Springer Series in Operations Research
and Financial Engineering

Randal Douc · Eric Moulines
Pierre Priouret · Philippe Soulier

Markov Chains

Springer

Springer Series in Operations Research and Financial Engineering

More information about this series at http://www.springer.com/series/3182

Randal Douc · Eric Moulines
Pierre Priouret · Philippe Soulier

Markov Chains

Springer

Randal Douc
Département CITI
Telecom SudParis
Évry, France

Eric Moulines
Centre de Mathématiques Appliquées
Ecole Polytechnique
Palaiseau, France

Pierre Priouret
Université Pierre et Marie Curie
Paris, France

Philippe Soulier
Université Paris Nanterre
Nanterre, France

ISSN 1431-8598 ISSN 2197-1773 (electronic)
Springer Series in Operations Research and Financial Engineering
ISBN 978-3-319-97703-4 ISBN 978-3-319-97704-1 (eBook)
https://doi.org/10.1007/978-3-319-97704-1

Library of Congress Control Number: 2018950197

Mathematics Subject Classification (2010): 60J05, 60-02, 60B10, 60J10, 60J22, 60F05

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Markov chains are a class of stochastic processes very commonly used to model random dynamical systems. Applications of Markov chains can be found in many fields, from statistical physics to financial time series. Examples of successful applications abound. Markov chains are routinely used in signal processing and control theory. Markov chains for storage and queueing models are at the heart of many operational research problems. Markov chain Monte Carlo methods and all their derivatives play an essential role in computational statistics and Bayesian inference.

The modern theory of discrete state-space Markov chains actually started in the 1930s with the work well ahead of its time of Doeblin (1938, 1940), and most of the theory (classification of states, existence of an invariant probability, rates of convergence to equilibrium, etc.) was already known by the end of the 1950s. Of course, there have been many specialized developments of discrete-state-space Markov chains since then, see for example Levin et al. (2009), but these developments are only taught in very specialized courses. Many books cover the classical theory of discrete-state-space Markov chains, from the most theoretical to the most practical. With few exceptions, they deal with almost the same concepts and differ only by the level of mathematical sophistication and the organization of the ideas.

This book deals with the theory of Markov chains on general state spaces. The foundations of general state-space Markov chains were laid in the 1940s, especially under the impulse of the Russian school (Yinnik, Yaglom, et al.). A summary of these early efforts can be found in Doob (1953). During the sixties and the seventies, some very significant results were obtained such as the extension of the notion of irreducibility, recurrence/transience classification, the existence of invariant measures, and limit theorems. The books by Orey (1971) and Foguel (1969) summarize these results.

Neveu (1972) brought many significant additions to the theory by introducing the taboo potential a function instead of a set. This approach is no longer widely used today in applied probability and will not be developed in this book (see, however, Chapter 4). The taboo potential approach was later expanded in the book by Revuz (1975). This latter book contains much more and essentially summarizes all that was known in the mid seventies.

A breakthrough was achieved in the works of Nummelin (1978) and Athreya and Ney (1978), which introduce the notion of the split chain and embedded renewal process. These methods allow one to reduce the study to the case of Markov chains that possess an atom, that is, a set in which a regeneration occurs. The theory of such chains can be developed in complete analogy with discrete state space. The renewal approach leads to many important results, such as geometric ergodicity of recurrent Markov chains (Nummelin and Tweedie 1978; Nummelin and Tuominen 1982, 1983) and limit theorems (central limit theorems, law of iterated logarithms). This program was completed in the book Nummelin (1984), which contains a considerable number of results but is admittedly difficult to read.

This preface would be incomplete if we did not quote Meyn and Tweedie (1993b), referred to as the bible of Markov chains by P. Glynn in his prologue to the second edition of this book (Meyn and Tweedie 2009). Indeed, it must be acknowledged that this book has had a profound impact on the Markov chain community and on the authors. Three of us learned the theory of Markov chains from Meyn and Tweedie (1993b), which has therefore shaped and biased our understanding of this topic.

Meyn and Tweedie (1993b) quickly became a classic in applied probability and is praised by both theoretically inclined researchers and practitioners. This book offers a self-contained introduction to general state-space Markov chains, based on the split chain and embedded renewal techniques. The book recognizes the importance of Foster–Lyapunov drift criteria to assess recurrence or transience of a set and to obtain bounds for the return time or hitting time to a set. It also provides, for positive Markov chains, necessary and sufficient conditions for geometric convergence to stationarity.

The reason we thought it would be useful to write a new book is to survey some of the developments made during the 25 years that have elapsed since the publication of Meyn and Tweedie (1993b). To save space while remaining self-contained, this also implied presenting the classical theory of general state-space Markov chains in a more concise way, eliminating some developments that we thought are more peripheral.

Since the publication of Meyn and Tweedie (1993b), the field of Markov chains has remained very active. New applications have emerged such as Markov chain Monte Carlo (MCMC), which now plays a central role in computational statistics and applied probability. Theoretical development did not lag behind. Triggered by the advent of MCMC algorithms, the topic of quantitative bounds of convergence became a central issue. Much progress has been achieved in this field, using either coupling techniques or operator-theoretic methods. This is one of the main themes of several chapters of this book and still an active field of research. Meyn and Tweedie (1993b) deals only with geometric ergodicity and the associated Foster–Lyapunov drift conditions. Many works have been devoted to subgeometric rates of convergence to stationarity, following the pioneering paper of Tuominen and Tweedie (1994), which appeared shortly after the first version of Meyn and

Tweedie (1993b). These results were later sharpened in a series of works of Jarner and Roberts (2002) and Douc et al. (2004a), where a new drift condition was introduced. There has also been substantial activity on sample paths, limit theorems, and concentration inequalities. For example, Maxwell and Woodroofe (2000) and Rio (2017) obtained conditions for the central limit theorems for additive functions of Markov chains that are close to optimal.

Meyn and Tweedie (1993b) considered exclusively irreducible Markov chains and total variation convergence. There are, of course, many practically important situations in which the irreducibility assumption fails to hold, whereas it is still possible to prove the existence of a unique stationary probability and convergence to stationarity in distances weaker than the total variation. This quickly became an important field of research.

Of course, there are significant omissions in this book, which is already much longer than we initially thought it would be. We do not cover large deviations theory for additive functionals of Markov chains despite the recent advances made in this field in the work of Balaji and Meyn (2000) and Kontoyiannis and Meyn (2005). Similarly, significant progress has been made in the theory of moderate deviations for additive functionals of Markov chains in a series of Chen (1999), Guillin (2001), Djellout and Guillin (2001), and Chen and Guillin (2004). These efforts are not reported in this book. We do not address the theory of fluid limit introduced in Dai (1995) and later refined in Dai and Meyn (1995), Dai and Weiss (1996) and Fort et al. (2006), despite its importance in analyzing the stability of Markov chains and its success in analyzing storage systems (such as networks of queues). There are other significant omissions, and in many chapters we were obliged sometimes to make difficult decisions.

The book is divided into four parts. In Part I, we give the foundations of Markov chain theory. All the results presented in these chapters are very classical. There are two highlights in this part: Kac's construction of the invariant probability in Chapter 3 and the ergodic theorems in Chapter 5 (where we also present a short proof of Birkhoff's theorem).

In Part II, we present the core theory of irreducible Markov chains, which is a subset of Meyn and Tweedie (1993b). We use the regeneration approach to derive most results. Our presentation nevertheless differs from that of Meyn and Tweedie (1993b). We first focus on the theory of atomic chains in Chapter 6. We show that the atoms are either recurrent or transient, establish solidarity properties for atoms, and then discuss the existence of an invariant measure. In Chapter 7, we apply these results to discrete state spaces. We would like to stress that this book can be read without any prior knowledge of discrete-state-space Markov chains: all the results are established as a special case of atomic chains. In Chapter 8, we present the key elements of discrete time-renewal theory. We use the results obtained for discrete-state-space Markov chains to provide a proof of Blackwell and Kendall's theorems, which are central to discrete-time renewal theory. As a first application, we obtain a version of Harris's theorem for atomic Markov chains (based on the first-entrance last-exit decomposition) as well as geometric and polynomial rates of convergence to stationarity.

For Markov chains on general state spaces, the existence of an atom is more the exception than the rule. The splitting method consists in extending the state space to construct a Markov chain that contains the original Markov chain (as its first marginal) and has an atom. Such a construction requires that one have first defined small sets and petite sets, which are introduced in Chapter 9. We have adopted a definition of irreducibility that differs from the more common usage. This avoids the delicate theorem of Jain and Jamison (1967) (which is, however, proved in the appendix of this chapter for completeness but is not used) and allows us to define irreducibility on arbitrary state spaces (whereas the classical assumption requires the use of a countably generated σ-algebra). In Chapter 10, we discuss recurrence, Harris recurrence, and transience of general state-space Markov chains. In Chapter 11, we present the splitting construction and show how the results obtained in the atomic framework can be translated for general state-space Markov chains. The last chapter of this part, Chapter 12, deals with Markov chains on complete separable metric spaces. We introduce the notions of Feller, strong-Feller, and T-chains and show how the notions of small and petite sets can be related in such cases to compact sets. This is a very short presentation of the theory of Feller chains, which are treated in much greater detail in Meyn and Tweedie (1993b) and Borovkov (1998).

The first two parts of the book can be used as a text for a one-semester course, providing the essence of the theory of Markov chains but avoiding difficult technical developments. The mathematical prerequisites are a course in probability, stochastic processes, and measure theory at no deeper level than, for instance, Billingsley (1986) and Taylor (1997). All the measure-theoretic results that we use are recalled in the appendix with precise references. We also occasionally use some results from martingale theory (mainly the martingale convergence theorem), which are also recalled in the appendix. Familiarity with Williams (1991) or the first three chapters of Neveu (1975) is therefore highly recommended. We also occasionally need some topology and functional analysis results for which we mainly refer to the books Royden (1988) and Rudin (1987). Again, the results we use are recalled in the appendix.

Part III presents more advanced results for irreducible Markov chains. In Chapter 13, we complement the results that we obtained in Chapter 8 for atomic Markov chains. In particular, we cover subgeometric rates of convergence. The proofs presented in this chapter are partly original. In Chapter 14 we discuss the geometric regularity of a Markov chain and obtain the equivalence of geometric regularity with a Foster–Lyapunov drift condition. We use these results to establish geometric rates of convergence in Chapter 15. We also establish necessary and sufficient conditions for geometric ergodicity. These results are already reported in Meyn and Tweedie (2009). In Chapter 16, we discuss subgeometric regularity and obtain the equivalence of subgeometric regularity with a family of drift conditions. Most of the arguments are taken from Tuominen and Tweedie (1994). We then discuss the more practical subgeometric drift conditions proposed in Douc et al. (2004a), which are the counterpart of the Foster–Lyapunov conditions for geometric regularity. In Chapter 17 we discuss the subgeometric rate of convergence to stationarity, using the splitting method.

In the last two chapters of this part, we reestablish the rates of convergence by two different types of methods that do not use the splitting technique.

In Chapter 18 we derive explicit geometric rates of convergence by means of operator-theoretic arguments and the fixed-point theorem. We introduce the uniform Doeblin condition and show that it is equivalent to uniform ergodicity, that is, convergence to the invariant distribution at the same geometric rate from every point of the state space. As a by-product, this result provides an alternative proof of the existence of an invariant measure for an irreducible recurrent kernel that does not use the splitting construction. We then prove nonuniform geometric rates of convergence by the operator method, using the ideas introduced in Hairer and Mattingly (2011).

In the last chapter of this part, Chapter 19, we discuss coupling methods that allow us to easily obtain quantitative convergence results as well as short and elegant proofs of several important results. We introduce different notions of coupling, starting almost from scratch: exact coupling, distributional coupling, and maximal coupling. This part owes much to the excellent treatises on coupling methods Lindvall and (1979) and Thorisson (2000), which of course cover much more than this chapter. We then show how exact coupling allows us to obtain explicit rates of convergence in the geometric and subgeometric cases. The use of coupling to obtain geometric rates was introduced in the pioneering work of Rosenthal (1995b) (some improvements were later supplied by Douc et al. (2004b). We also illustrate the use of the exact coupling method to derive subgeometric rates of convergence; we follow here the work of Douc et al. (2006, 2007). Although the content of this part is more advanced, part of it can be used in a graduate course on Markov chains. The presentation of the operator-theoretic approach of Hairer and Mattingly (2011), which is both useful and simple, is of course a must. I also think it interesting to introduce the coupling methods, because they are both useful and elegant.

In Part IV we focus especially on four topics. The choice we made was a difficult one, because there have been many new developments in Markov chain theory over the last two decades. There is, therefore, a great deal of arbitrariness in these choices and important omissions. In Chapter 20, we assume that the state space is a complete separable metric space, but we no longer assume that the Markov chain is irreducible. Since it is no longer possible to construct an embedded regenerative process, the techniques of proof are completely different; the essential difference is that convergence in total variation distance may no longer hold, and it must be replaced by Wasserstein distances. We recall the main properties of these distances and in particular the duality theorem, which allows us to use coupling methods. We have essentially followed Hairer et al. (2011) in the geometric case and Butkovsky (2014) and Durmus et al. (2016) for the subgeometric case. However, the methods of proof and some of the results appear to be original. Chapter 21 covers central limit theorems of additive functions of Markov chains. The most direct approach is to use a martingale decomposition (with a remainder term) of the additive functionals by introducing solutions of the Poisson equation. The approach is straightforward, and Poisson solutions exist under minimal technical assumptions (see Glynn and Meyn 1996), yet this method does not yield conditions close to optimal. A first approach to weaken these technical conditions was introduced in Kipnis and Varadhan (1985) and further developed by Maxwell and Woodroofe (2000): it keeps the martingale decomposition with remainder but replaces Poisson by resolvent solutions and uses tightness arguments. It yields conditions that are closer to being sufficient. A second approach, due to Gordin and Lifšic (1978) and later refined by many authors (see Rio 2017), uses another martingale decomposition and yields closely related (but nevertheless different) sets of conditions. We also discuss different expressions for the asymptotic variance, following Häggström and Rosenthal (2007). In Chapter 22, we discuss the spectral property of a Markov kernel P seen as an operator on an appropriately defined Banach space of complex functions and complex measures. We study the convergence to the stationary distribution using the particular structure of the spectrum of this operator; deep results can be obtained when the Markov kernel P is reversible (i.e., self-adjoint), as shown, for example, in Roberts and Tweedie (2001) and Kontoyiannis and Meyn (2012). We also introduce the notion of conductance and prove geometric convergence using conductance thorough Cheeger's inequalities, following Lawler and Sokal (1988) and Jarner and Yuen (2004). Finally, in Chapter 23 we give an introduction to sub-Gaussian concentration inequalities for Markov chains. We first show how McDiarmid's inequality can be extended to uniformly ergodic Markov kernels following Rio (2000a). We then discuss the equivalence between McDiarmid-typesub-Gaussian concentration inequalities and geometric ergodicity, using a result established in Dedecker and Gouëzel (2015). We finally obtain extensions of these inequalities for separately Lipschitz functions, following Djellout et al. (2004) and Joulin and Ollivier (2010).

We have chosen to illustrate the main results with simple examples. More substantial examples are considered in the exercises at the end of each chapter; the solutions of a majority of these exercises are provided. The reader is invited use these exercises (which are mostly fairly direct applications of the material) to test their understanding of the theory. We have selected examples from different fields, including signal processing and automatic control, time-series analysis and Markov chain Monte Carlo simulation algorithms.

We do not cite bibliographical references in the body of the chapters, but we have added at the end of each chapter bibliographical indications. We give precise bibliographical indications for the most recent developments. For former results, we do not necessarily seek to attribute authorship to the original results. Meyn and Tweedie (1993b) covers in much greater detail the genesis of the earlier works.

The authors would like to thank the large number of people who at times contributed to this book. Alain Durmus, Gersende Fort, and François Roueff gave us valuable advice and helped us to clarify some of the derivations. Their contributions were essential. Christophe Andrieu, Gareth Roberts, Jeffrey Rosenthal and Alexander Veretennikov also deserve special thanks. They have been a very valuable source of inspiration for years.

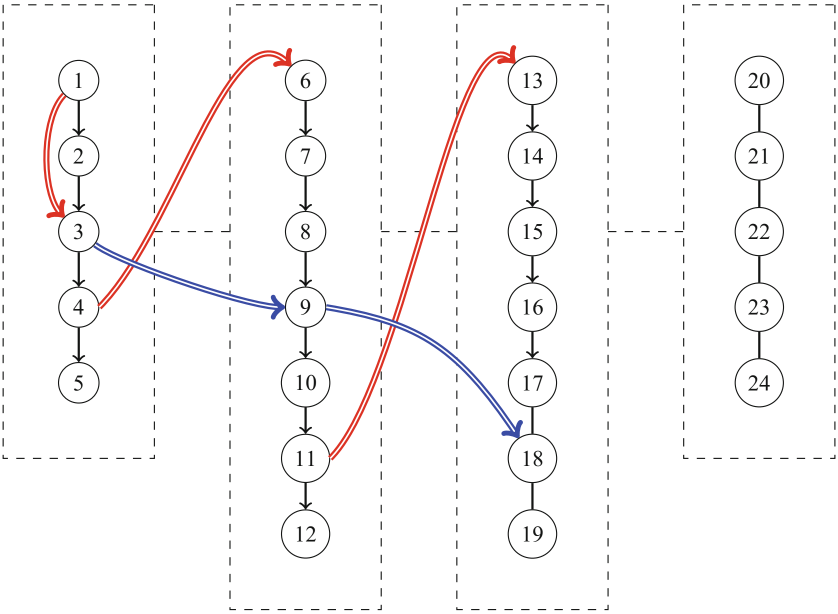

Fig. 1 Suggestion of playback order with respect to the different chapters of the book. The red arrows correspond to a possible path for a reader eager to focus only on the most fundamental results. The skipped chapters can then be investigated on a second reading. The blue arrows provide a fast track for a proof of the existence of an invariant measure and geometric rates of convergence for irreducible chains without the splitting technique. The chapters in the last part of the book are almost independent and can be read in any order.

We also benefited from the work of many colleagues who carefully reviewed parts of this book and helped us to correct errors and suggested improvements in the presentation: Yves Atchadé, David Barrera, Nicolas Brosse, Arnaud Doucet, Sylvain Le Corff, Matthieu Lerasle, Jimmy Olsson, Christian Robert, Claude Saint-Cricq, and Amandine Schreck.

We are also very grateful to all the students who for years helped us to polish what was at the beginning a set of rough lecture notes. Their questions and suggestions greatly helped us to improve the presentation and correct errors.

Évry, France Randal Douc
Palaiseau, France Eric Moulines
Paris, France Pierre Priouret
Nanterre, France Philippe Soulier

Contents

Part I Foundations

1 Markov Chains: Basic Definitions 3
1.1 Markov Chains 3
1.2 Kernels 6
1.3 Homogeneous Markov Chains 12
1.4 Invariant Measures and Stationarity 16
1.5 Reversibility 18
1.6 Markov Kernels on $\mathrm{L}^p(\pi)$ 20
1.7 Exercises 21
1.8 Bibliographical Notes 25

2 Examples of Markov Chains 27
2.1 Random Iterative Functions 27
2.2 Observation-Driven Models 35
2.3 Markov Chain Monte Carlo Algorithms 38
2.4 Exercises 49
2.5 Bibliographical Notes 51

3 Stopping Times and the Strong Markov Property 53
3.1 The Canonical Chain 54
3.2 Stopping Times 58
3.3 The Strong Markov Property 60
3.4 First-Entrance, Last-Exit Decomposition 64
3.5 Accessible and Attractive Sets 66
3.6 Return Times and Invariant Measures 67
3.7 Exercises 73
3.8 Bibliographical Notes 74

4 Martingales, Harmonic Functions and Poisson–Dirichlet Problems . . . 75
4.1 Harmonic and Superharmonic Functions . . . 75
4.2 The Potential Kernel . . . 77
4.3 The Comparison Theorem . . . 81
4.4 The Dirichlet and Poisson Problems . . . 85
4.5 Time-Inhomogeneous Poisson–Dirichlet Problems . . . 88
4.6 Exercises . . . 89
4.7 Bibliographical Notes . . . 95

5 Ergodic Theory for Markov Chains . . . 97
5.1 Dynamical Systems . . . 97
5.2 Markov Chain Ergodicity . . . 104
5.3 Exercises . . . 111
5.4 Bibliographical Notes . . . 115

Part II Irreducible Chains: Basics

6 Atomic Chains . . . 119
6.1 Atoms . . . 119
6.2 Recurrence and Transience . . . 121
6.3 Period of an Atom . . . 126
6.4 Subinvariant and Invariant Measures . . . 128
6.5 Independence of the Excursions . . . 134
6.6 Ratio Limit Theorems . . . 135
6.7 The Central Limit Theorem . . . 137
6.8 Exercises . . . 140
6.9 Bibliographical Notes . . . 144

7 Markov Chains on a Discrete State Space . . . 145
7.1 Irreducibility, Recurrence, and Transience . . . 145
7.2 Invariant Measures, Positive and Null Recurrence . . . 146
7.3 Communication . . . 148
7.4 Period . . . 150
7.5 Drift Conditions for Recurrence and Transience . . . 151
7.6 Convergence to the Invariant Probability . . . 154
7.7 Exercises . . . 159
7.8 Bibliographical Notes . . . 164

8 Convergence of Atomic Markov Chains . . . 165
8.1 Discrete-Time Renewal Theory . . . 165
8.2 Renewal Theory and Atomic Markov Chains . . . 175
8.3 Coupling Inequalities for Atomic Markov Chains . . . 180
8.4 Exercises . . . 187
8.5 Bibliographical Notes . . . 189

9 Small Sets, Irreducibility, and Aperiodicity 191
9.1 Small Sets 191
9.2 Irreducibility 194
9.3 Periodicity and Aperiodicity 201
9.4 Petite Sets 206
9.5 Exercises 211
9.6 Bibliographical Notes 215
9.A Proof of Theorem 9.2.6 215

10 Transience, Recurrence, and Harris Recurrence 221
10.1 Recurrence and Transience 221
10.2 Harris Recurrence 228
10.3 Exercises 236
10.4 Bibliographical Notes 239

11 Splitting Construction and Invariant Measures 241
11.1 The Splitting Construction 241
11.2 Existence of Invariant Measures 247
11.3 Convergence in Total Variation to the Stationary Distribution 251
11.4 Geometric Convergence in Total Variation Distance 253
11.5 Exercises 258
11.6 Bibliographical Notes 259
11.A Another Proof of the Convergence of Harris Recurrent Kernels 259

12 Feller and *T*-Kernels 265
12.1 Feller Kernels 265
12.2 *T*-Kernels 270
12.3 Existence of an Invariant Probability 274
12.4 Topological Recurrence 277
12.5 Exercises 279
12.6 Bibliographical Notes 285
12.A Linear Control Systems 285

Part III Irreducible Chains: Advanced Topics

13 Rates of Convergence for Atomic Markov Chains 289
13.1 Subgeometric Sequences 289
13.2 Coupling Inequalities for Atomic Markov Chains 291
13.3 Rates of Convergence in Total Variation Distance 303
13.4 Rates of Convergence in f-Norm 305
13.5 Exercises 311
13.6 Bibliographical Notes 312

14 Geometric Recurrence and Regularity . . . 313
14.1 f-Geometric Recurrence and Drift Conditions . . . 313
14.2 f-Geometric Regularity . . . 321
14.3 f-Geometric Regularity of the Skeletons . . . 327
14.4 f-Geometric Regularity of the Split Kernel . . . 332
14.5 Exercises . . . 334
14.6 Bibliographical Notes . . . 337

15 Geometric Rates of Convergence . . . 339
15.1 Geometric Ergodicity . . . 339
15.2 V-Uniform Geometric Ergodicity . . . 349
15.3 Uniform Ergodicity . . . 353
15.4 Exercises . . . 356
15.5 Bibliographical Notes . . . 358

16 (f, r)-Recurrence and Regularity . . . 361
16.1 (f, r)-Recurrence and Drift Conditions . . . 361
16.2 (f, r)-Regularity . . . 370
16.3 (f, r)-Regularity of the Skeletons . . . 377
16.4 (f, r)-Regularity of the Split Kernel . . . 381
16.5 Exercises . . . 382
16.6 Bibliographical Notes . . . 383

17 Subgeometric Rates of Convergence . . . 385
17.1 (f, r)-Ergodicity . . . 385
17.2 Drift Conditions . . . 392
17.3 Bibliographical Notes . . . 399
17.A Young Functions . . . 399

18 Uniform and V-Geometric Ergodicity by Operator Methods . . . 401
18.1 The Fixed-Point Theorem . . . 401
18.2 Dobrushin Coefficient and Uniform Ergodicity . . . 403
18.3 V-Dobrushin Coefficient . . . 409
18.4 V-Uniformly Geometrically Ergodic Markov Kernel . . . 412
18.5 Application of Uniform Ergodicity to the Existence of an Invariant Measure . . . 415
18.6 Exercises . . . 417
18.7 Bibliographical Notes . . . 419

19 Coupling for Irreducible Kernels . . . 421
19.1 Coupling . . . 422
19.2 The Coupling Inequality . . . 432
19.3 Distributional, Exact, and Maximal Coupling . . . 435
19.4 A Coupling Proof of V-Geometric Ergodicity . . . 441
19.5 A Coupling Proof of Subgeometric Ergodicity . . . 444

19.6 Exercises 449
19.7 Bibliographical Notes 451

Part IV Selected Topics

20 Convergence in the Wasserstein Distance 455
20.1 The Wasserstein Distance 456
20.2 Existence and Uniqueness of the Invariant Probability Measure 462
20.3 Uniform Convergence in the Wasserstein Distance 465
20.4 Nonuniform Geometric Convergence 471
20.5 Subgeometric Rates of Convergence for the Wasserstein Distance 476
20.6 Exercises 480
20.7 Bibliographical Notes 485
20.A Complements on the Wasserstein Distance 486

21 Central Limit Theorems 489
21.1 Preliminaries 490
21.2 The Poisson Equation 495
21.3 The Resolvent Equation 503
21.4 A Martingale Coboundary Decomposition 508
21.5 Exercises 517
21.6 Bibliographical Notes 519
21.A A Covariance Inequality 520

22 Spectral Theory 523
22.1 Spectrum 523
22.2 Geometric and Exponential Convergence in $L^2(\pi)$ 530
22.3 $L^p(\pi)$-Exponential Convergence 538
22.4 Cheeger's Inequality 545
22.5 Variance Bounds for Additive Functionals and the Central Limit Theorem for Reversible Markov Chains 553
22.6 Exercises 560
22.7 Bibliographical Notes 562
22.A Operators on Banach and Hilbert Spaces 563
22.B Spectral Measure 572

23 Concentration Inequalities 575
23.1 Concentration Inequality for Independent Random Variables 576
23.2 Concentration Inequality for Uniformly Ergodic Markov Chains 581
23.3 Sub-Gaussian Concentration Inequalities for V-Geometrically Ergodic Markov Chains 587

23.4 Exponential Concentration Inequalities Under Wasserstein Contraction 594
23.5 Exercises 599
23.6 Bibliographical Notes 601

Appendices 603

A Notations 605

B Topology, Measure and Probability 609
B.1 Topology 609
B.2 Measures 612
B.3 Probability 618

C Weak Convergence 625
C.1 Convergence on Locally Compact Metric Spaces 625
C.2 Tightness 626

D Total and V-Total Variation Distances 629
D.1 Signed Measures 629
D.2 Total Variation Distance 631
D.3 V-Total Variation 635

E Martingales 637
E.1 Generalized Positive Supermartingales 637
E.2 Martingales 638
E.3 Martingale Convergence Theorems 639
E.4 Central Limit Theorems 641

F Mixing Coefficients 645
F.1 Definitions 645
F.2 Properties 646
F.3 Mixing Coefficients of Markov Chains 653

G Solutions to Selected Exercises 657

References 733

Index 753

Part I
Foundations

Chapter 1
Markov Chains: Basic Definitions

Heuristically, a discrete-time stochastic process has the Markov property if the past and future are independent given the present. In this introductory chapter, we give the formal definition of a Markov chain and of the main objects related to this type of stochastic process and establish basic results. In particular, we will introduce in Section 1.2 the essential notion of a Markov kernel, which gives the distribution of the next state given the current state. In Section 1.3, we will restrict attention to time-homogeneous Markov chains and establish that a fundamental consequence of the Markov property is that the entire distribution of a Markov chain is characterized by the distribution of its initial state and a Markov kernel. In Section 1.4, we will introduce the notion of invariant measures, which play a key role in the study of the long-term behavior of a Markov chain. Finally, in Sections 1.5 and 1.6, which can be skipped on a first reading, we will introduce the notion of reversibility, which is very convenient and is satisfied by many Markov chains, and some further properties of kernels seen as operators and certain spaces of functions.

1.1 Markov Chains

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $(\mathsf{X}, \mathscr{X})$ a measurable space, and T a set. A family of X-valued random variables indexed by T is called an X-valued stochastic process indexed by T.

Throughout this chapter, we consider only the cases $T = \mathbb{N}$ and $T = \mathbb{Z}$.

A filtration of a measurable space $(\Omega, \mathscr{F})$ is an increasing sequence $\{\mathscr{F}_k,\ k \in T\}$ of sub-σ-fields of $\mathscr{F}$. A filtered probability space $(\Omega,\ \mathscr{F}, \{\mathscr{F}_k,\ k \in T\}, \mathbb{P})$ is a probability space endowed with a filtration.

A stochastic process $\{X_k,\ k \in T\}$ is said to be adapted to the filtration $\{\mathscr{F}_k,\ k \in T\}$ if for each $k \in T$, X_k is $\mathscr{F}_k$-measurable. The notation $\{(X_k, \mathscr{F}_k),\ k \in T\}$ will be used to indicate that the process $\{X_k,\ k \in T\}$ is adapted to the filtration $\{\mathscr{F}_k,\ k \in T\}$. The σ-field $\mathscr{F}_k$ can be thought of as the information available at time k. Requiring

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_1

the process to be adapted means that the probability of events related to X_k can be computed using solely the information available at time k.

The natural filtration of a stochastic process $\{X_k,\ k \in T\}$ defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is the filtration $\{\mathscr{F}_k^X,\ k \in T\}$ defined by

$$\mathscr{F}_k^X = \sigma(X_j\ , j \leq k\ , j \in T)\ , \quad k \in T\ .$$

By definition, a stochastic process is adapted to its natural filtration. The main definition of this chapter can now be stated.

Definition 1.1.1 (Markov chain) *Let $(\Omega, \mathscr{F}, \{\mathscr{F}_k,\ k \in T\}, \mathbb{P})$ be a filtered probability space. An adapted stochastic process $\{(X_k, \mathscr{F}_k),\ k \in T\}$ is a Markov chain if for all $k \in T$ and $A \in \mathscr{X}$,*

$$\mathbb{P}(X_{k+1} \in A \,|\, \mathscr{F}_k) = \mathbb{P}(X_{k+1} \in A \,|\, X_k) \qquad \mathbb{P} - \text{a.s.} \tag{1.1.1}$$

Condition (1.1.1) is equivalent to the following condition: for all $f \in \mathbb{F}_+(\mathsf{X}) \cup \mathbb{F}_b(\mathsf{X})$,

$$\mathbb{E}\left[f(X_{k+1}) \,|\, \mathscr{F}_k\right] = \mathbb{E}\left[f(X_{k+1}) | X_k\right] \qquad \mathbb{P} - \text{a.s.} \tag{1.1.2}$$

Let $\{\mathscr{G}_k,\ k \in T\}$ denote another filtration such that for all $k \in T$, $\mathscr{G}_k \subset \mathscr{F}_k$. If $\{(X_k, \mathscr{F}_k),\ k \in T\}$ is a Markov chain and $\{X_k,\ k \in T\}$ is adapted to the filtration $\{\mathscr{G}_k,\ k \in T\}$, then $\{(X_k, \mathscr{G}_k),\ k \in T\}$ is also a Markov chain. In particular, a Markov chain is always a Markov chain with respect to its natural filtration.

We now give other characterizations of a Markov chain.

Theorem 1.1.2. *Let $(\Omega, \mathscr{F}, \{\mathscr{F}_k,\ k \in T\}, \mathbb{P})$ be a filtered probability space and $\{(X_k, \mathscr{F}_k),\ k \in T\}$ an adapted stochastic process. The following properties are equivalent.*

(i) $\{(X_k, \mathscr{F}_k),\ k \in T\}$ is a Markov chain.

(ii) For every $k \in T$ and bounded $\sigma(X_j, j \geq k)$-measurable random variable Y,

$$\mathbb{E}\left[Y \,|\, \mathscr{F}_k\right] = \mathbb{E}\left[Y \,|\, X_k\right] \qquad \mathbb{P} - \text{a.s.} \tag{1.1.3}$$

(iii) For every $k \in T$, bounded $\sigma(X_j, j \geq k)$-measurable random variable Y, and bounded $\mathscr{F}_k^X$-measurable random variable Z,

$$\mathbb{E}\left[YZ | X_k\right] = \mathbb{E}\left[Y | X_k\right] \mathbb{E}\left[Z | X_k\right] \qquad \mathbb{P} - \text{a.s.} \tag{1.1.4}$$

Proof. (i) $\Rightarrow$ (ii) Fix $k \in T$ and consider the following property (where $\mathbb{F}_b(\mathsf{X})$ is the set of bounded measurable functions):

($\mathscr{P}_n$): (1.1.3) holds for all $Y = \prod_{j=0}^n g_j(X_{k+j})$, where $g_j \in \mathbb{F}_b(\mathsf{X})$ for all $j \geq 0$.

($\mathscr{P}_0$) is true. Assume that ($\mathscr{P}_n$) holds and let $\{g_j,\ j \in \mathbb{N}\}$ be a sequence of functions in $\mathbb{F}_b(\mathsf{X})$. The Markov property (1.1.2) yields

$$\begin{aligned}
&\mathbb{E}\left[g_0(X_k)\dots g_n(X_{k+n})g_{n+1}(X_{k+n+1})\middle|\,\mathscr{F}_k\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[g_0(X_k)\dots g_n(X_{k+n})g_{n+1}(X_{k+n+1})\middle|\,\mathscr{F}_{k+n}\right]\middle|\,\mathscr{F}_k\right] \\
&= \mathbb{E}\left[g_0(X_k)\dots g_n(X_{k+n})\mathbb{E}\left[g_{n+1}(X_{k+n+1})\middle|\,\mathscr{F}_{k+n}\right]\middle|\,\mathscr{F}_k\right] \\
&= \mathbb{E}\left[g_0(X_k)\dots g_n(X_{k+n})\mathbb{E}\left[g_{n+1}(X_{k+n+1})\middle|\,X_{k+n}\right]\middle|\,\mathscr{F}_k\right] .
\end{aligned}$$

The last term in the product being a measurable function of X_{n+k}, the induction assumption ($\mathscr{P}_n$) yields

$$\begin{aligned}
&\mathbb{E}\left[g_0(X_k)\dots g_n(X_{k+n})g_{n+1}(X_{k+n+1})\middle|\,\mathscr{F}_k\right] \\
&= \mathbb{E}\left[g_0(X_k)\dots g_n(X_{k+n})\mathbb{E}\left[g_{n+1}(X_{k+n+1})\middle|\,X_{k+n}\right]\middle|\,X_k\right] \\
&= \mathbb{E}\left[g_0(X_k)\dots g_n(X_{k+n})\mathbb{E}\left[g_{n+1}(X_{k+n+1})\middle|\,\mathscr{F}_{k+n}\right]\middle|\,X_k\right] \\
&= \mathbb{E}\left[g_0(X_k)\dots g_n(X_{k+n})g_{n+1}(X_{k+n+1})\middle|\,X_k\right] ,
\end{aligned}$$

which proves ($\mathscr{P}_{n+1}$). Therefore, ($\mathscr{P}_n$) is true for all $n \in \mathbb{N}$.
Consider the set

$$\mathscr{H} = \left\{Y \in \sigma(X_j, j \geq k) \,:\, \mathbb{E}\left[Y\middle|\,\mathscr{F}_k\right] = \mathbb{E}\left[Y\middle|\,X_k\right] \ \ \mathbb{P}-\text{a.s.}\right\} .$$

It is easily seen that $\mathscr{H}$ is a vector space. In addition, if $\{Y_n,\ n \in \mathbb{N}\}$ is an increasing sequence of nonnegative random variables in $\mathscr{H}$ and if $Y = \lim_{n\to\infty} Y_n$ is bounded, then by the monotone convergence theorem for conditional expectations,

$$\mathbb{E}\left[Y\middle|\,\mathscr{F}_k\right] = \lim_{n\to\infty} \mathbb{E}\left[Y_n\middle|\,\mathscr{F}_k\right] = \lim_{n\to\infty} \mathbb{E}\left[Y_n\middle|\,X_k\right] = \mathbb{E}\left[Y\middle|\,X_k\right] \qquad \mathbb{P}-\text{a.s.}$$

By Theorem B.2.4, the space $\mathscr{H}$ contains all $\sigma(X_j, j \geq k)$-measurable random variables.

(ii) $\Rightarrow$ (iii) If Y is a bounded $\sigma(X_j, j \geq k)$-measurable random variable and Z is a bounded $\mathscr{F}_k$-measurable random variable, an application of (ii) yields

$$\mathbb{E}\left[YZ\middle|\,\mathscr{F}_k\right] = Z\mathbb{E}\left[Y\middle|\,\mathscr{F}_k\right] = Z\mathbb{E}\left[Y\middle|\,X_k\right] \qquad \mathbb{P}-\text{a.s.}$$

Thus,

$$\begin{aligned}
\mathbb{E}\left[YZ\middle|\,X_k\right] &= \mathbb{E}\left[\mathbb{E}\left[YZ\middle|\,\mathscr{F}_k\right]\middle|\,X_k\right] = \mathbb{E}\left[Z\mathbb{E}\left[Y\middle|\,X_k\right]\middle|\,X_k\right] \\
&= \mathbb{E}\left[Z\middle|\,X_k\right]\mathbb{E}\left[Y\middle|\,X_k\right] \qquad \mathbb{P}-\text{a.s.}
\end{aligned}$$

(iii) $\Rightarrow$ (i) If Z is bounded and $\mathscr{F}_k$-measurable, we obtain

$$\begin{aligned}
\mathbb{E}\left[f(X_{k+1})Z\right] &= \mathbb{E}\left[\mathbb{E}\left[f(X_{k+1})Z\middle|\,X_k\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[f(X_{k+1})\middle|\,X_k\right]\mathbb{E}\left[Z\middle|\,X_k\right]\right] = \mathbb{E}\left[\mathbb{E}\left[f(X_{k+1})\middle|\,X_k\right]Z\right] .
\end{aligned}$$

This proves (i).

□

Heuristically, Condition (1.1.4) means that the future of a Markov chain is conditionally independent of its past, given its present state.

An important caveat must be made; the Markov property is not hereditary. If $\{(X_k,\mathscr{F}_k),\ k\in T\}$ is a Markov chain on X and f is a measurable function from $(\mathsf{X},\mathscr{X})$ to $(\mathsf{Y},\mathscr{Y})$, then, unless f is one-to-one, $\{(f(X_k),\mathscr{F}_k),k\in T\}$ need not be a Markov chain. In particular, if $\mathsf{X}=\mathsf{X}_1\times\mathsf{X}_2$ is a product space and $\{(X_k,\mathscr{F}_k),\ k\in T\}$ is a Markov chain with $X_k=(X_{1,k},X_{2,k})$ then the sequence $\{(X_{1,k},\mathscr{F}_k),k\in T\}$ may fail to be a Markov chain.

1.2 Kernels

We now introduce Markov kernels, which will be the core of the theory.

Definition 1.2.1 *Let* $(\mathsf{X},\mathscr{X})$ *and* $(\mathsf{Y},\mathscr{Y})$ *be two measurable spaces. A kernel N on* $\mathsf{X}\times\mathscr{Y}$ *is a mapping* $N:\mathsf{X}\times\mathscr{Y}\to[0,\infty]$ *satisfying the following conditions:*

(i) for every $x\in\mathsf{X}$, the mapping $N(x,\cdot):A\mapsto N(x,\ A)$ is a measure on $\mathscr{Y}$;
(ii) for every $A\in\mathscr{Y}$, the mapping $N(\cdot,A):x\mapsto N(x,A)$ is a measurable function from $(\mathsf{X},\mathscr{X})$ to $([0,\infty],\mathscr{B}([0,\infty])$.

- *N is said to be bounded if* $\sup_{x\in\mathsf{X}}N(x,\mathsf{Y})<\infty$.
- *N is called a Markov kernel if* $N(x,\mathsf{Y})=1$, *for all* $x\in\mathsf{X}$.
- *N is said to be sub-Markovian if* $N(x,\mathsf{Y})\le 1$, *for all* $x\in\mathsf{X}$.

Example 1.2.2 (Discrete state space kernel). Assume that X and Y are countable sets. Each element $x\in\mathsf{X}$ is then called a state. A kernel N on $\mathsf{X}\times\mathscr{P}(\mathsf{Y})$, where $\mathscr{P}(\mathsf{Y})$ is the set of all subsets of Y, is a (possibly doubly infinite) matrix $N=(N(x,y):x,y\in\mathsf{X}\times\mathsf{Y})$ with nonnegative entries. Each row $\{N(x,y)\ :\ y\in\mathsf{Y}\}$ is a measure on $(\mathsf{Y},\mathscr{P}(\mathsf{Y}))$ defined by

$$N(x,A)=\sum_{y\in A}N(x,y)\ ,$$

for $A\subset\mathsf{Y}$. The matrix N is said to be *Markovian* if every row $\{N(x,y)\ :\ y\in\mathsf{Y}\}$ is a probability on $(\mathsf{Y},\mathscr{P}(\mathsf{Y}))$, i.e., $\sum_{y\in\mathsf{Y}}N(x,y)=1$ for all $x\in\mathsf{X}$. The associated kernel is defined by $N(x,\{y\})=N(x,y)$ for all $x,y\in\mathsf{X}$. ◀

Example 1.2.3 (Measure seen as a kernel). A σ-finite measure ν on a space $(\mathsf{Y},\mathscr{Y})$ can be seen as a kernel on $\mathsf{X}\times\mathscr{Y}$ by defining $N(x,A)=\nu(A)$ for all $x\in\mathsf{X}$ and $A\in\mathscr{Y}$. It is a Markov kernel if ν is a probability measure. ◀

Example 1.2.4 (Kernel density). Let λ be a positive σ-finite measure on $(\mathsf{Y},\mathscr{Y})$ and $n : \mathsf{X}\times\mathsf{Y} \to \mathbb{R}_+$ a nonnegative function, measurable with respect to the product σ-field $\mathscr{X}\otimes\mathscr{Y}$. Then the mapping N defined on $\mathsf{X}\times\mathscr{Y}$ by

$$N(x,A) = \int_A n(x,y)\lambda(\mathrm{d}y) ,$$

is a kernel. The function n is called the density of the kernel N with respect to the measure λ. The kernel N is Markovian if and only if $\int_{\mathsf{Y}} n(x,y)\lambda(\mathrm{d}y) = 1$ for all $x \in \mathsf{X}$. ◀

Let N be a kernel on $\mathsf{X}\times\mathscr{X}$ and $f \in \mathbb{F}_+(\mathsf{Y})$. A function $Nf : \mathsf{X} \to \mathbb{R}_+$ is defined by setting, for $x \in \mathsf{X}$,

$$Nf(x) = N(x,\mathrm{d}y)f(y) .$$

For all functions f of $\mathbb{F}(\mathsf{Y})$ (where $\mathbb{F}(\mathsf{Y})$ stands for the set of measurable functions on $(\mathsf{Y},\mathscr{Y})$) such that Nf^+ and Nf^- are not both infinite, we define $Nf = Nf^+ - Nf^-$. We will also use the notation $N(x,f)$ for $Nf(x)$ and, for $A \in \mathscr{X}$, $N(x,\mathbb{1}_A)$ or $N\mathbb{1}_A(x)$ for $N(x,A)$.

Proposition 1.2.5 *Let N be a kernel on $\mathsf{X}\times\mathscr{Y}$. For all $f \in \mathbb{F}_+(\mathsf{Y})$, $Nf \in \mathbb{F}_+(\mathsf{X})$. Moreover, if N is a Markov kernel, then $|Nf|_\infty \le |f|_\infty$.*

Proof. Assume first that f is a simple nonnegative function, i.e., $f = \sum_{i\in I}\beta_i \mathbb{1}_{B_i}$ for a finite collection of nonnegative numbers β_i and sets $B_i \in \mathscr{Y}$. Then for $x \in \mathsf{X}$, $Nf(x) = \sum_{i\in I}\beta_i N(x,B_i)$, and by property (ii) of Definition 1.2.1, the function Nf is measurable. Recall that every function $f \in \mathbb{F}_+(\mathsf{X})$ is a pointwise limit of an increasing sequence of measurable nonnegative simple functions $\{f_n,\ n\in\mathbb{N}\}$, i.e., $\lim_{n\to\infty}\uparrow f_n(y) = f(y)$ for all $y \in \mathsf{Y}$. Then by the monotone convergence theorem, for all $x \in \mathsf{X}$,

$$Nf(x) = \lim_{n\to\infty} Nf_n(x) .$$

Therefore, Nf is the pointwise limit of a sequence of nonnegative measurable functions, hence is measurable. If, moreover, N is a Markov kernel on $\mathsf{X}\times\mathscr{Y}$ and $f \in \mathbb{F}_b(\mathsf{Y})$, then for all $x \in \mathsf{X}$,

$$Nf(x) = \int_{\mathsf{Y}} f(y)N(x,\mathrm{d}y) \le |f|_\infty \int_{\mathsf{Y}} N(x,\mathrm{d}y) = |f|_\infty N(x,\mathsf{Y}) = |f|_\infty .$$

This proves the last claim. □

With a slight abuse of notation, we will use the same symbol N for the kernel and the associated operator $N : \mathbb{F}_+(\mathsf{Y}) \to \mathbb{F}_+(\mathsf{X})$, $f \mapsto Nf$. This operator is additive and positively homogeneous: for all $f,g \in \mathbb{F}_+(\mathsf{Y})$ and $\alpha \in \mathbb{R}_+$, one has $N(f+g) = Nf + Ng$ and $N(\alpha f) = \alpha Nf$. The monotone convergence theorem shows that if

$\{f_n,\ n \in \mathbb{N}\} \subset \mathbb{F}_+(\mathsf{Y})$ is an increasing sequence of functions, then $\lim_{n\to\infty} \uparrow N f_n = N(\lim_{n\to\infty} \uparrow f_n)$. The following result establishes a converse.

Proposition 1.2.6 *Let $M : \mathbb{F}_+(\mathsf{Y}) \to \mathbb{F}_+(\mathsf{X})$ be an additive and positively homogeneous operator such that $\lim_{n\to\infty} M(f_n) = M(\lim_{n\to\infty} f_n)$ for every increasing sequence $\{f_n,\ n \in \mathbb{N}\}$ of functions in $\mathbb{F}_+(\mathsf{Y})$. Then*

(i) the function N defined on $\mathsf{X} \times \mathscr{Y}$ by $N(x,A) = M(\mathbb{1}_A)(x)$, $x \in \mathsf{X}$, $A \in \mathscr{Y}$, is a kernel;
(ii) $M(f) = Nf$ for all $f \in \mathbb{F}_+(\mathsf{Y})$.

Proof. (i) Since M is additive, for each $x \in \mathsf{X}$, the function $A \mapsto N(x,A)$ is additive. Indeed, for $n \in \mathbb{N}^*$ and mutually disjoint sets $A_1, \ldots, A_n \in \mathscr{Y}$, we obtain

$$N\left(x, \bigcup_{i=1}^n A_i\right) = M\left(\sum_{i=1}^n \mathbb{1}_{A_i}\right)(x) = \sum_{i=1}^n M(\mathbb{1}_{A_i})(x) = \sum_{i=1}^n N(x,A_i)\ .$$

Let $\{A_i,\ i \in \mathbb{N}\} \subset \mathscr{Y}$ be a sequence of mutually disjoints sets. Then, by additivity and the monotone convergence property of M, we get, for all $x \in \mathsf{X}$,

$$N\left(x, \bigcup_{i=1}^\infty A_i\right) = M\left(\sum_{i=1}^\infty \mathbb{1}_{A_i}\right)(x) = \sum_{i=1}^\infty M(\mathbb{1}_{A_i})(x) = \sum_{i=1}^\infty N(x,A_i)\ .$$

This proves that for all $x \in \mathsf{X}$, $A \mapsto N(x,A)$ is a measure on $(\mathsf{Y}, \mathscr{Y})$. For all $A \in \mathscr{X}$, $x \mapsto N(x,A) = M(\mathbb{1}_A)(x)$ belongs to $\mathbb{F}_+(\mathsf{X})$. Then N is a kernel on $\mathsf{X} \times \mathscr{Y}$.

(ii) If $f = \sum_{i\in I} \beta_i \mathbb{1}_{B_i}$ for a finite collection of nonnegative numbers β_i and sets $B_i \in \mathscr{Y}$, then the additivity and positive homogeneity of M shows that

$$M(f) = \sum_{i\in I} \beta_i M(\mathbb{1}_{B_i}) = \sum_{i\in I} \beta_i N \mathbb{1}_{B_i} = Nf\ .$$

Let now $f \in \mathbb{F}_+(\mathsf{Y})$ (where $\mathbb{F}_+(\mathsf{Y})$ is the set of measurable nonnegative functions) and let $\{f_n,\ n \in \mathbb{N}\}$ be an increasing sequence of nonnegative simple functions such that $\lim_{n\to\infty} f_n(y) = f(y)$ for all $y \in \mathsf{Y}$. Since $M(f) = \lim_{n\to\infty} M(f_n)$ and by the monotone convergence theorem $Nf = \lim_{n\to\infty} N f_n$, we obtain $M(f) = Nf$.

□

Kernels also act on measures. Let $\mu \in \mathbb{M}_+(\mathscr{X})$, where $\mathbb{M}_+(\mathscr{X})$ is the set of (nonnegative) measures on $(\mathsf{X}, \mathscr{X})$. For $A \in \mathscr{Y}$, define

$$\mu N(A) = \int_{\mathsf{X}} \mu(\mathrm{d}x)\, N(x,\ A)\ .$$

Proposition 1.2.7 *Let N be a kernel on $\mathsf{X} \times \mathscr{Y}$ and $\mu \in \mathbb{M}_+(\mathscr{X})$. Then $\mu N \in \mathbb{M}_+(\mathscr{Y})$. If N is a Markov kernel, then $\mu N(\mathsf{Y}) = \mu(\mathsf{X})$.*

Proof. Note first that $\mu N(A) \geq 0$ for all $A \in \mathscr{Y}$ and $\mu N(\emptyset) = 0$, since $N(x, \emptyset) = 0$ for all $x \in \mathsf{X}$. Therefore, it suffices to establish the countable additivity of μN. Let $\{A_i,\ i \in \mathbb{N}\} \subset \mathscr{Y}$ be a sequence of mutually disjoint sets. For all $x \in \mathsf{X}$, $N(x, \cdot)$ is a measure on $(\mathsf{Y}, \mathscr{Y})$; thus the countable additivity implies that $N(x, \bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} N(x, A_i)$. Moreover, the function $x \mapsto N(x, A_i)$ is nonnegative and measurable for all $i \in \mathbb{N}$; thus the monotone convergence theorem yields

$$\mu N \left(\bigcup_{i=1}^{\infty} A_i \right) = \int \mu(\mathrm{d}x) N \left(x, \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \int \mu(\mathrm{d}x) N(x, A_i) = \sum_{i=1}^{\infty} \mu N(A_i) .$$

□

1.2.1 Composition of Kernels

Proposition 1.2.8 (Composition of kernels) *Let $(\mathsf{X}, \mathscr{X})$, $(\mathsf{Y}, \mathscr{Y})$, $(\mathsf{Z}, \mathscr{Z})$ be three measurable sets and let M and N be two kernels on $\mathsf{X} \times \mathscr{Y}$ and $\mathsf{Y} \times \mathscr{Z}$. There exists a kernel on $\mathsf{X} \times \mathscr{Z}$, called the composition or the product of M and N, denoted by MN, such that for all $x \in \mathsf{X}$, $A \in \mathscr{Z}$ and $f \in \mathbb{F}_+(\mathsf{Z})$,*

$$MN(x, A) = \int_{\mathsf{Y}} M(x, \mathrm{d}y) N(y, A) , \tag{1.2.1}$$

$$MNf(x) = M[Nf](x) . \tag{1.2.2}$$

Furthermore, the composition of kernels is associative.

Proof. The kernels M and N define two additive and positively homogeneous operators on $\mathbb{F}_+(\mathsf{X})$. Let $\circ$ denote the usual composition of operators. Then $M \circ N$ is positively homogeneous, and for every nondecreasing sequence of functions $\{f_n,\ n \in \mathbb{N}\}$ in $\mathbb{F}_+(\mathsf{Z})$, by the monotone convergence theorem, $\lim_{n\to\infty} M \circ N(f_n) = \lim_{n\to\infty} M(Nf_n) = M \circ N(\lim_{n\to\infty} f_n)$. Therefore, by Proposition 1.2.6, there exists a kernel, denoted by MN, such that for all $x \in \mathsf{X}$ and $f \in \mathbb{F}_+(\mathsf{Z})$,

$$M \circ N(f)(x) = M(Nf)(x) = \int MN(x, \mathrm{d}z) f(z) .$$

Hence for all $x \in \mathsf{X}$ and $A \in \mathscr{Z}$, we get

$$MN(x,A) = M(N\mathbb{1}_A)(x) \int M(x,\mathrm{d}z)N\mathbb{1}_A(z) = \int M(x,\mathrm{d}z)N(z,A) .$$

□

Given a Markov kernel N on $\mathsf{X} \times \mathscr{X}$, we may define the nth power of this kernel iteratively. For $x \in \mathsf{X}$ and $A \in \mathscr{X}$, we set $N^0(x, A) = \delta_x(A)$, and for $n \geq 1$, we define N^n inductively by

$$N^n(x,A) = \int_{\mathsf{X}} N(x,\mathrm{d}y)\,N^{n-1}(y,A) . \tag{1.2.3}$$

For integers $k,n \geq 0$, this yields the Chapman–Kolmogorov equation:

$$N^{n+k}(x,A) = \int_{\mathsf{X}} N^n(x,\mathrm{d}y)\,N^k(y,A) . \tag{1.2.4}$$

In the case of a discrete state space X, a kernel N can be seen as a matrix with nonnegative entries indexed by X. Then the kth power of the kernel N^k defined in (1.2.3) is simply the kth power of the matrix N. The Chapman–Kolmogorov equation becomes, for all $x,y \in \mathsf{X}$,

$$N^{n+k}(x,y) = \sum_{z\in\mathsf{X}} N^n(x,z)\,N^k(z,y) . \tag{1.2.5}$$

1.2.2 Tensor Products of Kernels

Proposition 1.2.9 *Let $(\mathsf{X},\mathscr{X})$, $(\mathsf{Y},\mathscr{Y})$, and $(\mathsf{Z},\mathscr{Z})$ be three measurable spaces, and let M be a kernel on $\mathsf{X} \times \mathscr{Y}$ and N a kernel on $\mathsf{Y} \times \mathscr{Z}$. Then there exists a kernel on $\mathsf{X} \times (\mathscr{Y} \otimes \mathscr{Z})$, called the tensor product of M and N, denoted by $M \otimes N$, such that for all $f \in \mathbb{F}_+(\mathsf{Y} \times \mathsf{Z}, \mathscr{Y} \otimes \mathscr{Z})$,*

$$M \otimes Nf(x) = \int_{\mathsf{Y}} M(x,\mathrm{d}y) \int_{\mathsf{Z}} f(y,z)N(y,\mathrm{d}z) . \tag{1.2.6}$$

- *If the kernels M and N are both bounded, then $M \otimes N$ is a bounded kernel.*
- *If M and N are both Markov kernels, then $M \otimes N$ is a Markov kernel.*
- *If $(\mathsf{U},\mathscr{U})$ is a measurable space and P is a kernel on $\mathsf{Z} \times \mathscr{U}$, then $(M \otimes N) \otimes P = M \otimes (N \otimes P)$; i.e., the tensor product of kernels is associative.*

Proof. Define the mapping $I : \mathbb{F}_+(\mathsf{Y} \otimes \mathsf{Z}) \to \mathbb{F}_+(\mathsf{X})$ by

$$If(x) = \int_{\mathsf{Y}} M(x,\mathrm{d}y) \int_{\mathsf{Z}} f(y,z)N(y,\mathrm{d}z) .$$

The mapping I is additive and positively homogeneous. Since $I[\lim_{n\to\infty} f_n] = \lim_{n\to\infty} I(f_n)$ for every increasing sequence $\{f_n,\ n \in \mathbb{N}\}$, by the monotone convergence theorem, Proposition 1.2.6 shows that (1.2.6) defines a kernel on $\mathsf{X} \times (\mathscr{Y} \otimes \mathscr{X})$. The proofs of the other properties are left as exercises. □

For $n \geq 1$, the nth tensorial power $P^{\otimes n}$ of a kernel P on $X \times \mathscr{Y}$ is the kernel on $(\mathsf{X}, \mathscr{X}^{\otimes n})$ defined by

$$P^{\otimes n} f(x) = \int_{\mathsf{X}^n} f(x_1, \ldots, x_n) P(x, \mathrm{d}x_1) P(x_1, \mathrm{d}x_2) \cdots P(x_{n-1}, \mathrm{d}x_n) . \tag{1.2.7}$$

If ν is a σ-finite measure on $(\mathsf{X}, \mathscr{X})$ and N is a kernel on $\mathsf{X} \times \mathscr{Y}$, then we can also define the tensor product of ν and N, denoted by $\nu \otimes N$, which is a measure on $(\mathsf{X} \times \mathsf{Y}, \mathscr{X} \otimes \mathscr{Y})$ defined by

$$\nu \otimes N(A \times B) = \int_A \nu(\mathrm{d}x) N(x, B) . \tag{1.2.8}$$

1.2.3 Sampled Kernel, m-Skeleton, and Resolvent

Definition 1.2.10 (Sampled kernel, *m*-skeleton, resolvent kernel) *Let a be a probability on $\mathbb{N}$, that is, a sequence $\{a(n),\, n \in \mathbb{N}\}$ such that $a(n) \geq 0$ for all $n \in \mathbb{N}$ and $\sum_{k=0}^{\infty} a(k) = 1$. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. The sampled kernel K_a is defined by*

$$K_a = \sum_{n=0}^{\infty} a(n) P^n . \tag{1.2.9}$$

(i) For $m \in \mathbb{N}^*$ and $a = \delta_m$, $K_{\delta_m} = P^m$ is called the m-skeleton.
(ii) If $\varepsilon \in (0,1)$ and a_ε is the geometric distribution, i.e.,

$$a_\varepsilon(n) = (1-\varepsilon)\,\varepsilon^n , \quad n \in \mathbb{N} , \tag{1.2.10}$$

then K_{a_ε} is called the resolvent kernel.

Let $\{a(n),\, n \in \mathbb{N}\}$ and $\{b(n),\, n \in \mathbb{N}\}$ be two sequences of real numbers. We denote by $\{a * b(n),\, n \in \mathbb{N}\}$ the convolution of the sequences a and b defined, for $n \in \mathbb{N}$, by

$$a * b(n) = \sum_{k=0}^{n} a(k) b(n-k) .$$

Lemma 1.2.11 *If a and b are probabilities on $\mathbb{N}$, then the sampled kernels K_a and K_b satisfy the generalized Chapman–Kolmogorov equation*

$$K_{a*b} = K_a K_b . \tag{1.2.11}$$

Proof. Applying the definition of the sampled kernel and the Chapman–Kolmogorov equation (1.2.4) yields (note that all the terms in the sum below are nonnegative)

$$
\begin{aligned}
K_{a*b} &= \sum_{n=0}^{\infty} a*b(n)P^n = \sum_{n=0}^{\infty}\sum_{m=0}^{n} a(m)b(n-m)P^n \\
&= \sum_{n=0}^{\infty}\sum_{m=0}^{n} a(m)b(n-m)P^m P^{n-m} = \sum_{m=0}^{\infty} a(m)P^m \sum_{n=m}^{\infty} b(n-m)P^{n-m} = K_a K_b \,.
\end{aligned}
$$

□

1.3 Homogeneous Markov Chains

1.3.1 Definition

We can now define the main object of this book. Let $T=\mathbb{N}$ or $T=\mathbb{Z}$.

Definition 1.3.1 (Homogeneous Markov chain) *Let $(\mathsf{X},\mathscr{X})$ be a measurable space and let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$. Let $(\Omega,\mathscr{F},\{\mathscr{F}_k,\ k\in T\},\mathbb{P})$ be a filtered probability space. An adapted stochastic process $\{(X_k,\mathscr{F}_k),\ k\in T\}$ is called a homogeneous Markov chain with kernel P if for all $A\in\mathscr{X}$ and $k\in T$,*

$$
\mathbb{P}\left(X_{k+1}\in A \,|\, \mathscr{F}_k\right) = P(X_k,A) \quad \mathbb{P}-\text{a.s.} \tag{1.3.1}
$$

If $T=\mathbb{N}$, the distribution of X_0 is called the initial distribution.

Remark 1.3.2. Condition (1.3.1) is equivalent to $\mathbb{E}\left[f(X_{k+1})|\,\mathscr{F}_k\right] = Pf(X_k) \quad \mathbb{P}-$ a.s. for all $f\in\mathbb{F}_+(\mathsf{X})\cup\mathbb{F}_b(\mathsf{X})$. ▲

Remark 1.3.3. Let $\{(X_k,\mathscr{F}_k),\ k\in T\}$ be a homogeneous Markov chain. Then $\{(X_k,\mathscr{F}_k^X),\ k\in T\}$ is also a homogeneous Markov chain. Unless specified otherwise, we will always consider the natural filtration, and we will simply write that $\{X_k,\ k\in T\}$ is a homogeneous Markov chain. ▲

From now on, unless otherwise specified, we will consider $T=\mathbb{N}$. The most important property of a Markov chain is that its finite-dimensional distributions are entirely determined by the initial distribution and its kernel.

Theorem 1.3.4. *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$, and ν a probability measure on $(\mathsf{X},\mathscr{X})$. An X-valued stochastic process $\{X_k,\ k\in\mathbb{N}\}$ is a homogeneous*

Markov chain with kernel P and initial distribution ν if and only if the distribution of $(X_0,\dots,X_k)$ is $\nu\otimes P^{\otimes k}$ for all $k\in\mathbb{N}$.

Proof. Fix $k\geq 0$. Let $\mathscr{H}_k$ be the subspace $\mathbb{F}_b(\mathsf{X}^{k+1},\mathscr{X}^{\otimes(k+1)})$ of measurable functions f such that

$$\mathbb{E}\left[f(X_0,\dots,X_k)\right]=\nu\otimes P^{\otimes k}(f)\ . \tag{1.3.2}$$

Let $\{f_n,\ n\in\mathbb{N}\}$ be an increasing sequence of nonnegative functions in $\mathscr{H}_k$ such that $\lim_{n\to\infty}f_n=f$ with f bounded. By the monotone convergence theorem, f belongs to $\mathscr{H}_k$. By Theorem B.2.4, the proof will be concluded if we moreover check that $\mathscr{H}_k$ contains the functions of the form

$$f_0(x_0)\cdots f_k(x_k)\,,\quad f_0,\dots,f_k\in\mathbb{F}_b(\mathsf{X})\ . \tag{1.3.3}$$

We prove this by induction. For $k=0$, (1.3.2) reduces to $\mathbb{E}\left[f_0(X_0)\right]=\nu(f_0)$, which means that ν is the distribution of X_0. For $k\geq 1$, assume that (1.3.2) holds for $k-1$ and f of the form (1.3.3). Then

$$\begin{aligned}
\mathbb{E}\left[\prod_{j=0}^{k}f_j(X_j)\right]&=\mathbb{E}\left[\prod_{j=0}^{k-1}f_j(X_j)\mathbb{E}\left[f_k(X_k)\middle|\mathscr{F}_{k-1}\right]\right]\\
&=\mathbb{E}\left[\prod_{j=0}^{k-1}f_j(X_j)Pf_k(X_{k-1})\right]\\
&=\nu\otimes P^{\otimes(k-1)}(f_0\otimes\cdots\otimes f_{k-1}Pf_k)\\
&=\nu\otimes P^{\otimes k}(f_0\otimes\cdots\otimes f_k)\ .
\end{aligned}$$

The last equality holds because $P(fPg)=P\otimes P(f\otimes g)$. This concludes the induction and the direct part of the proof.

Conversely, assume that (1.3.2) holds. This obviously implies that ν is the distribution of X_0. We must prove that for each $k\geq 1$, $f\in\mathbb{F}_+(\mathsf{X})$, and each $\mathscr{F}^X_{k-1}$-measurable random variable Y,

$$\mathbb{E}\left[f(X_k)Y\right]=\mathbb{E}\left[Pf(X_{k-1})Y\right]\ . \tag{1.3.4}$$

Let $\mathscr{G}_k$ be the set of $\mathscr{F}^X_{k-1}$-measurable random variables Y satisfying (1.3.4). Then $\mathscr{G}_k$ is a vector space, and if $\{Y_n,\ n\in\mathbb{N}\}$ is an increasing sequence of nonnegative random variables such that $Y=\lim_{n\to\infty}Y_n$ is bounded, then $Y\in\mathscr{G}_k$ by the monotone convergence theorem. The property (1.3.2) implies (1.3.4) for $Y=\prod_{i=0}^{k-1}f_i(X_i)$, where for $j\geq 0$, $f_j\in\mathbb{F}_b(\mathsf{X})$. The proof is concluded as previously by applying Theorem B.2.4. □

Corollary 1.3.5 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let ν be a probability measure on $(\mathsf{X}, \mathscr{X})$. Let $\{X_k,\ k \in \mathbb{N}\}$ be a homogeneous Markov chain on X with kernel P and initial distribution ν. Then for all $n, k \geq 0$, the distribution of $(X_n, \ldots, X_{n+k})$ is $\nu P^n \otimes P^{\otimes k}$, and for all $n, m, k \geq 0$ and all bounded measurable functions f defined on X^k,*

$$\mathbb{E}\left[f(X_{n+m}, \ldots, X_{n+m+k}) \,|\, \mathscr{F}_n^X\right] = P^m \otimes P^{\otimes k} f(X_n) .$$

1.3.2 Homogeneous Markov Chains and Random Iterative Sequences

Under weak conditions on the structure of the state space X, every homogeneous Markov chain $\{X_k,\ k \in \mathbb{N}\}$ with values in X may be represented as a random iterative sequence, i.e., $X_{k+1} = f(X_k, Z_{k+1})$, where $\{Z_k,\ k \in \mathbb{N}\}$ is a sequence of i.i.d. random variables with values in a measurable space $(\mathsf{Z}, \mathscr{Z})$, X_0 is independent of $\{Z_k,\ k \in \mathbb{N}\}$, and f is a measurable function from $(\mathsf{X} \times \mathsf{Z}, \mathscr{X} \otimes \mathscr{Z})$ into $(\mathsf{X}, \mathscr{X})$.

This can be easily proved for a real-valued Markov chain $\{X_k,\ k \in \mathbb{N}\}$ with initial distribution ν and Markov kernel P. Let X be a real-valued random variable and let $F(x) = \mathbb{P}(X \leq x)$ be the cumulative distribution function of X. Let F^{-1} be the quantile function, defined as the generalized inverse of F by

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\} . \tag{1.3.5}$$

The right continuity of F implies that $u \leq F(x) \Leftrightarrow F^{-1}(u) \leq x$. Therefore, if Z is uniformly distributed on $[0,1]$, then $F^{-1}(Z)$ has the same distribution as X, since $\mathbb{P}(F^{-1}(Z) \leq t) = \mathbb{P}(Z \leq F(t)) = F(t) = \mathbb{P}(X \leq t)$.

Define $F_0(t) = \nu((-\infty, t])$ and $g = F_0^{-1}$. Consider the function F from $\mathbb{R} \times \mathbb{R}$ to $[0,1]$ defined by $F(x, x') = P(x, (-\infty, x'])$. Then for each $x \in \mathbb{R}$, $F(x, \cdot)$ is a cumulative distribution function. Let the associated quantile function $f(x, \cdot)$ be defined by

$$f(x,u) = \inf\left\{x' \in \mathbb{R} : F(x,x') \geq u\right\} . \tag{1.3.6}$$

The function $(x,u) \mapsto f(x,u)$ is Borel measurable, since $(x,x') \mapsto F(x,x')$ is itself a Borel measurable function. If Z is uniformly distributed on $[0,1]$, then for all $x \in \mathbb{R}$ and $A \in \mathscr{B}(\mathbb{R})$, we obtain

$$\mathbb{P}\left(f(x,Z) \in A\right) = P(x,A) .$$

Let $\{Z_k,\ k \in \mathbb{N}\}$ be a sequence of i.i.d. random variables, uniformly distributed on $[0,1]$. Define a sequence of random variables $\{X_k,\ k \in \mathbb{N}\}$ by $X_0 = g(Z_0)$, and for $k \geq 0$,

$$X_{k+1} = f(X_k, Z_{k+1}) .$$

Then $\{X_k,\ k \in \mathbb{N}\}$ is a Markov chain with Markov kernel P and initial distribution ν.

We state without proof a general result for reference only, since it will not be needed in the sequel.

Theorem 1.3.6. *Let $(\mathsf{X}, \mathscr{X})$ be a measurable space and assume that $\mathscr{X}$ is countably generated. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let ν be a probability on $(\mathsf{X}, \mathscr{X})$. Let $\{Z_k,\ k \in \mathbb{N}\}$ be a sequence of i.i.d. random variables uniformly distributed on $[0,1]$. There exist a measurable function g from $([0,1], \mathscr{B}([0,1]))$ to $(\mathsf{X}, \mathscr{X})$ and a measurable function f from $(\mathsf{X} \times [0,1], \mathscr{X} \otimes \mathscr{B}([0,1]))$ to $(\mathsf{X}, \mathscr{X})$ such that the sequence $\{X_k,\ k \in \mathbb{N}\}$ defined by $X_0 = g(Z_0)$ and $X_{k+1} = f(X_k, Z_{k+1})$ for $k \geq 0$ is a Markov chain with initial distribution ν and Markov kernel P.*

From now on, we shall deal almost exclusively with homogeneous Markov chains, and for simplicity, we shall omit mentioning "homogeneous" in the statements.

Definition 1.3.7 (Markov chain of order p) *Let $p \geq 1$ be an integer. Let $(\mathsf{X}, \mathscr{X})$ be a measurable space. Let $(\Omega, \mathscr{F}, \{\mathscr{F}_k,\ k \in \mathbb{N}\}, \mathbb{P})$ be a filtered probability space. An adapted stochastic process $\{(X_k, \mathscr{F}_k),\ k \in \mathbb{N}\}$ is called a Markov chain of order p if the process $\{(X_k, \ldots, X_{k+p-1}), k \in \mathbb{N}\}$ is a Markov chain with values in X^p.*

Let $\{X_k,\ k \in \mathbb{N}\}$ be a Markov chain of order $p \geq 2$ and let K_p be the kernel of the chain $\{\mathbf{X}_k,\ k \in \mathbb{N}\}$ with $\mathbf{X}_k = (X_k, \ldots, X_{k+p-1})$, that is,

$$\mathbb{P}\left(\mathbf{X}_1 \in A_1 \times \cdots \times A_p \,\middle|\, \mathbf{X}_0 = (x_0, \ldots, x_{p-1})\right) = K_p((x_0, \ldots, x_{p-1}), A_1 \times \cdots \times A_p) .$$

Since $\mathbf{X}_0$ and $\mathbf{X}_1$ have $p-1$ common components, the kernel K_p has a particular form. More precisely, defining the kernel K on $\mathsf{X}^p \times \mathscr{X}$ by

$$\begin{aligned} K_p(x_0, \ldots, x_{p-1}, A) &= K_p((x_0, \ldots, x_{p-1}), \mathsf{X}^{p-1} \times A) \\ &= \mathbb{P}\left(X_p \in A \,\middle|\, X_0 = x_0, \ldots, X_{p-1} = x_{p-1}\right) , \end{aligned}$$

we obtain that

$$K_p((x_0, \ldots, x_{p-1}), A_1 \times \cdots \times A_p) = \delta_{x_1}(A_1) \cdots \delta_{x_{p-1}}(A_{p-1}) K((x_0, \ldots, x_{p-1}), A_p) .$$

We thus see that an equivalent definition of a homogeneous Markov chain of order p is the existence of a kernel K on $\mathsf{X}^p \times \mathscr{X}$ such that for all $n \geq 0$,

$$\mathbb{E}\left[X_{n+p} \in A \,\middle|\, \mathscr{F}^X_{n+p-1}\right] = K((X_n, \ldots, X_{n+p-1}), A) .$$

1.4 Invariant Measures and Stationarity

Definition 1.4.1 (Invariant measure) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$.*

- *A nonzero measure μ is said to be subinvariant if μ is σ-finite and $\mu P \leq \mu$.*
- *A nonzero measure μ is said to be invariant if it is σ-finite and $\mu P = \mu$.*
- *A nonzero signed measure μ is said to be invariant if $\mu P = \mu$.*

A Markov kernel P is said to be positive if it admits an invariant probability measure.

A Markov kernel may admit one or more than one invariant measure, or none if X is not finite. Consider the kernel P on $\mathbb{N}$ such that $P(x, x+1) = 1$. Then P does not admit an invariant measure. Considered as a kernel on $\mathbb{Z}$, P admits the counting measure as its unique invariant measure. The kernel P on $\mathbb{Z}$ such that $P(x, x+2) = 1$ admits two invariant measures with disjoint supports: the counting measure on the even integers and the counting measure on the odd integers.

It must be noted that an invariant measure is σ-finite by definition. Consider again the kernel P defined by $P(x, x+1) = 1$, now as a kernel on $\mathbb{R}$. The counting measure on $\mathbb{R}$ satisfies $\mu P = \mu$, but it is not σ-finite. We will provide in Section 3.6 a criterion that ensures that a measure μ that satisfies $\mu = \mu P$ is σ-finite.

If an invariant measure is finite, it may be normalized to an invariant probability measure. The fundamental role of an invariant probability measure is illustrated by the following result. Recall that a stochastic process $\{X_k,\ k \in \mathbb{N}\}$ defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is said to be stationary if for all integers $k, p \geq 0$, the distribution of the random vector $(X_k, \ldots, X_{k+p})$ does not depend on k.

Theorem 1.4.2. *Let $(\Omega, \mathscr{F}, \{\mathscr{F}_k,\ k \in \mathbb{N}\}, \mathbb{P})$ be a filtered probability space and let P be a Markov kernel on a measurable space $(\mathsf{X}, \mathscr{X})$. A Markov chain $\{(X_k, \mathscr{F}_k),\ k \in \mathbb{N}\}$ defined on $(\Omega, \mathscr{F}, \{\mathscr{F}_k,\ k \in \mathbb{N}\}, \mathbb{P})$ with kernel P is a stationary process if and only if its initial distribution is invariant with respect to P.*

Proof. Let π denote the initial distribution. If the chain $\{X_k\}$ is stationary, then the marginal distribution is constant. In particular, the distribution of X_1 is equal to the distribution of X_0, which means precisely that $\pi P = \pi$. Thus π is invariant. Conversely, if $\pi P = \pi$, then $\pi P^h = \pi$ for all $h \geq 1$. Then for all integers h and n, by Corollary 1.3.5, the distribution of $(X_h, \ldots, X_{n+h})$ is $\pi P^h \otimes P^{\otimes n} = \pi \otimes P^{\otimes n}$. □

For a finite signed measure ξ on $(\mathsf{X}, \mathscr{X})$, we denote by ξ^+ and ξ^- the positive and negative parts of ξ (see Theorem D.1.3). Recall that ξ^+ and ξ^- are two mutually singular measures such that $\xi = \xi^+ - \xi^-$. A set S such that $\xi^+(S^c) = \xi^-(S) = 0$ is called a Jordan set for ξ.

Lemma 1.4.3 *Let P be a Markov kernel and λ an invariant signed measure. Then λ^+ is also invariant.*

Proof. Let S be a Jordan set for λ. For all $B \in \mathscr{X}$,

$$\begin{aligned}\lambda^+ P(B) \geq \lambda^+ P(B \cap S) &= \int P(x, B \cap S) \lambda^+(\mathrm{d}x) \\ &\geq \int P(x, B \cap S) \lambda(\mathrm{d}x) = \lambda(B \cap S) = \lambda^+(B) \,. \end{aligned} \tag{1.4.1}$$

Since $P(x, \mathsf{X}) = 1$ for all $x \in \mathsf{X}$, it follows that $\lambda^+ P(\mathsf{X}) = \lambda^+(\mathsf{X})$. This and the inequality (1.4.1) imply that $\lambda^+ P = \lambda^+$. □

Definition 1.4.4 (Absorbing set) *A set $B \in \mathscr{X}$ is called absorbing if $P(x, B)=1$ for all $x \in B$.*

This definition subsumes that the empty set is absorbing. Of course, the interesting absorbing sets are nonempty.

Proposition 1.4.5 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an invariant probability measure π. If $B \in \mathscr{X}$ is an absorbing set, then $\pi_B = \pi(B \cap \cdot)$ is an invariant finite measure. Moreover, if the invariant probability measure is unique, then $\pi(B) \in \{0, 1\}$.*

Proof. Let B be an absorbing set. Using that $\pi_B \leq \pi$, $\pi P = \pi$, and B is absorbing, we get that for all $C \in \mathscr{X}$,

$$\pi_B P(C) = \pi_B P(C \cap B) + \pi_B P(C \cap B^c) \leq \pi P(C \cap B) + \pi_B P(B^c) = \pi(C \cap B) = \pi_B(C) \,.$$

Replacing C by C^c and noting that $\pi_B P(\mathsf{X}) = \pi_B(\mathsf{X}) < \infty$, it follows that π_B is an invariant finite measure. To complete the proof, assume that P has a unique invariant probability measure. If $\pi(B) > 0$, then $\pi_B / \pi(B)$ is an invariant probability measure and is therefore equal to π. Since $\pi_B(B^c) = 0$, we get $\pi(B^c) = 0$. Thus $\pi(B) \in \{0, 1\}$. □

Theorem 1.4.6. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Then*

(i) The set of invariant probability measures for P is a convex subset of $\mathbb{M}_+(\mathscr{X})$.
(ii) For every two distinct invariant probability measures π, π' for P, the finite measures $(\pi - \pi')^+$ and $(\pi - \pi')^-$ are nontrivial, mutually singular, and invariant for P.

Proof. (i) P is an additive and positively homogeneous operator on $\mathbb{M}_+(\mathscr{X})$. Therefore, if π, π' are two invariant probability measures for P, then for every scalar $a \in [0,1]$, using first the linearity and then the invariance,

$$(a\pi + (1-a)\pi')P = a\pi P + (1-a)\pi' P = a\pi + (1-a)\pi' .$$

(ii) We apply Lemma 1.4.3 to the nonzero signed measure $\lambda = \pi - \pi'$. The measures $(\pi - \pi')^+$ and $(\pi - \pi')^-$ are singular, invariant, and nontrivial, since

$$(\pi - \pi')^+(\mathsf{X}) = (\pi - \pi')^-(\mathsf{X}) = \frac{1}{2}|\pi - \pi'|(\mathsf{X}) > 0 .$$

□

We will see in the forthcoming chapters that it is sometimes more convenient to study one iterate P^k of a Markov kernel than P itself. However, if P^k admits an invariant probability measure, then so does P.

Lemma 1.4.7 *Let P be a Markov kernel. For every $k \geq 1$, P^k admits an invariant probability measure if and only if P admits an invariant probability measure.*

Proof. If π is invariant for P, then it is obviously invariant for P^k for every $k \geq 1$. Conversely, if $\tilde{\pi}$ is invariant for P^k, set $\pi = k^{-1} \sum_{i=0}^{k-1} \tilde{\pi} P^i$. Then π is an invariant probability measure for P. Indeed, since $\tilde{\pi} = \tilde{\pi} P^k$, we obtain

$$\pi P = \frac{1}{k} \sum_{i=1}^{k} \tilde{\pi} P^i = \frac{1}{k} \sum_{i=1}^{k-1} \tilde{\pi} P^i + \tilde{\pi} P^k = \frac{1}{k} \sum_{i=1}^{k-1} \pi P^i + \tilde{\pi} = \pi .$$

□

1.5 Reversibility

Definition 1.5.1 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. A σ-finite measure ξ on $\mathscr{X}$ is said to be reversible with respect to P if the measure $\xi \otimes P$ on $\mathscr{X} \otimes \mathscr{X}$ is symmetric, i.e., for all $(A, B) \in \mathscr{X} \times \mathscr{X}$,*

$$\xi \otimes P(A \times B) = \xi \otimes P(B \times A) , \tag{1.5.1}$$

where $\xi \otimes P$ is defined in (1.2.8).

Equivalently, reversibility means that for all bounded measurable functions f defined on $(\mathsf{X} \times \mathsf{X}, \mathscr{X} \otimes \mathscr{X})$,

$$\iint_{\mathsf{X}\times\mathsf{X}} \xi(\mathrm{d}x)P(x,\mathrm{d}x')f(x,x') = \iint_{\mathsf{X}\times\mathsf{X}} \xi(\mathrm{d}x)P(x,\mathrm{d}x')f(x',x) \,. \tag{1.5.2}$$

If X is a countable state space, a (finite or σ-finite) measure ξ is reversible with respect to P if and only if for all $(x,x') \in \mathsf{X}\times\mathsf{X}$,

$$\xi(x)P(x,x') = \xi(x')P(x',x) \,, \tag{1.5.3}$$

a condition often referred to as the detailed balance condition.

If $\{X_k,\ k\in\mathbb{N}\}$ is a Markov chain with kernel P and initial distribution ξ, the reversibility condition (1.5.1) means precisely that (X_0,X_1) and (X_1,X_0) have the same distribution, i.e., for all $f\in\mathbb{F}_b(\mathsf{X}\times\mathsf{X},\mathscr{X}\otimes\mathscr{X})$,

$$\begin{aligned}\mathbb{E}_\xi[f(X_0,X_1)] &= \iint \xi(\mathrm{d}x_0)P(x_0,\mathrm{d}x_1)f(x_0,x_1) \\ &= \iint \xi(\mathrm{d}x_0)P(x_0,\mathrm{d}x_1)f(x_1,x_0) = \mathbb{E}_\xi[f(X_1,X_0)] \,.\end{aligned} \tag{1.5.4}$$

This implies in particular that the distribution of X_1 is the same as that of X_0, and this means that ξ is P-invariant: reversibility implies invariance. This property can be extended to all finite-dimensional distributions.

Proposition 1.5.2 *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$ and $\xi\in\mathbb{M}_1(\mathscr{X})$, where $\mathbb{M}_1(\mathscr{X})$ is the set of probability measures on $\mathscr{X}$. If ξ is reversible with respect to P, then*

(i) ξ is P-invariant;
(ii) the homogeneous Markov chain $\{X_k,\ k\in\mathbb{N}\}$ with Markov kernel P and initial distribution ξ is reversible, i.e., for all $n\in\mathbb{N}$, $(X_0,\ldots,X_n)$ and $(X_n,\ldots,X_0)$ have the same distribution.

Proof. (i) Using (1.5.1) with $A=\mathsf{X}$ and $B\in\mathscr{X}$, we get

$$\xi P(B) = \xi\otimes P(\mathsf{X}\times B) = \xi\otimes P(B\times\mathsf{X}) = \int \xi(\mathrm{d}x)\mathbb{1}_B(x)P(x,\mathsf{X}) = \xi(B) \,.$$

(ii) The proof is by induction. For $n=1$, (1.5.4) shows that (X_0,X_1) and (X_1,X_0) have the same distribution. Assume that such is the case for some $n\geq 1$. By the Markov property, X_0 and $(X_1,\ldots,X_n)$ are conditionally independent given X_1 and X_{n+1}, and $(X_n,\ldots,X_0)$ are conditionally independent given X_1. Moreover, by stationarity and reversibility, (X_{n+1},X_n) has the same distribution as (X_0,X_1), and by the induction assumption, $(X_1,\ldots,X_{n+1})$ and $(X_n,\ldots,X_0)$ have the same distribution. This proves that $(X_0,\ldots,X_{n+1})$ and $(X_{n+1},\ldots,X_0)$ have the same distribution.

□

1.6 Markov Kernels on $\mathrm{L}^p(\pi)$

Let $(\mathsf{X}, \mathscr{X})$ be a measurable space and $\pi \in \mathbb{M}_1(\mathscr{X})$. For $p \in [1, \infty)$ and f a measurable function on $(\mathsf{X}, \mathscr{X})$, we set

$$\|f\|_{\mathrm{L}^p(\pi)} = \left\{ \int |f(x)|^p \pi(\mathrm{d}x) \right\}^{1/p} ,$$

and for $p = \infty$, we set

$$\|f\|_{\mathrm{L}^\infty(\pi)} = \operatorname{esssup}_\pi (|f|) \ .$$

For $p \in [1, \infty]$, we denote by $\mathrm{L}^p(\pi)$ the space of all measurable functions on $(\mathsf{X}, \mathscr{X})$ for which $\|f\|_{\mathrm{L}^p(\pi)} < \infty$.

Remark 1.6.1. The maps $\|\cdot\|_{\mathrm{L}^p(\pi)}$ are not norms but only seminorms, since $\|f\|_{\mathrm{L}^p(\pi)} = 0$ implies $\pi(f = 0) = 1$ but not $f \equiv 0$. Define the relation $\backsim_\pi$ by $f \backsim_\pi g$ if and only if $\pi(f \neq g) = 0$. Then the quotient spaces $\mathrm{L}^p(\pi)/\backsim_\pi$ are Banach spaces, but the elements of these spaces are no longer functions, but equivalence classes of functions. For the sake of simplicity, as is customary, this distinction will be tacitly understood, and we will identify $\mathrm{L}^p(\pi)$ and its quotient by the relation $\backsim_\pi$ and treat it as a Banach space of functions. ▲

If $f \in \mathrm{L}^p(\pi)$ and $g \in \mathrm{L}^q(\pi)$, with $1/p + 1/q = 1$, then $fg \in \mathrm{L}^1(\pi)$, since by Hölder's inequality,

$$\|fg\|_{\mathrm{L}^1(\pi)} \leq \|f\|_{\mathrm{L}^p(\pi)} \|g\|_{\mathrm{L}^q(\pi)} \ . \tag{1.6.1}$$

Lemma 1.6.2 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ that admits an invariant probability measure π.*

(i) Let $f, g \in \mathbb{F}_+(\mathsf{X}) \cup \mathbb{F}_b(\mathsf{X})$. If $f = g$ $\pi - a.e.$, then $Pf = Pg$ $\pi - a.e.$
(ii) Let $p \in [1, \infty)$ and $f \in \mathbb{F}_+(\mathsf{X}) \cup \mathbb{F}_b(\mathsf{X})$. If $f \in \mathrm{L}^p(\pi)$, then $Pf \in \mathrm{L}^p(\pi)$ and

$$\|Pf\|_{\mathrm{L}^p(\pi)} \leq \|f\|_{\mathrm{L}^p(\pi)} \ .$$

Proof. (i) Write $N = \{x \in \mathsf{X} : f(x) \neq g(x)\}$. By assumption, $\pi(N) = 0$, and since $\int_{\mathsf{X}} \pi(\mathrm{d}x) P(x, N) = \pi(N) = 0$, it is also the case that $P(x, N) = 0$ for all x in a subset X_0 such that $\pi(\mathsf{X}_0) = 1$. Then for all $x \in \mathsf{X}_0$, we have

$$\int P(x, \mathrm{d}y) f(y) = \int_{N^c} P(x, \mathrm{d}y) f(y) = \int_{N^c} P(x, \mathrm{d}y) g(y) = \int P(x, \mathrm{d}y) g(y) \ .$$

This proves (i).

(ii) Applying Jensen's inequality and then Fubini's theorem, we obtain

$$\pi(|Pf|^p) = \int \left| \int f(y) P(x, \mathrm{d}y) \right|^p \pi(\mathrm{d}x) \leq \int \int |f(y)|^p P(x, \mathrm{d}y) \pi(\mathrm{d}x) = \pi(|f|^p) \ .$$

□

The next proposition then allows us to consider P as a bounded linear operator on the spaces $\mathrm{L}^p(\pi)$, where $p \in [1,\infty]$.

Proposition 1.6.3 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. For every $p \in [1,\infty]$, P can be extended to a bounded linear operator on $\mathrm{L}^p(\pi)$, and*

$$|||P|||_{\mathrm{L}^p(\pi)} = 1 \; . \tag{1.6.2}$$

Proof. For $f \in \mathrm{L}^1(\pi)$, define

$$A_f = \{x \in \mathsf{X} : \; P|f|(x) < \infty\} = \left\{x \in \mathsf{X} : \; f \in \mathrm{L}^1(P(x,\cdot))\right\} \; . \tag{1.6.3}$$

Since $\pi(P|f|) = \pi(|f|) < \infty$, we have $\pi(A_f) = 1$, and we may therefore define Pf on the whole space X by setting

$$Pf(x) = \begin{cases} \int_{\mathsf{X}} f(y) P(x,\mathrm{d}y) \; , & \text{if } x \in A_f \; , \\ 0 & \text{otherwise.} \end{cases} \tag{1.6.4}$$

This definition yields

$$\pi(|Pf|) = \pi(|Pf|\mathbb{1}_{A_f}) \leq \pi(P|f|\mathbb{1}_{A_f}) = \pi(P|f|) = \pi(|f|) \; .$$

That is $\|Pf\|_{\mathrm{L}^1(\pi)} \leq \|f\|_{\mathrm{L}^1(\pi)}$.

Furthermore, if $\pi(f = \tilde{f}) = 1$, then $\pi(P|f| = P|\tilde{f}|) = 1$ by Lemma 1.6.2 (i), and therefore $\pi(A_f \Delta A_{\tilde{f}}) = 0$. Hence it is also the case that $\pi(Pf = P\tilde{f}) = 1$. This shows that P acts on equivalence classes of functions and can be defined on the Banach space $\mathrm{L}^1(\pi)$. It is easily seen that for all $f, g \in \mathrm{L}^1(\pi)$ and $t \in \mathbb{R}$, $P(tf) = tPf$, $P(f+g) = Pf + Pg$, and we have just shown that $\|Pf\|_{\mathrm{L}^1(\pi)} \leq \|f\|_{\mathrm{L}^1(\pi)} < \infty$. Therefore, the relation (1.6.4) defines a bounded operator on the Banach space $\mathrm{L}^1(\pi)$.

Let $p \in [1,\infty)$ and $f \in \mathrm{L}^p(f)$. Then $f \in \mathrm{L}^1(\pi)$, and thus we can define Pf. Applying Lemma 1.6.2 (ii) to $|f|$ proves that $|||P|||_{\mathrm{L}^p(\pi)} \leq 1$ for $p < \infty$.

For $f \in \mathrm{L}^\infty(\pi)$, one has $\|f\|_{\mathrm{L}^\infty(\pi)} = \lim_{p\to\infty} \|f\|_{\mathrm{L}^p(\pi)}$, and so $\|Pf\|_{\mathrm{L}^\infty(\pi)} \leq \|f\|_{\mathrm{L}^\infty(\pi)}$, and thus it is also the case that $|||P|||_{\mathrm{L}^\infty(\pi)} \leq 1$.

Finally, $P\mathbb{1}_{\mathsf{X}} = \mathbb{1}_{\mathsf{X}}$, and thus (1.6.2) holds. □

1.7 Exercises

1.1. Let $(\mathsf{X}, \mathscr{X})$ be a measurable space, μ a σ-finite measure, and $n : \mathsf{X} \times \mathsf{X} \to \mathbb{R}_+$ a nonnegative function. For $x \in \mathsf{X}$ and $A \in \mathscr{X}$, define $N(x,A) = \int_A n(x,y)\mu(\mathrm{d}y)$. Show that for for every $k \in \mathbb{N}^*$, the kernel N^k has a density with respect to μ.

1.2. Let $\{Z_n,\ n \in \mathbb{N}\}$ be an i.i.d. sequence of random variables independent of X_0. Define recursively $X_n = \phi X_{n-1} + Z_n$.

1. Show that $\{X_n,\ n \in \mathbb{N}\}$ defines a time-homogeneous Markov chain.
2. Write its Markov kernel in the cases (i) Z_1 is a Bernoulli random variable with probability of success $1/2$ and (ii) the law of Z_1 has density q with respect to the Lebesgue measure.
3. Assume that Z_1 is Gaussian with zero mean and variance σ^2 and that X_0 is Gaussian with zero mean and variance σ_0^2. Compute the law of X_k for every $k \in \mathbb{N}$. Show that if $|\phi| < 1$, then there exists at least an invariant probability.

1.3. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $\{Z_k,\ k \in \mathbb{N}^*\}$ an i.i.d. sequence of real-valued random variables defined on $(\Omega, \mathscr{F}, \mathbb{P})$. Let U be a real-valued random variable independent of $\{Z_k,\ k \in \mathbb{N}\}$ and consider the sequence defined recursively by $X_0 = U$ and for $k \geq 1$, $X_k = X_{k-1} + Z_k$.

1. Show that $\{X_k,\ k \in \mathbb{N}\}$ is a homogeneous Markov chain.

Assume that the law of Z_1 has a density with respect to the Lebesgue measure.

2. Show that the kernel of this Markov chain has a density.

Consider now the sequence defined by $Y_0 = U^+$ and for $k \geq 1$, $Y_k = (Y_{k-1} + Z_k)^+$.

3. Show that $\{Y_k,\ k \in \mathbb{N}\}$ is a Markov chain.
4. Write the associated kernel.

1.4. In Section 1.2.3, the sampled kernel was introduced. We will see in this exercise how this kernel is related to a Markov chain sampled at random time instants. Let $(\Omega_0, \mathscr{F}, \{\mathscr{F}_n,\ n \in \mathbb{N}\}, \mathbb{P})$ be a filtered probability space and $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ a homogeneous Markov chain with Markov kernel P and initial distribution $\nu \in \mathbb{M}_1(\mathscr{X})$. Let $(\Omega_1, \mathscr{G}, \mathbb{Q})$ be a probability space and $\{Z_n,\ n \in \mathbb{N}^*\}$ a sequence of independent and identically distributed (i.i.d.) integer-valued random variables distributed according to $a = \{a(k),\ k \in \mathbb{N}\}$, i.e., for every $n \in \mathbb{N}^*$ and $k \in \mathbb{N}$, $\mathbb{Q}(Z_n = k) = a(k)$. Set $S_0 = 0$, and for $n \geq 1$, define recursively $S_n = S_{n-1} + Z_n$.

Put $\Omega = \Omega_0 \times \Omega_1$, $\mathscr{H} = \mathscr{F} \otimes \mathscr{G}$, and for every $n \in \mathbb{N}$,

$$\mathscr{H}_n = \sigma(A \times \{S_j = k\}, A \in \mathscr{F}_k, k \in \mathbb{N}, j \leq n) .$$

1. Show that $\{\mathscr{H}_n,\ n \in \mathbb{N}\}$ is a filtration.

Put $\bar{\mathbb{P}} = \mathbb{P} \otimes \mathbb{Q}$ and consider the filtered probability space $(\Omega, \mathscr{H}, \{\mathscr{H}_n,\ n \in \mathbb{N}\}, \bar{\mathbb{P}})$, where $\mathscr{H} = \bigvee_{n=0}^{\infty} \mathscr{H}_n$. For every $n \in \mathbb{N}$, set $Y_n = X_{S_n}$.

2. Show that for every $k, n \in \mathbb{N}$, $f \in \mathbb{F}_+(\mathsf{X})$ and $A \in \mathscr{F}_k$,

$$\bar{\mathbb{E}}[\mathbb{1}_{A \times \{S_n = k\}} f(Y_{n+1})] = \bar{\mathbb{E}}\left[\mathbb{1}_{A \times \{S_n = k\}} K_a f(Y_n)\right] ,$$

where K_a is the sampled kernel defined in Definition 1.2.10.

3. Show that $\{(Y_n, \mathscr{H}_n),\ n \in \mathbb{N}\}$ is a homogeneous Markov chain with initial distribution ν and transition kernel K_a.

1.5. Let $(\mathsf{X}, \mathscr{X})$ be a measurable space, $\mu \in \mathbb{M}_+(\mathscr{X})$ a σ-finite measure, and $p \in \mathbb{F}_+(X^2, \mathscr{X}^{\otimes 2})$ a positive function ($p(x,y) > 0$ for all $(x,y) \in \mathsf{X} \times \mathsf{X}$) such that for all $x \in \mathsf{X}$, $\int_{\mathsf{X}} p(x,y)\mu(\mathrm{d}y) = 1$.

For all $x \in \mathsf{X}$ and $A \in \mathscr{X}$, set $P(x,A) = \int_A p(x,y)\mu(\mathrm{d}y)$.

1. Let π be an invariant probability measure. Show that for all $f \in \mathbb{F}_+(\mathsf{X})$, $\pi(f) = \int_{\mathsf{X}} f(y)q(y)\mu(\mathrm{d}y)$ with $q(y) = \int_{\mathsf{X}} p(x,y)\pi(\mathrm{d}x)$.
2. Deduce that every invariant probability measure is equivalent to μ.
3. Show that P admits at most an invariant probability [Hint: Use Theorem 1.4.6 (ii).].

1.6. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let π be an invariant probability and $\mathsf{X}_1 \subset \mathsf{X}$ with $\pi(\mathsf{X}_1) = 1$. We will show that there exists $B \subset \mathsf{X}_1$ such that $\pi(B) = 1$ and $P(x,B) = 1$ for all $x \in B$ (i.e., B is absorbing for P).

1. Show that there exists a decreasing sequence $\{\mathsf{X}_i, i \geq 1\}$ of sets $\mathsf{X}_i \in \mathscr{X}$ such that $\pi(\mathsf{X}_i) = 1$ for all $i = 1, 2, \ldots$ and $P(x,\ \mathsf{X}_i) = 1$, for all $x \in \mathsf{X}_{i+1}$.
2. Define $B = \bigcap_{i=1}^{\infty} \mathsf{X}_i \in \mathscr{X}$. Show that B is not empty.
3. Show that B is absorbing and conclude.

1.7. Consider a Markov chain whose state space $\mathsf{X} = (0,1)$ is the open unit interval. If the chain is at x, then pick one of the two intervals $(0,\ x)$ and $(x,\ 1)$ with equal probability $1/2$ and move to a point y according to the uniform distribution on the chosen interval. Formally, let $\{U_k,\ k \in \mathbb{N}\}$ be a sequence of i.i.d. random variables uniformly distributed on $(0,1)$; let $\{\varepsilon_k,\ k \in \mathbb{N}\}$ be a sequence of i.i.d. Bernoulli random variables with probability of success $1/2$, independent of $\{U_k,\ k \in \mathbb{N}\}$; and let X_0 be independent of $\{(U_k, \varepsilon_k),\ k \in \mathbb{N}\}$ with distribution ξ on $(0,1)$. Define the sequence $\{X_k,\ k \in \mathbb{N}^*\}$ as follows:

$$X_k = \varepsilon_k X_{k-1} U_k + (1 - \varepsilon_k)\left\{X_{k-1} + U_k(1 - X_{k-1})\right\} . \tag{1.7.1}$$

1. Show that the kernel of this Markov chain has a density with respect to Lebesgue measure on the interval $(0,1)$, given by

$$k(x,y) = \frac{1}{2x}\mathbb{1}_{(0,x)}(y) + \frac{1}{2(1-x)}\mathbb{1}_{(x,1)}(y) . \tag{1.7.2}$$

Assume that this Markov kernel admits an invariant probability that possesses a density with respect to Lebesgue measure, which will be denoted by p.

2. Show that p must satisfy the following equation:

$$p(y) = \int_0^1 k(x,\ y)p(x)\mathrm{d}x = \frac{1}{2}\int_y^1 \frac{p(x)}{x}\mathrm{d}x + \frac{1}{2}\int_0^y \frac{p(x)}{1-x}\mathrm{d}x . \tag{1.7.3}$$

3. Assuming that p is positive, show that

$$\frac{p'(y)}{p(y)} = \frac{1}{2}\left(-\frac{1}{y} + \frac{1}{1-y}\right) .$$

4. Show that the solutions for this differential equation are given by

$$p_C(y) = \frac{C}{\sqrt{y(1-y)}} , \tag{1.7.4}$$

where $C \in \mathbb{R}$ is a constant, and that $C = \pi^{-1}$ yields a probability density function.

1.8. Let P be the Markov kernel defined on $[0,1] \times \mathscr{B}([0,1])$ by

$$P(x,\cdot) = \begin{cases} \delta_{x/2} & \text{if } x > 0 , \\ \delta_1 & \text{if } x = 0 . \end{cases}$$

Prove that P admits no invariant measure.

1.9. Show that if the Markov kernel P is reversible, then P^m is also reversible.

1.10. Prove (1.5.2).

1.11. The following model, called the Ehrenfest or dog–flea model, is a Markov chain on a finite state space $\{0,\dots,N\}$, where $N > 1$ is a fixed integer. Balls (or particles) numbered 1 to N are divided among two urns A and B. At each step, an integer i is drawn at random and the ball numbered i is moved to the other urn. Denote by X_n the number X_n of balls at time n in urn A.

1. Show that $\{X_n,\ n \in \mathbb{N}\}$ is a Markov chain on $\{0,\dots,N\}$ and compute its kernel P.
2. Prove that the binomial distribution $B(N,1/2)$ is reversible with respect to the kernel P.
3. Show that for $n \geq 1$, $\mathbb{E}\left[X_n \,\middle|\, \mathscr{F}^X_{n-1}\right] = (1-2/N)X_{n-1} + 1$.
4. Prove that $\lim_{n\to\infty} \mathbb{E}_x[X_n] = N/2$.

1.12. Let X be a finite set and π a probability on X such that $\pi(x) > 0$ for all $x \in \mathsf{X}$. Let M be a Markov transition matrix reversible with respect to π, i.e., $\pi(x)M(x,y) = \pi(y)M(y,x)$ for all $x,y \in \mathsf{X}$. Let D be a diagonal matrix whose diagonal elements are $\pi(x)$, $x \in \mathsf{X}$.

1. Show that $DM = M^T D$.
2. Show that for all $(x,y) \in \mathsf{X} \times \mathsf{X}$ and $k \in \mathbb{N}$, $\pi(x)M^k(x,y) = \pi(y)M^k(y,x)$.
3. Show that $T = D^{1/2}MD^{-1/2}$ can be orthogonally diagonalized, i.e., $T = \Gamma\beta\Gamma^T$, where β is a diagonal matrix (whose diagonal elements are the eigenvalues of T) and Γ is orthogonal.
4. Show that M can be diagonalized and has the same eigenvalues as T.
5. Compute the left and right eigenvectors of M as functions of Γ and D. Show that the right eigenvectors are orthogonal in $\mathrm{L}^2(\pi)$ and the left eigenvectors are orthogonal in $\mathrm{L}^2(\pi^{-1})$, where π^{-1} is the measure on X such that $\pi^{-1}(\{x\}) = 1/\pi(x)$.

1.13. Let $\mu \in \mathbb{M}_+(\mathscr{X})$ and $\varepsilon \in (0,1)$. Show that μ is invariant for P if and only if it is invariant for K_{a_ε}.

1.8 Bibliographical Notes

The concept of a Markov chain first appeared in a series of papers written between 1906 and 1910; see Markov (1910). The term Markov chain was coined by Bernstein (1927) twenty years after this invention. Basharin et al. (2004) contains much interesting information on the early days of Markov chains.

The theory of Markov chains over discrete state spaces was the subject of intense research activity that was triggered by the pioneering work of Doeblin (1938). Most of the theory of discrete-state-space Markov chains was developed in the 1950s and early 1960s. There are many nice monographs summarizing the state of the art in the mid-1960s; see, for example, Chung (1967), Kemeny et al. (1976), Taylor and Karlin (1998). As discrete state-space Markov chains continue to be taught in most applied mathematics courses, books continue to be published regularly on this topic. See, for example, Norris (1998), Brémaud (1999), Privault (2013), Sericola (2013), and Graham (2014). The research monograph Levin et al. (2009) describes the state of the art of research on discrete-state-space Markov chains and in particular the progress that has recently been made in quantifying the speed of convergence.

The theory of Markov chains on general state spaces was initiated in the late 1950s. The books by Orey (1971) and Revuz (1984) (first published in 1975) provide an overview of the early works. The book by Nummelin (1984) lays out the essential foundations of the modern theory. The influence of the book Meyn and Tweedie (1993b) (see also Meyn and Tweedie 2009) on current research in the field of Markov chains and all their applications cannot be overstated.

The theory of Markov chains was also developed by the Russian school, to which we owe major advances; see, for example, the research monographs by Kartashiov (1996) and Borovkov (1998). In particular, Theorem 1.3.6 is established in (Borovkov 1998, Theorem 11.8).

Chapter 2
Examples of Markov Chains

In this chapter we present various examples of Markov chains. We will often use these examples in the sequel to illustrate the results we will develop. Most of our examples are derived from time series models or Monte Carlo simulation methods.

Many time series models belong to the class of random iterative functions that are introduced in Section 2.1. We will establish in this section some properties of these models and in particular will provide conditions under which these models have an invariant probability. In Section 2.2, we introduce the so-called observation-driven models, which have many applications in econometrics in particular.

Finally, Section 2.3 is a short introduction to Markov chain Monte Carlo algorithms, which play a key role today in computational statistics. This section is only a very short overview of a vast domain that remains one of the most active fields of application of Markov chains.

2.1 Random Iterative Functions

Let (X,d) be a complete separable metric space and $(\mathsf{Z},\mathscr{Z})$ a measurable space. We consider X-valued stochastic processes $\{X_k,\ k\in\mathbb{N}\}$ that are defined by the recurrence

$$X_k = f(X_{k-1},Z_k)\ , \qquad k\geq 1\ , \tag{2.1.1}$$

where $f:\mathsf{X}\times\mathsf{Z}\to\mathsf{X}$ is a measurable function, and $\{Z_k,\ k\in\mathbb{N}\}$ is an i.i.d. sequence of random elements defined on a probability space $(\Omega,\mathscr{F},\mathbb{P})$ taking values in $(\mathsf{Z},\mathscr{Z})$, independent of the initial state X_0. Hereinafter, for convenience we will write

$$f_z(x) = f(x,z)\ ,$$

for all $(x,z)\in\mathsf{X}\times\mathsf{Z}$. It is assumed that the map $(z,x)\mapsto f_z(x)$ is measurable with respect to the product σ-field on $\mathscr{Z}\otimes\mathscr{X}$. Let μ be the distribution of Z_0. The process $\{X_k,\ k\in\mathbb{N}\}$ is a Markov chain with Markov kernel P given for $x\in\mathsf{X}$ and $h\in\mathbb{F}_+(\mathsf{X})$ by

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_2

$$Ph(x) = \mathbb{E}\left[h(f_{Z_0}(x))\right] = \int_{\mathsf{Z}} h(f(x,z))\mu(\mathrm{d}z) . \tag{2.1.2}$$

Note that every Markov chain $\{X_k,\ k \in \mathbb{N}\}$ has a representation (2.1.1) when (X, d) is a separable metric space equipped with its Borel σ-field. We will give several classical examples in Section 2.1.1 and prove the existence of a unique invariant distribution in Section 2.1.2.

2.1.1 Examples

Example 2.1.1 (Random walks). Let $\{Z_n,\ n \in \mathbb{N}^*\}$ be a sequence of i.i.d. random variables with values in $\mathsf{X} = \mathbb{R}^d$ and distribution μ. Let X_0 be a random variable in $\mathbb{R}^d$ independent of $\{Z_n,\ n \in \mathbb{N}^*\}$. A random walk with jump or increment distribution μ is a process $\{X_k,\ k \in \mathbb{N}\}$ defined by X_0 and the recurrence

$$X_k = X_{k-1} + Z_k , \qquad k \geq 1 .$$

This model follows the recurrence (2.1.1) with $f(x,z) = x + z$, and thus the process $\{X_k,\ k \in \mathbb{N}\}$ is a Markov chain with kernel given for $x \in \mathbb{R}^d$ and $A \in \mathscr{B}(\mathbb{R}^d)$ by $P(x,A) = \mu(A - x)$; that is, P is entirely determined by the increment distribution μ.

Example 2.1.2 (Autoregressive processes). Let $\{Z_n,\ n \in \mathbb{N}^*\}$ be a sequence of $\mathsf{Z} = \mathbb{R}^q$-valued i.i.d. random vectors and X_0 an $\mathsf{X} = \mathbb{R}^d$-valued random vector independent of $\{Z_n,\ n \in \mathbb{N}\}$. Let F be a $d \times d$ matrix, G a $d \times q$ matrix $(q \leq d)$, and μ a $d \times 1$ vector. The process $\{X_k,\ k \in \mathbb{N}\}$ defined by the recurrence equation

$$X_{n+1} = \mu + FX_n + GZ_{n+1} \tag{2.1.3}$$

is a first-order vector autoregressive process on $\mathbb{R}^d$. This is again an iterative model with $f(x,z) = \mu + Fx + Gz$ and Markov kernel P given for $x \in \mathbb{R}^d$ and $A \in \mathscr{B}(\mathbb{R}^d)$ by

$$P(x,A) = \mathbb{P}(\mu + Fx + GZ_1 \in A) .$$

The AR(1) process can be generalized by assuming that the current value is obtained as an affine combination of the p preceding values of the process and a random disturbance. For simplicity, we assume in the sequel that $d = 1$ and $\mu = 0$. Let $\{Z_k,\ k \in \mathbb{N}\}$ be a sequence of i.i.d. real-valued random variables, let $\phi_1, \ldots, \phi_p$ be real numbers, and let $X_0, X_{-1}, \ldots, X_{-p+1}$ be random variables, independent of the sequence $\{Z_k,\ k \in \mathbb{N}\}$. The scalar AR($p$) process $\{X_k,\ k \in \mathbb{N}\}$ is defined by the recurrence

$$X_k = \phi_1 X_{k-1} + \phi_2 X_{k-2} + \cdots + \phi_p X_{k-p} + Z_k , \qquad k \geq 0 . \tag{2.1.4}$$

The sequence $\{X_k,\ k \in \mathbb{N}\}$ is a Markov chain of order p in the sense of Definition 1.3.7, since the vector process $\mathbf{X}_k = (X_k, X_{k-1}, \ldots, X_{k-p+1})$ is a vector autoregressive process of order 1, defined by the recurrence

$$\mathbf{X}_k = \boldsymbol{\Phi}\mathbf{X}_{k-1} + BZ_k \;, \tag{2.1.5}$$

with

$$\boldsymbol{\Phi} = \begin{pmatrix} \phi_1 & \cdots & \cdots & \phi_p \\ 1 & 0 & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & & 1 & 0 \end{pmatrix} , \qquad B = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} .$$

Thus $\{\mathbf{X}_k,\ k \in \mathbb{N}\}$ is an $\mathbb{R}^p$-valued Markov chain with kernel P defined by

$$P(\mathbf{x}, A) = \mathbb{P}(\boldsymbol{\Phi}\mathbf{x} + BZ_0 \in A) \;, \tag{2.1.6}$$

for $\mathbf{x} \in \mathbb{R}^p$ and $A \in \mathscr{B}(\mathbb{R}^p)$.

Example 2.1.3 (ARMA(p,q)). A generalization of the AR(p) model is obtained by adding a moving average part to the autoregression:

$$X_k = \mu + \alpha_1 X_{k-1} + \cdots + \alpha_p X_{k-p} + Z_k + \beta_1 Z_{k-1} + \cdots + \beta_q Z_{k-q} \;, \tag{2.1.7}$$

where $\{Z_k,\ k \in \mathbb{Z}\}$ is a sequence of i.i.d. random variables with $\mathbb{E}[Z_0] = 0$. This yields a Markov chain of order $r = p \vee q$. Indeed, setting $\alpha_j = 0$ if $j > p$ and $\beta_j = 0$ if $j > q$ yields

$$\begin{pmatrix} X_{k+1} \\ \vdots \\ X_{k+r} \end{pmatrix} = \begin{pmatrix} 0 & 1 & \ldots & \\ \vdots & 0 & 1 & \ldots \\ \vdots & & \ddots & \ddots \\ 0 & \ldots & 0 & 1 \\ \alpha_r & & \ldots & \alpha_1 \end{pmatrix} \begin{pmatrix} X_k \\ \vdots \\ X_{k+r-1} \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \mu + Z_k + \beta_1 Z_{k-1} + \cdots + \beta_r Z_r \end{pmatrix} . \tag{2.1.8}$$

Example 2.1.4 (Functional autoregressive processes). In the AR(1) model, the conditional expectation of the value of the process at time k is an affine function of the previous value: $\mathbb{E}\left[X_k \,|\, \mathscr{F}_{k-1}^X\right] = \mu + FX_{k-1}$. In addition, provided that $\mathbb{E}\left[Z_1 Z_1^T\right] = \mathrm{I}$ in (2.1.4), the conditional variance is almost surely constant, since $\mathbb{E}\left[(X_k - \mathbb{E}\left[X_k \,|\, \mathscr{F}_{k-1}^X\right])(X_k - \mathbb{E}\left[X_k \,|\, \mathscr{F}_{k-1}^X\right])^T \,\middle|\, \mathscr{F}_{k-1}^X\right] = GG^T \ \mathbb{P}-\text{a.s.}$ We say that the model is conditionally homoscedastic. Of course, these assumptions can be relaxed in several directions. We might first consider models that are still conditionally homoscedastic, but for which the conditional expectation of X_k given the past is a nonlinear function of the past observation X_{k-1}, leading to the conditionally homoscedastic autoregressive functional, hereinafter FAR(1), given by

$$X_k = f(X_{k-1}) + GZ_k \;, \tag{2.1.9}$$

where $\{Z_k,\ k \in \mathbb{N}^*\}$ is a sequence of integrable zero-mean i.i.d. random vectors independent of X_0, and $f : \mathbb{R}^d \to \mathbb{R}^d$ is a measurable function. With this definition,

$f(X_{k-1}) = \mathbb{E}\left[X_k \,|\, \mathscr{F}_{k-1}^X\right]$ $\mathbb{P}-$a.s. The kernel of this chain is given, for $x \in \mathbb{R}^d$ and $A \in \mathscr{B}(\mathbb{R}^d)$, by

$$P(x,A) = \mathbb{P}(f(x) + GZ_1 \in A) .$$

Equivalently, for $x \in \mathbb{R}^d$ and $h \in \mathbb{F}_+(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$,

$$Ph(x) = \mathbb{E}\left[h(f(x) + GZ_1)\right] .$$

Compared to the AR(1) model, this model does not easily lend itself to a direct analysis, because expressing the successive iterates of the chain can be very involved.

The recurrence (2.1.9) can be seen as a general discrete-time dynamical model $x_k = f(x_{k-1})$ perturbed by the noise sequence $\{Z_k,\ k \in \mathbb{N}\}$. It is expected that the stability and other properties of the discrete-time dynamical system are related to the stability of (2.1.9).

It is also of interest to consider cases in which the conditional variance

$$\operatorname{Var}\left(X_k \,|\, \mathscr{F}_{k-1}^X\right) = \mathbb{E}\left[(X_k - \mathbb{E}\left[X_k \,|\, \mathscr{F}_{k-1}^X\right])(X_k - \mathbb{E}\left[X_k \,|\, \mathscr{F}_{k-1}^X\right])^T \,\middle|\, \mathscr{F}_{k-1}^X\right]$$

is a function of the past observation X_{k-1}; such models are said to be conditionally heteroscedastic. Heteroscedasticity can be modeled by considering the recurrence

$$X_k = f(X_{k-1}) + g(X_{k-1})Z_k , \tag{2.1.10}$$

where for each $x \in \mathbb{R}^d$, $g(x)$ is a $p \times q$ matrix. Assuming that $\mathbb{E}\left[Z_1 Z_1^T\right] = \mathrm{I}$, the conditional variance is given by $\operatorname{Var}\left(X_k \,|\, \mathscr{F}_{k-1}^X\right) = g(X_{k-1})\{g(X_{k-1})\}^T$ $\mathbb{P}-$a.s. The kernel of this Markov chain is given, for $x \in \mathbb{R}^d$ and $A \in \mathscr{B}(\mathbb{R}^d)$, by

$$P(x,A) = \mathbb{P}(f(x) + g(x)Z_1 \in A) ,$$

or equivalently for $x \in \mathsf{X}$ and $h \in \mathbb{F}_+(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$,

$$Ph(x) = \mathbb{E}\left[h(f(x) + g(x)Z_1)\right] .$$

As above, these models can be generalized by assuming that the conditional expectation $\mathbb{E}\left[X_k \,|\, \mathscr{F}_{k-1}^X\right]$ and the conditional variance $\operatorname{Var}\left(X_k \,|\, \mathscr{F}_{k-1}^X\right)$ are nonlinear functions of the p previous values of the process, $(X_{k-1}, X_{k-2}, \ldots, X_{k-p})$. Assuming again for simplicity that $d = 1$, we may consider the recurrence

$$X_k = f(X_{k-1}, \ldots, X_{k-p}) + \sigma(X_k, \ldots, X_{k-p})Z_k , \tag{2.1.11}$$

where $f : \mathbb{R}^p \to \mathbb{R}$ and $\sigma : \mathbb{R}^p \to \mathbb{R}_+$ are measurable functions.

Example 2.1.5 (ARCH(p)). It is generally acknowledged in the econometrics and applied financial literature that many financial time series such as log-returns of share prices, stock indices, and exchange rates exhibit stochastic volatility and heavy-tailedness. These features cannot be adequately simultaneously modeled by a linear time series model. Nonlinear models were proposed to capture these characteristics. In order for a linear time series model to possess heavy-tailed marginal distributions, it is necessary for the input noise sequence to be heavy-tailed. For nonlinear models, heavy-tailed marginals can be obtained even if the system is injected with a light-tailed input such as with normal noise. We consider here the autoregressive conditional heteroscedastic model of order p, the ARCH(p) model, defined as a solution to the recurrence

$$X_k = \sigma_k Z_k \tag{2.1.12a}$$

$$\sigma_k^2 = \alpha_0 + \alpha_1 X_{k-1}^2 + \cdots + \alpha_p X_{k-p}^2 , \tag{2.1.12b}$$

where the coefficients $\alpha_j \geq 0$, $j \in \{0, \ldots, p\}$, are nonnegative, and $\{Z_k,\ k \in \mathbb{Z}\}$ is a sequence of i.i.d. random variable with zero mean (often assumed to be standard Gaussian). The ARCH(p) process is a Markov chain of order p. Assume that Z_1 has a density g with respect to Lebesgue measure on $\mathbb{R}$. Then for $h \in \mathbb{F}_+(\mathbb{R}^p, \mathscr{B}(\mathbb{R}^p))$, we get

$$\begin{aligned} &Ph(x_1, \ldots, x_p) \\ &\quad = \mathbb{E}\left[h\left(\sqrt{\alpha_0 + \alpha_1 x_1^2 + \cdots + \alpha_p x_p^2}\, Z_1\right) \right] \\ &\quad = \int h(y) \frac{1}{\sqrt{\alpha_0 + \alpha_1 x_1^2 + \cdots + \alpha_p x_p^2}} g\left(\frac{y}{\sqrt{\alpha_0 + \alpha_1 x_1^2 + \cdots + \alpha_p x_p^2}} \right) \mathrm{d}y \, . \end{aligned}$$

The kernel therefore has a density with respect to Lebesgue measure given by

$$p(x_1, \ldots, x_p; y) = \frac{1}{\sqrt{\alpha_0 + \alpha_1 x_1^2 + \cdots + \alpha_p x_p^2}} g\left(\frac{y}{\sqrt{\alpha_0 + \alpha_1 x_1^2 + \cdots + \alpha_p x_p^2}} \right) .$$

We will latter see that it is relatively easy to discuss the properties of this model, which is used widely in financial econometrics.

Example 2.1.6 (Self-exciting threshold AR model). Self-exciting threshold AR (SETAR) models have been widely employed as a model for nonlinear time series. Threshold models are piecewise linear AR models for which the linear relationship varies according to delayed values of the process (hence the term self-exciting). In this class of models, different autoregressive processes may operate, and the change between the various AR processes is governed by threshold values and a time lag. An ℓ-regime TAR model has the form

$$X_k = \begin{cases} \phi_0^{(1)} + \sum_{i=1}^{p_1} \phi_i^{(1)} X_{k-i} + \sigma^{(1)} Z_k^{(1)} & \text{if } X_{k-d} \leq r_1 \, , \\ \phi_0^{(2)} + \sum_{i=1}^{p_2} \phi_i^{(2)} X_{k-i} + \sigma^{(2)} Z_k^{(2)} & \text{if } r_1 < X_{k-d} \leq r_2 \, , \\ \quad \vdots & \quad \vdots \\ \phi_0^{(\ell)} + \sum_{i=1}^{p_\ell} \phi_i^{(\ell)} X_{k-i} + \sigma^{(\ell)} Z_k^{(\ell)} & \text{if } r_{\ell-1} < X_{k-d} \, , \end{cases} \tag{2.1.13}$$

where $\{Z_k,\ k \in \mathbb{N}\}$ is an i.i.d. sequence of real-valued random variables, the positive integer d is a specified delay, and $-\infty < r_1 < \cdots < r_{\ell-1} < \infty$ is a partition of $\mathsf{X} = \mathbb{R}$. These models allow for changes in the AR coefficients over time, and these changes are determined by comparing previous values (back-shifted by a time lag equal to d) to fixed threshold values. Each different AR model is referred to as a regime. In the definition above, the values p_j of the order of AR models can differ in each regime, although in many applications, they are assumed to be equal.

The model can be generalized to include the possibility that the regimes depend on a collection of past values of the process, or that the regimes depend on an exogenous variable (in which case the model is not self-exciting).

The popularity of TAR models is due to their being relatively simple to specify, estimate, and interpret compared to many other nonlinear time series models. In addition, despite its apparent simplicity, the class of TAR models can reproduce many nonlinear phenomena such as stable and unstable limit cycles, jump resonance, harmonic distortion, modulation effects, chaos, and so on.

Example 2.1.7 (Random-coefficient autoregressive models). A process closely related to the AR(1) process is the random-coefficient autoregressive (RCA) process

$$X_k = A_k X_{k-1} + B_k \,, \tag{2.1.14}$$

where $\{(A_k, B_k),\ k \in \mathbb{N}^*\}$ is a sequence of i.i.d. random elements in $\mathbb{R}^{d\times d} \times \mathbb{R}^d$, independent of X_0. The Markov kernel P of this chain is defined by

$$Ph(x) = \mathbb{E}\left[h(A_1 x + B_1)\right] \,, \tag{2.1.15}$$

for $x \in \mathsf{X}$ and $h \in \mathbb{F}_+(\mathsf{X})$.

For instance, the volatility sequence $\{\sigma_k,\ k \in \mathbb{N}\}$ of the ARCH(1) process of Example 2.1.5 fits into the framework of (2.1.14) with $A_k = \alpha_1 Z_{k-1}^2$ and $B_k = \alpha_0$.

2.1.2 Invariant Distributions

The iterative representation $X_k = f(X_{k-1}, Z_k)$ is useful if the function $x \to f_z(x)$ has certain structural properties. We provide now conditions that ensure that the chain $\{X_k,\ k \in \mathbb{N}\}$ has a unique invariant distribution.

H 2.1.8 • *There exists a measurable function $K : \mathsf{Z} \to \mathbb{R}_+$ such that for all $(x,y,z) \in \mathsf{X} \times \mathsf{X} \times \mathsf{Z}$,*

$$d(f_z(x), f_z(y)) \le K(z) d(x,y) \,, \tag{2.1.16}$$

$$\mathbb{E}\left[\log^+ K(Z_1)\right] < \infty \,, \quad \mathbb{E}\left[\log K(Z_1)\right] < 0 \,. \tag{2.1.17}$$

• *There exists $x_0 \in \mathsf{X}$ such that*

$$\mathbb{E}\left[\log^+ d(x_0, f(x_0, Z_1))\right] < \infty . \tag{2.1.18}$$

It is easily seen that if (2.1.18) holds for some $x_0 \in \mathsf{X}$, then it holds for all $x_0' \in \mathsf{X}$. Indeed, for all $(x_0, x_0', z) \in \mathsf{X} \times \mathsf{X} \times \mathsf{Z}$, (2.1.16) implies

$$d(x_0', f_z(x_0')) \leq (1 + K(z)) d(x_0, x_0') + d(x_0, f_z(x_0)) .$$

Using the inequality $\log^+(x+y) \leq \log^+(x+1) + \log^+(y)$, the previous inequality yields

$$\log^+ d(x_0', f_z(x_0')) \leq \log^+(1 + K(z)) + \log^+\{1 + d(x_0, x_0')\} + \log^+ d(x_0, f_z(x_0) .$$

Taking expectations, we obtain that (2.1.18) holds for x_0'.

For $x \in \mathsf{X}$, define the forward chain $\{X_n^x,\ n \in \mathbb{N}\}$ and the backward process $\{Y_n^x,\ n \in \mathbb{N}\}$ starting from $X^x = Y_0^x = x$ by

$$X_k^x = f_{Z_k} \circ \cdots \circ f_{Z_1}(x_0) , \tag{2.1.19}$$
$$Y_k^x = f_{Z_1} \circ \cdots \circ f_{Z_k}(x_0) . \tag{2.1.20}$$

Since $\{Z_k,\ k \in \mathbb{N}\}$ is an i.i.d. sequence, Y_k^x has the same distribution as X_k^x for each $k \in \mathbb{N}$.

Theorem 2.1.9. *For every $x_0 \in \mathsf{X}$, the sequence $\{Y_k^{x_0},\ k \in \mathbb{N}\}$ converges almost surely to a $\mathbb{P}-$a.s. finite random variable Y_∞ that does not depend on x_0 and whose distribution is the unique invariant distribution of the kernel P defined in* (2.1.2).

Proof. For $x_0, x \in \mathsf{X}$, the Lipschitz condition (2.1.16) yields

$$\begin{aligned} d(Y_k^{x_0}, Y_k^x) &= d(f_{Z_1}[f_{Z_2} \circ \cdots \circ f_{Z_k}(x_0)], f_{Z_1}[f_{Z_2} \circ \cdots \circ f_{Z_k}(x)]) \\ &\leq K(Z_1) d(f_{Z_2} \circ \cdots \circ f_{Z_k}(x_0), f_{Z_2} \circ \cdots \circ f_{Z_k}(x)) . \end{aligned}$$

By induction, this yields

$$d(Y_k^{x_0}, Y_k^x) \leq d(x_0, x) \prod_{i=1}^k K(Z_i) . \tag{2.1.21}$$

Set $x = f_{Z_{k+1}}(x_0)$. Then $Y_{k+1}^{x_0} = Y_k^x$, and applying (2.1.21) yields

$$d(Y_k^{x_0}, Y_{k+1}^{x_0}) \leq d(x_0, f_{Z_{k+1}}(x_0)) \prod_{i=1}^k K(Z_i) . \tag{2.1.22}$$

Since we have assumed that $\mathbb{E}[\log K(Z_0)] \in [-\infty, 0)$, the strong law of large numbers yields

$$\limsup_{k\to\infty} k^{-1} \sum_{i=1}^{k} \log K(Z_i) < 0 \qquad \mathbb{P}-\text{a.s.}$$

Exponentiating, this yields

$$\limsup_{k\to\infty} \left\{ \prod_{i=1}^{k} K(Z_i) \right\}^{1/k} < 1 \qquad \mathbb{P}-\text{a.s.} \tag{2.1.23}$$

Applying the assumption (2.1.18), we obtain, for every $\delta > 0$,

$$\sum_{k=1}^{\infty} \mathbb{P}(k^{-1} \log^+ d(x_0, f_{Z_k}(x_0)) > \delta) \le \delta^{-1} \mathbb{E}\left[\log^+ d(x_0, f_{Z_0}(x_0))\right] < \infty .$$

Applying the Borel–Cantelli lemma yields $\lim_{k\to\infty} k^{-1} \log^+ d(x_0, f_{Z_k}(x_0)) = 0$ $\mathbb{P}-$a.s. and consequently $\limsup_{k\to\infty} k^{-1} \log d(x_0, f_{Z_k}(x_0)) \le 0$ $\mathbb{P}-$a.s., or equivalently

$$\limsup_{k\to\infty} \left\{ d(x_0, f_{Z_k}(x_0)) \right\}^{1/k} \le 1 \qquad \mathbb{P}-\text{a.s.} \tag{2.1.24}$$

Applying (2.1.23) and (2.1.24) and the Cauchy root test to (2.1.22) proves that the series $\sum_{k=1}^{\infty} d(Y_k^{x_0}, Y_{k+1}^{x_0})$ is almost surely convergent. Since (X, d) is complete, this in turn implies that $\{Y_k^{x_0},\ k \in \mathbb{N}\}$ is almost surely convergent to a $\mathbb{P}-$a.s. finite random variable, which we denote by $Y_\infty^{x_0}$. The bound (2.1.23) also implies that $\lim_{k\to\infty} \prod_{i=1}^{k} K(Z_i) = 0$ $\mathbb{P}-$a.s. This and (2.1.21) imply that $\lim_{k\to\infty} d(Y_k^{x_0}, Y_k^{x}) = 0$ $\mathbb{P}-$a.s. for all $x_0, x \in \mathsf{X}$, so that the distribution of Y_∞^x, say π, does not depend on x.

Since f_{Z_0} is continuous by (2.1.16), we have

$$f(Y_\infty^x, Z_0) = \lim_{k\to\infty} f(Y_k^x, Z_0) \stackrel{\text{law}}{=} \lim_{k\to\infty} Y_{k+1}^x = Y_\infty^x .$$

This proves that if the distribution of X_0 is π, then the distribution of X_1 is also π, that is, π is P-invariant.

We now prove that the distribution of Y_∞ is the unique invariant probability measure. Since X_k^x and Y_k^x have the same distribution for all $k \in \mathbb{N}$, for every bounded continuous function g we have

$$\lim_{k\to\infty} \mathbb{E}[g(X_k^x)] = \lim_{k\to\infty} \mathbb{E}[g(Y_k^x)] = \mathbb{E}[g(Y_\infty^x)] = \pi(g) .$$

This shows that $P^n(x, \cdot)$ converges weakly to π for all $x \in \mathsf{X}$. If ξ is an invariant measure, then for every bounded continuous function g, by Lebesgue's dominated convergence theorem, we have

$$\lim_{n\to\infty} \xi P^n(g) = \int_{\mathsf{X}} \lim_{n\to\infty} P^n g(x) \xi(\mathrm{d}x) = \int_{\mathsf{X}} \lim_{n\to\infty} \pi(g) \xi(\mathrm{d}x) = \pi(g) \; .$$

This proves that $\xi = \pi$. □

The use of weak convergence in metric spaces to obtain existence and uniqueness of invariant measures will be formalized and further developed in Chapter 12.

2.2 Observation-Driven Models

Definition 2.2.1 (Observation-driven model) *Let $(\mathsf{X}, \mathscr{X})$ and $(\mathsf{Y}, \mathscr{Y})$ be measurable spaces, Q a Markov kernel on $\mathsf{X} \times \mathscr{Y}$, and $f : \mathsf{X} \times \mathsf{Y} \to \mathsf{X}$ a measurable function. Let $(\Omega, \mathscr{F}, \{\mathscr{F}_k, k \in \mathbb{N}\}, \mathbb{P})$ be a filtered probability space. An observation-driven stochastic process $\{(X_k, Y_k, \mathscr{F}_k), k \in \mathbb{N}\}$ is an adapted process taking values in $\mathsf{X} \times \mathsf{Y}$ such that for all $k \in \mathbb{N}^*$ and all $A \in \mathscr{Y}$,*

$$\mathbb{P}(Y_k \in A \mid \mathscr{F}_{k-1}) = Q(X_{k-1}, A) \; , \tag{2.2.1a}$$

$$X_k = f(X_{k-1}, Y_k) \; . \tag{2.2.1b}$$

$$\begin{tikzcd}
{} \arrow[r] & X_{k-1} \arrow[r] \arrow[dr] & X_k \\
 & Y_{k-1} \arrow[u] & Y_k \arrow[u]
\end{tikzcd}$$

Fig. 2.1 Dependency graph of $\{(X_k, Y_k), k \in \mathbb{N}\}$.

The process $\{(X_k, Y_k), k \in \mathbb{N}\}$ is a Markov chain with kernel P characterized by

$$P((x,y), A \times B) = \int_B \mathbb{1}_A(f(x,z)) Q(x, \mathrm{d}z) \; , \tag{2.2.2}$$

for all $(x,y) \in \mathsf{X} \times \mathsf{Y}$, $A \in \mathscr{X}$, and $B \in \mathscr{Y}$. Note that Y_k is independent of Y_{k-1} conditionally on X_{k-1}, but $\{Y_k\}$ may not be a Markov chain. The sequence $\{X_k\}$ is a Markov chain with kernel P_1 defined for $x \in \mathsf{X}$ and $A \in \mathscr{X}$ by

$$P_1(x, A) = \int_{\mathsf{Y}} \mathbb{1}_A(f(x,z)) Q(x, \mathrm{d}z) \; . \tag{2.2.3}$$

We can express X_k as a function of the sequence $\{Y_1,\dots,Y_k\}$ and X_0. Writing $f_y(x)$ for $f(x,y)$ and $f_{y_2}\circ f_{y_1}(x)$ for $f(f(x,y_1),y_2)$, we have

$$X_k = f_{Y_k}\circ\cdots\circ f_{Y_1}(X_0)\,. \tag{2.2.4}$$

The name "observation-driven model" comes from the fact that in statistical applications, only the sequence $\{Y_k\}$ is observable. We know by Theorem 1.3.6 that every Markov chain can be represented in this way, with $\{Y_k\}$ i.i.d. random variables, independent of X_0. However, the latter representation may fail to be useful, in contrast to the more structured representation of Definition 2.2.1.

Example 2.2.2 (GARCH (p,q) model). The limitation of the ARCH model is that the squared process has the autocorrelation structure of an autoregressive process, which does not always fit the data. A generalization of the ARCH model is obtained by allowing the conditional variance to depend on the lagged squared returns $(X_{t-1}^2,\dots,X_{t-p}^2)$ and on the lagged conditional variances. This model is called the generalized autoregressive conditional heteroscedastic (GARCH) model, defined by the recurrence

$$X_k = \sigma_k Z_k \tag{2.2.5a}$$

$$\sigma_k^2 = \alpha_0 + \alpha_1 X_{k-1}^2 + \cdots + \alpha_p X_{k-p}^2 + \beta_1\sigma_{k-1}^2 + \cdots + \beta_q\sigma_{k-q}^2\,, \tag{2.2.5b}$$

where the coefficients $\alpha_0,\dots,\alpha_p,\beta_1,\dots,\beta_q$ are nonnegative and $\{Z_k,\ k\in\mathbb{Z}\}$ is a sequence of i.i.d. random variables with $\mathbb{E}[Z_0]=1$.

The GARCH(p,q) process is a Markov chain of order $r=p\vee q$. Indeed, setting $\alpha_j=0$ if $j>p$ and $\beta_j=0$ if $j>q$ yields

$$\begin{pmatrix}\sigma_{k+1}^2\\ \vdots\\ \sigma_{k+r}^2\end{pmatrix} = \begin{pmatrix} 0 & 1 & \dots & & \\ \vdots & 0 & 1 & & \dots \\ \vdots & & \ddots & & \ddots \\ 0 & \dots & 0 & & 1 \\ \alpha_r Z_k^2+\beta_r & & \dots & & \alpha_1 Z_{k+r-1}^2+\beta_1\end{pmatrix}\begin{pmatrix}\sigma_k^2\\ \vdots\\ \sigma_{k+r-1}^2\end{pmatrix} + \begin{pmatrix}0\\ \vdots\\ \alpha_0\end{pmatrix}\,. \tag{2.2.6}$$

These models do not allow for dependence between the volatility and the sign of the returns, since the volatility depends only on the squared returns. This property, which is often observed in financial time series, is the so-called leverage effect. To accommodate this effect, several modifications of the GARCH model have been considered. We give two such examples.

Example 2.2.3 (EGARCH). The EGARCH(p,q) models the log-volatility as an ARMA process that is not independent of the innovation of the returns. More precisely, it is defined by the recurrence

$$X_k = \sigma_k Z_k \;, \tag{2.2.7a}$$

$$\log \sigma_k^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j \eta_{k-j} + \sum_{j=1}^{q} \beta_j \log \sigma_{k-j}^2 \;, \tag{2.2.7b}$$

where $\{(Z_n, \eta_n,), n \in \mathbb{N}\}$ is a sequence of i.i.d. bivariate random vectors with possibly dependent components. The original specification of the sequence $\{\eta_n\}$ is $\eta_k = \theta Z_k + \lambda(|Z_k| - \mathbb{E}[|Z_0|])$.

Example 2.2.4 (TGARCH). The TGARCH models the volatility as a threshold ARMA process whereby the coefficient of the autoregressive part depends on the sign of the innovation. More precisely, it is defined by the recurrence

$$X_k = \sigma_k Z_k \;, \tag{2.2.8a}$$

$$\sigma_k^2 = \alpha_0 + \alpha X_{k-1}^2 + \phi X_{k-1}^2 \mathbb{1}_{Z_{k-1}>0} + \beta \sigma_{k-1}^2 \;. \tag{2.2.8b}$$

Building an integer-valued model with rich dynamics is not easy. Observation-driven models provide a convenient possibility.

Example 2.2.5 (Log-Poisson autoregression). Let $\{\mathscr{F}_k,\ k \in \mathbb{N}\}$ be a filtration and define an adapted sequence $\{(X_k, Y_k),\ k \in \mathbb{N}\}$ by

$$\mathscr{L}(Y_k|\mathscr{F}_{k-1}) = \text{Poisson}(\exp(X_{k-1})), \tag{2.2.9a}$$

$$X_k = \omega + bX_{k-1} + c\log(1+Y_k) \;, \qquad k \geq 1 \;, \tag{2.2.9b}$$

where ω, b, c are real-valued parameters. This process is of the form (2.2.1) with f defined on $\mathbb{R} \times \mathbb{N}$ and Q on $\mathbb{R} \times \mathscr{P}(\mathbb{N})$ by

$$f(x,z) = \omega + bx + c\log(1+z) \;,$$

$$Q(x,A) = \mathrm{e}^{-\mathrm{e}^x} \sum_{j \in A} \frac{\mathrm{e}^{jx}}{j!} \;,$$

for $x \in \mathbb{R}$, $z \in \mathbb{N}$, and $A \subset \mathbb{N}$.

The log-intensity X_k can be expressed as in (2.2.4) in terms of the lagged responses by expanding (2.2.9b):

$$X_k = \omega \frac{1-b^k}{1-b} + b^k X_0 + c\sum_{i=0}^{k-1} b^i \log(1+Y_{k-i-1}) \;.$$

This model can also be represented as a functional autoregressive model with an i.i.d. innovation. Let $\{N_k,\ k \in \mathbb{N}^*\}$ be a sequence of independent unit-rate homogeneous Poisson process on the real line, independent of X_0. Then $\{X_n,\ n \in \mathbb{N}\}$ may be expressed as $X_k = F(X_{k-1}, N_k)$, where F is the function defined on $\mathbb{R} \times \mathbb{N}^{\mathbb{R}}$ by

$$F(x,N) = \omega + bx + c\log\{1 + N(\mathrm{e}^x)\} \;. \tag{2.2.10}$$

The transition kernel P of the Markov chain $\{X_k,\ k \in \mathbb{N}\}$ can be expressed as

$$Ph(x) = \mathbb{E}\left[h(\omega + b + c\log\{1 + N(\mathrm{e}^x)\})\right] , \tag{2.2.11}$$

for all bounded measurable functions h, where N is a homogeneous Poisson process. We will investigate this model thoroughly in later chapters. We will see that the representation (2.2.10) sometimes does not yield optimal conditions for the existence and stability of the process, and the observation-driven model representation (2.2.9) will be useful.

2.3 Markov Chain Monte Carlo Algorithms

Markov chain Monte Carlo is a general method for the simulation of distributions known up to a multiplicative constant. Let ν be a σ-finite measure on a state space $(\mathsf{X}, \mathscr{X})$ and let $h_\pi \in \mathbb{F}_+(\mathsf{X})$ be such that $0 < \int_{\mathsf{X}} h_\pi(x)\nu(\mathrm{d}x) < \infty$. Typically, X is an open subset of $\mathbb{R}^d$ and ν is Lebesgue measure, or X is countable and ν is the counting measure. This function is associated with a probability measure π on X defined by

$$\pi(A) = \frac{\int_A h_\pi(x)\nu(\mathrm{d}x)}{\int_{\mathsf{X}} h_\pi(x)\nu(\mathrm{d}x)} . \tag{2.3.1}$$

We want to approximate expectations of functions $f \in \mathbb{F}_+(\mathsf{X})$ with respect to π,

$$\pi(f) = \frac{\int_{\mathsf{X}} f(x)h_\pi(x)\nu(\mathrm{d}x)}{\int_{\mathsf{X}} h_\pi(x)\nu(\mathrm{d}x)} .$$

If the state space X is high-dimensional and h_π is complex, then direct numerical integration is not an option. The classical Monte Carlo solution to this problem is to simulate i.i.d. random variables $Z_0, Z_1, \ldots, Z_{n-1}$ with distribution π and then to estimate $\pi(f)$ by the sample mean

$$\hat{\pi}(f) = n^{-1}\sum_{i=0}^{n-1} f(Z_i) . \tag{2.3.2}$$

This gives an unbiased estimate with standard deviation of order $O(n^{-1/2})$, provided that $\pi(f^2) < \infty$. Furthermore, by the central limit theorem, the normalized error $\sqrt{n}(\hat{\pi}(f) - \pi(f))$ has a limiting normal distribution, so that confidence intervals are easily obtained.

The problem often encountered in applications is that it might be very difficult to simulate i.i.d. random variables with distribution π. Instead, the Markov chain Monte Carlo (MCMC) solution is to construct a Markov chain on X that has π as invariant probability. The hope is that regardless of the initial distribution ξ, the law of large numbers will hold, i.e., $\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1} f(X_k) = \pi(f)\ \ \mathbb{P}_\xi - \text{a.s.}$ We will investigate the law of large numbers for Markov chains in Chapter 5 and subsequent chapters.

At first sight, it may seem even more difficult to find such a Markov chain than to estimate $\pi(f)$ directly. In the following subsections, we will exhibit several such constructions.

2.3.1 Metropolis–Hastings Algorithms

Let Q be a Markov kernel having density q with respect to ν i.e., $Q(x,A) = \int_A q(x,y)\nu(\mathrm{d}y)$ for every $x \in \mathsf{X}$ and $A \in \mathscr{X}$.

The Metropolis–Hastings algorithm proceeds in the following way. An initial starting value X_0 is chosen. Given X_k, a candidate move Y_{k+1} is sampled from $Q(X_k,\cdot)$. With probability $\alpha(X_k,Y_{k+1})$, it is accepted, and the chain moves to $X_{k+1} = Y_{k+1}$. Otherwise, the move is rejected, and the chain remains at $X_{k+1} = X_k$. The probability $\alpha(X_k,Y_{k+1})$ of accepting the move is given by

$$\alpha(x,y) = \begin{cases} \min\left(\frac{h_\pi(y)}{h_\pi(x)}\frac{q(y,x)}{q(x,y)}, 1\right) & \text{if } h_\pi(x)q(x,y) > 0\,, \\ 1 & \text{if } h_\pi(x)q(x,y) = 0\,. \end{cases} \tag{2.3.3}$$

The acceptance probability $\alpha(x,y)$ depends only on the ratio $h_\pi(y)/h_\pi(x)$; therefore, we need to know h_π only up to a normalizing constant. In Bayesian inference, this property plays a crucial role.

This procedure produces a Markov chain, $\{X_k,\ k \in \mathbb{N}\}$, with Markov kernel P given by

$$P(x,A) = \int_A \alpha(x,y)q(x,y)\nu(\mathrm{d}y) + \bar{\alpha}(x)\delta_x(A)\,, \tag{2.3.4}$$

with

$$\bar{\alpha}(x) = \int_{\mathsf{X}} \{1 - \alpha(x,y)\}q(x,y)\nu(\mathrm{d}y)\,. \tag{2.3.5}$$

The quantity $\bar{\alpha}(x)$ is the probability of remaining at the same point.

Proposition 2.3.1 *The distribution π is reversible with respect to the Metropolis–Hastings kernel P.*

Proof. Note first that for every $x,y \in \mathsf{X}$, one has

$$\begin{aligned} h_\pi(x)\alpha(x,y)q(x,y) &= \{h_\pi(x)q(x,y)\} \wedge \{h_\pi(y)q(y,x)\} \\ &= h_\pi(y)\alpha(y,x)q(y,x)\,. \end{aligned} \tag{2.3.6}$$

Thus for $C \in \mathscr{X} \times \mathscr{X}$,

$$\iint h_\pi(x)\alpha(x,y)q(x,y)\mathbb{1}_C(x,y)\nu(\mathrm{d}x)\nu(\mathrm{d}y)$$
$$= \iint h_\pi(y)\alpha(y,x)q(y,x)\mathbb{1}_C(x,y)\nu(\mathrm{d}x)\nu(\mathrm{d}y) \,. \quad (2.3.7)$$

On the other hand,

$$\begin{aligned}
&\iint h_\pi(x)\delta_x(\mathrm{d}y)\bar{\alpha}(x)\mathbb{1}_C(x,y)\nu(\mathrm{d}x) \\
&\quad = \int h_\pi(x)\bar{\alpha}(x)\mathbb{1}_C(x,x)\nu(\mathrm{d}x) = \int h_\pi(y)\bar{\alpha}(y)\mathbb{1}_C(y,y)\nu(\mathrm{d}y) \\
&\quad = \iint h_\pi(y)\delta_y(\mathrm{d}x)\bar{\alpha}(y)\mathbb{1}_C(x,y)\nu(\mathrm{d}y) \,. \qquad (2.3.8)
\end{aligned}$$

Hence, summing (2.3.7) and (2.3.8), we obtain

$$\iint h_\pi(x)P(x,\mathrm{d}y)\nu(\mathrm{d}x)\mathbb{1}_C(x,y) = \iint h_\pi(y)P(y,\mathrm{d}x)\mathbb{1}_C(x,y)\nu(\mathrm{d}y) \,.$$

This proves that π is reversible with respect to P. □

From Proposition 1.5.2, we obtain that π is an invariant probability for the Markov kernel P.

Example 2.3.2 (Random walk Metropolis algorithm). This is a particular case of the Metropolis–Hastings algorithm, where the proposal transition density is symmetric, i.e., $q(x,y) = q(y,x)$, for every $(x,y) \in \mathsf{X} \times \mathsf{X}$. Furthermore, assume that $\mathsf{X} = \mathbb{R}^d$ and let $\bar{q}$ be a symmetric density with respect to 0, i.e., $\bar{q}(-y) = \bar{q}(y)$ for all $y \in \mathsf{X}$. Consider the transition density q defined by $q(x,y) = \bar{q}(y-x)$. This means that if the current state is X_k, an increment Z_{k+1} is drawn from $\bar{q}$, and the candidate $Y_{k+1} = X_k + Z_{k+1}$ is proposed.

The acceptance probability (2.3.3) for the random walk Metropolis algorithm is given by

$$\alpha(x,y) = 1 \wedge \frac{h_\pi(y)}{h_\pi(x)} \,. \qquad (2.3.9)$$

If $h_\pi(Y_{k+1}) \geq h_\pi(X_k)$, then the move is accepted with probability one, and if $h_\pi(Y_{k+1}) < h_\pi(X_k)$, then the move is accepted with a probability strictly less than one.

The choice of the incremental distribution is crucial for the efficiency of the algorithm. A classical choice for $\bar{q}$ is the multivariate normal distribution with zero mean and covariance matrix Γ to be suitably chosen.

Example 2.3.3 (Independent Metropolis–Hastings sampler). Another possibility is to set the transition density to be $q(x,y) = \bar{q}(y)$, where $\bar{q}$ is a density on X. In this case, the next candidate is drawn independently of the current state of the chain. This yields the so-called independent sampler, which is closely related to the accept–reject algorithm for random variable simulation.

The acceptance probability (2.3.3) is given by

$$\alpha(x,\ y) = 1 \wedge \frac{h_\pi(y)\bar{q}(x)}{h_\pi(x)\bar{q}(y)} \ . \tag{2.3.10}$$

Candidate steps with a low weight $\bar{q}(Y_{k+1})/\pi(Y_{k+1})$ are rarely accepted, whereas candidates with a high weight are very likely to be accepted. Therefore, the chain will remain at these states for several steps with a high probability, thus increasing the importance of these states within the constructed sample.

Assume, for example, that h is the standard Gaussian density and that q is the density of the Gaussian distribution with zero mean and variance σ^2, so that $q(x) = h(y/\sigma)/\sigma$. Assume that $\sigma^2 > 1$, so that the values being proposed are sampled from a distribution with heavier tails than the objective distribution h. Then the acceptance probability is

$$\alpha(x,y) = \begin{cases} 1 & |y| \le |x| \ , \\ \exp(-(y^2 - x^2)(1-\sigma^{-2})/2) & |y| > |x| \ . \end{cases}$$

Thus the algorithm accepts all moves that decrease the norm of the current state but only some of those that increase it.

If $\sigma^2 < 1$, the values being proposed are sampled from a lighter-tailed distribution than h, and the acceptance probability becomes

$$\alpha(x,y) = \begin{cases} \exp(-(x^2 - y^2)(1-\sigma^{-2})/2) & |y| \le |x| \ , \\ 1 & |y| > |x| \ . \end{cases}$$

It is natural to inquire whether heavy-tailed or light-tailed proposal distributions should be preferred. This question will be partially answered in Example 15.3.3.

Example 2.3.4 (Langevin diffusion). More sophisticated proposals can be considered. Assume that $x \mapsto \log h_\pi(x)$ is everywhere differentiable. Consider the Langevin diffusion defined by the stochastic differential equation (SDE)

$$\mathrm{d}X_t = \frac{1}{2}\nabla \log h_\pi(X_t)\mathrm{d}t + \mathrm{d}W_t \ ,$$

where $\nabla \log h_\pi$ denotes the gradient of $\log h_\pi$. Under appropriate conditions, the Langevin diffusion has a stationary distribution with density h_π and is reversible. Assume that the proposal state Y_{k+1} in the Metropolis–Hastings algorithm corresponds to the Euler discretization of the Langevin SDE for some step size h:

$$Y_{k+1} = X_k + \frac{\gamma}{2}\nabla \log h_\pi(X_k) + \sqrt{\gamma} Z_{k+1} \ , \qquad Z_{k+1} \sim \mathrm{N}(0,\mathrm{I}) \ .$$

Such algorithms are known as Langevin Metropolis–Hastings algorithms. The gradient can be approximated numerically via finite differences and does not require knowledge of the normalizing constant of the target distribution h_π.

2.3.2 Data Augmentation

Throughout this section, $(\mathsf{X}, \mathscr{X})$ and $(\mathsf{Y}, \mathscr{Y})$ are Polish spaces equipped with their Borel σ-fields. Again, we wish to simulate from a probability measure π defined on $(\mathsf{X}, \mathscr{X})$ using a sequence $\{X_k,\ k \in \mathbb{N}\}$ of X-valued random variables. Data augmentation algorithms consist in writing the target distribution π as the marginal of the distribution π^* on the product space $(\mathsf{X} \times \mathsf{Y}, \mathscr{X} \otimes \mathscr{Y})$ defined by $\pi^* = \pi \otimes R$, where R is a kernel on $\mathsf{X} \times \mathscr{Y}$. By Theorem B.3.11, there exists also a kernel S on $\mathsf{Y} \times \mathscr{X}$ and a probability measure $\tilde{\pi}$ on $(\mathsf{Y}, \mathscr{Y})$ such that $\pi^*(C) = \iint \mathbb{1}_C(x,y)\tilde{\pi}(\mathrm{d}y)S(y, \mathrm{d}x)$ for $C \in \mathscr{X} \otimes \mathscr{Y}$. In other words, if (X,Y) is a pair of random variables with distribution π^*, then $R(x, \cdot)$ is the distribution of Y conditionally on $X = x$, and $S(y, \cdot)$ is the distribution of X conditionally on $Y = y$. The bivariate distribution π^* can then be expressed as follows:

$$\pi^*(\mathrm{d}x\mathrm{d}y) = \pi(\mathrm{d}x)R(x, \mathrm{d}y) = S(y, \mathrm{d}x)\tilde{\pi}(\mathrm{d}y) . \tag{2.3.11}$$

A data augmentation algorithm consists in running a Markov chain $\{(X_k, Y_k),\ k \in \mathbb{N}\}$ with invariant probability π^* and to use $n^{-1}\sum_{k=0}^{n-1} f(X_k)$ as an approximation of $\pi(f)$. A significant difference between this general approach and a Metropolis–Hastings algorithm associated with the target distribution π is that $\{X_k,\ k \in \mathbb{N}\}$ is no longer constrained to be a Markov chain. The transition from (X_k, Y_k) to (X_{k+1}, Y_{k+1}) is decomposed into two successive steps: Y_{k+1} is first drawn given (X_k, Y_k), and then X_{k+1} is drawn given (X_k, Y_{k+1}). Intuitively, Y_{k+1} can be used as an auxiliary variable, which directs the moves of X_k toward interesting regions with respect to the target distribution.

When sampling from R and S is feasible, a classical choice consists in following the two successive steps: given (X_k, Y_k),

(i) sample Y_{k+1} from $R(X_k, \cdot)$,
(ii) sample X_{k+1} from $S(Y_{k+1}, \cdot)$.

$$\begin{tikzcd}
X_k \arrow[dr, "R"] & X_{k+1} \\
Y_k \arrow[u, "S"] & Y_{k+1} \arrow[u, "S"']
\end{tikzcd}$$

Fig. 2.2 In this example, sampling from R and S is feasible.

It turns out that $\{X_k,\ k \in \mathbb{N}\}$ is a Markov chain with Markov kernel RS, and π is reversible with respect to RS.

Lemma 2.3.5 *The distribution π is reversible with respect to the kernel RS.*

Proof. By Definition 1.5.1, we must prove that the measure $\pi \otimes RS$ on X^2 is symmetric. For $A, B \in \mathscr{X}$, applying (2.3.11), we have

$$\begin{aligned}
&\pi \otimes RS(A \times B) \\
&= \int_{\mathsf{X}\times\mathsf{Y}} \pi(\mathrm{d}x) R(x, \mathrm{d}y) \mathbb{1}_A(x) S(y, B) = \int_{\mathsf{X}\times\mathsf{Y}} \mathbb{1}_A(x) S(y, B) \pi^*(\mathrm{d}x\mathrm{d}y) \\
&= \int_{\mathsf{X}\times\mathsf{Y}} \mathbb{1}_A(x) S(y, B) S(y, \mathrm{d}x) \tilde{\pi}(\mathrm{d}y) = \int_{\mathsf{Y}} S(y, A) S(y, B) \tilde{\pi}(\mathrm{d}y) \,.
\end{aligned}$$

This proves that $\pi \otimes RS$ is symmetric. □

Assume now that sampling from R or S is infeasible. In this case, we consider two instrumental kernels Q on $(\mathsf{X} \times \mathsf{Y}) \times \mathscr{Y}$ and T on $(\mathsf{X} \times \mathsf{Y}) \times \mathscr{X}$, which will be used to propose successive candidates for Y_{k+1} and X_{k+1}. For simplicity, assume that $R(x, \mathrm{d}y')$ and $Q(x, y; \mathrm{d}y')$ (resp. $S(y', \mathrm{d}x')$ and $T(x, y'; \mathrm{d}x')$) are dominated by the same measure, and call r and q (resp. s and t) the associated transition densities. We assume that r and s are known up to a normalizing constant. Define the Markov chain $\{(X_k, Y_k),\ k \in \mathbb{N}\}$ as follows. Given $(X_k, Y_k) = (x, y)$,

(DA1) draw a candidate $\tilde{Y}_{k+1}$ according to the distribution $Q(x, y; \cdot)$ and accept $Y_{k+1} = \tilde{Y}_{k+1}$ with probability $\alpha(x, y, \tilde{Y}_{k+1})$ defined by

$$\alpha(x, y, y') = \frac{r(x, y')q(x, y'; y)}{r(x, y)q(x, y; y')} \wedge 1 \,;$$

otherwise, set $Y_{k+1} = Y_k$; the Markov kernel on $\mathsf{X} \times \mathsf{Y} \times \mathscr{Y}$ associated with this transition is denoted by K_1;

(DA2) then draw a candidate $\tilde{X}_{k+1}$ according to the distribution $T(x, Y_{k+1}; \cdot)$ and accept $X_{k+1} = \tilde{X}_{k+1}$ with probability $\beta(x, Y_{k+1}, \tilde{X}_{k+1})$ defined by

$$\beta(x, y, x') = \frac{s(y, x')t(x', y; x)}{s(y, x)t(x, y; x')} \wedge 1 \,;$$

otherwise, set $X_{k+1} = X_k$; the Markov kernel on $\mathsf{X} \times \mathsf{Y} \times \mathscr{X}$ associated with this transition is denoted by K_2.

For $i = 1, 2$, let K_i^* be the kernels associated with K_1 and K_2 as follows: for $x \in \mathsf{X}$, $y \in \mathsf{Y}$, $A \in \mathscr{X}$, and $B \in \mathscr{Y}$,

$$K_1^*(x, y; A \times B) = \mathbb{1}_A(x) K_1(x, y; B) \,, \tag{2.3.12}$$

$$K_2^*(x, y; A \times B) = \mathbb{1}_B(y) K_2(x, y; A) \,. \tag{2.3.13}$$

Then the kernel of the chain $\{(X_n, Y_n),\ n \in \mathbb{N}\}$ is $K = K_1^* K_2^*$. The process $\{X_n,\ n \in \mathbb{N}\}$ is in general not a Markov chain, since the distribution of X_{k+1} conditionally on (X_k, Y_k) depends on (X_k, Y_k) and on X_k only, except in some special cases. Obviously, this construction includes the previous one, where sampling from R and S was feasible. Indeed, if $Q(x, y; \cdot) = R(x, \cdot)$ and $T(x, y; \cdot) = S(x, \cdot)$, then the acceptance probabilities α and β defined above simplify to one, the candidates are always accepted, and we are back to the previous algorithm.

Proposition 2.3.6 *The extended target distribution π^* is reversible with respect to the kernels K_1^* and K_2^* and invariant with respect to K.*

Proof. The reversibility of π^* with respect to K_1^* and K_2^* implies its invariance and consequently its invariance with respect to the product $K = K_1^* K_2^*$. Let us prove the reversibility of π^* with respect to K_1^*. For each $x \in \mathsf{X}$, the kernel $K_1(x,\cdot;\cdot)$ on $\mathsf{Y} \times \mathscr{Y}$ is the kernel of a Metropolis–Hastings algorithm with target density $r(x,\cdot)$, proposal kernel density $q(x,\cdot;\cdot)$, and acceptance probability $\alpha(x,\cdot,\cdot)$. By Proposition 2.3.1, this implies that the distribution $R(x,\cdot)$ is reversible with respect to the kernel $K_1(x,\cdot;\cdot)$. Applying the definition (2.3.13) of K_1^* and $\pi^* = \pi \otimes R$ yields, for $A, C \in \mathscr{X}$ and $B, D \in \mathscr{Y}$,

$$\begin{aligned}\pi^* \otimes K_1^*(A \times B \times C \times D) &= \iint_{A \times B} \pi(\mathrm{d}x\mathrm{d}y) K_1^*(x,y;C \times D) \\ &= \iint_{A \times B} \pi(\mathrm{d}x) R(x,\mathrm{d}y) \mathbb{1}_C(x) K_1(x,y,D) \\ &= \int_{A \cap C} \pi(\mathrm{d}x) [R(x,\cdot) \otimes K_1(x,\cdot;\cdot)](B \times D) \,.\end{aligned}$$

We have seen that for each $x \in \mathsf{X}$, the measure $R(x,\cdot) \otimes K_1(x,\cdot;\cdot)$ is symmetric; thus $\pi^* \otimes K^*$ is also symmetric. The reversibility of π^* with respect to K_2^* is proved similarly. $\square$

Example 2.3.7 (The slice sampler). Set $\mathsf{X} = \mathbb{R}^d$ and $\mathscr{X} = \mathscr{B}(\mathsf{X})$. Let μ be a σ-finite measure on $(\mathsf{X}, \mathscr{X})$ and let h be the density with respect to μ of the target distribution. We assume that for all $x \in \mathsf{X}$,

$$h(x) = C \prod_{i=0}^{k} f_i(x) \,,$$

where C is a constant (which is not necessarily known) and $f_i : \mathbb{R}^d \to \mathbb{R}_+$ are non-negative measurable functions. For $y = (y_1, \ldots, y_k) \in \mathbb{R}_+^k$, define

$$L(y) = \left\{ x \in \mathbb{R}^d \,:\, f_i(x) \geq y_i \,,\, i = 1, \ldots, k \right\} \,.$$

The f_0-slice-sampler algorithm proceeds as follows:

- given X_n, draw independently a k-tuple $Y_{n+1} = (Y_{n+1,1}, \ldots, Y_{n+1,k})$ of independent random variables such that $Y_{n+1,i} \sim \mathrm{Unif}(0, f_i(X_n))$, $i = 1, \ldots, k$.
- sample X_{n+1} from the distribution with density proportional to $f_0 \mathbb{1}_{L(Y_{n+1})}$.

Set $\mathsf{Y} = \mathbb{R}_+^k$ and for $(x,y) \in \mathsf{X} \times \mathsf{Y}$,

$$h^*(x,y) = Cf_0(x)\mathbb{1}_{L(y)}(x) = h(x)\prod_{i=1}^{k}\frac{\mathbb{1}_{[0,f_i(x)]}(y_i)}{f_i(x)} .$$

Let π^* be the probability measure with density h^* with respect to Lebesgue measure on $\mathsf{X}\times\mathsf{Y}$. Then $\int_{\mathsf{Y}} h^*(x,y)\mathrm{d}y = h(x)$ i.e., π is the first marginal of π^*. Let R be the kernel on $\mathsf{X}\times\mathscr{Y}$ with kernel density r defined by

$$r(x,y) = \frac{h^*(x,y)}{h(x)}\mathbb{1}\{h(x)>0\} .$$

Then $\pi^* = \pi\otimes R$. Define the distribution $\tilde{\pi} = \pi R$, its density $\tilde{h}(y) = \int_{\mathsf{X}} h^*(u,y)\mathrm{d}u$, and the kernel S on $\mathsf{Y}\times\mathscr{X}$ with density s by

$$s(y,x) = \frac{h^*(x,y)}{\tilde{h}(y)}\mathbb{1}\left\{\tilde{h}(y)>0\right\} .$$

If (X,Y) is a vector with distribution π^*, then $S(y,\cdot)$ is the conditional distribution of X given $Y=y$ and the Markov kernel of the chain $\{X_n,\ n\in\mathbb{N}\}$ is RS, and Lemma 2.3.5 can be applied to prove that π is reversible, hence invariant, with respect to RS.

2.3.3 *Two-Stage Gibbs Sampler*

The Gibbs sampler is a simple method that decomposes a complex multidimensional distribution into a collection of smaller-dimensional ones. Let $(\mathsf{X},\mathscr{X})$ and $(\mathsf{Y},\mathscr{Y})$ be complete separable metric spaces endowed with their Borel σ-fields. To construct the Markov chain $\{(X_n,Y_n),\ n\in\mathbb{N}\}$ with π^* as an invariant probability, we proceed exactly as in data-augmentation algorithms. Assume that π^* may be written as

$$\pi^*(\mathrm{d}x\mathrm{d}y) = \pi(\mathrm{d}x)R(x,\mathrm{d}y) = \tilde{\pi}(\mathrm{d}y)S(y,\mathrm{d}x), \tag{2.3.14}$$

where π and $\tilde{\pi}$ are probability measures on X and Y respectively, and R and S are kernels on $\mathsf{X}\times\mathscr{Y}$ and $\mathsf{Y}\times\mathscr{X}$ respectively.

The deterministic updating (two-stage) Gibbs (DUGS) sampler

When sampling from R and S is feasible, the DUGS sampler proceeds as follows: given (X_k,Y_k),

(DUGS1) sample Y_{k+1} from $R(X_k,\cdot)$,
(DUGS2) sample X_{k+1} from $S(Y_{k+1},\cdot)$.

For both the data augmentation algorithms and the two-stage Gibbs sampler we consider a distribution π^* on the product space $\mathsf{X}\times\mathsf{Y}$. In the former case, the distribution of interest is a marginal distribution of π^*, and in the latter case, the target distribution is π^* itself.

We may associate to each update (DUGS1) and (DUGS2) of the algorithm a transition kernel on $(\mathsf{X} \times \mathsf{Y}) \times (\mathscr{X} \otimes \mathscr{Y})$ defined for $(x,y) \in \mathsf{X} \times \mathsf{Y}$ and $A \times B \in \mathscr{X} \otimes \mathscr{Y}$ by

$$R^*(x,y;A \times B) = \mathbb{1}_A(x) R(x,B) \,, \tag{2.3.15}$$

$$S^*(x,y;A \times B) = \mathbb{1}_B(y) S(y,A) \,. \tag{2.3.16}$$

The transition kernel of the DUGS is then given by

$$P_{\mathrm{DUGS}} = R^* S^* \,. \tag{2.3.17}$$

Note that for $A \times B \in \mathscr{X} \otimes \mathscr{Y}$,

$$\begin{aligned} P_{\mathrm{DUGS}}(x,y;A \times B) &= \iint_{\mathsf{X}\times\mathsf{Y}} R^*(x,y;\mathrm{d}x'\mathrm{d}y') S^*(x',y';A \times B) \\ &= \iint_{\mathsf{X}\times\mathsf{Y}} R(x,\mathrm{d}y') \mathbb{1}_B(y') S(y',A) \\ &= \int_B R(x,\mathrm{d}y') S(y',A) = R \otimes S(x, B \times A) \,. \end{aligned} \tag{2.3.18}$$

As a consequence of Proposition 2.3.6, we obtain the invariance of π^*.

Lemma 2.3.8 *The distribution π^* is reversible with respect to the kernels R^* and S^* and invariant with respect to P_{DUGS}.*

The Random Scan Gibbs sampler (RSGS)

At each iteration, the RSGS algorithm consists in updating one component chosen at random. It proceeds as follows: given (X_k, Y_k),

(RSGS1) sample a Bernoulli random variable B_{k+1} with probability of success $1/2$.
(RSGS2) If $B_{k+1} = 0$, then sample Y_{k+1} from $R(X_k, \cdot)$; otherwise, sample X_{k+1} from $S(Y_{k+1}, \cdot)$.

The transition kernel of the RSGS algorithm can be written

$$P_{\mathrm{RSGS}} = \frac{1}{2} R^* + \frac{1}{2} S^* \,. \tag{2.3.19}$$

Lemma 2.3.8 implies that P_{RSGS} is reversible with respect to π^*, and therefore π^* is invariant for P_{RSGS}.

If sampling from R or S is infeasible, the Gibbs transitions can be replaced by a Metropolis–Hastings algorithm on each component as in the case of the DUGS algorithm. The algorithm is then called the two-stage Metropolis-within-Gibbs algorithm.

Example 2.3.9 (Scalar normal–inverse gamma). In a statistical problem one may be presented with a set of independent observations $\mathbf{y} = \{y_1, \ldots, y_n\}$, assumed to be normally distributed, but with unknown mean μ and variance τ^{-1} (τ is often referred to as the precision). The Bayesian approach to this problem is to assume that μ and τ are themselves random variables, with a given prior distribution. For example, we might assume that

$$\mu \sim \mathrm{N}(\theta_0, \phi_0^{-1}), \quad \tau \sim \Gamma(a_0, b_0), \tag{2.3.20}$$

i.e., μ is normally distributed with mean θ_0 and variance ϕ_0^{-1}, and τ has a gamma distribution with parameters a_0 and b_0.

The parameters θ_0, ϕ_0, a_0, and b_0 are assumed to be known. The posterior density h of (μ, τ), defined as the conditional density given the observations, is then given, using Bayes's formula, by

$$h(u,t) \propto \exp(-\phi_0(u-\theta_0)^2/2) \exp\left(-t\sum_{i=1}^n (y_i-u)^2/2\right) t^{a_0-1+n/2} \exp(-b_0 t) .$$

Conditioning on the observations introduces a dependence between μ and τ. Nevertheless, the conditional laws of μ given τ and τ given μ have a simple form. Write $\bar{y} = n^{-1}\sum_{i=1}^n y_i$ and $S^2 = n^{-1}\sum_{i=1}^n (y_i - \bar{y})^2$,

$$\begin{aligned}
&\theta_n(t) = (\phi_0\theta_0 + nt\bar{y})/(\phi_0 + nt), \quad \phi_n(t) = \phi_0 + nt, \\
&a_n = a_0 + n/2, \quad b_n(u) = b_0 + nS^2/2 + n(\bar{y}-u)^2/2 .
\end{aligned}$$

Then

$$\mathscr{L}(\mu|\tau) = \mathrm{N}(\theta_n(\tau), \phi_n^{-1}(\tau)), \qquad \mathscr{L}(\tau|\mu) = \Gamma(a_n, b_n(\mu)) .$$

The Gibbs sampler provides a simple approach to defining a Markov chain whose invariant probability has density h. First we simulate μ_0 and τ_0 independently with distribution as in (2.3.20). At the kth stage, given (μ_{k-1}, τ_{k-1}), we first simulate $N_k \sim \mathrm{N}(0,1)$ and $G_k \sim \Gamma(a_n, 1)$, and we set

$$\begin{aligned}
\mu_k &= \theta_n(\tau_{k-1}) + \phi_n^{-1/2}(\tau_{k-1}) N_k \\
\tau_k^{-1} &= b_n(\mu_k) G_k .
\end{aligned}$$

In the simple case that $\theta_0 = 0$ and $\phi_0 = 0$, which corresponds to a flat prior for μ (an improper distribution with a constant density on $\mathbb{R}$), the above equation can be rewritten as

$$\begin{aligned}
\mu_k &= \bar{y} + (n\tau_{k-1})^{-1/2} N_k \\
\tau_k^{-1} &= \left(b_0 + nS^2/2 + n(\bar{y}-\mu_k)^2/2\right) G_k .
\end{aligned}$$

Thus, $\{\tau_k^{-1},\ k \in \mathbb{N}\}$ and $\{(\mu_k - \bar{y})^2,\ k \in \mathbb{N}\}$ are Markov chains that follow the random coefficient autoregressions

$$\tau_k^{-1} = \frac{N_k^2 G_k}{2} \tau_{k-1}^{-1} + \left(b_0 + \frac{nS^2}{2} \right) G_k \,,$$
$$(\mu_k - \bar{y})^2 = \frac{N_k^2 G_{k-1}}{2} (\mu_{k-1} - \bar{y})^2 + \left(b_0 + \frac{nS^2}{2} \right) G_{k-1} \,.$$

2.3.4 Hit-and-Run Algorithm

Let K be a bounded subset of $\mathbb{R}^d$ with nonempty interior. Let $\rho : K \to [0, \infty)$ be a (not necessarily normalized) density, i.e., a nonnegative Lebesgue-integrable function. We define the probability measure π_ρ with density ρ by

$$\pi_\rho(A) = \frac{\int_A \rho(x) \mathrm{d}x}{\int_K \rho(x) \mathrm{d}x} \tag{2.3.21}$$

for all measurable sets $A \subset K$. For example, if $\rho(x) \equiv 1$, then π is simply the uniform distribution on K. The hit-and-run Markov kernel, presented below, can be used to sample approximately from π_ρ. The hit-and-run algorithm consists of two steps. Starting from $x \in K$, we first choose a random direction $\theta \in \mathrm{S}_{d-1}$, the unit sphere in $\mathbb{R}^d$ according to a uniform distribution on S_{d-1}. We then choose the next state of the Markov chain with respect to the density ρ restricted to the chord determined by the current state x and the direction $\theta \in \mathrm{S}_{d-1}$: for every function $f \in \mathbb{F}_+(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$,

$$H_\theta f(x) = \frac{1}{\ell_\rho(x,\theta)} \int_{s=-\infty}^{\infty} \mathbb{1}_K(x+s\theta) f(x+s\theta) \rho(x+s\theta) \mathrm{d}s \,, \tag{2.3.22}$$

where $\ell_\rho(x,\theta)$ is the normalizing constant defined as

$$\ell_\rho(x,\ \theta) = \int_{-\infty}^{\infty} \mathbb{1}_K(x+s\theta) \rho(x+s\theta) \mathrm{d}s \,. \tag{2.3.23}$$

The Markov kernel H that corresponds to the hit-and-run algorithm is therefore defined by, for all $f \in \mathbb{F}_+(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$,

$$Hf(x) = \int_{\mathrm{S}_{d-1}} H_\theta f(x) \sigma_{d-1}(\mathrm{d}\theta) \,, \tag{2.3.24}$$

where σ_{d-1} is the uniform distribution on S_{d-1}.

Lemma 2.3.10 *For all $\theta \in \mathrm{S}_{d-1}$, the Markov kernel H_θ is reversible with respect to π_ρ defined in (2.3.21). Furthermore, H is also reversible with respect to π_ρ.*

Proof. Let $c = \int_K \rho(x) \mathrm{d}x$ and $A, B \in \mathscr{B}(K)$. By elementary computations, we have

$$\begin{aligned}
\int_A H_\theta(x,\ B)\pi_\rho(\mathrm{d}x) &= \int_A \int_{-\infty}^{\infty} \frac{\mathbb{1}_B(x+s\theta)\rho(x+s\theta)\mathrm{d}s}{\ell_\rho(x,\theta)} \frac{\rho(x)\mathrm{d}x}{c} \\
&= \int_{\mathbb{R}^d} \int_{-\infty}^{\infty} \frac{\mathbb{1}_A(x)\mathbb{1}_B(x+s\theta)\rho(x+s\theta)\rho(x)}{\ell_\rho(x,\theta)} \frac{\mathrm{d}s\mathrm{d}x}{c} \\
&= \int_{\mathbb{R}^d} \int_{-\infty}^{\infty} \frac{\mathbb{1}_A(y-s\theta)\mathbb{1}_B(y)\rho(y)\rho(y-s\theta)}{\ell_\rho(y-s\theta,\theta)} \frac{\mathrm{d}s\mathrm{d}y}{c} \\
&= \int_B \int_{-\infty}^{\infty} \frac{\mathbb{1}_A(y-s\theta)\rho(y-s\theta)}{\ell_\rho(y-s\theta,\theta)} \mathrm{d}s\pi_\rho(\mathrm{d}y) \\
&= \int_B \int_{-\infty}^{\infty} \frac{\mathbb{1}_A(y-s\theta)\rho(y-s\theta)}{\ell_\rho(y,\theta)} \mathrm{d}s\pi_\rho(\mathrm{d}y) \\
&= \int_B \int_{-\infty}^{\infty} \frac{\mathbb{1}_A(y+t\theta)\rho(y+t\theta)}{\ell_\rho(y,\theta)} \mathrm{d}t\pi_\rho(\mathrm{d}y) \\
&= \int_B H_\theta(x,\ A)\pi_\rho(\mathrm{d}x)\ .
\end{aligned}$$

The reversibility of H_θ is proved. The reversibility of H_ρ follows from the reversibility of H_θ: for all $A, B \in \mathscr{B}(\mathbb{R}^d)$, we have

$$\begin{aligned}
\int_A H(x,B)\pi_\rho(\mathrm{d}x) &= \int_{\mathrm{S}_{d-1}} \int_A H_\theta(x,B)\pi_\rho(\mathrm{d}x)\sigma_{d-1}(\mathrm{d}\theta) \\
&= \int_{\mathrm{S}_{d-1}} \int_B H_\theta(x,A)\pi_\rho(\mathrm{d}x)\sigma_{d-1}(\mathrm{d}\theta) \\
&= \int_B H(x,A)\pi_\rho(\mathrm{d}x)\ .
\end{aligned}$$

□

2.4 Exercises

2.1 (Discrete autoregressive process). Consider the DAR(1) model defined by the recurrence

$$X_k = V_k X_{k-1} + (1-V_k)Z_k\ , \tag{2.4.1}$$

where $\{V_k,\ k \in \mathbb{N}\}$ is a sequence of i.i.d. Bernoulli random variables with $\mathbb{P}(V_k = 1) = \alpha \in [0,1)$, $\{Z_k,\ k \in \mathbb{N}\}$ are i.i.d. random variables with distribution π on a measurable space $(\mathsf{X}, \mathscr{X})$, and $\{V_k,\ k \in \mathbb{N}\}$ and $\{Z_k,\ k \in \mathbb{N}\}$ are mutually independent and independent of X_0, whose distribution is ξ.

1. Show that $Pf(x) = \alpha f(x) + (1-\alpha)\pi(f)$.
2. Show that π is the unique invariant probability.

Assume that $\mathsf{X} = \mathbb{N}$ and $\sum_{k=0}^{\infty} k^2\pi(k) < \infty$ and that the distribution of X_0 is π.

3. Show that for every positive integer h, $\operatorname{Cov}(X_h, X_0) = \alpha^h \operatorname{Var}(X_0)$.

2.2. Consider a scalar AR(1) process $\{X_k,\ k \in \mathbb{N}\}$ defined recursively as follows:

$$X_k = \phi X_{k-1} + Z_k \ , \tag{2.4.2}$$

where $\{Z_k,\ k \in \mathbb{N}\}$ is a sequence of i.i.d. random variables, independent of X_0, defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Assume that $\mathbb{E}[|Z_1|] < \infty$ and $\mathbb{E}[Z_1] = 0$.

1. Define the kernel P of this chain.
2. Show that for all $k \geq 1$, X_k has the same distribution as $\phi^k X_0 + B_k$, where $B_k = \sum_{j=0}^{k-1} \phi^j Z_j$.

Assume that $|\phi| < 1$.

3. Show that $B_k \xrightarrow{\mathbb{P}\text{-a.s.}} B_\infty = \sum_{j=0}^{\infty} \phi^j Z_j$.
4. Show that the distribution of B_∞ is the unique invariant probability of P.

Assume that $|\phi| > 1$ and the distribution of $\sum_{j=1}^{\infty} \phi^{-j} Z_j$ is continuous.

5. Show that for all $x \in \mathbb{R}$, $\mathbb{P}_x(\lim_{n\to\infty} |X_n| = \infty) = 1$.

2.3. Consider the bilinear process defined by the recurrence

$$X_k = aX_{k-1} + bX_{k-1}Z_k + Z_k \ , \tag{2.4.3}$$

where a and b are nonzero real numbers and $\{Z_k,\ k \in \mathbb{N}\}$ is a sequence of i.i.d. random variables that are independent of X_0. Assume that $\mathbb{E}[\ln(|a + bZ_0|)] < 0$ and $\mathbb{E}\left[\ln^+(|Z_0|)\right] < \infty$. Show that the bilinear model (2.4.3) has a unique invariant probability π and $\xi P^n \stackrel{\text{w}}{\Rightarrow} \pi$ for every initial distribution ξ.

2.4 (Exercise 1.7 continued). Consider the Markov chain given by the recurrence

$$X_k = \varepsilon_k X_{k-1} U_k + (1 - \varepsilon_k)\{X_{k-1} + U_k(1 - X_{k-1})\} \ , \tag{2.4.4}$$

where $\{\varepsilon_k,\ k \in \mathbb{N}\}$ is an i.i.d. sequence of Bernoulli random variables with probability of success $1/2$ and $\{U_k,\ k \in \mathbb{N}\}$ is an i.i.d. sequence of uniform random variables on $[0,1]$, both sequences being independent of X_0. Show that the Markov chain $\{X_k,\ k \in \mathbb{N}\}$ has a unique invariant stationary distribution.

2.5 (GARCH(1,1) process). Rewrite the GARCH$(1,1)$ process introduced in Example 2.2.2, which we can rewrite as

$$X_k = \sigma_k Z_k \ , \quad \sigma_k^2 = h(\sigma_{k-1}^2, Z_{k-1}^2) \ ,$$

with $h(x,z) = a_0 + (a_1 z + b_1)x$, $a_0 > 1$, $a_1, b_1 \geq 0$, and $\{Z_k,\ k \in \mathbb{Z}\}$ a sequence of i.i.d. random variables with zero mean and unit variance.

1. Prove that $\{\sigma_k^2\}$ is a Markov chain that satisfies the inequality (2.1.16) with $K(z) = a_1 z + b_1$.

2. Prove that a sufficient condition for the existence and uniqueness of an invariant probability is
$$\mathbb{E}\left[\log(a_1 Z_1^2 + b_1)\right] < 0 \,. \tag{2.4.5}$$
3. Prove that $a_1 + b_1 < 1$ is a necessary and sufficient condition for $\mathbb{E}\left[\sigma_k^2\right] < \infty$.
4. For $p \geq 1$, prove that a sufficient condition for the invariant probability of σ_k^2 to have a finite moment of order p is $\mathbb{E}\left[(a_1 Z_1^2 + b_1)^p\right] < 1$.

2.6 (Random coefficient autoregression). Consider the Markov chain $\{X_n,\ n \in \mathbb{N}\}$ on $\mathbb{R}^d$ defined by X_0 and the recurrence

$$X_k = A_k X_{k-1} + B_k \,, \tag{2.4.6}$$

where $\{(A_n, B_n),\ n \in \mathbb{N}\}$ is an i.i.d. sequence independent of X_0; A_n is a $d \times d$ matrix and B_n is a $d \times 1$ vector.

1. Prove that the forward and backward processes are given for $k \geq 1$ by

$$X_k^{x_0} = \sum_{j=0}^{k-1} \left(\prod_{i=k+1-j}^{k} A_i \right) B_{k-j} + \left(\prod_{i=1}^{k} A_i \right) x_0 \,,$$
$$Y_k^{x_0} = \sum_{j=0}^{k-1} \left(\prod_{i=1}^{j} A_i \right) B_{j+1} + \left(\prod_{i=1}^{k} A_i \right) x_0 \,,$$

where by convention, $\prod_{i=a}^{b} A_i = 1$ if $a > b$.

A subspace $L \subset \mathbb{R}^d$ is said to be invariant if for all $x \in L$, $\mathbb{P}(X_1 \in L \mid X_0 = x) = 1$. Assume that $\mathbb{E}\left[\log^+(|||A_1|||)\right] < \infty$, $\mathbb{E}\left[|B_1|\right] < \infty$, and that the only invariant subspace of $\mathbb{R}^d$ is $\mathbb{R}^d$ itself.

2. Show that if
$$\lim_{n\to\infty} n^{-1} \mathbb{E}\left[\log(|||A_1 \dots A_n|||)\right] < 0 \,,$$
then the random series $\sum_{j=1}^{\infty} \left(\prod_{i=1}^{j-1} A_j \right) B_j$ converges almost surely to a finite limit Y_∞.
3. Show that the distribution of Y_∞ is the unique invariant probability for the Markov chain $\{X_n^{x_0},\ n \in \mathbb{N}\}$ for all $x_0 \in \mathbb{R}^d$.

2.5 Bibliographical Notes

Random iterative functions produce a wealth of interesting examples of Markov chains. Diaconis and Freedman (1999) provides a survey with applications and very elegant convergence results. Exercise 2.4 is taken from Diaconis and Freedman (1999). Exercise 13.4 is taken from Diaconis and Hanlon (1992).

Markov chain Monte Carlo algorithms have received considerable attention since their introduction in statistics in the early 1980s. Most of the early work on this topic is summarized in Gelfand and Smith (1990), Geyer (1992), and Smith and Roberts (1993). The books Robert and Casella (2004), Gamerman and Lopes (2006), and Robert and Casella (2010) provide an in-depth introduction to Monte Carlo methods and their applications in Bayesian statistics. These books contain several chapters on MCMC methodology, which is illustrated with numerous examples. The handbook Brooks et al. (2011) provides an almost exhaustive account of the developments of MCMC methods up to 2010. It constitutes an indispensable source of references for MCMC algorithm design and applications in various fields. The surveys Roberts and Rosenthal (2004), Diaconis (2009), and Diaconis (2013) present many interesting developments.

There are many more examples of Markov chains that are not covered here. Examples of applications that are not covered in this book include queueing models (see Meyn (2006), Sericola (2013), Rubino and Sericola (2014)), stochastic control and Markov decision processes (see Hu and Yue (2008); Chang et al. (2013)), econometrics (see Frühwirth-Schnatter (2006)), and management sciences (see Ching et al. (2013)).

Chapter 3
Stopping Times and the Strong Markov Property

In this chapter, we will introduce what is arguably the single most important result of Markov chain theory, namely the strong Markov property.

To this end, we will first introduce the canonical space and the chain in Section 3.1. We will prove that it is always possible to consider that a Markov chain with state space $(\mathsf{X}, \mathscr{X})$ is defined on the product space $\mathsf{X}^{\mathbb{N}}$ endowed with the product σ-field $\mathscr{X}^{\otimes \mathbb{N}}$. This space is convenient for defining the shift operator, which is a key tool for the strong Markov property.

When studying the behavior of a Markov chain, it is often useful to decompose the sample paths into random subsequences depending on the successive visits to certain sets. The successive visits are examples of stopping times, that is, integer-valued random variables whose values depend only on the past of the trajectory up to this value. Stopping times will be formally introduced, and some of their general properties will be given in Section 3.2.

Having all the necessary tools in hand, we will be able to state and prove the Markov property and the strong Markov property in Section 3.3. Essentially, these properties mean that given a stopping time τ, the process $\{X_{\tau+k}, k \geq 0\}$ restricted to $\{\tau < \infty\}$ is a Markov chain with the same kernel as the original chain and independent of the history of the chain up to τ. Technically, they allow one to compute conditional expectations given the σ-field related to a stopping time. Thus the (strong) Markov property is easily understood to be of paramount importance in the theory of Markov chains seen from the sample path (or probabilistic) point of view, as opposed to a more operator-theoretic point of view, which will be only marginally taken in this book (see Chapters 18, 20 and 22). As an example of path decomposition using the successive visits to a given set, we will see in Section 3.4 the first-entrance last-exit decomposition, which will be used to derive several results in later chapters.

The fundamental notion of accessibility will be introduced in Definition 3.5.1. A set is said to be accessible if the probability to enter it is positive wherever the chain starts from. Chains that admit an accessible set will be considered in most parts of this book. In Section 3.6, which may be skipped on a first reading, very deep relations between return times to a set and invariant measures will be established.

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_3

3.1 The Canonical Chain

In this section, we show that given an initial distribution $\nu \in \mathbb{M}_1(\mathscr{X})$ and a Markov kernel P on $\mathsf{X} \times \mathscr{X}$, we can construct a Markov chain with initial distribution ν and transition kernel P on a specific filtered probability space, referred to as the canonical space. The following construction is valid for arbitrary measurable spaces $(\mathsf{X}, \mathscr{X})$.

Definition 3.1.1 (Coordinate process) *Let $\Omega = \mathsf{X}^{\mathbb{N}}$ be the set of X-valued sequences $\omega = (\omega_0, \omega_1, \omega_2, \dots)$ endowed with the σ-field $\mathscr{X}^{\otimes \mathbb{N}}$. The coordinate process $\{X_k,\ k \in \mathbb{N}\}$ is defined by*

$$X_k(\omega) = \omega_k\,, \quad \omega \in \Omega\,. \tag{3.1.1}$$

A point $\omega \in \Omega$ is referred to as a trajectory or a path.

For each $n \in \mathbb{N}$, define $\mathscr{F}_n = \sigma(X_m, m \le n)$. A set $A \in \mathscr{F}_n$ can be expressed as $A = A_n \times \mathsf{X}^{\mathbb{N}}$ with $A_n \in \mathscr{X}^{\otimes(n+1)}$. A function $f : \Omega \to \mathbb{R}$ is $\mathscr{F}_n$ measurable if it depends only on the first $n+1$ coordinates. In particular, the nth coordinate mapping X_n is measurable with respect to $\mathscr{F}_n$. The canonical filtration is $\{\mathscr{F}_n\,:\, n \in \mathbb{N}\}$. Define the algebra $\mathscr{A} = \cup_{n=0}^{\infty} \mathscr{F}_n$; the sets $A \in \mathscr{A}$ are called cylinders (or cylindrical sets) if

$$A = \prod_{n=0}^{\infty} A_n\,, \quad A_n \in \mathscr{X}\,,$$

where $A_n \neq \mathsf{X}$ for only finitely many n. Finally, we denote by $\mathscr{F}$ the σ-field generated by $\mathscr{A}$. By construction, $\mathscr{F} = \sigma(X_m, m \in \mathbb{N})$.

Theorem 3.1.2. *Let $(\mathsf{X}, \mathscr{X})$ be a measurable space and P a Markov kernel on $\mathsf{X} \times \mathscr{X}$. For every probability measure ν on $\mathscr{X}$, there exists a unique probability measure $\mathbb{P}_\nu$ on the canonical space $(\Omega, \mathscr{F}) = (\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}})$ such that the canonical process $\{X_n,\ n \in \mathbb{N}\}$ is a Markov chain with kernel P and initial distribution ν.*

Proof. We define a set function μ on $\mathscr{A}$ by

$$\mu(A) = \nu \otimes P^{\otimes n}(A_n)\,,$$

whenever $A = A_n \times \mathsf{X}^{\mathbb{N}}$ with $A_n \in \mathscr{X}^{\otimes(n+1)}$ (it is easy to see that this expression does not depend on the choice of n and A_n). Since $\nu \otimes P^{\otimes n}$ is a probability measure on $(\mathsf{X}^{n+1}, \mathscr{X}^{\otimes n})$ for every n, it is clear that μ is an additive set function on $\mathscr{A}$. We must now prove that μ is σ-additive. Let $\{F_n,\ n \in \mathbb{N}^*\}$ be a collection of mutually disjoint sets of $\mathscr{A}$ such that $\cup_{n=1}^{\infty} F_n \in \mathscr{A}$. For $n \geq 1$, define $B_n = \cup_{k=n}^{\infty} F_k$. Note that $B_n \in \mathscr{A}$, since $B_n = \left(\cup_{k=1}^{\infty} F_k\right) \setminus \left(\cup_{k=1}^{n-1} F_k\right)$. Since

$$\mu\left(\cup_{k=1}^{\infty} F_k\right) = \mu(B_n) + \mu\left(\cup_{k=1}^{n} F_k\right) = \mu(B_n) + \sum_{k=1}^{n} \mu(F_k) ,$$

this amounts to establishing that

$$\lim_{n\to\infty} \mu(B_n) = 0 . \tag{3.1.2}$$

Note that $B_{n+1} \subset B_n$ for all $n \in \mathbb{N}$ and $\cap_{n=1}^{\infty} B_n = \emptyset$. Set $B_0 = \mathsf{X}$. For each $n \in \mathbb{N}$, there exists $k(n)$ such that $B_n \in \mathscr{F}_{k(n)}$ and the σ-fields $\mathscr{F}_n$ are increasing. Thus we can assume that the sequence $\mathscr{F}_{k(n)}$ is also increasing. Moreover, by repeating if necessary the terms B_n in the sequence, we can assume that B_n is $\mathscr{F}_n$ measurable for all $n \in \mathbb{N}$.

For $0 \leq k \leq n$, we define kernels $Q_{k,n}$ on $\mathsf{X}^{k+1} \times \mathscr{F}_n$ as follows. If f is a nonnegative $\mathscr{F}_n$-measurable function, then as noted earlier we can identify it with an $\mathscr{X}^{\otimes(n+1)}$-measurable function $\bar{f}$, and we set for $k \geq 1$,

$$Q_{k,n} f(x_0,\ldots,x_k) = \int_{\mathsf{X}^{n-k}} P(x_k, \mathrm{d}x_{k+1}) \ldots P(x_{n-1}, \mathrm{d}x_n) \bar{f}(x_0,\ldots,x_n) .$$

By convention, $Q_{n,n}$ is the identity kernel. For every $n \geq 0$, B_n can be expressed as $B_n = C_n \times \mathsf{X}^{\mathbb{N}}$ with $C_n \in \mathscr{X}^{\otimes(n+1)}$. For $0 \leq k \leq n$, we define the $\mathscr{X}^{\otimes(k+1)}$-measurable function f_k^n by

$$f_k^n = Q_{k,n} \mathbb{1}_{B_n} .$$

For each fixed $k \geq 0$, the sequence $\left\{f_k^n : n \in \mathbb{N}\right\}$ is nonnegative and nonincreasing. Therefore, it is convergent. Moreover, it is uniformly bounded by 1. Set

$$f_k^{\infty} = \lim_{n\to\infty} f_k^n .$$

By construction, for each $k < n$ we get that $f_k^n = Q_{k,k+1} f_{k+1}^n$. Thus by Lebesgue's dominated convergence theorem, we have

$$f_k^{\infty} = Q_{k,k+1} f_{k+1}^{\infty} .$$

We now prove by contradiction that $\lim_n \mu(B_n) = 0$. Otherwise, since $\mu(B_n) = \nu Q_{0,n}(B_n)$ and $f_0^{\infty} = \lim_{n\to\infty} Q_{0,n}(B_n)$, the dominated convergence theorem yields

$$\nu(f_0^{\infty}) = \nu\left(\lim_{n\to\infty} Q_{0,n}(B_n)\right) = \lim_{n\to\infty} \nu Q_{0,n}(B_n) = \lim_{n\to\infty} \mu(B_n) > 0 .$$

Therefore, there exists $\bar{x}_0 \in \mathsf{X}$ such that $f_0^{\infty}(\bar{x}_0) > 0$. Then

$$0 < f_0^\infty(\bar{x}_0) = \int_{\mathsf{X}} P(\bar{x}_0, \mathrm{d}x_1) f_1^\infty(\bar{x}_0, x_1) \ .$$

Therefore, there exists $\bar{x}_1 \in \mathsf{X}$ such that $f_1^\infty(\bar{x}_0, \bar{x}_1) > 0$. This in turn yields

$$0 < f_1^\infty(\bar{x}_0, \bar{x}_1) = \int_{\mathsf{X}} P(\bar{x}_1, \mathrm{d}x_2) f_2^\infty(\bar{x}_0, \bar{x}_1, x_2) \ ,$$

and therefore there exists $\bar{x}_2 \in \mathsf{X}$ such that $f_2^\infty(\bar{x}_0, \bar{x}_1, \bar{x}_2) > 0$. By induction, we can build a sequence $\mathbf{x} = \{\bar{x}_n, \ n \in \mathbb{N}\}$ such that for all $k \geq 1$, $f_k^\infty(\bar{x}_0, \ldots, \bar{x}_k) > 0$. This yields, for all $k \geq 1$,

$$\mathbb{1}_{B_k}(\mathbf{x}) = f_k^k(\bar{x}_0, \ldots, \bar{x}_k) \geq f_k^\infty(\bar{x}_0, \ldots, \bar{x}_k) > 0 \ .$$

Since an indicator function takes only the values 0 and 1, this implies that $\mathbb{1}_{B_k}(\mathbf{x}) = 1$ for all $k \geq 0$. Thus $\mathbf{x} \in \cap_{k=0}^\infty B_k$, which contradicts the assumption $\cap_{k=0}^\infty B_k = \emptyset$. This proves (3.1.2). Therefore, μ is σ-additive on $\mathscr{A}$, and thus by Theorem B.2.8, it can be uniquely extended to a probability measure on $\mathscr{F}$. □

The expectation associated to $\mathbb{P}_\nu$ will be denoted by $\mathbb{E}_\nu$, and for $x \in \mathsf{X}$, $\mathbb{P}_x$ and $\mathbb{E}_x$ will be shorthand for $\mathbb{P}_{\delta_x}$ and $\mathbb{E}_{\delta_x}$.

Proposition 3.1.3 *For all $A \in \mathscr{X}^{\otimes \mathbb{N}}$,*

(i) the function $x \mapsto \mathbb{P}_x(A)$ is $\mathscr{X}$-measurable;
(ii) for all $\nu \in \mathbb{M}_1(\mathscr{X})$, $\mathbb{P}_\nu(A) = \int_{\mathsf{X}} \mathbb{P}_x(A) \nu(\mathrm{d}x)$.

Proof. Let $\mathscr{M}$ be the set of all $A \in \mathscr{X}^{\otimes \mathbb{N}}$ satisfying (i) and (ii). The set $\mathscr{M}$ is a monotone class and contains all the sets of the form $\prod_{i=1}^n A_i$, $A_i \in \mathscr{X}$, $n \in \mathbb{N}$, by (1.3.2). Hence, Theorem B.2.2 shows that $\mathscr{M} = \mathscr{X}^{\otimes \mathbb{N}}$. □

Definition 3.1.4 (Canonical Markov chain) *The canonical Markov chain with kernel P on $\mathsf{X} \times \mathscr{X}$ is the coordinate process $\{X_n, \ n \in \mathbb{N}\}$ on the canonical filtered space $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}}, \{\mathscr{F}_k^X, \ k \in \mathbb{N}\})$ endowed with the family of probability measures $\{\mathbb{P}_\nu, \nu \in \mathbb{M}_1(\mathscr{X})\}$ given by Theorem 3.1.2.*

In the sequel, unless explicitly stated otherwise, a Markov chain with kernel P on $\mathsf{X} \times \mathscr{X}$ will refer to the canonical chain on $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}})$.

One must be aware that with the canonical Markov chain on the canonical space $\mathsf{X}^{\mathbb{N}}$ comes a family of probability measures, indexed by the set of probability measures on $(\mathsf{X}, \mathscr{X})$. A property might be almost surely true with respect to one probability measure $\mathbb{P}_\mu$ and almost surely false with respect to another one.

Definition 3.1.5 *A property is true* $\mathbb{P}_* -$ a.s. *if it is almost surely true with respect to* $\mathbb{P}_\nu$ *for all initial distributions* $\nu \in \mathbb{M}_1(\mathscr{X})$*. Moreover, if for some* $A \in \mathscr{X}^{\otimes \mathbb{N}}$*, the probability* $\mathbb{P}_\nu(A)$ *does not depend on* ν*, then we simply write* $\mathbb{P}_*(A)$ *instead of* $\mathbb{P}_\nu(A)$*.*

By Proposition 3.1.3 (ii), a property is true $\mathbb{P}_* -$ a.s. if and only if it is almost surely true with respect to $\mathbb{P}_x$ for all x in X.

If a kernel P admits an invariant measure π, then by Theorem 1.4.2, the canonical chain $\{X_k,\ k \in \mathbb{N}\}$ is stationary under $\mathbb{P}_\pi$.

Remark 3.1.6. *Assume that the probability measure* π *is reversible with respect to* P *(see Definition 1.5.1). Define for all* $k \geq 0$*, the* σ*-field* $\mathscr{G}_k = \sigma(X_\ell\,,\ \ell \geq k)$*. It can be easily checked that for all* $A \in \mathscr{X}$ *and* $k \geq 0$*,*

$$\mathbb{P}\left(X_k \in A \,|\, \mathscr{G}_{k+1}\right) = P(X_{k+1}, A)\ ,$$

showing that P *is also the Markov kernel for the reverse-time Markov chain. From an initial distribution* π *at time* 0*, we can therefore use Theorem 3.1.2 applied to the Markov kernel* P *in both directions to construct a probability* $\mathbb{P}_\pi$ *on* $(\mathsf{X}^{\mathbb{Z}}, \mathscr{X}^{\otimes \mathbb{Z}})$ *such that the coordinate process* $\{X_k,\ k \in \mathbb{Z}\}$ *is a stationary Markov chain under* $\mathbb{P}_\pi$*.* ▲

Nevertheless, when π is no longer reversible with respect to P, the extension to a stationary process indexed by $\mathbb{Z}$ is not possible in full generality.

Theorem 3.1.7 (Stationary Markov chain indexed by $\mathbb{Z}$). *Let* X *be a Polish space endowed with its Borel* σ*-field. Let* P *be a Markov kernel on* $(\mathsf{X}, \mathscr{X})$ *that admits an invariant measure* π*. Then there exists a unique probability measure on* $(\mathsf{X}^{\mathbb{Z}}, \mathscr{X}^{\otimes \mathbb{Z}})$*, still denoted by* $\mathbb{P}_\pi$*, such that the coordinate process* $\{X_k,\ k \in \mathbb{Z}\}$ *is a stationary Markov chain under* $\mathbb{P}_\pi$*.*

Proof. For $k \leq n \in \mathbb{Z}$, let $\mu_{k,n}$ be the probability measure on X^{n-k} defined by

$$\mu_{k,n} = \pi \otimes P^{\otimes(n-k)}\ .$$

Then $\mu_{k,n}$ is a consistent family of probability measures. Therefore, by Theorem B.3.17, there exists a probability measure $\mathbb{P}_\pi$ on $(\mathsf{X}^{\mathbb{Z}}, \mathscr{X}^{\otimes \mathbb{Z}})$ such that for all $k \leq n \in \mathbb{Z}$, the distribution of $(X_k, \ldots, X_n)$ is $\mu_{k,n}$. Since $\mu_{k,n}$ depends only on $n-k$, the coordinate process is stationary. □

We now define the shift operator on the canonical space.

Definition 3.1.8 (Shift operator) *Let* $(\mathsf{X},\mathscr{X})$ *be a measurable space. The mapping* $\theta : \mathsf{X}^{\mathbb{N}} \to \mathsf{X}^{\mathbb{N}}$ *defined by*

$$\theta w = (w_0, w_1, w_2, \ldots) \mapsto \theta(w) = (w_1, w_2, \ldots) ,$$

is called the shift operator.

Proposition 3.1.9 *The shift operator* θ *is measurable with respect to* $\mathscr{X}^{\otimes\mathbb{N}}$.

Proof. For $n \in \mathbb{N}^*$ and $H \in \mathscr{X}^{\otimes n}$, consider the cylinder $H \times \mathsf{X}^{\mathbb{N}}$, that is,

$$H \times \mathsf{X}^{\mathbb{N}} = \{\omega \in \Omega \,:\, (\omega_0, \ldots, \omega_{n-1}) \in H\} .$$

Then

$$\theta^{-1}(H \times \mathsf{X}^{\mathbb{N}}) = \{\omega \in \Omega \,:\, (\omega_0, \ldots, \omega_n) \in \mathsf{X} \times H\} = \mathsf{X} \times H \times \mathsf{X}^{\mathbb{N}} ,$$

which is another cylinder, and since the cylinders generate the σ-field, it follows that $\mathscr{X}^{\otimes\mathbb{N}} = \sigma(\mathscr{C}_0)$, where $\mathscr{C}_0$ is the semialgebra of cylinders. Therefore, the shift operator is measurable. □

We define inductively θ_0 as the identity function, i.e., $\theta_0(w) = w$ for all $w \in \mathsf{X}^{\mathbb{N}}$, and for $k \geq 1$,

$$\theta_k = \theta_{k-1} \circ \theta .$$

Let $\{X_k,\ k \in \mathbb{N}\}$ be the coordinate process on $\mathsf{X}^{\mathbb{N}}$, as defined in (3.1.1). Then for $(j,k) \in \mathbb{N}^2$, one has

$$X_k \circ \theta_j = X_{j+k} .$$

Moreover, for all $p,k \in \mathbb{N}$ and $A_0, \ldots, A_p \in \mathscr{X}$,

$$\theta_k^{-1}\{X_0 \in A_0, \ldots, X_p \in A_p\} = \{X_k \in A_0, \ldots, X_{k+p} \in A_p\} ;$$

thus θ_k is measurable as a map from $(\mathsf{X}^{\mathbb{N}}, \sigma(X_j, j \geq k))$ to $(\mathsf{X}^{\mathbb{N}}, \mathscr{F}_\infty)$.

3.2 Stopping Times

In this section, we consider a filtered probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_k,\ k \in \mathbb{N}\}, \mathbb{P})$ and an adapted process $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$. We set $\mathscr{F}_\infty = \bigvee_{k\in\mathbb{N}} \mathscr{F}_k$, the sub-$\sigma$-field of

$\mathscr{F}$ generated by $\{\mathscr{F}_n,\ n \in \mathbb{N}\}$. In most applications, the sequence $\{\mathscr{F}_n,\ n \in \mathbb{N}\}$ is an increasing sequence. In many examples, for $n \in \mathbb{N}$, we have $\mathscr{F}_n = \sigma(Y_m, 0 \le m \le n)$, where $\{Y_m,\ m \in \mathbb{N}\}$ is a sequence of random variables; in this case, the σ-field $\mathscr{F}_\infty$ is the σ-field generated by the infinite sequence $\{Y_n,\ n \in \mathbb{N}\}$.

The term "stopping time" is an expression from gambling. A game of chance that evolves in time (for example, an infinite sequence of coin tosses) can be adequately represented by a filtered space $(\Omega, \mathscr{F}, \{\mathscr{F}_k,\ k \in \mathbb{N}\}, \mathbb{P})$, the sub-$\sigma$-fields $\mathscr{F}_n$ giving information on the results of the game available to the player at time n. A stopping rule is thus a set of conditions by which a player decides whether to leave the game at time n, based at each time on the information at his disposal at that time. A time ρ at which a game is terminated by such a rule is a stopping time. Note that stopping times may take the value $+\infty$, corresponding to the case in which the game never stops.

Definition 3.2.1 (Stopping times)

(i) A random variable τ from Ω to $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ is called a stopping time if for all $k \in \mathbb{N}$, $\{\tau = k\} \in \mathscr{F}_k$.
(ii) The family $\mathscr{F}_\tau$ of events $A \in \mathscr{F}$ such that for every $k \in \mathbb{N}$, $A \cap \{\tau = k\} \in \mathscr{F}_k$, is called the σ-field of events prior to time τ.

It can be easily checked that $\mathscr{F}_\tau$ is indeed a σ-field. Since $\{\tau = n\} = \{\tau \le n\} \setminus \{\tau \le n-1\}$, one can replace $\{\tau = n\}$ by $\{\tau \le n\}$ in the definition of the stopping time τ and in the definition of the σ-field $\mathscr{F}_\tau$. It may sometimes be useful to note that the constant random variables are also stopping times. In such a case, there exists $n \in \mathbb{N}$ such that $\tau(\omega) = n$ for every $\omega \in \Omega$ and $\mathscr{F}_\tau = \mathscr{F}_n$.

For every stopping time τ, the event $\{\tau = \infty\}$ belongs to $\mathscr{F}_\infty$, for it is the complement of the union of the events $\{\tau = n\}$, $n \in \mathbb{N}$, which all belong to $\mathscr{F}_\infty$. It follows that $B \cap \{\tau = \infty\} \in \mathscr{F}_\infty$ for all $B \in \mathscr{F}_\tau$, showing that $\tau : \Omega \to \bar{\mathbb{N}}$ is $\mathscr{F}_\infty$ measurable.

Definition 3.2.2 (Hitting times and return times) *For $A \in \mathscr{X}$, the first hitting time τ_A and return time σ_A of the set A by the process $\{X_n,\ n \in \mathbb{N}\}$ are defined respectively by*

$$\tau_A = \inf\{n \ge 0 : X_n \in A\} , \tag{3.2.1}$$
$$\sigma_A = \inf\{n \ge 1 : X_n \in A\} , \tag{3.2.2}$$

where by convention, $\inf \emptyset = +\infty$. *The successive return times $\sigma_A^{(n)}$, $n \ge 0$, are defined inductively by $\sigma_A^{(0)} = 0$ and for all $k \ge 0$,*

$$\sigma_A^{(k+1)} = \inf\left\{n > \sigma_A^{(k)} : X_n \in A\right\} . \tag{3.2.3}$$

It can be readily checked that return and hitting times are stopping times. For example,

$$\{\tau_A = n\} = \bigcap_{k=0}^{n-1} \{X_k \notin A\} \cap \{X_n \in A\} \in \mathscr{F}_n ,$$

so that τ_A is a stopping time.

We want to define the position of the process $\{X_n\}$ at time τ, i.e., $X_\tau(\omega) = X_{\tau(\omega)}(\omega)$. This quantity is not defined when $\tau(\omega) = \infty$. To handle this situation, we select an arbitrary $\mathscr{F}_\infty$-measurable random variable X_∞, and we set

$$X_\tau = X_k \text{ on } \{\tau = k\} , \quad k \in \bar{\mathbb{N}} .$$

Note that the random variable X_τ is $\mathscr{F}_\tau$-measurable, since for $A \in \mathscr{X}$ and $k \in \mathbb{N}$,

$$\{X_\tau \in A\} \cap \{\tau = k\} = \{X_k \in A\} \cap \{\tau = k\} \in \mathscr{F}_k .$$

3.3 The Strong Markov Property

Let τ be an integer-valued random variable. Define θ_τ on $\{\tau < \infty\}$ by

$$\theta_\tau(w) = \theta_{\tau(w)}(w) . \tag{3.3.1}$$

With this definition, we have $X_\tau = X_k$ on $\{\tau = k\}$ and $X_k \circ \theta_\tau = X_{\tau+k}$ on $\{\tau < \infty\}$.

Proposition 3.3.1 *Let $\{\mathscr{F}_n,\ n \in \mathbb{N}\}$ be the natural filtration of the coordinate process $\{X_n,\ n \in \mathbb{N}\}$. Let τ and σ be two stopping times with respect to $\{\mathscr{F}_n,\ n \in \mathbb{N}\}$.*

(i) For all integers $n, m \in \mathbb{N}$, $\theta_n^{-1}(\mathscr{F}_m) = \sigma(X_n,\ \ldots,\ X_{n+m})$.
(ii) The random variable defined by

$$\rho = \begin{cases} \sigma + \tau \circ \theta_\sigma & \text{on } \{\sigma < \infty\}, \\ \infty & \text{otherwise}, \end{cases}$$

is a stopping time. Moreover, on $\{\sigma < \infty\} \cap \{\tau < \infty\}$, $X_\tau \circ \theta_\sigma = X_\rho$.

Proof. (i) For all $A \in \mathscr{X}$ and all integers $k, n \in \mathbb{N}^2$,

$$\theta_n^{-1}\{X_k \in A\} = \{X_k \circ \theta_n \in A\} = \{X_{k+n} \in A\} .$$

Since the σ-field $\mathscr{F}_m$ is generated by the events of the form $\{X_k \in A\}$, where $A \in \mathscr{X}$ and $k \in \{0, \ldots, m\}$, the σ-field $\theta_n^{-1}(\mathscr{F}_m)$ is generated by the events $\{X_{k+n} \in A\}$, where $A \in \mathscr{X}$ and $k \in \{0, \ldots, m\}$, and by definition, the latter events generate the σ-field $\sigma(X_n,\ \ldots,\ X_{n+m})$.

(ii) We will first prove that for every positive integer k, $k+\tau\circ\theta_k$ is a stopping time. Since τ is a stopping time, $\{\tau=m-k\}\in\mathscr{F}_{m-k}$, and by (i), it is also the case that $\theta_k^{-1}\{\tau=m-k\}\in\mathscr{F}_m$. Thus,

$$\{k+\tau\circ\theta_k=m\}=\{\tau\circ\theta_k=m-k\}=\theta_k^{-1}\{\tau=m-k\}\in\mathscr{F}_m\ .$$

This proves that $k+\tau\circ\theta_k$ is a stopping time. We now consider the general case. From the definition of ρ, we obtain

$$\begin{aligned}\{\rho=m\}=\{\sigma+\tau\circ\theta_\sigma=m\}&=\bigcup_{k=0}^{m}\{k+\tau\circ\theta_k=m,\ \sigma=k\}\\&=\bigcup_{k=0}^{m}\{k+\tau\circ\theta_k=m\}\cap\{\sigma=k\}.\end{aligned}$$

Since σ is a stopping time and since $k+\tau\circ\theta_k$ is a stopping time for each k, we obtain that $\{\rho=m\}\in\mathscr{F}_m$. Thus ρ is a stopping time. By construction, if $\tau(\omega)$ and $\sigma(\omega)$ are finite, we have

$$X_\tau\circ\theta_\sigma(\omega)=X_{\tau\circ\theta_{\sigma(\omega)}}(\theta_\sigma(\omega))=X_{\sigma+\tau\circ\theta_\sigma}(\omega)\ .$$

□

Proposition 3.3.2 *The successive return times to a measurable set A are stopping times with respect to the natural filtration of the canonical process $\{X_n,\ n\in\mathbb{N}\}$. In addition, $\sigma_A=1+\tau_A\circ\theta_1$, and for $n\geq 0$,*

$$\sigma_A^{(n+1)}=\sigma_A^{(n)}+\sigma_A\circ\theta_{\sigma_A^{(n)}}\quad\text{on}\quad\{\sigma_A^{(n)}<\infty\}\ .$$

Proof. The proof is a straightforward application of Proposition 3.3.1 (ii). □

Theorem 3.3.3 (Markov property). *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$ and $\nu\in\mathbb{M}_1(\mathscr{X})$. For every $\mathscr{F}$-measurable positive or bounded random variable Y, initial distribution $\nu\in\mathbb{M}_1(\mathscr{X})$, and $k\in\mathbb{N}$, one has*

$$\mathbb{E}_\nu\left[Y\circ\theta_k\middle|\mathscr{F}_k\right]=\mathbb{E}_{X_k}[Y]\qquad\mathbb{P}_\nu-\text{a.s.}\tag{3.3.2}$$

Proof. We apply a monotone class theorem (see Theorem B.2.4). Let $\mathscr{H}$ be the vector space of bounded random variables Y such that (3.3.2) holds. By the monotone convergence theorem, if $\{Y_n,\ n\in\mathbb{N}\}$ is an increasing sequence of nonnegative random variables in $\mathscr{H}$ such that $\lim_{n\to\infty} Y_n = Y$ is bounded, then Y satisfies (3.3.2). It suffices to check that $\mathscr{H}$ contains the random variables $Y = g(X_0, \ldots, X_j)$, where $j \geq 0$ and g is any bounded measurable function on X^{j+1}; i.e., we need to prove that

$$\mathbb{E}_\nu[f(X_0,\ldots,X_k)g(X_k,\ldots,X_{k+j})] = \mathbb{E}_\nu[f(X_0,\ldots,X_k)\mathbb{E}_{X_k}[g(X_0,\ldots,X_j)]] .$$

This identity follows easily from (1.3.2). □

Remark 3.3.4 *A more general version of the Markov property whereby $\{\mathscr{F}_n : n \in \mathbb{N}\}$ is not necessarily the natural filtration can be obtained from Theorem 1.1.2 (ii).*

The Markov property can be significantly extended to random time shifts.

Theorem 3.3.5 (Strong Markov property). *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $\nu \in \mathbb{M}_1(\mathscr{X})$. For every $\mathscr{F}$-measurable positive or bounded random variable Y, initial distribution $\nu \in \mathbb{M}_1(\mathscr{X})$, and stopping time τ, one has*

$$\mathbb{E}_\nu\left[Y \circ \theta_\tau \mathbb{1}_{\{\tau<\infty\}}\middle| \mathscr{F}_\tau\right] = \mathbb{E}_{X_\tau}[Y]\mathbb{1}_{\{\tau<\infty\}} \qquad \mathbb{P}_\nu - \text{a.s.} \tag{3.3.3}$$

Proof. We will show that for all $A \in \mathscr{F}_\tau$,

$$\mathbb{E}_\nu\left[\mathbb{1}_A\, Y \circ \theta_\tau\, \mathbb{1}_{\{\tau<\infty\}}\right] = \mathbb{E}_\nu\left[\mathbb{1}_A\, \mathbb{E}_{X_\tau}[Y]\mathbb{1}_{\{\tau<\infty\}}\right] . \tag{3.3.4}$$

Since for all $k \in \mathbb{N}$, $A \cap \{\tau = k\} \in \mathscr{F}_k$, the Markov property (Theorem 3.3.3) implies

$$\begin{aligned}\mathbb{E}_\nu\left[\mathbb{1}_{A\cap\{\tau=k\}} Y \circ \theta_\tau\right] &= \mathbb{E}_\nu\left[\mathbb{1}_{A\cap\{\tau=k\}} Y \circ \theta_k\right] \\ &= \mathbb{E}_\nu\left[\mathbb{1}_{A\cap\{\tau=k\}} \mathbb{E}_{X_k}[Y]\right] = \mathbb{E}_\nu\left[\mathbb{1}_{A\cap\{\tau=k\}} \mathbb{E}_{X_\tau}[Y]\right] .\end{aligned}$$

Equation (3.3.4) follows by noting that

$$\begin{aligned}\mathbb{E}_\nu[\mathbb{1}_A\, Y \circ \theta_\tau\, \mathbb{1}_{\{\tau<\infty\}}] &= \sum_{k=0}^{\infty} \mathbb{E}_\nu[\mathbb{1}_{A\cap\{\tau=k\}}\, Y \circ \theta_k] \\ &= \sum_{k=0}^{\infty} \mathbb{E}_\nu\left[\mathbb{1}_{A\cap\{\tau=k\}} \mathbb{E}_{X_\tau}[Y]\right] = \mathbb{E}_\nu[\mathbb{1}_A \mathbb{1}_{\{\tau<\infty\}} \mathbb{E}_{X_\tau}[Y]] .\end{aligned}$$

□

We illustrate the use of the strong Markov property with some important properties of return times.

Proposition 3.3.6 *Let $C \in \mathscr{X}$.*

(i) If for all $x \in C$, $\mathbb{P}_x(\sigma_C < \infty) = 1$, then for all $x \in C$ and $n \in \mathbb{N}$, $\mathbb{P}_x(\sigma_C^{(n)} < \infty) = 1$.
(ii) If for all $x \in C^c$, $\mathbb{P}_x(\sigma_C < \infty) = 1$, then $\mathbb{P}_x(\sigma_C < \infty) = 1$ for all $x \in \mathsf{X}$.

Proof. (i) The proof is by induction on $n \geq 1$. First note that by assumption, $\mathbb{P}_x(\sigma_C^{(1)} < \infty) = 1$ for all $x \in C$. Assume that $\mathbb{P}_x(\sigma_C^{(n)} < \infty) = 1$ for all $x \in C$. By the strong Markov property, we get for all $x \in C$,

$$\begin{aligned}
\mathbb{P}_x(\sigma_C^{(n+1)} < \infty) &= \mathbb{P}_x(\sigma_C^{(n)} < \infty,\ \sigma_C \circ \theta_{\sigma_C^{(n)}} < \infty) \\
&= \mathbb{E}_x\left[\mathbb{1}_{\{\sigma_C^{(n)} < \infty\}} \mathbb{P}_{X_{\sigma_C^{(n)}}}(\sigma_C < \infty)\right] = \mathbb{P}_x(\sigma_C^{(n)} < \infty) = 1 \; .
\end{aligned}$$

(ii) For $x \in \mathsf{X}$, we have

$$\begin{aligned}
\mathbb{P}_x(\sigma_C < \infty) &= \mathbb{P}_x(X_1 \in C) + \mathbb{P}_x(X_1 \in C^c, \sigma_C \circ \theta < \infty) \\
&= \mathbb{P}_x(X_1 \in C) + \mathbb{P}_x(X_1 \notin C) = 1 \; .
\end{aligned}$$

□

For every set $C \in \mathscr{X}$, denote by $\mathscr{X}_C$ the subset of $\mathscr{X}$ defined as

$$\mathscr{X}_C = \{A \cap C : A \in \mathscr{X}\} \; . \tag{3.3.5}$$

It is easily seen that $\mathscr{X}_C$ is a σ-field, often called the trace σ-field on C or the induced σ-field on C.

Definition 3.3.7 (Induced kernel) *For all $C \in \mathscr{X}$, the induced kernel Q_C on $C \times \mathscr{X}_C$ is defined by*

$$Q_C(x, B) = \mathbb{P}_x(X_{\sigma_C} \in B,\ \sigma_C < \infty) \; , \qquad x \in C \, , \; B \in \mathscr{X}_C \; . \tag{3.3.6}$$

Let $\{X_n,\ n \in \mathbb{N}\}$ be a Markov chain on $(\mathsf{X}, \mathscr{X})$ and $C \in \mathscr{X}$. Assume that for all $x \in C$, $\mathbb{P}_x(\sigma_C < \infty) = 1$. Proposition 3.3.6 shows that $\mathbb{P}_x(\sigma_C^{(n)} < \infty) = 1$ for all $x \in C$ and $n \in \mathbb{N}$. We may then consider the process $\{\tilde{X}_n,\ n \in \mathbb{N}\}$ corresponding to the values of the Markov chain $\{X_n,\ n \in \mathbb{N}\}$ at the successive times of its returns to the set C. Theorem 3.3.8 shows that this process is again a Markov chain, called the induced chain on the set C.

Theorem 3.3.8. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $C \in \mathscr{X}$. Assume that $\mathbb{P}_x(\sigma_C < \infty) = 1$ for all $x \in C$. Then for all $x \in C$ and $n \in \mathbb{N}$, $\mathbb{P}_x(\sigma_C^{(n)} < \infty) = 1$. We set for all $n \in \mathbb{N}$,*

$$\tilde{X}_n = X_{\sigma_C^{(n)}} \mathbb{1}_{\{\sigma_C^{(n)} < \infty\}} + x_* \mathbb{1}_{\{\sigma_C^{(n)} = \infty\}}, \tag{3.3.7}$$

where x_* is an arbitrary element of C.

(i) For all $x \in C$, the process $\{\tilde{X}_n,\ n \in \mathbb{N}\}$ is under $\mathbb{P}_x$ a Markov chain on C with kernel Q_C (see Definition 3.3.7).

(ii) Let $A \subset C$ and denote by $\tilde{\sigma}_A$ the return time to the set A of the chain $\{\tilde{X}_n\}$. Then for all $x \in C$, $\mathbb{E}_x[\sigma_A] \le \mathbb{E}_x[\tilde{\sigma}_A] \sup_{y \in C} \mathbb{E}_y[\sigma_C]$.

Proof. By Proposition 3.3.6, we know that for all $n \in \mathbb{N}$ and $x \in C$, $\mathbb{P}_x(\sigma_C^{(n)} < \infty) = 1$.

(i) Let $x \in C$. Since $\mathbb{P}_x(\sigma_C^{(n)} < \infty) = 1$ for all $x \in C$ and $n \in \mathbb{N}$, the strong Markov property applied to the Markov chain $\{X_n\}$ yields, for all $B \in \mathscr{X}$,

$$\begin{aligned}\mathbb{P}_x\left(\tilde{X}_{n+1} \in B \,\middle|\, \mathscr{F}_{\sigma_C^{(n)}}\right) &= \mathbb{P}_x\left(X_{\sigma_C^{(n+1)}} \in B \,\middle|\, \mathscr{F}_{\sigma_C^{(n)}}\right) = \mathbb{P}_x\left(X_{\sigma_C} \circ \theta_{\sigma_C^{(n)}} \in B \,\middle|\, \mathscr{F}_{\sigma_C^{(n)}}\right) \\ &= \mathbb{P}_{X_{\sigma_C^{(n)}}}(X_{\sigma_C} \in B) = Q_C(\tilde{X}_n, B)\,.\end{aligned}$$

(ii) Since $A \subset C$, we have $\sigma_A = \sigma_C^{(\tilde{\sigma}_A)}$. Thus,

$$\sigma_A = \sum_{n=0}^{\tilde{\sigma}_A - 1} \{\sigma_C^{(n+1)} - \sigma_C^{(n)}\} = \sum_{n=0}^{\infty} \{\sigma_C^{(n+1)} - \sigma_C^{(n)}\} \mathbb{1}_{\{n < \tilde{\sigma}_A\}} = \sum_{n=0}^{\infty} \sigma_C \circ \theta_{\sigma_C^{(n)}} \mathbb{1}_{\{n < \tilde{\sigma}_A\}}\,.$$

Let $x \in C$. Note that $\{n < \tilde{\sigma}_A\} = \cap_{i=1}^{n} \{X_{\sigma^{(i)}} \notin A\} \in \mathscr{F}_{\sigma^{(n)}}$, and applying again Proposition 3.3.6, we have $\mathbb{P}_x(\sigma_C^{(n)} < \infty) = 1$. We then obtain by the strong Markov property,

$$\begin{aligned}\mathbb{E}_x[\sigma_A] &= \sum_{n=0}^{\infty} \mathbb{E}_x[\sigma_C \circ \theta_{\sigma_C^{(n)}} \mathbb{1}\{n < \tilde{\sigma}_A\}] \\ &= \sum_{n=0}^{\infty} \mathbb{E}_x[\mathbb{1}\{n < \tilde{\sigma}_A\} \mathbb{E}_{X_{\sigma_C^{(n)}}}[\sigma_C]] \le \mathbb{E}_x[\tilde{\sigma}_A] \sup_{y \in C} \mathbb{E}_y[\sigma_C]\,.\end{aligned}$$

□

3.4 First-Entrance, Last-Exit Decomposition

Given $A \in \mathscr{X}$, we define, for $n \ge 1$ and $B \in \mathscr{X}$,

$$ {}_A^nP(x,B) = \mathbb{P}_x(X_n \in B\,,\ n \le \sigma_A)\,. \tag{3.4.1}$$

Thus ${}_A^nP(x,B)$ is the probability that the chain goes from x to B in n steps without visiting the set A. It is called the n-step taboo probability. Note that ${}_A^1P = P$ and ${}_A^nP = (PI_{A^c})^{n-1}P$, where I_A is the kernel defined by $I_Af(x) = \mathbb{1}_A(x)f(x)$ for all $f \in \mathbb{F}_+(\mathsf{X})$.

Let $f \in \mathbb{F}_+(\mathsf{X})$ and $A \in \mathscr{X}$. For any given n, we may decompose $f(X_n)$ over the mutually exclusive events $\{\sigma_A \ge n\}$ and $\{\sigma_A = j\}$, $j \in \{1,\dots,n\}$. This yields the first entrance decomposition, which may be expressed with the taboo probabilities as follows, using the Markov property:

$$\begin{aligned} P^nf(x) = \mathbb{E}_x[f(X_n)] &= \mathbb{E}_x[\mathbb{1}\{n \le \sigma_A\}f(X_n)] + \sum_{j=1}^{n-1}\mathbb{E}_x[\mathbb{1}\{\sigma_A = j\}f(X_n)] \\ &= {}_A^nPf(x) + \sum_{j=1}^{n-1}\mathbb{E}_x\left[\mathbb{1}\{\sigma_A = j\}\mathbb{E}_{X_j}[f(X_{n-j})]\right] \\ &= {}_A^nPf(x) + \sum_{j=1}^{n-1}\mathbb{E}_x[\mathbb{1}\{\sigma_A \ge j\}\mathbb{1}_A(X_j)P^{n-j}f(X_j)] \\ &= {}_A^nPf(x) + \sum_{j=1}^{n-1}{}_A^jP(\mathbb{1}_A \times P^{n-j}f)(x)\,. \end{aligned} \tag{3.4.2}$$

The last exit decomposition is defined analogously:

$$\begin{aligned} P^nf(x) &= \mathbb{E}_x[f(X_n)] \\ &= \mathbb{E}_x[\mathbb{1}_{\{n\le\sigma_A\}}f(X_n)] + \sum_{j=1}^{n-1}\mathbb{E}_x[\mathbb{1}\{X_j \in A, X_{j+1} \notin A,\dots,X_{n-1} \notin A\}f(X_n)] \\ &= {}_A^nPf(x) + \sum_{j=1}^{n-1}\mathbb{E}_x[\mathbb{1}_A(X_j)\mathbb{E}_{X_j}[\mathbb{1}\{X_1 \notin A,\dots,X_{n-j-1} \notin A\}f(X_{n-j})]] \\ &= {}_A^nPf(x) + \sum_{j=1}^{n-1}\mathbb{E}_x[\mathbb{1}_A(X_j)\,{}_A^{n-j}Pf(X_j)] \\ &= {}_A^nPf(x) + \sum_{j=1}^{n-1}P^j(\mathbb{1}_A \times {}_A^{n-j}Pf)(x)\,. \end{aligned} \tag{3.4.3}$$

The first-entrance decomposition is clearly a decomposition that could be developed using the strong Markov property and the stopping time $\sigma_A \wedge n$. The last-exit decomposition, however, is not an example of the use of the strong Markov property: the last-exit time before n is not a stopping time. These decompositions do, however, illustrate the principles behind the (strong) Markov property, namely the decomposition of the probability space over the subevents on which the random time takes on the (countable) set of values.

Replacing P^j in the right-hand side of (3.4.3) by the expression obtained in (3.4.2) yields the so-called *first-entrance last-exit decomposition*:

$$\mathbb{E}_x[f(X_n)] = {}_A^nPf(x) + \sum_{1\le k\le j\le n-1} {}_A^kP[\mathbb{1}_A P^{j-k}(\mathbb{1}_A\ {}_A^{n-j}Pf)](x)\,. \tag{3.4.4}$$

The first-entrance last-exit formula is obtained by decomposing the probability space over the times of the first and last entrances to A prior to n. Taking $f = \mathbb{1}_B$ for $B \in \mathscr{X}$ in the previous relation leads to the following decomposition of $P^n(x,B)$:

$$P^n(x,B) = {}_A^nP(x,B) + \sum_{j=1}^{n-1} \int_A \left[\sum_{k=1}^{j} \int_A {}_A^kP(x,\mathrm{d}y)P^{j-k}(y,\mathrm{d}z)\right] {}_A^{n-j}P(z,B)\,. \tag{3.4.5}$$

3.5 Accessible and Attractive Sets

Definition 3.5.1 (Accessible set) *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$.*

(i) A set $A \in \mathscr{X}$ is said to be accessible if $\mathbb{P}_x(\sigma_A < \infty) > 0$ for all $x \in \mathsf{X}$.
(ii) The collection of accessible sets is denoted by $\mathscr{X}_P^+$.

The following lemma provides several equivalent characterizations of accessible sets.

Lemma 3.5.2 *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$. Let $A \in \mathscr{X}$. The following conditions are equivalent:*

(i) A is accessible.
(ii) For every $x \in \mathsf{X}$, there exists an integer $n \ge 1$ such that $P^n(x,A) > 0$.
(iii) For every $\mu \in \mathbb{M}_+(\mathscr{X})$, there exists an integer $n \ge 1$ such that $\mu P^n(A) > 0$.
(iv) For every $x \in A^c$, $\mathbb{P}_x(\sigma_A < \infty) > 0$.

Moreover, if A is accessible, then for all $a \in \mathbb{M}_+^1(\mathbb{N})$ with $a(k) > 0$ for $k \ge 1$, $K_a(x,A) > 0$ for all $x \in \mathsf{X}$. If there exists $a \in \mathbb{M}_+^1(\mathbb{N})$ such that $K_a(x,A) > 0$ for all $x \in \mathsf{X}$, then A is accessible.

Proof. The assertion (iv) $\Rightarrow$ (i) is the only nontrivial one. It means that if A can be reached from A^c, then it can be reached from A. Indeed, starting from A, either the chain remains in A, or it leaves A and then can reach it again. Formally, applying the Markov property yields

$$\begin{aligned}\mathbb{P}_x(\sigma_A < \infty) &= \mathbb{P}_x(X_1 \in A) + \mathbb{P}_x(X_1 \in A^c, \sigma_A \circ \theta < \infty) \\ &= \mathbb{P}_x(X_1 \in A) + \mathbb{E}_x[\mathbb{1}_{A^c}(X_1)\mathbb{P}_{X_1}(\sigma_A < \infty)]\,.\end{aligned}$$

For each $x \in \mathsf{X}$, either $\mathbb{P}_x(X_1 \in A) > 0$ or $\mathbb{P}_x(X_1 \in A) = 0$. In the latter case, one has $\mathbb{P}_x(\sigma_A < \infty) = \mathbb{E}_x[\mathbb{1}_{A^c}(X_1)\mathbb{P}_{X_1}(\sigma_A < \infty)] > 0$ if (iv) holds. Thus (iv) $\Rightarrow$ (i). $\square$

Definition 3.5.3 (Domain of attraction of a set, attractive set) *Let P be a Markov chain on $\mathsf{X} \times \mathscr{X}$. The domain of attraction C_+ of a nonempty set $C \in \mathscr{X}$ is the set of states $x \in \mathsf{X}$ from which the Markov chain returns to C with probability one:*

$$C_+ = \{x \in \mathsf{X} : \mathbb{P}_x(\sigma_C < \infty) = 1\} . \tag{3.5.1}$$

(i) If $C \subset C_+$, then the set C is said to be Harris recurrent.
(ii) If $C_+ = \mathsf{X}$, then the set C is said to be attractive.

If the domain of attraction C_+ of C contains C, then it may happen that $C_+ \subsetneq \mathsf{X}$. Nevertheless, as shown below, the set C_+ is absorbing.

Lemma 3.5.4 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let $C \in \mathscr{X}$ be a nonempty set such that $C \subset C_+$. Then the set C_+ is absorbing.*

Proof. Let $x \in C_+$. Then

$$\begin{aligned} 0 = \mathbb{P}_x(\sigma_C = \infty) &\geq \mathbb{P}_x(X_1 \in C^c , \sigma_C \circ \theta = \infty) \\ &\geq \mathbb{P}_x(X_1 \in C_+^c , \sigma_C \circ \theta = \infty) = \mathbb{E}_x[\mathbb{1}_{C_+^c}(X_1)\mathbb{P}_{X_1}(\sigma_C = \infty)] . \end{aligned}$$

Since $\mathbb{P}_y(\sigma_C = \infty) > 0$ for $y \in C_+^c$, this yields $P(x, C_+^c) = \mathbb{P}_x(X_1 \in C_+^c) = 0$. □

3.6 Return Times and Invariant Measures

Invariant and subinvariant measures were introduced in Section 1.4. Recall that a measure μ is subinvariant (resp. invariant) if μ is σ-finite and satisfies $\mu P \leq \mu$ (resp. $\mu P = \mu$). The next lemma gives a criterion for establishing that a measure satisfying $\mu P \leq \mu$ is σ-finite and hence subinvariant.

Lemma 3.6.1 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let $\mu \in \mathbb{M}_+(\mathscr{X})$ be such that $\mu P \leq \mu$. Assume that there exists an accessible set A such that $\mu(A) < \infty$. Then μ is σ-finite.*

Proof. Since $\mu P^k \leq \mu$ for all $k \in \mathbb{N}$, it is also the case that $\mu K_{a_\varepsilon} \leq \mu$. For every integer $m \geq 1$,

$$\begin{aligned} \infty > \mu(A) \geq \mu K_{a_\varepsilon}(A) &= \int \mu(\mathrm{d}x) K_{a_\varepsilon}(x, A) \\ &\geq m^{-1} \mu\left(\{x \in \mathsf{X} : K_{a_\varepsilon}(x, A) \geq 1/m\}\right) . \end{aligned}$$

Since A is accessible, the function $x \mapsto K_{a_\varepsilon}(x, A)$ is positive. Thus

$$\mathsf{X} = \bigcup_{m=1}^{\infty} \{x \in \mathsf{X} : K_{a_\varepsilon}(x, A) \geq 1/m\} .$$

This proves that μ is σ-finite. □

The next two theorems are the main results of this section. They provide expressions for an invariant measure in terms of the return time to a set C under certain conditions. These expressions will be used in later chapters to prove the existence and uniqueness of an invariant measure. For a measure $\mu \in \mathbb{M}_+(\mathscr{X})$ and $C \in \mathscr{X}$, we define the measures μ_C^0 and μ_C^1 by

$$\mu_C^0(B) = \int_C \mu(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} \mathbb{1}_B(X_k)\right] = \sum_{k=0}^{\infty}\int_C \mu(\mathrm{d}x)\mathbb{E}_x\left[\mathbb{1}\left\{k<\sigma_C\right\}\mathbb{1}_B(X_k)\right] , \tag{3.6.1}$$

$$\mu_C^1(B) = \int_C \mu(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_C} \mathbb{1}_B(X_k)\right] = \sum_{k=1}^{\infty}\int_C \mu(\mathrm{d}x)\mathbb{E}_x\left[\mathbb{1}\left\{k\le\sigma_C\right\}\mathbb{1}_B(X_k)\right] . \tag{3.6.2}$$

Lemma 3.6.2 *Let $C \in \mathscr{X}$ and $\mu \in \mathbb{M}_+(\mathscr{X})$. Then $\mu_C^1 = \mu_C^0 P$.*

Proof. For $B \in \mathscr{X}$, the Markov property implies

$$\begin{aligned}\mathbb{E}_x\left[\mathbb{1}\left\{k<\sigma_C\right\}P(X_k,B)\right] &= \mathbb{E}_x\left[\mathbb{1}\left\{k<\sigma_C\right\}\mathbb{E}_{X_k}\left[\mathbb{1}_B(X_1)\right]\right] \\ &= \mathbb{E}_x\left[\mathbb{1}\left\{k<\sigma_C\right\}\mathbb{E}_x\left[\mathbb{1}_B(X_1)\circ\theta_k \,|\, \mathscr{F}_k\right]\right] \\ &= \mathbb{E}_x\left[\mathbb{1}\left\{k+1\le\sigma_C\right\}\mathbb{1}_B(X_{k+1})\right] .\end{aligned}$$

Using this relation, we get

$$\begin{aligned}\mu_C^0 P(B) &= \sum_{k=0}^{\infty}\int_C \mu(\mathrm{d}x)\mathbb{E}_x\left[\mathbb{1}\left\{k<\sigma_C\right\}P(X_k,B)\right] \\ &= \sum_{k=0}^{\infty}\int_C \mu(\mathrm{d}x)\mathbb{E}_x\left[\mathbb{1}\left\{k+1\le\sigma_C\right\}\mathbb{1}_B(X_{k+1})\right] \\ &= \sum_{k=1}^{\infty}\int_C \mu(\mathrm{d}x)\mathbb{E}_x[\mathbb{1}\left\{k\le\sigma_C\right\}\mathbb{1}_B(X_k)] = \mu_C^1(B) .\end{aligned}$$

□

For $C \in \mathscr{X}$, recall that $\mathscr{X}_C$ denotes the induced σ-algebra, and $Q_C(x,B) = \mathbb{P}_x(\sigma_C < \infty, X_{\sigma_C} \in B)$ is the induced kernel.

Theorem 3.6.3. *Let $C \in \mathscr{X}$, let π_C be a probability measure on $\mathscr{X}_C$, and let π_C^0 be the measure on $\mathscr{X}$ defined, for $B \in \mathscr{X}$, by*

$$\pi_C^0(B) = \int_C \pi_C(\mathrm{d}x)\,\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1}\mathbb{1}_B(X_k)\right] . \tag{3.6.3}$$

Then the restriction of π_C^0 to the set C is π_C. Moreover, $\pi_C^0 = \pi_C^0 P$ if and only if $\pi_C = \pi_C Q_C$. If either of these properties holds, then $\mathbb{P}_x(\sigma_C < \infty) = 1 \quad \pi_C -$ a.s.

Proof. Replacing B by $B \cap C$ in (3.6.3) shows that $\pi_C^0(B \cap C) = \pi_C(B \cap C)$, which proves the first statement.

The identity $\pi_C^1 = \pi_C^0 P$ (see Lemma 3.6.2) implies

$$\begin{aligned}
\pi_C^0(B) + \pi_C Q_C(B \cap C) &= \pi_C^0(B) + \int_C \pi_C(\mathrm{d}x) \mathbb{E}_x[\mathbb{1}_B(X_{\sigma_C}) \mathbb{1}\{\sigma_C < \infty\}] \\
&= \pi_C(B \cap C) + \int_C \pi_C(\mathrm{d}x)\, \mathbb{E}_x \left[\sum_{k=1}^{\sigma_C} \mathbb{1}_B(X_k) \right] \\
&= \pi_C(B \cap C) + \pi_C^0 P(B) .
\end{aligned}$$

Since π_C is a probability measure on $\mathscr{X}_C$, if $\pi_C = \pi_C Q_C$, then $\pi_C^0 = \pi_C^0 P$. Conversely, assume that $\pi_C^0 = \pi_C^0 P$. Since $\pi_C^0(C) = \pi_C(C) = 1$, it follows that for all $B \in \mathscr{X}_C$, we have $\pi_C^0(B) \le \pi_C^0(C) = 1$, and therefore the relation

$$\pi_C^0(B) + \pi_C Q_C(B) = \pi_C(B) + \pi_C^0 P(B)$$

implies that $\pi_C Q_C = \pi_C$. Finally, if $\pi_C Q_C = \pi_C$, then

$$\pi_C(C) = \pi_C Q_C(C) = \int_C \pi_C(\mathrm{d}x) Q_C(x, C) = \int_C \pi_C(\mathrm{d}x) \mathbb{P}_x(\sigma_C < \infty) .$$

This implies that $\mathbb{P}_x(\sigma_C < \infty) = 1$ for π_C-almost all x. $\square$

Lemma 3.6.4 *Let μ be a P-subinvariant measure and $C \in \mathscr{X}$.*

(i) $\mu \ge \mu_C^0$ and $\mu \ge \mu_C^1$.
(ii) μ_C^0 and μ_C^1 are P-subinvariant if and only if $\mu(C) > 0$.
(iii) If $\mu\,|_C$ is Q_C-invariant, then $\mu_C^0 = \mu_C^1$, and both are P-invariant.
(iv) If μ is P-invariant and $\mu = \mu_C^0$, then $\mu = \mu_D^0 = \mu_D^1$ for every measurable set D that contains C.

Proof. (i) Recall that by definition, a P-subinvariant measure is σ-finite. Therefore, it suffices to prove that $\mu(B) \ge \mu_C^0(B)$ and $\mu(B) \ge \mu_C^1(B)$ for every $B \in \mathscr{X}$ satisfying $\mu(B) < \infty$. Let $B \in \mathscr{X}$ be such that $\mu(B) < \infty$. For every $k \ge 0$, define

$$u_{B,k}(x) = \mathbb{P}_x(X_k \in B, \sigma_C > k) .$$

Since $\{\sigma_C > k+1\} = \{\sigma_C \circ \theta > k\} \cap \{X_1 \notin C\}$, the Markov property yields

$$u_{B,k+1}(x) = \mathbb{P}_x(X_{k+1} \in B, \sigma_C > k+1) = \mathbb{E}_x[\mathbb{1}_{C^c}(X_1) u_{B,k}(X_1)] = P(u_{B,k} \mathbb{1}_{C^c})(x) .$$

Since μ is subinvariant, this yields $\mu(u_{B,k+1}) \le \mu(u_{B,k} \mathbb{1}_{C^c})$, with equality if μ is invariant. Note that

$$0 \le \mu(u_{B,k}) \le \mu(u_{B,0}) \le \mu(B) < \infty,$$

so that the difference $\mu(u_{B,k}) - \mu(u_{B,k+1})$ is well defined. This implies

$$\mu(u_{B,0}) - \mu(u_{B,n}) = \sum_{k=0}^{n-1} \{\mu(u_{B,k}) - \mu(u_{B,k+1})\} \ge \sum_{k=0}^{n-1} \mu(u_{B,k}\mathbb{1}_C) , \tag{3.6.4}$$

with equality if μ is invariant. This yields

$$\mu(B) = \mu(u_{B,0}) \ge \sum_{k=0}^{n-1} \mu(u_{B,k}\mathbb{1}_C) .$$

The series $\sum_{k=0}^{\infty} \mu(u_{B,k}\mathbb{1}_C)$ is therefore summable, and we have

$$\begin{aligned}
\mu(B) \ge \sum_{k=0}^{\infty} \mu(u_{B,k}\mathbb{1}_C) &= \sum_{k=0}^{\infty} \int_C \mu(\mathrm{d}x)\mathbb{P}_x(X_k \in B, \sigma_C > k) \\
&= \int_C \mu(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C - 1} \mathbb{1}_B(X_k)\right] = \mu_C^0(B) .
\end{aligned}$$

This proves that $\mu \ge \mu_C^0$, since μ is σ-finite. Since $\mu_C^1 = \mu_C^0 P$ and μ is subinvariant, this yields

$$\mu_C^1 = \mu_C^0 P \le \mu P \le \mu .$$

(ii) First note that (i) implies that μ_C^1 and μ_C^0 are σ-finite. By definition, we have

$$\mu_C^0(\mathsf{X}) = \mu_C^1(\mathsf{X}) = \int_C \mu(\mathrm{d}x)\mathbb{E}_x[\sigma_C] .$$

Thus μ_C^0 and μ_C^1 are nonzero if and only if $\mu(C) > 0$. Note now that for $k \ge 1$,

$$\mathbb{1}\{\sigma_C > k\} = \mathbb{1}_{C^c}(X_k)\mathbb{1}\{\sigma_C > k\} = \mathbb{1}_{C^c}(X_k)\mathbb{1}\{\sigma_C > k-1\} .$$

Since μ is subinvariant and $\mu \ge \mu_C^0$, this yields, for $f \in \mathbb{F}_+(\mathsf{X})$,

$$\begin{aligned}
\mu_C^0(f) &= \mu(f\mathbb{1}_C) + \sum_{k=1}^{\infty} \int_C \mu(\mathrm{d}x)\mathbb{E}_x[\mathbb{1}_{C^c}(X_k)f(X_k)\mathbb{1}\{\sigma_C > k\}] \\
&\ge \mu P(f\mathbb{1}_C) + \sum_{k=1}^{\infty} \int_C \mu(\mathrm{d}x)\mathbb{E}_x[\mathbb{1}_{C^c}(X_k)f(X_k)\mathbb{1}\{\sigma_C > k-1\}] \\
&\ge \mu_C^0 P(f\mathbb{1}_C) + \sum_{k=1}^{\infty} \int_C \mu(\mathrm{d}x)\mathbb{E}_x[P(f\mathbb{1}_{C^c})(X_{k-1})\mathbb{1}\{\sigma_C > k-1\}] \\
&= \mu_C^0 P(f\mathbb{1}_C) + \mu_C^0 P(f\mathbb{1}_{C^c}) = \mu_C^0 P(f) .
\end{aligned}$$

This proves that μ_C^0 is subinvariant. This in turn proves that μ_C^1 is subinvariant, since

$$\mu_C^1 P = (\mu_C^0 P)P \le \mu_C^0 P = \mu_C^1 .$$

(iii) Since $\mu_C^0 P(f \mathbb{1}_C) = \mu|_C Q_C$, if $\mu|_C Q_C = \mu|_C$, then all the inequalities above becomes equalities, and this yields $\mu_C^0 = \mu_C^0 P$ i.e., μ_C^0 is P-invariant. Since $\mu_C^1 = \mu_C^0 P$, this implies that $\mu_C^1 = \mu_C^0$.

(iv) If μ is invariant, starting from (3.6.4), we have, for all $n \ge 1$,

$$\mu(B) = \sum_{k=0}^{n-1} \mu(u_{B,k} \mathbb{1}_C) + \mu(u_{B,n}) .$$

We already know that the series is convergent; thus $\lim_{n\to\infty} \mu(u_{B,n})$ also exists, and this proves

$$\mu(B) = \int_C \mu(\mathrm{d}x) \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} \mathbb{1}_B(X_k) \right] + \lim_{n\to\infty} \int_{\mathsf{X}} \mu(\mathrm{d}x) \mathbb{P}_x(X_n \in B, \sigma_C > n) . \quad (3.6.5)$$

The identity (3.6.5) implies that $\mu = \mu_C^0$ if and only if

$$\lim_{n\to\infty} \int_{\mathsf{X}} \mu(\mathrm{d}x) \mathbb{P}_x(X_n \in B, \sigma_C > n) = 0 .$$

If $D \supset C$, then $\sigma_D \le \sigma_C$, and it is also the case that

$$\lim_{n\to\infty} \int_{\mathsf{X}} \mu(\mathrm{d}x) \mathbb{P}_x(X_n \in B, \sigma_D > n) = 0 .$$

Applying (3.6.5) with D instead of C then proves that $\mu(B) = \mu_D^0(B)$ for all B such that $\mu(B) < \infty$, and since μ is σ-finite, this proves that $\mu = \mu_D^0$. Since μ is invariant and $\mu_D^1 = \mu_D^0$, this also proves that $\mu = \mu_D^1$.

□

Theorem 3.6.5. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ that admits a subinvariant measure μ and let $C \in \mathscr{X}$ be such that $\mu(C) < \infty$ and $\mathbb{P}_x(\sigma_C < \infty) > 0$ for μ-almost all $x \in \mathsf{X}$. Then the following statements are equivalent:*

(i) $\mathbb{P}_x(\sigma_C < \infty) = 1$ for μ-almost all $x \in C$;
(ii) the restriction of μ to C is invariant with respect to Q_C;
(iii) for all $f \in \mathbb{F}_+(\mathsf{X})$,

$$\mu(f) = \int_C \mu(\mathrm{d}x) \mathbb{E}_x \left[\sum_{k=1}^{\sigma_C} f(X_k) \right] . \quad (3.6.6)$$

If any of these properties is satisfied, then μ is invariant, and for all $f \in \mathbb{F}_+(\mathsf{X})$,

$$\mu(f) = \int_C \mu(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C - 1} f(X_k)\right] . \tag{3.6.7}$$

Proof. First assume that (3.6.6) holds. Then, taking $f = \mathbb{1}_C$, we get

$$\mu(C) = \int_C \mu(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_C} \mathbb{1}_C(X_k)\right] = \int_C \mu(\mathrm{d}x)\mathbb{P}_x(\sigma_C < \infty) .$$

Since $\mu(C) < \infty$, this implies that $\mathbb{P}_x(\sigma_C < \infty) = 1$ for μ-almost all $x \in C$. This proves that (iii) implies (i).

Assume now that $\mathbb{P}_x(\sigma_C < \infty) = 1$ for μ-almost all $x \in C$. Define the measure μ_C^1 by

$$\mu_C^1(A) = \int_C \mu(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_C} \mathbb{1}_A(X_k)\right] .$$

Since $\mathbb{P}_x(\sigma_C < \infty) = 1$ for μ-almost all $x \in C$ by assumption, we have

$$\mu_C^1(C) = \int_C \mu(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_C} \mathbb{1}_C(X_k)\right] = \int_C \mu(\mathrm{d}x)\mathbb{P}_x(\sigma_C < \infty) = \mu(C) < \infty .$$

By Lemma 3.6.4 (i), $\mu \geq \mu_C^1$ and $\mu(C) = \mu_C^1(C) < \infty$; thus the respective restrictions to C of the measures μ and μ_C^1 must be equal. That is, for every $A \in \mathscr{X}$, $\mu(A \cap C) = \mu_C^1(A \cap C)$. This yields

$$\begin{aligned}\mu(A \cap C) = \mu_C^1(A \cap C) &= \int_C \mu(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_C} \mathbb{1}_{A\cap C}(X_k)\right] \\ &= \int_C \mu(\mathrm{d}x)\mathbb{P}_x(X_{\sigma_C} \in A) = \mu\,|_C\, Q_C(A \cap C) .\end{aligned}$$

This proves that the restriction of μ to C is invariant for Q_C. Thus (i) implies (ii).

Assume now that (ii) holds. Then Theorem 3.6.3 yields $\mu_C^1 = \mu_C^1 P$. The final step is to prove that $\mu = \mu_C^1$. For every $\varepsilon > 0$, μ is subinvariant and μ_C^1 is invariant with respect to the resolvent kernel K_{a_ε}. Let g be the measurable function defined on X by $g(x) = K_{a_\varepsilon}(x, C)$. Moreover,

$$\mu(g) = \mu K_{a_\varepsilon}(C) \leq \mu(C) = \mu_C^1(C) = \mu_C^1 K_{a_\varepsilon}(C) = \mu_C^1(g) .$$

Since it is also the case that $\mu \geq \mu_C^1$ and $\mu(C) < \infty$, this implies $\mu(g) = \mu_C^1(g)$, i.e., the measures $g \cdot \mu$ and $g \cdot \mu_C^1$ coincide. Since $g(x) > 0$ for μ-almost all $x \in \mathsf{X}$ and also for μ_C^1-almost all $x \in \mathsf{X}$ since $\mu \geq \mu_C^1$, this yields $\mu = \mu_C^1$. This proves (iii). The proof is completed by applying (3.6.6) combined with Lemma 3.6.4 (iii). $\square$

3.7 Exercises

3.1. Let $(\Omega, \mathscr{F}, \{\mathscr{F}_k, k \in \mathbb{N}\}, \mathbb{P})$ be a filtered probability space and let τ and σ be two stopping times for the filtration $\{\mathscr{F}_k, k \in \mathbb{N}\}$. Denote by $\mathscr{F}_\tau$ and $\mathscr{F}_\sigma$ the σ-fields of the events prior to τ and σ, respectively. Then

(i) $\tau \wedge \sigma$, $\tau \vee \sigma$, and $\tau + \sigma$ are stopping times;
(ii) if $\tau \le \sigma$, then $\mathscr{F}_\tau \subset \mathscr{F}_\sigma$;
(iii) $\mathscr{F}_{\tau \wedge \sigma} = \mathscr{F}_\tau \cap \mathscr{F}_\sigma$;
(iv) $\{\tau < \sigma\} \in \mathscr{F}_\tau \cap \mathscr{F}_\sigma$, $\{\tau = \sigma\} \in \mathscr{F}_\tau \cap \mathscr{F}_\sigma$.

3.2. Let $C \in \mathscr{X}$.

1. Assume that $\sup_{x\in C} \mathbb{E}_x[\sigma_C] < \infty$. Show that
$$\sup_{x\in C} \mathbb{E}_x[\sigma_C^{(n)}] \le n \sup_{x\in C} \mathbb{E}_x[\sigma_C] .$$
2. Let $p \ge 1$. Assume that $\sup_{x\in C} \mathbb{E}_x[\{\sigma_C\}^p] < \infty$. Show that
$$\sup_{x\in C} \mathbb{E}_x[\{\sigma_C^{(n)}\}^p] \le K(n,p) \sup_{x\in C} \mathbb{E}_x[\{\sigma_C\}^p]$$
for a constant $K(n,p) < \infty$.

3.3. For $A \in \mathscr{X}$, define by I_A, the multiplication operator by $\mathbb{1}_A$, $I_A f(x) = \mathbb{1}_A(x) f(x)$ for all $x \in \mathsf{X}$ and $f \in \mathbb{F}_+(\mathsf{X})$. Let $C \in \mathscr{X}$. Show that the induced kernel Q_C (see (3.3.7)) can be written as
$$Q_C = \sum_{n=0}^{\infty} (I_{C^c} P)^n I_C .$$

3.4. Let $A \in \mathscr{X}$.

1. For $x \in \mathsf{X}$, set $f(x) = \mathbb{P}_x(\tau_A < \infty)$. Show that for $x \in A^c$, $Ph(x) = h(x)$.
2. For $x \in \mathsf{X}$, set $f(x) = \mathbb{P}_x(\tau_A < \infty)$. Show that $Ph(x) \le h(x)$ for all $x \in \mathsf{X}$.

3.5. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let σ be a stopping time. Show that for every $A \in \mathscr{X}$ and $n \in \mathbb{N}$,
$$P^n(x,A) = \mathbb{E}_x[\mathbb{1}\{n \le \sigma\} \mathbb{1}_A(X_n)] + \mathbb{E}_x[\mathbb{1}\{\sigma < n\} P^{n-\sigma}(X_\sigma, A)] .$$

3.6. Let π be a P-invariant probability measure and let $C \in \mathsf{X}$ be such that $\mathbb{P}_x(\sigma_C < \infty) = 1$ for π-almost all $x \in \mathsf{X}$. Then for all $B \in \mathscr{X}$,
$$\pi(B) = \int_C \pi(\mathrm{d}x) \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C - 1} \mathbb{1}_B(X_k)\right] = \int_C \pi(\mathrm{d}x) \mathbb{E}_x\left[\sum_{k=1}^{\sigma_C} \mathbb{1}_B(X_k)\right] .$$

[Hint: Apply Lemma 3.6.4 to π and note that the limit in (3.6.5) is zero by Lebesgue's dominated convergence theorem.]

3.7. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an invariant probability measure π. Let $r = \{r(n),\ n \in \mathbb{N}\}$ be a positive sequence, $C \in \mathscr{X}$ an accessible set, and $f \in \mathbb{F}_+(\mathsf{X})$ a function. Define

$$C_+(r, f) = \left\{ x \in \mathsf{X} :\ \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} r(k) f(X_k) \right] < \infty \right\} . \tag{3.7.1}$$

Assume $\sup_{n\in\mathbb{N}} r(n)/r(n+1) < \infty$ and $C \subset C_+(r,f)$. Set $U = \sum_{k=0}^{\sigma_C - 1} r(k) f(X_k)$ and define $M = \sup_{n\in\mathbb{N}} r(n)/r(n+1) < \infty$.

1. Show that

$$\mathbb{1}_{C^c}(X_1) U \circ \theta \le M \mathbb{1}_{C^c}(X_1) U .$$

2. Show that

$$\mathbb{P}_x(\mathbb{1}_{C^c}(X_1)\mathbb{E}_{X_1}[U] < \infty) = 1 . \tag{3.7.2}$$

3. Show that $C_+(r, f)$ is accessible and absorbing, and that $\pi(C_+(r, f)) = 1$.

3.8. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an invariant probability measure π. Let $C \in \mathscr{X}$ be an accessible and absorbing set. Let $\varepsilon \in (0,1)$ and denote by K_{a_ε} the resolvent kernel given in Definition 1.2.10.

1. Show that $\int_C \pi(\mathrm{d}x) K_{a_\varepsilon}(x, C) = 1$ and $\int_{C^c} \pi(\mathrm{d}x) K_{a_\varepsilon}(x, C) = 0$.
2. Show that $\pi(C) = 1$.

3.8 Bibliographical Notes

The first-entrance last-exit decomposition is an essential tool that has been introduced and exploited in many different ways in Chung (1953, 1967).

Chapter 4
Martingales, Harmonic Functions and Poisson–Dirichlet Problems

In this chapter, we introduce several notions of potential theory for Markov chains. Harmonic and superharmonic functions on a set A are defined in Section 4.1, and Theorem 4.1.3 establishes links between these functions and the return (or hitting) times to the set A. In Section 4.2, we introduce the potential kernel and prove the maximum principle, Theorem 4.2.2, which will be very important in the study of recurrence and transience throughout Part II. In Section 4.3, we will state and prove a very simple but powerful result: the comparison theorem, Theorem 4.3.1. It will turn out to be the essential ingredient for turning drift conditions into bounds on moments of hitting times, the first example of such a use being given in Proposition 4.3.2. The Poisson and Dirichlet problems are introduced in Section 4.4, and solutions to these problems are given. The problems are boundary problems for the operator $I - P$, and their solutions are expressed in terms of the hitting time of the boundary. We then combine these problems into the Poisson–Dirichlet problem and provide in Theorem 4.4.5 a minimal solution. The Poisson–Dirichlet problem can be viewed as a potential-theoretic formulation of a drift-type condition.

4.1 Harmonic and Superharmonic Functions

We have seen that subinvariant and invariant measures, i.e., σ-finite measures $\lambda \in \mathbb{M}_+(\mathscr{X})$ satisfying $\lambda P \leq \lambda$ or $\lambda P = \lambda$, play a key role in the theory of Markov chains. Also of central importance are functions $f \in \mathbb{F}_+(\mathsf{X})$ that satisfy $Pf \leq f$ or $Pf = f$ outside a set A.

Definition 4.1.1 (Harmonic and superharmonic functions) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $A \in \mathscr{X}$.*

- *A function $f \in \mathbb{F}_+(\mathsf{X})$ is called superharmonic on A if $Pf(x) \leq f(x)$ for all $x \in A$.*

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_4

- *A function* $f \in \mathbb{F}_+(\mathsf{X}) \cup \mathbb{F}_b(\mathsf{X})$ *is called harmonic on* A *if* $Pf(x) = f(x)$ *for all* $x \in A$.

If $A = \mathsf{X}$ *and the function* f *satisfies one of the previous conditions, it is called simply superharmonic or harmonic.*

The following result shows that superharmonic and harmonic functions have deep connections with supermartingales and martingales. Together with classical limit theorems for martingales, this connection will provide relatively easy proofs for some nontrivial results.

Theorem 4.1.2. *Let* P *be a Markov kernel on* $\mathsf{X} \times \mathscr{X}$ *and* $A \in \mathscr{X}$.

(i) A function $f \in \mathbb{F}_+(\mathsf{X})$ *is superharmonic on* A^c *if and only if for all* $\xi \in \mathbb{M}_1(\mathscr{X})$, $\{f(X_{n\wedge\tau_A}),\ n \in \mathbb{N}\}$ *is a positive* $\mathbb{P}_\xi$*-supermartingale.*
(ii) A function $h \in \mathbb{F}_+(\mathsf{X}) \cup \mathbb{F}_b(\mathsf{X})$ *is harmonic on* A^c *if and only if for all* $\xi \in \mathbb{M}_1(\mathscr{X})$, $\{h(X_{n\wedge\tau_A}),\ n \in \mathbb{N}\}$ *is a* $\mathbb{P}_\xi$*-martingale.*

Proof. Set $M_n = f(X_{n\wedge\tau_A})$. Since τ_A is a stopping time, for every $n \in \mathbb{N}$,

$$f(X_{\tau_A})\,\mathbb{1}\{\tau_A \le n\} \quad \text{is}\ \mathscr{F}_n - \text{measurable.}$$

Assume first that f is superharmonic on A^c. Then for $\xi \in \mathbb{M}_1(\mathscr{X})$, we have $\mathbb{P}_\xi$ − a.s.,

$$\begin{aligned}
\mathbb{E}_\xi\left[M_{n+1}\middle|\mathscr{F}_n\right] &= \mathbb{E}_\xi\left[M_{n+1}(\mathbb{1}\{\tau_A \le n\} + \mathbb{1}\{\tau_A > n\})\middle|\mathscr{F}_n\right] \\
&= f(X_{\tau_A})\mathbb{1}\{\tau_A \le n\} + \mathbb{1}\{\tau_A > n\}\,\mathbb{E}_\xi\left[f(X_{n+1})\middle|\mathscr{F}_n\right] \\
&= f(X_{\tau_A})\mathbb{1}\{\tau_A \le n\} + \mathbb{1}\{\tau_A > n\}\,Pf(X_n)\,.
\end{aligned}$$

By assumption, f is superharmonic on A^c; moreover, if $\tau_A > n$, then $X_n \in A^c$. This implies that $Pf(X_n) \le f(X_n)$ on $\{\tau_A > n\}$. Therefore,

$$\mathbb{E}_\xi\left[M_{n+1}\middle|\mathscr{F}_n\right] \le f(X_{\tau_A})\mathbb{1}\{\tau_A \le n\} + \mathbb{1}\{\tau_A > n\}\,f(X_n) = f(X_{n\wedge\tau_A}) = M_n\,.$$

Thus $\{(M_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ is a positive $\mathbb{P}_\xi$-supermartingale.

Conversely, assume that for every $\xi \in \mathbb{M}_+(\mathscr{X})$, $\{(M_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ is a positive $\mathbb{P}_\xi$-supermartingale. If $x \in A^c$, then $\tau_A \ge 1 \quad \mathbb{P}_x$ − a.s. Therefore, for all $x \in A^c$,

$$f(x) \ge \mathbb{E}_x\left[f(X_{1\wedge\tau_A})\middle|\mathscr{F}_0\right] = \mathbb{E}_x\left[f(X_1)\middle|\mathscr{F}_0\right] = Pf(x)\,.$$

The case of a harmonic function is dealt with by replacing inequalities by equalities in the previous derivations. □

Theorem 4.1.3. *Let P be Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $A \in \mathscr{X}$. Then*

(i) the function $x \mapsto \mathbb{P}_x(\tau_A < \infty)$ is harmonic on A^c;
(ii) the function $x \mapsto \mathbb{P}_x(\sigma_A < \infty)$ is superharmonic.

Proof. (i) Define $f(x) = \mathbb{P}_x(\tau_A < \infty)$ and note that

$$Pf(x) = \mathbb{E}_x[f(X_1)] = \mathbb{E}_x[\mathbb{P}_{X_1}(\tau_A < \infty)] \, .$$

Using the relation $\sigma_A = 1 + \tau_A \circ \theta$ and applying the Markov property, we get

$$Pf(x) = \mathbb{E}_x[\mathbb{P}_x(\, \tau_A \circ \theta < \infty \,|\, \mathscr{F}_1)] = \mathbb{P}_x(\tau_A \circ \theta < \infty) = \mathbb{P}_x(\sigma_A < \infty) \, .$$

If $x \in A^c$, then $\mathbb{P}_x(\sigma_A < \infty) = \mathbb{P}_x(\tau_A < \infty)$, and hence $Pf(x) = f(x)$.

(ii) Define $g(x) = \mathbb{P}_x(\sigma_A < \infty)$. Along the same lines, we obtain

$$Pg(x) = \mathbb{E}_x[g(X_1)] = \mathbb{E}_x[\mathbb{P}_{X_1}(\sigma_A < \infty)] = \mathbb{P}_x(\sigma_A \circ \theta < \infty) \, .$$

Since $\{\sigma_A \circ \theta < \infty\} \subset \{\sigma_A < \infty\}$, the previous relation implies that $Pg(x) \leq g(x)$ for all $x \in \mathsf{X}$.

□

4.2 The Potential Kernel

Definition 4.2.1 (Number of visits, Potential kernel) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$.*

(i) The number of visits N_A to a set $A \in \mathscr{X}$ is defined by

$$N_A = \sum_{k=0}^{\infty} \mathbb{1}_A(X_k) \, . \tag{4.2.1}$$

(ii) For $x \in \mathsf{X}$ and $A \in \mathscr{X}$, the expected number $U(x,A)$ of visits to A starting from x is defined by

$$U(x,A) = \mathbb{E}_x[N_A] = \sum_{k=0}^{\infty} P^k(x,A) \, . \tag{4.2.2}$$

The kernel U is called the potential kernel associated to P.

For each $x \in \mathsf{X}$, the function $U(x,\cdot)$ defines a measure on $\mathscr{X}$ that is not necessarily σ-finite and can even be identically infinite.

It is easily seen that the potential kernel can be expressed in terms of the successive return times:

$$U(x,A) = \mathbb{1}_A(x) + \sum_{n=1}^{\infty} \mathbb{P}_x(\sigma_A^{(n)} < \infty) . \tag{4.2.3}$$

It is therefore natural to try to bound the expected number of visits to a set A when the chain starts from an arbitrary point in the space by the probability of hitting the set and the expected number of visits to the set when the chain starts within the set. This is done rigorously in the next result, referred to as the maximum principle, whose name comes from harmonic analysis.

Theorem 4.2.2 (Maximum principle). *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. For all $x \in \mathsf{X}$ and $A \in \mathscr{X}$,*

$$U(x,A) \le \mathbb{P}_x(\tau_A < \infty) \sup_{y \in A} U(y,A) .$$

Proof. By the strong Markov property, we get

$$\begin{aligned} U(x,A) &= \mathbb{E}_x\left[\sum_{n=0}^{\infty} \mathbb{1}_A(X_n)\right] = \mathbb{E}_x\left[\sum_{n=\tau_A}^{\infty} \mathbb{1}_A(X_n)\mathbb{1}\{\tau_A < \infty\}\right] \\ &= \sum_{n=0}^{\infty} \mathbb{E}_x\left[\mathbb{1}_A(X_n \circ \theta_{\tau_A})\mathbb{1}\{\tau_A < \infty\}\right] \\ &= \sum_{n=0}^{\infty} \mathbb{E}_x\left[\mathbb{1}\{\tau_A < \infty\}\mathbb{E}_{X_{\tau_A}}\left[\mathbb{1}_A(X_n)\right]\right] \le \mathbb{P}_x(\tau_A < \infty) \sup_{y \in A} U(y,A) . \end{aligned}$$

□

We state here another elementary property of the potential kernel as a lemma for further reference.

Lemma 4.2.3 *For every sampling distribution a on $\mathbb{N}$, $UK_a = K_aU \le U$.*

Proof. By definition, for all $x \in \mathsf{X}$ and $A \in \mathscr{X}$,

$$UP^k(x,A) = P^kU(x,A) = \sum_{n=0}^{\infty} P^{k+n}(x,A) \le U(x,A) .$$

For every distribution a on $\mathbb{N}$, this yields

$$K_aU(x,A) = \sum_{k=0}^{\infty} a(k)P^kU(x,A) \le \sum_{k=0}^{\infty} a(k)U(x,A) = U(x,A) .$$

□

Proposition 4.2.4 *For every $A \in \mathscr{X}$, the function $x \mapsto \mathbb{P}_x(N_A = \infty)$ is harmonic.*

Proof. Define $h(x) = \mathbb{P}_x(N_A = \infty)$. Then $Ph(x) = \mathbb{E}_x[h(X_1)] = \mathbb{E}_x[\mathbb{P}_{X_1}(N_A = \infty)]$, and applying the Markov property, we obtain

$$Ph(x) = \mathbb{E}_x[\mathbb{P}_x(N_A \circ \theta = \infty \,|\, \mathscr{F}_1)] = \mathbb{P}_x(N_A \circ \theta = \infty) = \mathbb{P}_x(N_A = \infty) = h(x) .$$

□

The following result is a first approach to the classification of the sets of a Markov chain. Let $A \in \mathscr{X}$. Assume first that $\sup_{x\in A} \mathbb{P}_x(\sigma_A < \infty) = \delta < 1$. We will then show that the probability of returning infinitely often to A is equal to zero and that the expected number of visits to A is finite. We will later call such sets uniformly transient. If, on the contrary, we assume that for all $x \in A$, $\mathbb{P}_x(\sigma_A < \infty)$, i.e., if with probability 1 a chain begun at $x \in A$ returns to A, then we will show that the chain begun from any $x \in A$ returns to A infinitely often with probability 1, and of course the expectation of the number of visits to A is infinite. Such sets will later be called recurrent.

Proposition 4.2.5 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let $A \in \mathscr{X}$.*

(i) Assume that there exists $\delta \in [0,1)$ such that $\sup_{x\in A} \mathbb{P}_x(\sigma_A < \infty) \le \delta$. Then for all $p \in \mathbb{N}^*$, one has $\sup_{x\in A} \mathbb{P}_x(\sigma_A^{(p)} < \infty) \le \delta^p$ and $\sup_{x\in \mathsf{X}} \mathbb{P}_x(\sigma_A^{(p)} < \infty) \le \delta^{p-1}$. Moreover,

$$\sup_{x\in\mathsf{X}} U(x,A) \le (1-\delta)^{-1} . \tag{4.2.4}$$

(ii) Assume that $\mathbb{P}_x(\sigma_A < \infty) = 1$ for all $x \in A$. Then for all $p \in \mathbb{N}^*$, one has $\inf_{x\in A} \mathbb{P}_x(\sigma_A^{(p)} < \infty) = 1$. Moreover, $\inf_{x\in A} \mathbb{P}_x(N_A = \infty) = 1$ for all $x \in A$.

Proof. (i) For $p \in \mathbb{N}$, one has $\sigma_A^{(p+1)} = \sigma_A^{(p)} + \sigma_A \circ \theta_{\sigma_A^{(p)}}$ on $\{\sigma_A^{(p)} < \infty\}$. Applying the strong Markov property yields

$$\begin{aligned}\mathbb{P}_x(\sigma_A^{(p+1)} < \infty) &= \mathbb{P}_x\left(\sigma_A^{(p)} < \infty, \sigma_A \circ \theta_{\sigma_A^{(p)}} < \infty\right) \\ &= \mathbb{E}_x\left[\mathbb{1}\left\{\sigma_A^{(p)} < \infty\right\} \mathbb{P}_{X_{\sigma_A^{(p)}}}(\sigma_A < \infty)\right] \le \delta \mathbb{P}_x(\sigma_A^{(p)} < \infty) .\end{aligned}$$

By induction, we obtain $\mathbb{P}_x(\sigma_A^{(p)} < \infty) \le \delta^p$ for every $p \in \mathbb{N}^*$ and $x \in A$. Thus for $x \in A$,

$$U(x,A) = \mathbb{E}_x[N_A] \leq 1 + \sum_{p=1}^{\infty} \mathbb{P}_x(\sigma_A^{(p)} < \infty) \leq (1-\delta)^{-1} .$$

Since by Theorem 4.2.2, one has that for all $x \in \mathsf{X}$, $U(x,A) \leq \sup_{y \in A} U(y,A)$, (4.2.4) follows.

(ii) By Proposition 3.3.6, $\mathbb{P}_x(\sigma_A^{(n)} < \infty) = 1$ for every $n \in \mathbb{N}$ and $x \in A$. Then

$$\mathbb{P}_x(N_A = \infty) = \mathbb{P}_x\left(\bigcap_{n=1}^{\infty} \{\sigma_A^{(n)} < \infty\}\right) = 1 .$$

□

Given $A, B \in \mathscr{X}$, it is of interest to give a condition ensuring that the number of visits to B will be infinite whenever the number of visits to A is infinite. The next result shows that this is true if A leads uniformly to B, i.e., the probability of returning to B from any $x \in A$ is bounded away from zero. The proof of this result uses the supermartingale convergence theorem.

Theorem 4.2.6. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let $A, B \in \mathscr{X}$ be such that* $\inf_{x \in A} \mathbb{P}_x(\sigma_B < \infty) > 0$. *For all $\xi \in \mathbb{M}_1(\mathscr{X})$,* $\{N_A = \infty\} \subset \{N_B = \infty\}$ $\mathbb{P}_\xi$ − a.s.

Proof. Let $\xi \in \mathbb{M}_1(\mathscr{X})$ and set $\delta = \inf_{x \in A} \mathbb{P}_x(\sigma_B < \infty)$. Since $\delta > 0$ by assumption, we have $\{N_A = \infty\} \subset \{\mathbb{P}_{X_n}(\sigma_B < \infty) \geq \delta \quad \text{i.o.}\}$. We will show that

$$\lim_{n \to \infty} \mathbb{P}_{X_n}(\sigma_B < \infty) = \mathbb{1}\{N_B = \infty\} \qquad \mathbb{P}_\xi - \text{a.s.} \tag{4.2.5}$$

Therefore, on the event $\{\mathbb{P}_{X_n}(\sigma_B < \infty) \geq \delta \quad \text{i.o.}\}$, we get $\lim_{n \to \infty} \mathbb{P}_{X_n}(\sigma_B < \infty) = 1$, showing that

$$\{N_A = \infty\} \subset \{\mathbb{P}_{X_n}(\sigma_B < \infty) \geq \delta \quad \text{i.o.}\} \subset \{N_B = \infty\} \;\; \mathbb{P}_\xi - \text{a.s.}$$

Let us now prove (4.2.5). By Theorem 4.1.3 (ii), the function $x \mapsto \mathbb{P}_x(\sigma_B < \infty)$ is superharmonic, and hence $\{\mathbb{P}_{X_n}(\sigma_B < \infty) : n \in \mathbb{N}\}$ is a bounded nonnegative supermartingale. By the supermartingale convergence theorem (Proposition E.1.3), the sequence $\{\mathbb{P}_{X_n}(\sigma_B < \infty), n \in \mathbb{N}\}$ converges $\mathbb{P}_\xi$ − a.s. and in $\mathrm{L}^1(\mathbb{P}_\xi)$. Thus, for every integer $p \in \mathbb{N}^*$ and $F \in \mathscr{F}_p$, we have by Lebesgue's dominated convergence theorem (considering only $n \geq p$),

$$\begin{aligned}
\mathbb{E}_\xi[\mathbb{1}_F \lim_{n \to \infty} \mathbb{P}_{X_n}(\sigma_B < \infty)] &= \lim_{n \to \infty} \mathbb{E}_\xi\left[\mathbb{1}_F \mathbb{P}_{X_n}(\sigma_B < \infty)\right] \\
&= \lim_{n \to \infty} \mathbb{E}_\xi\left[\mathbb{1}_F \mathbb{P}_\xi\left(\sigma_B \circ \theta_n < \infty \mid \mathscr{F}_n\right)\right] \\
&= \lim_{n \to \infty} \mathbb{P}_\xi(F \cap \{\sigma_B \circ \theta_n < \infty\}) .
\end{aligned}$$

Since

$$\{\sigma_B \circ \theta_n < \infty\} = \bigcup_{k>n} \{X_k \in B\} \downarrow_n \{X_n \in B \quad \text{i.o.}\} = \{N_B = \infty\} ,$$

Lebesgue's dominated convergence theorem implies

$$\mathbb{E}_\xi \left[\mathbb{1}_F \lim_{n\to\infty} \mathbb{P}_{X_n}(\sigma_B < \infty)\right] = \mathbb{P}_\xi(F \cap \{N_B = \infty\}) .$$

Since the above identity holds for every integer p and $F \in \mathscr{F}_p$, this proves (4.2.5). □

4.3 The Comparison Theorem

The general result below will be referred to as the comparison theorem. It is expressed in terms of a general stopping time τ, without specifying the nature of this stopping time, even though when it comes to applying this theorem, the stopping time is usually the hitting or the return time to a set C. It might be seen as a generalization of the optional stopping theorem for positive supermartingales. By convention, we set $\sum_{k=0}^{-1} = 0$.

Theorem 4.3.1 (Comparison Theorem). *Let $\{V_n,\ n \in \mathbb{N}\}$, $\{Y_n,\ n \in \mathbb{N}\}$, and $\{Z_n,\ n \in \mathbb{N}\}$ be three $\{\mathscr{F}_n,\ n \in \mathbb{N}\}$-adapted nonnegative processes such that for all $n \in \mathbb{N}$,*

$$\mathbb{E}\left[V_{n+1} \,|\, \mathscr{F}_n\right] + Z_n \le V_n + Y_n \qquad \mathbb{P} - \text{a.s.} \tag{4.3.1}$$

Then for every $\{\mathscr{F}_n,\ n \in \mathbb{N}\}$-stopping time τ,

$$\mathbb{E}\left[V_\tau \mathbb{1}\{\tau < \infty\}\right] + \mathbb{E}\left[\sum_{k=0}^{\tau-1} Z_k\right] \le \mathbb{E}\left[V_0\right] + \mathbb{E}\left[\sum_{k=0}^{\tau-1} Y_k\right] . \tag{4.3.2}$$

Proof. Let us prove by induction that for all $n \ge 0$,

$$\mathbb{E}\left[V_n\right] + \mathbb{E}\left[\sum_{k=0}^{n-1} Z_k\right] \le \mathbb{E}\left[V_0\right] + \mathbb{E}\left[\sum_{k=0}^{n-1} Y_k\right] . \tag{4.3.3}$$

The property is true for $n = 0$ (due to the above-mentioned convention). Assume that it is true for some $n \ge 0$. Then, applying (4.3.1) and the induction assumption, we obtain

$$
\begin{aligned}
\mathbb{E}[V_{n+1}] + \mathbb{E}\left[\sum_{k=0}^{n} Z_k\right] &= \mathbb{E}\left[\mathbb{E}[V_{n+1} \,|\, \mathscr{F}_n] + Z_n\right] + \mathbb{E}\left[\sum_{k=0}^{n-1} Z_k\right] \\
&\le \mathbb{E}[V_n + Y_n] + \mathbb{E}\left[\sum_{k=0}^{n-1} Z_k\right] = \mathbb{E}[V_n] + \mathbb{E}\left[\sum_{k=0}^{n-1} Z_k\right] + \mathbb{E}[Y_n] \\
&\le \mathbb{E}[V_0] + \mathbb{E}\left[\sum_{k=0}^{n} Y_k\right] .
\end{aligned}
$$

This proves (4.3.3). Note now that τ being an $\{\mathscr{F}_n\}$-stopping time, $\{\tau > n\} \in \mathscr{F}_n$; thus for $n \ge 0$,

$$
\begin{aligned}
&\mathbb{E}\left[V_{(n+1)\wedge\tau} \,\middle|\, \mathscr{F}_n\right] + Z_n \mathbb{1}\{\tau > n\} \\
&\qquad = \{\mathbb{E}[V_{n+1} \,|\, \mathscr{F}_n] + Z_n\} \mathbb{1}\{\tau > n\} + V_\tau \mathbb{1}\{\tau \le n\} \\
&\qquad \le (V_n + Y_n) \mathbb{1}\{\tau > n\} + V_\tau \mathbb{1}\{\tau \le n\} = V_{n\wedge\tau} + Y_n \mathbb{1}\{\tau > n\} .
\end{aligned}
$$

This means that the sequences $\{V_{n\wedge\tau}\}$, $\{Z_n \mathbb{1}\{\tau > n\}\}$, and $\{Y_n \mathbb{1}\{\tau > n\}\}$ satisfy assumption (4.3.1). Applying (4.3.3) to these sequences yields

$$
\begin{aligned}
\mathbb{E}[V_{n\wedge\tau} \mathbb{1}\{\tau < \infty\}] + \mathbb{E}\left[\sum_{k=0}^{n\wedge\tau-1} Z_k\right] &\le \mathbb{E}[V_{n\wedge\tau}] + \mathbb{E}\left[\sum_{k=0}^{n\wedge\tau-1} Z_k\right] \\
&\le \mathbb{E}[V_0] + \mathbb{E}\left[\sum_{k=0}^{n\wedge\tau-1} Y_k\right] \le \mathbb{E}[V_0] + \mathbb{E}\left[\sum_{k=0}^{\tau-1} Y_k\right] .
\end{aligned}
$$

Letting $n \to \infty$ on the left-hand side and applying Fatou's lemma yields (4.3.2). □

The comparison theorem is an essential tool for controlling the moments of the hitting or return times to a set. We will illustrate this through two examples. We first give a condition under which the expectation of the return time to a set is finite. This is the first instance of the use of a drift condition.

Proposition 4.3.2 *Assume that there exist measurable functions $V : \mathsf{X} \to [0,\infty]$ and $f : \mathsf{X} \to [0,\infty]$ and a set $C \in \mathscr{X}$ such that $PV(x) + f(x) \le V(x)$, $x \in C^c$. Then for all $x \in \mathsf{X}$,*

$$
\begin{aligned}
&\mathbb{E}_x[V(X_{\sigma_C}) \mathbb{1}_{\{\sigma_C < \infty\}}] + \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C - 1} f(X_k)\right] \\
&\qquad \le \{PV(x) + f(x)\} \mathbb{1}_C(x) + V(x) \mathbb{1}_{C^c}(x) . \quad (4.3.4)
\end{aligned}
$$

If $\sup_{x\in C}\{PV(x) + f(x)\} < \infty$, then

$$\sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} f(X_k) \right] < \infty ,$$

and

$$\mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} f(X_k) \right] < \infty$$

for all x such that $V(x) < \infty$. Furthermore, if π is an invariant probability measure and $\pi(\{V = \infty\}) = 0$, then $\pi(f) \leq \sup_{x \in C}\{PV(x) + f(x)\}$.

Proof. Write $d = \sup_{x \in C}\{PV(x) + f(x)\}$ (this quantity might be infinite). For $k \geq 0$, set $Z_k = f(X_k)$ and

$$V_0 = V(X_0)\mathbb{1}_{C^c}(X_0) , \qquad V_k = V(X_k) , \; k \geq 1$$
$$Y_0 = \{PV(X_0) + f(X_0)\}\mathbb{1}_C(X_0) , \qquad Y_k = d\mathbb{1}_C(X_k) , \; k \geq 1$$

with the convention $\infty \times 0 = 0$. Then (4.6.8) yields, for $k \geq 0$ and $x \in \mathsf{X}$,

$$\mathbb{E}_x [V_{k+1} \,|\, \mathscr{F}_k] + Z_k \leq V_k + Y_k \qquad \mathbb{P}_x - \text{a.s.}$$

Hence (4.3.1) holds, and (4.3.4) follows from the application of Theorem 4.3.1 with the stopping time σ_C. Assume now that $d < \infty$. Then by (4.3.4), for $x \in C$, we get

$$\mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} f(X_k) \right] \leq d ,$$

and if $x \notin C$, then $\mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} f(X_k) \right] \leq V(x)$. Let π be an invariant probability measure. Then for all $m \geq 0$, using Jensen's inequality and $PV(x) + f(x) \leq V(x) + d$, we get

$$\begin{aligned} \pi(f \wedge m) = n^{-1} \sum_{k=0}^{n-1} \pi P^k (f \wedge m) &\leq \pi \left(\left(n^{-1} \sum_{k=0}^{n-1} P^k f \right) \wedge m \right) \\ &\leq \pi \left[\left(n^{-1} V + d \right) \wedge m \right] . \end{aligned}$$

By letting $n \to \infty$ and then $m \to \infty$, we get $\pi(f) \leq d$. □

We will now give a condition under which the moment of the return time to a set admits a finite exponential moment.

Proposition 4.3.3 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $C \in \mathscr{X}$.*

(i) If $b = \sup_{x \in C} \mathbb{E}_x[\beta^{\sigma_C}] < \infty$ *for some* $\beta > 1$*, then* $V(x) = \mathbb{E}_x[\beta^{\tau_C}]$ *satisfies the geometric drift condition* $PV \leq \beta^{-1} V + b \mathbb{1}_C$.
(ii) If there exist a function $V : \mathsf{X} \to [1, \infty]$*,* $\lambda \in [0, 1)$*, and* $b < \infty$ *such that* $PV \leq \lambda V + b \mathbb{1}_C$*, then for all* $x \in \mathsf{X}$*,*

$$\mathbb{E}_x[\lambda^{-\sigma_C}] \leq V(x) + b\lambda^{-1} . \tag{4.3.5}$$

Proof. (i) Using the Markov property and the identity $\sigma_C = 1 + \tau_C \circ \theta$, we get

$$PV(x) = \mathbb{E}_x\left[\mathbb{E}_{X_1}[\beta^{\tau_C}]\right] = \mathbb{E}_x\left[\beta^{\tau_C \circ \theta}\right] = \beta^{-1}\mathbb{E}_x[\beta^{\sigma_C}] .$$

Hence $PV(x) = \beta^{-1} V(x)$ for $x \notin C$ and $\sup_{x \in C} PV(x) = \beta^{-1} \sup_{x \in C} \mathbb{E}_x[\beta^{\sigma_C}] < \infty$.

(ii) Set $V_n = V(X_n)$ for $n \geq 0$. Since $PV + (1-\lambda)V \leq V + b\mathbb{1}_C$, we get for $n \geq 0$,

$$\mathbb{E}\left[V_{n+1} \,|\, \mathscr{F}_n^X\right] = PV(X_n) + (1-\lambda)V(X_n) \leq V(X_n) + b\mathbb{1}_C(X_n) .$$

By applying Theorem 4.3.1, we therefore obtain

$$(1-\lambda)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C - 1} V(X_k)\right] \leq V(x) + b\mathbb{1}_C(x) .$$

Since $V \geq 1$, this implies that if $V(x) < \infty$, then $\mathbb{P}_x(\sigma_C < \infty) = 1$. Setting now $V_n = \lambda^{-n} V(X_n)$ for $n \geq 0$, we get

$$\mathbb{E}\left[V_{n+1} \,|\, \mathscr{F}_n^X\right] = \lambda^{-(n+1)} PV(X_n) \leq \lambda^{-n} V(X_n) + b\lambda^{-(n+1)}\mathbb{1}_C(X_n) .$$

Applying again Theorem 4.3.1, we get

$$\mathbb{E}_x[\lambda^{-\sigma_C} V(X_{\sigma_C})\mathbb{1}_{\{\sigma_C < \infty\}}] \leq V(x) + b\lambda^{-1}\mathbb{1}_C(x) .$$

If $V(x) < \infty$, then $\mathbb{P}_x(\sigma_C < \infty) = 1$, and (4.3.5) is thus satisfied. Equation (4.3.5) of course remains true if $V(x) = \infty$.

□

Surprisingly enough, the condition under which the moment of return time to a set C admits a finite exponential moment is equivalent to the existence of a geometric drift condition of the form $PV \leq \lambda V + b\mathbb{1}_C$, $\lambda \in [0, 1)$, and $b < \infty$. We will deepen these relationships in Chapters 14 and 16.

4.4 The Dirichlet and Poisson Problems

Definition 4.4.1 (Dirichlet Problem) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, $A \in \mathscr{X}$, and $g \in \mathbb{F}_+(\mathsf{X})$. A nonnegative function $u \in \mathbb{F}_+(\mathsf{X})$ is a solution to the Dirichlet problem if*

$$u(x) = \begin{cases} g(x) , & x \in A , \\ Pu(x) , & x \in A^c . \end{cases} \tag{4.4.1}$$

In words, we are looking for a function that is harmonic outside A and is equal to some positive function on A. Perhaps surprisingly, we will see below that it is fairly easy to find solutions to this problem. For $A \in \mathscr{X}$, we define a sub-Markovian kernel P_A for $x \in \mathsf{X}$ and $B \in \mathscr{X}$ by

$$P_A(x,B) = \mathbb{E}_x[\mathbb{1}_{\{\tau_A<\infty\}} \mathbb{1}_B(X_{\tau_A})] = \mathbb{P}_x(\tau_A < \infty, X_{\tau_A} \in B) , \tag{4.4.2}$$

which is the probability that the chain starting from x eventually hits the set $A \cap B$. For $f \in \mathbb{F}_+(\mathsf{X})$, we have

$$P_A f(x) = \mathbb{E}_x[\mathbb{1}_{\{\tau_A<\infty\}} f(X_{\tau_A})] . \tag{4.4.3}$$

The introduction of this kernel is motivated by the following result, which gives a solution to the Dirichlet problem.

Proposition 4.4.2 *For every $A \in \mathscr{X}$ and $g \in \mathbb{F}_+(\mathsf{X})$, the function $P_A g$ is a solution to the Dirichlet problem* (4.4.1).

Proof. If $x \in A$, then by definition, $P_A g(x) = g(x)$. For $x \in \mathsf{X}$, the identity $\sigma_A = 1 + \tau_A \circ \theta_1$ and the Markov property yield

$$\begin{aligned} PP_A g(x) &= \mathbb{E}_x[P_A g(X_1)] = \mathbb{E}_x[\{\mathbb{1}_{\{\tau_A<\infty\}} g(X_{\tau_A})\} \circ \theta_1] \\ &= \mathbb{E}_x[\mathbb{1}_{\{\tau_A \circ \theta_1<\infty\}} g(X_{1+\tau_A \circ \theta_1})] = \mathbb{E}_x[\mathbb{1}_{\{\sigma_A<\infty\}} g(X_{\sigma_A})] . \end{aligned}$$

For $x \notin A$, then $\sigma_A = \tau_A$ $\mathbb{P}_x -$a.s., and we obtain

$$PP_A g(x) = \mathbb{E}_x[\mathbb{1}_{\{\tau_A<\infty\}} g(X_{\tau_A})] = P_A g(x) .$$

$\square$

Definition 4.4.3 (Poisson problem) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, $A \in \mathscr{X}$, and let $f : A^c \to \mathbb{R}_+$ be a measurable function. A nonnegative function $u \in \mathbb{F}_+(\mathsf{X})$ is a solution to the Poisson problem if*

$$u(x) = \begin{cases} 0\,, & x \in A\,, \\ Pu(x) + f(x)\,, & x \in A^c\,. \end{cases} \tag{4.4.4}$$

In words, we are looking for a positive function that vanishes on the set A and is such that $u(x) = Pu(x) + f(x)$ on A^c. If $u(x)$ and $Pu(x)$ are both finite, this is equivalent to $\Delta u(x) = (\mathrm{I} - P)u(x) = f(x)$. If we interpret $\Delta = \mathrm{I} - P$ as the Laplacian, then this is the classical Poisson problem in potential theory. It is again possible to find an explicit solution to the Poisson problem. For $A \in \mathscr{X}$ and $h \in \mathbb{F}_+(\mathsf{X})$, define

$$G_A h(x) = \mathbb{1}_{A^c}(x)\mathbb{E}_x\left[\sum_{k=0}^{\tau_A - 1} h(X_k)\right] = \mathbb{E}_x\left[\sum_{k=0}^{\tau_A - 1} h(X_k)\right]\,, \tag{4.4.5}$$

where we have used the convention $\sum_{k=0}^{-1} \cdot = 0$. Note that $G_A h$ is nonnegative, but we do not assume that it is finite.

Proposition 4.4.4 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, $A \in \mathscr{X}$, and let $f : A^c \to \mathbb{R}_+$ be a measurable function. The function $G_A f$ is a solution to the Poisson problem* (4.4.4).

Proof. Set $u(x) = G_A f(x) = \mathbb{E}_x[S]$, where $S = \mathbb{1}_{A^c}(X_0)\sum_{k=0}^{\tau_A - 1} f(X_k)$. By convention, $u(x) = 0$ for $x \in A$. Applying the Markov property and the relation $\sigma_A = 1 + \tau_A \circ \theta_1$, we obtain

$$\begin{aligned} Pu(x) &= \mathbb{E}_x[u(X_1)] = \mathbb{E}_x[\mathbb{E}_{X_1}[S]] = \mathbb{E}_x[\mathbb{E}_x[S \circ \theta_1 \mid \mathscr{F}_1]] \\ &= \mathbb{E}_x[S \circ \theta_1] = \mathbb{E}_x\left[\mathbb{1}_{A^c}(X_1)\sum_{k=1}^{\tau_A \circ \theta_1} f(X_k)\right] = \mathbb{E}_x\left[\sum_{k=1}^{\sigma_A - 1} f(X_k)\right]\,, \end{aligned} \tag{4.4.6}$$

where the last equality follows from $\mathbb{1}_A(X_1)\sum_{k=1}^{\sigma_A - 1} f(X_k) = 0$. For $x \notin A$, one has $\sigma_A = \tau_A \;\; \mathbb{P}_x - \text{a.s.}$, and thus

$$f(x) + Pu(x) = f(x) + \mathbb{E}_x\left[\sum_{k=1}^{\sigma_A - 1} f(X_k)\right] = \mathbb{E}_x\left[\mathbb{1}_{A^c}(X_0)\sum_{k=0}^{\tau_A - 1} f(X_k)\right] = u(x)\,.$$

□

Combining Propositions 4.4.2 and 4.4.4 yields the solution to the Poisson–Dirichlet problem.

Theorem 4.4.5. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $A \in \mathscr{X}$. Given $f \in \mathbb{F}_+(A, \mathscr{X}_A)$ and $g \in \mathbb{F}_+(A^c, \mathscr{X}_{A^c})$, the function $P_A g + G_A f$ is a solution to the Poisson–Dirichlet problem*

$$u(x) = \begin{cases} g(x) \,, & x \in A, \\ Pu(x) + f(x) \,, & x \in A^c \,. \end{cases} \tag{4.4.7}$$

Furthermore, if $v \in \mathbb{F}_+(\mathsf{X})$ satisfies

$$v(x) \geq \begin{cases} g(x) \,, & x \in A, \\ Pv(x) + f(x) \,, & x \in A^c, \end{cases} \tag{4.4.8}$$

then $v \geq P_A g + G_A f$.

Remark 4.4.6. A function v that satisfies (4.4.8) is called a subsolution to the Poisson–Dirichlet problem (4.4.7). ▲

Proof. Equation (4.4.7) follows by combining Proposition 4.4.2 with Proposition 4.4.4. Assume now that (4.4.8) holds. Equation (4.4.8) implies

$$Pv + f\mathbb{1}_{A^c} + g\mathbb{1}_A \leq v + \mathbb{1}_A Pv \,.$$

Applying Theorem 4.3.1 with $V_n = v$, $Z_n = f\mathbb{1}_{A^c} + g\mathbb{1}_A$, $g = \mathbb{1}_A Pv$ and $\tau = \tau_A$, we obtain for all $x \in A^c$,

$$\begin{aligned} P_A g(x) + G_A f(x) &= \mathbb{E}_x\left[\mathbb{1}_{\{\tau_A<\infty\}} g(X_{\tau_A})\right] + \mathbb{E}_x\left[\sum_{k=0}^{\tau_A-1} f(X_k)\right] \\ &\leq \mathbb{E}_x\left[\mathbb{1}_{\{\tau_A<\infty\}} v(X_{\tau_A})\right] \\ &\quad + \mathbb{E}_x\left[\sum_{k=0}^{\tau_A-1} \{f(X_k)\mathbb{1}_{A^c}(X_k) + \mathbb{1}_A(X_k) g(X_k)\}\right] \\ &\leq v(x) + \mathbb{E}_x\left[\sum_{k=0}^{\tau_A-1} \mathbb{1}_A(X_k) Pv(X_k)\right] = v(x). \end{aligned}$$

On the other hand, $v(x) \geq g(x) = P_A g(x) + G_A f(x)$ for $x \in A$ by construction. □

We now state several useful consequences of Theorem 4.4.5.

Corollary 4.4.7 *The function $x \mapsto \mathbb{P}_x(\tau_A < \infty)$ is the smallest positive solution to the system*

$$v(x) \geq \begin{cases} 1 & \text{if } x \in A \;, \\ Pv(x) & \text{if } x \notin A \;. \end{cases}$$

Proof. Apply Theorem 4.4.5 with $g = \mathbb{1}_A$ and $f = 0$. □

Corollary 4.4.8 *The function $x \mapsto \mathbb{E}_x[\tau_A]$ is the smallest positive solution to the system*

$$v(x) \geq \begin{cases} 0 & \text{if } x \in A \;, \\ Pv(x) + 1 & \text{if } x \notin A \;. \end{cases}$$

Proof. We apply Theorem 4.4.5 with $g = 0$ and $f = \mathbb{1}_{A^c}$. In that case, the solution is given by

$$\mathbb{1}_{A^c}(x)\mathbb{E}_x\left[\sum_{k=0}^{\tau_A - 1} \mathbb{1}_{A^c}(X_k)\right] = \mathbb{1}_{A^c}(x)\mathbb{E}_x[\tau_A] = \mathbb{E}_x[\tau_A] \;.$$

□

Corollary 4.4.9 *Let $f \in \mathbb{F}_+(\mathsf{X})$. Then Uf is a solution to the equation $u = Pu + f$. If $w \in \mathbb{F}_+(\mathsf{X})$ satisfies the inequality*

$$w \geq Pw + f \;, \tag{4.4.9}$$

then $Uf \leq w$, i.e., Uf is the smallest solution to (4.4.9).

Proof. Apply Theorem 4.4.5 with $A = \emptyset$. □

4.5 Time-Inhomogeneous Poisson–Dirichlet Problems

We now introduce the time-inhomogeneous Poisson–Dirichlet problem. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. For a function h defined on the product space $\mathbb{N} \times \mathsf{X}$, we write $h_n(x) = h(n,x)$ for $n \in \mathbb{N}$ and $x \in \mathsf{X}$. Consider the Markov kernel $\tilde{P}$ on the product space $\mathbb{N} \times \mathsf{X}$ defined for $h \in \mathbb{F}_+(\mathbb{N} \times \mathsf{X}, \mathscr{P}(\mathbb{N}) \otimes \mathscr{X})$ by

$$\tilde{P}h(n,x) = \int h(n+1,y)P(x,\mathrm{d}y) = Ph_{n+1}(x) . \tag{4.5.1}$$

Let $\{(I_n, X_n),\ n \in \mathbb{N}\}$ be the coordinate process on the canonical space $(\mathbb{N} \times \mathsf{X})^{\mathbb{N}}$. For a probability measure ξ on $\mathscr{P}(\mathbb{N}) \otimes \mathscr{X}$, let $\tilde{\mathbb{P}}_\xi$ be the probability measure that makes the coordinate process a Markov chain with kernel $\tilde{P}$ and initial distribution ξ. Denote by $\tilde{\mathbb{E}}_\xi$ the associated expectation operator. Then for every $n, m \in \mathbb{N}$ and $x \in \mathsf{X}$ and $h \in \mathbb{F}_+(\mathbb{N} \times \mathsf{X}, \mathscr{P}(\mathbb{N}) \otimes \mathscr{X})$, we have

$$\tilde{P}^n h(m,x) = \tilde{\mathbb{E}}_{m,x}[h(I_n, X_n)] = P^n h_{n+m}(x) = \mathbb{E}_x[h_{n+m}(X_n)] .$$

We can now rewrite Theorem 4.4.5 in the time-inhomogeneous framework.

Theorem 4.5.1. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $A \in \mathscr{X}$. Given $\tilde{f} \in \mathbb{F}_+(\mathbb{N} \times A^c, \mathscr{P}(\mathbb{N}) \otimes \mathscr{X}_{A^c})$, and $\tilde{g} \in \mathbb{F}_+(\mathbb{N} \times A, \mathscr{P}(\mathbb{N}) \otimes \mathscr{X}_A)$, the function $\tilde{u} : (m,x) \in \mathbb{N} \times \mathsf{X}$ defined by*

$$\tilde{u}(m,x) = \mathbb{E}_x\left[\mathbb{1}_{\{\tau_A < \infty\}} \tilde{g}(m + \tau_A, X_{\tau_A})\right] + \mathbb{E}_x\left[\sum_{k=0}^{\tau_A - 1} \tilde{f}(m+k, X_k)\right] \tag{4.5.2}$$

is a solution to the inhomogeneous Poisson–Dirichlet problem

$$u(m,x) = \begin{cases} \tilde{g}(m,x) , & x \in A , \\ Pu_{m+1}(x) + \tilde{f}(m,x) , & x \notin A . \end{cases} \tag{4.5.3}$$

Furthermore, if $\tilde{v} \in \mathbb{F}_+(\mathbb{N} \times \mathsf{X}, \mathscr{P}(\mathbb{N}) \otimes \mathscr{X})$ is a subsolution to the inhomogeneous Poisson–Dirichlet problem

$$\tilde{v}(m,x) \geq \begin{cases} \tilde{g}(m,x) , & x \in A , \\ P\tilde{v}_{m+1}(x) + \tilde{f}(m,x) , & x \notin A , \end{cases} \tag{4.5.4}$$

then $\tilde{v}(m,x) \geq \tilde{u}(m,x)$ for every $(m,x) \in \mathbb{N} \times \mathsf{X}$.

Proof. Apply Theorem 4.4.5 to the Markov kernel $\tilde{P}$ defined in (4.5.1). □

4.6 Exercises

4.1 (Riesz decomposition). Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. We will show that a finite superharmonic function $f \in \mathbb{F}_+(\mathsf{X})$ can be decomposed uniquely as

$$f = h + Ug,$$

where h is a harmonic function and g is a positive measurable function. Furthermore, $h = \lim_{n\to\infty} P^n f$ and $g = f - Pf$.

1. Show that the sequence $\{P^n f \,:\, n \in \mathbb{N}\}$ converges.

Define $h(x) = \lim_{n\to\infty} P^n f(x)$.

2. Show that for every $x \in \mathsf{X}$, $Ph(x) = h(x)$.

Set $g = f - Pf$.

3. Show that g is nonnegative and that for all $x \in \mathsf{X}$, $Ug(x) = f(x) - h(x)$.

We now show that this decomposition is unique. Assume that $f = \bar{h} + U\bar{g}$, where $\bar{h}$ is a harmonic function and $\bar{g} \in \mathbb{F}_+(\mathsf{X})$.

4. Show that for all $n \geq 1$ and $x \in \mathsf{X}$,

$$\sum_{k=0}^{n-1} P^k g(x) = f(x) - P^n f(x) \; .$$

5. Show that $Ug = U\bar{g}$.
6. Conclude.

4.2. This exercise uses the results of Exercise 4.1. For $A \in \mathscr{X}$, define the functions

$$f_A(x) = \mathbb{P}_x(\tau_A < \infty) \; , \quad g_A(x) = \mathbb{1}_A(x)\mathbb{P}_x(\sigma_A = \infty), \quad \text{and} \quad h_A(x) = \mathbb{P}_x(N_A = \infty)$$

for $x \in \mathsf{X}$.

1. Show that the function h_A is harmonic.
2. Show that $f_A(x) = h(x) + Ug(x)$ with $h(x) = \lim_{n\to\infty} P^n f_A(x)$ and $g(x) = f_A(x) - Pf_A(x)$.
3. Show that for every $n \in \mathbb{N}$, $P^n f_A(x) = \mathbb{P}_x\left(\cup_{k\geq n}\{X_k \in A\}\right)$ and that $h = h_A$.
4. Show that

$$\begin{aligned} f_A(x) - Pf_A(x) &= \mathbb{P}_x\left(\bigcup_{k\geq 0}\{X_k \in A\}\right) - \mathbb{P}_x\left(\bigcup_{k\geq 1}\{X_k \in A\}\right) \\ &= \mathbb{1}_A(x)\mathbb{P}_x(\sigma_A = \infty) = g_A(x). \end{aligned}$$

5. Conclude that $f_A = h_A + Ug_A$.

4.3 (Dynkin formula). Let $\{(Z_n, \mathscr{F}_n),\, n \in \mathbb{N}\}$ be a bounded adapted process and τ an integrable stopping time. We will prove the following identity, due to Dynkin:

$$\mathbb{E}\left[Z_\tau\right] - \mathbb{E}\left[Z_0\right] = \mathbb{E}\left[\sum_{k=0}^{\tau-1} \mathbb{E}\left[Z_{k+1} - Z_k \,|\, \mathscr{F}_k\right]\right] \; .$$

We set $U_0 = 0$, and for $n \geq 1$,

$$U_n = Z_n - Z_0 - \sum_{k=0}^{n-1} \{\mathbb{E}[Z_{k+1} \mid \mathscr{F}_k] - Z_k\} .$$

1. Show that $\{(U_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ is a martingale.
2. Show that for all $n \in \mathbb{N}$,

$$\mathbb{E}[Z_{n\wedge\tau}] - \mathbb{E}[Z_0] = \mathbb{E}\left[\sum_{k=0}^{n\wedge\tau-1} \{\mathbb{E}[Z_{k+1} \mid \mathscr{F}_k] - Z_k\}\right] .$$

3. Conclude.

Let $f \in \mathbb{F}_b(\mathsf{X})$ and let P be a Markov kernel.

4. Show that for all $x \in \mathsf{X}$ and all stopping times τ such that $\mathbb{E}_x[\tau] < \infty$,

$$\mathbb{E}_x[f(X_\tau)] - f(x) = \mathbb{E}_x\left[\sum_{k=0}^{\tau-1} (P-I)f(X_k)\right] .$$

4.4. This exercise uses the results of Exercise 4.3. Let $\{(Z_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ be an adapted nonnegative process and τ a stopping time such that $\mathbb{P}(\tau < \infty) = 1$.

1. Show that

$$\mathbb{E}[Z_\tau] + \mathbb{E}\left[\sum_{k=0}^{\tau-1} Z_k\right] = \mathbb{E}[Z_0] + \mathbb{E}\left[\sum_{k=0}^{\tau-1} \mathbb{E}[Z_{k+1} \mid \mathscr{F}_k]\right] .$$

[Hint: Apply Exercise 4.3 to the finite stopping time $\tau \wedge n$ and the bounded process $\{Z_n^M,\ n \in \mathbb{N}\}$, where $Z_n^M = Z_n \wedge M$.]

2. Let $f \in \mathbb{F}_+(\mathsf{X})$. Show that for all $x \in \mathsf{X}$ and all stopping times τ such that $\mathbb{P}_x(\tau < \infty) = 1$,

$$\mathbb{E}_x[f(X_\tau)] + \mathbb{E}_x\left[\sum_{k=0}^{\tau-1} f(X_k)\right] = f(x) + \mathbb{E}_x\left[\sum_{k=0}^{\tau-1} Pf(X_k)\right] .$$

4.5 (Random walk on $\mathbb{Z}$). Consider the simple random walk on $\mathbb{Z}$, not necessarily symmetric, that is a Markov chain on $\mathbb{Z}$ with kernel P defined by $P(x, x+1) = p$ and $P(x, x-1) = q$ for all $x \in \mathbb{Z}$, where $p \in [0,1]$ and $p + q = 1$. Let $a < b \in \mathbb{Z}$ and let τ be the hitting time of $A = \{a+1, \ldots, b-1\}^c$, where $a+1 < b-1$ (we have dropped the dependence on the set A from the notation). The purpose of this exercise is to compute the moments of τ.

1. Show that for all $x \in A^c$, $\mathbb{P}_x(\tau \le b-a) \ge \gamma = p^{b-a}$.
2. Show that for all $x \in A^c$ and $n \in \mathbb{N}$, $\mathbb{P}_x(\tau > n) \le (1-\gamma)^{(n-(b-a))/(b-a)}$ and that $\mathbb{E}_x[\tau^s] < \infty$ for any $s > 0$.

Let g be a nonnegative function of A^c. Consider the system of equations

$$\begin{cases} u(x) = g(x) + Pu(x) , & a < x < b , \\ u(a) = \alpha , \quad u(b) = \beta . \end{cases} \tag{4.6.1}$$

3. Show that $u_1(x) = \mathbb{E}_x[\tau]$ is the minimal solution to (4.6.1) with $g(x) = \mathbb{1}_{A^c}(x)$, $\alpha = 0$, and $\beta = 0$.
4. Show that u_2 is the finite solution to the system (4.6.1) with $g(x) = 1 + 2Pu_1(x)$ for $x \in A^c$ and $\alpha = \beta = 0$.
5. Show that u_3 is the finite solution to the system (4.6.1) with $g(x) = 1 + 3Pu_1(x) + 3Pu_2(x)$ for $x \in A^c$, $\alpha = \beta = 0$.

We will finally show that the system of equations (4.6.1) has a unique finite solution on $\{a, \ldots, b\}$.

6. Show that for $x \in A^c$, the equation $u(x) - Pu(x) = g(x)$ is equivalent to

$$u(x+1) - u(x) = \rho\{u(x) - u(x-1)\} - p^{-1}g(x) , \tag{4.6.2}$$

where $\rho = (1-p)/p$.

7. Set $\Delta u(x+1) = u(x+1) - u(x)$. Show that (4.6.2) is equivalent, for $x = a + 1, \ldots, b$, to the system of equations

$$\Delta u(x) = \rho^{x-a-1}\Delta u(a+1) - p^{-1}\sum_{y=0}^{x-a-1}\rho^y g(x-y-1) , \tag{4.6.3}$$

and a solution u of (4.6.2) is uniquely determined by $u(a)$ and $u(a+1)$.

