Journal of Machine Learning Research 25 (2024) 110-117 Submitted 8/24; Revised 10/24; Published 12/24

Cycle Structure and Stationary Bias in Perturbed Metropolis–Hastings Kernels

Siddharth Nand SIDDHARTH.NAND@UBC.CA
Department of Statistics
University of British Columbia
Vancouver, Canada

Ortwin Athelstan
Department of Statistics
University of British Columbia
Vancouver, Canada

Editor: Csaba Szepesvari

Abstract

The same numerical error in a Metropolis–Hastings (MH) log ratio can either change the target or create probability currents. On a finite reciprocal proposal graph, the distinction is exact: an antisymmetric error is a target tilt if and only if its sum around every cycle is zero. We use this criterion to separate a state potential from a cycle residual. Under a Dobrushin contraction assumption, the stationary-law error is bounded by the sharp total-variation effect of the tilt plus a residual term measured at the acceptance-probability level. A three-state cycle shows why nonzero circulation need not alter the stationary marginal, even though it destroys reversibility. A two-state MH construction attains the inverse-contraction dependence in the residual term.

Keywords: Metropolis–Hastings, approximate Markov chains, perturbation bounds, total variation, invariant distributions

1 Introduction

Metropolis–Hastings (MH) algorithms depend on a log acceptance ratio. Errors in that quantity need not have the same effect on the sampler. If the error has the form $\psi(y)-\psi(x)$, the resulting transition kernel is exactly an MH kernel for a target tilted by e^{ψ}. An error that cannot be written as a state-potential difference does not have this interpretation in general. We use the standard MH construction and its general-state-space formulation (Hastings, 1970; Tierney, 1994).

This distinction suggests analyzing an acceptance-ratio error ε as

$$\varepsilon(x, y) = \psi(y) - \psi(x) + r(x, y),$$

where ψ describes a target change and r is a residual transition error. The decomposition alone is not a perturbation theorem: the residual must be measured in a norm that controls the transition kernel, and the effect of that kernel perturbation depends on stability of the reference chain.

©2024 Siddharth Nand and Ortwin Athelstan.

License: CC-BY 4.0, see https://creativecommons.org/licenses/by/4.0/. Attribution requirements are provided at http://jmlr.org/papers/v25/21-1205.html.

The results address two different questions. First, when does a pairwise log-ratio perturbation still define a reversible MH chain for some target? Second, when it does not, how much does a chosen potential–residual split say about its invariant law? The finite-graph criterion answers the first question exactly. The perturbation theorem answers the second under a strong one-step contraction assumption; it does not assert that random-walk or Langevin MH kernels on unbounded spaces satisfy that assumption.

2 Related literature

General perturbation theory controls the effect of replacing a Markov kernel by a nearby kernel. Total-variation sensitivity results for uniformly ergodic chains and Wasserstein bounds under weaker ergodicity conditions provide the stability mechanism used here (Mitrophanov, 2005; Rudolf and Schweizer, 2018). Approximate MH work, including noisy transitions and Monte Carlo within Metropolis, also supplies finite-time and invariant-law comparisons (Alquier et al., 2016; Medina-Aguayo et al., 2015, 2020). These results address general kernel error or approximate accept–reject steps; they do not distinguish an error that merely changes the invariant target from one that changes transition dynamics without preserving reversibility. That distinction is the organizing question here, not a replacement for existing perturbation bounds.

The reversible structure is governed by Kolmogorov's cycle criterion: on a finite connected graph, a positive edge flow is reversible precisely when its directed ratios multiply to one around every cycle. Standard Markov-chain treatments give this criterion in the broader language of detailed balance (Norris, 1997). The weighted least-squares projection used below is the graph Hodge decomposition that separates gradient flows from cyclic components, familiar from ranking and network-flow analyses (Jiang et al., 2011). We apply that geometry to pairwise errors in MH log ratios; the projection itself is not claimed as a new graph decomposition.

Finally, acceptance probabilities are nonlinear functions of the log-ratio, so a cycle residual must still be translated into a transition-kernel perturbation before invariant-law stability can be invoked. The standard MH construction and its posterior-sampling formulation are due to the classical algorithmic literature (Hastings, 1970; Tierney, 1994), and the accept–reject mechanism has a broader modern treatment in Glatt-Holtz et al. (2023). The results below combine these ingredients in a restricted setting: an exact finite-graph reversibility test, a weighted potential–residual split, and a contractive perturbation bound. In particular, they do not provide uniform ergodicity or contraction for arbitrary random-walk or Langevin MH chains.

