Mathematical Aspects of Mixing Times in Markov Chains

Mathematical Aspects of Mixing Times in Markov Chains

Ravi Montenegro

Department of Mathematical Sciences
University of Massachusetts Lowell
Lowell, Massachusetts 01854, USA
ravi_montenegro@uml.edu

Prasad Tetali

School of Mathematics
Georgia Institute of Technology
Atlanta, Georgia 30332, USA
tetali@math.gatech.edu

now
the essence of knowledge
Boston – Delft

Foundations and Trends® in Theoretical Computer Science

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

A Cataloging-in-Publication record is available from the Library of Congress

The preferred citation for this publication is R. Montenegro and P. Tetali, Mathematical Aspects of Mixing Times in Markov Chains, Foundation and Trends® in Theoretical Computer Science, vol 1, no 3, pp 237–354, 2006

Printed on acid-free paper

ISBN: 1-933019-29-8

Foundations and Trends® in Theoretical Computer Science
Volume 1 Issue 3, 2006

Editorial Board

Editor-in-Chief:
Madhu Sudan
Department of CS and EE
MIT, Stata Center, Room G640
32 Vassar Street, Cambridge
Massachusetts 02139,
USA
madhu@mit.edu

Editorial Scope

Foundations and Trends® in Theoretical Computer Science will publish survey and tutorial articles in the following topics:

- Algorithmic game theory
- Computational algebra
- Computational aspects of combinatorics and graph theory
- Computational aspects of communication
- Computational biology
- Computational complexity
- Computational geometry
- Computational learning
- Computational Models and Complexity
- Computational Number Theory
- Cryptography and information security
- Data structures
- Database theory
- Design and analysis of algorithms
- Distributed computing
- Information retrieval
- Operations Research
- Parallel algorithms
- Quantum Computation
- Randomness in Computation

Information for Librarians
Foundations and Trends® in Theoretical Computer Science, 2006, Volume 1, 4 issues. ISSN paper version 1551-305X.ISSN online version 1551-3068.Also available as a combined paper and online subscription.

Foundations and Trends® in
Theoretical Computer Science
Vol. 1, No 3 (2006) 237–354

DOI: 10.1561/0400000003

now
the essence of knowledge

Mathematical Aspects of Mixing Times in Markov Chains

Ravi Montenegro[1] and Prasad Tetali[2]

[1] *University of Massachusetts Lowell, Lowell, Massachusetts 01854, USA, ravi_montenegro@uml.edu*
[2] *Georgia Institute of Technology, Atlanta, Georgia 30332, USA, tetali@math.gatech.edu*

Abstract

In the past few years we have seen a surge in the theory of finite Markov chains, by way of new techniques to bounding the convergence to stationarity. This includes functional techniques such as logarithmic Sobolev and Nash inequalities, refined spectral and entropy techniques, and isoperimetric techniques such as the average and blocking conductance and the evolving set methodology. We attempt to give a more or less self-contained treatment of some of these modern techniques, after reviewing several preliminaries. We also review classical and modern lower bounds on mixing times. There have been other important contributions to this theory such as variants on coupling techniques and decomposition methods, which are not included here; our choice was to keep the analytical methods as the theme of this presentation. We illustrate the strength of the main techniques by way of simple examples, a recent result on the Pollard Rho random walk to compute the discrete logarithm, as well as with an improved analysis of the Thorp shuffle.

Contents

1

Introduction

Monte Carlo methods have been in use for a long time in statistical physics and other fields for sampling purposes. However, the computer scientists' novel idea [38] of reducing the problem of approximately counting the size of a large set of combinatorial objects to that of near-uniform sampling from the same set, gave the study of Markov chain Monte Carlo (MCMC) algorithms an entirely new purpose, and promptly spawned an active subtopic of research. We recall here that the work of [38] shows that in fact, under the technical assumption of so-called *self-reducibility*, approximate counting of the size of a set *in polynomial time* is feasible if and only if one is able to sample from the set with nearly uniform distribution, also in polynomial time. In terms of the finite Markov chain underlying an MCMC algorithm, the latter problem translates to designing and analyzing a Markov chain with a prescribed stationary measure, with a view (and hope) to providing rigorous estimates on the polynomial fastness of the rate of convergence to stationarity of the chain. Thus the classical subject of finite Markov chains has received much renewed interest and attention.

