Markov Chains and Mixing Times
Second Edition

David A. Levin
University of Oregon

Yuval Peres
Microsoft Research

With contributions by
Elizabeth L. Wilmer

With a chapter on “Coupling from the Past” by
James G. Propp and David B. Wilson

American Mathematical Society
Providence, Rhode Island
FRONT COVER: The figure on the bottom left of the front cover, courtesy of David B. Wilson, is a uniformly random lozenge tiling of a hexagon (see Section 25.2). The figure on the bottom center, also from David B. Wilson, is a random sample of an Ising model at its critical temperature (see Sections 3.3.5 and 25.2) with mixed boundary conditions. The figure on the bottom right, courtesy of Eyal Lubetzky, is a portion of an expander graph (see Section 13.6).

For additional information and updates on this book, visit www.ams.org/bookpages/mbk-107
Contents

Preface ix
  Preface to the Second Edition ix
  Preface to the First Edition ix
  Overview xi
  For the Reader xii
  For the Instructor xiii
  For the Expert xiv

Acknowledgements xvi

Part I: Basic Methods and Examples 1

Chapter 1. Introduction to Finite Markov Chains 2
  1.1. Markov Chains 2
  1.2. Random Mapping Representation 5
  1.3. Irreducibility and Aperiodicity 7
  1.4. Random Walks on Graphs 8
  1.5. Stationary Distributions 9
  1.6. Reversibility and Time Reversals 13
  1.7. Classifying the States of a Markov Chain* 15
  Exercises 17
  Notes 19

Chapter 2. Classical (and Useful) Markov Chains 21
  2.1. Gambler’s Ruin 21
  2.2. Coupon Collecting 22
  2.3. The Hypercube and the Ehrenfest Urn Model 23
  2.4. The Pólya Urn Model 25
  2.5. Birth-and-Death Chains 26
  2.6. Random Walks on Groups 27
  2.7. Random Walks on $\mathbb{Z}$ and Reflection Principles 30
  Exercises 34
  Notes 35

Chapter 3. Markov Chain Monte Carlo: Metropolis and Glauber Chains 38
  3.1. Introduction 38
  3.2. Metropolis Chains 38
  3.3. Glauber Dynamics 41
  Exercises 45
  Notes 45
### Contents

**Chapter 20. Continuous-Time Chains**
- 20.1. Definitions
- 20.2. Continuous-Time Mixing
- 20.3. Spectral Gap
- 20.4. Product Chains
- Exercises
- Notes

**Chapter 21. Countable State Space Chains**
- 21.1. Recurrence and Transience
- 21.2. Infinite Networks
- 21.3. Positive Recurrence and Convergence
- 21.4. Null Recurrence and Convergence
- 21.5. Bounds on Return Probabilities
- Exercises
- Notes

**Chapter 22. Monotone Chains**
- 22.1. Introduction
- 22.2. Stochastic Domination
- 22.3. Definition and Examples of Monotone Markov Chains
- 22.4. Positive Correlations
- 22.5. The Second Eigenfunction
- 22.6. Censoring Inequality
- 22.7. Lower Bound on $\bar{d}$
- 22.8. Proof of Strassen’s Theorem
- Exercises
- Notes

**Chapter 23. The Exclusion Process**
- 23.1. Introduction
- 23.2. Mixing Time of $k$-Exclusion on the $n$-Path
- 23.3. Biased Exclusion
- Exercises
- Notes

**Chapter 24. Cesàro Mixing Time, Stationary Times, and Hitting Large Sets**
- 24.1. Introduction
- 24.2. Equivalence of $t_{\text{stop}}$, $t_{\text{Ces}}$, and $t_G$ for Reversible Chains
- 24.3. Halting States and Mean-Optimal Stopping Times
- 24.4. Regularity Properties of Geometric Mixing Times
- 24.5. Equivalence of $t_G$ and $t_H$
- 24.6. Upward Skip-Free Chains
- 24.7. $t_H(\alpha)$ Are Comparable for $\alpha \leq 1/2$
- 24.8. An Upper Bound on $t_{\text{rel}}$
- 24.9. Application to Robustness of Mixing
- Exercises
- Notes
Preface

Preface to the Second Edition

Since the publication of the first edition, the field of mixing times has continued to enjoy rapid expansion. In particular, many of the open problems posed in the first edition have been solved. The book has been used in courses at numerous universities, motivating us to update it.

In the eight years since the first edition appeared, we have made corrections and improvements throughout the book. We added three new chapters: Chapter 22 on monotone chains, Chapter 23 on the exclusion process, and Chapter 24, which relates mixing times and hitting time parameters to stationary stopping times. Chapter 4 now includes an introduction to mixing times in $\ell^p$, which reappear in Chapter 10. The latter chapter has several new topics, including estimates for hitting times on trees and Eulerian digraphs. A bound for cover times using spanning trees has been added to Chapter 11, which also now includes a general bound on cover times for regular graphs. The exposition in Chapter 6 and Chapter 17 now employs filtrations rather than relying on the random mapping representation. To reflect the key developments since the first edition, especially breakthroughs on the Ising model and the cutoff phenomenon, the Notes at the end of chapters and the open problems have been updated.

