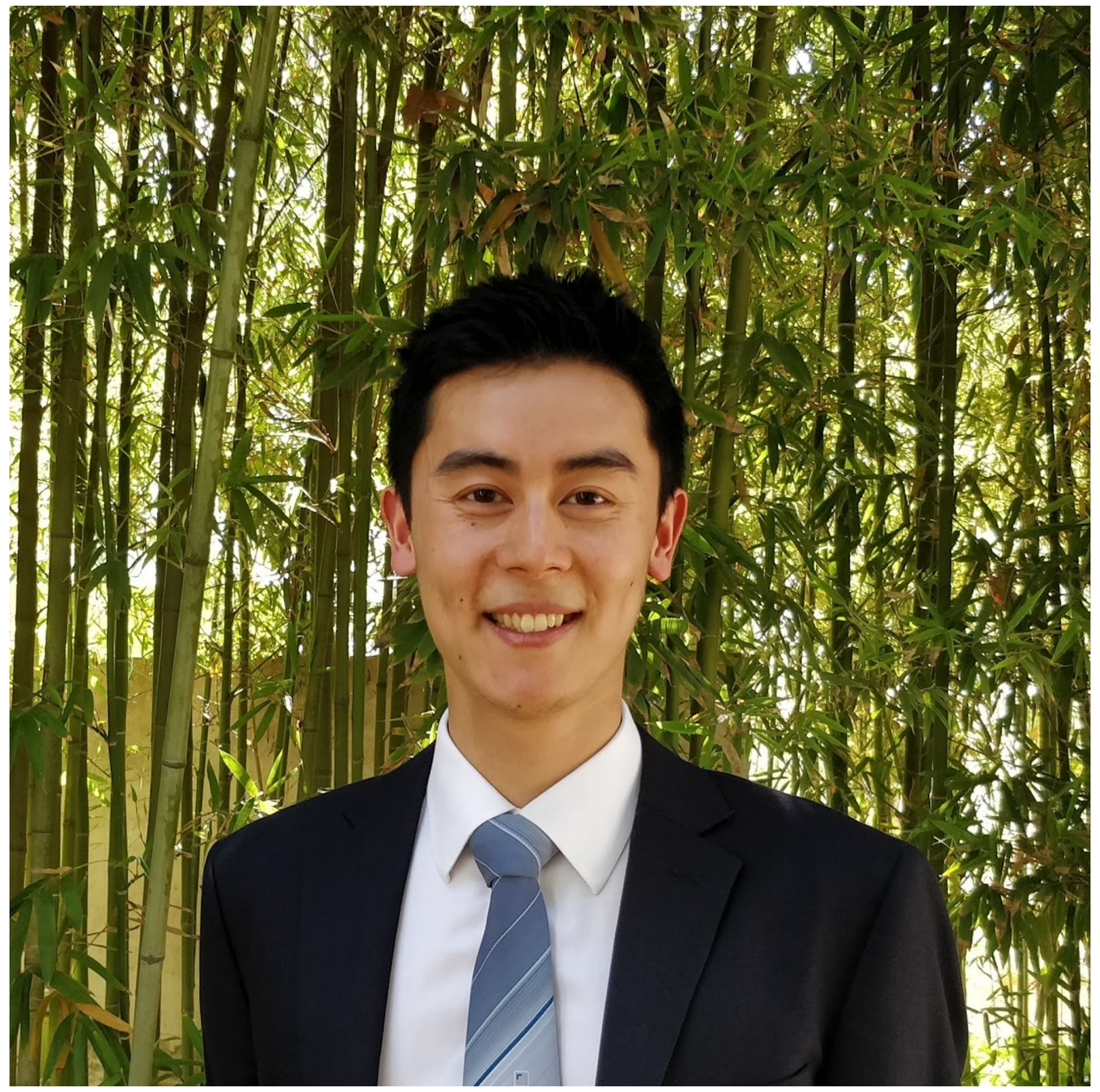
Convergence Bounds for the Random Walk Metropolis Algorithm

Perspectives from Isoperimetry

INI Programme: Stochastic Systems and Anomalous Diffusion
Workshop: Monte Carlo Sampling: Beyond the Diffusive Regime
25 November 2024

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.

