Uniform Ergodicity and Spectral Gap of Importance-tempered RWM

Quan Zhou

Department of Statistics, Texas A&M University

Main Results

Consider importance-tempered random-walk Metropolis (RWM) with

$$\pi_\beta(x) \propto \pi(x)^\beta, \qquad w(x) \propto \pi(x)^{1-\beta}.$$

For targets with tails of order $\|x\|^{-\gamma}$:

- We identify the exact range of β for uniform ergodicity.
- We reduce the global spectral-gap problem to a bounded domain.
- Simulations show less sensitivity to initialization and smaller variance.

Importance Sampling (IS)

π : target probability distribution; μ : trial probability distribution.

$$\int f \, \mathrm{d}\pi = \int \left(f \frac{\mathrm{d}\pi}{\mathrm{d}\mu} \right) \mathrm{d}\mu.$$

$w = \mathrm{d}\pi / \mathrm{d}\mu$ is the *importance weight*.

Estimating the expectation of f with samples from π
$\Longrightarrow$ estimating the expectation of fw with samples from μ

Importance Sampling Estimators

Let $X_i \sim \mu$. The IS estimator is

$$\widehat{\Pi}_{\mu,n}(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_i)w(X_i).$$

The self-normalized importance sampling (SNIS) estimator is

$$\widetilde{\Pi}_{\mu,n}(f) = \frac{\sum_{i=1}^{n} f(X_i)w(X_i)}{\sum_{i=1}^{n} w(X_i)}.$$

w only needs to be evaluated up to a normalizing constant.

Combining IS and MCMC

Deliberately choose a biased stationary distribution for an MCMC sampler and compute the SNIS estimator.
We call this scheme *importance-tempered MCMC* [3, 16].

A simple choice is

$$\pi_\beta(x) \propto \pi(x)^\beta, \qquad \beta \in (0, 1).$$

The same MCMC algorithm can usually be used for π and π_β.

Importance-tempered RWM

At iteration i, given $X_{i-1} = x$:

1. Draw $Z_i \sim \kappa$ and propose $x^* = x + \sigma Z_i$.
2. Accept x^* with probability

$$1 \wedge \frac{\pi(x^*)^\beta}{\pi(x)^\beta}.$$

3. Record X_i and the weight

$$\widetilde{w}_i = \pi(X_i)^{1-\beta}.$$

The only extra step is computing $\widetilde{w}_i$.

Setup for Theoretical Analysis

$$\widetilde{\Pi}_{\beta,n}(f) = \frac{\sum_{i=1}^{n} f(X_i) w(X_i)}{\sum_{i=1}^{n} w(X_i)}, \qquad w(x) \propto \pi(x)^{1-\beta}.$$

View $w(X_i)$ as the *time* the chain stays at X_i.
Replacing it by an exponential holding time with mean $w(X_i)$ gives a continuous-time Markov chain with generator

$$(\mathscr{A} g)(x) = \frac{1}{w(x)} \int_{\mathbb{R}^d} [g(y) - g(x)] \mathcal{T}(x, \mathrm{d}y),$$

where $\mathcal{T}$ is the transition kernel of the tempered RWM chain.

The same representation also appears in [16, 8].

Uniform and Geometric Ergodicity

Definition

A Markov process $(Y_t)_{t\geq 0}$ with invariant distribution Π is geometrically ergodic if, for each x, there exist $C(x) < \infty$ and $\theta \in (0, 1)$ such that

$$\|\mathrm{Law}(Y_t \mid Y_0 = x) - \Pi\|_{\mathrm{TV}} \leq C(x)\theta^t, \qquad t > 0.$$

If $\sup_x C(x) < \infty$, then $(Y_t)_{t\geq 0}$ is uniformly ergodic.

Uniform ergodicity uses the same bound for *all initial states*.

Ergodicity of RWM

For a positive continuous target on $\mathbb{R}^d$, RWM cannot be uniformly ergodic [9].

With heavy-tailed targets, it is generally not geometrically ergodic either [9, 12, 6].

Starting far in the tails, RWM may need an arbitrarily long time to reach the high-density region.

Langevin-type Samplers for Heavy-tiled Targets

Stramer and Tweedie [13] and Roberts and Stramer [11] accelerate a Langevin diffusion in the tails.

The corresponding diffusion matrix is proportional to

$$\pi(x)^{\beta-1}\boldsymbol{I}.$$

This is the same tail acceleration as the rate $w(x)^{-1}$ in our continuous-time chain. But discretization and Metropolis adjustment may change the chain's behavior significantly [14, 10].