We thank the many careful readers who sent us comments and corrections: Anselm Adelmann, Amitabha Bagchi, Nathanael Berestycki, Olena Bormashenko, Krzysztof Burdzy, Gerandy Brito, Darcy Camargo, Varsha Dani, Sukhada Fadnavis, Tertuliano Franco, Alan Frieze, Reza Gheissari, Jonathan Hermon, Ander Holroyd, Kenneth Hu, John Jiang, Svante Janson, Melvin Kianmanesh Rad, Yin Tat Lee, Zhongyang Li, Eyal Lubetzky, Abbas Mehrabian, R. Misturini, L. Morgado, Asaf Nachmias, Fedja Nazarov, Joe Neeman, Ross Pinsky, Anthony Quas, Miklos Racz, Dinah Shender, N. J. A. Sloane, Jeff Steif, Izabella Stuhl, Jan Swart, Ryokichi Tanaka, Daniel Wu, and Zhen Zhu. We are particularly grateful to Daniel Jerison, Pawel Pralat, and Perla Sousi, who sent us long lists of insightful comments.

Preface to the First Edition

Markov first studied the stochastic processes that came to be named after him in 1906. Approximately a century later, there is an active and diverse interdisciplinary community of researchers using Markov chains in computer science, physics, statistics, bioinformatics, engineering, and many other areas.

The classical theory of Markov chains studied fixed chains, and the goal was to estimate the rate of convergence to stationarity of the distribution at time $t$, as $t \to \infty$. In the past two decades, as interest in chains with large state spaces has increased, a different asymptotic analysis has emerged. Some target distance to
the stationary distribution is prescribed; the number of steps required to reach this
target is called the mixing time of the chain. Now, the goal is to understand how
the mixing time grows as the size of the state space increases.

The modern theory of Markov chain mixing is the result of the convergence, in
the 1980s and 1990s, of several threads. (We mention only a few names here; see
the chapter Notes for references.)

For statistical physicists Markov chains become useful in Monte Carlo simu-
lation, especially for models on finite grids. The mixing time can determine the
running time for simulation. However, Markov chains are used not only for sim-
ulation and sampling purposes, but also as models of dynamical processes. Deep
connections were found between rapid mixing and spatial properties of spin systems,
e.g., by Dobrushin, Shlosman, Stroock, Zegarlinski, Martinelli, and Olivieri.

In theoretical computer science, Markov chains play a key role in sampling and
approximate counting algorithms. Often the goal was to prove that the mixing
time is polynomial in the logarithm of the state space size. (In this book, we are
generally interested in more precise asymptotics.)

At the same time, mathematicians including Aldous and Diaconis were inten-
sively studying card shuffling and other random walks on groups. Both spectral
methods and probabilistic techniques, such as coupling, played important roles.
Alon and Milman, Jerrum and Sinclair, and Lawler and Sokal elucidated the con-
nection between eigenvalues and expansion properties. Ingenious constructions of
"expander" graphs (on which random walks mix especially fast) were found using
probability, representation theory, and number theory.

In the 1990s there was substantial interaction between these communities, as
computer scientists studied spin systems and as ideas from physics were used for
sampling combinatorial structures. Using the geometry of the underlying graph to
find (or exclude) bottlenecks played a key role in many results.

There are many methods for determining the asymptotics of convergence to
stationarity as a function of the state space size and geometry. We hope to present
these exciting developments in an accessible way.

We will only give a taste of the applications to computer science and statistical
physics; our focus will be on the common underlying mathematics. The prerequi-
sites are all at the undergraduate level. We will draw primarily on probability and
linear algebra, but we will also use the theory of groups and tools from analysis
when appropriate.

Why should mathematicians study Markov chain convergence? First of all, it is
a lively and central part of modern probability theory. But there are ties to several
other mathematical areas as well. The behavior of the random walk on a graph
reveals features of the graph's geometry. Many phenomena that can be observed in
the setting of finite graphs also occur in differential geometry. Indeed, the two fields
enjoy active cross-fertilization, with ideas in each playing useful roles in the other.
Reversible finite Markov chains can be viewed as resistor networks; the resulting
discrete potential theory has strong connections with classical potential theory. It
is amusing to interpret random walks on the symmetric group as card shuffles—and
real shuffles have inspired some extremely serious mathematics—but these chains
are closely tied to core areas in algebraic combinatorics and representation theory.
In the spring of 2005, mixing times of finite Markov chains were a major theme of the multidisciplinary research program *Probability, Algorithms, and Statistical Physics*, held at the Mathematical Sciences Research Institute. We began work on this book there.

**Overview**

We have divided the book into two parts.

In **Part I**, the focus is on techniques, and the examples are illustrative and accessible. Chapter 1 defines Markov chains and develops the conditions necessary for the existence of a unique stationary distribution. Chapters 2 and 3 both cover examples. In Chapter 2 they are either classical or useful—and generally both; we include accounts of several chains, such as the gambler’s ruin and the coupon collector, that come up throughout probability. In Chapter 3 we discuss Glauber dynamics and the Metropolis algorithm in the context of “spin systems.” These chains are important in statistical mechanics and theoretical computer science.

Chapter 4 proves that, under mild conditions, Markov chains do, in fact, converge to their stationary distributions and defines total variation distance and mixing time, the key tools for quantifying that convergence. The techniques of Chapters 5, 6, and 7 on coupling, strong stationary times, and methods for lower bounding distance from stationarity, respectively, are central to the area.

In Chapter 8 we pause to examine card shuffling chains. Random walks on the symmetric group are an important mathematical area in their own right, but we hope that readers will appreciate a rich class of examples appearing at this stage in the exposition.

Chapter 9 describes the relationship between random walks on graphs and electrical networks, while Chapters 10 and 11 discuss hitting times and cover times.

In **Part II**, we cover more sophisticated techniques and present several detailed case studies of particular families of chains. Much of this material appears here for the first time in textbook form.

Chapter 13 covers advanced spectral techniques, including comparison of Dirichlet forms and Wilson’s method for lower bounding mixing.

