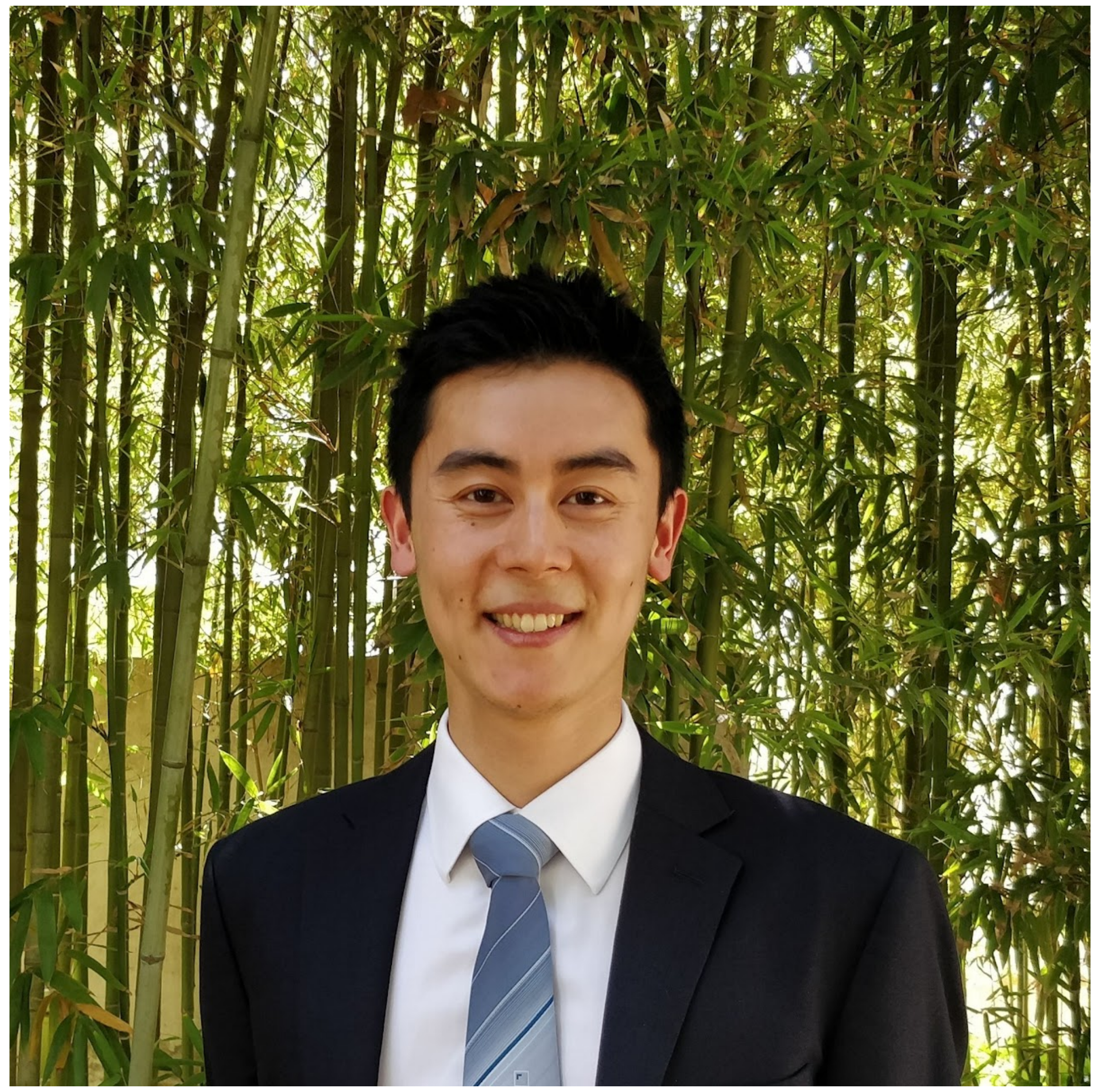
Convergence Bounds for the Random Walk Metropolis Algorithm

Perspectives from Isoperimetry

‘Advances in MCMC Sampling and Bayesian Computation’
IMS International Conference on Statistics and Data Science, 15 December 2025

Sam Power, University of Bristol

all joint work with: Christophe Andrieu, Anthony Lee (Bristol), Andi Wang (Warwick)



EXPLICIT CONVERGENCE BOUNDS FOR METROPOLIS MARKOV CHAINS: ISOPERIMETRY, SPECTRAL GAPS AND PROFILES

BY CHRISTOPHE ANDRIEU[1,a], ANTHONY LEE[1,b], SAM POWER[1,c] AND ANDI Q. WANG[2,d]

