properties. Uniqueness follows again from the fact that a tree has no cycles. □

Exercise 9.15. The first case (convergence to a vertex x) is clear, since the topology is discrete on T.

By definition, $w_n \to \xi \in \partial T$ if and only if for every $y \in \pi(o, \xi)$, there is n_y such that $w_n \in \widehat{T}_y$ for all $n \geq n_y$. Now, if $w_n \in \widehat{T}_y$ then $\pi(o, y)$ is part of $\pi(o, \xi)$ as well as of $\pi(o, \xi)$. That is, $w_n \wedge \xi \in \widehat{T}_y$, so that $|w_n \wedge \xi| \geq |y|$ for all $n \geq n_y$. Therefore $|w_n \wedge \xi| \to \infty$. Conversely, if $|w_n \wedge \xi| \to \infty$ then for each $y \in \pi(o, \xi)$ there is n_y such that $|w_n \wedge \xi| \geq |y|$ for all $n \geq n_y$, and in this case, we must have $w_n \in \widehat{T}_y$.

Again by definition of the topology, $w_n \to y^*$ if and only if for every finite set A of neighbours of y, there is n_A such that $w_n \in \widehat{T}_{x,y}$ for all $x \in A$ and all $n \geq n_A$. But for $x \in A$, one has $w_n \in \widehat{T}_{x,y}$ if and only if $x \notin \pi(y, w_n)$.

Finally, since $x_n^* \in \widehat{T}_{w,y}$ if and only if $x_n \in \widehat{T}_{w,y}$, and since all types of convergence are based on inclusions of the latter type, the statement about convergence of (x_n^*) follows.

We now prove compactness. Since the topology has a countable base, we just show sequential compactness. By construction, T is dense in $\widehat{T}$. Therefore it

is enough to show that every sequence (x_n) of elements of T has a subsequence that converges in $\widehat{X}$. If there is $x \in T$ such that $x_n = x$ for infinitely many n, then we have such a subsequence. So we may assume (passing to a subsequence, if necessary) that all x_n are distinct. We use an elementary inductive procedure, similar to Exercise 9.9.

If for every $y \sim o$, the cone T_y contains only finitely many x_n, then $o \in T^\infty$ and $x_n \to o^*$, and we are done.

Otherwise, there is $y_1 \sim o$ such that T_{y_1} contains infinitely many x_n, and we can pass to the next step.

If for every y with $y^- = y_1$, the cone T_y contains only finitely many x_n, then $y_1 \in T^\infty$ and $x_{n_k} \to y_1^*$ for the subsequence of those x_n that are in T_{y_1}, and we are done.

Otherwise, there is y_2 with $y_2^- = y_1$ such that T_{y_2} contains infinitely many x_n, and we pass to the third step.

Inductively, we either find $y_k \in T^\infty$ and a subsequence of (x_n) that converges to y_k^*, or else we find a sequence $o \sim y_1, y_2, y_3, \dots$ such that $y_{k+1}^- = y_k$ for all k, such that each T_{y_k} contains infinitely many x_n. The ray $[o, y_1, y_2, \dots]$ represents an end ξ of T, and it is immediate that (x_n) has a subsequence that converges to that end. □

Exercise 9.17. If $f_1, f_2 \in \mathcal{L}$ and $\lambda_1, \lambda_2 \in \mathbb{R}$ then

$$\big\{[x,y] \in E(T) : \lambda_1 f_1(x) + \lambda_2 f_2(x) \neq \lambda_1 f_1(y) + \lambda_2 f_2(y)\big\}$$
$$\subset \big\{[x,y] \in E(T) : f_1(x) \neq f_1(y)\big\} \cup \big\{[x,y] \in E(T) : f_2(x) \neq f_2(y)\big\},$$

which is finite.

Next, let $f \in \mathcal{L}$. In order to prove that f is in the linear span of $\mathcal{L}_0$, we proceed by induction on the cardinality n of the set $\{e \in E(T) : f(e^+) \neq f(e^-)\}$, which has to be even in our setting, since we have oriented edges. If $n = 0$ then $f \equiv c$, and we can write $f = c \cdot \mathbf{1}_{T_{x,y}} + c \cdot \mathbf{1}_{T_{y,x}}$.

Now suppose the statement is true for $n - 2$. There must be an edge $[x, y] \in \{e \in E(T) : f(e^+) \neq f(e^-)\}$ such that $f(u) = f(v)$ for all $[u, v] \in E(T_{x,y})$. That is, $f \equiv c$ is constant on $T_{x,y}$. Let $g = f + \big(f(x) - c\big) \cdot \mathbf{1}_{T_{x,y}}$. Then $g(x) = g(y)$, and the number of edges along which g differs is $n - 2$. By the induction hypothesis, g is a linear combination of functions $\mathbf{1}_{T_{u,v}}$, where $u \sim v$. Therefore also f has this property. □

Exercise 9.26. We use the first formula of Proposition 9.3 (b). If $F(y, x) = 1$ then

$$p(y,x) + \sum_{w \neq x \,:\, w \sim y} p(y,w) F(w,y) = 1.$$

Since $\sum_{w\,:\,w\sim y} p(y,w) = 1$, this yields $F(w, y) = 1$ for all $w \neq x$ with $w \sim y$. The reader can now proceed by induction on the distance from y in an obvious manner.

Now let ξ be a transient end, and let x be a new base point. Write $\pi = \pi(x, \xi) = [x = x_0, x_1, x_2, \dots]$. Then $x \wedge \xi = x_k$ for some k, so that $[x_k, x_{k+1}, \dots] \subset \pi(o, \xi)$ and $x_{n+1}^- = x_n$ for all $n \geq k$. Therefore $F(x_{n+1}, x_n) < 1$ for all $n \geq k$. The first part of the exercise implies that this holds for all $n \geq 0$, which shows that ξ is also a transient end with respect to the base point x. □

Exercise 9.31. Given x, let $w = x \wedge \xi$. This is a point on $\pi(o, \xi)$ as well as on $\pi(o, x)$. If $y \in T_w$ then $d(x, y) = d(x, w) + d(w, y)$ and $d(o, y) = d(o, w) + d(w, y)$, and $x \wedge y = w$. We see that $\mathrm{hor}(x, \xi) = \mathrm{hor}(x, y) = d(x, y) - d(o, y)$ is constant for all $y \in T_w$, when x is given. □

Exercise 9.39. If $x \sim o$ then

$$\begin{aligned} h(x) &= K(x, o)\, \nu(\partial T) + \big(K(x, x) - K(x, o)\big)\, \nu(\partial T_x) \\ &= F(x, o)\, \nu(\partial T) + \frac{1 - F(o, x)F(x, o)}{F(o, x)}\, \nu(\partial T_x) \\ &= F(x, o)\, \nu(\partial T) + \frac{1 - U(o, o)}{p(o, x)}\, \nu(\partial T_x) \end{aligned}$$

by (9.36).Therefore

$$\begin{aligned} \sum_{x \sim o} p(o, x) h(x) &= \Big(\sum_{x \sim o} p(o, x) F(x, o)\Big)\, \nu(\partial T) + \big(1 - U(o, o)\big) \sum_{x \sim o} \nu(\partial T_x) \\ &= \nu(\partial T) = h(o), \end{aligned}$$

as proposed.