To add concreteness to the above story, we briefly mention as examples of large sets of combinatorial objects, the set of matchings of a given

(as input) bipartite graph [34, 36], the set of proper colorings of a given graph using a fixed number of colors [28], the number of matrices having non-negative integer entries and with prescribed row and column sums [16], etc. Albeit combinatorial, a non-discrete estimation problem which received significant devotion, both by way of algorithms and analytical techniques, is that of (approximately) computing the volume of a high-dimensional convex body (see [47, 48] and references therein). There have already been some very good surveys focusing on such combinatorial, computational and statistical physics applications of finite Markov chains. For an elaboration of the above premise, and a crash course on several basic techniques, we recommend the excellent article of Jerrum [32]. Towards the end of this introduction, we provide other pointers to existing literature on this subject. However, much of the theory surveyed here is rather recent theoretical (analytical) development and is so far unavailable in a unified presentation. The significance of these new methods is as follows.

The rate of convergence to stationarity of a finite Markov chain is typically measured by the so-called mixing time, defined as the first time τ by which the L^1 (or more generally, L^p) distance between the distribution at time τ and the stationary distribution falls below a small threshold, such as $1/2e$. It is classical and elementary to show that the inverse spectral gap of a lazy reversible Markov chain captures the mixing time (in L^1 and L^2) up to a factor of $\log(1/\pi_*)$, where $\pi_* = \min_x \pi(x)$ denotes the smallest entry in the stationary probability (vector) π of the chain. While the more technical logarithmic Sobolev constant captures the L^2-mixing time up to a factor of $\log\log(1/\pi_*)$, it is typically much harder to bound – to mention a specific example, the exact constant is open for the 3-point space with arbitrary invariant measure; also in a few cases, the log-Sobolev constant is known not to give tight bounds on the L^1-mixing time. The main strength of the spectral profile techniques and the evolving set methodology considered in this survey seems to be that of avoiding extra penalty factors such as $\log\log(1/\pi_*)$. These extra pesky factors can indeed be non-negligible when the state space is of exponential (or worse) size in the size of the input. In the present volume, the above is illustrated with a couple of simple examples, and with the now-famous Thorp shuffle, for which an improved $O(d^{29})$ mixing time is described, building on the proof of

Morris that proved the first polynomial (in d) bound of $O(d^{44})$ – here the number of cards in the deck is 2^d, and hence the state space has size $2^d!$, resulting in a $\log\log(1/\pi_*)$ factor of only $O(d)$, while a $\log(1/\pi_*)$ factor would have yielded an all too costly $O(d2^d)$.

The approach to L^2-mixing time using the spectral profile has the additional advantage of yielding known (upper) estimates on mixing time, under a log-Sobolev inequality and/or a Nash-type inequality. Thus various functional analytic approaches to mixing times can be unified with the approach of bounding the spectral profile. The one exception to this is the approach to stationarity using relative entropy; the corresponding *entropy constant* capturing the rate of decay of entropy has also been hard to estimate.

A brief history of the above development can perhaps be summarized as follows. A fundamental contribution, by way of initiating several subsequent works, was made by Lovász and Kannan in [46] in which they introduced the notion of *average conductance* to bound the total variation mixing time. This result was further strengthened and developed by Morris and Peres using the so-called *evolving sets*, where they analyze a given chain by relating it to an auxiliary (dual) chain on subsets of the states of the original chain. While this was introduced in [61] in a (martingale-based) probabilistic language, it turns out to be, retrospectively, an independent and alternative view of the notion of a Doob transform introduced and investigated by Diaconis and Fill [18]. Further refinement and generalization of the evolving sets approach was done in detail by [56]. The functional analog of some of this is done via the spectral profile, developed for the present context of finite Markov chains, in [30], while having its origins in the developments by [4] and [14] in the context of manifolds.

Besides summarizing much of the above recent developments in this exciting topic, we address some classical aspects as well. In discrete-time, much of the literature uses laziness assumptions to avoid annoying technical difficulties. While laziness is a convenient assumption, it slows down the chain by a factor of 2, which may not be desirable in practice. We take a closer look at this issue and report bounds which reflect the precise dependence on laziness. The notion of modified conductance circumvents laziness altogether, and we discuss this aspect briefly and compare it to bounds derived from the functional approach.