[1]*School of Mathematics, University of Bristol,* [a]*c.andrieu@bristol.ac.uk,* [b]*anthony.lee@bristol.ac.uk,* [c]*sam.power@bristol.ac.uk*

[2]*Department of Statistics, University of Warwick,* [d]*andi.wang@warwick.ac.uk*

We derive the first explicit bounds for the spectral gap of a random walk Metropolis algorithm on $\mathbb{R}^d$ for any value of the proposal variance, which when scaled appropriately recovers the correct d^{-1} dependence on dimension for suitably regular invariant distributions. We also obtain explicit bounds on the L^2-mixing time for a broad class of models. In obtaining
we refine the use of isoperimetric profile inequalities to obtai
profile bounds, which also enable the derivation of explicit bou
broader class of models. We also obtain similar results for the
Crank–Nicolson Markov chain, obtaining dimension-independ
der suitable assumptions.

WEAK POINCARÉ INEQUALITIES FOR MARKOV CHAINS: THEORY AND APPLICATIONS

BY CHRISTOPHE ANDRIEU[1,a], ANTHONY LEE[1,b], SAM POWER[1,c] AND ANDI Q. WANG[2,d]

[1]*School of Mathematics, University of Bristol,* [a]*c.andrieu@bristol.ac.uk;* [b]*anthony.lee@bristol.ac.uk;* [c]*sam.power@bristol.ac.uk*

[2]*Department of Statistics, University of Warwick,* [d]*andi.wang@warwick.ac.uk*

We investigate the application of Weak Poincaré Inequalities (WPI) to Markov chains to study their rates of convergence and to derive complexity bounds. At a theoretical level we investigate the necessity of the existence of WPIs to ensure L^2-convergence, in particular by establishing equivalence with the Resolvent Uniform Positivity-Improving (RUPI) condition and pro-
nterexample. From a more practical perspective, we extend the
heeger's inequalities to the subgeometric setting, and further ap-
hniques to study random-walk Metropolis algorithms for heavy-
distributions and to obtain lower bounds on pseudo-marginal al-

Poincaré inequalities for Markov chains: a meeting with Cheeger, Lyapunov and Metropolis

Christophe Andrieu, Anthony Lee, Sam Power, Andi Q. Wang

School of Mathematics, University of Bristol

August 11, 2022

Abstract

We develop a theory of weak Poincaré inequalities to characterize convergence rates of ergodic Markov chains. Motivated by the application of Markov chains in the context of algorithms, we develop a relevant set of tools which enable the practical study of convergence rates in the setting of Markov chain Monte Carlo methods, but also well beyond.

August 2024

Explicit convergence bounds for Metropolis Markov chains: Isoperimetry, spectral gaps and profiles

Christophe Andrieu, Anthony Lee, Sam Power, Andi Q. Wang

Author Affiliations +

Ann. Appl. Probab. 34(4): 4022-4071 (August 2024). DOI: 10.1214/24-AAP2058

JOURNAL ARTICLE
50 PAGES

DOWNLOAD PDF

SAVE TO MY LIBRARY

Annals of Applied Probability Future Papers

Papers to Appear in Subsequent Issues

Weak Poincaré Inequalities for Markov Chains: Theory and Applications	Christophe Andrieu, Anthony Lee, Sam Power and Andi Q. Wang

arXiv > math > arXiv:2208.05239

Search...
Help |

Mathematics > Probability

[Submitted on 10 Aug 2022]

Poincaré inequalities for Markov chains: a meeting with Cheeger, Lyapunov and Metropolis

Christophe Andrieu, Anthony Lee, Sam Power, Andi Q. Wang

We develop a theory of weak Poincaré inequalities to characterize convergence rates of ergodic Markov chains. Motivated by the application of Markov chains in the context of algorithms, we develop a relevant set of tools which enable the practical study of convergence rates in the setting of Markov chain Monte Carlo methods, but also well beyond.