Exercise (Corollary) 9.44. If the Green kernel vanishes at infinity then it vanishes at every boundary point, and the Dirichlet problem at infinity admits a solution.

Conversely, if $\lim_{x \to \xi} G(x, o) = 0$ for every $\xi \in \partial^* T$, then the function g on $\widehat{T}$ is continuous, where $g(x) = G(x, o)$ for $x \in T$ and $g(\xi) = 0$ for every $\xi \in \partial^* T$. Given $\varepsilon > 0$, every $\xi \in \partial^* T$ has an open neighbourhood V_ξ in $\widehat{T}$ on which $g < \varepsilon$. Then $V = \bigcup_{\xi \in \partial^* T} V_\xi$ is open and contains $\partial^* T$. Thus, the complement $\widehat{T} \setminus V$ is compact and contains no boundary point. Thus, it is a finite set of vertices, outside of which $g < \varepsilon$. □

Exercise 9.52. For $k \in \mathbb{N}$ and any end $\xi \in \partial T$, let $\varphi_k(\xi) = h\big(v_k(\xi)\big)$. This is a continuous function on ∂T. It is a standard fact that the set of all points where a sequence of continuous functions converges to a given Borel measurable function is a Borel set. □

Exercise 9.54.

$$\begin{aligned}\mathsf{Pr}_x[Z_n \in T_x \setminus \{x\} \text{ for all } n \geq 1] &= \sum_{y\,:\,y^-=x} p(x,y)\big(1 - F(y,x)\big) \\ &= 1 - p(x,x^-) - \sum_{y\,:\,y^-=x} p(x,y)F(y,x) \\ &= -p(x,x^-) + \frac{p(x,x^-)}{F(x,x^-)}\end{aligned}$$

by Proposition 9.3 (b), and the formula follows. □

Exercise 9.58. We know that on $\mathbb{T}_q$, the function $F(y,x|z)$ is the same for all pairs of neighbours x, y. It coincides with the function $F(1,0|z)$ of the factor chain $(|Z_n|)_{n\geq 0}$ on $\mathbb{N}_0$, which is the infinite drunkard's walk with reflecting barrier at 0 and "forward" transition probability $p = q/(q+1)$. Therefore

$$F(y,x|z) = \frac{q+1}{2qz}\Big(1 - \sqrt{1-\rho^2 z^2}\Big), \quad \text{where } \rho = \frac{2\sqrt{q}}{q+1}.$$

Using binomial expansion, we get

$$F(y,x|z) = \frac{1}{\sqrt{q}} \sum_{n=1}^{\infty} (-1)^{n-1} \binom{1/2}{n} (\rho\, z)^{2n-1}.$$

Therefore

$$f^{(2n-1)}(y,x) = \frac{(\rho/2)^{2n-1}}{n\sqrt{q}} \binom{2n-2}{n-1}.$$

With these values,

$$\mathsf{Pr}_o[W_{k+1} = y,\ \epsilon_{k+1} = m + 2n - 1 \mid W_k = x,\ \epsilon_k = m] = f^{(2n-1)}(y,x). \quad \square$$

Exercise 9.62. Let ξ be an end that satisfies the criterion of Corollary 9.61. Let $\pi(o,\xi) = [o = x_0, x_1, x_2, \dots]$. Suppose that ξ is recurrent. Then there is k such that $F(x_{n+1}, x_n) = 1$ for all $n \geq 1$. Since all involved properties are independent of the base point, we may suppose without loss of generality that $k = 0$. Write $x_1 = x$. Then the random walk on the branch $B_{[o,x]}$ with transition matrix $P_{[o,x]}$ is recurrent. By reversibility, we can view that branch as a recurrent network in the sense of Section 4.D. It contains the ray $\pi(o,\xi)$ as an induced subnetwork, recurrent by Exercise 4.54. The latter inherits the conductances $a(x_{i-1}, x_i)$ of its edges from the branch, and the transition probabilities on the ray satisfy

$$\frac{p_{\text{ray}}(x_i, x_{i-1})}{p_{\text{ray}}(x_i, x_{i+1})} = \frac{a(x_i, x_{i-1})}{a(x_i, x_{i+1})} = \frac{p(x_i, x_{i-1})}{p(x_i, x_{i+1})}.$$

But since

$$\sum_{k=1}^{\infty} \prod_{i=1}^{k} \frac{p_{\mathrm{ray}}(x_i, x_{i-1})}{p_{\mathrm{ray}}(x_i, x_{i+1})} < \infty,$$

this birth-and-death chain is transient by Theorem 5.9 (i), a contradiction. □

Exercise 9.70. We can consider the branch $B_{[v,w]}$ and the associated random walk $P_{[v,w]}$. We know that we can replace $g = g_o$ with g_v in the assumption of the exercise. With v in the place of the root, we can apply Theorem 9.69 (a): finiteness of g_v on $\partial T_{v,w}$ implies that $P_{[v,w]}$ is transient. Therefore $F(w, v) < 1$, so that also P on T is transient.

