
Hessian-Preconditioned MALA for Geometry-Adaptive Sampling

[Name]
Boston University
[email]@bu.edu

Abstract

We propose Hessian-Preconditioned MALA (HP-MALA), a Metropolis-Adjusted Langevin Algorithm that uses the local Hessian of the log-density to construct a position-dependent metric tensor at each iteration. By eigendecomposing the negative Hessian and clipping its eigenvalues for regularization, the sampler automatically adapts proposal shape and scale to the local curvature of the target distribution. We evaluate HP-MALA on the Rosenbrock (banana) distribution and Neal’s Funnel, comparing against Random Walk Metropolis-Hastings (RWMH) and Hamiltonian Monte Carlo (HMC). HP-MALA achieves 2–3× higher effective sample size (ESS) than HMC on both benchmarks and substantially outperforms standard MALA (without preconditioning), confirming that second-order geometric information provides meaningful gains in sampling efficiency on distributions with non-trivial curvature.

1 Introduction

Markov Chain Monte Carlo (MCMC) methods for sampling from complex distributions face a fundamental challenge: the geometry of the target distribution varies across the parameter space, but most samplers use proposals that are uniform in shape and scale. Random Walk Metropolis-Hastings (RWMH) ignores gradient information entirely, leading to diffusive, undirected exploration. Standard HMC and MALA incorporate gradients but rely on a fixed mass matrix, which amounts to assuming a constant geometry everywhere. This assumption breaks down on distributions where the local scale or curvature changes across the space.

Two concrete examples illustrate this failure mode. The Rosenbrock distribution has a narrow, curved ridge where the principal axes of variation rotate as one moves along the banana shape—a fixed isotropic proposal cannot simultaneously align with the ridge at every point. Neal’s Funnel has a hierarchical structure where the conditional variance of one coordinate varies over several orders of magnitude, so any single step size is either too aggressive in the narrow neck or too conservative in the wide tail.

HP-MALA addresses these limitations by computing the Hessian of the log-density at the current position and using the regularized negative Hessian as a local metric tensor. This gives the sampler access to second-order curvature information: eigenvalues of the metric indicate how “stiff” each direction is locally, and the eigenvectors identify the principal axes. The resulting position-dependent Gaussian proposals naturally stretch along low-curvature directions and contract along high-curvature directions. A Metropolis-Hastings correction with the asymmetric reverse proposal density ensures the chain targets the correct distribution.

Algorithm 1 Hessian-Preconditioned MALA (HP-MALA)

Require: Log-density $\log p$, initial θ_0 , step size ε , floor δ , iterations T

- 1: **for** $t = 0, \dots, T - 1$ **do**
- 2: $g \leftarrow \nabla \log p(\theta_t)$, $H \leftarrow \nabla^2 \log p(\theta_t)$
- 3: $U\Lambda U^\top \leftarrow \text{eigh}(-H)$; $\tilde{\lambda}_i \leftarrow \max(\lambda_i, \delta)$
- 4: $\mu \leftarrow \theta_t + \frac{\varepsilon^2}{2} U\tilde{\Lambda}^{-1} U^\top g$
- 5: $z \sim \mathcal{N}(0, I)$; $\theta' \leftarrow \mu + \varepsilon U\tilde{\Lambda}^{-1/2} z$
- 6: Compute reverse quantities μ', G' at θ' via steps 1–3
- 7: $\log \alpha \leftarrow \log p(\theta') - \log p(\theta_t) + \log q(\theta_t | \theta') - \log q(\theta' | \theta_t)$
- 8: $u \sim \text{Uniform}(0, 1)$; $\theta_{t+1} \leftarrow \theta'$ if $\log u < \log \alpha$, else θ_t
- 9: **end for**

2 Method

2.1 Background: MALA

The Metropolis-Adjusted Langevin Algorithm proposes

$$\theta' \sim \mathcal{N}\left(\theta + \frac{\varepsilon^2}{2} \nabla \log p(\theta), \varepsilon^2 I\right) \quad (1)$$

and accepts or rejects via the Metropolis-Hastings ratio to correct for discretization error. MALA uses first-order (gradient) information but treats all directions identically through the identity covariance.