Further details on the modified conductance and its usefulness can be found in [57]. Another issue is that of the role of reversibility (a.k.a. detailed balance conditions). We tried to pay particular attention to it, due to current trend in the direction of analyzing various nonreversible Markov chains. Although often a convenient assumption, we avoid as much as possible this additional assumption. In particular, we include a proof of the lower bound on the total variation mixing time in terms of the second eigenvalue in the general case. Besides providing upper and lower bounds for the mixing time of reversible and non-reversible chains, we report recent successes (with brief analysis) in the analysis of some non-reversible chains; see for example, the Pollard Rho random walk for the discrete logarithm problem and the Thorp shuffle.

In Section 2 we introduce notions of mixing times and prove the basic upper bounds on these notions using Poincaré and logarithmic Sobolev type functional constants. In Section 3 we move on to recent results using the spectral profile, as opposed to using simply the second eigenvalue. In Section 4 we review the evolving set methods. Our treatment of lower bounds on mixing times is provided in Section 5. We consider several examples for illustration in Section 6. In the penultimate section, we gather a few recent results together. This includes recent results on the so-called fastest mixing Markov chain problem, and a recent theorem [52] from perturbation theory of finite Markov chains; this theorem relates the stability of a stochastic matrix (subject to perturbations) to the rate of convergence to equilibrium of the matrix. We also recall here an old but not so widely known characterization of the spectral gap, which seems worth revisiting due to recent results utilizing this formulation. The Appendix contains a discussion on the relations between the distances considered in this survey, and others such as relative pointwise (L^∞) distance.

We mention here a few additional sources, by way of survey articles, for the interested reader. For a good overview of the basic techniques in estimating the mixing times of finite Markov chains, see [32, 33, 35]. Other updates include the tutorial lectures of [40, 65]. Also a recent manuscript of Dyer et al. [25] describes several comparison theorems for reversible as well as nonreversible Markov chains.

References

[1] D. J. Aldous and J. Fill, *Reversible Markov Chains and Random Walks on Graphs,* (book to appear). URL for draft at http://stat-www.berkeley.edu/users/aldous/RWG/book.html.

[2] N. Alon, R. Boppana, and J. Spencer, "An asymptotic isoperimetric inequality," *Geom. and Funct. Anal.*, vol. 8, pp. 411–436, 1998.

[3] D. Bakry and M. Emery, "Diffusions hypercontractives," *Séminaire de Probabilités XIX, Lecture Notes in Math. 1123*, pp. 177–206, Springer-Verlag, 1985.

[4] M. Barlow, T. Coulhon, and A. Grigor'yan, "Manifolds and graphs with slow heat kernel decay," *Invent. Math.*, vol. 144, pp. 609–649, 2001.

[5] S. Bochner, "Vector fields and Ricci Curvature," *Bull. Amer. Math. Soc.*, vol. 52, pp. 776–797, 1946.

[6] C. Borgs, *Statistical Physics Expansion Methods in Combinatorics and Computer Science.* CBMS-SIAM Lecture Notes, in preparation.

[7] C. Borgs, J. Chayes, A. Frieze, J. H. Kim, P. Tetali, E. Vigoda, and V. Vu, "Torpid mixing of some MCMC algorithms in statistical physics," *Proceedings of the 40th IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 218–229, 1999.

[8] A.-S. Boudou, P. Caputo, P. D. Pra, and G. Posta, "Spectral gap estimates for interacting particle systems via a Bochner type identity," *Journal of Functional Analysis*, vol. 232, pp. 222–258, 2005.

[9] S. Boyd, P. Diaconis, J. Sun, and L. Xiao, "Fastest mixing Markov chain on a path," *The American Mathematical Monthly*, vol. 113, no. 1, pp. 70–74, 2006.

[10] S. Boyd, P. Diaconis, and L. Xiao, "Fastest mixing Markov chain on a graph," *SIAM Review*, vol. 46, no. 4, pp. 667–689, 2004.