An Overview

- Sampling with the Random Walk Metropolis (RWM)
- Markov chain Mixing by Isoperimetric Analysis
- Applications to { Specific Targets, Other Samplers }

Sampling with RWM

Markov Chain Monte Carlo

- “target” distribution π on $\mathbf{R}^d$
- want samples from π to answer questions
- MCMC: use *iterative* strategy to obtain *approximate* samples
    - practically: want to converge in few iterations

$$X_0 \to X_1 \to X_2 \to \cdots \to X_N \overset{d}{\approx} \pi$$

$$\frac{1}{N} \sum_{0<n\leq N} f(X_n) \approx \int \pi(\mathrm{d}x) f(x) =: \pi(f)$$

Random Walk à la Metropolis

- take $Q\left(x, \mathrm{d}y\right) = \mathcal{N}\left(\mathrm{d}y; x, \sigma^2 \cdot \mathbf{I}_d\right)$
- 'accept' moves (from Q) with probability

$$\alpha\left(x, y\right) = \min\left\{1, \frac{\pi\left(y\right)}{\pi\left(x\right)}\right\}$$

- call this kernel $\mathsf{RWM}\left(\pi, \sigma^2\right)$; this samples π correctly
- only needs i) π up to a constant and ii) samples from $\mathcal{N}\left(0,1\right)$

Diffusion Limits for RWM

- taking $\sigma \to 0^{+}$ and rescaling $t \propto \sigma^2 \cdot n$, obtain limiting process

$$\mathrm{d}X_t = \nabla \log \pi \left(X_t \right) \, \mathrm{d}t + \sqrt{2} \mathrm{d}W_t$$

which is the **Overdamped Langevin Diffusion**, $\mathsf{OLD}\left(\pi\right)$

for $\sigma > 0$, can we infer that $\sigma^2 \cdot N_{\mathrm{mix}}^{\mathsf{RWM}\left(\pi, \sigma^2\right)} \lesssim T_{\mathrm{mix}}^{\mathsf{OLD}(\pi)}$?

Connecting RWM and OLD

- there appears to be a strong 'resemblance' between $\mathsf{RWM}(\pi)$, $\mathsf{OLD}(\pi)$
    - one expects that if $\alpha \gtrsim 1$, then indeed $\sigma^2 \cdot N_{\text{mix}}^{\mathsf{RWM}(\pi, \sigma^2)} \lesssim T_{\text{mix}}^{\mathsf{OLD}(\pi)}$
- what nature of 'resemblance' could make this rigorous?
    - for e.g. pathwise behaviour, not true 'uniformly enough' (c.f. ULA)
    - key similarity: *exit behavior*, *boundary behaviour*