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

Weak Poincaré Inequalities for Markov chains: theory and applications

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

School of Mathematics, University of Bristol and University of Warwick

December 20, 2023

Abstract

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 equiva-
lence with the Resolvent Uniform Positivity-Improving (RUPI) condition
g a counterexample. From a more practical perspective, we
elebrated Cheeger's inequalities to the subgeometric setting,
apply these techniques to study random-walk Metropolis al-
heavy-tailed target distributions and to obtain lower bounds
arginal algorithms.

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.

arXiv > math > arXiv:2312.11689

Search...

Help |

Mathematics > Probability

[Submitted on 18 Dec 2023]

Weak Poincaré Inequalities for Markov chains: theory and applications

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

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 \mathrm{L}^{2}-convergence, in particular by establishing equivalence with the Resolvent Uniform Positivity-Improving (RUPI) condition and providing a counterexample. From a more practical perspective, we extend the celebrated Cheeger's inequalities to the subgeometric setting, and further apply these techniques to study random-walk Metropolis algorithms for heavy-tailed target distributions and to obtain lower bounds on pseudo-marginal algorithms.

arXiv > math > arXiv:2208.05239

Search...

Help |

Mathematics > Probability

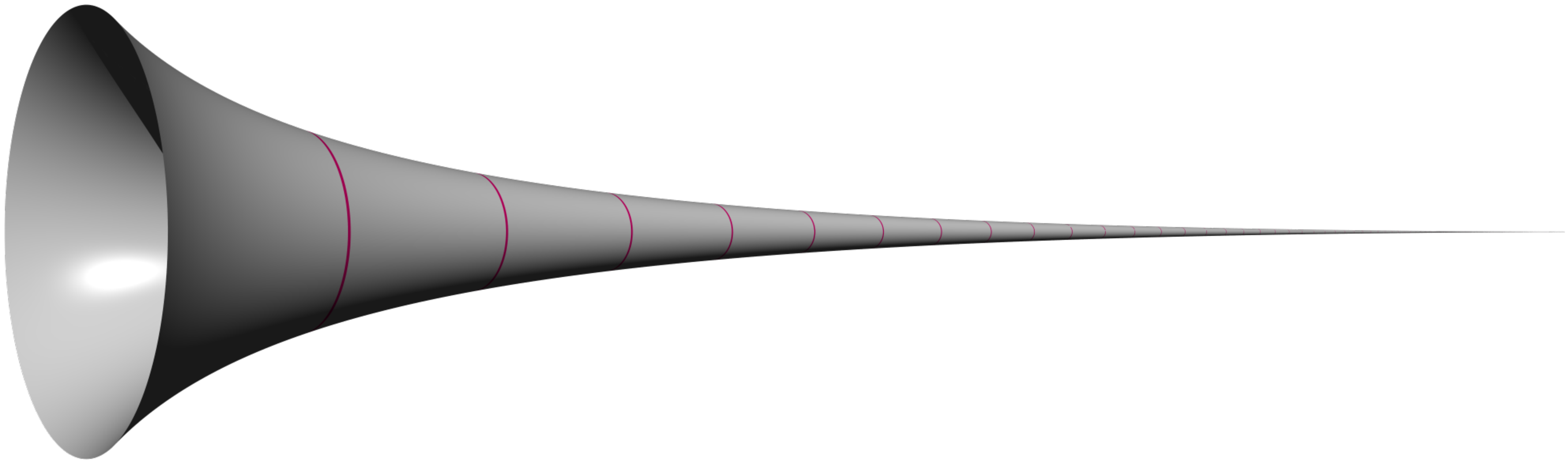
[Submitted on 10 Aug 2022]