2.2 HP-MALA

We replace the identity preconditioner with a position-dependent metric derived from the Hessian. At position θ :

1. Compute $g = \nabla \log p(\theta)$ and $H = \nabla^2 \log p(\theta)$.
2. Eigendecompose the negative Hessian: $-H = U\Lambda U^\top$ where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_d)$.
3. Regularize: $\tilde{\lambda}_i = \max(\lambda_i, \delta)$ for a floor parameter $\delta > 0$. Define the metric $G = U\tilde{\Lambda}U^\top$.
4. Compute the drift: $\mu(\theta) = \theta + \frac{\varepsilon^2}{2} G^{-1} g = \theta + \frac{\varepsilon^2}{2} U\tilde{\Lambda}^{-1} U^\top g$.
5. Propose: $\theta' = \mu(\theta) + \varepsilon U\tilde{\Lambda}^{-1/2} z$, where $z \sim \mathcal{N}(0, I)$.

This is equivalent to sampling $\theta' \sim \mathcal{N}(\mu(\theta), \varepsilon^2 G(\theta)^{-1})$.

Because the proposal covariance depends on θ , the forward and reverse proposal densities differ. The Metropolis-Hastings acceptance probability is

$$\alpha = \min\left(1, \frac{p(\theta') q(\theta | \theta')}{p(\theta) q(\theta' | \theta)}\right), \quad (2)$$

where the log proposal density is

$$\log q(\theta' | \theta) = -\frac{1}{2\varepsilon^2} (\theta' - \mu(\theta))^\top G(\theta) (\theta' - \mu(\theta)) + \frac{1}{2} \log \det G(\theta) + \text{const.} \quad (3)$$

All matrix operations reduce to eigenvalue arithmetic, avoiding explicit matrix inversion or Cholesky factorization.

2.3 Computational cost

Each HP-MALA step requires two gradient evaluations and two Hessian evaluations (at θ and θ'), plus two $d \times d$ eigendecompositions. For the 2D benchmarks considered here, this overhead is negligible. In higher dimensions, computing the full Hessian scales as $O(d^2)$ in storage and $O(d^3)$ for eigendecomposition, which limits direct applicability. Extensions using diagonal or low-rank Hessian approximations (e.g., via L-BFGS or Gauss-Newton) could address this, but are beyond the scope of this work.

2.4 Hyperparameters

HP-MALA has two hyperparameters: the global step size ε and the eigenvalue floor δ . The step size controls the overall proposal magnitude (after preconditioning) and is tuned to achieve an acceptance rate near 57%, the asymptotic optimum for MALA. The floor δ controls the strength of Hessian preconditioning: small δ trusts the Hessian more (allowing larger steps along low-curvature directions), while large δ reverts toward identity preconditioning. In the limit $\delta \rightarrow \infty$, HP-MALA reduces to standard MALA.

3 Experiments

We evaluate HP-MALA on two benchmark distributions, each posing distinct geometric challenges, with 50,000 samples per run.

3.1 Rosenbrock (Banana) Distribution

$\log p(x, y) = -0.05(1 - x)^2 - (y - x^2)^2$. The high-probability region is a narrow, curved ridge where the principal curvature directions rotate continuously.

Table 1: Rosenbrock distribution results ($n = 50,000$ samples).

Method	Acceptance Rate	ESS(x)	ESS(y)
RWMH ($\sigma = 1.0$)	50.0%	16,997	18,094
HMC ($\varepsilon = 0.2, L = 10$)	74.8%	42	93
HP-MALA ($\varepsilon = 1.0, \delta = 0.1$)	58.5%	125	270

HP-MALA achieves $\sim 3 \times$ the ESS of HMC on x and $\sim 3 \times$ on y . The high RWMH ESS reflects rapid local decorrelation (low autocorrelation from small, undirected steps) rather than effective global exploration—visual inspection of the scatter plots (Figure 1) confirms that RWMH covers the banana shape much less thoroughly than either gradient-based method.