[11] P. Caputo and P. Tetali, *Walks, Transpositions, and Exclusion*. Preprint (July 2004).
[12] F. Chung, "Laplacians and the Cheeger inequality for directed graphs," *Annals of Combinatorics*, vol. 9, no. 1, pp. 1–19, 2005.
[13] T. Coulhon, "Ultracontractivity and Nash-type inequalities," *J. Funct. Anal.*, vol. 141, pp. 510–539, 1996.
[14] T. Coulhon, A. Grigor'yan, and C. Pittet, "A geometric approach to on-diagonal heat kernel lower bound on groups," *Ann. Inst. Fourier*, vol. 51, pp. 1763–1827, 2001.
[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons: New York, 1991.
[16] M. Cryan, M. Dyer, and D. Randall, "Approximately Counting Integral Flows and Cell-Bounded Contingency Tables," *Proc. of the 37th Annual ACM Symp. on Theory of Computing (STOC)*, pp. 413–422, 2005.
[17] J.-D. Deuschel and D. W. Stroock, *Large Deviations*. Academic Press, 1989.
[18] P. Diaconis and J. Fill, "Strong stationary times via a new form of duality," *The Annals of Probability*, vol. 18, no. 4, pp. 1483–1522, 1990.
[19] P. Diaconis, S. Holmes, and R. Neal, "Analysis of a Non-reversible Markov Chain Sampler," *Annals of Applied Probability*, vol. 10, no. 3, pp. 726–752, 2000.
[20] P. Diaconis and L. Saloff-Coste, "Comparison Theorems for Reversible Markov Chains," *The Annals of Applied Probability*, vol. 3, no. 3, pp. 696–730, 1993.
[21] P. Diaconis and L. Saloff-Coste, "Logarithmic Sobolev inequalities for finite Markov chains," *The Annals of Applied Probability*, vol. 6, no. 3, pp. 695–750, 1996.
[22] P. Diaconis and L. Saloff-Coste, "Nash Inequalities for Finite Markov Chains," *Journal of Theoretical Probability*, vol. 9, pp. 459–510, 1996.
[23] P. Diaconis and M. Shahshahani, "Time to reach," *SIAM J. Math. Anal.*, vol. 18, no. 1, pp. 208–218, 1987.
[24] P. Diaconis and D. Stroock, "Geometric bounds for eigenvalues of Markov chains," *The Annals of Applied Probability*, vol. 1, pp. 36–61, 1991.
[25] M. Dyer, L. Goldberg, M. Jerrum, and R. Martin, *Markov chain comparison*. 2005. Preprint (arXiv:math.PR/0410331).
[26] J. Fill, "Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process," *The Annals of Applied Probability*, vol. 1, no. 1, pp. 62–87, 1991.
[27] N. Fountoulakis and B. A. Reed, "The evolution of the conductance of a random graph," in preparation.
[28] A. Frieze and E. Vigoda, "Survey of Markov Chains for Randomly Sampling Colorings," To appear in Festschrift for Dominic Welsh, (2006).
[29] S. Goel, "Analysis of top to bottom-k shuffles," *The Annals of Applied Probability*, vol. 16, no. 1, pp. 30–55, 2006.
[30] S. Goel, R. Montenegro, and P. Tetali, "Mixing time bounds and the spectral profile," *Electronic Journal of Probability*, vol. 11, pp. 1–26, 2006.
[31] A. Grigor'yan, "Heat kernel upper bounds on a complete non-compact manifold," *Revista Matemá,tica Iberoamericaná*, vol. 10, pp. 395–452, 1994.

[32] M. Jerrum, "Mathematical Foundations of the Markov Chain Monte Carlo Method," in *Probabilistic Methods for Algorithmic Discrete Mathematics*, (Habib et al., ed.), pp. 116–165, Springer-Verlag, Germany, 1998.

[33] M. Jerrum, *Counting, Sampling and Integrating : Algorithms & Complexity*. Birkhäuser Verlag, Basel, 2003.

[34] M. Jerrum and A. Sinclair, "Conductance and the rapid mixing property for Markov chains: the approximation of the permanent resolved," *Proceedings of the 20th Annual ACM Symposium on Theory of Computing (STOC 1988)*, pp. 235–243, 1988.