Chapters 14 and 15 cover some of the most important families of “large” chains studied in computer science and statistical mechanics and some of the most important methods used in their analysis. Chapter 14 introduces the path coupling method, which is useful in both sampling and approximate counting. Chapter 15 looks at the Ising model on several different graphs, both above and below the critical temperature.

Chapter 16 revisits shuffling, looking at two examples—one with an application to genomics—whose analysis requires the spectral techniques of Chapter 13.

Chapter 17 begins with a brief introduction to martingales and then presents some applications of the evolving sets process.

Chapter 18 considers the cutoff phenomenon. For many families of chains where we can prove sharp upper and lower bounds on mixing time, the distance from stationarity drops from near 1 to near 0 over an interval asymptotically smaller than the mixing time. Understanding why cutoff is so common for families of interest is a central question.
Chapter 19 on lamplighter chains, brings together methods presented throughout the book. There are many bounds relating parameters of lamplighter chains to parameters of the original chain: for example, the mixing time of a lamplighter chain is of the same order as the cover time of the base chain.

Chapters 20 and 21 introduce two well-studied variants on finite discrete time Markov chains: continuous time chains and chains with countable state spaces. In both cases we draw connections with aspects of the mixing behavior of finite discrete-time Markov chains.

Chapter 25, written by Propp and Wilson, describes the remarkable construction of coupling from the past, which can provide exact samples from the stationary distribution.

Chapter 26 closes the book with a list of open problems connected to material covered in the book.

For the Reader

Starred sections, results, and chapters contain material that either digresses from the main subject matter of the book or is more sophisticated than what precedes them and may be omitted.

Exercises are found at the ends of chapters. Some (especially those whose results are applied in the text) have solutions at the back of the book. We of course encourage you to try them yourself first!

The Notes at the ends of chapters include references to original papers, suggestions for further reading, and occasionally “complements.” These generally contain related material not required elsewhere in the book—sharper versions of lemmas or results that require somewhat greater prerequisites.

The Notation Index at the end of the book lists many recurring symbols.

Much of the book is organized by method, rather than by example. The reader may notice that, in the course of illustrating techniques, we return again and again to certain families of chains—random walks on tori and hypercubes, simple card shuffles, proper colorings of graphs. In our defense we offer an anecdote.

In 1991 one of us (Y. Peres) arrived as a postdoc at Yale and visited Shizuo Kakutani, whose rather large office was full of books and papers, with bookcases and boxes from floor to ceiling. A narrow path led from the door to Kakutani’s desk, which was also overflowing with papers. Kakutani admitted that he sometimes had difficulty locating particular papers, but he proudly explained that he had found a way to solve the problem. He would make four or five copies of any really interesting paper and put them in different corners of the office. When searching, he would be sure to find at least one of the copies.

Cross-references in the text and the Index should help you track earlier occurrences of an example. You may also find the chapter dependency diagrams below useful.

We have included brief accounts of some background material in Appendix A. These are intended primarily to set terminology and notation, and we hope you will consult suitable textbooks for unfamiliar material.

Be aware that we occasionally write symbols representing a real number when an integer is required (see, e.g., the \( \binom{n}{\delta k} \)'s in the proof of Proposition 13.37).
hope the reader will realize that this omission of floor or ceiling brackets (and the details of analyzing the resulting perturbations) is in her or his best interest as much as it is in ours.

For the Instructor

The prerequisites this book demands are a first course in probability, linear algebra, and, inevitably, a certain degree of mathematical maturity. When introducing material which is standard in other undergraduate courses—e.g., groups—we provide definitions but often hope the reader has some prior experience with the concepts.

In Part I, we have worked hard to keep the material accessible and engaging for students. (Starred material is more sophisticated and is not required for what follows immediately; they can be omitted.)

Here are the dependencies among the chapters of Part I:

Chapters 1 through 7, shown in gray, form the core material, but there are several ways to proceed afterwards. Chapter 8 on shuffling gives an early rich application but is not required for the rest of Part I. A course with a probabilistic focus might cover Chapters 9, 10, and 11. To emphasize spectral methods and combinatorics, cover Chapters 8 and 12 and perhaps continue on to Chapters 13 and 16.

While our primary focus is on chains with finite state spaces run in discrete time, continuous-time and countable-state-space chains are both discussed—in Chapters 20 and 21, respectively.

We have also included Appendix B, an introduction to simulation methods, to help motivate the study of Markov chains for students with more applied interests. A course leaning towards theoretical computer science and/or statistical mechanics might start with Appendix B, cover the core material, and then move on to Chapters 14, 15, and 25.

Of course, depending on the interests of the instructor and the ambitions and abilities of the students, any of the material can be taught! Below we include a full diagram of dependencies of chapters. Its tangled nature results from the interconnectedness of the area: a given technique can be applied in many situations, while a particular problem may require several techniques for full analysis.
The logical dependencies of chapters. The core Chapters 1 through 7 are in dark gray, the rest of Part I is in light gray, and Part II is in white.

For the Expert

Several other recent books treat Markov chain mixing. Our account is more comprehensive than those of H"aggstr"om (2002), Jerrum (2003), or Montenegro and Tetali (2006), yet not as exhaustive as Aldous and Fill (1999). Norris (1998) gives an introduction to Markov chains and their applications but does not focus on mixing. Since this is a textbook, we have aimed for accessibility and comprehensibility, particularly in Part I.

What is different or novel in our approach to this material?