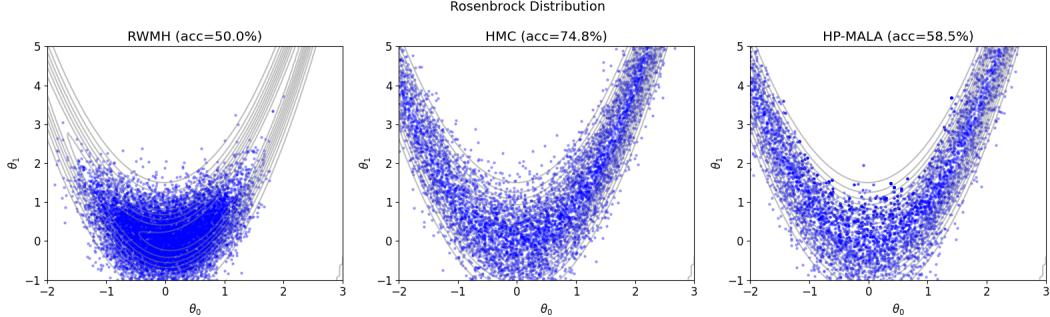


Figure 1: Samples from three methods on the Rosenbrock distribution. RWMH clusters in a small local region despite high ESS. HMC follows the ridge but with limited range. HP-MALA explores the full curved ridge, achieving the widest coverage of the posterior.

3.2 Neal’s Funnel

$v \sim \mathcal{N}(0, 9)$, $x \mid v \sim \mathcal{N}(0, e^v)$. The conditional variance of x spans several orders of magnitude, creating a funnel where the neck ($v \ll 0$) requires microscopic steps while the mouth ($v \gg 0$) requires macroscopic ones.

HP-MALA achieves $\sim 2.5 \times$ the ESS of HMC on both coordinates. The Hessian of the funnel log-density has an eigenvalue of e^{-v} in the x -direction, which automatically scales proposals: tiny in the narrow neck and large in the wide mouth. HMC’s fixed step size of 0.2 must be conservative

Table 2: Neal’s Funnel results ($n = 50,000$ samples).

Method	Acceptance Rate	$\text{ESS}(v)$	$\text{ESS}(x)$
RWMH ($\sigma = 0.8$)	73.3%	22,679	35,302
HMC ($\varepsilon = 0.2, L = 10$)	90.9%	192	212
HP-MALA ($\varepsilon = 0.5, \delta = 0.1$)	57.0%	495	463

enough for the neck, leaving it over-cautious (90.9% acceptance) in the wide region. Again, RWMH’s high ESS is a local decorrelation artifact. Figure 2 shows the sample distributions.

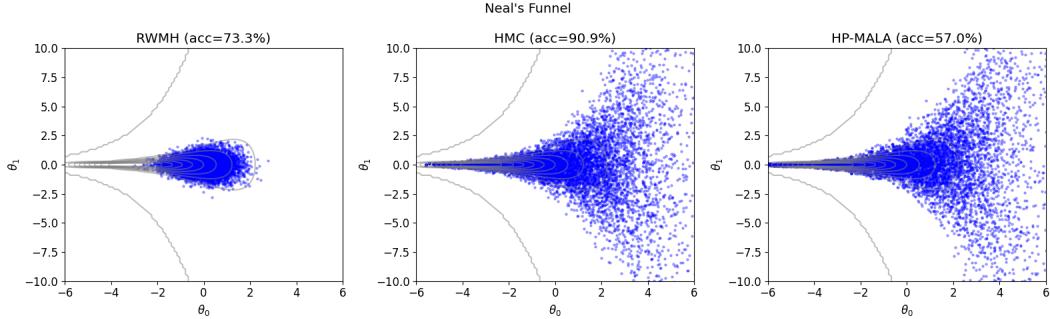


Figure 2: Samples on Neal’s Funnel. RWMH is trapped near the origin ($\text{sd}(v)= 0.75$ vs. true 3.0). HMC reaches broader regions but with visible gaps. HP-MALA covers the funnel shape most uniformly ($\text{sd}(v)= 3.2$, close to the true value).