[35] M. Jerrum and A. Sinclair, "The Markov chain Monte Carlo method: an approach to approximate counting and integration," in *Approximation Algorithms for NP-hard Problems*, (D. Hochbaum, ed.), ch. 12, PWS Publishing, Boston, 1996.

[36] M. Jerrum, A. Sinclair, and E. Vigoda, "A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries," *Journal of the ACM*, vol. 51, pp. 671–697, 2004.

[37] M. Jerrum and J.-B. Son, "Spectral Gap and log-Sobolev constant for balanced matroids," *Proc. of the 43rd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2002)*, pp. 721–729, 2002.

[38] M. Jerrum, L. Valiant, and V. Vazirani, "Random generation of combinatorial structures from a uniform distribution," *Theoretical Computer Science*, vol. 43, pp. 169–188, 1998.

[39] W. Johnson and J. Lindenstrauss, "Extensions of Lipschitz maps into a Hilbert space," *Contemp. Math.*, vol. 26, pp. 189–206, 1984.

[40] R. Kannan, "Markov chains and polynomial time algorithms," Plenary Talk at *Proc. of 35th Annual IEEE Symp. on the Foundations of Computer Science*, pp. 656–671, 1994.

[41] R. Kannan, L. Lovász, and R. Montenegro, "Blocking conductance and mixing in random walks," *Combinatorics, Probability and Computing*, to appear, 2006.

[42] H. J. Landau and A. M. Odlyzko, "Bounds for eigenvalues of certain stochastic matrices," *Linear Algebra Appl.*, vol. 38, pp. 5–15, 1981.

[43] G. Lawler and A. Sokal, "Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality," *Transactions of the American Mathematical Society*, vol. 309, pp. 557–580, 1988.

[44] T.-Y. Lee and H.-T. Yau, "Logarithmic Sobolev inequalities for some models of random walks," *The Annals of Probability*, vol. 26, no. 4, pp. 1855–1873, 1998.

[45] A. Lichnérowicz, *Géométrie des groupes de transformations*. Dunod, Paris, 1958.

[46] L. Lovász and R. Kannan, "Faster mixing via average conductance," *Proc. of the 31st Annual ACM Symp. on Theory of Computing*, pp. 282–287, 1999.

[47] L. Lovász and S. Vempala, "Simulated annealing in convex bodies and an $O*(n^4)$ volume algorithm," *Proc. of the 44th IEEE Found. of Computer Science, Boston*, To appear in Jour. Comp. Sys. Sci.(FOCS '03 special issue), 2003.

[48] L. Lovász and S. Vempala, "Hit-and-run from a corner," *Proc. of the 36th ACM Symp. on the Theory of Computing, Chicago, SIAM J. Computing (STOC '04 special issue)*, vol. 35, no. 4, pp. 985–1005, 2006.

[49] L. Miclo, "Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies," *Séminaire de probabilités de Strasbourg*, vol. 31, pp. 136–167, 1997.

[50] M. Mihail, "Conductance and Convergence of Markov Chains-A Combinatorial Treatment of Expanders," *Proc. of the 30th Annual Symposium on Foundations of Computer Science*, pp. 526–531, 1989.

[51] S. Miller and R. Venkatesan, "Spectral Analysis of Pollard Rho Collisions," *Proc. of the 7th Algorithmic Number Theory Symposium (ANTS VII)* in series *Lecture Notes in Computer Science (LNCS)*, Springer, to appear, 2006.

[52] A. Mitrophanov, "Stability and exponential convergence of continuous-time Markov chains," *J. Appl. Probab.*, vol. 40, pp. 970–979, 2003.

[53] A. Mitrophanov, "Sensitivity and convergence of uniformly ergodic Markov chains," *J. Appl. Probab.*, vol. 42, pp. 1003–1014, 2005.

[54] A. Mitrophanov, A. Lomsadze, and M. Borodovsky, "Sensitivity of hidden Markov models," *J. Appl. Probab.*, vol. 42, pp. 632–642, 2005.

[55] R. Montenegro, "Vertex and edge expansion properties for rapid mixing," *Random Structures & Algorithms*, vol. 26, no. 1–2, pp. 52–68, 2005.