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

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

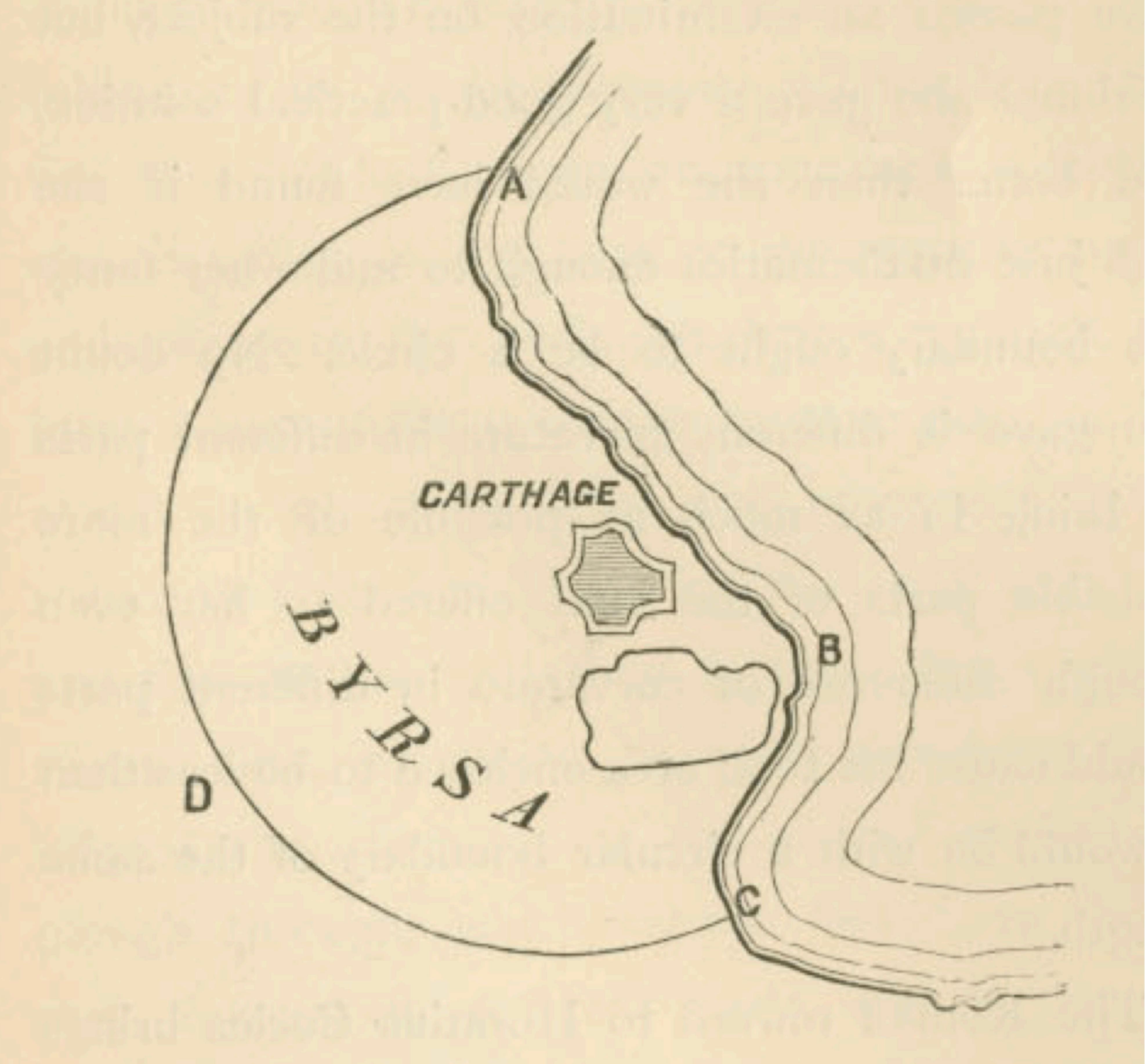
We develop a theory of weak Poincaré inequalities to characterize convergence rates of ergodic Markov chains. Motivated by the application of Markov chains in the context of algorithms, we develop a relevant set of tools which enable the practical study of convergence rates in the setting of Markov chain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