8. Determine the unique solution ϕ of (4.6.3) in the case that $\phi(a+1) = 1$, $\phi(a) = 0$, and $g(x) = 0$ for every $x \in \{-a+1, \ldots, b\}$.
9. Determine the unique solution ψ of (4.6.2) such that $\psi(a) = \psi(a+1) = 0$ for an arbitrary function g.
10. Determine the unique solution to (4.6.1) for an arbitrary function g and arbitrary initial conditions.

4.6 (Birth-and-death chain). A level-dependent quasi birth-and-death process is a Markov chain on the finite set $\mathbb{Z}$ with kernel P defined by

$$P(x, x+1) = p_x , \quad P(x, x-1) = q_x , \quad P(x, x) = r_x ,$$

with $p_x + q_x + r_x = 1$ for all $x \in \mathbb{Z}$. Denote for $x \in \mathbb{N}$ by $h(x)$ the extinction probability starting from x, i.e., $h(x) = \mathbb{P}_x(\tau_0 < \infty)$.

1. Show that h is the smallest solution to $h(0) = 1$ and $Ph(x) = h(x)$ for $x \in \mathbb{N}^*$.
2. Show that h is nonincreasing and that for every $x \in \mathbb{N}^*$,

$$h(x) = p_x h(x+1) + q_x h(x-1) .$$

Define $u(x) = h(x-1) - h(x)$.

3. Show that for all $x \in \mathbb{N}$, $u(x+1) = \gamma(x)u(1)$ with $\gamma(0) = 1$ and

$$\gamma(x) = \frac{q_x q_{x-1} \cdots q_1}{p_x p_{x-1} \cdots p_1} \ .$$

4. Deduce that for all $x \in \mathbb{N}^*$, $h(x) = 1 - u(1)\{\gamma(0) + \cdots + \gamma(x-1)\}$.

Assume first that $\sum_{x=0}^{\infty} \gamma(x) = \infty$.

5. Show that $h(x) = 1$ for all $x \in \mathbb{N}$.

Assume now that $\sum_{x=0}^{\infty} \gamma(x) < \infty$.

6. Show that $h(x) = \sum_{y=x}^{\infty} \gamma(y) / \sum_{y=0}^{\infty} \gamma(y)$.

In the latter case, for $x \in \mathbb{N}^*$, we have $h(x) < 1$, so the population survives with positive probability.

4.7. This is a follow-up on Exercise 4.6. Let P be a Markov kernel on $\mathbb{N}$ with transition probability given by $P(0,1) = 1$ and for $x \geq 1$,

$$P(x,x+1) + P(x,x-1) = 1, \quad P(x,x+1) = \left(\frac{x+1}{x}\right)^2 P(x,x-1) \ .$$

Show that $\mathbb{P}_0(\sigma_0 = \infty) = 6/\pi^2$.

4.8 (The gambler's ruin). Let $\{Z_k,\ k \in \mathbb{N}^*\}$ be a sequence of i.i.d. random variables taking values in $\{-1,1\}$ with probability $\mathbb{P}(Z_k = 1) = \mathbb{P}(Z_k = -1) = 1/2$. Denote by X_n the current wealth of a gambler, i.e.,

$$X_n = X_0 + Z_1 + Z_2 + \cdots + Z_n \ ,$$

where X_0 is the gambler's initial wealth. Assume that the gambler stops the game when her wealth reaches either the upper barrier a or the lower barrier $-b$, a and b being positive integers. The gambler's wealth is a Markov chain on the state space $\mathsf{X} = \{-b, \ldots, a\}$. Let τ be the hitting time of the set $\{-b, a\}$, i.e., $\tau = \inf\{k \geq 0, X_k \in \{-b, a\}\}$. We want to compute the probability that the game ends in finite time. Define the function u on X by $u(x) = \mathbb{P}_x(\tau < \infty)$.

1. Show that u is harmonic on $\mathsf{X} \setminus \{-b, a\}$ and that for $x \in \mathsf{X} \setminus \{-b, a\}$,

$$u(x) = Pu(x) = \frac{1}{2} u(x-1) + \frac{1}{2} u(x+1) \ . \tag{4.6.4}$$

2. Deduce that

$$u(x) = u(-b) + (x+b)\{u(-b+1) - u(-b)\} \tag{4.6.5}$$

for all $x \in \mathsf{X} \setminus \{-b, a\}$.

3. Show that $\mathbb{P}_x(\tau < \infty) = 1$ for all $x \in \mathsf{X}$, i.e., the game ends in finite time almost surely for every initial wealth $x \in \{-b, \ldots, a\}$.

We now compute the probability $u(x) = \mathbb{P}_x(\tau_a < \tau_{-b})$ of winning. We can also write $u(x) = \mathbb{E}_x[\mathbb{1}_a(X_\tau)]$.

4. Show that u is the smallest nonnegative solution to the equations

$$\begin{cases} u(x) = Pu(x) \;, & x \in \mathsf{X} \setminus \{-b, a\} \;, \\ u(-b) = 0 \;, \quad u(a) = 1 \;. \end{cases}$$

5. Show that the probability of winning when the initial wealth is x is equal to $u(x) = (x+b)/(a+b)$.

We will now compute the expected time of a game.

6. Show that $u(x) = \mathbb{E}_x[\tau]$ is the smallest solution to the Poisson problem (4.4.4) and that for $x \in \{-b+1, \ldots, a-1\}$,

$$u(x) = \tfrac{1}{2}u(x-1) + \tfrac{1}{2}u(x+1) + 1 \;, \tag{4.6.6}$$

with boundary conditions $u(-b) = 0$ and $u(a) = 0$.
7. Show that $u(x)$ is given by

$$u(x) = (a-x)(x+b) \;,\; x = -b, \ldots, a \;. \tag{4.6.7}$$

4.9. Let P be a Markov kernel on a discrete state space X.

1. Show that for all $x \in \mathsf{X}$, $\mathbb{P}_x(\sigma_x^{(n)} < \infty) = \{\mathbb{P}_x(\sigma_x < \infty)\}^n$.
2. Show that for all $x \in \mathsf{X}$, $\mathbb{P}_x(N_x = \infty) = \lim_{n\to\infty}\{\mathbb{P}_x(\sigma_x < \infty)\}^n$.
3. Show that $\mathbb{E}_x[N_x] = \sum_{n=0}^{\infty}\{\mathbb{P}_x(\sigma_x < \infty)\}^n$.
4. Show that the following conditions are equivalent:

$$\mathbb{P}_x(\sigma_x < \infty) = 1 \Longleftrightarrow \mathbb{P}_x(N_x = \infty) = 1 \Longleftrightarrow U(x,x) = \infty \;.$$

4.10. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that there exist a sequence of measurable functions $V_n : \mathsf{X} \to [0,\infty]$, $n \geq 0$, a measurable function $h : \mathsf{X} \to [0,\infty]$, a nonnegative sequence r, and a set $C \in \mathscr{X}$ such that

$$PV_{n+1}(x) + r(n)h(x) \leq V_n(x) \;, \qquad x \in C^c \;,\; n \in \mathbb{N} \;. \tag{4.6.8}$$

Show that for every $x \in \mathsf{X}$,

$$\mathbb{E}_x\left[V_{\sigma_C}(X_{\sigma_C})\mathbb{1}_{\{\sigma_C<\infty\}}\right] + \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k)h(X_k)\right] \\ \leq \{PV_1(x) + r(0)h(x)\}\mathbb{1}_C(x) + V_0(x)\mathbb{1}_{C^c}(x) \;. \tag{4.6.9}$$

4.11. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let r be a nonnegative sequence, $g, h : \mathsf{X} \to [0,\infty]$ measurable functions, and $C \in \mathscr{X}$. Define for $n \geq 0$,

$$W_n(x) = \mathbb{E}_x\left[r(n+\tau_C)g(X_{\tau_C})\mathbb{1}_{\{\tau_C<\infty\}}\right] + \mathbb{E}_x\left[\sum_{k=0}^{\tau_C-1} r(n+k)h(X_k)\right] \;. \tag{4.6.10}$$

1. Show that for all $x \in \mathsf{X}$,

$$PW_{n+1}(x) + r(n)h(x) \\ = \mathbb{E}_x[r(n+\sigma_C)g(X_{\sigma_C})\mathbb{1}_{\{\sigma_C<\infty\}}] + \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(n+k)h(X_k)\right] . \quad (4.6.11)$$

2. Let $\{V_n,\ n \in \mathbb{N}\}$ be a sequence of nonnegative functions such that for all $n \geq 0$,

$$PV_{n+1}(x) + r(n)h(x) \leq V_n(x) , \qquad x \notin C , \quad (4.6.12a)$$
$$r(n)g(x) \leq V_n(x) , \qquad x \in C . \quad (4.6.12b)$$

Show that $V_n \geq W_n$ for all $n \geq 0$ and if $\sup_C\{PV_1 + r(0)h\} < \infty$, then

$$\sup_C\{PW_1 + r(0)h\} < \infty. \quad (4.6.13)$$

4.12. Let $f : \mathsf{X} \to [1,\infty]$, $g : \mathsf{X} \to [0,\infty]$ be measurable functions, $\delta > 1$ a constant, and $C \in \mathscr{X}$.

1. Find the minimal solution $W_C^{f,g,\delta}$ of

$$PV(x) + f(x) \leq \delta^{-1}V(x) , \quad x \notin C, \quad (4.6.14)$$

and $V(x) \geq g(x)$ for $x \in C$.

2. Prove that if $\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1}\delta^k f(X_k)\right] < \infty$, then

$$\sup_{x\in C}\{PW_C^{f,g,\delta}(x) + f(x)\} < \infty .$$

4.7 Bibliographical Notes

The potential theory of Markov chains is developed in (Revuz 1984, Chapter 2). This chapter presents only a few elements of the theory that gave rise to a great deal of work following the early work of Kemeny and Snell (1961a,b, 1963), Orey (1964), Chung (1967), Neveu (1964, 1972). Privault (2008) is a very didactic introduction to the links between Markov chain potential theory and classical potential theory. Some additional results (such as the Riesz decomposition theorem, which is a central result in potential theory) are established in the exercises.

The important result Theorem 4.2.6, which will find many applications, is due to Orey (1971) (an earlier version appeared in Orey (1959)).

The comparison theorem (Theorem 4.3.1) is an easy consequence of a (discrete-time) version of Dynkin's formula (see Exercise 4.3). Since this is the only place where Dynkin's formula plays a role, we have nevertheless decided to provide a "direct" proof. The use of drift criteria for general state-space chains and the use of Theorem 4.3.1 seem to appear for the first time in Kalashnikov (1968, 1971, 1977). We have closely followed here (Meyn and Tweedie 2009, Chapter 11).

Chapter 5
Ergodic Theory for Markov Chains

This chapter is concerned with the asymptotic behavior of sample averages of stationary ergodic Markov chains. For this purpose, it is convenient to link the Markov chain to a certain dynamical system. The law of large numbers for Markov chains is then obtained as a consequence of the classical Birkhoff theorem. It turns out that under appropriate assumptions, this approach still holds for functions that actually depend on the whole trajectory, such as $n^{-1}\sum_{k=0}^{n-1} f(\{X_{k+\ell},\ \ell \in \mathbb{N}\})$ or $n^{-1}\sum_{k=0}^{n-1} f(\{X_{k-\ell},\ \ell \in \mathbb{N}\})$. A key result of this chapter is Theorem 5.2.6, which shows that the existence of a unique invariant probability measure implies the ergodicity of the associated dynamical system, which in turn allows one to apply the Birkhoff ergodic theorem. Still, the price to pay for using dynamical system theory is the stationarity assumption. Typically in this chapter, the law of large numbers will be proved $\mathbb{P}_\pi -$ a.s. (where π is the unique invariant probability measure for P) and will be extended to other initial distributions. Sufficient conditions are given in this chapter, but a more thorough treatment for other initial distributions requires notions that will be introduced in later chapters.

5.1 Dynamical Systems

We first briefly introduce some basic definitions and properties of dynamical systems that will be useful in applying such systems to Markov chains.

5.1.1 Definitions

Definition 5.1.1 (Dynamical system) *Let* $(\Omega, \mathscr{B}, \mathbb{P})$ *be a probability space.*

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_5

- *A measurable map* T *from* $(\Omega, \mathscr{B})$ *to* $(\Omega, \mathscr{B})$ *is a measure-preserving transformation if for all* $A \in \mathscr{B}$,
$$\mathbb{P}(\mathrm{T}^{-1}(A)) = \mathbb{P}(A) \; .$$
The probability $\mathbb{P}$ *is then said to be invariant under the transformation* T, *and* $(\Omega, \mathscr{B}, \mathbb{P}, \mathrm{T})$ *is said to be a dynamical system.*
- *The application* T *is said to be an invertible measure-preserving transformation if it is measure-preserving, invertible, and its inverse* T^{-1} *is measurable.*

If the transformation T is measure preserving and invertible, then T^{-1} is also measure-preserving, since for all $A \in \mathscr{B}$,

$$\mathbb{P}((\mathrm{T}^{-1})^{-1}(A)) = \mathbb{P}(\mathrm{T}(A)) = \mathbb{P}(\mathrm{T}^{-1}\{\mathrm{T}(A)\}) = \mathbb{P}(A) \; .$$

Note also that if T is measure-preserving, then for all integers $n \in \mathbb{N}$ and $A \in \mathscr{B}$,

$$\mathbb{P}(\mathrm{T}^{-n}(A)) = \mathbb{P}(A) \; .$$

Let $(\mathsf{X}, \mathscr{X})$ be a measurable space. Denote by $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}})$ the associated canonical space and by $\{X_n,\ n \in \mathbb{N}\}$ the coordinate process. The shift operator θ (see Definition 3.1.8) is defined, for $\omega = (\omega_k)_{k \in \mathbb{N}} \in \mathsf{X}^{\mathbb{N}}$, by

$$\theta(\omega_0, \omega_1, \dots) = (\omega_1, \omega_2, \dots) \; .$$

Note that $X_k \circ \theta = X_{k+1}$ for all $k \geq 0$. By Proposition 3.1.9, θ is a measurable map from $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}})$ to $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}})$, but it is not invertible. Recall that $\{X_n,\ n \in \mathbb{N}\}$ is stationary if the distribution of $(X_k, \dots, X_{k+n})$ is independent of k for all $n \in \mathbb{N}$. The next lemma shows the connection between the stationarity of the coordinate process and the invariance of $\mathbb{P}$ under the shift operator θ.

Lemma 5.1.2 *A probability measure* $\mathbb{P}$ *on* $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}})$ *is invariant under the shift operator* θ *if and only if the coordinate process is stationary under* $\mathbb{P}$.

Proof. It suffices to note (Theorem B.2.6) that $\mathbb{P}$ is measure-preserving if and only if for all $n \geq 1$ and all $f \in \mathbb{F}_b(\mathsf{X}^n, \mathscr{X}^{\otimes n})$, $\mathbb{E}[f(X_0, \dots, X_{n-1})] = \mathbb{E}[f(X_1, \dots, X_n)]$. □

Example 5.1.3 (One-sided Markov shift). Let P be a kernel on $(\mathsf{X}, \mathscr{X})$ that admits an invariant probability π on $(\mathsf{X}, \mathscr{X})$. By Theorem 3.1.2, there exists a unique probability measure $\mathbb{P}_\pi$ on the canonical space such that the coordinate process is a Markov chain with kernel P and initial distribution π. Then by Theorem 1.4.2, the canonical chain is a stationary process. Lemma 5.1.2 then shows that $\mathbb{P}_\pi$ is invariant under θ, i.e., $\mathbb{P}_\pi \circ \theta^{-1} = \mathbb{P}_\pi$.

5.1.2 Invariant Events

Definition 5.1.4 (Invariant random variable, invariant event) *Let* T *be a measurable map from* $(\Omega, \mathscr{B})$ *to* $(\Omega, \mathscr{B})$.

- *An* $\bar{\mathbb{R}}$*-valued random variable* Y *on* $(\Omega, \mathscr{B})$ *is invariant for* T *if* $Y \circ \mathrm{T} = Y$.
- *An event* A *is invariant for* T *if* $A = \mathrm{T}^{-1}(A)$, *or equivalently if its indicator function* $\mathbb{1}_A$ *is invariant for* T.

Proposition 5.1.5 *Let* T *be a measurable map from* $(\Omega, \mathscr{B})$ *to* $(\Omega, \mathscr{B})$.

(i) The collection $\mathscr{I}$ *of invariant sets is a sub-*σ*-field of* $\mathscr{B}$.
(ii) Let Y *be an* $\bar{\mathbb{R}}$*-valued random variable;* Y *is invariant if and only if* Y *is* $\mathscr{I}$*-measurable.*

Proof. The proof of (i) is elementary and is therefore omitted. Consider now (ii). If $Y \circ \mathrm{T} = Y$, then for all $B \in \mathscr{B}(\bar{\mathbb{R}})$,

$$\mathrm{T}^{-1}(Y^{-1}(B)) = (Y \circ \mathrm{T})^{-1}(B) = Y^{-1}(B) .$$

Thus $Y^{-1}(B) \in \mathscr{I}$, and Y is $\mathscr{I}$-measurable.

Conversely, if Y is $\mathscr{I}$-measurable, define $A_{k,n} = \{\frac{k}{n} \le Y < \frac{k+1}{n}\} \in \mathscr{I}$, $n \ge 1$, $k \in \mathbb{Z}$. Then with the convention $\infty \times 0 = 0$, Y is the pointwise limit of the sequence $\{Y_n,\ n \in \mathbb{N}^*\}$ defined by

$$Y_n = \sum_{k \in \mathbb{Z}} \frac{k}{n} \mathbb{1}_{A_{k,n}} + \infty \mathbb{1}_{\{Y=+\infty\}} - \infty \mathbb{1}_{\{Y=-\infty\}} .$$

Since Y is $\mathscr{I}$-measurable, the sets $A_{k,n}$, $\{Y = -\infty\}$, and $\{Y = +\infty\}$ belong to $\mathscr{I}$; hence the functions Y_n are invariant for all n, and

$$Y \circ T = (\lim_{n\to\infty} Y_n) \circ T = \lim_{n\to\infty} (Y_n \circ T) = \lim_{n\to\infty} Y_n = Y .$$

□

The most important examples of invariant random variables that will be considered in the sequel are defined as limits.

Lemma 5.1.6 *Let* $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}})$ *be the canonical space,* $\{X_n,\ n \in \mathbb{N}\}$ *the coordinate process, and* θ *the shift operator. Then* $\mathscr{I} \subset \cap_{k \ge 0} \sigma(X_\ell,\ \ell \ge k)$. *Moreover, for all* $f \in \mathbb{F}\left(\bar{\mathbb{R}}, \mathscr{B}(\bar{\mathbb{R}})\right)$, $\limsup_{n\to\infty} f(X_n)$, $\liminf_{n\to\infty} f(X_n)$, $\limsup_{n\to\infty} n^{-1}(f(X_0) + \cdots + f(X_{n-1}))$, *and* $\liminf_{n\to\infty} n^{-1}(f(X_0) + \cdots + f(X_{n-1}))$ *are invariant random variables.*

Proof. Set $\mathscr{G}_k = \sigma(X_\ell, \ell \geq k)$ and $\mathscr{G}_\infty = \cap_{k\geq 0}\mathscr{G}_k$. Let A be an invariant set. Then $A \in \mathscr{G}_0 = \mathscr{X}^{\otimes \mathbb{N}}$. Since we have the implication if $A \in \mathscr{G}_k$, then $A = \theta^{-1}(A) \in \mathscr{G}_{k+1}$, we obtain by induction that $A \in \mathscr{G}_k$ for all k and thus $A \in \cap_{k\geq 0}\mathscr{G}_k$.

The remaining statements of the lemma are straightforward. □

Let now $(\Omega, \mathscr{B}, \mathbb{P}, \mathrm{T})$ be a dynamical system, that is, T is measure-preserving for $\mathbb{P}$. An $\bar{\mathbb{R}}$-valued random variable Y defined on Ω is said to be $\mathbb{P}-\text{a.s.}$ invariant (for T) if $Y \circ \mathrm{T} = Y$, $\mathbb{P}-\text{a.s.}$ Similarly, an event $A \in \mathscr{B}$ is $\mathbb{P}-\text{a.s.}$ invariant (for T) if its indicator function $\mathbb{1}_A$ is $\mathbb{P}-\text{a.s.}$ invariant.

Lemma 5.1.7 *If Y is $\mathbb{P}-\text{a.s.}$ invariant, then there exists an invariant random variable Z such that $Y = Z \;\; \mathbb{P}-\text{a.s.}$ In particular, if $A \in \mathscr{B}$ is $\mathbb{P}-\text{a.s.}$ invariant, there exists $B \in \mathscr{I}$ such that $\mathbb{1}_A = \mathbb{1}_B \;\; \mathbb{P}-\text{a.s.}$*

Proof. The random variable $Z = \limsup_{n\to\infty} Y \circ \mathrm{T}^n$ is invariant. Since Y is $\mathbb{P}-\text{a.s.}$ invariant, $Y = Y \circ \mathrm{T} \;\; \mathbb{P}-\text{a.s.}$; hence $Y = Y \circ \mathrm{T}^n$ $\mathbb{P}-\text{a.s.}$ for all $n \geq 1$. This yields $Y = Z$ $\mathbb{P}-\text{a.s.}$ If $Y = \mathbb{1}_A$, then there exists an invariant random variable Z such that $\mathbb{1}_A = Z \;\; \mathbb{P}-\text{a.s.}$ The set $B = \{Z = 1\}$ is therefore invariant, and $\mathbb{1}_A = \mathbb{1}_B \;\; \mathbb{P}-\text{a.s.}$
□

It is easy to check that the family of $\mathbb{P}-\text{a.s.}$ invariant sets for T is a σ-algebra $\mathscr{I}_{\mathbb{P}}$. Lemma 5.1.7 shows that $\mathscr{I}_{\mathbb{P}}$ is the $\mathbb{P}$-completion of the invariant σ-algebra $\mathscr{I}$ (the σ-algebra generated by $\mathscr{I}$ and the family of sets that are $\mathbb{P}$-negligible).

Denote by T^n the transformation T iterated n times, and by convention, we let T^0 be the identity function. The behavior of time averages is given by the following fundamental result.

Theorem 5.1.8 (Birkhoff's ergodic theorem). *Let $(\Omega, \mathscr{B}, \mathbb{P}, \mathrm{T})$ be a dynamical system and Y a random variable such that $\mathbb{E}[|Y|] < \infty$. Then*

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y \circ \mathrm{T}^k = \mathbb{E}[Y \mid \mathscr{I}] \qquad \mathbb{P}-\text{a.s.} \tag{5.1.1}$$

Moreover, convergence also holds in $\mathrm{L}^1(\mathbb{P})$.

The proof is based on the following lemma.

Lemma 5.1.9 *Let Z be a random variable such that $\mathbb{E}[|Z|] < \infty$. If $\mathbb{E}[Z \mid \mathscr{I}] > 0 \;\; \mathbb{P}-\text{a.s.}$, then*

$$\liminf_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Z \circ \mathrm{T}^k \geq 0 \qquad \mathbb{P}-\text{a.s.}$$

Proof. For all $n \in \mathbb{N}^*$, write $S_n = \sum_{k=0}^{n-1} Z \circ \mathrm{T}^k$. Note that for all $n \geq 1$, $\mathbb{E}[|S_n|] \leq n\mathbb{E}[|Z|] < \infty$. Set $L_n = \inf\{S_k : 1 \leq k \leq n\}$ and $A = \{\inf_{n\in\mathbb{N}^*} L_n = -\infty\}$. Since

$|Z| < \infty$ $\mathbb{P}-$a.s., $\{\inf_{n\ge 1} S_n = -\infty\} = \{\inf_{n\ge 1} S_n \circ \mathrm{T} = -\infty\}$ $\mathbb{P}-$a.s., the set A is $\mathbb{P}-$a.s. invariant. Since $L_{n-1} \ge L_n$,

$$\begin{aligned} L_n &= Z + \inf\{S_k - Z \,:\, 1 \le k \le n\} \\ &= Z + \inf(0, L_{n-1} \circ \mathrm{T}) \ge Z + \inf(0, L_n \circ \mathrm{T})\,. \end{aligned} \tag{5.1.2}$$

Since $\mathbb{E}[|S_k|] < \infty$ for all $k \in \mathbb{N}$, for all $n \ge 1$, $\mathbb{E}[|L_n|] \le \sum_{k=0}^{n-1} \mathbb{E}[|S_k|] < \infty$, and (5.1.2) implies that $Z \le L_n + (L_n \circ \mathrm{T})^- = L_n + L_n^- \circ \mathrm{T}$ $\mathbb{P}-$a.s. Then, using $\mathbb{1}_A = \mathbb{1}_A \circ \mathrm{T}$ $\mathbb{P}-$a.s., we get

$$\begin{aligned} \mathbb{E}[\mathbb{1}_A Z] &\le \mathbb{E}[\mathbb{1}_A L_n] + \mathbb{E}\left[\mathbb{1}_A L_n^- \circ \mathrm{T}\right] = \mathbb{E}[\mathbb{1}_A L_n] + \mathbb{E}\left[\mathbb{1}_A \circ \mathrm{T}\; L_n^- \circ \mathrm{T}\right] \\ &\le \mathbb{E}[\mathbb{1}_A L_n] + \mathbb{E}\left[\mathbb{1}_A L_n^-\right] = \mathbb{E}\left[\mathbb{1}_A L_n^+\right]\,. \end{aligned} \tag{5.1.3}$$

Since $L_n^+ \le Z^+$ with $\mathbb{E}[Z^+] < \infty$ and $\lim_{n\to\infty} \mathbb{1}_A L_n^+ = 0$ $\mathbb{P}-$a.s., Lebesgue's dominated convergence theorem shows that $\mathbb{E}[\lim_{n\to\infty} \mathbb{1}_A L_n^+] = 0$. Therefore, since $0 \le \mathbb{E}[\mathbb{1}_A \mathbb{E}[Z \,|\, \mathscr{I}]] = \mathbb{E}[\mathbb{1}_A Z]$, we finally get, using (5.1.3),

$$\mathbb{E}[\mathbb{1}_A \mathbb{E}[Z \,|\, \mathscr{I}]] = \mathbb{E}[\mathbb{1}_A Z] \le \mathbb{E}\left[\lim_{n\to\infty} \mathbb{1}_A L_n^+\right] = 0\,.$$

By assumption, $\mathbb{E}[Z \,|\, \mathscr{I}] > 0$ $\mathbb{P}-$a.s., and so the previous inequality shows that $\mathbb{P}(A) = 0$. We conclude that $\liminf_{n\to\infty} n^{-1} S_n \ge 0$ $\mathbb{P}-$a.s.

□

Proof (of Theorem 5.1.8). Let $\varepsilon > 0$ and set $Z = Y - \mathbb{E}[Y \,|\, \mathscr{I}] + \varepsilon$. Note that $\mathbb{E}[|Z|] \le 2\mathbb{E}[|Y|] + \varepsilon$, showing that $\mathbb{E}[Z \,|\, \mathscr{I}]$ is welldefined and by construction, $\mathbb{E}[Z \,|\, \mathscr{I}] > 0$. Using that $\mathbb{E}[Y \,|\, \mathscr{I}]$ is $\mathscr{I}$-measurable, it is invariant according to Proposition 5.1.5, and Lemma 5.1.9 implies that

$$\liminf_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y \circ \mathrm{T}^k \ge \mathbb{E}[Y \,|\, \mathscr{I}] - \varepsilon \qquad \mathbb{P}-\text{a.s.}$$

Replacing Y by $-Y$, we finally obtain

$$-\varepsilon + \limsup_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y \circ \mathrm{T}^k \le \mathbb{E}[Y \,|\, \mathscr{I}] \le \liminf_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y \circ \mathrm{T}^k + \varepsilon \qquad \mathbb{P}-\text{a.s.}$$

This proves (5.1.1), since $\varepsilon > 0$ is arbitrary.

We now turn to the $\mathrm{L}^1(\mathbb{P})$ convergence. Define $M_n(Y) = n^{-1} \sum_{k=0}^{n-1} Y \circ \mathrm{T}^k$. If Y bounded, then Lebesgue's dominated convergence theorem shows that the convergence in (5.1.1) also holds in $\mathrm{L}^1(\mathbb{P})$. For a bounded random variable $\bar{Y}$, consider the decomposition

$$|M_n(Y) - \mathbb{E}[Y \,|\, \mathscr{I}]| \le |M_n(Y) - M_n(\bar{Y})| + |M_n(\bar{Y}) - \mathbb{E}[\bar{Y} \,|\, \mathscr{I}]| + \mathbb{E}[|\bar{Y} - Y| \,|\, \mathscr{I}]\,.$$

Let us define $\|U\|_1 = \mathbb{E}[|U|]$. Note that $\|\mathbb{E}[|\bar{Y} - Y| \,|\, \mathscr{I}]\|_1 \le \|\bar{Y} - Y\|_1$ and

$$\||M_n(Y) - M_n(\bar{Y})\|_1 \leq n^{-1} \sum_{k=0}^{n-1} \left\|(Y - \bar{Y}) \circ \mathrm{T}^k\right\|_1 = \|Y - \bar{Y}\|_1 \ ,$$

where we have used that the transformation T is measure-preserving and thus $\left\|(Y - \bar{Y}) \circ \mathrm{T}^k\right\|_1 = \|Y - \bar{Y}\|_1$ for all $k \in \mathbb{N}$. Therefore,

$$\limsup_{n\to\infty} \|M_n(Y) - \mathbb{E}\left[Y \,|\, \mathscr{I}\right]\|_1 \leq 2\,\|\bar{Y} - Y\|_1 \ .$$

The proof is complete, since bounded random variables are dense in $\mathrm{L}^1(\mathbb{P})$. □

The most interesting case of application of Theorem 5.1.8 is that in which the σ-field $\mathscr{I}$ is trivial, in which case the conditional expectation $\mathbb{E}\left[Y \,|\, \mathscr{I}\right]$ can be replaced by $\mathbb{E}\left[Y\right]$ in (5.1.1).

Definition 5.1.10 (Ergodic dynamical system) *A dynamical system* $(\Omega, \mathscr{B}, \mathbb{P}, \mathrm{T})$ *is ergodic if the invariant* σ*-field* $\mathscr{I}$ *is trivial for* $\mathbb{P}$*, i.e., for all* $A \in \mathscr{I}$*,* $\mathbb{P}(A) \in \{0,1\}$*.*

Corollary 5.1.11 *Let* $(\Omega, \mathscr{B}, \mathbb{P}, \mathrm{T})$ *be an ergodic dynamical system and* Y *an* $\bar{\mathbb{R}}$*-valued random variable such that* $\mathbb{E}\left[|Y|\right] < \infty$*. Then*

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y \circ \mathrm{T}^k = \mathbb{E}\left[Y\right] \qquad \mathbb{P}-\text{a.s.} \tag{5.1.4}$$

5.1.2.1 Dynamical Systems Associated with One-Sided and Two-Sided Sequences

In the context of Markov chains on a measurable space $(\mathsf{X}, \mathscr{X})$, the dynamical systems will be associated with either one-sided sequences $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}})$, or two-sided sequences $(\mathsf{X}^{\mathbb{Z}}, \mathscr{X}^{\otimes\mathbb{Z}})$. Denote by $\bar{\theta} : \mathsf{X}^{\mathbb{Z}} \to \mathsf{X}^{\mathbb{Z}}$ the shift operator on $\mathsf{X}^{\mathbb{Z}}$: for every two-sided sequence $\omega = (\omega_n)_{n\in\mathbb{Z}} \in \mathsf{X}^{\mathbb{Z}}$,

$$[\bar{\theta}(\omega)]_n = \omega_{n+1} \ , \qquad \text{for all } n \in \mathbb{Z} \ . \tag{5.1.5}$$

Moreover, set $\bar{\theta}_1 = \bar{\theta}$, and for all $n > 1$, $\bar{\theta}_n = \bar{\theta}_{n-1} \circ \bar{\theta}$. Let $\bar{\mathbb{P}}$ be a probability measure on $(\mathsf{X}^{\mathbb{Z}}, \mathscr{X}^{\otimes\mathbb{Z}})$ and denote by $\bar{\mathbb{E}}$ the associated expectation operator. Let Π be the measurable map from $(\mathsf{X}^{\mathbb{Z}}, \mathscr{X}^{\otimes\mathbb{Z}})$ to $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}})$ defined by

$$\Pi(\omega) = (\omega_n)_{n\in\mathbb{N}} \ , \qquad \text{for all } \omega = (\omega_n)_{n\in\mathbb{Z}} \in \mathsf{X}^{\mathbb{Z}} \ . \tag{5.1.6}$$

We denote by $\mathbb{P} = \bar{\mathbb{P}} \circ \Pi^{-1}$ the probability induced on $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}})$ by the probability $\bar{\mathbb{P}}$ and the map Π. Let θ be the shift operator on $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}})$ defined in Definition 3.1.8, and note that $\theta \circ \Pi = \Pi \circ \bar{\theta}$.

Lemma 5.1.12 *If* $(\mathsf{X}^{\mathbb{Z}}, \mathscr{X}^{\otimes\mathbb{Z}}, \bar{\mathbb{P}}, \bar{\theta})$ *is a dynamical system, then* $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}}, \mathbb{P}, \theta)$ *is also a dynamical system.*

Proof. By assumption, $\bar{\mathbb{P}} \circ \bar{\theta}^{-1} = \bar{\mathbb{P}}$. Combining with $\theta \circ \Pi = \Pi \circ \bar{\theta}$, this implies for all $A \in \mathscr{X}^{\otimes \mathbb{N}}$,

$$\begin{aligned} \mathbb{P} \circ \theta^{-1}(A) &= (\bar{\mathbb{P}} \circ \Pi^{-1}) \circ \theta^{-1}(A) = \bar{\mathbb{P}} \circ (\theta \circ \Pi)^{-1}(A) = \bar{\mathbb{P}} \circ (\Pi \circ \bar{\theta})^{-1}(A) \\ &= (\bar{\mathbb{P}} \circ \bar{\theta}^{-1}) \circ \Pi^{-1}(A) = \bar{\mathbb{P}} \circ \Pi^{-1}(A) = \mathbb{P}(A) \; . \end{aligned}$$

□

Proposition 5.1.13 *Assume that* $(\mathsf{X}^{\mathbb{Z}}, \mathscr{X}^{\otimes \mathbb{Z}}, \bar{\mathbb{P}}, \bar{\theta})$ *is a dynamical system. If the dynamical system* $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}}, \mathbb{P}, \theta)$ *is ergodic, then* $(\mathsf{X}^{\mathbb{Z}}, \mathscr{X}^{\otimes \mathbb{Z}}, \bar{\mathbb{P}}, \bar{\theta})$ *is ergodic.*

Proof. Let A be an invariant set for the dynamical system $(\mathsf{X}^{\mathbb{Z}}, \mathscr{X}^{\otimes \mathbb{Z}}, \bar{\mathbb{P}}, \bar{\theta})$, that is, $\mathbb{1}_A = \mathbb{1}_A \circ \bar{\theta}$. We will show that $\bar{\mathbb{P}}(A) = 0$ or 1.

Note first that $\mathscr{X}^{\otimes \mathbb{Z}} = \sigma(\mathscr{F}^+_{-k}, k \in \mathbb{N})$, where $\mathscr{F}^+_\ell = \sigma(X_i, \ell \le i < \infty)$ and $\{X_i, \, i \in \mathbb{Z}\}$ are the coordinate processes on $\mathsf{X}^{\mathbb{Z}}$. This allows us to apply the approximation lemma, Lemma B.2.5, showing that for all $\varepsilon > 0$, there exist $k_\varepsilon \in \mathbb{N}^*$ and an $\mathscr{F}^+_{-k_\varepsilon}$-measurable random variable Z_ε such that $\bar{\mathbb{E}}[|Z_\varepsilon|] < \infty$ and $\bar{\mathbb{E}}[|\mathbb{1}_A - Z_\varepsilon|] \le \varepsilon$. Set $Y_\varepsilon = Z_\varepsilon \circ \bar{\theta}_{k_\varepsilon}$. By construction, Y_ε is $\mathscr{F}^+_0$-measurable. Using that A is an invariant set, we obtain

$$\bar{\mathbb{E}}[|\mathbb{1}_A - Y_\varepsilon|] = \bar{\mathbb{E}}[|\mathbb{1}_A \circ \bar{\theta}_k - Z_\varepsilon \circ \bar{\theta}_k|] = \bar{\mathbb{E}}[|\mathbb{1}_A - Z_\varepsilon|] \le \varepsilon \; .$$

Since ε is arbitrary, there exists an $\mathscr{F}^+_0$-measurable random variable Y satisfying $\bar{\mathbb{E}}[|Y|] < \infty$ and $\mathbb{1}_A = Y$, $\bar{\mathbb{P}} -$ a.s. Since $1 = \bar{\mathbb{P}}(\mathbb{1}_A = Y) \le \bar{\mathbb{P}}(Y \in \{0,1\}) \le 1$, there exists $B \in \mathscr{F}^+_0$ such that

$$\mathbb{1}_B = Y = \mathbb{1}_A \, , \quad \bar{\mathbb{P}} - \text{a.s.} \tag{5.1.7}$$

Equation (5.1.7) and the invariance of A then show that

$$\bar{\mathbb{P}}(\mathbb{1}_B \circ \bar{\theta} = \mathbb{1}_A \circ \bar{\theta} = \mathbb{1}_A = \mathbb{1}_B) = 1 \; .$$

Now note that $\mathscr{F}^+_0 = \sigma(\Pi)$, the σ-algebra generated by Π, where the canonical projection $\Pi : \mathsf{X}^{\mathbb{Z}} \to \Omega$ is defined in (5.1.6). Then since $B \in \mathscr{F}^+_0$, there exists $C \in \mathscr{F}^+$ such that $B = \Pi^{-1}(C)$, and thus

$$\begin{aligned} 1 &= \bar{\mathbb{P}}(\mathbb{1}_B = \mathbb{1}_B \circ \bar{\theta}) \\ &= \bar{\mathbb{P}}(\mathbb{1}_C \circ \Pi = \mathbb{1}_C \circ \Pi \circ \bar{\theta}) \\ &\overset{(i)}{=} \bar{\mathbb{P}}(\mathbb{1}_C \circ \Pi = \mathbb{1}_C \circ \theta \circ \Pi) = \bar{\mathbb{P}} \circ \Pi^{-1}(\mathbb{1}_C = \mathbb{1}_C \circ \theta) \overset{(ii)}{=} \mathbb{P}(\mathbb{1}_C = \mathbb{1}_C \circ \theta) \, , \end{aligned}$$

where $\overset{(i)}{=}$ follows from $\Pi \circ \bar{\theta} = \theta \circ \Pi$ and $\overset{(ii)}{=}$ from $\mathbb{P} = \bar{\mathbb{P}} \circ \Pi^{-1}$. The dynamical system $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}}, \mathbb{P}, \theta)$ being ergodic, this implies that $\mathbb{P}(C) = 0$ or 1, which concludes the proof, since

$$\mathbb{P}(C) = \bar{\mathbb{P}} \circ \Pi^{-1}(C) = \bar{\mathbb{P}}(B) = \bar{\mathbb{P}}(A) .$$

□

Proposition 5.1.13 allows us to study the ergodicity of dynamical systems only on one-sided sequences (instead of two-sided sequences) and then to apply Birkhoff's ergodic theorem either to functions depending on the future $n^{-1}\sum_{k=0}^{n-1} f(\{X_{k+\ell},\ \ell \in \mathbb{N}\})$ or even on the whole past $n^{-1}\sum_{k=0}^{n-1} f(\{X_{k-\ell},\ \ell \in \mathbb{N}\})$. From now on, we consider only the ergodicity of dynamical systems associated to one-sided sequences.

5.2 Markov Chain Ergodicity

We specialize the results of the previous section in the context of Markov chains. Here and subsequently, we consider a Markov kernel P on a measurable space $(\mathsf{X}, \mathscr{X})$ and the coordinate process $\{X_k,\ k \in \mathbb{N}\}$ on the canonical space $(\Omega, \mathscr{F}) = (\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}})$, endowed with the family of probability measures $\mathbb{P}_\xi$, $\xi \in \mathbb{M}_1(\mathscr{X})$ under which the coordinate process is a Markov chain with kernel P and initial distribution ξ.

As a consequence of Birkhoff's ergodic theorem and Corollary 5.1.11, we obtain the ergodic theorem for Markov chains.

Theorem 5.2.1. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P admits an invariant probability measure π and that the associated dynamical system $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, \theta)$ is ergodic. Then for all random variables $Y \in \mathrm{L}^1(\mathbb{P}_\pi)$,*

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y \circ \theta_k = \mathbb{E}_\pi[Y] \qquad \mathbb{P}_\pi - \text{a.s.}$$

Moreover, convergence also holds in $\mathrm{L}^1(\mathbb{P}_\pi)$.

The condition $Y \in \mathrm{L}^1(\mathbb{P}_\pi)$ may be relaxed to $\mathbb{E}_\pi[Y^+] < \infty$, as shown in Exercise 5.6. We now relate harmonic functions (defined in Chapter 4) to invariant random variables for the shift transformation θ.

Proposition 5.2.2 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$.*

(i) Let Y be a bounded invariant random variable for the shift transformation θ. Then the function $h_Y : x \mapsto h_Y(x) = \mathbb{E}_x[Y]$ is a bounded harmonic function.

(ii) Let h be a bounded harmonic function and define $Y = \limsup_{n\to\infty} h(X_n)$. Then Y is an invariant random variable for θ, and for all $\xi \in \mathbb{M}_1(\mathscr{X})$, the

sequence $\{h(X_n),\ n \in \mathbb{N}\}$ *converges to* Y $\mathbb{P}_\xi - \text{a.s.}$ *and in* $\mathrm{L}^1(\mathbb{P}_\xi)$*. Moreover,* $h(x) = \mathbb{E}_x[Y]$ *for all* $x \in \mathsf{X}$.

(iii) Let π *be an invariant probability measure and* $Y \in \mathrm{L}^1(\mathbb{P}_\pi)$ *an invariant random variable for* θ*. Then* $\mathbb{E}_x[|Y|] < \infty$ $\pi - a.e.$*, the function* $x \mapsto \mathbb{E}_x[Y]$ *is* π*-integrable, and* $Y = \mathbb{E}_{X_0}[Y]$ $\mathbb{P}_\pi - \text{a.s.}$

Proof. (i) Assume that $Y : \mathsf{X}^{\mathbb{N}} \to \mathbb{R}$ is a bounded invariant random variable, i.e., $Y \circ \theta = Y$. By the Markov property, for all $x \in \mathsf{X}$,

$$Ph_Y(x) = \mathbb{E}_x[h_Y(X_1)] = \mathbb{E}_x[\mathbb{E}_{X_1}[Y]] = \mathbb{E}_x[Y \circ \theta_1] = \mathbb{E}_x[Y] = h_Y(x) ,$$

showing that h_Y is harmonic.

(ii) Let h be a bounded harmonic function: $Ph(x) = h(x)$ for all $x \in \mathsf{X}$. Then $\{(h(X_n), \mathscr{F}_n),\ n \in \mathbb{N}\}$ is a bounded $\mathbb{P}_\xi$-martingale, for every initial distribution $\xi \in \mathbb{M}_1(\mathscr{X})$. By Doob's martingale convergence theorem, the sequence $\{h(X_n), n \in \mathbb{N}\}$ converges $\mathbb{P}_\xi - \text{a.s.}$ and in $\mathrm{L}^1(\mathbb{P}_\xi)$ to a limit. Hence we get

$$Y = \lim_{n\to\infty} h(X_n)\mathbb{P}_\xi - \text{a.s. and } \mathbb{E}_\xi[Y] = \lim_{n\to\infty} \mathbb{E}_\xi[h(X_n)]. \tag{5.2.1}$$

The function h being harmonic, we have $h(x) = P^n h(x) = \mathbb{E}_x[h(X_n)]$ for all $x \in \mathsf{X}$ and $n \in \mathbb{N}$. Applying (5.2.1) with $\xi = \delta_x$ yields

$$\mathbb{E}_x[Y] = \lim_{n\to\infty} \mathbb{E}_x[h(X_n)] = h(x) .$$

(iii) Since $Y \in \mathrm{L}^1(\mathbb{P}_\pi)$, $\mathbb{E}_\pi[|Y|] = \int_{\mathsf{X}} \pi(\mathrm{d}x)\mathbb{E}_x[|Y|]$, showing that $\mathbb{E}_x[|Y|] < \infty$ π-*a.e.* and that the function $x \mapsto \mathbb{E}_x[Y]$ is integrable with respect to π. By the Markov property and the invariance of Y, we get

$$\mathbb{E}_{X_k}[Y] = \mathbb{E}_\pi\left[Y \circ \theta_k \middle| \mathscr{F}_k\right] = \mathbb{E}_\pi\left[Y \middle| \mathscr{F}_k\right] \qquad \mathbb{P}_\pi - \text{a.s.}$$

Therefore, $\{(\mathbb{E}_{X_k}[Y], \mathscr{F}_k), k \in \mathbb{N}\}$ is a uniformly integrable $\mathbb{P}_\pi$-martingale. By Theorem E.3.7,

$$\lim_{k\to\infty} \mathbb{E}_{X_k}[Y] = \lim_{k\to\infty} \mathbb{E}_\pi\left[Y \middle| \mathscr{F}_k\right] = \mathbb{E}_\pi\left[Y \middle| \mathscr{F}\right] = Y \qquad \mathbb{P}_\pi - \text{a.s.} \tag{5.2.2}$$

and in $\mathrm{L}^1(\mathbb{P}_\pi)$. Moreover, applying successively that the translation operator θ is measure-preserving for $\mathbb{P}_\pi$ and $Y = Y \circ \theta_k$, we obtain for all $k \in \mathbb{N}$,

$$\begin{aligned}\mathbb{E}_\pi[|Y - \mathbb{E}_{X_0}[Y]|] &= \mathbb{E}_\pi[|Y - \mathbb{E}_{X_0}[Y]| \circ \theta_k] \\ &= \mathbb{E}_\pi[|Y \circ \theta_k - \mathbb{E}_{X_k}[Y]|] = \mathbb{E}_\pi[|Y - \mathbb{E}_{X_k}[Y]|] .\end{aligned}$$

On taking the limit as k goes to infinity, (5.2.2) yields

$$\mathbb{E}_\pi[|Y - \mathbb{E}_{X_0}[Y]|] = \lim_{k\to\infty} \mathbb{E}_\pi[|Y - \mathbb{E}_{X_k}[Y]|] = 0 .$$

□

Remark 5.2.3. Proposition 5.2.2 shows that the map $Y \mapsto h_Y$, where $h_Y(x) = \mathbb{E}_x[Y]$, $x \in \mathsf{X}$, defines a one-to-one correspondence between the bounded harmonic functions and the bounded invariant random variables. If Y is a bounded invariant random variable, then $h_Y : x \mapsto h_Y(x) = \mathbb{E}_x[Y]$ is a bounded harmonic function. If h is a bounded harmonic function, then $h(x) = \mathbb{E}_x[Y]$, where $Y = \limsup_{n\to\infty} h(X_n)$, is an invariant random variable (hence $h = h_Y$). ▲

Corollary 5.2.4 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. The following statements are equivalent.*

(i) The bounded harmonic functions are constant.
(ii) The invariant σ-field $\mathscr{I}$ is trivial up to an equivalence, i.e., for all $A \in \mathscr{I}$, we get for all $\xi \in \mathbb{M}_1(\mathscr{X})$, $\mathbb{P}_\xi(A) = 0$ or $\mathbb{P}_\xi(A) = 1$.

Proof. (i) ⇒ (ii): Let A be an invariant set. Then $h_A : x \mapsto h_A(x) = \mathbb{P}_x(A)$ is a harmonic function by Proposition 5.2.2, which is constant under (i), i.e., $h_A(x) = c$ for all $x \in \mathsf{X}$. By the Markov property, we get that for all $\xi \in \mathbb{M}_1(\mathscr{X})$, $h_A(X_n) = \mathbb{E}_{X_n}[\mathbb{1}_A] = \mathbb{E}_\xi[\mathbb{1}_A \circ \theta_n \,|\, \mathscr{F}_n]$ $\mathbb{P}_\xi$ − a.s. Since $A \in \mathscr{I}$, it follows that $\mathbb{E}_\xi[\mathbb{1}_A \circ \theta_n \,|\, \mathscr{F}_n] = \mathbb{E}_\xi[\mathbb{1}_A \,|\, \mathscr{F}_n]$ $\mathbb{P}_\xi$ − a.s., and Theorem E.3.7 shows that $\mathbb{E}_\xi[\mathbb{1}_A \,|\, \mathscr{F}_n] \xrightarrow{\mathbb{P}_\xi\text{-a.s.}} \mathbb{1}_A$ $\mathbb{P}_\xi$ − a.s., which implies that $c \in \{0,1\}$.

(ii) ⇒ (i): Let h be a bounded harmonic function and $\xi \in \mathbb{M}_1(\mathscr{X})$. The random variable $Y = \limsup_{n\to\infty} h(X_n)$ is invariant, and under (ii), there exists a constant $c < \infty$ (possibly depending on ξ) such that $\limsup_{n\to\infty} h(X_n) = c$ $\mathbb{P}_\xi$ − a.s. By Proposition 5.2.2, $\mathbb{E}_x[Y] = c = h(x)$ for all $x \in \mathsf{X}$. Therefore, h is constant, which establishes (i). □

If we wish to obtain the law of large numbers for a particular Markov chain by applying Theorem 5.2.1, we have to check the ergodicity assumption. It is therefore convenient to have sufficient conditions ensuring ergodicity.

We now give a sufficient condition for $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}}, \mathbb{P}_\pi, \theta)$ to be ergodic expressed in terms of absorbing sets.

Lemma 5.2.5 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an invariant probability measure π. If for all absorbing sets $B \in \mathscr{X}$, $\pi(B) \in \{0,1\}$, then the dynamical system $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}}, \mathbb{P}_\pi, \theta)$ is ergodic.*

Proof. Let $A \in \mathscr{I}$ and define $h(x) = \mathbb{E}_x[\mathbb{1}_A]$ and $B = \{x \in \mathsf{X} : h(x) = 1\}$. By Proposition 5.2.2 (i), h is a nonnegative harmonic function bounded by 1. For all $x \in B$, we have $\mathbb{E}_x[h(X_1)] = Ph(x) = h(x) = 1$, which implies $\mathbb{P}_x(h(X_1) = 1) = 1$. Therefore, for all $x \in B$, we get $\mathbb{P}_x(X_1 \in B) = \mathbb{P}_x(h(X_1) = 1) = 1$. Therefore, B is absorbing, and hence under the stated assumption, we have $\pi(B) \in \{0,1\}$.

By Proposition 5.2.2 (iii), we know that $\mathbb{P}_\pi(\mathbb{E}_{X_0}[\mathbb{1}_A] = \mathbb{1}_A) = 1$, which implies that $\mathbb{P}_\pi(h(X_0) \in \{0,1\}) = 1$. This yields

$$\begin{aligned}\mathbb{P}_\pi(A) = \mathbb{E}_\pi[\mathbb{E}_{X_0}[\mathbb{1}_A]] &= \int_{\mathsf{X}} \pi(\mathrm{d}x) h(x) \\ &= \int_{\mathsf{X}} \pi(\mathrm{d}x) \mathbb{1}\{h(x) = 1\} = \int_{\mathsf{X}} \pi(\mathrm{d}x) \mathbb{1}_B(x) = \pi(B) .\end{aligned}$$

Thus $\mathbb{P}_\pi(A) \in \{0,1\}$, and $\mathscr{I}$ is trivial for $\mathbb{P}_\pi$. □

It turns out that the sufficient condition in Lemma 5.2.5 is also a necessary condition. Before showing the necessary part, we first draw an easy and useful consequence of Lemma 5.2.5.

Theorem 5.2.6. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting a unique invariant probability measure π. The dynamical system $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, \theta)$ is ergodic.*

Proof. By Proposition 1.4.5, since P has unique invariant probability π, it follows that for every absorbing set B, $\pi(B) \in \{0,1\}$. We conclude by Lemma 5.2.5. □

The uniqueness of the invariant probability measure is a sufficient but not a necessary condition for ergodicity, as illustrated in Exercise 5.8. Compared to Lemma 5.2.5, the following lemma goes one step further. When the dynamical system is not ergodic, the state space X contains not only one but at least two disjoint absorbing sets that are not trivial with respect to π.

Lemma 5.2.7 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an invariant probability measure π. If the dynamical system $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, \theta)$ is not ergodic, then there exist two disjoint absorbing sets B and B' in $\mathscr{X}$ such that $\pi(B) = 1 - \pi(B') \in (0,1)$, and $\pi_B(\cdot) = \pi(B \cap \cdot)/\pi(B)$ and $\pi_{B'}(\cdot) = \pi(B' \cap \cdot)/\pi(B')$ are invariant probability measures.*

Proof. Since the dynamical system $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, \theta)$ is not ergodic, there exists $A \in \mathscr{I}$ such that $\mathbb{P}_\pi(A) = \alpha \in (0,1)$. Since $\mathscr{I}$ is a σ-field, we also have $A^c \in \mathscr{I}$. Define $B = \{x \in \mathsf{X} : \mathbb{E}_x[\mathbb{1}_A] = 1\}$ and $B' = \{x \in \mathsf{X} : \mathbb{E}_x[\mathbb{1}_{A^c}] = 1\}$. As noted in the proof of Lemma 5.2.5 (see also Exercise 5.4), the sets B and B' are absorbing and

$$\pi(B) = \mathbb{P}_\pi(A) = 1 - \mathbb{P}_\pi(A^c) = 1 - \pi(B') \in (0,1) .$$

By Proposition 1.4.5, π_B and $\pi_{B'}$ are invariant probability measures. □

Without ergodicity assumption, the generalized version of Birkhoff's ergodic theorem in Theorem 5.1.8 shows that the normalized partial sums still converge, but the limit is a random variable that is not necessarily almost surely constant (see also an illustration in Exercise 5.8). In the context of Markov chains, this limit turns out to be a function of X_0. More precisely, we have the following theorem.

Proposition 5.2.8 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an invariant probability measure π and let $Y \in \mathrm{L}^1(\mathbb{P}_\pi)$. Let Z be a version of the conditional expectation $\mathbb{E}_\pi[Y \mid \mathscr{I}]$, i.e., $Z = \mathbb{E}_\pi[Y \mid \mathscr{I}]$ $\mathbb{P}_\pi - \text{a.s.}$ Then there exists a set $S \in \mathscr{X}$ such that $\pi(S) = 1$ and for each $x \in S$,*

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y \circ \theta_k = \mathbb{E}_x[Z] \qquad \mathbb{P}_x - \text{a.s.} \tag{5.2.3}$$

Proof. Define $\phi(x) = \mathbb{E}_x[Z]$. It follows from Proposition 5.2.2 (iii) that $\mathbb{E}_\pi[Y \mid \mathscr{I}] = \phi(X_0)$, $\mathbb{P}_\pi - \text{a.s.}$ Hence Theorem 5.1.8 yields

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} Y \circ \theta_k = \mathbb{E}_\pi[Y|\mathscr{I}] = \phi(X_0) \qquad \mathbb{P}_\pi - \text{a.s.}$$

Set $A = \left\{\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} Y \circ \theta_k = \phi(X_0)\right\}$. The previous relation implies $\mathbb{P}_\pi(A) = 1$, i.e., $\int \pi(\mathrm{d}x)\mathbb{P}_x(A) = 1$. Since $\mathbb{P}_x(A) \leq 1$ for all $x \in \mathsf{X}$, this implies that $\mathbb{P}_x(A) = 1$ for π-almost all $x \in \mathsf{X}$. Setting $S = \{x \in \mathsf{X} : \mathbb{P}_x(A) = 1\}$ concludes the proof. □

In Proposition 5.2.8, the limit $\mathbb{E}_x[Z]$ in (5.2.3) is expressed in terms of Z, a version of the conditional expectation $\mathbb{E}_\pi[Y \mid \mathscr{I}]$. If we choose another version of $\mathbb{E}_\pi[Y \mid \mathscr{I}]$, say Z', under $\mathbb{P}_\pi$, then obviously, $Z = Z'$ $\mathbb{P}_\pi - \text{a.s.}$, but we do not necessarily have $\mathbb{E}_x[Z'] = \mathbb{E}_x[Z]$ $\mathbb{P}_x - \text{a.s.}$, since without additional assumptions, $\mathbb{P}_x$ is not necessarily dominated by $\mathbb{P}_\pi$. The situation is different when the dynamical system is ergodic, since the limit is then $\mathbb{P}_\pi - \text{a.s.}$ constant.

Theorem 5.2.9 (Birkhoff's theorem for Markov chains). *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and assume that P admits an invariant probability measure π such that $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}}, \mathbb{P}_\pi, \theta)$ is ergodic. Let $Y \in \mathrm{L}^1(\mathbb{P}_\pi)$. Then for π-almost all $x \in \mathsf{X}$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} Y \circ \theta_k = \mathbb{E}_\pi[Y] \qquad \mathbb{P}_x - \text{a.s.}$$

Proof. Since π is ergodic, the invariant σ-field $\mathscr{I}$ is trivial for $\mathbb{P}_\pi$. This implies $\mathbb{E}_\pi[Y \mid \mathscr{I}] = \mathbb{E}_\pi[Y]$ $\mathbb{P}_\pi - \text{a.s.}$ □

Theorem 5.2.10. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. If π_1 and π_2 are distinct invariant probability measures such that $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}}, \mathbb{P}_{\pi_1}, \theta)$ and $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}}, \mathbb{P}_{\pi_2}, \theta)$ are ergodic, then π_1 and π_2 are mutually singular.*

Proof. Note first that if $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}}, \mathbb{P}_\pi, \theta)$ is ergodic and $f \in \mathbb{F}_b(\mathsf{X})$, then applying Theorem 5.2.9 to the random variable $Y = f(X_0)$ and invoking the dominated convergence theorem, we obtain that there exists a set $S \in \mathscr{X}$ such that $\pi(S) = 1$ and for all $x \in S$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P^k f(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{E}_x[f(X_k)] = \pi(f) \; .$$

Now assume that π_1 and π_2 are distinct invariant probability measures such that $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}}, \mathbb{P}_{\pi_1}, \theta)$ and $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}}, \mathbb{P}_{\pi_2}, \theta)$ are ergodic. Let $C \in \mathscr{X}$ be such that $\pi_1(C) \neq \pi_2(C)$, and set, for $i = 1, 2$,

$$S_i = \left\{ x \in \mathsf{X} : \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} P^k \mathbb{1}_C(x) = \pi_i(C) \right\} .$$

We have $S_1 \cap S_2 = \emptyset$, $\pi_1(S_1) = 1$, and $\pi_2(S_2) = 1$, which means that π_1 and π_2 are mutually singular. □

We have now all the tools for obtaining a necessary and sufficient condition for the dynamical system to be ergodic.

Theorem 5.2.11. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an invariant probability measure π. The dynamical system $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}}, \mathbb{P}_\pi, \theta)$ is ergodic if and only if for all absorbing sets $B \in \mathscr{X}$, $\pi(B) \in \{0, 1\}$.*

Proof. The sufficient condition follows from Lemma 5.2.5. We now consider an ergodic dynamical system $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}}, \mathbb{P}_\pi, \theta)$, and we let B be an absorbing set. Assume first that $\pi(B) > 0$. Then by Proposition 1.4.5, $\bar{\pi}_B(\cdot) = \pi(B \cap \cdot)/\pi(B)$ is an invariant probability measure. Moreover, note that for all $A \in \mathscr{X}^{\otimes\mathbb{N}}$,

$$\mathbb{P}_{\bar{\pi}_B}(A) = \int \pi(\mathrm{d}x) \mathbb{P}_x(A) \frac{\mathbb{1}_B(x)}{\pi(B)} \leq \frac{\mathbb{P}_\pi(A)}{\pi(B)} \; .$$

Combining with the ergodicity of the dynamical system $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}}, \mathbb{P}_\pi, \theta)$, we deduce that every invariant set $A \in \mathscr{I}$ satisfies either $0 = \mathbb{P}_\pi(A) = \mathbb{P}_{\bar{\pi}_B}(A)$ or $0 = \mathbb{P}_\pi(A^c) = \mathbb{P}_{\bar{\pi}_B}(A^c)$. The dynamical system $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}}, \mathbb{P}_{\bar{\pi}_B}, \theta)$ is therefore ergodic, and by Theorem 5.2.10, $\bar{\pi}_B = \pi$, since they are not mutually singular. This implies that $\pi(B^c) = \bar{\pi}_B(B^c) = 0$. Finally, $\pi(B) \in \{0, 1\}$, which concludes the proof. □

Proposition 5.2.12 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting a unique invariant probability measure π, and let $h \in \mathrm{L}^1(\pi)$ be a harmonic function. Then $h(x) = \pi(h)$ for π-almost every $x \in \mathsf{X}$.*

Proof. Since $h \in \mathrm{L}^1(\pi)$ is a harmonic function, Theorem 4.1.2 shows that the process $\{h(X_n),\, n \in \mathbb{N}\}$ is a $\mathbb{P}_\pi$-martingale. Moreover, $\sup_n \mathbb{E}_\pi[|h(X_n)|] = \pi(|h|) < \infty$. Theorem E.3.1 then shows that this martingale is $\mathbb{P}_\pi -$a.s. convergent. If h is not $\pi -$a.s. constant, then there exists $a < b$ such that $\pi(\{h \le a\}) > 0$ and $\pi(\{h \ge b\}) > 0$. Theorem 5.2.9 applied to $Y = \mathbb{1}\{h(X_0) \le a\}$ and $Y = \mathbb{1}\{h(X_0)) \ge b\}$ implies that the sequence $\{X_n,\, n \in \mathbb{N}\}$ visits $\mathbb{P}_\pi -$a.s. the sets $\{h < a\}$ and $\{h > b\}$ infinitely often. This contradicts the $\mathbb{P}_\pi -$a.s. convergence of $\{h(X_n),\, n \in \mathbb{N}\}$. □

Corollary 5.2.13 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting a unique invariant probability measure π. Let $A \in \mathscr{X}$ be such that $\pi(A) > 0$. Then, $\mathbb{P}_x(N_A = \infty) = 1$, for π-almost every $x \in \mathsf{X}$.*

Proof. Proposition 4.2.4 shows that the function $h(x) = \mathbb{P}_x(N_A = \infty)$ is harmonic. The result follows from Proposition 5.2.12 by noting that if $\pi(A) > 0$, then $\mathbb{P}_\pi(N_A = \infty) = 1$, since $n^{-1}\sum_{k=0}^{n-1} \mathbb{1}_A(X_k) = \pi(A) > 0 \qquad \mathbb{P}_\pi -$a.s. □

In Proposition 5.2.8 and Theorem 5.2.9, the law of large numbers is obtained under $\mathbb{P}_x$ for all x belonging to a set S such that $\pi(S) = 1$ and that may depend on the random variable Y under consideration. This is unsatisfactory, since it does not tell whether the law of large numbers holds for a given $x \in \mathsf{X}$ or more generally for a given initial distribution ξ. We now give a criterion to obtain the law of large numbers when the chain does not start from stationarity. Recall that the total variation distance between two probability measures $\mu, \nu \in \mathbb{M}_1(\mathscr{X})$ is defined by

$$\|\mu - \nu\|_{\mathrm{TV}} = \sup_{h \in \mathbb{F}_b(\mathsf{X}), |h|_\infty \le 1} |\mu(h) - \nu(h)| \,.$$

More details and basic properties on the total variation distance are given in Appendix D.2.

Proposition 5.2.14 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an invariant probability measure π. If the dynamical system $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, \theta)$ is ergodic and if $\xi \in \mathbb{M}_1(\mathscr{X})$ is such that $\lim_{n\to\infty} \left\| n^{-1}\sum_{k=1}^n \xi P^k - \pi \right\|_{\mathrm{TV}} = 0$, then for all $Y \in \mathrm{L}^1(\mathbb{P}_\pi)$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} Y \circ \theta_k = \mathbb{E}_\pi[Y] \qquad \mathbb{P}_\xi -\text{a.s.}$$

Proof. Set $A = \{\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} Y \circ \theta_k = \mathbb{E}_\pi[Y]\}$. Since $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, \theta)$ is ergodic, we already know that $\mathbb{P}_\pi(A) = 1$. We show that $\mathbb{P}_\xi(A) = 1$. Define the function h by $h(x) = \mathbb{E}_x[\mathbb{1}_A]$. Since $A \in \mathscr{I}$, Proposition 5.2.2 implies that h is harmonic. Then for all $n \in \mathbb{N}$, $n^{-1}\sum_{k=1}^n \xi P^k h = \xi(h)$. Moreover, noting that $h \le 1$, we have, by assumption, $\lim_{n\to\infty} n^{-1}\sum_{k=1}^n \xi P^n(h) = \pi(h)$. Thus $\xi(h) = \pi(h)$ and

$$\mathbb{P}_\xi(A) = \int_{\mathsf{X}} \xi(\mathrm{d}x)\mathbb{E}_x[\mathbb{1}_A] = \xi(h) = \pi(h) = \mathbb{P}_\pi(A) = 1 \;.$$

□

The condition $\lim_n \|\xi P^n - \pi\|_{\mathrm{TV}} = 0$ is not mandatory for having a law of large numbers under $\mathbb{P}_\xi$. In some situations, one can get the same result without any straightforward information in the decrease of $\|\xi P^n - \pi\|_{\mathrm{TV}}$ toward 0. This is the case, for example, for Metropolis–Hastings kernels, as illustrated in Exercise 5.9. Another illustration can be found in Exercise 5.14, where the law of large numbers is extended to different initial distributions in the case in which (X, d) is a complete separable metric space.

5.3 Exercises

5.1. Let $(\Omega, \mathscr{B}, \mathbb{P}, \mathrm{T})$ be a dynamical system. Show that $\mathscr{I} \neq \cap_{k \geq 0} \sigma(X_l, l > k)$.

5.2. Let $(\Omega, \mathscr{B}, \mathbb{P})$ be a probability space and $\theta : \Omega \to \Omega$ a measurable transformation. Let $\mathscr{B}_0$ be a family of sets, stable under finite intersection and generating $\mathscr{B}$. If for all $B \in \mathscr{B}_0$, $\mathbb{P}[\theta^{-1}(B)] = \mathbb{P}(B)$, then T is measure-preserving.

5.3. Let $(\Omega, \mathscr{B}, \mathbb{P}, \mathrm{T})$ be a dynamical system. Let Y be an $\bar{\mathbb{R}}$-valued random variable such that $\mathbb{E}[Y^+] < \infty$. Show that for all $k \geq 0$,

$$\mathbb{E}\left[Y \circ \mathrm{T}^k \,\middle|\, \mathscr{I} \right] = \mathbb{E}[Y \,|\, \mathscr{I}] \qquad \mathbb{P} - \text{a.s.}$$

5.4. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let $(\Omega, \mathscr{B})$ be the canonical space. For $A \in \mathscr{I}$, define $B = \{x \in \mathsf{X} : \mathbb{P}_x(A) = 1\}$.

1. Show that B is absorbing.
2. Let π be an invariant probability. Show that $\pi(A) = \mathbb{P}_\pi(B)$.

5.5. The following exercise provides the converse of Theorem 5.2.1. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let π be a probability measure, $\pi \in \mathbb{M}_1(\mathscr{X})$. Assume that for all $f \in \mathbb{F}_b(\mathsf{X})$, we get

$$\lim_{n \to \infty} n^{-1} \sum_{k=0}^{n-1} f(X_k) = \pi(f) \qquad \mathbb{P}_\pi - \text{a.s.}$$

1. Show that π is invariant.
2. Let $A \in \mathscr{I}$ and set $B = \{x \in \mathsf{X} : \mathbb{P}_x(A) = 1\}$. Show that
$$\mathbb{1}_A = \mathbb{P}_{X_0}(A) = \mathbb{1}_B(X_0) \qquad \mathbb{P}_\pi - \text{a.s.}$$
and that for all $k \in \mathbb{N}$, $\mathbb{1}_A = \mathbb{1}_B(X_0) = \cdots = \mathbb{1}_B(X_k)$ $\mathbb{P}_\pi - \text{a.s.}$
3. Show that the dynamical system $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, \theta)$ is ergodic.

5.6. In this exercise, we prove various extensions of Birkhoff's ergodic theorem. Let $(\Omega, \mathscr{B}, \mathbb{P}, \mathrm{T})$ be a dynamical system. In the first two questions, we assume that the dynamical system is ergodic.

1. Let Y be a nonnegative random variable such that $\mathbb{E}[Y] = \infty$. Show that

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y \circ \mathrm{T}^k = \infty \qquad \mathbb{P}-\text{a.s.}$$

[Hint: Use Corollary 5.1.11 with $Y_M = Y \wedge M$ and let M tend to infinity.]

2. Let Y be a random variable such that $\mathbb{E}[Y^+] < \infty$. Show that

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y \circ \mathrm{T}^k = \mathbb{E}[Y] \qquad \mathbb{P}-\text{a.s.}$$

3. In what follows, we do not assume any ergodicity of the dynamical system $(\Omega, \mathscr{B}, \mathbb{P}, \mathrm{T})$. Let Y be a nonnegative random variable. Set

$$A = \{\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y \circ \mathrm{T}^k = \mathbb{E}[Y \mid \mathscr{I}]\}$$

Let $M > 0$. Using Theorem 5.1.8 with $Y\mathbb{1}\{\mathbb{E}[Y \mid \mathscr{I}] \le M\}$, show that on $\{\mathbb{E}[Y \mid \mathscr{I}] \le M\}$,

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y \circ \mathrm{T}^k = \mathbb{E}[Y \mid \mathscr{I}] \qquad \mathbb{P}-\text{a.s.}$$

Deduce that $\mathbb{P}(A^c \cap \{\mathbb{E}[Y \mid \mathscr{I}] < \infty\}) = 0$. Moreover, show that for all $M > 0$,

$$\liminf_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y \circ \mathrm{T}^k \ge \mathbb{E}[Y \wedge M \mid \mathscr{I}] \qquad \mathbb{P}-\text{a.s.}$$

Deduce that $\mathbb{P}(A^c \cap \{\mathbb{E}[Y \mid \mathscr{I}] = \infty\}) = 0$.

4. Let Y be a random variable such that $\mathbb{E}[Y^+] < \infty$. Show that

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y \circ \mathrm{T}^k = \mathbb{E}[Y \mid \mathscr{I}] \qquad \mathbb{P}-\text{a.s.}$$

[Hint: Write $Y = Y^+ - Y^-$ and use the previous question with Y^-.]

5.7. Let P be the kernel on $\mathsf{X} \times \mathscr{X}$ defined for all $(x,A) \in \mathsf{X} \times \mathscr{X}$, $P(x,A) = \delta_x(A)$. Find all the probability measures $\pi \in \mathbb{M}_1(\mathscr{X})$ such that $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, \theta)$ is ergodic.

5.8. Let μ_0 (resp. μ_1) be a probability measure on $\mathbb{R}^+$ (resp. $\mathbb{R}^- \setminus \{0\}$). Let P be a Markov kernel on $\mathbb{R} \times \mathscr{B}(\mathbb{R})$ defined by $P(x,\cdot) = \mu_0$ if $x \ge 0$ and $P(x,\cdot) = \mu_1$ otherwise. Set for all $\alpha \in (0,1)$, $\mu_\alpha = (1-\alpha)\mu_0 + \alpha\mu_1$.

1. Show that for all $\alpha \in (0,1)$, μ_α is an invariant probability measure but the dynamical system $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}}, \mathbb{P}_{\mu_\alpha}, \theta)$ is not ergodic.
2. Show that for $\alpha \in \{0,1\}$, $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}}, \mathbb{P}_{\mu_\alpha}, \theta)$ is ergodic.
3. Let $f \in \mathrm{L}^1(\mu_0) \cap \mathrm{L}^1(\mu_1)$. Find a function ϕ such that for all $x \in \mathbb{R}$,

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} f(X_k) = \phi(X_0) \qquad \mathbb{P}_{\mu_\alpha} - \text{a.s.}$$

   For all $\alpha \in (0,1)$, find a version of $\mathbb{E}_{\mu_\alpha}[f|\mathscr{I}]$.

5.9. In this exercise, we will find a sufficient condition for a Metropolis–Hastings kernel to satisfy the law of large numbers starting from any initial distribution. We make use of the notation of Section 2.3.1. Let π be a target distribution on a measurable space $(\mathsf{X}, \mathscr{X})$ and assume that π has a positive density h with respect to a measure $\mu \in \mathbb{M}_+(\mathscr{X})$. Let Q be a proposal kernel on $\mathsf{X} \times \mathscr{X}$ and assume that Q has a positive kernel density $y \mapsto q(x,y)$ with respect to μ. The Metropolis–Hastings kernel P is then defined by (2.3.4).

1. Show that
$$P(x,A) \geq \int_A \alpha(x,y) q(x,y) \mu(\mathrm{d}y) .$$
   Deduce that π is the unique invariant probability of P.
2. Let $A \in \mathscr{I}$ and assume that $\mathbb{P}_\pi(A) = 0$. Set $\phi(x) = \mathbb{P}_x(A)$. For all $x \in \mathsf{X}$, show that
$$\phi(x) = P\phi(x) = \int \frac{\alpha(x,y)q(x,y)}{h(y)} \pi(\mathrm{d}y)\phi(y) + \bar{\alpha}(x)\phi(x) ,$$
   and deduce that $\mathbb{P}_x(A) = 0$.
3. Let $\xi \in \mathbb{M}_1(\mathscr{X})$. Deduce from the previous question that for all random variables $Y \in \mathrm{L}^1(\mathbb{P}_\pi)$,

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y \circ \theta_k = \mathbb{E}_\pi[Y] \qquad \mathbb{P}_\xi - \text{a.s.}$$

5.10. Let P a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an invariant probability measure π_1 such that $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}}, \mathbb{P}_{\pi_1}, \theta)$ is ergodic. Let μ be another invariant probability measure.

1. Show that $\pi_1 \wedge \mu$ is an invariant finite measure.
2. If $\pi_1 \wedge \mu(\mathsf{X}) \neq 0$, show, using Theorem 5.2.10, that $\pi_1 \wedge \mu / \pi_1 \wedge \mu(\mathsf{X}) = \pi_1$.
3. Show that there exists an invariant probability measure π_2 satisfying that π_1 and π_2 are mutually singular and there exists $\alpha \in [0,1]$ such that $\mu = \alpha\pi_1 + (1-\alpha)\pi_2$.

5.11. Let $\{X_n,\ n \in \mathbb{Z}\}$ be a canonical stationary Markov chain on $(\mathsf{X}^{\mathbb{Z}}, \mathscr{X}^{\otimes \mathbb{Z}})$. We set
$$\overline{\mathscr{F}^0_{-\infty}} = \overline{\sigma(X_k, k \leq 0)}^P , \quad \overline{\mathscr{F}^\infty_0} = \overline{\sigma(X_k, k \geq 0)}^P .$$
We consider an invariant bounded random variable Y.

(i) Show that Y is $\overline{\mathscr{F}_0^\infty}$ and $\overline{\mathscr{F}_{-\infty}^0}$ measurable.
(ii) Deduce from the previous question that $Y = \mathbb{E}\left[Y \,|\, X_0\right] \quad \mathbb{P}-\text{a.s.}$
(iii) Show that the previous identity holds for every $\mathbb{P}$-integrable or positive invariant random variable Y.

The following exercises deal with subadditive sequences. A sequence of random variables $\{Y_n,\ n \in \mathbb{N}^\star\}$ is said to be subadditive for the dynamical system $(\Omega, \mathscr{B}, \mathbb{P}, \mathrm{T})$ if for all $(n,p) \in \mathbb{N}^\star$, $Y_{n+p} \le Y_n + Y_p \circ \mathrm{T}^n$. The sequence is said to be additive if for all $(n,p) \in \mathbb{N}^\star$, $Y_{n+p} = Y_n + Y_p \circ \mathrm{T}^n$.

5.12 (Fekete's lemma). Consider $\{a_n,\ n \in \mathbb{N}^\star\}$, a sequence in $[-\infty, \infty)$ such that, for all $(m,n) \in \mathbb{N}^\star \times \mathbb{N}^\star$, $a_{n+m} \le a_n + a_m$. Then

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{m \in \mathbb{N}^\star} \frac{a_m}{m} \;;$$

in other words, the sequence $\{n^{-1} a_n, n \in \mathbb{N}^\star\}$ either converges to its lower bound or diverges to $-\infty$.

5.13. Let $(\Omega, \mathscr{B}, \mathbb{P}, \mathrm{T})$ be a dynamical system and $\{Y_n,\ n \in \mathbb{N}^\star\}$ a subadditive sequence of functions such that $\mathbb{E}\left[Y_1^+\right] < \infty$. Show that for all $n \in \mathbb{N}^\star$, $\mathbb{E}\left[Y_n^+\right] \le n\mathbb{E}\left[Y_1^+\right] < \infty$ and

$$\lim_{n\to\infty} n^{-1}\mathbb{E}\left[Y_n\right] = \inf_{n \in \mathbb{N}^\star} n^{-1}\mathbb{E}\left[Y_n\right] \,, \tag{5.3.1}$$

$$\lim_{n\to\infty} n^{-1}\mathbb{E}\left[Y_n \,|\, \mathscr{I}\right] = \inf_{n \in \mathbb{N}^\star} n^{-1}\mathbb{E}\left[Y_n \,|\, \mathscr{I}\right] \,, \quad \mathbb{P}-\text{a.s.}, \tag{5.3.2}$$

where $\mathscr{I}$ is the invariant σ-field. [Hint: Use Exercise 5.3.]

5.14. Let P be a Markov kernel on a complete separable metric space (X, d) that admits a unique invariant probability measure π. We assume that there exist a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a stochastic process $\{(X_n, X_n'),\ n \in \mathbb{N}\}$ such that $\{X_n,\ n \in \mathbb{N}\}$ and $\{X_n',\ n \in \mathbb{N}\}$ are Markov chains with kernel P and initial distribution $\xi \in \mathbb{M}_1(\mathscr{X})$ and π, respectively. Assume that $\mathsf{d}(X_n, X_n') \xrightarrow{\mathbb{P}\text{-a.s.}} 0$.

1. Show that there exists a countable set $H \subset \mathrm{U}_b(\mathsf{X})$ such that $\mathbb{P}(A) = 1$, where

$$A = \left\{ \omega \in \Omega \,:\, \forall h \in H \,,\ \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} h(X_k'(\omega)) = \pi(h) \right\} .$$

2. Deduce that there exists a set $\tilde{\Omega}$ such that $\mathbb{P}(\tilde{\Omega}) = 1$ and for all bounded continuous functions h on X and all $\tilde{\omega} \in \tilde{\Omega}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} h(X_k(\tilde{\omega})) = \pi(h) .$$

3. Let V be a nonnegative and uniformly continuous function such that $\pi(V) < \infty$. Show that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} V(X_k) = \pi(V) , \qquad \mathbb{P} - \text{a.s.}$$

4. Under the assumptions of the previous question, show that there exists $\bar{\Omega}$ such that for all $\omega \in \bar{\Omega}$ and all measurable functions f such that $\sup_{x\in\mathsf{X}} |f(x)|/V(x) < \infty$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k(\omega)) = \pi(f) , \qquad \mathbb{P} - \text{a.s.}$$

5.4 Bibliographical Notes

Ergodic theory is a very important area of probability theory that has given rise to a great deal of work. The interested reader will find an introduction to this field in the books Walters (1982) and Billingsley (1978). The application of ergodic theory to Markov chains is a very classical subject. A detailed study of the ergodic theory of Markov chains can be found in (Revuz 1984, Chapter 4) and (Hernández-Lerma and Lasserre 2003, Chapter 2); These books contain many references to works on this subject that began in the early 1960s.

The proof of Birkhoff's theorem (Theorem 5.1.8) is borrowed from unpublished notes by B. Delyon and extended to possibly nonergodic dynamical systems. This approach is closely related to the very short proof of the law of large numbers for i.i.d. random variables written by Jacques Neveu in his unpublished lecture notes at École Polytechnique. Several other proofs of the ergodic theorem are also given in Billingsley (1978).

Theorem 5.2.6 is essentially borrowed from (Hernández-Lerma and Lasserre (2003), Proposition 2.4.3) even if the statements of the two results are slightly different.

Part II
Irreducible Chains: Basics

Chapter 6
Atomic Chains

In this chapter, we enter the core of the theory of Markov chains. We will encounter for the first time the fundamental notions of state classification, dichotomy between transience and recurrence, period, existence, uniqueness (up to scale), and characterization of invariant measures, as well as the classical limit theorems: the law of large numbers and the central limit theorem. These notions will be introduced, and the results will be obtained by means of the simplifying assumption that the state space contains an accessible atom. An atom is a set of states out of which the chain exits under a distribution common to all its individual states. A singleton is thus an atom, but if the state space is not discrete, it will in most cases be useless by failing to be accessible. Let us recall that a set is accessible if the chains eventually enter this set wherever it starts from with positive probability. If the state space is discrete, then accessible singletons usually exist, and the theory elaborated in the present chapter for chains with an accessible atom can be applied directly: this will be done in the next Chapter 7. However, most Markov chains on general state spaces do not possess an accessible atom, and therefore this chapter might seem of limited interest. Fortunately, we will see in Chapter 11 that it is possible to create an artificial atom by enlarging the state space of an irreducible Markov chain. The notion of irreducibility will also be first met in this chapter and then fully developed in Chapter 9. Therefore, this chapter is essential for the theory of irreducible Markov chains, and it has only Chapter 3 as a prerequisite.

6.1 Atoms

Definition 6.1.1 (Atom) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. A subset $\alpha \in \mathscr{X}$ is called an atom if there exists $\nu \in \mathbb{M}_1(\mathscr{X})$ such that $P(x, \cdot) = \nu$ for all $x \in \alpha$.*

R. Douc et al., *Markov Chains*, Springer Series in Operations Research
and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_6

A singleton is an atom. If X is not discrete, singletons are in general not very interesting atoms, because they are typically not accessible, and only when an accessible atom exists can meaningful results be obtained. Recall that a set A is said to be accessible if $\mathbb{P}_x(\sigma_A < \infty) > 0$ for all $x \in \mathsf{X}$. In this chapter, we will see examples of Markov chains on general state spaces that admit an accessible atom and even an accessible singleton, as shown in our first example.

Example 6.1.2 (Reflected random walk). Consider the random walk on $\mathbb{R}_+$ reflected at 0 defined by $X_k = (X_{k-1} + Z_k)^+$, where $\{Z_k, k \in \mathbb{N}\}$ is a real-valued i.i.d. sequence. The singleton $\{0\}$ is an atom. Let ν be the distribution of Z_1 and assume that there exists $a > 0$ such that $\nu((-\infty, -a)) > 0$. Then for all $n \in \mathbb{N}$ and $x \le na$,

$$P^n(x, \{0\}) \ge \mathbb{P}(Z_1 \le -a, \dots, Z_n \le -a) \ge \nu((-\infty, -a))^n > 0 .$$

Since n is arbitrary, the atom $\{0\}$ is accessible. ◀

We now introduce an important notation that is specific to atomic chains. If a function h defined on X is constant on α, then we write $h(\alpha)$ instead of $h(x)$ for all $x \in \alpha$. This convention will be used mainly in the following examples.

- For a measurable nonnegative function $f : \mathsf{X} \to \mathbb{R}_+$ and $k \ge 1$, we will write $P^k f(\alpha)$ instead of $P^k f(x)$ for $x \in \alpha$. If $f = \mathbb{1}_A$, we write $P^k(\alpha, A)$.
- If $A \in \mathscr{X}^{\mathbb{N}}$ is such that the function $x \to \mathbb{P}_x(A)$ is constant on α, then we will write $\mathbb{P}_\alpha(A)$ instead of $\mathbb{P}_x(A)$ for $x \in \alpha$.
- For every positive $\mathscr{X}^{\mathbb{N}}$-measurable random variable Y such that $\mathbb{E}_x[Y]$ is constant on α, we will write $\mathbb{E}_\alpha[Y]$ instead of $\mathbb{E}_x[Y]$ for $x \in \alpha$.
- The potential $U(x, \alpha)$ is constant on α, so we write $U(\alpha, \alpha)$.

Here is an example of this situation. Let $g \in \mathbb{F}_+(\mathsf{X})$, and let Y be $\sigma(X_s, s \ge 1)$-positive. Assume that g is constant on the set α. Then for all $x, x' \in \alpha$,

$$\mathbb{E}_x[g(X_0)Y] = \mathbb{E}_{x'}[g(X_0)Y] .$$

Thus the function $x \to \mathbb{E}_x[g(X_0)Y]$ is constant on α, and therefore it will be written $\mathbb{E}_\alpha[g(X_0)Y]$. An interesting consequence of this elementary remark is that equality holds in the maximum principle, Theorem 4.2.2.

Lemma 6.1.3 (Atomic maximum principle) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ that admits an atom α. Then for all $x \in \mathsf{X}$,*

$$U(x, \alpha) = \mathbb{P}_x(\tau_\alpha < \infty) U(\alpha, \alpha) .$$

Proof. Applying the strong Markov property, we get

$$
\begin{aligned}
U(x,\alpha) &= \mathbb{E}_x\left[\sum_{n=0}^{\infty} \mathbb{1}_\alpha(X_n)\right] = \mathbb{E}_x\left[\sum_{n=\tau_\alpha}^{\infty} \mathbb{1}_\alpha(X_n)\right] \\
&= \sum_{n=0}^{\infty} \mathbb{E}_x\left[\mathbb{1}_\alpha(X_n \circ \theta_{\tau_\alpha})\mathbb{1}\{\tau_\alpha < \infty\}\right] = \sum_{n=0}^{\infty} \mathbb{E}_x\left[\mathbb{1}\{\tau_\alpha < \infty\}\mathbb{E}_{X_{\tau_\alpha}}\left[\mathbb{1}_\alpha(X_n)\right]\right] \\
&= \sum_{n=0}^{\infty} \mathbb{P}_x(\tau_\alpha < \infty) P^n(\alpha,\alpha) = \mathbb{P}_x(\tau_\alpha < \infty) U(\alpha,\alpha).
\end{aligned}
$$

□

An important property of an accessible atom is that it can be used to characterize accessible sets.

Lemma 6.1.4 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ that admits an accessible atom α.*

(i) A set $A \in \mathscr{X}$ is accessible if and only if $\mathbb{P}_\alpha(\sigma_A < \infty) > 0$.
(ii) Let $A \in \mathscr{X}$. If A is not accessible, then A^c is accessible.

Proof. By definition, if A is accessible, then $\mathbb{P}_x(\sigma_A < \infty) > 0$ for every $x \in \alpha$. Since the function $x \mapsto \mathbb{P}_x(\sigma_A < \infty)$ is constant on α, this means that $\mathbb{P}_x(\sigma_A < \infty) > 0$. Conversely, if $\mathbb{P}_\alpha(\sigma_A < \infty) > 0$, then there exists $n \geq 1$ such that $P^n(\alpha, A) > 0$. Since α is accessible, it follows that for every $x \in \mathsf{X}$, there exists $k \geq 1$ such that $P^k(x,\alpha) > 0$. Then

$$
P^{n+k}(x,A) \geq \int_\alpha P^k(x,\mathrm{d}y) P^n(y,A) = P^k(x,\alpha) P^n(\alpha,A) > 0\,.
$$

Finally, if A is not accessible, then $P(\alpha,A) = 0$, and thus $P(\alpha,A^c) = 1$. □

6.2 Recurrence and Transience

Definition 6.2.1 (Atomic Recurrence and transience) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let α be an atom. The atom α is said to be recurrent if $U(\alpha,\alpha) = \infty$ and transient if $U(\alpha,\alpha) < \infty$.*

By definition, every atom is either recurrent or transient. Assume that a chain started from α returns to α with probability 1. It is then a simple application of the strong Markov property to show that the chain returns to α infinitely often with probability 1, i.e., the atom is recurrent. If, to the contrary, there is a positive probability that the chain started in α never returns to α, then it is not obvious that the atom is transient. This is indeed the case, and the dichotomy between recurrence and transience is also a dichotomy between almost surely returning to the atom or never returning to it with a positive probability. This fact is formally stated and proved in the next theorem.

Theorem 6.2.2. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let $\alpha \in \mathscr{X}$ be an atom.*

(i) The atom α is recurrent if one of the following equivalent properties is satisfied:

(a) $\mathbb{P}_\alpha(\sigma_\alpha < \infty) = 1$;
(b) $\mathbb{P}_\alpha(N_\alpha = \infty) = 1$;
(c) $U(\alpha,\alpha) = \mathbb{E}_\alpha[N_\alpha] = \infty$.

Moreover, for all $x \in \mathsf{X}$, $\mathbb{P}_x(\sigma_\alpha < \infty) = \mathbb{P}_x(N_\alpha = \infty)$.

(ii) The atom α is transient if one of the following equivalent properties is satisfied:

(a) $\mathbb{P}_\alpha(\sigma_\alpha < \infty) < 1$;
(b) $\mathbb{P}_\alpha(N_\alpha < \infty) = 1$;
(c) $U(\alpha,\alpha) = \mathbb{E}_\alpha[N_\alpha] < \infty$.

In that case, $U(\alpha,\alpha) = \{1 - \mathbb{P}_\alpha(\sigma_\alpha < \infty)\}^{-1}$, and under $\mathbb{P}_\alpha$, the number of visits N_α to α has a geometric distribution with mean $1/\mathbb{P}_\alpha(\sigma_\alpha < \infty)$.

Proof. By definition of the successive return times and the strong Markov property, we get, for $n \geq 1$,

$$\begin{aligned}
\mathbb{P}_\alpha(\sigma_\alpha^{(n)} < \infty) &= \mathbb{P}_\alpha(\sigma_\alpha^{(n-1)} < \infty,\ \sigma_\alpha \circ \theta_{\sigma_\alpha^{(n-1)}} < \infty) \\
&= \mathbb{E}_\alpha\left[\mathbb{1}_{\{\sigma_\alpha^{(n-1)} < \infty\}} \mathbb{P}_{X_{\sigma_\alpha^{(n-1)}}}(\sigma_\alpha < \infty)\right] \\
&= \mathbb{P}_\alpha(\sigma_\alpha^{(n-1)} < \infty)\mathbb{P}_\alpha(\sigma_\alpha < \infty)\,.
\end{aligned}$$

By induction, this yields, for $n \geq 1$,

$$\mathbb{P}_\alpha(\sigma_\alpha^{(n)} < \infty) = \{\mathbb{P}_\alpha(\sigma_\alpha < \infty)\}^n\,. \tag{6.2.1}$$

This yields, with the convention $\sigma_\alpha^{(0)} = 0$,

$$\begin{aligned}
\mathbb{P}_\alpha(N_\alpha = \infty) &= \lim_{n\to\infty} \mathbb{P}_\alpha(N_\alpha \geq n) \\
&= \lim_{n\to\infty} \mathbb{P}_\alpha(\sigma_\alpha^{(n)} < \infty) = \lim_{n\to\infty} \{\mathbb{P}_\alpha(\sigma_\alpha < \infty)\}^n\,,
\end{aligned} \tag{6.2.2}$$

and

$$U(\alpha,\alpha) = \mathbb{E}_\alpha[N_\alpha] = \sum_{n=0}^{\infty} \mathbb{P}_\alpha(\sigma_\alpha^{(n)} < \infty) = \sum_{n=0}^{\infty} \{\mathbb{P}_\alpha(\sigma_\alpha < \infty)\}^n\,. \tag{6.2.3}$$

For $x \in \mathsf{X}$, the strong Markov property implies that

$$\begin{aligned}\mathbb{P}_x(N_\alpha = +\infty) &= \mathbb{P}_x(N_\alpha \circ \theta_{\sigma_\alpha} = +\infty, \sigma_\alpha < +\infty) \\ &= \mathbb{P}_x(\sigma_\alpha < +\infty)\mathbb{P}_\alpha(N_\alpha = +\infty)\,. \end{aligned} \tag{6.2.4}$$

(i) The identity (6.2.2) yields that $\mathbb{P}_\alpha(\sigma_\alpha < \infty) = 1$ if and only if $\mathbb{P}_\alpha(N_\alpha = \infty) = 1$, and the identity (6.2.3) yields that $\mathbb{P}_\alpha(\sigma_\alpha < \infty) = 1$ if and only if $U(\alpha, \alpha) = \infty$. The last assertion follows from (6.2.4).

(ii) Similarly, (6.2.2) yields that $\mathbb{P}_\alpha(\sigma_\alpha < \infty) < 1$ if and only if $\mathbb{P}_\alpha(N_\alpha < \infty) = 1$, and (6.2.3) yields that $\mathbb{P}_\alpha(\sigma_\alpha < \infty) < 1$ if and only if $U(\alpha, \alpha) < \infty$. Moreover, (6.2.1) yields

$$\mathbb{P}_\alpha(N_\alpha > n) = \mathbb{P}_\alpha(\sigma_\alpha^{(n)} < \infty) = \mathbb{P}_\alpha(\sigma_\alpha < \infty)^n\,.$$

This proves that the distribution of N_α is geometric with mean $\mathbb{P}_\alpha(\sigma_\alpha < \infty)$ and $U(\alpha, \alpha) = 1/\mathbb{P}_\alpha(\sigma_\alpha = \infty)$.

□

A recurrent atom is one to which the chain returns infinitely often. A transient atom is one that will eventually be left forever.

Lemma 6.2.3 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let α be an accessible recurrent atom. Then the set α_∞ defined by $\alpha_\infty = \{x \in \mathsf{X} : \mathbb{P}_x(N_\alpha = \infty) = 1\}$ is absorbing.*

Proof. By Proposition 4.2.4, the function $h(x) = \mathbb{P}_x(N_\alpha = \infty)$ is harmonic. For $x \in \alpha_\infty$, we get

$$\begin{aligned}1 = h(x) = Ph(x) &= \mathbb{E}_x[\mathbb{1}_{\alpha_\infty}(X_1)\mathbb{P}_{X_1}(N_\alpha = \infty)] + \mathbb{E}_x[\mathbb{1}_{\alpha_\infty^c}(X_1)\mathbb{P}_{X_1}(N_\alpha = \infty)] \\ &= P(x, \alpha_\infty) + \mathbb{E}_x[\mathbb{1}_{\alpha_\infty^c}(X_1)\mathbb{P}_{X_1}(N_\alpha = \infty)].\end{aligned}$$

The previous relation may be rewritten as

$$\mathbb{E}_x[\mathbb{1}_{\alpha_\infty^c}(X_1)\{1 - \mathbb{P}_{X_1}(N_\alpha = \infty)\}] = 0\,.$$

For $x \in \alpha_\infty^c$, $\mathbb{P}_x(N_\alpha = \infty) < 1$; thus the previous relation implies $P(x, \alpha_\infty^c) = 0$. □

Proposition 6.2.4 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let α be an atom.*

(i) If α is accessible recurrent, then every atom β satisfying $\mathbb{P}_\alpha(\sigma_\beta < \infty) > 0$ is accessible recurrent and

$$\mathbb{P}_\alpha(N_\beta = \infty) = \mathbb{P}_\beta(N_\alpha = \infty) = 1\,. \tag{6.2.5}$$

(ii) If α is recurrent and if there exists an accessible atom β, then α is accessible.

Proof. (i) The atom β is accessible by Lemma 6.1.4. Applying Theorem 4.2.6 with $A = \alpha$ and $B = \beta$, we obtain, for all $x \in \alpha$, $1 = \mathbb{P}_x(N_\alpha = \infty) \le \mathbb{P}_x(N_\beta = \infty) = \mathbb{P}_\alpha(N_\beta = \infty)$. On the other hand, applying again the strong Markov property, we obtain, for all $x \in \alpha$,

$$\begin{aligned}1 = \mathbb{P}_x(N_\beta = \infty) &= \mathbb{P}_x(\tau_\beta < \infty, N_\beta \circ \theta_{\tau_\beta} = \infty) \\ &= \mathbb{P}_x(\tau_\beta < \infty)\mathbb{P}_\beta(N_\beta = \infty) \le \mathbb{P}_\beta(N_\beta = \infty)\,.\end{aligned}$$

This proves that β is recurrent. Interchanging the roles of α and β proves (6.2.5).

(ii) Let α be a recurrent atom and β an accessible atom. Then

$$\begin{aligned}\mathbb{P}_\alpha(\sigma_\beta < \infty) = \mathbb{P}_\alpha(\sigma_\beta < \infty, N_\alpha = +\infty) &= \mathbb{P}_\alpha(\sigma_\beta < \infty, N_\alpha \circ \sigma_\beta = +\infty) \\ &= \mathbb{P}_\alpha(\sigma_\beta < \infty)\mathbb{P}_\beta(N_\alpha = +\infty)\,.\end{aligned}$$

Since $\mathbb{P}_\alpha(\sigma_\beta < \infty) > 0$, this implies that $\mathbb{P}_\beta(N_\alpha = +\infty) = 1$, and α is accessible.

□

As an immediate consequence of Proposition 6.2.4, we obtain that accessible atoms are either all recurrent or all transient. We can extend the definition of recurrence to all sets.

Definition 6.2.5 (Recurrent set, Recurrent kernel) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$.*

- *A set $A \in \mathscr{X}$ is said to be recurrent if $U(x,A) = \infty$ for all $x \in A$.*
- *The kernel P is said to be recurrent if every accessible set is recurrent.*

Definition 6.2.6 (Uniformly Transient set, Transient set) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$.*

- *A set $A \in \mathscr{X}$ is called uniformly transient if $\sup_{x\in A} U(x,A) < \infty$.*
- *A set $A \in \mathscr{X}$ is called transient if $A = \bigcup_{n=1}^\infty A_n$, where A_n is uniformly transient.*
- *A Markov kernel P is said to be transient if X is transient.*

Theorem 6.2.7. *Let P be a Markov kernel on $(\mathsf{X}, \mathscr{X})$ that admits an accessible atom α.*

(i) P is recurrent if and only if α is recurrent.
(ii) P is transient if and only if α is transient.

Proof. (i) If P is recurrent, then an accessible atom α is recurrent by definition. Conversely, let α be a recurrent accessible atom. Then α_∞ is absorbing by Lemma 6.2.3 and $\alpha \subset \alpha_\infty$. Let A be an accessible set, and set $B = A \cap \alpha_\infty$. Since α_∞ is absorbing and A is accessible, $\mathbb{P}_\alpha(\sigma_B < \infty) > 0$ and $\mathbb{P}_\alpha(N_B = \infty) = 1$ by Theorem 4.2.6. Therefore, for all $x \in A$,

$$U(x,A) \geq \mathbb{E}_x[N_B] \geq \mathbb{E}_x[\mathbb{1}\{\sigma_\alpha < \infty\} N_B \circ \theta_{\sigma_\alpha}] = \mathbb{P}_x(\sigma_\alpha < \infty)\mathbb{E}_\alpha[N_B] = \infty\,.$$

This proves that A is recurrent.

(ii) Assume first the atom α is transient. We will show that there exists a countable covering of X by uniformly transient sets, i.e., a family $\{\mathsf{X}_n, n \in \mathbb{N}^*\} \subset \mathscr{X}$ such that $\sup_{x\in\mathsf{X}} U(x,\mathsf{X}_m) < \infty$ for all $m \geq 1$ and $\mathsf{X} = \bigcup_{m=1}^\infty \mathsf{X}_m$.
For $m \in \mathbb{N}^*$, define $\mathsf{X}_m = \left\{x \in \mathsf{X} : \sum_{i=0}^m P^i(x,\alpha) \geq m^{-1}\right\}$. Since $U(x,\alpha) \geq UP^i(x,\alpha)$ for all $i \geq 0$ and $x \in \mathsf{X}$ and since α is transient, the atomic version of the maximum principle, Lemma 6.1.3, yields for all $x \in \mathsf{X}$,

$$\begin{aligned}\infty > &(m+1)U(\alpha,\alpha) \geq (m+1)U(x,\alpha) \geq \sum_{i=0}^m UP^i(x,\alpha)\\ &\geq \sum_{i=0}^m \int_{\mathsf{X}_m} U(x,\mathrm{d}y)P^i(y,\alpha) = \int_{\mathsf{X}_m} U(x,\mathrm{d}y)\sum_{i=0}^m P^i(y,\alpha) \geq m^{-1}U(x,\mathsf{X}_m)\,.\end{aligned}$$

Therefore, X_m is uniformly transient. Moreover, $\mathsf{X}_m \subseteq \mathsf{X}_{m+1}$, and since α is accessible,

$$\bigcup_{m=1}^\infty \mathsf{X}_m = \{x \in \mathsf{X} : U(x,\alpha) > 0\} = \mathsf{X}\,.$$

Conversely, if P is transient, then $\mathsf{X} = \cup_{m\geq 1}\mathsf{X}_m$ with X_m uniformly transient for all $m \geq 1$. Since $P(\alpha,\mathsf{X}) = 1$, there exists r such that $P(\alpha,\mathsf{X}_r) > 0$. By Lemma 6.1.4, X_r is accessible and transient, and therefore α cannot be recurrent in view of Proposition 6.2.4, so it is transient.

□

In some cases, an invariant probability measure for P can be exhibited. The following proposition provides a simple criterion for recurrence.

Proposition 6.2.8 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P admits an atom α and an invariant probability measure π.*

(i) If $\pi(\alpha) > 0$, then α is recurrent.
(ii) If α is accessible, then $\pi(\alpha) > 0$ and α (and hence P) are recurrent.

Proof. (i) Since π is invariant, we have

$$\pi U(\alpha) = \sum_{n=0}^{\infty} \pi P^n(\alpha) = \sum_{n=0}^{\infty} \pi(\alpha) . \tag{6.2.6}$$

Therefore, if $\pi(\alpha) > 0$, the atomic version of the maximum principle (Lemma 6.1.3) yields

$$\infty = \pi U(\alpha) = \int_{\mathsf{X}} \pi(\mathrm{d}y) U(y,\alpha) \leq U(\alpha,\alpha) \int_{\mathsf{X}} \pi(\mathrm{d}y) = U(\alpha,\alpha) .$$

(ii) Since α is an accessible atom, $K_{a_\varepsilon}(x,\alpha) > 0$ for all $x \in \mathsf{X}$ and $\varepsilon \in (0,1)$. Therefore, we get that

$$\pi(\alpha) = \pi K_{a_\varepsilon}(\alpha) = \int_{\mathsf{X}} \pi(\mathrm{d}x) K_{a_\varepsilon}(x,\alpha) > 0 .$$

Therefore α is recurrent by (i), and P is recurrent by Theorem 6.2.7.

□

6.3 Period of an Atom

Consider the Markov kernel P on $\{0,1\}$ defined as follows:

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

The Markov chain associated with this kernel behaves as follows: if it is in state 0, then it jumps to 1 and vice versa. Therefore, starting, for instance, from 0, the chain will be back in state 0 at all even integers. The unique invariant probability for this kernel is the uniform distribution $\pi = (1/2, 1/2)$. However, the periodic behavior precludes the convergence of the chain to its stationary distribution, an important and desirable feature. We will further discuss this in Chapter 7 and later chapters. We focus here only on the definition of periodicity.

Definition 6.3.1 (Period) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let α be an atom. Define by E_α the subset*

$$E_\alpha = \{n > 0 : P^n(\alpha,\alpha) > 0\} . \tag{6.3.1}$$

The period $d(\alpha)$ of the atom α is the greatest common divisor (g.c.d.) of E_α, with the convention g.c.d.$(\emptyset) = \infty$. An atom is said to be aperiodic if its period is 1.

It is easily seen that the set E_α is stable by addition, i.e., if $n_1,\ldots,n_s \in E_\alpha$ and $b_1,\ldots,b_s$ are nonnegative integers such that $\sum_{i=1}^{s} b_i > 0$, then $\sum_{i=1}^{s} b_i n_i > 0$. To go further, we need an elementary result that is a straightforward consequence of Bézout's theorem.

Lemma 6.3.2 *Let E be a subset of $\mathbb{N}^*$, which is stable by addition, and let $d = g.c.d.(E)$. There exists $n_0 \in \mathbb{N}^*$ such that $dn \in E$ for all $n \geq n_0$.*

Proof. There exist $n_1, \ldots, n_s \in E$ such that $d = g.c.d.(n_1, \ldots, n_s)$, and by Bézout's theorem, there exist $a_1, \ldots, a_s \in \mathbb{Z}$ such that $\sum_{i=1}^s a_i n_i = d$. Setting $p = \sum_{i=1}^s a_i^- n_i$, we get

$$\sum_{i=1}^s a_i^+ n_i = \sum_{i=1}^s (a_i + a_i^-) n_i = p + d \,.$$

Since E is stable by addition, p and $p+d$ belong to E. Since $p \in E$, there exists $k \in \mathbb{N}$ such that $p = kd$. For $n \geq k^2$, we may write $n = mk + r$ with $r \in \{0, \ldots, k-1\}$ and $m \geq k$. Then, using again that E is stable by addition, we get

$$dn = d(mk+r) = mkd + rd = (m-r)p + r(p+d) \in E \,.$$

□

Proposition 6.3.3 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an accessible atom α. There exists an integer $n_0 \in \mathbb{N}^*$ such that $nd(\alpha) \in E_\alpha$ for all $n \geq n_0$.*

Proof. If $n, m \in E_\alpha$, then $P^{n+m}(\alpha, \alpha) \geq P^n(\alpha, \alpha) P^m(\alpha, \alpha) > 0$. The set E_α is therefore stable by addition. Applying Lemma 6.3.2 concludes the proof. □

The analysis of atomic chains would be difficult if different atoms might have different periods. Fortunately, this cannot happen.

Proposition 6.3.4 *Let α and β be two accessible atoms. Then $d(\alpha) = d(\beta)$.*

Proof. Assume that α and β are two accessible atoms. Then there exist positive integers ℓ and m such that $P^\ell(\alpha, \beta) > 0$ and $P^m(\beta, \alpha) > 0$. Then $P^{\ell+m}(\alpha, \alpha) \geq P^\ell(\alpha, \beta) P^m(\beta, \alpha) > 0$. This implies that $d(\alpha)$ divides $\ell + m$. Moreover, for every $n \in E_\beta$,

$$P^{\ell+n+m}(\alpha, \alpha) \geq P^\ell(\alpha, \beta) P^n(\beta, \beta) P^m(\beta, \alpha) > 0 \,.$$

This implies that $d(\alpha)$ also divides $\ell + n + m$, and therefore $d(\alpha)$ divides n. Since $n \in E_\beta$ is arbitrary, this implies that $d(\alpha)$ divides $d(\beta)$. Similarly, $d(\beta)$ divides $d(\alpha)$, and thus $d(\alpha) = d(\beta)$. □

Definition 6.3.5 (Period of an atomic Markov chain) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ that admits an accessible atom α. The common period of all accessible atoms is called the period of P. If one and hence all accessible atoms are aperiodic, then P is called an aperiodic Markov kernel.*

We conclude with a characterization of aperiodicity.

Proposition 6.3.6 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ that admits an accessible atom α. The kernel P is aperiodic if and only if for all accessible atoms α there exists $N \in \mathbb{N}$ such that $P^n(\alpha, \alpha) > 0$ for all $n \geq N$.*

Moreover, if P is aperiodic, then for all accessible atoms α, β, there exists $N \in \mathbb{N}$ such that $P^n(\alpha, \beta) > 0$ for all $n \geq N$.

Proof. Assume that P is aperiodic. Let α be an accessible atom. Since E_α is stable by addition and $d(\alpha) = 1$, Lemma 6.3.2 implies that there exists an integer n_0 such that $P^n(\alpha, \alpha) > 0$ for all $n \geq n_0$. The converse is obvious.

Let α, β be accessible atoms. Then there exist m and p such that $P^m(\alpha, \alpha) > 0$, $P^p(\alpha, \beta) > 0$, and for all $n \geq n_0$,

$$P^{m+n+p}(\alpha, \beta) \geq P^m(\alpha, \alpha) P^n(\alpha, \alpha) P^p(z, \beta) > 0 .$$

The statement follows with $N = m + n_0 + p$. □

6.4 Subinvariant and Invariant Measures

In this section, we will use the results of Section 3.6 and in particular Theorem 3.6.5 to prove the existence of invariant measures with respect to a Markov kernel that admits an accessible and recurrent atom α.

Definition 6.4.1 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. An atom $\alpha \in \mathscr{X}$ is said to be*

(i) positive if $\mathbb{E}_\alpha[\sigma_\alpha] < \infty$;
(ii) null recurrent if it is recurrent and $\mathbb{E}_\alpha[\sigma_\alpha] = \infty$.

By definition, a positive atom is recurrent. We define the measure λ_α on $\mathscr{X}$ by

$$\lambda_\alpha(A) = \mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} \mathbb{1}_A(X_k)\right], \qquad A \in \mathscr{X}. \tag{6.4.1}$$

Theorem 6.4.2. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let α be an accessible atom.*

(i) If α is recurrent, then λ_α is invariant.
(ii) If λ_α is invariant, then α is recurrent.
(iii) If α is recurrent, then every subinvariant measure λ is invariant, proportional to λ_α, satisfies $\lambda(\alpha) < \infty$, and for all $B \in \mathscr{X}$,

$$\lambda(B) = \lambda(\alpha)\lambda_\alpha(B) = \lambda(\alpha)\int_\alpha \lambda_\alpha(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_\alpha - 1} \mathbb{1}_B(X_k)\right].$$

(iv) Assume that α is recurrent. Then α is positive if and only if P admits a unique invariant probability measure π. If α is positive, then the unique invariant probability measure can be expressed as $\pi = (\mathbb{E}_\alpha[\sigma_\alpha])^{-1}\lambda_\alpha$.

Proof. (i) Since $\lambda_\alpha(\alpha) = \mathbb{P}_\alpha(\sigma_\alpha < \infty) = 1$ and α is accessible, Lemma 3.6.1 ensures that λ_α is σ-finite. Let the trace σ-field of $\mathscr{X}$ on α be denoted by $\mathscr{X}_\alpha$ (see (3.3.5)). Let ν be the measure defined on α by $\nu_\alpha(B) = \mathbb{P}_\alpha(X_{\sigma_\alpha} \in B, \sigma_\alpha < \infty)$, $B \in \mathscr{X}_\alpha$. Finally, let Q_α be the induced kernel on the atom α (see Definition 3.3.7):

$$Q_\alpha(x,B) = \mathbb{P}_x(X_{\sigma_\alpha} \in B, \sigma_\alpha < \infty) = \nu_\alpha(B), \quad x \in \alpha, B \in \mathscr{X}_\alpha.$$

If $\mathbb{P}_\alpha(\sigma_\alpha < \infty) = 1$, then $\nu_\alpha(\alpha) = 1$, and obviously ν_α is Q_α invariant. Moreover, by Theorem 3.6.3 applied with $C = \alpha$, the measure ν_α^0 defined for $B \in \mathscr{X}$ by

$$\nu_\alpha^0(B) = \int_\alpha \nu_\alpha(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_\alpha - 1} \mathbb{1}_B(X_k)\right]$$

is invariant. By Lemma 3.6.2, we have for all $B \in \mathscr{X}$,

$$\nu_\alpha^0(B) = \nu_\alpha^0 P(B) = \int_\alpha \nu_\alpha(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_\alpha} \mathbb{1}_B(X_k)\right] = \lambda_\alpha(B).$$

This proves that λ_α is invariant.

(ii) Assume that λ_α is invariant. We apply again Theorem 3.6.5. Note first that $\lambda_\alpha(\alpha) \le 1$ and $\mathbb{P}_\alpha(\sigma_\alpha < \infty) > 0$, since α is accessible. Since λ_α is invariant, Theorem 3.6.5 implies that $\mathbb{P}_\alpha(\sigma_\alpha < \infty) = 1$, which shows that α is recurrent.

(iii) Assume now that α is recurrent and let λ be a subinvariant measure. Since $\mathbb{P}_\alpha(\sigma_\alpha < \infty) = 1$, Theorem 3.6.5 shows that

$$\mu(B) = \int_\alpha \mu(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_\alpha} \mathbb{1}_B(X_k)\right]$$

is invariant. Since the function $x \mapsto \mathbb{E}_x\left[\sum_{k=1}^{\sigma_\alpha} \mathbb{1}_B(X_k)\right]$ is invariant on α, we obtain that $\mu(B) = \mu(\alpha)\lambda_\alpha(B)$ for all $B \in \mathscr{X}$. This proves that all the subinvariant measures are invariant and proportional to λ_α.

(iv) If α is positive, then $\lambda_\alpha(\mathsf{X}) = \mathbb{E}_\alpha[\sigma_\alpha] < \infty$, and $\lambda_\alpha/\lambda_\alpha(\mathsf{X})$ is thus the unique invariant probability measure. Conversely, if α is recurrent and P admits an invariant probability measure π, then by (i), π is proportional to λ_α. Since $\pi(\mathsf{X}) = 1 < \infty$, which implies

$$\mathbb{E}_\alpha[\sigma_\alpha] = \lambda_\alpha(\mathsf{X}) < \infty\,,$$

showing that α is positive.

□

Let α be an accessible positive atom and let π be the unique invariant probability measure. We now turn our attention to modulated moments of the return times to the atom. For a sequence $\{r(k), k \in \mathbb{N}\}$, we define the integrated sequence $\{r^*(k), k \in \mathbb{N}\}$ by $r^*(0) = 0$ and

$$r^*(k) = \sum_{j=1}^{k} r(j)\,, \qquad k \geq 1\,. \tag{6.4.2}$$

Lemma 6.4.3 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an accessible and positive atom α and let π be its unique invariant probability measure. Let $\{r(n), n \in \mathbb{N}\}$ be a nonnegative sequence and let $f \in \mathbb{F}_+(\mathsf{X})$. Then*

$$\mathbb{E}_\pi\left[\sum_{k=1}^{\sigma_\alpha} r(k) f(X_k)\right] = \pi(\alpha)\mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} r^*(k) f(X_k)\right]\,.$$

Proof. Define the function h on X by $h(x) = \mathbb{E}_x\left[\sum_{j=1}^{\sigma_\alpha} r(j) f(X_j)\right]$. Applying Theorem 6.4.2 and noting that h is constant on α, we obtain

$$\begin{aligned}
\mathbb{E}_\pi\left[\sum_{k=1}^{\sigma_\alpha} r(k) f(X_k)\right] &= \pi(\alpha)\mathbb{E}_\alpha\left[\sum_{k=0}^{\sigma_\alpha-1} r(k) h(X_k)\right] \\
&= \pi(\alpha)\mathbb{E}_\alpha\left[\sum_{k=0}^{\sigma_\alpha-1} \mathbb{E}_{X_k}\left[\sum_{j=1}^{\sigma_\alpha} r(j) f(X_j)\right]\right] \\
&= \pi(\alpha)\sum_{k=0}^{\infty}\sum_{j=1}^{\infty} \mathbb{E}_\alpha\left[\mathbb{1}_{\{k<\sigma_\alpha\}}\mathbb{1}_{\{j\leq\sigma_\alpha\circ\theta_k\}} r(j) f(X_{k+j})\right]\,.
\end{aligned}$$

Since $\{k < \sigma_\alpha\} \cap \{j \leq \sigma_\alpha \circ \theta_k\} = \{k + j \leq \sigma_\alpha\}$, it follows that

$$\mathbb{E}_\pi\left[\sum_{k=1}^{\sigma_\alpha} r(k)f(X_k)\right] = \sum_{k=0}^{\infty}\sum_{j=1}^{\infty}\mathbb{E}_\alpha\left[\mathbb{1}_{\{k+j\le\sigma_\alpha\}}r(j)f(X_{k+j})\right]$$
$$= \sum_{n=1}^{\infty}\sum_{j=1}^{n}\mathbb{E}_\alpha\left[\mathbb{1}_{\{n\le\sigma_\alpha\}}r(j)f(X_n)\right] = \mathbb{E}_\alpha\left[\sum_{n=1}^{\sigma_\alpha} r^*(n)f(X_n)\right].$$

□

In the case of a geometric sequence, this identity yields a necessary and sufficient condition for the finiteness of the geometric moment.

Corollary 6.4.4 *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$ admitting an accessible and positive atom α. Let $f \in \mathbb{M}_+(\mathscr{X})$ and $r > 0$. Then*

$$\mathbb{E}_\pi\left[\sum_{k=1}^{\sigma_\alpha} r^k f(X_k)\right] < \infty \Leftrightarrow \mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} r^k f(X_k)\right] < \infty.$$

Proof. If $r(k) = r^k$, then $r^*(k) = r(r^k - 1)/(r-1)$. □

This last result is in sharp contrast to the polynomial moments of the return times. Indeed, for all $r > 0$, $\mathbb{E}_\pi[\sigma_\alpha^r] < \infty$ if and only if $\mathbb{E}_\alpha[\sigma_\alpha^{r+1}] < \infty$. For geometric moments, there is no such discrepancy. To illustrate this, we relate the first two moments of the return time to the state x under the stationary distribution $\mathbb{P}_\pi$ to higher moments under $\mathbb{P}_\alpha$. For instance, we have, for every $\alpha \in \mathsf{X}$,

$$\mathbb{E}_\pi[\sigma_\alpha] = \pi(\alpha)\mathbb{E}_\alpha\left[\frac{\sigma_\alpha(\sigma_\alpha+1)}{2}\right],$$
$$\mathbb{E}_\pi[\sigma_\alpha^2] = \pi(\alpha)\mathbb{E}_\alpha\left[\frac{\sigma_\alpha(\sigma_\alpha+1)(2\sigma_\alpha+1)}{6}\right].$$

Recall that the attraction set α_+ of α was defined in Section 3.6 as $\alpha_+ = \{x \in \mathsf{X} : \mathbb{P}_x(\sigma_\alpha < \infty) = 1\}$.

Lemma 6.4.5 *Let P be a Markov kernel admitting an accessible and recurrent atom α. Then α_+ is absorbing. If, moreover, α is positive and π denotes the unique invariant probability measure, then $\pi(\alpha_+) = 1$.*

Proof. Since α is recurrent, it follows that $\alpha \subset \alpha_+$, and the set α_+ is absorbing by Lemma 3.5.4. This implies that $\mathbb{P}_\alpha\left(\bigcap_{k=1}^\infty\{X_k \in \alpha_+\}\right) = 1$ and

$$\lambda_\alpha(\alpha_+) = \mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha}\mathbb{1}_{\alpha_+}(X_k)\right] = \mathbb{E}_\alpha[\sigma_\alpha].$$

Thus if α is positive, Theorem 6.4.2 yields $\pi(\alpha_+) = \lambda_\alpha(\alpha_+)/\mathbb{E}_\alpha[\sigma_\alpha] = 1$. □

We end this section with a developed example of a particular Markov chain that is not itself atomic but that can be closely associated to an atomic Markov chain.

Example 6.4.6 (Stochastic Unit Root). Let $\{Z_k, k \in \mathbb{N}\}$ and $\{U_k, k \in \mathbb{N}\}$ be two independent sequences of i.i.d. random variables defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, the distribution of U_1 being uniform on $[0,1]$. Let $r : \mathbb{R} \to [0,1]$ be a càdlàg nondecreasing function, and let the sequence $\{X_k, k \in \mathbb{N}\}$ be defined recursively by X_0 and for $k \geq 1$,

$$X_k = \begin{cases} X_{k-1} + Z_k & \text{if } U_k \leq r(X_{k-1}), \\ Z_k & \text{otherwise}. \end{cases} \tag{6.4.3}$$

We assume that ν_Z, the distribution of Z_k, has a continuous positive density f_Z with respect to the Lebesgue measure. Clearly, $\{X_n, n \in \mathbb{N}\}$ defines a Markov Chain on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Its Markov kernel P can be expressed as follows: for all $(x, A) \in \mathbb{R} \times \mathscr{B}(\mathbb{R})$,

$$P(x, A) = r(x)\mathbb{E}\left[\mathbb{1}_A(x + Z_0)\right] + (1 - r(x))\mathbb{E}\left[\mathbb{1}_A(Z_0)\right]. \tag{6.4.4}$$

Without any further assumption, $\{X_n, n \in \mathbb{N}\}$ is not an atomic Markov chain, but it actually can be embedded into an atomic Markov chain. Define $\mathscr{F}_0 = \sigma(X_0)$, and for $k \geq 1$, $\mathscr{F}_k = \sigma(X_0, U_\ell, Z_\ell, 1 \leq \ell \leq k)$. Then $\{X_k, k \in \mathbb{N}\}$ is adapted to the filtration $\mathscr{F} = \{\mathscr{F}_k, k \in \mathbb{N}\}$. Define for $k \geq 1$,

$$V_k = \mathbb{1}\left\{U_k \leq r(X_{k-1})\right\}. \tag{6.4.5}$$

Then (6.4.3) reads

$$X_k = X_{k-1}V_k + Z_k. \tag{6.4.6}$$

Setting $W_k = (X_k, V_{k+1})$, the sequence $\{W_k, k \in \mathbb{N}\}$ is then a Markov chain with

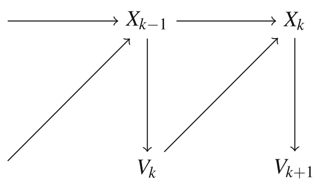

Fig. 6.1 Dependency graph of $\{(X_k, V_{k+1}), k \in \mathbb{N}\}$.

kernel $\bar{P}$ defined for $w = (x, v) \in \mathbb{R} \times \{0,1\}$ and $A \in \mathscr{B}(\mathbb{R})$ by

$$\begin{aligned} \bar{P}(w, A \times \{1\}) &= \mathbb{E}\left[\mathbb{1}_A(xv + Z_0) r(xv + Z_0)\right] \\ \bar{P}(w, A \times \{0\}) &= \mathbb{E}\left[\mathbb{1}_A(xv + Z_0)(1 - r(xv + Z_0))\right]. \end{aligned}$$

The Markov kernel $\bar{P}$ admits the set $\alpha = \mathbb{R} \times \{0\}$ as an atom. Let $\bar{\mathbb{P}}_\mu$ be the probability induced on $((\mathsf{X} \times \{0,1\})^{\mathbb{N}}, (\mathscr{X} \otimes \mathscr{P}(\{0,1\}))^{\otimes \mathbb{N}})$ by the Markov kernel $\bar{P}$ and the initial distribution μ. We denote by $\mathbb{E}_\mu$ the associated expectation operator. By definition of $\bar{P}$, for all $w = (x,v) \in \mathbb{R} \times \{0,1\}$,

$$\bar{P}(w, \alpha) = 1 - \mathbb{E}\left[r(xv + Z_0)\right] .$$

Therefore, if $\mathrm{Leb}(\{r < 1\}) > 0$ and Z_0 has a positive density with respect to Lebesgue measure, then $1 - \mathbb{E}\left[r(xv + Z_0)\right] > 0$. This in turn implies that $\bar{P}(w, \alpha) > 0$ and α is accessible and aperiodic.

We now investigate the existence of an invariant probability measure for P. For $k \in \mathbb{N}^*$, define $S_k = Z_1 + \cdots + Z_k$ and

$$p_k = \mathbb{E}\left[\prod_{i=1}^{k} r(S_k)\right] .$$

Then $\bar{\mathbb{P}}_\alpha(\sigma_\alpha > k) = p_k$ and

$$\bar{\mathbb{E}}_\alpha[f(X_1, \ldots, X_k)\mathbb{1}\{\sigma_\alpha \geq k\}] = \mathbb{E}\left[r(S_1) \cdots r(S_{k-1}) f(S_1, \ldots, S_k)\right] .$$

Lemma 6.4.7 *If $\sum_{k=1}^{\infty} p_k < \infty$, then P admits an invariant probability measure π defined for $A \in \mathscr{B}(\mathsf{X})$ by*

$$\pi(A) = \frac{\sum_{k=1}^{\infty} \mathbb{E}\left[r(Z_1) \cdots r(Z_1 + \cdots + Z_{k-1}) \mathbb{1}_A(Z_1 + \cdots + Z_k)\right]}{\sum_{k=1}^{\infty} p_k} .$$

Proof. The condition $\sum_k p_k < \infty$ implies that $\mathrm{Leb}(\{r < 1\} > 0)$; hence α is accessible for $\bar{P}$. Moreover, since $\bar{\mathbb{P}}_{(x,0)}(\sigma_\alpha > k) = p_k$, the same condition also implies that α is positive. Thus Theorem 6.4.2 implies that $\bar{P}$ admits a unique invariant probability measure $\bar{\pi}$ defined for $A \in \mathscr{B}(\mathsf{X})$ and $i \in \{0,1\}$ by

$$\bar{\pi}(A \times \{i\}) = \frac{\bar{\mathbb{E}}_\alpha[\sum_{k=1}^{\sigma_\alpha} \mathbb{1}_{A \times \{i\}}(W_k)]}{\bar{\mathbb{E}}_\alpha[\sigma_\alpha]} .$$

The measure π defined by $\pi(A) = \bar{\pi}(A \times \{0,1\})$ is invariant for P. Indeed, π is by definition the distribution of X_k for all k under $\bar{\mathbb{P}}_\pi$. This means that π is a stationary distribution of the chain $\{X_k\}$, and thus it is invariant. Moreover,

$$\begin{aligned}
\pi(A) &= \frac{\sum_{k=1}^{\infty} \bar{\mathbb{P}}_\alpha[k \leq \sigma_\alpha, X_k \in A]}{\sum_{k=1}^{\infty} \bar{\mathbb{P}}_\alpha[k \leq \sigma_\alpha]} \\
&= \frac{\sum_{k=1}^{\infty} \mathbb{E}\left[r(Z_1) \cdots r(Z_1 + \cdots + Z_{k-1}) \mathbb{1}_A(Z_1 + \cdots + Z_k)\right]}{\sum_{k=1}^{\infty} p_k} .
\end{aligned}$$

□

6.5 Independence of the Excursions

Let $\sigma_\alpha^{(i)}$ be the successive returns to the atom α and recall the convention $\sigma_\alpha^{(0)} = 0$ and $\sigma_\alpha^{(1)} = \sigma_\alpha$.

Proposition 6.5.1 *Let P be a Markov kernel that admits a recurrent atom α. Let $Z_0, \dots, Z_k$ be $\mathscr{F}_{\sigma_\alpha}$ measurable random variables such that for $i = 1, \dots, k$. The function $x \mapsto \mathbb{E}_x[Z_i]$ is constant on α. Then for every initial distribution $\lambda \in \mathbb{M}_1(\mathscr{X})$ such that $\mathbb{P}_\lambda(\sigma_\alpha < \infty) = 1$,*

$$\mathbb{E}_\lambda\left[\prod_{i=0}^{k} Z_i \circ \theta_{\sigma_\alpha^{(i)}}\right] = \mathbb{E}_\lambda[Z_0] \prod_{i=1}^{k} \mathbb{E}_\alpha[Z_i]\,. \tag{6.5.1}$$

Proof. For $k = 1$, the assumption that the function $x \to \mathbb{E}_x[Z_1]$ is constant on α and the strong Markov property yield

$$\mathbb{E}_\lambda[Z_0 Z_1 \circ \theta_{\sigma_\alpha}] = \mathbb{E}_\lambda[Z_0 \mathbb{E}_{X_{\sigma_\alpha}}[Z_1]] = \mathbb{E}_\lambda[Z_0 \mathbb{E}_\alpha[Z_1]] = \mathbb{E}_\lambda[Z_0]\mathbb{E}_\alpha[Z_1]\,.$$

Assume now that (6.5.1) holds for some $k \geq 1$. Then the induction assumption, the identity $\theta_{\sigma_\alpha^{(k)}} = \theta_{\sigma_\alpha^{(k-1)}} \circ \theta_{\sigma_\alpha}$ on $\{\theta_{\sigma_\alpha^{(k)}} < \infty\}$, and the strong Markov property yield

$$\begin{aligned}
\mathbb{E}_\lambda\left[\prod_{i=0}^{k} Z_i \circ \theta_{\sigma_\alpha^{(i)}}\right] &= \mathbb{E}_\lambda\left[Z_0 \left(\prod_{i=1}^{k} Z_i \circ \theta_{\sigma_\alpha^{(i-1)}}\right) \circ \theta_{\sigma_\alpha}\right] \\
&= \mathbb{E}_\lambda\left[Z_0 \mathbb{E}_{X_{\sigma_\alpha}}\left[\prod_{i=1}^{k} Z_i \circ \theta_{\sigma_\alpha^{(i-1)}}\right]\right] = \mathbb{E}_\lambda[Z_0] \prod_{i=1}^{k} \mathbb{E}_\alpha[Z_i]\,.
\end{aligned}$$

□

As an application, for $f \in \mathbb{F}(\mathsf{X})$, define $\mathscr{E}_1(\alpha, f) = \sum_{k=1}^{\sigma_\alpha} f(X_k)$, and for $n \in \mathbb{N}$,

$$\mathscr{E}_{n+1}(\alpha, f) = \mathscr{E}_1(\alpha, f) \circ \theta_{\sigma_\alpha^{(n)}} = \sum_{k=\sigma_\alpha^{(n)}+1}^{\sigma_\alpha^{(n+1)}} f(X_k)\,. \tag{6.5.2}$$

Corollary 6.5.2 *Let P be a Markov kernel admitting a recurrent atom α. Then under $\mathbb{P}_\alpha$, the sequence $\{\mathscr{E}_n(\alpha, f), n \in \mathbb{N}^*\}$ is i.i.d. For every $\mu \in \mathbb{M}_1(\mathscr{X})$ such that $\mathbb{P}_\mu(\sigma_\alpha < \infty) = 1$, the random variables $\mathscr{E}_n(\alpha, f)$, $n \geq 1$, are independent and $\mathscr{E}_n(\alpha, f)$, $n \geq 2$, are i.i.d.*

6.6 Ratio Limit Theorems

Let P be a Markov kernel admitting a recurrent atom α. In this section, we consider the convergence of functionals of the Markov chains such as

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k)\,, \qquad \frac{\sum_{k=1}^{n} f(X_k)}{\sum_{k=1}^{n} g(X_k)}\,.$$

To obtain the limits of these quantities, when they exist, an essential ingredient is Proposition 6.5.1, i.e., the independence of the excursions between successive visits to a recurrent atom.

Lemma 6.6.1 *Let P be a Markov kernel that admits a recurrent atom α and let f be a finite λ_α-integrable function. Then for every initial distribution μ such that $\mathbb{P}_\mu(\sigma_\alpha < \infty) = 1$,*

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} f(X_k)}{\sum_{k=1}^{n} \mathbb{1}_\alpha(X_k)} = \lambda_\alpha(f) \;\; \mathbb{P}_\mu - \text{a.s.} \tag{6.6.1}$$

Proof. We first show that for every $f \in \mathrm{L}^1(\lambda_\alpha)$,

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} f(X_k)}{\sum_{k=1}^{n} \mathbb{1}_\alpha(X_k)} = \lambda_\alpha(f) \qquad \mathbb{P}_\alpha - \text{a.s.} \tag{6.6.2}$$

Let $f \in \mathrm{L}^1(\lambda_\alpha)$ be a nonnegative function and let $\{\mathscr{E}_k(\alpha,f),\, k\in\mathbb{N}^*\}$ be as in (6.5.2). The random variable $\mathscr{E}_1(\alpha,f)$ is $\mathscr{F}_{\sigma_\alpha}$-measurable, and by definition of λ_α, we have

$$\mathbb{E}_\alpha[\mathscr{E}_1(\alpha,f)] = \mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} f(X_k)\right] = \lambda_\alpha(f)\,.$$

By Corollary 6.5.2, the random variables $\{\mathscr{E}_k(\alpha,f), k\in\mathbb{N}^*\}$, are i.i.d. under $\mathbb{P}_\alpha$. Thus, the strong law of large numbers yields

$$\frac{1}{n}\sum_{k=1}^{\sigma_\alpha^{(n)}} f(X_k) = \frac{\mathscr{E}_1(\alpha,f)+\cdots+\mathscr{E}_n(\alpha,f)}{n} \xrightarrow{\mathbb{P}_\alpha\text{-a.s.}} \lambda_\alpha(f)\,.$$

The same convergence holds if we replace n by any integer-valued random sequence $\{\nu_n,\, n\in\mathbb{N}\}$ such that $\lim_{n\to\infty}\nu_n = \infty$ $\mathbb{P}_\alpha - \text{a.s.}$ For $n\in\mathbb{N}$, define $\nu_n = \sum_{k=1}^{n} \mathbb{1}_\alpha(X_k)$, the number of visits to the atom α before time n. Since α is a recurrent atom, it follows that $\nu_n \to \infty$ $\mathbb{P}_\alpha - \text{a.s.}$ Moreover,

$$\frac{\sum_{k=1}^{\sigma_\alpha^{(\nu_n)}} f(X_k)}{\nu_n} \le \frac{\sum_{k=1}^{n} f(X_k)}{\sum_{k=1}^{n} \mathbb{1}_\alpha(X_k)} \le \left(\frac{\nu_n+1}{\nu_n}\right) \frac{\sum_{k=1}^{\sigma_\alpha^{(\nu_n+1)}} f(X_k)}{\nu_n+1}\,.$$

Since the leftmost and rightmost terms have the same limit, we obtain (6.6.2). Writing $f = f^+ - f^-$, we obtain the same conclusion for $f \in \mathrm{L}^1(\lambda_\alpha)$.

Let now μ be an initial distribution such that $\mathbb{P}_\mu(\sigma_\alpha < \infty) = 1$. Since α is recurrent, it is also the case that $\mathbb{P}_\mu(N_\alpha = \infty) = 1$. This implies that $\lim_n \nu_n = \infty$ $\mathbb{P}_\mu -$a.s. Write, then, for $n \geq \sigma_\alpha$,

$$\frac{\sum_{k=1}^n f(X_k)}{\sum_{k=1}^n \mathbb{1}_\alpha(X_k)} = \frac{\sum_{k=1}^{\sigma_\alpha} f(X_k)}{\nu_n} + \frac{\sum_{k=\sigma_\alpha+1}^n f(X_k)}{1 + \sum_{k=\sigma_\alpha+1}^n \mathbb{1}_\alpha(X_k)} .$$

Since $\mathbb{P}_\mu(\sigma_\alpha < \infty) = 1$ and $\mathbb{P}_\mu(\lim_n \nu_n = \infty) = 1$, we have

$$\limsup_{n\to\infty} \frac{\sum_{k=1}^{\sigma_\alpha} f(X_k)}{\nu_n} = 0 \qquad \mathbb{P}_\mu -\text{a.s.} \tag{6.6.3}$$

Since $\mathbb{P}_\mu(\sigma_\alpha < \infty) = 1$, the strong Markov property and (6.6.2) yield

$$\mathbb{P}_\mu\left(\lim_{\ell\to\infty} \frac{\sum_{k=\sigma_\alpha+1}^{\sigma_\alpha+\ell} f(X_k)}{\sum_{k=\sigma_\alpha+1}^{\sigma_\alpha+\ell} \mathbb{1}_\alpha(X_k)} = \lambda_\alpha(f) \right) = \mathbb{P}_\alpha\left(\lim_{\ell\to\infty} \frac{\sum_{k=1}^{\ell} f(X_k)}{\sum_{k=1}^{\ell} \mathbb{1}_\alpha(X_k)} = \lambda_\alpha(f) \right) = 1 ,$$

showing that

$$\lim_{n\to\infty} \frac{\sum_{k=\sigma_\alpha+1}^n f(X_k)}{\sum_{k=\sigma_\alpha+1}^n \mathbb{1}_\alpha(X_k)} \mathbb{1}\{n \geq \sigma_\alpha\} = \lim_{\ell\to\infty} \frac{\sum_{k=\sigma_\alpha+1}^{\sigma_\alpha+\ell} f(X_k)}{\sum_{k=\sigma_\alpha+1}^{\sigma_\alpha+\ell} \mathbb{1}_\alpha(X_k)} = \lambda_\alpha(f) \qquad \mathbb{P}_\mu -\text{a.s.}$$

This relation and (6.6.3) prove (6.6.1). □

Theorem 6.6.2. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let α be an accessible and recurrent atom. Let λ be a nontrivial invariant measure for P. Then for every initial distribution μ such that $\mathbb{P}_\mu(\sigma_\alpha < \infty) = 1$ and all finite λ-integrable functions f, g such that $\lambda(g) \neq 0$,*

$$\lim_{n\to\infty} \frac{\sum_{k=1}^n f(X_k)}{\sum_{k=1}^n g(X_k)} = \frac{\lambda(f)}{\lambda(g)} \qquad \mathbb{P}_\mu -\text{a.s.}$$

Proof. By Theorem 6.4.2, $\lambda = \lambda(\alpha)\lambda_\alpha$ with $0 < \lambda(\alpha) < \infty$. This implies that $\mathrm{L}^1(\lambda_\alpha) = \mathrm{L}^1(\lambda)$ and

$$\frac{\lambda_\alpha(f)}{\lambda_\alpha(g)} = \frac{\lambda(f)}{\lambda(g)} .$$

Thus we can apply Lemma 6.6.1 to the functions f and g and take the ratio, since we have assumed that $\lambda_\alpha(g) \neq 0$. □

For a positive atom, we obtain the usual law of large numbers, and for a null recurrent atom, dividing by n instead of the number of visits to the atom in (6.6.1) yields a degenerate limit.

Corollary 6.6.3 *Let P be a Markov kernel with an accessible and recurrent atom α and let μ be a probability measure such that $\mathbb{P}_\mu(\sigma_\alpha < \infty) = 1$.*

(i) If α is positive and π is the unique invariant probability measure, then for every finite π-integrable function f,

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \stackrel{\mathbb{P}_\mu\text{-a.s.}}{\longrightarrow} \pi(f)\,.$$

(ii) If α is null recurrent and λ is a nontrivial invariant measure, then for every finite λ-integrable function f,

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \stackrel{\mathbb{P}_\mu\text{-a.s.}}{\longrightarrow} 0\,. \tag{6.6.4}$$

Proof. If P is positive recurrent, then the conclusion follows from Theorems 6.4.2 and 6.6.2 on setting $g \equiv 1$. Assume now that α is null recurrent. Then $\lambda(\mathsf{X}) = \lambda(\alpha)\lambda_\alpha(\mathsf{X}) = \lambda(\alpha)\mathbb{E}_\alpha[\sigma_\alpha] = \infty$. Let f be a nonnegative function such that $\lambda(f) < \infty$. Since λ is a σ-finite measure, for every $\varepsilon > 0$, we may choose a set F in such a way that $0 < \lambda(F) < \infty$ and $\lambda(f)/\lambda(F) \le \varepsilon$. Then setting $g = \mathbb{1}_F$ in Theorem 6.6.2, we obtain

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(X_k) \le \limsup_{n\to\infty} \frac{\sum_{k=1}^{n} f(X_k)}{\sum_{k=1}^{n} \mathbb{1}_F(X_k)} = \frac{\lambda(f)}{\lambda(F)} \le \varepsilon \qquad \mathbb{P}_\mu - \text{a.s.}$$

Since ε is arbitrary, this proves (6.6.4). □

6.7 The Central Limit Theorem

Let P be a Markov kernel with invariant probability measure π and let $f \in \mathbb{F}(\mathsf{X})$ be such that $\pi(|f|) < \infty$. We say that the sequence $\{f(X_k), k \in \mathbb{N}\}$ satisfies a central limit theorem (CLT) if there exists a constant $\sigma^2(f) \ge 0$ such that $n^{-1/2}\sum_{k=1}^{n}\{f(X_k) - \pi(f)\}$ converges in distribution to a Gaussian distribution with zero mean and variance $\sigma^2(f)$ under $\mathbb{P}_\mu$ for every initial distribution $\mu \in \mathbb{M}_1(\mathscr{X})$. Note that we allow the special case $\sigma^2(f) = 0$, which corresponds to weak convergence to 0. For an i.i.d. sequence, a CLT holds as soon as $\pi(|f|^2) < \infty$. This is no longer true in general for a Markov chain, and additional assumptions are needed.

Theorem 6.7.1. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let α be an attractive atom. Denote by π the unique invariant probability measure of P. Let $f \in \mathbb{F}(\mathsf{X})$ be a function satisfying*

$$\pi(|f|) < \infty, \qquad \mathbb{E}_\alpha\left[\left(\sum_{k=1}^{\sigma_\alpha}\{f(X_k) - \pi(f)\}\right)^2\right] < \infty. \tag{6.7.1}$$

Then for every initial distribution $\mu \in \mathbb{M}_1(\mathscr{X})$, we have

$$n^{-1/2}\sum_{k=1}^{n}\{f(X_k) - \pi(f)\} \stackrel{\mathbb{P}_\mu}{\Longrightarrow} \mathrm{N}(0, \sigma^2(f)), \tag{6.7.2}$$

with

$$\sigma^2(f) = \frac{1}{\mathbb{E}_\alpha[\sigma_\alpha]}\mathbb{E}_\alpha\left[\left(\sum_{k=1}^{\sigma_\alpha}\{f(X_k) - \pi(f)\}\right)^2\right]. \tag{6.7.3}$$

Proof. Without loss of generality, we assume that $\pi(f) = 0$. We decompose the sum $\sum_{k=1}^{n} f(X_k)$ into excursions between successive visits to the state α. Let $\mathscr{E}_j(\alpha, f)$, $j \geq 1$, be defined as in (6.5.2), and let $\nu_n = \sum_{k=1}^{n} \mathbb{1}_\alpha(X_k)$ be the number of visits to the atom α before n.

Applying Corollary 6.6.3 with $f = \mathbb{1}_\alpha$, we obtain

$$\frac{\nu_n}{n} \stackrel{\mathbb{P}_\mu\text{-a.s.}}{\longrightarrow} \pi(\alpha) = \frac{1}{\mathbb{E}_\alpha[\sigma_\alpha]}. \tag{6.7.4}$$

Thus $\nu_n \to \infty$ $\mathbb{P}_\mu$ $-$a.s., and we can consider only the event $\nu_n \geq 2$. Then

$$\sum_{k=1}^{n} f(X_k) = \mathscr{E}_1(\alpha, f) + \sum_{k=2}^{\nu_n} \mathscr{E}_j(\alpha, f) + \sum_{i=\sigma_C^{(\nu_n)}+1}^{n} f(X_i).$$

Since α is attractive and positive, by Corollary 6.5.2 the random variables $\mathscr{E}_j(\alpha, f)$, $j \geq 1$, are independent under $\mathbb{P}_\mu$ for every initial distribution μ, and $\mathscr{E}_j(\alpha, f)$, $j \geq 2$, are i.i.d. under $\mathbb{P}_\mu$. Theorem E.4.5, (6.7.1), and (6.7.4) imply that $n^{-1/2}\sum_{j=2}^{\nu_n} \mathscr{E}_j(\alpha, f)$ converges weakly under $\mathbb{P}_\mu$ to $\mathrm{N}(0, \sigma^2(f))$. The theorem will be proved if we show that

$$\lim_{n\to\infty} n^{-1/2}\left|\sum_{k=0}^{n-1} f(X_k) - \sum_{j=2}^{\nu_n} \mathscr{E}_j(\alpha,f)\right|$$
$$\leq \limsup_{n\to\infty} n^{-1/2}|\mathscr{E}_1(\alpha,f)| + \limsup_{n\to\infty} n^{-1/2} \sum_{k=\sigma_\alpha^{(\nu_n)}+1}^{n} f(X_k) = 0\,, \quad (6.7.5)$$

where the limits must hold in $\mathbb{P}_\mu$ probability for every initial distribution μ. Since the atom α is attractive and recurrent, $\mathbb{P}_\mu(\sigma_\alpha < \infty) = 1$. Therefore, the sum $\mathscr{E}_1(\alpha,|f|)$ has a $\mathbb{P}_\mu$ $-$ a.s. finite number of terms, and $\lim_{n\to\infty} n^{-1/2}\mathscr{E}_1(\alpha,f) = 0$ $\mathbb{P}_\mu$ $-$ a.s. To conclude, it remains to prove that as $n \to \infty$,

$$n^{-1/2}\textstyle\sum_{i=\sigma_\alpha^{(\nu_n)}+1}^{n} f(X_i) \stackrel{\mathbb{P}_\mu\ -\text{prob}}{\longrightarrow} 0\,. \quad (6.7.6)$$

To prove this convergence, we will use the following lemma.

Lemma 6.7.2 *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$, and let α be an attractive atom satisfying $\mathbb{E}_\alpha[\sigma_\alpha] < \infty$ and $\mu \in \mathbb{M}_1(\mathscr{X})$. Let $\nu_n = \sum_{k=1}^n \mathbb{1}_\alpha(X_k)$ be the number of visits to the atom α before n. Then for all $\varepsilon > 0$, there exists an integer $k > 0$ such that*

$$\sup_{n\in\mathbb{N}} \mathbb{P}_\mu(n - \sigma_\alpha^{(\nu_n)} > k) \leq \varepsilon\,.$$

Proof. Using the Markov property, we have for all $n \geq 1$ and $k \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}_\mu(n - \sigma_\alpha^{(\nu_n)} = k) &\leq \mathbb{P}_\mu(X_{n-k} \in \alpha, \sigma_\alpha \circ \theta_{n-k} > k) \\ &= \mathbb{P}_\mu(X_{n-k} \in \alpha)\mathbb{P}_\alpha(\sigma_\alpha > k) \leq \mathbb{P}_\alpha(\sigma_\alpha > k)\,.\end{aligned}$$

Since $\mathbb{E}_\alpha[\sigma_\alpha] < \infty$, this bound yields, for all $n \in \mathbb{N}$,

$$\mathbb{P}_\mu(n - \sigma_\alpha^{(\nu_n)} > k) \leq \sum_{j=k+1}^{\infty} \mathbb{P}_\alpha(\sigma_\alpha > j) \to_{k\to\infty} 0\,.$$

$\square$

We can now conclude the proof of Theorem 6.7.1. Let $\varepsilon > 0$. By Lemma 6.7.2, we may choose $k \in \mathbb{N}$ such that $\mathbb{P}_\mu(n - \sigma_\alpha^{(\nu_n)} > k) < \varepsilon/2$ for all $n \in \mathbb{N}$. Then

$$\begin{aligned}A_n &= \mathbb{P}_\mu\left(n^{-1/2}\left|\textstyle\sum_{k=\sigma_\alpha^{(\nu_n)}+1}^{n} f(X_k)\right| > \eta\right) \\ &\leq \mathbb{P}_\mu(n - \sigma_\alpha^{(\nu_n)} > k) + \mathbb{P}_\mu\left(n^{-1/2}\left|\textstyle\sum_{j=\sigma_\alpha^{(\nu_n)}+1}^{n} f(X_j)\right| > \eta, n - \sigma_\alpha^{(\nu_n)} \leq k\right) \\ &\leq \varepsilon/2 + \sum_{s=1}^{k} \mathbb{P}_\mu\left(n^{-1/2}\textstyle\sum_{j=\sigma_\alpha^{(\nu_n)}}^{\sigma_\alpha^{(\nu_n)}+k} |f(X_j)| > \eta, n - \sigma_\alpha^{(\nu_n)} = s\right) \\ &\leq \varepsilon/2 + \sum_{s=1}^{k} \mathbb{P}_\mu\left(n^{-1/2}\sum_{j=n-s}^{n-s+k} |f(X_j)| > \eta, X_{n-s} \in \alpha, \sigma_\alpha \circ \theta_{n-s} > s\right).\end{aligned}$$

Therefore, by the Markov property, we get $A_n \le \varepsilon/2 + \sum_{s=1}^{\infty} a_n(s)$, with

$$a_n(s) = \mathbb{P}_\alpha\left(n^{-1/2}\sum_{k=1}^{k}|f(X_k)| > \eta, \sigma_\alpha > s\right)\mathbb{1}_{\{s\le n\}}.$$

Note that for all $s \in \mathbb{N}$, $\lim_{n\to\infty} a_n(s) = 0$ and $a_n(s) \le \mathbb{P}_\alpha(\sigma_\alpha > s)$. Furthermore, since $\sum_{s=1}^{\infty}\mathbb{P}_\alpha(\sigma_\alpha > s) < \infty$, Lebesgue's dominated convergence theorem shows that $\lim_{n\to\infty}\sum_{s=1}^{\infty} a_n(s) = 0$ and therefore, since ε is arbitrary, that

$$\lim_{n\to\infty}\mathbb{P}_\mu\left(n^{-1/2}\left|\sum_{k=\sigma_\alpha^{(\nu_n)}+1}^{n} f(X_k)\right| > \eta\right) = 0.$$

This establishes (6.7.6) and concludes the proof of Theorem 6.7.1. □

Remark 6.7.3. Let f be a measurable function such that $\pi(f^2) < \infty$ and $\pi(f) = 0$. Since $\{X_k, k \in \mathbb{N}\}$ is a stationary sequence under $\mathbb{P}_\pi$, we obtain

$$\begin{aligned}\mathbb{E}_\pi\left[\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n} f(X_k)\right)^2\right] &= \frac{1}{n}\sum_{k=0}^{n-1}\mathbb{E}_\pi[f^2(X_k)] + \frac{2}{n}\sum_{k=0}^{n-1}\sum_{\ell=0}^{k-1}\mathbb{E}_\pi[f(X_k)f(X_\ell)] \\ &= \mathbb{E}_\pi[f^2(X_0)] + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\mathbb{E}_\pi[f(X_0)f(X_k)].\end{aligned}$$

If the series $\sum_{k=1}^{\infty}|\mathbb{E}_\pi[f(X_0)f(X_k)]|$ is convergent, then

$$\lim_{n\to\infty}\mathbb{E}_\pi\left[\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n} f(X_k)\right)^2\right] = \mathbb{E}_\pi[f^2(X_0)] + 2\sum_{k=1}^{\infty}\mathbb{E}_\pi[f(X_0)f(X_k)]. \tag{6.7.7}$$

▲

6.8 Exercises

6.1. A Galton–Watson process is a stochastic process $\{X_n, n \in \mathbb{N}\}$ that evolves according to the recurrence $X_0 = 1$ and

$$X_{n+1} = \sum_{j=1}^{X_n}\xi_j^{(n+1)}, \tag{6.8.1}$$

where $\{\xi_j^{(n+1)} : n, j \in \mathbb{N}\}$ is a set of i.i.d. nonnegative integer-valued random variables with distribution ν. The random variable X_n can be thought of as the number of descendants in the nth generation, and $\{\xi_j^{(n+1)}, j = 1, \ldots, X_n\}$ represents the number of (male) children of the jth descendant of the nth generation.

The conditional distribution of X_{n+1} given the past depends only on the current size of the population X_n and the number of offspring of each individual $\{\xi_j^{(n+1)}\}_{j=1}^{X_n}$, which are conditionally independent given the past.

1. Show that the process $\{X_n, n \in \mathbb{N}\}$ is a homogeneous Markov chain and determine its transition matrix.

We assume that the offspring distribution has a finite mean $\mu = \sum_{k=1}^{\infty} k\nu(k) < \infty$ and that $\nu(0) > 0$. Denote by $\mathbb{E}_x[X_k]$ the average number of individuals at the kth generation.

2. Show that the state $\{0\}$ is accessible. Show that all the states except 0 are transient.
3. Show that for all $x \in \mathbb{N}$ and $k \in \mathbb{N}$, $\mathbb{E}_x[X_k] = x\mu^k$.
4. Show that if $\mu < 1$, then $\mathbb{P}_x(\tau_0 < \infty) = 1$, and that the Markov kernel is recurrent.

6.2. We pursue here the study of the Galton–Watson process. We assume that the offspring distribution ν satisfies $\nu(0) > 0$ and $\nu(0) + \nu(1) < 1$ (it places some positive probability on some integer $k \geq 2$). We assume in this exercise that the initial size of the population is $X_0 = 1$. Denote by $\Phi_k(u) = \mathbb{E}\left[u^{X_k}\right]$ ($|u| \leq 1$) the generating function of the random variable X_k and by $\varphi(u) = \mathbb{E}\left[u^{\xi_1^{(1)}}\right]$ the generating function of the offspring distribution.

1. Show that $\Phi_0(u) = u$ and for $k \geq 0$,

$$\Phi_{k+1}(u) = \varphi(\Phi_k(u)) = \Phi_k(\varphi(u)).$$

2. Show that if the mean offspring number is given by $\mu = \sum_{k=1}^{\infty} k\nu(k) < \infty$, then the expected size of the nth generation is $\mathbb{E}[X_n] = \mu^n$.
3. Show that if the variance satisfies $\sigma^2 = \sum_{k=0}^{\infty}(k-\mu)^2\nu(k) < \infty$, then the variance of X_k is finite, and give a formula for it.
4. Show that $\varphi : [0,1] \to \mathbb{R}^+$ is strictly increasing, strictly convex with strictly increasing first derivative, and $\varphi(1) = 1$.
5. Show that $\Phi_n(0) = \mathbb{P}(Z_n = 0)$ and that $\lim_{n\to\infty} \Phi_n(0)$ exists and is equal to the extinction probability $\rho = \mathbb{P}(\sigma_0 < \infty)$.
6. Show that the extinction probability ρ is the smallest nonnegative root of the fixed-point equation $\phi(r) = r$.

A Galton–Watson process with mean offspring number μ is said to be supercritical if $\mu > 1$, critical if $\mu = 1$, and subcritical if $\mu < 1$.

7. In the supercritical case, show that the fixed-point equation has a unique root $\rho < 1$ less than one. [Hint: Use Exercise 6.3.]
8. In the critical and subcritical cases, show that the only root is $\rho = 1$.

This implies that the extinction is certain if and only if the Galton–Watson process is critical or subcritical. If, on the other hand, it is supercritical, then the probability of extinction is $\rho < 1$. (Nevertheless, the Markov kernel remains recurrent!)

6.3. Let $\{b_k, k \in \mathbb{N}\}$ be a probability on $\mathbb{N}$ such that $b_0 > 0$ and $b_0 + b_1 < 1$. For $s \in [0,1]$, set $\phi(s) = \sum_{k=0}^{\infty} b_k s^k$, the generating function of $\{b_k, k \in \mathbb{N}\}$. Show that:

1. If $\sum_{k=1}^{\infty} k b_k \leq 1$, then $s = 1$ is the unique solution to $\phi(s) = s$ in $[0,1]$.
2. If $\sum_{k=1}^{\infty} k b_k > 1$, then there exists a single $s_0 \in (0,1)$ such that $\phi(s_0) = s_0$.

6.4 (Simple random walk on $\mathbb{Z}$). The Bernoulli random walk on $\mathbb{Z}$ is defined by

$$P(x,x+1) = p, \quad P(x,x-1) = q, \quad p \geq 0, \quad q \geq 0, \quad p+q = 1.$$

1. Show that $P^n(0,x) = p^{(n+x)/2} q^{(n-x)/2} \binom{n}{(n+x)/2}$ when the sum $n+x$ is even and $|x| \leq n$ and $P^n(0,x) = 0$ otherwise.
2. Deduce that for all $n \in \mathbb{N}$, $P^{2n}(0,0) = \binom{2n}{n} p^n q^n$.
3. Show that the expected number of visits to $\{0\}$ is $U(0,0) = \sum_{k=0}^{\infty} \binom{2k}{k} p^k q^k$.
4. Show that

$$P^{2k}(0,0) = \binom{2k}{k} p^k q^k \sim_{k \to \infty} (4pq)^k (\pi k)^{-1/2}.$$

Assume first that $p \neq 1/2$.

5. Show that the state $\{0\}$ is transient.

Assume now that $p = 1/2$.

6. Show that the state $\{0\}$ is recurrent.
7. Show that the counting measure on $\mathbb{Z}$ is an invariant measure and that the state $\{0\}$ is null recurrent.

6.5 (Simple symmetric random walk on $\mathbb{Z}^2$ and $\mathbb{Z}^3$). A random walk on $\mathbb{Z}^d$ is called simple and symmetric if its increment distribution gives equal weight $1/(2d)$ to the points $z \in \mathbb{Z}^d$ satisfying $|z| = 1$. The transition kernel of the d-dimensional simple random walk is given by $P(x,y) = 1/(2d)$ if $|y-x| = 1$ and $P(x,y) = 0$ otherwise.

Consider a symmetric random walk of $\mathsf{X} = \mathbb{Z}^2$, i.e., $P(x,y) = 1/4$ if $|x-y| = 1$ and $P(x,y) = 0$ otherwise, where $|\cdot|$ stands here for the Euclidean norm. This means that the chain may jump from a point $x = (x^1, x^2)$ to one of its four neighbors, $x \pm e_1$ and $x \pm e_2$, where $e_1 = (1,0)$ and $e_2 = (0,1)$.

Let X_n^+ and X_n^- be the orthogonal projections of X_n on the diagonal lines $y = x$ and $y = -x$, respectively.

1. Show that $\{X_n^+, n \in \mathbb{N}\}$ and $\{X_n^-, n \in \mathbb{N}\}$ are independent simple symmetric random walks on $2^{-1/2}\mathbb{Z}$ and $X_n = (0,0)$ if and only if $X_n^+ = X_n^- = 0$.
2. Show that $P^{(2n)}((0,0),(0,0)) = \left(\binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \right)^2 \sim_{n \to \infty} 1/(\pi n)$. [Hint: Stirling's formula.]
3. Show that the state $\{0,0\}$ is recurrent. Is it positive or null recurrent?

Consider now the case $d = 3$. The transition probabilities are given by $P(x,y) = 1/6$ when $|x-y| = 1$ and $P(x,y) = 0$ otherwise. Thus the chain jumps from one state to one of its nearest neighbors with equal probability.

4. Show that $\sum_{m=0}^{\infty} P^{6m}(0,0) < \infty$. [Hint: Show that $P^{2n}(0,0) = O(n^{-3/2})$.]
5. Show that $U(0,0) < \infty$ and therefore that the state $\{0\}$ is transient.

6.6. Let S be a subset of $\mathbb{Z}$. Assume that g.c.d.$(S) = 1$, $S_+ = S \cap \mathbb{Z}_+^* \neq \emptyset$, $S_- = S \cap \mathbb{Z}_-^* \neq \emptyset$. Show that

$$I = \{x \in \mathbb{Z} : x = x_1 + \cdots + x_n, \text{ for some } n \in \mathbb{N}^*, \; x_1, \ldots, x_n \in S\} = \mathbb{Z}\,.$$

6.7 (Random walks on $\mathbb{Z}$). Let $\{Z_n, n \in \mathbb{N}\}$ be an i.i.d. sequence on $\mathbb{Z}$ with distribution ν and consider the random walk $X_n = X_{n-1} + Z_n$. The kernel P is defined by

$$P(x,y) = \nu(y-x) = P(0, y-x)\,.$$

We assume that $\nu \neq \delta_0$ and $\sum_{z \in \mathbb{Z}} |z| \nu(z) < \infty$ and set $m = \sum_{z \in \mathbb{Z}} z \nu(z)$. We set $\mathsf{S} = \{z \in \mathbb{Z}, \; \nu(z) > 0\}$ and assume that $1 = g.c.d.(\mathsf{S})$.

1. Show that if $m \neq 0$, the Markov kernel P is transient.

In the sequel we assume that $m = 0$.

2. Show that for all $x, y \in \mathbb{Z}$, $\mathbb{P}_x(\sigma_y < \infty) > 0$. [Hint: Use Exercise 6.6.]
3. Let $\varepsilon > 0$. Show that

$$1 = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{P}_0(|X_k| \leq \varepsilon k) \leq \liminf_{n \to \infty} \frac{1}{n} U(0, [-\lfloor \varepsilon n \rfloor, \lfloor \varepsilon n \rfloor]).$$

4. Show that

$$\liminf_{n \to \infty} \frac{1}{n} U(0, [-\lfloor \varepsilon n \rfloor, \lfloor \varepsilon n \rfloor]) \leq 2\varepsilon U(0,0)\,.$$

5. Show that the Markov kernel P is recurrent.

6.8. This is a follow-up to Exercise 4.7. Let P be a Markov kernel on $\mathbb{N}$ with transition probability given by $P(0,1) = 1$ and for $x \geq 1$,

$$P(x, x+1) + P(x, x-1) = 1, \quad P(x, x+1) = \left(\frac{x+1}{x}\right)^2 P(x, x-1)\,.$$

1. Show that the states $x \in \mathbb{N}$ are accessible.
2. Show that the Markov kernel P is transient.
3. Show that for all $x \in \mathsf{X}$, $\mathbb{P}_x(\liminf_{n \to \infty} X_n = \infty) = 1$.

6.9. Let P be a Markov kernel on $\mathbb{N}$ with transition probability given by $P(0,1) = 1$, and for $x \geq 1$,

$$P(x, x+1) + P(x, x-1) = 1, \quad P(x, x+1) = \left(\frac{x+1}{x}\right)^{\alpha} P(x, x-1)\,,$$

where $\alpha \in (0, \infty)$.

1. Show that the states $x \in \mathbb{N}$ are accessible.
2. Determine the values of α for which the Markov kernel P is recurrent or transient.

6.10. Let $\{X_k,\, k \in \mathbb{N}\}$ be a stochastic unit root process, as defined in Example 6.4.6, that satisfies (6.4.3) with $r = \mathbb{1}_{\mathbb{R}^+}$ and assume that $X_0 = 0$. We use the notation of Example 6.4.6. Show that $\{Y_k = X_k^+,\, k \in \mathbb{N}\}$ is a reflected Markov chain as in Example 6.1.2, that is, it satisfies $Y_k = (Y_{k-1} + Z_k)^+$.

6.9 Bibliographical Notes

All the results presented in this chapter are very classical. They were mostly developed for discrete-valued Markov chains, but the translation of these results for atomic chains is immediate.

The recurrence–transience dichotomy Theorem 6.2.7 for discrete-state-space Markov chains appears in Feller (1971) (see also Chung (1967), Çinlar (1975)).

The stochastic unit root model given in Example 6.4.6 was proposed by Granger and Swanson (1997) and further studied in Gourieroux and Robert (2006).

The essential idea of studying the excursions of the chain between two successive visits to an atom is already present in Doeblin (1938) and is implicit in some earlier works of Kolmogorov. (For French-speaking readers, we recommend Charmasson et al. (2005) on the short and tragic story of Wolfgang Doeblin, the son of the german writer Alfred Doeblin, a mathematical genius who revolutionized Markov's chain theory just before being killed in the early days of the Second World War.) This decomposition has many consequences in Markov chain theory. The first is the law of large numbers for empirical averages, which becomes an elementary consequence of the theory of stopped random walks. It is almost equally easy to get the ratio limit theorem Theorem 6.6.2: these results were established in Chung (1953, 1954) for discrete-state-space Markov chains (see Port (1965) and (Orey, 1971, Chapter 3), which give a good overview of ratio limit theorems, which also play an important role in ergodic theory.)

The proof of the central limit theorem (Theorem 6.7.1) is based on Anscombe's theorem (Theorem Anscombe (1952, 1953)), stated and proved in Theorem E.4.5 (this is a special case of Anscombes's theorem, which in this form with a direct proof is due to Rényi (1957)). The proof of Theorem 6.7.1 is borrowed from Chung (1967) (see also (Meyn and Tweedie, 2009, Chapter 17)). A beautiful discussion of Anscombe's theorem with many references is given in Gut (2012). The same approach can be extended to obtaina functional version of the CLT or the law of the iterated logarithm.

Chapter 7
Markov Chains on a Discrete State Space

In this chapter we will discuss the case in which the state space X is discrete, which means either finite or countably infinite. In this case, it will always be assumed that $\mathscr{X} = \mathscr{P}(\mathsf{X})$, the set of all subsets of X. Since every state is an atom, we will first apply the results of Chapter 6 and then highlight the specificities of Markov chains on countable state spaces. In particular, in Section 7.5 we will obtain simple drift criteria for transience and recurrence, and in Section 7.6 we will make use for the first time of coupling arguments to prove the convergence of the iterates of the kernel to the invariant probability measure.

7.1 Irreducibility, Recurrence, and Transience

Let $P = \{P(x,y) : (x,y) \in \mathsf{X}\}$ be a Markov kernel on a denumerable state space. Theorem 6.2.2 applied to the present framework yields that for every state $\mathsf{a} \in \mathsf{X}$,

$$U(\mathsf{a},\mathsf{a}) = \infty \Leftrightarrow \mathbb{P}_{\mathsf{a}}(\sigma_{\mathsf{a}} < \infty) = 1 \Leftrightarrow \mathbb{P}_{\mathsf{a}}(N_{\mathsf{a}} = \infty) = 1\,,$$
$$U(\mathsf{a},\mathsf{a}) < \infty \Leftrightarrow \mathbb{P}_{\mathsf{a}}(\sigma_{\mathsf{a}} < \infty) < 1 \Leftrightarrow \mathbb{P}_{\mathsf{a}}(N_{\mathsf{a}} = \infty) < 1\,.$$

We now introduce the following definition, which anticipates those of Chapter 9 for general state-space Markov chains.

Definition 7.1.1 (Irreducibility, strong irreducibility) *Let X be a discrete state space and P a Markov kernel on X.*

(i) P is irreducible if it admits an accessible state.
(ii) P is strongly irreducible if all the states are accessible.

It should be stressed that Definition 7.1.1 is not the usual notion of irreducibility for a Markov kernel on a discrete state space. In most books, a Markov kernel on

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_7

a discrete state space is said to be irreducible if all the states are accessible: in this book, this property is referred to as strong irreducibility (this notion is specific to Markov kernels on a discrete state space and does not have a natural extension to general state spaces).

Definition 7.1.1 is in line with the definition of irreducibility for general state spaces, which will be introduced in Chapter 9, and therefore we do not comply with the usual terminology in order to avoid having two conflicting definitions of irreducibility.

We now turn to transience and recurrence, which were introduced in Definition 6.2.1. The following result is simply a repetition of Theorem 6.2.7.

Theorem 7.1.2. *Let P be an irreducible kernel on a discrete state space X. Then P is either transient or recurrent but not both.*

(i) P is recurrent if and only if it admits an accessible recurrent state. If P is recurrent, then for all accessible states $x, y \in \mathsf{X}$,

$$\mathbb{P}_x(N_y = \infty) = \mathbb{P}_y(N_x = \infty) = 1. \tag{7.1.1}$$

(ii) P is transient if and only if it admits an accessible transient state. If P is transient, then $U(x,y) < \infty$ for all $x, y \in \mathsf{X}$.

Remark 7.1.3 The condition of Theorem 7.1.2 (i) can be slightly improved: assuming that the recurrent state is accessible is not required. Indeed, Proposition 6.2.4 implies that if the Markov kernel P is irreducible (i.e., admits an accessible atom), then every recurrent state is also accessible. Moreover, if P is strongly irreducible, then (7.1.1) is satisfied for all $x, y \in \mathsf{X}$. ▲

Theorem 7.1.4. *Assume that P has an invariant probability measure π.*

(i) Every $x \in \mathsf{X}$ such that $\pi(x) > 0$ is recurrent.
(ii) If P is irreducible, then P is recurrent.

Proof. The result follows from Proposition 6.2.8 and Theorem 7.1.2. □

7.2 Invariant Measures, Positive and Null Recurrence

Let P be a recurrent irreducible Markov kernel on a discrete state space X. Let $\mathsf{a} \in \mathsf{X}$ be an accessible and recurrent state. By Theorem 6.4.2, P admits an invariant measure, and all invariant measures are proportional to the measure λ_{a} defined by

$$\lambda_{\mathsf{a}}(x) = \mathbb{E}_{\mathsf{a}}\left[\sum_{k=1}^{\sigma_{\mathsf{a}}} \mathbb{1}\{X_k = x\}\right], \qquad x \in \mathsf{X}. \tag{7.2.1}$$

Note that $\lambda_{\mathsf{a}}(\mathsf{a}) = 1$ and $\lambda_{\mathsf{a}}(\mathsf{X}) = \mathbb{E}_{\mathsf{a}}[\sigma_{\mathsf{a}}]$, which is not necessarily finite. We now restate Theorem 6.4.2 in the present context.

Theorem 7.2.1. *If P is a recurrent irreducible Markov kernel on a discrete state space X, then there exists a nontrivial invariant measure λ, unique up to multiplication by a positive constant. For every accessible state $\mathsf{a} \in \mathsf{X}$, the measure λ_{a} defined in (7.2.1) is the unique invariant measure λ such that $\lambda(\mathsf{a}) = 1$.*

(i) If $\mathbb{E}_{\mathsf{a}}[\sigma_{\mathsf{a}}] < \infty$ for one accessible state a, then the same property holds for all accessible states, and there exists a unique invariant probability π given for all $x \in \mathsf{X}$ by

$$\pi(x) = \frac{\mathbb{E}_{\mathsf{a}}\left[\sum_{k=0}^{\sigma_{\mathsf{a}}-1} \mathbb{1}\{X_k = x\}\right]}{\mathbb{E}_{\mathsf{a}}[\sigma_{\mathsf{a}}]} = \frac{\mathbb{E}_{\mathsf{a}}\left[\sum_{k=1}^{\sigma_{\mathsf{a}}} \mathbb{1}\{X_k = x\}\right]}{\mathbb{E}_{\mathsf{a}}[\sigma_{\mathsf{a}}]}. \tag{7.2.2}$$

(ii) If $\mathbb{E}_{\mathsf{a}}[\sigma_{\mathsf{a}}] = \infty$ for one accessible state a, then the same property holds for all accessible states, and all the invariant measures are infinite.

We formalize in the next definition the dichotomy stated in Theorem 7.2.1.

Definition 7.2.2 (Positive and null-recurrent Markov kernels) *Let P be an irreducible Markov kernel on a discrete state space X. The Markov kernel is positive if it admits an invariant probability. The Markov kernel P is null recurrent if it is recurrent and all its invariant measures are infinite.*

When P is positive, the previous result provides an explicit relation between the unique invariant probability and the mean of the first return time to a given accessible set a. Indeed, applying (7.2.2) with $x = \mathsf{a}$ yields

$$\pi(\mathsf{a}) = \frac{\lambda_{\mathsf{a}}(\mathsf{a})}{\lambda_{\mathsf{a}}(\mathsf{X})} = \frac{\lambda_{\mathsf{a}}(\mathsf{a})}{\mathbb{E}_{\mathsf{a}}[\sigma_{\mathsf{a}}]} = \frac{1}{\mathbb{E}_{\mathsf{a}}[\sigma_{\mathsf{a}}]}. \tag{7.2.3}$$

Corollary 7.2.3 (Finite state space) *If the state space X is finite, then every irreducible Markov kernel is positive.*

Proof. By definition, for every $x \in \mathsf{X}$,

$$\sum_{y \in \mathsf{X}} U(x,y) = \mathbb{E}_x \left[\sum_{y \in \mathsf{X}} N_y \right] = \infty \,.$$

Therefore, if X is finite, there must exist a state $y \in \mathsf{X}$ satisfying $U(x,y) = \infty$. By the maximum principle (see Lemma 6.1.3), $U(y,y) \geq U(x,y) = \infty$, and the state y is recurrent. Therefore, P admits a nontrivial invariant measure, which is necessarily finite, since X is finite. □

7.3 Communication

We now introduce the notion of communication between states, which yields a classification of states into classes of recurrent and transient states. The notion of communication has no equivalent in the theory of general state-space Markov chains.

Definition 7.3.1 (Communication) *A state x leads to the state y, which we write $x \to y$, if $\mathbb{P}_x(\tau_y < \infty) > 0$. Two states x and y communicate, which we write $x \leftrightarrow y$, if $x \to y$ and $y \to x$.*

Equivalently, $x \to y$ if $U(x,y) > 0$ or if there exists $n \geq 0$ such that $P^n(x,y) = \mathbb{P}_x(X_n = y) > 0$. The most important property of the communication relation is that it is an equivalence relation.

Proposition 7.3.2 *The relation of communication between states is an equivalence relation.*

Proof. By definition, $x \leftrightarrow x$ for all x and $x \leftrightarrow y$ if and only if $y \leftrightarrow x$. If $x \to y$ and $y \to z$, then there exist integers n and m such that $P^n(x,y) > 0$ and $P^m(y,z) > 0$. Then the Chapman–Kolmogorov equation (1.2.5) implies

$$P^{n+m}(x,z) \geq P^n(x,y)P^m(y,z) > 0 \,.$$

This proves that $x \to z$. □

Therefore, the state space X may be partitioned into equivalence classes for the communication relation. The equivalence class of the state x is denoted by $C(x)$ i.e., $C(x) = \{y \in \mathsf{X} : x \leftrightarrow y\}$. Note that by definition, a state communicates with itself.

If the kernel P is not irreducible, there may exist transient and recurrent states. A transient state may communicate only with itself. Moreover, we already know by Proposition 6.2.4 that a recurrent state may lead only to another recurrent state. As a consequence, an equivalence class for the communication relation contains only recurrent states or only transient states. A class that contains only recurrent states will be called a recurrent class.

Theorem 7.3.3. *Let P be a Markov kernel on a discrete state space X. Then there exists a partition $\mathsf{X} = (\cup_{i \in I} \mathsf{R}_i) \cup \mathsf{T}$ of X such that*

- *if $x \in \mathsf{T}$, then x is transient;*
- *for every $i \in I$, R_i is absorbing, and the trace of P on R_i is strongly irreducible and recurrent.*

Assume that P is irreducible.

(i) If P is transient, then $\mathsf{X} = \mathsf{T}$.

(ii) P is recurrent, then $\mathsf{X} = \mathsf{T} \cup \mathsf{R}$, $\mathsf{R} \neq \emptyset$, and the trace of P on R is strongly irreducible and recurrent. Moreover, there exists a unique (up to a multiplicative constant) P-invariant measure λ, and $\mathsf{R} = \{x \in \mathsf{X} : \lambda(x) > 0\}$.

Proof. Since communication is an equivalence relation and an equivalence class contains either transient states or recurrent states, we can define T as the set of all transient states, and the sets R_i are the recurrent classes.

If C is a recurrent class and $y \in C$, then $U(y,y) = \infty$. Applying the maximum principle for atomic chains, Lemma 6.1.3, we obtain $U(x,y) = \mathbb{P}_x(\tau_y < \infty) U(y,y) = \infty$ for all $x \in C$.

Let us finally prove that a recurrent class C is absorbing. Let $x \in C$ and $y \in \mathsf{X}$ be such that $x \to y$. By Proposition 6.2.4, y is recurrent and $y \to x$. Thus $y \in C$, and this proves that $P(x,C) = 1$.

Assume now that P is irreducible. By Theorem 7.1.2, P is either transient or recurrent. If P is transient, then all the states are transient and $\mathsf{X} = \mathsf{T}$. If P is recurrent, then there exists a recurrent accessible state a, and $\mathsf{X} = \mathsf{T} \cup \mathsf{R}$, where R is the equivalence class of a. Denote by λ_{a} the unique invariant measure such that $\lambda_{\mathsf{a}}(\mathsf{a}) = 1$ given by (7.2.1). Every $x \in \mathsf{R}$ is recurrent, and Theorem 7.2.1 implies that $\lambda_{\mathsf{a}}(x) > 0$. Conversely, let x be a state such that $\lambda_{\mathsf{a}}(x) > 0$. Then by definition of λ_{a}, we have

$$0 < \lambda_{\mathsf{a}}(x) = \mathbb{E}_{\mathsf{a}}\left[\sum_{k=1}^{\sigma_{\mathsf{a}}} \mathbb{1}\{X_k = x\}\right] .$$

This implies that $\mathbb{P}_{\mathsf{a}}(\sigma_x < \infty) > 0$, and thus x is accessible and recurrent by Proposition 6.2.4. This proves that $\mathsf{R} = \{x \in \mathsf{X} : \lambda(x) > 0\}$ is the set of accessible recurrent states. $\square$

Example 7.3.4. Let $\mathsf{X} = \mathbb{N}$ and let P be the kernel on X defined by $P(0,0) = 1$, and for $n \geq 1$,

$$P(n, n+1) = p_n \,, \qquad P(n,0) = 1 - p_n \,,$$

where $\{p_n, n \in \mathbb{N}^*\}$ is a sequence of positive real numbers such that $0 < p_n < 1$. Then P is irreducible, since the state 0 is accessible, and all states are transient except the absorbing state 0. The state space can be decomposed as $\mathsf{X} = \mathsf{R} \cup \mathsf{T}$ with $\mathsf{R} = \{0\}$ and $\mathsf{T} = \mathbb{N}^*$. Every invariant measure for P is a multiple of the Dirac mass at 0.

Moreover, for all $k \geq n \geq 1$,

$$\mathbb{P}_n(\sigma_0 > k) = p_n \cdots p_k\,.$$

This yields, for all $n \geq 1$,

$$\mathbb{P}_n(\sigma_0 = \infty) = \prod_{k=n}^{\infty} p_k\,.$$

If $\sum_{k=1}^{\infty} \log(1/p_k) < \infty$, then $\mathbb{P}_n(\sigma_0 = \infty) > 0$ for every $n \geq 1$. Therefore, with positive probability, the Markov chain started at n does not hit the state 0. ◀

7.4 Period

Since every singleton is an atom, Definition 6.3.1 is still applicable: the period $d(x)$ of the state x is the greatest common divisor of the set $E_x = \{n > 0 : P^n(x,x) > 0\}$. If P is irreducible, then every accessible state x has a finite period, since by definition there exists at least one $n > 0$ for which $P^n(x,x) > 0$. By Proposition 6.3.3, if x is accessible, then the set E_x is stable by addition, and there exists an integer n_0 such that $nd(x) \in E_x$ for all $n \geq n_0$. Moreover, by Proposition 6.3.4, all accessible states have the same period, and the period of an irreducible kernel is the common period of all accessible states; the kernel is said to be aperiodic if its period is 1. For a discrete state space, the only additional result is the following cyclical decomposition of the state space.

Theorem 7.4.1. *Assume that P is an irreducible Markov kernel on a discrete state space X with period d. Let X_P^+ be the set of all the accessible states for P. Then there exist disjoint sets $D_0, \ldots, D_{d-1}$ such that $\mathsf{X}_P^+ = \bigcup_{i=0}^{d-1} D_i$ and $P(x, D_{i+1}) = 1$ for every $x \in D_i$ and $i = 0, 1, \ldots, d-1$, with the convention $D_d = D_0$. This decomposition is unique up to permutation.*

Proof. If $d = 1$, there is nothing to prove. Fix $\mathsf{a} \in \mathsf{X}_P^+$. For $i = 0, \ldots, d-1$, let D_i be the set defined by

$$D_i = \left\{ x \in \mathsf{X}_P^+ \,:\, \sum_{n=1}^{\infty} P^{nd-i}(x, \mathsf{a}) > 0 \right\}.$$

Since a is accessible, $\cup_{i=0}^{d-1} D_i = \mathsf{X}_P^+$. For $i, i' \in \{0, \dots, d-1\}$, if $y \in D_i \cap D_{i'}$, then there exist $m, n \in \mathbb{N}^*$ such that $P^{md-i}(y, \mathsf{a}) > 0$ and $P^{nd-i'}(y, \mathsf{a}) > 0$. Since y is accessible, there also exists $\ell \in \mathbb{N}$ such that $P^{\ell}(\mathsf{a}, y) > 0$. This yields $P^{md-i+\ell}(\mathsf{a}, \mathsf{a}) > 0$ and $P^{nd-i'+\ell}(\mathsf{a}, \mathsf{a}) > 0$. Therefore, d divides $i - i'$, and thus $i = i'$. This shows that the sets D_i, $i = 0, \dots, d-1$, are mutually disjoint. Let $x, y \in \mathsf{X}_P^+$ be such that $P(y, x) > 0$. If $x \in D_i$, then there exists $k \in \mathbb{N}^*$ such that $P^{kd-i}(x, \mathsf{a}) > 0$, and thus

$$P^{kd-i+1}(y, \mathsf{a}) \geq P(y, x) P^{kd-i}(x, \mathsf{a}) > 0.$$

Thus $y \in D_{i-1}$ if $i \geq 1$, and $y \in D_{d-1}$ if $i = 0$.

Let now $F_0, \dots, F_{d-1}$ be a partition of X_P^+ such that $P(x, F_{i+1}) = 1$ for all $x \in F_i$ and $i = 0, \dots, d-1$, with the convention $F_d = F_0$. Up to a permutation, we can assume that $\mathsf{a} \in F_0$. It suffices to prove that $D_0 = F_0$. Let $x \in F_0$ and $n \in \mathbb{N}$ be such that $P^n(x, \mathsf{a}) > 0$. Since x and a are both elements of F_0, n must be a multiple of d. Thus $F_0 \subset D_0$. Conversely, if $x \in D_0$, then there exists $k \geq 1$ such that $P^{kd}(x, a) > 0$, and thus $x \in F_0$. This proves that $D_0 = F_0$ and that the decomposition is unique. $\square$

7.5 Drift Conditions for Recurrence and Transience

In this section, we make use for the first time of the so-called drift conditions to prove the recurrence or the transience of a Markov kernel. A drift condition is a relation between a nonnegative measurable function f and Pf.

Theorem 7.5.1. *Assume that there exist an accessible set F and a function $W : \mathsf{X} \to \mathbb{R}_+^*$ satisfying*

(i) $PW(x) \leq W(x)$ for all $x \in F^c$;
(ii) there exists an accessible state $x_0 \in F^c$ such that $W(x_0) < \inf_{x \in F} W(x)$.

Then P is transient.

Proof. Since $\inf_{x \in F} W(x) > W(x_0) > 0$, we can assume without loss of generality that $\inf_{x \in F} W(x) = 1 > W(x_0) > 0$. Then the assumptions become $PW(x) \leq W(x)$ for $x \in F^c$, $W(x_0) < 1$ for some $x_0 \in F^c$, and $W(x) \geq 1$ for all $x \in F$. By Corollary 4.4.7, these properties imply that $\mathbb{P}_x(\tau_F < \infty) \leq W(x)$ for all $x \in \mathsf{X}$. Thus

$$\mathbb{P}_{x_0}(\sigma_F < \infty) = \mathbb{P}_{x_0}(\tau_F < \infty) \leq W(x_0) < 1\,.$$

Since x_0 is accessible, it follows that $\mathbb{P}_x(\sigma_{x_0} < \infty) > 0$ for all $x \in F$. The previous arguments and the strong Markov property yield, for $x \in F$,

$$\mathbb{P}_x(\sigma_F = \infty) \geq \mathbb{P}_x(\sigma_F = \infty, \sigma_{x_0} < \infty) = \mathbb{P}_{x_0}(\sigma_F = \infty)\mathbb{P}_x(\sigma_{x_0} < \infty) > 0\,.$$

Hence $\mathbb{P}_x(\sigma_F < \infty) < 1$ for all $x \in F$, and thus all the states in F are transient. By assumption, F contains at least one accessible state; thus P is transient. □

We now provide a drift condition for recurrence.

Theorem 7.5.2. *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that there exist a finite subset $F \subset \mathsf{X}$ and a finite nonnegative function V such that*

(i) the level sets $\{V \leq r\}$ are finite for all $r \in \mathbb{N}$;
(ii) $PV(x) \leq V(x)$ for all $x \in F^c$.

Then P is recurrent.

Proof. By assumption, the function V is superharmonic outside F. Thus Theorem 4.1.2 (i) shows that the sequence $\{V(X_{n\wedge\tau_F}), n \in \mathbb{N}\}$ is a nonnegative $\mathbb{P}_x$-supermartingale for all $x \in \mathsf{X}$. Using the supermartingale convergence theorem, Proposition E.1.3, the sequence $\{V(X_{n\wedge\tau_F}), n \in \mathbb{N}\}$ converges $\mathbb{P}_x$ − a.s. to a finite random variable, and

$$\mathbb{E}_x\left[\lim_{n\to\infty} V(X_{n\wedge\tau_F})\right] \leq V(x) < \infty\,.$$

Therefore, for all $x \in \mathsf{X}$,

$$\mathbb{E}_x\left[\mathbb{1}_{\{\tau_F=\infty\}} \lim_{n\to\infty} V(X_n)\right] \leq \mathbb{E}_x\left[\lim_{n\to\infty} V(X_{n\wedge\tau_F})\right] \leq V(x) < \infty\,. \tag{7.5.1}$$

The proof proceeds by contradiction. Assume that the Markov kernel P is transient. For $r \in \mathbb{N}$, set $G = \{V \leq r\}$. Since P is transient, $\mathbb{E}_x[N_y] < \infty$ for all $x, y \in \mathsf{X}$ by Theorem 7.1.2. Hence we have

$$U(x,G) = \mathbb{E}_x[N_G] = \sum_{y\in G} \mathbb{E}_x[N_y] < \infty\,.$$

Therefore, $\mathbb{P}_x(N_G < \infty) = 1$ and $\mathbb{P}_x(\liminf_{n\to\infty} V(X_n) \geq r) = 1$ for all $x \in \mathsf{X}$. Since r is arbitrary, this yields

$$\mathbb{P}_x\left(\lim_{n\to\infty} V(X_n) = \infty\right) = 1\,. \tag{7.5.2}$$

The bound (7.5.1) implies that $\mathbb{P}_x\left(\mathbb{1}_{\{\tau_F=\infty\}} \liminf_{n\to\infty} V(X_n) < \infty\right) = 1$. This is compatible with (7.5.2) only if $\mathbb{P}_x(\tau_F = \infty) = 0$. This, in turn, implies that for all $x \in \mathsf{X}$,

$$\mathbb{P}_x(\sigma_F < \infty) = \mathbb{P}_x(\tau_F \circ \theta < \infty) = \mathbb{E}_x[\mathbb{P}_{X_1}(\tau_F < \infty)] = 1\,.$$

Applying Proposition 3.3.6, we obtain that $\mathbb{P}_x(\sigma_F^{(n)} < \infty) = 1$ for all $x \in F$ and $n \in \mathbb{N}$, so that $\mathbb{P}_x(N_F = \infty) = 1$ for all $x \in F$, and consequently,

$$\mathbb{E}_x[N_F] = \sum_{y\in F} U(x,y) = \infty\,.$$

Since F is finite, $U(x,y) = \infty$ for at least one $(x,y) \in F \times F$. By Theorem 7.1.2, this contradicts the transience of P. $\square$

We continue with a drift criterion for positive recurrence.

Theorem 7.5.3. *Let P be an irreducible Markov kernel and F an accessible finite subset of X. Then P is positive if and only if there exists a function $V : \mathsf{X} \to [0,\infty)$ satisfying*

(i) $\sup_{x\in F} PV(x) < \infty$*;*
(ii) $PV(x) \le V(x) - 1$ *for all* $x \in F^c$*.*

Proof. Let $V : \mathsf{X} \to [0,\infty)$ be a function satisfying (i) and (ii). By Corollary 4.4.8, $V(x) \ge \mathbb{E}_x[\tau_F]$ for all $x \in \mathsf{X}$. Therefore, for all $x \in F$,

$$\mathbb{E}_x[\sigma_F] = 1 + \mathbb{E}_x[\tau_F \circ \theta] = 1 + \mathbb{E}_x[\mathbb{E}_{X_1}(\tau_F)] \le 1 + \mathbb{E}_x[V(X_1)] \le 1 + PV(x) < \infty .$$

By Theorem 3.3.8, this implies that $\mathbb{P}_x(\sigma_F^{(n)} < \infty) = 1$ for all $n \in \mathbb{N}$ and $x \in \mathsf{X}$, and the sequence $\{\tilde{X}_n, n \in \mathbb{N}\}$ defined for $n \in \mathbb{N}$ by $\tilde{X}_n = X_{\sigma_F^{(n)}}$ is a Markov chain on F with kernel Q_F. The set F being finite, there exists a state $\mathsf{a} \in F$ that is recurrent for Q_F and a fortiori for P. By Proposition 6.2.4 (i), a is accessible, and therefore P is recurrent.

We now show that the Markov kernel Q_F is irreducible. Since F is accessible, there exists an accessible state $\mathsf{a} \in F$. Let $x \in F$, $x \ne \mathsf{a}$. The state a being accessible, there exist $n \ge 1$ and $x_1,\ldots,x_{n-1} \in \mathsf{X}$ such that $\mathbb{P}_x(X_1 = x_1,\ldots,X_{n-1} = x_{n-1}, X_n = \mathsf{a}) > 0$. Let $1 \le j_1 < j_2 < \cdots < j_m < n$ be the time indices j_k such that $x_{j_k} \in F$, $k \in \{1,\ldots,m\}$. Then

$$\begin{aligned} Q_F^{m+1}(x,\mathsf{a}) &\ge \mathbb{P}_x(\tilde{X}_1 = x_{j_1},\ldots,\tilde{X}_m = x_{j_m}, \tilde{X}_{m+1} = \mathsf{a}) \\ &\ge \mathbb{P}_x(X_1 = x_1,\ldots,X_{n-1} = x_{n-1}, X_n = \mathsf{a}) > 0 . \end{aligned}$$

Thus a is accessible for Q_F, which is therefore irreducible. Since F is finite, Q_F is positive by Corollary 7.2.3. Thus for every accessible state b in F, we have $\mathbb{E}_{\mathsf{b}}[\tilde{\sigma}_{\mathsf{b}}] < \infty$, where $\tilde{\sigma}_{\mathsf{b}} = \inf\{n \ge 1 : \tilde{X}_n = \mathsf{b}\}$. Applying Theorem 3.3.8 (ii) with $A = \{\mathsf{b}\}$, we obtain

$$\mathbb{E}_{\mathsf{b}}[\sigma_{\mathsf{b}}] \le \mathbb{E}_{\mathsf{b}}[\tilde{\sigma}_{\mathsf{b}}] \sup_{y\in F} \mathbb{E}_y[\sigma_F] < \infty .$$

This implies that the kernel P is positive by Theorem 7.2.1.

Conversely, assume that P is positive. Fix $\mathsf{a} \in \mathsf{X}$ and set $V_{\mathsf{a}}(x) = \mathbb{E}_x[\tau_{\mathsf{a}}]$. We first show that $V_{\mathsf{a}}(x)$ is finite for all $x \in \mathsf{X}$. Let $x \ne \mathsf{a}$. The strong Markov property combined with $\{\tau_x < \sigma_{\mathsf{a}}\} \in \mathscr{F}_{\tau_x}$ implies

$$\begin{aligned}\infty > \mathbb{E}_{\mathsf{a}}[\sigma_{\mathsf{a}}] &\geq \mathbb{E}_{\mathsf{a}}[(\tau_x + \tau_{\mathsf{a}} \circ \theta_{\tau_x})\mathbb{1}_{\{\tau_x<\sigma_{\mathsf{a}}\}}] \\ &\geq \mathbb{E}_{\mathsf{a}}[\tau_{\mathsf{a}} \circ \theta_{\tau_x}\mathbb{1}_{\{\tau_x<\sigma_{\mathsf{a}}\}}] = \mathbb{E}_x[\tau_{\mathsf{a}}]\mathbb{P}_{\mathsf{a}}(\tau_x < \sigma_{\mathsf{a}}) .\end{aligned}$$

It suffices to prove that $V_{\mathsf{a}}(x)$ is finite, provided that $\mathbb{P}_{\mathsf{a}}(\tau_x < \sigma_{\mathsf{a}}) > 0$. Since a is recurrent and $x \neq \mathsf{a}$, we have $\mathbb{P}_{\mathsf{a}}(\cup_{k=0}^{\infty}\{\sigma_{\mathsf{a}}^{(k)} < \tau_x < \sigma_{\mathsf{a}}^{(k+1)}\}) = 1$. Thus there exists $k \geq 0$ such that

$$0 < \mathbb{P}_{\mathsf{a}}(\sigma_{\mathsf{a}}^{(k)} < \tau_x < \sigma_{\mathsf{a}}^{(k+1)}) = \mathbb{P}_{\mathsf{a}}(\sigma_{\mathsf{a}}^{(k)} < \tau_x, \tau_x \circ \theta_{\sigma_{\mathsf{a}}^{(k)}} < \sigma_{\mathsf{a}} \circ \theta_{\sigma_{\mathsf{a}}^{(k)}}) .$$

Since $X_{\sigma_{\mathsf{a}}^{(k)}} = \mathsf{a}$, conditioning on $\mathscr{F}_{\sigma_{\mathsf{a}}^{(k)}}$ and applying the strong Markov property yields

$$\mathbb{P}_{\mathsf{a}}(\sigma_{\mathsf{a}}^{(k)} < \tau_{\mathsf{a}})\mathbb{P}_{\mathsf{a}}(\tau_x < \sigma_{\mathsf{a}}) > 0 .$$

This proves that $\mathbb{P}_{\mathsf{a}}(\tau_x < \sigma_{\mathsf{a}}) > 0$. Therefore, V_{a} takes values in $[0,\infty)$. By Corollary 4.4.8, V_{a} is a solution to $V_{\mathsf{a}}(\mathsf{a}) = 0$ and $PV_{\mathsf{a}}(x) = V_{\mathsf{a}}(x) - 1$ for $x \neq \mathsf{a}$, and thus V_{a} satisfies (i) and (ii). □

7.6 Convergence to the Invariant Probability

Let P be a Markov kernel that admits a unique invariant probability π. We investigate in this section the convergence of the iterates $\{\xi P^n, n \in \mathbb{N}\}$ started from a given initial distribution ξ. There are many different ways to assess the convergence of ξP^n to π. We will consider here convergence in the total variation distance. The main properties of the total variation distance are presented in Appendix D for general state spaces. We briefly recall here the definition and main and properties in the discrete setting.

Definition 7.6.1 *Let ξ and ξ' be two probabilities on a finite or countable set X. The total variation distance between ξ and ξ' is defined by*

$$\mathrm{d_{TV}}(\xi, \xi') = \frac{1}{2}\sum_{x \in \mathsf{X}} |\xi(x) - \xi'(x)| . \tag{7.6.1}$$

The total variation distance is bounded by 1: for all probability measures ξ, ξ', $\mathrm{d_{TV}}(\xi,\xi') \leq 1$ and $\mathrm{d_{TV}}(\xi,\xi') = 1$ if and only if ξ and ξ' are supported by disjoint subsets of X. The total variation distance can be characterized as the operator norm of the bounded signed measure $\xi - \xi'$ acting on the space of functions on X equipped with the oscillation seminorm, i.e.,

$$\mathrm{d_{TV}}(\xi, \xi') = \sup\{\xi(f) - \xi'(f) : \mathrm{osc}\,(f) \leq 1\} , \tag{7.6.2}$$

where

$$\operatorname{osc}(f) = \sup_{x,y \in \mathsf{X}} |f(x) - f(y)| . \tag{7.6.3}$$

To prove the convergence, we will use a coupling method. We will use the coupling method on many occasions in the book: the discrete case is particularly simple and provides a good introduction to this technique. Define the Markov kernel $\bar{P}$ on X^2 by

$$\bar{P}((x,x'),(y,y')) = P(x,y)P(x',y') , \qquad x,y,x',y' \in \mathsf{X} . \tag{7.6.4}$$

For any initial distribution $\bar{\xi}$ on $\mathsf{X} \times \mathsf{X}$, let $\bar{\mathbb{P}}_{\bar{\xi}}$ be the probability measure on the canonical space $(\mathsf{X} \times \mathsf{X})^{\mathbb{N}}$ such that the canonical process $\{(X_n, X_n'), n \in \mathbb{N}\}$ is a Markov chain with kernel $\bar{P}$. By definition of the kernel $\bar{P}$, for $x, x' \in \mathsf{X}$, the two components are under $\bar{\mathbb{P}}_{x,x'}$ independent Markov chains with kernel P started from x and x', respectively.

Lemma 7.6.2 *If P is irreducible and aperiodic, then $\bar{P}$ is irreducible and aperiodic. If P is strongly irreducible, then $\bar{P}$ is strongly irreducible. If P is transient, then $\bar{P}$ is transient.*

Proof. Let x, y be accessible states for P. By Proposition 6.3.3, since P is aperiodic, there exists an integer n_0 such that $P^n(x,x) > 0$ and $P^n(y,y) > 0$ for all $n \geq n_0$. Moreover, since x and y are accessible, for all $x', y' \in \mathsf{X}$ there exist m and p such that $P^m(x',x) > 0$ and $P^p(y',y) > 0$. Thus, for $n \geq n_1 = n_0 + m \vee p$,

$$\begin{aligned} P^n(x',x) &\geq P^m(x',x)P^{n-m}(x,x) > 0 , \\ P^n(y',y) &\geq P^p(y',y)P^{n-p}(y,y) > 0 . \end{aligned}$$

This yields $\bar{P}^n((x',y'),(x,y)) = P^n(x',x)P^n(y',y) > 0$ for all $n \geq n_1$. This proves that $\bar{P}$ is irreducible, since (x,y) is accessible and aperiodic by Proposition 6.3.6. Exactly the same argument shows that $\bar{P}$ is strongly irreducible if P is strongly irreducible.

Assume that P is transient. For all $(x,y) \in \mathsf{X}^2$, $U(x,y) < \infty$, which implies that $\lim_{n\to\infty} P^n(x,y) = 0$ and

$$\sum_{n=0}^{\infty} \{P^n(x,y)\}^2 = \sum_{n=0}^{\infty} \bar{P}^n((x,y),(x,y)) < \infty ,$$

which proves that $\bar{P}$ is transient. □

From now on, we fix one state $\mathsf{a} \in \mathsf{X}$ and let T be the hitting time of (a,a) by $\{(X_n, X_n'), n \in \mathbb{N}\}$, i.e.,

$$T = \inf\left\{n \geq 0 : (X_n, X_n') = (\mathsf{a},\mathsf{a})\right\} . \tag{7.6.5}$$

The usefulness of the coupling method derives from the following result.

Proposition 7.6.3 (Coupling inequality) *Let ξ and ξ' be two probability measures on* X. *Then*

$$\mathrm{d_{TV}}(\xi P^n, \xi' P^n) \le \bar{\mathbb{P}}_{\xi\otimes\xi'}(T > n)\,. \tag{7.6.6}$$

Proof. For $f \in \mathbb{F}_b(\mathsf{X})$, write

$$\begin{aligned}\xi P^n(f) - \xi' P^n(f) &= \bar{\mathbb{E}}_{\xi\otimes\xi'}[f(X_n) - f(X'_n)] \\ &= \bar{\mathbb{E}}_{\xi\otimes\xi'}[\{f(X_n) - f(X'_n)\}\mathbb{1}_{\{T>n\}}] + \sum_{k=0}^{n-1} \bar{\mathbb{E}}_{\xi\otimes\xi'}[\{f(X_n) - f(X'_n)\}\mathbb{1}_{\{T=k\}}]\,.\end{aligned}$$

For $k \le n-1$, the Markov property and the fact that the chains $\{X_n\}$ and $\{X'_n\}$ have the same distribution if they start from the same initial value yield

$$\bar{\mathbb{E}}_{\xi\otimes\xi'}[\{f(X_n) - f(X'_n)\}\mathbb{1}_{\{T=k\}}] = \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbb{1}_{\{T=k\}}\bar{\mathbb{E}}_{(\mathsf{a},\mathsf{a})}[f(X_{n-k}) - f(X'_{n-k})]\right] = 0\,.$$

Altogether, we obtain

$$|\xi P^n(f) - \xi' P^n(f)| \le \mathrm{osc}\,(f)\,\bar{\mathbb{P}}_{\xi\otimes\xi'}(T > n)\,.$$

Applying the characterization (7.6.2) yields (7.6.6). □

The coupling inequality provides an easy way to establish convergence to the stationary probability. Recall that for a given set $\mathsf{R} \in \mathscr{X}$, the set R_+ is defined in (3.5.1).

Theorem 7.6.4. *Let P be a Markov kernel on a discrete state space* X. *Assume that P is irreducible, aperiodic, and positive. Denote by π its unique invariant probability and set $\mathsf{R} = \{x \in \mathsf{X} : \pi(x) > 0\}$. Then for every probability measure ξ on X such that $\xi(\mathsf{R}_+^c) = 0$,*

$$\lim_{n\to\infty} \mathrm{d_{TV}}(\xi P^n, \pi) = 0\,. \tag{7.6.7}$$

If P is strongly irreducible, then $\mathsf{R}_+ = \mathsf{X}$, and (7.6.7) holds for every probability measure ξ.

Proof. Assume first that the Markov kernel P is strongly irreducible and aperiodic. Hence by Lemma 7.6.2, $\bar{P}$ is strongly irreducible. Since $(\pi\otimes\pi)\bar{P} = \pi P \otimes \pi P = \pi\otimes\pi$, the probability measure $\pi\otimes\pi$ is invariant for $\bar{P}$. Since $\bar{P}$ is strongly irreducible and $\pi\otimes\pi$ is an invariant probability measure, Theorem 7.1.4 implies that $\bar{P}$ is positive. Denote by T the hitting time of the set (a,a). Since P is strongly irreducible and recurrent, Theorem 7.1.2 shows that for all $x, x' \in \mathsf{X}$, $\bar{\mathbb{P}}_{x,x'}(T < \infty) = 1$.

Hence $\bar{\mathbb{P}}_{\xi\otimes\xi'}(T<\infty)=1$, and the limit (7.6.7) follows from (7.6.6) applied with $\xi'=\pi$.

We now consider the general case. By Theorem 7.3.3, R is absorbing, the trace of P to R is strongly irreducible, and π is the unique invariant distribution of $P|_{\mathsf{R}}$. Let $\mathsf{a}\in\mathsf{R}$ be a recurrent state. It follows from the first part of the proof that for every $\mathsf{a}\in\mathsf{R}$, $\lim_{n\to\infty}\|\delta_{\mathsf{a}}P^n-\pi\|_{\mathrm{TV}}=0$. It is also easily seen that $\mathsf{R}_+=\{x\in\mathsf{X}:\mathbb{P}_x(\sigma_{\mathsf{a}}<\infty)=1\}$. Let ξ be a probability measure such that $\xi(\mathsf{R}_+^c)=0$. Since $\mathbb{P}_x(\sigma_{\mathsf{a}}<\infty)=1$ for all $x\in\mathsf{R}$, we get $\mathbb{P}_\xi(\sigma_{\mathsf{a}}<\infty)=1$. For every $\varepsilon>0$, we may choose n_0 large enough that $\mathbb{P}_\xi(\sigma_{\mathsf{a}}\geq n_0)<\varepsilon$ and $\|\delta_{\mathsf{a}}P^n-\pi\|_{\mathrm{TV}}\leq\varepsilon$ for all $n\geq n_0$. For every function f such that $|f|_\infty\leq 1$, we get

$$\begin{aligned}\mathbb{E}_\xi[f(X_n)]-\pi(f)=\sum_{k=1}^{n}\mathbb{P}_\xi(\sigma_{\mathsf{a}}=p)\{\mathbb{E}_{\mathsf{a}}[f(X_{n-p})]-\pi(f)\}\\ +\mathbb{E}_\xi[\{f(X_n)-\pi(f)\}\mathbb{1}_{\{\sigma_{\mathsf{a}}>n\}}]\,.\end{aligned}\tag{7.6.8}$$

Since $\mathbb{P}_\xi(\sigma_{\mathsf{a}}<\infty)=1$ and

$$|\mathbb{E}_\xi[\{f(X_n)-\pi(f)\}\mathbb{1}_{\{\sigma_{\mathsf{a}}>n\}}]|\leq 2|f|_\infty\mathbb{P}_\xi(\sigma_{\mathsf{a}}>n)\,,$$

on the other hand, for all $n\geq 2n_0$, we obtain

$$\left|\sum_{k=1}^{n}\mathbb{P}_\xi(\sigma_{\mathsf{a}}=p)\{\mathbb{E}_{\mathsf{a}}[f(X_{n-p})]-\pi(f)\}\right|\leq\varepsilon+2|f|_\infty\mathbb{P}_\xi(\sigma_{\mathsf{a}}\geq n_0)\,.$$

Since ε is arbitrary, this concludes the proof. □

It is also interesting to study the forgetting of the initial distribution $\xi\in\mathbb{M}_1(\mathscr{X})$ when P is null recurrent. In that case, aperiodicity is not needed, since the iterates of the kernel converge to zero.

Theorem 7.6.5. *Let P be an irreducible null-recurrent Markov kernel on a discrete state space* X. *Then for all $\xi\in\mathbb{M}_1(\mathscr{X})$ and $y\in\mathsf{X}$,*

$$\lim_{n\to\infty}\xi P^n(y)=0\,.\tag{7.6.9}$$

Proof. By Theorem 7.3.3, $\mathsf{X}=\mathsf{T}\cup\mathsf{R}$: all the states in T are transient, and the trace of P on R is strongly irreducible and recurrent. If $y\in\mathsf{T}$, then $U(x,y)<\infty$ for all $x\in\mathsf{X}$, showing that $\lim_{n\to\infty}P^n(x,y)=0$. Therefore, by Lebesgue's dominated convergence theorem, (7.6.9) holds for all $y\in\mathsf{T}$ and $\xi\in\mathbb{M}_1(\mathscr{X})$.

Assume now that P is strongly irreducible and recurrent. Assume first that P is aperiodic. Then by Lemma 7.6.2, $\bar{P}$ is irreducible and aperiodic. By Theorem 7.1.2, $\bar{P}$ is thus either transient or recurrent.

(i) If $\bar{P}$ is transient, then for all $x,y \in \mathsf{X}$,

$$\infty > \bar{U}((x,x),(y,y)) = \sum_{n=0}^{\infty} \bar{P}^n((x,x),(y,y)) = \sum_{n=0}^{\infty} [P^n(x,y)]^2 .$$

Therefore, $\lim_{n\to\infty} P^n(x,y) = 0$ and (7.6.9) holds by Lebesgue's dominated convergence theorem.

(ii) Assume now that $\bar{P}$ is recurrent. By Lemma 7.6.2, $\bar{P}$ is strongly irreducible. Let $\mathsf{a} \in \mathsf{X}$ and let T be the hitting time of (a,a). Since $\bar{P}$ is strongly irreducible and recurrent, Theorem 7.1.2 shows that $\bar{\mathbb{P}}_{x,x'}(T < \infty) = 1$ for all $x,x' \in \mathsf{X}$. Thus $\bar{\mathbb{P}}_{\xi\otimes\xi'}(T < \infty) = 1$ for all probability measures ξ, ξ', and by Proposition 7.6.3, this yields $\lim_{n\to\infty} \mathrm{d}_{\mathrm{TV}}(\xi P^n, \xi' P^n) = 0$. This, in turn, implies, for all $y \in \mathsf{X}$,

$$\lim_{n\to\infty} |\xi P^n(y) - \xi' P^n(y)| = 0 . \tag{7.6.10}$$

We must now prove that (7.6.10) implies (7.6.9). Let μ be an invariant measure for P (such measures are unique up to a scaling factor). For every finite set A, define the probability measure μ_A on the set A by

$$\mu_A(x) = \begin{cases} \frac{\mu(x)}{\mu(A)} & \text{if } x \in A , \\ 0 & \text{otherwise} . \end{cases}$$

Fix $y \in \mathsf{X}$. Then for $n \in \mathbb{N}^*$, we get

$$\mu_A P^n(y) \le \frac{\mu P^n(y)}{\mu(A)} = \frac{\mu(y)}{\mu(A)} .$$

Since $\mu(\mathsf{X}) = \infty$, the right-hand side can be made less than some arbitrarily small ε by choosing a sufficiently large A. Applying (7.6.10) with a probability measure ξ and μ_A yields

$$\limsup_{n\to\infty} \xi P^n(y) \le \varepsilon + \lim_{n\to\infty} |\xi P^n(y) - \mu_A P^n(y)| = \varepsilon .$$

This proves (7.6.9).

Assume now that the kernel P has period $d \ge 2$. Let $D_0, \ldots, D_{d-1}$ be a partition of X as in Theorem 7.4.1. Then the restriction of P^d to each class D_i is strongly irreducible, aperiodic, and null recurrent. Thus the first part of the proof shows that if $x,y \in D_i$, then $\lim_{n\to\infty} P^{nd}(x,y) = 0$. If $x \in D_k$ and $y \in D_j$ for some $j \ne k$, then there exists $m < d$ such that $P^m(x, D_j) = 1$ and $P^{kd+r}(x,y) = 0$ for $r \ne m$. This implies that $P^n(x,y) = 0$ if $n \ne kd + m$ and thus, by Lebesgue's dominated convergence theorem,

$$\lim_{n\to\infty} P^n(x,y) = \lim_{k\to\infty} P^{kd+m}(x,y) = \sum_{z\in D_j} P^m(x,z) \lim_{k\to\infty} P^{kd}(z,y) = 0 .$$

This proves that (7.6.9) holds for $\xi = \delta_x$ for all $x \in \mathsf{X}$, hence for every initial distribution by applying again Lebesgue's dominated convergence theorem. □

7.7 Exercises

7.1. Let P be an irreducible Markov kernel on a discrete state space X. The set X_P^+ of accessible states is absorbing, and inaccessible states are transient.

7.2. Show that Theorem 7.1.4 does not hold if we assume only that π is an invariant measure (instead of an invariant probability measure).

7.3. Identify the communication classes of the following transition matrix:

$$\begin{pmatrix} 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 1/2 & 0 & 0 & 0 & 1/2 \end{pmatrix}.$$

Find the recurrent classes.

7.4 (Wright–Fisher model). The Wright–Fisher model is an ideal genetics model used to investigate the fluctuation of gene frequency in a population of constant size under the influence of mutation and selection. The model describes a simple haploid random reproduction disregarding selective forces and mutation pressure. The size of the population is set to N individuals of two types 1 and 2. Let X_n be the number of individuals of type 1 at time n. Then $\{X_n, n \in \mathbb{N}\}$ is a Markov chain with state space $\mathsf{X} = \{0, 1, \ldots, N\}$ and transition matrix

$$P(j,k) = \binom{N}{k} \left(\frac{j}{N}\right)^k \left(1 - \frac{j}{N}\right)^{N-k},$$

with the usual convention $0^0 = 1$. In words, given that the number of type-1 individuals at the current generation is j, the number of type-1 individuals at the next generation follows a binomial distribution with success probability j/N. Looking backward, this can be interpreted as having each individual in the next generation "choose their parents at random" from the current population. A basic phenomenon of the Wright–Fisher model without mutation is fixation, which is the elimination of all but one type of individual after an almost surely finite random time. This phenomenon is shown in this exercise.

1. Show that the Markov kernel P is not irreducible and the states 0 and N are absorbing.
2. Show that for all $x \in \{1, \ldots, N-1\}$, $\mathbb{P}_x(\sigma_x < \infty) < 1$ and $\mathbb{E}_x[N_x] < \infty$.
3. Show that $\{X_n, n \in \mathbb{N}\}$ is a martingale that converges to X_∞ $\mathbb{P}_x$ − a.s. and in $\mathrm{L}^1(\mathbb{P}_x)$ for all $x \in \{0, \ldots, N\}$.
4. Show that $\mathbb{P}_x(X_\infty = N) = x/N$ and $\mathbb{P}_x(X_\infty = 0) = 1 - x/N$ for $x \in \{1, \ldots, N\}$.

7.5. Let P be the Markov kernel on $\mathbb{N}$ defined by $P(x, x+1) = p > 0$, $P(x, x-1) = q > 0$, $P(x,x) = r \geq 0$, for $x \geq 1$, $P(0,0) = 1 - p$, $P(0,1) = p$.

1. Show that the state $\{0\}$ is accessible.
2. Show that all the states communicate.
3. Show that $f(x) = \mathbb{P}_x(\tau_0 < \infty)$ is the smallest nonnegative function on $\mathbb{N}$ satisfying $f(0) = 1$ and for $x \geq 1$,

$$f(x) = qf(x-1) + rf(x) + pf(x+1), \quad x \geq 1. \tag{7.7.1}$$

4. Show that the solutions for (7.7.1) are given by $f(x) = c_1 + c_2(q/p)^x$ if $p \neq q$ and $f(x) = c_1 + c_2 x$ if $p = q$ for constants c_1 and c_2 to be determined.
5. Assume that $p < q$. Show that the Markov kernel P is recurrent.
6. Assume that $p > q$. Show that $\mathbb{P}_x(\tau_0 < \infty) < 1$ for all $x \geq 1$ and that the Markov kernel P is transient.
7. Assume that $p = q$. Show that the Markov kernel P is recurrent.

7.6 (Simple symmetric random walk on $\mathbb{Z}^d$). Consider the symmetric simple random walk on $\mathbb{Z}^d$ with kernel $P(x,y) = 1/2d$ if $|x-y| = 1$ and $P(x,y) = 0$ otherwise. Set $V(x) = |x|^{2\alpha}$, where $\alpha \in (-\infty, 1]$.

1. Show that all the states communicate.
2. Show that there exists a constant $C(\alpha,d)$ such that for all $|x| \geq 2$,

$$PV(x) - V(x) = 2\alpha\{2\alpha - 2 + d + r(x)\}\,|x|^{2\alpha-2},$$

with $|r(x)| \leq C(\alpha,d)\,|x|^{-1}$.
3. Assume that $d = 1$. Using the drift condition above, show that P is recurrent.
4. Assume that $d \geq 3$. Using the drift condition above, show that P is transient.

It remains to consider $d = 2$, which is more subtle. Consider $W(x) = \{\log(|x|^2)\}^\alpha$ with $0 < \alpha < 1$.

5. Compute $PW(x) - W(x)$.
6. Show that the symmetric simple random walk is transient.

7.7 (INAR process). An INAR (INteger AutoRegressive) process is a Galton–Walton process with immigration defined by the recurrence

$$X_0 = 1, \qquad X_{n+1} = \sum_{j=1}^{X_n} \xi_j^{(n+1)} + Y_{n+1},$$

where $\{\xi_j^{(n)},\, j,n \in \mathbb{N}^*\}$ are i.i.d. Bernoulli random variables and $\{Y_n,\, n \in \mathbb{N}^*\}$ is a sequence of i.i.d. integer-valued random variables independent of $\{\xi_j^{(n)}\}$ and X_0. We denote by ν the distribution of Y_1. We assume that $\nu(0) > 0$ and $m = \sum_{k=0}^\infty k\nu(k) < \infty$. We set $\alpha = \mathbb{P}(\xi_1^{(1)} = 1) \in (0,1)$.

1. Show that the state $\{0\}$ is accessible and that P is irreducible.
2. Show that for all $k \geq 1$,

$$\mathbb{E}_x[X_k] = \alpha^k x + m \sum_{j=0}^{k-1} \alpha^j ,$$

where $m = \mathbb{E}[Y_1]$.
3. Assume that $\mathrm{Var}(Y_1) = \sigma^2$. Show that for all $k \geq 1$,

$$\mathrm{Var}_x(X_k) = (1-\alpha) \sum_{j=1}^{k} \alpha^{2j-1} \mathbb{E}_x[X_{k-j}] + \sigma^2 \sum_{j=1}^{k} \alpha^{2(j-1)}.$$

4. Show that P is positive.
5. For $|s| \leq 1$ and $x \in \mathbb{N}$, denote by $\phi_{n,x}(s) = \mathbb{E}_x[s^{X_n}]$ the moment-generating function of X_n. Show that for all $n \geq 1$,

$$\phi_{n,x}(s) = \phi_{n-1,x}(s)\psi(s) ,$$

where ψ is the moment-generating function of Y_1.
6. Show that for all $n \geq 1$,

$$\phi_{n,x}(s) = (1-\alpha^n + \alpha^n s)^x \prod_{k=0}^{n-1} \psi(1-\alpha^k + \alpha^k s).$$

7. Show that for all $x \in \mathbb{N}$, $\lim_{n\to\infty} \phi_{n,x}(s) = \phi(s)$ (which does not depend on x) and that

$$\phi(s) = \psi(s)\phi(1-\alpha+\alpha s) .$$

7.8 (Discrete time queueing system). Clients arrive for service and enter a queue. During each time interval, a single customer is served, provided that at least one customer is present in the queue. If no customer is awaiting service, then during this period no service is performed. During a service period, new clients may arrive. We assume that the number of arrivals during the nth service period is a sequence of i.i.d. integer-valued random variables $\{Z_n, n \in \mathbb{N}\}$, independent of the initial state X_0 and whose distribution is given by

$$\mathbb{P}(Z_n = k) = a_k \geq 0 , \quad k \in \mathbb{N} , \quad \sum_{k=0}^{\infty} a_k = 1 .$$

The state of the queue at the start of each period is defined to be the number of clients waiting for service, which is given by

$$X_{n+1} = (X_n - 1)^+ + Z_{n+1} .$$

1. Show that $\{X_n, n \in \mathbb{N}\}$ is a Markov chain. Determine its kernel.
2. Describe the behavior of the chain when $a_0 = 1$ and $a_0 + a_1 = 1$.

In the sequel, it is assumed that the arrival distribution is nondegenerate, in the sense that $0 < a_0 < 1$ and $a_0 + a_1 < 1$.

3. Show that all the states communicate.

Denote by m the mean number of clients entering into service $m = \sum_{k=0}^{\infty} k a_k$. Assume first that $m > 1$.

4. Fix $b > 0$ and set $W(x) = b^x$, $x \in \mathbb{N}$. Show that $PW(x) = \varphi(b) b^{x-1}$, where φ is the moment-generating function of the distribution $\{a_k, k \in \mathbb{N}\}$.
5. Show that there exists a unique $b_0 \in (0,1)$ such that $\phi(b_0) = b_0$.
6. Show that P is transient.

Assume now that $m = \sum_{k=0}^{\infty} k a_k \leq 1$.

7. Set $V(x) = x$. Show that for every $x > 0$,

$$PV(x) \leq V(x) - (1-m)\,. \tag{7.7.2}$$

8. Show that P is positive.

7.9. Let P be the transition matrix on $\mathsf{X} = \{0,1\}$,

$$P = \begin{pmatrix} 1-\ \alpha & \alpha \\ \beta & 1-\ \beta \end{pmatrix}$$

with $0 < \alpha \leq 1$ and $0 < \beta \leq 1$. Assume that $\min(\alpha,\ \beta) < 1$.

1. Show that
$$P^n = \frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta}\begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}$$
and determine $\lim_{n\to\infty} P^n$.
2. Compute the stationary distribution π of P.
3. Compute $\mathrm{Cov}_\pi(X_n, X_{n+p})$.
4. Set $S_n = X_1 + \ldots + X_n$. Compute $\mathbb{E}_\pi[S_n]$ and $\mathrm{Var}_\pi(S_n)$. Give a bound for
$$\mathbb{P}_\pi\left(|n^{-1}S_n - \alpha/(\alpha+\beta)| \geq \delta\right).$$

7.10. Let $\{p(x) : x \in \mathbb{N}\}$ be a probability on $\mathsf{X} = \mathbb{N}$. Assume that $p(x) > 0$ for every $x > 0$. Define $G(x) = \sum_{y>x} p(y)$. We consider the Markov kernel P given by $P(0,\ y) = p(y)$ for all $y \geq 0$ and for all $x \geq 1$,

$$P(x,\ y) = \begin{cases} 1/x & \text{if } y < x, \\ 0 & \text{otherwise.} \end{cases}$$

1. Show that the Markov kernel P is strongly irreducible and recurrent.
2. Let μ be an invariant measure. Show that $\sum_{y>0} \mu(y) y^{-1} < +\infty$.
3. Set $\phi(x) = \sum_{y=x+1}^{\infty} y^{-1}\mu(y)$. Express φ and then μ as functions of G and $\mu(0)$.

Assume $p(0) = 0$ and $p(x) = 1/x - 1/(x+1)$ for $x \geq 1$.

4. Is this chain positive?
5. Determine the limit $\lim_{n\to\infty}\ n^{-1}\sum_{k=0}^{n-1} \mathbb{1}_{\{0\}}(X_k)$.

7.11. Let P be an irreducible Markov kernel on a discrete state space X. Assume that there exist a bounded function $V : \mathsf{X} \to \mathbb{R}_+$ and $r \geq 0$ such that

(i) the level sets $\{V \leq r\}$ and $\{V > r\}$ are accessible;
(ii) the level set $\{V \leq r\}$ is finite;
(iii) $PV(x) \geq V(x)$ for all $x \in \{V > r\}$.

Show that P is transient.

7.12. Let P be an irreducible Markov kernel. Assume that there exist a nonempty finite subset $F \subset \mathsf{X}$, a constant $b < \infty$, and functions $V : \mathsf{X} \to [0,\infty)$ and $f : \mathsf{X} \to [1,\infty)$ such that

$$PV(x) \leq V(x) - f(x) + b\mathbb{1}_F(x) . \tag{7.7.3}$$

Show that P is positive and that its unique invariant probability measure π satisfies $\pi(f) < \infty$.

7.13. Let us consider $d > 1$ balls numbered from 1 to d and two urns A and B. At the nth round of the game, a number i is sampled uniformly in $\{1,\ldots,d\}$, and one of the urns is chosen with probability $1/2$, independently from the past. The ball numbered i is placed in the selected urn. Denote by X_n the number of balls in urn A after n successive rounds of the game.

1. Determine the Markov kernel P associated with this process. Show that this Markov kernel is strongly irreducible and positive. Is it aperiodic?
2. Show that there exist two real constants a and b such that for every $x \in \mathsf{X} = \{1,\ldots,d\}$, $\sum_{y\in\mathsf{X}} yP(x,y) = ax+b$. Compute $\mathbb{E}_x[X_n]$ and $\lim_{n\to\infty} \mathbb{E}_x[X_n]$.
3. Assume that X_0 has a binomial distribution with parameters d and parameter of success $1/2$. What is the law of X_1?
4. Determine the invariant probability of this chain. Compute $\mathbb{E}_d[\sigma_{\{d\}}]$ and for all $x,y \in \mathsf{X}$, $\lim_{n\to\infty} P^n(x,\ y)$.

7.14. Let P be a Markov kernel on a countable set X. Assume that for every $x \in \mathsf{X}$, $P(x,x) < 1$. Define $\tau = \inf\{n \geq 1 \,:\, X_n \neq X_0\}$.

1. Compute $\mathbb{P}_x(\tau = n)$ for all $x \in \mathsf{X}$ and $n \in \mathbb{N}$. Show that $\mathbb{P}_x(\tau < \infty) = 1$.
2. Determine $\mathbb{P}_x(X_\tau = y)$ for $x,y \in \mathsf{X}$.

Define recursively the stopping times $\tau_0 = 0$ and for $n \geq 0$, $\tau_{n+1} = \tau_n + \tau \circ \theta_{\tau_n}$.

3. Show that for every $x \in \mathsf{X}$ and $n \in \mathbb{N}$, $\mathbb{P}_x(\tau_n < \infty) = 1$.
4. Show that $Y_n = X_{\tau_n}$ is a Markov chain. Determine the associated Markov kernel Q.
5. Assume that P is strongly irreducible and positive with invariant probability π. Show that Q is strongly irreducible and positive with invariant measure $\tilde{\mu}(y) = \{1 - P(y,\ y)\}\mu(y)$, $y \in \mathsf{X}$.

7.8 Bibliographical Notes

All of the results we have presented in this chapter can be found in the classic books on Markov chain theory with discrete state space, which include Chung (1967), Kemeny et al. (1976), Seneta (1981), and Taylor and Karlin (1998). The original editions of these books are from the 1960s, but the references we have given correspond to their latest editions. Since the study of Markov chains with discrete state spaces is unavoidable in all applied probability formations, many books on this subject continue to be published. The books by Norris (1998), Brémaud (1999), Privault (2013), Sericola (2013), Graham (2014) present a modern account of the theory together with many examples from a variety of fields.

Drift conditions for recurrence and transience were introduced by Foster (1952b) and Foster (1953). This criterion was later studied by Holmes (1967), Pakes (1969), and Tweedie (1975) (these early works are reviewed in Sennot et al. (1983)). Mertens et al. (1978) have shown that the drift conditions introduced in Theorems 7.5.1 and 7.5.2 are also sufficient.

The convergence of positive and aperiodic Markov kernels in total variation to their stationary distributions (Theorem 7.6.4) was first established in Kolmogorov (1931) (and was later refined by Feller (1971) and Chung (1967) using analytic proofs). The coupling proof approach presented here was introduced by Doeblin (1938).

Chapter 8
Convergence of Atomic Markov Chains

The main object of this chapter is to prove the convergence of the iterates of a positive recurrent atomic Markov kernel to its invariant probability measure. This will be done by two different methods: application of renewal theory and coupling techniques.

In Section 8.1 we will provide a concise introduction to the theory of discrete-time renewal processes. Renewal theory can be applied to the study of a stochastic process that exhibits a certain recurrent pattern and starts anew with a fixed distribution after each occurrence of this pattern. This gives rise to a renewal process that models the recurrence property of this pattern. For a discrete-time Markov chain, the typical pattern is the visit to a state, and the times between each visit are i.i.d. by the strong Markov property. The main results of renewal theory that we will present are Blackwell's and Kendall's theorems. Blackwell's theorem, Theorem 8.1.7, states that the probability that an event occurs at time n converges to the inverse mean waiting time. Kendall's theorem, Theorem 8.1.9, provides a geometric rate of convergence to Blackwell's theorem under a geometric moment condition for the waiting time distribution.

We will apply these results in Section 8.2 to prove the convergence in total variation of the iterates of a positive and aperiodic Markov kernel to its invariant distribution and establish conditions under which the rate of convergence is geometric.

8.1 Discrete-Time Renewal Theory

Definition 8.1.1 (Renewal process) *Let $\{Y_n,\, n \in \mathbb{N}^*\}$ be a sequence of i.i.d. positive integer-valued random variables with distribution $b = \{b(n),\, n \in \mathbb{N}^*\}$ and let Y_0 be a nonnegative integer-valued random variable with distribution $a = \{a(n),\, n \in \mathbb{N}\}$, independent of the sequence $\{Y_n,\, n \in \mathbb{N}^*\}$.*

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_8

- *The process* $\{S_n,\ n \in \mathbb{N}\}$ *defined by*

$$S_n = \sum_{i=0}^{n} Y_i \tag{8.1.1}$$

is called a renewal process. The random times S_n*,* $n \geq 0$*, are called the renewals or the epochs of the renewal process. The common distribution* b *of the random variables* Y_n*,* $n \geq 1$*, is called the waiting time distribution.*
- *The first renewal time* Y_0 *is called the delay of the process, and its distribution* a *is called the delay distribution. A renewal process is called pure or zero-delayed if* $Y_0 \equiv 0$ *(i.e., if* a *is concentrated at* 0*) and delayed otherwise.*
- *The renewal process and its waiting time distribution* b *are said to be aperiodic if g.c.d.* $\{n > 0 \,:\, b(n) > 0\} = 1$*.*

The sequence $\{S_n,\ n \in \mathbb{N}\}$ is a random walk with positive jumps, and it is thus a Markov chain on $\mathbb{N}$ with initial distribution a (the delay distribution) and transition kernel P given by

$$P(i,j) = \begin{cases} b(j-i) & \text{if } j > i\,, \\ 0 & \text{otherwise.} \end{cases} \tag{8.1.2}$$

As usual, we consider the canonical realization of the renewal process, which means that the canonical space $(\mathbb{N}^{\mathbb{N}}, \mathscr{P}(\mathbb{N})^{\otimes \mathbb{N}})$ is endowed with the probability measure $\mathbb{P}_a$, which makes the coordinate process a renewal process with waiting time distribution b and delay distribution a; $\mathbb{P}_i$ is shorthand for $\mathbb{P}_{\delta_i}$, and $\mathbb{E}_i$ denote the corresponding expectations. Other Markov chains associated with the renewal process will be defined later, so we stress here this notation.

It is often convenient to indicate the epochs by a random sequence $\{V_n,\ n \in \mathbb{N}\}$ such that $V_n = 1$ if n is a renewal time and $V_n = 0$ otherwise i.e.,

$$V_n = \sum_{m=0}^{\infty} \mathbb{1}\{S_m = n\}\ .$$

The delayed renewal sequence $\{v_a(k),\ k \in \mathbb{N}\}$ associated with the delay distribution a is defined by

$$v_a(k) = \mathbb{P}_a(V_k = 1) = \sum_{m=0}^{\infty} \mathbb{P}_a(S_m = k)\ , \qquad k \geq 0\ . \tag{8.1.3}$$

For $i \in \mathbb{N}$, we write v_i for v_{δ_i}, and for $i = 0$, we write u for v_0. The sequence u is called a pure renewal sequence:

$$u(k) = \mathbb{P}_0(V_k = 1)\ , \qquad k \geq 0\ . \tag{8.1.4}$$

Since $Y_1, Y_2, \ldots, Y_m$ are i.i.d. positive random variables with common distribution b, the distribution of $Y_1 + \cdots + Y_m$ is b^{*m}, the m-fold convolution of b, defined recursively by

$$b^{*0} = \delta_0 \,,\; b^{*1} = b \,,\; b^{*m} = b^{*(m-1)} * b \,,\; m \geq 1 \,, \tag{8.1.5}$$

where δ_0 is identified to the sequence $\{\delta_0(n),\, n \in \mathbb{N}\}$. For the zero-delayed renewal process, $b^{*m}(k)$ is the probability that the $(m+1)$th epoch occurs at time k, i.e., $\mathbb{P}_0(S_m = k) = b^{*m}(k)$. Note that $b^{*m}(n) = 0$ if $m > n$. This yields

$$u(k) = \mathbb{P}_0(V_k = 1) = \sum_{n=0}^{\infty} \mathbb{P}_0(S_n = k) = \sum_{n=0}^{\infty} b^{*n}(k) \,. \tag{8.1.6}$$

Since $\mathbb{P}_0(S_0 = 0) = 1$, we have $V_0 = 1\;\; \mathbb{P}_0 -$a.s., i.e., $u(0) = 1$. The delayed renewal sequence v_a can be expressed in terms of the pure renewal sequence and the delay distribution. Indeed,

$$\begin{aligned} v_a(n) = \mathbb{P}_a(V_n = 1) &= \mathbb{P}_a(Y_0 = n) + \sum_{k=1}^{n-1} \mathbb{P}_a(Y_0 = k) \sum_{m=1}^{n-1} \mathbb{P}_0(Y_1 + \cdots + Y_m = n-k) \\ &= a(n) + \sum_{k=1}^{n-1} a(k) u(n-k) = \sum_{k=0}^{n} a(k) u(n-k) = a * u(n) \,. \end{aligned} \tag{8.1.7}$$

Theorem 8.1.2. *For every distribution a on $\mathbb{N}$, the delayed renewal sequence v_a is the unique positive solution to*

$$v_a = a + b * v_a \,. \tag{8.1.8}$$

In particular, the pure renewal sequence u is the unique positive solution to

$$u = \delta_0 + b * u \,. \tag{8.1.9}$$

Proof. We first prove (8.1.9). Since $u(0) = 1$, applying (8.1.6), we obtain

$$\begin{aligned} u(n) = \sum_{k=0}^{\infty} b^{*k}(n) &= \delta_0(n) + \sum_{k=1}^{\infty} b * b^{*(k-1)}(n) \\ &= \delta_0(n) + b * \sum_{k=1}^{\infty} b^{*(k-1)}(n) = \delta_0(n) + b * u(n) \,. \end{aligned}$$

Let ν be a positive sequence that satisfies (8.1.8). Iterating, we obtain, for all $n \geq 0$,

$$\nu = a * \sum_{j=0}^{n} b^{*j} + \nu * b^{*(n+1)} \,.$$

Since $\nu * b^{*(n+1)}(k) = 0$ for $k \leq n$, this yields, for every $k \in \mathbb{N}$ and $n \geq k$,

$$v(k) = a * \sum_{j=0}^{n} b^{*j}(k) = a * \sum_{j=0}^{\infty} b^{*j}(k) = a * u(k) = v_a(k) .$$

□

For $z \in \mathbb{C}$, let U and B be the generating functions of the zero-delayed renewal sequence $\{u(n),\, n \in \mathbb{N}\}$ and of the waiting time distribution $\{b(n),\, n \in \mathbb{N}^*\}$, respectively, that is,

$$U(z) = \sum_{n=0}^{\infty} u(n) z^n , \quad B(z) = \sum_{n=1}^{\infty} b(n) z^n .$$

These series are absolutely convergent on the open unit disk $\{z \in \mathbb{C} : |z| < 1\}$. The renewal equation (8.1.9) implies that these generating functions satisfy

$$U(z) = 1 + B(z)U(z) ,$$

or equivalently, since on $\{z \in \mathbb{C} : |z| < 1\}$, $|B(z)| < 1$,

$$U(z) = \frac{1}{1 - B(z)} . \tag{8.1.10}$$

Set $V_a(z) = \sum_{k=0}^{\infty} v_a(k) z^k$ and $A(z) = \sum_{k=0}^{\infty} a(k) z^k$, the generating functions of the delayed renewal sequence and of the delay, respectively. These generating functions are absolutely convergent on $\{z \in \mathbb{C} : |z| < 1\}$. Then (8.1.7) and the expression for the generating function of the zero-delayed renewal U yield

$$V_a(z) = A(z)U(z) = \frac{A(z)}{1 - B(z)} .$$

If the mean waiting time (or mean recurrence time) is finite, then the delay distribution a may be chosen in such a way that the delayed renewal sequence v_a is constant.

Proposition 8.1.3 *Assume that the mean waiting time is finite, i.e., $m = \sum_{j=1}^{\infty} j b(j) < \infty$. The unique delay distribution a_{s} yielding a constant delayed renewal sequence, called the stationary delay distribution, is given by*

$$a_{\mathrm{s}}(k) = m^{-1} \sum_{j=k+1}^{\infty} b(j) , \qquad k \geq 0 . \tag{8.1.11}$$

In that case, for all $n \geq 0$, $v_{a_{\mathrm{s}}}(n) = m^{-1}$, which is called the renewal intensity.

Proof. Suppose that $v_a(k) = c$ for all $k \in \mathbb{N}$, where c is some positive constant that will be chosen later. Then $V_a(z) = c(1 - z)^{-1}$, $|z| < 1$, and

$$A(z) = c\{1 - B(z)\}(1-z)^{-1} . \tag{8.1.12}$$

A direct direct identification of the coefficients in (8.1.12) yields, for all $k \geq 0$,

$$a(k) = c\left(1 - \sum_{j=1}^{k} b(j)\right) = c \sum_{j=k+1}^{\infty} b(j) = c\mathbb{P}_0(Y_1 \geq k+1) .$$

Since $1 = \sum_{k=0}^{\infty} a(k)$, the constant c must satisfy $1 = c\sum_{k=1}^{\infty} \mathbb{P}_0(Y_1 \geq k) = c\mathbb{E}_0[Y_1]$, and thus $c = 1/m$. It is easy to conclude the statement of the proposition. □

To a renewal process is naturally associated the counts of the number of renewals that occurred in a given subset of $\mathbb{N}$. Formally, for $A \subset \mathbb{N}$,

$$N_A = \sum_{k=0}^{\infty} \mathbb{1}_A(S_k) .$$

Corollary 8.1.4 *If the mean waiting time m is finite and if the delay distribution is the stationary delay distribution a_{s}, then for all $A \subset \mathbb{N}$, $\mathbb{E}_{a_{\mathrm{s}}}[N_A] = m^{-1}\mathrm{card}(A)$.*

Proof. This is a straightforward consequence of Proposition 8.1.3, since for every delay distribution,

$$\mathbb{E}_a[N_A] = \sum_{n\in A} \mathbb{E}_a\left[\sum_{k=0}^{\infty} \mathbb{1}_{\{S_k=n\}}\right] = \sum_{n\in A} v_a(n) .$$

If $a = a_{\mathrm{s}}$, then the delayed renewal sequence is constant and equal to m^{-1}, and the result follows. □

8.1.1 Forward Recurrence Time Chain

Let $\{\rho_k,\ k \in \mathbb{N}\}$ be the sequence of stopping times with respect to the filtration $\{\mathscr{F}_n^S = \sigma(S_k, k \leq n),\ n \in \mathbb{N}\}$ defined by

$$\rho_k = \inf\{n \geq 0 : S_n > k\} , \tag{8.1.13}$$

the time of the first renewal epoch after time k. The forward recurrence time chain (also called the residual lifetime) is the sequence $\{A_k,\ k \in \mathbb{N}\}$, defined by

$$A_k = S_{\rho_k} - k , \qquad k \in \mathbb{N} , \tag{8.1.14}$$

the number of time steps before the next renewal epoch after k. By definition, the residual lifetime is never 0. In particular, for $k = 0$, we have

$$A_0 = \begin{cases} S_0 = Y_0 & \text{if } Y_0 > 0 \,, \\ S_1 = Y_1 & \text{if } Y_0 = 0 \,. \end{cases}$$

Thus the distribution of A_0, denoted by μ_a, is given by

$$\mu_a(k) = \mathbb{P}_a(A_0 = k) = a(k) + b(k)a(0) \,, \quad k \geq 1 \,. \tag{8.1.15}$$

Observe also that $A_k > 1$ implies $S_{\rho_k} > k+1$; hence $\rho_{k+1} = \rho_k$ and $A_{k+1} = A_k - 1$. If $A_k = 1$, then a renewal occurs at time $k+1$.

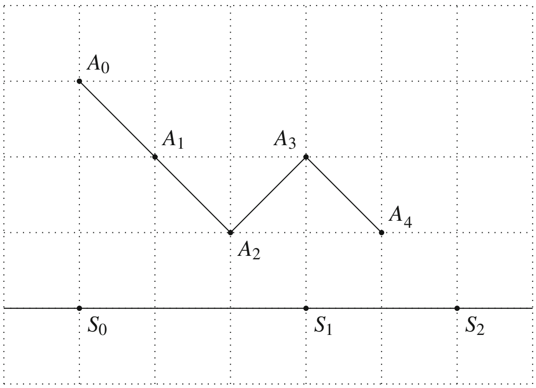

Fig. 8.1 An example of a residual lifetime process.

Proposition 8.1.5 *Under $\mathbb{P}_a$, the forward recurrence time chain $\{A_k,\ k \in \mathbb{N}\}$ is a Markov chain on $\mathbb{N}^*$ with initial distribution μ_a defined in (8.1.15) and kernel Q defined by*

$$Q(1, j) = b(j) \,, \quad j \geq 1 \,, \tag{8.1.16a}$$

$$Q(j, j-1) = 1 \,, \quad j \geq 2 \,. \tag{8.1.16b}$$

The Markov kernel Q is irreducible on $\mathsf{X} = \{0, \ldots, \sup\{n \in \mathbb{N} : b(n) \neq 0\}\}$ and recurrent. If the mean waiting time m is finite, then Q is positive recurrent with invariant distribution μ_s given by

$$\mu_s(k) = m^{-1} \sum_{j=k}^{\infty} b(j) \,, \ k \geq 1 \,. \tag{8.1.17}$$

Proof. For $\ell \geq 0$, set $\mathscr{F}_\ell^A = \sigma(A_j, j \leq \ell)$. Let $C \in \mathscr{F}_k^A$. Since $\rho_j \leq \rho_k$ for $j \leq k$, if $C \in \mathscr{F}_k^A$, then $C \cap \{\rho_k = \ell\} \in \mathscr{F}_\ell^S$. This yields

$$\begin{aligned}
\mathbb{P}_a(C, A_k = 1, A_{k+1} = j) &= \sum_{\ell=0}^{k+1} \mathbb{P}_a(C, A_k = 1, \rho_k = \ell, Y_{\ell+1} = j) \\
&= \sum_{\ell=0}^{k+1} \mathbb{P}_a(C, A_k = 1, \rho_k = \ell)\mathbb{P}_a(Y_{\ell+1} = j) \\
&= \mathbb{P}_a(C, A_k = 1)b(j) \,.
\end{aligned}$$

This proves (8.1.16a). The equality (8.1.16b) follows from the observation already made that if $A_k > 1$, then $A_{k+1} = A_k - 1$.

Set $k_0 = \sup\{n \in \mathbb{N} : b(n) > 0\} \in \bar{\mathbb{N}}$. For all $k \in \{1, \ldots, k_0\}$, $Q^k(k,1) = 1$, and if $\ell \geq k$ is such that $b(\ell) > 0$, then $Q^\ell(1,k) \geq b(\ell) > 0$. Thus Q is irreducible. Since $\mathbb{P}_1(\tau_1 = k) = b(k)$, we have $\mathbb{P}_1(\tau_1 < \infty) = 1$, showing that Q is recurrent. To check that μ_s is invariant for Q, note that for $j \geq 1$,

$$\begin{aligned}
\mu_s Q(j) &= \sum_{i=1}^{\infty} \mu_s(i)Q(i,j) = \mu_s(1)b(j) + \mu_s(j+1) \\
&= m^{-1}\left(b(j) + \sum_{\ell=j+1}^{\infty} b(\ell)\right) = \mu_s(j) \,.
\end{aligned}$$

□

We now provide a useful uniform bound on the distribution of the forward recurrence time chain of the pure renewal process.

Lemma 8.1.6 *For all $n \in \mathbb{N}$ and $k \geq 1$, $\mathbb{P}_0(A_n = k) \leq \mathbb{P}_0(Y_1 \geq k)$.*

Proof. For $n \geq 0$ and $k \geq 1$, we have

$$\begin{aligned}
\mathbb{P}_0(A_n = k) &= \sum_{j=0}^{\infty} \mathbb{P}_0(S_j \leq n < S_{j+1}, A_n = k) \\
&= \sum_{j=0}^{\infty} \sum_{i=0}^{n} \mathbb{P}_0(S_j = i, Y_{j+1} = n + k - i) \\
&= \sum_{i=0}^{n} \left(\sum_{j=0}^{\infty} \mathbb{P}_0(S_j = i)\right) \mathbb{P}_0(Y_1 = k + i) \\
&= \sum_{i=0}^{n} u(i)\mathbb{P}_0(Y_1 = k + i) \leq \mathbb{P}_0(Y_1 \geq k) \,.
\end{aligned}$$

□

8.1.2 Blackwell's and Kendall's Theorems

In this section, we assume that the delay distribution is periodic. Blackwell's theorem (Theorem 8.1.7) shows that for every delay distribution a, the delayed renewal sequence $\{v_a(n), n \in \mathbb{N}\}$ converges to the renewal intensity $1/m$, where $m \in [1, \infty]$ is the (possibly infinite) mean of the delay distribution.

Theorem 8.1.7. *Assume that b is aperiodic. Then for every delay distribution a,*

$$\lim_{n\to\infty} v_a(n) = 1/m \,, \tag{8.1.18}$$

with $m = \sum_{k=1}^{\infty} kb(k) \in (0, \infty]$.

Proof. Let Q be the kernel of the forward recurrence time chain, defined in (8.1.16). The Markov kernel Q is irreducible on $F = \{1, \ldots, \sup\{j \in \mathbb{N}^* : b(j) \neq 0\}\}$. For $k \geq 1$ such that $b(k) > 0$, we have

$$Q^k(1,1) \geq Q(1,k)Q^{k-1}(k,1) = Q(1,k) = b(k) > 0 \,.$$

Since the distribution b is aperiodic, this proves that the state 1 is aperiodic. Since the kernel Q is, moreover, irreducible, this implies that Q aperiodic by Proposition 6.3.4. By definition, $A_{n-1} = 1$ if and only if there is a renewal at time n; thus for $n \geq 1$,

$$v_a(n) = \mathbb{P}_a(A_{n-1} = 1) \,. \tag{8.1.19}$$

If $m < \infty$, we have seen in Proposition 8.1.5 that the probability measure μ_s defined in (8.1.17) is invariant for Q. Thus, applying Theorem 7.6.4, we obtain

$$\lim_{n\to\infty} v_a(n) = \lim_{n\to\infty} \mathbb{P}_a(A_{n-1} = 1) = \mu_s(1) = \frac{1}{m} \,.$$

If $m = \infty$, the chain is null recurrent, and Theorem 7.6.5 yields

$$\lim_{n\to\infty} v_a(n) = \lim_{n\to\infty} \mathbb{P}_a(A_{n-1} = 1) = 0 \,.$$

This proves (8.1.18) when $m = \infty$. □

Before investigating rates of convergence, we state an interesting consequence of Theorem 8.1.7.

Lemma 8.1.8 *Assume that the waiting time distribution b is aperiodic and that the mean waiting time m is finite. Let* N *be a subset of* $\mathbb{N}$. *Assume that* $\operatorname{card}(\mathsf{N}) = \infty$. *Then for every delay distribution a,* $\mathbb{P}_a(\sum_{k=0}^{\infty} \mathbb{1}_{\mathsf{N}}(S_k) = \infty) = 1$.

Proof. Let τ_n be the hitting time of the state n by the renewal process $\{S_k,\ k \in \mathbb{N}\}$. Note that

$$\mathbb{P}_0(\tau_n < \infty) = \mathbb{P}_0\left(\bigcup_{\ell=0}^{\infty}\{S_\ell = n\}\right) = u(n) ,$$

where u is the pure renewal sequence. Let $\eta \in (0, 1/m)$. By Theorem 8.1.7, there exists $\ell \geq 1$ such that for all $n \geq \ell$, $\mathbb{P}_0(\tau_n < \infty) = u(n) \geq \eta$. Since $\mathrm{card}(\mathsf{N}) = \infty$, for each $i \in \mathbb{N}$, we can choose $j \in \mathsf{N}$ such that $j \geq i + \ell$. Then

$$0 < \eta \leq \mathbb{P}_0(\tau_{j-i} < \infty) = \mathbb{P}_i(\tau_j < \infty) \leq \mathbb{P}_i(\sigma_{\mathsf{N}} < \infty) .$$

For every delay distribution a, $\mathbb{P}_a(\sum_{j=0}^{\infty} \mathbb{1}_{\mathsf{N}}(S_j) = \infty) = 1$. Since $\inf_{i \in \mathbb{N}} \mathbb{P}_i(\sigma_{\mathsf{N}} < \infty) \geq \eta$, by applying Theorem 4.2.6 to the Markov chain $\{S_n,\ n \in \mathbb{N}\}$, we finally obtain that $\mathbb{P}_a(\sum_{j=0}^{\infty} \mathbb{1}_{\mathsf{N}}(S_j) = \infty) = 1$. $\square$

If the waiting distribution has geometric moments, Kendall's theorem shows that the convergence in Theorem 8.1.7 holds at a geometric rate.

Theorem 8.1.9. *Assume that the waiting distribution b is aperiodic. Then the following properties are equivalent:*

(i) There exists $\beta > 1$ such that the series $\sum_{n=1}^{\infty} b(n) z^n$ is absolutely summable for all $|z| < \beta$.
(ii) There exist $\tilde{\beta} > 1$ and $\lambda > 0$ such that the series $\sum_{n=0}^{\infty}\{u(n) - \lambda\} z^n$ is absolutely summable for all $|z| < \tilde{\beta}$.

In both cases, the mean waiting time is finite and equal to λ^{-1}.

Proof. We first prove that both assumptions imply that the mean waiting time $m = \sum_{k=1}^{\infty} k b(k)$ is finite. This is a straightforward consequence of (i), and it is also implied by (ii), since in that case, $\lim_{n\to\infty} u(n) = \lambda$, and this limit is also equal to m^{-1} by Blackwell's theorem. Thus $\lambda > 0$ implies $m < \infty$.

Set now $w(0) = 1$, $w(n) = u(n) - u(n-1)$ for $n \geq 1$, and for $|z| < 1$,

$$F(z) = \sum_{n=0}^{\infty} (u(n) - m^{-1}) z^n , \; W(z) = \sum_{n=0}^{\infty} w(n) z^n .$$

Then for all $|z| < 1$, we get

$$W(z) = (1-z)F(z) + 1/m . \tag{8.1.20}$$

Recall that $B(z)$ and $U(z)$ denote the generating functions of the waiting distribution b and the renewal sequence u, respectively, and that $U(z) = (1 - B(z))^{-1}$ for $|z| < 1$ by (8.1.10). Thus for $|z| < 1$,

$$W(z) = (1-z)U(z) = (1-z)(1 - B(z))^{-1} . \tag{8.1.21}$$

Assume first that (i) holds. Note that $|B(z)| < 1$ for $|z| < 1$. The aperiodicity of b implies that $B(z) \neq 1$ for all $|z| \leq 1$, $z \neq 1$. The proof is by contradiction. Assume that there exists $\theta \in (0, 2\pi)$ such that $B(\mathrm{e}^{\mathrm{i}\theta}) = 1$. Then

$$\sum_{k=1}^{\infty} b(k)\cos(k\theta) = 1 \,.$$

If $\theta \notin \{2\pi/k : k \in \mathbb{N}^*\}$, then $|\cos(k\theta)| < 1$ for all $k \in \mathbb{N}$ and $\sum_{k=1}^{\infty} b(k)\cos(k\theta) < 1$, which is a contradiction. Therefore, there exists an integer $k_0 > 1$ such that $\theta = 2\pi/k_0$. Since b is aperiodic, there exists k such that $b(k) \neq 0$ and $\cos(2\pi k/k_0) < 1$. This implies that $\sum_{k=1}^{\infty} b(k)\cos(2\pi k/k_0) < 1$, which is again a contradiction.

Thus 1 is the only root of $B(z) = 1$, and moreover, it is a single root, since $B'(1) = \sum_{n=1}^{\infty} nb(n) = m \neq 0$. The function $z \mapsto 1 - B(z)$ is analytic on the disk $\{z \in \mathbb{C} : |z| < \beta\}$ and does not vanish on $\{z \in \mathbb{C} : |z| \leq 1\} \setminus \{1\}$. Since the zeros of an analytic function are isolated, there exists $\tilde{\beta} \in (1, \beta)$ such that $1 - B(z) \neq 0$ for all $z \neq 1$ such that $|z| < \tilde{\beta}$.

Thus the function $z \mapsto W(z) = (1-z)(1-B(z))^{-1}$ is analytic on the disk $\left\{z \in \mathbb{C} : |z| < \tilde{\beta}\right\}$. This implies that $\sum_{k=1}^{\infty} r^k |w(k)| < \infty$ for $r \in \left(1, \tilde{\beta}\right)$. Since $\lim_{k\to\infty} u(n) = 1/m$ by Blackwell's theorem (Theorem 8.1.7), we obtain for $r \in \left(1, \tilde{\beta}\right)$,

$$\begin{aligned} r^n |u(n) - 1/m| &= r^n \left| \lim_{k\to\infty} \{u(n) - u(k)\} \right| = r^n \left| \lim_{k\to\infty} \sum_{j=n+1}^{k} w(j) \right| \\ &\leq r^n \sum_{k=n+1}^{\infty} |w(k)| \leq \sum_{k=n+1}^{\infty} |w(k)| r^k < \infty \,. \end{aligned}$$

This proves (ii).

Conversely, if (ii) holds, then the function $z \mapsto W(z)$ is analytic on the disk $\left\{z \in \mathbb{C} : |z| < \tilde{\beta}\right\}$. Indeed, for all $r < \tilde{\beta}$,

$$\begin{aligned} \sum_{n=1}^{\infty} |u(n) - u(n-1)| r^n &\leq \sum_{n=1}^{\infty} |u(n) - m^{-1}| r^n + \sum_{n=1}^{\infty} |u(n-1) - m^{-1}| r^n \\ &\leq (1+r) \sum_{n=0}^{\infty} |u(n) - m^{-1}| r^n \,. \end{aligned}$$

This implies that $\sum_{n=1}^{\infty} |u(n) - u(n-1)| < \infty$; hence applying Blackwell's theorem, we obtain

$$W(1) = 1 + \sum_{n=1}^{\infty} \{u(n) - u(n-1)\} = \lim_{n\to\infty} u(n) = 1/m \,.$$

By (8.1.21), we have, for $|z| < 1$, $B(z) = 1 + (z-1)/W(z)$. Since $B(z)$ is bounded on $\{|z| \leq 1\}$, this implies that $W(z) \neq 0$ for $|z| \leq 1$, $z \neq 1$. This implies that

$r_0 = \inf\{|z| : W(z) = 0\} > 1$; hence B is analytic on $\{|z| < \beta\}$ with $\beta = \tilde{\beta} \wedge r_0 > 1$. This proves (i). □

8.2 Renewal Theory and Atomic Markov Chains

In this section, P is a Markov kernel on $\mathsf{X} \times \mathscr{X}$ having a recurrent atom α, i.e., $\mathbb{P}_\alpha(\sigma_\alpha < \infty) = 1$. We will now build a renewal process based on the successive visits to the atom α. Define recursively

$$S_0 = \sigma_\alpha \;, \qquad S_{k+1} = S_k + \sigma_\alpha \circ \theta_{S_k} \;, \qquad k \geq 0 \;. \tag{8.2.1}$$

By construction, the random variables S_k are the successive visits of the chain $\{X_n\}$ to the atom α: $S_k = \sigma_\alpha^{(k+1)}$ for all $k \geq 0$. Define

$$Y_0 = S_0 = \sigma_\alpha \;, \qquad Y_k = S_k - S_{k-1} \;, \qquad k \geq 1 \;.$$

Proposition 8.2.1 *Let α be a recurrent atom. Let $\xi \in \mathbb{M}_1(\mathscr{X})$ be an initial distribution such that $\mathbb{P}_\xi(\sigma_\alpha < \infty) = 1$. Then under $\mathbb{P}_\xi$, $\{S_n,\ n \in \mathbb{N}\}$ is a renewal process with waiting time distribution b and delay distribution a_ξ defined by*

$$b(k) = \mathbb{P}_\alpha(\sigma_\alpha = k) \;, \;\; a_\xi(k) = \mathbb{P}_\xi(\sigma_\alpha = k) \;, \;\; k \in \mathbb{N}.$$

Proof. This is a direct application of Proposition 6.5.1, since $Y_0 = \sigma_\alpha$ is $\mathscr{F}_{\sigma_\alpha}$ measurable by definition, and for $k \geq 1$, $Y_k = \sigma_\alpha \circ \theta_{S_{k-1}}$. □

The renewal process $\{S_n,\ n \in \mathbb{N}\}$ is called the renewal process associated to the Markov chain $\{X_n,\ n \in \mathbb{N}\}$. The pure and delayed renewal sequences associated to this renewal process can be related to the original chain as follows. For $\xi \in \mathbb{M}_1(\mathscr{X})$ and $n \geq 0$,

$$u(n) = \mathbb{P}_\alpha(X_n \in \alpha) \;, \tag{8.2.2a}$$
$$v_{a_\xi}(n) = \mathbb{P}_\xi(X_n \in \alpha) \;. \tag{8.2.2b}$$

The initial distribution of the forward recurrence process associated to the renewal process is given by

$$\mu_{a_\xi}(k) = \mathbb{P}_\xi(\sigma_\alpha = k) \;, \;\; k \geq 1 \;. \tag{8.2.3}$$

If the atom α is accessible and positive, then Theorem 6.4.2 shows that the Markov kernel P admits a unique invariant probability π. As shown in the following result, if the Markov chain is started from stationarity, then the renewal sequence associated to the atom α is also stationary.

Proposition 8.2.2 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an accessible and positive atom α. Denote by π the unique invariant probability. Then a_π is the stationary delay distribution, and the invariant probability μ_{a_π} of the associated forward recurrence time chain is given by $\mu_{a_\pi}(k) = \mathbb{P}_\pi(\sigma_\alpha = k)$, $k \geq 1$.*

Proof. By (8.2.2b), $v_{a_\pi}(n) = \mathbb{P}_\pi(X_n \in \alpha) = \pi(\alpha)$ for all $n \in \mathbb{N}$. Thus v_{a_π} is constant. The second statement is an immediate consequence of (8.2.2), (8.2.3), and Proposition 8.1.5. □

Corollary 8.2.3 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an accessible, aperiodic, and positive atom α. Then*

$$\lim_{n\to\infty} P^n(\alpha, \alpha) = \frac{1}{\mathbb{E}_\alpha[\sigma_\alpha]} .$$

Proof. Since the atom is aperiodic, the waiting-time distribution $b(n) = \mathbb{P}_\alpha(\sigma_\alpha = n)$ $n \in \mathbb{N}^*$ is also aperiodic. Since the atom is positive, we have $\sum_{j=1}^\infty j\mathbb{P}_\alpha(\sigma_\alpha = j) = \mathbb{E}_\alpha[\sigma_\alpha] < \infty$. Applying Blackwell's theorem (see Theorem 8.1.7), we get

$$\lim_{n\to\infty} P^n(\alpha, \alpha) = \lim_{n\to\infty} u(n) = \frac{1}{\mathbb{E}_\alpha[\sigma_\alpha]} > 0 .$$

□

8.2.1 Convergence in Total Variation Distance

In this section, we will study the convergence of the iterates of a Markov kernel P using the renewal theory. We will show that if the Markov kernel admits an aperiodic positive atom, then the iterates of the chain converge in total variation toward the invariant law. The key to the proof is to combine the first-entrance last-exit decomposition (see Section 3.4) and Blackwell's and Kendall's theorems.

Proposition 8.2.4 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with an accessible atom α satisfying $\mathbb{P}_\alpha(\sigma_\alpha < \infty) = 1$. Then for all $f \in \mathbb{F}_+(\mathsf{X}) \cup \mathbb{F}_b(\mathsf{X})$, $\xi \in \mathbb{M}_1(\mathscr{X})$, and $n \geq 1$, we get*

$$\mathbb{E}_\xi[f(X_n)] = \mathbb{E}_\xi[\mathbb{1}_{\{\sigma_\alpha \geq n\}} f(X_n)] + a_\xi * u * \psi_f(n), \tag{8.2.4}$$

where for $n \geq 1$,

$$\psi_f(0) = 0 \quad \text{and} \quad \psi_f(n) = \mathbb{E}_\alpha[f(X_n)\mathbb{1}_{\{\sigma_\alpha \geq n\}}] \,, n \geq 1 \,. \tag{8.2.5}$$

Proof. Recall that for $k \geq 1$, $\mathbb{P}_\xi(X_k = \alpha) = [a_\xi * u](k)$. Then

$$\begin{aligned}
&\mathbb{E}_\xi[f(X_n)] \\
&= \mathbb{E}_\xi[\mathbb{1}_{\{\sigma_\alpha \geq n\}} f(X_n)] + \sum_{k=1}^{n-1} \mathbb{E}_\xi[\mathbb{1}\{X_k \in \alpha, X_{k+1} \notin \alpha, \ldots, X_{n-1} \notin \alpha\} f(X_n)] \\
&= \mathbb{E}_\xi[\mathbb{1}_{\{\sigma_\alpha \geq n\}} f(X_n)] + \sum_{k=1}^{n-1} \mathbb{P}_\xi(X_k \in \alpha) \mathbb{E}_\alpha[\mathbb{1}\{X_1 \notin \alpha, \ldots, X_{n-k-1} \notin \alpha\} f(X_{n-k})] \\
&= \mathbb{E}_\xi[\mathbb{1}_{\{\sigma_\alpha \geq n\}} f(X_n)] + \sum_{k=1}^{n-1} \mathbb{P}_\xi(X_k \in \alpha) \psi_f(n-k) \\
&= \mathbb{E}_\xi[\mathbb{1}_{\{\sigma_\alpha \geq n\}} f(X_n)] + a_\xi * u * \psi_f(n) \,,
\end{aligned}$$

where we have used in the last identity $\psi_f(0) = 0$ and $a_\xi(0) = 0$. □

Corollary 8.2.5 *Assume that P is a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with an accessible positive atom α. Denote by π the invariant probability. Then for all $\xi \in \mathbb{M}_1(\mathscr{X})$, we get*

$$\|\xi P^n - \pi\|_{\mathrm{TV}} \leq \mathbb{P}_\xi(\sigma_\alpha \geq n) + |a_\xi * u - \pi(\alpha)| * \psi(n) + \pi(\alpha) \sum_{k=n+1}^{\infty} \psi(k) \,, \tag{8.2.6}$$

where $\psi(0) = 0$ *and* $\psi(n) = \psi_1(n) = \mathbb{P}_\alpha(n \leq \sigma_\alpha)$ *for* $n \geq 1$.

Proof. Let $f \in \mathbb{F}_b(\mathsf{X})$. By Theorem 6.4.2, the invariant probability may be expressed as

$$\begin{aligned}
\pi(f) &= \pi(\alpha) \mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} f(X_k)\right] \\
&= \pi(\alpha) \sum_{k=1}^{\infty} \mathbb{E}_\alpha\left[\mathbb{1}_{\{\sigma_\alpha \geq k\}} f(X_k)\right] = \pi(\alpha) \sum_{k=1}^{\infty} \psi_f(k) < \infty \,,
\end{aligned}$$

where ψ_f is defined in (8.2.5). Since $\pi(\alpha) \sum_{k=1}^{n} \psi_f(n) = \pi(\alpha) * \psi_f(n)$, Proposition 8.2.4, with (8.2.4), implies

$$\begin{aligned}
&\xi P^n(f) - \pi(f) \\
&\qquad = \mathbb{E}_\xi[f(X_n)\mathbb{1}_{\{n \leq \sigma_\alpha\}}] + [a_\xi * u - \pi(\alpha)] * \psi_f(n) - \pi(\alpha) \sum_{k=n+1}^{\infty} \psi_f(k) \,.
\end{aligned} \tag{8.2.7}$$

Therefore, by taking the supremum over $f \in \mathbb{F}_b(\mathsf{X})$ satisfying $|f|_\infty \le 1$, we get

$$\begin{aligned}\|\xi P^n - \pi\|_{\mathrm{TV}} &= \sup_{|f|_\infty \le 1} |\xi P^n f - \pi(f)| \\ &\le \mathbb{P}_\xi(n \le \sigma_\alpha)) + |a_\xi * u - \pi(\alpha)| * \psi(n) + \pi(\alpha) \sum_{k=n+1}^{\infty} \psi(k) .\end{aligned}$$

□

We now apply Blackwell's theorem (Theorem 8.1.7) to show the convergence of the iterates of the Markov chain to its invariant probability measure.

Theorem 8.2.6. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P admits an accessible, aperiodic, and positive atom α and an invariant probability measure π. If $\xi \in \mathbb{M}_1(\mathscr{X})$ is such that $\mathbb{P}_\xi(\sigma_\alpha < \infty) = 1$, then* $\lim_{n\to\infty} \mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) = 0$.

Proof. We will use Corollary 8.2.5 and show that each term on the right-hand side of the inequality (8.2.6) tends to 0. Note first that

$$\sum_{k=1}^{\infty} \psi(k) = \sum_{k=1}^{\infty} \mathbb{P}_\alpha(k \le \sigma_\alpha) = \mathbb{E}_\alpha[\sigma_\alpha] < \infty .$$

This implies that $\lim_{n\to\infty} \sum_{k=n+1}^{\infty} \psi(k) = 0$. On the other hand, since $\mathbb{P}_\xi(\sigma_\alpha < \infty) = 1$, we get that $\lim_{n\to\infty} \mathbb{P}_\xi(\sigma_\alpha \ge n) = 0$. By Corollary 8.2.3, $\lim_{n\to\infty} u(n) = \{\mathbb{E}_\alpha(\sigma_\alpha)\}^{-1} = \pi(\alpha)$. Recall that if $\{v(n),\ n \in \mathbb{N}\}$ and $\{w(n),\ n \in \mathbb{N}\}$ are two sequences such that $\lim_{n\to\infty} v(n) = 0$ and $\sum_{n=0}^{\infty} |w(n)| < \infty$, then $\lim_{n\to\infty} v * w(n) = 0$. Therefore, we get that $\lim_{n\to\infty} [a_\xi * \{u - \pi(\alpha)\}](n) = 0$ and

$$\lim_{n\to\infty} \{a_\xi * \pi(\alpha)\}(n) = \lim_{n\to\infty} \sum_{k=1}^{n} a_\xi(k) = \pi(\alpha) .$$

We conclude by decomposing the difference $a_\xi * u - \pi(\alpha)$ as follows:

$$a_\xi * u - \pi(\alpha) = a_\xi * \{u - \pi(\alpha)\} + a_\xi * \pi(\alpha) - \pi(\alpha) .$$

□

Remark 8.2.7. Recall that $\alpha_+ = \{x \in \mathsf{X} : \mathbb{P}_x(\sigma_\alpha < \infty) = 1\})$ and that $\pi(\alpha_+) = 1$ by Lemma 6.4.5. Hence for every $\xi \in \mathbb{M}_1(\mathscr{X})$ such that $\xi(\alpha_+^c) = 0$, we get that $\mathbb{P}_\xi(\sigma_\alpha < \infty) = 1$. ▲

8.2.2 Geometric Convergence in Total Variation Distance

We now apply Kendall's theorem (Theorem 8.1.9) to prove the geometric convergence of the iterates of the Markov kernel to its stationary distribution.

Lemma 8.2.8 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P admits an aperiodic positive atom α. Denote by π its unique invariant probability. Then $\mathbb{E}_\alpha[\beta^{\sigma_\alpha}] < \infty$ for some $\beta > 1$ if and only if $\sum_{n=1}^\infty \delta^n |P^n(\alpha,\alpha) - \pi(\alpha)| < \infty$ for some $\delta > 1$.*

Proof. We apply Theorem 8.1.9 with $b(n) = \mathbb{P}_\alpha(\sigma_\alpha = n)$, $u(n) = \mathbb{P}_\alpha(X_n = \alpha) = P^n(\alpha,\alpha)$, and $\pi(\alpha) = 1/\mathbb{E}_\alpha[\sigma_\alpha]$. □

Theorem 8.2.9. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P admits an accessible aperiodic atom $\alpha \in \mathscr{X}$ and $\beta > 1$ such that $\mathbb{E}_\alpha[\beta^{\sigma_\alpha}] < \infty$. Then P has a unique invariant probability π, and there exist $\delta \in (1,\beta]$ and $\varsigma < \infty$ such that for all $\xi \in \mathbb{M}_1(\mathscr{X})$,*

$$\sum_{n=1}^\infty \delta^n \mathrm{d_{TV}}(\xi P^n, \pi) \le \varsigma \mathbb{E}_\xi[\delta^{\sigma_\alpha}] . \tag{8.2.8}$$

Proof. Set $b(n) = \mathbb{P}_\alpha(\sigma_\alpha = n)$, $u(n) = \mathbb{P}_\alpha(X_n = \alpha) = P^n(\alpha,\alpha)$. Since $\psi(n) = \mathbb{P}_\alpha(\sigma_\alpha \ge n)$, we have for $\delta > 1$,

$$\sum_{n=1}^\infty \psi(n)\delta^n \le \frac{\delta}{\delta-1}\mathbb{E}_\alpha[\delta^{\sigma_\alpha}] . \tag{8.2.9}$$

The assumption implies that $\mathbb{E}_\alpha[\sigma_\alpha] < \infty$, and by Theorem 6.4.2, it admits a unique invariant probability π. We will use the well-known property that the moment-generating function of a convolution is the product of the moment-generating functions of the terms of the product, i.e., for all nonnegative sequences $\{c(n),\, n \in \mathbb{N}\}$ and $\{d(n),\, n \in \mathbb{N}\}$ and for every $\delta > 0$,

$$\sum_{n=0}^\infty c * d(n)\delta^n = \left(\sum_{i=0}^\infty c(i)\delta^i\right)\left(\sum_{j=0}^\infty d(j)\delta^j\right) . \tag{8.2.10}$$

The bound (8.2.6) in Corollary 8.2.5 and (8.2.9) yield $2\sum_{n=1}^\infty \delta^n \mathrm{d_{TV}}(\xi P^n, \pi) \le \sum_{i=1}^3 A_i$ with $A_1 = \sum_{n=1}^\infty \mathbb{P}_\xi(\sigma_\alpha \ge n)\delta^n$, $A_2 = \pi(\alpha)\sum_{n=1}^\infty \sum_{k=n+1}^\infty \psi(k)\delta^n$, and $A_3 = B_1 \cdot B_2$ with

$$B_1 = \sum_{n=1}^\infty |a_\xi * u - \pi(\alpha)|\delta^n \quad \text{and} \quad B_2 = \sum_{n=1}^\infty \psi(n)\delta^n. \tag{8.2.11}$$

We will now consider each of these terms. Note first that

$$A_1 = \mathbb{E}_\xi \left[\sum_{n=1}^{\sigma_\alpha} \delta^n \right] \le \frac{\delta}{\delta - 1} \mathbb{E}_\xi [\delta^{\sigma_\alpha}] \,. \tag{8.2.12}$$

Now consider A_2:

$$A_2 = \pi(\alpha) \sum_{k=1}^{\infty} \psi(k) \sum_{n=1}^{k-1} \delta^n \le \frac{\delta\pi(\alpha)}{\delta - 1} \sum_{k=1}^{\infty} \psi(k) \delta^k \le \frac{\delta\pi(\alpha)}{(\delta-1)^2} \mathbb{E}_\alpha [\delta^{\sigma_\alpha}] \,. \tag{8.2.13}$$

Finally, we consider A_3. Note first that

$$A_3 = B_1 \cdot B_2 \le \frac{\delta}{\delta - 1} \mathbb{E}_\alpha [\delta^{\sigma_\alpha}] \sum_{n=0}^{\infty} |a_\xi * u(n) - \pi(\alpha)| \delta^n \,\,. \tag{8.2.14}$$

Using for $n \ge 0$ the bound

$$|a_\xi * u(n) - \pi(\alpha)| \le a_\xi * |u(n) - \pi(\alpha)| + \pi(\alpha) \sum_{k=n+1}^{\infty} a_\xi(k) \,,$$

we obtain, using again (8.2.10),

$$\begin{aligned} B_1 &\le \left(\sum_{n=0}^{\infty} a_\xi(n) \delta^n \right) \left(\sum_{n=0}^{\infty} |u(n) - \pi(\alpha)| \delta^n \right) + \sum_{n=0}^{\infty} \sum_{k=n+1}^{\infty} a_\xi(k) \delta^n \\ &\le \mathbb{E}_\xi [\delta^{\sigma_\alpha}] \sum_{n=0}^{\infty} |u(n) - \pi(\alpha)| \delta^n + \frac{\delta}{\delta - 1} \mathbb{E}_\xi [\delta^{\sigma_\alpha}] \,. \end{aligned} \tag{8.2.15}$$

By Lemma 8.2.8, there exists $\delta > 1$ such that $\sum_{n=0}^{\infty} |u(n) - \pi(\alpha)| \delta^n < \infty$. Hence using (8.2.14), there exists $\varsigma < \infty$ such that $A_3 \le \varsigma \mathbb{E}_\xi [\delta^{\sigma_\alpha}]$. Equation (8.2.8) follows from (8.2.12) and (8.2.13). □

8.3 Coupling Inequalities for Atomic Markov Chains

In this section, we obtain rates of convergence for $\mathrm{d_{TV}}(\xi P^n, \xi' P^n)$ using an approach based on coupling techniques. We have already used the coupling technique in Section 7.6 for Markov kernels on a discrete state space. We will show in this section how these techniques can be adapted to a Markov kernel P on a general state space admitting an atom.

Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an accessible atom α. Define the Markov kernel $\bar{P}$ on $\mathsf{X}^2 \times \mathscr{X}^{\otimes 2}$ as follows: for all $(x, x') \in \mathsf{X}^2$ and $A \in \mathscr{X}^{\otimes 2}$,

$$\bar{P}((x,x'),A) = \int P(x,\mathrm{d}y) P(x',\mathrm{d}y') \mathbb{1}_A(y,y') \,. \tag{8.3.1}$$

Let $\{(X_n, X'_n),\ n \in \mathbb{N}\}$ be the canonical process on the canonical product space $\Omega = (\mathsf{X} \times \mathsf{X})^{\mathbb{N}}$. For $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$, let $\bar{\mathbb{P}}_{\xi \otimes \xi'}$ be the probability measure on Ω such that $\{(X_n, X'_n),\ n \in \mathbb{N}\}$ is a Markov chain with kernel $\bar{P}$ and initial distribution $\xi \otimes \xi'$. The notation $\bar{\mathbb{E}}_{\xi \otimes \xi'}$ stands for the associated expectation operator. An important feature is that $\alpha \times \alpha$ is an atom for $\bar{P}$. Indeed, for all $x, x' \in \alpha$ and $A, A' \in \mathscr{X}$,

$$\bar{P}((x,x'), A \times A') = P(x,A)P(x',A) = P(\alpha,A)P(\alpha,A') .$$

For an initial distribution $\xi' \in \mathbb{M}_1(\mathscr{X})$ and a random variable Y on Ω, if the function $x \mapsto \bar{\mathbb{E}}_{\delta_x \otimes \xi'}[Y]$ does not depend on $x \in \alpha$, then we write $\bar{\mathbb{E}}_{\alpha \otimes \xi'}[Y]$ for $\bar{\mathbb{E}}_{\delta_x \otimes \xi'}[Y]$ when $x \in \alpha$. Similarly, for $x, x' \in \alpha$, we write $\bar{\mathbb{E}}_{\alpha \otimes \alpha}[Y]$ for $\bar{\mathbb{E}}_{\delta_x \otimes \delta_{x'}}[Y]$ if the latter quantity is constant on $\alpha \times \alpha$.

Let ξ and ξ' be two probability measures on X. Denote by T the return time to $\alpha \times \alpha$ for the Markov chain $\{(X_n, X'_n),\ n \in \mathbb{N}\}$, i.e.,

$$T = \sigma_{\alpha \times \alpha} = \inf\left\{n \geq 1 \,:\, (X_n, X'_n) \in \alpha \times \alpha\right\} . \tag{8.3.2}$$

The fundamental result about the coupling time T is stated in the following lemma (which is an atomic version of Proposition 7.6.3).

Lemma 8.3.1 *Let P be a Markov kernel with an atom α. For all $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$ and all $n \in \mathbb{N}$,*

$$\mathrm{d}_{\mathrm{TV}}(\xi P^n, \xi' P^n) \leq \bar{\mathbb{P}}_{\xi \otimes \xi'}(T \geq n) . \tag{8.3.3}$$

Moreover, for every nonnegative sequence $\{r(n),\ n \in \mathbb{N}\}$,

$$\sum_{n=0}^{\infty} r(n) \mathrm{d}_{\mathrm{TV}}(\xi P^n, \xi' P^n) \leq \bar{\mathbb{E}}_{\xi \otimes \xi'}\left[r^0(T)\right] , \tag{8.3.4}$$

where $r^0(n) = \sum_{k=0}^{n} r(k)$ for all $n \in \mathbb{N}$.

Proof. Let $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$. Then for all $f \in \mathbb{F}_b(\mathsf{X})$,

$$\begin{aligned}
\xi P^n(f) &= \bar{\mathbb{E}}_{\xi \otimes \xi'}[f(X_n)] \\
&= \bar{\mathbb{E}}_{\xi \otimes \xi'}[f(X_n)\mathbb{1}\{n \leq T\}] + \bar{\mathbb{E}}_{\xi \otimes \xi'}[f(X_n)\mathbb{1}\{n > T\}] \\
&= \bar{\mathbb{E}}_{\xi \otimes \xi'}[f(X_n)\mathbb{1}\{n \leq T\}] + \bar{\mathbb{E}}_{\xi \otimes \xi'}[\mathbb{1}\{n > T\} P^{n-T} f(\alpha)] .
\end{aligned}$$

Similarly,

$$\xi' P^n(f) = \bar{\mathbb{E}}_{\xi \otimes \xi'}[f(X'_n)\mathbb{1}\{n \leq T\}] + \bar{\mathbb{E}}_{\xi \otimes \xi'}[\mathbb{1}\{n > T\} P^{n-T} f(\alpha)] .$$

Altogether, this implies that

$$|\xi P^n(f) - \xi' P^n(f)| \leq \operatorname{osc}(f)\, \bar{\mathbb{P}}_{\xi \otimes \xi'}(n \leq T) .$$

The bound (8.3.3) follows by application of Proposition D.2.4. Applying (8.3.3) yields

$$\sum_{n=0}^{\infty} r(n) \mathrm{d}_{\mathrm{TV}}(\xi P^n, \xi' P^n) \le \sum_{n=0}^{\infty} \bar{\mathbb{E}}_{\xi \otimes \xi'}[r(n) \mathbb{1}\{T \ge n\}] = \bar{\mathbb{E}}_{\xi \otimes \xi'}\left[r^0(T)\right] .$$

□

Lemma 8.3.1 suggests that rates of convergence in total variation distance of the iterates of the kernel to the invariant probability will be obtained by finding conditions under which $\bar{\mathbb{E}}_{\xi \otimes \xi'}[r^0(T)] < \infty$. We first give a proof of Theorem 8.3.2) using the coupling method.

Theorem 8.3.2. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P admits an accessible, aperiodic, and positive atom α and invariant probability measure π. If $\xi \in \mathbb{M}_1(\mathscr{X})$ is such that $\mathbb{P}_\xi(\sigma_\alpha < \infty) = 1$, then* $\lim_{n\to\infty} \mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) = 0$.

Proof. Recall that $\alpha_+ = \{x \in \mathsf{X} : \mathbb{P}_x(\sigma_\alpha < \infty) = 1\}$ is the domain of attraction of the atom α. By Lemma 6.4.5, $\pi(\alpha_+) = 1$, whence $\mathbb{P}_\pi(\sigma_\alpha < \infty) = 1$. Write $\mathsf{N}' = \left\{k \in \mathbb{N} : X'_k \in \alpha\right\}$ (N is a random set). By Theorem 6.2.2, for every probability measure ξ' such that $\mathbb{P}_{\xi'}(\sigma_\alpha < \infty) = 1$, we have

$$\mathbb{P}_{\xi'}(\mathrm{card}(\mathsf{N}') = \infty) = \mathbb{P}_{\xi'}\left(\sum_{k=0}^{\infty} \mathbb{1}_\alpha(X'_k) = \infty\right) = 1 .$$

By the strong Markov property, the successive visits $\sigma_\alpha^{(n)}$ to α define an aperiodic renewal process with delay distribution $a(n) = \mathbb{P}_\xi(\tau_\alpha = n)$. Therefore, for each ω' such that $\mathrm{card}(\mathsf{N}'(\omega')) = \infty$, by Lemma 8.1.8, we have

$$\mathbb{P}_\xi\left(\sum_{n=1}^{\infty} \mathbb{1}_{\mathsf{N}'(\omega')}(\sigma_\alpha^{(n)}) = \infty\right) = 1.$$

This yields that

$$\bar{\mathbb{P}}_{\xi \otimes \xi'}\left(\sum_{n=1}^{\infty} \mathbb{1}_{\mathsf{N}'}(\sigma_\alpha^{(n)}) = \infty\right) = \mathbb{P}_{\xi'}(\mathrm{card}(\mathsf{N}') = \infty) = 1 .$$

Thus for every initial distribution ξ' such that $\mathbb{P}_{\xi'}(\sigma_\alpha < \infty) = 1$, we have

$$\bar{\mathbb{P}}_{\xi \otimes \xi'}(\sigma_{\alpha \times \alpha} < \infty) \ge \bar{\mathbb{P}}_{\xi \otimes \xi'}(\mathrm{card}(\mathsf{N}) = \infty) = 1 .$$

The proof is concluded by applying (8.3.3). □

We now state two technical lemmas that will be used to obtain polynomial rates of convergence.

Lemma 8.3.3 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with a positive atom α. For all $\xi \in \mathbb{M}_1(\mathscr{X})$ and $k, n \in \mathbb{N}^*$,*

$$\mathbb{E}_\xi\left[\{\sigma_\alpha^{(n)}\}^k\right] \le n^{k-1}[\mathbb{E}_\xi[\sigma_\alpha^k] + (n-1)\mathbb{E}_\alpha(\sigma_\alpha^k)] \,.$$

Proof. Since $\sigma_\alpha^{(n)} = \sigma_\alpha^{(n-1)} + \sigma_\alpha \circ \theta_{\sigma_\alpha^{(n-1)}}$, we have

$$\begin{aligned}\left\{\mathbb{E}_\xi\left[\{\sigma_\alpha^{(n)}\}^k\right]\right\}^{1/k} &\le \left\{\mathbb{E}_\xi\left[\{\sigma_\alpha^{(n-1)}\}^k\right]\right\}^{1/k} + \left\{\mathbb{E}_\xi\left[\sigma_\alpha^k \circ \theta_{\sigma_\alpha^{(n-1)}}\right]\right\}^{1/k} \\ &= \left\{\mathbb{E}_\xi\left[\{\sigma_\alpha^{(n-1)}\}^k\right]\right\}^{1/k} + \left\{\mathbb{E}_\alpha[\sigma_\alpha^k]\right\}^{1/k},\end{aligned}$$

and the result follows by induction. Using Jensen's inequality, we obtain

$$\begin{aligned}\mathbb{E}_\xi\left[\{\sigma_\alpha^{(n)}\}^k\right] &= n^k\left[\frac{1}{n}\left\{\mathbb{E}_\xi[\sigma_\alpha^k]\right\}^{1/k} + \frac{n-1}{n}\left\{\mathbb{E}_\alpha[\sigma_\alpha^k]\right\}^{1/k}\right]^k \\ &\le n^{k-1}\left\{\mathbb{E}_\xi[\sigma_\alpha^k] + (n-1)\mathbb{E}_\alpha[\sigma_\alpha^k]\right\} \,.\end{aligned}$$

□

Lemma 8.3.4 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with a positive aperiodic atom α. Assume that $\mathbb{E}_\alpha[\sigma_\alpha^k] < \infty$ for some $k \in \mathbb{N}^*$. Then there exists a constant $\varsigma < \infty$ such that for all ξ, $\xi' \in \mathbb{M}_1(\mathscr{X})$,*

$$\bar{\mathbb{E}}_{\xi\otimes\xi'}[T^k] \le \varsigma\mathbb{E}_\xi[\sigma_\alpha^k]\mathbb{E}_{\xi'}[\sigma_\alpha^k] \,. \tag{8.3.5}$$

Proof. Denote by π the unique invariant probability. Set $\rho_n = P^n(\alpha, \alpha)$. By Proposition 6.3.6, we may choose m large enough that $\rho_n > 0$ for all $n \ge m$. Since $P^n(\alpha, \alpha) > 0$ for all $n \ge m$ and by Corollary 8.2.3, $\lim_{n\to\infty}\rho_n = 1/\pi(\alpha)$, we obtain $\sup_{n\ge m}\rho_n^{-1} \le \varsigma < \infty$.

Let $T = \sigma_{\alpha\times\alpha}$, $S = \sigma_{\alpha\times\mathsf{X}}^{(m)}$. Let $r \in \mathbb{N}^*$. Since $T \le S + T \circ \theta_S$, we get

$$\bar{\mathbb{E}}_{\xi\otimes\xi'}[T^r] \le 2^{r-1}\left\{\mathbb{E}_\xi[\{\sigma_\alpha^{(m)}\}^r] + \bar{\mathbb{E}}_{\xi\otimes\xi'}[T^r \circ \theta_S]\right\} \,. \tag{8.3.6}$$

Then

$$\begin{aligned}\bar{\mathbb{E}}_{\xi\otimes\xi'}[T^r \circ \theta_S] &= \sum_{n=m}^{\infty}\bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbb{1}_{\{S=n\}}\bar{\mathbb{E}}_{\alpha,X'_n}[T^r]\right] \\ &= \sum_{n=m}^{\infty}\mathbb{P}_\xi(\sigma_\alpha^{(m)} = n)\bar{\mathbb{E}}_{\xi\otimes\xi'}[\bar{\mathbb{E}}_{\alpha,X'_n}[T^r]] \\ &= \sum_{n=m}^{\infty}\mathbb{P}_\xi(\sigma_\alpha^{(m)} = n)\rho_n^{-1}\mathbb{P}_\alpha(X_n \in \alpha)\bar{\mathbb{E}}_{\xi\otimes\xi'}[\bar{\mathbb{E}}_{\alpha,X'_n}[T^r]] \,.\end{aligned} \tag{8.3.7}$$

Note that $\bar{\mathbb{E}}_{\xi\otimes\xi'}[\bar{\mathbb{E}}_{\alpha,X'_n}[T^r]] = \int \xi'(\mathrm{d}x')\bar{\mathbb{E}}_{\alpha,x'}[T^r]$ does not depend on the initial distribution $\xi \in \mathbb{M}_1(\mathscr{X})$. Hence $\bar{\mathbb{E}}_{\xi\otimes\xi'}[\bar{\mathbb{E}}_{\alpha,X'_n}[T^r]] = \bar{\mathbb{E}}_{\alpha\otimes\xi'}[\bar{\mathbb{E}}_{\alpha,X'_n}[T^r]]$. Plugging this expression into (8.3.7) yields

$$
\begin{aligned}
\bar{\mathbb{E}}_{\xi\otimes\xi'}[T^r\circ\theta_S] &\le \sum_{n=m}^{\infty}\mathbb{P}_\xi(\sigma_\alpha^{(m)}=n)\rho_n^{-1}\bar{\mathbb{E}}_{\alpha\otimes\xi'}[\mathbb{1}_\alpha(X_n)\bar{\mathbb{E}}_{\alpha,X'_n}[T^r]] \\
&\le \sum_{n=m}^{\infty}\mathbb{P}_\xi(\sigma_\alpha^{(m)}=n)\rho_n^{-1}\bar{\mathbb{E}}_{\alpha\otimes\xi'}[\bar{\mathbb{E}}_{X_n,X'_n}[T^r]]\,. \qquad (8.3.8)
\end{aligned}
$$

By the Markov property, we get $\bar{\mathbb{E}}_{\alpha\otimes\xi'}[\bar{\mathbb{E}}_{X_n,X'_n}[T^r]] = \bar{\mathbb{E}}_{\alpha\otimes\xi'}[T^r\circ\theta_n]$. Since $T\circ\theta_n \le \sigma_{\alpha\times\alpha}^{(n)}$, we get by applying Lemma 8.3.3 with $T=\sigma_{\alpha\times\alpha}$ instead of σ_α,

$$
\begin{aligned}
\bar{\mathbb{E}}_{\alpha\otimes\xi'}[\bar{\mathbb{E}}_{X_n,X'_n}[T^r]] &= \bar{\mathbb{E}}_{\alpha\otimes\xi'}[T^r\circ\theta_n] \\
&\le \bar{\mathbb{E}}_{\alpha\otimes\xi'}[\{\sigma_{\alpha\times\alpha}^{(n+1)}\}^r] \le (n+1)^{r-1}\{\bar{\mathbb{E}}_{\alpha\otimes\xi'}[T^r]+n\bar{\mathbb{E}}_{\alpha\otimes\alpha}[T^r]\}.
\end{aligned}
$$

Plugging this bound into (8.3.8) yields

$$
\bar{\mathbb{E}}_{\xi\otimes\xi'}[T^r\circ\theta_S] \le 2^r\varsigma\mathbb{E}_\xi[\sigma_\alpha^r]\left\{\bar{\mathbb{E}}_{\alpha\otimes\xi'}[T^r]+\bar{\mathbb{E}}_{\alpha\otimes\alpha}[T^r]\right\}\,,
$$

which, combined with (8.3.6), implies

$$
\bar{\mathbb{E}}_{\xi\otimes\xi'}[T^r] \le \varsigma_r\mathbb{E}_\xi[\{\sigma_\alpha^{(m)}\}^r]\left\{\bar{\mathbb{E}}_{\alpha\otimes\xi'}[T^r]+\bar{\mathbb{E}}_{\alpha\otimes\alpha}[T^r]\right\}\,, \qquad (8.3.9)
$$

where $\varsigma_r<\infty$. By interchanging ξ and ξ', we obtain along the same lines

$$
\bar{\mathbb{E}}_{\xi\otimes\xi'}[T^r] \le \varsigma_r\mathbb{E}_{\xi'}[\{\sigma_\alpha^{(m)}\}^r]\left\{\bar{\mathbb{E}}_{\xi\otimes\alpha}[T^r]+\bar{\mathbb{E}}_{\alpha\otimes\alpha}[T^r]\right\}\,,
$$

showing that

$$
\bar{\mathbb{E}}_{\alpha\otimes\xi'}[T^r] \le 2\varsigma_r\mathbb{E}_{\xi'}[\{\sigma_\alpha^{(m)}\}^r]\,\bar{\mathbb{E}}_{\alpha\otimes\alpha}[T^r]\,. \qquad (8.3.10)
$$

Plugging this bound into (8.3.9) and using Lemma 8.3.3 yields that there exists a constant $\kappa_r<\infty$ such that

$$
\bar{\mathbb{E}}_{\xi\otimes\xi'}[T^r] \le \kappa_r\mathbb{E}_\xi[\sigma_\alpha^r]\mathbb{E}_{\xi'}[\sigma_\alpha^r]\bar{\mathbb{E}}_{\alpha\otimes\alpha}[T^r]\,. \qquad (8.3.11)
$$

To prove (8.3.5), it remains to show that $\bar{\mathbb{E}}_{\alpha\otimes\alpha}[T^k]<\infty$. We proceed by induction. The set $\alpha\times\alpha$ is an accessible atom for $\bar{P}$, and $\pi\otimes\pi$ is an invariant probability for $\bar{P}$. Since $\pi\otimes\pi(\alpha\times\alpha)=\{\pi(\alpha)\}^2>0$, the atom $\alpha\times\alpha$ is recurrent by Proposition 6.2.8 and positive by Theorem 6.4.2: $\bar{\mathbb{E}}_{\alpha\otimes\alpha}[T]<\infty$.

Assume now that $\bar{\mathbb{E}}_{\alpha\otimes\alpha}[T^r]<\infty$ for $r\in\mathbb{N}^*$ with $r<k$. Lemma 6.4.3 implies that $\mathbb{E}_\pi[\sigma_\alpha^r]<\infty$. Applying (8.3.11) with $\xi=\xi'=\pi$ shows that

$$
\bar{\mathbb{E}}_{\pi\otimes\pi}[T^r] \le \kappa_r\{\mathbb{E}_\pi[\sigma_\alpha^r]\}^2\,\bar{\mathbb{E}}_{\alpha\otimes\alpha}[T^r].
$$

Applying now Lemma 6.4.3 to the coupling kernel $\bar{P}$ yields $\bar{\mathbb{E}}_{\alpha\otimes\alpha}[T^{r+1}]<\infty$. □

Theorem 8.3.5. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P admits an accessible aperiodic and positive atom α. Denote by π the invariant probability.*

(i) Assume that $\mathbb{E}_\alpha[\sigma_\alpha^k] < \infty$ for some $k \in \mathbb{N}^*$. Then there exists a constant $\varsigma < \infty$ such that for all $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$ and $n \in \mathbb{N}$,

$$n^k \mathrm{d_{TV}}(\xi P^n, \xi' P^n) \le \varsigma \mathbb{E}_\xi[\sigma_\alpha^k] \mathbb{E}_{\xi'}[\sigma_\alpha^k] \,, \tag{8.3.12}$$

$$\sum_{n=1}^{\infty} n^{k-1} \mathrm{d_{TV}}(\xi P^n, \xi' P^n) \le \varsigma \mathbb{E}_\xi[\sigma_\alpha^k] \mathbb{E}_{\xi'}[\sigma_\alpha^k] \,. \tag{8.3.13}$$

(ii) Assume that $\mathbb{E}_\alpha[\sigma_\alpha^{k+1}] < \infty$. Then there exists a constant $\varsigma < \infty$ such that for all $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\mathrm{d_{TV}}(\delta_x P^n, \pi) \le \varsigma \mathbb{E}_\xi[\sigma_\alpha^k] n^{-k} \,. \tag{8.3.14}$$

Proof. Note that if $\mathbb{E}_\alpha[\sigma_\alpha^k] < \infty$, then Lemma 6.4.3 shows that $\mathbb{E}_\pi[\sigma_\alpha^k] < \infty$. The bounds (8.3.12), (8.3.13), and (8.3.14) follow directly from Lemmas 8.3.1 and 8.3.4. □

We now extend these results to geometric convergence.

Lemma 8.3.6 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with an attractive positive atom α satisfying $P(\alpha, \alpha) > 0$. Assume that $\mathbb{E}_\alpha[\beta^{\sigma_\alpha}] < \infty$ for some $\beta > 1$. Then there exist $\delta > 1$ and a constant $\varsigma < \infty$ such that for all ξ, $\xi' \in \mathbb{M}_1(\mathscr{X})$,*

$$\bar{\mathbb{E}}_{\xi \otimes \xi'}[\delta^T] \le \varsigma \mathbb{E}_\xi[\beta^{\sigma_\alpha}] \mathbb{E}_{\xi'}[\beta^{\sigma_\alpha}] \,. \tag{8.3.15}$$

Proof. As in Lemma 8.3.4, we may choose $m \in \mathbb{N}^*$ such that $\sup_{n \in m} \rho_n^{-1} \le \varsigma < \infty$, where $\rho_n = P^n(\alpha, \alpha)$. Set $T = \sigma_{\alpha \times \alpha}$ and $S = \sigma_{\alpha \times \mathsf{X}}^{(m)}$. Lemma 8.2.8 shows that $\sum_{n=1}^\infty \kappa^n |P^n(\alpha, \alpha) - \pi(\alpha)| < \infty$ for some $\kappa > 1$, which implies

$$\sum_{n=1}^{\infty} \kappa^n |P^n(\alpha, \alpha) P^n(\alpha, \alpha) - \pi(\alpha)\pi(\alpha)| < \infty.$$

Applying again Lemma 8.2.8 to the Markov kernel $\bar{P}$ on $\mathsf{X}^2 \times \mathscr{X}^{\otimes 2}$ (noting that $\alpha \times \alpha$ is an atom for $\bar{P}$), we get that there exists $\gamma > 1$ such that

$$\bar{\mathbb{E}}_{\alpha \times \alpha}[\gamma^T] < \infty \,. \tag{8.3.16}$$

We can choose γ such that $\bar{\mathbb{E}}_{\alpha \times \alpha}[\gamma^T] \le \beta$ and $\delta > 1$ such that $\delta^2 \le \beta \wedge \gamma$. Using that $T \le S + T \circ \theta_S$ and $uv \le (1/2)(u^2 + v^2)$, we get

$$\bar{\mathbb{E}}_{\xi \otimes \xi'}[\delta^T] \le \bar{\mathbb{E}}_{\xi \otimes \xi'}[\delta^{S + T \circ \theta_S}] \le \frac{1}{2} \mathbb{E}_\xi[\delta^{2\sigma_\alpha^{(m)}}] + \frac{1}{2} \bar{\mathbb{E}}_{\xi \otimes \xi'}[\delta^{2T \circ \theta_S}]. \tag{8.3.17}$$

We now compute a bound for $\bar{\mathbb{E}}_{\xi\otimes\xi'}[\delta^{2T\circ\theta_S}]$. By the Markov property,

$$\begin{aligned}\bar{\mathbb{E}}_{\xi\otimes\xi'}[\delta^{2T\circ\theta_S}] &= \sum_{n=m}^{\infty} \mathbb{P}_\xi(\sigma_\alpha^{(m)}=n)\bar{\mathbb{E}}_{\xi\otimes\xi'}[\bar{\mathbb{E}}_{\alpha,X'_n}[\delta^{2T}]] \\ &= \sum_{n=m}^{\infty} \mathbb{P}_\xi(\sigma_\alpha^{(m)}=n)\rho_n^{-1}\mathbb{P}_\alpha(X_n\in\alpha)\bar{\mathbb{E}}_{\xi\otimes\xi'}[\bar{\mathbb{E}}_{\alpha,X'_n}[\delta^{2T}]] . \end{aligned} \tag{8.3.18}$$

Note that $\bar{\mathbb{E}}_{\xi\otimes\xi'}[\bar{\mathbb{E}}_{\alpha,X'_n}[\delta^{2T}]] = \int \xi'(\mathrm{d}x')\bar{\mathbb{E}}_{\alpha,x'}[\delta^{2T}]$ does not depend on the initial distribution $\xi\in\mathbb{M}_1(\mathscr{X})$. Hence $\bar{\mathbb{E}}_{\xi\otimes\xi'}[\bar{\mathbb{E}}_{\alpha,X'_n}[\delta^{2T}]] = \bar{\mathbb{E}}_{\alpha\otimes\xi'}[\bar{\mathbb{E}}_{\alpha,X'_n}[\delta^{2T}]]$. Plugging this expression into (8.3.18) yields

$$\begin{aligned}\bar{\mathbb{E}}_{\xi\otimes\xi'}[\delta^{2T\circ\theta_S}] &= \sum_{n=m}^{\infty} \mathbb{P}_\xi(\sigma_\alpha^{(m)}=n)\rho_n^{-1}\bar{\mathbb{E}}_{\alpha\otimes\xi'}[\mathbb{1}_\alpha(X_n)\bar{\mathbb{E}}_{\alpha,X'_n}[\delta^{2T}]] \\ &\le \sum_{n=m}^{\infty} \mathbb{P}_\xi(\sigma_\alpha^{(m)}=n)\rho_n^{-1}\bar{\mathbb{E}}_{\alpha\otimes\xi'}[\bar{\mathbb{E}}_{X_n,X'_n}[\delta^{2T}]]. \end{aligned} \tag{8.3.19}$$

The Markov property implies $\bar{\mathbb{E}}_{\alpha\otimes\xi'}[\bar{\mathbb{E}}_{X_n,X'_n}[\delta^{2T}]] = \bar{\mathbb{E}}_{\alpha\otimes\xi'}[\delta^{2T}\circ\theta_n]$. Since $\sigma_{\alpha\times\alpha}\ge 1$, we get $T\circ\theta_n \le \sigma_{\alpha\times\alpha}^{(n+1)}$. Using $\sigma_{\alpha\times\alpha}^{(n+1)} = \sigma_{\alpha\times\alpha}^{(n)} + T\circ\theta_{\sigma_{\alpha\times\alpha}^{(n)}}$ recursively, we finally obtain

$$\bar{\mathbb{E}}_{\alpha\otimes\xi'}[\bar{\mathbb{E}}_{X_n,X'_n}[\delta^{2T}]] \le \bar{\mathbb{E}}_{\alpha\otimes\xi'}[\delta^{2T}][\bar{\mathbb{E}}_{\alpha\otimes\alpha}[\delta^{2T}]]^n \le \bar{\mathbb{E}}_{\alpha\otimes\xi'}[\delta^{2T}]\beta^n .$$

Plugging this relation into (8.3.19) and using (8.3.17) yields

$$\bar{\mathbb{E}}_{\xi\otimes\xi'}[\delta^T] \le (1/2)\left\{\mathbb{E}_\xi[\beta^{\sigma_\alpha^{(m)}}] + \varsigma\mathbb{E}_\xi[\beta^{\sigma_\alpha^{(m)}}]\bar{\mathbb{E}}_{\alpha\otimes\xi'}[\gamma^T]\right\} . \tag{8.3.20}$$

By interchanging ξ and ξ' we obtain similarly

$$\bar{\mathbb{E}}_{\xi\otimes\xi'}[\delta^T] \le (1/2)\left\{\mathbb{E}_{\xi'}[\beta^{\sigma_\alpha^{(m)}}] + \varsigma\mathbb{E}_{\xi'}[\beta^{\sigma_\alpha^{(m)}}]\bar{\mathbb{E}}_{\xi\otimes\alpha}[\gamma^T]\right\} . \tag{8.3.21}$$

Setting $\xi=\delta_\alpha$ in (8.3.21) implies that

$$\bar{\mathbb{E}}_{\alpha\otimes\xi'}[\delta^T] \le 1/2\left\{\mathbb{E}_{\xi'}[\beta^{\sigma_\alpha^{(m)}}] + \varsigma\mathbb{E}_{\xi'}[\beta^{\sigma_\alpha^{(m)}}]\bar{\mathbb{E}}_{\alpha\otimes\alpha}[\gamma^T]\right\} .$$

Note that $\mathbb{E}_\xi[\beta^{\sigma_\alpha^{(m)}}] \le \{\mathbb{E}_\alpha[\beta^{\sigma_\alpha}]\}^{m-1}\mathbb{E}_\xi[\beta^{\sigma_\alpha}]$. The proof is concluded by plugging this relation into (8.3.20) and then (8.3.16). $\square$

As an immediate consequence of Lemmas 8.3.1 and 8.3.6, we obtain the following result.

Theorem 8.3.7. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ that admits an accessible aperiodic atom α and $\beta > 1$ such that $\mathbb{E}_\alpha[\beta^{\sigma_\alpha}] < \infty$. Then P has a unique invariant distribution π, and there exist $\delta \in (1, \beta]$ and $\varsigma < \infty$ such that for all $\xi \in \mathbb{M}_1(\mathscr{X})$,*

$$\sum_{n=1}^{\infty} \delta^n \mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) \le \varsigma \mathbb{E}_\xi[\delta^{\sigma_\alpha}] . \tag{8.3.22}$$

8.4 Exercises

8.1. Consider a sequence of independent Bernoulli trials with success probability p. A renewal V_n occurs at time $n \in \mathbb{N}$ if a success occurs. Show using (8.1.10) that the waiting time distribution is geometric with mean $1/p$.

8.2. Consider a zero-delayed renewal process, i.e., $S_0 = Y_0 = 0$, and define the sequence of random times by $\{\eta_k,\ k \in \mathbb{N}\}$,

$$\eta_k = \sup\{n \in \mathbb{N} :\ S_n \le k\} . \tag{8.4.1}$$

That is, η_k is the last renewal time before time k. Note that η_k is not a stopping time. There is a simple relation between η_k and ρ_k defined in (8.1.13):

$$\eta_k = \sup\{n \in \mathbb{N} :\ S_n \le k\} = \inf\{n \in \mathbb{N} :\ S_n > k\} - 1 = \rho_k - 1 .$$

The backward recurrence time chain (also called the age process) $\{B_k,\ k \in \mathbb{N}\}$ is defined for $k \in \mathbb{N}$ by

$$B_k = k - S_{\eta_k} . \tag{8.4.2}$$

The total lifetime is the sum of the residual lifetime A_k and the age B_k:

$$C_k = S_{\rho_k} - k + k - S_{\eta_k} = S_{\eta_k+1} - S_{\eta_k} = Y_{\eta_k} ,$$

which is the total duration of the current renewal interval.

Show that the backward recurrence time chain $\{B_k,\ k \in \mathbb{N}\}$ is a nonnegative-integer-valued Markov chain. Determine its Markov kernel.

8.3. We use the notation and definitions of Exercise 8.2. Show that the kernel R is strongly irreducible and recurrent on $\{0, \ldots, \sup\{n \in \mathbb{N} :\ b(n) \neq 0\}\}$. Assume that the mean waiting time is finite: $m = \sum_{j=1}^{\infty} j b(j) < \infty$. Show that R is positive recurrent and admits an invariant probability measure $\bar{\pi}$ on $\mathbb{N}$ defined by

$$\bar{\pi}(j) = m^{-1}\mathbb{P}_0(Y_1 > j) = m^{-1}\sum_{\ell=j+1}^{\infty} b(\ell)\,, \quad j \geq 0\,.$$

8.4. This exercise provides an analytical proof of Blackwell's theorem.

1. Set $L = \limsup_n u(n)$ and let $\{n_k,\ k \in \mathbb{N}\}$ be a subsequence that converges to L. Show that there exists a sequence $\{q(j),\ j \in \mathbb{Z}\}$ such that

$$\lim_{k\to\infty} u(n_k + j)\mathbb{1}_{\{j \geq -n_k\}} = q(j)\,,$$

for all $j \in \mathbb{Z}$.
2. Show that $q(p) = \sum_{j=1}^{\infty} b(j)q(p-j)$.
3. Set $S = \{n \geq 1 \,:\, b(n) > 0\}$. Show that $q(-p) = L$ for all $p \in S$.
4. Show that $q(-p) = L$ if $p = p_1 + \cdots + p_n$ with $p_i \in S$ for $i = 1,\ldots,n$.
5. Show that $q(j) = L$ for all $j \in \mathbb{Z}$.
6. Set $\bar{b}(j) = \sum_{i=j+1}^{\infty} b(i)$, so that $\bar{b}(0) = 1$, $b(j) = \bar{b}(j-1) - \bar{b}(j)$, $j \geq 1$, and $\sum_{j=0}^{\infty} \bar{b}(j) = m$. Show that for all $n \geq 1$,

$$\sum_{j=0}^{n} \bar{b}(j)u(n-j) = \sum_{j=0}^{n-1} \bar{b}(j)u(n-1-j).$$

7. Show that for all $k \geq 0$,

$$\sum_{j=0}^{\infty} \bar{b}(j)u(n_k - j)\mathbb{1}_{\{j \leq n_k\}} = 1\,. \tag{8.4.3}$$

8. If $m = \infty$, show that $L = 0$.
9. If $m < \infty$, show that $\limsup_{n\to\infty} u(n) = \liminf_{n\to\infty} u(n) = 1/m$.
10. Conclude.

8.5. Consider a recurrent irreducible aperiodic Markov kernel P over a discrete state space X. Fix one arbitrary state $\mathsf{a} \in \mathsf{X}$ and set, for $n \geq 0$ and $x \in \mathsf{X}$,

$$b(n) = \mathbb{P}_{\mathsf{a}}(\sigma_{\mathsf{a}} = n)\,, \quad a_x(n) = \mathbb{P}_x(\sigma_{\mathsf{a}} = n)\,, \tag{8.4.4}$$

$$u(n) = \mathbb{P}_{\mathsf{a}}(X_n = \mathsf{a})\,, \quad v_{a_x}(n) = a_x * u(n) = \mathbb{P}_x(X_n = \mathsf{a})\,. \tag{8.4.5}$$

Show that u defined in (8.4.5) is the pure renewal sequence associated with b considered as a waiting time distribution and v_{a_x} is the delayed renewal sequence associated with the delay distribution a_x.

8.6. The bounds obtained in (8.2.6) can be used to obtain an alternative proof of Theorem 7.6.4 for aperiodic and positive recurrent Markov kernels.

Let P be an irreducible aperiodic positive recurrent Markov kernel on a discrete state space X and let π be its invariant probability. Show that for all $x \in \mathsf{X}$,

$$\lim_{n\to\infty} \mathrm{d_{TV}}(\delta_x P^n, \pi) = 0\,.$$

8.7. Let P be a Markov kernel on a finite space X. Assume that P is strongly irreducible.

1. Show that there exist a finite integer r and $\varepsilon > 0$ such that for all $x, y \in \mathsf{X}$, $\mathbb{P}_x(\sigma_y \leq r) \geq \varepsilon$.
2. Show that $\mathbb{P}_x(\sigma_y > kr) \leq (1-\varepsilon)^k$.
3. Show that there exists $b > 1$ such that for all $x, y \in \mathsf{X} \times \mathsf{X}$, $\mathbb{E}_x[b^{\sigma_y}] < \infty$.

8.8. Let P be a Markov kernel on a discrete state space X. Let $C \subset \mathsf{X}$ be a finite set. Assume that P is strongly irreducible and that there exists $\beta > 1$ such that $\sup_{x \in C} \mathbb{E}_x[\beta^{\sigma_C}] < \infty$.

1. Set $\nu_x = \inf\left\{n \geq 1 \,:\, X_{\sigma_C^{(n)}} = x\right\}$. Show that there exists $r > 1$ such that $\mathbb{E}_x[r^{\nu_x}] < \infty$ for all $x \in C$. [Hint: Consider the induced Markov chain $\left\{X_{\sigma_C^{(n)}} \,:\, n \in \mathbb{N}\right\}$ on C and apply Exercise 8.7.]
2. Choose $s > 0$ such that $M^s \leq \beta^{1/2}$. Show that
$$\mathbb{P}_x(\sigma_x \geq n) \leq \left(\sup_{x \in C} \mathbb{E}_x[r^{\nu_x}]\right) r^{-sn} + (\sqrt{\beta})^{-n}.$$
3. Show that there exists $\delta > 1$ such that $\mathbb{E}_x[\delta^{\sigma_x}] < \infty$ for all $x \in C$.

8.5 Bibliographical Notes

The basic facts on renewal theory can be found in Feller (1971) and Cox (1962). Blackwell's theorem (Theorem 8.1.7) was first proved by Blackwell (1948). Several proofs were proposed, most of them not probabilistic. The simple coupling proof presented here is due to Lindvall (1977) (see also Thorisson (1987)).

Kendall's theorem (Theorem 8.1.9) was first established in Kendall (1959). The proof given here closely follows the original derivation. One weakness of this proof is that it does not provide a quantitative estimate of the rate of convergence. Improvements to Kendall's theorem were proposed in Meyn and Tweedie (1994) and later by Baxendale (2005). Sharp estimates were introduced in (Bednorz 2013, Theorem 2.8).

The proof of convergence using the first-entrance last-exit decomposition presented in Section 8.2.1 was introduced in Nummelin (1978) and refined in Nummelin and Tweedie (1978). Our presentation follows closely (Meyn and Tweedie 2009, Chapter 13).

Chapter 9
Small Sets, Irreducibility, and Aperiodicity

So far, we have considered only atomic and discrete Markov chains. When the state space is not discrete, many Markov chains do not admit accessible atoms. Recall that a set C is an atom if each time the chain visits C, it regenerates, i.e., it leaves C under a probability distribution that is constant over C. If the state space does not possess an atom, we may require instead that the chain restart anew from C with some fixed probability (strictly less than one) that is constant over C. Then this property is satisfied by many more Markov chains. Such sets will be called small sets. The purpose of this chapter is to provide the first basic properties of Markov kernels that admit accessible small sets.

9.1 Small Sets

Definition 9.1.1 (Small Set) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. A set $C \in \mathscr{X}$ is called a small set if there exist $m \in \mathbb{N}^*$ and a nonzero measure $\mu \in \mathbb{M}_+(\mathscr{X})$ such that for all $x \in C$ and $A \in \mathscr{X}$,*

$$P^m(x,A) \geq \mu(A) . \tag{9.1.1}$$

The set C is then said to be an (m,μ)-small set.

The definition entails that μ is a finite measure and $0 < \mu(\mathsf{X}) \leq 1$. Hence it can be written $\mu = \varepsilon\nu$ with $\varepsilon = \mu(\mathsf{X})$, and ν is a probability measure. If $\varepsilon = 1$, then equality must hold in (9.1.1), and thus C is an atom. Hereinafter, when we write "C is an $(m, \varepsilon\nu)$-small set," it will be always assumed that $\varepsilon \in (0,1]$ and ν is a probability measure. When we do not need to mention the associated measure, we may simply write "C is an m-small set." As we did for atoms, we further define certain specific properties of small sets.

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_9

Definition 9.1.2 *An (m,μ)-small set C is said to be*

- *strongly aperiodic if $m = 1$ and $\mu(C) > 0$;*
- *positive if $\mathbb{E}_x[\sigma_C] < \infty$ for all $x \in C$.*

Example 9.1.3. An atom is a 1-small set. A small set is not necessarily strongly aperiodic. Consider, for instance, the forward recurrence chain introduced in Section 8.1.1. The state $\{1\}$ is an atom, and every finite subset of integers C is small. However, if the waiting time distribution b puts zero mass on C, then C is not strongly aperiodic. ◀

Recall from Definition 3.5.1 that a set A is said to be accessible if $\mathbb{P}_x(\sigma_A < \infty) > 0$ for all $x \in \mathsf{X}$. The set of accessible sets is denoted by $\mathscr{X}_P^+$.