- Our approach is probabilistic whenever possible. We also integrate “classical” material on networks, hitting times, and cover times and demonstrate its usefulness for bounding mixing times.
- We provide an introduction to several major statistical mechanics models, most notably the Ising model, and collect results on them in one place.
– We give expository accounts of several modern techniques and examples, including evolving sets, the cutoff phenomenon, lamplighter chains, and the $L$-reversal chain.
– We systematically treat lower bounding techniques, including several applications of Wilson’s method.
– We use the transportation metric to unify our account of path coupling and draw connections with earlier history.
– We present an exposition of coupling from the past by Propp and Wilson, the originators of the method.
Acknowledgements

The authors thank the Mathematical Sciences Research Institute, the National Science Foundation VIGRE grant to the Department of Statistics at the University of California, Berkeley, and National Science Foundation grants DMS-0244479 and DMS-0104073 for support. We also thank Hugo Rossi for suggesting we embark on this project. Thanks to Blair Ahlquist, Tonci Antunovic, Elisa Celis, Paul Cuff, Jian Ding, Ori Gurel-Gurevich, Tom Hayes, Itamar Landau, Yun Long, Karola Mészáros, Shobhana Murali, Weiyang Ning, Tomoyuki Shirai, Walter Sun, Sivapran Vanniasegaram, and Ariel Yadin for corrections to an earlier version and making valuable suggestions. Yelena Shvets made the illustration in Section 6.5.4. The simulations of the Ising model in Chapter 15 are due to Raissa D’Souza. We thank László Lovász for useful discussions. We are indebted to Alistair Sinclair for his work co-organizing the M.S.R.I. program Probability, Algorithms, and Statistical Physics in 2005, where work on this book began. We thank Robert Calhoun for technical assistance.

Finally, we are greatly indebted to David Aldous and Persi Diaconis, who initiated the modern point of view on finite Markov chains and taught us much of what we know about the subject.
Bibliography

The pages on which a reference appears follow the symbol ↑.


425
BIBLIOGRAPHY


Diaconis, P. 1988a. Group representations in probability and statistics. 93
Diaconis, P. 2013. Some things we’ve learned (about Markov chain Monte Carlo), Bernoulli 19, no. 4, 1294–1305. 372
Ding, J., J. R. Lee, and Y. Peres. 2012. Cover times, blanket times, and majorizing measures, Ann. of Math. (2) 175, no. 3, 1409–1471. 168
Dubins, L. E. 1968.
at arxiv:1107.2612
Doob, J. L. 1953.
Fini d’états
Doeblin, W. 1938.
Einstein, A. 1934.
On the method of theoretical physics, Philosophy of Science 1, no. 2, 163–169.


Hermon, J., H. Lacoin, and Y. Peres. 2016. *Total variation and separation cutoffs are not equivalent and neither one implies the other*, Electron. J. Probab. 21, Paper No. 44.


Peres, Y. 2002. *Brownian intersections, cover times and thick points via trees*, (Beijing, 2002), Higher Ed. Press, Beijing, pp. 73–78.


BIBLIOGRAPHY 437


Subag, E. 2013. A lower bound for the mixing time of the random-to-random insertions shuffle, Electron. J. Probab. 18, no. 20. 363


Notation Index

The symbol := means defined as.

The set \{\ldots, -1, 0, 1, \ldots\} of integers is denoted \(\mathbb{Z}\) and the set of real numbers is denoted \(\mathbb{R}\).

For sequences \((a_n)\) and \((b_n)\), the notation \(a_n = O(b_n)\) means that for some \(c > 0\) we have \(a_n/b_n \leq c\) for all \(n\), while \(a_n = o(b_n)\) means that \(\lim_{n \to \infty} a_n/b_n = 0\), and \(a_n \asymp b_n\) means both \(a_n = O(b_n)\) and \(b_n = O(a_n)\) are true.