Some Vignettes on Isoperimetry



Isoperimetry, Take 1

Gabriel's Horn / Torricelli's Trumpet



Isoperimetry, Take 2

Dido’s Problem

Isoperimetry, Take 3

'Holding' Brownian Motion on a Sphere

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

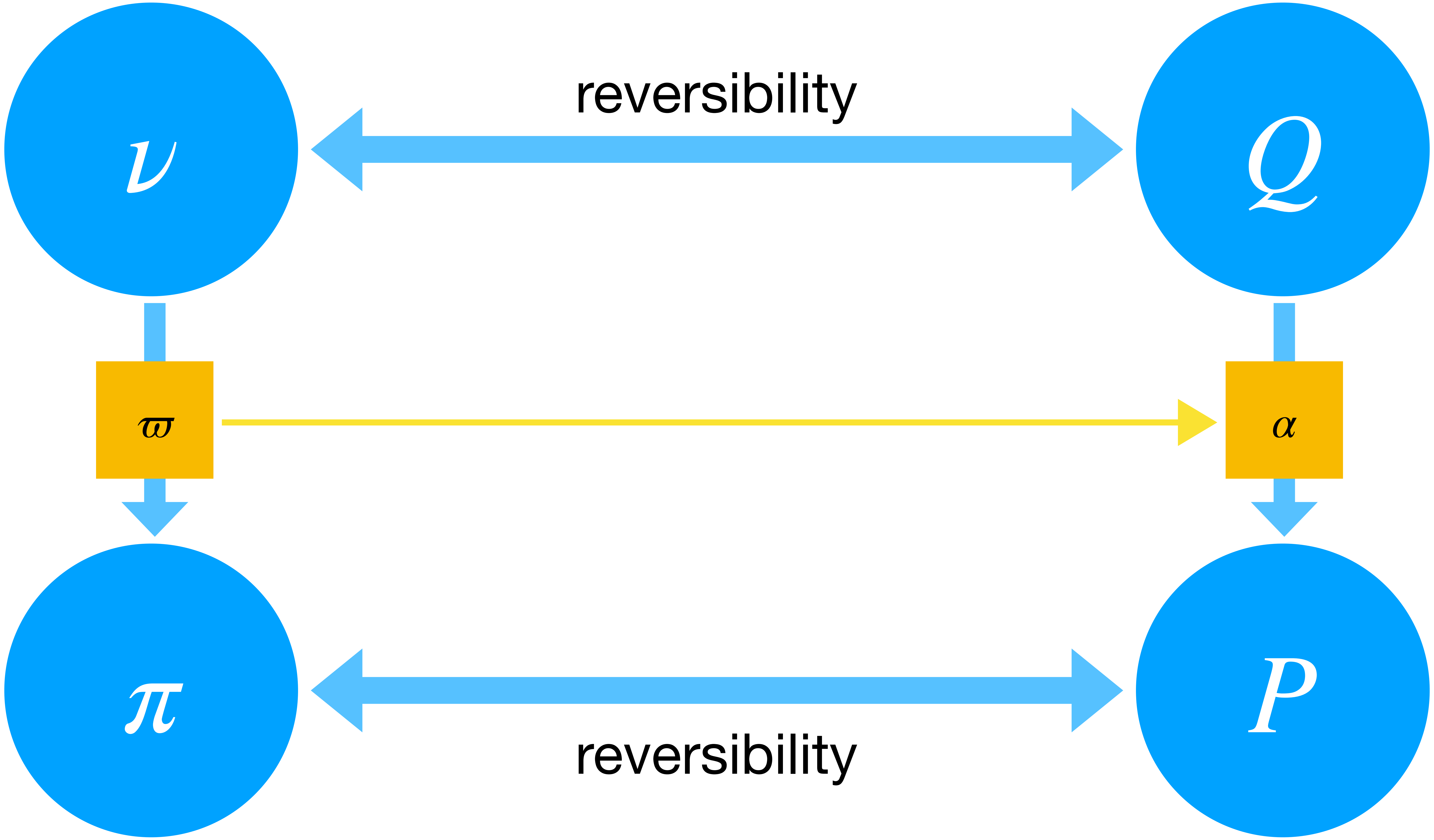
$$X_0 \to X_1 \to X_2 \to \cdots \to X_N \overset{d}{\approx} \pi$$