Example 9.1.4. Consider the scalar autoregressive AR(1) model $X_k = \alpha X_{k-1} + Z_k$, $k \geq 1$, where $\{Z_k,\ k \in \mathbb{N}^*\}$ is an i.i.d. sequence, independent of X_0 and $\alpha \in \mathbb{R}$. Assume that the distribution of the innovation has a continuous everywhere positive density f with respect to Lebesgue measure. Let $C \subset \mathbb{R}$ be a compact set such that $\mathrm{Leb}(C) > 0$. Then for all Borel sets A and $x \in C$, we get

$$
\begin{aligned}
P(x,A) &= \int_A f(y-\alpha x)\mathrm{d}y \\
&\geq \int_{A\cap C} f(y-\alpha x)\mathrm{d}y \geq \inf_{(x,y)\in C\times C} f(y-\alpha x)\mathrm{Leb}(A\cap C)\ .
\end{aligned}
$$

This shows that C is a small set. Of course, this set is accessible, since for all $x \in \mathbb{R}$,

$$
P(x,C) = \int_C f(y-\alpha x)\mathrm{d}y > 0\ .
$$

◀

Example 9.1.5. We can generalize Example 9.1.4. Let P be a Markov kernel on $\mathbb{R}^d \times \mathscr{B}\left(\mathbb{R}^d\right)$ such that

$$
P(x,A) \geq \int_A q(x,y)\mathrm{Leb}_d(\mathrm{d}y)\ , \qquad A \in \mathscr{B}\left(\mathbb{R}^d\right)\ ,
$$

where q is a positive lower semicontinuous function on $\mathbb{R}^d \times \mathbb{R}^d$. Then every compact set C with positive Lebesgue measure is small. Indeed, for $x \in C$ and $A \in \mathscr{B}\left(\mathbb{R}^d\right)$,

$$
\begin{aligned}
P(x,A) &\geq \int_A q(x,y)\,\text{`Leb}_d(\mathrm{d}y) \\
&\geq \int_{A\cap C} q(x,y)\mathrm{Leb}_d(\mathrm{d}y) \geq \inf_{(x,y)\in C\times C} q(x,y)\mathrm{Leb}_d(A\cap C)\,.
\end{aligned}
$$

This proves that C is an $\varepsilon\nu$-small set with $\varepsilon = \inf_{(x,y)\in C\times C} q(x,y)$ and $\nu = \mathrm{Leb}_d(\cdot\cap C)$. Furthermore, C is accessible, since $P(x,C) = \int_C q(x,y)\mathrm{Leb}_d(\mathrm{d}y) > 0$ for all $x \in \mathbb{R}^d$. Such kernels will be further investigated in Chapter 12. ◀

Lemma 9.1.6 *If C is an accessible (m,μ)-small set, then there exist $m' \geq m$ and $\mu' \in \mathbb{M}_+(\mathscr{X})$ such that C is an (m',μ')-small set and $\mu'(C) > 0$.*

Proof. Since $C \in \mathscr{X}_P^+$, Lemma 3.5.2 (iii) shows that there exists $n \in \mathbb{N}^*$ such that $\mu P^n(C) > 0$. Since C is an (m,μ)-small set, this yields, for every $A \in \mathscr{X}$ and $x \in C$,

$$
P^{m+n}(x,A) = \int_{\mathsf{X}} P^m(x,\mathrm{d}y)P^n(y,A) \geq \int_{\mathsf{X}} \mu(\mathrm{d}y)P^n(y,A) = \mu P^n(A)\,.
$$

This proves that C is an (m',μ')-small set with $m' = m+n$ and $\mu' = \mu P^n$. Moreover, $\mu'(C) = \mu P^n(C) > 0$. □

The following lemma will be very useful. It states formally the idea that a small set leads uniformly to every accessible set and that a set that leads uniformly to a small set is also a small set.

Lemma 9.1.7 *Let C be an (m,μ)-small set.*

(i) For every $A \in \mathscr{X}_P^+$, there exists an integer $q \geq m$ such that $\inf_{x\in C} P^q(x,A) > 0$.
(ii) Let $D \in \mathscr{X}$. If there exists $n \geq 1$ such that $\inf_{x\in D} P^n(x,C) \geq \delta$, then D is an $(n+m,\delta\mu)$-small set.

Proof. (i) Since $A \in \mathscr{X}_P^+$, by Lemma 3.5.2, there exists $n\geq 1$ such that $\mu P^n(A)>0$. Thus for $x \in C$, we get

$$
P^{m+n}(x,A) = \int_{\mathsf{X}} P^m(x,\mathrm{d}y)P^n(y,A) \geq \int_{\mathsf{X}} \mu(\mathrm{d}y)P^n(y,A) = \mu P^n(A) > 0\,.
$$

(ii) For $x \in D$ and $A \in \mathscr{X}$,

$$
\begin{aligned}
P^{n+m}(x,A) &= \int_{\mathsf{X}} P^n(x,\mathrm{d}y)P^m(y,A) \geq \int_C P^n(x,\mathrm{d}y)P^m(y,A) \geq \int_C P^n(x,\mathrm{d}y)\mu(A) \\
&= P^n(x,C)\mu(A) \geq \delta\mu(A)\,.
\end{aligned}
$$

□

Proposition 9.1.8 *If there exists an accessible small set, then X is a countable union of small sets.*

Proof. Let C be an accessible small set and for $n,m \geq 1$, define

$$C_{n,m} = \left\{x \in \mathsf{X} : P^n(x, C) \geq m^{-1}\right\} .$$

Since C is accessible, for every $x \in \mathsf{X}$, there exists $n \in \mathbb{N}^*$ such that $P^n(x,C) > 0$; thus $\mathsf{X} = \bigcup_{n,m\geq 1} C_{n,m}$. Moreover, each set $C_{n,m}$ is small, because by construction, $C_{n,m}$ yields uniformly to the small set C (see Lemma 9.1.7 (ii)). □

The following result is extremely important. It gives a convenient criterion for checking the accessibility of a set, expressed in terms of the minorization measure of an accessible small set.

Proposition 9.1.9 *Assume that C is an accessible (m,μ)-small set. Then for every $A \in \mathscr{X}$, $\mu(A) > 0$ implies that A is accessible.*

Proof. Since C is accessible, for every $x \in \mathsf{X}$, there exists $n \geq 1$ such that $P^n(x,C) > 0$. If $\mu(A) > 0$, then

$$P^{n+m}(x,A) \geq \int_C P^n(x,\mathrm{d}x')P^m(x',A) \geq P^n(x,C)\mu(A) > 0 .$$

Thus A is accessible. □

9.2 Irreducibility

Mimicking Definition 7.1.1, we now introduce the general definition of irreducible kernel, in which accessible small sets replace accessible states.

Definition 9.2.1 (Irreducible kernel) *A Markov kernel P on $\mathsf{X} \times \mathscr{X}$ is said to be irreducible if it admits an accessible small set.*

Although seemingly weak, the assumption of irreducibility has some important consequences. The definition guarantees that a small set is always reached by a chain with some positive probability from any starting point.

We are now going to see that there exists an equivalent characterization of irreducibility in terms of measures. There actually exist many measures other than those introduced in Proposition 9.1.9 that provide a sufficient condition for accessibility and possibly also a necessary condition. We will therefore introduce the following definition.

Definition 9.2.2 (Irreducibility measure) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let $\phi \in \mathbb{M}_+(\mathscr{X})$ be a nontrivial σ-finite measure.*

- *ϕ is said to be an irreducibility measure if $\phi(A) > 0$ implies $A \in \mathscr{X}_P^+$.*
- *ϕ is said to be a maximal irreducibility measure if ϕ is an irreducibility measure and $A \in \mathscr{X}_P^+$ implies $\phi(A) > 0$.*

Remark 9.2.3. Since every irreducibility measure ϕ is σ-finite by definition, we can assume when needed that ϕ is a probability measure. Indeed, let $\{A_n,\ n \in \mathbb{N}^*\}$ be a measurable partition of X such that $0 < \phi(A_n) < \infty$ for all $n \geq 1$. Then ϕ is equivalent to the probability measure ϕ' defined for $A \in \mathscr{X}$ by

$$\phi'(A) = \sum_{n=1}^{\infty} 2^{-n} \frac{\phi(A \cap A_n)}{\phi(A_n)} .$$

Proposition 9.1.9 can now be rephrased in the language of irreducibility: the minorizing measure of an accessible small set is an irreducibility measure. The following result shows that maximal irreducibility measures exist and are all equivalent.

Theorem 9.2.4. *If ϕ is an irreducibility measure, then for every $\varepsilon > 0$, ϕK_{a_ε} is a maximal irreducibility measure. All irreducibility measures are absolutely continuous with respect to every maximal irreducibility measure, and all maximal irreducibility measures are equivalent.*

Proof. Set $\psi = \phi K_{a_\varepsilon}$. If A is accessible, then for every $\varepsilon \in (0,1)$ and for all $x \in \mathsf{X}$, $K_{a_\varepsilon}(x,A) > 0$. This implies $\psi(A) = \phi K_{a_\varepsilon}(A) > 0$. Consider now the converse. Let $A \in \mathscr{X}$ be such that $\psi(A) = \phi K_{a_\varepsilon}(A) > 0$. Define

$$\bar{A} = \{x \in \mathsf{X} : \mathbb{P}_x(\tau_A < \infty) > 0\} = \{x \in \mathsf{X} : K_{a_\varepsilon}(x,A) > 0\} . \tag{9.2.1}$$

Then by definition of $\bar{A}$, we have

$$0 < \psi(A) = \int_{\bar{A}} \phi(\mathrm{d}x) K_{a_\varepsilon}(x,A) .$$

Hence $\phi(\bar{A}) > 0$, and thus $\bar{A}$ is accessible, since ϕ is an irreducibility measure. The strong Markov property implies that for all $x \in \mathsf{X}$,

$$\mathbb{P}_x(\sigma_A < \infty) \geq \mathbb{P}_x(\sigma_{\bar{A}} < \infty, \tau_A \circ \theta_{\sigma_{\bar{A}}} < \infty) = \mathbb{E}_x[\mathbb{1}_{\{\sigma_{\bar{A}} < \infty\}} \mathbb{P}_{X_{\sigma_{\bar{A}}}}(\tau_A < \infty)] > 0 ,$$

showing that A is also accessible.

To prove the second statement, let ψ' be an irreducibility measure and ψ a maximal irreducibility measure. If $\psi(A) = 0$, then A is not accessible, which implies

$\psi'(A) = 0$ by definition. Therefore, ψ' is absolutely continuous with respect to ψ. This completes the proof. □

Theorem 9.2.5. *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Then every accessible set contains an accessible $(m, \varepsilon\nu)$-small set C with $\nu(C) > 0$.*

Proof. Since P is irreducible, there exists an accessible $(n, \varepsilon\nu)$-small set C. Let ψ be a maximal irreducibility measure and A an accessible set. For $p, q \in \mathbb{N}^*$, write

$$A_{p,q} = \left\{x \in A : \mathbb{P}_x(\sigma_C = p) \geq q^{-1}\right\} .$$

Since A is accessible, $A = \bigcup_{p,q\in\mathbb{N}^*} A_{p,q}$, and there exist p, q such that $\psi(A_{p,q}) > 0$. For all $x \in A_{p,q}$ and $B \in \mathscr{X}$, we get

$$P^{n+p}(x, B) \geq \mathbb{P}_x(\sigma_C = p, X_n \circ \theta_p \in B) = \mathbb{E}_x\left[\mathbb{1}_{\{\sigma_C = p\}} \mathbb{P}_{X_p}(X_n \in B)\right] \geq q^{-1}\varepsilon\nu(B) ,$$

showing that $A_{p,q}$ is an $(n+p, \varepsilon\nu)$-small set. We conclude by invoking Lemma 9.1.6. □

We have seen that the minorizing measure of a small set is an irreducibility measure. We now prove the converse: if a Markov kernel P admits an irreducibility measure, then it admits an accessible small set. Therefore, irreducibility can be defined equivalently by the existence of a small set or an irreducibility measure.

Theorem 9.2.6. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume in addition that the σ-algebra $\mathscr{X}$ is countably generated. The Markov kernel P is irreducible if and only if it admits an irreducibility measure.*

Proof. The proof is postponed to Section 9.A and may be omitted on a first reading. □

Example 9.2.7 (Example 9.1.4 continued). Consider the scalar AR(1) model $X_k = \alpha X_{k-1} + Z_k$, $k \geq 1$, where $\{Z_k,\ k \in \mathbb{N}^*\}$ is an i.i.d. sequence, independent of X_0 and $\alpha \in \mathbb{R}$. Assume that the distributions of the innovation has a density that is positive in a neighborhood of zero. Assume for simplicity that Z_1 is uniform on $[-1, 1]$.

• If $|\alpha| < 1$, then the restriction of Lebesgue measure to $[-1/2, 1/2]$ is an irreducibility measure. Indeed, for $B \subset [-1/2, 1/2]$ and $x \in [-1/2, 1/2]$,

$$P(x, B) = \frac{1}{2}\int_B \mathbb{1}_{[-1,1]}(y - \alpha x)\mathrm{d}y = \frac{1}{2}\mathrm{Leb}(B) . \tag{9.2.2}$$

This proves that every B such that $\mathrm{Leb}(B) > 0$ is accessible from $[-1/2, 1/2]$. To check accessibility from an arbitrary x, note that $X_n = \alpha^n x + \sum_{j=0}^{n-1} \alpha^j Z_{n-j}$. For $M > 0$, if $\max_{1\leq j\leq n} |Z_j| \leq M$, then

$$\sum_{j=0}^{n-1} |\alpha|^j |Z_{n-j}| \leq M/(1-|\alpha|) .$$

Taking $M = (1-|\alpha|)/4$, we obtain

$$\begin{aligned}\mathbb{P}_x(X_n \in [\alpha^n x - 1/4, \alpha^n x + 1/4]) &\geq \{\mathbb{P}(Z_1 \in [-(1-|\alpha|)/4, (1-|\alpha|)/4])\}^n \\ &\geq \{(1-|\alpha|)/4\}^n .\end{aligned}$$

Thus for $n(x)$ such that $|\alpha|^{n(x)}|x| \leq 1/4$, this yields $\mathbb{P}_x(X_n \in [-1/2, 1/2]) > 0$. This proves that $[-1/2, 1/2]$ is accessible. Together with (9.2.2), this proves that every set $B \subset [-1/2, 1/2]$ with positive Lebesgue measure is accessible. Thus the chain is irreducible.

• Assume now that $|\alpha| > 1$. Then $|X_n| \geq |\alpha|^n|x| - |\alpha|^{n+1}/(|\alpha|-1)$, and we obtain, for every $k > 0$, that if $x > (k+1)|\alpha|/(|\alpha|-1)$, then $|X_n| > k|\alpha|/(|\alpha|-1) > k$ for all $n \geq 0$, and thus $[-k, k]$ is not accessible. This proves that the chain is not irreducible. ◀

Theorem 9.2.5 shows that if a chain is irreducible, then there exists a maximal irreducibility measure ψ such that $\mathscr{X}_P^+ = \{A \in \mathscr{X} : \psi(A) > 0\}$. The next result shows that a set that is not accessible ($\psi(A) = 0$) is avoided from ψ-almost every starting point. It is essential here, of course, to take for ψ a maximal irreducibility measure; it is of course no longer true for an irreducibility measure. Indeed, every nontrivial restriction of an irreducibility measure is still an irreducibility measure.

Proposition 9.2.8 *Let P be an irreducible kernel on $\mathsf{X} \times \mathscr{X}$. A set $A \in \mathscr{X}$ is not accessible for P if and only if the set $\{x \in \mathsf{X} : \mathbb{P}_x(\tau_A < \infty) > 0\}$ is not accessible for P.*

Proof. For every $\varepsilon \in (0,1)$, we have

$$\{x \in \mathsf{X} : \mathbb{P}_x(\tau_A < \infty) > 0\} = \{x \in \mathsf{X} : K_{a_\varepsilon}(x, A) > 0\} .$$

Let ψ be a maximal irreducibility measure. Then ψK_{a_ε} is also a maximal irreducibility measure and is therefore equivalent to ψ. Hence $\psi(A) = 0$ if and only if $\psi K_{a_\varepsilon}(A) = 0$, showing that $\psi(\{x \in \mathsf{X} : K_{a_\varepsilon}(x, A) > 0\}) = 0$. Since ψ is maximal, the set $\{x \in \mathsf{X} : K_{a_\varepsilon}(x, A) > 0\}$ is not accessible. □

Proposition 9.2.9 *Let P be an irreducible kernel on $\mathsf{X} \times \mathscr{X}$. If $A \in \mathscr{X}$ and $A \notin \mathscr{X}_P^+$, then $A^c \in \mathscr{X}_P^+$. A countable union of inaccessible sets is not accessible.*

Proof. Let ψ be a maximal irreducibility measure. If $A \in \mathscr{X}$, then either $\psi(A) > 0$ or $\psi(A^c) > 0$, which means that at least one of A and A^c is accessible. If $\{A_n,\ n \in \mathbb{N}\}$ is a countable union of inaccessible sets, then $\psi(A_n) = 0$ for all $n \geq 0$, and thus $\psi(\cup_{n\geq 0} A_n) = 0$. □

Remark 9.2.10. This provides a criterion for nonirreducibility: if there exists $A \in \mathscr{X}$ such that neither A nor A^c is accessible, then P is not irreducible. In particular, if there exist two disjoint absorbing sets, then the chain is not irreducible. ▲

Definition 9.2.11 (Full set) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. A set $F \in \mathscr{X}$ is said to be full if F^c is not accessible.*

If P is an irreducible kernel, then a set F is full if $\psi(F^c) = 0$ for every maximal irreducibility measure ψ. A full set is nearly the same thing as an absorbing set.

Proposition 9.2.12 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Then every nonempty absorbing set is full, and every full set contains an absorbing full set.*

Proof. The first statement is obvious: if A is an absorbing set, then by definition, its complement is inaccessible, and thus A is full. Now let A be a full set and define

$$C = \{x \in A : \mathbb{P}_x(\sigma_{A^c} = \infty) = 1\} = \{x \in \mathsf{X} : \mathbb{P}_x(\tau_{A^c} = \infty) = 1\} .$$

Note first that the set C is not empty, since otherwise, A^c would be accessible from A, i.e., $\mathbb{P}_x(\sigma_{A^c} < \infty) > 0$ for all $x \in A$, and this would imply that A^c was accessible by Lemma 3.5.2, which contradicts the assumption that A is full. For $x \in C$, applying the Markov property, we obtain

$$\begin{aligned} 1 = \mathbb{P}_x(\sigma_{A^c} = \infty) &= \mathbb{P}_x(\tau_{A^c} \circ \theta_1 = \infty) \\ &= \mathbb{E}_x[\mathbb{1}_{C^c}(X_1)\mathbb{P}_{X_1}(\tau_{A^c} = \infty)] + \mathbb{E}_x[\mathbb{1}_C(X_1)\mathbb{P}_{X_1}(\tau_{A^c} = \infty)] \\ &= \mathbb{E}_x[\mathbb{1}_{C^c}(X_1)\mathbb{P}_{X_1}(\tau_{A^c} = \infty)] + \mathbb{P}_x(X_1 \in C) . \end{aligned}$$

If $x \in C^c$, then $\mathbb{P}_x(\tau_{A^c} = \infty) < 1$, and thus the previous identity implies that $\mathbb{P}_x(X_1 \in C^c) = 0$, since otherwise, the sum of the two terms would be strictly less than 1. Thus C is absorbing and hence full by the first statement. □

Proposition 9.2.13. *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that there exist two measurable functions $V_0, V_1 : \mathsf{X} \to [0,\infty]$ satisfying*

(i) $V_0(x) = \infty \;\Rightarrow\; V_1(x) = \infty$,
(ii) $V_0(x) < \infty \;\Rightarrow\; PV_1(x) < \infty$.

Then the set $\{V_0 < \infty\}$ *is either empty or full and absorbing. If* $\{V_0 < \infty\} \neq \emptyset$, *then there exists* $n_0 \in \mathbb{N}$ *such that* $\{V_0 \leq n_0\}$ *is accessible.*

Proof. Assume that the set $S = \{V_0 < \infty\}$ is not empty. Note that

$$S^c = \{x \in \mathsf{X} : V_0(x) = \infty\} \subset \{x \in \mathsf{X} : V_1(x) = \infty\} .$$

For all $x \in S$, we get $\mathbb{E}_x[\mathbb{1}_{S^c}(X_1)V_1(X_1)] \leq \mathbb{E}_x[V_1(X_1)] = PV_1(x) < \infty$. Therefore, the set S is absorbing and hence full by Proposition 9.2.12. Now, since S is full and P is irreducible, Proposition 9.2.9 implies that $S \in \mathscr{X}_P^+$. Combining this with $S = \cup_{n \in \mathbb{N}}\{V_0 \leq n\}$ and applying again Proposition 9.2.9, we get the last statement of the proposition. □

Corollary 9.2.14 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let r be a positive increasing sequence such that* $\lim_{n\to\infty} r(n) = \infty$ *and* $A \in \mathscr{X}$, $A \neq \emptyset$. *Assume that* $\sup_{x \in A} \mathbb{E}_x[r(\sigma_A)] < \infty$. *Then the set*

$$\{x \in \mathsf{X} : \mathbb{E}_x[r(\sigma_A)] < \infty\}$$

is full and absorbing, and A is accessible.

Proof. Set $W(x) = \mathbb{E}_x[r(\sigma_A)]$ (with the convention $r(\infty) = \infty$). On the event $\{X_1 \notin A\}$, the relation $\sigma_A = 1 + \sigma_A \circ \theta_1 \geq \sigma_A \circ \theta_1$ holds, whence

$$\begin{aligned} PW(x) &= \mathbb{E}_x[\mathbb{1}_A(X_1)\mathbb{E}_{X_1}[r(\sigma_A)]] + \mathbb{E}_x[\mathbb{1}_{A^c}(X_1)r(\sigma_A \circ \theta_1)] \\ &\leq M + \mathbb{E}_x[r(\sigma_A)] \leq M + W(x) , \end{aligned}$$

where $M = \sup_{x \in A} \mathbb{E}_x[r(\sigma_A)]$. Applying Proposition 9.2.13 with $V_0 = V_1 = W$ shows that $S = \{W < \infty\}$ is full absorbing. For all $x \in S$, since $\mathbb{E}_x[r(\sigma_A)] < \infty$, we have $\mathbb{P}_x(\sigma_A < \infty) = 1$. Now note that since S is full and P is irreducible, Proposition 9.2.9 shows that S is accessible. Then for all $x \in \mathsf{X}$,

$$\begin{aligned} \mathbb{P}_x(\sigma_A < \infty) &\geq \mathbb{P}_x(\sigma_S < \infty, \sigma_A \circ \theta_{\sigma_S} < \infty) \\ &= \mathbb{E}_x[\mathbb{1}_{\{\sigma_S < \infty\}}\mathbb{P}_{X_{\sigma_S}}(\sigma_A < \infty)] = \mathbb{P}_x(\sigma_S < \infty) > 0 , \end{aligned}$$

showing that A is accessible. □

Theorem 9.2.15. *Let P be an irreducible Markov kernel. An invariant probability measure for P is a maximal irreducibility measure.*

Proof. Let π be an invariant probability measure. We must prove that $\pi(A) > 0$ if and only if A is accessible. Fix $\varepsilon > 0$. The invariance of π with respect to P implies that it is invariant with respect to K_{a_ε}. If A is accessible, then by Lemma 3.5.2, $K_{a_\varepsilon}(x,A) > 0$ for all $x \in \mathsf{X}$. This implies $\pi(A) = \pi K_{a_\varepsilon}(A) > 0$.

We now prove the converse implication. Let $A \in \mathscr{X}$ be such that $\pi(A) > 0$. Set $\bar{A} = \{x \in \mathsf{X} : \mathbb{P}_x(\tau_A < \infty) > 0\}$. By the strong Markov property, we have for $x \in \bar{A}^c$,

$$0 = \mathbb{P}_x(\tau_A < \infty) \geq \mathbb{E}_x\left[\mathbb{1}\left\{\sigma_{\bar{A}} < \infty\right\} \mathbb{P}_{X_{\sigma_{\bar{A}}}}(\sigma_A < \infty)\right] .$$

Since $\mathbb{P}_{X_{\sigma_{\bar{A}}}}(\tau_A < \infty) > 0$ if $\sigma_{\bar{A}} < \infty$, the previous identity implies that for $x \in \bar{A}^c$, $\mathbb{P}_x(\sigma_{\bar{A}} < \infty) = \mathbb{P}_x(\tau_{\bar{A}} < \infty) = 0$. This means that $K_{a_\varepsilon}(x,\bar{A}) = 0$ for all $x \in \bar{A}^c$. Then since π is invariant,

$$\pi(\bar{A}) = \pi K_{a_\varepsilon}(\bar{A}) = \int_{\mathsf{X}} K_{a_\varepsilon}(x,\bar{A})\pi(\mathrm{d}x) = \int_{\bar{A}} K_{a_\varepsilon}(x,\bar{A})\pi(\mathrm{d}x) . \tag{9.2.3}$$

Noting that $A \subset \bar{A}$ and π is a probability measure, we have $0 < \pi(A) \leq \pi(\bar{A}) \leq 1 < \infty$. Combining this with (9.2.3) yields $K_{a_\varepsilon}(x,\bar{A}) = 1$ for π-almost all $x \in \bar{A}$. Equivalently, $K_{a_\varepsilon}(x,\bar{A}^c) = 0$ for π-almost all $x \in \bar{A}$. Therefore, $\bar{A}^c$ is inaccessible, and by Proposition 9.2.9, $\bar{A}$ is accessible, and consequently A is also accessible by Proposition 9.2.8. □

This property of invariant probability measures for an irreducible kernel has an important corollary.

Corollary 9.2.16 *If P is irreducible, then it admits at most one invariant probability measure.*

Proof. Assume that P admits two distinct invariant probability measures. Then by Theorem 1.4.6 there exist two mutually singular invariant probability measures π_1 and π_2. Since invariant probability measures are maximal irreducibility measures and maximal irreducibility measures are equivalent by Theorem 9.2.4, we obtain a contradiction. □

It is worthwhile to note that according to Corollary 9.2.16, an irreducible kernel admits at most one invariant probability measure, but it may admit more than one invariant measure. This is illustrated in the following example.

Example 9.2.17. Let $p \in (0,1) \setminus \{1/2\}$ and consider the Markov kernel P on $\mathbb{Z}$ defined by

$$P(x,y) = p\mathbb{1}\{y = x+1\} + (1-p)\mathbb{1}\{y = x-1\} \ .$$

The measures λ_0 and λ_1 defined respectively by $\lambda_0(k) = 1$ and $\lambda_1(k) = ((1-p)/p)^k$ for all $k \in \mathbb{Z}$ are both invariant with respect to P and are maximal irreducibility measures. ◀

9.3 Periodicity and Aperiodicity

For an irreducible kernel, it is possible to extend the notion of period to accessible small sets. Let C be an accessible small set and define the set E_C by

$$E_C = \left\{ n \in \mathbb{N}^* \ : \ \inf_{x \in C} P^n(x,C) > 0 \right\} \ .$$

By Lemma 9.1.6, there exist an integer m and a measure μ such that C is an (m,μ)-small set with $\mu(C) > 0$. Then $m \in E_C$, since by definition, for all $x \in C$,

$$P^m(x,C) \geq \mu(C) > 0 \ .$$

Thus the set E_C is not empty.

Definition 9.3.1 (Period of an accessible small set) *The period of an accessible small set C is the positive integer $d(C)$ defined by*

$$d(C) = g.c.d. \left\{ n \in \mathbb{N}^* \ : \ \inf_{x \in C} P^n(x,C) > 0 \right\} \ . \tag{9.3.1}$$

Lemma 9.3.2 *Let C be an accessible small set.*

(i) E_C is stable under addition.
(ii) There exists an integer n_0 such that $nd(C) \in E_C$ for all $n \geq n_0$.

Proof. (i) If $n,m \in E_C$, then for $x \in C$,

$$\begin{aligned} P^{n+m}(x,C) &\geq \int_C P^n(x,\mathrm{d}y) P^m(y,C) \geq P^n(x,C) \inf_{y \in C} P^m(y,C) \\ &\geq \inf_{z \in C} P^n(z,C) \inf_{y \in C} P^m(y,C) > 0 \ . \end{aligned}$$

(ii) This follows from Lemma 6.3.2, since E_C is stable under addition.

□

Lemma 9.3.3 *Let C be an accessible $(m,\varepsilon\nu)$-small set with $\nu(C)>0$.*

(i) For every $n\in E_C$, C is an $(n+m,\varepsilon\eta_n\nu)$-small set with $\eta_n=\inf_{z\in C}P^n(z,C)$.
(ii) There exists an integer n_0 such that for all $n\geq n_0$, C is an $(nd(C),\varepsilon_n\nu)$-small set with $\varepsilon_n>0$.

Proof. (i) For $x\in C$ and $A\in\mathscr{X}$,

$$P^{m+n}(x,A)\geq\int_C P^n(x,\mathrm{d}y)P^m(y,A)\geq\varepsilon P^n(x,C)\nu(A)\geq\varepsilon\eta_n\nu(A)\,.$$

(ii) Since $m\in E_C$, $d(C)$ divides m, the result follows from Lemma 9.3.2 (ii) and Lemma 9.3.3 (i).

□

As in the countable space case, it can be shown that the value of $d(C)$ is in fact a property of the Markov kernel P and does not depend on the particular small set C chosen.

Lemma 9.3.4. *Let C and C' be accessible small sets. Then $d(C)=d(C')$.*

Proof. Assume that C and C' are (m,μ)- and (m',μ')-small sets and $\mu(C)>0$ $\mu'(C')>0$. By Lemma 9.1.7, accessible sets are uniformly accessible from small sets, i.e., there exist $k,k'\in\mathbb{N}^*$ such that

$$\inf_{x\in C}P^k(x,C')>0\quad\text{and}\quad\inf_{x\in C'}P^{k'}(x,C)>0\,.$$

For $n\in E_C$ and $n'\in E_{C'}$, we have $\inf_{x\in C}P^n(x,C)>0$ and $\inf_{x\in C'}P^{n'}(x,C')>0$. Then for $x\in C$, we have

$$\begin{aligned}P^{k+n'+k'+n}(x,C)&\geq\int_{C'}P^k(x,\mathrm{d}x')\int_{C'}P^{n'}(x',\mathrm{d}y')\int_C P^{k'}(y',\mathrm{d}y)P^n(y,C)\\&\geq\inf_{x\in C}P^n(x,C)\ \inf_{x\in C'}P^{k'}(x,C)\ \inf_{x\in C'}P^{n'}(x,C')\ \inf_{x\in C}P^k(x,C')>0\,.\end{aligned}$$

Thus $k+n'+k'+n\in E_C$. Since $n'\in E_{C'}$ is arbitrary and $E_{C'}$ is closed under addition, the same holds with $2n'$, i.e., $k+2n'+k'+n\in E_C$. Thus $n'=(k+2n'+k'+n)-(k+n'+k'+n)$ is a multiple of $d(C)$, and this implies that $d(C)$ divides $d(C')$. Similarly, $d(C')$ divides $d(C)$, and this yields $d(C)=d(C')$. □

Let us introduce the following definition.

Definition 9.3.5 (Period, aperiodicity, strong aperiodicity) *Let P be an irreducible Markov kernel on $\mathsf{X}\times\mathscr{X}$.*

- *The common period of all accessible small sets is called the period of P.*
- *If the period is equal to one, the kernel is said to be aperiodic.*
- *If there exists an accessible $(1,\mu)$-small set C with $\mu(C)>0$, the kernel is said to be strongly aperiodic.*

If P is an irreducible Markov kernel with period d, then the state space can be partitioned similarly to what happens for a denumerable state space; see Theorem 7.4.1. We will see later that this decomposition is essentially unique.

Theorem 9.3.6. *Let P be an irreducible Markov kernel with period d. There exists a sequence $C_0, C_1, \ldots, C_{d-1}$ of mutually disjoint accessible sets such that for $i = 0, \ldots, d-1$ and $x \in C_i$, $P(x, C_{i+1[d]}) = 1$. Consequently, $\bigcup_{i=0}^{d-1} C_i$ is absorbing.*

Proof. Let C be an accessible small set. For $i = 0, \ldots, d-1$, define

$$\bar{C}_i = \left\{ x \in \mathsf{X} : \sum_{n=1}^{\infty} P^{nd-i}(x, C) > 0 \right\} .$$

Note that since C has period d, $C \subset \bar{C}_0$. Since C is accessible, $\mathsf{X} = \bigcup_{i=0}^{d-1} \bar{C}_i$. Let $i, j \in \{0, \ldots, d-1\}$ and assume that $\bar{C}_i \cap \bar{C}_j$ is accessible. For $n, p, k \in \mathbb{N}^*$, define

$$A_{n,p,k} = \left\{ x \in \bar{C}_i \cap \bar{C}_j \,:\, P^{nd-i}(x, C) \wedge P^{pd-j}(x, C) > 1/k \right\} .$$

Since $\bigcup_{n,p,k \geq 1} A_{n,p,k} = \bar{C}_i \cap \bar{C}_j$ is assumed to be accessible, there exist $n, p, k \geq 1$ such that $A_{n,p,k}$ is accessible. Lemma 9.1.7 shows that there exist $\alpha > 0$ and $r \in \mathbb{N}$ such that $\inf_{x \in C} P^r(x, A_{n,p,k}) \geq \alpha$. Set $\eta_n = \inf_{x \in C} P^n(x, C)$. This yields, for $x \in C$ and $s \in E_C$,

$$\begin{aligned}
P^{s+r+nd-i+s}(x, C) &\geq \int_C P^s(x, \mathrm{d}y) \int_{A_{n,p,k}} P^r(y, \mathrm{d}x') \int_C P^{nd-i}(x', \mathrm{d}y') P^s(y', C) \\
&\geq \eta_s \int_C P^s(x, \mathrm{d}y) \int_{A_{n,p,k}} P^r(y, \mathrm{d}x') P^{nd-i}(x', C) \\
&\geq k^{-1} \eta_s \int_C P^s(x, \mathrm{d}y) P^r(y, A_{n,p,k}) \geq k^{-1} \eta_s \alpha P^s(x, C) \geq k^{-1} \eta_s^2 \alpha .
\end{aligned}$$

This implies that $s + r + nd - i + s \in E_C$. Similarly, $s + r + nd - j + s \in E_C$, and this implies that d divides $(i - j)$. Since $i, j \in \{0, 1, \ldots, d-1\}$, this implies that $i = j$.

We have thus proved that if $i \neq j$, then $\bar{C}_i \cap \bar{C}_j$ is inaccessible. Set

$$G = \bigcup_{i,j=0}^{d-1} (\bar{C}_i \cap \bar{C}_j) \quad \text{and} \quad F = \bigcup_{i=0}^{d-1} \bar{C}_i \setminus G = \mathsf{X} \setminus G .$$

Then F is full, and thus by Proposition 9.2.12, there exists an absorbing full set $D \subset F$. For $i = 0, \ldots, d-1$, set $C_i = (\bar{C}_i \setminus G) \cap D$. The sets $\{C_i\}_{i=0}^{d-1}$ are mutually disjoint, and $\bigcup_{i=0}^{d-1} C_i = D$ is full and absorbing. Thus for $x \in D$, there exists $i \in \{0, \ldots, d-1\}$ such that $P(x, C_i) > 0$. Then

$$\sum_{n\geq 1} P^{nd-(i-1)}(x,C) \geq \int_{C_i} P(x,\mathrm{d}y) \sum_{n\geq 1} P^{nd-i}(y,C) > 0 \; .$$

Thus $x \in C_{i-1}$ if $i > 0$ and $x \in C_{d-1}$ if $i = 0$. Since the sets C_i are mutually disjoint, this in turn implies that $P(x,C_i) = 1$. □

The sets $\{C_i\}_{i=0}^{d-1}$ in Theorem 9.3.6 are called periodicity classes, and the decomposition $\{C_i\}_{i=0}^{d-1}$ is called a cyclic decomposition. This is a deep result that ensures that we can think of cycles in general spaces exactly as we think of them in countable spaces. The following corollary shows that up to an inaccessible set, this decomposition is unique.

Corollary 9.3.7 *Let $(C_0,\dots,C_{d-1})$ and $(D_0,\dots,D_{d-1})$ be two cyclic decompositions. Then there exists $j \in \{0,\dots,d-1\}$ such that $\left\{C_i \cap D_{i+j} \,:\, i = 0,\dots,d-1\right\}$ (where addition is modulo d) is a cyclic decomposition.*

Proof. Since $\cup_{i=0}^{d-1} D_i$ is absorbing, it is full. Then setting $N = (\cup D_j)^c$, N is inaccessible. Write

$$C_0 = (C_0 \cap N) \bigcup_{j=0}^{d-1} (C_0 \cap D_j) \; .$$

Since C_0 is accessible, there exists by Proposition 9.2.9 at least one j such that $C_0 \cap D_j$ is accessible. Up to a permutation, we can assume without loss of generality that $C_0 \cap D_0$ is accessible. For $i = 0,\dots,d-1$, set $E_i = C_i \cap D_i$ and $E_d = E_0$. Then the sets E_i $i = 0,\dots,d-1$ are distinct, and for $i = 0,\dots,d-1$ and $x \in E_i$, $P(x,C_{i+1}) = P(x,D_{i+1}) = 1$. Thus

$$P(x,E_{i+1}) = P(x,C_{i+1} \cap D_{i+1}) \geq 1 - P(x,C_{i+1}^c) - P(x,D_{i+1}^c) = 1 \; .$$

This implies that $\cup_{i=0}^{d-1} E_i$ is absorbing and that $(E_0,\dots,E_{d-1})$ is a cyclic decomposition. □

An important consequence of this cycle decomposition is that an accessible small set must be included in a periodicity class.

Corollary 9.3.8 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. If C is an accessible small set and $(D_0,\dots,D_{d-1})$ is a cyclic decomposition, then there exists a unique $j \in \{0,\dots,d-1\}$ such that $C \subset D_j \cup N$, where $N = (\bigcup_{i=0}^{d-1} D_i)^c$ is inaccessible.*

Proof. By Lemma 9.1.6, we can assume that C is an accessible (m,μ)-small set with $\mu(C) > 0$. By Proposition 9.1.9, $\mu(S^c) = 0$. Thus there exists $k \in \{0,\dots,d-1\}$ such that $\mu(C \cap D_k) > 0$. If $x \in C \cap D_k$, then $P^m(x, C \cap D_k) \geq \mu(C \cap D_k) > 0$, so $m = rd$ for some $r \in \mathbb{N}*$. Now if $x \in C$, then $P^{rd}(x, C \cap D_k) > 0$, which implies that $x \notin D_j$ for $j \neq k$. This shows that $C \subset D_k \cup S^c$. □

Another interesting consequence of the cyclic decomposition is a simple condition for P to be aperiodic.

Lemma 9.3.9 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. If for all $A \in \mathscr{X}_P^+$, $\lim_{n\to\infty} P^n(x, A) = \pi(A)$ for π-almost all $x \in \mathsf{X}$, then P is aperiodic.*

Proof. The proof is by contradiction. Assume that the period d is larger than 2. Let $C_0, \ldots, C_{d-1}$ be a cyclic decomposition as stated in Theorem 9.3.6 and note that $\pi(C_0) > 0$, since C_0 is accessible and π is a maximal irreducibility measure (see Theorem 9.2.15). By assumption, there exists $x \in C_0$ such that $\lim_{n\to\infty} P^n(x, C_0) = \pi(C_0) > 0$. But since $P^{1+kd}(x, C_1) = 1$ for all $k \in \mathbb{N}$, we must have $P^{1+kd}(x, C_0) = 0$, which contradicts the fact that $\lim_{n\to\infty} P^n(x, C_0) = \pi(C_0) > 0$. □

Theorem 9.3.10. *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. The chain is aperiodic if and only if for all $A \in \mathscr{X}_P^+$ and $x \in \mathsf{X}$, there exists k_0 such that $P^k(x, A) > 0$ for all $k \geq k_0$.*

Proof. Assume first that P is aperiodic. Choose $A \in \mathscr{X}_P^+$ and $x \in \mathsf{X}$. By Theorem 9.2.5, there exists an accessible $(m, \varepsilon\nu)$-small set $B \subset A$ with $\nu(B) > 0$. By Lemma 9.3.3, there exist $r \geq 1$ and a sequence of constants $\varepsilon_k > 0$ such that B is a $(k, \varepsilon_k \nu)$-small set for all $k \geq r$. Since B is accessible, there exists $n \geq 1$ such that $P^n(x, B) > 0$. Thus for $k \geq n + r$,

$$\begin{aligned} P^k(x, A) \geq P^k(x, B) &\geq \int_B P^n(x, \mathrm{d}y) P^{k-n}(y, B) \\ &\geq \varepsilon_{k-n} \nu(B) P^n(x, B) > 0 \,. \end{aligned}$$

Conversely, if P is not aperiodic, then Theorem 9.3.6 implies that the condition of the proposition cannot be satisfied. □

It will become clear as we proceed that it is much easier to deal with strongly aperiodic chains. Regrettably, this condition is not satisfied in general; however, we can often prove results for strongly aperiodic chains and then use special methods to extend them to general chains through the m-skeleton chain or the resolvent kernel.

Theorem 9.3.11. *Let P be an irreducible and aperiodic kernel on $\mathsf{X} \times \mathscr{X}$ and $m \in \mathbb{N}^*$. Then:*

(i) A set A is accessible for P if and only if A is accessible for P^m, i.e., $\mathscr{X}_P^+ = \mathscr{X}_{P^m}^+$.

(ii) P^m is irreducible and aperiodic.

(iii) If C is an accessible small set for P, then C is an accessible small set for P^m. Moreover, there exists $m_0 \in \mathbb{N}$ such that C is an accessible 1-small set for P^{m_0}.

Proof. (i) Obviously, $\mathscr{X}_{P^m}^+ \subset \mathscr{X}_P^+$. We now establish that $\mathscr{X}_P^+ \subseteq \mathscr{X}_{P^m}^+$ for every $m \in \mathbb{N}$. Let $A \in \mathscr{X}_P^+$. Applying Theorem 9.3.10, for all $x \in \mathsf{X}$ there exists k_0 such that $P^k(x,A) > 0$ for all $k \geq k_0$. This implies that $A \in \mathscr{X}_{P^m}^+$ for all $m \in \mathbb{N}^*$.

(ii) Let ϕ be an irreducibility measure for P and let $A \in \mathscr{X}$ be such that $\phi(A) > 0$. Then $A \in \mathscr{X}_P^+$, and by (i), $A \in \mathscr{X}_{P^m}^+$. Hence ϕ is also an irreducibility measure for P^m, and consequently, P^m is irreducible. Since $\mathscr{X}_{P^m}^+ \subset \mathscr{X}_P^+$, Theorem 9.3.10 shows that for all $A \in \mathscr{X}_{P^m}^+$ and $x \in \mathsf{X}$, there exists $k_0 > 0$ such that $P^k(x,A) > 0$ for all $k \geq k_0$, and thus $A \in \mathscr{X}_P^+$. This, of course, implies that for $\ell \geq \lceil k_0/m \rceil$, $P^{\ell m}(x,A) > 0$. Theorem 9.3.10 then shows that P^m is aperiodic.

(iii) Let C be an accessible $(r, \varepsilon\nu)$-small set. Since P is aperiodic, Lemma 9.3.2 shows that there exists n_0 such that for all $n \geq n_0$, $\inf_{x \in C} P^n(x,C) > 0$. For all $n \geq n_0$, all $x \in C$, and $A \in \mathscr{X}$,

$$P^{n+r}(x,A) \geq \int_C P^n(x,\mathrm{d}y)P^r(y,A) \geq \varepsilon\nu(A) \inf_{x \in C} P^n(x,C) \, .$$

On choosing $n \geq n_0$ such that $n + r$ is a multiple of m, this relation shows that C is a small set for the skeleton P^m. Hence there exists $k \in \mathbb{N}$ such that C is 1-small for P^{km}. Moreover, C is accessible for P^{km} by (i).

□

9.4 Petite Sets

Small sets are very important in the theory of Markov chains, but unfortunately, the union of two small sets is not necessarily small. For example, setting $\mathsf{X} = \{0,1\}$ and $P(0,1) = P(1,0) = 1$, we have that $\{0\}$ and $\{1\}$ are small, but the whole state space $\{0,1\}$ is not small. Therefore, we introduce a generalization of small sets, called petite sets, which will be a convenient substitute for small sets. We will later see that for aperiodic kernels, petite sets and small sets coincide. First we introduce the set of probabilities on $\mathbb{N}$ that put zero mass at zero:

$$\mathbb{M}_1^*(\mathbb{N}) = \{a = \{a(n),\, n \in \mathbb{N}\} \in \mathbb{M}_1(\mathbb{N}) : a(0) = 0\} \, . \tag{9.4.1}$$

Note that $\mathbb{M}_1^*(\mathbb{N})$ is stable by convolution, i.e., if $a \in \mathbb{M}_1^*(\mathbb{N})$ and $b \in \mathbb{M}_1^*(\mathbb{N})$, then $c = a * b \in \mathbb{M}_1^*(\mathbb{N})$.

Definition 9.4.1 (Petite set) *A set $C \in \mathscr{X}$ is called petite if there exist $a \in \mathbb{M}_1(\mathbb{N})$ and a nonzero measure $\mu \in \mathbb{M}_+(\mathscr{X})$ such that for all $x \in C$ and $A \in \mathscr{X}$,*

$$K_a(x,A) \geq \mu(A) .$$

The set C is then said to be an (a,μ)-petite set.

In other words, a petite set is a 1-small set for a sampled kernel K_a. The empty set is petite. An (m,μ)-small set is a petite set for the sampling distribution a that puts mass 1 at m. The converse is generally false, as will be shown after Proposition 9.4.5.

Lemma 9.4.2 *If the Markov kernel P admits an accessible (a,μ)-petite set, then μ is an irreducibility measure and P is irreducible.*

Proof. Let C be an accessible (a,μ)-petite set, $\varepsilon \in (0,1)$, $x \in \mathsf{X}$, and $A \in \mathscr{X}$ such that $\mu(A) > 0$. Since C is accessible, $K_{a_\varepsilon}(x,C) > 0$. Using the generalized Chapman–Kolmogorov formula (Lemma 1.2.11), we have

$$K_{a*a_\varepsilon}(x,A) \geq \int_C K_{a_\varepsilon}(x,\mathrm{d}y)K_a(y,A) \geq \mu(A)K_{a_\varepsilon}(x,C) > 0 .$$

This shows that A is accessible and μ is an irreducibility measure. □

Lemma 9.4.3 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let $C,D \in \mathscr{X}$. Assume that C is an (a,μ)-petite set and that there exists $b \in \mathbb{M}_1(\mathbb{N})$ such that $\delta = \inf_{x\in D} K_b(x,C) > 0$. Then D is a petite set.*

Proof. For all $x \in C$ and $A \in \mathscr{X}$, we get that

$$K_{b*a}(x,A) \geq \int_C K_b(x,\mathrm{d}y)K_a(y,A) \geq \mu(A)K_b(x,C) \geq \delta\mu(A) .$$

□

The following lemma shows that the minorization measure in the definition of petite set may always be chosen to be a maximal irreducibility measure when P is irreducible. For $p \in \mathbb{N}$, define the probability $\gamma_p \in \mathbb{M}_1(\mathbb{N})$:

$$\gamma_p(k) = 1/p \,\text{for}\, k \in \{1,\ldots,p\} \text{ and } \gamma_p(k) = 0 \text{ otherwise.} \tag{9.4.2}$$

Proposition 9.4.4 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let C be a petite set.*

(i) There exist a sampling distribution $b \in \mathbb{M}_1^*(\mathbb{N})$ and a maximal irreducibility measure ψ such that C is a (b,ψ)-petite set.

(ii) There exist $p \in \mathbb{N}$ and a nontrivial measure μ such that C is a (γ_p, μ)-petite set.

Proof. (i) Let C be an (a, μ_C)-petite set and D an accessible (m, μ_D)-small set (such a set always exists by Definition 9.2.1). By Proposition 9.1.9, μ_D is an irreducibility measure. Since D is accessible, for every $\varepsilon \in (0,1)$, Lemma 3.5.2 implies $\mu_C K_{a_\varepsilon}(D) > 0$. The measure $\psi = \mu_C K_{a_\varepsilon}(D) \mu_D K_{a_\varepsilon}$ is then a maximal irreducibility measure by Theorem 9.2.4. Therefore, for $x \in C$ and $A \in \mathscr{X}$, using again the generalized Chapman–Kolmogorov equations, we get

$$
\begin{aligned}
K_{a*a_\varepsilon*\delta_m*a_\varepsilon}(x,A) &\geq \int_D K_{a*a_\varepsilon}(x,\mathrm{d}y) \int_{\mathsf{X}} P^m(y,\mathrm{d}z) K_{a_\varepsilon}(z,A) \\
&\geq \int_D K_{a*a_\varepsilon}(x,\mathrm{d}y) \mu_D K_{a_\varepsilon}(A) = K_{a*a_\varepsilon}(x,D) \mu_D K_{a_\varepsilon}(A) \\
&\geq \int K_a(x,\mathrm{d}y) K_{a_\varepsilon}(y,D) \mu_D K_{a_\varepsilon}(A) \geq \mu_C K_{a_\varepsilon}(D) \mu_D K_{a_\varepsilon}(A) = \psi(A) \, .
\end{aligned}
$$

This proves that C is a (b, ψ)-petite set, with $b = a * a_\varepsilon * \delta_m * a_\varepsilon$. Note that $b(0) = 0$.

(ii) By (i), we can assume that C is a (b, ψ)-petite set, where ψ is a maximal irreducibility measure and $b \in \mathbb{M}_1(\mathbb{N})$. For all $x \in C$, $K_b(x,D) \geq \psi(D) > 0$. Choose N such that $\sum_{k=N+1}^{\infty} b(k) \leq (1/2)\psi(D)$. Then for all $x \in C$,

$$
\sum_{k=0}^{N} P^k(x,D) \geq \sum_{k=0}^{N} b(k) P^k(x,D) \geq \frac{1}{2}\psi(D) \, ,
$$

and for all $A \in \mathscr{X}$,

$$
\begin{aligned}
\sum_{k=1}^{N+m} P^k(x,A) &\geq \sum_{k=0}^{N} P^{k+m}(x,A) \geq \sum_{k=0}^{N} \int_D P^k(x,\mathrm{d}y) P^m(y,A) \\
&\geq \mu_D(A) \sum_{k=0}^{N} P^k(x,D) \geq \frac{1}{2}\psi(D)\mu_D(A) \, .
\end{aligned}
$$

□

A main difference between small and petite sets is that the union of two petite sets is petite, whereas the union of two small sets is not necessarily small.

Proposition 9.4.5 *Let P be an irreducible kernel on $\mathsf{X} \times \mathscr{X}$. Then a finite union of petite sets is petite, and X is covered by a denumerable union of increasing petite sets.*

Proof. Let C be an (a,μ)-petite set and D a (b,ν)-petite set. By Proposition 9.4.4, we can assume that μ and ν are maximal irreducibility measures. Set $c = (a+b)/2$. Since the Markov kernel P is irreducible, there exists an accessible small set B. Then $\mu(B) > 0$, $\nu(B) > 0$, and for $x \in C \cup D$, we have

$$
\begin{aligned}
K_c(x,B) &= \frac{1}{2} K_a(x,B) + \frac{1}{2} K_b(x,B) \geq \frac{1}{2}\mu(B)\mathbb{1}_C(x) + \frac{1}{2}\nu(B)\mathbb{1}_D(x) \\
&\geq \frac{1}{2}\{\mu(B) \wedge \nu(B)\} > 0 \, .
\end{aligned}
$$

Thus $C \cup D$ is petite by Lemma 9.4.3.

By definition, P admits at least one accessible small set. By Proposition 9.1.8, X is covered by a countable union of small sets $\{C_j,\ j \in \mathbb{N}^*\}$, i.e., $\mathsf{X} = \bigcup_{i=1}^{\infty} C_i$. For $j \geq 1$, set $D_j = \bigcup_{i=1}^{j} C_i$. Then $D_j \subset D_{j+1}$ and $\mathsf{X} = \cup_{j \geq 1} D_j$. Moreover, for each $j \geq 1$, D_j is petite as a finite union of small sets. $\square$

Definition 9.4.6 (Uniform accessibility) *A set B is uniformly accessible from A if there exists $m \in \mathbb{N}^*$ such that* $\inf_{x \in A} \mathbb{P}_x(\sigma_B \leq m) > 0$.

Lemma 9.4.7 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ and C, $D \in \mathscr{X}$.*

(i) Let $m \in \mathbb{N}$. Then

$$
\inf_{x \in C} \mathbb{P}_x(\sigma_D \leq m) > 0 \Longleftrightarrow \inf_{x \in C} K_{\gamma_m}(x,D) > 0 \, ,
$$

where γ_m is defined in (9.4.2).

(ii) If D is petite and uniformly accessible from $C \in \mathscr{X}$, then C is petite.

Proof. (i) Each of these conditions is equivalent to the existence of $\delta > 0$ satisfying the following condition: for all $x \in C$, there exists $r(x) \in \{1, \ldots, m\}$ such that $P^{r(x)}(x,D) \geq \delta/m$.

(ii) This follows from (i) and Lemma 9.4.3.

$\square$

Lemma 9.4.8 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$, and C a petite set. Let r be a nonnegative increasing sequence such that* $\lim_{n \to \infty} r(n) = \infty$.

(i) For every $d > 0$, the set $\{x \in \mathsf{X} : \mathbb{E}_x[r(\tau_C)] \leq d\}$ is petite.
(ii) Every set $B \in \mathscr{X}$ such that $\sup_{x \in B} \mathbb{E}_x[r(\tau_C)] < \infty$ *is petite.*

Proof. (i) Set $D = \{x \in \mathsf{X} : \mathbb{E}_x[r(\tau_C)] \leq d\}$. Since C is petite, $D \cap C$ is also petite. Consider now $x \in D \cap C^c$. For all $k \in \mathbb{N}^*$, we have

$$
\mathbb{P}_x(\sigma_C \geq k) = \mathbb{P}_x(\tau_C \geq k) = \mathbb{P}_x(r(\tau_C) \geq r(k)) \leq [r(k)]^{-1}\mathbb{E}_x[r(\tau_C)] \leq [r(k)]^{-1} d \, .
$$

Thus for k sufficiently large, $\inf_{x\in D\cap C^c} \mathbb{P}_x(\sigma_C \le k) \ge 1/2$. This proves that C is uniformly accessible from $D\cap C^c$; hence $D\cap C^c$ is petite by Lemma 9.4.7. Since the union of two petite sets is petite by Proposition 9.4.5, this proves that D is petite.

(ii) By assumption, there exists $b > 0$ such that $B \subset \{x \in \mathsf{X} : \mathbb{E}_x[r(\tau_C)] \le b\}$, which is petite by (i), and therefore B is also petite.

□

Proposition 9.4.9 *Let P be an irreducible Markov kernel. A set C is petite if and only if every accessible set A is uniformly accessible from C.*

Proof. Assume that C is a petite set. By Proposition 9.4.4, we can assume that C is (b,ψ)-petite, where ψ is a maximal irreducibility measure and $b \in \mathbb{M}_1(\mathbb{N})$ is a probability on $\mathbb{N}$ satisfying $b(0) = 0$. Let $A \in \mathscr{X}_P^+$. We have for all $x \in C$, $K_b(x,A) \ge \psi(A) > 0$. There exists $m \in \mathbb{N}$ such that

$$\sum_{k=1}^{m} P^k(x,A) \ge \sum_{k=1}^{m} b(k)P^k(x,A) \ge \frac{1}{2}\psi(A)\ .$$

Conversely, assume that every accessible set is uniformly accessible from C. Since P is irreducible, there exists an accessible (n,μ)-small set D. This implies that $\inf_{x\in C} \mathbb{P}_x(\sigma_D \le m) > 0$ for some $m > 0$, and by Lemma 9.4.7, C is petite. □

We have now all the ingredients to prove that for an aperiodic kernel, petite sets and small sets coincide.

Theorem 9.4.10. *If P is irreducible and aperiodic, then every petite set is small.*

Proof. Let C be a petite set and D an accessible (r,μ)-small set with $\mu(D) > 0$. By Proposition 9.4.9, we can also choose m_0 and $\delta > 0$ such that

$$\inf_{x\in C} \mathbb{P}_x(\tau_D \le m_0) \ge \inf_{x\in C} \mathbb{P}_x(\sigma_D \le m_0) \ge \delta\ .$$

Since P is aperiodic, Lemma 9.3.3 shows that we can choose $\varepsilon > 0$ and then $m \ge m_0$ large enough that $\inf_{x\in D} P^k(x,\cdot) \ge \varepsilon\mu(\cdot)$ for $k = m, m+1, \ldots, 2m$. Then for $x \in C$ and $B \in \mathscr{X}$,

$$\begin{aligned} P^{2m}(x,B) &= \mathbb{P}_x(X_{2m} \in B) \ge \sum_{k=0}^{m} \mathbb{P}_x(\tau_D = k, X_{2m} \in B) \\ &= \sum_{k=0}^{m} \mathbb{E}_x\left[\mathbb{1}\{\tau_D = k\}\mathbb{P}_{X_k}(X_{2m-k} \in B)\right] \ge \varepsilon\mu(B)\sum_{k=0}^{m} \mathbb{P}_x(\tau_D = k) \ge \delta\varepsilon\mu(B)\ . \end{aligned}$$

Therefore, C is a $(2m,\mu)$-small set. □

Proposition 9.4.11 *An irreducible Markov kernel P on $\mathsf{X}\times\mathscr{X}$ is aperiodic if and only if X is covered by an increasing denumerable union of small sets.*

Proof. By Proposition 9.4.5, X is covered by an denumerable union of increasing petite sets $\{D_j,\ j\in\mathbb{N}\}$. Since P is aperiodic, by Theorem 9.4.10, D_j is small for all $j\in\mathbb{N}$.

Conversely, assume that $\mathsf{X}=\cup_{j\geq 1}D_j$, where $\{D_j,\ j\in\mathbb{N}^*\}$ is an increasing sequence of small sets. By Proposition 9.2.9, there exists j such that D_j is accessible. Applying Corollary 9.3.8, there exists a periodicity class C_0 such that $D_j\subset C_0\cup N_j$, where N_j is inaccessible. Since D_k is also an accessible small set for all for all $k\geq j$ and contains D_j, Corollary 9.3.8 also implies $D_k\subset C_0\cup N_k$, where N_k is inaccessible. Finally, $\cup_{k\geq j}D_k=\mathsf{X}$ is therefore included in a periodicity class up to an inaccessible set, and P is aperiodic. □

We conclude with a very important property of petite sets: invariant measures give finite mass to petite sets.

Lemma 9.4.12 *Let P be an irreducible Markov kernel on $\mathsf{X}\times\mathscr{X}$. Let $\mu\in\mathbb{M}_+(\mathscr{X})$ be a σ-finite measure such that $\mu P\leq\mu$. Then $\mu(C)<\infty$ for every petite set C.*

Proof. Let C be an (a,ν)-petite set, where $a\in\mathbb{M}_1(\mathbb{N})$. Since μ is σ-finite, there exists $B\in\mathscr{X}$ such that $\mu(B)<\infty$ and $\nu(B)>0$. Then

$$\infty>\mu(B)\geq\int_C\mu(\mathrm{d}x)K_a(x,B)\geq\nu(B)\mu(C)\;.$$

Thus $\mu(C)$ is finite. □

9.5 Exercises

9.1. Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$. Assume that P is irreducible. For every $A\in\mathscr{X}_P^+$ and irreducibility measure ϕ, show that $\phi_A(\cdot)=\phi(A\cap\cdot)$ is an irreducibility measure.

9.2. Assume that P admits an accessible atom α. Construct a maximal irreducibility measure.

9.3. Let $\{Z_k,\ k\in\mathbb{N}^*\}$ be an i.i.d. sequence of real-valued random variables with cumulative distribution function F. Let X_0 be a real-valued random variable independent of $\{Z_k,\ k\in\mathbb{N}^*\}$. Consider the Markov chain defined by $X_n=(X_{n-1}+Z_n)^+$ for $n\geq 1$. Denote by P the Markov kernel associated to this Markov chain.

1. Show that P is irreducible if and only if $F((-\infty,0)) > 0$.

Assume now that $F((-\infty,0)) > 0$.

2. Show that $\{0\}$ is uniformly accessible from every compact $C \subset \mathbb{R}$.

9.4. Let P be an irreducible and aperiodic Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. Show that for every accessible set B there is a small set $C \subset B$ such that for some n and $\delta > 0$,

$$P^n(x,A) \geq \delta\pi(A) , \quad x \in C, A \subset C .$$

9.5. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that there exist a σ-finite measure λ and a sequence $\{p_n,\ n \in \mathbb{N}^*\}$ of nonnegative functions defined on X^2 such that for every $n \geq 1$ and $x \in \mathsf{X}$, $p_n(x,\cdot)$ is measurable and for every $n \geq 1$, $x \in \mathsf{X}$ and $A \in \mathscr{X}$,

$$P^n(x,A) \geq \int_A p_n(x,y)\lambda(\mathrm{d}y) .$$

Assume that there exists a set $C \in \mathscr{X}$ such that $\lambda(C) > 0$ and for all $x \in \mathsf{X}$, $\sum_{n=1}^{\infty} p_n(x,y) > 0$ for λ-almost every $y \in C$. Show that λ restricted to C is an irreducibility measure.

9.6. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that there exist an integer m, a set $C \in \mathscr{X}$, a probability measure ν, and for every $B \in \mathscr{X}$, a measurable function $y \mapsto \varepsilon(y,B)$ such that for all $x \in C$, $P^m(x,B) \geq \varepsilon(x,B)\nu(B)$. Show that if C is accessible and for every $B \in \mathscr{X}$, $\varepsilon(x,B) > 0$ for every $x \in C$, then P is irreducible.

9.7. Let $\lambda \in \mathbb{M}_+(\mathscr{X})$ a σ-finite measure. Let π be a probability density with respect to λ. Let $q : \mathsf{X} \times \mathsf{X} \mapsto \mathbb{R}_+$ be a probability transition kernel with respect to λ. Consider the Metropolis–Hastings kernel with target density π and transition density q. Assume that

$$\pi(y) > 0 \ \Rightarrow \ (q(x,y) > 0 \text{ for all} x \in \mathsf{X}) .$$

Show that P is irreducible.

9.8. Let P be an irreducible aperiodic Markov chain on $\mathsf{X} \times \mathscr{X}$, where $\mathscr{X}$ is countably generated. Assume that P has an invariant probability denoted by π. Call a subset $S \in \mathscr{X}$ hypersmall if $\pi(S) > 0$ and there exist $\varepsilon > 0$ and $m \in \mathbb{N}$ such that S is $(m,\varepsilon\pi)$-small. Show that every set of positive π-measure contains a hypersmall set.

9.9. Let P be Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that there exist two disjoint absorbing sets $A_1, A_2 \in \mathscr{X}$ (i.e., A_1 and A_2 are absorbing and $A_1 \cap A_2 = \emptyset$). Show that the Markov kernel P is not irreducible.

9.10. Let $\{Z_k,\ k \in \mathbb{N}^*\}$ be a sequence of i.i.d. real-valued random variables. Let X_0 be a real-valued random variable, independent of $\{Z_k,\ k \in \mathbb{N}\}$. Consider the unrestricted random walk defined by $X_k = X_{k-1} + Z_k$ for $k \geq 1$.

Assume that the increment distribution (the distribution of Z_1) has an absolutely continuous part with respect to Lebesgue measure Leb on $\mathbb{R}$ with a density γ that is positive and bounded away from zero at the origin; that is, for some for some $\beta > 0, \delta > 0$,

$$\mathbb{P}(Z_1 \in A) \geq \int_A \gamma(x) \mathrm{d}x \,,$$

and $\gamma(x) \geq \delta > 0$ for all $|x| < \beta$.

1. Show that $C = [-\beta/2, \beta/2]$ is an accessible small set.
2. Show that $\mathrm{Leb}(\cdot \cap C)$ is an irreducibility measure.

Assume now that the increment distribution is concentrated on $\mathbb{Q}$, more precisely, for all $r \in \mathbb{Q}$, $\mathbb{P}(Z_1 = r) > 0$.

3. Show that the Markov chain is not irreducible.

9.11. Let P_1, P_2 be two Markov kernels defined on $\mathsf{X} \times \mathscr{X}$. Let $\alpha \in (0,1)$. Assume that $C \in \mathscr{X}$ is a accessible small set for the Markov kernel P_1. Show that C is an accessible small set for $P = \alpha P_1 + (1-\alpha) P_2$.

9.12. Let P_1, P_2 be two Markov kernels defined on $\mathsf{X} \times \mathscr{X}$. Suppose that C is a $(1, \varepsilon\nu)$-accessible small set for P_1. Show that C is also a small set for $P_1 P_2$ and $P_2 P_1$.

9.13. Consider the Metropolis–Hastings algorithm on a topological space X. Let ν be a σ-finite measure. Assume that the target distribution and the proposal kernel are dominated by ν, i.e., $\pi = h_\pi \cdot \nu$ and $Q(x,A) = \int_A q(x,y) \nu(\mathrm{d}y)$, where $q : \mathsf{X} \times \mathsf{X} \to \mathbb{R}_+$.

Assume that

(i) The density h_π is bounded above on compact sets of X.
(ii) The transition density q is bounded from below on compact sets of $\mathsf{X} \times \mathsf{X}$.

Show that P is irreducible with ν as an irreducibility measure, is strongly aperiodic, and every nonempty compact set C is small. This shows that every compact set is small and that $\pi|_C$, the restriction of π to the set C, is an irreducibility measure.

9.14. Consider the Metropolis–Hastings algorithm on a metric space (X, d). Let ν be a σ-finite measure. Assume that the target distribution and the proposal kernel are dominated by ν, i.e., $\pi = h_\pi \cdot \nu$ and $Q(x,A) = \int_A q(x,y) \nu(\mathrm{d}y)$, where $q : \mathsf{X} \times \mathsf{X} \to \mathbb{R}_+$. Assume that h_π is bounded away from 0 and ∞ on compact sets and that there exist $\delta > 0$ and $\varepsilon > 0$ such that for every $x \in \mathsf{X}$, $\inf_{\mathrm{B}(x,\delta)} q(x, \cdot) \geq \varepsilon$. Show that the Metropolis–Hastings kernel P is irreducible with ν as an irreducibility measure, is strongly aperiodic, and every nonempty compact set is small.

9.15. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let $C \in \mathscr{X}$ be an $(m, \varepsilon\nu)$-small set, where $m \in \mathbb{N}$ and $\varepsilon > 0$. Show that for all $(x, x') \in C \times C$,

$$\left\| P^m(x, \cdot) - P^m(x', \cdot) \right\|_{\mathrm{TV}} \leq 2(1-\varepsilon) \,.$$

9.16. Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Denote by d the period of P. Let $\{C_i\}_{i=0}^{d-1}$ be a cyclic decomposition. Show that every small set C must be essentially contained inside one specific member of the cyclic class, i.e., that there exists $i_0 \in \{0, \ldots, d-1\}$ such that $\psi(C \Delta C_{i_0}) = 0$, where ψ is a maximal irreducibility measure.

9.17. Consider the independent Metropolis–Hastings sampler introduced in Example 2.3.3 (we use the notation introduced in that example). Show that if $h(x) \le c\bar{q}(x)$ for some constant $c > 0$, then the state space X is small.

9.18. Let $\pi \in \mathbb{M}_1(\mathscr{X})$ be a probability and P a Metropolis–Hastings kernel on the state space $\mathsf{X} \times \mathscr{X}$ with proposal kernel $Q(x, \cdot)$. Show that for all $m \in \mathbb{N}$, $A \in \mathscr{X}$ and $x \in A$,

$$P^m(x,A) \le \sum_{i=0}^{m} \binom{m}{i} Q^i(x,A) \,. \tag{9.5.1}$$

9.19. Consider the random walk Metropolis algorithm introduced in Example 2.3.2. Denote by π the target distribution on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. We assume the following:

(i) π has a density, denoted by h_π, with respect to d-dimensional Lebesgue measure. Moreover, the density h_π is a continuous positive density function.

(ii) The increment distribution has density $\bar{q}$ with respect to d-dimensional Lebesgue measure that is continuous, positive, and bounded.

Let P be the kernel of a d-dimensional random walk Metropolis algorithm with target distribution π. Show that the unbounded sets are not small for P.

9.20. Let P be an irreducible and aperiodic kernel on $\mathsf{X} \times \mathscr{X}$ and let $m \in \mathbb{N}^*$. Show that there exists $m_0 \in \mathbb{N}^*$ such that P^r is strongly aperiodic for all integers $r \ge m_0$.

9.21. We use the notation of Exercise 9.19. Let $C \in \mathscr{B}(\mathbb{R}^d)$.

1. Show that for all $x \in \mathsf{X}$, $P(x, \{x\}^c) \le |q|_\infty / \pi(x)$.

Let C be an $(m, \varepsilon\nu)$-small set. Assume that π is unbounded on C.

1. Show that we may choose $x_1 \neq x_2 \in C$ such that $P^m(x_i, \{x_i\}) > (1 - \varepsilon/2)$, $i = 1, 2$.
2. Show that the set C is not small. [Hint: Use Exercise 9.15 to prove that for all $(x, x') \in C \times C$, $\|P^m(x, \cdot) - P^m(x', \cdot)\|_{\mathrm{TV}} \le (1 - \varepsilon)$.]

9.22. Let P be an irreducible kernel on $\mathsf{X} \times \mathscr{X}$. Assume that for some measurable function $V : \mathsf{X} \to [0, \infty]$, $\{V < \infty\}$ is accessible. Show that there exists n such that for all $k \ge n$, $\{V \le k\}$ is accessible.

9.6 Bibliographical Notes

The use of irreducibility as a basic concept for general state space Markov chains was initially developed by Doeblin (1940) in one of the very first studies devoted to the analysis of Markov chains of general state spaces. This idea was later developed by Doob (1953), Harris (1956), Chung (1964), Orey (1959), and Orey (1971). The concept of maximal irreducibility measure was introduced in Tweedie (1974a) and Tweedie (1974b). The results on full sets (Proposition 9.2.12) was established in Nummelin (1984).

A precursor of the notion of small set was introduced by Doeblin (1940) for Markov chains over general state spaces (see Meyn and Tweedie (1993a) for a modern exposition of these results). The existence of small sets as they are defined in this chapter was established by Jain and Jamison (1967). Our proof of Theorem 9.2.6 is borrowed from the monograph Orey (1971). Most of the existing proofs of these results reproduce the arguments given in this work.

Petite sets, as defined Section 9.4, were introduced in Meyn and Tweedie (1992). This definition generalizes previous versions introduced in Nummelin and Tuominen (1982) and Duflo (1997). These authors basically introduced the notion of petite set but with specific sampling distributions.

Note that our definition of irreducibility is not classical. A kernel is said to be irreducible if there exists a nontrivial irreducibility measure. We have adopted another equivalent point of view: a Markov kernel is said to be irreducible if there exists an accessible small set. These two definitions are equivalent by Theorem 9.2.6.

9.A Proof of Theorem 9.2.6

We preface the proof by several preliminary results and auxiliary lemmas.

Lemma 9.A.1 *Let $(E, \mathscr{B}, \mu)$ be a probability space and let $\{\mathscr{P}_n : n \in \mathbb{N}\}$ be a sequence of finite increasing partitions of E such that $\mathscr{B} = \sigma\{\mathscr{P}_n : n \in \mathbb{N}\}$. Let $A \in \mathscr{B}$ with $\mu(A) > 0$. Then for all $\varepsilon > 0$, there exists $U \in \bigcup_n \mathscr{P}_n$ such that $\mu(A \cap U) \geq (1-\varepsilon)\mu(U) > 0$.*

Proof. For $x \in E$, let E_x^n be the element of $\mathscr{P}_n$ that contains x, and set

$$X_n(x) = \begin{cases} \mu(A \cap E_x^n)/\mu(E_x^n) & \text{if } \mu(E_x^n) \neq 0 \,, \\ 0 & \text{if } \mu(E_x^n) = 0 \,. \end{cases}$$

Then X_n is a version of $\mathbb{E}\left[\mathbb{1}_A \,|\, \sigma(\mathscr{P}_n)\right]$. By the martingale theorem, $\lim_{n\to\infty} X_n = \mathbb{1}_A$ μ −a.s. Thus there exists $x \in A$ such that $\lim_n X_n(x) = 1$, and one can choose $U = E_x^n$ for an integer n sufficiently large. □

Proposition 9.A.2 *Let $(\mathsf{X}, \mathscr{X})$ be a measurable space such that the σ-field $\mathscr{X}$ is countably generated. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let $\mu \in \mathbb{M}_1(\mathscr{X})$. Then* *there exist a measurable function $f \in \mathbb{F}_+\left(\mathsf{X}^2, \mathscr{X}^{\otimes 2}\right)$ and a kernel N on $\mathsf{X} \times \mathscr{X}$ such that for all $x \in \mathsf{X}$, $N(x,\cdot)$ and μ are mutually singular and for all $x \in \mathsf{X}$, $P(x,\cdot) = f(x,\cdot)\cdot\mu + N(x,\cdot)$.*

Proof. For every $a \in \mathsf{X}$, by the Radon–Nikodym theorem, there exist a measurable function f_a on $(\mathsf{X}, \mathscr{X})$ and a measure $N(a,\cdot)$ such that μ and $N(a,\cdot)$ are mutually singular and

$$P(a,\cdot) = f_a \cdot \mu + N(a,\cdot)\,. \tag{9.A.1}$$

Since $\mathscr{X}$ is countably generated, there exists an increasing sequence of finite partitions $\mathscr{P}_n = \left\{A_{k,n} : 1 \le k \le m_n\right\}$, $n \ge 1$, such that $\mathscr{X} = \sigma(\mathscr{P}_n\,,\ n \in \mathbb{N})$. Set $\mathscr{X}_n = \sigma(\mathscr{P}_n)$ and write

$$f_n(a,x) = \sum_{k=1}^{m_n} \frac{P(a,A_{k,n})}{\mu(A_{k,n})} \mathbb{1}_{A_{k,n}}(x)\,.$$

Then (a,x) being fixed, $\{(f_n(a,x), \mathscr{X}_n),\ n \in \mathbb{N}\}$ is a nonnegative μ-martingale, and thus $\lim_{n\to\infty} f_n(a,x)$ exists μ − a.s. Denote by $f_\infty(a,x)$ this limit. Since $f_n \in \mathbb{F}_+\left(\mathsf{X}^2, \mathscr{X}^{\otimes 2}\right)$ for each $n \in \mathbb{N}$, we also have $f_\infty \in \mathbb{F}_+\left(\mathsf{X}^2, \mathscr{X}^{\otimes 2}\right)$. To complete the proof, it thus remains to show that $f_a = f_\infty(a,\cdot)$, μ − a.s. and that N is a kernel on $\mathsf{X} \times \mathscr{X}$.

For all $g \in \mathbb{F}_+(\mathsf{X})$, define $\mathbb{E}_\mu[g] = \int g \mathrm{d}\mu$. Then for all $A \in \mathscr{P}_n$, we get by (9.A.1),

$$\mathbb{E}_\mu\left[f_a \,|\, A\right] = \frac{1}{\mu(A)} \int_A f_a \mathrm{d}\mu \le \frac{P(a,A)}{\mu(A)}\,,$$

so that $\mathbb{E}_\mu\left[f_a \,|\, \mathscr{X}_n\right] \le f_n(a,\cdot)$, and letting n to infinity, we get $f_a \le f_\infty(a,\cdot)$ μ − a.s. We now turn to the converse inequality. By Fatou's lemma, for all $A \in \mathscr{P}_\ell$ and hence for all for all $A \in \cup_\ell \mathscr{P}_\ell$,

$$\int_A f_\infty(a,u)\mu(\mathrm{d}u) = \int_A \liminf_{n\to\infty} f_n(a,u)\mu(\mathrm{d}u) \le \liminf_{n\to\infty} \int_A f_n(a,u)\mu(\mathrm{d}u) \le P(a,A)\,. \tag{9.A.2}$$

To extend (9.A.2) to all $A \in \mathscr{X}$, note that $\cup_\ell \mathscr{P}_\ell$ is an algebra, $\mathscr{X} = \sigma(\cup_\ell \mathscr{P}_\ell)$, and

$$\int_{\mathsf{X}} f_\infty(a,u)\mu(\mathrm{d}u) + P(a,\mathsf{X}) \le 2P(a,\mathsf{X}) = 2 < \infty\,.$$

Then for all $A \in \mathscr{X}$ and all $\varepsilon > 0$, there exists $A_\varepsilon \in \cup_\ell \mathscr{P}_\ell$ such that

$$\int_{A\Delta A_\varepsilon} f_\infty(a,u)\mu(\mathrm{d}u) + P(a, A\Delta A_\varepsilon) \le \varepsilon\,.$$

Combining this with (9.A.2) yields that (9.A.2) holds for all $A \in \mathscr{X}$. Now choose $B \in \mathscr{X}$ such that $\mu(B^c) = 0$ and $N(a,B) = 0$. Then

$$\int_A f_\infty(a,u)\mu(\mathrm{d}u) = \int_{A\cap B} f_\infty(a,u)\mu(\mathrm{d}u) \le P(a, A\cap B) = \int_A f_a(u)\mu(\mathrm{d}u)\,.$$

Thus $f_\infty(a,\cdot) \le f_a \quad \mu -$a.s. Finally, $f_\infty(a,\cdot) = f_a \quad \mu -$a.s., and (9.A.1) holds with f_a replaced by $f_\infty(a,\cdot)$. The fact that N is indeed a kernel follows from $f_\infty \in \mathbb{F}_+\left(\mathsf{X}^2, \mathscr{X}^{\otimes 2}\right)$, P is a kernel on $\mathsf{X} \times \mathscr{X}$, and for all $A \in \mathscr{X}$,

$$N(\cdot, A) = P(\cdot, A) - \int_A f_\infty(\cdot, u)\mu(\mathrm{d}u) \; .$$

This completes the proof. □

Corollary 9.A.3 *Assume that the space $(\mathsf{X}, \mathscr{X})$ is separable. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $\mu \in \mathbb{M}_1(\mathscr{X})$. If for all $x \in \mathsf{X}$, $P(x,\cdot) \ll \mu$, then there exists $f \in \mathbb{F}_+(\mathscr{X}^{\otimes 2})$ such that for all $x \in \mathsf{X}$, $P(x,\cdot) = f(x,\cdot)\cdot\mu$.*

Lemma 9.A.4 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let ϕ be an irreducibility measure. For all $n \in \mathbb{N}$, there exist a bimeasurable function p_n on X^2 and a kernel S_n on $\mathsf{X} \times \mathscr{X}$ such that $S_n(x,\cdot)$ and ϕ are mutually singular for all $x \in \mathsf{X}$,*

$$P^n(x, \mathrm{d}y) = p_n(x,y)\phi(\mathrm{d}y) + S_n(x, \mathrm{d}y) \; , \tag{9.A.3}$$

and for all $m, n \in \mathbb{N}$ and $x, y \in \mathsf{X}$,

$$p_{m+n}(x,y) \ge \int_{\mathsf{X}} P^m(x, \mathrm{d}z) p_n(z,y) \ge \int_{\mathsf{X}} p_m(x,z) p_n(z,y) \phi(\mathrm{d}z) \; . \tag{9.A.4}$$

Moreover, for all $x \in \mathsf{X}$,

$$\sum_{n \ge 1} p_n(x,\cdot) > 0 \qquad \phi - a.e. \tag{9.A.5}$$

Proof. By Proposition 9.A.2, for every $n \ge 1$, there exist a bimeasurable function $p_n^0 : \mathsf{X}^2 \to \mathbb{R}_+$ and a kernel S_n on $\mathsf{X} \times \mathscr{X}$ such that for all $x \in \mathsf{X}$ and $A \in \mathscr{X}$,

$$P^n(x, A) = \int_{\mathsf{X}} p_n^0(x,y)\phi(\mathrm{d}x) + S_n(x, A) \; .$$

Define inductively the sequence of positive measurable functions $\{p_n,\ n \in \mathbb{N}\}$ on X^2 in the following way: set $p_1 = p_1^0$, and for all $n > 1$ and $x, y \in \mathsf{X}$, set

$$p_n(x,y) = p_n^0(x,y) \vee \sup_{1 \le k < n} \int_{\mathsf{X}} P^{n-k}(x, \mathrm{d}z) p_k(z,y) \; . \tag{9.A.6}$$

By construction, p_n satisfies the first inequality in (9.A.4). We now show by induction on $n \ge 1$ that for every $x \in \mathsf{X}$, $p_n(x,y) = p_n^0(x,y)$, for $\phi -$almost all y. Indeed, this is true for $n = 1$ by definition of p_1. For $n \ge 2$, assume that the induction assumption is true for $n-1$, i.e., for all $k = 1, \ldots, n-1$, $p_k(x,y) = p_k^0(x,y)$, $\phi -$almost all y. Then we have, for all $k = 1, \ldots, n-1$ and $(x, A) \in \mathsf{X} \times \mathscr{X}$,

$$
\begin{aligned}
P^n(x,A) &= \int P^{n-k}(x,\mathrm{d}z)P^k(z,A) \\
&\geq \int P^{n-k}(x,\mathrm{d}z)\int_A p_k^0(z,y)\phi(\mathrm{d}y) = \int_A \int P^{n-k}(x,dz)p_k^0(z,y)\phi(\mathrm{d}y) \,.
\end{aligned}
$$

Let the set $B_n^x \in \mathscr{X}$ be such that $\phi(B_n^x) = 0$ and $S_n(x,B_n^x) = S_n(s,\mathsf{X})$. Then applying the previous inequality and the induction assumption, we obtain

$$
\begin{aligned}
\int_A p_n^0(x,y)\phi(\mathrm{d}y) = P^n(x,A\setminus B_n^x) &\geq \int_{A\setminus B_n^x}\left(\int P^{n-k}(x,\mathrm{d}z)p_k^0(z,y)\right)\phi(\mathrm{d}y) \\
&= \int_A \left(\int P^{n-k}(x,\mathrm{d}z)p_k^0(z,y)\right)\phi(\mathrm{d}y) \\
&= \int_A \left(\int P^{n-k}(x,\mathrm{d}z)p_k(z,y)\right)\phi(\mathrm{d}y) \,.
\end{aligned}
$$

This implies that for all $1 \leq k \leq n$ and $x \in \mathsf{X}$,

$$
p_n^0(x,\cdot) \geq \int P^{n-k}(x,\mathrm{d}z)p_k(z,\cdot) \qquad \phi-\text{a.e.}
$$

Since the set A is arbitrary, this proves that the induction assumption is true for n. Therefore, for all $x \in \mathsf{X}$, (9.A.3) and the first inequality in (9.A.4) hold. This in turn proves the second inequality in (9.A.4).

We now prove the last statement. Fix one particular $x_0 \in \mathsf{X}$. Set $F = \cap_{n\geq 1}(B_n^{x_0})^c$, where $B_n^{x_0}$ was defined above. Then $\phi(F^c) \leq \sum_{n\geq 1}\phi(B_n^{x_0}) = 0$, and for every $n \geq 1$, $S_n(x_0,F) = 0$. Since ϕ is an irreducibility measure for P, for $B \subset F$ such that $\phi(B) > 0$, there exists m such that $P^m(x_0,B) = \int_B p_m(x_0,y)\phi(\mathrm{d}y) > 0$. This implies

$$
\int_B \sum_{n\geq 1} p_n(x_0,y)\phi(\mathrm{d}y) \geq \int_B p_m(x_0,y)\phi(\mathrm{d}y) > 0 \,.
$$

Since B is an arbitrary subset of F and $\phi(F^c) = 0$, this implies $\sum_{n\geq 1} p_n(x_0,\cdot) > 0$ ϕ-*a.e.* This proves (9.A.5). □

Let $H \in \mathscr{X}_{\mathbb{P}}^+$. We are going to prove that there exists an accessible small set D such that $D \subset H$. Let ψ be a maximal irreducibility measure. Then $\psi_H(\cdot) = \psi(\cdot \cap H)$ is an irreducibility measure. Therefore, by Remark 9.2.3, we can choose an irreducibility measure ϕ such that $\phi(H) = 1$ and $\phi(H^c) = 0$.

Let $F \in \mathscr{X}^{\otimes 2}$ be a set. Given $x,y \in \mathsf{X}$, we define the sections $F_1(x)$ and $F_2(y)$ by $F_1(x) = \{y \in \mathsf{X} : (x,y) \in F\}$ and $F_2(y) = \{x \in \mathsf{X} : (x,y) \in F\}$. This definition entails the identity $\mathbb{1}_{F_1(x)}(y) = \mathbb{1}_{F_2(y)}(x) = \mathbb{1}_F(x,y)$. For $A,B \in \mathscr{X}^{\otimes 2}$, define

$$
E_{A,B} = \left\{(x,y,z) \in \mathsf{X}^3 : (x,y) \in A, (y,z) \in B\right\} \,.
$$

Lemma 9.A.5 *Assume that* $\phi^{\otimes 3}(E_{A,B}) > 0$. *Then there exist* $C,D \in \mathscr{X}$ *such that* $\phi(C) > 0$, $\phi(D) > 0$ *and*

$$\inf_{x \in C, z \in D} \phi(A_1(x) \cap B_2(z)) > 0 .$$

Proof. Since $\mathscr{X}$ is countably generated, there exists a sequence of finite and increasing partitions $\mathscr{P}_n$ such that $\mathscr{X}^{\otimes 3} = \sigma(\bigcup_n \mathscr{P}_n^3)$. By Lemma 9.A.1, there exist an integer n and $U, V, W \in \mathscr{P}_n$ such that

$$\begin{aligned}\phi^{\otimes 3}(E_{A,B} \cap (U \times V \times W)) &= \int_{U \times V \times W} \mathbb{1}_A(x,y) \mathbb{1}_B(y,z) \phi(\mathrm{d}x) \phi(\mathrm{d}y) \phi(\mathrm{d}z) \\ &> \frac{3}{4} \phi(U) \phi(V) \phi(W) > 0 .\end{aligned}$$

This yields

$$\begin{aligned}\phi(W) \int_{U \times V} \mathbb{1}_A(x,y) \phi(\mathrm{d}x) \phi(\mathrm{d}y) &\geq \int_{U \times V \times W} \mathbb{1}_A(x,y) \mathbb{1}_B(y,z) \phi(\mathrm{d}x) \phi(\mathrm{d}y) \phi(\mathrm{d}z) \\ &> \frac{3}{4} \phi(U) \phi(V) \phi(W) .\end{aligned}$$

Since $\phi(W) > 0$, this implies

$$\int_{U \times V} \mathbb{1}_A(x,y) \phi(\mathrm{d}x) \phi(\mathrm{d}y) > \frac{3}{4} \phi(U) \phi(V) . \tag{9.A.7}$$

Similarly,

$$\int_{V \times W} \mathbb{1}_B(y,z) \phi(\mathrm{d}y) \phi(\mathrm{d}z) > \frac{3}{4} \phi(V) \phi(W) . \tag{9.A.8}$$

Define the sets C and D by

$$\begin{aligned}C &= \left\{ x \in U \,:\, \phi(A_1(x) \cap V) > \frac{3}{4} \phi(V) \right\} , \\ D &= \left\{ z \in W \,:\, \phi(B_2(z) \cap V) > \frac{3}{4} \phi(V) \right\} .\end{aligned}$$

Since $\int_{U \times V} \mathbb{1}_A(x,y) \phi(\mathrm{d}x) \phi(\mathrm{d}y) = \int_U \phi(A_1(x) \cap V) \phi(\mathrm{d}x)$, (9.A.7) yields $\phi(C) > 0$. Similarly, (9.A.8) yields $\phi(D) > 0$. Let μ be the probability measure on $\mathscr{X}$ defined by $\mu(G) = \phi(G \cap V)/\phi(V)$. Then for $x \in C$ and $z \in D$, $\mu(A_1(x)) > 3/4$ and $\mu(A_1(x)) > 3/4$, and since μ is a probability measure, this implies that $\mu(A_1(x) \cap A_1(x)) > 1/2$. Finally, this shows that for all $x \in C$ and $z \in D$,

$$\phi(A_1(x) \cap B_2(z)) \geq \phi(A_1(x) \cap B_2(z) \cap V) > \frac{1}{2} \phi(V) > 0 .$$

□

Proof (of Theorem 9.2.6). For every $x \in \mathsf{X}$, $\sum_{n \geq 1} p_n(x, \cdot) > 0$, $\phi -$ a.e. Therefore, there exist $r, s \in \mathbb{N}$ such that

$$\int_{\mathsf{X}^3} p_r(x,y)p_s(y,z)\phi(\mathrm{d}x)\phi(\mathrm{d}y)\phi(\mathrm{d}z) > 0.$$

For $\eta > 0$, define

$$F^\eta = \left\{(x,y) \in \mathsf{X}^2 \, : \, p_r(x,y) > \eta\right\} ,$$
$$G^\eta = \left\{(y,z) \in \mathsf{X}^2 \, : \, p_s(y,z) > \eta\right\} .$$

Then for $\eta > 0$ sufficiently small,

$$\phi^{\otimes 3}(\{(x,y,z) \, : \, (x,y) \in F^\eta, \, (y,z) \in G^\eta\}) > 0 .$$

Applying Lemma 9.A.5 with $A = F^\eta$ and $B = G^\eta$, we obtain that there exist $C, D \in \mathscr{X}$ such that $\phi(C) > 0$, $\phi(D) > 0$, and $\gamma > 0$ such that for all $x \in C$ and $z \in D$, $\phi(F_1^\eta(x) \cap G_2^\eta(z)) \geq \gamma > 0$. Then for all $u \in C$ and $v \in D$, by definition of F^η and G^η, we obtain

$$\begin{aligned} p_{r+s}(u,v) &\geq \int_{\mathsf{X}} p_r(u,y)p_s(y,v)\phi(\mathrm{d}y) \\ &\geq \int_{F_1^\eta(x)\cap G_2^\eta(z)} p_r(u,y)p_s(y,v)\phi(\mathrm{d}y) \geq \eta^2\gamma > 0 . \end{aligned} \tag{9.A.9}$$

Since ϕ is an irreducibility measure and $\phi(C) > 0$, C is accessible, thus for all $x \in \mathsf{X}$, $\sum_{k\geq 1} P^k(x,C) > 0$. Since $\phi(D) > 0$, we obtain $\int_D \phi(\mathrm{d}x) \sum_{k\geq 1} P^k(x,C) > 0$. Thus there exists $k \in \mathbb{N}^*$ such that $\int_D \phi(\mathrm{d}x)P^k(x,C) > 0$. This in turn implies that there exist $G \subset D$ and $\delta > 0$ such that $\phi(G) > 0$ and $P^k(x,C) \geq \delta$ for all $x \in G$. Since $\phi(G) > 0$, the set G is accessible. To conclude, we prove that G is a small set. Using (9.A.4) and (9.A.9), we have for all $x \in G$ and $z \in G \subset D$,

$$p_{r+s+k}(x,z) \geq \int_C P^k(x,\mathrm{d}y)p_{r+s}(y,z) \geq \delta\eta^2\gamma > 0 . \tag{9.A.10}$$

Finally, define $m = r + s + k$ and $\mu(\cdot) = \delta\eta^2\gamma\phi(\cdot \cap C)$. Then applying (9.A.3) and (9.A.10), we obtain, for all $x \in G$ and $B \in \mathscr{X}$,

$$P^m(x,B) \geq \int_B p_m(x,z)\phi(\mathrm{d}z) \geq \int_{B\cap C} p_m(x,z)\phi(\mathrm{d}z) \geq \mu(B) .$$

This proves that G is an (m,μ)-small set. Since $\phi(H^c) = 0$ and $\phi(G) > 0$, it follows that $\phi(H \cap G) > 0$. Since $H \cap G$ is a small set, $H \cap G$ is an accessible small set.

Therefore, we have established that if P admits an irreducibility measure, then every accessible set H contains an accessible small set G and therefore that P is irreducible. Conversely, if P is irreducible, then it admits an irreducibility measure by Proposition 9.1.9. □

Chapter 10
Transience, Recurrence, and Harris Recurrence

Recurrence and transience properties have already been examined in Chapters 6 and 7 for atomic and discrete Markov chains. We revisit these notions for irreducible Markov chains. Some of the properties we have shown for atomic chains extend quite naturally to irreducible chains. This is in particular true of the dichotomy between recurrent and transient chains (compare Theorem 6.2.7 or Theorem 7.1.2 with Theorem 10.1.5 below). Other properties are more specific such as Harris recurrence, which will be introduced in Section 10.2.

10.1 Recurrence and Transience

Recall the definitions of recurrent sets and kernels given in Definition 6.2.5.

Definition 10.1.1 (Recurrent set, recurrent kernel)

- *A set $A \in \mathscr{X}$ is said to be recurrent if $U(x,A) = \infty$ for all $x \in A$.*
- *A Markov kernel P on $\mathsf{X} \times \mathscr{X}$ is said to be recurrent if P is irreducible and every accessible set is recurrent.*

Theorem 10.1.2. *A Markov kernel P on $\mathsf{X} \times \mathscr{X}$ is recurrent if and only if it admits an accessible recurrent petite set.*

Proof. Let P be an irreducible Markov kernel. Then it admits an accessible small set C, and if P is recurrent, then by definition, C is recurrent.

Conversely, let C be a recurrent petite set. The kernel P being irreducible, Proposition 9.4.4 shows that there exist $a \in \mathbb{M}_1^*(\mathbb{N})$ and a maximal irreducibility measure ψ such that C is (a,ψ)-petite. Let A be an accessible set. Define

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_10

$\hat{A} = \{x \in A : U(x,A) = \infty\}$, and for $r \geq 1$, $A_r = \{x \in A : U(x,A) \leq r\}$. Then

$$A = \left(\bigcup_{r \geq 1} A_r \right) \cup \hat{A} \, .$$

By the maximum principle, we have

$$\sup_{x \in \mathsf{X}} U(x,A_r) \leq \sup_{x \in A_r} U(x,A_r) \leq \sup_{x \in A_r} U(x,A) \leq r \, .$$

Since by Lemma 4.2.3, $UK_a = K_a U \leq U$, we get that

$$r \geq U(x,A_r) \geq UK_a(x,A_r) \geq \int_C U(x,\mathrm{d}y) K_a(y,A_r) \geq U(x,C) \psi(A_r) \, .$$

Since $U(x,C) = \infty$ for $x \in C$, this implies that $\psi(A_r) = 0$ for all $r \geq 1$. Since A is accessible, we have $\psi(A) > 0$, and thus it must be the case that $\psi(\hat{A}) > 0$. Hence since ψ is an irreducibility measure, the set $\hat{A}$ is accessible. By Lemma 3.5.2, this implies that $K_{a_\varepsilon}(x,\hat{A}) > 0$ for all $x \in \mathsf{X}$ and $\varepsilon \in (0,1)$. Using again $U \geq K_{a_\varepsilon} U$, we obtain, for all $x \in \mathsf{X}$,

$$U(x,A) \geq K_{a_\varepsilon} U(x,A) \geq \int_{\hat{A}} K_{a_\varepsilon}(x,\mathrm{d}y) U(y,A) = \infty \times K_{a_\varepsilon}(x,\hat{A}) \, .$$

This implies that $U(x,A) = \infty$ for all $x \in A$, and A is recurrent. □

We have seen in Chapter 6 that a Markov kernel admitting an accessible atom is either recurrent or transient. In order to obtain a dichotomy between transient and recurrent irreducible kernels, we introduce the following definition.

Definition 10.1.3 (Uniformly transient set, transient set)

- *A set $A \in \mathscr{X}$ is called uniformly transient if* $\sup_{x \in A} U(x,A) < \infty$.
- *A set $A \in \mathscr{X}$ is called transient if $A = \bigcup_{n=1}^{\infty} A_n$, where A_n is uniformly transient.*
- *A Markov kernel P is said to be transient if X is transient.*

Proposition 10.1.4 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Then P is transient if and only if there exists an accessible uniformly transient set. Furthermore, if P is transient, every petite set is uniformly transient.*

Proof. Assume first that P is transient. By Definition 10.1.3, $\mathsf{X} = \bigcup_{n=1}^{\infty} A_n$, where for each $n \in \mathbb{N}^*$, the set A_n is uniformly transient. Among the collection $\{A_n, \; n \in \mathbb{N}\}$, there is at least one accessible set, showing that there exists an accessible uniformly transient set.

Assume now that there exists a uniformly transient set A. We will show that every petite set is uniformly transient. We will then conclude that P is transient, since we know by Proposition 9.4.5 that X is covered by an denumerable union of increasing petite sets. Let C be a petite set. By Proposition 9.4.4, we can choose the sampling distribution $a \in \mathbb{M}_1^*(\mathbb{N})$ and the measure ψ in such a way that C is an (a,ψ)-petite set, where ψ is a maximal irreducibility measure. By Lemma 4.2.3, we get for all $x \in \mathsf{X}$,

$$U(x,A) \geq UK_a(x,A) \geq \int_C U(x,\mathrm{d}y)K_a(y,A) \geq \psi(A)U(x,C) \; .$$

Thus since A is accessible and uniformly transient, the maximum principle yields

$$\sup_{x\in C} U(x,C) \leq \sup_{x\in\mathsf{X}} U(x,A)/\psi(A) = \sup_{x\in A} U(x,A)/\psi(A) < \infty \; .$$

Hence C is uniformly transient.

Since P is irreducible, by Proposition 9.1.8, X is a countable union of small, hence petite, sets. By the first statement, these sets are also uniformly transient. Hence X is transient. □

Theorem 10.1.5. *An irreducible kernel P on $\mathsf{X} \times \mathscr{X}$ is either recurrent or transient. Let C be an accessible (a,μ)-petite set with $\mu(C) > 0$.*

(i) If $\mu U(C) < \infty$, then P is transient.
(ii) If $\mu U(C) = \infty$, then P is recurrent.

Proof. Let ψ be a maximal irreducibility measure. Assume that P is not recurrent. Then there exist an accessible set A and $x_0 \in A$ such that $U(x_0,A) < \infty$. Set $\bar{A} = \{x \in \mathsf{X} : U(x,A) < \infty\}$. Since $PU \leq U$, we get for all $x \in \bar{A}$,

$$\int_{\bar{A}^c} P(x,\mathrm{d}y)U(y,A) \leq PU(x,A) \leq U(x,A) < \infty,$$

showing that $P(x,\bar{A}^c) = 0$. Hence $\bar{A}$ is absorbing, and since it is nonempty, $\bar{A}$ is full by Proposition 9.2.12. This implies that $\psi(\{x \in \mathsf{X} : U(x,A) = \infty\}) = 0$. Let $A_n = \{x \in A : U(x,A) \leq n\}$. Then $\psi(A_n) \uparrow \psi(A) > 0$. For sufficiently large n, A_n is accessible and uniformly transient, since

$$\sup_{x\in A_n} U(x,A_n) \leq \sup_{x\in A_n} U(x,A) \leq n \; .$$

Proposition 10.1.4 shows that the kernel P is transient.

Conversely, assume that the Markov kernel P is transient. In this case, $\mathsf{X} = \bigcup_{n=1}^{\infty} \mathsf{X}_n$, where for every $n \in \mathbb{N}^*$, X_n is uniformly transient. There exists $n \in \mathbb{N}^*$

such that $\psi(\mathsf{X}_n) > 0$. Hence X_n is accessible and uniformly transient, and therefore the kernel P cannot be recurrent.

Let C be an accessible (a,μ)-small set with $\mu(C) > 0$. If P is recurrent, then for all $x \in C$, $U(x,C) = \infty$, and hence since $\mu(C) > 0$, $\mu U(C) = \infty$. If P is transient, then $\mathsf{X} = \bigcup_{n=1}^{\infty} \mathsf{X}_n$, where X_n is uniformly transient. Proposition 10.1.4 shows that C is uniformly transient. Since $\sup_{x\in\mathsf{X}} U(x,C) \le \sup_{x\in C} U(x,C) < \infty$ (by the maximum principle, Theorem 4.2.2), we obtain that $\mu U(C) < \infty$. ☐

Theorem 10.1.6. *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. If P admits an invariant probability measure, then P is recurrent.*

Proof. Assume that P is transient. Then $\mathsf{X} = \bigcup_{m=1}^{\infty} B_m$, where B_m are uniformly transient sets. By the maximum principle, this implies that $\sup_{x\in\mathsf{X}} U(x,B_m) < \infty$. Let π be an invariant probability measure. Then for all integers $m,n \ge 1$, we have

$$n\pi(B_m) = \sum_{k=0}^{n-1} \pi P^k(B_m) \le \pi U(B_m) \le \sup_{x\in\mathsf{X}} U(x,B_m) < \infty\,.$$

Since the left-hand side remains bounded as n increases, this implies $\pi(B_m) = 0$ for all m and $\pi(\mathsf{X}) = 0$. This is a contradiction, since $\pi(\mathsf{X}) = 1$ by assumption. ☐

Proposition 10.1.7 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Every inaccessible set is transient.*

Proof. Let N be inaccessible. Then by definition, N^c is full, and by Proposition 9.2.12, N^c contains a full absorbing and accessible set H. Since $N \subset H^c$, it suffices to prove that H^c is transient. Set $A_{m,r} = \{x \in H^c : P^m(x,H) \ge 1/r\}$ for $m,r \ge 1$. Since the set H is accessible, it follows that

$$H^c = \bigcup_{m,r=1}^{\infty} A_{m,r}\,.$$

We now show that $A_{m,r}$ is uniformly transient. Since H is absorbing, for all $m \ge 1$ and $x \in \mathsf{X}$, we have $\mathbb{P}_x(X_m \in H) \le \mathbb{P}_x(\sigma_{H^c}^{(m)} = \infty)$, or equivalently, $\mathbb{P}_x(\sigma_{H^c}^{(m)} < \infty) \le \mathbb{P}_x(X_m \notin H)$. Therefore, for $m,r \ge 1$ and $x \in A_{m,r}$,

$$\mathbb{P}_x(\sigma_{A_{m,r}}^{(m)} < \infty) \le \mathbb{P}_x(\sigma_{H^c}^{(m)} < \infty) \le \mathbb{P}_x(X_m \notin H) \le 1 - 1/r = \delta\,.$$

As in the proof of Proposition 4.2.5, on applying the strong Markov property, we obtain

$$\sup_{x \in A_{m,r}} \mathbb{P}_x(\sigma_{A_{m,r}}^{(mn)} < \infty) \leq \delta^n .$$

This yields, for $x \in A_{m,r}$,

$$U(x,A_{m,r}) = 1 + \sum_{n=1}^{\infty} \mathbb{P}_x(\sigma_{A_{m,r}}^{(n)} < \infty) \leq 1 + m \sum_{n=1}^{\infty} \mathbb{P}_x(\sigma_{A_{m,r}}^{(mn)} < \infty) \leq 1 + \frac{m\delta}{1-\delta} .$$

Thus the sets $A_{m,r}$ are uniformly transient, and H^c and N are transient. □

The next result parallels Proposition 9.2.8 for transient sets.

Lemma 10.1.8 *Let $A \in \mathscr{X}$.*

(i) If A is uniformly transient and there exists $a \in \mathbb{M}_1(\mathbb{N})$ such that $\inf_{x \in B} K_a(x,A) > 0$*, then B is uniformly transient.*
(ii) If A is transient, then the set $\tilde{A}$ defined by

$$\tilde{A} = \{x \in \mathsf{X} : \mathbb{P}_x(\sigma_A < \infty) > 0\} \tag{10.1.1}$$

is transient.

Proof. (i) Set $\delta = \inf_{x \in B} K_a(x,A)$. Let A be a uniformly transient set. Lemma 4.2.3 implies that for all $x \in B$,

$$U(x,A) \geq UK_a(x,A) \geq \int_B U(x,\mathrm{d}y)K_a(y,A) \geq \delta U(x,B) .$$

By the maximum principle, Theorem 4.2.2, this yields

$$\sup_{x \in B} U(x,B) \leq \delta^{-1} \sup_{x \in B} U(x,A) \leq \delta^{-1} \sup_{x \in A} U(x,A) < \infty .$$

Thus B is uniformly transient.

(ii) If A is transient, it can be expressed as $A = \bigcup_{n=1}^{\infty} A_n$, where the set A_n is uniformly transient for each n. For $n,i,j \geq 1$, set

$$\tilde{A}_{n,i,j} = \left\{ x \in \mathsf{X} : \sum_{k=1}^{j} P^k(x,A_n) > 1/i \right\}$$

and $\tilde{A}_n = \bigcup_{i,j=1}^{\infty} \tilde{A}_{n,i,j}$. Applying (i) with the sampling distribution $a_j = j^{-1} \sum_{k=1}^{j} \delta_k$ yields that $\tilde{A}_{n,i,j}$ is uniformly transient and consequently that $\tilde{A}_n$ is transient. Since $A = \bigcup_{n \geq 1} A_n$, we have

$$\tilde{A} = \bigcup_{n=1}^{\infty} \{x \in \mathsf{X} : \mathbb{P}_x(\sigma_{A_n} < \infty) > 0\} = \bigcup_{n=1}^{\infty} \tilde{A}_n = \bigcup_{n,i,j \geq 1} \tilde{A}_{n,i,j} .$$

Therefore, the set $\tilde{A}$ is also transient.

□

Lemma 10.1.9 *Let P be a recurrent irreducible kernel on $\mathsf{X} \times \mathscr{X}$, ψ a maximal irreducibility measure, and $A \in \mathscr{X}_P^+$. Then $\psi(\{x \in A : \mathbb{P}_x(\sigma_A < \infty) < 1\}) = 0$.*

Proof. Set $B = \{x \in A : \mathbb{P}_x(\sigma_A < \infty) < 1\}$ and define for every $n \in \mathbb{N}^*$, $B_n = \{x \in A : \mathbb{P}_x(\sigma_A < \infty) < 1 - 1/n\}$. Since $B_n \subset A$, we have for all $x \in B_n$,

$$\mathbb{P}_x(\sigma_{B_n} < \infty) \le \mathbb{P}_x(\sigma_A < \infty) \le 1 - 1/n\,.$$

Applying then Proposition 4.2.5 (i), we get $\sup_{x \in B_n} U(x, B_n) \le n$, so that B_n is not recurrent. Since P is recurrent, this implies that $B_n \notin \mathscr{X}_P^+$. Using that $\bigcup_{n=1}^{\infty} B_n = B$, Proposition 9.2.9 therefore yields $\psi(B) = 0$. □

Theorem 10.1.10. *Let P be a recurrent Markov kernel on $\mathsf{X} \times \mathscr{X}$ and A an accessible set.*

(i) For every maximal irreducibility measure ψ, there exists $\tilde{A} \subset A$ such that $\psi(A \setminus \tilde{A}) = 0$ and $\mathbb{P}_x(N_{\tilde{A}} = \infty) = 1$ for all $x \in \tilde{A}$.
(ii) The set $A_\infty = \{x \in \mathsf{X} : \mathbb{P}_x(N_A = \infty) = 1\}$ is absorbing and full.

Proof. (i) We set

$$A_1 = \{x \in A : \mathbb{P}_x(\sigma_A < \infty) = 1\}\,, \qquad B_1 = \{x \in A : \mathbb{P}_x(\sigma_A < \infty) < 1\}\,.$$

Lemma 10.1.9 shows that $\psi(B_1) = 0$ for every maximal irreducibility measure. Set

$$\tilde{A} = \{x \in A_1 : \mathbb{P}_x(\sigma_{B_1} < \infty) = 0\}\,, \qquad B_2 = \{x \in A_1 : \mathbb{P}_x(\sigma_{B_1} < \infty) > 0\}\,.$$

Since $\psi(B_1) = 0$ and ψ is a maximal irreducibility measure, Proposition 9.2.8 shows that $\psi(B_2) = 0$. Furthermore, for $x \in \tilde{A}$, we get

$$0 = \mathbb{P}_x(\sigma_{B_1} < \infty) \ge \mathbb{P}_x(\sigma_{B_2} < \infty, \sigma_{B_1} \circ \theta_{\sigma_{B_2}} < \infty) = \mathbb{E}_x[\mathbb{1}_{\{\sigma_{B_2} < \infty\}} \mathbb{P}_{X_{\sigma_{B_2}}}(\sigma_{B_1} < \infty)]\,.$$

This proves that $\mathbb{P}_x(\sigma_{B_2} < \infty) = 0$. Hence for $x \in \tilde{A}$, we get $\mathbb{P}_x(\sigma_{B_i} < \infty) = 0$, $i = 1, 2$, and $\mathbb{P}_x(\sigma_A < \infty) = 1$, which implies $\mathbb{P}_x(\sigma_{\tilde{A}} < \infty) = 1$ and hence, by Proposition 3.3.6, $\mathbb{P}_x(N_{\tilde{A}} = \infty) = 1$.

(ii) We first prove that A_∞ is absorbing. For $x \in A_\infty$, we have

$$\begin{aligned} 1 = \mathbb{P}_x(N_A = \infty) &= \mathbb{P}_x(X_1 \in A_\infty, N_A = \infty) + \mathbb{P}_x(X_1 \in A_\infty^c, N_A = \infty) \\ &= \mathbb{P}_x(X_1 \in A_\infty) + \mathbb{P}_x(X_1 \in A_\infty^c, N_A = \infty)\,. \end{aligned}$$

This implies $\mathbb{P}_x(X_1 \in A_\infty^c) = \mathbb{P}_x(X_1 \in A_\infty^c, N_A = \infty)$, and by definition of A_∞, this is impossible unless $\mathbb{P}_x(X_1 \in A_\infty^c) = 0$. This proves that A_∞ is absorbing.
Since $\tilde{A} \subset A_\infty$ and $\psi(\tilde{A}) > 0$, it follows that $A_\infty \neq \emptyset$, and thus it is full.

□

We give here a drift criterion for transience. In the following section, we will exhibit a drift criterion for recurrence.

Theorem 10.1.11. *Let P be an irreducible kernel on $\mathsf{X} \times \mathscr{X}$. Assume that there exist a nonnegative bounded function V and $r \geq 0$ such that*

(i) the level sets $\{V \leq r\}$ and $\{V > r\}$ are both accessible,
(ii) $PV(x) \geq V(x)$ for all $x \in \{V > r\}$.

Then P is transient.

Proof. Define $C = \{V \leq r\}$ and

$$h(x) = \begin{cases} \{|V|_\infty - V(x)\}/\{|V|_\infty - r\} & x \notin C\,, \\ 1 & x \in C\,. \end{cases} \tag{10.1.2}$$

By construction, the function h is nonnegative, and for all $x \in \mathsf{X}$, we get

$$\begin{aligned} Ph(x) &= \int_C P(x,\mathrm{d}y)h(y) + \int_{C^c} P(x,\mathrm{d}y)h(y) \\ &= \int_{\mathsf{X}} P(x,\mathrm{d}y)\frac{|V|_\infty - V(y)}{|V|_\infty - r} + \int_C P(x,\mathrm{d}y)\left(1 - \frac{|V|_\infty - V(y)}{|V|_\infty - r}\right) \\ &= \frac{|V|_\infty - PV(x)}{|V|_\infty - r} + \int_C P(x,\mathrm{d}y)\frac{V(y) - r}{|V|_\infty - r} \leq \frac{|V|_\infty - PV(x)}{|V|_\infty - r}\,. \end{aligned}$$

Since $PV(x) \geq V(x)$ for all $x \in C^c$, the previous inequality implies that $Ph(x) \leq h(x)$ for all $x \in C^c$. Corollary 4.4.7 shows that $h(x) \geq \mathbb{P}_x(\tau_C < \infty)$ for all $x \in \mathsf{X}$. Since $h(x) < 1$ for all $x \in C^c$, this implies that

$$\mathbb{P}_x(\tau_C < \infty) = \mathbb{P}_x(\sigma_C < \infty) < 1\,,$$

for all $x \in C^c$. On the other hand, for $x \in C$, since $\mathbb{P}_x(\sigma_{C^c} < \infty) > 0$,

$$\mathbb{P}_x(\sigma_C = \infty) \geq \mathbb{P}_x(\sigma_C = \infty, \sigma_{C^c} < \infty) = \mathbb{E}_x\left[\mathbb{1}\{\sigma_{C^c} < \infty\}\mathbb{P}_{X_{\sigma_{C^c}}}(\sigma_C = \infty)\right] > 0\,.$$

Therefore, $\mathbb{P}_x(\sigma_C < \infty) < 1$ for all $x \in C$.

If P is recurrent, then Lemma 10.1.9 shows that $\mathbb{P}_x(\sigma_C < \infty) = 1$ for ψ-almost every $x \in C$ and every maximal irreducibility measure ψ, which contradicts the previous statement. □

We conclude this section by showing that if the kernel P is aperiodic, then the recurrence and transience of P are equivalent to the recurrence and transience of any of its skeletons P^m.

Proposition 10.1.12 *Let P be an irreducible and aperiodic Markov kernel on $\mathsf{X} \times \mathscr{X}$.*

1. *The Markov kernel P is transient if and only if one (and therefore every) m-skeleton P^m is transient.*
2. *The Markov kernel P is recurrent if and only if one (and therefore every) m-skeleton P^m is recurrent.*

Proof. The Chapman–Kolmogorov equations show that for every $A \in \mathscr{X}$ and $x \in \mathsf{X}$,

$$\sum_{j=1}^{\infty} P^j(x,\, A) = \sum_{r=1}^{m} \int P^r(x, \mathrm{d}y) \sum_j P^{jm}(y,\, A) \leq mM\,. \tag{10.1.3}$$

This elementary relation is the key equation in the proof.

(i) If A is a uniformly transient set for the m-skeleton P^m, with $\sum_j P^{jm}(x,\, A) \leq M$, then (10.1.3) implies that $\sum_{j=1}^{\infty} P^j(x,\, A) \leq mM$. Thus A is uniformly transient for P. Hence P is transient whenever a skeleton is transient. Conversely, if P is transient, then every P^k is transient, since

$$\sum_{j=1}^{\infty} P^j(x,\, A) \geq \sum_{j=1}^{\infty} P^{jk}(x,\, A)\,.$$

(ii) If the m-skeleton P^m is recurrent, then from (10.1.3), we again have that

$$\sum P^j(x,\, A) = \infty, \quad x \in \mathsf{X}, A \in \mathscr{X}_P^+\,,$$

so that the Markov kernel P is recurrent.

(iii) Conversely, suppose that P is recurrent. For every m, it follows from aperiodicity and Theorem 9.3.11 that P^m is irreducible; hence by Theorem 10.1.5, this skeleton is either recurrent or transient. If it were transient, we would have P transient from the previous question, which would lead to a contradiction.

□

10.2 Harris Recurrence

For atomic and discrete chains, we have seen in Theorem 6.2.2 that the recurrence in the sense of Definition 10.1.1 of an atom is equivalent to the property that the number of visits to the atom is infinite when starting from the atom. In the general case, this is no longer true, and we have to introduce the following definition.

Definition 10.2.1 (Harris recurrence) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$.*

(i) A set $A \in \mathscr{X}$ is said to be Harris recurrent if for all $x \in A$, $\mathbb{P}_x(N_A = \infty) = 1$.
(ii) The kernel P is said to be Harris recurrent if all accessible sets are Harris recurrent.

It is obvious from the definition that if a set is Harris recurrent, then it is recurrent. Harris recurrence is a strengthening of recurrence in the sense that it requires an almost sure infinite number of visits instead of an infinite expected number of visits to a set.

By Proposition 4.2.5, if for some $A \in \mathscr{X}$, $\mathbb{P}_x(\sigma_A < \infty) = 1$ for all $x \in A$, then $\mathbb{P}_x(\sigma_A^{(p)} < \infty) = 1$ for all $p \in \mathbb{N}^*$ and $x \in A$ and $\mathbb{P}_x(N_A = \infty) = 1$ for all $x \in A$. Then the set A is Harris recurrent. Conversely, if for all $x \in A$, $\mathbb{P}_x(N_A = \infty) = 1$, then $\mathbb{P}_x(\sigma_A < \infty) = 1$ for all $x \in A$. Therefore, a set A is Harris recurrent if and only if for all $x \in A$, $\mathbb{P}_x(\sigma_A < \infty) = 1$. The latter definition is often used.

We prove next that if P is Harris recurrent, then the number of visits to an accessible set is almost surely infinite starting from any point in the space.

Proposition 10.2.2 *If P is a Harris recurrent Markov kernel on $\mathsf{X} \times \mathscr{X}$, then for all $A \in \mathscr{X}_P^+$ and $x \in \mathsf{X}$, $\mathbb{P}_x(N_A = \infty) = 1$.*

Proof. Let A be an accessible Harris recurrent set and x_0 an arbitrary element of X. Set $B = \{x_0\} \cup A$. We have $\inf_{x \in B} \mathbb{P}_x(\sigma_A < \infty) = \delta > 0$, since $\inf_{x \in A} \mathbb{P}_x(\sigma_A < \infty) = 1$ and $\mathbb{P}_{x_0}(\sigma_A < \infty) > 0$. Thus by Theorem 4.2.6, $\mathbb{P}_{x_0}(N_B = \infty) \le \mathbb{P}_{x_0}(N_A = \infty)$. Since B is accessible, it is Harris recurrent under the stated assumption, which implies that $1 = \mathbb{P}_{x_0}(N_B = \infty) = \mathbb{P}_{x_0}(N_A = \infty)$. □

A Harris recurrent kernel is of course recurrent, but as illustrated by the next example, the converse does not hold.

Example 10.2.3 (Recurrent but not Harris recurrent). Let $\{a(n),\ n \in \mathbb{N}\}$ be a sequence of positive numbers such that $a(n) > 0$ for all $n \in \mathbb{N}$. We define a Markov kernel on $\mathsf{X} = \mathbb{N}$ by

$$P(0,0) = 1\ ,\ \ P(n,n+1) = \mathrm{e}^{-a(n)}\ ,\ \ P(n,0) = 1 - \mathrm{e}^{-a(n)}\ ,\ \ n \ge 1\ .$$

In words, this Markov chain either moves to the right with probability $\mathrm{e}^{-a(n)}$ or jumps back to zero, where it is absorbed. For $n \ge 1$, an easy calculation shows that

$$\mathbb{P}_n(\sigma_0 = \infty) = \mathrm{e}^{-\sum_{k=n}^{\infty} a(k)}\ , \qquad \mathbb{P}_n(\sigma_0 < \infty) = 1 - \mathrm{e}^{-\sum_{k=n}^{\infty} a(k)}\ .$$

The Markov kernel P is irreducible and $\{0\}$ is absorbing. Therefore, δ_0 is a maximal irreducibility measure, and every accessible set must contain 0. Let B be an accessible set and $1 \le n \in B$. Then $\mathbb{P}_n(N_B = \infty) = \mathbb{P}_n(\sigma_0 < \infty)$. If $\sum_{k=1}^{\infty} a(k) = \infty$, then for all $n \in \mathbb{N}$, $\mathbb{P}_n(\sigma_0 < \infty) = 1$, and the Markov kernel P is Harris recurrent. If $\sum_{k=1}^{\infty} a(k) < \infty$, then $0 < \mathbb{P}_n(\sigma_0 < \infty) = \mathbb{P}_n(N_B = \infty) < 1$, and hence $\mathbb{E}_n[N_B] = \infty$ for all $n \in \mathbb{N}$ and P is recurrent, but not Harris recurrent.

Proposition 10.2.4 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. If there exists a petite set C such that $\mathbb{P}_x(\sigma_C < \infty) = 1$ for all $x \notin C$, then P is Harris recurrent.*

Proof. By Proposition 3.3.6, the condition $\mathbb{P}_x(\sigma_C < \infty) = 1$ for all $x \notin C$ implies that $\mathbb{P}_x(\sigma_C < \infty) = 1$ and $\mathbb{P}_x(N_C = \infty) = 1$ for all $x \in \mathsf{X}$.

Let A be an accessible set. Since C is petite, Proposition 9.4.9 shows that A is uniformly accessible from C, so Theorem 4.2.6 implies $1 = \mathbb{P}_x(N_C = \infty) \le \mathbb{P}_x(N_A = \infty)$ for all $x \in \mathsf{X}$. Every accessible set is thus Harris recurrent, and P is Harris recurrent.
□

Definition 10.2.5 (Maximal absorbing set) *Let A be an absorbing set. The set A is said to be maximal absorbing if $A = \{x \in \mathsf{X} : \mathbb{P}_x(\sigma_A < \infty) = 1\}$.*

Recall that the set $A_+ = \{x \in \mathsf{X} : \mathbb{P}_x(\sigma_A < \infty) = 1\}$ is called the domain of attraction of the set A (see Definition 3.5.3). Then A is a maximal absorbing set if $A = A_+$.

Example 10.2.6. We continue with Example 10.2.3. The set $\{0\}$ is absorbing. If $\sum_{k=1}^{\infty} a(k) = \infty$, then for all $n \in \mathbb{N}$, $\mathbb{P}_n(\sigma_A < \infty) = 1$. Therefore, $\{0\}$ is not maximal absorbing. It is easy to see that $\mathbb{N}$ is the only maximal absorbing set. If $\sum_{k=1}^{\infty} a(k) = \ell < \infty$, then $\mathbb{P}_n(\sigma_A < \infty) < 1$ for all $n \in \mathbb{N}^*$. Hence $\{0\}$ is maximal absorbing.

Even though all Markov kernels may not be Harris recurrent, the following theorem provides a very useful decomposition of the state space of a recurrent Markov kernel.

Theorem 10.2.7. *Let P be a recurrent irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Then there exists a unique partition $\mathsf{X} = H \cup N$ such that*

(i) H is maximal absorbing,
(ii) N is transient,
(iii) the restriction of P to H is Harris recurrent.

For every accessible petite set C, we have

$$H = \{x \in \mathsf{X} : \mathbb{P}_x(N_C = \infty) = 1\} . \tag{10.2.1}$$

If P is not Harris recurrent, then the set N is nonempty and $\mathbb{P}_x(\sigma_H = \infty) > 0$ for all $x \in N$. Furthermore, for all petite sets $C \subset N$ and $x \in N$, $\mathbb{P}_x(N_C = \infty)$=0.

Proof. Let C be an accessible petite set. Define H by (10.2.1). By Theorem 10.1.10, H is absorbing and full. By definition, for every $x \in H_+$, $\mathbb{P}_x(\sigma_H < \infty) = 1$; thus

$$\mathbb{P}_x(N_C = \infty) \geq \mathbb{P}_x(N_C \circ \theta_{\sigma_H} = \infty) = \mathbb{E}_x\left[\mathbb{1}_{\{\sigma_H<\infty\}}\mathbb{P}_{X_{\sigma_H}}(N_C = \infty)\right] = 1 .$$

Therefore $x \in H$, and thus $H_+ \subset H$. Conversely, since H is absorbing, $\mathbb{P}_x(\sigma_H < \infty) = 1$ for all $x \in H$. Thus $H \subset H_+$. Therefore, H is maximal absorbing. Now let A be an accessible set for the restriction of P to H. This implies that A is accessible for P from every $x \in H$. Since H is itself accessible for the Markov kernel P from every $x \in \mathsf{X}$, this shows that A is also accessible for P (from every $x \in \mathsf{X}$). By Proposition 9.4.9, accessible sets A are uniformly accessible from petite sets. By Theorem 4.2.6, this implies that for all $x \in \mathsf{X}$, $\mathbb{P}_x(N_C = \infty) \leq \mathbb{P}_x(N_A = \infty)$. This yields, for all $x \in H$ and all accessible sets A,

$$1 = \mathbb{P}_x(N_C = \infty) \leq \mathbb{P}_x(N_A = \infty) .$$

Thus the restriction of P to H is Harris recurrent. The set $N = \mathsf{X} \setminus H$ is inaccessible, since H is full and is therefore transient by Proposition 10.1.7.

We will now establish that the decomposition is unique. Consider a partition $\mathsf{X} = H' \cup N'$ satisfying conditions (i)–(iii). The sets H and H' are full and absorbing; hence $H \cap H'$ is also full and absorbing, and $H \cap H'$ is accessible. Since the restriction of P to H' is Harris recurrent, it follows that for every $x \in H'$, $\mathbb{P}_x(\sigma_{H\cap H'} < \infty) = 1$. This shows that $H' \subset (H \cap H')_+ \subset H_+ = H$ (where the last equality holds because H is maximal absorbing). Reversing the roles of H and H', we finally get $H = H'$ and hence $N = N'$.

We finally prove the last statement. Assume that P is not Harris recurrent, i.e., N is not empty. If $x \in N$, then $\mathbb{P}_x(\sigma_H = \infty) > 0$, since H is maximal absorbing by assumption. Let $C \subset N$ be a petite set. By Proposition 9.4.9, the set H being accessible, H is uniformly accessible from C, i.e., $\inf_{x\in C}\mathbb{P}_x(\sigma_H < \infty) > 0$. For all $x \in N$, we have

$$\mathbb{P}_x(N_C = \infty) = \mathbb{P}_x(N_C = \infty, \sigma_H < \infty) + \mathbb{P}_x(N_C = \infty, \sigma_H = \infty) .$$

Since H is absorbing and $C \cap H = \emptyset$, we have that for every $x \in N$, $\mathbb{P}_x(N_C = \infty, \sigma_H < \infty) = 0$. On the other hand, since $\inf_{x\in C}\mathbb{P}_x(\sigma_H < \infty) > 0$, Theorem 4.2.6 yields that for all $x \in N$,

$$\mathbb{P}_x(N_C = \infty, \sigma_H = \infty) \leq \mathbb{P}_x(N_H = \infty, \sigma_H = \infty) = 0 .$$

Finally, for all $x \in N$, $\mathbb{P}_x(N_C = \infty) = 0$. The proof is complete. □

Corollary 10.2.8 *Let P be a recurrent irreducible Markov kernel. Every accessible set A contains an accessible Harris recurrent set B such that $A \setminus B$ is inaccessible.*

Proof. Write $\mathsf{X} = H \cup N$, where H and N are defined in Theorem 10.2.7, and choose $B = A \cap H$. □

Example 10.2.9. Consider again Example 10.2.3. The Dirac mass at 0 is a maximal irreducibility measure, and the set $\{0\}$ is full and absorbing. Moreover, P restricted to $\{0\}$ is Harris recurrent, and $\mathbb{N}^*$ is transient. This example shows that the decomposition of Theorem 10.2.7 is not always informative. ◀

We now give a criterion for Harris recurrence in terms of harmonic functions (which were introduced in Section 4.1). We preface the proof by a lemma, which is of independent interest.

Lemma 10.2.10 *Let P be a Harris recurrent irreducible kernel on $\mathsf{X} \times \mathscr{X}$. Let ψ be a maximal irreducibility measure. If h is a positive superharmonic function, then there exists $c \geq 0$ such that $h = c$ ψ-a.e. and $h \geq c$ everywhere.*

Proof. If h is not constant ψ-*a.e.*, then there exists $a < b$ such that $\psi(\{h < a\}) > 0$, $\psi(\{h > b\}) > 0$. For all initial distributions $\mu \in \mathbb{M}_1(\mathscr{X})$, $\{(h(X_n), \mathscr{F}_n^X), n \in \mathbb{N}\}$ is a positive $\mathbb{P}_\mu$-supermartingale, so $\{h(X_n), n \in \mathbb{N}\}$ converges $\mathbb{P}_\mu$ − a.s. to $Z = \limsup_{n\to\infty} h(X_n)$. Since P is Harris recurrent, every accessible set is visited infinitely often with probability one, for every initial distribution. Hence under $\mathbb{P}_\mu$, $\{h(X_n), n \in \mathbb{N}\}$ visits infinitely often the sets $\{h < a\}$ and $\{h > b\}$,

$$\mathbb{P}_\mu(\{h(X_n) < a, \text{i.o.}\}) = 1 = \mathbb{P}_\mu(\{h(X_n) > b, \text{i.o.}\}),$$

which results in a contradiction. Hence $\psi(\{x \in \mathsf{X} : h(x) = c\}^c) = 0$.

For every $\varepsilon > 0$, define $D_\varepsilon = \{x \in \mathsf{X} : c - \varepsilon < h(x) < c + \varepsilon\}$. The set D_ε is accessible, since $\{x \in \mathsf{X} : h(x) = c\}$ is accessible. Hence $\{h(X_n), n \in \mathbb{N}\}$ visits infinitely often D_ε, $\mathbb{P}_\mu(N_{D_\varepsilon} = \infty) = 1$, which implies that the limit of the sequence Z belongs to D_ε with probability 1: $\mathbb{P}_\mu(c - \varepsilon < Z < c + \varepsilon) = 1$. Since this result holds for every $\varepsilon > 0$, this also implies $\mathbb{P}_\mu(Z = c) = 1$. By Fatou's lemma, for all $x \in \mathsf{X}$,

$$c = \mathbb{E}_x\left[\lim_{n\to\infty} h(X_n)\right] \leq \liminf_{n\to\infty} \mathbb{E}_x[h(X_n)] \leq \mathbb{E}_x[h(X_0)] = h(x) .$$

□

Theorem 10.2.11. *Let P be an irreducible kernel on $\mathsf{X} \times \mathscr{X}$.*

(i) If every bounded harmonic function is constant, then P is either transient or Harris recurrent.
(ii) If P is Harris recurrent, then every bounded harmonic function is constant.

Proof. (i) Assume that P is not transient. Then by Theorem 10.1.5, it is recurrent. By Theorem 10.2.7, there exists a full absorbing set H such that for all $A \in \mathscr{X}_P^+$, $h(x) = \mathbb{P}_x(N_A = \infty) = 1$ for all $x \in H$.

The function $x \mapsto h(x) = \mathbb{P}_x(N_A = \infty)$ is harmonic by Proposition 4.2.4. If every harmonic function is constant, then $h(x) = \mathbb{P}_x(N_A = \infty) = 1$ for all $x \in \mathsf{X}$, that is, P is Harris recurrent.

(ii) Let h be a bounded harmonic function. The two functions $h + |h|_\infty$ and $|h|_\infty - h$ are positive and superharmonic on X. By Lemma 10.2.10, there exist c and c' such that $h + |h|_\infty = c$ ψ-*a.e.*, $h + |h|_\infty \geq c$, $|h|_\infty - h = c'$ ψ-*a.e.*, and $|h|_\infty - h \geq c'$. This implies $c - |h|_\infty = |h|_\infty - c'$ and $h \geq c - |h|_\infty = |h|_\infty - c' \geq h$. Therefore, h is constant. □

Theorem 10.2.12. *Let P be an irreducible Harris recurrent Markov kernel. Then the following hold:*

(i) If $A \in \mathscr{X}_P^+$, then $\mathbb{P}_x(N_A = \infty) = 1$ for all $x \in \mathsf{X}$.
(ii) If $A \notin \mathscr{X}_P^+$, then $\mathbb{P}_x(N_A = \infty) = 0$ for all $x \in \mathsf{X}$.

Proof. Assertion (i) is a restatement of the definition. Consider assertion (ii). Let $A \in \mathscr{X}$. The set $F = \{N_A = \infty\}$ is invariant. By Proposition 5.2.2, the function $x \mapsto h(x) = \mathbb{P}_x(F)$ is bounded and harmonic, and $h(X_n) \xrightarrow{\mathbb{P}_*\text{-a.s.}} \mathbb{1}_F$. Hence by Theorem 10.2.11, the function h is constant, and we have either $\mathbb{P}_x(F) \equiv 1$ or $\mathbb{P}_x(F) \equiv 0$. If $A \notin \mathscr{X}_P^+$, then Proposition 9.2.8 shows that $\{x \in \mathsf{X} : \mathbb{P}_x(\sigma_A < \infty) > 0\}$ is also inaccessible. Therefore, there exists $x \in \mathsf{X}$ such that $\mathbb{P}_x(F) = 0$, whence $\mathbb{P}_x(F) = 0$ for all $x \in \mathsf{X}$. □

We conclude this section by providing a sufficient drift condition for the Markov kernel P to be Harris recurrent.

Theorem 10.2.13. *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that there exist a function $V : \mathsf{X} \to [0, \infty)$ and a petite set C such that*

(i) the function V is superharmonic outside C, i.e., for all $x \notin C$, $PV(x) \leq V(x)$;
(ii) for all $r \in \mathbb{N}$, the level sets $\{V \leq r\}$ are petite.

Then P is Harris recurrent.

Proof. By Theorems 10.1.5 and 10.2.7, P is either transient or recurrent, and we can write $\mathsf{X} = H \cup N$ with $H \cap N = \emptyset$, N is transient, and H is either empty (if P is transient) or a maximal absorbing set (if P is recurrent), and in the latter case, the restriction of P to H is Harris recurrent. By Proposition 10.1.4 and Theorem 10.2.7, the set N has the following properties:

(a) for all $x \in N$, $\mathbb{P}_x(\sigma_H = \infty) > 0$;
(b) for all $x \in N$ and all petite sets $G \subset N$, $\mathbb{P}_x(N_G = \infty) = 0$.

Define the sequence $\{U_n,\ n \in \mathbb{N}\}$ by $U_n = V(X_n)\mathbb{1}\{\tau_C \geq n\}$. Since $PV(x) \leq V(x)$ for $x \notin C$, we get for $n \geq 1$,

$$\mathbb{E}\left[U_n \,|\, \mathscr{F}_{n-1}\right] = \mathbb{1}\{\tau_C \geq n\}\mathbb{E}\left[V(X_n) \,|\, \mathscr{F}_{n-1}\right] \leq \mathbb{1}\{\tau_C \geq n-1\}V(X_{n-1}) = U_{n-1} \,.$$

This implies that $\{U_n,\ n \in \mathbb{N}\}$ is a nonnegative supermartingale and therefore converges $\mathbb{P}_x$ − a.s. to a finite limit for all $x \in \mathsf{X}$.

For $r > 0$, set $G = \{V \leq r\} \cap N$. Then G is a petite set by assumption (ii). Thus property (b) implies that for all $x \in N$,

$$\begin{aligned}\mathbb{P}_x\left(\sum_{n=0}^{\infty}\mathbb{1}\{V(X_n) \leq r\} = \infty, \tau_H = \infty\right) &= \mathbb{P}_x\left(N_G = \infty, \tau_H = \infty\right) \\ &\leq \mathbb{P}_x\left(N_G = \infty\right) = 0 \,.\end{aligned}$$

This proves that for all $x \in N$ and all $r > 0$,

$$\mathbb{P}_x\left(\limsup_{n\to\infty} V(X_n) > r, \tau_H = \infty\right) = \mathbb{P}_x(\tau_H = \infty) \,. \tag{10.2.2}$$

By the monotone convergence theorem, this yields, for all $x \in N$,

$$\mathbb{P}_x\left(\limsup_{n\to\infty} V(X_n) = \infty, \tau_H = \infty\right) = \mathbb{P}_x(\tau_H = \infty) \,. \tag{10.2.3}$$

This equality obviously holds for $x \in H$, since both sides are then equal to zero, whence it holds for all $x \in \mathsf{X}$. Since $\mathbb{P}_x(\limsup_{n\to\infty} U_n < \infty) = 1$ for all $x \in \mathsf{X}$ and $U_n = V(X_n)$ on $\tau_C = \infty$, we have

$$\begin{aligned}&\mathbb{P}_x\left(\limsup_{n\to\infty} V(X_n) = \infty, \tau_C = \infty, \tau_H = \infty\right) \\ &\qquad = \mathbb{P}_x\left(\limsup_{n\to\infty} U_n = \infty, \tau_C = \infty, \tau_H = \infty\right) = 0 \,.\end{aligned} \tag{10.2.4}$$

Combining (10.2.3) and (10.2.4) yields, for all $x \in \mathsf{X}$,

$$\begin{aligned}\mathbb{P}_x(\tau_H = \infty) &= \mathbb{P}_x\left(\limsup_{n\to\infty} V(X_n) = \infty, \tau_C < \infty, \tau_H = \infty\right) \\ &\leq \mathbb{P}_x\left(\tau_C < \infty, \tau_H = \infty\right) \leq \mathbb{P}_x\left(\tau_H = \infty\right) .\end{aligned}$$

Therefore, for all $x \in \mathsf{X}$, $\mathbb{P}_x\left(\tau_C < \infty, \tau_H = \infty\right) = \mathbb{P}_x\left(\tau_H = \infty\right)$ and

$$\mathbb{P}_x\left(\tau_C = \infty, \tau_H = \infty\right) = 0 .$$

If $H \neq \emptyset$, then it is full and absorbing, and P restricted to H is Harris recurrent by assumption. Thus there is an accessible petite set $D \subset H$ such that $\mathbb{P}_x(\tau_D < \infty) = 1$ for all $x \in H$, which further implies that $\mathbb{P}_x(\tau_D = \infty, \tau_H < \infty) = 0$ for all $x \in \mathsf{X}$. If $H = \emptyset$ is empty, set $D = \emptyset$. Then in both cases, we have for all $x \in \mathsf{X}$,

$$\begin{aligned}\mathbb{P}_x(\tau_{C\cup D} = \infty) &= \mathbb{P}_x(\tau_{C\cup D} = \infty, \tau_H = \infty) + \mathbb{P}_x(\tau_{C\cup D} = \infty, \tau_H < \infty) \\ &\leq \mathbb{P}_x(\tau_C = \infty, \tau_H = \infty) + \mathbb{P}_x(\tau_D = \infty, \tau_H < \infty) = 0 .\end{aligned}$$

Since $C \cup D$ is petite by Proposition 9.4.5, we have proved that there exists a petite set F such that $\mathbb{P}_x(\tau_F = \infty) = 1$ for all $x \in \mathsf{X}$. By Proposition 10.2.4, this proves that P is Harris recurrent. □

We conclude this section by showing that if the kernel P is aperiodic, then P is Harris recurrent if and only if all its skeletons are Harris recurrent.

Proposition 10.2.14 *Let P be an irreducible and aperiodic Markov kernel on $\mathsf{X} \times \mathscr{X}$. The kernel P is Harris recurrent if and only if each m-skeleton P^m is Harris recurrent for every $m \geq 1$.*

Proof. (i) Assume that P^m is Harris recurrent. Since $m\sigma_{A,m} \geq \sigma_A$ for every $A \in \mathscr{X}$, where

$$\sigma_{A,m} = \inf\{k \geq 1 \,:\, X_{km} \in A\} , \tag{10.2.5}$$

it follows that P is also Harris recurrent.

(ii) Suppose now that P is Harris recurrent. For all $m \geq 2$, we know from Proposition 10.1.12 that P^m is recurrent; hence by Theorem 10.2.7, there exists a maximal absorbing set H_m for the m-skeleton P^m such that the restriction of P^m to H_m is Harris recurrent.

By Theorem 9.3.11, since P is aperiodic, $\mathscr{X}_P^+ = \mathscr{X}_{P^m}^+$. Since $H_m^c \notin \mathscr{X}_{P^m}^+$, it follows that $H_m \notin \mathscr{X}_P^+$, showing that H_m is full for P. Proposition 9.2.12 shows that since H_m is full, there exists a subset $H \subset H_m$ that is absorbing and full for P.

Since P is Harris recurrent, we have that for all $x \in \mathsf{X}$, $\mathbb{P}_x(\sigma_H < \infty) = 1$, and since H is absorbing, we know that $m\sigma_{H,m} \leq \sigma_H + m$ (where $\sigma_{H,m}$ is defined in (10.2.5)). This shows that for all $x \in \mathsf{X}$, $\mathbb{P}_x(\sigma_{H,m} < \infty) = \mathbb{P}_x(\sigma_H < \infty) = 1$.

Let $A \in \mathscr{X}_{P^m}^+ = \mathscr{X}_P^+$. By the strong Markov property, for every $x \in \mathsf{X}$, we have

$$\begin{aligned}\mathbb{P}_x(\sigma_{A,m} < \infty) &\geq \mathbb{P}_x(\sigma_{H,m} + \sigma_{A,m} \circ \sigma_{H,m} < \infty) \\ &= \mathbb{E}_x[\mathbb{1}_{\{\sigma_{H,m}<\infty\}} \mathbb{P}_{X_{\sigma_{H,m}}}(\sigma_{A,m} < \infty)] = \mathbb{P}_x(\sigma_{H,m} < \infty) = 1 .\end{aligned}$$

This shows that the m-skeleton P^m is Harris recurrent.

□

10.3 Exercises

10.1 (Random walk on $\mathbb{R}^d$). Consider $\{X_n,\ n \in \mathbb{N}\}$, a random walk on $\mathsf{X} = \mathbb{R}^d$, i.e., $X_n = X_{n-1} + Z_n$, where $\{Z_n,\ n \in \mathbb{N}^*\}$ is an i.i.d. sequence of $\mathbb{R}^d$-valued random variables defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Assume that the increment distribution μ is absolutely continuous with respect to the Lebesgue measure $\mu \ll \text{Leb}$ and that its support contains a ball centered at the origin $\text{supp}(\mu) \supset \text{B}(0,a)$ for some $a > 0$. Denote by g the density of μ with respect to the Lebesgue measure: $\mu = g \cdot \text{Leb}$. Let h be a bounded harmonic function. For all $x \in \mathbb{R}^d$ and $n \in \mathbb{N}$, set $M_n(x) = h(x + Z_1 + \ldots + Z_n)$.

1. Show that h is uniformly continuous on $\mathbb{R}^d$.
2. Show that $\limsup_{n\to\infty} M_n(x) = \liminf_{n\to\infty} M_n(x) = H(x)$ $\mathbb{P}-$a.s. and $M_n(x) = \mathbb{E}\left[H(x) | \mathscr{F}_n^Z\right]$ $\mathbb{P}-$a.s.
3. Show that $H(x) = h(x)$ $\mathbb{P}-$a.s. and that $h(x + Z_1) = h(x)$ $\mathbb{P}-$a.s. [Hint: Use the zero–one law.]
4. Show that every bounded harmonic function h is constant.
5. Show that P is either transient or Harris recurrent.

10.2. Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an invariant probability π. Assume that P admits a density p with respect to a σ-finite measure ν.

1. Show that $\pi \ll \nu$.
2. Show that P is Harris recurrent.

10.3. We use the notation of Section 2.3. Let $(\mathsf{X}, \mathscr{X})$ be a measurable space and ν a σ-finite measure on $\mathscr{X}$. Let h_π be a positive function satisfying $\nu(h_\pi) < \infty$. Let Q be a Markov kernel having density q with respect to ν i.e., $Q(x,A) = \int_A q(x,y)\nu(\mathrm{d}y)$ for every $x \in \mathsf{X}$ and $A \in \mathscr{X}$. Consider the Metropolis–Hastings kernel given by

$$P(x,A) = \int_A \alpha(x,y)q(x,y)\nu(\mathrm{d}y) + \bar{\alpha}(x)\delta_x(A),$$

where $\bar{\alpha}(x) = \int_{\mathsf{X}} \{1 - \alpha(x,y)\} q(x,y)\nu(\mathrm{d}y)$. Denote by π the measure $\pi = h_\pi \cdot \nu / \nu(h_\pi)$. Let h be a bounded harmonic function for P.

1. Show that P is recurrent and that $h = \pi(h)$ π-almost everywhere.
2. Show that $\{1 - \bar{\alpha}(x)\}\{h(x) - \pi(h)\} = 0$ for all $x \in \mathsf{X}$.
3. Show that $\bar{\alpha}(x) < 1$ for all $x \in \mathsf{X}$.

4. Show that P is Harris recurrent.

10.4. Suppose that π is a mixture of Gaussian distributions:

$$\pi = \sum_{i=1}^{\infty} 6\pi^{-2} i^{-2} \mathrm{N}(i,\, \mathrm{e}^{-i^2}).$$

Consider the Metropolis–Hastings algorithm that uses the following proposal. For $x \notin \mathbb{Z}_+$,

$$Q(x, \mathrm{d}y) = \frac{1}{\sqrt{2\pi}} \exp\{-(y-x)^2/2\} \mathrm{d}y\,,$$

which is an ordinary Gaussian random walk proposal. However, for $x \in \mathbb{Z}_+$, instead we propose

$$Q(x, \mathrm{d}y) = \frac{1}{x^2} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-(y-x)^2/2} \mathrm{d}y + (1 - \frac{1}{x^2}) \frac{1}{2} \{\delta_{x-1}(\mathrm{d}y) + \delta_{x+1}(\mathrm{d}y)\},$$

where $a \in (0,1)$.

1. Show that $N = \mathbb{Z}_+$ is transient.
2. Show that $H = \mathbb{R} \setminus \mathbb{Z}_+$ is maximal absorbing and that the restriction of P to H is Harris recurrent.

10.5 (Random walk on $\mathbb{R}^+$). Consider the Markov chain on $\mathbb{R}^+$ defined by

$$X_n = (X_{n-1} + W_n)^+, \tag{10.3.1}$$

where $\{W_n,\ n \in \mathbb{N}\}$ is an i.i.d. sequence of random variables with density q with respect to the Lebesgue measure. Assume that q is positive and lower semicontinuous and $\mathbb{E}\left[|W_1|\right] < \infty$.

1. Show that δ_0 is an irreducibility measure for P and that compact sets are small.

Assume first that $\mathbb{E}\left[W_1\right] < 0$.

2. Set $V(x) = x$ and let $x_0 < \infty$ be such that $\int_{-x_0}^{\infty} w\gamma(w)\mathrm{d}w < \mathbb{E}\left[W_1\right]/2 < 0$. Show that $V(x) = x$, for $x > x_0$,

$$PV(x) - V(x) \le \int_{-x_0}^{\infty} wq(w)\mathrm{d}w\,.$$

3. Show that the Markov kernel P is recurrent.

Assume now that $\mathbb{E}\left[W_1\right] = 0$ and $\mathbb{E}\left[W_1^2\right] = \sigma^2 < \infty$. We use the test function

$$V(x) = \begin{cases} \log(1+x) & \text{if } x > R\,, \\ 0 & \text{if } 0 \le x \le R\,, \end{cases} \tag{10.3.2}$$

where R is a positive constant to be chosen.

4. Show that for $x > R$,

$$PV(x) \le (1 - Q(R-x)) \log(1+x) + U_1(x) - U_2(x) ,$$

where Q is the cumulative distribution function of the increment distribution and

$$\begin{aligned} U_1(x) &= (1/(1+x))\mathbb{E}\left[W_1 \mathbb{1}\{W_1 > R - x\}\right] , \\ U_2(x) &= (1/(2(1+x)^2))\mathbb{E}\left[W_1^2 \mathbb{1}\{R - x < W_1 < 0\}\right] . \end{aligned}$$

5. Show that $U_1(x) = o(x^{-2})$ and

$$U_2(x) = (1/(2(1+x)^2))\mathbb{E}\left[W_1^2 \mathbb{1}\{W_1 < 0\}\right] - o(x^{-2}) .$$

6. Show that the Markov kernel is recurrent.

10.6 (Functional autoregressive models). Consider the first-order functional autoregressive model on $\mathbb{R}^d$ defined iteratively by

$$X_k = m(X_{k-1}) + Z_k , \tag{10.3.3}$$

where $\{Z_k,\ k \in \mathbb{N}\}$ is an i.i.d. sequence of random vectors independent of X_0 and $m : \mathbb{R}^d \to \mathbb{R}^d$ is a measurable function, bounded on every compact set. Assume that the distribution of Z_0 has density q with respect to Lebesgue measure on $\mathbb{R}^d$ that is bounded away from zero on every compact set.

Assume that $\mu_\beta = \mathbb{E}\left[\exp(\beta Z_1)\right] < \infty$ and that $\liminf_{|x|\to\infty} |m(x)| / |x| > 1$. Set $V(x) = 1 - \exp(-\beta |x|)$.

1. Show that the Markov kernel P associated with the recurrence (10.3.3) is given for all $x \in \mathbb{R}^d$ and $A \in \mathscr{B}(\mathbb{R}^d)$ by $P(x,A) = \int_A q(y - m(x))\mathrm{d}y$.
2. Show that every compact set with nonempty interior is 1-small.
3. Show that $PV(x) \le V(x) - W(x)$, where $\lim_{|x|\to\infty} W(x) = \infty$.
4. Show that the Markov kernel P is transient.

10.7. Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Show that P is transient if and only if there exist a bounded nonnegative function V and $C \in \mathscr{X}_P^+$ such that $PV(x) \ge V(x)$ for $x \notin C$ and $D = \left\{x \in \mathsf{X} :\ V(x) > \sup_{y \in C} V(y)\right\} \in \mathscr{X}_P^+$.

10.8. Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Suppose that P admits an invariant probability, denoted by π, i.e., $\pi P = \pi$.

1. Show that P is recurrent.
2. Let $A \in \mathscr{X}_P^+$. Show that $\mathbb{P}_y(N_A = \infty) = 1$ for π-almost every $y \in \mathsf{X}$.

Assume that there exists $m \in \mathbb{N}$ such that for all $x \in \mathbb{R}^d$, $P^m(x,\cdot) \ll \pi$, and let $r(x,y)$ denote the Radon–Nikodym derivative of $P^m(x,\cdot)$ with respect to π.

3. Show that P is Harris recurrent.

10.9. Let P be an irreducible Harris recurrent Markov kernel that admits a unique invariant probability measure π. Show that for all $Y \in \mathrm{L}^1(\Omega, \mathscr{F}, \mathbb{P}_\pi)$ and $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\frac{1}{n} \sum_{k=0}^{n-1} Y \circ \theta_k \stackrel{\mathbb{P}_\xi\text{-a.s.}}{\longrightarrow} \mathbb{E}_\pi[Y] .$$

10.4 Bibliographical Notes

We have closely followed in this Chapter the presentation given in Meyn and Tweedie (2009, Chapters 8 and 9). A great deal of work was devoted to characterizing the recurrence and transience of irreducible Markov kernels, and the presentation we give in this chapter focuses on the more important results and ignores many possible refinements.

On a countable state space, the recurrence–transience dichotomy that we generalize here is classical. Detailed expositions can be found in Chung (1967), Kemeny et al. (1976), and Norris (1998) among many others. Extensions to Markov chains on general state spaces was initiated in the 1960s. The early book by Orey (1971) already contains most of the results presented in this chapter, even though the exact terminology has changed a bit.

The notion of uniformly transient was introduced in Meyn and Tweedie (2009). Many closely related concepts appeared earlier in Tweedie (1974a), Tweedie (1974b). Some of the proof techniques are inherited from Nummelin (1978) and Nummelin (1984).

The concept of Harris recurrence was introduced in Harris (1956). The decomposition theorem, Theorem 10.2.7, which shows that recurrent kernels are “almost” Harris (the restriction to full absorbing set is Harris) was proved by Tuominen (1976) (earlier versions of this result can be found in Jain and Jamison (1967)).

The proof of the drift condition for Harris recurrence, Theorem 10.2.13, is borrowed from Fralix (2006).

Chapter 11
Splitting Construction and Invariant Measures

Chapter 6 was devoted to the study of Markov kernels admitting an accessible atom. The existence of an accessible atom had very important consequences, in particular for the existence and characterization of invariant measures. These results cannot be used if the state space does not admit an atom, which is the most frequent case for Markov kernels on a general state space.

The main goal of this chapter is to show that if P is an irreducible Markov kernel, that is, if P admits an accessible small set, then it is possible to define a kernel $\check{P}$ on an extended state space $(\check{\mathsf{X}}, \check{\mathscr{X}})$ that admits an atom and is such that P is the projection of $\check{P}$ onto X. This means that we can build a Markov chain $\{(X_k, D_k),\ k \in \mathbb{N}\}$ with kernel $\check{P}$ admitting an accessible atom and whose first component process $\{X_k,\ k \in \mathbb{N}\}$ is a Markov chain with kernel P. The chain $\{(X_k, D_k),\ k \in \mathbb{N}\}$ is referred to as the split chain, and its properties are directly related to those of the original chain. Most importantly, since $\check{P}$ admits an accessible atom, it admits a unique (up to scaling) invariant measure. In Section 11.2, we will use this measure to prove that P also admits a unique (up to scaling) invariant measure. In Sections 11.3 and 11.4, we will give results on the convergence in total variation distance of the iterates of the kernel by means of this splitting construction.

11.1 The Splitting Construction

Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting a $(1, \mu)$-small set C with $\mu(C) = 1$. Without loss of generality, we may assume that C is a $(1, 2\varepsilon\nu)$-small set with $\varepsilon \in (0, 1)$ and $\nu(C) = 1$. Using 2ε as a constant may seem arbitrary here, but we will see later in the construction the importance of this choice. Define the residual kernel R for $x \in \mathsf{X}$ and $A \in \mathscr{X}$ by

$$R(x, A) = \begin{cases} \{P(x, A) - \varepsilon\nu(A)\}/(1 - \varepsilon), & x \in C\,, \\ P(x, A), & x \notin C\,. \end{cases} \tag{11.1.1}$$

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_11

The splitting construction is based on the following decomposition of the Markov kernel P: for $x \in \mathsf{X}$ and $A \in \mathscr{X}$,

$$P(x,A) = \{1 - \varepsilon \mathbb{1}_C(x)\} R(x,A) + \varepsilon \mathbb{1}_C(x) \nu(A) \,. \tag{11.1.2}$$

Hence the kernel P is a mixture of two kernels with weights depending on x. It is worthwhile to note that the second kernel on the right-hand side of the previous equation does not depend on x. We will use this fundamental property to construct an atom.

The construction requires that one consider the extended state space $\check{\mathsf{X}} = \mathsf{X} \times \{0,1\}$, equipped with the associated product σ-field $\check{\mathscr{X}} = \mathscr{X} \otimes \mathscr{P}(\{0,1\})$. We first provide an *informal description* of a transition step of the split chain $\{(X_n, D_n),\ n \in \mathbb{N}\}$ associated to $\check{P}$.

- If $X_n \notin C$, then X_{n+1} is sampled from $P(X_n, \cdot)$.
- If $X_n \in C$ and $D_n = 0$, then X_{n+1} is sampled from $R(X_n, \cdot)$.
- If $X_n \in C$ and $D_n = 1$, then X_{n+1} is sampled from ν.
- The bell variable D_{n+1} is sampled from a Bernoulli distribution with success probability ε, independent of the past.

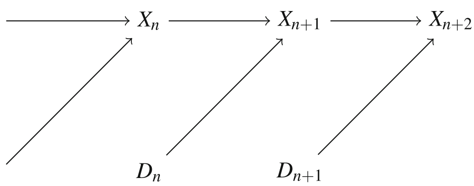

Fig. 11.1 Dependence graph of $\{(X_n, D_n),\ n \in \mathbb{N}\}$.

We now proceed with a rigorous construction of the split kernel $\check{P}$. Let b_ε be the Bernoulli distribution with success probability ε,

$$\mathrm{b}_\varepsilon = (1-\varepsilon)\delta_{\{0\}} + \varepsilon \delta_{\{1\}} \,. \tag{11.1.3}$$

For $f \in \mathbb{F}_+(\check{\mathsf{X}}, \check{\mathscr{X}}) \cup \mathbb{F}_b(\check{\mathsf{X}}, \check{\mathscr{X}})$, we define a function $\bar{f}_\varepsilon$ on X by

$$\bar{f}_\varepsilon(x) = [\delta_x \otimes \mathrm{b}_\varepsilon] f = (1-\varepsilon) f(x,0) + \varepsilon f(x,1) \,. \tag{11.1.4}$$

If $\check{\xi} \in \mathbb{M}_+(\check{\mathscr{X}})$ is a measure defined on the product space, we define the measure $\check{\xi}_0$ on $\mathscr{X}$ by

$$\check{\xi}_0(A) = \check{\xi}(A \times \{0,1\}) \,, \qquad A \in \mathscr{X} \,. \tag{11.1.5}$$

If for all $x \in \mathsf{X}$, $f(x,0) = f(x,1)$ (in words, f does not depend on the second component), then $\check{\xi}(f) = \check{\xi}_0(\bar{f}_\varepsilon)$. This definition also entails that for $\xi \in \mathbb{M}_+(\mathscr{X})$, $[\xi \otimes \mathrm{b}_\varepsilon]_0 = \xi$. Moreover, for $f \in \mathbb{F}_+(\check{\mathsf{X}}, \check{\mathscr{X}}) \cup \mathbb{F}_b(\check{\mathsf{X}}, \check{\mathscr{X}})$ and $\xi \in \mathbb{M}_+(\mathscr{X})$,

$$\xi(\bar{f}_\varepsilon) = [\xi \otimes \mathrm{b}_\varepsilon](f) \,. \tag{11.1.6}$$

We now define the split Markov kernel $\check{P}$ on $\check{\mathsf{X}} \times \check{\mathscr{X}}$ as follows. For $(x,d) \in \check{\mathsf{X}}$ and $\check{A} \in \check{\mathscr{X}}$, set

$$\check{P}(x,d;\check{A}) = Q(x,d;\cdot) \otimes \mathrm{b}_\varepsilon(\check{A}) \,, \tag{11.1.7}$$

where Q is the Markov kernel on $\check{\mathsf{X}} \times \mathscr{X}$ defined by, for all $B \in \mathscr{X}$,

$$Q(x,d;B) = \mathbb{1}_C(x)\left(\mathbb{1}_{\{0\}}(d)R(x,B) + \mathbb{1}_{\{1\}}(d)\nu(B)\right) + \mathbb{1}_{C^c}(x)P(x,B) \,. \tag{11.1.8}$$

Equivalently, for all $g \in \mathbb{F}_+(\mathsf{X},\mathscr{X}) \cup \mathbb{F}_b(\mathsf{X},\mathscr{X})$, we get

$$\begin{aligned} Qg(x,0) &= \mathbb{1}_C(x)Rg(x) + \mathbb{1}_{C^c}(x)Pg(x) \\ Qg(x,1) &= \mathbb{1}_C(x)\nu(g) + \mathbb{1}_{C^c}(x)Pg(x) \,. \end{aligned}$$

To stress the dependence of the splitting kernel on (ε,ν), we write $\check{P}_{\varepsilon,\nu}$ instead of $\check{P}$ whenever there is ambiguity.

It follows immediately from these definitions that for all $f \in \mathbb{F}_+(\check{\mathsf{X}},\check{\mathscr{X}}) \cup \mathbb{F}_b(\check{\mathsf{X}},\check{\mathscr{X}})$,

$$\check{P}f(x,d) = Q\bar{f}_\varepsilon(x,d) \,. \tag{11.1.9}$$

An important feature of this construction is that $\{D_n,\ n \in \mathbb{N}^*\}$ is an i.i.d. sequence of Bernoulli random variables with success probability ε that is independent of $\{X_n,\ n \in \mathbb{N}\}$. The essential property of the split chain is that if X_0 and D_0 are independent, then $\{(X_k, \mathscr{F}_k^X),\ k \in \mathbb{N}\}$ is a Markov chain with kernel P.

Lemma 11.1.1 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let C be a $(1,\varepsilon\nu)$-small set. For all $\xi \in \mathbb{M}_+(\mathscr{X})$ and $n \in \mathbb{N}$,*

$$[\xi \otimes \mathrm{b}_\varepsilon]\check{P}^n = \xi P^n \otimes \mathrm{b}_\varepsilon \,. \tag{11.1.10}$$

Proof. For $f \in \mathbb{F}_+(\check{\mathsf{X}},\check{\mathscr{X}})$, Fubini's theorem, (11.1.9), (11.1.2), and (11.1.6) yield

$$\begin{aligned} [\xi \otimes \mathrm{b}_\varepsilon]\check{P}f &= (1-\varepsilon)\xi(\mathbb{1}_C R\bar{f}_\varepsilon) + \varepsilon\xi(\mathbb{1}_C \nu(\bar{f}_\varepsilon)) + \xi(\mathbb{1}_{C^c}P\bar{f}_\varepsilon) \\ &= \xi(\mathbb{1}_C P\bar{f}_\varepsilon) + \xi(\mathbb{1}_{C^c}P\bar{f}_\varepsilon) = \xi P(\bar{f}_\varepsilon) = [\xi P \otimes \mathrm{b}_\varepsilon](f) \,. \end{aligned}$$

An easy induction yields the general result. □

We now consider the canonical chain associated with the kernel $\check{P}$ on $\check{\mathsf{X}} \times \check{\mathscr{X}}$. We adapt the notation of Section 3.1. For $\check{\mu} \in \mathbb{M}_1(\check{\mathscr{X}})$, we denote by $\check{\mathbb{P}}_{\check{\mu}}$ the probability measure on the canonical space $(\check{\mathsf{X}}^{\mathbb{N}}, \check{\mathscr{X}}^{\otimes\mathbb{N}})$ such that the coordinate process, denoted here by $\{(X_k,D_k),\ k \in \mathbb{N}\}$, is a Markov chain with initial distribution $\check{\mu}$ and Markov kernel $\check{P}$, called the split chain. We also denote by $\{\check{\mathscr{F}}_k,\ k \in \mathbb{N}\}$ and $\{\mathscr{F}_k^X,\ k \in \mathbb{N}\}$ the natural filtration of the canonical process $\{(X_k,D_k),\ k \in \mathbb{N}\}$ and of the process $\{X_k,\ k \in \mathbb{N}\}$, respectively.

In what follows, for all $g \in \mathbb{F}_+(\mathsf{X})$, define the function $g \otimes \mathbf{1} \in \mathbb{F}_+(\check{\mathsf{X}}, \check{\mathscr{X}})$ by

$$g \otimes \mathbf{1}(x,d) = g(x) \quad \text{for any } (x,d) \in \check{\mathsf{X}}. \tag{11.1.11}$$

Proposition 11.1.2 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let C be a $(1, \varepsilon\nu)$-small set. Set $\check{P} = \check{P}_{\varepsilon,\nu}$. Then for every $\xi \in \mathbb{M}_1(\mathscr{X})$, $\{(X_k, \mathscr{F}_k^X),\ k \in \mathbb{N}\}$ is under $\check{\mathbb{P}}_{\xi \otimes \mathrm{b}_\varepsilon}$ a Markov chain on $\mathsf{X} \times \mathscr{X}$ with initial distribution ξ and Markov kernel P.*

Proof. Write $\check{\xi} = \xi \otimes \mathrm{b}_\varepsilon$. For $g \in \mathbb{F}_+(\mathsf{X})$ and $n \geq 0$, we get, using (11.1.9), (11.1.2), and the obvious identity $\overline{\{g \otimes \mathbf{1}\}}_\varepsilon(x) = g(x)$,

$$
\begin{aligned}
&\check{\mathbb{E}}_{\check{\xi}}\left[g(X_{n+1}) \,|\, \mathscr{F}_n^X\right] \\
&= \check{\mathbb{E}}_{\check{\xi}}\left[\check{\mathbb{E}}_{\check{\xi}}\left[\{g \otimes \mathbf{1}\}(X_{n+1}) \,|\, \check{\mathscr{F}}_n\right] \,\Big|\, \mathscr{F}_n^X\right] = \check{\mathbb{E}}_{\check{\xi}}\left[\check{P}[g \otimes \mathbf{1}](X_n, D_n) \,|\, \mathscr{F}_n^X\right] \\
&= \mathbb{1}_C(X_n)\left[Rg(X_n)\check{\mathbb{P}}_{\check{\xi}}\left(D_n = 0 \,|\, \mathscr{F}_n^X\right) + \nu(g)\check{\mathbb{P}}_{\check{\xi}}\left(D_n = 1 \,|\, \mathscr{F}_n^X\right)\right] + \mathbb{1}_{C^c}(X_n)Pg(X_n) \\
&= \mathbb{1}_C(X_n)[(1-\varepsilon)Rg(X_n) + \varepsilon\nu(g)] + \mathbb{1}_{C^c}(X_n)Pg(X_n) = Pg(X_n)\,.
\end{aligned}
$$

□

We show that every invariant measure for $\check{P}$ can always be written as the product of an invariant measure for P and b_ε.

Proposition 11.1.3 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let C be a $(1, \varepsilon\nu)$-small set. Setting $\check{P} = \check{P}_{\varepsilon,\nu}$, we have the two following properties:*

(i) If $\lambda \in \mathbb{M}_+(\mathscr{X})$ is P-invariant, then $\lambda \otimes \mathrm{b}_\varepsilon$ is $\check{P}$-invariant.
(ii) If $\check{\lambda} \in \mathbb{M}_+(\check{\mathscr{X}})$ is $\check{P}$-invariant, then $\check{\lambda} = \check{\lambda}_0 \otimes \mathrm{b}_\varepsilon$, where $\check{\lambda}_0$ is defined in (11.1.5). In addition, $\check{\lambda}_0$ is P-invariant.

Proof. (i) If $\lambda \in \mathbb{M}_+(\mathscr{X})$ is P-invariant, then applying Lemma 11.1.1 yields $[\lambda \otimes \mathrm{b}_\varepsilon]\check{P} = \lambda P \otimes \mathrm{b}_\varepsilon = \lambda \otimes \mathrm{b}_\varepsilon$, showing that $\lambda \otimes \mathrm{b}_\varepsilon$ is $\check{P}$-invariant.

(ii) Assume now that $\check{\lambda}$ is $\check{P}$-invariant. Let $f, h \in \mathbb{F}_+(\check{\mathsf{X}}, \check{\mathscr{X}})$ be such that $\bar{f}_\varepsilon = \bar{h}_\varepsilon$. It follows from (11.1.9) that $\check{P}f = \check{P}h$, since these two quantities depend on f and h through $\bar{f}_\varepsilon$ and $\bar{h}_\varepsilon$ only. Since $\check{\lambda}$ is $\check{P}$-invariant, on applying the previous identity with $h = \bar{f}_\varepsilon \otimes \mathbf{1}$, we get

$$\check{\lambda}(f) = \check{\lambda}\check{P}(f) = \check{\lambda}\check{P}(\bar{f}_\varepsilon \otimes \mathbf{1}) = \check{\lambda}(\bar{f}_\varepsilon \otimes \mathbf{1}) = \check{\lambda}_0(\bar{f}_\varepsilon) = [\check{\lambda}_0 \otimes \mathrm{b}_\varepsilon](f)\,.$$

This identity holds for all $f \in \mathbb{F}_+(\check{\mathsf{X}}, \check{\mathscr{X}})$ thus $\check{\lambda} = \check{\lambda}_0 \otimes \mathrm{b}_\varepsilon$. Since $\check{\lambda}_0 \otimes \mathrm{b}_\varepsilon$ is $\check{P}$-invariant, Lemma 11.1.1 yields, for $g \in \mathbb{F}_+(\mathsf{X})$,

$$\check{\lambda}_0(g) = [\check{\lambda}_0 \otimes \mathrm{b}_\varepsilon](g \otimes \mathbf{1}) = [\check{\lambda}_0 \otimes \mathrm{b}_\varepsilon]\check{P}(g \otimes \mathbf{1}) = [\check{\lambda}_0 P \otimes \mathrm{b}_\varepsilon](g \otimes \mathbf{1}) = \check{\lambda}_0 P(g) ,$$

showing that $\check{\lambda}_0$ is P-invariant. □

The essential property of the split kernel $\check{P}$ stems from the fact that $\check{\alpha} = C \times \{1\}$ is an atom for the split kernel $\check{P}$, which inherits some properties of the set C.

Proposition 11.1.4 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let C be a $(1, 2\varepsilon\nu)$-small set with $\nu(C) = 1$. Setting $\check{P} = \check{P}_{\varepsilon,\nu}$, we have the following results:*

(i) The set $\check{\alpha} = C \times \{1\}$ is an aperiodic atom for $\check{P}$.
(ii) The set $C \times \{0, 1\}$ is small for the kernel $\check{P}$.
(iii) If C is accessible, then the atom $\check{\alpha}$ is accessible for $\check{P}$, and hence $\check{P}$ is irreducible.
(iv) For all $k \geq 1$, $\check{P}^k(\check{\alpha}, \check{\alpha}) = \varepsilon\nu P^{k-1}(C)$.
(v) If C is recurrent for P, then $\check{\alpha}$ is recurrent for $\check{P}$.
(vi) If C is Harris recurrent for P, then for all $\xi \in \mathbb{M}_1(\mathscr{X})$ satisfying $\mathbb{P}_\xi(\sigma_C < \infty) = 1$, $\check{\mathbb{P}}_{\xi\otimes\delta_d}(\sigma_{\check{\alpha}} < \infty) = 1$ for all $d \in \{0, 1\}$. Moreover, if P is Harris recurrent, then $\check{P}$ is Harris recurrent.
(vii) If C is accessible and P admits an invariant probability measure π, then $\check{\alpha}$ is positive for $\check{P}$.

Proof. (i) By definition, for every $(x, d) \in \check{\alpha}$ and $\check{A} \in \check{\mathscr{X}}$, we get $\check{P}(x, d; \check{A}) = [\nu \otimes \mathrm{b}_\varepsilon](\check{A})$. Thus $\check{\alpha}$ is an atom for $\check{P}$. Taking $\check{A} = \check{\alpha}$, we get for every $(x, d) \in \check{\alpha}$, $\check{P}(x, d; \check{\alpha}) = \varepsilon\nu(C) = \varepsilon > 0$, showing that the atom $\check{\alpha}$ is aperiodic.

(ii) For all $x \in C$ and $A \in \mathscr{X}$, $R(x, A) \geq \varepsilon\nu(A)$. Applying the identity (11.1.7) yields, for $x \in C$, $d \in \{0, 1\}$, and $\check{A} \in \check{\mathscr{X}}$,

$$\check{P}(x, d; \check{A}) \geq \varepsilon \mathbb{1}_{\{0\}}(d)\{\nu \otimes \mathrm{b}_\varepsilon\}(\check{A}) + \mathbb{1}_{\{1\}}(d)\{\nu \otimes \mathrm{b}_\varepsilon\}(\check{A}) \geq \varepsilon\{\nu \otimes \mathrm{b}_\varepsilon\}(\check{A}) .$$

(iii) For every $k \geq 1$ and $x \in \mathsf{X}$, since $\check{\mathbb{P}}_{(x,d)}(D_k = 1 | \mathscr{F}_k^X) = \varepsilon$, we get

$$\check{\mathbb{P}}_{(x,d)}((X_k, D_k) \in \check{\alpha}) = \check{\mathbb{P}}_{(x,d)}(X_k \in C, D_k = 1) = \varepsilon\check{\mathbb{P}}_{(x,d)}(X_k \in C) . \tag{11.1.12}$$

Since under $\check{\mathbb{P}}_{(x,d)}$ the law of (X_1, D_1) is $Q(x, d; \cdot) \otimes \mathrm{b}_\varepsilon$ (with Q defined in (11.1.8)), the Markov property implies

$$\check{\mathbb{P}}_{(x,d)}(X_k \in C) = \check{\mathbb{E}}_{(x,d)}[\check{\mathbb{E}}_{(X_1,D_1)}[\mathbb{1}_C(X_{k-1})]] = \check{\mathbb{P}}_{Q(x,d;\cdot)\otimes\mathrm{b}_\varepsilon}(X_{k-1} \in C) .$$

Applying Proposition 11.1.2, we finally get

$$\check{\mathbb{P}}_{(x,d)}(X_k \in C) = \mathbb{P}_{Q(x,d;\cdot)}(X_{k-1} \in C) = \int Q(x, d; \mathrm{d}x_1) P^{k-1}(x_1, C) . \tag{11.1.13}$$

Since $C \in \mathscr{X}_P^+$, Lemma 3.5.2 shows that for all $(x,d) \in \check{\mathsf{X}}$, there exists $k \in \mathbb{N}^*$ such that $\int Q(x,d;\mathrm{d}x_1)P^{k-1}(x_1,C) > 0$, which implies $\check{\mathbb{P}}_{(x,d)}((X_k,D_k) \in \check{\alpha}) > 0$, showing that the set $\check{\alpha}$ is accessible.

(iv) By (11.1.12), we have for all $x \in \mathsf{X}$ and $k \geq 1$, $\check{\mathbb{P}}_{(x,1)}((X_k,D_k) \in \check{\alpha}) = \varepsilon\check{\mathbb{P}}_{(x,1)}(X_k \in C)$. It follows from (11.1.8) that for $x \in C$, $Q(x,1;\cdot) = \nu(\cdot)$. Therefore, for all $x \in C$, (11.1.13) shows that for $k \geq 1$, $\check{\mathbb{P}}_{(x,1)}(X_k \in C) = \mathbb{P}_\nu(X_{k-1} \in C) = \nu P^{k-1}(C)$.

(v) Assume that C is recurrent for P. Since $\nu(C) = 1$ and C is recurrent, summing over $k \geq 1$ yields, for all $(x,1) \in \check{\alpha}$,

$$\sum_{k=1}^{\infty} \check{P}^k(\check{\alpha},\check{\alpha}) = \varepsilon \sum_{k=0}^{\infty} \nu P^k(C) = \varepsilon \int_C \nu(\mathrm{d}x)U(x,C) = \infty ,$$

showing that $\check{\alpha}$ is recurrent.

(vi) Recall that $\inf_{x\in C} P(x,C) \geq 2\varepsilon$. Then it follows from the definitions that

$$\inf_{x\in C} \check{P}(x,0;C\times\{1\}) = \varepsilon \inf_{x\in C} R(x,C) \geq \varepsilon^2 ,$$

and $\inf_{x\in C} \check{P}(x,1;C\times\{1\}) = \varepsilon$. Hence $\inf_{(x,d)\in C\times\{0,1\}} \check{\mathbb{P}}_{(x,d)}(X_1 \in \check{\alpha}) \geq \varepsilon^2$. Now assume that $\mathbb{P}_\xi(\sigma_C < \infty) = 1$. Proposition 11.1.2 shows that $\check{\mathbb{P}}_{\xi\otimes\mathrm{b}_\varepsilon}(\sigma_{C\times\{0,1\}} < \infty) = \mathbb{P}_\xi(\sigma_C < \infty) = 1$, and for all $(x,d) \in C\times\{0,1\}$, $\check{\mathbb{P}}_{\delta_x\otimes\mathrm{b}_\varepsilon}(\sigma_{C\times\{0,1\}} < \infty) = \mathbb{P}_x(\sigma_C < \infty) = 1$. For all $x \in C$, we have

$$\begin{aligned}
\check{\mathbb{P}}_{\xi\otimes\mathrm{b}_\varepsilon}(\sigma_{C\times\{0,1\}} < \infty) &= (1-\varepsilon)\check{\mathbb{P}}_{\xi\otimes\delta_0}(\sigma_{C\times\{0,1\}} < \infty) + \varepsilon\check{\mathbb{P}}_{\xi\otimes\delta_1}(\sigma_{C\times\{0,1\}} < \infty)\\
\check{\mathbb{P}}_{\delta_x\otimes\mathrm{b}_\varepsilon}(\sigma_{C\times\{0,1\}} < \infty) &= (1-\varepsilon)\check{\mathbb{P}}_{(x,0)}(\sigma_{C\times\{0,1\}} < \infty) + \varepsilon\check{\mathbb{P}}_{(x,1)}(\sigma_{C\times\{0,1\}} < \infty) .
\end{aligned}$$

Thus for $d \in \{0,1\}$, we have $\check{\mathbb{P}}_{\xi\otimes\delta_d}(\sigma_{C\times\{0,1\}} < \infty) = 1$ and $\check{\mathbb{P}}_{(x,d)}(\sigma_{C\times\{0,1\}} < \infty) = 1$ for all $(x,d) \in C\times\{0,1\}$. This, in turn, implies that $\check{\mathbb{P}}_{\xi\otimes\delta_d}(N_{C\times\{0,1\}} = \infty) = 1$. Since $\inf_{(x,d)\in C\times\{0,1\}} \check{P}(x,d;\check{\alpha}) \geq \varepsilon^2 > 0$, Theorem 4.2.6 implies that

$$1 = \check{\mathbb{P}}_{\xi\otimes\delta_d}(N_{C\times\{0,1\}} = \infty) = \check{\mathbb{P}}_{\xi\otimes\delta_d}(N_{\check{\alpha}} = \infty) .$$

(vii) By (ii) and Proposition 11.1.3, $\check{P}$ is irreducible and admits $\pi\otimes\mathrm{b}_\varepsilon$ as an invariant probability measure. Then by Theorem 10.1.6, the Markov kernel $\check{P}$ is recurrent. Then (ii) implies that $\check{\alpha}$ is accessible for the recurrent kernel $\check{P}$, hence recurrent. Applying Theorem 6.4.2 (iv) shows that the atom $\check{\alpha}$ is positive.

□

11.2 Existence of Invariant Measures

In this section we prove the existence and uniqueness (up to a scaling factor) of an invariant measure for a Markov kernel P admitting an accessible recurrent petite set. We start with the case in which the kernel P admits a strongly aperiodic accessible small set.

Proposition 11.2.1 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. If there exists an accessible, recurrent $(1,\mu)$-small set C with $\mu(C) > 0$, then P admits an invariant measure λ, unique up to multiplication by a positive constant and such that $0 < \lambda(C) < \infty$.*

Proof. Let C be an accessible, recurrent $(1,\mu)$-small set with $\mu(C) > 0$. Without loss of generality, we can assume that C is $(1,2\varepsilon\nu)$-small with $\nu(C) = 1$, which in particular implies that $\inf_{x \in C} P(x,C) \geq 2\varepsilon$. Consider $\check{P} = \check{P}_{\varepsilon,\nu}$, the split kernel defined in (11.1.7). According to Proposition 11.1.4, $\check{\alpha} = C \times \{1\}$ is an atom for $\check{P}$ that is accessible and recurrent for $\check{P}$. By Theorem 6.4.2, this implies the existence of an invariant measure $\check{\lambda}$ for $\check{P}$. Without loss of generality, we can assume that $\check{\lambda}(\check{\alpha})$=1. Define a measure λ on $\mathscr{X}$ by $\lambda(A) = \check{\lambda}_0(A) = \check{\lambda}(A \times \{0,1\})$. By Proposition 11.1.3, λ is invariant for P and $\check{\lambda} = \lambda \otimes \mathrm{b}_\varepsilon$. Let now λ' be another invariant measure for P. Then $\check{\lambda}' = \lambda' \otimes \mathrm{b}_\varepsilon$ is invariant for $\check{P}$ by Proposition 11.1.3. By Theorem 6.4.2, $\check{\lambda}'$ must then be proportional to $\check{\lambda}$, i.e., there exists $c > 0$ such that $\check{\lambda}' = c\check{\lambda}$. This yields, for every $A \in \mathscr{X}$,

$$\lambda'(A) = \check{\lambda}'(A \times \{0,1\}) = c\check{\lambda}(A \times \{0,1\}) = c\lambda(A) .$$

We now show that $0 < \lambda(C) < \infty$. Since $\check{\lambda}(\check{\alpha}) = 1$, we have $\lambda(C) = \check{\lambda}(C \times \{0,1\}) \geq \check{\lambda}(\check{\alpha}) = 1$. Thus $\lambda(C) > 0$. Moreover, since λ is P-invariant and C is $(1,\varepsilon\nu)$-small, $\lambda(C) < \infty$ by Lemma 9.4.12. □

We now extend this result to the case of an accessible recurrent m-small set. For this purpose, we need the following lemmas.

Lemma 11.2.2 *Let C be an accessible small set. Then C is an accessible $(1,\mu)$-small set with $\mu(C) > 0$ for the resolvent kernel K_{a_η} for all $\eta > 0$. Moreover, if C is recurrent for P, then it is also recurrent for K_{a_η}.*

Proof. Without loss of generality, we can assume by Lemma 9.1.6 that C is (m,μ)-small with $\mu(C) > 0$. For $\eta \in (0,1)$, $x \in C$, and $A \in \mathscr{X}$, we have

$$K_{a_\eta}(x,A) \geq (1-\eta)\eta^m P^m(x,A) \geq (1-\eta)\eta^m \mu(A) .$$

Thus C is a strongly aperiodic small set for K_{a_η}. Moreover, if C is accessible for P, then it is also accessible for K_{a_η} by Lemma 3.5.2.

Assume now that C is recurrent for P. We prove below that it is also recurrent for K_{a_η}. We first establish the identity

$$\sum_{n=1}^{\infty} K_{a_\eta}^n = \frac{1-\eta}{\eta} U \,, \tag{11.2.1}$$

where U is the potential kernel; see Definition 4.2.1. Indeed, by Lemma 1.2.11, $K_{a_\eta}^n = K_{a_\eta^{*n}}$, which implies

$$\sum_{n=1}^{\infty} K_{a_\eta}^n = \sum_{n=1}^{\infty} K_{a_\eta^{*n}} = \sum_{n=1}^{\infty}\sum_{k=0}^{\infty} a_\eta^{*n}(k) P^k = \sum_{k=0}^{\infty}\left(\sum_{n=1}^{\infty} a_\eta^{*n}(k)\right) P^k \,. \tag{11.2.2}$$

Moreover, for all $z \in (0,1)$,

$$\begin{aligned}\sum_{k=0}^{\infty}\left(\sum_{n=1}^{\infty} a_\eta^{*n}(k)\right) z^k &= \sum_{n=1}^{\infty}\left(\sum_{k=0}^{\infty} a_\eta^{*n}(k) z^k\right) = \sum_{n=1}^{\infty}\left((1-\eta)\sum_{k=0}^{\infty}\eta^k z^k\right)^n \\ &= \sum_{n=1}^{\infty}\left(\frac{1-\eta}{1-\eta z}\right)^n = \frac{1-\eta}{\eta(1-z)} = \frac{1-\eta}{\eta}\sum_{n=0}^{\infty} z^n \,.\end{aligned}$$

Thus for all $k \in \mathbb{N}$, $\sum_{n=1}^{\infty} a_\eta^{*n}(k) = (1-\eta)/\eta$. Plugging this identity into (11.2.2) proves (11.2.1). If C is recurrent for P, then $U(x,C) = \infty$ for all $x \in C$, and thus by (11.2.1), the set C is also recurrent for K_{a_η}. □

Lemma 11.2.3 *Let $\lambda \in \mathbb{M}_+(\mathscr{X})$ and $\eta \in (0,1)$. Then λ is invariant for P if and only if it is invariant for K_{a_η}.*

Proof. If $\lambda = \lambda P$, then $\lambda = \lambda K_{a_\eta}$. Conversely, assume that $\lambda = \lambda K_{a_\eta}$. The identity $K_{a_\eta} = (1-\eta) I + \eta K_{a_\eta} P$ yields $\lambda = (1-\eta)\lambda + \eta\lambda P$. Thus $\lambda(A) = \lambda P(A)$ for all $A \in \mathscr{X}$ such that $\lambda(A) < \infty$. Since by definition λ is σ-finite, this yields $\lambda P = \lambda$. □

Lemma 11.2.4 *Let P be an irreducible and recurrent Markov kernel on $\mathsf{X} \times \mathscr{X}$. Then every subinvariant measure is invariant. Let λ be an invariant measure and A an accessible set. Then for all $h \in \mathbb{F}_+(\mathsf{X})$,*

$$\lambda(h) = \int_A \lambda(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_A - 1} h(X_k)\right] = \int_A \lambda(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_A} h(X_k)\right] \,. \tag{11.2.3}$$

Proof. The proof consists in checking the assumptions of Theorem 3.6.5.

Let B be an accessible set. Using that λ is σ-finite, there exists $A \subset B$ such that $\lambda(A) < \infty$ and A is accessible. By Lemma 3.6.4 (iv), it suffices to prove that λ is invariant and that $\lambda = \lambda_A^0$, where λ_A^0 is defined in (3.6.1).

Let λ be a subinvariant measure and let A be an accessible set such that $\lambda(A) < \infty$. By Theorem 10.1.10, the set $A_\infty = \{x \in \mathsf{X} : \mathbb{P}_x(N_A = \infty) = 1\}$ is full and absorbing. Define $\tilde{A} = A \cap A_\infty$. Then for $x \in \tilde{A}$, $\mathbb{P}_x(\sigma_A < \infty) = 1$ and $\mathbb{P}_x(X_{\sigma_A} \in \tilde{A}) = 1$, since A_∞ is absorbing. Thus $\mathbb{P}_x(\sigma_A = \sigma_{\tilde{A}}) = 1$ for all $x \in \tilde{A}$. This implies that $\mathbb{P}_x(\sigma_{\tilde{A}} < \infty) = 1$

for all $x \in \tilde{A}$. Note also that since A_∞ is full and A is accessible, $\tilde{A} = A \cap A_\infty$ is accessible. We can therefore apply Theorem 3.6.5, and we obtain that λ is invariant and $\lambda = \lambda^0_{\tilde{A}}$, where $\lambda^0_{\tilde{A}}$ is defined in (3.6.1). Since $\tilde{A} \subset A$, by Lemma 3.6.4, this implies that $\lambda = \lambda^0_A = \lambda^1_A$. □

Theorem 11.2.5. *Let P be an irreducible and recurrent Markov kernel on $\mathsf{X} \times \mathscr{X}$. Then P admits a nonzero invariant measure λ, unique up to multiplication by a positive constant and such that $\lambda(C) < \infty$ for all petite sets C. Moreover, for every accessible set A and $h \in \mathbb{F}_+(\mathsf{X})$,*

$$\lambda(h) = \int_A \lambda(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_A-1} h(X_k)\right] = \int_A \lambda(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_A} h(X_k)\right]. \tag{11.2.4}$$

Proof. Since the kernel P is irreducible and recurrent, it admits a recurrent and accessible small set C. Then by Lemma 11.2.2, C is a $(1,\mu)$-small set with $\mu(C) > 0$ for the kernel K_{a_η} for every fixed $\eta \in (0,1)$. According to Proposition 11.2.1, K_{a_η} admits an invariant measure λ that is unique up to scaling and $0 < \lambda(C) < \infty$. By Lemma 11.2.3, this implies that λ is also the unique (up to scaling) invariant measure for P. Lemma 9.4.12 yields $\lambda(B) < \infty$ for all petite sets B, and (11.2.4) follows from Lemma 11.2.4. □

We have shown in Theorem 9.2.15 that an invariant probability measure is a maximal irreducibility measure. We now extend this property to possibly nonfinite measures.

Corollary 11.2.6 *Let P be an irreducible and recurrent Markov kernel on $\mathsf{X} \times \mathscr{X}$. Then an invariant measure is a maximal irreducibility measure.*

Proof. Let λ be an invariant measure. We show that $A \in \mathscr{X}_P^+$ if and only if $\lambda(A) > 0$. If A is an accessible set, then $K_{a_\varepsilon}(x,A) > 0$ for all $x \in \mathsf{X}$ and $\varepsilon \in (0,1)$. Since λ is invariant, $\lambda = \lambda K_{a_\varepsilon}$, showing that $\lambda(A) = \lambda K_{a_\varepsilon}(A) > 0$.

Conversely, assume that A is inaccessible. Then by Proposition 9.2.8, the set $\bar{A} = \{x \in \mathsf{X} : \mathbb{P}_x(\tau_A < \infty) > 0\}$ is also inaccessible. Set $A^0 = \bar{A}^c$. Hence A^0 is accessible, and we can therefore apply Theorem 11.2.5 to show that

$$\lambda(A) = \int_{A^0} \lambda(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_{A^0}} \mathbb{1}_A(X_k)\right].$$

Since $\mathbb{P}_x(\sigma_A = \infty) = 1$ for all $x \in A^0$, we obtain that $\mathbb{E}_x\left[\sum_{k=1}^{\sigma_{A^0}} \mathbb{1}_A(X_k)\right] = 0$, whence $\lambda(A) = 0$. □

We now address the existence of an invariant probability measure. We begin with a definition.

Definition 11.2.7 (Positive and null Markov kernel) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. If P is irreducible and admits an invariant probability measure π, the Markov kernel P is called positive. If P does not admit such a measure, then we call P null.*

Theorem 11.2.8. *Let P be an irreducible and recurrent Markov kernel on $\mathsf{X} \times \mathscr{X}$. Denote by λ a nonzero invariant measure for P. If there exists an accessible petite set C such that*

$$\int_C \lambda(\mathrm{d}x)\mathbb{E}_x[\sigma_C] < \infty \,, \tag{11.2.5}$$

then P is positive. Moreover, if $h \in \mathbb{F}_+(\mathsf{X})$ and $\int_C \lambda(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} h(X_k)\right] < \infty$, then $\lambda(h) < \infty$.

Proof. Since P is irreducible and recurrent, then by Theorem 11.2.5, P admits an invariant measure λ with $0 < \lambda(C) < \infty$, unique up to multiplication by a constant. Taking $h \equiv 1$ in (11.2.4) yields

$$\lambda(\mathsf{X}) = \int_C \lambda(\mathrm{d}x)\mathbb{E}_x[\sigma_C] \,.$$

This proves that λ is a finite measure and can be normalized to be a probability measure. Applying again Theorem 11.2.5, we obtain, for $h \in \mathbb{F}_+(\mathsf{X})$,

$$\lambda(h) = \int_C \lambda(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} h(X_k)\right] < \infty.$$

This proves the second statement. □

Corollary 11.2.9 *If P is an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ and if there exists a petite set C such that*

$$\sup_{x\in C} \mathbb{E}_x[\sigma_C] < \infty, \tag{11.2.6}$$

then P is positive.

Proof. First note that (11.2.6) implies that for all $x \in C$, $\mathbb{P}_x(\sigma_C < \infty) = 1$, and hence $\mathbb{P}_x(N_C = \infty) = 1$. Then for all $x \in C$, $U(x,C) = \infty$, and the set C is recurrent. On the other hand, Corollary 9.2.14 shows that the set C is also accessible. Then Theorem 10.1.2 applies, and the Markov kernel P is recurrent. By Theorem 11.2.5, P admits a nonzero invariant measure λ satisfying $\lambda(C) < \infty$ (since C is petite), unique up to multiplication by a constant. Together with (11.2.6), this implies (11.2.5). The proof is then completed by applying Theorem 11.2.8. □

11.3 Convergence in Total Variation to the Stationary Distribution

Theorem 8.2.6 shows that if P admits an accessible aperiodic positive atom α, then for all $\xi \in \mathbb{M}_1(\mathscr{X})$ satisfying $\mathbb{P}_\xi(\sigma_\alpha < \infty) = 1$, we have $\lim_{n\to\infty} \mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) = 0$, where π is the unique invariant probability measure. Using the splitting construction, we now extend this result to irreducible positive Markov kernels. We use below the notation introduced in the splitting construction (see Section 11.1).

Theorem 11.3.1. *Let P be a positive aperiodic Markov kernel on $\mathsf{X} \times \mathscr{X}$. Denote by π the unique invariant probability measure and H the maximal absorbing set such that the restriction of P to H is Harris recurrent (see Theorem 10.2.7). For all $\xi \in \mathbb{M}_1(\mathscr{X})$ satisfying $\xi(H^c)=0$, $\lim_{n\to\infty} \mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi)=0$.*

Proof. Since the restriction of P to H is Harris recurrent and H is maximal absorbing, we may assume without loss of generality that P is Harris recurrent. There exists an (m,μ)-accessible small set C with $\mu(C) > 0$ that is Harris recurrent. We consider separately two cases.

(I) Assume that C is $(1,\mu)$-small with $\mu(C) > 0$. Setting $\check{P} = \check{P}_{\varepsilon,\nu}$, Proposition 11.1.4 shows that $\check{P}$ is irreducible, and on applying Proposition 11.1.3, it turns out that $\pi \otimes \mathrm{b}_\varepsilon$ is the (unique) invariant probability measure of $\check{P}$. Proposition 11.1.4 then shows that $\check{\alpha} = C \times \{1\}$ is an accessible aperiodic positive atom for $\check{P}$. Let $\xi \in \mathbb{M}_1(\mathscr{X})$ be a probability measure. Since P is Harris recurrent, $\mathbb{P}_\xi(\sigma_C < \infty) = 1$ and Proposition 11.1.4 (vi) show that $\check{\mathbb{P}}_{\xi\otimes \mathrm{b}_\varepsilon}(\sigma_{\check{\alpha}} < \infty) = 1$. Theorem 8.2.6 then implies

$$\lim_{n\to\infty} \mathrm{d}_{\mathrm{TV}}([\xi \otimes \mathrm{b}_\varepsilon]\check{P}^n, \pi \otimes \mathrm{b}_\varepsilon) = 0 \,. \tag{11.3.1}$$

From Lemma 11.1.1 and Proposition 11.1.3, we get that $\mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) \le \mathrm{d}_{\mathrm{TV}}([\xi \otimes \mathrm{b}_\varepsilon]\check{P}^n, [\pi \otimes \mathrm{b}_\varepsilon])$. Hence $\lim_{n\to\infty} \mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) = 0$.

(II) Theorem 9.3.11 and Proposition 10.2.14 show that P^m is irreducible, aperiodic, and Harris recurrent. Moreover, the kernel P^m is positive with invariant distribution π. We can therefore apply the first part (I) to P^m, and we obtain $\lim_{n\to\infty} \mathrm{d}_{\mathrm{TV}}(\xi P^{nm}, \pi) = 0$. For all $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$, $\mathrm{d}_{\mathrm{TV}}(\xi P, \xi' P) \le \mathrm{d}_{\mathrm{TV}}(\xi, \xi')$ (see Lemma D.2.10). Thus we have, for all $r \in \{0, \ldots, m-1\}$,

$$\mathrm{d_{TV}}(\xi P^{nm+r}, \pi) \leq \mathrm{d_{TV}}(\xi P^{nm} P^r, \pi P^r) \leq \mathrm{d_{TV}}(\xi P^{nm}, \pi) ,$$

which concludes the proof.

□

We now extend this result to periodic Markov kernels.

Corollary 11.3.2 *Let P be a d-periodic Harris recurrent Markov kernel on $\mathsf{X} \times \mathscr{X}$ with an invariant probability π. Let $C_0, \ldots, C_{d-1}$ be a cyclic decomposition. For $k \in \{0, \ldots, d-1\}$, denote by π_k the probability on C_k given for all $A \in \mathscr{X}$ by $\pi_k(A) = \pi(A \cap C_k)/\pi(C_k)$.*

(i) For all $\xi \in \mathbb{M}_1(\mathscr{X})$ such that $\xi([\cup_{k=0}^{d-1} C_k]^c) = 0$ and all $j \geq 0$,

$$\lim_{n\to\infty} \left\| \xi P^{nd+j} - \sum_{k=0}^{d-1} \xi(C_k) \pi_{(k+j)[d]} \right\|_{\mathrm{TV}} = 0 .$$

(ii) For all $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\lim_{n\to\infty} \left\| d^{-1} \sum_{j=0}^{d-1} \xi P^{nd+j} - \pi \right\|_{\mathrm{TV}} = 0 . \tag{11.3.2}$$

If P is recurrent (but not necessarily Harris recurrent), then there exists an accessible Harris recurrent small set C, and (11.3.2) holds for every $\xi \in \mathbb{M}_1(\mathscr{X})$ satisfying $\mathbb{P}_\xi(\sigma_C < \infty) = 1$.

Proof. (i) Applying Theorem 11.3.1 to P^d on C_k, for $k \in \{0, \ldots, d-1\}$, we obtain for all $\nu \in \mathbb{M}_1(\mathscr{X})$ satisfying $\nu(C_k^c) = 0$,

$$\lim_{n\to\infty} \left\| \nu P^{dn} - \pi_k \right\|_{\mathrm{TV}} = 0 . \tag{11.3.3}$$

Note that for all $j \in \{0, \ldots d-1\}$, $\nu P^{dn+j} = \nu P^j P^{dn}$, and for $A \in \mathscr{X}$,

$$\begin{aligned}
\pi_k P^j(A) &= \frac{1}{\pi(C_k)} \int_{C_k} \pi(\mathrm{d}y) P^j(y, A) = \frac{1}{\pi(C_k)} \int_{C_k} \pi(\mathrm{d}y) P^j(y, A \cap C_{(k+j)[d]}) \\
&= \frac{1}{\pi(C_k)} \int \pi(\mathrm{d}y) P^j(y, A \cap C_{(k+j)[d]}) = \frac{\pi(A \cap C_{(k+j)[d]})}{\pi(C_k)} .
\end{aligned}$$

Since $\pi(C_k) = \pi(C_{(k+j)[d]})$, we get $\pi_k P^j = \pi_{(k+j)[d]}$, and (11.3.3) therefore implies $\lim_{n\to\infty} \left\| \nu P^{dn+j} - \pi_{k[d]} \right\|_{\mathrm{TV}} = 0$. Setting $\xi_k(A) = \xi(A \cap C_k)/\xi(C_k)$ if $\xi(C_k) > 0$, we obtain

$$\xi P^{dn+j} = \sum_{k = \xi(C_k) > 0} \xi(C_k) \xi_k P^{dn+j} ,$$

and the result follows.

(ii) If $\xi([\cup_{k=0}^{d-1} C_k]^c) = 0$, then (ii) follows from (i) by summation. Set

$$u_k(x) = \left\| d^{-1} \sum_{j=0}^{d-1} \delta_x P^{(k+j)} - \pi \right\|_{\mathrm{TV}} .$$

It follows from this definition that $u_k \leq 2$ and $\lim_{n\to\infty} u_{dn}(x) = 0$ for all $x \in C = \cup_{k=0}^{d-1} C_k$. Let $\xi \in \mathbb{M}_1(\mathscr{X})$. Since the Markov kernel P is Harris recurrent, $\mathbb{P}_\xi(\tau_C < \infty) = 1$. We have, for $h \in \mathbb{F}_b(\mathsf{X})$ such that $|h|_\infty \leq 1$,

$$\begin{aligned}
&\left| d^{-1} \sum_{j=0}^{d-1} \xi P^{nd+j}(h) - \pi(h) \right| = \left| \mathbb{E}_\xi \left[d^{-1} \sum_{j=0}^{d-1} h(X_{nd+j}) - \pi(h) \right] \right| \\
&\leq 2\mathbb{P}_\xi(\tau_C > nd) + \left| \mathbb{E}_\xi \left[\mathbb{1}_{\{\tau_C \leq nd\}} \left\{ d^{-1} \sum_{j=0}^{d-1} h(X_{nd-\tau_C+j}) \circ \theta_{\tau_C} - \pi(h) \right\} \right] \right| \\
&= 2\mathbb{P}_\xi(\tau_C > nd) + \left| \mathbb{E}_\xi \left[\mathbb{1}_{\{\tau_C \leq nd\}} \left\{ d^{-1} \sum_{j=0}^{d-1} P^{nd-\tau_C+j} h(X_{\tau_C}) - \pi(h) \right\} \right] \right| \\
&\leq 2\mathbb{P}_\xi(\tau_C > nd) + \mathbb{E}_\xi \left[\mathbb{1}_{\{\tau_C \leq nd\}} u_{nd-\tau_C}(X_{\tau_C}) \right] \to_{n\to\infty} 0
\end{aligned}$$

by Lebesgue's dominated convergence theorem uniformly in $|h| \leq 1$.

□

11.4 Geometric Convergence in Total Variation Distance

We have shown Section 8.2.2 that if the kernel P is aperiodic and admits an atom α such that $\mathbb{E}_\alpha[\beta^{\sigma_\alpha}] < \infty$ for some $\beta > 1$, then there exists $\delta \in (1, \beta)$ such that $\sum_{k=1}^{\infty} \delta^n \mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) < \infty$ for all initial distributions satisfying $\mathbb{E}_\xi[\delta^{\sigma_\alpha}] < \infty$. We will show that this result extends to the irreducible and aperiodic Markov kernels on a general state space using the splitting method.

Before going further, we will establish a result that will play a crucial role in the proof.

Theorem 11.4.1. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, $C \in \mathscr{X}$, and let ρ, τ be two stopping times with $\tau \geq 1$. Assume that for all $n \in \mathbb{N}$,*

$$\rho \leq n + \rho \circ \theta_n , \qquad \text{on } \{\rho > n\} . \tag{11.4.1}$$

Moreover, assume that there exists $\gamma > 0$ such that for all $x \in C$,

$$\mathbb{P}_x(\tau < \infty, X_\tau \in C) = 1 , \qquad \mathbb{P}_x(\rho \leq \tau) \geq \gamma . \tag{11.4.2}$$

Then the following hold:

(i) For all $x \in C$, $\mathbb{P}_x(\rho < \infty) = 1$.

(ii) If $\sup_{x \in C} \mathbb{E}_x[\beta^\tau] < \infty$ for some $\beta > 1$, then there exist $\delta \in (1, \beta)$ and $\varsigma < \infty$ such that for all $h \in \mathbb{F}_+(\mathsf{X})$,

$$\sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\rho-1} \delta^k h(X_k) \right] \le \varsigma \sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\tau-1} \beta^k h(X_k) \right] . \tag{11.4.3}$$

Proof. Define $\tau^{(0)} = 0$, $\tau^{(1)} = \tau$, and for $n \ge 1$, $\tau^{(n)} = \tau^{(n-1)} + \tau \circ \theta_{\tau^{(n-1)}}$. Using (11.4.2), the strong Markov property shows that for all $k \in \mathbb{N}$ and $x \in C$, $\mathbb{P}_x(\tau^{(k)} < \infty, X_{\tau^{(k)}} \in C) = 1$.

(i) Using (11.4.1), we get

$$\{\rho > \tau^{(k)}, \tau^{(k)} < \infty\} \subset \{\rho > \tau^{(k-1)}, \rho \circ \theta_{\tau^{(k-1)}} > \tau \circ \theta_{\tau^{(k-1)}}, \tau^{(k-1)} < \infty\} . \tag{11.4.4}$$

The strong Markov property then yields, for $x \in C$,

$$\begin{aligned} \mathbb{P}_x(\rho > \tau^{(k)}) &\le \mathbb{P}_x(\rho > \tau^{(k-1)}, \rho \circ \theta_{\tau^{(k-1)}} > \tau \circ \theta_{\tau^{(k-1)}}) \\ &\le (1-\gamma)\mathbb{P}_x(\rho > \tau^{(k-1)}) . \end{aligned} \tag{11.4.5}$$

By induction, this yields for $x \in C$,

$$\mathbb{P}_x(\rho > \tau^{(k)}) \le (1-\gamma)^k . \tag{11.4.6}$$

Therefore, $\mathbb{P}_x(\rho = \infty) \le \lim_{k \to \infty} \mathbb{P}_x(\rho > \tau^{(k)}) = 0$, i.e., $\mathbb{P}_x(\rho < \infty) = 1$ for all $x \in C$.

(ii) For $h \in \mathbb{F}_+(\mathsf{X})$ and $\delta \in (1, \beta]$, we set

$$M(h, \delta) = \sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\tau-1} \delta^k h(X_k) \right] . \tag{11.4.7}$$

Using the strong Markov property, we get

$$\begin{aligned} \mathbb{E}_x \left[\sum_{k=0}^{\rho-1} \delta^k h(X_k) \right] &\le \sum_{k=0}^{\infty} \mathbb{E}_x \left[\mathbb{1}_{\{\rho > \tau^{(k)}\}} \delta^{\tau^{(k)}} \mathbb{E}_{X_{\tau^{(k)}}} \left[\sum_{j=0}^{\tau-1} \delta^j h(X_j) \right] \right] \\ &\le M(h, \beta) \sum_{k=0}^{\infty} \mathbb{E}_x \left[\mathbb{1}_{\{\rho > \tau^{(k)}\}} \delta^{\tau^{(k)}} \right] . \end{aligned} \tag{11.4.8}$$

Note that this inequality remains valid even if $M(h, \beta) = \infty$. By Jensen's inequality, for all $x \in C$,

$$\mathbb{E}_x[\delta^\tau] \le \{\mathbb{E}_x[\beta^\tau]\}^{\log(\delta)/\log(\beta)} \le \Phi(\delta) := \left\{\sup_{x\in C}\mathbb{E}_x[\beta^\tau]\right\}^{\log(\delta)/\log(\beta)} .$$

By the strong Markov property, we further have $\sup_{x\in C}\mathbb{E}_x\left[\delta^{\tau^{(k)}}\right] \le \{\Phi(\delta)\}^k$. Since $\lim_{\delta\to 1}\Phi(\delta)=1$, we can choose $1<\delta\le\sqrt{\beta}$ such that $(1-\gamma)\Phi(\delta^2)<1$. For every $x\in C$, applying (11.4.6) and the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}\sum_{k=0}^{\infty}\mathbb{E}_x\left[\mathbb{1}\left\{\rho>\tau^{(k)}\right\}\delta^{\tau^{(k)}}\right] &\le \sum_{k=0}^{\infty}\left\{\mathbb{P}_x(\rho>\tau^{(k)})\right\}^{1/2}\left\{\mathbb{E}_x\left[\delta^{2\tau^{(k)}}\right]\right\}^{1/2}\\ &\le \sum_{k=0}^{\infty}(1-\gamma)^{k/2}[\Phi(\delta^2)]^{k/2}<\infty .\end{aligned}$$

Plugging this bound into (11.4.8) proves (11.4.3) with $\varsigma=\sum_{k=0}^{\infty}\left\{(1-\gamma)\Phi(\delta^2)\right\}^{k/2}$.

□

Theorem 11.4.2. *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$. Assume that there exist an accessible (m,μ) small set C and $\beta>1$ such that $\mu(C)>0$ and $\sup_{x\in C}\mathbb{E}_x[\beta^{\sigma_C}]<\infty$. Then P has a unique invariant probability measure π, and there exist $\delta>1$ and $\varsigma<\infty$ such that for all $\xi\in\mathbb{M}_1(\mathscr{X})$,*

$$\sum_{k=1}^{\infty}\delta^k \mathrm{d}_{\mathrm{TV}}(\xi P^k,\pi)\le\varsigma\mathbb{E}_\xi[\beta^{\sigma_C}] .$$

Proof. The Markov kernel P is positive by Corollary 11.2.9.

(i) Assume first that the set C is $(1,\mu)$-small and hence that the Markov kernel P is strongly aperiodic. We consider the split chain $\check{P}$ associated to C. By Proposition 11.1.4 (vi), for all $(x,d)\in C\times\{0,1\}$, $\check{P}_{(x,d)}(\sigma_{\check{\alpha}}<\infty)=1$. Furthermore, by Lemma 11.1.1,

$$(1-\varepsilon)\check{\mathbb{E}}_{(x,0)}[\beta^{\sigma_{C\times\{0,1\}}}]+\varepsilon\check{\mathbb{E}}_{(x,0)}[\beta^{\sigma_{C\times\{0,1\}}}]=\mathbb{E}_x[\beta^{\sigma_C}] ,$$

which implies that

$$\sup_{(x,d)\in C\times\{0,1\}}\check{\mathbb{E}}_{(x,d)}[\beta^{\sigma_{C\times\{0,1\}}}]=M<\infty .$$

We apply Theorem 11.4.1 to the set $C\times\{0,1\}$, $\rho=\sigma_{\check{\alpha}}$, $\tau=\sigma_{C\times\{0,1\}}$, and $h\equiv 1$. We obtain that for some $\gamma\in(1,\beta]$, $\sup_{(x,d)\in C\times\{0,1\}}\check{\mathbb{E}}_{(x,d)}[\gamma^{\sigma_{\check{\alpha}}}]<\infty$, which implies that $\check{\mathbb{E}}_{\check{\alpha}}[\gamma^{\check{\alpha}}]<\infty$.

By Theorem 8.2.9, there exists $\delta\in(1,\gamma]$ and $\varsigma<\infty$ such that for all $\check{\xi}\in\mathbb{M}_1(\check{\mathsf{X}},\check{\mathscr{X}})$,

$$\sum_{k=1}^{\infty} \delta^k \mathrm{d_{TV}}(\check{\xi}\check{P}^k, \pi \otimes \mathrm{b}_\varepsilon) \le \varsigma \mathbb{E}_{\check{\xi}}[\gamma^{\sigma_{\check{\alpha}}}] \,.$$

From Lemma 11.1.1 and Proposition 11.1.3, we get that $\mathrm{d_{TV}}(\xi P^n, \pi) \le \mathrm{d_{TV}}([\xi \otimes \mathrm{b}_\varepsilon]\check{P}^n, [\pi \otimes \mathrm{b}_\varepsilon])$. On the other hand, since $\sigma_\alpha \le \sigma_{C\times\{0,1\}} + \sigma_{\check{\alpha}} \circ \theta_{\sigma_{C\times\{0,1\}}}$, we get

$$\begin{aligned}
\check{\mathbb{E}}_{\xi\otimes \mathrm{b}_\varepsilon}[\gamma^{\sigma_{\check{\alpha}}}] &\le \check{\mathbb{E}}_{\xi\otimes \mathrm{b}_\varepsilon}[\gamma^{\sigma_{C\times\{0,1\}}} \check{\mathbb{E}}_{X_{\sigma_{C\times\{0,1\}}}}[\gamma^{\check{\alpha}}]] \\
&\le M \check{\mathbb{E}}_{\xi\otimes \mathrm{b}_\varepsilon}[\gamma^{\sigma_{C\otimes\{0,1\}}}] = M \mathbb{E}_\xi[\beta^{\sigma_C}] \,.
\end{aligned}$$

(ii) We are now going to extend this result for irreducible Markov kernels that are aperiodic but not strongly aperiodic. We set

$$V(x) = \mathbb{E}_x[\lambda^{-\tau_C}] \tag{11.4.9}$$

with $\lambda = \beta^{-1} \in [0,1)$. Proposition 4.3.3 shows that

$$PV \le \lambda V + b\mathbb{1}_C \,, \tag{11.4.10}$$

with $b \sup_{x\in C} \mathbb{E}_x[\lambda^{-\sigma_C}]$. By Lemma 9.4.8, $\{V \le d\}$ is, for all $d > 0$, a petite set and hence a small set, because P is aperiodic; see Theorem 9.4.10. By Corollary 9.2.14, the set $\{V < \infty\}$ is full absorbing, so that $\pi(\{V < \infty\}) = 1$. We can choose $d \ge 2b/(1-\lambda)$ such that $\{V \le d\}$ is an accessible (m,μ)-small set with $\mu(C) > 0$. Iterating m times the inequality (11.4.10), we obtain

$$P^m V \le \lambda^m V + b_m \,, \qquad b_m = b\frac{1-\lambda^m}{1-\lambda} \,. \tag{11.4.11}$$

Set $\eta = (1+\lambda^m)/2$. Since $d > 2b/(1-\lambda)$, we get

$$(\eta - \lambda^m)d = \frac{1-\lambda^m}{2} d \ge b_m \,. \tag{11.4.12}$$

From (11.4.11), we get

$$P^m V \le \eta V + b_m - (\eta - \lambda^m)V,$$

and (11.4.12) shows that on $\{V \ge d\}$, $P^m V \le \eta V$. Therefore, $\{V \le d\}$ is a 1-small set for P^m, and

$$P^m V \le \eta V + b_m \mathbb{1}_{\{V\le d\}} \,. \tag{11.4.13}$$

Set $\sigma_{D,m} = \inf\{k \ge 1 \,:\, X_{km} \in D\}$. Applying Proposition 4.3.3 (ii) to the Markov kernel P^m, we obtain

$$\sup_{x\in D} \mathbb{E}_x[\eta^{-\sigma_{D,m}}] < \infty \,.$$

Since the Markov kernel P^m is strongly aperiodic, we can apply the first part of the proof to P^m to show that there exist $\delta \in \left[1, \eta^{-1}\right)$ and $\varsigma_0 < \infty$ such that for all $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{k=1}^{\infty} \delta^k \mathrm{d}_{\mathrm{TV}}(\xi P^{mk}, \pi) \leq \varsigma_0 \mathbb{E}_\xi [\eta^{-\sigma_{D,m}}] \,.$$

Since $\mathrm{d}_{\mathrm{TV}}(\xi P, \xi' P) \leq \mathrm{d}_{\mathrm{TV}}(\xi, \xi')$ for all $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$, the previous inequality implies

$$\sum_{k=1}^{\infty} \delta^{k/m} \mathrm{d}_{\mathrm{TV}}(\delta_x P^k, \pi) \leq m\delta^m \sum_{k=1}^{\infty} \delta^k \mathrm{d}_{\mathrm{TV}}(\delta_x P^{mk}, \pi) \leq \varsigma_0 \mathbb{E}_\xi [\eta^{-\sigma_{D,m}}] \,.$$

Applying again Proposition 4.3.3 (ii) to P^m, (11.4.13) shows that for all $x \in \mathsf{X}$,

$$\mathbb{E}_x[\eta^{-\sigma_{D,m}}] \leq V(x) + b_m \eta^{-1} \leq (1 + b_m \eta^{-1}) V(x),$$

where $V(x) = \mathbb{E}_x[\lambda^{-\tau_C}]$ (see (11.4.9)). The proof is concluded by noting that $V(x) \leq \mathbb{E}_x[\lambda^{-\sigma_C}] = \mathbb{E}_x[\beta^{\sigma_C}]$.

□

Example 11.4.3 (Functional autoregressive model). The first-order functional autoregressive model on $\mathbb{R}^d$ is defined iteratively by $X_k = m(X_{k-1}) + Z_k$, where $\{Z_k,\ k \in \mathbb{N}^*\}$ is an i.i.d. sequence of random vectors independent of X_0 and $m : \mathbb{R}^d \to \mathbb{R}^d$ is a locally bounded measurable function satisfying

$$\limsup_{|x| \to \infty} \frac{|m(x)|}{|x|} < 1 \,. \tag{11.4.14}$$

Assume that the distribution of Z_1 has density q with respect to Lebesgue measure on $\mathbb{R}^d$ that is bounded away from zero on every compact set and that $\mathbb{E}[|Z_1|] < \infty$. Let K be a compact set with nonempty interior. Then $0 < \mathrm{Leb}(K) < \infty$, and for every $x \in K$,

$$P(x, A) = \int_A q(y - m(x)) \mathrm{d}y \geq \int_{A \cap K} q(y - m(x)) \mathrm{d}y \geq \varepsilon_K \nu(A) \,,$$

with

$$\nu_K(A) = \frac{\mathrm{Leb}(A \cap K)}{\mathrm{Leb}(K)} \,, \quad \varepsilon_K = \mathrm{Leb}(K) \min_{(t,x) \in K \times K} q(t - m(x)) \,.$$

Therefore, every compact subset K with nonempty interior is $(1, \varepsilon_K \nu_K)$-small. Setting $V(x) = 1 + |x|$, we thus obtain for all $d > 0$ that the sets $\{V \leq d\}$ are compact with nonempty interior. They are hence small, and since q is positive, these sets are also accessible. Moreover,

$$PV(x) = 1 + \mathbb{E}[|m(x) + Z_1|] \leq 1 + |m(x)| + \mathbb{E}[|Z_1|] \,. \tag{11.4.15}$$

By (11.4.14), there exist $\lambda \in [0, 1)$ and $r \in \mathbb{R}_+$ such that for all $|x| \geq r$, $|m(x)|/|x| \leq \lambda$. For $|x| \geq r$, this implies

$$PV(x) \le 1 + \lambda|x| + \mathbb{E}|Z_1| = \lambda V(x) + 1 - \lambda + \mathbb{E}|Z_1| \,.$$

Since m is bounded on compact sets, (11.4.15) implies that PV is also bounded on compact sets. Thus setting $b = (1 - \lambda + \mathbb{E}|Z_1|) \vee \sup_{|x| \le r} PV(x)$, we obtain $PV(x) \le \lambda V(x) + b$.

11.5 Exercises

11.1. Consider a Markov chain on $\mathsf{X} = \{0,1\}$ with transition matrix given by

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} .$$

Define $A = \{X_{2k} = 1\,,\ \text{i.o.}\}$.

1. Show that A is asymptotic but not invariant.
2. Show that the asymptotic σ-field is not trivial.

11.2. Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let $f : \mathsf{X} \to \mathbb{R}_+$ be a measurable function. Assume that there exists a $(1, \varepsilon\nu)$-small set with $\varepsilon \in (0,1)$. Set $\check{P} = \check{P}_{\varepsilon,\nu}$.

1. Show that for all $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$ and $k \in \mathbb{N}$ such that $\xi P^k f < \infty$ and $\xi' P^k f < \infty$,

$$\left\| \xi P^k - \xi' P^k \right\|_f \le \left\| [\xi \otimes \mathrm{b}_\varepsilon] \check{P}^k - [\xi' \otimes \mathrm{b}_\varepsilon] \check{P}^k \right\|_{f \otimes \mathbf{1}} . \tag{11.5.1}$$

2. Assume that P admits an invariant probability measure π satisfying $\pi(f) < \infty$. Show that for all $\xi \in \mathbb{M}_1(\mathscr{X})$ and $k \in \mathbb{N}$ such that $\xi P^k f < \infty$, we have

$$\left\| \xi P^k - \pi \right\|_f \le \left\| [\xi \otimes \mathrm{b}_\varepsilon] \check{P}^k - [\pi \otimes \mathrm{b}_\varepsilon] \right\|_{f \otimes \mathbf{1}} . \tag{11.5.2}$$

11.3. Let P be an aperiodic recurrent Markov kernel. If there exist a petite set C and $\varepsilon > 0$ such that $\lim_{n\to\infty} P^n(x, C) = \varepsilon$ for all $x \in \mathsf{X}$, then P is positive and Harris recurrent.

The following exercises use definitions and results introduced in Section 11.A.

11.4. Assume that the asymptotic σ-field $\mathscr{A}$ is almost surely trivial. Then for all $A \in \mathscr{A}$, the mapping defined on $\mathbb{M}_1(\mathscr{X})$ by $\mu \mapsto \mathbb{P}_\mu(A)$ is constant, and this constant (which may depend on A) is either equal to 0 or equal to 1.

11.5. Let P be a Harris null recurrent Markov kernel with invariant measure μ. The aim of this exercise is to prove that $\lim_{n\to\infty} P^n(x, A) = 0$ for all $x \in \mathsf{X}$ and $A \in \mathscr{X}$ such that $\mu(A) < \infty$. Assume that there exist $A \in \mathscr{X}$ and $x \in \mathsf{X}$ such that $\mu(A) < \infty$ and $\limsup_{n\to\infty} P^n(x, A) > 0$. Fix $\varepsilon > 0$ and choose $\delta > 0$ such that

$$\limsup_{n\to\infty} P^n(x, A) \ge \delta(\mu(A) + \varepsilon) . \tag{11.5.3}$$

Assume first that P is aperiodic.

1. Show that there exists B such that $\mu(B) \geq 1/\delta$ and $\lim_{n\to\infty} \sup_{y\in B} |P^n(x,A) - P^n(y,A)| = 0$.
2. Show that we may choose n_0 large enough that for all $n \geq n_0$,

$$\mu(A) \geq \mu(B)(P^n(x,A) - \varepsilon\delta/2) .$$

3. Conclude.

In the general case, consider the cyclic decomposition $C_0, C_1, \ldots, C_{d-1}$ such that $C = \cup_{i=0}^{d-1} C_i$ is full and absorbing, and thus $\mu(C^c) = 0$, since μ is a maximal irreducibility measure by Corollary 11.2.6.

4. Show that $\lim_{n\to\infty} P^n(x,A) = 0$ for every $x \in C$.
5. Show that $\lim_{n\to\infty} P^n(x,A) = 0$ for every $x \notin C$.
6. Conclude.

11.6 Bibliographical Notes

The concept of regeneration plays a central role in the theory of recurrent Markov chains. Splitting techniques were introduced by Nummelin (1978). In that foundational paper, the author deduces various basic results using the renewal methods previously employed for atomic Markov chains. In particular, Nummelin (1978) provides the construction of the invariant measure (Theorem 11.2.5). Essentially the same technique was introduced in Athreya and Ney (1978). The splitting construction introduced here is slightly different. We learned it from Dedecker and Gouëzel (2015).

The renewal representation of Markov chains can be extended to Markov chains that are not irreducible. See, for example, Nummelin (1991) and Nummelin (1997).

Theorem 11.A.4 is due to Orey (1971) (an earlier version is given in Orey (1962) for discrete-state-space Markov chains; see also Blackwell and Freedman (1964)). The proof is not based on the renewal decomposition and made use of the concept of tail σ-algebra.

11.A Another Proof of the Convergence of Harris Recurrent Kernels

Definition 11.A.1 (Asymptotic or tail σ-algebra) *Let P be a Markov kernel on* $\mathsf{X} \times \mathscr{X}$.

- *The σ-algebra $\mathscr{A}$ defined by*

$$\mathscr{A} = \cap_{n\in\mathbb{N}} \mathscr{A}_n \;, \qquad \mathscr{A}_n = \sigma\{X_k \,:\, k \geq n\} \;,$$

  is called the asymptotic or tail σ-field.
- *An event A belonging to $\mathscr{A}$ is said to be an asymptotic or tail event.*
- *A random variable measurable with respect to $\mathscr{A}$ is said to be an asymptotic or tail random variable.*
- *The asymptotic σ-field $\mathscr{A}$ is said to be trivial if for all $\mu \in \mathbb{M}_1(\mathsf{X})$ and $A \in \mathscr{A}$, $\mathbb{P}_\mu(A) = 0$ or 1.*

The event A is asymptotic if and only if for all $n \in \mathbb{N}$, there exists $A_n \in \mathscr{A}_n$ such that $A = \theta_n^{-1}(A_n)$. The σ-algebra $\mathscr{I}$ of invariant events is thus included in $\mathscr{A}$. The converse is not true.

We extend the definition of $\mathbb{P}_\mu$ to all bounded measures $\mu \in \mathbb{M}_b(\mathscr{X})$: define

$$\mathbb{P}_\mu(A) = \int \mu(\mathrm{d}x)\mathbb{P}_x(A) \;.$$

Lemma 11.A.2 *If $\mathscr{A}$ is a.s. trivial, then for all $B \in \mathscr{X}^{\otimes\mathbb{N}}$ and $\nu \in \mathbb{M}_b(\mathsf{X})$,*

$$\lim_{n\to\infty} \sup_{A\in\mathscr{A}_n} |\mathbb{P}_\nu(A\cap B) - \mathbb{P}_\nu(A)\mathbb{P}_\nu(B)| = 0 \;.$$

Proof. Write

$$\begin{aligned}
\sup_{A\in\mathscr{A}_n} |\mathbb{P}_\nu(A\cap B) - \mathbb{P}_\nu(A)\mathbb{P}_\nu(B)| &= \sup_{A\in\mathscr{A}_n} |\mathbb{E}_\nu[\mathbb{1}_A(\mathbb{1}_B - \mathbb{P}_\nu(B))]| \\
&= \sup_{A\in\mathscr{A}_n} |\mathbb{E}_\nu[\mathbb{1}_A(\mathbb{P}_\nu(B\,|\,\mathscr{A}_n) - \mathbb{P}_\nu(B))]| \\
&\leq \mathbb{E}_\nu[|\mathbb{P}_\nu(B\,|\,\mathscr{A}_n) - \mathbb{P}_\nu(B)|] \;.
\end{aligned}$$

The last term tends to 0 by Lebesgue's dominated convergence theorem, since $\lim_{n\to\infty}\mathbb{P}_\nu(B\,|\,\mathscr{A}_n) = \mathbb{P}_\nu(B\,|\,\mathscr{A})$ by Theorem E.3.9, and $\mathbb{P}_\nu(B\,|\,\mathscr{A}) = \mathbb{P}_\nu(B)$, since $\mathscr{A}$ is trivial. □

Let Q be the Markov kernel on the measurable space $(\mathsf{X}\times\mathbb{N}, \mathscr{X}\otimes\mathscr{P}(\mathbb{N}))$ defined by

$$Qf(x,m) = \int f(y,m+1)P(x,\mathrm{d}y), \quad f \in \mathbb{F}_+(\mathsf{X}\times\mathbb{N}) \;.$$

By iterating the kernel Q, we obtain for all $n \geq 0$,

$$Q^n f(x,m) = \int f(y,m+n)P^n(x,\mathrm{d}y) \;.$$

Denote by $(Z_n, \mathbb{P}^Q_z)$ the canonical chain associated to the Markov kernel Q. If $z = (x,m)$, then $\mathbb{P}_z$ is the distribution of $\{(X_n, m+n) : n \geq 0\}$ under $\mathbb{P}_x$.

Proposition 11.A.3 *Let P be a kernel on $\mathsf{X} \times \mathscr{X}$. The following assertions are equivalent:*

(i) the asymptotic σ-field $\mathscr{A}$ is a.s. trivial;
(ii) the bounded Q-harmonic functions are constant;
(iii) for all $\lambda, \mu \in \mathbb{M}_1(\mathsf{X})$, $\lim_{n\to\infty} \|\lambda P^n - \mu P^n\|_{\mathrm{TV}} = 0$.

Proof. (i) $\Rightarrow$ (iii). Assume that (i) holds. Let $\lambda, \mu \in \mathbb{M}_1(\mathscr{X})$ and assume that $\lambda \neq \mu$. Denote by ν^+ and ν^- the positive and negative parts of $\nu = \lambda - \mu$ and let S be a Jordan set for ν, i.e., $S \in \mathscr{X}$ and $\nu^+(S^c) = \nu^-(S) = 0$. Since $\{X_n \in D\} \in \mathscr{A}_n$ for all $D \in \mathscr{X}$ and by definition of the Jordan set S, $\mathbb{P}_{\nu^+}(X_n \in D) = \mathbb{P}_{|\nu|}(X_n \in D, X_0 \in S)$, we obtain

$$\begin{aligned}
&\sup_{D\in\mathscr{X}} |\mathbb{P}_{\nu^+}(X_n \in D) - \mathbb{P}_{\nu^+}(X_0 \in S)\mathbb{P}_{|\nu|}(X_n \in D)| \\
&\qquad = \sup_{D\in\mathscr{X}} |\mathbb{P}_{|\nu|}(X_n \in D, X_0 \in S) - \mathbb{P}_{|\nu|}(X_0 \in S)\mathbb{P}_{|\nu|}(X_n \in D)| \\
&\qquad \leq \sup_{A\in\mathscr{A}_n} |\mathbb{P}_{|\nu|}(A \cap \{X_0 \in S\}) - \mathbb{P}_{|\nu|}(X_0 \in S)\mathbb{P}_{|\nu|}(A)| \,.
\end{aligned}$$

Applying Lemma 11.A.2 to $|\nu|$ and $B = \{X_0 \in S\}$ yields

$$\lim_{n\to\infty} \sup_{D\in\mathscr{X}} |\mathbb{P}_{\nu^+}(X_n \in D) - \mathbb{P}_{\nu^+}(X_0 \in S)\mathbb{P}_{|\nu|}(X_n \in D)| = 0 \,.$$

Replacing $\{X_0 \in S\}$ by $\{X_0 \in S^c\}$, we obtain similarly

$$\lim_{n\to\infty} \sup_{D\in\mathscr{X}} |\mathbb{P}_{\nu^-}(X_n \in D) - \mathbb{P}_{\nu^-}(X_0 \in S^c)\mathbb{P}_{|\nu|}(X_n \in D)| = 0 \,.$$

Since $\mathbb{P}_{\nu^+}(X_0 \in S) = \nu^+(S) = \nu^-(S^c) = \mathbb{P}_{\nu^-}(X_0 \in S^c)$, the previous limits imply that

$$\lim_{n\to\infty} \sup_{D\in\mathscr{X}} |\mathbb{P}_{\nu^+}(X_n \in D) - \mathbb{P}_{\nu^-}(X_n \in D)| = 0 \,,$$

which is equivalent to $\lim_{n\to\infty} \|\lambda P^n - \mu P^n\|_{\mathrm{TV}} = 0$. This proves (iii).

(iii) $\Rightarrow$ (ii). Assume that (iii) holds. Let h be a bounded Q-harmonic function. We have

$$\begin{aligned}
|h(x,m) - h(y,m)| &= |Q^n h(x,m) - Q^n h(y,m)| \\
&= \left| \int_{\mathsf{X}} h(z, m+n) P^n(x, \mathrm{d}z) - \int_{\mathsf{X}} h(z, m+n) P^n(y, \mathrm{d}z) \right| \\
&\leq |h|_\infty \, \left\| \delta_x P^n - \delta_y P^n \right\|_{\mathrm{TV}} .
\end{aligned}$$

By assumption, the right-hand side tends to 0 as n tends to infinity. This implies that $(x,m) \mapsto h(x,m)$ does not depend on x, and we can thus write $h(x,m) = g(m)$ for a bounded function $g : \mathbb{N} \to \mathbb{R}$. The assumption that h is Q-harmonic implies that $g(m) = g(m+1)$ for all $m \in \mathbb{N}$. Hence g, and consequently h, is constant. This proves (ii).

(ii) $\Rightarrow$ (i) Assume that (ii) holds. Fix a distinguished $x_0 \in \mathsf{X}$ and define a mapping θ_{-1} on $\mathsf{X}^{\mathbb{N}}$ by

$$\theta_{-1}(\omega_0, \omega_1, \ldots) = (x_0, \omega_0, \omega_1, \ldots) \ .$$

Then for all $p \geq 1$, we define θ_{-p} by the following recurrence: $\theta_{-(p+1)} = \theta_{-p} \circ \theta_{-1}$. If $A \in \mathscr{A}_n$, then $\theta_{-n}^{-1}(A)$ does not depend on the choice of x_0. Indeed, writing $\mathbb{1}_A = f_n(X_n, X_{n+1}, \ldots)$, it follows that

$$\mathbb{1}_{\theta_{-n}^{-1}(A)} = \mathbb{1}_A \circ \theta_{-n} = f_n(X_0, X_1, \ldots) \ .$$

Note that θ_{-1} is not a left inverse of the shift θ. However, for $A \in \mathscr{A}_n$ and $n > m+1$, $\mathbb{1}_A \circ \theta_{-(m+1)} \circ \theta = \mathbb{1}_A \circ \theta_{-m}$. For $A \in \mathscr{A}$, define the function h on $\mathsf{X} \times \mathbb{N}$ by $h(x,m) = \mathbb{P}_x(\theta_{-m}^{-1}(A))$ and $h(x,0) = \mathbb{P}_x(A)$. We have

$$\begin{aligned} Qh(x,m) &= \int h(y, m+1) P(x, \mathrm{d}y) = \mathbb{E}_x[\mathbb{E}_{X_1}[\mathbb{1}_A \circ \theta_{-(m+1)}]] \\ &= \mathbb{E}_x[\mathbb{1}_A \circ \theta_{-(m+1)} \circ \theta] = \mathbb{E}_x[\mathbb{1}_A \circ \theta_{-m}] = h(x,m) \ . \end{aligned} \tag{11.A.1}$$

This proves that h is a bounded Q-harmonic function, hence is constant by assumption. Then there exists β such that $h(x,m) = \beta$ for all $(x,m) \in \mathsf{X} \times \mathbb{N}$. In particular, $\beta = h(x,0) = \mathbb{P}_x(A)$. Now for all $n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}_{X_n}(A) &= h(X_n, 0) = h(X_n, n) = \mathbb{E}_{X_n}[\mathbb{1}_A \circ \theta_{-n}] \\ &= \mathbb{E}_x\left[\mathbb{1}_A \circ \theta_{-n} \circ \theta_n \,\middle|\, \mathscr{F}_n^X\right] = \mathbb{E}_x\left[\mathbb{1}_A \,\middle|\, \mathscr{F}_n^X\right] \ . \end{aligned}$$

By Theorem E.3.7, $\mathbb{P}_x(A|\mathscr{F}_n^X)$ converges $\mathbb{P}_x -$ a.s. to $\mathbb{1}_A$ as n tends to infinity, so that $\beta \in \{0,1\}$. We have thus proved that for all $A \in \mathscr{A}$, the function $x \mapsto \mathbb{P}_x(A)$ is constant and equal either to 1 or 0, i.e., (i) holds.

□

Theorem 11.A.4. *Let P be an aperiodic Harris recurrent kernel on $\mathsf{X} \times \mathscr{X}$. Then for all $\lambda, \mu \in \mathbb{M}_1(\mathsf{X})$, $\lim_{n\to\infty} \|\lambda P^n - \mu P^n\|_{\mathrm{TV}} = 0$. If P is positive with invariant probability measure π, then $\lim_{n\to\infty} \|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} = 0$ for all $x \in \mathsf{X}$.*

Proof. By Proposition 11.A.3, it is sufficient to prove that if h is a bounded Q-harmonic function, then h is constant. Let h be a bounded Q-harmonic function and set $\tilde{h}(x,m) = h(x,m+1)$. If we can prove that $h = \tilde{h}$, then $(x,m) \mapsto h(x,m)$ does not depend on m and can thus be written, by abuse of notation, $h(x,m) = h(x)$, where h is a bounded P-harmonic function, so that h is constant by Theorem 10.2.11.

The proof is by contradiction. Assume that there exists $z_0 = (x_0, m_0)$ such that $h(z_0) \neq \tilde{h}(z_0)$. Let $\{Z_n,\ n \in \mathbb{N}\}$ be a Markov chain with transition kernel Q. It can be easily checked that $\tilde{h}$ is also a Q-harmonic function, so that $h(Z_n)$ and $\tilde{h}(Z_n)$ are bounded martingales that converge to H and $\tilde{H}$, $\mathbb{P}_{z_0}$ − a.s. We have $\mathbb{E}_{z_0}[H] = h(z_0) \neq \tilde{h}(z_0) = \mathbb{E}_{z_0}[\tilde{H}]$, so that $\mathbb{P}_{z_0}(H \neq \tilde{H}) > 0$. Assume, for instance, that $\mathbb{P}_{z_0}(H < \tilde{H}) > 0$. The case $\mathbb{P}_{z_0}(H > \tilde{H}) > 0$ can be treated in the same way. Note first that there exist $a < b$ such that $\mathbb{P}_{z_0}(H < a < b < \tilde{H}) > 0$. Let $A = \{z :\ h(z) < a\}$, $B = \left\{z :\ \tilde{h}(z) > b\right\}$. Since $a < b$, it follows that

$$\begin{aligned} &\{(x,n) \in \mathsf{X} \times \mathbb{N} :\ (x,n) \in A \cap B, (x,n+1) \in A \cap B\} \\ &\qquad \subset \{(x,n) \in \mathsf{X} \times \mathbb{N} :\ h(x,n+1) < a < b < h(x,n+1)\} = \emptyset\,. \end{aligned} \tag{11.A.2}$$

Since $h(Z_n)$ converges to H and $\tilde{h}(Z_n)$ converges to $\tilde{H}$, it follows that $\mathbb{P}_{z_0}$ − a.s., and since $\mathbb{P}_{z_0}(H < a < b < \tilde{H}) > 0$, we have

$$\begin{aligned} &\mathbb{P}_{z_0}(\exists k,\ \forall n \geq k,\ Z_n \in A \cap B) \\ &\qquad = \mathbb{P}_{z_0}(\exists k,\ \forall n \geq k,\ h(Z_n) < a < b < \tilde{h}(Z_n)) > 0. \end{aligned} \tag{11.A.3}$$

Define

$$D_k = \bigcap_{n=k}^{\infty} \{Z_n \in A \cap B\}\,, \qquad D = \bigcup_{k \geq 0} D_k\,.$$

Then (11.A.3) implies that $\mathbb{P}_{z_0}(D) > 0$. Define $g(z) = \mathbb{P}_z(\bigcap_{n=0}^{\infty}\{Z_n \in A \cap B\}) = \mathbb{P}_z(D_0)$. By the Markov property,

$$g(Z_k) = \mathbb{P}_{Z_k}(D_0) = \mathbb{P}\left(\bigcap_{n=k}^{\infty} \{Z_n \in A \cap B\} \,\middle|\, \mathscr{F}_k^Z \right) = \mathbb{P}\left(D_k \,|\, \mathscr{F}_k^Z \right)\,.$$

We first show that $\mathbb{P}_{z_0}(\lim_{k\to\infty} g(Z_k) = \mathbb{1}_D) = 1$. Indeed, since D_k is increasing, we have for all $m \leq k$,

$$\begin{aligned} |\mathbb{P}(D_k|\mathscr{F}_k^Z) - \mathbb{P}(D|\mathscr{F}_\infty^Z)| &\leq \mathbb{P}(D \setminus D_k|\mathscr{F}_k^Z) + |\mathbb{P}(D|\mathscr{F}_k^Z) - \mathbb{P}(D|\mathscr{F}_\infty^Z)| \\ &\leq \mathbb{P}(D \setminus D_m|\mathscr{F}_k^Z) + |\mathbb{P}(D|\mathscr{F}_k^Z) - \mathbb{P}(D|\mathscr{F}_\infty^Z)|\,. \end{aligned}$$

Letting k then m tend to infinity yields

$$\limsup_{k\to\infty} |\mathbb{P}(D_k|\mathscr{F}_k^Z) - \mathbb{P}(D|\mathscr{F}_\infty^Z)| \leq \lim_{m\to\infty} \mathbb{P}(D \setminus D_m|\mathscr{F}_\infty^Z) = 0 \qquad \mathbb{P}_{z_0} - \text{a.s.}$$

We obtain

$$\lim_{k\to\infty} g(Z_k) = \lim_{k\to\infty} \mathbb{P}(D_k|\mathscr{F}_k^Z) = \mathbb{P}(D|\mathscr{F}_\infty^Z) = \mathbb{1}_D \qquad \mathbb{P}_{z_0} - \text{a.s.}$$

Thus

$$\mathbb{P}_{z_0}\left(\lim_{n\to\infty} g(Z_n) = 1\right) = \mathbb{P}_{z_0}(\mathbb{1}_D = 1) = \mathbb{P}_{z_0}(D) > 0 \,. \tag{11.A.4}$$

Let C be an accessible small set. By Lemma 9.3.3, there exist a probability measure ν, $\varepsilon \in (0,1]$, and $m \in \mathbb{N}$ such that C is both an $(m, \varepsilon\nu)$- and an $(m+1, \varepsilon\nu)$-small set, i.e., for all $x \in C$,

$$P^m(x,\cdot) \ge \varepsilon\nu \,, \qquad P^{m+1}(x,\cdot) \ge \varepsilon\nu \,. \tag{11.A.5}$$

Since C is accessible, Proposition 10.2.2 implies that $\mathbb{P}_{x_0}(X_n \in C \text{ i.o.}) = 1$. By (11.A.4), it is also the case that $\mathbb{P}_{z_0}(\lim_{n\to\infty} g(Z_n) = 1, X_n \in C \text{ i.o.}) > 0$. Therefore, there exists $z_1 = (x_1, n_1)$ such that $x_1 \in C$ and $g(z_1) > 1 - (\varepsilon/4)\nu(C)$, i.e.,

$$\mathbb{P}_{z_1}(\exists n,\, Z_n \notin A \cap B) = 1 - g(z_1) < (\varepsilon/4)\nu(C) \,.$$

Define

$$\begin{aligned} C_0 &= \{x \in C :\ (x, n_1 + m) \notin A \cap B\} \,, \\ C_1 &= \{x \in C :\ (x, n_1 + m + 1) \notin A \cap B\} \,. \end{aligned}$$

We have, using the first inequality in (11.A.5),

$$\begin{aligned} \varepsilon\nu(C_0) &\le \mathbb{P}_{x_1}(X_m \in C_0) \le \mathbb{P}_{x_1}((X_m, n_1 + m) \notin A \cap B) = \mathbb{P}_{z_1}(Z_m \notin A \cap B) \\ &\le \mathbb{P}_{z_1}(\exists n,\, Z_n \notin A \cap B) = 1 - g(z_1) \le (\varepsilon/4)\nu(C) \,. \end{aligned}$$

This yields $\nu(C_0) < \nu(C)/4$. Similarly, using the second inequality in (11.A.5), we obtain

$$\begin{aligned} \varepsilon\nu(C_1) &\le \mathbb{P}_{x_1}(X_{m+1} \in C_0) \le \mathbb{P}_{x_1}((X_{m+1}, n_1 + m + 1) \notin A \cap B) \\ &= \mathbb{P}_{z_1}(Z_{m+1} \notin A \cap B) \le \mathbb{P}_{z_1}(\exists n,\, Z_n \notin A \cap B) = 1 - g(z_1) \le (\varepsilon/4)\nu(C) \,. \end{aligned}$$

Thus $\nu(C_1) < \nu(C)/4$, and altogether these two bounds yield $\nu(C_0 \cup C_1) \le \nu(C)/2 < \nu(C)$, and C contains a point x that does not belong to $C_0 \cup C_1$, i.e., $(x, n_1 + m) \in A \cap B$ and $(x, n_1 + m + 1) \in A \cap B$. This contradicts (11.A.2). $\square$

Chapter 12
Feller and T-Kernels

So far, we have considered Markov kernels on abstract state spaces without any topological structure. In the overwhelming majority of examples, the state space will be a metric space endowed with its Borel σ-field, and in this chapter, we will take advantage of this structure.

Throughout this chapter, (X, d) will be a metric space endowed with its Borel σ-field, denoted by $\mathscr{X}$.

In Sections 12.1 and 12.2, we will introduce Feller kernels, strong Feller kernels, and T-kernels; examples include most of the usual Markov chains on $\mathbb{R}^d$. These types of kernels have certain smoothness properties that can be used to obtain convenient criteria for irreducibility. Another convenient property is that compact sets are petite for an irreducible T-kernel and also for a Feller kernel under an additional topological condition.

In Section 12.3, we will investigate topological conditions for the existence of an invariant probability measure. These conditions can in some cases be applied to certain nonirreducible Feller kernels for which the existence results of Chapter 11 do not apply.

12.1 Feller Kernels

Recall that a sequence of probability measures $\{\mu_n, n \in \mathbb{N}\}$ on a metric space (X, d) is said to converge weakly to a probability measure μ (which we denote by $\mu_n \stackrel{\mathrm{w}}{\Rightarrow} \mu$) if $\lim_{n\to\infty} \mu_n(f) = \mu(f)$ for all functions $f \in \mathrm{C}_b(\mathsf{X})$, the space of real-valued bounded continuous functions on X. The space $\mathrm{C}_b(\mathsf{X})$ endowed with the supremum norm $|\cdot|_\infty$ and the induced topology of uniform convergence is a Banach space. We have already seen in Proposition 1.2.5 that a Markov kernel P maps bounded

R. Douc et al., *Markov Chains*, Springer Series in Operations Research
and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_12

functions onto bounded functions. Thus a Markov kernel maps $\mathrm{C}_b(\mathsf{X})$ into $\mathbb{F}_b(\mathsf{X})$ but not necessarily into $\mathrm{C}_b(\mathsf{X})$ itself. This property must be assumed.

Definition 12.1.1 (Feller kernels and strong Feller kernels) *Let P be a Markov kernel on a metric space* (X, d).

(i) P is called a Feller kernel if $Pf \in \mathrm{C}_b(\mathsf{X})$ for all $f \in \mathrm{C}_b(\mathsf{X})$.
(ii) P is called a strong Feller kernel if $Pf \in \mathrm{C}_b(\mathsf{X})$ for all $f \in \mathbb{F}_b(\mathsf{X})$.

Alternatively, a Markov kernel P is Feller if for every sequence $\{x_n, n \in \mathbb{N}\}$ in X such that $\lim_{n\to\infty} x_n = x$, the sequence of probability measures $\{P(x_n, \cdot), n \in \mathbb{N}\}$ converges weakly to $P(x, \cdot)$, i.e., for all $f \in \mathrm{C}_b(\mathsf{X})$, $\lim_{n\to\infty} Pf(x_n) = Pf(x)$.

A Markov kernel P is strong Feller if and only if for every sequence $\{x_n, n \in \mathbb{N}\}$ in X such that $\lim_{n\to\infty} x_n = x \in \mathsf{X}$, the convergence $\lim_{n\to\infty} Pf(x_n) = Pf(x)$ holds for every $f \in \mathbb{F}_b(\mathsf{X})$. This mode of convergence of the sequence of probability measures $P(x_n, \cdot)$ to $P(x, \cdot)$ is called setwise convergence. Hence, P is strong Feller if $\{P(x_n, \cdot), n \in \mathbb{N}\}$ converges setwise to $P(x, \cdot)$ for every sequence $\{x_n\}$ converging to x.

Proposition 12.1.2 *Let P be a Markov kernel on a metric space* (X, d).

(i) If the kernel P is Feller, then P^n is a Feller kernel for all $n \in \mathbb{N}$. The sampled kernel K_a is also Feller for every sampling distribution a.
(ii) If P is strong Feller, then P^n is strong Feller for all $n \in \mathbb{N}$. The sampled kernel K_a is strong Feller for every sampling distribution $a = \{a(n), n \in \mathbb{N}\}$ such that $a(0) = 0$.

Proof. (i) If P is Feller, then P^n is Feller for all $n \in \mathbb{N}$. For every bounded continuous function $f \in \mathrm{C}_b(\mathsf{X})$, the function $K_a f$ is bounded continuous by Lebesgue's dominated convergence theorem, showing that K_a is Feller.

(ii) The proof is the same. Note that the result is not true if $a(0) > 0$ (for all $x \in \mathsf{X}$, the kernel $Q(x, A) = \delta_x(A)$ for $A \in \mathscr{X}$ is Feller but not strong Feller).

□

Remark 12.1.3. If X is a countable set equipped with the discrete topology, then $\mathbb{F}_b(\mathsf{X}) = \mathrm{C}_b(\mathsf{X})$, and P satisfies the strong Feller property. ▲

Since $\mathrm{C}_b(\mathsf{X}) \subset \mathbb{F}_b(\mathsf{X})$, a strong Feller kernel is a Feller kernel. The converse is not true.

Example 12.1.4 (A Feller kernel that is not strong Feller). Consider the Markov kernel on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ given by $P(x, A) = \delta_{x+1}(A)$ for all $x \in \mathbb{R}$ and $A \in \mathscr{B}(\mathbb{R})$. Then for every Borel function f, $Pf(x) = f(x+1)$. The Markov kernel P is clearly Feller (Pf is continuous if f is continuous), but it is not strong Feller. ◀

We have seen in Theorem 1.3.6 that a Markov chain can be expressed as a random iterative system $X_{n+1} = F(X_n, Z_{n+1})$, where $\{Z_n, n \in \mathbb{N}^*\}$ is a sequence of i.i.d. random elements on a probability space $(\mathsf{Z}, \mathscr{Z})$, independent of X_0, and $F : \mathsf{X} \times \mathsf{Z} \to \mathsf{X}$ is a measurable function. If the function F has a certain smoothness property, then it defines a Feller kernel.

Lemma 12.1.5 *Let (X, d) be a metric space and $(\mathsf{Z}, \mathscr{Z})$ a measurable space, $\mu \in \mathbb{M}_1(\mathscr{Z})$, Z a random variable with distribution μ, and let $F : \mathsf{X} \times \mathsf{Z} \to \mathsf{X}$ be a measurable function. Let P be the Markov kernel associated to the function F and the measure μ, defined for $x \in \mathsf{X}$ and $f \in \mathbb{F}_b(\mathsf{X})$ by*

$$Pf(x) = \mathbb{E}\left[f(F(x,Z))\right] = \int f(F(x,z))\mu(\mathrm{d}z)\,.$$

If the function $x \to F(x,z)$ is continuous with respect to x for μ almost all $z \in \mathsf{Z}$, then P is a Feller kernel.

Proof. Let $f \in \mathrm{C}_b(\mathsf{X})$ and $x \in \mathsf{X}$. By assumption, the function $f(F(x,z))$ is bounded and continuous with respect to x for μ-almost all z. By Lebesgue's dominated convergence theorem, this implies that $Pf(x) = \mathbb{E}\left[f(F(x,Z))\right]$ is continuous. □

Proposition 12.1.6 *A Feller kernel on $\mathsf{X} \times \mathscr{X}$ is a bounded linear operator on the Banach space $(\mathrm{C}_b(\mathsf{X}), |\cdot|_\infty)$ and a sequentially continuous operator on $\mathbb{M}_1(\mathscr{X})$ endowed with the topology of weak convergence.*

Proof. Let P be a Feller kernel. For $f \in \mathrm{C}_b(\mathsf{X})$ and $x \in \mathsf{X}$,

$$|Pf(x)| \le \int_{\mathsf{X}} |f(y)| P(x, \mathrm{d}y) \le |f|_\infty P(x, \mathsf{X}) \le |f|_\infty\,.$$

This proves the first statement. Let $\{\mu_n, n \in \mathbb{N}\}$ be a sequence of probability measures on $(\mathsf{X}, \mathscr{X})$ such that μ_n converges weakly to a probability measure μ. Then for every $f \in \mathrm{C}_b(\mathsf{X})$, since Pf is also in $\mathrm{C}_b(\mathsf{X})$, we have

$$\lim_{n\to\infty} (\mu_n P)(f) = \lim_{n\to\infty} \mu_n(Pf) = \mu(Pf) = (\mu P)(f)\,.$$

This proves that the sequence $\{\mu_n P, n \in \mathbb{N}\}$ converges weakly to μP. □

Proposition 12.1.7 *Let (X, d) be a complete separable metric space and let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. If P is strong Feller, then there exists a probability measure μ on $\mathscr{B}(\mathsf{X})$ such that $P(x, \cdot)$ is absolutely continuous with respect to μ for all $x \in \mathsf{X}$. In addition, there exists a bimeasurable function $(x,y) \mapsto p(x,y)$ such that $p(x,y) = \mathrm{d}P(x,\cdot)/\mathrm{d}\mu(y)$.*

Proof. Since (X, d) is a complete separable metric space, there exists a sequence $\{x_n, n \in \mathbb{N}^*\}$ that is dense in X. Define the measure μ on $\mathscr{B}(\mathsf{X})$ by $\mu = \sum_{n=1}^{\infty} 2^{-n} P(x_n, \cdot)$. For $A \in \mathscr{B}(\mathsf{X})$ such that $\mu(A) = 0$, we have $P(x_n, A) = 0$ for all $n \geq 1$. Since P is strong Feller, the function $x \mapsto P(x, A)$ is continuous. Since it vanishes on a dense subset of X, this function is identically equal to 0. Thus $P(x, \cdot)$ is absolutely continuous with respect to μ for all x in X. The existence of the bimeasurable version of the Radon–Nykodym derivative is given by Corollary 9.A.3. □

We now give a characterization of the Feller and strong Feller properties in terms of lower semicontinuous functions (see Definition B.1.5).

Proposition 12.1.8 *Let (X, d) be a metric space.*

(i) A Markov kernel P is Feller if and only if the function $P(\cdot, U)$ is lower semicontinuous for every open set U.
(ii) A Markov kernel P is strong Feller if and only if the function $P(\cdot, A)$ is lower semicontinuous for every Borel set $A \in \mathscr{X}$.

Proof. (i) Assume that the Markov kernel P is Feller. If U is an open set, then there exists an increasing sequence $\{f_n, n \in \mathbb{N}\} \subset \mathrm{C}_b(\mathsf{X})$ such that $\mathbb{1}_U = \lim_{n\to\infty} f_n$. (Take $f_n(x) = 1 \wedge n d(x, U^c)$, for instance). By the monotone convergence theorem, it is also the case that $P(\cdot, U) = \lim_{n\to\infty} P f_n$, and for every $n \in \mathbb{N}$, $P f_n$ is continuous (and therefore lower semicontinuous). Hence $P(\cdot, U)$ is a pointwise increasing limit of lower semicontinuous functions and is therefore lower semicontinuous by Proposition B.1.7 (iii).

Conversely, let $f \in \mathrm{C}_b(\mathsf{X})$ be such that $0 \leq f \leq 1$. Then $f = \lim_{n\to\infty} f_n$ with

$$f_n = 2^{-n} \sum_{k=1}^{2^n} \mathbb{1}_{\{f > k2^{-n}\}} = \sum_{k=1}^{2^n} \frac{k-1}{2^n} \mathbb{1}_{[(k-1)2^{-n}, k2^{-n})}(f) . \tag{12.1.1}$$

Since f is continuous, for each $k \in \{1, \ldots, 2^n - 1\}$, $\{f < k2^{-n}\}$ is an open set. Hence the function $P(\cdot, \{f < k2^{-n}\})$ is lower semicontinuous, and so is $P f_n$ (a finite sum of lower semicontinuous functions being lower semicontinuous; see Proposition B.1.7 (iv)). By the monotone convergence theorem, $P f = \lim_{n\to\infty} P f_n$, which is lower semicontinuous by Proposition B.1.7 (iii). Similarly, $P(1 - f) = 1 - P f$ is also lower semicontinuous. This implies that $P f$ is both lower semicontinuous and upper semicontinuous, hence continuous. This proves that P is Feller.

(ii) The direct implication is obvious; the proof of the converse is a verbatim repetition of the previous proof except that the argument that $\{f < k2^{-n}\}$ is an open set is replaced by $\{f < k2^{-n}\}$ being a Borel set.

□

A very important property of irreducibility is that compact sets are petite. To prove this property, we first need to prove that the closure of a petite set is petite.

Lemma 12.1.9 *Let P be an irreducible Feller kernel. Then the closure of a petite set is petite.*

Proof. Let A be an (a,μ)-petite set, i.e., $K_a(x,B) \geq \mu(B)$ for all $x \in A$ and $B \in \mathscr{X}$. Let $\overline{A}$ be the closure of A. We will show that there exists a petite set H such that $\inf_{x\in\overline{A}} K_a(x,H) > 0$. The set H being petite, this implies by Lemma 9.4.7 that the set $\overline{A}$ is also petite.

Since P is irreducible, Proposition 9.1.8 shows that X is the union of a countable collection of small sets. Since μ is nontrivial, $\mu(C) > 0$ for some small set $C \in \mathscr{B}(\mathsf{X})$.

By Theorem B.2.17, μ is inner regular, and thus there exists a closed set $H \subset C$ such that $\mu(H) > 0$. Since P is Feller, the sampled kernel K_a is also Feller. Since H^c is an open set, by Proposition 12.1.8, $K_a(\cdot,H^c)$ is lower semicontinuous, so $K_a(\cdot,H) = 1 - K_a(\cdot,H^c)$ is upper semicontinuous. By Proposition B.1.9 (ii), we have

$$\inf_{x\in\overline{A}} K_a(x,H) = \inf_{x\in A} K_a(x,H) \geq \mu(H) > 0\,,$$

which concludes the proof. □

Theorem 12.1.10. *Let P be an irreducible Feller kernel. If there exists an accessible open petite set, then all compact sets are petite. If there exists a maximal irreducibility measure whose topological support has nonempty interior, then there exists an accessible open petite set, and all compact sets are petite.*

Proof. Assume that there exists an accessible open petite set. Then $K_{a_\varepsilon}(x,U) > 0$ for all $x \in \mathsf{X}$. Since P is a Feller kernel, the function $K_{a_\varepsilon}(\cdot,U)$ is lower semicontinuous by Proposition 12.1.8. Thus for every compact set H, $\inf_{x\in H} K_{a_\varepsilon}(x,U) > 0$ by Proposition B.1.7 (v). Therefore, H is petite by Lemma 9.4.7.

Let now ψ be a maximal irreducibility measure whose support has nonempty interior. Since P is irreducible, there exists an accessible small set A. For $\varepsilon \in (0,1)$ and $k \in \mathbb{N}^*$, set

$$B_k = \{x \in \mathsf{X} : K_{a_\varepsilon}(x,A) \geq 1/k\}\,.$$

Since A is accessible, $\mathsf{X} = \bigcup_{k=1}^{\infty} B_k$. Each B_k leads uniformly to the small set A, and thus B_k is also petite by Lemma 9.4.7. By Lemma 12.1.9, $\overline{B_k}$ is also petite, and the set C_k defined by $C_k = \overline{B_k} \cap \operatorname{supp}(\psi)$ is petite and closed, since $\operatorname{supp}(\psi)$ is closed by definition. By construction, $\operatorname{supp}(\psi) = \bigcup_{k=1}^{\infty} C_k$, and by assumption, $\operatorname{supp}(\psi)$ has nonempty interior. By Baire's Theorem B.1.1, there must exist at least one k such that C_k has nonempty interior, say U, which is a petite set (as a subset of a petite set). Moreover, $\psi(U) > 0$ by definition of the support of ψ (see Proposition B.2.15). This implies that U is accessible. □

12.2 T-Kernels

We now introduce the notion of T-kernel, which is a significant generalization of the strong Feller property that holds in many applications.

Definition 12.2.1 (T-kernel, continuous component) *A Markov kernel P is called a T-kernel if there exist a sampling distribution $a \in \mathbb{M}_1(\mathbb{N})$ and a sub-Markovian kernel T such that*

(i) $T(x,\mathsf{X}) > 0$ for all $x \in \mathsf{X}$;
(ii) for all $A \in \mathscr{X}$, the function $x \mapsto T(x,A)$ is lower semicontinuous;
(iii) for all $x \in \mathsf{X}$ and $A \in \mathscr{X}$, $K_a(x,A) \geq T(x,A)$.

The sub-Markovian kernel T is called the continuous component of P.

A strong Feller kernel is a T-kernel: simply take $T = P$ and $a = \delta_1$. A Feller kernel is not necessarily a T-kernel. The T-kernels form a larger class of Markov kernels than strong Feller kernels. For instance, it will be shown in Exercise 12.6 that the Markov kernel associated to a random walk on $\mathbb{R}^d$ is always Feller, but is strong Feller if and only if its increment distribution is absolutely continuous with respect to Lebesgue measure. However, it is a T-kernel under a much weaker condition. For instance, a Metropolis–Hastings MCMC sampler is generally not a strong Feller kernel but is a T-kernel under weak additional conditions.

Lemma 12.2.2 *Let T be a sub-Markovian kernel such that for all $A \in \mathscr{X}$, the function $x \mapsto T(x,A)$ is lower semicontinuous. Then for all $f \in \mathbb{F}_+(\mathsf{X})$, the function $x \mapsto Tf(x)$ is lower semicontinuous.*

Proof. Every $f \in \mathbb{F}_+(\mathsf{X})$ is an increasing limit of simple functions $\{f_n, n \in \mathbb{N}\}$. For every $n \in \mathbb{N}$, f_n is simple, and Tf_n is therefore lower semicontinuous, as a finite sum of lower semicontinuous functions. By the monotone convergence theorem, $Tf = \lim_{n\to\infty} Tf_n$, and by Proposition B.1.7 (iii), an increasing limit of lower semicontinuous functions is lower semicontinuous; therefore, Tf is is lower semicontinuous. □

For a T-kernel (and a fortiori for a strong Feller kernel), we have a stronger result than that for Feller kernels: all compact sets are petite without any additional assumption.

Theorem 12.2.3. *Let P be an irreducible T-kernel. Then every compact set is petite.*

Proof. Since P is irreducible, there exists an accessible petite set A satisfying $K_{a_\varepsilon}(x,A) > 0$ for all $x \in \mathsf{X}$ and $\varepsilon \in (0,1)$ (see Lemma 9.1.6). Since P is a T-kernel, there exist a sampling distribution a and a continuous component T such that $K_a \geq T$

and $T(x,\mathsf{X})>0$ for all $x\in\mathsf{X}$. By the generalized Chapman–Kolmogorov formula (Lemma 1.2.11), this implies that for all $x\in\mathsf{X}$,

$$K_{a*a_\varepsilon}(x,A)=K_aK_{a_\varepsilon}(x,A)\geq TK_{a_\varepsilon}(x,A)>0\,.$$

Let C be a compact set. By Lemma 12.2.2, the function $TK_{a_\varepsilon}(\cdot,A)$ is lower semicontinuous. Moreover, it is positive everywhere on X, so it is uniformly bounded from below on C. This implies that $\inf_{x\in C}K_{a*a_\varepsilon}(x,A)>0$ and by Lemma 9.4.7 that C is petite. □

Theorem 12.2.3 admits a converse. On a locally compact separable metric space, if every compact set is petite, then P is a T-kernel. To prove this result we need the following lemma.

Lemma 12.2.4 *Let P be a Markov kernel. If X is a countable union of open petite sets, then P is a T-kernel.*

Proof. Let $\{U_k,k\in\mathbb{N}\}$ be a collection of open petite sets such that $\mathsf{X}=\bigcup_{k=1}^\infty U_k$. By definition of a petite set, for every integer k, there exist a sampling distribution $a^{(k)}\in\mathbb{M}_1(\mathbb{N})$ and a nontrivial measure $\nu_k\in\mathbb{M}_+(\mathsf{X})$ such that $K_{a^{(k)}}\geq\mathbb{1}_{U_k}\nu_k$. We then set $T_k=\mathbb{1}_{U_k}\nu_k$, $T=\sum_{k=1}^\infty 2^{-k}T_k$, and $a=\sum_{k\geq1}2^{-k}a^{(k)}$. The function T is well defined, since $T_k(x,\mathsf{X})\leq1$ for all $k\in\mathbb{N}$ and $x\in\mathsf{X}$. This yields

$$K_a=\sum_{k\geq1}2^{-k}K_{a^{(k)}}\geq T\,.$$

The indicator function of an open set being lower semicontinuous, the function $x\mapsto T_k(x,A)$ is lower semicontinuous for every $A\in\mathscr{X}$. Thus the function $x\mapsto T(x,A)$ is lower semicontinuous as an increasing limit of lower semicontinuous functions. Finally, since $T_k(x,\mathsf{X})>0$ for all $x\in U_k$ and $\mathsf{X}=\bigcup_{k\geq1}U_k$, we have that $T(x,\mathsf{X})>0$ for all $x\in\mathsf{X}$. □

Theorem 12.2.5. *Assume that (X,d) is a locally compact separable metric space. Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$. If every compact set is petite, then P is a T-kernel.*

Proof. Let $\{x_k,\,k\in\mathbb{N}\}$ be a dense sequence in X. For every integer k, there exists a relatively compact open neighborhood of x_k. Then $\overline{U}_k$ is compact, hence petite, and therefore U_k is also petite, since a subset of a petite set is petite. Hence X is a countable union of open petite sets, and P is a T-kernel by Lemma 12.2.4. □

Combining Theorems 12.1.10 and 12.2.5, we obtain the following criterion for a Feller kernel to be a T kernel.

Corollary 12.2.6 *Let P be an irreducible Feller kernel on a locally compact separable metric space (X,d). If there exists a maximal irreducibility measure whose topological support has nonempty interior, then P is a T-kernel.*

This result also provides a criterion to check whether a Feller kernel is not a T chain. See Exercise 12.10.

Example 12.2.7 (Vector autoregressive process). Consider the vector autoregressive process (see Example 2.1.2) defined by the recurrence

$$X_{n+1} = FX_n + GZ_{n+1}, \tag{12.2.1}$$

where $\{Z_n, n \in \mathbb{N}^*\}$ is a sequence of $\mathbb{R}^q$-valued i.i.d. random vectors, X_0 is an $\mathbb{R}^p$-valued random vector independent of $\{Z_n, n \in \mathbb{N}\}$, F is a $p \times p$ matrix, and G is a $p \times q$ matrix $(p \geq q)$. Assume that the pair (F,G) is controllable (see Section 12.A) and that the distribution μ of the random vector Z_1 is nonsingular with respect to the Lebesgue measure, i.e., there exists a nonnegative function g such that $\mathrm{Leb}(g) > 0$ and $\mu \geq g \cdot \mathrm{Leb}$.

Assume first that $p = q$ and $G = \mathrm{I}_q$. For $A \in \mathscr{B}(\mathbb{R}^p)$, define

$$T(x,A) = \int \mathbb{1}_A(y) g(y - Fx) \mathrm{d}y \,.$$

Note that for all $x \in \mathbb{R}^q$, $T(x, \mathbb{R}^q) = \mathrm{Leb}(g) > 0$. Since the function $z \mapsto \int |g(y-z) - g(y)| \mathrm{d}y$ is continuous, it follows that for all $A \in \mathscr{B}(\mathbb{R}^q)$, $x \mapsto T(x,A)$ is continuous. Hence P is a T-kernel.

We now consider the general case. By iterating (2.1.3), we get

$$X_n = F^n X_0 + \sum_{k=1}^{n} F^{n-k} G Z_k \,. \tag{12.2.2}$$

We assume again that the pair (F,G) is controllable. This means that the matrix $C_m = [G\,|FG\,|\cdots|F^{m-1}G]$ has full rank for some sufficiently large m (it suffices to take for m the degree of the minimal polynomial of F). Define $\tilde{X}_n = X_{nm}$, $\tilde{F} = F^m$, and

$$\tilde{Z}_{n+1} = F^{m-1} G Z_{nm+1} + F^{m-2} G Z_{nm+2} + \cdots + FGZ_{nm+m-1} + GZ_{nm+m} \,.$$

we may rewrite the recurrence (12.2.1) as follows:

$$\tilde{X}_{n+1} = \tilde{F} \tilde{X}_n + \tilde{Z}_{n+1} \,.$$

Denote by $\boldsymbol{\Phi} : \mathbb{R}^{mq} \to \mathbb{R}^p$ the linear map

$$(z_1, z_2, \ldots, z_m) \to F^{m-1} G z_1 + F^{m-2} G z_2 + \ldots + FG z_{m-1} + G z_m \,.$$

The rank of Φ is p, since the pair (F,G) is controllable. The distribution of the random vector $(Z_{nm+1}^T, Z_{nm+1}^T, \ldots, Z_{nm+m-1}^T)^T$ over $\mathbb{R}^{mp}$ is $\mu^{\otimes m}$, which by assumption satisfies $g^{\otimes m} \cdot \mathrm{Leb}$. It can be shown (see Exercise 12.12) that there exists a function g such that the distribution $\nu = \mu^{\otimes m} \circ \Phi^{-1}$ of the random vector $\tilde{Z}_1$ has a continuous component, i.e., there exists a nonnegative function $\tilde{g}$ such that $\mathrm{Leb}(g) > 0$ and $\nu \geq \tilde{g} \cdot \mathrm{Leb}$. Using the first part of the proof, we obtain $P^m(x,A) \geq \int_A \tilde{g}(y - \tilde{F}x)\mathrm{d}y$, where $x \mapsto T(x,A)$ is continuous and $T(x,\mathsf{X}) = \mathrm{Leb}(\tilde{g}) > 0$. Hence P is a T-kernel.

We now introduce reachable points, which will in particular provide a characterization of irreducibility.

Definition 12.2.8 (Reachable point) *A point x^* is reachable if every open neighborhood of x^* is accessible.*

Theorem 12.2.9. *Let P be a T-kernel. If there exists a reachable point x^*, then P is irreducible and $\phi = T(x^*,\cdot)$ is an irreducibility measure. In addition, $T(x^*,\cdot) \ll \mu$ for every invariant measure μ, and there exists at most one invariant probability measure.*

Proof. Let T be a continuous component of K_a. Then by definition, $T(x^*,\mathsf{X}) > 0$. Let $A \in \mathscr{X}$ be such that $T(x^*,A) > 0$. Since the function $x \mapsto T(x,A)$ is lower semicontinuous, there exists $U \in \mathscr{V}_{x^*}$ such that $T(x,A) \geq \delta > 0$ for all $x \in U$. Since x^* is assumed to be reachable, this implies that $K_{a_\varepsilon}(x,U) > 0$ for all $x \in \mathsf{X}$ and $\varepsilon \in (0,1)$. Then by Lemma 1.2.11, for all $x \in \mathsf{X}$,

$$\begin{aligned} K_{a_\varepsilon * a}(x,A) &= \int_{\mathsf{X}} K_{a_\varepsilon}(x,\mathrm{d}y) K_a(y,A) \geq \int_U K_{a_\varepsilon}(x,\mathrm{d}y) K_a(y,A) \\ &\geq \int_U K_{a_\varepsilon}(x,\mathrm{d}y) T(y,A) \geq \delta K_{a_\varepsilon}(x,U) > 0 \,. \end{aligned}$$

Therefore, A is accessible, and hence $T(x^*,\cdot)$ is an irreducibility measure. If μ is an invariant measure and $T(x^*,A) > 0$, then $\mu(A) = \int \mu(\mathrm{d}x) K_{a_\varepsilon * a}(x,A) > 0$. Thus $T(x^*,\cdot)$ is absolutely continuous with respect to μ.

The last statement is a consequence of Corollary 9.2.16: an irreducible kernel has at most one invariant probability measure. □

Example 12.2.10. We pursue the investigation of the first-order vector autoregressive process $X_{n+1} = FX_n + GZ_{n+1}$, and we use the notation introduced in Example 12.2.7. We will find sufficient conditions under which the associated kernel possesses a reachable state. Denote by $\rho(F)$ the spectral radius for F, i.e., $\rho(F)$ is the maximal modulus of the eigenvalues of F. It is well known that if the spectral radius $\rho(F)$ is less than 1, then there exist constants c and $\bar{\rho} < 1$ such that for every $n \in \mathbb{N}$, $|||F^n||| \leq c\bar{\rho}^n$.

Assume that the pair (F,G) is controllable and that the distribution μ of the random vector Z_1 satisfies $\mu \geq \rho_0 \mathbb{1}_{\mathrm{B}(z_*,\varepsilon_0)} \cdot \mathrm{Leb}_p$ for some $z_* \in \mathbb{R}^p$, $\rho_0 > 0$, and $\varepsilon_0 > 0$. Denote by $x_* \in \mathbb{R}^q$ the state given by

$$x_* = \sum_{k=0}^{\infty} F^k G z_* \,. \tag{12.2.3}$$

For all $n \in \mathbb{N}$, $X_n = F^n X_0 + \sum_{k=1}^n F^{n-k} G Z_k$, and thus for all $x \in \mathbb{R}^p$ and all open neighborhoods O of x^*, there exist n large enough and ε sufficiently small such that on the event $\bigcap_{k=1}^n \{|Z_k - z_*| \leq \varepsilon\}$,

$$X_n = F^n x + \sum_{k=1}^n F^{n-k} G Z_k \in O \,,$$

showing that $P^n(x,O) \geq \mu^n(\mathrm{B}(z_*,\varepsilon)) > 0$. Hence the state x_* is reachable. If in addition the pair (F,G) is controllable, then as shown in Example 12.2.7, P is a T-kernel. Hence P is an irreducible T-kernel: Theorem 12.2.3 shows that every compact set is petite (as shown in Exercise 12.13, compact sets are even small). ◄

12.3 Existence of an Invariant Probability

For $\mu \in \mathbb{M}_1(\mathscr{X})$, consider the probability measures π_n^μ, $n \geq 1$, defined by

$$\pi_n^\mu = n^{-1} \sum_{k=0}^{n-1} \mu P^k \,. \tag{12.3.1}$$

This probability is the expected n-step occupation measure with initial distribution μ, i.e., for every $A \in \mathscr{B}(\mathsf{X})$, $\pi_n^\mu(A) = n^{-1} \mathbb{E}_\mu \left[\sum_{k=0}^{n-1} \mathbb{1}_A(X_k) \right]$. By definition of π_n^μ, the following relation between π_n^μ and $\pi_n^\mu P$ holds:

$$\pi_n^\mu P = \pi_n^\mu + \frac{1}{n} \{\mu P^n - \mu\} \,. \tag{12.3.2}$$

This relation is the key to the following result.

Proposition 12.3.1 *Let P be a Feller kernel on a metric space (X,d). For $\mu \in \mathbb{M}_1(\mathscr{X})$, all the weak limits of $\{\pi_n^\mu, n \in \mathbb{N}^*\}$ along subsequences are P-invariant.*

Proof. Let π be a weak limit along a subsequence $\{\pi_{n_k}^\mu, k \in \mathbb{N}\}$. Since P is Feller, it follows that $Pf \in \mathrm{C}_b(\mathsf{X})$ for all $f \in \mathrm{C}_b(\mathsf{X})$. Thus, using (12.3.2),

$$|\pi P(f) - \pi(f)| = |\pi(Pf) - \pi(f)| = \lim_{k\to\infty} |\pi^{\mu}_{n_k}(Pf) - \pi^{\mu}_{n_k}(f)|$$
$$= \lim_{k\to\infty} \frac{1}{n_k} |\mu P^{n_k}(f) - \mu(f)| \le \lim_{k\to\infty} \frac{2|f|_\infty}{n_k} = 0\,,$$

this proves that $\pi = \pi P$ by Corollary B.2.18. □

This provides a method for proving the existence of an invariant probability measure. However, to be of any practical use, this method requires a practical way to prove relative compactness. Such a criterion is provided by tightness. The family Π of probability measures on X is tight if for every $\varepsilon > 0$, there exists a compact set K such that for all $\xi \in \Pi$, $\xi(K) \ge 1 - \varepsilon$; see Section C.2.

Theorem 12.3.2. *Let P be a Feller kernel. Assume that there exists $\mu \in \mathbb{M}_1(\mathscr{X})$ such that the family of probability measures $\{\pi^{\mu}_n, n \in \mathbb{N}\}$ is tight. Then P admits an invariant probability measure.*

Proof. By Prohorov's theorem, Theorem C.2.2, if $\{\pi^{\mu}_n, n \in \mathbb{N}\}$ is tight, then it is relatively compact, and thus there exist $\pi \in \mathbb{M}_1(\mathscr{X})$ and a sequence $\{n_k, k \in \mathbb{N}\}$ such that $\{\pi_{n_k}, k \in \mathbb{N}\}$ converges weakly to π. By Proposition 12.3.1, the probability measure π is P-invariant. □

An efficient way to check the tightness of the sequence $\{\pi^{\mu}_n, n \in \mathbb{N}\}$ is by means of Lyapunov functions.

Theorem 12.3.3. *Let P be a Feller kernel on a metric space (X, d). Assume that there exist a measurable function $V : \mathsf{X} \to [0, \infty]$ such that $V(x_0) < \infty$ for at least one $x_0 \in \mathsf{X}$, a measurable function $f : \mathsf{X} \to [1, \infty)$ such that the level sets $\{x \in \mathsf{X} : f(x) \le c\}$ are compact for all $c > 0$, and a constant $b < \infty$ such that*

$$PV + f \le V + b\,. \tag{12.3.3}$$

Then P admits an invariant probability measure.

Proof. For all $n \in \mathbb{N}$, we obtain by induction

$$P^n V + \sum_{k=0}^{n-1} P^k f \le V + (n+1)b\,.$$

Therefore, we get for all $n \in \mathbb{N}$ that $\pi^{\delta_{x_0}}_n(f) \le V(x_0) + b$, which implies that for all $c > 0$ and $n \in \mathbb{N}$, $\pi^{\delta_{x_0}}_n(\{f \ge c\}) \le \{V(x_0) + b\}/c$. □

The drift condition (12.3.3) does not always hold. It is thus of interest to derive a weaker criterion for the existence of an invariant probability. This can be achieved if the space (X, d) is a locally compact separable metric space; see Section B.1.3. A function $f \in \mathrm{C}_b(\mathsf{X})$ is said to vanish at infinity if for every $\varepsilon > 0$, there exists a compact set K such that $|f(x)| \leq \varepsilon$ for all $x \notin K$. The set of continuous functions vanishing at infinity is denoted by $\mathrm{C}_0(\mathsf{X})$; see Definition B.1.11. This function space induces a new form of weak convergence, namely *weak** convergence. A sequence of bounded measures $\{\mu_n, n \in \mathbb{N}\}$ converges *weakly** to $\mu \in \mathbb{M}_b(\mathscr{X})$, which we write $\mu_n \stackrel{\mathrm{w}^*}{\Rightarrow} \mu$, if $\lim_{n\to\infty} \mu_n(f) = \mu(f)$ for all $f \in \mathrm{C}_0(\mathsf{X})$. Note that *weak** convergence is weaker than weak convergence and that the *weak** limit of a sequence of probability measures is a bounded measure but not necessarily a probability measure. See Appendix C.

We first extend Proposition 12.3.1 to *weak** convergence.

Proposition 12.3.4 *Let (X, d) be a locally compact separable metric space and P a Feller kernel. If π is a weak* limit of $\{\pi_n^\mu, n \in \mathbb{N}\}$ along a subsequence, then π is invariant.*

Proof. Assume that the subsequence $\{\pi_{n_k}^\mu, k \in \mathbb{N}\}$ converges *weakly** to π. Since P is Feller, $Pf \in \mathrm{C}_b^+(\mathsf{X})$ for all $f \in \mathrm{C}_0^+(\mathsf{X})$. Applying Proposition C.1.2 and the bound $|\pi_{n_k}^\mu(Pf) - \pi_{n_k}^\mu(f)| \leq 2n_k^{-1}|f|_\infty$ (see (12.3.2)), we obtain that for all $f \in \mathrm{C}_0^+(\mathsf{X})$,

$$\pi(Pf) \leq \liminf_{k\to\infty} \pi_{n_k}^\mu(Pf) = \liminf_{k\to\infty} \pi_{n_k}^\mu(f) = \pi(f)\,. \tag{12.3.4}$$

Therefore, $\pi P(f) = \pi(Pf) \leq \pi(f)$ for all $f \in \mathrm{C}_0^+(\mathsf{X})$. By B.2.21, this implies that $\pi P \leq \pi$, and since $\pi P(\mathsf{X}) = \pi(\mathsf{X})$, we conclude that then $\pi P = \pi$. ☐

As mentioned above, *weak** limits of a sequence of probability measures are bounded measures but not necessarily probability measures and can even be the trivial measure (identically equal to zero), in which case we would have achieved very little. We need an additional assumption to ensure the existence of an invariant probability measure.

Theorem 12.3.5. *Let (X, d) be a locally compact separable metric space and P a Feller kernel. Assume that there exist $f_0 \in \mathrm{C}_0^+(\mathsf{X})$ and $\mu \in \mathbb{M}_1(\mathscr{X})$ such that* $\liminf_{n\to\infty} \pi_n^\mu(f_0) > 0$. *Then P admits an invariant probability measure.*

Proof. By Proposition C.1.3, $\mathbb{M}_1(\mathscr{X})$ is *weak** sequentially compact. Therefore, there is a subsequence $\{\pi_{n_k}^\mu, k \in \mathbb{N}\}$ that converges *weakly** to a bounded measure ν that is invariant by Proposition 12.3.4. Under the stated assumption, $\nu(f_0) > 0$, and therefore ν is nontrivial. Since ν is bounded, the measure $\nu/\nu(\mathsf{X})$ is an invariant probability measure. □

Theorem 12.3.6. *Let P be a Markov kernel on a locally compact separable metric space (X, d). Assume that there exist $k \geq 1$ such that P^k is Feller, a function $V : \mathsf{X} \to [1, \infty]$ finite for at least one $x_0 \in \mathsf{X}$, a compact set K, and a positive real number b such that*

$$P^k V \leq V - 1 + b\mathbb{1}_K(x). \tag{12.3.5}$$

Then P admits an invariant probability measure.

Proof. We begin with the case $k = 1$. Write the drift condition as $V \geq PV + 1 - b\mathbb{1}_K$ and iterate n times to obtain, setting $\pi_n^x = \pi_n^{\delta_x}$,

$$V(x_0) \geq P^n V(x_0) + n - b \sum_{k=0}^{n-1} P^k(x_0, K) = P^n V(x_0) + n - nb\pi_n^{x_0}(K).$$

Since $V(x_0) < \infty$, rearranging terms and multiplying by n^{-1} yields

$$-\frac{1}{n} V(x_0) + 1 \leq \frac{1}{n} \{P^n V(x_0) - V(x_0)\} + 1 \leq b\pi_n^{x_0}(K). \tag{12.3.6}$$

By Proposition C.1.3, a bounded sequence of measures admits a *weak** limit point. Thus there exist $\pi \in \mathbb{M}_b(\mathscr{X})$ and a subsequence $\{n_k, k \in \mathbb{N}\}$ such that $\pi_{n_k}^{x_0} \stackrel{\mathrm{w}^*}{\Rightarrow} \pi$, and by Proposition 12.3.4, $\pi P = \pi$. By (12.3.6) and Proposition C.1.2, we obtain

$$b^{-1} \leq \limsup_{k \to \infty} \pi_{n_k}^{x_0}(K) \leq \pi(K),$$

which implies that $\pi(K) > 0$. This π is a bounded nonzero invariant measure, so it can be normalized into an invariant probability measure.

In the case $k > 1$, the previous part implies that P^k admits an invariant probability measure, and thus P admits an invariant probability measure by Lemma 1.4.7. □

12.4 Topological Recurrence

Definition 12.4.1 (Topological recurrence)

(i) A point x^* is said to be topologically recurrent if $\mathbb{E}_{x^*}[N_O] = \infty$ for all $O \in \mathscr{V}_{x^*}$.

(ii) A point x^* is said to be topologically Harris recurrent if $\mathbb{P}_{x^*}(N_O = \infty) = 1$ for all $O \in \mathscr{V}_{x^*}$.

Reachable topologically recurrent points can be used to characterize recurrence.

Theorem 12.4.2. *Let P be an irreducible Markov kernel on a complete separable metric space (X, d).*

(i) If P is recurrent, then every reachable point is topologically recurrent.
(ii) If P is a T-kernel and if there exists a reachable and topologically recurrent point, then P is recurrent.

Proof. (i) If x^* is reachable, then by definition, every $O \in \mathscr{V}_{x^*}$ is accessible. If P is recurrent, then every accessible set is recurrent, and thus $U(x^*, O) = \infty$ for every $O \in \mathscr{V}_{x^*}$, i.e., x^* is topologically recurrent.

(ii) If P is a T-kernel, then there exists a sampling distribution a such that $K_a(x, \cdot) \geq T(x, \cdot)$ for all $x \in \mathsf{X}$, and by Theorem 12.2.9, $T(x^*, \cdot)$ is an irreducibility measure. The proof is by contradiction. If P is transient, then X is a countable union of uniformly transient sets. Since $T(x^*, \cdot)$ is nontrivial, there exists a uniformly transient set B such that $T(x^*, B) > 0$. The function $x \mapsto T(x, B)$ being lower semicontinuous, by Lemma B.1.6 there exists $F \in \mathscr{V}_{x^*}$ such that $\inf_{x \in F} T(x, B) > 0$, which in turn implies that $\inf_{x \in F} K_a(x, B) = \delta > 0$. By Lemma 10.1.8 (i), this yields that F is uniformly transient. This contradicts the assumption that x_* is topologically recurrent. Therefore, P is recurrent.

□

We now provide a convenient criterion to prove the topological Harris recurrence of a point.

Theorem 12.4.3. *Let P be a Markov kernel. If $\mathbb{P}_{x^*}(\sigma_O < \infty) = 1$ for all $O \in \mathscr{V}_{x^*}$, then x^* is topologically Harris recurrent.*

Proof. We prove by induction that $\mathbb{P}_{x^*}(\sigma_V^{(j)} < \infty) = 1$ for all $j \geq 1$ and all $V \in \mathscr{V}_{x^*}$. This is true for $j = 1$ by assumption. Assume that it is true for one $j \geq 1$. For $O \in \mathscr{V}_{x^*}$, we have

$$
\begin{aligned}
\mathbb{P}_{x^*}(X_{\sigma_O} = x^*, \sigma_O^{(j+1)} < \infty) &= \mathbb{P}_{x^*}(X_{\sigma_O} = x^*, \sigma_O^{(j)} \circ \theta_{\sigma_O} < \infty) \\
&= \mathbb{P}_{x^*}(X_{\sigma_O} = x^*) . \qquad (12.4.1)
\end{aligned}
$$

Let $V_n \in \mathscr{V}_{x^*}$ be a decreasing sequence of open neighborhoods of x^* such that $V_{\subset}O$ for all $n \in \mathbb{N}$ and $\{x^*\} = \bigcap_{n\geq 1} V_n$. Then for all $n \geq 1$, the induction assumption yields

$$\mathbb{P}_{x^*}(X_{\sigma_O} \in O \setminus V_n, \sigma_O^{(j+1)} < \infty) \geq \mathbb{P}_{x^*}(X_{\sigma_O} \in O \setminus V_n, \sigma_{V_n}^{(j)} < \infty) = \mathbb{P}_{x^*}(X_{\sigma_O} \in O \setminus V_n).$$

The latter inequality implies that

$$\begin{aligned}\mathbb{P}_{x^*}(X_{\sigma_O} \in O \setminus \{x^*\}, \sigma_O^{(j+1)} < \infty) &\geq \liminf_n \mathbb{P}_{x^*}(X_{\sigma_O} \in O \setminus V_n, \sigma_O^{(j+1)} < \infty) \\ &\geq \liminf_n \mathbb{P}_{x^*}(X_{\sigma_O} \in O \setminus V_n) \\ &= \mathbb{P}_{x^*}(X_{\sigma_O} \in O \setminus \{x^*\}) = 1 - \mathbb{P}_{x^*}(X_{\sigma_O} = x^*). \end{aligned} \quad (12.4.2)$$

Combining (12.4.1) and (12.4.2) yields $\mathbb{P}_{x^*}(\sigma_O^{(j+1)} < \infty) = 1$. □

12.5 Exercises

12.1. Consider the functional autoregressive model

$$X_k = m(X_{k-1}) + \sigma(X_{k-1})Z_k, \quad k \in \mathbb{N}^*, \quad (12.5.1)$$

where $\{Z_k, k \in \mathbb{N}\}$ is an i.i.d. sequence taking values in $\mathbb{R}^p$ independent of X_0, $m : \mathbb{R}^q \mapsto \mathbb{R}^q$ is a continuous function, and $\sigma : \mathbb{R}^q \mapsto \mathbb{R}^{q\times q}$ is a matrix-valued continuous function. Denote by P the Markov kernel associated with this Markov chain.

1. Show that P is Feller.

Assume that for each $x \in \mathbb{R}^q$, $\sigma(x)$ is invertible and the function $x \mapsto \sigma^{-1}(x)$ is continuous. Assume in addition that μ admits a density g with respect to Lebesgue measure on $\mathbb{R}^q$.

2. Show that P is strong Feller.

12.2. Let $\{Z_k, k \in \mathbb{N}^*\}$ be a sequence of i.i.d. Bernoulli random variables with mean $p \in (0,1)$, independent of the random variable X_0 with values in $[0,1]$, and let $\{X_k, k \in \mathbb{N}\}$ be the Markov chain defined by the following recurrence:

$$X_{n+1} = \frac{1}{3}(X_n + Z_{n+1}), \quad n \geq 0.$$

Denote by P the Markov kernel associated with this chain. Show that P is Feller but not strong Feller.

12.3. Let P be the Markov kernel defined in Exercise 12.2. In this exercise, we show by contradiction that P is not a T-kernel. Assume to the contrary that there exist a sampling distribution $a \in \mathbb{M}_1^*(\mathbb{N})$ and a sub-Markovian kernel T such that

(i) $T(x, \mathsf{X}) > 0$ for all $x \in \mathsf{X}$;
(ii) for all $A \in \mathscr{X}$, the function $x \mapsto T(x, A)$ is lower semicontinuous;
(iii) for all $x \in \mathsf{X}$ and $A \in \mathscr{X}$, $K_a(x, A) \geq T(x, A)$.

1. Show that for all $x \in \mathbb{Q} \cap [0,1]$, $T(x, \mathbb{Q}^c \cap [0,1]) = 0$.
2. Deduce using (ii) that for all $x \in [0,1]$, $T(x, \mathbb{Q}^c \cap [0,1]) = 0$. Conclude.

12.4. Let P be the Markov kernel defined in Exercise 12.2. In this exercise, we show that P is not irreducible.

1. Show that for all $x \in \mathbb{Q} \cap [0,1]$, $\mathbb{P}_x(\sigma_{\mathbb{Q}^c \cap [0,1]} < \infty) = 0$.
2. Similarly, show that for all $x \in \mathbb{Q}^c \cap [0,1]$, $\mathbb{P}_x(\sigma_{\mathbb{Q} \cap [0,1]} < \infty) = 0$.
3. Conclude.

12.5. Let P be a Markov kernel on a metric space (X, d). Assume that there exist $\mu \in \mathbb{M}_+(\mathscr{X})$ and a bounded measurable function g on $\mathsf{X} \times \mathsf{X}$, continuous with respect to its first argument, such that $Pf(x) = \int g(x,y) f(y) \mu(\mathrm{d}y)$ for all $f \in \mathbb{F}_b(\mathsf{X})$ and $x \in \mathsf{X}$. Prove that P is strong Feller.

12.6. Let $\{Z_k, k \in \mathbb{N}\}$ be a sequence of i.i.d. random variables with common distribution μ on $\mathbb{R}^q$, independent of the $\mathbb{R}^q$-valued random variable X_0, and define, for $k \geq 1$, $X_k = X_{k-1} + Z_k$. The kernel of this Markov chain is given by $P(x, A) = \mu(A - x)$ for $x \in \mathbb{R}^q$ and $A \in \mathscr{B}(\mathbb{R}^q)$.

1. Show that the kernel P is Feller.
2. Assume that μ has a density with respect to the Lebesgue measure on $\mathbb{R}^q$. Show that P is strong Feller.

We will now prove the converse. Assume that the Markov kernel P is strong Feller.

3. Let A be a measurable set such that $\mu(A) = \delta > 0$. Show that we may choose an open set $O \in \mathscr{V}_0$ such that $P(x, A) = \mu(A - x) \geq \delta/2$ for all $x \in O$.
4. Show that $\mathrm{Leb}(A) \geq \frac{\delta}{2} \mathrm{Leb}(O) > 0$ and conclude.

12.7. A probability measure μ on $\mathscr{B}\left(\mathbb{R}^d\right)$ is said to be spread out if there exists p such that μ^{*p} is nonsingular with respect to Lebesgue measure.

Show that the following properties are equivalent.

(i) μ is spread out.
(ii) There exist $q \in \mathbb{N}^*$ and a compactly supported nonidentically zero and continuous function g such that $\mu^{*q} \geq g \cdot \mathrm{Leb}$.
(iii) There exist an open set O, $\alpha > 0$, and $q \in \mathbb{N}^*$ such that $\mathbb{1}_O \cdot \mu^{*q} \geq \alpha \mathbb{1}_O \cdot \mathrm{Leb}$.

Let P be the Markov kernel of the random walk with increment distribution μ defined by $P(x, A) = \mu(A - x)$, $x \in \mathsf{X}$, $A \in \mathscr{X}$.

4. Show that if μ is spread out, then there exist $q \in \mathbb{N}^*$ and a nonzero function $g \in \mathrm{C}_c^+(\mathbb{R}^\mathrm{d})$ such that $P^q(x, A) \geq \mathrm{Leb}(\mathbb{1}_A * g(x))$ and P is a T-kernel.

We finally show the converse: if P is a T-kernel, the increment measure is spread out. The proof is by contradiction. Assume that P is a T-kernel (i.e., there exists $a \in \mathbb{M}_1(\mathbb{N}^*)$ such that $T(x,A) \geq K_a(x,A)$ for all $x \in \mathsf{X}$ and $A \in \mathscr{X}$) and that μ is not spread out.

1. Show that there exists $A \in \mathscr{B}(\mathbb{R}^d)$ such that for all $n \geq 1$, $\mu^{*n}(A) = 1$ and $\mathrm{Leb}(A) = 0$.
2. Show that there exists a neighborhood O of 0 such that $\inf_{x \in O} K_a(x,A) \geq \delta > 0$.
3. Show that $\mathrm{Leb}(A) = \int P^n(x,A)\mathrm{d}x$.
4. Show that $\mathrm{Leb}(A) \geq \delta \mathrm{Leb}(O) > 0$ and conclude.

12.8. Consider the autoregressive process $Y_k = \alpha_1 Y_{k-1} + \cdots + \alpha_p Y_{k-p} + Z_k$ of order p, where $\{Z_k,\, k \in \mathbb{N}\}$ is an i.i.d. sequence; see Example 2.1.2. Define $\alpha(z) = 1 - \alpha_1 z^1 - \cdots - \alpha_p z^p$ and let A be the companion matrix of the polynomial $\alpha(z)$:

$$A = \begin{bmatrix} \alpha_1 & \cdots & \cdots & \alpha_p \\ 1 & 0 & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & & 1 & 0 \end{bmatrix} . \tag{12.5.2}$$

1. Show that the AR(p) model can be rewritten as a first-order vector autoregressive sequence $X_k = AX_{k-1} + BZ_k$ with $X_k = [Y_k, \ldots, Y_{k-p+1}]'$, A is the companion matrix of $\alpha(z)$, and $B = [1, 0, \ldots, 0]'$.
2. Show that the pair (A, B) is controllable.

Denote by P the Markov kernel associated with the Markov chain $\{X_k,\, k \in \mathbb{N}\}$. Assume that the distribution of Z_1 has a nontrivial continuous component with respect to Lebesgue measure.

3. Show that P is a T-kernel.
4. Assume that the zeros of the characteristic polynomials lie outside the unit circle. Show that P is an irreducible T-kernel that admits a reachable point

12.9. Let $\mathsf{X} = [0,1]$ endowed with the usual topology, $\alpha \in (0,1)$, and let P be defined by

$$P(x,0) = 1 - P(x,x) = x\,, \qquad x\; in\, [0,1]\,.$$

1. Prove that P is irreducible and Feller.
2. Prove that $\lim_{x \to 0} \mathbb{P}_x(\sigma_0 \leq n) = 0$.
3. Prove that P is Harris recurrent.
4. Prove that the state space is not petite and that X is not a T-chain.

12.10. Let $\mathsf{X} = [0,1]$ endowed with the usual topology, $\alpha \in (0,1)$, and let P be defined by

$$P(x,0) = 1 - P(x,\alpha x) = x\,, \qquad P(0,0) = 1\,.$$

1. Prove that P is irreducible and Feller.
2. Prove that $\lim_{x \to 0} \mathbb{P}_x(\sigma_0 \leq n) = 0$.
3. Prove that P is recurrent but not Harris recurrent.
4. Prove that the state space is not petite and that X is not a T-chain.

12.11. Consider the recurrence $X_k = FX_{k-1} + GZ_k$, where F is a $p \times p$ matrix, G is a $p \times q$ matrix, and $\{Z_k, k \in \mathbb{N}\}$ is an i.i.d. sequence of random Gaussian vectors in $\mathbb{R}^q$ with zero mean and identity covariance matrix. Denote by P the associated Markov kernel.

1. Show that the kernel P is irreducible, that every nontrivial measure ϕ that possesses a density on $\mathbb{R}^p$ is an irreducibility measure, and that Lebesgue measure is a maximal irreducibility measure.
2. Show that for every compact set A and set B with positive Lebesgue measure, we have $\inf_{x \in A} \mathbb{P}_x(\sigma_B < \infty) > 0$.

12.12. Let $\Phi : \mathbb{R}^s \mapsto \mathbb{R}^q$ be a linear map from $(s \geq q)$ with rank q. Let $\xi \in \mathbb{M}_1(\mathbb{R}^s)$ be a probability on $\mathbb{R}^s$ such that $\xi \geq f \cdot \mathrm{Leb}_s \neq 0$, for some nonnegative integrable function f. Show that there exists a nonnegative integrable function g such that $\xi \circ \Phi^{-1} \geq g \cdot \mathrm{Leb}_q \neq 0$.

12.13. We use the assumptions and notation of Example 12.2.10.

1. Show that there exist a small set C and an open set O containing x^* such that $\inf_{x \in O} T(x, C) = \delta > 0$.
2. Show that O is a small set.
3. If A is a compact set, then $\inf_{x \in A} P^n(x, O) = \gamma > 0$.
4. Show that every compact set is small.

12.14. Assume that there exists $\mu \in \mathbb{M}_1(\mathscr{X})$ such that the family of probability measures $\{\mu P^n, n \in \mathbb{N}\}$ is tight. Show that P admits an invariant probability.

12.15. Let P be a Feller kernel on a compact metric space. Show that P admits an invariant probability.

12.16. Let P be a Feller kernel on a metric space (X, d). Assume that there exists a nonnegative function $V \in \mathrm{C}(\mathsf{X})$ such that the sets $\{V \leq c\}$ are compact for all $c > 0$. Assume further that there exist $\lambda \in [0, 1)$ and $b \in \mathbb{R}^+$ such that

$$PV \leq \lambda V + b\,. \tag{12.5.3}$$

Show that there exists a P-invariant probability measure and each invariant probability measure π satisfies $\pi(V) < \infty$. [Hint: Use Exercise 12.14.]

12.17. Let P be a Feller kernel on (X, d).

1. Let $\mu, \pi \in \mathbb{M}_1(\mathscr{X})$. Show that if $\mu P^n \stackrel{\mathrm{w}}{\Rightarrow} \pi$, then π is P-invariant.
2. Let $\pi \in \mathbb{M}_1(\mathscr{X})$. Assume that for every $x \in \mathsf{X}$, $\delta_x P^n \stackrel{\mathrm{w}}{\Rightarrow} \pi$. Prove that π is the unique P-invariant probability and $\xi P^n \stackrel{\mathrm{w}}{\Rightarrow} \pi$ for every $\xi \in \mathbb{M}_1(\mathscr{X})$.

12.18. Consider the log-Poisson autoregressive process defined in Example 2.2.5. Assume that $|b+c| \vee |b| \vee |c| < 1$.

1. Prove that its Markov kernel P defined in (2.2.11) is Feller.
2. Prove that the drift condition (12.3.3) holds with $V(x) = \mathrm{e}^{|x|}$.
3. Conclude that an invariant probability exists under this condition.

12.19. We consider the Metropolis–Hastings algorithm introduced in Section 2.3.1. We use the notation introduced in this section: $h_\pi \in \mathbb{F}_+(\mathsf{X})$ is the unnormalized density of the target distribution π with respect to a σ-finite measure ν, $(x,y) \mapsto q(x,y)$ is the proposal density kernel. We assume below that h_π is continuous and $q : \mathsf{X} \to \mathsf{X} \to \mathbb{R}_+$ is continuous. We must define the state space of the Markov chain to be the set $\mathsf{X}_\pi = \{x \in \mathsf{X} : h_\pi(x) > 0\}$. The assumption that h_π is continuous means that X_π is an open set. Show that the Metropolis–Hastings kernel is a T-kernel.

12.20. Let $\{Z_k, k \in \mathbb{N}\}$ be an i.i.d. sequence of scalar random variables. Assume that the distribution of Z_1 has a density with respect to Lebesgue measure denoted p. Assume that p is positive and lower semicontinuous. Consider a Markov chain on $\mathbb{R}$ defined by the recurrence $X_k = F(X_{k-1}, Z_k)$, where $F : \mathbb{R} \to \mathbb{R}$ is a $C^\infty(\mathbb{R})$ function. We denote by P the associated Markov kernel.

For all $x^0 \in \mathbb{R}$ and every sequence of real numbers $\{z_k, k \in \mathbb{N}\}$, define recursively

$$F_k(x_0, z_1, \ldots, z_k) = F(F_{k-1}(x_0, z_1, \ldots, z_{k-1}), z_k)\,.$$

Assume that for each initial condition $x_0 \in \mathbb{R}$, there exist $k \in \mathbb{N}^*$ and a sequence $(z_1^0, \ldots, z_k^0)$ such that the derivative

$$\left[\, \frac{\partial F_k}{\partial u_1}(x_0, z_1^0, \ldots, z_k^0) \;\cdots\; \frac{\partial F_k}{\partial u_k}(x_0, z_1^0, \ldots, z_k^0) \,\right]$$

is nonzero. Show that P is a T-kernel.

12.21. Let P be a Markov kernel on a complete separable metric space and R the set of points that are topologically Harris recurrent. Let $\{V_n, n \in \mathbb{N}\}$ be the set of open balls with rational radius and center in a countable dense subset. Set

$$A_n(j) = \left\{ y \in V_n : \mathbb{P}_y(\sigma_{V_n} < \infty) \le 1 - 1/j \right\}\,.$$

1. Show that $\mathbb{R}^c = \bigcup_{n,j} A_n(j)$.
2. Show that $A_n(j)$ is uniformly transient and that R^c is transient.

12.22. Let ν be a probability measure that is equivalent to Lebesgue measure on $\mathbb{R}$ (e.g., the standard Gaussian distribution) and let μ be a distribution on the rational numbers such that $\mu(q) > 0$ for all $q \in \mathbb{Q}$ (which is possible, since $\mathbb{Q}$ is countable). Let P be the Markov kernel on $\mathbb{R}$ such that

$$P(x, \cdot) = \begin{cases} \nu \text{ if } x \in \mathbb{R} \setminus \mathbb{Q}\,, \\ \mu \text{ if } x \in \mathbb{Q}\,. \end{cases}$$

Prove that the kernel P is topologically Harris recurrent and admits two invariant measures.

In the following exercises, we will discuss the notion of evanescence. In all that follows, X is a locally compact separable metric space. We say that an X-valued sequence $\{u_n, n \in \mathbb{N}\}$ tends to infinity if for every compact set K of X, the set $\{n \in \mathbb{N} : u_n \in K\}$ is finite. A function $f : \mathsf{X} \to \mathbb{R}_+$ is said to tend to infinity if for all $A > 0$, there exists a compact set K such that $f(x) \geq A$ for all $x \notin K$. If $\{X_n, n \in \mathbb{N}\}$ is a stochastic process, we denote by $\{X_n \to \infty\}$ the set of paths that tend to infinity.

Since X is a locally compact separable metric space, there exists an increasing sequence $\{K_n, n \in \mathbb{N}\}$ of compact sets with nonempty interior such that $\mathsf{X} = \cup_{n \geq 0} K_n$. The event $\{X_n \to \infty\}$ is the set of paths that visit each compact set finitely many times:

$$\{X_n \to \infty\} = \bigcap_{j \geq 0} \{X_n \in K_j, \text{ i.o.}\}^c .$$

Equivalently, $X_n \not\to \infty$ if and only if there exists a compact set K that is visited infinitely often by $\{X_n\}$. In $\mathbb{R}^d$ endowed with any norm, this notion corresponds to the usual one: $X_n \to \infty$ if and only if $\lim_{n \to \infty} |X_n| = \infty$ in the usual sense.

Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$.

(i) P is said to be evanescent if for all $x \in \mathsf{X}$, $\mathbb{P}_x(X_n \to \infty) = 1$.
(ii) P is said to be nonevanescent if for all $x \in \mathsf{X}$, $\mathbb{P}_x(X_n \to \infty) = 0$.

12.23. Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P is evanescent.

1. Show that there exists an accessible compact set K and that for all $x \in \mathsf{X}$, $\mathbb{P}_x(N_K = \infty) = 0$.
2. Show that P is transient. [Hint: Proceed by contradiction: if P is recurrent, then K contains an accessible Harris recurrent set $\tilde{K}$.]

12.24. Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P is Harris recurrent.

1. Show that there exists $x_0 \in \mathsf{X}$ such that $h(x_0) < 1$, where $h(x) = \mathbb{P}_x(X_n \to \infty)$. [Hint: Use Exercise 12.23.]
2. Show that $h(x) = h(x_0)$ for all $x \in \mathsf{X}$.
3. Show that P is nonevanescent. [Hint: Show that $\mathbb{P}_{X_n}(A) = \mathbb{P}_x(A \,|\, \mathscr{F}_n)$ converges $\mathbb{P}_x -$a.s. to $\mathbb{1}_A$ for all $x \in \mathsf{X}$.]

12.25. Assume that there exist a nonnegative finite measurable function V on X and a compact set C such that $PV(x) \leq V(x)$ if $x \notin C$ and V tends to infinity.

1. Show that for all $x \in \mathsf{X}$, there exists a random variable M_∞ that is $\mathbb{P}_x$ almost surely finite for all $x \in \mathsf{X}$ such that for all $n \in \mathbb{N}$, $V(X_{n \wedge \tau_C}) \to M_\infty$.
2. Prove that $\mathbb{P}_x(\sigma_C = \infty, X_n \to \infty) = 0$ for all $x \in \mathsf{X}$.
3. Show that $\{X_n \to \infty\} = \lim_{p \to \infty} \uparrow \{X_n \to \infty, \sigma_C \circ \theta_p = \infty\}$.
4. Show that P is nonevanescent.

12.26. Assume that P is a T-kernel. Let $A \in \mathscr{X}$ be a transient set. Define

$$A^0 = \{x \in \mathsf{X} : \mathbb{P}_x(\sigma_A < \infty) = 0\} .$$

1. Let $\tilde{A} = \{x \in \mathsf{X} : \mathbb{P}_x(\sigma_A < \infty) > 0\}$. Show that $\tilde{A} = \bigcup_{i=1}^{\infty} \tilde{A}_i$, where the sets $\tilde{A}_i$ are uniformly transient.

For $i, j \in \mathbb{N}^*$, set $U_j = \left\{x : T(x, A^0) > 1/j\right\}$ and $U_{i,j} = \{x : T(x, A_i) > 1/j\}$.

2. Show that $\left\{(U_i, U_{i,j}) : i, j > 0\right\}$ is an open covering of X.
3. Let K be a compact set. Show that there exists $k \geq 1$ such that $K \subset U_k \cup \bigcup_{i=1}^{k} U_{i,k}$.
4. Show that $\{X_n \in K \text{ i.o.}\} \subset \{X_n \in U_k \text{ i.o.}\}$ $\mathbb{P}_x -$a.s.
5. Let a be a sampling distribution such that $K_a \geq T$. Show that for all $y \in U_k$, $\mathbb{P}_y(\sigma_{A^0} < \infty) \geq K_a(y, A^0) \geq T(y, A^0) = 1/k$.
6. Show that $\{X_n \in K \text{ i.o.}\} \subset \{\sigma_{A^0} < \infty\}$ $\mathbb{P}_x -$a.s. for all $x \in \mathsf{X}$ and every compact set K.
7. Show that for all $x \in \mathsf{X}$, $\mathbb{P}_x(\{X_n \to \infty\} \cup \{\sigma_{A^0} < \infty\}) = 1$.

12.27. This exercise uses the results obtained in Exercises 12.23, 12.24, and 12.26. Let P be an irreducible T-kernel.

1. P is transient if and only if P is evanescent.
2. P is recurrent if and only if there exists $x \in \mathsf{X}$ such that $\mathbb{P}_x(X_n \to \infty) < 1$.
3. P is Harris recurrent if and only if P is nonevanescent. [Hint: If P is nonevanescent, then P is recurrent by question 2, and by Theorem 10.2.7, we can write $\mathsf{X} = H \cup N$ with H maximal absorbing, N transient, and $H \cap N = \emptyset$, where H is maximal absorbing and N is transient. Prove that N is empty.]

12.6 Bibliographical Notes

The concept of Feller chains was introduced by W. Feller. Numerous results on Feller chains were obtained in the works of Foguel (1962, 1968, 1969) and Lin (1970, 1971); see Foguel (1973) for a review of these early references.

Most of the results in Section 12.3 were first established in Foguel (1962, 1968). The presentation of the results and the proofs in this section follow closely (Hernández-Lerma and Lasserre, 2003, Chapter 7).

12.A Linear Control Systems

Let p, q be integers and $\{u_k, k \in \mathbb{N}\}$ a deterministic sequence of vectors in $\mathbb{R}^q$. Denote by F a $p \times p$ matrix and let G be a $q \times q$ matrix. Consider the sequence $\{x_k, k \in \mathbb{N}\}$ of vectors in $\mathbb{R}^p$ defined recursively for $k \geq 1$ by

$$x_k = F x_{k-1} + G u_k \, . \tag{12.A.1}$$

These equations define a linear system. The sequence $\{u_k, k \in \mathbb{N}\}$ is called the input. The solution to the difference equation (12.A.1) can be expressed explicitly as follows:

$$x_k = F^k x_0 + \sum_{\ell=0}^{k-1} F^\ell G u_{k-\ell} \, . \tag{12.A.2}$$

The pair of matrices (F,G) is controllable if for each pair of states $x_0, \; x^\star \in \mathsf{X} = \mathbb{R}^p$, there exist an integer m and a sequence $(u_1^\star, \ldots, u_m^\star) \in \mathbb{R}^q$ such that $x_m = x^\star$ when $(u_1, \ldots, u_m) = (u_1^\star, \ldots, u_m^\star)$ and the initial condition is equal to x_0. In words, controllability asserts that the inputs u_k can be chosen in such a way that every terminal state $x^\star$ can be reached from any starting point x_0.

For every integer k, using some control sequence $(u_1, \ldots, u_m)$, we have

$$x_m = F^m x_0 + [G|FG|\cdots|F^{m-1}G] \begin{pmatrix} u_m \\ \vdots \\ u_1 \end{pmatrix} .$$

The linear recurrence is controllable if for some integer r, the range space of the matrix

$$C_r = [G|FG|\cdots|F^{r-1}G] \tag{12.A.3}$$

is equal to $\mathbb{R}^p$. Define

$$m(F,G) = \inf\{r > 0 : \operatorname{rank}(C_r) = p\} \, , \tag{12.A.4}$$

with the usual convention $\inf \emptyset = \infty$. The pair (F,G) is said to be controllable if $m(F,G) < \infty$. Clearly, if $\operatorname{rank}(G) = p$, then $m(F,G) = 1$.

Moreover, if the pair (F,G) is controllable, then $m(F,G) \le m_0 \le n$, where m_0 is the degree of the minimal polynomial of F. Note that the minimal polynomial is the monic polynomial α of lowest degree for which $\alpha(F) = 0$. For every $r > m_0$, F^{r-1} can be expressed as a linear combination of $F^{r-2}, \ldots, I$; hence $C_r = C_{m_0}$.

Part III
Irreducible Chains: Advanced Topics

Chapter 13
Rates of Convergence for Atomic Markov Chains

In this chapter we will complement the results obtained in Chapter 8 on the convergence of the distribution of the nth iterate of a positive recurrent atomic Markov chain to its invariant distribution. We will go beyond the geometric and polynomial rates of convergence considered in Section 8.3. In Section 13.1, we will introduce general subgeometric rates, which include the polynomial rate. We will also extend the results of Section 8.3 (which dealt only with convergence in total variation distance) to convergence in the f-total variation distance for certain unbounded functions $f \geq 1$.

These results will be obtained by means of the same coupling method as in Section 8.3: given a kernel P that admits an accessible atom, we consider two independent copies of a Markov chain with kernel P, and the coupling time T will simply be the first time when both chains simultaneously visit the atom. In Section 13.2 we will recall this construction and give a number of (very) technical lemmas whose purpose will be to relate modulated moments of the return time to the atom to similar moments for the coupling time T. As a reward for our efforts, we will easily obtain in Sections 13.3 and 13.4 our main results.

13.1 Subgeometric Sequences

A subgeometric sequence increases to infinity more slowly than any exponential sequence, that is, it satisfies $\limsup_{n\to\infty} \log r(n)/n = 0$. For instance, a polynomial sequence is subgeometric. This first definition is not always sufficiently precise and must be refined. We first introduce the following notation which will be often used. Given a sequence $r : \mathbb{N} \to \mathbb{R}$, we define its primitive r^0 by

$$r^0(n) = \sum_{k=0}^{n} r(k) \;, \quad n \geq 0 \;. \tag{13.1.1}$$

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_13

We now introduce the sets of subgeometric sequences. We will obviously impose restrictions on the type of sequences we can consider, the mildest being the log-subadditivity.