Furthermore, we can apply the same argument to any sub-branch $B_{[x,y]}$ of $B_{[v,w]}$, where x is closer to v than y, and get $F(y, x) < 1$. Thus, if $\xi \in \partial T_{v,w}$ and $\pi(v, \xi) = [v = x_0, x_1, x_2, \dots]$ then $F(x_n, x_{n-1}) < 1$ for all $n \geq 1$: the end ξ is transient. □

Exercise 9.73. We can order the p_i such that $p_{i+1} \geq p_i$ for all i. Then also $p_i/(1 - p_i) \geq p_j/(1 - p_j)$ whenever $i \geq j$. In particular, we have

$$\lambda = \frac{p_1}{1 - p_1} \frac{p_2}{1 - p_2}, \quad \text{where } \lambda = \max\Big\{ \frac{p_i}{1 - p_i} \frac{p_j}{1 - p_j} : i, j \in \mathcal{I}, i \neq j \Big\}.$$

Now the inequality $p_1 + p_2 < 1$ readily implies that $\lambda < 1$. If $x \in T \setminus \{o\}$ and $\pi(0, x) = [o = x_0, x_1, \dots, x_n]$ then for $i = 1, \dots, n - 1$,

$$\frac{p(x_i, x_{i-1})}{1 - p(x_i, x_{i-1})} \frac{p(x_{i+1}, x_i)}{1 - p(x_{i+1}, x_i)} \leq \lambda,$$

so that

$$\prod_{i=1}^{k} \frac{p(x_i, x_{i-1})}{1 - p(x_i, x_{i-1})} \leq \begin{cases} \lambda^{k/2}, & \text{if } k \text{ is even,} \\ \frac{p_1}{1-p_1} \lambda^{(k-1)/2}, & \text{if } k \text{ is odd.} \end{cases}$$

It follows that g is bounded by $M = 1\big/\big((1 - p_1)(1 - \sqrt{\lambda}\,)\big)$. □

Exercise 9.80. This computation of the largest eigenvalue of 2×2 matrices is left entirely to the reader.

Exercise 9.82. Variant 1. In Example 9.47, we have $s = q + 1$ and $p_i = 1/(q + 1)$. The function $\Phi(t)$ becomes

$$\Phi(t) = \frac{1}{2}\Big(\sqrt{(q + 1)^2 + 4t^2} - (q - 1) \Big)$$

and the unique positive solution of the equation $\Phi'(t) = \Phi(t)/t$ is easily computed:

$$\rho(P) = 2\sqrt{q}\big/(q + 1).$$

Variant 2. $\bar{Z}_n = |Z_n|$ is the infinite drunkard's walk on $\mathbb{N}_0$ (reflecting at state 0) with "forward probability" $p = q/(q+1)$, where $q+1$ is the vertex degree. In this example, $G(o,o|z) = \bar{G}(0,0|z)$, because in our factor chain, the class corresponding to the state 0 has only the vertex o in its preimage under the natural projection. But $\bar{G}(0,0|z)$ is the function $G_{\mathbb{N}}(0,0|z)$ computed in Example 5.23. (Attention: the q of that example is $1-p$.) That is,

$$G(o,o|z) = \frac{2q}{q-1+\sqrt{(q+1)^2-4qz^2}}.$$

The smallest positive singularity of this function is $\mathsf{r}(P) = (q+1)\big/\big(2\sqrt{q}\big)$, and $\rho(P) = 1/\mathsf{r}(P)$. □

Exercise 9.84. Let $C_n = \{x \in T : |x| = n\}$, $n \geq 0$, be the classes corresponding to the projection $x \mapsto |x|$. Then $C_0 = \{o\}$ and $\bar{p}(0,1) = p(o, C_1) = 1$. For $n \geq 1$ and $x \in C_n$, we have

$$\bar{p}(n, n-1) = p(x, x^-) = 1 - \alpha$$

and

$$\bar{p}(n, n+1) = p(x, C_{n+1}) = 1 - p(x, x^-) = \alpha.$$

These numbers are independent of the specific choice of $x \in C_n$, so that we have indeed a factor chain. The latter is the infinite drunkard's walk on $\mathbb{N}_0$ (reflecting at state 0) with "forward probability" $p = \alpha$. The relevant computations can be found in Example 5.23. □

Exercise 9.88. (1) We prove the following by induction on n.

- For all $k_1, \dots, k_n \in \mathbb{N}$ and $x \in X \subset X_{(N)}$,

$$\begin{aligned}\mathsf{Pr}_x[\boldsymbol{t}_1 = k_1,\ \boldsymbol{t}_2 = k_1 + k_2, \dots, \boldsymbol{t}_n = k_1 + \cdots + k_n] \\ = f^{(k_1)}(0,N)\, f^{(k_2)}(0,N) \cdots f^{(k_n)}(0,N),\end{aligned}$$

where $f^{(k_i)}(0,N)$ refers to SRW on the integer interval $\{0, \dots, N\}$. This implies immediately that the increments of the stopping times are i.i.d.

For $n = 1$, we have already proved that formula. Suppose that it is true for $n-1$. We have, using the strong Markov property,

$$\begin{aligned}&\mathsf{Pr}_x[\boldsymbol{t}_1 = k_1,\ \boldsymbol{t}_2 = k_1 + k_2, \dots, \boldsymbol{t}_n = k_1 + \cdots + k_n] \\ &\quad = \sum_{y \in X} \mathsf{Pr}_x[\boldsymbol{t}_1 = k_1, Z_{k_1} = y] \times \\ &\qquad \times \mathsf{Pr}_x[\boldsymbol{t}_2 = k_1 + k_2, \dots, \boldsymbol{t}_n = k_1 + \cdots + k_n \mid \boldsymbol{t}_1 = k_1, Z_{k_1} = y] \overset{(*)}{=}\end{aligned}$$

$$
\begin{aligned}
&\overset{(*)}{=} \sum_{y\in X} \mathsf{Pr}_x[\boldsymbol{t}_1 = k_1, Z_{k_1} = y] \\
&\quad \times \mathsf{Pr}_y[\boldsymbol{t}_1 = k_2,\ \boldsymbol{t}_2 = k_2 + k_3, \ldots, \boldsymbol{t}_{n-1} = k_2 + \cdots + k_n] \\
&= \sum_{y\in X} \mathsf{Pr}_x[\boldsymbol{t}_1 = k_1, Z_{k_1} = y]\ f^{(k_2)}(0,N)\cdots f^{(k_n)}(0,N) \\
&= \mathsf{Pr}_x[\boldsymbol{t}_1 = k_1]\, f^{(k_2)}(0,N)\cdots f^{(k_n)}(0,N) \\
&= f^{(k_1)}(0,N)\, f^{(k_2)}(0,N)\cdots f^{(k_n)}(0,N).
\end{aligned}
$$

We remark that $(*)$ holds since $\boldsymbol{t}_k$ is the stopping time of the $(k-1)$-st visit in a point in X distinct from the previous one after the time $\boldsymbol{t}_1$.

If the subdivision is arbitrary, then the distribution of $\boldsymbol{t}_1$ depends on the starting point in X. Therefore the increments cannot be identically distributed. They also cannot be independent, since in this situation, the distribution of $\boldsymbol{t}_2 - \boldsymbol{t}_1$ depends on the point $Z_{\boldsymbol{t}_1}$.