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

Sampling à la Metropolis

- generic recipe for constructing π-reversible kernels P:
  - start with simple $\nu \gg \pi$, and a ν-reversible kernel Q; write $\varpi = \mathrm{d}\pi/\mathrm{d}\nu$
  - propose moves with $y \sim Q\left(x, \mathrm{d}y\right)$
  - evaluate moves via $r\left(x, y\right) = \varpi\left(y\right)/\varpi\left(x\right)$

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


reversibility
ν
Q
ϖ
α
π
P
reversibility

Random Walk à la Metropolis

- take $\nu = \mathrm{Leb}$, $Q\left(x, \mathrm{d}y\right) = \mathcal{N}\left(\mathrm{d}y; x, \sigma^2 \cdot \mathbf{I}_d\right)$

- 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 RWM $\left(\pi, \sigma^2\right)$

- only needs π up to a constant and samples from $\mathcal{N}\left(0,1\right)$

First Results on RWM $\left(\pi, \sigma^2\right)$

- π-reversible under almost no conditions
- original focus more qualitative than quantitative:
  - ergodic under quite mild conditions
  - *exponentially* ergodic under *lighter-than-exponential* tails (roughly)
  - *slower-than-exponentially* ergodic under *heavier-than-exponential* tails

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 $T_{\mathrm{mix}}^{\mathsf{RWM}\left(\pi, \sigma^2\right)} \lesssim \sigma^{-2} \cdot T_{\mathrm{mix}}^{\mathsf{OLD}(\pi)}$?

The Overdamped Langevin Diffusion

- write target as $\pi \propto \exp(-U)$; call U the 'potential'

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

- straightforward to check that this process is π-reversible, hence invariant
- perhaps embarrassingly (for me), at least in the context of sampling …

there is almost no diffusion *less anomalous.*

The Overdamped Langevin Diffusion

- OLD (π) is somehow a ‘canonical’ Markov process
  - { geometry, concentration of measure, transport, … }
- many aspects are very well-understood by now

Crash Course on OLD (π)

(all convergence described is in an L^2 sense)