Definition 13.1.1 (Log-subadditive sequences) *A sequence* $r : \mathbb{N} \to [1,\infty)$ *is said to be log-subadditive if* $r(n+m) \leq r(n)r(m)$ *for all* $n,m \in \mathbb{N}$. *The set* $\mathscr{S}$ *is the set of nondecreasing log-subbaditive sequence.*

The set $\bar{\mathscr{S}}$ *is the set of sequences* r *such that there exist a sequence* $\tilde{r} \in \mathscr{S}$ *and constants* $c_1, c_2 \in (0,\infty)$ *that satisfy* $c_1\tilde{r} \leq r \leq c_2\tilde{r}$.

If $r \in \bar{\mathscr{S}}$, then r is not necessarily increasing but there exists a constant M_r such that $r(n+m) \leq M_r r(n)r(m)$ for all $n,m \geq 0$. Geometric sequences $\{\beta^n,\ n \in \mathbb{N}\}$ with $\beta \geq 1$ belong to $\mathscr{S}$.

Definition 13.1.2

(i) Λ_0 *the set of sequences* $r \in \mathscr{S}$ *such that the sequence* $n \mapsto n^{-1}\log r(n)$ *is non-increasing and* $\lim_{n\to\infty} n^{-1}\log r(n) = 0$.
(ii) Λ_1 *is the set of sequences* $r \in \mathscr{S}$ *such that* $\lim_{n\to\infty} r(n+1)/r(n) = 1$.
(iii) Λ_2 *is the set of sequences* $r \in \mathscr{S}$ *such that* $\limsup_{n\to\infty} r(n)/r^0(n) = 0$.

For $i \in 1,2$, $\bar{\Lambda}_i$ *is the set of sequences* r *such that there exist* $0 < c_1 < c_2 < \infty$ *and* $r_i \in \Lambda_i$ *satisfying* $c_1 r_i \leq r \leq c_2 r_i$.

It is easily shown that following sequences belong to Λ_0

(a) Logarithmic sequences: $\log^\beta(1+n)$, $\beta > 0$.
(b) Polynomial sequences: $(1+n)^\beta$, $\beta > 0$.
(c) Subexponential sequences: $\{1+\log(1+n)\}^\alpha(n+1)^\beta \mathrm{e}^{cn^\gamma}$, for $\alpha,\beta \in \mathbb{R}$, $\gamma \in (0,1)$ and $c > 0$.

Lemma 13.1.3 *(i)* $\Lambda_0 \subset \Lambda_1 \subset \Lambda_2$.
(ii) Let $r \in \Lambda_1$. *For every* $\varepsilon > 0$ *and* $m_0 \in \mathbb{N}$, *there exists* $M < \infty$ *such that, for all* $n \geq 0$ *and* $m \leq m_0$, $r(n+m) \leq (1+\varepsilon)r(n) + M$.

Proof. (i) Let $r \in \Lambda_0$. Since $n \mapsto n^{-1}\log r(n)$ is decreasing, we have

$$0 \leq \log\frac{r(n+1)}{r(n)} = (n+1)\frac{\log r(n+1)}{n+1} - (n+1)\frac{\log r(n)}{n} + \frac{\log r(n)}{n} \leq \frac{\log r(n)}{n} .$$

Since moreover $\lim_{n\to\infty} n^{-1}\log r(n) = 0$, then $\lim_{n\to\infty} r(n+1)/r(n) = 1$. This establishes the inclusion $\Lambda_0 \subset \Lambda_1$.

Let $r \in \Lambda_1$. By induction, it obviously holds that

$$\lim_{n\to\infty} r(n+k)/r(n) = \lim_{n\to\infty} r(n-k)/r(n) = 1 ,$$

for every $k \geq 1$. Thus, for every $m \geq 1$, we have

$$\liminf_{n\to\infty} \sum_{k=0}^{n} \frac{r(k)}{r(n)} \geq \liminf_{n\to\infty} \sum_{k=n-m}^{n} \frac{r(k)}{r(n)} = m+1 .$$

Since m is arbitrary, this proves that $\lim_{n\to\infty} r^0(n)/r(n) = \infty$. This proves that $\Lambda_1 \subset \Lambda_2$

(ii) Fix $\varepsilon > 0$ and $m_0 \geq 1$. There exists $n_0 \in \mathbb{N}$ such that, for all $n \geq n_0$, $r(n+m_0) \leq r(n)\{1+\varepsilon\}$. Set $M = r(n_0+m_0)$. Then, for all $n \geq 0$ and $m \leq m_0$, since r is increasing, we obtain

$$\begin{aligned} r(n+m) &\leq r(n+m_0) \leq (1+\varepsilon) r(n) \mathbb{1}\{n \geq n_0\} + r(n_0+m_0)\mathbb{1}\{n < n_0\} \\ &\leq r(n)(1+\varepsilon) + M . \end{aligned}$$

□

Lemma 13.1.4 *(i) If $r \in \mathscr{S}$, then $r^0 \in \bar{\mathscr{S}}$.*
(ii) If $r \in \Lambda_i$, then $r^0 \in \bar{\Lambda}_i$, $i = 1,2$.

Proof. (i) If $r \in \mathscr{S}$, then

$$\begin{aligned} r^0(m+n) &= r^0(m) + \sum_{i=1}^{n} r(m+i) \leq r^0(m-1) + r(m) r^0(n) \\ &\leq r^0(m) + r(m) r^0(n) \leq 2 r^0(m) r^0(n) . \end{aligned}$$

Thus $2r^0 \in \mathscr{S}$.

(ii) If $r \in \Lambda_2$ (which includes the case $r \in \Lambda_1$),

$$\frac{r^0(n+1)}{r^0(n)} = 1 + \frac{r(n+1)}{r^0(n)} \leq 1 + r(1)\frac{r(n)}{r^0(n)} \to 1 ,$$

as $n \to \infty$ and thus $r^0 \in \Lambda_1 \subset \Lambda_2$. This also proves that $r \in \Lambda_1$ implies $r^0 \in \Lambda_1$.

□

13.2 Coupling Inequalities for Atomic Markov Chains

Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an accessible atom α. For convenience, we reintroduce here the notation and definitions of Section 8.3. Define the Markov kernel $\bar{P}$ on $\mathsf{X}^2 \times \mathscr{X}^{\otimes 2}$ as follows: for all $(x,x') \in \mathsf{X}^2$ and $A \in \mathscr{X}^{\otimes 2}$

$$\bar{P}((x,x'),A) = \int P(x,\mathrm{d}y) P(x',\mathrm{d}y') \mathbb{1}_A(y,y') . \tag{13.2.1}$$

Let $\{(X_n, X'_n),\ n \in \mathbb{N}\}$ be the canonical process on the canonical product space $\Omega = (\mathsf{X} \times \mathsf{X})^{\mathbb{N}}$. For $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$, let $\bar{\mathbb{P}}_{\xi\otimes\xi'}$ be the probability measure on Ω such that

$\{(X_n, X'_n),\ n \in \mathbb{N}\}$ is a Markov chain with kernel P and initial distribution $\xi \otimes \xi'$. The notation $\bar{\mathbb{E}}_{\xi \otimes \xi'}$ stands for the associated expectation operator. An important feature is that $\alpha \times \alpha$ is an atom for $\bar{P}$. Indeed, for all $x, x' \in \alpha$ and $A, A' \in \mathscr{X}$,

$$\bar{P}((x,x'), A \times A') = P(x,A)P(x',A) = P(\alpha,A)P(\alpha,A') .$$

For an initial distribution $\xi' \in \mathbb{M}_1(\mathscr{X})$ and a random variable Y on Ω, if the function $x \mapsto \bar{\mathbb{E}}_{\delta_x \otimes \xi'}[Y]$ does not depend on $x \in \alpha$, then we write $\bar{\mathbb{E}}_{\alpha \times \xi'}[Y]$ for $\bar{\mathbb{E}}_{\delta_x \otimes \xi'}[Y]$ when $x \in \alpha$. Similarly, for $x, x' \in \alpha$, we write $\bar{\mathbb{E}}_{\alpha \times \alpha}[Y]$ for $\bar{\mathbb{E}}_{\delta_x \otimes \delta_{x'}}[Y]$ if the latter quantity is constant on $\alpha \times \alpha$.

Denote by T the return time to $\alpha \times \alpha$ for the Markov chain $\{(X_n, X'_n),\ n \in \mathbb{N}\}$, i.e.

$$T = \sigma_{\alpha \times \alpha} = \inf\left\{n \geq 1 \,:\, (X_n, X'_n) \in \alpha \times \alpha\right\} . \tag{13.2.2}$$

By Lemma 8.3.1, we know that

- For all $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$ and all $n \in \mathbb{N}$,

$$\mathrm{d_{TV}}(\xi P^n, \xi' P^n) \leq \bar{\mathbb{P}}_{\xi \otimes \xi'}(T \geq n) , \tag{13.2.3}$$

- For every nonnegative sequence $\{r(n),\ n \in \mathbb{N}\}$,

$$\sum_{n \geq 0} r(n) \mathrm{d_{TV}}(\xi P^n, \xi' P^n) \leq \bar{\mathbb{E}}_{\xi \otimes \xi'}\left[r^0(T)\right] , \tag{13.2.4}$$

  where $r^0(n) = \sum_{k=0}^{n} r(k)$ for all $n \in \mathbb{N}$.

We will establish bounds on the coupling time T by considering the following sequence of stopping times. Fix a positive integer q and let $\bar{\theta}$ be the shift operator on $(\mathsf{X} \times \mathsf{X})^{\mathbb{N}}$: for all $x = \{(x_k, x'_k),\ k \in \mathbb{N}\}$, $\bar{\theta}(x) = y$ where $y = \left\{(x_{k+1}, x'_{k+1}) \,:\, k \in \mathbb{N}\right\}$. Now, define

$$\nu_{-1} = \sigma_{\alpha \times \mathsf{X}} \wedge \sigma_{\mathsf{X} \times \alpha} , \quad \nu_0 = \sigma_{\alpha \times \mathsf{X}} \vee \sigma_{\mathsf{X} \times \alpha}$$

and for $k \geq 0$,

$$\nu_{k+1} = \begin{cases} \infty & \text{if } \nu_k = \infty , \\ \nu_k + q + \tau_{\mathsf{X} \times \alpha} \circ \bar{\theta}_{\nu_k + q} \mathbb{1}\left\{X_{\nu_k} \in \alpha\right\} & \\ \qquad + \tau_{\alpha \times \mathsf{X}} \circ \bar{\theta}_{\nu_k + q} \mathbb{1}\left\{X_{\nu_k} \notin \alpha\right\} , & \text{if } \nu_k < \infty . \end{cases}$$

For all $k \geq 0$, set

$$U_k = \nu_k - \nu_{k-1} . \tag{13.2.5}$$

For $j \in \mathbb{N}$, let $\mathscr{B}_j$ be the σ-algebra defined by

$$\mathscr{B}_j = \bar{\mathscr{F}}_{\nu_{j-1}} \vee \sigma(U_j) , \tag{13.2.6}$$

Obviously, $\bar{\mathscr{F}}_{\nu_{j-1}} \subset \mathscr{B}_j \subset \bar{\mathscr{F}}_{\nu_j}$.

By construction, at time ν_k, if finite, then at least one of the components of the chain (X_n, X'_n) is in α. If both components are in α, then weak coupling occurs.

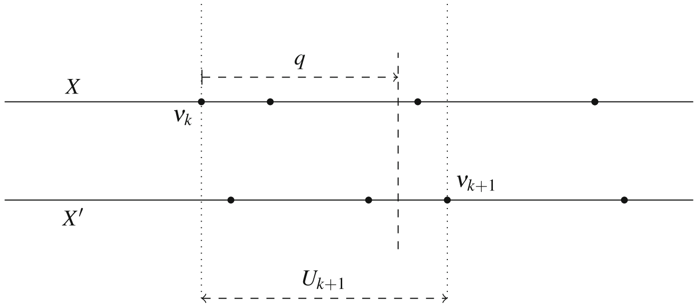

Fig. 13.1 The dots stand for the time indices when X_k or X'_k enters the atom α. In this particular example, the event $\{X_{\nu_k} \in \alpha\}$ holds and ν_{k+1} is the first time index after $\nu_k + q$ that $X'_k \in \alpha$.

If only one component is in α at time ν_k, then ν_{k+1} is the return time to α of the other component after time $\nu_k + q$, for a time lag q. If the atom is recurrent, all the stopping times ν_k are almost surely finite.

Lemma 13.2.1 *Let α be a recurrent atom. Then for all initial distributions ξ and ξ' such that $\mathbb{P}_\xi(\sigma_\alpha < \infty) = \mathbb{P}_{\xi'}(\sigma_\alpha < \infty) = 1$ and all $k \in \mathbb{N}$,*

$$\bar{\mathbb{P}}_{\xi \otimes \xi'}(\nu_k < \infty) = 1 .$$

Proof. Since α is a recurrent atom and $\mathbb{P}_\xi(\sigma_\alpha < \infty) = 1$, we have

$$\bar{\mathbb{P}}_{\xi \otimes \xi'}(N_{\alpha \times \mathsf{X}} = \infty) = \mathbb{P}_\xi(N_\alpha = \infty) = 1 .$$

Similarly, $\bar{\mathbb{P}}_{\xi \otimes \xi'}(N_{\mathsf{X} \times \alpha} = \infty) = \mathbb{P}_{\xi'}(N_\alpha = \infty) = 1$. Noting that $\{N_{\mathsf{X} \times \alpha} = \infty, N_{\alpha \times \mathsf{X}} = \infty\} \subset \{\nu_k < \infty\}$, we obtain $\bar{\mathbb{P}}_{\xi \otimes \xi'}(\nu_k < \infty) = 1$ for all $k \in \mathbb{N}$. □

Remark 13.2.2 *The dependence in q is implicit in the notation but is crucial and should be kept in mind. If α is an accessible, aperiodic and positive atom, Corollary 8.2.3 implies that for every $\gamma \in (0, \pi(\alpha))$, there exists $q \in \mathbb{N}^*$ such that for all $n \geq q$, $\mathbb{P}_\alpha(X_n \in \alpha) = P^n(\alpha, \alpha) > \gamma$. In the rest of the chapter, we fix one arbitrary $\gamma \in (0, \pi(\alpha))$ and q is chosen in such a way in the definition of the stopping times ν_k.*

Consider the first time κ in the sequence $\{\nu_k,\ k \in \mathbb{N}\}$ where both chains are simultaneously in α, that is,

$$\kappa = \inf\left\{n \geq 0 :\ (X_{\nu_n}, X'_{\nu_n}) \in \alpha \times \alpha\right\} .$$

By construction, T is bounded by ν_κ:

$$T \leq \nu_\kappa . \tag{13.2.7}$$

The following lemma will be used several times.

Lemma 13.2.3 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P admits an accessible, aperiodic and positive atom α. Let h be a nonnegative function on $\mathbb{N}$ and define $H(u) = \mathbb{E}_\alpha[h(q + \tau_\alpha \circ \theta_{u+q})]$. Then, for all $j \in \mathbb{N}$,*

$$\bar{\mathbb{E}}\left[h(U_{j+1}) \,\middle|\, \mathscr{B}_j\right] = H(U_j) ,$$

where U_j and $\mathscr{B}_j$ are defined in (13.2.5) and (13.2.6), respectively. Moreover, for all $f \in \mathbb{F}_+(\mathsf{X})$ and $j \in \mathbb{N}$,

$$\mathbb{1}_\alpha(X_{\nu_{j-1}})\bar{\mathbb{E}}\left[f(X_{\nu_j}) \,\middle|\, \mathscr{B}_j\right] = \mathbb{1}_\alpha(X_{\nu_{j-1}})P^{U_j}f(\alpha) ,$$
$$\mathbb{1}_{\alpha^c}(X_{\nu_{j-1}})\bar{\mathbb{E}}\left[f(X'_{\nu_j}) \,\middle|\, \mathscr{B}_j\right] = \mathbb{1}_{\alpha^c}(X_{\nu_{j-1}})P^{U_j}f(\alpha) .$$

Proof. Let $j \in \mathbb{N}$ be fixed. Since $\mathscr{B}_j = \bar{\mathscr{F}}_{\nu_{j-1}} \vee \sigma(U_j)$, it is sufficient to show that, for all $A \in \bar{\mathscr{F}}_{\nu_{j-1}}$ and all $k \geq q$,

$$\bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}_A \mathbb{1}_{\{U_j=k\}} h(U_{j+1})] = \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}_A \mathbb{1}_{\{U_j=k\}} H(U_j)] \tag{13.2.8}$$

$$\bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}_\alpha(X_{\nu_{j-1}})\mathbb{1}_A \mathbb{1}_{\{U_j=k\}} f(X_{\nu_j})] = \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}_\alpha(X_{\nu_{j-1}})\mathbb{1}_A \mathbb{1}_{\{U_j=k\}} P^{U_j}f(\alpha)] \tag{13.2.9}$$

$$\begin{aligned}&\bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}_{\alpha^c}(X_{\nu_{j-1}})\mathbb{1}_A \mathbb{1}_{\{U_j=k\}} f(X'_{\nu_j})] \\ &\qquad = \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}_{\alpha^c}(X_{\nu_{j-1}})\mathbb{1}_A \mathbb{1}_{\{U_j=k\}} P^{U_j}f(\alpha)]\end{aligned} \tag{13.2.10}$$

By Lemma 13.2.1, we have $\bar{\mathbb{P}}(\nu_{j-1} < \infty) = 1$. Thus, applying the strong Markov property yields

$$\begin{aligned}&\bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}_\alpha(X_{\nu_{j-1}})\mathbb{1}_A \mathbb{1}\left\{U_j = k\right\} h(U_{j+1})] \\ &= \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbb{1}_\alpha(X_{\nu_{j-1}})\mathbb{1}_A \mathbb{1}\left\{\tau_{\mathsf{X}\times\alpha} \circ \bar{\theta}_{\nu_{j-1}+q} = k-q\right\} h(q + \tau_{\alpha\times\mathsf{X}} \circ \bar{\theta}_{\nu_{j-1}+q+k})\right] \\ &= \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbb{1}_\alpha(X_{\nu_{j-1}})\mathbb{1}_A \bar{\mathbb{E}}_{(\alpha, X'_{\nu_{j-1}})}[\mathbb{1}\left\{\tau_{\mathsf{X}\times\alpha} \circ \bar{\theta}_q = k-q\right\} h(q + \tau_{\alpha\times\mathsf{X}} \circ \bar{\theta}_{q+k})]\right]\end{aligned}$$

which implies that

$$\begin{aligned}\bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}_\alpha(X_{\nu_{j-1}})\mathbb{1}_A \mathbb{1}\left\{U_j = k\right\} h(U_{j+1})] &= \mathbb{E}_\alpha[h(q + \tau_\alpha \circ \theta_{q+k})] \\ &\times \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbb{1}_\alpha(X_{\nu_{j-1}})\mathbb{1}_A \mathbb{E}_{X'_{\nu_{j-1}}}[\mathbb{1}\left\{\tau_\alpha \circ \theta_q = k-q\right\}]\right] .\end{aligned} \tag{13.2.11}$$

Using this equality with $h \equiv 1$, we get

$$\begin{aligned}&\bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}_\alpha(X_{\nu_{j-1}})\mathbb{1}_A \mathbb{1}\left\{U_j = k\right\}] \\ &\qquad = \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbb{1}_\alpha(X_{\nu_{j-1}})\mathbb{1}_A \mathbb{E}_{X'_{\nu_{j-1}}}[\mathbb{1}\left\{\tau_\alpha \circ \theta_q = k-q\right\}]\right]\end{aligned} \tag{13.2.12}$$

Finally, plugging (13.2.12) into (13.2.11) and using the definition of H,

$$\bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}_\alpha(X_{\nu_{j-1}})\mathbb{1}_A\mathbb{1}\left\{U_j=k\right\}h(U_{j+1})]=\bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}_\alpha(X_{\nu_{j-1}})\mathbb{1}_A\mathbb{1}\left\{U_j=k\right\}H(U_j)]\,.$$

Similarly,

$$\begin{aligned}&\bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbb{1}_{\alpha^c}(X_{\nu_{j-1}})\mathbb{1}_A\mathbb{1}\left\{U_j=k\right\}h(U_{j+1})\right]\\&\qquad=\bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbb{1}_{\alpha^c}(X_{\nu_{j-1}})\mathbb{1}_A\mathbb{1}\left\{U_j=k\right\}H(U_j)\right]\,.\end{aligned}$$

Thus, (13.2.9) is shown. The proof of (13.2.8) and (13.2.10) follow the same lines and are omitted for brevity. □

Lemma 13.2.4 *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$. Assume that P admits an accessible, aperiodic and positive atom α. For every nonnegative sequence $\{r(n),\,n\in\mathbb{N}\}$ and $j\geq 1$,*

$$\bar{\mathbb{E}}\left[\mathbb{1}\left\{\kappa>j\right\}r(\nu_j)\,\middle|\,\bar{\mathscr{F}}_{\nu_{j-1}}\right]\leq(1-\gamma)\mathbb{1}\left\{\kappa>j-1\right\}\bar{\mathbb{E}}\left[r(\nu_j)\,\middle|\,\bar{\mathscr{F}}_{\nu_{j-1}}\right]\,. \quad (13.2.13)$$

Proof. First note that by Lemma 13.2.1, $\bar{\mathbb{P}}_{\xi\otimes\xi'}(\nu_k<\infty)=1$ for all $k\in\mathbb{N}$. Assume now for instance that $X_{\nu_{j-1}}\in\alpha$. Applying Lemma 13.2.3 and recalling that by construction $U_j\geq q$, we obtain

$$\begin{aligned}&\bar{\mathbb{E}}\left[\mathbb{1}\left\{\kappa>j\right\}r(\nu_j)\,\middle|\,\bar{\mathscr{F}}_{\nu_{j-1}}\right]\mathbb{1}_\alpha(X_{\nu_{j-1}})\\&=\bar{\mathbb{E}}\left[\mathbb{1}_{\alpha^c}(X_{\nu_j})r(\nu_j)\,\middle|\,\bar{\mathscr{F}}_{\nu_{j-1}}\right]\mathbb{1}_\alpha(X_{\nu_{j-1}})\mathbb{1}\left\{\kappa>j-1\right\}\\&=\bar{\mathbb{E}}\left[\bar{\mathbb{E}}\left[\mathbb{1}_{\alpha^c}(X_{\nu_j})\,\middle|\,\mathscr{B}_j\right]r(\nu_j)\,\middle|\,\bar{\mathscr{F}}_{\nu_{j-1}}\right]\mathbb{1}_\alpha(X_{\nu_{j-1}})\mathbb{1}\left\{\kappa>j-1\right\}\\&=\bar{\mathbb{E}}\left[P^{U_j}(\alpha,\alpha^c)r(\nu_j)\,\middle|\,\bar{\mathscr{F}}_{\nu_{j-1}}\right]\mathbb{1}_\alpha(X_{\nu_{j-1}})\mathbb{1}\left\{\kappa>j-1\right\}\\&\leq(1-\gamma)\bar{\mathbb{E}}\left[r(\nu_j)\,\middle|\,\bar{\mathscr{F}}_{\nu_{j-1}}\right]\mathbb{1}_\alpha(X_{\nu_{j-1}})\mathbb{1}\left\{\kappa>j-1\right\}\,.\end{aligned}$$

This proves (13.2.13). □

Taking $r\equiv 1$ in (13.2.13) yields

$$\bar{\mathbb{P}}\left(\kappa>j\,|\,\bar{\mathscr{F}}_{\nu_0}\right)\leq(1-\gamma)^j\,. \quad (13.2.14)$$

Lemma 13.2.5 *Let P be a Markov kernel with a recurrent atom α. Then, for all $n\in\mathbb{N}$ and $k\in\mathbb{N}$,*

$$\mathbb{P}_\alpha(\tau_\alpha\circ\theta_n=k)\leq\mathbb{P}_\alpha(\sigma_\alpha>k)\,. \quad (13.2.15)$$

(i) For every nonnegative sequence $\{r(n),\,n\in\mathbb{N}\}$,

$$\mathbb{E}_\alpha[r(\tau_\alpha\circ\theta_n)]\leq\mathbb{E}_\alpha[r^0(\sigma_\alpha-1)]\,,$$

(ii) For a nonnegative sequence r, set $\bar{r}(n) = \max_{0 \le j \le n} r(j)$. If $\mathbb{E}_\alpha[\bar{r}(\sigma_\alpha)] < \infty$, then,

$$\mathbb{E}_\alpha[r(\tau_\alpha \circ \theta_n)] \le n \mathbb{E}_\alpha[\bar{r}(\sigma_\alpha)] , \tag{13.2.16}$$

$$\lim_{n\to\infty} n^{-1} \mathbb{E}_\alpha[r(\tau_\alpha \circ \theta_n)] = 0 . \tag{13.2.17}$$

(iii) Assume that there exists $\beta > 1$ such that $\mathbb{E}_\alpha[\beta^{\sigma_\alpha}] < \infty$. Then, for all $q \ge 0$ and $\varepsilon > 0$, there exists $\delta \in (1, \beta)$ such that

$$\sup_{n\in\mathbb{N}} \mathbb{E}_\alpha[\delta^{q+\tau_\alpha \circ \theta_n}] \le 1 + \varepsilon .$$

Proof. Set $\sigma_\alpha^{(0)} = 0$ and for $k \ge 0$, $p_k = \mathbb{P}_\alpha(\sigma_\alpha = k)$ and $q_k = \mathbb{P}_\alpha(\sigma_\alpha > k) = \sum_{j>k} p_j$. Then,

$$\begin{aligned}
\mathbb{P}_\alpha(\tau_\alpha \circ \theta_n = k) &= \sum_{j=0}^{\infty} \mathbb{P}_\alpha(\sigma_\alpha^{(j)} < n \le \sigma_\alpha^{(j+1)}, \tau_\alpha \circ \theta_n = k) \\
&= \sum_{j=0}^{\infty} \sum_{i=0}^{n-1} \mathbb{P}_\alpha(\sigma_\alpha^{(j)} = i, \sigma_\alpha \circ \theta_{\sigma_\alpha^{(j)}} = k + n - i) \\
&= \sum_{i=0}^{n-1} \sum_{j=0}^{\infty} \mathbb{P}_\alpha(\sigma_\alpha^{(j)} = i) \mathbb{P}_\alpha(\sigma_\alpha = k + n - i) \le \sum_{i=0}^{n-1} p_{k+n-i} = \sum_{j=1}^{n} p_{k+j} .
\end{aligned} \tag{13.2.18}$$

This proves (13.2.15).

(i) Follows from (13.2.15) by summation by parts.
(ii) Using (13.2.18) and the fact that $\bar{r}$ is increasing,

$$\begin{aligned}
\mathbb{E}_\alpha[\bar{r}(\tau_\alpha \circ \theta_n)] &\le \sum_{j=1}^{n} \sum_{k=0}^{\infty} \bar{r}(k) p_{k+j} \le \sum_{j=1}^{n} \sum_{k=0}^{\infty} \bar{r}(k+j) p_{k+j} \\
&\le \sum_{j=1}^{n} \mathbb{E}_\alpha[\bar{r}(\sigma_\alpha)] \le n \mathbb{E}_\alpha[\bar{r}(\sigma_\alpha)] .
\end{aligned}$$

This proves (13.2.16) by noting that $r \le \bar{r}$. Since $\bar{r}$ is increasing, using again (13.2.18), we obtain

$$\begin{aligned}
\frac{1}{n} \mathbb{E}_\alpha[\bar{r}(\tau_\alpha \circ \theta_n)] &\le \frac{1}{n} \sum_{k=1}^{\infty} \bar{r}(k) \sum_{j=1}^{n} p_{k+j} \\
&= \frac{1}{n} \sum_{j=1}^{n} \left\{ \sum_{k=j+1}^{\infty} \bar{r}(k-j) p_k \right\} \le \frac{1}{n} \sum_{j=1}^{n} \left\{ \sum_{k=j+1}^{\infty} \bar{r}(k) p_k \right\} .
\end{aligned}$$

Since $\lim_{j\to\infty} \sum_{k=j+1}^{\infty} \bar{r}(k) p_k = 0$, this proves (13.2.17) by noting that $r \le \bar{r}$.
(iii) Set $q_j = \sum_{k>j} p_k$. For $\varepsilon > 0$, we have

$$\sum_{k=0}^{\infty} \beta^k q_k = \sum_{k=0}^{\infty} \beta^k \sum_{j>k} p_j = \sum_{j=1}^{\infty} p_j \sum_{k=0}^{j-1} \beta^k$$
$$\leq \frac{1}{\beta-1} \sum_{j=1}^{\infty} \beta^j p_j = \mathbb{E}_\alpha[\beta^{\sigma_\alpha}]/(\beta-1) < \infty \,.$$

Now, choose ℓ sufficiently large so that $\beta^q \sum_{k=\ell}^{\infty} \beta^k q_k \leq \varepsilon/2$. This integer ℓ being fixed, pick $\delta \in (1,\beta)$ such that $\delta^{q+\ell} \leq 1+\varepsilon/2$. The proof is completed by using again (13.2.18),

$$\mathbb{E}_\alpha[\delta^{q+\tau_\alpha\circ\theta_n}] = \mathbb{E}_\alpha\left[\delta^{q+\tau_\alpha\circ\theta_n}\mathbb{1}_{\{\tau_\alpha\circ\theta_n\leq\ell\}}\right] + \mathbb{E}_\alpha\left[\delta^{q+\tau_\alpha\circ\theta_n}\mathbb{1}_{\{\tau_\alpha\circ\theta_n>\ell\}}\right]$$
$$\leq \delta^{q+\ell} + \beta^q \sum_{k=\ell}^{\infty} \beta^k \mathbb{P}_\alpha(\tau_\alpha\circ\theta_n = k) \leq \delta^{q+\ell} + \beta^q \sum_{k=\ell}^{\infty} \beta^k q_k \leq 1+\varepsilon \,.$$

□

Lemma 13.2.6 *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$. Assume that P admits an accessible, aperiodic and positive atom α. Let $r = \{r(n),\ n\in\mathbb{N}\}$ be a positive sequence and set $\bar{r}(n) = \max_{0\leq j\leq n} r(j)$.*

(i) If $\mathbb{E}_\alpha[\bar{r}(\sigma_\alpha+q)] < \infty$, then, for every $\rho \in (1-\gamma,1)$, there exists a constant C such that for all $j\in\mathbb{N}$,

$$\bar{\mathbb{E}}\left[r(U_{j+1})\mathbb{1}\{\kappa>j\}\,\middle|\,\bar{\mathscr{F}}_{\nu_0}\right] \leq C\rho^j U_0 \,. \tag{13.2.19}$$

(ii) If $\mathbb{E}_\alpha[r^0(\sigma_\alpha+q-1)] < \infty$, then for every $\varepsilon>0$, there exists an integer ℓ such that for all $n\geq\ell$ and all $j\in\mathbb{N}$,

$$\bar{\mathbb{E}}\left[r(U_{j+1})\mathbb{1}\left\{U_{j+1}\geq n\right\}\,\middle|\,\mathscr{B}_j\right] \leq \varepsilon \,. \tag{13.2.20}$$

Proof. (i) Since $\{\kappa > j-1\} \in \bar{\mathscr{F}}_{\nu_{j-1}} \subset \mathscr{B}_j$, we obtain, by Lemma 13.2.3,

$$\bar{\mathbb{E}}\left[r(U_{j+1})\mathbb{1}\{\kappa>j\}\,\middle|\,\bar{\mathscr{F}}_{\nu_0}\right] \leq \bar{\mathbb{E}}\left[\mathbb{1}\{\kappa>j-1\}\bar{\mathbb{E}}\left[r(U_{j+1})\,\middle|\,\mathscr{B}_j\right]\,\middle|\,\bar{\mathscr{F}}_{\nu_0}\right]$$
$$= \bar{\mathbb{E}}\left[\mathbb{1}\{\kappa>j-1\}H_r(U_j)\,\middle|\,\bar{\mathscr{F}}_{\nu_0}\right] \tag{13.2.21}$$

with $H_r(u) = \mathbb{E}_\alpha[r(q+\tau_\alpha\circ\theta_{u+q})]$. Applying Lemma 13.2.5 (ii) yields

$$H_r(u) \leq (u+q)\mathbb{E}_\alpha[\bar{r}(q+\sigma_\alpha)] \,.$$

Set $w_j = \bar{\mathbb{E}}\left[U_j\mathbb{1}\{\kappa>j-1\}\,\middle|\,\bar{\mathscr{F}}_{\nu_0}\right]$. Combining this inequality with (13.2.14) yields

$$\bar{\mathbb{E}}\left[r(U_{j+1})\mathbb{1}\{\kappa>j\}\,\middle|\,\bar{\mathscr{F}}_{\nu_0}\right]$$
$$\leq \mathbb{E}_\alpha[\bar{r}(q+\sigma_\alpha)]\bar{\mathbb{E}}\left[U_j\mathbb{1}\{\kappa>j-1\}\,\middle|\,\bar{\mathscr{F}}_{\nu_0}\right] + q\mathbb{E}_\alpha[\bar{r}(q+\sigma_\alpha)]\bar{\mathbb{P}}\left(\kappa>j-1\,\middle|\,\bar{\mathscr{F}}_{\nu_0}\right)$$
$$\leq \mathbb{E}_\alpha[\bar{r}(q+\sigma_\alpha)]w_j + q\mathbb{E}_\alpha[\bar{r}(q+\sigma_\alpha)](1-\gamma)^{j-1} \,, \tag{13.2.22}$$

Using again (13.2.21) with $r(u)=u$, we obtain

$$w_{j+1} = \bar{\mathbb{E}}\left[U_{j+1}\mathbb{1}\{\kappa > j\} \,\middle|\, \bar{\mathscr{F}}_{\nu_0}\right] \le \bar{\mathbb{E}}\left[\mathbb{1}\{\kappa > j-1\}H(U_j) \,\middle|\, \bar{\mathscr{F}}_{\nu_0}\right] , \quad (13.2.23)$$

where we now define $H(u) = q + \mathbb{E}_\alpha[\tau_\alpha \circ \theta_{u+q}]$. Lemma 13.2.5 (ii) (applied to the identity sequence $r(n) = n$) implies that $\lim_{u\to\infty} H(u)/u = 0$. Thus, there exists a constant M_1 such that for all $u \in \mathbb{N}$, $H(u) \le (1-\gamma)u + M_1$. Combining this with (13.2.23) yields

$$\begin{aligned} w_{j+1} &\le (1-\gamma)\bar{\mathbb{E}}\left[U_j \mathbb{1}\{\kappa > j-1\} \,\middle|\, \bar{\mathscr{F}}_{\nu_0}\right] + M_1 \bar{\mathbb{P}}\left(\kappa > j-1 \,\middle|\, \bar{\mathscr{F}}_{\nu_0}\right) \\ &\le (1-\gamma)w_j + M_1(1-\gamma)^{j-1} . \end{aligned}$$

Noting that $w_0 = \bar{\mathbb{E}}\left[U_0 \,\middle|\, \bar{\mathscr{F}}_{\nu_0}\right] = U_0$, we obtain $w_j \le (1-\gamma)^j U_0 + jM_1(1-\gamma)^{j-2}$. Plugging this inequality into (13.2.22) completes the proof of (13.2.19).
(ii) Fix now $\varepsilon > 0$. Applying Lemma 13.2.3, we have, for all $j \in \mathbb{N}$ and all $n \ge q$,

$$\begin{aligned} \bar{\mathbb{E}}\left[r(U_{j+1})\mathbb{1}\left\{U_{j+1} \ge n\right\} \,\middle|\, \mathscr{B}_j\right] &\le \sup_u \mathbb{E}_\alpha[r(q+\tau_\alpha\circ\theta_{q+u})\mathbb{1}\left\{q+\tau_\alpha\circ\theta_{q+u} \ge n\right\}] \\ &\le \sum_{k\ge n-q} r(k+q)\mathbb{P}_\alpha(\sigma_\alpha > k) \end{aligned}$$

Since $\sum_{k=0}^\infty r(k+q)\mathbb{P}_\alpha(\sigma_\alpha > k) = \mathbb{E}_\alpha[r^0(\sigma_\alpha + q - 1)] < \infty$ by assumption, we can choose ℓ such that for all $n \ge \ell$, $\sum_{k\ge n-q} r(k+q)\mathbb{P}_\alpha(\sigma_\alpha > k) \le \varepsilon$.

□

13.2.1 Coupling Bounds

Proposition 13.2.7 *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$. Assume that P admits an accessible, aperiodic and positive atom α . If $\xi,\xi' \in \mathbb{M}_1(\mathscr{X})$ satisfy $\mathbb{P}_\xi(\sigma_\alpha < \infty) = \mathbb{P}_{\xi'}(\sigma_\alpha < \infty) = 1$, then $\bar{\mathbb{P}}_{\xi\otimes\xi'}(T < \infty) = 1$.*

Proof. First note that by Lemma 13.2.1, $\bar{\mathbb{P}}_{\xi\otimes\xi'}(\nu_k < \infty) = 1$ for all $k \in \mathbb{N}$. Moreover, by (13.2.14),

$$\bar{\mathbb{P}}_{\xi\otimes\xi'}(\kappa > n) \le (1-\gamma)^n , \quad (13.2.24)$$

whence $\bar{\mathbb{P}}_{\xi\otimes\xi'}(\kappa < \infty) = 1$ and $\bar{\mathbb{P}}_{\xi\otimes\xi'}(T < \infty) = 1$ by (13.2.7). □

We now give a bound for geometric moments of the coupling time T.

Proposition 13.2.8 *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$. Assume that P admits an accessible, aperiodic and positive atom α and that there exists $\beta > 1$ such* *that $\mathbb{E}_\alpha[\beta^{\sigma_\alpha}] < \infty$. Then there exist $\delta \in (1,\beta)$ and $\varsigma < \infty$ such that for all initial distributions ξ and ξ',*

$$\mathbb{E}_{\xi\otimes\xi'}[\delta^T] \le \varsigma\{\mathbb{E}_\xi[\beta^{\sigma_\alpha}] + \mathbb{E}_{\xi'}[\beta^{\sigma_\alpha}]\} .$$

Proof. For every nonnegative increasing sequence r, we have

$$\begin{aligned}\bar{\mathbb{E}}_{\xi\otimes\xi'}[r(T)] &\le \bar{\mathbb{E}}_{\xi\otimes\xi'}[r(\nu_\kappa)] = \sum_{j=0}^{\infty} \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}\{\kappa = j\}\, r(\nu_j)] \\ &\le \bar{\mathbb{E}}_{\xi\otimes\xi'}[r(\nu_0)] + \sum_{j=1}^{\infty} \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}\{\kappa > j-1\}\, r(\nu_j)] . \end{aligned} \tag{13.2.25}$$

Applying (13.2.25) to $r(k) = \delta^k$ and then using the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}\bar{\mathbb{E}}_{\xi\otimes\xi'}[\delta^T] &\le \bar{\mathbb{E}}_{\xi\otimes\xi'}[\delta^{\nu_0}] + \sum_{j=0}^{\infty} \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}\{\kappa > j\}\, \delta^{\nu_{j+1}}] \\ &\le \bar{\mathbb{E}}_{\xi\otimes\xi'}[\beta^{\nu_0}] + \sum_{j=0}^{\infty} \{\bar{\mathbb{P}}_{\xi\otimes\xi'}(\kappa > j)\bar{\mathbb{E}}_{\xi\otimes\xi'}[\delta^{2\nu_{j+1}}]\}^{1/2} . \end{aligned}$$

We now bound each term of the right-hand side. Recall that by (13.2.14), $\bar{\mathbb{P}}_{\xi\otimes\xi'}(\kappa > j) \le (1-\gamma)^j$. Choose $\varepsilon > 0$ such that $(1+\varepsilon)(1-\gamma) < 1$. Combining Lemma 13.2.3 and Lemma 13.2.5 (iii), there exists $\delta \in (1,\beta)$ such that for all $j \in \mathbb{N}$,

$$\bar{\mathbb{E}}\left[\delta^{2U_{j+1}} \,\middle|\, \mathscr{B}_j\right] \le \sup_{u\in\mathbb{N}} \mathbb{E}_\alpha\left[\delta^{(2q+2\tau_\alpha\circ\bar{\theta}_{u+q})}\right] \le 1+\varepsilon ,$$

Then, for all $j \in \mathbb{N}$,

$$\bar{\mathbb{E}}_{\xi\otimes\xi'}[\delta^{2\nu_{j+1}}] = \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\delta^{2\nu_j}\bar{\mathbb{E}}\left[\delta^{2U_{j+1}} \,\middle|\, \mathscr{B}_j\right]\right] \le (1+\varepsilon)\bar{\mathbb{E}}_{\xi\otimes\xi'}[\delta^{2\nu_j}] ,$$

and by induction,

$$\bar{\mathbb{E}}_{\xi\otimes\xi'}[\delta^{2\nu_j}] \le (1+\varepsilon)^j \bar{\mathbb{E}}_{\xi\otimes\xi'}[\delta^{\nu_0}] \le (1+\varepsilon)^j \bar{\mathbb{E}}_{\xi\otimes\xi'}[\beta^{\nu_0}] .$$

Finally,

$$\bar{\mathbb{E}}_{\xi\otimes\xi'}[\delta^T] \le \bar{\mathbb{E}}_{\xi\otimes\xi'}[\beta^{\nu_0}] + (1+\varepsilon)^{1/2}\bar{\mathbb{E}}^{1/2}_{\xi\otimes\xi'}[\beta^{\nu_0}] \sum_{j=0}^{\infty} \{(1-\gamma)(1+\varepsilon)\}^{j/2} .$$

The series is convergent because of the choice of ε. The proof is completed by noting that

$$1 \leq \bar{\mathbb{E}}_{\xi\otimes\xi'}[\beta^{\nu_0}] = \bar{\mathbb{E}}_{\xi\otimes\xi'}[\beta^{\sigma_{\alpha\times\mathsf{X}}} \vee \beta^{\sigma_{\mathsf{X}\times\alpha}}] \leq \mathbb{E}_{\xi}[\beta^{\sigma_\alpha}] + \mathbb{E}_{\xi'}[\beta^{\sigma_\alpha}] .$$

□

We now turn to the case of subgeometric moments.

Proposition 13.2.9 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P admits an accessible, aperiodic and positive atom α. Let $r \in \bar{\Lambda}_1$ be such that $\mathbb{E}_\alpha[r^0(\sigma_\alpha)] < \infty$. Then, there exists a constant $\varsigma < \infty$ such that for all initial distributions ξ and ξ',*

$$\bar{\mathbb{E}}_{\xi\otimes\xi'}[r(T)] \leq \varsigma\{\mathbb{E}_\xi[r(\sigma_\alpha)] + \mathbb{E}_{\xi'}[r(\sigma_\alpha)]\} .$$

Proof. Without loss of generality, we assume that $r \in \Lambda_1$. Set

$$w_j = \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}\{\kappa > j-1\} r(\nu_j)] .$$

Applying (13.2.25), we obtain

$$\bar{\mathbb{E}}_{\xi\otimes\xi'}[r(T)] \leq \sum_{j=0}^{\infty} w_j . \tag{13.2.26}$$

Set $\varepsilon > 0$ such that $\tilde{\varepsilon} := (1+\varepsilon)(1-\gamma)+\varepsilon < 1$. By Lemma 13.2.6, there exists a constant n_0 such that for all $n \geq n_0$ and all $j \in \mathbb{N}$,

$$\bar{\mathbb{E}}\left[r(U_{j+1})\mathbb{1}\left\{U_{j+1} \geq n\right\} \middle| \mathscr{B}_j\right] \leq \varepsilon .$$

Such an integer n_0 being chosen and there exists a constant ς such that $r(m+n) \leq (1+\varepsilon)r(m)+\varsigma+\mathbb{1}\{n \geq n_0\} r(m)r(n)$ (see Lemma 13.1.3). Plugging this inequality into $w_{j+1} = \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}\{\kappa > j\} r(\nu_j + U_j)]$, we get

$$\begin{aligned} w_{j+1} \leq &(1+\varepsilon)\bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}\{\kappa > j\} r(\nu_j)] \\ &+ \varsigma \mathbb{P}_{\xi\otimes\xi'}(\kappa > j) + \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbb{1}\{\kappa > j\}\mathbb{1}\left\{U_{j+1} > n_0\right\} r(\nu_j) r(U_{j+1})\right] \end{aligned}$$

We now bound each term of the right-hand side. By Lemma 13.2.4,

$$\begin{aligned} \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}\{\kappa > j\} r(\nu_j)] &= \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\bar{\mathbb{E}}\left[\mathbb{1}\{\kappa > j\} r(\nu_j) \middle| \bar{\mathscr{F}}_{\nu_{j-1}}\right]\right] \\ &\leq (1-\gamma)\bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbb{1}\{\kappa > j-1\}\bar{\mathbb{E}}\left[r(\nu_j) \middle| \bar{\mathscr{F}}_{\nu_{j-1}}\right]\right] \\ &= (1-\gamma) w_j . \end{aligned}$$

Applying (13.2.20) yields

$$
\begin{aligned}
&\bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbb{1}\{\kappa > j\}\mathbb{1}\left\{U_{j+1} > n_0\right\} r(\nu_j) r(U_{j+1})\right] \\
&\qquad = \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbb{1}\{\kappa > j-1\} r(\nu_j) \bar{\mathbb{E}}\left[r(U_{j+1})\mathbb{1}\left\{U_{j+1} > n_0\right\} \,\middle|\, \mathscr{B}_j\right]\right] \le \varepsilon w_j \,.
\end{aligned}
$$

Finally, applying (13.2.24), we obtain

$$
w_{j+1} \le (1+\varepsilon)(1-\gamma) w_j + \varsigma(1-\gamma)^j + \varepsilon w_j \le \tilde{\varepsilon} w_j + \varsigma \tilde{\varepsilon}^j \,.
$$

This implies that $w_j \le \tilde{\varepsilon}^j w_0 + \varsigma j \tilde{\varepsilon}^{j-1}$. The proof is completed by plugging this inequality into (13.2.26) together with

$$
w_0 = \bar{\mathbb{E}}_{\xi\otimes\xi'}[r(\nu_0)] \le \bar{\mathbb{E}}_{\xi\otimes\xi'}[r(\sigma_{\alpha\times\mathsf{X}}) \vee r(\sigma_{\mathsf{X}\times\alpha})] \le \mathbb{E}_\xi[r(\sigma_\alpha)] + \mathbb{E}_{\xi'}[r(\sigma_\alpha)] \,.
$$

□

Proposition 13.2.10 *Let $r \in \bar{\Lambda}_1$ such that $\mathbb{E}_\alpha[r^0(\sigma_\alpha)] < \infty$. Then, there exists a constant $\varsigma < \infty$ such that for all initial distributions ξ and ξ',*

$$
\bar{\mathbb{E}}_{\xi\otimes\xi'}[r^0(T)] \le \varsigma\{\mathbb{E}_\xi[r^0(\sigma_\alpha)] + \mathbb{E}_{\xi'}[r^0(\sigma_\alpha)]\} \,.
$$

Proof. Without loss of generality, we assume that $r \in \Lambda_1$. Applying (13.2.25) with r replaced by r^0 and defining now $w_j = \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}\{\kappa > j-1\} r^0(\nu_j)]$, we get

$$
\bar{\mathbb{E}}_{\xi\otimes\xi'}[r^0(T)] \le \sum_{j=0}^{\infty} w_j \,. \tag{13.2.27}
$$

Set $\tilde{\varepsilon} = (1+\varepsilon)(1-\gamma)+\varepsilon$ and choose $\varepsilon > 0$ such that $\tilde{\varepsilon} < 1$. We now prove that the right-hand side of (13.2.27) is a convergent series. According to Lemma 13.2.6, there exists m such that $\bar{\mathbb{E}}\left[r(U_{j+1})\mathbb{1}\left\{U_{j+1} > m\right\} \,\middle|\, \mathscr{B}_j\right] \le \varepsilon$ for all $j \in \mathbb{N}$. This m being chosen, write $w_{j+1} = w_{j+1}^{(0)} + w_{j+1}^{(1)}$ where

$$
\begin{aligned}
w_{j+1}^{(0)} &:= \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}\left\{\kappa > j, U_{j+1} > m\right\} r^0(U_{j+1}+\nu_j)] \,, \\
w_{j+1}^{(1)} &:= \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}\left\{\kappa > j, U_{j+1} \le m\right\} r^0(\nu_j+U_{j+1})] \,.
\end{aligned}
$$

Since $r \in \Lambda_1 \subset \mathscr{S}$ (see Definition 13.1.1), by Lemma 13.1.4, $r^0 \in \bar{\mathscr{S}}$. This allows to apply Lemma 13.2.6 (i) with r replaced by $r^0 \in \bar{\Lambda}_1$: for all $\rho \in (1-\gamma, 1)$, there exists a finite constant ς_0 such that for all $j \in \mathbb{N}$,

$$
\bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}\{\kappa > j\} r^0(U_{j+1})] \le \varsigma_0 \rho^j \bar{\mathbb{E}}_{\xi\otimes\xi'}[U_0] \,. \tag{13.2.28}
$$

Then, using (13.2.28) and $r^0 \in \bar{\mathscr{S}}$, we have

$$
\begin{aligned}
&w_{j+1}^{(0)} \\
&\leq \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}\{\kappa > j\} r^0(U_{j+1})] + \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}\{\kappa > j\} r^0(\nu_j) r(U_{j+1}) \mathbb{1}\{U_{j+1} > m\}] \\
&\leq \varsigma_0 \rho^j \bar{\mathbb{E}}_{\xi\otimes\xi'}[U_0] + \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\mathbb{1}\{\kappa > j-1\} r^0(\nu_j) \bar{\mathbb{E}}\left[r(U_{j+1}) \mathbb{1}\{U_{j+1} > m\} \,\middle|\, \mathscr{B}_j \right]\right] \\
&\leq \varsigma_0 \rho^j \bar{\mathbb{E}}_{\xi\otimes\xi'}[U_0] + \varepsilon w_j \; .
\end{aligned}
$$

Moreover, since $\lim_{k\to\infty} r(k)/r^0(k) = 0$, there exists a finite constant ς_1 such that for all $k \in \mathbb{N}$, $r(k) r^0(m) \leq \varepsilon r^0(k) + \varsigma_1$. Then, using again (13.2.28), we obtain

$$
\begin{aligned}
w_{j+1}^{(1)} &\leq \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}\{\kappa > j, U_{j+1} \leq m\} r^0(\nu_j + m)] \\
&\leq \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}\{\kappa > j\} \{r^0(\nu_j) + r(\nu_j) r^0(m)\}] \\
&\leq (1-\gamma) \bar{\mathbb{E}}_{\xi\otimes\xi'}[\mathbb{1}\{\kappa > j-1\} \{(1+\varepsilon) r^0(\nu_j) + \varsigma_1\}] \\
&\leq (1-\gamma)(1+\varepsilon) w_j + \varsigma_1 (1-\gamma)^j \; .
\end{aligned}
$$

Finally, there exists a finite constant M such that for all $j \in \mathbb{N}$,

$$
w_{j+1} \leq \{(1-\gamma)(1+\varepsilon) + \varepsilon\} w_j + \{\varsigma_0 \bar{\mathbb{E}}_{\xi\otimes\xi'}[U_0] + \varsigma_1\} (1-\gamma)^j \leq \tilde{\varepsilon} w_j + M \tilde{\varepsilon}^j \; .
$$

This implies that $w_j \leq \tilde{\varepsilon}^j w_0 + M j \tilde{\varepsilon}^{j-1}$. The proof is completed by plugging this inequality into (13.2.27) and noting that

$$
\begin{aligned}
w_0 = \bar{\mathbb{E}}_{\xi\otimes\xi'}[r^0(\nu_0)] &\leq \bar{\mathbb{E}}_{\xi\otimes\xi'}[r^0(\sigma_{\alpha\times\mathsf{X}}) \vee r^0(\sigma_{\mathsf{X}\times\alpha})] \\
&\leq \mathbb{E}_{\xi}[r^0(\sigma_\alpha)] + \mathbb{E}_{\xi'}[r^0(\sigma_\alpha)] \; .
\end{aligned}
$$

□

We conclude this section with a bound on a polynomial moment $\bar{\mathbb{E}}_{\xi\otimes\xi'}[T^s]$ of the coupling time. When $s > 1$ this is simply a particular case of 13.2.10. However, when $s \in (0,1)$, the function $r(n) = (n+1)^{s-1}$ is decreasing and thus does not belong to $\bar{\Lambda}_1$ so that Proposition 13.2.10 does not apply.

Proposition 13.2.11 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P admits an accessible, aperiodic and positive atom α. Let $s > 0$ and assume that $\mathbb{E}_\alpha[\sigma_\alpha^s] < \infty$. Then, there exists $\varsigma > 0$ such that for all initial distributions ξ and ξ',*

$$
\bar{\mathbb{E}}_{\xi\otimes\xi'}[T^s] \leq \varsigma \{\mathbb{E}_{\xi}[\sigma_\alpha^s] + \mathbb{E}_{\xi'}[\sigma_\alpha^s]\} \; .
$$

Proof. If $s \geq 1$, we apply Proposition 13.2.10 to $r(n) = (n+1)^{s-1}$. If $s \in (0,1)$, we can apply the bound $(a_0 + \ldots + a_n)^s \leq a_0^s + \ldots + a_n^s$. Using the convention $\sum_{j=1}^{0} a_i = 0$, we obtain

$$
\begin{aligned}
\bar{\mathbb{E}}_{\xi\otimes\xi'}[T^s] \le \bar{\mathbb{E}}_{\xi\otimes\xi'}[\nu_\kappa^s] &\le \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\left(\nu_0^s + \sum_{j=1}^{\kappa} U_j^s\right)\right] \\
&\le \bar{\mathbb{E}}_{\xi\otimes\xi'}[\nu_0^s] + \sum_{j=1}^{\infty} \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[U_j^s \mathbb{1}\{\kappa > j-1\}\right] . \qquad (13.2.29)
\end{aligned}
$$

By Lemma 13.2.6 (i) applied with $r(k) = k$, for every $\rho \in (1-\gamma, 1)$, there exists a constant M such that for all $j \ge 1$,

$$
\bar{\mathbb{E}}\left[\mathbb{1}\{\kappa > j\} U_{j+1} \,\middle|\, \mathscr{F}_{\nu_0}\right] \le M\rho^j U_0
$$

This implies, using the concavity of the function $u \to u^s$,

$$
\begin{aligned}
\bar{\mathbb{E}}_{\xi\otimes\xi'}\left[U_{j+1}^s \mathbb{1}\{\kappa > j\}\right] &= \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\bar{\mathbb{E}}\left[U_{j+1}^s \mathbb{1}\{\kappa > j\} \,\middle|\, \bar{\mathscr{F}}_0\right]\right] \\
&\le \bar{\mathbb{E}}_{\xi\otimes\xi'}\left[\bar{\mathbb{E}}\left[U_{j+1} \mathbb{1}\{\kappa > j\} \,\middle|\, \bar{\mathscr{F}}_0\right]^s\right] \le \varsigma^s \rho^{sj} \bar{\mathbb{E}}_{\xi\otimes\xi'}[U_0^s] .
\end{aligned}
$$

Plugging this into (13.2.29) and noting that $1 \le w_0 = \bar{\mathbb{E}}_{\xi\otimes\xi'}[\nu_0^s] \le \mathbb{E}_\xi[\sigma_\alpha^s] + \mathbb{E}_{\xi'}[\sigma_\alpha^s]$ complete the proof. □

13.3 Rates of Convergence in Total Variation Distance

In this section, we show how the coupling inequalities (Lemma 8.3.1) combined with Propositions 13.2.8–13.2.10 yield rates of convergence in the total variation distance of $\delta_x P^n$ to the invariant probability measure π (whose existence is ensured by the existence of a positive atom α).

Theorem 13.3.1. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and α an accessible aperiodic and positive atom. Denote by π the unique invariant probability. Assume that there exists $\beta > 1$ such that $\mathbb{E}_\alpha[\beta^{\sigma_\alpha}] < \infty$. Then $\mathbb{E}_\pi[\beta^{\sigma_\alpha}] < \infty$, and there exist $\delta \in (1, \beta)$ and $\varsigma < \infty$ such that for every initial distribution ξ,*

$$
\sum_{n=0}^{\infty} \delta^n \mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) \le \varsigma \mathbb{E}_\xi[\beta^{\sigma_\alpha}] .
$$

Remark 13.3.2. *Since $\mathbb{E}_\pi[\beta^{\sigma_\alpha}] < \infty$, the series $\sum_{n=0}^{\infty} \delta^n \mathrm{d}_{\mathrm{TV}}(P^n(x, \cdot), \pi)$ is summable for π-almost all $x \in \mathsf{X}$.* ▲

Proof. By Corollary 6.4.4, $\mathbb{E}_\pi[\beta^{\sigma_\alpha}] < \infty$. Applying the bound (8.3.4) and Proposition 13.2.8 with $\mu = \pi$, we obtain that there exist $\delta \in (1, \beta)$ and $\varsigma < \infty$ such that for all $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{n=0}^{\infty} \delta^n \mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) \le \mathbb{E}_{\xi\otimes\pi}\left[\sum_{n=0}^{T} \delta^n\right]$$
$$\le \delta(\delta-1)^{-1}\mathbb{E}_{\xi\otimes\pi}[\delta^T] \le \varsigma\left\{\mathbb{E}_{\xi}[\beta^{\sigma_\alpha}] + \mathbb{E}_{\pi}[\beta^{\sigma_\alpha}]\right\} .$$

□

To state the results in the subgeometric case, we introduce the following convention. The first difference Δr of a sequence r is defined by

$$\Delta r(n) = r(n) - r(n-1) \;,\;\; n \ge 1 \;,\;\; \Delta r(0) = r(0) \;.$$

Note that with this convention, we have, for all $n \ge 0$,

$$r(n) = \sum_{k=0}^{n} \Delta r(k) = (\Delta r)^0(n) \;.$$

Theorem 13.3.3. *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$ and α an accessible, aperiodic, and positive atom α. Denote by π the unique invariant probability measure. Assume that there exists $r \in \bar{\Lambda}_1$ such that $\mathbb{E}_\alpha[r^0(\sigma_\alpha)] < \infty$.*

(i) There exists a constant $\varsigma < \infty$ such that for all initial distributions ξ and μ,

$$\sum_{n=0}^{\infty} r(n)\mathrm{d}_{\mathrm{TV}}(\xi P^n, \mu P^n) \le \varsigma\left(\mathbb{E}_\xi[r^0(\sigma_\alpha)] + \mathbb{E}_\mu[r^0(\sigma_\alpha)]\right) . \tag{13.3.1}$$

(ii) There exists a constant $\varsigma < \infty$ such that for every initial distribution ξ and all $n \in \mathbb{N}$,

$$r(n)\mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) \le \varsigma\mathbb{E}_\xi[r(\sigma_\alpha)] \,. \tag{13.3.2}$$

(iii) If either $\lim_{n\to\infty} \uparrow r(n) = \infty$ and $\mathbb{E}_\xi[r(\sigma_\alpha)] < \infty$ or $\lim_{n\to\infty} r(n) < \infty$ and $\mathbb{P}_\xi(\sigma_\alpha < \infty) = 1$, then

$$\lim_{n\to\infty} r(n)\mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) = 0 \,. \tag{13.3.3}$$

(iv) If, in addition, $\Delta r \in \bar{\Lambda}_1$, then there exists a constant $\varsigma < \infty$ such that for every initial distribution ξ,

$$\sum_{n=1}^{\infty} \Delta r(n)\mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) \le \varsigma\mathbb{E}_\xi[r(\sigma_\alpha)] \,.$$

Proof. Without loss of generality, we assume that $r \in \Lambda_1$.

(i) The bound (13.3.1) is obtained by applying (8.3.4) and Proposition 13.2.10.

(ii) By Lemma 6.4.3, the condition $\mathbb{E}_\alpha[r^0(\sigma_\alpha)]$ implies that $\mathbb{E}_\pi[r(\sigma_\alpha)] < \infty$; hence $\mathbb{P}_\pi(\sigma_\alpha < \infty) = 1$. Proposition 13.2.9 shows that there exists $\varsigma < \infty$ such that $\bar{\mathbb{E}}_{\xi\otimes\pi}[r(T)] \le \varsigma\{\mathbb{E}_\xi[r(\sigma_\alpha)] + \mathbb{E}_\pi[r(\sigma_\alpha)]\}$. By Lemma 8.3.1, we get

$$r(n)\mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) \le \bar{\mathbb{P}}_{\xi\otimes\pi}(T \ge n) \le \bar{\mathbb{E}}_{\xi\otimes\pi}[r(T)] .$$

(iii) This is a refinement of (ii). The case $\limsup r(n) < \infty$ and $\mathbb{P}_\xi(\sigma_\alpha < \infty) = 1$ is dealt with in Theorem 8.2.6. Note that $\lim_{n\to\infty} \uparrow r(n) = \infty$; the condition $\mathbb{E}_\xi[r(\sigma_\alpha)] < \infty$ implies that $\mathbb{P}_\xi(\sigma_\alpha < \infty) = 1$. Since $\mathbb{E}_\pi[r(\sigma_\alpha)] < \infty$, we also have $\mathbb{P}_\pi(\sigma_\alpha < \infty) = 1$. Proposition 13.2.7 shows that $\bar{\mathbb{P}}_{\xi\otimes\pi}(T < \infty) = 1$.

On the other hand, by Proposition 13.2.9, we have $\bar{\mathbb{E}}_{\xi\otimes\pi}[r(T)] \le \varsigma\{\mathbb{E}_\xi[r(\sigma_\alpha)] + \mathbb{E}_\pi[r(\sigma_\alpha)]\}$, and thus $\mathbb{E}_\xi[r(\sigma_\alpha)] < \infty$ and $\mathbb{E}_\pi[r(\sigma_\alpha)] < \infty$ imply $\bar{\mathbb{E}}_{\xi\otimes\pi}[r(T)] < \infty$. Since the sequence $\{r(n),\ n \in \mathbb{N}\}$ is nondecreasing, it follows that

$$r(n)\bar{\mathbb{P}}_{\xi\otimes\pi}(T \ge n) \le \bar{\mathbb{E}}_{\xi\otimes\pi}[r(T)\mathbb{1}_{\{T\ge n\}}] .$$

Since $\bar{\mathbb{P}}_{\xi\otimes\pi}(T < \infty) = 1$ and $\bar{\mathbb{E}}_{\xi\otimes\pi}[r(T)] < \infty$, Lebesgue's dominated convergence theorem shows that $\lim_{n\to\infty} r(n)\bar{\mathbb{P}}_{\xi\otimes\pi}(T \ge n) = 0$. The proof is concluded by Lemma 8.3.1, which shows that for all $n \in \mathbb{N}$, $\mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) \le \bar{\mathbb{P}}_{\xi\otimes\pi}(T \ge n)$.

(iv) We assume without loss of generality that $\Delta r \in \Lambda_1$. Applying (8.3.4), we get that for all $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{n=1}^{\infty} \Delta r(n)\mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) \le \bar{\mathbb{E}}_{\xi\otimes\pi}[r(T)] . \tag{13.3.4}$$

Applying now Proposition 13.2.9 to the sequence Δr, there exists $\varsigma_1 < \infty$ such that $\bar{\mathbb{E}}_{\xi\otimes\pi}[r(T)] \le \varsigma\left\{\mathbb{E}_\xi[r(\sigma_\alpha)] + \mathbb{E}_\pi[r(\sigma_\alpha)]\right\}$. The proof is concluded on noting that by Lemma 6.4.3, $\mathbb{E}_\pi[r(\sigma_\alpha)] < \infty$.

□

13.4 Rates of Convergence in f-Norm

Let $f : \mathsf{X} \to [1, \infty)$ be a measurable function fixed once and for all throughout this section. Define the f-norm of a measure $\xi' \in \mathbb{M}_\pm(\mathscr{X})$ as follows:

$$\|\xi'\|_f = \sup_{\substack{g\in\mathbb{F}(\mathsf{X})\\ |g|\le f}} \xi'(g) . \tag{13.4.1}$$

Properties of the f-norm are given in Appendix D.3. The next result, which fundamentally relies on the fact that α is an atom, provides a very simple link between the rate of convergence in total variation norm and in f-norm.

Proposition 13.4.1 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P admits an accessible, aperiodic, and positive atom α.*

(i) For all $n \in \mathbb{N}^*$,

$$\left\|\xi P^n - \xi' P^n\right\|_f \le \mathbb{E}_\xi\left[f(X_n)\mathbb{1}_{\{\sigma_\alpha \ge n\}}\right] + \mathbb{E}_{\xi'}\left[f(X_n)\mathbb{1}_{\{\sigma_\alpha \ge n\}}\right] + \sum_{j=1}^{n-1} |\xi P^j(\alpha) - \xi' P^j(\alpha)| \mathbb{E}_\alpha\left[f(X_{n-j})\mathbb{1}_{\{\sigma_\alpha \ge n-j\}}\right]. \quad (13.4.2)$$

(ii) For every sequence $r \in \mathscr{S}$ and initial distributions ξ and ξ',

$$\sum_{n=1}^{\infty} r(n)\left\|\xi P^n - \xi' P^n\right\|_f \le \mathbb{E}_\xi\left[\sum_{j=1}^{\sigma_\alpha} r(j) f(X_j)\right] + \mathbb{E}_{\xi'}\left[\sum_{j=1}^{\sigma_\alpha} r(j) f(X_j)\right] + \mathbb{E}_\alpha\left[\sum_{j=1}^{\sigma_\alpha} r(j) f(X_j)\right] \sum_{n=1}^{\infty} r(n)|\xi P^n(\alpha) - \xi' P^n(\alpha)|. \quad (13.4.3)$$

Remark 13.4.2. By definition of the total variation distance, in (13.4.2) and (13.4.3), the terms $|\xi P^n(\alpha) - \xi' P^n(\alpha)|$ can be further bounded by $\mathrm{d_{TV}}(\xi P^n, \xi' P^n)$. ▲

Proof (*of Proposition* 13.4.1).

(i) Let $g \in \mathbb{F}_b(\mathsf{X})$. Then

$$\begin{aligned}
\mathbb{E}_\xi[g(X_n)] &= \mathbb{E}_\xi[g(X_n)\mathbb{1}\{\sigma_\alpha \ge n\}] + \sum_{j=1}^{n-1} \mathbb{E}_\xi[\mathbb{1}_\alpha(X_j)\mathbb{1}_{\alpha^c}(X_{j+1})\cdots\mathbb{1}_{\alpha^c}(X_n)g(X_n)] \\
&= \mathbb{E}_\xi[g(X_n)\mathbb{1}\{\sigma_\alpha \ge n\}] + \sum_{j=1}^{n-1} \mathbb{E}_\xi\left[\mathbb{1}_\alpha(X_j)\mathbb{E}_{X_j}[g(X_{n-j})\mathbb{1}\{\sigma_\alpha \ge n-j\}]\right] \\
&= \mathbb{E}_\xi[g(X_n)\mathbb{1}\{\sigma_\alpha \ge n\}] + \sum_{j=1}^{n-1} \mathbb{P}_\xi(X_j \in \alpha)\mathbb{E}_\alpha[g(X_{n-j})\mathbb{1}\{\sigma_\alpha \ge n-j\}] \\
&= \mathbb{E}_\xi[g(X_n)\mathbb{1}\{\sigma_\alpha \ge n\}] + \sum_{j=1}^{n-1} \xi P^j(\alpha)\mathbb{E}_\alpha[g(X_{n-j})\mathbb{1}\{\sigma_\alpha \ge n-j\}].
\end{aligned}$$

In the previous computations, we used the last-exit decomposition, and the fact that α is an atom was crucial. This yields, for all $g \in \mathbb{F}_b(\mathsf{X})$ such that $|g| \le f$,

$$|\xi P^n g - \xi' P^n g| \le \mathbb{E}_\xi[f(X_n)\mathbb{1}\{\sigma_\alpha \ge n\}] + \mathbb{E}_{\xi'}[f(X_n)\mathbb{1}\{\sigma_\alpha \ge n\}] + \sum_{j=1}^{n-1} |\xi P^j(\alpha) - \xi' P^j(\alpha)|\mathbb{E}_\alpha[f(X_{n-j})\mathbb{1}\{\sigma_\alpha \ge n-j\}].$$

Taking the supremum over all $g \in \mathbb{F}_b(\mathsf{X})$ such that $|g| \le f$ yields (13.4.2), by Theorem D.3.2.

(ii) For every initial distribution ξ, we have

$$\sum_{n=1}^{\infty} r(n)\mathbb{E}_\xi[f(X_n)\mathbb{1}\{\sigma_\alpha \ge n\}] = \mathbb{E}_\xi\left[\sum_{j=1}^{\sigma_\alpha} r(j)f(X_j)\right].$$

Thus, multiplying both sides of (13.4.2) by $r(n)$ and summing over n, we obtain

$$\begin{aligned}\sum_{n=1}^{\infty} r(n)\left\|\xi P^n - \xi' P^n\right\|_f &\le \mathbb{E}_\xi\left[\sum_{j=1}^{\sigma_\alpha} r(j)f(X_j)\right] + \mathbb{E}_{\xi'}\left[\sum_{j=1}^{\sigma_\alpha} r(j)f(X_j)\right] \\ &+ \sum_{n=1}^{\infty} r(n)\sum_{j=1}^{n-1}|\xi P^j(\alpha) - \xi' P^j(\alpha)|\mathbb{E}_\alpha[f(X_{n-j})\mathbb{1}\{\sigma_\alpha \ge n-j\}].\end{aligned}$$

Since $r \in \mathscr{S}$ (see Definition 13.1.1), we can write

$$\begin{aligned}&\sum_{n=1}^{\infty} r(n)\sum_{j=1}^{n-1}|\xi P^j(\alpha) - \xi' P^j(\alpha)|\mathbb{E}_\alpha[f(X_{n-j})\mathbb{1}\{\sigma_\alpha \ge n-j\}] \\ &\quad = \sum_{j=1}^{\infty}|\xi P^j(\alpha) - \xi' P^j(\alpha)|\sum_{n=j+1}^{\infty} r(n)\mathbb{E}_\alpha[f(X_{n-j})\mathbb{1}\{\sigma_\alpha \ge n-j\}] \\ &\quad = \sum_{j=1}^{\infty}|\xi P^j(\alpha) - \xi' P^j(\alpha)|\sum_{n=1}^{\infty} r(n+j)\mathbb{E}_\alpha[f(X_n)\mathbb{1}\{\sigma_\alpha \ge n\}] \\ &\quad \le \sum_{j=1}^{\infty} r(j)|\xi P^j(\alpha) - \xi' P^j(\alpha)|\sum_{n=1}^{\infty} r(n)\mathbb{E}_\alpha[f(X_n)\mathbb{1}\{\sigma_\alpha \ge n\}] \\ &\quad = \mathbb{E}_\alpha\left[\sum_{j=1}^{\sigma_\alpha} r(j)f(X_j)\right]\sum_{j=1}^{\infty} r(j)|\xi P^j(\alpha) - \xi' P^j(\alpha)|.\end{aligned}$$

This proves (13.4.3).

□

Combining Theorems 13.3.1 and 13.3.3 and Proposition 13.4.1, we obtain rates of convergence in f-norm for atomic chains.

Theorem 13.4.3. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and α an accessible, aperiodic, and positive atom. Denote by π the unique invariant probability. Assume that there exists $\delta > 1$ such that*

$$\mathbb{E}_\alpha\left[\sum_{n=1}^{\sigma_\alpha} \delta^n f(X_n)\right] < \infty.$$

Then there exist $\beta \in (1,\delta)$ and a constant ς such that for every $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{n=1}^{\infty} \beta^n \|\xi P^n - \pi\|_f \le \varsigma \mathbb{E}_\xi \left[\sum_{n=1}^{\sigma_\alpha} \delta^n f(X_n) \right] . \tag{13.4.4}$$

Proof. By Lemma 6.4.3, we have $\mathbb{E}_\pi[\sum_{n=1}^{\sigma_\alpha} \delta^n f(X_n)] < \infty$. Thus the bound (13.4.4) is a consequence of Theorem 13.3.1 and Proposition 13.4.1. □

Theorem 13.4.4. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and α an accessible, aperiodic, and positive atom. Denote by π the unique invariant probability. Assume that there exists $r \in \bar{\Lambda}_1$ such that*

$$\mathbb{E}_\alpha \left[\sum_{k=1}^{\sigma_\alpha} r(k) f(X_k) \right] < \infty . \tag{13.4.5}$$

(i) There exists $\varsigma < \infty$ such that for every initial distribution $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{n=0}^{\infty} r(n) \left\|\xi P^n - \xi' P^n\right\|_f \le \varsigma \left\{ \mathbb{E}_\xi \left[\sum_{k=1}^{\sigma_\alpha} r(k) f(X_k) \right] + \mathbb{E}_{\xi'} \left[\sum_{k=1}^{\sigma_\alpha} r(k) f(X_k) \right] \right\} . \tag{13.4.6}$$

(ii) There exists $\varsigma < \infty$ such that for every initial distribution $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$r(n) \|\xi P^n - \pi\|_f \le \varsigma \mathbb{E}_\xi \left[\sum_{k=1}^{\sigma_\alpha} r(k) f(X_k) \right] . \tag{13.4.7}$$

(iii) If $\mathbb{E}_\xi[\sum_{k=1}^{\sigma_\alpha} r(k) f(X_k)] < \infty$, then

$$\lim_{n\to\infty} r(n) \|\xi P^n - \pi\|_f = 0 . \tag{13.4.8}$$

(iv) If $\Delta r \in \bar{\Lambda}_1$, then there exists a finite constant C such that for every $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{n=0}^{\infty} \Delta r(n) \|\xi P^n - \pi\|_f \le C \mathbb{E}_\xi \left[\sum_{k=1}^{\sigma_\alpha} \Delta r(k) f(X_k) \right] . \tag{13.4.9}$$

Proof. Without loss of generality, we assume that $r \in \Lambda_1$. Since $f \ge 1$, the assumption (13.4.5) implies that $\mathbb{E}_\alpha[r^0(\sigma_\alpha)] < \infty$, where $r^0(n) = \sum_{k=0}^{n} r(k)$.

(i) Combining Theorem 13.3.3 (i) and Proposition 13.4.1 (ii) yields (13.4.6).

(ii) By Proposition 13.4.1 (i),

$$
\begin{aligned}
r(n)\left\|\xi P^n-\pi\right\|_f \leq\ & r(n)\mathbb{E}_\xi[\mathbb{1}\{\sigma_\alpha \geq n\} f(X_n)]+r(n)\mathbb{E}_\pi[\mathbb{1}\{\sigma_\alpha \geq n\} f(X_n)] \\
&+r(n)\sum_{j=1}^{n-1}\left\|\xi P^j-\pi\right\|_{\mathrm{TV}}\mathbb{E}_\alpha[\mathbb{1}\{\sigma_\alpha \geq n-j\} f(X_{n-j})]\,.
\end{aligned}
\tag{13.4.10}
$$

The first term of the right-hand side of (13.4.10) is bounded by $\mathbb{E}_\xi\left[\sum_{k=1}^{\sigma_\alpha} r(k)f(X_k)\right]$. Consider now the second term on the right-hand side of (13.4.10). Applying the same computations as in the proof of Lemma 6.4.3, we have

$$
\begin{aligned}
r(n)\mathbb{E}_\pi[\mathbb{1}\{\sigma_\alpha \geq n\} f(X_n)] &= r(n)\pi(\alpha)\mathbb{E}_\alpha\left[\sum_{k=0}^{\sigma_\alpha-1}\mathbb{E}_{X_k}[f(X_n)\mathbb{1}\{\sigma_\alpha \geq n\}]\right] \\
&= r(n)\pi(\alpha)\sum_{k=0}^{\infty}\mathbb{E}_\alpha\left[\mathbb{1}\{\sigma_\alpha > k\}f(X_{n+k})\mathbb{1}\{\sigma_\alpha \geq n+k\}\right] \\
&= r(n)\pi(\alpha)\mathbb{E}_\alpha\left[\sum_{k=n}^{\sigma_\alpha} f(X_k)\right] \leq \pi(\alpha)\mathbb{E}_\alpha\left[\sum_{k=n}^{\sigma_\alpha} r(k)f(X_k)\right] \\
&\leq \pi(\alpha)\mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} r(k)f(X_k)\right]\,.
\end{aligned}
$$

The last term is finite by assumption (13.4.5). Consider now the last term on the right-hand side of (13.4.10). Using $r(n) \leq r(j)r(n-j)$ for $1 \leq j \leq n-1$ and applying Theorem 13.3.3 (ii), we obtain

$$
\begin{aligned}
& r(n)\sum_{j=1}^{n-1}\left\|\xi P^j-\pi\right\|_{\mathrm{TV}}\mathbb{E}_\alpha[\mathbb{1}\{\sigma_\alpha \geq n-j\} f(X_{n-j})] \\
&\leq \left\{\sup_{j\in\mathbb{N}^*} r(j)\left\|\xi P^j-\pi\right\|_{\mathrm{TV}}\right\}\sum_{j=1}^{n-1} r(n-j)\mathbb{E}_\alpha[\mathbb{1}\{\sigma_\alpha \geq n-j\} f(X_{n-j})] \\
&\leq \varsigma\mathbb{E}_\xi\left[\sum_{k=1}^{\sigma_\alpha} r(k)f(X_k)\right]\sum_{k=1}^{\infty} r(k)\mathbb{E}_\alpha[\mathbb{1}\{\sigma_\alpha \geq k\} f(X_k)] \\
&= \varsigma\mathbb{E}_\xi\left[\sum_{k=1}^{\sigma_\alpha} r(k)f(X_k)\right]\mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} r(k)f(X_k)\right]\,.
\end{aligned}
$$

Since the last expectation is finite by assumption (13.4.5), the proof of (13.4.7) is complete.

(iii) We now assume $\mathbb{E}_\xi[\sum_{k=1}^{\sigma_\alpha} r(k)f(X_k)] < \infty$ and turn to the proof of (13.4.8). We will use again the bound (13.4.10). The first term on the right-hand side of (13.4.10) tends to zero, since it is the general term of a summable series by assumption. Consider the second term on the right-hand side of (13.4.10). As previously, it can be bounded by

$$r(n)\mathbb{E}_\pi[\mathbb{1}\{\sigma_\alpha \geq n\} f(X_n)] \leq \pi(\alpha)\mathbb{E}_\alpha\left[\sum_{k=n}^{\sigma_\alpha} r(k) f(X_k)\right] .$$

Since the last expectation is finite by assumption (13.4.5), this shows that $\lim_{n\to\infty} = 0$ by Lebesgue's dominated convergence theorem. We finally consider the last term of the right-hand side of (13.4.10). Set $a(j) = r(j)\left\|\xi P^j - \pi\right\|_{\mathrm{TV}}$ and $b(j) = r(j)\mathbb{E}_\alpha[\mathbb{1}\{\sigma_\alpha \geq j\} f(X_j)]$. Using $r(n) \leq r(j)r(n-j)$ for $1 \leq j \leq n-1$, we have

$$\begin{aligned} r(n)\sum_{j=1}^{n-1}\left\|\xi P^j - \pi\right\|_{\mathrm{TV}} \mathbb{E}_\alpha[\mathbb{1}\{\sigma_\alpha \geq n-j\} f(X_{n-j})] \\ \leq \sum_{j=1}^{n-1} a(j)b(n-j) = \sum_{k=1}^{\infty} b(k)a(n-k)\mathbb{1}\{1 \leq k < n\} . \end{aligned}$$

By Theorem 13.3.3 (ii),

$$\lim_{n\to\infty} a(n-k)\mathbb{1}\{1 \leq k \leq n\} = \lim_{n\to\infty} r(n)\left\|\xi P^n - \pi\right\|_{\mathrm{TV}} = 0 .$$

Moreover, $\sum_{k=1}^{\infty} b(k) < \infty$ by (13.4.5); thus Lebesgue's dominated convergence theorem yields

$$\begin{aligned} \lim_{n\to\infty} r(n)\sum_{j=1}^{n-1}\left\|\xi P^j - \pi\right\|_{\mathrm{TV}}\mathbb{E}_\alpha[\mathbb{1}\{\sigma_\alpha \geq n-j\} f(X_{n-j})] \\ = \lim_{n\to\infty}\sum_{k=1}^{\infty} b(k)a(n-k)\mathbb{1}\{1 \leq k < n\} = 0 . \end{aligned} \tag{13.4.11}$$

The proof of (13.4.8) is complete.

(iv) Without loss of generality, we assume that $\Delta r \in \bar{\mathscr{S}}$. Since $f \geq 1$, the assumption (13.4.5) implies that $\mathbb{E}_\alpha[r^0(\sigma_\alpha)] < \infty$. Applying Proposition 13.4.1 (ii) with r replaced by Δr and ξ' replaced by π yields

$$\begin{aligned} \sum_{n=1}^{\infty}\Delta r(n)\left\|\xi P^n - \pi\right\|_f \leq \mathbb{E}_\xi\left[\sum_{j=1}^{\sigma_\alpha}\Delta r(j) f(X_j)\right] + \mathbb{E}_\pi\left[\sum_{j=1}^{\sigma_\alpha}\Delta r(j) f(X_j)\right] \\ + \mathbb{E}_\alpha\left[\sum_{j=1}^{\sigma_\alpha}\Delta r(j) f(X_j)\right]\sum_{n=1}^{\infty}\Delta r(n)\left\|\xi P^n - \pi\right\|_{\mathrm{TV}} . \end{aligned} \tag{13.4.12}$$

The second term of the right-hand side is finite according to Lemma 6.4.3 (with r replaced by Δr) and (13.4.9). Since $r \in \Lambda_1$, Theorem 13.3.3 (iv) together (13.4.5) implies that the last term of the right-hand side of (13.4.12) is finite.

□

13.5 Exercises

13.1. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P admits an accessible, aperiodic, and positive atom α. Let $s > 0$ and assume that $\mathbb{E}_\alpha[\sigma_\alpha^s] < \infty$.

1. Show that there exists $M > 0$ such that for all initial distributions λ and μ,

$$\sum_{n=0}^{\infty} n^{s-1} \mathrm{d}_{\mathrm{TV}}(\lambda P^n, \mu P^n) \leq M\{\mathbb{E}_\lambda[\sigma_\alpha^s] + \mathbb{E}_\mu[\sigma_\alpha^s]\} . \tag{13.5.1}$$

2. Assume that $\mathbb{E}_\lambda[\sigma_\alpha^s] + \mathbb{E}_\mu[\sigma_\alpha^s] < \infty$. Show that $\lim_{n\to\infty} n^s \mathrm{d}_{\mathrm{TV}}(\lambda P^n, \mu P^n) = 0$.

13.2. In this exercise, we use the notation of Chapter 8. We consider a renewal process $\{S_k,\, k \in \mathbb{N}\}$ with aperiodic waiting time distribution b and delay distribution a. We want to investigate the rate of convergence of the pure and delayed renewal sequences $\{u(n),\, n \in \mathbb{N}\}$ and $\{v_a(n),\, n \in \mathbb{N}\}$ to its limit m^{-1}, where $m = \sum_{n=1}^{\infty} nb(n)$, assumed to be finite.

1. Assume that there exists $\beta > 1$ such that $\sum_{n=1}^{\infty} \beta^n a(n) < \infty$. Show that there exist $\delta > 1$ and a constant M such that

$$\sum_{n=1}^{\infty} \delta^n |v_a(n) - u(n)| \leq M \sum_{n=1}^{\infty} \beta^n a(n) .$$

2. Assume that there exists $r \in \bar{\Lambda}_1$ such that $\sum_{n=1}^{\infty} r^0(n) b(n) < \infty$. Show that there exists a constant M such that for every delay distribution a,

$$\sum_{n=1}^{\infty} \Delta r(n) |v_a(n) - u(n)| < M \sum_{n=1}^{\infty} r(n) a(n) ,$$

$$\sum_{n=1}^{\infty} r(n) |v_a(n) - u(n)| < M \sum_{n=1}^{\infty} r^0(n) a(n) .$$

13.3. Let $\{p(n),\, n \in \mathbb{N}\}$ be a sequence of positive real numbers such that $p(0) = 1$, $p(n) \in (0,1)$ for all $n \geq 1$, and $\lim_{n\to\infty} \prod_{i=1}^{n} p(i) = 0$. Consider the backward recurrence time chain with transition kernel P defined as $P(n, n+1) = 1 - P(n,\, 0) = p_n$, for all $n \geq 0$ (see Exercise 8.2). Assume that

$$\sum_{n=1}^{\infty} \prod_{j=1}^{n} p_j < \infty .$$

Let σ_0 be the return time to $\{0\}$.

1. Show that P is irreducible, aperiodic, and positive recurrent and that the unique invariant probability π is given for all $j \in \mathbb{N}$ by

$$\pi(j) = \frac{p_0 \cdots p_{j-2}}{\sum_{n=1}^{\infty} p_1 \dots p_n} .$$

2. Show that for all functions $f_k : \mathbb{N} \to \mathbb{R}_+$,

$$\mathbb{E}_0 \left[\sum_{k=0}^{\sigma_0 - 1} f_k(X_k) \right] = \mathbb{E}_0 \left[\sum_{k=0}^{\sigma_0 - 1} f_k(k) \right].$$

(Therefore, there is no loss of generality to consider only $(1,r)$-modulated moments of the return time to zero.)

3. Assume that $\sup_{n \geq 1} p_n \leq \lambda < 1$. Show that there exists $\beta > 1$ such that for all initial distributions λ, $\limsup_{n \to \infty} \beta^n \|\lambda P^n - \pi\|_{\mathrm{TV}} = 0$.
4. Assume that for some $\theta > 0$, $p_n = 1 - (1+\theta)n^{-1} + o(n^{-1})$. Show that for all $\beta \in [0, \theta)$, there exists a constant C such that for all initial distributions λ,

$$\sum_{n=1}^{\infty} n^{-1+\beta} \|\lambda P^n - \pi\|_{\mathrm{TV}} \leq C \,.$$

13.4 (Continuation of Exercise 1.12). Let X be a finite set and π a probability on X such that $\pi(x) > 0$ for all $x \in \mathsf{X}$. Let M be a Markov transition matrix reversible with respect to π, i.e., $\pi(x)M(x,y) = \pi(y)M(y,x)$ for all $x, y \in \mathsf{X}$. In this exercise, we derive bounds on the rate of convergence in total variation distance in terms of the eigenvalues of M.

Let $(\beta_y)_{y \in \mathsf{X}}$ be the eigenvalues of M, $(f_y)_{y \in \mathsf{X}}$ an orthonormal basis in $\mathrm{L}^2(\pi)$ consisting of right eigenfunctions of M, and $(g_y)_{y \in \mathsf{X}}$ an orthonormal basis in $\mathrm{L}^2(1/\pi)$ consisting of left eigenfunctions of M.

Show that for every initial state x, $4\|M(x,\,\cdot) - \pi(\cdot)\|_{\mathrm{TV}}^2$ is bounded by each of the following three quantities:

$$\sum_y \beta_y^{2k} f_y^2(x) - 1 \,, \qquad \frac{1}{\pi^2(x)} \sum_y \beta_y^{2k} g_y(x) - 1 \,, \qquad \frac{1}{\pi(x)} (\beta^*)^{2k} \,.$$

13.6 Bibliographical Notes

Important references on coupling include Lindvall (1992) and Thorisson (2000). Proposition 13.2.9 is due to Lindvall (1979) (see also Lindvall (1992)). A preliminary version of this result was reported in Pitman (1974).

Chapter 14
Geometric Recurrence and Regularity

We have already seen that successive visits to petite sets play a crucial role in the study of the stability of an irreducible Markov chain. In Chapter 11, the existence of an invariant measure and its expression were obtained in terms of the return time to an accessible petite set. In this chapter, we will begin the study of rates of convergence to an invariant distribution by means of modulated moments of the return time to a petite set. However, in practice, it is with few exceptions difficult to compute these modulated moments. In this chapter, we introduce drift conditions that involve only the kernel P or one of its iterates P^n rather than the return or hitting times and relate them to the modulated moments of the excursions outside a petite set C. We first consider geometric moments and geometric drift conditions. The corresponding rates of convergence will be obtained in Chapter 15. Subgeometric moments and rates of convergence will be investigated in a parallel way in Chapters 16 and 17.

14.1 f-Geometric Recurrence and Drift Conditions

Definition 14.1.1 (f-Geometric recurrence) *Let $f : \mathsf{X} \to [1,\infty)$ be a measurable function and $\delta > 1$. A set $C \in \mathscr{X}$ is said to be (f,δ)-geometrically recurrent if*

$$\sup_{x\in C} \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} \delta^k f(X_k)\right] < \infty\,. \tag{14.1.1}$$

The set C is said to be f-geometrically recurrent if it is (f,δ)-geometrically recurrent for some $\delta > 1$. The set C is said to be geometrically recurrent if it is f-geometrically recurrent for some $f \geq 1$.

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_14

Note that C is geometrically recurrent if and only if there exists $\delta > 1$ such that $\sup_{x\in C} \mathbb{E}_x[\delta^{\sigma_C}] < \infty$. The f-geometric recurrence property is naturally associated with drift conditions. The main results of this section provide necessary and sufficient conditions for f-geometric recurrence in terms of drift conditions. The key tool is the comparison theorem (Theorem 4.3.1), which is essential to establish f-geometric recurrence of a set C from a sequence of drift conditions.

For $f : \mathsf{X} \to [1,\infty)$ a measurable function and $\delta \geq 1$, define

$$W_C^{f,\delta}(x) = \mathbb{E}_x\left[\sum_{k=0}^{\tau_C-1} \delta^{k+1} f(X_k)\right] , \tag{14.1.2}$$

with the convention $\sum_0^{-1} = 0$, so that $W_C^{f,\delta}(x) = 0$ for $x \in C$.

Proposition 14.1.2 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that there exist a measurable function $V : \mathsf{X} \to [0,\infty]$, a measurable function $f : \mathsf{X} \to [1,\infty)$, $\delta \geq 1$, and a set $C \in \mathscr{X}$ such that*

$$PV(x) + f(x) \leq \delta^{-1} V(x) , \qquad x \in C^c . \tag{14.1.3}$$

Then

(i) for all $x \in \mathsf{X}$,

$$\mathbb{E}_x\left[V(X_{\sigma_C})\delta^{\sigma_C}\mathbb{1}\{\sigma_C < \infty\}\right] + \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} \delta^{k+1} f(X_k)\right] \leq \delta\{PV(x) + f(x)\}\mathbb{1}_C(x) + V(x)\mathbb{1}_{C^c}(x) , \tag{14.1.4}$$

where we use the convention $0 \times \infty = 0$ on the right-hand side of the inequality;

(ii) the function $W_C^{f,\delta}$ given by (14.1.2) satisfies the drift condition (14.1.3).

Proof. The proof of (14.1.4) is an application of Theorem 4.3.1 with $\tau = \sigma_C$, $Z_n = \delta^{n+1} f(X_n)$,

$$V_0 = V(X_0)\mathbb{1}_{C^c}(X_0) , \qquad V_n = \delta^n V(X_n) n \geq 1 ,$$
$$Y_0 = \delta\{PV(X_0) + f(X_0)\} , \qquad Y_n = d\delta^n \mathbb{1}_C(X_n) n \geq 1 .$$

The proof of (ii) follows from elementary calculations. □

Proposition 14.1.3 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, $C \in \mathscr{X}$, $\delta > 1$, and let $f : \mathsf{X} \to [1,\infty)$ be a measurable function. The following conditions are equivalent:*

(i) The set C is (f,δ)-geometrically recurrent.
(ii) There exist a measurable function $V : \mathsf{X} \to [0,\infty]$ and $b \in [0,\infty)$ such that

$$PV + f \leq \delta^{-1} V + b\mathbb{1}_C \,, \tag{14.1.5}$$

and $\sup_{x\in C} V(x) < \infty$.

Moreover, if any, hence all, of these conditions holds, then the function $V = W_C^{f,\delta}$ satisfies (14.1.5).

Proof. (i) ⇒ (ii). Assume that C is (f,δ)-geometrically recurrent. Then for all $x \in \mathsf{X}$, we get

$$\delta P W_C^{f,\delta}(x) + \delta f(x) = PW_1(x) + r(0)h(x) = \delta \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} \delta^k f(X_k) \right] ,$$

and the function $W_C^{f,\delta}$ satisfies (14.1.5) with $b = \sup_{x\in C} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} \delta^k f(X_k) \right] < \infty$. Moreover, $\sup_{x\in C} V(x) = \sup_{x\in C} W_C^{f,\delta}(x) = 0 < \infty$.

(ii) ⇒ (i). Assume that $V : \mathsf{X} \to [0,\infty]$ is a function satisfying (14.1.5) and $\sup_{x\in C} V(x) < \infty$. Proposition 14.1.2 (i) shows that C is (f,δ)-geometrically recurrent.

□

We now examine these conditions for an irreducible Markov kernel.

Theorem 14.1.4. *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let $f : \mathsf{X} \to [1,\infty)$ be a measurable function and $V : \mathsf{X} \to [0,\infty]$ a measurable function such that $\{V < \infty\} \neq \emptyset$. The following conditions are equivalent:*

(i) There exist $\lambda \in [0,1)$ and $b \in [0,\infty)$ such that

$$PV + f \leq \lambda V + b \,. \tag{14.1.6}$$

Moreover, for all $d > 0$, the sets $\{V \leq d\}$ are petite, and there exists $d_0 \in [0,\infty)$ such that for all $d \geq d_0$, $\{V \leq d\}$ is accessible.
(ii) There exist $\lambda \in [0,1)$ and $b, d_1 \in [0,\infty)$ such that

$$PV + f \leq \lambda V + b\mathbb{1}_{\{V \leq d_1\}} \,, \tag{14.1.7}$$

and for all $d \geq d_1$, the sets $\{V \leq d\}$ are petite and accessible.
(iii) There exist a petite set C, $\lambda \in [0,1)$, and $b \in [0,\infty)$ such that

$$PV + f \leq \lambda V + b\mathbb{1}_C \,. \tag{14.1.8}$$

Proof. (i) $\Rightarrow$ (ii) Let d_0 be as in (i). Choose $\tilde{\lambda} \in (\lambda, 1)$ and $d_1 \geq d_0 \vee b(\tilde{\lambda} - \lambda)^{-1}$. The level set $C = \{V \leq d_1\}$ is accessible and petite by assumption. For $x \in C$, (14.1.6) yields $PV(x) + f(x) \leq \tilde{\lambda} V(x) + b$. For $x \notin C$, $-(\tilde{\lambda} - \lambda)V(x) < -b$ and (14.1.6) imply

$$PV(x) + f(x) \leq \tilde{\lambda} V(x) + b - (\tilde{\lambda} - \lambda)V(x) < \tilde{\lambda} V(x) .$$

(ii) $\Rightarrow$ (iii) We obtain (14.1.8) from (14.1.7) by setting $C = \{V \leq d_1\}$.

(iii) $\Rightarrow$ (i) Condition (14.1.8) obviously implies (14.1.6). We next prove that the level set $\{V \leq d\}$ is petite for every $d > 0$. By (14.1.4), we get that

$$\begin{aligned}\lambda^{-1} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} \lambda^{-k} f(X_k) \right] &\leq \lambda^{-1} \{PV(x) + f(x)\} \mathbb{1}_C(x) + V(x) \mathbb{1}_{C^c}(x) \\ &\leq V(x) + \lambda^{-1} b \mathbb{1}_C(x) ,\end{aligned}$$

showing that

$$\{x \in \mathsf{X} : V(x) \leq d\} \subset \left\{x \in \mathsf{X} : \mathbb{E}_x[\lambda^{-\sigma_C}] \leq (d\lambda + b)(\lambda^{-1} - 1) + 1\right\} .$$

Since C is petite, the set on the right-hand side is petite by Lemma 9.4.8, and therefore $\{x \in \mathsf{X} : V(x) \leq d\}$ is also petite. Using (14.1.8), Proposition 9.2.13 applied with $V = V_0 = V_1$ shows that (noting $\{V < \infty\}$) the set $\{V < \infty\}$ is full and absorbing and there exists d_0 such that $\{V \leq d_0\}$ is accessible, which implies that for all $d \geq d_0$, $\{V \leq d\}$ is accessible.

□

We now introduce a drift condition that covers many cases of interest.

Definition 14.1.5 (Condition $\mathrm{D_g}(V, \lambda, b, C)$:Geometric drift toward C) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. The Markov kernel P is said to satisfy the condition $\mathrm{D_g}(V, \lambda, b, C)$ if $V : \mathsf{X} \to [1, \infty)$ is a measurable function, $\lambda \in [0, 1)$, $b \in [0, \infty)$, $C \in \mathscr{X}$, and*

$$PV \leq \lambda V + b \mathbb{1}_C . \tag{14.1.9}$$

If $C = \mathsf{X}$, we simply write $\mathrm{D_g}(V, \lambda, b)$.

The function V is called a drift or test or Lyapunov function. If (14.1.9) holds, then for every $a > 0$, it also holds with V and b replaced by aV and ab. Therefore, there is no restriction in assuming $V \geq 1$ rather than an arbitrary positive lower bound. A bounded function V always satisfies condition $\mathrm{D_g}(V, \lambda, b)$ for every $\lambda \in (0, 1)$. It suffices to choose the constant b appropriately. Therefore, condition

$D_g(V, \lambda, b)$ is meaningful mostly when V is an unbounded function. In that case, the geometric drift condition is typically satisfied when

$$\limsup_{R\to\infty} \sup_{V(x)\geq R} \frac{PV(x)}{V(x)} < 1 ,$$
$$\text{for every } R > 0, \sup_{V(x)\leq R} PV(x) < \infty .$$

Condition $D_g(V, \lambda, b, C)$ obviously implies condition $D_g(V, \lambda, b)$. The converse is true if we set $\{V \leq d\}$ with d such that $\lambda + b/d < 1$ (see Exercise 14.1).

Note that the function f that appears in Proposition 14.1.2, Proposition 14.1.3, and Theorem 14.1.4 satisfies $f \geq 1$, and therefore, it is useless to write $D_g(V, \lambda, b, C)$ as $PV + f \leq \lambda V + b\mathbb{1}_C$ with $f \equiv 0$ for applying these results. Instead, the following remark allows us to derive a drift condition with a function $f \geq 1$.

Assume that condition $D_g(V, \lambda, b, C)$ is satisfied for some nonempty petite set C. Then for every $\tilde{\lambda} \in (\lambda, 1)$, we have $PV + (\tilde{\lambda} - \lambda)V \leq \tilde{\lambda}V + b\mathbb{1}_C$, or equivalently,

$$P\tilde{V} + f \leq \tilde{\lambda}\tilde{V} + \tilde{b}\mathbb{1}_C ,$$

where we have used the notation $\tilde{V} = V/(\tilde{\lambda} - \lambda)$, $f = V \geq 1$, and $\tilde{b} = b/(\tilde{\lambda} - \lambda)$. This shows that if C is petite, then $D_g(V, \lambda, b, C)$ implies Theorem 14.1.4 (iii), and hence that $\tilde{V}$ and f satisfy all of the equivalent conditions listed in Theorem 14.1.4.

This remark immediately implies the following corollary.

Corollary 14.1.6 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let $V : \mathsf{X} \to [1, \infty)$ be a measurable function. The following conditions are equivalent:*

(i) There exist $\lambda \in [0, 1)$ and $b \in [0, \infty)$ such that

$$PV \leq \lambda V + b . \tag{14.1.10}$$

Moreover, for all $d > 0$, the sets $\{V \leq d\}$ are petite and there exists $d_0 \in [0, \infty)$ such that $\{V \leq d_0\}$ is accessible.

(ii) There exist $\lambda \in [0, 1)$ and $b, d_1 \in [0, \infty)$ such that

$$PV \leq \lambda V + b\mathbb{1}_{\{V\leq d_1\}},$$

and for all $d \geq d_1$, the sets $\{V \leq d\}$ are petite and accessible.

(iii) There exist a petite set C, $\lambda \in [0, 1)$, and $b \in [0, \infty)$ such that

$$PV \leq \lambda V + b\mathbb{1}_C . \tag{14.1.11}$$

Example 14.1.7 (Random walk Metropolis algorithm). Let π be a probability measure on $\mathbb{R}$ and assume that it has a density function h_π over $\mathbb{R}$ with respect to Lebesgue measure. Assume that h_π is continuous, positive ($h_\pi(x) > 0$ for all $x \in \mathbb{R}$), and h_π is log-concave in the tails, i.e., there exists $\alpha > 0$ and some x_1 such that for all $y \geq x \geq x_1$,

$$\log h_\pi(x) - \log h_\pi(y) \geq \alpha(y-x) , \tag{14.1.12}$$

and similarly, for all $y \leq x \leq -x_1$,

$$\log h_\pi(x) - \log h_\pi(y) \geq \alpha(x-y) . \tag{14.1.13}$$

Denote by $\bar{q}$ a continuous, positive, and symmetric density on $\mathbb{R}$ and consider the random walk Metropolis (RWM) algorithm (see Example 2.3.2) associated with the increment distribution $\bar{q}$. We denote by P the associated Markov kernel. For each $x \in \mathbb{R}$, define the sets

$$A_x = \{y \in \mathbb{R} : h_\pi(x) \leq h_\pi(y)\} , \quad R_x = \{y \in \mathbb{R} : h_\pi(x) > h_\pi(y)\} ,$$

for the acceptance and (possible) rejection regions for the chain started from $x \in \mathbb{R}$. It is easily seen that P is irreducible. It is not difficult to show that every compact set $C \subset \mathbb{R}$ such that $\mathrm{Leb}(C) > 0$ is small. Indeed, by positivity and continuity, we have $\sup_{x \in C} h_\pi(x) < \infty$ and $\inf_{x,y \in C} \bar{q}(|y-x|) > 0$. For a fixed $x \in C$ and $B \subset C$,

$$\begin{aligned}
P(x,B) &\geq \int_{R_x \cap B} \bar{q}(|y-x|)\alpha(x,y)\mathrm{d}y + \int_{A_x \cap B} \bar{q}(|y-x|)\alpha(x,y)\mathrm{d}y \\
&= \int_{R_x \cap B} \frac{h_\pi(y)}{h_\pi(x)} \bar{q}(|y-x|)\mathrm{d}y + \int_{A_x \cap B} \bar{q}(|y-x|)\mathrm{d}y \\
&\geq \frac{\varepsilon}{d} \int_{R_x \cap B} h_\pi(y)\mathrm{d}y + \frac{\varepsilon}{d} \int_{A_x \cap B} h_\pi(y)\mathrm{d}y = \varepsilon d^{-1} \pi(B) ,
\end{aligned}$$

with $\varepsilon = \inf_{x,y \in C} \bar{q}(|y-x|)$ and $d = \sup_{x \in C} h_\pi(x)$. Hence for all $B \in \mathscr{B}(\mathbb{R})$ and $x \in C$,

$$P(x,B) \geq P(x, B \cap C) \geq \frac{\varepsilon \pi(C)}{d} \frac{\pi(B \cap C)}{\pi(C)} ,$$

which shows that C is 1-small and hence that P is strongly aperiodic.

We next establish the geometric drift condition. Assume first that h_π is symmetric. In this case, by (14.1.12), there exists x_0 such that $A_x = \{y \in \mathbb{R} : |y| \leq |x|\}$ for $|x| > x_0$. Let us choose a $x^* \geq x_0 \vee x_1$ and consider the Lyapunov function $V(x) = \mathrm{e}^{s|x|}$ for all $s < \alpha$. Define $Q(x, \mathrm{d}y) = \bar{q}(y-x)\mathrm{d}y$. Identifying moves to A_x, R_x, and $\{x\}$ separately, we can write, for $x \geq x^*$,

$$\lambda_x := \frac{PV(x)}{V(x)} = 1 + \int_{\{|y|\le x\}} Q(x,\mathrm{d}y)[\exp(s(|y|-x))-1] \\ + \int_{\{|y|>x\}} Q(x,\mathrm{d}y)[\exp(s(|y|-x))-1][h_\pi(y)/h_\pi(x)]. \qquad (14.1.14)$$

The log-concavity implies that for $y \ge x \ge x^*$, $h_\pi(y)/h_\pi(x) \le \mathrm{e}^{-\alpha(y-x)}$. Therefore, we have, for $x \ge x^*$ and $s < \alpha$,

$$\lambda_x \le 1 + Q(x,(2x,\ \infty)) + Q(x,(-\infty,0)) + \int_0^x Q(x,\mathrm{d}y)[\exp(s(y-x))-1] \\ + \int_x^{2x} Q(x,\mathrm{d}y)\exp(-\alpha(y-x))[\exp(s(y-x))-1]\,. \qquad (14.1.15)$$

The terms $Q(x,(2x,\ \infty))$ and $Q(x,(-\infty,0))$ are bounded by $\int_x^\infty \bar{q}(z)\mathrm{d}z$ and can therefore be made arbitrarily small by taking x^* large enough, since it is assumed that $x \ge x^*$. We will have a drift toward $C = [-x^*,\ x^*]$ if the sum of the second and third terms in (14.1.15) is strictly bounded below 0 for all $x \ge x^*$. These terms may be expressed as

$$\int_0^x \bar{q}(z)[\mathrm{e}^{-sz} - 1 + \mathrm{e}^{-(\alpha-s)z} - \mathrm{e}^{-\alpha z}]\mathrm{d}z = -\int_0^x \bar{q}(z)[1-\mathrm{e}^{-sz}][1-\mathrm{e}^{-(\alpha-s)z}]\mathrm{d}z\,.$$

Since the integrand on the right is positive and increasing as z increases, we find that for suitably large x^*, λ_x in (14.1.15) is strictly less than 1.

For $0 \le x \le x^*$, the right-hand side of (14.1.15) is bounded by

$$1 + 2\int_{x^*}^\infty \bar{q}(z)\mathrm{d}z + 2\exp(sx^*)\int_0^{x^*} \bar{q}(z)\mathrm{d}z\,.$$

For negative x, the same calculations are valid by symmetry. Therefore, $\mathrm{D_g}(V,\lambda,b,C)$ holds with $V(x) = \mathrm{e}^{s|x|}$ and $C = [-x^*,x^*]$, which is small. Thus condition (iii) in Corollary 14.1.6 is satisfied.

Consider now the general case. We have immediately from the construction of the algorithm that there exists $x_0 \in \mathbb{R}$ such that for $x > x_0$, the set $(x,\ \infty)$ is a subset of R_x and the set $(-x,x)$ is a subset of A_x; similarly for $x < -x_0$, the set $(-\infty,x)$ is a subset of R_x and the set $(x,-x)$ is a subset of A_x. Again set $V(x) = \mathrm{e}^{s|x|}$. The only difference stems from the fact that we need to control the term, for $x > 0$,

$$\int_{y\le -x} Q(x,\mathrm{d}y)[\exp(s(|y|-x))-1][1 \vee h_\pi(y)/h_\pi(x)]$$

in (14.1.14). This term is negligible if $q(x) \le b\exp(-\alpha|x|)$. Under this additional condition, the condition $\mathrm{D_g}(V,\lambda,b,C)$ is satisfied, and condition (iii) in Corollary 14.1.6 is again satisfied. ◀

Proposition 14.1.8 *Let P be a Markov kernel satisfying condition* $D_g(V,\lambda,b)$*. Then for each positive integer m,*

$$P^m V \le \lambda^m V + \frac{b(1-\lambda^m)}{1-\lambda} \le \lambda^m V + \frac{b}{1-\lambda} . \tag{14.1.16}$$

Conversely, if there exists $m \ge 2$ such that P^m satisfies condition $D_g(V_m,\lambda_m,b_m)$*, then P satisfies condition* $D_g(V,\lambda,b)$ *with*

$$V = V_m + \lambda_m^{-1/m} P V_m + \cdots + \lambda_m^{-(m-1)/m} P^{m-1} V_m ,$$
$$\lambda = \lambda_m^{1/m} \quad \text{and} \quad b = \lambda_m^{-(m-1)/m} b_m .$$

Proof. Assume that $PV \le \lambda V + b$ with $\lambda \in (0,1)$ and $b \in [0,\infty)$. By straightforward induction, we obtain, for $m \ge 1$,

$$P^m V \le \lambda^m V + b \sum_{k=0}^{m-1} \lambda^k \le \lambda^m V + b(1-\lambda^m)/(1-\lambda) .$$

This proves the first part. Conversely, if $P^m V_m \le \lambda_m V_m + b_m$, set

$$V = V_m + \lambda_m^{-1/m} P V_m + \cdots + \lambda_m^{-(m-1)/m} P^{m-1} V_m .$$

Then

$$\begin{aligned} PV &= PV_m + \lambda_m^{-1/m} P^2 V_m + \cdots + \lambda_m^{-(m-1)/m} P^m V_m \\ &\le PV_m + \lambda_m^{-1/m} P^2 V_m + \cdots + \lambda_m^{-(m-2)/m} P^{m-1} V_m + \lambda_m^{-(m-1)/m} (\lambda_m V_m + b_m) , \\ &= \lambda_m^{1/m} V + \lambda_m^{-(m-1)/m} b_m . \end{aligned}$$

□

Remark 14.1.9. Since $V \ge 1$ (provided that it is not identically equal to infinity), letting m tend to infinity in (14.1.16) yields $1 \le b/(1-\lambda)$, i.e., λ and b must always satisfy $\lambda + b \ge 1$. ▲

Lemma 14.1.10 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P satisfies the drift condition* $D_g(V,\lambda,b)$*. If P admits an invariant probability measure π such that $\pi(\{V=\infty\}) = 0$, then $\pi(V) < \infty$.*

Proof. By Proposition 14.1.8, for all positive integers m,

$$P^m V \le \lambda^m V + \frac{b}{1-\lambda} .$$

The concavity of the function $x \mapsto x \wedge c$ yields for all $n \in \mathbb{N}$ and $c > 0$,

$$\pi(V \wedge c) = \pi P^n(V \wedge c) \le \pi(\{P^n V\} \wedge c) \le \pi(\{\lambda^n V + b/(1-\lambda)\} \wedge c) .$$

Letting n and then c tend to infinity yields $\pi(V) \le b/(1-\lambda)$. □

14.2 f-Geometric Regularity

Definition 14.2.1 (f-Geometrically regular sets and measures) *Let P be an irreducible kernel on $\mathsf{X} \times \mathscr{X}$ and $f : \mathsf{X} \to [1,\infty)$ a measurable function.*

(i) A set $A \in \mathscr{X}$ is said to be f-geometrically regular if for every $B \in \mathscr{X}_P^+$, there exists $\delta > 1$ (possibly depending on A and B) such that

$$\sup_{x \in A} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_B - 1} \delta^k f(X_k) \right] < \infty .$$

(ii) A probability measure $\xi \in \mathbb{M}_1(\mathscr{X})$ is said to be f-geometrically regular if for every $B \in \mathscr{X}_P^+$, there exists $\delta > 1$ (possibly depending on ξ and B) such that

$$\mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_B - 1} \delta^k f(X_k) \right] < \infty .$$

(iii) A point $x \in \mathsf{X}$ is f-geometrically regular if δ_x is f-geometrically regular.
(iv) The Markov kernel P is said to be f-geometrically regular if there exists an accessible f-geometrically regular set.

When $f \equiv 1$ in the preceding definition, we will simply say geometrically regular instead of 1-geometrically regular. If A is geometrically regular, then every probability measure ξ such that $\xi(A) = 1$ is geometrically regular.

Recall that a set C is (f,δ)-geometrically recurrent if there exists $\delta > 1$ such that

$$\sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} \delta^k f(X_k) \right] < \infty . \tag{14.2.1}$$

It is therefore straightforward to see that an f-geometrically regular accessible set is f-geometrically recurrent. At first sight, regularity seems to be a much stronger requirement than recurrence. In particular, the intersection and the union of two f-geometrically regular sets are still f-geometrically regular sets, whereas the intersection of two f-geometrically recurrent sets is not necessarily f-geometrically recurrent.

We preface the proof with a technical lemma. Recall the set $\bar{\mathscr{S}}$ from Definition 13.1.1.

Lemma 14.2.2 *Let $r \in \bar{\mathscr{S}}$ be such that $\kappa = \sup_{x\in A} \mathbb{E}_x[r(\sigma_A)] < \infty$. Then for every $n \geq 1$ and $h \in \mathbb{F}_+(\mathsf{X})$, we get*

$$\sup_{x\in A} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_A^{(n)}-1} r(k)h(X_k) \right] \leq \left(\sum_{k=0}^{n-1} \kappa^k \right) \sup_{x\in A} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_A-1} r(k)h(X_k) \right] .$$

Proof. Without loss of generality, we assume that $r \in \mathscr{S}$. Set $S_n = \sum_{k=0}^{\sigma_A^{(n)}-1} r(k)h(X_k)$. Then using $r(n+m) \leq r(n)r(m)$, we get

$$S_n = \sum_{k=0}^{\sigma_A-1} r(k)h(X_k) + \sum_{k=\sigma_A}^{\sigma_A+\sigma_A^{(n-1)}\circ\theta_{\sigma_A}} r(k)h(X_k) \leq S_1 + r(\sigma_A) S_{n-1} \circ \theta_{\sigma_A}$$

on the set $\{\sigma_A^{(n-1)} < \infty\}$, which implies that $\mathbb{E}_x[S_n] \leq \mathbb{E}_x[S_1] + \kappa \sup_{x\in A} \mathbb{E}_x[S_{n-1}]$. Setting for $n \geq 1$, $B_n = \sup_{x\in A} \mathbb{E}_x[S_n]$, we obtain the recurrence $B_n \leq B_1 + \kappa B_{n-1}$, which yields $B_n \leq B_1(1+\kappa+\cdots+\kappa^{n-1})$. ☐

Theorem 14.2.3. *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$ and $A, B \in \mathscr{X}$. Assume that*

(i) there exists $q \in \mathbb{N}^*$ such that $\inf_{x\in A} \mathbb{P}_x(\sigma_B \leq q) > 0$;
(ii) $\sup_{x\in A} \mathbb{E}_x[\delta^{\sigma_A}] < \infty$ for some $\delta > 1$.

Then there exist $\beta \in (1, \delta)$ and $\varsigma < \infty$ such that for all $h \in \mathbb{F}_+(\mathsf{X})$,

$$\sup_{x\in A} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_B-1} \beta^k h(X_k) \right] \leq \varsigma \sup_{x\in A} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_A-1} \delta^k h(X_k) \right] .$$

Proof. We apply Theorem 11.4.1 with $\tau = \sigma_A^{(q)}$ and $\rho = \sigma_B$. It is easily seen that $\rho = \sigma_B$ satisfies (11.4.1). Since $q \leq \tau$, we get $0 < \inf_{x\in A} \mathbb{P}_x(\sigma_B \leq q) \leq \inf_{x\in A} \mathbb{P}_x(\sigma_B \leq \tau)$. Moreover, since $\mathbb{P}_x(\sigma_A < \infty) = 1$ for all $x \in A$, Proposition 4.2.5 (ii) implies $\mathbb{P}_x(\sigma_A^{(q)} < \infty) = 1$, and thus

$$\mathbb{P}_x(\tau < \infty,\, X_\tau \in A) = \mathbb{P}_x(\tau < \infty) = 1 ,$$

showing that (11.4.2) is satisfied. The proof follows from Lemma 14.2.2. ☐

Theorem 14.2.4. *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$, let $f : \mathsf{X} \to [1, \infty)$ be a measurable function, and let $C \in \mathscr{X}$ be a set. The following conditions are equivalent:*

(i) The set C is accessible and f-geometrically regular.
(ii) The set C is petite and f-geometrically recurrent.

Proof. (i) ⇒ (ii) Assume that the set C is accessible and f-geometrically regular. Then by definition, there exists $\delta > 1$ such that C is f-geometrically recurrent. Let D be an accessible petite set. The definition of geometric regularity implies that $\sup_{x \in C} \mathbb{E}_x[\sigma_D] < \infty$; therefore, the set D is uniformly accessible from C, which implies that C is petite by Lemma 9.4.8.

(ii) ⇒ (i) Assume that C is an f-geometrically recurrent petite set. Then C is accessible by Corollary 9.2.14. Let A be an accessible set. By Proposition 9.4.9, there exist $q \in \mathbb{N}^*$ and $\gamma > 0$ such that $\inf_{x \in C} \mathbb{P}_x(\sigma_A \le q) \ge \gamma$. By Theorem 14.2.3, there exists $\beta > 1$ such that

$$\sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_A - 1} \beta^k f(X_k) \right] < \infty \, .$$

This proves that C is f-geometrically regular. □

We have seen in Lemma 9.4.3 that a set that leads uniformly to a petite set is itself petite. There exists a similar criterion for geometric regularity.

Lemma 14.2.5 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$, let $f : \mathsf{X} \to [1, \infty)$ be a measurable function, and let C be an accessible f-geometrically regular set. Then*

(i) for all $B \in \mathscr{X}_P^+$ and $\delta \in (1, \infty)$, there exist constants $(\beta, \varsigma) \in (1, \delta) \times \mathbb{R}$ such that for all $x \in \mathsf{X}$,

$$\mathbb{E}_x \left[\sum_{k=0}^{\sigma_B - 1} \beta^k f(X_k) \right] \le \varsigma \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} \delta^k f(X_k) \right] ;$$

(ii) every set $A \in \mathscr{X}$ satisfying $\sup_{x \in A} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} \delta^k f(X_k) \right] < \infty$ for some $\delta > 1$ is f-geometrically regular;

(iii) every probability measure $\xi \in \mathbb{M}_1(\mathscr{X})$ satisfying $\mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_C - 1} \delta^k f(X_k) \right] < \infty$ for some $\delta > 1$ is f-geometrically regular.

Proof. First note that (ii) and (iii) are immediate from (i). We now prove (i). Since C is f-geometrically regular, for all $B \in \mathscr{X}_P^+$, there exists $\beta > 1$ such that

$$\sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_B - 1} \beta^k f(X_k) \right] < \infty \, .$$

Replacing β by a smaller value if necessary, we may assume that $\beta \in (1, \delta)$. Since $\sigma_B \leq \sigma_C \mathbb{1}\{\sigma_C = \infty\} + (\sigma_C + \sigma_B \circ \theta_{\sigma_C})\mathbb{1}\{\sigma_C < \infty\}$, we have for all $x \in \mathsf{X}$,

$$
\begin{aligned}
&\mathbb{E}_x\left[\sum_{k=0}^{\sigma_B-1} \beta^k f(X_k)\right] \\
&\leq \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} \beta^k f(X_k)\right] + \mathbb{E}_x\left[\mathbb{1}_{\{\sigma_C<\infty\}}\beta^{\sigma_C}\left\{\sum_{k=0}^{\sigma_B-1} \beta^k f(X_k)\right\} \circ \theta_{\sigma_C}\right] \\
&\leq \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} \beta^k f(X_k)\right] + \mathbb{E}_x[\beta^{\sigma_C}] \sup_{x\in C} \mathbb{E}_x\left[\sum_{k=0}^{\sigma_B-1} \beta^k f(X_k)\right] .
\end{aligned}
$$

The result follows, since $\mathbb{E}_x[\beta^{\sigma_C}] \leq (\beta-1)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} \beta^k f(X_k)\right] + 1$ and $\beta < \delta$. $\square$

Theorem 14.2.6. *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $f : \mathsf{X} \to [1, \infty)$ a measurable function. The Markov kernel P is f-geometrically regular if and only if it satisfies one of the following equivalent conditions:*

(i) There exists an f-geometrically recurrent petite set.
(ii) There exist a function $V : \mathsf{X} \to [0, \infty]$ such that $\{V < \infty\} \neq \emptyset$, a nonempty petite set C, $\lambda \in [0, 1)$, and $b < \infty$ such that

$$PV + f \leq \lambda V + b\mathbb{1}_C .$$

(iii) There exists an accessible f-geometrically regular set.
(iv) There exists a full and absorbing set S that can be covered by a countable number of accessible f-geometrically regular sets.

If any of these conditions holds, the Markov kernel P satisfies the following properties, with V as in (ii):

(a) A probability measure $\xi \in \mathbb{M}_1(\mathscr{X})$ is f-geometrically regular if and only if there exist an f-geometrically recurrent petite set C and $\delta > 1$ such that $\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} \delta^k f(X_k)\right] < \infty$.
(b) For every $A \in \mathscr{X}_P^+$, there exist constants $\varsigma < \infty$ and $\beta > 1$ such that for all $x \in \mathsf{X}$,

$$\mathbb{E}_x\left[\sum_{k=0}^{\sigma_A-1} \beta^k f(X_k)\right] \leq \varsigma\{V(x)+1\} . \tag{14.2.2}$$

(c) Every probability measure $\xi \in \mathbb{M}_1(\mathscr{X})$ such that $\xi(V) < \infty$ is f-geometrically regular.
(d) The set $S_P(f)$ of f-geometrically regular points is full and absorbing and contains the full and absorbing set $\{V < \infty\}$.

Proof. (i) $\Rightarrow$ (ii) This is immediate from Proposition 14.1.3 applied to the f-geometrically recurrent petite set C (and since $\sup_C V < \infty$ implies in particular that $\{V < \infty\} \neq \emptyset$).

(ii) $\Rightarrow$ (iii) By Theorem 14.1.4, there exists a petite and accessible level set $C = \{V \leq d\}$ such that (14.1.7) holds. Together with Proposition 14.1.3, this implies that C is an f-geometrically recurrent set. Since C is in addition petite, Theorem 14.2.4 then shows that C is an accessible f-geometrically regular set.

(iii) $\Rightarrow$ (iv) Let C be an accessible f-geometrically regular set. Using Theorem 14.2.4, the set C is (f,δ)-geometrically recurrent. Since by Proposition 14.1.3, the function $V = W_C^{f,\delta}$ defined in (14.1.2) satisfies $PV + f \leq \delta^{-1}V + b\mathbb{1}_C$, Proposition 9.2.13 with $V_0 = V_1 = W_C^{f,\delta}$ shows that the nonempty set $\{W_C^{f,\delta} < \infty\}$ is full and absorbing and that there exists n_0 such that for all $n \geq n_0$, the sets $\{W_C^{f,\delta} \leq n\}$ are accessible. Moreover, Lemma 14.2.5 (ii) shows that the sets $\{W_C^{f,\delta} \leq n\}$ are f-geometrically regular. Since their union covers $\{W_C^{f,\delta} < \infty\}$, the result follows.

(iii) $\Rightarrow$ (i) Obvious by Theorem 14.2.4.

(a) By Lemma 14.2.5 (iii), every $\xi \in \mathbb{M}_1(\mathscr{X})$ satisfying $\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1}\delta^k f(X_k)\right] < \infty$ for some petite set C and $\delta > 1$ is f-geometrically regular. This proves that the condition is sufficient.

Conversely, assume that ξ is f-geometrically regular. Since P is f-geometrically regular, there exists an f-geometrically regular and accessible set C. Since ξ is f-geometrically regular, $\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1}\delta^k f(X_k)\right] < \infty$. By Theorem 14.2.4, the set C is also f-geometrically recurrent and petite. This proves the necessary part.

(b) Under (ii), Theorem 14.1.4 shows that for some petite set $D = \{V \leq d\}$ and some constant $b < \infty$, the inequality $PV + f \leq \lambda V + b\mathbb{1}_{\{V \leq d\}}$ holds. Moreover, by Proposition 14.1.3, D is (f,λ^{-1})-geometrically recurrent. Finally, D is petite and (f,λ^{-1})-geometrically recurrent, and Lemma 14.2.5 (i) shows that there exist finite constants $\beta > 1$ and $\varsigma \in \left(1,\lambda^{-1}\right]$ such that

$$\mathbb{E}_x\left[\sum_{k=0}^{\sigma_A-1}\beta^k f(X_k)\right] \leq \varsigma \mathbb{E}_x\left[\sum_{k=0}^{\sigma_D-1}\lambda^{-k} f(X_k)\right] . \tag{14.2.3}$$

Since $PV + f \leq \lambda V + b\mathbb{1}_{\{V \leq d\}}$, (14.1.4) in Proposition 14.1.2 shows that for all $x \in \mathsf{X}$, $\mathbb{E}_x\left[\sum_{k=0}^{\sigma_D-1}\lambda^{-k} f(X_k)\right] \leq \lambda V(x) + b\mathbb{1}_D(x)$. Plugging this bound into (14.2.3) yields (14.2.2).

(c) This follows by integrating (14.2.2) with respect to $\xi \in \mathbb{M}_1(\mathscr{X})$.

(d) By (b), if $V(x) < \infty$, then $\xi = \delta_x$ is f-geometrically regular, and thus $S_P(f)$ contains $\{V < \infty\}$. Now, under (i), there exists an f-geometrically regular and accessible set C. Define $W_C^{f,\delta}$ as in (14.1.2) and note that

$$S_P(f) = \bigcup_{\delta \in \mathbb{Q} \cap [1,\infty]} \{W_C^{f,\delta} < \infty\} .$$

Since by Proposition 14.1.3, the function $V = W_C^{f,\delta}$ satisfies $PV + f \le \delta^{-1}V + b\mathbb{1}_C$, Proposition 9.2.13 (applied with $V_0 = V_1 = W_C^{f,\delta}$) shows that the nonempty set $\{W_C^{f,\delta} < \infty\}$ is full and absorbing. Thus $S_P(f)$ is full and absorbing as a countable union of full absorbing sets.

□

We conclude this section with conditions under which the invariant measure is f-geometrically regular.

Theorem 14.2.7. *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $f : \mathsf{X} \to [1,\infty)$ a measurable function. If P is f-geometrically regular, then P has a unique invariant probability measure π. In addition, π is f-geometrically regular.*

Proof. Since P is f-geometrically regular, Theorem 14.2.6 shows that there exists an f-geometrically recurrent petite set C, i.e., $\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1}\beta^k f(X_k)\right] < \infty$, for some $\beta > 1$. By Theorem 14.2.4, the set C is accessible and f-geometrically regular. Since $f \ge 1$, C satisfies $\sup_{x\in C}\mathbb{E}_x[\sigma_C] < \infty$, and Corollary 11.2.9 implies that P is positive (and recurrent) and admits a unique invariant probability measure π.

We will now establish that the invariant probability π is f-geometrically regular. By Theorem 14.2.6 (a), it suffices to show that $\mathbb{E}_\pi\left[\sum_{n=0}^{\sigma_C-1}\beta^n f(X_n)\right] < \infty$. Set $g(x) = \mathbb{E}_x\left[\sum_{n=0}^{\sigma_C-1}\beta^n f(X_n)\right]$ and $h(x) = \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} g(X_k)\right]$. Since C is accessible, Theorem 11.2.5 yields

$$\mathbb{E}_\pi\left[\sum_{n=0}^{\sigma_C-1}\beta^n f(X_n)\right] = \pi(g) = \int_C \pi(\mathrm{d}x)h(x)\,. \tag{14.2.4}$$

Setting $Z = \sum_{n=0}^{\infty}\mathbb{1}_{\{n<\sigma_C\}}\beta^n f(X_n)$, we have $g(x) = \mathbb{E}_x[Z]$ and

$$\begin{aligned} h(x) &= \sum_{k=0}^{\infty}\mathbb{E}_x[\mathbb{1}_{\{k<\sigma_C\}}Z\circ\theta_k] = \sum_{k=0}^{\infty}\sum_{n=0}^{\infty}\mathbb{E}_x[\mathbb{1}_{\{n+k<\sigma_C\}}\beta^n f(X_{n+k})] \\ &= \sum_{j=0}^{\infty}\sum_{\ell=0}^{j}\beta^\ell\mathbb{E}_x[\mathbb{1}_{\{j<\sigma_C\}}f(X_j)] \le \frac{\beta}{\beta-1}\mathbb{E}_x\left[\sum_{j=0}^{\sigma_C-1}\beta^j f(X_j)\right]. \end{aligned}$$

Since C is f-geometrically recurrent, we have

$$\mathbb{E}_\pi\left[\sum_{n=0}^{\sigma_C-1}\beta^n f(X_n)\right] \le \frac{\beta}{\beta-1}\pi(C)\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1}\beta^k f(X_k)\right] < \infty\,. \tag{14.2.5}$$

□

14.3 f-Geometric Regularity of the Skeletons

A natural issue is to relate the f-geometric regularity of a Markov kernel P and its skeletons. We will show below that if P is irreducible and aperiodic, then P is f-geometrically regular if and only if for all $m \in \mathbb{N}^*$, its skeleton P^m is $f^{(m)}$-geometrically regular, where

$$f^{(m)} = \sum_{i=0}^{m-1} P^i f \,. \tag{14.3.1}$$

Before we proceed with the proof, we need to obtain preparatory technical results. For every integer $m \in \mathbb{N}^*$ and $C \in \mathscr{X}$, define by $\sigma_{C,m}$ the first return time to the set C for the m-skeleton chain:

$$\sigma_{C,m} = \inf\{k \geq 1 \,:\, X_{km} \in C\} \,. \tag{14.3.2}$$

Set for $i \in \{0, \ldots, m-1\}$,

$$\vartheta_{C,m,i} = \inf\{n \geq 1 \,:\, n \equiv i\,[m], \quad X_n \in C\} \,, \tag{14.3.3}$$

and define

$$\vartheta_{C,m} = \max_{0 \leq i < m} \vartheta_{C,m,i} \,. \tag{14.3.4}$$

The following lemma summarizes the properties of $\vartheta_{C,m}$ that we will systematically exploit in the sequel.

Lemma 14.3.1 *Let P be an irreducible and aperiodic Markov kernel on $\mathsf{X} \times \mathscr{X}$, $m \geq 1$ an integer, and C an $(r, \varepsilon\nu)$-small set such that $\nu(C) > 0$ for some integer r. Then the following hold:*

(i) $\vartheta_{C,m}$ is a stopping time, and for all $n \in \mathbb{N}$, $\vartheta_{C,m} \leq n + \vartheta_{C,m} \circ \theta_n$.
(ii) There exists $q > 0$ such that

$$\inf_{x \in C} \mathbb{P}_x(\vartheta_{C,m} \leq q) > 0 \,. \tag{14.3.5}$$

(iii) Assume that C is Harris recurrent, i.e., for ally $x \in C$, $\mathbb{P}_x(\sigma_C < \infty) = 1$. Then for every $\xi \in \mathbb{M}_1(\mathscr{X})$ such that $\mathbb{P}_\xi(\sigma_C < \infty) = 1$, $\mathbb{P}_\xi(\vartheta_{C,m} < \infty) = 1$. Moreover, $\mathbb{P}_\xi(\sigma_{C,m} < \infty) = 1$, and C is Harris recurrent for P^m.

Proof. (i) Obvious.

(ii) By Lemma 9.3.3, there exist n_0 such that $n_0 \equiv 0\,[m]$ and a sequence of constants $\varepsilon_n > 0$ such that $\inf_{x \in C} P^n(x, \cdot) \geq \varepsilon_n \nu$ for all $n \geq n_0$. Define the events A_i, $i = 0, \ldots, m-1$, by

$$A_i = \{X_{n_0} \in C,\, X_{2n_0+1} \in C, \ldots,\, X_{(i+1)n_0+i} \in C\} \,.$$

By the Markov property, we have, for all $x \in C$ and $i \in \{1, \ldots, m-1\}$,

$$\mathbb{P}_x(A_i) = \mathbb{E}_x[\mathbb{1}_{A_{i-1}} P^{n_0+1}(X_{in_0+i-1}, C)] \geq \varepsilon_{n_0+1} \nu(C) \mathbb{P}_x(A_{i-1}) .$$

By induction, we get for all $x \in C$, $\mathbb{P}_x(A_{m-1}) \geq \gamma$ where $\gamma = \varepsilon_{n_0+1}^{m-1} \varepsilon_{n_0} \nu^m(C) > 0$. Set $q = mn_0 + m - 1$. Since $(i+1)n_0 + i \equiv i\,[m]$ for $i \in \{0, \ldots, m-1\}$, we have by construction $A_{m-1} \subset \{\vartheta_{C,m} \leq q\}$. Therefore, we get

$$\inf_{x \in C} \mathbb{P}_x(\vartheta_{C,m} \leq q) \geq \inf_{x \in C} \mathbb{P}_x(A_{m-1}) \geq \gamma > 0 .$$

(iii) We will apply Theorem 11.4.1 (i) with $\rho = \vartheta_{C,m}$ and $\tau = \sigma_C^{(q)}$. Since C is Harris recurrent, for all $x \in C$, $\mathbb{P}_x(\sigma_C^{(q)} < \infty, X_{\sigma_C^{(q)}} \in C) = 1$ and $\mathbb{P}_x(\vartheta_{C,m} \leq \sigma_C^{(q)}) \geq \inf_{x \in C} \mathbb{P}_x(\vartheta_{C,m} \leq q) > 0$. Hence for all $x \in C$, $\mathbb{P}_x(\vartheta_{C,m} < \infty) = 1$.

Since $\{\sigma_C < \infty, \vartheta_{C,m} \circ \theta_{\sigma_C} < \infty\} \subset \{\vartheta_{C,m} < \infty\}$, the strong Markov property then implies

$$\mathbb{P}_\xi(\sigma_C < \infty) = \mathbb{E}_\xi\left[\mathbb{1}_{\{\sigma_C < \infty\}} \mathbb{P}_{X_{\sigma_C}}(\vartheta_{C,m} < \infty)\right] \leq \mathbb{P}_\xi(\vartheta_{C,m} < \infty) .$$

Hence, since by assumption $\mathbb{P}_\xi(\sigma_C < \infty) = 1$, we obtain $\mathbb{P}_\xi(\vartheta_{C,m} < \infty) = 1$. Using now $\{\sigma_C < \infty, \vartheta_{C,m} \circ \theta_{\sigma_C} < \infty\} \subset \{\sigma_{C,m} < \infty\}$, the strong Markov property then implies for all $\xi' \in \mathbb{M}_1(\mathscr{X})$,

$$\mathbb{P}_{\xi'}(\sigma_C < \infty) = \mathbb{E}_{\xi'}\left[\mathbb{1}_{\{\sigma_C < \infty\}} \mathbb{P}_{X_{\sigma_C}}(\vartheta_{C,m} < \infty)\right] \leq \mathbb{P}_{\xi'}(\sigma_{C,m} < \infty) .$$

Taking $\xi' = \delta_x$ for all $x \in C$ then shows that the set C being Harris recurrent for P, it is also Harris recurrent for P^m. Taking now $\xi' = \xi$ yields $\mathbb{P}_\xi(\sigma_{C,m} < \infty) = 1$, since $\mathbb{P}_\xi(\sigma_C < \infty) = 1$.

□

Proposition 14.3.2 *Let P be an irreducible aperiodic Markov kernel on $\mathsf{X} \times \mathscr{X}$, $f : \mathsf{X} \to [1, \infty)$ a measurable function, and m an integer.*

(i) Let C be an (f, δ)-geometrically recurrent petite set. Then there exist $\beta \in (1, \delta)$ and $\varsigma < \infty$ such that for all $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_{C,m}-1} \beta^{mk} f^{(m)}(X_{mk})\right] \leq \varsigma \mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C - 1} \delta^k f(X_k)\right] ,$$

where $\sigma_{C,m}$ and $f^{(m)}$ are defined in (14.3.2) and (14.3.1). Moreover, the set C is $f^{(m)}$-geometrically recurrent for P^m.

(ii) For every $C \in \mathscr{X}$, $\delta > 1$, and $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C - 1} \delta^k f(X_k)\right] \leq \delta^m \mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_{C,m}-1} \delta^{mk} f^{(m)}(X_{mk})\right] .$$

If the set C is $f^{(m)}$-geometrically recurrent for P^m, then C is f-geometrically recurrent.

Proof. (i) For every initial distribution $\xi \in \mathbb{M}_1(\mathscr{X})$ and $\beta > 1$, we have

$$\begin{aligned}\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_{C,m}-1} \beta^{mk} f^{(m)}(X_{mk})\right] &\leq \sum_{i=0}^{m-1}\sum_{k=0}^{\infty} \beta^{mk+i}\mathbb{E}_\xi\left[f(X_{mk+i})\mathbb{1}\{mk < m\sigma_{C,m}\}\right] \\ &= \mathbb{E}_\xi\left[\sum_{k=0}^{m\sigma_{C,m}-1} \beta^k f(X_k)\right]. \end{aligned} \tag{14.3.6}$$

Since by construction, $m\sigma_{C,m} \leq \vartheta_{C,m}$ (see (14.3.4)), (14.3.6) yields

$$\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_{C,m}-1} \beta^{mk} f^{(m)}(X_{mk})\right] \leq \mathbb{E}_\xi\left[\sum_{k=0}^{\vartheta_{C,m}-1} \beta^k f(X_k)\right]. \tag{14.3.7}$$

Since the set C is petite and P is aperiodic, Theorem 9.4.10 implies that C is also $(r, \varepsilon\nu)$-small. By Lemma 9.1.6, without loss of generality, we may assume that $\nu(C) > 0$. By Lemma 14.3.1, there exists $q > 0$ such that

$$\inf_{x\in C}\mathbb{P}_x(\vartheta_{C,m} \leq q) > 0\,. \tag{14.3.8}$$

We use Theorem 11.4.1 with $\rho = \vartheta_{C,m}$ and $\tau = \sigma_C^{(q)}$. Lemma 14.3.1 (i) implies condition (11.4.1). Since C is f-geometrically recurrent, we have for all $x \in C$, $\mathbb{P}_x(\sigma_C^{(q)} < \infty) = 1$, which implies $\mathbb{P}_x(\tau < \infty, X_\tau \in C) = \mathbb{P}_x(\tau < \infty) = 1$. Moreover, using $\sigma_C^{(q)} \geq q$ and (14.3.8), we have $\inf_{x\in C}\mathbb{P}_x(\vartheta_{C,m} \leq \sigma_C^{(q)}) \geq \inf_{x\in C}\mathbb{P}_x(\vartheta_{C,m} \leq q) > 0$, showing that (11.4.2). Theorem 11.4.1 shows that there exist $\varsigma_1 < \infty$ and $\beta \in (1,\delta)$ such that

$$\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\vartheta_{C,m}-1} \beta^k f(X_k)\right] \leq \varsigma_1 \sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C^{(q)}-1} \delta^k f(X_k)\right].$$

Moreover, by Lemma 14.2.2, there exists $\varsigma_2 < \infty$ such that

$$\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C^{(q)}-1} \delta^k f(X_k)\right] \leq \varsigma_2 \sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} \delta^k f(X_k)\right].$$

Finally, there exist $\varsigma_3 < \infty$ and $\beta \in (1,\delta)$ such that

$$\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\vartheta_{C,m}-1} \beta^k f(X_k)\right] \leq \varsigma_3 \sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} \delta^k f(X_k)\right]. \tag{14.3.9}$$

Combining this with (14.3.7), where $\xi = \delta_x$, and taking the supremum on C, we get

$$\sup_{x\in C} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_{C,m}-1} \beta^{mk} f^{(m)}(X_{mk}) \right] \le \varsigma_3 \sup_{x\in C} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C-1} \delta^k f(X_k) \right] < \infty . \tag{14.3.10}$$

Thus C is $f^{(m)}$-geometrically recurrent for P^m. Moreover, the strong Markov property, together with $\vartheta_{C,m} \le \sigma_C + \mathbb{1}_{\{\sigma_C<\infty\}} \vartheta_{C,m} \circ \theta_{\sigma_C}$, yields

$$\begin{aligned}
&\mathbb{E}_\xi \left[\sum_{k=0}^{\vartheta_{C,m}-1} \beta^k f(X_k) \right] \\
&\le \mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_C-1} \beta^k f(X_k) \right] + \mathbb{E}_\xi \left[\mathbb{1}_{\{\sigma_C<\infty\}} \beta^{\sigma_C} \left\{ \sum_{k=0}^{\vartheta_{C,m}-1} \beta^k f(X_k) \right\} \circ \theta_{\sigma_C} \right] \\
&\le \mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_C-1} \delta^k f(X_k) \right] + \mathbb{E}_\xi \left[\delta^{\sigma_C} \right] \sup_{x\in C} \mathbb{E}_x \left[\sum_{k=0}^{\vartheta_{C,m}-1} \beta^k f(X_k) \right] .
\end{aligned}$$

Combining this with (14.3.9), there exists a constant $\varsigma < \infty$ such that for every $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_{C,m}-1} \beta^{mk} f^{(m)}(X_{mk}) \right] \le \varsigma \mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_C-1} \delta^k f(X_k) \right] .$$

(ii) By the Markov property, using that $m\sigma_{C,m}$ is a stopping time, we get

$$\begin{aligned}
\mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_{C,m}-1} \delta^{mk} f^{(m)}(X_{mk}) \right] &= \sum_{k=0}^{\infty} \mathbb{E}_\xi \left[\mathbb{1}_{\{mk<m\sigma_{C,m}\}} \delta^{mk} \sum_{j=0}^{m-1} P^j f(X_{mk}) \right] \\
&= \sum_{k=0}^{\infty} \mathbb{E}_\xi \left[\mathbb{1}_{\{mk<m\sigma_{C,m}\}} \delta^{mk} \sum_{j=0}^{m-1} f(X_{mk+j}) \right] .
\end{aligned}$$

Since $\delta^{mk} \le \delta^{-m} \delta^{mk+j}$ for $j \in \{0,\ldots,m-1\}$, we obtain

$$\begin{aligned}
&\mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_{C,m}-1} \delta^{mk} f^{(m)}(X_{mk}) \right] \\
&\qquad \ge \delta^{-m} \sum_{k=0}^{\infty} \mathbb{E}_\xi \left[\mathbb{1}_{\{mk<m\sigma_{C,m}\}} \sum_{j=0}^{m-1} \delta^{mk+j} f(X_{mk+j}) \right] \\
&\qquad = \delta^{-m} \mathbb{E}_\xi \left[\sum_{k=0}^{m\sigma_{C,m}-1} \delta^k f(X_k) \right] \ge \delta^{-m} \mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_C-1} \delta^k f(X_k) \right] .
\end{aligned}$$

Taking $\xi = \delta_x$ and summing over $x \in \mathsf{X}$ shows that if C is $f^{(m)}$-geometrically recurrent for P^m, then it is f-geometrically recurrent.

□

Theorem 14.3.3. *Let P be an irreducible aperiodic Markov kernel on $\mathsf{X} \times \mathscr{X}$, $f : \mathsf{X} \to [1, \infty)$ a measurable function, and $m \geq 2$.*

(i) A set C is accessible and f-geometrically regular if and only if C is accessible and $f^{(m)}$-geometrically regular for P^m.
(ii) The Markov kernel P is f-geometrically regular if and only if P^m is $f^{(m)}$-geometrically regular.
(iii) A probability measure ξ is f-geometrically regular for P if and only if ξ is $f^{(m)}$-geometrically regular for P^m.

Proof. (i) Assume first that C is an accessible f-geometrically regular set. By Theorem 14.2.4, the set C is petite (and hence small by Theorem 9.4.10, since P is aperiodic) and f-geometrically recurrent, i.e., $\sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} \delta^k f(X_k) \right] < \infty$ for some $\delta > 1$. By Theorem 9.3.11 (iii), the set C is accessible and small for P^m. By Proposition 14.3.2, there exist $\beta \in (1, \delta)$ and $\varsigma < \infty$ such that for every $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_{C,m} - 1} \beta^k f^{(m)}(X_{mk}) \right] \leq \varsigma \mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_C - 1} \delta^k f(X_k) \right] .$$

Setting $\xi = \delta_x$ and taking the supremum over $x \in C$ yields

$$\sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_{C,m} - 1} \beta^k f^{(m)}(X_{mk}) \right] < \infty .$$

Thus C is accessible, small, and $f^{(m)}$-geometrically recurrent for the kernel P^m. It is thus accessible and $f^{(m)}$-geometrically regular by Theorem 14.2.4.

Conversely, assume that C is accessible and an $f^{(m)}$-geometrically regular set for P^m. By Theorem 14.2.4, C is a nonempty petite, hence small, $f^{(m)}$-geometrically recurrent set for P^m. Applying Proposition 14.3.2 (ii) shows that the set C is f-geometrically recurrent for P. Since C is small for P^m, it is also small for P, and Theorem 14.2.4 shows that C is accessible and f-geometrically regular.

(ii) The Markov kernel P is f-geometrically regular if and only if there exists an accessible f-geometrically regular set C for P. Such a set is also accessible and $f^{(m)}$-geometrically regular for P^m. The proof follows from (i).

(iii) Let $\xi \in \mathbb{M}_1(\mathscr{X})$ be f-geometrically regular. By Theorem 14.2.6 (a), there exist a nonempty f-geometrically recurrent petite set C and $\delta > 1$ such that $\mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_C - 1} \delta^k f(X_k) \right] < \infty$. Proposition 14.3.2 (i) shows that there exists $\beta > 1$ such that

$$\mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_{C,m} - 1} \beta^{mk} f^{(m)}(X_{mk}) \right] < \infty ,$$

and C is $f^{(m)}$-geometrically recurrent for P^m. Applying again Theorem 14.2.6 (a), ξ is $f^{(m)}$-geometrically regular for P^m.

Conversely, if ξ is $f^{(m)}$-geometrically regular for P^m, then there exist, by Theorem 14.2.6 (a), a nonempty $f^{(m)}$-recurrent petite set C for P^m and $\beta > 1$ such that

$$\mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_{C,m}-1} \beta^{mk} f^{(m)}(X_{mk}) \right] < \infty .$$

Then Proposition 14.3.2 (ii) shows that C is f-geometrically regular for P and

$$\mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_C-1} \beta^k f(X_k) \right] < \infty .$$

Applying again Theorem 14.2.6 (a), we conclude that ξ is f-geometrically regular.
□

14.4 f-Geometric Regularity of the Split Kernel

Proposition 14.4.1 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let C be a $(1, 2\varepsilon\nu)$-small set with $\nu(C) = 1$ and $\inf_{x \in C} P(x, C) \geq 2\varepsilon$. Set $\check{P} = \check{P}_{\varepsilon,\nu}$. Let $f : \mathsf{X} \to [1, \infty)$ be a measurable function and r a positive sequence.*

(i) If C is f-geometrically regular for the kernel P, then $C \times \{0, 1\}$ is $\bar{f}$-geometrically regular for the kernel $\check{P}$, where $\bar{f}(x, d) = f(x)$ for all $x \in \mathsf{X}$ and $d \in (0, 1)$.

(ii) If the split chain $\check{P}$ is $\bar{f}$-geometrically regular and f is bounded on C, then P is f-geometrically regular.

Proof. (i) Let $A \in \mathscr{X}_P^+$. Since $\sum_{k=0}^{\sigma_{A\times\{0,1\}}-1} r(k) f(X_k) \in \mathscr{F}_\infty^X$, Proposition 11.1.2 shows that

$$\check{\mathbb{E}}_{\delta_x \otimes \mathrm{b}_\varepsilon} \left[\sum_{k=0}^{\sigma_{A\times\{0,1\}}-1} \delta^k f(X_k) \right] = \mathbb{E}_x \left[\sum_{k=0}^{\sigma_A-1} \delta^k f(X_k) \right] .$$

Since $\delta_x \otimes \mathrm{b}_\varepsilon = (1-\varepsilon)\delta_{(x,0)} + \varepsilon \delta_{(x,1)}$ for all $x \in \mathsf{X}$, this implies

$$\sup_{(x,d)\in C\times\{0,1\}} \check{\mathbb{E}}_{(x,d)}\left[\sum_{k=0}^{\sigma_{A\times\{0,1\}}-1} \delta^k \bar{f}(X_k,D_k)\right]$$
$$\leq \max(\varepsilon^{-1},(1-\varepsilon)^{-1}) \sup_{x\in C} \mathbb{E}_x\left[\sum_{k=0}^{\sigma_A-1}\delta^k f(X_k)\right] < \infty\,.$$

The proof follows easily.

(ii) If $\check{P}$ is $\bar{f}$-geometrically regular, then $\check{P}$ admits an invariant probability measure (see Theorem 14.2.6). By Proposition 11.1.3, this invariant probability is of the form $\pi\otimes \mathrm{b}_\varepsilon$, where π is an invariant probability for P. By Theorem 9.2.15, $\pi\otimes\mathrm{b}_\varepsilon$ is a maximal irreducibility measure for $\check{P}$, and π is a maximal irreducibility measure for P. Moreover, by Theorem 14.2.6, there exists an increasing sequence $\{\check{D}_n,\ n\in\mathbb{N}\}$ of $\bar{f}$-geometrically regular sets for $\check{P}$ such that $\bigcup_{n=0}^\infty \check{D}_n$ is full and absorbing.

We will now establish that there exists $D\subset C$ such that D is accessible (i.e., $\pi(D)>0$) and f-geometrically regular. Define $\check{F}_n = \check{D}_n\cap(C\times\{0,1\})$. For every $n\in\mathbb{N}$, the set $\check{F}_n$ is $\bar{f}$-geometrically regular for $\check{P}$. Furthermore, the sequence $\{\check{F}_n,\ n\in\mathbb{N}\}$ is increasing and

$$\pi\otimes\mathrm{b}_\varepsilon\left((C\times\{0,1\})\setminus\bigcup_{n=0}^\infty \check{F}_n\right) = \pi\otimes\mathrm{b}_\varepsilon\left((C\times\{0,1\})\cap\left\{\bigcup_{n=0}^\infty \check{D}_n\right\}^c\right) = 0\,, \tag{14.4.1}$$

where we have used that $\bigcup_{n=0}^\infty \check{D}_n$ is full and $\pi\otimes\mathrm{b}_\varepsilon$ is a maximal irreducibility measure. For $i\in\{0,1\}$ and every $n\in\mathbb{N}$, define

$$F_{n,i}\times\{i\} = \check{F}_n\cap(\mathsf{X}\times\{i\})\subset C\times\{i\}\,.$$

Obviously, we have $\check{F}_n = (F_{n,0}\times\{0\})\cup(F_{n,1}\times\{1\})$. Moreover, $\{F_{n,i}, n\in\mathbb{N}\}$, $i=0,1$, are two increasing sequences of sets in $\mathscr{X}$, and (14.4.1) shows that

$$\lim_{n\to\infty}\pi(C\setminus F_{n,0}) = \lim_{n\to\infty}\pi(C\setminus F_{n,1}) = 0,$$

which implies that $\lim_{n\to\infty}\pi(C\setminus(F_{n,0}\cap F_{n,1}))=0$.

Choose n large enough that the set $F_{n,0}\cap F_{n,1}$ is accessible and put $D=F_{n,0}\cap F_{n,1}$. By construction, $\pi(D)>0$, and therefore $D\times\{0,1\}$ is accessible for $\check{P}$. Moreover, $D\times\{0,1\}$ is $\bar{f}$-geometrically regular for $\check{P}$ (as a subset of a regular set). If $A\in\mathscr{X}_P^+$, then $A\times\{0,1\}$ is accessible for $\check{P}$, and thus for all $x\in D$, using Proposition 11.1.2,

$$\mathbb{E}_x\left[\sum_{k=0}^{\sigma_A-1}\delta^k f(X_k)\right] = \check{\mathbb{E}}_{\delta_x\otimes\mathrm{b}_\varepsilon}\left[\sum_{k=0}^{\sigma_{A\times\{0,1\}}-1}\delta^k f(X_k)\right]$$
$$\leq \sup_{(x,d)\in D\times\{0,1\}}\check{\mathbb{E}}_{(x,d)}\left[\sum_{k=0}^{\sigma_{A\times\{0,1\}}-1}\delta^k f(X_k)\right] < \infty\,.$$

This proves that D is f-geometrically regular for P.

□

14.5 Exercises

14.1. Assume that $\mathrm{D_g}(V,\lambda,b)$ holds. Then condition $\mathrm{D_g}(V,\lambda+b/d,b,C)$ holds with $C=\{V\le d\}$ if $\lambda+b/d<1$.

14.2. Let P be an irreducible Markov kernel on $\mathsf{X}\times\mathscr{X}$. Let C be an (f,δ)-regular petite set. Show that the set $\{W_C^{f,\delta}<\infty\}$ is full and absorbing and for every $d\ge 0$, the sets $\{W_C^{f,\delta}\le d\}$ are accessible for d large enough and petite.

14.3. Assume that there exist a measurable function $V:\mathsf{X}\to[1,\infty)$ and a set $C\in\mathscr{X}$ such that

$$PV\le\lambda V+b\mathbb{1}_C\ \text{,for some constants}\ \lambda\in[0,1)\ \text{and}\ b<\infty. \tag{14.5.1}$$

Show that the following hold:

1. For all $x\in\mathsf{X}$ such that $V(x)<\infty$, $\mathbb{P}_x(\sigma_C<\infty)=1$ and
$$\mathbb{E}_x[\lambda^{-\sigma_C}]\le\mathbb{E}_x\left[\lambda^{-\sigma_C}V(X_{\sigma_C})\right]\le V(x)+b\lambda^{-1}\mathbb{1}_C(x)\ . \tag{14.5.2}$$
2. For all $\delta\in(1,1/\lambda)$ and $x\in\mathsf{X}$,
$$\mathbb{E}_x[V_{\sigma_C}(X_{\sigma_C})\mathbb{1}\{\sigma_C<\infty\}]+(1-\delta\lambda)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1}\delta^kV(X_k)\right] \le V(x)+b\delta\mathbb{1}_C(x)\ . \tag{14.5.3}$$
3. If π is an invariant measure such that $\pi(\{V=\infty\})=0$, then $\mathbb{E}_\pi[\lambda^{-\sigma_C}]<\infty$.

14.4. An INAR (INteger AutoRegressive) process is a Galton–Walton process with immigration, defined by the recurrence $X_0=1$ and

$$X_{n+1}=\sum_{j=1}^{X_n}\xi_j^{(n+1)}+Y_{n+1}\ , \tag{14.5.4}$$

where $\{\xi_j^{(n)},\ j,n\in\mathbb{N}^*\}$ are i.i.d. integer-valued random variables and $\{Y_n,\ n\in\mathbb{N}^*\}$ is a sequence of i.i.d. integer-valued random variables, independent of $\{\xi_j^{(n)}\}$. The random variable Y_{n+1} represents the immigrants, that is, the part of the $(n+1)$th generation that does not descend from the nth generation.

Let ν be the distribution of ξ_1^1 and μ the distribution of Y_1.

1. Show that P, the transition kernel of this Markov chain, is given by

$$P(j,k) = \mu * \nu^{*j}(k)\ .$$

Set for $x \in \mathbb{N}$, $V(x) = x$.

2. Find $\lambda < 1$ and a finite set C such that $PV(x) \leq \lambda V(x) + b\mathbb{1}_C(x)$.
3. Show that $\sup_{x\in C} \mathbb{E}_x[\lambda^{-\sigma_C}] < \infty$.

14.5. Consider a functional autoregressive model $X_{k+1} = h(X_k) + Z_{k+1}$, where $h : \mathbb{R} \to \mathbb{R}$ is a measurable function, $\{Z_k,\ k \in \mathbb{N}^*\}$ is an i.i.d. sequence of integrable random variables, independent of X_0. We set $m = \mathbb{E}[|Z_1|]$ and assume the following:

(i) There exist $\ell > 0$ and $M < \infty$ such that $|h(x)| \leq |x| - \ell$ for all $|x| \geq M$.
(ii) There exist $\beta > 0$ and $K < \infty$ such that $K = \mathbb{E}\left[e^{\beta|Z_1|}\right] < \infty$ and $Ke^{-\beta\ell} = \lambda < 1$.
(iii) $\sup_{|x|\leq M} |h(x)| < \infty$.

Set $W(x) = e^{\beta|x|}$ and $C = [-M, +M]$.

1. Show that $PW(x) \leq Ke^{\beta|h(x)|}$.
2. Show that for $x \notin C$, $PW(x) \leq Ke^{-\beta\ell}W(x) = \lambda W(x)$.
3. Show that for $\sup_{x\in C}$, $PW(x) < \infty$.
4. Show that for all $x \in \mathbb{R}$, $\mathbb{E}_x[\lambda^{-\sigma_C}] < \infty$ and $\sup_{x\in C} \mathbb{E}_x[\lambda^{-\sigma_C}] < \infty$.

14.6 (ARCH(1) model). Consider the Markov chain defined on $\mathbb{R}$ by

$$X_k = \sqrt{\alpha_0 + \alpha_1 X_{k-1}^2}\, Z_k\ , \quad \alpha_0 > 0, \alpha_1 > 0,$$

where $\{Z_k,\ k \in \mathbb{N}\}$ is an independent sequence of real-valued random variables having a density with respect to Lebesgue measure denoted by g. Assume that there exists $s \in (0,1]$ such that $\alpha_1^{-s} > \mathbb{E}\left[Z_0^{2s}\right]$.

1. Write the Markov kernel P of this Markov chain.
2. Show that for all $s \in (0,1]$ and $\alpha \geq 0$, we have $(1+\alpha)^s \leq 1 + \alpha^s$. Deduce that for all $x, y > 0$, $(x+y)^s \leq x^s + y^s$.
3. Obtain a geometric drift condition using the function $V(x) = 1 + x^{2s}$.

14.7 (Random walk Metropolis algorithm on $\mathbb{R}$). In the random walk Metropolis algorithm on $\mathbb{R}$ (see Example 2.3.2), a candidate is drawn from the transition density $q(x,y) = \bar{q}(y-x)$, where $\bar{q}$ is a symmetric density $\bar{q}(y) = \bar{q}(-y)$, and is accepted with probability $\alpha(x,y)$ given by

$$\alpha(x,y) = \frac{\pi(y)}{\pi(x)} \wedge 1,$$

where π is the target density, which is assumed to be positive. Assume in addition that π is symmetric and log-concave in the tails, i.e., there exist $\beta > 0$ and some $x_0 > 0$ such that

$$\log \pi(x) - \log \pi(y) \geq \beta(y-x) , \qquad y \geq x \geq x_0 ,$$
$$\log \pi(x) - \log \pi(y) \geq \beta(x-y) , \qquad y \leq x \leq -x_0 .$$

1. Show that the Markov kernel P associated to the Metropolis algorithm is given, for $x \in \mathbb{R}$ and $A \in \mathscr{B}(\mathbb{R})$, by

$$P(x,A) = \int_A q(x,y)\alpha(x,y)\mathrm{d}y + \mathbb{1}_A(x) \int_{\mathbb{R}} q(x,y)\,[1-\alpha(x,y)]\,\mathrm{d}y.$$

2. Set $V(x) = \mathrm{e}^{s|x|}$ for all $s \in (0,\beta)$. Show that there exist a compact set $C \in \mathscr{B}(\mathbb{R})$ and constants $\lambda \in [0,1)$ and $b < \infty$ such that $PV \leq \lambda V + b\mathbb{1}_C$.
3. Show that there exists a constant $\delta > 1$ such that $\sup_{x\in C} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C-1} \delta^k V(X_k)\right] < \infty$.
4. Show that there exist constants $\delta > 1$ and $\varsigma < \infty$ such that for all $x \in \mathbb{R}$, $\mathbb{E}_x \left[\sum_{k=0}^{\sigma_C-1} \delta^k V(X_k)\right] \leq \kappa V(x)$.

14.8. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let $C \in \mathscr{X}$ be a nonempty set, $b \in [0,\infty)$, $f, V : \mathsf{X} \to [1,\infty]$ such that for all $x \in \mathsf{X}$,

$$f(x)PV(x) \leq V(x) + b\mathbb{1}_C(x) . \tag{14.5.5}$$

(i) Prove that

$$\mathbb{E}_x\left[\prod_{i=0}^{\sigma_C-1} f(X_i)\right] \leq V(x) + b\mathbb{1}_C(x) . \tag{14.5.6}$$

[Hint: $\pi_{-1} = 1$, $\pi_n = \prod_{i=0}^n f(X_i)$ and $V_n = V(X_n)$ for $n \geq 0$. Prove by induction using (14.5.5) that for all $n \geq 0$,

$$\mathbb{E}_x[\pi_{n\wedge\sigma_C-1} V_{n\wedge\sigma_C}] \leq V(x) + b\mathbb{1}_C(x) .]$$

(ii) Conversely, assume that $\sup_{x\in C} \mathbb{E}_x[\prod_{i=0}^{\sigma_C-1} f(X_i)] < \infty$ and set

$$V(x) = \mathbb{E}_x[\prod_{i=0}^{\tau_C-1} f(X_i)] .$$

Prove that there exists b such that (14.5.5) holds.

14.9. Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$, $f : \mathsf{X} \to [1,\infty)$ a measurable function, and $m \geq 1$. Assume that there exist a function $V : \mathsf{X} \to [0,\infty]$ such that $\{V < \infty\} \neq \emptyset$, a nonempty petite set C, $\lambda \in [0,1)$ and $b < \infty$ such that $PV + f \leq \lambda V + b\mathbb{1}_C$.

1. Show that

$$P^m V + \sum_{k=0}^{m-1} \lambda^{m-1-k} P^k f \leq \lambda^m V + b(1-\lambda^m)/(1-\lambda) . \tag{14.5.7}$$

2. Show that if P is aperiodic, then there exist a petite set D and $\lambda^{(m)} \in [0,1)$ and $b^{(m)} < \infty$ such that

$$P^m V^{(m)} + f^{(m)} \leq \lambda^{(m)} V^{(m)} + b^{(m)} \mathbb{1}_D \,, \tag{14.5.8}$$

where $f^{(m)}$ is defined in (14.3.1) and $V^{(m)} = \lambda^{-(m-1)} V$.
3. Using the drift condition (14.5.8), show that if P is f-regular, then P^m is $f^{(m)}$-regular.

14.6 Bibliographical Notes

The use of drift conditions to control the return times to a set was introduced for Markov chains over discrete state spaces by Foster (1953, 1952a). Early references evidencing the links between geometric drift conditions and regularity for discrete-state-space Markov chains include Kendall (1960), Vere-Jones (1962), Miller (1965/1966), Popov (1977), and Popov (1979). Extensions of these results to general state spaces was carried out in Nummelin and Tweedie (1976) and Nummelin and Tweedie (1978). The theory of geometric recurrence and geometric regularity is fully developed in the books of Nummelin (1984) and Meyn and Tweedie (1993b).

Chapter 15
Geometric Rates of Convergence

We have seen in Chapter 11 that a positive recurrent irreducible kernel P on $\mathsf{X} \times \mathscr{X}$ admits a unique invariant probability measure, say π. If the kernel is, moreover, aperiodic, then the iterates of the kernel $P^n(x,\cdot)$ converge to π in total variation distance for π-almost all $x \in \mathsf{X}$. Using the characterizations of Chapter 14, we will in this chapter establish conditions under which the rate of convergence is geometric in f-norm, i.e., $\lim_{n\to\infty} \delta^n \|P^n(x,\cdot) - \pi\|_f = 0$ for some $\delta > 1$ and positive measurable function f. We will also consider the related problems of finding nonasymptotic bounds of convergence, i.e., functions $M : \mathsf{X} \to \mathbb{R}_+$ such that for all $n \in \mathbb{N}$ and $x \in \mathscr{X}$, $\delta^n \|P^n(x,\cdot) - \pi\|_f \le M(x)$. We will provide different expressions for the bound $M(x)$ either in terms of (f,δ)-modulated moments of the return time to a small set $\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} \delta^k f(X_k)\right]$ or in terms of appropriately defined drift functions. We will also see the possible interplays between these different expressions of the bounds.

15.1 Geometric Ergodicity

Definition 15.1.1 (f-geometric ergodicity) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and let $f : \mathsf{X} \to [1,\infty)$ be a measurable function. The kernel P is said to be f-geometrically ergodic if it is irreducible, positive with invariant probability π, and if there exist*

(i) a measurable function $M : \mathsf{X} \to [0,\infty]$ such that $\pi(\{M < \infty\}) = 1$,
(ii) a measurable function $\beta : \mathsf{X} \to [1,\infty)$ such that $\pi(\{\beta > 1\}) = 1$

satisfying for all $n \in \mathbb{N}$ and $x \in \mathsf{X}$,

$$\beta^n(x)\, \|P^n(x,\cdot) - \pi\|_f \le M(x)\ .$$

If $f \equiv 1$, then P is simply said to be geometrically ergodic.

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_15

In Chapter 13, we considered atomic kernels and obtained rates of convergence of the iterates of the kernel to the invariant probability. In this section, we will extend these results to aperiodic irreducible kernels by means of the splitting construction. We use the same notation as in Section 11.1, in particular for the split kernel $\check{P}$, which can also be written $\check{P}_{\varepsilon,\nu}$ whenever there is ambiguity.

Lemma 15.1.2 *Let P be an irreducible Markov kernel on $\mathsf{X}\times\mathscr{X}$, $f:\mathsf{X}\to[1,\infty)$ a measurable function, and C a $(1,2\varepsilon\nu)$-small set with $\nu(C)=1$. Set $\check{P}=\check{P}_{\varepsilon,\nu}$ and $\check{\alpha}=C\times\{1\}$. Assume that for some $\delta>1$,*

$$\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1}\delta^k f(X_k)\right]<\infty\,. \tag{15.1.1}$$

Then there exist $\beta\in(1,\delta)$ and $\varsigma<\infty$ such that

$$\sup_{(x,d)\in C\times\{0,1\}}\check{\mathbb{E}}_{(x,d)}\left[\sum_{k=0}^{\sigma_{\check{\alpha}}}\beta^k f(X_k)\right]\le\varsigma\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1}\delta^k f(X_k)\right]\,, \tag{15.1.2}$$

and for all $\xi\in\mathbb{M}_1(\mathscr{X})$,

$$\check{\mathbb{E}}_{\xi\otimes\mathrm{b}_\varepsilon}\left[\sum_{k=0}^{\sigma_{\check{\alpha}}}\beta^k f(X_k)\right]\le\varsigma\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1}\delta^k f(X_k)\right]\,. \tag{15.1.3}$$

Proof. The condition (15.1.1) implies that $M=\sup_{x\in C}f(x)<\infty$ and $\inf_{x\in C}\mathbb{P}_x(\sigma_C<\infty)=1$, so that C is Harris recurrent for P. We define $\check{C}=C\times\{0,1\}$. Proposition 11.1.4 implies that for all $(x,d)\in\check{C}$, $\check{\mathbb{P}}_{(x,d)}(\sigma_{\check{C}}<\infty)=1$ and $\check{\mathbb{P}}_{(x,d)}(\sigma_{\check{\alpha}}<\infty)=1$. For $(x,d)\in\check{\mathsf{X}}$ such that $\check{\mathbb{P}}_{(x,d)}(\sigma_{\check{\alpha}}<\infty)=1$ and for all $\beta\in(0,1)$, we have

$$\check{\mathbb{E}}_{(x,d)}[\beta^{\sigma_{\check{\alpha}}}f(X_{\sigma_{\check{\alpha}}})]\le M\beta\check{\mathbb{E}}_{(x,d)}[\beta^{\sigma_{\check{\alpha}}-1}]\le M\beta\check{\mathbb{E}}_{(x,d)}\left[\sum_{k=0}^{\sigma_{\check{\alpha}}-1}\beta^k f(X_k)\right],$$

which implies that

$$\check{\mathbb{E}}_{(x,d)}\left[\sum_{k=0}^{\sigma_{\check{\alpha}}}\beta^k f(X_k)\right]\le(1+M\beta)\check{\mathbb{E}}_{(x,d)}\left[\sum_{k=0}^{\sigma_{\check{\alpha}}-1}\beta^k f(X_k)\right]\,. \tag{15.1.4}$$

On the other hand, for every $x\in C$, we have by Proposition 11.1.2,

$$\check{\mathbb{E}}_{\delta_x\otimes\mathrm{b}_\varepsilon}\left[\sum_{k=0}^{\sigma_{\check{C}}-1}\delta^k f(X_k)\right]=\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1}\delta^k f(X_k)\right]\,. \tag{15.1.5}$$

Note also that for every positive random variable Y,

$$\sup_{(x,D)\in\check{C}} \check{\mathbb{E}}_{(x,d)}[Y] \leq \varsigma_\varepsilon \sup_{x\in C} \check{\mathbb{E}}_{\delta_x\otimes \mathrm{b}_\varepsilon}[Y] \,,$$

with $\varsigma_\varepsilon = \varepsilon^{-1} \vee (1-\varepsilon)^{-1}$. Applying this bound to (15.1.5) and then using that $f \geq 1$ implies that $\sup_{(x,d)\in\check{C}} \check{\mathbb{E}}_{(x,d)}[\delta^{\sigma_{\check{C}}}] < \infty$.

By Proposition 11.1.4 (vi), we get $\inf_{(x,d)\in\check{C}} \check{\mathbb{P}}_{(x,d)}(X_1 \in \check{\alpha}) > 0$.

We may therefore apply Theorem 14.2.3 with $A = \check{C}$, $B = \check{\alpha}$, and $q = 1$, which shows there exist $\beta \in (1, \delta)$ and a finite constant ς_0 such that

$$\begin{aligned}
\sup_{(x,d)\in\check{C}} \check{\mathbb{E}}_{(x,d)} \left[\sum_{k=0}^{\sigma_{\check{\alpha}}-1} \beta^k f(X_k)\right] &\leq \varsigma_0 \sup_{(x,d)\in\check{C}} \check{\mathbb{E}}_{(x,d)} \left[\sum_{k=0}^{\sigma_{\check{C}}-1} \delta^k f(X_k)\right] \\
&\leq \varsigma_0 \varsigma_\varepsilon \sup_{x\in C} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C-1} \delta^k f(X_k)\right] .
\end{aligned}$$

Combining with (15.1.4) yields (15.1.2). Noting that $\sigma_{\check{\alpha}} \leq \sigma_{\check{C}} + \sigma_{\check{\alpha}} \circ \theta_{\sigma_{\check{C}}}$ on the event $\{\sigma_{\check{C}} < \infty\}$, we get

$$\begin{aligned}
\check{\mathbb{E}}_{\xi\otimes \mathrm{b}_\varepsilon} \left[\sum_{k=0}^{\sigma_{\check{\alpha}}} \delta^k f(X_k)\right] &\leq \check{\mathbb{E}}_{\xi\otimes \mathrm{b}_\varepsilon} \left[\sum_{k=0}^{\sigma_{\check{C}}-1} \delta^k f(X_k)\right] + \check{\mathbb{E}}_{\xi\otimes \mathrm{b}_\varepsilon} \left[\sum_{k=\sigma_{\check{C}}}^{\sigma_{\check{\alpha}}\circ\theta_{\sigma_{\check{C}}}} \delta^k f(X_k)\right] \\
&\leq \check{\mathbb{E}}_{\xi\otimes \mathrm{b}_\varepsilon} \left[\sum_{k=0}^{\sigma_{\check{C}}-1} \delta^k f(X_k)\right] + \check{\mathbb{E}}_{\xi\otimes \mathrm{b}_\varepsilon} [\delta^{\sigma_{\check{C}}}] \sup_{(x,d)\in\check{C}} \check{\mathbb{E}}_{(x,d)} \left[\sum_{k=0}^{\sigma_{\check{\alpha}}} \delta^k f(X_k)\right] \\
&= \mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_C-1} \delta^k f(X_k)\right] \left\{1 + \delta \sup_{(x,d)\in\check{C}} \check{\mathbb{E}}_{(x,d)} \left[\sum_{k=0}^{\sigma_{\check{\alpha}}} \delta^k f(X_k)\right]\right\} ,
\end{aligned} \tag{15.1.6}$$

which proves (15.1.3). □

We first provide sufficient conditions under which the Markov kernel P is f-geometrically ergodic.

Theorem 15.1.3. *Let P be an irreducible aperiodic Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $f : \mathsf{X} \to [1,\infty)$ a measurable function. Assume that P is f-geometrically regular, that is, one of the following equivalent conditions is satisfied (see Theorem 14.2.6):*

(i) There exists an f-geometrically recurrent petite set C, i.e., for some $\delta > 1$,

$$\sup_{x\in C} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C-1} \delta^k f(X_k)\right] < \infty \,. \tag{15.1.7}$$

(ii) There exist a function $V : \mathsf{X} \to [0,\infty]$ such that $\{V < \infty\} \neq \emptyset$, a nonempty petite set C, $\lambda \in [0,1)$, and $b < \infty$ such that

$$PV + f \leq \lambda V + b\mathbb{1}_C .$$

Then, denoting by π the invariant probability measure, the following properties hold:

(a) There exist a set $S \in \mathscr{X}$ such that $\pi(S) = 1$, $\{V < \infty\} \subset S$, with V as in (ii), and $\beta > 1$ such that for all $x \in S$,

$$\sum_{n=0}^{\infty} \beta^n \|P^n(x, \cdot) - \pi\|_f < \infty . \tag{15.1.8}$$

(b) For every f-geometrically regular distribution $\xi \in \mathbb{M}_1(\mathscr{X})$, there exists $\gamma > 1$ such that

$$\sum_{n=0}^{\infty} \gamma^n \|\xi P^n - \pi\|_f < \infty . \tag{15.1.9}$$

(c) There exist constants $\vartheta < \infty$ and $\beta > 1$ such that for all initial distributions $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{n=0}^{\infty} \beta^n \|\xi P^n - \pi\|_f \leq \vartheta M(\xi) \tag{15.1.10}$$

with $M(\xi) = \mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_C - 1} \delta^k f(X_k)\right]$ and δ as in (15.1.7) or $M(\xi) = \xi(V) + 1$ with V as in (ii).

Proof. Since C is small and $\sup_{x \in C} \mathbb{E}_x[\sigma_C] < \infty$, the existence and uniqueness of the invariant probability π follow from Corollary 11.2.9.

We assume (i) and will prove that there exist $\beta \in (1, \delta)$ and a finite constant $\varsigma < \infty$ such that for every $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{n=0}^{\infty} \beta^n \|\xi P^n - \pi\|_f \leq \varsigma \mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_C - 1} \delta^k f(X_k)\right] . \tag{15.1.11}$$

This is the central part of the proof. The other assertions follow almost immediately. The proof proceeds in two steps. We will first establish the result for a strongly aperiodic kernel and use for that purpose the splitting construction introduced in Chapter 11. We then extend the result to the general case using the m-skeleton.

(I) We first assume that P admits an f-geometrically recurrent and $(1, \mu)$-small set C with $\mu(C) > 0$. Since C is petite and f-geometrically recurrent, it is also accessible by Theorem 14.2.4. By Proposition 11.1.4, the set $\check{\alpha} = C \times \{1\}$ is an accessible, aperiodic, and positive atom for the split kernel $\check{P} = \check{P}_{\varepsilon,\nu}$ defined in (11.1.7). Using (15.1.2) in Lemma 15.1.2, the condition (15.1.7) implies that there exists $\gamma \in (1, \delta)$ such that $\check{\mathbb{E}}_{\check{\alpha}} \left[\sum_{k=0}^{\sigma_{\check{\alpha}}} \gamma^k f(X_k)\right] < \infty$. By Proposition 11.1.3, $\check{P}$ admits a unique invariant probability measure, which may be expressed as $\pi \otimes \mathrm{b}_\varepsilon$, where we recall that π is the unique invariant probability measure for P. In addition, Lemma 11.1.1 implies

$$\left\|\xi P^k - \pi\right\|_f \le \left\|(\xi \otimes \mathrm{b}_\varepsilon)\check{P}^k - \pi \otimes \mathrm{b}_\varepsilon\right\|_{f\otimes \mathbf{1}} . \tag{15.1.12}$$

Combining with Theorem 13.4.3 and (15.1.3) in Lemma 15.1.2, we obtain that there exist $\beta \in (1,\gamma)$ and $\varsigma_1,\ \varsigma_2 < \infty$ such that

$$\begin{aligned}\sum_{k=1}^{\infty} \beta^k \left\|\xi P^k - \pi\right\|_f &\le \sum_{k=1}^{\infty} \beta^k \left\|(\xi \otimes \mathrm{b}_\varepsilon)\check{P}^k - \pi \otimes \mathrm{b}_\varepsilon\right\|_{\bar{f}} \\ &\le \varsigma_1 \check{\mathbb{E}}_{\xi\otimes \mathrm{b}_\varepsilon}\left[\sum_{k=1}^{\sigma_{\check{\alpha}}} \gamma^k f(X_k)\right] \le \varsigma_1\varsigma_2 \mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} \delta^k f(X_k)\right] .\end{aligned}$$

(II) Assume now that P admits an f-geometrically recurrent petite set C. Applying Theorem 14.2.4, the set C is accessible. Moreover, since P is irreducible and aperiodic, the set C is also small by Theorem 9.4.10. Then by Lemma 9.1.6, we may assume without loss of generality that C is (m,μ)-small with $\mu(C) > 0$, and hence C is an accessible $(1,\mu)$-small set with $\mu(C) > 0$ for the kernel P^m. To apply (I), it remains to show that C is $f^{(m)}$-geometrically recurrent for the kernel P^m, where $f^{(m)} = \sum_{i=0}^{m-1} P^i f$. By Proposition 14.3.2, there exists $\gamma \in (1,\delta)$ and $\varsigma_1 < \infty$ such that for every $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_{C,m}-1} \gamma^{mk} f^{(m)}(X_{mk})\right] \le \varsigma_1 \mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} \delta^k f(X_k)\right] . \tag{15.1.13}$$

Using (15.1.7), this implies that $\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_{C,m}-1} \gamma^{mk} f^{(m)}(X_{mk})\right] < \infty$. We may therefore apply (I) to the kernel P^m to show that there exist $\beta \in (1,\delta)$ and $\varsigma_2 < \infty$ such that for all $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{k=1}^{\infty} \beta^k \left\|\xi P^{mk} - \pi\right\|_{f^{(m)}} \le \varsigma_2 \mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_{C,m}-1} \gamma^{mk} f^{(m)}(X_{mk})\right] . \tag{15.1.14}$$

To conclude, we need to relate $\sum_{k=1}^{\infty} \delta^k \left\|\xi P^k - \pi\right\|_f$ and $\sum_{k=1}^{\infty} \beta^k \left\|\xi P^{mk} - \pi\right\|_{f^{(m)}}$. This is not a difficult task. Note first that if $|g| \le f$, then for $i \in \{0,\ldots,m-1\}$, $|P^i g| \le f^{(m)}$, which implies

$$\left\|\xi P^{mk+i} - \pi\right\|_f = \sup_{|g|\le f} |\xi P^{mk+i} g - \pi(g)| \le \left\|\xi P^{mk} - \pi\right\|_{f^{(m)}} .$$

Therefore, we get

$$\begin{aligned}\sum_{k=0}^{\infty} \beta^{k/m} \left\|\xi P^k - \pi\right\|_f &\le \sum_{i=0}^{m-1}\sum_{\ell=0}^{\infty} \beta^{(\ell m+i)/m} \left\|\xi P^{mk} - \pi\right\|_{f^{(m)}} \\ &\le m\beta \sum_{\ell=0}^{\infty} \delta^\ell \left\|\xi P^{mk} - \pi\right\|_{f^{(m)}} .\end{aligned}$$

The bound (15.1.11) then follows from (15.1.13) and (15.1.14).

The rest of the proof is elementary, given all the previous results.

(a) The set $S_{\delta,C} := \left\{x \in \mathsf{X} : \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} \delta^k f(X_k)\right] < \infty\right\}$ is full and absorbing by Corollary 9.2.14. Since π is a maximal irreducibility measure, $\pi(S_{\delta,C}) = 1$. For all $x \in S_0$, (15.1.8) follows from (15.1.11) with $\xi = \delta_x$. If (ii) is satisfied, then we may choose the petite set C and the function V such that $\sup_C V < \infty$. By Theorem 14.2.6 (b), there exist a constant $\varsigma < \infty$ and $\delta > 1$ such that

$$\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} \delta^k f(X_k)\right] \le \varsigma\{V(x)+1\} . \tag{15.1.15}$$

Therefore, we get $\{V < \infty\} \subset S_{\delta,C}$, which concludes the proof of (a).

(b) If ξ is f-geometrically regular, then $\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} \kappa^k f(X_k)\right] < \infty$ for some $\kappa > 1$, and (15.1.9) follows from (15.1.11).

(c) If (i) is satisfied, then (15.1.11) gives the desired result. If (ii) is satisfied, the conclusion follows from (15.1.11) and (15.1.15).

□

Specializing Theorem 15.1.3 to the case $f \equiv 1$, we extend Theorem 8.2.9 to irreducible and aperiodic Markov chains.

Corollary 15.1.4 *Let P be an irreducible and aperiodic Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that there exists a geometrically recurrent small set C, i.e., $\sup_{x\in C} \mathbb{E}_x[\delta^{\sigma_C}] < \infty$ for some $\delta > 1$. Then P is geometrically ergodic with invariant probability π. In addition,*

(a) there exist $S \in \mathscr{X}$ with $\pi(S) = 1$ and $\beta > 1$ such that for all $x \in S$,

$$\sum_{k=1}^{\infty} \beta^k \left\| P^k(x,\cdot) - \pi \right\|_{\mathrm{TV}} < \infty ;$$

(b) there exist $\beta > 1$ and $\varsigma < \infty$ such that for every initial distribution $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{k=1}^{\infty} \beta^k \left\| \xi P^k - \pi \right\|_{\mathrm{TV}} \le \varsigma \mathbb{E}_\xi\left[\delta^{\sigma_C}\right] .$$

Proof. This follows directly from Theorem 15.1.3 on setting $f \equiv 1$. By Corollary 9.2.4, the set $\{x \in \mathsf{X} : \mathbb{E}_x[\delta^{\sigma_C}] < \infty\}$ is full and absorbing, which establishes the second assertion. □

If we set $f \equiv 1$, then the sufficient conditions of Theorem 15.1.3 for a Markov kernel P to be f-geometrically ergodic may be shown to be also necessary.

Theorem 15.1.5. *Let P be an irreducible, aperiodic, and positive Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. The following assertions are equivalent:*

(i) P is geometrically ergodic.
(ii) There exist a small set C and constants $\varsigma < \infty$ and $0 < \rho < 1$ such that for all $n \in \mathbb{N}$,

$$\sup_{x \in C} |P^n(x,C) - \pi(C)| \leq \varsigma \rho^n \,. \tag{15.1.16}$$

(iii) There exist an $(m, \varepsilon\nu)$-accessible small set C such that $\nu(C) > 0$ and constants $\varsigma < \infty$, $\rho \in [0,1)$ satisfying

$$\left| \int_C \nu(\mathrm{d}x)\{P^n(x,C) - \pi(C)\} \right| \leq \varsigma \rho^n \,.$$

(iv) There exist an accessible small set C and $\beta > 1$ such that $\sup_{x \in C} \mathbb{E}_x[\beta^{\sigma_C}] < \infty$.
(v) There exist $\rho < 1$ and a measurable function $M : \mathsf{X} \to [0,\infty]$ such that $\pi(M) < \infty$ and for all $x \in \mathsf{X}$ and $n \in \mathbb{N}$,

$$\|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} \leq M(x)\rho^n \,.$$

Proof. (i) $\Rightarrow$ (ii) If P is geometrically ergodic, then there exist measurable functions $M : \mathsf{X} \to [0,\infty]$ and $\rho : \mathsf{X} \to [0,1]$ satisfying $\pi(\{M < \infty\}) = \pi(\{\rho < 1\}) = 1$ such that for all $x \in \mathsf{X}$ and $n \in \mathbb{N}$,

$$\|\delta_x P^n - \pi\|_{\mathrm{TV}} \leq M(x)\rho^n(x) \,. \tag{15.1.17}$$

Since P is irreducible, it admits an accessible small set D. By Theorem 9.2.15, the invariant probability π is a maximal irreducibility measure; hence $\pi(D) > 0$. For $m > 0$ and $r \in [0,1)$, define the set

$$C(m,r) := D \cap \{x \in \mathsf{X} : M(x) \leq m\} \cap \{x \in \mathsf{X} : \rho(x) \leq r\} \,.$$

For every $m > 0$ and $r \in [0,1)$, the set $C(m,r)$ is a small set as a subset of a small set. Moreover, the set

$$\{x \in \mathsf{X} : M(x) < \infty\} \cap \{x \in \mathsf{X} : \rho(x) < 1\}$$

being full, m and r may be chosen large enough that $\pi(C(m,r)) > 0$. Then by (15.1.17), for all $x \in C(m,r)$, we have

$$|P^n(x, C(m,r)) - \pi(C(m,r))| \leq \|\delta_x P^n - \pi\|_{\mathrm{TV}} \leq m r^n \,.$$

(ii) $\Rightarrow$ (iii) Assume that C is an accessible (ℓ,μ)-small set satisfying (15.1.16). By Lemma 9.1.6, we may choose m and ν such that C is an $(m,\varepsilon\nu)$-small set with

$\varepsilon > 0$ and $\nu(C) > 0$. Moreover, we get

$$\left| \int_C \nu(\mathrm{d}x)\{P^n(x,C) - \pi(C)\} \right| \leq \nu(C) \sup_{x \in C} |P^n(x,C) - \pi(C)| \leq \nu(C) \varsigma \rho^n .$$

(iii) $\Rightarrow$ (iv) The proof is in two steps. We first assume the existence of a strongly aperiodic small set. We will then extend the result to general aperiodic kernels by considering a skeleton.

(I) Assume first that there exist a $(1, \varepsilon\nu)$ small set C satisfying $\nu(C) > 0$ and constants $\varsigma < \infty$ and $\rho \in [0,1)$ such that $|\nu P^n(C) - \pi(C)| \leq \varsigma \rho^n$. Consider the split kernel $\check{P}$ introduced in Section 11.1. Set $\check{\alpha} = C \times \{1\}$. By Proposition 11.1.4 (ii), $\check{\alpha}$ is an accessible atom for $\check{P}$. By Proposition 11.1.4 (iv), we have for all $n \geq 1$, $\check{P}^n(\check{\alpha}, \check{\alpha}) = \varepsilon \nu P^{n-1}(C)$. Therefore, for every $z \in \mathbb{C}$ such that $|z| \leq \rho^{-1}$, the series

$$\sum_{n=1}^{\infty} \{\check{P}^n(\check{\alpha}, \check{\alpha}) - \varepsilon \pi(C)\} z^n = \varepsilon \sum_{n=1}^{\infty} \{\nu P^{n-1}(C) - \pi(C)\} z^n < \infty \tag{15.1.18}$$

is absolutely convergent. Kendall's theorem, Theorem 8.1.9, shows that (15.1.18) is equivalent to the existence of an exponential moment for the return time to the atom $\check{\alpha}$, i.e., there exists $\delta > 1$ such that

$$\check{\mathbb{E}}_{\check{\alpha}}[\delta^{\sigma_{\check{\alpha}}}] < \infty .$$

Since by Proposition 11.1.4 the set $\check{\alpha}$ is accessible for $\check{P}$, the kernel $\check{P}$ is geometrically regular by Theorem 14.2.6. The kernel P is therefore geometrically regular by Proposition 14.4.1 (ii). Hence the kernel P admits an accessible geometrically regular set D (see Definition 14.2.1). By Theorem 14.2.4, the set D is petite and geometrically recurrent, i.e., $\sup_{x \in D} \mathbb{E}_x[\beta^{\sigma_D}] < \infty$ for some $\beta > 1$, hence accessible by Theorem 14.2.4. Furthermore, since P is aperiodic, every petite set is small by Theorem 9.4.10.

(II) Assume now that C is an accessible $(m, \varepsilon\nu)$-small set for some $m > 1$ and $\nu(C) > 0$. Without loss of generality, we may assume that $\nu \in \mathbb{M}_1(\mathscr{X})$, $\nu(C) = 1$, and $\inf_{x \in C} P^m(x,C) \geq 2\varepsilon$. Applying (I), there exists an accessible small set D such that $\sup_{x \in D} \mathbb{E}_x[\delta^{\sigma_{D,m}}] < \infty$ for some $\delta > 1$, where $\sigma_{D,m}$ is the return time to C for the skeleton chain P^m. The set D is also small for the kernel P, and since $\sigma_D \leq m\sigma_{D,m}$, it follows that $\sup_{x \in D} \mathbb{E}_x\left[\delta^{\sigma_D/m}\right] \leq \sup_{x \in D} \mathbb{E}_x[\delta^{\sigma_{D,m}}] < \infty$. Hence, the Markov kernel P admits a small accessible geometrically recurrent set.

The rest of the proof is immediate. [(iv) $\Rightarrow$ (v)] follows from Corollary 15.1.4 on choosing $M(x) = \mathbb{E}_x[\delta^{\tau_C}]$ for an appropriate $\delta > 1$. [(v) $\Rightarrow$ (i)] is obvious.

$\square$

Theorem 15.1.6. *Let P be an irreducible and positive Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. Assume that P is geometrically ergodic. Then for every $p \geq 1$, there exist a function $V : \mathsf{X} \to [1, \infty]$ $\kappa \in [1, \infty)$ and $\varsigma < \infty$ such that $\pi(V^p) < \infty$ and for all $n \in \mathbb{N}$ and $x \in \mathsf{X}$,*

$$\|P^n(x, \cdot) - \pi\|_{\mathrm{TV}} \leq \|P^n(x, \cdot) - \pi\|_{V} \leq \varsigma V(x) \kappa^{-n} .$$

Proof. By Lemma 9.3.9, the Markov kernel P is aperiodic. By Theorem 15.1.5 (iv), there exist an accessible small set C and $\beta > 1$ such that $\sup_{x \in C} \mathbb{E}_x[\beta^{\sigma_C}] < \infty$. For $\delta \in (1, \beta]$ and $x \in \mathsf{X}$, we set $V_\delta(x) = \mathbb{E}_x[\delta^{\tau_C}]$. We have $PV_\delta(x) = \delta^{-1} \mathbb{E}_x[\delta^{\sigma_C}]$ and therefore

$$PV_\delta(x) \leq \delta^{-1} V_\delta(x) + b_\delta \mathbb{1}_C(x) \quad \text{ where } b_\delta = \delta^{-1} \sup_{x \in C} \mathbb{E}_x[\delta^{\sigma_C}]. \tag{15.1.19}$$

By Corollary 9.2.14, the set $\{x \in \mathsf{X} : \mathbb{E}_x[\beta^{\sigma_C}] < \infty\}$ is full and absorbing. Since π is a maximal irreducibility measure, $\pi(\{V_\beta = \infty\}) = 0$, and thanks to (15.1.19), by Lemma 14.1.10, $\pi(V_\beta) < \infty$. Set $\delta = \beta^{1/p}$. Applying Jensen's inequality, we get for $p \geq 1$,

$$\pi(V_\delta^p) = \int \{\mathbb{E}_x[\delta^{\tau_C}]\}^p \pi(\mathrm{d}x) \leq \int \mathbb{E}_x[\delta^{p\tau_C}] \pi(\mathrm{d}x) = \pi(V_\beta) < \infty .$$

Let $\alpha > \delta$ and set $\gamma^{-1} = \delta^{-1} - \alpha^{-1}$, $V = V_\delta \mathbb{1}_{\{V_\delta < \infty\}} + \mathbb{1}_{\{V_\delta = \infty\}}$, and $W_\delta = \gamma V_\delta$. Note that $V \geq 1$ and $W_\delta \leq \gamma V$ π-*a.e.* We get, using (15.1.19), for all $\in \mathsf{X}$,

$$PW_\delta(x) + V(x) \leq \alpha^{-1} W_\delta(x) + \gamma b_\delta \mathbb{1}_C(x) .$$

We apply Theorem 15.1.3 (c) (with $f \leftarrow V$, $V \leftarrow W_\delta$); there exist $\kappa > 1$ and $\varsigma_1 < \infty$ such that

$$\|P^n(x, \cdot) - \pi\|_{V} \leq \varsigma_1 \kappa^{-n} \{W_\delta(x) + 1\} \leq \varsigma_2 \kappa^{-n} V(x) \quad \pi - a.e.$$

□

Theorem 15.1.5 may be used to establish that some seemingly innocuously Markov kernels P may fail to be geometrically ergodic.

Example 15.1.7. Let P be a positive Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. Assume that the invariant probability π is not concentrated at a single point and that the essential supremum of the function $x \mapsto P(x, \{x\})$ with respect to π is equal to 1:

$$\operatorname{esssup}_\pi(P) = \inf\{\delta > 0 : \pi(\{x \in \mathsf{X} : P(x, \{x\}) \geq \delta\}) = 0\} = 1 .$$

We will prove by contradiction that the Markov kernel P cannot be geometrically ergodic. Assume that the Markov kernel P is geometrically ergodic. From Theorem

15.1.5 (iv), there exist an $(m, \varepsilon\nu)$-small set C and $\beta > 1$ such that

$$\sup_{x\in C} \mathbb{E}_x \left[\beta^{\sigma_C}\right] < \infty \,. \tag{15.1.20}$$

Because the stationary distribution is not concentrated at a point and P is irreducible,

$$\text{for all } x \in \mathsf{X}, \quad P(x, \{x\}) < 1 \,. \tag{15.1.21}$$

(Recall that π is a maximal irreducibility measure; hence if there exists $x \in \mathsf{X}$ such that $P(x, \{x\}) = 1$, then the set $\{x\}$ is absorbing and hence full.) Because C is $(m, \varepsilon\nu)$-small, we may write for all $x \in C$,

$$P^m(x, \cdot) = \varepsilon\nu + (1-\varepsilon)R(x, \cdot) \,, \quad R(x, \cdot) = (1-\varepsilon)^{-1}\left\{P^m(x, \cdot) - \varepsilon\nu\right\} .$$

Hence for all $x, x' \in C$, we have $P^m(x, \cdot) - P^m(x', \cdot) = (1-\varepsilon)\{R(x, \cdot) - R(x', \cdot)\}$, which implies

$$\left\|P^m(x, \cdot) - P^m(x', \cdot)\right\|_{\mathrm{TV}} \le 2(1-\varepsilon) \,. \tag{15.1.22}$$

For $j \ge 1$, denote by A_j the set $A_j := \left\{x \in \mathsf{X} : P(x, \{x\}) \ge 1 - j^{-1}\right\}$; under the stated assumption, $\pi(A_j) > 0$ for all $j \ge 1$. We will show that

$$\sup_{x\in C} P(x, \{x\}) < 1 \,, \tag{15.1.23}$$

which implies that for large enough $j \ge 1$, we must have have $A_j \cap C = \emptyset$.

The proof of (15.1.23) is also by contradiction. Assume that $\sup_{x\in C} P(x, \{x\}) = 1$. Since $P(x, \{x\}) < 1$ (see (15.1.21)), there must be two distinct points x_0 and $x_1 \in C$ satisfying $P(x_i, \{x_i\}) > (1-\varepsilon/2)^{1/m}$, or equivalently $P^m(x_i, \{x_i\}) > (1-\varepsilon/2)$, $i = 0, 1$. By Proposition D.2.3, we have

$$\|P^m(x_0, \cdot) - P^m(x_1, \cdot)\|_{\mathrm{TV}} = \sup \sum_{i=0}^{I} |P^m(x_0, B_i) - P^m(x_1, B_i)| \,,$$

where the supremum is taken over all finite measurable partitions $\{B_i\}_{i=0}^{I}$. Taking $B_0 = \{x_0\}$, $B_1 = \{x_1\}$, and $B_2 = \mathsf{X} \setminus (B_0 \cup B_1)$, we therefore have

$$\begin{aligned}
&\|P^m(x_0, \cdot) - P^m(x_1, \cdot)\|_{\mathrm{TV}} \\
&\quad \ge |P^m(x_0, \{x_0\}) - P^m(x_1, \{x_0\})| + |P^m(x_0, \{x_1\}) - P^m(x_1, \{x_1\})| \ge 2(1-\varepsilon) \,,
\end{aligned}$$

where we have used $P^m(x_i, \{x_i\}) > (1-\varepsilon/2)$, $i = 0, 1$, and $P^m(x_i, \{x_j\}) < \varepsilon/2$, $i \ne j \in \{0, 1\}$. This gives a contradiction to (15.1.22). Hence $\sup_{x\in C} P(x, \{x\}) < 1$ and $A_j \cap C = \emptyset$ for large enough j.

Choose j large enough that $1 - j^{-1} > \beta^{-1}$, where β is defined in (15.1.20) and $A_j \cap C = \emptyset$. By Theorem 9.2.15, π is a maximal irreducibility measure. Since $\pi(A_j) > 0$, then A_j is accessible. Let $x \in C$. Then there exists an integer n such that $P^n(x, A_j) > 0$. By the last exit decomposition from C (see Section 3.4), we get that

$$P^n(x,A_j) = \mathbb{E}_x[\mathbb{1}_{A_j}(X_n)\mathbb{1}\{\sigma_C \geq n\}] \\ + \sum_{i=1}^{n-1} \int_C P^i(x,\mathrm{d}x')\mathbb{E}_{x'}[\mathbb{1}_{A_j}(X_{n-i})\mathbb{1}\{\sigma_C \geq n-i\}] > 0 .$$

Therefore, there exist $x_0 \in C$ and $\ell \in \{1,\ldots,n\}$ such that $\mathbb{E}_{x_0}[\mathbb{1}_{A_j}(X_\ell)\mathbb{1}\{\sigma_C \geq \ell\}] > 0$. For all $k \geq 0$, we have, using that $A_j \cap C = \emptyset$,

$$\mathbb{P}_{x_0}(\sigma_C \geq \ell + k) \geq \mathbb{E}_{x_0}[\mathbb{1}_{A_j}(X_\ell)\mathbb{1}\{\sigma_C \geq \ell\}](1-j^{-1})^k ,$$

giving the contradiction that $\mathbb{E}_{x_0}[\beta^{\sigma_C}] = \infty$. Therefore, P cannot be geometrically ergodic.

15.2 V-Uniform Geometric Ergodicity

Definition 15.2.1 (V-uniform geometric ergodicity) *Let $V : \mathsf{X} \to [1,\infty)$ be a measurable function and P a Markov kernel on $\mathsf{X} \times \mathscr{X}$.*

(i) The Markov kernel P is said to be V-uniformly ergodic if P admits an invariant probability measure π such that $\pi(V) < \infty$ and there exists a nonnegative sequence $\{\varsigma_n,\ n \in \mathbb{N}\}$ such that $\lim_{n\to\infty} \varsigma_n = 0$ and for all $x \in \mathsf{X}$,

$$\|P^n(x,\cdot) - \pi\|_V \leq \varsigma_n V(x) . \tag{15.2.1}$$

(ii) The Markov kernel P is said to be V-uniformly geometrically ergodic if P is V-uniformly ergodic and there exist constants $\varsigma < \infty$ and $\beta > 1$ such that for all $n \in \mathbb{N}$, $\varsigma_n \leq \varsigma\beta^{-n}$.

(iii) If $V \equiv 1$, the Markov kernel P is said to be uniformly (geometrically) ergodic.

Lemma 15.2.2 *Let $V : \mathsf{X} \to [1,\infty)$ be a measurable function. Let P be a positive Markov kernel on $\mathsf{X} \times \mathscr{X}$ with stationary distribution π satisfying $\pi(V) < \infty$. Assume that there exists a sequence $\{\zeta_k,\ k \in \mathbb{N}\}$ such that for all $x \in \mathsf{X}$, $\left\|P^k(x,\cdot) - \pi\right\|_V \leq \zeta_k V(x)$. Then for all $n,m \in \mathbb{N}$ and $x \in \mathsf{X}$, $\|P^{n+m}(x,\cdot) - \pi\|_V \leq \zeta_n \zeta_m V(x)$.*

Proof. Since P is V-uniformly ergodic, it follows that for all $k \in \mathbb{N}$, $x \in \mathsf{X}$, and every measurable function satisfying $\sup_{y\in\mathsf{X}} |f(y)|/V(y) < \infty$ and $x \in \mathsf{X}$, we get

$$|\delta_x P^k(f) - \pi(f)| \leq \{\sup_{y\in\mathsf{X}} |f(y)|/V(y)\}\zeta_k V(x) .$$

For all $n,m \in \mathbb{N}$, we have

$$P^{n+m} - \mathbf{1} \otimes \pi = (P^n - \mathbf{1} \otimes \pi)(P^m - \mathbf{1} \otimes \pi) .$$

Furthermore, it is also the case that $|P^m f(y) - \pi(f)| \leq \zeta_m V(y)$ for all $y \in \mathsf{X}$. Thus we get, for all $f \in \mathbb{F}(\mathsf{X})$ such that $\sup_{y \in \mathsf{X}} |f(y)|/V(y) \leq 1$ and all $x \in \mathsf{X}$,

$$\delta_x P^{n+m}(f) - \pi(f) = \delta_x [P^n - \mathbf{1} \otimes \pi][(P^m - \mathbf{1} \otimes \pi)(f)] \leq \zeta_n \zeta_m V(x) .$$

□

Proposition 15.2.3 *Let $V : \mathsf{X} \to [1, \infty)$ be a measurable function. Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. The Markov kernel P is V-uniformly ergodic if and only if P is V-uniformly geometrically ergodic.*

Proof. Assume that P is V-uniformly ergodic. Denote by π the invariant probability. There exists a sequence $\{\varsigma_n, n \in \mathbb{N}\}$ such that $\lim_{n \to \infty} \varsigma_n = 0$ and for all $x \in \mathsf{X}$, $\|\delta_x P^n - \pi\|_V \leq \varsigma_n V(x)$. Let $\beta > 1$ and choose $m \in \mathbb{N}$ such that $\varsigma_m = \beta^{-1} < 1$. Applying Lemma 15.2.2 with $\zeta_{km} = \beta^{-k}$, we get that $\left\|\delta_x P^{km} - \pi\right\|_V \leq \beta^{-k} V(x)$ for all $x \in \mathsf{X}$. Let $n \in \mathbb{N}$. We have $n = km + r$, $r < m$. Then using again Lemma 15.2.2 and setting $\varsigma = \max_{1 \leq j < m} \varsigma_j$, we get for all $x \in \mathsf{X}$,

$$\|\delta_x P^n - \pi\|_V = \left\|\delta_x P^{mk+r} - \pi\right\|_V \leq \varsigma \beta^{-k} V(x) \leq \varsigma \beta \beta^{-n/m} V(x) ,$$

showing that P is V-uniformly geometrically ergodic. The converse implication is obvious. □

We now state equivalences that parallel the results of Chapter 14.

Theorem 15.2.4. *Let $V : \mathsf{X} \to [1, \infty)$ be a measurable function. Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Then the following conditions are equivalent:*

(i) P is V-uniformly geometrically ergodic.
(ii) P is positive and aperiodic, and there exist $\varsigma < \infty$, $\beta > 1$, and a petite set C such that $\sup_{x \in C} V(x) < \infty$ and for all $x \in \mathsf{X}$,

$$\mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} \beta^k V(X_k) \right] \leq \varsigma V(x) .$$

Moreover, the following properties hold:

(a) If P is aperiodic and condition $\mathrm{D_g}(V, \lambda, b, C)$ holds for some petite set C, then P is V-uniformly geometrically ergodic.

(b) If P is V-uniformly geometrically ergodic, then P is positive and aperiodic, and condition $\mathrm{D_g}(V_0, \lambda, b, C)$ is satisfied for some petite set C and some function V_0 satisfying $V \le V_0 \le \varsigma V$ and constants $\varsigma < \infty$, $b < \infty$, $\lambda \in [0,1)$.

Proof. (I) Assume that P is aperiodic and that the condition $\mathrm{D_g}(V, \lambda, b, C)$ holds for some petite set C. We will first prove that (ii) is satisfied. By Corollary 14.1.6, we may assume without loss of generality that V is bounded on C. We may choose $\varepsilon > 0$ and $\tilde{\lambda} \in [0,1)$ such that

$$PV + \varepsilon V \le \tilde{\lambda} V + b \mathbb{1}_C \,.$$

By Proposition 14.1.3, the set C is $(V, \tilde{\lambda}^{-1})$-geometrically recurrent. By Proposition 14.1.2, for all $x \in \mathsf{X}$,

$$\begin{aligned} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} \tilde{\lambda}^{-k} V(X_k) \right] &\le \varepsilon^{-1} \left\{ \sup_C V + b\tilde{\lambda}^{-1} \right\} \mathbb{1}_C(x) + \varepsilon^{-1} V(x) \mathbb{1}_{C^c}(x) \\ &\le \left\{ \varepsilon^{-1} \left\{ \sup_C V + b\tilde{\lambda}^{-1} \right\} + \varepsilon^{-1} \right\} \varsigma V(x) \,. \end{aligned}$$

Therefore, condition (ii) is satisfied.

(II) We will now establish (b). Since P is V-uniformly geometrically ergodic, P admits an invariant probability measure π satisfying $\pi(V) < \infty$ and there exist $\rho < 1$ and $M < \infty$ such that for all $n \in \mathbb{N}$ and $x \in \mathsf{X}$, $\|P^n(x,\cdot) - \pi\|_V \le M\rho^n V(x)$. Then for all $A \in \mathscr{X}$ and $x \in \mathsf{X}$, we get

$$|P^n(x,A) - \pi(A)| \le \|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} \le \|P^n(x,\cdot) - \pi\|_V \le M\rho^n V(x) \,. \tag{15.2.2}$$

For all $A \in \mathscr{X}$ and $x \in \mathsf{X}$, we therefore have

$$P^n(x,A) \ge \pi(A) - M\rho^n V(x). \tag{15.2.3}$$

If $\pi(A) > 0$, we may therefore choose n large enough that $P^n(x,A) > 0$, showing that P is irreducible and π is an irreducibility measure. Since π is invariant for P, Theorem 9.2.15 shows that π is a maximal irreducibility measure.

Let C be an accessible small set. Since π is a maximal irreducibility measure, $\pi(C) > 0$, and for all d and for all $x \in \mathsf{X}$ satisfying $V(x) \le d$, we may choose n large enough that

$$P^n(x,C) \ge \pi(C) - M\rho^n d \ge 1/2\pi(C) \,.$$

Therefore, $\inf_{x \in \{V \le d\}} P^n(x,C) > 0$, and since C is a small set, $\{V \le d\}$ is also a small set by Lemma 9.1.7.

Since $\mathsf{X} = \{V < \infty\}$ and $\pi(\mathsf{X}) = 1$, we may choose d_0 large enough that $\pi(\{V \le d\}) > 0$ for all $d \ge d_0$. Since π is a maximal irreducibility measure, for all $d \ge d_0$, $\{V \le d\}$ is an accessible small set. Applying (15.2.3) with $D = \{V \le d\}$, we may find n large enough that $\inf_{x \in D} P^m(x,D) \geq \pi(D)/2 > 0$ for all $m > n$. This implies that the period of D is equal to 1 and hence that P is aperiodic.

Equation 15.2.2 also implies that for all $x \in \mathsf{X}$ and $k \in \mathbb{N}$,

$$P^k V(x) \leq M \rho^k V(x) + \pi(V) . \tag{15.2.4}$$

We may therefore choose m large enough that $M\rho^m \leq \lambda < 1$, so that P^m satisfies the condition $\mathrm{D_g}(V, \lambda, \pi(V))$. Hence by Proposition 14.1.8, P satisfies the condition $\mathrm{D_g}(V_0, \lambda^{1/m}, \lambda^{-(m-1)/m}\pi(V))$, where

$$V_0 = \sum_{k=0}^{m-1} \lambda^{-k/m} P^k V .$$

Clearly, for all $x \in \mathsf{X}$, $V(x) \leq V_0(x)$. On the other hand, by (15.2.4), for all $x \in \mathsf{X}$, we get

$$V_0(x) \leq \left\{ M \sum_{k=0}^{m-1} \lambda^{-k/m} \rho^k \right\} V(x) + \sum_{k=0}^{m-1} \pi(V) \lambda^{-k/m} ,$$

showing that $V_0(x) \leq \varsigma V(x)$, with $\varsigma = M \sum_{k=0}^{m-1} \lambda^{-k/m} \{\rho^k + \pi(V)\}$. For all $d > 0$, the set $\{V \leq d\}$ is petite, and therefore $\{V_0 \leq d\}$ is also petite for all d. We conclude by applying Corollary 14.1.6.

(III) We will show that (i) $\Rightarrow$ (ii). Assume that P is V-uniformly geometrically ergodic. Then (II) shows that (b) is satisfied, i.e., P is positive and aperiodic, and the drift condition $\mathrm{D_g}(V_0, \lambda, b, C)$ is satisfied for some petite set C and some function $V \leq V_0 \leq \varsigma V$ and $\sup_{x \in C} V(x) < \infty$. Condition (iii) follows from (I).

(IV) We will show that (ii) $\Rightarrow$ (i). Since $\sup_C V < \infty$, the set C is petite and V-geometrically recurrent. Since P is aperiodic, we may apply Theorem 15.1.3, which shows that there exist constants $\rho \in (1, \beta)$ and $M < \infty$ such that

$$\rho^k \left\| P^k(x, \cdot) - \pi \right\|_V \leq M \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} \beta^k V(X_k) \right] \leq M \varsigma V(x) ,$$

showing that P is V-uniformly geometrically ergodic.

(V) We will finally prove (a). By (I), we already have that (iii) is satisfied. Since (ii) $\Rightarrow$ (i), this shows that the Markov kernel P is V-uniformly geometrically ergodic, which is (a).

□

Example 15.2.5 (Example 11.4.3 (continued)). We consider in this example the first-order functional autoregressive model studied in Example 11.4.3. We recall briefly the results obtained in that example. The first-order functional autoregressive model on $\mathbb{R}^d$ is defined iteratively by $X_k = m(X_{k-1}) + Z_k$, where $\{Z_k,\ k \in \mathbb{N}\}$ is an i.i.d. sequence of random vectors independent of X_0 and $m : \mathbb{R}^d \to \mathbb{R}^d$ is a locally bounded measurable function satisfying $\limsup_{|x| \to \infty} |m(x)|/|x| < 1$. We assume that the distribution of Z_0 has a density q with respect to Lebesgue measure on $\mathbb{R}^d$ that is bounded away from zero on every compact set and that $\mathbb{E}[|Z_0|] < \infty$.

Under these assumptions, we have shown that every compact set is $(1, \varepsilon\nu)$-small and thus strongly aperiodic. In addition, we have shown, setting $V(x) = 1 + |x|$, that $PV(x) \le \lambda V(x) + b$ for all $\lambda \in \left(\limsup_{|x|\to\infty} |m(x)|/|x|, 1\right)$. Hence by applying Theorem 15.2.4, the Markov kernel P is V-uniformly geometrically ergodic, i.e., there exist a unique stationary distribution π, $\beta > 1$, and $\varsigma < \infty$ such that for all $x \in \mathbb{R}^d$,

$$\beta^n \|P^n(x,\cdot) - \pi\|_V \le \kappa V(x) .$$

Example 15.2.6 (Random walk Metropolis algorithm). We again consider the random walk Metropolis algorithm over the real line. We briefly summarize the models and the main results obtained so far. Let h_π be a positive and continuous density function over $\mathbb{R}$ that is log-concave in the tails; see (14.1.12), (14.1.3). Let $\bar{q}$ be a continuous, positive, and symmetric density on $\mathbb{R}$. We denote by P the Markov kernel associated to the random walk Metropolis (RWM) algorithm (see Example 2.3.2) with increment distribution $\bar{q}$. We have established that P is irreducible and that every compact set $C \subset \mathbb{R}$ such that $\mathrm{Leb}(C) > 0$ is $(1, \varepsilon\nu)$-small. We have also established that $\mathrm{D_g}(V, \lambda, b, C)$ holds with $V(x) = \mathrm{e}^{s|x|}$ and $C = [-x_*, x_*]$. Hence by applying Theorem 15.2.4, the random walk Metropolis–Hastings kernel P is V-uniformly geometrically ergodic, i.e., there exist $\beta > 1$ and $\varsigma < \infty$ such that for all $x \in \mathbb{R}^d$,

$$\beta^n \|P^n(x,\cdot) - \pi\|_V \le \kappa V(x) ,$$

where π is the target distribution. ◀

15.3 Uniform Ergodicity

We now specialize the results above to the case that the Markov kernel P is uniformly geometrically ergodic. Recall that P is uniformly geometrically ergodic if it admits an invariant probability π and if there exist $\beta \in (1, \infty]$ and a $\varsigma < \infty$ such that $\sup_{x\in\mathsf{X}} \|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} \le \varsigma\beta^{-n}$ for all $n \in \mathbb{N}$. We already know from Proposition 15.2.3 that uniform geometric ergodicity is equivalent to the apparently weaker uniform ergodicity that states that $\lim_{n\to\infty} \sup_{x\in\mathsf{X}} \|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} = 0$.

Most of the results that we obtained immediately translate to this case by simply setting $V \equiv 1$. Nevertheless, uniform ergodicity remains a remarkable property. This is linked to the fact that the convergence of the iterates of the Markov kernel to the stationary distribution π does not depend on the initial distribution.

When a Markov kernel is uniformly ergodic, there are many properties that hold uniformly over the whole space. It turns out that these properties are in fact equivalent to uniform ergodicity. This provides many criteria for checking uniform ergodicity, which we will use to give conditions for uniform ergodicity of a Markov kernel and to give conditions for nonuniform ergodicity.

Theorem 15.3.1. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. The following statements are equivalent:*

(i) P is uniformly geometrically ergodic.

(ii) P is a positive aperiodic Markov kernel, and there exist a small set C and $\beta > 1$ such that

$$\sup_{x \in \mathsf{X}} \mathbb{E}_x[\beta^{\sigma_C}] < \infty .$$

(iii) The state space X is small.

(iv) P is a positive aperiodic Markov kernel, and there exist a bounded function $V : \mathsf{X} \to [1, \infty)$, a petite set C, and constants $\lambda \in [0,1)$, $b < \infty$ such that the condition $\mathrm{D_g}(V, \lambda, b, C)$ is satisfied.

Proof. (i) $\Rightarrow$ (ii) Follows from Theorem 15.2.4 (ii) with $V \equiv 1$.

(ii) $\Rightarrow$ (iii) Lemma 9.4.8 shows that the set $\{x \in \mathsf{X} : \mathbb{E}_x[\beta^{\tau_C}] \le d\}$ is petite. Since $\sup_{x \in \mathsf{X}} \mathbb{E}_x[\beta^{\tau_C}] < \infty$, the state space X is petite and hence small, since P is aperiodic.

(iii) $\Rightarrow$ (i) Assume that the state space X is $(m, \varepsilon\nu)$-small, i.e., for all $x \in \mathsf{X}$ and $A \in \mathscr{X}$, $P^m(x,A) \ge \varepsilon\nu(A)$. Hence for all $A \in \mathscr{X}$ such that $\nu(A) > 0$, one has $P^m(x,A) > 0$, which shows that P is irreducible and ν is an irreducibility measure. Since X is an accessible small set and $\sigma_{\mathsf{X}} = 1 \mathbb{P}_x -$a.s. for all $x \in \mathsf{X}$, Theorem 10.1.2 shows that P is recurrent. By applying Theorem 11.2.5, P admits a unique (up to a multiplication by a positive constant) measure μ. This measure satisfies $\mu(C) < \infty$ for every petite set C; since X is small, this implies that $\mu(\mathsf{X}) < \infty$, showing that P is positive. Denote by π the unique invariant probability.

It remains to prove that P is aperiodic. The proof is by contradiction. Assume that P is an irreducible Markov kernel with period d. There exists a sequence $C_0, \ldots, C_{d-1}$ of mutually disjoint accessible sets such that for all $i = 0, \ldots, d-1$ and $x \in C_i$, $P(x, C_{\bmod i+1 d}) = 0$. Note that $\bigcup_{i=0}^{d-1} C_i$ is absorbing, and hence there exists $i_0 \in \{0, \ldots, d-1\}$ such that $\nu(C_{i_0}) > 0$. Therefore, we should have, for all $x \in \mathsf{X}$, $P^m(x, C_{i_0}) > 0$, which contradicts $P^m(x, C_{i_0}) = 0$ for $x \notin C_i$, for $i \neq \bmod (i_0 - m)d$.

Since P is irreducible, positive, and aperiodic, we may conclude the proof by applying Theorem 15.2.4 (ii) with $V \equiv 1$ and $C = \mathsf{X}$.

(i) $\Rightarrow$ (iv) Follows from Theorem 15.2.4 (a) and (iv) $\Rightarrow$ (i) from Theorem 15.2.4 (b).

□

Example 15.3.2 (Compact state space). Let (X, d) be a compact metric space and P a Markov kernel with transition density. Assume that there exists a function $t : \mathsf{X} \times \mathsf{X} \to \mathbb{R}_+$ with respect to a σ-finite reference measure ν such that

(i) for all $x \in \mathsf{X}$ and $A \in \mathscr{X}$, $P(x,A) \ge T(x,A) := \int_A t(x,y)\nu(\mathrm{d}y)$;
(ii) for all $y \in \mathsf{X}$, $x \mapsto t(x,y)$ is continuous;

(iii) $t(x,y) > 0$ for all $x, y \in \mathsf{X}$.

Since the space X is compact, $\inf_{x\in\mathsf{X}} t(x,y) = \min_{x\in\mathsf{X}} t(x,y) \geq g(y) > 0$ for all $y \in \mathsf{X}$. Hence for all $x \in \mathsf{X}$ and $A \in \mathscr{X}$, we get that

$$P(x,A) \geq \int_A t(x,y)\nu(\mathrm{d}y) \geq \int_A g(y)\nu(\mathrm{d}y),$$

showing that the space is $(1, \varepsilon\varphi)$-small with $\varphi(A) = \int_A g(y)\nu(\mathrm{d}y)/\int_{\mathsf{X}} g(y)\nu(\mathrm{d}y)$ and $\varepsilon = \int_{\mathsf{X}} g(y)\nu(\mathrm{d}y)$.

In the case of the Metropolis–Hastings algorithm, such conditions hold if, for example, the proposal density $q(x,\ y)$ is continuous in x for all y, is positive for all x, y, and the target probability has a density π that is continuous and positive everywhere with

$$t(x,y) = q(x,y) 1 \wedge \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}.$$

◀

Example 15.3.3 (Independent Metropolis–Hastings sampler). We consider again the independent Metropolis–Hastings algorithm (see Section 2.3.1, Example 2.3.3). Let μ be a σ-finite measure on $(\mathsf{X}, \mathscr{X})$. Let h be the density with respect to μ of the target distribution π. Denote by q the proposal density. Assume that $\sup_{x\in\mathsf{X}} h(x)/q(x) < \infty$. For $k \geq 1$, given X_{k-1}, a proposal Y_k is drawn from the distribution q, independently of the past. Then set $X_k = Y_k$ with probability $\alpha(X_k, Y_k)$, where

$$\alpha(x,y) = \frac{h(y)q(x)}{h(x)q(y)} \wedge 1 \ .$$

Otherwise, set $X_{k+1} = X_k$. The transition kernel P of the Markov chain is defined, for $(x,A) \in \mathsf{X} \times \mathscr{X}$, by

$$P(x,A) = \int_A q(y)\alpha(x,y)\mu(\mathrm{d}y) + \left[1 - \int q(y)\alpha(x,y)\mu(\mathrm{d}y)\right] \delta_x(A) \ .$$

As shown in Proposition 2.3.1, π is reversible with respect to P. Hence π is a stationary distribution for P. Assume now that there exists $\varepsilon > 0$ such that

$$\inf_{x\in\mathsf{X}} \frac{q(x)}{h(x)} \geq \varepsilon \ . \tag{15.3.1}$$

Then for all $x \in \mathsf{X}$ and $A \in \mathscr{X}$, we have

$$\begin{aligned} P(x,A) &\geq \int_A \left(\frac{h(y)q(x)}{h(x)q(y)} \wedge 1\right) q(y)\mu(\mathrm{d}y) \\ &= \int_A \left(\frac{q(x)}{h(x)} \wedge \frac{q(y)}{h(y)}\right) h(y)\mu(\mathrm{d}y) \geq \varepsilon\pi(A) \ . \end{aligned} \tag{15.3.2}$$

Thus X is a small set, and the kernel P is uniformly geometrically ergodic by Theorem 15.3.1 (iii).

Consider the situation in which $\mathsf{X} = \mathbb{R}$ and the target density is a zero-mean standard Gaussian $\pi = \mathrm{N}(0,1)$. Assume that the proposal density is chosen to be the density of the $\mathrm{N}(1,1)$ distribution. The acceptance ratio is given by

$$\alpha(x,y) = 1 \wedge \frac{h(y)}{h(x)} \frac{q(x)}{q(y)} = 1 \wedge \mathrm{e}^{x-y} .$$

This choice implies that moves to the right may be rejected, but moves to the left are always accepted. Condition (15.3.1) is not satisfied in this case. It is easily shown that the algorithm does not converge at a geometric rate to the target distribution (see Exercise 15.10).

If, on the other hand, the mean is known but the variance (which is equal to 1) is unknown, then we may take the proposal density q to be $\mathrm{N}(0,\sigma^2)$ for some known $\sigma^2 > 1$. Then $q(x)/h(x) \geq \sigma^{-1}$, and (15.3.1) holds. This shows that the state space is small and hence that the Markov kernel P is uniformly geometrically ergodic. ◀

15.4 Exercises

15.1. Consider the Markov chain in $\mathbb{R}_+$ defined by $X_{k+1} = (X_k + Z_{k+1})^+$, where $\{Z_k,\ k \in \mathbb{N}\}$ is a sequence of random variables such that $\mathbb{E}[Z_1] < \infty$ and for $M < \infty$ and $\beta > 0$, $\mathbb{P}(Z_1 > y) \leq M\mathrm{e}^{-\beta y}$ for all $y \in \mathbb{R}_+$. Show that $\mathrm{D_g}(V, \lambda, b, C)$ is satisfied with $V(x) = \mathrm{e}^{tx} + 1$ for some positive t and C chosen as $[0,c]$ for some $c > 0$.

15.2. Consider the Metropolis–Hastings kernel P defined in (2.3.4). Let $\bar{\alpha}(x) = \int_{\mathsf{X}} \{1 - \alpha(x,y)\} q(x,y) \nu(\mathrm{d}y)$ be the rejection probability from each point $x \in \mathsf{X}$. Show that if $\operatorname{esssup}_\pi(\bar{\alpha}) = 1$ and $\pi(\{x\}) < 1$ for all $x \in \mathsf{X}$, then the Metropolis–Hastings kernel is not geometrically ergodic.

15.3. Consider the functional autoregressive model $X_k = f(X_{k-1}) + \sigma(X_{k-1}) Z_k$, where $\{Z_k,\ k \in \mathbb{N}\}$ are i.i.d. standard Gaussian random variables, f and σ are bounded measurable functions, and there exist $a, b > 0$ such that $a \leq \sigma^2(x) \leq b$ for all $x \in \mathbb{R}$. Show that the associated kernel is uniformly geometrically ergodic.

15.4. We use the notation of Section 2.3. Consider the independent sampler introduced in Example 2.3.3.

(i) Assume that $\bar{q}(x)/h(x) \geq c$, π-*a.e.* Show that the Markov kernel P is uniformly ergodic.

(ii) Assume that $\sup\{c > 0 : \pi(\{x \in \mathsf{X} : \bar{q}(x)/h(x) \leq c\}) = 0\} = 0$. Show that the Markov kernel P is not geometrically ergodic.

15.5. Let P and Q be two Markov kernels on $\mathsf{X} \times \mathscr{X}$. Assume that P is uniformly ergodic. Let $\alpha \in (0,1)$. Show that $\alpha P + (1-\alpha) Q$ is uniformly ergodic.

15.6. Show that a Markov kernel on a finite state space for which all the states are accessible and that is aperiodic is always uniformly geometrically ergodic.

15.7. Let $X_k = (\alpha_0 + \alpha_1 X_{k-1}^2)^{1/2} Z_k$, where $\{Z_k,\ k \in \mathbb{N}\}$ is an i.i.d. sequence, be an ARCH(1) sequence. Assume that $\alpha_0 > 0$, $\alpha_1 > 0$, and that the random variable Z_1 has a density g that is bounded away from zero on a neighborhood of 0, i.e., $g(z) \geq g_{\min} \mathbb{1}_{[-a,a]}(z)$ for some $a > 0$. Assume also that there exists $s \in (0,1]$ such that $\mu_{2s} = \mathbb{E}\left[Z_0^{2s}\right] < \infty$. Set $V(x) = 1 + x^{2s}$.

1. Assume $\alpha_1^s \mu_{2s} < 1$. Show that $PV(x) \leq \lambda V(x) + b$ for some $\lambda \in (0,1]$ and $b < \infty$.
2. Show that every interval $[-c,c]$ with $c > 0$ is small.

15.8. Consider the INAR (or Galton–Watson process with immigration) $\{X_n,\ n \in \mathbb{N}\}$ introduced in Exercise 14.4, defined by X_0 and

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_{n,i}^{(n+1)} + Y_{n+1} \; .$$

Set $m = \mathbb{E}\left[\xi_1^{(1)}\right]$. Assume that $m < 1$. Show that this Markov chain is geometrically ergodic.

15.9. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P is uniformly ergodic. Show that for every accessible set A, there exists $\delta_A \in (1,\infty)$ such that $\sup_{x \in \mathsf{X}} \mathbb{E}_x[\delta_A^{\sigma_A}] < \infty$.

15.10. Consider an independent Metropolis–Hastings sampler on $\mathsf{X} = \mathbb{R}$. Assume that the target density is a zero-mean standard Gaussian $\pi = \mathrm{N}(0,1)$ and the proposal density is the $\mathrm{N}(1,1)$ distribution. Show that the state space is not small.

15.11. Let P be a random walk Metropolis algorithm on $\mathbb{R}^d$ with target distribution $\pi = h_\pi \cdot \mathrm{Leb}$ and proposal density $q(x,y) = \bar{q}(|y-x|)$, where $\bar{q}$ is a bounded function. If $\operatorname{esssup}_\pi (h_\pi) = \infty$, then P is not geometrically ergodic. [Hint: Use Example 15.1.7.]

15.12. Consider the following count model:

$$X_k = \beta + \gamma(N_{k-1} - \mathrm{e}^{X_{k-1}})\mathrm{e}^{-X_{k-1}} \; , \tag{15.4.1}$$

where, conditionally on $(X_0,\ldots,X_k)$, N_k has a Poisson distribution with intensity e^{X_k}. Show that the chain is geometrically uniformly ergodic.

15.13. We want to sample the distribution on $\mathbb{R}^2$ with density with respect to Lebesgue measure proportional to

$$\pi(\mu,\theta) \propto \theta^{-(m+1)/2} \exp\left(-\frac{1}{2\theta}\sum_{j=1}^{m}(y_j - \mu)^2\right) , \tag{15.4.2}$$

where $\{y_j\}_{j=1}^m$ are constants. This might be seen as the posterior distribution in Bayesian analysis of the parameters in a model where $Y_1,\ldots,Y_m$ are i.i.d. $\mathrm{N}(\mu,\theta)$

and the prior (μ,θ) belongs to $\theta^{-1/2}\mathbb{1}_{\mathbb{R}_+}(\theta)$ (this prior is improper, but the posterior distribution is proper as long as $m \geq 3$). We use a two-stage Gibbs sampler (Section 2.3.3) to make draws from (15.4.2), which amounts to drawing

- $\mu_{k+1} \sim R(\theta_k,\cdot)$ with

$$R(\theta,A) = \int_A \frac{1}{\sqrt{2\pi\theta/m}} \exp\left(-\frac{m}{2\theta}(\mu-\bar{y})^2\right) \mathrm{d}\mu\ , \quad \text{with}\, \bar{y} = m^{-1}\sum_{i=1}^m y_i\ ;$$

- $\theta_{k+1} \sim S(\mu_{k+1},\cdot)$ with

$$C(\mu,A) = \int g\left(\frac{m-1}{2}, \frac{s^2+m(\bar{y}-\mu)^2}{2}; \theta\right) \mathrm{d}\theta,$$

  where for $(\alpha,\beta) \in \mathbb{R}_+ \times \mathbb{R}_+$,

$$g(\alpha,\beta;\theta) \propto \theta^{-(\alpha+1)} \mathrm{e}^{-\beta/\theta} \mathbb{1}_{\mathbb{R}_+}(\theta)$$

  and $s^2 = \sum_{i=1}^m (y_i - \bar{y})^2$.

Denote by P the transition kernel associated to this Markov chain. Assume that $m \geq 5$. Define $V(\mu,\ \theta) = (\mu-\bar{y})^2$.

1. Show that

$$\begin{aligned}\mathbb{E}\left[V(\mu_{k+1},\theta_{k+1}) \mid \mu_k,\theta_k\right] &= \mathbb{E}\left[V(\mu_{k+1},\theta_{k+1}) \mid \mu_k\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[V(\mu_{k+1},\theta_{k+1}) \mid \theta_{k+1}\right] \mid \mu_k\right]\ .\end{aligned}$$

2. Show that $\mathbb{E}\left[V(\mu_{k+1},\ \theta_{k+1}) \mid \theta_{k+1}\right] = \frac{\theta_{k+1}}{m}$.
3. Show that

$$\mathbb{E}\left[V(\mu_{k+1},\ \theta_{k+1}) \mid \mu_k,\ \theta_k\right] = \frac{1}{m-3} V(\mu_k,\ \theta_k) + \frac{s^2}{m(m-3)}.$$

4. Prove the following drift condition:

$$PV(\mu',\ \theta') \leq \gamma V(\mu',\ \theta') + L,$$

   where $\gamma \in (1/(m-3),\ 1)$ and $L = s^2/(m(m-3))$.

15.5 Bibliographical Notes

Uniform ergodicity dates back to the earliest work on Markov chains on general state spaces by Doeblin (1938) and Doob (1953).

Geometric ergodicity of nonlinear time series models have been studied by many authors (see Tjostheim (1990), Tong (1990), Tjøstheim (1994) and the references therein). It is difficult to give proper credit to all these research efforts.

Geometric ergodicity of functional autoregressive processes was studied, among many references, by Doukhan and Ghindès (1983), Bhattacharya and Lee (1995), An and Chen (1997). The stability of the self-exciting threshold autoregression (SETAR) model of order 1 was completely characterized by Petruccelli and Woolford (1984), Chan et al. (1985), Guo and Petruccelli (1991)). Cline and Pu (1999) (see also Cline and Pu (2002, 2004)) develop general conditions under which nonlinear time series with state-dependent errors ($X_k = \alpha(X_{k-1}) + \gamma(X_{k-1}; Z_k)$, where $\{Z_k,\ k \in \mathbb{N}\}$ is an i.i.d. sequence) are geometrically ergodic (extending the conditions given in Example 15.2.5).

Similarly, numerous works have been devoted to finding conditions under which MCMC algorithms are geometrically ergodic. It is clearly impossible to cite all these works here. The uniform geometric ergodicity of the independence sampler (see Example 15.3.3) was established in Tierney (1994) and Mengersen and Tweedie (1996). The geometric ergodicity of the random walk Metropolis (Example 15.2.6) is discussed in Roberts and Tweedie (1996), Jarner and Hansen (2000), and Saksman and Vihola (2010). Geometric ergodicity of hybrid Monte Carlo methods (including the Gibbs sampler; see Example 15.13) is studied in Roberts and Rosenthal (1997), Hobert and Geyer (1998), Roberts and Rosenthal (1998). Many results can be found in Rosenthal (1995a), Rosenthal (2001), Roberts and Rosenthal (2004), and Rosenthal (2009). Example 15.1.7 is borrowed from Roberts and Tweedie (1996).

Explicit bounds using the splitting construction and regenerations are discussed in Meyn and Tweedie (1994) and Baxendale (2005). Hobert et al. (2002) discusses a way to use regeneration techniques as a simulation method.

Chapter 16
(f,r)-Recurrence and Regularity

In Chapter 14, we introduced the notions of f-geometric recurrence and f-geometric regularity. We showed that these two conditions coincided for petite sets. We also established a drift condition and showed that it is, under mild conditions, equivalent to f-geometric recurrence and regularity. In this chapter we will establish parallel results for subgeometric rates of convergence. In Section 16.1, we will define (f,r)-recurrence. The main difference with geometric recurrence is that (f,r)-recurrence is equivalent to an infinite sequence of drift conditions, rather than a single one. However, we will introduce the sufficient condition $\mathrm{D}_{\mathrm{sg}}(V,\phi,b,C)$, which is in practice more convenient to obtain than the aforementioned sequence of drift conditions. Following the path of Chapter 14, we will then introduce (f,r)-regularity and establish its relation to (f,r)-recurrence in Section 16.2. The regularity of the skeletons and split kernel will be investigated in Sections 16.2–16.4.

16.1 (f,r)-Recurrence and Drift Conditions

We now introduce the notion of (f,r)-recurrence, where the rate function r is not necessarily geometric. This generalizes the (f,δ)-geometric recurrence defined in Definition 14.1.1.

Definition 16.1.1 **($(f,r)-$recurrence)** *Let $f : \mathsf{X} \to [1,\infty)$ be a measurable function and $r = \{r(n),\ n \in \mathbb{N}\}$ a sequence such that $r(n) \geq 1$ for all $n \in \mathbb{N}$. A set $C \in \mathscr{X}$ is said to be (f,r)-recurrent if*

$$\sup_{x\in C} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C-1} r(k) f(X_k) \right] < \infty\,. \tag{16.1.1}$$

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_16

The definition of (f,r)-recurrence implies that $f \geq 1$ and $r \geq 1$. This implies $\sup_{x\in C} \mathbb{E}_x[\sigma_C] < \infty$, which in turn yields $\mathbb{P}_x(\sigma_C < \infty) = 1$ for all $x \in C$. An (f,r)-recurrent set is therefore necessarily Harris recurrent and recurrent (see Definitions 10.1.1 and 10.2.1). The (f,r)-recurrence property will be used (with further conditions on the set C) to prove the existence of an invariant probability measure, to control moments of this invariant probability, and to obtain rates of convergence of the iterates of the Markov kernel to its stationary distribution (when such a distribution exists and is unique).

Again, the (f,r) recurrence property will be shown to be equivalent to drift conditions. We first introduce the following sequence of drift conditions.

Definition 16.1.2 (Condition $\mathrm{D_{sg}}(\{V_n\}, f, r, b, C)$) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. The Markov kernel P is said to satisfy condition $\mathrm{D_{sg}}(\{V_n\}, f, r, b, C)$ if $V_n : \mathsf{X} \to [0,\infty]$, $n \in \mathbb{N}$, are measurable functions, $f : \mathsf{X} \to [1,\infty)$ is a measurable function, $\{r(n),\ n \in \mathbb{N}\}$ is a sequence such that $\inf_{n\in\mathbb{N}} r(n) \geq 1$, $b > 0$, $C \in \mathscr{X}$, and for all $n \in \mathbb{N}$,*

$$PV_{n+1} + r(n)f \leq V_n + br(n)\mathbb{1}_C\,. \tag{16.1.2}$$

Remark 16.1.3. For simplicity, we have adopted the convention $\inf_{x\in\mathsf{X}} f(x) \geq 1$. In fact, it is enough to assume that $\inf_{x\in\mathsf{X}} f(x) > 0$. It suffices to rescale the drift condition (16.1.2). ▲

Let $f : \mathsf{X} \to [1,\infty)$ be a measurable function, $C \in \mathscr{X}$ a set, and $r = \{r(n),\ n \in \mathbb{N}\}$ a nonnegative sequence. Define

$$W^{f,r}_{n,C}(x) = \mathbb{E}_x\left[\sum_{k=0}^{\tau_C - 1} r(n+k)f(X_k)\right], \tag{16.1.3}$$

with the convention $\sum_0^{-1} = 0$, so that $W^{f,r}_{n,C}(x) = 0$ for $x \in C$. The set of log-subbaditive sequences $\bar{\mathscr{S}}$ and related sequences are defined in Section 13.1.

Proposition 16.1.4 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let $C \in \mathscr{X}$, $f : \mathsf{X} \to [1,\infty)$ a measurable function, and $\{r(n),\ n \in \mathbb{N}\} \in \bar{\mathscr{S}}$. The following conditions are equivalent:*

(i) The set C is (f,r)-recurrent.
(ii) Condition $\mathrm{D_{sg}}(\{V_n\}, f, r, b, C)$ holds and $\sup_{x\in C} V_0(x) < \infty$.

Moreover, if the set C is (f,r)-recurrent, then condition $\mathrm{D_{sg}}(\{V_n\}, f, r, b, C)$ is satisfied with $V_n = W^{f,r}_{n,C}$ and $b = \sup_{x\in C} \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C - 1} r(k)f(X_k)\right]$. In addition, if condition $\mathrm{D_{sg}}(\{V_n\}, f, r, b, C)$ is satisfied, then

$$\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right] \le V_0(x) + br(0)\mathbb{1}_C(x) . \tag{16.1.4}$$

Proof. We can assume without loss of generality that $r \in \mathscr{S}$.

(i) $\Rightarrow$ (ii) Assume that C is (f,r)-recurrent. For all $x \in \mathsf{X}$, we get

$$PW_{1,C}^{f,r}(x) + r(0)f(x) = \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right] .$$

Hence $PW_{1,C}^{f,r} + r(0)f \le W_{0,C}^{f,r} + b\mathbb{1}_C$ with $b = \sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right]$, showing that condition $\mathrm{D_{sg}}(\{V_n\}, f, r, b, C)$ holds with $V_n = W_{n,C}^{f,r}$ (see (16.1.3)). Moreover, in that case, $\sup_{x\in C} V_0(x) = 0 < \infty$.

(ii) $\Rightarrow$ (i) Assume that $\mathrm{D_{sg}}(\{V_n\}, f, r, b, C)$ holds. For every $x \in \mathsf{X}$, we get

$$\mathbb{E}_x\left[V_{\sigma_C}(X_{\sigma_C})\mathbb{1}_{\{\sigma_C<\infty\}}\right] + \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right] \le V_0(x) + br(0)\mathbb{1}_C(x) , \tag{16.1.5}$$

which implies (16.1.4). If, in addition, $\sup_{x\in C} V_0(x) < \infty$, then (16.1.5) ensures that C is (f,r)-recurrent.

□

Example 16.1.5. If $r \equiv 1$, then Proposition 16.1.4 shows that the set C is $(f,1)$-recurrent if and only if there exists a function V such that $PV + f \le V + b\mathbb{1}_C$ and $\sup_C V < \infty$. Indeed, if the latter condition holds, then condition $\mathrm{D_{sg}}(\{V_n\}, f, r, b, C)$ holds with $V_n = V$ for all n.

Example 16.1.6 (Random walk on the half-line). Let P be the Markov transition kernel for the random walk on $[0,\infty)$ given for all $n \in \mathbb{N}$ by

$$X_{n+1} = (X_n + W_{n+1})^+ , \tag{16.1.6}$$

where $\{W_n,\ n \in \mathbb{N}\}$ is a sequence of i.i.d. real-valued random variables with common distribution ν. We assume that $\mathbb{E}[W_1] < 0$ and that there exists an integer $m \ge 2$ such that

$$\mathbb{E}\left[\{W_1^+\}^m\right] < \infty . \tag{16.1.7}$$

It is easily shown that the Markov kernel chain is δ_0-irreducible, aperiodic, and positive, and all compact sets are petite. We first assume that the support of the distribution ν is included in $[-x_0,\infty)$ for some $x_0 \in \mathbb{R}_+$. Choose $a > 0$ in such a way that $c := -(m/2)\mathbb{E}[W + a] > 0$. We define for $x \in \mathbb{R}_+$ and $n \in \mathbb{N}$,

$$V_n(x) = (x + an)^m .$$

For all $x > x_0$, we get

$$PV_{n+1}(x) = \int_{-x_0}^{\infty} (x+an+a+y)^m \nu(\mathrm{d}y)$$
$$\leq V_n(x) - 2b(x+an)^{-1} + (x+an)^{m-2}\varsigma(m),$$

where $\varsigma(m) < \infty$. We may now choose $z_0 \geq 1$ large enough that for $x \notin C := [0, z_0]$,

$$PV_{n+1}(x) \leq V_n(x) - c(x+an)^{-1} \leq V_n(x) - r_k(n) f_k(x) ,$$

where $r_k(n) = n^{m-k}$ and $f_k(x) = c\binom{m}{k} a^{m-k} x^k \vee 1$, for all $k \in \{0, \ldots, m-1\}$. Note that $\inf f_k(x) > 0$ (see Remark 16.1.3). The set C is petite for P and $\sup_C V_0(x) < \infty$ and $\sup_{x \in C}(x+an)^{m-2} < \infty$.

To handle the general case, we may truncate the distribution ν at $-x_0$, so that the truncated distribution still has a negative mean. The Markov kernel $\tilde{P}$ satisfies the condition above. Therefore, we have

$$\sup_{x \in C} \mathbb{E}_x^{\tilde{P}} \left[\sum_{n=0}^{\sigma_C - 1} r_k(n) f_k(X_n) \right] < \infty ,$$

where for Q a Markov kernel on $(\mathsf{X}, \mathscr{X})$ and $\xi \in \mathbb{M}_1(\mathscr{X})$, $\mathbb{P}_\xi^Q$ and $\mathbb{E}_\xi^Q$ denote the distribution and expectation of a Markov chain started at x with transition kernel Q. By a stochastic domination argument (which is in this case a straightforward application of coupling; see Chapter 19), it may be shown that

$$\mathbb{E}_x^P \left[\sum_{n=0}^{\sigma_C - 1} r_k(n) f_k(X_n) \right] \leq \mathbb{E}_x^{\tilde{P}} \left[\sum_{n=0}^{\sigma_C - 1} r_k(n) f_k(X_n) \right] .$$

The proof follows. ◀

In practice, it may be relatively hard to find a sequence of functions $\{V_n\}$, a function f, and a sequence $r \in \bar{\mathscr{S}}$ such that condition $\mathrm{D_{sg}}(\{V_n\}, f, r, b, C)$ holds. We now introduce another drift condition that may appear a bit more restrictive, but in practice, it provides most usual subgeometric rates.

Definition 16.1.7 (Condition $\mathrm{D_{sg}}(V, \phi, b, C)$) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. The Markov kernel P is said to satisfy the subgeometric drift condition subgeometricdrift if $V : \mathsf{X} \to [1, \infty)$ is a measurable function, $\phi : [1, \infty) \to (0, \infty)$ is a concave, increasing function, continuously differentiable on $(0, \infty)$ such that $\lim_{v \to \infty} \phi'(v) = 0$, $b > 0$, $C \in \mathscr{X}$, and*

$$PV + \phi \circ V \leq V + b\mathbb{1}_C . \tag{16.1.8}$$

If $C = \mathsf{X}$, we simply write $\mathrm{D_{sg}}(V, \phi, b)$.

Remark 16.1.8. Recall that in the condition $D_{sg}(V,\phi,b,C)$ it is assumed that ϕ is concave and continuously differentiable, and $\lim_{v\to\infty}\phi'(v)=0$. Since ϕ' is nonincreasing, if we do not assume that $\lim_{v\to\infty}\phi'(v)=0$, then there exists $c\in(0,1)$ such that $\lim_{v\to\infty}\phi'(v)=c>0$. This yields $v-\phi(v)\le(1-c)v+c-\phi(1)$, and in this case, condition $D_{sg}(V,\phi,b,C)$ implies $D_g(V,1-c,b')$ for some suitable constant b'. ▲

Theorem 16.1.9. *Let P be an irreducible Markov kernel on $\mathsf{X}\times\mathscr{X}$, $V:\mathsf{X}\to[1,\infty)$ a measurable function, $\phi:[1,\infty)\to(0,\infty)$ a concave, increasing function, continuously differentiable on $(0,\infty)$ such that $\lim_{v\to\infty}\phi(v)=\infty$ and $\lim_{v\to\infty}\phi'(v)=0$. The following conditions are equivalent:*

(i) There exists $b\in[0,\infty)$ such that

$$PV+\phi\circ V\le V+b\,. \tag{16.1.9}$$

Moreover, for all $d>0$, the sets $\{V\le d\}$ are petite, and there exists d_0 such that for all $d\ge d_0$, $\{V\le d\}$ is accessible.

(ii) There exist $b,d_1\in[0,\infty)$ such that

$$PV+\phi\circ V\le V+b\mathbb{1}_{\{V\le d_1\}}\,, \tag{16.1.10}$$

and for all $d\ge d_1$, the set $\{V\le d\}$ is petite and accessible.

(iii) There exist a petite set C and $b\in[0,\infty)$ such that

$$PV+\phi\circ V\le V+b\mathbb{1}_C\,. \tag{16.1.11}$$

Proof. (i) $\Rightarrow$ (ii) We have only to prove (16.1.9). Choose d such that $\phi(d)\ge 2b$. The level set $C=\{V\le d\}$ is petite by assumption. For $x\in C$, $PV(x)+(1/2)\phi\circ V(x)\le V(x)+b$. For $x\notin C$,

$$\begin{aligned}PV(x)+(1/2)\phi\circ V(x)&\le PV(x)+\phi\circ V(x)-(1/2)\phi(d)\\&\le V(x)+b-(1/2)\phi(d)\le V(x)\,.\end{aligned}$$

(ii) $\Rightarrow$ (iii) We obtain (16.1.11) from (16.1.10) by taking $C=\{V\le d\}$.

(iii) $\Rightarrow$ (i) Since ϕ is nondecreasing and $V\ge 1$, we have $PV+\phi(1)\le PV+\phi\circ V\le V+b\mathbb{1}_C$. By applying Proposition 4.3.2 with $f\equiv\phi(1)$, we get

$$\phi(1)\mathbb{E}_x[\sigma_C]\le V(x)+b\mathbb{1}_C(x)\,,$$

showing that $\{x\in\mathsf{X}:\ V(x)\le d\}\subset\{x\in\mathsf{X}:\ \phi(1)\mathbb{E}_x[\sigma_C]\le d+b\}$. Since the set C is petite, $\{x\in\mathsf{X}:\ \phi(1)\mathbb{E}_x[\sigma_C]\le d+b\}$ is petite by Lemma 9.4.8, and therefore $\{V\le d\}$ is petite. Using (16.1.11), Proposition 9.2.13 applied with $V=V_0=V_1$ shows that (noting $\{V<\infty\}$) the set $\{V<\infty\}$ is full and absorbing, there exists d_0 such that $\{V\le d_0\}$ is accessible, which implies that for all $d\ge d_0$, $\{V\le d\}$ is accessible.

We now introduce a subclass of subgeometric rate functions indexed by concave functions, related to $\mathrm{D_{sg}}(V, \phi, b, C)$.

Let $\psi : [1, \infty) \to (0, \infty)$ be a concave increasing differentiable function. Let H_ψ be the primitive of $1/\psi$ that cancels at 1, i.e.,

$$H_\psi(v) = \int_1^v \frac{\mathrm{d}x}{\psi(x)} . \tag{16.1.12}$$

Then H_ψ is an increasing concave differentiable function on $[1, \infty)$. Moreover, since ψ is concave, ψ' is decreasing. Hence $\psi(v) \leq \psi(1) + \psi'(1)(v-1)$ for all $v \geq 1$, which implies that H_ψ increases to infinity.

We can thus define its inverse $H_\psi^{-1} : [0, \infty) \to [1, \infty)$, which is also an increasing and differentiable function, with derivative $(H_\psi^{-1})'(v) = \psi \circ H_\psi^{-1}(v)$. Define the function r_ψ on $[0, \infty)$ by

$$r_\psi(t) = (H_\psi^{-1})'(t) = \psi \circ H_\psi^{-1}(t) . \tag{16.1.13}$$

For simplicity, whenever it is useful, we still denote by r_ψ the restriction of the function r_ψ on $\mathbb{N}$ (depending on the context, r_ψ is either a function or a sequence).

Lemma 16.1.10 *Let* $\psi : [1, \infty) \to (0, \infty)$ *be a concave increasing function, continuously differentiable on* $[1, \infty)$ *and such that* $\lim_{v\to\infty} \psi'(v) = 0$*. Then*

(i) the sequence $\{n^{-1} \log r_\psi(n), n \in \mathbb{N}\}$ *is decreasing to zero;*
(ii) for all n, m*,* $r_\psi(n+m) \leq r_\psi(n) r_\psi(m)/r_\psi(0)$*;*
(iii) $r_\psi \in \bar{\Lambda}_1$*, where* $\bar{\Lambda}_1$ *is defined in Definition 13.1.2.*

Proof. By definition of r_ψ, for all $t \geq 0$,

$$s(\log r_\psi)'(t) = \frac{r_\psi'(t)}{r_\psi(t)} = \frac{(H_\psi^{-1})'(t)\psi' \circ H_\psi^{-1})(t)}{(H_\psi^{-1})'(t)} = \psi' \circ H_\psi^{-1}(t) .$$

Thus the function $(\log r_\psi)'$ decreases to 0, since H_ψ^{-1} increases to infinity and ψ' decreases to 0.

(i) Note that

$$\frac{\log r_\psi(n)}{n} = \frac{\log r_\psi(0)}{n} + \frac{1}{n}\int_0^n (\log r_\psi)'(s)\,\mathrm{d}s .$$

This implies that the sequence $\{\log r_\psi(n)/n, n \in \mathbb{N}\}$ decreases to 0.

(ii) The concavity of $\log r_\psi$ implies that for all $n, m \geq 0$,

$$\log r_\psi(n+m) - \log r_\psi(n) \leq \log r_\psi(m) - \log r_\psi(0) .$$

(iii) The sequence $\{\tilde{r}_n,\ n \in \mathbb{N}\}$, where $\tilde{r}_\psi(n) = (r_\psi(0) \vee 1)^{-1} r_\psi(n)$, belongs to $\mathscr{S}$ and $\lim_{n\to\infty} \log \tilde{r}_\psi(n)/n = 0$. Hence $\tilde{r}_\psi \in \Lambda_0$ (where Λ_0 is defined in Definition 13.1.2), and the proof follows by Lemma 13.1.3, which shows that $\Lambda_0 \subset \Lambda_1$.

□

Of course, only the behavior of ψ at infinity is of interest. If $\psi : [1,\infty) \to (0,\infty]$ is a concave increasing function, continuously differentiable on $[1,\infty)$ and such that $\lim_{v\to\infty} \psi'(v) = 0$, we can always find a concave increasing function, continuously differentiable on $[1,\infty)$ and taking values in $[1,\infty)$, that coincides with ψ on $[v_1,\infty)$ for some sufficiently large v_1 (see Exercise 16.3). Examples of subgeometric rates are given in Exercise 16.4.

We now prove that the subgeometric drift condition $\mathrm{D_{sg}}(V,\phi,b,C)$ implies that there exist a sequence $\{V_n,\ n \in \mathbb{N}\}$ and a constant b' such that $\mathrm{D_{sg}}(\{V_n\},1,r_\phi,b',C)$ holds. For this purpose, we define, for $k \in \mathbb{N}$, the function H_k on $[1,\infty)$ and V_k on X by

$$H_k = H_\phi^{-1}(k + H_\phi) - H_\phi^{-1}(k) , \qquad V_k = H_k \circ V . \tag{16.1.14}$$

Since r_ϕ is the derivative of H_ϕ^{-1}, this yields

$$H_k(v) = \int_0^{H_\phi(v)} r_\phi(z+k)\,\mathrm{d}z .$$

For $k = 0$, this yields $H_0(v) = v - 1$ and $V_0 = V - 1$. Since r_ϕ is increasing, this also yields that the sequence $\{H_k\}$ is increasing and $H_k(v) \geq v$ for all $v \geq 1$.

Proposition 16.1.11 *The subgeometric drift condition* $\mathrm{D_{sg}}(V,\phi,b,C)$ *implies* $\mathrm{D_{sg}}(\{V_n\},1,r_\phi,br_\phi(1)/r_\phi^2(0),C)$.

Proof. The function r_ϕ being increasing and log-concave, this implies that H_k is concave for all $k \geq 0$ and

$$H_k'(v) = \frac{r_\phi(H_\phi(v)+k)}{\phi(v)} = \frac{r_\phi(H_\phi(v)+k)}{r_\phi(H_\phi(v))} . \tag{16.1.15}$$

This yields

$$\begin{aligned}
H_{k+1}(v) - H_k(v) &= \int_0^{H_\phi(v)} \{r_\phi(z+k+1) - r_\phi(z+k)\}\,\mathrm{d}z \\
&= \int_0^{H_\phi(v)} \int_0^1 r_\phi'(z+k+s)\,\mathrm{d}s\,\mathrm{d}z \\
&= \int_0^1 \{r_\phi(H_\phi(v)+k+s) - r_\phi(k+s)\}\,\mathrm{d}s \\
&\leq r_\phi(H_\phi(v)+k+1) - r_\phi(k) = \phi(v)H_{k+1}'(v) - r_\phi(k) .
\end{aligned}$$

Composing with V, we obtain

$$V_{k+1} - \phi \circ V \times H'_{k+1} \circ V \le V_k - r_\phi(k) \; . \tag{16.1.16}$$

Let g be a concave differentiable function on $[1,\infty)$. Since g' is decreasing, for all $v \ge 1$ and $x \in \mathbb{R}$ such that $v + x \ge 1$, we have

$$g(v+x) \le g(v) + x g'(v) \; . \tag{16.1.17}$$

Using the concavity of H_{k+1}, the drift condition $\mathrm{D_{sg}}(V,\phi,b,C)$, (16.1.17), and (16.1.16), we obtain, for all $k \ge 0$ and $x \in \mathsf{X}$,

$$\begin{aligned} PV_{k+1} &\le H_{k+1}(PV) \le H_{k+1}\left(V - \phi \circ V + b\mathbb{1}_C\right) \\ &\le H_{k+1} \circ V + (-\phi \circ V + b\mathbb{1}_C) H'_{k+1} \circ V \\ &\le V_{k+1} - \phi \circ V \times H'_{k+1} \circ V + bH'_{k+1}(1)\mathbb{1}_C \\ &\le V_k - r_\phi(k) + bH'_{k+1}(1)\mathbb{1}_C \; . \end{aligned}$$

Applying (16.1.15), we obtain that $H'_{k+1}(1) = r_\phi(k+1)/r_\phi(0) \le r_\phi(k) r_\phi(1)/r^2_\phi(0)$, which proves our claim. □

Theorem 16.1.12. *Let P be an irreducible kernel on $\mathsf{X} \times \mathscr{X}$. Assume that $\mathrm{D_{sg}}(V,\phi,b,C)$ holds. Then for all $x \in \mathsf{X}$,*

$$\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C - 1} \phi \circ V(X_k)\right] \le V(x) + b\mathbb{1}_C(x) \; , \tag{16.1.18}$$

$$\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C - 1} r_\phi(k)\right] \le V(x) + b\frac{r_\phi(1)}{r_\phi(0)}\mathbb{1}_C(x) \; . \tag{16.1.19}$$

If, moreover, π is an invariant probability measure, then $\pi(\phi \circ V) < \infty$.

Proof. The bound (16.1.18) follows from Proposition 4.3.2. By Proposition 16.1.11, condition $\mathrm{D_{sg}}(V,\phi,b,C)$ implies $\mathrm{D_{sg}}(\{V_n\},1,r_\phi,br_\phi(1)/r^2_\phi(0),C)$, where V_n is defined in (16.1.14). Thus we can apply Proposition 16.1.4, (16.1.4), to obtain the bound (16.1.19). The last statement follows from Proposition 4.3.2. □

Example 16.1.13 (Random walk on $[0,\infty)$; Example 16.1.6 (continued).). We will follow a different method here. Instead of checking $\mathrm{D_{sg}}(\{V_n\},f,r,b,C)$, we will instead use $\mathrm{D_{sg}}(V,\phi,b,C)$. As we will see, this approach has several distinctive advantages over the first method. More precisely, we will show that there exist a finite interval $C = [0, z_0]$ and constants $0 < c, b < \infty$ such that for all $x \in \mathbb{R}_+$,

$$PV(x) \le V(x) - cV^\alpha(x) + b\mathbb{1}_C(x) \; , \tag{16.1.20}$$

where we have set $V(x) = (x+1)^m$ for $x \in \mathbb{R}_+$ and $\alpha = (m-1)/m$.

Take $x_0 > 0$ so large that $\int_{-x_0}^{\infty} y\nu(\mathrm{d}y) < 0$. For $x > x_0$, we bound $PV(x)$ by considering jumps smaller than $-x_0$ and jumps larger than $-x_0$ separately,

$$PV(x) \leq V(x-x_0)\nu((-\infty,-x_0)) + \int_{-x_0}^{\infty} V(x+y)\nu(\mathrm{d}y) \,. \tag{16.1.21}$$

First, we bound $V(x-x_0)$ in terms of $V(x)$ and $(V(x))^{\alpha}$,

$$\begin{aligned} V(x) - V(x-x_0) &= \int_{x-x_0}^{x} m(y+1)^{m-1}\mathrm{d}y \geq x_0 m(x-x_0+1)^{m-1} \\ &\geq x_0 m \left(\frac{x-x_0+1}{x+1}\right)^{m-1} (x+1)^{m-1} \geq c_1(x+1)^{m-1} \,, \end{aligned}$$

where $c_1 = x_0 m(1/(x_0+1))^{m-1}$. We now bound the second term on the right-hand side of (16.1.21). First note that for $x \geq 0$ and $y \geq 0$, we get

$$(x+y+1)^{m-2} \leq (x+1)^{m-2}(y+1)^{m-2} \,, \tag{16.1.22}$$

since

$$\log(x+y+1) - \log(x+1) = \int_{x+1}^{x+1+y} \frac{1}{z}\mathrm{d}z \leq \int_{1}^{1+y} \frac{1}{z}\mathrm{d}z = \log(y+1) \,.$$

For $y > 0$, we then get

$$\begin{aligned} V(x+y) &\leq V(x) + m(x+1)^{m-1}y + \frac{1}{2}m(m-1)(x+y+1)^{m-2}y^2 \\ &\leq V(x) + m(x+1)^{m-1}y + \frac{1}{2}m(m-1)(x+1)^{m-2}(y+1)^m, \end{aligned}$$

and for $-x_0 \leq y \leq 0$, we get

$$V(x+y) \leq V(x) + m(x+1)^{m-1}y + \frac{1}{2}m(m-1)(x+1)^{m-2}x_0^2.$$

Plugging these bounds into (16.1.21) and using the assumption (16.1.7), we obtain, for $x > x_0$,

$$PV(x) \leq V(x) - c_2(x+1)^{m-1} + c_3(x+1)^{m-2}$$

for some constants $0 < c_2$, $c_3 < \infty$, which can be explicitly computed. Hence there exist a positive constant c and a real number $z_0 \geq x_0$ such that for $x > z_0$, $PV(x) \leq V(x) - c(x+1)^{m-1}$. Finally, since $PV(x)$ and $(x+1)^{m-1}$ are both bounded on $C = [0, z_0]$, there exists a constant b such that (16.1.20) holds.

For comparison, we note that when $\mathbb{E}\left[\mathrm{e}^{sW_1}\right] < \infty$ for some $s > 0$, the chain is geometrically ergodic, and there is a solution to the drift equation $PV \leq \lambda V + b$ with $\lambda \in [0,1)$ and Lyapunov function $V(x) = \mathrm{e}^{tx}$ for $t < s$. ◀

16.2 (f,r)-Regularity

Definition 16.2.1 (**$(f,r)-$regular sets and measures**) *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$, $f : \mathsf{X} \to [1,\infty)$ a measurable function, and $\{r(n),\, n \in \mathbb{N}\}$ a sequence such that* $\inf_{n\in\mathbb{N}} r(n) \geq 1$.

(i) A set $A \in \mathscr{X}$ is said to be (f,r)-regular if for all $B \in \mathscr{X}_P^+$,

$$\sup_{x\in A} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_B - 1} r(k) f(X_k) \right] < \infty .$$

(ii) A probability measure $\xi \in \mathbb{M}_1(\mathscr{X})$ is said to be (f,r)-regular if for all $B \in \mathscr{X}_P^+$,

$$\mathbb{E}_\xi \left[\sum_{k=0}^{\sigma_B - 1} r(k) f(X_k) \right] < \infty .$$

(iii) A point $x \in \mathsf{X}$ is said to be (f,r)-regular if $\{x\}$ (or δ_x) is (f,r)-regular. The set of (f,r)-regular points for P is denoted by $S_P(f,r)$.

(iv) The Markov kernel P is said to be (f,r)-regular if there exists an accessible (f,r)-regular set.

When $f \equiv 1$ and $r \equiv 1$, we will simply say regular instead of $(1,1)$-regular. There is an important difference between the definitions of (f,r)-regularity and f-geometric regularity. In the former, the sequence r is fixed and the same for all accessible sets. In the latter, the geometric rate may depend on the set.

Before going further, it is necessary to extend our considerations to the subgeometric case, Theorem 11.4.1.

Theorem 16.2.2. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, $C \in \mathscr{X}$, and ρ, τ two stopping times with $\tau \geq 1$. Assume that for all $n \in \mathbb{N}$,*

$$\rho \leq n + \rho \circ \theta_n , \qquad \text{on } \{\rho > n\} . \tag{16.2.1}$$

Moreover, assume that there exists $\gamma > 0$ such that for all $x \in C$,

$$\mathbb{P}_x(\tau < \infty,\, X_\tau \in C) = 1 , \qquad \mathbb{P}_x(\rho \leq \tau) \geq \gamma . \tag{16.2.2}$$

Then the following hold:

(i) For all $x \in C$, $\mathbb{P}_x(\rho < \infty) = 1$.

(ii) If $\sup_{x\in C}\mathbb{E}_x[r^0(\tau)]<\infty$ *(where* $r^0(n)=\sum_{k=0}^n r(k)$*) for a sequence* $r\in\bar{\Lambda}_2$*, then there exists* $\varsigma<\infty$ *such that for all* $h\in\mathbb{F}_+(\mathsf{X})$*,*

$$\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\rho-1} r(k)h(X_k)\right]\le\varsigma\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\tau-1} r(k)h(X_k)\right]. \tag{16.2.3}$$

Proof. The proof of (i) is identical to that of Theorem 11.4.1 (i). Define $\tau^{(0)}=0$, $\tau^{(1)}=\tau$, and for $n\ge1$, $\tau^{(n)}=\tau^{(n-1)}+\tau\circ\theta_{\tau^{(n-1)}}$. Set

$$M(h,r)=\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\tau-1} r(k)h(X_k)\right]. \tag{16.2.4}$$

For $r\in\mathscr{S}$ (see Definition 13.1.1), the strong Markov property implies

$$\begin{aligned}\mathbb{E}_x\left[\sum_{k=0}^{\rho-1} r(k)h(X_k)\right]&\le\sum_{k=0}^{\infty}\mathbb{E}_x\left[\mathbb{1}\left\{\rho>\tau^{(k)}\right\}\sum_{j=\tau^{(k)}}^{\tau^{(k+1)}-1} r(j)h(X_j)\right]\\&\le\sum_{k=0}^{\infty}\mathbb{E}_x\left[\mathbb{1}\left\{\rho>\tau^{(k)}\right\}r(\tau^{(k)})\mathbb{E}_{X_{\tau^{(k)}}}\left[\sum_{j=0}^{\tau-1} r(j)h(X_j)\right]\right]\\&\le M(h,r)\sum_{k=0}^{\infty}\mathbb{E}_x\left[\mathbb{1}\left\{\rho>\tau^{(k)}\right\}r(\tau^{(k)})\right].\end{aligned} \tag{16.2.5}$$

Note that this inequality remains valid even if $M(h,r)=\infty$.

Without loss of generality, we assume that $r\in\Lambda_2$. Set

$$M_1=\sup_{x\in C}\mathbb{E}_x\left[r^0(\tau)\right]<\infty, \tag{16.2.6}$$

which is finite by assumption. We prove by induction that for all $p\in\mathbb{N}^*$,

$$\sup_{x\in C}\mathbb{E}_x[r^0(\tau^{(p)})]=M_p<\infty. \tag{16.2.7}$$

Let $p\ge2$ and assume that $M_{p-1}<\infty$. Note that on the event $\{\tau^{(p-1)}<\infty\}$,

$$r^0(\tau^{(p)})\le r^0(\tau^{(p-1)})+r(\tau^{(p-1)})r^0(\tau)\circ\theta_{\tau^{(p-1)}}. \tag{16.2.8}$$

The strong Markov property implies that for all $x\in C$,

$$\begin{aligned}\mathbb{E}_x[r^0(\tau^{(p)})]&\le\mathbb{E}_x[r^0(\tau^{(p-1)})]+\mathbb{E}_x\left[r(\tau^{(p-1)})\mathbb{E}_{X_{\tau^{(p-1)}}}\left[r^0(\tau)\right]\right]\\&\le M_{p-1}+M_1\,M_{p-1}<\infty.\end{aligned}$$

Set $u_k(x) = \mathbb{E}_x\left[\mathbb{1}_{\{\rho>\tau^{(k)}\}} r^0(\tau^{(k)})\right]$. Using (16.2.8), the strong Markov property, we obtain $u_k(x) \le a_k(x) + b_k(x)$ with

$$\begin{aligned} a_k(x) &\le \mathbb{E}_x\left[\mathbb{1}\left\{\rho > \tau^{(k-1)}\right\} r^0(\tau^{(k-1)})\, \mathbb{P}_{X_{\tau^{(k-1)}}}(\rho > \tau)\right] \\ b_k(x) &\le \mathbb{E}_x\left[\mathbb{1}\left\{\rho > \tau^{(k-1)}\right\} r(\tau^{(k-1)})\, \mathbb{E}_{X_{\tau^{(k-1)}}}[r^0(\tau)]\right] . \end{aligned}$$

Applying (16.2.2), we obtain, using (11.4.4),

$$a_k(x) \le (1-\gamma)\mathbb{E}_x\left[\mathbb{1}\left\{\rho > \tau^{(k-1)}\right\} r^0(\tau^{(k-1)})\right] . \tag{16.2.9}$$

Since $r \in \Lambda_2$, we have $\lim_{k\to\infty} r(k)/r^0(k) = 0$, and there exists k_0 such that for all $k \ge k_0$, $M_1 r(k) \le (\gamma/2) r^0(k)$, where M_1 is defined in (16.2.6). Thus for all $x \in C$ and $k \ge k_0$,

$$\begin{aligned} b_k(x) &\le M_1\, \mathbb{E}_x\left[\mathbb{1}\left\{\rho > \tau^{(k-1)}\right\} r(\tau^{(k-1)})\right] \\ &\le (\gamma/2)\mathbb{E}_x\left[\mathbb{1}\left\{\rho > \tau^{(k-1)}\right\} r^0(\tau^{(k-1)})\right] . \end{aligned} \tag{16.2.10}$$

Combining (16.2.9) and (16.2.10), we obtain that $u_k(x) \le (1-\gamma/2)u_{k-1}(x)$ for all $x \in C$ and $k \ge k_0$. Since $\sup_{x\in C} u_k(x) \le M_{k_0}$ for $k \le k_0$, we get that for all $k \in \mathbb{N}$, $u_k(x) \le (1-\gamma/2)^{k-k_0} M_{k_0}$, which yields

$$\sup_{x\in C} \sum_{k=0}^{\infty} \mathbb{E}_x\left[\mathbb{1}\left\{\rho > \tau^{(k)}\right\} r(\tau^{(k)})\right] \le r(1) M_{k_0} \sum_{k=0}^{\infty} (1-\gamma/2)^{k-k_0} < \infty .$$

This proves (16.2.3) with $\varsigma = 2\gamma^{-1} r(1) M_{k_0} (1-\gamma/2)^{-k_0}$. □

We also need to extend Theorem 14.2.3.

Theorem 16.2.3. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $A, B \in \mathscr{X}$. Assume the following:*

(i) There exists $q \in \mathbb{N}^*$ such that $\inf_{x\in A} \mathbb{P}_x(\sigma_B \le q) > 0$.
(ii) $\sup_{x\in A} \mathbb{E}_x[r^0(\sigma_A)] < \infty$ for $r \in \bar{\Lambda}_2$ (where $r^0(n) = \sum_{k=0}^{n} r(k)$, and $\bar{\Lambda}_2$ is as defined in Definition 13.1.2).

Then there exists $\varsigma < \infty$ such that for all $h \in \mathbb{F}_+(\mathsf{X})$,

$$\sup_{x\in A} \mathbb{E}_x\left[\sum_{k=0}^{\sigma_B - 1} r(k) h(X_k)\right] \le \varsigma \sup_{x\in A} \mathbb{E}_x\left[\sum_{k=0}^{\sigma_A - 1} r(k) h(X_k)\right] .$$

Proof. The proof is along the same lines as that of Theorem 14.2.3 (using Theorem 16.2.2 instead of Theorem 11.4.1). □

Theorem 16.2.4. *Let P be an irreducible Markov kernel on $\mathsf{X}\times\mathscr{X}$, $f:\mathsf{X}\to[1,\infty)$ a measurable function, and $\{r(n),\ n\in\mathbb{N}\}\in\bar{\Lambda}_2$ (see Definition 13.1.2). The following conditions are equivalent:*

(i) The set C is accessible and (f,r)-regular.
(ii) The set C is petite and (f,r)-recurrent.

Proof. We can assume without loss of generality that $r\in\Lambda_2$.

(i) $\Rightarrow$ (ii) Assume that C is accessible and (f,r)-regular. Then C is $(1,1)$-regular. Let A be an accessible petite set. Then $\sup_{x\in C}\mathbb{E}_x[\sigma_A]<\infty$ by definition, and Lemma 9.4.8 implies that C is petite.

(ii) $\Rightarrow$ (i) First, the set C is accessible by Corollary 9.2.14. Moreover, since $f\geq 1$, $\sup_{x\in C}\mathbb{E}_x[r^0(\sigma_C-1)]<\infty$ where $r^0(n)=\sum_{k=0}^n r(k)$. Therefore, since $r\in\mathscr{S}$,

$$\sup_{x\in C}\mathbb{E}_x[r^0(\sigma_C)]<r(0)+r(1)\sup_{x\in C}\mathbb{E}_x[r^0(\sigma_C-1)]<\infty\,.$$

Let A be an accessible set. By Proposition 9.4.9 A is uniformly accessible from C and Theorem 16.2.3 implies that C is then (f,r)-regular.

Note that the assumption $r\in\bar{\Lambda}_2$ was implicitly used to invoke Theorem 16.2.3.

□

Similarly to Lemma 14.2.5, we now prove that a set that leads to an accessible (f,r)-regular set is also (f,r)-regular.

Lemma 16.2.5 *Let P be an irreducible Markov kernel on $\mathsf{X}\times\mathscr{X}$, $f:\mathsf{X}\to[1,\infty)$ a measurable function, and $r\in\bar{\Lambda}_2$ (see Definition 13.1.2). Assume that there exists an accessible (f,r)-regular set C. Then*

(i) for every $B\in\mathscr{X}_P^+$, there exists a constant $\varsigma<\infty$ such that for all $x\in\mathsf{X}$,

$$\mathbb{E}_x\left[\sum_{k=0}^{\sigma_B-1}r(k)f(X_k)\right]\leq\varsigma\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1}r(k)f(X_k)\right]\ ;$$

(ii) every set $A\in\mathscr{X}$ satisfying $\sup_{x\in A}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1}r(k)f(X_k)\right]<\infty$ is (f,r)-regular;

(iii) every probability measure $\xi\in\mathbb{M}_1(\mathscr{X})$ satisfying $\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1}r(k)f(X_k)\right]<\infty$ is (f,r)-regular.

Proof. Without loss of generality, we assume that $r\in\Lambda_2$. First note that (ii) and (iii) are immediate from (i). Since C is (f,r)-regular, for every $B\in\mathscr{X}_P^+$, we get

$$\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_B-1}r(k)f(X_k)\right]<\infty\,.$$

Since $\sigma_B \leq \sigma_C \mathbb{1}\{\sigma_C = \infty\} + (\sigma_C + \sigma_B \circ \theta_{\sigma_C})\mathbb{1}\{\sigma_C < \infty\}$, the strong Markov property shows that for all $x \in \mathsf{X}$,

$$
\begin{aligned}
&\mathbb{E}_x\left[\sum_{k=0}^{\sigma_B-1} r(k)f(X_k)\right] \\
&\quad \leq \mathbb{E}_x\left[\sum_{k=0}^{\tau_C-1} r(k)f(X_k)\right] + \mathbb{E}_x\left[\mathbb{1}_{\{\sigma_C<\infty\}}\left\{\sum_{k=0}^{\sigma_B-1} r(k)f(X_k)\right\}\circ\theta_{\sigma_C}\right] \\
&\quad \leq \mathbb{E}_x\left[\sum_{k=0}^{\tau_C-1} r(k)f(X_k)\right] + \mathbb{E}_x[r(\sigma_C)]\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_B-1} r(k)f(X_k)\right].
\end{aligned}
$$

The result follows because $\mathbb{E}_x[r(\sigma_C)] \leq r(1)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right] < \infty$. ☐

Theorem 16.2.6. *Let P be an irreducible Markov kernel on $\mathsf{X}\times\mathscr{X}$, $f:\mathsf{X}\to[1,\infty)$ a measurable function, and $r\in\bar{\Lambda}_2$. The Markov kernel P is (f,r)-regular if and only if it satisfies one of the following equivalent conditions:*

(i) There exists a nonempty (f,r)-recurrent petite set.
(ii) The condition $\mathrm{D_{sg}}(\{V_n\},f,r,b,C)$ holds for a nonempty petite set C and functions $\{V_n,\ n\in\mathbb{N}\}$ such that $\sup_{x\in C}V_0(x)<\infty$.
(iii) There exists an accessible (f,r)-regular set.
(iv) There exists an absorbing full set S that can be covered by a countable number of accessible (f,r)-regular sets.

If any of these conditions holds, the Markov kernel P satisfies the following properties, with the sequence $\{V_n,\ n\in\mathbb{N}\}$ as in item (ii):

(a) A probability measure $\xi\in\mathbb{M}_1(\mathscr{X})$ is (f,r)-regular if and only if there exists an (f,r)-recurrent petite set C such that $\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right]<\infty$.
(b) For every $A\in\mathscr{X}_P^+$, there exists a constant $\varsigma<\infty$ such that for all $x\in\mathsf{X}$,

$$
\mathbb{E}_x\left[\sum_{k=0}^{\sigma_A-1} r(k)f(X_k)\right] \leq \varsigma\{V_0(x)+1\}. \tag{16.2.11}
$$

(c) Every probability measure $\xi\in\mathbb{M}_1(\mathscr{X})$ such that $\xi(V_0)<\infty$ is (f,r)-regular.
(d) The set $S_P(f,r)$ of (f,r)-regular points is full and absorbing and is equal to $\{V_0<\infty\}$.

Proof. Without loss of generality, we assume that $r\in\Lambda_2$.

(i) $\Rightarrow$ (ii) Let C be a nonempty (f,r)-recurrent petite set. By Proposition 16.1.4, the condition $\mathrm{D_{sg}}(\{V_n\},f,r,b,C)$ is satisfied with $V_n = W_{n,C}^{f,r}$ (see (16.1.3)) and $b =$

$\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right] < \infty$. By construction, $W^{f,r}_{0,C}(x) = 0$ for $x \in C$, so that $\sup_{x\in C} V_0(x) < \infty$.

(ii) $\Rightarrow$ (iii) By Proposition 16.1.4, if $\mathrm{D_{sg}}(\{V_n\}, f, r, b, C)$ holds for a nonempty petite set C, $\sup_C V_0 < \infty$, then C is (f,r)-recurrent. Since C is petite, Theorem 16.2.4 shows that C is an accessible (f,r)-regular set.

(iii) $\Rightarrow$ (iv) Let C be an accessible (f,r)-regular set. By Theorem 16.2.4, C is also an (f,r)-recurrent petite set. For $d > 0$, set $C_d = \left\{x \in \mathsf{X} : W^{f,r}_{0,C}(x) \le d\right\}$. Since $\{W^{f,r}_{0,C} = \infty\} \subset \{W^{f,r}_{1,C} = \infty\}$ and $\{W^{f,r}_{0,C} < \infty\} \subset \{PW^{f,r}_{1,C} < \infty\}$, Proposition 9.2.13 shows that the set $\{W^{f,r}_{0,C} < \infty\}$ is full and absorbing (since $C \subset \{W^{f,r}_{0,C} < \infty\}$, this set is not empty) and the sets $\left\{x \in \mathsf{X} : W^{f,r}_{0,C}(x) \le n\right\}$ for $n \ge n_0$ are accessible. Lemma 16.2.5 show that the sets $\left\{x \in \mathsf{X} : W^{f,r}_{0,C}(x) \le n\right\}$ are (f,r)-regular.

(iv) $\Rightarrow$ (i) Obvious by Theorem 16.2.4.

(a) By Lemma 16.2.5 (iii), every $\xi \in \mathbb{M}_1(\mathscr{X})$ satisfying $\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right] < \infty$, where C is a petite set, is (f,r)-regular. Hence the condition is sufficient.

Conversely, assume that ξ is (f,r)-regular. Since P is (f,r)-regular, there exists an accessible (f,r)-regular set C. Since ξ is (f,r)-regular, $\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right] < \infty$. By Theorem 16.2.4, the set C is also (f,r)-recurrent and petite. This proves the necessary part.

(b) Assume that condition $\mathrm{D_{sg}}(\{V_n\}, f, r, b, C)$ holds for a nonempty petite set C and $\sup_{x\in C} V_0(x) < \infty$. By (16.1.4), we get

$$\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right] \le V_0(x) + br(0)\mathbb{1}_C(x)\,. \tag{16.2.12}$$

By Proposition 16.1.4, the set C is (f,r)-recurrent, and since it is also petite, we get that C is also accessible and (f,r)-regular. Then Lemma 16.2.5 (i) shows that for everyy $A \in \mathscr{X}_P^+$, there exists $\varsigma < \infty$ such that

$$\mathbb{E}_x\left[\sum_{k=0}^{\sigma_A-1} r(k)f(X_k)\right] \le \varsigma\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right]\,. \tag{16.2.13}$$

The proof of (16.2.11) follows by combining (16.2.12) and (16.2.13).

(c) This follows by integrating (16.2.11) with respect to $\xi \in \mathbb{M}_1(\mathscr{X})$.

(d) Since P is (f,r)-regular, there exists an accessible (f,r)-regular set C. Define $\left\{x \in \mathsf{X} : W^{f,r}_{0,C}(x) < \infty\right\}$. By Lemma 16.2.5, the sets $\{W^{f,r}_{0,C} \le n\}$ are (f,r)-regular for all $n \ge 0$. Hence $\{W^{f,r}_{0,C} \le n\} \subset S_P(f,r)$ for all $n \in \mathbb{N}$, and therefore $\{W^{f,r}_{0,C} < \infty\} = \bigcup_{n=1}^\infty \{W^{f,r}_{0,C} \le n\} \subset S_P(f,r)$.

Conversely, if $x \notin \{W^{f,r}_{0,C} < \infty\}$, then $x \notin C$ and $\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right] = \infty$. Since C is accessible, this implies that $x \notin S_P(f,r)$. Hence $\{W^{f,r}_{0,C} = \infty\} \subset S^c_P(f,r)$. $\square$

We conclude this section by studying the subgeometric regularity of the invariant probability measure.

Theorem 16.2.7. *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$, $f : \mathsf{X} \to [1,\infty)$ a measurable function, and $r = \{r(n),\, n \in \mathbb{N}\}$ a sequence such that $r(n) \geq 1$ for all $n \in \mathbb{N}$. Assume that P is (f,r)-regular. Then P has a unique invariant probability measure π. In addition, if $r \in \bar{\Lambda}_2$ (see Definition 13.1.2) is an increasing sequence, then for all $A \in \mathscr{X}_P^+$, $\mathbb{E}_\pi\left[\sum_{k=0}^{\sigma_A-1} \Delta r(k) f(X_k)\right] < \infty$.*

Remark 16.2.8. It would be tempting to say that π is $(f,\Delta r)$-regular. We nevertheless refrain from doing so, because the assumption $\inf_{n\geq 0} \Delta r(n) > 0$ is not necessarily fulfilled. For example, setting $r(k) = (k+1)^{1/2}$, we have $r \in \bar{\Lambda}_2$, but $\inf_{n\geq 0} \Delta r(n) = 0$. ▲

Proof. The existence and uniqueness of the invariant probability π follows from Corollary 11.2.9 along the same lines as in Theorem 14.2.7. We now assume without loss of generality that $r \in \Lambda_2$ and prove that for every accessible set A, we have $\mathbb{E}_\pi\left[\sum_{k=0}^{\sigma_A-1} \Delta r(k) f(X_k)\right] < \infty$. The ideas are similar to those used in Theorem 14.2.7 with some additional technical difficulties.

Let $A \in \mathscr{X}_P^+$. By Theorem 16.2.6, there exists an accessible full set S that is covered by a countable number of accessible (f,r)-regular sets, $S = \bigcup_{n=1}^\infty S_n$. Since π is a (maximal) irreducibility measure, $0 < \pi(A) = \pi(A \cap S)$, and therefore there exists n_0 such that $\pi(A \cap S_{n_0}) > 0$. Set $B = A \cap S_{n_0}$. Then B is a subset of A that is accessible (since $\pi(B) > 0$) and (f,r)-regular (as a subset of the (f,r)-regular set S_{n_0}). Moreover, since $\sigma_A \leq \sigma_B$, we have

$$\mathbb{E}_\pi\left[\sum_{n=0}^{\sigma_A-1} \Delta r(n) f(X_n)\right] \leq \mathbb{E}_\pi\left[\sum_{n=0}^{\sigma_B-1} \Delta r(n) f(X_n)\right] .$$

As above, consider the following functions: $g(x) = \mathbb{E}_x\left[\sum_{n=0}^{\sigma_B-1} \Delta r(n) f(X_n)\right]$ and $h(x) = \mathbb{E}_x\left[\sum_{n=0}^{\sigma_B-1} g(X_n)\right]$. Since B is accessible, Theorem 11.2.5 yields

$$\mathbb{E}_\pi\left[\sum_{n=0}^{\sigma_B-1} \Delta r(n) f(X_n)\right] = \pi(g) = \int_B \pi(\mathrm{d}x) h(x) . \tag{16.2.14}$$

Setting $Z = \sum_{n=0}^\infty \mathbb{1}_{\{n<\sigma_B\}} \Delta r(n) f(X_n)$, we have $g(x) = \mathbb{E}_x[Z]$ and

$$
\begin{aligned}
h(x) &= \mathbb{E}_x\left[\sum_{k=0}^{\infty} \mathbb{1}_{\{k<\sigma_B\}} \mathbb{E}_{X_k}[Z]\right] = \sum_{k=0}^{\infty} \mathbb{E}_x[\mathbb{1}_{\{k<\sigma_B\}} Z \circ \theta_k] \\
&= \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} \mathbb{E}_x[\mathbb{1}_{\{n+k<\sigma_B\}} \Delta r(n) f(X_{n+k})] = \sum_{j=0}^{\infty}\sum_{\ell=0}^{j} \mathbb{E}_x[\mathbb{1}_{\{j<\sigma_B\}} \Delta r(\ell) f(X_j)] \,.
\end{aligned}
$$

Hence we get

$$
h(x) = \sum_{j=0}^{\infty} \mathbb{E}_x[\mathbb{1}_{\{j<\sigma_B\}}\, r(j) f(X_j)] = \mathbb{E}_x\left[\sum_{j=0}^{\sigma_B-1} r(j) f(X_j)\right] .
$$

Therefore, using (16.2.14), we get

$$
\begin{aligned}
\mathbb{E}_\pi\left[\sum_{n=0}^{\sigma_B-1} \Delta r(n) f(X_n)\right] &= \int_B \pi(\mathrm{d}x) \mathbb{E}_x\left[\sum_{n=0}^{\sigma_B-1} r(n) f(X_n)\right] \\
&\le \pi(B) \sup_{x\in B} \mathbb{E}_x\left[\sum_{n=0}^{\sigma_B-1} r(n) f(X_n)\right] < \infty \,,
\end{aligned}
$$

since B is also (f,r)-recurrent (as an accessible (f,r)-regular set). Finally, for all $A \in \mathscr{X}_P^+$,

$$
\mathbb{E}_\pi\left[\sum_{n=0}^{\sigma_A-1} \Delta r(n) f(X_n)\right] < \infty \,.
$$

□

16.3 (f,r)-Regularity of the Skeletons

We now relate the (f,r)-regularity of the Markov kernel P and that of its skeletons P^m, $m \in \mathbb{N}^*$. We will show below that if P is irreducible and aperiodic, then P is (f,r)-regular if and only if each of its skeletons P^m is (f,r)-regular. We preface the proof by the following key technical result.

Proposition 16.3.1 *Let P be an irreducible aperiodic Markov kernel on $\mathsf{X} \times \mathscr{X}$, $f : \mathsf{X} \to [1,\infty)$ a measurable function, $r \in \bar{\Lambda}_1$, and $m \ge 2$.*

(i) Let C be an (f,r)-recurrent petite set for P. Then there exists a constant $\varsigma < \infty$ such that for all $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$
\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_{C,m}-1} r^{(m)}(k) f^{(m)}(X_{mk})\right] \le \varsigma \mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} r(k) f(X_k)\right] ,
$$

where $f^{(m)}$ is defined in (14.3.1) and $r^{(m)}(n) = r(mn)$. The set C is $(f^{(m)}, r^{(m)})$-recurrent for P^m.

(ii) There exists $\varsigma < \infty$ such that for every $C \in \mathscr{X}$ and $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right] \leq \varsigma \mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_{C,m}-1} r^{(m)}(k)f^{(m)}(X_{mk})\right].$$

If the set C is $(f^{(m)}, r^{(m)})$-recurrent for P^m, then C is (f,r)-recurrent.

Proof. Without loss of generality, we assume that $r \in \Lambda_1$.

(i) For every initial distribution $\xi \in \mathbb{M}_1(\mathscr{X})$, using that $m\sigma_{C,m}$ is a stopping time and the strong Markov property, we obtain

$$\begin{aligned}\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_{C,m}-1} r(mk)f^{(m)}(X_{mk})\right] &\leq \sum_{i=0}^{m-1}\sum_{k=0}^{\infty} r(mk+i)\mathbb{E}_\xi\left[f(X_{mk+i})\mathbb{1}\{mk < m\sigma_{C,m}\}\right] \\ &= \mathbb{E}_\xi\left[\sum_{k=0}^{m\sigma_{C,m}-1} r(k)f(X_k)\right]. \end{aligned} \tag{16.3.1}$$

Since by construction $m\sigma_{C,m} \leq \vartheta_{C,m}$ (see (14.3.4)), (16.3.1) yields

$$\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_{C,m}-1} r(mk)f^{(m)}(X_{mk})\right] \leq \mathbb{E}_\xi\left[\sum_{k=0}^{\vartheta_{C,m}-1} r(k)f(X_k)\right]. \tag{16.3.2}$$

The set C being petite and P aperiodic, C is $(r, \varepsilon\nu)$-small by Theorem 9.4.10. Without loss of generality, we may assume that $\nu(C) > 0$ (see Lemma 9.1.6). By Lemma 14.3.1, there exists $q > 0$ such that

$$\inf_{x\in C} \mathbb{P}_x(\vartheta_{C,m} \leq q) > 0. \tag{16.3.3}$$

We apply Theorem 16.2.2 with $\rho = \vartheta_{C,m}$ and $\tau = \sigma_C^{(q)}$. Lemma 14.3.1 (i) implies (16.2.1). Since C is an (f,r)-recurrent set, we get for all $x \in C$, $\mathbb{P}_x(\sigma_C^{(q)} < \infty) = 1$, and thus $\mathbb{P}_x(\tau < \infty, X_\tau \in C) = \mathbb{P}_x(\tau < \infty) = 1$. Moreover, using $\tau \geq q$ and by (16.3.3), we obtain $\inf_{x\in C} \mathbb{P}_x(\vartheta_{C,m} \leq \tau) \geq \inf_{x\in C} \mathbb{P}_x(\vartheta_{C,m} \leq q) > 0$, proving (16.2.2). Theorem 16.2.2 shows that there exist constants $\varsigma_1, \varsigma_2 < \infty$ such that

$$\begin{aligned}\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\vartheta_{C,m}-1} r(k)f(X_k)\right] &\le \varsigma_1 \sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C^{(q)}-1} r(k)f(X_k)\right] \\ &\le \varsigma_2 \sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right],\end{aligned}$$

where the last inequality follows from Lemma 14.2.2. Using (16.3.2) with $\xi=\delta_x$ and taking the supremum on C, we get

$$\begin{aligned}\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_{C,m}-1} r(mk)f^{(m)}(X_{mk})\right] &\le \sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\vartheta_{C,m}-1} r(k)f(X_k)\right] \\ &\le \varsigma_2 \sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right] < \infty\,. \qquad (16.3.4)\end{aligned}$$

Therefore, the set C is $(f^{(m)},r^{(m)})$-recurrent for P^m. Note that by the strong Markov property and by $\vartheta_{C,m}\le\sigma_C+\mathbb{1}_{\{\sigma_C<\infty\}}\vartheta_{C,m}\circ\theta_{\sigma_C}$,

$$\begin{aligned}&\mathbb{E}_\xi\left[\sum_{k=0}^{\vartheta_{C,m}-1} r(k)f(X_k)\right] \\ &\le \mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right] + \mathbb{E}_\xi\left[\mathbb{1}_{\{\sigma_C<\infty\}} r(\sigma_C)\left\{\sum_{k=0}^{\vartheta_{C,m}-1} r(k)f(X_k)\right\}\circ\theta_{\sigma_C}\right] \\ &\le \mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right] + \mathbb{E}_\xi\left[r(\sigma_C)\right]\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\vartheta_{C,m}-1} r(k)f(X_k)\right].\end{aligned}$$

Combining this with (16.3.4) and (16.3.2), there exists a constant $\varsigma<\infty$ such that for all $\xi\in\mathbb{M}_1(\mathscr{X})$,

$$\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_{C,m}-1} r^{(m)}(k)f^{(m)}(X_{mk})\right] \le \varsigma\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right].$$

(ii) By the Markov property, using that $m\sigma_{C,m}$ is a stopping time, we get

$$\begin{aligned}\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_{C,m}-1} r^{(m)}(k)f^{(m)}(X_{mk})\right] &= \sum_{k=0}^{\infty}\mathbb{E}_\xi\left[\mathbb{1}\left\{km<m\sigma_{C,m}\right\} r(km)\sum_{j=0}^{m-1}P^j f(X_{km})\right] \\ &= \sum_{k=0}^{\infty}\mathbb{E}_\xi\left[\mathbb{1}\left\{km<m\sigma_{C,m}\right\} r(km)\sum_{j=0}^{m-1} f(X_{km+j})\right].\end{aligned}$$

Using that for $j\in\{0,\dots,m-1\}$, $r^{-1}(m)r(km+j)\le r(km)$, we get

$$
\begin{aligned}
&\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_{C,m}-1} r^{(m)}(k) f^{(m)}(X_{mk})\right] \\
&\quad \geq r^{-1}(m) \sum_{k=0}^{\infty} \mathbb{E}_\xi\left[\mathbb{1}\left\{km < m\sigma_{C,m}\right\} \sum_{j=0}^{m-1} r(km+j) f(X_{km+j})\right] \\
&\quad = r^{-1}(m) \mathbb{E}_\xi\left[\sum_{k=0}^{m\sigma_{C,m}-1} r(k) f(X_k)\right] \geq r^{-1}(m) \mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} r(k) f(X_k)\right] .
\end{aligned}
$$

□

Theorem 16.3.2. *Let P be an irreducible aperiodic Markov kernel on $\mathsf{X} \times \mathscr{X}$, $f : \mathsf{X} \to [1,\infty)$ a measurable function, $r \in \bar{\Lambda}_1$, and $m \geq 2$.*

(i) A set C is accessible and (f,r)-regular if and only if C is accessible and $(f^{(m)}, r^{(m)})$-regular for P^m.
(ii) The Markov kernel P is (f,r)-regular if and only if P^m is $(f^{(m)}, r^{(m)})$-regular.
(iii) A probability measure ξ is (f,r)-regular for P if and only if ξ is $(f^{(m)}, r^{(m)})$-regular for P^m.

Proof. (i) Assume first that C is an accessible (f,r)-regular set. By Theorem 16.2.4, the set C is petite (and hence small, since P is aperiodic) and (f,r)-recurrent, i.e., $\sup_{x\in C} \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k) f(X_k)\right] < \infty$. By Theorem 9.3.11 (iii), the set C is accessible and small for P^m. By Proposition 16.3.1, there exists $\varsigma < \infty$ such that for all $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$
\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_{C,m}-1} r^{(m)}(k) f^{(m)}(X_{mk})\right] \leq \varsigma \mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} r(k) f(X_k)\right] .
$$

Setting $\xi = \delta_x$ and taking the supremum over $x \in C$, we get that

$$
\sup_{x\in C} \mathbb{E}_x\left[\sum_{k=0}^{\sigma_{C,m}-1} r^{(m)}(k) f^{(m)}(X_{mk})\right] < \infty .
$$

Thus C is accessible, small, and $(f^{(m)}, r^{(m)})$-recurrent. It is thus accessible and $(f^{(m)}, r^{(m)})$-regular by Theorem 16.2.4.

Conversely, assume that the set C is accessible and $(f^{(m)}, r^{(m)})$-regular for P^m. By Theorem 16.2.4, C is a nonempty petite and hence small $(f^{(m)}, r^{(m)})$-recurrent set for P^m, i.e.,

$$
\sup_{x\in C} \mathbb{E}_x\left[\sum_{k=0}^{\sigma_{C,m}-1} r^{(m)}(k) f^{(m)}(X_{mk})\right] < \infty .
$$

Obviously, the set C is an accessible small set for P. The set C is (f,r)-recurrent for P by proposition 16.3.1 (ii). Hence the set C is small (f,r)-recurrent. It is (f,r)-regular by Theorem 16.2.4.

(ii) The Markov kernel P is (f,r)-regular if and only if there exists an accessible (f,r)-regular set C for P. Such a set is also accessible and $(f^{(m)}, r^{(m)})$-regular for P^m. The proof follows from (i).

(iii) By Theorem 16.2.6 (a), a probability measure ξ is (f,r)-regular if and only if there exists a nonempty petite set C such that $\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} r(k) f(X_k)\right] < \infty$. Proposition 16.3.1 (i) shows that $\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_{C,m}-1} r^{(m)}(k) f(X_{mk})\right] < \infty$. Thus by Theorem 16.2.6 (a), ξ is $(f^{(m)}, r^{(m)})$-regular.

Conversely, if ξ is $(f^{(m)}, r^{(m)})$-regular for P^m, then there exists a nonempty petite set C for P^m such that $\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_{C,m}-1} r^{(m)}(k) f^{(m)}(X_{mk})\right] < \infty$. Clearly, C is petite for P, and by Proposition 16.3.1 (ii), C is f-regular. By Theorem 16.2.6 (a), ξ is (f,r)-regular. □

16.4 (f,r)-Regularity of the Split Kernel

Proposition 16.4.1 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let C be a $(1, 2\varepsilon\nu)$-small set with $\nu(C) = 1$ and $\inf_{x \in C} P(x, C) \geq 2\varepsilon$. Set $\check{P} = \check{P}_{\varepsilon,\nu}$. Let $f : \mathsf{X} \to [1, \infty)$ be a measurable function and r a positive sequence.*

(i) If C is (f,r)-regular for the kernel P, then $C \times \{0,1\}$ is $(\bar{f}, r)$-regular for the kernel $\check{P}$, where $\bar{f}(x,d) = sf(x)$ for all $x \in \mathsf{X}$ and $d \in (0,1)$.

(ii) If the split chain $\check{P}$ is $(\bar{f}, r)$-regular and f is bounded on C, then P is (f,r)-regular.

Proof. The proof is along the same lines as that of Proposition 14.4.1 (replacing Theorem 14.2.6 by Theorem 16.2.6). □

Theorem 16.4.2. *Let P be an irreducible recurrent kernel on $\mathsf{X} \times \mathscr{X}$. The following conditions are equivalent:*

(i) P is regular.
(ii) P is positive.

Proof. By Corollary 11.2.9, the existence of a petite and positive set is a sufficient condition for P to be positive. We now show that it is a necessary condition.

(I) Assume first that P is positive and that there exists an accessible strongly aperiodic small set C (hence P is strongly aperiodic). Set $\check{P} = \check{P}_{\varepsilon,\nu}$. Let π be the unique invariant probability of P. By Propositions 11.1.4 (ii) and 11.1.3 (i), the split chain $\check{P}$ is positive with invariant probability $\pi \otimes \mathrm{b}_\varepsilon$. By Proposition 11.1.4, $\check{\alpha} = C \times \{1\}$ is an accessible atom. Applying Proposition 6.2.8 (ii), $\check{\alpha}$ is recurrent. Then $\check{\mathbb{E}}_{\check{\alpha}}[\sigma_{\check{\alpha}}] < \infty$ by Theorem 6.4.2 (iv). Since $\check{\alpha}$ is small and $(1,1)$-recurrent, this implies that $\check{P}$ is regular, and this, in turn, implies that P is regular by Proposition 16.4.1.

(II) Assume now that P is positive and aperiodic. Denote by π the unique invariant probability measure. Let $C \in \mathscr{X}_P^+$ be a small set for P. By Theorem 9.3.11, we can actually choose m such that C is an accessible $(1, \varepsilon\nu)$-small set for P^m. By (I), P^m is regular, and hence P is regular by Theorem 16.3.2.

(III) Assume now that P is d-periodic and positive. By Theorem 9.3.6, there exists a sequence $C_0, C_1, \ldots, C_{d-1}$ of mutually disjoint accessible sets such that for $i = 0, \ldots, d-1$ and $x \in C_i$, $P(x, C_{i+1}) = 1$ with $C_d = C_0$. The restriction $P^d|C_0$ of the kernel P^d to C_0 is positive and aperiodic. Applying (II), $P^d|C_0$ is regular. Therefore, there exists a small set $C \subset C_0$ for $P^d|C_0$ such that $\sup_{x \in C} \mathbb{E}_x[\sigma_{C,d}] < \infty$, where $\sigma_{C,d} = \inf\{n \in \mathbb{N} : X_{dn} \in C\}$ (see (14.3.2)). Then C is also a small set for P and $\sup_{x \in C} \mathbb{E}_x[\sigma_C] \le \sup_{x \in C} \mathbb{E}_x[\sigma_{C,d}]$, showing that P is regular.

□

16.5 Exercises

16.1. Consider a functional autoregressive model $X_{k+1} = h(X_k) + Z_{k+1}$, where $h : \mathbb{R} \to \mathbb{R}$ is a measurable function, $\{Z_k,\ k \in \mathbb{N}^*\}$ is a sequence of i.i.d. integrable random variables, independent of X_0. We set $m = \mathbb{E}[|Z_1|]$ and make the following assumptions:

(i) There exist $\ell > m$ and $M < \infty$ such that for $|x| \ge M$, $|h(x)| \le |x| - \ell$.
(ii) $\sup_{|x| \le M} |h(x)| < \infty$.

Set $W(x) = |x|$ and $C = [-M, +M]$.

1. Show that for $x \notin C$, $PW(x) \le |h(x)| + m \le |x| - (\ell - m)$.
2. Show that for $x \in C$, $PW(x) \le |x| - (\ell - m) + \sup_{|x| \le M}\{|h(x)| - |x| + \ell\}$.
3. Set $V(x) = W(x)/(\ell - m)$. Show that $PV(x) \le V(x) - 1 + b\mathbb{1}_C(x)$, with $b < \infty$.
4. Show that for all $x \in \mathsf{X}$, $\mathbb{E}_x[\sigma_C] < \infty$, and that $\sup_{x \in C} \mathbb{E}_x[\sigma_C] < \infty$.

16.2. Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let $f : \mathsf{X} \to [1, \infty)$ be a measurable function and $\{r(n),\ n \in \mathbb{N}\} \in \bar{\mathscr{S}}$ a log-subadditive sequence (see Definition 13.1.1). Let C be a nonempty (f,r)-recurrent petite set. Show that

1. the set $\{W_{0,C}^{f,r} < \infty\}$ is full and absorbing;
2. there exists $d_0 > 0$ such that the sets $\{W_{0,C}^{f,r} \le d\}$ are accessible and petite for all $d \ge d_0$.

16.3. Let $v_0 \geq 1$ and let $\psi : [v_0, \infty] \to (0, \infty]$ be a concave increasing, continuously differentiable function on $[v_0, \infty)$ such that $\lim_{v \to \infty} \psi'(v) = 0$. Then there exists $v_1 \in [v_0, \infty)$ such that $\psi(v_1) - v_1 \psi'(v_1) > 0$. Consider the function $\phi : [1, \infty) \to [1, \infty)$ given, for $v \in [1, v_1)$, by

$$\phi(v) = 1 + \{2\psi'(v_1)(v_1 - 1) - \psi(v_1)\} \frac{v-1}{v_1 - 1} \\ + 2\{\psi(v_1) - (v_1 - 1)\psi'(v_1)\} \left(\frac{v-1}{v_1-1}\right)^{1/2} \quad (16.5.1)$$

and $\phi(v) = \psi(v)$ for $v \geq v_1$. The function ϕ is a concave increasing function, continuously differentiable on $[1, \infty)$, $\phi(1) = 1$. Moreover, the two sequences r_ϕ and r_ψ are equivalent, i.e., $\lim_{n \to \infty} r_\phi(n)/r_\psi(n) = 1$.

16.4. 1. Compute r_ϕ for $\phi(v) = v^\alpha$ with $0 < \alpha < 1$.
2. Let $\phi_0(v) = v \log^{-\delta}(v)$, where $\delta > 0$. Show that there exists a constant v_0 such that ϕ_0 is concave on $[v_0, \infty)$. Set $\phi(v) = \phi_0(v + v_0)$ and give an expression for r_ϕ.

16.5. Let $r : [0, \infty) \to (0, \infty)$ be a continuous increasing log-concave function. Define $h(x) = 1 + \int_0^x r(t) \mathrm{d}t$ and let $h^{-1} : [1, \infty) \to [0, \infty)$ be its inverse. Define the function ϕ on $[1, \infty)$ by

$$\phi(v) = \frac{1}{(h^{-1})'(v)} = r \circ h^{-1}(v) \, .$$

1. Show that $r = r_\phi$, ϕ is concave, and

$$\lim_{v \to \infty} \phi'(v) = 0 \Leftrightarrow \lim_{x \to \infty} \frac{r'(x)}{r(x)} = 0 \, .$$

2. Compute ϕ to obtain a polynomial rate $r(t) = (1 + ct)^\gamma$, $c, \gamma > 0$.
3. Compute ϕ to obtain a subexponential rate $r(t) = (1+t)^{\beta - 1} \mathrm{e}^{c\{(1+t)^\beta - 1\}}$, $\beta \in (0, 1)$, $c > 0$.

16.6. Let P be a strongly irreducible recurrent irreducible kernel on a discrete state space X. Show that if there exist $s > 0$ and $x \in \mathsf{X}$ such that $\mathbb{E}_x[\sigma_x^{s \vee 1}] < \infty$, then for all $y, z \in \mathsf{X}$, $\mathbb{E}_y[\sigma_z^{s \vee 1}] < \infty$ and $\lim_{n \to \infty} n^s \mathrm{d}_{\mathrm{TV}}(P^n(x, \cdot), P^n(y, \cdot)) = 0$.

16.6 Bibliographical Notes

The (f, r)-regularity results for subgeometric sequences are borrowed from the works of Nummelin and Tuominen (1982), Nummelin and Tuominen (1983), and Tuominen and Tweedie (1994).

The drift condition for (f, r)-recurrence $\mathrm{D}_{\mathrm{sg}}(V, \phi, b, C)$ was introduced in Douc et al. (2004a), building on earlier results in Fort and Moulines (2000), Fort (2001), Jarner and Roberts (2002), and Fort and Moulines (2003b).

Chapter 17
Subgeometric Rates of Convergence

We have seen in Chapter 11 that a recurrent irreducible kernel P on $\mathsf{X} \times \mathscr{X}$ admits a unique invariant measure that is a probability measure π if the kernel is positive. If the kernel is, moreover, aperiodic, then the iterates of the kernel $P^n(x,\cdot)$ converge to π in f-norm for π-almost all $x \in \mathsf{X}$, where f is a measurable function. In this chapter, we will establish convergence rates, which amounts to finding increasing sequences r such that $\lim_{n\to\infty} r(n) \|P^n(x,\cdot) - \pi\|_f = 0$. We will also consider the related problems of finding nonasymptotic bounds of convergence, i.e., functions $M : \mathsf{X} \to \mathbb{R}_+$ such that for all $n \in \mathbb{N}$ and $x \in \mathscr{X}$, $r(n) \|P^n(x,\cdot) - \pi\|_f \leq M(x)$. We will provide different expressions for the bound $M(x)$ either in terms of (f,r)-modulated moments of the return time to a small set $\mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} r(k) f(X_k)\right]$ or in terms of appropriately defined drift functions. We will also see the possible interplays between these different expressions of the bounds.

17.1 (f,r)-Ergodicity

We now consider subgeometric rates of convergence to the stationary distribution. The different classes of subgeometric rate sequences are defined in Section 13.1.

Definition 17.1.1 **((f,r)-ergodicity)** *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, $f : \mathsf{X} \to [1,\infty)$ a measurable function, and $r = \{r(n),\ n \in \mathbb{N}\} \in \Lambda_1$. The Markov kernel P is said to be (f,r)-ergodic if P is irreducible, positive with invariant probability π, and if there exists a full and absorbing set $S(f,r) \in \mathscr{X}$ satisfying*

$$\lim_{n\to\infty} r(n) \|P^n(x,\cdot) - \pi\|_f = 0\,, \quad \textit{for all } x \in S(f,r)\,.$$

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_17

In this section, we will derive sufficient conditions under which a Markov kernel P is (f,r)-ergodic. More precisely, we will show that if the Markov kernel P is (f,r)-regular, then P also is (f,r)-ergodic. The path to establishing these results parallels the one used for geometric ergodicity. It is based on the renewal approach for atomic Markov chains and the splitting construction. We preface the proof of the main result by a preparatory lemma, which is a subgeometric version of Lemma 15.1.2. In all this section, we use the notation introduced in Chapter 11.

Lemma 17.1.2 *Let P be an irreducible Markov kernel on $\mathsf{X}\times\mathscr{X}$, $f:\mathsf{X}\to[1,\infty)$ a measurable function, and C a $(1,\varepsilon\nu)$-small set satisfying $\nu(C)=1$ and* $\inf_{x\in C}P(x,C)\geq 2\varepsilon$. *Set $\check{\alpha}=C\times\{1\}$. Let $r\in\bar{\Lambda}_1$ be a sequence. Assume that*

$$\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1}r(k)f(X_k)\right]<\infty\,. \tag{17.1.1}$$

Then there exists $\varsigma<\infty$ such that

$$\sup_{(x,d)\in\check{C}}\check{\mathbb{E}}_{(x,d)}\left[\sum_{k=0}^{\sigma_{\check{\alpha}}}r(k)f(X_k)\right]\leq\varsigma\sup_{x\in C}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1}r(k)f(X_k)\right]\,, \tag{17.1.2}$$

and for every $\xi\in\mathbb{M}_1(\mathscr{X})$,

$$\check{\mathbb{E}}_{\xi\otimes\mathrm{b}_\varepsilon}\left[\sum_{k=0}^{\sigma_{\check{\alpha}}}r(k)f(X_k)\right]\leq\varsigma\mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1}r(k)f(X_k)\right]\,. \tag{17.1.3}$$

Proof. Without loss of generality, we assume that $r\in\Lambda_1$. Condition (17.1.1) implies that $M=\sup_{x\in C}f(x)<\infty$ and $\inf_{x\in C}\mathbb{P}_x(\sigma_C<\infty)=1$. Proposition 11.1.4 implies that $\mathbb{P}_{\check{\alpha}}(\sigma_{\check{\alpha}}<\infty)=1$ and for all $(x,d)\in\check{C}$, $\check{\mathbb{P}}_{(x,d)}(\sigma_{\check{C}}<\infty)=1$ and $\check{\mathbb{P}}_{(x,d)}(\sigma_{\check{\alpha}}<\infty)=1$. For $(x,d)\in\check{\mathsf{X}}$ such that $\check{\mathbb{P}}_{(x,d)}(\sigma_{\check{\alpha}}<\infty)=1$, we get

$$\check{\mathbb{E}}_{(x,d)}[r(\sigma_{\check{\alpha}})f(X_{\sigma_{\check{\alpha}}})]\leq Mr(1)\check{\mathbb{E}}_{(x,d)}[r(\sigma_{\check{\alpha}}-1)]\leq Mr(1)\check{\mathbb{E}}_{(x,d)}\left[\sum_{k=0}^{\sigma_{\check{\alpha}}-1}r(k)f(X_k)\right],$$

which implies that

$$\check{\mathbb{E}}_{(x,d)}\left[\sum_{k=0}^{\sigma_{\check{\alpha}}}r(k)f(X_k)\right]\leq(1+Mr(1))\check{\mathbb{E}}_{(x,d)}\left[\sum_{k=0}^{\sigma_{\check{\alpha}}-1}r(k)f(X_k)\right]\,. \tag{17.1.4}$$

On the other hand, for every $x\in C$, Proposition 11.1.2 shows that

$$\check{\mathbb{E}}_{\delta_x\otimes\mathrm{b}_\varepsilon}\left[\sum_{k=0}^{\sigma_{\check{C}}-1}r(k)f(X_k)\right]=\mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1}r(k)f(X_k)\right]. \tag{17.1.5}$$

Note also that for every nonnegative random variable Y, we get $\sup_{(x,d)\in\check{C}} \check{\mathbb{E}}_{(x,d)}[Y] \le \varsigma_\varepsilon \sup_{x\in C} \check{\mathbb{E}}_{\delta_x\otimes \mathrm{b}_\varepsilon}[Y]$ with $\varsigma_\varepsilon = \varepsilon^{-1} \vee (1-\varepsilon)^{-1}$. Applying this bound to (17.1.5) and then using that $f \ge 1$ shows that $\sup_{(x,d)\in\check{C}} \check{\mathbb{E}}_{(x,d)}[r^0(\sigma_{\check{C}})] < \infty$.

By Proposition 11.1.4 (vi), we get $\inf_{(x,d)\in\check{C}} \check{\mathbb{P}}_{(x,d)}(X_1 \in \check{\alpha}) > 0$.