(2) We know that the statement is true for $k = 1$. Suppose it holds for $k-1$. Then, again by the strong Markov property,

$$
\begin{aligned}
\mathsf{Pr}_x[\boldsymbol{t}_k = n,\ Z_n = y] &= \sum_{m=0}^{n} \sum_{v\in X} \mathsf{Pr}_x[\boldsymbol{t}_1 = m,\ Z_m = v,\ \boldsymbol{t}_k = n,\ Z_n = y] \\
&= \sum_{m=0}^{n} \sum_{v\in X} \mathsf{Pr}_x[\boldsymbol{t}_1 = m,\ Z_m = v]\ \mathsf{Pr}_v[\boldsymbol{t}_{k-1} = n-m,\ Z_{n-m} = y].
\end{aligned}
$$

(Note that in reality, we cannot have $m = 0$ or $n = 0$; the associated probabilities are 0.) We deduce, using the product formula for power series,

$$
\begin{aligned}
&\sum_{n=0}^{\infty} \mathsf{Pr}_x\big[\boldsymbol{t}_k = n,\ Z_n = y\big]\, z^n \\
&\quad = \sum_{v\in X} \sum_{n=0}^{\infty} \sum_{m=0}^{n} \mathsf{Pr}_x[\boldsymbol{t}_1 = m,\ Z_m = v]\ \mathsf{Pr}_v[\boldsymbol{t}_{k-1} = n-m,\ Z_{n-m} = y]\, z^n \\
&\quad = \sum_{v\in X} \Big(\sum_{m=0}^{\infty} \mathsf{Pr}_x[\boldsymbol{t}_1 = m,\ Z_m = v]\, z^m \Big) \Big(\sum_{n=0}^{\infty} \mathsf{Pr}_v[\boldsymbol{t}_{k-1} = n,\ Z_n = y]\, z^n \Big) \\
&\quad = \sum_{v\in X} \big(p(x,v)\,\phi(z) \big)\, \big(p^{(k-1)}(v,y)\,\phi(z)^{k-1} \big),
\end{aligned}
$$

which yields the stated formula. □

Exercise 9.91. We notice that the reversing measure $\widetilde{\mathsf{m}}$ of SRW on $\widetilde{X}$ satisfies $\widetilde{\mathsf{m}}|_X = \mathsf{m}$. The resistance of every edge is $= 1$. Therefore

$$(\tilde f, \tilde f)_{\widetilde\Gamma} = \sum_{\tilde x \in \widetilde X} \tilde f(\tilde x)^2\, \widetilde{\mathsf{m}}(\tilde x) \ge \sum_{x \in X \subset \widetilde X} \tilde f(\tilde x)^2\, \widetilde{\mathsf{m}}(\tilde x) = (f, f)_\Gamma.$$

Along any edge of the inserted path of the subdivision that replaces an original edge $[x, y]$ of X, the difference of $\tilde f$ is 0, with precisely one exception, where the difference is $f(y) - f(x)$. Thus, the contribution of that inserted path to $D_{\widetilde\Gamma}(\tilde f)$ is $\big(f(y) - f(x)\big)^2$. Therefore $D_{\widetilde\Gamma}(\tilde f) = D_\Gamma(f)$. □

Exercise 9.93. Conditions (i) and (ii) imply yield that

$$D_P(f) \ge \varepsilon \cdot D_T(f) \quad \text{and} \quad (f, f)_P \le M \cdot (f, f)_T \quad \text{for every } f \in \ell_0(X).$$

Here, the index P obviously refers to the reversible Markov chain with transition matrix P and associated reversing measure m, while the index T refers to SRW. We have already seen that $(f, f)_P - (Pf, f)_P = D_P(f)$. Therefore, since

$$\rho(P) = \sup\left\{\frac{(Pf, f)_P}{(f, f)_P} : f \in \ell_0(X),\ f \ne 0\right\},$$

we get

$$\begin{aligned} 1 - \rho(P) &= \inf\left\{\frac{D_P(f)}{(f, f)_P} : f \in \ell_0(X),\ f \ne 0\right\} \\ &\ge \frac{\varepsilon}{M} \inf\left\{\frac{D_T(f)}{(f, f)_T} : f \in \ell_0(X),\ f \ne 0\right\} = \frac{\varepsilon}{M}\big(1 - \rho(T)\big), \end{aligned}$$

as proposed. □

Exercise 9.96. We use the cone types. There are two of them. Type -1 corresponds to T_x, where $x \in \pi(o, \varpi)$, $x \ne o$. That is, ∂T_x contains ϖ. Type 1 corresponds to any T_y that "looks downwards" in Figure 37, so that T_y is a q-ary tree rooted at x, and $\varpi \notin \partial T_x$.

If x has type -1, then T_x contains precisely one neighbour of x with the same type, namely $x^\sim$, and $p(x, x^\sim) = 1 - \alpha$. Also, T_x contains $q - 1$ neighbours y of x with type 1, and $p(x, y) = \alpha/q$ for each of them.

If y has type 1, then all of its q neighbours $w \in T_y$ also have type one, and $p(y, w) = \alpha/q$ for each of them.

Therefore

$$A = \begin{pmatrix} q\frac{1-\alpha}{\alpha} & q - 1 \\ 0 & \frac{\alpha}{1-\alpha} \end{pmatrix}.$$

The two eigenvalues are the elements in the principal diagonal of A, and at least one of them is > 1. This yields transience. □

Exercise 9.97. The functions $F_-(z) = F(x, x^\sim|z)$ and $F_+(z) = F(x^\sim, x|z)$ are independent of x, compare with Exercise 2.12. Proposition 9.3 leads to the two quadratic equations

$$F_-(z) = (1-\alpha)\, z + \alpha\, z\, F_-(z)^2 \quad \text{and}$$
$$F_+(z) = \frac{\alpha}{q}\, z + (q-1)\, \frac{\alpha}{q}\, z\, F_-(z) F_+(z) + (1-\alpha)\, z\, F\, +\, (z)^2.$$

Since $F_-(0) = 0$, the right one of the two solutions of the quadratic equation for $F_-(z)$ is

$$F_-(z) = \frac{1}{2\alpha z}\Big(1 - \sqrt{1 - 4\alpha(1-\alpha)z^2}\Big).$$

One next has to solve the quadratic equation for $F_+(z)$. By the same argument, the right solution is the one with the minus sign in front of the square root. We let $F_- = F_-(1)$ and $F_+ = F_+(1)$. In the end, we find after elementary computations

$$F_- = \begin{cases} \dfrac{1-\alpha}{\alpha} & \text{if } \alpha \geq \dfrac{1}{2}, \\ 1 & \text{if } \alpha \leq \dfrac{1}{2}, \end{cases} \quad \text{and} \quad F_+ = \begin{cases} \dfrac{1}{q} & \text{if } \alpha \geq \dfrac{1}{2}, \\ \dfrac{\alpha}{(1-\alpha)q} & \text{if } \alpha \leq \dfrac{1}{2}. \end{cases}$$

We see that when $\alpha > 1/2$ then $F_-(1) < 1$ and $F_+(1) < 1$. Then the Green kernel vanishes at infinity, and the Dirichlet problem is solvable. On the other hand, when $\alpha \leq 1/2$ then the random walk converges almost surely to ϖ, so that supp ν_o does not coincide with the full boundary: the Dirichlet problem at infinity does not admit solution in this case.