3 Setup

Let $(\mathsf{X}, \mathcal{X})$ be a standard Borel space, π a probability measure, and Q a proposal Markov kernel. Let P be an MH kernel with proposal Q and acceptance probability

$$a(x, y) = h(\ell(x, y)), \qquad h(t) = 1 \wedge e^t,$$

where ℓ is an exact log acceptance ratio for π and Q. Self-proposals may be included in Q; they do not affect the arguments.

For a measurable log-ratio error ε, define the perturbed kernel P^{ε} by using

$$a^{\varepsilon}(x,y) = h\big(\ell(x,y) + \varepsilon(x,y)\big)$$

with the same proposal Q. Assume throughout that P^{ε} has an invariant probability measure μ^{ε}.

For a bounded measurable potential ψ, define

$$Z_{\psi} = \int e^{\psi(x)}\,\pi(dx), \qquad \pi_{\psi}(dx) = Z_{\psi}^{-1} e^{\psi(x)}\,\pi(dx).$$

The MH log-ratio for π_{ψ} with proposal Q is $\ell(x,y) + \psi(y) - \psi(x)$. Write P^{ψ} for that MH kernel. For a Markov kernel K, its Dobrushin coefficient is

$$\delta(K) = \sup_{x,x' \in \mathsf{X}} \mathrm{TV}\big(K(x,\cdot), K(x',\cdot)\big).$$

We use $\mathrm{TV}(\nu,\rho) = \sup_{A \in \mathcal{X}} |\nu(A) - \rho(A)|$.

4 When is a log-ratio error a potential?

For a finite state space, form the undirected proposal graph by joining x and y whenever $Q(x,y) > 0$ and $Q(y,x) > 0$. Assume this graph is connected and the target gives positive mass to every state. A pairwise error is called antisymmetric if $\varepsilon(x,y) = -\varepsilon(y,x)$ on every proposal edge.

Proposition 1 (Cycle criterion for a potential error) *Under these assumptions, the perturbed MH kernel with log-ratio error ε is reversible with respect to a probability measure if and only if*

$$\sum_{i=0}^{k-1} \varepsilon(x_i, x_{i+1}) = 0$$

for every graph cycle $x_0, x_1, \dots, x_k = x_0$. When this condition holds, there is a state function ψ, unique up to an additive constant, such that $\varepsilon(x,y) = \psi(y) - \psi(x)$ on every edge, and the reversible law is $\pi_{\psi} \propto e^{\psi}\pi$.

Proof For an oriented edge, the exact MH ratio satisfies

$$\frac{\pi(x)Q(x,y)}{\pi(y)Q(y,x)} = e^{-\ell(x,y)}.$$

The identity $h(s)/h(-s) = e^{s}$ therefore gives the ratio of perturbed off-diagonal flows:

$$\frac{\pi(x)Q(x,y)h(\ell(x,y) + \varepsilon(x,y))}{\pi(y)Q(y,x)h(\ell(y,x) + \varepsilon(y,x))} = e^{\varepsilon(x,y)}.$$

If a reversible law ν exists, irreducibility implies $\nu(x) > 0$. Writing $\psi(x) = \log(\nu(x)/\pi(x))$, detailed balance forces $\varepsilon(x,y) = \psi(y) - \psi(x)$. Summing around a cycle proves necessity. Conversely, if every cycle sum is zero, define $\psi(y)$ by summing ε along any path from a fixed root to y. The cycle condition makes this definition path-independent. The displayed flow ratio then verifies detailed balance for $e^{\psi}\pi$. ∎

A circulating three-state error. Consider a three-state cycle with uniform target and symmetric nearest-neighbor proposals. Set $\varepsilon(0,1) = \varepsilon(1,2) = \varepsilon(2,0) = c$ and use $-c$ on the reverse edges, where $c > 0$. The cycle sum is $3c$, so the error is not a potential. In state order $0, 1, 2$, the transition matrix is

$$\widetilde{P} = \frac{1}{2}\begin{pmatrix} 1-e^{-c} & 1 & e^{-c} \\ e^{-c} & 1-e^{-c} & 1 \\ 1 & e^{-c} & 1-e^{-c} \end{pmatrix}.$$

It is doubly stochastic, so the invariant law remains uniform. But the clockwise stationary current on each edge is $\frac{1}{6}(1-e^{-c})$, which is nonzero for $c \neq 0$. The example separates stationary bias from irreversibility: a non-potential component may change the dynamics without changing the invariant marginal.