We may therefore apply Theorem 16.2.3 with $A = \check{C}$, $B = \check{\alpha}$, and $q = 1$ to show that there exists a finite constant ς_0 satisfying

$$\begin{aligned}
\sup_{(x,d)\in\check{C}} \check{\mathbb{E}}_{(x,d)}\left[\sum_{k=0}^{\sigma_{\check{\alpha}}-1} r(k)f(X_k)\right] &\le \varsigma_0 \sup_{(x,d)\in\check{C}} \check{\mathbb{E}}_{(x,d)}\left[\sum_{k=0}^{\sigma_{\check{C}}-1} r(k)f(X_k)\right] \\
&\le \varsigma_0 \varsigma_\varepsilon \sup_{x\in C} \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right] .
\end{aligned}$$

Equation (17.1.2) results from (17.1.4). Noting that $\sigma_{\check{\alpha}} \le \sigma_{\check{C}} + \sigma_{\check{\alpha}} \circ \theta_{\sigma_{\check{C}}}$ on $\{\sigma_{\check{C}} < \infty\}$ and using

$$\begin{aligned}
\check{\mathbb{E}}_{\xi\otimes \mathrm{b}_\varepsilon}\left[\sum_{k=0}^{\sigma_{\check{\alpha}}} r(k)f(X_k)\right] &\le \check{\mathbb{E}}_{\xi\otimes \mathrm{b}_\varepsilon}\left[\sum_{k=0}^{\sigma_{\check{C}}-1} r(k)f(X_k)\right] + \check{\mathbb{E}}_{\xi\otimes \mathrm{b}_\varepsilon}\left[\sum_{k=\sigma_{\check{C}}}^{\sigma_{\check{\alpha}}\circ\theta_{\sigma_{\check{C}}}} r(k)f(X_k)\right] \\
&\le \check{\mathbb{E}}_{\xi\otimes \mathrm{b}_\varepsilon}\left[\sum_{k=0}^{\sigma_{\check{C}}-1} r(k)f(X_k)\right] + \check{\mathbb{E}}_{\xi\otimes \mathrm{b}_\varepsilon}\left[r(\sigma_{\check{C}})\right] \sup_{(x,d)\in\check{C}} \check{\mathbb{E}}_{(x,d)}\left[\sum_{k=0}^{\sigma_{\check{\alpha}}} r(k)f(X_k)\right] \\
&= \mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right]\left\{1 + r(1) \sup_{(x,d)\in\check{C}} \check{\mathbb{E}}_{(x,d)}\left[\sum_{k=0}^{\sigma_{\check{\alpha}}} r(k)f(X_k)\right]\right\} , \qquad (17.1.6)
\end{aligned}$$

we obtain (17.1.3). □

Theorem 17.1.3. *Let P be an irreducible and aperiodic Markov kernel on $\mathsf{X}\times\mathscr{X}$, $f : \mathsf{X} \to [1,\infty)$ a measurable function, and $r \in \bar{\Lambda}_1$. Assume that one of the following equivalent conditions of Theorem 16.2.6 is satisfied:*

(i) There exists a nonempty (f,r)-recurrent small set,

$$\sup_{x\in C} \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right] < \infty . \qquad (17.1.7)$$

(ii) The condition $\mathrm{D_{sg}}(\{V_n\}, f, r, b', C)$ holds for a nonempty petite set C and functions $\{V_n,\ n\in\mathbb{N}\}$ that satisfy $\sup_C V_0 < \infty$, $\{V_0 = \infty\} \subset \{V_1 = \infty\}$.

Then P is (f,r)-ergodic with unique invariant probability measure π. In addition, setting

$$M(\xi) = \mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C-1} r(k)f(X_k)\right], \qquad (17.1.8)$$

the following properties hold:

(a) *There exists a full and absorbing set $S(f,r)$ containing the set $\{V_0 < \infty\}$ (with V_0 as in (ii)) such that for all $x \in S(f,\ r)$,*

$$\lim_{n\to\infty} r(n)\,\|P^n(x,\ \cdot) - \pi\|_f = 0\,. \tag{17.1.9}$$

(b) *For every (f,r)-regular initial distribution ξ,*

$$\lim_{n\to\infty} r(n)\,\|\xi P^n - \pi\|_f = 0\,. \tag{17.1.10}$$

(c) *There exists a constant $\varsigma < \infty$ such that for all initial distributions $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$,*

$$\sum_{n=1}^{\infty} r(n)\left\|\xi P^n - \xi' P^n\right\|_f \le \varsigma\{M(\xi) + M(\xi')\}\,. \tag{17.1.11}$$

(d) *There exists $\varsigma < \infty$ such that for every initial distribution ξ and all $n \in \mathbb{N}$,*

$$r(n)\,\|\xi P^n - \pi\|_f \le \varsigma M(\xi)\,. \tag{17.1.12}$$

(e) *If $\Delta r \in \bar{\Lambda}_1$, then there exists $\varsigma < \infty$ such that for every $\xi \in \mathbb{M}_1(\mathscr{X})$,*

$$\sum_{k=1}^{\infty} \Delta r(k)\left\|\xi P^k - \pi\right\|_f \le \varsigma \mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C - 1} \Delta r(k) f(X_k)\right]\,. \tag{17.1.13}$$

Equations (17.1.11) *and* (17.1.12) *also hold with $M(\xi) = \xi(V_0) + 1$.*

Proof. Without loss of generality, we assume that $r \in \Lambda_1$. Since C is small and $\sup_{x\in C} \mathbb{E}_x[\sigma_C] < \infty$, the existence and uniqueness of the invariant probability π follow from Corollary 11.2.9.

(I) Assume first that the Markov kernel P admits a $(1,\mu)$-small set P. By Proposition 11.1.4, the set $\check{\alpha} = C \times \{1\}$ is an aperiodic atom for the split kernel $\check{P}$.

Using Lemma 17.1.2 ((17.1.2) and (17.1.3)), condition (17.1.7) implies that there exists $\varsigma_1 < \infty$ such that $\check{\mathbb{E}}_{\check{\alpha}}\left[\sum_{k=0}^{\sigma_{\check{\alpha}}} r(k) f(X_k)\right] < \infty$ and for every $\xi \in \mathbb{M}_1(\mathscr{X})$, $\check{M}(\xi) \le \varsigma_1 M(\xi)$, where

$$M(\xi) = \mathbb{E}_\xi\left[\sum_{k=0}^{\sigma_C - 1} r(k) f(X_k)\right] \quad \text{and} \quad \check{M}(\xi) = \check{\mathbb{E}}_{\xi\otimes \mathrm{b}_\varepsilon}\left[\sum_{k=1}^{\sigma_{\check{\alpha}}} r(k) f(X_k)\right]\,.$$

By Proposition 11.1.3, $\check{P}$ admits a unique invariant probability measure, which may be expressed as $\pi \otimes \mathrm{b}_\varepsilon$, where π is the unique invariant probability measure for P. Then by Lemma 17.1.2, we have $\check{M}(\xi) \le \varsigma_1 M(\xi)$, and applying Theorem 13.4.4, 13.4.6, we obtain

$$\sum_{k=1}^{\infty} r(k) \left\| (\xi \otimes \mathrm{b}_\varepsilon)\check{P}^k - (\xi' \otimes \mathrm{b}_\varepsilon)\check{P}^k \right\|_{f \otimes \mathbf{1}} \leq \varsigma_1 \varsigma_2 \{M(\xi) + M(\xi')\} .$$

The proof of (17.1.11) follows from Lemma 11.1.1, which implies

$$\left\| \xi P^k - \xi' P^k \right\|_f \leq \left\| [\xi \otimes \mathrm{b}_\varepsilon]\check{P}^k - [\xi' \otimes \mathrm{b}_\varepsilon]\check{P}^k \right\|_{f \otimes \mathbf{1}} . \tag{17.1.14}$$

The bound (17.1.12) and the limit (17.1.10) are obtained similarly using Theorem 13.4.4, Equations (13.4.7) and (13.4.8) by applying (17.1.14) and Proposition 11.1.3.

The bound (17.1.13) is a consequence of Theorem 13.4.4, (iv).

(II) The method to extend the result from the strongly aperiodic case to the general aperiodic case is exactly along the same lines as in Theorem 15.1.3. Using Proposition 16.3.1 instead of Proposition 14.3.2 in the derivations, we get that there exists a constant $\varsigma < \infty$ such that for all $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{k=1}^{\infty} r(mk) \left\| \xi P^{mk} - \xi' P^{mk} \right\|_{f^{(m)}} \leq \varsigma \{M(\xi) + M(\xi')\} , \tag{17.1.15}$$

where $f^{(m)} = \sum_{i=0}^{m-1} P^i f$. For $i \in \{0, \ldots, m-1\}$ and $|g| \leq f$, we have $|P^i g| \leq f^{(m)}$, whence for $k \geq 0$,

$$\sup_{|g| \leq f} |\xi P^{mk+i} g - \xi' P^{mk+i} g| \leq \sup_{|h| \leq f^{(m)}} |\xi P^{mk} h - \xi' P^{mk} h| .$$

This yields $\left\| \xi P^{mk+i} - \xi' P^{mk+i} \right\|_f \leq \left\| \xi P^{mk} - \xi' P^{mk} \right\|_{f^{(m)}}$. Since the sequence r is increasing and log-subadditive, we obtain

$$\begin{aligned} \sum_{k=1}^{\infty} r(k) \left\| \xi P^k - \xi' P^k \right\|_f &\leq \sum_{i=0}^{m-1} \sum_{k=0}^{\infty} r(mk+i) \left\| \xi P^{mk+i} - \xi' P^{mk+i} \right\|_f \\ &\leq m r(m) \sum_{k=0}^{\infty} r(mk) \left\| \xi P^{mk} - \xi' P^{mk} \right\|_{f^{(m)}} , \end{aligned}$$

which concludes the proof.

It remains to prove (a). Let C be an (f,r)-recurrent small set. Denote by $S_P(f,r)$ the set of (f,r)-regular points. For all $x \in S_P(f,r)$, δ_x is (f,r)-regular, and hence (17.1.10) implies that $\lim_{n \to \infty} r(n) \left\| P^n(x, \cdot) - \pi \right\|_f = 0$. Theorem 16.2.6 shows that the set $S_P(f,r)$ is full and absorbing and contains the set $\{V_0 < \infty\}$, where V_0 is as in (ii). □

We now specialize these results to total variation convergence and extend the results introduced in Theorem 13.3.3 to aperiodic $(1,r)$-regular kernels.

Corollary 17.1.4 *Let P be an irreducible and aperiodic Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that there exist a sequence $r \in \bar{\Lambda}_1$ and a small set C such that*

$$\sup_{x \in C} \mathbb{E}_x \left[r^0(\sigma_C) \right] < \infty \,, \quad \text{where } r^0(n) = \sum_{k=0}^{n} r(k) \,.$$

Then the kernel P admits a unique invariant probability π. Moreover, we have the following:

(i) If $\Delta r \in \bar{\Lambda}_1$ and either $\lim_{n \to \infty} \uparrow r(n) = \infty$ and $\mathbb{E}_\xi [r(\sigma_C)] < \infty$ or $\lim_{n \to \infty} r(n) < \infty$ and $\mathbb{P}_\xi (\sigma_C < \infty) = 1$, then

$$\lim_{n \to \infty} r(n) \left\| \xi P^n - \pi \right\|_{\mathrm{TV}} = 0 \,. \tag{17.1.16}$$

Moreover, there exists a set $S \in \mathscr{X}$ such that $\pi(S) = 1$ and for all $x \in S$,

$$\lim_{n \to \infty} r(n) \left\| P^n(x, \cdot) - \pi \right\|_{\mathrm{TV}} = 0 \,. \tag{17.1.17}$$

(ii) If $\Delta r \in \bar{\Lambda}_1$, then there exists $\varsigma < \infty$ such that for every initial distribution ξ and all $n \in \mathbb{N}$,

$$r(n) \left\| \xi P^n - \pi \right\|_{\mathrm{TV}} \le \varsigma \mathbb{E}_\xi \left[r(\sigma_C) \right] \,. \tag{17.1.18}$$

(iii) There exists $\varsigma < \infty$ such that for every initial distribution $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{k=1}^{\infty} r(k) \left\| \xi P^k - \xi' P^k \right\|_{\mathrm{TV}} \le \varsigma \left\{ \mathbb{E}_\xi \left[r^0(\sigma_C) \right] + \mathbb{E}_{\xi'} \left[r^0(\sigma_C) \right] \right\} \,. \tag{17.1.19}$$

(iv) If $\Delta r \in \bar{\Lambda}_1$, then there exists $\varsigma < \infty$ such that for every $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{k=1}^{\infty} \Delta r(k) \left\| \xi P^k - \pi \right\|_{\mathrm{TV}} \le \varsigma \mathbb{E}_\xi \left[r(\sigma_C) \right] \,. \tag{17.1.20}$$

Proof. (a) Equations (17.1.19) and (17.1.20) follow from (17.1.11) and (17.1.13) with $f \equiv 1$.

(b) The proof of (17.1.16) and (17.1.18) requires more attention. Indeed, setting $f \equiv 1$ in (17.1.12) shows that there exists $\varsigma < \infty$ such that $r(n) \left\| \xi P^n - \pi \right\|_{\mathrm{TV}} \le \varsigma \mathbb{E}_\xi [r^0(\sigma_C)]$, which is not the desired result. To obtain (17.1.18), we will use Theorem 13.3.3 instead of Theorem 13.4.4. Assume first that P admits a $(1, \mu)$-small set P. By Proposition 11.1.4, the set $\check{\alpha} = C \times \{1\}$ is an aperiodic atom for the split kernel $\check{P}$. Applying Lemma 17.1.2, (17.1.2), with $f \equiv 1$ implies that $\check{\mathbb{E}}_{\check{\alpha}} \left[r^0(\sigma_{\check{\alpha}}) \right] < \infty$. Applying Lemma 15.1.2, (17.1.3), with $f \equiv 1$ and the sequence Δr shows that there exists $\varsigma < \infty$ such that for every $\xi \in \mathbb{M}_1(\mathscr{X})$, $\check{\mathbb{E}}_{\xi \otimes \mathrm{b}_\varepsilon} [r(\sigma_{\check{\alpha}})] \le \varsigma \mathbb{E} [r(\sigma_C)]$. Equations (17.1.16 and 17.1.18) follow from Theorem 13.3.3 (ii) and (iii).

Assume now that the Markov kernel admits an $(m,\varepsilon\nu)$-small set C. By Lemma 9.1.6, we may assume without loss of generality that $\nu(C)=1$. Theorem 9.3.11 shows that C is an accessible strongly aperiodic small set for the kernel P^m. Applying the result above to the kernel P^m shows that there exists $\varsigma_1<\infty$ such that for every $\xi\in\mathbb{M}_1(\mathscr{X})$ and $k\in\mathbb{N}$, $r^{(m)}(k)\left\|\xi P^{mk}-\pi\right\|_{\mathrm{TV}}\leq\varsigma_1\mathbb{E}_\xi[r(\sigma_{C,m})]$. Since $\|\xi P-\xi' P\|_{\mathrm{TV}}\leq\|\xi-\xi'\|_{\mathrm{TV}}$ and for $n=mk+q$, $q\in\{0,\ldots,m-1\}$, $r(n)\leq r^{(m)}(k)r(m-1)$, we get that

$$r(n)\,\|\xi P^n-\pi\|_{\mathrm{TV}}\leq r(m-1)\varsigma_1\mathbb{E}_\xi[r(\sigma_{C,m})]\,.$$

By applying Proposition 16.3.1 to the sequence Δr and $f\equiv 1$, there exists $\varsigma_2<\infty$ such that for all $\xi\in\mathbb{M}_1(\mathscr{X})$, $\mathbb{E}_\xi[r(\sigma_{C,m})]\leq\varsigma_2\mathbb{E}_\xi[r(\sigma_C)]$. The proof of (17.1.18) follows. The proof of (17.1.16) is along the same lines.

When $\lim_{n\to\infty}r(n)=\infty$, (17.1.17) follows from (17.1.16) by Corollary 9.2.14, which shows that the set $S:=\{x\in\mathsf{X}\,:\,\mathbb{E}_x[r(\sigma_C)]<\infty\}$ is full and absorbing. Since by Theorem 9.2.15 an invariant probability measure is a maximal irreducibility measure, $\pi(S)=1$. When $\limsup r(n)<\infty$, we set $S=\{x\in\mathsf{X}\,:\,\mathbb{P}_x(\sigma_C<\infty)=1\}$. Theorem 10.1.10 shows that this set is full and absorbing, and hence $\pi(S)=1$. □

Example 17.1.5 (Backward recurrence time chain). Let $\{p_n,\ n\in\mathbb{N}\}$ be a sequence of positive real numbers such that $p_0=1$, $p_n\in(0,1)$ for all $n\geq 1$ and $\lim_{n\to\infty}\prod_{i=1}^n p_i=0$. Consider the backward recurrence time chain with transition kernel P defined as $P(n,n+1)=1-P(n,0)=p_n$, for all $n\geq 0$. The Markov kernel P is irreducible and strongly aperiodic, and $\{0\}$ is an atom. Let σ_0 be the return time to $\{0\}$. We have for all $n\geq 1$,

$$\mathbb{P}_0(\sigma_0=n+1)=(1-p_n)\prod_{j=0}^{n-1}p_j \quad\text{and}\quad \mathbb{P}_0(\sigma_0>n)=\prod_{j=0}^{n-1}p_j\,.$$

By Theorem 7.2.1, the Markov kernel P is positive recurrent if and only if $\mathbb{E}_0[\sigma_0]<\infty$, i.e.,

$$\sum_{n=1}^{\infty}\prod_{j=1}^{n}p_j<\infty,$$

and the stationary distribution π is given by $\pi(0)=\pi(1)=1/\mathbb{E}_0[\sigma_0]$ and for $j\geq 2$,

$$\pi(j)=\frac{\mathbb{E}_0\left[\sum_{k=1}^{\sigma_0}\mathbb{1}_{\{X_k=j\}}\right]}{\mathbb{E}_0[\sigma_0]}=\frac{\mathbb{P}_0(\sigma_0\geq j)}{\mathbb{E}_0[\sigma_0]}=\frac{p_0\cdots p_{j-2}}{\sum_{n=1}^{\infty}p_1\cdots p_n}.$$

Because the distribution of the return time to the atom $\{0\}$ has such a simple expression in terms of the transition probability $\{p_n,\ n\in\mathbb{N}\}$, we are able to exhibit the largest possible rate function r such that the $(1,r)$ modulated moment of the return time $\mathbb{E}_0\left[\sum_{k=0}^{\sigma_0-1}r(k)\right]$ is finite. We will also prove that the drift condition $D_{\mathrm{sg}}(V,\phi,b)$ holds for appropriately chosen functions V and ϕ and yields the optimal rate of convergence. Note also that for every function h,

$$\mathbb{E}_0\left[\sum_{k=0}^{\sigma_0-1} h(X_k)\right] = \mathbb{E}_0\left[\sum_{k=0}^{\sigma_0-1} h(k)\right].$$

Therefore, there is no loss of generality to consider only $(1,r)$ modulated moments of the return time to zero.

If $\sup_{n\ge 1} p_n \le \lambda < 1$, then for all $\lambda < \mu < 1$, $\mathbb{E}_0[\mu^{-\sigma_0}] < \infty$, and $\{0\}$ is thus a geometrically recurrent atom. Subgeometric rates of convergence in total variation norm are obtained when $\limsup_{n\to\infty} p_n = 1$. Depending on the rate at which the sequence $\{p_n,\ n \in \mathbb{N}\}$ approaches 1, different behaviors can be obtained covering essentially the three typical rates (polynomial, logarithmic, and subexponential) discussed above.

Polynomial rates: Assume that for $\theta > 0$ and large n, $p_n = 1-(1+\theta)n^{-1}$. Then $\prod_{i=1}^n p_i \asymp n^{-1-\theta}$. Thus $\mathbb{E}_0\left[\sum_{k=0}^{\sigma_0-1} r(k)\right] < \infty$ if and only if $\sum_{k=1}^\infty r(k)k^{-1-\theta} < \infty$. For instance, $r(n) = n^\beta$ with $0 \le \beta < \theta$ is suitable.

Subgeometric rates: If for large n, $p_n = 1-\theta\beta n^{\beta-1}$ for $\theta > 0$ and $\beta \in (0,1)$, then $\prod_{i=1}^n p_i \asymp \mathrm{e}^{-\theta n^\beta}$. Thus $\mathbb{E}_0[\sum_{k=0}^{\sigma_0-1} \mathrm{e}^{ak^\beta}] < \infty$ if $a < \theta$ and $\mathbb{E}_0[\sum_{k=0}^{\sigma_0-1} \mathrm{e}^{ak^\beta}] = \infty$ if $a \ge \theta$.

Logarithmic rates: If for $\theta > 0$ and large n, $p_n = 1-1/n-(1+\theta)/(n\log(n))$, then $\prod_{j=1}^n p_j \asymp n^{-1}\log^{-1-\theta}(n)$, which is a summable series. Hence if r is nondecreasing and $\sum_{k=1}^\infty r(k)\prod_{j=1}^n p_j < \infty$, then $r(k) = o(\log^\theta(k))$. In particular, $r(k) = \log^\beta(k)$ is suitable for all $0 \le \beta < \theta$.

◀

17.2 Drift Conditions

We will now translate this result into terms of the drift condition $\mathrm{D_{sg}}(V,\phi,b,C)$, where $\phi : [1,\infty) \to (0,\infty)$ is a concave increasing differentiable function. Recall that H_ϕ denotes the primitive of $1/\phi$ that cancels at 1, $H_\phi(v) = \int_1^v \mathrm{d}x/\phi(x)$ (see Equation (16.1.12). It is an increasing concave differentiable function on $[1,\infty)$, and $\lim_{x\to\infty} \uparrow H_\phi(x) = \infty$. The inverse $H_\phi^{-1} : [0,\infty) \to [1,\infty)$ is also an increasing and differentiable function. Finally, we set $r_\phi(t) = (H_\phi^{-1})'(t) = \phi \circ H_\phi^{-1}(t)$. (see Equation (16.1.13)).

Theorem 17.2.1. *Let P be an irreducible and aperiodic Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that $\mathrm{D_{sg}}(V,\phi,b,C)$ holds for some small set C satisfying $\sup_C V < \infty$. Then P has a unique invariant probability measure π, and for all $x \in \mathsf{X}$,*

$$\lim_{n\to\infty} r_\phi(n)\,\|P^n(x,\cdot)-\pi\|_{\mathrm{TV}} = 0\,, \quad \lim_{n\to\infty} \|P^n(x,\cdot)-\pi\|_{\phi\circ V} = 0\,. \tag{17.2.1}$$

There exists a constant $\varsigma < \infty$ such that for all initial conditions $\xi, \xi' \in \mathbb{M}_!(\mathscr{X})$,

$$\sum_{n=0}^{\infty} r_\phi(n) \left\| \xi P^n - \xi' P^n \right\|_{\mathrm{TV}} \leq \varsigma \left\{ \xi(V) + \xi'(V) + 2b r_\phi(1)/r_\phi(0) \right\} \tag{17.2.2}$$

$$\sum_{n=0}^{\infty} \left\| \xi P^n - \xi' P^n \right\|_{\phi \circ V} \leq \varsigma \left\{ \xi(V) + \xi'(V) + 2b \right\} . \tag{17.2.3}$$

Proof. By Proposition 16.1.11, $\mathrm{D_{sg}}(V, \phi, b, C)$ implies that $\mathrm{D_{sg}}(\{V_n\}, 1, r_\phi, b', C)$ holds with $V_n = H_n \circ V$, $H_n = H_\phi^{-1}(n + H_\phi) - H_\phi^{-1}(n)$, and $b' = b r_\phi(1)/r_\phi^2(0)$. Moreover, $\mathrm{D_{sg}}(V, \phi, b, C)$ also implies that $\mathrm{D_{sg}}(\{V_n\}, 1, \phi \circ V, b, C)$ holds with $V_n = V$ for all $n \in \mathbb{N}$. The result then follows from Theorem 17.1.3 combined with Theorem 16.1.12. □

Example 17.2.2 (Backward recurrence time chain; Example 17.1.5 (continued)). We consider again the backward recurrence time chain, but this time we will use $\mathrm{D_{sg}}(V, \phi, b, C)$. For $\gamma \in (0,1)$ and $x \in \mathbb{N}^*$, define $V(0) := 1$ and $V(x) := \prod_{j=0}^{x-1} p_j^{-\gamma}$. Then for all $x \geq 0$, we have

$$\begin{aligned} PV(x) &= p_x V(x+1) + (1 - p_x) V(0) = p_x^{1-\gamma} V(x) + 1 - p_x \\ &\leq V(x) - (1 - p_x^{1-\gamma}) V(x) + 1 - p_x \,. \end{aligned}$$

Thus for $0 < \delta < 1 - \gamma$ and large enough x,

$$PV(x) \leq V(x) - \delta (1 - p_x) V(x). \tag{17.2.4}$$

Polynomial rates: Assume that $p_n = 1 - (1+\theta) n^{-1}$ for some $\theta > 0$. Then $V(x) \asymp x^{\gamma(1+\theta)}$ and $(1 - p_x) V(x) \asymp V(x)^{1 - 1/(\gamma(1+\theta))}$. Thus condition $\mathrm{D_{sg}}(V, \phi, b)$ holds with $\phi(v) = c v^\alpha$ for $\alpha = 1 - 1/(\gamma(1+\theta))$ for every $\gamma \in (0,1)$. Theorem 17.2.1 yields the rate of convergence in total variation distance $n^{\alpha/(1-\alpha)} = n^{\gamma(1+\theta)-1}$, i.e., n^β for all $0 \leq \beta < \theta$.

Subgeometric rates: Assume that $p_n = 1 - \theta \beta n^{\beta - 1}$ for some $\theta > 0$ and $\beta \in (0,1)$. Then for large enough x, (17.2.4) yields

$$PV(x) \leq V(x) - \theta \beta \delta x^{\beta - 1} V(x) \leq c V(x) \{\log(V(x))\}^{1 - 1/\beta} ,$$

for $c < \theta^{1/\beta} \beta \delta$. Defining $\alpha := 1/\beta - 1$, Theorem 17.2.1 yields the following rate of convergence in total variation distance:

$$n^{-\alpha/(1+\alpha)} \exp\left(\{c(1+\alpha) n\}^{1/(1+\alpha)} \right) = n^{\beta - 1} \exp\left(\theta \delta^\beta n^\beta \right) .$$

Since δ is arbitrarily close to 1, we recover the fact that $\mathbb{E}_0[\sum_{k=0}^{\sigma_0 - 1} \mathrm{e}^{a k^\beta}] < \infty$ for every $a < \theta$.

Logarithmic rates: Assume finally that $p_n = 1 - n^{-1} - (1+\theta)n^{-1}\log^{-1}(n)$ for some $\theta > 0$. Choose $V(x) := \left(\prod_{j=0}^{x-1} p_j\right) / \log^{\varepsilon}(x)$ for $\varepsilon > 0$ arbitrarily small. Then for constants $c < c' < c'' < 1$ and large x, we obtain

$$\begin{aligned} PV(x) &= \frac{\log^{\varepsilon}(x)}{\log^{\varepsilon}(x+1)} V(x) + 1 - p_x = V(x) - c''\varepsilon \frac{V(x)}{x\log(x)} + 1 - p_x \\ &\le V(x) - c'\varepsilon \log^{\theta-\varepsilon}(x) \le V(x) - c\varepsilon \log^{\theta-\varepsilon}(V(x)). \end{aligned}$$

Here again Theorem 17.2.1 yields the optimal rate of convergence.

◀

Example 17.2.3 (Independent Metropolis–Hastings sampler). Suppose that $\pi \in \mathbb{M}_1(\mathscr{X})$ and let $Q \in \mathbb{M}_1(\mathscr{X})$ be another probability measure such that π is absolutely continuous with respect to Q with Radon–Nikodym derivative

$$\frac{d\pi}{dQ}(x) = \frac{1}{q(x)} \quad \text{for } x \in \mathsf{X}. \tag{17.2.5}$$

If the chain is currently at $x \in \mathsf{X}$, a move is proposed to y drawn from Q and accepted with probability

$$\alpha(x,y) = \frac{q(x)}{q(y)} \wedge 1 \,. \tag{17.2.6}$$

If the proposed move is not accepted, the chain remains at x. Denote by P the Markov kernel associated with the independence sampler. It can easily be verified that the chain is irreducible and has unique stationary measure π. If π and Q both have densities denoted by π and q, respectively, with respect to some common reference measure and if there exists $\beta > 0$ such that

$$\frac{q(x)}{\pi(x)} \ge \varepsilon \,, \quad \text{for all } x \in \mathsf{X}, \tag{17.2.7}$$

then the independence sampler is uniformly ergodic (see Example 15.3.3), and if (17.2.7) does not hold π-almost surely, then the independence sampler is not geometrically ergodic. However, it is still possible to obtain a subgeometric rate of convergence when (17.2.7) is violated.

First consider the case that π is the uniform distribution on $[0,1]$ and Q has density q with respect to Lebesgue measure on $[0,1]$ of the form

$$q(x) = (r+1)x^r \,, \quad \text{for some } r > 0 \,. \tag{17.2.8}$$

For each $x \in [0,1]$, define the regions of acceptance and possible rejection by

$$A_x = \{y \in [0,1] \,:\, q(y) \le q(x)\} \,, \qquad R_x = \{y \in [0,1] \,:\, q(y) > q(x)\} \,.$$

We will show that for each $r < s < r+1$, the independence sampler P satisfies

$$PV \le V - cV^{\alpha} + b\mathbb{1}_C\,, \quad \text{where } V(x) = 1/x^s, \alpha = 1 - r/s \text{ and } C \text{ is a petite set.} \tag{17.2.9}$$

The acceptance and rejection regions are $A_x = [0,x]$ and $R_x = (x,1]$. Furthermore, all sets of the form $[y,1]$ are petite. Using straightforward algebra, we get

$$\begin{aligned}
PV(x) &= \int_0^x V(y)q(y)\mathrm{d}y + \int_x^1 V(y)\alpha(x,y)q(y)\mathrm{d}y \\
&\quad + V(x)\int_x^1 (1-\alpha(x,y))q(y)\mathrm{d}y \\
&= \int_0^x (r+1)y^{r-s}\mathrm{d}y + \int_x^1 \frac{1}{y^s}(r+1)x^r\mathrm{d}y + \frac{1}{x^s}\int_x^1 (r+1)(y^r - x^r)\mathrm{d}y \\
&= \frac{r+1}{r-s+1}x^{r-s+1} + \frac{r+1}{-s+1}x^r - \frac{r+1}{-s+1}x^{r-s+1} + \frac{1}{x^s} - x^{r-s+1} \\
&\quad - (r+1)x^{r-s}(1-x) \\
&= V(x) - (r+1)V(x)^{1-r/s}(1-x) + c1x^{r-s+1} + c2x^r\,.
\end{aligned}$$

Since $r-s+1$ and r are both positive, x^{r-s+1} and x^r tend to 0 as x tends to 0, while $V(x)^{1-r/s} = x^{r-s}$ tends to ∞ as x tends to 0. Thus (17.2.9) is satisfied with $C = [x_0,1]$ for x_0 sufficiently small and some constants b and c.

The choice of s leading to the best rate of convergence is $r+1-\varepsilon$, which gives $\alpha \approx 1 - r/(r+1)$. Hence the independence sampler converges in total variation at a polynomial rate of order $1/r$.

We consider the general case. For simplicity, we assume that the two probabilities π and Q are equivalent, which is no restriction. We assume that for some $r > 0$,

$$\pi(\mathscr{A}_\varepsilon) =_{\varepsilon\to 0} O(\varepsilon^{1/r}) \quad \text{where } \mathscr{A}_\varepsilon = \{x \in \mathsf{X} : \; q(x) \le \varepsilon\}. \tag{17.2.10}$$

We will show that for each $r < s < r+1$, the independence sampler P satisfies

$$PV \le V - cV^{\alpha} + b\mathbb{1}_C\,, \quad \text{where } V(x) = (1/q(x))^{s/r}, \alpha = 1 - r/s, \tag{17.2.11}$$

and C is a petite set. Note that $A_x = \mathscr{A}_{q(x)}$ and that all the sets $\mathscr{A}_\varepsilon^c$ are petite:

$$\begin{aligned}
PV(x) &= \int_{\mathscr{A}_{q(x)}} V(y)q(y)\pi(\mathrm{d}y) + \int_{\mathscr{A}^c_{q(x)}} V(y)\alpha(x,y)q(y)\pi(\mathrm{d}y) \\
&\quad + V(x)\int_{\mathscr{A}^c_{q(x)}} (1-\alpha(x,y))q(y)\pi(\mathrm{d}y) \\
&= \int_{\mathscr{A}_{q(x)}} q(y)^{1-s/r}\pi(\mathrm{d}y) + \int_{\mathscr{A}^c_{q(x)}} q(x)q(y)^{-s/r}\pi(\mathrm{d}y) \\
&\quad + V(x)\int_{\mathscr{A}^c_{q(x)}} (q(y)-q(x))\pi(\mathrm{d}y)\,.
\end{aligned}$$

Therefore, denoting by F the cumulative distribution function of q under π, we get

$$
\begin{aligned}
PV(x) &\le \int_{\mathscr{A}_{q(x)}} q(y)^{1-s/r}\pi(\mathrm{d}y) + \int_{\mathscr{A}^c_{q(x)}} q(x)q(y)^{-s/r}\pi(\mathrm{d}y) - V(x)^{\alpha}\pi(\mathscr{A}^c_{q(x)}) + V(x) \\
&= \int_{[0,q(x)]} y^{1-s/r}F(\mathrm{d}y) + \int_{(q(x),\infty)} q(x)y^{-s/r}F(\mathrm{d}y) - V(x)^{\alpha}\pi(\mathscr{A}^c_{q(x)}) + V(x)\,,
\end{aligned}
$$

where the inequality stems from $\int_{\mathscr{A}^c_{q(x)}} q(y)\pi(\mathrm{d}y) = Q(\mathscr{A}^c_{q(x)}) \le 1$. Under (17.2.11), there exist positive K and y_0 such that $F(y) \le Ky^{1/r}$ for $y \le y_0$. Since $Ky^{1/r}$ is the cumulative distribution function for the measure with density $K_1 y^{1/r-1}$ with respect to Lebesgue measure and since $y^{1-s/r}$ and $y^{-s/r}$ are decreasing functions, we get, for all $q(x) \le y_0$,

$$
\begin{aligned}
\int_{[0,q(x)]} y^{1-s/r}F(\mathrm{d}y) &\le K_1 \int_{[0,q(x)]} y^{1-s/r}y^{1/r-1}\mathrm{d}y, \\
\int_{(q(x),\infty)} q(x)y^{-s/r}F(\mathrm{d}y) &\le K_1 \int_{(q(x),y_0]} q(x)y^{-s/r}y^{1/r-1}\mathrm{d}y + \int_{(y_0,\infty)} q(x)y^{-s/r}F(\mathrm{d}y).
\end{aligned}
$$

Therefore, the two integrals in (17.2.12) both tend to 0 as $q(x)$ tends to 0. Since $V(x)^{\alpha}$ tends to ∞ and $\pi(A^c_{q(x)})$ tends to 1 as $q(x)$ tends to 0, (17.2.11) is satisfied with $C = \mathscr{A}^c_{\varepsilon}$ for ε sufficiently small. ◀

A pair of strictly increasing continuous functions (Υ, Ψ) defined on $\mathbb{R}_+$ is called a pair of inverse Young functions if for all $x, y \ge 0$,

$$
\Upsilon(x)\Psi(y) \le x + y\,. \tag{17.2.12}
$$

A typical example is $\Upsilon(x) = (px)^{1/p}$ and $\Psi(y) = (qy)^{1/q}$, where $p, q > 0$, $1/p + 1/q = 1$. Indeed, the concavity of the logarithm yields, for $x, y > 0$,

$$
\begin{aligned}
(px)^{1/p}(qy)^{1/q} &= \exp\{p^{-1}\log(px) + q^{-1}\log(qx)\} \\
&\le \exp\log\{px/p + qy/q\} = x + y\,.
\end{aligned}
$$

The use of inverse Young functions allows us to obtain a tradeoff between rates and f-norm using the following interpolation lemma.

Lemma 17.2.4 *Let (Υ, Φ) be a pair of inverse Young functions, r a sequence of nonnegative real numbers, and $f \in \mathbb{F}_+(\mathsf{X})$. Then for all $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$ and all $k \in \mathbb{N}$,*

$$
\Upsilon(r(k))\left\|\xi - \xi'\right\|_{\Psi(f)} \le r(k)\left\|\xi - \xi'\right\|_{\mathrm{TV}} + \left\|\xi - \xi'\right\|_f\,.
$$

Proof. The proof follows from

$$
\begin{aligned}
\Upsilon(r(k))\left\|\xi - \xi'\right\|_{\Psi(f)} &= \int |\xi - \xi'|(\mathrm{d}x)\,[\Upsilon \circ r(k)\,\Psi \circ f(x)] \\
&\le \int |\xi - \xi'|(\mathrm{d}x)\,[r(k) + f(x)] = r(k)\left\|\xi - \xi'\right\|_{\mathrm{TV}} + \left\|\xi - \xi'\right\|_f\,.
\end{aligned}
$$

□

We now extend the previous results to weighted total variation distances by interpolation using Young functions.

Theorem 17.2.5. *Let P be an irreducible and aperiodic Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that $\mathrm{D_{sg}}(V, \phi, b, C)$ holds for some small set C satisfying $\sup_C V < \infty$. Let (Υ, Ψ) be a pair of inverse Young functions. Then there exists an invariant probability measure π, and for all $x \in \mathsf{X}$,*

$$\lim_{n\to\infty} \Upsilon(r_\phi(n)) \quad \|P^n(x,\cdot) - \pi\|_{\Psi(\phi\circ V)} = 0. \tag{17.2.13}$$

There exists a constant $\varsigma < \infty$ such that for all initial conditions $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{n=0}^{\infty} \Upsilon(r_\phi(n)) \left\|\xi P^n - \xi' P^n\right\|_{\Psi(\phi\circ V)} \le \varsigma \left(\xi(V) + \xi'(V) + 2b\{1 + r_\phi(1)/r_\phi(0)\}\right) . \tag{17.2.14}$$

Proof. Lemma 17.2.4 shows that for all $x \in \mathsf{X}$ and $k \in \mathbb{N}$,

$$\Upsilon(r(k)) \|P^n(x,\cdot) - \pi\|_{\Psi(\phi\circ V)} \le r(k) \|\xi P^n(x,\cdot) - \pi\|_{\mathrm{TV}} + \|P^n(x,\cdot) - \phi\|_{\phi\circ V} .$$

The proof of (17.2.13) follows from Theorem 17.2.1, (17.2.1). Equation 17.2.14 follows similarly from Theorem 17.2.1, ((17.2.2), (17.2.3)). $\square$

We provide below some examples of rates of convergence obtained using Theorem 17.2.5. We assume in this discussion that P is an irreducible and aperiodic Markov kernel on $\mathsf{X} \times \mathscr{X}$ and that $\mathrm{D_{sg}}(V, \phi, b, C)$ holds for some small set C satisfying $\sup_C V < \infty$.

Polynomial rates of convergence are associated with the functions $\phi(v) = cv^\alpha$ for some $\alpha \in [0,1)$ and $c \in (0,1]$. The rate of convergence in total variation distance is $r_\phi(n) \propto n^{\alpha/(1-\alpha)}$. Set $\Upsilon(x) = ((1-p)x)^{(1-p)}$ and $\Psi(x) = (px)^p$ for some p, $0 < p < 1$. Theorem 17.2.5 yields, for all $x \in \{V < \infty\}$,

$$\lim_{n\to\infty} n^{(1-p)\alpha/(1-\alpha)} \ \|P^n(x,\cdot) - \pi\|_{V^{\alpha p}} = 0. \tag{17.2.15}$$

This convergence remains valid for $p = 0, 1$ by Theorem 17.2.1. Set $\kappa = 1 + (1 - p)\alpha/(1-\alpha)$, so that $1 \le \kappa \le 1/(1-\alpha)$. With this notation, (17.2.15) reads

$$\lim_{n\to\infty} n^{\kappa-1} \ \|P^n(x,\cdot) - \pi\|_{V^{1-\kappa(1-\alpha)}} = 0. \tag{17.2.16}$$

It is possible to extend this result using more general interpolation functions. We can, for example, obtain nonpolynomial rates of convergence with control functions that are not simply powers of the drift functions. To illustrate this point, set for $b > 0$, $\Upsilon(x) = (1 \vee \log(x))^b$ and $\Psi(x) = x(1 \vee \log(x))^{-b}$. It is not difficult to check that we have

$$\sup_{(x,y)\in[1,\infty)\times[1,\infty)} (x+y)^{-1}\Upsilon(x)\Psi(y) < \infty,$$

so that for all $x \in \{V < \infty\}$, we have

$$\lim_{n\to\infty} \log^b(n)\ \|P^n(x,\cdot)-\pi\|_{V^\alpha(1+\log(V))^{-b}} = 0, \tag{17.2.17}$$

$$\lim_{n\to\infty} n^{\alpha/(1-\alpha)} \log^{-b}(n)\ \|P^n(x,\cdot)-\pi\|_{(1+\log(V))^b} = 0, \tag{17.2.18}$$

and for all $0 < p < 1$,

$$\lim_{n\to\infty} n^{(1-p)\alpha/(1-\alpha)} \log^b n\ \|P^n(x,\cdot)-\pi\|_{V^{\alpha p}(1+\log V)^{-b}} = 0.$$

Logarithmic rates of convergence: Such rates are obtained when the function ϕ increases to infinity more slowly than polynomially. We consider here only the case $\phi(v) = c(1+\log(v))^\alpha$ for some $\alpha \geq 0$ and $c \in (0,1]$. A straightforward calculation shows that $r_\phi(n) \asymp_{n\to\infty} \log^\alpha(n)$.

Applying Theorem 17.2.5, intermediate rates can be obtained along the same lines as above. Choosing, for instance, $\Upsilon(x) = ((1-p)x)^{1-p}$ and $\Psi(x) = (px)^p$ for $0 \leq p \leq 1$, then for all $x \in \{V < \infty\}$,

$$\lim_{n\to\infty} (1+\log(n))^{(1-p)\alpha}\ \|P^n(x,\cdot)-\pi\|_{(1+\log(V))^{p\alpha}} = 0\,.$$

Subexponential rates of convergence: It is also of interest to consider rate functions that increase faster than polynomially, e.g., rate functions of the form

$$r(n)\{1+\log(n)\}^\alpha(n+1)^\beta \mathrm{e}^{cn^\gamma},\ \alpha,\beta\in\mathbb{R}, \gamma\in(0,1)\,\text{and}\,c>0. \tag{17.2.19}$$

Such rates are obtained when the function ϕ is such that $v/\phi(v)$ goes to infinity more slowly than polynomially. More precisely, assume that ϕ is concave and differentiable on $[1,+\infty)$ and that for large v, $\phi(v) = cv/\log^\alpha(v)$ for some $\alpha > 0$ and $c > 0$. A simple calculation yields

$$r_\phi(n) \asymp_{n\to\infty} n^{-\alpha/(1+\alpha)} \exp\left(\{c(1+\alpha)n\}^{1/(1+\alpha)}\right).$$

Applying Theorem 17.2.5 with $\Upsilon(x) = x^{1-p}(1\vee\log(x))^{-b}$ and $\Psi(x) = x^p(1\vee\log(x))^b$ for $p\in(0,1)$ and $b\in\mathbb{R}$; $p=0$ and $b>0$; or $p=1$ and $b<-\alpha$ yields, for all $x\in\{V<\infty\}$,

$$\lim_{n\to\infty} n^{-(\alpha+b)/(1+\alpha)} \exp\left((1-p)\{c(1+\alpha)n\}^{1/(1+\alpha)}\right) \|P^n(x,\cdot)-\pi\|_{V^p(1+\log V)^b} = 0. \tag{17.2.20}$$

17.3 Bibliographical Notes

Polynomial and subgeometric ergodicity of Markov chains was systematically studied in Tuominen and Tweedie (1994), from which we have borrowed the formulation of Theorem 17.1.3.

Several practical drift conditions to derive polynomial rates of convergence were proposed in the works of Veretennikov (1997, 1999), Fort and Moulines (2000), Tanikawa (2001), Jarner and Roberts (2002), and Fort and Moulines (2003a). Rates of convergence using the drift condition $\mathrm{D}_{\mathrm{sg}}(V,\phi,b,C)$ are discussed in Douc et al. (2004a). Further connections between these two drift conditions can be found in Andrieu and Vihola (2015) and Andrieu et al. (2015).

Subexponential rates of convergence were studied by means of coupling techniques under different conditions by Klokov and Veretennikov (2004b) (see also Malyshkin (2000), Klokov and Veretennikov (2004a), and Veretennikov and Klokov (2004)).

Subgeometric drift conditions have also been obtained through state-dependent drift conditions, which are not introduced in this book. These drift conditions are investigated, for example, in Connor and Kendall (2007) and Connor and Fort (2009).

17.A Young Functions

We briefly recall in this appendix Young's inequality and Young functions.

Lemma 17.A.1 *Let $\alpha : [0,M] \to \mathbb{R}$ be a strictly increasing continuous function such that $\alpha(0) = 0$. Denote by β its inverse. For all $0 \le x \le M$, $0 \le y \le \alpha(M)$,*

$$xy \le A(x) + B(y) , \quad A(x) = \int_0^x \alpha(u)\mathrm{d}u \quad \textit{and} \quad B(y) = \int_0^y \beta(u)\mathrm{d}u , \tag{17.A.1}$$

with equality if $y = \alpha(x)$.

Proof It is easily shown that for $z \in [0,M]$,

$$\int_0^z \alpha(u)\mathrm{d}u + \int_0^{\alpha(z)} \beta(u)\mathrm{d}u = z\alpha(z) . \tag{17.A.2}$$

Indeed, the graph of α divides the rectangle with diagonal $(0,0) - (x,\alpha(x))$ into lower and upper parts, and the integrals correspond to the respective areas. Since α is strictly increasing, A is strictly convex. Hence for every $0 < z \ne x \le M$, we have

$$\int_0^x \alpha(u)\mathrm{d}u \ge \int_0^z \alpha(u)\mathrm{d}u + \alpha(z)(x-z) .$$

In particular, if $z = \beta(y)$, we obtain

$$\int_0^x \alpha(u)\mathrm{d}u \geq \int_0^{\beta(y)} \alpha(u)\mathrm{d}u + xy - y\beta(y) .$$

The proof is concluded by applying (17.A.2), which shows that

$$\int_0^{\beta(y)} \alpha(u)\mathrm{d}u = y\beta(y) - \int_0^y \beta(u)\mathrm{d}u .$$

□

The pair (A, B) defined in (17.A.1) is called a pair of Young functions.

Lemma 17.A.2 *Let* $\alpha : [0, \infty) \to [0, \infty)$ *be a strictly increasing continuous function such that* $\alpha(0) = 0$ *and* $\lim_{t\to\infty} \alpha(t) = \infty$*. Denote by* β *the inverse of* α*, by* A *and* B *the primitives of* α *and* β *that vanish at zero, and* $\Upsilon = A^{-1}$ *and* $\Psi = B^{-1}$*. Then* (Υ, Ψ) *is a pair of inverse Young functions, i.e., for all* $x, y \in \mathbb{R}_+$*,* $\Upsilon(x)\Psi(y) \leq x + y$*.*

Proof. For a fixed $v > 0$, define the function $h_v(u) = uv - A(u)$, where $u \geq 0$. Then $h_v'(u) = v - \alpha(u)$ vanishes for $u = \beta(v)$, and h_v' is decreasing, since α is increasing. Thus h_v is concave and attains its maximum value at $\beta(v)$. Therefore, for all $u, v \geq 0$,

$$uv \leq A(u) + h_v(\beta(v)) .$$

Since $h_v \circ \beta(0) = h_v(0) = -A(0) = 0 = B(0)$ and since $h_v \circ \beta(v) = v\beta(v) - A \circ \beta(v)$,

$$\begin{aligned}(h_v \circ \beta)'(v) &= \beta(v) + v\beta'(v) - A'(\beta(v))\beta'(v) = \beta(v) + v\beta'(v) - \alpha \circ \beta(v)\beta'(v) \\ &= \beta(v) + v\beta'(v) - v\beta'(v) = \beta(v) = B'(v) .\end{aligned}$$

We conclude that $h_v \circ \beta = B$. □

Chapter 18
Uniform and V-Geometric Ergodicity by Operator Methods

In this chapter, we will obtain new characterizations and proofs of the uniform ergodicity properties established in Chapter 15. We will consider a Markov kernel P as a linear operator on a set of probability measures endowed with a certain metric. An invariant probability measure is a fixed point of this operator, and therefore, a natural idea is to use a fixed-point theorem to prove convergence of the iterates of the kernel to the invariant distribution. To do so, in Section 18.1, we will first state and prove a version of the fixed-point theorem that suits our purposes. As appears in the fixed-point theorem, the main restriction of this method is that it can provide only geometric rates of convergence. These techniques will be again applied in Chapter 20, where we will be dealing with other metrics on the space of probability measures.

In order to apply this fixed-point theorem, we must prove that P is a contraction with respect to the chosen distance, or in other words, a Lipschitz map with Lipschitz coefficient strictly less than one. The Lipschitz coefficient of the Markov kernel P with respect to the total variation distance is called its Dobrushin coefficient. The fixed-point theorem and the Dobrushin coefficient will be used in Section 18.2 to obtain uniform ergodicity. In Section 18.3 we will consider the V-norm introduced in Section 13.4 (see also Appendix D.3), which induces the V-Dobrushin coefficient, which will be used in Section 18.4 to obtain geometric rates of convergence in the V-norm.

As a by-product of Theorem 18.2.4, we will give in Section 18.5 a new proof of Theorem 11.2.5 (which states the existence and uniqueness of the invariant measure of a recurrent irreducible Markov kernel) that does not use the splitting construction.

18.1 The Fixed-Point Theorem

The set of probability measures $\mathbb{M}_1(\mathscr{X})$ endowed with the total variation distance is a complete metric space (see Appendix D). A Markov kernel is an operator on this space, and an invariant probability measure is a fixed point of this operator. It is thus

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_18

natural to use the classical fixed-point theorem to find conditions for the existence of an invariant measure and to identify the convergence rate of the sequence of iterates of the kernel to the invariant probability measure. Therefore, we restate here the fixed-point theorem for a Markov kernel P, in a general framework in which we consider P as an operator on a subset $\mathbb{F}$ of $\mathbb{M}_1(\mathscr{X})$, endowed with a metric ρ that can possibly be different from the total variation distance.

Theorem 18.1.1. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, $\mathbb{F}$ a subspace of $\mathbb{M}_1(\mathscr{X})$, and ρ a metric on $\mathbb{F}$ such that $(\mathbb{F}, \rho)$ is complete. Suppose in addition that $\delta_x \in \mathbb{F}$ for all $x \in \mathsf{X}$ and that $\mathbb{F}$ is stable by P. Assume that there exist an integer $m > 0$ and constants $A_r > 0$, $r \in \{1, \ldots, m-1\}$, and $\alpha \in [0,1)$ such that for all $\xi, \xi' \in \mathbb{F}$,*

$$\rho(\xi P^r, \xi' P^r) \leq A_r \rho(\xi, \xi') , \qquad r \in \{1, \ldots, m-1\} , \tag{18.1.1}$$

$$\rho(\xi P^m, \xi' P^m) \leq \alpha \rho(\xi, \xi') . \tag{18.1.2}$$

Then there exists a unique invariant probability measure $\pi \in \mathbb{F}$, and for all $\xi \in \mathbb{F}$ and $n \in \mathbb{N}$,

$$\rho(\xi P^n, \pi) \leq \left(1 \vee \max_{1 \leq r < m} A_r\right) \rho(\xi, \pi) \alpha^{\lfloor n/m \rfloor} . \tag{18.1.3}$$

Assume that one of the following conditions is satisfied:

(i) The convergence of a sequence of probability measures in $(\mathbb{F}, \rho)$ implies setwise convergence.
(ii) The set X is a metric space endowed with its Borel σ-field $\mathscr{X}$, and the convergence of a sequence of probability measures in $(\mathbb{F}, \rho)$ implies its weak convergence.

Then π is the unique P-invariant probability measure in $\mathbb{M}_1(\mathscr{X})$.

Proof. Let us first prove the uniqueness. If π and π' are such that $\pi P = \pi$ and $\pi' P = \pi$, then $\pi P^m = \pi$ and $\pi' P^m = \pi$; thus

$$\rho(\pi, \pi') = \rho(\pi P^m, \pi' P^m) \leq \alpha \rho(\pi, \pi') < \rho(\pi, \pi') ,$$

the last inequality being a consequence of $\alpha \in (0,1)$. This proves that $\pi = \pi'$. To prove the existence, consider $\xi, \xi' \in \mathbb{F}$ and an integer n. Write $n = km + r$ with $r \in \{0, \ldots, m-1\}$ and $k \in \mathbb{N}$. Then

$$\rho(\xi P^n, \xi' P^n) = \rho(\xi P^{km+r}, \xi' P^{km+r}) \leq \alpha^k \rho(\xi P^r, \xi' P^r) .$$

Taking $\xi' = \xi P$, we obtain

$$\begin{aligned}\rho(\xi P^n, \xi P^{n+1}) &\le \alpha^k \rho(\xi P^r, \xi P^{r+1}) = \alpha^{\lfloor n/m \rfloor} \rho(\xi P^r, \xi P^{r+1}) \\ &\le \alpha^{\lfloor n/m \rfloor} \max_{0 \le r < m} \rho(\xi P^r, \xi P^{r+1}) .\end{aligned}$$

This implies that $\{\xi P^n\}$ is a Cauchy sequence, and since $(\mathbb{F}, \rho)$ is complete, it converges to a limit $\pi \in \mathbf{F}$. Assumption (18.1.2) (if $m = 1$) or (18.1.1) (if $m > 1$) implies that P is continuous, and thus $\pi = \pi P$ is a fixed point. Therefore,

$$\begin{aligned}\rho(\xi P^n, \pi) = \rho(\xi P^n, \pi P^n) &\le \alpha^{\lfloor n/m \rfloor} \max_{0 \le r < m} \rho(\xi P^r, \pi P^r) \\ &\le \alpha^{\lfloor n/m \rfloor} \left(1 \vee \max_{1 \le r \le m-1} A_r\right) \rho(\xi, \pi) .\end{aligned}$$

This proves (18.1.3). We now prove the last part of the theorem. Let $\pi \in \mathbb{F}$ be the unique invariant probability in $\mathbb{F}$ and let $\tilde{\pi}$ be an invariant probability in $\mathbb{M}_1(\mathscr{X})$. Then for all $f \in \mathbb{F}_b(\mathsf{X})$ (or $f \in \mathrm{C}_b(\mathsf{X})$), we have

$$\tilde{\pi}(f) = \tilde{\pi} P^n(f) = \int P^n f(x) \tilde{\pi}(\mathrm{d}x) .$$

By the first part of the theorem, the sequence $\{\delta_x P^n,\ n \in \mathbb{N}\}$ converges with respect to the distance ρ, hence either setwise or weakly to the probability π. Thus $\lim_{n\to\infty} P^n f(x) = \pi(f)$ for all $x \in \mathsf{X}$ and all $f \in \mathbb{F}_b(\mathsf{X})$ (or $f \in \mathrm{C}_b(\mathsf{X})$). Since in addition, $|P^n f(x)| \le |f|_\infty$ Lebesgue's dominated convergence theorem implies that $\lim_{n\to\infty} \int P^n f(x) \tilde{\pi}(\mathrm{d}x) = \pi(f)$, this yields $\tilde{\pi}(f) = \pi(f)$. Therefore, $\tilde{\pi} = \pi$, which concludes the proof. $\square$

The second part of the theorem means that if convergence with respect to ρ implies either setwise or weak convergence (i.e., the topology induced by ρ is finer than the topology of weak convergence), then the invariant probability is unique not only in $\mathbb{F}$, but also in $\mathbb{M}_1(\mathscr{X})$. If $\mathbb{F} = \mathbb{M}_1(\mathscr{X})$, then this condition is superfluous to obtain the uniqueness of the invariant probability in $\mathbb{M}_1(\mathscr{X})$.

18.2 Dobrushin Coefficient and Uniform Ergodicity

We have already introduced in Theorem 15.3.1 a set of conditions that are equivalent to uniform geometric ergodicity, the most striking of which is without doubt that the whole state space must be small. In this section we will introduce another necessary and sufficient condition, which is directly related to the strong contraction of the iterates in the total variation distance. For this purpose, we introduce the Dobrushin coefficient, which is the modulus of continuity of a Markov kernel P on $\mathsf{X} \times \mathscr{X}$, considered as an operator on $\mathbb{M}_1(\mathscr{X})$ endowed with the total variation distance.

Definition 18.2.1 (Dobrushin coefficient) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. The Dobrushin coefficient $\Delta(P)$ is the Lipschitz coefficient of P with respect to the total variation distance, i.e.,*

$$\Delta(P) = \sup_{\xi \neq \xi' \in \mathbb{M}_1(\mathscr{X})} \frac{\mathrm{d}_{\mathrm{TV}}(\xi P, \xi' P)}{\mathrm{d}_{\mathrm{TV}}(\xi, \xi')} = \sup_{\xi \neq \xi' \in \mathbb{M}_1(\mathscr{X})} \frac{\|\xi P - \xi' P\|_{\mathrm{TV}}}{\|\xi - \xi'\|_{\mathrm{TV}}} . \tag{18.2.1}$$

The kernel P can also be considered a linear operator on the linear space $\mathbb{M}_0(\mathscr{X})$ of bounded signed measures μ satisfying $\mu(\mathsf{X}) = 0$. Endowed with the total variation norm, $\mathbb{M}_0(\mathscr{X})$ is a Banach space. In this setting, the Dobrushin coefficient $\Delta(P)$ is the operator norm of P. This yields straightforwardly that if P and Q are two Markov kernels on $\mathsf{X} \times \mathscr{X}$, then

$$\Delta(PQ) \le \Delta(P)\Delta(Q) . \tag{18.2.2}$$

We now prove that the Dobrushin coefficient is always less than 1, and while so doing, we provide a more convenient expression for it.

Lemma 18.2.2 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Then*

$$\Delta(P) = \sup_{(x,x') \in \mathsf{X} \times \mathsf{X}} \mathrm{d}_{\mathrm{TV}}(P(x,\cdot), P(x',\cdot)) \le 1 . \tag{18.2.3}$$

Proof. By definition, the right-hand side of (18.2.3) is less than or equal to $\Delta(P)$. We now prove the converse inequality. Applying the definition of the total variation distance and homogeneity, we have

$$\Delta(P) = \sup\{\|\xi P\|_{\mathrm{TV}} : \xi \in \mathbb{M}_0(\mathscr{X}), \|\xi\|_{\mathrm{TV}} \le 1\} . \tag{18.2.4}$$

Using Proposition D.2.4 and the bound (D.2.4), we have, for $\xi \in \mathbb{M}_0(\mathscr{X})$, since $\xi P \in \mathbb{M}_0(\mathscr{X})$,

$$\|\xi P\|_{\mathrm{TV}} = 2 \sup_{\mathrm{osc}(f) \le 1} |(\xi P)(f)| = 2 \sup_{\mathrm{osc}(f) \le 1} |\xi(Pf)| \le \|\xi\|_{\mathrm{TV}} \sup_{\mathrm{osc}(f) \le 1} \mathrm{osc}(Pf) .$$

Note now that

$$\begin{aligned}
\sup_{\mathrm{osc}(f) \le 1} \mathrm{osc}(Pf) &= \sup_{\mathrm{osc}(f) \le 1} \sup_{x,x'} |Pf(x) - Pf(x')| \\
&= \sup_{x,x'} \sup_{\mathrm{osc}(f) \le 1} \left|\{P(x,\cdot) - P(x',\cdot)\} f\right| \\
&= \frac{1}{2} \sup_{x,x'} \left\|P(x,\cdot) - P(x',\cdot)\right\|_{\mathrm{TV}} = \sup_{x,x'} \mathrm{d}_{\mathrm{TV}}(P(x,\cdot), P(x',\cdot)) .
\end{aligned}$$

Thus for $\xi \in \mathbb{M}_0(\mathscr{X})$ such that $\|\xi\|_{\mathrm{TV}} \le 1$, we obtain

$$\|\xi P\|_{\mathrm{TV}} \le \sup_{x,x'} \mathrm{d}_{\mathrm{TV}}(P(x,\cdot),P(x',\cdot)) \;.$$

Recalling (18.2.4), this proves the converse inequality. □

Lemma 18.2.2 and Corollary D.2.5 yield the following bound, for all $f \in \mathbb{F}_b(\mathsf{X})$ and $x,y \in \mathsf{X}$:

$$|Pf(x) - Pf(y)| \le \Delta\,(P)\,\mathrm{osc}\,(f) \;. \tag{18.2.5}$$

Lemma 18.2.3 *For all $\xi,\xi' \in \mathbb{M}_1(\mathscr{X})$, the sequence $\{\mathrm{d}_{\mathrm{TV}}(\xi P^n, \xi' P^n),\ n \in \mathbb{N}\}$ is decreasing and*

$$\mathrm{d}_{\mathrm{TV}}(\xi P^n, \xi' P^n) \le \{\Delta(P)\}^n \mathrm{d}_{\mathrm{TV}}(\xi,\xi') \;. \tag{18.2.6}$$

If π is an invariant probability measure, then for every $\xi \in \mathbb{M}_1(\mathscr{X})$, the sequence $\{\mathrm{d}_{\mathrm{TV}}(\xi P^n,\pi),\ n \in \mathbb{N}\}$ is decreasing and $\mathrm{d}_{\mathrm{TV}}(\xi P^n,\pi) \le \{\Delta(P)\}^n \mathrm{d}_{\mathrm{TV}}(\xi,\pi)$.

Proof. By definition of the Dobrushin coefficient, we have

$$\mathrm{d}_{\mathrm{TV}}(\xi P^{n+1}, \xi' P^{n+1}) \le \Delta(P)\mathrm{d}_{\mathrm{TV}}(\xi P^n, \xi' P^n) \;.$$

This proves that the sequence is decreasing, since $\Delta(P) \le 1$ (see Lemma 18.2.2) and (18.2.6) follows by induction. If π is an invariant probability measure, then $\pi P^n = \pi$, and the second part of the lemma is obtained by replacing ξ' and $\xi' P^n$ by π. □

Of crucial importance are the situations in which the kernel P or one of its iterates is a strict contraction, i.e., there exists an integer $m \ge 1$ such that $\Delta(P^m) < 1$. In this case, Lemma 18.2.3 implies that the initial distributions ξ and $\xi' \in \mathbb{M}_1(\mathscr{X})$ will be forgotten exponentially fast.

Theorem 18.2.4. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ satisfying $\Delta\,(P^m) \le 1-\varepsilon$. Then P admits a unique invariant probability measure π. In addition, for all $\xi \in \mathbb{M}_1(\mathscr{X})$,*

$$\|\xi P^n - \pi\|_{\mathrm{TV}} \le \|\xi - \pi\|_{\mathrm{TV}}\ (1-\varepsilon)^{\lfloor n/m \rfloor} \;. \tag{18.2.7}$$

Proof. By Theorem D.2.7, $(\mathbb{M}_1(\mathscr{X}), \mathrm{d}_{\mathrm{TV}})$ is a complete metric space. Thus we can apply Theorem 18.1.1, which proves that there exists a unique invariant probability measure π, and

$$\begin{aligned}
\|\xi P^n - \pi\|_{\mathrm{TV}} &\le (1 \vee \max_{1 \le r < m} \Delta\,(P^r))\, \|\xi - \pi\|_{\mathrm{TV}}\ (1-\varepsilon)^{\lfloor n/m \rfloor} \\
&\le \|\xi - \pi\|_{\mathrm{TV}}\ (1-\varepsilon)^{\lfloor n/m \rfloor} \;,
\end{aligned}$$

where the last inequality follows from $\Delta(P^r) \le 1$ for all r. □

Since the total variation distance of two probability measures is always less than 1, we have

$$d_{\mathrm{TV}}(\xi P^n, \pi) \le (1-\varepsilon)^{\lfloor n/m \rfloor} . \tag{18.2.8}$$

This means that the convergence is uniform with respect to the initial distribution and holds at a geometric rate.

We already know from Theorem 15.3.1 (iii) that the Markov kernel P is uniformly (geometrically) ergodic if and only if the state space X is m-small. We can now add another equivalent condition.

Theorem 18.2.5. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. The following statements are equivalent.*

(i) P is uniformly geometrically ergodic.
(ii) $\Delta(P^m) < 1$ for some $m \in \mathbb{N}$.

Proof. We already know that (ii) ⇒ (i). Assume that P is uniformly geometrically ergodic. By definition, P admits an invariant probability π, and there exist constants $\varsigma < \infty$ and $\rho < 1$ such that $\sup_{x \in \mathsf{X}} \|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} \le \varsigma \rho^n$. By the triangle inequality, this implies

$$\frac{1}{2} \sup_{(x,x') \in \mathsf{X} \times \mathsf{X}} \left\|P^n(x,\cdot) - P^n(x',\cdot)\right\|_{\mathrm{TV}} \le \sup_{x \in \mathsf{X}} \|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} \le \varsigma \rho^n .$$

By Lemma 18.2.2, this means that $\Delta(P^n) < 1$ for some n. Thus (i) ⇒ (ii). □

We will now state sufficient conditions under which the Dobrushin coefficient of the Markov kernel P or one of its iterates P^m is strictly less than 1.

Definition 18.2.6 (Doeblin set and uniform Doeblin condition) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, $m \ge 1$ an integer, and $\varepsilon > 0$. A set $C \in \mathscr{X}$ is an (m,ε)-Doeblin set if for every $(x,x') \in C \times C$,*

$$d_{\mathrm{TV}}(P^m(x,\cdot), P^m(x',\cdot)) \le 1 - \varepsilon . \tag{18.2.9}$$

If the state space X is a Doeblin set, we say that P^m satisfies the uniform Doeblin condition.

If P is uniformly ergodic, then it satisfies the uniform Doeblin condition. Doeblin sets and small sets are closely related, as shown below.

Lemma 18.2.7 *(i) If C is an $(m, \varepsilon\nu)$-small set, then C is an (m, ε)-Doeblin set.*
(ii) If P is irreducible and aperiodic, then any Doeblin set is small.

Proof. (i) Set $Q(x,\cdot) = (1-\varepsilon)^{-1}(P^m(x,\cdot) - \varepsilon\nu)$ for $x \in C$. Note that $Q(x,\cdot)$ is a probability measure for every $x \in C$ and for all $x \in C$,

$$P^m(x,\cdot) = (1-\varepsilon)Q(x,\cdot) + \varepsilon\nu .$$

Therefore, for $x, x' \in C$, since the total variation distance is bounded by 1, we have

$$\mathrm{d_{TV}}(P^m(x,\cdot), P^m(x',\cdot)) = (1-\varepsilon)\mathrm{d_{TV}}(Q(x,\cdot), Q(x',\cdot)) \le 1-\varepsilon .$$

(ii) Let C be an (m,ε)-Doeblin set i.e., $\mathrm{d_{TV}}(P^m(x,\cdot), P^m(x',\cdot)) \le 1-\varepsilon$ for all $x, x' \in C$. Choose an arbitrary point $x_0 \in C$. By Proposition 9.4.11, X is an increasing union of small sets, and thus there exists an (n,μ)-small set $S \subset \mathsf{X}$ such that $P^m(x_0, S) \ge 1 - \varepsilon/2$. By Corollary D.2.5, we then have, for all $x \in C$,

$$P^m(x,S) \ge P^m(x_0,S) - \mathrm{d_{TV}}(P^m(x,\cdot), P^m(x_0,\cdot)) \ge 1-\varepsilon/2 - 1 + \varepsilon = \varepsilon/2 .$$

Therefore, for every $x \in C$ and $A \in \mathscr{X}$,

$$P^{n+m}(x,A) \ge \int_S P^m(x,\mathrm{d}y)P^n(y,A) \ge \frac{\varepsilon}{2}\mu(A) .$$

This proves that C is a small set. □

Example 18.2.8. If X is finite or countable, we get

$$\mathrm{d_{TV}}(P^m(x,\cdot), P^m(x',\cdot)) = 1 - \sum_{z\in\mathsf{X}} P^m(x,z) \wedge P^m(x',z) .$$

The set C is an (m,ε)-Doeblin set if $\min_{x,x'\in C}\sum_{z\in\mathsf{X}} P^m(x,z) \wedge P^m(x',z) \ge \varepsilon$. Set $\eta_m = \sum_{y\in\mathsf{X}} \inf_{x\in C} P^m(x,y)$. If $\eta_m > 0$, then C is an $(m, \eta_m\nu_m)$-small set with $\nu_m(z) = \eta_m^{-1}\inf_{x\in C} P^m(x,z)$. It is always the case that $\eta_m \le \varepsilon_m$. ◀

Example 18.2.9. We will show by means of a simple example that the results obtained in Theorem 18.2.4 cannot be improved in general. We consider the independent Metropolis–Hastings sampler on a discrete state space $\mathsf{X} = \{1,\ldots,m\}$ for some finite m. We denote by π the target distribution and by q the proposal distribution. To simplify the notation, we set $\pi(x) = \pi_x$ and $q(x) = q_x$. We assume that $\pi_x > 0$ and $q_x > 0$ for all $x \in \mathsf{X}$, and we denote by $w_x = \pi_x/q_x$ the importance weight associated with state $x \in \mathsf{X}$.

Without loss of generality, we assume that the states are sorted according to the magnitudes of their importance ratio, i.e., the elements are labeled so that $\{w_1 \ge w_2 \ge \cdots \ge w_m\}$. The acceptance probability of a move from x to y is given by

$$\alpha(x,y) = 1 \wedge \frac{\pi_y q_x}{\pi_x q_y} = 1 \wedge \frac{w_y}{w_x} ,$$

and the ordering of the states therefore implies that $\alpha(x,y) = 1$ for $y \le x$ and $\alpha(x,y) = w_y/w_x$ for $y > x$. Define $\eta_0 = 1$, $\eta_m = 0$, and for $x \in \{1,\dots,m-1\}$,

$$\eta_x = \sum_{y>x}(q_y - \pi_y/w_x) = \sum_{y>x} q_y \frac{w_x - w_y}{w_x} ,$$

which is the probability of being rejected in the next step if the chain is at state x. The transition matrix P of the independent Metropolis–Hastings sampler can be written in the form

$$P(x,y) = \begin{cases} q_y, & y < x, \\ w_y q_y / w_x, & x < y, \\ q_x + \eta_x, & x = y. \end{cases} \tag{18.2.10}$$

In words, if the chain is state x, then all the moves to states $y < x$ are accepted. The moves to states $y > x$ are accepted with probability w_y/w_x. The probability of staying at x is the sum of the probability of proposing x and the probability of rejecting a move outside x.

Denote by $\mathrm{L}^2(\pi)$ the set of functions $g : \mathsf{X} \to \mathbb{R}$ satisfying $\sum_{x\in\mathsf{X}} \pi(x) g^2(x) < \infty$. We equip this space with the scalar product $\langle g,h\rangle_{\mathrm{L}^2(\pi)} = \sum_{x\in\mathsf{X}} g(x)h(x)\pi(x)$. The probability measure π is reversible with respect to the Markov transition P, which implies that P is self-adjoint, $\langle Pg,h\rangle_{\mathrm{L}^2(\pi)} = \langle g,Ph\rangle_{\mathrm{L}^2(\pi)}$. The spectral theorem says that there is an orthonormal basis of eigenvectors. It can be shown by direct calculation that $\{\eta_0,\ \eta_1,\dots,\eta_{m-1}\}$ are the eigenvalues of P in decreasing order and the corresponding eigenvectors are

$$\begin{aligned} &\psi_0 = (1,\ 1,\ \dots,\ 1)^T , \\ &\psi_i = (0,0,\ \dots,\ 0,\ S_{i+1},\ -\pi_i,\ \dots,\ -\pi_i)^T ,\ 1 \le i \le m-1, \end{aligned}$$

where $S_{i+1} = \sum_{k=i+1}^m \pi_i$ is the ith component of the vector ψ_i. By elementary manipulations, for all $x,y \in \mathsf{X}\times\mathsf{X}$,

$$P(x,y) = \pi_y \sum_{i=0}^{m-1} \eta_i \psi_i(x)\psi_i(y) , \quad P^n(x,y) = \pi_y \sum_{i=0}^{m-1} \eta_i^n \psi_i(x)\psi_i(y) . \tag{18.2.11}$$

When applied to the Markov kernel P^n, these formulas yield, for all $x,y \in \mathsf{X}\times\mathsf{X}$,

$$P^n(x,y) = \begin{cases} \pi_y \left(1 + \sum_{k=1}^{x-1} (\eta_k^n \pi_k)/(S_k S_{k+1}) - \eta_x^n/S_x\right) , & x < y, \\ \pi_y \left(1 + \sum_{k=1}^{x-1} (\eta_k^n \pi_k)/(S_k S_{k+1}) - \eta_x^n/S_x\right) + \eta_y^n , & x = y, \\ \pi_y \left(1 + \sum_{k=1}^{y-1} (\eta_k^n \pi_k)/(S_k S_{k+1}) - \eta_x^n/S_x\right) , & x > y. \end{cases}$$

Using the Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
\|P^n(x,\cdot)-\pi\|_{\mathrm{TV}} &= \sum_{y\in\mathsf{X}} |P^n(x,y)-\pi(y)| \\
&\le \sum_{y\in\mathsf{X}} \frac{\{P^n(x,y)-\pi(y)\}^2}{\pi(y)} = \sum_{y\in\mathsf{X}} \frac{\{P^n(x,y)\}^2}{\pi(y)} - 1 \,.
\end{aligned}
$$

Since P is self-adjoint in $\mathrm{L}^2(\pi)$, P^n is also self-adjoint in $\mathrm{L}^2(\pi)$, which implies, for all $x,y\in\mathsf{X}$ and $n\in\mathbb{N}$, $\pi(x)P^n(x,y)=\pi(y)P^n(y,x)$. Hence

$$
\begin{aligned}
\sum_{y\in\mathsf{X}} \frac{\{P^n(x,y)\}^2}{\pi(y)} &= \sum_{y\in\mathsf{X}} \frac{P^n(x,y)\pi(x)P^n(x,y)}{\pi(x)\pi(y)} \\
&= \sum_{y\in\mathsf{X}} \frac{P^n(x,y)P^n(y,x)}{\pi(x)} = \frac{P^{2n}(x,x)}{\pi(x)} \,.
\end{aligned}
$$

Using (18.2.11), we therefore obtain

$$
\|P^n(x,\cdot)-\pi\|_{\mathrm{TV}} \le \sum_{i=1}^{m} \eta_i^n \psi_i^2(x) \,. \tag{18.2.12}
$$

Using this eigenexpansion, we see that the exact rate of convergence of this algorithm is given by the second eigenvalue, namely

$$
\begin{aligned}
\eta_1 &= \sum_{k>1}(q_k - \pi_k/w_1) = (1-q_1)-(1-\pi_1)/w_1 \\
&= 1-q_1/\pi_1 = 1-\min_{x\in\mathsf{X}}(q_x/\pi_x) \,,
\end{aligned}
$$

if we recall the ordering on this chain. Applying the bound of Equation (15.3.2) obtained in Example 15.3.3, we know that $P(x,A)\ge\varepsilon\pi(A)$ for all $A\in\mathscr{X}$ and $x\in\mathsf{X}$ with $\varepsilon=\min_{x\in\mathsf{X}} q_x/\pi_x$. Thus η_1^n is exactly the rate of convergence ensured by Theorem 18.2.4.

18.3 V-Dobrushin Coefficient

To prove nonuniform convergence, we must replace the total variation distance on $\mathbb{M}_1(\mathscr{X})$ by the V-distance and the Dobrushin coefficient by the V-Dobrushin coefficient. Before going further, some additional notation and definitions are required. Let $V\in\mathbb{F}(\mathsf{X})$ with values in $[1,\infty)$. The V-norm $|f|_V$ of a function $f\in\mathbb{F}(\mathsf{X})$ is defined by

$$
|f|_V = \sup_{x\in\mathsf{X}} \frac{|f(x)|}{V(x)} \,.
$$

The V-norm of a bounded signed measure $\xi\in\mathbb{M}_s(\mathscr{X})$ is defined by

$$\|\xi\|_V = |\xi|(V) .$$

The space of finite signed measures ξ such that $\|\xi\|_V < \infty$ is denoted by $\mathbb{M}_V(\mathscr{X})$. The V-oscillation seminorm of the function $f \in \mathbb{F}_b(\mathsf{X})$ is defined by

$$\operatorname{osc}_V(f) = \sup_{(x,x')\in\mathsf{X}\times\mathsf{X}} \frac{|f(x)-f(x')|}{V(x)+V(x')} . \tag{18.3.1}$$

Finally, we define the spaces of measures

$$\mathbb{M}_{0,V}(\mathscr{X}) = \{\xi \in \mathbb{M}_0(\mathscr{X}) : \xi(V) < \infty\} , \tag{18.3.2}$$

$$\mathbb{M}_{1,V}(\mathscr{X}) = \{\xi \in \mathbb{M}_1(\mathscr{X}) : \xi(V) < \infty\} . \tag{18.3.3}$$

Theorem D.3.2 shows that for $\xi \in \mathbb{M}_{0,V}(\mathscr{X})$,

$$\|\xi\|_V = \sup\{\xi(f) : \operatorname{osc}_V(f) \le 1\} . \tag{18.3.4}$$

The V-norm induces on $\mathbb{M}_{1,V}(\mathscr{X})$ a distance d_V defined for $\xi,\xi' \in \mathbb{M}_{1,V}(\mathscr{X})$ by

$$\mathrm{d}_V(\xi,\xi') = \frac{1}{2}\|\xi-\xi'\|_V . \tag{18.3.5}$$

The set $\mathbb{M}_{1,V}(\mathscr{X})$ equipped with the distance d_V is a complete metric space; see Corollary D.3.4. If X is a metric space endowed with its Borel σ-field, then convergence with respect to the distance d_V implies weak convergence. Further properties of the V-norm and the associated V-distance are given in Appendix D.3. We will need only the following property.

Lemma 18.3.1 *Let $\xi,\xi' \in \mathbb{M}_1(\mathscr{X})$ and $\varepsilon \in (0,1)$. If $\mathrm{d}_{\mathrm{TV}}(\xi,\xi') \le 1-\varepsilon$, then*

$$\|\xi-\xi'\|_V \le \xi(V)+\xi'(V)-2\varepsilon . \tag{18.3.6}$$

Proof. Set $\nu = \xi+\xi'-|\xi-\xi'|$. Then

$$\|\xi-\xi'\|_{\mathrm{TV}} = |\xi-\xi'|(\mathsf{X}) = 2-\nu(\mathsf{X}) .$$

Thus $\mathrm{d}_{\mathrm{TV}}(\xi,\xi') \le 1-\varepsilon$ if and only if $\nu(\mathsf{X}) \ge 2\varepsilon$. Since $V \ge 1$, we have

$$\begin{aligned}\|\xi-\xi'\|_V &= |\xi-\xi'|(V) = \xi(V)+\xi'(V)-\nu(V)\\ &\le \xi(V)+\xi'(V)-\nu(\mathsf{X}) \le \xi(V)+\xi'(V)-2\varepsilon .\end{aligned}$$

□

Definition 18.3.2 (V-Dobrushin Coefficient) *Let $V : \mathsf{X} \to [1,\infty)$ be a measurable function. Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$ such that for every $\xi \in \mathbb{M}_{1,V}(\mathscr{X})$,*

$\xi P \in \mathbb{M}_{1,V}(\mathscr{X})$. *The V-Dobrushin coefficient of the Markov kernel P, denoted by $\Delta_V(P)$, is defined by*

$$\Delta_V(P) = \sup_{\xi \neq \xi' \in \mathbb{M}_{1,V}(\mathscr{X})} \frac{\mathrm{d}_V(\xi P, \xi' P)}{\mathrm{d}_V(\xi, \xi')} = \sup_{\xi \neq \xi' \in \mathbb{M}_{1,V}(\mathscr{X})} \frac{\|\xi P - \xi' P\|_V}{\|\xi - \xi'\|_V}. \tag{18.3.7}$$

If the function V is not bounded, then in contrast to the Dobrushin coefficient, the V-Dobrushin coefficient is not necessarily finite. When $\Delta_V(P) < \infty$, then P can be seen as a bounded linear operator on the space $\mathbb{M}_{0,V}(\mathscr{X})$, and $\Delta_V(P)$ is its operator norm, i.e.,

$$\Delta_V(P) = \sup_{\substack{\xi \in \mathbb{M}_{0,V}(\mathscr{X}) \\ \xi \neq 0}} \frac{\|\xi P\|_V}{\|\xi\|_V} = \sup_{\substack{\xi \in \mathbb{M}_{0,V}(\mathscr{X}) \\ \|\xi\|_V \leq 1}} \|\xi P\|_V . \tag{18.3.8}$$

This yields in particular the submultiplicativity of the Dobrushin coefficient, i.e., if P, Q are Markov kernels on $\mathsf{X} \times \mathscr{X}$, then

$$\Delta_V(PQ) \leq \Delta_V(P)\,\Delta_V(Q) . \tag{18.3.9}$$

An equivalent expression for the V-Dobrushin coefficient in terms of the V-oscillation seminorm extending (18.2.3) is available.

Lemma 18.3.3 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Then*

$$\Delta_V(P) = \sup_{\mathrm{osc}_V(f) \leq 1} \mathrm{osc}_V(Pf) = \sup_{(x,x') \in \mathsf{X} \times \mathsf{X}} \frac{\|P(x,\cdot) - P(x',\cdot)\|_V}{V(x) + V(x')} . \tag{18.3.10}$$

Proof. Since $\|\delta_x - \delta_{x'}\|_V = V(x) + V(x')$ for $x \neq x'$, the right-hand side of (18.3.10) is obviously less than or equal to $\Delta_V(P)$. By Theorem D.3.2 and (D.3.5), we have, for all ξ, $\xi' \in \mathbb{M}_1(\mathscr{X})$,

$$\begin{aligned}
\left\|\xi P - \xi' P\right\|_V &= \sup_{\mathrm{osc}_V(f) \leq 1} |\xi P(f) - \xi' P(f)| \\
&= \sup_{\mathrm{osc}_V(f) \leq 1} |\xi(Pf) - \xi'(Pf)| \leq \left\|\xi - \xi'\right\|_V \sup_{\mathrm{osc}_V(f) \leq 1} \mathrm{osc}_V(Pf) .
\end{aligned}$$

To conclude, we apply again Theorem D.3.2 to obtain

$$\begin{aligned}
\sup_{\mathrm{osc}_V(f) \leq 1} \mathrm{osc}_V(Pf) &= \sup_{\mathrm{osc}_V(f) \leq 1} \sup_{x,x'} \frac{|Pf(x) - Pf(x')|}{V(x) + V(x')} \\
&= \sup_{x,x'} \sup_{\mathrm{osc}_V(f) \leq 1} \frac{|[P(x,\cdot) - P(x',\cdot)]f|}{V(x) + V(x')} = \sup_{x,x'} \frac{\|P(x,\cdot) - P(x',\cdot)\|_V}{V(x) + V(x')} .
\end{aligned}$$

$\square$

It is important to have a condition that ensures that the V-Dobrushin coefficient is finite.

Lemma 18.3.4 *Assume that P satisfies the $D_g(V,\lambda,b)$ drift condition. Then for all $r \in \mathbb{N}^*$,*

$$\Delta_V(P^r) \le \lambda^r + b\frac{1-\lambda^r}{1-\lambda} . \tag{18.3.11}$$

Proof. For all $x, x' \in \mathsf{X}$, we have

$$\frac{\|P^r(x,\cdot) - P^r(x',\cdot)\|_V}{V(x)+V(x')} \le \frac{P^rV(x)+P^rV(x')}{V(x)+V(x')} \le \lambda^r + \frac{2b(1-\lambda^r)}{(1-\lambda)\{V(x)+V(x')\}} .$$

The bound (18.3.11) follows from Lemma 18.3.3 using $V(x)+V(x') \ge 2$. □

18.4 V-Uniformly Geometrically Ergodic Markov Kernel

We now state and prove the equivalent of Theorem 18.2.5 for a V-uniformly geometrically ergodic Markov kernel.

Theorem 18.4.1. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. The following statements are equivalent:*

(i) P is V-uniformly geometrically ergodic.
(ii) There exist $m \in \mathbb{N}^*$ and $\varepsilon > 0$ such that

$$\Delta_V(P^m) \le 1-\varepsilon , \tag{18.4.1}$$

and $\Delta_V(P^r) < \infty$ for all $r \in \{0,\dots,m-1\}$.

Proof. We first show that (i) ⇒ (ii). By (15.2.1), there exist $\rho \in [0,1)$ and $\varsigma < \infty$ such that for all measurable functions f satisfying $|f|_V \le 1$, $x, x' \in \mathsf{X}$, and $n \in \mathbb{N}$, we have

$$|P^n f(x) - P^n f(x')| \le \varsigma(V(x)+V(x'))\rho^n ,$$

which implies $\Delta_V(P^n) \le \varsigma\rho^n$ for all $n \in \mathbb{N}$. For every $\varepsilon \in (0,1)$, we may therefore choose m large enough that $\Delta_V(P^m) \le 1-\varepsilon$.

Conversely, (ii) ⇒ (i) follows directly from Theorem 18.1.1. □

In this section we will establish that the drift condition $D_g(V,\lambda,b)$ implies the V-uniform geometric ergodicity property, providing a different proof of Theorem 15.2.4. We will first prove that under condition $D_g(V,\lambda,b)$, we can bound the Dobrushin coefficient related to a modification of the function V. For $\beta > 0$, define

$$V_\beta = 1 + \beta(V-1) . \tag{18.4.2}$$

We first give conditions that ensure that $\Delta_{V_\beta}(P) < 1$. Recall that if condition $\mathrm{D_g}(V, \lambda, b)$ holds, then $\lambda + b \geq 1$; cf. Remark 14.1.9.

Lemma 18.4.2 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ satisfying the geometric drift condition $\mathrm{D_g}(V, \lambda, b)$. Then*

$$\Delta_{V_\beta}(P) \leq 1 + \beta(b + \lambda - 1) . \tag{18.4.3}$$

Assume, moreover, that there exists d such that the level set $\{V \leq d\}$ is a $(1, \varepsilon)$-Doeblin set and

$$\lambda + 2b/(1+d) < 1 . \tag{18.4.4}$$

Then for all $\beta \in \left(0, \varepsilon(b + \lambda - 1)^{-1} \wedge 1\right)$,

$$\Delta_{V_\beta}(P) \leq \gamma_1(\beta, b, \lambda, \varepsilon) \vee \gamma_2(\beta, b, \lambda) < 1 , \tag{18.4.5}$$

with

$$\gamma_1(\beta, b, \lambda, \varepsilon) = 1 - \varepsilon + \beta(b + \lambda - 1) , \tag{18.4.6}$$

$$\gamma_2(\beta, b, \lambda) = 1 - \beta \frac{(1-\lambda)(1+d) - 2b}{2(1-\beta) + \beta(1+d)} . \tag{18.4.7}$$

Proof. Since P satisfies condition $\mathrm{D_g}(V, \lambda, b)$, we have

$$PV_\beta = 1 - \beta + \beta PV \leq 1 - \beta + \beta\lambda V + \beta b = \lambda V_\beta + b_\beta \tag{18.4.8}$$

with $b_\beta = (1-\lambda)(1-\beta) + b\beta$. Thus P also satisfies condition $\mathrm{D_g}(V_\beta, \lambda, b_\beta)$, and applying Lemma 18.3.4 with $r = 1$ yields (18.4.3).

Set $C = \{V \leq d\}$. Since C is a $(1, \varepsilon)$-Doeblin set, for all $x, x' \in \mathsf{X}$,

$$\mathrm{d_{TV}}(P(x, \cdot), P(x', \cdot)) \leq 1 - \varepsilon \mathbb{1}_{C \times C}(x, x') .$$

Since $V_\beta \geq 1$, we can apply Lemma 18.3.1 and the drift condition (18.4.8); we obtain

$$\begin{aligned} \left\| P(x, \cdot) - P(x', \cdot) \right\|_{V_\beta} &\leq PV_\beta(x) + PV_\beta(x') - 2\varepsilon \mathbb{1}_{C \times C}(x, x') \\ &\leq \lambda\{V_\beta(x) + V_\beta(x')\} + 2b_\beta - 2\varepsilon \mathbb{1}_{C \times C}(x, x') . \end{aligned}$$

This yields

$$\frac{\|P(x, \cdot) - P(x', \cdot)\|_{V_\beta}}{V_\beta(x) + V_\beta(x')} \leq \lambda + 2 \frac{b_\beta - \varepsilon \mathbb{1}_C(x, x')}{V_\beta(x) + V_\beta(x')} . \tag{18.4.9}$$

If $(x, x') \notin C \times C$, then $V(x) + V(x') \geq 1 + d$, and hence by (18.4.9),

$$\frac{\|P(x,\cdot)-P(x',\cdot)\|_{V_\beta}}{V_\beta(x)+V_\beta(x')} \le \lambda + \frac{2(1-\lambda)(1-\beta)+2b\beta}{2(1-\beta)+\beta(1+d)} = \gamma_2(\beta,b,\lambda)\,.$$

The function $\beta \mapsto \gamma_2(\beta,b,\lambda)$ is monotone, and since $\gamma_2(0,b,\lambda)=1$, $\gamma_2(1,b,\lambda)=\lambda+2b/(1+d)<1$, it is strictly decreasing, showing that $\gamma_2(\beta,b,\lambda)<1$ for all $\beta\in(0,1]$. If $(x,x')\in C\times C$ and $(1-\lambda)(1-\beta)+b\beta-\varepsilon<0$, then

$$\frac{\|P(x,\cdot)-P(x',\cdot)\|_{V_\beta}}{V_\beta(x)+V_\beta(x')} \le \lambda + 2\frac{(1-\lambda)(1-\beta)+b\beta-\varepsilon}{V_\beta(x)+V_\beta(x')} \le \lambda \le \gamma_2(\beta,b,\lambda)\,.$$

If $(x,x')\in C\times C$ and $(1-\lambda)(1-\beta)+b\beta-\varepsilon>0$, we obtain, using $V_\beta\ge 1$,

$$\begin{aligned}\frac{\|P(x,\cdot)-P(x',\cdot)\|_{V_\beta}}{V_\beta(x)+V_\beta(x')} &\le \lambda+(1-\lambda)(1-\beta)+b\beta-\varepsilon\\ &= 1+\beta(\lambda+b-1)-\varepsilon=\gamma_1(\beta,b,\lambda,\varepsilon)\,,\end{aligned}$$

with $\gamma_1(\beta)<1$ for $\beta\in\big(0,\varepsilon(\lambda+b-1)^{-1}\wedge 1\big)$. □

Theorem 18.4.3. *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$ satisfying the drift condition $D_g(V,\lambda,b)$. Assume, moreover, that there exist $d\ge 1$ and $m\in\mathbb{N}$ such that the level set $\{V\le d\}$ is an (m,ε)-Doeblin set and*

$$\lambda+2b/(1+d)<1\,. \tag{18.4.10}$$

Then $\Delta_V(P^n)<\infty$ for all $n\ge 1$, and there exists an integer $r\ge 1$ such that $\Delta_V(P^r)<1$. Consequently, there exists a unique invariant probability measure π, and P is V-uniformly geometrically ergodic. Moreover, for all $\beta\in\big(0,\varepsilon(b_m+\lambda^m-1)^{-1}\wedge 1\big)$, $n\in\mathbb{N}$, and $\xi\in\mathbb{M}_{1,V}(\mathscr{X})$, we have

$$\mathrm{d}_V(\xi P^n,\pi)\le\beta^{-1}(1+\varepsilon)\,\|\pi-\xi\|_V\,\rho^{\lfloor n/m\rfloor}\,,$$

with

$$\rho=\gamma_1(\beta,b_m,\lambda^m,\varepsilon)\vee\gamma_2(\beta,b_m,\lambda^m)<1\,, \tag{18.4.11}$$
$$b_m=b(1-\lambda^m)(1-\lambda)^{-1}\,, \tag{18.4.12}$$

γ_1, and γ_2 as in (18.4.6) and (18.4.7).

Proof. By Proposition 14.1.8, for $m\ge 1$, P^m satisfies the geometric drift condition $D_g(V,\lambda^m,b_m)$, where b_m is defined in (18.4.12). Note that $b_m/(1-\lambda^m)=b/(1-\lambda)$, and thus $\lambda+2b/(1+d)<1$ if and only if $\lambda^m+2b_m/(1+d)<1$. Moreover, as noted in Remark 14.1.9, $\lambda+b>1$, we have also $\lambda_m+b_m>1$. By Lemma 18.4.2, for every $q=1,\dots,m-1$ and $\beta\in\big(0,\varepsilon(b_m+\lambda^m-1)^{-1}\wedge 1\big)$, we have

$$\Delta_{V_\beta}(P^m) \le \gamma_1(\beta, b_m, \lambda^m, \varepsilon) \vee \gamma_2(\beta, b_m, \lambda^m) < 1 ,$$
$$\Delta_{V_\beta}(P^q) \le 1 + \beta(b_q + \lambda^q - 1) .$$

The condition $b + \lambda > 1$ implies that $b_q + \lambda^q$ is increasing with respect to q and for $q = 1, \ldots, m-1$,

$$\Delta_{V_\beta}(P^q) \le 1 + \beta(b_q + \lambda^q - 1) \le 1 + \varepsilon .$$

We now apply Theorem 18.1.1 to obtain that there exists a unique invariant probability π, $\pi(V_\beta) < \infty$, and for every $n \in \mathbb{N}^\star$ and $\xi \in \mathbb{M}_{1,V_\beta}(\mathscr{X})$, we have

$$\|\xi P^n - \pi\|_{V_\beta} \le (1 + \varepsilon) \|\xi - \pi\|_{V_\beta} \rho^{\lfloor n/m \rfloor} .$$

Since $\|\cdot\|_{V_\beta} = (1-\beta)\|\cdot\|_{\mathrm{TV}} + \beta\|\cdot\|_V$ and $\|\cdot\|_{\mathrm{TV}} \le \|\cdot\|_V$, we have

$$\beta\|\cdot\|_V \le \|\cdot\|_{V_\beta} \le \|\cdot\|_V . \tag{18.4.13}$$

Thus

$$\|\xi P^n - \pi\|_V \le \beta^{-1} \|\xi P^n - \pi\|_{V_\beta} \le \beta^{-1}(1+\varepsilon)\|\xi - \pi\|_V \rho^{\lfloor n/m \rfloor} .$$

Using again (18.4.13), we get $\Delta_V(P^n) \le \beta^{-1}\Delta_{V_\beta}(P^n)$. Thus $\Delta_V(P^n) < \infty$ for all $n \ge 1$, and there exists an integer $r \ge 1$ such that $\Delta_V(P^r) < 1$. □

18.5 Application of Uniform Ergodicity to the Existence of an Invariant Measure

In this section, we apply the result of the previous section to obtain a new proof of the existence and uniqueness of the invariant measure of a recurrent irreducible Markov kernel, Theorem 11.2.5, which does not use the splitting construction. We restate it here for convenience.

Theorem 18.5.1. *Let P be a recurrent irreducible Markov kernel. Then P admits a nonzero invariant measure μ, unique up to multiplication by a positive constant and such that $\mu(C) > 0$ for all petite sets C. Moreover, μ is a maximal irreducibility measure, and for every accessible set A and all $f \in \mathbb{F}_+(\mathsf{X})$,*

$$\mu(f) = \int_A \mu(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_A} f(X_k)\right] = \int_A \mu(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=0}^{\sigma_A - 1} f(X_k)\right] . \tag{18.5.1}$$

Proof. We will prove the existence and uniqueness only up to scaling of a nonzero invariant measure. The proof is, as usual, in several steps, from the case that there exists a Harris-recurrent accessible 1-small set to the general case.

(i) Assume first that P admits an accessible $(1,\varepsilon\nu)$-small set C such that $\mathbb{P}_x(\sigma_C < \infty) = 1$ for all $x \in C$. Then the induced kernel Q_C (see Definition 3.3.7) is a Markov kernel on $C \times \mathscr{X}_C$ given by $Q_C(x,B) = \mathbb{P}_x(X_{\sigma_C} \in B)$ for $x \in C$ and $B \in \mathscr{X}_C$. Define the kernel P_C on C by

$$P_C(x,B) = \mathbb{P}_x(\tau_C \in B) , \qquad x \in C , \; B \in \mathscr{X}_C .$$

Then for $x \in C$ and $B \in \mathscr{X}_C$, using once again the identity $\sigma_C = \tau_C \circ \theta + 1$, the Markov property, and the fact that C is a $(1,\varepsilon\nu)$-small set, we obtain

$$\begin{aligned} Q_C(x,B) &= \mathbb{P}_x(X_{\sigma_C} \in B) = \mathbb{P}_x(X_{\tau_C} \circ \theta \in B) \\ &= \mathbb{E}_x\left[P_C(X_1,B)\right] = PP_C(x,B) \geq \varepsilon\nu P_C(B) . \end{aligned}$$

This proves that Q_C satisfies the uniform Doeblin condition, i.e., C is small for Q_C. Therefore, by Theorem 18.2.4, there exists a unique Q_C-invariant probability measure, which we denote by π_C. Applying Theorem 3.6.3, we obtain that the measure μ defined by

$$\mu(A) = \int_C \pi_C(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_C} \mathbb{1}_A(X_k)\right]$$

is P-invariant, and the restriction of μ to C, denoted by $\mu\,|_C$, is equal to π_C. This proves the existence of an invariant measure for P, and we now prove the uniqueness up to scaling. Let $\tilde{\mu}$ be another P-invariant measure. Then $\tilde{\mu}(C) < \infty$ by Lemma 9.4.12, and thus we can assume without loss of generality that $\tilde{\mu}(C) = 1$. Applying Theorem 3.6.5 yields that the restriction $\tilde{\mu}\,|_C$ of $\tilde{\mu}$ to C is invariant for Q_C. Thus $\tilde{\mu}\,|_C = \pi_C$, since we have just seen that Q_C admits a unique invariant probability measure. Applying Theorem 3.6.3 yields $\mu = \tilde{\mu}$.

(ii) Assume now that C is an accessible strongly aperiodic small set. Then the set $C_\infty = \{x \in \mathsf{X} : \mathbb{P}_x(N_C = \infty) = 1\}$ is full and absorbing by Lemma 10.1.9. Define $\tilde{C} = C \cap C_\infty$. Then for $x \in \tilde{C}$, $\mathbb{P}_x(\sigma_C < \infty) = 1$, and since C_∞ is absorbing, $\mathbb{P}_x(\sigma_C = \sigma_{\tilde{C}}) = 1$ for $x \in \tilde{C}$. This yields that $\mathbb{P}_x(\sigma_{\tilde{C}} = 1)$ for $x \in \tilde{C}$, and we can apply the first part of the proof. Then there exists a unique invariant probability measure $\pi_{\tilde{C}}$ for $Q_{\tilde{C}}$, and the measure μ defined by

$$\mu(f) = \int_{\tilde{C}} \pi_{\tilde{C}}(\mathrm{d}x)\mathbb{E}_x\left[\sum_{k=1}^{\sigma_{\tilde{C}}} f(X_k)\right]$$

is a nonzero invariant measure for P. The uniqueness of the invariant measure is obtained as previously.

(iii) Let us now turn to the general case. If C is a recurrent accessible (m,μ)-small set, we can assume without loss of generality that $\mu(C) > 0$. By Lemma 11.2.2, C is then a recurrent strongly aperiodic accessible small set for the kernel K_{a_η}. The previous step shows that K_{a_η} admits a unique nonzero invariant measure up to scaling. By Lemma 11.2.3, P also has a unique invariant measure up to scaling. (Note that Lemmas 11.2.2 and 11.2.3 are independent of the splitting construction.)

□

18.6 Exercises

18.1. Let $\{X_n,\ n \in \mathbb{N}\}$ be a Markov chain with kernel P and initial distribution μ. Let $1 \le \ell < n$ and let Y be a bounded nonnegative $\sigma(X_j, j \ge n)$-measurable random variable. Prove that

$$\left|\mathbb{E}\left[Y \,|\, \mathscr{F}_\ell^X\right] - \mathbb{E}_\mu[Y]\right| \le \Delta\left(P^{n-\ell}\right) \|Y\|_\infty .$$

[Hint: Write $h(X_n) = \mathbb{E}\left[Y \,|\, \mathscr{F}_n^X\right]$ and note that $|h|_\infty \le \|Y\|_\infty$. Write then

$$\begin{aligned}
\mathbb{E}\left[Y \,|\, \mathscr{F}_\ell^X\right] - \mathbb{E}_\mu[Y] &= \mathbb{E}\left[h(X_n) \,|\, \mathscr{F}_\ell^X\right] - \mathbb{E}_\mu[h(X_n)] \\
&= \mathbb{E}_{X_\ell}[h(X_{n-\ell})] - \int_{\mathsf{X}} \mu P^\ell(\mathrm{d}y)\mathbb{E}_y[h(X_{n-\ell})] \\
&= \int_{\mathsf{X}} \mu P^\ell(\mathrm{d}y)\{P^{n-\ell}h(X_\ell) - P^{n-\ell}h(y)\} .
\end{aligned}$$

Use then the bound (18.2.5) and the fact that the oscillation of a nonnegative function is at most equal to its sup-norm.]

18.2. This exercise provides an example of a Markov chain for which the state space X is $(1,\varepsilon)$-Doeblin but for which X is not 1-small. Consider the Markov chain on $\mathsf{X} = \{1,2,3\}$ with transition probabilities given by

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

1. Show that the stationary distribution is $\pi = [1/3, 1/3, 1/3]$.
2. Show that X is a $(1, 1/2)$-Doeblin set but is not 1-small.
3. Show that for all $n \in \mathbb{N}$, $\sup_{x\in\mathsf{X}} \mathrm{d_{TV}}(P^n(x,\ \cdot), \pi) \le (1/2)^n$.
4. Show that X is $(2, 3/4\pi)$-small and that $\Delta(P^2) = 1/4$ and compute a bound for $\mathrm{d_{TV}}(P^n(x,\ \cdot), \pi)$.

18.3. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that there exist an integer $m \in \mathbb{N}^*$, $\mu \in \mathbb{M}_+(\mathscr{X})$, a measurable function p_m on X^2, and $C \in \mathscr{X}$ such that for all $x \in C$ and $A \in \mathscr{X}$,

$$P^m(x,A) \ge \int_A p_m(x,y)\mu(\mathrm{d}y) . \tag{18.6.1}$$

1. Assume that

$$\varepsilon = \inf_{(x,x') \in C \times C} \int_{\mathsf{X}} p_m(x,y) \wedge p_m(x',y) \mu(\mathrm{d}y) > 0 . \tag{18.6.2}$$

   Show that C is a Doeblin set.
2. Assume that there exists a nonnegative measurable function g_m such that $g_m(y) \leq \inf_{x \in C} p_m(x,y)$ for μ-almost all $y \in \mathsf{X}$ and $\hat{\varepsilon} = \int_{\mathsf{X}} g_m(y) \mu(\mathrm{d}y) > 0$. Show that C is an m-small set.

18.4 (Slice Sampler). Consider the slice sampler as described in Example 2.3.7 in the particular situation in which $k = 1$ and $f_0 = 1$, $f_1 = \pi$.

1. Show that the Markov kernel P of $\{X_n,\ n \in \mathbb{N}\}$ may thus be written as follows: for all $(x,B) \in \mathsf{X} \times \mathscr{X}$,

$$P(x,B) = \frac{1}{\pi(x)} \int_0^{\pi(x)} \frac{\mathrm{Leb}(B \cap L(y))}{\mathrm{Leb}(L(y))} \mathrm{d}y , \tag{18.6.3}$$

   where $L(y) := \{x' \in \mathsf{X} :\ \pi(x') \geq y\}$.
2. Assume that π is bounded and that the topological support $\mathscr{S}_\pi$ of π is such that $\mathrm{Leb}(\mathscr{S}_\pi) < \infty$. Under these assumptions, we will show that P is uniformly ergodic.

18.5. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an invariant probability π. Let $\{\varepsilon_n,\ n \in \mathbb{N}\}$ be a sequence satisfying $\lim_{n\to\infty} \varepsilon_n = 0$. Assume that for all $x \in \mathsf{X}$, $\mathrm{d}_{\mathrm{TV}}(P^n(x,\cdot),\pi) \leq M(x)\varepsilon_n$, where M is a nonnegative function satisfying $M(x) < \infty$ for all $x \in \mathsf{X}$. Then P admits an (m,ε)-Doeblin set.

18.6. Let $\{Z_k,\ k \in \mathbb{N}\}$ and $\{U_k,\ k \in \mathbb{N}\}$ be two independent sequences of i.i.d. random variables on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, the distribution of U_1 being uniform on $[0,1]$. Let $r : \mathbb{R} \to [0,1]$ be a càdlàg nondecreasing function, X_0 a real-valued random variable, and define recursively the sequence $\{X_k,\ k \in \mathbb{N}\}$ by

$$X_k = \begin{cases} X_{k-1} + Z_k & \text{if } U_k \leq r(X_{k-1}) , \\ Z_k & \text{otherwise} . \end{cases} \tag{18.6.4}$$

We assume that ν_Z, the distribution of Z_k, has a continuous positive density f_Z with respect to Lebesgue measure.

1. Show that for all $\varepsilon \in (0,1]$, the set $\{r \leq 1 - \varepsilon\}$ is a $(1, \varepsilon\nu_Z)$-small set.
2. Assume that $\sup_{x \in \mathbb{R}} r(x) < 1$. Show that P is uniformly ergodic.
3. Assume that $\mathbb{E}[Z_0] < 0$ and $\mathbb{E}[\exp(tZ_0)] < \infty$ for some $t > 0$. Show that there exists $s \in (0,t)$ such that the Markov kernel P is V_s-geometrically ergodic, where $V_s(x) = \exp(sx) + 1$.

18.7 Bibliographical Notes

The Dobrushin contraction coefficient was introduced by R. Dobrushin in a series of papers Dobrushin (1956c), Dobrushin (1956b), Dobrushin (1956a). These papers dealt with homogeneous and inhomogeneous Markov chains (on discrete state spaces), the motivation being to obtain limit theorems for additive functionals. The use of Dobrushin contraction coefficients to study convergence of inhomogeneous Markov chains on general state spaces was later undertaken by Madsen (1971); Madsen and Isaacson (1973). Lipschitz contraction properties of Markov kernels over general state spaces (equipped with entropy-like distances) are further studied in Del Moral et al. (2003), where generalizations of the Dobrushin coefficient are introduced.

The analysis of the discrete-state-space independent sampler, Example 18.2.9, is taken from Liu (1996) and uses results from Diaconis and Hanlon (1992) and Diaconis (2009).

The extension to V-norm follows closely the simple and very elegant ideas developed in Hairer and Mattingly (2011), from which we have borrowed Theorem 18.4.3. The definition of V-Dobrushin coefficient is implicit in this work, but this terminology is, to the best of our knowledge, novel.

Chapter 19
Coupling for Irreducible Kernels

This chapter deals with coupling techniques for Markov chains. We will use these technique to obtain bounds for

$$\Delta_n(f,\xi,\xi') = |\xi P^n f - \xi' P^n f| \,,$$

where f belongs to an appropriate class of functions and $\xi,\xi' \in \mathbb{M}_1(\mathscr{X})$. Using the canonical space $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}})$ and the notation of Chapter 3, we can write $\Delta_n(f,\xi,\xi') = |\mathbb{E}_\xi[f(X_n)] - \mathbb{E}_{\xi'}[f(X_n)]|$. In this expression, two different probability measures $\mathbb{P}_\xi$ and $\mathbb{P}_{\xi'}$ are used on the canonical space, and the expectation of the same function $f(X_n)$ is considered under these two probability measures. In contrast, when using coupling techniques, we construct a common probability measure, say $\mathbb{P}_{\xi,\xi'}$, on an extended state space, and denoting by $\mathbb{E}_{\xi,\xi'}$ the associated expectation operator, we show that

$$\Delta_n(f,\xi,\xi') = |\mathbb{E}_{\xi,\xi'}[f(X_n) - f(X'_n)]| \,,$$

where in this case two different random variables $f(X_n)$ and $f(X'_n)$ are involved under a common probability space. The problem then boils down to evaluating on the same probability space the closeness of the two random variables $f(X_n)$ and $f(X'_n)$ in a sense to be defined. There are actually many variations around coupling techniques and many different ways for constructing the common probability space. We first introduce in Section 19.1 general results on the coupling of two probability measures and then introduce the notion of kernel coupling, which will be essential for coupling Markov chains. In Section 19.2, we then state and prove the most fundamental ingredient of this chapter, known as the coupling inequality. We then introduce different variations around coupling, and we conclude this chapter by obtaining bounds for geometric and subgeometric Markov kernels. The expressions of the geometric bounds are of the same flavor as in Chapter 18, but they are obtained here through coupling techniques instead of operator methods and are of a more quantitative nature.

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_19

19.1 Coupling

19.1.1 Coupling of Probability Measures

In this section, we introduce the basics of the coupling technique used to obtain bounds for the total variation distance (or more generally for the V-total variation distance) between two probability measures. For this purpose, it is convenient to express the total variation distance of $\xi, \xi' \in \mathbb{M}_+(\mathscr{X})$ as a function of the total mass of the infimum of measures, denoted by $\xi \wedge \xi'$. The infimum $\xi \wedge \xi'$ is characterized as follows. If η is any measure satisfying $\eta(A) \leq \xi(A) \wedge \xi(A)$ for all $A \in \mathscr{X}$, then $\eta \leq \xi \wedge \xi'$. Moreover, the measures $\xi - \xi \wedge \xi'$ and $\xi' - \xi \wedge \xi'$ are mutually singular. These properties are established in Proposition D.2.8.

Lemma 19.1.1 *For* $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$,

$$\mathrm{d_{TV}}(\xi, \xi') = 1 - (\xi \wedge \xi')(\mathsf{X}) \,. \tag{19.1.1}$$

Proof. Define $\nu = \xi - \xi \wedge \xi'$ and $\nu' = \xi' - \xi \wedge \xi'$. Then ν and ν' are positive and mutually singular measures, and $\nu(\mathsf{X}) = \nu'(\mathsf{X}) = 1 - \xi \wedge \xi'(\mathsf{X})$. Therefore,

$$\mathrm{d_{TV}}(\xi, \xi') = \frac{1}{2}|\xi - \xi'|(\mathsf{X}) = \frac{1}{2}|\nu - \nu'|(\mathsf{X}) = \frac{\nu(\mathsf{X}) + \nu'(\mathsf{X})}{2} = 1 - \xi \wedge \xi'(\mathsf{X}) \,.$$

□

We may interpret this expression of the total variation distance in terms of coupling of probability measures, which we define now. Recall that the Hamming distance is defined on every nonempty set X by $(x, y) \mapsto \mathbb{1}\{x \neq y\}$.

To avoid measurability issues, the following assumption will be in force throughout the chapter.

A 19.1.2 *The diagonal* $\Delta = \{(x,x) \,:\, x \in \mathsf{X}\}$ *is measurable in* $\mathsf{X} \times \mathsf{X}$*, i.e.,* $\Delta \in \mathscr{X} \otimes \mathscr{X}$.

This assumption holds if X is a metric space endowed with its Borel σ-field.

Definition 19.1.3 (Coupling of probability measures)

- *A coupling of two probability measures* $(\xi, \xi') \in \mathbb{M}_1(\mathscr{X}) \times \mathbb{M}_1(\mathscr{X})$ *is a probability measure* γ *on the product space* $(\mathsf{X} \times \mathsf{X}, \mathscr{X} \otimes \mathscr{X})$ *whose marginals are* ξ *and* ξ'*, i.e.,* $\gamma(A \times \mathsf{X}) = \xi(A)$ *and* $\gamma(\mathsf{X} \times A) = \xi'(A)$ *for all* $A \in \mathscr{X}$.
- *The set of all couplings of* ξ *and* ξ' *is denoted by* $\mathscr{C}(\xi, \xi')$.
- *The measure* $\xi \otimes \xi'$ *is called the independent coupling of* ξ *and* ξ'.

- *A coupling $\gamma \in \mathscr{C}(\xi, \xi')$ is said to be optimal for the Hamming distance (or for the total variation distance) if $\gamma(\Delta^c) = \mathrm{d}_{\mathrm{TV}}(\xi, \xi')$.*

It is often convenient to interpret a coupling of probability measures $(\xi, \xi') \in \mathbb{M}_1(\mathscr{X}) \times \mathbb{M}_1(\mathscr{X})$ in terms of the joint distribution of two random variables. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $X, X' : \Omega \to \mathsf{X}$ be X-valued random variables such that $\mathscr{L}_{\mathbb{P}}(X) = \xi$ and $\mathscr{L}_{\mathbb{P}}(X') = \xi'$. Then the joint distribution γ of (X, X') is a coupling of (ξ, ξ'). By a slight abuse of terminology, we will say that (X, X') is a coupling of (ξ, ξ') and write $(X, X') \in \mathscr{C}(\xi, \xi')$.

Example 19.1.4. Let $\xi = \mathrm{N}(-1,1)$ and $\xi' = \mathrm{N}(1,1)$. Let $X \sim \mathrm{N}(-1,1)$ and set $X' = X + 2$. Then (X, X') is a coupling of (ξ, ξ'), but it is not the optimal coupling for the Hamming distance, since $\mathbb{P}(X \neq X') = 1$, whereas by Proposition D.2.8 and Lemma 19.1.1,

$$\mathrm{d}_{\mathrm{TV}}(\xi, \xi') = 1 - \int_{-\infty}^{\infty} \phi(x+1) \wedge \phi(x-1) \mathrm{d}x = 1 - 2 \int_1^{\infty} \frac{\mathrm{e}^{-u^2/2}}{\sqrt{2\pi}} \mathrm{d}u \,,$$

where ϕ is the density of the standard Gaussian distribution.

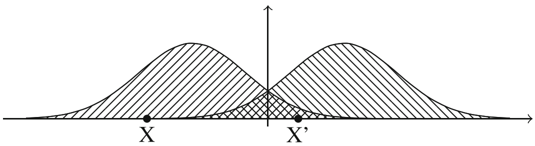

Fig. 19.1 An example of coupling of two probability measures.

Lemma 19.1.5 *Assume* **H** 19.1.2. *If $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$ are mutually singular, then every $\gamma \in \mathscr{C}(\xi, \xi')$ satisfies*

$$\gamma(\Delta) = 1 - \gamma(\Delta^c) = 1 - \mathrm{d}_{\mathrm{TV}}(\xi, \xi') = 0 \,.$$

Equivalently, every coupling (X, X') of (ξ, ξ') defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ satisfies $\mathbb{P}(X = X') = 0$.

Proof. Since ξ and ξ' are singular, there exists a set $A \in \mathscr{X}$ such that $\xi(A^c) = \xi'(A) = 0$. Thus

$$\begin{aligned} \gamma(\Delta) &= \gamma(\{(x,x) : x \in A^c\}) + \gamma(\{(x,x) : x \in A\}) \\ &\leq \gamma(A^c \times \mathsf{X}) + \gamma(\mathsf{X} \times A) = \xi(A^c) + \xi'(A) = 0 \,. \end{aligned}$$

Equivalently,

$$\begin{aligned}\mathbb{P}(X = X') &= \mathbb{P}(X = X', X \in A^c) + \mathbb{P}(X = X', X' \in A) \\ &\le \mathbb{P}(X \in A^c) + \mathbb{P}(X' \in A) = \xi(A^c) + \xi'(A) = 0 .\end{aligned}$$

□

The next result will be used to get a bound for the total variation between two probability measures via coupling.

Theorem 19.1.6. *Assume* **H** 19.1.2 *and let* $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$. *Then*

$$\mathrm{d_{TV}}(\xi, \xi') = \inf_{\gamma \in \mathscr{C}(\xi,\xi')} \gamma(\Delta^c) = \inf_{(X,X') \in \mathscr{C}(\xi,\xi')} \mathbb{P}(X \neq X') . \tag{19.1.2}$$

The probability measures η *and* η' *defined by*

$$\eta = \frac{\xi - \xi \wedge \xi'}{1 - \xi \wedge \xi'(\mathsf{X})} , \qquad \eta' = \frac{\xi' - \xi \wedge \xi'}{1 - \xi \wedge \xi'(\mathsf{X})} . \tag{19.1.3}$$

are mutually singular. A measure $\gamma \in \mathscr{C}(\xi, \xi')$ *is an optimal coupling of* (ξ, ξ') *for the Hamming distance if and only if there exists* $\beta \in \mathscr{C}(\eta, \eta')$ *such that*

$$\gamma(B) = \{1 - \xi \wedge \xi'(\mathsf{X})\}\beta(B) + \int_B \xi \wedge \xi'(\mathrm{d}x)\delta_x(\mathrm{d}x') , \qquad B \in \mathscr{X}^{\otimes 2} . \tag{19.1.4}$$

Proof. (a) Let (X, X') be a coupling of (ξ, ξ') defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\gamma = \mathscr{L}_{\mathbb{P}}(X, X')$. For $f \in \mathbb{F}_b(\mathsf{X})$, we have

$$\xi(f) - \xi'(f) = \mathbb{E}\left[\{f(X) - f(X')\}\mathbb{1}_{\{X \neq X'\}}\right] \le \operatorname{osc}(f)\mathbb{P}(X \neq X') .$$

On the other hand, applying Proposition D.2.4 yields

$$\mathrm{d_{TV}}(\xi, \xi') = \frac{1}{2}\left\|\xi - \xi'\right\|_{\mathrm{TV}} \le \sup\left\{(\xi - \xi')(f) : f \in \mathbb{F}_b(\mathsf{X}), \operatorname{osc}(f) \le 1\right\} ,$$

Hence, for any coupling (X, X') of (ξ, ξ'), we obtain

$$\mathrm{d_{TV}}(\xi, \xi') \le \mathbb{P}(X \neq X') = \gamma(\Delta^c) . \tag{19.1.5}$$

(b) The probability measures η and η' defined in (19.1.3) are mutually singular by Proposition D.2.8 (ii).

(c) We now show that the lower-bound in (19.1.5) is achieved for any coupling γ satisfying (19.1.4). Since η and η' are mutually singular and $\beta \in \mathscr{C}(\eta, \eta')$, Lemma 19.1.5 implies that $\beta(\Delta^c) = 1$. Then, by (19.1.4) and Lemma 19.1.1, we get

$$\gamma(\Delta^c) = \{1 - \xi \wedge \xi'(\mathsf{X})\}\beta(\Delta^c) + \int_{\Delta^c} \xi \wedge \xi'(\mathrm{d}x)\delta_x(\mathrm{d}x') = \mathrm{d_{TV}}(\xi, \xi') . \tag{19.1.6}$$

This shows that the coupling γ is optimal for the Hamming distance.

(d) We finally prove that if γ is an optimal coupling of (ξ,ξ'), then it can be written as in (19.1.4). Define the measure μ on $(\mathsf{X},\mathscr{X})$ by

$$\mu(A)=\gamma(\Delta\cap(A\times\mathsf{X}))\,,\quad A\in\mathscr{X}\,.$$

Since by (19.1.6) $\gamma(\Delta^c)=\mathrm{d}_{\mathrm{TV}}(\xi,\xi')$, we get

$$\mu(\mathsf{X})=\gamma(\Delta)=\xi\wedge\xi'(\mathsf{X})\,. \tag{19.1.7}$$

Moreover,

$$\mu(A)=\gamma(\Delta\cap(A\times\mathsf{X}))\le\gamma(A\times\mathsf{X})=\xi(A)$$

and similarly,

$$\mu(A)=\gamma(\Delta\cap(A\times\mathsf{X}))=\gamma(\Delta\cap(A\times A))=\gamma(\Delta\cap(\mathsf{X}\times A))\le\xi'(A)\,.$$

Thus $\mu\le\xi\wedge\xi'$. Combining with (19.1.7), we obtain that $\mu=\xi\wedge\xi'$. Then, γ can be written as: for all $B\in\mathscr{X}^{\otimes 2}$,

$$\gamma(B)=\gamma(\Delta^c\cap B)+\gamma(\Delta\cap B)=\gamma(\Delta^c)\frac{\gamma(\Delta^c\cap B)}{\gamma(\Delta^c)}+\int_B\xi\wedge\xi'(\mathrm{d}x)\delta_x(\mathrm{d}x')\,. \tag{19.1.8}$$

Plugging $B=A\times\mathsf{X}$ into (19.1.8) and using (19.1.7) yield

$$\frac{\gamma(\Delta^c\cap(A\times\mathsf{X}))}{\gamma(\Delta^c)}=\frac{1}{\gamma(\Delta^c)}\left(\gamma(A\times\mathsf{X})-\int_{A\times\mathsf{X}}\xi\wedge\xi'(\mathrm{d}x)\delta_x(\mathrm{d}x')\right)=\eta(A)\,,$$

where η is defined in (19.1.3). Similarly,

$$\frac{\gamma(\Delta^c\cap(\mathsf{X}\times A))}{\gamma(\Delta^c)}=\eta'(A)\,.$$

Thus $\gamma(\Delta^c\cap\cdot)/\gamma(\Delta^c)\in\mathscr{C}(\eta,\eta')$. The proof is completed.

□

The V-norm can also be characterized in terms of coupling.

Theorem 19.1.7. *Let $\xi,\xi'\in\mathbb{M}_{1,V}(\mathscr{X})$ (see (18.3.3)). Then*

$$\|\xi-\xi'\|_V=\inf_{\gamma\in\mathscr{C}(\xi,\xi')}\int_{\mathsf{X}\times\mathsf{X}}\{V(x)+V(y)\}\mathbb{1}\{x\neq y\}\,\gamma(\mathrm{d}x\mathrm{d}y) \tag{19.1.9}$$

Moreover the infimum is achieved by any coupling $\gamma\in\mathscr{C}(\xi,\xi')$ which is optimal for the Hamming distance.

Proof. Let S be a Jordan set for $\xi - \xi'$ and set $f = V\mathbb{1}_S - V\mathbb{1}_{S^c}$. Then $|f(x)| = |V(x)|$ and $|f(x) - f(y)| \leq (V(x) + V(y))\mathbb{1}\{x \neq y\}$. Set $\rho_V(x,y) = \{V(x) + V(y)\}\mathbb{1}\{x \neq y\}$. Applying the definition of a Jordan set, we obtain, for any $\gamma \in \mathscr{C}(\xi, \xi')$,

$$
\begin{aligned}
\|\xi - \xi'\|_V &= |\xi - \xi'|(V) = (\xi - \xi')(V\mathbb{1}_S) - (\xi - \xi')(V\mathbb{1}_{S^c}) = (\xi - \xi')(f) \\
&= \iint_{\mathsf{X}\times\mathsf{X}} \{f(x) - f(y)\}\gamma(\mathrm{d}x\mathrm{d}y) \leq \iint_{\mathsf{X}\times\mathsf{X}} \rho_V(x,y)\gamma(\mathrm{d}x\mathrm{d}y) .
\end{aligned}
$$

Taking the infimum over $\gamma \in \mathscr{C}(\xi, \xi')$ yields that $\|\xi - \xi'\|_V$ is smaller than the right hand side of (19.1.9). Conversely, let $\gamma \in \mathscr{C}(\xi, \xi')$ be an optimal coupling for the Hamming distance, i.e.,

$$
\gamma(B) = \{1 - \xi \wedge \xi'(\mathsf{X})\}\beta(B) + \int_B \xi \wedge \xi'(\mathrm{d}x)\delta_x(\mathrm{d}x') , \quad B \in \mathscr{X}^{\otimes 2} , \tag{19.1.10}
$$

where $\beta \in \mathscr{C}(\eta, \eta')$ and η and η' are defined by (19.1.3). Then by definition of ρ_V,

$$
\iint_{\mathsf{X}\times\mathsf{X}} \rho_V(x,y)\gamma(\mathrm{d}x\mathrm{d}y) = (\xi - \xi \wedge \xi')(V) + (\xi' - \xi \wedge \xi')(V) = \|\xi - \xi'\|_V .
$$

This shows that the right hand side of (19.1.9) is smaller than and therefore equal to $\|\xi - \xi'\|_V$. □

19.1.2 Kernel Coupling

We now extend the notion of coupling to Markov kernels. We must first define the infimum of two kernels.

Definition 19.1.8 (Infimum of two kernels) *Let P and Q be two kernels on $\mathsf{X} \times \mathscr{X}$. The infimum of the kernels P and Q, denoted $P \wedge Q$, is defined on $\mathsf{X}^2 \times \mathscr{X}$ by*

$$
P \wedge Q(x,x';A) = [P(x,\cdot) \wedge Q(x',\cdot)](A) , \quad x,x' \in \mathsf{X} , \quad A \in \mathscr{X} . \tag{19.1.11}
$$

We do not exclude the case $P = Q$, in which case attention must be paid to the fact that $P \wedge P$ is not equal to P, since these two kernels are not even defined on the same space.

Proposition 19.1.9 *Assume that $\mathscr{X}$ is countably generated. Then, $P \wedge Q$ is a kernel on $(\mathsf{X} \times \mathsf{X}) \times \mathscr{X}$ and the function $(x,x') \mapsto \mathrm{d}_{\mathrm{TV}}(P(x,\cdot), Q(x',\cdot))$ is measurable.*

Proof. By Definition 1.2.1, we only have to prove that, for every $A \in \mathscr{X}$, the function $(x,x') \mapsto (P \wedge Q)(x,x';A)$ is measurable. Since $\mathscr{X}$ is countably generated, by Lemma D.1.4, there exists a countable algebra $\mathscr{A}$ such that for all $x,x' \in \mathsf{X}$ and $A \in \mathscr{X}$,

$$[P(x,\cdot) - Q(x',\cdot)]^+(A) = \sup_{B \in \mathscr{A}} [P(x,\cdot) - Q(x',\cdot)](A \cap B) .$$

The supremum is taken over a countable set, so that the function $(x,x') \mapsto [P(x,\cdot) - Q(x',\cdot)]^+(A)$ is measurable. Similarly, $(x,x') \mapsto [P(x,\cdot) - Q(x',\cdot)]^-(A)$ and $(x,x') \mapsto |P(x,\cdot) - Q(x',\cdot)|(A)$ are measurable. Since

$$P \wedge Q(x,x',\cdot) = P(x,\cdot) + Q(x',\cdot) - |P(x,\cdot) - Q(x',\cdot)| ,$$

we obtain that $P \wedge Q$ is a kernel. Moreover, $\|P(x,\cdot) - Q(x',\cdot)\|_{\mathrm{TV}} = |P(x,\cdot) - Q(x',\cdot)|(\mathsf{X})$ and the function $(x,x') \mapsto \mathrm{d}_{\mathrm{TV}}(P(x,\cdot), Q(x',\cdot))$ is therefore measurable. □

We now extend the coupling of measures to the coupling of kernels.

Definition 19.1.10 (Kernel coupling) *Let P and Q be two Markov kernels on $\mathsf{X} \times \mathscr{X}$. A Markov kernel K on $\mathsf{X}^2 \times \mathscr{X}^{\otimes 2}$ is said to be a* kernel coupling *of (P,Q) if, for all $x,x' \in \mathsf{X}$ and $A \in \mathscr{X}$,*

$$K(x,x';A \times \mathsf{X}) = P(x,A) , \quad K(x,x';\mathsf{X} \times A) = Q(x',A) . \tag{19.1.12}$$

It is said to be an optimal kernel coupling of (P,Q) for the Hamming distance if for all $x,x' \in \mathsf{X}$,

$$K(x,x';\Delta^c) = \int_{\mathsf{X} \times \mathsf{X}} \mathbb{1}\{y \neq y'\} K(x,x';\mathrm{d}y\mathrm{d}y') = \mathrm{d}_{\mathrm{TV}}(P(x,\cdot), Q(x',\cdot)) . \tag{19.1.13}$$

A trivial example of kernel coupling is the independent coupling $K(x,x';C) = \iint P(x,\mathrm{d}y)Q(x',\mathrm{d}y')\mathbb{1}_C(y,y')$. A kernel K on $\mathsf{X}^2 \times \mathscr{X}^{\otimes 2}$ is an optimal kernel coupling of (P,Q) if for all $x,x' \in \mathsf{X}$, the measure $K(x,x';\cdot)$ is an optimal coupling of $(P(x,\cdot), Q(x',\cdot))$.

We now construct an optimal coupling of kernels based on the optimal coupling of measures given in Theorem 19.1.6. To this end, since the total variation distance involves a supremum, we must carefully address the issue of measurability. Define, for $x,x' \in \mathsf{X}$,

$$\varepsilon(x,x') = 1 - \mathrm{d}_{\mathrm{TV}}(P(x,\cdot), Q(x',\cdot)) = (P \wedge Q)(x,x';\mathsf{X}) .$$

Define the kernels R and R' on $\mathsf{X}^2 \times \mathscr{X}$ by

$$R(x,x';\cdot) = \{1-\varepsilon(x,x')\}^{-1}\{P(x,\cdot) - (P\wedge Q)(x,x';\cdot)\} ,$$
$$R'(x,x';\cdot) = \{1-\varepsilon(x,x')\}^{-1}\{Q(x,\cdot) - (P\wedge Q)(x,x';\cdot)\} ,$$

if $\varepsilon(x,x') < 1$ and let $R(x,x';\cdot)$ and $R'(x,x';\cdot)$ be two arbitrary mutually singular probability measures on X if $\varepsilon(x,x') = 1$. Let $\tilde{R}$ be a kernel on $\mathsf{X}^2 \times \mathscr{X}^{\otimes 2}$ such that

$$\tilde{R}(x,x';\cdot) \in \mathscr{C}(R(x,x';\cdot), R'(x,x';\cdot)) , \tag{19.1.14}$$

for all $(x,x') \in \mathsf{X}^2$. Note that $\tilde{R}$ is not a kernel coupling of R and R' in the sense of Definition 19.1.10. One possible choice is defined by $\tilde{R}(x,x';A\times B) = R(x,x';A)R'(x,x';B)$ for all $A,B \in \mathscr{X}$. By construction, the measures $R(x,x';\cdot)$ and $R(x,x';\cdot)$ are mutually singular thus $\bar{R}(x,x';\Delta) = 0$ for all $x,x' \in \mathsf{X}$. Define finally the kernel K on $\mathsf{X}^2 \times \mathscr{X}^{\otimes 2}$ for $x,x' \in \mathsf{X}$ and $B \in \mathscr{X}^{\otimes 2}$ by

$$K(x,x';B) = \{1-\varepsilon(x,x')\}\tilde{R}(x,x';B) + \int_B (P\wedge Q)(x,x';\mathrm{d}u)\delta_u(\mathrm{d}v) . \tag{19.1.15}$$

Proposition 19.1.9 ensures that K is indeed a kernel.

Remark 19.1.11. If $\{(X_k,X'_k),\ k\in\mathbb{N}\}$ is a Markov chain with Markov kernel K, then the transition may be described as follows. For $k \geq 0$, conditionally on (X_k,X'_k), draw conditionally independently random variables $U_{k+1}, Y_{k+1}, Y'_{k+1}, Z_{k+1}$ such that U_{k+1}, (Y_{k+1},Y'_{k+1}) and Z_{k+1} are independent; U_{k+1} is a Bernoulli random variable with mean $\varepsilon(X_k,X'_k)$; (Y_{k+1},Y'_{k+1}) follows the distribution $\tilde{R}(X_k,X'_k;\cdot)$ if $\varepsilon(X_k,X'_k) < 1$ and has an arbitrary distribution otherwise; Z_{k+1} follows the distribution $P\wedge Q(X_k,X'_k;\cdot)/\varepsilon(X_k,X'_k)$ if $\varepsilon(X_k,X'_k) > 0$ and has an arbitrary distribution otherwise. Then, set

$$X_{k+1} = (1-U_{k+1})Y_{k+1} + U_{k+1}Z_{k+1} , \qquad X'_{k+1} = (1-U_{k+1})Y'_{k+1} + U_{k+1}Z_{k+1} .$$

▲

We will now establish that the kernel K defined in (19.1.15) is indeed an optimal kernel coupling of (P,Q) and that in the case $P = Q$, the diagonal is an absorbing set. Thus, in the latter case, if $\{(X_k,X'_k),\ k\in\mathbb{N}\}$ is a bivariate Markov chain with Markov kernel K and if for some $k \geq 0$, $X_k = X'_k$, then the two components remain forever equal.

Theorem 19.1.12. *Let $(\mathsf{X},\mathscr{X})$ be a measurable space such that $\mathscr{X}$ is countably generated and assume* **H** *19.1.2. Let P and Q be Markov kernels on $\mathsf{X}\times\mathscr{X}$ and let K be defined by (19.1.15).*

(i) The kernel K is an optimal kernel coupling of (P,Q) for the Hamming distance.
(ii) If $P = Q$, the diagonal is an absorbing set for K, i.e., $K(x,x;\Delta) = 1$ for all $x \in \mathsf{X}$.
(iii) If $P = Q$, a set C is a $(1,\varepsilon)$-Doeblin set if and only if $K(x,x';\Delta) \geq \varepsilon$ for all $x,x' \in C$.

Proof. (i) By Theorem 19.1.6, K is an optimal kernel coupling of (P,Q).

(ii) If $P = Q$, then by (19.1.15), for all $x \in \mathsf{X}$,

$$K(x,x;\Delta) = \int_{\Delta} (P \wedge P)(x,x;\mathrm{d}u)\delta_u(\mathrm{d}v) = P(x,\mathsf{X}) = 1 .$$

(iii) By definition, $\mathrm{d}_{\mathrm{TV}}(P(x,\cdot),P(x',\cdot)) = K(x,x';\Delta)$, and thus C is a $(1,\varepsilon)$-Doeblin set if and only if $K(x,x';\Delta) \geq \varepsilon$ for all $x \in C$.

□

We now state further properties of the iterates of a kernel coupling K, where the kernel K is not necessarily of the form (19.1.15).

Proposition 19.1.13 *Let P, Q be two Markov kernels on $\mathsf{X} \times \mathscr{X}$ and ξ, ξ' two probability measures on X. Let K be a kernel coupling of (P,Q). Then:*

(i) for every $n \geq 1$, K^n is a kernel coupling of (P^n,Q^n);
(ii) if $\gamma \in \mathscr{C}(\xi,\xi')$ then for every $n \geq 1$, γK^n is a coupling of ξP^n and $\xi' Q^n$.

Proof. (i) The proof is by induction. The property is satisfied for $n = 1$. For $n \geq 1$, we have

$$\begin{aligned} K^{n+1}(x,x';A \times \mathsf{X}) &= \int_{\mathsf{X}\times\mathsf{X}} K(x,x';\mathrm{d}y\mathrm{d}y')K^n(y,y';A \times \mathsf{X}) \\ &= \int_{\mathsf{X}\times\mathsf{X}} K(x,x';\mathrm{d}y\mathrm{d}y')P^n(y,A) \\ &= \int_{\mathsf{X}} P(x,\mathrm{d}y)P^n(y,A) = P^{n+1}(x,A) . \end{aligned}$$

We prove similarly that $K^{n+1}(x,x';\mathsf{X} \times A) = Q^{n+1}(x',A)$, which implies that K^{n+1} is a kernel coupling of (P^{n+1},Q^{n+1}), and this concludes the induction.

(ii) Let us prove that the first marginal of γK^n is ξP^n. For $A \in \mathscr{X}$, we have

$$\begin{aligned} \gamma K^n(A \times \mathsf{X}) &= \int_{\mathsf{X}\times\mathsf{X}} \gamma(\mathrm{d}x\mathrm{d}x')K^n(x,x';A \times \mathsf{X}) \\ &= \int_{\mathsf{X}\times\mathsf{X}} \gamma(\mathrm{d}x\mathrm{d}x')P^n(x,A) = \int_{\mathsf{X}} \xi(\mathrm{d}x)P^n(x,A) = \xi P^n(A) . \end{aligned}$$

The proof that the second marginal of γK^n is $\xi' Q^n$ is similar.

□

19.1.3 Examples of Kernel Coupling

The optimal kernel coupling defined in (19.1.15) is optimal for the Hamming distance, but is not always easy to construct in practice. In the case that there exists a $(1,\varepsilon\nu)$-small set C, one may try to define a kernel coupling in terms of a bivariate chain $\{(X_n,X'_n),\ n\in\mathbb{N}\}$ that in particular has the following properties:

- $\{X_n,\ n\in\mathbb{N}\}$ and $\{X'_n,\ n\in\mathbb{N}\}$ are both Markov chains with kernel P;
- each time X_k and X'_k are simultaneously in C, there is a probability at least ε that coupling occurs, i.e., $X_{k+1}=X'_{k+1}$.

We now give examples of practical constructions.

Example 19.1.14 (Independent coupling). Both chains start independently with kernel P until they reach C simultaneously. That is, if $(X_k,X'_k)=(x,x')\notin C\times C$, then X_{k+1} and X'_{k+1} are drawn independently of each other and from the past with the distributions $P(x,\cdot)$ and $P(x',\cdot)$, respectively. If $(X_k,X_k)=(x,x')\in C\times C$, a coin is tossed with probability of heads ε.

- If the coin comes up heads, then X_{k+1} is sampled from the distribution ν, and we set $X'_{k+1}=X_{k+1}$.
- If the coin comes up tails, then X_{k+1} and X'_{k+1} are sampled independently of each other and from the past from the distributions $(1-\varepsilon)^{-1}(P(x,\cdot)-\varepsilon\nu)$ and $(1-\varepsilon)^{-1}(P(x',\cdot)-\varepsilon\nu)$, respectively.

The chains may remain coupled for a certain amount of time, but there is a positive probability that they will split and evolve again independently until the next coupling.

Formally, the kernel coupling K_1 corresponding to this construction is defined as follows. Set $\bar{P}(x,\cdot)=(1-\varepsilon)^{-1}\{P(x,\cdot)-\varepsilon\nu\}$, $\bar{C}=C\times C$ and let $\tilde{\nu}$ be the measure on $\mathsf{X}\times\mathsf{X}$, concentrated on the diagonal such that $\tilde{\nu}(B)=\int_B \nu(\mathrm{d}x)\delta_x(\mathrm{d}x')$. Then

$$K_1(x,x';\cdot)=P(x,\cdot)\otimes P(x',\cdot)\mathbb{1}_{\bar{C}^c}(x,x') \\ +\varepsilon\tilde{\nu}\mathbb{1}_{\bar{C}}(x,x')+(1-\varepsilon)\bar{P}(x,\cdot)\otimes\bar{P}(x',\cdot)\mathbb{1}_{\bar{C}}(x,x')\ .$$

Then for $x,x'\in\mathsf{X}$ and $A\in\mathscr{X}$,

$$K_1(x,x';A\times\mathsf{X})=P(x,A)\mathbb{1}_{\bar{C}^c}(x,x')+\{\varepsilon\nu(A)+(1-\varepsilon)\bar{P}(x,A)\}\mathbb{1}_{\bar{C}}(x,x')=P(x,A)\ ,$$

and similarly, $K_1(x,x';\mathsf{X}\times A)=P(x',A)$, so that K_1 is a kernel coupling of (P,P). Moreover, $K_1(x,x';\Delta)\geq\varepsilon$ for $(x,x')\in C\times C$. ◀

Example 19.1.15 (Independent, then forever, coupling). The previous construction is modified as follows. If $(X_k,X'_k)=(x,x')$, then the following hold:

- If $x=x'$, then X_{k+1} is sampled from the distribution $P(x,\cdot)$, independently of the past. Then set $X'_{k+1}=X_{k+1}$.
- If $x\neq x'$, then proceed as previously in the independent coupling case.

Formally, the kernel K_2 of the Markov chain $\{(X_n, X'_n),\ n \in \mathbb{N}\}$ is defined by

$$K_2(x,x';\cdot) = \tilde{P}(x,\cdot)\mathbb{1}\left\{x = x'\right\} + K_1(x,x';\cdot)\mathbb{1}\left\{x \neq x'\right\} ,$$

where $\tilde{P}$ is the kernel on $\mathsf{X} \times \mathscr{X}^{\otimes 2}$ defined by $\tilde{P}(x,B) = \int_B P(x,\mathrm{d}x_1)\delta_{x_1}(\mathrm{d}x'_1)$. Then we have, for all $x,x' \in \mathsf{X}$ and $A \in \mathscr{X}$,

$$K_2(x,x';A\times\mathsf{X}) = P(x,A)\mathbb{1}\left\{x = x'\right\} + K_1(x,x';A\times\mathsf{X})\mathbb{1}\left\{x \neq x'\right\} = P(x,A) ,$$

and similarly, $K_2(x,x';\mathsf{X}\times A) = P(x',A)$, showing that K_2 is again a kernel coupling of (P,P). For $(x,x') \in C \times C$,

$$K(x,x';\Delta) = \mathbb{1}\left\{x = x'\right\} + K_1(x,x';\Delta)\mathbb{1}\left\{x \neq x'\right\} \geq \mathbb{1}\left\{x = x'\right\} + \varepsilon\mathbb{1}\left\{x \neq x'\right\} \geq \varepsilon .$$

Moreover, the diagonal is absorbing, i.e., $K_2(x,x;\Delta) = 1$. ◀

Example 19.1.16. Monotone coupling Let $(\mathsf{X},\preceq)$ be a totally ordered set. For $a \in \mathsf{X}$, set $(-\infty,a] = \{x \in \mathsf{X} : x \preceq a\}$ and $[a,\infty) = \{x \in \mathsf{X} : a \preceq x\}$. Let $\mathscr{X}$ be a σ-field that contains the intervals. A measurable real-valued function V on $(\mathsf{X},\mathscr{X})$ is called increasing if $V(x) \leq V(x')$ for all pairs (x,x') such that $x \preceq x'$. A Markov kernel P on $\mathsf{X}\times\mathscr{X}$ is called stochastically monotone if for every $y \in \mathsf{X}$, the map $x \mapsto P(x,(-\infty,y])$ is decreasing. This means that if $x \preceq x'$, then a random variable X with distribution $P(x,\cdot)$ is stochastically dominated by a random variable Y with distribution $P(x',\cdot)$. If P is a stochastically monotone Markov kernel, it is possible to define a kernel coupling K in such a way that the two components $\{X_n,\ n \in \mathbb{N}\}$ and $\{X'_n,\ n \in \mathbb{N}\}$ are pathwise ordered, i.e., their initial orders are preserved at all times, until they eventually merge after coupling.

Assume that C is a $(1,\varepsilon)$-Doeblin set and let K be the optimal kernel coupling given in (19.1.15), that is,

$$K(x,x';B) = \{1-\varepsilon(x,x')\}\tilde{R}(x,x';B) + \int_B (P\wedge P)(x,x';\mathrm{d}u)\delta_u(\mathrm{d}v) ,$$

for $B \in \mathscr{X}^{\otimes 2}$, where $\tilde{R}$ is a kernel on $\mathsf{X}^2 \times \mathscr{X}^{\otimes 2}$ that satisfies (19.1.14). We provide a particular choice of $\tilde{R}$ that preserves the order of the initial conditions. Since the Markov kernel P is stochastically monotone, if $x \preceq x'$, then for all $y \in \mathsf{X}$,

$$R(x,x';(-\infty,y]) \geq R'(x,x';(-\infty,y]) = R(x',x;(-\infty,y]) , \tag{19.1.16}$$

where the last equality follows from $R'(x,x';\cdot) = R(x',x;\cdot)$. For $x,x' \in \mathsf{X}$, let $G(x,x',\cdot)$ be the quantile function of the distribution $R(x,x';\cdot)$. The monotonicity property (19.1.16) implies that for all $u \in (0,1)$ and $x \preceq x'$,

$$G(x,x';u) \preceq G(x',x;u) .$$

Let U be uniformly distributed on $[0,1]$ and define the kernel $\tilde{R}$ on $\mathsf{X}^2 \times \mathscr{X}^{\otimes 2}$ by

$$\tilde{R}(x,x';A) = \mathbb{P}((G(x,x';U),G(x',x;U)) \in A) .$$

Then $\tilde{R}$ satisfies (19.1.14) and preserves the order. Consequently, the associated optimal coupling kernel K also preserves the order of the initial conditions, that is, if $x \preceq x'$, then $X_n \preceq X'_n$ for all $n \geq 0$. ◀

19.2 The Coupling Inequality

We now have all the elements to state the coupling inequality, which is a key tool in Markov chain analysis. Let $\{(X_n, X'_n),\ n \in \mathbb{N}\}$ be the coordinate process on $(\mathsf{X} \times \mathsf{X})^{\mathbb{N}}$. Denote by T the coupling time of $\{(X_n, X'_n),\ n \in \mathbb{N}\}$ defined by

$$T = \inf\left\{n \geq 1 : X_n = X'_n\right\} = \inf\left\{n \geq 1 : (X_n, X'_n) \in \Delta\right\} . \tag{19.2.1}$$

Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$. Let K be a kernel coupling of (P, P) and $\gamma \in \mathscr{C}(\xi, \xi')$. As usual, we denote by $\mathbb{P}_\gamma$ the probability measure on the canonical space that makes $\{(X_n, X'_n),\ n \in \mathbb{N}\}$ a Markov chain with kernel K and initial distribution γ. As usual, the dependence of $\mathbb{P}_\gamma$ on the choice of the kernel coupling K is implicit.

Theorem 19.2.1. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, and K a kernel coupling of (P, P). Let $V : \mathsf{X} \to [1, \infty)$ be a measurable function on X. Then for all $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$ and $\gamma \in \mathscr{C}(\xi, \xi')$,*

$$\left\|\xi P^n - \xi' P^n\right\|_V \leq \mathbb{E}_\gamma[\{V(X_n) + V(X'_n)\}\mathbb{1}\{T > n\}] . \tag{19.2.2}$$

Proof. Let $f \in \mathbb{F}_b(\mathsf{X})$ be such that $|f| \leq V$. For all $0 \leq k \leq n$,

$$\begin{aligned}\mathbb{E}_\gamma[\mathbb{1}\{T = k\}(f(X_n) - f(X'_n))] &= \mathbb{E}_\gamma[\mathbb{1}\{T = k\}\mathbb{E}_{(X_k, X'_k)}[f(X_{n-k}) - f(X'_{n-k})]] \\ &= \mathbb{E}_\gamma[\mathbb{1}\{T = k\}(P^{n-k}f(X_k) - P^{n-k}f(X'_k))] = 0 ,\end{aligned}$$

where the last equality follows from $X_k = X'_k$ on $\{T = k\}$. Thus for all $n \in \mathbb{N}$,

$$\begin{aligned}|\xi P^n(f) - \xi' P^n(f)| = |\mathbb{E}_\gamma[f(X_n) - f(X'_n)]| &= |\mathbb{E}_\gamma[(f(X_n) - f(X'_n))\mathbb{1}\{T > n\}]| \\ &\leq \mathbb{E}_\gamma[\{V(X_n) + V(X'_n)\}\mathbb{1}\{T > n\}] .\end{aligned}$$

The proof is completed by applying the characterization of the V-norm (18.3.4). □

The coupling inequality will be used to obtain rates of convergence in the following way. For a nonnegative sequence r and a measurable function W defined on X^2 such that $V(x) + V(x') \leq W(x, x')$, we have

$$\begin{aligned}\sum_{k=0}^{\infty} r(k) \left\|\xi P^n - \xi' P^n\right\|_V &\le \sum_{k=0}^{\infty} r(k)\mathbb{E}_\gamma[\{V(X_n)+V(X'_n)\}\mathbb{1}\{T>n\}] \\ &\le \mathbb{E}_\gamma\left[\sum_{k=0}^{T-1} r(k)\{V(X_k)+V(X'_k)\}\right] \\ &\le \mathbb{E}_\gamma\left[\sum_{k=0}^{T-1} r(k)W(X_k,X'_k)\right] .\end{aligned}$$

For $V \equiv 1$, choosing $W \equiv 2$ yields

$$\sum_{k=0}^{\infty} r(k)\mathrm{d}_{\mathrm{TV}}(\xi P^n, \xi' P^n) \le \mathbb{E}_\gamma\left[r^0(T-1)\right] .$$

Bounds on the coupling time can be obtained if there exists a set $\bar{C} \in \mathscr{X}^{\otimes 2}$ such that the coupling is successful after a visit of (X_n, X'_n) to $\bar{C}$ with a probability greater than ε; formally, if

$$K(x,x',\Delta) \ge \varepsilon \quad for\ all \quad (x,x') \in \bar{C} . \tag{19.2.3}$$

Define the time of the nth visit to $\bar{C}$ by $\tau_n = \tau_{\bar{C}} + \sigma_{\bar{C}}^{(n-1)} \circ \theta_{\tau_{\bar{C}}}$, $n \ge 1$. At this point we do not know whether the return times to $\bar{C}$ are finite. This will be guaranteed later by drift conditions. Applying the strong Markov property, this yields on the event $\{\tau_n < \infty\}$,

$$\begin{aligned}&\mathbb{P}(T = \tau_n + 1 \,|\, \mathscr{F}_{\tau_n}) = K(X_{\tau_n}, X'_{\tau_n}, \Delta)\mathbb{1}\{T > \tau_n\} \ge \varepsilon \mathbb{1}\{T > \tau_n\} , \\ &\mathbb{P}(T > \tau_n + 1 \,|\, \mathscr{F}_{\tau_n}) = K(X_{\tau_n}, X'_{\tau_n}, \Delta^c)\mathbb{1}\{T > \tau_n\} \le (1-\varepsilon)\mathbb{1}\{T > \tau_n\} .\end{aligned}$$

Therefore, if $\mathbb{P}_{x,x'}(\tau_n < \infty) = 1$, we have for every bounded and $\mathscr{F}_{\tau_n}$-measurable random variable H_n,

$$\mathbb{E}_{x,x'}[H_n \mathbb{1}\{T > \tau_n\}] \le \varepsilon^{-1}\mathbb{E}_{x,x'}[H_n \mathbb{1}\{T = \tau_n + 1\}] , \tag{19.2.4}$$

$$\mathbb{E}_{x,x'}[H_n \mathbb{1}\{T > \tau_n + 1\}] \le (1-\varepsilon)\mathbb{E}_{x,x'}[H_n \mathbb{1}\{T > \tau_n\}] . \tag{19.2.5}$$

In particular, taking $H_n = 1$ yields

$$\mathbb{P}_{x,x'}(T > \tau_{n+1}) \le \mathbb{P}_{x,x'}(T > \tau_n + 1) \le (1-\varepsilon)\mathbb{P}_{x,x'}(T > \tau_n) .$$

This yields inductively

$$\mathbb{P}_{x,x'}(T > \tau_{n+1}) \le \mathbb{P}_{x,x'}(T > \tau_n + 1) \le (1-\varepsilon)^n . \tag{19.2.6}$$

This further implies that if $\mathbb{P}_{x,x'}(\tau_n < \infty) = 1$ for all $n \ge 1$, then $\mathbb{P}_{x,x'}(T < \infty) = 1$, and the number of visits to $\bar{C}$ before coupling occurs is stochastically dominated by a geometric random variable with mean $1/\varepsilon$.

A change of measure formula

Set $\varepsilon(x,x') = K(x,x';\Delta)$. There exist kernels Q and R such that $Q(x,x';\Delta) = 1$ and

$$K(x,x';\cdot) = \varepsilon(x,x')Q(x,x';\cdot) + (1-\varepsilon(x,x'))R(x,x';\cdot) \ . \tag{19.2.7}$$

The kernel Q must satisfy

$$Q(x,x';A) = \frac{K(x,x';A\cap\Delta)}{K(x,x';\Delta)}$$

if $K(x,x';\Delta) \neq 0$, and it can be taken as an arbitrary measure on the diagonal if $\varepsilon(x,x') = K(x,x';\Delta) = 0$. The kernel R is then defined by (19.2.7) if $\varepsilon(x,x') < 1$, and it can be taken as an arbitrary measure on $\mathscr{X}^{\otimes 2}$ if $\varepsilon(x,x') = 1$.

Let $\bar{\mathbb{P}}_{x,x'}$ be the probability measure on the canonical space that makes the canonical process $\{(X_n,X'_n),\ n\in\mathbb{N}\}$ a Markov chain with kernel R starting from (x,x').

Lemma 19.2.2 *Let $n \geq 0$ and let H_n be a bounded $\mathscr{F}_n$-measurable random variable. Then*

$$\mathbb{E}_{x,x'}[H_n \mathbb{1}\{T>n\}] = \bar{\mathbb{E}}_{x,x'}\left[H_n \prod_{i=0}^{n-1}(1-\varepsilon(X_i,X'_i))\right] \ . \tag{19.2.8}$$

Let $\bar{C}$ be a set such that (19.2.3) holds and define $\eta_n = \sum_{i=0}^{n}\mathbb{1}_{\bar{C}}(X_i,X'_i)$, $n\geq 0$, and $\eta_{-1} = 0$. Then

$$\mathbb{E}_{x,x'}[H_n \mathbb{1}\{T>n\}] \leq \bar{\mathbb{E}}_{x,x'}\left[H_n(1-\varepsilon)^{\eta_{n-1}}\right] \ . \tag{19.2.9}$$

Proof. Let h be a bounded measurable function defined on X^2 and let $\bar{h}(x,x') = h(x,x')\mathbb{1}\{x\neq x'\}$. Then $Q\bar{h}\equiv 0$, and since $R(x,x';\Delta)=0$,

$$K\bar{h}(x,x') = (1-\varepsilon(x,x'))R\bar{h}(x,x') = (1-\varepsilon(x,x'))Rh(x,x') \ .$$

By definition of the coupling time T, we then have for all $n\geq 0$,

$$\begin{aligned}
&\mathbb{E}\left[h(X_{n+1},X'_{n+1})\mathbb{1}\{T>n+1\}\,\middle|\,\mathscr{F}_n\right] \\
&\quad = \mathbb{E}\left[h(X_{n+1},X'_{n+1})\mathbb{1}\left\{X_{n+1}\neq X'_{n+1}\right\}\,\middle|\,\mathscr{F}_n\right]\mathbb{1}\{T>n\} \\
&\quad = K\bar{h}(X_n,X'_n)\mathbb{1}\{T>n\} = (1-\varepsilon(X_n,X'_n))Rh(X_n,X'_n)\mathbb{1}\{T>n\} \ .
\end{aligned}$$

The desired property is true for $n=0$. Assume that it is true for some $n\geq 0$. Let H_n be $\mathscr{F}_n$-measurable let and h and $\bar{h}$ be as above. Applying the previous identity and the induction assumption, we obtain

$$
\begin{aligned}
&\mathbb{E}_{x,x'}[H_n h(X_{n+1}, X'_{n+1}) \mathbb{1}\{T > n+1\}] \\
&\quad = \mathbb{E}_{x,x'}[H_n Rh(X_n, X'_n) \mathbb{1}\{T > n\}] \\
&\quad = \bar{\mathbb{E}}_{x,x'}\left[H_n (1-\varepsilon(X_n, X'_n)) Rh(X_n, X'_n) \prod_{i=0}^{n-1} (1-\varepsilon(X_i, X'_i)) \right] \\
&\quad = \bar{\mathbb{E}}_{x,x'}\left[H_n h(X_{n+1}, X'_{n+1}) \prod_{i=0}^{n} (1-\varepsilon(X_i, X'_i)) \right] .
\end{aligned}
$$

This concludes the induction. The bound (19.2.9) follows straightforwardly from (19.2.8), since $1-\varepsilon(x,x') \le 1-\varepsilon \mathbb{1}_{\bar{C}}(x,x')$. □

19.3 Distributional, Exact, and Maximal Coupling

There are coupling techniques more general than the kernel coupling described in Section 19.1.2. To be specific, we now introduce distributional and exact couplings for two general stochastic processes (not only Markov chains), and then we will define coupling times T, which are more general than those in (19.2.1) and for which the classical coupling inequality (19.2.1) still holds. Importantly, we show the existence of maximal distributional coupling times (in a sense given by Definition 19.3.5 below), which therefore implies that the coupling inequalities can be made tight.

Let $(\mathsf{X}, \mathscr{X})$ be a measurable space. In all this section, $\mathbb{Q}$ and $\mathbb{Q}'$ denote two probability measures on the canonical space $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}})$.

Fix $x^* \in \mathsf{X}$. For every X-valued stochastic process $Z = \{Z_n, n \in \mathbb{N}\}$ and $\bar{\mathbb{N}}$-valued random variable T, define the X-valued stochastic process $\theta_T Z$ by $\theta_T Z = \{Z_{T+k}, k \in \mathbb{N}\}$ on $\{T < \infty\}$ and $\theta_T Z = (x^*, x^*, x^*, \dots)$ on $\{T = \infty\}$.

Definition 19.3.1 (Distributional coupling.) *Let $Z = \{Z_n, n \in \mathbb{N}\}$, let $Z' = \{Z'_n, n \in \mathbb{N}\}$ be X-valued stochastic processes, and let T, T' be $\bar{\mathbb{N}}$-valued random variables defined on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$.*

We say that $\{(\Omega, \mathscr{F}, \mathbb{P}, Z, T, Z', T')\}$ is a distributional coupling of $(\mathbb{Q}, \mathbb{Q}')$ if

- *for all $A \in \mathscr{X}^{\otimes \mathbb{N}}$, $\mathbb{P}(Z \in A) = \mathbb{Q}(A)$ and $\mathbb{P}(Z' \in A) = \mathbb{Q}'(A)$;*
- *$(\theta_T Z, T)$ and $(\theta_{T'} Z', T')$ have the same law.*

The random variables T and T' are called the coupling times. The distributional coupling is said to be successful if $\mathbb{P}(T < \infty) = 1$.

From the definition of the distributional coupling, the coupling times T and T' have the same law, and in particular, $\mathbb{P}(T < \infty) = \mathbb{P}(T' < \infty)$. Before stating the classical coupling inequality, we need to introduce some additional notation. For every measure μ on $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes \mathbb{N}})$ and σ-field $\mathscr{G} \subset \mathscr{X}^{\otimes \mathbb{N}}$, we denote by $(\mu)_{\mathscr{G}}$ the restriction of the measure μ to $\mathscr{G}$. Moreover, for all $n \in \mathbb{N}$, define the σ-field $\mathscr{G}_n = \{\theta_n^{-1}(A) : A \in \mathscr{X}^{\otimes \mathbb{N}}\}$.

Lemma 19.3.2 *Let $(\Omega, \mathscr{F}, \mathbb{P}, Z, T, Z', T')$ be a distributional coupling of $(\mathbb{Q}, \mathbb{Q}')$. For all $n \in \mathbb{N}$,*

$$\left\| (\mathbb{Q})_{\mathscr{G}_n} - (\mathbb{Q}')_{\mathscr{G}_n} \right\|_{\mathrm{TV}} \leq 2\mathbb{P}(T > n) . \tag{19.3.1}$$

Proof. Using Definition 19.3.1, we have that for all $A \in \mathscr{X}^{\otimes \mathbb{N}}$,

$$\begin{aligned} \mathbb{P}(\theta_n Z \in A, T \leq n) &= \sum_{k=0}^{n} \mathbb{P}(\theta_k(\theta_T Z) \in A, T = n-k) \\ &= \sum_{k=0}^{n} \mathbb{P}(\theta_k(\theta_{T'} Z') \in A, T' = n-k) = \mathbb{P}(\theta_n Z' \in A, T' \leq n) . \end{aligned}$$

Then, noting that $\mathbb{Q}(\theta_n^{-1}(A)) = \mathbb{P}(\theta_n Z \in A)$,

$$\begin{aligned} \mathbb{Q}(\theta_n^{-1}(A)) - \mathbb{Q}'(\theta_n^{-1}(A)) &= \mathbb{P}(\theta_n Z \in A) - \mathbb{P}(\theta_n Z' \in A) \\ &= \mathbb{P}(\theta_n Z \in A, T > n) - \mathbb{P}(\theta_n Z' \in A, T' > n) \\ &\leq \mathbb{P}(\theta_n Z \in A, T > n) \leq \mathbb{P}(T > n) . \end{aligned}$$

Interchanging (Z, T) and (Z', T') in the previous inequality and noting that T and T' have the same law completes the proof. □

Definition 19.3.3 (Exact coupling) *We say that $(\Omega, \mathscr{F}, \mathbb{P}, Z, Z', T)$ is an exact coupling of $(\mathbb{Q}, \mathbb{Q}')$ if*

- *for all $A \in \mathscr{X}^{\otimes \mathbb{N}}$, $\mathbb{P}(Z \in A) = \mathbb{Q}(A)$ and $\mathbb{P}(Z' \in A) = \mathbb{Q}'(A)$;*
- $\theta_T Z = \theta_T Z'$, $\mathbb{P}-\text{a.s.}$

An exact coupling (Z, Z') of $(\mathbb{Q}, \mathbb{Q}')$ with coupling time T is also a distributional coupling with coupling times (T, T). We now examine the converse when X is a Polish space.

Lemma 19.3.4 *Let $(\mathsf{X}, \mathscr{X})$ be a Polish space. Assume that there exists a successful distributional coupling of $(\mathbb{Q}, \mathbb{Q}')$. Then there exists a successful exact coupling of $(\mathbb{Q}, \mathbb{Q}')$.*

Proof. Let $(\Omega, \mathscr{F}, \mathbb{P}, Z, T, Z', T')$ be a successful distributional coupling of $(\mathbb{Q}, \mathbb{Q}')$ and set $U = (\theta_T Z, T)$ and $U' = (\theta_{T'} Z', T')$. Since the coupling is successful, we can assume without loss of generality that U and U' take values on $\mathsf{X}^{\mathbb{N}} \times \mathbb{N}$. Setting $\mu_1 = \mathscr{L}_{\mathbb{P}}(Z)$, $\mu_2 = \mathscr{L}_{\mathbb{P}}(U) = \mathscr{L}_{\mathbb{P}}(U')$ and $\mu_3 = \mathscr{L}_{\mathbb{P}}(Z')$, we have

$$\mathscr{L}_{\mathbb{P}}(Z, U) \in \mathscr{C}(\mu_1, \mu_2) , \qquad \mathscr{L}_{\mathbb{P}}(U', Z') \in \mathscr{C}(\mu_2, \mu_3) .$$

Since $(\mathsf{X}, \mathscr{X})$ is a Polish space, μ_1 and μ_3 are probability measures on the Polish space $\mathsf{X}^{\mathbb{N}}$. We can therefore apply the gluing lemma, Lemma B.3.12, combined with Remark B.3.13: there exist random variables $(\bar{Z}, \bar{U}, \bar{Z}')$ taking values on $\mathsf{X}^{\mathbb{N}} \times (\mathsf{X}^{\mathbb{N}} \times \mathbb{N}) \times \mathsf{X}^{\mathbb{N}}$ such that

$$\mathscr{L}_{\bar{\mathbb{P}}}(\bar{Z}, \bar{U}) = \mathscr{L}_{\mathbb{P}}(Z, U) \;, \qquad \mathscr{L}_{\bar{\mathbb{P}}}\left(\bar{U}, \bar{Z}'\right) = \mathscr{L}_{\mathbb{P}}\left(U', Z'\right) \;. \tag{19.3.2}$$

Using these two equalities and noting that $\bar{U} = (\bar{V}, \bar{T})$ is an $\mathsf{X}^{\mathbb{N}} \times \mathbb{N}$-valued random variable, we get

$$\bar{\mathbb{P}}(\theta_{\bar{T}} \bar{Z} = \bar{V}) = 1 = \bar{\mathbb{P}}(\theta_{\bar{T}} \bar{Z}' = \bar{V}) \;,$$

which implies $\theta_{\bar{T}} \bar{Z} = \theta_{\bar{T}} \bar{Z}'$ $\mathbb{P} -$a.s. Moreover, using again (19.3.2),

$$\mathscr{L}_{\bar{\mathbb{P}}}(\bar{Z}) = \mathscr{L}_{\mathbb{P}}(Z) = \mathbb{Q} \;, \qquad \mathscr{L}_{\bar{\mathbb{P}}}\left(\bar{Z}'\right) = \mathscr{L}_{\mathbb{P}}\left(Z'\right) = \mathbb{Q}' \;,$$

which shows that $(\bar{Z}, \bar{Z}')$ is an exact successful coupling of $(\mathbb{Q}, \mathbb{Q}')$ with coupling time $\bar{T}$. □

Definition 19.3.5 *A distributional coupling (Z, Z') of $(\mathbb{Q}, \mathbb{Q}')$ with coupling times (T, T') is maximal if for all $n \in \mathbb{N}$,*

$$\left\| (\mathbb{Q})_{\mathscr{G}_n} - (\mathbb{Q}')_{\mathscr{G}_n} \right\|_{\mathrm{TV}} = 2\mathbb{P}(T > n) \;.$$

In words, a distributional coupling is maximal if equality holds in (19.3.1) for all $n \in \mathbb{N}$. Note that

$$\left\| (\mathbb{Q})_{\mathscr{G}_n} - (\mathbb{Q}')_{\mathscr{G}_n} \right\|_{\mathrm{TV}} = 2\left(1 - (\mathbb{Q})_{\mathscr{G}_n} \wedge (\mathbb{Q}')_{\mathscr{G}_n}(\mathsf{X}^{\mathbb{N}})\right) \;,$$

and thus a distributional coupling (Z, Z') of $(\mathbb{Q}, \mathbb{Q}')$ with coupling times (T, T') is maximal if and only if for all $n \in \mathbb{N}$,

$$\mathbb{P}(T \le n) = (\mathbb{Q})_{\mathscr{G}_n} \wedge (\mathbb{Q})_{\mathscr{G}_n}(\mathsf{X}^{\mathbb{N}}) \;. \tag{19.3.3}$$

We now turn to the specific case of Markov chains. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Denote by $\{X_n,\ n \in \mathbb{N}\}$ the coordinate process and define as previously $\mathscr{G}_n = \left\{\theta_n^{-1}(A) : A \in \mathscr{X}^{\otimes \mathbb{N}}\right\}$.

Lemma 19.3.6 *For all $\mu, \nu \in \mathbb{M}_1(\mathscr{X})$,*

$$\left\| \left(\mathbb{P}_\mu\right)_{\mathscr{G}_n} - \left(\mathbb{P}_\nu\right)_{\mathscr{G}_n} \right\|_{\mathrm{TV}} = \|\mu P^n - \nu P^n\|_{\mathrm{TV}} \;.$$

Proof. For all $A \in \mathscr{X}^{\otimes \mathbb{N}}$, by the Markov property,

$$\begin{aligned}\mathbb{P}_\mu(\theta_n^{-1}(A)) - \mathbb{P}_\nu(\theta_n^{-1}(A)) &= \mathbb{E}_\mu[\mathbb{1}_A \circ \theta_n] - \mathbb{E}_\nu[\mathbb{1}_A \circ \theta_n] \\ &= \mathbb{E}_\mu[\mathbb{E}_{X_n}[\mathbb{1}_A]] - \mathbb{E}_\nu[\mathbb{E}_{X_n}[\mathbb{1}_A]].\end{aligned}$$

Since the nonnegative function $x \mapsto \mathbb{E}_x[\mathbb{1}_A]$ is bounded above by 1, the previous inequality implies $\left\| (\mathbb{P}_\mu)_{\mathscr{G}_n} - (\mathbb{P}_\nu)_{\mathscr{G}_n} \right\|_{\mathrm{TV}} \le \|\mu P^n - \nu P^n\|_{\mathrm{TV}}$. Conversely, for all $B \in \mathscr{X}$, set $A = B \times \mathsf{X}^{\mathbb{N}^*}$. Then

$$\begin{aligned}\mu P^n(B) - \nu P^n(B) &= \mathbb{E}_\mu[\mathbb{1}_B(X_n)] - \mathbb{E}_\nu[\mathbb{1}_B(X_n)] \\ &= \mathbb{E}_\mu[\mathbb{1}_A \circ \theta_n] - \mathbb{E}_\nu[\mathbb{1}_A \circ \theta_n] = \mathbb{P}_\mu(\theta_n^{-1}(A)) - \mathbb{P}_\nu(\theta_n^{-1}(A)),\end{aligned}$$

which implies $\|\mu P^n - \nu P^n\|_{\mathrm{TV}} \le \left\| (\mathbb{P}_\mu)_{\mathscr{G}_n} - (\mathbb{P}_\nu)_{\mathscr{G}_n} \right\|_{\mathrm{TV}}$. □

The coupling theorem for Markov chains follows directly from Lemma 19.3.2.

Theorem 19.3.7. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $\mu, \nu \in \mathbb{M}_1(\mathscr{X})$. If $(\Omega, \mathscr{F}, \mathbb{P}, Z, T, Z', T')$ is a distributional coupling of $(\mathbb{P}_\mu, \mathbb{P}_\nu)$, then for all $n \in \mathbb{N}$,*

$$\|\mu P^n - \nu P^n\|_{\mathrm{TV}} \le 2\mathbb{P}(T > n) . \tag{19.3.4}$$

We now turn to the question of maximal coupling for Markov chains.

Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $\mu, \nu \in \mathbb{M}_1(\mathscr{X})$. Set $\gamma_0^{\langle \mu,\nu \rangle} = \mu \wedge \nu$ and $\chi_0^{\langle \mu,\nu \rangle} = \mu$. We now define $\gamma_n^{\langle \mu,\nu \rangle}$ and $\chi_n^{\langle \mu,\nu \rangle}$ for $n \ge 1$. Since for all $n \in \mathbb{N}^*$,

$$(\mu P^{n-1} \wedge \nu P^{n-1}) P \le \mu P^n \wedge \nu P^n \le \mu P^n , \tag{19.3.5}$$

we can define the (nonnegative) measures $\gamma_n^{\langle \mu,\nu \rangle}$ and $\chi_n^{\langle \mu,\nu \rangle}$ on $(\mathsf{X}, \mathscr{X})$ by

$$\begin{aligned}\gamma_n^{\langle \mu,\nu \rangle} &= \mu P^n \wedge \nu P^n - (\mu P^{n-1} \wedge \nu P^{n-1}) P , \\ \chi_n^{\langle \mu,\nu \rangle} &= \mu P^n - (\mu P^{n-1} \wedge \nu P^{n-1}) P = (\mu P^{n-1} - \nu P^{n-1})^+ P .\end{aligned}$$

Above, we made use of the identity $(\lambda - \lambda')^+ = \lambda - \lambda \wedge \lambda'$, valid for all pairs (λ, λ') of probability measures; see Proposition D.2.8 (v). We will make repeated use of this identity. Using again (19.3.5), we have for all $n \ge 0$, $\gamma_n^{\langle \mu,\nu \rangle} \le \chi_n^{\langle \mu,\nu \rangle}$, and we can therefore define the Radon–Nikodym derivative functions $r_n^{\langle \mu,\nu \rangle}$ as follows: for all $n \ge 0$,

$$r_n^{\langle \mu,\nu \rangle} = \frac{\mathrm{d}\gamma_n^{\langle \mu,\nu \rangle}}{\mathrm{d}\chi_n^{\langle \mu,\nu \rangle}} \in [0,1] .$$

Set for all $n \ge 0$, $s_n^{\langle \mu,\nu \rangle} = 1 - r_n^{\langle \mu,\nu \rangle}$. Therefore, $s_0^{\langle \mu,\nu \rangle} = 1 - \frac{\mathrm{d}(\mu \wedge \nu)}{\mathrm{d}\mu} = \frac{\mathrm{d}(\mu - \nu)^+}{\mathrm{d}\mu}$, and for all $n \ge 1$,

$$s_n^{\langle\mu,\nu\rangle} = 1 - \frac{\mathrm{d}\gamma_n^{\langle\mu,\nu\rangle}}{\mathrm{d}\chi_n^{\langle\mu,\nu\rangle}} = \frac{\mathrm{d}(\chi_n^{\langle\mu,\nu\rangle} - \gamma_n^{\langle\mu,\nu\rangle})}{\mathrm{d}\chi_n^{\langle\mu,\nu\rangle}} = \frac{\mathrm{d}(\mu P^n - \nu P^n)^+}{\mathrm{d}\chi_n^{\langle\mu,\nu\rangle}} \in [0,1] \; . \quad (19.3.6)$$

Set $\mathsf{Y} = \mathsf{X} \times [0,1]$ and $\mathscr{Y} = \mathscr{X} \otimes \mathscr{B}([0,1])$. Define the kernel Q on $\mathsf{Y} \times \mathscr{Y}$ as follows: for all $A \in \mathscr{Y}$,

$$Q((x,u),A) = \int P(x,\mathrm{d}x')\mathbb{1}_A(x',u')\mathbb{1}_{[0,1]}(u')\mathrm{d}u' \; .$$

In words, a transition according to Q may be described by moving the first component according to the Markov kernel P and by drawing the second component independently according to a uniform distribution on $[0,1]$.

Let $\{Y_n = (X_n, U_n),\ n \in \mathbb{N}\}$ be the coordinate process associated with the canonical space $(\mathsf{Y}^{\mathbb{N}}, \mathscr{Y}^{\otimes\mathbb{N}})$ equipped with a family of probability measures $(\bar{\mathbb{P}}_\xi)$ induced by the Markov kernel Q and initial distributions ξ on $(\mathsf{Y},\mathscr{Y})$. The notation $\bar{\mathbb{E}}_\xi$ stands for the associated expectation operator. Define the stopping times T and T' by

$$T = \inf\left\{ i \in \mathbb{N} : U_i \le r_i^{\langle\mu,\nu\rangle}(X_i) \right\} , \qquad T' = \inf\left\{ i \in \mathbb{N} : U_i \le r_i^{\langle\nu,\mu\rangle}(X_i) \right\} .$$

Lemma 19.3.8 *For all nonnegative or bounded measurable functions V on $(\mathsf{X},\mathscr{X})$ and all $n \ge 0$,*

$$\bar{\mathbb{E}}_{\mu\otimes\mathrm{Unif}(0,1)}[V(X_n)\mathbb{1}\{T > n\}] = (\mu P^n - \nu P^n)^+(V) \; . \quad (19.3.7)$$

Proof. The proof is by induction on n. For $n = 0$,

$$\begin{aligned}\bar{\mathbb{E}}_{\mu\otimes\mathrm{Unif}(0,1)}[V(X_0)\mathbb{1}\{T > 0\}] &= \bar{\mathbb{E}}_{\mu\otimes\mathrm{Unif}(0,1)}[V(X_0)\mathbb{1}\left\{U_0 > r_0^{\langle\mu,\nu\rangle}(X_0)\right\}] \\ &= \bar{\mathbb{E}}_{\mu\otimes\mathrm{Unif}(0,1)}[V(X_0)\{1 - r_0^{\langle\mu,\nu\rangle}(X_0)\}] \\ &= (\mu - \mu\wedge\nu)(V) = (\mu - \nu)^+(V) \; .\end{aligned}$$

Assume that (19.3.7) holds for $n \ge 0$. Then, applying successively the Markov property, the induction assumption, and the change of measures (19.3.6), we have

$$\begin{aligned}&\bar{\mathbb{E}}_{\mu\otimes\mathrm{Unif}(0,1)}[V(X_{n+1})\mathbb{1}\{T > n+1\}] \\ &\quad= \bar{\mathbb{E}}_{\mu\otimes\mathrm{Unif}(0,1)}[V(X_{n+1})\mathbb{1}\left\{U_{n+1} > r_{n+1}^{\langle\mu,\nu\rangle}(X_{n+1})\right\}\mathbb{1}\{T > n\}] \\ &\quad= \bar{\mathbb{E}}_{\mu\otimes\mathrm{Unif}(0,1)}[V(X_{n+1})s_{n+1}^{\langle\mu,\nu\rangle}(X_{n+1})\mathbb{1}\{T > n\}] \\ &\quad= \bar{\mathbb{E}}_{\mu\otimes\mathrm{Unif}(0,1)}[P(Vs_{n+1}^{\langle\mu,\nu\rangle})(X_n)\mathbb{1}\{T > n\}] \\ &\quad= (\mu P^n - \mu P^n \wedge \nu P^n)P(Vs_{n+1}^{\langle\mu,\nu\rangle}) \qquad \text{(by the induction assumption)} \\ &\quad= \chi_{n+1}^{\langle\mu,\nu\rangle}(Vs_{n+1}^{\langle\mu,\nu\rangle}) \qquad \text{(by definition of} \chi_{n+1}^{\langle\mu,\nu\rangle}) \\ &\quad= (\mu P^{n+1} - \nu P^{n+1})^+(V) . \qquad \text{(by definition of } s_{n+1}^{\langle\mu,\nu\rangle})\end{aligned}$$

This proves that (19.3.7) holds with n replaced by $n+1$. □

Theorem 19.3.9. *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$ and $\mu,\nu\in\mathbb{M}_1(\mathscr{X})$. There exists a maximal distributional coupling of $(\mathbb{P}_\mu,\mathbb{P}_\nu)$. If $(\mathsf{X},\mathscr{X})$ is Polish, then there exists a maximal and exact coupling of $(\mathbb{P}_\mu,\mathbb{P}_\nu)$.*

Proof. By definition, $\bar{\mathbb{P}}_{\mu\otimes\mathrm{Unif}(0,1)}(X\in\cdot)=\mathbb{P}_\mu$ and $\bar{\mathbb{P}}_{\nu\otimes\mathrm{Unif}(0,1)}(X\in\cdot)=\mathbb{P}_\nu$. Moreover,

$$\bar{\mathbb{E}}_{\mu\otimes\mathrm{Unif}(0,1)}[V(X_0)\mathbb{1}\{T=0\}]=\mu(Vr_0^{\langle\mu,\nu\rangle})=\mu\left(V\frac{\mathrm{d}\gamma_0^{\langle\mu,\nu\rangle}}{\mathrm{d}\chi_0^{\langle\mu,\nu\rangle}}\right)=\gamma_0^{\langle\mu,\nu\rangle}(V)\,.$$

Applying Lemma 19.3.8, we get for all bounded measurable functions V on $(\mathsf{X},\mathscr{X})$,

$$\begin{aligned}
&\bar{\mathbb{E}}_{\mu\otimes\mathrm{Unif}(0,1)}[V(X_n)\mathbb{1}\{T=n\}]\\
&\qquad=\bar{\mathbb{E}}_{\mu\otimes\mathrm{Unif}(0,1)}[V(X_n)\mathbb{1}\{T>n-1\}]-\bar{\mathbb{E}}_{\mu\otimes\mathrm{Unif}(0,1)}[V(X_n)\mathbb{1}\{T>n\}]\\
&\qquad=(\mu P^{n-1}-\nu P^{n-1})^+(PV)-(\mu P^n-\nu P^n)^+(V)\\
&\qquad=\mu P^nV-(\mu P^{n-1}\wedge\nu P^{n-1})PV-\mu P^nV+(\mu P^n\wedge\nu P^n)V=\gamma_n^{\langle\mu,\nu\rangle}V\,.
\end{aligned}$$

Let $A\in\mathscr{X}^{\otimes\mathbb{N}}$ and set $V(x)=\mathbb{E}_x[\mathbb{1}_A]$. Applying the previous equality yields

$$\bar{\mathbb{P}}_{\mu\otimes\mathrm{Unif}(0,1)}(\theta^nX\in A,T=n)=\bar{\mathbb{E}}_{\mu\otimes\mathrm{Unif}(0,1)}[V(X_n)\mathbb{1}\{T=n\}]=\gamma_n^{\langle\mu,\nu\rangle}V\,.$$

Similarly,

$$\bar{\mathbb{P}}_{\nu\otimes\mathrm{Unif}(0,1)}(\theta^nX\in A,T'=n)=\gamma_n^{\langle\nu,\mu\rangle}V\,.$$

Since $\gamma_n^{\langle\mu,\nu\rangle}=\gamma_n^{\langle\nu,\mu\rangle}$, this shows that the laws of (X,T) under $\bar{\mathbb{P}}_{\mu\otimes\mathrm{Unif}(0,1)}$ and (X,T') under $\bar{\mathbb{P}}_{\nu\otimes\mathrm{Unif}(0,1)}$ are a distributional coupling of $(\mathbb{P}_\mu,\mathbb{P}_\nu)$. Moreover, taking $V=\mathbb{1}_{\mathsf{X}}$ in (19.3.7) yields

$$2\bar{\mathbb{E}}_{\mu\otimes\mathrm{Unif}(0,1)}[\mathbb{1}\{T>n\}]=2(\mu P^n-\nu P^n)^+(\mathsf{X})=\|\mu P^n-\nu P^n\|_{\mathrm{TV}}\,,$$

i.e., the distributional coupling is maximal. The last part of the theorem follows from Lemma 19.3.4. □

Remark 19.3.10. Note that by Lemma 19.3.8, for all nonnegative measurable functions V on $(\mathsf{X},\mathscr{X})$,

$$\begin{aligned}
\|\mu P^n-\nu P^n\|_V&=(\mu P^n-\nu P^n)^+V+(\nu P^n-\mu P^n)^+V\\
&=\bar{\mathbb{E}}_{\mu\otimes\mathrm{Unif}(0,1)}[V(X_n)\mathbb{1}\{T>n\}]+\bar{\mathbb{E}}_{\nu\otimes\mathrm{Unif}(0,1)}[V(X_n)\mathbb{1}\{T'>n\}]\,.
\end{aligned}$$

This implies

$$\sum_{n=0}^{\infty} r(n) \|\mu P^n - \nu P^n\|_V$$
$$= \bar{\mathbb{E}}_{\mu\otimes\mathrm{Unif}(0,1)}\left[\sum_{n=0}^{T-1} r(n)V(X_n)\right] + \bar{\mathbb{E}}_{\nu\otimes\mathrm{Unif}(0,1)}\left[\sum_{n=0}^{T'-1} r(n)V(X_n)\right] .$$

▲

19.4 A Coupling Proof of V-Geometric Ergodicity

In this section, we use coupling techniques to give a new proof of Theorem 18.4.3 with more explicit constants.

Theorem 19.4.1. *Let P be a Markov kernel satisfying the drift condition $\mathrm{D_g}(V,\lambda,b)$. Assume, moreover, that there exist an integer $m \geq 1$, $\varepsilon \in (0,1]$, and $d > 0$ such that the level set $\{V \leq d\}$ is an (m,ε)-Doeblin set and $\lambda + 2b/(1+d) < 1$. Then P admits a unique invariant probability measure π, $\pi(V) < \infty$, and for all $\xi \in \mathbb{M}_{1,V}(\mathscr{X})$ and $n \geq 1$,*

$$\mathrm{d}_V(\xi P^n, \pi) \leq c_m\{\pi(V) + \xi(V)\}\rho^{\lfloor n/m \rfloor} , \tag{19.4.1}$$

with

$$\log \rho = \frac{\log(1-\varepsilon)\log \bar{\lambda}_m}{\log(1-\varepsilon) + \log \bar{\lambda}_m - \log \bar{b}_m} , \tag{19.4.2a}$$
$$\bar{\lambda}_m = \lambda^m + 2b_m/(1+d) , \quad \bar{b}_m = \lambda^m b_m + d \quad , \quad b_m = b/(1-\lambda) , \tag{19.4.2b}$$
$$c_m = \{\lambda^m + (1-\lambda^m)/(1-\lambda)\}\{1 + \bar{b}_m/[(1-\varepsilon)(1-\bar{\lambda}_m)]\} . \tag{19.4.2c}$$

By Lemma 18.3.4, $\Delta_V(P^q) \leq \lambda^m + b(1-\lambda^m)/(1-\lambda)$ for all $q < m$ with b_m as in (19.4.2b). Thus for $n = mk + q$, $0 \leq q < m$,

$$\mathrm{d}_V(P^n(x,\cdot), P^n(x',\cdot)) \leq \Delta_V(P^q)\, \mathrm{d}_V(P^{km}(x,\cdot), p^{km}(x',\cdot)) .$$

Moreover, by Proposition 14.1.8, $P^m V \leq \lambda^m V + b_m$. Thus it suffices to prove Theorem 19.4.1 for $m = 1$. We will first obtain bounds for the kernel coupling under a bivariate drift condition (Lemma 19.4.2) and then extend the drift condition $\mathrm{D_g}(V,\lambda,b)$ to the kernel coupling.

Lemma 19.4.2 *Let P be a kernel on $(\mathsf{X},\mathscr{X})$, K a kernel coupling of (P,P), and $\bar{C} \in \mathscr{X}^{\otimes 2}$ a set such that $K(x,x';\Delta) \geq \varepsilon$ for all $(x,x') \in \bar{C}$. Assume that there exist a measurable function $\bar{V} : \mathsf{X}^2 \to [1,\infty]$ and constants $\bar{\lambda} \in (0,1)$ and $\bar{b} > 0$ such that*

$$K\bar{V} \le \bar{\lambda}\bar{V}\mathbb{1}_{\bar{C}^c} + \bar{b}\mathbb{1}_{\bar{C}} . \tag{19.4.3}$$

Then for all $x, x' \in \mathsf{X}$ and $n \ge 0$,

$$\mathbb{P}_{(x,x')}(T > n) \le \left(\bar{V}(x,x') + 1\right)\rho^n , \tag{19.4.4a}$$

$$\mathbb{E}_{(x,x')}[\bar{V}(X_n, X'_n)\mathbb{1}\{T > n\}] \le \left(2\bar{V}(x,x') + \bar{b}(1-\bar{\lambda})^{-1}\right)\rho^n , \tag{19.4.4b}$$

where T is the coupling time defined in (19.2.1) and

$$\log \rho = \frac{\log(1-\varepsilon)\log\bar{\lambda}}{\log(1-\varepsilon) + \log\bar{\lambda} - \log\bar{b}} . \tag{19.4.5}$$

Proof. The drift condition (19.4.3) yields

$$(1-\varepsilon(x,x'))R\bar{V}(x,x') \le KV(x,x') \le \bar{\lambda}^{\mathbb{1}_{\bar{C}^c}(x,x')}\bar{b}^{\mathbb{1}_{\bar{C}}(x,x')}\bar{V}(X_n, X'_n) .$$

Set $H_0 = 1$, $H_n = \prod_{i=0}^{n-1}(1-\varepsilon(X_i, X'_i))$, $n \ge 1$, and $Z_n = \bar{\lambda}^{-n+\eta_{n-1}}\bar{b}^{-\eta_{n-1}}H_n\bar{V}(X_n, X'_n)$, $n \ge 0$. This yields

$$\begin{aligned}
\bar{\mathbb{E}}\left[Z_{n+1} \,|\, \mathscr{F}_n\right] &= \bar{\lambda}^{-n-1+\eta_n}\bar{b}^{-\eta_n}H_{n+1}R\bar{V}(X_n, X'_n) \\
&\le \bar{\lambda}^{-n-1+\eta_n}\bar{b}^{-\eta_n}H_{n+1}\frac{\bar{\lambda}^{\mathbb{1}_{\bar{C}^c}(X_n,X'_n)}\bar{b}^{\mathbb{1}_{\bar{C}}(X_n,X'_n)}}{1-\varepsilon(X_n, X'_n)}\bar{V}(X_n, X'_n) = Z_n .
\end{aligned}$$

Thus $\{Z_n,\ n \in \mathbb{N}\}$ is a positive supermartingale under $\bar{\mathbb{P}}$. Let $m > 0$ (not necessarily an integer). Then applying the change of measure formula (19.2.8), we obtain

$$\begin{aligned}
&\mathbb{E}_{x,x'}[\bar{V}(X_n, X'_n)\mathbb{1}\{T > n\}] \\
&\quad = \bar{\mathbb{E}}_{x,x'}[\bar{V}(X_n, X'_n)H_n] \\
&\quad = \bar{\mathbb{E}}_{x,x'}[\bar{V}(X_n, X'_n)H_n\mathbb{1}\{\eta_{n-1} > m\}] + \bar{\mathbb{E}}_{x,x'}[\bar{V}(X_n, X'_n)H_n\mathbb{1}\{\eta_{n-1} \le m\}] \\
&\quad \le (1-\varepsilon)^m\bar{\mathbb{E}}_{x,x'}[\bar{V}(X_n, X'_n)] + \bar{\lambda}^{n-m}\bar{b}^m\bar{\mathbb{E}}[Z_n] \\
&\quad \le (1-\varepsilon)^m\{\bar{\lambda}^n\bar{V}(x,x') + \bar{b}/(1-\bar{\lambda})\} + \bar{\lambda}^{n-m}\bar{b}^m\bar{\mathbb{E}}[Z_0] \\
&\quad \le (1-\varepsilon)^m\{\bar{\lambda}^n\bar{V}(x,x') + \bar{b}/(1-\bar{\lambda})\} + \bar{\lambda}^{n-m}\bar{b}^m\bar{V}(x,x') .
\end{aligned}$$

Similarly, replacing $\bar{V}$ by 1 on the left-hand side and using $1 \le \bar{V}$, we obtain

$$\begin{aligned}
\mathbb{P}_{x,x'}(T > n) = \bar{\mathbb{E}}_{x,x'}[H_n] &\le \bar{\mathbb{E}}_{x,x'}[H_n\mathbb{1}\{\eta_{n-1} \ge m\}] \\
&\quad + \bar{\mathbb{E}}_{x,x'}[\bar{V}(X_n, X'_n)H_n\mathbb{1}\{\eta_{n-1} < m\}] \\
&\le (1-\varepsilon)^m + \bar{\lambda}^{n-m}\bar{b}^m\bar{V}(x,x') .
\end{aligned}$$

We now choose m such that

$$\frac{m}{n} = \frac{\log\bar{\lambda}}{\log(1-\varepsilon) + \log\bar{\lambda} - \log\bar{b}} .$$

Then $(1-\varepsilon)^m = \bar{b}^m \bar{\lambda}^{n-m} = \rho^n$ with ρ as in (19.4.5), and the bounds (19.4.4) follow from the previous ones. □

We now have all the ingredients to prove Theorem 19.4.1.

Proof (of Theorem 19.4.1). Let K be the optimal kernel coupling of (P^m, P^m) defined in (19.1.15). Set $C = \{V \le d\}$ and $\bar{C} = C \times C$. Then for all $x, x' \in C$,

$$K(x,x';\Delta) = (P^m \wedge P^m)(x,x';\mathsf{X}) \ge \varepsilon .$$

Define $\bar{V}(x,x') = \{V(x) + V(x')\}/2$. By Proposition 14.1.8, $P^m V \le \lambda^m V + b_m$, where b_m is defined in (19.4.2b). Since K is a kernel coupling of (P^m, P^m), we obtain that for all $x, x' \in \mathsf{X}$,

$$K\bar{V}(x,x') = \frac{P^m V(x) + P^m V(x')}{2} \le \lambda^m \bar{V}(x,x') + b_m .$$

If $(x,x') \notin C \times C$, then $\bar{V}(x,x') \ge (1+d)/2$ and

$$K\bar{V}(x,x') \le \lambda^m \bar{V}(x,x') + \frac{2b_m}{1+d} \bar{V}(x,x') = \bar{\lambda}_m \bar{V}(x,x') . \tag{19.4.6}$$

If $(x,x') \in C \times C$, then $\bar{V}(x,x') \le d$ and

$$K\bar{V}(x,x') \le \lambda^m d + b_m . \tag{19.4.7}$$

Thus the drift condition (19.4.3) holds with $\bar{\lambda} = \bar{\lambda}_m$ and $\bar{b} = \bar{b}_m$ defined in (19.4.2b). The assumptions of Lemma 19.4.2 hold, and thus we can apply it. Combining with Theorem 19.2.1, this yields, for all $x, x' \in \mathsf{X}$ and all $n \ge 1$,

$$\begin{aligned} \mathrm{d}_V(P^{nm}(x,\cdot), P^{nm}(x',\cdot)) \mathbb{E}_{x,x'}[(V(X_n) + V(X'_n)) \mathbb{1}\{T > n\}] \\ \le \left(2\bar{V}(x,x') + \bar{b}_m \{(1-\varepsilon)(1-\bar{\lambda}_m)\}^{-1}\right) \rho^n . \end{aligned}$$

By Lemma 18.3.4, $\Delta_V(P^q) \le \lambda^m + b(1-\lambda^m)/(1-\lambda)$ for all $q \le m$. Thus for $n = mk + q$, $0 \le q < m$, this yields

$$\begin{aligned} \mathrm{d}_V(P^n(x,\cdot), P^n(x',\cdot)) &\le \Delta_V(P^q)\, \mathrm{d}_V(P^{km}(x,\cdot), p^{km}(x',\cdot)) \\ &\le \left(\lambda^m + \frac{b(1-\lambda^m)}{1-\lambda}\right) \left(2\bar{V}(x,x') + \frac{\bar{b}_m}{(1-\varepsilon)(1-\bar{\lambda}_m)}\right) \rho^{[n/m]} . \end{aligned} \tag{19.4.8}$$

Applying Theorem 18.1.1 yields the existence and uniqueness of the invariant measure π and $\pi(V) < \infty$. Integrating (19.4.8) with respect to π and ξ yields (19.4.1).
□

19.5 A Coupling Proof of Subgeometric Ergodicity

The main result of this section provides subgeometric rates of convergence under the drift condition $D_{sg}(V,\phi,b,C)$ introduced in Definition 16.1.7. Recall that if ϕ is concave, the subgeometric sequence r_ϕ is defined in (16.1.13) by $r_\phi(t) = \phi \circ H_\phi^{-1}(t)$, where H_ϕ is the primitive of $1/\phi$ that vanishes at 1. Rates slower than r_ϕ will be obtained by interpolation. Let Ψ_1 and Ψ_2 defined on $[0,\infty)$ be a pair of inverse Young functions, that is, such that $\Psi_1(x)\Psi_2(y) \le x+y$ for all $x,y \ge 0$. For simplicity, we will consider only the case in which C is a $(1,\varepsilon)$-Doeblin set.

Theorem 19.5.1. *Let C be a $(1,\varepsilon)$-Doeblin set. Assume that condition $D_{sg}(V,\phi,b,C)$ holds with $\sup_C V < \infty$ and $d = \inf_{C^c} \phi \circ V > b$. Let (Ψ_1,Ψ_2) be a pair of inverse Young functions, let $\kappa \in (0, 1-b/d)$, and set $r(n) = \Psi_1(r_\phi(\kappa n))$ and $f = \Psi_2(\phi \circ V)$.*

(i) There exists a constant ϑ such that for every $\xi,\xi' \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{n=0}^{\infty} r(n) \mathrm{d}_f(\xi P^n, \xi' P^n) < \vartheta\{\xi(V) + \xi'(V)\} . \tag{19.5.1}$$

(ii) There exists a unique invariant probability measure π, $\pi(\phi \circ V) < \infty$, and there exists ϑ such that for every $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{n=0}^{\infty} \Delta r(n) \mathrm{d}_f(\xi P^n, \pi) < \vartheta \xi(V) . \tag{19.5.2}$$

Moreover, for every $x \in \mathsf{X}$, $\lim_{n\to\infty} r_\phi(\kappa n) \mathrm{d}_{TV}(P^n(x,\cdot),\pi) = 0$.

(iii) If $\pi(V) < \infty$, then there exists ϑ such that for all initial distributions ξ,

$$\sum_{n=0}^{\infty} r(n) \mathrm{d}_f(\xi P^n, \pi) < \vartheta \xi(V) . \tag{19.5.3}$$

The proof of Theorem 19.5.1 follows the same path as the proof of Theorem 19.4.1. Let $W : \mathsf{X} \times \mathsf{X} \to [1,\infty]$ be a measurable function and define

$$\bar{W}_r(x,x') = \mathbb{E}_{x,x'}\left[\sum_{k=0}^{\tau_{\bar{C}}} r(k) W(X_k, X_k')\right] , \tag{19.5.4}$$

$$\bar{W}_r^* = \sup_{(x,x')\in\bar{C}} \mathbb{E}_{x,x'}\left[\sum_{k=1}^{\sigma_{\bar{C}}} r(k) W(X_k, X_k')\right] . \tag{19.5.5}$$

Lemma 19.5.2 *Let K be a kernel coupling of (P,P) and let $\bar{C}$ be such that $K(x,x';\Delta) \ge \varepsilon$ for all $(x,x') \in \bar{C}$. Assume, moreover, that $r \in \Lambda_2$ and $\bar{W}_r^* < \infty$. Then* *for all* $\delta \in (0,1)$, *there exists a finite constant* C_δ *such that for all* $x,x' \in \mathsf{X}$,

$$\mathbb{E}_{x,x'}\left[\sum_{k=0}^{T-1} r(k)W(X_k,X'_k)\right] \le \frac{\bar{W}_r(x,x')+C_\delta}{1-\delta} \, . \tag{19.5.6}$$

Proof. If $\bar{W}_r(x,x') = \infty$, there is nothing to prove, so we can assume that $\bar{W}_r(x,x') < \infty$. Since $\bar{W}_r^* < \infty$ by assumption, this implies that $\mathbb{P}_{x,x'}(\sigma_{\bar{C}}^{(n)} < \infty) = 1$ for all $n \ge 1$ by Proposition 3.3.6. This further implies that $\mathbb{P}_{x,x'}(T < \infty) = 1$ (see (19.2.6) and the comments thereafter). Applying successively the strong Markov property (19.2.5) and (19.2.4), we obtain

$$\begin{aligned}
&\sum_{k=0}^{n} r(k)\mathbb{E}_{x,x'}[W(X_k,X'_k)\mathbb{1}\{T>k\}] \\
&\quad\le \bar{W}_r(x,x') + \sum_{i=1}^{\infty}\mathbb{E}_{x,x'}\left[\sum_{k=\tau_i+1}^{\tau_{i+1}} r(k)W(X_k,X'_k)\mathbb{1}\{T>k\}\mathbb{1}\{\tau_i \le n\}\right] \\
&\quad\le \bar{W}_r(x,x') \\
&\qquad + \sum_{i=1}^{\infty}\mathbb{E}_{x,x'}\left[r(\tau_i)\mathbb{1}\{T>\tau_i+1\}\left(\sum_{k=1}^{\sigma_{\bar{C}}} r(k)W(X_k,X'_k)\right)\circ\theta_{\tau_i}\mathbb{1}\{\tau_i\le n\}\right] .
\end{aligned}$$

Hence we get

$$\begin{aligned}
&\sum_{k=0}^{n} r(k)\mathbb{E}_{x,x'}[W(X_k,X'_k)\mathbb{1}\{T>k\}] \\
&\quad\le \bar{W}_r(x,x') + \bar{W}_r^*\sum_{i=1}^{\infty}\mathbb{E}_{x,x'}[r(\tau_i)\mathbb{1}\{T>\tau_i+1\}\mathbb{1}\{\tau_i\le n\}] \\
&\quad\le \bar{W}_r(x,x') + (1-\varepsilon)\bar{W}_r^*\sum_{i=1}^{\infty}\mathbb{E}_{x,x'}[r(\tau_i)\mathbb{1}\{T>\tau_i\}\mathbb{1}\{\tau_i\le n\}] \\
&\quad\le \bar{W}_r(x,x') + \varepsilon^{-1}(1-\varepsilon)\bar{W}_r^*\sum_{i=1}^{\infty}\mathbb{E}_{x,x'}[r(\tau_i)\mathbb{1}\{T=\tau_i+1\}\mathbb{1}\{\tau_i\le n\}] \\
&\quad\le \bar{W}_r(x,x') + \varepsilon^{-1}(1-\varepsilon)\bar{W}_r^*\sum_{k=0}^{n} r(k)\mathbb{P}_{x,x'}(T=k+1) \, .
\end{aligned}$$

Since $r \in \Lambda_2$, for every $\delta > 0$ there exists a finite constant C_δ such that

$$C_\delta = \sup_{k\ge 0}\{\varepsilon^{-1}(1-\varepsilon)\bar{W}_r^* r(k) - \delta r^0(k)\} \, ,$$

where $r^0(n) = \sum_{k=0}^{n} r(k)$. Since $W \ge 1$, this yields

$$
\begin{aligned}
&\sum_{k=0}^{n} r(k)\mathbb{E}_{x,x'}[W(X_k,X'_k)\mathbb{1}\{T>k\}] \\
&\qquad\qquad \leq \bar{W}_r(x,x') + C_\delta + \delta\sum_{k=0}^{n} r^0(k)\mathbb{P}_{x,x'}(T=k+1) \\
&\qquad\qquad \leq \bar{W}_r(x,x') + C_\delta + \delta\sum_{j=0}^{n} r(j)\mathbb{E}_{x,x'}[W(X_j,X'_j)\mathbb{1}\{T>j\}] \; .
\end{aligned}
$$

This proves that

$$
\sum_{k=0}^{n} r(k)\mathbb{E}_{x,x'}[W(X_k,X'_k)\mathbb{1}\{T>k\}] \leq \frac{\bar{W}_r(x,x')+C_\delta}{1-\delta} \; .
$$

Letting n tend to infinity yields (19.5.6). □

We now prove that if C is a $(1,\varepsilon)$-Doeblin set and if $\mathrm{D_{sg}}(V,\phi,b,C)$ holds, then the kernel K satisfies condition $\mathrm{D_{sg}}(\bar{V},\bar{\phi},\bar{b},\bar{C})$ with $\bar{C}=C\times C$ and for a suitable choice of the functions $\bar{\phi}$ and $\bar{V}$ and the constant $\bar{b}$.

Lemma 19.5.3 *Assume that the drift condition* $\mathrm{D_{sg}}(V,\phi,b,C)$ *holds with V bounded on C. Set $d=\inf_{x\notin C}\phi\circ V(x)$ and $\bar{C}=C\times C$. If $d>b$, then for $\kappa\in(0,1-b/d)$,*

$$
K\bar{V}+\bar{\phi}\circ\bar{V} \leq \bar{V}+\bar{b}\mathbb{1}_{\bar{C}} \tag{19.5.7}
$$

with $\bar{V}(x,x')=V(x)+V(x')-1$, $\bar{\phi}=\kappa\phi$ and $\bar{b}=2b$.

Proof. First consider the case $(x,x')\notin\bar{C}$. Then $\mathbb{1}_C(x)+\mathbb{1}_C(x')\leq 1$, and since ϕ is increasing and $\bar{V}(x,x')\geq V(x)\vee V(x')$,

$$
\phi\circ\bar{V}(x,x') \geq \phi\circ V(x)\vee\phi\circ V(x') \geq d \; .
$$

The choice of κ then implies that $b\leq(1-\kappa)d\leq(1-\kappa)\phi\circ\bar{V}(x,x')$ for $(x,x')\notin C\times C$. The function ϕ being concave, it follows that for all $u,v\geq 1$,

$$
\phi(u+v-1) \leq \phi(u)+\phi(v)-\phi(1) \leq \phi(u)+\phi(v) \; .
$$

Since K is a coupling kernel of (P,P), we have, for $(x,x')\notin\bar{C}$,

$$
\begin{aligned}
K\bar{V}(x,x')+\bar{\phi}\circ\bar{V}(x,x') &\leq K\bar{V}(x,x')+(\kappa\phi\circ\bar{V}(x,x')+b)-b \\
&\leq K\bar{V}(x,x')+\phi\circ\bar{V}(x)-b \\
&\leq PV(x)+PV(x')-1+\phi\circ V(x)+\phi\circ V(x')-b \\
&\leq \bar{V}(x,x')+b\left\{\mathbb{1}_C(x)+\mathbb{1}_C(x')\right\}-b \leq \bar{V}(x,x') \; .
\end{aligned}
$$

If $(x,x')\in\bar{C}$, we have, using that $\bar{\phi}\leq\phi$,

$$
\begin{aligned}
K\bar{V}(x,x')+\bar{\phi}\circ\bar{V}(x,x') &\leq PV(x)+PV(x')-1+\phi\circ V(x)+\phi\circ V(x') \\
&\leq \bar{V}(x,x')+2b \; .
\end{aligned}
$$

This proves that $\mathrm{D_{sg}}(\bar{V},\bar{\phi},\bar{b},\bar{C})$ holds. □

Lemma 19.5.4 *(i) (Toeplitz's lemma) Let $\{a_n,\ n \in \mathbb{N}\}$ be a sequence of positive numbers such that $b_n = \sum_{i=1}^n a_i \to \infty$. Let $\{x_n,\ n \in \mathbb{N}\}$ be a sequence such that $\lim_{n\to\infty} x_n = x_\infty$. Then $\lim_{n\to\infty} b_n^{-1} \sum_{i=1}^n a_i x_i = x_\infty$.*
(ii) (Kronecker's lemma) Let $\{x_n,\ n \in \mathbb{N}\}$ be a sequence of numbers such that the series $\sum x_n$ converges. Let $\{b_n,\ n \in \mathbb{N}\}$ be an increasing sequence of positive numbers such that $\lim_{n\to\infty} b_n = \infty$. Then $b_n^{-1} \sum_{i=1}^n b_i x_i \to 0$.

Proof. (Hall and Heyde 1980, Section 2.6). □

Proof (of Theorem 19.5.1).
(i) Let K an optimal kernel coupling of (P,P) as defined in (19.1.15). Set $\bar{C} = C \times C$. Then $K(x,x';\Delta) = (P \wedge P)(x,x';\mathsf{X}) \geq \varepsilon$ if $x,x' \in C$. Define $\bar{V}(x,x') = V(x) + V(x') - 1$. Set $f = \Psi_2(\phi \circ V)$. Applying Theorem 19.2.1 and Lemma 19.5.2 with $r(n) = \Psi_1(r_\phi(\kappa n))$ and $W(x,x') = f(x) + f(x')$ yields

$$\begin{aligned}\sum_{k=0}^{\infty} r(k) \mathrm{d}_f(P^k(x,\cdot),P^k(x',\cdot)) &\leq \sum_{k=0}^{\infty} r(k) \mathbb{E}_{(x,x')}[\{f(X_k)+f(X'_k)\}\mathbb{1}\{T>k\}] \\ &= \sum_{k=0}^{\infty} r(k) \mathbb{E}_{(x,x')}[W(X_k,X'_k)\mathbb{1}\{T>k\}] \leq \vartheta \bar{W}_r(x,x')\,,\end{aligned} \tag{19.5.8}$$

where the function $\bar{W}_r$ is defined in (19.5.4) and ϑ is a finite constant, provided that the quantity $\bar{W}_r^*$ defined in (19.5.5) is finite. Since $\bar{V}(x,x') \geq V(x) \vee V(x')$ and ϕ is increasing, it is also the case that

$$\phi \circ V(x) + \phi \circ V(x') \leq 2\phi \circ \bar{V}(x,x') = 2\kappa^{-1}\bar{\phi} \circ \bar{V}(x,x')\,.$$

Since (Ψ_1,Ψ_2) is a pair of inverse Young functions, this yields

$$\begin{aligned}\bar{W}_r(x,x') &\leq \mathbb{E}_{(x,x')}\left[\sum_{k=0}^{\tau_{\bar{C}}} r_\phi(\kappa k)\right] + \mathbb{E}_{(x,x')}\left[\sum_{k=0}^{\tau_{\bar{C}}} \{\phi \circ V(X_k) + \phi \circ V(X'_k)\}\right] \\ &\leq \mathbb{E}_{(x,x')}\left[\sum_{k=0}^{\tau_{\bar{C}}} r_\phi(\kappa k)\right] + 2\kappa^{-1}\mathbb{E}_{(x,x')}\left[\sum_{k=0}^{\tau_{\bar{C}}} \bar{\phi} \circ \bar{V}(X_k,X'_k)\right]\,.\end{aligned}$$

Similarly,

$$\bar{W}_r^* \leq \sup_{x,x' \in C} \mathbb{E}_{(x,x')}\left[\sum_{k=1}^{\sigma_{\bar{C}}} r_\phi(\kappa k)\right] + 2\kappa^{-1} \sup_{x,x' \in C} \mathbb{E}_{(x,x')}\left[\sum_{k=1}^{\sigma_{\bar{C}}} \bar{\phi} \circ \bar{V}(X_k,X'_k)\right]\,.$$

By Lemma 19.5.3 and Theorem 16.1.12 and since $r_\phi \in \Lambda_2$ and $r_{\bar{\phi}}(k) = \kappa r_\phi(\kappa k)$, we have

$$\mathbb{E}_{(x,x')}\left[\sum_{k=1}^{\sigma_{\bar{C}}} r_\phi(\kappa k)\right] \le r_\phi(\kappa)\kappa^{-1}\mathbb{E}_{(x,x')}\left[\sum_{k=0}^{\sigma_{\bar{C}}-1} \kappa r_\phi(\kappa k)\right]$$
$$\le r_\phi(\kappa)\kappa^{-1}\bar{V}(x,x') + \bar{b}\frac{\kappa^{-1}r_\phi^2(\kappa)}{\phi(1)}\mathbb{1}_{\bar{C}}(x,x')$$

and

$$\mathbb{E}_{(x,x')}\left[\sum_{k=1}^{\sigma_{\bar{C}}} \bar{\phi}\circ\bar{V}(X_k,X'_k)\right] \le \mathbb{E}_{(x,x')}\left[\sum_{k=0}^{\sigma_{\bar{C}}-1} \bar{\phi}\circ\bar{V}(X_k,X'_k)\right] + \sup_{x,x'\in C}\bar{\phi}\circ\bar{V}(x,x')$$
$$\le \bar{V}(x,x') + \bar{b}\mathbb{1}_{\bar{C}}(x,x') + 2\kappa\sup_{x\in C}\phi\circ V(x)\,.$$

Since V is bounded on C and $\bar{V} \ge 1$, this yields $\bar{W}_r^* < \infty$, and there exists a constant ϑ' such that

$$\bar{W}_r(x,x') \le \vartheta'\bar{V}(x,x')\,.$$

Plugging this bound into (19.5.8) and integrating the resulting bound with respect to the initial distributions yields (19.5.1).

(ii) Taking $\Psi_1(u) = u$, $\Psi_2(v) = 1$, $\xi = \delta_x$, and successively $\xi' = \delta_x P$ and $\xi' = \delta_{x'}$, we obtain, for all $x,x' \in \mathsf{X}$,

$$\sum_{n=0}^{\infty} r_\phi(\kappa n)\mathrm{d}_{\mathrm{TV}}(\delta_x P^n, \delta_x P^{n+1}) < \infty\,,$$
$$\sum_{n=0}^{\infty} r_\phi(\kappa n)\mathrm{d}_{\mathrm{TV}}(\delta_x P^n, \delta_{x'} P^n) < \infty\,.$$

This implies that for each $x \in \mathsf{X}$, $\{P^n(x,\cdot),\ n \ge 0\}$ is a Cauchy sequence in $\mathbb{M}_1(\mathscr{X})$ endowed with the total variation distance that is a complete metric space by Theorem D.2.7. Therefore, there exists a probability measure π such that $P^n(x,\cdot)$ converges to π in total variation distance and π does not depend on the choice of $x \in \mathsf{X}$. Then for all $A \in \mathsf{X}$ and all $x \in \mathsf{X}$, $\pi(A) = \lim_{n\to\infty} P^{n+1}(x,A) = \lim_{n\to\infty} P^n(x, P\mathbb{1}_A) = \pi P(A)$, showing that π is invariant. Moreover, if $\tilde{\pi}$ is an invariant probability measure, then $\tilde{\pi}(A) = \tilde{\pi}P^n(A) = \lim_{n\to\infty}\int\tilde{\pi}(\mathrm{d}x)P^n(x,A)$. Since $P^n(x,A)$ is bounded and converges to $\pi(A)$ as n tends to infinity, Lebesgue's dominated convergence theorem shows that $\tilde{\pi}(A) = \pi(A)$. Thus the invariant probability measure is unique. Since $\mathrm{D_{sg}}(V,\phi,b,C)$ holds and $\sup_{x\in C} V(x) < \infty$, Proposition 4.3.2 implies that $\pi(\phi\circ V) < \infty$. Moreover, applying (19.5.1) with $\xi' = \xi P$ (and noting that $\xi PV \le \xi(V) + b$) yields

$$\sum_{n=0}^{\infty} \Delta r(n) \mathrm{d}_f(\xi P^n, \pi) \le \sum_{n=0}^{\infty} \Delta r(n) \sum_{k=n}^{\infty} \mathrm{d}_f(\xi P^k, \xi P^{k+1})$$
$$= \sum_{k=0}^{\infty} r(k) \mathrm{d}_f(\xi P^k, \xi P^{k+1}) < \vartheta \xi(V) ,$$

for some constant ϑ. This proves (19.5.2). Taking again $\Psi_1(u) = u$ and $\Psi_2(v) = 1$, we get

$$\sum_{n=0}^{\infty} \Delta r_\phi(n) \mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) < \infty . \tag{19.5.9}$$

Since $n \mapsto \mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi)$ is decreasing, Lemma 19.5.4 with $x_n = \Delta r_\phi(n) \mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi)$ and $b_n = \mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi)^{-1}$ implies that

$$\lim_{n\to\infty} b_n^{-1} \sum_{i=0}^{n} b_i x_i = \lim_{n\to\infty} r_\phi(n) \mathrm{d}_{\mathrm{TV}}(\xi P^n, \pi) = 0 .$$

(iii) Finally, if $\pi(V) < \infty$, then (19.5.3) follows from (19.5.1) with $\xi' = \pi$. □

Remark 19.5.5. The pairs (r, f) for which we can prove (19.5.1) or (19.5.3) are obtained by interpolation between r_ϕ and $\phi \circ V$ by means of pairs of inverse Young functions. There is a noticeable trade-off between the rate of convergence to the invariant probability and the size of functions that can be controlled at this rate: the faster the rate, the flatter the function. This is in sharp contrast to the situation in which the kernel P is V-geometrically ergodic.

The assumption $\inf_{C^c} \phi \circ V > b$ is not restrictive if V is unbounded. Indeed, if $C \subset C'$, then $\mathrm{D_{sg}}(V, \phi, b, C)$ implies $\mathrm{D_{sg}}(V, \phi, b, C')$, and enlarging the set C increases $\inf_{C^c} \phi \circ V$. ▲

19.6 Exercises

19.1. Let $\alpha > \beta$, $\xi = \mathrm{Pn}(\alpha)$, and $\xi' = \mathrm{Pn}(\beta)$. Let (X, Y) be two independent random variables such that $X \sim \mathrm{Pn}(\alpha)$ and $Y \sim \mathrm{Pn}(\beta - \alpha)$. Set $X' = X + Y$.

1. Show that (X, X') is a coupling of (ξ, ξ').
2. Show that this coupling is not optimal.

19.2. Let $\varepsilon \in (0,1)$ and let $\xi = \mathrm{Unif}([0,1])$, $\xi' = \mathrm{Unif}([\varepsilon, 1+\varepsilon])$. Construct an optimal coupling of ξ and ξ'.

19.3. Let $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$ be such that $\xi \wedge \xi'(\mathsf{X}) = \varepsilon \in (0,1)$. Let η, η' be the probability measures defined in (19.1.3) and $\beta \in \mathscr{C}(\eta, \eta')$. Let U be a Bernoulli random variable with mean ε, Z a random variable independent of U with distribution $\varepsilon^{-1}\xi \wedge \xi'$, and (Y,Y) a random pair with distribution β, independent of U and Z. Let (X,X') be defined by

$$X = (1-U)Y + UZ\,, \quad X' = (1-U)Y' + UZ\,.$$

Show that (X,X') is an optimal (for the Hamming distance) coupling of (ξ,ξ').

19.4. Let ξ, ξ' be two probability measures on a measurable space $(\mathsf{E},\mathscr{E})$. Show that for $f \in \mathrm{L}^p(\xi) \cap \mathrm{L}^p(\xi')$, we have

$$|\xi(f) - \xi'(f)| \leq (\|f\|_{L^p(\xi)} + \|f\|_{L^p(\xi')})\mathrm{d}_{\mathrm{TV}}^{1/q}(\xi,\xi')\,.$$

19.5. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Let $\varepsilon \in (0,1)$. Show that $\Delta(P) \leq 1-\varepsilon$ if and only if there exists a kernel coupling K of (P,P) such that for all $x,x' \in \mathsf{X}$,

$$K(x,x';\Delta) \geq \varepsilon\,. \tag{19.6.1}$$

19.6. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that the state space X is a $(1,\varepsilon\nu)$-small set for P. Provide an alternative proof of Theorem 18.2.4 by coupling.

19.7. We consider Ehrenfest's urn, which was introduced in Exercise 1.11. Recall that the chain $\{X_n,\ n \in \mathbb{N}\}$ counts the number of red balls in an urn containing red and green balls. At each instant, a ball is randomly drawn and replaced by a ball of the other color. It is a periodic chain with period 2. In order to simplify the discussion, we will make it aperiodic by assuming that instead of always jumping from one state to an adjacent one, it may remain at the same state with probability $1/2$.

1. Write the associated Markov kernel P.
2. For simplicity, we consider only the case N even. Using Exercise 19.5, show that $\Delta\left(P^{N/2}\right) \leq 1-(2N)^{-N}(N!/(N/2)!)^2$.

19.8. Consider the random scan Gibbs sampler for a positive distribution π on the state space $\mathsf{X} = \{0,1\}^d$, i.e., the vertices of a d-dimensional hypercube (so that $|\mathsf{X}| = 2^d$). Given $X_k = x = (x_1,\ldots,x_d)$, the next value $X_{k+1} = z = (z_1,\ldots,z_d)$ is obtained by the following algorithm:

(a) Choose I_{k+1} uniformly in $\{1,2,\ldots,d\}$, independently of the past;
(b) set $z_i = x_i$ for $i \neq I_{k+1}$;
(c) for $i = I_{k+1}$, z_{k+1} is drawn independently of the past as a Bernoulli random variable with success probability

$$\pi_{i,x}(1) = \frac{\pi(x_1,\ldots,x_{i-1},1,x_{i+1},\ldots,x_d)}{\pi(x_1,\ldots,x_{i-1},0,x_{i+1},\ldots,x_d) + \pi(x_1,\ldots,x_{i-1},1,x_{i+1},\ldots,x_d)}\,.$$

Set $\pi_{i,x}(0) = 1-\pi_{i,x}(1)$ and for $i = 1,\ldots,d$ and $\zeta \in \{0,1\}$,

$$x_i^{\zeta} = (x_1,\ldots,x_{i-1},\zeta,x_{i+1},\ldots,x_d)\,.$$

The kernel P of this chain is given by $P(x,x_i^\zeta) = d^{-1}\pi_{i,x}(\zeta)$ for $i \in \{1,\dots,d\}$ and $\zeta \in \{0,1\}$ and $P(x,z) = 0$ if $\sum_{i=1}^d |x_i - z_i| > 1$. Set

$$M = \frac{\min_{x\in\mathsf{X}} \pi(x)}{\max_{x\in\mathsf{X}} \pi(x)} .$$

Assume for simplicity that d is even. We will prove that $\Delta\left(P^{d/2}\right) \le 1-\varepsilon$ by a coupling construction.

1. Show that for all $x \in \mathsf{X}$, $M/(1+M) \le \pi_{i,x}(\zeta) \le 1/(1+M)$.
2. Let $\{(X_k, X'_k),\ k \in \mathbb{N}\}$ be two chains starting from x and x'. Update X_k into X_{k+1} by the previous algorithm, and if $I_{k+1} = i$, then set $I'_{k+1} = d-i+1$ and proceed with the update of X'_k. Give an expression for the kernel K of this Markov chain.
3. Compute a lower bound for the probability of coupling after $d/2$ moves.
4. Compute an upper bound of $\Delta\left(P^{d/2}\right)$.

Let us now compare the bound of $\Delta\left(P^{d/2}\right)$ obtained using the coupling construction and the bound that can be deduced from a uniform minorization of the Markov kernel over the whole state space.

5. Show that X cannot be m-small if $m < d$.
6. Show that X is $(d, M^{d-1}d!d^{-d}\pi)$-small.

19.9. Let $(\mathsf{X}, \preceq)$ be a totally ordered set. Assume that P is a stochastically monotone Markov kernel on $\mathsf{X}\times\mathscr{X}$ and that there exists an increasing function $V : \mathsf{X} \to [1,\infty)$ such that the drift condition $\mathrm{D_g}(V,\lambda,b)$ holds. Assume that there exists $x_0 \in \mathsf{X}$ such that $(-\infty, x_0]$ is a $(1,\varepsilon)$-Doeblin set and $\lambda + b/V(x_0) < 1$. Suppose that P has an invariant probability measure π such that $\pi(V) < \infty$. Using the optimal kernel coupling K described in Example 19.1.16 and the function $\bar{V}$ defined by $\bar{V}(x,x') = V(x \vee x')$, show that for all $\xi \in \mathbb{M}_{1,V}(\mathscr{X})$ and $n \ge 1$,

$$\mathrm{d}_V(\xi P^n, \pi) \le \{\pi(V) + \xi(V) + \bar{b}(1-\bar{\lambda})^{-1}\rho^n \tag{19.6.2}$$

with $\bar{\lambda} = \lambda + b/V(x_0)$, $\bar{b} = \lambda V(x_0) + b$ and ρ as in (19.4.5).

19.7 Bibliographical Notes

The use of coupling for Markov chains can be traced to the early work of Doeblin (1938). The coupling method was then popularized to get rates of convergence of Markov chains by Pitman (1974), Griffeath (1975), Griffeath (1978), Lindvall (1979). The books Lindvall (1992) and Thorisson (2000) provide a very complete account of coupling methods with many applications to Markov chains.

The coupling method to establish geometric ergodicity of general state space Markov chains, Theorem 19.4.1, is essentially taken from Rosenthal (1995b) with the minor improvement presented in Rosenthal (2002). The definition of the coupling kernel and the change of measure formula, Lemma 19.2.2, is taken from Douc et al. (2004b)). We have also borrowed some technical tricks that appeared earlier in Roberts and Tweedie (1999). The surveys Rosenthal (2001), Roberts and Rosenthal (2004) contain many examples of the use of coupling to assess the convergence of the MCMC algorithm.

The monotone coupling technique to study stochastically ordered Markov chains using (Example 19.1.16) is inspired by the work of Lund and Tweedie (1996) (see also Lund et al. (1996) for extension to Markov processes). The details of the proofs are nevertheless rather different.

The coupling construction was used to establish subgeometric rates of convergence in Douc et al. (2006) and Douc et al. (2007). Theorem 19.5.1 is adapted from these two publications.

The paper Roberts and Rosenthal (2011), which presents a different coupling construction adapted to the analysis of the independence sampler, is also of great interest.

Part IV
Selected Topics

Chapter 20
Convergence in the Wasserstein Distance

In the previous chapters, we obtained rates of convergence in the total variation distance of the iterates P^n of an irreducible positive Markov kernel P to its unique invariant measure π for π-almost every $x \in \mathsf{X}$ and for all $x \in \mathsf{X}$ if the kernel P is irreducible and positive Harris recurrent. Conversely, convergence in the total variation distance for all $x \in \mathsf{X}$ entails irreducibility and that π be a maximal irreducibility measure.

Therefore, if P is not irreducible, convergence in the total variation distance cannot hold, and it is necessary to consider weaker distances on the space of probability measures. For this purpose, as in Chapter 12, we will consider Markov kernels on metric spaces, and we will investigate the convergence of the iterates of the kernel in the Wasserstein distances. We will begin this chapter with a minimal introduction to the Wasserstein distance in Section 20.1. The main tool will be the duality theorem, Theorem 20.1.2, which requires the following assumption.

Throughout this chapter, unless otherwise indicated, (X, d) is a complete separable metric space endowed with its Borel σ-field, denoted by $\mathscr{X}$.

In Section 20.2, we will provide a criterion for the existence and uniqueness of an invariant distribution that can be applied to certain nonirreducible chains. In the following sections, we will prove rates of convergence in the Wasserstein distance. The geometric rates will be obtained in Sections 20.3 and 20.4 by methods very similar to those used in Chapter 18. Subgeometric rates of convergence will be obtained by a coupling method close to the one used in Section 19.5. In all these results, small sets and Doeblin sets, which irreducible chains do not possess, are replaced by sets in which the Markov kernel has appropriate contractivity properties with respect to the Wasserstein distance. The drift conditions used in this chapter are the same as those considered in Part III.

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_20

20.1 The Wasserstein Distance

Let $\mathsf{c} : \mathsf{X} \times \mathsf{X} \to \mathbb{R}_+$ be a symmetric measurable function such that $\mathsf{c}(x,y) = 0$ if and only if $x = y$. Such a function c is called distance-like. Recall that $\mathscr{C}(\xi, \xi')$ denotes the set of couplings of two probability measures ξ and ξ' and define the possibly infinite quantity $\mathbf{W}_{\mathsf{c}}(\xi, \xi')$ by

$$\mathbf{W}_{\mathsf{c}}(\xi, \xi') = \inf_{\gamma \in \mathscr{C}(\xi, \xi')} \int_{\mathsf{X} \times \mathsf{X}} \mathsf{c}(x,y) \gamma(\mathrm{d}x\mathrm{d}y) .$$

The quantity $\mathbf{W}_{\mathsf{c}}(\xi, \xi')$ wil be called the Wasserstein distance between ξ and ξ' associated to the cost function c. This is actually an abuse of terminology, since $\mathbf{W}_{\mathsf{c}}$ may not be a distance when the cost function c does not satisfy the triangular inequality. However, when $\mathsf{c} = \mathsf{d}$, we will see that $\mathbf{W}_{\mathsf{d}}$ is actually a distance on an appropriate subset of $\mathbb{M}_1(\mathscr{X})$. The main examples of general cost functions are the following:

- $\mathsf{c}(x,y) = \mathsf{d}^p(x,y)$ for $p \geq 1$.
- $\mathsf{c}(x,y) = \mathsf{d}(x,y)\{V(x) + V(y)\}$ where V is a measurable nonnegative function.

An important feature of the Wasserstein distance is that it is achieved by one particular coupling.

Theorem 20.1.1. *Let $\mathsf{c} : \mathsf{X} \times \mathsf{X} \to \mathbb{R}_+$ be a symmetric, nonnegative lower semi-continuous function. Then there exists a probability measure $\gamma \in \mathscr{C}(\xi, \xi')$ such that*

$$\mathbf{W}_{\mathsf{c}}(\xi, \xi') = \int_{\mathsf{X} \times \mathsf{X}} \mathsf{c}(x,y) \gamma(\mathrm{d}x\mathrm{d}y) . \tag{20.1.1}$$

A coupling $\gamma \in \mathscr{C}(\xi, \xi')$ which satisfies (20.1.1) is called optimal with respect to $\mathbf{W}_{\mathsf{c}}$.

Proof (of Theorem 20.1.1). For $n \geq 1$, define $a_n = \mathbf{W}_{\mathsf{c}}(\xi, \xi') + 1/n$. Then there exists $\gamma_n \in \mathscr{C}(\xi, \xi')$ such that

$$\mathbf{W}_{\mathsf{c}}(\xi, \xi') \leq \int_{\mathsf{X} \times \mathsf{X}} \mathsf{c}(x,y) \gamma_n(\mathrm{d}x\mathrm{d}y) \leq a_n .$$

Since (X, d) is a complete separable metric space, the probability measures ξ and ξ' are tight by Prokhorov's theorem C.2.2, i.e., for every $\varepsilon > 0$, there exist a compact set K such that $\xi(K) \geq 1 - \varepsilon/2$ and $\xi'(K) \geq 1 - \varepsilon/2$. Since $\gamma_n \in \mathscr{C}(\xi, \xi')$ for each $n \in \mathbb{N}$, this yields

$$\gamma_n((K \times K)^c) \leq \gamma_n((K^c \times \mathsf{X}) \cup (\mathsf{X} \times K^c)) \leq \xi(K^c) + \xi'(K^c) \leq \varepsilon .$$

This proves that the sequence $\{\gamma_n,\ n \in \mathbb{N}^*\}$ is tight hence relatively compact by Theorem C.2.2. Since $\mathscr{C}(\xi,\xi')$ is closed for the topology of weak convergence, there exist $\zeta \in \mathscr{C}(\xi,\xi')$ and a subsequence $\{\gamma_{n_k}\}$ which converges weakly to ζ. Since c is lower-semicontinuous and bounded from below (by 0), the Portmanteau Lemma yields

$$\int_{\mathsf{X}\times\mathsf{X}} \mathsf{c}(x,y)\, \zeta(\mathrm{d}x\mathrm{d}y) \le \liminf_{k\to\infty} \int_{\mathsf{X}\times\mathsf{X}} \mathsf{c}(x,y)\, \gamma_{n_k}(\mathrm{d}x\mathrm{d}y) \le \lim\inf_{k\to\infty} a_{n_k} = \mathbf{W}_{\mathsf{c}}\left(\xi,\xi'\right) .$$

Since the converse inequality holds by definition, this proves that the coupling ζ achieves the Wasserstein distance. □

Let $\xi,\xi' \in \mathbb{M}_1(\mathscr{X})$. By Theorem 19.1.6 and Proposition D.2.4, the total variation distance satisfies

$$\mathrm{d}_{\mathrm{TV}}(\xi,\xi') = \inf_{\gamma\in\mathscr{C}(\xi,\xi')} \int_{\mathsf{X}\times\mathsf{X}} \mathbb{1}_{\{x\neq x'\}}\gamma(\mathrm{d}x\mathrm{d}x') = \sup_{\substack{f\in\mathbb{F}_b(\mathsf{X})\\ \mathrm{osc}(f)\le 1}} |\xi(f)-\xi'(f)| . \qquad (20.1.2)$$

Since the total variation distance is the Wasserstein distance relatively to the Hamming distance $\mathbb{1}\{x\neq y\}$, a natural question is whether a duality formula similar to (20.1.2) continues to hold for $\mathbf{W}_{\mathsf{c}}$ for more general cost functions c. The answer is positive if the cost function c is lower semi-continuous. The following duality theorem will not be proved and we refer to Section 20.7 for references.

Theorem 20.1.2. *Let $\mathsf{c} : \mathsf{X}\times\mathsf{X} \to \mathbb{R}_+$ be asymmetric, nonnegative lower semi-continuous function. Then, for all probability measures on X, we have*

$$\mathbf{W}_{\mathsf{c}}\left(\xi,\xi'\right) = \sup\left\{\xi(f)+\xi'(g) \,:\, f,g\in \mathrm{C}_b(\mathsf{X})\,,\ f(x)+g(x')\le \mathsf{c}(x,x')\right\} . \qquad (20.1.3)$$

In the case $\mathsf{c} = \mathsf{d}$, the duality formula (20.1.3) can be expressed in terms of Lipschitz functions. Let $\mathrm{Lip}_{\mathsf{d}}(\mathsf{X})$ be the set of Lipschitz functions on X and for $f \in \mathrm{Lip}_{\mathsf{d}}(\mathsf{X})$,

$$|f|_{\mathrm{Lip}(\mathsf{d})} = \sup_{x\neq x'} \frac{f(x)-f(x')}{\mathsf{d}(x,x')} .$$

Then,

$$\mathbf{W}_{\mathsf{d}}\left(\xi,\xi'\right) = \sup\left\{\xi(f)-\xi'(f) \,:\, f \text{ bounded}\ ,\ |f|_{\mathrm{Lip}(\mathsf{d})}\le 1\right\} . \qquad (20.1.4)$$

To see that (20.1.4) follows from (20.1.3), consider $f,g \in \mathrm{C}_b(\mathsf{X})$ such that $f(x)+g(x') \le \mathsf{d}(x,x')$. We will show that there exists a bounded function $\varphi \in \mathrm{Lip}_{\mathsf{d}}(\mathsf{X})$ such that

$$f \le \varphi \;, \qquad g \le -\varphi \;. \tag{20.1.5}$$

Indeed, define successively

$$\tilde{f}(x) = \inf_{x' \in \mathsf{X}} \left\{ \mathsf{d}(x,x') - g(x') \right\} \;, \tag{20.1.6}$$

$$\tilde{g}(x') = \inf_{x \in \mathsf{X}} \left\{ \mathsf{d}(x,x') - \tilde{f}(x) \right\} \;, \tag{20.1.7}$$

$$\varphi(x) = \left[\tilde{f}(x) - \tilde{g}(x) \right] /2 \;. \tag{20.1.8}$$

Since $f(x) \le \mathsf{d}(x,x') - g(x')$ for all $x' \in \mathsf{X}$ by assumption, the definition of $\tilde{f}$ implies that $f \le \tilde{f} \le -g$. Since f and g are bounded, this implies that $\tilde{f}$ is bounded. By definition, $g(x') \le \mathsf{d}(x,x') - \tilde{f}(x)$ for all $x,x' \in \mathsf{X}$, thus $g \le \tilde{g} \le -\tilde{f}$. Thus $\tilde{g}$ is also bounded.

It follows from (20.1.7) that for all $x,x' \in \mathsf{X}$,

$$\tilde{f}(x) + \tilde{g}(x') \le \mathsf{d}(x,x') \;.$$

Choosing $x = x'$, we get $\tilde{f}(x) + \tilde{g}(x) \le 0$. By definition of φ, this implies $\tilde{f} \le \varphi$ and $\tilde{g} \le -\varphi$.

Altogether, we have proved that $f \le \tilde{f} \le \varphi$ and $g \le \tilde{g} \le -\varphi$ thus (20.1.5) holds. It remains to show that φ is a bounded function in $\mathrm{Lip}_{\mathsf{d}}(\mathsf{X})$. In view of the definition (20.1.8) of φ, it suffices to show that $\tilde{f}, \tilde{g}$ belong to $\mathrm{Lip}_{\mathsf{d}}(\mathsf{X})$. We will only prove $\tilde{f} \in \mathrm{Lip}_{\mathsf{d}}(\mathsf{X})$ since the same arguments for $\tilde{g}$ are simlilar. For all $x_0, x_1, x \in \mathsf{X}$, the triangular inequality yields

$$\mathsf{d}(x_0,x) - g(x) \le \mathsf{d}(x_0,x_1) + \mathsf{d}(x_1,x) - g(x) \;.$$

Taking the infimum with respect to $x \in \mathsf{X}$ on both sides of the inequality yields $\tilde{f}(x_0) \le \mathsf{d}(x_0,x_1) + \tilde{f}(x_1)$. Since x_0, x_1 are arbitrary, this proves that $\tilde{f} \in \mathrm{Lip}_{\mathsf{d}}(\mathsf{X})$.

For a general cost function c, $\mathrm{Lip}_{\mathsf{c}}(\mathsf{X})$ is the set of c-Lipschitz functions, i.e., functions f for which there exists a finite constant ϑ such that for all $x,x' \in \mathsf{X}$, $|f(x) - f(x')| \le \vartheta \mathsf{c}(x,x')$. The c-Lipschitz norm is then defined by

$$|f|_{\mathrm{Lip}(\mathsf{c})} = \sup_{\substack{x,x' \in \mathsf{X} \\ x \neq x'}} \frac{f(x) - f(x')}{\mathsf{c}(x,x')} \;.$$

Then the duality Theorem 20.1.2 yields

$$|\xi(f) - \xi'(f)| \le |f|_{\mathrm{Lip}(\mathsf{c})} \, \mathbf{W}_{\mathsf{c}}(\xi, \xi)' \;. \tag{20.1.9}$$

However, there is no characterization similar to (20.1.4) for general cost functions.

As shown in Theorem 19.1.12, there exists a kernel coupling of a Markov kernel P with itself which is optimal for the total variation distance, that is a kernel coupling K of (P,P) such that, for all $(x,x') \in \mathsf{X} \times \mathsf{X}$,

$$\mathrm{d_{TV}}(P(x,\cdot),P(x',\cdot)) = \int K(x,x';\mathrm{d}y\mathrm{d}y')\mathbb{1}\left\{y \neq y'\right\} .$$

We now investigate the existence of coupling kernel associated general cost functions. As for Theorem 20.1.2, the following result will not be proved and we again refer to Section 20.7 for references.

Theorem 20.1.3. *Let* $\mathsf{c} : \mathsf{X} \times \mathsf{X} \to \mathbb{R}_+$ *be a symmetric, nonnegative lower semi-continuous function. There exists a kernel coupling* K *of* (P,P) *such that for all* $(x,x') \in \mathsf{X} \times \mathsf{X}$*,*

$$\mathbf{W}_{\mathsf{c}}\left(P(x,\cdot),P(x',\cdot)\right) = \int_{\mathsf{X}\times\mathsf{X}} \mathsf{c}(y,y')K(x,x';\mathrm{d}y\mathrm{d}y') . \tag{20.1.10}$$

Consequently, the application $(x,x') \mapsto \mathbf{W}_{\mathsf{c}}\left(P(x,\cdot),P(x',\cdot)\right)$ *is measurable.*

A kernel coupling of (P,P) which satisfies (20.1.10) is said to be optimal with respect to the cost function c.

The existence of an optimal kernel coupling (satisfying (20.1.10)) yields the following corollary.

Corollary 20.1.4 *For all probability measures* ξ, ξ' *and* $\gamma \in \mathscr{C}(\xi,\xi')$*,*

$$\mathbf{W}_{\mathsf{c}}\left(\xi P, \xi' P\right) \leq \int_{\mathsf{X}\times\mathsf{X}} \mathbf{W}_{\mathsf{c}}\left(P(x,\cdot),P(x',\cdot)\right) \gamma(\mathrm{d}x\mathrm{d}x') . \tag{20.1.11}$$

Moreover, if K *is a kernel coupling of* (P,P)*, then for all* $n \in \mathbb{N}$*,*

$$\mathbf{W}_{\mathsf{c}}\left(\xi P^n, \xi' P^n\right) \leq \int_{\mathsf{X}\times\mathsf{X}} K^n \mathsf{c}(x,x')\gamma(\mathrm{d}x\mathrm{d}x') . \tag{20.1.12}$$

Proof. Let K be an optimal kernel coupling of (P,P) for $\mathbf{W}_{\mathsf{c}}$. Then, for $\gamma \in \mathscr{C}(\xi,\xi')$, γK is a coupling of ξP and $\xi' P$ and (20.1.11) follows from

$$\begin{aligned}
\mathbf{W}_{\mathsf{c}}\left(\xi P, \xi' P\right) &\leq \int_{\mathsf{X}\times\mathsf{X}} \mathsf{c}(u,v)\gamma K(\mathrm{d}u\mathrm{d}v) = \int_{\mathsf{X}\times\mathsf{X}} \gamma(\mathrm{d}x\mathrm{d}x') \int_{\mathsf{X}\times\mathsf{X}} \mathsf{c}(u,v)K(x,x';\mathrm{d}u\mathrm{d}v) \\
&= \int_{\mathsf{X}\times\mathsf{X}} \mathbf{W}_{\mathsf{c}}\left(P(x,\cdot),P(x',\cdot)\right) \gamma(\mathrm{d}x\mathrm{d}x') .
\end{aligned}$$

If K is a kernel coupling of (P,P), then for all $n \in \mathbb{N}$, K^n is a kernel coupling of (P^n,P^n) and γK^n is a coupling of $(\xi P^n, \xi' P^n)$; (20.1.12) follows. □

Throughout the rest of the chapter, the following assumption on the cost function c will be in force.

A 20.1.5 *The function* $\mathsf{c} : \mathsf{X} \times \mathsf{X} \to \mathbb{R}_+$ *is symmetric, lower semi-continuous and* $\mathsf{c}(x,y) = 0$ *if and only if* $x = y$. *Moreover, there exists an integer* $p \geq 1$ *such that* $\mathsf{d}^p \leq \mathsf{c}$.

If c is symmetric, lower semi-continuous and distance-like, the existence of an optimal coupling yields that $\mathbf{W}_{\mathsf{c}}$ is also distance-like i.e., $\mathbf{W}_{\mathsf{c}}(\xi, \xi') = 0$ implies $\xi = \xi'$.

Before going further, we briefly recall the essential definitions and properties of the Wasserstein distance associated to the particular cost functions $\mathsf{c} = \mathsf{d}^p$.

Definition 20.1.6 *For* $p \geq 1$ *and* $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$, *the Wasserstein distance of order* p *between* ξ *and* ξ' *denoted by* $\mathbf{W}_{\mathsf{d},p}(\xi,\xi')$, *is defined by*

$$\mathbf{W}^p_{\mathsf{d},p}(\xi,\xi') = \inf_{\gamma \in \mathscr{C}(\xi,\xi')} \int_{\mathsf{X}\times\mathsf{X}} \mathsf{d}^p(x,y)\gamma(\mathrm{d}x\mathrm{d}y) , \tag{20.1.13}$$

where $\mathscr{C}(\xi,\xi')$ *is the set of coupling of* ξ *and* ξ'. *For* $p = 1$, *we simply write* $\mathbf{W}_{\mathsf{d}}$.

The Wasserstein distance can be expressed in terms of random variables as:

$$\mathbf{W}_{\mathsf{d},p}(\xi,\xi') = \inf_{(X,X')\in\mathscr{C}(\xi,\xi')} \left\{\mathbb{E}\left[\mathsf{d}^p(X,X')\right]\right\}^{1/p} ,$$

where $(X,X') \in \mathscr{C}(\xi,\xi')$ means as in Section 19.1.1 that the distribution of the pair (X,X') is a coupling of ξ and ξ'. By Hölder's inequality, it obviously holds that if $p \leq q$, then for all $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$,

$$\mathbf{W}_{\mathsf{d},p}(\xi,\xi') \leq \mathbf{W}_{\mathsf{d},q}(\xi,\xi') . \tag{20.1.14}$$

If $\mathsf{d}(x,y) = \mathbb{1}\{x \neq y\}$, then Theorem 19.1.6 shows that $\mathbf{W}_{\mathsf{d}} = \mathrm{d}_{\mathrm{TV}}$. Similarly, if we choose the distance $\mathsf{d}(x,y) = \{V(x)+V(y)\}\mathbb{1}\{x \neq y\}$, Theorem 19.1.7 shows that the associated distance is the distance associated to the V-norm. Hence, the Wasserstein distance can be seen as an extension of the total variation distance to more general distances d.

It is easily seen that $\mathbf{W}_{\mathsf{d},p}(\delta_x,\delta_y) = \mathsf{d}(x,y)$ for all $x,y \in \mathsf{X}$ and for $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\mathbf{W}^p_{\mathsf{d},p}(\delta_x,\xi) = \int_{\mathsf{X}} \mathsf{d}^p(x,y)\xi(\mathrm{d}y) \in [0,\infty] . \tag{20.1.15}$$

Thus, the distance $\mathbf{W}_{\mathsf{d},p}(\xi,\xi')$ can be infinite.

Definition 20.1.7 (Wasserstein space) *The Wasserstein space of order p is defined by*

$$\mathbb{S}_p(\mathsf{X},\mathsf{d}) = \left\{ \xi \in \mathbb{M}_1(\mathscr{X}) : \int_{\mathsf{X}} \mathsf{d}^p(x,y)\, \xi(\mathrm{d}y) < \infty \text{ for all } x \in \mathsf{X} \right\} . \tag{20.1.16}$$

For $p = 1$, we simply write $\mathbb{S}(\mathsf{X},\mathsf{d})$.

Of course, if d is bounded then $\mathbb{S}_{\mathsf{d},p}(\mathsf{X},\mathsf{d}) = \mathbb{M}_1(\mathscr{X})$. If d is not bounded, then the distance $\tilde{\mathsf{d}} = \mathsf{d} \wedge m$ defines the same topology as d on X and $(\mathsf{X},\tilde{\mathsf{d}})$ is still complete and separable. Applying Minkowski's inequality, we have

$$\left\{ \int_{\mathsf{X}} \mathsf{d}^p(x,y)\, \xi(\mathrm{d}y) \right\}^{1/p} \le \mathsf{d}(x_0,x) + \left\{ \int_{\mathsf{X}} \mathsf{d}^p(x_0,y)\, \xi(\mathrm{d}y) \right\}^{1/p} < \infty .$$

Therefore $\int_{\mathsf{X}} \mathsf{d}^p(x,y)\, \xi(\mathrm{d}y)$ is finite for one $x \in \mathsf{X}$ if and only if it is finite for all $x \in \mathsf{X}$. If $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$ and $\gamma \in \mathscr{C}(\xi,\xi')$, then for all $x_0 \in \mathsf{X}$,

$$\left\{ \int_{\mathsf{X}\times\mathsf{X}} \mathsf{d}^p(x,y) \gamma(\mathrm{d}x\mathrm{d}y) \right\}^{1/p} \le \left\{ \int_{\mathsf{X}} \mathsf{d}^p(x_0,x) \xi(\mathrm{d}x) \right\}^{1/p} + \left\{ \int_{\mathsf{X}} \mathsf{d}^p(x_0,y) \xi'(\mathrm{d}y) \right\}^{1/p} .$$

This implies that for all $\xi, \xi' \in \mathbb{S}_p(\mathsf{X},\mathsf{d})$,

$$\mathbf{W}_{\mathsf{d},p}(\xi,\xi') \le \left\{ \int_{\mathsf{X}} \mathsf{d}^p(x_0,x) \xi(\mathrm{d}x) \right\}^{1/p} + \left\{ \int_{\mathsf{X}} \mathsf{d}^p(x_0,y) \xi'(\mathrm{d}y) \right\}^{1/p} < \infty .$$

The Wasserstein space and distance have the following properties.

Theorem 20.1.8. $(\mathbb{S}_p(\mathsf{X},\mathsf{d}), \mathbf{W}_{\mathsf{d},p})$ *is a complete separable metric space and the distributions with finite support are dense in* $\mathbb{S}_p(\mathsf{X},\mathsf{d})$. *If* $\{\mu_n,\ n \in \mathbb{N}\}$ *is a sequence of probability measures in* $\mathbb{S}_p(\mathsf{X},\mathsf{d})$, *the following statements are equivalent:*

(i) $\lim_{n\to\infty} \mathbf{W}_{\mathsf{d},p}(\mu_n,\mu_0) = 0$;

(ii) $\mu_n \overset{w}{\Rightarrow} \mu_0$ *and* $\lim_{M\to\infty} \limsup_{n\to\infty} \int_{\mathsf{X}} \mathsf{d}^p(x_0,x) \mathbb{1}\{\mathsf{d}(x_0,x) > M\}\, \mu_n(\mathrm{d}x) = 0$.

See Section 20.A for a proof. If $\mathsf{d}_1, \mathsf{d}_2$ are two distances on X such that $\mathsf{d}_1 \le \mathsf{d}_2$, then it follows from the definition that $\mathbf{W}_{\mathsf{d}_1,p} \le \mathbf{W}_{\mathsf{d}_2,p}$. In particular, if $\mathsf{d} \le 1$, then $\mathbf{W}_{\mathsf{d}} \le \mathrm{d}_{\mathrm{TV}}$. An important consequence is that the topology induced by the Wasserstein distance is coarser than the topology of total variation when the distance d is bounded. This means that more sequences will converge in the Wasserstein distance than in total variation. This suits our purpose to study non irreducible Markov kernels whose iterates may not converge in total variation to the invariant probability. When the distance d is not bounded, neither convergence implies the other (see Exercise 20.1).

20.2 Existence and Uniqueness of the Invariant Probability Measure

In this section, we will provide a sufficient condition for the existence and uniqueness of an invariant probability measure. We have already obtained such results under the assumption that the kernel is irreducible. In the next results, we do not assume irreducibility.

Theorem 20.2.1. *Assume that* **A** 20.1.5 *and the following conditions hold.*

(i) There exist a kernel coupling K of (P,P), a set $\bar{C} \in \mathscr{X} \otimes \mathscr{X}$, a measurable function $\bar{V} : \mathsf{X} \times \mathsf{X} \to [0,\infty)$ and constants $(\varepsilon, \bar{b}) \in (0,1) \times (0,\infty)$ such that $\mathsf{c} \le \bar{V}$ and

$$K\mathsf{c} \le (1 - \varepsilon \mathbb{1}_{\bar{C}})\mathsf{c} \,, \qquad K\bar{V} + 1 \le \bar{V} + \bar{b}\mathbb{1}_{\bar{C}} \,. \tag{20.2.1}$$

(ii) There exist $x_0 \in \mathsf{X}$, a nondecreasing concave function $\psi : [0,\infty) \to [0,\infty)$ such that $\lim_{v\to\infty} \psi(v) = \infty$, a subsequence $\{n_k,\ k \in \mathbb{N}\}$ such that $\lim_{k\to\infty} n_k = \infty$ and

$$\sup_{k \in \mathbb{N}} P^{n_k}(\psi \circ V_{x_0})(x_0) < \infty \,, \qquad PV_{x_0}(x_0) < \infty \,, \tag{20.2.2}$$

where $V_{x_0}(x) = \bar{V}(x_0, x)$.

Then P admits a unique invariant probability measure π and for all $\xi \in \mathbb{M}_1(\mathscr{X})$,

$$\lim_{n\to\infty} \mathbf{W}_{\mathsf{c}\wedge 1}(\xi P^n, \pi) = 0 \,. \tag{20.2.3}$$

Proof. Replacing if needed $\bar{V}$ by $\bar{V} + 1$ in (20.2.1), we assume that $\bar{V} \ge 1$. For $n \in \mathbb{N}$, set $S_n = \bar{V} + n$. Then (20.2.1) implies

$$KS_{n+1} \le S_n + \bar{b}\mathbb{1}_{\bar{C}} \le (1 + \bar{b}\mathbb{1}_{\bar{C}})S_n \,.$$

Pick $\alpha \in (0,1)$ such that $(1 - \varepsilon \mathbb{1}_{\bar{C}})^{1-\alpha}(1 + \bar{b}\mathbb{1}_{\bar{C}})^{\alpha} \le 1$. Hölder's inequality yields for all $n \ge 0$,

$$K(\mathsf{c}^{1-\alpha} S_{n+1}^{\alpha}) \le (K\mathsf{c})^{1-\alpha}(KS_{n+1})^{\alpha} \le \mathsf{c}^{1-\alpha} S_n^{\alpha} \,.$$

Applying the previous inequality repeatedly yields

$$n^{\alpha} K^n \mathsf{c}^{1-\alpha} \le K^n(\mathsf{c}^{1-\alpha} S_n^{\alpha}) \le \mathsf{c}^{1-\alpha} S_0^{\alpha} = \mathsf{c}^{1-\alpha} \bar{V}^{\alpha} \,. \tag{20.2.4}$$

This implies that

$$\lim_{n\to\infty}[K^n(\mathsf{c}\wedge 1)](x,x') \le \lim_{n\to\infty} K^n \mathsf{c}^{1-\alpha}(x,x') = 0 ,$$

for all $x,x' \in \mathsf{X}$. Since moreover $K^n(\mathsf{c}\wedge 1) \le 1$, we obtain by Lebesgue's dominated convergence theorem that for all $\xi,\xi' \in \mathbb{M}_1(\mathscr{X})$ and $\gamma \in \mathscr{C}(\xi,\xi')$,

$$\lim_{n\to\infty}\int_{\mathsf{X}\times\mathsf{X}}[K^n(\mathsf{c}\wedge 1)](x,x')\gamma(\mathrm{d}x\mathrm{d}x') = 0 .$$

Combining this limit with Corollary 20.1.4 yields

$$\lim_{n\to\infty} \mathbf{W}_{\mathsf{c}\wedge 1}\left(\xi P^n, \xi' P^n\right) = 0 . \tag{20.2.5}$$

If π and π' are two invariant probability measures, (20.2.5) yields $\mathbf{W}_{\mathsf{c}\wedge 1}(\pi,\pi') = 0$ which implies that $\pi = \pi'$ since $\mathsf{c}\wedge 1$ is lower semicontinuous and distance-like.

We now prove that P admits at least one invariant probability measure. To this end, we will find a subsequence of $\{P^n(x_0,\cdot),\ n\in\mathbb{N}\}$ which converges weakly to a probability measure π and we will show that $\pi = \pi P$. For $M>0$, set $A_M = \{V_{x_0} \le M\}$ and let $\gamma \in \mathscr{C}(\delta_{x_0}, \delta_{x_0}P^{n_k})$ where the sequence $\{n_k\}$ is defined in (ii). Applying successively Corollary 20.1.4, the bound (20.2.4) combined with $\mathsf{c} \le \bar{V}$ and $K^n(\mathsf{c}\wedge 1) \le 1$ yields

$$\begin{aligned}
&\mathbf{W}_{\mathsf{c}\wedge 1}\left(P^n(x_0,\cdot), P^{n+n_k}(x_0,\cdot)\right) \\
&\quad\le \int \left(\mathbb{1}_{\mathsf{X}\times A_M}(x,y) + \mathbb{1}_{\mathsf{X}\times A_M^c}(x,y)\right) K^n(\mathsf{c}\wedge 1)(x,y)\gamma(\mathrm{d}x\mathrm{d}y) \\
&\quad\le n^{-\alpha}\gamma(\bar{V}\mathbb{1}_{\mathsf{X}\times A_M}) + \gamma(\mathsf{X}\times A_M^c) \\
&\quad= n^{-\alpha}P^{n_k}(V_{x_0}\mathbb{1}_{A_M})(x_0) + P^{n_k}(\mathbb{1}_{A_M^c})(x_0) .
\end{aligned} \tag{20.2.6}$$

Replacing ψ by $\psi - \psi(0)$ if necessary, we may assume that $\psi(0) = 0$. In this case, the function ψ being concave, $u \mapsto \psi(u)/u$ is nonincreasing, or equivalently, $u \mapsto u/\psi(u)$ is nondecreasing. This implies $V_{x_0} \le M\psi\circ V_{x_0}/\psi(M)$ on $A_M = \{V_{x_0} \le M\}$. In addition, $A_M^c \subset \{\psi\circ V_{x_0} > \psi(M)\}$. Therefore, writing $M_\psi = \sup_{k\in\mathbb{N}} P^{n_k}(\psi\circ V_{x_0})(x_0)$, which is finite by assumption (20.2.2), we obtain

$$P^{n_k}(V_{x_0}\mathbb{1}_{A_M})(x_0) \le \frac{MM_\psi}{\psi(M)} , \qquad P^{n_k}\mathbb{1}_{A_M^c}(x_0) \le \frac{P^{n_k}(\psi\circ V_{x_0})(x_0)}{\psi(M)} \le \frac{M_\psi}{\psi(M)} .$$

Taking now $M = n^\alpha$ and plugging these inequalities into (20.2.6) yield: for all $n,k\in\mathbb{N}$,

$$\mathbf{W}_{\mathsf{c}\wedge 1}\left(P^n(x_0,\cdot), P^{n+n_k}(x_0,\cdot)\right) \le 2M_\psi/\psi(n^\alpha) .$$

Set $u_0 = 1$, for $k \ge 1$, $u_k = \inf\left\{n_\ell :\ \psi(n_\ell^\alpha) > 2^k\right\}$ and $m_k = \sum_{i=0}^k u_i$. Then,

$$\mathbf{W}_{\mathsf{c}\wedge 1}\left(P^{m_k}(x_0,\cdot),P^{m_{k+1}}(x_0,\cdot)\right) \le 2M_\psi/\psi(m_k^\alpha) \le 2M_\psi/\psi(u_k^\alpha) \le 2^{-k+1}M_\psi \;.$$

Unfortunately, $\mathbf{W}_{\mathsf{c}\wedge 1}$ is not a metric, but since $(\mathsf{d}\wedge 1)^p \le \mathsf{c}\wedge 1$, the previous inequality shows that $\{P^{m_k}(x_0,\cdot),\ k\in\mathbb{N}\}$ is a Cauchy sequence of probability measures in the complete metric space $(\mathbb{M}_1(\mathscr{X}),\mathbf{W}_{\mathsf{d}\wedge 1,p})$; see Theorem 20.1.8. Therefore, there exists a probability measure π such that $\lim_{k\to\infty}\mathbf{W}_{\mathsf{d}\wedge 1,p}\left(P^{m_k}(x_0,\cdot),\pi\right)=0$. It remains to show that $\pi=\pi P$. First note that Corollary 20.1.4, Jensen's inequality and (20.2.1) imply for all $\xi,\xi'\in\mathbb{M}_1(\mathscr{X})$,

$$\begin{aligned}
\mathbf{W}_{\mathsf{c}\wedge 1}\left(\xi P,\xi' P\right) &\le \inf_{\gamma\in\mathscr{C}(\xi,\xi')}\int_{\mathsf{X}\times\mathsf{X}} K(\mathsf{c}\wedge 1)(x,x')\gamma(\mathrm{d}x\mathrm{d}x') \\
&\le \inf_{\gamma\in\mathscr{C}(\xi,\xi')}\int_{\mathsf{X}\times\mathsf{X}} [K\mathsf{c}(x,x')\wedge 1]\gamma(\mathrm{d}x\mathrm{d}x') \\
&\le \inf_{\gamma\in\mathscr{C}(\xi,\xi')}\int_{\mathsf{X}\times\mathsf{X}} [\mathsf{c}(x,x')\wedge 1]\gamma(\mathrm{d}x\mathrm{d}x') = \mathbf{W}_{\mathsf{c}\wedge 1}\left(\xi,\xi'\right) \;.
\end{aligned}$$

Combining this inequality with $\mathbf{W}_{\mathsf{d}\wedge 1,p}(\xi,\xi') \le [\mathbf{W}_{\mathsf{c}\wedge 1}(\xi,\xi')]^{1/p}$ and the triangular inequality for the metric $\mathbf{W}_{\mathsf{d}\wedge 1,p}$ yields

$$\begin{aligned}
&\mathbf{W}_{\mathsf{d}\wedge 1,p}(\pi,\pi P) \\
&\quad\le \mathbf{W}_{\mathsf{d}\wedge 1,p}(\pi,P^{m_k}(x_0,\cdot)) + \mathbf{W}_{\mathsf{d}\wedge 1,p}\left(P^{m_k+1}(x_0,\cdot),\pi P\right) \\
&\qquad\qquad + \mathbf{W}_{\mathsf{d}\wedge 1,p}\left(P^{m_k}(x_0,\cdot),P^{m_k+1}(x_0,\cdot)\right) \\
&\quad\le 2\left[\mathbf{W}_{\mathsf{c}\wedge 1}(\pi,P^{m_k}(x_0,\cdot))\right]^{1/p} + \left[\mathbf{W}_{\mathsf{c}\wedge 1}\left(P^{m_k}(x_0,\cdot),P^{m_k+1}(x_0,\cdot)\right)\right]^{1/p} \;.
\end{aligned}$$

We have seen that the first term of the right-hand side converges to 0. The second term also converges to 0 by applying (20.2.5) to $\xi=\delta_{x_0}$ and $\xi'=\delta_{x_0}P$. Finally, $\mathbf{W}_{\mathsf{d}\wedge 1,p}(\pi,\pi P)=0$ which implies $\pi=\pi P$. The proof of (20.2.3) is then completed by applying (20.2.5) with $\xi'=\pi$. □

Remark 20.2.2. Let us give a sufficient condition for the condition (ii) of Theorem 20.2.1. Assume there exist a measurable function $W:\mathsf{X}\to[0,\infty)$ and a constant b'such that

$$PW + \psi\circ V_{x_0} \le W + b' \;. \tag{20.2.7}$$

Then, by Theorem 4.3.1,

$$\sum_{k=0}^{n-1} P^k\psi\circ V_{x_0}(x_0) \le W(x_0) + nb'$$

which yields after dividing by n,

$$\sup_{n\ge 1} n^{-1}\sum_{k=0}^{n-1} P^k\psi\circ V_{x_0}(x_0) \le W(x_0) + b' \;.$$

Therefore, there exists an infinite number of n_k such that $P^{n_k}\psi\circ V_{x_0}(x_0)\le W(x_0)+b'+1$ and (20.2.2) holds. In particular, if the function $\bar{V}$ that appears in (20.2.1) is of the form $\bar{V}(x,x')=V(x)+V(x')-1$ and if V satisfies the subgeometric drift condition $PV+\psi\circ V\le V+\tilde{b}\mathbb{1}_{\tilde{C}}$, then (20.2.7) holds with $W=V$, $b'=\psi\circ V(x_0)-\psi(1)+\tilde{b}$. Indeed, by concavity of ψ, we have $\psi(a+b-1)-\psi(a)\le\psi(b)-\psi(1)$ for $a\ge 1$ and $b\ge 0$. Then,

$$\begin{aligned}PW(x)+\psi\circ V_{x_0}(x)&=PV(x)+\psi(V(x)+V(x_0)-1)\\&\le PV(x)+\psi\circ V(x)+\psi\circ V(x_0)-\psi(1)\\&\le V(x)+\tilde{b}\mathbb{1}_{\tilde{C}}(x)+\psi\circ V(x_0)-\psi(1)\le W(x)+b'\;.\end{aligned}$$

20.3 Uniform Convergence in the Wasserstein Distance

Theorem 20.2.1 provides sufficient conditions for the convergence of ξP^n to the invariant probability measure π with respect to $\mathbf{W}_{\mathsf{c}\wedge 1}$, but it does not give information on the rate of convergence. We now turn to conditions that imply either geometric or subgeometric decreasing bounds. We begin by introducing the Dobrushin coefficient associated with a cost function c. This is a generalization of the V-Dobrushin coefficient seen in Definition 18.3.2.

Definition 20.3.1 (c-Dobrushin Coefficient) *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$. The c-Dobrushin coefficient $\Delta_{\mathsf{c}}(P)$ of P is defined by*

$$\Delta_{\mathsf{c}}(P)=\sup\left\{\frac{\mathbf{W}_{\mathsf{c}}(\xi P,\xi' P)}{\mathbf{W}_{\mathsf{c}}(\xi,\xi')}\;:\;\xi,\xi'\in\mathbb{M}_1(\mathscr{X})\,,\;\mathbf{W}_{\mathsf{c}}\left(\xi,\xi'\right)<\infty\,,\;\xi\neq\xi'\right\}\,.$$

For $p\ge 1$, we write $\Delta_{\mathsf{d},p}(P)=\left[\Delta_{\mathsf{d}^p}(P)\right]^{1/p}$. If $\Delta_{\mathsf{c}}(P)<1$, the Markov kernel P is said to be $\mathbf{W}_{\mathsf{c}}$-uniformly ergodic.

In contrast to the Dobrushin coefficient relative to the total variation distance introduced in Definition 18.2.1, which always satisfies $\Delta(P)\le 1$, the c-Dobrushin coefficient is not necessarily finite. From the definition, for all $\xi,\xi'\in\mathbb{M}_1(\mathscr{X})$, $\mathbf{W}_{\mathsf{c}}(\xi P,\xi' P)\le\Delta_{\mathsf{c}}(P)\,\mathbf{W}_{\mathsf{c}}(\xi,\xi')$ holds even if $\mathbf{W}_{\mathsf{c}}(\xi,\xi')=\infty$ (using the convention $0\times\infty=0$).

The measurability of the optimal kernel coupling (see Theorem 20.1.3) yields the following expression for the c-Dobrushin coefficient. This result parallels Lemmas 18.2.2 and 18.3.3.

Lemma 20.3.2 *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$. Then*

$$\Delta_{\mathsf{c}}(P) = \sup_{x \neq x'} \frac{\mathbf{W}_{\mathsf{c}}(P(x,\cdot), P(x',\cdot))}{\mathsf{c}(x,x')} . \tag{20.3.1}$$

Proof. Let the right-hand side of (20.3.1) be denoted by $\tilde{\Delta}_{\mathsf{c}}(P)$. If $\tilde{\Delta}_{\mathsf{c}}(P) = 0$, then $P(x,\cdot) = P(x',\cdot)$ for all $x \neq x'$, which clearly implies $\Delta_{\mathsf{c}}(P) = 0 = \tilde{\Delta}_{\mathsf{c}}(P)$. We now assume $\tilde{\Delta}_{\mathsf{c}}(P) > 0$. Since $\mathbf{W}_{\mathsf{c}}(\delta_x, \delta_{x'}) = \mathsf{c}(x,x')$ for all $x,x' \in \mathsf{X}$, we have $\tilde{\Delta}_{\mathsf{c}}(P) \leq \Delta_{\mathsf{c}}(P)$. To prove the converse inequality, let $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$ and take an arbitrary $\gamma \in \mathscr{C}(\xi,\xi')$. Applying Corollary 20.1.4, we obtain

$$\begin{aligned}
\mathbf{W}_{\mathsf{c}}(\xi P, \xi' P) &\leq \int_{\mathsf{X}\times\mathsf{X}} \gamma(\mathrm{d}x\mathrm{d}x') \mathbf{W}_{\mathsf{c}}\left(P(x,\cdot), P(x',\cdot)\right) \\
&\leq \tilde{\Delta}_{\mathsf{c}}(P) \int_{\mathsf{X}\times\mathsf{X}} \mathsf{c}(x,x')\gamma(\mathrm{d}x\mathrm{d}x') .
\end{aligned}$$

This yields $\mathbf{W}_{\mathsf{c}}(\xi P, \xi' P) \leq \tilde{\Delta}_{\mathsf{c}}(P) \mathbf{W}_{\mathsf{c}}(\xi, \xi')$. Since ξ and ξ' are arbitrary, this in turn implies that $\Delta_{\mathsf{c}}(P) \leq \tilde{\Delta}_{\mathsf{c}}(P)$. $\square$

If $\Delta_{\mathsf{c}}(P) < \infty$ and $f \in \mathrm{Lip}_{\mathsf{c}}(\mathsf{X})$, then by (20.1.9), $Pf \in \mathrm{Lip}_{\mathsf{c}}(\mathsf{X})$ and

$$|Pf|_{\mathrm{Lip}(\mathsf{c})} \leq \Delta_{\mathsf{c}}(P) \, |f|_{\mathrm{Lip}(\mathsf{c})} . \tag{20.3.2}$$

This implies

$$\sup\left\{ |Pf|_{\mathrm{Lip}(\mathsf{c})} : f \in \mathrm{Lip}_{\mathsf{c}}(\mathsf{X}), \; |f|_{\mathrm{Lip}(\mathsf{c})} \leq 1 \right\} \leq \Delta_{\mathsf{c}}(P) .$$

Equality holds in the above expression if c is a distance by the duality (20.1.4), but such is not the case for a general cost function c.

Proposition 20.3.3 *Let P and Q be two Markov kernels on $(\mathsf{X}, \mathscr{X})$. Then* $\Delta_{\mathsf{c}}(PQ) \leq \Delta_{\mathsf{c}}(P)\Delta_{\mathsf{c}}(Q)$.

Proof. For $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$, we have, by definition,

$$\mathbf{W}_{\mathsf{c}}(\xi PQ, \xi' PQ) \leq \Delta_{\mathsf{c}}(Q) \mathbf{W}_{\mathsf{c}}(\xi P, \xi' P) \leq \Delta_{\mathsf{c}}(Q)\Delta_{\mathsf{c}}(P) \mathbf{W}_{\mathsf{c}}(\xi, \xi') .$$

$\square$

If the Dobrushin coefficient of an iterate of the kernel P is strictly contracting, then we can adapt the fixed-point theorem, Theorem 18.1.1, to obtain the existence and uniqueness of the invariant probability measure and a uniform geometric rate of convergence.

Theorem 20.3.4. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, c a cost function satisfying* **A** 20.1.5*, and $x_0 \in \mathsf{X}$ such that $\int \mathsf{c}(x_0,x)P(x_0,\mathrm{d}x) < \infty$. Assume that there exist an* *integer $m \geq 1$ and a constant $\varepsilon \in [0,1)$ such that $\Delta_{\mathsf{c}}(P^m) \leq 1-\varepsilon$. Assume in addition that $\Delta_{\mathsf{c}}(P) \vee \Delta_{\mathsf{d},p}(P) < \infty$ for all $p \geq 0$ such that $d^p \leq c$.*

Then P admits a unique invariant probability measure π. Moreover, $\pi \in \mathbb{S}_p(\mathsf{X},\mathsf{d})$, and for all $\xi \in \mathbb{M}_1(\mathsf{X})$ and $n \in \mathbb{N}$,

$$\mathbf{W}_{\mathsf{d}^p}(\xi P^n, \pi) \leq \mathbf{W}_{\mathsf{c}}(\xi P^n, \pi) \leq \kappa(1-\varepsilon)^{\lfloor n/m \rfloor} \mathbf{W}_{\mathsf{c}}(\xi, \pi) , \tag{20.3.3}$$

with $\kappa = 1 \vee \sup_{1 \leq r < m} \Delta_{\mathsf{c}}(P^r)$.

Proof. The proof is adapted from Theorem 18.1.1, which cannot be applied directly, since c is not necessarily a distance. For $n \in \mathbb{N}$, write $n = m\lfloor n/m \rfloor + r$, where $r \in \{0,\dots,m-1\}$. Using $\mathsf{d}^p \leq \mathsf{c}$ and the submultiplicativity property of Proposition 20.3.3 yields

$$\begin{aligned}\mathbf{W}_{\mathsf{d}^p}\left(\xi P^n, \xi' P^n\right) &\leq \mathbf{W}_{\mathsf{c}}\left(\xi P^n, \xi' P^n\right) \leq \Delta_{\mathsf{c}}(P^n)\, \mathbf{W}_{\mathsf{c}}\left(\xi, \xi'\right) \\ &\leq \kappa [\Delta_{\mathsf{c}}(P^m)]^{\lfloor n/m \rfloor} \mathbf{W}_{\mathsf{c}}\left(\xi, \xi'\right) \leq \kappa(1-\varepsilon)^{\lfloor n/m \rfloor} \mathbf{W}_{\mathsf{c}}\left(\xi, \xi'\right) .\end{aligned} \tag{20.3.4}$$

Since $\mathbf{W}_{\mathsf{c}}\left(\delta_{x_0}, \delta_{x_0} P\right) = \int \mathsf{c}(x_0,x) P(x_0, \mathrm{d}x) < \infty$ by assumption, applying (20.3.4) with $(\xi, \xi') = (\delta_{x_0}, \delta_{x_0} P)$, we obtain

$$\mathbf{W}_{\mathsf{d}^p}\left(\delta_{x_0} P^n, \delta_{x_0} P^{n+1}\right) \leq \kappa(1-\varepsilon)^{\lfloor n/m \rfloor} \int \mathsf{c}(x_0,x) P(x_0, \mathrm{d}x) < \infty . \tag{20.3.5}$$

Consequently, $\{P^n(x_0,\cdot),\ n \in \mathbb{N}\}$ is a Cauchy sequence of probability measures in the complete metric space $(\mathbb{S}_p(\mathsf{X},\mathsf{d}), \mathbf{W}_{\mathsf{d},p})$. Therefore, there exists a probability measure $\pi \in \mathbb{S}_p(\mathsf{X},\mathsf{d})$ such that $\lim_{n\to\infty} \mathbf{W}_{\mathsf{d},p}\left(\delta_{x_0} P^n, \pi\right) = 0$. Now, for all $n \geq 1$,

$$\begin{aligned}\mathbf{W}_{\mathsf{d},p}(\pi P, \pi) &\leq \mathbf{W}_{\mathsf{d},p}\left(\pi P, \delta_{x_0} P^{n+1}\right) + \mathbf{W}_{\mathsf{d},p}\left(\delta_{x_0} P^{n+1}, \delta_{x_0} P^n\right) + \mathbf{W}_{\mathsf{d},p}\left(\delta_{x_0} P^n, \pi\right) \\ &\leq \left(\Delta_{\mathsf{d},p}(P) + 1\right) \mathbf{W}_{\mathsf{d},p}\left(\pi, \delta_{x_0} P^n\right) + \mathbf{W}_{\mathsf{d},p}\left(\delta_{x_0} P^{n+1}, \delta_{x_0} P^n\right) \to 0\end{aligned}$$

as $n \to 0$. This proves that $\pi = \pi P$. Applying (20.3.4) with $\xi' = \pi$ yields (20.3.3). To complete the proof, it remains to show the uniqueness of the invariant probability measure. For all $x \in \mathsf{X}$, taking $\xi = \delta_{x_0}$ and $\xi' = \delta_x$ in (20.3.4) and combining it with $\lim_{n\to\infty} \mathbf{W}_{\mathsf{d},p}\left(\delta_{x_0} P^n, \pi\right) = 0$ yields

$$\lim_{n\to\infty} \mathbf{W}_{\mathsf{d},p}(\delta_x P^n, \pi) = 0 , \qquad \text{for all } x \in \mathsf{X} .$$

This in turn implies that $\delta_x P^n \stackrel{\text{w}}{\Rightarrow} \pi$ for all $x \in \mathsf{X}$, and consequently, for all bounded continuous functions f and all $x \in \mathsf{X}$, $\lim_{n\to\infty} P^n f(x) = \pi(f)$. By the dominated convergence theorem, we have for all bounded continuous functions f that

$$\pi'(f) = \pi' P^n(f) = \int \pi'(\mathrm{d}x) P^n f(x) \to \pi(f) ,$$

as $n \to \infty$. Thus $\pi = \pi'$, and the proof is complete. □

In view of Theorem 20.1.3 and Lemma 20.3.2, the way to prove that $\Delta_{\mathsf{c}}(P) < 1$ is to construct, for all $x, x' \in \mathsf{X}$, a pair of random variables $(X_1^x, X_1^{x'})$ on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ whose joint distribution is a coupling of $P(x, \cdot)$ and $P(x', \cdot)$. If there exists $\varepsilon \in (0,1)$ such that $\mathbb{E}\left[\mathsf{c}(X_1^x, X_1^{x'})\right] \le (1-\varepsilon)\mathsf{c}(x,x')$ for all $x, x' \in \mathsf{X}$, then $\Delta_{\mathsf{c}}(P) \le 1-\varepsilon$. The following examples provide two different types of coupling. We will apply Theorem 20.3.4 either to $\mathsf{c} = \mathsf{d}$ or to $\mathsf{c} = \mathsf{d}^p$.

Example 20.3.5. Let $\{Z_n,\ n \in \mathbb{N}^*\}$ be a sequence of i.i.d. Bernoulli random variables with mean $1/2$, independent of the random variable X_0 with values in $[0,1]$, and define the Markov chain $\{X_n,\ n \in \mathbb{N}\}$ on $[0,1]$ by

$$X_{n+1} = \frac{1}{2}(X_n + Z_{n+1})\ , n \ge 0\ .$$

Let P be the Markov kernel of the chain $\{X_n\}$. For $x, y \in [0,1]$ such that $x - y$ is not rational number, $P^n(x,\cdot)$ and $P^n(y,\cdot)$ are singular for every $n \ge 0$, whence $\mathsf{d}_{\mathrm{TV}}(P^n(x,\cdot), P^n(y,\cdot)) = 1$, and thus $\Delta(P) = 1$. For $x, y \in [0,1]$,

$$\mathbf{W}_{\mathsf{d},p}(P(x,\cdot), P(y,\cdot)) \le \{\mathbb{E}[|(x+Z_1)/2 - (y+Z_1)/2|^p]\}^{1/p} = \frac{1}{2}|x-y|\ .$$

This proves that $\Delta_{\mathsf{d},p}(P) \le 1/2$. Since $\mathbf{W}_{\mathsf{d}}(P(x,\cdot), P(y,\cdot)) \le \mathbf{W}_{\mathsf{d},p}(P(x,\cdot), P(y,\cdot))$ for all $p \ge 1$, it suffices to prove that $\mathbf{W}_{\mathsf{d}}(P(x,\cdot), P(y,\cdot)) \ge \frac{1}{2}|x-y|$. By the duality theorem, Theorem 20.1.2, a lower bound is given by $Pf(x) - Pf(y)$ with $f(x) = x$, that is,

$$\mathbf{W}_{\mathsf{d}}(P(x,\cdot), P(y,\cdot)) \ge Pf(x) - Pf(y) = \frac{x-y}{2}\ .$$

Altogether, we have proved that $\mathbf{W}_{\mathsf{d},p}(P(x,\cdot), P(y,\cdot)) = \mathbf{W}_{\mathsf{d}}(P(x,\cdot), P(y,\cdot)) = \frac{1}{2}|x-y|$. Thus $\Delta_{\mathsf{d},p}(P) = 1/2$ for all $p \ge 1$. ◀

The coupling method used in the previous example consists in using the same sequence $\{Z_n\}$ for the two chains starting from different points. This simple idea may not always be successful, as illustrated in Exercise 20.7.

We conclude this section by applying this result to the random iterative functions introduced in Section 2.1 and defined on X by the recurrence

$$X_k = f(X_{k-1}, Z_k)\ , \quad k \ge 1\ , \tag{20.3.6}$$

where $f : \mathsf{X} \times \mathsf{Z} \to \mathsf{X}$ is a measurable function, $\{Z_k,\ k \in \mathbb{N}\}$ is an i.i.d. sequence of random elements defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ taking values in $(\mathsf{Z}, \mathscr{Z})$, independent of the initial condition X_0. Hereinafter, we define

$$f(x,z) = f_z(x)\ , \quad \text{for all } (x,z) \in \mathsf{X} \times \mathsf{Z}\ .$$

It is assumed that the map $(z,x) \mapsto f_z(x)$ is measurable with respect to the product sigma-field on $\mathscr{Z} \otimes \mathscr{X}$. Denote by μ the distribution of Z_0. The process $\{X_k,\ k \in \mathbb{N}\}$

is a Markov chain with Markov kernel P given for $x \in \mathsf{X}$ and $h \in \mathbb{F}_+(\mathsf{X})$ by

$$Ph(x) = \mathbb{E}\left[h(f_{Z_0}(x))\right] = \int_{\mathsf{Z}} h(f(x,z))\mu(\mathrm{d}z) . \tag{20.3.7}$$

For $x \in \mathsf{X}$, define the forward chain $\{X_n^x,\ n \in \mathbb{N}\}$ and the backward process $\{Y_n^x,\ n \in \mathbb{N}\}$ starting from $X_0^x = Y_0^x = x$ by

$$X_k^x = f_{Z_k} \circ \cdots \circ f_{Z_1}(x_0) , \tag{20.3.8}$$
$$Y_k^x = f_{Z_1} \circ \cdots \circ f_{Z_k}(x_0) . \tag{20.3.9}$$

By varying the starting point x but using the same maps, we define a family of Markov chains, one for each starting state, on the same probability space. We can thus consider the joint behavior of the Markov chains started at x and y and the distance $\mathsf{d}(X_n^x, X_n^y)$ between the chains after n time steps. An important property is that since $\{Z_k,\ k \in \mathbb{N}\}$ is an i.i.d. sequence, Y_k^x has the same distribution as X_k^x for each $k \in \mathbb{N}$. For a random variable Y, we write $\|Y\|_p = \{\mathbb{E}[|Y|^p]\}^{1/p}$.

Theorem 20.3.6. *Assume that there exist $\varepsilon \in (0,1)$, $p \geq 1$, and $x_0 \in \mathsf{X}$ such that for all $(x,y) \in \mathsf{X} \times \mathsf{X}$,*

$$\left\|\mathsf{d}(f_{Z_0}(x), f_{Z_0}(y))\right\|_p \leq (1-\varepsilon)\mathsf{d}(x,y) . \tag{20.3.10}$$

Assume, moreover, that there exists $x_0 \in \mathsf{X}$ such that

$$\left\|\mathsf{d}(x_0, f_{Z_0}(x_0))\right\|_p < \infty . \tag{20.3.11}$$

Let P be the Markov kernel given by (20.3.7). Then $P(x,\cdot) \in \mathbb{S}(\mathsf{X},p)$ for all $x \in \mathsf{X}$, $\Delta_{\mathsf{d},p}(P) \leq 1-\varepsilon$, and the unique invariant probability π is in $\mathbb{S}_p(\mathsf{X},\mathsf{d})$. Moreover, for all $x \in \mathsf{X}$, the sequence $\{f_{Z_1} \circ \cdots \circ f_{Z_n}(x), n \in \mathbb{N}\}$ converges almost surely as n tends to infinity and in the pth mean to a random variable Y_∞ whose distribution is π.

Proof. By the triangle inequality, the Minkowski inequality, and (20.3.10), we get that

$$\begin{aligned}\left\|\mathsf{d}(x, f_{Z_0}(x))\right\|_p &\leq \mathsf{d}(x,x_0) + \left\|\mathsf{d}(x_0, f_{Z_0}(x_0))\right\|_p + \left\|\mathsf{d}(f_{Z_0}(x), f_{Z_0}(x_0))\right\|_p \\ &\leq (2-\varepsilon)\mathsf{d}(x,x_0) + \left\|\mathsf{d}(x_0, f_{Z_0}(x_0))\right\|_p .\end{aligned}$$

By definition of the kernel P, (20.3.11) means that $P(x_0,\cdot) \in \mathbb{S}_p(\mathsf{X},\mathsf{d})$.

Condition (20.3.10) implies that $\mathbf{W}_{\mathsf{d},p}(P(x,\cdot),P(y,\cdot)) \leq (1-\varepsilon)\mathsf{d}(x,y)$; hence by Lemma 20.3.2, we get that $\Delta_{\mathsf{d},p}(P) \leq 1-\varepsilon$. By Theorem 20.3.4, this proves the existence and uniqueness of the invariant measure $\pi \in \mathbb{S}_p(\mathsf{X},\mathsf{d})$.

We now establish the expression for the limiting distribution. From (20.3.10), we get for all $n \geq 1$ and $x, y \in \mathsf{X}$ that

$$
\begin{aligned}
&\|\mathsf{d}(Y_n^x, Y_n^y)\|_p \\
&= \{\mathbb{E}\,[\mathbb{E}\,[\,\mathsf{d}^p(f_{Z_1} \circ f_{Z_2} \circ \cdots \circ f_{Z_n}(x), f_{Z_1} \circ f_{Z_2} \circ \cdots \circ f_{Z_n}(y)) \,|\, Z_2, \ldots, Z_n]]\}^{1/p} \\
&\leq (1-\varepsilon)\,\|\mathsf{d}(f_{Z_2} \circ \cdots \circ f_{Z_n}(x), f_{Z_2} \circ \cdots \circ f_{Z_n}(y))\|_p = (1-\varepsilon)\,\left\|\mathsf{d}(Y_{n-1}^x, Y_{n-1}^y)\right\|_p ,
\end{aligned}
$$

where we have used that (Y_{n-1}^x, Y_{n-1}^y) and $(f_{Z_2} \circ \cdots \circ f_{Z_n}(x), f_{Z_2} \circ \cdots \circ f_{Z_n}(y))$ have the same distributions. By iterating this inequality, we therefore obtain $x, y \in \mathsf{X}$,

$$
\|\mathsf{d}(Y_n^x, Y_n^y)\|_p \leq (1-\varepsilon)^n \mathsf{d}(x, y) .
$$

Since Z_{n+1} is independent of Y_n^x, the latter inequality implies that for all $n \geq 1$,

$$
\begin{aligned}
\left\|\mathsf{d}(Y_n^x, Y_{n+1}^x)\right\|_p &= \left\{\mathbb{E}\left[\mathbb{E}\left[\,\mathsf{d}^p(f_{Z_1} \circ \cdots \circ f_{Z_n}(x), f_{Z_1} \circ \ldots f_{Z_n}(f_{Z_{n+1}}(x))) \,\middle|\, Z_{n+1}\right]\right]\right\}^{1/p} \\
&\leq (1-\varepsilon)^n \left\|\mathsf{d}(x, f_{Z_{n+1}}(x))\right\|_p = (1-\varepsilon)^n \left\|\mathsf{d}(x, f_{Z_0}(x))\right\|_p .
\end{aligned}
$$

This proves that the series $\sum_n \mathsf{d}(Y_n^x, Y_{n+1}^x)$ is convergent in pth mean and almost surely. Let this limit be denoted by Y_∞^x. Since $\{Z_n,\ n \in \mathbb{N}\}$ is an i.i.d. sequence, for each n the distribution of Y_n^x is $P^n(x, \cdot)$. Since convergence in the Wasserstein distance implies weak convergence, we have $P^n(x, \cdot) \stackrel{\text{w}}{\Rightarrow} \pi$, and therefore the distribution of Y_n^x is π for all $x \in \mathsf{X}$. ☐

Example 20.3.7. Consider the bilinear process defined by the recurrence

$$
X_k = aX_{k-1} + bX_{k-1}Z_k + Z_k , \tag{20.3.12}
$$

where a and b are nonzero real numbers and $\{Z_k,\ k \in \mathbb{N}\}$ is a sequence of i.i.d. random variables that are independent of X_0 and such that $\mathbb{E}\,[|Z_0|^p] < \infty$ for some $p \geq 1$. Writing $f_z(x) = (a + bx)z$, we have

$$
\mathbf{W}_{\mathsf{d},p}\,(P(x, \cdot), P(y \cdot)) \leq \left\|f_{Z_0}(x) - f_{Z_0}(y)\right\|_p = \|a + bZ_0\|_p\,|x - y| .
$$

Thus $\Delta_{\mathsf{d},p}\,(P) \leq \|a + bZ_0\|_p$. If $\|a + bZ_0\|_p < 1$, then (20.3.10) and (20.3.11) hold. The invariant probability can be expressed as

$$
Y_\infty = Z_0 \sum_{k=0}^{\infty} \prod_{j=1}^{k} (a + bZ_j) ,
$$

the series being convergent almost surely and in pth mean. ◀

20.4 Nonuniform Geometric Convergence

We pursue here the parallel with Chapter 18. The results of Section 18.4 were obtained under a geometric drift condition and the assumption that an (m,ε)-Doeblin set exists. In the present context, Doeblin sets will be replaced by $(\mathsf{c},m,\varepsilon)$-contracting sets on which the restriction of P has certain contractivity properties with respect to the Wasserstein distance $\mathbf{W}_{\mathsf{c}}$.

Definition 20.4.1 ($(\mathsf{c},m,\varepsilon)$-contracting set) *A set $\bar{C} \subset \mathscr{X} \otimes \mathscr{X}$ is called a $(\mathsf{c},m,\varepsilon)$-contracting set if for all $(x,y) \in \bar{C}$,*

$$\mathbf{W}_{\mathsf{c}}\left(P^m(x,\cdot),P^m(y,\cdot)\right) \leq (1-\varepsilon)\mathsf{c}(x,y)\ .$$

Given the existence of such a set and a drift condition, we can prove the geometric convergence in a Wasserstein distance of the iterates of the kernel to the invariant probability. For simplicity, we consider only the case $m = 1$; the extension to $m \geq 1$ is straightforward. The result is based on the following technical proposition.

Proposition 20.4.2 *Assume that there exist a kernel coupling K of (P,P), a measurable function $\bar{V} : \mathsf{X}\times\mathsf{X} \to [1,\infty)$, a set $\bar{C} \in \mathscr{X}\otimes\mathscr{X}$, $\bar{d} > 0$, and $\varepsilon \in (0,1)$ such that*

$$K\mathsf{c} \leq (1-\varepsilon\mathbb{1}_{\bar{C}})\mathsf{c}\ , \qquad \{\bar{V} \leq \bar{d}\} \subset \bar{C}\ .$$

If K satisfies the geometric drift condition $\mathrm{D_g}(\bar{V},\bar{\lambda},\bar{b},\bar{C})$, then for all $(\alpha,\beta) \in (0,1)\times[0,\infty)$, $x,y \in \mathsf{X}$ and $n \in \mathbb{N}$,

$$\mathbf{W}_{\mathsf{c}^{1-\alpha}\bar{V}^{\alpha}}\left(P^n(x,\cdot),P^n(y,\cdot)\right) \leq \rho^n_{\alpha,\beta}\mathsf{c}^{1-\alpha}(x,y)\left[\bar{V}(x,y)+\beta\right]^{\alpha}\ , \tag{20.4.1}$$

with

$$\rho_{\alpha,\beta} = \left[(1-\varepsilon)^{1-\alpha}\left(\frac{\bar{\lambda}+\bar{b}+\beta}{1+\beta}\right)^{\alpha}\right] \vee \left(\frac{\bar{\lambda}\bar{d}+\beta}{\bar{d}+\beta}\right)^{\alpha}\ . \tag{20.4.2}$$

Moreover, for all $\beta \geq 0$, there exists $\alpha \in (0,1)$ such that $\rho_{\alpha,\beta} < 1$, and conversely, for all $\alpha \in (0,1)$, there exists $\beta \geq 0$ such that $\rho_{\alpha,\beta} < 1$.

Remark 20.4.3. We actually prove that for each $\alpha \in (0,1)$, there exist $\beta > 0$ and $\rho_{\alpha,\beta} \in (0,1)$ given by (20.4.2) such that $Kc_{\alpha,\beta} \leq \rho_{\alpha,\beta}c_{\alpha,\beta}$ with $c_{\alpha,\beta} = c^{1-\alpha}(\bar{V}+\beta)^{\alpha}$. Thus we can apply Theorem 20.3.4, and by Theorem 20.1.3, there exists a coupling $\{(X_n,X'_n),\ n\in\mathbb{N}\}$ such that $\mathbb{E}_{x,x'}[c_{\alpha,\beta}(X_n,X'_n)] \leq \rho^n c_{\alpha,\beta}(x,x')$. Moreover, if $c \le \bar{V}$ and $\bar{V}(x,x') = \{V(x)+V(x')\}/2$, this yields, for all $n \ge 0$,

$$\mathbb{E}_{x,x'}[c(X_n,X'_n)] \le \mathbb{E}_{x,x'}[c_{\alpha,\beta}(X_n,X'_n)] \le \tfrac{1}{2}(1+\beta)\rho^n(V(x)+V(x')) \,. \tag{20.4.3}$$

Proof (of Proposition 20.4.2). For $\beta \ge 0$, set $\bar{V}_\beta = \bar{V} + \beta$ and

$$\rho = \sup_{x,y\in\mathsf{X}} \left[(1-\varepsilon \mathbb{1}_{\bar{C}}(x,y))^{1-\alpha} \left(\frac{K\bar{V}_\beta(x,y)}{\bar{V}_\beta(x,y)} \right)^{\alpha} \right] ,$$

which is finite, since K satisfies condition $\mathrm{D_g}(\bar{V},\bar{\lambda},\bar{b},\bar{C})$. Furthermore, Hölder's inequality yields

$$\begin{aligned} K(\mathsf{c}^{1-\alpha}\bar{V}_\beta^\alpha) &\le (K\mathsf{c})^{1-\alpha}(K\bar{V}_\beta)^\alpha \\ &\le \left[(1-\varepsilon\mathbb{1}_{\bar{C}})^{1-\alpha} \left(K\bar{V}_\beta/\bar{V}_\beta \right)^\alpha \right] \mathsf{c}^{1-\alpha}\bar{V}_\beta^\alpha \le \rho\, \mathsf{c}^{1-\alpha}\bar{V}_\beta^\alpha \,. \end{aligned}$$

Using $\bar{V} \le \bar{V}_\beta$ and a straightforward induction, we obtain

$$K^n(\mathsf{c}^{1-\alpha}\bar{V}^\alpha) \le K^n(\mathsf{c}^{1-\alpha}\bar{V}_\beta^\alpha) \le \rho^n \mathsf{c}^{1-\alpha}\bar{V}_\beta^\alpha = \rho^n \mathsf{c}^{1-\alpha}(\bar{V}+\beta)^\alpha \,. \tag{20.4.4}$$

Moreover, Corollary 20.1.4, (20.1.12), yields, for all $x,y \in \mathsf{X}$,

$$\mathbf{W}_{\mathsf{c}^{1-\alpha}\bar{V}^\alpha}(P^n(x,\cdot),P^n(y,\cdot)) \le K^n(\mathsf{c}^{1-\alpha}\bar{V}^\alpha)(x,y) \,.$$

Combining this bound with (20.4.4) proves (20.4.1) provided that $\rho \le \rho_{\alpha,\beta}$, where $\rho_{\alpha,\beta}$ is defined in (20.4.2). We will now establish this inequality. Since the geometric drift condition $\mathrm{D_g}(\bar{V},\bar{\lambda},\bar{b},\bar{C})$ holds for the Markov kernel K, it follows that

$$\frac{K\bar{V}_\beta}{\bar{V}_\beta} \le \varphi(\bar{V}) \,, \tag{20.4.5}$$

with

$$\varphi(v) = \frac{\bar{\lambda} v + \bar{b}\mathbb{1}_{\bar{C}} + \beta}{v+\beta} \,.$$

The function φ is monotone, $\varphi(0) \ge 1$, and $\lim_{v\to\infty}\varphi(v) = \bar{\lambda} < 1$. It is thus nonincreasing, and since $\bar{V} \ge \mathbb{1}_{\bar{C}} + \bar{d}\mathbb{1}_{\bar{C}^c}$, we obtain

$$\varphi(\bar{V}) \le \varphi(\mathbb{1}_{\bar{C}} + \bar{d}\mathbb{1}_{\bar{C}^c}) = \frac{\bar{\lambda}+\bar{b}+\beta}{1+\beta}\mathbb{1}_{\bar{C}} + \frac{\bar{\lambda}\bar{d}+\beta}{\bar{d}+\beta}\mathbb{1}_{\bar{C}^c} \,.$$

Combining with (20.4.5) yields

$$
\begin{aligned}
&(1-\varepsilon \mathbb{1}_{\bar{C}})^{1-\alpha} \left(K\bar{V}_\beta / \bar{V}_\beta\right)^\alpha \\
&\qquad \le \left[(1-\varepsilon)^{1-\alpha} \left(\frac{\bar{\lambda}+\bar{b}+\beta}{1+\beta}\right)^\alpha\right] \mathbb{1}_{\bar{C}} + \left(\frac{\bar{\lambda}\bar{d}+\beta}{\bar{d}+\beta}\right)^\alpha \mathbb{1}_{\bar{C}^c} \le \rho_{\alpha,\beta} \;.
\end{aligned}
$$

Finally, since ρ is the supremum of the left-hand side over $\mathsf{X} \times \mathsf{X}$, we obtain that $\rho \le \rho_{\alpha,\beta}$. The last part of the theorem follows from Lemma 20.4.4 below. □

The bound (20.4.1) in Proposition 20.4.2 is useful only if $\rho_{\alpha,\beta} < 1$ for a suitable choice of $(\alpha,\beta) \in (0,1) \times [0,\infty)$. We may first fix $\beta \ge 0$ and search for $\alpha^\star(\beta) = \arg\min_{\alpha \in (0,1)} \rho_{\alpha,\beta}$. Optimizing (20.4.2), we obtain

$$
\begin{aligned}
&\alpha^\star(\beta) \\
&= [\log(1-\varepsilon)] \left[\log(1-\varepsilon) + \log\left(\frac{\bar{\lambda}\bar{d}+\beta}{\bar{d}+\beta}\right) - \log\left(\frac{\bar{\lambda}+\bar{b}+\beta}{1+\beta}\right)\right]^{-1} \;. \qquad (20.4.6)
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
&\log \rho_{\alpha^\star(\beta),\beta} \\
&= \inf_{\alpha \in (0,1)} \log \rho_{\alpha,\beta} = \frac{\log(1-\varepsilon)\log\left(\frac{\bar{\lambda}\bar{d}+\beta}{\bar{d}+\beta}\right)}{\log(1-\varepsilon) + \log\left(\frac{\bar{\lambda}\bar{d}+\beta}{\bar{d}+\beta}\right) - \log\left(\frac{\bar{\lambda}+\bar{b}+\beta}{1+\beta}\right)} < 0 \;, \qquad (20.4.7)
\end{aligned}
$$

where the strict inequality follows from $(\varepsilon, \bar{\lambda}) \in (0,1)^2$ and $\bar{\lambda} + \bar{b} \ge 1$. To get the optimal rate, we take the infimum of (20.4.7) with respect to β, that is,

$$
\log \rho_{\alpha^\star(\beta^\star),\beta^\star} = \inf_{\beta \in \mathbb{R}} \log \rho_{\alpha^\star(\beta),\beta} \;,
$$

but unfortunately, the expression for $\log \rho_{\alpha^\star(\beta^\star),\beta^\star}$ is not explicit. Instead, we can consider particular values of $\log \rho_{\alpha^\star(\beta),\beta}$, since they are strictly negative for all $\beta \ge 0$. For example, taking $\beta = 0$, we get

$$
\log \rho_{\alpha^\star(0),0} = \inf_{\alpha \in (0,1)} \log \rho_{\alpha,0} = \frac{\log(1-\varepsilon)\log \bar{\lambda}}{\log(1-\varepsilon) + \log \bar{\lambda} - \log(\bar{\lambda} + \bar{b})} < 0 \;, \qquad (20.4.8)
$$

which can be compared to the bounds obtained in Theorem 19.4.1. Still, the bound in Proposition 20.4.2 is associated with the cost function $\mathsf{c}^{1-\alpha}\bar{V}^\alpha$, and in some cases, it may be interesting to obtain a geometrically decreasing bound when α is fixed and arbitrarily close to 1. The rationale for this is that for $(x,y) \in \mathsf{X}^2$, $\lim_{\alpha \to 1} \mathsf{c}^{1-\alpha}\bar{V}^\alpha = \mathbb{1}\{x \ne y\}\bar{V}(x,y)$. While it is not possible under the assumptions of Proposition 20.4.2 to obtain a bound for the cost function $(x,y) \mapsto \mathbb{1}\{x \ne y\}\bar{V}(x,y)$, it is interesting instead to get a bound for any cost function $\mathsf{c}^{1-\alpha}\bar{V}^\alpha$, where α is arbitrarily close to 1. Fix now any $\alpha \in (0,1)$ and let us show that $\rho_{\alpha,\beta}$ can be made strictly less than 1 with a convenient choice of β. Note that the function

$$\beta \mapsto \psi(\beta) = (1-\varepsilon)^{1-\alpha} \left(\frac{\bar{\lambda} + \bar{b} + \beta}{1+\beta} \right)^{\alpha}$$

tends to $(1-\varepsilon)^{1-\alpha} < 1$ as β tends to infinity. This allows us to take $\beta = \beta_0$ large enough that $\psi(\beta_0) < 1$. We then obtain

$$\rho_{\alpha,\beta_0} = \psi(\beta_0) \vee \left(\frac{\bar{\lambda}\bar{d} + \beta_0}{\bar{d} + \beta_0} \right)^{\alpha} < 1 .$$

An optimal choice of β for a given α can theoretically be obtained by solving $\psi(\beta) = \left(\frac{\bar{\lambda}\bar{d}+\beta}{\bar{d}+\beta} \right)^{\alpha}$, but once again, this equation does not admit any closed-form solution in general. We summarize our findings in the following lemma.

Lemma 20.4.4 *For all $\beta \geq 0$, there exists $\alpha \in (0,1)$ such that $\rho_{\alpha,\beta} < 1$, and conversely, for all $\alpha \in (0,1)$, there exists $\beta \geq 0$ such that $\rho_{\alpha,\beta} < 1$.*

We now state and prove the main result of this section. It provides geometric rates of convergence in the Wasserstein distance and parallels the results of Theorems 18.4.3 and 19.4.1, which were established for the V-norm.

Theorem 20.4.5. *Let P be a Markov kernel satisfying the drift condition $D_g(V, \lambda, b)$ and assume that the cost function c satisfies* **A** 20.1.5 *and for all $x, y \in \mathsf{X}$,*

$$\mathbf{W}_{\mathsf{c}}\left(P(x,\cdot), P(y,\cdot)\right) \leq \mathsf{c}(x,y) .$$

Assume, moreover, that there exist $\varepsilon \in (0,1)$ and $d > 0$ such that $\lambda + 2b/(1+d) < 1$ and that $\{V \leq d\} \times \{V \leq d\}$ is a $(\mathsf{c}, 1, \varepsilon)$-contracting set. Moreover, assume that $\mathsf{c}(x,y) \leq \bar{V}(x,y) := \{V(x) + V(y)\}/2$ for all $x, y \in \mathsf{X}$. Then

(i) P admits a unique invariant measure π and $\pi(V) < \infty$;
(ii) for every $\alpha \in (0,1)$, there exist $\rho \in (0,1)$ and $\vartheta < \infty$ such that for all initial distributions ξ and all $n \in \mathbb{N}$,

$$\mathbf{W}_{\mathsf{c}}\left(\xi P^n, \pi\right) \leq \mathbf{W}_{\mathsf{c}^{1-\alpha}\bar{V}^{\alpha}}\left(\xi P^n, \pi\right) \leq \vartheta \rho^n \left[\xi(V^{\alpha}) + \pi(V^{\alpha})\right] . \tag{20.4.9}$$

Remark 20.4.6. Theorem 20.4.5 could be proved by applying Theorem 20.3.4 with c replaced by $\mathsf{c}^{1-\alpha}\bar{V}^{\alpha}$, since (20.4.1) implies that for every α in $(0,1)$, there exists a sufficiently large m such that P^m is $\mathbf{W}_{\mathsf{c}^{1-\alpha}\bar{V}^{\alpha}}$-uniformly ergodic. Indeed, for a fixed $\alpha \in (0,1)$, choose β such that $\rho_{\alpha,\beta} < 1$ (this can be done by Lemma 20.4.4). Then using (20.4.1) and $\bar{V} \geq 1$,

$$\begin{aligned} \mathbf{W}_{\mathsf{c}^{1-\alpha}\bar{V}^{\alpha}}\left(P^m(x,\cdot), P^m(y,\cdot)\right) &\leq \rho_{\alpha,\beta}^m \mathsf{c}^{1-\alpha}(x,y) \left[\bar{V}(x,y) + \beta\right]^{\alpha} \\ &\leq \left(\rho_{\alpha,\beta}^m (1+\beta)^{\alpha}\right) \left(\mathsf{c}^{1-\alpha}\bar{V}^{\alpha}\right)(x,y) . \end{aligned}$$

Since $\rho_{\alpha,\beta} < 1$, we can choose m sufficiently large that $\rho^m_{\alpha,\beta}(1+\beta)^\alpha < 1$. And for such m, P^m is $\mathbf{W}_{\mathsf{c}^{1-\alpha}\bar{V}^\alpha}$-uniformly ergodic. However, we decide to prove Theorem 20.4.5 along another path, which highlights how Theorem 20.2.1 can be used to obtain the existence and uniqueness of the invariant probability measure.

Proof (of Theorem 20.4.5). Set $\bar{C} = \{V \le d\} \times \{V \le d\}$. According to Theorem 20.1.3, there exists a kernel coupling K of (P,P) such that

$$K\mathsf{c} \le (1-\varepsilon\mathbb{1}_{\bar{C}})\mathsf{c} .$$

(i) This kernel K being chosen, we have (as in the proof of Theorem 19.4.1)

$$K\bar{V} \le \bar{\lambda}\bar{V} + b\mathbb{1}_{\bar{C}} ,$$

with $\bar{\lambda} = \lambda + 2b/(1+d)$. Thus K satisfies the drift condition $\mathrm{D_g}(\bar{V},\bar{\lambda},b,\bar{C})$, whence

$$K\bar{V} + 1 - \bar{\lambda} \le \bar{V} + \bar{b}\mathbb{1}_{\bar{C}} .$$

Thus the condition theorem, Theorem 20.2.1 (i), holds. By Proposition 14.1.8, $\sup_{n\ge 0} P^n V \le V + b/(1-\lambda)$, so that Theorem 20.2.1 (ii) also holds (with $\psi(v) = v$). We can therefore apply Theorem 20.2.1 to prove that P admits a unique invariant probability π. Moreover, we know by Lemma 14.1.10 that $\pi(V) < \infty$.

(ii) Since $\mathsf{c} \le \bar{V}$, we get $\mathsf{c} \le \mathsf{c}^{1-\alpha}\bar{V}^\alpha \le \bar{V}^\alpha$. By definition of $\bar{C}$, if $(x,x') \notin \bar{C}$, then $\bar{V}(x,x') \ge (d+1)/2$. Setting $\bar{d} = (d+1)/2$, we have $\{\bar{V} < \bar{d}\} \subset \bar{C}$. We can thus apply Proposition 20.4.2. Using Corollary 20.1.4 and (20.4.1), we get for all $\gamma \in \mathscr{C}(\xi,\pi)$,

$$\mathbf{W}_{\mathsf{c}}(\xi P^n,\pi) \le \mathbf{W}_{\mathsf{c}^{1-\alpha}\bar{V}^\alpha}(\xi P^n,\pi) \le \int_{\mathsf{X}\times\mathsf{X}} \mathbf{W}_{\mathsf{c}^{1-\alpha}\bar{V}^\alpha}(P^n(x,\cdot),P^n(y,\cdot))\,\gamma(\mathrm{d}x\mathrm{d}y) ,$$

proving (20.4.9).

□

Corollary 20.4.7 *Under the assumptions of Theorem 20.4.5, for all $\alpha \in (0,1)$, there exist a finite constant ϑ and $\rho \in (0,1)$ such that for all measurable functions $f \in \mathrm{Lip}_{\mathsf{c}^{1-\alpha}\bar{V}^\alpha}(\mathsf{X})$ and all $n \in \mathbb{N}$,*

$$|\xi P^n(f) - \pi(f)| \le \vartheta\rho^n\,[\xi(V^\alpha) + \pi(V^\alpha)]\,|f|_{\mathrm{Lip}(\mathsf{c}^{1-\alpha}\bar{V}^\alpha)} . \qquad (20.4.10)$$

Proof. Fix $x_0 \in \mathsf{X}$. Since $\mathsf{c}^{1-\alpha}\bar{V}^\alpha \le \bar{V}$, it follows that for all $x \in \mathsf{X}$,

$$|f(x)| \le |f(x_0)| + \bar{V}(x_0,x)\,|f|_{\mathrm{Lip}(\mathsf{c}^{1-\alpha}\bar{V}^\alpha)} ,$$

and thus $f \in \mathrm{L}^1(\pi)$, since $\pi(V) < \infty$. By (20.1.9) applied to the cost function $\mathsf{c}^{1-\alpha}\bar{V}^\alpha$,

$$|\xi P^n(f) - \pi(f)| \leq \mathbf{W}_{\mathsf{c}^{1-\alpha}\bar{V}^\alpha}(\xi P^n, \pi) \, |f|_{\mathrm{Lip}(\mathsf{c}^{1-\alpha}\bar{V}^\alpha)} \, .$$

Theorem 20.4.5 combined with Lemma 20.4.4 completes the proof. □

20.5 Subgeometric Rates of Convergence for the Wasserstein Distance

In this section we establish subgeometric rates of convergence in the Wasserstein distance under the drift condition $D_{\mathrm{sg}}(V,\phi,b,C)$. Recall that for an increasing concave function ϕ, the subgeometric sequence r_ϕ is defined in (16.1.13) by $r_\phi(t) = \phi \circ H_\phi^{-1}(t)$, where H_ϕ is the primitive of $1/\phi$ that vanishes at 1. Set $\bar{V}(x,y) = V(x) + V(y) - 1$ and recall that for a sequence r, we define $r^0(n) = \sum_{j=0}^n r(j)$. Our main result will be a consequence of the following technical lemma.

Lemma 20.5.1 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that* **A** 20.1.5 *and the drift condition $D_{\mathrm{sg}}(V,\phi,b,C)$ hold with $\sup_C V < \infty$, $d = \inf_{C^c} \phi \circ V > b$. Assume, moreover, that there exists a kernel coupling K of (P,P) such that*

$$K\mathsf{c} \leq (1 - \varepsilon \mathbb{1}_{C \times C})\mathsf{c} \, . \tag{20.5.1}$$

Let $\alpha \in (0,1)$, $\kappa \in (0, 1 - b/d)$, and set $\bar{\phi} = \kappa\phi$ and $\tilde{\mathsf{c}} = \mathsf{c}^{1-\alpha}\bar{\phi}^\alpha \circ \bar{V}$. Then there exists a finite constant ϑ such that for all $n \in \mathbb{N}$ and $x, y \in \mathsf{X}$,

$$K^n \tilde{\mathsf{c}} \leq \vartheta \tilde{\mathsf{c}} \, , \tag{20.5.2}$$

$$[r^0_{\bar{\phi}}(n)]^\alpha K^n \mathsf{c}^{1-\alpha} \leq \vartheta \mathsf{c}^{1-\alpha}\bar{V}^\alpha \, , \tag{20.5.3}$$

$$r^\alpha_{\bar{\phi}}(n) K^n \mathsf{c}^{1-\alpha}(x,y) \leq \vartheta \mathsf{c}^{1-\alpha}(x,y)\{V^\alpha(x) + \bar{\phi}^\alpha \circ V(y)\} \, . \tag{20.5.4}$$

Proof. By Lemma 19.5.3, for $\kappa \in (0, 1 - b/d)$, the drift condition $D_{\mathrm{sg}}(\bar{V}, \bar{\phi}, \bar{b}, \bar{C})$ holds for the kernel K with $\bar{\phi} = \kappa\phi$ and $\bar{b} = 2b$. Fix $\alpha \in (0,1)$ and pick $\delta \in (0,1)$ such that $(1 - \varepsilon \mathbb{1}_{\bar{C}})^{1-\alpha}(1 + \delta \mathbb{1}_{\bar{C}})^\alpha \leq 1$.

(i) Choose $M \geq 0$ such that $\bar{\phi}(\bar{b}+1) - \bar{\phi}(1) \leq M\delta$. By concavity, for all $v \geq 1$, we have $\bar{\phi}(v + \bar{b}) - \bar{\phi}(v) \leq \bar{\phi}(\bar{b}+1) - \bar{\phi}(1)$, whence

$$\begin{aligned} K(\bar{\phi} \circ \bar{V} + M) &\leq \bar{\phi}(K\bar{V}) + M \leq \bar{\phi}(\bar{V} + \bar{b}\mathbb{1}_{\bar{C}}) + M \\ &\leq \bar{\phi} \circ \bar{V} + M + [\bar{\phi}(\bar{b}+1) - \bar{\phi}(1)]\mathbb{1}_{\bar{C}} \leq (1 + \delta \mathbb{1}_{\bar{C}})(\bar{\phi} \circ \bar{V} + M) \, . \end{aligned}$$

Combining this bound with (20.5.1) and Hölder's inequality yields

$$K[\mathsf{c}^{1-\alpha}(\bar{\phi} \circ \bar{V} + M)^\alpha] \leq (K\mathsf{c})^{1-\alpha}[K(\bar{\phi} \circ \bar{V} + M)]^\alpha \leq \mathsf{c}^{1-\alpha}(\bar{\phi} \circ \bar{V} + M)^\alpha \, .$$

We apply the previous inequality repeatedly, and since ϕ is bounded away from zero, we obtain

$$K^n \tilde{\mathsf{c}} \le K^n [\mathsf{c}^{1-\alpha}(\bar{\phi} \circ \bar{V} + M)^\alpha] \le \mathsf{c}^{1-\alpha}(\bar{\phi} \circ \bar{V} + M)^\alpha \le \vartheta \tilde{\mathsf{c}} \,,$$

for a constant ϑ independent of n. This proves (20.5.2).