\begin{itemize}
  \item \(A_n\) (alternating group), 100
  \item \(B\) (congestion ratio), 188
  \item \(C(a \leftrightarrow z)\) (effective conductance), 118
  \item \(E(f, h)\) (Dirichlet form), 181
  \item \(E(f)\) (Dirichlet form), 181
  \item \(E\) (edge set), 8
  \item \(E\) (expectation), 367
  \item \(E_\mu\) (expectation from initial distribution \(\mu\)), 4
  \item \(E_x\) (expectation from initial state \(x\)), 4
  \item \(E_\mu\) (expectation w.r.t. \(\mu\)), 92, 392
  \item \(G\) (graph), 8
  \item \(G^\circ\) (lamplighter graph), 273
  \item \(I\) (current flow), 117
  \item \(P\) (transition matrix), 2
  \item \(P_A\) (transition matrix of induced chain), 186
  \item \(\hat{P}\) (time reversal), 12
  \item \(P\{X \in B\}\) (probability of event), 366
  \item \(P_\mu\) (probability from initial distribution \(\mu\)), 4
  \item \(P_x\) (probability from initial state \(x\)), 4
  \item \(P_{x,y}\) (probability w.r.t. coupling started from \(x\) and \(y\)), 61
  \item \(Q(x, y)\) (edge measure), 58
  \item \(R(a \leftrightarrow z)\) (effective resistance), 118
  \item \(S_n\) (symmetric group), 75
  \item \(S^V\) (configuration set), 41
  \item \(V\) (vertex set), 8
  \item \(Var\) (variance), 661
  \item \(Var_\mu\) (variance w.r.t. \(\mu\)), 92
  \item \(W\) (voltage), 117
  \item \(Z_n\) (\(n\)-cycle), 63
  \item \(Z_n^d\) (torus), 61
  \item \(c(\varepsilon)\) (conductance), 115
  \item \(d(t)\) (total variation distance), 53
  \item \(d(t)\) (total variation distance), 53
  \item \(d_H\) (Hellinger distance), 58, 287
  \item \(id\) (identity element), 27
  \item \(i.i.d.\) (independent and identically distributed), 109, 60
  \item \(r(\varepsilon)\) (resistance), 115
  \item \(s_x(t)\) (separation distance started from \(x\)), 79
  \item \(s(t)\) (separation distance), 79
  \item \(t_{cov}\) (worst case mean cover time), 149
  \item \(t_{hit}\) (maximal hitting time), 128
  \item \(t_{mix}(\varepsilon)\) (mixing time), 53
  \item \(t_{Ces}\) (Cesàro mixing time), 58
  \item \(t_{cont}^{mix}\) (continuous mixing time), 283
  \item \(t_{rel}\) (relaxation time), 162
  \item \(t_{\odot}\) (target time), 128
  \item \(\beta\) (inverse temperature), 44
  \item \(\delta_x\) (Dirac delta), 4
  \item \(\Delta\) (maximum degree), 70
  \item \(\Gamma_{xy}\) (path), 188
  \item \(\gamma\) (spectral gap), 102
  \item \(\gamma_*\) (absolute spectral gap), 102
  \item \(\lambda_j\) (eigenvalue of transition matrix), 162
  \item \(\lambda_*\) (maximal non-trivial eigenvalue), 112
  \item \(\mathcal{X}\) (state space), 2
  \item \(\omega\) (root of unity), 164
  \item \(\Phi(S)\) (bottleneck ratio of set), 58
  \item \(\Phi_*\) (bottleneck ratio), 58
  \item \(\pi\) (stationary distribution), 9
  \item \(\rho\) (metric), 201, 375
  \item \(\rho_K(\mu, \nu)\) (transportation metric), 201
\end{itemize}
\( \rho_{i,j} \) (reversal), \(^{237}\)
\( \sigma \) (Ising spin), \(^{17}\)
\( \tau_A \) (hitting time for set), \(^{77\,127}\)
\( \tau_{a,b} \) (commute time), \(^{130}\)
\( \tau_{\text{couple}} \) (coupling time), \(^{62}\)
\( \tau_{\text{cov}} \) (cover time variable), \(^{149}\)
\( \tau_{A}^\text{cov} \) (cover time for set), \(^{150}\)
\( \tau_x \) (hitting time), \(^{10\,127}\)
\( \tau^+_x \) (first return time), \(^{10\,127}\)
\( \theta \) (flow), \(^{117}\)

\( \wedge \) (min), \(^{39}\)
\( (ijk) \) (cycle (permutation)), \(^{100}\)
\( \partial S \) (boundary of \( S \)), \(^{89}\)
\( \ell^2(\pi) \) (inner product space), \(^{160}\)
\( [x] \) (equivalence class), \(^{25}\)
\( \langle \cdot, \cdot \rangle \) (standard inner product), \(^{160}\)
\( \langle \cdot, \cdot \rangle_\pi \) (inner product w.r.t. \( \pi \)), \(^{160}\)
\( \hat{\mu} \) (reversed distribution), \(^{55}\)
\( 1_A \) (indicator function), \(^{14}\)
\( \sim \) (adjacent to), \(^{18}\)
\( \|\mu - \nu\|_{TV} \) (total variation distance), \(^{47}\)
Index

Italics indicate that the reference is to an exercise.

absolute spectral gap, 162
absorbing state, 16
acceptance-rejection sampling, 380
alternating group, 100, 109
aperiodic chain, 7
approximate counting, 210
averaging over paths, 190

ballot theorem, 39
binary tree, 65
  Ising model on, 221
random walk on
  bottleneck ratio lower bound, 91
  commute time, 132
  coupling upper bound, 66
  cover time, 151
  hitting time, 146
  no cutoff, 268
birth-and-death chain, 26, 260, 300
  stationary distribution, 26
block dynamics
  for Ising model, 221, 362
bottleneck ratio, 88, 89
  bounds on relaxation time, 183
  lower bound on mixing time, 88
boundary, 89
Bounded Convergence Theorem, 371

Catalan number, 32
Cayley graph, 29
censoring inequality, 315
Central Limit Theorem, 369
Cesàro mixing time, 83, 230
CFTP, see also coupling from the past
Chebyshev’s inequality, 367
Cheeger constant, 98
children (in tree), 90
coin tossing patterns, see also patterns in coin tossing
colorings, 45
  approximate counting of, 210
  Glauber dynamics for, 42, 361
  exponential lower bound on star, 90
  lower bound on empty graph, 97
  path coupling upper bound, 207
  Metropolis dynamics for
    grand coupling upper bound, 70
    relaxation time, 180
  communicating classes, 15
  commute time, 130
  Identity, 131
  comparison of Markov chains, 185
  canonical paths, 188
  on groups, 190
  randomized paths, 190
  theorem, 188, 223, 233, 240
complete graph, 81
  Ising model on, 219
  lamplighter chain on, 278
conductance, 115
  bottleneck ratio, 88
configuration, 11
congestion ratio, 158, 190
connected graph, 74
connective constant, 224
continuous-time chain, 281
  Convergence Theorem, 283
  product chains, 286
  relation to lazy chain, 283
  relaxation time, 285
Convergence Theorem, 62
  continuous time, 283
  coupling proof, 73
  null recurrent chain, 301
  positive recurrent chain, 291
     convolution, 441
  counting lower bound, 87
  coupling
    bound on d(t), 62
    characterization of total variation distance, 50
    from the past, 349
    grand, 60, 61, 322, 355
Markovian, 61, 76
  of distributions, 49, 50, 201
  of Markov chains, 61
  of random variables, 49, 201
  optimal, 60, 202
441
coupon collector, 22, 62, 81, 82, 94
cover time variable, 139
current flow, 117
cutoff, 262
  open problems, 361
  window, 263
cutset edge, 122
cycle
  biased random walk on, 14
  Ising model on
    mixing time pre-cutoff, 220
random walk on, 5, 8, 17, 23, 32, 78
  bottleneck ratio, 83
  coupling upper bound, 68
  cover time, 149, 157
  eigenvalues and eigenfunctions, 164
  hitting time upper bound, 112
  last vertex visited, 86
  lower bound, 83
  no cutoff, 268
  spectral gap, 155
  strong stationary time upper bound,
    82, 86
  cycle law, 118
  cycle notation, 100
  cyclic-to-random shuffle, 116