5 A least-squares potential on a finite graph

The cycle criterion also gives a canonical finite-state decomposition when the error is not itself a potential. Let $G = (V, E)$ be the connected reciprocal proposal graph and choose one orientation for each undirected edge. Write ε_e for the oriented antisymmetric error and let B be the vertex–edge incidence matrix, with $(B\psi)_e = \psi(\mathrm{head}(e)) - \psi(\mathrm{tail}(e))$. For positive edge weights w_e, define

$$J(\psi) = \frac{1}{2}\sum_{e\in E} w_e\big(\varepsilon_e - (B\psi)_e\big)^2.$$

The weights can reflect proposal frequency or another chosen measure of edge importance. They are part of the decomposition and must be reported; there is no weight-free projection in a general graph.

Proposition 2 (Weighted graph decomposition) *Fix the gauge $\sum_{x\in V}\psi(x) = 0$. The minimizer $\psi^\star$ of J is unique and satisfies*

$$B^\top W B\psi^\star = B^\top W\varepsilon, \qquad B^\top W r^\star = 0, \qquad r^\star = \varepsilon - B\psi^\star,$$

where $W = \mathrm{diag}(w_e)$. In particular, the residual is orthogonal to every potential gradient in the weighted edge inner product, and

$$\|\varepsilon\|_W^2 = \|B\psi^\star\|_W^2 + \|r^\star\|_W^2.$$

Proof Differentiating J gives the normal equations $B^\top W(B\psi^\star - \varepsilon) = 0$. On a connected graph, $B^\top W B$ is the weighted graph Laplacian with constants as its null space; the zero-sum gauge therefore makes the solution unique. The normal equations give the residual orthogonality. Expanding $\|\varepsilon\|_W^2 = \|B\psi^\star + r^\star\|_W^2$ then yields the Pythagorean identity. ■

This projection minimizes a quadratic edge error, not the acceptance-sensitive quantity in Theorem 3. It is therefore a principled finite-graph way to choose a candidate potential, but does not by itself optimize the invariant-law bound. For a general pairwise error, the symmetric component $[\varepsilon(x,y) + \varepsilon(y,x)]/2$ is already incompatible with any potential difference; the graph projection applies to the antisymmetric component.

6 A potential–residual perturbation bound

Theorem 3 (Potential–residual bound) *Suppose ψ is bounded and*

$$\varepsilon(x,y) = \psi(y) - \psi(x) + r(x,y).$$

Define the acceptance-sensitive residual

$$\eta_{\mathrm{acc}} = \sup_{x\in\mathsf{X}} \int |h(\ell(x,y)+\psi(y)-\psi(x)+r(x,y)) - h(\ell(x,y)+\psi(y)-\psi(x))|\, Q(x,dy).$$

Suppose also that $\delta(P^\psi) \le \tau < 1$. Then

$$\mathrm{TV}(\mu^\varepsilon, \pi) \le \tanh\!\left(\frac{\mathrm{osc}(\psi)}{4}\right) + \frac{\eta_{\mathrm{acc}}}{1-\tau}.$$

Proof By definition, the proposal-averaged absolute difference between the two acceptance probabilities is bounded by η_{acc}. The two transition kernels have the same off-diagonal proposal measure apart from the difference in acceptance probabilities; the diagonal mass is determined by the off-diagonal mass. It follows that

$$\sup_x \mathrm{TV}\big(P^\varepsilon(x,\cdot), P^\psi(x,\cdot)\big) \le \eta_{\mathrm{acc}}.$$

Using invariance of μ^ε for P^ε and of π_ψ for P^ψ,

$$\begin{aligned}\mathrm{TV}(\mu^\varepsilon, \pi_\psi) &\le \mathrm{TV}\big(\mu^\varepsilon P^\varepsilon, \mu^\varepsilon P^\psi\big) + \mathrm{TV}\big(\mu^\varepsilon P^\psi, \pi_\psi P^\psi\big)\\ &\le \eta_{\mathrm{acc}} + \delta(P^\psi)\,\mathrm{TV}(\mu^\varepsilon, \pi_\psi).\end{aligned}$$