(ii) Since $\mathrm{D}_{\mathrm{sg}}(\bar{V}, \bar{\phi}, \bar{b}, \bar{C})$ holds, we can apply Proposition 16.1.11: the sequence of nonnegative functions $\{V_k,\, k \in \mathbb{N}\}$ on $\mathsf{X} \times \mathsf{X}$ defined by $V_k = H_{\bar{\phi}}^{-1}(H_{\bar{\phi}} \circ \bar{V} + k) - H_{\bar{\phi}}^{-1}(k)$ satisfies

$$KV_{k+1} + r_{\bar{\phi}}(k) \le V_k + b' r_{\bar{\phi}}(k) \mathbb{1}_{\bar{C}} \,, \tag{20.5.5}$$

with $b' = \bar{b} r_{\bar{\phi}}(1)/r_{\bar{\phi}}^2(0)$. Set $M_\delta = \sup_{k \in \mathbb{N}} \{\delta^{-1} b' r_{\bar{\phi}}(k) - r_{\bar{\phi}}^0(k-1)\}$, which is finite, since $\lim_{n \to \infty} r_{\bar{\phi}}(n)/r_{\bar{\phi}}^0(n) = 0$. Setting $S_k = V_k + r_{\bar{\phi}}^0(k-1) + M_\delta$, (20.5.5) can be reexpressed as

$$\begin{aligned} KS_{k+1} &\le S_k + b' r_{\bar{\phi}}(k) \mathbb{1}_{\bar{C}} = \left(1 + \frac{b' r_{\bar{\phi}}(k)}{S_k} \mathbb{1}_{\bar{C}}\right) S_k \\ &\le \left(1 + \frac{b' r_{\bar{\phi}}(k)}{r_{\bar{\phi}}^0(k-1) + M_\delta} \mathbb{1}_{\bar{C}}\right) S_k \le (1 + \delta \mathbb{1}_{\bar{C}}) S_k \,. \end{aligned} \tag{20.5.6}$$

Combining (20.5.1) and (20.5.6) and applying Hölder's inequality yields for all $k \ge 0$,

$$K(\mathsf{c}^{1-\alpha} S_{k+1}^\alpha) \le (K\mathsf{c})^{1-\alpha} (KS_{k+1})^\alpha \le \mathsf{c}^{1-\alpha} S_k^\alpha \,.$$

Applying repeatedly the previous inequality and $r_{\bar{\phi}}^0(n-1) + M_\delta \le S_n$, we obtain

$$\begin{aligned} [r_{\bar{\phi}}^0(n-1)]^\alpha K^n(\mathsf{c}^{1-\alpha}) &\le (r_{\bar{\phi}}^0(n-1) + M_\delta)^\alpha K^n(\mathsf{c}^{1-\alpha}) \le K^n(\mathsf{c}^{1-\alpha} S_n^\alpha) \le \mathsf{c}^{1-\alpha} S_0^\alpha \\ &\le \mathsf{c}^{1-\alpha}(V_0 + M_\delta)^\alpha \le \mathsf{c}^{1-\alpha}(\bar{V} + M_\delta)^\alpha \le (1 + M_\delta)^\alpha \mathsf{c}^{1-\alpha} \bar{V}^\alpha \,. \end{aligned}$$

This proves (20.5.3).

(iii) In order to obtain (20.5.4), we use (20.5.1) if $V(y) > M$ and (20.5.3) if $V(y) \le M$, which yields

$$\begin{aligned} K\mathsf{c}^{1-\alpha}(x,y) &\le \mathsf{c}^{1-\alpha}(x,y) \mathbb{1}_{\{V(y)>M\}} + \vartheta [r_{\bar{\phi}}^0(n)]^{-\alpha} \mathsf{c}^{1-\alpha}(x,y) \bar{V}^\alpha(x,y) \mathbb{1}_{\{V(y) \le M\}} \\ &\le \mathsf{c}^{1-\alpha}(x,y) \left[\frac{\bar{\phi}^\alpha \circ V(y)}{\bar{\phi}^\alpha(M)} + \frac{\vartheta}{[r_{\bar{\phi}}^0(n)]^\alpha} \bar{V}^\alpha(x,y) \mathbb{1}_{\{V(y) \le M\}} \right] . \end{aligned} \tag{20.5.7}$$

Recalling that $\bar{V}(x,y) = V(x) + V(y) - 1$, we have

$$\bar{V}^\alpha(x,y) \mathbb{1}_{\{V(y) \le M\}} \le (V(y) - 1)^\alpha \mathbb{1}_{\{V(y) \le M\}} + V^\alpha(x) \,. \tag{20.5.8}$$

By concavity of $\bar\phi$, we have for $1 \le v \le M$,

$$v-1 \le (M-1)\frac{\bar\phi(v)-\bar\phi(1)}{\bar\phi(M)-\bar\phi(1)} \le \frac{(M-1)\bar\phi(v)}{\bar\phi(M)-\bar\phi(1)}\,.$$

Replacing v by $V(y)$ and plugging this bound into (20.5.8) yields

$$\bar V^{\alpha}(x,y)\mathbb{1}_{\{V(y)\le M\}} \le \left(\frac{(M-1)\bar\phi\circ V(y)}{\bar\phi(M)-\bar\phi(1)}\right)^{\alpha} + V^{\alpha}(x)\,.$$

Combining this with (20.5.7), we finally get

$$\begin{aligned}
&K\mathsf{c}^{1-\alpha}(x,y)\\
&\le \mathsf{c}^{1-\alpha}(x,y)\left[\frac{\bar\phi^{\alpha}\circ V(y)}{\bar\phi^{\alpha}(M)} + \vartheta\left(\frac{M-1}{r^{0}_{\bar\phi}(n)}\right)^{\alpha}\frac{\bar\phi^{\alpha}\circ V(y)}{[\bar\phi(M)-\bar\phi(1)]^{\alpha}} + \vartheta\frac{V^{\alpha}(x)}{[r^{0}_{\bar\phi}(n)]^{\alpha}}\right].
\end{aligned}$$

Now choose $M = H^{-1}_{\bar\phi}(n)$ and note that $\bar\phi(M) = \bar\phi\circ H^{-1}_{\bar\phi}(n) = r_{\bar\phi}(n) \le r^{0}_{\bar\phi}(n)$. Since

$$\begin{aligned}
M-1 = H^{-1}_{\bar\phi}(n)-1 &= H^{-1}_{\bar\phi}(n) - H^{-1}_{\bar\phi}(0)\\
&= \int_0^n \bar\phi\circ H^{-1}_{\bar\phi}(t)\mathrm{d}t \le \sum_{k=0}^{n}\bar\phi\circ H^{-1}_{\bar\phi}(k) = r^{0}_{\bar\phi}(n)\,,
\end{aligned}$$

we finally obtain, for $n \ge 1$,

$$\begin{aligned}
&K\mathsf{c}^{1-\alpha}(x,y)\\
&\le \frac{\mathsf{c}^{1-\alpha}(x,y)}{r^{\alpha}_{\bar\phi}(n)}\left[\bar\phi^{\alpha}\circ V(y)\left(1+\frac{\vartheta}{\left\{1-\bar\phi(1)/\bar\phi\circ H^{-1}_{\bar\phi}(1)\right\}^{\alpha}}\right) + \vartheta V^{\alpha}(x)\right].
\end{aligned}$$

This completes the proof.

□

Recall that $\bar V(x,y) := V(x)+V(y)-1$.

Theorem 20.5.2. *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$. Assume that* **A** 20.1.5 *and the subgeometric drift condition $\mathrm{D}_{\mathrm{sg}}(V,\phi,b,C)$ hold with $\sup_C V < \infty$, $d = \inf_{C^c}\phi\circ V > b$, and $\mathsf{c} \le \bar V$. Assume, moreover, that $\bar C = C\times C$ is a $(\mathsf{c},1,\varepsilon)$-contracting set and for all $x,y\in\mathsf{X}$,*

$$\mathbf{W}_{\mathsf{c}}(P(x,\cdot),P(y,\cdot)) \le \mathsf{c}(x,y)\,.$$

Then P admits a unique invariant probability measure π and $\pi(\phi\circ V) < \infty$.

Let (Ψ_1,Ψ_2) be inverse Young functions, $(\alpha,\kappa)\in(0,1)\times(0,1-b/d)$, and set

$$\bar{\phi} = \kappa\phi \ , \qquad r(n) = \Psi_1\left[r^{\alpha}_{\bar{\phi}}(n)\right] \ , \qquad f = \Psi_2\left[\bar{\phi}^{\alpha} \circ \bar{V}\right] \ .$$

Then there exists a constant ϑ such that for all $n \in \mathbb{N}$ and $x \in \mathsf{X}$,

$$r(n)\mathbf{W}_{\mathsf{c}^{1-\alpha}f}(P^n(x,\cdot),\pi) \le \vartheta \int_{\mathsf{X}} \mathsf{c}^{1-\alpha}(x,y)\{V^{\alpha}(x) + \phi^{\alpha} \circ V(y)\}\pi(\mathrm{d}y) \ . \quad (20.5.9)$$

Remark 20.5.3. The bound (20.5.9) is useless unless the integral is finite. This is the case if c is bounded or more generally if $\mathsf{c}(x,y) \le \phi \circ \bar{V}(x,y)$, since $\pi(\phi \circ V) < \infty$. The integral is also finite if $\pi(V) < \infty$ (since by assumption, we already know that $c(x,y) \le \bar{V}(x,y)$). In all these cases, we obtain

$$r(n)\mathbf{W}_{\mathsf{c}^{1-\alpha}f}(P^n(x,\cdot),\pi) \le \vartheta' V(x) \ . \quad (20.5.10)$$

Proof (of Theorem 20.5.2).

(i) By Theorem 20.1.3, there exists a kernel coupling K of (P,P) such that (20.1.10) holds and

$$K\mathsf{c} \le (1 - \varepsilon \mathbb{1}_{\bar{C}})\mathsf{c} \ . \quad (20.5.11)$$

Lemma 19.5.3 implies that for all $\kappa \in (0, 1-b/d)$, the drift condition $\mathrm{D_{sg}}(\bar{V}, \bar{\phi}, \bar{b}, \bar{C})$ holds for the kernel K with $\bar{\phi} = \kappa\phi$ and $\bar{b} = 2b$. Thus

$$K\bar{V} + \bar{\phi}(1) \le K\bar{V} + \bar{\phi} \circ \bar{V} \le \bar{V} + \bar{b}\mathbb{1}_{\bar{C}} \ .$$

Finally, condition (i) of Theorem 20.2.1 holds. By Remark 20.2.2, condition (ii) also holds, and thus we can apply Theorem 20.2.1 to obtain the existence and uniqueness of an invariant probability measure π. Moreover, $\pi(\phi \circ V) < \infty$ by Theorem 16.1.12.

(ii) Since (Ψ_1, Ψ_2) is a pair of inverse Young functions, we have $r(n)f \le r^{\alpha}_{\bar{\phi}}(n) + \bar{\phi}^{\alpha} \circ \bar{V}$. This implies

$$\begin{aligned} r(n)\mathbf{W}_{\mathsf{c}^{1-\alpha}f}(P^n(x,\cdot),\pi) &\le r(n)\int_{\mathsf{X}} K^n(\mathsf{c}^{1-\alpha}f)(x,y)\pi(\mathrm{d}y) \\ &\le r^{\alpha}_{\bar{\phi}}(n)\int_{\mathsf{X}} K^n\mathsf{c}^{1-\alpha}(x,y)\pi(\mathrm{d}y) + \int_{\mathsf{X}} K^n\tilde{\mathsf{c}}(x,y)\pi(\mathrm{d}y) \ . \end{aligned} \quad (20.5.12)$$

The first term of the right-hand side can be bounded by (20.5.4). To complete the proof, we have to prove the following bound for the second term of the right-hand side:

$$\int_{\mathsf{X}} K^n\tilde{\mathsf{c}}(x,y)\pi(\mathrm{d}y) \le \vartheta \int_{\mathsf{X}} \mathsf{c}^{1-\alpha}(x,y)\{V^{\alpha}(x) + \phi^{\alpha} \circ V(y)\}\pi(\mathrm{d}y) \ ,$$

where ϑ is some constant that does not depend on n. Indeed, (20.5.2) shows that

$$\int_{\mathsf{X}} K^n\tilde{\mathsf{c}}(x,y)\pi(\mathrm{d}y) \le \int \tilde{\mathsf{c}}(x,y)\pi(\mathrm{d}y) \ .$$

Now since $\bar{\phi}$ is concave, $\bar{\phi}(a+b-1) \le \bar{\phi}(a) + (b-1)\bar{\phi}'(a) \le \bar{\phi}(a) + (b-1)\bar{\phi}'(1)$, and thus, recalling that $\bar{V}(x,y) = V(x) + V(y) - 1$ and $\bar{\phi} = \kappa\phi$, there exists a constant ϑ such that

$$\begin{aligned}\tilde{\mathsf{c}}(x,y) = \mathsf{c}^{1-\alpha}(x,y)\bar{\phi}^{\alpha} \circ \bar{V}(x,y) &\le \mathsf{c}^{1-\alpha}(x,y)\left[\bar{\phi} \circ V(y) + (V(x)-1)\bar{\phi}'(1)\right]^{\alpha} \\ &\le \vartheta \mathsf{c}^{1-\alpha}(x,y)\left[V^{\alpha}(x) + \phi^{\alpha} \circ V(y)\right] .\end{aligned}$$

This concludes the proof. □

Remark 20.5.4. Assume that one of the conditions of Remark 20.5.3 holds and choose $\psi_1(u) = u$ and $\psi_2 \equiv 1$ in Theorem 20.5.2. Then for every $\alpha \in (0,1)$, there exists a finite constant ϑ such that for all $n \in \mathbb{N}$ and $x \in \mathsf{X}$,

$$r_{\bar{\phi}}^{\alpha}(n)\mathbf{W}_{\mathsf{c}^{1-\alpha}}(P^n(x,\cdot),\pi) \le \vartheta V(x) .$$

This rate of convergence can be improved when $\phi(u) = u^{\alpha_0}$ for some $\alpha_0 \in (0,1)$. In that case, we have $\pi(V^{\alpha_0}) < \infty$, $r_{\bar{\phi}}(n) = \phi \circ H_{\bar{\phi}}^{-1}(n) = \left[H_{\bar{\phi}}^{-1}(n)\right]^{\alpha_0} = O(n^{\alpha_0/(1-\alpha_0)})$. Thus if c is bounded, (20.5.3) yields

$$\mathbf{W}_{\mathsf{c}^{1-\alpha_0}}(P^n(x,\cdot),\pi) \le \vartheta V^{\alpha_0}(x) n^{-\alpha_0/(1-\alpha_0)} .$$

If $c \le \bar{V}$ and $\pi(V) < \infty$, then (20.5.3) yields

$$\mathbf{W}_{\mathsf{c}^{1-\alpha_0}}(P^n(x,\cdot),\pi) \le \vartheta V(x) n^{-\alpha_0/(1-\alpha_0)} .$$

20.6 Exercises

20.1. Consider $\mathbb{R}$ equipped with the Euclidean distance.

1. Consider the sequence $\{\mu_n,\ n \in \mathbb{N}\}$ of probabilities on $\mathbb{N}$ such that $\mu_n(\{0\}) = 1 - n^{-1} = 1 - \mu_n(\{n\})$. Show that $\{\mu_n,\ n \in \mathbb{N}\}$ converges to the point mass at $\{0\}$ in total variation, but not in the Wasserstein distance.
2. Consider the sequence $\{\mu_n,\ n \in \mathbb{N}\}$ of probability distributions on $[0,1]$ with density $1 + \sin(2\pi n x)$. Show that $\{\nu_n,\ n \in \mathbb{N}\}$ converges to the uniform distribution on $[0,1]$ in the Wasserstein distance but not in total variation.

20.2. Let $\xi_1, \xi_1', \xi_2, \xi_2' \in \mathbb{M}_1(\mathscr{X})$ and $\alpha \in (0,1)$. Show that

$$\begin{aligned}&\mathbf{W}_{\mathsf{d},p}^{p}\left(\alpha\xi_1 + (1-\alpha)\xi_2, \alpha\xi_1' + (1-\alpha)\xi_2'\right) \\ &\qquad \le \alpha \mathbf{W}_{\mathsf{d},p}^{p}\left(\xi_1,\xi_1'\right) + (1-\alpha)\mathbf{W}_{\mathsf{d},p}^{p}\left(\xi_2,\xi_2'\right) . \qquad (20.6.1)\end{aligned}$$

20.3. Let $\{\mu_n,\, n \in \mathbb{N}\}$ and $\{\nu_n,\, n \in \mathbb{N}\}$ be two sequences of probability measures in $\mathbb{S}_p(\mathsf{X},\mathsf{d})$ such that $\lim_{n\to\infty} \mathbf{W}_{\mathsf{d},p}(\mu_n,\mu_0) = \lim_{n\to\infty} \mathbf{W}_{\mathsf{d},p}(\nu_n,\nu_0) = 0$. Let $\{\gamma_n,\, n \in \mathbb{N}^*\}$ be a sequence of optimal couplings of (μ_n,ν_n).

1. Prove that $\{\gamma_n,\, n \in \mathbb{N}^*\}$ is tight.
2. Prove that weak limits along subsequences are optimal couplings of (μ_0,ν_0).

20.4. Let $\mathbb{H}$ be the Hilbert space of square summable sequences:

$$\mathbb{H} = \left\{ u \in \mathbb{R}^{\mathbb{N}} : \textstyle\sum_{n=0}^{\infty} u_n^2 < \infty \right\} .$$

Let $\alpha \in (0,1)$ and let $\boldsymbol{\Phi}$ be the linear operator defined on $\mathbb{H}$ by

$$\boldsymbol{\Phi}(u_0,u_1,\dots) = (0,\alpha u_0,\alpha u_1,\dots) .$$

Let $\{Z_n,\, n \in \mathbb{N}^*\}$ be a sequence of i.i.d. real-valued random variables and define

$$\mathbb{Z}_n = (Z_n,0,0\dots),$$

that is, $\mathbb{Z}$ is a random sequence whose first term is Z_n and all other are equal to zero. Define now the sequence $\{\mathbb{X}_n,\, n \in \mathbb{N}\}$ by $\mathbb{X}_0$, independent of $\{\mathbb{Z}_n\}$ and

$$\mathbb{X}_{n+1} = \boldsymbol{\Phi}\mathbb{X}_n + \mathbb{Z}_{n+1} .$$

Let θ be the shift operator, i.e., $\theta(u_0,u_1,\dots) = (u_1,u_2,\dots)$.

1. Prove that $\mathbb{X}_n = \boldsymbol{\Phi}^n\mathbb{X}_0 + \sum_{k=1}^{n} \boldsymbol{\Phi}^{n-k}\mathbb{Z}_k$ and $\mathbb{X}_0 = \alpha^{-n}\theta^n\mathbb{X}_n$.
2. Prove that that the kernel is not irreducible.
3. Prove that $\Delta_{\mathsf{d},p}(P) \le \alpha$ for all $p \ge 1$.

20.5. Consider the count model $\{X_k\}$ introduced in Example 2.2.5 that satisfies the iterative representation (2.2.10). Prove that (20.3.10) holds if $|b|+|c|<1$. [Hint: (Prove and) use the inequality

$$\mathbb{E}\left[\log\left(\frac{1+N(\mathrm{e}^y)}{1+N(\mathrm{e}^x)}\right)\right] \le y-x, \tag{20.6.2}$$

where N is a homogeneous Poisson process on $\mathbb{R}$ and $x \le y$.]

20.6. Consider the functional autoregressive process $\{X_n,\, n \in \mathbb{N}\}$ defined by X_0 and the recurrence

$$X_{n+1} = g(X_n) + Z_{n+1} ,$$

where $\{Z_n,\, n \in \mathbb{N}\}$ is a sequence of i.i.d. random vectors in $\mathbb{R}^d$, independent of X_0, and $g : \mathbb{R}^d \to \mathbb{R}^d$ is a locally bounded measurable function. For $x,y \in \mathbb{R}^d$, define $\mathsf{d}(x,y) = |x-y|$. Denote by P the Markov kernel associated with this Markov chain. Assume the following:

(a) There exists $a > 0$ such that $\mathbb{E}\left[e^{a|Z_0|}\right] < \infty$.
(b) For every compact set K of $\mathbb{R}^d$, there exists $\varepsilon_K > 0$ such that $|g(x) - g(y)| \leq (1 - \varepsilon_K)|x - y|$.
(c) There exists a measurable function $h : \mathbb{R}^d \to [0,1]$ such that $\lim_{|x|\to\infty} |x|h(x) = \infty$ and $|g(x)| \leq |x|(1 - h(x))$ for large enough x.

1. Show that for all $x, x' \in \mathbb{R}^d$, $\mathbf{W}_{\mathsf{d}}(P(x,\cdot), P(x',\cdot)) \leq \mathsf{d}(x,x')$.
2. Let K be a compact set. Show that $K \times K$ is a $(\mathsf{d}, 1, \varepsilon)$-contracting set.
3. Set $V(x) = \{a^{-1} \vee 1\}e^{a|x|}$. Show that there exist $A > 0$, $\lambda < 1$, and $b < \infty$ such that for all $|x| \geq A$, $PV(x) \leq \lambda V(x)$, and for all $x \in \mathbb{R}^d$, $PV(x) \leq \lambda V(x) + b$.

For $\delta > 0$, define

$$\bar{C} = \left\{(x,y) \in \mathbb{R}^d \,:\, V(x) + V(y) \leq 2(b+\delta)/(1-\lambda)\right\} \subset K \times K \,.$$

4. Show that $\bar{C}$ is a $(\mathsf{d}, 1, \varepsilon)$-contracting set.
5. Show that the Markov kernel P has a unique stationary distribution and that there exists $\rho \in [0,1)$ such that for every probability measure ξ, $\mathbf{W}_{\mathsf{d}}(\xi P^n, \pi) \leq \rho^n\{\xi(V) + \pi(V)\}$.

20.7 (A lazy random walk on a discrete cube). For $N \geq 2$, consider the Markov kernel P defined on $\{0,1\}^N$ as follows. If $X_0 = x$, then with probability $1/2$, do nothing, i.e., set $X_1 = x$; with probability $1/2$, choose one coordinate of x at random and flip it. Formally, let $\{B_k,\, k \in \mathbb{N}^*\}$ and $\{I_k,\, k \in \mathbb{N}^*\}$ be independent sequences of i.i.d. random variables such that B_k is a Bernoulli random variable with mean $1/2$ and I_k is uniformly distributed on $\{1,\ldots,N\}$. Let $\oplus$ denote Boolean addition in $\{0,1\}^N$, that is, $1 \oplus 0 = 0 \oplus 1 = 1$ and $0 \oplus 0 = 1 \oplus 1 = 0$ if $\mathsf{d} = 1$, and extend this operation componentwise if $d > 1$. Finally, for $i = 1,\ldots,n$, let e_i be the ith basis vector, with a single component equal to 1 in the ith position and all other components equal to 0. The sequence $\{X_n,\, n \in \mathbb{N}\}$ satisfies the recurrence

$$X_n = F(X_{n-1}; B_n, I_n)\ ,\ n \geq 1\ , \tag{20.6.3}$$

where F is defined on $\{0,1\}^N \times \{0,1\} \times \{1,\ldots,N\}$ by

$$F(x, \varepsilon, i) = x \oplus \varepsilon e_i \,.$$

Let d be the Hamming distance on $\{0,1\}^N$, that is, $\mathsf{d}(x,y)$ is the number of different coordinates in x and y, i.e., $\mathsf{d}(x,y) = \sum_{i=1}^N \mathbb{1}_{\{x \neq y\}}$.

1. Show that the function F is an isometry with respect to x.

Therefore, the simple coupling used in Example 20.3.5 fails here. Using a different coupling, we can prove that the d-Dobrushin coefficient $\Delta_{\mathsf{d}}(P)$ is less than 1. For $x, x' \in \{0,1\}^N$, define

$$(X_1, X_1') = \left(x \oplus B_1 e_{I_1}, x' \oplus B_1 e_{I_1} \mathbb{1}_{\{x_{I_1} = x'_{I_1}\}} + x' \oplus (1 - B_1) e_{I_1} \mathbb{1}_{\{x_{I_1} \neq x'_{I_1}\}}\right) \,.$$

2. Show that (X_1, X_1') is a coupling of $P(x, \dot{)}$ and $P(x', \cdot)$.
3. Show that $\Delta_{\mathsf{d}}(P) \leq 1 - 1/N$.
[Hint: Show that $\mathbb{P}(x_{I_1} = x'_{I_1}) = 1 - \mathsf{d}(x, x')/N$.]

In Exercises 20.8–20.10 we give an alternative proof of Theorem 20.4.5 that uses coupling and that is very close to the proof of Theorem 19.4.1. Let P be a Markov kernel on a complete separable metric space (X, d). Assume that $\mathbf{W}_{\mathsf{d},p}(P(x,\cdot), P(x',\cdot)) \leq \mathsf{d}(x,x')$ for all $x, x' \in \mathsf{X}$ and that there exist a measurable function $V : \mathsf{X} \to [1, \infty)$, $\lambda \in (0,1)$, and $b > 0$ such that condition $\mathrm{D}(V, \lambda, b)$ holds and for all $x, y \in \mathsf{X}$,

$$\mathsf{d}^p(x,y) \leq V(x) + V(y) \; . \tag{20.6.4}$$

Assume, moreover, that there exists $\delta > 0$ such that

$$\bar{C} = \{(x,y) \in \mathsf{X} \times \mathsf{X} : \; V(x) + V(y) \leq 2(b+\delta)/(1-\lambda)\} \tag{20.6.5}$$

is a $(\mathsf{d}^p, 1, \varepsilon)$-contracting set. The assumptions and Theorem 20.1.3 imply that there exists a kernel coupling K of (P, P) such that

$$K\mathsf{d}^p(x,y) \leq \{1 - \varepsilon \mathbb{1}_{\bar{C}}(x,y)\}^p \mathsf{d}^p(x,y) \; . \tag{20.6.6}$$

Let $\{(X_n, X_n'),\ n \in \mathbb{N}\}$ be the coordinate process on the canonical space $(\mathsf{X} \times \mathsf{X})^{\mathbb{N}}$ and let $\mathbb{P}_\gamma$ be the probability measure on the canonical space that makes the coordinate process a Markov chain with kernel K and initial distribution γ. Set $\mathscr{F}_n = \sigma(X_0, X_0', \ldots, X_n, X_n')$.

20.8. 1. Define $Z_n = \mathsf{d}^p(X_n, X_n')$. Show that $\{Z_n,\ n \in \mathbb{N}\}$ is a positive supermartingale.
2. Set $\sigma_{\bar{C}}^{(m)} = \sigma_m$. Show that $\mathbb{E}_\gamma[Z_{\sigma_m}] \leq (1-\varepsilon)^{pm} \mathbb{E}_\gamma[Z_0]$.
3. Let $\eta_n = \sum_{i=0}^n \mathbb{1}_{\bar{C}}(X_i, X_i')$ be the number of visits to the set $\bar{C}$ before time n. Show that for every $n \geq 0$,

$$\mathbb{E}_\gamma[Z_n] \leq (1-\varepsilon)^{pm} \mathbb{E}_\gamma[Z_0] + \mathbb{E}_\gamma[Z_n \mathbb{1}\{\eta_{n-1} < m\}] \; . \tag{20.6.7}$$

Define $\bar{V} : \mathsf{X} \times \mathsf{X} \to [1, \infty]$ by $\bar{V}(x,y) = \{V(x) + V(y)\}/2$, so that we can write $\bar{C} = \{\bar{V} \leq (b+\delta)/(1-\lambda)\}$.

20.9. 1. Show that $K\bar{V} \leq \bar{\lambda} \bar{V} \mathbb{1}_{\bar{C}^c} + \bar{b} \mathbb{1}_{\bar{C}}$. where $\bar{\lambda} = \lambda + b(1-\lambda)/(b+\delta) < 1$ and $\bar{b} = b + \lambda(b+\delta)/(1-\lambda) \geq 1$.
2. Define the sequence $\{S_n,\ n \in \mathbb{N}\}$ by $S_0 = \bar{V}(X_0, X_0')$ and for $n \geq 1$,

$$S_n = \bar{\lambda}^{-n+\eta_{n-1}} \bar{b}^{-\eta_{n-1}} \bar{V}(X_n, X_n') \; , \tag{20.6.8}$$

with the convention $\eta_{-1} = 0$. Show that $\{S_n,\ n \in \mathbb{N}\}$ is a positive supermartingale.
3. Show that $\mathbb{E}_\gamma[Z_n \mathbb{1}_{\{\eta_{n-1} < m\}}] \leq 2\bar{\lambda}^{n-m} \bar{b}^m \mathbb{E}_\gamma[S_0]$.

4. Using (20.6.7), show that for all $(x,x') \in \mathsf{X} \times \mathsf{X}$ and $n \in \mathbb{N}$,

$$\mathbb{E}_{x,x'}[\mathsf{d}^p(X_n, X_n')] \le 2\{(1-\varepsilon)^{pm} + \bar{b}^m \bar{\lambda}^{n-m}\}\bar{V}(x,x') . \tag{20.6.9}$$

20.10. 1. Show that there exists $\tau \in (0,1)$ such that for all probability measures ξ, ξ' and $\gamma \in \mathscr{C}(\xi, \xi')$,

$$\mathbf{W}_{\mathsf{d},p}\left(\xi P^n, \xi' P^n\right) \le \mathbb{E}_\gamma[\mathsf{d}^p(X_n, X_n')] \le 2\{\xi(V) + \xi'(V)\}\tau^n . \tag{20.6.10}$$

[Hint: Use (20.6.9) and optimize in m.]

2. Show that there exists a unique invariant probability measure π and that $\pi(V) < \infty$.
3. Show that for every probability measure ξ, $\mathbf{W}_{\mathsf{d},p}(\xi P^n, \pi) \le \tau^n\{\xi(V) + \pi(V)\}$.

20.11. Let $\{d_n,\ n \in \mathbb{N}\}$ be a nondecreasing sequence of metrics on X that are continuous with respect to the topology of X and such that $\lim_{n\to\infty} \mathsf{d}_n(x,y) = \mathbb{1}\{x \ne y\}$. Prove that for all $\mu, \nu \in \mathbb{M}_1(\mathscr{X})$, $\lim_{n\to\infty} \mathbf{W}_{\mathsf{d}_n}(\mu, \nu) = \mathsf{d}_{\mathrm{TV}}(\mu, \nu)$.

20.12 (Asymptotically ultra-Feller kernels). Let X be a complete separable metric space. A Markov kernel P on $\mathsf{X} \times \mathscr{X}$ is said to be asymptotically ultra-Feller if there exist an increasing sequence $\{n_k,\ k \in \mathbb{N}\}$ and a nondecreasing sequence of metrics $\{\mathsf{d}_n,\ n \in \mathbb{N}\}$, continuous with respect to the topology of X, such that $\lim_{n\to\infty} \mathsf{d}_n(x,y) = \mathbb{1}_{\{x \ne y\}}$ and for all $x^* \in \mathsf{X}$,

$$\inf_{A \in \mathscr{V}_{x^*}} \limsup_{k\to\infty} \sup_{x \in A} \mathbf{W}_{\mathsf{d}_k}\left(P^{n_k}(x,\cdot), P^{n_k}(x^*,\cdot)\right) = 0 . \tag{20.6.11}$$

Let P be an asymptotically ultra-Feller Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting a reachable point x^*. Assume that the Markov kernel P admits two distinct invariant probability measures.

1. Show that without loss of generality, the two invariant probabilities μ and ν may be chosen to be mutually singular.
2. Show that $\mu(A) > 0$ and $\nu(A) > 0$ for all $A \in \mathscr{V}_{x^*}$.
3. For all $\varepsilon > 0$, show that there exists a set $A \in \mathscr{V}_{x^*}$ such that

$$\lim_{k\to\infty} \sup_{x,x' \in A} \mathbf{W}_{\mathsf{d}_k}\left(P^{n_k}(x,\cdot), P^{n_k}(x',\cdot)\right) \le \varepsilon . \tag{20.6.12}$$

Set $\alpha = \mu(A) \wedge \nu(A) \in (0,1]$. Define the probability measures μ_A and ν_A by $\mu_A(B) = [\mu(A)]^{-1}\mu(A \cap B)$, $\nu_A(B) = [\nu(A)]^{-1}\nu(A \cap B)$, $B \in \mathscr{X}$. Finally, define the probability measures $\bar{\mu}$ and $\bar{\nu}$ by $\mu = (1-\alpha)\bar{\mu} + \alpha\mu_A$ and $\nu = (1-\alpha)\bar{\nu} + \alpha\nu_A$ (if $\alpha = 1$, then $\bar{\mu}$ and $\bar{\nu}$ may be chosen arbitrarily).

4. Show that

$$\mathbf{W}_{\mathsf{d}}(\mu, \nu) \le \mathbf{W}_{\mathsf{d}}(\mu P^n, \nu P^n) \le 1 - \alpha + \alpha \sup_{(x,y) \in A \times A} \mathbf{W}_{\mathsf{d}}\left(P^n(x,\cdot), P^n(y,\cdot)\right) . \tag{20.6.13}$$

[Hint: Use Exercise 20.2.]

5. Show that P admits at most one invariant probability. [Hint: Use Exercise 20.11.]

Remark 20.6.1. The notion of asymptotically ultra-Feller kernels extends that of ultra-Feller kernels. A kernel is ultra-Feller at x^* if $\lim_{x \to x^*} \|P(x,\cdot) - P(x^*,\cdot)\|_{\mathrm{TV}} = 0$. If the kernel P is ultra-Feller, then it is also asymptotically ultra-Feller. Indeed, choose $n_k = 1$ for all $k \in \mathbb{N}$ and $\mathsf{d}_n(x,y) = \mathbb{1}_{\{x \neq y\}}$ for all $n \in \mathbb{N}$. Then since P is ultra-Feller, we have $\lim_{x \to x^*} \|P(x,\cdot) - P(x^*,\cdot)\|_{\mathrm{TV}} = 0$, so that

$$\inf_{A \in \mathscr{V}_{x^*}} \limsup_{k \to \infty} \sup_{x \in A} \|P^{n_k}(x,\cdot) - P^{n_k}(x^*,\cdot)\|_{\mathrm{TV}} = 0 \,.$$

This implies that P satisfies (20.6.11), and P is thus asymptotically ultra-Feller.

20.7 Bibliographical Notes

The Monge–Kantorovich problem has undergone many recent developments, and its use in probability theory has been extremely successful. The monograph Rachev and Rüschendorf (1998) is devoted to various types of Monge–Kantorovich mass transportation problems with applications. The classical theory of optimal mass transportation is given in Ambrosio (2003) and Bogachev and Kolesnikov (2012) and Ambrosio and Gigli (2013). Villani (2009) provide an impressive number of results on optimal transport, the geometry of Wasserstein's space, and its applications in probability.

Theorem 20.1.2 and Theorem 20.1.3 are established in (Villani (2009), Theorem 5.10 and Corollary 5.22). This statement can also be found in essentially the same form in (Rachev and Rüschendorf (1998), Chapter 3) and (Dudley (2002), Chapter 11).

Geometric convergence results are adapted from Hairer et al. (2011). The coupling proof introduced in Exercises 20.8 and 20.10 is adapted from the work of Durmus and Moulines (2015). Earlier convergence results are reported in Gibbs (2004) and Madras and Sezer (2010).

Subgeometric convergence in Wasserstein distance was studied in the works of Butkovsky and Veretennikov (2013) and Butkovsky (2014); these results were later improved by Durmus et al. (2016).

This chapter provides only a very quick introduction to the numerous uses of optimal transport and Wasserstein's spaces to Markov chains. There are many notable omissions. Ollivier (2009) (see also Ollivier (2010)) define the Ricci curvature on a metric space in terms of the Wasserstein distance with respect to the underlying distance. This notion is closely connected to our definition of c-Dobrushin coefficient (Definition 20.3.1). Joulin and Ollivier (2010) presents a detailed analysis of nonasymptotic error estimates using the coarse Ricci curvature. Ideas closely related to those developed in this chapter have been developed for continuous-time Markov processes; see, for example, Guillin et al. (2009) and Cattiaux and Guillin (2014).

20.A Complements on the Wasserstein Distance

In this section, we complement and prove some results of Section 20.1.

Theorem 20.A.1. $\mathbf{W}_{\mathsf{d},p}$ *is a distance on the Wasserstein space* $\mathbb{S}_p(\mathsf{X},\mathsf{d})$.

Proof. If $\xi = \xi'$, then $\mathbf{W}_{\mathsf{d},p}(\xi,\xi') = 0$, since we can choose the diagonal coupling, that is, γ is the distribution of (X,X) where X has distribution ξ. Conversely, if $\mathbf{W}_{\mathsf{d},p}(\xi,\xi') = 0$, then there exists a pair of random variables (X,X') defined on a probability space $(\Omega,\mathscr{F},\mathbb{P})$ with marginal distribution ξ and ξ' such that $\mathbb{E}[\mathsf{d}^p(X,X')] = 0$, which implies $X = X'$ $\mathbb{P}-$a.s., hence $\xi = \xi'$.

Since $\mathbf{W}_{\mathsf{d},p}(\xi,\xi') = \mathbf{W}_{\mathsf{d},p}(\xi',\xi)$ obviously holds, the proof will be completed if we prove the triangle inequality. Let $\varepsilon > 0$ and $\mu_1,\mu_2,\mu_3 \in \mathbb{S}_p(\mathsf{X},\mathsf{d})$. By definition, there exist $\gamma_1 \in \mathscr{C}(\mu_1,\mu_2)$ and $\gamma_2 \in \mathscr{C}(\mu_2,\mu_3)$ such that

$$\left\{\int_{\mathsf{X}\times\mathsf{X}} \mathsf{d}^p(x,y)\gamma_1(\mathrm{d}x\mathrm{d}y)\right\}^{1/p} \leq \mathbf{W}_{\mathsf{d},p}(\mu_1,\mu_2) + \varepsilon \,,$$
$$\left\{\int_{\mathsf{X}\times\mathsf{X}} \mathsf{d}^p(y,z)\gamma_2(\mathrm{d}y\mathrm{d}z)\right\}^{1/p} \leq \mathbf{W}_{\mathsf{d},p}(\mu_2,\mu_3) + \varepsilon \,.$$

By the gluing lemma, Lemma B.3.12 (which assumes that X is a Polish space), we can choose (Z_1,Z_2,Z_3) such that $\mathscr{L}_{\mathbb{P}}(Z_1,Z_2) = \gamma_1$ and $\mathscr{L}_{\mathbb{P}}(Z_2,Z_3) = \gamma_2$. This implies that $\mathscr{L}_{\mathbb{P}}(Z_1) = \mu_1$ and $\mathscr{L}_{\mathbb{P}}(Z_3) = \mu_3$. Thus

$$\begin{aligned}
\mathbf{W}_{\mathsf{d},p}(\mu_1,\mu_3) &\leq (\mathbb{E}[\mathsf{d}^p(Z_1,Z_3)])^{1/p} \leq (\mathbb{E}[\mathsf{d}^p(Z_1,Z_2)])^{1/p} + (\mathbb{E}[\mathsf{d}^p(Z_2,Z_3)])^{1/p} \\
&= \left\{\int_{\mathsf{X}\times\mathsf{X}} \mathsf{d}^p(x,y)\gamma_1(\mathrm{d}x\mathrm{d}y)\right\}^{1/p} + \left\{\int_{\mathsf{X}\times\mathsf{X}} \mathsf{d}^p(y,z)\gamma_2(\mathrm{d}y\mathrm{d}z)\right\}^{1/p} \\
&\leq \mathbf{W}_{\mathsf{d}}(\mu_1,\mu_2) + \mathbf{W}_{\mathsf{d}}(\mu_2,\mu_3) + 2\varepsilon \,.
\end{aligned}$$

Since ε is arbitrary, the triangle inequality holds. □

The following result relates the Wasserstein distance and the Prokhorov metric $\boldsymbol{\rho}_{\mathsf{d}}$ and shows that convergence in the Wasserstein distance implies weak convergence.

Proposition 20.A.2 *Let* μ, ν *be two probability measures on* X. *Then*

$$\boldsymbol{\rho}_{\mathsf{d}}^2(\mu,\nu) \leq \mathbf{W}_{\mathsf{d}}(\mu,\nu) \,. \tag{20.A.1}$$

Let $\{\mu_n,\ n \in \mathbb{N}\}$ *be a sequence of probability measures on* X. *For* $p \geq 1$, *if* $\lim_{n\to\infty} \mathbf{W}_{\mathsf{d},p}(\mu_n,\mu) = 0$, *then* $\{\mu_n,\ n \in \mathbb{N}\}$ *converges weakly to* μ.

Proof. Without loss of generality, we assume that $\mathbf{W}_{\mathsf{d}}(\mu,\nu) < \infty$ and set $a = \sqrt{\mathbf{W}_{\mathsf{d}}(\mu,\nu)}$. For $A \in \mathscr{X}$, define $f_a(x) = (1 - \mathsf{d}(x,A)/a)^+$ and let A^a be the a-enlargement of A. Then $\mathbb{1}_A \leq f_a \leq \mathbb{1}_{A^a}$ and $|f_a(x) - f_a(y)| \leq \mathsf{d}(x,y)/a$ for all $(x,y) \in \mathsf{X} \times \mathsf{X}$. Let γ be the optimal coupling of μ and ν. This yields

$$\begin{aligned}\nu(A) &\leq \nu(f_a) \leq \mu(f_a) + \int_{\mathsf{X}\times\mathsf{X}} |f_a(x) - f_a(y)|\gamma(\mathrm{d}x\mathrm{d}y) \\ &\leq \mu(A^a) + a^{-1}\mathbf{W}_{\mathsf{d}}(\mu,\nu) \leq \mu(A^a) + a\,.\end{aligned}$$

By definition of the Prokhorov metric, this proves that $\boldsymbol{\rho}_{\mathsf{d}}(\mu,\nu) \leq a$ and hence (20.A.1) by the choice of a. Since the Prokhorov metric metrizes weak convergence by Theorem C.2.7 and $\mathbf{W}_{\mathsf{d}} \leq \mathbf{W}_{d,p}$ for all $p \geq 1$ by (20.1.14), we obtain that convergence with respect to the Wasserstein distance implies weak convergence. □

Proof (of Theorem 20.1.8). Let $\{\mu_n,\ n \in \mathbb{N}\}$ be a Cauchy sequence for $\mathbf{W}_{\mathsf{d},p}$. By Proposition 20.A.2, it is also a Cauchy sequence for the Prokhorov metric, and by Theorem C.2.7, there exists a probability measure μ such that $\mu_n \stackrel{\mathrm{w}}{\Rightarrow} \mu$. We must prove that $\mu \in \mathbb{S}_p(\mathsf{X},\mathsf{d})$. Fix $x_0 \in \mathsf{X}$. For every $M > 0$, the function $x \mapsto \mathsf{d}(x_0,x) \wedge M$ is continuous. Thus, there exists N such that

$$\begin{aligned}\int_{\mathsf{X}} (\mathsf{d}^p(x_0,x) \wedge M)\mu(\mathrm{d}x) &\leq \int_{\mathsf{X}} (\mathsf{d}^p(x_0,x) \wedge M)\mu_N(\mathrm{d}x) + 1 \\ &\leq \int_{\mathsf{X}} \mathsf{d}^p(x_0,x)\mu_N(\mathrm{d}x) + 1 < \infty\,.\end{aligned}$$

By the monotone convergence theorem, this proves that $\mu \in \mathbb{S}_p(\mathsf{X},\mathsf{d})$, and thus $(\mathbb{S}_p(\mathsf{X},\mathsf{d}), \mathbf{W}_{\mathsf{d},p})$ is complete.

We now prove the density of the distributions with finite support. Fix an arbitrary $a_0 \in \mathsf{X}$. For all $n \geq 1$, by Lemma B.1.3, there exists a partition $\{A_{n,k},\ k \geq 1\}$ of X by Borel sets such that $\mathrm{diam}(A_{n,k}) \leq 1/n$ for all k. Choose now, for each $n,k \geq 1$, a point $a_{n,k} \in A_{n,k}$. Set $B_{n,k} = \bigcup_{j=1}^k A_{n,j}$. Then $B_{n,k}^c$ is a decreasing sequence of Borel sets and $\bigcap_{k\geq 0} B_{n,k}^c = \emptyset$. Let $\mu \in \mathbb{S}_p(\mathsf{X},\mathsf{d})$. Then by dominated convergence, $\lim_{k\to\infty} \int_{B_{n,k}^c} \mathsf{d}^p(a_0,x)\mu(\mathrm{d}x) = 0$. We may thus choose k_0 large enough that $\int_{B_{n,k_0}^c} \mathsf{d}^p(a_0,x)\mu(\mathrm{d}x) < 1/n$. Let X be a random variable with distribution μ. Define the random variable Y_n by

$$Y_n = a_0 \mathbb{1}_{B_{n,k_0}^c}(X) + \sum_{j=1}^{k_0} a_{n,j} \mathbb{1}_{A_{n,j}}(X)\,.$$

Let ν_n be the distribution of Y_n. Then

$$
\begin{aligned}
\mathbf{W}^p_{\mathsf{d},p}(\mu,\nu_n) &\le \mathbb{E}\left[\mathsf{d}^p(X,Y_n)\right] \\
&= \sum_{j=1}^{k_0} \mathbb{E}\left[\mathsf{d}^p(X,Y_n)\mathbb{1}_{A_{n,j}}(X)\right] + \mathbb{E}\left[\mathsf{d}^p(X,Y_n)\mathbb{1}_{B^c_{n,k_0}}(X)\right] \\
&\le \frac{1}{n^p}\sum_{j=1}^{k_0} \mathbb{P}(X\in A_{n,j}) + \int_{B^c_k} \mathsf{d}^p(a_0,x)\mu(\mathrm{d}x) \le 2/n \,.
\end{aligned}
$$

This proves that the set of probability measures that are finite convex combinations of the Dirac measures δ_{a_0} and $\delta_{a_{n,k}}$, $n,k\ge 1$, are dense in $\mathbb{S}_p(\mathsf{X},\mathsf{d})$. Restricting to combinations with rational weights proves that $\mathbb{S}_p(\mathsf{X},\mathsf{d})$ is separable.

Assume now that (i) holds. Then $\mu_n \stackrel{\mathrm{w}}{\Rightarrow} \mu_0$ by Proposition 20.A.2. Applying (20.1.15) and the triangle inequality, we obtain

$$
\begin{aligned}
\limsup_{n\to\infty}\int_{\mathsf{X}} \mathsf{d}^p(x_0,y)\mu_n(\mathrm{d}y) &= \limsup_{n\to\infty} \mathbf{W}_{\mathsf{d},p}\left(\delta_{x_0},\mu_n\right) \\
&\le \mathbf{W}_{\mathsf{d},p}\left(\delta_{x_0},\mu_0\right) + \lim_{n\to\infty}\mathbf{W}_{\mathsf{d},p}(\mu_n,\mu_0) \\
&= \int_{\mathsf{X}} \mathsf{d}^p(x_0,y)\mu_0(\mathrm{d}y)\,.
\end{aligned}
$$

Since $\mu_n \stackrel{\mathrm{w}}{\Rightarrow} \mu_0$, it follows that

$$
\lim_{n\to\infty}\int_{\mathsf{X}} \mathsf{d}^p(x_0,y)\mathbb{1}\left\{\mathsf{d}(x_0,y)\le M\right\}\mu_n(\mathrm{d}y) = \int_{\mathsf{X}} \mathsf{d}^p(x_0,y)\mathbb{1}\left\{\mathsf{d}(x_0,y)|\le M\right\}\mu_n(\mathrm{d}y)
$$

for all M such that $\mu_0(\{y\in\mathsf{X}:\ \mathsf{d}(x_0,y)=M\})=0$. This proves (ii).

Conversely, if (ii) holds, then by Skorokhod's representation theorem, Theorem B.3.18, there exists a sequence $\{X_n,\ n\in\mathbb{N}^*\}$ of random elements defined on a common probability space $(\Omega,\mathscr{A},\mathbb{P})$ such that the distribution of X_n is μ_n for all $n\in\mathbb{N}$ and $X_n\to X_0$ $\mathbb{P}-$a.s. This yields by Lebesgue's dominated convergence theorem,

$$
\limsup_{n\to\infty}\mathbb{E}\left[\mathsf{d}^p(X_n,X_0)\mathbb{1}\left\{\mathsf{d}(x_0,X_n)\le M\right\}\right]=0\,.
$$

By (ii), we also have

$$
\begin{aligned}
&\lim_{M\to\infty}\limsup_{n\to\infty}\mathbb{E}^{1/p}[\mathsf{d}^p(X_n,X_0)\mathbb{1}\left\{\mathsf{d}(x_0,X_n)>M\right\}] \\
&\qquad \le \lim_{M\to\infty}\limsup_{n\to\infty}\mathbb{E}^{1/p}[\mathsf{d}^p(X_n,x_0)\mathbb{1}\left\{\mathsf{d}(x_0,X_n)>M\right\}] \\
&\qquad\qquad + \lim_{M\to\infty}\limsup_{n\to\infty}\mathbb{E}^{1/p}[\mathsf{d}^p(x_0,X_0)\mathbb{1}\left\{\mathsf{d}(x_0,X_n)>M\right\}]=0\,.
\end{aligned}
$$

Altogether, we have shown that

$$
\lim_{n\to\infty}\mathbf{W}^p_{\mathsf{d},p}(\mu_n,\mu_0)\le\lim_{n\to\infty}\mathbb{E}\left[\mathsf{d}^p(X_n,X_0)\right]=0\,.
$$

This proves (i).

□

Chapter 21
Central Limit Theorems

Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ that admits an invariant probability measure π and let $\{X_n,\ n \in \mathbb{N}\}$ be the canonical Markov chain. Given a function $h \in \mathrm{L}^2_0(\pi) = \left\{h \in \mathrm{L}^2(\pi)\ :\ \pi(h) = 0\right\}$, consider the partial sum

$$S_n(h) = \sum_{k=0}^{n-1} h(X_k)\ .$$

For an initial distribution $\xi \in \mathbb{M}_1(\mathscr{X})$, we say that the central limit theorem (CLT) holds for h under $\mathbb{P}_\xi$ if there exists a positive constant $\sigma^2(h)$ such that

$$n^{-1/2} S_n(h) \stackrel{\mathbb{P}_\xi}{\Longrightarrow} \mathrm{N}(0, \sigma^2(h))\ .$$

In this chapter, we will prove the CLT under several sets of conditions, under both stationary distributions $\mathbb{P}_\pi$ and $\mathbb{P}_\xi$ for certain initial distributions ξ. A fruitful approach to obtain a central limit theorem is to represent the sum $S_n(h)$ as the sum of a martingale and a reminder term and to apply a central limit theorem for martingales. In the preliminary Section 21.1, we will prove a generalization of the martingale CLT. This CLT is proved for stationary Markov chains. In Section 21.1.2, we will give a condition under which the central limit holds when the chain does not start under the invariant distribution.

A first method to obtain the martingale decomposition is to use the Poisson equation, which will be introduced in Section 21.2. The Poisson equation plays an important role in many areas, and its use is not limited to the CLT.

When the Poisson equation does not admit a solution, other approaches must be considered. We will discuss two possible approaches in the vast literature dedicated to the central limit theorems for Markov chains (and more generally for dependent sequences).

The first idea is to replace the Poisson equation by the resolvent equation, which always has a solution in $\mathrm{L}^2(\pi)$. Based on the solution to the resolvent equation, we will establish in Section 21.3 a CLT under a set of conditions that are close to optimal.

R. Douc et al., *Markov Chains*, Springer Series in Operations Research
and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_21

The second approach, which will be developed in Section 21.4, is based on a different martingale decomposition. This technique, which was initially developed for stationary weakly dependent sequences, yields a CLT, Theorem 21.4.1, under a single imprimitive sufficient condition. This condition, in turn, provides optimal conditions for geometrically and polynomially ergodic irreducible chains. In the final section, Section 21.4.2, we will apply Theorem 21.4.1 to nonirreducible kernels for which convergence holds in the Wasserstein distance. For these kernels, the functions h considered must satisfy additional Lipschitz-type conditions.

21.1 Preliminaries

In this section, we establish a version of the martingale central limit theorem and other results that will be used to prove the CLT for Markov chains.

21.1.1 Application of the Martingale Central Limit Theorem

We first state an auxiliary central limit theorem for martingales whose increments are functions of a Markov chain. For $m \geq 1$, let π_m be the joint distribution of $(X_0, \ldots, X_m)$ under $\mathbb{P}_\pi$, i.e., $\pi_m = \pi \otimes P^{\otimes m}$.

Lemma 21.1.1 *Let P be a Markov kernel that admits a unique invariant probability measure π. Let $G \in \mathrm{L}^2(\pi_m)$. Assume that $\mathbb{E}\left[G(X_0, \ldots, X_m) \,|\, \mathscr{F}_0\right] = 0$ $\mathbb{P}_\pi -$a.s. Then*

$$n^{-1/2} \sum_{k=m}^{n} G(X_{k-m}, \ldots, X_k) \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0, s^2)$$

with

$$s^2 = \mathbb{E}_\pi \left[\left(\sum_{j=0}^{m-1} \mathbb{E}\left[G(X_j, \ldots, X_{j+m}) \,\middle|\, \mathscr{F}_m \right] - \mathbb{E}\left[G(X_j, \ldots, X_{j+m}) \,\middle|\, \mathscr{F}_{m-1} \right] \right)^2 \right] .$$

If $s^2 = 0$, weak convergence simply means convergence in probability to 0.

Proof. We first express the sum $S_n = \sum_{k=m}^{n} G(X_{k-m}, \ldots, X_k)$ as the sum of a martingale difference sequence and remainder terms. For $k = m, \ldots, n$ and $q = k - m + 1, \ldots, k$, write $Y_k = G(X_{k-m}, \ldots, X_k)$ and

$$\xi_k^{(q)} = \mathbb{E}\left[G(X_{k-m}, \ldots, X_k) \,|\, \mathscr{F}_q\right] - \mathbb{E}\left[G(X_{k-m}, \ldots, X_k) \,|\, \mathscr{F}_{q-1}\right] .$$

Write also $S_n = Y_m + \cdots + Y_n$. Then $\mathbb{E}\left[G(X_{k-m}, \ldots, X_k) \,|\, \mathscr{F}_{k-m}\right] = 0$ $\mathbb{P}_\pi -$a.s. for all $k \geq m$, and we have

$$S_n = \sum_{k=m}^{n} \sum_{q=k-m+1}^{k} \xi_k^{(q)} = \sum_{q=1}^{n} \sum_{k=q}^{q+m-1} \xi_k^{(q)} \mathbb{1}\{m \le k \le n\}$$
$$= \sum_{q=1}^{n} \sum_{j=0}^{m-1} \xi_{q+j}^{(q)} \mathbb{1}\{m \le q+j \le n\} \; .$$

If $m \le q \le n-m+1$, then the indicator is equal to 1 for all $j = 0, \ldots, m-1$, i.e., only the first and last $m-1$ terms are affected by the indicator. Write

$$\zeta_q = \sum_{j=0}^{m-1} \xi_{q+j}^{(q)} \, , \qquad M_n = \sum_{q=m}^{n-m+1} \zeta_q \, .$$

Since $G \in \mathrm{L}^2(\pi_m)$, we may therefore write $S_n = M_n + R_n$, and the sequence $\{R_n\}$ satisfies $\sup_{n\in\mathbb{N}} \mathbb{E}_\pi[R_n^2] < \infty$, since the random variables $\xi_k^{(q)}$ are uniformly bounded in $\mathrm{L}^2(\pi)$, and R_n is a sum of at most $2m$ terms of this form. The sequence $\{\zeta_q, \, q \in \mathbb{N}\}$ is a stationary square-integrable martingale difference sequence. Therefore, to prove the central limit theorem for $\{M_n, \, n \in \mathbb{N}\}$, we apply the central limit theorem for stationary martingale difference sequences (see Corollary E.4.2). We must check the following conditions: there exists $s > 0$ such that

$$n^{-1} \sum_{q=m}^{n-m+1} \mathbb{E}\left[\zeta_q^2 \,\middle|\, \mathscr{F}_{q-1} \right] \stackrel{\mathbb{P}_\pi-\text{prob}}{\longrightarrow} s^2, \tag{21.1.1}$$

and for all $\varepsilon > 0$,

$$n^{-1} \sum_{q=m}^{n-m+1} \mathbb{E}\left[\zeta_q^2 \mathbb{1}\left\{ |\zeta_q| > \varepsilon\sqrt{n} \right\} \,\middle|\, \mathscr{F}_{q-1} \right] \stackrel{\mathbb{P}_\pi-\text{prob}}{\longrightarrow} 0, \tag{21.1.2}$$

By stationarity, the expectation of the left-hand side is $\mathbb{E}_\pi\left[\zeta_m^2 \mathbb{1}\left\{ |\zeta_m| > \varepsilon\sqrt{n} \right\} \right]$. By monotone convergence theorem since $\mathbb{E}_\pi[\xi_m^2] < \infty$, we obtain

$$\lim_{n\to\infty} \mathbb{E}_\pi\left[\zeta_m^2 \mathbb{1}\left\{ |\zeta_m| > \varepsilon\sqrt{n} \right\} \right] = 0 \, .$$

This proves (21.1.2). Now the left-hand side of (21.1.1) might be expressed as $n^{-1} \sum_{q=m}^{n-m+1} H(X_{q-1})$ with

$$H(x) = \mathbb{E}_x\left[\left(\sum_{j=0}^{m-1} \xi_{m+j}^{(m)} \right)^2 \right] \, .$$

Since the invariant probability is unique, we can apply Theorems 5.2.6 and 5.2.9, which yield (21.1.1), since $s^2 = \pi(H)$. □

Remark 21.1.2. For all $g \in \mathrm{L}^2(\pi)$, set $G(x_0, \ldots, x_m) = g(x_m) - P^m g(x_0)$. In that case, the limiting variance takes the simpler form

$$s^2 = \mathbb{E}_\pi \left[\left\{ \sum_{j=0}^{m-1} P^j g(X_1) - P^{j+1} g(X_0) \right\}^2 \right] . \tag{21.1.3}$$

For $m = 1$, this simply yields $s^2 = \mathbb{E}_\pi[\{g(X_1) - Pg(X_0)\}^2]$. ▲

21.1.2 From the Invariant to an Arbitrary Initial Distribution

We will derive below conditions under which the central limit theorems holds under $\mathbb{P}_\pi$ (when the Markov chain is started from its invariant distribution and is therefore stationary), i.e., for some $h \in \mathrm{L}_0^2(\pi)$,

$$n^{-1/2} S_n(h) \overset{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0, \sigma_\pi(h)) . \tag{21.1.4}$$

Assuming that (21.1.4) holds, it is natural to ask for which initial distributions $\xi \in \mathbb{M}_1(\mathscr{X})$ the CLT still holds. Perhaps surprisingly, these conditions are very different from those that ensure the CLT under the stationary distribution. As we will see in this section, it is possible to deduce from (21.1.4) that $n^{-1/2} S_n(h) \overset{\mathbb{P}_\xi}{\Longrightarrow} \mathrm{N}(0, \sigma_\pi(h))$ under quite weak conditions, at least when the Markov kernel P is irreducible. In particular, if P is a positive irreducible Harris recurrent Markov kernel and if the central limit holds under $\mathbb{P}_\pi$, then it holds under every initial distribution.

Proposition 21.1.3 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, $h : \mathsf{X} \to \mathbb{R}$ a measurable function, and $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$. Assume that*

$$\lim_{m \to \infty} \left\| (\xi - \xi') P^m \right\|_{\mathrm{TV}} = 0 . \tag{21.1.5}$$

Then if $n^{-1/2} S_n(h) \overset{\mathbb{P}_\xi}{\Longrightarrow} \mu$ for some probability measure μ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, then $n^{-1/2} S_n(h) \overset{\mathbb{P}_{\xi'}}{\Longrightarrow} \mu$.

Proof. For $n > m$, we get

$$\begin{aligned}
\mathbb{E}_\xi[\exp(\mathrm{i}tn^{-1/2} S_n(h))] &= \mathbb{E}_\xi[\exp(\mathrm{i}tn^{-1/2} S_{n-m}(h) \circ \theta_m)] + r_{m,n}(\xi) \\
&= \mathbb{E}_\xi[\mathbb{E}_{X_m}[\exp(\mathrm{i}tn^{-1/2} S_{n-m}(h))]] + r_{m,n}(\xi) \\
&= \xi P^m u_{m,n} + r_{m,n}(\xi)
\end{aligned}$$

with $u_{m,n}(x) = \mathbb{E}_x[\exp(\mathrm{i}tn^{-1/2} S_{n-m}(h))]$ and $r_{m,n}(\xi) \le \mathbb{E}_\xi[|1 - \exp(\mathrm{i}tn^{-1/2} S_m(h))|]$. For every $m \in \mathbb{N}$, we have $\lim_{n\to\infty} r_{m,n}(\xi) = 0$. Furthermore, since $|u_{m,n}|_\infty \le 1$, it follows that

$$
\begin{aligned}
&|\mathbb{E}_{\xi}[\exp(\mathrm{i}tn^{-1/2}S_n(h))] - \mathbb{E}_{\xi'}[\exp(\mathrm{i}tn^{-1/2}S_n(h))]| \\
&\qquad \le \left\|\xi P^m - \xi' P^m\right\|_{\mathrm{TV}} + r_{m,n}(\xi) + r_{m,n}(\xi') \; .
\end{aligned}
$$

Therefore, $\lim_{n\to\infty} |\mathbb{E}_{\xi}[\exp(\mathrm{i}tn^{-1/2}S_n(h))] - \mathbb{E}_{\xi}[\exp(\mathrm{i}tn^{-1/2}S_n(h))]| = 0$, and the result follows. □

We now replace the condition (21.1.5) with a weaker condition under which the existence of the limit is replaced by a convergence in Cesàro mean. As we will see below, this makes it possible to deal in particular with the case of periodic Markov kernels. We first need a preliminary result that is of independent interest.

Lemma 21.1.4 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. For $h \in \mathrm{L}^2(\pi)$, set $A_\infty(h) := \left\{\lim_{n\to\infty} n^{-1/2}|h(X_n)| = 0\right\}$. Then $\mathbb{P}_\pi(A_\infty(h)) = 1$ and $\mathbb{P}_x(A_\infty(h)) = 1$ for π almost all $x \in \mathsf{X}$. If P is positive, irreducible, and Harris recurrent, then $\mathbb{P}_\xi(A_\infty(h)) = 1$ for all $\xi \in \mathbb{M}_1(\mathscr{X})$.*

Proof. For all $\varepsilon > 0$, we get

$$
\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}_\pi(n^{-1/2}|h(X_n)| > \varepsilon) &= \sum_{n=1}^{\infty} \mathbb{P}_\pi(\varepsilon^{-2}|h(X_n)|^2 > n) \\
&= \sum_{n=1}^{\infty} \pi\left(\{\varepsilon^{-2}|h|^2 > n\}\right) \le \varepsilon^{-2}\pi(h^2) < \infty \; .
\end{aligned}
$$

Therefore, by the Borel–Cantelli lemma we obtain $\mathbb{P}_\pi(A_\infty(h)) = 1$. Set $g(x) = \mathbb{P}_x(A_\infty(h))$. Then

$$
Pg(x) = \mathbb{E}_x[\mathbb{E}_{X_1}[\mathbb{1}_{A_\infty(h)}]] = \mathbb{E}_x[\mathbb{1}_{A_\infty(h)} \circ \theta] = \mathbb{E}_x[\mathbb{1}_{A_\infty(h)}] = g(x) \; .
$$

Therefore, the function g is harmonic and $\pi(g) = \mathbb{P}_\pi(A_\infty(h)) = 1$. This implies that $g(x) = 1$ for π almost all $x \in \mathsf{X}$. If P is a positive, irreducible, and Harris recurrent, then the function g is constant by Theorem 10.2.11, and therefore $g(x) = 1$ for all $x \in \mathsf{X}$, which concludes the proof. □

Proposition 21.1.5 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, $h : \mathsf{X} \to \mathbb{R}$ a measurable function, and $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$. Assume that $n^{-1/2}h(X_n) \overset{\mathbb{P}_\xi-\text{prob}}{\longrightarrow} 0$, $n^{-1/2}h(X_n) \overset{\mathbb{P}_{\xi'}-\text{prob}}{\longrightarrow} 0$, and*

$$
\lim_{m\to\infty} \left\| \frac{1}{m} \sum_{k=0}^{m-1} (\xi - \xi')P^k \right\|_{\mathrm{TV}} = 0 \; . \tag{21.1.6}
$$

Then if $n^{-1/2}S_n(h) \overset{\mathbb{P}_\xi}{\Longrightarrow} \mu$ for some probability measure μ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, then $n^{-1/2}S_n(h) \overset{\mathbb{P}_{\xi'}}{\Longrightarrow} \mu$.

Proof. For $j,k \in \mathbb{N}^2$, set $S_{j,k} = \sum_{i=j}^{k} h(X_i)$, with the convention $S_{j,k} = 0$ if $j > k$. The dependence of $S_{j,k}$ on h is implicit. For all $t \in \mathbb{R}$ and $n > m$, using the Markov property, we get

$$
\begin{aligned}
\mathbb{E}_\xi\left[\exp(\mathrm{i}tn^{-1/2}S_{0,n-1})\right] &= \frac{1}{m}\sum_{k=0}^{m-1}\mathbb{E}_\xi\left[\exp(\mathrm{i}tn^{-1/2}S_{k,n+k-1})\right] + r_{m,n}(\xi) \\
&= \frac{1}{m}\sum_{k=0}^{m-1}\mathbb{E}_{\xi P^k}\left[\exp(\mathrm{i}tn^{-1/2}S_{0,n-1})\right] + r_{m,n}(\xi) \\
&= \frac{1}{m}\sum_{k=0}^{m-1}\xi P^k u_n + r_{m,n}(\xi),
\end{aligned}
$$

where we have set $u_n(x) = \mathbb{E}_x\left[\exp(\mathrm{i}tn^{-1/2}S_{0,n-1})\right]$ and

$$
|r_{m,n}(\xi)| \leq \frac{1}{m}\sum_{k=0}^{m-1}\left\{\mathbb{E}_\xi\left[|1-\exp(\mathrm{i}tn^{-1/2}S_{n,n+k-1})| + |1-\exp(\mathrm{i}tn^{-1/2}S_{0,k-1})|\right]\right\} .
$$

By Lemma 21.1.4, $n^{-1/2}|h(X_n)| \stackrel{\mathbb{P}_\xi-\text{prob}}{\longrightarrow} 0$. This implies for each $k \in \mathbb{N}$,

$$
\lim_{n\to\infty}\mathbb{E}_\xi\left[|1-\exp(\mathrm{i}tn^{-1/2}S_{n,n+k-1})|\right] = 0 ,
$$

and thus for each m, $\lim_{n\to\infty} r_{m,n}(\xi) = 0$. Now since $|u_n|_\infty \leq 1$, we get

$$
\begin{aligned}
&\left|\mathbb{E}_\xi\left[\exp(\mathrm{i}tn^{-1/2}S_{0,n-1})\right] - \mathbb{E}_{\xi'}\left[\exp(\mathrm{i}tn^{-1/2}S_{0,n-1})\right]\right| \\
&\leq \left|\frac{1}{m}\sum_{k=0}^{m-1}\xi P^k u_n - \frac{1}{m}\sum_{k=0}^{m-1}\xi' P^k u_n\right| + |r_{m,n}(\xi)| + |r_{m,n}(\xi')| \\
&\leq \left\|\frac{1}{m}\sum_{k=0}^{m-1}(\xi-\xi')P^k\right\|_{\mathrm{TV}} + |r_{m,n}(\xi)| + |r_{m,n}(\xi')| ,
\end{aligned}
$$

which concludes the proof. □

Corollary 21.1.6 *Let P be an irreducible positive and Harris recurrent kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π and let $h \in \mathrm{L}_0^2(\pi)$. If $n^{-1/2}S_n(h) \stackrel{\mathbb{P}_\xi}{\Longrightarrow} \mathrm{N}(0,\sigma^2(h))$ for some $\xi \in \mathbb{M}_1(\mathscr{X})$, then $n^{-1/2}S_n(h) \stackrel{\mathbb{P}_{\xi'}}{\Longrightarrow} \mathrm{N}(0,\sigma^2(h))$ for all $\xi' \in \mathbb{M}_1(\mathscr{X})$.*

Proof. By Corollary 11.3.2, we get $\lim_{m\to\infty}\left\|m^{-1}\sum_{k=0}^{m-1}(\xi-\xi')P^k\right\|_{\mathrm{TV}} = 0$. The proof is then completed by applying Proposition 21.1.5. □

We now extend this result to an irreducible recurrent positive Markov kernel P. For C an accessible small set, define

$$H = \{x \in \mathsf{X} : \mathbb{P}_x(\sigma_C < \infty) = 1\} . \tag{21.1.7}$$

According to Theorem 10.2.7, the set H does not depend on the choice of the small set C and is maximal absorbing. If we denote by π the unique invariant probability of P, we get $\pi(H) = 1$.

Corollary 21.1.7 *Let P be an irreducible recurrent positive Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π and let $h \in \mathrm{L}^2_0(\pi)$. Let H be given by (21.1.7). Assume that for some $\xi \in \mathbb{M}_1(\mathscr{X})$ satisfying $\xi(H^c) = 0$, we get $n^{-1/2}S_n(h) \overset{\mathbb{P}_\xi}{\Longrightarrow} \mathrm{N}(0, \sigma^2(h))$. Then $n^{-1/2}S_n(h) \overset{\mathbb{P}_{\xi'}}{\Longrightarrow} \mathrm{N}(0, \sigma^2(h))$ for all $\xi' \in \mathbb{M}_1(\mathscr{X})$ satisfying $\xi'(H^c) = 0$.*

Proof. By Theorem 10.2.7, the restriction of P to H is Harris recurrent. Since P is positive, this restriction is irreducible positive and Harris recurrent. We conclude by Corollary 21.1.6. □

We now generalize the results above to a Markov kernel P that admits an invariant probability π but is not necessarily irreducible.

Proposition 21.1.8 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π, $\xi \in \mathbb{M}_1(\mathscr{X})$, and $h \in \mathrm{L}^2_0(\pi)$. Assume that $n^{-1/2}S_n(h) \overset{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0, \sigma^2(h))$. If in addition*

(i) the function h is finite,
(ii) either $\lim_{n\to\infty} \|\xi P^n - \pi\|_{\mathrm{TV}} = 0$ or $\lim_{n\to\infty} \left\| n^{-1} \sum_{k=0}^{n-1} P^k - \pi \right\|_{\mathrm{TV}} = 0$ and $\lim_{n\to\infty} n^{-1/2} h(X_n) = 0$ $\mathbb{P}_\xi$ $-$ a.s.,

then $n^{-1/2}S_n(h) \overset{\mathbb{P}_\xi}{\Longrightarrow} \mathrm{N}(0, \sigma^2(h))$.

Proof. Follows from Proposition 21.1.3 and Proposition 21.1.5. □

21.2 The Poisson Equation

In this section we will prove a central limit theorem for $n^{-1/2}S_n(h)$ using the Poisson equation.

Definition 21.2.1 (Poisson equation) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with a unique invariant probability π. For $h \in \mathbb{F}(\mathsf{X})$ such that $\pi(|h|) < \infty$, the equation*

$$\hat{h} - P\hat{h} = h - \pi(h) \tag{21.2.1}$$

is called the Poisson equation associated with the function h.

A solution to the Poisson equation (21.2.1) is a function $\hat{h} \in \mathbb{F}(\mathsf{X})$ satisfying for π-a.e. $x \in \mathsf{X}$, $P|\hat{h}|(x) < \infty$ and $\hat{h}(x) - P\hat{h}(x) = h(x) - \pi(h)$.

The solution to the Poisson equation allows us to relate $S_n(h)$ to a martingale. If $\hat{h} - P\hat{h} = h$, we have the decomposition

$$S_n(h) = M_n(h) + \hat{h}(X_0) - \hat{h}(X_n) \;, \tag{21.2.2}$$

with

$$M_n(h) = \sum_{k=1}^{n} \left\{ \hat{h}(X_k) - P\hat{h}(X_{k-1}) \right\} \;. \tag{21.2.3}$$

The asymptotic behavior of the sequence $\{S_n(h), n \in \mathbb{N}\}$ will be derived from that of the martingale $\{M_n(h), n \in \mathbb{N}\}$.

Lemma 21.2.2 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. If π is the unique invariant probability measure and $\hat{h}_1$ and $\hat{h}_2$ are solutions to Poisson equations such that $\pi(|\hat{h}_i|) < \infty$, $i = 1, 2$, then there exists $c \in \mathbb{R}$ such that $\hat{h}_2(x) = c + \hat{h}_1(x)$ for π-a.e. $x \in \mathsf{X}$.*

Proof. If $\hat{h}_1$ and $\hat{h}_2$ are two solutions to the Poisson equation, then $\hat{h}_1 - \hat{h}_2 = P(\hat{h}_1 - \hat{h}_2)$, i.e., $\hat{h}_1 - \hat{h}_2$ is harmonic. Since $\pi(|\hat{h}_1 - \hat{h}_2|) < \infty$ and since π is now assumed to be the unique invariant probability measure, Proposition 5.2.12 implies that $\hat{h}_1 - \hat{h}_2$ is π-almost surely constant. □

Proposition 21.2.3 *Let P be a Markov kernel with a unique invariant probability measure π. Let $h \in \mathrm{L}^p(\pi)$ be such that $\pi(h) = 0$. Assume that $\sum_{k=0}^{\infty} \left\| P^k h \right\|_{\mathrm{L}^p(\pi)} < \infty$ for some $p \geq 1$. Then $\hat{h} = \sum_{k=0}^{\infty} P^k h$ is a solution to the Poisson equation and $\hat{h} \in \mathrm{L}^p(\pi)$.*

Proof. Note first that since $\sum_{k=0}^{\infty} \left\| P^k h \right\|_{\mathrm{L}^p(\pi)} < \infty$, the series $\sum_{k=0}^{\infty} P^k h$ is normally convergent in $\mathrm{L}^p(\pi)$. We denote by $\hat{h}$ the sum of this series. Because P is a bounded linear operator on $\mathrm{L}^p(\pi)$, we have $P\hat{h} = \sum_{k=1}^{\infty} P^k h$ and $\hat{h} - P\hat{h} = h$ in $\mathrm{L}^p(\pi)$. □

Proposition 21.2.4 *Let P be an irreducible and aperiodic Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. Assume that there exist $V : \mathsf{X} \to [0,\infty]$, $f : \mathsf{X} \to [1,\infty)$, $b < \infty$, and a nonempty petite set C such that $\sup_C V < \infty$ and $PV + f \le V + b\mathbb{1}_C$.*

Every function $h \in \mathbb{F}(\mathsf{X})$ satisfying $|h|_f < \infty$ is π-integrable, and there exists a solution $\hat{h}$ to the Poisson equation such that $|\hat{h}|_V < \infty$.

Proof. Theorem 17.1.3 (a) shows that the set $\{V < \infty\}$ is full and absorbing. Since π is a maximal irreducibility measure, $\pi(\{V = \infty\}) = 0$, and Proposition 4.3.2 shows that $\pi(f) < \infty$. By Theorem 17.1.3 (c), there exists $\varsigma < \infty$ such that for every $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$,

$$\sum_{n=0}^{\infty} \left\| \xi P^n - \xi' P^n \right\|_f \le \varsigma \left\{ \xi(V) + \xi'(V) + 1 \right\} . \tag{21.2.4}$$

Let $x \in \mathsf{X}$ be such that $V(x) < \infty$, which implies $PV(x) < \infty$. Let h be a function such that $|h|_f < \infty$ and $\pi(h) = 0$. Applying (21.2.4) with $\xi = \delta_x$ and $\xi' = \delta_x P$ yields

$$\begin{aligned} \sum_{n=0}^{\infty} |P^n h(x) - P^{n+1} h(x)| &\le |h|_f \sum_{n=0}^{\infty} \left\| P^n(x,\cdot) - P^{n+1}(x,\cdot) \right\|_f \\ &\le \varsigma \, |h|_f \left\{ V(x) + PV(x) + 1 \right\} < \infty . \end{aligned} \tag{21.2.5}$$

By Theorem 17.1.3 (a), we know that $\lim_{n\to\infty} P^n h(x) = 0$, which implies

$$h(x) = \sum_{n=0}^{\infty} \{ P^n h(x) - P^{n+1} h(x) \} ,$$

for all x such that $V(x) < \infty$. Choose $x_0 \in \{V < \infty\}$. Then (21.2.4) shows that for all $x \in \mathsf{X}$,

$$\sum_{n=0}^{\infty} |P^n h(x) - P^n h(x_0)| \le \varsigma \{ V(x) + V(x_0) + 1 \} . \tag{21.2.6}$$

Consider the function defined on X by

$$\tilde{h}(x) = \sum_{n=0}^{\infty} \{ P^n h(x) - P^n h(x_0) \} \tag{21.2.7}$$

if $V(x) < \infty$ and $\tilde{h}(x) = 0$ otherwise. Then if $V(x) < \infty$, the absolute summability of the series in (21.2.6) yields

$$\tilde{h}(x) - P\tilde{h}(x) = \sum_{n=0}^{\infty} \{ P^n h(x) - P^{n+1} h(x) \} = h(x) .$$

This proves that $\tilde{h}$ is a solution to the Poisson equation. $\square$

Theorem 21.2.5. *Let P be a Markov kernel with a unique invariant probability measure π. Let $h \in \mathrm{L}^2(\pi)$ be such that $\pi(h) = 0$. Assume that there exists a solution $\hat{h} \in \mathrm{L}^2(\pi)$ to the Poisson equation $\hat{h} - P\hat{h} = h$. Then*

$$n^{-1/2} \sum_{k=0}^{n-1} h(X_k) \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0, \sigma_\pi^2(h)) ,$$

where

$$\sigma_\pi^2(h) = \mathbb{E}_\pi[\{\hat{h}(X_1) - P\hat{h}(X_0)\}^2] = 2\pi(h\hat{h}) - \pi(h^2) . \tag{21.2.8}$$

Proof. The sequence $\{M_n(h), n \in \mathbb{N}\}$ defined in (21.2.3) is a martingale under $\mathbb{P}_\pi$ and satisfies the assumptions of Lemma 21.1.1 with $m = 1$ and $G(x,y) = \hat{h}(y) - P\hat{h}(x)$. By Markov's property,

$$\mathbb{E}\left[G(X_{k-1}, X_k) \,|\, \mathscr{F}_{k-1}\right] = \mathbb{E}\left[\hat{h}(X_k) \,\middle|\, X_{k-1}\right] - P\hat{h}(X_{k-1}) = 0 \qquad \mathbb{P}_\pi - \text{a.s.}$$

Lemma 21.1.1 shows that

$$n^{-1/2} M_n(\hat{h}) \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}\left(0, \mathbb{E}_\pi[\{\hat{h}(X_1) - P\hat{h}(X_0)\}^2]\right) .$$

We will now establish (21.2.8). Since the Markov chain $\{X_k,\ k \in \mathbb{N}\}$ is stationary under $\mathbb{P}_\pi$, we get $\mathbb{E}_\pi[|\hat{h}(X_0) + \hat{h}(X_n)|] \le 2\pi(|\hat{h}|)$, which implies that

$$n^{-1/2}\{\hat{h}(X_0) + \hat{h}(X_n)\} \stackrel{\mathbb{P}_\pi\text{ -prob}}{\longrightarrow} 0 .$$

Let us now prove the equality of the expressions (21.2.8) for the variance. Since $P\hat{h}(X_0) = \mathbb{E}\left[\hat{h}(X_1) \,\middle|\, \mathscr{F}_0\right]$, we have $\mathbb{E}_\pi[\hat{h}(X_1) P\hat{h}(X_0)] = \mathbb{E}_\pi[\{P\hat{h}(X_0)\}^2]$ and thus

$$\mathbb{E}_\pi[\{\hat{h}(X_1) - P\hat{h}(X_0)\}^2] = \mathbb{E}_\pi[\{\hat{h}(X_1)\}^2 - \{P\hat{h}(X_0)\}^2] = \pi(\hat{h}^2 - (P\hat{h})^2) .$$

Since $h = \hat{h} - P\hat{h}$, we further have $\hat{h}^2 - (P\hat{h})^2 = (\hat{h} - P\hat{h})(\hat{h} + P\hat{h}) = h(2\hat{h} - h)$. Therefore, $\pi(\hat{h}^2 - (P\hat{h})^2) = 2\pi(h\hat{h}) - \pi(h^2)$. $\square$

Theorem 21.2.6. *Let P be a Markov kernel with a unique invariant probability measure π. Let $h \in \mathrm{L}^2(\pi)$ be such that $\pi(h) = 0$. Assume that $\sum_{k=0}^{\infty} \left\| P^k h \right\|_{\mathrm{L}^2(\pi)} < \infty$. Then $\sum_{k=0}^{\infty} |\pi(hP^k h)| < \infty$ and*

$$n^{-1/2}\sum_{k=0}^{n-1} h(X_k) \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0,\sigma_\pi^2(h)) ,$$

where

$$\sigma_\pi^2(h) = \pi(h^2) + 2\sum_{k=1}^{\infty}\pi(hP^k h) = \lim_{n\to\infty} n^{-1}\mathbb{E}_\pi[S_n^2(h)] . \tag{21.2.9}$$

Proof. By Proposition 21.2.3, the series $\hat{h} = \sum_{k=0}^{\infty} P^k h$ is a solution to the Poisson equation, and $\hat{h} \in \mathrm{L}^2(\pi)$. By the Cauchy–Schwarz inequality, we have

$$\sum_{k=0}^{\infty} |\pi(hP^k k)| \le \|h\|_{\mathrm{L}^2(\pi)} \sum_{k=0}^{\infty} \left\| P^k h \right\|_{\mathrm{L}^2(\pi)} < \infty .$$

We conclude by applying Theorem 21.2.5 using the identity $2\pi(h\hat{h}) - \pi(h^2) = \pi(h^2) + 2\sum_{k=1}^{\infty}\pi(hP^k h)$.

The identity $\sigma_\pi^2(h) = \lim_{n\to\infty} n^{-1}\mathbb{E}_\pi[S_n^2(h)]$ follows from Lemma 21.2.7 below. □

Lemma 21.2.7 *Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$ with invariant probability measure π. Let $h \in \mathrm{L}^2(\pi)$. If the limit*

$$\lim_{n\to\infty} \sum_{k=1}^{n} \pi(hP^k h) \tag{21.2.10}$$

exists in $\mathbb{R}\cup\{+\infty\}$, then

$$\lim_{n\to\infty} n^{-1}\mathbb{E}_\pi\left[\left(\sum_{k=0}^{n-1} h(X_k)\right)^2\right] = \pi(h^2) + 2\lim_{n\to\infty}\sum_{k=1}^{n}\pi(hP^k h) . \tag{21.2.11}$$

Proof. By stationarity, we have

$$\begin{aligned}\frac{1}{n}\mathbb{E}_\pi\left[\left(\sum_{k=0}^{n-1} h(X_k)\right)^2\right] &= \pi(h^2) + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\pi(hP^k h) \\ &= \pi(h^2) + \frac{2}{n}\sum_{\ell=1}^{n}\left\{\sum_{k=1}^{\ell-1}\pi(hP^k h)\right\} .\end{aligned}$$

We conclude the proof by Cesàro's theorem. □

Remark 21.2.8. If the limit in (21.2.10) exists, it is usual to denote it by $\sum_{k=1}^{\infty}\pi(hP^k h)$, but it is important to remember in that case that this notation does not imply that the series is absolutely summable. ▲

We will now illustrate the use of summability condition $\sum_{k=0}^{\infty} \left\| P^k h \right\|_{\mathrm{L}^2(\pi)} < \infty$. The following lemma is instrumental in the sequel.

Lemma 21.2.9 *Let* $(\mathsf{X}, \mathscr{X})$ *be a measurable space,* $(\xi, \xi') \in \mathbb{M}_1(\mathscr{X})$*,* $p \geq 1$*, and* $h \in \mathrm{L}^p(\xi) \cap \mathrm{L}^p(\xi')$*. Then*

$$|\xi(h) - \xi'(h)| \leq \|\xi - \xi'\|_{\mathrm{TV}}^{(p-1)/p} \left\{ \xi(|f|^p) + \xi'(|f|^p) \right\}^{1/p} . \tag{21.2.12}$$

Proof. Without loss of generality, we assume that $\|\xi - \xi'\|_{\mathrm{TV}} \neq 0$. Note first that

$$\begin{aligned} |\xi(h) - \xi'(h)|^p &= \left| \int \{\xi(\mathrm{d}x) - \xi'(\mathrm{d}x)\} h(x) \right|^p \\ &\leq \left(\int \frac{|\xi - \xi'|(\mathrm{d}x)}{\|\xi - \xi'\|_{\mathrm{TV}}} |h(x)| \right)^p \|\xi - \xi'\|_{\mathrm{TV}}^p , \end{aligned}$$

where $|\xi - \xi'|$ denotes the total variation of the finite signed measure $\xi - \xi'$. Since $|\xi - \xi'| / \|\xi - \xi'\|_{\mathrm{TV}}$ is a probability measure, Jensen's inequality implies

$$\begin{aligned} |\xi(h) - \xi'(h)|^p &\leq \|\xi - \xi'\|_{\mathrm{TV}}^p \int \frac{|\xi(\mathrm{d}x) - \xi'(\mathrm{d}x)|}{\|\xi - \xi'\|_{\mathrm{TV}}} |h(x)|^p \\ &\leq \|\xi - \xi'\|_{\mathrm{TV}}^{p-1} \left\{ \xi(|h|^p) + \xi'(|h|^p) \right\} . \end{aligned}$$

The proof of (21.2.12) is complete. □

Theorem 21.2.10. *Let* P *be Markov kernel on* $\mathsf{X} \times \mathscr{X}$ *with invariant probability measure* π*. If the Markov kernel* P *is* π*-a.e. uniformly ergodic, i.e., there exist* $\varsigma < \infty$ *and* $\rho \in [0, 1)$ *such that for* π*-a.e.* $x \in \mathsf{X}_0$*,*

$$\|P^n(x, \cdot) - \pi\|_{\mathrm{TV}} \leq \varsigma \rho^n , \quad \textit{for all } n \in \mathbb{N},$$

then for all $h \in \mathrm{L}_0^2(\pi)$*,* $\sum_{k=0}^{\infty} |\pi(hP^k h)| < \infty$*, we get*

$$n^{-1/2} \sum_{k=0}^{n-1} h(X_k) \overset{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}\left(0, \sigma_\pi^2(h)\right) ,$$

where

$$\sigma_\pi^2(h) = \pi(h^2) + 2 \sum_{k=1}^{\infty} \pi(hP^k h) . \tag{21.2.13}$$

Proof. Let $h \in \mathrm{L}_0^2(\pi)$. Since $\pi(h^2) < \infty$ and π is invariant, we have that for all $n \in \mathbb{N}$, $\pi(P^n h^2) < \infty$, showing that $\pi(\mathsf{X}_0) = 1$, where

$$\mathsf{X}_0 = \bigcap_{n=0}^{\infty} \left\{ x \in \mathsf{X} : P^n h^2(x,\cdot) < \infty, \|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} \le \varsigma\rho^n \right\} .$$

For $x \in \mathsf{X}_0$, we apply Lemma 21.2.9 with $\xi = P^n(x,\cdot)$, $\xi' = \pi$, and $p = 2$. Since $\pi(h) = 0$, this implies

$$|P^n h(x)| \le \|P^n(x,\cdot) - \pi\|_{\mathrm{TV}}^{1/2} \{P^n h^2(x) + \pi(h^2)\}^{1/2} \le \varsigma\rho^n \{P^n h^2(x) + \pi(h^2)\}^{1/2} .$$

Taking the square and integrating with respect to π, the latter inequality implies $\pi(\{P^n h\}^2) \le 2\varsigma^2\rho^{2n}\pi(h^2)$, showing that $\sum_{n=0}^{\infty} \|P^n h\|_{\mathrm{L}^2(\pi)} < \infty$. We conclude by applying Theorem 21.2.6. □

Theorem 21.2.11. *Let P be an aperiodic irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. Assume that there exist $V : \mathsf{X} \to [0,\infty]$, $f : \mathsf{X} \to [1,\infty)$, $b < \infty$, and a nonempty petite set C such that $\sup_C V < \infty$ and $PV + f \le V + b\mathbb{1}_C$. Assume in addition that $\pi(V^2) < \infty$. Then every function $h \in \mathbb{F}(\mathsf{X})$ satisfying $|h|_f < \infty$ is π-integrable, and*

$$n^{-1/2} \sum_{k=0}^{n-1} \bar{h}(X_k) \overset{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0, \sigma_\pi^2(\bar{h})) , \quad \bar{h} = h - \pi(h) ,$$

where $\sigma_\pi^2(\bar{h}) = \pi(\bar{h}^2) + 2\sum_{k=1}^{\infty} \pi(\bar{h}P^k\bar{h})$.

Proof. First note that since $f \le PV + f \le V + b\mathbb{1}_C$ and $\pi(V^2) < \infty$, we get $f \in \mathrm{L}^2(\pi)$ and hence $h \in \mathrm{L}^2(\pi)$. According to Proposition 21.2.4, there exists a solution to the Poisson equation $\hat{h}_0(x) = \sum_{k=0}^{\infty}\{P^k h(x) - P^k h(x_0)\}$, where x_0 is an arbitrary point in $\{V < \infty\}$. The condition $\pi(V^2) < \infty$ implies $\pi(V) < \infty$, and Theorem 17.1.3 (c) then shows that there exists $\varsigma < \infty$ such that

$$\sum_{n=0}^{\infty} \|P^n(x,\cdot) - \pi\|_f \le \varsigma\{V(x) + \pi(V) + 1\} .$$

Since $|h|_f < \infty$, $|P^n h(x_0) - \pi(h)| \le |h|_f \|P^n(x_0,\cdot) - \pi\|_f$, and setting $\bar{h} = h - \pi(h)$, the latter inequality implies

$$\sum_{n=0}^{\infty} |P^n \bar{h}(x_0)| \le \varsigma\, |h|_f \{V(x_0) + \pi(V) + 1\} .$$

Since by Lemma 21.2.2, Poisson solutions are defined up to an additive constant, $\hat{h}(x) = \sum_{k=0}^{\infty} P^k \bar{h}(x)$ is a Poisson solution, which satisfies $|\hat{h}(x)| \le \sum_{k=0}^{\infty} |P^k \bar{h}(x)| \le \varsigma\{V(x) + \pi(V) + 1\}$. Since $\pi(V^2) < \infty$, this implies $\hat{h} \in \mathrm{L}^2(\pi)$, and we may apply Theorem 21.2.5 to prove that

$$n^{-1/2}\sum_{k=0}^{n-1}\bar{h}(X_k) \overset{\mathbb{P}_\xi}{\Longrightarrow} \mathrm{N}(0,\sigma_\pi^2(\bar{h})),$$

where $\sigma_\pi^2(\bar{h}) = 2\pi(\bar{h}\hat{h}) - \pi(\bar{h}^2)$. Moreover, since

$$\sum_{k=0}^{\infty}|\bar{h}(x)P^k\bar{h}(x)| \le \varsigma \left|\bar{h}\right|_f f(x)\{V(x)+\pi(V)+1\}$$

we get $\sum_{k\ge 1}\pi(|\bar{h}P^k\bar{h}|) < \infty$. This shows that

$$\sigma_\pi^2(\bar{h}) = 2\pi(\bar{h}\hat{h}) - \pi(\bar{h}^2) = \pi(\bar{h}^2) + 2\sum_{k=1}^{\infty}\pi(\bar{h}P^k\bar{h}).$$

□

Example 21.2.12. Assume that the Markov kernel P is irreducible, aperiodic, and satisfies the geometric drift condition $D_g(V,\lambda,b,C)$ and that C is a small set. Then the central limit theorem holds for every measurable function g such that $\left|g^2\right|_V < \infty$ if $\pi(V) < \infty$ or simply $|g|_V < \infty$ if $\pi(V^2) < \infty$. ◀

Example 21.2.13. Assume that the Markov kernel P is irreducible and aperiodic and that there exist a small set C and a measurable function $V : \mathsf{X} \to [1,\infty)$ such that $\sup_C V < \infty$, and constants $b, c > 0$ and $\tau \in [0,1)$ such that

$$PV + cV^\tau \le V + b\mathbb{1}_C\,. \tag{21.2.14}$$

If $\pi(V^2) < \infty$, then the central limit theorem holds for all $g \in \mathbb{F}(\mathsf{X})$ such that $|h|_{V^\tau} < \infty$. The condition $\pi(V^2) < \infty$ can be relaxed at the cost of a stronger condition on g. Let $\eta \in (0,1)$. Then for $x \notin C$, $PV(x) \le V(x) - cV^\tau(x)$, and using the inequality $\varphi(a) \le \varphi(x) - (x-a)\varphi'(x)$ for the concave function $\varphi(x) = x^\eta$, we get for $x \notin C$,

$$P(V^\eta)(x) \le [PV(x)]^\eta \le [V(x) - cV^\tau(x)]^\eta \le V^\eta(x) - \eta cV^{\tau+\eta-1}(x)\,.$$

Also, using Jensen's inequality, we get

$$\sup_{x\in C} P(V^\eta)(x) \le \sup_{x\in C}(PV(x))^\eta \le \{\sup_{x\in C} V(x) + b\}^\eta < \infty\,.$$

Thus there exist constants $b_\eta < \infty$ and $c_\eta < \infty$ satisfying

$$PV^\eta + c_\eta V^{\tau+\eta-1} \le V^\eta + b_\eta\mathbb{1}_C\,.$$

Thus if $\tau + \eta - 1 \ge 0$ and $\pi(V^{2\eta}) < \infty$, the central limit theorem holds for all functions g such that $|g|_{V^{\tau+\eta-1}} < \infty$. ◀

21.3 The Resolvent Equation

The existence of a Poisson solution in $\mathrm{L}^2(\pi)$ may be too restrictive. It is possible to keep a decomposition of the sum $S_n(h)$ in the form of a martingale M_n and a remainder R_n using Poisson solutions based on the resolvent, which is defined for $h \in \mathrm{L}^2_0(\pi)$ and $\lambda > 0$ by the resolvent equation

$$(1+\lambda)\hat{h}_\lambda - P\hat{h}_\lambda = h \,. \tag{21.3.1}$$

In contrast to the classical Poisson equation, the resolvent equation always has a solution $\hat{h}_\lambda$ in $\mathrm{L}^2(\pi)$, because $(1+\lambda)\mathrm{I} - P$ is invertible for all $\lambda > 0$. This solution is given by

$$\hat{h}_\lambda = (1+\lambda)^{-1} \sum_{j=0}^{\infty} (1+\lambda)^{-j} P^j h \,. \tag{21.3.2}$$

By Proposition 1.6.3, $\|Ph\|_{\mathrm{L}^2(\pi)} \le \|h\|_{\mathrm{L}^2(\pi)}$, and therefore

$$\left\|\hat{h}_\lambda\right\|_{\mathrm{L}^2(\pi)} \le (1+\lambda)^{-1} \sum_{j=0}^{\infty} (1+\lambda)^{-j} \left\|P^j h\right\|_{\mathrm{L}^2(\pi)} \tag{21.3.3}$$

$$\le (1+\lambda)^{-1} \sum_{j=0}^{\infty} (1+\lambda)^{-j} \|h\|_{\mathrm{L}^2(\pi)} = \lambda^{-1} \|h\|_{\mathrm{L}^2(\pi)} \,. \tag{21.3.4}$$

Define

$$H_\lambda(x_0, x_1) = \hat{h}_\lambda(x_1) - P\hat{h}_\lambda(x_0) \,. \tag{21.3.5}$$

Since $\hat{h}_\lambda \in \mathrm{L}^2(\pi)$ and P is a weak contraction in $\mathrm{L}^2(\pi)$, it follows that $H_\lambda \in \mathrm{L}^2(\pi_1)$, where $\pi_1 = \pi \otimes P$. Define

$$M_n(\hat{h}_\lambda) := \sum_{j=1}^{n} H_\lambda(X_{j-1}, X_j) \,, \tag{21.3.6}$$

$$R_n(\hat{h}_\lambda) := \hat{h}_\lambda(X_0) - \hat{h}_\lambda(X_n) \,. \tag{21.3.7}$$

Lemma 21.3.1 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with a unique invariant probability π. For each fixed $\lambda > 0$ and all $n \ge 1$,*

$$S_n(h) = M_n(\hat{h}_\lambda) + R_n(\hat{h}_\lambda) + \lambda S_n(\hat{h}_\lambda) \,. \tag{21.3.8}$$

Moreover, $\{M_n(\hat{h}_\lambda),\, n \ge 0\}$ is a $\mathbb{P}_\pi$-martingale,

$$n^{-1/2} M_n(\hat{h}_\lambda) \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0, \mathbb{E}_\pi[H_\lambda^2(X_0, X_1)]), \tag{21.3.9}$$

and

$$\mathbb{E}_\pi[R_n^2(\lambda)] \le 4 \left\|\hat{h}_\lambda\right\|^2_{\mathrm{L}^2(\pi)} \,. \tag{21.3.10}$$

Proof. Since $\hat{h}_\lambda$ is the solution to the resolvent equation, we get

$$
\begin{aligned}
S_n(h) &= \sum_{k=0}^{n-1} \{(1+\lambda)\hat{h}_\lambda(X_k) - P\hat{h}_\lambda(X_k)\} \\
&= \sum_{j=0}^{n-1} \{\hat{h}_\lambda(X_j) - P\hat{h}_\lambda(X_j)\} + \lambda S_n(\hat{h}_\lambda) \\
&= \sum_{j=1}^{n} \{\hat{h}_\lambda(X_j) - P\hat{h}_\lambda(X_{j-1})\} + \hat{h}_\lambda(X_0) - \hat{h}_\lambda(X_n) + \lambda S_n(\hat{h}_\lambda) \\
&= M_n(\hat{h}_\lambda) + R_n(\hat{h}_\lambda) + \lambda S_n(\hat{h}_\lambda) .
\end{aligned}
$$

This proves (21.3.8). Equation (21.3.9) follows from Lemma 21.1.1 combined with Remark 21.1.2. The bound (21.3.10) follows from $\mathbb{E}_\pi[\hat{h}_\lambda^2(X_n)] = \left\|\hat{h}_\lambda\right\|_{\mathrm{L}^2(\pi)}$. □

Theorem 21.3.2. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with a unique invariant probability π. Let h be a measurable function such that $\pi(h^2) < \infty$ and $\pi(h) = 0$. Assume that there exist a function $H \in \mathrm{L}^2(\pi_1)$ with $\pi_1 = \pi \otimes P$ and a sequence $\{\lambda_k,\ k \in \mathbb{N}\}$ such that*

$$
0 < \liminf_{k\to\infty} k\lambda_k \leq \limsup_{k\to\infty} k\lambda_k < \infty , \tag{21.3.11a}
$$

$$
\lim_{k\to\infty} \sqrt{\lambda_k}\, \|\hat{h}_{\lambda_k}\|_{\mathrm{L}^2(\pi)} = 0 , \tag{21.3.11b}
$$

$$
\lim_{k\to\infty} \left\|H_{\lambda_k} - H\right\|_{\mathrm{L}^2(\pi_1)} = 0 . \tag{21.3.11c}
$$

Then $n^{-1/2}S_n(h) \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0, \|H\|^2_{\mathrm{L}^2(\pi_1)})$. Moreover, the limit $\lim_{n\to\infty} n^{-1}\mathbb{E}_\pi[S_n^2(h)]$ exists and is equal to $\|H\|^2_{\mathrm{L}^2(\pi_1)}$.

Proof. Since $\pi_1(H_\lambda) = 0$ and H_{λ_k} converges to H in $\mathrm{L}^2(\pi_1)$, we have $\pi_1(H) = 0$. Since $\int P(x_0, \mathrm{d}x_1)H_\lambda(x_0,x_1) = 0$, we have

$$
\begin{aligned}
&\int \pi(\mathrm{d}x_0) \left[\int P(x_0, \mathrm{d}x_1) H(x_0,x_1)\right]^2 \\
&\qquad = \int \pi(\mathrm{d}x_0) \left\{\int P(x_0,\mathrm{d}x_1)[H(x_0,x_1) - H_{\lambda_k}(x_0,x_1)]\right\}^2 \\
&\qquad \leq \int \pi_1(\mathrm{d}x_0,\mathrm{d}x_1)[H(x_0,x_1) - H_{\lambda_k}(x_0,x_1)]^2 = \left\|H_{\lambda_k} - H\right\|^2_{\mathrm{L}^2(\pi_1)} .
\end{aligned}
$$

By assumption (21.3.11c), this proves that $\int P(x_0,\mathrm{d}x_1)H(x_0,x_1) = 0$, π-*a.e.* Hence $\mathbb{E}\left[H(X_j,X_{j+1}) \,\middle|\, \mathscr{F}_j\right] = 0$, $\mathbb{P}_\pi -$ a.s.

For $n \geq 1$, set $M_n = \sum_{j=1}^n H(X_{j-1},X_j)$. Then $\{M_n,\ n \in \mathbb{N}\}$ is a martingale, and by Lemma 21.1.1, we have

$$n^{-1/2}M_n \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0, \|H\|^2_{\mathrm{L}^2(\pi_1)}) . \tag{21.3.12}$$

Since $\mathbb{E}_\pi[\{M_n(\hat{h}_{\lambda_k}) - M_n\}^2] = n \left\|H_{\lambda_k} - H\right\|^2_{\mathrm{L}^2(\pi_1)}$ for each n, condition (21.3.11c) implies that

$$\lim_{k\to\infty} \mathbb{E}_\pi[\{M_n(\hat{h}_{\lambda_k}) - M_n\}^2] = 0 . \tag{21.3.13}$$

Next, condition (21.3.11b) implies that, still for fixed n,

$$\lim_{k\to\infty} \lambda_k \mathbb{E}_\pi[S_n^2(\hat{h}_{\lambda_k})] \le n \lim_{k\to\infty} \lambda_k \mathbb{E}_\pi[S_n(\hat{h}^2_{\lambda_k})] = 0 . \tag{21.3.14}$$

Using the decomposition (21.3.8), we have for $j,k > 0$,

$$\begin{aligned}
&\mathbb{E}_\pi[(R_n(\hat{h}_{\lambda_j}) - R_n(\hat{h}_{\lambda_k}))^2] \\
&\qquad \le 2\mathbb{E}_\pi[\{M_n(\hat{h}_{\lambda_j}) - M_n(\hat{h}_{\lambda_k})\}^2] + 4\lambda_j^2 \mathbb{E}_\pi[S_n^2(\hat{h}_{\lambda_j})] + 4\lambda_k^2 \mathbb{E}_\pi[S_n^2(\hat{h}_{\lambda_k})] .
\end{aligned}$$

Then (21.3.13) and (21.3.14) show that for every fixed n, $\{R_n(\hat{h}_{\lambda_k}),\ k \in \mathbb{N}\}$ is a Cauchy sequence in $\mathrm{L}^2(\pi)$ and there exists a random variable $R_n \in \mathrm{L}^2(\pi)$ such that

$$\lim_{k\to\infty} \mathbb{E}_\pi[\{R_n(\hat{h}_{\lambda_k}) - R_n\}^2] = 0 . \tag{21.3.15}$$

Therefore, letting $\lambda \to 0$ along the subsequence λ_k in the decomposition (21.3.8) yields

$$S_n(h) = M_n + R_n , \quad \mathbb{P}_\pi - \text{a.s.} \tag{21.3.16}$$

It remains to show that $\mathbb{E}_\pi[R_n^2] = o(n)$ as $n \to \infty$. Applying the decompositions (21.3.8) and (21.3.16) and the conditions (21.3.11), we obtain

$$\begin{aligned}
\mathbb{E}_\pi[R_n^2] &= \mathbb{E}_\pi[\left\{M_n(\hat{h}_{\lambda_n}) - M_n + \lambda_n S_n(\hat{h}_{\lambda_n}) + R_n(\hat{h}_{\lambda_n})\right\}^2] \\
&\le 3\mathbb{E}_\pi[\{M_n(\hat{h}_{\lambda_n}) - M_n\}^2] + 3\lambda_n^2 \mathbb{E}_\pi[S_n^2(\hat{h}_{\lambda_n})] + 3\mathbb{E}_\pi[R_n^2(\hat{h}_{\lambda_n})] \\
&\le 3n \left\{ \left\|H_{\lambda_n} - H\right\|^2_{\mathrm{L}^2(\pi_1)} + \left(n\lambda_n + \frac{4}{n\lambda_n}\right) \lambda_n \left\|\hat{h}_{\lambda_n}\right\|^2_{\mathrm{L}^2(\pi)} \right\} = o(n) .
\end{aligned}$$

Combining this inequality with (21.3.16) and (21.3.12) yields $n^{-1/2}S_n(h) \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0, \|H\|^2_{\mathrm{L}^2(\pi_1)})$. The fact that the limit $\lim_{n\to\infty} n^{-1}\mathbb{E}_\pi[S_n^2(h)]$ exists and is equal to $\|H\|^2_{\mathrm{L}^2(\pi_1)}$ follows from the decomposition (21.3.16) and $\mathbb{E}_\pi[R_n^2] = o(n)$. $\square$

The challenge now is to find sufficient conditions for the verification of the conditions (21.3.11). Let P be a Markov kernel that admits a unique invariant probability measure π. For $n \ge 1$, define the kernel V_n by

$$V_n h(x) = \mathbb{E}_x\left[\sum_{k=0}^{n-1} h(X_k)\right] = \sum_{k=0}^{n-1} P^k h(x) , \quad x \in \mathsf{X} . \tag{21.3.17}$$

By Proposition 1.6.3, $\|Ph\|_{\mathrm{L}^2(\pi)} \leq \|h\|_{\mathrm{L}^2(\pi)}$, and therefore V_n is a bounded linear operator on $\mathrm{L}^2(\pi)$ for each n. Consider the Maxwell–Woodroofe condition:

$$\sum_{n=1}^{\infty} n^{-3/2} \|V_n h\|_{\mathrm{L}^2(\pi)} < \infty \,, \tag{21.3.18}$$

where $h \in \mathrm{L}^2_0(\pi)$. Assume first that the Poisson equation $\hat{h} - P\hat{h} = h$ admits a solution $\hat{h}$ in $\mathrm{L}^2(\pi)$. Then

$$V_n h = \sum_{k=0}^{n-1} P^k h = \sum_{k=0}^{n-1} P^k (\mathrm{I} - P)\hat{h} = \hat{h} - P^n \hat{h} \,.$$

Since $\left\|P^n \hat{h}\right\|_{\mathrm{L}^2(\pi)} \leq \left\|\hat{h}\right\|_{\mathrm{L}^2(\pi)}$, this proves that $\|V_n h\|_{\mathrm{L}^2(\pi)} \leq 2\left\|\hat{h}\right\|_{\mathrm{L}^2(\pi)}$ and the series (21.3.18) is summable. Conversely, Jensen's inequality shows that

$$\|V_n h\|^2_{\mathrm{L}^2(\pi)} = \mathbb{E}_\pi[\left(\mathbb{E}_{X_0}[S_n(h)]\right)^2] \leq \mathbb{E}_\pi[S_n^2(h)] \,.$$

Thus if $\limsup_{n\to\infty} n^{-1}\mathbb{E}_\pi[S_n^2(h)] < \infty$, then $\limsup_{n\to\infty} n^{-1/2} \|V_n h\|_{\mathrm{L}^2(\pi)} < \infty$, and we can therefore say that condition (21.3.18) is (within a logarithmic term) not far from being necessary.

Theorem 21.3.3. *Let P be a Markov kernel that admits a unique invariant probability measure π. Let $h \in \mathrm{L}^2_0(\pi)$ be such that (21.3.18) holds. Then the limit*

$$\sigma^2(h) = \lim_{n\to\infty} n^{-1}\mathbb{E}_\pi[S_n^2(h)] \tag{21.3.19}$$

exists and is finite, and $n^{-1/2}S_n(h) \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0, \sigma^2(h))$.

Proof. The proof amounts to checking the conditions of Theorem 21.3.2. Set $\mu_k = 2^{-k}$. We will first establish that

$$\sum_{k=0}^{\infty} \sqrt{\mu_k} \left\|\hat{h}_{\mu_k}\right\|_{\mathrm{L}^2(\pi)} < \infty \,. \tag{21.3.20}$$

Applying summation by parts, we have, for $\lambda > 0$,

$$\hat{h}_\lambda = \sum_{k=1}^{\infty} \frac{P^{k-1}h}{(1+\lambda)^k} = \lambda \sum_{n=1}^{\infty} \frac{V_n h}{(1+\lambda)^{n+1}} \,.$$

This identity and the Minkowski inequality yield

$$\left\|\hat{h}_{\mu_k}\right\|_{\mathrm{L}^2(\pi)} \leq \mu_k \sum_{n=1}^{\infty} (1+\mu_k)^{-n-1} \|V_n h\|_{\mathrm{L}^2(\pi)} \,.$$

This implies, by changing the order of summation,

$$\sum_{k=0}^{\infty} \sqrt{\mu_k} \left\| \hat{h}_{\mu_k} \right\|_{\mathrm{L}^2(\pi)} \leq \sum_{n=1}^{\infty} \left[\sum_{k=0}^{\infty} \frac{\mu_k^{3/2}}{(1+\mu_k)^{n+1}} \right] \|V_n h\|_{\mathrm{L}^2(\pi)} .$$

The quantity in the brackets is equal to $\sum_{k=1}^{\infty}(\mu_{k-1}-\mu_k)h_n(\mu_k)$ with $h_n(x) = \sqrt{x}/(1+x)^{n+1}$. Setting $a_n = 1/(2n+1)$, the function h_n is increasing on $[0,a_n]$ and decreasing on $(a_n,1]$. The series is then bounded above by

$$\begin{aligned} a_n h(a_n) + \int_{a_n}^{1} h_n(x)\mathrm{d}x &\leq O(n^{-3/2}) + \int_0^1 \frac{\sqrt{x}}{1+x} \mathrm{e}^{-nx/2}\mathrm{d}x \\ &\leq O(n^{-3/2}) + n^{-3/2} \int_0^{\infty} \sqrt{u}\mathrm{e}^{-u/2}\mathrm{d}u = O(n^{-3/2}) , \end{aligned}$$

proving (21.3.20).

We will then show that there exists a function $H \in \mathrm{L}^2(\pi_1)$ such that

$$\lim_{k\to\infty} \left\| H_{\mu_k} - H \right\|_{\mathrm{L}^2(\pi_1)} = 0 . \tag{21.3.21}$$

For ν a measure on an arbitrary measurable space, let $\langle \cdot,\cdot \rangle_{\mathrm{L}^2(\nu)}$ denote the scalar product of the space $\mathrm{L}^2(\nu)$. Since $\hat{h}_\lambda$ is a solution to the resolvent equation, we have $P\hat{h}_\lambda = (1+\lambda)\hat{h}_\lambda - h$, and thus for $\lambda,\mu > 0$,

$$\begin{aligned} \langle H_\lambda, H_\mu \rangle_{\mathrm{L}^2(\pi_1)} &= \langle \hat{h}_\lambda, \hat{h}_\mu \rangle_{\mathrm{L}^2(\pi)} - \langle P\hat{h}_\lambda, P\hat{h}_\mu \rangle_{\mathrm{L}^2(\pi)} \\ &= -(\lambda+\mu+\lambda\mu) \langle \hat{h}_\lambda, \hat{h}_\mu \rangle_{\mathrm{L}^2(\pi)} \\ &\quad + (1+\lambda) \langle \hat{h}_\lambda, h \rangle_{\mathrm{L}^2(\pi)} + (1+\mu) \langle \hat{h}_\mu, h \rangle_{\mathrm{L}^2(\pi)} - \|h\|^2_{\mathrm{L}^2(\pi)} . \end{aligned}$$

This yields, on applying the Cauchy–Schwarz inequality,

$$\begin{aligned} \left\| H_\lambda - H_\mu \right\|^2_{\mathrm{L}^2(\pi_1)} &= \|H_\lambda\|^2_{\mathrm{L}^2(\pi_1)} - 2\langle H_\lambda, H_\mu \rangle_{\mathrm{L}^2(\pi_1)} + \left\|H_\mu\right\|^2_{\mathrm{L}^2(\pi_1)} \\ &= -(2\lambda+\lambda^2) \left\| \hat{h}_\lambda \right\|^2_{\mathrm{L}^2(\pi)} + 2(\lambda+\mu+\lambda\mu) \langle \hat{h}_\lambda, \hat{h}_\mu \rangle_{\mathrm{L}^2(\pi)} \\ &\qquad - (2\mu+\mu^2) \left\| \hat{h}_\mu \right\|^2_{\mathrm{L}^2(\pi)} \\ &\leq 2(\lambda+\mu) \left\| \hat{h}_\lambda \right\|_{\mathrm{L}^2(\pi)} \left\| \hat{h}_\mu \right\|_{\mathrm{L}^2(\pi)} \\ &\leq (\lambda+\mu) \left\{ \left\| \hat{h}_\lambda \right\|^2_{\mathrm{L}^2(\pi)} + \left\| \hat{h}_\mu \right\|^2_{\mathrm{L}^2(\pi)} \right\} . \end{aligned}$$

Applying this bound with $\lambda = \mu_k$ and $\mu = \mu_{k-1}$ and using (21.3.20) yields

$$\sum_{k=1}^{\infty} \left\| H_{\mu_k} - H_{\mu_{k-1}} \right\|_{\mathrm{L}^2(\pi_1)} \le \sqrt{3} \sum_{k=1}^{\infty} \sqrt{\mu_k} \left\| \hat{h}_{\mu_k} \right\|^2_{\mathrm{L}^2(\pi)} + \sqrt{3/2} \sum_{k=1}^{\infty} \sqrt{\mu_{k-1}} \left\| \hat{h}_{\mu_{k-1}} \right\|_{\mathrm{L}^2(\pi)}$$
$$\le (\sqrt{3} + \sqrt{3/2}) \sum_{k=0}^{\infty} \sqrt{\mu_k} \left\| \hat{h}_{\mu_k} \right\|_{\mathrm{L}^2(\pi)} < \infty .$$

This proves (21.3.21).

Let k_n be the unique integer such that $2^{k_n - 1} \le n < 2^{k_n}$ and define $\lambda_n = 2^{-k_n}$ for $n \ge 1$. Then $1/2 \le n\lambda_n \le 1$, i.e., (21.3.11a) holds. Moreover, $\{\lambda_k, \ k \in \mathbb{N}^*\} \subset \{2^{-k}, \ k \in \mathbb{N}^*\}$. Thus (21.3.20) and (21.3.21) yield (21.3.11b) and (21.3.11c), and Theorem 21.3.2 applies. □

21.4 A Martingale Coboundary Decomposition

In this section, we prove a central limit theorem based on yet another martingale decomposition. It originates in the general theory of CLT for stationary weakly dependent sequences.

Theorem 21.4.1. *Let P be a Markov kernel that admits a unique invariant probability measure π. Let $h \in \mathrm{L}^2_0(\pi)$ and assume that*

$$\lim_{m \to \infty} \sup_{n \ge 0} \left| \sum_{k=0}^{n} \pi(P^m h P^k h) \right| = 0 . \tag{21.4.1}$$

Then $n^{-1/2} S_n(h) \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0, \sigma^2(h))$ with

$$\sigma^2(h) = \pi(h^2) + 2 \sum_{k=1}^{\infty} \pi(h P^k h) = \lim_{n \to \infty} n^{-1} \mathbb{E}_\pi[S_n^2(h)] .$$

Remark 21.4.2. We do not exclude the possibility that the limiting variance is zero, in which case weak convergence simply means convergence in probability to 0. A sufficient condition for condition (21.4.1) to hold is given by

$$\lim_{m \to \infty} \sum_{k=0}^{\infty} |\pi(P^m h P^k h)| = 0 . \tag{21.4.2}$$

▲

Proof (of Theorem 21.4.1). Fix $m \ge 1$. Define the sequence $\{(Y_k, Z_k), \ k \ge m\}$ by

$$Z_k = P^m h(X_{k-m}) = \mathbb{E}\left[h(X_k) \,|\, \mathscr{F}_{k-m} \right] , \qquad Y_k = h(X_k) - Z_k .$$

Applying Lemma 21.1.1 with $G(x_0,\dots,x_m) = h(x_m) - P^m h(x_0)$, we obtain that there exists σ_m^2 such that $n^{-1/2}\sum_{k=m}^n Y_k \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0,\sigma_m^2)$. For $n > m$, define $R_{m,n} = \sum_{k=m+1}^n Z_k$. It remains to show that

$$\limsup_{m\to\infty} \sup_{n>m} n^{-1}\mathbb{E}_\pi[R_{m,n}^2] = 0 \,, \tag{21.4.3}$$

$$\lim_{m\to\infty} \sigma_m^2 = \sigma^2(h) \,. \tag{21.4.4}$$

We consider first (21.4.3):

$$
\begin{aligned}
&n^{-1}\mathbb{E}_\pi[R_{m,n}^2] \\
&= \frac{n-m}{n}\pi((P^m h)^2) + \frac{2}{n}\sum_{k=m+1}^{n-1}\sum_{j=k+1}^{n} \mathbb{E}_\pi\left[\mathbb{E}\left[h(X_k)\,\middle|\,\mathscr{F}_{k-m}\right]\mathbb{E}\left[h(X_j)\,\middle|\,\mathscr{F}_{k-m}\right]\right] \\
&= \frac{n-m}{n}\pi((P^m h)^2) + \frac{2}{n}\sum_{k=m+1}^{n-1}\sum_{j=k+1}^{n} \mathbb{E}_\pi\left[P^m h(X_{k-m}) P^{j+m-k} h(X_{k-m})\right] \\
&= \frac{n-m}{n}\pi((P^m h)^2) + \frac{2}{n}\sum_{k=m+1}^{n-1}\sum_{j=k+1}^{n} \mathbb{E}_\pi\left[P^m h(X_0) P^{j+m-k} h(X_0)\right] \\
&= \frac{n-m}{n}\pi((P^m h)^2) + \frac{2}{n}\sum_{k=m+1}^{n-1}\sum_{j=1}^{n-k} \mathbb{E}_\pi\left[P^m h(X_0) P^{j+m} h(X_0)\right] \\
&= \frac{n-m}{n}\pi((P^m h)^2) + \frac{2}{n}\sum_{j=1}^{n-m-1} (n-j-m)\pi(P^m h P^{j+m} h) \,.
\end{aligned}
$$

Set for $n > m$, $S_{m,n}(q) = \sum_{j=q}^{n-m}\pi(P^m h P^{j+m} h)$. Applying summation by parts, we obtain

$$\sum_{q=1}^{n-m-1}(n-q-m)\pi(P^m h P^{q+m} h) = (n-1-m)S_{m,n}(1) - \sum_{q=1}^{n-m-1} S_{m,n}(q) \,.$$

Altogether, we obtain

$$n^{-1}\mathbb{E}_\pi[R_{m,n}^2] = \frac{n-m}{n}\pi((P^m h)^2) + \frac{2(n-1-m)}{n}S_{m,n}(1) - \frac{2}{n}\sum_{q=1}^{n-m-1} S_{m,n}(q) \,,$$

which implies that

$$\sup_{n\ge m+1} n^{-1}\mathbb{E}_\pi[R_{m,n}^2] \le \pi\left((P^m h)^2\right) + 4\sup_{g\in\mathbb{N}}\sup_{n\ge m+1} |S_{m,n}(q)| \,.$$

Since condition (21.4.1) implies that $\lim_{m\to\infty}\sup_{q\in\mathbb{N}}\sum_{n\ge m}|S_{m,n}(q)| = 0$, (21.4.3) follows.

We must now prove that $\lim_{m\to\infty}\sigma_m^2 = \sigma^2(h)$. The identity (21.1.3) in Remark 21.1.2 shows that

$$\sigma_m^2 = \mathbb{E}_\pi\left[\left(\sum_{j=0}^{m-1} P^j(h(X_1) - Ph(X_0))\right)^2\right] \sum_{j=0}^{m-1} \mathbb{E}_\pi\left[\{P^j h(X_1) - P^{j+1}h(X_0)\}^2\right]$$
$$+ 2\sum_{j=0}^{m-1}\sum_{q=1}^{m-j-1} \mathbb{E}_\pi\left[\left\{P^j h(X_1) - P^{j+1}h(X_0)\right\}\left\{P^{j+q}h(X_1) - P^{j+q+1}h(X_0)\right\}\right] . \tag{21.4.5}$$

For $j,q \geq 0$, using the stationarity and the Markov property, we can show that

$$\mathbb{E}_\pi\left[\left\{P^j h(X_1) - P^{j+1}h(X_0)\right\}\left\{P^{j+q}h(X_1) - P^{j+q+1}h(X_0)\right\}\right]$$
$$= \pi\left(P^j h P^{j+q} h\right) - \pi\left(P^{j+1} h P^{j+q+1} h\right) .$$

Plugging this expression into (21.4.5) and then rearranging the terms in the summation yields

$$\sigma_m^2 = \pi(h^2) + 2\sum_{j=1}^{m-1} \pi(hP^j h) - \pi((P^m h)^2) - 2\sum_{j=1}^{m-1} \pi(P^m h P^j h) .$$

By assumption, we can choose m such that

$$\left|\sigma^2(h) - \sigma_m^2\right| \leq \varepsilon .$$

Since for every fixed m, $\lim_{n\to\infty} n^{-1}\mathrm{Var}_\pi\left(\sum_{k=1}^m h(X_k)\right) = 0$, the central limit theorem follows with the limiting variance as stated. □

If h is bounded and P is uniformly ergodic, the convergence of the series in (21.4.2) is trivial. In other circumstances, this requires more work. This will be done in the following subsections for irreducible geometrically and subgeometrically ergodic kernels and for nonirreducible kernels.

21.4.1 Irreducible Geometrically and Subgeometrically Ergodic Kernels

The main tool is the following lemma, which relies on a general covariance inequality that will proved in Section 21.A for the sake of completeness. We first introduce some notation. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space and X a random variable. We denote by F_X the cumulative distribution function and by $\bar{F}_X$ the survival function of the random variable $|X|$, i.e., for $x \in \mathbb{R}_+$, $F_X(x) = \mathbb{P}(|X| \leq x)$ and $\bar{F}_X = 1 - F_X$. The function $\bar{F}_X$ is nonincreasing and continuous to the right with limits to the left. We denote by Q_X the tail quantile function of X, defined for all $u \in [0,1]$ by

$$Q_X(u) = \inf\{x \in \mathbb{R}_+ : \bar{F}_X(x) \leq u\} \tag{21.4.6}$$

with the convention $\inf\emptyset = +\infty$. Note that for all $u \in [0,1]$, $Q_X(u) = \bar{Q}_X(1-u)$, where $\bar{Q}_X$ is the quantile function of $|X|$,

$$\bar{Q}_X(u) = \inf\{x \in \mathbb{R}_+ \,:\, \bar{F}_X(x) \geq u\} \ .$$

The quantile function $\bar{Q}_X$ being nondecreasing and left continuous with limits to the right, the tail quantile Q_X is nonincreasing and right continuous with limits to the left. Moreover,

$$\mathbb{P}(|X| > x) \leq u \quad \text{if and only if} \quad Q_X(u) \leq x \ .$$

Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. For $h \in \mathbb{F}(\mathsf{X})$ and $m \in \mathbb{N}$, define the tail quantile function Q_m of $|P^m h(X_0)|$ under $\mathbb{P}_\pi$ by

$$Q_m(u) = \inf\{x \geq 0 \,:\, \mathbb{P}_\pi(|P^m h(X_0)| > x) \leq u\} \ . \tag{21.4.7}$$

To apply Theorem 21.4.1, it is required to obtain a bound of $\sum_{k=1}^\infty |\pi(P^m h P^k h)|$. For this purpose, we will use a covariance inequality that is stated and proved in Section 21.A.

Lemma 21.4.3 *Let P be a Markov kernel that admits an invariant probability measure π. Assume that there exists a sequence $\{\rho_n,\, n \in \mathbb{N}\}$ such that for all $n \geq 1$,*

$$\int_{\mathsf{X}} \pi(\mathrm{d}x) \mathrm{d}_{\mathrm{TV}}(P^n(x,\cdot),\pi) \leq \rho_n \ . \tag{21.4.8}$$

Let H be the function defined on $[0,1]$ by

$$H(u) = \sum_{k=1}^\infty \mathbb{1}\{u < \rho_k\} \ . \tag{21.4.9}$$

Then for all $h \in \mathrm{L}_0^2(\pi)$,

$$\sum_{k=1}^\infty |\pi(P^m h P^k h)| \leq \int_0^1 Q_0^2(u) H(u) \mathbb{1}\{u \leq \rho_m\} \mathrm{d}u \ . \tag{21.4.10}$$

Consequently, if $\lim_{n\to\infty} \rho_m = 0$ and $\int_0^1 Q_0^2(u)H(u)\mathrm{d}u < \infty$, then (21.4.2) holds.

Proof. For $m \geq 0$, set $g = P^m h$. We apply Lemma 21.A.1 to $X = g(X_0)$ and $Y = h(X_k)$. For this purpose, it is required to compute for all $(x,y) \in \mathbb{R}^2$,

$$\begin{aligned}
&|\mathbb{P}_\pi(g(X_0) > x, h(X_k) > y) - \mathbb{P}_\pi(g(X_0) > x)\mathbb{P}_\pi(h(X_k) > y)| \\
&\qquad = \left|\mathbb{E}_\pi\left[\mathbb{1}_{\{g(X_0)>x\}}\left\{\mathbb{E}\left[\mathbb{1}_{\{h(X_k)>y\}} - \mathbb{P}_\pi(h(X_k)>y) \,\middle|\, \mathscr{F}_0\right]\right\}\right]\right| \\
&\qquad \leq \mathbb{E}_\pi\left[\left|P^k[\mathbb{1}_{(y,\infty)} \circ h](X_0) - \pi(\mathbb{1}_{(y,\infty)} \circ h)\right|\right] \leq \rho_k \ .
\end{aligned}$$

Define

$$a_k = \rho_k \wedge 1 \ . \tag{21.4.11}$$

Since $h(X_k)$ has the same distribution as $h(X_0)$, its tail quantile function is Q_0, and we obtain

$$|\pi(gP^k h)| \leq 2\left(\int_0^{a_k} Q_0^2(u)\mathrm{d}u\right)^{1/2}\left(\int_0^{a_k} Q_m^2(u)\mathrm{d}u\right)^{1/2} .$$

Furthermore, by Lemma 21.A.3 applied to $Y = P^m h(X_0) = \mathbb{E}\left[h(X_m)\,|\,\mathscr{F}_0\right]$ and $X = h(X_m)$, which is distributed as $h(X_0)$ under $\mathbb{P}_\pi$, we have

$$\int_0^{a_k} Q_m^2(u)\mathrm{d}u \leq \int_0^{a_k} Q_0^2(u)\mathrm{d}u .$$

We have thus obtained, for all $k,m \geq 0$,

$$|\pi(P^k h P^m h)| \leq 2\int_0^{a_k} Q_0^2(u)\mathrm{d}u .$$

Interchanging the roles of k and m, we obtain

$$|\pi(P^k h P^m h)| \leq 2\int_0^{a_m} Q_0^2(u)\mathrm{d}u .$$

These two bounds yield

$$|\pi(P^k h P^m h)| \leq 2\int_0^{1} Q_0^2(u)\mathbb{1}\left\{u \leq \rho_k\right\}\mathbb{1}\left\{u \leq \rho_m\right\}\mathrm{d}u .$$

Summing over the indices k yields (21.4.10). □

Combining Theorem 21.4.1 and Lemma 21.4.3, we obtain central limit theorems for polynomially or geometrically ergodic Markov kernels.

Theorem 21.4.4. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. Assume that there exists a sequence $\{\rho_k,\ k \in \mathbb{N}\}$ such that for all $n \geq 1$,*

$$\int \pi(\mathrm{d}x)\mathrm{d}_{\mathrm{TV}}(P^n(x,\cdot),\pi) \leq \rho_n , \tag{21.4.12}$$

$$\sum_{k=1}^{\infty} \rho_k^{\delta/(2+\delta)} < \infty , \quad \textit{for some } \delta > 0. \tag{21.4.13}$$

Then for every $h \in \mathrm{L}^{2+\delta}(\pi)$ and $\pi(h) = 0$, we get

$$n^{-1/2}\sum_{k=0}^{n-1} h(X_k) \stackrel{\mathbb{P}_\xi}{\Longrightarrow} \mathrm{N}\left(0,\sigma_\pi^2(h)\right) , \tag{21.4.14}$$

where $\sigma_\pi^2(h)$ is given by (21.2.13).

Proof. In order to apply Lemma 21.4.3, we must prove that $\int_0^1 Q_0^2(u)H(u)\mathrm{d}u < \infty$, where

$$H(u) = \sum_{k=1}^{\infty} \mathbb{1}\left\{u \leq \rho_k\right\} .$$

If $h \in \mathrm{L}^{2+\delta}(\pi)$, then by Markov's inequality, we have $Q_0(u) \leq \varsigma u^{-1/(2+\delta)}$ for all $u \in [0,1]$. Thus

$$\int_0^1 Q_0^2(u)H(u)\mathrm{d}u \leq \varsigma^2 \sum_{k=1}^{\infty} \int_0^{\rho_k} u^{-2/(2+\delta)}\mathrm{d}u \leq \bar{\varsigma} \sum_{k=1}^{\infty} \rho_k^{\delta/(2+\delta)} < \infty .$$

□

Corollary 21.4.5 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. Assume that there exists $a > 1$ such that for all $n \geq 1$,*

$$\int \pi(\mathrm{d}x)\mathrm{d}_{\mathrm{TV}}(P^n(x,\cdot),\pi) \leq n^{-a} . \tag{21.4.15}$$

Then for every $h \in \mathrm{L}^{2+\delta}(\pi)$ with $\delta > 2/(a-1)$ (i.e., $a > 1 + 2/\delta$) and $\pi(h) = 0$, (21.4.14) holds.

Proof. The result follows from Theorem 21.4.4 by setting $\rho_n = n^{-a}$. □

Remark 21.4.6. The condition $h \in \mathrm{L}^{2+\delta}(\pi)$ with $\delta > 2/(a-1)$ is sharp. There exist polynomially ergodic chains with rate n^{-a} with $a > 1$ and functions $h \in \mathrm{L}^2(\pi)$ such that the CLT does not hold. ▲

For a geometrically ergodic kernel (i.e., one that satisfies one of the equivalent conditions of Theorem 15.1.5), we obtain the following result.

Theorem 21.4.7. *Let P be a Markov kernel with invariant probability π. Assume that P is geometrically ergodic (see Definition 15.1.1). Then for every measurable function h such that $\pi(h^2 \log(1+|h|)) < \infty$ and $\pi(h) = 0$,*

$$n^{-1/2} \sum_{k=0}^{n-1} h(X_k) \stackrel{\mathbb{P}_x}{\Longrightarrow} \mathrm{N}\left(0, \sigma^2(h)\right) ,$$

where the asymptotic variance is given by (21.2.13).

Proof. By Theorem 15.1.6, there exist $\rho \in [0,1)$ and a function $V : \mathsf{X} \to [1,\infty)$ satisfying $\pi(V) < \infty$ such that for all $n \in \mathbb{N}$ and π-*a.e.* $x \in \mathsf{X}$,

$$\mathrm{d}_{\mathrm{TV}}(P^n(x,\cdot),\pi) \leq V(x)\rho^n \ .$$

Hence for all $n \geq 1$, $\int \pi(\mathrm{d}x)\mathrm{d}_{\mathrm{TV}}(P^n(x,\cdot),\pi) \leq \pi(V)\rho^n$ for all $n \geq 1$. We must prove that $\int Q_0^2(u)H(u)\mathrm{d}u < \infty$, where

$$H(u) = \sum_{k=1}^{\infty} \mathbb{1}\left\{u \leq \pi(V)\rho^k\right\} = \sum_{k=1}^{\infty} \mathbb{1}\left\{k \leq \frac{\log(\pi(V)/u)}{\log(1/\rho)}\right\} \leq \frac{\log(\pi(V)/u)}{\log(1/\rho)} \ .$$

Set $\phi(x) = (1+x)\log(1+x) - x$ and $\psi(y) = \mathrm{e}^y - 1 - y$. Then (ϕ,ψ) is a pair of Young functions (see Lemma 17.A.2), and thus $xy \leq \phi(cx) + \psi(y/c)\}$ for all $x, y \geq 0$ and $c > 0$. This yields

$$\begin{aligned}
\int_0^1 Q_0^2(u)H(u)\mathrm{d}u &\leq \int_0^1 \phi(cQ_0^2(u))\mathrm{d}u + \int_0^1 \psi(H(u)/c)\mathrm{d}u \\
&\leq \pi(\{1+ch^2\}\log\{1+ch^2\}) + \int_0^1 \left(\frac{u}{\pi(V)}\right)^{-1/\{c\log(1/\rho)\}} \mathrm{d}u.
\end{aligned}$$

Choosing $c > 1/\log(1/\rho)$ proves that the function $\int_0^1 Q_0^2(u)H(u)\mathrm{d}u$ is finite. On the other hand, the condition $\pi(\{1+h^2\}\log(1+|h|)) < \infty$ implies that for all $c > 0$, $\pi(\{1+ch^2\}\log(1+ch^2)) < \infty$. □

Remark 21.4.8. For geometrically ergodic Markov chains, the moment condition $\pi(h^2\log(1+|h|)) < \infty$ cannot be further refined to a second moment without additional assumptions. One may construct a geometrically ergodic Markov chain and a function h such that $\pi(h^2) < \infty$, and yet a CLT fails. ▲

We can also prove the central limit theorem when the rate sequence is a general subgeometric rate. The following result subsumes the previous ones.

Theorem 21.4.9. *Let P be a Markov kernel with invariant probability π. Let (ϕ,ψ) be a pair of inverse Young functions. Assume that there exists $\varsigma < \infty$ such that for all $n \in \mathbb{N}$,*

$$\int \pi(\mathrm{d}x)\mathrm{d}_{\mathrm{TV}}(P(x,\cdot),\pi) \leq \varsigma/\phi(n) \ . \tag{21.4.16}$$

If there exists $c > 0$ such that $\int_0^1 \phi(\phi^{-1}(\varsigma/u)/c)\mathrm{d}u < \infty$, then for every measurable function h such that $\pi(\psi(h^2)) < \infty$ and $\pi(h) = 0$,

$$n^{-1/2}\sum_{k=0}^{n-1} h(X_k) \stackrel{\mathbb{P}_\xi}{\Longrightarrow} \mathrm{N}\left(0,\sigma_\pi^2(h)\right),$$

where $\sigma_\pi^2(h)$ is given by (21.2.13).

Proof. We check that under the stated condition, the function Q^2h is integrable on $[0,1]$ with

$$H(u) = \sum_{n=1}^{\infty} \mathbb{1}\{u \le \varsigma/\phi(n)\} = \sum_{n=1}^{\infty} \mathbb{1}\left\{n \le \phi^{-1}(\varsigma/u)\right\} \le \phi^{-1}(\varsigma/u) \, .$$

Therefore,

$$\int Q_0^2(u)H(u)\mathrm{d}u \le \mathbb{E}_\pi[\psi(h^2(X_0))] + c\int_0^1 \phi(\phi^{-1}(\varsigma/u)/c)\mathrm{d}u \, .$$

□

Example 21.4.10. If $\phi(n) = \mathrm{e}^{an^\beta}$ for some $a > 0$ and $\beta \in (0,1)$, then $\phi^{-1}(y) = \log^{1/\beta}(y/a)$ and

$$\phi(\phi^{-1}(m/u)/c) = \mathrm{e}^{a\log(m/u)/c} = Cu^{-a/c} \, .$$

This is an integrable function on $[0,1]$ if $c > a$. The inverse Young conjugate ψ of ϕ satisfies

$$\psi(x) \sim Cx\log^{1/\beta}(x)$$

as $x \to \infty$. Therefore, the central limit theorem holds for functions h such that

$$\pi(h^2 \log^{1/\beta}(1+|h|)) < \infty \, .$$

◀

21.4.2 Nonirreducible Kernels

In this section we check that the conditions of Theorem 21.4.1 hold for a nonirreducible kernel that satisfies the contractivity properties with respect to the Wasserstein distance of Theorem 20.4.5. For a function V, set as usual $\bar{V}(x,y) = \{V(x)+V(y)\}/2$.

Theorem 21.4.11. *Let P be a Markov kernel on a complete separable metric space satisfying the drift condition $\mathrm{D_g}(V,\lambda,b)$. Let c be a cost function that satisfies* **A** 20.1.5*, $\mathsf{c} \le \bar{V}$, and for all $x,y \in \mathsf{X}$,*

$$\mathbf{W}_{\mathsf{c}}(P(x,\cdot),P(y,\cdot)) \le \mathsf{c}(x,y) \, .$$

Assume, moreover, that there exist $\varepsilon \in (0,1)$ and $d > 0$ such that $\lambda + 2b/(1+d) < 1$ and $\{V \le d\} \times \{V \le d\}$ is a $(\mathsf{c},1,\varepsilon)$-contracting set. Then for every $\alpha \in (0,1/2)$, every function h such that $|h|_{\mathrm{Lip}(\mathsf{c}^{1-\alpha}\bar{V}^{\alpha})} < \infty$ and $\pi(h) = 0$, and every initial dis-

tribution ξ such that $\xi(V) < \infty$, the central limit theorem holds under $\mathbb{P}_\xi$, i.e., $n^{-1/2}\sum_{k=0}^{n-1} h(X_k) \stackrel{\mathbb{P}_\xi}{\Longrightarrow} \mathrm{N}(0,\sigma^2(h))$ with $\sigma^2(h) = \pi(h^2) + 2\sum_{k=1}^{\infty} \pi(hP^k h)$.

Proof. Fix $\alpha \in (0,1/2)$ and set $\mathsf{c}_\alpha = \mathsf{c}^{1-\alpha}\bar{V}^\alpha$. By Theorem 20.4.5 and Corollary 20.4.7, there exist $\rho \in (0,1)$ and a constant ϑ such that

$$|P^n h(x)| \leq \vartheta\, |h|_{\mathrm{Lip}(c_\alpha)} \rho^n V^\alpha(x) \,.$$

Since $\pi(V) < \infty$ by Theorem 20.4.5, this implies that $\|P^n h\|_{\mathrm{L}^2(\pi)} = O(\rho^n)$. Therefore, (21.4.2) holds, and this proves the central limit theorem under $\mathbb{P}_\pi$ with the limiting variance as stated. Let ξ be an initial distribution. As noted in Remark 20.4.3, Proposition 20.4.2 implies that there exists a coupling $\{(X_n, X_n'),\ n \in \mathbb{N}\}$ such that $\{X_n,\ n \in \mathbb{N}\}$ and $\{X_n',\ n \in \mathbb{N}\}$ are Markov chains with kernel P and initial distributions ξ and π and for all $\gamma \in \mathscr{C}(\xi,\pi)$,

$$\mathbb{E}_\gamma[\mathsf{c}_\alpha(X_n, X_n')] \leq \vartheta \rho^n \{\xi(V) + \pi(V)\} \,.$$

This yields for $h \in \mathrm{Lip}_{\mathsf{c}_\alpha}(\mathsf{X})$,

$$\mathbb{E}\left[\left| n^{-1/2}\sum_{i=0}^{n-1} h(X_i) - n^{-1/2}\sum_{i=0}^{n-1} h(X_i')\right|\right] \\ \leq \vartheta\, |h|_{\mathrm{Lip}(\mathsf{c}_\alpha)} \{\xi(V) + \pi(V)\} n^{-1/2}\sum_{k=0}^{n-1}\rho^k = O(n^{-1/2}) \,.$$

This proves that the limiting distribution of $n^{-1/2}\sum_{i=1}^n h(X_i')$ is the same as that of $n^{-1/2}\sum_{i=1}^n h(X_i')$, i.e., the CLT holds under $\mathbb{P}_\xi$. $\square$

When the rate of convergence is polynomial, we also obtain a central limit theorem under more stringent restrictions on the functions considered.

Theorem 21.4.12. *Let P be a Markov kernel on a complete separable metric space satisfying the drift condition $\mathrm{D}_{\mathrm{sg}}(V,\phi,b,C)$ with V unbounded, $\pi(V^2) < \infty$, $\sup_C V < \infty$, $d = \inf_{C^c} \phi \circ V > b$, $\phi(u) = u^{\alpha_0}$ with $\alpha_0 \in (1/2,1)$ and $\mathsf{c} \leq \bar{V}$. Assume, moreover, that $\bar{C} = C \times C$ is a $(\mathsf{c},1,\varepsilon)$-contracting set and for all $x,y \in \mathsf{X}$,*

$$\mathbf{W}_{\mathsf{c}}(P(x,\cdot),P(y,\cdot)) \leq \mathsf{c}(x,y) \,.$$

Then for all initial distributions ξ such that $\xi(V) < \infty$ and all functions h such that $|h|_{\mathrm{Lip}(\mathsf{c}^{1-\alpha_0})} < \infty$ and $\pi(h) = 0$, the central limit theorem holds under $\mathbb{P}_\xi$, i.e., $n^{-1/2}\sum_{k=0}^{n-1} h(X_k) \stackrel{\mathbb{P}_\xi}{\Longrightarrow} \mathrm{N}(0,\sigma^2(h))$ with $\sigma^2(h) = \pi(h^2) + 2\sum_{k=1}^{\infty} \pi(hP^k h)$.

Proof. By Theorem 20.5.2 and Remark 20.5.3 and (20.1.9), if $\pi(V) < \infty$, $\pi(h) = 0$, and $|h|_{\mathrm{Lip}(\mathsf{c}^{1-\alpha_0})} < \infty$, then $|Ph(x)| \le \vartheta n^{-\alpha_0/(1-\alpha_0)} V(x)$. Since $\alpha_0 > 1/2$, $\alpha_0/(1-\alpha_0) > 1$, and if moreover $\pi(V^2) < \infty$, then (21.4.2) holds. This proves the central limit theorem under $\mathbb{P}_\pi$ with the stated variance. The bound (20.5.3) implies that there exists a coupling $\{(X_n, X'_n),\ n \in \mathbb{N}\}$ such that $\{X_n,\ n \in \mathbb{N}\}$ and $\{X'_n,\ n \in \mathbb{N}\}$ are Markov chains with kernel P and initial distributions ξ and π and for all $\gamma \in \mathscr{C}(\xi, \pi)$,

$$\mathbb{E}_\gamma[\mathsf{c}^{1-\alpha_0}(X_n, X'_n)] \le \vartheta n^{-\alpha_0/(1-\alpha_0)} \{\xi(V) + \pi(V)\} .$$

This yields for $h \in \mathrm{Lip}_{\mathsf{c}^{1-\alpha_0}}(\mathsf{X})$,

$$\mathbb{E}\left[\left| n^{-1/2} \sum_{i=0}^{n-1} h(X_i) - n^{-1/2} \sum_{i=0}^{n-1} h(X'_i) \right|\right] \\ \le \vartheta\, |h|_{\mathrm{Lip}(\mathsf{c}_\alpha)} \{\xi(V) + \pi(V)\} n^{-1/2} \sum_{k=0}^{n-1} k^{-\alpha_0/(1-\alpha_0)} = O(n^{-1/2}) .$$

This proves that the limiting distribution of $n^{-1/2} \sum_{i=1}^n h(X'_i)$ is the same as that of $n^{-1/2} \sum_{i=1}^n h(X'_i)$, i.e., the CLT holds under $\mathbb{P}_\xi$. □

21.5 Exercises

21.1. Let P be a Markov kernel that admits a positive recurrent attractive atom α and let $h \in \mathrm{L}^1(\pi)$ with $\pi(h) = 0$. Show that a solution to the Poisson equation (21.2.1) is given for all $x \in \mathsf{X}$ by

$$\hat{h}(x) = \mathbb{E}_x\left[\sum_{k=0}^{\tau_\alpha} h(X_k)\right] = \mathbb{E}_x\left[\sum_{k=0}^{\sigma_\alpha} h(X_k)\right] . \tag{21.5.1}$$

21.2. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ that admits a positive recurrent attractive atom α and let $h \in \mathrm{L}^1(\pi)$ be such that $\pi(h) = 0$ and $\mathbb{E}_\alpha\left[\left(\sum_{k=1}^{\sigma_\alpha} h(X_k)\right)^2\right] < \infty$.

1. Show that $n^{-1/2} \sum_{k=1}^n h(X_k) \stackrel{\mathbb{P}_\mu}{\Longrightarrow} \mathrm{N}(0, \sigma^2(h))$ for every initial distribution $\mu \in \mathbb{M}_1(\mathscr{X})$, with

$$\sigma^2(h) = \frac{1}{\mathbb{E}_\alpha[\sigma_\alpha]} \mathbb{E}_\alpha\left[\left(\sum_{k=1}^{\sigma_\alpha} h(X_k)\right)^2\right] . \tag{21.5.2}$$

2. Assume that

$$\mathbb{E}_\alpha\left[\left(\sum_{i=1}^{\sigma_\alpha}|h(X_i)|\right)^2\right] < \infty\,.$$

Prove that $\pi(h^2)+\pi(|h\hat{h}|)<\infty$ and

$$2\pi(h\hat{h})-\pi(h^2)=\frac{1}{\mathbb{E}_\alpha[\sigma_\alpha]}\,\mathbb{E}_\alpha\left[\left(\sum_{k=1}^{\sigma_\alpha}h(X_k)\right)^2\right]\,. \tag{21.5.3}$$

21.3. Let P be a Markov kernel that admits a positive recurrent attractive atom α and let $h\in \mathrm{L}^2_0(\pi)$ be such that $\sum_{k=1}^\infty \pi(|hP^kh|)<\infty$. Let $\sigma^2(h)$ be as in (21.5.2). Show that

$$\sigma^2(h)=\pi(h^2)+2\sum_{k=1}^\infty \pi(hP^kh)=\lim_{n\to\infty}\frac{1}{n}\mathbb{E}_\pi\left[\left(\sum_{k=1}^n h(X_k)\right)^2\right]\,.$$

21.4. Provide an alternative proof of Proposition 21.1.3 based on the maximal distributional coupling of $(\mathbb{P}_\lambda,\mathbb{P}_\pi)$ (see Theorem 19.3.9).

21.5. This exercise provides an example of a Markov chain for which a CLT holds but $\lim_{n\to\infty} n\,\mathrm{Var}_\pi(S_n(h)^2)=\infty$. Let $\mathsf{X}=\mathbb{N}$ and let h be the identity function. Consider the Markov chain on X with transition matrix defined by $P(0,0)=1/2$ and for $j\geq 1$, $P(j,-j)=P(-j,0)=1$ and $P(0,j)=c/j^3$ with $c=\zeta(3)^{-1}/2$, and $\zeta(s)=\sum_{n=1}^\infty n^{-s}$ is the Riemann zeta function. Whenever the chain leaves 0, it cycles to some positive integer j, then to $-j$, and then back to 0.

1. Show that P has a unique invariant probability π, that $n^{-1/2}\sum_{i=0}^{n-1}h(X_i)\stackrel{\mathbb{P}_\pi}{\Longrightarrow}\mathrm{N}(0,\sigma^2)$, but that $\lim_{n\to\infty} n\,\mathrm{Var}_\pi(S_n(h)^2)=\infty$.
2. Modify this construction to obtain a nondegenerate CLT.

21.6 (Continuation of Example 20.3.5). Let $\{\varepsilon_n,\,n\in\mathbb{N}\}$ be i.i.d. random variables taking the values 0 and 1 with probability $1/2$ each and define

$$X_n=\frac{1}{2}(X_{n-1}+\varepsilon_n)\,,\quad n\geq 1\,.$$

Set $D_k=\{j2^{-k}:j=0,\,\ldots,\,2^k-1\}$. Let f be a square-integrable function defined on $[0,1]$ such that $\int_0^1 f(x)\mathrm{d}x=0$. Denote by $\|\cdot\|_2$ the L^2 norm with respect to Lebesgue measure on $[0,1]$.

4. Show that

$$P^kf(x)=2^{-k}\sum_{z\in D_k}\int_0^1\left[f\left(\frac{x}{2^k}+z\right)-f\left(\frac{y}{2^k}+z\right)\right]\mathrm{d}y\,.$$

5. Show that

$$\left\|P^k f\right\|_2^2 \leq 2^k \int \int_{|x-y|\leq 2^{-k}} [f(x)-f(y)]^2 \mathrm{d}x\mathrm{d}y \,.$$

6. Prove that if f is Hölder continuous with exponent $\gamma > 1/2$, i.e., there exists a constant ϑ such that $|f(x)-f(y)| \leq C|x-y|^{\gamma}$, then Condition (21.4.2) holds.

21.6 Bibliographical Notes

Early proofs of the CLT for Markov chains were obtained in Dobrushin (1956c), Dobrushin (1956a), Nagaev (1957), Billingsley (1961), and Cogburn (1972). A detailed account of the theory is given in Jones and Hobert (2001), Jones (2004), Häggström (2005), and Häggström and Rosenthal (2007).

The Poisson equation and the general potential theory of positive kernels is developed in Neveu (1972) and Revuz (1984). Existence and properties of Poisson equations are presented in Glynn and Meyn (1996) (the statement of Proposition 21.2.4 is similar to (Glynn and Meyn 1996, Theorem 2.3), but the proof is different). The decomposition (21.2.2) was used by Maigret (1978) and Duflo (1997) to derive a CLT for Harris recurrent Markov chains. For irreducible Markov chains, Meyn and Tweedie (2009) used the Poisson equation to derive a central limit theorem and law of iterated logarithm for geometrically ergodic Markov chains. Jarner and Roberts (2002) extended these results to polynomial ergodicity (Example 21.2.13 is taken from (Jarner and Roberts 2002, Theorem 4.2)).

The proof of Theorem 21.3.3 is due to Maxwell and Woodroofe (2000) (see also Tóth (1986) and Tóth (2013)). The resolvent equation was introduced earlier in Kipnis and Varadhan (1985) and Kipnis and Varadhan (1986). Necessary and sufficient conditions (not discussed here) for additive functionals of a Markov chains to be asymptotically normal are given in Wu and Woodroofe (2004).

The proof of Theorem 21.4.1 is originally due to Gordin (1969) (see also Eagleson (1975), Durrett and Resnick (1978), Dedecker and Rio (2000), Dedecker and Rio (2008)). The version we use here is an adaptation of the proof of (Hall and Heyde 1981, Theorem 5.2) to the context of Markov chains. The main arguments of the proof of Lemma 21.4.3 can be found in Rio (1993) (see also Rio (1994, 2000b, 2017)). The counterexample alluded to in Remark 21.4.6 is developed in (Rio (2017), Section 9.7).

Theorem 21.4.4 is taken from Jones (2004), where it is obtained as a consequence of the CLT for strongly mixing sequencees established in Ibragimov (1959) and Ibragimov (1963) (see also Ibragimov and Linnik (1971), Doukhan (1994); Doukhan et al. (1994); Dedecker et al. (2007)). Exercise 21.5 is taken from (Häggström and Rosenthal 2007, Examples 11 and 12).

The comparison of the different possible expressions for the variance in the CLT is discussed in Häggström and Rosenthal (2007); see also Häggström (2005). Construction of confidence intervals for additive functionals of Markov chains using the CLT is discussed in Flegal and Jones (2010), Atchadé (2011), and Flegal and Jones

(2011). These papers also discuss different estimators of the asymptotic variance, which is an important topic in practice. Confidence intervals for additive functionals are also discussed in Atchadé (2016) and Rosenthal (2017).

21.A A Covariance Inequality

Lemma 21.A.1 *Let $(\Omega,\mathscr{A},\mathbb{P})$ be a probability space and X,Y two square-integrable random variables defined on $(\Omega,\mathscr{A},\mathbb{P})$. Define*

$$\alpha = \alpha(X,Y) = 2 \sup_{(x,y)\in\mathbb{R}^2} |\mathbb{P}(X > x, Y > y) - \mathbb{P}(X > x)\mathbb{P}(Y > y)| \,. \tag{21.A.1}$$

Then

$$|\mathrm{Cov}(X,Y)| \le 2\int_0^\alpha Q_X(u)Q_Y(u)\mathrm{d}u \le 2\left(\int_0^a Q_X^2(u)\mathrm{d}u\right)^{1/2}\left(\int_0^\alpha Q_Y^2(u)\mathrm{d}u\right)^{1/2} .$$

Proof. For X,Y two square-integrable random variables defined on a probability space $(\Omega,\mathscr{A},\mathbb{P})$, one has

$$\begin{aligned}&\mathrm{Cov}\,(X,Y)\\ &\quad = \int_0^\infty\int_0^\infty \mathrm{Cov}\,(\mathbb{1}\{X > x\} - \mathbb{1}\{X < -x\}, \mathbb{1}\{Y > y\} - \mathbb{1}\{Y < -y\})\,\mathrm{d}x\mathrm{d}y \,.\end{aligned} \tag{21.A.2}$$

Note indeed that every random variable X can be written as

$$X = X^+ - X^- = \int_0^\infty \left[\mathbb{1}_{\{X>x\}} - \mathbb{1}_{\{X<-x\}}\right]\mathrm{d}x \,.$$

Writing Y similarly and applying Fubini's theorem yields (21.A.2). For $x \in \mathbb{R}$, set $I_x = \mathbb{1}_{(x,\infty)} - \mathbb{1}_{(-\infty,-x)}$. Since the functions I_x are uniformly bounded by 1, we obtain

$$|\mathrm{Cov}\,(I_x(X), I_y(Y))\,| = \left|\mathbb{E}\left[I_x(X)\{I_y(Y) - \mathbb{E}\left[I_y(Y)\right]\}\right]\right| \le 2\alpha \,.$$

On the other hand, using that $\mathbb{E}\left[|I_x(X)|\right] = \mathbb{P}(|X| > x)$, we get

$$|\mathrm{Cov}\,(I_x(X), I_y(Y))\,| \le 2\mathbb{P}(|X| > x) \wedge \mathbb{P}(|Y| > y) \,.$$

Plugging these bounds into (21.A.2), we obtain

$$
\begin{aligned}
|\mathrm{Cov}\,(X,Y)| &\le \int_0^\infty \int_0^\infty |\mathrm{Cov}\,(I_x(X), I_y(Y))|\,\mathrm{d}x\mathrm{d}y \\
&\le 2\int_0^\infty \int_0^\infty \min\{\alpha, \mathbb{P}(|X|>x), \mathbb{P}(|Y|>y)\}\mathrm{d}x\mathrm{d}y \\
&\le 2\int_0^\alpha \left(\int_0^\infty \mathbb{1}\{u<\mathbb{P}(|X|>x)\}\,\mathrm{d}x\right)\left(\int_0^\infty \mathbb{1}\{u<\mathbb{P}(|Y|>y)\}\,\mathrm{d}y\right)\mathrm{d}u \\
&= 2\int_0^\alpha \mathrm{d}u \int_0^\infty \mathbb{1}\{Q_X(u)>x\}\,\mathrm{d}x \int_0^\infty \mathbb{1}\{Q_Y(u)>y\}\,\mathrm{d}y \\
&= 2\int_0^\alpha Q_X(u)Q_Y(u)\mathrm{d}u\,.
\end{aligned}
$$

The proof is concluded by applying Hölder's inequality. □

Lemma 21.A.2 *Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space. Let X be a real-valued random variable and V a uniformly distributed random variable independent of X defined on $(\Omega, \mathscr{A}, \mathbb{P})$. Define $F_X(x^-) = \lim_{\substack{y\to x\\ y<x}} F_X(y)$, $\Delta F_X(x) = F_X(x) - F_X(x^-)$, where F_X is the cumulative distribution function and*

$$
U = 1 - F_X(X^-) - V\Delta F_X(X)\,.
$$

Then U is uniformly distributed and $Q_X(U) = X$ $\mathbb{P}-\text{a.s.}$, where Q_X is the tail quantile function.

Proof. That $Q_X(U) = X$ is straightforward, since by definition, $Q_X(v) = x$ for all $v \in [1-F_X(x^-), 1-F_X(x)]$, whether there is a jump at x or not. To check that U is uniformly distributed over $[0,1]$, note that $\mathbb{P}(X>x)>u$ if and only if $Q_X(u)>x$. Since V is uniformly distributed on $[0,1]$, this yields

$$
\begin{aligned}
&\mathbb{P}(U>u) \\
&= \mathbb{P}(1-F_X(X)>u) + \mathbb{P}\left(X = Q_X(u), F_X(F_X^{\leftarrow}(u)^-) + V\Delta F_X(F_X(F_X^{\leftarrow}(u)^-)) \le u\right) \\
&= F_X(Q_X(u)^-) + \mathbb{P}(X = Q_X(u))\frac{1-F_X(Q_X(u)^-)-u}{F_X(Q_X(u)^-)-F_X(Q_X(u))} = 1-u\,.
\end{aligned}
$$

□

Lemma 21.A.3 *Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space and $\mathscr{B}$ a sub-σ-algebra of $\mathscr{A}$. Let X be a square-integrable random variable and $Y = \mathbb{E}[X\,|\,\mathscr{B}]$. Then for all $a \in [0,1]$,*

$$
\int_0^a Q_Y^2(u)\mathrm{d}u \le \int_0^a Q_X^2(u)\mathrm{d}u\,.
$$

Proof. By Lemma 21.A.2, let V be a uniformly distributed random variable, independent of $\mathscr{B}$ and X, and define $U = 1 - F_Y(Y^-) - V\{F_Y(Y) - F_Y(Y^-)\}$. Set $\mathscr{G} = \mathscr{B} \vee \sigma(V)$. Then $Q_Y(U) = Y$ is $\mathscr{B}$-measurable and $\mathbb{E}[X\,|\,\mathscr{B}] = \mathbb{E}[X\,|\,\mathscr{G}]$ $\mathbb{P}-\text{a.s.}$ Applying Jensen's inequality, we obtain

$$\begin{aligned}\int_0^a Q_Y^2(u)\mathrm{d}u &= \mathbb{E}\left[Q_Y^2(U)\mathbb{1}\{U \le a\}\right] = \mathbb{E}\left[(\mathbb{E}[X \,|\, \mathscr{G}])^2 \mathbb{1}\{U \le a\}\right] \\ &\le \mathbb{E}\left[\mathbb{E}\left[X^2 \,\middle|\, \mathscr{G}\right] \mathbb{1}\{U \le a\}\right] = \mathbb{E}\left[X^2 \mathbb{1}\{U \le a\}\right] \\ &= \int_0^\infty \mathbb{P}(X^2 > x, U \le a)\mathrm{d}x \le \int_0^\infty [\mathbb{P}(X^2 > x) \wedge a]\mathrm{d}x \,.\end{aligned}$$

Noting that $\mathbb{P}(X^2 > x) > u$ if and only if $Q_X^2(u) > x$ and applying Fubini's theorem, we obtain

$$\begin{aligned}\int_0^\infty [\mathbb{P}(X^2 > x) \wedge a]\mathrm{d}x &= \int_0^\infty \left(\int_0^a \mathbb{1}\left\{\mathbb{P}(X^2 > x) > u\right\} \mathrm{d}u\right) \mathrm{d}x \\ &= \int_0^\infty \left(\int_0^a \mathbb{1}\left\{Q_X^2(u) > x\right\} \mathrm{d}u\right) \mathrm{d}x \\ &= \int_0^a \left(\int_0^\infty \mathbb{1}\left\{Q_X^2(u) > x\right\} \mathrm{d}x\right) \mathrm{d}u = \int_0^a Q_X^2(u)\mathrm{d}u \,.\end{aligned}$$

$\square$

Chapter 22
Spectral Theory

Let P be a positive Markov kernel on $\mathsf{X} \times \mathscr{X}$ admitting an invariant distribution π. We have shown in Section 1.6 that P defines an operator on the Banach space $\mathrm{L}^p(\pi)$. Therefore, a natural approach to the properties of P consists in studying the spectral properties of this operator. This is the main theme of this chapter. In Section 22.1, we first define the spectrum of P seen as an operator both on $\mathrm{L}^p(\pi)$, $p \geq 1$, and on an appropriately defined space of complex measures. We will also define the adjoint operator and establish some key relations between the operator norm of the operator and that of its adjoint. In Section 22.2, we discuss geometric and exponential convergence in $\mathrm{L}^2(\pi)$. We show that the existence of an $\mathrm{L}^2(\pi)$-spectral gap implies $\mathrm{L}^2(\pi)$-geometric ergodicity; these two notions are shown to be equivalent if the operator P is self-adjoint in $\mathrm{L}^2(\pi)$ (or equivalently that π is reversible with respect to P). We extend these notions to cover $\mathrm{L}^p(\pi)$ exponential convergence in Section 22.3. In Section 22.4, we introduce the notion of conductance and establish the Cheeger inequality for reversible Markov kernels.

22.1 Spectrum

Let $(\mathsf{X}, \mathscr{X})$ be a measurable space and $\pi \in \mathbb{M}_1(\mathscr{X})$. For $p \geq 1$, denote by $\mathrm{L}^p(\pi)$ the space of **complex-valued** functions such that $\pi(|f|^p) < \infty$, where $|f|$ denotes the modulus of f.

In this chapter, unless otherwise stated, all vector spaces are defined over the field $\mathbb{C}$.

Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. For all $f \in \mathrm{L}^p(\pi)$, $Pf = Pf_R + \mathrm{i}Pf_I$, where f_R and f_I denote the real and imaginary parts of f. Let H be a closed subspace of $\mathrm{L}^p(\pi)$ stable by P, i.e., for all $f \in \mathsf{H}$, $Pf \in \mathsf{H}$. We denote by $P|_{\mathsf{H}}$ the restriction of the Markov kernel P to H. The operator norm of $P|_{\mathsf{H}}$ is given by

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_22

$$|||P|||_{\mathsf{H}} = |||P|_{\mathsf{H}}|||_{\mathrm{L}^p(\pi)} = \sup\left\{ \|Pf\|_{\mathrm{L}^p(\pi)} \,:\, f \in \mathsf{H}, \|f\|_{\mathrm{L}^p(\pi)} \le 1 \right\} . \tag{22.1.1}$$

Denote by $\mathsf{BL}(\mathsf{H})$ the space of bounded linear operators on H. According to Proposition 1.6.3, $P|_{\mathsf{H}} \in \mathsf{BL}(\mathsf{H})$.

Definition 22.1.1 (Resolvent and spectrum) *Let* H *be a closed subspace of* $\mathrm{L}^p(\pi)$ *stable by* P*. The resolvent set of* $P|_{\mathsf{H}}$ *is the set of* $\lambda \in \mathbb{C}$ *for which the operator* $(\lambda\mathrm{I} - P|_{\mathsf{H}})$ *has an inverse in* $\mathsf{BL}(\mathsf{H})$*. The spectrum, denoted by* $\mathrm{Spec}(P|\mathsf{H})$*, of* $P|_{\mathsf{H}}$ *is the complement of the resolvent set.*

A complex number λ *is called an eigenvalue of* $P|_{\mathsf{H}} \in \mathsf{BL}(\mathsf{H})$ *if there exists* $h \in \mathsf{H} \setminus \{0\}$ *such that* $P|_{\mathsf{H}} h = \lambda h$*, or equivalently* $\mathrm{Ker}(\lambda\mathrm{I} - P|_{\mathsf{H}}) \neq \{0\}$*. The vector* h *is called an eigenvector of* P *associated with the eigenvalue* λ*.*

The point spectrum of $\mathrm{Spec}_p(P|\mathsf{H})$ *is the set of the eigenvalues of* $P|_{\mathsf{H}}$*. The dimension of* $\mathrm{Ker}(\lambda\mathrm{I} - P|_{\mathsf{H}})$ *is called the multiplicity of the eigenvalue* λ*.*

It is easily seen that the point spectrum is a subset of the spectrum.

Proposition 22.1.2 *Let* P *be a Markov kernel on* $\mathsf{X} \times \mathscr{X}$*. Assume that* P *admits a unique invariant probability measure* π*. Then for all* $p \ge 1$*, 1 is an eigenvalue of* P*, i.e.,* $1 \in \mathrm{Spec}_p(P|\,\mathrm{L}^p(\pi))$*, with multiplicity* 1.

Proof. Obviously $h = \mathbf{1}$ is an eigenvector of P associated with the eigenvalue 1. If $h \in \mathrm{L}^p(\pi)$ is an eigenvector associated with the eigenvalue 1, then $Ph = h$, i.e., the function h is harmonic. Since $p \ge 1$, this implies $h \in \mathrm{L}^1(\pi)$, which implies that h_R and h_I, the real and imaginary parts of h, are harmonic functions in $\mathrm{L}^1(\pi)$. Then Proposition 5.2.12 shows that $h(x) = \pi(h)$ for π-almost every $x \in \mathsf{X}$. □

Denote by Π the Markov kernel defined by $\Pi(x,A) = \pi(A)$ for every $x \in \mathsf{X}$ and $A \in \mathscr{X}$. The kernel of the operator Π,

$$\mathrm{L}^p_0(\pi) = \{f \in \mathrm{L}^p(\pi) \,:\, \Pi f = 0\} , \tag{22.1.2}$$

plays an important role. Note that Π is a bounded linear operator on $\mathrm{L}^p(\pi)$. It is therefore continuous (by Theorem 22.A.2), and $\mathrm{L}^p_0(\pi) = \mathrm{Ker}(\Pi)$ is closed.

Let $\nu \in \mathbb{M}_{\mathbb{C}}(\mathscr{X})$, where $\mathbb{M}_{\mathbb{C}}(\mathscr{X})$ denotes the set of complex measures on $(\mathsf{X}, \mathscr{X})$. We say that ν is dominated by π, which we denote by $\nu \ll \pi$, if for all $A \in \mathscr{X}$, $\pi(A) = 0$ implies $\nu(A) = 0$. The Radon–Nikodym theorem shows that ν admits a density $\mathrm{d}\nu/\mathrm{d}\pi$ with respect to π. For $q \in [1, \infty]$, define

$$\|\nu\|_{\mathbb{M}_q(\pi)} = \begin{cases} \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\pi}\right\|_{\mathrm{L}^q(\pi)} , & |\nu| \ll \pi , \\ \infty , & \text{otherwise} , \end{cases} \tag{22.1.3}$$

and define

$$\mathbb{M}_q(\pi) = \left\{ \nu \in \mathbb{M}_{\mathbb{C}}(\mathscr{X}) : \|\nu\|_{\mathbb{M}_q(\pi)} < \infty \right\}. \tag{22.1.4}$$

For notational simplicity, the dependence of $\mathbb{M}_q(\pi)$ on $\mathscr{X}$ is implicit. The space $\mathbb{M}_q(\pi)$ is a Banach space that is isometrically isomorphic to $\mathrm{L}^q(\pi)$:

$$\|\nu\|_{\mathbb{M}_q(\pi)} = \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\pi} \right\|_{\mathrm{L}^q(\pi)} .$$

For $\nu \in \mathbb{M}_q(\pi)$, we get

$$\|\nu\|_{\mathrm{TV}} = \int \left| \frac{\mathrm{d}\nu}{\mathrm{d}\pi}(x) \right| \pi(\mathrm{d}x) \leq \left\{ \int \left| \frac{\mathrm{d}\nu}{\mathrm{d}\pi}(x) \right|^q \pi(\mathrm{d}x) \right\}^{1/q} = \|\nu\|_{\mathbb{M}_q(\pi)} < \infty . \tag{22.1.5}$$

Denote by $\mathbb{M}_q^0(\pi)$ the subset of signed measures whose total mass is equal to zero:

$$\mathbb{M}_q^0(\pi) = \left\{ \nu \in \mathbb{M}_q(\pi) : \nu(\mathsf{X}) = 0 \right\} . \tag{22.1.6}$$

Lemma 22.1.3 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. Let $q \in [1, \infty]$. For all $\nu \in \mathbb{M}_q(\pi)$ we have $\nu P \in \mathbb{M}_q(\pi)$ and*

$$\|\nu P\|_{\mathbb{M}_q(\pi)} \leq \|\nu\|_{\mathbb{M}_q(\pi)} .$$

Proof. Let ν be a finite signed measure dominated by π. We first show that νP is dominated by π. Let $N \in \mathscr{X}$ be such that $\pi(N) = 0$. Since π is invariant for P, we have $\int \pi(\mathrm{d}x) P(x,N) = \pi(N) = 0$, showing that $P(x,N) = 0$ for π-almost every $x \in \mathsf{X}$. We have

$$\nu P(N) = \int \nu(\mathrm{d}x) P(x,N) = \int \pi(\mathrm{d}x) \frac{\mathrm{d}\nu}{\mathrm{d}\pi}(x) P(x,N) = 0 ,$$

showing that νP is also dominated by π.

We now prove that if $\mathrm{d}\nu/\mathrm{d}\pi \in \mathrm{L}^q(\pi)$, then $\mathrm{d}(\nu P)/\mathrm{d}\pi \in \mathrm{L}^q(\pi)$ and

$$\|\mathrm{d}(\nu P)/\mathrm{d}\pi\|_{\mathrm{L}^q(\pi)} \leq \|\mathrm{d}\nu/\mathrm{d}\pi\|_{\mathrm{L}^q(\pi)} .$$

For all $f \in \mathbb{M}_{\mathbb{C}}(\mathscr{X})$, we get

$$\begin{aligned}
\int |f(y)| \left| \frac{\mathrm{d}(\nu P)}{\mathrm{d}\pi}(y) \right| \pi(\mathrm{d}y) &= \int |f(y)| |\nu P|(\mathrm{d}y) \\
&\leq \iint |f(y)| \left| \frac{\mathrm{d}\nu}{\mathrm{d}\pi}(x) \right| \pi(\mathrm{d}x) P(x, \mathrm{d}y) = \int \left| \frac{\mathrm{d}\nu}{\mathrm{d}\pi}(x) \right| P|f|(x) \pi(\mathrm{d}x) .
\end{aligned} \tag{22.1.7}$$

Assume first that $q \in (1,\infty)$. Let p be such that $p^{-1} + q^{-1} = 1$. Choose $f \in \mathrm{L}^p(\pi)$. Applying Hölder's inequality to (22.1.7) and using Proposition 1.6.3 yields

$$\int |f(y)| \left| \frac{\mathrm{d}(\nu P)}{\mathrm{d}\pi}(y) \right| \pi(\mathrm{d}y) \le \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\pi} \right\|_{\mathrm{L}^q(\pi)} \|P|f|\|_{\mathrm{L}^p(\pi)} \le \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\pi} \right\|_{\mathrm{L}^q(\pi)} \|f\|_{\mathrm{L}^p(\pi)} < \infty .$$

Using Lemma B.2.13, $\mathrm{d}(\nu P)/\mathrm{d}\pi$ belongs to $\mathrm{L}^q(\pi)$ and

$$\|\mathrm{d}(\nu P)/\mathrm{d}\pi\|_{\mathrm{L}^q(\pi)} \le \|\mathrm{d}\nu/\mathrm{d}\pi\|_{\mathrm{L}^q(\pi)} .$$

Consider now the case $q = 1$. Applying (22.1.7) with $f = \mathbf{1}$, we directly obtain

$$\|\mathrm{d}(\nu P)/\mathrm{d}\pi\|_{\mathrm{L}^1(\pi)} \le \|\mathrm{d}\nu/\mathrm{d}\pi\|_{\mathrm{L}^1(\pi)} .$$

To complete the proof, we now consider the case $q = \infty$. Let $A \in \mathscr{X}$. Applying (22.1.7) with $f = \mathbb{1}_A$ and noting that $\pi P \mathbb{1}_A = \pi(A)$ yields

$$\int \mathbb{1}_A(y) \left| \frac{\mathrm{d}(\nu P)}{\mathrm{d}\pi}(y) \right| \pi(\mathrm{d}y) \le \int \left| \frac{\mathrm{d}\nu}{\mathrm{d}\pi}(x) \right| P\mathbb{1}_A(x)\pi(\mathrm{d}x) \le \operatorname{esssup}_\pi (\mathrm{d}\nu/\mathrm{d}\pi)\, \pi(A) .$$

Then taking $A_\delta = \{y \in \mathsf{X} : |\mathrm{d}(\nu P)/\mathrm{d}\pi(y)| > \delta\}$ where $\delta < \operatorname{esssup}_\pi (\mathrm{d}(\nu P)/\mathrm{d}\pi)$, we get

$$\delta \pi(A_\delta) \le \operatorname{esssup}_\pi (\mathrm{d}\nu/\mathrm{d}\pi)\, \pi(A_\delta) .$$

The proof is complete, since $\pi(A_\delta) \ne 0$ and δ is an arbitrary real number strictly less than $\operatorname{esssup}_\pi (\mathrm{d}(\nu P)/\mathrm{d}\pi)$. □

Lemma 22.1.3 shows that the Markov kernel P can also be considered a bounded linear operator on the measure space $\mathbb{M}_q(\pi)$, where $q \in [1, \infty]$. The operator norm of P, considered as a bounded linear operator on the measure space $\mathbb{M}_q(\pi)$, is given by

$$|||P|||_{\mathbb{M}_q(\pi)} = \sup \left\{ \|\nu P\|_{\mathbb{M}_p(\pi)} : \nu \in \mathbb{M}_q(\pi), \|\nu\|_{\mathbb{M}_p(\pi)} \le 1 \right\} .$$

We now provide an explicit expression for $\mathrm{d}(\nu P)/\mathrm{d}\pi$ for all $\nu \in \mathbb{M}_1(\pi)$. This requires that we introduce the adjoint operator of P. We shall use the following definition.

We say that (p, q) are *conjugate real numbers* if $p, q \in [1, \infty]$ and $p^{-1} + q^{-1} = 1$, where we use the convention $\infty^{-1} = 0$.

If $f \in \mathrm{L}^1(\pi)$ and $f \ge 0$, we may define the finite measure π_f by

$$\pi_f(A) = \int \pi(\mathrm{d}x) f(x) P(x, A) \qquad A \in \mathscr{X} . \tag{22.1.8}$$

We can now extend the definition of π_f to any real-valued function $f \in \mathrm{L}^1(\pi)$ (f is no longer assumed to be nonnegative) by setting $\pi_f = \pi_{f^+} - \pi_{f^-}$. For f a complex-valued function in $\mathrm{L}^1\, \pi$, we set $\pi_f = \pi_{f_R} + \mathrm{i}\pi_{f_I}$, where (f_R, f_I) are the real and imaginary parts of f, respectively.

If $\pi(A) = 0$, then $P(x,A) = 0$ for π-almost all $x \in \mathsf{X}$, and this implies $\pi_f(A) = 0$. The complex measure π_f is thus dominated by π, and we can define the adjoint operator P^* as follows.

Definition 22.1.4 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. The adjoint operator $P^* : \mathrm{L}^1(\pi) \to \mathrm{L}^1(\pi)$ of P is defined for all $f \in \mathrm{L}^1(\pi)$ by*

$$P^* f = \frac{\mathrm{d}\pi_f}{\mathrm{d}\pi} , \tag{22.1.9}$$

where π_f is a complex measure defined by

$$\pi_f(A) = \int \pi(\mathrm{d}x) f(x) P(x,A) , \quad A \in \mathscr{X} . \tag{22.1.10}$$

By definition, $P^*\mathbf{1} = \mathbf{1}$. Since $|\mathrm{d}\pi_f/\mathrm{d}\pi| \leq \mathrm{d}\pi_{|f|}/\mathrm{d}\pi \quad \pi - \text{a.s.}$, (22.1.9) implies that $|P^* f| \leq P^*|f|$. Note that P^* is actually a bounded linear operator, since it is clearly linear and

$$\int \pi(\mathrm{d}x)|P^* f(x)| = \int \pi(\mathrm{d}x) P^*|f|(x) \leq \int \pi(\mathrm{d}x) \frac{\mathrm{d}\pi_{|f|}}{\mathrm{d}\pi}(x) = \|f\|_{\mathrm{L}^1(\pi)} .$$

Since $|\pi_f| \leq \pi_{|f|}$, we have the inclusion $\mathrm{L}^1(\pi_{|f|}) \subset \mathrm{L}^1(|\pi_f|)$. Then by definition of the Radon–Nikodym derivative, we obtain for all $g \in \mathrm{L}^1(\pi_{|f|})$, the duality equality

$$\begin{aligned} \int \pi(\mathrm{d}x) f(x) \overline{Pg(x)} &= \int \pi(\mathrm{d}x) f(x) P\bar{g}(x) = \pi_f(\bar{g}) \\ &= \int \pi(\mathrm{d}x) P^* f(x) \overline{g(x)} . \end{aligned} \tag{22.1.11}$$

Proposition 22.1.5 *(i) Let $(f,g) \in \mathrm{L}^p(\pi) \times \mathrm{L}^q(\pi)$, where (p,q) are conjugate. Then*

$$\int \pi(\mathrm{d}x) f(x) \overline{Pg(x)} = \int \pi(\mathrm{d}x) P^* f(x) \overline{g(x)} . \tag{22.1.12}$$

(ii) For all $p \in [1,\infty]$, $\mathrm{L}^p(\pi)$ is stable by P^* and $P^* \in \mathsf{BL}(\mathrm{L}^p(\pi))$.
(iii) For all conjugate real numbers (p,q), $\alpha, \beta \in \mathbb{C}$, and Markov kernels P, Q on $\mathsf{X} \times \mathscr{X}$ with π as invariant probability measure, we have

$$|||\bar{\alpha} P^* + \bar{\beta} Q^*|||_{\mathrm{L}^q(\pi)} = |||\alpha P + \beta Q|||_{\mathrm{L}^p(\pi)} .$$

Moreover, for all $n \in \mathbb{N}$, $(P^*)^n = (P^n)^*$.

Proof. (i) Using (22.1.11), it is sufficient to prove that $g \in \mathrm{L}^1(\pi_{|f|})$. More precisely, we will establish that for all $(f,g) \in \mathrm{L}^p(\pi) \times \mathrm{L}^q(\pi)$, where (p,q) are conjugate,

$$\pi_{|f|}(|g|) \leq \|f\|_{\mathrm{L}^p(\pi)} \|g\|_{\mathrm{L}^q(\pi)} < \infty \,. \tag{22.1.13}$$

If $p \in (1,\infty]$, then using Hölder's inequality and Lemma 1.6.2, we obtain for all $(f,g) \in \mathrm{L}^p(\pi) \times \mathrm{L}^q(\pi)$,

$$\pi_{|f|}(|g|) = \int \pi(\mathrm{d}x)|f|(x)P|g|(x) \leq \|f\|_{\mathrm{L}^p(\pi)} \|P|g|\|_{\mathrm{L}^q(\pi)} \leq \|f\|_{\mathrm{L}^p(\pi)} \|g\|_{\mathrm{L}^q(\pi)} \,.$$

To complete the proof of (22.1.13), consider the case $(f,g) \in \mathrm{L}^1(\pi) \times \mathrm{L}^\infty(\pi)$. For all $\rho > \operatorname{esssup}_\pi(|g|)$, set $A_\rho = \{|g| > \rho\}$. Then $\pi(A_\rho) = 0$, and thus $\pi_{|f|}(A_\rho) = 0$. This implies

$$\pi_{|f|}(|g|) \leq \rho \pi_{|f|}(\mathsf{X}) + \pi_{|f|}(|g| \mathbb{1}_{A_\rho}) = \rho \|f\|_{\mathrm{L}^1(\pi)} \,,$$

and since ρ is an arbitrary real number strictly greater than $\operatorname{esssup}_\pi(|g|)$, this implies $\pi_{|f|}(|g|) \leq \|g\|_{\mathrm{L}^\infty(\pi)} \|f\|_{\mathrm{L}^1(\pi)} < \infty$. This completes the proof of (i).

(ii) Applying (22.1.13), for all $(f,g) \in \mathrm{L}^p(\pi) \times \mathrm{L}^q(\pi)$, where (p,q) are conjugate,

$$\left| \int \pi(\mathrm{d}x) P^* f(x) \overline{g(x)} \right| = \left| \int \pi(\mathrm{d}x) f(x) \overline{Pg(x)} \right| \leq \pi_{|f|}(|g|) \leq \|f\|_{\mathrm{L}^p(\pi)} \|g\|_{\mathrm{L}^q(\pi)} < \infty \,.$$

Applying Lemma B.2.13, we get $\|P^* f\|_{\mathrm{L}^p(\pi)} \leq \|f\|_{\mathrm{L}^p(\pi)}$, showing that $\mathrm{L}^p(\pi)$ is stable by P^* and $P^* \in \mathsf{BL}(\mathrm{L}^p(\pi))$.

(iii) By Lemma B.2.13, if $g \in \mathrm{L}^q(\mu) < \infty$, then

$$\|g\|_{\mathrm{L}^q(\mu)} = \sup \left\{ \left| \int f \bar{g} \mathrm{d}\mu \right| \,:\, \|f\|_{\mathrm{L}^p(\mu)} \leq 1 \right\} \,.$$

This implies that

$$\begin{aligned}
|||\alpha P^* + \beta Q^*|||_{\mathrm{L}^q(\pi)} &= \sup \left\{ \|(\alpha P^* + \beta Q^*) g\|_{\mathrm{L}^q(\pi)} \,:\, \|g\|_{\mathrm{L}^q(\pi)} \leq 1 \right\} \\
&= \sup \left\{ \bar{\alpha} \int f \overline{P^* g} \mathrm{d}\pi + \bar{\beta} \int f \overline{Q^* g} \mathrm{d}\pi \,:\, \|f\|_{\mathrm{L}^p(\pi)} \leq 1, \|g\|_{\mathrm{L}^q(\pi)} \leq 1 \right\} \\
&= \sup \left\{ \bar{\alpha} \int P f \bar{g} \mathrm{d}\pi + \bar{\beta} \int Q f \bar{g} \mathrm{d}\pi \,:\, \|f\|_{\mathrm{L}^p(\pi)} \leq 1, \|g\|_{\mathrm{L}^q(\pi)} \leq 1 \right\} \\
&= \sup \left\{ \left\| (\bar{\alpha} P + \bar{\beta} Q) f \right\|_{\mathrm{L}^p(\pi)} \,:\, \|f\|_{\mathrm{L}^p(\pi)} \leq 1 \right\} = |||\bar{\alpha} P + \bar{\beta} Q|||_{\mathrm{L}^p(\pi)} \,.
\end{aligned}$$

If $f \in \mathrm{L}^p(\pi)$, then $P^* f$ is the only element in $\mathrm{L}^p(\pi)$ that satisfies (22.1.12) for all $g \in \mathrm{L}^q(\pi)$. Using this property, we obtain, by an easy recurrence, $(P^*)^n = (P^n)^*$. □

As a consequence of (22.1.12), for all $p \in [1,\infty]$ and $f \in \mathrm{L}^p(\pi)$, $P^* f$ is the only element in $\mathrm{L}^p(\pi)$ such that for all $g \in \mathrm{L}^q(\pi)$,

$$\int \pi(\mathrm{d}x) f(x) \overline{Pg(x)} = \int \pi(\mathrm{d}x) P^* f(x) \overline{g(x)} \,.$$

Remark 22.1.6. Assume that the Markov kernel P is dominated by a σ-finite measure μ, that is, suppose that for all $(x,A) \in \mathsf{X} \times \mathscr{X}$, $P(x,A) = \int_A p(x,y)\mu(\mathrm{d}y)$, where $(x,y) \mapsto p(x,y)$ is $(\mathscr{X} \times \mathscr{X}, \mathscr{B}(\mathbb{R}))$-measurable. In this case, π is dominated by μ. Denoting by h_π the density of π with respect to μ, we have for π-almost all $y \in \mathsf{X}$,

$$P^* f(y) = \frac{\mathrm{d}\pi_f}{\mathrm{d}\pi}(y) = \frac{h_\pi(x) f(x) p(x,y)}{h_\pi(y)} \,.$$

▲

If the complex measure ν is dominated by π, that is, if $\nu \in \mathbb{M}_1(\pi)$, then plugging $f = \mathrm{d}\nu/\mathrm{d}\pi$ into (22.1.10) yields $\pi_f = \nu P$, and by (22.1.9), we get

$$P^* \left(\frac{\mathrm{d}\nu}{\mathrm{d}\pi} \right) = P^* f = \frac{\mathrm{d}(\nu P)}{\mathrm{d}\pi} \,. \tag{22.1.14}$$

Lemma 22.1.7 *Let P, Q be Markov kernels on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. For all conjugate real numbers (p,q),*

$$|||P - Q|||_{\mathrm{L}^p(\pi)} = |||P - Q\,|||_{\mathbb{M}_q(\pi)} \,.$$

Proof. Using (22.1.3), (22.1.14), and Proposition 22.1.5 (iii), we obtain

$$\begin{aligned}
|||P - Q|||_{\mathbb{M}_q(\pi)} &= \sup \left\{ \left\| \frac{\mathrm{d}(\mu P)}{\mathrm{d}\pi} - \frac{\mathrm{d}(\mu Q)}{\mathrm{d}\pi} \right\|_{\mathrm{L}^q(\pi)} : \mu \in \mathbb{M}_q(\pi), \|\mu\|_{\mathbb{M}_q(\pi)} \le 1 \right\} \\
&= \sup \left\{ \left\| (P^* - Q^*) \left(\frac{\mathrm{d}\mu}{\mathrm{d}\pi} \right) \right\|_{\mathrm{L}^q(\pi)} : \frac{\mathrm{d}\mu}{\mathrm{d}\pi} \in \mathrm{L}^q(\pi), \left\| \frac{\mathrm{d}\mu}{\mathrm{d}\pi} \right\|_{\mathrm{L}^q(\pi)} \le 1 \right\} \\
&= |||P^* - Q^*|||_{\mathrm{L}^q(\pi)} = |||P - Q\,|||_{\mathrm{L}^p(\pi)} \,.
\end{aligned}$$

□

Recall that Π is the Markov kernel defined by $\Pi(x,A) = \pi(A)$, where $A \in \mathscr{X}$. By Lemma 22.1.7, for all conjugate real numbers (p,q) and all $n \in \mathbb{N}$,

$$|||P^n - \Pi|||_{\mathrm{L}^p(\pi)} = |||P^n - \Pi\,|||_{\mathbb{M}_q(\pi)} \,.$$

Since $\Pi P = P\Pi = \Pi$, we get $P^n - \Pi = (P - \Pi)^n$, and the previous identity also implies that

$$|||(P - \Pi)^n|||_{\mathrm{L}^p(\pi)} = |||P^n - \Pi|||_{\mathrm{L}^p(\pi)} = |||P^n - \Pi|||_{\mathbb{M}_q(\pi)} = |||(P - \Pi)^n\,|||_{\mathbb{M}_q(\pi)} \,.$$

We formalize these results in the following theorem for future reference.

Theorem 22.1.8. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. For all conjugate real numbers (p,q) and all $n \in \mathbb{N}$,*

$$|||(P-\Pi)^n|||_{\mathrm{L}^p(\pi)} = |||P^n - \Pi|||_{\mathrm{L}^p(\pi)} = |||P^n - \Pi|||_{\mathbb{M}_q(\pi)} = |||(P-\Pi)^n\,|||_{\mathbb{M}_q(\pi)} \; .$$

In particular, for all probability measures $\nu \in \mathbb{M}_q(\pi)$ and all $n \in \mathbb{N}$,

$$\|\nu P^n - \pi\|_{\mathrm{TV}} \le \|\nu P^n - \pi\|_{\mathbb{M}_q(\pi)} \le |||(P-\Pi)^n\,|||_{\mathrm{L}^p(\pi)} \, \|\nu - \pi\|_{\mathbb{M}_q(\pi)} \; .$$

We conclude this section with a link between self-adjointness and reversibility. We first need the following definition.

Definition 22.1.9 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. We say that P is* self-adjoint *on* $\mathrm{L}^2(\pi)$ *if for all $f \in \mathrm{L}^2(\pi)$, $Pf = P^* f$, that is, $P = P^*$.*

Lemma 22.1.10 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. Then P is self-adjoint if and only if π is reversible with respect to P.*

Proof. Assume that P is self-adjoint. Then applying (22.1.12) with $f = \mathbb{1}_A$ and $g = \mathbb{1}_B$, we get for all $A, B \in \mathscr{X}$,

$$\pi \otimes P(A \times B) = \int \pi(\mathrm{d}x) \mathbb{1}_A(x) P(x,B) = \int \pi(\mathrm{d}x) \mathbb{1}_B(x) P(x,A) = \pi \otimes P(B \times A) \; ,$$

showing that π is reversible for P. Conversely, assume that π is reversible with respect to P. Then for $f \in \mathrm{L}^1(\pi)$ and $B \in \mathscr{X}$,

$$\pi_f(B) = \int \pi(\mathrm{d}x) f(x) P(x,B) = \int \pi(\mathrm{d}x) \mathbb{1}_B(x) Pf(x) \; .$$

This shows that $Pf = \mathrm{d}\pi_f / \mathrm{d}\pi = P^* f$, and the proof is concluded. □

22.2 Geometric and Exponential Convergence in $\mathrm{L}^2(\pi)$

In this section, we consider P to be an operator on the Hilbert space $\mathrm{L}^2(\pi)$ equipped with the scalar product

$$\langle f, g \rangle_{\mathrm{L}^2(\pi)} = \int f(x) \overline{g(x)} \pi(\mathrm{d}x) \; , \qquad \|f\|^2_{\mathrm{L}^2(\pi)} = \langle f, f \rangle_{\mathrm{L}^2(\pi)} \; . \tag{22.2.1}$$

Because $\mathrm{L}^2_0(\pi)$ is closed, the space $\mathrm{L}^2(\pi)$ may be decomposed as

$$\mathrm{L}^2(\pi) = \mathrm{L}^2_0(\pi) \overset{\perp}{\oplus} \{\mathrm{L}^2_0(\pi)\}^{\perp} . \tag{22.2.2}$$

For all $f \in \mathrm{L}^2_0(\pi)$, we get $\pi P(f) = \pi(f) = 0$, showing that $\mathrm{L}^2_0(\pi)$ is stable by P. And since $P\mathbf{1} = \mathbf{1}$, the subspace $\{\mathrm{L}^2_0(\pi)\}^{\perp}$ is also stable by P. The orthogonal projection on these two spaces is then explicitly given as follows: for $f \in \mathrm{L}^2(\pi)$,

$$f = \{f - \langle f, \mathbf{1}\rangle_{\mathrm{L}^2(\pi)} \mathbf{1}\} + \langle f, \mathbf{1}\rangle_{\mathrm{L}^2(\pi)} \mathbf{1} .$$

The operator $(\lambda \mathrm{I} - P)$ is invertible if and only if $\lambda \mathrm{I} - P|_{\mathrm{L}^2_0(\pi)}$ and $\lambda \mathrm{I} - P|_{\{\mathrm{L}^2_0(\pi)\}^{\perp}}$ are invertible. As a consequence,

$$\begin{aligned} \mathrm{Spec}(P|\mathrm{L}^2(\pi)) &= \mathrm{Spec}(P|\mathrm{L}^2_0(\pi)) \cup \mathrm{Spec}(P|\{\mathrm{L}^2_0(\pi)\}^{\perp}) \\ &= \{1\} \cup \mathrm{Spec}(P|\mathrm{L}^2_0(\pi)) . \end{aligned} \tag{22.2.3}$$

By Proposition 22.1.5 (i), the adjoint operator P^* of P satisfies for all $f, g \in \mathrm{L}^2(\pi)$,

$$\langle Pf, g\rangle_{\mathrm{L}^2(\pi)} = \langle f, P^* g\rangle_{\mathrm{L}^2(\pi)} . \tag{22.2.4}$$

Lemma 22.2.1 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. For $p \in [1, \infty]$,*

$$|||P|||_{\mathrm{L}^p_0(\pi)} \leq |||P - \Pi|||_{\mathrm{L}^p(\pi)} \leq 2\, |||\, P\, |||_{\mathrm{L}^p_0(\pi)} . \tag{22.2.5}$$

Moreover,

$$|||P|||_{\mathrm{L}^2_0(\pi)} = |||P - \Pi\, |||_{\mathrm{L}^2(\pi)} . \tag{22.2.6}$$

Proof. Note first that

$$\begin{aligned} |||P|||_{\mathrm{L}^p_0(\pi)} &= \sup_{\|g\|_{\mathrm{L}^p(\pi)} \leq 1, \Pi(g) = 0} \|Pg\|_{\mathrm{L}^p(\pi)} = \sup_{\|g\|_{\mathrm{L}^p(\pi)} \leq 1, \Pi(g) = 0} \|Pg - \Pi(g)\|_{\mathrm{L}^p(\pi)} \\ &\leq \sup_{\|f\|_{\mathrm{L}^p(\pi)} \leq 1} \|(P - \Pi) f\|_{\mathrm{L}^p(\pi)} = |||P - \Pi\, |||_{\mathrm{L}^p(\pi)} . \end{aligned}$$

To establish the upper bound in (22.2.5), it suffices to notice that

$$\begin{aligned} |||P - \Pi|||_{\mathrm{L}^p(\pi)} &= 2 \sup_{\|f\|_{\mathrm{L}^p(\pi)} \leq 1} \|P\{(1/2)(f - \Pi f)\}\|_{\mathrm{L}^p(\pi)} \\ &\leq 2 \sup_{\|g\|_{\mathrm{L}^p(\pi)} \leq 1, \Pi(g) = 0} \|Pg\|_{\mathrm{L}^p(\pi)} = 2\, |||\, P\, |||_{\mathrm{L}^p_0(\pi)} . \end{aligned}$$

If $p = 2$, the decomposition (22.2.2) yields $\|f - \Pi(f)\|_{\mathrm{L}^2(\pi)} \leq \|f\|_{\mathrm{L}^2(\pi)}$, which in turn implies

$$
\begin{aligned}
|||P-\Pi|||_{\mathrm{L}^2(\pi)} &= \sup_{\|f\|_{\mathrm{L}^2(\pi)}\le 1} \|(P-\Pi)f\|_{\mathrm{L}^2(\pi)} = \sup_{\|f\|_{\mathrm{L}^2(\pi)}\le 1} \|P\{f-\Pi(f)\}\|_{\mathrm{L}^2(\pi)} \\
&\le \sup_{\|g\|_{\mathrm{L}^2(\pi)}\le 1, \Pi(g)=0} \|Pg\|_{\mathrm{L}^2(\pi)} = |||P|||_{\mathrm{L}^2_0(\pi)} .
\end{aligned}
$$

This proves (22.2.6). □

The space $\mathbb{M}_2(\pi)$ is a Hilbert space equipped with the inner product

$$
(\nu,\mu)_{\mathbb{M}_2(\pi)} = \int \frac{\mathrm{d}\nu}{\mathrm{d}\pi}(x)\overline{\frac{\mathrm{d}\mu}{\mathrm{d}\pi}}(x)\pi(\mathrm{d}x) = \left\langle \frac{\mathrm{d}\nu}{\mathrm{d}\pi}, \frac{\mathrm{d}\mu}{\mathrm{d}\pi} \right\rangle_{\mathrm{L}^2(\pi)} .
$$

Using this notation, we may decompose the space $\mathbb{M}_2(\pi)$ as follows:

$$
\mathbb{M}_2(\pi) = \mathbb{M}_2^0(\pi) \overset{\perp}{\oplus} \left\{\mathbb{M}_2^0(\pi)\right\}^{\perp} ,
$$

where $\mathbb{M}_2^0(\pi)$ is defined in (22.1.6). The orthogonal projections on these two subspaces are again explicit by writing for $\mu \in \mathbb{M}_2(\pi)$, $\mu = \{\mu - \mu(\mathsf{X})\pi\} + \mu(\mathsf{X})\pi$. Then

$$
\left\{\mathbb{M}_2^0(\pi)\right\}^{\perp} = \{\nu \in \mathbb{M}_2(\pi) : \ \nu = c\cdot\pi,\ c \in \mathbb{R}\} .
$$

The space $\mathbb{M}_2^0(\pi)$ (respectively $\left\{\mathbb{M}_2^0(\pi)\right\}^{\perp}$) is also a Hilbert space that is isometrically isomorphic to $\mathrm{L}^2_0(\pi)$ (respectively $\left\{\mathrm{L}^2_0(\pi)\right\}^{\perp}$). Recall that for $\nu \in \mathbb{M}_2(\pi)$, we have by (22.1.14),

$$
\frac{\mathrm{d}(\nu P)}{\mathrm{d}\pi} = P^* \frac{\mathrm{d}\nu}{\mathrm{d}\pi} .
$$

Moreover, denoting by νP^* the measure $A \mapsto \nu(P^* \mathbb{1}_A)$, we get by (22.1.12),

$$
\nu P^*(A) = \int \pi(\mathrm{d}x)\frac{\mathrm{d}\nu}{\mathrm{d}\pi}(x)P^*\mathbb{1}_A(x) = \int \pi(\mathrm{d}x)\mathbb{1}_A(x)P\overline{\frac{\mathrm{d}\nu}{\mathrm{d}\pi}}(x) ,
$$

showing that $\mathrm{d}(\nu P^*)/\mathrm{d}\pi = P\mathrm{d}\nu/\mathrm{d}\pi$. Now for all $\mu,\nu \in \mathbb{M}_2(\pi)$, we then have

$$
\begin{aligned}
(\nu P,\mu)_{\mathbb{M}_2(\pi)} &= \left\langle \frac{\mathrm{d}(\nu P)}{\mathrm{d}\pi}, \frac{\mathrm{d}\mu}{\mathrm{d}\pi} \right\rangle_{\mathrm{L}^2(\pi)} \\
&= \left\langle P^*\frac{\mathrm{d}\nu}{\mathrm{d}\pi}, \frac{\mathrm{d}\mu}{\mathrm{d}\pi} \right\rangle_{\mathrm{L}^2(\pi)} = \left\langle \frac{\mathrm{d}\nu}{\mathrm{d}\pi}, P\frac{\mathrm{d}\mu}{\mathrm{d}\pi} \right\rangle_{\mathrm{L}^2(\pi)} \\
&= \left\langle \frac{\mathrm{d}\nu}{\mathrm{d}\pi}, \frac{\mathrm{d}(\mu P^*)}{\mathrm{d}\pi} \right\rangle_{\mathrm{L}^2(\pi)} = (\nu,\mu P^*)_{\mathbb{M}_2(\pi)} .
\end{aligned}
\tag{22.2.7}
$$

Definition 22.2.2 **($\mathrm{L}^2(\pi)$-geometric ergodicity and exponential convergence)**
Let P be a Markov kernel on $\mathsf{X}\times\mathscr{X}$ with invariant probability π.

(i) P is said to be $\mathrm{L}^2(\pi)$-geometrically ergodic if there exist $\rho \in [0,1)$ and for all probability measures $\nu \in \mathbb{M}_2(\pi)$, a constant $C(\nu) < \infty$ satisfying

$$\|\nu P^n - \pi\|_{\mathbb{M}_2(\pi)} \le C(\nu)\rho^n , \qquad \text{for all } n \in \mathbb{N} .$$

(ii) P is said to be $\mathrm{L}^2(\pi)$-exponentially convergent if there exist $\alpha \in [0,1)$ and $M < \infty$ such that

$$|||P^n - \Pi|||_{\mathrm{L}^2(\pi)} = |||P^n|||_{\mathrm{L}^2_0(\pi)} \le M\alpha^n , \qquad \text{for all } n \in \mathbb{N} . \tag{22.2.8}$$

Note that the equality in (22.2.8) is a consequence of Lemma 22.2.1.

Definition 22.2.3 ($\mathrm{L}^2(\pi)$**-absolute spectral gap)** *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. The Markov kernel P has an $\mathrm{L}^2(\pi)$-absolute spectral gap if*

$$\mathrm{Abs.Gap}_{\mathrm{L}^2(\pi)}(P) := 1 - \sup\left\{|\lambda| \,:\, \lambda \in \mathrm{Spec}(P|\mathrm{L}^2_0(\pi))\right\} > 0 .$$

Proposition 22.2.4 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π.*

(i) P has an $\mathrm{L}^2(\pi)$-absolute spectral gap if and only if there exists $m > 1$ such that $|||P^m|||_{\mathrm{L}^2_0(\pi)} < 1$.
(ii) If P has an $\mathrm{L}^2(\pi)$-absolute spectral gap, then the Markov kernel P is $\mathrm{L}^2(\pi)$-geometrically ergodic.

Proof. (i) Proposition 22.A.13 shows that

$$\sup\left\{|\lambda| \,:\, \lambda \in \mathrm{Spec}(P|\mathrm{L}^2_0(\pi))\right\} = \lim_{m\to\infty} \left\{|||P^m|||_{\mathrm{L}^2_0(\pi)}\right\}^{1/m} .$$

(ii) By (i), there exists $m > 1$ such that $|||P^m|||_{\mathrm{L}^2_0(\pi)} < 1$. By Theorem 22.1.8 and (22.2.6), we get for every probability measure $\nu \in \mathbb{M}_2(\pi)$

$$\begin{aligned}\|\nu P^n - \pi\|_{\mathbb{M}_2(\pi)} &= \|\nu[P^n - \Pi]\|_{\mathbb{M}_2(\pi)} \le |||P^n - \Pi\,|||_{\mathbb{M}_2(\pi)}\, \|\nu\|_{\mathbb{M}_2(\pi)} \\ &= |||P^n - \Pi\,|||_{\mathrm{L}^2(\pi)}\, \|\nu\|_{\mathbb{M}_2(\pi)} = |||P^n\,|||_{\mathrm{L}^2_0(\pi)}\, \|\nu\|_{\mathbb{M}_2(\pi)} .\end{aligned}$$

The proof is completed by noting that $|||P^n|||_{\mathrm{L}^2_0(\pi)} \le |||P^m|||^{\lfloor n/m \rfloor}_{\mathrm{L}^2_0(\pi)}$. $\square$

We now specialize the results to the case that the probability measure π is reversible with respect to the Markov kernel P. According to Lemma 22.1.10, P is self-adjoint in $\mathrm{L}^2(\pi)$. And using (22.1.12), we get for all functions $f,g \in \mathrm{L}^2(\pi)$,

$$\langle Pf, g\rangle_{\mathrm{L}^2(\pi)} = \int_{\mathsf{X}} \pi(\mathrm{d}x) Pf(x) g(x) = \int_{\mathsf{X}} \pi(\mathrm{d}x) f(x) Pg(x) = \langle Pg, f\rangle_{\mathrm{L}^2(\pi)} .$$

Hence the operator P is self-adjoint in $\mathrm{L}^2(\pi)$. Moreover, (22.2.7) shows that the operator P is also self-adjoint in $\mathbb{M}_2(\pi)$, i.e., for all $\mu, \nu \in \mathbb{M}_2(\pi)$,

$$(\nu P, \mu)_{\mathbb{M}_2(\pi)} = (\nu, \mu P)_{\mathbb{M}_2(\pi)} .$$

Let $\mathsf{H} \subset \mathrm{L}^2(\pi)$ be a subspace of $\mathrm{L}^2(\pi)$ stable by P. By Theorem 22.A.19, the spectrum of the restriction of a self-adjoint operator P to H is included in a segment of the real line defined by

$$\operatorname{Spec}(P|\mathsf{H}) \subset \Big[\inf_{f\in \mathsf{H}, \|f\|_{\mathrm{L}^2(\pi)} \leq 1} \langle Pf, f\rangle , \sup_{f\in \mathsf{H}, \|f\|_{\mathrm{L}^2(\pi)} \leq 1} \langle Pf, f\rangle \Big] . \tag{22.2.9}$$

Proposition 22.2.5 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. Then*

$$1 - |||P|||_{\mathrm{L}^2_0(\pi)} = 1 - |||P - \Pi|||_{\mathrm{L}^2(\pi)} \leq \operatorname{Abs.Gap}_{\mathrm{L}^2(\pi)}(P) ,$$

with equality if P is reversible with respect to π.

Proof. By Proposition 22.A.13,

$$1 - \operatorname{Abs.Gap}_{\mathrm{L}^2(\pi)}(P) = \lim_{m\to\infty} \left\{ |||P^m|||_{\mathrm{L}^2_0(\pi)} \right\}^{1/m} \leq |||P|||_{\mathrm{L}^2_0(\pi)} .$$

This proves the first part of the proposition. Assume now that P is reversible with respect to π and define

$$\lambda_{\min} = \inf\left\{\lambda : \lambda \in \operatorname{Spec}(P|\mathrm{L}^2_0(\pi))\right\}$$
$$\lambda_{\max} = \sup\left\{\lambda : \lambda \in \operatorname{Spec}(P|\mathrm{L}^2_0(\pi))\right\} .$$

By applying (22.2.9) with $\mathsf{H} = \mathrm{L}^2_0(\pi)$, we get

$$1 - \operatorname{Abs.Gap}_{\mathrm{L}^2(\pi)}(P) = \sup\left\{|\lambda| : \lambda \in \operatorname{Spec}(P|\mathrm{L}^2_0(\pi))\right\} = \max\{|\lambda_{\min}|, \lambda_{\max}\} .$$

Moreover, since P is a self-adjoint operator on $\mathrm{L}^2(\pi)$, Theorem 22.A.17 together with Theorem 22.A.19 yields

$$|||P|||_{\mathrm{L}^2_0(\pi)} = \sup\left\{ |\langle Pf, f\rangle| \,:\, \|f\|_{\mathrm{L}^2(\pi)} \le 1, f \in \mathrm{L}^2_0(\pi) \right\} = \max\{|\lambda_{\min}|,\ \lambda_{\max}\},$$

which therefore implies $1 - \mathrm{Abs.Gap}_{\mathrm{L}^2(\pi)}(P) = |||P|||_{\mathrm{L}^2_0(\pi)}$. □

Assume that P is a self-adjoint operator on $\mathrm{L}^2(\pi)$. Theorem 22.B.3 shows that to every function $f \in \mathrm{L}^2(\pi)$ we may associate a unique finite measure on $[-1,1]$ (the spectral measure) satisfying for all $k \in \mathbb{N}$,

$$\mathbb{E}_\pi[f(X_0)\overline{f(X_k)}] = \left\langle f, P^k f \right\rangle_{\mathrm{L}^2(\pi)} = \left\langle P^k f, f \right\rangle_{\mathrm{L}^2(\pi)} = \int_{-1}^{1} t^k \mu_f(\mathrm{d}t) \,. \tag{22.2.10}$$

Applying this relation with $k = 0$ shows that

$$\|f\|_{\mathrm{L}^2(\pi)} = \mathbb{E}_\pi[|f(X_0)|^2] = \mu_f([-1,1]) \,. \tag{22.2.11}$$

Theorem 22.2.6. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ reversible with respect to the probability measure π.*

(i) If for all $f \in \mathrm{L}^2_0(\pi)$, the support of the spectral measure μ_f is included in the interval $[-\rho, \rho]$, $\rho \in [0,1)$, then $\mathrm{Abs.Gap}_{\mathrm{L}^2(\pi)}(P) \ge 1 - \rho$.

(ii) If the Markov kernel P has an $\mathrm{L}^2(\pi)$-absolute spectral gap $\mathrm{Abs.Gap}_{\mathrm{L}^2(\pi)}(P)$*, then for all $f \in \mathrm{L}^2_0(\pi)$, the support of the spectral measure μ_f is included in the interval* $\left[-1 + \mathrm{Abs.Gap}_{\mathrm{L}^2(\pi)}(P), 1 - \mathrm{Abs.Gap}_{\mathrm{L}^2(\pi)}(P)\right]$.

Proof. (i) Let $f \in \mathrm{L}^2_0(\pi)$. By the definition of the spectral measure, we obtain, using (22.2.10) and (22.2.11),

$$\begin{aligned}\|Pf\|^2_{\mathrm{L}^2(\pi)} &= \langle Pf, Pf\rangle_{\mathrm{L}^2(\pi)} = \langle f, P^2 f\rangle_{\mathrm{L}^2(\pi)} \\ &= \int_{-1}^{1} t^2 \mu_f(\mathrm{d}t) \le \rho^2 \mu_f([-1,1]) = \rho^2 \|f\|_{\mathrm{L}^2(\pi)} \,.\end{aligned}$$

Combining this with Proposition 22.2.5 yields $1 - \mathrm{Abs.Gap}_{\mathrm{L}^2(\pi)}(P) = |||P|||_{\mathrm{L}^2_0(\pi)} \le \rho$.

(ii) Conversely, assume $|||P|||_{\mathrm{L}^2_0(\pi)} \le \rho$. By definition, this implies that for all $f \in \mathrm{L}^2_0(\pi)$ and $n \in \mathbb{N}$,

$$\|P^n f\|_{\mathrm{L}^2(\pi)} \le \rho^n \|f\|_{\mathrm{L}^2(\pi)} \,. \tag{22.2.12}$$

We now prove by contradiction that μ_f is supported by $[-\rho, \rho]$. Assume that there exist $f \in \mathrm{L}^2_0(\pi)$ and $r \in (\rho, 1]$ such that $\mu_f(I_r) > 0$, where $I_r = [-1, -r] \cup [r, 1]$. Then since π is reversible with respect to P,

$$\|P^n f\|^2_{\mathrm{L}^2(\pi)} = \int_{-1}^{1} t^{2n} \mu_f(\mathrm{d}t) \ge \int_{I_r} t^{2n} \mu_f(\mathrm{d}t) \ge r^{2n} \mu_f(I_r) \,.$$

This contradicts (22.2.12). □

Theorem 22.2.7. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ reversible with respect to the probability measure π. The following statements are equivalent:*

(i) P is $\mathrm{L}^2(\pi)$-geometrically ergodic.
(ii) P has an $\mathrm{L}^2(\pi)$-absolute spectral gap.

Proof. (i) ⇒ (ii) Assume that P is $\mathrm{L}^2(\pi)$-geometrically ergodic. We first prove an apparently stronger result: for every complex measure $\nu \in \mathbb{M}_2^0(\pi)$, there exist a finite constant $C(\nu)$ and $\rho \in [0,1)$ such that

$$\|\nu P^n\|_{\mathbb{M}_2(\pi)} \leq C(\nu)\rho^n \,, \qquad \text{for all } n \in \mathbb{N} \,. \tag{22.2.13}$$

It suffices to prove that this result holds for real-valued signed measures. Let ν be a nontrivial signed measure belonging to $\mathbb{M}_2^0(\pi)$. Define $f = \mathrm{d}\nu/\mathrm{d}\pi$, which by definition belongs to $\mathrm{L}_0^2(\pi)$. Define $g^+ = f^+/Z$, $g^- = f^-/Z$, where $Z = \pi(f^+) = \pi(f^-)$. Note that $\mu_+ = g^+ \cdot \pi$ and $\mu_- = g^- \cdot \pi$ are two probability measures that belong to $\mathbb{M}_2(\pi)$. Moreover, by (22.1.14), $\mathrm{d}(\nu P^n)/\mathrm{d}\pi = P^n(\mathrm{d}\nu/\mathrm{d}\pi)$. This implies

$$\begin{aligned}\|\nu P^n\|_{\mathbb{M}_2(\pi)} &= \|P^n(\mathrm{d}\nu/\mathrm{d}\pi)\|_{\mathrm{L}^2(\pi)} = \left\|P^n\{Zg^+ - Zg^-\}\right\|_{\mathrm{L}^2(\pi)} \\ &\leq Z\left\|P^n\{g^+ - 1\}\right\|_{\mathrm{L}^2(\pi)} + Z\left\|P^n\{g^- - 1\}\right\|_{\mathrm{L}^2(\pi)} \\ &= Z\|\mu_+ P^n - \pi\|_{\mathbb{M}_2(\pi)} + Z\|\mu_- P^n - \pi\|_{\mathbb{M}_2(\pi)} \,.\end{aligned}$$

Since P is $\mathrm{L}^2(\pi)$-geometrically ergodic, there exist two constants $C(\mu_+) < \infty$ and $C(\mu_-) < \infty$ such that for all $n \in \mathbb{N}$,

$$\|\nu P^n\|_{\mathbb{M}_2(\pi)} \leq Z\{C(\mu_+) + C(\mu_-)\}\rho^n \,,$$

showing that (22.2.13) is satisfied.

Let now $\nu \in \mathbb{M}_2^0(\pi)$ and set $f = \mathrm{d}\nu/\mathrm{d}\pi$, which belongs to $\mathrm{L}_0^2(\pi)$. For all $n \in \mathbb{N}$, we get, using the reversibility of P and (22.2.10),

$$\|\nu P^n\|_{\mathbb{M}_2(\pi)} = \|P^n f\|_{\mathrm{L}^2(\pi)} = \left(\langle f, P^{2n} f\rangle_{\mathrm{L}^2(\pi)}\right)^{1/2} = \left(\int_{-1}^1 t^{2n}\mu_f(\mathrm{d}t)\right)^{1/2} \,,$$

where μ_f is the spectral measure associated with the function f. We must have for all $1 > r > \rho$, $\mu_f([-1,-r] \cup [r,1]) = 0$, for otherwise, we could choose $r \in (\rho, 1)$ such that

$$\|\nu P^n\|_{\mathbb{M}_2(\pi)} = \int_{-1}^1 t^{2n}\mu_f(\mathrm{d}t) \geq r^{2n}\mu_f([-1,-r] \cup [r,1]) \,,$$

which contradicts (22.2.13). Therefore, if P is $\mathrm{L}^2(\pi)$-geometrically ergodic, then for all $\nu \in \mathbb{M}_2^0(\pi)$, the spectral measure of the function $\mathrm{d}\nu/\mathrm{d}\pi$ is included in $[-\rho,\rho]$. Since the space $\mathbb{M}_2^0(\pi)$ is isometrically isomorphic to $\mathrm{L}_0^2(\pi)$, if P is $\mathrm{L}^2(\pi)$-geometrically ergodic, then the spectral measure associated with every function $f \in \mathrm{L}_0^2(\pi)$ must be included in $[-\rho,\rho]$. We conclude by applying Theorem 22.2.6.

(ii) $\Rightarrow$ (i) Follows from Proposition 22.2.4 (note that in this case, the reversibility does not play a role). □

We conclude this section with an extension of this result to a possibly irreversible kernel, provided that the identity $P^*P = PP^*$ is satisfied.

Proposition 22.2.8 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. Assume that the Markov kernel P is normal, i.e., $P^*P = PP^*$. Then $|||P|||_{\mathrm{L}_0^2(\pi)} = \lim_{n\to\infty} |||P^n|||_{\mathrm{L}_0^2(\pi)}^{1/n}$, and the following statements are equivalent:*

(i) P is $\mathrm{L}^2(\pi)$-exponentially convergent.
(ii) P has an $\mathrm{L}^2(\pi)$-absolute spectral gap.

Moreover, if there exist $M < \infty$ and $\alpha \in [0,1)$ such that for all $n \in \mathbb{N}$, $|||P^n|||_{\mathrm{L}_0^2(\pi)} \leq M\alpha^n$, then $1-\alpha \leq \mathrm{Abs.Gap}_{\mathrm{L}^2(\pi)}(P)$.

Proof. Since P is normal, $PP^* = P^*P$, and consequently, for all $n \in \mathbb{N}$, $P^n(P^*)^n = (PP^*)^n$. By Corollary 22.A.18, we get for all $n \geq 1$,

$$|||P^n|||_{\mathrm{L}_0^2(\pi)}^2 = |||P^n(P^*)^n|||_{\mathrm{L}_0^2(\pi)} = |||(PP^*)^n|||_{\mathrm{L}_0^2(\pi)} .$$

Now, using again Corollary 22.A.18 and applying successively Proposition 22.2.5 and Proposition 22.A.13 to the self-adjoint operator PP^*, we get

$$\begin{aligned} |||P|||_{\mathrm{L}_0^2(\pi)}^2 &= |||PP^*|||_{\mathrm{L}_0^2(\pi)} = 1 - \mathrm{Abs.Gap}_{\mathrm{L}^2(\pi)}(PP^*) \\ &= \lim_{n\to\infty} |||(PP^*)^n|||_{\mathrm{L}_0^2(\pi)}^{1/n} \\ &= \lim_{n\to\infty} |||P^n|||_{\mathrm{L}_0^2(\pi)}^{2/n}, \end{aligned} \tag{22.2.14}$$

which concludes the first part of the proof.

(i) $\Rightarrow$ (ii) Assume that P is $\mathrm{L}^2(\pi)$-exponentially convergent. Then there exist $M < \infty$ and $\alpha \in [0,1)$ such that for all $n \in \mathbb{N}$, $|||P^n|||_{\mathrm{L}_0^2(\pi)} \leq M\alpha^n$, which implies by (22.2.14) that $|||P|||_{\mathrm{L}_0^2(\pi)} \leq \alpha < 1$, and the proof of the first implication follows from Proposition 22.2.5.

(ii) $\Rightarrow$ (i) Follows from Proposition 22.2.4. □

22.3 $\mathrm{L}^p(\pi)$-Exponential Convergence

We will now generalize Definition 22.2.2 to $\mathrm{L}^p(\pi)$ for $p \geq 1$.

Definition 22.3.1 (**$\mathrm{L}^p(\pi)$-exponential convergence**) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. Let $p \in [1, \infty]$. The Markov kernel P is said to be $\mathrm{L}^p(\pi)$-exponentially convergent if there exist $\alpha \in [0,1)$ and $M < \infty$ such that for all $n \in \mathbb{N}$,*

$$|||P^n|||_{\mathrm{L}_0^p(\pi)} \leq M\alpha^n .$$

By Proposition 22.A.13, if $|||P^m|||_{\mathrm{L}_0^p(\pi)} < \mathbb{1}$ for some $m \geq 1$, then P is $\mathrm{L}^p(\pi)$-exponentially convergent.

Let (p,q) be conjugate real numbers. As shown in the next proposition, the $\mathrm{L}^p(\pi)$-exponential convergence turns out to imply the convergence of the operator P^n, acting on measures in $\mathbb{M}_q^0(\pi)$, in the following sense.

Proposition 22.3.2 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. Let p,q be conjugate real numbers. Assume that the Markov kernel P is $\mathrm{L}^p(\pi)$-exponentially convergent. Then there exists a finite positive constant M and a real number $\alpha \in (0,1)$ such that for all $\nu \in \mathbb{M}_q^0(\pi)$ and for all $n \in \mathbb{N}$,*

$$\|\nu P^n\|_{\mathbb{M}_q(\pi)} \leq M\alpha^n \|\nu\|_{\mathbb{M}_q(\pi)} .$$

Proof. Since $\nu \in \mathbb{M}_q^0(\pi)$, we have $\nu P^n = \nu(P^n - \Pi)$. Combining this with Theorem 22.1.8 and Lemma 22.2.1 yields for all $n \in \mathbb{N}$,

$$\begin{aligned}\|\nu P^n\|_{\mathbb{M}_q(\pi)} &\leq |||P^n - \Pi|||_{\mathbb{M}_q(\pi)} \|\nu\|_{\mathbb{M}_q(\pi)} \\ &= |||P^n - \Pi|||_{\mathrm{L}^p(\pi)} \|\nu\|_{\mathbb{M}_q(\pi)} \leq 2\,|||P^n|||_{\mathrm{L}_0^p(\pi)} \|\nu\|_{\mathbb{M}_q(\pi)} .\end{aligned}$$

The proof is completed by noting that P is $\mathrm{L}^p(\pi)$-exponentially convergent. □

Quite surprisingly, the existence of an $\mathrm{L}^2(\pi)$-absolute spectral gap implies $\mathrm{L}^p(\pi)$-exponential convergence for all $p \in (1,\infty)$.

Proposition 22.3.3 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. Assume that P has an $\mathrm{L}^2(\pi)$-absolute spectral gap. Then for all $p \in [1,\infty]$, the Markov kernel P is $\mathrm{L}^p(\pi)$-exponentially convergent, and for all $n \in \mathbb{N}$, we have*

$$|||(P-\Pi)^n|||_{\mathrm{L}^p(\pi)} \leq \begin{cases} 2^{(2-p)/p}\,|||\,P^n|||^{2(p-1)/p}_{\mathrm{L}^2_0(\pi)}, & p \in [1,2]\;, \\ 2^{1-2/p}\;|||\,P^n|||^{2/p}_{\mathrm{L}^2_0(\pi)}, & p \in [2,\;\infty]\;. \end{cases} \tag{22.3.1}$$

Proof. Let $p \in [1,2]$. We first use the Riesz–Thorin interpolation theorem, Theorem 22.A.3, for $p \in [1,2]$. By Proposition 1.6.3,

$$|||(P-\Pi)^n|||_{\mathrm{L}^1(\pi)} \leq |||P-\Pi|||_{\mathrm{L}^1(\pi)} \leq 2\,.$$

Moreover, by (22.2.6), $|||(P-\Pi)^n|||_{\mathrm{L}^2(\pi)} = |||P^n-\Pi|||_{\mathrm{L}^2(\pi)} = |||P^n|||_{\mathrm{L}^2_0(\pi)}$. Noting that

$$p^{-1} = (1-\theta).1^{-1} + \theta.2^{-1} \qquad \text{with} \qquad \theta = 2(p-1)/p\,,$$

we then obtain the first upper bound in (22.3.1) by applying Theorem 22.A.3.

Let $p \in [2,\infty)$. We use again the Riesz–Thorin theorem to interpolate between 2 and ∞. As before, we have $|||(P-\Pi)^n|||_{\mathrm{L}^2(\pi)} = |||P^n|||_{\mathrm{L}^2_0(\pi)}$. Applying again Proposition 1.6.3 yields

$$|||(P-\Pi)^n|||_{\mathrm{L}^\infty(\pi)} \leq |||P-\Pi|||_{\mathrm{L}^\infty(\pi)} \leq 2\,.$$

Using the convention $\infty^{-1} = 0$, we have

$$p^{-1} = (1-\theta).\infty^{-1} + \theta.2^{-1} \qquad \text{with} \qquad \theta = 2/p\,,$$

and the Riesz–Thorin interpolation theorem then concludes the proof. □

By an interpolation argument, we get a partial converse of Proposition 22.3.3 in the case that P is normal, i.e., $PP^* = P^*P$, where P^* is the adjoint of P.

Proposition 22.3.4 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. Assume that P is normal and that the Markov kernel P is $\mathrm{L}^p(\pi)$-exponentially convergent. Then*

$$|||P|||_{\mathrm{L}^2_0(\pi)} \leq \begin{cases} \left\{\lim_{n\to\infty} |||P^n|||^{1/n}_{\mathrm{L}^p_0(\pi)}\right\}^{p/2}, & p \in [1,2]\,, \\ \left\{\lim_{n\to\infty} |||P^n|||^{1/n}_{\mathrm{L}^p_0(\pi)}\right\}^{1/\{2(1-p^{-1})\}}, & p \in [2,\infty]\,. \end{cases}$$

Proof. By Proposition 22.2.8, since P is normal, $|||P|||_{\mathrm{L}^2_0(\pi)} = \lim_{n\to\infty} |||P^n|||^{1/n}_{\mathrm{L}^2_0(\pi)}$. Let $\alpha \in (\lim_{n\to\infty} |||P^n|||^{1/n}_{\mathrm{L}^p_0(\pi)}, 1)$. Assume first that $p \in [1,2]$. There exists $M < \infty$ such that for all $n \in \mathbb{N}$, $|||(P-\Pi)^n|||_{\mathrm{L}^p(\pi)} \leq M\alpha^n$. Using Proposition 1.6.3, we get $|||(P-\Pi)^n|||_{\mathrm{L}^\infty(\pi)} \leq 2$. We use the Riesz–Thorin interpolation theorem (Theorem 22.A.3) to show that for all $n \in \mathbb{N}$,

$$|||P^n|||_{\mathrm{L}^2_0(\pi)} = |||(P-\Pi)^n|||_{\mathrm{L}^2(\pi)} \leq 2^{1-p/2} M^{p/2} \alpha^{pn/2} .$$

Then applying Proposition 22.2.8 to the normal kernel P, we get $|||P|||_{\mathrm{L}^2_0(\pi)} = \lim_{n\to\infty} |||P^n|||^{1/n}_{\mathrm{L}^2_0(\pi)} \leq \alpha^{p/2}$, and the proof is complete for $p \in [1,2]$.

Assume now that $p \in [2,\infty]$. By Proposition 1.6.3, we have $|||(P-\Pi)^n|||_{\mathrm{L}^1(\pi)} \leq 1$, and the proof follows again by the Riesz–Thorin interpolation theorem and Proposition 22.2.8. □

It is of course interesting to relate $\mathrm{L}^p(\pi)$-exponential convergence for some $p \in [1,\infty]$ with the different definitions of ergodicity that we have introduced in Chapter 15. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. Recall from Definition 15.2.1 that the Markov kernel P is uniformly geometrically ergodic if there exist constants $C < \infty$ and $\rho \in [0,1)$ such that for all $n \in \mathbb{N}$ and $x \in \mathsf{X}$,

$$\|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} \leq C\rho^n . \tag{22.3.2}$$

We will say that the Markov kernel P is π-*a.e.* uniformly geometrically ergodic if the inequality holds for π-*a.e.* x. As shown below, uniform geometric ergodicity is equivalent to $\mathrm{L}^\infty(\pi)$-exponential convergence, which by Proposition 22.3.4 implies that $|||P|||_{\mathrm{L}^2_0(\pi)} < 1$ if the Markov kernel P is normal.

Proposition 22.3.5 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. The following statements are equivalent:*

(i) The Markov kernel P is π-a.e. uniformly geometrically ergodic.
(ii) The Markov kernel P is $\mathrm{L}^\infty(\pi)$-exponentially convergent.

In addition, if one of these conditions is satisfied, then P is $\mathrm{L}^p(\pi)$-exponentially convergent for all $p \in (1,\infty]$.

Proof. To establish (i) $\iff$ (ii), it suffices to show that for all $n \in \mathbb{N}$ and π-*a.e.* $x \in \mathsf{X}$,

$$\sup_{\|h\|_{\mathrm{L}^\infty(\pi)} \leq 1} |P^n h(x) - \pi(h)| = \sup_{|h|_\infty \leq 1} |P^n h(x) - \pi(h)| . \tag{22.3.3}$$

We have only to show that the left-hand side is smaller than the right-hand side, the reverse inequality being obvious.

For $n \in \mathbb{N}$ and $N \in \mathscr{X}$, set $\mathsf{X}[n,N] = \{x \in \mathsf{X} : P^n(x,N) = 0\}$. Since π is an invariant probability, we get for all $n \in \mathbb{N}$ and $N \in \mathscr{X}$,

$$\pi(N) = 0 \Leftrightarrow \pi(\mathsf{X}[n,N]) = 1 .$$

Let h be a function in $\mathrm{L}^\infty(\pi)$ satisfying $\|h\|_{\mathrm{L}^\infty(\pi)} \le 1$. If $N \in \mathscr{X}$ and $\pi(N) = 0$, then for all $x \in \mathsf{X}[n,N]$, we have

$$|P^n h(x) - \pi(h)| = |P^n(\mathbb{1}_{N^c} h)(x) - \pi(\mathbb{1}_{N^c} h)| . \tag{22.3.4}$$

Set $\tilde{h}(x) = h(x)\mathbb{1}_{\{|h(x)|\le 1\}}$. Hence $|\tilde{h}|_\infty \le 1$. Since $\pi(\{|h| > 1\}) = 0$, applying (22.3.4) with $N = \{|h| > 1\}$ shows that for all $x \in \mathsf{X}[n,\{|h| > 1\}]$, we get

$$|P^n h(x) - \pi(h)| = |P^n \tilde{h}(x) - \pi(\tilde{h})| \le \sup_{|g|_\infty \le 1} |P^n g(x) - \pi(g)| .$$

Finally, (i) is equivalent to (ii).

We now turn to the proof of the last part of the proposition. More specifically, we will show that if P is π-*a.e.* uniformly geometrically ergodic, then P is exponentially convergent in $\mathrm{L}^p(\pi)$, where $p \in (1,\infty]$. Set $Q = P - \Pi$. For $k \in \mathbb{N}^*$ and $x \in \mathsf{X}$ such that $\left\|Q^k(x,\cdot)\right\|_{\mathrm{TV}} > 0$, we get for all $h \in \mathrm{L}^p(\pi)$,

$$|Q^k h(x)| \le \int \left|Q^k(x,\cdot)\right|(\mathrm{d}y)|h|(y) = \left\|Q^k(x,\cdot)\right\|_{\mathrm{TV}} \int \frac{\left|Q^k(x,\cdot)\right|}{\|Q^k(x,\cdot)\|_{\mathrm{TV}}}(\mathrm{d}y)|h|(y) . \tag{22.3.5}$$

Using Jensen's inequality, we obtain

$$|Q^k h(x)|^p \le \left\|Q^k(x,\cdot)\right\|_{\mathrm{TV}}^{p-1} \int |Q^k(x,\cdot)|(\mathrm{d}y)|h|^p(\mathrm{d}y) . \tag{22.3.6}$$

There exist $\varsigma < \infty$ and $\rho \in [0,1)$ such that $\left\|Q^k(x,\cdot)\right\|_{\mathrm{TV}} \le \varsigma\rho^k$ for π-*a.e.* $x \in \mathsf{X}$ and all $k \in \mathbb{N}$. Since $|Q^k(x,\cdot)| \le P^k(x,\cdot) + \pi$, we get

$$\begin{aligned} |Q^k h(x)|^p &\le \left\|Q^k(x,\cdot)\right\|_{\mathrm{TV}}^{p-1} \left\{P^k|h|^p(x) + \pi(|h|^p)\right\} \\ &\le \{\varsigma\rho^k\}^{p-1} \left\{P^k|h|^p(x) + \pi(|h|^p)\right\} . \end{aligned}$$

This implies that

$$\left\|Q^k h\right\|_{\mathrm{L}^p(\pi)} \le 2^{1/p} \varsigma^{(p-1)/p} \rho^{(p-1)k/p} \|h\|_{\mathrm{L}^p(\pi)} .$$

The proof is complete. □

Corollary 22.3.6 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that P is reversible with respect to π and is uniformly geometrically ergodic. Then P has an $\mathrm{L}^2(\pi)$-absolute spectral gap.*

Proof. The result follows from Propositions 22.3.4 and 22.3.5. □

The next example shows that a reversible Markov kernel P may have an absolute spectral gap without being uniformly geometrically ergodic. Therefore, the uniform geometric ergodicity for reversible Markov kernel is a stronger property than the existence of a spectral gap.

Example 22.3.7 Consider the Gaussian autoregressive process of order 1 given by the recurrence $X_{k+1} = \phi X_k + \sigma Z_{k+1}$, where $\{Z_k,\ k \in \mathbb{N}^*\}$ is a sequence of i.i.d. standard Gaussian random variables independent of X_0, $\phi \in (-1,1)$, and $\sigma > 0$. The associated Markov kernel chain is given, for all $A \in \mathscr{B}(\mathbb{R})$, by

$$P(x,A) = \int_A \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\phi x)^2}{2\sigma^2}\right) \mathrm{d}y \,. \tag{22.3.7}$$

This Markov kernel is reversible to $\mathrm{N}(0,\sigma_\infty^2)$, the Gaussian distribution with zero mean and variance $\sigma_\infty^2 = \sigma^2/(1-\phi^2)$. For all $x \in \mathbb{R}$, $n \in \mathbb{N}$, $A \in \mathscr{B}(\mathbb{R})$, and $x \in \mathbb{R}$, we have

$$P^n(x,A) = \int_A \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(y-\phi^n x)^2}{2\sigma_n^2}\right) \mathrm{d}y \,, \qquad \sigma_n^2 = \sigma^2 \frac{1-\phi^{2n}}{1-\phi^2} \,. \tag{22.3.8}$$

For all $\delta > 0$, we have $\liminf_{n\to\infty} P^n(\phi^{-n/2}, [-\delta,\delta]) = 0$, whereas $\mathrm{N}(0,\sigma_\infty^2)([\delta,\delta]) > 0$, showing that the Markov kernel P is not uniformly (geometrically) ergodic. We will nevertheless show that P has positive absolute spectral gap. For every function $f \in \mathrm{L}_0^2(\pi)$, we get

$$\|Pf\|_{\mathrm{L}^2(\pi)}^2 = \langle Pf, Pf \rangle_{\mathrm{L}^2(\pi)} = \left\langle f, P^2 f \right\rangle_{\mathrm{L}^2(\pi)} = \mathrm{Cov}_\pi(f(X_0), f(X_2)) \,.$$

To bound the right-hand side of the previous inequality, we use Gebelein's inequality, which states that if (U,V) is a centered Gaussian vector in $\mathbb{R}^2$ with $\mathbb{E}\left[U^2\right] = 1$ and $\mathbb{E}\left[V^2\right] = 1$ and if f,g are two complex-valued functions such that $\mathbb{E}\left[f(U)\right] = 0$ and $\mathbb{E}\left[g(V)\right] = 0$, then $|\mathbb{E}\left[f(U)\overline{g(V)}\right]| \leq \rho \{\mathbb{E}\left[|f(U)|^2\right]\}^{1/2} \{\mathbb{E}\left[|g(V)|^2\right]\}^{1/2}$, where $\rho = |\mathbb{E}\left[UV\right]|$ is the correlation coefficient. Applying this inequality with $U = X_0/\sigma_\infty$ and $V = X_2/\sigma_\infty$, we obtain

$$\mathrm{Cov}_\pi(f(X_0), f(X_2)) \leq \phi^2 \|f\|_{\mathrm{L}^2(\pi)} \,.$$

Hence the Markov kernel P has an absolute $\mathrm{L}^2(\pi)$-spectral gap that is larger than $1-\phi^2$.

Recall from Theorem 15.1.5 that if P is irreducible, aperiodic, and positive with invariant probability measure π, then P is geometrically ergodic if and only if there exist a measurable function $V : \mathsf{X} \to [1,\infty]$ and a constant $\rho \in [0,1)$ such that $\pi(\{V < \infty\}) = 1$ and for all $n \in \mathbb{N}$ and $x \in \mathsf{X}$,

$$\|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} \leq V(x)\rho^n \,. \tag{22.3.9}$$

Lemma 22.3.8 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. Assume in addition that P is geometrically ergodic. Then for all $p \in [1,\infty)$, there exists $\varsigma < \infty$ such that for all $f \in \mathbb{F}_b(\mathsf{X})$ and $n \in \mathbb{N}$,*

$$\|P^n f - \pi(f)\|_{\mathrm{L}^p(\pi)} \leq \varsigma |f|_\infty \rho^n .$$

Proof. Note that P is necessarily aperiodic by Lemma 9.3.9. By Theorem 15.1.6, there exists a function $V : \mathsf{X} \to [1,\infty]$ satisfying $\|V\|_{\mathrm{L}^p(\pi)} < \infty$, $\rho \in [0,1)$ and $\varsigma_0 < \infty$, such that for all $n \in \mathbb{N}$, $\|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} \leq \varsigma_0 V(x) \rho^n$ for π-*a.e.* $x \in \mathsf{X}$. For all $f \in \mathbb{F}_b(\mathsf{X})$, we therefore have $|P^n f(x) - \pi(f)| \leq \|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} |f|_\infty$ for π-*a.e.* $x \in \mathsf{X}$, which implies that

$$\|P^n f - \pi(f)\|_{\mathrm{L}^p(\pi)} \leq \varsigma_0 \|V\|_{\mathrm{L}^p(\pi)} |f|_\infty \rho^n .$$

□

Lemma 22.3.9 *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. If P is $\mathrm{L}^2(\pi)$-geometrically ergodic, then P is aperiodic.*

Proof. The proof is by contradiction. Assume that the period d is larger than 2. Let $C_0, \ldots, C_{d-1}$ be a cyclic decomposition as stated in Theorem 9.3.6 and note that $\pi(C_0) > 0$ since C_0 is accessible and π is a maximal irreducibility measure (see Theorem 9.2.15). Set for $A \in \mathscr{X}$, $\mu(A) = \pi(A \cap C_0)/\pi(C_0)$, and note that $\mu \in \mathbb{M}_2(\pi)$. Note that for all $k \in \mathbb{N}$, $\mu P^{kd+1}(C_1) = 1$, so that $\mu P^{kd+1}(C_0) = 0$. Using (22.1.5) and the fact that P is $\mathrm{L}^2(\pi)$-geometrically ergodic, we obtain

$$\limsup_{n\to\infty} \|\mu P^n - \pi\|_{\mathrm{TV}} \leq \limsup_{n\to\infty} \|\mu P^n - \pi\|_{\mathbb{M}_2(\pi)} = 0 .$$

This implies $\lim_{n\to\infty} \mu P^n(C_0) = \pi(C_0) > 0$, which contradicts $\mu P^{kd+1}(C_0) = 0$ for all $k \in \mathbb{N}$. □

Theorem 22.3.10. *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. If the Markov kernel P is $\mathrm{L}^2(\pi)$-geometrically ergodic, then P is geometrically ergodic.*

Proof. Let $\mu \in \mathbb{M}_2(\pi)$. By (22.1.5), $\|\mu\|_{\mathrm{TV}} \leq \|\mu\|_{\mathbb{M}_2(\pi)}$. Since the Markov kernel P is $\mathrm{L}^2(\pi)$-geometrically ergodic, there exist $\rho \in [0,1)$ and for all probability measures $\mu \in \mathbb{M}_2(\pi)$ a constant $C(\mu) < \infty$ such that for all $n \in \mathbb{N}$,

$$\|\mu P^n - \pi\|_{\mathrm{TV}} \leq \|\mu P^n - \pi\|_{\mathbb{M}_2(\pi)} \leq C(\mu) \rho^n . \tag{22.3.10}$$

We need to extend this relation to π-*a.e.* starting points $x \in \mathsf{X}$. Since P is irreducible, Theorem 9.2.15 shows that π is a maximal irreducibility measure. By Proposition 9.4.4 (i) and Theorem 9.4.10, there exists an accessible $(m, \varepsilon\pi)$-small set S. Note that $\pi(S) > 0$, since S is accessible and π is a maximal irreducibility measure.

Define μ to be π restricted to S and normalized to be a probability measure, i.e., $\mu = \{\pi(S)\}^{-1}\mathbb{1}_S(x) \cdot \pi$. Then

$$\int \left(\frac{\mathrm{d}\mu}{\mathrm{d}\pi}\right)^2 \mathrm{d}\pi = \frac{1}{\pi(S)} < \infty ,$$

showing that μ is in $\mathbb{M}_2(\pi)$. Using (22.3.10), we get for all $n \in \mathbb{N}$,

$$\left| \int_S \mu(\mathrm{d}y)\{P^n(y,S) - \pi(S)\} \right| \le \left\| \int_S \mu(\mathrm{d}y) P^n(y,\cdot) - \pi \right\|_{\mathrm{TV}} \le C(\mu)\rho^n .$$

By Lemma 22.3.9, P is aperiodic. We conclude with the characterization Theorem 15.1.5 (iii). □

We now consider the converse application.

Theorem 22.3.11. *Let P be an irreducible Markov kernel on $\mathsf{X} \times \mathscr{X}$ reversible with respect to the probability measure π. Then the following statements are equivalent:*

(i) P has an absolute $\mathrm{L}^2(\pi)$-spectral gap.
(ii) P is geometrically ergodic.

Proof. (i) ⇒ (ii) From Proposition 22.2.4, the existence of an absolute $\mathrm{L}^2(\pi)$-spectral gap implies that the Markov kernel P is $\mathrm{L}^2(\pi)$-geometrically ergodic, and the conclusion follows from Theorem 22.3.10.

(ii) ⇒ (i) Since P is geometrically ergodic, Lemma 22.3.8 shows that there exists a constant $\rho \in [0,1)$ such that for all $f \in \mathbb{F}_b(\mathsf{X})$ satisfying $\pi(f) = 0$,

$$\|P^n f\|_{\mathrm{L}^2(\pi)} \le C(f)\rho^n , \qquad \text{for some constant } C(f) < \infty. \tag{22.3.11}$$

Since the Markov kernel P is self-adjoint in $\mathrm{L}^2(\pi)$, for all $n \in \mathbb{N}$,

$$\begin{aligned} \|P^n f\|_{\mathrm{L}^2(\pi)} &= \langle P^n f, P^n f\rangle_{\mathrm{L}^2(\pi)} = \left\langle f, P^{2n} f\right\rangle \\ &= \int_{-1}^{1} t^{2n} \nu_f(\mathrm{d}t) , \end{aligned} \tag{22.3.12}$$

where ν_f is the spectral measure associated with P (see Theorem 22.B.3). We now use the same argument as in the proof of Theorem 22.2.7 to show that the support of the spectral measure is included in $[-\rho,\rho]$. To be specific, taking $a \in (\rho,1)$, we get

$$\|P^n f\|_{\mathrm{L}^2(\pi)} \ge a^{2n} \nu_f([-1,-a] \cup [a,1]),$$

and (22.3.11) therefore implies that $\nu_f([-1,-a] \cup [a,1]) = 0$. Using again (22.3.12), we get that for every function $f \in \mathbb{F}_b(\mathsf{X})$ with $\pi(f) = 0$,

$$\|P^n f\|_{L^2(\pi)} \le \rho^n \int_{-1}^{1} \nu_f(dt) = \rho^{2n} \|f\|_{L^2(\pi)} \ .$$

Since $\{f \in \mathbb{F}_b(\mathsf{X}) : \pi(f) = 0\}$ is dense in $L^2_0(\pi)$, we have for all $f \in L^2_0(\pi)$ and $n \in \mathbb{N}$, $\|P^n f\|_{L^2(\pi)} \le \rho^n \|f\|_{L^2(\pi)}$. Therefore, $\mathrm{Abs.Gap}_{L^2(\pi)}(P) \ge 1 - \rho$.

□

22.4 Cheeger's Inequality

In most of this section, we consider a Markov kernel P that is reversible with respect to the probability measure π. We set

$$\lambda_{\max}(P) = \sup\{\lambda : \lambda \in \mathrm{Spec}(P|L^2_0(\pi))\} \ , \tag{22.4.1}$$

and we define

$$\mathrm{Gap}_{L^2(\pi)}(P) = 1 - \lambda_{\max}(P)$$

as the spectral gap of P. The objective of this section is to establish bounds on $\mathrm{Gap}_{L^2(\pi)}(P)$. We begin with the definition of the *Cheeger constant* (also called the *conductance*), which is valid for every Markov kernel P with invariant probability measure π.

Definition 22.4.1 (Conductance) *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. The Cheeger constant is*

$$\mathrm{k}_P = \inf\{\mathrm{k}_P(A) : A \in \mathscr{X}, 0 < \pi(A) < 1\} \ , \tag{22.4.2}$$

with

$$\mathrm{k}_P(A) = \frac{\int \pi(\mathrm{d}x)\mathbb{1}_A(x)P(x,A^c)}{\pi(A)\pi(A^c)}, \qquad A \in \mathscr{X} \ . \tag{22.4.3}$$

In words, the Cheeger constant $\mathrm{k}_P(A)$ associated with a Markov kernel P and a set A is the probability flow from A to its complement A^c, normalized by the invariant probabilities of A and A^c. If for some set $A \in \mathscr{X}$, the flow from A to A^c is very small compared to the invariant distribution of A and A^c, then it is sensible to expect that the mixing time of the Markov kernel will be large.

Lemma 22.4.2 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ reversible with respect to π. Then*

$$\mathrm{k}_P = \inf_{A \in \mathscr{X}, 0 < \pi(A) \le 1/2} \mathrm{k}_P(A) \le 2 \ .$$

Proof. Since P is self-adjoint, we have $\langle \mathbb{1}_A, P\mathbb{1}_{A^c}\rangle_{L^2(\pi)} = \langle \mathbb{1}_{A^c}, P\mathbb{1}_A\rangle_{L^2(\pi)}$, which implies for all $A \in \mathscr{X}$,

$$\mathrm{k}_P(A) = \frac{\langle \mathbb{1}_A, P\mathbb{1}_{A^c} \rangle_{\mathrm{L}^2(\pi)}}{\pi(A)\pi(A^c)} = \frac{\langle \mathbb{1}_{A^c}, P\mathbb{1}_A \rangle_{\mathrm{L}^2(\pi)}}{\pi(A^c)\pi(A)} = \mathrm{k}_P(A^c).$$

The proof of the equality is complete, since for all $A \in \mathscr{X}$, either $\pi(A) \leq 1/2$ or $\pi(A^c) \leq 1/2$. We now turn to the upper bound. First, note that

$$\pi(A^c)\,\mathrm{k}_P(A) = \frac{\int \pi(\mathrm{d}x)\mathbb{1}_A(x)P(x,A^c)}{\pi(A)} \leq 1 \ .$$

Replacing A by A^c, we also have $\pi(A)\,\mathrm{k}_P(A^c) \leq 1$. Combining this with $\mathrm{k}_P(A^c) = \mathrm{k}_P(A)$, we deduce $\mathrm{k}_P(A) = \pi(A^c)\,\mathrm{k}_P(A) + \pi(A)\,\mathrm{k}_P(A^c) \leq 2$, and the proof is finished. □

Theorem 22.4.3. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, reversible with respect to π. Then*

$$\frac{\mathrm{k}_P^2}{8} \leq \mathrm{Gap}_{\mathrm{L}^2(\pi)}(P) \leq \mathrm{k}_P \ . \tag{22.4.4}$$

Proof. Using the notation $Q = \mathrm{I} - P$, we can express $\mathrm{k}_P(A)$ defined in (22.4.3) as follows:

$$\mathrm{k}_P(A) = -\frac{\langle \mathbb{1}_A, Q\mathbb{1}_{A^c} \rangle_{\mathrm{L}^2(\pi)}}{\pi(A)\pi(A^c)} = \frac{\langle \mathbb{1}_A, Q\mathbb{1}_A \rangle_{\mathrm{L}^2(\pi)}}{\pi(A)\pi(A^c)} \ . \tag{22.4.5}$$

Moreover, for all $f \in \mathrm{L}^2(\pi)$,

$$1 - \frac{\langle f, Pf \rangle_{\mathrm{L}^2(\pi)}}{\|f\|^2_{\mathrm{L}^2(\pi)}} = \frac{\langle f, f \rangle_{\mathrm{L}^2(\pi)} - \langle f, Pf \rangle_{\mathrm{L}^2(\pi)}}{\|f\|^2_{\mathrm{L}^2(\pi)}} = \frac{\langle f, Qf \rangle_{\mathrm{L}^2(\pi)}}{\|f\|^2_{\mathrm{L}^2(\pi)}} \ .$$

Since P is self-adjoint, Theorem 22.A.19 shows that

$$\begin{aligned} \mathrm{Gap}_{\mathrm{L}^2(\pi)}(P) = 1 - \lambda_{\max}(P) &= 1 - \sup_{f \in \mathrm{L}^2_0(\pi), f \neq 0} \frac{\langle f, Pf \rangle_{\mathrm{L}^2(\pi)}}{\|f\|^2_{\mathrm{L}^2(\pi)}} \\ &= \inf_{f \in \mathrm{L}^2_0(\pi), f \neq 0} \frac{\langle f, Qf \rangle_{\mathrm{L}^2(\pi)}}{\|f\|^2_{\mathrm{L}^2(\pi)}} \ . \end{aligned} \tag{22.4.6}$$

Combining this identity with the fact that $\langle f, Qg \rangle_{\mathrm{L}^2(\pi)} = 0$ if f is constant and $g \in \mathrm{L}^2_0(\pi)$ or if $f \in \mathrm{L}^2(\pi)$ and g is constant, we get

$$\begin{aligned} \mathrm{Gap}_{\mathrm{L}^2(\pi)}(P) &\leq \inf_{A \in \mathscr{X}, 0 < \pi(A) < 1} \frac{\langle \mathbb{1}_A - \pi(A), Q(\mathbb{1}_A - \pi(A)) \rangle_{\mathrm{L}^2(\pi)}}{\left\|\mathbb{1}_A - \pi(A)\right\|^2_{\mathrm{L}^2(\pi)}} \\ &= \inf_{A \in \mathscr{X}, 0 < \pi(A) < 1} \frac{\langle \mathbb{1}_A, Q\mathbb{1}_A \rangle_{\mathrm{L}^2(\pi)}}{\pi(A)\pi(A^c)} = \mathrm{k}_P \ , \end{aligned}$$

where the last equality follows from (22.4.5). This establishes the upper bound in (22.4.4). We next turn to the lower bound. The Markov kernel Q being real and self-adjoint, it suffices to consider real functions. Since π is invariant for P, we obtain for every real-valued function $f \in \mathrm{L}^2(\pi)$,

$$\begin{aligned}\langle f, Qf\rangle_{\mathrm{L}^2(\pi)} &= \frac{1}{2}\int \pi(\mathrm{d}x)P(x,\mathrm{d}y)\{f^2(x)+f^2(y)-2f(x)f(y)\} \\ &= \frac{1}{2}\int \pi(\mathrm{d}x)P(x,\mathrm{d}y)\{f(x)-f(y)\}^2 .\end{aligned}$$

Set now $g = f + c$. By the Cauchy–Schwarz inequality,

$$\begin{aligned}\langle f, Qf\rangle_{\mathrm{L}^2(\pi)} &= \frac{1}{2}\int \pi(\mathrm{d}x)P(x,\mathrm{d}y)\{g(x)-g(y)\}^2 \\ &\geq \frac{1}{2}\frac{\{\int \pi(\mathrm{d}x)P(x,\mathrm{d}y)|g^2(x)-g^2(y)|\}^2}{\int \pi(\mathrm{d}x)P(x,\mathrm{d}y)\{g(x)+g(y)\}^2} \\ &\geq \frac{1}{2}\frac{\{\int \pi(\mathrm{d}x)P(x,\mathrm{d}y)|g^2(x)-g^2(y)|\}^2}{\int \pi(\mathrm{d}x)P(x,\mathrm{d}y)\{2g^2(x)+2g^2(y)\}} \\ &= \frac{1}{8}\frac{\{\int \pi(\mathrm{d}x)P(x,\mathrm{d}y)|g^2(x)-g^2(y)|\}^2}{\int \pi(\mathrm{d}x)g^2(x)}\end{aligned} \tag{22.4.7}$$

where we have used in the last equality that $\pi P = \pi$. Using again the invariance of π, we have $\int \pi(\mathrm{d}x)P(x,\mathrm{d}y)\{g^2(x)-g^2(y)\} = 0$, which implies

$$\begin{aligned}N &:= \int \pi(\mathrm{d}x)P(x,\mathrm{d}y)|g^2(x)-g^2(y)| \\ &= 2\int \pi(\mathrm{d}x)P(x,\mathrm{d}y)\mathbb{1}\left\{g^2(x)>g^2(y)\right\}\left\{g^2(x)-g^2(y)\right\} \\ &= 2\int \pi(\mathrm{d}x)P(x,\mathrm{d}y)\mathbb{1}\left\{g^2(x)>g^2(y)\right\}\int_0^\infty \mathbb{1}\left\{g^2(x)>u\geq g^2(y)\right\}\mathrm{d}u .\end{aligned}$$

Using first Fubini's theorem and then writing $\mathbb{1}\left\{g^2(x)>u\geq g^2(y)\right\} = \mathbb{1}_{A_u}(x)\mathbb{1}_{A_u^c}(y)$, where $A_u = \left\{x \in \mathsf{X} :\ g^2(x) > u\right\}$, we may express N as

$$N = 2\int_0^\infty \mathrm{d}u \int \pi(\mathrm{d}x)P(x,\mathrm{d}y)\mathbb{1}\left\{g^2(x)>u\geq g^2(y)\right\} = 2\int_0^\infty \mathrm{d}u \left\langle \mathbb{1}_{A_u}, P\mathbb{1}_{A_u^c}\right\rangle_{\mathrm{L}^2(\pi)} .$$

Therefore,

$$\begin{aligned}N &\geq 2\,\mathrm{k}_P \int_0^\infty \mathrm{d}u\, \pi(A_u)\pi(A_u^c) = 2\,\mathrm{k}_P\int_0^\infty \mathrm{d}u \int \pi(\mathrm{d}x)\pi(\mathrm{d}y)\mathbb{1}\left\{g^2(x)>u\geq g^2(y)\right\} \\ &= 2\,\mathrm{k}_P \int \pi(\mathrm{d}x)\pi(\mathrm{d}y)\mathbb{1}\left\{g^2(x)>g^2(y)\right\}\left\{g^2(x)-g^2(y)\right\} \\ &= \mathrm{k}_P \int \pi(\mathrm{d}x)\pi(\mathrm{d}y)|g^2(x)-g^2(y)| .\end{aligned}$$

Plugging this inequality into (22.4.7) yields for all $c \in \mathbb{R}$,

$$\langle f, Qf \rangle_{L^2(\pi)} \geq \frac{\mathrm{k}_P^2}{8} \frac{\{\int \pi(\mathrm{d}x)\pi(\mathrm{d}y)|\{f(x)+c\}^2 - \{f(y)+c\}^2|\}^2}{\int \pi(\mathrm{d}x)\{f(x)+c\}^2(x)} . \tag{22.4.8}$$

Let $f \in L_0^2(\pi)$ be such that $\|f\|_{L^2(\pi)} = 1$. Let U_0, U_1 be two i.i.d. random variables on $(\mathsf{X}, \mathscr{X})$ such that π is the distribution of U_i, $i \in \{0,1\}$. Setting $X = f(U_0)$ and $Y = f(U_1)$, we obtain that X and Y are real-valued i.i.d. random variables with zero mean and unit variance. Moreover, (22.4.8) shows that

$$\langle f, Qf \rangle_{L^2(\pi)} \geq \frac{\mathrm{k}_P^2}{8} \sup_{c \in \mathbb{R}} \frac{\{\mathbb{E}\left[|(X+c)^2 - (Y+c)^2|\right]\}^2}{\mathbb{E}\left[(Y+c)^2\right]} .$$

Combining this with Lemma 22.4.4 below and (22.4.6) yields $\mathrm{Gap}_{L^2(\pi)}(P) \geq \mathrm{k}_P^2/8$, which completes the proof.

□

Lemma 22.4.4. *Let X,Y be two i.i.d. centered real-valued random variables with variance 1. Then*

$$K := \sup_{c \in \mathbb{R}} \frac{\{\mathbb{E}\left[|(X+c)^2 - (Y+c)^2|\right]\}^2}{\mathbb{E}\left[(Y+c)^2\right]} \geq 1 .$$

Proof. First note that by definition of K,

$$\begin{aligned} K &\geq \limsup_{c \to \infty} \frac{\{\mathbb{E}\left[|(X+c)^2 - (Y+c)^2|\right]\}^2}{\mathbb{E}\left[(Y+c)^2\right]} \\ &= \limsup_{c \to \infty} \frac{\{\mathbb{E}\left[|(X-Y)(X+Y+2c)|\right]\}^2}{\mathbb{E}\left[(Y+c)^2\right]} = 4\{\mathbb{E}\left[|X-Y|\right]\}^2 . \end{aligned}$$

Moreover, using that X,Y are independent and $\mathbb{E}[Y] = 0$, we get $\mathbb{E}\left[\,|X-Y|\,|\,\sigma(X)\right] \geq |\mathbb{E}\left[X-Y\,|\,\sigma(X)\right]| = |X|$, which in turn implies

$$K \geq 4\{\mathbb{E}\left[|X|\right]\}^2 . \tag{22.4.9}$$

By choosing $c = 0$ in the definition of K, we get $K^{1/2} \geq \mathbb{E}\left[|X^2 - Y^2|\right]$. Using that X,Y are independent and $\mathbb{E}\left[Y^2\right] = 1$, we get

$$\mathbb{E}\left[|X^2 - Y^2|\,\middle|\,\sigma(X)\right] \geq |\mathbb{E}\left[X^2 - Y^2\,\middle|\,\sigma(X)\right]| = |X^2 - 1| .$$

Then, noting that $\mathbb{E}\left[X^2\right] = 1$ and $u^2 \wedge 1 \leq u$ for all $u \geq 0$, we see that

$$K^{1/2} \geq \mathbb{E}\left[|X^2 - 1|\right] = \mathbb{E}\left[X^2 + 1 - 2(X^2 \wedge 1)\right] \geq 2 - 2\mathbb{E}\left[|X|\right] . \tag{22.4.10}$$

Then using either (22.4.9) if $\mathbb{E}[|X|] \geq 1/2$ or (22.4.10) if $\mathbb{E}[|X|] < 1/2$, we finally obtain $K \geq 1$ in all cases. □

Example 22.4.5. Let $\mathsf{G} \subset \mathbb{R}^d$ be a bounded Borel set with $\mathrm{Leb}(\mathsf{G}) > 0$ and let $\rho : \mathsf{G} \to [0, \infty)$ be an integrable function with respect to Lebesgue measure. Assume that we are willing to sample the distribution π_ρ on $(\mathsf{G}, \mathscr{B}(\mathsf{G}))$ with density defined by

$$h_\rho(x) = \frac{\rho(x)\mathbb{1}_{\mathsf{G}}(x)}{\int_G \rho(x)\mathrm{d}x}\,, \quad \pi_\rho = h_\rho \cdot \mathrm{Leb}_d\,. \tag{22.4.11}$$

Assume that there exist $0 < c_1 < c_2 < \infty$ such that $c_1 \leq \rho(x) \leq c_2$ for all $x \in \mathsf{G}$. We consider an independent Metropolis–Hastings sampler (see Example 2.3.3) with uniform proposal distribution over G, i.e., with density

$$\bar{q}(x) = \frac{\mathbb{1}_{\mathsf{G}}(x)}{\mathrm{Leb}(\mathsf{G})},$$

i.e., a state is proposed with the uniform distribution on G. The independent sampler kernel is given, for $x \in \mathsf{G}$ and $A \in \mathscr{B}(\mathsf{G})$, by

$$P(x,A) = \int_A \alpha(x,y)\frac{\mathrm{d}y}{\mathrm{Leb}(\mathsf{G})} + \mathbb{1}_A(x)\left(1 - \int_{\mathsf{G}} \alpha(x,\, y)\frac{\mathrm{d}y}{\mathrm{Leb}(\mathsf{G})}\right),$$

where for $(x,y) \in \mathsf{G} \times \mathsf{G}$,

$$\alpha(x,y) = \min\left(1, \frac{h_\rho(y)}{h_\rho(x)}\right) = \min\left(1, \frac{\rho(y)}{\rho(x)}\right)\,. \tag{22.4.12}$$

Recall that the Markov kernel P is reversible with respect to the target distribution π_ρ. For all $x \in \mathsf{G}$ and $A \in \mathscr{B}(\mathsf{G})$, we get

$$\begin{aligned} P(x,A) &\geq \frac{1}{\mathrm{Leb}(\mathsf{G})}\int_A \left\{\frac{1}{\rho(y)} \wedge \frac{1}{\rho(x)}\right\}\rho(y)\mathrm{d}y \geq \frac{1}{c_2\mathrm{Leb}(\mathsf{G})}\int_A \rho(y)\mathrm{d}y \\ &\geq \frac{c_1}{c_2}\int_A \bar{q}(y)\mathrm{d}y\,. \end{aligned}$$

Hence by applying Theorem 15.3.1, we see that the Markov kernel P is uniformly ergodic and

$$\left\|P^n(x,\cdot) - \pi_\rho\right\|_{\mathrm{TV}} \leq (1 - c_1/c_2)^n\,.$$

Let us apply Theorem 22.4.3. We estimate the Cheeger constant: for $A \in \mathscr{B}(\mathsf{G})$, we get

$$
\begin{aligned}
\int_A P(x, A^c)\pi_\rho(\mathrm{d}x) &= \int_A \left(\int_{A^c} \alpha(x, y) \frac{\mathrm{d}y}{\mathrm{Leb}(\mathsf{G})} \right) \pi_\rho(\mathrm{d}x) \\
&= \frac{1}{\mathrm{Leb}(\mathsf{G})} \int_A \left(\int_{A^c} \min\left\{ \int_{\mathsf{G}} \frac{\rho(z)}{\rho(x)} \mathrm{d}z, \int_{\mathsf{G}} \frac{\rho(z)}{\rho(y)} \mathrm{d}z \right\} \pi_\rho(\mathrm{d}y) \right) \pi_\rho(\mathrm{d}x) \\
&\geq \frac{c_1}{c_2} \pi_\rho(A) \pi_\rho(A^c) ,
\end{aligned}
$$

and therefore $\mathrm{k}_P \geq c_1/c_2$. ◀

Cheeger's theorem makes it possible to calculate a bound of the spectral gap, $\mathrm{Gap}_{\mathrm{L}^2(\pi)}(P)$ in terms of the conductance. Considering (22.4.4), we can see that a necessary and sufficient condition for the existence of a spectral gap is that Cheeger's constant be positive. Unfortunately, bounds on the conductance allow bounds only on the maximum of $\mathrm{Spec}(P|\mathrm{L}^2_0(\pi))$. The convergence results we developed in Section 22.2 require us to obtain bounds of the absolute spectral gap. It is therefore also necessary to consider the minimum of $\mathrm{Spec}(P|\mathrm{L}^2_0(\pi))$. When the Markov kernel P is reversible, there is always a simple way to get rid of $\mathrm{Spec}(P|\mathrm{L}^2_0(\pi)) \cap [-1, 0)$ by considering the *lazy version* of the Markov chain. At each step of the algorithm, the Markov chain either remains at the current position, with probability $1/2$, or moves according to P. The Markov kernel of the lazy chain therefore is $Q = 1/2(\mathrm{I} + P)$. The spectrum of this operator is nonnegative, which implies that the negative values of the spectrum of P do not matter much in practice. When the Markov chain is used for Monte Carlo simulations, this strategy has almost no influence on the computational cost: at each iteration, it is simply necessary to sample an additional binomial random variable. In some cases, it is, however, possible to avoid such a modification.

Definition 22.4.6. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ reversible with respect to π. We say that P defines a positive operator on $\mathrm{L}^2(\pi)$ if for all $f \in \mathrm{L}^2(\pi)$, $\langle f, Pf \rangle_{\mathrm{L}^2(\pi)} \geq 0$.*

It follows from Theorem 22.A.19 that the spectrum of the positive Markov kernel is a subset of $[0, 1]$. Therefore, if P is reversible and defines a positive operator on $\mathrm{L}^2(\pi)$, then $\mathrm{Gap}_{\mathrm{L}^2(\pi)}(P) = \mathrm{Abs.Gap}_{\mathrm{L}^2(\pi)}(P)$. In other cases, the absolute spectral gap for P can possibly be different from the spectral gap for P, depending on the relative value of the infimum of the spectrum associated with $\mathrm{L}^2_0(\pi)$ with respect to the supremum.

Example 22.4.7 (Positivity of the DUGS kernel). We consider the DUGS algorithm described in Section 2.3.3. Let $(\mathsf{X}, \mathscr{X})$ and $(\mathsf{Y}, \mathscr{Y})$ be complete separable metric spaces endowed with their Borel σ-fields $\mathscr{X}$ and $\mathscr{Y}$. Recall that we assume that there exist probability measures π and $\tilde{\pi}$ on $(\mathsf{X}, \mathscr{X})$ and Markov kernels R on $\mathsf{X} \times \mathscr{Y}$ and S on $\mathsf{Y} \times \mathscr{X}$ such that

$$\pi^*(\mathrm{d}x\mathrm{d}y) = \pi(\mathrm{d}x)R(x,\mathrm{d}y) = \tilde{\pi}(\mathrm{d}y)S(y,\mathrm{d}x)\,, \tag{22.4.13}$$

where $\tilde{\pi}$ is a probability measure on Y. Recall that the DUGS sampler is a two-step procedure, which can be described as follows. Given (X_k, Y_k),

(DUGS1) sample Y_{k+1} from $R(X_k, \cdot)$,
(DUGS2) sample X_{k+1} from $S(Y_{k+1}, \cdot)$.

The sequence $\{X_n,\ n \in \mathbb{N}\}$ therefore defines a Markov chain with Markov kernel $P = RS$. Note that for all $(f,g) \in \mathrm{L}^2(\pi) \times \mathrm{L}^2(\tilde{\pi})$,

$$\langle f, Rg\rangle_{\mathrm{L}^2(\pi)} = \int \pi(\mathrm{d}x) f(x) R(x,\mathrm{d}y) g(y) = \int \tilde{\pi}(\mathrm{d}y) f(x) S(y,\mathrm{d}x) g(y) = \langle g, Sf\rangle_{\mathrm{L}^2(\tilde{\pi})}\,,$$

showing that $S = R^*$. Therefore, Lemma 22.A.20 implies that $P = RR^*$ is a positive operator on $\mathrm{L}^2(\pi)$. ◀

Example 22.4.8 (Positivity of the hit and run Markov kernel). Let K be a bounded subset of $\mathbb{R}^d$ with nonempty interior. Let $\rho : K \to [0,\infty)$ be a (not necessarily normalized) density, i.e., a nonnegative Lebesgue-integrable function. We define the measure with density ρ by

$$\pi_\rho(A) = \frac{\int_A \rho(x)\mathrm{d}x}{\int_K \rho(x)\mathrm{d}x}\,, A \in \mathscr{B}(\mathbb{R}^d)\,. \tag{22.4.14}$$

The hit-and-run method, introduced in Section 2.3.4, is an algorithm to sample π_ρ. It consists of two steps: starting from $x \in K$, choose a random direction $\theta \in \mathrm{S}_{d-1}$ (the unit sphere in $\mathbb{R}^d$) and then choose the next state of the Markov chain with respect to the density ρ restricted to the chord determined by $x \in K$ and $\theta \in \mathrm{S}_{d-1}$. The Markov operator H that corresponds to the hit-and-run chain is defined by

$$Hf(x) = \int_{\mathrm{S}_{d-1}} \frac{1}{\ell_\rho(x,\theta)} \int_{-\infty}^{\infty} f(x+s\theta)\rho(x+s\theta)\mathrm{d}s\,\sigma_{d-1}(\mathrm{d}\theta)\,,$$

where σ_{d-1} is the uniform distribution of the $(d-1)$-dimensional unit sphere and

$$\ell_\rho(x,\ \theta) = \int_{-\infty}^{\infty} \mathbb{1}_K(x+s\theta)\rho(x+s\theta)\mathrm{d}s\,. \tag{22.4.15}$$

We have shown in Lemma 2.3.10 that the Markov kernel H_ρ is reversible with respect to π_ρ. Let μ be the product measure of π_ρ and the uniform distribution on S_{d-1}, and let $\mathrm{L}^2(\mu)$ be the Hilbert space of functions $g : K \times \mathrm{S}_{d-1} \to \mathbb{R}$ equipped with the inner product

$$\langle g,h\rangle_{\mathrm{L}^2(\mu)} = \int_K \int_{\mathrm{S}_{d-1}} g(x,\theta)h(x,\theta)\sigma_{d-1}(\mathrm{d}\theta)\pi_\rho(\mathrm{d}x), \quad \text{for } g,h \in \mathrm{L}^2(\mu)\ .$$

Define the operators $M : \mathrm{L}^2(\mu) \mapsto \mathrm{L}^2(\pi)$ and $T : \mathrm{L}^2(\mu) \to \mathrm{L}^2(\mu)$ as follows:

$$Mg(x) = \int_{\mathrm{S}_{d-1}} g(x,\theta)\sigma_{d-1}(\mathrm{d}\theta)\ , \tag{22.4.16}$$

and

$$Tg(x,\theta) = \frac{1}{\ell_\rho(x,\theta)} \int_{-\infty}^{\infty} g(x+s\theta,\ \theta)\rho(x+s\theta)\mathrm{d}s\ . \tag{22.4.17}$$

Recall that the adjoint operator of M is the unique operator M^* that satisfies $\langle f, Mg\rangle_{\mathrm{L}^2(\pi)} = \langle M^* f, g\rangle_{\mathrm{L}^2(\mu)}$ for all $f \in \mathrm{L}^2(\pi)$, $g \in \mathrm{L}^2(\mu)$. Since

$$\langle f, Mg\rangle_{\mathrm{L}^2(\pi)} = \int_K \int_{\mathrm{S}_{d-1}} f(x)g(x,\theta)\sigma_{d-1}(\mathrm{d}\theta)\pi_\rho(\mathrm{d}x)\ ,$$

we obtain that for all $\theta \in \mathrm{S}_{d-1}$ and $x \in K$, $M^* f(x,\ \theta) = f(x)$. This implies

$$MTM^* f(x) = \int_{\mathrm{S}_{d-1}} \frac{1}{\ell_\rho(x,\theta)} \int_{-\infty}^{\infty} f(x+s\theta)\rho(x+s\theta)\mathrm{d}s\sigma_{d-1}(\mathrm{d}\theta) = Hf(x)\ . \tag{22.4.18}$$

First of all, note that by Fubini's theorem, the operator T is self-adjoint in $\mathrm{L}^2(\mu)$. It remains to show that the operator T is positive. For all $s \in \mathbb{R}$, $x \in K$, and $\theta \in \mathrm{S}_{d-1}$, we have

$$\begin{aligned} Tg(x+s\theta,\theta) &= \frac{\int_{-\infty}^{\infty} g(x+(s+s')\theta)\rho(x+(s+s')\theta)\mathbb{1}_K(x+(s+s')\theta)\mathrm{d}s'}{\int_{-\infty}^{\infty} \rho(x+(s+s')\theta,\theta)\mathbb{1}_K(x+(s+s')\theta)\mathrm{d}s'} \\ &= \frac{\int_{-\infty}^{\infty} g(x+s'\theta)\rho(x+s'\theta)\mathbb{1}_K(x+(s+s')\theta)\mathrm{d}s'}{\int_{-\infty}^{\infty} \rho(x+s'\theta,\theta)\mathbb{1}_K(x+s'\theta)\mathrm{d}s'} = Tg(x,\theta). \end{aligned}$$

It follows that

$$\begin{aligned} T^2 g(x,\ \theta) &= \frac{1}{\ell_\rho(x,\theta)} \int_{-\infty}^{\infty} Tg(x+s\theta,\ \theta)\rho(x+s\theta)\mathrm{d}s \\ &= Tg(x,\ \theta). \end{aligned}$$

Thus T is a self-adjoint and idempotent operator on $\mathrm{L}^2(\mu)$, which implies that T is a projection, and in particular, that it is positive. By Lemma 22.A.20, the relation $H = MTM^*$ established in (22.4.18) shows that the Markov operator H is positive.
◀

22.5 Variance Bounds for Additive Functionals and the Central Limit Theorem for Reversible Markov Chains

Proposition 22.5.1 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. Assume that the probability measure π is reversible with respect to P. Set*

$$\lambda_{\min} = \inf\left\{\lambda \,:\, \lambda \in \mathrm{Spec}(P|\mathrm{L}^2_0(\pi))\right\} ,$$
$$\lambda_{\max} = \sup\left\{\lambda \,:\, \lambda \in \mathrm{Spec}(P|\mathrm{L}^2_0(\pi))\right\} .$$

Then for all $h \in \mathrm{L}^2_0(\pi)$,

$$\mathrm{Var}_\pi(S_n(h)) = \int_{\lambda_{\min}}^{\lambda_{\max}} w_n(t)\nu_h(\mathrm{d}t) , \tag{22.5.1}$$

where $S_n(h) = \sum_{j=0}^{n-1} h(X_j)$ and ν_h denotes the spectral measure associated with h (see Theorem 22.B.3) and $w_n : [-1,1] \to \mathbb{R}$ defined by $w_n(1) = n^2$ and

$$w_n(t) = n\frac{1+t}{1-t} - \frac{2t(1-t^n)}{(1-t)^2} , \qquad \text{for } t \in [-1,\ 1) . \tag{22.5.2}$$

If $\lambda_{\max} < 1$, then

$$\begin{aligned} \mathrm{Var}_\pi(S_n(h)) &\le \left\{ n\frac{1+\lambda_{\max}}{1-\lambda_{\max}} - \frac{2\lambda_{\max}(1-\lambda_{\max}^n)}{(1-\lambda_{\max})^2} \right\} \|h\|^2_{\mathrm{L}^2(\pi)} \\ &\le \frac{2n}{(1-\lambda_{\max})} \|h\|^2_{\mathrm{L}^2(\pi)} . \end{aligned} \tag{22.5.3}$$

Proof. Since $h \in \mathrm{L}^2_0(\pi)$, we have

$$\begin{aligned} \mathrm{Var}_\pi(S_n(h)) &= \sum_{j=0}^{n-1} \mathbb{E}_\pi[|h(X_j)|^2] + 2\sum_{j=0}^{n-1}\sum_{i=j+1}^{n-1} \mathbb{E}_\pi[h(X_j)\overline{h(X_i)}] , \\ &= n\mathbb{E}_\pi[|h(X_0)|^2] + 2\sum_{\ell=1}^{n-1}(n-\ell)\mathbb{E}_\pi[h(X_0)\overline{h(X_\ell)}] , \end{aligned} \tag{22.5.4}$$

where we have used that for $j \ge i$, $\mathbb{E}_\pi[h(X_i)\overline{h(X_j)}] = \mathbb{E}_\pi[h(X_0)\overline{h(X_{j-i})}]$. For $\ell \in \mathbb{N}$, the definition of the spectral measure ν_h implies

$$\mathbb{E}_\pi[h(X_0)\overline{h(X_\ell)}] = \left\langle h, P^\ell h \right\rangle_{\mathrm{L}^2(\pi)} = \int_{\lambda_{\min}}^{\lambda_{\max}} t^\ell \nu_h(\mathrm{d}t) .$$

Altogether, this gives

$$\mathrm{Var}_\pi(S_n(h)) = \int_{\lambda_{\min}}^{\lambda_{\max}} \left\{ n + 2\sum_{\ell=1}^{n-1} (n-\ell)t^\ell \right\} \nu_h(\mathrm{d}t) = \int_{\lambda_{\min}}^{\lambda_{\max}} w_n(t)\nu_h(\mathrm{d}t) \; .$$

If $\lambda_{\max} < 1$, since the function $t \mapsto w_n(t)$ is increasing, it follows that

$$\mathrm{Var}_\pi(S_n(h)) \le w_n(\lambda_{\max}) \int_{\lambda_{\min}}^{\lambda_{\max}} \nu_h(\mathrm{d}t) \; .$$

The proof of (22.5.3) follows from $\|h\|^2_{\mathrm{L}^2(\pi)} = \int_{-1}^{1} \nu_h(\mathrm{d}t)$. □

Proposition 22.5.2 *Assume that the probability measure π is reversible with respect to P and $h \in \mathrm{L}^2(\pi)$. Then we have the following properties:*

(i) The limit

$$\sigma^2_\pi(h) = \lim_{n\to\infty} n^{-1}\mathbb{E}_\pi[|S_n(h)|^2] \tag{22.5.5}$$

exists in $[0,\infty]$. This limit $\sigma^2_\pi(h)$ is finite if and only if

$$\int_0^1 \frac{1}{1-t}\nu_h(\mathrm{d}t) < \infty \; , \tag{22.5.6}$$

where ν_h is the spectral measure associated with h, and in this case,

$$\sigma^2_\pi(h) = \int_{-1}^{1} \frac{1+t}{1-t}\nu_h(\mathrm{d}t) \; . \tag{22.5.7}$$

(ii) If $\nu_h(\{-1\}) = 0$, then $\lim_{n\to\infty} \sum_{k=1}^n \left\langle h, P^k h\right\rangle_{\mathrm{L}^2(\pi)}$ exists in $[0,\infty]$. This limit is finite if and only if the condition (22.5.6) holds, and in this case,

$$0 < \sigma^2_\pi(h) = \|h\|^2_{\mathrm{L}^2(\pi)} + 2\lim_{n\to\infty} \sum_{k=1}^n \left\langle h, P^k h\right\rangle_{\mathrm{L}^2(\pi)} \; . \tag{22.5.8}$$

Proof. (i) Proposition 22.5.1 shows that

$$n^{-1}\mathbb{E}_\pi\left[|S_n(h)|^2\right] = \int_{-1}^{1} n^{-1} w_n(t)\nu_h(\mathrm{d}t) \; ,$$

where the function w_n is defined in (22.5.2) and ν_h is the spectral measure associated with h. For $t \in [-1,0)$, we get $\lim_{n\to\infty} n^{-1}w_n(t) = (1+t)/(1-t)$ and $|n^{-1}w_n(t)| \le 5$, and Lebesgue's dominated convergence theorem implies

$$\lim_{n\to\infty} \int_{-1}^{0-} n^{-1}w_n(t)\nu_h(\mathrm{d}t) = \int_{-1}^{0-} \frac{1+t}{1-t}\nu_h(\mathrm{d}t) \; .$$

On the interval $[0,1]$, the sequence $\{n^{-1}w_n(t), n \in \mathbb{N}\}$ is increasing and converges to $(1+t)/(1-t)$ (with the convention $1/0 = \infty$). By the monotone convergence theorem, we therefore obtain

$$\lim_{n\to\infty} \uparrow \int_0^1 n^{-1} w_n(t) \nu_h(\mathrm{d}t) = \int_0^1 \frac{1+t}{1-t} \nu_h(\mathrm{d}t) \ .$$

The proof of (ii) follows.

(ii) Assume now that $\nu_h(\{-1\}) = 0$. We show that $\lim_{n\to\infty} \sum_{k=1}^n \left\langle h, P^k h \right\rangle_{\mathrm{L}^2(\pi)}$ exists. Applying (22.2.10), we have

$$R_n = \sum_{k=1}^{n-1} \left\langle h, P^k h \right\rangle_{\mathrm{L}^2(\pi)} = \int_{-1}^1 h_n(t) \nu_h(\mathrm{d}t) \ , \tag{22.5.9}$$

with the convention $R_1 = 0$ and for $n > 1$,

$$h_n(t) = \sum_{k=1}^{n-1} t^k = \begin{cases} t\frac{1-t^{n-1}}{1-t} \ , & t \in [-1,1) \ , \\ n-1, & t = 1 \ . \end{cases} \tag{22.5.10}$$

For $t \in [-1,0)$, we have $0 \le |h_n(t)| \le 2$ and $\lim_{n\to\infty} h_n(t) = t/(1-t)$ for $t \in (-1,0)$. Since $\nu_h(\{-1\}) = 0$, Lebesgue's dominated convergence theorem shows that

$$\lim_{n\to\infty} \int_{-1}^{0-} h_n(t) \nu_h(\mathrm{d}t) = \int_{-1}^{0-} \frac{t}{1-t} \nu_h(\mathrm{d}t) \ .$$

If $t \in [0,1)$, then $h_n(t)$ is nonnegative and $h_n(t) < h_{n+1}(t)$; therefore, the monotone convergence theorem yields

$$\lim_{n\to\infty} \int_0^1 h_n(t) \nu_h(\mathrm{d}t) = \int_0^1 \frac{t}{1-t} \nu_h(\mathrm{d}t) \ ,$$

the latter limit being in $[0,\infty]$.

Since $\|h\|^2_{\mathrm{L}^2(\pi)} = \int_{-1}^1 \nu_h(\mathrm{d}t)$, we get that

$$\|h\|^2_{\mathrm{L}^2(\pi)} + 2 \lim_{n\to\infty} \sum_{k=1}^n \left\langle h, P^k h \right\rangle_{\mathrm{L}^2(\pi)} = \int_{-1}^1 \frac{1+t}{1-t} \nu_h(\mathrm{d}t) \ .$$

The proof is concluded by applying Lemma 21.2.7.

□

Example 22.5.3. Set $\mathsf{X} = \{-1, \ 1\}$, $\pi(\{-1\}) = \pi(\{1\}) = 1/2$ and $P(1, \ \{-1\}) = P(-1, \ \{1\}) = 1$. It is easily seen that π is reversible with respect to P. For all $h \in \mathrm{L}^2(\pi)$, it can be easily checked that

$$\nu_h = \frac{1}{4}|h(1) + h(-1)|^2 \delta_1 + \frac{1}{4}|h(1) - h(-1)|^2 \delta_{-1}$$

satisfies (22.2.10), and by uniqueness, it is the spectral measure ν_h. Now let h be the identity function. Then $|\sum_{i=0}^{n-1} h(X_i)| \le 1$, so $\lim_{n\to\infty} n^{-1}\mathbb{E}_\pi\left[S_n^2(h)\right] = 0$. On the other hand, $\mathrm{Cov}_\pi(h(X_0), h(X_k)) = \left\langle h, P^k h\right\rangle_{\mathrm{L}^2(\pi)} = (-1)^k$, which implies that

$$-1 = \liminf_{n\to\infty} \sum_{k=1}^{n} \left\langle h, P^k h\right\rangle_{\mathrm{L}^2(\pi)} < \limsup_{n\to\infty} \sum_{k=1}^{n} \left\langle h, P^k h\right\rangle_{\mathrm{L}^2(\pi)} = 0\,.$$

Nevertheless, Proposition 22.5.2 is not violated, since $\nu_h(\{-1\}) = 1$. ◀

We will now find conditions under which $\nu_h(\{-1,1\}) = 0$, where ν_h is the spectral measure associated with h.

Lemma 22.5.4 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $\pi \in \mathbb{M}_1(\mathscr{X})$. Assume that π is reversible with respect to P. Then:*

(i) For all $h \in \mathrm{L}_0^2(\pi)$ such that $\nu_h(\{-1,1\}) = 0$, we have

$$\lim_{k\to\infty} \left\langle h, P^k h\right\rangle_{\mathrm{L}^2(\pi)} = 0\,.$$

(ii) For all $h \in \mathrm{L}_0^2(\pi)$, if $\lim_{k\to\infty} \left\langle h, P^k h\right\rangle_{\mathrm{L}^2(\pi)} = 0$, then $\nu_h(\{-1,1\}) = 0$.

(iii) If $\lim_{k\to\infty} \left\|P^k(x,\cdot) - \pi\right\|_{\mathrm{TV}} = 0$ for π-almost all $x \in \mathsf{X}$, then for all $h \in \mathrm{L}_0^2(\pi)$,

$$\lim_{k\to\infty} \left\langle h, P^k h\right\rangle_{\mathrm{L}^2(\pi)} = 0\,.$$

Proof. (i) By Lebesgue's dominated convergence theorem (since $|t^k| \le 1$ for $-1 \le t \le 1$, $\int_{-1}^{1} \nu_h(\mathrm{d}h) = \|h\|^2_{\mathrm{L}^2(\pi)} < \infty$), we have

$$\lim_{k\to\infty} \left\langle h, P^k h\right\rangle = \lim_{k\to\infty} \int_{-1}^{1} t^k \nu_h(\mathrm{d}t) = \int_{-1}^{1} (\lim_{k\to\infty} t^k)\nu_h(\mathrm{d}t) = 0\,,$$

where we have used that $\lim_{k\to\infty} t^k = 0$, ν_h-*a.e.*

(ii) We may write $\nu_h = \nu_h^0 + \nu_h^{\mathrm{a}}$, where $\nu_h^0(\{-1,1\}) = 0$ and $\nu_h^{\mathrm{a}}(\{-1,1\}) = \nu_h(\{-1,1\})$. For all k, we have

$$\left\langle h, P^k h\right\rangle_{\mathrm{L}^2(\pi)} = (-1)^k \nu_h(\{-1\}) + \nu_h(\{1\}) + \int_{-1}^{1} t^k \nu_h^0(\mathrm{d}t)\,.$$

Since $\lim_{k\to\infty} \int_{-1}^{1} t^k \nu_h^0(\mathrm{d}t) = 0$, we have

$$0 = \lim_{k\to\infty} \left\langle h, P^{2k} h\right\rangle_{\mathrm{L}^2(\pi)} = \nu_h(\{-1\}) + \nu_h(\{1\})$$

$$0 = \lim_{k\to\infty} \left\langle h, P^{2k+1} h\right\rangle_{\mathrm{L}^2(\pi)} = -\nu_h(\{-1\}) + \nu_h(\{1\}).$$

The proof follows.

(iii) For every bounded measurable complex-valued function h satisfying $\pi(h) = 0$, we get $\lim_{k\to\infty} P^k h(x) = 0$ for π-almost all $x \in \mathsf{X}$. Since

$$\left\langle h, P^k h \right\rangle_{\mathrm{L}^2(\pi)} = \mathbb{E}_\pi[h(X_0)\overline{h(X_k)}] = \mathbb{E}_\pi[h(X_0)\overline{P^k h(X_0)}] ,$$

Lebesgue's dominated convergence theorem implies

$$\lim_{k\to\infty} \left\langle h, P^k h \right\rangle_{\mathrm{L}^2(\pi)} = \lim_{k\to\infty} \mathbb{E}_\pi[h(X_0)\overline{P^k h(X_0)}] = 0 .$$

The proof follows, since the space of bounded measurable complex-valued functions is dense in $\mathrm{L}^2(\pi)$ and P is a bounded linear operator in $\mathrm{L}^2(\pi)$.

□

Theorem 22.5.5. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, $\pi \in \mathbb{M}_1(\mathscr{X})$, and h a real-valued function in $\mathrm{L}^2_0(\pi)$. Assume that π is reversible with respect to P and $\int_0^1 (1-t)^{-1} \nu_h(\mathrm{d}t) < \infty$, where ν_h is the spectral measure associated with h defined in (22.2.10). Then*

$$n^{-1/2} \sum_{j=0}^{n-1} h(X_j) \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0, \sigma^2(h))$$

with

$$\sigma^2(h) = \int_{-1}^{1} \frac{1+t}{1-t} \nu_h(\mathrm{d}t) \tag{22.5.11}$$

$$= \|h\|^2_{\mathrm{L}^2(\pi)} + 2 \lim_{n\to\infty} \sum_{k=1}^{n} \left\langle h, P^k h \right\rangle_{\mathrm{L}^2(\pi)} = \lim_{n\to\infty} n^{-1} \mathbb{E}_\pi[\{S_n(h)\}^2] < \infty . \tag{22.5.12}$$

Remark 22.5.6. Under the additional assumption $\nu_h(\{-1\}) = 0$, the proof of Theorem 22.5.5 is a simple consequence of Theorem 21.4.1. Indeed, (22.2.10) and reversibility show that for all $m, k \geq 0$,

$$\left\langle P^m h, P^k h \right\rangle_{\mathrm{L}^2(\pi)} = \left\langle h, P^{m+k} h \right\rangle_{\mathrm{L}^2(\pi)} = \int_{-1}^{1} t^{m+k} \nu_h(\mathrm{d}t) .$$

Therefore, the condition (21.4.1) of Theorem 21.4.1 is implied by the existence of the limit $\lim_{n\to\infty} \sum_{k=1}^n \left\langle h, P^k h \right\rangle_{\mathrm{L}^2(\pi)}$, which was shown to hold under the stated assumptions in Proposition 22.5.2. ▲

The proof of Theorem 22.5.5 is an application of Theorem 21.3.2. Before proceeding with the proof of Theorem 22.5.5, we will establish two bounds on the solutions for the resolvent equation (see (21.3.1)),

$$(1+\lambda)\hat{h}_\lambda - P\hat{h}_\lambda = h \ , \quad \lambda > 0 \ .$$

For $\lambda > 0$, the resolvent equation has a unique solution, which is given by

$$\hat{h}_\lambda = \{(1+\lambda)\mathrm{I} - P\}^{-1} h = (1+\lambda)^{-1} \sum_{k=0}^{\infty} (1+\lambda)^{-k} P^k h \ . \tag{22.5.13}$$

Lemma 22.5.7 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, $\pi \in \mathbb{M}_1(\mathscr{X})$, and $h \in \mathrm{L}^2_0(\pi)$. Assume that π is reversible with respect to P. If $\int_0^1 (1-t)^{-1} \nu_h(\mathrm{d}t)$, then*

$$\lim_{\lambda \to 0} \lambda \left\langle \hat{h}_\lambda, \hat{h}_\lambda \right\rangle_{\mathrm{L}^2(\pi)} = 0 \ . \tag{22.5.14}$$

Proof. Consider first (22.5.14). Since $\hat{h}_\lambda$ is the solution to the resolvent equation (21.3.1), it follows that

$$\begin{aligned}
\lambda \left\langle \hat{h}_\lambda, \hat{h}_\lambda \right\rangle_\pi &= \lambda (1+\lambda)^{-2} \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} (1+\lambda)^{-k} (1+\lambda)^{-\ell} \left\langle P^k h, P^\ell h \right\rangle_{\mathrm{L}^2(\pi)} \\
&= \lambda (1+\lambda)^{-2} \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \int_{-1}^{1} (1+\lambda)^{-k} t^k (1+\lambda)^{-\ell} t^\ell \nu_h(\mathrm{d}t) \\
&= \int_{-1}^{1} \frac{\lambda}{(1+\lambda-t)^2} \nu_h(\mathrm{d}t) \ .
\end{aligned}$$

Since $\lambda/(1+\lambda-t)^2 \leq (1-t)^{-1}$ and $\int_0^1 (1-t)^{-1} \nu_h(\mathrm{d}t) < \infty$, we conclude by Lebesgue's dominated convergence theorem. □

Define (see (21.3.5))

$$H_\lambda(x_0, x_1) = \hat{h}_\lambda(x_1) - P\hat{h}_\lambda(x_0) \ .$$

Lemma 22.5.8 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, $\pi \in \mathbb{M}_1(\mathscr{X})$, and $h \in \mathrm{L}^2_0(\pi)$. Assume that π is reversible with respect to P. Set $\pi_1 = \pi \otimes P$. If $\int_0^1 (1-t)^{-1} \nu_h(\mathrm{d}t)$, then there exists a function $H \in \mathrm{L}^2(\pi_1)$ such that $\lim_{\lambda \downarrow 0} \|H_\lambda - H\|_{\mathrm{L}^2(\pi_1)} = 0$. In addition,*

$$\int_{-1}^{1} \frac{1+t}{1-t} \nu_h(\mathrm{d}t) = \|H\|^2_{\mathrm{L}^2(\pi_1)} \ . \tag{22.5.15}$$

Proof. First, we observe that for $g \in \mathrm{L}^2(\pi)$,

$$\begin{aligned}
\mathbb{E}_\pi[|g(X_1) - Pg(X_0)|^2] &= \mathbb{E}_\pi[|g(X_1)|^2] - \mathbb{E}_\pi[|Pg(X_0)|^2] \\
&= \langle g, g \rangle_{\mathrm{L}^2(\pi)} - \langle Pg, Pg \rangle_\pi = \left\langle g, (\mathrm{I} - P^2) g \right\rangle_{\mathrm{L}^2(\pi)} \ .
\end{aligned}$$

Let $0 < \lambda_1, \lambda_2$. Applying this identity with

$$H_{\lambda_1}(X_0,X_1) - H_{\lambda_2}(X_0,X_1) = \{\hat{h}_{\lambda_1} - \hat{h}_{\lambda_2}\}(X_1) - \{\hat{h}_{\lambda_1} - \hat{h}_{\lambda_2}\}(X_0)$$

and using that $\hat{h}_{\lambda_i} = (1+\lambda_i)\mathrm{I} - P$, we get

$$\|H_{\lambda_1} - H_{\lambda_2}\|^2_{\mathrm{L}^2(\pi_1)} = \int_{-1}^{1} (1-t^2)\left(\frac{1}{1+\lambda_1 - t} - \frac{1}{1+\lambda_2 - t}\right)^2 \nu_h(\mathrm{d}t) \; .$$

The integrand is bounded by $4(1+t)/(1-t)$, which is μ_f integrable and goes to 0 as $\lambda_1, \lambda_2 \to 0$, showing that H_λ has a limit in $\mathrm{L}^2(\pi_1)$, i.e., there exists $H \in \mathrm{L}^2(\pi_1)$ such that $\|H_\lambda - H\|_{\mathrm{L}^2(\pi_1)} \to_{\lambda \downarrow 0} 0$.

Along the same lines, we obtain

$$\|H_\lambda\|_{\mathrm{L}^2(\pi_1)} = \left\langle \hat{h}_\lambda, (\mathrm{I} - P^2)\hat{h}_\lambda \right\rangle_{\mathrm{L}^2(\pi)} = \int_{-1}^{1} \frac{1-t^2}{(1+\lambda - t)^2} \nu_h(\mathrm{d}t) \; .$$

Since the integrand is bounded above by $(1+t)/(1-t)^{-1}$, by Lebesgue's dominated convergence theorem, as $\lambda \downarrow 0$, the previous expression converges to

$$\|H\|_{\mathrm{L}^2(\pi_1)} = \lim_{\lambda \downarrow 0} \|H_\lambda\|_{\mathrm{L}^2(\pi_1)} = \int_{-1}^{1} \frac{1+t}{1-t} \nu_h(\mathrm{d}t) \; .$$

□

Proof (of Theorem 22.5.5). We use Theorem 21.3.2. Lemma 22.5.7 shows that $\lim_{\lambda \downarrow 0} \sqrt{\lambda} \left\| \hat{f}_\lambda \right\|_{\mathrm{L}^2(\pi)} = 0$. Lemma 22.5.8 shows that there exists $H \in \mathrm{L}^2(\pi_1)$ such that $\lim_{\lambda \downarrow 0} \|H_\lambda - H\|_{\mathrm{L}^2(\pi_1)} = 0$.

Theorem 21.3.2 implies that $\sqrt{n} S_n(h) \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0, \|H\|_{\mathrm{L}^2(\pi_1)})$. The proof follows from Lemma 22.5.8, Eq. (22.5.15). □

Corollary 22.5.9 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ and $\pi \in \mathbb{M}_1(\mathscr{X})$. Assume that π is reversible with respect to P. Let $h \in \mathrm{L}^2_0(\pi)$ be a real-valued function satisfying*

$$\lim_{n \to \infty} n^{-1} \mathbb{E}_\pi[|S_n(h)|^2] < \infty \text{ and } \lim_{n \to \infty} \langle h, P^n h \rangle_{\mathrm{L}^2(\pi)} = 0. \qquad (22.5.16)$$

Then

$$n^{-1/2} \sum_{j=0}^{n-1} h(X_j) \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0, \sigma^2(h)) \; ,$$

where $\sigma^2(h) > 0$ is given by (22.5.11). If $\lambda_{\max} = \sup\left\{ \lambda \,:\, \lambda \in \mathrm{Spec}(P | \mathrm{L}^2_0(\pi)) \right\} < 1$, in particular, if P is geometrically ergodic, then the condition (22.5.16) is satisfied for all $h \in \mathrm{L}^2_0(\pi)$.

Proof. Apply Proposition 22.5.2, Lemma 22.5.4, and Theorems 22.3.11 and 22.5.5. □

22.6 Exercises

22.1. Let f,g be two π-integrable functions. Show that $P[f+g] = Pf + Pg$, π-*a.e.*

22.2. 1. Let P be the Markov kernel given in (22.3.7) with $|\phi| < 1$. Set $\pi = \mathrm{N}(0, \sigma^2/(1-\phi^2))$. Show that π is reversible with respect to P.
2. Prove (22.3.8).
3. Provide a lower bound for $\|P^n(x,\cdot) - \pi\|_{\mathrm{TV}}$.

22.3. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ with unique invariant probability π. For $n_0 \in \mathbb{N}$ and $f \in \mathbb{F}(\mathsf{X})$, set

$$S_{n,n_0}(f) = n^{-1} \sum_{j=n_0}^{n} f(X_{j+n_0}) \,.$$

For $\nu \in \mathbb{M}_1(\mathscr{X})$, define

$$e_\nu(S_{n,n_0}, f) = \left\{ \mathbb{E}_\nu \left[\{ S_{n,n_0}(f) - \pi(f) \}^2 \right] \right\}^{1/2} .$$

Let $r \in [1,2]$. Assume that $f \in \mathrm{L}_0^r(\pi)$ and let $\nu \in \mathbb{M}_{r/(r-1)}(\pi)$ be a probability measure.

1. Show that

$$e_\nu(S_{n,n_0}, f)^2 = e_\pi(S_n, f)^2 + \frac{1}{n^2} \sum_{j=1}^{n} L_{j+n_0}(f^2) + \frac{2}{n^2} \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} L_{j+n_0}(f P^{k-j} f), \tag{22.6.1}$$

where for $h \in \mathrm{L}^r(\pi)$ and $i \in \mathbb{N}$,

$$L_i(h) = \left\langle (P^i - \Pi)h, \left(\frac{\mathrm{d}\nu}{\mathrm{d}\pi} - 1 \right) \right\rangle . \tag{22.6.2}$$

2. Show that if $r \in [1,2)$, then for all $h \in \mathrm{L}_0^r(\pi)$ and $k \in \mathbb{N}$, we have

$$|L_k(h)| \le 2^{2/r} \{1 - \mathrm{Abs.Gap}_{\mathrm{L}^2(\pi)}(R)\}^{2k\frac{r-1}{r}} \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\pi} - 1 \right\|_{\mathrm{L}^{\frac{r}{r-1}}(\pi)} \|h\|_{\mathrm{L}^r(\pi)} . \tag{22.6.3}$$

3. Show that if $r = 1$ and the transition kernel is $\mathrm{L}^1(\pi)$-exponentially convergent (for all $|||P^n|||_{\mathrm{L}_0^1(\pi)} \le M\alpha^n$), then for all $h \in \mathrm{L}_0^1(\pi)$ and $k \in \mathbb{N}$,

$$|L_k(h)| \le M\alpha^k \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\pi} - 1 \right\|_{\mathrm{L}^\infty(\pi)} \|h\|_{\mathrm{L}^1(\pi)} . \tag{22.6.4}$$

22.4. In this exercise, we construct a reversible Markov kernel P and a function h such that $n^{-1/2}\sum_{i=0}^{n-1} h(X_i) \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0,\sigma^2)$ but $\lim_{n\to\infty} n^{-1}\mathbb{E}_\pi[S_n(h)^2] = \infty$.

Let X be the integers, with h the identity function. Consider the Markov kernel P on X with transition probabilities given by $P(0,\ 0) = 0, P(0,\ y) = c|y|^{-4}$ for $y \neq 0$ (with $c = 45/\pi^4$) and for $x \neq 0$,

$$P(x,\ y) = \begin{cases} |x|^{-1}, & y = 0, \\ 1 - |x|^{-1}, & y = -x, \\ 0, & \text{otherwise.} \end{cases}$$

That is, the chain jumps from 0 to a random site x and then oscillates between $-x$ and x for a geometric amount of time with mean $|x|$, before returning to 0.

1. Show that this Markov kernel is positive and identify its unique invariant probability.
2. Prove that $n^{-1/2}\sum_{i=0}^{n-1} h(X_i) \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0,\sigma^2)$. [Hint: Use Theorem 6.7.1.]
3. Show that $\mathrm{Var}_\pi(X_0) = \infty$.
4. For $n \geq 2$, set $S_n = \sum_{i=0}^{n} X_n$ and $D_n = \{\tau_\alpha \leq n\}$. Show that $0 < \mathbb{P}_\pi(D_n) < 1$ and that for even n, $S_n \mathbb{1}_{\{\tau_\alpha > n\}} = X_0$.
5. Show that $\mathrm{Var}_\pi\left(\sum_{i=0}^{n} X_i\right)$ is infinite for n even.

22.5. Let P_0 and P_1 be Markov transition kernels on $(\mathsf{X},\mathscr{X})$ with invariant probability π. We say that P_1 dominates P_0 on the off-diagonal, written $P_0 \leq P_1$, if for all $A \in \mathscr{X}$ and π-*a.e.* all x in X,

$$P_0(x, A \setminus \{x\}) \leq P_1(x, A \setminus \{x\}) \ .$$

Let P_0 and P_1 be Markov transition kernels on $(\mathsf{X},\mathscr{X})$ with invariant probability π. We say that P_1 dominates P_0 in the covariance ordering, written $P_0 \preceq P_1$, if for all $f \in \mathrm{L}^2(\pi)$,

$$\langle f, P_1 f\rangle \leq \langle f, P_0 f\rangle \ .$$

Let P_0 and P_1 be Markov transition kernels on $(\mathsf{X},\mathscr{X})$, with invariant probability π. Assume that $P_0 \leq P_1$. For all $x \in \mathsf{X}$ and $A \in \mathscr{X}$, define

$$P(x,A) = \delta_x(A) + P_1(x,A) - P_0(x,A) \ .$$

1. Show that P is a Markov kernel.
2. Show that for all $f \in \mathrm{L}^2(\pi)$,

$$\langle f, P_0 f\rangle - \langle f, P_1 f\rangle = \iint \pi(\mathrm{d}x) P(x, \mathrm{d}y) \left(f(x) - f(y)\right)^2 / 2,$$

   and that $P_0 \preceq P_1$.

22.6. We use the notation of Exercise 22.5. Let P_0 and P_1 be Markov kernels on $\mathsf{X} \times \mathscr{X}$ and $\pi \in \mathbb{M}_1(\mathscr{X})$. Assume that π is reversible with respect to P_0 and P_1. The objective of this exercise is to show that if $P_0 \preceq P_1$, then for all $f \in \mathrm{L}^2_0(\pi)$,

$$v_1(f,P_1) \le v_0(f,P_0) ,$$

where for $i \in \{0,1\}$,

$$v_i(f,P_i) = \pi(f^2) + 2 \lim_{n\to\infty} \sum_{k=1}^{n} \pi(fP_i^k f) .$$

For all $\alpha \in (0,1)$, set $P_\alpha = (1-\alpha)P_0 + \alpha P_1$. For $\lambda \in (0,1)$, define

$$w_\lambda(\alpha) = \sum_{k=0}^{\infty} \lambda^k \left\langle f, P_\alpha^k f \right\rangle .$$

1. Show that for all $\alpha \in (0,1)$,

$$\frac{\mathrm{d}w_\lambda(\alpha)}{\mathrm{d}\alpha} = \sum_{k=0}^{\infty} \lambda^k \sum_{i=1}^{k} \left\langle f, P_\alpha^{i-1}(P_1 - P_0)P_\alpha^{k-i} f \right\rangle .$$

2. Show that $w_\lambda(1) \le w_\lambda(0)$ for all $\lambda \in (0,1)$.

22.7 Bibliographical Notes

Spectral theory is a very active area, and this chapter provides only a very short and incomplete introduction to the developments in this field. Our presentation closely follows Rudolf (2012) (see also Rudolf and Schweizer (2015)).

Theorem 22.3.11 was established in (Roberts and Rosenthal 1997, Theorem 2.1) and Roberts and Tweedie (2001). These results were further improved in Kontoyiannis and Meyn (2012). Proposition 22.3.3 is established in (Rudolf 2012, Proposition 3.17) (see also Rudolf (2009) and Rudolf (2010)).

The derivation of Cheeger's inequality is taken from Lawler and Sokal (1988). Applications of Cheeger's inequality to computing mixing rates of Markov chains were considered by, among many others, Lovász and Simonovits (1993), Kannan et al. (1995), Yuen (2000; 2001; 2002), and Jarner and Yuen (2004).

Proposition 22.5.2 is borrowed from Häggström and Rosenthal (2007) (the proof presented here is different). Theorem 22.5.5 was first established in Varadhan (1985; 1986) (see also Tóth (1986), Tóth (2013), Cuny and Lin (2016), Cuny (2017)). The proof given here is borrowed from Maxwell and Woodroofe (2000).

The use of spectral theory is explored in depth in Huang et al. (2002) and in the series of papers Kontoyiannis and Meyn (2003), Kontoyiannis and Meyn (2005), and Kontoyiannis and Meyn (2012).

We have not covered the theory of quasicompact operators. The book Hennion and Hervé (2001) is a worthwhile introduction to the subject with an emphasis on limit theorems. Recent developments are presented in Hennion and Hervé (2001), Hervé and Ledoux (2014a), Hervé and Ledoux (2014b), Hervé and Ledoux (2016).

22.A Operators on Banach and Hilbert Spaces

We introduce the basic definitions and notation that we have used in this book. A basic introduction to operator theory is given in Gohberg and Goldberg (1981), covering most of what is needed for the development of Chapter 22. A much more detailed account is given in Simon (2015). Let $(\mathsf{H}, \|\cdot\|_{\mathsf{H}})$ and $(\mathsf{G}, \|\cdot\|_{\mathsf{G}})$ be complex Banach spaces. Whenever there is no ambiguity on the space, we write $\|\cdot\|$ instead of $\|\cdot\|_{\mathsf{H}}$. A function $A : \mathsf{H} \to \mathsf{G}$ is a linear operator from H to G if for all $x, y \in \mathsf{H}$ and $\alpha \in \mathbb{C}$, $A(x+y) = A(x) + A(y)$ and $A(\alpha x) = \alpha A(x)$. For convenience, we often write Ax instead of $A(x)$. The linear operator A is said to be bounded if $\sup_{\|x\|_{\mathsf{H}} \leq 1} \|Ax\|_{\mathsf{G}} < \infty$. The (operator) norm of A, written $|||A|||_{\mathsf{H}\to\mathsf{G}}$, is given by

$$|||A|||_{\mathsf{H}\to\mathsf{G}} := \sup_{\|y\|_{\mathsf{H}} \leq 1} \|Ay\|_{\mathsf{G}} = \sup_{\|y\|_{\mathsf{H}} = 1} \|Ay\|_{\mathsf{G}} \, . \tag{22.A.1}$$

The identity operator $\mathrm{I} : \mathsf{H} \to \mathsf{H}$, defined by $\mathrm{I}x = x$, is a bounded linear operator, and its norm is 1. Denote by $\mathsf{BL}(\mathsf{H}, \mathsf{G})$ the set of bounded linear operators from H to G. For simplicity, $\mathsf{BL}(\mathsf{H}, \mathsf{H})$ will be abbreviated $\mathsf{BL}(\mathsf{H})$. If $A \in \mathsf{BL}(\mathsf{H})$, we will use the shorthand notation $|||A|||_{\mathsf{H}}$ instead of $|||A|||_{\mathsf{H}\to\mathsf{H}}$. If A and B are in $\mathsf{BL}(\mathsf{H}, \mathsf{G})$, it is easy to check that

(i) $\alpha A + \beta B \in \mathsf{BL}(\mathsf{H}, \mathsf{G})$, for all $\alpha, \beta \in \mathbb{C}$;
(ii) $|||\alpha A|||_{\mathsf{H}\to\mathsf{G}} = |\alpha| \, |||A|||_{\mathsf{H}\to\mathsf{G}}$, for all $\alpha \in \mathbb{C}$;
(iii) $|||A + B|||_{\mathsf{H}\to\mathsf{G}} \leq |||A|||_{\mathsf{H}\to\mathsf{G}} + |||B|||_{\mathsf{H}\to\mathsf{G}}$;
(iv) if $A, B \in \mathsf{BL}(\mathsf{H})$ then defining AB by $ABx = A(Bx)$, we have $AB, BA \in \mathsf{BL}(\mathsf{H})$ and $|||CA|||_{\mathsf{H}} \leq |||C|||_{\mathsf{H}} \, |||A|||_{\mathsf{H}}$.

Theorem 22.A.1. *The set* $\mathsf{BL}(\mathsf{H}, \mathsf{G})$ *equipped with its operator norm is a Banach space.*

Theorem 22.A.2. *Let* $A \in \mathsf{BL}(\mathsf{H}, \mathsf{G})$. *The following statements are equivalent:*

(i) A is continuous at a point.
(ii) A is uniformly continuous on H.
(iii) A is bounded.

Proof. See (Gohberg and Goldberg 1981, Theorem 3.1). □

An operator $A \in \mathsf{BL}(\mathsf{H})$ is called invertible if there exists an operator $A^{-1} \in \mathsf{BL}(\mathsf{H})$ such that $AA^{-1}x = A^{-1}Ax$ for every $x \in \mathsf{H}$. The operator A^{-1} is called the inverse of A. If A and B are invertible operators in $\mathsf{BL}(\mathsf{H})$, then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.

Theorem 22.A.3 (Riesz–Thorin interpolation theorem). *Let* $(\mathsf{X}, \mathscr{X}, \mu)$ *be a* σ-*finite measure space,* $p_0, p_1, q_0, q_1 \in [1, \infty]$, *and* $T \in \mathsf{BL}(\mathrm{L}^{p_j}(\mu), \mathrm{L}^{q_j}(\mu))$ *for* $j \in \{0, 1\}$. *For* $\theta \in [0, 1]$, *we define* $p_\theta^{-1} = (1-\theta)p_0^{-1} + \theta p_1^{-1}$ *and* $q_\theta^{-1} = (1-\theta)q_0^{-1} + \theta q_1^{-1}$. *Then* $T \in \mathsf{BL}(\mathrm{L}^{p_\theta}(\mu), \mathrm{L}^{q_\theta}(\mu))$ *and*

$$|||T|||_{\mathrm{L}^{p_\theta}(\mu) \to \mathrm{L}^{q_\theta}(\mu)} \leq \left\{ |||T|||_{\mathrm{L}^{p_0}(\mu) \to \mathrm{L}^{q_0}(\mu)} \right\}^{1-\theta} \left\{ |||T|||_{\mathrm{L}^{p_1}(\mu) \to \mathrm{L}^{q_1}(\mu)} \right\}^{\theta} .$$

Proof. See (Lerner 2014, Theorem 9.1.2). □

The kernel of $A \in \mathsf{BL}(\mathsf{H})$ is denoted by $\mathrm{Ker}(A)$. It is the closed subspace defined by $\{x \in \mathsf{H} : Ax = 0\}$. The operator A is said to be injective if $\mathrm{Ker}(A) = \{0\}$. The *range* (or *image*) of A, written $\mathrm{Ran}(A)$, is the subspace $\{Ax : x \in \mathsf{H}\}$. If $\mathrm{Ran}(A)$ is finite-dimensional, A is called an operator of *finite rank* and $\dim \mathrm{Ran}(A)$ is the rank of A.