degree of vertex, 8
density function, 366
depth (of tree), 65
descendant (in tree), 91
detailed balance equations, 139
diameter, 88, 201
diameter lower bound, 88
dimer system, 385
  Dirichlet form, 181
  distinguishing statistic, 91
  distribution function, 366
  divergence
    of flow, 174
  Dominated Convergence Theorem, 371
  domino tiling, 385
  Doob h-transform, 257
  Doob decomposition, 266
  Durrett chain
    comparison upper bound, 240
    distinguishing statistic lower bound, 238
  East model, 393
    lower bound, 37
    edge cutset, 122
    edge measure, 85
  effective conductance, 118
  effective resistance, 118
    gluing nodes, 119, 122
    of grid graph, 125
    of tree, 150
    Parallel Law, 119
  energy
    of flow, 121
    of Ising configuration, 14
    ergodic theorem, 392
    escape probability, 118
    essential state, 15
    Eulerian graphs, 135
    even permutation, 100
    event, 400
    evolving-set process, 260
    exclusion process, 324
    biased, 330
    monotonicity of, 326
    on path
      mixing time, 329
    expander graph, 115
    Ising model on, 227
    ExpanderMixingLemma, 177
    expectation, 367

  Fibonacci numbers, 213
  FIFO queue, 304
  “fifteen” puzzle, 110
  first return time, 10, 127
  flow, 117
  fugacity, 43
  fully polynomial randomized approximation scheme, 210
  gambler’s ruin, 24, 34, 126, 247
  Gaussian elimination chain, 363
  generating function, 141
  generating set, 28
  geometric mixing time, 336
  Gibbs distribution, 44
  Gibbs sampler, 41
  Glauber dynamics
    definition, 12
    for colorings, 42, 301
      path coupling upper bound, 207
    for hardcore model, 41, 72
      coupling from the past, 356
      relaxation time, 181
    for Ising model, 44, 155, 216
      coupling from the past, 351
    for product measure, 169
    glued graphs, 143
      complete, 51
      lower bound, 36
      strong stationary time upper bound,
        51
    hypercube
      hitting time upper bound, 143
      strong stationary time, 146
INDEX 443

torus
  bottleneck ratio lower bound, 90
  hitting time upper bound, 127, 144
gluing (in networks), 119, 122
grand coupling, 69, 70, 84, 85
graph, 8
  Cayley, 20
colorings, see also colorings
  complete, 81
  connected, 88
degree of vertex, 8
diameter, 88
easy, 9
expander, 196, 229
glued, see also glued graphs
grid, 123
ladder, 225
loop, 9
multiple edges, 9
oriented, 117
proper coloring of, 38, see also colorings
  regular, 10
  counting lower bound, 87
  simple random walk on, 8
Green’s function, 119, 203
  grid graph, 123
  Ising model on, 227
group, 27
  generating set of, 28
  random walk on, 28, 75, 99, 190
  symmetric, 75
halting state, 20
Hamming weight, 20
hardcore model, 12
  Glauber dynamics for, 44
  coupling from the past, 350
  relaxation time, 181
  monotone, 306
  on complete graph
    mixing time bounds, 249
    on cycle
      mixing time pre-cutoff, 220
      relaxation time, 315
  on expander, 225
  on grid
    relaxation time lower bound, 227
    mixing time upper bound, 221
  on tree, 229
  open problems, 360
  partial order on configurations, 351
  partition function, 44
  positive correlations, 312
  isoperimetric constant, 98
k-fuzz, 303
Kac lemma, 297
Kirchhoff’s node law, 117

$\ell^p(\pi)$ distance, 174
L-reversal chain, see also Durrett chain
ladder graph, 225
lamplighter chain, 273, 363
  mixing time, 276
  on cycle, 276
INDEX