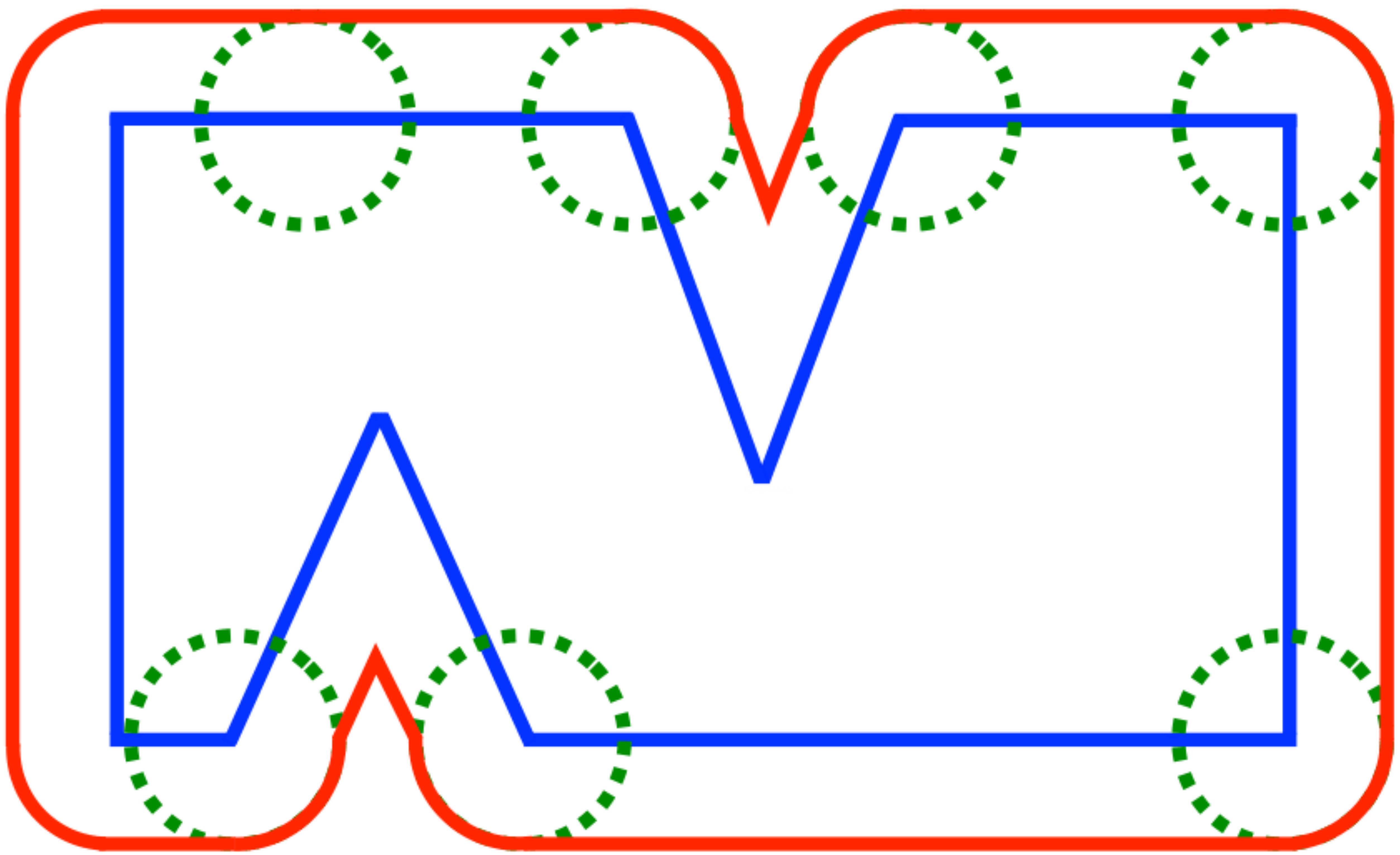
Isoperimetry for Markov Chains

Isoperimetry, Take 0

Enclosing a large area with a finite string

Isoperimetry 101

- in one sentence: *‘compare the mass of a set to the mass of its boundary’*
- a moment’s thought: small mass, large boundary is *easy*
- so, materially we are asking
  - *for a given mass, how small can the boundary be?*
  - *for a given boundary length, how large can the enclosed mass be?*



Isoperimetry for Probability Measures

{ r-enlargements, Minkowski content, … }

Probabilistic Isoperimetry

- with $A \subseteq \mathbf{R}^d$, take $A^r = \left\{ x \in \mathbf{R}^d : \text{dist}\,(x, A) \leq r \right\}$, and define

$$\pi^+ (A) := \lim\inf_{r \to 0^+} \frac{\pi \left(A^r \setminus A \right)}{r}$$

- let $I_\pi = I$ be maximal such that for any $0 \leq p \leq 1/2$,

$$\pi (A) = p \qquad \Longrightarrow \qquad \pi^+ (A) \geq I \left(p \right)$$

  - 'if mass $= p$, then boundary $\geq I \left(p \right)$'
    - (tough) exercise: what sort of sets A will be extremal here?

Dynamical Picture of Isoperimetry

- the definition of I_π at first seems quite 'static' …
  - but it equally furnishes a 'dynamic' interpretation:
    - let $X_0 \sim \pi_{\upharpoonright A}$ evolve by OLD (π).
    - then, what is the probability that $X_t \in A^\complement$, as $t \sim 0^+$?

isoperimetry characterises the difficulty for a diffusion to escape a set!

Mixing Time of OLD (π) via Isoperimetry

- under reasonable conditions on π, one can bound

$$T_{\mathrm{mix}}^{\mathsf{OLD}(\pi)}(\varepsilon) \lesssim \int_{\varepsilon/\Delta_0}^{1/2} \frac{p}{I_\pi\left(p\right)^2}\,\mathrm{d}p$$

where Δ_0 relates to the initialisation

- unified description for { faster-than, slower-than, … } exponential rates
- observe that rates are dictated by behaviour of I_π as $p \to 0^+$

From OLD (π) to RWM $\left(\pi, \sigma^2\right)$

- we see that isoperimetric analysis can be highly informative for OLD (π)

  can it also be informative for the convergence of RWM $\left(\pi, \sigma^2\right)$*?*

- our analysis ought to account for the 'discreteness' of RWM $\left(\pi, \sigma^2\right)$
  - we will see this is essentially the *only* additional obstacle

An Extra Ingredient

- for $\delta > 0$, $\tau \in (0,1)$, say that P is '(δ, τ)-close coupling' if

$$\mathsf{d}\left(x, y\right) \leq \delta \implies \mathsf{TV}\left(P_x, P_y\right) \leq 1 - \tau\,.$$