We next compute

$$U(x,x) = U = \alpha\, F_- + (1-\alpha)\, F_+ = \min\{\alpha, 1-\alpha\}\frac{q+1}{q}.$$

For $x, y \in \mathbb{T}_q$, let $v = v(x,y)$ be the point on $\pi(x,y)$ which minimizes $\mathrm{hor}(v) = \mathrm{hor}(v, \varpi)$. In other words, this is the first common point on the geodesic rays $\pi(x, \varpi)$ and $\pi(y, \varpi)$ – the *confluent* of x and y with respect to ϖ. (Recall that $x \wedge y$ is the confluent with respect to o.) With this notation,

$$G(x,y) = F_-^{d(x,v)}\, F_+^{d(y,v)}\, \frac{1}{1-U}.$$

We now compute the Martin kernel. It is immediate that

$$K(x,\varpi) = F_-(1)^{\mathrm{hor}(x)} = \begin{cases} \left(\dfrac{1-\alpha}{\alpha}\right)^{\mathrm{hor}(x)} & \text{if } \alpha \geq \dfrac{1}{2}, \\ 1 & \text{if } \alpha \leq \dfrac{1}{2}. \end{cases}$$

We leave to the reader the geometric reasoning that leads to the formula for the Martin kernel at $\xi \in \partial\mathbb{T}_q \setminus \{\varpi\}$: setting $\lambda = \sqrt{F_- F_+}$,

$$K(x,\xi) = K(x,\varpi)\,\lambda^{\mathrm{hor}(x,\xi)-\mathrm{hor}(x)},$$

where $\lambda = \min\left\{\frac{1-\alpha}{\alpha q},\ \frac{\alpha}{(1-\alpha)q}\right\}^{1/2}$ □

Exercise 9.101. We can rewrite

$$D(1)^{-1}D'(1) = \big(I - D(1)AD(1)\big)^{-1}DB^{-1}$$

and, with a few transformations,

$$\big(I - D(1)\big)^{-1}\big(I - D(1)AD(1)\big) = \big(I - QD(1)\big)\big(I - D(1)\big)^{-1}.$$

The last identity implies

$$\sigma\,\big(I - D(1)\big)^{-1}\big(I - D(1)AD(1)\big) = \sigma.$$

Therefore, if $\mathbf{1}$ denotes the column vector over $\mathcal{I}$ with all entries $= 1$, then

$$\frac{1}{\ell} = \sigma\,D(1)^{-1}D'(1)\,\mathbf{1} = \sigma\,\big(I - D(1)\big)^{-1}DB^{-1}\,\mathbf{1},$$

which is the proposed formula. □

Bibliography

A Textbooks and other general references

[A-N] Athreya, K. B., Ney, P. E., *Branching processes.* Grundlehren Math. Wiss. 196, Springer-Verlag, New York 1972. 134

[B-M-P] Baldi, P., Mazliak, L., Priouret, P., *Martingales and Markov chains. Solved exercises and elements of theory.* Translated from the 1998 French original, Chapman & Hall/CRC, Boca Raton 2002. xiii

[Be] Behrends, E., *Introduction to Markov chains. With special emphasis on rapid mixing.* Adv. Lectures Math., Vieweg, Braunschweig 2000. xiii

[Br] Brémaud, P., *Markov chains. Gibbs fields, Monte Carlo simulation, and queues*. Texts Appl. Math. 31, Springer-Verlag, New York 1999. xiii, 6

[Ca] Cartier, P., Fonctions harmoniques sur un arbre. *Symposia Math.* **9** (1972), 203–270. xii, 250, 262

[Ch] Chung, K. L., *Markov chains with stationary transition probabilities*. 2nd edition. Grundlehren Math. Wiss. 104, Springer-Verlag, New York 1967. xiii

[Di] Diaconis, P., *Group representations in probability and statistics*. IMS Lecture Notes Monogr. Ser. 11, Institute of Mathematical Statistics, Hayward, CA, 1988. 83

[D-S] Doyle, P. G., Snell, J. L., *Random walks and electric networks*. Carus Math. Monogr. 22, Math. Assoc. America, Washington, DC, 1984. xv, 78, 80, 231

[Dy] Dynkin, E. B., Boundary theory of Markov processes (the discrete case). *Russian Math. Surveys* **24** (1969), 1–42. xv, 191, 212

[F-M-M] Fayolle, G., Malyshev, V. A., Men'shikov, M. V., *Topics in the constructive theory of countable Markov chains.* Cambridge University Press, Cambridge 1995. x

[F1] Feller, W., *An introduction to probability theory and its applications, Vol.* I. 3rd edition, Wiley, New York 1968. x

[F2] Feller, W., *An introduction to probability theory and its applications, Vol.* II. 2nd edition, Wiley, New York 1971. x

[F-S] Flajolet, Ph., Sedgewick, R., *Analytic combinatorics.* Cambridge University Press, Cambridge 2009. xiv

[Fr] Freedman, D., *Markov chains.* Corrected reprint of the 1971 original, Springer-Verlag, New York 1983. xiii

[Hä] Häggström, O., *Finite Markov chains and algorithmic applications*. London Math. Soc. Student Texts 52, Cambridge University Press, 2002. xiii

[H-L] Hernández-Lerma, O., Lasserre, J. B., *Markov chains and invariant probabilities.* Progr. Math. 211, Birkhäuser, Basel 2003. xiv

[Hal] Halmos, P., *Measure theory*. Van Nostrand, New York, N.Y., 1950. 8, 9

[Har] Harris, T. E., *The theory of branching processes*. Corrected reprint of the 1963 original, Dover Phoenix Ed., Dover Publications, Inc., Mineola, NY, 2002. 134