Lemma 22.A.4 *Let* $A \in \mathsf{BL}(\mathsf{H})$ *be such that for all* $x \in \mathsf{H}$, $\|Ax\| \geq c\|x\|$, *where* c *is a positive constant. Then for all* $n \in \mathbb{N}$, *the range of* A^n *(denoted by* $\mathrm{Ran}(A^n)$*) is closed.*

Proof. Since $\|A^n x\| \geq c^n \|x\|$ for all $x \in \mathsf{H}$, it suffices to establish the property with $n = 1$. Let $\{y_n,\ n \in \mathbb{N}\}$ be a convergent sequence of elements of $\mathrm{Ran}(A)$ converging to y. Then $y_n = Ax_n$ for some sequence $\{x_n,\ n \in \mathbb{N}\}$, and we need to show that $y = Ax$ for some x. Since $\{y_n,\ n \in \mathbb{N}\}$ is convergent, it is a Cauchy sequence. Now

$$\|x_n - x_m\| \leq \frac{1}{c}\|A(x_n - x_m)\| = \frac{1}{c}\|y_n - y_m\| ,$$

so $\{x_n,\ n \in \mathbb{N}\}$ is a Cauchy sequence, and it therefore converges to some element x. Then since A is continuous, we have $y = \lim_n y_n = \lim_n Ax_n = A \lim_n x_n = Ax$, as required. □

Theorem 22.A.5. *Assume that* $T \in \mathsf{BL}(\mathsf{H})$ *and* $|||T|||_{\mathsf{H}} < 1$. *Then* $\mathrm{I} - T$ *is invertible, and for every* $y \in \mathsf{H}$, $(\mathrm{I} - T)^{-1} y = \sum_{k=0}^{\infty} T^k y$ *with the convention* $T^0 = \mathrm{I}$. *Moreover,*

$$\lim_{n \to \infty} |||(\mathrm{I} - T)^{-1} - \sum_{k=0}^{n} T^k |||_{\mathsf{H}} = 0 ,$$

and $|||(\mathrm{I} - T)^{-1}|||_{\mathsf{H}} \leq (1 - |||T|||_{\mathsf{H}})^{-1}$.

Proof. Given $y \in \mathsf{H}$, the series $\sum_{k=0}^{\infty} T^k y$ converges. Indeed, let $s_n = \sum_{k=0}^{n} T^k y$. Then for $n > m$,

$$\|s_n - s_m\| \le \sum_{k=m+1}^{n} \|T^k y\| \le \|y\| \sum_{k=m+1}^{n} \{|||T|||_{\mathsf{H}}\}^k \to 0$$

as $m,n \to \infty$. Since H is complete, the sequence $\{s_n,\ n \in \mathbb{N}\}$ converges. Define $S : \mathsf{H} \to \mathsf{H}$ by $Sy = \sum_{k=0}^{\infty} T^k y$. The operator S is linear and

$$\|Sy\| \le \sum_{k=0}^{\infty} \|T^k y\| \le \sum_{k=0}^{\infty} \{|||T|||_{\mathsf{H}}\}^k \|y\| = (1 - |||T|||_{\mathsf{H}})^{-1} \|y\| \,.$$

Hence $|||S|||_{\mathsf{H}} \le (1 - |||T|||_{\mathsf{H}})^{-1}$ and

$$\begin{aligned}(\mathrm{I} - T)Sy &= (\mathrm{I} - T) \sum_{k=0}^{\infty} T^k y = \sum_{k=0}^{\infty} (\mathrm{I} - T) T^k y = \sum_{k=0}^{\infty} T^k (\mathrm{I} - T) y \\ &= S(\mathrm{I} - T)y = \sum_{k=0}^{\infty} T^k y - \sum_{k=0}^{\infty} T^{k+1} y = y.\end{aligned}$$

Therefore, $\mathrm{I} - T$ is invertible and $(\mathrm{I} - T)^{-1} = S$. Finally, as $n \to \infty$,

$$|||(\mathrm{I} - T)^{-1} - \sum_{k=0}^{n} T^k |||_{\mathsf{H}} \le \sum_{k=n+1}^{\infty} \{|||T|||_{\mathsf{H}}\}^k \to 0.$$

□

Corollary 22.A.6 *Let $T \in \mathsf{BL}(\mathsf{H})$ be invertible. Assume that $S \in \mathsf{BL}(\mathsf{H})$ and $|||T - S|||_{\mathsf{H}} < \left\{|||T^{-1}|||_{\mathsf{H}}\right\}^{-1}$. Then S is invertible,*

$$S^{-1} = \sum_{k=0}^{\infty} [T^{-1}(T - S)]^k T^{-1},$$

and

$$|||T^{-1} - S^{-1}|||_{\mathsf{H}} \le \frac{\left\{|||T^{-1}|||_{\mathsf{H}}\right\}^2 |||\,T - S|||_{\mathsf{H}}}{1 - |||T^{-1}\,|||_{\mathsf{H}}\,|||T - S|||_{\mathsf{H}}}. \tag{22.A.2}$$

Proof. Since $S = T - (T - S) = T[\mathrm{I} - T^{-1}(T - S)]$ and $|||T^{-1}(T - S)|||_{\mathsf{H}} < |||T^{-1}\,|||_{\mathsf{H}}$ $|||T - S|||_{\mathsf{H}} < 1$, it follows from Theorem 22.A.5 that S is invertible and

$$S^{-1} y = [\mathrm{I} - T^{-1}(T - S)]^{-1} T^{-1} y = \sum_{k=0}^{\infty} [T^{-1}(T - S)]^k T^{-1} y \,.$$

Hence $|||T^{-1} - S^{-1}|||_{\mathsf{H}} \le \{|||T|||_{\mathsf{H}}\}^{-1} \sum_{k=1}^{\infty} \left\{|||T^{-1}|||_{\mathsf{H}}\right\}^k |||\,T - S|||_{\mathsf{H}}$, which proves (22.A.2). □

If we define the distance $d(T,S)$ between the operators T and S to be $|||T - S|||_{\mathsf{H}}$, then Corollary 22.A.6 shows that the set X of invertible operators in $\mathsf{BL}(\mathsf{H})$ is an open set in the sense that if T is in X, then there exists $r > 0$ such that $d(T,S) < r$ implies $S \in \mathsf{X}$.

Also, the inverse operation is continuous with respect to d, i.e., if T is invertible and $d(T,T_n) \to 0$, then T_n is invertible for all n sufficiently large and $d(T^{-1},T_n^{-1}) \to 0$.

Given a linear operator T that maps a finite-dimensional vector space H into H, it is well known from linear algebra that the equation $\lambda x - Tx = y$ has a unique solution for every $y \in \mathsf{H}$ if and only if $\det(\lambda \mathrm{I} - T) \neq 0$, where by abuse of notation, T is the matrix associated to the operator T in a given basis of H. Therefore, $\lambda \mathrm{I} - T$ is invertible for all but a finite number of λ. If H is an infinite-dimensional Banach space, then the set of those λ for which $\lambda \mathrm{I} - T$ is not invertible is a set that is usually more difficult to determine.

Definition 22.A.7 *Given $T \in \mathsf{BL}(\mathsf{H})$, a point $\lambda \in \mathbb{C}$ is* regular *if $\lambda I - T$ is invertible, i.e., there exists a bounded linear operator $R_\lambda(T)$ such that*

$$(\lambda \mathrm{I} - T)R_\lambda(T) = R_\lambda(T)(\lambda \mathrm{I} - T) = \mathrm{I}\,.$$

The set $\mathrm{Res}(T|\mathsf{H})$ *of regular points is called the* resolvent set *of T, i.e.,*

$$\mathrm{Res}(T|\mathsf{H}) := \{\lambda \in \mathbb{C} : \lambda \mathrm{I} - T \textit{ is invertible}\}\,. \tag{22.A.3}$$

The spectrum $\mathrm{Spec}(T|\mathsf{H})$ *of T is the complement of* $\mathrm{Res}(T|\mathsf{H})$.

$$\mathrm{Spec}(T|\mathsf{H}) = \mathbb{C} \setminus \mathrm{Res}(T|\mathsf{H})\,. \tag{22.A.4}$$

If $\lambda \in \mathrm{Res}(T|\mathsf{H})$, then for all $x \in \mathsf{H}$,

$$\|(T - \lambda \mathrm{I})R_\lambda(T)x\|_{\mathsf{H}} = \|x\|_{\mathsf{H}}\,,$$

so if $y = R_\lambda(T)x$, then $\|y\|_{\mathsf{H}} \le |||R_\lambda(T)\,|||_{\mathsf{H}}\,\|x\|_{\mathsf{H}}$, which implies

$$\|(T - \lambda \mathrm{I})y\|_{\mathsf{H}} = \|x\|_{\mathsf{H}} \ge \{|||R_\lambda(T)|||_{\mathsf{H}}\}^{-1}\,\|y\|_{\mathsf{H}}\,,$$

showing that

$$\inf_{\|y\|_{\mathsf{H}}=1} \|(T - \lambda \mathrm{I})y\|_{\mathsf{H}} \ge \{|||R_\lambda(T)|||_{\mathsf{H}}\}^{-1}\,. \tag{22.A.5}$$

Theorem 22.A.8. *The resolvent set of every $T \in \mathsf{BL}(\mathsf{H})$ is an open set. The closed set* $\mathrm{Spec}(T|\mathsf{H})$ *is included in the ball $\{\lambda \in \mathbb{C} : |\lambda| < |||T|||_{\mathsf{H}}\}$.*

Proof. Assume $\lambda_0 \in \mathrm{Res}(T|\mathsf{H})$. Since $\lambda_0 I - T$ is invertible, it follows from Corollary 22.A.6 that there exists $\varepsilon > 0$ such that if $|\lambda - \lambda_0| < \varepsilon$, then $\lambda \mathrm{I} - T$ is invertible. Hence $\mathrm{Res}(T|\mathsf{H})$ is open. If $|\lambda| > |||T|||_{\mathsf{H}}$, then $\mathrm{I} - T/\lambda$ is invertible, since $|||T/\lambda|||_{\mathsf{H}} < 1$. Therefore, $\lambda\mathrm{I} - T = \lambda(\mathrm{I} - T/\lambda)$ is also invertible. □

As a consequence, the spectrum $\mathrm{Spec}(T|\mathsf{H})$ is a nonempty compact subset of $\mathbb{C}$ and $\mathrm{Spec}(T|\mathsf{H}) \subset \overline{\mathrm{B}(0, |||T|||_{\mathsf{H}})}$.

Definition 22.A.9 (Analytic operator-valued function) *An operator-valued function $\lambda \mapsto A(\lambda)$ that maps a subset of $\mathbb{C}$ into $\mathsf{BL}(\mathsf{H})$ is analytic at λ_0 if*

$$A(\lambda) = \sum_{k=0}^{\infty} A_k(\lambda - \lambda_0)^k ,$$

where each A_k belongs to $\mathsf{BL}(\mathsf{H})$ and the series converges for each λ in some neighborhood of λ_0.

Theorem 22.A.10. *The function $\lambda \mapsto R_T(\lambda) = (\lambda\mathrm{I} - T)^{-1}$ is analytic at each point in the open set* $\mathrm{Res}(T|\mathsf{H})$.

Proof. Suppose $\lambda_0 \in \mathrm{Res}(T|\mathsf{H})$. We have

$$\lambda\mathrm{I} - T = (\lambda_0\mathrm{I} - T) - (\lambda_0 - \lambda)\mathrm{I} = (\lambda_0\mathrm{I} - T)[\mathrm{I} - (\lambda_0 - \lambda)R_T(\lambda_0)] . \qquad (22.A.6)$$

Since $\mathrm{Res}(T|\mathsf{H})$ is open, we may choose $\varepsilon > 0$ such that $|\lambda - \lambda_0| < \varepsilon$ implies $\lambda \in \mathrm{Res}(T|\mathsf{H})$ and $|||(\lambda - \lambda_0)R_T(\lambda_0)|||_{\mathsf{H}} < 1$. In this case, it follows from (22.A.6) that

$$\begin{aligned} R_T(\lambda) &= [\mathrm{I} - (\lambda_0 - \lambda)R_T(\lambda_0)]^{-1}(\lambda_0\mathrm{I} - T)^{-1} \\ &= \sum_{k=0}^{\infty} (\lambda_0 - \lambda)^k R_T^{k+1}(\lambda_0) , \end{aligned}$$

which completes the proof. □

Definition 22.A.11 *The function $\lambda \mapsto= R_T(\lambda) = (\lambda\mathrm{I} - T)^{-1}$ is called the resolvent function of T, or simply the resolvent of T.*

A complex number λ is called an eigenvalue of $T \in \mathsf{BL}(\mathsf{H})$ if there exists $y \neq 0 \in \mathsf{H}$ such that $Ty = \lambda y$, or equivalently, $\mathrm{Ker}(\lambda\mathrm{I} - T) \neq \{0\}$. The vector y is called an eigenvector of T corresponding to the eigenvalue λ. Every linear operator on a finite-dimensional Euclidean space over $\mathbb{C}$ has at least one eigenvalue. However, an operator on an infinite-dimensional Banach space may have no eigenvalues.

Definition 22.A.12 (Point spectrum, spectral radius) *The point spectrum* $\operatorname{Spec}_p(T|\mathsf{H})$ *is a subset of the spectrum*

$$\operatorname{Spec}_p(T|\mathsf{H}) := \{\lambda \in \mathbb{C} : \operatorname{Ker}(\lambda \mathrm{I} - T) \neq \{0\}\}. \tag{22.A.7}$$

The elements λ *of* $\operatorname{Spec}_p(T|\mathsf{H})$ *are called the* eigenvalues *of* T*, and* $\operatorname{Ker}(\lambda \mathrm{I} - T)$ *is called the* eigenspace *associated with* λ*. The dimension of* $\operatorname{Ker}(\lambda \mathrm{I} - T)$ *is called the multiplicity of the eigenvalue* λ*.*

The spectral radius of T *is defined by*

$$\operatorname{Sp.Rad.}(T|\mathsf{H}) := \sup\{|\lambda| : \lambda \in \operatorname{Spec}(T|\mathsf{H})\} . \tag{22.A.8}$$

Proposition 22.A.13 *For all* $T \in \mathsf{BL}(\mathsf{H})$*, we have*

$$\operatorname{Sp.Rad.}(T|\mathsf{H}) = \lim_{n\to\infty} \{|||T^n|||_{\mathsf{H}}\}^{1/n} \leq |||T|||_{\mathsf{H}} . \tag{22.A.9}$$

Proof. Once one knows that $\lim_{n\to\infty}\{|||T^n|||_{\mathsf{H}}\}^{1/n}$ exists, this is essentially a version of the Cauchy formula, that the radius of convergence of $(\mathrm{I} - \lambda T)^{-1}$ is $\limsup_{n\to\infty}\{|||T^n|||_{\mathsf{H}}\}^{1/n}$.

Since for every $n \in \mathbb{N}$, $|||T^n|||_{\mathsf{H}} \leq \{|||T|||_{\mathsf{H}}\}^n$, the inequality in (22.A.9) holds. We first show that $\lim_{n\to\infty}\{|||T^n|||_{\mathsf{H}}\}^{1/n}$ exists. Indeed, defining $\alpha_n := \log|||T^n|||_{\mathsf{H}}$, we obtain that $\{\alpha_n,\ n \in \mathbb{N}^*\}$ is subadditive: $\alpha_{n+m} \leq \alpha_n + \alpha_m$ for all $(m,n) \in \mathbb{N}^* \times \mathbb{N}^*$. Then, by Fekete's subadditive lemma (see, for example, Exercise 5.12), $\lim_{n\to\infty}\alpha_n/n$ exists and is equal to $\inf_{n\in\mathbb{N}}\alpha_n/n$. This implies that $\{|||T^n|||_{\mathsf{H}}\}^{1/n}$ converges as n goes to infinity.

Note that $R = \{\lim_{n\to\infty}\{|||T^n|||_{\mathsf{H}}\}^{1/n}\}^{-1}$ is the radius of convergence of the series $\sum_{n=0}^{\infty}\lambda^n T^n$. For all $|\lambda| < R$, the series $\sum_{n=0}^{\infty}\lambda^n T^n$ is convergent, and the operator $\mathrm{I} - \lambda T$ is therefore invertible. Writing $\mathrm{I} - \lambda T = \lambda(\lambda^{-1}\mathrm{I} - T)$, we then obtain that $\lambda^{-1} \in \operatorname{Res}(T|\mathsf{H})$ for all $|\lambda^{-1}| > R^{-1}$, and thus $\operatorname{Sp.Rad.}(T|\mathsf{H}) \leq \lim_{n\to\infty}\{|||T^n|||_{\mathsf{H}}\}^{1/n}$.

On the other hand, if $r > \operatorname{Sp.Rad.}(T|\mathsf{H})$, then the function $\mu \mapsto (\mathrm{I} - \mu T)^{-1}$ is analytic on the disc, $\left\{\mu : |\mu| \leq r^{-1}\right\}$. Thus by a Cauchy estimate, $|||T^n|||_{\mathsf{H}} \leq C(r^{-n})^{-1}$ and $\lim_{n\to\infty}\{|||T^n|||_{\mathsf{H}}\}^{1/n} \leq r$. Since this inequality is satisfied for any $r > \operatorname{Sp.Rad.}(T|\mathsf{H})$, we get $\operatorname{Sp.Rad.}(T|\mathsf{H}) \geq \lim_{n\to\infty}\{|||T^n|||_{\mathsf{H}}\}^{1/n}$. □

Assume that $(\mathsf{H}, \langle\cdot,\cdot\rangle_{\mathsf{H}})$ is a Hilbert space. For each $y \in \mathsf{H}$, the functional f_y defined on H by $f_y(x) = \langle x, y\rangle_{\mathsf{H}}$ is linear. Moreover,

$$|f_y(x)| = |\langle x, y\rangle_{\mathsf{H}}| \leq \|x\|_{\mathsf{H}}\|y\|_{\mathsf{H}} .$$

Thus f_y is bounded and $|||f_y|||_{\mathsf{H}} \leq \|y\|_{\mathsf{H}}$. Since

$$|||f_y|||_{\mathsf{H}} \, \|y\|_{\mathsf{H}} \geq |f_y(y)| = |\langle y, y\rangle_{\mathsf{H}}| = \|y\|_{\mathsf{H}}^2,$$

we have $|||f_y||| \geq \|y\|$. The Riesz representation theorem shows that every bounded linear functional is an f_y.

Theorem 22.A.14 (Riesz representation theorem). *Let* $(\mathsf{H}, \langle\cdot,\cdot\rangle_{\mathsf{H}})$ *be a Hilbert space. For all* $f \in \mathsf{H}^*$*, there exists a unique* $y \in \mathsf{H}$ *such that for all* $x \in H$*,* $f(x) = \langle x, y\rangle_{\mathsf{H}}$*. Therefore,* $f = f_y$ *and* $|||f_y|||_{\mathsf{H}} = \|y\|_{\mathsf{H}}$*.*

Proof. See (Gohberg and Goldberg 1981, Theorem 5.2). □

Let $(\mathsf{H}, \langle\cdot,\cdot\rangle_{\mathsf{H}})$ and $(\mathsf{G}, \langle\cdot,\cdot\rangle_{\mathsf{G}})$ be two Hilbert spaces and $T \in \mathsf{BL}(\mathsf{H}, \mathsf{G})$. For each $y \in \mathsf{G}$, the functional $x \mapsto f_y(x) = \langle Tx, y\rangle_{\mathsf{G}}$ is a bounded linear functional on H. Hence the Riesz representation theorem guarantees the existence of a unique $y^* \in \mathsf{H}$ such that for all $x \in \mathsf{H}$, $\langle Tx, y\rangle_{\mathsf{H}} = f_y(x) = \langle x, y^*\rangle_{\mathsf{H}}$. This gives rise to an operator $T^* \in \mathsf{BL}(\mathsf{G}, \mathsf{H})$ defined by $T^*y = y^*$ satisfying

$$\langle Tx, y\rangle_{\mathsf{G}} = \langle x, y^*\rangle_{\mathsf{H}} = \langle x, T^*y\rangle_{\mathsf{H}} \, , \quad \text{for all } x \in \mathsf{H} \, . \tag{22.A.10}$$

The operator T^* is called the *adjoint* of T.

Now we will consider the case in which H, G are Banach spaces and $T \in \mathsf{BL}(\mathsf{H}, \mathsf{G})$. For $\mu \in \mathsf{G}^*$, $\nu \in \mathsf{H}^*$, we use the notation $\mu(x) = \langle \mu, x\rangle_{\mathsf{G}}$, $x \in \mathsf{G}$, $\nu(y) = \langle \nu, y\rangle_{\mathsf{H}}$, $y \in \mathsf{H}$. There exists a unique adjoint $T^* \in \mathsf{BL}(\mathsf{G}^*, \mathsf{H}^*)$ that is defined by an equation that generalizes (22.A.10) to the setting of Banach spaces: for all $\mu \in \mathsf{G}^*$ and $x \in \mathsf{H}$,

$$T^*\mu(x) := \langle T^*\mu, x\rangle_{\mathsf{H}} = \langle \mu, Tx\rangle_{\mathsf{G}} \, .$$

Note, however, that the adjoint is a map $T^* : \mathsf{G}^* \to \mathsf{H}^*$, whereas $T : \mathsf{H} \to \mathsf{G}$. In particular, in contrast to the Hilbert space case, we cannot consider compositions of T with T^*. We have that

$$|||T|||_{\mathsf{H}\to\mathsf{G}} = |||T^*|||_{\mathsf{G}^*\to\mathsf{H}^*} \, . \tag{22.A.11}$$

Theorem 22.A.15. *If T and S are in* $\mathsf{BL}(\mathsf{H}, \mathsf{G})$*, then*

(i) $(T+S)^* = T^* + S^*$, $(\alpha T)^* = \bar{\alpha} T^*$ *for all* $\alpha \in \mathbb{C}$*;*
(ii) If H *and* G *are Hilbert spaces, then* $(TS)^* = S^*T^*$*.*

Proof. See (Gohberg and Goldberg 1981, Theorem 11.3). □

Definition 22.A.16 (Self-adjoint) *Let* $(\mathsf{H}, \langle , \rangle_{\mathsf{H}})$ *be a Hilbert space. An operator* $T \in \mathsf{BL}(\mathsf{H})$ *is said to be self-adjoint if* $T = T^*$, *i.e., for all* $x, y \in \mathsf{H}$,

$$\langle Tx, y\rangle_{\mathsf{H}} = \langle x, Ty\rangle_{\mathsf{H}} \ .$$

Theorem 22.A.17. *Let* T *be a self-adjoint operator on the Hilbert space* $(\mathsf{H}, \langle\cdot,\cdot\rangle_{\mathsf{H}})$. *Then* $|||T|||_{\mathsf{H}} = \sup_{\|x\|_{\mathsf{H}}\leq 1} |\langle Tx, x\rangle_{\mathsf{H}}|$.

Proof. See (Gohberg and Goldberg 1981, Theorem 4.1, Chapter III). □

Corollary 22.A.18 *Let* $(\mathsf{H}, \langle\cdot,\cdot\rangle_{\mathsf{H}})$ *be a Hilbert space. For all* $T \in \mathsf{BL}(\mathsf{H})$,

$$|||T^*T|||_{\mathsf{H}} = |||T\,|||^2_{\mathsf{H}} \ .$$

Proof. Since T^*T is self-adjoint,

$$|||T^*T|||_{\mathsf{H}} = \sup_{\|x\|\leq 1} |\langle T^*Tx,\ x\rangle_{\mathsf{H}}| = \sup_{\|x\|\leq 1} \|Tx\|^2_{\mathsf{H}} = |||T\,|||^2_{\mathsf{H}} \ .$$

□

Theorem 22.A.19. *Let* T *be a self-adjoint operator on the Hilbert space* $(\mathsf{H}, \langle\cdot,\cdot\rangle_{\mathsf{H}})$. *Set*

$$m = \inf_{\|x\|_{\mathsf{H}}=1} \langle Tx, x\rangle_{\mathsf{H}} \quad \textit{and} \quad M = \sup_{\|x\|_{\mathsf{H}}=1} \langle Tx, x\rangle_{\mathsf{H}} \ .$$

(i) $\mathrm{Spec}(T|\mathsf{H}) \subseteq [m, M]$;
(ii) $m \in \mathrm{Spec}(T|\mathsf{H})$ *and* $M \in \mathrm{Spec}(T|\mathsf{H})$.

Proof. (i) Suppose $\lambda \notin [m, M]$. Denote by d the distance of λ to the segment $[m, M]$. Let $x \in \mathsf{H}$ be any unit vector and write $\alpha = \langle Tx, x\rangle_{\mathsf{H}} \in [m, M]$. Then $\langle(\alpha\mathrm{I} - T)x, x\rangle_{\mathsf{H}} = \langle x, (\alpha\mathrm{I} - T)x\rangle_{\mathsf{H}} = 0$ and

$$\begin{aligned}\|(\lambda\mathrm{I} - T)x\|^2_{\mathsf{H}} &= \|[\lambda\mathrm{I} - \alpha\mathrm{I} + (\alpha\mathrm{I} - T)]x\|^2_{\mathsf{H}} \\ &= \langle[\lambda\mathrm{I} - \alpha\mathrm{I} + (\alpha\mathrm{I} - T)]x, [\lambda\mathrm{I} - \alpha\mathrm{I} + (\alpha\mathrm{I} - T)]x\rangle_{\mathsf{H}} \\ &= |\lambda - \alpha|^2\|x\|^2 + \|(\alpha\mathrm{I} - T)x\|^2_{\mathsf{H}} \geq |\lambda - \alpha|^2 \geq d^2.\end{aligned}$$

It follows that for all $x \in \mathsf{H}$, $\|(\lambda\mathrm{I} - T)x\|_{\mathsf{H}} \geq d\|x\|_{\mathsf{H}}$ (apply the above for $x/\|x\|_{\mathsf{H}}$). Hence $\lambda\mathrm{I} - T$ is injective, and by Lemma 22.A.4, it has closed range. Further, if $0 \neq$

$z \perp \mathrm{Ran}(\lambda\mathrm{I}-T)$, then $0 = \langle(\lambda\mathrm{I}-T)x, z\rangle = \langle x, (\bar{\lambda}\mathrm{I}-T)z\rangle$ for all $x \in \mathsf{H}$, and so $(\bar{\lambda}\mathrm{I}-T)z = 0$. But this is impossible, since from above, noting that $d = \mathsf{d}(\lambda, [m,M]) = \mathsf{d}(\bar{\lambda}, [m,M])$, we have $\|(\bar{\lambda}\mathrm{I}-T)z\|_{\mathsf{H}} \geq d\|z\|_{\mathsf{H}}$. Therefore, $\mathrm{Ran}(\lambda\mathrm{I}-T) = \mathsf{H}$ (being both dense and closed).

Therefore, for all $y \in \mathsf{H}$, there is a unique $x \in \mathsf{H}$ such that $y = (\lambda\mathrm{I}-T)x$. Define $(\lambda\mathrm{I}-T)^{-1}y = x$. Then $\|y\|_{\mathsf{H}} \geq d\|x\|_{\mathsf{H}}$, so

$$\|(\lambda\mathrm{I}-T)^{-1}y\|_{\mathsf{H}} = \|x\|_{\mathsf{H}} \leq \frac{1}{d}\|y\|_{\mathsf{H}},$$

showing that $(\lambda\mathrm{I}-T)^{-1}$ is a bounded operator. Thus $\lambda \notin \mathrm{Spec}(T|\mathsf{H})$, proving (i).

(ii) From Theorem 22.A.17, $|||T|||$ is either M or $-m$. We consider only the case $|||T||| = M = \sup_{\|x\|=1}\langle Tx, x\rangle$; if $|||T||| = -m$, it suffices to apply the proof to $-T$. There exists a sequence $\{x_n,\ n \in \mathbb{N}\}$ of unit vectors such that $\lim_{n\to\infty}\langle Tx_n, x_n\rangle = M$. Then

$$\|(T-M\mathrm{I})x_n\|^2 = \|Tx_n\|^2 + M^2 - 2M\langle Tx_n, x_n\rangle \leq 2M^2 - 2M\langle Tx_n, x_n\rangle \to 0\,.$$

Hence $T - M\mathrm{I}$ has no inverse in $\mathsf{BL}(\mathsf{H})$ (since if X were such an operator, then $1 = \|x_n\| = \|X(T-M\mathrm{I})x_n\| \leq |||X|||\,\|X(T-M\mathrm{I})x_n\| \to 0$), and so $M \in \mathrm{Spec}(T|\mathsf{H})$. For m, note that by Theorem 22.A.17,

$$\sup_{\|x\|=1}\langle(M\mathrm{I}-T)x, x\rangle = M - m = \|M\mathrm{I}-T\|,$$

since $\inf_{\|x\|=1}\langle(M\mathrm{I}-T)x, x\rangle = 0$. Applying the result just proved to the operator $M\mathrm{I}-T$ shows that $M-m \in \mathrm{Spec}(M\mathrm{I}-T|\mathsf{H})$, that is, $(M-m)\mathrm{I}-(M\mathrm{I}-T) = T-m\mathrm{I}$ has no inverse. Hence $m \in \mathrm{Spec}(T|\mathsf{H})$.

□

A self-adjoint operator on the Hilbert space H is positive if for all $f \in \mathsf{H}$, $\langle f, Pf\rangle \geq 0$.

Lemma 22.A.20 *Let H_1 and H_2 be Hilbert spaces and $M : \mathsf{H}_1 \to \mathsf{H}_2$ a bounded linear operator. Let M^* be the adjoint operator of M, and let $T : \mathsf{H}_2 \to \mathsf{H}_2$ be a bounded linear and positive operator. Then $MTM^* : \mathsf{H}_1 \to \mathsf{H}_1$ is also positive.*

Proof. We denote the inner product of H_i by $\langle\cdot,\cdot\rangle_i$ for $i = 1, 2$. By the definition of the adjoint operator and positivity of T,

$$\langle MTM^* f, f\rangle_1 = \langle TM^* f, M^* f\rangle_2 \geq 0.$$

□

22.B Spectral Measure

Denote by $\mathbb{C}[X]$ (resp. $\mathbb{R}[X]$) the set of polynomials with complex (resp. real) coefficients. Let H be a Hilbert space on $\mathbb{C}$ equipped with a sesquilinear product $\langle\cdot,\cdot\rangle$. If $p \in \mathbb{C}[X]$ or $\mathbb{R}[X]$, write $p(T)$ for the operator on H defined by $p(T) = \sum_{k=0}^m a_k T^m$, where the coefficients $\{a_i\,,\ i = 0,\dots,m\}$ are such that $p(X) = \sum_{k=0}^m a_k X^m$. We now review some basic properties of the self-adjoint operator T.

Lemma 22.B.1 *Let T, T_1, and T_2 be bounded linear operators on H and assume that $T = T_1T_2 = T_2T_1$. Then T is invertible if and only if T_1 and T_2 are invertible.*

Proof. If T_1 and T_2 are invertible, then T is also invertible. Assume now that T is invertible. Then we have $T_1(T_2T^{-1}) = \mathrm{I}$ and $T^{-1}T_1T_2 = (T^{-1}T_2)T_1 = \mathrm{I}$, but $T^{-1}T_2 = (T^{-1}T_2)T_1(T_2T^{-1}) = T_2T^{-1}$, so that T_2T^{-1} is an inverse for T_1. In the same way, T_1T^{-1} is an inverse for T_2. □

Proposition 22.B.2 *Let T be a self-adjoint operator on the Hilbert space $(\mathsf{H},\langle\cdot,\cdot\rangle)$.*

(i) $\mathrm{Sp.Rad.}(T|\mathsf{H}) = |||T|||_{\mathsf{H}}$, *where* $\mathrm{Sp.Rad.}(T|\mathsf{H})$ *is defined in* (22.A.8).

(ii) For all $p \in \mathbb{C}[X]$, $p(T)$ is a bounded linear operator on H. For all $p_1, p_2 \in \mathbb{C}[X]$, $p_1(T)$ and $p_2(T)$ commute. If $p \in \mathbb{R}[X]$, then $p(T)$ is self-adjoint. Moreover, for all $p \in \mathbb{C}[X]$,

$$p(\mathrm{Spec}(T|\mathsf{H})) = \mathrm{Spec}(p(T)|\mathsf{H})\,.$$

(iii) For all $p \in \mathbb{R}[X]$, $|||p(T)||| = \sup\{|p(\lambda)|\,:\,\lambda \in \mathrm{Spec}(T|\mathsf{H})\}$.

Proof. (i) By Proposition 22.A.13, we know that

$$\mathrm{Sp.Rad.}(T|\mathsf{H}) = \lim_{n\to\infty}\{|||T^n|||_{\mathsf{H}}\}^{1/n} \le |||T|||\ .$$

Since T is self-adjoint, we have for all $x \in \mathsf{H}$ such that $\|x\| = 1$,

$$\|Tx\|^2 = \langle Tx, Tx\rangle = \left\langle T^2x, x\right\rangle \le \|T^2x\|\,\|x\| \le |||T^2|||\,,$$

so that $|||T|||^2 \le |||T^2|||$. Finally, $|||T^2||| = |||T|||^2$. By a straightforward induction, we then obtain that for all $n \in \mathbb{N}$, $|||T^n||| = |||T|||^n$, and thus

$$\mathrm{Sp.Rad.}(T|\mathsf{H}) = \lim_{n\to\infty}\{|||T^n|||_{\mathsf{H}}\}^{1/n} = |||T|||_{\mathsf{H}}\ .$$

(ii) The first assertions are obvious. We have only to prove that $p(\mathrm{Spec}(T|\mathsf{H})) = \mathrm{Spec}(p(T)|\mathsf{H})$ for all $p \in \mathbb{C}[X]$. Write $\lambda - p(z) = a_0\prod_{k=1}^m(z_k - z)$, where $a_0 \in \mathbb{C}$ and $z_1,\dots,z_m$ are the (not necessarily distinct) roots of $\lambda - p(z)$. Then

$$\lambda \mathrm{I} - p(T) = a_0 \prod_{k=1}^{m} (z_k \mathrm{I} - T) .$$

According to Lemma 22.B.1, $\lambda \mathrm{I} - p(T)$ is invertible if and only if for all $k \in \{1,\dots,m\}$, $z_k \mathrm{I} - T$ is invertible. Therefore, $\lambda \in \mathrm{Spec}(p(T)|\mathsf{H})$ if and only if for some $k \in \{1,\dots,m\}$, $z_k \in \mathrm{Spec}(T|\mathsf{H})$. Finally, $\lambda \in \mathrm{Spec}(p(T)|\mathsf{H})$ if and only if there exists $\mu \in \mathrm{Spec}(T|\mathsf{H})$ such that $\lambda = p(\mu)$. Hence $\mathrm{Spec}(p(T)|\mathsf{H}) = p(\mathrm{Spec}(T|\mathsf{H}))$.

(iii) Noting that $p(T)$ is self-adjoint and using (i) and (ii), we obtain

$$\begin{aligned}|||p(T)|||_{\mathsf{H}} = \mathrm{Sp.Rad.}(p(T)|\mathsf{H}) &= \sup_{\lambda \in \mathrm{Spec}(T|\mathsf{H})} |\lambda| \\ &= \sup_{\lambda \in p(\mathrm{Spec}(T|\mathsf{H}))} |\lambda| = \sup_{\mu \in \mathrm{Spec}(T|\mathsf{H})} |p(\mu)| .\end{aligned}$$

□

Theorem 22.B.3 (Spectral measure). *Let T be a self-adjoint operator on H and $f \in \mathsf{H}$. There exists a unique measure ν_f on $\mathbb{R}$, supported by $\mathrm{Spec}(T|\mathsf{H})$, such that for all $n \geq 0$,*

$$\int x^n \nu_f(\mathrm{d}x) = \langle T^n f, f \rangle . \tag{22.B.1}$$

Furthermore, $\nu_f(\mathbb{R}) = \nu_f(\mathrm{Spec}(T|\mathsf{H})) = \|f\|^2$.

Proof. We consider only $\mathbb{R}$-valued functions. If $p,q \in \mathbb{R}[X]$ with $p(\lambda) = q(\lambda)$ for all $\lambda \in \mathrm{Spec}(T|\mathsf{H})$, then by Proposition 22.B.2 (iii),

$$|||p(T) - q(T)||| = \sup_{\lambda \in \mathrm{Spec}(T|\mathsf{H})} |p(\lambda) - q(\lambda)| = 0 .$$

Thus $p(T) = q(T)$. Set $\mathscr{I}(p) = \langle p(T)f, f \rangle$; $\mathscr{I}$ is a well-defined linear form on

$$\mathscr{D} = \left\{ \varphi \in \mathrm{C}_b(\mathrm{Spec}(\mathsf{T}|\mathsf{H})) \,:\, \varphi = p|_{\mathrm{Spec}(T|\mathsf{H})}, p \in \mathbb{R}[X], \right\}$$

and it is continuous, since by Proposition 22.B.2 (iii),

$$\begin{aligned}|\mathscr{I}(p)| &= |\langle p(T)f, f \rangle| \leq \|p(T)f\| . \|f\| \\ &\leq |||p(T)|||\, \|f\|^2 = \|f\|^2 . \sup_{\lambda \in \mathrm{Spec}(T|\mathsf{H})} |p(\lambda)| .\end{aligned}$$

Since by the Stone–Weierstrass theorem, $\mathscr{D}$ is dense in $\mathrm{C}_b(\mathrm{Spec}(\mathsf{T}|\mathsf{H}))$, $\mathscr{I}$ extends to a continuous linear form on $\mathrm{C}_b(\mathrm{Spec}(\mathsf{T}|\mathsf{H}))$. Moreover, let φ be a nonnegative function in $\mathrm{C}_b(\mathrm{Spec}(\mathsf{T}|\mathsf{H}))$. Using again the Stone–Weierstrass theorem, there exists a sequence of polynomials $\{p_n,\ n \in \mathbb{N}\}$ such that $\lim_{n\to\infty} \sup_{\lambda \in \mathrm{Spec}(T|\mathsf{H})} |p_n^2(\lambda) - \varphi(\lambda)| = 0$. Since

$$\mathscr{I}(p_n^2) = \left\langle p_n^2(T)f, f \right\rangle = \langle p_n(T)f, p_n(T)f \rangle \geq 0,$$

we have $\mathscr{I}(\varphi) = \lim_{n\to\infty} \mathscr{I}(p_n^2) \geq 0$. Therefore, $\mathscr{I}$ is a nonnegative continuous linear form on $\mathrm{C}_b(\mathrm{Spec}(\mathsf{T}|\mathsf{H}))$, and the Riesz theorem shows that there exists a measure ν_f on $\mathrm{Spec}(T|\mathsf{H})$ such that $\mathscr{I}(f) = \int f \mathrm{d}\nu_f$. We can then extend this measure to $\mathbb{R}$, setting $\nu_f(A) = \nu_f(A \cap \mathrm{Spec}(T|\mathsf{H}))$ for all $A \in \mathscr{B}(\mathbb{R})$. Finally, taking $n = 0$ in (22.B.1), we have $\nu_f(\mathbb{R}) = \|f\|^2$. The uniqueness of the spectral measure is obvious due to the density of the polynomials. □

Chapter 23
Concentration Inequalities

Let $(\mathsf{X}, \mathscr{X})$ be a measurable space, $(\Omega, \mathscr{F}, \mathbb{P})$ a probability space, and $\{X_k,\ k \in \mathbb{N}\}$ an X-valued stochastic process. The concentration of measure phenomenon occurs when a function $f(X_0, \ldots, X_n)$ takes values that are close to the mean value of $\mathbb{E}[f(X_0, \ldots, X_n)]$ (provided such a quantity exists). This phenomenon has been extensively studied in the case that $\{X_k,\ k \in \mathbb{N}\}$ is a sequence of independent random variables. We consider in this chapter the case of a homogeneous Markov chain.

In Section 23.1, we introduce sub-Gaussian concentration inequalities for functions of independent random variables. We will consider functions of bounded difference, which means that oscillations of such functions with respect to each variable are uniformly bounded. These functions include additive functions and suprema of additive functions and are sufficient for most statistical applications. We will state and prove McDiarmid's inequality for independent random variables in order to introduce in this simple context the main idea of the proof, which is a martingale decomposition based on a sequential conditioning.

The same method of proof, with increasingly involved technical details, will be applied in Sections 23.2 and 23.3 to obtain sub-Gaussian concentration inequalities for uniformly ergodic and V-geometrically ergodic Markov chains.

In Section 23.4, we will consider possibly nonirreducible kernels that are contracting in the Wasserstein distance. In that case, functions of bounded difference must be replaced by separately Lipschitz functions.

Throughout this chapter, we will use the following shorthand notation for tuples.

> For $k \le n$ and $x_k, \ldots, x_n \in \mathsf{X}$, we write x_k^n for $(x_k, \ldots, x_n)$.

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1_23

23.1 Concentration Inequality for Independent Random Variables

We first define the class of functions of interest.

Definition 23.1.1 (Functions of bounded difference) *A measurable function $f : \mathsf{X}^n \to \mathbb{R}$ is said to have the bounded difference property if there exist nonnegative constants $(\gamma_0, \dots, \gamma_{n-1})$ such that for all $x_0^{n-1} \in \mathsf{X}^n$ and $y_0^{n-1} \in \mathsf{X}^n$,*

$$|f(x_0^{n-1}) - f(y_0^{n-1})| \le \sum_{i=0}^{n-1} \gamma_i \mathbb{1}_{\{x_i \ne y_i\}} . \tag{23.1.1}$$

The class of all functions f that satisfy (23.1.1) is denoted by $\mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$.

In words, if $f \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$, if we change the ith component x_i to y_i while keeping all other x_j fixed, the value of f changes by at most γ_i.

Conversely, let f be a function such that for all $i \in \{0, \dots, n-1\}$, $x_0^{n-1} \in \mathsf{X}^n$, $y_i \in \mathsf{X}$,

$$|f(x_0^{n-1}) - f(x_0^{i-1}, y_i, x_{i+1}^{n-1})| \le \gamma_i \mathbb{1}_{\{x_i \ne y_i\}} , \tag{23.1.2}$$

with the convention $x_p^q = \emptyset$ if $p > q$. Since for all $x_0^{n-1} \in \mathsf{X}^n$ and $y_0^{n-1} \in \mathsf{X}^n$,

$$f(x_0^{n-1}) - f(y_0^{n-1}) = \sum_{i=0}^{n-1} \{f(x_0^i, y_{i+1}^{n-1}) - f(x_0^{i-1}, y_i^{n-1})\},$$

we get that if (23.1.2) is satisfied, then $\mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$.

Example 23.1.2. Let $f_0, \dots, f_{n-1}$ be n functions with bounded oscillations, $\gamma_i = \mathrm{osc}\,(f_i) < \infty$, and let f be the sum $f(x_0^{n-1}) = \sum_{i=0}^{n-1} f_i(x_i)$, $x_0^{n-1} \in \mathsf{X}^n$. The function f belongs to $\mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$. ◀

Example 23.1.3. Let $(\mathsf{X}, \mathscr{X})$ be a measurable space and $X_0, \dots, X_{n-1}$ n independent X-valued random variables with common distribution μ. For each $x_0^{n-1} \in \mathsf{X}^n$, let $\hat{\mu}_{x_0^{n-1}}$ be the empirical measure defined by

$$\hat{\mu}_{x_0^{n-1}} = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{x_i} . \tag{23.1.3}$$

Let $\mathscr{G}$ be a collection of functions defined on X and assume that $\sup_{g \in \mathscr{G}} |g|_\infty \le M$. Consider the function f defined on X by

$$f(x_0^{n-1}) = \sup_{g \in \mathscr{G}} |\hat{\mu}_{x_0^{n-1}}(g) - \mu(g)| . \tag{23.1.4}$$

By changing only one coordinate x_j, the value of $f(x_0^{n-1})$ can change by at most M/n. Indeed, for $x_0^{n-1} \in \mathsf{X}^n$, $i \in \{0,\dots,n-1\}$, and $x_i' \in \mathsf{X}$, let $x_0^{n-1,(i)}$ denote x_0^{n-1} with x_i replaced by x_i':

$$x_0^{n-1,(i)} = (x_0^{i-1}, x_i', x_{i+1}^{n-1}) .$$

Then

$$\begin{aligned}
f(x_0^{n-1}) - f(x_0^{n-1,(i)}) &= \sup_{g\in\mathscr{G}} |\hat{\mu}_{x_0^{n-1}}(g) - \mu(g)| - \sup_{g'\in\mathscr{G}} |\hat{\mu}_{x_0^{n-1,(i)}}(g') - \mu(g')| \\
&= \sup_{g\in\mathscr{G}} \inf_{g'\in\mathscr{G}} \{|\hat{\mu}_{x_0^{n-1}}(g) - \mu(g)| - |\hat{\mu}_{x_0^{n-1,(i)}}(g') - \mu(g')|\} \\
&\le \sup_{g\in\mathscr{G}} \{|\hat{\mu}_{x_0^{n-1}}(g) - \mu(g)| - |\hat{\mu}_{x_0^{n-1,(i)}}(g) - \mu(g)|\} \\
&\le \sup_{g\in\mathscr{G}} |\hat{\mu}_{x_0^{n-1}}(g) - \hat{\mu}_{x_0^{n-1,(i)}}(g)| \\
&= \frac{1}{n} \sup_{g\in\mathscr{G}} |g(x_i) - g(x_i')| \le 2M/n .
\end{aligned}$$

Swapping x_0^{n-1} and $x_0^{n-1,(i)}$, we obtain that $|f(x_0^{n-1}) - f(x_0^{n-1,(i)})| \le 2M/n$. Thus $f \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$ with $\gamma_i = 2M/n$ for all $i \in \{0,\dots,n-1\}$. ◀

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $\{X_k,\ k \in \mathbb{N}\}$ an X-valued stochastic process. The process $\{X_k,\ k \in \mathbb{N}\}$ satisfies a sub-Gaussian concentration inequality if there exists a constant κ such that for all $n \in \mathbb{N}^*$, $\gamma_0^{n-1} \in \mathbb{R}_+^n$, $f \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$, and $s \ge 0$, one has

$$\mathbb{E}\left[\exp\left(s\left\{f(X_0^{n-1}) - \mathbb{E}\left[f(X_0^{n-1})\right]\right\}\right)\right] \le \mathrm{e}^{s^2\kappa\sum_{j=0}^{n-1}\gamma_j^2} . \tag{23.1.5}$$

This inequality might be used to bound the probability $\mathbb{P}(Y \ge t)$, where $Y = f(X_0^{n-1}) - \mathbb{E}\left[f(X_0^{n-1})\right]$ for all $t > 0$, using Chernoff's technique. Observe that for all $s > 0$, we have

$$\mathbb{P}(Y \ge t) = \mathbb{P}(\mathrm{e}^{sY} \ge \mathrm{e}^{st}) \le \mathrm{e}^{-st}\mathbb{E}\left[\mathrm{e}^{sY}\right], \tag{23.1.6}$$

where the first step results from monotonicity of the function $\psi(x) = \mathrm{e}^{sx}$, and the second step is from Markov's inequality. The Chernoff bound is obtained by choosing $s > 0$ that makes the right-hand side of (23.1.6) suitably small. If (23.1.6) holds for all $s > 0$, the optimal bound is

$$\mathbb{P}(Y \ge t) \le \inf_{s>0} \mathrm{e}^{-st}\mathbb{E}\left[\mathrm{e}^{sY}\right] .$$

Using (23.1.5) for $\mathbb{E}\left[\mathrm{e}^{sY}\right]$ and optimizing with respect to s shows that

$$\mathbb{P}\left[f(X_0^{n-1}) - \mathbb{E}\left[f(X_0^{n-1})\right] \ge t\right] \le \exp\left(-\frac{t^2}{4\kappa\sum_{j=0}^{n-1}\gamma_j^2}\right) . \tag{23.1.7}$$

The same inequality is satisfied if we replace f by $-f$, so that the bound (23.1.5) also implies that

$$\mathbb{P}\left[\left|f(X_0^{n-1}) - \mathbb{E}\left[f(X_0^{n-1})\right]\right| \geq t\right] \leq 2\exp\left(-\frac{t^2}{4\kappa\sum_{j=0}^{n-1}\gamma_j^2}\right) . \tag{23.1.8}$$

In some cases, (23.1.5) is satisfied only for $s \in [0, s_*)$, where $s_* < \infty$, in which case the optimization leading to (23.1.7) should be adapted. If (23.1.5) is not satisfied for any $s \geq 0$, one might consider a polynomial concentration inequality. We will not deal with that case in this chapter.

The method for obtaining exponential inequalities in this chapter is based on a martingale decomposition of the difference $f(X_0^{n-1}) - \mathbb{E}\left[f(X_0^{n-1})\right]$. The argument is easily described when $\{X_k,\, k \in \mathbb{N}\}$ is a sequence of independent random variables. For each $k \in \mathbb{N}$, we denote by μ_k the law of X_k. Without loss of generality, we assume that $\mathbb{E}\left[f(X_0^{n-1})\right] = 0$. Define $g_{n-1}(x_0^{n-1}) = f(x_0^{n-1})$, and for $\ell \in \{0, \ldots, n-2\}$ and $x_0^\ell \in \mathsf{X}^{\ell+1}$, set

$$g_\ell(x_0^\ell) = \int f(x_0^\ell, x_{\ell+1}^{n-1}) \prod_{k=\ell+1}^{n-1} \mu_k(\mathrm{d}x_k) . \tag{23.1.9}$$

With these definitions, we get

$$g_{n-1}(x_0^{n-1}) = \sum_{\ell=1}^{n-1} \left\{g_\ell(x_0^\ell) - g_{\ell-1}(x_0^{\ell-1})\right\} + g_0(x_0) . \tag{23.1.10}$$

For all $\ell \in \{1, \ldots, n-1\}$ and all $x_0^{\ell-1} \in \mathsf{X}^\ell$, we have

$$g_{\ell-1}(x_0^{\ell-1}) = \int g_\ell(x_0^{\ell-1}, x_\ell)\mu_\ell(\mathrm{d}x_\ell) .$$

Thus we obtain that $g_{\ell-1}(X_0^{\ell-1}) = \mathbb{E}\left[g_\ell(X_0^\ell) \,\middle|\, \mathscr{F}_{\ell-1}^X\right] \quad \mathbb{P}-\text{a.s.}$ for $\ell \geq 1$ where $\mathscr{F}_\ell^X = \sigma(X_0, \ldots, X_\ell)$. Setting by convention $\mathscr{F}_{-1}^X = \{\emptyset, \mathsf{X}^{\mathbb{N}}\}$, we have

$$\mathbb{E}\left[g_0(X_0) \,|\, \mathscr{F}_{-1}^X\right] = \mathbb{E}\left[g_0(X_0)\right] = \mathbb{E}\left[f(X_0^{n-1})\right] = 0 .$$

Hence the sequence $\{(g_\ell(X_0^\ell), \mathscr{F}_\ell^X),\, \ell = 0, \ldots, n-1\}$ is a zero-mean $\mathbb{P}$-martingale. Furthermore, for each $\ell \in \{1, \ldots, n-1\}$, $x_0^\ell \in \mathsf{X}^{\ell+1}$, and $x \in \mathsf{X}$,

$$\inf_{x \in \mathsf{X}} g_\ell(x_0^{\ell-1}, x) \leq g_\ell(x_0^\ell) \leq \inf_{x \in \mathsf{X}} g_\ell(x_0^{\ell-1}, x) + \gamma_\ell , \tag{23.1.11}$$

$$\inf_{u \in \mathsf{X}} g_0(u) \leq g_0(x) \leq \inf_{u \in \mathsf{X}} g_0(u) + \gamma_0 . \tag{23.1.12}$$

Indeed, for $\ell \geq 1$, the inequality $\inf_{x \in \mathsf{X}} g_\ell(x_0^{\ell-1}, x) \leq g_\ell(x_0^\ell)$ obviously holds. Moreover, by (23.1.9), we have for all $x^* \in \mathsf{X}$,

$$g_\ell(x_0^\ell) \le \int f(x_0^{\ell-1}, x^*, x_{\ell+1}^{n-1}) \prod_{k=\ell+1}^{n-1} \mu_k(\mathrm{d}x_k) + \gamma_k = g_\ell(x_0^{\ell-1}, x^*) + \gamma_k ,$$

which proves (23.1.11), since x^* is arbitrary. The proof of (23.1.12) is along the same lines.

The next ingredient is the following lemma (which is often used to establish Hoeffding's inequality).

Lemma 23.1.4 *Let* $(\mathsf{X}, \mathscr{X})$ *be a measurable space,* $\mu \in \mathbb{M}_1(\mathscr{X})$*, and* $h \in \mathbb{F}(\mathsf{X})$*. Assume that*

(i) there exists $\gamma \ge 0$ *such that for all* $x \in \mathsf{X}$*,*

$$-\infty < \inf_{x' \in \mathsf{X}} h(x') \le h(x) \le \inf_{x' \in \mathsf{X}} h(x') + \gamma ;$$

(ii) $\int |h(x)| \mu(\mathrm{d}x) < \infty$.

Then for all $s \ge 0$ *and* $x \in \mathsf{X}$*,*

$$\int \mathrm{e}^{s[h(x) - \int h(u)\mu(\mathrm{d}u)]} \mu(\mathrm{d}x) \le \mathrm{e}^{s^2\gamma^2/8} .$$

Proof. Without loss of generality, we assume $\int h(x)\mu(\mathrm{d}x) = 0$. For $s \ge 0$, we set

$$L(s) = \log \int \mathrm{e}^{sh(x)} \mu(\mathrm{d}x) . \tag{23.1.13}$$

Since h is bounded, the function L is (infinitely) differentiable, and we have $L(0) = 0$ and $L'(0) = 0$. Define $\mu_s \in \mathbb{M}_1(\mathscr{X})$ by

$$\mu_s(A) = \frac{\int_A \mathrm{e}^{sh(x)} \mu(\mathrm{d}x)}{\int_{\mathsf{X}} \mathrm{e}^{sh(x)} \mu(\mathrm{d}x)} , \qquad A \in \mathscr{X} . \tag{23.1.14}$$

Then

$$\begin{aligned} L'(s) &= \int_{\mathsf{X}} h(x) \mu_s(\mathrm{d}x) , \\ L''(s) &= \int_{\mathsf{X}} h^2(x) \mu_s(\mathrm{d}x) - \left\{ \int h^2(x) \mu_s(\mathrm{d}x) \right\}^2 . \end{aligned}$$

Set $c = \inf_{x \in \mathsf{X}} h(x) + \gamma/2$. Note that for all $x \in \mathsf{X}$, we have $|h(x) - c| \le \gamma/2$. Therefore, for all $s \ge 0$,

$$\begin{aligned} L''(s) &= \int \left[h(x) - \int h(x') \mu_s(\mathrm{d}x') \right]^2 \mu_s(\mathrm{d}x) \\ &\le \int [h(x) - c]^2 \mu_s(\mathrm{d}x) \le \gamma^2/4 . \end{aligned}$$

Since $L(0) = 0$ and $L'(0) = 0$, we conclude by applying Taylor's theorem that $L(s) \le s^2\gamma^2/8$ for all $s \ge 0$. $\square$

Since (23.1.11) and (23.1.12) hold, we can apply Lemma 23.1.4, and we have for all $s \ge 0$,

$$8 \log \mathbb{E}\left[\mathrm{e}^{s\{g_\ell(X_0^\ell) - g_{\ell-1}(X_0^{\ell-1})\}} \,\middle|\, \mathscr{F}_{\ell-1}^X \right] \le s^2 \gamma_\ell^2 \,. \tag{23.1.15}$$

Thus for $\ell \le n-1$, we have

$$\begin{aligned}
\mathbb{E}\left[\mathrm{e}^{sg_\ell(X_0^\ell)}\right] &= \mathbb{E}\left[\mathrm{e}^{sg_{\ell-1}(X_0^{\ell-1})}\mathrm{e}^{sg_\ell(X_0^\ell)-sg_{\ell-1}(X_0^{\ell-1})}\right] \\
&= \mathbb{E}\left[\mathrm{e}^{sg_{\ell-1}(X_0^{\ell-1})}\mathbb{E}\left[\mathrm{e}^{s\{g_\ell(X_0^\ell)-g_{\ell-1}(X_0^{\ell-1})\}} \,\middle|\, \mathscr{F}_{\ell-1}^X\right]\right] \\
&\le \mathbb{E}\left[\mathrm{e}^{sg_{\ell-1}(X_0^{\ell-1})}\right]\mathrm{e}^{s^2\gamma_\ell^2/8} \,.
\end{aligned}$$

A straightforward induction, using (23.1.10) and (23.1.15), yields for all $s \ge 0$,

$$\mathbb{E}_\xi\left[\mathrm{e}^{sg_{n-1}(X_0^{n-1})}\right] \le \exp\left((s^2/8) \sum_{\ell=0}^{n-1} \gamma_\ell^2 \right) . \tag{23.1.16}$$

Applying Markov's inequality yields

$$\mathbb{P}_\xi\left(f(X_0^{n-1}) > t\right) \le \exp\left(-st + s^2 \sum_{\ell=0}^{n-1} \gamma_\ell^2/8 \right) .$$

By choosing $s = 4t/\sum_{\ell=0}^{n-1}\gamma_\ell^2$, we obtain McDiarmid's inequality for independent random variables, stated in the following theorem.

Theorem 23.1.5. *Let $(\mathsf{X}, \mathscr{X})$ be a measurable space, $(\Omega, \mathscr{F}, \mathbb{P})$ a probability space, and $(X_0, \ldots, X_{n-1})$ an n-tuple of independent X-valued random variables defined on $(\Omega, \mathscr{F}, \mathbb{P})$. Let $(\gamma_0, \ldots, \gamma_{n-1})$ be nonnegative constants and $f \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$. Then for all $t > 0$,*

$$\mathbb{P}(f(X_0^{n-1}) - \mathbb{E}\left[f(X_0^{n-1})\right] \ge t) \le \exp\left(-\frac{2t^2}{\sum_{i=0}^{n-1}\gamma_i^2} \right) . \tag{23.1.17}$$

We now illustrate the usefulness of McDiarmid's inequality.

Example 23.1.6. Let $(\mathsf{X}, \mathscr{X})$ be a measurable space and let $X_0, \ldots, X_{n-1}$ be n independent X-valued random variables with common distribution μ. Let $\mathscr{G} \subset \mathbb{F}_b(\mathsf{X})$ be a countable collection of functions such that $\sup_{g\in\mathscr{G}} |g|_\infty \le M$ and consider the function

$$f(x_0^{n-1}) = \sup_{g\in\mathscr{G}} |\hat{\mu}_{x_0^{n-1}}(g) - \mu(g)| \,.$$

We have shown in Example 23.1.3 that $f \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$ with $\gamma_i = 2M/n$ for all $i \in \{0, \ldots, n-1\}$. Consequently, by applying Theorem 23.1.5, we get that for all $\varepsilon > 0$,

$$\mathbb{P}(|f(X_0^{n-1}) - \mathbb{E}\left[f(X_0^{n-1})\right]| \geq \varepsilon) \leq 2\mathrm{e}^{-2n\varepsilon^2/M^2} .$$

This shows that the uniform deviation $f(x_0^{n-1}) = \sup_{g \in \mathscr{G}} |\hat{\mu}_{x_0^{n-1}}(g) - \mu(g)|$ concentrates around its mean $\mathbb{E}\left[f(X_0^{n-1})\right]$ with an exponential rate. ◀

23.2 Concentration Inequality for Uniformly Ergodic Markov Chains

We will now prove a concentration inequality similar to McDiarmid's inequality (23.1.17) for uniformly ergodic Markov chains. Let P be a positive Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. We will use a martingale decomposition similar to the one obtained in (23.1.10). Assume $\mathbb{E}\left[f(X_0^{n-1})\right] = 0$. Set $g_{n-1}(x_0^{n-1}) = f(x_0^{n-1})$, and for $\ell = 0, \ldots, n-2$ and $x_0^\ell \in \mathsf{X}^{\ell+1}$, set

$$g_\ell(x_0^\ell) = \int f(x_0^{n-1}) \prod_{i=\ell+1}^{n-1} P(x_{i-1}, \mathrm{d}x_i) = \mathbb{E}_{x_\ell}\left[f(x_0^\ell, X_1^{n-\ell-1})\right] . \tag{23.2.1}$$

With these definitions, we get

$$g_{n-1}(x_0^{n-1}) = \sum_{\ell=1}^{n-1} \left\{g_\ell(x_0^\ell) - g_{\ell-1}(x_0^{\ell-1})\right\} + g_0(x_0) , \tag{23.2.2}$$

and for $\ell \in \{1, \ldots, n-1\}$ and $x_0^{\ell-1} \in \mathsf{X}^\ell$,

$$g_{\ell-1}(x_0^{\ell-1}) = \int g_\ell(x_0^{\ell-1}, x_\ell) P(x_{\ell-1}, \mathrm{d}x_\ell) . \tag{23.2.3}$$

This shows that $g_{\ell-1}(X_0^{\ell-1}) = \mathbb{E}\left[g_\ell(X_0^\ell) \,\middle|\, \mathscr{F}_{\ell-1}^X\right]$ $\mathbb{P}_\xi$ − a.s. for $\ell \geq 1$, where $\mathscr{F}_\ell^X = \sigma(X_0, \ldots, X_\ell)$. Hence $\{(g_\ell(X_0^\ell), \mathscr{F}_\ell^X),\ \ell = 0, \ldots, n-1\}$ is a $\mathbb{P}_\xi$-martingale for every $\xi \in \mathbb{M}_1(\mathscr{X})$.

When considering (23.2.1), a first crucial step is to bound $\mathbb{E}_\xi\left[h(X_0^{n-1})\right] - \mathbb{E}_{\xi'}\left[h(X_0^{n-1})\right]$, where $h \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$, and ξ, ξ' are two arbitrary initial distributions.

Considering the inequality (23.1.1), a natural idea is to use exact coupling techniques. Consider $Z = \{Z_n,\ n \in \mathbb{N}\}$, $Z' = \{Z'_n,\ n \in \mathbb{N}\}$, two X-valued stochastic processes, and T an $\bar{\mathbb{N}}$-valued random variable defined on the same probability space $(\Omega, \mathscr{G}, \mathbb{P})$ such that (Z, Z') is an *exact coupling* of $(\mathbb{P}_\xi, \mathbb{P}_{\xi'})$ with coupling time T (see Definition 19.3.3). Recall that the shift operator $\theta : \mathsf{X}^\mathbb{N} \to \mathsf{X}^\mathbb{N}$ is defined as follows: for $z = \{z_k,\ k \in \mathbb{N}\} \in \mathsf{X}^\mathbb{N}$, θz is the sequence $\theta z = \{z_{k+1},\ k \in \mathbb{N}\}$. We then set $\theta_1 = \theta$, and for $n \in \mathbb{N}^*$, we define inductively $\theta_n = \theta_{n-1} \circ \theta$. We also need to define θ_∞. To this end, fix an arbitrary $x^* \in \mathsf{X}$ and define $\theta_\infty : \mathsf{X}^{\mathbb{N}} \to \mathsf{X}^{\mathbb{N}}$ such that for $z = \{z_k,\ k \in \mathbb{N}\} \in \mathsf{X}^{\mathbb{N}}$, $\theta_\infty z \in \mathsf{X}^{\mathbb{N}}$ is the constant sequence $(\theta_\infty z)_k = x^*$ for all $k \in \mathbb{N}$. With this notation, we recall the two properties of an exact coupling:

(i) for all $A \in \mathscr{X}^{\otimes \mathbb{N}}$, we have $\mathbb{P}(Z \in A) = \mathbb{P}_\xi(A)$ and $\mathbb{P}(Z' \in A) = \mathbb{P}_{\xi'}(A)$;
(ii) $\theta_T Z = \theta_T Z'$ $\mathbb{P} - \text{a.s.}$

Then for every exact coupling of $(\mathbb{P}_\xi, \mathbb{P}_{\xi'})$ with coupling time T,

$$\begin{aligned} \left|\mathbb{E}_\xi\left[h(X_0^{n-1})\right] - \mathbb{E}_{\xi'}\left[h(X_0^{n-1})\right]\right| &= \left|\mathbb{E}\left[h(Z_0^{n-1}) - h(\{Z'\}_0^{n-1})\right]\right| \\ &\le \mathbb{E}\left[\sum_{i=0}^{n-1} \gamma_i \mathbb{1}\left\{Z_i \neq Z_i'\right\}\right] \le \sum_{i=0}^{n-1} \gamma_i \mathbb{P}(T > i)\,. \end{aligned} \tag{23.2.4}$$

If the coupling is in addition *maximal*, that is, if $\mathrm{d}_{\mathrm{TV}}(\xi P^n, \xi' P^n) = \mathbb{P}(T > n)$ for all $n \in \mathbb{N}$, then

$$\left|\mathbb{E}_\xi\left[h(X_0^{n-1})\right] - \mathbb{E}_{\xi'}\left[h(X_0^{n-1})\right]\right| \le \sum_{i=0}^{n-1} \gamma_i \mathrm{d}_{\mathrm{TV}}(\xi P^i, \xi' P^i)\,. \tag{23.2.5}$$

For example Theorem 19.3.9 shows that maximal exact couplings always exist if (X, d) is a complete separable metric space. Unfortunately, for a general state space $(\mathsf{X}, \mathscr{X})$, a maximal exact coupling may not exist without further assumptions. The following lemma provides sufficient conditions for getting an upper bound similar to (23.2.5) up to a multiplicative constant β.

Lemma 23.2.1 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Then there exists a constant β such that for all $n \in \mathbb{N}$, nonnegative constants $(\gamma_0, \ldots, \gamma_{n-1})$, $h \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$, and $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$,*

$$\left|\mathbb{E}_\xi\left[h(X_0^{n-1})\right] - \mathbb{E}_{\xi'}\left[h(X_0^{n-1})\right]\right| \le \beta \sum_{i=0}^{n-1} \gamma_i \mathrm{d}_{\mathrm{TV}}(\xi P^i, \xi' P^i)\,. \tag{23.2.6}$$

Moreover, (23.2.6) holds with $\beta = 1$ if one of the following conditions holds:

(i) For all $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$, there exists a maximal exact coupling for $(\mathbb{P}_\xi, \mathbb{P}_{\xi'})$.
(ii) For all $n \in \mathbb{N}$ and $i \in \{0, \ldots, n-1\}$, $x_i^{n-1} \mapsto \inf_{u_0^{i-1} \in \mathsf{X}^i} h(u_0^{i-1}, x_i^{n-1})$ is measurable.

Otherwise, the inequality (23.2.6) holds with $\beta = 2$.

Proof. If for all $(\xi, \xi') \in \mathbb{M}_1(\mathscr{X})$ there exists a maximal exact coupling for $(\mathbb{P}_\xi, \mathbb{P}_{\xi'})$, then (23.2.6) with $\beta = 1$ follows from (23.2.5).

We now consider the case in which a maximal coupling might not exist. Set $h_0 = h$, and for $i \in \{1, \ldots, n-1\}$,

$$h_i(x_i^{n-1}) = \inf_{u_0^{i-1} \in \mathsf{X}^i} h(u_0^{i-1}, x_i^{n-1})\,,$$

and by convention, we set h_n to the constant function $h_n(x_0^{n-1}) = \inf_{u_0^{n-1} \in \mathsf{X}^n} h(u_0^{n-1})$. Then

$$h(x_0^{n-1}) = \sum_{i=0}^{n-1} \{h_i(x_i^{n-1}) - h_{i+1}(x_{i+1}^{n-1})\} + h_n \, .$$

Assume that h_i is measurable for all $i \in \{0, \ldots, n-1\}$. Then we can set for all $i \in \{0, \ldots, n-1\}$ and $x_0^i \in \mathsf{X}^{i+1}$,

$$\begin{aligned} w_i(x_i) &= \int \left\{ h_i(x_i^{n-1}) - h_{i+1}(x_{i+1}^{n-1}) \right\} \prod_{\ell=i+1}^{n-1} P(x_{\ell-1}, \mathrm{d}x_\ell) \, , \\ &= \int \left\{ \inf_{u_0^{i-1} \in \mathsf{X}^i} h(u_0^{i-1}, x_i^{n-1}) - \inf_{u_0^i \in \mathsf{X}^{i+1}} h(u_0^i, x_{i+1}^{n-1}) \right\} \prod_{\ell=i+1}^{n-1} P(x_{\ell-1}, \mathrm{d}x_\ell) \, . \end{aligned} \tag{23.2.7}$$

With this notation, we get $\mathbb{E}\left[\{h_i(X_i^{n-1}) - h_{i+1}(X_{i+1}^{n-1})\} \,\middle|\, \mathscr{F}_i^X \right] = w_i(X_i) \;\; \mathbb{P}_\xi - \text{a.s.}$, which implies that

$$\mathbb{E}_\xi \left[h(X_0^{n-1}) \right] = \sum_{i=0}^{n-1} \xi P^i w_i + h_n \, .$$

Since $h \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$, the expression (23.2.7) shows that $0 \le w_i \le \gamma_i$. Since for $i \in \{0, \ldots, n-1\}$, $\mathrm{d}_{\mathrm{TV}}(\xi P^i, \xi' P^i) = \sup\left\{ \xi P^i f - \xi' P^i f \,:\, f \in \mathbb{F}_b(\mathsf{X}), |f|_\infty \le 1 \right\}$, we obtain

$$|\mathbb{E}_\xi \left[h(X^{n-1}) \right] - \mathbb{E}_{\xi'} \left[h(X^{n-1}) \right]| \le \sum_{i=0}^{n-1} |\xi P^i w_i - \xi' P^i w_i| \le \sum_{i=0}^{n-1} \gamma_i \mathrm{d}_{\mathrm{TV}}(\xi P^i, \xi' P^i) \, .$$

This proves (23.2.6) with $\beta = 1$.

If we no longer assume that h_i is measurable for all $i \in \{0, \ldots, n-1\}$, then we still can prove (23.2.6), but the upper bound is less tight, i.e., $\beta = 2$. Fix an arbitrary $x^* \in \mathsf{X}$. The proof follows the same lines as above, but for $i \in \{1, \ldots, n-1\}$, we replace h_i by $\bar{h}_i(x_i^{n-1}) = h(x^*, \ldots, x^*, x_i^{n-1})$. By convention, we set $\bar{h}_n$ to the constant function $\bar{h}_n = h(x^*, \ldots, x^*)$ and $\bar{h}_0 = h$. With this notation, we again have the decomposition

$$h(x_0^{n-1}) = \sum_{i=0}^{n-1} \{\bar{h}_i(x_i^{n-1}) - \bar{h}_{i+1}(x_{i+1}^{n-1})\} + \bar{h}_n \, .$$

Setting for all $i \in \{0, \ldots, n-1\}$ and all $x_0^i \in \mathsf{X}^{i+1}$,

$$\begin{aligned} \bar{w}_i(x_i) &= \int \left\{ \bar{h}_i(x_i^{n-1}) - \bar{h}_{i+1}(x_{i+1}^{n-1}) \right\} \prod_{\ell=i+1}^{n-1} P(x_{\ell-1}, \mathrm{d}x_\ell) \\ &= \int \left\{ h(x^*, \ldots, x^*, x_i^{n-1}) - h(x^*, \ldots, x^*, x_{i+1}^{n-1}) \right\} \prod_{\ell=i+1}^{n-1} P(x_{\ell-1}, \mathrm{d}x_\ell) \, , \end{aligned} \tag{23.2.8}$$

it is easily seen that $\mathbb{E}\left[\{\bar{h}_i(X_i^{n-1}) - \bar{h}_{i+1}(X_{i+1}^{n-1})\} \,\middle|\, \mathscr{F}_i^X\right] = \bar{w}_i(X_i)$ $\mathbb{P}_\xi$ − a.s., which implies that

$$\mathbb{E}_\xi\left[h(X_0^{n-1})\right] = \sum_{i=0}^{n-1} \xi P^i \bar{w}_i + \bar{h}_n \, .$$

Since $h \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$, (23.2.8) shows that $|\bar{w}_i|_\infty \le \gamma_i$. Then, using (D.2.2),

$$|\mathbb{E}_\xi\left[h(X^{n-1})\right] - \mathbb{E}_{\xi'}\left[h(X^{n-1})\right]| \le \sum_{i=0}^{n-1} |\xi P^i \bar{w}_i - \xi' P^i \bar{w}_i| \le 2 \sum_{i=0}^{n-1} \gamma_i \mathrm{d_{TV}}(\xi P^i, \xi' P^i) \, .$$

□

We now extend McDiarmid's inequality to uniformly ergodic Markov kernels. Recall that $\Delta(P)$ denotes the Dobrushin coefficient of a Markov kernel P.

Theorem 23.2.2. *Let P be Markov kernel on $\mathsf{X} \times \mathscr{X}$. Then there exists $\beta > 0$ such that for all $n \in \mathbb{N}^*$, $\gamma_0^{n-1} \in \mathbb{R}_+^n$, $f \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$, $\xi \in \mathbb{M}_1(\mathscr{X})$, and $t > 0$,*

$$\mathbb{P}_\xi\left(f(X_0^{n-1}) - \mathbb{E}_\xi[f(X_0^{n-1})] > t\right) \le \mathrm{e}^{-2t^2/D_n(\beta)} \, , \tag{23.2.9}$$

with

$$D_n(\beta) = \sum_{\ell=0}^{n-1} \left(\gamma_\ell + \beta \sum_{m=\ell+1}^{n-1} \gamma_m \Delta\left(P^{m-\ell}\right)\right)^2 \, . \tag{23.2.10}$$

Moreover, we may set $\beta = 1$ if one of the following two conditions is satisfied:

(i) For all $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$, there exists a maximal exact coupling of $(\mathbb{P}_\xi, \mathbb{P}_{\xi'})$.
(ii) For all $i \in \{0, \ldots, n-1\}$, $x_i^{n-1} \mapsto \inf_{u_0^{i-1} \in \mathsf{X}^i} h(u_0^{i-1}, x_i^{n-1})$ is measurable.

In all other cases, (23.2.9) is satisfied with $\beta = 2$.

Remark 23.2.3. For an n-tuple of independent random variables X_0^{n-1}, we can apply the previous result with $\Delta\left(P^j\right) = 0$ for all $j \ge 1$, and we recover McDiarmid's inequality:

$$\mathbb{P}\left(|f(X_0^{n-1}) - \mathbb{E}\left[f(X_0^{n-1})\right]| \ge t\right) \le 2 \exp\left(-2t^2 / \sum_{i=0}^{n-1} \gamma_i^2\right) \, . \tag{23.2.11}$$

▲

Proof (of Theorem 23.2.2). Without loss of generality, we assume $\mathbb{E}_\xi[f(X_0^{n-1})] = 0$. We need to prove only the one-sided inequality

$$\mathbb{P}_\xi\left(f(X_0^{n-1}) > t\right) \le \mathrm{e}^{-2t^2/D_n} \, . \tag{23.2.12}$$

For $\ell = 0, \dots, n-1$, set

$$A_\ell = \gamma_\ell + \beta \sum_{m=\ell+1}^{n-1} \gamma_m \Delta\left(P^{m-\ell}\right) .$$

Consider the functions g_ℓ, $\ell \in \{0, \dots, n-1\}$ defined in (23.2.1). We will prove that for each $\ell \in \{1, \dots, n-1\}$ and $x_0^\ell \in \mathsf{X}^{\ell+1}$,

$$\inf_{x \in \mathsf{X}} g_\ell(x_0^{\ell-1}, x) \le g_\ell(x_0^\ell) \le \inf_{x \in \mathsf{X}} g_\ell(x_0^{\ell-1}, x) + A_\ell , \tag{23.2.13}$$

$$\inf_{x \in \mathsf{X}} g_0(x) \le g_0(x_0) \le \inf_{x \in \mathsf{X}} g_0(x) + A_0 . \tag{23.2.14}$$

If (23.2.13) holds, then applying Hoeffding's lemma, Lemma 23.1.4, yields for all $s \ge 0$,

$$8 \log \mathbb{E}\left[\mathrm{e}^{s\{g_\ell(X_0^\ell) - g_{\ell-1}(X_0^{\ell-1})\}} \,\middle|\, \mathscr{F}_{\ell-1}^X \right] \le s^2 A_\ell^2 .$$

This and (23.2.2) yield for all $s \ge 0$,

$$8 \log \mathbb{E}_\xi \left[\mathrm{e}^{s f(X_0^{n-1})} \right] = 8 \log \mathbb{E}_\xi \left[\mathrm{e}^{s g_{n-1}(X_0^{n-1})} \right] \le s^2 \sum_{\ell=0}^{n-1} A_\ell^2 = s^2 D_n .$$

Applying Markov's inequality, we obtain

$$\mathbb{P}_\xi \left(f(X_0^{n-1}) > t \right) \le \exp\left(-st + s^2 D_n / 8 \right) .$$

Choosing $s = 4t/D_n$ yields (23.2.12). To complete the proof, it remains to establish (23.2.13) and (23.2.14). The first inequality in (23.2.13), $\inf_{x \in \mathsf{X}} g_\ell(x_0^{\ell-1}, x) \le g_\ell(x_0^\ell)$, obviously holds. We now turn to the second inequality in (23.2.13). For an arbitrary $x^* \in \mathsf{X}$,

$$\begin{aligned} g_\ell(x_0^\ell) = \int f(x_0^{n-1}) \prod_{i=\ell+1}^{n-1} P(x_{i-1}, \mathrm{d}x_i) &\le \int f(x_0^{\ell-1}, x^*, x_{\ell+1}^{n-1}) \prod_{i=\ell+1}^{n-1} P(x_{i-1}, \mathrm{d}x_i) + \gamma_\ell \\ &= \mathbb{E}_{\delta_{x_\ell} P} \left[h_{\ell+1}(X_0^{n-\ell-2}) \right] + \gamma_\ell , \end{aligned}$$

where $h_{\ell+1} : \mathsf{X}^{n-\ell-1} \to \mathbb{R}$ is defined by $h_{\ell+1}(x_{\ell+1}^{n-1}) = f(x_0^{\ell-1}, x^*, x_{\ell+1}^{n-1})$. Since $h_{\ell+1} \in \mathbb{BD}(\mathsf{X}^{n-\ell-1}, \gamma_{\ell+1}^{n-1})$, (23.2.6) shows that

$$\begin{aligned} &\mathbb{E}_{\delta_{x_\ell} P} \left[h_{\ell+1}(X_0^{n-\ell-2}) \right] - \mathbb{E}_{\delta_{x^*} P} \left[h_{\ell+1}(X_0^{n-\ell-2}) \right] \\ &\qquad \le \beta \sum_{m=\ell+1}^{n-1} \gamma_m \mathrm{d_{TV}}(\delta_{x_\ell} P^{m-\ell}, \delta_{x^*} P^{m-\ell}) . \end{aligned}$$

By Definition 18.2.1, $\mathrm{d_{TV}}(\xi P^k, \xi' P^k) \le \Delta\left(P^k\right) \mathrm{d_{TV}}(\xi, \xi')$ for all $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$, and thus we get

$$\begin{aligned}
g_\ell(x_0^\ell) &\leq \mathbb{E}_{\delta_{x^*}P}\left[h_{\ell+1}(X_0^{n-\ell-2})\right] + \beta \sum_{m=\ell+1}^{n-1} \gamma_m \Delta\left(P^{m-\ell}\right) + \gamma_\ell \\
&= g_\ell(x_0^{\ell-1}, x^*) + A_\ell \; .
\end{aligned}$$

Since x^* is arbitrary, we finally obtain $g_\ell(x_0^\ell) \leq \inf_{x^* \in \mathsf{X}} g_\ell(x_0^{\ell-1}, x^*) + A_\ell$, which completes the proof of (23.2.13). The proof of (23.2.14) is along the same lines. □

As a byproduct of Theorem 23.2.2, we obtain Hoeffding's inequality for uniformly ergodic Markov kernels. We consider the functional $x_0^{n-1} \mapsto \sum_{i=0}^{n-1} f(x_i)$, where f has bounded oscillations, and we study the deviation of the sum centered at $\pi(f)$.

Corollary 23.2.4 *Let $\{X_k,\ k \in \mathbb{N}\}$ be a uniformly ergodic Markov chain with kernel P and invariant probability measure π. Set*

$$\varDelta = \sum_{\ell=1}^{\infty} \Delta\left(P^\ell\right) < \infty \; . \tag{23.2.15}$$

Let $f : \mathsf{X} \to \mathbb{R}$ be a measurable function with bounded oscillations. Then for all $\xi \in \mathbb{M}_1(\mathscr{X})$ and $t \geq n^{-1} \mathrm{d}_{\mathrm{TV}}(\xi, \pi)(1+\varDelta)\operatorname{osc}(f)$,

$$\begin{aligned}
&\mathbb{P}_\xi\left(\left|\sum_{i=0}^{n-1} f(X_i) - \pi(f)\right| > nt\right) \\
&\qquad \leq 2\exp\left\{-\frac{2n\left(t - n^{-1}\mathrm{d}_{\mathrm{TV}}(\xi,\pi)(1+\varDelta)\operatorname{osc}(f)\right)^2}{\operatorname{osc}^2(f)\,(1+\varDelta)^2}\right\} \; . \qquad (23.2.16)
\end{aligned}$$

Remark 23.2.5. The restriction $t \geq n^{-1}\mathrm{d}_{\mathrm{TV}}(\xi,\pi)(1+\varDelta)\operatorname{osc}(f)$ is the cost of centering at $\pi(f)$. It is zero if $\xi = \pi$, in which case we simply recover (23.2.9).

Proof. Note first that the convergence of the series (23.2.15) is ensured by Theorem 18.2.5. Applying the bound (18.2.5), we obtain

$$\sum_{i=0}^{n-1} \left|\mathbb{E}_\xi[f(X_i)] - \pi(f)\right| \leq \mathrm{d}_{\mathrm{TV}}(\xi,\pi)(1+\varDelta)\operatorname{osc}(f) \; .$$

Next, applying Theorem 23.2.2 to the function $(x_0, \ldots, x_{n-1}) \mapsto \sum_{i=0}^{n-1} f(x_i)$, which satisfies (23.1.1) with $\gamma_i = \operatorname{osc}(f)$ for all $i = 0, \ldots, n-1$, we obtain

$$\begin{aligned}
&\mathbb{P}_\xi\left(\left|\sum_{i=0}^{n-1} f(X_i)-\pi(f)\right|>n\delta\right)\\
&\leq \mathbb{P}_\xi\left(\left|\sum_{i=0}^{n-1} f(X_i)-\mathbb{E}_\xi\left[f(X_i)\right]\right|+\mathrm{d}_{\mathrm{TV}}(\xi,\pi)(1+\Delta)\operatorname{osc}(f)>n\delta\right)\\
&\leq 2\mathrm{e}^{-2\left[n\delta-\mathrm{d}_{\mathrm{TV}}(\xi,\pi)(1+\Delta)\operatorname{osc}(f)\right]^2/D_n}\,,
\end{aligned}$$

with

$$D_n=\sum_{\ell=0}^{n-1}\left(\gamma_\ell+\sum_{s=\ell+1}^{n-1}\Delta\left(P^{s-\ell}\right)\gamma_s\right)^2\leq n\operatorname{osc}^2(f)(1+\Delta)^2\,.$$

□

23.3 Sub-Gaussian Concentration Inequalities for V-Geometrically Ergodic Markov Chains

The results presented in the previous section apply to uniformly ergodic Markov kernels. In this section, we will study how these results can be extended to V-uniformly ergodic Markov kernels.

Theorem 23.3.1. *Let P be an irreducible, aperiodic, and geometrically regular Markov kernel on $\mathsf{X}\times\mathscr{X}$ with invariant probability π. Then for every geometrically recurrent small set C, there exists a positive constant β such that for all $n\in\mathbb{N}^*$, $\gamma_0^{n-1}\in\mathbb{R}_+^n$, $f\in\mathbb{BD}(\mathsf{X}^n,\gamma_0^{n-1})$, $t\geq 0$ and $x\in C$,*

$$\mathbb{P}_x\left(\left|f(X_0^{n-1})-\mathbb{E}_x\left[f(X_0^{n-1})\right]\right|>t\right)\leq 2\mathrm{e}^{-\beta t^2/\sum_{\ell=0}^{n-1}\gamma_\ell^2}\,. \tag{23.3.1}$$

Remark 23.3.2. It is possible to obtain an explicit expression of the constant β as a function of the tail distribution for the return time to the set C. ▲

Remark 23.3.3. Denote by π the invariant probability measure. By Theorems 15.1.3 and 15.1.5, there exists an increasing family of geometrically recurrent small sets $\{C_n,\, n\in\mathbb{N}\}$ such that $\pi(\cup_{n\geq 0}C_n)=1$. Therefore, the exponential inequality (23.3.1) holds for π-almost all $x\in\mathsf{X}$ with a constant β that may depend on x but is uniform on geometrically recurrent small sets. In many applications, the sets C can be chosen to be the level sets $\{V\leq d\}$ of a drift function V such that $\mathrm{D_g}(V,\lambda,b,C)$ holds. ▲

Proof (of Theorem 23.3.1). Similarly to the proof of Theorem 23.2.2, the inequality (23.3.1) will be a consequence of a bound for the Laplace transform. We will prove in Lemma 23.3.4 that there exists a constant κ such that for all $x \in C$, $\gamma_0^{n-1} \in \mathbb{R}_+^n$, and $f \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$,

$$\mathbb{E}_x\left[\mathrm{e}^{f(X_0^{n-1})-\mathbb{E}_x[f(X_0^{n-1})]}\right] \leq \mathrm{e}^{\kappa \sum_{i=1}^{n-1} \gamma_i^2} . \tag{23.3.2}$$

Let $s > 0$. Applying Markov's inequality and (23.3.2) to the function $s \cdot f$ yields

$$\begin{aligned}\mathbb{P}_x(f(X_0^{n-1}) - \mathbb{E}_x[f(X_0^{n-1})] > t) &= \mathbb{P}_x\left(\mathrm{e}^{s\left(f(X_0^{n-1})-\mathbb{E}_x[f(X_0^{n-1})]\right)} > \mathrm{e}^{st}\right) \\ &\leq \mathrm{e}^{-st}\mathbb{E}_x\left[\mathrm{e}^{s(f(X_0^{n-1})-\mathbb{E}_x[f(X_0^{n-1})])}\right] \\ &\leq \mathrm{e}^{-st+\kappa s^2 \sum_{i=1}^{n-1} \gamma_i^2} .\end{aligned}$$

Taking $s = t/(2\kappa \sum_{i=1}^{n-1} \gamma_i^2)$ yields

$$\mathbb{P}_x(f(X_0^{n-1}) - \mathbb{E}_x[f(X_0^{n-1})] > t) \leq \mathrm{e}^{-t^2/(4\kappa \sum_{i=1}^{n-1} \gamma_i^2)} .$$

Applying again this inequality with f replaced by $-f$ proves (23.3.1) with $\beta = (4\kappa)^{-1}$. □

Our main task is now to prove (23.3.2).

Lemma 23.3.4 *Let P be an irreducible, aperiodic, geometrically regular Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. Then for every geometrically recurrent small set C, there exists a constant κ such that for all $\gamma_0^{n-1} \in \mathbb{R}_+^n$, $f \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$, and $x \in C$,*

$$\mathbb{E}_x\left[\mathrm{e}^{f(X_0^{n-1})-\mathbb{E}_x\left[f(X_0^{n-1})\right]}\right] \leq \exp\left(\kappa \sum_{i=1}^{n-1} \gamma_i^2\right) . \tag{23.3.3}$$

Proof. Define the stopping times $\{\tau_i,\ i \in \mathbb{N}\}$ by

$$\tau_i = \inf\{n \geq i : X_n \in C\} = i + \tau_C \circ \theta_i . \tag{23.3.4}$$

In words, for al $i \in \mathbb{N}$, τ_i is the first hitting time of the set C after i. Note that $\tau_{n-1} \geq n-1$ and $\tau_0 = 0$ if $X_0 \in C$. Thus we have, for all $x \in C$,

$$\mathbb{E}_x[f(X_0^{n-1})|\mathscr{F}_{\tau_{n-1}}^X] = f(X_0^{n-1}) , \quad \mathbb{E}_x[f(X_0^{n-1})|\mathscr{F}_{\tau_0}^X] = \mathbb{E}_x[f(X_0^{n-1})], \qquad \mathbb{P}_x - \text{a.s.}$$

Setting $G_i = \mathbb{E}_x[f(X_0^{n-1})|\mathscr{F}_{\tau_i}^X]$ for $i \in \{0,\ldots,n-1\}$, we obtain

$$f(X_0^{n-1}) - \mathbb{E}_x[f(X_0^{n-1})] = \sum_{i=1}^{n-1}\{G_i - G_{i-1}\} . \tag{23.3.5}$$

Note that for $i \in \{1,\ldots,n-1\}$, if $\tau_{i-1} > i-1$, then $\tau_i = \tau_{i-1}$, which implies $G_i - G_{i-1} = 0$. Therefore, we get

$$G_i - G_{i-1} = \{G_i - G_{i-1}\}\mathbb{1}_{\{\tau_{i-1}=i-1\}} . \tag{23.3.6}$$

We will prove in Lemma 23.3.6 that we may choose a constant $\alpha_0 \in [0,1)$ in such a way that for all $\alpha \in [\alpha_0, 1)$, there exist ς_1, ς_2 such that for all $n \in \mathbb{N}$, $\gamma_0^{n-1} \in \mathbb{R}_+^n$, $i \in \{1,\dots,n-1\}$, and $x \in C$,

$$|G_i - G_{i-1}| \le \varsigma_1 \mathbb{1}_{\{\tau_{i-1}=i-1\}}\{\max_{1\le i\le n} \gamma_i\}\sigma_C \circ \theta_{i-1} , \qquad \mathbb{P}_x - \text{a.s.} \tag{23.3.7}$$

$$|G_i - G_{i-1}|^2 \le \varsigma_2 \mathbb{1}_{\{\tau_{i-1}=i-1\}}\alpha^{-2\sigma_C\circ\theta_{i-1}} \sum_{k=i}^{n-1} \gamma_k^2 \alpha^{k-i} , \qquad \mathbb{P}_x - \text{a.s.} \tag{23.3.8}$$

For $i \in \{1,\dots,n-1\}$, we have $\mathbb{E}_x\left[G_i - G_{i-1} \,|\, \mathscr{F}^X_{\tau_{i-1}}\right] = 0$. Thus applying (23.3.6), (23.3.7), (23.3.8), and the bound $\mathrm{e}^t \le 1 + t + t^2\mathrm{e}^{|t|}$, we obtain

$$\begin{aligned}
&\mathbb{E}_x\left[\mathrm{e}^{G_i - G_{i-1}} \,|\, \mathscr{F}^X_{\tau_{i-1}}\right] \\
&\le 1 + \mathbb{E}_x\left[(G_i - G_{i-1})^2\mathrm{e}^{|G_i-G_{i-1}|} \,\middle|\, \mathscr{F}^X_{i-1}\right]\mathbb{1}_{\{\tau_{i-1}=i-1\}} \\
&\le 1 + \varsigma_2\left(\sum_{k=i}^{n-1}\gamma_k^2\alpha^{k-i}\right)\mathbb{E}_x\left[\mathrm{e}^{(-2\log\alpha+\varsigma_1\{\max_{1\le i\le n}\gamma_i\})\sigma_C}\circ\theta_{i-1} \,\middle|\, \mathscr{F}^X_{i-1}\right]\mathbb{1}_{\{\tau_{i-1}=i-1\}} \\
&\le 1 + \varsigma_2\left(\sum_{k=i}^{n-1}\gamma_k^2\alpha^{k-i}\right)\mathbb{E}_{X_{i-1}}\left[\mathrm{e}^{(-2\log\alpha+\varsigma_1\{\max_{1\le i\le n}\gamma_i\})\sigma_C}\right]\mathbb{1}_{\{\tau_{i-1}=i-1\}} .
\end{aligned}$$

The small set C is geometrically recurrent, and thus there exists $\delta > 1$ such that $\sup_{x\in C}\mathbb{E}_x[\delta^{\sigma_C}] < \infty$. Choose

$$\varepsilon \in \left(0, \varsigma_1^{-1}\log(\delta)\right) , \tag{23.3.9}$$

and then choose $\alpha \in [\alpha_0, 1)$ such that

$$-2\log(\alpha) + \varsigma_1\varepsilon \le \log(\delta) . \tag{23.3.10}$$

Since $\tau_{i-1} = i-1$ implies $X_{i-1} \in C$, this choice of α implies that for all $i \in \{1,\cdots,n\}$ and $\gamma_0^{n-1} \in \mathbb{R}_+^n$ satisfying $\max_{0\le k\le n-1}\gamma_k \le \varepsilon$,

$$\begin{aligned}
\mathbb{E}_x\left[\mathrm{e}^{G_i-G_{i-1}} \,|\, \mathscr{F}^X_{\tau_{i-1}}\right] &\le 1 + \varsigma_2\left(\sum_{k=i}^{n-1}\gamma_k^2\alpha^{k-i}\right)\sup_{x\in C}\mathbb{E}_x[\delta^{\sigma_C}] \\
&\le \exp\left(\varsigma_2\sup_{x\in C}\mathbb{E}_x[\delta^{\sigma_C}]\sum_{k=i}^{n-1}\gamma_k^2\alpha^{k-i}\right) .
\end{aligned}$$

By the decomposition (23.3.5) and successive conditioning, we obtain that for all $n \in \mathbb{N}$, $\gamma_0^{n-1} \in \mathbb{R}_+^n$ such that $\max_{0\le k\le n-1}\gamma_k \le \varepsilon$, $f \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$,

$$\mathbb{E}_x\left[\mathrm{e}^{f(X_0^{n-1})-\mathbb{E}_x\left[f(X_0^{n-1})\right]}\right]=\mathbb{E}_x\left[\mathrm{e}^{\sum_{i=1}^{n-1}\{G_i-G_{i-1}\}}\right]\le \mathrm{e}^{\kappa_\varepsilon\sum_{k=1}^{n-1}\gamma_k^2}, \tag{23.3.11}$$

where

$$\kappa_\varepsilon=\varsigma_2(1-\alpha)^{-1}\sup_{x\in C}\mathbb{E}_x[\delta^{\sigma_C}]. \tag{23.3.12}$$

We now extend this result to all $n\in\mathbb{N}$ and $\gamma_0^{n-1}\in\mathbb{R}_+^n$. Choose an arbitrary $x^*\in\mathsf{X}$ and define the function $\tilde{f}:\mathsf{X}^n\to\mathbb{R}$ by

$$\tilde{f}(x_0^{n-1})=f\left(x_0\mathbb{1}_{\{\gamma_0\le\varepsilon\}}+x^*\mathbb{1}_{\{\gamma_0>\varepsilon\}},\dots,x_{n-1}\mathbb{1}_{\{\gamma_{n-1}\le\varepsilon\}}+x^*\mathbb{1}_{\{\gamma_{n-1}>\varepsilon\}}\right).$$

Then $\tilde{f}\in\mathbb{BD}(\mathsf{X}^n,\gamma_0\mathbb{1}_{\{\gamma_0\le\varepsilon\}},\dots,\gamma_{n-1}\mathbb{1}_{\{\gamma_{n-1}\le\varepsilon\}})$, and (23.3.11) shows that

$$\mathbb{E}_x\left[\mathrm{e}^{\tilde{f}(X_0^{n-1})-\mathbb{E}_x\left[\tilde{f}(X_0^{n-1})\right]}\right]\le \mathrm{e}^{\kappa_\varepsilon\sum_{k=1}^{n-1}\gamma_k^2\mathbb{1}_{\{\gamma_k\le\varepsilon\}}}.$$

Moreover,

$$|f(x_0^{n-1})-\tilde{f}(x_0^{n-1})|\le\sum_{i=1}^{n-1}\gamma_i\mathbb{1}_{\{\gamma_i>\varepsilon\}}\le\varepsilon^{-1}\sum_{i=1}^{n-1}\gamma_i^2.$$

This implies

$$\begin{aligned}\mathbb{E}_x\left[\mathrm{e}^{f(X_0^{n-1})-\mathbb{E}_x\left[f(X_0^{n-1})\right]}\right]&\le \mathrm{e}^{2\varepsilon^{-1}\sum_{i=1}^{n-1}\gamma_i^2}\,\mathbb{E}_x\left[\mathrm{e}^{\tilde{f}(X_0^{n-1})-\mathbb{E}_x\left[\tilde{f}(X_0^{n-1})\right]}\right]\\&\le \mathrm{e}^{(2\varepsilon^{-1}+\kappa_\varepsilon)\sum_{i=1}^{n-1}\gamma_i^2}.\end{aligned}$$

Thus (23.3.3) holds for all $n\in\mathbb{N}$, $\gamma_0^{n-1}\in\mathbb{R}_+^n$, $f\in\mathbb{BD}(\mathsf{X}^n,\gamma_0^{n-1})$ with $\kappa=\kappa_\varepsilon+2\varepsilon^{-1}$. □

There remains only to prove the bounds (23.3.7) and (23.3.8). A preliminary lemma is needed.

Lemma 23.3.5 *Let P be an irreducible, aperiodic, geometrically regular Markov kernel on $\mathsf{X}\times\mathscr{X}$ with invariant probability π. For every geometrically recurrent small set C, there exist $\alpha_0\in[0,1)$ and $\varsigma_0<\infty$ such that for all $n\in\mathbb{N}$, $\gamma_0^{n-1}\in\mathbb{R}_+^n$, $f\in\mathbb{BD}(\mathsf{X}^n,\gamma_0^{n-1})$, and $x\in C$,*

$$|\mathbb{E}_x[f(X_0^{n-1})]-\mathbb{E}_\pi[f(X_0^{n-1})]|\le\varsigma_0\sum_{k=0}^{n-1}\gamma_k\alpha_0^k. \tag{23.3.13}$$

Proof. By Lemma 23.2.1, we get

$$\left|\mathbb{E}_x\left[f(X_0^{n-1})\right]-\mathbb{E}_\pi\left[f(X_0^{n-1})\right]\right|\le 2\sum_{k=0}^{n-1}\gamma_k\mathrm{d_{TV}}(\delta_xP^k,\pi).$$

The small set C is geometrically recurrent; thus there exists $\delta>1$ such that $\sup_{x\in C}\mathbb{E}_x[\delta^{\sigma_C}]<\infty$, and by Theorem 15.1.3 (c), there exist $\alpha_0\in[0,1)$ and ς such that for all $x \in C$, $k \in \mathbb{N}$,

$$\mathrm{d}_{\mathrm{TV}}(\delta_x P^k, \pi) \le \varsigma \alpha_0^k \sup_{x \in C} \mathbb{E}_x[\delta^{\sigma_C}] < \infty .$$

□

Lemma 23.3.6 *Let P be an irreducible, aperiodic, geometrically regular Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability π. Then for every geometrically recurrent small set C, there exists a constant $\alpha_0 \in [0,1)$ such that for all $\alpha \in [\alpha_0, 1)$, there exist constant ς_1, ς_2 satisfying for all $n \in \mathbb{N}$, $i \in \{1, \ldots, n-1\}$, $\gamma_0^{n-1} \in \mathbb{R}_+^n$, $f \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$, and $x \in C$,*

$$|G_i - G_{i-1}| \le \varsigma_1 \mathbb{1}_{\{\tau_{i-1} = i-1\}} \{\max_{1 \le i \le n} \gamma_i\} \sigma_C \circ \theta_{i-1} \qquad \mathbb{P}_x - \text{a.s.}, \tag{23.3.14}$$

$$|G_i - G_{i-1}|^2 \le \varsigma_2 \mathbb{1}_{\{\tau_{i-1} = i-1\}} \alpha^{-2\sigma_C \circ \theta_{i-1}} \sum_{k=i}^{n-1} \gamma_k^2 \alpha^{k-i} \qquad \mathbb{P}_x - \text{a.s.}, \tag{23.3.15}$$

where $G_i = \mathbb{E}_x[f(X_0^{n-1}) | \mathscr{F}_{\tau_i}^X]$.

Proof. For $i \in \{1, \ldots, n\}$ define

$$G_{i,1} = \mathbb{E}_x[f(X_0^{n-1}) | \mathscr{F}_{\tau_{i-1}}^X] \mathbb{1}_{\{\tau_{i-1} = i-1\}} ,$$
$$G_{i,2} = \mathbb{E}_x[f(X_0^{n-1}) | \mathscr{F}_{\tau_i}^X] \mathbb{1}_{\{\tau_{i-1} = i-1\}} .$$

With this notation, we get for $i \in \{1, \ldots, n-1\}$,

$$G_i - G_{i-1} = \{G_i - G_{i-1}\} \mathbb{1}_{\{\tau_{i-1} = i-1\}} = G_{i,2} - G_{i,1} .$$

Consider first $G_{i,1}$. Define $g_{n-1} = f$ and for $i = 0, \ldots, n-2$,

$$g_i(x_0^i) = \mathbb{E}_{x_i}[f(x_0^i, X_1^{n-i-1})] .$$

By the Markov property, we have for all $0 \le i \le n-1$, for all $x \in \mathsf{X}$,

$$\mathbb{E}_x[f(X_0^{n-1}) | \mathscr{F}_i^X] = g_i(X_0^i) , \quad \mathbb{P}_x - \text{a.s.}$$

Define also $g_{n-1,\pi} = f$ and for $0 \le i < n-1$,

$$g_{i,\pi}(x_0^i) = \mathbb{E}_\pi[f(x_0^i, X_1^{n-i-1})] = \mathbb{E}_\pi[f(x_0^i, X_{i+1}^{n-1})] . \tag{23.3.16}$$

Lemma 23.3.5 shows that there exist ς_0 and $\alpha_0 \in [0,1)$ such that for all $n \in \mathbb{N}$, $i \in \{0, \ldots, n-1\}$, $\gamma_0^{n-1} \in \mathbb{R}_+^n$, $f \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$, and $x_i \in C$,

$$|g_i(x_0^i) - g_{i,\pi}(x_0^i)| \le \varsigma_0 \sum_{k=i+1}^{n-1} \gamma_k \alpha_0^{k-i} . \tag{23.3.17}$$

This implies that

$$G_{i,1} = g_{i-1}(X_0^{i-1})\mathbb{1}_{\{\tau_{i-1}=i-1\}} = R_{i,1} + g_{i-1,\pi}(X_0^{i-1})\mathbb{1}_{\{\tau_{i-1}=i-1\}} ,$$

where by (23.3.17), $|R_{i,1}| \leq \varsigma_0 \sum_{k=i}^{n-1} \gamma_k \alpha_0^{k-i+1}$. Consider now $G_{i,2}$:

$$G_{i,2} = f(X_0^{n-1})\mathbb{1}_{\{\tau_{i-1}=i-1,\tau_i \geq n-1\}} + \sum_{j=i}^{n-2} g_j(X_0^j)\mathbb{1}_{\{\tau_{i-1}=i-1,\tau_i=j\}} .$$

Then noting that if $\tau_i < \infty$, $X_{\tau_i} \in C$, and using again (23.3.17), we obtain

$$\sum_{j=i}^{n-2} g_j(X_0^j)\mathbb{1}_{\{\tau_{i-1}=i-1,\tau_i=j\}} = R_{i,2} + \sum_{j=i}^{n-2} g_{j,\pi}(X_0^j)\mathbb{1}_{\{\tau_{i-1}=i-1,\tau_i=j\}} ,$$

where $|R_{i,2}| \leq \varsigma_0 \sum_{k=\tau_i+1}^{n-1} \gamma_k \alpha_0^{k-\tau_i}$ with the convention $\sum_{k=s}^{t} = 0$ if $t < s$. Thus for $i \in \{1,\ldots,n-1\}$, we get

$$\begin{aligned}
&\{G_i - G_{i-1}\}\mathbb{1}_{\{\tau_{i-1}=i-1\}} \\
&\quad = R_{i,1} + R_{i,2} + \left[f(X_0^{n-1}) - g_{i-1,\pi}(X_0^{i-1})\right]\mathbb{1}_{\{\tau_{i-1}=i-1,\tau_i \geq n-1\}} \\
&\qquad + \sum_{j=i}^{n-2} \left[g_{j,\pi}(X_0^j) - g_{i-1,\pi}(X_0^{i-1})\right]\mathbb{1}_{\{\tau_{i-1}=i-1,\tau_i=j\}} .
\end{aligned}$$

Then

$$|f(x_0^{n-1}) - g_{i-1,\pi}(x_0^{i-1})| \leq \mathbb{E}_\pi[|f(x_0^{n-1}) - f(x_0^{i-1}, X_i^{n-1})|] \leq \sum_{k=i}^{n-1} \gamma_k .$$

And similarly, for all $1 \leq i \leq j \leq n-2$, using (23.3.16), we get

$$|g_{j,\pi}(x_0^j) - g_{i-1,\pi}(x_0^{i-1})| \leq \mathbb{E}_\pi[|f(x_0^j, X_{j+1}^{n-1}) - f(x_0^{i-1}, X_i^{n-1})|] \leq \sum_{k=i}^{j} \gamma_k .$$

Altogether, we have obtained that for all $i \in \{1,\ldots,n-2\}$,

$$\begin{aligned}
&|G_i - G_{i-1}|\mathbb{1}_{\{\tau_{i-1}=i-1\}} \\
&\qquad \leq \varsigma_0 \sum_{k=i}^{n-1} \gamma_k \alpha_0^{k-i+1} + \varsigma_0 \sum_{k=\tau_i+1}^{n-1} \gamma_k \alpha_0^{k-\tau_i} + \sum_{k=i}^{\tau_i \wedge (n-1)} \gamma_k .
\end{aligned} \tag{23.3.18}$$

This yields, with $\bar{\gamma} = \max_{1 \leq i \leq n} \gamma_i$,

$$|G_i| \leq 2\varsigma_0 \bar{\gamma}(1-\alpha_0)^{-1} + \bar{\gamma}(\tau_i - i + 1) .$$

Since $\tau_i = i + \tau_C \circ \theta_i$ and $\sigma_C = 1 + \tau_C \circ \theta$, we get for $i \geq 1$ that $\tau_i - i + 1 = \sigma_C \circ \theta_{i-1}$, and the previous equation yields (23.3.7) with $\varsigma_1 = 2\varsigma_0(1 + (1-\alpha_0)^{-1})$.

We now consider (23.3.15). Note first that for $\alpha \in (\alpha_0, 1)$, (23.3.4) and (23.3.18) yield

$$|G_i - G_{i-1}| \le \varsigma_0 \sum_{k=i}^{n-1} \gamma_k \alpha^{k-i} + \varsigma_0 \alpha^{-\tau_i+i} \sum_{k=\tau_i+1}^{n-1} \gamma_k \alpha^{k-i+1} + \alpha^{-\tau_i+i} \sum_{k=i}^{\tau_i \wedge (n-1)} \gamma_k \alpha^{k-i}$$
$$\le \varsigma_0 \sum_{k=i}^{n-1} \gamma_k \alpha^{k-i} + \varsigma_0 \alpha^{-\sigma_C \circ \theta_{i-1}} \sum_{k=i}^{n-1} \gamma_k \alpha^{k-i} \le 2\varsigma_0 \alpha^{-\sigma_C \circ \theta_{i-1}} \sum_{k=i}^{n-1} \gamma_k \alpha^{k-i} .$$

The latter bound and the Cauchy–Schwarz inequality yield

$$|G_i - G_{i-1}|^2 \le 4\varsigma_0^2 (1-\alpha)^{-1} \alpha^{-2\sigma_C \circ \theta_{i-1}} \sum_{k=i}^{n-1} \gamma_k^2 \alpha^{k-i} .$$

This proves (23.3.8) with $\varsigma_2 = 4\varsigma_0^2(1-\alpha)^{-1}$. □

Since P is not uniformly ergodic, we cannot obtain a deviation inequality for all initial distributions. However, we can extend 23.3.1 to the case that the initial distribution is the invariant probability.

Corollary 23.3.7 *Let P be a geometrically ergodic Markov kernel. Then there exists a constant β such that for all $f \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$ and all $t \ge 0$,*

$$\mathbb{P}_\pi\left(\left|f(X_0^{n-1}) - \mathbb{E}_\pi\left[f(X_0^{n-1})\right]\right| > t\right) \le 2\mathrm{e}^{-\beta t^2 / \sum_{i=0}^{n-1} \gamma_i^2} .$$

Proof. Assume that P is geometrically ergodic. Then by Lemma 9.3.9, the Markov kernel P is aperiodic. By Theorem 15.1.5, there exists an accessible geometrically recurrent small set C. Let $x^* \in C$. For $k > 0$ and $f \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$, define

$$\Delta_k = \left|\mathbb{E}_{x^*}\left[f(X_k^{n+k-1})\right] - \mathbb{E}_\pi\left[f(X_0^{n-1})\right]\right| .$$

The function $x_0^{n+k-1} \mapsto f(x_k^{n+k-1})$ belongs to $\mathbb{BD}(\mathsf{X}^{n+k}, 0_0^{k-1}, \gamma_0^{n-1})$, where 0_0^{k-1} is the k-dimensional null vector. Thus applying Theorem 23.3.1, there exists $\beta > 0$ such that

$$\mathbb{P}_{x^*}\left(\left|f(X_k^{n+k-1}) - \mathbb{E}_\pi\left[f(X_0^{n-1})\right]\right| > t\right)$$
$$\le \mathbb{P}_{x^*}\left(\left|f(X_k^{n+k-1}) - \mathbb{E}_{x^*}\left[f(X_k^{n+k-1})\right]\right| > (t-\Delta_k)^+\right)$$
$$\le 2\mathrm{e}^{-\beta((t-\Delta_k)^+)^2 / \sum_{i=0}^{n-1} \gamma_i^2} .$$

Moreover, for all $x \in C$, $\lim_{n\to\infty} \mathrm{d}_{\mathrm{TV}}(P^n(x,\cdot), \pi) = 0$. Thus $\lim_{k\to\infty} \Delta_k = 0$, and for every bounded measurable function h, $\lim_{k\to\infty} P^k h(x^*) = \pi(h)$. Setting $h(x) = \mathbb{P}_x\left(\left|f(X_k^{n+k-1}) - \mathbb{E}_\pi\left[f(X_0^{n-1})\right]\right| > t\right)$, we therefore get

$$\begin{aligned}\pi(h) &= \mathbb{P}_\pi\left(\left|f(X_0^{n-1}) - \mathbb{E}_\pi\left[f(X_0^{n-1})\right]\right| > t\right) \\ &= \lim_{k\to\infty} P^k h(x^*) = \lim_{k\to\infty} \mathbb{P}_{x^*}\left(\left|f(X_k^{n+k-1}) - \mathbb{E}_\pi\left[f(X_0^{n-1})\right]\right| > t\right) \\ &\le \lim_{k\to\infty} 2\mathrm{e}^{-\beta((t-\Delta_k)^+)^2/\sum_{i=0}^{n-1}\gamma_i^2} = 2\mathrm{e}^{-\beta t^2/\sum_{i=0}^{n-1}\gamma_i^2}\;.\end{aligned}$$

The proof is completed. □

Similar to Corollary 23.2.4, we obtain as a corollary an exponential inequality for the empirical measure $\hat{\pi}_n$ centered at $\pi(f)$.

Corollary 23.3.8 *Let P be a geometrically recurrent Markov kernel. Then for every geometrically recurrent small set C, there exist constants $\beta > 0$ and κ such that for all $x \in C$ and $t > \kappa n^{-1}\operatorname{osc}(f)$,*

$$\mathbb{P}_x\left(|\hat{\pi}_n(f) - \pi(f)| > t\right) \le 2\exp\left\{-\frac{2n\left(t - \kappa n^{-1}\operatorname{osc}(f)\right)^2}{\kappa^2\operatorname{osc}^2(f)}\right\}\;. \tag{23.3.19}$$

Proof. By Lemma 23.3.5, there exists a constant κ such that for all $x \in C$,

$$|\mathbb{E}_x[\hat{\pi}_n(f)] - \pi(f)| \le \kappa n^{-1}\operatorname{osc}(f)\;.$$

The rest of the proof is along the same lines as the proof of Corollary 23.2.4. □

23.4 Exponential Concentration Inequalities Under Wasserstein Contraction

In this section, (X,d) is a complete separable metric space endowed with its Borel σ-field, denoted by $\mathscr{X}$. We cannot extend McDiarmid's inequality to functions of bounded differences applied to a nonirreducible Markov chain. Recall from Chapter 18 that functions of bounded differences are closely related to the total variation distance. As seen in Chapter 20, in order to use the Wasserstein distance, we can consider only Lipschitz functions. Therefore, we introduce the following definition, which parallels Definition 23.1.1.

Definition 23.4.1 (Separately Lipschitz functions) *A function $f : \mathsf{X}^n \to \mathbb{R}$ is separately Lipschitz if there exist nonnegative constants $(\gamma_0, \ldots, \gamma_{n-1})$ such that for all $x_0^{n-1} \in \mathsf{X}^n$ and $y_0^{n-1} \in \mathsf{X}^n$,*

$$|f(x_0^{n-1}) - f(y_0^{n-1})| \le \sum_{i=0}^{n-1} \gamma_i \mathsf{d}(x_i, y_i) . \tag{23.4.1}$$

The class of all functions f that satisfy (23.4.1) *is denoted by* $\mathrm{Lip}_{\mathsf{d}}(\mathsf{X}^n, \gamma_0^{n-1})$.

Note that if d is the Hamming distance, then $\mathrm{Lip}_{\mathsf{d}}(\mathsf{X}^n, \gamma_0^{n-1}) = \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$. We first state a technical result, which is similar to Lemma 23.2.1.

Lemma 23.4.2 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$ such that $\Delta_{\mathsf{d}}(P) < \infty$. Then for all $n \in \mathbb{N}^*$, $\gamma_0^0[n-1] \in \mathbb{R}_+^n$, $f \in \mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$, and $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$,*

$$|\mathbb{E}_\xi[f(X_0^{n-1})] - \mathbb{E}_{\xi'}[f(X_0^{n-1})]| \le \sum_{i=0}^{n-1} \gamma_i \Delta_{\mathsf{d}}^i(P) \mathbf{W}_{\mathsf{d}}(\xi, \xi') .$$

Proof. Theorem 20.1.3 shows that there exists a kernel coupling K of (P,P) such that for all $(x,x') \in \mathsf{X} \times \mathsf{X}$, $\mathbf{W}_{\mathsf{d}}(P(x,\cdot), P(x',\cdot)) = K\mathsf{d}(x,x')$. Using Theorem 20.3.2 and Proposition 20.3.3, we have for all $i \in \mathbb{N}$ and $(x,x') \in \mathsf{X} \times \mathsf{X}$,

$$K^i \mathsf{d}(x,x') \le \Delta_{\mathsf{d}}^i(P)\, \mathsf{d}(x,x') . \tag{23.4.2}$$

Let $\eta \in \mathscr{C}(\xi, \xi')$ and let $\bar{\mathbb{P}}_\eta$ be the probability measure on $((\mathsf{X} \times \mathsf{X})^{\mathbb{N}}, (\mathscr{X} \otimes \mathscr{X})^{\otimes \mathbb{N}})$ that makes the coordinate process $\{(X_n, X_n'),\ n \in \mathbb{N}\}$ a Markov chain with the Markov kernel K and initial distribution η, and let $\bar{\mathbb{E}}_\eta$ be the associated expectation operator. Then applying (23.4.2) yields

$$\begin{aligned}
&|\mathbb{E}_\xi[f(X_0^{n-1})] - \mathbb{E}_{\xi'}[f(X_0^{n-1})]| = |\bar{\mathbb{E}}_\eta[f(X_0^{n-1}) - f(\{X'\}_0^{n-1})]| \\
&\quad \le \bar{\mathbb{E}}_\eta\left[\sum_{i=0}^{n-1} \gamma_i \mathsf{d}(X_i, X_i')\right] = \sum_{i=0}^{n-1} \gamma_i \eta(K^i \mathsf{d}) \le \sum_{i=0}^{n-1} \gamma_i \Delta_{\mathsf{d}}^i(P) \int \eta(\mathrm{d}x\mathrm{d}y) \mathsf{d}(x,y) ,
\end{aligned}$$

which completes the proof, since η is arbitrary in $\mathscr{C}(\xi, \xi')$. □

In order to get exponential concentration inequalities for

$$\mathbb{P}_x(|f(X_0^{n-1}) - \mathbb{E}_x[f(X_0^{n-1})]| > t)$$

where $f \in \mathrm{Lip}_{\mathsf{d}}(\mathsf{X}^n, \gamma_0^{n-1})$, we again make use of the functions g_ℓ, $\ell \in \{0, \ldots, n-1\}$ defined in (23.2.1):

$$g_\ell(x_0^\ell) = \int f(x_0^{n-1}) \prod_{i=\ell+1}^{n-1} P(x_{i-1}, \mathrm{d}x_i) = \mathbb{E}_{x_\ell}\left[f(x_0^\ell, X_1^{n-\ell-1})\right] . \tag{23.4.3}$$

Combining Lemma 23.4.2 and (23.4.3) shows that for all $\ell \in \{0,\dots,n-1\}$, $x_0^{\ell-1} \in \mathsf{X}^\ell$, and $x,y \in \mathsf{X}$,

$$\begin{aligned}
&g_\ell(x_0^{\ell-1},x) - g_\ell(x_0^{\ell-1},y) \\
&\qquad \le \gamma_\ell \mathsf{d}(x,y) + \mathbb{E}_x\left[f(x_0^{\ell-1},y,X_1^{n-\ell-1})\right] - \mathbb{E}_y\left[f(x_0^{\ell-1},y,X_1^{n-\ell-1})\right] \\
&\qquad \le \gamma_\ell \mathsf{d}(x,y) + \sum_{i=\ell+1}^{n-1} \gamma_i \Delta_{\mathsf{d}}^{i-\ell-1}(P)\, \mathbf{W}_{\mathsf{d}}(P(x,\cdot),P(y,\cdot)) \\
&\qquad \le \left\{\sum_{i=\ell}^{n-1} \gamma_i \Delta_{\mathsf{d}}^{i-\ell}(P)\right\} \mathsf{d}(x,y)\,.
\end{aligned} \tag{23.4.4}$$

Theorem 23.4.3. *Let* (X,d) *be a complete separable metric space and let* P *be a Markov kernel on* $\mathsf{X}\times\mathscr{X}$*. Assume that there exist constants* $(\beta,\delta) \in \mathbb{R}_+ \times \bar{\mathbb{R}}_+$ *such that for all measurable functions* h *such that* $|h|_{\mathrm{Lip}(\mathsf{d})} \le \delta$ *and all* $x \in \mathsf{X}$*,*

$$P(\mathrm{e}^h)(x) \le \mathrm{e}^{2\beta^2|h|^2_{\mathrm{Lip}(\mathsf{d})}}\mathrm{e}^{Ph(x)}\,. \tag{23.4.5}$$

Let $n \ge 1$ *and let* $\gamma_0,\dots,\gamma_{n-1}$ *be nonnegative real numbers (at least one of which is positive). Define for* $\ell \in \{0,\dots,n-1\}$*,*

$$\alpha_\ell = \sum_{i=\ell}^{n-1} \gamma_i \Delta_{\mathsf{d}}^{i-\ell}(P)\,, \qquad \alpha^* = \max_{0\le k\le n-1} \alpha_k\,, \qquad \alpha^2 = \sum_{k=0}^{n-1} \alpha_k^2\,.$$

Then for all $f \in \mathrm{Lip}_{\mathsf{d}}(\mathsf{X}^n,\gamma_0^{n-1})$ *and all* $x \in \mathsf{X}$*,*

$$\mathbb{P}_x(|f(X_0^{n-1}) - \mathbb{E}_x[f(X_0^{n-1})]| > t) \le \begin{cases} 2\mathrm{e}^{-\frac{t^2}{8\beta^2\alpha^2}} & \text{if} \quad 0 \le t \le \delta(2\beta\alpha)^2/\alpha^*\,, \\ 2\mathrm{e}^{-\frac{\delta t}{2\alpha^*}} & \text{if} \quad t > \delta(2\beta\alpha)^2/\alpha^*\,. \end{cases} \tag{23.4.6}$$

Proof. Without loss of generality, we assume $\mathbb{E}_x[f(X_0^{n-1})] = 0$. We again consider the functions g_ℓ, $\ell \in \{0,\dots,n-1\}$ defined in (23.4.3). By (23.4.4), for all $s \ge 0$ and $x_0^{\ell-1} \in \mathsf{X}^\ell$, the function

$$x \mapsto s g_\ell(x_0^{\ell-1},x)$$

is Lipschitz with constant $s\alpha_\ell = s\sum_{i=\ell}^{n-1}\gamma_i\Delta_{\mathsf{d}}^{i-\ell}(P)$. Thus if $s\alpha^* \le \delta$, then combining (23.4.5) with (23.2.3), we obtain for all $\ell \in \{1,\dots,n-1\}$,

$$\mathbb{E}\left[\mathrm{e}^{sg_\ell(X_0^\ell)} \,\middle|\, \mathscr{F}_{\ell-1}^X\right] \le \mathrm{e}^{2s^2\beta^2\alpha_\ell^2}\mathrm{e}^{sg_{\ell-1}(X_0^{\ell-1})}\,,$$
$$\mathbb{E}\left[\mathrm{e}^{sg_0(X_0)}\right] \le \mathrm{e}^{2s^2\beta^2\alpha_0^2}\,.$$

This implies, using the decomposition (23.2.2),

$$\log \mathbb{E}_x \left[\mathrm{e}^{sf(X_0^{n-1})} \right] = \log \mathbb{E}_x \left[\mathrm{e}^{sg_{n-1}(X_0^{n-1})} \right] \le 2s^2\beta^2 \sum_{\ell=0}^{n-1} \alpha_\ell^2 \,.$$

Applying Markov's inequality and setting $\alpha^2 = \sum_{\ell=0}^{n-1} \alpha_\ell^2$, we obtain

$$\mathbb{P}_x \left(f(X_0^{n-1}) - \mathbb{E}_x[f(X_0^{n-1})] > t \right) \le \exp\left(-st + 2s^2\beta^2\alpha^2 \right) \,. \tag{23.4.7}$$

If $0 \le t \le \delta(2\beta\alpha)^2/\alpha^*$, we can choose $s = (2\beta\alpha)^{-2}t$, which implies $s\alpha^* \le \delta$, and consequently,

$$\mathbb{P}_x \left(f(X_0^{n-1}) - \mathbb{E}_x[f(X_0^{n-1})] > t \right) \le \mathrm{e}^{-\frac{t^2}{8\beta^2\alpha^2}} \,.$$

If $t > \delta(2\beta\alpha)^2/\alpha^*$, we choose $s = \delta/\alpha^*$. Then

$$2s^2\beta^2\alpha^2 = 2\frac{\delta^2\beta^2\alpha^2}{(\alpha^*)^2} \le \frac{\delta t}{2\alpha^*} \,.$$

Plugging this inequality and $s = \delta/\alpha^*$ into (23.4.7) yields

$$\mathbb{P}_x \left(f(X_0^{n-1}) - \mathbb{E}_x[f(X_0^{n-1})] > t \right) \le \mathrm{e}^{-\frac{\delta t}{2\alpha^*}} \,.$$

This proves (23.4.6). □

For us to be able to apply Theorem 23.4.3, the Markov kernel P must satisfy property (23.4.5). This inequality can be proved when P satisfies the so-called logarithmic Sobolev inequality. See Exercise 23.5. We will here illustrate this result by considering a Markov kernel P on $\mathsf{X} \times \mathscr{X}$ with finite granularity, defined as

$$\sigma_\infty = \frac{1}{2} \sup_{x \in \mathsf{X}} \operatorname{diam} \{ \operatorname{supp}(\mathsf{S}(x)) \} \,, \qquad \mathsf{S}(x) = \operatorname{supp}(P(x, \cdot)) \,. \tag{23.4.8}$$

For every Lipschitz function h we get

$$\begin{aligned} \mathbb{E}_x \left[\{h(X_1) - Ph(x)\}^2 \right] &= \frac{1}{2} \iint_{\mathsf{X}^2} \{h(y) - h(z)\}^2 P(x, \mathrm{d}y) P(x, \mathrm{d}z) \\ &\le \frac{1}{2} \iint_{\mathsf{S}(x) \times \mathsf{S}(x)} \{h(y) - h(z)\}^2 P(x, \mathrm{d}y) P(x, \mathrm{d}z) \\ &\le \frac{1}{2} |h|^2_{\mathrm{Lip}(\mathsf{d})} \iint_{\mathsf{S}(x) \times \mathsf{S}(x)} \mathsf{d}^2(y, z) P(x, \mathrm{d}y) P(x, \mathrm{d}z) \,. \end{aligned}$$

Since $\mathsf{d}(y, z) \le 2\sigma_\infty$ for all $y, z \in \mathsf{S}(x)$, the previous bound implies

$$\mathbb{E}_x \left[\{h(X_1) - Ph(x)\}^2 \right] \le |h|^2_{\mathrm{Lip}(\mathsf{d})} \, 2\sigma_\infty^2 \,. \tag{23.4.9}$$

Lemma 23.4.4 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that the granularity $\sigma_\infty < \infty$ is finite. Let $h : \mathsf{X} \to \mathbb{R}$ be a measurable function such that $|h|_{\mathrm{Lip}(\mathsf{d})} \in (0, 1/(3\sigma_\infty)]$. Then for all $x \in \mathsf{X}$,*

$$P(\mathrm{e}^h)(x) \leq \mathrm{e}^{2|h|^2_{\mathrm{Lip}(\mathsf{d})}\sigma^2_\infty} \mathrm{e}^{Ph(x)} . \tag{23.4.10}$$

Proof. Note that for all $u \in [0,1]$ and $x \in \mathsf{X}$,

$$\begin{aligned}\{(1-u)Ph(x) + uh(X_1)\} &\leq Ph(x) + |h|_{\mathrm{Lip}(\mathsf{d})} \int_{\mathsf{S}(x)} \mathsf{d}(X_1, y) P(x, \mathrm{d}y) \\ &\leq Ph(x) + 2\,|h|_{\mathrm{Lip}(\mathsf{d})}\,\sigma_\infty \qquad \mathbb{P}_x - \text{a.s.} ,\end{aligned} \tag{23.4.11}$$

where we have used that $X_1 \in \mathsf{S}(x)$ $\mathbb{P}_x$ − a.s. and for all $y, z \in \mathsf{S}(x)$, $\mathsf{d}(y,z) \leq 2\sigma_\infty$, where $\mathsf{S}(x)$ is defined in (23.4.8). Set $\varphi(u) = \exp\{[(1-u)Ph(x) + uh(X_1)]\}$, $u \in [0,1]$. Writing $\varphi(1) \leq \varphi(0) + \varphi'(0) + \sup_{u \in [0,1]} \varphi''(u)/2$ and taking the expectation with respect to $\mathbb{P}_x$ yields

$$\mathbb{E}_x[\mathrm{e}^{h(X_1)}] \leq \mathrm{e}^{Ph(x)} + \frac{1}{2}\mathbb{E}_x\left[\{h(X_1) - Ph(x)\}^2\right] \mathrm{e}^{Ph(x) + 2|h|_{\mathrm{Lip}(\mathsf{d})}\sigma_\infty} ,$$

where the last term on the right-hand side follows from (23.4.11). Combining this with the bound (23.4.9), we finally get

$$P(\mathrm{e}^h)(x) = \mathbb{E}_x[\mathrm{e}^{h(X_1)}] \leq \mathrm{e}^{Ph(x)}\left(1 + |h|^2_{\mathrm{Lip}(\mathsf{d})}\,\sigma^2_\infty \mathrm{e}^{2|h|_{\mathrm{Lip}(\mathsf{d})}\sigma_\infty}\right) .$$

If $|h|_{\mathrm{Lip}(\mathsf{d})} < 1/(3\sigma_\infty)$, then $\mathrm{e}^{2|h|_{\mathrm{Lip}(\mathsf{d})}\sigma_\infty} \leq \mathrm{e}^{2/3} \leq 2$, and this proves (23.4.10) using $1 + u \leq \mathrm{e}^u$. □

Theorem 23.4.5. *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Assume that $\Delta_{\mathsf{d}}(P) < \infty$ and $\sigma_\infty < \infty$, where σ_∞ is defined in (23.4.8). Let $n \geq 1$ and let $\gamma_0, \ldots, \gamma_{n-1}$ be nonnegative real numbers (at least one of which is positive). Define*

$$\alpha_k = \sum_{i=k}^{n-1} \gamma_i \Delta^{i-k}_{\mathsf{d}}(P) , \qquad \alpha^* = \max_{0 \leq k \leq n-1} \alpha_k , \qquad \alpha^2 = \sum_{k=0}^{n-1} \alpha_k^2 .$$

Then for all $f \in \mathrm{Lip}_{\mathsf{d}}(\mathsf{X}^n, \gamma_0^{n-1})$,

$$\mathbb{P}_x(|f(X_0^{n-1}) - \mathbb{E}_x[f(X_0^{n-1})]| > t) \leq \begin{cases} 2\mathrm{e}^{-\frac{t^2}{8\alpha^2\sigma^2_\infty}} & \text{if} \quad 0 \leq t \leq 4\alpha^2\sigma_\infty/(3\alpha^*) , \\ 2\mathrm{e}^{-\frac{t}{6\alpha^*\sigma_\infty}} & \text{if} \quad t > 4\alpha^2\sigma_\infty/(3\alpha^*) . \end{cases} \tag{23.4.12}$$

Proof. Lemma 23.4.4 shows that (23.4.5) is satisfied with $\delta = 1/3\sigma_\infty$ and $\beta^2 = \sigma_\infty^2$. The result then follows from Theorem 23.4.3. □

23.5 Exercises

23.1 (Hoeffding's lemma). In this exercise, we derive another proof of Hoeffding's inequality. Let V be an integrable random variable on $(\Omega, \mathscr{F}, \mathbb{P})$ and let $\mathscr{G} \subset \mathscr{F}$ be a σ-field such that

$$\mathbb{E}\left[V \middle| \mathscr{G}\right] = 0 \quad \text{and} \quad A \leq V \leq B \qquad \mathbb{P} - \text{a.s.},$$

where A, B are two $\mathscr{G}$-measurable random variables.

1. Set $p = -A/(B-A)$ and $\phi(u) = -pu + \log(1-p+pe^u)$. Show that

$$\mathbb{E}\left[\mathrm{e}^{sV} \middle| \mathscr{G}\right] = (1-p+p\mathrm{e}^{s(B-A)})\mathrm{e}^{-ps(B-A)} = \mathrm{e}^{\phi(s(B-A))}.$$

2. Show that for all $s > 0$,

$$\mathbb{E}\left[\mathrm{e}^{sV} \middle| \mathscr{G}\right] \leq \mathrm{e}^{\frac{1}{8}s^2(B-A)^2} .$$

23.2. Let P be a uniformly ergodic Markov kernel on $\mathbb{R}$ with invariant probability π and let $x \mapsto F_\pi(x) = \pi((-\infty, x])$ be the associated distribution function. Let $\{X_t,\ t \in \mathbb{N}\}$ be the canonical chain associated with the kernel P, and let F_n be the corresponding empirical distribution function, i.e., $\hat{F}_n(x) = n^{-1}\sum_{t=0}^{n-1} \mathbb{1}_{\{X_t \leq x\}}$. Let the Kolmogorov–Smirnov statistic K_n be defined by

$$K_n = \sup_{x \in \mathbb{R}} \left|\hat{F}_n(x) - F_\pi(x)\right| .$$

Prove that for every initial distribution ξ, $n \geq 1$, and $t \geq n^{-1}\mathrm{d}_{\mathrm{TV}}(\xi, \pi)(1+\Delta)$, we have

$$\mathbb{P}_\xi\left(K_n \geq t\right) \leq 2\exp\left\{-\frac{2n\left(t - n^{-1}\mathrm{d}_{\mathrm{TV}}(\xi,\pi)(1+\Delta)\right)^2}{(1+2\Delta)^2}\right\} ,$$

where Δ is defined in (23.2.15).
[Hint: Write K_n as a function of $(X_0, \ldots, X_{n-1})$ that satisfies the assumptions of Theorem 23.2.2.]

23.3. Let P be a uniformly ergodic Markov kernel on $\mathbb{R}^d$ and let $\{X_k,\ k \in \mathbb{N}\}$ be the associated canonical chain. Assume that the (unique) invariant probability π has a density h with respect to Lebesgue measure on $\mathbb{R}^d$. Let K be a measurable nonnegative function on $\mathbb{R}^d$ such that $\int_{\mathbb{R}^d} K(u)\mathrm{d}u = 1$. Given $X_0, \ldots, X_{n-1}$, a nonparametric kernel density estimator h_n of the density h is defined by

$$h_n(x) = \frac{1}{n}\sum_{i=0}^{n-1}\frac{1}{b_n^d}K\left(\frac{x-X_i}{b_n}\right), \quad x \in \mathbb{R}^d , \tag{23.5.1}$$

where $\{b_n,\, n \in \mathbb{N}^*\}$ is a sequence of positive real numbers (called bandwidths). The integrated error is defined by

$$J_n = \int |h_n(x) - h(x)|\mathrm{d}x .$$

Prove that for all $n \geq 1$ and $t > 0$,

$$\mathbb{P}_\pi(|J_n - \mathbb{E}_\pi[J_n]| > t) \leq 2\exp\left(-\frac{nt^2}{2(1+2\Delta)^2}\right) . \tag{23.5.2}$$

[Hint: Write J_n as a function of $X_0,\dots,X_{n-1}$ that satisfy the assumptions of Theorem 23.2.2.]

23.4. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, where (X,Υ) is a complete separable metric space and $\mathscr{X}$ is the associated σ-field. Assume that $\Delta_{\mathsf{d}}(P) \leq 1-\kappa$ with $\kappa \in [0,1)$. Recall that by Theorem 20.3.4, P admits a unique invariant measure π such that

$$E(x) = \int_{\mathsf{X}} \mathsf{d}(x,y)\pi(\mathrm{d}y) < \infty$$

for all $x \in \mathsf{X}$. Note that $E(x) \leq \operatorname{diam}(\mathsf{X})$. Let f be a 1-Lipschitz function. Define

$$\hat{\pi}_n(f) = \frac{1}{n}\sum_{i=0}^{n-1} f(X_i) .$$

1. Show that

$$|\mathbb{E}_x[\hat{\pi}_n(f)] - \pi(f)| \leq \frac{E(x)}{n\kappa} \leq \frac{\operatorname{diam}(\mathsf{X})}{n\kappa} . \tag{23.5.3}$$

2. Assume that $\sigma_\infty < \infty$ (see (23.4.8)) and that $\operatorname{diam}(\mathsf{X}) < \infty$. Show that for every Lipschitz function f and $t > (n\kappa)^{-1}|f|_{\mathrm{Lip}(\mathsf{d})}\operatorname{diam}(\mathsf{X})$,

$$\mathbb{P}_x(|\hat{\pi}_n(f) - \pi(f)| > t) \leq \begin{cases} 2\mathrm{e}^{-n\kappa^2(t-(n\kappa)^{-1}|f|_{\mathrm{Lip}(\mathsf{d})}\operatorname{diam}(\mathsf{X}))^2/(8\sigma_\infty^2)} & \text{if } 0 \leq t \leq 4\sigma_\infty/(3\kappa) , \\ 2\mathrm{e}^{-n\kappa(t-(n\kappa)^{-1}|f|_{\mathrm{Lip}(\mathsf{d})}\operatorname{diam}(\mathsf{X}))/(6\sigma_\infty)} & \text{if } t > 4\sigma_\infty/(3\kappa) . \end{cases} \tag{23.5.4}$$

23.5. Assume that the kernel P on $\mathbb{R}^d$ satisfies the logarithmic Sobolev inequality, that is, for all continuously differentiable functions $f : \mathbb{R}^d \to \mathbb{R}$ and all $x \in \mathbb{R}^d$,

$$P(f^2\log(f^2))(x) - (Pf^2)(x)\log(Pf^2)(x) \leq 2CP(|\nabla f|^2)(x) . \tag{23.5.5}$$

1. Set $f_t^2 = \mathrm{e}^{th-\frac{1}{2}t^2C|h|^2_{\mathrm{Lip(d)}}}$. Prove that for all $t \in \mathbb{R}$ and all $x \in \mathbb{R}$,

$$P(|\nabla f_t|^2)(x) \le \frac{t^2}{4}\,|h|^2_{\mathrm{Lip(d)}}\,P(f_t^2)(x)\,. \tag{23.5.6}$$

2. Set $\Lambda(t,x) = Pf_t(x)$. Use (23.5.5) and (23.5.6) to prove that for all $x \in \mathbb{R}^d$ and $t \in \mathbb{R}$,

$$t\Lambda'(t,x) \le \Lambda(t,x)\log\Lambda(t,x)\,. \tag{23.5.7}$$

3. Deduce that $\Lambda(t) \le 1$ for all $t \in \mathbb{R}_+$ and $x \in \mathbb{R}^d$. Conclude that (23.4.5) holds with $\beta^2 = C/4$ and $\delta = \infty$.

23.6 Bibliographical Notes

The concentration of measure phenomenon was evidenced by V. Milman in the 1970s while he was studying the asymptotic geometry of Banach spaces. The monographs Ledoux (2001) and Boucheron et al. (2013) present numerous examples and probabilistic, analytical, and geometric techniques related to this notion. A short but insightful introduction is given in Bercu et al. (2015).

Hoeffding's inequality (see Lemma 23.1.4) was first established in Hoeffding (1963). McDiarmid's inequality (Theorem 23.1.5) was established in McDiarmid (1989) which introduced the method of proof based on a martingale decomposition. McDiarmid's inequality for uniformly ergodic Markov chains was first established in Rio (2000a), whose result is (slightly) improved in Theorem 23.2.2. See also Samson (2000) for other types of concentration inequalities for uniformly ergodic Markov chains. For additive functionals, Lezaud (1998) proved a Prokhorov-type inequality under a spectral gap condition in L^2, from which a sub-Gaussian concentration inequality follows.

There are far fewer results for geometrically ergodic chains. Adamczak (2008) established a sub-Gaussian concentration inequality for geometrically ergodic Markov chains under the additional assumptions that the Markov kernel is strongly aperiodic and that the functional is invariant under permutations of variables. The sub-Gaussian concentration inequality of V-geometrically ergodic Markov chains stated in Theorem 23.3.1 is adapted from Dedecker and Gouëzel (2015) (see also Chazottes and Gouëzel (2012)). The proof essentially follows the original derivation Dedecker and Gouëzel (2015) but simplifies (and clarifies) some arguments using distributional coupling.

The exponential concentration inequality for uniformly contractive Markov chains in the Wasserstein distance (Theorem 23.4.5) is due to Joulin and Ollivier (2010). Many important results had already been presented in Djellout et al. (2004). The use of the logarithmic Sobolev inequality (illustrated in Exercise 23.5) to prove concentration inequalities is the subject of a huge literature. See, e.g., Ledoux (2001).

Appendices

Appendix A
Notations

Sets and Numbers

- $\mathbb{N}$: the set of natural numbers including zero, $\mathbb{N} = \{0, 1, 2, \dots\}$.
- $\mathbb{N}^*$: the set of natural numbers excluding zero, $\mathbb{N}^* = \{1, 2, \dots\}$.
- $\bar{\mathbb{N}}$: the extended set of natural numbers, $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$.
- $\mathbb{Z}$: the set of relative integers, $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$.
- $\mathbb{R}$: the set of real numbers.
- $\mathbb{R}^d$: Euclidean space consisting of all column vectors $x = (x_1, \dots, x_d)'$.
- $\bar{\mathbb{R}}$: the extended real line, i.e., $\mathbb{R} \cup \{-\infty, \infty\}$.
- $\lceil x \rceil$: the smallest integer greater than or equal to x.
- $\lfloor x \rfloor$: the greatest integer less than or equal to x.
- If $a = \{a(n),\ n \in \mathbb{Z}\}$ and $b = \{b(n),\ n \in \mathbb{Z}\}$ are two sequences, $a * b$ denotes the convolution of a and b, defined formally by $a * b(n) = \sum_{k \in \mathbb{Z}} a(k) b(n-k)$. The jth power of convolution of the sequence a is denoted by a^{*j} with $a^{*0}(0) = 1$ and $a^{*0}(k) = 0$ if $k \neq 0$.

Metric Spaces

- (X, d): a metric space.
- $\mathrm{B}(x, r)$: the open ball of radius $r > 0$ centered at x,

$$\mathrm{B}(x, r) = \{y \in \mathsf{X} : d(x, y) < r\} .$$

- $\overline{U}$: closure of the set $U \subset \mathsf{X}$.
- ∂U: boundary of the set $U \subset \mathsf{X}$.

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1

Binary Relations

- $a \wedge b$: the minimum of a and b.
- $a \vee b$: the maximum of a and b.

Let $\{a_n,\ n \in \mathbb{N}\}$ and $\{b_n,\ n \in \mathbb{N}\}$ be two positive sequences.

- $a_n \asymp b_n$: the ratio of the two sides is bounded from above and below by positive constants that do not depend on n.
- $a_n \sim b_n$: the ratio of the two sides converges to 1.

Vectors, Matrices

- $\mathbb{M}_d(\mathbb{R})$ (resp. $\mathbb{M}_d(\mathbb{C})$): the set of $d \times d$ matrices with real (resp. complex) coefficients.
- For $M \in \mathbb{M}_d(\mathbb{C})$ and $|\cdot|$, any norm on $\mathbb{C}^d$, $|||M|||$, is the operator norm, defined as
$$|||M||| = \sup\left\{ \frac{|Mx|}{|x|},\ x \in \mathbb{C}^d,\ x \neq 0 \right\} .$$
- Id $d \times d$: the identity matrix.
- Let A and B be $m \times n$ and $p \times q$ matrices, respectively. Then the Kronecker product $A \otimes B$ of A and B is the $mp \times nq$ matrix whose (i,j)th block is the $p \times q$ matrix $A_{i,j}B$, where $A_{i,j}$ is the (i,j)th element of A. Note that the Kronecker product is associative: $(A \otimes B) \otimes C = A \otimes (B \otimes C)$ and $(A \otimes B)(C \otimes B) = (AC \otimes BD)$ (for matrices with compatible dimensions).
- Let A be an $m \times n$ matrix. Then $\mathrm{Vec}(A)$ is the $(mn \times 1)$ vector obtained from A by stacking the columns of A (from left to right). Note that $\mathrm{Vec}(ABC) = (C^T \otimes A)\mathrm{Vec}(B)$.

Functions

- $\mathbb{1}_A$: indicator function with $\mathbb{1}_A(x) = 1$ if $x \in A$ and 0 otherwise. $\mathbb{1}\{A\}$ is used if A is a composite statement.
- f^+: the positive part of the function f, i.e., $f^+(x) = f(x) \vee 0$.
- f^-: the negative part of the function f, i.e., $f^-(x) = -(f(x) \wedge 0)$.
- $f^{-1}(A)$: the inverse image of the set A under f.
- For f a real-valued function on X, $|f|_\infty = \sup\{f(x) : x \in \mathsf{X}\}$ is the supremum norm and $\mathrm{osc}\,(f)$ is the oscillation seminorm, defined as
$$\mathrm{osc}\,(f) = \sup_{(x,y) \in \mathsf{X} \times \mathsf{X}} |f(x) - f(y)| = 2 \inf_{c \in \mathbb{R}} |f - c|_\infty . \tag{A.0.1}$$

- A nonnegative (resp. positive) function is a function with values in $[0,\infty]$ (resp. $(0,\infty]$).
- A nonnegative (resp. positive) real-valued function is a function with values in $[0,\infty)$ (resp. $(0,\infty)$).
- If $f : \mathsf{X} \to \mathbb{R}$ and $g : \mathsf{Y} \to \mathbb{R}$ are two functions, then $f \otimes g$ is the function from $\mathsf{X} \times \mathsf{Y}$ to $\mathbb{R}$ defined for all $(x,y) \in \mathsf{X} \times \mathscr{Y}$ by $f \otimes g(x,y) = f(x)g(y)$.

Function Spaces

Let $(\mathsf{X}, \mathscr{X})$ be a measurable space.

- $\mathbb{F}(\mathsf{X})$: the vector space of measurable functions from $(\mathsf{X}, \mathscr{X})$ to $(-\infty,\infty)$.
- $\mathbb{F}_+(\mathsf{X})$: the cone of measurable functions from $(\mathsf{X}, \mathscr{X})$ to $[0,\infty]$.
- $\mathbb{F}_b(\mathsf{X})$: the subset of $\mathbb{F}(\mathsf{X})$ of bounded functions.
- For all $\xi \in \mathbb{M}_s(\mathscr{X})$ and $f \in \mathbb{F}_b(\mathsf{X})$, $\xi(f) = \int f \mathrm{d}\xi$.
- If X is a topological space, then
  - $\mathrm{C}_b(\mathsf{X})$ is the space of all bounded continuous real functions defined on X;
  - $\mathrm{C}(\mathsf{X})$ is the space of all continuous real functions defined on X;
  - $\mathrm{U}_b(\mathsf{X})$ is the space of all bounded uniformly continuous real functions defined on X;
  - $\mathrm{U}(\mathsf{X})$ is the space of all uniformly continuous real functions defined on X;
  - $\mathrm{Lip}_b(\mathsf{X})$ is the space of all bounded real Lipschitz functions defined on X;
  - $\mathrm{Lip}(\mathsf{X})$ is the space of all Lipschitz real functions defined on X.
- If X is a locally compact separable metric space:
  - $\mathrm{C}_c(\mathsf{X})$ is the space of all continuous functions with compact support.
  - $\mathbb{F}_0(\mathsf{X})$ is the space of all continuous functions that converge to zero at infinity.
- $\mathscr{L}^p(\mu)$: the space of measurable functions f such that $\int |f|^p \mathrm{d}\mu < \infty$.

Measures

Let $(\mathsf{X}, \mathscr{X})$ be a measurable space.

- δ_x: Dirac measure with mass concentrated on x, i.e., $\delta_x(A) = 1$ if $x \in A$ and 0 otherwise.
- Leb: Lebesgue measure on $\mathbb{R}^d$.
- $\mathbb{M}_s(\mathscr{X})$ is the set of finite signed measures on $(\mathsf{X}, \mathscr{X})$.
- $\mathbb{M}_+(\mathscr{X})$ is the set of measures on $(\mathsf{X}, \mathscr{X})$.
- $\mathbb{M}_1(\mathscr{X})$ denotes the set of probabilities on $(\mathsf{X}, \mathscr{X})$.
- $\mathbb{M}_0(\mathscr{X})$: the set of finite signed measures ξ on $(\mathsf{X}, \mathscr{X})$ satisfying $\xi(\mathsf{X}) = 0$.

- $\mathbb{M}_b(\mathscr{X})$: the set of bounded measures ξ on $(\mathsf{X}, \mathscr{X})$.
- $\mu \ll \nu$: μ is absolutely continuous with respect to ν.
- $\mu \sim \nu$: μ is equivalent to ν, i.e., $\mu \ll \nu$ and $\nu \ll \mu$.

If X is a topological space (in particular a metric space), then $\mathscr{X}$ is always taken to be the Borel sigma-field generated by the topology of X. If $\mathsf{X} = \bar{\mathbb{R}}^d$, its Borel sigma-field is denoted by $\mathscr{B}\left(\bar{\mathbb{R}}^d\right)$.

- $\mathrm{supp}(\mu)$: the (topological) support of a measure μ on a metric space.
- $\mu_n \stackrel{\mathrm{w}}{\Rightarrow} \mu$: the sequence of probability measures $\{\mu_n,\ n \in \mathbb{N}\}$ converges weakly to μ, i.e., for all $h \in \mathrm{C}_b(\mathsf{X})$, $\lim_{n\to\infty} \mu_n(h) = \mu(h)$.

The topological space X is locally compact if every point $x \in \mathsf{X}$ has a compact neighborhood.

- $\mathrm{C}_0(\mathsf{X})$: the Banach space of continuous functions that vanish at infinity.
- $\mu_n \stackrel{\mathrm{w}^*}{\Rightarrow} \mu$: The sequence of σ-finite measures $\{\mu_n,\ n \in \mathbb{N}\}$ converges to μ **weakly*, i.e., $\lim_{n\to\infty} \mu_n(h) = \mu(h)$ for all $h \in \mathrm{C}_0(\mathsf{X})$.

Probability Spaces

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. A random variable X is a measurable mapping from $(\Omega, \mathscr{F})$ to $(\mathsf{X}, \mathscr{X})$.

- $\mathbb{E}(X)$, $\mathbb{E}[X]$: the expectation of a random variable X with respect to the probability $\mathbb{P}$.
- $\mathrm{Cov}(X,Y)$: covariance of the random variables X and Y.
- Given a sub-σ-field $\mathscr{F}$ and $A \in \mathscr{A}$, $\mathbb{P}(A \mid \mathscr{F})$ is the conditional probability of A given $\mathscr{F}$ and $\mathbb{E}[X \mid \mathscr{F}]$ is the conditional expectation of X given $\mathscr{F}$.
- $\mathscr{L}_{\mathbb{P}}(X)$: the distribution of X on $(\mathsf{X}, \mathscr{X})$ under $\mathbb{P}$, i.e., the image of $\mathbb{P}$ under X.
- $X_n \stackrel{\mathbb{P}}{\Longrightarrow} X$: the sequence of random variables (X_n) converges to X in distribution under $\mathbb{P}$.
- $X_n \stackrel{\mathbb{P}-\mathrm{prob}}{\longrightarrow} X$: the sequence of random variables (X_n) converges to X in probability under $\mathbb{P}$.
- $X_n \stackrel{\mathbb{P}\text{-a.s.}}{\longrightarrow} X$: the sequence of random variables (X_n) converges to X $\mathbb{P}$-almost surely.

Usual Distributions

- $B(n,p)$: binomial distribution of n trials with success probability p.
- $\mathrm{N}(\mu,\sigma^2)$: normal distribution with mean μ and variance σ^2.
- $\mathrm{Unif}(a,b)$: uniform distribution of $[a,b]$.
- χ^2: chi-square distribution.
- χ^2_n: chi-square distribution with n degrees of freedom.

Appendix B
Topology, Measure and Probability

B.1 Topology

B.1.1 Metric Spaces

Theorem B.1.1 (Baire's theorem). *Let* (X,d) *be a complete metric space. A countable union of closed sets with empty interior has an empty interior.*

Proof. (Rudin 1991, Theorem 2.2). □

Let (X,d) be a metric space. The distance from a point to a set and the distance between two sets are defined, for $x \in \mathsf{X}$ and $E, F \subset \mathsf{X}$, by

$$d(x,E) = \inf\{d(x,y) \,:\, y \in E\} \;,$$
$$d(E,F) = \inf\{d(x,y) \,:\, x \in E, y \in F\} = \inf\{d(x,F) \,:\, x \in E\} \;.$$

Observe that $d(x,E) = 0$ if and only if $x \in \overline{E}$. The diameter of $E \subset \mathsf{X}$ is defined as $\operatorname{diam}(E) = \sup\{d(x,y) \,:\, x,y \in E\}$. A set E is said to be bounded if $\operatorname{diam}(E) < \infty$.

Lemma B.1.2 *Let* X *be a metric space. Let* F *be a closed set and* W *an open set such that* $F \subset W$*. Then there exists* $f \in \mathrm{C}(\mathsf{X})$ *such that* $0 \le f \le 1$*,* $f = 1$ *on* F*, and* $f = 0$ *on* W^c*.*

Proof. For all $x \in \mathsf{X}$, $d(x,F) + d(x,W^c) > 0$. Therefore, the function f defined by $f(x) = 1 - d(x,F)/\{d(x,F) + d(x,W^c)\}$ has the required properties. □

Lemma B.1.3 *Let* (X,d) *be a separable metric space. Then for every* $\varepsilon > 0$*, there exists a partition* $\{A_k,\ k \in \mathbb{N}^*\} \subset \mathscr{B}(\mathsf{X})$ *of* X *such that* $\operatorname{diam}(A_k) \le \varepsilon$ *for all* $k \ge 1$*.*

Proof. Since X is separable, there exists a countable covering $\{B_n,\ n \in \mathbb{N}^*\}$ of X by open balls of radius $\varepsilon/2$. We set $A_1 = B_1$ and, for $k \ge 2$, $A_k = B_k \setminus \bigcup_{j=1}^{k} B_j$, so that

R. Douc et al., *Markov Chains*, Springer Series in Operations Research
and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1

$\{A_k,\ k \in \mathbb{N}^*\}$ defines a partition of X by Borel sets with diameters equal to at most ε. □

Definition B.1.4 (Polish space)
A Polish space X is a topological space that is separable and completely metrizable, i.e., there exists a metric d inducing the topology of X such that (X, d) is a complete separable metric space.

B.1.2 Lower and Upper Semicontinuous Functions

Definition B.1.5 *Let X be a topological space.*

(i) A function $f : \mathsf{X} \to (-\infty, \infty]$ is said to be lower semicontinuous at x_0 if for all $a < f(x_0)$, there exists $V \in \mathscr{V}_{x_0}$ such that for all $x \in V$, $f(x) \geq a$. A function f is lower semicontinuous on X (which we denote by f is lower semicontinuous) if f is lower semicontinuous at x for all $x \in \mathsf{X}$.
(ii) A function $f : \mathsf{X} \to [-\infty, \infty)$ is upper semicontinuous at x_0 if $-f$ is lower semicontinuous at x_0. A function f is upper semicontinuous on X if f is upper semicontinuous at x for all $x \in \mathsf{X}$.
(iii) A function $f : \mathsf{X} \to (-\infty, \infty)$ is continuous if it is both lower and upper semicontinuous.

Lemma B.1.6 *Let X be a topological space. A function $f : \mathsf{X} \to \mathbb{R}$ is lower semicontinuous if and only if the set $\{f > a\}$ is open for all $a \in \mathbb{R}$; it is upper semicontinuous if and only if $\{f < a\}$ is open for all $a \in \mathbb{R}$.*

Proposition B.1.7 *Let X be a topological space.*

(i) If U is open in X, then $\mathbb{1}_U$ is lower semicontinuous.
(ii) If f is lower semicontinuous and $c \in [0, \infty)$, then cf is lower semicontinuous.
(iii) If $\mathscr{G}$ is a family of lower semicontinuous functions and $f(x) = \sup\{g(x) : g \in \mathscr{G}\}$, then f is lower semicontinuous.
(iv) If f and g are lower semicontinuous, then $f + g$ is lower semicontinuous.
(v) If f is lower semicontinuous and K is compact, then $\inf\{x \in K : f(x)\} = f(x_0)$ for some $x_0 \in K$.

(vi) If X *is a metric space and* f *is lower semicontinuous and nonnegative, then*

$$f(x) = \sup\{g(x) : g \in \mathrm{C}(\mathsf{X}), 0 \le g \le f\} .$$

Proposition B.1.8 *Let* X *be a separable metric space. A function* $f : \mathsf{X} \to \mathbb{R}$ *is lower semicontinuous if and only if there exists an increasing sequence* $\{f_n, n \in \mathbb{N}\}$ *of continuous functions such that* $f = \lim_{n\to\infty} f_n$.

Proposition B.1.9 *Let* X *be a topological space and* $A \in \mathsf{X}$.

(i) If f *is lower semicontinuous, then* $\sup_{x\in A} f(x) = \sup_{x\in \overline{A}} f(x)$.
(ii) If f *is upper semicontinuous, then* $\inf_{x\in A} f(x) = \inf_{x\in \overline{A}} f(x)$.

B.1.3 Locally Compact Separable Metric Spaces

Definition B.1.10 *A metric space* (X, d) *is said to be a locally compact separable metric space if it is separable and if each point admits a relatively compact neighborhood.*

A compact space is locally compact but there exist locally compact sets that are not compact, such as the Euclidean space $\mathbb{R}^n$ and every space locally homeomorphic to $\mathbb{R}^n$.

Definition B.1.11 *Let* (X, d) *be a locally compact separable metric space.*

(i) The support of a real-valued continuous function f*, denoted by* $\mathrm{supp}(f)$*, is the closure of the set* $\{|f| > 0\}$.
(ii) $\mathrm{C}_c(\mathsf{X})$ *is the space of continuous functions with compact support.*
(iii) $\mathrm{C}_0(\mathsf{X})$ *is the space of continuous functions vanishing at infinity, i.e., for every* $\varepsilon > 0$*, there exists a compact* K_ε *such that* $|f(x)| \le \varepsilon$ *if* $x \notin K_\varepsilon$.
(iv) $\mathrm{C}_b(\mathsf{X})$ *is the space of bounded continuous functions.*

The following inclusions obviously hold: $\mathrm{C}_c(\mathsf{X}) \subset \mathrm{C}_0(\mathsf{X}) \subset \mathrm{C}_b(\mathsf{X})$. Moreover, $\mathrm{C}_0(\mathsf{X})$ is the closure of $\mathrm{C}_c(\mathsf{X})$ with respect to the uniform topology. The next result is known as Urysohn's lemma.

Lemma B.1.12 *If X is a locally compact separable metric space and $K \subset U \subset \mathsf{X}$, where K is compact and U is open, then there exist a relatively compact open set W such that $K \subset W \subset U$ and a function $f \in \mathrm{C}_c(\mathsf{X})$ such that $\mathbb{1}_K \le f \le \mathbb{1}_W$.*

Proposition B.1.13 *Let (X, d) be a locally compact separable metric space.*

(i) Let U be an open set. There exists a sequence $\{V_n,\ n \in \mathbb{N}^*\}$ of relatively compact open sets such that $V_n \subset \bar{V}_n \subset V_{n+1}$ and $U = \bigcup_n V_n$.
(ii) Let U be an open set. There exists an increasing sequence $\{f_n,\ n \in \mathbb{N}^*\}$, $f_n \in \mathrm{C}_c(\mathsf{X})$, $0 \le f_n \le 1$, such that $f_n \uparrow \mathbb{1}_U$ ($\mathbb{1}_U$ is the pointwise increasing limit of elements of $\mathrm{C}_c(\mathsf{X})$).
(iii) Let K be a compact set. There exists a sequence $\{V_n,\ n \in \mathbb{N}^*\}$ of relatively compact open sets such that $V_{n+1} \subset \bar{V}_{n+1} \subset V_n$ and $K = \bigcap_n V_n$.
(iv) Let K be a compact set. There exists a decreasing sequence $\{f_n,\ n \in \mathbb{N}^*\}$, $f_n \in \mathrm{C}_c(\mathsf{X})$, $0 \le f_n \le 1$, such that $f_n \downarrow \mathbb{1}_K$ ($\mathbb{1}_K$ is the pointwise decreasing limit of functions in $\mathrm{C}_c(\mathsf{X})$).

Lemma B.1.14 *Let $f \ge 0$ be a lower semicontinuous function. Then there exists an increasing sequence $\{f_n,\ n \in \mathbb{N}\} \in \mathrm{C}_c^+(\mathsf{X})$ such that $f = \lim_{n\to\infty} f_n$.*

Theorem B.1.15. *Let (X, d) be a locally compact separable metric space. Then $\mathrm{C}_0(\mathsf{X})$ equipped with the uniform norm is separable.*

B.2 Measures

An algebra on a set X is a nonempty set of subsets of X that is closed under finite union, finite intersection, and complement (hence contains X). A σ-field on X is a set of subsets of X that is closed under countable union, countable intersection, and complement. A measurable space is a pair $(\mathsf{X}, \mathscr{X})$, where X is a nonempty set and $\mathscr{X}$ a σ-field.

A σ-field $\mathscr{B}$ is said to be generated by a collection of sets $\mathscr{C}$, written $\mathscr{B} = \sigma(\mathscr{C})$, if $\mathscr{B}$ is the smallest σ-field containing all the sets of $\mathscr{C}$. A σ-field $\mathscr{B}$ is countably generated if it is generated by a countable collection $\mathscr{C}$.

If X is a topological space, its Borel σ-field is the σ-field generated by its topology.

B.2.1 Monotone Class Theorems

Definition B.2.1 *Let Ω be a set. A collection $\mathscr{M}$ of subsets of Ω is called a monotone class if*

(i) $\Omega \in \mathscr{M}$;
(ii) $A, B \in \mathscr{M}$, $A \subset B \Longrightarrow B \setminus A \in \mathscr{M}$;
(iii) $\{A_n,\ n \in \mathbb{N}\} \subset \mathscr{M}$, $A_n \subset A_{n+1} \Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathscr{M}$.

A σ-field is a monotone class. The intersection of an arbitrary family of monotone classes is also a monotone class. Hence for every family of subsets $\mathscr{C}$ of Ω, there is a smallest monotone class containing $\mathscr{C}$, which is the intersection of all monotone classes containing $\mathscr{C}$.

If $\mathscr{N}$ is a monotone class that is stable under finite intersection, then $\mathscr{N}$ is an algebra. Indeed, since $\Omega \subset \mathscr{N}$, $\mathscr{N}$ is stable by proper difference, and if $A \in \mathscr{N}$, then $A^c = \Omega \setminus A \in \mathscr{N}$. Stability under finite intersection implies stability under finite union.

Theorem B.2.2. *Let $\mathscr{C} \subset \mathscr{M}$. Assume that $\mathscr{C}$ is stable under finite intersection and that $\mathscr{M}$ is a monotone class. Then $\sigma(\mathscr{C}) \subset \mathscr{M}$.*

Proof. (Billingsley 1986, Theorem 3.4). □

Theorem B.2.3. *Let $\mathscr{H}$ be a vector space of bounded functions on Ω and $\mathscr{C}$ a class of subsets of Ω stable under finite intersection. Assume that $\mathscr{H}$ satisfies*

(i) $\mathbb{1}_{\Omega} \in \mathscr{H}$ *and for all* $A \in \mathscr{C}$, $\mathbb{1}_A \in \mathscr{H}$;
(ii) if $\{f_n,\ n \in \mathbb{N}\}$ *is a bounded increasing sequence of functions of* $\mathscr{H}$, *then* $\sup_{n\in\mathbb{N}} f_n \in \mathscr{H}$.

Then $\mathscr{H}$ contains all the bounded $\sigma(\mathscr{C})$-measurable functions.

Theorem B.2.4. *Let $\mathscr{H}$ be a vector space of bounded real-valued functions on a measurable space $(\Omega, \mathscr{A})$, and $\{X_i,\ i \in I\}$ a family of measurable functions from $(\Omega, \mathscr{A})$ to $(\mathsf{X}_i, \mathscr{F}_i)$. Assume the following:*

(i) If $\{Y_n,\ n \in \mathbb{N}\}$ *is a bounded increasing sequence of functions of* $\mathscr{H}$, *then* $\sup_{n\in\mathbb{N}} Y_n \in \mathscr{H}$.

(ii) For all J a finite subset of I and $A_i \in \mathscr{F}_i$, $i \in J$,

$$\prod_{i \in J} \mathbb{1}_{A_i} \circ X_i \in \mathscr{H} \ .$$

Then $\mathscr{H}$ contains all the bounded $\sigma(X_i, i \in I)$-measurable functions.

B.2.2 Measures

Let A and B be two subsets of a set E. Let $A \Delta B$ be the set of elements of $A \cup B$ that are not in $A \cap B$, i.e.,

$$A \Delta B = (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A) \ .$$

Lemma B.2.5 *Let μ be a bounded measure on a measurable set $(\mathsf{X}, \mathscr{X})$ and let $\mathscr{A}$ be a subalgebra of $\mathscr{X}$. If $\mathscr{X} = \sigma(\mathscr{A})$, then for every measure μ on $\mathscr{X}$, for every $\varepsilon > 0$ and $B \in \mathscr{X}$, there exists $A \in \mathscr{A}$ such that $\mu(A \Delta B) \le \varepsilon$.*

Proof. Let $\mathscr{M}$ be the set of $B \in \mathscr{X}$ having the desired property. Let us prove that $\mathscr{M}$ is a monotone class. By definition, $\mathscr{A} \subset \mathscr{M}$, and thus $\mathsf{X} \in \mathscr{M}$. Let $B, C \in \mathscr{M}$ be such that $B \subset C$. For $\varepsilon > 0$, let $A, A' \in \mathscr{A}$ be such that $\mu(A \Delta B) \le \varepsilon$ and $\mu(A' \Delta C) \le \varepsilon$. Then

$$\begin{aligned}\mu((C \setminus B) \Delta (A' \setminus A)) &= \mu(|\mathbb{1}_C \mathbb{1}_{B^c} - \mathbb{1}_{A'} \mathbb{1}_{A^c}|) \\ &\le \mu\left(|\mathbb{1}_C - \mathbb{1}_{A'}| \mathbb{1}_{B^c}\right) + \mu\left(\mathbb{1}_{A'} |\mathbb{1}_{B^c} - \mathbb{1}_{A^c}|\right) \le \mu(C \Delta A') + \mu(B \Delta A) \le 2\varepsilon \ .\end{aligned}$$

This proves that $C \setminus B \in \mathscr{M}$. Now let $\{B_n, \ n \in \mathbb{N}\}$ be an increasing sequence of elements of $\mathscr{M}$ and set $B = \cup_{i=1}^{\infty} B_i$, $\bar{B}_n = B \setminus B_n$. Then $\bar{B}_n$ decreases to $\emptyset$. Thus by Lebesgue's dominated convergence theorem for $\varepsilon > 0$ there exists $n \ge 1$ such that $\mu(\bar{B}_n) \le \varepsilon$. Let also $A \in \mathscr{A}$ be such that $\mu(A \Delta B_n) \le \varepsilon$. Then

$$\begin{aligned}\mu(A \Delta B) &= \mu(A \setminus \{B_n \cup \bar{B}_n\}) + \mu(\{B_n \cup \bar{B}_n\} \setminus A) \\ &\le \mu(A \setminus B_n) + \mu(B_n \setminus A) + \mu(\bar{B}_n) = \mu(A \Delta B_n) + \mu(\bar{B}_n) \le 2\varepsilon \ .\end{aligned}$$

We have proved that $\mathscr{M}$ is a monotone class, whence $\mathscr{M} = \sigma(\mathscr{A}) = \mathscr{X}$. □

Theorem B.2.6. *Let μ and ν be two measures on a measurable space $(\mathsf{X}, \mathscr{X})$, and let $\mathscr{C} \subset \mathscr{X}$ be stable under finite intersection. If for all $A \in \mathscr{C}$, $\mu(A) = \nu(A) < \infty$ and $\mathsf{X} = \bigcup_{n=1}^{\infty} \mathsf{X}_n$ with $\mathsf{X}_n \in \mathscr{C}$, then $\mu = \nu$ on $\sigma(\mathscr{C})$.*

Definition B.2.7 (Image measure) *Let μ be a measure on $(\mathsf{X}, \mathscr{X})$ and $f : (\mathsf{X}, \mathscr{X}) \to (\mathsf{Y}, \mathscr{Y})$ a measurable function. The image measure $\mu \circ f^{-1}$ of μ under f is a measure on Y, defined by $\mu \circ f^{-1}(B) = \mu(f^{-1}(B))$ for all $B \in \mathscr{Y}$.*

A set function μ defined on an algebra $\mathscr{A}$ is said to be σ-additive if for each collection $\{A_n,\ n \in \mathbb{N}\}$ of mutually disjoint sets of $\mathscr{A}$ such that $\bigcup_{n=0}^{\infty} A_n \in \mathscr{A}$,

$$\mu\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} \mu(A_n)\ .$$

It is necessary to assume that $\bigcup_{n=0}^{\infty} A_n \in \mathscr{A}$, since $\mathscr{A}$ is an algebra, not a σ-field.

Theorem B.2.8 (Carathéodory's extension theorem). *Let X be a set and $\mathscr{A}$ an algebra. Let μ be a σ-additive nonnegative set function on an algebra $\mathscr{A}$ of a set X. Then there exists a measure $\bar{\mu}$ on $\sigma(\mathscr{A})$. If μ is σ-finite, this extension is unique.*

Proof. See (Billingsley 1986, Theorem 3.1). □

B.2.3 Integrals

Theorem B.2.9 (Monotone convergence theorem). *Let μ be a measure on $(\mathsf{X}, \mathscr{X})$ and let $\{f_i,\ i \in \mathbb{N}\} \subset \mathbb{F}_+(\mathsf{X})$ be an increasing sequence of functions. Let $f = \sup_{n\to\infty} f_n$. Then*

$$\lim_{n\to\infty} \mu(f_n) = \mu(f)\ . \tag{B.2.1}$$

Theorem B.2.10 (Fatou's lemma). *Let μ be a measure on $(\mathsf{X}, \mathscr{X})$ and $\{f_n : n \in \mathbb{N}\} \subset \mathbb{F}_+(\mathsf{X})$. Then*

$$\int_{\mathsf{X}} \liminf_{n\to\infty} f_n(x)\mu(\mathrm{d}x) \le \liminf_{n\to\infty} \int_{\mathsf{X}} f_n(x)\mu(\mathrm{d}x)\ . \tag{B.2.2}$$

Theorem B.2.11 (Lebesgue's dominated convergence theorem). *Let μ be a measure on $(\mathsf{X},\mathscr{X})$ and $g \in \mathbb{F}_+(\mathsf{X})$ a μ-integrable function. Let $\{f, f_n : n \in \mathbb{N}\} \subset \mathbb{F}(\mathsf{X})$ be such that $|f_n(x)| \leq g(x)$ for all n and $\lim_{n\to\infty} f_n(x) = f(x)$ for μ-a.e. $x \in \mathsf{X}$. Then all the functions f_n and f are μ-integrable and*

$$\lim_{n\to\infty} \mu(f_n) = \mu(f) \, .$$

Theorem B.2.12 (Egorov's theorem). *Let $\{f_n, \ n \in \mathbb{N}\}$ be a sequence of measurable real-valued functions defined on a measured space $(\mathsf{X},\mathscr{X},\mu)$. Let $A \in \mathscr{X}$ be such that $\mu(A) < \infty$ and $\{f_n, \ n \in \mathbb{N}\}$ converges μ-a.e. to a function f on A. Then for every $\varepsilon > 0$, there exists a set $B \in \mathscr{X}$ such that $\mu(A \setminus B) \leq \varepsilon$ and $\{f_n, \ n \in \mathbb{N}\}$ converges uniformly to f on B.*

Lemma B.2.13 *Let $(\mathsf{X},\mathscr{X},\mu)$ be a probability space and let $p, q \in [1,\infty]$ be such that (p,q) are conjugate. Then for all measurable functions g,*

$$\|g\|_{\mathrm{L}^q(\mu)} = \sup\left\{ \left| \int f g \mathrm{d}\mu \right| \, : \, f \right\} \textit{ is measurable and } \|f\|_{\mathrm{L}^p(\mu)} \leq 1 \, .$$

Proof. (Royden 1988, Proposition 6.5.11). □

B.2.4 Measures on a Metric Space

Let (X,d) be a metric space endowed with its Borel σ-field.

Definition B.2.14 (Topological support of a measure) *The topological support of a measure μ is the smallest closed set whose complement has zero μ-measure.*

Proposition B.2.15 *Let (X,d) be a separable metric space. The support of a measure μ is the set of all points $x \in \mathsf{X}$ for which every open neighborhood $U \in \mathscr{V}_x$ of x has positive measure.*

Proof. See (Parthasarathy 1967, Theorem 2.2.1). □

Definition B.2.16 (Inner and outer regularity) *Let* (X, d) *be a metric space. A measure* μ *on the Borel sigma-field* $\mathscr{X} = \mathscr{B}(\mathsf{X})$ *is called*

(i) inner regular on $A \in \mathscr{B}(\mathsf{X})$ *if* $\mu(A) = \sup\{\mu(F) : F \subset A, F \text{ closed set}\}$*;*
(ii) outer regular on $A \in \mathscr{B}(\mathsf{X})$ *if* $\mu(A) = \inf\{\mu(U) : U \text{ open set } \supset A\}$.
(iii) regular if it is both inner and outer regular on all $A \in \mathscr{B}(\mathsf{X})$.

Theorem B.2.17. *Let* (X, d) *be a metric space. Every bounded measure is inner regular.*

Proof. (Billingsley 1999, Theorem 1.1). □

Corollary B.2.18 *Let* μ, ν *be two measures on* (X, d)*. If* $\mu(f) \leq \nu(f)$ *for all nonnegative bounded uniformly continuous functions* f*, then* $\mu \leq \nu$*. In particular, if* $\mu(f) = \nu(f)$ *for all nonnegative bounded uniformly continuous functions* f*, then* $\mu = \nu$*.*

Proof. (Billingsley 1999, Theorem 1.2). □

Definition B.2.19 *Let* X *be a locally compact separable metric space. A measure* μ *on the Borel sigma-field* $\mathscr{X}$ *is called a Radon measure if* μ *is finite on every compact set. The set of Radon measures on* $\mathscr{X}$ *is denoted by* $\mathbb{M}_{\mathrm{r}}(\mathscr{X})$.

A bounded measure on a locally compact separable metric space is a Radon measure. Lesbesgue measure on $\mathbb{R}^d$ equipped with the Euclidean distance is a Radon measure.

Theorem B.2.20. *A Radon measure on a locally compact separable metric space* X *is regular, and moreover, for every Borel set* A,

$$\mu(A) = \sup\{\mu(K) : K \subset A, K \text{ compact set}\} .$$

Corollary B.2.21 *Let μ and ν be two Radon measures on a locally compact separable metric space X. If $\mu(f) \leq \nu(f)$ for all $f \in \mathrm{C}_c^+(\mathsf{X})$, then $\mu \leq \nu$.*

Corollary B.2.22 *Let μ be a Radon measure on a locally compact separable metric space X. For all p, $1 \leq p < +\infty$, $\mathrm{C}_c(\mathsf{X})$ is dense in $\mathrm{L}^p(\mathsf{X}, \mu)$.*

B.3 Probability

B.3.1 Conditional Expectation

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space.

Lemma B.3.1 *Let X be a nonnegative random variable and $\mathscr{G}$ a sub-σ-field of $\mathscr{F}$. There exists a nonnegative $\mathscr{G}$-measurable random variable Y such that*

$$\mathbb{E}[XZ] = \mathbb{E}[YZ] \tag{B.3.1}$$

for all nonnegative $\mathscr{G}$-measurable random variables Z. If $\mathbb{E}[X] < \infty$, then $\mathbb{E}[Y] < \infty$. If Y' is a nonnegative $\mathscr{G}$-measurable random variable that also satisfies (B.3.1), then $Y = Y'$ $\mathbb{P} -$ a.s.

A random variable with the above properties is called a version of the conditional expectation of X given $\mathscr{G}$, and we write $Y = \mathbb{E}[X | \mathscr{G}]$. Conditional expectations are thus defined up to $\mathbb{P}$-almost sure equality. Hence when writing $\mathbb{E}[X|\mathscr{G}] = Y$, for instance, we always mean that this relation holds $\mathbb{P} -$ a.s., that is, Y is a version of the conditional expectation.

Define $X^- = \max(-X, 0)$.

Definition B.3.2 (Conditional expectation) *Let $\mathscr{G}$ be a sub-σ-field and X a random variable such that $\mathbb{E}[X^-] \wedge \mathbb{E}[X^+] < \infty$. A version of the conditional expectation of X given $\mathscr{G}$ is defined by*

$$\mathbb{E}[X|\mathscr{G}] = \mathbb{E}\left[X^+ \middle| \mathscr{G}\right] - \mathbb{E}\left[X^- \middle| \mathscr{G}\right] .$$

If X is an indicator $\mathbb{1}_A$, we define $\mathbb{P}(A|\mathscr{G}) = \mathbb{E}[\mathbb{1}_A | \mathscr{G}]$.

Lemma B.3.3 *Let $\mathscr{B}$ be a σ-field $\mathscr{B}$ generated by a countable partition of measurable sets $\{B_n,\ n \in \mathbb{N}\}$ with $\mathbb{P}(B_n) > 0$ for all $n \in \mathbb{N}$. Then for every nonnegative random variable X,*

$$\mathbb{E}[X \mid \mathscr{B}] = \sum_{j=0}^{\infty} \frac{\mathbb{E}\left[X \mathbb{1}_{B_j}\right]}{\mathbb{P}(B_j)} \mathbb{1}_{B_j} .$$

Conditional expectation has the same properties as the expectation operator; in particular, it is a positive linear operator and satisfies Jensen's inequality.

Proposition B.3.4 *Let $\mathscr{G}$ be a sub-σ-field of $\mathscr{F}$, and X a random variable such that $\mathbb{E}[X^-] < \infty$. All equalities below hold $\mathbb{P}-$a.s.:*

(i) If $\mathscr{G} = \{\emptyset, \Omega\}$, then $\mathbb{E}[X|\mathscr{G}] = \mathbb{E}[X]$.
(ii) If $\mathscr{H}$ is a sub-σ-field $\mathscr{G}$, then $\mathbb{E}[\mathbb{E}[X|\mathscr{H}]|\mathscr{G}] = \mathbb{E}[X|\mathscr{G}]$.
(iii) If X is independent of $\mathscr{G}$, then

$$\mathbb{E}[X|\mathscr{G}] = \mathbb{E}[X] . \tag{B.3.2}$$

(iv) If X is $\mathscr{G}$-measurable and either $Y \geq 0$ or $\mathbb{E}[|XY|] < \infty$ and $\mathbb{E}[|Y|] < \infty$, then $\mathbb{E}[XY|\mathscr{G}] = X\mathbb{E}[Y|\mathscr{G}]$.
(v) If ϕ is a convex function and $\mathbb{E}[\phi(X)^-] < \infty$, then $\phi(\mathbb{E}[X|\mathscr{G}]) \leq \mathbb{E}[\phi(X)|\mathscr{G}]$.

The monotone convergence theorem and Lebesgue's dominated convergence theorem hold for conditional expectations.

Proposition B.3.5 *Let $\{X_n,\ n \in \mathbb{N}\}$ be a sequence of random variables.*

(i) If $X_n \geq 0$ and $X_n \uparrow X$, then $\lim_{n\to\infty} \mathbb{E}[X_n|\mathscr{G}] = \mathbb{E}[X|\mathscr{G}]$.
(ii) If $|X_n| \leq Z$, $\mathbb{E}[Z|\mathscr{G}] < \infty$, and $\lim_{n\to\infty} X_n = X$, then $\lim_{n\to\infty} \mathbb{E}[|X_n - X||\mathscr{G}] = 0$.

Lemma B.3.6 *Let X be an integrable random variable such that $\mathbb{E}[X] = 0$. Then*

$$\sup_{C\in\mathscr{C}} |\mathbb{E}[X\mathbb{1}_C]| = \mathbb{E}\left[\mathbb{E}[X|\mathscr{C}]^+\right] = \frac{1}{2}\mathbb{E}[|\mathbb{E}[X|\mathscr{C}]|] .$$

Proof. Since $\mathbb{E}[X] = 0$, it is also the case that $\mathbb{E}[\mathbb{E}[X|\mathscr{C}]] = 0$, which yields

$$\mathbb{E}\left[\mathbb{E}[X|\mathscr{C}]^+\right] = \mathbb{E}\left[\mathbb{E}[X|\mathscr{C}]^-\right] = \frac{1}{2}\mathbb{E}[|\mathbb{E}[X|\mathscr{C}]|] .$$

Observe first that

$$\mathbb{E}[X\mathbb{1}_C] = \mathbb{E}[\mathbb{E}[X\,|\,\mathscr{C}]\,\mathbb{1}_C] \le \mathbb{E}\left[(\mathbb{E}[X\,|\,\mathscr{C}])^+\mathbb{1}_C\right] \le \mathbb{E}\left[(\mathbb{E}[X\,|\,\mathscr{C}])^+\right] .$$

We prove similarly that $-\mathbb{E}\left[X\mathbb{1}_C\right] \le \mathbb{E}[(\mathbb{E}[X\,|\,\mathscr{C}])^-]$, and since $\mathbb{E}[(\mathbb{E}[X\,|\,\mathscr{C}])^-] = \mathbb{E}[(\mathbb{E}[X\,|\,\mathscr{C}])^+]$, this proves that $\sup_{C\in\mathscr{C}} |\mathbb{E}\left[X\mathbb{1}_C\right]| \le \mathbb{E}[(\mathbb{E}[X\,|\,\mathscr{C}])^+]$. The equality is seen to hold by taking $C = \{\mathbb{E}[X\,|\,\mathscr{C}] \ge 0\}$. □

B.3.2 Conditional Expectation Given a Random Variable

Let Y be a random variable such that $\mathbb{E}[Y^+] \wedge \mathbb{E}[Y^-] < \infty$ and let $\sigma(X)$ be the sub-σ-field generated by a random variable X. We write $\mathbb{E}[Y|X]$ for $\mathbb{E}[Y|\sigma(X)]$, and we call it the conditional expectation of Y given X. By construction, $\mathbb{E}[Y\,|X]$ is a $\sigma(X)$-measurable random variable. Thus, there exists a real-valued measurable function g on X such that $\mathbb{E}[Y|X] = g(X)\mathbb{P}^X$ – a.s. The function g is defined up to $\mathbb{P}^X$ equivalence: if $\tilde{g}$ satisfies this equality, then $\mathbb{P}(g(X) = \tilde{g}(X)) = 1$. Therefore, we write $\mathbb{E}[Y|X=x]$ for $g(x)$. If Y is an indicator $\mathbb{1}_A$, we write $\mathbb{P}(A\,|X=x)$ for $\mathbb{E}[A\,|X=x]$.

B.3.3 Conditional Distribution

Definition B.3.7 (Regular conditional probability) *Let* $(\Omega,\mathscr{F},\mathbb{P})$ *be a probability space and* $\mathscr{G}$ *a sub-*σ*-field of* $\mathscr{F}$*. A regular version of the conditional probability given* $\mathscr{G}$ *is a map* $\mathbb{P}^{\mathscr{G}} : \Omega\times\mathscr{F} \to [0,1]$ *such that*

(i) for all $F\in\mathscr{F}$, $\mathbb{P}^{\mathscr{G}}(\cdot,F)$ *is* $\mathscr{G}$*-measurable and for every* $\omega\in\Omega$, $\mathbb{P}^{\mathscr{G}}(\omega,\cdot)$ *is a probability on* $\mathscr{F}$*;*
(ii) for all $F\in\mathscr{F}$, $\mathbb{P}^{\mathscr{G}}(\cdot,F) = \mathbb{P}(F\,|\mathscr{G})$ $\mathbb{P}$ – a.s.

Definition B.3.8 (Regular conditional distribution of Y **given** $\mathscr{G}$**)** *Let* $\mathscr{G}$ *be a sub-*σ*-field of* $\mathscr{F}$, $(\mathsf{Y},\mathscr{Y})$ *a measurable space, and* Y *a* Y*-valued random variable. A regular version of the conditional distribution of* Y *given* $\mathscr{G}$ *is a function* $\mathbb{P}^{Y|\mathscr{G}}$: $\Omega\times\mathscr{Y} \to [0,1]$ *such that*

(i) for all $E\in\mathscr{Y}$, $\mathbb{P}^{Y|\mathscr{G}}(\cdot,E)$ *is* $\mathscr{G}$*-measurable and for every* $\omega\in\Omega$, $\mathbb{P}^{Y|\mathscr{G}}(\omega,\cdot)$ *is a probability measure on* $\mathscr{Y}$*;*
(ii) for all $E\in\mathscr{Y}$, $\mathbb{P}^{Y|\mathscr{G}}(\cdot,E) = \mathbb{P}(E\,|\mathscr{G})$, $\mathbb{P}$ – a.s.

A regular version of the conditional distribution exists if Y takes values in a Polish space.

Theorem B.3.9. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, Y a Polish space, $\mathscr{G}$ a sub-σ-field of $\mathscr{F}$, and Y a Y-valued random variable. Then there exists a regular version of the conditional distribution of Y given $\mathscr{G}$.*

When a regular version of a conditional distribution of Y given $\mathscr{G}$ exists, conditional expectations can be written as integrals for each ω: if g is integrable with respect to $\mathbb{P}^Y$, then

$$\mathbb{E}\left[g(Y)\,|\,\mathscr{G}\right] = \int_{\mathsf{Y}} g(y)\mathbb{P}^{Y|\mathscr{G}}(\cdot, \mathrm{d}y) \qquad \mathbb{P}-\text{a.s.}$$

Definition B.3.10 (Regular conditional distribution of Y given X) *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let X and Y be random variables with values in the measurable spaces $(\mathsf{X}, \mathscr{X})$ and $(\mathsf{Y}, \mathscr{Y})$, respectively. Then a regular version of the conditional distribution of Y given $\sigma(X)$ is a function $N : \mathsf{X} \times \mathscr{Y} \to [0,1]$ such that the following conditions hold:*

(i) For all $E \in \mathscr{Y}$, $N(\cdot, E)$ is $\mathscr{X}$-measurable; for all $x \in \mathsf{X}$, $N(x, \cdot)$ is a probability on $(\mathsf{Y}, \mathscr{Y})$.
(ii) For all $E \in \mathscr{Y}$,

$$N(X, E) = \mathbb{P}\left(Y \in E\,|\,X\right) \ \ \mathbb{P}-\text{a.s.} \tag{B.3.3}$$

Theorem B.3.11. *Let X be a random variable with values in $(\mathsf{X}, \mathscr{X})$ and let Y be a random variable with values in a Polish space Y. Then there exists a regular version of the conditional distribution of Y given X.*

If μ and ν are two probabilities on a measurable space $(\mathsf{X}, \mathscr{X})$, we denote by $\mathscr{C}(\mu, \nu)$ the set of all couplings of μ and ν; see Definition 19.1.3.

Lemma B.3.12 (Gluing lemma) *Let $(\mathsf{X}_i, \mathscr{X}_i)$, $i \in \{1,2,3\}$, be three measurable spaces. For $i \in \{1,2,3\}$, let μ_i be a probability measure on X_i and set $\mathsf{X} = \mathsf{X}_1 \times \mathsf{X}_2 \times \mathsf{X}_3$ and $\mathscr{X} = \mathscr{X}_1 \otimes \mathscr{X}_2 \otimes \mathscr{X}_3$. Assume that $(\mathsf{X}_1, \mathscr{X}_1)$ and $(\mathsf{X}_3, \mathscr{X}_3)$ are Polish spaces. Then for every $\gamma_1 \in \mathscr{C}(\mu_1, \mu_2)$ and $\gamma_2 \in \mathscr{C}(\mu_2, \mu_3)$, there exists a probability measure π on $(\mathsf{X}, \mathscr{X})$ such that $\pi(\cdot \times \mathsf{X}_3) = \gamma_1(\cdot)$ and $\pi(\mathsf{X}_1 \times \cdot) = \gamma_2(\cdot)$.*

Proof. Since X_1 and X_3 are Polish spaces, we can apply Theorem B.3.11. There exist two kernels K_1 on $\mathsf{X}_2 \times \mathscr{X}_1$ and K_3 on $\mathsf{X}_2 \times \mathscr{X}_3$ such that for all $A \in \mathscr{X}_1 \otimes \mathscr{X}_2$

and all $B \in \mathscr{X}_2 \otimes \mathscr{X}_3$,

$$\gamma_1(A) = \int_{\mathsf{X}_1 \times \mathsf{X}_2} \mathbb{1}_A(x,y)\mu_2(\mathrm{d}y)K_1(y,\mathrm{d}x) \; , \;\; \gamma_2(B) = \int_{\mathsf{X}_2 \times \mathsf{X}_3} \mathbb{1}_B(y,z)\mu_2(\mathrm{d}y)K_3(y,\mathrm{d}z) \; .$$

Then define the probability measure π on X by

$$\pi(f) = \int_{\mathsf{X}_1 \times \mathsf{X}_2 \times \mathsf{X}_3} f(x,y,z)K_1(y,\mathrm{d}x)K_3(y,\mathrm{d}z)\mu_2(\mathrm{d}y) \; ,$$

for all bounded and measurable functions f. Then for all $(A,B,C) \in \mathscr{X}_1 \times \mathscr{X}_2 \times \mathscr{X}_3$,

$$\pi(A \times B \times \mathsf{X}_3) = \int_B \mu_2(\mathrm{d}y)K_1(y,A)K_3(y,\mathsf{X}_3) = \int_B \mu_2(\mathrm{d}y)K_1(y,A) = \gamma_1(A \times B) \; ,$$
$$\pi(\mathsf{X}_1 \times B \times C) = \int_B \mu_2(\mathrm{d}y)K_1(y,\mathsf{X}_1)K_3(y,C) = \int_B \mu_2(\mathrm{d}y)K_3(y,C) = \gamma_3(B \times C) \; .$$

□

Remark B.3.13. An equivalent formulation of the gluing lemma, Lemma B.3.12, is that when X_1 and X_3 are Polish spaces, then for every $\mu_i \in \mathbb{M}_1(\mathscr{X}_i)$, $i \in \{1,2,3\}$, $\gamma_1 \in \mathscr{C}(\mu_1,\mu_2)$, and $\gamma_2 \in \mathscr{C}(\mu_2,\mu_3)$, there exist a probability space $(\Omega,\mathscr{F},\mathbb{P})$ and X_i-valued random variables Z_i, $i \in \{1,2,3\}$, such that $\mathscr{L}_{\mathbb{P}}(Z_1,Z_2) = \gamma_1$, $\mathscr{L}_{\mathbb{P}}(Z_2,Z_3) = \gamma_2$. ▲

B.3.4 Conditional Independence

Definition B.3.14 (Conditional independence) *Let $(\Omega,\mathscr{F},\mathbb{P})$ be a probability space, and let $\mathscr{G}$ and $\mathscr{G}_1,\ldots,\mathscr{G}_n$, be sub-$\sigma$-fields of $\mathscr{F}$. Then $\mathscr{G}_1,\ldots,\mathscr{G}_n$ are said to be conditionally independent given $\mathscr{G}$ if for all bounded random variables $X_1,\ldots,X_n$ measurable with respect to $\mathscr{G}_1,\ldots,\mathscr{G}_n$, respectively,*

$$\mathbb{E}\left[X_1 \cdots X_n \middle| \mathscr{G}\right] = \prod_{i=1}^n \mathbb{E}\left[X_i \middle| \mathscr{G}\right] \; .$$

If $Y_1,\ldots,Y_n$ and Z are random variables, then $Y_1,\ldots,Y_n$ are said to be conditionally independent given Z if the sub-σ-fields $\sigma(Y_1),\ldots,\sigma(Y_n)$ are $\mathbb{P}$-conditionally independent given $\sigma(Z)$.

Proposition B.3.15 *Let $(\Omega,\mathscr{F},\mathbb{P})$ be a probability space and let $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$ be sub-σ-fields of $\mathscr{F}$. Then $\mathscr{A}$ and $\mathscr{B}$ are $\mathbb{P}$-conditionally independent given $\mathscr{C}$*

if and only if for every bounded $\mathscr{A}$-measurable random variable X,

$$\mathbb{E}\,[X|\mathscr{B}\vee\mathscr{C}] = \mathbb{E}\,[X|\mathscr{C}]\ , \tag{B.3.4}$$

where $\mathscr{B}\vee\mathscr{C}$ denotes the σ-field generated by $\mathscr{B}\cup\mathscr{C}$.

Proposition B.3.15 can be used as an alternative definition of conditional independence: it means that $\mathscr{A}$ and $\mathscr{B}$ are conditionally independent given $\mathscr{C}$ if for all $\mathscr{A}$-measurable nonnegative random variables X, there exists a version of the conditional expectation $\mathbb{E}\,[X|\mathscr{B}\vee\mathscr{C}]$ that is $\mathscr{C}$-measurable.

Lemma B.3.16 *Let $\mathscr{A},\mathscr{B}$ be conditionally independent given $\mathscr{C}$. For every random variable $X \in \mathrm{L}^1(\mathscr{A})$ such that $\mathbb{E}\,[X] = 0$,*

$$\sup_{B\in\mathscr{B}\vee\mathscr{C}} |\mathbb{E}\,[X\mathbb{1}_B]| = \sup_{B\in\mathscr{C}} |\mathbb{E}\,[X\mathbb{1}_B]| = \frac{1}{2}\mathbb{E}\,[|\mathbb{E}\,[X\,|\,\mathscr{C}]\,|]\ .$$

Proof. We already know that the second equality holds by Lemma B.3.6. By the conditional independence assumption and Proposition B.3.15, $\mathbb{E}\,[X\,|\,\mathscr{B}\vee\mathscr{C}] = \mathbb{E}\,[X\,|\,\mathscr{C}]\,\mathbb{P}-$a.s. Thus applying Lemma B.3.6 yields

$$\begin{aligned}\sup_{B\in\mathscr{B}\vee\mathscr{C}} |\mathbb{E}\,[X\mathbb{1}_B]\,| &= \sup_{B\in\mathscr{B}\vee\mathscr{C}} |\mathbb{E}\,[\mathbb{E}\,[X\,|\,\mathscr{B}\vee\mathscr{C}]\,\mathbb{1}_B]\,| = \sup_{B\in\mathscr{B}\vee\mathscr{C}} |\mathbb{E}\,[\mathbb{E}\,[X\,|\,\mathscr{C}]\,\mathbb{1}_B]\,| \\ &= \frac{1}{2}\mathbb{E}\,[|\mathbb{E}\,[\mathbb{E}\,[X\,|\,\mathscr{C}]\,|\,\mathscr{B}\vee\mathscr{C}]|] = \frac{1}{2}\mathbb{E}\,[|\mathbb{E}\,[X\,|\,\mathscr{C}]|]\ .\end{aligned}$$

□

B.3.5 Stochastic Processes

Let $(\Omega,\mathscr{F},\mathbb{P})$ be a probability space and $(\mathsf{X},\mathscr{X})$ a measurable space. Let T be a set and $\{X_t,\ t\in T\}$ an X-valued stochastic process, that is, a collection of X-valued random variables indexed by T. For every finite subset $S\subset T$, let μ_S be the distribution of $(X_s, s\in S)$. Denote by $\mathscr{S}$ the set of all finite subsets of T. The set of probability measures $\{\mu_S,\ S\in\mathscr{S}\}$ is called the set of finite-dimensional distributions of the process $\{X_t,\ t\in T\}$. For $S\subset S'\subset T$, the canonical projection $p_{S,S'}$ of X^S on $\mathsf{X}^{S'}$ is defined by $p_{S,S'}(x_s, s\in S) = (x_s, s\in S')$. The finite-dimensional distributions satisfy the following consistency conditions:

$$\mu_S = \mu_{S'}\circ p_{S,S'}^{-1}\ . \tag{B.3.5}$$

Conversely, let $\{\mu_S,\ S\in\mathscr{S}\}$ be a family of probability measures such that for all $S\in\mathscr{S}$, μ_S is a probability on $\mathscr{X}^{\otimes S}$. We say that this family is consistent if it satisfies (B.3.5). Introduce the canonical space $\Omega = \mathsf{X}^T$, whose elements are denoted by

$\omega = (\omega_t, t \in T)$, and the coordinate process $\{X_t,\ t \in T\}$ defined by

$$X_t(\omega) = \omega_t\ ,\ \ t \in T\ .$$

The product space Ω is endowed with the product σ-field $\mathscr{F} = \mathscr{X}^{\otimes T}$.

Theorem B.3.17 (Kolmogorov). *Assume that* X *is a Polish space. Let* $\{\mu_S, S \subset T, S \text{ finite}\}$ *be a consistent family of measures. Then there exists a unique probability measure on the canonical space under which the family of finite-dimensional distributions of the canonical process process is* $\{\mu_S,\ S \in \mathscr{S}\}$.

Proof. (Kallenberg 2002, Theorem 5.16). □

Theorem B.3.18 (Skorohod's representation theorem). *Let* $\{\xi_n,\ n \in \mathbb{N}\}$ *be a sequence of random elements in a complete separable metric space* (S, ρ) *such that* $\xi_n \stackrel{w}{\Rightarrow} \xi_0$. *Then on a suitable probability space* $(\hat{\Omega}, \hat{\mathscr{F}}, \hat{\mathbb{P}})$, *there exists a sequence* $\{\eta_n,\ n \in \mathbb{N}\}$ *of random elements such that* $\mathscr{L}_{\hat{\mathbb{P}}}(\eta_n) = \mathscr{L}_{\mathbb{P}}(\xi_n)$ *for all* $n \geq 0$ *and* $\eta_n \to \eta_0$ $\hat{\mathbb{P}}$*-almost surely.*

Proof. (Kallenberg 2002, Theorem 3.30). □

Appendix C
Weak Convergence

Throughout this chapter, (X, d) is a metric space, and all measures are defined on its Borel σ-field, that is, the smallest σ-field containing the open sets. Additional properties of the metric space (completeness, separability, local compactness, etc.) will be made precise for each result as needed. The essential reference is Billingsley (1999).

Definition C.0.1 (Weak convergence) *Let* (X, d) *be a metric space. Let* $\{\mu, \mu_n, n \in \mathbb{N}\} \subset \mathbb{M}_1(\mathsf{X})$. *The sequence* $\{\mu_n,\ n \in \mathbb{N}\}$ *converges weakly to* μ *if for all* $f \in \mathrm{C}_b(\mathsf{X})$, $\lim_{n\to\infty} \mu_n(f) = \mu(f)$. *This is denoted by* $\mu_n \overset{w}{\Rightarrow} \mu$.

C.1 Convergence on Locally Compact Metric Spaces

In this section, (X, d) is assumed to be a locally compact separable metric space space.

Definition C.1.1 (Weak* convergence) *Let* (X, d) *be a locally compact separable metric space. A sequence of bounded measures* $\{\mu_n,\ n \in \mathbb{N}\}$ *converges weakly∗ to* $\mu \in \mathbb{M}_b(\mathscr{X})$, *written* $\mu_n \overset{w^*}{\Rightarrow} \mu$, *if* $\lim_{n\to\infty} \mu_n(f) = \mu(f)$ *for all* $f \in \mathrm{C}_0(\mathsf{X})$.

Proposition C.1.2 *Let* X *be a locally compact separable metric space. If* $\mu_n \overset{w^*}{\Rightarrow} \mu$ *on* X, *then* $\mu(f) \le \liminf_{n\to\infty} \mu_n(f)$ *for every* $f \in \mathrm{C}_b(\mathsf{X})$ *and* $\limsup_{n\to\infty} \mu_n(K) \le \mu(K)$ *for every compact set* K.

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1

One fundamental difference between weak and weak* convergence is that the latter does not preserve the total mass. However, the previous result shows that if the sequence of measures $\{\mu_n,\ n \in \mathbb{N}\}$ converges weakly* to μ, then $\mu(\mathsf{X}) \leq \liminf_{n\to\infty} \mu_n(\mathsf{X})$. Thus in particular, if a sequence of probability measures weakly* converges to a bounded measure μ, then $\mu(\mathsf{X}) \leq 1$. It may even happen that $\mu = 0$.

Proposition C.1.3 *Let (X, d) be a locally compact separable metric space and let $\{\mu_n,\ n \in \mathbb{N}\}$ be a sequence in $\mathbb{M}_b(\mathscr{X})$ such that $\sup_{n\in\mathbb{N}} \mu_n(\mathsf{X}) \leq B < \infty$. Then there exist a subsequence $\{n_k,\ k \in \mathbb{N}\}$ and $\mu \in \mathbb{M}_+(\mathscr{X})$ such that $\mu(\mathsf{X}) \leq B$ and $\{\mu_{n_k}, k \in \mathbb{N}\}$ converges weakly$*$ to μ.*

C.2 Tightness

Definition C.2.1 *Let (X, d) be a metric space. Let Γ be a subset of $\mathbb{M}_1(\mathsf{X})$.*

(i) The set Γ is said to be tight if for all $\varepsilon > 0$, there exists a compact set $K \subset \mathsf{X}$ such that for all $\mu \in \Gamma$, $\mu(K) \geq 1 - \varepsilon$.
(ii) The set Γ is said to be relatively compact if every sequence of elements in Γ contains a weakly convergent subsequence, or equivalently, if $\overline{\Gamma}$ is compact.

Theorem C.2.2 (Prohorov). *Let (X, d) be a metric space. If $\Gamma \subset \mathbb{M}_1(\mathsf{X})$ is tight, then it is relatively compact. If (X, d) is separable and complete, then the converse is true.*

Proof. (Billingsley 1999, Theorems 5.1 and 5.2). □

As a consequence, a finite family of probability measures on a complete separable metric space is tight.

Corollary C.2.3 *Let (X, d) be a complete separable metric space and $\mu \in \mathbb{M}_1(\mathsf{X})$. Then for all $A \in \mathscr{B}(\mathsf{X})$,*

$$\mu(A) = \sup\{\mu(K) : K \text{ compact set} \subset A\} .$$

Lemma C.2.4 *Let* (X, d) *be a metric space. Let* $\{\nu_n,\ n \in \mathbb{N}\}$ *be a sequence in* $\mathbb{M}_1(\mathscr{X})$ *and* V *a nonnegative function in* $\mathrm{C}(\mathsf{X})$*. If the level sets* $\{V \le c\}$ *are compact for all* $c > 0$ *and* $\sup_{n \ge 1} \nu_n(V) < \infty$*, then the sequence* $\{\nu_n,\ n \in \mathbb{N}\}$ *is tight.*

Proof. Set $M = \sup_{n \ge 1} \nu_n(V)$. By Markov's inequality, we have, for $\varepsilon > 0$,

$$\nu_n(\{V > M/\varepsilon\}) \le (\varepsilon/M)\nu_n(V) \le \varepsilon .$$

By assumption, $\{V \le M/\varepsilon\}$ is compact; thus $\{\nu_n,\ n \in \mathbb{N}\}$ is tight. □

Lemma C.2.5 *Let* (X, d) *be a metric space. Let* $\Gamma \subset \mathbb{M}_1(\mathscr{X}^{\otimes 2})$*. For* $\lambda \in \Gamma$ *and* $A \in \mathscr{X}$*, define* $\lambda_1(A) = \lambda(A \times \mathsf{X})$ *and* $\lambda_2(A) = \lambda(\mathsf{X} \times A)$*. If* $\Gamma_1 = \{\lambda_1 : \lambda \in \Gamma\}$ *and* $\Gamma_2 = \{\lambda_2 : \lambda \in \Gamma\}$ *are tight in* $\mathbb{M}_1(\mathsf{X})$*, then* Γ *is tight.*

Proof. Simply observe that

$$\mathsf{X}^2 \setminus (K_1 \times K_2) \subset ((\mathsf{X} \setminus K_1) \times \mathsf{X}) \cup (\mathsf{X} \times (\mathsf{X} \setminus K_2)) .$$

Moreover, if K_1 and K_2 are compact subsets of X, then $K_1 \times K_2$ is a compact subset of X^2. These two facts yield the result. □

For $A \subset \mathsf{X}$ and $\alpha > 0$, we define $A^\alpha = \{x \in \mathsf{X},\ d(x, A) < \alpha\}$.

Definition C.2.6 *Let* (X, d) *be a metric space. The Prokhorov metric* $\boldsymbol{\rho}_{\mathsf{d}}$ *is defined on* $\mathbb{M}_1(\mathscr{X})$ *by*

$$\boldsymbol{\rho}_{\mathsf{d}}(\lambda, \mu) = \inf\{\alpha > 0 : \lambda(F) \le \mu(F^\alpha) + \alpha \text{ for all closed } F\} . \tag{C.2.1}$$

Theorem C.2.7. *Let* (X, d) *be a separable metric space. Then* $(\mathbb{M}_1(\mathsf{X}), \boldsymbol{\rho}_{\mathsf{d}})$ *is separable, and* $\boldsymbol{\rho}_{\mathsf{d}}$ *metrizes the weak convergence. If, moreover,* (X, d) *is complete, then* $(\mathbb{M}_1(\mathsf{X}), \boldsymbol{\rho}_{\mathsf{d}})$ *is complete.*

Proof. (Dudley 2002, Theorem 11.3.3) and (Billingsley 1999, Theorem 6.8). □

Lemma C.2.8 *Let* (X, d) *be a metric space. Let* $\mu, \nu \in \mathbb{M}_1(\mathsf{X})$*. If there exist random variables* X*,* Y *defined on a probability space* $(\Omega, \mathscr{F}, \mathbb{P})$ *such that* $\mathscr{L}_{\mathbb{P}}(X) = \mu$*,* $\mathscr{L}_{\mathbb{P}}(Y) = \nu$*, and* $\alpha > 0$ *such that* $\mathbb{P}(d(X, Y) > \alpha) \le \alpha$*, then* $\boldsymbol{\rho}_{\mathsf{d}}(\mu, \nu) \le \alpha$*.*

Proof. For every closed set F,

$$\mu(F) = \mathbb{P}(X \in F) \le \mathbb{P}(X \in F, d(X,Y) < \alpha) + \mathbb{P}(d(X,Y) \ge \alpha) \le \mathbb{P}(Y \in F^{\alpha}) + \alpha \,.$$

Thus $\boldsymbol{\rho}_{\mathsf{d}}(\mu, \nu) \le \alpha$. $\square$

Appendix D
Total and V-Total Variation Distances

Given the importance of total variation in this book, we provide an almost self-contained introduction. Unlike the other chapters in this appendix, we will establish most of the results, except the most classical ones.

D.1 Signed Measures

Definition D.1.1 (Finite signed measure) *A finite signed measure on* $(\mathsf{X}, \mathscr{X})$ *is a function* $\nu : \mathscr{X} \to \mathbb{R}$ *such that if* $\{A_n,\ n \in \mathbb{N}\} \subset \mathscr{X}$ *is a sequence of mutually disjoints sets, then* $\sum_{n=1}^{\infty} |\nu(A_n)| < \infty$ *and* $\nu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \nu(A_n)$. *The set of finite signed measures on* $(\mathsf{X}, \mathscr{X})$ *is denoted by* $\mathbb{M}_{\pm}(\mathscr{X})$.

Definition D.1.2 (Singular measures) *Two measures* μ, ν *on a measurable space* $(\mathsf{X}, \mathscr{X})$ *are singular if there exists a set* A *in* $\mathscr{X}$ *such that* $\mu(A^c) = \nu(A) = 0$.

Theorem D.1.3 (Hahn–Jordan). *Let* ξ *be a finite signed measure. There exists a unique pair of finite singular measures* (ξ^+, ξ^-) *such that* $\xi = \xi^+ - \xi^-$.

Proof. (Rudin 1987, Theorem 6.14). □

The pair (ξ^+, ξ^-) is called the Jordan decomposition of the signed measure ξ. The finite (positive) measure $|\xi| = \xi^+ + \xi^-$ is called the total variation of ξ. It is the smallest measure ν such that for all $A \in \mathscr{X}$, $\nu(A) \geq |\xi(A)|$. A set S such that

R. Douc et al., *Markov Chains*, Springer Series in Operations Research
and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1

$\xi^+(S^c) = \xi^-(S) = 0$ is called a Jordan set for ξ. If S and S' are two Jordan sets for ξ, then $|\xi|(S\Delta S') = 0$.

Lemma D.1.4 *Let μ be a signed measure on $\mathscr{X}$. Then for all $B \in \mathscr{X}$,*

$$\mu^+(B) = \sup_{A\in\mathscr{X}} \mu(B\cap A) . \tag{D.1.1}$$

If $\mathscr{X}$ is generated by an algebra $\mathscr{A}$, then

$$\mu^+(B) = \sup_{A\in\mathscr{A}} \mu(B\cap A) . \tag{D.1.2}$$

Proof. Let S be a Jordan set for μ. Then for all $B \in \mathscr{X}$, $\mu^+(B) = \mu(B\cap S)$. Moreover, for all $A \in \mathscr{X}$,

$$\mu(B\cap A) \le \mu^+(B\cap A) \le \mu^+(B) .$$

This proves (D.1.1).

Assume now that $\mathscr{X} = \sigma(\mathscr{A})$ and let $B \in \mathscr{X}$. By (D.1.1), for all $\varepsilon > 0$, there exists $C \in \mathscr{X}$ such that $\mu^+(B) \le \mu(B\cap C) + \varepsilon$. The approximation lemma, Lemma B.2.5, applied to the (positive) measure $\mu^+ + \mu^-$ implies that there exists $A \in \mathscr{A}$ such that $\mu^+(C\Delta A) + \mu^-(C\Delta A) \le \varepsilon$. Then (D.1.2) follows from

$$\mu^+(B) \le \mu(B\cap C) + \varepsilon \le \mu(B\cap A) + 2\varepsilon .$$

$\square$

Lemma D.1.5 *Let X be a set and $\mathscr{B}$ a countably generated σ-field on X. There exists a sequence $\{B_n,\ n \in \mathbb{N}\} \subset \mathscr{B}$ such that for all signed measures on $\mathscr{B}$,*

$$\sup_{B\in\mathscr{B}} |\mu(B)| = \sup_{n\in\mathbb{N}} |\mu(B_n)| , \qquad \sup_{B\in\mathscr{B}} \mu(B) = \sup_{n\in\mathbb{N}} \mu(B_n) . \tag{D.1.3}$$

Proof. Since $\mathscr{B}$ is countably generated, there exists a countable algebra $\mathscr{A} = \{B_n,\ n \in \mathbb{N}\}$ such that $\sigma(\mathscr{A}) = \mathscr{B}$. Let μ be a signed measure and S a Jordan set for μ. Then $\sup_{B\in\mathscr{B}} \mu(B) = \mu(S)$, and by Lemma B.2.5, there exists $A \in \mathscr{A}$ such that $\mu(S\setminus A) \le \varepsilon$. This yields $\mu(A) \ge \mu(S) - \varepsilon$, and therefore the second statement in (D.1.3) holds. Since $\sup_{B\in\mathscr{B}} |\mu(B)| = \max(\mu(S), -\mu(S^c))$, the first statement in (D.1.3) is proved similarly. $\square$

D.2 Total Variation Distance

Proposition D.2.1 *A set function ξ is a signed measure if and only if there exist $\mu \in \mathbb{M}_+(\mathscr{X})$ and $h \in \mathrm{L}^1(\mu)$ such that $\xi = h \cdot \mu$. Then $S = \{h \geq 0\}$ is a Jordan set for ξ, $\xi^+ = h^+ \cdot \mu$, $\xi^- = h^- \cdot \mu$, and $|\xi| = |h| \cdot \mu$.*

Proof. The direct implication is straightforward. Let us now establish the converse. Let ξ be a signed measure, (ξ^+, ξ^-) its Jordan decomposition, and S a Jordan set. We have for all $A \in \mathscr{X}$,

$$\xi^+(A) = \xi(A \cap S), \quad \xi^-(A) = -\xi(A \cap S^c) .$$

Then for $A \in \mathscr{X}$,

$$\xi(A) = \xi(A \cap S) - \xi(A \cap S^c) = |\xi|(A \cap S) - |\xi|(A \cap S^c) = \int_A (\mathbb{1}_S - \mathbb{1}_{S^c}) \mathrm{d}|\xi| ,$$

showing that $\xi = (\mathbb{1}_S - \mathbb{1}_{S^c}) \cdot |\xi|$ and concluding the proof. □

Definition D.2.2 (Total variation distance) *Let ξ be a finite signed measure on $(\mathsf{X}, \mathscr{X})$ with Jordan decomposition (ξ^+, ξ^-). The total variation norm of ξ is defined by*

$$\|\xi\|_{\mathrm{TV}} = |\xi|(\mathsf{X}) .$$

The total variation distance between two probability measures $\xi, \xi' \in \mathbb{M}_1(\mathsf{X})$ is defined by

$$\mathrm{d_{TV}}(\xi, \xi') = \frac{1}{2} \|\xi - \xi'\|_{\mathrm{TV}} .$$

Note that $\mathrm{d_{TV}}(\xi, \xi') = (\xi - \xi')(S)$, where S is a Jordan set for $\xi - \xi'$. This definition entails straightforwardly the following equivalent one.

Proposition D.2.3 *Let ξ be a finite signed measure on $(\mathsf{X}, \mathscr{X})$,*

$$\|\xi\|_{\mathrm{TV}} = \sup \sum_{i=1}^{I} |\xi(A_i)| , \qquad \text{(D.2.1)}$$

where the supremum is taken over all finite measurable partitions $\{A_1, \ldots, A_I\}$ of X.

Proof. Let S be a Jordan set for μ. Then $\|\mu\|_{\mathrm{TV}} = \mu(S) - \mu(S^c)$. Thus $\|\xi\|_{\mathrm{TV}} \leq \sup \sum_{i=1}^{I} |\mu(A_i)|$. Conversely,

$$\sum_{i=1}^{I} |\mu(A_i)| = \sum_{i=1}^{I} \mu(A_i \cap S) - \sum_{i=1}^{I} \mu(A_i \cap S^c) \leq \|\xi\|_{\mathrm{TV}} \ .$$

□

Let $\mathbb{M}_0(\mathscr{X})$ be the set of finite signed measures ξ such that $\xi(\mathsf{X}) = 0$. We now give equivalent characterizations of the total variation norm for signed measures. Let the oscillation $\operatorname{osc}(f)$ of a bounded function f be defined by

$$\operatorname{osc}(f) = \sup_{x,x' \in \mathsf{X}} |f(x) - f(x')| = 2 \inf_{c \in \mathbb{R}} |f - c|_\infty \ .$$

Proposition D.2.4 *For* $\xi \in \mathbb{M}_s(\mathscr{X})$,

$$\|\xi\|_{\mathrm{TV}} = \sup\{\xi(f) \, : \, f \in \mathbb{F}_b(\mathsf{X}), |f|_\infty \leq 1\} \ . \tag{D.2.2}$$

If, moreover, $\xi \in \mathbb{M}_0(\mathscr{X})$, *then*

$$\|\xi\|_{\mathrm{TV}} = 2 \sup\{\xi(f) \, : \, f \in \mathbb{F}_b(\mathsf{X}), \ \operatorname{osc}(f) \leq 1\} \ . \tag{D.2.3}$$

Proof. By Proposition D.2.1, $\xi = h \cdot \mu$ with $h \in \mathrm{L}^1(\mu)$ and $\mu \in \mathbb{M}_+(\mathscr{X})$. The proof of (D.2.2) follows from the identity

$$\|\xi\|_{\mathrm{TV}} = \int_{\mathsf{X}} |h| \mathrm{d}\mu = \int_{\mathsf{X}} \{\mathbb{1}_{h>0} - \mathbb{1}_{h<0}\} h \mathrm{d}\mu = \sup_{|f| \leq 1} \int f h \mathrm{d}\mu \ .$$

Now let $\xi \in \mathbb{M}_0(\mathscr{X})$. Then $\xi(f) = \xi(f + c)$ for all $c \in \mathbb{R}$, and thus for all $c \in \mathbb{R}$,

$$|\xi(f)| = |\xi(f - c)| \leq \|\xi\|_{\mathrm{TV}} |f - c|_\infty \ .$$

Since this inequality is valid for all $c \in \mathbb{R}$, this yields

$$|\xi(f)| \leq \|\xi\|_{\mathrm{TV}} \inf_{c \in \mathbb{R}} |f - c|_\infty = \frac{1}{2} \|\xi\|_{\mathrm{TV}} \operatorname{osc}(f) \ . \tag{D.2.4}$$

Conversely, if we set $f = (1/2)(\mathbb{1}_S - \mathbb{1}_{S^c})$, where S is a Jordan set for ξ, then $\operatorname{osc}(f) = 1$ and

$$\xi(f) = \frac{1}{2}\{\xi^+(S) + \xi^-(S^c)\} = \frac{1}{2}\{\xi^+(\mathsf{X}) + \xi^-(\mathsf{X})\} = \frac{1}{2} \|\xi\|_{\mathrm{TV}} \ .$$

Combining this with (D.2.4) proves (D.2.3). □

Corollary D.2.5 *If $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$, then $\xi - \xi' \in \mathbb{M}_0(\mathscr{X})$ and for all $f \in \mathbb{F}_b(\mathsf{X})$,*

$$|\xi(f) - \xi'(f)| \leq \mathrm{d}_{\mathrm{TV}}(\xi, \xi') \operatorname{osc}(f) . \tag{D.2.5}$$

In particular, for every $A \in \mathscr{X}$, $|\xi(A) - \xi'(A)| \leq \mathrm{d}_{\mathrm{TV}}(\xi, \xi')$.

Proposition D.2.6 *If X is a metric space and $\mathscr{X}$ is its Borel σ-field, the convergence in total variation of a sequence of probability measures on $(\mathsf{X}, \mathscr{X})$ implies its weak convergence.*

Proof. Convergence in total variation implies that $\lim_{n\to\infty} \xi_n(h) = \xi(h)$ for all bounded measurable functions h. This is a stronger property than weak convergence, which requires this convergence only for bounded continuous functions h defined on X. □

Theorem D.2.7. *The space $(\mathbb{M}_{\mathrm{s}}(\mathscr{X}), \|\cdot\|_{\mathrm{TV}})$ is a Banach space.*

Proof. Let $\{\xi_n,\ n \in \mathbb{N}\}$ be a Cauchy sequence in $\mathbb{M}_{\mathrm{s}}(\mathscr{X})$. Define

$$\lambda = \sum_{n=0}^{\infty} \frac{1}{2^n} |\xi_n| ,$$

which is a measure, as a limit of an increasing sequence of measures. By construction, $|\xi_n| \ll \lambda$ for all $n \in \mathbb{N}$. Therefore, there exist functions $f_n \in \mathrm{L}^1(\lambda)$ such that $\xi_n = f_n.\lambda$ and $\|\xi_n - \xi_m\|_{\mathrm{TV}} = \int |f_n - f_m| \mathrm{d}\lambda$. This implies that $\{f_n,\ n \in \mathbb{N}\}$ is a Cauchy sequence in $\mathrm{L}^1(\lambda)$, which is complete. Thus there exists $f \in \mathrm{L}^1(\lambda)$ such that $f_n \to f$ in $\mathrm{L}^1(\lambda)$. Setting $\xi = f.\lambda$, we obtain that $\xi \in \mathbb{M}_{\mathrm{s}}(\mathscr{X})$ and $\lim_{n\to\infty} \|\xi_n - \xi\|_{\mathrm{TV}} = \lim_{n\to\infty} \int |f_n - f| \mathrm{d}\lambda = 0$. □

We now define and characterize the minimum of two measures.

Proposition D.2.8 *Let $\xi, \xi' \in \mathbb{M}_+(\mathscr{X})$ be two measures.*

(i) The set of measures η such that $\eta \leq \xi$ and $\eta \leq \xi'$ admits a maximal element, denoted by $\xi \wedge \xi'$ and called the minimum of ξ and ξ'.
(ii) The measures $\xi - \xi \wedge \xi'$ and $\xi' - \xi \wedge \xi'$ are positive and mutually singular.
(iii) Conversely, if there exist measures η, ν, and ν' such that $\xi = \eta + \nu$, $\xi' = \eta + \nu'$, and ν and ν' are mutually singular, then $\eta = \xi \wedge \xi'$.

(iv) If $\xi = f \cdot \mu$ and $\xi' = f' \cdot \mu$, then $\xi \wedge \xi' = (f \wedge f') \cdot \mu$.
(v) If $\xi(\mathsf{X}) \vee \xi'(\mathsf{X}) < \infty$, then $(\xi - \xi')^+ = \xi - \xi \wedge \xi'$ and

$$|\xi - \xi'| = \xi + \xi' - 2\xi \wedge \xi' .$$

Proof. Let μ be a σ-finite measure such that $\xi = f \cdot \mu$ and $\xi' = f' \cdot \mu$ (take, for instance, $\mu = \xi + \xi'$). Let $\rho = (f \wedge f') \cdot \mu$. If $\eta \le \xi$ and $\eta \le \xi'$, then

$$\begin{aligned} \eta(A) &= \eta(A \cap \{f \ge f'\}) + \eta(A \cap \{f < f'\}) \\ &\le \xi'(A \cap \{f \ge f'\}) + \xi(A \cap \{f < f'\}) \\ &= \rho(A \cap \{f \ge f'\}) + \rho(A \cap \{f < f'\}) = \rho(A) . \end{aligned}$$

This proves (i), (ii), and (iv). Let now η, ν, and ν' be as in (iii) and let g, h, and h' be their densities with respect to μ. Then $f = g + h$, $f' = g + h'$, and $hh' = 0\ \ \mu -$a.s., since ν and ν' are mutually singular. This implies that $g = f \wedge f'\ \mu -$a.s., and hence $\eta = \xi \wedge \xi'$. This proves (iii). Finally, using the identities $(p - q)^+ = p - p \wedge q$ and $|p - q| = p + q - 2p \wedge q$ ($p, q \ge 0$), we obtain, for all $A \in \mathscr{X}$,

$$\begin{aligned} (\xi - \xi')^+(A) &= \int_A (f - f')^+ \mathrm{d}\mu = \int_A f \mathrm{d}\mu - \int_A f \wedge f' \mathrm{d}\mu = (\xi(A) - \xi \wedge \xi')(A) \ , \\ |\xi - \xi'|(A) &= \int_A |f - f'| \mathrm{d}\mu = \int_A f \mathrm{d}\mu + \int_A f' \mathrm{d}\mu - 2\int_A f \wedge f' \mathrm{d}\mu \\ &= \xi(A) + \xi'(A) - 2(\xi \wedge \xi')(A) . \end{aligned}$$

This yields (v). □

Remark D.2.9 *It must be noted that $\xi \wedge \xi'$ is not defined by $(\xi \wedge \xi')(A) = \xi(A) \wedge \xi'(A)$, since that would not even define an additive set function.*

Lemma D.2.10 *Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$. Then for all $\xi, \xi' \in \mathbb{M}_1(\mathscr{X})$,*

$$\mathrm{d_{TV}}(\xi P, \xi' P) \le \mathrm{d_{TV}}(\xi, \xi') .$$

Proof. Note that if $h \in \mathbb{F}(\mathsf{b})X$, then $|Ph|_\infty \le |h|_\infty$. Therefore,

$$\begin{aligned} \mathrm{d_{TV}}(\xi P, \xi' P) &= (1/2) \sup_{|h| \le 1} |\xi Ph - \xi' Ph| \\ &= (1/2) \sup_{|h| \le 1} |\xi(Ph) - \xi'(Ph)| \le \mathrm{d_{TV}}(\xi, \xi') . \end{aligned}$$

□

D.3 V-Total Variation

Let $(\mathsf{X}, \mathscr{X})$ be a measurable space. In this section, we consider a function $V \in \mathbb{F}(\mathsf{X})$ taking values in $[1, \infty]$. We define $D_V = \{x \in \mathsf{X} : V(x) < \infty\}$.

Definition D.3.1 (V-norm) *The space of finite signed measures ξ such that $|\xi|(V) < \infty$ is denoted by $\mathbb{M}_V(\mathscr{X})$.*

(i) The V-norm of a measure $\xi \in \mathbb{M}_V(\mathscr{X})$ is

$$\|\xi\|_V = |\xi|(V) .$$

(ii) The V-norm of a function $f \in \mathbb{F}(\mathsf{X})$ is

$$|f|_V = \sup_{x \in D_V} \frac{|f(x)|}{V(x)} .$$

(iii) The V-oscillation of a function $f \in \mathbb{F}(\mathsf{X})$ is

$$\operatorname{osc}_V(f) := \sup_{(x,x') \in D_V \times D_V} \frac{|f(x) - f(x')|}{V(x) + V(x')} . \tag{D.3.1}$$

Of course, when $V = \mathbb{1}_{\mathsf{X}}$, then $\|\xi\|_{\mathbb{1}_{\mathsf{X}}} = \|\xi\|_{\mathrm{TV}}$ by (D.2.2). It also holds that $\|\xi\|_V = \|V \cdot \xi\|_{\mathrm{TV}}$. We now give characterizations of the V-norm similar to the characterizations of the TV-norm provided in Proposition D.2.4.

Theorem D.3.2. *For $\xi \in \mathbb{M}_V(\mathscr{X})$,*

$$\|\xi\|_V = \sup\{\xi(f) : f \in \mathbb{F}_b(\mathsf{X}), |f|_V \le 1\} . \tag{D.3.2}$$

Let $\xi \in \mathbb{M}_0(\mathscr{X}) \cap \mathbb{M}_V(\mathscr{X})$. Then

$$\|\xi\|_V = \sup\{\xi(f) : \operatorname{osc}_V(f) \le 1\} . \tag{D.3.3}$$

Proof. Equation (D.3.2) follows from

$$\|\xi\|_V = \|V \cdot \xi\|_{\mathrm{TV}} = \sup_{|f|_\infty \le 1} \xi(Vf) = \sup_{|g|_V \le 1} \xi(g) , \tag{D.3.4}$$

since $\{g \in \mathbb{F}(\mathsf{X}) : |g(x)| \le V(x), x \in D_V\} = \{g = f \cdot V : f \in \mathbb{F}(\mathsf{X}), |f|_\infty \le 1\}$.

Assume now that $\xi \in \mathbb{M}_0(\mathscr{X}) \cap \mathbb{M}_V(\mathscr{X})$ and let S be a Jordan set for ξ. Since $\xi(V\mathbb{1}_S - V\mathbb{1}_{S^c}) = |\xi|(V)$ and $\operatorname{osc}_V(V\mathbb{1}_S - V\mathbb{1}_{S^c}) = 1$, we obtain that

$$\|\xi\|_V = |\xi|(V) \le \sup\{|\xi(f)| \,:\, \mathrm{osc}_V(f) \le 1\} \ .$$

For $\xi \in \mathbb{M}_0(\mathscr{X}) \cap \mathbb{M}_V(\mathscr{X})$, we have $\xi^+(D_V) = \xi^-(D_V)$ and $|\xi|(D_V^c) = 0$. Hence for every measurable function f such that $\mathrm{osc}_V(f) < \infty$, we obtain

$$\begin{aligned}
\xi(f) &= \frac{1}{\xi^+(D_V)} \iint_{D_V \times D_V} \xi^+(\mathrm{d}x)\xi'^-(\mathrm{d}x')\{f(x) - f(x')\} \\
&= \frac{1}{\xi^+(D_V)} \iint_{D_V \times D_V} \xi^+(\mathrm{d}x)\xi'^-(\mathrm{d}x')\{V(x) + V(x')\} \frac{f(x) - f(x')}{V(x) + V(x')} \\
&\le \|\xi\|_V \,\mathrm{osc}_V(f) \ .
\end{aligned}$$

This yields $\sup\{|\xi(f)| \,:\, \mathrm{osc}_V(f) \le 1\} \le \|\xi\|_V$. □

Note that when $V = \mathbb{1}_{\mathsf{X}}$, then $\mathrm{osc}_V(f) = \mathrm{osc}(f)/2$, and thus Proposition D.2.4 can be seen as a particular case of Theorem D.3.2. We also have the following bound, which is similar to (D.2.4):

$$|\xi(f)| \le \|\xi\|_V \,\mathrm{osc}_V(f) \ . \tag{D.3.5}$$

Proposition D.3.3 *The space* $(\mathbb{M}_V(\mathscr{X}), \|\cdot\|_V)$ *is complete.*

Proof. Let $\{\xi_n,\ n \in \mathbb{N}\}$ be a Cauchy sequence in $\mathbb{M}_V(\mathscr{X})$. Define

$$\lambda = \sum_{n=0}^{\infty} \frac{1}{2^n |\xi_n|(V)} |\xi_n| \ ,$$

which is a measure, as a limit of an increasing sequence of measures. By construction, $\lambda(V) < \infty$ and $|\xi_n| \ll \lambda$ for all $n \in \mathbb{N}$. Therefore, there exist functions $f_n \in \mathrm{L}^1(V \cdot \lambda)$ such that $\xi_n = f_n.\lambda$ and $\|\xi_n - \xi_m\|_V = \int |f_n - f_m| V \mathrm{d}\lambda$. This implies that $\{f_n,\ n \in \mathbb{N}\}$ is a Cauchy sequence in $\mathrm{L}^1(V \cdot \lambda)$, which is complete. Thus, there exists $f \in \mathrm{L}^1(V \cdot \lambda)$ such that $f_n \to f$ in $\mathrm{L}^1(V \cdot \lambda)$. Setting $\xi = f.\lambda$, we obtain that $\xi \in \mathbb{M}_V(\mathscr{X})$ and $\lim_{n\to\infty} \|\xi_n - \xi\|_V = \lim_{n\to\infty} \int |f_n - f| V \mathrm{d}\lambda = 0$. □

Proposition D.3.3 yields the following corollary, which is the crux in applying the fixed-point theorem.

Corollary D.3.4 *The space* $(\mathbb{M}_{1,V}(\mathscr{X}), \mathrm{d}_V)$*, where* $\mathbb{M}_{1,V}(\mathscr{X}) = \mathbb{M}_V(\mathscr{X}) \cap \mathbb{M}_1(\mathscr{X})$*, is complete. If* X *is a metric space endowed with its Borel* σ*-field, then convergence with respect to the distance* d_V *implies weak convergence.*

Proof. We need to prove only the second statement. Since $\mathrm{d}_V(\mu,\nu) \le \mathrm{d}_{\mathrm{TV}}(\mu,\nu)$, convergence in d_V distance implies convergence in total variation, which implies weak convergence by Proposition D.2.6. □

Appendix E
Martingales

We recall here the definitions and main properties of martingales, submartingales, and supermartingales that are used in this book. There are many excellent books on martingales, which is an essential topic in probability theory. We will use in this chapter Neveu (1975) and Hall and Heyde (1980).

E.1 Generalized Positive Supermartingales

Definition E.1.1 (Generalized positive supermartingales) *Let* $(\Omega, \mathscr{F}, \{\mathscr{F}_n,\ n \in \mathbb{N}\}, \mathbb{P})$ *be a filtered probability space and* $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ *a positive adapted process.* $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ *is a generalized positive supermartingale if for all* $0 \le m < n$*,* $\mathbb{E}\left[X_n \,|\, \mathscr{F}_m\right] \le X_m$*,* $\mathbb{P} - \text{a.s.}$

Proposition E.1.2 *Let* $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ *be a generalized positive supermartingale. For all* $a > 0$*,*

$$\mathbb{P}\left(\sup_{n \ge 0} X_n \ge a\right) \le a^{-1} \mathbb{E}\left[X_0 \wedge a\right] .$$

Proof. See (Neveu 1975, Proposition II-2-7). □

Proposition E.1.3 *Let* $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ *be a generalized positive supermartingale.* $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ *converges almost surely to a variable* $X_\infty \in$

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1

$[0,\infty]$. *The limit* $X_\infty = \lim_{n\to\infty} X_n$ *satisfies the inequality* $\mathbb{E}[X_\infty \,|\, \mathscr{F}_n] \le X_n$, *for all* $n \in \mathbb{N}$. *Furthermore,* $\{X_\infty < \infty\} \subset \{X_0 < \infty\}$ $\mathbb{P}$*-almost surely.*

Proof. See (Neveu 1975, Theorem II-2-9). For all $M > 0$, using Fatou's lemma and the supermartingale property, we get

$$\mathbb{E}\left[\mathbb{1}_{\{X_0 \le M\}} X_\infty\right] \le \liminf_{n\to\infty} \mathbb{E}\left[\mathbb{1}_{\{X_0 \le M\}} X_n\right] \le \mathbb{E}\left[\mathbb{1}_{\{X_0 \le M\}} X_0\right] \le M \,.$$

□

Proposition E.1.4 *Let* $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ *be a generalized positive supermartingale that converges* $\mathbb{P}-\mathrm{a.s.}$ *to* X_∞. *Then for every pair of stopping times* ν_1, ν_2 *such that* $\nu_1 \le \nu_2 \quad \mathbb{P}-\mathrm{a.s.}$, *we have*

$$X_{\nu_1} \ge \mathbb{E}[X_{\nu_2} \,|\, \mathscr{F}_{\nu_1}] \qquad \mathbb{P}-\mathrm{a.s.}$$

Proof. See (Neveu 1975, Theorem II-2-13). □

E.2 Martingales

Definition E.2.1 (Martingale, submartingale, supermartingale) *Let* $(\Omega, \mathscr{F}, \{\mathscr{F}_n,\ n \in \mathbb{N}\}, \mathbb{P})$ *be a filtered probability space and* $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ *a real-valued integrable adapted process.* $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ *is*

(i) *a martingale if for all* $0 \le m < n$, $\mathbb{E}[X_n \,|\, \mathscr{F}_m] = X_m$, $\mathbb{P}-\mathrm{a.s.}$
(ii) *a submartingale if for all* $0 \le m < n$, $\mathbb{E}[X_n \,|\, \mathscr{F}_m] \ge X_m$, $\mathbb{P}-\mathrm{a.s.}$
(iii) *a supermartingale if for all* $0 \le m < n$, $\mathbb{E}[X_n \,|\, \mathscr{F}_m] \le X_m$, $\mathbb{P}-\mathrm{a.s.}$

If $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ is a submartingale, then $\{(-X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ is a supermartingale; it is a martingale if and only if it is a submartingale and a supermartingale.

Definition E.2.2 (Martingale difference) *A sequence* $\{Z_n,\ n \in \mathbb{N}^*\}$ *is a martingale difference sequence with respect to the filtration* $\{\mathscr{F}_n,\ n \in \mathbb{N}\}$ *if* $\{(Z_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ *is an integrable adapted process and* $\mathbb{E}[Z_n \,|\, \mathscr{F}_{n-1}] = 0$ $\mathbb{P}-\mathrm{a.s.}$ *for all* $n \in \mathbb{N}^*$.

Proposition E.2.3 *Let $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ be a submartingale. Then for all $a > 0$ and all $n \geq 0$,*

$$a\mathbb{P}\left(\max_{k\leq n} X_k \geq a\right) \leq \mathbb{E}\left[X_n \mathbb{1}\left\{\max_{k\leq n} X_k \geq a\right\}\right] \leq \mathbb{E}\left[X_n\right] .$$

Proof. See (Neveu 1975, Proposition II-2-7) and (Hall and Heyde 1980, Theorem 2.1). □

Theorem E.2.4 (Doob's inequalities). *Let $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ be a martingale or a positive submartingale. Then for all $p \in (1, \infty)$ and $m \in \mathbb{N}^*$,*

$$\|X_m\|_p \leq \left\|\max_{k\leq m} |X_k|\right\|_p \leq \frac{p}{p-1} \|X_m\|_p .$$

Proof. See (Neveu 1975, Proposition IV-2-8) and (Hall and Heyde 1980, Theorem 2.2). □

E.3 Martingale Convergence Theorems

Theorem E.3.1. *If $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ is a submartingale satisfying $\sup_n \mathbb{E}\left[X_n^+\right] < \infty$, then there exists a random variable X such that $X_n \stackrel{\mathbb{P}\text{-a.s.}}{\longrightarrow} X$ and $\mathbb{E}\left[|X|\right] < \infty$.*

Proof. See (Neveu 1975, Theorem IV-1-2) and (Hall and Heyde 1980, Theorem 2.5). □

Definition E.3.2 (Uniform integrability) *A family $\{X_i,\ i \in I\}$ of random variables is said to be uniformly integrable if*

$$\lim_{A\to\infty} \sup_{i\in I} \mathbb{E}\left[|X_i| \mathbb{1}\{|X_i| \geq A\}\right] = 0 .$$

Proposition E.3.3 *Let $\{X_n,\ n \in \mathbb{N}\}$ be a sequence of integrable random variables. The following statements are equivalent:*

(i) There exists a random variable X_∞ such that $\lim_{n\to\infty} \mathbb{E}[|X_n - X_\infty|] = 0$.
(ii) There exists a random variable X_∞ such that $X_n \xrightarrow{\mathbb{P}-\mathrm{prob}} X_\infty$ and the sequence $\{X_n, n \in \mathbb{N}\}$ is uniformly integrable.

Proof. (Billingsley 1999, Theorems 3.5 and 3.6). □

Theorem E.3.4. *Let $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ be a uniformly integrable submartingale. There exists a random variable X_∞ such that X_n converges almost surely and in L^1 to X_∞ and $X_n \leq \mathbb{E}[X_\infty | \mathscr{F}_n]$ $\mathbb{P}-\mathrm{a.s.}$ for all $n \in \mathbb{N}$.*

Proof. See (Neveu 1975, Proposition IV-5-24). □

Corollary E.3.5 *Let $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ be a martingale or a nonnegative submartingale. Assume that there exists $p > 1$ such that*

$$\sup_{n\geq 0} \mathbb{E}[|X_n|^p] < \infty\,. \tag{E.3.1}$$

Then there exists a random variable X_∞ such that X_n converges in L^p and almost surely to X_∞.

Theorem E.3.6. *Let $\{(X_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ be a martingale. The following statements are equivalent:*

(i) The sequence $\{X_n,\ n \in \mathbb{N}\}$ is uniformly integrable.
(ii) The sequence $\{X_n,\ n \in \mathbb{N}\}$ converges in L^1.
(iii) There exists $X \in \mathrm{L}^1$ such that for all $n \in \mathbb{N}$, $X_n = \mathbb{E}[X \mid \mathscr{F}_n]$ $\mathbb{P}-\mathrm{a.s.}$

Proof. See (Neveu 1975, Proposition IV-2-3). □

Theorem E.3.7. *Let* $X \in \mathrm{L}^1$, *let* $\{\mathscr{F}_n,\ n \in \mathbb{N}\}$ *be a filtration, and let* $\mathscr{F}_\infty = \sigma\left(\bigcup_{n=0}^\infty \mathscr{F}_n\right)$. *Then the sequence* $\{\mathbb{E}\left[X \,|\, \mathscr{F}_n\right], n \in \mathbb{N}\}$ *converges* $\mathbb{P} - \mathrm{a.s.}$ *and in* L^1 *to* $\mathbb{E}\left[X \,|\, \mathscr{F}_\infty\right]$.

The sequence $\{\mathbb{E}\left[X \,|\, \mathscr{F}_n\right], n \in \mathbb{N}\}$ is called a regular martingale.

Definition E.3.8 (Reversed martingale) *Let* $\{\mathscr{B}_n,\ n \in \mathbb{N}\}$ *be a decreasing sequence of* σ*-fields. A sequence* $\{X_n,\ n \in \mathbb{N}\}$ *of positive or integrable random variables is called a reverse supermartingale relative to the sequence* $\{\mathscr{B}_n,\ n \in \mathbb{N}\}$ *if for all* $n \in \mathbb{N}$, *the random variable* X_n *is* $\mathscr{B}_n$ *measurable and* $\mathbb{E}\left[X_n \,|\, \mathscr{B}_{n+1}\right] \leq X_{n+1}$ *for all* $n \in \mathbb{N}$. *A reverse martingale or submartingale is defined accordingly.*

Theorem E.3.9. *Let* $X \in \mathrm{L}^1$ *and let* $\{\mathscr{B}_n,\ n \in \mathbb{N}\}$ *be a nonincreasing sequence of* σ*-fields. Then the sequence* $\{\mathbb{E}\left[X \,|\, \mathscr{B}_n\right], n \in \mathbb{N}\}$ *converges* $\mathbb{P} - \mathrm{a.s.}$ *and in* L^1 *to* $\mathbb{E}\left[X \,|\, \bigcap_{n=0}^\infty \mathscr{B}_n\right]$.

Proof. See (Neveu 1975, Theorem V-3-11). □

E.4 Central Limit Theorems

Theorem E.4.1. *Let* $\{p_n,\ n \in \mathbb{N}\}$ *be a sequence of integers. For each* $n \in \mathbb{N}$, *let* $\{(M_{n,k}, \mathscr{F}_{n,k}),\ 0 \leq k \leq p_n\}$ *with* $M_{n,0} = 0$ *a square-integrable martingale. Assume that*

$$\sum_{k=1}^{p_n} \mathbb{E}\left[\left.(M_{n,k} - M_{n,k-1})^2\right| \mathscr{F}_{n,k-1}\right] \stackrel{\mathbb{P}\text{ -prob}}{\longrightarrow} \sigma^2\,, \tag{E.4.1}$$

$$\sum_{k=1}^{p_n} \mathbb{E}\left[\left.|M_{n,k} - M_{n,k-1}|^2 \mathbb{1}_{\{|M_{n,k} - M_{n,k-1}| > \varepsilon\}}\right| \mathscr{F}_{n,k-1}\right] \stackrel{\mathbb{P}\text{ -prob}}{\longrightarrow} 0\,. \tag{E.4.2}$$

for all $\varepsilon > 0$. *Then* $M_{n,p_n} \stackrel{\mathbb{P}}{\Longrightarrow} \mathrm{N}(0, \sigma^2)$.

Proof. (Hall and Heyde 1980, Corollary 3.1). □

Applying the previous result in the case that $M_{n,k} = M_k/\sqrt{n}$ and $\{M_n,\ n \in \mathbb{N}\}$ is a square-integrable martingale yields the following corollary.

Corollary E.4.2 *Let $\{(Z_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ be a square-integrable martingale difference sequence. Assume that there exists $\sigma > 0$ such that*

$$n^{-1} \sum_{j=1}^{n} \mathbb{E}\left[Z_j^2 \,\middle|\, \mathscr{F}_{j-1}\right] \overset{\mathbb{P}\text{-prob}}{\longrightarrow} \sigma^2 , \tag{E.4.3}$$

$$n^{-1} \sum_{k=1}^{n} \mathbb{E}\left[Z_k^2 \mathbb{1}_{\{|Z_k| > \varepsilon\sqrt{n}\}} \,\middle|\, \mathscr{F}_{k-1}\right] \overset{\mathbb{P}\text{-prob}}{\longrightarrow} 0 , \tag{E.4.4}$$

for all $\varepsilon > 0$. Then $n^{-1/2}\sum_{k=1}^{n} Z_k \overset{\mathbb{P}}{\Longrightarrow} \mathrm{N}(0, \sigma^2)$.

Theorem E.4.3. *Let $\{Z_n,\ n \in \mathbb{N}^*\}$ be a stationary sequence adapted to the filtration $\{\mathscr{F}_n,\ n \in \mathbb{N}\}$ and such that $\mathbb{E}\left[Z_1^2\right] < \infty$, $\mathbb{E}\left[Z_n \,|\, \mathscr{F}_{n-m}\right] = 0$ for all $n \geq m$ and*

$$\frac{1}{n} \sum_{q=1}^{n} \mathbb{E}\left[\left(\sum_{j=0}^{m-1} \mathbb{E}\left[Z_{q+j} \,\middle|\, \mathscr{F}_q\right] - \mathbb{E}\left[Z_{q+j} \,\middle|\, \mathscr{F}_{q-1}\right]\right)^2 \,\middle|\, \mathscr{F}_{q-1}\right] \overset{\mathbb{P}\text{-prob}}{\longrightarrow} s^2 . \tag{E.4.5}$$

Then $n^{-1/2}\sum_{k=1}^{n} Z_k \overset{\mathbb{P}}{\Longrightarrow} \mathrm{N}(0, s^2)$.

Proof. Since $\mathbb{E}\left[Z_1^2\right] < \infty$, it suffices to prove the central limit theorem for the sum $S_n = \sum_{k=m}^{n} Z_k$. For $k = 1, \ldots, n$ and $q \geq 1$, write

$$\xi_k^{(q)} = \mathbb{E}\left[Z_k \,|\, \mathscr{F}_q\right] - \mathbb{E}\left[Z_k \,|\, \mathscr{F}_{q-1}\right] .$$

Then using the assumption $\mathbb{E}\left[Z_m \,|\, \mathscr{F}_0\right] = 0$, we have

$$\begin{aligned} S_n = \sum_{k=m}^{n} \sum_{q=k-m+1}^{k} \xi_k^{(q)} &= \sum_{q=1}^{n} \sum_{k=q}^{q+m-1} \xi_k^{(q)} \mathbb{1}\left\{m \leq k \leq n\right\} \\ &= \sum_{q=1}^{n} \sum_{j=0}^{m-1} \xi_{q+j}^{(q)} \mathbb{1}\left\{m \leq q+j \leq n\right\} . \end{aligned}$$

If $m \leq q \leq n - m + 1$, then the indicator is 1, i.e., only $2m - 2$ terms are affected by the indicator. Write

$$\zeta_q = \sum_{j=0}^{m-1} \xi_{q+j}^{(q)} , \qquad M_n = \sum_{q=m}^{n-m+1} \zeta_q .$$

Since $\mathbb{E}\left[Z_1^2\right] < \infty$, $n^{-1/2}(S_n - M_n) \stackrel{\mathbb{P}-\text{prob}}{\longrightarrow} 0$. The sequence $\{\zeta_q\}$ is a stationary square-integrable martingale difference sequence. Therefore, to prove the central limit theorem for M_n, we apply Corollary E.4.2. Condition (E.4.3) holds by assumption. By stationarity, the expectation of the left-hand side of (E.4.4) is here

$$\mathbb{E}\left[\zeta_m^2 \mathbb{1}\left\{|\zeta_m| > \varepsilon\sqrt{n}\right\}\right] \to 0$$

as n tends to infinity, since an integrable random variable is uniformly integrable. □

A stationary sequence $\{X_n,\ n \in \mathbb{N}\}$ is said to be m-dependent for a given integer m if $(X_1,\ldots,X_i)$ and $(X_j,X_{j+1},\ldots)$ are independent whenever $j-i > m$.

Corollary E.4.4 *Let $\{Y_n,\ n \in \mathbb{N}\}$ be a stationary m-dependent process on a probability space $(\Omega,\mathscr{F},\mathbb{P})$ such that $\mathbb{E}\left[Y_0^2\right] < \infty$ and $\mathbb{E}[Y_0] = 0$. Then*

$$n^{-1/2} \sum_{k=0}^{n-1} Y_k \stackrel{\mathbb{P}}{\Longrightarrow} \mathrm{N}(0,\sigma^2)$$

with

$$\sigma^2 = \mathbb{E}\left[Y_0^2\right] + 2\sum_{k=1}^{m} \mathbb{E}[Y_0 Y_k] \ . \tag{E.4.6}$$

Theorem E.4.5. *Let m be an integer and $\{Y_n,\ n \in \mathbb{N}^*\}$ a stationary m-dependent process with mean 0. Let $\{\eta_n,\ n \in \mathbb{N}^*\}$ be a sequence of random variables taking only strictly positive integer values such that*

$$\frac{\eta_n}{n} \stackrel{\mathbb{P}-\text{prob}}{\longrightarrow} \vartheta \in (0,\infty) \ . \tag{E.4.7}$$

Then

$$\eta_n^{-1/2} \sum_{k=0}^{\eta_n} Y_k \stackrel{\mathbb{P}}{\Longrightarrow} \mathrm{N}(0,\sigma^2) \quad \text{and} \quad n^{-1/2} \sum_{k=0}^{\eta_n} Y_k \stackrel{\mathbb{P}}{\Longrightarrow} \mathrm{N}(0,\vartheta\sigma^2) ,$$

where σ^2 is given in (E.4.6).

Proof. Set $S_n = \sum_{k=1}^n Y_k$. Without loss of generality, we may assume that $\sigma^2 = 1$. By Corollary E.4.4, $S_{\lfloor \vartheta n \rfloor}/\sqrt{\lfloor \vartheta n \rfloor}$ converges weakly to the standard Gaussian distribution. Write

$$\frac{S_{\eta_n}}{\sqrt{\eta_n}} = \sqrt{\frac{\lfloor \vartheta n \rfloor}{\eta_n}} \frac{S_{\lfloor \vartheta n \rfloor}}{\sqrt{\lfloor \vartheta n \rfloor}} + \sqrt{\frac{\vartheta n}{\eta_n}} \frac{S_{\eta_n} - S_{\lfloor \vartheta n \rfloor}}{\sqrt{\vartheta n}} .$$

By assumption (E.4.7), $\eta_n/\vartheta n \xrightarrow{\mathbb{P}-\text{prob}} 1$ and $\eta_n/\lfloor \vartheta n \rfloor \xrightarrow{\mathbb{P}-\text{prob}} 1$. The theorem will be proved if we show that

$$\frac{S_{\eta_n} - S_{\lfloor \vartheta n \rfloor}}{\sqrt{\vartheta n}} \xrightarrow{\mathbb{P}-\text{prob}} 0 . \tag{E.4.8}$$

Let $\varepsilon \in (0,1)$ be fixed and set

$$a_n = (1-\varepsilon^3)\vartheta n , \; b_n = (1+\varepsilon^3)\vartheta n .$$

Then

$$\begin{aligned}
&\mathbb{P}\left(|S_{\eta_n} - S_{\lfloor \vartheta n \rfloor}| > \varepsilon\sqrt{\vartheta n}\right) \\
&\le \mathbb{P}\left(|S_{\eta_n} - S_{\lfloor \vartheta n \rfloor}| > \varepsilon\sqrt{\vartheta n}, \eta_n \in [a_n,b_n]\right) + \mathbb{P}(\eta_n \notin [a_n,b_n]) \\
&\le \mathbb{P}\left(\max_{a_n \le j \le b_n} |S_j - S_{\lfloor \vartheta n \rfloor}| > \varepsilon\sqrt{\vartheta n}\right) + \mathbb{P}\left(|\eta_n - \vartheta n| \ge \varepsilon^3 n\right) .
\end{aligned} \tag{E.4.9}$$

For $i \in \{0,1,\dots,m-1\}$ and $j \in \mathbb{N}$, set $S_j^{(i)} = \sum_{k=0}^{\lfloor j/m \rfloor} Y_{km+i}$. Note that

$$\begin{aligned}
\mathbb{P}\left(\max_{a_n \le j \le b_n} |S_j - S_{\lfloor \vartheta n \rfloor}| > \varepsilon\sqrt{\vartheta n}\right) &\le \mathbb{P}\left(\max_{1 \le j \le b_n - a_n} |S_j| > \varepsilon\sqrt{\vartheta n}\right) \\
&\le \sum_{i=0}^{m-1} \mathbb{P}\left(\max_{1 \le j \le b_n - a_n} |S_j^{(i)}| > \varepsilon\sqrt{\vartheta n}/m\right) .
\end{aligned} \tag{E.4.10}$$

Since for each $i \in \{0,\dots,m-1\}$, the random variables $\{Y_{km+i}, k \in \mathbb{N}\}$ are i.i.d., Kolmogorov's maximal inequality (Proposition E.2.3) yields

$$\begin{aligned}
\mathbb{P}\left(\max_{1 \le j \le b_n - a_n} |S_j^{(i)}| > \varepsilon\sqrt{\vartheta n}/m\right) &\le \frac{\operatorname{Var}(S_{b_n - a_n})}{\varepsilon^2 \vartheta n} \\
&\le \frac{(2\varepsilon^3 \vartheta n) m^2 \mathbb{E}\left[Y_0^2\right]}{m \varepsilon^2 \vartheta n} = 2m\varepsilon .
\end{aligned} \tag{E.4.11}$$

Assumption (E.4.7) implies that $\limsup_{n\to\infty} \mathbb{P}\left(|\eta_n - \vartheta n| \ge \varepsilon^3 n\right) = 0$. Combining this inequality with (E.4.9) and (E.4.10) shows that

$$\limsup_{n\to\infty} \mathbb{P}\left(|S_{\eta_n} - S_{\lfloor \vartheta n \rfloor}|/\sqrt{\lfloor \vartheta n \rfloor} > \varepsilon\right) \le 2m^2\varepsilon .$$

Since ε is arbitrary, this proves (E.4.8) and consequently the theorem. □

Appendix F
Mixing Coefficients

In this appendix, we briefly recall the definitions and the main properties of mixing coefficients for stationary sequences in Appendices F.1 and F.2 and show that they have particularly simple expressions for Markov chains under the invariant distribution in Appendix F.3. These mixing coefficient are not particularly useful for Markov chains, since taking advantage of the Markov property usually provides similar or better results than using more general methods. Furthermore, Markov chains are often used to build counterexamples or to show that the results obtained with these mixing coefficients are optimal in some sense. Therefore, this appendix is for reference only and is not used elsewhere in this book. Bradley (2005) provides a survey of the main results. The books Doukhan (1994), Rio (2017), and Bradley (2007a, b, c) are authoritative in this field.

F.1 Definitions

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $\mathscr{A}, \mathscr{B}$ two sub-σ-fields of $\mathscr{F}$. Different coefficients were proposed to measure the strength of the dependence between $\mathscr{A}$ and $\mathscr{B}$.

Definition F.1.1 (Mixing coefficients) *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $\mathscr{A}, \mathscr{B}$ two sub-σ-fields of $\mathscr{F}$.*

(i) The α-mixing coefficient is defined by

$$\alpha(\mathscr{A}, \mathscr{B}) = \sup \{ |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|, A \in \mathscr{A}, B \in \mathscr{B} \} . \qquad \text{(F.1.1)}$$

(ii) The β-mixing coefficient is defined by

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1

$$\beta(\mathscr{A},\ \mathscr{B}) = \frac{1}{2}\sup\sum_{i=1}^{I}\sum_{j=1}^{J}|\mathbb{P}(A_i\cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)|\ , \tag{F.1.2}$$

where the supremum is taken over all pairs of (finite) partitions $\{A_1,\ \ldots,\ A_I\}$ *and* $\{B_1,\ \ldots,\ B_J\}$ *of* Ω *such that* $A_i \in \mathscr{A}$ *for each* i *and* $B_j \in \mathscr{B}$ *for each* j.

(iii) The ϕ*-mixing coefficient is defined by*

$$\phi(\mathscr{A},\ \mathscr{B}) = \sup\{|\mathbb{P}(B\,|\,A) - \mathbb{P}(B)|,\ A\in\mathscr{A},\ B\in\mathscr{B},\ \mathbb{P}(A) > 0\}\ . \tag{F.1.3}$$

(iv) The ρ*-mixing coefficient is defined by*

$$\rho(\mathscr{A},\mathscr{B}) = \sup \mathrm{Corr}(f,\ g)\ , \tag{F.1.4}$$

where the supremum is taken over all pairs of square-integrable random variables f *and* g *such that* f *is* $\mathscr{A}$*-measurable and* g *is* $\mathscr{B}$*-measurable.*

F.2 Properties

These coefficients share many common properties. In order to avoid repetition, when stating a property valid for all these coefficients, we will let $\delta(\cdot,\cdot)$ denote any one of them. The coefficients α, β, and ρ are symmetric, whereas the coefficient ϕ is not, but all of them are increasing.

Proposition F.2.1 *If* $\mathscr{A}\subset\mathscr{A}'$ *and* $\mathscr{B}\subset\mathscr{B}'$, *then* $\delta(\mathscr{A},\mathscr{B}) \le \delta(\mathscr{A}',\mathscr{B}')$. *Moreover,*

$$\delta(\mathscr{A},\mathscr{B}) = \sup(\delta(\mathscr{U},\mathscr{V}),\ \mathscr{U},\mathscr{V}\ \textit{finite}\ \sigma\textit{-field}, \mathscr{U}\subset\mathscr{A},\mathscr{V}\subset\mathscr{B})\ . \tag{F.2.1}$$

Proof. Let $\tilde{\delta}(\mathscr{A},\mathscr{B})$ denote the right-hand side of (F.2.1). By the increasing property of the coefficients, $\tilde{\delta}(\mathscr{A},\mathscr{B}) \le \delta(\mathscr{A},\mathscr{B})$. The converse inequality is trivial for α, ϕ, and ρ coefficients. It suffices to consider the finite σ-fields $\{\emptyset, A, A^c, \Omega\}$. We now prove it for the β coefficient. Let $(A_i)_{i\in I}$ and $(B_j)_{j\in J}$ be two partitions of Ω with elements in $\mathscr{A}$ and $\mathscr{B}$, respectively. Let $\mathscr{U}$ and $\mathscr{V}$ be the σ-fields generated by these partitions. Then the desired inequality follows from the identity

$$\beta(\mathscr{U},\mathscr{V}) = \frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}|\mathbb{P}(A_i\cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)|\ .$$

To check that this is true, note that if $A_1\cap A_2 = \emptyset$, then

$$
\begin{aligned}
&|\mathbb{P}((A_1 \cup A_2) \cap B) - \mathbb{P}(A_1 \cup A_2)\mathbb{P}(B)| \\
&\qquad = |\mathbb{P}(A_1 \cap B) - \mathbb{P}(A_1)\mathbb{P}(B) + \mathbb{P}(A_2 \cap B) - \mathbb{P}(A_2)\mathbb{P}(B)| \\
&\qquad \le |\mathbb{P}(A_1 \cap B) - \mathbb{P}(A_1)\mathbb{P}(B)| + |\mathbb{P}(A_2 \cap B) - \mathbb{P}(A_2)\mathbb{P}(B)| \; .
\end{aligned}
$$

This proves our claim, since a partition of Ω measurable with respect to $\mathscr{U}$ consists of sets that are unions of A_i. □

The β coefficient can be characterized in terms of total variation distance. Let $\overline{\mathbb{P}}_{\mathscr{A},\mathscr{B}}$ be the probability measure on $(\Omega \times \Omega, \mathscr{A} \otimes \mathscr{B})$ defined by

$$
\overline{\mathbb{P}}_{\mathscr{A},\mathscr{B}}(A \times B) = \mathbb{P}(A \cap B) \; , \;\; A \in \mathscr{A} \; , \;\; B \in \mathscr{B} \; .
$$

Proposition F.2.2 $\beta(\mathscr{A},\mathscr{B}) = \mathrm{d}_{\mathrm{TV}}(\overline{\mathbb{P}}_{\mathscr{A},\mathscr{B}}, \mathbb{P}_{\mathscr{A}} \otimes \mathbb{P}_{\mathscr{B}})$.

Proof. Let μ be a finite signed measure on a product space $(\mathsf{A} \times \mathsf{B}, \mathscr{A} \otimes \mathscr{B})$. We will prove that

$$
\|\mu\|_{\mathrm{TV}} = \sup \sum_{i=1}^{I} \sum_{j=1}^{J} |\mu(A_i \times B_j)| \, , \tag{F.2.2}
$$

where the supremum is taken over all finite unions of disjoint measurable rectangles. Applying this identity to $\mu = \overline{\mathbb{P}}_{\mathscr{A},\mathscr{B}} - \mathbb{P}_{\mathscr{A}} \otimes \mathbb{P}_{\mathscr{B}}$ will prove our claim. Let the right-hand side of (F.2.2) be denoted by m. By Proposition D.2.3,

$$
\|\mu\|_{\mathrm{TV}} = \sup \sum_{k} |\mu(C_k)| \, ,
$$

where the supremum is taken over finite partitions $\{C_k\}$ of $A \times B$, measurable with respect to $\mathscr{A} \otimes \mathscr{B}$. Thus $\|\mu\|_{\mathrm{TV}} \ge m$. Let D be a Jordan set for μ, i.e., $D \in \mathscr{A} \otimes \mathscr{B}$ satisfying $\|\mu\|_{\mathrm{TV}} = \mu(D) - \mu(D^c)$. For every $\varepsilon > 0$, there exists $E \in \mathscr{E}$ such that $|\mu(D) - \mu(E)| < \varepsilon$ and $|\mu(D^c) - \mu(E^c)| < \varepsilon$. Let $(A_i, i \in I)$ and $(B_j, j \in J)$ be two finite partitions of $(\mathsf{A}, \mathscr{A})$ and $(\mathsf{B}, \mathscr{B})$, respectively, such that $E \in \sigma(A_i \times B_j, (i,j) \in I \times J)$. Then there exists a subset $K \subset I \times J$ such that

$$
\mu(E) = \sum_{(i,j) \in K} \mu(A_i \times B_j) \; , \;\; \mu(E^c) = \sum_{(i,j) \in I \times J \setminus K} \mu(A_i \times B_j) \; .
$$

Therefore,

$$
\begin{aligned}
\|\mu\|_{\mathrm{TV}} - 2\varepsilon &\le |\mu(E)| + |\mu(E^c)| \\
&\le \sum_{(i,j) \in K} |\mu(A_i \times B_j)| + \sum_{(i,j) \in I \times J \setminus K} |\mu(A_i \times B_j)| \\
&= \sum_{(i,j) \in I \times J} |\mu(A_i \times B_j)| \; .
\end{aligned}
$$

Since ε is arbitrary, this implies that $\|\mu\|_{\mathrm{TV}} \le m$. □

Example F.2.3. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let (X,Y) be a random pair. Then

$$\beta(\sigma(X), \sigma(Y)) = \mathrm{d}_{\mathrm{TV}}(\mathscr{L}_{\mathbb{P}}((X,Y)), \mathscr{L}_{\mathbb{P}}(X) \otimes \mathscr{L}_{\mathbb{P}}(Y)) .$$

◀

The α, β, and ϕ coefficients define increasingly strong measures of dependence.

Proposition F.2.4 $2\alpha(\mathscr{A},\mathscr{B}) \le \beta(\mathscr{A},\mathscr{B}) \le \phi(\mathscr{A},\mathscr{B})$.

Proof. The first inequality is a straightforward consequence of the definitions. Let $A \in \mathscr{A}$, $B \in \mathscr{B}$. Note that $|\mathbb{P}(A^c \cap B^c) - \mathbb{P}(A^c)\mathbb{P}(B^c)| = |\mathbb{P}(A \cap B^c) - \mathbb{P}(A)\mathbb{P}(B^c)| = |\mathbb{P}(A^c \cap B) - \mathbb{P}(A^c)\mathbb{P}(B)| = |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)$. Thus

$$\beta(\mathscr{A},\mathscr{B}) \ge \frac{1}{2} \times 4|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| = 2|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| .$$

Since A and B are arbitrary, this yields $\beta(\mathscr{A},\mathscr{B}) \ge 2\alpha(\mathscr{A},\mathscr{B})$.

Let $\{A_i, i \in I\}$ and $\{B_j, j \in J\}$ be two finite partitions of $(\Omega, \mathscr{A})$ and $(\Omega, \mathscr{B})$. For $i \in I$, set

$$J(i) = \{j \in J,\ \mathbb{P}(A_i \cap B_j) \ge \mathbb{P}(A_i)\mathbb{P}(B_j)\} , \quad B(i) = \bigcup_{j \in J(i)} B_j .$$

Since $\{B_j, j \in J\}$ is a partition of Ω, it follows that $\sum_{j\in J}(\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)) = 0$, and hence $\sum_{j \in J} |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)| = \sum_{j \in J(i)} \{\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)\}$ for all $i \in I$. Thus

$$\begin{aligned}
\frac{1}{2}\sum_{j \in J} |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)| &= \sum_{j \in J(i)} \{\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)\} \\
&= \sum_{j \in J(i)} \mathbb{P}(A_i)\{\mathbb{P}\left(B_j \,\middle|\, A_i\right) - \mathbb{P}(B_j)\} \\
&= \mathbb{P}(A_i)\{\mathbb{P}(B(i)\,|\,A_i) - \mathbb{P}(B(i))\} \\
&\le \mathbb{P}(A_i)\phi(\mathscr{A},\mathscr{B}) .
\end{aligned}$$

Summing over i, this yields

$$\frac{1}{2}\sum_{i \in I}\sum_{j \in J} |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)| \le \phi(\mathscr{A},\mathscr{B}) .$$

By Proposition F.2.1, this proves that $\beta(\mathscr{A},\mathscr{B}) \le \phi(\mathscr{A},\mathscr{B})$. □

We now give a characterization of the α coefficient in terms of conditional probabilities.

Proposition F.2.5 $\alpha(\mathscr{A},\mathscr{B}) = \frac{1}{2}\sup_{B\in\mathscr{B}} \mathbb{E}\left[\left|\mathbb{P}\left(B \,|\, \mathscr{A}\right) - \mathbb{P}(B)\right|\right]$.

Proof. For an integrable random variable X such that $\mathbb{E}[X] = 0$, we have the following characterization:

$$\mathbb{E}[|X|] = 2\sup_{A\in\mathscr{A}} \mathbb{E}[X\mathbb{1}_A] \ . \tag{F.2.3}$$

This is easily seen by considering the set $A = \{X > 0\}$. For $A \in \mathscr{A}$, $B \in \mathscr{B}$, $\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{E}[\{\mathbb{P}(B \,|\, \mathscr{A}) - \mathbb{P}(B)\}\mathbb{1}_A]$,

$$\begin{aligned}\alpha(\mathscr{A},\mathscr{B}) &= \sup_{A\in\mathscr{A}}\sup_{B\in\mathscr{B}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| \\ &= \sup_{B\in\mathscr{B}}\sup_{A\in\mathscr{A}} \mathbb{E}[\{\mathbb{P}(B \,|\, \mathscr{A}) - \mathbb{P}(B)\}\mathbb{1}_A] \\ &= \frac{1}{2}\sup_{B\in\mathscr{B}} \mathbb{E}[|\mathbb{P}(B \,|\, \mathscr{A}) - \mathbb{P}(B)|] \ .\end{aligned}$$

□

In order to give more convenient characterizations of the β and ϕ coefficients, we will need the following assumption.

A F.2.6 Let $\mathscr{A}$ and $\mathscr{B}$ be two sub-σ-fields of $\mathscr{F}$. The σ-field $\mathscr{B}$ is countably generated, and there exists a Markov kernel $N : \Omega\times\mathscr{B} \to [0,1]$ such that for every $B \in \mathscr{B}$, $\omega \mapsto N(\omega,B)$ is $\mathscr{A}$-measurable and $\mathbb{P}(B \,|\, \mathscr{A}) = N(\cdot,B)$ $\mathbb{P}-$a.s.

The following result provides alternative expressions for the ϕ coefficients and shows the importance of **A** F.2.6. For a real-valued random variable X defined on a probability space $(\Omega,\mathscr{F},\mathbb{P})$, let $\operatorname{esssup}_{\mathbb{P}}(X)$ be the smallest number $M \in (-\infty,\infty]$ such that $\mathbb{P}(X \le M) = 1$.

Proposition F.2.7 *For all sub-σ-fields $\mathscr{A}$ and $\mathscr{B}$,*

$$\phi(\mathscr{A},\mathscr{B}) = \sup_{B\in\mathscr{B}} \operatorname{esssup}_{\mathbb{P}}|\mathbb{P}(B \,|\, \mathscr{A}) - \mathbb{P}(B)| \ . \tag{F.2.4}$$

Moreover, if **A** F.2.6 *holds, then*

$$\phi(\mathscr{A},\mathscr{B}) = \operatorname{esssup}_{\mathbb{P}} \sup_{B\in\mathscr{B}} |\mathbb{P}(B \,|\, \mathscr{A}) - \mathbb{P}(B)| \ . \tag{F.2.5}$$

Proof. Set $\phi'(\mathscr{A},\mathscr{B}) = \sup_{B\in\mathscr{B}} \operatorname{esssup}_{\mathbb{P}}|\mathbb{P}(B\,|\,\mathscr{A}) - \mathbb{P}(B)|$. For every $B \in \mathscr{B}$, we have

$$\begin{aligned}|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| &= |\mathbb{E}[\mathbb{1}_A\{\mathbb{P}(B\,|\,\mathscr{A}) - \mathbb{P}(B)\}]| \\ &\le \mathbb{P}(A)\operatorname{esssup}_{\mathbb{P}}(|\mathbb{P}(B\,|\,\mathscr{A}) - \mathbb{P}(B)|) \\ &\le \mathbb{P}(A)\phi'(\mathscr{A},\mathscr{B})\,.\end{aligned}$$

Thus $\phi(\mathscr{A},\mathscr{B}) \le \phi'(\mathscr{A},\mathscr{B})$. Conversely, for every $\varepsilon > 0$, there exists B_ε such that

$$\mathbb{P}(|\mathbb{P}(B_\varepsilon\,|\,\mathscr{A}) - \mathbb{P}(B_\varepsilon)| > \phi'(\mathscr{A},\mathscr{B}) - \varepsilon) > 0\,.$$

Define the $\mathscr{A}$-measurable sets A_ε and A'_ε by

$$\begin{aligned}A_\varepsilon &= \{\mathbb{P}(B_\varepsilon\,|\,\mathscr{A}) - \mathbb{P}(B_\varepsilon) > \phi'(\mathscr{A},\mathscr{B}) - \varepsilon\}\,, \\ A'_\varepsilon &= \{\mathbb{P}(B_\varepsilon) - \mathbb{P}(B_\varepsilon\,|\,\mathscr{A}) > \phi'(\mathscr{A},\mathscr{B}) - \varepsilon\}\,.\end{aligned}$$

Note that either $\mathbb{P}(A_\varepsilon) > 0$ or $\mathbb{P}(A'_\varepsilon) > 0$. Assume that $\mathbb{P}(A_\varepsilon) > 0$. Then,

$$\begin{aligned}\phi(\mathscr{A},\mathscr{B}) &\ge \frac{\mathbb{P}(A_\varepsilon\cap B_\varepsilon) - \mathbb{P}(A_\varepsilon)\mathbb{P}(B_\varepsilon)}{\mathbb{P}(A_\varepsilon)} \\ &= \frac{1}{\mathbb{P}(A_\varepsilon)}\int_{A_\varepsilon}[\mathbb{P}(B_\varepsilon\,|\,\mathscr{A}) - \mathbb{P}(B_\varepsilon)]\mathrm{d}\mathbb{P} \ge (\phi'(\mathscr{A},\mathscr{B}) - \varepsilon)\,.\end{aligned}$$

Since ε is arbitirary, this implies that $\phi(\mathscr{A},\mathscr{B}) \ge \phi'(\mathscr{A},\mathscr{B})$. This proves (F.2.4), and we now prove (F.2.5). Under F.2.6, $\mathbb{P}(B\,|\,\mathscr{A}) = N(\cdot,B)$. Since $\mathscr{B}$ is countably generated, by Lemma D.1.5, there exists a sequence $\{B_n,\ n\in\mathbb{N}\}$ of $\mathscr{B}$-measurable sets such that

$$\sup_{B\in\mathscr{B}}|N(\cdot,B) - P(B)| = \sup_{n\in\mathbb{N}}|N(\cdot,B_n) - P(B_n)|\,.$$

Therefore, $\sup_{B\in\mathscr{B}}|N(\cdot,B) - P(B)|$ is measurable, and we can define the random variable $\operatorname{esssup}_{\mathbb{P}}(\sup_{B\in\mathscr{B}}|\mathbb{P}(B\,|\,\mathscr{A}) - P(B)|)$. For every $C \in \mathscr{B}$,

$$|\mathbb{P}(C\,|\,\mathscr{A}) - P(C)| \le \sup_{B\in\mathscr{B}}|\mathbb{P}(B\,|\,\mathscr{A}) - P(B)|\,,$$

which implies that

$$\operatorname{esssup}_{\mathbb{P}}(|\mathbb{P}(C\,|\,\mathscr{A}) - P(C)|) \le \operatorname{esssup}_{\mathbb{P}}\left(\sup_{B\in\mathscr{B}}|\mathbb{P}(B\,|\,\mathscr{A}) - P(B)|\right)\,.$$

Thus

$$\phi(\mathscr{A},\mathscr{B}) = \phi'(\mathscr{A},\mathscr{B}) \le \operatorname{esssup}_{\mathbb{P}}\left(\sup_{B\in\mathscr{B}}|\mathbb{P}(B\,|\,\mathscr{A}) - P(B)|\right)\,.$$

Conversely, for every $C \in \mathscr{B}$,

$$
\begin{aligned}
|\mathbb{P}(C\,|\,\mathscr{A}) - P(C)| &\le \operatorname{esssup}_{\mathbb{P}}(|\mathbb{P}(C\,|\,\mathscr{A}) - P(C)|) \\
&\le \sup_{B\in\mathscr{B}} \operatorname{esssup}_{\mathbb{P}}(|\mathbb{P}(B\,|\,\mathscr{A}) - P(B)|) = \phi(\mathscr{A},\mathscr{B}) .
\end{aligned}
$$

This proves that $\operatorname{esssup}_{\mathbb{P}}(\sup_{B\in\mathscr{B}} |\mathbb{P}(B\,|\,\mathscr{A}) - P(B)|) \le \phi(\mathscr{A},\mathscr{B})$, which concludes the proof of (F.2.5). □

Proposition F.2.8 *For all sub-σ-fields $\mathscr{A}$ and $\mathscr{B}$ such that* **A** F.2.6 *holds,*

$$
\beta(\mathscr{A},\mathscr{B}) = \mathbb{E}\left[\sup_{B\in\mathscr{B}} |\mathbb{P}(B\,|\,\mathscr{A}) - \mathbb{P}(B)|\right] .
$$

Proof. Set $Q = \overline{\mathbb{P}}_{\mathscr{A},\mathscr{B}} - \mathbb{P}_{\mathscr{A}} \otimes \mathbb{P}_{\mathscr{B}}$ and $\beta_1 = (1/2)\,\|Q\|_{\mathrm{TV}} = \beta(\mathscr{A},\mathscr{B})$ by Proposition F.2.2. Under **A** F.2.6, $\sup_{B\in\mathscr{B}} |\mathbb{P}(B\,|\,\mathscr{A}) - \mathbb{P}(B)|$ is measurable. Therefore, we can set $\beta_2 = \mathbb{E}[\sup_{B\in\mathscr{B}} |\mathbb{P}(B\,|\,\mathscr{A}) - \mathbb{P}(B)|]$, and we must prove that $\beta_1 = \beta_2$. Let $D \in \mathscr{A} \otimes \mathscr{B}$ be a Jordan set for Q (which is a signed measure on $(\Omega^2, \mathscr{A} \otimes \mathscr{B})$). Then $\beta_1 = Q(D)$. Set $D_{\omega_1} = \{\omega_2 \in \Omega : (\omega_1,\omega_2) \in D\}$. Then $D_{\omega_1} \in \mathscr{B}$. Under **A** F.2.6, the identity $\overline{\mathbb{P}}_{\mathscr{A},\mathscr{B}} = \mathbb{P} \otimes N$ holds. Indeed, for $A \in \mathscr{A}$ and $B \in \mathscr{B}$, we have

$$
\begin{aligned}
\mathbb{P} \otimes N(A \times B) &= \int_A \mathbb{P}(\mathrm{d}\omega) N(\omega, B) = \mathbb{E}[\mathbb{1}_A \mathbb{P}(B\,|\,\mathscr{A})] \\
&= \mathbb{P}(A \times B) = \overline{\mathbb{P}}_{\mathscr{A},\mathscr{B}}(A \times B) .
\end{aligned}
$$

Define the signed kernel M by setting, for $\omega \in \Omega$ and $B \in \mathscr{B}$,

$$
M(\omega, B) = N(\omega, B) - \mathbb{P}(B) . \tag{F.2.6}
$$

With this notation, $Q = \mathbb{P} \otimes M$. Since $\mathscr{B}$ is countably generated, $\sup_{B\in\mathscr{B}} M(\cdot, B)$ is measurable. Thus applying Fubini's theorem, we obtain

$$
\begin{aligned}
\beta_1 = Q(D) &= \iint \mathbb{P}(\mathrm{d}\omega_1) M(\omega_1, \mathrm{d}\omega_2) \mathbb{1}_D(\omega_1,\omega_2) = \int \mathbb{P}(\mathrm{d}\omega_1) M(\omega_2, D_{\omega_1}) \\
&\le \int \mathbb{P}(\mathrm{d}\omega_1) \sup_{B\in\mathscr{B}} M(\omega_1, B) = \mathbb{E}\left[\sup_{B\in\mathscr{B}} |\mathbb{P}(B\,|\,\mathscr{A}) - \mathbb{P}(B)|\right] = \beta_2 .
\end{aligned}
$$

By Lemma D.1.5, there exists a sequence $\{B_n,\ n \in \mathbb{N}\}$ such that $\sup_{B\in\mathscr{B}} M(\cdot, B) = \sup_{n\ge 0} M(\cdot, B_n)$. Set $Z = \sup_{n\ge 0} M(\cdot, B_n)$. For $\varepsilon > 0$, define the $\mathscr{A}$-measurable random variable N by

$$
N(\omega_1) = \inf\{n \ge 0 : M(\omega_1, B_n) \ge Z(\omega_1) - \varepsilon\} .
$$

Then we can define a set $C \in \mathscr{A} \otimes \mathscr{B}$ by

$$
C = \{(\omega_1,\omega_2),\ \omega_2 \in B_{N(\omega_1)}\} = \bigcup_k \{(\omega_1,\omega_2),\ n(\omega_1) = k, \omega_2 \in B_k\} .
$$

We then have

$$\begin{aligned}\beta_1 &\geq Q(C) = \mathbb{P} \otimes M(C) = \int \mathbb{P}(\mathrm{d}\omega_1) M(\omega_1, B_{N(\omega_1)}) \\ &\geq \mathbb{E}\left[\sup_{B \in \mathscr{B}} M(\cdot, B)\right] - \varepsilon = \beta_2 - \varepsilon \,.\end{aligned}$$

Since ε is arbitrary, we obtain that $\beta_1 \geq \beta_2$. □

The following result is the key to proving the specific properties of the mixing coefficients of Markov chains that we will state in the next section.

Proposition F.2.9 *Let $\mathscr{A}$, $\mathscr{B}$, and $\mathscr{C}$ be sub-σ-fields of $\mathscr{F}$. If $\mathscr{A}$ and $\mathscr{C}$ are conditionally independent given $\mathscr{B}$, then $\delta(\mathscr{A} \vee \mathscr{B}, \mathscr{C}) = \delta(\mathscr{B}, \mathscr{C})$ and $\phi(\mathscr{A}, \mathscr{B} \vee \mathscr{C}) = \phi(\mathscr{A}, \mathscr{B})$.*

Proof. Write $|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| = \mathbb{E}[\{\mathbb{1}_A - \mathbb{P}(A)\}\mathbb{1}_B]$. Lemma B.3.16 implies that for all $A \in \mathscr{A}$,

$$\sup_{B \in \mathscr{B} \vee \mathscr{C}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| = \sup_{B \in \mathscr{B}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

This establishes that $\alpha(\mathscr{A}, \mathscr{B} \vee \mathscr{C}) = \alpha(\mathscr{A}, \mathscr{B})$ and $\phi(\mathscr{A}, \mathscr{B} \vee \mathscr{C}) = \phi(\mathscr{A}, \mathscr{B})$. Applying Propositions B.3.15 and F.2.7, we obtain

$$\begin{aligned}\phi(\mathscr{A} \vee \mathscr{B}, \mathscr{C}) &= \sup_{C \in \mathscr{C}} \operatorname{esssup}_{\mathbb{P}} (|\mathbb{P}(C \mid \mathscr{A} \vee \mathscr{B}) - \mathbb{P}(B)|) \\ &= \sup_{C \in \mathscr{C}} \operatorname{esssup}_{\mathbb{P}} |\mathbb{P}(C \mid \mathscr{B}) - \mathbb{P}(B)| = \phi(\mathscr{B}, \mathscr{C}) \,.\end{aligned}$$

Assume now that $\mathscr{A}$, $\mathscr{B}$, and $\mathscr{C}$ are generated by finite partitions $\{A_i,\ i \in I\}$, $\{B_j,\ j \in J\}$, and $\{C_k,\ k \in K\}$. Then $\mathscr{B} \vee \mathscr{C}$ is generated by the finite partition $\{B_j \cap C_k, j \in J, k \in K\}$. Therefore, using Lemma B.3.3, we obtain, for every $i \in I$,

$$\begin{aligned}&\sum_{j \in J, k \in K} |\mathbb{P}(A_i \cap B_j \cap C_k) - \mathbb{P}(A_i)\mathbb{P}(B_j \cap C_k)| \\ &\qquad = \mathbb{E}\left[\left|\mathbb{E}\left[\mathbb{1}_{A_i} - \mathbb{P}(A_i) \,\middle|\, \mathscr{B} \vee \mathscr{C}\right]\right|\right] \\ &\qquad = \mathbb{E}\left[\left|\mathbb{E}\left[\mathbb{1}_{A_i} - \mathbb{P}(A_i) \,\middle|\, \mathscr{B}\right]\right|\right] \\ &\qquad = \sum_{j \in J} |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)| \,.\end{aligned}$$

Summing this identity over i yields $\beta(\mathscr{A}, \mathscr{B} \vee \mathscr{C}) = \beta(\mathscr{A}, \mathscr{B})$ when the σ-fields are generated by finite partitions of Ω. Applying Proposition F.2.1 concludes the proof. □

F.3 Mixing Coefficients of Markov Chains

In this section, we will discuss the mixing properties of Markov chains. Let $(\mathsf{X}, \mathscr{X})$ be a measurable space and assume that $\mathscr{X}$ is countably generated. Let P be a Markov kernel on $\mathsf{X} \times \mathscr{X}$, $(\Omega, \mathscr{F}, \mathbb{P})$ be the canonical space and $\{X_n,\ n \in \mathbb{N}\}$ the coordinate process. For $0 \leq m \leq n$, define

$$\mathscr{F}_m^n = \sigma(X_k, m \leq k \leq n)\ ,\quad \mathscr{F}_n^\infty = \sigma(X_k, n \leq k \leq \infty)\ .$$

These σ-fields are also countably generated. We are interested in the mixing coefficients of the σ-fields $\mathscr{F}_0^n$ and $\mathscr{F}_{n+k}^\infty$ under the probability measure $\mathbb{P}_\mu$ on the canonical space. In order to stress the initial distribution, we will add the subscript μ to the notation: $\delta_\mu(\mathscr{F}_0^n, \mathscr{F}_{n+k}^\infty)$ is the δ coefficient of $\mathscr{F}_0^n$ and $\mathscr{F}_{n+k}^\infty$ under $\mathbb{P}_\mu$.

Lemma F.3.1 *For all $n, k \geq 0$, the pair of σ-fields $(\mathscr{F}_0^n, \mathscr{F}_{n+k}^\infty)$ satisfy* **A** F.2.6.

Proof. Let θ be the shift operator. If $B \in \mathscr{F}_{n+k}^\infty$, then $\mathbb{1}_B \circ \theta^{-n}$ is the indicator of an event $B_k \in \mathscr{F}_k^\infty$. By the Markov property,

$$\mathbb{P}(B \,|\, \mathscr{F}_0^n) = \mathbb{E}\left[\mathbb{1}_B \circ \theta^{-n} \circ \theta^n \,\middle|\, \mathscr{F}_0^n \right] = \mathbb{E}_{X_n}[\mathbb{1}_B \circ \theta^{-n}] = P(X_n, B_k)\ .$$

This defines a kernel on $\Omega \times \mathscr{F}_{n+k}^\infty$, and thus **A** F.2.6 holds. □

The Markov property entails a striking simplification of the mixing coefficients of a Markov chain.

Proposition F.3.2 *For every initial distribution μ on X,*

$$\delta_\mu(\mathscr{F}_0^n, \mathscr{F}_{n+k}^\infty) = \delta_\mu(\sigma(X_n), \sigma(X_{n+k}))\ .$$

Proof. By the Markov property, $\mathscr{F}_0^n$ and $\mathscr{F}_{n+k+1}^\infty$ are conditionally independent given X_n; similarly, $\mathscr{F}_0^{n-1}$ and X_{n+k} are conditionally independent given X_n. Applying Proposition F.2.9, we have, for every coefficient δ_μ,

$$\begin{aligned}\delta_\mu(\mathscr{F}_0^n, \mathscr{F}_{n+k}^\infty) &= \delta_\mu(\mathscr{F}_0^n, \sigma(X_{n+k}) \vee \mathscr{F}_{n+k+1}^\infty) = \delta_\mu(\mathscr{F}_0^{n-1}, \sigma(X_{n+k}))\\ &= \delta_\mu(\mathscr{F}_0^n \vee \sigma(X_n), \sigma(X_{n+k})) = \delta_\mu(\sigma(X_n), \sigma(X_{n+k}))\ .\end{aligned}$$

□

We can now state the main result of this section.

Theorem F.3.3. *For every initial distribution μ,*

$$\alpha_\mu(\mathscr{F}_0^n, \mathscr{F}_{n+k}^\infty) = \sup_{A \in \mathscr{X}} \int \mu P^n(\mathrm{d}x) |P^k(x,A) - \mu P^{n+k}(A)| \,, \tag{F.3.1}$$

$$\beta_\mu(\mathscr{F}_0^n, \mathscr{F}_{n+k}^\infty) = \int \mu P^n(\mathrm{d}x) \mathrm{d}_{\mathrm{TV}}(P^k(x,\cdot), \mu P^{n+k}) \,, \tag{F.3.2}$$

$$\phi_\mu(\mathscr{F}_0^n, \mathscr{F}_{n+k}^\infty) = \operatorname{esssup}_{\mu P^n} \left(\mathrm{d}_{\mathrm{TV}}(P^k(x,\cdot), \mu P^{n+k}) \right) \,. \tag{F.3.3}$$

Proof. Applying Propositions F.3.2 and F.2.5, we have

$$\begin{aligned}
\alpha_\mu(\mathscr{F}_0^n, \mathscr{F}_{n+k}^\infty) &= \alpha_\mu(\sigma(X_n), \sigma(X_{n+k})) \\
&= \frac{1}{2} \sup_{B \in \sigma(X_{n+k})} \mathbb{E}_\mu[|\mathbb{P}_\mu(B|\sigma(X_n)) - \mathbb{P}(B)|] \\
&= \frac{1}{2} \sup_{C \in \mathscr{X}} \mathbb{E}_\mu(|\mathbb{P}_\mu(X_{n+k} \in C|X_n) - \mathbb{P}(X_{n+k} \in C)|) \\
&= \frac{1}{2} \sup_{C \in \mathscr{X}} \int \mu P^n(dx) |P^k(x,C) - \mu P^{n+k}(C)| \,.
\end{aligned}$$

This proves (F.3.1). Applying now Proposition F.2.8, we obtain

$$\begin{aligned}
\beta_\mu(\mathscr{F}_0^n, \mathscr{F}_{n+k}^\infty) &= \beta_\mu(\sigma(X_n), \sigma(X_{n+k})) \\
&= \mathbb{E}_\mu\left[\sup_{B \in \sigma(X_{n+k})} |\mathbb{P}_\mu(B|\sigma(X_n)) - \mathbb{P}_\mu(B)| \right] \\
&= \mathbb{E}_\mu\left[\sup_{C \in \mathscr{X}} |\mathbb{P}_\mu(X_{n+k} \in C|X_n) - \mathbb{P}_\mu(X_{n+k} \in C)| \right] \\
&= \mathbb{E}_\mu\left[\sup_{C \in \mathscr{X}} |P^k(X_n, C) - \mu P^{n+k}(C)| \right] \\
&= \int \mu P^n(\mathrm{d}x) \sup_{C \in \mathscr{X}} |P^k(x,C) - \mu P^{n+k}(C)| \\
&\le \int \mu P^n(\mathrm{d}x) \mathrm{d}_{\mathrm{TV}}(P^k(x,\cdot), \mu P^{n+k}) \,.
\end{aligned}$$

This proves (F.3.2). Using Proposition F.2.7, we have

$$
\begin{aligned}
\phi_\mu(\mathscr{F}_0^n, \mathscr{F}_{n+k}^\infty) &= \phi_\mu(\sigma(X_n), \sigma(X_{n+k})) \\
&= \operatorname{esssup}_{\mathbb{P}} \left(\sup_{B \in \sigma(X_{n+k})} |\mathbb{P}_\mu(B|\sigma(X_n)) - \mathbb{P}(B)| \right) \\
&= \operatorname{esssup}_{\mathbb{P}} \left(\sup_{C \in \mathscr{X}} |P^k(X_n, C) - \mathbb{P}(X_{n+k} \in C)| \right) \\
&= \operatorname{esssup}_{\mu P^n} \left(\sup_{C \in \mathscr{X}} |P^k(x, C) - \mu P^{n+k}(C)| \right) \\
&= \operatorname{esssup}_{\mu P^n} \left(\mathrm{d_{TV}}(P^k(x, \cdot), \mu P^{n+k}) \right) .
\end{aligned}
$$

This proves (F.3.3). □

Corollary F.3.4 *Let P be a positive Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. Assume that there exist a function $V : \mathsf{X} \to [0, \infty]$ such that $\pi(V) < \infty$ and a nonincreasing sequence $\{\beta_n,\ n \in \mathbb{N}^*\}$ satisfying $\lim_{n\to\infty} \beta_n = 0$ such that $\mathrm{d_{TV}}(P^n(x, \cdot), \pi) \le V(x)\beta_n$ for all $n \ge 0$ and $x \in \mathsf{X}$. Then the canonical chain $\{X_n,\ n \in \mathbb{N}\}$ is β-mixing under $\mathbb{P}_\pi$:*

$$
\beta_\pi(\mathscr{F}_0^n, \mathscr{F}_{n+k}^\infty) \le \pi(V)\beta_n . \tag{F.3.4}
$$

If V is bounded, then the canonical chain $\{X_n,\ n \in \mathbb{N}\}$ is π-mixing with geometric rate under $\mathbb{P}_\pi$.

Proof. The bound (F.3.4) is an immediate consequence of (F.3.2). The last statement is a consequence of (F.3.3) and Proposition 15.2.3). □

We now turn to the ρ-mixing coefficients under stationarity. For notational clarity, we set $\rho_k = \rho_\pi(\mathscr{F}_0^n, \mathscr{F}_{n+k}^\infty)$.

Proposition F.3.5 *Let P be a positive Markov kernel on $\mathsf{X} \times \mathscr{X}$ with invariant probability measure π. Then for all $k \ge 1$,*

$$
\rho_k = |||P|||_{\mathrm{L}_0^2(\pi)} ,
$$

and for all $k \ge 1$, $\rho_k \le \rho_1^k$. Furthermore, if P is reversible with respect to π, then P is geometrically ergodic if and only if $\rho_1 < 1$.

Proof. The first two claims are straightforward consequences of the definition of the ρ-mixing coefficients, and the last one is a consequence of Theorem 22.3.11. □

Appendix G
Solutions to Selected Exercises

Solutions to Exercises of Chapter 1

1.4 1. We have

$$\begin{aligned}\bar{\mathbb{E}}[\mathbb{1}_{A\times\{S_n=k\}}f(Y_{n+1})] &= \bar{\mathbb{E}}[\mathbb{1}_{A\times\{S_n=k\}}f(X_{k+Z_{n+1}})] \\ &= \sum_{j=0}^{\infty} a(j)\bar{\mathbb{E}}[\mathbb{1}_{\{S_n=k\}}\mathbb{1}_A f(X_{k+j})] \\ &= \sum_{j=0}^{\infty} a(j)\bar{\mathbb{E}}[\mathbb{1}_{\{S_n=k\}}\mathbb{1}_A P^j f(X_k)] \\ &= \bar{\mathbb{E}}[\mathbb{1}_{A\times\{S_n=k\}}K_a f(X_k)] = \bar{\mathbb{E}}\left[\mathbb{1}_{A\times\{S_n=k\}}K_a f(Y_n)\right] .\end{aligned}$$

2. This identity shows that for all $n \in \mathbb{N}$ and $f \in \mathbb{F}_+(\mathsf{X})$, $\bar{\mathbb{E}}\left[f(Y_{n+1}\,|\,\mathscr{H}_n\right] = f(Y_n)$.

1.5 Let π be an invariant probability. Then for all $f \in \mathbb{F}_+(\mathsf{X})$, by Fubini's theorem,

$$\begin{aligned}\int_{\mathsf{X}} f(x)\pi(\mathrm{d}x) = \int_{\mathsf{X}} Pf(x)\pi(\mathrm{d}x) &= \int_{\mathsf{X}}\left[\int_{\mathsf{X}} p(x,y)f(y)\mu(\mathrm{d}y)\right]\pi(\mathrm{d}x) \\ &= \int_{\mathsf{X}}\left[\int_{\mathsf{X}} p(x,y)\pi(\mathrm{d}x)\right] f(y)\mu(\mathrm{d}y) .\end{aligned}$$

This implies that $\pi(f) = \int_{\mathsf{X}} f(y)q(y)\mu(\mathrm{d}y)$ with $q(y) = \int_{\mathsf{X}} p(x,y)\pi(\mathrm{d}x) > 0$. Hence the probabilities π and μ are equivalent.

Assume that there are two distinct invariant probabilities. By Theorem 1.4.6 (ii), there exist two singular invariant probabilities π and π', say. Since we have just proved that $\pi \sim \mu$ and $\pi' \sim \mu$, this is a contradiction.

1.6 1. The invariance of π implies that

$$\pi(\mathsf{X}_1) = 1 = \int_{\mathsf{X}_1} P(x,\ \mathsf{X}_1)\pi(\mathrm{d}x) .$$

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1

Therefore, there exists a set $\mathsf{X}_2 \in \mathscr{X}$ such that

$$\mathsf{X}_2 \subset \mathsf{X}_1\,,\ \pi(\mathsf{X}_2) = 1 \quad \text{and} \quad P(x,\ \mathsf{X}_1) = 1\,,\ \text{for all } x \in \mathsf{X}_2.$$

Repeating the above argument, we obtain a decreasing sequence $\{\mathsf{X}_i, i \geq 1\}$ of sets $\mathsf{X}_i \in \mathscr{X}$ such that $\pi(\mathsf{X}_i) = 1$ for all $i = 1,2,\ \ldots$, and $P(x,\ \mathsf{X}_i) = 1$ for all $x \in \mathsf{X}_{i+1}$.

2. The set B is nonempty, because

$$\pi(B) = \pi\left(\bigcap_{i=1}^{\infty} \mathsf{X}_i\right) = \lim_{i\to\infty} \pi(\mathsf{X}_i) = 1\ .$$

3. The set B is absorbing for P, because for all $x \in B$,

$$P(x,\ B) = P\left(x,\ \bigcap_{i=1}^{\infty} \mathsf{X}_i\right) = \lim_{i\to\infty} P(x,\ \mathsf{X}_i) = 1.$$

1.8 The proof is by contradiction. Assume that μ is invariant. Clearly, one must have $\mu(\{0\}) = 0$, since $P(x,\{0\}) = 0$ for every $x \in [0,1]$. Since for $x \in [0,1]$, $P(x,(1/2,1]) = 0$, one must also have $\mu((1/2,1]) = 0$. Proceeding by induction, we must have $\mu((1/2^n,1]) = 0$ for every n, and therefore $\mu((0,1]) = 0$. Therefore, $\mu([0,1]) = 1$.

1.11 1. The transition matrix is given by

$$P(i,i+1) = \frac{N-i}{N}\,, \quad i = 0,\ldots,N-1\ ,$$
$$P(i,i-1) = \frac{i}{N}\,, \quad i = 1,\ldots,N\ .$$

2. For all $i = 0,\ldots,N-1$,

$$\binom{N}{i}\frac{N-i}{N} = \frac{N!(N-i)}{i!(N-i)!N} = \binom{N}{i+1}\frac{i+1}{N}\ .$$

This is the detailed balance condition of Definition 1.5.1. Thus the binomial distribution $B(N,1/2)$ is invariant.

3. For $n \geq 1$,

$$\mathbb{E}\left[X_n\,|\,X_{n-1}\right] = (X_{n-1}+1)\frac{N-X_{n-1}}{N} + (X_{n-1}-1)\frac{X_{n-1}}{N} = X_{n-1}(1-2/N)+1\ .$$

4. Set $m_n(x) = \mathbb{E}_x[X_n]$ for $x \in \{0,\ldots,N\}$ and $a = 1-2/N$, which yields

$$m_n(x) = a m_{n-1}(x) + 1\ .$$

The solution to this recurrence equation is

$$m_n(x) = xa^n + \frac{1-a^n}{1-a} \;,$$

and since $0 < a < 1$, this yields that $\lim_{n\to\infty} \mathbb{E}_x[X_n] = 1/(1-a) = N/2$, which is the expectation of the stationary distribution.

1.12 1. For all $(x,y) \in \mathsf{X}\times\mathsf{X}$, $[DM]_{x,y} = \pi(x)M(x,y)$ and $[M^T D]_{x,y} = M(y,x)\pi(y)$, and hence $[DM]_{x,y} = [M^D]x,y$.

2. The proof is by induction. Assume that $DM^{k-1} = (M^{k-1})^T D$. Then

$$DM^k = DM^{k-1}M = (M^{k-1})^T DM = (M^{k-1})^T M^T D = (M^k)^T D \,.$$

3. Premultiplying by $D^{-1/2}$ and postmultiplying by $D^{1/2}$ the relation $DM = M^T D$, we get $T = D^{1/2}MD^{-1/2} = D^{-1/2}M^T D^{1/2}$. Thus T can be orthogonally diagonalized, $T = \Gamma\beta\Gamma^t$, with Γ orthogonal and β a diagonal matrix having the eigenvalues of T, and therefore M, on the diagonal. Thus $M = V\beta V^{-1}$ with $V = D^{-1/2}\Gamma, V^{-1} = \Gamma^T D^{1/2}$.
4. The right eigenvectors of M are the columns of $V : V_{xy} = \Gamma_{xy}/\sqrt{\pi(x)}$. These are orthonormal in $\mathrm{L}^2(\pi)$. The left eigenvectors are the rows of V^{-1}: $V^{-1}_{xy} = \Gamma_{yx}\sqrt{\pi(y)}$. These are orthonormal in $\mathrm{L}^2(1/\pi)$.

1.13 If $\mu = \mu P$, then $\mu = \mu K_{a_\eta}$. Conversely, assume that $\mu = \mu K_{a_\eta}$. The identity $K_{a_\eta} = (1-\eta)I + \eta K_{a_\eta} P$ yields $\mu = (1-\eta)\mu + \eta\mu P$. Thus $\mu(A) = \mu P(A)$ for all $A \in \mathscr{X}$ such that $\mu(A) < \infty$. Since by definition μ is σ-finite, this yields $\mu P = \mu$.

Solutions to Exercises of Chapter 2

2.1 1. For every bounded measurable function f, we get

$$\begin{aligned}\mathbb{E}[f(X_1)] &= \mathbb{E}[f(V_1X_0 + (1-V_1)Z_1)] \\ &= \alpha\mathbb{E}[f(X_0)] + (1-\alpha)\mathbb{E}[f(Z_1)] = \alpha\xi(f) + (1-\alpha)\pi(f) \,.\end{aligned}$$

This implies that the $\{X_n,\, n\in\mathbb{N}\}$ is a Markov chain with kernel P defined by

$$Pf(x) = \alpha f(x) + (1-\alpha)\pi(f) \,.$$

2. Since $Pf = \alpha f + (1-\alpha)\pi(f)$, we get

$$\xi P = \alpha\xi + (1-\alpha)\pi \,,$$

for every probability measure ξ on $\mathbb{R}$. This yields that π is the unique invariant probability.
3. for every positive integer h, we get

$$\mathrm{Cov}(X_h, X_0) = \mathrm{Cov}(V_h X_{h-1} + (1-V_h)Z_h, X_0) = \alpha\,\mathrm{Cov}(X_{h-1}, X_0) \,,$$

which implies that $\mathrm{Cov}(X_h, X_0) = \alpha^h \mathrm{Var}(X_0)$.

2.2 1. The kernel P is defined by $Ph(x) = \mathbb{E}[h(\phi x + Z_0)]$.

2. Iterating (2.4.2) yields for all $k \geq 1$,

$$X_k = \phi^k X_0 + \sum_{j=0}^{k-1} \phi^j Z_{k-j} = \phi^k X_0 + A_k \,, \tag{G.1}$$

with $A_k = \sum_{j=0}^{k-1} \phi^j Z_{k-j}$. Since $\{Z_k,\ k \in \mathbb{N}\}$ is an i.i.d. sequence, A_k and B_k have the same distribution for all $k \geq 1$.

3. Assume that $|\phi| < 1$. Then $\{B_k,\ k \in \mathbb{N}\}$ is a martingale and is bounded in $\mathrm{L}^1(\mathbb{P})$, i.e.,

$$\sup_{k \geq 0} \mathbb{E}[|B_k|] \leq \mathbb{E}[|Z_0|] \sum_{j=0}^{\infty} |\phi^j| < \infty \,.$$

Hence, by the martingale convergence theorem (Theorem E.3.1),

$$B_k \xrightarrow{\mathbb{P}\text{-a.s.}} B_\infty = \sum_{j=0}^{\infty} \phi^j Z_j \,.$$

4. Let π be the distribution of B_∞ and let Z_{-1} have the same distribution as Z_0 and be independent of all other variables. Then π is invariant, since $\phi B_\infty + Z_{-1}$ has the same distribution as B_∞ and has distribution πP by definition of P. Let ξ be an invariant distribution and X_0 have distribution ξ. Then for every $n \geq 1$, the distribution of $X_n = \phi^n X_0 + \sum_{j=1}^{n} \phi^j + Z_{n-j}$ is also ξ. On the other hand, we have seen that the distribution of X_n is the same as that of $\phi^n X_0 + B_k$. Since $B_k \xrightarrow{\mathbb{P}\text{-a.s.}} B_\infty$ and $\phi^n X_0 \xrightarrow{\mathbb{P}\text{-a.s.}} 0$, we obtain that $\xi = \pi$.

5. If $X_0 = x$, applying (G.1), we obtain for all $n \geq 1$,

$$\begin{aligned} \phi^{-n} X_n &= x + \phi^{-n} \frac{\phi^n - 1}{\phi - 1} \mu + \sum_{j=0}^{n-1} \phi^{j-n} Z_{n-j} \\ &= x + \frac{1 - \phi^{-n}}{\phi - 1} \mu + \sum_{j=1}^{n} \phi^{-j} Z_j \,. \end{aligned}$$

Thus since $C_j = \sum_{j=1}^{n} \phi^{-j} Z_j$ is a martingale bounded in $\mathrm{L}^1(\mathbb{P})$, we obtain

$$\lim_{n \to \infty} \phi^{-n} X_n = x + \frac{1}{\phi - 1} \mu + \sum_{j=1}^{\infty} \phi^{-j} Z_j \ \ \mathbb{P}_x - \text{a.s.}$$

Thus $\lim_{n \to \infty} |X_n| = +\infty$ unless possibly if $x + \frac{1}{\phi - 1} \mu + \sum_{j=1}^{\infty} \phi^{-j} Z_j = 0$, which happens with zero $\mathbb{P}_x$ probability for all x if the distribution of $\sum_{j=1}^{\infty} \phi^{-j} Z_j$ is continuous.

2.3 Defining $f(x,z) = (a + bz)x + z$ yields $X_k = f(X_{k-1}, Z_k)$.

(i) For $(x,y,z) \in \mathbb{R}^3$, $|f(x,z) - f(y,z)| \le |a+bz||x-y|$. If $\mathbb{E}[\ln(|a+bZ_0|)] < 0$, then (2.1.16) holds with $K(z) = |a+bz|$. If, in addition, $\mathbb{E}\left[\ln^+(|Z_0|)\right] < \infty$, then (2.1.18) also holds, and Theorem 2.1.9 holds. Thus the bilinear process defined by (2.4.3) has a unique invariant probability π and $\xi P^n \stackrel{\mathrm{w}}{\Rightarrow} \pi$ for every initial distribution ξ.

2.4 Set $\mathsf{Z} = [0,1] \times \{0,1\}$ and $\mathscr{Z} = \mathscr{B}([0,1]) \otimes \mathscr{P}\{0,1\}$. Then $X_k = f_{Z_k}(X_{k-1})$ with $Z_k = (U_k, \varepsilon_k)$ and $f_{u,\varepsilon}(x) = xu\varepsilon + (1-\varepsilon)[x + u(1-x)]$. For all $(x,y) \in [0,1] \times [0,1]$, $|f_{u,\varepsilon}(x) - f_{u,\varepsilon}(y)| \le K(u,\varepsilon)|x-y|$ with

$$K(u,\varepsilon) = \varepsilon u + (1-\varepsilon)(1-u)\ . \tag{G.2}$$

Equation (2.1.16) is satisfied, since $\mathbb{E}[|\log(K(U,\varepsilon))|] = \mathbb{E}[|\log(U)|] < \infty$ and $\mathbb{E}[\log(U)] = -1$. And (2.1.18) is also satisfied, since for all $x \in [0,1]$ and $z \in \mathsf{Z}$, $f_z(x) \in [0,1]$. Therefore, Theorem 2.1.9 shows that $\{X_k,\ k \in \mathbb{N}\}$ has a unique invariant probability.

Solutions to Exercises of Chapter 3

3.1 1. We must show that the events $\{\tau \wedge \sigma \le n\}$, $\{\tau \vee \sigma \le n\}$ and $\{\tau + \sigma \le n\}$ belong to $\mathscr{F}_n$ for every $n \in \mathbb{N}$. Since

$$\{\tau \wedge \sigma \le n\} = \{\tau \le n\} \cup \{\sigma \le n\}$$

and τ and σ are stopping times, $\{\tau \le n\}$ and $\{\sigma \le n\}$ belong to $\mathscr{F}_n$; therefore, $\{\tau \wedge \sigma \le n\} \in \mathscr{F}_n$. Similarly, $\{\tau \vee \sigma \le n\} = \{\tau \le n\} \cap \{\sigma \le n\} \in \mathscr{F}_n$. Finally,

$$\{\tau + \sigma \le n\} = \bigcup_{k=0}^{n} \{\tau \le k\} \cap \{\sigma \le n-k\}\ .$$

For $0 \le k \le n$, we have $\{\tau \le k\} \in \mathscr{F}_k \subset \mathscr{F}_n$ and $\{\sigma \le n-k\} \in \mathscr{F}_{n-k} \subset \mathscr{F}_n$; hence $\{\tau + \sigma \le n\} \in \mathscr{F}_n$.

2. Let $A \in \mathscr{F}_\tau$ and $n \in \mathbb{N}$. Since $\{\sigma \le n\} \subset \{\tau \le n\}$, it follows that

$$A \cap \{\sigma \le n\} = A \cap \{\tau \le n\} \cap \{\sigma \le n\}\ .$$

Since $A \in \mathscr{F}_\tau$ and $\{\sigma \le n\} \in \mathscr{F}_n$ (σ begins a stopping time), we have $A \cap \{\tau \le n\} \in \mathscr{F}_n$. Therefore, $A \cap \{\tau \le n\} \cap \{\sigma \le n\} \in \mathscr{F}_n$ and $A \cap \{\sigma \le n\} \in \mathscr{F}_n$. Thus $A \in \mathscr{F}_\sigma$.

3. It follows from (i) and (ii) that $\mathscr{F}_{\tau\wedge\sigma} \subset \mathscr{F}_\tau \cap \mathscr{F}_\sigma$. Conversely, let $A \in \mathscr{F}_\tau \cap \mathscr{F}_\sigma$. Obviously $A \subset \mathscr{F}_\infty$. To prove that $A \in \mathscr{F}_{\tau\wedge\sigma}$, one must show that for every $k \ge 0$, $A \cap \{\tau \wedge \sigma \le k\} \in \mathscr{F}_k$. We have $A \cap \{\tau \le k\} \in \mathscr{F}_k$ and $A \cap \{\sigma \le k\} \in \mathscr{F}_k$. Hence since $\{\tau \wedge \sigma \le k\} = \{\tau \le k\} \cup \{\sigma \le k\}$, we get

$$
\begin{aligned}
A \cap \{\tau \wedge \sigma \le k\} &= A \cap (\{\tau \le k\} \cup \{\sigma \le k\}) \\
&= (A \cap \{\tau \le k\}) \cup (A \cap \{\sigma \le k\}) \in \mathscr{F}_k \, .
\end{aligned}
$$

4. Let $n \in \mathbb{N}$. Then

$$
\{\tau < \sigma\} \cap \{\tau \le n\} = \bigcup_{k=0}^{n} \{\tau = k\} \cap \{\sigma > k\} \, .
$$

For $0 \le k \le n$, $\{\tau = k\} = \{\tau \le k\} \cap \{\tau \le k-1\}^c \in \mathscr{F}_k \subset \mathscr{F}_n$ and $\{\sigma > k\} = \{\sigma \le k\}^c \in \mathscr{F}_k \subset \mathscr{F}_n$. Therefore, $\{\tau < \sigma\} \cap \{\tau \le n\} \in \mathscr{F}_n$, showing that $\{\tau < \sigma\} \in \mathscr{F}_\tau$. Similarly,

$$
\{\tau < \sigma\} \cap \{\sigma \le n\} = \bigcup_{k=0}^{n} \{\sigma = k\} \cap \{\tau < k\},
$$

and since for $0 \le k \le n$, $\{\sigma = k\} \in \mathscr{F}_k \subset \mathscr{F}_n$ and $\{\tau < k\} = \{\tau \le k-1\} \in \mathscr{F}_{k-1} \subset \mathscr{F}_n$, one also has $\{\tau < \sigma\} \cap \{\sigma \le n\} \in \mathscr{F}_n$, so that $\{\tau < \sigma\} \in \mathscr{F}_\sigma$. Finally, $\{\tau < \sigma\} \in \mathscr{F}_\tau \cap \mathscr{F}_\sigma$. The last statement of the proposition follows from

$$
\{\tau = \sigma\} = \{\tau < \sigma\}^c \cap \{\sigma < \tau\}^c \in \mathscr{F}_\tau \cap \mathscr{F}_\sigma \, .
$$

3.5

$$
\begin{aligned}
P^n(x,A) = \mathbb{E}_x[\mathbb{1}_A(X_n)] &= \mathbb{E}_x[\mathbb{1}_{\{\sigma \le n\}} \mathbb{1}_A(X_n)] + \mathbb{E}_x[\mathbb{1}_{\{\sigma > n\}} \mathbb{1}_A(X_n)] \\
&= \sum_{k=1}^{n} \mathbb{E}_x[\mathbb{1}_{\{\sigma = k\}} \mathbb{1}_A(X_n)] + \mathbb{E}_x[\mathbb{1}_{\{\sigma > n\}} \mathbb{1}_A(X_n)] \, .
\end{aligned}
$$

By the Markov property, for $k \le n$, we get

$$
\mathbb{E}_x[\mathbb{1}_{\{\sigma = k\}} \mathbb{1}_A(X_n)] = \mathbb{E}_x[\mathbb{1}_{\{\sigma = k\}} \mathbb{1}_A(X_{n-k}) \circ \theta^k] = \mathbb{E}_x[\mathbb{1}_{\{\sigma = k\}} P^{n-k} \mathbb{1}_A(X_k)] \, .
$$

The proof follows.

3.7 1. First note that the assumption $C \subset C_+(r,f)$ implies $\mathbb{P}_x(\mathbb{1}_C(X_1)\mathbb{E}_{X_1}[U] < \infty) = 1$.
Combining $U \circ \theta = \sum_{k=1}^{\sigma_C \circ \theta} r(k-1) f(X_k)$ with the fact that on the event $\{X_1 \notin C\}$, $\sigma_C = 1 + \sigma_C \circ \theta$, we get

$$
\begin{aligned}
\mathbb{1}_{C^c}(X_1) U \circ \theta = \mathbb{1}_{C^c}(X_1) \left(\sum_{k=1}^{\sigma_C - 1} r(k-1) f(X_k) \right) &\le M \mathbb{1}_{C^c}(X_1) \left(\sum_{k=1}^{\sigma_C - 1} r(k) f(X_k) \right) \\
&\le M \mathbb{1}_{C^c}(X_1) U \, .
\end{aligned}
$$

2. by the Markov property, for every $x \in C_+(r,f)$,

$$
\mathbb{E}_x[\mathbb{1}_{C^c}(X_1) \mathbb{E}_{X_1}[U]] = \mathbb{E}_x[\mathbb{1}_{C^c}(X_1) U \circ \theta] \le M \mathbb{E}_x[\mathbb{1}_{C^c}(X_1) U] \le M \mathbb{E}_x[U] < \infty \, .
$$

This implies $\mathbb{P}_x(\mathbb{1}_{C^c}(X_1)\mathbb{E}_{X_1}[U] < \infty) = 1$, and (3.7.2) is proved.