  - not a 'for all τ, there exists δ …' condition
    - … but still morally encodes 'continuity' / 'smoothness' of P
- operational interpretation:

"if we get within δ, then we can coalesce in one step w.p. $\geq \tau$"

Mixing Times via Isoperimetry

- **Proposition**: Let π have isoperimetric profile I_π, and let P be a π-reversible, positive Markov kernel which is (δ, τ)-close coupling. Then,

$$N_{\text{mix}}^{P}(\varepsilon) \lesssim \delta^{-2} \cdot \tau^{-2} \cdot \int_{\varepsilon/\Delta_0}^{1/2} \frac{p}{I_\pi\left(p\right)^2} \, \mathrm{d}p$$

- (all implied constants are made fully explicit in the papers)

- (usually, take $\tau \in \Theta(1)$ and ignore)

Acceptance Rate Control for RWM $\left(\pi, \sigma^2\right)$

- for $P = \mathrm{RWM}\left(\pi, \sigma^2\right)$, the close coupling property is related to *accepting moves*
- operationally, if we can lower bound

$$\begin{aligned}\alpha(x) &= \int Q\left(x, \mathrm{d}y\right) \cdot \alpha\left(x, y\right) \\ &= \int Q\left(x, \mathrm{d}y\right) \cdot \exp\left(-\left[U\left(y\right) - U(x)\right]_{+}\right)\end{aligned}$$

uniformly in x, then it follows with $(\delta, \tau) \asymp (\sigma, 1)$

- under some quantitative smoothness assumption on U, not too hard to prove

Obtaining Explicit Bounds for $\mathsf{RWM}\left(\pi, \sigma^2\right)$

- there is a nice 'division of labour' here: first, you write down π, and then
  - ask one friend to study the isoperimetry π, estimate I_π
  - ask another friend to study the smoothness of U, find σ so that $\alpha(x) \gtrsim 1$
- Finally, combine the estimates as

$$N_{\mathrm{mix}}^{\mathsf{RWM}\left(\pi, \sigma^2\right)}(\varepsilon) \lesssim \sigma^{-2} \cdot \int_{\varepsilon/\Delta_0}^{1/2} \frac{p}{I_\pi\left(p\right)^2} \, \mathrm{d}p$$

Applications of Main Theorems

- with $0 < m \leq \nabla^2 U(x) \leq L < \infty$, take $\sigma \asymp \left(L \cdot d\right)^{-\frac{1}{2}}$; get $N_{\mathrm{mix}}^{\mathsf{RWM}_\star(\pi)} \lesssim L \cdot d/m$
- with $\alpha > 0$, $0 < p < 2$, $U(x) = \|x\|_p^\alpha$, take $\sigma \sim d^{-1/p}$; get $N_{\mathrm{mix}}^{\mathsf{RWM}_\star(\pi)} \lesssim d^{2/p+2/\alpha-1}$
- with $\eta > 0$, product of `η-Cauchy', get $N_{\mathrm{mix}}^{\mathsf{RWM}_\star(\pi)} \lesssim d \cdot \left(d \cdot \frac{\Delta_0}{\varepsilon}\right)^{2/\eta}$
- With $\tau \gtrsim d$, multivariate Student-t, get $N_{\mathrm{mix}}^{\mathsf{RWM}_\star(\pi)} \lesssim d^2 \cdot \left(\frac{\Delta_0}{\varepsilon}\right)^{2/\tau}$
    - (in latter cases, Δ_0 captures (large) influence of initialisation)

Applications of Proof Techniques

- understanding RWM neatly feeds into results for more advanced algorithms
  - used towards guarantees for { RWM-within-Gibbs, Multiple-Proposal Methods, ‘Hybrid’ Chains, Parallel-in-Time RWM, Preconditioned RWM, pCN, … }
- general ‘isoperimetric’ approach (which is **not** ours!) is surprisingly widely applicable
  - same local-global decomposition used for { Gibbs Sampling (large-block and small-block), Hit-and-Run, Langevin Monte Carlo, Hamiltonian Monte Carlo, … }

Take-Aways

- Metropolis Algorithms for Monte Carlo Simulation
- Connections to the Langevin Diffusion
- Isoperimetric Problems for Probability Measures
- Non-Asymptotic Analysis of RWM Algorithm in Several Regimes
    - Global Picture: Isoperimetric Profile of π
    - Local Picture: Acceptance Rate Control from Smoothness of U

🎄 - - - Bonus Material - - - 🎅🏻

“How on earth can I get this talk down to twenty minutes?”

–Sam Power, 13 December 2025