[Hi] Hille, E., *Analytic function theory, Vols.* I–II. Chelsea Publ. Comp., New York 1962 130

[I-M] Isaacson, D. L., Madsen, R. W., *Markov chains. Theory and applications*. Wiley Ser. Probab. Math. Statist., Wiley, New York 1976. xiii, 6, 53

[J-T] Jones, W. B., Thron, W. J., *Continued fractions. Analytic theory and applications.* Encyclopedia Math. Appl. 11, Addison-Wesley, Reading, MA, 1980. 126

[K-S] Kemeny, J. G., Snell, J. L., *Finite Markov chains*. Reprint of the 1960 original, Undergrad. Texts Math., Springer-Verlag, New York 1976. ix, xiii, 1

[K-S-K] Kemeny, J. G., Snell, J. L., Knapp, A. W., *Denumerable Markov chains*. 2nd edition, Grad. Texts in Math. 40, Springer-Verlag, New York 1976. xiii, xiv, 189

[La] Lawler, G. F., *Intersections of random walks.* Probab. Appl., Birkhäuser, Boston, MA, 1991. x, xv

[Le] Lesigne, Emmanuel, *Heads or tails. An introduction to limit theorems in probability.* Translated from the 2001 French original by A. Pierrehumbert, Student Math. Library 28, Amer. Math. Soc., Providence, RI, 2005. 115

[Li] Lindvall, T., *Lectures on the coupling method.* Corrected reprint of the 1992 original, Dover Publ., Mineola, NY, 2002. x, 63

[L-P] Lyons, R. with Peres, Y., *Probability on trees and networks*. Book in preparation; available online at http://mypage.iu.edu/~rdlyons/prbtree/prbtree.html xvi, 78, 80, 134

[No] Norris, J. R., *Markov chains*. Cambridge Ser. Stat. Probab. Math. 2, Cambridge University Press, Cambridge 1998. xiii, 131

[Nu] Nummelin, E., *General irreducible Markov chains and non-negative operators.* Cambridge Tracts in Math. 83, Cambridge University Press, Cambridge 1984. xiv

[Ol] Olver, F. W. J., *Asymptotics and special functions*. Academic Press, San Diego, CA, 1974. 128

[Pa] Parthasarathy, K. R., *Probability measures on metric spaces*. Probab. and Math. Stat. 3, Academic Press, New York, London 1967. 10, 202, 217

[Pe] Petersen, K., *Ergodic theory*. Cambridge University Press, Cambridge 1990. 69

[Pi] Pintacuda, N., *Catene di Markov*. Edizioni ETS, Pisa 2000. xiii

[Ph] Phelps, R. R., *Lectures on Choquet's theorem*. Van Nostrand Math. Stud. 7, New York, 1966. 184

[Re] Revuz, D., *Markov chains*. Revised edition, Mathematical Library 11, North-Holland, Amsterdam 1984. xiii, xiv

[Ré] Révész, P., *Random walk in random and nonrandom environments*. World Scientific, Teaneck, NJ, 1990. x

[Ru] Rudin, W., *Real and complex analysis*. 3rd edition, McGraw-Hill, New York 1987. 105

[SC] Saloff-Coste, L., Lectures on finite Markov chains. In *Lectures on probability theory and statistics,* École d'eté de probabilités de Saint-Flour XXVI-1996 (ed. by P. Bernard), Lecture Notes in Math. 1665, Springer-Verlag, Berlin 1997, 301–413. x, xiii, 78, 83

[Se] Seneta, E., *Non-negative matrices and Markov chains*. Revised reprint of the second (1981) edition. Springer Ser. Statist. Springer-Verlag, New York 2006. ix, x, xiii, 6, 40, 53, 57, 75

[So] Soardi, P. M., *Potential theory on infinite networks.* Lecture Notes in Math. 1590, Springer-Verlag, Berlin 1994. xv

[Sp] Spitzer, F., *Principles of random walks*. 2nd edition, Grad. Texts in Math. 34, Springer-Verlag, New York 1976. x, xv, 115

[St1] Stroock, D., *Probability theory, an analytic view*. 2nd ed., Cambridge University Press, New York 2000. xiv

[St2] Stroock, D., *An introduction to Markov processes*. Grad. Texts in Math. 230, Springer-Verlag, Berlin 2005. xiii

[Wa] Wall, H. S., *Analytic theory of continued fractions*. Van Nostrand, Toronto 1948. 126

[W1] Woess, W., *Catene di Markov e teoria del potenziale nel discreto*. Quaderni dell'Unione Matematica Italiana 41, Pitagora Editrice, Bologna 1996. v, x, xi, xiv

[W2] Woess, W., *Random walks on infinite graphs and groups*. Cambridge Tracts in Mathematics 138, Cambridge University Press, Cambridge 2000 (paperback reprint 2008). vi, x, xv, xvi, 78, 151, 191, 207, 225, 287

B Research-specific references

[1] Amghibech, S., Criteria of regularity at the end of a tree. In *Séminaire de probabilités XXXII,* Lecture Notes in Math. 1686, Springer-Verlag, Berlin 1998, 128–136. 261

[2] Askey, R., Ismail, M., Recurrence relations, continued fractions, and orthogonal polynomials. *Mem. Amer. Math. Soc.* **49** (1984), no. 300. 126

[3] Bajunaid, I., Cohen, J. M., Colonna, F., Singman, D., Trees as Brelot spaces. *Adv. in Appl. Math.* **30** (2003), 706–745. 273

[4] Bayer, D., Diaconis, P., Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.* **2** (1992), 294–313. 98

[5] Benjamini, I., Peres, Y., Random walks on a tree and capacity in the interval. *Ann. Inst. H. Poincaré Prob. Stat.* **28** (1992), 557–592. 261

[6] Benjamini, I., Peres, Y., Markov chains indexed by trees. *Ann. Probab.* **22** (1994), 219–243. 147

[7] Benjamini, I., Schramm, O., Random walks and harmonic functions on infinite planar graphs using square tilings. *Ann. Probab.* **24** (1996), 1219–1238. 269

[8] Blackwell, D., On transient Markov processes with a countable number of states and stationary transition probabilities. *Ann. Math. Statist.* **26** (1955), 654–658. 219

[9] Cartwright, D. I., Kaimanovich, V. A., Woess, W., Random walks on the affine group of local fields and of homogeneous trees. *Ann. Inst. Fourier (Grenoble)* **44** (1994), 1243–1288. 292

[10] Cartwright, D. I., Soardi, P. M., Woess, W., Martin and end compactifications of non locally finite graphs. *Trans. Amer. Math. Soc.* **338** (1993), 679–693. 250, 261