on hypercube, 278
on torus, 278
relaxation time, 274
separation cutoff, 279
Laws of Large Numbers, 368
lazy version of a Markov chain, 8, 176 283
leaf, 18 65
level (of tree), 65
linear congruential sequence, 384
Lipschitz constant, 180, 213
loop, 9
lower bound methods
bottleneck ratio, 88 89
counting bound, 57
diameter bound, 88
distinguishing statistic, 94
Wilson’s method, 193
lozenge tiling, 352
lumped chain, see also projection
Markov chain
aperiodic, 7
birth-and-death, 26
communicating classes of, 15
comparison of, see also comparison of Markov chains
continuous time, 254
Convergence Theorem, 52 75
coupling, 61
definition of, 2
ergodic theorem, 392
exclusion process, see also exclusion process
irreducible, 7
lamplighter, see also lamplighter chain
lazy version of, 8
mixing time of, 53
Monte Carlo method, 38, 358
null recurrent, 297
periodic, 71 179
positive recurrent, 297
product, see also product chain
projection of, 26 33
random mapping representation of, 61
reversible, 14 116
stationary distribution of, 9
time averages, 172
time reversal of, 14 74
time-inhomogeneous, 49 113 208
transient, 298
transitive, 21 94
unknown, 355
Markov property, 2
Markov’s inequality, 267
Markovian coupling, 61 73
martingale, 244
Matthews method
lower bound on cover time, 150
upper bound on cover time, 150
maximum principle, 176 116
MCMC, see also Markov chain, Monte Carlo method
metric space, 201 374
Metropolis algorithm, 38
arbitrary base chain, 10
for colorings, 70 130
for Ising model, 185
symmetric base chain, 38
minimum expectation of a stationary time, 387
mixing time, 54
\(\ell^2\) upper bound, 171
Cesaro, 83
continuous time, 283
coupling upper bound, 82
hitting time upper bound, 139
path coupling upper bound, 204
relaxation time lower bound, 102
relaxation time upper bound, 102
monotone chains, 208
positive correlations, 112
Monotone Convergence Theorem, 371
monotone spin system, 110
Monte Carlo method, 38
null recurrent, 294
integer, 294
node, 116
node law, 117
odd permutation, 100
Ohm’s law, 118
optimal coupling, 50 202
Optional Stopping Theorem, 246
order statistic, 389
oriented edge, 117
Parallel Law, 119
parity (of permutation), 100
partition function, 44
path
metric, 208
random walk on, see also
birth-and-death chain, see also gambler’s ruin, 69 120 263
eigenvalues and eigenfunctions, 166
167
path coupling, 201
upper bound on mixing time, 201 210
patterns in coin tossing
cover time, 110
hitting time, 144 248
perfect sampling, see also sampling, exact
INDEX

periodic chain, 7
  eigenvalues of, 176
pivot chain for self-avoiding walk, 386
Pólya’s urn, 28, 124, 125, 138
positive correlations
  definition of, 310
  of product measures, 311
positive recurrent, 296
pre-cutoff, 263, 271
  mixing time of Ising model on cycle, 220
previsible sequence, 245
probability
  distribution, 366
  measure, 366
  space, 365
product chain
  eigenvalues and eigenfunctions of, 168
  in continuous time, 286
  spectral gap, 169
  Wilson’s method lower bound, 195
projection, 25, 34, 165
onto coordinate, 201
proper colorings, see also colorings
random adjacent transpositions, 235
  comparison upper bound, 235
  coupling upper bound, 234
  single card lower bound, 235
  Wilson’s method lower bound, 236
random colorings, 90
random mapping representation, 6, 69
random number generator, see also pseudorandom number generator
random sample, 38
Random Target Lemma, 128
random transposition shuffle, 101, 111
  coupling upper bound, 103
  lower bound, 102
  relaxation time, 164
  strong stationary time upper bound, 104
random walk
  on $\mathbb{R}$, 244
  on $\mathbb{Z}$, 30, 293, 307
    biased, 245
    null recurrent, 296
  on $\mathbb{Z}^d$, 292
    recurrent for $d = 2$, 296
    transient for $d = 3$, 296
on binary tree
  bottleneck ratio lower bound, 91
  commute time, 132
  coupling upper bound, 66
  cover time, 151
  hitting time, 159
  no cutoff, 268
  on cycle, 5, 8, 17, 28, 34, 78
    bottleneck ratio, 183
    coupling upper bound, 63
    cover time, 149, 157
    eigenvalues and eigenfunctions, 164
    hitting time upper bound, 142
    last vertex visited, 86
    lower bound, 83
    no cutoff, 268
    spectral gap, 165
    strong stationary time upper bound, 82, 86
  on group, 24, 75, 99, 190
  on hypercube, 23, 28
    $\ell^2$ upper bound, 172
    bottleneck ratio, 183
    coupling upper bound, 62
    cover time, 157
    cutoff, 169, 266
    distinguishing statistic lower bound, 94
    eigenvalues and eigenfunctions of, 170
    hitting time, 145
    relaxation time, 181
    separation cutoff, 299
    strong stationary time upper bound, 163, 171
  Wilson’s method lower bound, 193
  on path, see also birth-and-death chain,
    see also gambler’s ruin, 60, 120, 263
    eigenvalues and eigenfunctions, 166
  on torus, 61
    coupling upper bound, 63, 75
    cover time, 152, 157
    hitting time, 140
    perturbed, 190, 198
    self-avoiding, 385
    simple, 8, 14, 115, 189
    weighted, 185
  randomized paths, 190
  Rayleigh’s Monotonicity Law, 122, 205
  Rayleigh-Ritz theorem, 376
  recurrent, 293, 303
  reflection principle, 30, 51, 75
  regular graph, 10
    counting lower bound, 81
    relaxation time, 162
    bottleneck ratio bounds, 183
    continuous time, 255
    coupling upper bound, 180
    mixing time lower bound, 162
    mixing time upper bound, 163
    variational characterization of, 182
    resistance, 115
    return probability, 141, 255, 302
  reversal, see also Durrett chain, 237
  reversed chain, see also time reversal
reversed distribution, 55
reversibility, 13 116
detailed balance equations, 13
riffle shuffle, 106 113
counting lower bound, 109
generalized, 117
strong stationary time upper bound, 108
rising sequence, 108
rooted tree, 65
roots of unity, 164
sampling, 379
and counting, 209
exact, 209 365
self-avoiding walk, 385 386 391
semi-random transpositions, 113
separation distance, 20 50 83 363
total variation upper bound, 80
upper bound on total variation, 80
Series Law, 119
shift chain, see also patterns in coin tossing shuffle:
cyclic-to-random, 114
move-to-front, 82
open problems, 301
random adjacent transposition, 233
comparison upper bound, 236
coupling upper bound, 237
single card lower bound, 238
Wilson’s method lower bound, 239
random transposition, 104 177
coupling upper bound, 103
lower bound, 102
relaxation time, 164
strong stationary time upper bound, 104 112
riffle, 106 113
counting lower bound, 109
generalized, 117
strong stationary time upper bound, 108
semi-random transpositions, 113
top-to-random, 75
cutoff, 202
lower bound, 95
strong stationary time upper bound, 78 82 85
simple random walk, 51 115 159
stationary distribution of, 9
simplex, 383
simulation
of random variables, 379 390
sink, 117
source, 117
spectral gap, 182 see also relaxation time absolute, 102
bottleneck ratio bounds, 183
variational characterization of, 183
spectral theorem for symmetric matrices, 376
spin system, 14
montone, 310
star, 90
stationary distribution, 9
uniqueness of, 13 17
stationary time, 77 83
strong
Stirling’s formula, 376
stochastic domination, 307
stochastic flow, see also grand coupling stopping time, 86 248
Strassen’s theorem, 308
strength of flow, 117
Strong Law of Large Numbers, 368
strong stationary time, 78 258
submartingale, 244
submultiplicativity of $d(t)$, 54 of $s(t)$, 55
supermartingale, 244 260
support, 366
symmetric group, 76 99
symmetric matrix, 376
systematic updates, 361
target time, 128 129
tensor product, 168
Thomson’s Principle, 121 294
tiling
domino, 385
lozenge, 352
time averages, 172
time reversal, 14 57 55 58 88 82 107
time-inhomogeneous Markov chain, 119 113
208
top-to-random shuffle, 75
cutoff, 202
lower bound, 95
strong stationary time upper bound, 78 82 85
torus
definition of, 74
glued
bottleneck ratio lower bound, 90
hitting time upper bound, 144
lamplighter chain on, 278
random walk on
coupling upper bound, 63 75
cover time, 152 177
hitting time, 138
perturbed, 136 178
total variation distance, 44
coupling characterization of, 50
Hellinger distance upper bound, 288
monotonicity of, 77
separation distance upper bound, 80
standardized \((d(t), \bar{d}(t))\), 53
upper bound on separation distance, 80
transient, 293
transition matrix
definition of, 2
eigenvalues of, 160, 176
multiply on left, 5
multiply on right, 5
spectral representation of, 160
transition probabilities, \(t\)-step, 5
transition times, 281
transitive
chain, 29, 34
network, 131
transportation metric, 201, 213
transpose (of a matrix), 376
transposition, 100
tree, 77, 95
binary, see also binary tree, 86
effective resistance, 120
Ising model on, 221, 229
rooted, 85
triangle inequality, 375
unbiasing
von Neumann, 378
unit flow, 117
unity
roots of, 164
unknown chain
sampling from, 358
up-right path, 33
urn model
Ehrenfest, 24, 34, 266
Pólya, 25, 124, 125, 138
variance, 307
voltage, 117
von Neumann unbiasing, 378
Wald’s identity, 60
Weak Law of Large Numbers, 368
weighted random walk, 115
Wilson’s method, 193, 220, 236
window (of cutoff), 263
winning streak, 55, 66
time reversal, 68
wreath product, 273
This book is an introduction to the modern theory of Markov chains, whose goal is to determine the rate of convergence to the stationary distribution, as a function of state space size and geometry. This topic has important connections to combinatorics, statistical physics, and theoretical computer science. Many of the techniques presented originate in these disciplines.