Rearranging gives

$$\mathrm{TV}(\mu^\varepsilon, \pi_\psi) \le \frac{\eta_{\mathrm{acc}}}{1-\tau}.$$

Finally, let $g = d\pi_\psi/d\pi = e^\psi/Z_\psi$ and set $a = \operatorname{ess\,inf}_\pi g$ and $b = \operatorname{ess\,sup}_\pi g$. Then $\int g\, d\pi = 1$ and $b/a \le e^{\mathrm{osc}(\psi)}$. By convexity, $|g-1|$ lies below the chord joining its values at a and b, which gives

$$\mathrm{TV}(\pi_\psi, \pi) = \frac{1}{2}\int |g-1|\, d\pi \le \frac{(1-a)(b-1)}{b-a}.$$

If $a = b = 1$, the desired bound is immediate. Writing $r = b/a$, the right-hand side is maximized over $a \in [1/r, 1]$ at $a = r^{-1/2}$, where it equals $(\sqrt{r}-1)/(\sqrt{r}+1) = \tanh(\log(r)/4)$. Since $r \le e^{\mathrm{osc}(\psi)}$, this proves

$$\mathrm{TV}(\pi_\psi, \pi) \le \tanh\!\left(\frac{\mathrm{osc}(\psi)}{4}\right).$$

The triangle inequality completes the proof. ∎

Proposition 4 (Sharpness of the residual contraction factor) *For every $\tau \in [0,1)$ and $0 < \delta < (1-\tau)/2$, there is a two-state MH example with $\delta(P) = \tau$, $\psi = 0$, and $\eta_{\mathrm{acc}} = \delta$ for which*

$$\mathrm{TV}(\mu^\varepsilon, \pi) = \frac{\eta_{\mathrm{acc}}}{1-\tau}.$$

Proof Let π be uniform on $\{0,1\}$, set $a = (1-\tau)/2$, and use proposal probabilities $Q(0,1) = 2a$ and $Q(1,0) = a$, with the remaining mass assigned to self-proposals. The exact MH acceptance probabilities are $1/2$ from 0 to 1 and 1 from 1 to 0, so both accepted transition probabilities equal a and $\delta(P) = |1-2a| = \tau$.

Change these acceptance probabilities to

$$\widetilde{a}(0,1) = \frac{1}{2} + \frac{\delta}{2a}, \qquad \widetilde{a}(1,0) = 1 - \frac{\delta}{a}.$$

They lie in $[0,1]$ because $0 < \delta < a$. The perturbed transition probabilities are $a+\delta$ and $a-\delta$, respectively, so

$$\mu^{\varepsilon}(\{1\}) = \frac{a+\delta}{2a} = \frac{1}{2} + \frac{\delta}{1-\tau}.$$

The rowwise proposal-averaged acceptance discrepancies are both δ, hence $\eta_{\mathrm{acc}} = \delta$ and the claimed equality follows. The induced log-ratio errors are not antisymmetric:

$$\varepsilon(0,1) = \log(1+\delta/a), \qquad \varepsilon(1,0) = \log(1-\delta/a) - \log 2,$$

whose sum is $\log(1-(\delta/a)^2) - \log 2 \neq 0$. Thus the example uses a genuinely pair-dependent residual. ∎

Remark 5 (Simple residual bounds) *The acceptance-sensitive quantity satisfies*

$$\eta_{\mathrm{acc}} \leq \sup_x \int \min\{1, |r(x,y)|\}\, Q(x,dy) \leq \sup_x \int |r(x,y)|\, Q(x,dy),$$

because h is 1-Lipschitz and takes values in $[0,1]$. The first bound can be strictly smaller when the residual is large on a small proposal set. Computing η_{acc} still requires control of the exact log ratio and the proposed potential; the inequality does not make the bound computable unless those quantities can themselves be controlled.

Corollary 6 (Pure potential error) *If $r = 0$, then $P^{\varepsilon} = P^{\psi}$ and its invariant measure is π_ψ. In particular,*

$$\mathrm{TV}(\pi_\psi, \pi) \leq \tanh\left(\frac{\mathrm{osc}(\psi)}{4}\right).$$

Remark 7 (Sharpness of the target-tilt term) *The bound is sharp. For $c > 0$, consider a two-point space with*

$$\pi(x_-) = \frac{e^{c/2}}{1+e^{c/2}}, \qquad \pi(x_+) = \frac{1}{1+e^{c/2}}, \qquad \psi(x_-) = -\frac{c}{2}, \quad \psi(x_+) = \frac{c}{2}.$$

Then $\mathrm{osc}(\psi) = c$ and equality holds.