[11] Choquet, G., Deny, J., Sur l'équation de convolution $\mu = \mu * \sigma$. *C. R. Acad. Sci. Paris* **250** (1960), 799–801. 224

[12] Chung K. L., Ornstein, D., On the recurrence of sums of random variables. *Bull. Amer. Math. Soc.* **68** (1962), 30–32. 113

[13] Derriennic, Y., Marche aléatoire sur le groupe libre et frontière de Martin. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** (1975), 261–276. 250

[14] Diaconis, P., Shahshahani, M., Generating a random permutation with random transpositions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **57** (1981), 159–179. 101

[15] Diaconis, P., Stroock, D., Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1** (1991), 36–61. x, 93

[16] Di Biase, F., *Fatou type theorems. Maximal functions and approach regions*. Progr. Math. 147, Birkhäuser, Boston 1998. 262

[17] Doob, J. L., Discrete potential theory and boundaries. *J. Math. Mech.* **8** (1959), 433–458. 182, 189, 206, 217, 261

[18] Doob, J. L., Snell, J. L.,Williamson, R. E., Application of boundary theory to sums of independent random variables. In *Contributions to probability and statistics*, Stanford University Press, Stanford 1960, 182–197. 224

[19] Dynkin, E. B., Malyutov, M. B., Random walks on groups with a finite number of generators. *Soviet Math. Dokl.* **2** (1961), 399–402. 219, 250

[20] Erdös, P., Feller, W., Pollard, H., A theorem on power series. *Bull. Amer. Math. Soc.* **55** (1949), 201–204. x

[21] Figà-Talamanca, A., Steger, T., Harmonic analysis for anisotropic random walks on homogeneous trees. *Mem. Amer. Math. Soc.* **110** (1994), no. 531. 250

[22] Gairat, A. S., Malyshev, V. A., Men'shikov, M. V., Pelikh, K. D., Classification of Markov chains that describe the evolution of random strings. *Uspekhi Mat. Nauk* **50** (1995), 5–24; English translation in *Russian Math. Surveys* **50** (1995), 237–255. 278

[23] Gantert, N., Müller, S., The critical branching Markov chain is transient. *Markov Process. Related Fields* **12** (2006), 805–814. 147

[24] Gerl, P., Continued fraction methods for random walks on $\mathbb{N}$ and on trees. In *Probability measures on groups, VII*, Lecture Notes in Math. 1064, Springer-Verlag, Berlin 1984, 131–146. 126

[25] Gerl, P., Rekurrente und transiente Bäume. In *Sém. Lothar. Combin.* (IRMA Strasbourg) 10 (1984), 80–87; available online at http://www.emis.de/journals/SLC/ 269

[26] Gilch, L., Rate of escape of random walks on free products. *J. Austral. Math. Soc.* **83** (2007), 31–54. 267

[27] Gilch, L., Müller, S., Random walks on directed covers of graphs. Preprint, 2009. 273

[28] Good, I. J., Random motion and analytic continued fractions. *Proc. Cambridge Philos. Soc.* **54** (1958), 43–47. 126

[29] Guivarc'h, Y., Sur la loi des grands nombres et le rayon spectral d'une marche aléatoire. *Astérisque* **74** (1980), 47–98. 151

[30] Harris T. E., First passage and recurrence distributions. *Trans. Amer. Math. Soc.* **73** (1952), 471–486. 140

[31] Hennequin, P. L., Processus de Markoff en cascade. *Ann. Inst. H. Poincaré* **18** (1963), 109–196. 224

[32] Hunt, G. A., Markoff chains and Martin boundaries. *Illinois J. Math.* **4** (1960), 313–340. 189, 191, 206, 218

[33] Kaimanovich, V. A., Lyapunov exponents, symmetric spaces and a multiplicative ergodic theorem for semisimple Lie groups. *J. Soviet Math.* **47** (1989), 2387–2398. 292

[34] Karlin, S., McGregor, J., Random walks. *Illinois J. Math.* **3** (1959), 66–81. 126

[35] Kemeny, J. G., Snell, J. L., Boundary theory for recurrent Markov chains. *Trans. Amer. Math. Soc.* **106** (1963), 405–520. 187

[36] Kingman, J. F. C., The ergodic theory of Markov transition probabilities. *Proc. London Math. Soc.* **13** (1963), 337–358.

[37] Knapp, A. W., Regular boundary points in Markov chains. *Proc. Amer. Math. Soc.* **17** (1966), 435–440. 261

[38] Korányi, A., Picardello, M. A., Taibleson, M. H., Hardy spaces on nonhomogeneous trees. With an appendix by M. A. Picardello and W. Woess. *Symposia Math.* **29** (1987), 205–265. 250

[39] Lalley, St., Finite range random walk on free groups and homogeneous trees. *Ann. Probab.* **21** (1993), 2087–2130. 267

[40] Ledrappier, F., Asymptotic properties of random walks on free groups. In *Topics in probability and Lie groups: boundary theory,* CRM Proc. Lecture Notes 28, Amer. Math. Soc., Providence, RI, 2001, 117–152. 267

[41] Lyons, R., Random walks and percolation on trees. *Ann. Probab.* **18** (1990), 931–958. 278

[42] Lyons, T., A simple criterion for transience of a reversible Markov chain. *Ann. Probab.* **11** (1983), 393–402. 107

[43] Mann, B., How many times should you shuffle a deck of cards? In *Topics in contemporary probability and its applications,* Probab. Stochastics Ser., CRC, Boca Raton, FL, 1995, 261–289. 98

[44] Nagnibeda, T., Woess, W., Random walks on trees with finitely many cone types. *J. Theoret. Probab.* **15** (2002), 383–422. 267, 278, 295

[45] Ney, P., Spitzer, F., The Martin boundary for random walk. *Trans. Amer. Math. Soc.* **121** (1966), 116–132. 224, 225

[46] Picardello, M. A., Woess, W., Martin boundaries of Cartesian products of Markov chains. *Nagoya Math. J.* **128** (1992), 153–169. 207

[47] Pólya, G., Über eine Aufgabe der Wahrscheinlichkeitstheorie betreffend die Irrfahrt im Straßennetz. *Math. Ann.* **84** (1921), 149–160. 112

[48] Soardi, P. M., Yamasaki, M., Classification of infinite networks and its applications. *Circuits, Syst. Signal Proc.* **12** (1993), 133–149. 107

[49] Vere-Jones, D., Geometric ergodicity in denumerable Markov chains. *Quart. J. Math. Oxford* **13** (1962), 7–28. 75

[50] Yamasaki, M., Parabolic and hyperbolic infinite networks. *Hiroshima Math. J.* **7** (1977), 135–146. 107

[51] Yamasaki, M., Discrete potentials on an infinite network. *Mem. Fac. Sci., Shimane Univ.* **13** (1979), 31–44.