1. For U convex, convergence is *exponentially fast*

2. For U uniformly convex, convergence is initially *faster-than-exponential*.

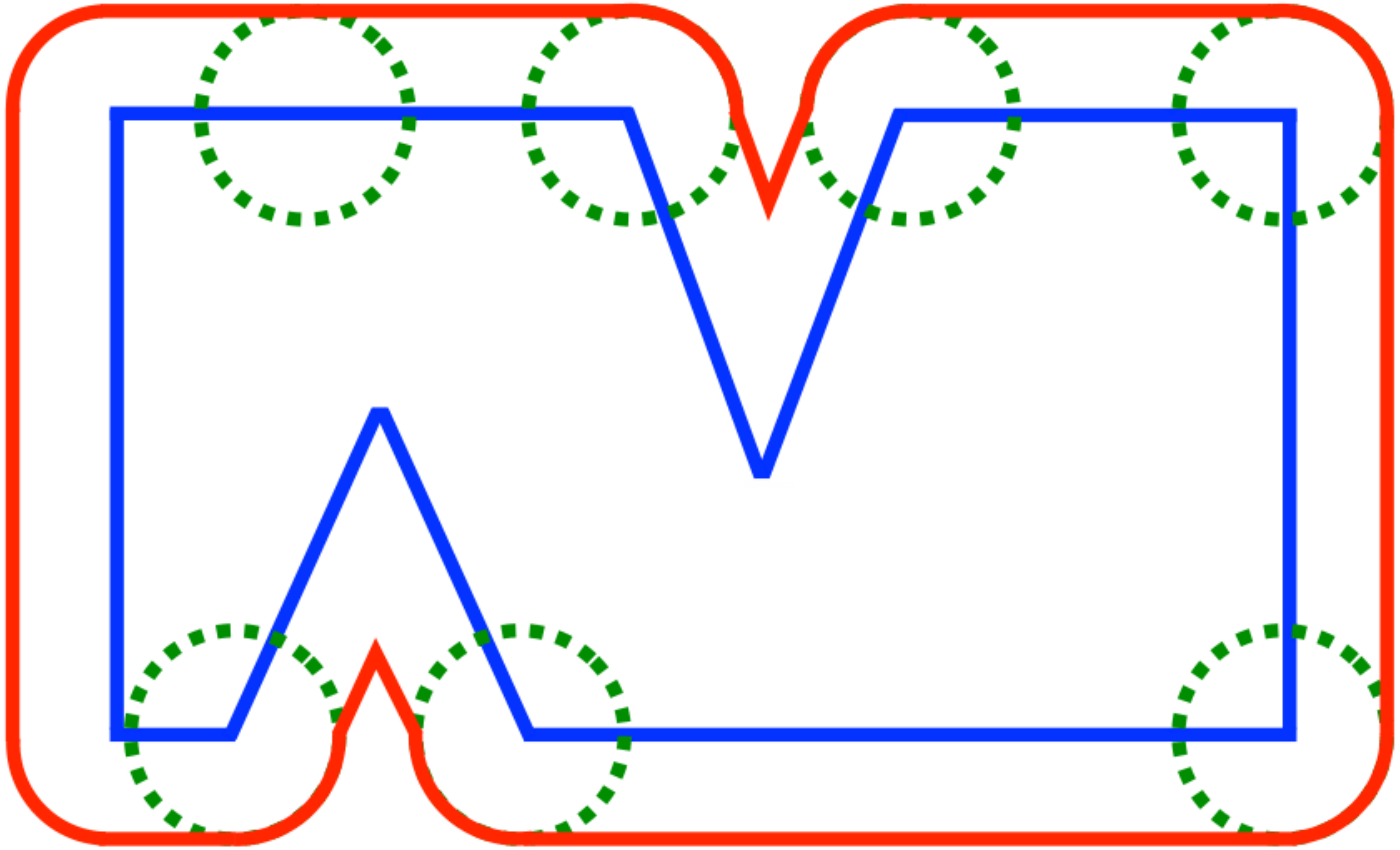
3. For U of sublinear growth, convergence must be *slower-than-exponential*.

4. For U convex, exponential convergence rate is *conjecturally dimension-free.*

5. $\exists$ transfer principles (bounded change-of-measure, Lipschitz transport, $\ldots$)

Connecting RWM and OLD

- there appears to be a strong ‘resemblance’ between RWM (π), OLD (π)
    - one expects that if $\alpha \gtrsim 1$, then indeed $T_{\mathrm{mix}}^{\mathsf{RWM}\left(\pi, \sigma^2\right)} \lesssim \sigma^{-2} \cdot T_{\mathrm{mix}}^{\mathsf{OLD}(\pi)}$
- 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*



Back to Isoperimetry

{ 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

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

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

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

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

Examples of Isoperimetric Profiles

- model problem: polynomial potentials, light tails

$$\pi(x) \propto \exp\left(-|x|\right) \implies I\left(p\right) \propto p$$

$$\pi(x) \propto \exp\left(-|x|^2\right) \implies I\left(p\right) \gtrsim p \cdot \sqrt{\log\left(1/p\right)}$$

$$\pi(x) \propto \exp\left(-|x|^\alpha\right) \implies I\left(p\right) \gtrsim p \cdot \left(\log\left(1/p\right)\right)^{1-1/\alpha}$$