Proposition 8 (First- and second-order response) *Let f and ψ be bounded and define $\pi_t(dx) \propto e^{t\psi(x)}\pi(dx)$ for $0 \leq t \leq 1$. Then*

$$\frac{d}{dt}\int f\, d\pi_t = \mathrm{Cov}_{\pi_t}(f,\psi), \qquad \frac{d^2}{dt^2}\int f\, d\pi_t = \mathbb{E}_{\pi_t}\left[(f - \mathbb{E}_{\pi_t} f)(\psi - \mathbb{E}_{\pi_t}\psi)^2\right].$$

In particular,

$$\left| \int f \, d\pi_\psi - \int f \, d\pi - \mathrm{Cov}_\pi(f, \psi) \right| \le \frac{\|f\|_\infty}{4} \operatorname{osc}(\psi)^2.$$

Proof Write $\mathbb{E}_{\pi_t} f = \mathbb{E}_\pi[f e^{t\psi}]/\mathbb{E}_\pi[e^{t\psi}]$ and differentiate the quotient. This gives the covariance formula; a second differentiation gives the stated centered third-moment expression. Since $|f - \mathbb{E}_{\pi_t} f| \le 2\|f\|_\infty$ and $\mathrm{Var}_{\pi_t}(\psi) \le \operatorname{osc}(\psi)^2/4$, we have $|d^2 \mathbb{E}_{\pi_t} f / dt^2| \le \|f\|_\infty \operatorname{osc}(\psi)^2/2$. Taylor's theorem at $t = 0$ yields the remainder bound. ■

7 Limitations

The invariant-law inequality assumes a one-step Dobrushin contraction of P^ψ. On an unbounded state space this is generally much stronger than geometric ergodicity and is not implied by strong log-concavity alone. Replacing it by drift–minorization or a weighted Wasserstein condition would require a correspondingly weighted residual norm and additional control of the approximate chain's tails. We have not made that extension here.

On finite graphs, the least-squares projection depends on its edge weights and minimizes a quadratic error, whereas the invariant-law bound depends on the acceptance change. The projection is therefore a decomposition tool, not an optimizer of the bound. Determining useful weights or an acceptance-weighted projection is a separate problem. The results in this paper are distributional statements and make no claim about the error of a finite-run empirical average.

Acknowledgments and Disclosure of Funding

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors declare no competing interests.

References

Pierre Alquier, Nial Friel, Richard Everitt, and Aidan Boland. Noisy Monte Carlo: convergence of markov chains with approximate transition kernels. *Statistics and Computing*, 26(1–2):29–47, 2016. doi: 10.1007/s11222-014-9521-x.

Nathan Glatt-Holtz, Justin Krometis, and Cecilia Mondaini. On the accept-reject mechanism for metropolis–hastings algorithms. *The Annals of Applied Probability*, 33(6B): 5279–5333, 2023. doi: 10.1214/23-AAP1948.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. doi: 10.1093/biomet/57.1.97.

Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, and Yinyu Ye. Statistical ranking and combinatorial Hodge theory. *Mathematical Programming*, 127(1):203–244, 2011. doi: 10.1007/s10107-010-0419-x.

Felipe Medina-Aguayo, Daniel Rudolf, and Nikolaus Schweizer. Perturbation bounds for monte carlo within metropolis via restricted approximations. *Stochastic Processes and their Applications*, 130(4):2200–2227, 2020. doi: 10.1016/j.spa.2019.06.015.

Felipe J. Medina-Aguayo, Anthony Lee, and Gareth O. Roberts. Stability of noisy metropolis–hastings, 2015.

A. Yu. Mitrophanov. Sensitivity and convergence of uniformly ergodic markov chains. *Journal of Applied Probability*, 42(4):1003–1014, 2005.

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

Daniel Rudolf and Nikolaus Schweizer. Perturbation theory for markov chains via wasserstein distance. *Bernoulli*, 24(4A):2610–2639, 2018. doi: 10.3150/17-BEJ938.

Luke Tierney. Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4):1701–1728, 1994. doi: 10.1214/aos/1176325750.