[52] Woess, W., Random walks and periodic continued fractions. *Adv. in Appl. Probab.* **17** (1985), 67–84. 126

[53] Woess, W., Generating function techniques for random walks on graphs. In *Heat kernels and analysis on manifolds, graphs, and metric spaces* (Paris, 2002), Contemp. Math. 338, Amer. Math. Soc., Providence, RI, 2003, 391–423. xiv

List of symbols and notation

This list contains a selection of the most important symbols and notation.

Numbers

$\mathbb{N} = \{1, 2, \dots\}$	the natural numbers (positive integers)
$\mathbb{N}_0 = \{0, 1, 2, \dots\}$	the non-negative integers
$\mathbb{N}_{\text{odd}} = \{1, 3, 5, \dots\}$	the odd natural numbers
$\mathbb{Z}$	the integers
$\mathbb{R}$	the real numbers
$\mathbb{C}$	the complex numbers

Markov chains

X	typical symbol for the state space of a Markov chain
P, also Q	typical symbols for transition matrices
C, $C(x)$	irreducible class (of state x)
$G(x, y\|z)$	Green function (1.32)
$F(x, y\|z)$, $U(x, y\|z)$	first passage time generating functions (1.37)
$L(x, y\|z)$	generating function of "last exit" probabilities (3.56)

Probability space ingredients

Ω	trajectory space, see (1.8)
$\mathcal{A}$	σ-algebra generated by all cylinder sets (1.9)
Pr_x and Pr_ν	probability measure on the trajectory space with respect to starting point x, resp. initial distribution ν, see (1.10)
$\mathsf{Pr}(\cdot)$	probability of a set in the σ-algebra
$\mathsf{Pr}[\cdot]$	probability of an event described by a logical expression
$\mathsf{Pr}[\cdot \mid \cdot]$	conditional probability
E	expectation (expected value)

Random times

$\boldsymbol{s}, \boldsymbol{t}$	typical symbols for stopping times (1.24)
$\boldsymbol{s}^W, \boldsymbol{s}^x$	first passage time = time (≥ 0) of first visit to the set W, resp. the point x (1.26)
$\boldsymbol{t}^W, \boldsymbol{t}^x$	time (≥ 1) of first visit to the set W, resp. the point x *after the start*
$\boldsymbol{\epsilon}_V, \boldsymbol{\epsilon}_k$	exit time from the set V, resp. (in a tree) from the ball with radius k around the starting point (7.34), (9.19)

Measures

μ, ν	typical symbols for measures on X, $\mathbb{R}$, $\mathbb{Z}^d$, a group, etc.
$\bar{\mu}$	mean or mean vector of the probability measure μ on $\mathbb{Z}$ or $\mathbb{Z}^d$
supp	support of a measure or function
m_C	the stationary probability measure of a positive recurrent essential class of a Markov chain (3.19)
m	(a) the stationary probability measure of a positive recurrent, irreducible Markov chain, (b) the reversing measure of a reversible Markov chain, not necessarily with finite mass (4.1)

Graphs, trees

Γ	typical notation for a (usually directed) graph
$V(\Gamma)$	vertex set of Γ
$E(\Gamma)$	(oriented) edge set of Γ
$\Gamma(P)$	graph of the Markov chain with transition matrix P (1.6)
T	typical notation for a tree
$\mathbb{T}_q$	homogenous tree with degree $q+1$
π	according to context, (a) path in a graph, (b) projection map, or (c) 3.14159...
Π	a set of paths
$\pi[\xi, \eta]$	geodesic arc, ray or two-way infinite geodesic from ξ to η in a tree
$\widehat{T}$	end compactification of the tree T (9.14)
∂T	space of ends of the tree T
T^∞	set of vertices of the tree T with infinite degree
T^*	set of improper vertices
$\partial^* T$	boundary of the non locally finite tree T, consisting of ends and improper vertices (9.14)

Reversible Markov chains

m	reversing measure (4.1)
∇	difference operator associated with a Markov chain (4.6)
$\mathfrak{L} = P - I$	Laplace operator associated with a Markov chain
$\mathrm{spec}(P)$	spectrum of P
λ	typical notation for an eigenvalue of P

Groups, vectors, functions

$\mathfrak{G}$	typical notation for a group
$\mathfrak{S}$	permutation group (symmetric group)
$\boldsymbol{e}_i$	unit vector in $\mathbb{Z}^d$
$\mathbf{0}$	according to context (a) constant function with value 0, (b) zero column vector in $\mathbb{Z}^d$
$\mathbf{1}$	according to context (a) constant function with value 1, (b) column vector in $\mathbb{Z}^d$ with all coordinates equal to 1
$\mathbf{1}_A$	indicator function of the set or event A
$f(t_0-)$	limit of $f(t)$ as $t \to t_0$ from below (t real)

Galton–Watson process and branching Markov chains

GW	abbreviation for Galton–Watson
BMC	abbreviation for branching Markov chain
BMC(X, P, μ)	branching Markov chain with state space X, transition matrix P and offspring distribution μ
$\Sigma = \{1, \dots, N\}$ or $\Sigma = \mathbb{N}$	set of possible offspring numbers, interpreted as an alphabet
Σ^*	set of all words over Σ, viewed as the N-ary tree (possibly with $N = \infty$)
T	full genealogical tree, a random or deterministic subtree of Σ^* with property (5.32), typically a GW tree
$\boldsymbol{\tau}$	deterministic finite subtree of Σ^* containing the root ϵ

Potential and boundary theory

$\mathcal{H}$	space of harmonic functions
$\mathcal{H}^+$	cone of non-negative harmonic functions
$\mathcal{H}^\infty$	space of bounded harmonic functions
$\mathcal{S}$	set of superharmonic functions
$\mathcal{S}^+$	cone of non-negative superharmonic functions
$\mathcal{E}$	cone of excessive measures
$\widehat{X}$	a compactification of the discrete set X
ξ, η	typical notation for elements of the boundary $\widehat{X} \setminus X$ and for elements of the boundary of a tree
$\widehat{X}(P)$	the Martin compactification of (X, P)
$\mathcal{M} = \widehat{X}(P) \setminus X$	the Martin boundary

Index