- actually, this is all still true for products of the above

- note that behaviour for small sets can be much better (i.e. $I\left(p\right) \gg p$)

Examples of Isoperimetric Profiles

- model problem: heavy-tailed problems ($\alpha \in (0,1)$, $\eta > 0$)

$$\pi(x) \propto \prod_i \exp\left(-\left|x_i\right|^\alpha\right) \implies I\left(p\right) \gtrsim p \cdot \left(\log\left(d/p\right)\right)^{1-1/\alpha}$$

$$\pi(x) \propto \exp\left(-\left|x\right|_2^\alpha\right) \implies I\left(p\right) \gtrsim c_{d,\alpha} \cdot p \cdot \left(\log\left(1/p\right)\right)^{1-1/\alpha}$$

$$\pi(x) \propto \prod_i \left(1 + \left|x_i\right|\right)^{-(1+\eta)} \implies I\left(p\right) \gtrsim d^{-1/\eta} \cdot p^{1+1/\eta}$$

$$\pi(x) \propto \left(1 + \left|x\right|_2\right)^{-(1+\eta)} \implies I\left(p\right) \gtrsim c_{d,\eta} \cdot p^{1+1/\eta}$$

- note that behaviour for small sets can be much worse (i.e. $I\left(p\right) \ll p$)
- note also that tensorisation makes things worse (c.f. convex case)

Obtaining Isoperimetric Estimates

- in one dimension, picture is rather complete (Bobkov, Houdré, …)
- Lyapunov conditions (Cattiaux, Guillin, F.Y. Wang, L.M. Wu, …)
- generic transfer principles (change-of-measure, transport, …)
- under convexity,
  - I_π tensorises nicely (Bobkov, …)
  - closely related to functional inequalities (E. Milman, M. Ledoux, …)

Dynamical Picture of Isoperimetry

- the definition of I_π at first seems quite ‘static’ …
  - but it equally furnishes a ‘dynamic’ interpretation:
    - let $X_0 \sim \pi_{\restriction A}$ 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

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

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 is still morally a 'continuity' / 'smoothness' condition on 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,

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

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

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

Close Coupling for RWM $\left(\pi, \sigma^2\right)$

- can we show that RWM $\left(\pi, \sigma^2\right)$ satisfies a close coupling condition?

   1. for **any** Metropolis kernel, bound $\mathsf{TV}\left(P_x, P_y\right) \leq \mathsf{TV}\left(Q_x, Q_y\right) + \bar{\rho}$

      where $\bar{\rho}$ is the worst-case rejection rate.

   2. for $Q\left(x, \mathrm{d}y\right) = \mathcal{N}\left(\mathrm{d}y; x, \sigma^2 \cdot \mathbf{I}_d\right)$, bound $\mathsf{TV}\left(Q_x, Q_y\right) \leq \dfrac{\mathsf{d}\left(x, y\right)}{2 \cdot \sigma}$

- take-away: it will be sufficient to bound $\underline{\alpha} = 1 - \bar{\rho}$ away from 0

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

- define

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

- we want to lower-bound $\alpha(x) \geq \underline{\alpha} > 0$, uniformly in x

- natural to make some quantitative smoothness assumption on U

From Smoothness to $\underline{\alpha}$

- assume that for some symmetric $\psi \geq 0$, it holds that

$$\forall x, h \in \mathbf{R}^d, \qquad U\left(x+h\right) \leq U(x) + U'(x)\,h + \psi\left(h\right)$$

- it then follows that

$$\underline{\alpha} \geq \frac{1}{2} \cdot \exp\left(- \int \mathcal{N}\left(\mathrm{d}z; 0, \mathbf{I}_d\right) \cdot \psi\left(\sigma \cdot z\right) \right)$$