The central tools for estimating convergence times, including coupling, strong stationary times, and spectral methods, are developed. The authors discuss many examples, including card shuffling and the Ising model, from statistical mechanics, and present the connection of random walks to electrical networks and apply it to estimate hitting and cover times.

The first edition has been used in courses in mathematics and computer science departments of numerous universities. The second edition features three new chapters (on monotone chains, the exclusion process, and stationary times) and also includes smaller additions and corrections throughout. Updated notes at the end of each chapter inform the reader of recent research developments.

Markov Chains and Mixing Times is a magical book, managing to be both friendly and deep. It gently introduces probabilistic techniques so that an outsider can follow. At the same time, it is the first book covering the geometric theory of Markov chains and has much that will be new to experts. It is certainly THE book that I will use to teach from. I recommend it to all comers, an amazing achievement.

—Persi Diaconis, Mary V. Sunseri Professor of Statistics and Mathematics, Stanford University

Mixing times are an active research topic within many fields from statistical physics to the theory of algorithms, as well as having intrinsic interest within mathematical probability and exploiting discrete analogs of important geometry concepts. The first edition became an instant classic, being accessible to advanced undergraduates and yet bringing readers close to current research frontiers. This second edition adds chapters on monotone chains, the exclusion process and hitting time parameters. Having both exercises and citations to important research papers it makes an outstanding basis for either a lecture course or self-study.

—David Aldous, University of California, Berkeley

Mixing time is the key to Markov chain Monte Carlo, the queen of approximation techniques. With new chapters on monotone chains, exclusion processes, and set-hitting, Markov Chains and Mixing Times is more comprehensive and thus more indispensable than ever. Prepare for an eye-opening mathematical tour!

—Peter Winkler, Dartmouth College

The study of finite Markov chains has recently attracted increasing interest from a variety of researchers. This is the second edition of a very valuable book on the subject. The main focus is on the mixing time of Markov chains, but there is a lot of additional material.

In this edition, the authors have taken the opportunity to add new material and bring the reader up to date on the latest research. I have used the first edition in a graduate course and I look forward to using this edition for the same purpose in the near future.

—Alan Frieze, Carnegie Mellon University