Other Methods for Heavy-tailed Targets

Several approaches modify how the sampler moves in the tails:

- heavy-tailed proposal distributions [5];
- smooth variable transformations [7];
- stereographic projection [15];
- sub-Cauchy projection [4].

Stereographic methods *compress the space*; importance tempering *compresses the time*.

Assumptions

Suppose

$$\pi(x) = \exp\{\psi(\|x\|)\}, \qquad r\psi'(r) \longrightarrow -\gamma,$$

with $\gamma > d$ and a mild second-derivative bound.

Thus $\pi(x)$ has polynomial tails of order γ, and π_β is integrable only if $\beta > d/\gamma$.

The proposal density κ is symmetric, positive near zero, and satisfies

$$\mathbb{E}(ZZ^\top) = \boldsymbol{I}_d, \qquad \mathbb{E}\|Z\|^4 < \infty.$$

Ergodicity of Importance-tempered RWM

Let $(Y_t)_{t\geq 0}$ be the continuous-time chain associated with importance-tempered RWM.

Theorem

Let $\gamma > d+2$. Then $(Y_t)_{t\geq 0}$ is uniformly ergodic if and only if

$$\frac{d}{\gamma} < \beta < \frac{\gamma - 2}{\gamma}.$$

Multivariate t-distribution

For the d-dimensional t-distribution with m degrees of freedom,

$$\pi(x) \propto \left(1 + \frac{\|x\|^2}{m}\right)^{-(m+d)/2}, \qquad \gamma = m + d.$$

Hence, the continuous-time chain is uniformly ergodic if and only if

$$\frac{d}{d+m} < \beta < \frac{d+m-2}{d+m}.$$

This interval is nonempty exactly when $m > 2$.

Effect of Initialization

Let τ_R be the first time $(Y_t)_{t\geq 0}$ enters $\mathcal{B}_R$. The drift condition gives

$$\sup_x \mathbb{E}_x[e^{\alpha_R \tau_R}] \leq 2, \qquad \sup_x \mathbb{E}_x[\tau_R] \leq \alpha_R^{-1}.$$

The importance weight accumulated in the tails is uniformly controlled over all initial states.

The number of RWM iterations needed to enter $\mathcal{B}_R$ is *not* uniformly controlled.

Exponentially Light Tails

Corollary

Suppose

$$\pi(x) \propto \exp(-\|x\|^a), \qquad a > 0.$$

Then $(Y_t)_{t\geq 0}$ *is uniformly ergodic for every* $\beta \in (0,1)$.

Standard RWM is generally only geometrically ergodic in this setting.

Proof of Uniform Ergodicity

For sufficiency, find a bounded function V such that

$$(\mathscr{A}V)(x) \leq -\alpha_R V(x), \qquad \|x\| > R.$$

The ball $\mathcal{B}_R$ is petite, so the bounded drift function yields uniform ergodicity [2].

For necessity, uniform ergodicity would imply

$$\sup_x \mathbb{E}_x[\tau_R] < \infty.$$

A second drift function shows that this is impossible above the threshold.

Drift Functions

For sufficiency, use

$$V(x) = 1 + (R/2)^{-\nu} - \max\{\|x\|, R/2\}^{-\nu},$$

where

$$\nu = \frac{1}{2}\min\{1, \gamma(1-\beta) - 2\} > 0.$$

For necessity, use

$$\widetilde{V}(x) = \log(1 + \|x\|^2).$$

The two functions give the exact upper threshold $\beta = (\gamma - 2)/\gamma$.

Spectral Gap

For a reversible rate kernel $\mathcal{Q}$, define

$$\mathcal{E}_{\mathcal{Q}}(f,f) = \frac{1}{2}\int [f(y) - f(x)]^2 \mathcal{Q}(x, \mathrm{d}y)\Pi(\mathrm{d}x),$$

and

$$\mathrm{Gap}(\mathcal{Q}) = \inf_{\mathrm{Var}_\Pi(f)>0} \frac{\mathcal{E}_{\mathcal{Q}}(f,f)}{\mathrm{Var}_\Pi(f)}.$$

For heavy-tailed targets, the unbounded tail region is usually the main difficulty.