[56] R. Montenegro, *Duality and evolving set bounds on mixing times.* 2006. Preprint.

[57] R. Montenegro, *Eigenvalues of non-reversible Markov chains: their connection to mixing times, reversible Markov chains, and Cheeger inequalities.* 2006. Preprint (arXiv:math.PR/0604362).

[58] R. Montenegro, *On the mixing time of the simple random walk and max-degree walk on a directed graph.* 2006. In preparation.

[59] B. Morris, "The mixing time for simple exclusion," *Annals of Applied Probability*, to appear, 2006.

[60] B. Morris, "The mixing time of the Thorp shuffle," *SIAM Journal on Computing (SICOMP)*, to appear, 2006.

[61] B. Morris and Y. Peres, "Evolving sets, mixing and heat kernel bounds," *Probability Theory and Related Fields*, vol. 133, no. 2, pp. 245–266, 2005.

[62] A. Naor, M. Sammer, and P. Tetali, *The Fastest Mixing Markov Process and the Subgaussian Constant.* 2005. Preprint.

[63] J. R. Norris, *Markov Chains.* Cambridge University Press, 1997.

[64] Y. Peres and D. Revelle, "Mixing times for random walks on finite lamplighter groups," *Electronic J. on Probab.*, vol. 9, pp. 825–845, 2004.

[65] D. Randall, "Rapidly mixing Markov chains with applications in computer science and physics," *Computing in Science & Engineering*, vol. 8, no. 2, pp. 30–41, March 2006.

[66] D. Randall and P. Tetali, "Analyzing Glauber dynamics using comparison of Markov chains," *J. Math. Physics*, vol. 41, pp. 1598–1615, 2000.

[67] M. Reed and B. Simon, *Methods of Modern Mathematical Physics II: Fourier Analysis, Self-Adjointness.* Academic Press Inc, New York, 1975.

[68] L. Saloff-Coste, "Total variation lower bounds for finite Markov chains: Wilson's lemma," in *Random walks and geometry*, pp. 515–532, Walter de Gruyter GmbH & Co. KG, Berlin, 2004.

[69] E. Seneta, "Coefficients of ergodicity: structure and applications," *Adv. Appl. Prob.*, vol. 11, pp. 576–590, 1979.

[70] A. Sinclair, "Improved bounds for mixing rates of Markov chains and multicommodity flow," *Combinatorics, Probability and Computing*, vol. 1, no. 4, pp. 351–370, 1992.

[71] E. Stein, "Interpolation of Linear Operators," *Trans. Amer. Math. Soc.*, vol. 83, pp. 482–492, 1956.

[72] J. Sun, S. Boyd, L. Xiao, and P. Diaconis, "The fastest mixing Markov process on a graph and a connection to a maximum variance unfolding problem," *SIAM Review*, to appear, 2006.

[73] D. Wilson, "Mixing Time of the Rudvalis Shuffle," *Electronic Communications in Probability*, vol. 8, no. 77–85, 2003.

[74] D. Wilson, "Mixing times of lozenge tiling and card shuffling Markov chains," *The Annals of Applied Probability*, vol. 14, no. 1, pp. 274–325, 2004.

Appendix

Our main focus has been on bounds for L^2 distance. Our bounds on mixing times in L^2 and relative entropy also yield bounds on the total variation mixing time using the following well-known inequality relating probability measures ν and μ.

$$\|\nu - \mu\|_{\mathrm{TV}} = \frac{1}{2}\left\|\frac{\nu}{\mu} - 1\right\|_{1,\mu} \leq \frac{1}{2}\left\|\frac{\nu}{\mu} - 1\right\|_{2,\mu}. \tag{8.1}$$