3. Therefore, the set $C_+(r, f)$ is absorbing. The set C being accessible and $C \subset C_+(r, f)$, the set $C_+(r, f)$ is in turn accessible. The proof is then completed by applying Exercise 3.8.

3.8 1. Since π is invariant,

$$\pi(C) = \pi K_{a_\varepsilon}(C) = \int_C \pi(\mathrm{d}x) K_{a_\varepsilon}(x, C) + \int_{C^c} \pi(\mathrm{d}x) K_{a_\varepsilon}(x, C) . \tag{G.3}$$

Since C is absorbing, $K_{a_\varepsilon}(x, C) = 1$ for all $x \in C$. The first term on the right-hand side of (G.3) is then equal to $\pi(C)$. Finally,

$$\int_{C^c} \pi(\mathrm{d}x) K_{a_\varepsilon}(x, C) = 0 .$$

2. The set C being accessible, the function $x \mapsto K_{a_\varepsilon}(x, C)$ is positive. The previous equation then implies $\pi(C^c) = 0$.

Solutions to Exercises of Chapter 4

4.1 1. Since f is superharmonic, $\{P^n f : n \in \mathbb{N}\}$ is a decreasing sequence of positive functions, hence convergent.

2. Since $P^n f \le f$ for all $n \ge 1$ and $Pf \le f < \infty$, applying Lebesgue's dominated convergence theorem yields, for every $x \in \mathsf{X}$,

$$Ph(x) = P\left(\lim_{n\to\infty} P^n f(x)\right) = \lim_{n\to\infty} P^{n+1} f(x) = h(x) .$$

3. Since f is superharmonic, g is nonnegative. Therefore, $P^k g \ge 0$ for all $k \in \mathbb{N}$, and $Ug = \lim_{n\to\infty} \sum_{k=0}^{n-1} P^k g$ is well defined. Moreover, we have, for all $n \ge 1$ and $x \in \mathsf{X}$,

$$\sum_{k=0}^{n-1} P^k g(x) = f(x) - P^n f(x) .$$

Taking limits on both sides yields $Ug(x) = f(x) - h(x)$.

4. Since $\bar{h}$ is harmonic, we have

$$P^n f = \bar{h} + \sum_{k=n}^{\infty} P^k \bar{g} .$$

5. Since $U\bar{g}(x) < \infty$ for all $x \in \mathsf{X}$, we have $\lim_{n\to\infty} \sum_{k=n}^{\infty} P^k \bar{g}(x) = 0$. This yields

$$\bar{h} = \lim_{n\to\infty} P^n f = h .$$

This in turn implies that $Ug = U\bar{g}$. Since $Ug = g + PUg$ and $U\bar{g} = \bar{g} + PU\bar{g}$, we also conclude that $g = \bar{g}$.

4.2 1. Applying Exercise 4.1, we can write $f_A(x) = h(x) + Ug(x)$ with $h(x) = \lim_{n\to\infty} P^n f_A(x)$ and $g(x) = f_A(x) - Pf_A(x)$.

2. The Markov property yields

$$\begin{aligned} P^n f_A(x) &= \mathbb{E}_x[f_A(X_n)] = \mathbb{E}_x[\mathbb{P}_{X_n}(\tau_A < \infty)] \\ &= \mathbb{P}_x(\tau_A \circ \theta_n < \infty) = \mathbb{P}_x\left(\bigcup_{k\geq n}\{X_k \in A\}\right) . \end{aligned}$$

This yields that the harmonic part of f_A in the Riesz decomposition is given by

$$\begin{aligned} h(x) &= \lim_{n\to\infty} P^n f_A(x) = \lim_{n\to\infty} \mathbb{P}_x\left(\bigcup_{k\geq n}\{X_k \in A\}\right) \\ &= \mathbb{P}_x\left(\limsup_{k\to\infty}\{X_k \in A\}\right) = \mathbb{P}_x(N_A = \infty) = h_A(x) . \end{aligned}$$

3. We finally have to compute $f_A - Pf_A$:

$$\begin{aligned} f_A(x) - Pf_A(x) &= \mathbb{P}_x\left(\bigcup_{k\geq 0}\{X_k \in A\}\right) - \mathbb{P}_x\left(\bigcup_{k\geq 1}\{X_k \in A\}\right) \\ &= \mathbb{P}_x\left(\{X_0 \in A\} \cap \bigcap_{n=1}^{\infty}\{X_n \notin A\}\right) = \mathbb{1}_A(x)\mathbb{P}_x(\sigma_A = \infty) = g_A(x) . \end{aligned}$$

4.3 1. We have $U_{n+1} - U_n = Z_{n+1} - \mathbb{E}\left[Z_{n+1} \,|\, \mathscr{F}_n\right]$, and thus $\mathbb{E}\left[U_{n+1} - U_n \,|\, \mathscr{F}_n\right] = 0$. Therefore, $\{(U_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ is a martingale.

2. Since $\{(U_n, \mathscr{F}_n),\ n \in \mathbb{N}\}$ is a martingale, $\mathbb{E}\left[U_{n\wedge\tau}\right] = \mathbb{E}\left[U_0\right]$. This implies

$$\mathbb{E}\left[Z_{n\wedge\tau}\right] - \mathbb{E}\left[Z_0\right] = \mathbb{E}\left[\sum_{k=0}^{n\wedge\tau-1}\{\mathbb{E}\left[Z_{k+1} \,|\, \mathscr{F}_k\right] - Z_k\}\right] .$$

3. We conclude by applying Lebesgue's dominated convergence theorem, since $\{Z_n,\ n \in \mathbb{N}\}$ is bounded and the stopping time τ is integrable.

4.4 Applying Exercise 4.3 to the finite stopping time $\tau \wedge n$ and the bounded process $\{Z_n^M,\ n \in \mathbb{N}\}$, where $Z_n^M = Z_n \wedge M$, we get

$$\mathbb{E}\left[Z_{\tau\wedge n}^M\right] + \mathbb{E}\left[\sum_{k=0}^{\tau\wedge n-1} Z_k^M\right] = \mathbb{E}\left[Z_0^M\right] + \mathbb{E}\left[\sum_{k=0}^{\tau\wedge n-1} \mathbb{E}\left[Z_{k+1}^M \,\middle|\, \mathscr{F}_k\right]\right] .$$

Using Lebesgue's dominated convergence theorem, $\lim_{n\to\infty} \mathbb{E}\left[Z_{\tau\wedge n}^M\right] = \mathbb{E}\left[Z^M\right]$. Using the monotone convergence theorem, we get

$$\mathbb{E}\left[Z_\tau^M\right] + \mathbb{E}\left[\sum_{k=0}^{\tau-1} Z_k^M\right] = \mathbb{E}\left[Z_0^M\right] + \mathbb{E}\left[\sum_{k=0}^{\tau-1} \mathbb{E}\left[Z_{k+1}^M \,\middle|\, \mathscr{F}_k\right]\right] .$$

We conclude by using again the monotone convergence theorem as M goes to infinity.

4.5 1. For $x \notin A$, $\mathbb{P}_x(\tau = 0) = 1$. For $x \in A^c$, we have

$$\begin{aligned}\mathbb{P}_x(\tau \le (b-a)) &\ge \mathbb{P}_x(X_1 = x+1,\, X_2 = x+2,\, \dots,\, X_{b-x} = b) \\ &\ge p^{b-x} \ge p^{b-a} > 0 .\end{aligned}$$

This implies that $\mathbb{P}_x(\tau > b-a) \le 1-\gamma$ for all $x \in A^c$, where $\gamma = p^{b-a}$.

2. For all $x \in A^c$ and $k \in \mathbb{N}^*$, the Markov property implies

$$\begin{aligned}\mathbb{P}_x(\tau > k(b-a)) &= \mathbb{P}_x(\tau > (k-1)(b-a),\; \tau \circ \theta_{(k-1)(b-a)} > (b-a)) \\ &= \mathbb{E}_x[\mathbb{1}_{\{\tau > (k-1)(b-a)\}} \mathbb{P}_{X_{(k-1)(b-a)}}(\tau > (b-a))] \\ &\le (1-\gamma)\mathbb{P}_x(\tau > (k-1)(b-a)) ,\end{aligned}$$

which by induction yields, for every $x \in A^c$,

$$\mathbb{P}_x(\tau > k(b-a)) \le (1-\gamma)^k .$$

For $n \ge (b-a)$, setting $n = k(b-a) + r$, with $r \in \{0, \dots, (b-a)-1\}$, we get for all $x \in A^c$,

$$\mathbb{P}_x(\tau > n) \le \mathbb{P}_x(\tau > k(b-a)) \le (1-\gamma)^k \le (1-\gamma)^{(n-(b-a))/(b-a)} .$$

3. Proposition 4.4.4 shows that $u_1(x) = \mathbb{E}_x[\tau]$ is the minimal solution to (4.6.1) with $g(x) = \mathbb{1}_{A^c}(x)$, $\alpha = 0$, and $\beta = 0$.
4. For $s = 2$ and every $x \in A^c$, we have

$$\begin{aligned}u_2(x) &= \mathbb{E}_x[\sigma^2] = \mathbb{E}_x[(1+\tau\circ\theta)^2] \\ &= 1 + 2\mathbb{E}_x[\tau\circ\theta] + \mathbb{E}_x[\tau^2\circ\theta] \\ &= 1 + 2\mathbb{E}_x[\mathbb{E}\left[\tau\circ\theta \,|\, \mathscr{F}_1\right]] + \mathbb{E}_x[\mathbb{E}\left[\tau^2\circ\theta \,\middle|\, \mathscr{F}_1\right]] \\ &= 1 + 2\mathbb{E}_x[\mathbb{E}_{X_1}[\tau]] + \mathbb{E}_x[\mathbb{E}_{X_1}[\tau^2]] \\ &= 1 + 2Pu_1(x) + Pu_2(x) .\end{aligned}$$

Therefore, u_2 is the finite solution to the system (4.6.1) with $g(x) = 1 + 2Pu_1(x)$ for $x \in A^c$ and $\alpha = \beta = 0$.

5. Similarly, for $x \in A^c$,

$$\begin{aligned}u_3(x) &= \mathbb{E}_x[\tau^3] = \mathbb{E}_x[(1+\tau\circ\theta_1)^3] \\ &= 1 + 3\mathbb{E}_x[\tau\circ\theta_1] + 3\mathbb{E}_x[\tau^2\circ\theta_1] + \mathbb{E}_x[\tau^3\circ\theta_1] \\ &= 1 + 3Pu_1(x) + 3Pu_2(x) + Pu_3(x),\end{aligned}$$

which implies that u_3 is the finite solution to the system (4.6.1) with $g(x) = 1 + 3Pu_1(x) + 3Pu_2(x)$ for $x \in A^c$, $\alpha = \beta = 0$.

6. Direct on writing the definitions.
7. By straightforward algebraic manipulations.
8. Applying (4.6.3), we obtain, for $x \in \{a+1,\dots,b\}$, $\phi(x) - \phi(x-1) = \rho^{x-a+1}$, which implies

$$\phi(x) = \sum_{y=a+1}^{x} \rho^{y-a+1} = \begin{cases} (1-\rho^{x-a})/(1-\rho) & \text{if } \rho \neq 1, \\ x-a & \text{otherwise.} \end{cases}$$

9. Equation (4.6.3) becomes, for $x \in \{a+1,\dots,b\}$,

$$\Delta\psi(x) = -p^{-1} \sum_{y=0}^{x-a-1} \rho^y g(x-y-1)\,,$$

and this yields

$$\psi(x) = -p^{-1} \sum_{z=a+1}^{x} \sum_{y=0}^{z-a-2} \rho^y g(x-y-1)\,. \tag{G.4}$$

10. Set

$$w = \alpha + \gamma\phi + \psi \tag{G.5}$$

with $\gamma = \{\phi(b)\}^{-1}(\beta - \alpha - \psi(b))$ (which is well defined, since $\phi(b) > 0$). By construction, $w(a) = \alpha$, $w(x) = Pw(x) + g(x)$ for all $x \in \{a+1,\dots,b-1\}$, and $w(b) = \alpha + \gamma\phi(b) + \psi(b) = \beta$.

4.6 Level dependent birth-and-death processes can be used to describe the position of a particle moving on a grid that at each step may remain at the same state or move to an adjacent state with a probability possibly depending on the state.

If $P(0,0) = 1$ and $p_x + q_x = 1$ for $x > 0$, this process may be considered a model for the size of a population, recorded each time it changes, p_x being the probability that a birth occurs before a death when the size of the population is x. Birth-and-death processes have many applications in demography, queueing theory, performance engineering, and biology. They may be used to study the size of a population, the number of diseases within a population, and the number of customers waiting in a queue for a service.

1. By Proposition 4.4.2, the function h is the smallest solution to the Dirichlet problem (4.4.1) with $f = \mathbf{1}$ and $A = \{0\}$. The equation $Ph(x) = h(x)$ for $x > 0$ yields

$$h(x) = p_x h(x+1) + q_x h(x-1)\,.$$

2. Note that h is decreasing and define $u(x) = h(x-1) - h(x)$. Then $p_x u(x+1) = q_x u(x)$, and we obtain by induction that $u(x+1) = \gamma(x)u(1)$ with $\gamma(0) = 1$ and

$$\gamma(x) = \frac{q_x q_{x-1} \cdots q_1}{p_x p_{x-1} \cdots p_1} \,.$$

This yields, for $x \geq 1$,

$$h(x) = h(0) - u(1) - \cdots - u(x) = 1 - u(1)\{\gamma(0) + \cdots + \gamma(x-1)\} \,.$$

3. If $\sum_{x=0}^{\infty} \gamma(x) = \infty$, the restriction $0 \leq h(x) \leq 1$ imposes $u(1) = 0$ and $h(x) = 1$ for all $x \in \mathbb{N}$.
4. If $\sum_{x=0}^{\infty} \gamma(x) < \infty$, we can choose $u(1) > 0$ such that $1 - u(1)\sum_{x=0}^{\infty} \gamma(x) \geq 0$. Therefore, the minimal nonnegative solution to the Dirichlet problem is obtained by setting $u(1) = (\sum_{x=0}^{\infty} \gamma(x))^{-1}$, which yields the solution

$$h(x) = \frac{\sum_{y=x}^{\infty} \gamma(y)}{\sum_{y=0}^{\infty} \gamma(y)} \,.$$

In this case, for $x \in \mathbb{N}^*$, we have $h(x) < 1$, so the population survives with positive probability.

4.8 1. u is harmonic on $\mathsf{X} \setminus \{-b, a\}$ by Theorem 4.1.3 (i). Thus for $x \in \mathsf{X} \setminus \{-b, a\}$,

$$u(x) = Pu(x) = \frac{1}{2}u(x-1) + \frac{1}{2}u(x+1) \,. \tag{G.6}$$

This implies that $u(x+1) - u(x) = u(x-1) - u(x)$ and

$$u(x) = u(-b) + (x+b)\{u(-b+1) - u(-b)\} \tag{G.7}$$

for all $x \in \mathsf{X} \setminus \{-b, a\}$. Since $u(a) = u(-b) = 1$, this yields $u(-b+1) = u(-b)$ and thus $u(x) = 1$, i.e., $\mathbb{P}_x(\tau < \infty) = 1$ for all $x \in \mathsf{X}$. Therefore, the game ends in finite time almost surely for every initial wealth $x \in \{-b, \ldots, a\}$.

2. We now compute the probability $u(x) = \mathbb{P}_x(\tau_a < \tau_{-b})$ of winning. We can also write $u(x) = \mathbb{E}_x[\mathbb{1}_a(X_\tau)]$. Theorem 4.4.5 (with $\beta = 1$ and $f = \mathbb{1}_a$) shows that u is the smallest nonnegative solution to the equations

$$\begin{cases} u(x) = Pu(x) \,, & x \in \mathsf{X} \setminus \{-b, a\} \,, \\ u(-b) = 0 \,, & u(a) = 1 \,. \end{cases}$$

3. We have established in (4.6.5) that the harmonic functions on $\mathsf{X} \setminus \{-b, a\}$ are given by $u(x) = u(-b) + (x+b)\{u(-b+1) - u(-b)\}$. Since $u(-b) = 0$, this yields $u(x) = (x+b)u(-b+1)$ for all $x \in \{-b, \ldots, a\}$. The boundary condition $u(a) = 1$ implies that $u(-b+1) = 1/(a+b)$. Therefore, the probability of winning when the initial wealth is x is equal to $u(x) = (x+b)/(a+b)$.
4. We will now compute the expected time of a game. Denote by $\tau = \tau_a \wedge \tau_{-b}$ the hitting time of the set $\{-b, a\}$. By Theorem 4.4.5, $u(x) = \mathbb{E}_x[\tau_C]$ is the smallest solution to the Poisson problem (4.4.4). This yields the following recurrence equation (which differs from (G.6) by an additional constant term): for $x \in \{-b+1, \ldots, a-1\}$,

$$u(x) = \tfrac{1}{2}u(x-1) + \tfrac{1}{2}u(x+1) + 1 \,. \tag{G.8}$$

5. The boundary conditions are $u(-b) = 0$ and $u(a) = 0$. Define $\Delta u(x-1) = u(x) - u(x-1)$ and $\Delta^2 u(x-1) = u(x+1) - 2u(x) + u(x-1)$. Equation (G.8) implies that for $x \in \{-b+1, \ldots, a-1\}$,

$$\Delta^2 u(x-1) = -2 \,. \tag{G.9}$$

6. The boundary conditions imply that the only solution to (G.9) is given by

$$u(x) = (a-x)(x+b) \,, \; x = -b, \ldots, a \,. \tag{G.10}$$

4.10 For $k \geq 0$, set $Z_k = r(k)h(X_k)$ and

$$\begin{aligned} &U_0 = \{PV_1(X_0) + r(0)h(X_0)\}\mathbb{1}_C(X_0) + V_0(X_0)\mathbb{1}_{C^c}(X_0) \,, && U_k = V_k(X_k) \,, \; k \geq 1, \\ &Y_0 = 0 \,, && Y_k = \infty \times \mathbb{1}_C(X_k) \,, \; k \geq 1, \end{aligned}$$

with the convention $\infty \times 0 = 0$. Then (4.6.8) yields, for $k \geq 0$ and $x \in \mathsf{X}$,

$$\mathbb{E}_x\left[U_{k+1} \,|\, \mathscr{F}_k\right] + Z_k \leq U_k + Y_k \qquad \mathbb{P}_x - \text{a.s.}$$

Hence (4.3.1) holds and (4.6.9) follows from the application of Theorem 4.3.1 with σ_C:

$$\begin{aligned} &\mathbb{E}_x\left[U_{\sigma_C}\mathbb{1}\{\sigma_C < \infty\}\right] + \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k)h(X_k)\right] \\ &\quad = \mathbb{E}_x\left[V_{\sigma_C}(X_{\sigma_C})\mathbb{1}\{\sigma_C < \infty\}\right] + \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} r(k)h(X_k)\right] \\ &\quad \leq \mathbb{E}_x[U_0] + \mathbb{E}_x\left[\sum_{k=0}^{\sigma_C-1} Y_k\right] = \{PV_1(x) + r(0)h(x)\}\mathbb{1}_C(x) + V_0(x)\mathbb{1}_{C^c}(x) \,. \end{aligned}$$

4.11 1. To prove (4.6.11), recall that $\sigma_C = 1 + \tau_C \circ \theta$ and $X_{\tau_C} \circ \theta = X_{\sigma_C}$. Applying the Markov property (Theorem 3.3.3) and these relations, we get

$$\begin{aligned} &PW_{n+1}(x) \\ &\quad = \mathbb{E}_x[\mathbb{E}_{X_1}[r(n+1+\tau_C)g(X_{\tau_C})\mathbb{1}_{\{\tau_C<\infty\}}]] + \mathbb{E}_x\left[\mathbb{E}_{X_1}\left[\sum_{k=0}^{\tau_C-1} r(n+k+1)h(X_k)\right]\right] \\ &\quad = \mathbb{E}_x[r(n+1+\tau_C\circ\theta)g(X_{\tau_C}\circ\theta)\mathbb{1}_{\{\tau_C\circ\theta<\infty\}}] + \mathbb{E}_x\left[\sum_{k=0}^{\tau_C\circ\theta-1} r(n+k+1)h(X_k\circ\theta)\right] \\ &\quad = \mathbb{E}_x[r(n+\sigma_C)g(X_{\sigma_C})\mathbb{1}_{\{\sigma_C<\infty\}}] + \mathbb{E}_x\left[\sum_{k=1}^{\sigma_C-1} r(n+k)h(X_k)\right] . \end{aligned}$$

Adding $r(n)h(x) = r(n)\mathbb{E}_x[h(X_0)]$ on both sides proves (4.6.11).

2. Applying Theorem 4.5.1 with $\tilde{h}(n,x) = r(n)h(x)$, $\tilde{g}(n,x) = g(n)r(x)$ (noting that $\tilde{U}(n,x) = W_n(x)$ for $n \geq 0$ and $x \in \mathsf{X}$) yields $V_n \geq W_n$ for all $n \geq 0$. The bound (4.6.13) follows from (4.6.11) with $n = 0$ and $W_0 \leq V_0$.

4.12 For $x \in \mathsf{X}$, define

$$W_C^{f,g,\delta}(x) = \mathbb{E}_x[g(X_{\tau_C})\mathbb{1}_{\{\tau_C<\infty\}}] + \mathbb{E}_x\left[\sum_{k=0}^{\tau_C-1} \delta^{k+1} f(X_k)\right] .$$

Applying Theorem 4.5.1, (4.5.3), with $m = 0$, $\tilde{g}(k,x) = \delta^k$, $\tilde{f}(k,x) = \delta^{k+1} f(x)$, we obtain

$$\delta^{-1} W_C^{f,g,\delta}(x) = \begin{cases} g(x), & x \in C, \\ PW_C^{f,g,\delta}(x) + f(x), & x \notin C . \end{cases}$$

Let V be a function satisfying (4.6.14) and $V(x) \geq g(x)$ for $x \in C$. Then (4.5.4) holds with $\tilde{v}(k,x) = \delta^k V(x)$, and we conclude by applying Theorem 4.5.1 that $V \geq W_C^{f,g,\delta}$.

Solutions to Exercises of Chapter 5

5.1 For $A \in \mathscr{X}$, it can be easily checked that $\{X_{2n} \in A, \text{ i.o.}\} \in \cap_{k \geq 0} \sigma(X_l, l > k)$ but $\{X_{2n} \in A, \text{ i.o.}\} \notin \mathscr{I}$.

5.2 The probabilities $\mathbb{P} \circ \mathrm{T}^{-1}$ and $\mathbb{P}$ coincide on $\mathscr{B}_0$, which is stable under finite intersection. The proof follows from Theorem B.2.6. We will show that they coincide on the sigma-field generated by $\mathscr{B}_0$, that is, $\mathscr{B}$. To achieve this aim, consider $\mathscr{C} = \{B \in \mathscr{B}, \mathbb{P}[\mathrm{T}^{-1}(B)] = \mathbb{P}[B]\}$. Under the assumptions of the lemma, $\mathscr{B}_0 \subset \mathscr{C}$. We now show that $\mathscr{C} = \mathscr{B}$ by applying the monotone class theorem. Note first that $\Omega \in \mathscr{C}$, since $\mathrm{T}^{-1}(\Omega) = \Omega$. Let $A \in \mathscr{C}$ and $B \in \mathscr{C}$ be such that $A \subset B$; since $\mathbb{P} \circ \mathrm{T}^{-1}$ and $\mathbb{P}$ are probabilities,

$$\mathbb{P}[\mathrm{T}^{-1}(B \setminus A)] = \mathbb{P}[\mathrm{T}^{-1}(B)] - \mathbb{P}[\mathrm{T}^{-1}(A)] = \mathbb{P}[B] - \mathbb{P}[A] = \mathbb{P}[B \setminus A] .$$

Finally, let $\{A_n, n \in \mathbb{N}\}$ be an increasing sequence of elements of $\mathscr{C}$. Then using classical properties of measures,

$$\mathbb{P} \circ \mathrm{T}^{-1}\left(\bigcup_{n\in\mathbb{N}} A_n\right) == \lim_{n\to\infty} \mathbb{P}[\mathrm{T}^{-1}(A_n)] = \lim_{n\to\infty} \mathbb{P}[A_n] = \mathbb{P}\left[\bigcup_{n\in\mathbb{N}} A_n\right] .$$

Therefore, $\mathscr{C}$ is a monotone class containing $\mathscr{B}_0$. Thus by the monotone class theorem, $\mathscr{B} = \sigma(\mathscr{B}_0) \subset \mathscr{C}$.

5.3 Let $A \in \mathscr{I}$. Since $\mathbb{1}_A = \mathbb{1}_A \circ \mathrm{T}^k$ and T is measure-preserving, we have

$$\mathbb{E}\left[\mathbb{1}_A\, Y \circ \mathrm{T}^k\right] = \mathbb{E}\left[\mathbb{1}_A \circ \mathrm{T}^k\; Y \circ \mathrm{T}^k\right] = \mathbb{E}\left[\mathbb{1}_A Y\right] = \mathbb{E}\left[\mathbb{1}_A \mathbb{E}\left[Y \,\middle|\, \mathscr{I}\right]\right] .$$

This implies that $\mathbb{E}\left[Y \circ \mathrm{T}^k \,\middle|\, \mathscr{I}\right] = \mathbb{E}\left[Y \,\middle|\, \mathscr{I}\right]$ $\mathbb{P}-\text{a.s.}$

5.4 1. Let $A \in \mathscr{I}$ and define $h(x) = \mathbb{E}_x[\mathbb{1}_A]$, $B = \{x \in \mathsf{X} : h(x) = 1\}$. By Proposition 5.2.2 (i), h is a nonnegative harmonic function bounded by 1. This implies that for $x \in B$, $\mathbb{E}_x[h(X_1)] = Ph(x) = h(x) = 1$. Thus for all $x \in B$, we get (using that if Z is a random variable taking values in $[0,1]$ and $\mathbb{E}[Z] = 1$, then $Z = 1$ $\mathbb{P}-\text{a.s.}$)

$$\mathbb{P}_x(X_1 \in B) = \mathbb{P}_x(h(X_1) = 1) = 1 .$$

Thus B is absorbing.

2. By Proposition 5.2.2 (iii), we know that $\mathbb{P}_\pi(\mathbb{E}_{X_0}[\mathbb{1}_A] = \mathbb{1}_A) = 1$, which implies that $\mathbb{P}_\pi(h(X_0) \in \{0,1\}) = 1$. This yields

$$\begin{aligned}\mathbb{P}_\pi(A) = \mathbb{E}_\pi[\mathbb{E}_{X_0}[\mathbb{1}_A]] &= \int_{\mathsf{X}} \pi(\mathrm{d}x) h(x) \\ &= \int_{\mathsf{X}} \pi(\mathrm{d}x) \mathbb{1}\{h(x) = 1\} = \int_{\mathsf{X}} \pi(\mathrm{d}x) \mathbb{1}_B(x) = \pi(B) .\end{aligned}$$

5.5 1. Since f is bounded, the convergence also holds on $\mathrm{L}^1(\mathbb{P}_\pi)$. Then since by the Markov property, $\mathbb{E}_\pi[Pf(X_k)] = \mathbb{E}_\pi[f(X_{k+1})]$, we have

$$\begin{aligned}\pi(Pf) &= \mathbb{E}_\pi\left[\lim_n \frac{1}{n}\sum_{k=0}^{n-1} Pf(X_k)\right] \\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{E}_\pi[Pf(X_k)] = \mathbb{E}_\pi\left[\lim_n \frac{1}{n}\sum_{k=1}^{n} f(X_k)\right] = \pi(f) ,\end{aligned}$$

which shows that π is invariant.

2. Let $A \in \mathscr{I}$. By Proposition 5.2.2 (iii), $\mathbb{1}_A = \mathbb{P}_{X_0}(A) = \mathbb{1}_B(X_0)$ $\mathbb{P}_\pi - \text{a.s.}$, where $B = \{x \in \mathsf{X} : \mathbb{P}_x(A) = 1\}$. Since for all $k \in \mathbb{N}$, $\mathbb{1}_A = \mathbb{1}_B(X_0) = \cdots = \mathbb{1}_B(X_k)$ $\mathbb{P}_\pi - \text{a.s.}$, we obtain

$$\mathbb{1}_B(X_0) = \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_B(X_k) \stackrel{\mathbb{P}_\pi\text{-a.s.}}{\longrightarrow} \pi(B) .$$

3. Therefore, $\pi(B) = \mathbb{P}_\pi(A) = 0$ or 1, i.e., the invariant σ-field is trivial for $\mathbb{P}_\pi$, and thus $(\mathsf{X}^{\mathbb{N}}, \mathscr{X}^{\otimes\mathbb{N}}, \mathbb{P}_\pi, \theta)$ is ergodic.

5.11 (i) Let $\varepsilon > 0$. There exist $m \in \mathbb{N}$ and a $\sigma(X_k, -m \le k \le m)$-measurable random variable, denoted by Z, satisfying $\mathbb{E}[|Y - Z|] < \varepsilon$. We also have $\mathbb{E}[|Y - Z\circ\theta^m|] < \varepsilon$ and $Z \circ \theta^m \in \sigma(X_k, k \ge 0)$. Therefore, we can construct a sequence $\{Z_n,\ n \in \mathbb{N}\}$ of $\sigma(X_k, k \ge 0)$-measurable random variables such that $\lim_{n\to\infty}\mathbb{E}[|Z_n - Y|] = 0$. Taking, if necessary, a subsequence, we can assume that $Z_n \stackrel{[\text{-a.s.}}{\longrightarrow} \mathbb{P}]Y$. Then

$U = \limsup_{n\to\infty} Z_n$ is $\sigma(X_k, k \geq 0)$-measurable and $Y = U \quad \mathbb{P}-$a.s. Hence Y is $\overline{\mathscr{F}_0^\infty}$-measurable. We treat the negative case in the same manner.

(ii) The σ-algebras $\overline{\mathscr{F}_0^\infty}$ and $\overline{\mathscr{F}_{-\infty}^0}$ are independent conditionally to X_0. This implies that

$$\mathbb{E}\left[Y^2 \,\middle|\, X_0\right] = \mathbb{E}\left[Y.Y \,|\, X_0\right] = \mathbb{E}\left[Y \,|\, X_0\right]\mathbb{E}\left[Y \,|\, X_0\right] = \{\mathbb{E}\left[Y \,|\, X_0\right]\}^2 .$$

The Cauchy–Schwarz inequality shows that

$$\begin{aligned}\{\mathbb{E}\left[Y^2\right]\}^2 &= \{\mathbb{E}\left[Y\mathbb{E}\left[Y \,|\, X_0\right]\right]\}^2 \leq \mathbb{E}\left[Y^2\right]\mathbb{E}\left[\{\mathbb{E}\left[Y \,|\, X_0\right]\}^2\right] \\ &= \mathbb{E}\left[Y^2\right]\mathbb{E}\left[\mathbb{E}\left[Y^2 \,\middle|\, X_0\right]\right] = \{\mathbb{E}\left[Y^2\right]\}^2 .\end{aligned}$$

Therefore, equality holds in the Cauchy–Schwarz inequality, and $Y = \lambda\mathbb{E}\left[Y \,|\, X_0\right]$ $\mathbb{P}-$a.s. Taking the expectation, we obtain $\lambda = 1$.

(iii) The proof is elementary and left to the reader.

5.12 Set $m \in \mathbb{N}^\star$. Every number n can be written in the form $n = q(n)m + r(n)$, where $r(n) \in \{0, \ldots, m-1\}$. We define $a_0 = 0$. Then we have $a_n = a_{q(n)m+r(n)} \leq q(n)a_m + a_{r(n)}$, and then we have

$$\frac{a_n}{n} = \frac{a_{q(n)m+r(n)}}{q(n)m+r(n)} \leq \frac{q(n)m}{q(n)m+r(n)}\,\frac{a_m}{m} + \frac{a_{r(n)}}{n} ,$$

which implies that

$$\inf_{n\in\mathbb{N}^\star} \frac{a_n}{n} \leq \liminf_{n\to\infty} \frac{a_n}{n} \leq \limsup_{n\to\infty} \frac{a_n}{n} \leq \frac{a_m}{m} .$$

Since this inequality is valid for all $m \in \mathbb{N}^\star$, the result follows.

5.13 By subadditivity of the sequence,

$$Y_n^+ \leq \left(\sum_{k=0}^{n-1} Y_1 \circ \mathrm{T}^k\right)^+ \leq \sum_{k=0}^{n-1} Y_1^+ \circ \mathrm{T}^k .$$

Now take the expectation on both sides of the previous inequality and use that T is measure-preserving. We obtain that $\mathbb{E}\left[Y_n^+\right] \leq n\mathbb{E}\left[Y_1^+\right] < \infty$. With a similar argument, $\mathbb{E}\left[Y_p^+ \circ \mathrm{T}^n\right] < \infty$. This implies that $\mathbb{E}\left[Y_{n+p}\right]$ and $\mathbb{E}\left[Y_p \circ \mathrm{T}^n\right]$ are well defined and

$$\mathbb{E}\left[Y_{n+p}\right] \leq \mathbb{E}\left[Y_n + Y_p \circ \mathrm{T}^n\right] = \mathbb{E}\left[Y_n\right] + \mathbb{E}\left[Y_p \circ \mathrm{T}^n\right] = \mathbb{E}\left[Y_n\right] + \mathbb{E}\left[Y_p\right] .$$

The proof of (5.3.1) and (5.3.2) follows from Fekete's lemma (see Exercise 5.12) applied to $u_n = \mathbb{E}\left[Y_n\right]$ and $u_n = \mathbb{E}\left[Y_n \,|\, \mathscr{I}\right]$ and Exercise 5.3.

5.14 We recall Parthasaraty's theorem (see (Parthasarathy 1967, Theorem 6.6)): Let (X, Υ) be a separable metric space. Then there exist a metric δ on X that induces the same topology on X as d and a countable set H of bounded and uniformly continuous (with respect to δ) functions such that for all $\{\mu, \mu_n, n \geq 1\} \subset \mathbb{M}_1(\mathsf{X})$, the following assertions are equivalent:

(i) μ_n converges weakly to μ.
(ii) For all $h \in H$, $\lim_{n\to\infty} \mu_n(h) = \mu(h)$.

1. By Parthasaraty's theorem, there exists a countable set $H \subset \mathrm{U}_b(\mathsf{X})$ such that for all $\mu, \mu_n \in \mathbb{M}_1(\mathscr{X}), n \geq 1$, the following two statements are equivalent:

   (i) $\mu_n \stackrel{\text{w}}{\Rightarrow} \mu$;
   (ii) for all $h \in H$, $\lim_{n\to\infty} \mu_n(h) = \mu(h)$.

   Now for all $h \in H$, since h is bounded, $\pi|h| < \infty$, and therefore by Theorem 5.2.9, $\mathbb{P}\left(\left\{\omega \in \Omega \,:\, \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} h(X'_k(\omega)) = \pi(h)\right\}\right) = 1$. The proof follows since H is countable.
2. Set $B = \{\omega \in \Omega \,:\, \lim_{n\to\infty} \mathsf{d}(X_n(\omega), X'_n(\omega)) = 0\}$ and

$$A' = \left\{ \omega \in \Omega \,:\, \forall h \in H \,,\ \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} h(X'_k(\omega)) = \pi(h) \right\} .$$

   Then $A \cap B \subset A'$. Since $\mathbb{P}(A) = \mathbb{P}(B) = 1$, we deduce $\mathbb{P}(A') = 1$, and applying again Parthasaraty's theorem, for all $\omega \in A'$, the sequence of measures $\mu_n(\omega) = n^{-1} \sum_{k=1}^{n} \delta_{X_k(\omega)}$ converges weakly to π. The result follows.
3. By Theorem 5.2.9,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} V(X'_k) = \pi(V) \,, \qquad \mathbb{P} - \text{a.s.}$$

   For every $\alpha > 0$,

$$|n^{-1} \sum_{k=1}^{n} \{V(X'_k) - V(X_k)\}| \leq \sup\left\{|V(x) - V(x')| \,:\, \mathsf{d}(x,x') \leq \alpha\right\} \\ + n^{-1} \sum_{k=1}^{n} |V(X'_k) - V(X_k)| \mathbb{1}\left\{\mathsf{d}(X'_k, X_k) > \alpha\right\} .$$

   The first term on the right-hand side can be made arbitrarily small, since f is uniformly continuous. Moreover, since $\mathsf{d}(X_n, X'_n) \xrightarrow{\mathbb{P}\text{-a.s.}} 0$, the series in the second term on the right-hand side contains only a finite number of positive terms $\mathbb{P} -$ a.s., and therefore the second term on the right-hand side tends to 0 $\mathbb{P} -$ a.s. as n goes to infinity. Finally,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} V(X_k) = \pi(V) \,, \qquad \mathbb{P} - \text{a.s.}$$

4. There exists $\bar{\Omega}$ such that $\mathbb{P}(\bar{\Omega}) = 1$ and for all $\omega \in \bar{\Omega}$,

$$\mu_n(\omega) = n^{-1} \sum_{k=1}^{n} \delta_{X_k(\omega)}$$

converges weakly to π and $\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} V(X_k(\omega)) = \pi(V)$. For all $\omega \in \bar{\Omega}$,

$$\begin{aligned}\frac{1}{n}\sum_{k=0}^{n-1}\{V - V \wedge M\}(X_k(\omega)) &= \frac{1}{n}\sum_{k=0}^{n-1} V(X_k(\omega)) - \frac{1}{n}\sum_{k=0}^{n-1} V \wedge M(X_k(\omega)) \\ &\to \pi(V) - \pi(V \wedge M) \;,\end{aligned}$$

showing that

$$\lim_{M\to\infty}\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1}\{V - V \wedge M\}(X_k(\omega)) = 0 \;. \tag{G.0.11}$$

Without loss of generality, we assume that $0 \le f \le V$. Note that

$$\begin{aligned}\frac{1}{n}\sum_{k=0}^{n-1} f(X_k(\omega)) - \pi(f) &= \frac{1}{n}\sum_{k=0}^{n-1}\{f - f \wedge M\}(X_k(\omega)) \\ &+ \frac{1}{n}\sum_{k=0}^{n-1}\{f \wedge M(X_k(\omega)) - \pi(f \wedge M)\} + \pi(f \wedge M) - \pi(f) \;.\end{aligned}$$

Since the function $x \mapsto x - x \wedge M$ is nondecreasing, we have $\{f - f \wedge M\}(X_k) \le \{V - V \wedge M\}(X_k)$ and (G.11) implies

$$\lim_{M\to\infty}\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1}\{f - f \wedge M\}(X_k(\omega)) = 0 \;.$$

On the other hand, since the function $f \wedge M$ is bounded and continuous, we obtain

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1}\{f \wedge M(X_k(\omega)) - \pi(f \wedge M)\} = 0.$$

The proof is complete by noting that $\lim_{M\to\infty} \pi(f \wedge M) = \pi(f)$ and $\mathbb{P}(\bar{\Omega}) = 1$.

Solutions to Exercises of Chapter 6

6.1 1. The transition matrix is given by $P(0,0) = 1$, and for $j \in \mathbb{N}^*$ and $k \in \mathbb{N}$,

$$P(j,k) = \sum_{(k_1,\ldots,k_j)\in\mathbb{N}^j, k_1+\cdots+k_j=k} \nu(k_1)\nu(k_2)\cdots\nu(k_j) = \nu^{*j}(k) \;.$$

The state 0 is absorbing and the population is forever extinct if it reaches zero.

2. The state 0 is absorbing and hence recurrent. Since we assume that $\nu(0) > 0$, we have $P(x,0) = \nu^x(0) > 0$ for every $x \in \mathbb{N}^*$ and thus $\mathbb{P}_x(\sigma_0 < \infty) > 0$ for all $x \in \mathbb{N}$, which implies that $\mathbb{P}_x(\sigma_x = \infty) > 0$. Thus 0 is the only recurrent state.
3. The Markov property yields $\mathbb{E}_x[X_{k+1}] = \mathbb{E}_x[\mathbb{E}_x[X_{k+1} \,|\, X_k]] = \mu \mathbb{E}_x[X_k]$, and by induction we obtain that $\mathbb{E}_x[X_k] = x\mu^k$ for all $k \geq 0$.
4. Note indeed that $\{\tau_0 = \infty\} = \bigcap_{k=0}^{\infty}\{X_k \geq 1\}$, and the population does not disappear if there is at least one individual in the population at each generation. Since $\{X_{k+1} \geq 1\} \subset \{X_k \geq 1\}$, we get for all $x \in \mathbb{N}$,
$$\mathbb{P}_x(\tau_0 = \infty) = \lim_{k\to\infty} \mathbb{P}_x(X_k \geq 1) \leq \lim_{k\to\infty} \mathbb{E}_x[X_k] = \lim_{k\to\infty} x\mu^k = 0 \,.$$
5. $p_0 = 0$ and $p_1 < 1$: in this case, the population diverges to infinity with probability 1.

6.2 1. By the Markov property, we get
$$\begin{aligned}\Phi_{k+1}(u) &= \mathbb{E}\left[u^{X_{k+1}}\right] = \mathbb{E}\left[\mathbb{E}\left[u^{X_{k+1}} \,\middle|\, X_k\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[u^{\sum_{j=1}^{X_k}\xi_j^{(k+1)}} \,\middle|\, X_k\right]\right] = \mathbb{E}\left[\prod_{k=1}^{X_k}\mathbb{E}\left[u^{\xi_j^{(k+1)}}\right]\right] = \Phi_k(\phi(u))\,.\end{aligned}$$
By induction, we obtain that
$$\Phi_k(u) = \underbrace{\phi \circ \cdots \circ \phi(u)}_{k\,\text{times}}\,,$$
and thus it is also the case that $\Phi_{k+1}(u) = \phi(\Phi_k(u))$.
2. $\Phi_n(0) = \mathbb{E}\left[0^{X_n}\right] = \sum_{k=0}^{\infty}\mathbb{E}\left[0^k \mathbb{1}_{\{X_n = k\}}\right] = \mathbb{P}(X_n = 0)$. By the nature of the Galton–Watson process, these probabilities are nondecreasing in n, because if $X_n = 0$, then $X_{n+1} = 0$. Therefore, we have the limit $\lim_{n\to\infty}\Phi_n(0) = 0$. Finally, $\{\sigma_0 < \infty\} = \bigcup_{n=1}^{\infty}\{X_n = 0\}$.
3. By the continuity of φ, we have
$$\begin{aligned}\varphi(\rho) &= \varphi\left(\lim_{n\to\infty}\Phi_n(0)\right) = \lim_{n\to\infty}\varphi(\Phi_n(0)) \\ &= \lim_{n\to\infty}\Phi_{n+1}(0) = \rho\,.\end{aligned}$$
Finally, it remains to show that ρ is the smallest nonnegative root of the fixed-point equation. This follows from the monotonicity of the probability generating functions $\Phi_n(0)$: since $\zeta \geq 0$,
$$\Phi_n(0) \leq \Phi_n(\zeta) = \zeta\,.$$
Taking the limit of each side as $n \to \infty$ shows that $\rho \leq \zeta$.

6.3 Since $\sum_{n=2}^{\infty} b_n > 0$, we have

$$\phi(0) = b_0 \,,\ \phi(1) = 1 \,,\ \phi'(s) = \sum_{n\geq 1} n b_n s^{n-1} > 0 \,,\ \phi''(s) = \sum_{n\geq 2} n(n-1) b_n s^{n-2} > 0 \,.$$

Thus the function ϕ is continuous and strictly convex on $[0,1]$. Note also that the left derivative of ϕ at 1 is $\phi'(1) = \sum_{n\geq 0} n b_n = \mu$ (and by convexity, this makes sense also if $\mu = \infty$).

- If $\mu \leq 1$, then by convexity, the graph of ϕ stands above the diagonal on $[0,1)$.
- If $\mu > 1$, then by convexity, the graph of ϕ is below the diagonal on an interval $(1-\varepsilon, 1]$, and since $\phi(0) > 0$, by the mean value theorem, there must exist $s \in (0,1)$ such that $\phi(s) = s$, that is, the graph of ϕ crosses the diagonal at s.

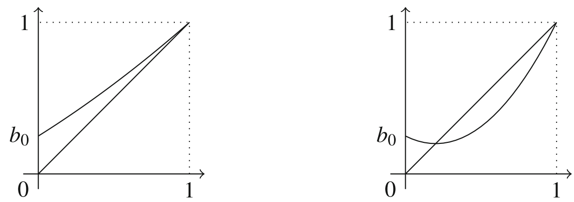

Fig. G.0.1 The cases $\mu < 1$ (left panel) and $\mu > 1$ (right panel).

6.4 1. For $n \in \mathbb{N}$, $P^n(0,x)$ is the probability of an n-step transition from 0 to x (the probability that a "particle," starting at zero, finds itself after n iterations at x). Suppose that n and x are both even or odd and that $|x| \leq n$ (otherwise, $P^n(0,x) = 0$). Then $P^n(0,x)$ is the probability of $(x+n)/2$ successes in n independent Bernoulli trials, where the probability of success is p. Therefore,

$$P^n(0,x) = p^{(n+x)/2} q^{(n-x)/2} \binom{n}{(n+x)/2} \,,$$

where the sum $n+x$ is even and $|x| \leq n$ and $P^n(0,x) = 0$ otherwise.

2. If the chain starts at 0, then it cannot return at 0 after an odd number of steps, so $P^{2n+1}(0,0) = 0$. Any given sequence of steps of length $2n$ from 0 to 0 occurs with probability $p^n q^n$, there being n steps to the right and n steps to the left, and the number of such sequences is the number of way of choosing n steps to the right in $2n$ moves. Thus

$$P^{2n}(0,0) = \binom{2n}{n} p^n q^n \,.$$

3. The expected number of visits to state 0 for the random walk started at 0 is therefore given by

$$U(0,0) = \sum_{k=0}^{\infty} P^k(0,0) = \sum_{k=0}^{\infty} \binom{2k}{k} p^k q^k \,.$$

4. Applying Stirling's formula $(2k)! \sim \sqrt{4\pi k}(2k/\mathrm{e})^{2k}$ $k! \sim \sqrt{2\pi k}(k/\mathrm{e})^k$ yields
$$P^{2k}(0,0) = \binom{2k}{k} p^k q^k \sim_{k\to\infty} (4pq)^k (\pi k)^{-1/2} .$$
5. If $p \neq 1/2$, then $4pq < 1$, and the series $U(0,0)$ is summable. The expected number of visits to 0 when the random walk is started at 0 is finite. The state $\{0\}$ is transient, and all the atoms are transient.
6. If $p = 1/2$, then $4pq = 1$ and $P^{2k}(0,0) \sim_{k\to\infty} (\pi k)^{-1/2}$, so that $U(0,0) = \sum_{n=0}^{\infty} P^n(0,0) = +\infty$. The state 0 is therefore recurrent, and all the accessible sets are recurrent.
7. The counting measure on $\mathbb{Z}$ is an invariant measure. Since $\lambda(\mathsf{X}) = \infty$, the Markov kernel is therefore null recurrent.

6.5 1. Then $\{X_n^+, n \in \mathbb{N}\}$ and $\{X_n^-, n \in \mathbb{N}\}$ are independent simple symmetric random walks on $2^{-1/2}\mathbb{Z}$, and $X_n = (0,0)$ if and only if $X_n^+ = X_n^- = 0$. Therefore,
$$P^{(2n)}((0,0),(0,0)) = \left(\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\right)^2 \sim \frac{1}{\pi n} ,$$
as $n \to \infty$ by Stirling's formula.
2. $\sum_{n=0}^{\infty} P^n(0,0) = \infty$, and the simple symmetric random walk on $\mathbb{Z}^2$ is recurrent.
3. If the chain starts at 0, then it can return to zero only after an even number of steps, say $2n$. Of these $2n$ steps, there must be i up, i down, j north, j south, k east, k west for some $i,j,k \geq 0$ such that $i+j+k = n$. This yields
$$P^{2n}(0,0) = \sum_{\substack{i,j,k\geq 0\\ i+j+k=n}} \frac{(2n)!}{(i!j!k!)^2}\frac{1}{6^{2n}} = \binom{2n}{n}\frac{1}{2^{2n}} \sum_{\substack{i+j+k=n\\ i,j,k\geq 0}} \binom{n}{i\,j\,k}^2 \frac{1}{3^{2n}} .$$
Note now that
$$\sum_{\substack{i+j+k=n\\ i,j,k\geq 0}} \binom{n}{i\,j\,k} = 3^n ,$$
and if $n = 3m$, then
$$\binom{n}{i\,j\,k} = \frac{n!}{i!j!k!} \leq \frac{n!}{(m!)^3} ,$$
for all i,j,k such that $i+j+k = 3m$. Thus applying Stirling's formula, we obtain
$$P^{2n}(0,0) \leq \binom{2n}{n}\left(\frac{1}{2}\right)^{2n} \frac{n!}{(m!)^3}\left(\frac{1}{3}\right)^n \sim \frac{1}{2\sqrt{2\pi}^3}\left(\frac{6}{n}\right)^{3/2} .$$
Hence $\sum_{m=0}^{\infty} P^{6m}(0,0) < \infty$.

4. Since $P^{6m}(0,0) \geq (1/6)^2 P^{(6m-2)}(0,0)$ and $P^{6m}(0,0) \geq (1/6)^4 P^{(6m-4)}(0,0)$, this proves that $U(0,0) < \infty$, and the three-dimensional simple random walk is transient. In fact, it can be shown that the probability of return to the origin is about 0.340537329544.

6.6 Note first that the set I is stable by addition. We set $d_+ = g.c.d.(S_+)$ and $d_- = g.c.d.(S_-)$. Under the stated assumption, $g.c.d.(d_+, d_-) = 1$. Let $I_+ = \{x \in \mathbb{Z}_+,\, x = x_1 + \ldots + x_n,\, n \in \mathbb{N}^*,\, x_1, \ldots, x_n \in S_+\}$. By Lemma 6.3.2, there exists n_+ such that for all $n \geq n_+$, $d_+ n \in I_+ \subset I$; similarly, there exists n_- such that for all $n \geq n_-$, $-d_- n \in I$. Now, by Bézout's identity, $pd_+ + qd_- = 1$ for some $p, q \in \mathbb{Z}$. Then for all $r \in \mathbb{Z}^*$ and $k \in \mathbb{Z}$,

$$r = r(p - kd_-)d_+ + r(q + kd_+)d_- \;.$$

If $r > 0$ (resp. $r < 0$) for $-k$ large (resp. for k large), $r(p - kd_-) \geq n_+$ and $-r(q + kd_+) \geq n_-$, showing that $r \in I$. Furthermore, $0 = r - r \in I$.

6.7 1. If $m \neq 0$, then $\lim_{n\to\infty} X_n/n = \operatorname{sign}(m) \times \infty$ $\mathbb{P}$ – a.s. by the law of large numbers. This implies that $S_n \to \infty$ with the sign of m almost surely, and the Markov kernel P is therefore transient.

2. Since $\nu \neq \delta_0$, this implies that there exist integers $z_1 > 0$ and $z_2 < 0$ such that $\nu(z_1)\nu(z_2) > 0$. By Exercise 6.6, for every $x \in \mathbb{Z}$, there exist $z_1, \ldots, z_k \in \mathsf{S}$ such that $x = z_1 + \cdots + z_n$. Therefore,

$$\begin{aligned}\mathbb{P}_0(X_n = x) &\geq P(0, z_1) P(z_1, z_1 + z_2) \ldots P(z_1 + \cdots + z_{n-1}, x) \\ &= \nu(z_1) \ldots \nu(z_n) > 0 \,,\end{aligned}$$

which proves that $\mathbb{P}_0(\sigma_x < \infty) > 0$. Since for all $x, y \in \mathbb{Z}$, $\mathbb{P}_x(\sigma_y < \infty) = \mathbb{P}_0(\sigma_{y-x} < \infty)$, the proof follows.

3. Let $\varepsilon > 0$. By the law of large numbers, $\lim_{k\to\infty} \mathbb{P}_0(k^{-1}|X_k| \leq \varepsilon) = 1$ for all $\varepsilon > 0$. Hence by Cesàro's theorem, $\lim_{n\to\infty} n^{-1} \sum_{k=1}^n \mathbb{P}_0(|X_k| \leq \varepsilon k) = 1$, for all $\varepsilon > 0$. Since $\mathbb{P}_0(|X_k| \leq \varepsilon k) \leq \mathbb{P}_0(|X_k| \leq \lfloor \varepsilon n \rfloor)$ for all $k \in \{0, \ldots, n\}$, we get

$$\begin{aligned}1 &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n \mathbb{P}_0(|X_k| \leq \varepsilon k) \\ &\leq \liminf_{n\to\infty} \frac{1}{n} \sum_{k=1}^n \mathbb{P}_0(|X_k| \leq \lfloor \varepsilon n \rfloor) \leq \liminf_{n\to\infty} \frac{1}{n} U(0, [-\lfloor \varepsilon n \rfloor, \lfloor \varepsilon n \rfloor]) \,.\end{aligned}$$

4. By the maximum principle, for all $i \in \mathbb{Z}$, we have $U(0,i) \leq U(i,i) = U(0,0)$. Therefore, we obtain

$$\begin{aligned}1 &= \liminf_{n\to\infty} \frac{1}{n} \sum_{i=-\lfloor \varepsilon n \rfloor}^{\lfloor \varepsilon n \rfloor - 1} U(0,i) \\ &\leq \liminf_{n\to\infty} \frac{2\lfloor \varepsilon n \rfloor}{n} U(0,0) = 2\varepsilon U(0,0) \,.\end{aligned}$$

Since ε is arbitrary, this implies $U(0,0) = \infty$, and the chain is recurrent.

6.10 Define $Y_0 = 0$ and for $k \geq 0$, $Y_{k+1} = (Y_k + Z_{k+1})^+$. The proof is by a recurrence. We have $X_0^+ = Y_0 = 0$. Now assume that $X_{k-1}^+ = Y_{k-1}$. Then since $r = \mathbb{1}_{\mathbb{R}^+}$,

$$X_k = X_{k-1} \mathbb{1}\{X_{k-1} \geq 0\} + Z_k = Y_{k-1} + Z_k \ .$$

This implies $X_k^+ = (Y_{k-1} + Z_k)^+ = Y_k$.

Solutions to Exercises of Chapter 7

7.1 Let $\mathsf{a} \in \mathsf{X}_P^+$ and $x \in \mathsf{X}$ be such that there exists $n \geq 1$ such that $P^n(\mathsf{a}, x) > 0$. Let $y \in \mathsf{X}$. Since a is accessible, there exists $k \in \mathbb{N}$ such that $P^k(y, \mathsf{a}) > 0$. This yields

$$P^{n+k}(y,x) \geq P^k(y,\mathsf{a})P^n(\mathsf{a},x) > 0 \ .$$

Thus x is accessible, and so X_P^+ is absorbing. Let now x be an inaccessible state a,nd let a be an accessible state. Then $\mathbb{P}_x(\sigma_{\mathsf{a}} < \infty) > 0$. Since X_P^+ is absorbing, we have $\mathbb{P}_x(\sigma_x = \infty) \geq \mathbb{P}_x(\sigma_{\mathsf{a}} < \infty) > 0$. Thus x is transient.

7.2 A transient kernel may indeed have an invariant probability, which is necessarily infinite. Consider, for instance, the symmetric simple random walk on Z that is irreducible, transient, and admits an invariant infinite measure.

7.4
1. Recall that by definition, $P(0,0) = 1$ and $P(N,N) = 1$, i.e., both states 0 and N are absorbing. The chain is therefore not irreducible (there is no accessible atom).
2. For $x \in \{1,\ldots,N-1\}$, $\mathbb{P}_x(\sigma_x = \infty) \geq \mathbb{P}_x(X_1 = 0) + \mathbb{P}_x(X_1 = N) > 0$.
3. The distribution of X_{n+1} given X_n is $\mathrm{Bin}(N, X_n/N)$, and thus $\mathbb{E}[X_{n+1} \mid X_n] = X_n$. Thus $\{X_n,\ n \in \mathbb{N}\}$ is a martingale. Since it is uniformly bounded, by the martingale convergence theorem, Theorem E.3.4, it converges $\mathbb{P}_x -$ a.s. and in $\mathrm{L}^1(\mathbb{P}_x)$ for all $x \in \{0,\ldots,N\}$.
4. Since the total number of visits to $x \in \{1,\ldots,N-1\}$ is finite, $X_\infty \in \{0,N\}$ $\mathbb{P} -$ a.s. necessarily takes its values in $\{0,\ N\}$. Since $\{X_n\}$ converges to X_∞ $\mathrm{L}^1(\mathbb{P}_x)$, we obtain

$$x = \mathbb{E}_x[X_0] = \mathbb{E}_x[X_\infty] = N \cdot \mathbb{P}_x(X_\infty = N) \ ,$$

so that $\mathbb{P}_x(X_\infty = N) = x/N$ and $\mathbb{P}_x(X_\infty = 0) = 1 - x/N$.

7.5
1. We have $P^2(0,0) = qp > 1$, and for $x \geq 1$, $P^x(x,0) = q^x$, showing that $\{0\}$ is accessible. On the other hand, for all $x \geq 1$, $P^x(0,x) = p^x$. Hence all the states communicate.

2. By Corollary 4.4.7, the function f defined on $\mathbb{N}$ by $f(x) = \mathbb{P}_x(\tau_0 < \infty)$, $x \in \mathbb{N}$, is the smallest nonnegative function on $\mathbb{N}$ such that $f(0) = 1$ and $Pf(x) = f(x)$ on $\mathbb{Z}_+$. This is exactly (7.7.1).
3. Since $p+q+r=1$, this relation implies that for $x \geq 1$,
$$q\{f(x)-f(x-1)\} = p\{f(x+1)-f(x)\},$$
showing that $f(x)-f(x-1) = (p/q)^{x-1}(f(1)-1)$ for $x \geq 1$. Therefore, $f(x) = c_1 + c_2(q/p)^x$ if $p \neq q$ and $f(x) = c_1 + c_2 x$ if $p = q$ for constants c_1 and c_2 to be determined.
4. Assume first that $p < q$. Then $c_2 \geq 0$, since otherwise, $f(x)$ would be strictly negative for large values of x. The smallest positive solution is therefore of the form $f(x) = c_1$; the condition $f(0) = 1$ implies that the smallest positive solution to (7.7.1) taking the value 1 at 0 is $f(x) = 1$ for $x > 0$. Therefore, if $p < q$, then for all $x > 1$, the chain starting at x visits 0 with probability 1.
5. Assume now that $p > q$. In this case, $c_1 \geq 0$, since $\lim_{x\to\infty} f(x) = c_1$. The smallest nonnegative solution is therefore of the form $f(x) = c_2(q/p)^x$, and the condition $f(0) = 1$ implies $c_2 = 1$. Therefore, starting at $x \geq 1$, the chain visits the state 0 with probability $f(x) = (q/p)^x$ and never visits 0 with probability $1-(q/p)^x$.
6. If $p = q$, $f(x) = c_1 + c_2 x$. Since $f(x) = 0$, we have $c_1 = 1$, and the smallest positive solution is obtained for $c_2 = 0$. The hitting probability of 0 is therefore $f(x) = 1$ for every $x = 1, 2, \ldots$, as in the case $p < q$.

7.6 1. Applying a third-order Taylor expansion to $V(y)-V(x)$ for $y \in \mathbb{Z}^d$ such that $|y-x| \leq 1$ and summing over the $2d$ neighbors, we obtain, for $x \in \mathbb{Z}^d$ such that $|x| \geq 2$,
$$PV(x) - V(x) = 4\alpha(2\alpha - 2 + d)\,|x|^{2\alpha-2} + r(x)\ ,$$
where (constants may take different values on each appearance)
$$\begin{aligned} |r(x)| &\leq C \sup_{\|y-x\|\leq 1} |V^{(3)}(y)| \leq \sup_{\|y-x\|\leq 1} \|y\|^{2\alpha-3} \\ &\leq C\,|x|^{2\alpha-3}\ . \end{aligned}$$
If $|y-x| \leq 1$ and $|x| \geq 2$, then $\|y\| \geq \|x\|/2$,
$$PV(x) - V(x) = 2\alpha\{2\alpha - 2 + d + r(x)\}\,|x|^{2\alpha-2}\ ,$$
with $|r(x)| \leq C(\alpha, d)\,|x|^{-1}$.
2. If $d = 1$, then for each $\alpha \in (0, 1/2)$, we can choose M such that $PV(x) - V(x) \leq 0$ for $|x| \geq M$. Applying Theorem 7.5.2 yields that the one-dimensional simple symmetric random walk is recurrent.
3. If $d \geq 3$, then for each $\alpha \in (-1/2, 0)$, we can choose M such that $PV(x) - V(x) \leq 0$ if $|x| \geq M$. Moreover, since $\alpha < 0$, $\inf_{|x|\leq M} W(x) \geq M^\alpha$, and for each

x_0 such that $|x_0| > M$, $V(x_0) = |x_0|^\alpha < M^\alpha$. So we can apply Theorem 7.5.1 to obtain that the d-dimensional simple symmetric random walk is transient.

4. By a Taylor expansion, we can show that if $|x| \geq 2$, then

$$PW(x) - W(x) = \{4\alpha(\alpha-1) + O(|x|^{-1})\}\{\log(|x|^2)\}^{\alpha-2}|x|^{-2} .$$

Therefore, we can choose M such that $PW(x) - W(x) \leq 0$ if $|x| \geq M$, and Theorem 7.5.2 shows that the chain is recurrent.

7.8 1. The transition kernel of the chain is given by $P(0,y) = a_y$ for $y \in \mathbb{N}$ and for $x \geq 1$,

$$P(x,y) = \begin{cases} a_{y-x+1} & \text{if } y \geq x-1 , \\ 0 & \text{otherwise.} \end{cases}$$

2. If $a_0 = 1$, there is no client entering into service. If $a_0 + a_1 = 1$, then there is at most one client entering into service, and the number of clients in service will always decrease, unless $a_0 = 0$, in which case the number of clients remains constant.
3. By assumption, there exists $k_0 > 1$ such that $a_{k_0} > 0$. For $k \in \mathbb{N}$, let m be the unique integer such that $k_0 + m(k_0-1) \geq k > k_0 + (m-1)(k_0-1)$ and set $r = k_0 + m(k_0-1) - k$. Then $0 \to k_0 \to 2k_0 - 1 \to \cdots \to k_0 + m(k_0-1) \to k_0 + m(k_0-1) - 1 \to \cdots \to k_0 + m(k_0-1) - r = k$. Formally,

$$P(0,k) \geq P(0,k_0)P(k_0, 2k_0-1)\cdots P(k_0 + m(k_0-1))a_0^r = a_{k_0}^m a_0^r > 0 .$$

Thus $0 \to k$ and $k \to i$ for all $i \leq k$. This proves that all the states communicate and the Markov kernel is irreducible.

4. We have

$$\begin{aligned} PW(x) &= \sum_{y=0}^{\infty} P(x,y)b^y = \sum_{y=x-1}^{\infty} a_{y-x+1}b^y \\ &= b^{x-1}\sum_{y=x-1}^{\infty} a_{y-x+1}b^{y-x+1} = b^{x-1}\sum_{y=0}^{\infty} a_y b^y = b^{x-1}\varphi(b) . \end{aligned}$$

If $m > 1$, the mean number of clients entering into service is strictly greater than the number of clients processed in one unit of time. In that case, we will prove that the chain is transient and the number of clients in the queue diverges to infinity, whence each individual state is visited almost surely a finite number of times. If $m < 1$, then we will prove that the chain is recurrent. Consider first the case $m > 1$.

5. Exercise 6.3 shows that there exists a unique $b_0 \in (0,1)$ such that $\phi(b_0) = b_0$.
6. Set $F = \{0\}$ and $W(x) = b_0^x$ for $x \in \mathbb{N}$. Then $PW(x) = W(x)$ and $W(x) < W(0) = 1$ for all $x \in F^c$. Thus the assumptions of Theorem 7.5.1 are satisfied, and we can conclude that the Markov kernel P is transient.
7. Recall that $m > 1$ and $V(x) = x$. For all $x > 0$, we have

$$PV(x) = \sum_{y=x-1}^{\infty} a_{y-x+1} y = \sum_{y=x-1}^{\infty} a_{y-x+1}(y-x+1) + x - 1$$
$$= \sum_{k=0}^{\infty} k a_k - 1 + x = V(x) - (1-m) \leq V(x) \; . \tag{G.11}$$

8. Define $V_m(x) = x/(1-m)$ for $x \geq 0$. Then (7.7.2) can be rewritten as $PV_m(x) \leq V_m(x) - 1$ for every $x > 0$. Moreover,

$$PV_m(0) = (1-m)^{-1} \sum_{k=0}^{\infty} k a_k \leq m/(1-m) \; .$$

Thus we can apply Theorem 7.5.3 to conclude that the Markov kernel P is positive if $m < 1$.

7.11 Set $F = \{V \leq r\}$ and

$$W(x) = \begin{cases} (|V|_\infty - V(x))/(|V|_\infty - r) \; , & x \in F^c \; , \\ 1 \; , & x \in F \; . \end{cases} \tag{G.12}$$

Since by assumption, $\{V > r\}$ is nonempty and V is bounded, we have $|V|_\infty > r$. Thus W is well defined, nonnegative, and

$$\begin{aligned} PW(x) &= \mathbb{E}_x[W(X_1)] = \mathbb{E}_x[\mathbb{1}_{F^c}(X_1)W(X_1)] + \mathbb{E}_x[\mathbb{1}_F(X_1)W(X_1)] \\ &= \mathbb{E}_x\left[\frac{|V|_\infty - V(X_1)}{|V|_\infty - r}\right] + \mathbb{E}_x\left[\mathbb{1}_F(X_1)\left(1 - \frac{|V|_\infty - V(X_1)}{|V|_\infty - r}\right)\right] \\ &= \frac{|V|_\infty - PV(x)}{|V|_\infty - r} + \mathbb{E}_x\left[\mathbb{1}_F(X_1)\frac{V(X_1) - r}{|V|_\infty - r}\right] \leq \frac{|V|_\infty - PV(x)}{|V|_\infty - r} \; . \end{aligned}$$

By assumption, if $x \in F^c$, then $PV(x) \geq V(x)$. Thus the previous inequality implies that $PW(x) \leq W(x)$ for $x \notin F$. On the other hand, $W(x) = 1$ for $x \in F$, and since $\{V > r\}$ is accessible, there exists $x_0 \in F^c$ such that $W(x_0) < 1 = \inf_{x \in F} W(x)$. Therefore, Theorem 7.5.1 applies, and P is transient.

7.12 Since $f \geq 1$, the kernel P is positive by Theorem 7.5.3. Applying Proposition 4.3.2 and Theorem 7.2.1 yields, for every $x \in \mathsf{X}$,

$$\pi(f) = \frac{1}{\mathbb{E}_x[\sigma_x]}\mathbb{E}_x\left[\sum_{k=0}^{\sigma_F - 1} f(X_k)\right] < \frac{V(x) + b}{\mathbb{E}_x[\sigma_x]} < \infty \; .$$

Solutions to Exercises of Chapter 8

8.1 We have $u(n) = p$ for $n \geq 1$, showing that $u(z) = (1-(1-p)z)/(1-z)$. Hence $B(z) = pz/(1-(1-p)z)$ by virtue of (8.1.10).

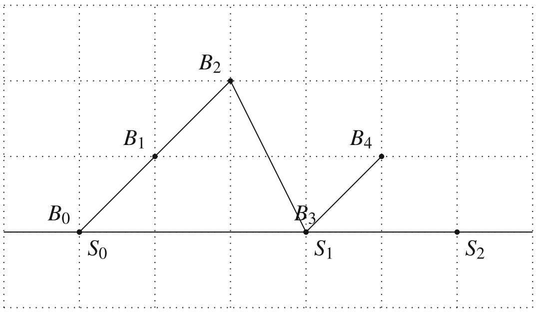

Fig. G.0.2 An example of an age process. If for some $k \in \mathbb{N}$ and $i \in \{0,\dots,k\}$ we have $B_k = i > 0$, then either $B_{k+1} = i+1$ or 0, $B_{k-1} = i-1,\dots,B_{k-i} = 0$, and $k-i$ is a renewal time.

8.2 For $k \in \mathbb{N}$ and $i \in \{0,\dots,k\}$, for all $D \in \mathscr{F}_k^B = \sigma(B_\ell, \ell \le k)$, we get

$$\begin{aligned}\mathbb{P}_0(D, B_k = i, B_{k+1} = i+1) &= \sum_{\ell=0}^{k} \mathbb{P}_0(D, B_k = i, \eta_k = \ell, Y_{\ell+1} > i+1) \\ &= \mathbb{P}_0(Y_1 > i+1) \sum_{\ell=0}^{k} \mathbb{P}_0(D, B_k = i, \eta_k = \ell) \\ &= \mathbb{P}_0(Y_1 > i+1)\mathbb{P}_0(D, B_k = i) ,\end{aligned}$$

where we have used that $D \cap \{\eta_k = \ell\} \in \mathscr{F}_\ell^S$ and $Y_{\ell+1}$ is independent of $\mathscr{F}_\ell^S$. Along the same lines, we obtain

$$\mathbb{P}_0(D, B_k = i, B_{k+1} = 0) = \mathbb{P}_0(Y_1 = i+1) \sum_{\ell=0}^{k} \mathbb{P}_0(D, B_k = i, \eta_k = \ell) .$$

The Markov kernel R is thus defined for $n \in \mathbb{N}$ by

$$R(n, n+1) = \mathbb{P}_{0,b}\left(Y_1 > n+1 \,|\, Y_1 > n\right) = \frac{\sum_{j=n+2}^{\infty} b(j)}{\sum_{j=n+1}^{\infty} b(j)} , \tag{G.13a}$$

$$R(n, 0) = \mathbb{P}_{0,b}\left(Y_1 = n+1 \,|\, Y_1 > n\right) = \frac{b(n+1)}{\sum_{j=n+1}^{\infty} b(j)} . \tag{G.13b}$$

8.3 For all $k \in \{0,\dots,\sup\{n \in \mathbb{N} : b(n) \ne 0\} - 1\}$, $R^k(0,k) > 0$, and $R^\ell(k,0) > 0$, where $\ell = \inf\{n \ge k : b(n) \ne 0\} + 1$. The kernel R is recurrent, since $\mathbb{P}_1(\sigma_1 < \infty) = 1$. For $j \ge 1$, we have

$$\begin{aligned}\bar{\pi}R(j) = \bar{\pi}(j-1)R(j-1,j) &= m^{-1}\mathbb{P}_0(Y_1 > j-1)\frac{\mathbb{P}_0(Y_1 > j)}{\mathbb{P}_0(Y_1 > j-1)} \\ &= m^{-1}\mathbb{P}_0(Y_1 > j) = \bar{\pi}(j) .\end{aligned}$$

For $j = 0$, we get

$$\begin{aligned}\bar{\pi}R(0) &= \sum_{j=0}^{\infty} \bar{\pi}(j)R(j,0) = m^{-1} \sum_{j=0}^{\infty} \mathbb{P}_0(Y_1 > j) \frac{\mathbb{P}_0(Y = j+1)}{\mathbb{P}_0(Y_1 > j)} \\ &= m^{-1} \sum_{j=0}^{\infty} \mathbb{P}_0(Y = j+1) = m^{-1} = \bar{\pi}(0) .\end{aligned}$$

8.4 1. Set $L = \limsup_n u(n)$. There exists a subsequence $\{n_k,\ k \in \mathbb{N}\}$ such that $\lim_{k\to\infty} u(n_k) = L$. Using the diagonal extraction procedure, we can assume without loss of generality that there exists a sequence $\{q(j),\ j \in \mathbb{Z}\}$ such that

$$\lim_{k\to\infty} u(n_k + j)\mathbb{1}_{\{j \geq -n_k\}} = q(j) ,$$

for all $j \in \mathbb{Z}$.

2. It then follows that $q(0) = L$ and $q(j) \leq L$ for all $j \in \mathbb{Z}$. By the renewal equation (8.1.9), for all $p \in \mathbb{Z}$,

$$u(n_k + p) = \sum_{j=1}^{\infty} b(j)u(n_k + p - j)\mathbb{1}_{\{j \leq n_k + p\}} .$$

Since $u(k) \leq 1$, we obtain, by Lebesgue's dominated convergence theorem

$$q(p) = \sum_{j=1}^{\infty} b(j)q(p-j) . \tag{G.14}$$

3. Since $q(j) \leq L$ for all $j \in \mathsf{Z}$, (G.14) yields, for $p \in S$,

$$\begin{aligned}L = q(0) &= \sum_{j=1}^{\infty} b(j)q(-j) = b(p)q(-p) + \sum_{j \neq p} b(j)q(-j) \\ &\geq b(p)q(-p) + \{1 - b(p)\}L .\end{aligned}$$

This implies that $q(-p) = L$ for all $p \in S$.

4. Let now p be such that $q(-p) = L$. Then arguing as previously,

$$L = \sum_{j=1}^{\infty} b(j)q(-p-j) \geq b(q)q(-p-h) + \{1 - b(h)\}L ,$$

and thus $q(-p-h) = L$ for every $h \in S$. By induction, we obtain that $q(-p) = L$ if $p = p_1 + \ldots + p_n$ with $p_i \in S$ for $i = 1, \ldots, n$.

5. Since the sequence $\{b(j),\ j \in \mathbb{N}\}$ is aperiodic, by Lemma 6.3.2, there exists $p_0 \geq 1$ such that $q(-p) = L$ for all $p \geq p_0$. By (G.14), this yields

$$q(-p_0 + 1) = \sum_{j=1}^{\infty} b(j)q(-p_0 + 1 - j) = L .$$

By induction, this yields that $q(j) = L$ for all $j \in \mathbb{Z}$.

6. Set $\bar{b}(j) = \sum_{i=j+1}^{\infty} b(i)$, so that $\bar{b}(0) = 1$, $b(j) = \bar{b}(j-1) - \bar{b}(j)$, $j \geq 1$, and $\sum_{j=0}^{\infty} \bar{b}(j) = m$. Applying the identity (8.1.9) and summation by parts, we obtain, for $n \geq 1$,

$$u(n) = b * u(n) = \sum_{j=1}^{n} \{\bar{b}(j-1) - \bar{b}(j)\} u(n-j)$$
$$= \sum_{j=0}^{n-1} \bar{b}(j) u(n-j-1) - \sum_{j=0}^{n} \bar{b}(j) u(n-j) + \bar{b}(0) u(n) .$$

Since $\bar{b}(0) = 1$, this yields, for all $n \geq 1$,

$$\sum_{j=0}^{n} \bar{b}(j) u(n-j) = \sum_{j=0}^{n-1} \bar{b}(j) u(n-1-j) .$$

7. By induction, this leads to

$$\sum_{j=0}^{n} \bar{b}(j) u(n-j) = \bar{b}(0) u(0) = 1 .$$

Therefore, for all $k \geq 0$, we obtain (8.4.3).

8. If $m = \infty$, applying Fatou's lemma, (8.4.3) yields

$$1 = \lim_{k \to \infty} \sum_{j=0}^{\infty} \bar{b}(j) u(n_k - j) \mathbb{1}_{\{j \leq n_k\}} \geq L \sum_{j=0}^{\infty} \bar{b}(j) = L \times \infty ,$$

which implies that $L = 0$.

9. If $m < \infty$, then (8.4.3) and Lebesgue's dominated convergence theorem yield $Lm = 1$, i.e., $\limsup_{n \to \infty} u(n) = 1/m$. Setting $\tilde{L} = \liminf_{n \to \infty} u(n)$ and arguing along the same lines, we obtain $\liminf_{n \to \infty} u(n) \geq 1/m$. This proves (8.1.18) for the pure renewal sequence.

10. To prove (8.1.18) in the general case of a delayed renewal sequence, write

$$v_a(n) = a * u(n) = \sum_{k=0}^{n} a(k) u(n-k) = \sum_{k=0}^{\infty} a(k) u(n-k) \mathbb{1}_{\{k \leq n\}} .$$

The proof is concluded by applying the result for the pure renewal sequence and Lebesgue's dominated convergence theorem.

8.5 Decomposing the event $\{X_n = \alpha\}$ according to the first entrance to the state α and applying the Markov property yields, for $n \geq 1$,

$$
\begin{aligned}
u(n) = \mathbb{P}_\alpha(X_n = \alpha) &= \mathbb{P}_\alpha(X_n = \alpha, \sigma_\alpha = n) + \sum_{k=1}^{n-1} \mathbb{P}_\alpha(X_n = \alpha, \sigma_\alpha = k) \\
&= \mathbb{P}_\alpha(\sigma_\alpha = n) + \sum_{k=1}^{n-1} \mathbb{E}_\alpha\left[\mathbb{1}\{\sigma_\alpha = k\}\mathbb{E}_\alpha\left[\mathbb{1}\{X_{n-k} = \alpha\} \circ \theta_k \,|\, \mathscr{F}_k\right]\right] \\
&= \mathbb{P}_\alpha(\sigma_\alpha = n) + \sum_{k=1}^{n-1} \mathbb{P}_\alpha(\sigma_\alpha = k)\mathbb{P}_\alpha(X_{n-k} = \alpha) \\
&= b(n) + \sum_{k=1}^{n-1} u(n-k)b(k) \,.
\end{aligned}
$$

Since $u(0) = 1$, this yields

$$
u(n) = \delta_0(n) + b * u(n) \,. \tag{G.15}
$$

This means that the sequence u satisfies the pure renewal equation (8.1.9). Moreover, applying the strong Markov property, we obtain

$$
\begin{aligned}
a_x * u(n) = \sum_{k=1}^{n} a_x(k)u(n-k) &= \sum_{k=1}^{n} \mathbb{P}_x(\sigma_\alpha = k)\mathbb{P}_\alpha(X_{n-k} = \alpha) \\
&= \mathbb{E}_x\left[\sum_{k=1}^{n} \mathbb{1}\{\sigma_\alpha = k\}\mathbb{P}_{X_{\sigma_\alpha}}(X_{n-k} = \alpha)\right] \\
&= \mathbb{E}_x\left[\sum_{k=1}^{n} \mathbb{1}\{\sigma_\alpha = k\}\mathbb{1}\{X_n = \alpha\}\right] = \mathbb{P}_x(X_n = \alpha) \,.
\end{aligned}
$$

This identity and (G.15) prove that $a_x * u$ is the delayed renewal sequence associated with the delay distribution a_x.

8.6 Applying 8.2.5, we must prove that

$$
\lim_{n\to\infty} \mathbb{P}_x(\sigma_\alpha \geq n) = 0 \,, \tag{G.16a}
$$

$$
\lim_{n\to\infty} |a_x * u - \pi(\alpha)| * \psi(n) = 0 \,, \tag{G.16b}
$$

$$
\lim_{n\to\infty} \sum_{k=n+1}^{\infty} \psi(k) = 0 \,. \tag{G.16c}
$$

Since P is irreducible, for all $x \in \mathsf{X}$, we have $\mathbb{P}_x(\sigma_\alpha < \infty) = 1$; thus (G.16a) holds. Since P is positive recurrent, $\mathbb{E}_\alpha[\sigma_\alpha] = \sum_{n=1}^{\infty} \psi(n) < \infty$, so (G.16c) also holds. Since P is aperiodic, the distribution b defined in (8.4.5) is also aperiodic. Thus we can apply Blackwell's theorem (Theorem 8.1.7) with $\pi(\alpha) = 1/\mathbb{E}_\alpha[\sigma_\alpha] = 1/\sum_{k=1}^{\infty} kb(k)$, and we have $\lim_{n\to\infty} a_x * u(n) = \pi(\alpha)$. Since $\sum_{k\geq 1} \psi(k) < \infty$, by Lebesgue's dominated convergence theorem, we finally obtain that (G.16b) holds.

8.7 1. Since all the states communicate, for all $x,y \in \mathsf{X}$, there exists an integer $nr(x,y)$ such that $\mathbb{P}_x(\sigma_y \le r(x,y)) > 0$. Define $r = \sup_{x,y\in\mathsf{X}} r(x,y)$ and $\varepsilon = \inf_{x,y\in\mathsf{X}} \mathbb{P}_x(\sigma_y \le r(x,y))$. Since X is finite, r is a finite integer, $\varepsilon > 0$, and for all $x,y \in \mathsf{X}$, $\mathbb{P}_x(\sigma_y \le r) \ge \varepsilon$.

2. We have

$$\begin{aligned}\mathbb{P}_x(\sigma_y > kr) &= \mathbb{P}_x(\sigma_y > (k-1)r, \sigma_y \circ \theta_{(k-1)r} > r)\\ &= \mathbb{E}_x\left[\mathbb{1}_{\{\sigma_y > (k-1)r\}} \mathbb{P}_{X_{(k-1)r}}(\sigma_y > r)\right] \le (1-\varepsilon)\mathbb{P}_x(\sigma_y > (k-1)r)\,.\end{aligned}$$

Thus for all x,y, $\mathbb{P}_x(\sigma_y > kr) \le (1-\varepsilon)^k$.

3. For $b > 1$, it follows that

$$\begin{aligned}\mathbb{E}_x[b^{\sigma_y}] &= \sum_{k=1}^{\infty} b^k \mathbb{P}_x(\sigma_y = k) = b\sum_{k=0}^{\infty} b^k \mathbb{P}_x(\sigma_y = k+1) \le b\sum_{k=0}^{\infty} b^k \mathbb{P}_x(\sigma_y > k)\\ &\le rb^{r+1}\sum_{k=0}^{\infty} b^{kr}\mathbb{P}_x(\sigma_y > kr) \le rb^{r+1}\sum_{k=0}^{\infty}[(1-\varepsilon)b^r]^k\,.\end{aligned}$$

If b is such that $(1-\varepsilon)b^r < 1$, then the series is summable and $\mathbb{E}_x[b^{\sigma_y}] < \infty$.

8.8 1. Set $M = \sup_{x\in C}\mathbb{E}_x[a^{\sigma_C^{(n)}}]$. Then by induction, for all $n \ge 1$, $\sup_{x\in C}\mathbb{E}_x[a^{\sigma_C^{(n)}}] \le M^n$. Then $\sigma_x = \sigma_C^{(\nu_x)}$. Applying Exercise 8.7 to the induced Markov chain on the set C, $\left\{X_{\sigma_C^{(n)}} \,:\, n \in \mathbb{N}\right\}$ (see Definition 3.3.7), we obtain that there exists $r > 1$ such that $\mathbb{E}_x[r^{\nu_x}] < \infty$ for all $x \in C$.

2. Choose $s > 0$ such that $M^s \le \beta^{1/2}$. Then

$$\begin{aligned}\mathbb{P}_x(\sigma_x \ge n) &\le \mathbb{P}_x(\nu_x \ge sn) + \mathbb{P}_x(\sigma_C^{(\nu_x)} \ge n,\ \nu_x < sn)\\ &\le \mathbb{P}_x(\nu_x \ge sn) + \mathbb{P}_x(\sigma_C^{([sn])} \ge n) \le \mathbb{E}_x[r^{\nu_x}]r^{-sn} + \beta^{-n}\mathbb{E}_x\left[\beta^{\sigma_C^{([sn])}}\right]\\ &\le \mathbb{E}_x[r^{\nu_x}]r^{-sn} + \beta^{-n}M^{sn} \le \left(\sup_{x\in C}\mathbb{E}_x[r^{\nu_x}]\right)r^{-sn} + (\sqrt{\beta})^{-n}\,.\end{aligned}$$

3. Choosing $\delta = r \wedge \sqrt{\beta}$ yields $\mathbb{E}_x[\delta^{\sigma_x}]$ for all $x \in C$.

Solutions to Exercises of Chapter 9

9.3 1. Assume first that $F((-\infty,0)) = 0$. Then for all $k > 0$, the set $[0,k)$ is inaccessible. Conversely, suppose for some $\delta,\varepsilon > 0$, $F((-\infty,\ -\varepsilon)) > \delta$. Then for all n, if $x/\varepsilon < n$, then

$$P^n(x,\ \{0\}) \ge \delta^n > 0,$$

showing that $\{0\}$ is an accessible atom and therefore an accessible small set.

2. If $C = [0,\ c]$ for some c, then this implies for all $x \in C$ that

$$\mathbb{P}_x(\sigma_0 \leq c/\varepsilon) \geq \delta^{1+c/\in},$$

showing that $\{0\}$ is uniformly accessible from every compact subset.

9.5 Let $A \in \mathscr{X}$ be such that $\lambda(A \cap C) > 0$. Then for every $x \in \mathsf{X}$,

$$\sum_{n=1}^{\infty} P^n(x, A \cap C) \geq \int_{A \cap C} \sum_{n=1}^{\infty} p_n(x,y)\lambda(\mathrm{d}y) > 0 \,.$$

Thus A is accessible.

9.6 Since C is accessible, it follows that for all $x \in \mathsf{X}$, $P^n(x,C) > 0$. By the Chapman–Kolmogorov equations, we get

$$P^{n+m}(x,B) \geq \int_C P^n(x,\mathrm{d}y)P^m(y,B) \geq \nu(B) \int_C \varepsilon(y,B)P^n(x,\mathrm{d}y) > 0 \,.$$

9.7 Define $p(x,y) = q(x,y)\alpha(x,y)$. Suppose first that $\pi(y) > 0$. Consider two cases. If $\pi(y)q(y,x) \geq \pi(x)q(x,y)$, then we simply have $p(x,y) = q(x,y) > 0$ by assumption; this is also the case if $\pi(x) = 0$. If, on the other hand, $\pi(y)q(y,x) < \pi(x)q(x,y)$, then $p(x,y) = q(y,x)\pi(y)/\pi(x)$, which is positive also, since this case requires $\pi(x) > 0$, and our assumption then implies that $q(y,x) > 0$ for all $y \in \mathsf{X}$.

Thus if $\pi(A) > 0$, we must also have $\int_A p(x,y)\lambda(\mathrm{d}y) > 0$ for all $x \in \mathsf{X}$, and the chain is "one-step" irreducible.

9.10 1. if $B \subseteq C$ and $x \in C$, then

$$\begin{aligned} P(x,\ B) &= \mathbb{P}(Z_1 \in B - x) \\ &\geq \int_{B-x} \gamma(y)\mathrm{d}y \geq\ \delta \mathrm{Leb}(B) \,. \end{aligned}$$

2. From every x we can reach C in at most $n = 2|x|/\beta$ steps with positive probability.
3. $\mathrm{Leb}(\cdot \cap C)$ is an irreducibility measure by Proposition 9.1.9.
4. For all $x \in \mathbb{R}$, the set $\{x + q\ :\ q \in \mathbb{Q}\} = x + \mathbb{Q}$ is absorbing. The state space $\mathbb{R}$ is covered by an uncountably infinite number of absorbing sets.

9.13 Let C be a nonempty compact set. By hypothesis, we have $M = \sup_{x \in C} h_\pi(x) < \infty$ and $\varsigma = \inf_{x,y \in C} q(x,y) > 0$. Choose $A \subseteq C$, and for fixed x denote the region where moves might be rejected by

$$R_x = \left\{ y \in A\ :\ \frac{\pi(y)}{\pi(x)} \frac{q(y,x)}{q(x,y)} < 1 \right\},$$

and set $A_x = A \setminus R_x$ as the region where all moves are accepted.

By construction, for $x \in C$,

$$
\begin{aligned}
P(x,\ A) &\geq \int_{R_x} q(x,\ y) \min\left\{\frac{h_\pi(y)}{h_\pi(x)}\frac{q(y,x)}{q(x,y)},\ 1\right\} \nu(\mathrm{d}y) \\
&\quad + \int_{A_x} q(x,\ y) \min\left\{\frac{h_\pi(y)}{h_\pi(x)}\frac{q(y,x)}{q(x,y)},\ 1\right\} \nu(\mathrm{d}y) \\
&= \int_{R_x} \frac{h_\pi(y)}{h_\pi(x)} q(y,\ x)\nu(\mathrm{d}y) + \int_{A_x} q(x,\ y)\nu(\mathrm{d}y) \\
&\geq (\varsigma/M)\int_{R_x} h_\pi(y)\nu(\mathrm{d}y) + \varsigma \int_{A_x} h_\pi(y)/M\nu(\mathrm{d}y) \\
&= (\varsigma/M)\pi(A)\ .
\end{aligned}
$$

9.15 For $x \in C$, we use the following decomposition:

$$
P^m(x,\cdot) = (1-\varepsilon)R_m(x,\dot{}) + \varepsilon\nu\ , \quad R_m(x,\cdot) = \frac{1}{1-\varepsilon}\{P^m(x,\cdot - \varepsilon\nu\}\ .
$$

Therefore, we get, for $(x,x') \in C \times C$,

$$
\left\|P^m(x,\cdot) - P^m(x',\cdot)\right\|_{\mathrm{TV}} \leq (1-\varepsilon)\left\|R_m(x,\cdot) - R_m(x',\cdot)\right\|_{\mathrm{TV}}\ ,
$$

and we conclude by noting that $\|R_m(x,\cdot) - R_m(x',\cdot)\|_{\mathrm{TV}} \leq 2$.

9.18 The result follows by an induction argument. The statement (9.5.1) is trivial for $m = 0$. Moreover, suppose the statement is true for $m = k-1$. Then

$$
\begin{aligned}
P^k(x,A) &= \int_{\mathsf{X}} P^{k-1}(x,\mathrm{d}y)P(y,\ A) \\
&\leq \int_{\mathsf{X}} \left\{\sum_{i=0}^{k-1}\binom{k-1}{i} Q^i(x,\mathrm{d}y)\right\}\{\mathbb{1}_A(y) + Q(y,A)\} \\
&= \sum_{i=0}^{k-1}\binom{k-1}{i} Q^i(x,A) + \sum_{i=0}^{k-1}\binom{k-1}{i} Q^{i+1}(x,A) \\
&= \sum_{i=0}^{k-1}\left\{\binom{k-1}{i} + \binom{k-1}{i-1}\right\} Q^i(x,\ A) + Q^k(x,A) \\
&= \sum_{i=0}^{k}\binom{k}{i} Q^i(x,\ A)\ .
\end{aligned}
$$

9.19 We use the notation introduced in Example 2.3.2. To show that unbounded sets are not small, it is sufficient to prove that for all bounded Borel sets A and for all $m \in \mathbb{N}^*$, $\lim_{|x|\to\infty} P^m(x,A) = 0$. This will done by induction on m. First set $m = 1$ and let A be a bounded Borel set. Denoting by $r(x)$ the probability of staying at the same position x, that is, $r(x) = 1 - \int \bar{q}(z)\alpha(x,x+z)\mathrm{d}z$, we have

$$
\begin{aligned}
P(x,A) &= \int \bar{q}(z)\alpha(x,x+z)\mathbb{1}_A(x+z)\mathrm{d}z + r(x)\mathbb{1}_A(x) \\
&\leq \int \bar{q}(z)\mathbb{1}_A(x+z)\mathrm{d}z + \mathbb{1}_A(x)\,. \qquad \text{(G.17)}
\end{aligned}
$$

Since A is bounded, applying Lebesgue's dominated convergence theorem proves that $\lim_{|x|\to\infty} P(x,A) = 0$. Assume that $\lim_{|x|\to\infty} P^m(x,A) = 0$ for some $m \geq 1$. Then using again (G.17), we obtain

$$
P^{m+1}(x,A) \leq \int \bar{q}(z)P^m(x+z,A)\mathrm{d}z + \mathbb{1}_A(x)\,.
$$

The induction assumption together with Lebesgue's dominated convergence theorem shows that $\lim_{|x|\to\infty} P^{m+1}(x,A) = 0$. This finishes the proof.

9.20 By Definition 9.2.1 and Lemma 9.1.6, there exists an accessible $(r,\varepsilon\nu)$-small set C with $r \in \mathbb{N}^*$, $\varepsilon > 0$, $\nu \in \mathbb{M}_1(\mathscr{X})$, and $\nu(C) > 0$. Since the kernel P is aperiodic, Lemma 9.3.3 (ii) shows that there exists an integer n_0 such that C is an $(n,\varepsilon_n\nu)$-small set for all $n \geq n_0$. Provided that C is accessible for P^n, the kernel P^n is strongly aperiodic. We will show that C is actually accessible for P^m for all $m \in \mathbb{N}^*$. Since C is accessible, for all $x \in \mathsf{X}$, there exists $k > 0$ such that $P^k(x,C) > 0$. Hence for all $n \geq n_0$, we get

$$
P^{k+n}(x,C) \geq \int_C P^k(x,\mathrm{d}y)P^n(y,C) \geq \varepsilon_n\nu(C)P^k(x,C) > 0\,.
$$

Thus C is accessible for P^m for all $m \in \mathbb{N}^*$, and the proof is complete.

9.21 1. We will first compute an upper bound for the probability of accepting a move started at x:

$$
\begin{aligned}
P(x,\{x\}^c) &= \int q(x,y)(1 \wedge \frac{\pi(y)}{\pi(x)})\mathrm{d}y \\
&\leq \frac{M}{\pi(x)}\int \pi(y)\mathrm{d}y = \frac{M}{\pi(x)}.
\end{aligned}
$$

2. Let C be a set on which π is unbounded. Then $\inf_{x\in C} P(x,\{x\}^c) = 0$. We may just choose x_0 and x_1 such that $P(x_i,\{x_i\}) > (1-\varepsilon/2)^{1/m}$, $i = 0,1$.
3. By Proposition D.2.3, we have

$$
\|P^m(x_0,\cdot) - P^m(x_1,\cdot)\|_{\mathrm{TV}} = \sup \sum_{i=0}^{I} |P^m(x_0,B_i) - P^m(x_1,B_i)|,
$$

where the supremum is taken over all finite measurable partitions $\{B_i\}_{i=0}^I$. Taking $B_0 = \{x_0\}$, $B_1 = \{x_1\}$, and $B_2 = \mathsf{X} \setminus (B_0 \cup B_1)$, we therefore have

$$
\begin{aligned}
&\|P^m(x_0,\cdot) - P^m(x_1,\cdot)\|_{\mathrm{TV}} \\
>&|P^m(x_0,\{x_0\}) - P^m(x_1,\{x_0\})| + |P^m(x_0,\{x_1\}) - P^m(x_1,\{x_1\})| \geq 2(1-\varepsilon)\,,
\end{aligned}
$$

where we have used $P^m(x_i, \{x_i\}) > (1-\varepsilon/2)$, $i = 0, 1$ and $P^m(x_i, \{x_j\}) < \varepsilon/2$, $i \neq j \in \{0, 1\}$.

Solutions to Exercises of Chapter 10

10.1 1. For every bounded function h, we have

$$Ph(x) = \int h(x+y)\mu(\mathrm{d}y) = \int h(x+y)g(y)\mathrm{d}y = \int h(y)g(y-x)\mathrm{d}y \,.$$

Since $h \in \mathrm{L}^\infty(\mathrm{Leb})$, $g \in \mathrm{L}^1(\mathrm{Leb})$, the function Ph is uniformly continuous on $\mathbb{R}^d$.

2. The sequence $\{(M_n(x), \mathscr{F}_n^Z), n \in \mathbb{N}\}$ is a bounded martingale.
3. Obviously the random variable $H(x)$ is invariant under finite permutation of the sequence $\{Z_n,\, n \in \mathbb{N}^*\}$, and the zero–one law shows that there exists a constant c such that $H(x) = c$ $\mathbb{P}-$a.s. Therefore, $H(x) = \mathbb{E}[H(x)] = h(x)$, $\mathbb{P}-$a.s. We have $h(x+Z_1) = M_1(x) = \mathbb{E}\left[H(x)\,|\,\mathscr{F}_1^Z\right] = h(x)$ $\mathbb{P}-$a.s.
4. It follows that $h(x+y) = h(x)$ μ-*a.e.*, and by continuity, $h(x+y) = h(x)$ for all $y \in \mathrm{supp}(\mu)$, and since $\mathrm{supp}(\mu) \supset \mathrm{B}(0,a)$, for all $y \in \mathbb{R}^d$.

10.2 1. Let $A \in \mathscr{X}$ be such that $\nu(A) = 0$. We have

$$\pi(A) = \int \pi(\mathrm{d}x)P(x,A) = \int \pi(\mathrm{d}x) \int \mathbb{1}_A(y)p(x,y)\nu(\mathrm{d}y) = 0 \,.$$

2. Since P admits an invariant probability, P is recurrent by Theorem 10.1.6. Let h be a bounded harmonic function. By Proposition 5.2.12, $h(x) = \pi(h)$ π-*a.e.*Since $Ph(x) = h(x)$ for all $x \in \mathsf{X}$, we get

$$Ph(x) = \int p(x,y)h(y)\nu(\mathrm{d}y) = \int p(x,y)\pi(h)\nu(\mathrm{d}y) = \pi(h) \,.$$

Thus $h(x) = \pi(h)$ for all $x \in \mathsf{X}$, and Theorem 10.2.11 (ii) shows that P is Harris recurrent.

10.3 1. P admits π as its unique invariant probability. Hence P is recurrent by Theorem 10.1.6. By Proposition 5.2.12, $h(x) = \pi(h)$ π-*a.e.*
2. We have

$$\int q(x,y)\alpha(x,y)h(y)\mu(\mathrm{d}y) = \{1 - \bar{\alpha}(x)\}\pi(h)$$

and thus

$$\int P(x,\mathrm{d}y)h(y) = \{1-\bar{\alpha}(x)\}\pi(h) + \bar{\alpha}(x)h(x) = h(x) \,,$$

which implies $\{1-\bar{\alpha}(x)\}\{h(x) - \pi(h)\} = 0$.

3. Since π is not concentrated at a single point, π-irreducibility implies that $\bar{\alpha}(x) < 1$ for all $x \in \mathsf{X}$.
4. $h(x) = \pi(h)$ for all $x \in \mathsf{X}$. Theorem 10.2.11 (ii) shows that P is Harris recurrent.

10.5 1. For $a > 0$, $x \in [0,a]$, and a measurable set $A \subset \mathbb{R}_+$, we have

$$P(x,A) = \mathbb{P}((x+W)_+ \in A) \geq \mathbb{P}(x+W \leq 0, 0 \in A) \geq \mathbb{P}(W \leq -a)\delta_0(A) .$$

Since q is positive, $\mathbb{P}(W < -a) > 0$ for all $a > 0$, and thus compact sets are small. This also proves that δ_0 is an irreducibility measure by Proposition 9.1.9.

2. For $x > x_0$, we have

$$PV(x) - V(x) = \mathbb{E}\left[W_1 \mathbb{1}_{W_1 \geq -x}\right] - x\mathbb{P}(W_1 \leq -x) \leq \int_{-x_0}^{\infty} wq(w)\mathrm{d}w .$$

3. The assumptions of Theorem 10.2.13 hold with $C = [0,\ x_0]$ and $V(x) = x$, and thus P is Harris recurrent.
4. For all $y > -1$, we have $\log(1+y) \leq y - (y^2/2)\mathbb{1}\{y < 0\}$, which implies

$$\begin{aligned}\log(1+x+W_1)&\mathbb{1}\{x+W_1 \geq R\} \\ &= [\log(1+x) + \log(1+W_1/(1+x))]\mathbb{1}\{x+W_1 \geq R\} \\ &\leq [\log(1+x) + W_1/(1+x)]\mathbb{1}\{x+W_1 \geq R\} \\ &\qquad - (W_1^2/(2(1+x)^2))\mathbb{1}\{R-x \leq W_1 < 0\} .\end{aligned}$$

If $x > R$, then $1+x > 0$, and by taking expectations in the previous inequality, we obtain

$$\begin{aligned}PV(x) &= \mathbb{E}[\log(1+x+W_1)\mathbb{1}\{x+W_1 > R\}] \\ &\leq (1-Q(R-x))\log(1+x) + U_1(x) - U_2(x) .\end{aligned}$$

5. Since $\mathbb{E}[W_1] = 0$, we have $\mathbb{E}[W_1\mathbb{1}\{W_1 > R-x\}] = -\mathbb{E}[W_1\mathbb{1}\{W_1 \leq R-x\}]$, and thus for $x > R$,

$$\mathbb{E}[|W_1|\mathbb{1}\{W_1 \leq R-x\}] \leq \frac{\mathbb{E}\left[W_1^2\right]}{x-R} .$$

This shows that $U_1(x) = o(x^{-2})$. On the other hand, since $\mathbb{E}\left[W_1^2\right] < \infty$,

$$U_2(x) = (1/(2(1+x)^2))\mathbb{E}\left[W_1^2\mathbb{1}\{W_1 < 0\}\right] - o(x^{-2}) .$$

6. Thus by choosing R large enough, we obtain for $x > R$,

$$PV(x) \leq V(x) - (1/(2(1+x)^2))\mathbb{E}\left[W_1^2\mathbb{1}\{W_1 < 0\}\right] + o(x^{-2}) \leq V(x) .$$

Since the function V is unbounded off petite sets, the kernel is recurrent by Theorem 10.2.13.

10.6 1. Let K be a compact set with nonempty interior. Then $\mathrm{Leb}(K) > 0$, and for every $x \in K$,

$$P(x,A) = \int_A q(y-m(x))\mathrm{d}y \geq \int_{A\cap K} q(y-m(x))\mathrm{d}y \geq \varepsilon_K \nu(A) ,$$

with

$$\nu_K(A) = \frac{\mathrm{Leb}(A\cap K)}{\mathrm{Leb}(K)} , \quad \varepsilon_K = \mathrm{Leb}(K) \min_{(t,x)\in K\times K} q(t-m(x)) .$$

2. Using that $|m(x)+Z_1| \geq |m(x)| - |Z_1|$, we obtain

$$\begin{aligned} PV(x) &= 1 - \mathbb{E}\left[\exp(-\beta\,|m(x)+Z_1|)\right] \\ &\geq 1 - \mu_\beta \exp(-\beta\,|m(x)|) = V(x) - W(x) , \end{aligned}$$

where

$$W(x) = \mu_\beta \exp(-\beta\,|m(x)|) + \exp(-\beta\,|x|) .$$

Under the stated conditions, $\lim_{|x|\to\infty} W(x) = \infty$.

3. For $r \in (0,1)$, $\{V \leq r\} = \{|x| \leq -\alpha^{-1}\log(1-r)\}$, and we may choose r small enough that for all $x \in \mathbb{R}^d$ such that $|x| > r$, $W(x) < 0$. Therefore, $PV > V$ on $\{V > r\}$. If Z_1 has a positive density with respect to Lebesgue measure on $\mathbb{R}^d$, then Leb is an irreducibility measure, and the sets $\{V \leq r\}$ and $\{V > r\}$ are both accessible. Therefore, by Theorem 10.1.11, the chain is transient.

10.8 1. P is recurrent by application of Theorem 10.1.6. (If P admits an invariant probability measure π, then P is recurrent.)

2. Set $A_\infty = \{x \in \mathsf{X} : \mathbb{P}_x(N_A = \infty) = 1\}$. By applying Theorem 10.1.10, this set is absorbing and full. Since π is a maximal irreducibility measure by Theorem 9.2.15, this implies that $\pi(A_\infty) = 1$, i.e., $\mathbb{P}_y(N_A = \infty) = 1$ for π almost all $y \in \mathsf{X}$.

3. For all $x \in \mathsf{X}$,

$$\begin{aligned} \mathbb{P}_x(N_A = \infty) &= \mathbb{P}_x(N_A \circ \theta_m = \infty) = \mathbb{E}_x[\mathbb{P}_{X_m}(N_A = \infty)] \\ &= \int_{\mathsf{X}} r(x,y)\mathbb{P}_y(N_A = \infty)\pi(\mathrm{d}y) = \int_{\mathsf{X}} r(x,y)\pi(\mathrm{d}y) = 1 . \end{aligned}$$

Therefore, P is Harris recurrent.

10.9 Set $A = \{\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1} Y\circ\theta_k = \mathbb{E}_\pi[Y]\}$. The set A is invariant, and the function $h(x) = \mathbb{P}_x(A)$ is harmonic (see Proposition 5.2.2 (iii)) and hence is constant by Theorem 10.2.11. By Theorem 5.2.6, π is ergodic and $\mathbb{P}_\pi(A) = 1$. Therefore, $\mathbb{P}_x(A) = 1$ for all $x \in \mathsf{X}$ and $\mathbb{P}_\xi(A) = 1$ for all $\xi \in \mathbb{M}_1(\mathscr{X})$.

Solutions to Exercises of Chapter 11

11.1 The first assertion is obvious. To prove the second assertion, note that $\mathbb{P}_0(A) = 0$ and $\mathbb{P}_1(A) = 1$. Hence we have $\mathbb{P}_\mu(A) = 1/2$ if $\mu = (\delta_0 + \delta_1)/2$, showing that the asymptotic σ-field $\mathscr{A}$ is not trivial.

11.2 We get using Lemma 11.1.1 that

$$\begin{aligned}\left\|\xi P^k - \xi' P^k\right\|_f &= \sup_{|h|\leq f} |\xi P^k h - \xi' P^k h| \\ &= \sup_{|h|\leq f} |[\xi P^k \otimes \mathrm{b}_\varepsilon](h\otimes \mathbf{1}) - [\xi' P^k \otimes \mathrm{b}_\varepsilon](h\otimes \mathbf{1})| \\ &= \sup_{|h|\leq f} |[\xi \otimes \mathrm{b}_\varepsilon]\check{P}^k(h\otimes \mathbf{1}) - [\xi' \otimes \mathrm{b}_\varepsilon]\check{P}^k(h\otimes \mathbf{1})| \,.\end{aligned}$$

Since the condition $|h| \leq f$ implies that $|h\otimes \mathbf{1}| \leq f \otimes \mathbf{1}$, (11.5.1) follows. Applying (11.5.1) with $\xi' = \pi$ and using Proposition 11.1.3, we deduce (11.5.2).

11.3 For all $x \in \mathsf{X}$, we have

$$\mathbb{P}_x(\sigma_C < \infty) = \lim_n \mathbb{P}_x(\sigma_C \leq n) \geq \lim_n P^n(x,C) = \varepsilon > 0 \,.$$

By Theorem 4.2.6, this implies that $\mathbb{P}_x(N_C = \infty) = 1$ for all $x \in \mathsf{X}$. Therefore, the chain is Harris recurrent by Proposition 10.2.4 and positive by Exercise 11.5.

11.4 Assume that for all $\mu \in \mathbb{M}_1(\mathsf{X})$ and all $A \in \mathscr{A}$, $\mathbb{P}_\mu(A) = 0$ or 1. If the mapping $\mu \to \mathbb{P}_\mu(A)$ is not constant, then there exist $\mu_1, \mu_2 \in \mathsf{X}$ such that $\mathbb{P}_{\mu_1}(A) = 1$ and $\mathbb{P}_{\mu_2}(A) = 0$ and by setting $\mu = (\mu_1 + \mu_2)/2$, we obtain that $\mathbb{P}_\mu(A) = 1/2$, which is a contradiction.

11.5
1. By Theorem 11.A.4, $\lim_{n\to\infty} |P^n(x,A) - P^n(y,A)| = 0$ for all $y \in \mathsf{X}$. Since P is null recurrent, $\mu(\mathsf{X}) = \infty$ is infinite. Therefore, by Egorov's theorem, Theorem B.2.12, there exists B such that $\mu(B) \geq 1/\delta$ and $\lim_{n\to\infty} \sup_{y\in B} |P^n(x,A) - P^n(y,A)| = 0$.
2. We can choose n_0 large enough that $\sup_{y\in B} |P^n(x,A) - P^n(y,A)| \leq \varepsilon\delta/2$ for $n \geq n_0$. This yields, for $n \geq n_0$,

$$\begin{aligned}\mu(A) &\geq \int_B \mu(\mathrm{d}y) P^n(y,A) \geq \int_B \mu(\mathrm{d}y)(P^n(x,A) - \varepsilon\delta/2) \\ &= \mu(B)(P^n(x,A) - \varepsilon\delta/2) \,.\end{aligned}$$

3. Letting $n \to \infty$ and using (11.5.3) yields $\mu(A) \geq \delta^{-1}\{\limsup_{n\to\infty} P^n(x,A)\} - \varepsilon/2 = \mu(A) + \varepsilon/2$, which is impossible.
4. Elementary.
5. If $x \in C^c$, then $\mathbb{P}_x(\sigma_C < \infty) = 1$, since the chain is Harris recurrent. By the Markov property (see Exercise 3.5), we get

$$P^n(x,A) = \mathbb{E}_x[\mathbb{1}\{n \le \sigma_C\}\, \mathbb{1}_A(X_n)] + \mathbb{E}_x[\mathbb{1}\{\sigma_C < n\}\, P^{n-\sigma_C}(X_{\sigma_C}, A)] \to 0$$

as $n \to \infty$ by Lebesgue's dominated convergence theorem.

Solutions to Exercises of Chapter 12

12.1 1. For $f \in \mathbb{F}_b(\mathbb{R}^q)$,

$$Pf(x) = \int_{\mathbb{R}^q} f(m(x) + \sigma(x)z)\mu(\mathrm{d}z) \,. \tag{G.18}$$

By Lemma 12.1.5, P is Feller if m and σ are continuous.

2. Applying the change of variable $y = m(x) + \sigma(x)z$, (G.18) may be rewritten as

$$Pf(x) = \int_{\mathbb{R}^q} f(y)|\det \sigma^{-1}(x)| g\left(\sigma^{-1}(x)\{y - m(x)\}\right) \mathrm{d}y \,.$$

For every $\varepsilon > 0$, there exists a continuous function $g_\varepsilon : \mathbb{R}^q \mapsto \mathbb{R}^+$ with compact support such that $\int_{\mathbb{R}^q} |g(z) - g_\varepsilon(z)|\mathrm{d}z \le \varepsilon$. For a $f \in \mathbb{F}_b(\mathbb{R}^q)$, define the kernel P_ε by

$$\begin{aligned} P_\varepsilon f(x) &= \int_{\mathbb{R}^q} f(m(x) + \sigma(x)z) g_\varepsilon(z)\mathrm{d}z \\ &= \int_{\mathbb{R}^q} f(y)|\det \sigma^{-1}(x)| g_\varepsilon\left(\sigma^{-1}(x)\{y - m(x)\}\right) \mathrm{d}y \,. \end{aligned}$$

Since g_ε is continuous with compact support, for every $x_0 \in \mathbb{R}$,

$$\lim_{x \to x_0} P_\varepsilon f(x) = P_\varepsilon f(x_0) \,.$$

That is, the kernel P_ε is strong Feller. Moreover, for every $f \in \mathbb{F}_b(\mathbb{R}^q)$ such that $|f|_\infty \le 1$,

$$\sup_{x \in \mathbb{R}} |Pf(x) - P_\varepsilon f(x)| \le \varepsilon \,.$$

This yields

$$\begin{aligned} |Pf(x) - Pf(x_0)| &\le |Pf(x) - P_\varepsilon f(x)| + |P_\varepsilon f(x) - P_\varepsilon f(x_0)| + |P_\varepsilon f(x_0) - Pf(x_0)| \\ &\le 2\varepsilon + |P_\varepsilon f(x) - P_\varepsilon f(x_0)| \,. \end{aligned}$$

Thus $\limsup_{x \to x_0} |Pf(x) - Pf(x_0)| \le 2\varepsilon$. Since ε is arbitrary, this proves that P is strong Feller.

12.2 The kernel P of this chain is defined by

$$P(x,A) = p\mathbb{1}_A((x+1)/3) + (1-p)\mathbb{1}_A(x/3) .$$

If f is continuous on $[0,1]$, then for all $x \in [0,1]$, $Pf(x) = pf((x+1)/3) + (1-p)f(x/3)$, which defines a continuous function. Thus P is Feller. However, it is not strong Feller. Consider, for instance, $f = \mathbb{1}_{[0,1/2]}$. Then $Pf = pf + 1 - p$, which is discontinuous.

12.6 1. For all $f \in \mathrm{C}_b(\mathsf{X})$, we have $Pf(x) = \int f(x+z)\mu(\mathrm{d}z)$, which is continuous by Lebesgue's dominated convergence theorem. Thus P is Feller.

2. Assume that μ has a density h with respect to Lebesgue measure on $\mathbb{R}^q$. Then for $f \in \mathbb{F}(\mathsf{X})$ and $x, x' \in \mathbb{R}^q$,

$$\begin{aligned}
|Pf(x) - Pf(x')| &= \left| \int_{\mathbb{R}^q} \{f(x+y) - f(x'+y)\} h(y)\mathrm{d}y \right| \\
&= \left| \int_{\mathbb{R}^q} \{h(y-x) - h(y-x')\} f(y)\mathrm{d}y \right| \\
&\leq |f|_\infty \int_{\mathbb{R}^q} |h(y) - h(y-(x-x'))|\mathrm{d}y .
\end{aligned}$$

The function $x \mapsto \int |h(y) - h(y-x)|\mathrm{d}y$ is continuous at 0. (To see this, approximate h by a compactly supported continuous function g_ε such that $\int |g_\varepsilon - h| \leq \varepsilon$.) This yields $\lim_{x' \to x} |Pf(x) - Pf(x')| = 0$, and P is strong Feller.

3. Conversely, assume that P is strong Feller. Let A be a measurable set such that $\mu(A) = \delta > 0$. Since $x \to P(x,A)$ is continuous and $P(0,A) = \mu(A) = \delta$, we may choose an open set $O \in \mathscr{V}_0$ such that $P(x,A) = \mu(A-x) \geq \delta/2$ for all $x \in O$.

4. Using Fubini's theorem, symmetry, and the translation-invariance of Lebesgue measure, we obtain

$$\begin{aligned}
\mathrm{Leb}(A) &= \int_{\mathbb{R}^q} \mu(\mathrm{d}y) \int_{\mathbb{R}^q} \mathbb{1}_A(x)\mathrm{d}x = \int_{\mathbb{R}^q} \mu(\mathrm{d}y) \int_{\mathbb{R}^q} \mathbb{1}_A(y-x)\mathrm{d}x \\
&= \int_{\mathbb{R}^q} \mathrm{d}x \int_{\mathbb{R}^q} \mathbb{1}_A(y-x)\mu(\mathrm{d}y) = \int_{\mathbb{R}^q} \mu(A-x)\mathrm{d}x \\
&\geq \int_O \mu(A-x)\mathrm{d}x \geq \frac{\delta}{2}\mathrm{Leb}(O) > 0 .
\end{aligned}$$

This proves that $\mu(A) > 0$ implies $\mathrm{Leb}(A) > 0$, and hence μ is absolutely continuous with respect to Lebesgue measure.

12.7 1. (i) $\Rightarrow$ (ii) If μ^{*p} is nonsingular with respect to Lebesgue measure, then there exists a function $g \in \mathrm{L}^1(\mathrm{Leb}) \cap \mathrm{L}^\infty(\mathrm{Leb})$ such that $\mu^{*p} \geq g.\mathrm{Leb}$ and g is not identically equal to zero. Then $\mu^{*2p} \geq g * g.\mathrm{Leb}$, and $g * g$ is continuous and is not identically equal to zero, which implies (ii) for $q = 2p$.
(ii) $\Rightarrow$ (iii) Since g is continuous and nonzero, there exist an open set O and $\alpha > 0$ such that $g \geq \alpha\mathbb{1}_O$. (iii) follows.
(iii) $\Rightarrow$ (i) is obvious.

2. If μ is spread out, we have by Exercise 12.7, $\mu^{*q} \geq g \cdot \mathrm{Leb}$, where $g \in \mathrm{C}_c^+(\mathbb{R}^\mathsf{d}, \mathscr{B}(\mathbb{R}^\mathsf{d}))$ and g is nonzero. We set for $x \in \mathbb{R}^d$ and $A \in \mathscr{B}(\mathbb{R}^d)$, $T(x,A) = \mathrm{Leb}(\mathbb{1}_A * g(x))$. It is easily shown that $x \mapsto T(\cdot,A)$ is continuous.
   Conversely, assume that μ is not spread out and that P is a T-kernel, i.e., there exists $a \in \mathbb{M}_1(\mathbb{N}^*)$ such that $T(x,A) \geq K_a(x,A)$ for all $x \in \mathsf{X}$ and $A \in \mathscr{X}$.
3. For all $n \geq 1$, there exists A_n such that $\mu^{*n}(A_n) = 1$ and $\mathrm{Leb}(A_n) = 0$. If we set $A = \bigcap_{n \geq 1} A_n$, we have, for all $n \geq 1$, $\mu^{*n}(A) = 1$ and $\mathrm{Leb}(A) = 0$.
4. Since P is a T-kernel,
$$T(0,A^c) \leq K_a(0,A^c) = \sum_{k=1}^{\infty} a(k)\mu^{*n}(A^c) = 0\,.$$
   Hence $T(\cdot,A) > 0$, and since T is lower semicontinuous, $\inf_{x \in O} T(x,A) = \delta > 0$ for some $O \in \mathscr{V}_0$. This implies that $\inf_{x \in O} K_a(x,A) \geq \delta > 0$.
5. By the symmetry and invariance of Lebesgue measure, we get
$$\begin{aligned}\mathrm{Leb}(A) &= \int_{\mathbb{R}^q} \mu^{*n}(\mathrm{d}y) \int_{\mathbb{R}^q} \mathbb{1}_A(x)\mathrm{d}x = \int_{\mathbb{R}^q} \mu^{*n}(\mathrm{d}y) \int_{\mathbb{R}^q} \mathbb{1}_A(y-x)\mathrm{d}x \\ &= \int_{\mathbb{R}^q} \mathrm{d}x \int_{\mathbb{R}^q} \mathbb{1}_A(y-x)\mu(\mathrm{d}y) = \int_{\mathbb{R}^q} \mu^{*n}(A-x)\mathrm{d}x\,.\end{aligned}$$
6. We have
$$\begin{aligned}\mathrm{Leb}(A) &= \sum_{n \geq 1} a(n)\mathrm{Leb}(A) = \sum_{n \geq 1} a(n) \int P^n(x,A)\mathrm{d}x \\ &= \int_O K_a(x,A)\mathrm{d}x \geq \delta \mathrm{Leb}(O) > 0,\end{aligned}$$
   and we thereby obtain a contradiction.

12.8 1. Compute the controllability matrix C_p.
$$C_p = [B|AB|\ldots|A^{p-1}B] = \begin{bmatrix} 1 & \eta_1 & \eta_2 & \cdots & \eta_{p-1} \\ 0 & 1 & \eta_1 & & \vdots \\ \vdots & \ddots & 1 & \ddots & \vdots \\ \vdots & & & \ddots & \eta_1 \\ 0 & 0 & \cdots & \cdots & 1 \end{bmatrix},$$
   where we define $\eta_0 = 1,\ \eta_i = 0$ for $i < 0$, and for $j \geq 2$,
$$\eta_j = \sum_{i=1}^{k} \alpha_i \eta_{j-i}.$$
   The triangular structure of the controllability matrix now implies that the pair (A,B) is controllable.
2. By Example 12.2.7, P is a T-kernel.

3. If the zeros of the polynomial $\alpha(z)$ lie outside the closed unit disk, the spectral radius $\rho(F)$ is strictly less than one, and then Example 12.2.10 shows that P is an irreducible T-kernel that admits a reachable point.

12.9 1. Let f be continuous and bounded on $[0,1]$. For all $x \in [0,1]$, we have

$$Pf(x) = xf(0) + (1-x)f(x) .$$

Thus P is Feller. Since the chain is nonincreasing starting from any value, the only accessible sets are those containing 0, and P is δ_0 irreducible.

2. For $x > 0$, we have

$$\mathbb{P}_x(\sigma_0 > n) = (1-x)^n \to 1 .$$

3. Since the only accessible sets are those that contain zero and zero is absorbing, the kernel is Harris recurrent, since the probability of eventually reaching $\{0\}$ starting from $x \neq 0$ is 1.
4. The accessible state $\{0\}$ is not uniformly accessible from X, and thus X is compact but not petite.

12.10 1. Let f be continuous and bounded on $[0,1]$. For all $x \in [0,1]$, we have

$$Pf(x) = xf(0) + (1-x)f(\alpha x) .$$

Thus P is Feller. Since the chain is decreasing starting from any value, the only accessible sets are those containing 0, and P is δ_0 irreducible.

2. For $x > 0$, we have

$$\mathbb{P}_x(\sigma_0 > n) = \prod_{k=1}^{n}(1-\alpha^k x) \to 1 .$$

3. Since the only accessible sets are those that contain zero and zero is absorbing, the kernel is recurrent. It is not Harris recurrent, since the probability of reaching $\{0\}$ starting from $x \neq 0$ is not zero.
4. The accessible state $\{0\}$ is not uniformly accessible from X, and thus X is compact but not petite.

12.11 1. We need to prove that the distribution $P^k(x,\cdot)$ is absolutely continuous with respect to Lebesgue measure and has a density that is everywhere positive on $\mathbb{R}^p$. For each deterministic initial condition $x \in \mathbb{R}^p$, the distribution of X_k is Gaussian for each $k \in \mathbb{N}$ (a linear combination of independent Gaussian vectors is also Gaussian). It is required to prove only that $P^k(x,\cdot)$ is not concentrated on some lower-dimensional subspace of $\mathbb{R}^p$. This will happen if we can show that the covariance of X_k (or equivalently of the distribution $P^k(x,\cdot)$) is of full rank for each $x \in \mathbb{R}^p$.
We compute the mean and variance of X_k for each initial condition $x \in \mathbb{R}^p$. The mean is given by $\mu_k(x) = F^k x$, and the covariance matrix is

$$\mathbb{E}_x[(X_k - \mu_k(x))(X_k - \mu_k(x))^T] = \Sigma_k := \sum_{i=0}^{k-1} F^i G G^T \{F^i\}^T .$$

The covariance is therefore of full rank if and only if the pair (F,G) is controllable. Therefore, for all $k>0$ and $x\in\mathbb{R}^p$, $P^k(x,\cdot)$ has density $p_k(x,\cdot)$ given by

$$p_k(x,y)=(2\pi|\Sigma_k|)^{-p/2}\exp\left\{-\frac{1}{2}(y-F^kx)\Sigma_k^{-1}(y-F^kx)\right\}.$$

The density is everywhere positive, as required.

2. For every compact set A, every set B of positive Lebesgue measure, and all $k\in\mathbb{N}^*$, one has $\inf_{x\in A}P^k(x,B)>0$. This proves the claim.

12.12 We set $\nu=f\cdot\mathrm{Leb}_q$ and $\mathbb{R}^s=\mathbb{R}^q\bigoplus\mathbb{R}^{s-q}$, and we choose a linear map Ψ from $\mathbb{R}^s$ to $\mathbb{R}^{s-q}$ with rank $s-q$. The linear map $\Delta=\Phi+\Psi$ is one-to-one from $\mathbb{R}^s$ to $\mathbb{R}^s$. By the change of variables formula, $\nu\circ\Delta^{-1}$ has a density proportional to $f\circ\Delta^{-1}$ with respect to Leb_s. Since $\Phi=\pi_{\mathbb{R}^q}\circ\Delta$, where $\pi_{\mathbb{R}^q}$ is the canonical projection from $\mathbb{R}^s$ to $\mathbb{R}^q$, $\nu\circ\Phi^{-1}$ has a density g with respect to Leb_q. Finally, $\xi\circ\Phi^{-1}\geq g\cdot\mathrm{Leb}_q\neq0$.

12.13 1. By Example 12.2.7, the m-skeleton P^m possesses a continuous component T that is everywhere nontrivial. By Theorem 9.2.5, there exists a small set C for which $T(x^*,C)>0$, and hence by the Feller property, an open set O containing x^* satisfying $\inf_{x\in O}T(x,C)=\delta>0$.

2. By Lemma 9.1.7 (ii), O is a small set.
3. Since for all $n\in\mathbb{N}$, $X_n=F^nX_0+\sum_{k=1}^nF^{n-k}GZ_k$, for every $x\in A$ and open neighborhood O of x^*, there exist n large enough and ε sufficiently small that on the event $\bigcap_{k=1}^n\{|Z_k-z_*|\leq\varepsilon\}$,

$$X_n=F^nx+\sum_{k=1}^nF^{n-k}GZ_k\in O,$$

showing that $\inf_{x\in A}P^n(x,O)\geq\mu^n(\mathrm{B}(z_*,\varepsilon))>0$.

4. Hence applying again Lemma 9.1.7 (ii) shows that A is a small set.

12.14 It suffices to show that $\{\pi_n^\mu,n\in\mathbb{N}\}$ is tight. By assumption, $\{\mu P^n,n\in\mathbb{N}\}$ is tight, and thus for each $\varepsilon>0$, there exists a compact set K such that $\mu P^n(K^c)\leq\varepsilon$ for all $n\geq0$. This yields $\pi_n^\mu(K^c)\leq\varepsilon$ for all $n\in\mathbb{N}$, and thus π_n^μ is tight.

12.15 If the state space is compact, then $\{\pi_n^\mu,n\in\mathbb{N}\}$ is tight for all $\mu\in\mathbb{M}_1(\mathscr{X})$.

12.16 1. Let $\mu\in\mathbb{M}_1(\mathscr{X})$ be such that $\mu(V)<\infty$ (take, for instance, $\mu=\delta_x$ for all $x\in\mathsf{X}$). By induction, (12.5.3) yields the bound $\mu P^nV\leq\mu(V)+b/(1-\lambda)$. Thus $\{\mu P^n,n\in\mathbb{N}\}$ is tight by Lemma C.2.4 and hence admits limit points that are invariant probability measures by Exercise 12.14.

2. Let π be an invariant measure. Then by concavity of the function $x\to x\wedge M$, we have, for every $M>0$,

$$\pi(V\wedge M)=\pi P^n(V\wedge M)\leq\pi((P^nV)\wedge M)\leq\pi(\{\lambda^nV+b/(1-\lambda)\}\wedge M).$$

Letting n first and then M tend to infinity yields $\pi(V)\leq b/(1-\lambda)$.

12.17 1. Since P is Feller, if $f \in \mathrm{C}_b(\mathsf{X})$, then $Pf \in \mathrm{C}_b(\mathsf{X})$. Therefore,

$$\pi P(f) = \pi(Pf) = \lim_{n\to\infty} (\mu P^n)Pf = \lim_{n\to\infty} \mu P^{n+1}(f) = \pi(f)\ .$$

Thus πP and π take equal values on all bounded continuous functions and are therefore equal by Corollary B.2.18.

2. π is invariant by item 1 above. For all $f \in \mathrm{C}_b(\mathsf{X})$ and $x \in \mathsf{X}$, we get $\lim_{n\to\infty} P^n f(x) = \lim_{n\to\infty} \delta_x P^n(f) = \pi(f)$ and $|P^n f(x)| \leq |f|_\infty$. Therefore, for $\xi \in \mathbb{M}_1(\mathscr{X})$, Lebesgue's dominated convergence theorem yields

$$\lim_{n\to\infty} \xi P^n(f) = \lim_{n\to\infty} \int_{\mathsf{X}} P^n f(x)\, \xi(\mathrm{d}x) = \int_{\mathsf{X}} \lim_{n\to\infty} P^n f(x)\, \xi(\mathrm{d}x) = \pi(f)\ .$$

Thus $\xi P^n \stackrel{\mathrm{w}}{\Rightarrow} \pi$. If, moreover, ξ is invariant, then $\xi = \xi P^n$ for all n, whence $\xi = \pi$.

12.18 1. The homogeneous Poisson point process is stochastically continuous, so $Ph(x) = \mathbb{E}\,[h(\omega + bx + c\log(1+N(\mathrm{e}^x)))]$ is a continuous function of x. Thus P is Feller.

2. We use the bound $|u+v| = |u|+|v|$ if $uv \geq 0$ and $|u+v| \leq |u| \vee |v|$ otherwise. If $bc \geq 0$, this yields

$$\begin{aligned} PV(x) &= \mathbb{E}\left[\mathrm{e}^{|\omega+bx+c\log(1+N(\mathrm{e}^x))|}\right] \leq \mathrm{e}^{|\omega|}\mathrm{e}^{|b||x|}\mathbb{E}\left[(1+N(\mathrm{e}^x))^{|c|}\right] \\ &\leq \mathrm{e}^{|\omega|}\mathrm{e}^{|b||x|}(1+\mathbb{E}\,[N(\mathrm{e}^x)])^{|c|} \leq \vartheta \mathrm{e}^{|b+c||x|}\ . \end{aligned}$$

If $bc < 0$, we obtain

$$\begin{aligned} PV(x) &= \mathbb{E}\left[\mathrm{e}^{|\omega+bx+c\log(1+N(\mathrm{e}^x))|}\right] \leq \mathrm{e}^{|\omega|}\left(\mathrm{e}^{|b||x|} + \mathbb{E}\left[(1+N(\mathrm{e}^x))^{|c|}\right]\right) \\ &\leq \mathrm{e}^{|\omega|}(\mathrm{e}^{|b||x|} + \mathrm{e}^{|c||x|}) \leq \vartheta \mathrm{e}^{(|b|\vee|c|)|x|}\ . \end{aligned}$$

This proves that PV/V tends to zero at infinity and is bounded on compact sets, and therefore the drift condition (12.3.3) holds.

3. By Exercise 12.16, the properties we have just shown prove that the kernel P admits an invariant probability measure.

12.19 We take the continuous component to be the part of the kernel corresponding to accepted updates, that is,

$$T(x,A) = \int_A q(x,\ y)\alpha(x,y)\mathrm{d}y\ , \tag{G.19}$$

where we define

$$\alpha(x,\ y) = \begin{cases} 1, & h_\pi(y)q(y,\ x) \geq h_\pi(x)q(x,\ y), \\ \frac{h_\pi(y)q(y,x)}{h_\pi(x)q(x,y)}, & \text{otherwise}. \end{cases} \tag{G.20}$$

Fix y and consider a sequence $x_n \to x$ with $x \in \mathsf{X}_\pi$. It is clear that if $q(x,\ y) > 0$, then

$$\alpha(x_n,\ y)q(x_n,\ y) \to \alpha(x,\ y)q(x,\ y)$$

by the continuity assumptions of the theorem. In case $q(x,\ y) = 0$, we have

$$0 \le \alpha(x_n,\ y)q(x_n,\ y) \le q(x_n,\ y) \to 0$$

by the continuity assumptions of the theorem and our definition of $\alpha(x,\ y)$. The integrand in (G.19) being an lower semicontinuous function for each fixed value of the variable of integration, so is the integral by Fatou's lemma. It remains only to be shown that $T(x, \mathsf{X}_\pi) > 0$ for every $x \in \mathsf{X}_\pi$, but if this failed for any x, it would mean that the chain could never move from x to anywhere.

12.20 Without loss of generality, assume that

$$\frac{\partial F_k}{\partial z_k}(x_0^0, z_1^0, \ldots, z_k^0) \neq 0 \tag{G.21}$$

with $(z_1^0, \ldots, z_k^0) \in \mathbb{R}^k$. Consider the function $F^k : \mathbb{R}^{k+1} \to \mathbb{R}^{k+1}$ given by

$$F^k(x_0, z_1, \ldots, z_k) = (x_0, z_1, \ldots, z_{k-1}, x_k)^T,$$

where $x_k = F_k(x_0, z_1, \ldots, z_k)$. The Jacobian of F^k is given by

$$DF^k := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & 1 & 0 \\ \frac{\partial F_k}{\partial x_0} & \frac{\partial F_k}{\partial z_1} & \cdots & \frac{\partial F_k}{\partial z_k} \end{pmatrix}, \tag{G.22}$$

which is of full rank at $(x_0^0, z_1^0, \ldots, z_k^0)$. By the inverse function theorem, there exist an open set $B = B_{x_0^0} \times B_{z_1^0} \times \cdots \times B_{z_k^0}$ containing $(x_0^0, z_1^0, \ldots, z_k^0)$ and a smooth function $G^k : \{F^k\{B\}\} \to \mathbb{R}^{k+1}$ such that

$$G^k(F^k(x_0, z_1, \ldots, z_k)) = (x_0, z_1, \ldots, z_k),$$

for all $(x_0, z_1, \ldots, z_k) \in B$. Taking G_k to be the last component of G^k, we get, for all $(x_0, z_1, \ldots, z_k) \in B$,

$$G_k(x_0, z_1, \ldots, z_{k-1}, x_k) = G_k(x_0, z_1, \ldots, z_{k-1}, F_k(x_0, z_1, \ldots, z_k)) = z_k.$$

For all $x_0 \in B_{x_0^0}$ and positive nonnegative Borel functions f, define

$$\begin{aligned} P^k f(x_0) &= \int \cdots \int f(F_k(x_0, z_1, \ldots, z_k))p(z_k)\cdots p(z_1)\mathrm{d}z_1 \cdots \mathrm{d}z_k \\ &\ge \int_{B_{z_1^0}} \cdots \int_{B_{z_k^0}} f(F_k(x_0, z_1, \ldots, z_k))p(z_k)\cdots p(z_1)\mathrm{d}z_1 \cdots \mathrm{d}z_k. \end{aligned} \tag{G.23}$$

We integrate first over z_k, the remaining variables being fixed. Using the change of variables

$$x_k = F_k(x_0, z_1, \dots, z_k), z_k = G_k(x_0, z_1, \dots, z_{k-1}, x_k),$$

we obtain for $(x_0, z_1, \dots, z_{k-1}) \in B_{x_0^0} \times \cdots \times B_{z_{k-1}^0}$,

$$\int_{B_{z_k^0}} f(F_k(x_0, z_1, \dots, z_k)) p(z_k) \mathrm{d}z_k = \int_{\mathbb{R}} f(x_k) q_k(x_0, z_1, \dots, z_{k-1}, x_k) \mathrm{d}x_k, \tag{G.24}$$

where setting $\xi := (x_0, z_1, \dots, z_{k-1}, x_k)$, $q_k(\xi)$ is given by

$$q_k(\xi) := \mathbb{1}_B(G^k(\xi)) p(G_k(\xi)) \left| \frac{\partial G_k}{\partial x_k}(\xi) \right| .$$

Since q_k is positive and lower semicontinuous on the open set $F^k\{B\}$ and zero on $F^k\{B\}^c$, it follows that q_k is lower semicontinuous on $\mathbb{R}^{k+1}$. Define the kernel T_0 for an arbitrary bounded function f as

$$T_0 f(x_0) := \int \cdots \int f(x_k) q_k(\xi) p(z_1) \cdots p(z_{k-1}) \mathrm{d}z_1 \cdots \mathrm{d}z_{k-1} \mathrm{d}x_k. \tag{G.25}$$

The kernel T_0 is nontrivial at x_0^0, since

$$q_k(\xi^0) p(z_1^0) \cdots p(z_{k-1}^0) = \left| \frac{\partial G_k}{\partial x_k}(\xi^0) \right| p(z_k^0) p(z_1^0) \cdots p(z_{k-1}^0) > 0,$$

where $\xi^0 = (x_0^0, z_1^0, \dots, z_{k-1}^0, x_k^0)$. We will show that $T_0 f$ is lower semicontinuous on $\mathbb{R}$ whenever f is positive and bounded.

Since $q_k(x_0, z_1, \dots, z_{k-1}, x_k) p(z_1) \cdots p(z_{k-1})$ is lower semicontinuous, there exists a sequence of nonnegative, continuous functions $r_i : \mathbb{R}^{k+1} \to \mathbb{R}_+$, $i \in \mathbb{N}$, such that for each i, the function r_i has bounded support and as $i \uparrow \infty$,

$$r_i(x_0, z_1, \dots, z_{k-1}, x_k) \uparrow q_k(x_0, z_1, \dots, z_{k-1}, x_k) p(z_1) \cdots p(z_{k-1})$$

for each $(x_0, z_1, \dots, z_{k-1}, x_k) \in \mathbb{R}^{k+1}$. Consider the kernel T_i,

$$T_i f(x_0) := \int_{\mathbb{R}^k} f(x_k) r_i(x_0, z_1, \dots, z_{k-1}, x_k) \mathrm{d}z_1 \cdots \mathrm{d}z_{k-1} \mathrm{d}x_k.$$

It follows from Lebesgue's dominated convergence theorem that $T_i f$ is continuous for every bounded function f. If f is also positive, then as $i \uparrow \infty$, $T_i f(x_0) \uparrow T_0 f(x_0)$, $x_0 \in \mathbb{R}$, showing that $T_0 f$ is lower semicontinuous.

Using (G.23) and (G.24), it follows that T_0 is a continuous component of P^k and P is a T-kernel.

12.21 1. By applying Theorem 12.4.3 with $x \notin R$, there exists V_n such that $\mathbb{P}_x(\sigma_{V_n} < \infty) < 1$.

2. For $y \in A_n(j)$, we have $\mathbb{P}_y(\sigma_{A_n(j)} < \infty) \le \mathbb{P}_y(\sigma_{V_n} < \infty) \le 1 - 1/j$. By Proposition 4.2.5, this implies that $\sup_{y \in \mathsf{X}} U(y, A_n(j)) < \infty$ and $A_n(j)$ is uniformly transient. Thus R^c is transient.

12.22 If $X_0 = x_0 \in \mathbb{R} \setminus \mathbb{Q}$, then the sequence $\{X_n,\ n \in \mathbb{N}\}$ is a sequence of i.i.d. random variables with distribution ν. By assumption, $\nu(U) > 0$ for all open sets U. Thus by the strong law of large numbers, every open set is visited infinitely open starting from any irrational number. If $X_0 = x_0 \in \mathbb{Q}$, then the sequence $\{X_n,\ n \in \mathbb{N}\}$ is a sequence of i.i.d. random variables with values in $\mathbb{Q}$ and distribution μ. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, $\mu(U) > 0$ for every open set U, and thus by the strong law of large numbers, it is also the case that every open set is visited infinitely often starting from any rational number. Thus the kernel P is topologically Harris recurrent.

The measures μ and ν are both invariant. If $X_0 \sim \mu$, then $\{X_n,\ n \in \mathbb{N}\}$ is a sequence of i.i.d. random variables with distribution μ. If $X_0 \sim \nu$, then $\mathbb{P}_\nu(\exists n \ge 0\,,\ X_n \in \mathbb{Q}) = 0$; therefore, $\{X_n,\ n \in \mathbb{N}\}$ is a sequence of i.i.d. random variables with distribution μ.

12.23 1. Since X is an increasing union of compact sets, there exists an accessible compact set K. Since P is evanescent, $\mathbb{P}_x(N_K = \infty) = 0$ for all $x \in \mathsf{X}$.
2. Assume that P is recurrent. By Corollary 10.2.8, K contains an accessible Harris-recurrent set $\tilde{K}$. For all $x \in \tilde{K}$, $1 = \mathbb{P}_x(N_{\tilde{K}} = \infty) \le \mathbb{P}_x(N_K = \infty)$, and this is a contradiction. Hence P is transient.

12.24 1. By Theorem 10.1.5, P is not transient. Therefore, by Exercise 12.23, P is not evanescent, and there exists $x_0 \in \mathsf{X}$ such that $0 \le h(x_0) < 1$, where $h(x) = \mathbb{P}_x(X_n \to \infty) < 1$.
2. The set $A = \{X_n \to \infty\}$ being invariant, the function h is harmonic by Proposition 5.2.2. By Theorem 10.2.11, since P is Harris recurrent, bounded harmonic functions are constants, which implies $h(x) = h(x_0)$ for every $x \in \mathsf{X}$.
3. By the martingale convergence theorem, Theorem E.3.1, $\mathbb{P}_{X_n}(A) = \mathbb{P}_x(A \mid \mathscr{F}_n)$ converges $\mathbb{P}_x$ almost surely to $\mathbb{1}_A$ for all $x \in \mathsf{X}$. Therefore, $h(x) = 0$ for all $x \in \mathsf{X}$.

12.25 1. The assumption means that V is superharmonic outside C (see Definition 4.1.1). By Theorem 4.1.2, $\{V(X_{n \wedge \tau_C}), n \ge 0\}$ is a positive supermartingale. Since $V(X_0) < \infty$, by the supermartingale convergence theorem (Proposition E.1.3), there exists a random variable M_∞ that is $\mathbb{P}_x$ almost surely finite for all $x \in \mathsf{X}$ such that for all $n \in \mathbb{N}$, $V(X_{n \wedge \tau_C}) \to M_\infty$.
2. Since V tends to infinity, this implies that $\mathbb{P}_x(\sigma_C = \infty,\ X_n \to \infty) = 0$ for all $x \in \mathsf{X}$.
3. If $X_n \to \infty$, then there exists an integer p such that $\sigma_C \circ \theta_p = \infty$, i.e., the chain does not return to C after p. The events $\{\sigma_C \circ \theta_p = \infty\}$ are increasing, and thus

$$\{X_n \to \infty\} = \bigcup_{p \ge 0} \{X_n \to \infty, \sigma_C \circ \theta_p = \infty\}\ .$$

4. Since obviously $\{X_n \to \infty\} = \{X_n \circ \theta_p \to \infty\}$, we obtain

$$\begin{aligned}\mathbb{P}_x(X_n \to \infty) &= \lim_{p\to\infty} \mathbb{P}_x(X_n \to \infty, \sigma_C \circ \theta_p = \infty) \\ &= \lim_{p\to\infty} \mathbb{E}_x\left[\mathbb{P}_{X_p}(X_n \to \infty, \sigma_C = \infty)\right] = 0\,.\end{aligned}$$

12.26 1. By Lemma 10.1.8 (ii), $\tilde{A}$ is transient, so we can write $\tilde{A} = \bigcup_{i=1}^{\infty} \tilde{A}_i$, where the sets $\tilde{A}_i$ are uniformly transient.

2. By definition of the sets $\tilde{A}$ and A^0, $\mathsf{X} = \tilde{A} \cup A^0$, and by definition of a T-kernel, $T(x,\mathsf{X}) > 0$ for all $x \in \mathsf{X}$. Thus there exists $j > 0$ such that either $T(x,A^0) > 1/j$ or there exists $i > 0$ such that $T(x,\tilde{A}_i) > 1/j$. So if we set

$$U_j = \left\{x\,:\, T(x,A^0) > 1/j\right\}\,, \quad U_{i,j} = \left\{x\,:\, T(x,A_i) > 1/j\right\}\,,$$

we obtain that $(U_i, U_{i,j},\, i,j > 0)$ is a covering of X, and moreover, these sets are open, since $T(\cdot,A)$ is lower semicontinuous for every measurable set A by definition of a T-kernel.

3. Since $K \subset \bigcup_{j\ge 1, i\ge 1}(U_i \cup U_{i,j})$, the compactness property implies that K can be covered by finitely many U_j and $U_{i,j}$. Since the sequences $U_j, U_{i,j}$ are increasing with respect to j, there exists $k \ge 1$ such that $K \subset U_k \cup \bigcup_{i=1}^{k} U_{i,k}$.

4. By Lemma 10.1.8 (i), each set $U_{i,k}$ is uniformly transient, thus visited only a finite number of times; therefore, $\{X_n \in K \text{ i.o.}\} \subset \{X_n \in U_k \text{ i.o.}\}$ $\mathbb{P}_x - \text{a.s.}$

5. For $y \in U_k$, we have

$$\begin{aligned}\mathbb{P}_y(\sigma_{A^0} < \infty) &= \sum_{k=0}^{\infty} a(k)\mathbb{P}_y(\sigma_{A^0} < \infty) \\ &\ge \sum_{k=0}^{\infty} a(k)\mathbb{P}_y(X_k \in A^0) = K_a(y,A^0) \ge T(y,A^0) = 1/k\,.\end{aligned}$$

6. By Theorem 4.2.6, this implies that $\{N_{U_k} = \infty\} \subset \{N_{A^0} = \infty\} \subset \{\sigma_{A^0} < \infty\}$ $\mathbb{P}_x - \text{a.s.}$

12.27 1. If P is evanescent, then P is transient by Exercise 12.23. Conversely, if P is transient, we apply Exercise 12.26 with $A = \mathsf{X}$, and then $A^0 = \emptyset$ and $\mathbb{P}_x(X_n \to \infty) = 1$ for all $x \in \mathsf{X}$.

2. By Theorem 10.1.5, P is not transient if and only if it is recurrent, and thus the statements 1 and 2 are equivalent.

3. If P is Harris recurrent, then P is nonevanescent by Exercise 12.24. Conversely, assume that P is nonevanescent. Then it is recurrent by question 2, and by Theorem 10.2.7, we can write $\mathsf{X} = H \cup N$ with H maximal absorbing, N transient, and $H \cap N = \emptyset$. We must prove that N is empty. Since H is maximal absorbing, if $x \in N$, then $\mathbb{P}_x(\sigma_H < \infty) < 1$, and hence $\mathbb{P}_x(\sigma_N < \infty) > 0$, since $H \cup N = \mathsf{X}$. This means that $N^0 = H$, where $N^0 = \{x \in \mathsf{X}\,:\, \mathbb{P}_x(\sigma_N < \infty) = 0\}$. Since P is nonevanescent, $\mathbb{P}_x(X_n \to \infty) = 0$. By Exercise 12.26, this implies that $\mathbb{P}_x(\sigma_H < \infty) = 1$, which is impossible, since H is maximal absorbing and $H \cap N = \emptyset$. Therefore, N is empty and P is Harris recurrent.

Solutions to Exercises of Chapter 13

13.1 1. The bound (13.5.1) follows from (8.3.4) and Proposition 13.2.11.

2. Note that the conditions $\mathbb{E}_\lambda[\sigma_\alpha^s] + \mathbb{E}_\mu[\sigma_\alpha^s] < \infty$ imply that $\mathbb{P}_\lambda(\sigma_\alpha < \infty) = 1$ and $\mathbb{P}_\mu(\sigma_\alpha) = 1$. Proposition 13.2.7 shows that $\bar{\mathbb{P}}_{\lambda\otimes\pi}(T < \infty) = 1$.
By Proposition 13.2.9, we have

$$\bar{\mathbb{E}}_{\lambda\otimes\mu}[T^s] \leq C\{\mathbb{E}_\lambda[\sigma_\alpha^s] + \mathbb{E}_\mu[\sigma_\alpha^s]\}.$$

The condition $\mathbb{E}_\lambda[\sigma_\alpha^s] + \mathbb{E}_\mu[\sigma_\alpha^s] < \infty$ implies that $\bar{\mathbb{E}}_{\lambda\otimes\mu}[T^s] < \infty$. Note that

$$n^s \bar{\mathbb{P}}_{\lambda\otimes\mu}(T \geq n) \leq \bar{\mathbb{E}}_{\lambda\otimes\mu}[T^s \mathbb{1}_{\{T\geq n\}}] \,.$$

Since $\bar{\mathbb{P}}_{\lambda\otimes\mu}(T < \infty) = 1$ and $\bar{\mathbb{E}}_{\lambda\otimes\mu}[T^s] < \infty$, Lebesgue's dominated convergence theorem shows that $\lim_{n\to\infty} n^s \bar{\mathbb{P}}_{\lambda\otimes\mu}(T \geq n) = 0$. The proof is concluded by Lemma 8.3.1, which shows that for all $n \in \mathbb{N}$, $\mathrm{d}_{\mathrm{TV}}(\lambda P^n, \mu) \leq \bar{\mathbb{P}}_{\lambda\otimes\mu}(T \geq n)$.

13.2 Consider the forward recurrence time chain $\{A_n,\ n \in \mathbb{N}\}$ on $\mathbb{N}^*$.

The state 1 is an accessible positive recurrent atom, and it is aperiodic, since b is aperiodic. The distribution of the return time to 1 is the waiting distribution b if the chain starts at 1; hence for every sequence $\{r(n),\ n \in \mathbb{N}\}$,

$$\mathbb{E}_1\left[r(\sigma_1)\right] = \sum_{n=1}^{\infty} r(n)b(n) \,.$$

Without loss of generality, we can assume that the delay distribution a puts no mass at zero. Then applying the identity (8.1.15), the distribution of A_0 is a, and since $\sigma_1 = A_0 - 1$ if $A_0 \geq 2$, we have, for $n \geq 1$,

$$\mathbb{P}_a(\sigma_1 = n) = a(n+1) + a(1)b(n) \,.$$

This yields the equivalence, for every sequence $\{r(n),\ n \in \mathbb{N}\}$,

$$\mathbb{E}_a[r(\sigma_1)] = \sum_{n=1}^{\infty} r(n)a(n+1) + a(1)\sum_{n=1}^{\infty} r(n)b(n) \,.$$

The pure and delayed renewal sequences u and v_a are given by

$$u(n) = \mathbb{P}_1(A_n = 1) \,, \quad v_a(n) = \mathbb{P}_a(A_n = 1) \,.$$

With Q the kernel of the forward recurrence time chain, this yields

$$|v_a(n) - u(n)| \leq \|aQ^n - Q^n(1,\cdot)\|_{\mathrm{TV}} \,.$$

If a is the invariant probability for Q given in (8.1.17), then $v_a(n) = m^{-1}$ for all $n \geq 1$. We can now translate Theorems 13.3.1 and 13.3.3 into the language of renewal theory.

13.3 1. Clearly all the states communicate and $\{0\}$ is an aperiodic atom. Easy computations show that for all $n \geq 1$, $\mathbb{P}_0(\sigma_0 = n+1) = (1-p_n)\prod_{j=0}^{n-1} p_j$ and $\mathbb{P}_0(\sigma_0 > n) = \prod_{j=0}^{n-1} p_j$. By Theorem 6.4.2, P is positive recurrent since $\mathbb{E}_0[\sigma_0] < \infty$, and the stationary distribution π is given, by $\pi(0) = \pi(1) = 1/\mathbb{E}_0[\sigma_0]$ and for $j \geq 2$,

$$\pi(j) = \frac{\mathbb{E}_0\left[\sum_{k=1}^{\sigma_0} \mathbb{1}_{\{j\}}(X_k)\right]}{\mathbb{E}_0[\sigma_0]} = \frac{\mathbb{P}_0(\sigma_0 \geq j)}{\mathbb{E}_0[\sigma_0]} = \frac{p_0 \cdots p_{j-2}}{\sum_{n=1}^{\infty} p_1 \ldots p_n}.$$

2. It suffices to note that $\mathbb{P}_0\left(X_k = k \,|\, \sigma_0 > k\right) = 1$.
3. For all $\lambda < \mu < 1$, $\mathbb{E}_0[\mu^{-\sigma_0}] < \infty$, and $\{0\}$ is thus a geometrically ergodic atom.
4. It is easily seen that $\prod_{i=1}^{n} p_i = O(n^{-1-\theta})$. We have $\mathbb{E}_0\left[\sum_{k=0}^{\sigma_0-1} r(k)\right] < \infty$ if and only if $\sum_{k=1}^{\infty} r(k) k^{-1-\theta} < \infty$. This shows that $\mathbb{E}_0\left[\sum_{k=0}^{\tau_0-1} r(k)\right] < \infty$ for $r(k) = O(k^{\beta})$ for all $\beta \in [0,\theta)$. The statement follows by noting that $\mathbb{E}_\lambda[r(\sigma_0)] \leq \mathbb{E}_0[r(\sigma_0)]$ for every initial distribution λ and applying Theorem 13.3.3.

13.4 For every fixed x,

$$\begin{aligned}
\left(\sum_y |M^k(x,\, y) - \pi(y)|\right)^2 &= \left(\sum_y \frac{|M^k(x,y) - \pi(y)|}{\sqrt{\pi(y)}} \sqrt{\pi(y)}\right)^2 \\
&\leq \sum_y \frac{|M^k(x,y) - \pi(y)|^2}{\pi(y)} \\
&= \sum_y \frac{\{M^k(x,y)\}^2}{\pi(y)} - 1 = \frac{M^{2k}(x,x)}{\pi(x)} - 1 \\
&= \frac{1}{\pi(x)} \sum_y \beta_y^{2k} f_y^2(x) - 1.
\end{aligned}$$

Solutions to Exercises of Chapter 14

14.1 By definition of the set C, for $x \in \mathsf{X}$, we have

$$b = b\mathbb{1}_{C^c}(x) + b\mathbb{1}_C(x) \leq \frac{b}{d} V(x) + b\mathbb{1}_C(x)\,.$$

Thus $\mathrm{D_g}(V, \lambda, b)$ implies

$$PV \le \lambda V + b \le \lambda V + \frac{b}{d} V + b \mathbb{1}_C = \bar{\lambda} V + b \mathbb{1}_C \,,$$

where $\bar{\lambda} = \lambda + b/d$.

14.2 An application of Proposition 9.2.13 with $V_0 = V_1 = W_C^{f,\delta}$ proves that the set $\{W_C^{f,\delta} < \infty\}$ is full and absorbing and $\{W_C^{f,\delta} \le d\}$ is accessible for all sufficiently large d. The level sets $\{W_C^{f,\delta} \le d\}$ are petite by Lemma 9.4.8.

14.3 1. Since $V \ge 1$, the drift condition (14.5.1) implies that

$$PV + 1 - \lambda \le \lambda V + (1-\lambda) + b\mathbb{1}_C \le V + b\mathbb{1}_C \,.$$

Applying Proposition 4.3.2 with $f = 1$, we obtain $\mathbb{P}_x(\sigma_C < \infty) = 1$ for all $x \in \mathsf{X}$ such that $V(x) < \infty$. The bound (14.5.2) follows from Proposition 14.1.2 (i) with $\delta = \lambda^{-1}$ and $f \equiv 0$.

2. For $\delta \in (1, 1/\lambda)$, the drift condition (14.5.1) yields

$$PV + \delta^{-1}(1 - \delta\lambda)V \le \delta^{-1} V + b\mathbb{1}_C \,. \tag{G.26}$$

Thus the bound (14.5.3) follows from Proposition 14.1.2 (i), with $f = \delta^{-1}(1 - \delta\lambda)V$.

3. Follows from Lemma 14.1.10 and the bound (14.5.2).

14.5 1. Recall that $W(x) = \mathrm{e}^{\beta|x|}$. Then

$$PW(x) = \mathbb{E}_x[\mathrm{e}^{\beta|X_1|}] \le \mathbb{E}\left[\mathrm{e}^{\beta|h(x)|}\mathrm{e}^{\beta|Z_1|}\right] = K\mathrm{e}^{\beta|h(x)|} \,.$$

2. For $|x| > M$, $\mathrm{e}^{\beta|h(x)|} \le \mathrm{e}^{\beta|x|}\mathrm{e}^{-\beta\ell}$, which implies

$$PW(x) \le K\mathrm{e}^{-\beta\ell} W(x) = \lambda W(x) \,.$$

3. For $|x| \le M$, $PW(x) \le b$, where $b = K \sup_{|x|\le M} \mathrm{e}^{\beta|h(x)|}$.

14.8 True for $n = 0$. If $n \ge 0$, (14.5.5) yields

$$\begin{aligned}
&\mathbb{E}_x[\pi_{(n+1)\wedge\sigma_C - 1} V_{(n+1)\wedge\sigma_C}] \\
&\quad = \mathbb{E}_x[\pi_{n-1} f(X_n) V(X_{n+1}) \mathbb{1}\{n < \sigma_C\}] \\
&\qquad\quad + \mathbb{E}_x[\pi_{\sigma_C - 1} V_{\sigma_C} \mathbb{1}\{\sigma_C \le n\}] + b\mathbb{1}\{n = 0\}\mathbb{1}_C(x) \\
&\quad \le \mathbb{E}_x[\pi_{n-1} V(X_n) \mathbb{1}\{n < \sigma_C\}] + \mathbb{E}_x[\pi_{\sigma_C - 1} V_{\sigma_C} \mathbb{1}\{\sigma_C \le n\}] \\
&\quad \le \mathbb{E}_x[\pi_{n-1} V(X_n) \mathbb{1}\{n < \sigma_C\}] + \mathbb{E}_x[\pi_{n-1} V(X_n) \mathbb{1}\{\sigma_C = n\}] \\
&\qquad\quad + \mathbb{E}_x[\pi_{\sigma_C - 1} V_{\sigma_C} \mathbb{1}\{\sigma_C \le n-1\}] + b\mathbb{1}\{n = 0\}\mathbb{1}_C(x) \\
&\quad \le \mathbb{E}_x[\pi_{n-1} V(X_n) \mathbb{1}\{n - 1 < \sigma_C\}] \\
&\qquad\quad + \mathbb{E}_x[\pi_{\sigma_C - 1} V_{\sigma_C} \mathbb{1}\{\sigma_C \le n-1\}] + b\mathbb{1}\{n = 0\}\mathbb{1}_C(x) \\
&\quad = \mathbb{E}_x[\pi_{(n\wedge\sigma_C - 1)} V_{n\wedge\sigma_C}] + b\mathbb{1}\{n = 0\}\mathbb{1}_C(x) \,.
\end{aligned}$$

By induction and since $V \geq 1$, this yields

$$\mathbb{E}_x[\pi_{(n+1)\wedge\sigma_C - 1}] \leq \mathbb{E}_x[\pi_{(n+1)\wedge\sigma_C - 1} V_{(n+1)\wedge\sigma_C}] \leq V(x) + b\mathbb{1}_C(x) \,.$$

Letting $n \to \infty$ yields (14.5.6).

14.9 1. The proof of (14.5.7) follows by an easy induction. Equation (14.5.7) implies that $P^m V^{(m)} + f^{(m)} \leq \lambda V^{(m)} + \lambda^{-(m-1)} b(1-\lambda^m)/(1-\lambda)$.

2. Since P is irreducible and aperiodic, Theorem 9.3.11 shows that P^m is irreducible and that the level sets $\{V^{(m)} \leq d\}$ are accessible and petite for d large enough. The proof of (14.5.8) follows from Theorem 14.1.14 applied to P^m.
3. If P is f-geometrically regular, then Theorem 14.2.6 (ii) shows that there exist a function $V : \mathsf{X} \to [0,\infty]$ such that $\{V < \infty\} \neq \emptyset$, a nonempty petite set C, $\lambda \in [0,1)$, and $b < \infty$ such that $PV + f \leq \lambda V + b\mathbb{1}_C$. If, moreover, P is aperiodic, Exercise 14.9 shows that then $P^m V^{(m)} + f^{(m)} \leq \lambda^{(m)} V^{(m)} + b^{(m)} \mathbb{1}_D$, where $\lambda^{(m)} \in [0,1)$, D is a nonempty petite set, and $V^{(m)} = \lambda^{-(m-1)}$. Using again Theorem 14.2.6 (ii), the Markov kernel P^m is therefore $f^{(m)}$-geometrically regular.
4. Using Theorem 14.2.6 (b) again for P^m, we get that every probability measure ξ satisfying $\xi(V) < \infty$ is f-geometrically regular for P and $f^{(m)}$-geometrically regular for P^m.

Solutions to Exercises of Chapter 15

15.3 The Markov kernel of this chain has a density with respect to the Lebesgue measure given by

$$p(x,y) = \frac{1}{\sqrt{2\pi\sigma^2(x)}} \exp\left(-\frac{1}{2\sigma^2(x)}(y - f(x))^2\right) \,.$$

Then for $y \in [-1,1]$, we have

$$\inf_{x\in\mathbb{R}} p(x,y) \geq \frac{1}{\sqrt{2\pi b}} \exp\left(-\frac{1}{2a}(y - \inf_{x\in\mathbb{R}} f(x))^2 \vee (y - \sup_{x\in\mathbb{R}} f(x))^2\right) > 0 \,.$$

Hence the state space $\mathbb{R}$ is a 1-small set, and the Markov kernel P is uniformly geometrically ergodic by Theorem 15.3.1 (iii).

15.6 Assume that all the states are accessible and that P aperiodic. Choose $x_0 \in \mathsf{X}$. By Proposition 6.3.6, there exists an integer m_0 such that $P^n(x_0,x_0) > 0$ for all $n > m_0$. Since all the states are accessible, for every $x \in \mathsf{X}$ there exists an integer $m(x)$ such that $P^{m(x)}(x,x_0) > 0$. This in turn implies that for $m = \max_{x\in\mathsf{X}}(m(x)+m_0)$, we have $P^m(x,x_0) > 0$. Set then $\zeta_m = \inf_{x'\in\mathsf{X}} P^m(x',x_0) > 0$. This yields $P^m(x,A) \geq \zeta_m \delta_{x_0}(A)$ for every $x \in \mathsf{X}$. Hence the state space is small, and the Markov kernel is therefore uniformly geometrically ergodic.

15.7 1. Since $(x+y)^s \le x^s + y^s$ for all $x, y > 0$, we have

$$\begin{aligned} PV(x) &= \mathbb{E}_x[V(X_1)] = 1 + (\alpha_0 + \alpha_1 x^2)^s \mu_{2s} \\ &\le 1 + \alpha_0^s \mu_{2s} + \alpha_1^s \mu_{2s} x^{2s} \le \lambda V(x) + b \,, \end{aligned}$$

with $\lambda = \alpha_1^s \mu_{2s}$ and $b = 1 - \alpha_1^s \mu_{2s} + \alpha_0^s \mu_{2s}$. Thus, provided that $\alpha_1^s \mu_{2s} < 1$, the transition kernel P satisfies the geometric drift condition $\mathrm{D_g}(V, \lambda, b)$.

2. For $A \in \mathscr{B}(\mathbb{R})$ and $x \in [-c, c]$, we have

$$\begin{aligned} P(x,A) &= \int_{-\infty}^{\infty} \mathbb{1}_A\left((\alpha_0 + \alpha_1 x^2)^{1/2} z\right) g(z)\mathrm{d}z \\ &= (\alpha_0 + \alpha_1 x^2)^{-1/2} \int_{-\infty}^{\infty} \mathbb{1}_A(v) g((\alpha_0 + \alpha_1 x^2)^{-1/2} v)\mathrm{d}v \\ &\ge (\alpha_0 + \alpha_1 c^2)^{-1/2} g_{\min} \int_{-\infty}^{\infty} \mathbb{1}_A(v) \mathbb{1}_{[-a,a]}(\alpha_0^{-1/2} v)\mathrm{d}v \\ &= 2a\alpha_0^{1/2} (\alpha_0 + \alpha_1 c^2)^{-1/2} \frac{1}{2a\alpha_0^{1/2}} \int_{-a\alpha_0^{1/2}}^{a\alpha_0^{1/2}} \mathbb{1}_A(v)\mathrm{d}v \,. \end{aligned}$$

If we set $\varepsilon = 2a\alpha_0^{1/2}(\alpha_0 + \alpha_1 b^2)^{-1/2} g_{\min}$ and define the measure ν by

$$\nu(A) = \frac{1}{2a\sqrt{\alpha_0}} \mathrm{Leb}(A \cap [-a\sqrt{\alpha_0}, a\sqrt{\alpha_0}]) \,,$$

we obtain that $P(x,A) \ge \varepsilon \nu(A)$ for all $x \in [-a, a]$. Thus every bounded interval and hence every compact set of $\mathbb{R}$ is small.

15.8 Let P be the Markov kernel of the INAR process and let V be the identity function on $\mathbb{N}$, i.e., $V(x) = x$ for all $x \in \mathbb{N}$. Then the kernel P satisfies a geometric drift condition with Lyapunov function V. Indeed,

$$PV(x) = mx + \mathbb{E}[Y_1] = mV(x) + \mathbb{E}[Y_1] \,.$$

Fix $\eta \in (0,1)$ and let k_0 be the smallest integer such that $k_0 > \mathbb{E}[Y_1]/\eta$ (assuming implicitly the latter expectation to be finite). Define $C = \{0, \ldots, k_0\}$ and $b = \mathbb{E}[Y_1]$. These choices yield

$$PV \le (m + \eta)V + b\mathbb{1}_C \,.$$

Let ν denote the distribution of Y_1. Then for $x, y \in \mathbb{N}$, we have

$$\begin{aligned} P(x,y) &= \mathbb{P}\left(\sum_{i=1}^{x} \xi_i^{(1)} + Y_1 = y\right) \\ &\ge \mathbb{P}\left(\sum_{i=1}^{x} \xi_i^{(1)} + Y_1 = y \,, \xi_1^{(1)} = 0, \ldots, \xi_1^{(x)} = 0\right) = \mu(y) \,. \end{aligned}$$

Since $m < 1$ implies that $\mathbb{P}(\xi_1^{(1)} = 0) > 0$, this yields, for $x \le k_0$,

$$P(x,y) \ge \varepsilon \mu(y) ,$$

with $\varepsilon = \{\mathbb{P}(\xi_1^{(1)} = 0)\}^{k_0}$. Thus C is a $(1,\varepsilon)$-small set.

Note also that the INAR process is stochastically monotone, since given $X_0 = x_0$ and $x > x_0$,

$$X_1 = \sum_{j=1}^{x_0} \xi_j^{(1)} + Y_1 \le \sum_{j=1}^{x} \xi_j^{(1)} + Y_1 \;\; \mathbb{P}-\text{a.s.}$$

15.11 Suppose that P is geometrically ergodic, and is therefore π-irreducible. We obtain a contradiction by showing that the conditions for Example 15.1.7 hold here. Specifically, suppose that for some arbitrary $\varepsilon > 0$, x is such that $h_\pi(x) \ge \varepsilon^{-2}$. (We know that $\pi(\{x : h_\pi(x) \ge \varepsilon^{-2}\}) > 0$.) Now set $A = \{y : h_\pi(y) \ge \varepsilon^{-1}\}$, and note that since h_π is a probability density function, $\text{Leb}(A) \le \varepsilon$, where μ_d^{Leb} denotes d-dimensional Lebesgue measure. Setting $M = \sup_z q(z)$, we have the bound

$$\begin{aligned} P(x, \{x\}^c) &= \int_{\mathbb{R}^d} q(x, y)(1 \wedge \frac{\pi(y)}{\pi(x)})\mathrm{d}y \\ &= \int_A q(x, y)(1 \wedge \frac{\pi(y)}{\pi(x)})\mathrm{d}y + \int_{A^c} q(x, y)(1 \wedge \frac{\pi(y)}{\pi(x)})\mathrm{d}y \\ &\le \int_A q(x, y)\mathrm{d}y + \int_{A^c} \frac{\pi(y)}{\pi(x)} q(x, y)\mathrm{d}y \\ &\le \varepsilon M + \varepsilon = (M+1)\varepsilon . \end{aligned}$$

Since ε is arbitrary, it follows that $\operatorname{esssup}_\pi (P(x,\{x\})) = 1$, so that by Example 15.1.7, geometric ergodicity fails.

15.12 By definition, for $k \ge 1$, we have

$$X_k \le \beta - \gamma \quad \text{if} \quad \gamma < 0 , \tag{G.27}$$
$$X_k \ge \beta - \gamma \quad \text{if} \quad \gamma \ge 0 . \tag{G.28}$$

We consider separately two cases.

Case $\gamma < 0$.

In this case, (G.27) shows that the state space is $(-\infty, \beta - \gamma]$. Let B be a Borel set such that $\beta - \gamma \in B$. Then for all $x \le \beta - \gamma$,

$$\begin{aligned} P(x,B) &= \mathbb{P}(X_1 \in B \,|\, X_0 = x) \ge \mathbb{P}(X_1 = \beta - \gamma \,|\, X_0 = x) \\ &= \mathbb{P}(N_1 = 0 \,|\, X_0 = x) = \mathrm{e}^{-\mathrm{e}^x} \ge \mathrm{e}^{-\mathrm{e}^{\beta-\gamma}} = \mathrm{e}^{-\mathrm{e}^{\beta-\gamma}} \delta_{\beta-\gamma}(B) . \end{aligned}$$

Thus the state space is 1-small.

Case $\gamma > 0$.

Now (G.28) shows that the state space is $[\beta-\gamma,\ \infty)$. Let $C = [\beta-\gamma,\ \beta+\gamma]$. Then for $x \in C$ and every Borel set B containing $\beta-\gamma$,

$$\begin{aligned}
P(x,B) &= \mathbb{P}(X_1 \in B \,|\, X_0 = x) \ \geq \mathbb{P}(X_1 \ = \beta-\gamma \,|\, X_0 = x) \\
&= \mathbb{P}(N_1 = 0 \,|\, X_0 = x) \ = \mathrm{e}^{-\mathrm{e}^{x}} \geq \mathrm{e}^{-\mathrm{e}^{\beta+\gamma}} \\
P^2(x,B) &\geq \mathbb{P}(X_{t+1} = \beta-\gamma,\ X_t = \beta-\gamma \,|\, X_{k-1} = x) \geq \mathrm{e}^{-2\mathrm{e}^{\beta+\gamma}} .
\end{aligned} \tag{G.29}$$

On the other hand, if $x > \beta+\gamma$, then noting that $\mathbb{E}[X_1 \,|\, X_0 = x] = \beta$, we have

$$\begin{aligned}
P(x,C) &= \mathbb{P}(\beta-\gamma \leq X_1 \leq \beta+\gamma \,|\, X_0 = x) \\
&= \mathbb{P}(|X_1-\beta| \leq \gamma \,|\, X_0 = x) \\
&\geq 1-\gamma^{-2}\mathrm{Var}(X_1 \,|\, X_0 = x) \\
&= 1-\gamma^{-2}\gamma^2\mathrm{e}^{-x} \geq 1-\mathrm{e}^{-(\beta+\gamma)} .
\end{aligned}$$

Then for $x \geq \beta+\gamma$, we have

$$\begin{aligned}
P^2(x,B) &= \mathbb{P}(X_2 \in B \,|\, X_0 = x) \geq \mathbb{P}(X_2 \in B,\ X_1 \in C \,|\, X_0 = x) \\
&= \mathbb{E}[\mathbb{1}_B(X_2)\mathbb{1}_C(X_1) | X_0 = x] \\
&= \mathbb{E}[\mathbb{1}_C(X_1)P(X_1,B) | X_0 = x] \\
&\geq \mathrm{e}^{-\mathrm{e}^{\beta+\gamma}} P(x,C) \geq \mathrm{e}^{-\mathrm{e}^{\beta+\gamma}}(1-\mathrm{e}^{-(\beta+\gamma)}) \\
&= \mathrm{e}^{-\mathrm{e}^{\beta+\gamma}}(1-\mathrm{e}^{-(\beta+\gamma)})\delta_{\beta-\gamma}(B) .
\end{aligned} \tag{G.30}$$

Finally, (G.29) and (G.30) show that the state space is 2-small.

Solutions to Exercises of Chapter 16

16.1 1. For $x \notin C$, we get

$$PW(x) = \mathbb{E}_x[|h(x)+Z_1|] \leq |h(x)|+m \leq |x|-(\ell-m) .$$

2. For $x \in C$, we similarly obtain

$$\begin{aligned}
PW(x) &\leq |h(x)|+m \leq |x|+|h(x)|-|x|+m \\
&\leq |x|-(\ell-m)+\sup_{|x|\leq M}\{|h(x)|-|x|+\ell\} .
\end{aligned}$$

3. Setting $V(x) = W(x)/(\ell - m)$ and $b = (\ell - m)^{-1} \sup_{|x| \leq M} \{|h(x)| - |x| + \ell\}$, we get
$$PV(x) \leq V(x) - 1 + b\mathbb{1}_C(x) .$$

16.2 We essentially repeat the arguments of the proof of Proposition 16.1.4. It suffices to consider the case $r \in \mathscr{S}$.

1. Exercise 4.11 (with $g \equiv 0$ and $h \equiv f$) implies that
$$PW_{1,C}^{f,r} + r(0)f = \sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} r(k) f(X_k) \right] . \tag{G.31}$$
Hence we have
$$PW_{1,C}^{f,r} + r(0)f = W_{0,C}^{f,r} + b\mathbb{1}_C ,$$
where $b = \sup_{x \in C} \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} r(k) f(X_k) \right]$. This proves that $W_{0,C}^{f,r}(x) < \infty$ implies $PW_{1,C}^{f,r}(x) < \infty$. Moreover, $W_{0,C}^{f,r} \leq W_{1,C}^{f,r}$, because r is nondecreasing and therefore $W_{0,C}^{f,r}(x) = \infty$ implies $W_{1,C}^{f,r}(x) = \infty$. Hence by Proposition 9.2.13, the set $\{W_{0,C}^{f,r} < \infty\}$ is full and absorbing, and $\{W_{0,C}^{f,r} \leq d\}$ is accessible for all sufficiently large d.
2. We can write $\{W_{0,C}^{f,r} \leq d\} = C \cup C_d$ with
$$C_d = \left\{ x \in \mathsf{X} : \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} r(k) f(X_k) \right] \leq d \right\} .$$
By Proposition 9.4.5, the union of two petite sets is petite, and thus it suffices to show that the set C_d is petite. This follows from Lemma 9.4.8, since if $x \in C_d$,
$$\mathbb{E}_x[r(\sigma_C)] \leq r(1)\mathbb{E}_x[r(\sigma_C - 1)] \leq r(1)\mathbb{E}_x \left[\sum_{k=0}^{\sigma_C - 1} r(k) f(X_k) \right] \leq r(1)d .$$

16.3 Since ψ is concave and continuously differentiable on $[v_0, \infty)$, we have for all $v \in [v_0, \infty)$, $\psi(v_0) \leq \psi(v) + \psi'(v)(v_0 - v)$. Hence $\psi(v_0) - \psi'(v)v_0 \leq \psi(v) - v\psi'(v)$, and since $\lim_{v \to \infty} \psi'(v) = 0$ and $\psi(v_0) \geq 1$, we may choose v_1 large enough that $\psi(v_1) - v_1\psi'(v_1) > 0$. It is easily seen that $\phi(1) = 1$, $\phi(v_1) = \psi(v_1)$, and $\phi'(v) \geq \phi'(v_1) = \psi'(v_1) \geq 0$ for all $v \in [1, v_1]$. Since $\psi(v_1) - (v_1 - 1)\psi'(v_1) \geq 0$, the function ϕ is concave on $[1, \infty)$.

16.4 1. Then for $v \geq 1$ and $t \geq 0$,
$$H_\phi(v) = \int_1^v \frac{\mathrm{d}s}{s^\alpha} = \frac{v^{1-\alpha} - 1}{1 - \alpha} , \qquad H_\phi^{-1}(t) = (1 + (1-\alpha)t)^{1/(1-\alpha)} .$$
Then $r_\phi(t) = (1-\alpha)^{-1}(1 + (1-\alpha)t)^\delta$ with $\delta = \alpha/(1-\alpha)$.

2. Differentiating ϕ_0 twice, we obtain

$$\phi_0''(v) = -\delta v^{-1} \log^{-\delta-2}(v) \left(\log v - \delta - 1\right) ,$$

which is negative for $\log v \geq \delta + 1$. This proves the first claim with $v_0 = \exp(\delta + 1)$. Now,

$$\begin{aligned} H_\phi(v) &= \int_1^v \frac{1}{\phi_0(u+v_0)} \mathrm{d}u = \int_{1+v_0}^{v+v_0} \frac{1}{\phi_0(u)} \mathrm{d}u \\ &= (\delta+1)^{-1} \left(\log^{\delta+1}(v+v_0) - \log^{\delta+1}(1+v_0) \right) . \end{aligned}$$

Then by straightforward algebra,

$$r_\phi(t) = \left(H_\phi^{-1}\right)'(t) = (\delta+1)(\alpha t+\beta)^{-\delta/(\delta+1)} \exp\left\{ (\alpha t+\beta)^{1/(\delta+1)} \right\} ,$$

where $\alpha = \delta + 1$ and $\beta = \log^{1+\delta}(v_0 + 1)$.

16.5 1. To obtain the polynomial rate $r(t) = (1+ct)^\gamma$, $c, \gamma > 0$, choose

$$\phi(v) = \{1 + c(1+\gamma)(v-1)\}^{\gamma/(1+\gamma)} , \; v \geq 1 .$$

2. To obtain the subexponential rate $r(t) = (1+t)^{\beta-1} \mathrm{e}^{c\{(1+t)^\beta - 1\}}$, $\beta \in (0,1)$, $c > 0$, choose

$$\phi(v) = \frac{v}{\{1 + c^{-1} \log(c\beta v)\}^{(1-\beta)/\beta}} .$$

The rate r is log-concave for large enough t (for all $t \geq 0$ if $c\beta \geq 1$), and the function ϕ is concave for v large enough (for all $v \geq 1$ if $c\beta \geq 1$).

16.6 By irreducibility, for all $x, z \in \mathsf{X}$, $\mathbb{P}_x(\sigma_z < \infty) > 0$, and thus there exists $q \in \mathbb{N}^*$ such that $\mathbb{P}_x(X_q = z) > 0$. Applying Theorem 16.2.3 with $A = \{x\}$, $B = \{z\}$, $f \equiv 1$, and $r(n) = n^{s \vee 1}$, we obtain that $\mathbb{E}_x[\sigma_z^{s \vee 1}] < \infty$ for all $z \in \mathsf{X}$. By Corollary 9.2.14, the set

$$S_x := \left\{ y \in \mathsf{X} : \mathbb{E}_y[\sigma_x^{s \vee 1}] < \infty \right\}$$

is full and absorbing. Therefore, $\pi(S_x) = 1$, and thus $S_x = \mathsf{X}$ by irreducibility. For $y, z \in \mathsf{X}$, we have $\sigma_z \leq \sigma_x + \sigma_z \circ \theta_{\sigma_x}$, and thus applying the strong Markov property, we obtain

$$\begin{aligned} \mathbb{E}_y[\sigma_z^{s \vee 1}] &\leq 2^{(s-1)^+} \mathbb{E}_y[\sigma_x^{s \vee 1}] + 2^{(s-1)^+} \mathbb{E}_y[\sigma_z \circ \theta_{\sigma_x}^{s \vee 1}] \\ &= 2^{(s-1)^+} \mathbb{E}_y[\sigma_x^{s \vee 1}] + 2^{(s-1)^+} \mathbb{E}_x[\sigma_z^{s \vee 1}] < \infty . \end{aligned}$$

The last statement is then a consequence of Exercise 13.1.

Solutions to Exercises of Chapter 18

18.2 1. for all $x \neq x' \in \{1,2,3\}$, we have

$$d_{TV}(P(x,\cdot),P(x',\cdot)) = \frac{1}{2}\sum_{y\in\mathsf{X}} |P(x,y) - P(x',y)| = \frac{1}{2}\,.$$

The state space X is not 1-small. The only measure μ for which $P(x,\{y\}) \geq \mu(\{y\})$ for all $x,y \in \mathsf{X}$ is the zero measure.

2. Applying Lemma 18.2.7 to X, we get that $\Delta(P) = 1/2$, and Theorem 18.2.4 implies $\sup_{x\in\mathsf{X}} d_{TV}(P^n(x,\ \cdot),\pi) \leq (1/2)^n$, for all $n \in \mathbb{N}$.
3. For $m = 2$,

$$P^2 = \begin{pmatrix} 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}.$$

Then $P^2(x,\cdot) \geq (3/4)\nu$ holds with $\varepsilon = 3/4$ and $\Delta(P^2) \leq 1/4$. Theorem 18.2.4 yields that

$$d_{TV}(P^n(x,\ \cdot),\pi) \leq (1/4)^{\lfloor n/2\rfloor} = \begin{cases} (1/2)^n & \text{if } n \text{ is even,} \\ (1/2)^{n-1} & \text{if } n \text{ is odd.} \end{cases}$$

This second bound is essentially the same as the first one, and both are (nearly) optimal, since the modulus of the second-largest eigenvalue of the matrix P is equal to $1/2$.

18.3 1. For $x,x' \in \mathsf{X}$ and $A \in \mathscr{X}$, define $Q(x,x',A) = \int_A p_m(x,y) \wedge p_m(x',y)\mu(\mathrm{d}y)$. Then $P^m(x,\cdot) - Q(x,x',\cdot)$ and $P^m(x,\cdot) - Q(x,x',\cdot)$ are measures on $\mathscr{X}$. Thus

$$\begin{aligned} &d_{TV}(P^m(x,\cdot),P^m(x',\cdot)) \\ &= \frac{1}{2}|P^m(x,\cdot) - Q(x,x',\cdot) - \{P^m(x',\cdot) - Q(x,x',\cdot)\}|(\mathsf{X}) \\ &\leq \frac{1}{2}|P^m(x,\cdot) - Q(x,x',\cdot)|(\mathsf{X}) + \frac{1}{2}|P^m(x',\cdot) - Q(x,x',\cdot)|(\mathsf{X}) \\ &\leq 1 - \int_{\mathsf{X}} p_m(x,y) \wedge p_m(x',y)\mu(\mathrm{d}y) \leq 1 - \varepsilon\,. \end{aligned}$$

2. Define the measure ν on $\mathscr{X}$ by $\nu(A) = \hat{\varepsilon}^{-1}\int_A g_m(y)\mu(\mathrm{d}y)$. Then for every $x \in C$ and $A \in \mathscr{X}$, $P^m(x,A) \geq \int_A p_m(x,y)\mu(\mathrm{d}y) \geq \hat{\varepsilon}\nu(A)$, showing that C is a small set.

18.4 1. Given X_n, the random variable Y_{n+1} is drawn according to $\mathrm{Unif}(0,\pi(X_n))$, and then X_{n+1} is drawn according to a density proportional to $x \mapsto \mathbb{1}\{\pi(x) \geq Y_{n+1}\}$.

2. Set $M := \sup_{x\in\mathsf{X}} \pi(x)$. By combining (18.6.3) with Fubini's theorem, we obtain for all $x \in \mathsf{X}$,

$$
\begin{aligned}
P(x,B) &= \int_B \left(\frac{1}{\pi(x)} \int_0^{\pi(x)} \frac{\mathbb{1}\{\pi(x') \geq y\}}{\operatorname{Leb}(L(y))} \mathrm{d}y \right) \mathrm{d}x' \\
&\geq \frac{1}{\operatorname{Leb}(\mathscr{S}_\pi)} \int_B \frac{\pi(x) \wedge \pi(x')}{\pi(x)} \mathrm{d}x' \\
&\geq \frac{1}{\operatorname{Leb}(\mathscr{S}_\pi)} \int_B 1 \wedge \frac{\pi(x')}{M} \mathrm{d}x' .
\end{aligned}
$$

Thus the whole state space X is small, and the kernel P is seen to be uniformly ergodic by applying Theorem 15.3.1.

18.5 Pick d sufficiently large that the set $C = \{x \in \mathsf{X} : M(x) \leq d\}$ is nonempty. Fix $\eta > 0$. Then for all sufficiently large m and $x, x' \in C$, $\mathrm{d_{TV}}(P^m(x,\cdot),\pi) \leq \eta$ and $\mathrm{d_{TV}}(P^m(x',\cdot),\pi) \leq \eta$, so that $\mathrm{d_{TV}}(P^m(x,\cdot),P^m(x',\cdot)) \leq 2\eta$. Thus C is an $(m, 1-2\eta)$-Doeblin set.

18.6 The Markov kernel P can be expressed as follows: for all $(x,A) \in \mathbb{R} \times \mathscr{B}(\mathbb{R})$,

$$
P(x,A) = r(x)\mathbb{E}\left[\mathbb{1}_A(x+Z_0)\right] + (1-r(x))\mathbb{E}\left[\mathbb{1}_A(Z_0)\right] . \tag{G.32}
$$

1. Indeed, for all $x \in \{r \leq 1-\varepsilon\}$, we have

$$
P(x,A) \geq (1-r(x))\mathbb{E}\left[\mathbb{1}_A(Z_0)\right] \geq \varepsilon \nu_Z(A) .
$$

2. Since $\sup_{x \in \mathbb{R}} r(x) < 1$, the whole state space X is small, and P is therefore uniformly ergodic.
3. Define the function φ on $[0,t]$ by $\varphi(s) = \mathbb{E}\left[\exp(sZ_0)\right]$. Since $\mathbb{E}\left[\exp(tZ_0)\right] < \infty$, the function φ is differentiable at $s=0$ and $\varphi'(0) = \mathbb{E}\left[Z_0\right] < 0$. Thus there exists $s \in (0,t)$ such that $\varphi(s) < \varphi(0) = 1$. This particular value of s being chosen, set $V_s(x) = 1 + \exp(sx)$ and $\lambda = \varphi(s) < 1$. With P defined in (G.32), we get

$$
\begin{aligned}
PV_s(x) &= r(x)\{\exp(sx)\mathbb{E}\left[\exp(sZ_0)\right]\} + (1-r(x))\mathbb{E}\left[\exp(sZ_0)\right] + 1 \\
&\leq \lambda \exp(sx) + \lambda + 1 \leq \lambda V_s(x) + 1 .
\end{aligned}
$$

Since for all $M > 1$, the level set $\{V_s \leq M\} = \{x \in \mathbb{R} : x \leq \log(M-1)/s\}$ is a subset of $\{r \leq r(\log(M-1)/s)\}$ that is small, Theorem 18.4.3 shows that the Markov kernel P is V_s-geometrically ergodic.

Solutions to Exercises of Chapter 19

19.1 1. $X' = X + Y$ follows a Poisson distribution with parameter $\alpha + (\beta - \alpha) = \beta$, and thus (X, X') is a coupling of (ξ, ξ').
2. We have $\mathbb{P}(X \neq X') = \mathbb{P}(Y \neq 0) = 1 - \exp(\beta - \alpha)$. Applying Theorem 19.1.6 yields $\mathrm{d_{TV}}(\xi, \xi') \leq 1 - \exp(\beta - \alpha)$. This coupling is not optimal, since

$$\int_{\{x=x'=0\}} \xi \wedge \xi'(\mathrm{d}x)\delta_x(\mathrm{d}x') = \mathrm{e}^{-\alpha} ,$$
$$\mathbb{P}(X = X' = 0) = \mathbb{P}(X = 0, Y = 0) = \mathrm{e}^{-\alpha}\mathrm{e}^{-\beta+\alpha} = \mathrm{e}^{-\beta} \neq \mathrm{e}^{-\alpha} .$$

These two quantities should be equal for an optimal coupling by (19.1.4) applied to $B = \{X = X' = 0\}$.

19.2 Draw independently a Bernoulli random variable U with probability of success $1-\varepsilon$, $Y \sim \mathrm{Unif}([0,\varepsilon])$, $Y' \sim \mathrm{Unif}([1,1+\varepsilon])$, and $Z \sim \mathrm{Unif}([1-\varepsilon,1])$, and set

$$(X,X') = \begin{cases} (Y,Y') & \text{if } U = 0 , \\ (Z,Z) & \text{otherwise.} \end{cases}$$

Then (X,X') is an optimal coupling of (ξ,ξ').

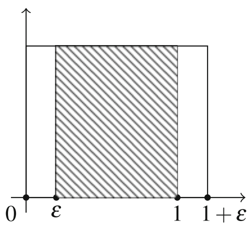

Fig. G.0.3 An example of optimal coupling.

19.3 For $A \in \mathscr{X}$, we have

$$\mathbb{P}(X \in A) = (1-\varepsilon)\eta(A) + \xi \wedge \xi'(A) = \xi(A) .$$

Similarly, $\mathbb{P}(X' \in A) = \xi'(A)$. Thus (X,X') is a coupling of (ξ,ξ'). Since η and η' are mutually singular, Lemma 19.1.5 yields $\mathbb{P}(Y = Y') = 0$. Thus on applying Lemma 19.1.1, we obtain

$$\mathbb{P}(X = X') = (1-\varepsilon)\mathbb{P}(Y = Y') + \varepsilon = \varepsilon = 1 - \mathrm{d}_{\mathrm{TV}}(\xi,\xi') .$$

Thus $\mathbb{P}(X \neq X') = \mathrm{d}_{\mathrm{TV}}(\xi,\xi')$, and (X,X) is an optimal coupling of (ξ,ξ').

19.4 Let (X,Y) be a coupling of (ξ,ξ') defined on a probability space $(\Omega,\mathscr{F},\mathbb{P})$. Applying Hölder's inequality and Minkowski's inequality, we obtain

$$\begin{aligned}\left|\int f\mathrm{d}\xi - \int f\mathrm{d}\xi'\right| &= |\mathbb{E}\left[(f(X)-f(Y))\mathbb{1}(X\neq Y)\right]| \\ &\leq \|f(X)-f(Y)\|_{\mathrm{L}^p(\mathbb{P})}\left\{\mathbb{P}(X\neq Y)\right\}^{1/q} \\ &\leq (\|f(X)\|_{\mathrm{L}^p(\mathbb{P})}+\|f(Y)\|_{\mathrm{L}^p(\mathbb{P})})\{\mathbb{P}(X\neq Y)\}^{1/q} \\ &= (\|f\|_{\mathrm{L}^p(\xi)}+\|f\|_{\mathrm{L}^p(\xi)})\{\mathbb{P}(X\neq Y)\}^{1/q}.\end{aligned}$$

Taking the infimum over all the coupling (X,Y) of ξ and ξ' yields the desired bound by Theorem 19.1.6.

19.5 By Theorem 19.1.12, if $\Delta(P)\leq 1-\varepsilon$, the optimal kernel coupling defined in (19.1.15) satisfies (19.6.1). Conversely, if (19.6.1) holds, then Theorem 19.1.6 yields

$$\mathrm{d}_{\mathrm{TV}}(P(x,\cdot),P(x',\cdot))\leq K(x,x';\Delta^c)\leq 1-\varepsilon\ .$$

By Lemma 18.2.2, this yields $\Delta(P)\leq 1-\varepsilon$.

19.6 Run two copies of the chain, one from an initial distribution concentrated at x and the other from the initial (invariant) distribution π (π exists, since the Markov kernel P is uniformly ergodic). At every time instant, do the following:

(i) With probability ε, choose for both chains the same next position from the distribution ν, after which they will be coupled and then can be run with identical sample paths.
(ii) With probability $1-\varepsilon$, draw independently for each chain an independent position from the distribution

$$R(x,\cdot)=\{P(x,\cdot)-\varepsilon\nu\}/(1-\varepsilon).$$

This Markov kernel is well defined, since for all $x\in\mathsf{X}$, $P(x,\cdot)\geq\varepsilon\nu$. The marginal distributions of these chains are identical to the original distributions for every n (this is a special instance of independent and then forever coupling described in Example 19.1.15). If we let T deonte the coupling time (see (19.2.1)), then using the coupling inequality, Theorem 19.2.1, we have $\mathrm{d}_{\mathrm{TV}}(P^n(x,\dot{)},\pi)\leq(1-\varepsilon)^n$ (at each time step, the coupling is successful with probability ε).

19.7 1. The state space is $\{0,\ldots,N\}$ and transition matrix P is given by

$$P(i,i+1)=\frac{N-i}{2N}\ ,\ \ i=0,\ldots,N-1\ ,$$
$$P(i,i)=\frac{1}{2}\ ,\ \ i=0,\ldots,N\ ,$$
$$P(i,i-1)=\frac{i}{2N}\ ,\ \ i=1,\ldots,N\ .$$

2. Consider the independent coupling, that is, the kernel K defined by $K(x,x';\cdot)=P(x,\cdot)P(x',\cdot)$. Let $\{(X_n,X_n'),\ n\in\mathbb{N}\}$ be the canonical chain with kernel K. Then

starting from any $x,x' \in \{0,\dots,N\}$, there is a positive probability of coupling in fewer than $N/2$ steps if the chains move toward each other at each step. More precisely,

$$\mathbb{P}_{x,x'}(X_{N/2} = X'_{N/2}) \geq 2^{-N} \prod_{i=0}^{N/2-1} \frac{N-i}{N} \prod_{i=N/2+1}^{N} \frac{i}{N} = \frac{(N!)^2}{2^N N^N((N/2)!)^2} .$$

The result follows from Exercise 19.5.

19.8 1. The kernel K of this chain is given by

$$K(x,x';x_i^{\varepsilon},x_{d-i+1}^{\varepsilon'}) = \frac{1}{d}\pi_{i,x}(\varepsilon)\pi_{d-i+1,\varepsilon'}(x') ,$$

and $K(x,x';z,z') = 0$ for other values z,z'. It is readily checked that K is a coupling kernel of (P,P).

2. • After $d/2$ moves, all the sites may have been updated by either chain. This happens only if each site i or $d-i+1$ was updated only once by each chain. Since each site is chosen at random, this event has probability

$$p_d = \frac{d(d-2)\cdots 2}{(d/2)!d^{d/2}} = \frac{(d/2)!2^{d/2}}{d^{d/2}} .$$

• At each move, two different coordinates are updated. They are made (or remain) equal with probability equal to at least $\{M/(1+M)\}^2$.

This implies that $\mathbb{P}_{x,x'}(X_{d/2} = X'_{d/2}) \geq \varepsilon$ with

$$\varepsilon \geq \frac{M^d}{(1+M)^d} \frac{(d/2)!2^{d/2}}{d^{d/2}} .$$

3. If all the coordinates of x and x' are distinct, then $P^m(x,x') = 0$ if $m < d$. It is impossible for the chain to update all its components and hence move to an arbitrary state.
4. For $m = d$, the probability that the sites $I_1,\dots,I_d$ that are updated during the first d moves are distinct is $d!d^{-d}$. Given that they are, the probability of hitting a given site $x' \in \mathsf{X}$ after d steps starting from x is greater than $M^{d-1}\pi(x')$. Therefore, choosing $\tilde{\varepsilon} \in \left[M^d d!d^{-d}, M^{-d}d!d^{-d}\right]$, we obtain $P^d(x,x') \geq \tilde{\varepsilon}\pi(x')$.

Assume that π is uniform. In the case that π is uniform ($M = 1$), Stirling's formula gives, for large d, $\varepsilon \sim (\pi d)^{1/2}\mathrm{e}^{-d/2}$ and $\tilde{\varepsilon} = d!d^{-d} \sim (2\pi d)^{1/2}\mathrm{e}^{-d}$, and thus $\varepsilon > \tilde{\varepsilon}$ for sufficiently large d.

19.9 Set $C = (-\infty, x_0]$ and $\bar{C} = C \times C$. Since C is a Doeblin set, the optimal kernel coupling K described in Example 19.1.16 satisfies $K(x,x';\Delta) \geq \varepsilon$. Let us check that $\bar{V}$ satisfies the drift condition (19.4.3). Since V is increasing and since K preserves the order, if $x \preceq x' \in \mathsf{X}$, we have

$$
\begin{aligned}
K\bar{V}(x,x') &= \mathbb{E}_{x,x'}[V(X_1 \vee X_1')] = \mathbb{E}_{x,x'}[V(X_1')] \\
&= PV(x') \le \lambda V(x') + b = \lambda \bar{V}(x,x') + b \,.
\end{aligned}
$$

If $(x,x') \notin C \times C$, then necessarily $x_0 \preceq x'$ and $V(x') \ge V(x_0)$. Thus

$$
K\bar{V}(x,x') \le (\lambda + b/V(x_0))V(x') = \bar{\lambda}\bar{V}(x,x') \,.
$$

If $(x,x') \in C \times C$, then $V(x') \le V(x_0)$ and $K\bar{V}(x,x') \le \lambda V(x_0) + b = \bar{b}$. Thus (19.4.3) holds. We can apply Lemma 19.4.2, and (19.6.2) is obtained by integration of (19.4.4b).

Solutions to Exercises of Chapter 20

20.2 Let $\gamma_i \in \mathscr{C}(\xi_i, \xi_i')$, $i = 1,2$. Then $\alpha\gamma_1 + (1-\alpha)\gamma_2$ is a coupling of $(\alpha\xi_1 + (1-\alpha)\xi_2, \alpha\xi_1' + (1-\alpha)\xi_2')$. Thus

$$
\begin{aligned}
&\mathbf{W}^p_{\mathsf{d},p}\left(\alpha\xi_1 + (1-\alpha)\xi_2, \alpha\xi_1' + (1-\alpha)\xi_2'\right) \\
&\qquad \le \alpha \int_{\mathsf{X}\times\mathsf{X}} \mathsf{d}^p(x,y)\gamma_1(\mathrm{d}x\mathrm{d}y) + (1-\alpha)\int_{\mathsf{X}\times\mathsf{X}} \mathsf{d}^p(x,y)\gamma_2(\mathrm{d}x\mathrm{d}y) \,.
\end{aligned}
$$

The result follows by taking the infimum over γ_1 and γ_2.

20.3 1. Note first that by the triangle inequality, we have

$$
|\mathbf{W}_{\mathsf{d},p}(\mu_n, \nu_n) - \mathbf{W}_{\mathsf{d},p}(\mu, \nu)| \le \mathbf{W}_{\mathsf{d},p}(\mu_n, \mu) + \mathbf{W}_{\mathsf{d},p}(\nu_n, \nu) \,.
$$

Thus $\lim_{n\to\infty} \mathbf{W}_{\mathsf{d},p}(\mu_n, \nu_n) = \mathbf{W}_{\mathsf{d},p}(\mu, \nu)$. The sequence $\{\gamma_n\}$ is tight (cf. the proof of Theorem 20.1.1).

2. Let γ be a weak limit along a subsequence $\{\gamma_{n_k}\}$. Then for $M > 0$,

$$
\begin{aligned}
\int_{\mathsf{X}\times\mathsf{X}} (\mathsf{d}^p(x,y) \wedge M)\gamma(\mathrm{d}x\mathrm{d}y) &= \lim_{k\to\infty} \int_{\mathsf{X}\times\mathsf{X}} (\mathsf{d}^p(x,y) \wedge M)\gamma_{n_k}(\mathrm{d}x\mathrm{d}y) \\
&\le \limsup_{k\to\infty} \int_{\mathsf{X}\times\mathsf{X}} \mathsf{d}^p(x,y)\gamma_{n_k}(\mathrm{d}x\mathrm{d}y) \\
&= \limsup_{k\to\infty} \mathbf{W}^p_{\mathsf{d},p}\left(\mu_{n_k}, \nu_{n_k}\right) = \mathbf{W}^p_{\mathsf{d},p}(\mu, \nu) \,.
\end{aligned}
$$

Letting M tend to ∞, this yields $\int_{\mathsf{X}\times\mathsf{X}} \mathsf{d}^p(x,y)\gamma(\mathrm{d}x\mathrm{d}y) \le \mathbf{W}^p_{\mathsf{d},p}(\mu,\nu)$, which implies that γ is an optimal coupling of μ and ν.

20.4 1. As in the finite-dimensional case, a simple recurrence shows that $\mathbb{X}_n$ can be expressed as $\mathbb{X}_n = \boldsymbol{\Phi}^n\mathbb{X}_0 + \sum_{k=1}^n \boldsymbol{\Phi}^{n-k}\mathbb{Z}_k$. To prove the second identity, it suffices to note that $\theta^n\boldsymbol{\Phi}^n\mathbb{X}_0 = \alpha^n\mathbb{X}_0$ and for $k \ge 1$, $\theta^n\boldsymbol{\Phi}^{n-k}\mathbb{Z}_k = 0$.

2. Therefore, for any coupling $(\mathbb{X}_n, \mathbb{X}_n')$ of $P^n(x,\cdot)$ and $P^n(x',\cdot)$, $\mathbb{P}(\mathbb{X}_n \ne \mathbb{X}_n') = 1$ if $x \ne x'$. By Theorem 19.1.6, this implies that $\mathrm{d}_{\mathrm{TV}}(P^n(x,\cdot), P(x',\cdot)) = 1$ if $x \ne x'$.

3. Consider now the Wasserstein distance $\mathbf{W}_{\mathsf{d},p}$ with respect to the distance $\mathsf{d}^2(u,v) = \sum_{n=0}^{\infty}(u_n - v_n)^2$. As in the finite-dimensional case, consider the simple coupling $(\mathbb{X}_n, \mathbb{X}'_n) = (\Phi^n x + \sum_{k=1}^n \Phi^{n-k}\mathbb{Z}_k, \Phi^n x' + \sum_{k=1}^n \Phi^{n-k}\mathbb{Z}_k)$. Then

$$\mathbf{W}^p_{\mathsf{d},p}\left(P(x,\cdot),P(x',\cdot)\right) \le \mathbb{E}\left[\mathsf{d}^p(\mathbb{X}_n,\mathbb{X}'_n)\right] = \mathbb{E}\left[\mathsf{d}^p(\Phi^n x, \Phi^n x')\right] = \alpha^n \mathsf{d}^p(x,x') .$$

Thus $\Delta_{\mathsf{d},p}(P) \le \alpha < 1$.

20.5 By Minkowski's inequality, we have, for $v \ge u \ge 0$,

$$\|F(u,N) - F(v,N)\|_1 \le |b||u-v| + |c| \left\| \log\left(\frac{1+N(\mathrm{e}^v)}{1+N(\mathrm{e}^u)}\right)\right\|_1 .$$

Applying (20.6.2) yields, for $x,y \ge 0$,

$$\|F(x,N) - F(y,N)\|_1 \le (|b|+|c|)|x-y| ,$$

and the contraction property (20.3.10) holds with $p=1$ if $|b|+|c|<1$.

We now prove (20.6.2). Since N has independent increments, we can write $(1+N(\mathrm{e}^y))/(1+N(\mathrm{e}^x)) = 1 + V/(1+W)$, where V and W are independent Poisson random variables with respective means $\mathrm{e}^y - \mathrm{e}^x$ and e^x. The function $t \mapsto \log(1+t)$ is concave, and thus by Jensen's inequality, we obtain

$$\begin{aligned}\mathbb{E}\left[\log\left(\frac{1+N(\mathrm{e}^y)}{1+N(\mathrm{e}^x)}\right)\right] &= \mathbb{E}\left[\log\left(1+\frac{V}{1+W}\right)\right] \\ &\le \log\left(1+\mathbb{E}\left[\frac{V}{1+W}\right]\right) = \log\left(1+(\mathrm{e}^y-\mathrm{e}^x)\mathbb{E}\left[\frac{1}{1+W}\right]\right) .\end{aligned}$$

The last expectation can be computed and bounded:

$$\mathbb{E}\left[\frac{1}{1+W}\right] = \mathrm{e}^{-\mathrm{e}^x}\sum_{k=0}^{\infty}\frac{1}{1+k}\frac{\mathrm{e}^{kx}}{k!} = \mathrm{e}^{-x}\mathrm{e}^{-\mathrm{e}^x}\sum_{k=1}^{\infty}\frac{\mathrm{e}^{kx}}{k!} \le \mathrm{e}^{-x} .$$

This yields

$$\mathbb{E}\left[\log\left(\frac{1+N(\mathrm{e}^y)}{1+N(\mathrm{e}^x)}\right)\right] \le \log\left(1+(\mathrm{e}^y-\mathrm{e}^x)\mathrm{e}^{-x}\right) = y - x .$$

This proves (20.6.2).

20.6 1. Assumption (b) implies that $|g(x)-g(y)| \le |x-y|$ for all $x,y \in \mathbb{R}^d$. Therefore,

$$\begin{aligned}\mathbf{W}_{\mathsf{d}}\left(P(x,\cdot),P(x',\cdot)\right) &\le \mathbb{E}\left[\mathsf{d}(g(x)+Z_0, g(x')+Z_0)\right] = |g(x)-g(x')| \\ &\le |x-x'| = \mathsf{d}(x,x') .\end{aligned}$$

2. Assumption (b) implies that

$$\mathbf{W}_{\mathsf{d}}\left(P(x,\cdot),P(x',\cdot)\right) \le \mathbb{E}\left[\mathsf{d}(g(x)+Z_0,g(x')+Z_0)\right]$$
$$= |g(x)-g(x')| \le (1-\varepsilon_K)\mathsf{d}(x,x') .$$

3. Fix $B>0$ such that $\mathbb{E}\left[\mathrm{e}^{a|Z_0|}\right]\mathrm{e}^{-aB}<1$ and set $\lambda=\mathbb{E}\left[\mathrm{e}^{a|Z_0|}\right]\mathrm{e}^{-aB}$. Condition (c) implies that there exists $A>0$ such that $|x|\ge A$ implies $|g(x)|\le|x|-B$. Thus for $|x|\ge A$, we obtain

$$PV(x) = \{a^{-1}\vee 1\}\mathbb{E}\left[\mathrm{e}^{a|g(x)+Z_0|}\right] \le \{a^{-1}\vee 1\}\mathbb{E}\left[\mathrm{e}^{a|Z_0|}\right]\mathrm{e}^{a|g(x)|} \le \lambda V(x) .$$

Since g is locally bounded, define $M=\sup_{|x|\le A}|g(x)|$. Then for $|x|\le A$,

$$PV(x) \le \{(2/a)\vee 1\}\mathrm{e}^{aM}\mathbb{E}\left[\mathrm{e}^{a|Z_0|}\right] .$$

Setting $b=\{(2/a)\vee 1\}\mathrm{e}^{aM}\mathbb{E}\left[\mathrm{e}^{a|Z_0|}\right]$ yields $PV\le\lambda V+b$.

4. Define $K=\{V\le 2(b+\delta)/(1-\lambda)\}$. Then

$$\bar{C}=\left\{(x,y)\in\mathbb{R}^d \,:\, V(x)+V(y)\le 2(b+\delta)/(1-\lambda)\right\}\subset K\times K .$$

Thus $\bar{C}$ is a subset of a $(\mathsf{d},1,\varepsilon)$-contracting set and hence is itself a $(\mathsf{d},1,\varepsilon)$-contracting set.

5. Since

$$|x-y|\le|x|+|y|\le\frac{1}{a}\{\mathrm{e}^{a|x|}+\mathrm{e}^{a|y|}\}\le V(x)+V(y) ,$$

we conclude by applying Theorem 20.4.5.

20.7 1. For all $x,x'\in\{0,1\}^N$, $\varepsilon\in\{0,1\}$, and $i\in\{1,\dots,N\}$,

$$\mathsf{d}(F(x,\varepsilon,i),F(x',\varepsilon,i))=\mathsf{d}(x\oplus\varepsilon e_i,x'\oplus\varepsilon e_i)=\mathsf{d}(x,x') .$$

2. Since B_1 is independent of I_1 and has the same distribution as $1-B_1$, we get

$$\begin{aligned}
\mathbb{E}\left[g(X_1')\right] &= \mathbb{E}\left[g\left(x'\oplus B_1e_{I_1}\mathbb{1}_{\{x_{I_1}=x'_{I_1}\}}+x'\oplus(1-B_1)e_{I_1}\mathbb{1}_{\{x_{I_1}\ne x'_{I_1}\}}\right)\right]\\
&= \mathbb{E}\left[g(x'\oplus B_1e_{I_1})\mathbb{1}_{\{x_{I_1}=x'_{I_1}\}}\right]+\mathbb{E}\left[g(x'\oplus(1-B_1)e_{I_1})\mathbb{1}_{\{x_{I_1}\ne x'_{I_1}\}}\right]\\
&= \mathbb{E}\left[g(x'\oplus B_1e_{I_1})\mathbb{1}_{\{x_{I_1}=x'_{I_1}\}}\right]+\mathbb{E}\left[g(x'\oplus B_1e_{I_1})\mathbb{1}_{\{x_{I_1}\ne x'_{I_1}\}}\right]\\
&= \mathbb{E}\left[g(x'\oplus B_1e_{I_1})\right]=Pg(x') .
\end{aligned}$$

3. Since I_1 is uniformly distributed on $\{1,\dots,N\}$, $\mathbb{P}(x_{I_1}=x'_{I_1})=1-\mathsf{d}(x,x')/N$. Thus

$$\begin{aligned}\mathbb{E}\left[\mathsf{d}(X_1,X_1')\right] &= \mathsf{d}(x,x')\mathbb{P}(x_{I_1}=x'_{I_1})+(\mathsf{d}(x,x')-1)\mathbb{P}(x_{I_1}\neq x'_{I_1})\\ &= \mathsf{d}(x,x')(1-\mathsf{d}(x,x')/N)+(\mathsf{d}(x,x')-1)\mathsf{d}(x,x')/N\\ &= \mathsf{d}(x,x')(1-1/N)\,.\end{aligned}$$

This yields, by definition of the Wasserstein distance,

$$\mathbf{W}_{\mathsf{d}}\left(P(x,\cdot),P(x',\cdot)\right)\le(1-1/N)\mathsf{d}(x,x')\,,$$

and this proves that $\Delta_d\,(P)\le 1-1/N$ by Lemma 20.3.2.

20.8 1. Applying (20.6.6), we obtain

$$\begin{aligned}\mathbb{E}\left[\mathsf{d}^p(X_n,X_n')\,\middle|\,\mathscr{F}_{n-1}\right] &= K\mathsf{d}^p(X_{n-1},X'_{n-1})\\ &\le \mathsf{d}^p(X_{n-1},X'_{n-1})\{1-\varepsilon\mathbb{1}_{\bar{C}}(X_{n-1},X'_{n-1})\}^p\\ &\le \mathsf{d}^p(X_{n-1},X'_{n-1})\,.\end{aligned}$$

Defining $Z_n=\mathsf{d}^p(X_n,X_n')$, this proves that $\{Z_n,\ n\in\mathbb{N}\}$ is a supermartingale.

2. Applying the strong Markov property yields

$$\begin{aligned}\mathbb{E}\left[Z_{\sigma_{m+1}}\,\middle|\,\mathscr{F}_{\sigma_m}\right] &= \mathbb{E}\left[\mathbb{E}\left[Z_{\sigma_{m+1}}\,\middle|\,\mathscr{F}_{\sigma_m+1}\right]\,\middle|\,\mathscr{F}_{\sigma_m}\right]\\ &\le \mathbb{E}\left[Z_{\sigma_m+1}\,|\,\mathscr{F}_{\sigma_m}\right]\le(1-\varepsilon)^pZ_{\sigma_m}\,.\end{aligned}$$

Inductively, this yields $\mathbb{E}_\gamma[Z_{\sigma_m}]\le(1-\varepsilon)^{pm}\mathbb{E}_\gamma[Z_0]$.

3. For $n\ge 0$, we obtain

$$\begin{aligned}\mathbb{E}_\gamma[Z_n] &= \mathbb{E}_\gamma[Z_n\mathbb{1}\{\sigma_m\le n\}]+\mathbb{E}_\gamma[Z_n\mathbb{1}\{\sigma_m>n\}]\\ &\le \mathbb{E}_\gamma[Z_{\sigma_m}]+\mathbb{E}_\gamma[Z_n\mathbb{1}\{\eta_{n-1}<m\}]\\ &\le (1-\varepsilon)^{pm}\mathbb{E}_\gamma[Z_0]+\mathbb{E}_\gamma[Z_n\mathbb{1}\{\eta_{n-1}<m\}]\,. \end{aligned}\tag{G.33}$$

20.9 1. Using straightforward computations, we get

$$\begin{aligned}K\bar{V}(x,y) &\le \lambda\bar{V}(x,y)+b\\ &= \{\lambda\bar{V}(x,y)+b\}\mathbb{1}_{\bar{C}^c}(x,y)+\{\lambda\bar{V}(x,y)+b\}\mathbb{1}_{\bar{C}}(x,y)\\ &\le \left\{\lambda+\frac{b(1-\lambda)}{b+\delta}\right\}\bar{V}(x,y)\mathbb{1}_{\bar{C}^c}(x,y)+\left\{b+\frac{(b+\delta)\lambda}{1-\lambda}\right\}\mathbb{1}_{\bar{C}}(x,y)\,.\end{aligned}$$

Set $\bar{\lambda}=\lambda+b(1-\lambda)/(b+\delta)<1$ and $\bar{b}=b+\lambda(b+\delta)/(1-\lambda)\ge 1$. This yields

$$K\bar{V}\le\bar{\lambda}\bar{V}\mathbb{1}_{\bar{C}^c}+\bar{b}\mathbb{1}_{\bar{C}}\,.$$

2. Using the relation $\eta_n=\eta_{n-1}+\mathbb{1}_{\bar{C}}(X_n,X_n')$ and $\bar{V}\ge 1$, we obtain

$$\begin{aligned}\mathbb{E}\left[S_{n+1} \,|\, \mathscr{F}_n\right] &= \bar{\lambda}^{-n-1+\eta_n} \bar{b}^{-\eta_n} K\bar{V}(X_n, X_n') \\ &\leq \bar{\lambda}^{-n-1+\eta_n} \bar{b}^{-\eta_n} \{\bar{\lambda}\bar{V}(X_n, X_n') \mathbb{1}_{\bar{C}^c}(X_n, X_n') + \bar{b}\mathbb{1}_{\bar{C}}(X_n, X_n')\} \\ &= \bar{\lambda}^{-n+\eta_{n-1}} \bar{b}^{-\eta_{n-1}} \{\bar{V}(X_n, X_n') \mathbb{1}_{\bar{C}^c}(X_n, X_n') + \mathbb{1}_{\bar{C}}(X_n, X_n')\} \leq S_n \,.\end{aligned}$$

Thus $\{S_n\}$ is a supermartingale.

3. By (20.6.4), we get that $Z_n \leq 2\bar{V}(X_n, X_n')$, which implies

$$\mathbb{E}_\gamma[Z_n \mathbb{1}_{\{\eta_{n-1}<m\}}] \leq 2\bar{\lambda}^{n-m}\bar{b}^m \mathbb{E}_\gamma[S_n \mathbb{1}\{\eta_{n-1} < m\}] \leq 2\bar{\lambda}^{n-m}\bar{b}^m \mathbb{E}_\gamma[S_0] \,.$$

4. Plugging this bound into (20.6.7) yields (20.6.9).

20.10 1. We use (20.6.9). Set $m = n\log\bar{\lambda}/\{\log\bar{\lambda} - \log\bar{b} + \log(1-\varepsilon)\}$ and

$$\log\tau = \frac{p\log(1-\varepsilon)\log\bar{\lambda}}{\log\bar{\lambda} - \log\bar{b} + \log(1-\varepsilon)} < 0 \,.$$

This yields, for all $x, x' \in \mathsf{X}\times\mathsf{X}$,

$$\mathbb{E}_{x,x'}[\mathsf{d}^p(X_n, X_n')] \leq 4\bar{V}(x,x')\tau^n \,.$$

If ξ, ξ' are probability measures on X such that $\xi(V) + \xi'(V) < \infty$, then integrating the previous bound with respect to $\gamma \in \mathscr{C}(\xi, \xi')$ yields (20.6.10).

2. Choosing $\mu = \delta_x$ and $\nu = P(x,\cdot)$ yields

$$\mathbf{W}_{\mathsf{d},p}\left(P^n(x,\cdot), P^{n+1}(x,\cdot)\right) \leq \{V(x) + PV(x)\}\tau^n \leq \{2V(x) + b\}\tau^n \,.$$

Since $(\mathbb{S}_p(\mathsf{X},\mathsf{d}), \mathbf{W}_{\mathsf{d},p})$ is a complete metric space, this proves that there exists a probability measure π such that $\mathbf{W}_{\mathsf{d},p}(P^n(x,\cdot),\pi) = O(\tau^n)$. Since P is weakly contracting for $\mathbf{W}_{\mathsf{d},p}$, this yields $\mathbf{W}_{\mathsf{d},p}\left(P^{n+1}(x,\cdot), \pi P\right) \leq \mathbf{W}_{\mathsf{d},p}(P^n(x,\cdot),\pi) = O(\tau^n)$, which implies that π is invariant. By Lemma 14.1.10, this implies that $\pi(V) < \infty$. Let π' be another invariant probability measure. Then we also have that $\pi'(V) < \infty$ and (20.6.10) yields

$$\mathbf{W}_{\mathsf{d},p}\left(\pi,\pi'\right) = \mathbf{W}_{\mathsf{d},p}\left(\pi P^n, \pi' P^n\right) \leq 2\tau^n\{\pi(V) + \pi'(V)\} \,.$$

This proves that $\mathbf{W}_{\mathsf{d},p}(\pi,\pi') = 0$, and the invariant probability measure is unique.

20.11 Since the sequence $\{\mathbf{W}_{\mathsf{d}_n},\, n \geq 1\}$ is nondecreasing and $\mathbf{W}_{\mathsf{d}_n} \leq \mathrm{d_{TV}}$, it follows that $\lim_{n\to\infty}\mathbf{W}_{\mathsf{d}_n}(\mu,\nu) \leq \mathrm{d_{TV}}(\mu,\nu)$. Fix an arbitrary $L_1 > \lim_{n\to\infty}\mathbf{W}_{\mathsf{d}_n}(\mu,\nu)$. We will prove that $\mathrm{d_{TV}}(\mu,\nu) \leq L_1$, which yields $\mathrm{d_{TV}}(\mu,\nu) \leq \lim_{n\to\infty}\mathbf{W}_{\mathsf{d}_n}(\mu,\nu)$. For all $n \in \mathbb{N}$, there exists $\gamma_n \in \mathscr{C}(\mu,\nu)$ such that $\int \mathsf{d}_n(x,y)\gamma_n(\mathrm{d}x,\mathrm{d}y) \leq L_1$. For all compact sets K_1, K_2, we have $\gamma_n((K_1\times K_2)^c) \leq \mu(K_1^c) + \nu(K_2^c)$. Since μ and ν are tight, this implies that the sequence of probabilities $\{\gamma_n\}$ is tight and hence relatively compact by Prokhorov's theorem (Theorem C.2.2). Therefore, there exists a subsequence $\{\gamma_{n_k}\}$ converging weakly to $\gamma_\infty \in \mathscr{C}(\mu,\nu)$. For a given $n \in \mathbb{N}$, for all k such that $n \leq n_k$, we have by assumption $\mathsf{d}_n \leq \mathsf{d}_{n_k}$, and hence

$$\int \mathsf{d}_n(x,y)\gamma_{n_k}(\mathrm{d}x\mathrm{d}y) \le \int \mathsf{d}_{n_k}(x,y)\gamma_{n_k}(\mathrm{d}x\mathrm{d}y) \le L_1 \; .$$

Since $\mathsf{d}_n \le 1$ for all n and we have assumed that the distances d_n are continuous (for the topology of X), we obtain for every $n \in \mathbb{N}$,

$$\int \mathsf{d}_n(x,y)\gamma_\infty(\mathrm{d}x\mathrm{d}y) = \lim_{k\to\infty} \int \mathsf{d}_n(x,y)\gamma_{n_k}(\mathrm{d}x\mathrm{d}y) \le L_1 \; .$$

By Theorem 19.1.6, we have $\mathsf{d}_{\mathrm{TV}}(\mu,\nu) \le \int \mathbb{1}\{x \ne y\}\,\gamma_\infty(\mathrm{d}x\mathrm{d}y)$. Since $\mathsf{d}_n(x,y)$ increases to $\mathbb{1}\{x \ne y\}$, the monotone convergence theorem yields

$$\mathsf{d}_{\mathrm{TV}}(\mu,\nu) \le \int \mathbb{1}\{x \ne y\}\,\gamma_\infty(\mathrm{d}x\mathrm{d}y) = \lim_{n\to\infty} \int \mathsf{d}_n(x,y)\gamma_\infty(\mathrm{d}x\mathrm{d}y) \le L_1 \; .$$

Since L_1 is arbitrary, this concludes the proof.

20.12 1. By Theorem 1.4.6, we may assume that μ and ν are mutually singular.

2. Let $A \in \mathscr{V}_{x^*}$. There exists an open set $O \subset A$ that is accessible and contains x^* (since x^* is assumed to be reachable). Since μ is invariant, this implies $\mu(A) \ge \mu(O) = \mu K_{a_\varepsilon}(O) > 0$. Thus $\mu(A) > 0$, and similarly $\nu(A) > 0$.
3. Because P is asymptotically ultra-Feller, for every $\varepsilon > 0$ there exists a set $A \in \mathscr{V}_{x^*}$ such that

$$\lim_{k\to\infty} \sup_{x\in A} \mathbf{W}_{\mathsf{d}_k}\left(P^{n_k}(x,\cdot), P^{n_k}(x^*,\cdot)\right) \le \varepsilon/2 \; .$$

This implies (20.6.12).
4. The probability measures μ and ν being invariant, (20.6.1) and Exercise 20.2 yield, for every distance d on X,

$$\begin{aligned}
\mathbf{W}_{\mathsf{d}}(\mu,\nu) &= \mathbf{W}_{\mathsf{d}}(\mu P^n, \nu P^n) \\
&\le (1-\alpha)\mathbf{W}_{\mathsf{d}}(\bar\mu P^n, \bar\nu P^n) + \alpha \mathbf{W}_{\mathsf{d}}(\mu_A P^n, \nu_A P^n) \\
&\le (1-\alpha) + \alpha \iint_{A\times A} \mathbf{W}_{\mathsf{d}}\left(P^n(x,\cdot), P^n(y,\cdot)\right) \mu_A(\mathrm{d}x)\nu_A(\mathrm{d}y) \\
&\le 1-\alpha+\alpha \sup_{(x,y)\in A\times A} \mathbf{W}_{\mathsf{d}}\left(P^n(x,\cdot), P^n(y,\cdot)\right) \; .
\end{aligned}$$

5. Combining (20.6.12) and (20.6.13) and applying Exercise 20.11 (where the assumption that the distances d_k are continuous is used) yields

$$\mathsf{d}_{\mathrm{TV}}(\mu,\nu) = \limsup_{k\to\infty} \mathbf{W}_{\mathsf{d}_k}(\mu,\nu) \le 1-\alpha+\varepsilon\alpha < 1 \; .$$

This is a contradiction, since μ and ν are mutually singular by assumption.

Solutions to Exercises of Chapter 21

21.1 Note first that for $x \notin \alpha$, $\mathbb{E}_x\left[\sum_{k=0}^{\tau_\alpha} h(X_k)\right] = \mathbb{E}_x\left[\sum_{k=0}^{\sigma_\alpha} h(X_k)\right]$. A function $h \in \mathbb{F}(\mathsf{X})$ is integrable and has zero mean with respect to π if and only if

$$\mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} |h(X_k)|\right] < \infty\,, \qquad \mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} h(X_k)\right] = 0\,. \tag{G.34}$$

If $x \in \alpha$, then $\mathbb{E}_x\left[\sum_{k=0}^{\tau_\alpha} h(X_k)\right] = h(x)$ and

$$\mathbb{E}_x\left[\sum_{k=0}^{\sigma_\alpha} h(X_k)\right] = h(x) + \mathbb{E}_x\left[\sum_{k=1}^{\sigma_\alpha} h(X_k)\right] = h(x) + \mathbb{E}_\alpha[\sigma_\alpha]\pi(h) = h(x)\,.$$

This proves the second equality in (21.5.1) for all $x \in \mathsf{X}$. By definition, $\hat{h}(x) = h(x)$ for $x \in \alpha$. Applying the Markov property and the identity $\sigma_\alpha = 1 + \tau_\alpha \circ \theta$, we get

$$P\hat{h}(x) = \mathbb{E}_x\left[\mathbb{E}_{X_1}\left[\sum_{k=0}^{\tau_\alpha} h(X_k)\right]\right] = \mathbb{E}_x\left[\sum_{k=1}^{\sigma_\alpha} h(X_k)\right]\,.$$

Since $\mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} h(X_k)\right] = 0$, we have $P\hat{h}(x) = 0$ for $x \in \alpha$, and thus $\hat{h}(x) - P\hat{h}(x) = \hat{h}(x) = h(x)$. For $x \notin \alpha$,

$$P\hat{h}(x) = \mathbb{E}_x\left[\sum_{k=1}^{\sigma_\alpha} h(X_k)\right] = \mathbb{E}_x\left[\sum_{k=1}^{\tau_\alpha} h(X_k)\right] = \hat{h}(x) - h(x)\,.$$

21.2 1. This is Theorem 6.7.1.

2. By the strong Markov property and the identity $\sigma_\alpha = \tau_\alpha \circ \theta_i + i$ on $\sigma_\alpha \geq i$ for $i \geq 1$, we have

$$\begin{aligned}
\mathbb{E}_\alpha[\sigma_\alpha]\pi(|h\hat{h}|) &= \mathbb{E}_\alpha\left[\sum_{i=1}^{\sigma_\alpha} |h(X_i)\hat{h}(X_i)|\right] = \mathbb{E}_\alpha\left[\sum_{i=1}^{\sigma_\alpha} |h(X_i)|\mathbb{E}_{X_i}\left[\sum_{j=0}^{\sigma_\alpha} |h(X_j)|\right]\right] \\
&= \mathbb{E}_\alpha\left[\sum_{i=1}^{\sigma_\alpha} |h(X_i)|\mathbb{E}\left[\sum_{j=0}^{\sigma_\alpha} |h(X_j)| \circ \theta_i \,\middle|\, \mathscr{F}_i\right]\right] \\
&= \mathbb{E}_\alpha\left[\sum_{i=1}^{\sigma_\alpha} |h(X_i)| \sum_{j=0}^{\sigma_\alpha - i} |h(X_{i+j})|\right] \leq \mathbb{E}_\alpha\left[\left(\sum_{i=1}^{\sigma_\alpha} |h(X_i)|\right)^2\right]\,.
\end{aligned}$$

Since

$$\mathbb{E}_\alpha[\sigma_\alpha]\pi(h^2) = \mathbb{E}_\alpha\left[\sum_{i=1}^{\sigma_\alpha} h^2(X_i)\right] \leq \mathbb{E}_\alpha\left[\left(\sum_{i=1}^{\sigma_\alpha} |h(X_i)|\right)^2\right]\,,$$

the first claim is proved (and is actually an if and only if). Starting from the last line without the absolute values, we obtain

$$
\begin{aligned}
2\mathbb{E}_\alpha[\sigma_\alpha]\pi(h\hat{h}) &= 2\mathbb{E}_\alpha\left[\sum_{i=1}^{\sigma_\alpha} h(X_i)\sum_{j=i}^{\sigma_\alpha} h(X_j)\right] \\
&= \mathbb{E}_\alpha\left[\left(\sum_{i=1}^{\sigma_\alpha} h(X_i)\right)^2\right] + \mathbb{E}_\alpha\left[\sum_{i=0}^{\sigma_\alpha} h^2(X_j)\right] \\
&= \mathbb{E}_\alpha[\sigma_\alpha]\{\sigma^2(h) + \pi(h^2)\} .
\end{aligned}
$$

21.3

$$
\begin{aligned}
\sum_{k=1}^{\infty} \pi(hP^k h) &= \sum_{k=1}^{\infty} \mathbb{E}_\pi[h(X_0)h(X_k)] \\
&= \mathbb{E}_\pi\left[h(X_0)\sum_{k=1}^{\sigma_\alpha} h(X_k)\right] + \sum_{j=1}^{\infty} \mathbb{E}_\pi\left[h(X_0)\sum_{k=\sigma_\alpha^{(j)}+1}^{\sigma_\alpha^{(j+1)}} h(X_k)\right] .
\end{aligned}
$$

By the strong Markov property, for $j \geq 1$,

$$
\begin{aligned}
\mathbb{E}_\pi\left[h(X_0)\sum_{k=\sigma_\alpha^{(j)}+1}^{\sigma_\alpha^{(j+1)}} h(X_k)\right] &= \mathbb{E}_\pi\left[h(X_0)\mathbb{E}\left[\sum_{k=\sigma_\alpha^{(j)}+1}^{\sigma_\alpha^{(j+1)}} h(X_k)\,\middle|\,\mathscr{F}_{\sigma_\alpha^{(j)}}\right]\right] \\
&= \mathbb{E}_\pi\left[h(X_0)\mathbb{E}_\alpha\left[\sum_{k=1}^{\sigma_\alpha} h(X_k)\right]\right] = 0 .
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\sum_{k=1}^{\infty} \pi(hP^k h) &= \mathbb{E}_\pi[h(X_0)\sum_{k=1}^{\sigma_\alpha} h(X_k)] \\
&= \mathbb{E}_\pi\left[h(X_0)\mathbb{E}\left[\sum_{k=1}^{\sigma_\alpha} h(X_k)\,\middle|\,\mathscr{F}_0\right]\right] \\
&= \mathbb{E}_\pi[h(X_0)\{\hat{h}(X_0) - h(X_0)\}] = \pi(h\hat{h}) - \pi(h^2) .
\end{aligned}
$$

This yields $\pi(h^2) + 2\sum_{k=1}^{\infty} \pi(hP^k h) = \pi(h\hat{h}) - \pi(h^2) = \sigma^2(h)$ by Exercise 21.2. The second equality follows from Lemma 21.2.7.

21.5 1. This Markov chain is irreducible and aperiodic, with stationary distribution given by $\pi(0) = 1/2$ and $\pi(j) = \pi(-j) = c'/j^3$, where $c' = \zeta(3)/4$. Furthermore, $\pi(h) = 0$.
Since $h(j) + h(-j) = 0$ and $h(0) = 0$, it is easy to see that for $n \geq 2$, we have

$$\sum_{i=0}^{n-1} h(X_i) = \mathbb{1}_{\{X_0<0\}} X_1 + \mathbb{1}_{\{X_{n-1}>0\}} X_{n-1} \ .$$

In particular, $\sum_{i=0}^{n-1} h(X_i) \le |X_0| + |X_{n-1}|$, and since by stationarity $\mathbb{E}_\pi[|X_0|] = \mathbb{E}_\pi|X_{n-1}| = \sum_{x\neq 0} |x|c'|x|^{-3} < \infty$, it follows immediately that $n^{-1/2}\sum_{i=0}^{n-1} h(X_i)$ converges in distribution to 0, i.e., to $\mathrm{N}(0,\ 0)$. It also follows that for $n \ge 2$,

$$\mathrm{Var}_\pi\left(\sum_{i=0}^{n-1} h(X_i)\right) = 2\mathbb{E}_\pi[X_0^2 \mathbb{1}_{\{X_0>0\}}] = 2\sum_{j=1}^{\infty} j^2(c'/j^3) = \infty.$$

2. Replace the state space X by $\mathsf{X} \times \{-1,\ 1\}$, let the first coordinate $\{X_n,\ n\in\mathbb{N}\}$ evolve as before, let the second coordinate $\{Y_n,\ n\in\mathbb{N}\}$ evolve independently of $\{X_n,\ n\in\mathbb{N}\}$ such that each $\{Y_n,\ n\in\mathbb{N}\}$ is i.i.d. equal to -1 or 1 with probability $1/2$ each, and redefine h as $h(x,\ y) = x+y$. Then $n^{-1/2}\sum_{i=0}^{n-1} h(X_i, Y_i)$ will converge in distribution to $\mathrm{N}(0,1)$.

21.6 1. Let f be a 1-Lipschitz function. Then

$$\mathbb{E}\left[f((x+\varepsilon_1)/2) - f((y+\varepsilon_1)/2)\right] \le |x-y|/2 \ .$$

By the duality theorem this proves that $\Delta_d(P) \le 1/2$, and thus there exists a unique invariant probability by Theorem 20.3.4. The invariant measure is Lebesgue measure on $[0,1]$.

2. Since Lebesgue measure is invariant for π, it follows that $\int_0^1 P^k f(x)\mathrm{d}x = 0$ for all $k \ge 1$. Therefore,

$$\begin{aligned} P^k f(x) &= 2^{-k} \sum_{z\in D_k} f\left(\frac{x}{2^k}+z\right) \\ &= 2^{-k} \sum_{z\in D_k} \int_0^1 \left[f\left(\frac{x}{2^k}+z\right) - f\left(\frac{y}{2^k}+z\right)\right] \mathrm{d}y \ . \end{aligned}$$

3. The previous identity yields

$$\begin{aligned} \left\|P^k f\right\|_2^2 &\le 2^{-k} \sum_{z\in D_k} \int_0^1 \int_0^1 \left[f\left(\frac{x}{2^k}+z\right) - f\left(\frac{y}{2^k}+z\right)\right]^2 \mathrm{d}y\mathrm{d}x \\ &\le 2^k \iint_{|x-y|\le 2^{-k}} [f(x)-f(y)]^2 \mathrm{d}x\mathrm{d}y \ . \end{aligned}$$

4. If f is Hölder continuous, then

$$\left\|P^k f\right\|_2^2 \le 2^k \iint_{|x-y|\le 2^{-k}} [f(x)-f(y)]^2 \mathrm{d}x\mathrm{d}y$$
$$\le C2^k \iint_{|x-y|\le 2^{-k}} |x-y|^{2\gamma} \mathrm{d}x\mathrm{d}y \le C2^{k(2\gamma-1)} .$$

This proves that $\sum_{k=1}^{\infty} \left\|P^k f\right\|_2^2 < \infty$, which implies that (21.4.2) holds.

Solutions to Exercises of Chapter 22

22.1

$$P[f+g](x) = \mathbb{1}_{A_{f+g}}(x) \int \{f(y)+g(y)\} P(x,\mathrm{d}y)$$

and

$$Pf(x)+Pg(x) = \mathbb{1}_{A_f}(x) \int f(y)P(x,\mathrm{d}y) + \mathbb{1}_{A_g} \int g(y)P(x,\mathrm{d}y) .$$

These two functions coincide on $A_f \cap A_g$, and since $\pi(A_f)=1$ and $\pi(A_g)=1$, we have

$$P[f+g](x) = Pf(x)+Pg(x) , \quad \pi - a.e.$$

22.3 1.

$$\mathbb{E}_\nu[|\pi(f)-S_{n,n_0}(f)|^2] = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n} \mathbb{E}_\nu[f(X_{n_0+j})f(X_{n_0+i})]$$
$$= \frac{1}{n^2}\sum_{j=1}^{n} \int P^{n_0+j}(f^2)(x)\nu(\mathrm{d}x) + \frac{2}{n^2}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} \int P^{n_0+j}(fP^{k-j}f)(x)\nu(\mathrm{d}x) .$$

For $h \in \mathrm{L}^r(\pi)$ and $\nu \in \mathbb{M}_{r/(r-1)}(\pi)$, we have for all $i \in \mathbb{N}$ that $\frac{\mathrm{d}\nu}{\mathrm{d}\pi} \cdot P^i h$ is integrable with respect to π. Then for every $h \in \mathrm{L}_0^r(\pi)$, we get

$$\int P^i h(x)\nu(\mathrm{d}x) = \left\langle P^i h, \frac{\mathrm{d}\nu}{\mathrm{d}\pi} \right\rangle_{\mathrm{L}^2(\pi)} = \langle P^i h, \mathbf{1}\rangle_{\mathrm{L}^2(\pi)} + \left\langle P^i h, \left(\frac{\mathrm{d}\nu}{\mathrm{d}\pi}-1\right)\right\rangle_{\mathrm{L}^2(\pi)} .$$

2. Applying Hölder's inequality with conjugate parameter r and $s = \dfrac{r}{r-1}$ to $L_k(h) = \langle P^k h, \frac{\mathrm{d}\nu}{\mathrm{d}\pi}-1\rangle_{\mathrm{L}^2(\pi)}$, one has

$$|L_k(h)| \le \left\|P^k h\right\|_{\mathrm{L}^r(\pi)} \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\pi}-1\right\|_{\mathrm{L}^s(\pi)} \le |||P^k|||_{\mathrm{L}_0^r(\pi)} \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\pi}-1\right\|_{\mathrm{L}^s(\pi)} \|h\|_{\mathrm{L}^r(\pi)} .$$

The proofs of items 2 and 3 follow.

22.4 1. This Markov kernel is irreducible and aperiodic and is reversible with respect to the stationary distribution given by $\pi(x) = c'|x|^{-3}$ and $\pi(0) = c'/c$, where $c' = [c^{-1} + 2\zeta(3)]^{-1}$. Hence the chain is positive recurrent.

2. We use Theorem 6.7.1. We consider the state $\mathsf{a} = \{0\}$. The sum over a single tour, $\sum_{i=1}^{\sigma_\alpha} X_i$, is $X_{\sigma_\alpha+1}$, $-X_{\sigma_\alpha+1}$, or 0. Furthermore, $\mathbb{P}_\alpha(X_1 = y) = P(0,y) = c|y|^{-4}$, so $\mathbb{E}_\alpha[X_1^2] = \sum_{y\neq 0} y^2 c|y|^{-4} < \infty$. This implies that $\sum_{i=1}^{\sigma_\alpha} X_i$ has finite variance, say V. It then follows from Theorem 6.7.1 that

$$n^{-1/2}\sum_{i=0}^{n} X_i \stackrel{\mathbb{P}_\pi}{\Longrightarrow} \mathrm{N}(0,\ V/\mathbb{E}_\alpha[\sigma_\alpha]),$$

where $\mathbb{E}_\alpha[\sigma_\alpha] = 1/\pi(0)$ (by Theorem 6.4.2).

3.

$$\mathrm{Var}_\pi(X_0) = \sum_{x\in\mathsf{X}} \mathbb{P}_\pi(X_0 = x)x^2 = \sum_{x\in\mathsf{X}} c'|x|^{-3}x^2 = \sum_{x\in\mathsf{X}} c'|x|^{-1} = \infty\,.$$

4. $\mathbb{P}_\pi(\tau_\alpha \le n) = \sum_{x\in\mathsf{X}} \pi(x)\mathbb{P}_x(\tau_\alpha \le n)$. For $x \neq 0$, $\mathbb{P}_x(\tau_\alpha \le n) \in (0,1)$. For n even, we have $S_n = X_0$ on the event D_n^c because of cancellation.

5.

$$\begin{aligned}\mathbb{E}_\pi[S_n^2 \mathbb{1}_{D_n^c}] = \mathbb{E}_\pi[X_0^2 \mathbb{1}_{D_n^c}] &= \sum_{x\in\mathsf{X}} x^2\pi(x)\mathbb{P}_\pi(\tau_\alpha > n)\\ &= \sum_{x\in\mathsf{X}} c'|x|^{-3}(1-|x|^{-1})^n x^2 = \sum_{x\in\mathsf{X}} c'|x|^{-1}(1-|x|^{-1})^n = \infty.\end{aligned}$$

Hence $\mathrm{Var}_\pi(S_n) = \mathbb{E}_\pi[S_n^2] = \infty$. In particular, the limit in the definition of $\lim_{n\to\infty} n^{-1}\mathbb{E}_\pi[S_n^2]$ is either infinite or undefined.

22.5 1. P is a nonnegative kernel, since for all $x \in \mathsf{X}$ and $A \in \mathscr{X}$,

$$\begin{aligned}P(x,\{x\}) &= 1 + P_1(x,\{x\}) - P_0(x,\{x\}) \ge 0\,,\\ P(x,A\setminus\{x\}) &= P_1(x,A\setminus\{x\}) - P_0(x,A\setminus\{x\}) \ge 0\,.\end{aligned}$$

Combining with $P(x,\mathsf{X}) = 1$, this implies that P is a Markov kernel.

2.

$$\begin{aligned}\langle f, P_0 f\rangle - \langle f, P_1 f\rangle &= \iint \pi(\mathrm{d}x) f(x)(P_0(x,\mathrm{d}y) - P_1(x,\mathrm{d}y))f(y)\\ &= \iint \pi(\mathrm{d}x) f(x)(\delta_x(\mathrm{d}y) - P(x,\mathrm{d}y))f(y)\\ &= \int \pi(\mathrm{d}x) f^2(x) - \iint \pi(\mathrm{d}x)P(x,\mathrm{d}y)f(x)f(y)\\ &= \iint \pi(\mathrm{d}x)P(x,\mathrm{d}y)\left[\frac{f^2(x)-f^2(y)}{2} + f(x)f(y)\right],\end{aligned}$$

where the last inequality follows from the fact that P is clearly π-invariant. Finally,

$$\langle f, P_0 f\rangle - \langle f, P_1 f\rangle = \iint \pi(\mathrm{d}x)P(x,\mathrm{d}y)\left(f(x)-f(y)\right)^2/2 \geq 0\,.$$

The proof is complete.

22.6 Assume that the spectral measure of P_1 and P_2 is not concentrated at -1. Then for $i \in \{0,1\}$, using Proposition 22.5.2, we get

$$v_i(f,P_i) = \int_{-1}^{1} \frac{1+x}{1-x}\mu_{f,P_i}(\mathrm{d}x)\,.$$

1. For all $1 \leq \ell \leq k$ and all $\alpha_1,\ldots,\alpha_k$,

$$\begin{aligned}\left\langle f, P_{\alpha_1}\ldots P_{\alpha_k} f\right\rangle = (1-\alpha_\ell)\left\langle f, P_{\alpha_1}\ldots P_{\alpha_{\ell-1}}P_0 P_{\alpha_{\ell+1}}\ldots P_{\alpha_k} f\right\rangle \\ + \alpha_\ell \left\langle f, P_{\alpha_1}\ldots P_{\alpha_{\ell-1}}P_1 P_{\alpha_{\ell+1}}\ldots P_{\alpha_k} f\right\rangle\,,\end{aligned}$$

so that

$$\frac{\partial}{\partial \alpha_\ell}\left\langle f, P_{\alpha_1}\ldots P_{\alpha_k} f\right\rangle = \left\langle f, P_{\alpha_1}\ldots P_{\alpha_{\ell-1}}(P_1-P_0)P_{\alpha_{\ell+1}}\ldots P_{\alpha_k} f\right\rangle\,,$$

and thus we obtain by differentiating $\alpha \mapsto w_\lambda(\alpha)$,

$$\frac{\mathrm{d}w_\lambda(\alpha)}{\mathrm{d}\alpha} = \sum_{k=0}^{\infty}\lambda^k \sum_{i=1}^{k}\left\langle f, P_\alpha^{i-1}(P_1-P_0)P_\alpha^{k-i} f\right\rangle\,.$$

2. Using that π is reversible for the kernel P_α, we have

$$\begin{aligned}\frac{\mathrm{d}w_\lambda(\alpha)}{\mathrm{d}\alpha} &= \sum_{i=1}^{\infty}\sum_{k\geq i}^{\infty}\lambda^k\left\langle P_\alpha^{i-1}f, (P_1-P_0)P_\alpha^{k-i}f\right\rangle \\ &= \lambda\left\langle \sum_{\ell=0}^{\infty}\lambda^\ell P_\alpha^\ell f, (P_1-P_0)\sum_{\ell=0}^{\infty}\lambda^\ell P_\alpha^\ell f\right\rangle \leq 0\,,\end{aligned}$$

which completes the proof.

Solutions to Exercises of Chapter 23

23.1 1. Note that by convexity of the exponential function,

$$\mathrm{e}^{sx} \leq \frac{x-A}{B-A}\,\mathrm{e}^{sB} + \frac{B-x}{B-A}\,\mathrm{e}^{sA}\,, \quad \text{for } A \leq x \leq B.$$

Since $\mathbb{E}\left[V\,|\,\mathscr{G}\right] = 0$, we get

$$\mathbb{E}\left[\mathrm{e}^{sV}\,\middle|\,\mathscr{G}\right] \le \frac{B}{B-A}\,\mathrm{e}^{sA} - \frac{A}{B-A}\,\mathrm{e}^{sB}$$
$$= (1-p+p\mathrm{e}^{s(B-A)})\mathrm{e}^{-ps(B-A)} = \mathrm{e}^{\phi(s(B-A))}.$$

2. The derivative of ϕ is

$$\phi'(u) = -p + \frac{p}{p+(1-p)\mathrm{e}^{-u}};$$

therefore $\phi(0) = \phi'(0) = 0$. In addition,

$$\phi''(u) = \frac{p(1-p)\mathrm{e}^{-u}}{(p+(1-p)\mathrm{e}^{-u})^2} \le \frac{1}{4}\,.$$

Thus by the Taylor–Lagrange theorem, $\phi(u) \le u^2/8$, which concludes the proof.

23.4 1. It follows from Lemma 20.3.2 that if f is a Lipschitz function, then Pf is also Lipschitz and $|Pf|_{\mathrm{Lip(d)}} \le \Delta_{\mathsf{d}}(P)\,|f|_{\mathrm{Lip(d)}}$. Since π is invariant for P, we have, for every $i \ge 1$,

$$|P^i f(x) - \pi(f)| = |P^i f(x) - \pi(P^i f)| \le \int_{\mathsf{X}} |P^i f(x) - P^i f(y)|\pi(\mathrm{d}y)$$
$$\le (1-\kappa)^i \int_{\mathsf{X}} \mathsf{d}(x,y)\pi(\mathrm{d}y) = (1-\kappa)^i E(x)\,.$$

Summing over i yields (23.5.3), since $\sum_{i=0}^{\infty}(1-\kappa)^i = \kappa^{-1}$.

2. By applying question 1, we obtain

$$\mathbb{P}_x(|\hat{\pi}_n(f) - \pi(f)| > t) \le \mathbb{P}_x(|\hat{\pi}_n(f) - \mathbb{E}_x[\hat{\pi}_n(f)]| + (n\kappa)^{-1}\,|f|_{\mathrm{Lip(d)}}\,\mathrm{diam}(\mathsf{X}) > t)\,.$$

We conclude by applying Theorem 23.4.5 with $\gamma_i = n^{-1}\,|f|_{\mathrm{Lip(d)}}$, $i=0,\ldots,n-1$.

23.5 1. If h is continuously differentiable on $\mathbb{R}^d$, then $|\nabla h|_\infty \le |h|_{\mathrm{Lip(d)}}$. Thus for all $x \in \mathbb{R}^d$,

$$P(|\nabla f_t|^2)(x) = \frac{t^2}{4} P\left(|\nabla h|^2 f^2\right)(x) \le \frac{t^2\,|h|^2_{\mathrm{Lip(d)}}}{4} P\left(f_t^2\right)(x)\,.$$

2. Applying (23.5.5) and (23.5.6) to f_t and the definition of Λ yields

$$P\left(\left\{th - \frac{1}{2}t^2 C\,|h|^2_{\mathrm{Lip(d)}}\right\} f_t^2\right)(x) - \Lambda(t,x)\log\Lambda(t,x) \le \frac{Ct^2\,|h|^2_{\mathrm{Lip(d)}}}{2} P\left(f_t^2\right)(x)\,.$$

This yields

$$P\left(\left\{th - C\,|h|^2_{\mathrm{Lip(d)}}\right\} f_t^2\right)(x) \le \Lambda(t,x)\log\Lambda(t,x)\;.$$

It is easily checked that the left-hand side is exactly $t\Lambda'(t,x) - \Lambda(t,x)\log\Lambda(t,x)$. This proves (23.5.7).

3. The inequality (23.5.7) implies that the function $t \to t^{-1}\log\Lambda(t,x)$ is nonincreasing. Since it vanishes at zero, this yields $\Lambda(t,x) \le 1$ for all $t \ge 0$ and $x \in \mathbb{R}^d$. By definition of Λ, this means that (23.4.5) holds with $\beta^2 = C/4$ and $\delta = \infty$.

References

Adamczak R (2008) A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. Electron J Probab 13(34):1000–1034

Ambrosio L (2003) Lecture notes on optimal transport problems. In: Mathematical aspects of evolving interfaces (Funchal, 2000), Lecture Notes in Math., vol 1812, Springer, Berlin, pp 1–52, DOI https://doi.org/10.1007/978-3-540-39189-0_1

Ambrosio L, Gigli N (2013) A user's guide to optimal transport. In: Modelling and optimisation of flows on networks, Lecture Notes in Math., vol 2062, Springer, Heidelberg, pp 1–155, DOI https://doi.org/10.1007/978-3-642-32160-3_1

An HZ, Chen SG (1997) A note on the ergodicity of non-linear autoregressive model. Statist Probab Lett 34(4):365–372, DOI https://doi.org/10.1016/S0167-7152(96)00204-0

Andrieu C, Vihola M (2015) Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. Ann Appl Probab 25(2):1030–1077, DOI https://doi.org/10.1214/14-AAP1022

Andrieu C, Fort G, Vihola M (2015) Quantitative convergence rates for subgeometric Markov chains. J Appl Probab 52(2):391–404, DOI https://doi.org/10.1239/jap/1437658605

Anscombe FJ (1952) Large-sample theory of sequential estimation. Proc Cambridge Philos Soc 48:600–607

Anscombe FJ (1953) Sequential estimation. J Roy Statist Soc Ser B 15:1–21; discussion, 21–29, URL http://links.jstor.org/sici?sici=0035-9246(1953)15:1⟨1:SE⟩2.0.CO;2-R&origin=MSN

Atchadé YF (2011) Kernel estimators of asymptotic variance for adaptive Markov chain Monte Carlo. Ann Statist 39(2):990–1011, DOI https://doi.org/10.1214/10-AOS828

Atchadé YF (2016) Markov chain Monte Carlo confidence intervals. Bernoulli 22(3):1808–1838, DOI https://doi.org/10.3150/15-BEJ712

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1

Athreya KB, Ney P (1978) A new approach to the limit theory of recurrent Markov chains. Trans Am Math Soc 245:493–501

Balaji S, Meyn SP (2000) Multiplicative ergodicity and large deviations for an irreducible Markov chain. Stochastic Process Appl 90(1):123–144, DOI https://doi.org/10.1016/S0304-4149(00)00032-6

Basharin GP, Langville AN, Naumov VA (2004) The life and work of A. A. Markov. Linear Algebra Appl 386:3–26, DOI https://doi.org/10.1016/j.laa.2003.12.041

Baxendale PH (2005) Renewal theory and computable convergence rates for geometrically ergodic Markov chains. Ann Appl Probab 15(1B):700–738

Bednorz W (2013) The Kendall theorem and its application to the geometric ergodicity of Markov chains. Appl Math (Warsaw) 40(2):129–165, DOI https://doi.org/10.4064/am40-2-1

Bercu B, Delyon B, Rio E (2015) Concentration inequalities for sums and martingales. Springer Briefs in Mathematics, Springer, Cham, DOI https://doi.org/10.1007/978-3-319-22099-4

Bernstein S (1927) Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. Math Ann 97(1):1–59, DOI https://doi.org/10.1007/BF01447859

Bhattacharya R, Lee C (1995) On geometric ergodicity of nonlinear autoregressive models. Statist Probab Lett 22(4):311–315, DOI https://doi.org/10.1016/0167-7152(94)00082-J

Billingsley P (1961) Statistical inference for Markov processes. Statistical Research Monographs, Vol. II. The University of Chicago Press, Chicago, Ill.

Billingsley P (1978) Ergodic theory and information. Robert E. Krieger Publishing Co., Huntington, N.Y., reprint of the 1965 original

Billingsley P (1986) Probability and measure, 2nd edn. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, a Wiley-Interscience Publication

Billingsley P (1999) Convergence of probabilities, 2nd edn. Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons Inc., New York, a Wiley-Interscience Publication

Blackwell D (1948) A renewal theorem. Duke Math J 15:145–150, URL http://projecteuclid.org/euclid.dmj/1077474668

Blackwell D, Freedman D (1964) The tail σ-field of a Markov chain and a theorem of Orey. Ann Math Statist 35:1291–1295, DOI https://doi.org/10.1214/aoms/1177703284

Bogachev VI, Kolesnikov AV (2012) The Monge-Kantorovich problem: achievements, connections, and prospects. Uspekhi Mat Nauk 67(5(407)):3–110

Borovkov AA (1998) Ergodicity and stability of stochastic processes. Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Ltd., Chichester, translated from the 1994 Russian original by V. Yurinsky [V. V. Yurinskiĭ]

Boucheron S, Lugosi G, Massart P (2013) Concentration inequalities. Oxford University Press, Oxford, DOI https://doi.org/10.1093/acprof:oso/9780199535255.001.0001, a nonasymptotic theory of independence, With a foreword by Michel Ledoux

Bradley RC (2005) Basic properties of strong mixing conditions. A survey and some open questions. Probab Surv 2:107–144, DOI https://doi.org/10.1214/154957805100000104, update of, and a supplement to, the 1986 original

Bradley RC (2007a) Introduction to strong mixing conditions. Vol. 1. Kendrick Press, Heber City, UT

Bradley RC (2007b) Introduction to strong mixing conditions. Vol. 2. Kendrick Press, Heber City, UT

Bradley RC (2007c) Introduction to strong mixing conditions. Vol. 3. Kendrick Press, Heber City, UT

Brémaud P (1999) Markov chains, Texts in Applied Mathematics, vol 31. Springer, New York, DOI https://doi.org/10.1007/978-1-4757-3124-8, gibbs fields, Monte Carlo simulation, and queues

Brooks S, Gelman A, Jones GL, Meng XL (eds) (2011) Handbook of Markov chain Monte Carlo. Chapman & Hall/CRC Handbooks of Modern Statistical Methods, CRC Press, Boca Raton, FL, DOI https://doi.org/10.1201/b10905

Butkovsky O (2014) Subgeometric rates of convergence of Markov processes in the Wasserstein metric. Ann Appl Probab 24(2):526–552, DOI https://doi.org/10.1214/13-AAP922

Butkovsky OA, Veretennikov AY (2013) On asymptotics for Vaserstein coupling of Markov chains. Stochastic Process Appl 123(9):3518–3541, DOI https://doi.org/10.1016/j.spa.2013.04.016

Cattiaux P, Guillin A (2014) Functional inequalities via Lyapunov conditions. In: Optimal transportation, London Math. Soc. Lecture Note Ser., vol 413, Cambridge Univ. Press, Cambridge, pp 274–287

Çinlar E (1975) Introduction to stochastic processes. Prentice-Hall, Inc., Englewood Cliffs, N.J.

Chan KS, Petruccelli JD, Tong H, Woolford SW (1985) A multiple-threshold AR(1) model. J Appl Probab 22(2):267–279

Chang H, Hu J, Fu MC, Marcus SI (2013) Simulation-based algorithms for Markov decision processes, 2nd edn. Communications and Control Engineering Series, Springer, London, DOI https://doi.org/10.1007/978-1-4471-5022-0

Charmasson T, Petit M, Méchine S (2005) Archives et manuscrits de Wolfgang Doeblin. Revue d'histoire des sciences 58(1):225–236, DOI https://doi.org/10.3406/rhs.2005.2245, URL http://www.persee.fr/doc/rhs_0151-4105_2005_num_58_1_2245

Chazottes JR, Gouëzel S (2012) Optimal concentration inequalities for dynamical systems. Comm Math Phys 316(3):843–889, DOI https://doi.org/10.1007/s00220-012-1596-7

Chen X (1999) Limit theorems for functionals of ergodic Markov chains with general state space. Mem Amer Math Soc 139(664):xiv+203

Chen X, Guillin A (2004) The functional moderate deviations for Harris recurrent Markov chains and applications. Ann Inst H Poincaré Probab Statist 40(1):89–124, DOI https://doi.org/10.1016/S0246-0203(03)00061-X

Ching WK, Huang X, Ng MK, Siu TK (2013) Markov chains, International Series in Operations Research & Management Science, vol 189, 2nd edn. Springer, New York, DOI https://doi.org/10.1007/978-1-4614-6312-2, models, algorithms and applications

Chung KL (1953) Contributions to the theory of Markov chains. J Research Nat Bur Standards 50:203–208

Chung KL (1954) Contributions to the theory of Markov chains. II. Trans Amer Math Soc 76:397–419, DOI https://doi.org/10.2307/1990789

Chung KL (1964) The general theory of Markov processes according to Doeblin. Z Wahrscheinlichkeitstheorie und Verw Gebiete 2:230–254 (1964), DOI https://doi.org/10.1007/BF00533381

Chung KL (1967) Markov chains with stationary transition probabilities. Second edition. Die Grundlehren der mathematischen Wissenschaften, Band 104, Springer, New York, Inc.

Cline DB, Pu HH (1999) Geometric ergodicity of nonlinear time series. Statist Sinica 9(4):1103–1118

Cline DB, Pu HH (2002) A note on a simple Markov bilinear stochastic process. Statist Probab Lett 56(3):283–288, DOI https://doi.org/10.1016/S0167-7152(01)00192-4

Cline DBH, Pu HH (2004) Stability and the Lyapounov exponent of threshold AR-ARCH models. Ann Appl Probab 14(4):1920–1949, DOI https://doi.org/10.1214/105051604000000431

Cogburn R (1972) The central limit theorem for markov processes. In: Le Cam LM, Neyman J, Scott EL (eds) Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, Calif., pp i+605, held at the Statistical Laboratory, University of California, Berkeley, Calif., June 21–July 18, April 9–12, June 16–21 and July 19–22, 1971, Volume II: Probability theory

Connor S, Kendall W (2007) Perfect simulation for a class of positive recurrent Markov chains. Adv Appl Probab to appear

Connor SB, Fort G (2009) State-dependent Foster-Lyapunov criteria for subgeometric convergence of Markov chains. Stochastic Process Appl 119(12):4176–4193, DOI https://doi.org/10.1016/j.spa.2009.10.001

Cox DR (1962) Renewal theory. Methuen & Co. Ltd., London; John Wiley & Sons, Inc., New York

Cuny C (2017) Invariance principles under the Maxwell-Woodroofe condition in Banach spaces. Ann Probab 45(3):1578–1611, DOI https://doi.org/10.1214/16-AOP1095

Cuny C, Lin M (2016) Limit theorems for Markov chains by the symmetrization method. J Math Anal Appl 434(1):52–83, DOI https://doi.org/10.1016/j.jmaa.2015.07.061

Dai J (1995) On positive Harris recurrence for multiclass queuing networks: a unified approach via fluid limit models. Ann Appl Probab 5:49–77

Dai J, Weiss G (1996) Stability and instability of fluid models for reentrant lines. Math Oper Res 21(1):115–134

Dai JG, Meyn SP (1995) Stability and convergence of moments for multiclass queueing networks via fluid limit models. IEEE Trans Automat Control 40(11):1889–1904

Dedecker J, Gouëzel S (2015) Subgaussian concentration inequalities for geometrically ergodic Markov chains. Electron Commun Probab 20(64):12, DOI https://doi.org/10.1214/ECP.v20-3966

Dedecker J, Rio E (2000) On the functional central limit theorem for stationary processes. Ann Inst H Poincaré Probab Statist 36(1):1–34, DOI https://doi.org/10.1016/S0246-0203(00)00111-4

Dedecker J, Rio E (2008) On mean central limit theorems for stationary sequences. Ann Inst Henri Poincaré Probab Stat 44(4):693–726, DOI https://doi.org/10.1214/07-AIHP117

Dedecker J, Doukhan P, Lang G, León R, Louhichi S, Prieur C (2007) Weak dependence: with examples and applications, Lecture Notes in Statistics, vol 190. Springer, New York

Del Moral P, Ledoux M, Miclo L (2003) On contraction properties of Markov kernels. Probab Theory Related Fields 126(3):395–420

Diaconis P (2009) The Markov chain Monte Carlo revolution. Bull Amer Math Soc (NS) 46(2):179–205, DOI https://doi.org/10.1090/S0273-0979-08-01238-X

Diaconis P (2013) Some things we've learned (about Markov chain Monte Carlo). Bernoulli 19(4):1294–1305, DOI https://doi.org/10.3150/12-BEJSP09

Diaconis P, Freedman D (1999) Iterated random functions. SIAM Rev 47(1):45–76

Diaconis P, Hanlon P (1992) Eigen-analysis for some examples of the Metropolis algorithm. In: Hypergeometric functions on domains of positivity, Jack polynomials, and applications (Tampa, FL, 1991), Contemp. Math., vol 138, Amer. Math. Soc., Providence, RI, pp 99–117, DOI https://doi.org/10.1090/conm/138/1199122

Djellout H, Guillin A (2001) Moderate deviations for Markov chains with atom. Stochastic Process Appl 95(2):203–217, DOI https://doi.org/10.1016/S0304-4149(01)00100-4

Djellout H, Guillin A, Wu L (2004) Transportation cost-information inequalities and applications to random dynamical systems and diffusions. Ann Probab 32(3B):2702–2732, DOI https://doi.org/10.1214/009117904000000531

Dobrushin RL (1956a) Central limit theorem for nonstationary Markov chains. II. Teor Veroyatnost i Primenen 1:365–425

Dobrushin RL (1956b) An example of a countable homogeneous Markov process all states of which are instantaneous. Teor Veroyatnost i Primenen 1:481–485

Dobrushin RL (1956c) On the condition of the central limit theorem for inhomogeneous Markov chains. Dokl Akad Nauk SSSR (NS) 108:1004–1006

Doeblin W (1938) Sur deux problèmes de M. Kolmogoroff concernant les chaînes dénombrables. Bull Soc Math France 66:210–220, URL http://www.numdam.org/item?id=BSMF_1938__66__210_0

Doeblin W (1940) éléments d'une théorie générale des chaînes simples constantes de Markoff. Ann École Norm (3) 57:61–111, URL http://www.numdam.org/item?id=ASENS_1940_3_57__61_0

Doob J (1953) Stochastic Processes. Wiley, London

Douc R, Fort G, Moulines E, Soulier P (2004a) Practical drift conditions for subgeometric rates of convergence. Ann Appl Probab 14(3):1353–1377

Douc R, Moulines E, Rosenthal J (2004b) Quantitative bounds for geometric convergence rates of Markov chains. Ann Appl Probab 14(4):1643–1665

Douc R, Moulines E, Soulier P (2006) Subgeometric ergodicity of Markov chains. In: Dependence in probability and statistics, Lect. Notes Stat., vol 187, Springer, New York, pp 55–64, DOI https://doi.org/10.1007/0-387-36062-X_2

Douc R, Moulines E, Soulier P (2007) Computable convergence rates for subgeometric ergodic Markov chains. Bernoulli 13(3):831–848, DOI https://doi.org/10.3150/07-BEJ5162

Doukhan P (1994) Mixing, Lecture Notes in Statistics, vol 85. Springer, New York, DOI https://doi.org/10.1007/978-1-4612-2642-0, properties and examples

Doukhan P, Ghindès M (1983) Estimation de la transition de probabilité d'une chaîne de Markov Doëblin-récurrente. Étude du cas du processus autorégressif général d'ordre 1. Stochastic Process Appl 15(3):271–293, DOI https://doi.org/10.1016/0304-4149(83)90036-4

Doukhan P, Massart P, Rio E (1994) The functional central limit theorem for strongly mixing processes. Ann Inst H Poincaré Probab Statist 30(1):63–82, URL http://www.numdam.org/item?id=AIHPB_1994__30_1_63_0

Dudley RM (2002) Real Analysis and Probability. Cambridge University Press

Duflo M (1997) Random Iterative Models, vol 34. Springer, Berlin, translated from the 1990 French original by S. S. Wilson and revised by the author

Durmus A, Moulines E (2015) Quantitative bounds of convergence for geometrically ergodic Markov chain in the Wasserstein distance with application to the Metropolis adjusted Langevin algorithm. Stat Comput 25(1):5–19, DOI https://doi.org/10.1007/s11222-014-9511-z

Durmus A, Fort G, Moulines E (2016) Subgeometric rates of convergence in Wasserstein distance for Markov chains. Ann Inst Henri Poincaré Probab Stat 52(4):1799–1822, DOI https://doi.org/10.1214/15-AIHP699

Durrett R, Resnick SI (1978) Functional limit theorems for dependent variables. Ann Probab 6(5):829–846, URL http://links.jstor.org/sici?sici=0091-1798(197810)6:5⟨829:FLTFDV⟩2.0.CO;2-J&origin=MSN

Eagleson GK (1975) On Gordin's central limit theorem for stationary processes. J Appl Probability 12:176–179

Feller W (1971) An Introduction to Probability Theory and its Applications. Wiley

Flegal JM, Jones G (2010) Batch means and spectral variance estimators in markov chain monte carlo. Ann Statist 38(2):1034–1070, DOI https://doi.org/10.1214/09-AOS735

Flegal JM, Jones GL (2011) Implementing MCMC: estimating with confidence. In: Handbook of Markov chain Monte Carlo, Chapman & Hall/CRC Handb. Mod. Stat. Methods, CRC Press, Boca Raton, FL, pp 175–197

Foguel SR (1962) Existence of invariant measures for Markov processes. Proc Amer Math Soc 13:833–838, DOI https://doi.org/10.2307/2034070

Foguel SR (1968) Existence of a σ finite invariant measure for a Markov process on a locally compact space. Israel J Math 6:1–4, DOI https://doi.org/10.1007/BF02771598

Foguel SR (1969) The ergodic theory of Markov processes. Van Nostrand Mathematical Studies, No. 21, Van Nostrand Reinhold Co., New York-Toronto, Ont.-London

Foguel SR (1973) The ergodic theory of positive operators on continuous functions. Ann Scuola Norm Sup Pisa (3) 27:19–51

Fort G (2001) Contrôle explicite d'ergodicité de chaînes de Markov: applications àl'analyse de convergence de l'algorithme Monte Carlo EM. PhD thesis, Université de Paris VI

Fort G, Moulines E (2000) V-subgeometric ergodicity for a Hastings-Metropolis algorithm. Statist Probab Lett 49(4):401–410

Fort G, Moulines E (2003a) Convergence of the Monte Carlo expectation maximization for curved exponential families. Ann Statist 31(4):1220–1259

Fort G, Moulines E (2003b) Polynomial ergodicity of Markov transition kernels. Stoch Process Appl 103:57–99

Fort G, Meyn S, Moulines E, Priouret P (2006) ODE methods for Markov Chain stability with applications to MCMC. In: Proceedings of the 1st international conference on Performance evaluation methodolgies and tools, valuetools 2006

Foster FG (1952a) A Markov chain derivation of discrete distributions. Ann Math Statistics 23:624–627

Foster FG (1952b) On Markov chains with an enumerable infinity of states. Proc Cambridge Philos Soc 48:587–591

Foster FG (1953) On the stochastic matrices associated with certain queuing processes. Ann Math Statistics 24:355–360

Fralix BH (2006) Foster-type criteria for Markov chains on general spaces. J Appl Probab 43(4):1194–1200, DOI https://doi.org/10.1239/jap/1165505219

Frühwirth-Schnatter S (2006) Finite mixture and Markov switching models. Springer Series in Statistics, Springer, New York

Gamerman D, Lopes H (2006) Markov chain Monte Carlo, 2nd edn. Texts in Statistical Science Series, Chapman & Hall/CRC, Boca Raton, FL, stochastic simulation for Bayesian inference

Gelfand AE, Smith AFM (1990) Sampling based approaches to calculating marginal densities. J Am Statist Assoc 85:398–409

Geyer CJ (1992) Practical Markov chain Monte Carlo. Statist Sci 7(4):473–483, DOI https://doi.org/10.1214/ss/1177011137

Gibbs AL (2004) Convergence in the Wasserstein metric for Markov chain Monte Carlo algorithms with applications to image restoration. Stoch Models 20(4):473–492, DOI https://doi.org/10.1081/STM-200033117

Glynn PW, Meyn SP (1996) A Liapounov bound for solutions of the Poisson equation. Ann Probab 24(2):916–931

Gohberg I, Goldberg S (1981) Basic operator theory. Birkhäuser, Boston, Mass.

Gordin MI (1969) The central limit theorem for stationary processes. Dokl Akad Nauk SSSR 188:739–741

Gordin MI, Lifšic BA (1978) Central limit theorem for stationary Markov processes. Dokl Akad Nauk SSSR 239(4):766–767

Gourieroux C, Robert C (2006) Stochastic unit root models. Econometric Theory 22(6):1052–1090, DOI https://doi.org/10.1017/S0266466606060518

Graham C (2014) Markov chains. Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester; Dunod, Paris, DOI https://doi.org/10.1002/9781118881866, analytic and Monte Carlo computations

Granger CWJ, Swanson NR (1997) An introduction to stochastic unit-root processes. J Econometrics 80(1):35–62, DOI https://doi.org/10.1016/S0304-4076(96)00016-4

Griffeath D (1975) Uniform coupling of non-homogeneous Markov chains. J Appl Probability 12(4):753–762

Griffeath D (1978) Coupling methods for Markov processes. In: Studies in probability and ergodic theory, Adv. in Math. Suppl. Stud., vol 2, Academic Press, New York-London, pp 1–43

Guillin A (2001) Moderate deviations of inhomogeneous functionals of Markov processes and application to averaging. Stochastic Process Appl 92(2):287–313, DOI https://doi.org/10.1016/S0304-4149(00)00081-8

Guillin A, Léonard C, Wu L, Yao N (2009) Transportation-information inequalities for Markov processes. Probab Theory Related Fields 144(3–4):669–695, DOI https://doi.org/10.1007/s00440-008-0159-5

Guo M, Petruccelli JD (1991) On the null recurrence and transience of a first-order SETAR model. J Appl Probab 28(3):584–592

Gut A (2012) Anscombe's theorem 60 years later. Sequential Analysis 31(3):368–396, DOI https://doi.org/10.1080/07474946.2012.694349

Häggström O (2005) On the central limit theorem for geometrically ergodic Markov chains. Probab Theory Related Fields 132(1):74–82, DOI https://doi.org/10.1007/s00440-004-0390-7

Häggström O, Rosenthal JS (2007) On variance conditions for Markov chain CLTs. Electron Comm Probab 12:454–464 (electronic), DOI https://doi.org/10.1214/ECP.v12-1336

Hairer M, Mattingly JC (2011) Yet another look at Harris' ergodic theorem for Markov chains. In: Seminar on Stochastic Analysis, Random Fields and Applications VI, Progr. Probab., vol 63, Birkhäuser/Springer Basel AG, Basel, pp 109–117

Hairer M, Mattingly JC, Scheutzow M (2011) Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. Probab Theory Related Fields 149(1–2):223–259

Hall P, Heyde CC (1980) Martingale Limit Theory and its Application. Academic Press, New York, London

Hall P, Heyde CC (1981) Rates of convergence in the martingale central limit theorem. Ann Probab 9(3):395–404

Harris TE (1956) The existence of stationary measures for certain Markov processes. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. II, University of California Press, Berkeley and Los Angeles, pp 113–124

Hennion H, Hervé L (2001) Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness, Lecture Notes in Mathematics, vol 1766. Springer, Berlin, DOI https://doi.org/10.1007/b87874

Hernández-Lerma O, Lasserre JB (2003) Markov chains and invariant probabilities, Progress in Mathematics, vol 211. Birkhäuser Verlag, Basel

Hervé L, Ledoux J (2014a) Approximating Markov chains and V-geometric ergodicity via weak perturbation theory. Stochastic Process Appl 124(1):613–638, DOI https://doi.org/10.1016/j.spa.2013.09.003

Hervé L, Ledoux J (2014b) Spectral analysis of Markov kernels and application to the convergence rate of discrete random walks. Adv in Appl Probab 46(4):1036–1058, DOI https://doi.org/10.1239/aap/1418396242

Hervé L, Ledoux J (2016) A computable bound of the essential spectral radius of finite range Metropolis-Hastings kernels. Statist Probab Lett 117:72–79, DOI https://doi.org/10.1016/j.spl.2016.05.007

Hobert JP, Geyer CJ (1998) Geometric ergodicity of Gibbs and block Gibbs samplers for a hierarchical random effects model. J Multivariate Anal 67(2):414–430, DOI https://doi.org/10.1006/jmva.1998.1778

Hobert JP, Jones G, Presnell B, Rosenthal JS (2002) On the applicability of regenerative simulation in Markov chain Monte Carlo. Biometrika 89(4):731–743

Hoeffding W (1963) Probability inequalities for sums of bounded random variables. J Am Statist Assoc 58(301):13–30

Holmes PT (1967) On non-dissipative Markov chains. Sankhyā Ser A 29:383–390

Hu Q, Yue W (2008) Markov decision processes with their applications, Advances in Mechanics and Mathematics, vol 14. Springer, New York

Huang J, Kontoyiannis I, Meyn SP (2002) The ODE method and spectral theory of Markov operators. In: Stochastic theory and control (Lawrence, KS, 2001), Lect. Notes Control Inf. Sci., vol 280, Springer, Berlin, pp 205–221, DOI https://doi.org/10.1007/3-540-48022-6_15

Ibragimov IA (1959) Some limit theorems for stochastic processes stationary in the strict sense. Dokl Akad Nauk SSSR 125:711–714

Ibragimov IA (1963) A central limit theorem for a class of dependent random variables. Teor Verojatnost i Primenen 8:89–94

Ibragimov IA, Linnik YV (1971) Independent and stationary sequences of random variables. Wolters-Noordhoff Publishing, Groningen, with a supplementary chapter by I. A. Ibragimov and V. V. Petrov, Translation from the Russian edited by J. F. C. Kingman

Jain N, Jamison B (1967) Contributions to Doeblin's theory of Markov processes. Z Wahrsch Verw Geb 8:19–40

Jarner S, Hansen E (2000) Geometric ergodicity of Metropolis algorithms. Stoch Process Appl 85:341–361

Jarner SF, Roberts GO (2002) Polynomial convergence rates of Markov chains. Ann Appl Probab 12(1):224–247

Jarner SF, Yuen WK (2004) Conductance bounds on the L^2 convergence rate of Metropolis algorithms on unbounded state spaces. Adv in Appl Probab 36(1):243–266, DOI https://doi.org/10.1239/aap/1077134472

Jones G, Hobert J (2001) Honest exploration of intractable probability distributions via Markov chain Monte Carlo. Statist Sci 16(4):312–334, DOI https://doi.org/10.1214/ss/1015346317

Jones GL (2004) On the Markov chain central limit theorem. Probab Surv 1:299–320, DOI https://doi.org/10.1214/154957804100000051

Joulin A, Ollivier Y (2010) Curvature, concentration and error estimates for Markov chain Monte Carlo. Ann Probab 38(6):2418–2442, DOI https://doi.org/10.1214/10-AOP541

Kalashnikov VV (1968) The use of Lyapunov's method in the solution of queueing theory problems. Izv Akad Nauk SSSR Tehn Kibernet, pp 89–95

Kalashnikov VV (1971) Analysis of ergodicity of queueing systems by means of the direct method of Lyapunov. Avtomat i Telemeh 32(4):46–54

Kalashnikov VV (1977) Analysis of stability in queueing problems by a method of trial functions. Teor Verojatnost i Primenen 22(1):89–105

Kallenberg O (2002) Foundations of modern probability, 2nd edn. Probability and its Applications (New York), Springer, New York, DOI https://doi.org/10.1007/978-1-4757-4015-8

Kannan R, Lovász L, Simonovits M (1995) Isoperimetric problems for convex bodies and a localization lemma. Discrete Comput Geom 13(3-4):541–559, DOI https://doi.org/10.1007/BF02574061

Kartashiov N (1996) Stong stable Markov Chains. VSP International publisher

Kemeny J, Snell JL (1961a) Potentials for denumerable Markov chains. J Math Anal Appl 3:196–260, DOI https://doi.org/10.1016/0022-247X(61)90054-3

Kemeny JG, Snell JL (1961b) On Markov chain potentials. Ann Math Statist 32:709–715, DOI https://doi.org/10.1214/aoms/1177704966

Kemeny JG, Snell JL (1963) Boundary theory for recurrent Markov chains. Trans Amer Math Soc 106:495–520, DOI https://doi.org/10.2307/1993756

Kemeny JG, Snell JL, Knapp AW (1976) Denumerable Markov chains, 2nd edn. Springer, New York-Heidelberg-Berlin, with a chapter on Markov random fields, by David Griffeath, Graduate Texts in Mathematics, No. 40

Kendall DG (1959) Unitary dilations of Markov transition operators, and the corresponding integral representations for transition-probability matrices. In: Probability and statistics: The Harald Cramér volume (edited by Ulf Grenander), Almqvist & Wiksell, Stockholm; John Wiley & Sons, New York, pp 139–161

Kendall DG (1960) Geometric ergodicity and theory of queues. In: Mathematical methods in the social sciences, Stanford Univ. Press, Stanford, Calif., pp 176–195

Kipnis C, Varadhan SRS (1985) Central limit theorems for additive functionals of reversible Markov chains and applications. Astérisque 132:65–70, colloquium in honor of Laurent Schwartz, vol 2 (Palaiseau, 1983)

Kipnis C, Varadhan SRS (1986) Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. Comm Math Phys 104(1):1–19, URL http://projecteuclid.org/euclid.cmp/1104114929

Klokov SA, Veretennikov AY (2004a) On the sub-exponential mixing rate for a class of Markov diffusions. J Math Sci (N Y) 123(1):3816–3823, DOI https://doi.org/10.1023/B:JOTH.0000036322.50269.39

Klokov SA, Veretennikov AY (2004b) Sub-exponential mixing rate for a class of Markov chains. Math Commun 9(1):9–26

Kolmogorov A (1931) über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. Math Ann 104(1):415–458, DOI https://doi.org/10.1007/BF01457949

Kontoyiannis I, Meyn SP (2003) Spectral theory and limit theorems for geometrically ergodic Markov processes. Ann Appl Probab 13(1):304–362, DOI https://doi.org/10.1214/aoap/1042765670

Kontoyiannis I, Meyn SP (2005) Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. Electron J Probab 10(3):61–123 (electronic)

Kontoyiannis I, Meyn SP (2012) Geometric ergodicity and the spectral gap of non-reversible Markov chains. Probab Theory Related Fields 154(1–2):327–339, DOI https://doi.org/10.1007/s00440-011-0373-4

Lawler GF, Sokal AD (1988) Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. Trans Amer Math Soc 309(2):557–580, DOI https://doi.org/10.2307/2000925

Ledoux M (2001) The concentration of measure phenomenon, Mathematical Surveys and Monographs, vol 89. American Mathematical Society, Providence, RI

Lerner N (2014) A course on integration theory. Birkhäuser/Springer, Basel, including more than 150 exercises with detailed answers

Levin DA, Peres Y, Wilmer EL (2009) Markov chains and mixing times. American Mathematical Society, Providence, RI, with a chapter by James G. Propp and David B. Wilson

Lezaud P (1998) Chernoff-type bound for finite Markov chains. Ann Appl Probab 8(3):849–867, DOI https://doi.org/10.1214/aoap/1028903453

Lin M (1970) Conservative Markov processes on a topological space. Israel J Math 8:165–186, DOI https://doi.org/10.1007/BF02771312

Lin M (1971) Mixing for Markov operators. Z Wahrscheinlichkeitstheorie und Verw Gebiete 19:231–242, DOI https://doi.org/10.1007/BF00534111

Lindvall T (1977) A probabilistic proof of Blackwell's renewal theorem. Ann Probability 5(3):482–485

Lindvall T (1979) On coupling of discrete renewal sequences. Z Wahrsch Verw Gebiete 48(1):57–70

Lindvall T (1992) Lectures on the Coupling Method. Wiley, New-York

Liu JS (1996) Metropolized independent sampling with comparisons to rejection sampling and importance sampling. Stat Comput 6:113–119

Lovász L, Simonovits M (1993) Random walks in a convex body and an improved volume algorithm. Random Structures Algorithms 4(4):359–412, DOI https://doi.org/10.1002/rsa.3240040402

Lund RB, Tweedie RL (1996) Geometric convergence rates for stochastically ordered Markov chains. Mathematics of Operation Research 21:182–194

Lund RB, Meyn SP, Tweedie R (1996) Computable exponential convergence rates for stochastically ordered Markov processes. Annals of Applied Probability 6:218–237

Madras N, Sezer D (2010) Quantitative bounds for Markov chain convergence: Wasserstein and total variation distances. Bernoulli 16(3):882–908, DOI https://doi.org/10.3150/09-BEJ238

Madsen RW (1971) A note on some ergodic theorems of A. Paz. Ann Math Statist 42:405–408, DOI https://doi.org/10.1214/aoms/1177693534

Madsen RW, Isaacson DL (1973) Strongly ergodic behavior for non-stationary Markov processes. Ann Probability 1:329–335

Maigret N (1978) Théorème de limite centrale fonctionnel pour une chaîne de Markov récurrente au sens de Harris et positive. Ann Inst H Poincaré Sect B (NS) 14(4):425–440 (1979)

Malyshkin MN (2000) Subexponential estimates for the rate of convergence to the invariant measure for stochastic differential equations. Teor Veroyatnost i Primenen 45(3):489–504, DOI https://doi.org/10.1137/S0040585X97978403

Markov AA (1910) Recherches sur un cas remarquable d'épreuves dépendantes. Acta Math 33(1):87–104, DOI https://doi.org/10.1007/BF02393213

Maxwell M, Woodroofe M (2000) Central limit theorems for additive functionals of Markov chains. The Annals of Probability 28(2):713–724

McDiarmid C (1989) On the method of bounded differences. In: Surveys in combinatorics, 1989 (Norwich, 1989), London Math. Soc. Lecture Note Ser., vol 141, Cambridge Univ. Press, Cambridge, pp 148–188

Mengersen K, Tweedie RL (1996) Rates of convergence of the Hastings and Metropolis algorithms. Ann Statist 24:101–121

Mertens J, Samuel-Cahn E, Zamir S (1978) Necessary and sufficient conditions for recurrence and transience of Markov chains, in terms of inequalities. J Appl Probab 15(4):848–851

Meyn SP (2006) Control Techniques for Complex Networks. Cambridge University Press, Cambridge. 562 pp. ISBN 978-0-521-88441-9

Meyn SP, Tweedie RL (1992) Stability of Markovian processes. I. Criteria for discrete-time chains. Adv in Appl Probab 24(3):542–574, DOI https://doi.org/10.2307/1427479

Meyn SP, Tweedie RL (1993a) The Doeblin decomposition. In: Doeblin and modern probability (Blaubeuren, 1991), Contemp. Math., vol 149, Amer. Math. Soc., Providence, RI, pp 211–225, DOI https://doi.org/10.1090/conm/149/01272

Meyn SP, Tweedie RL (1993b) Markov Chains and Stochastic Stability. Springer, London

Meyn SP, Tweedie RL (1994) Computable bounds for convergence rates of Markov chains. Annals of Applied Probability 4:981–1011

Meyn SP, Tweedie RL (2009) Markov Chains and Stochastic Stability. Cambridge University Press, London

Miller HD (1965/1966) Geometric ergodicity in a class of denumerable Markov chains. Z Wahrscheinlichkeitstheorie und Verw Gebiete 4:354–373

Nagaev SV (1957) Some limit theorems for stationary Markov chains. Teor Veroyatnost i Primenen 2:389–416

Neveu J (1964) Chaînes de Markov et théorie du potentiel. Annales Scientifiques de l'Université de Clermont-Ferrand 2(24):37–89

Neveu J (1972) Potentiel Markovien récurrent des chaînes de Harris. Ann Inst Fourier (Grenoble) 22(2):85–130, URL http://www.numdam.org/item?id=AIF_1972__22_2_85_0

Neveu J (1975) Discrete-Time Martingales. North-Holland

Norris JR (1998) Markov chains, Cambridge Series in Statistical and Probabilistic Mathematics, vol 2. Cambridge University Press, Cambridge, reprint of 1997 original

Nummelin E (1978) A splitting technique for Harris recurrent Markov chains. Z Wahrscheinlichkeitstheorie und Verw Gebiete 4:309–318

Nummelin E (1984) General Irreducible Markov Chains and Non-Negative Operators. Cambridge University Press

Nummelin E (1991) Renewal representations for Markov operators. Adv Math 90(1):15–46, DOI https://doi.org/10.1016/0001-8708(91)90018-3

Nummelin E (1997) On distributionally regenerative Markov chains. Stochastic Process Appl 72(2):241–264, DOI https://doi.org/10.1016/S0304-4149(97)00088-4

Nummelin E, Tuominen P (1982) Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory. Stochastic Process Appl 12(2):187–202, DOI https://doi.org/10.1016/0304-4149(82)90041-2

Nummelin E, Tuominen P (1983) The rate of convergence in Orey's theorem for Harris recurrent Markov chains with applications to renewal theory. Stochastic Processes and Their Applications 15:295–311

Nummelin E, Tweedie RL (1976) Geometric ergodicity for a class of Markov chains. Ann Sci Univ Clermont 61(Math. No. 14):145–154, École d'Été de Calcul des Probabilités de Saint-Flour (Saint-Flour, 1976)

Nummelin E, Tweedie RL (1978) Geometric ergodicity and R-positivity for general Markov chains. Ann Probability 6(3):404–420

Ollivier Y (2009) Ricci curvature of Markov chains on metric spaces. J Funct Anal 256(3):810–864, DOI https://doi.org/10.1016/j.jfa.2008.11.001

Ollivier Y (2010) A survey of Ricci curvature for metric spaces and Markov chains. In: Probabilistic approach to geometry, Adv. Stud. Pure Math., vol 57, Math. Soc. Japan, Tokyo, pp 343–381

Orey S (1959) Recurrent Markov chains. Pacific J Math 9:805–827, URL http://projecteuclid.org/euclid.pjm/1103039121

Orey S (1962) An ergodic theorem for Markov chains. Z Wahrscheinlichkeitstheorie Verw Gebiete 1:174–176, DOI https://doi.org/10.1007/BF01844420

Orey S (1964) Potential kernels for recurrent Markov chains. J Math Anal Appl 8:104–132, DOI https://doi.org/10.1016/0022-247X(64)90088-5

Orey S (1971) Lecture Notes on Limit Theorems for Markov Chain Transition Probabilities. Springer

Pakes AG (1969) Some conditions for ergodicity and recurrence of Markov chains. Operations Res 17:1058–1061

Parthasarathy KR (1967) Probability measures on metric spaces. Probability and Mathematical Statistics, No. 3, Academic Press, Inc., New York-London

Petruccelli JD, Woolford SW (1984) A threshold AR(1) model. J Appl Probab 21(2):270–286

Pitman JW (1974) Uniform rates of convergence for Markov chain transition probabilities. Z Wahrscheinlichkeitstheorie und Verw Gebiete 29:193–227, DOI https://doi.org/10.1007/BF00536280

Popov NN (1977) Geometric ergodicity conditions for countable Markov chains. Dokl Akad Nauk SSSR 234(2):316–319

Popov NN (1979) Geometric ergodicity of Markov chains with an arbitrary state space. Dokl Akad Nauk SSSR 247(4):798–802

Port SC (1965) Ratio limit theorems for Markov chains. Pacific J Math 15:989–1017, URL http://projecteuclid.org/euclid.pjm/1102995584

Privault N (2008) Potential Theory in Classical Probability, Springer, Berlin, Heidelberg, pp 3–59. DOI https://doi.org/10.1007/978-3-540-69365-9_2

Privault N (2013) Understanding Markov chains. Springer Undergraduate Mathematics Series, Springer Singapore, Singapore, DOI https://doi.org/10.1007/978-981-4451-51-2, examples and applications

Rachev ST, Rüschendorf L (1998) Mass transportation problems. Vol. I. Probability and its Applications (New York), Springer, New York, theory

Rényi A (1957) On the asymptotic distribution of the sum of a random number of independent random variables. Acta Math Acad Sci Hungar 8:193–199, DOI https://doi.org/10.1007/BF02025242

Revuz D (1975) Markov chains. North-Holland Publishing Co., Amsterdam-Oxford; American Elsevier Publishing Co., Inc., New York, North-Holland Mathematical Library, Vol. 11

Revuz D (1984) Markov Chains, 2nd edn. North-Holland Publishing, Amsterdam

Rio E (1993) Covariance inequalities for strongly mixing processes. Ann Inst H Poincaré Probab Statist 29(4):587–597, URL http://www.numdam.org/item?id=AIHPB_1993__29_4_587_0

Rio E (1994) Inégalités de moments pour les suites stationnaires et fortement mélangeantes. C R Acad Sci Paris Sér I Math 318(4):355–360

Rio E (2000a) Inégalités de hoeffding pour des fonctions lipshitziennes de suites dépendantes. Comptes Rendus de l'Académie des Sciences pp 905–908

Rio E (2000b) Théorie asymptotique des processus aléatoires faiblement dépendants, Mathématiques & Applications (Berlin) [Mathematics & Applications], vol 31. Springer, Berlin

Rio E (2017) Asymptotic theory of weakly dependent random processes, Probability Theory and Stochastic Modelling, vol 80. Springer, Berlin, DOI https://doi.org/10.1007/978-3-662-54323-8, translated from the 2000 French edition [MR2117923]

Robert CP, Casella G (2004) Monte Carlo Statistical Methods, 2nd edn. Springer, New York

Robert CP, Casella G (2010) Introducing Monte Carlo methods with R. Use R!, Springer, New York, DOI https://doi.org/10.1007/978-1-4419-1576-4

Roberts GO, Rosenthal JS (1997) Geometric ergodicity and hybrid Markov chains. Electron Comm Probab 2(2):13–25, DOI https://doi.org/10.1214/ECP.v2-981

Roberts GO, Rosenthal JS (1998) Markov chain Monte Carlo: Some practical implications of theoretical results. Canad J Statist 26:5–32

Roberts GO, Rosenthal JS (2004) General state space Markov chains and MCMC algorithms. Probab Surv 1:20–71

Roberts GO, Rosenthal JS (2011) Quantitative non-geometric convergence bounds for independence samplers. Methodol Comput Appl Probab 13(2):391–403, DOI https://doi.org/10.1007/s11009-009-9157-z

Roberts GO, Tweedie RL (1996) Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. Biometrika 83(1):95–110

Roberts GO, Tweedie RL (1999) Bounds on regeneration times and convergence rates for Markov chains. Stochastic Processes and Their Applications 80:211–229

Roberts GO, Tweedie RL (2001) Geometric L^2 and L^1 convergence are equivalent for reversible Markov chains. J Appl Probab 38A:37–41, DOI https://doi.org/10.1239/jap/1085496589, probability, statistics and seismology

Rosenthal JS (1995a) Convergence rates for Markov chains. SIAM Rev 37(3):387–405, DOI https://doi.org/10.1137/1037083

Rosenthal JS (1995b) Minorization conditions and convergence rates for Markov chain Monte Carlo. J Amer Statist Assoc 90(430):558–566

Rosenthal JS (2001) A review of asymptotic convergence for general state space Markov chains. Far East J Theor Stat 5(1):37–50

Rosenthal JS (2002) Quantitative convergence rates of Markov chains: a simple account. Electron Comm Probab 7:123–128, DOI https://doi.org/10.1214/ECP.v7-1054

Rosenthal JS (2009) Markov chain Monte Carlo algorithms: theory and practice. In: Monte Carlo and quasi-Monte Carlo methods 2008, Springer, Berlin, pp 157–169, DOI https://doi.org/10.1007/978-3-642-04107-5_9

Rosenthal JS (2017) Simple confidence intervals for MCMC without CLTs. Electron J Stat 11(1):211–214, DOI https://doi.org/10.1214/17-EJS1224

Royden HL (1988) Real analysis, 3rd edn. Macmillan Publishing Company, New York

Rubino G, Sericola B (2014) Markov chains and dependability theory. Cambridge University Press, Cambridge, DOI https://doi.org/10.1017/CBO9781139051705

Rudin W (1987) Real and complex analysis, 3rd edn. McGraw-Hill Book Co., New York

Rudin W (1991) Functional analysis, 2nd edn. International Series in Pure and Applied Mathematics, McGraw-Hill, Inc., New York

Rudolf D (2009) Explicit error bounds for lazy reversible Markov chain Monte Carlo. J Complexity 25(1):11–24, DOI https://doi.org/10.1016/j.jco.2008.05.005

Rudolf D (2010) Error bounds for computing the expectation by Markov chain Monte Carlo. Monte Carlo Methods Appl 16(3–4):323–342, DOI https://doi.org/10.1515/MCMA.2010.012

Rudolf D (2012) Explicit error bounds for Markov chain Monte Carlo. Dissertationes Math (Rozprawy Mat) 485:1–93, DOI https://doi.org/10.4064/dm485-0-1

Rudolf D, Schweizer N (2015) Error bounds of MCMC for functions with unbounded stationary variance. Statist Probab Lett 99:6–12, DOI https://doi.org/10.1016/j.spl.2014.07.035

Saksman E, Vihola M (2010) On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. Ann Appl Probab 20(6):2178–2203, DOI https://doi.org/10.1214/10-AAP682

Samson PM (2000) Concentration of measure inequalities for Markov chains and Φ-mixing processes. Ann Probab 28(1):416–461, DOI https://doi.org/10.1214/aop/1019160125

Seneta E (1981) Non-negative Matrices and Markov Chains, 2nd edn. Springer Series in Statistics, Springer, New York, DOI https://doi.org/10.1007/0-387-32792-4

Sennot LI, Humblet PA, Tweedie RL (1983) Technical note—mean drifts and the non-ergodicity of Markov chains. Operations Research 31(4):783–789

Sericola B (2013) Markov chains. Applied Stochastic Methods Series, ISTE, London; John Wiley & Sons, Inc., Hoboken, NJ, DOI https://doi.org/10.1002/9781118731543, theory, algorithms and applications

Simon B (2015) Operator theory. A Comprehensive Course in Analysis, Part 4, American Mathematical Society, Providence, RI, DOI https://doi.org/10.1090/simon/004

Smith AFM, Roberts GO (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. J Roy Statist Soc Ser B 55(1):3–23, URL http://links.jstor.org/sici?sici=0035-9246(1993)55:1⟨3:BCVTGS⟩2.0.CO;2-#&origin=MSN

Tanikawa A (2001) Markov chains satisfying simple drift conditions for subgeometric ergodicity. Stoch Models 17(2):109–120, DOI https://doi.org/10.1081/STM-100002059

Taylor HM, Karlin S (1998) An introduction to stochastic modeling, 3rd edn. Academic Press, Inc., San Diego, CA

Taylor JC (1997) An introduction to measure and probability. Springer, New York, DOI https://doi.org/10.1007/978-1-4612-0659-0

Thorisson H (1987) A complete coupling proof of Blackwell's renewal theorem. Stochastic Process Appl 26(1):87–97, DOI https://doi.org/10.1016/0304-4149(87)90052-4

Thorisson H (2000) Coupling, Stationarity and Regeneration. Probability and its Applications, Springer, New-York

Tierney L (1994) Markov chains for exploring posterior disiributions (with discussion). Ann Statist 22(4):1701–1762

Tjostheim D (1990) Nonlinear time series and Markov chains. Adv in Appl Probab 22(3):587–611, DOI https://doi.org/10.2307/1427459

Tjøstheim D (1994) Non-linear time series: a selective review. Scand J Statist 21(2):97–130

Tong H (1990) Non-linear Time Series: A Dynamical System Approach. Oxford University Press

Tóth B (1986) Persistent random walks in random environment. Probab Theory Relat Fields 71(4):615–625, DOI https://doi.org/10.1007/BF00699043

Tóth B (2013) Comment on a theorem of M. Maxwell and M. Woodroofe. Electron Commun Probab 18(13):4, DOI https://doi.org/10.1214/ECP.v18-2366

Tuominen P (1976) Notes on 1-recurrent Markov chains. Z Wahrscheinlichkeitstheorie und Verw Gebiete 36(2):111–118, DOI https://doi.org/10.1007/BF00533994

Tuominen P, Tweedie R (1994) Subgeometric rates of convergence of f-ergodic Markov Chains. Advances in Applied Probability 26:775–798

Tweedie RL (1974a) R-theory for Markov chains on a general state space. I. Solidarity properties and R-recurrent chains. Ann Probability 2:840–864

Tweedie RL (1974b) R-theory for Markov chains on a general state space. II. r-subinvariant measures for r-transient chains. Ann Probability 2:865–878

Tweedie RL (1975) Relations between ergodicity and mean drift for Markov chains. Austral J Statist 17(2):96–102

Vere-Jones D (1962) Geometric ergodicity in denumerable Markov chains. Quart J Math Oxford Ser (2) 13:7–28

Veretennikov A (1997) On polynomial mixing bounds for stochastic differential equations. Stochastic Process Appl 70:115–127

Veretennikov A (1999) On polynomial mixing and the rate of convergence for stochastic differential and difference equations. Theory of probability and its applications pp 361–374

Veretennikov AY, Klokov SA (2004) On the subexponential rate of mixing for Markov processes. Teor Veroyatn Primen 49(1):21–35, DOI https://doi.org/10.1137/S0040585X97980841

Villani C (2009) Optimal transport, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol 338. Springer, Berlin, old and new

Walters P (1982) An introduction to ergodic theory, Graduate Texts in Mathematics, vol 79. Springer, New York-Berlin

Williams D (1991) Probability with martingales. Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, DOI https://doi.org/10.1017/CBO9780511813658

Wu WB, Woodroofe M (2004) Martingale approximations for sums of stationary processes. Ann Probab 32(2):1674–1690, DOI https://doi.org/10.1214/009117904000000351

Yuen WK (2000) Applications of geometric bounds to the convergence rate of Markov chains on $\mathbf{R}^n$. Stochastic Process Appl 87(1):1–23, DOI https://doi.org/10.1016/S0304-4149(99)00101-5

Yuen WK (2001) Application of geometric bounds to convergence rates of Markov chains and Markov processes on R(n). ProQuest LLC, Ann Arbor, MI, URL http://gateway.proquest.com/openurl?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&res_dat=xri:pqdiss&rft_dat=xri:pqdiss:NQ58619, thesis (Ph.D.)–University of Toronto (Canada)

Yuen WK (2002) Generalization of discrete-time geometric bounds to convergence rate of Markov processes on $\mathbb{R}^n$. Stoch Models 18(2):301–331, DOI https://doi.org/10.1081/STM-120004469

Index

Symbols
$(\mathsf{c}, m, \varepsilon)$-contracting set, 471
C_+, 67
K_a, 11
Q_C, 63
$S_P(f, r)$, 370
V-norm, 425, 635
V-oscillation, 635
k_P, 545
$\mathrm{k}_P(A)$, 545
Λ_0, 290
Λ_1, 290
Λ_2, 290
$\mathbb{P}_*$, 57
$|\cdot|_V$, 635
$||\cdot||_f$, 305, 635
$\mathrm{Abs.Gap}_{\mathrm{L}^2(\pi)}(P)$, 533
$\mathscr{X}_P^+$, 66, 192
X_P^+, 150
$\alpha(\mathscr{A}, \mathscr{B})$, 645
$\beta(\mathscr{A}, \mathscr{B})$, 645
$\Delta_{\mathsf{d},p}(P)$, 465
$\Delta_{\mathsf{c}}(P)$, 465
$\Delta(P)$, 403
$\mathscr{X}_C$, 63
$\stackrel{\mathrm{w}^*}{\Rightarrow}$, 625
$\stackrel{\mathrm{w}}{\Rightarrow}$, 625
$\mathrm{L}^2{}_0(\pi)$, 489
$\mathsf{BL}(\mathsf{H})$, 524
$\mathbb{F}_+(\mathsf{Y})$, 8
$\mathbb{F}_b(\mathsf{X})$, 4
$\mathrm{C}_b(\mathsf{X})$, 611
$\mathrm{C}_0(\mathsf{X})$, 611
$\mathrm{C}_c(\mathsf{X})$, 611
$\mathbb{F}(\mathsf{Y})$, 7
$\mathrm{Gap}_{\mathrm{L}^2(\pi)}(P)$, 545
$\mathrm{L}^p(\pi)$, 20
$||\cdot||_{\mathrm{L}^p(\pi)}$, 20
$||\cdot||_{\mathrm{L}^\infty(\pi)}$, 20
$\mathrm{Lip}_{\mathsf{d}}(\mathsf{X}^n, \gamma_0^{n-1})$, 595
$\bar{\Lambda}_0$, 290
$\bar{\Lambda}_1$, 290
$\bar{\Lambda}_2$, 290
$\mathbb{BD}(\mathsf{X}^n, \gamma_0^{n-1})$, 576
$\mathscr{C}(\xi, \xi')$, 422
$\mathscr{S}$, 290
$\mathbb{M}_+(\mathscr{X})$, 8
$\mathbb{M}_0(\mathscr{X})$, 632
$\mathbb{M}_1^*(\mathbb{N})$, 206
$\mathbb{M}_1(\mathscr{X})$, 19
$\mathbb{M}_V(\mathscr{X})$, 635
$\mathbb{M}_{\mathbb{C}}(\mathscr{X})$, 524
$\mathbb{M}_\pm(\mathscr{X})$, 629
μ_C^0, 68
μ_C^1, 68
$\mathrm{osc}_V(\cdot)$, 635
$\mathrm{osc}(f)$, 632
$\phi(\mathscr{A}, \mathscr{B})$, 646
π_C^0, 68
ρ_{d}, 627
$\rho(\mathscr{A}, \mathscr{B})$, 646
$\sigma_\pi^2(h)$, 554
$\mathbb{S}(\mathsf{X}, d)$, 461
$\mathbb{S}_p(\mathsf{X}, d)$, 460
$\mathrm{d_{TV}}(\cdot,\cdot)$, 631
$||\cdot||_{\mathrm{TV}}$, 631
$\mathbf{W}_{\mathsf{d}}$, 460
$\mathbf{W}_{\mathsf{d},p}$, 486
$\mathbf{W}_{\mathsf{c}}$, 456
r^0, 289

R. Douc et al., *Markov Chains*, Springer Series in Operations Research and Financial Engineering, https://doi.org/10.1007/978-3-319-97704-1

$\mathrm{D_g}(V,\lambda,b,C)$, 316
$\mathrm{D_g}(V,\lambda,b)$, 316
$\mathrm{D_{sg}}(\{V_n\},f,r,b,C)$, 362
$\mathrm{D_{sg}}(V,\phi,b,C)$, 364
$\mathrm{L}^2(\pi)$-absolute spectral gap, 533, 535–538, 541, 542, 544, 550
$\mathrm{L}^2(\pi)$-exponential convergence, 532
$\mathrm{L}^2(\pi)$-geometric ergodicity, 532, 533, 536, 543
$\mathrm{L}^\infty(\pi)$-exponential convergence, 540
$\mathrm{L}^p(\pi)$-exponential convergence, 538, 539

A
adjoint operator, 527, 531, 539, 552, 569, 571
analytic function, 567
aperiodicity, 128, 155, 156, 165, 173, 176, 178, 179, 185, 187, 202, 205, 210, 211, 228, 235, 245, 251, 262, 298, 300, 302, 306, 328, 331, 341, 344, 345, 350, 354, 377, 380, 387, 390, 392, 397, 407, 497, 501, 543, 587
  strong, 202
asymptotic σ -field, 260
atom, 119
  aperiodic, 126, 298, 300, 302, 306
  null recurrent, 128, 137
  positive, 128, 137
  recurrent, 121, 122, 124, 129
  transient, 121, 122, 124

B
Birkhoff's ergodic theorem, 100, 104, 108
Blackwell's theorem, 172, 178
bounded difference, 576

C
canonical filtration, 54
canonical process, 54
central limit theorem, 498, 500, 501, 504, 506, 508, 512–516
  atomic, 138
Chapman–Kolmogorov equation, 10
Cheeger's inequality, 546
Cheegers constant, 545
communication, 148
comparison theorem, 81
concentration inequality, 580, 584, 587, 593, 596, 598
conductance, 545
conjugate
  real numbers, 526
convergence
  weak*, 625
  weak, 625
coordinate process, 54
coupling
  distributional, 435, 438
  exact, 436, 440
  maximal distributional coupling, 437, 440
  of probability measures, 422
  of two kernels, 427
  optimal coupling for the Wasserstein distance, 456
  optimal coupling with respect to a cost function, 459
  optimal coupling for the V-norm, 425
  optimal coupling for the total variation distance, 422, 424
  successful, 435
  times, 435
coupling inequality, 156, 180, 291, 432, 436
cyclic decomposition, 204

D
data augmentation, 42
Dirichlet problem, 85
distributional coupling, 435
Dobrushin coefficient, 403
  V-Dobrushin coefficient, 410
  c-Dobrushin coefficient, 465
Doeblin set
  (m,ε)-Doeblin set, 406, 414
domain of attraction, 67
drift condition
  condition $\mathrm{D_g}(V,\lambda,b,C)$, 316
  condition $\mathrm{D_g}(V,\lambda,b)$, 316
  condition $\mathrm{D_{sg}}(\{V_n\},f,r,b,C)$, 362
  condition $\mathrm{D_{sg}}(V,\phi,b,C)$, 364
  geometric drift toward C, 316
dynamical system, 97
Dynkin formula, 90

E
eigenvalue, 524, 568
eigenvector, 524
ergodic dynamical system, 102, 104, 107, 109
ergodicity, 102
  f-geometric, 339
  geometric, 339
ergodicity geometric, 345, 347
event
  asymptotic, 260
  tail, 260
exact coupling, 436

F
first-entrance decomposition, 65

first-entrance last-exit decomposition, 64, 176
fixed-point theorem, 401, 402
functions of bounded difference, 576

G
gluing lemma, 621

H
Hahn–Jordan decomposition, 629
harmonic function, 75, 76, 232
hitting time, 59

I
infimum
  of two measures, 422
  of two kernels, 426
invariant
  event, 99
  measure, 16
  probability measure, 17
  random variable, 99
invariant probability measure, 104, 107, 108, 129, 200, 224, 255, 273, 275–277, 368, 376, 392, 405, 414, 444, 462, 466, 469, 474, 478

J
Jordan
  decomposition, 629
  set, 630

K
Kac formula, 71, 248, 249
Kendall's theorem, 173, 179
kernel
  (f,r)-ergodic, 385, 387
  (f,r)-regular, 370, 374, 376, 380
  T-kernel, 270, 271
  V uniformly ergodic, 349
  V uniformly geometrically ergodic, 349, 350, 412, 414, 441
  f-geometrically regular, 321, 324, 326, 331, 341
  aperiodic, 128, 150, 155, 156, 202, 205, 210, 211, 228, 235, 251, 262, 328, 331, 341, 344, 345, 350, 354, 377, 380, 387, 390, 392, 397, 407, 497, 501, 543, 587
  bounded, 6
  continuous component, 270
  coupling, 427, 428, 459
  density, 7
  Feller, 266, 269, 279
  geometrically ergodic, 345, 347
  geometrically uniformly ergodic, 354
  Harris recurrent, 229, 230, 232, 233
  homogeneous, 12
  induced, 63
  irreducible, 145, 194, 196, 200, 205, 233
  Markov, 6
  null, 250
  null-recurrent, 147
  optimal coupling, 428
  positive, 16, 147, 153, 250, 381
  potential, 77
  recurrent, 124, 146, 152, 221, 223, 224
  regular, 381
  resolvent, 11
  sampled, 11
  split, 241, 381
  strong Feller, 266
  strongly aperiodic, 202
  strongly irreducible, 145
  transient, 124, 146, 151, 222, 223, 227, 232
  uniformly ergodic, 349
  uniformly geometrically ergodic, 349, 406

L
last-exit decomposition, 65
Lyaponov function, 316

M
m-skeleton, 11
McDiarmid's inequality, 580, 584, 587
Markov chain
  canonical, 56
  homogeneous, 12
  order p, 15
  reversible, 18
  stationary, 57
Markov property, 61
martingale, 638
  difference, 638
  regular, 641
  submartingale, 638
  supermartingale, 638
maximum principle, 78, 120
measure
  (f,r)-regular, 370, 380
  f-geometrically regular, 321, 323, 326, 331
  image, 615
  inner regular, 617
  invariant, 16, 129, 249, 415
  irreducibility, 194–196

maximal irreducibility, 195, 200, 226, 249, 269, 415
outer regular, 617
Radon, 617
spread out, 280
subinvariant, 16
topological support, 616
measure invariant, 147
mixing coefficient
α, 645
β, 645
ϕ, 646
ρ, 646
Models
AR(p), 28, 281
AR(1), 28, 196
ARCH(p), 30
ARMA((p,q), 29
bilinear process, 470
birth and death chain, 92
DAR(1), 49
deterministic updating Gibbs sampler, 45
EGARCH, 36
FAR, 29, 257, 279, 352
Galton–Watson process, 141
gambler's ruin, 93
GARCH, 36
GARCH$(1,1)$, 50
hit-and-run algorithm, 48
Hit and Run sampler, 551
INAR process, 334
independent Metropolis–Hastings sampler, 40, 214, 355, 357, 394, 407, 549
Langevin diffusion, 41
log-Poisson autoregression, 37, 283, 481
Metropolis–Hastings algorithm, 39, 113, 213, 214, 236, 237, 283, 355, 356
Metropolis-Hastings algorithm, 212
observation-driven models, 35
random iterative functions, 27
random scan Gibbs sampler, 46
random walk, 28
random walk Metropolis algorithm, 40, 214, 318, 335, 353, 357
random walk on $\mathbb{R}^+$, 237
RCA, 32, 51
SETAR, 31
slice sampler, 44
TGARCH, 37
two-stage Gibbs sampler, 45, 358
vector autoregressive process, 28, 272, 273, 282
monotone class, 613

N
number of visits, 77

O
Observation-driven model, 35
operator
adjoint, 527
conductance, 545
positive, 550
self-adjoint, 570

P
period
of an accessible small set, 201
of an irreducible kernel, 202
of an atom, 126
of an atomic kernel, 128
periodicity classes, 204
petite set, 206
point
(f,r)-regular, 370, 374
f-geometrically regular, 321, 324
point spectrum, 568
Poisson–Dirichlet problem, 87, 89
time-inhomogeneous, 88
Poisson equation, 496, 498
Poisson problem, 85
Prohorov
metric, 627
theorem, 626
proper space, 568

R
random iterative functions, 27
random variable
asymptotic, 260
tail, 260
random walk
simple, 91
reachable point, 273, 278
recurrence
(f,r)-recurrence, 361
(f,r)-recurrent set, 361
f-geometric, 313
regular point, 566
renewal process, 165
aperiodic, 166
delay distribution, 166
delayed, 166, 167
epochs, 166
pure, 166, 167
renewals, 166
waiting time distribution, 166
zero-delayed, 166

resolvent, 524
  equation, 503
  kernel, 11
resolvent set, 566
return time, 59
reversibility, 18
Riesz-Thorin interpolation theorem, 564
Riesz decomposition, 89

S
self-adjoint on $L^2(\pi)$, 530
semi-continuous
  lower, 610
  upper, 610
separately Lipschitz functions, 594
sequences
  log-subbaditive, 290, 362
set
  (f,r)-recurrent, 361, 373, 387
  (f,r)-regular, 370, 373, 374, 380
  f-geometrically recurrent, 323
  f-geometrically regular, 321, 323, 324, 331
  absorbing, 17, 109
  accessible, 66, 192, 323, 373
  attractive, 67
  full, 198
  Harris recurrent, 67, 229
  maximal absorbing, 230
  petite, 323, 373
  recurrent, 124, 221
  transient, 124, 222
  uniformly transient, 124, 222
shift operator, 58
skeleton, 11
small set, 191
  Harris recurrent, 192
  positive, 192
  strongly aperiodic, 192
space
  locally compact metric, 611
  Polish, 610
  separable measurable, 612
spectral gap, 545
spectral measure, 573
spectral radius, 568
spectrum, 524
splitting construction, 241
stopping time, 59
strong Markov property, 62
subgeometric
  drift condition, 364
  ergodicity, 397, 444, 478
  sequences, 289, 366
superharmonic function, 75, 76, 233
support of a continuous function, 611

T
tail σ-field, 260
tightness, 626
Toeplitz lemma, 447
topological recurrence, 277
total variation
  f-norm, 305
  distance, 154, 423, 424, 631
  norm, 631
total variation of a measure, 629

U
uniform accessibility, 209
uniform Doeblin condition, 406
uniform integrability, 639

W
Wasserstein
  distance, 456, 457, 459, 460, 478, 486, 515, 516
  distance of order p, 460, 486
  space, 460
weak*-convergence, 625

Y
Young functions, 396