Local and Global Spectral Gaps

Let $\mathcal{Q}_R$ be the restriction of $\mathcal{Q}$ to $\mathcal{B}_R$, and let

$$p_R = \Pi(\mathcal{B}_R), \qquad \lambda_R = \mathrm{Gap}(\mathcal{Q}_R).$$

Suppose a drift condition holds:

$$\int_{\mathbb{R}^d} \left[\frac{V(y)}{V(x)} - 1 \right] \mathcal{Q}(x, \mathrm{d}y) \leq -\alpha_R \mathbf{1}_{\mathcal{B}_R^c}(x) + b_R \mathbf{1}_{\mathcal{B}_R}(x).$$

If $2b_R(1 - p_R) \leq \alpha_R p_R$, then

$$\mathrm{Gap}(\mathcal{Q}) \geq \lambda_R \left[1 + \frac{2(\lambda_R + b_R)}{\alpha_R p_R} \right]^{-1}.$$

This extends a localization result for Langevin diffusions [1].

Spectral Gap of Importance-tempered RWM

Let Π be the multivariate t-distribution with m degrees of freedom and set $\gamma = m + d$.

Theorem

Suppose $m > 2 + 2\epsilon$ *and choose*

$$\beta = \frac{\gamma - 2 - \epsilon}{\gamma}.$$

For all sufficiently large $R = \mathrm{poly}(d)$,

$$\mathrm{Gap}(\mathcal{K}) \geq \frac{\sigma^2 \epsilon}{65 R^2} \mathrm{Gap}(\mathcal{K}_R).$$

Polynomial Localization

If m is fixed and $\sigma^2 = d^{-\eta}$, we may choose

$$R = O(d^a), \qquad a = \max\left\{ \frac{m+2\eta}{2(m-2)},\ 4 - \frac{\eta}{2} \right\}.$$

Thus the global problem on $\mathbb{R}^d$ is reduced to a ball of polynomial radius.

If $\mathrm{Gap}(\mathcal{K}_R)$ is polynomial in d, so is $\mathrm{Gap}(\mathcal{K})$.

Standard RWM has no positive spectral gap for polynomial-tailed targets.

Simulation I

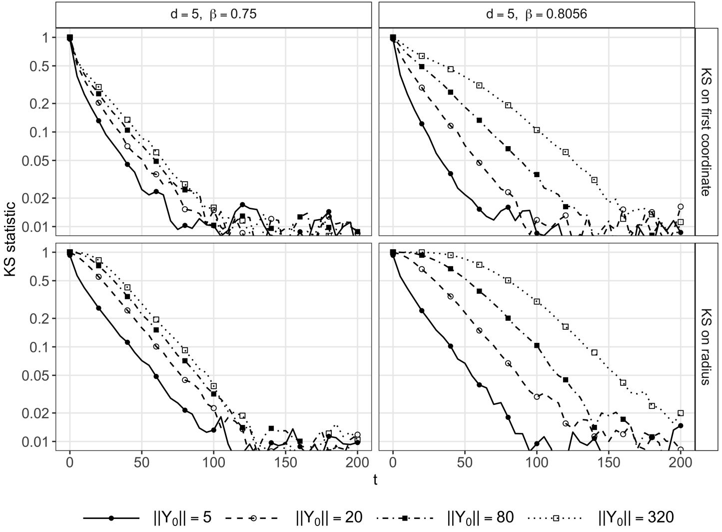
Let Π be the d-dimensional t-distribution with $m = 4$. The threshold is

$$\beta^* = \frac{d+2}{d+4}.$$

We compare values just below and above β^* for $d = 1, 5, 10$.

- Four initial radii growing geometrically.
- 10^4 independent trajectories.
- KS statistics for $Y_{t,1}$ and $\|Y_t\|$.

Uniform Ergodicity in Simulation



Below β^*, convergence is much less sensitive to the initial state.

Simulation II

Let Π be the d-dimensional t-distribution with $m = 5$ and

$$\beta = \frac{d + 0.5k}{d + 5}, \qquad k = 1, \ldots, 10.$$

The uniform-ergodicity threshold is $\beta^* = (d + 3)/(d + 5)$.
We consider three initializations and four test functions:

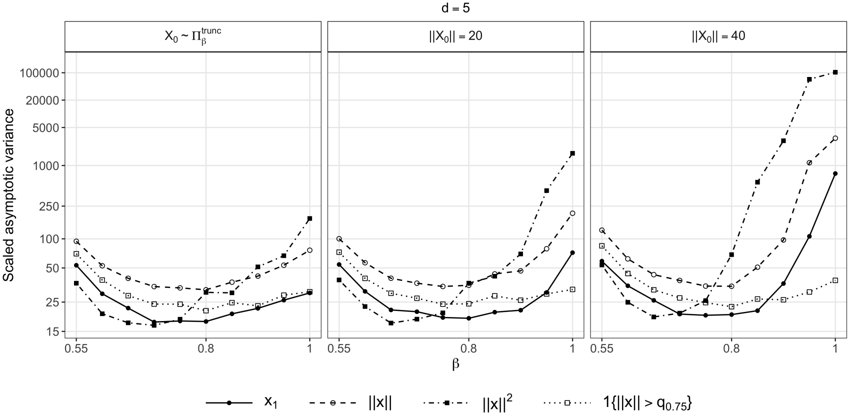
$$x_1, \qquad \|x\|, \qquad \|x\|^2, \qquad 1_{\{\|x\| > q_{0.75}\}}.$$

Using $1,000$ runs, we estimate

$$\mathrm{AV}(f) = \frac{n}{\mathrm{Var}_{\Pi}(f)} \widehat{\mathrm{Var}}\{\widetilde{\Pi}_{\beta,n}(f)\}.$$

No burn-in samples are removed.

Asymptotic Variance



An appropriate β substantially reduces variance, especially with poor initialization.

Choice of β

- For most test functions, the best β is close to β^*.
- For $f(x) = \|x\|^2$, a smaller β is preferable.
- Below β^*, the estimator is much less sensitive to initialization.

If β is too small, the discrete-time chain may need many iterations to leave the tails.

Values *slightly below β^** appear to give a good trade-off.

Concluding Remarks

- Importance tempering is a nearly zero-cost modification of RWM.
- The continuous-time chain can be uniformly ergodic for heavy-tailed targets.
- Its global spectral gap is controlled by a restricted spectral gap.
- A suitable β improves precision and robustness to initialization.

The same idea can be used with other MCMC samplers.

Acknowledgment

This work was supported by the National Science Foundation through grants DMS-2245591 and DMS-2311307.

References I

[1] Dominique Bakry, Franck Barthe, Patrick Cattiaux, and Arnaud Guillin. A simple proof of the Poincaré inequality for a large class of probability measures. *Electronic Communications in Probability*, 13: 60–66, 2008.

[2] Douglas Down, Sean P Meyn, and Richard L Tweedie. Exponential and uniform ergodicity of Markov processes. *The Annals of Probability*, 23(4):1671–1691, 1995.

[3] Robert Gramacy, Richard Samworth, and Ruth King. Importance tempering. *Statistics and Computing*, 20:1–7, 2010.

[4] Sebastiano Grazzi, Sifan Liu, Gareth O Roberts, and Jun Yang. Sub-Cauchy sampling: Escaping the dark side of the moon. *arXiv preprint arXiv:2601.11066*, 2026.

[5] Søren F Jarner and Gareth O Roberts. Convergence of heavy-tailed Monte carlo Markov chain algorithms. *Scandinavian Journal of Statistics*, 34(4):781–815, 2007.

[6] Søren Fiig Jarner and Ernst Hansen. Geometric ergodicity of Metropolis algorithms. *Stochastic processes and their applications*, 85 (2):341–361, 2000.

[7] Leif T Johnson and Charles J Geyer. Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm. *The Annals of Statistics*, pages 3050–3076, 2012.

[8] Samuel Livingstone, Giorgos Vasdekis, and Giacomo Zanella. Foundations of locally-balanced Markov processes. *arXiv preprint arXiv:2504.13322*, 2025.

[9] Kerrie L Mengersen and Richard L Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *The annals of Statistics*, 24 (1):101–121, 1996.

[10] Gareth O Roberts and Jeffrey S Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71, 2004.

References III

[11] Gareth O Roberts and Osnat Stramer. Langevin diffusions and Metropolis-Hastings algorithms. *Methodology and computing in applied probability*, 4:337–357, 2002.

[12] Gareth O Roberts and Richard L Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1):95–110, 1996.

[13] O Stramer and RL Tweedie. Langevin-type models i: Diffusions with given stationary distributions and their discretizations. *Methodology and Computing in Applied Probability*, 1(3):283–306, 1999.

[14] O Stramer and RL Tweedie. Langevin-type models ii: Self-targeting candidates for MCMC algorithms. *Methodology and Computing in Applied Probability*, 1(3):307–328, 1999.

[15] Jun Yang, Krzysztof Łatuszyński, and Gareth O Roberts. Stereographic Markov chain Monte Carlo. *The Annals of Statistics*, 52(6):2692–2713, 2024.

References IV

[16] Giacomo Zanella and Gareth Roberts. Scalable importance tempering and Bayesian variable selection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 81(3):489–517, 2019.