- given a specific ψ, tune $\sigma = \sigma\left(d; \psi\right)$ to stabilise $\underline{\alpha} \geq 1/4$ (e.g.); easy

Obtaining Explicit Bounds for 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 $\underline{\alpha} \gtrsim 1$
- Finally, combine the estimates as

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

Application: Log-Concave Sampling

- consider 'well-conditioned' convex U,
  - i.e. for some $0 < m \leq L < \infty$, uniformly in $x \in \mathbf{R}^d$,

$$\text{eigs}\left(\text{Hess}U\left(x\right)\right) \subseteq [m, L]$$

  - then

$$I\left(p\right) \gtrsim m^{1/2} \cdot p \cdot \sqrt{\log\left(1/p\right)}, \qquad \psi\left(h\right) \leq \frac{L}{2} |h|_2^2$$

  - taking $\sigma \sim \left(L \cdot d\right)^{-1/2}$ yields $T_{\text{mix}}^{\text{RWM}(\pi)} \lesssim \kappa \cdot d$, where $\kappa = L/m$

Application: ℓ^p-type Targets

- consider $U(x) = \|x\|_p^\alpha$ for $\alpha, p > 0$, taking $\sigma \sim d^{-1/p}$
  - $\alpha = p = 2$ (Gaussian) gives $T_{\mathrm{mix}}^{\mathsf{RWM}(\pi)} \lesssim d$
  - $\alpha \in [1,2], p = 2$ ('spherical Subbotin') gives $T_{\mathrm{mix}}^{\mathsf{RWM}(\pi)} \lesssim d^{2/\alpha}$
  - $\alpha, p \in [1,2]$ ('ℓ^p-type Subbotin') gives $T_{\mathrm{mix}}^{\mathsf{RWM}(\pi)} \lesssim d^{2/p+2/\alpha-1}$
- one factor for 'concentration', one factor for 'roughness' (dimensional)

Application: Heavy-Tailed Targets

- consider product-form targets with $U(x) = \sum_i U_0(x_i)$
  - $U_0(\bar{x}) = \frac{1+\eta}{2} \cdot \log\left(1 + |\bar{x}|^2\right)$ gives $T_{\text{mix}}^{\mathsf{RWM}(\pi)} \lesssim d \cdot \left(d \cdot \frac{\Delta_0}{\varepsilon}\right)^{2/\eta}$
  - $U_0(\bar{x}) = \left(\tau + |\bar{x}|^2\right)^{\eta/2}$ gives $T_{\text{mix}}^{\mathsf{RWM}(\pi)} \lesssim d \cdot \left(\log\left(d \cdot \frac{\Delta_0}{\varepsilon}\right)\right)^{2/\eta-1}$
- Student-t: $U(x) = \frac{d+\tau}{2} \cdot \log\left(\tau + |x|^2\right)$ with $\tau \gtrsim d$ gives $T_{\text{mix}}^{\mathsf{RWM}(\pi)} \lesssim d^2 \cdot \left(\frac{\Delta_0}{\varepsilon}\right)^{2/\eta}$
- in general, ’feasible start’ has $\Delta_0 = \exp\left(\Theta\left(d\right)\right)$, so initialisation ‘hurts more’ with heavy tails

Take-Aways

- Isoperimetric Problems for Probability Measures
- Metropolis Algorithms for Monte Carlo Simulation
- Connections to the Langevin Diffusion
- 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 - - -

RWM Stuff

- other product-form proposals
- analysis of preconditioned Crank-Nicolson sampler
- other consequences (e.g. asymptotic variance, CLT)
- sharpness of estimates in various settings
- open: super-quadratic U

Functional Inequalities Stuff

- more on Poincaré: Standard, Weak, and Super
- connections to { conductance methods, Cheeger, … }
- applications to other MCMC algorithms (PMMH, ABC-MCMC, …)