Further assuming that ν is absolutely continuous with respect to μ, the so-called Pinsker inequality (see Lemma 12.6.1 in [15] for a proof), asserts that:

$$\|\nu - \mu\|_{\mathrm{TV}}^2 \leq \frac{1}{2}D(\nu\|\mu) \tag{8.2}$$

Finally the general inequality $(\mathbb{E}_\mu f)\mathrm{Ent}_\mu(f) \leq \mathrm{Var}_\mu(f)$, valid for all measurable functions on an arbitrary probability space (since $\log \frac{f}{\mathbb{E}_\mu f} \leq \frac{f}{\mathbb{E}_\mu f} - 1$), when applied to $f = \nu/\mu$ implies that,

$$D(\nu\|\mu) \leq \left\|\frac{\nu}{\mu} - 1\right\|_{2,\mu}^2. \tag{8.3}$$

In a sense the L^∞, or relative pointwise distance, is the strongest of all distances. Our L^2 bounds also induce L^∞ bounds. Observe that if $t = t_1 + t_2$ then

$$\begin{aligned}\left|\frac{H_t(x,y)-\pi(y)}{\pi(y)}\right| &= \left|\frac{\sum_z (H_{t_1}(x,z)-\pi(z))\,(H_{t_2}(z,y)-\pi(y))}{\pi(y)}\right| \\ &= \left|\sum_z \pi(z)\left(\frac{H_{t_1}(x,z)}{\pi(z)}-1\right)\left(\frac{H^*_{t_2}(y,z)}{\pi(z)}-1\right)\right| \\ &\le \left\|h^x_{t_1}-1\right\|_2 \left\|h^{*y}_{t_2}-1\right\|_2 \qquad (8.4)\end{aligned}$$

where the inequality follows from Cauchy-Schwartz. Several bounds on L_∞ mixing then follow immediately, including

$$\tau_2(\epsilon) \le \tau_\infty(\epsilon) \le \tau_2\left(\epsilon\sqrt{\frac{\pi_*}{1-\pi_*}}\right)$$

and

$$\tau_2(\epsilon) \le \tau_\infty(\epsilon) \le \tau_2^{\mathsf{P}}(\sqrt{\epsilon}) + \tau_2^{\mathsf{P}^*}(\sqrt{\epsilon})\,.$$

The first of these two is somewhat unappealing because the asymptotic portion of $\tau_2(\epsilon)$ is of the form $\lambda^{-1}\log(1/\epsilon)$, and so taking $\tau_2\left(\epsilon\sqrt{\frac{\pi_*}{1-\pi_*}}\right)$ adds an extra factor of $\lambda^{-1}\log(1/\pi_*)$ to the $\tau_2(\epsilon)$ bound, potentially large relative to spectral profile bounds. The second bound unfortunately requires study of both P and P^*. However, if P is reversible then this last bound becomes

$$\tau_2(\epsilon) \le \tau_\infty(\epsilon) \le 2\tau_2(\sqrt{\epsilon})\,. \qquad (8.5)$$

More generally, most bounds in this survey were the same for P and P^*. For instance, (8.5) holds for the spectral profile bounds on L^2 mixing in terms of $\Lambda(r)$. In particular, the Dirichlet form satisfies

$$\mathcal{E}_{\mathsf{P}}(f,f) = \mathcal{E}_{\mathsf{P}^*}(f,f)$$

and so $\lambda(\mathsf{P}) = \lambda(\mathsf{P}^*)$, $\Lambda_{\mathsf{P}}(r) = \Lambda_{\mathsf{P}^*}(r)$, $\rho(\mathsf{P}) = \rho(\mathsf{P}^*)$ and $\Phi_{\mathsf{P}}(r) = \Phi_{\mathsf{P}^*}(r)$ (as $\mathsf{Q}(A,A^c) = \mathcal{E}(1_A,1_A)$).

It is not as clear how the f-congestion bounds behave for P^*. However, if π is uniform then

$$\begin{aligned}\Psi(A) = \Psi(A^c) &= \min_{\pi(B)=\pi(A)} \mathsf{Q}(A^c, B) \\ &= \min_{\pi(B)=\pi(A)} \mathsf{Q}_{\mathsf{P}^*}(B, A^c) \geq \min_{\pi(B)=\pi(A)} \Psi_{\mathsf{P}^*}(B)\end{aligned}$$

and so $\tilde{\phi}(r) \geq \tilde{\phi}_{\mathsf{P}^*}(r)$. The converse follows similarly, so $\tilde{\phi}_{\mathsf{P}}(r) = \tilde{\phi}_{\mathsf{P}^*}(r)$ when π is uniform, and (8.5) holds for modified-conductance bounds.