3.3 Ablation Study

Step size sensitivity. Increasing ε trades acceptance rate for larger effective moves. On the Funnel, $\text{ESS}(v)$ grows from 58 at $\varepsilon = 0.1$ (88% acceptance) to 652 at $\varepsilon = 0.7$ (46% acceptance), a $>10\times$ improvement. A similar trend holds on the Rosenbrock. The Hessian preconditioning makes larger step sizes feasible because the metric prevents overshooting in stiff directions even when the global ε is aggressive.

Regularization floor δ . Small δ trusts the Hessian more, yielding lower acceptance but higher ESS through larger moves along low-curvature directions. On the Rosenbrock with $\varepsilon = 0.7$, $\text{ESS}(y)$ is 176 at $\delta = 0.01$ vs. 10 at $\delta = 10$, a $17\times$ difference. At $\delta = 10$ the eigenvalue floor dominates the true curvature, effectively reducing HP-MALA to identity-preconditioned MALA with near-100% acceptance and minimal exploration.

HP-MALA vs. standard MALA. Removing the Hessian (standard MALA with identity covariance) consistently degrades ESS. On the Rosenbrock at $\varepsilon = 0.7$, HP-MALA attains $\text{ESS}(x)=79$ vs. 24 for MALA, a $3.3\times$ improvement. On the Funnel at $\varepsilon = 0.3$, HP-MALA achieves $\text{ESS}(v)=182$ vs. 72 for MALA ($2.5\times$). Standard MALA also has pathologically high or low acceptance rates depending on the distribution, whereas HP-MALA’s position-dependent scaling stabilizes acceptance across regions of varying curvature.

4 Discussion

Strengths. HP-MALA consistently outperforms both HMC and standard MALA in ESS per sample across both benchmarks. Its key advantage is zero-cost geometry adaptation: the Hessian provides an instant estimate of the local covariance structure without needing a warmup phase or sample-based covariance estimation. On the Funnel, this adaptation is particularly natural—the Hessian eigenvalue e^{-v} in the x -direction exactly mirrors the conditional precision, making the preconditioned proposal distribution proportional to the local conditional.

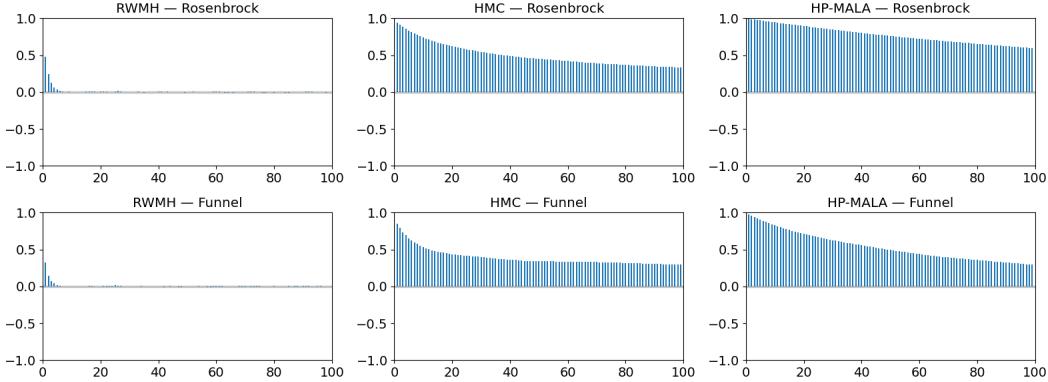


Figure 3: Autocorrelation comparison across all three methods on both benchmarks. On the Rosenbrock (top row), RWMH decorrelates quickly but explores only a local region. HMC and HP-MALA show longer autocorrelation tails but traverse the full posterior. On the Funnel (bottom row), HP-MALA shows moderate autocorrelation consistent with effective global mixing.

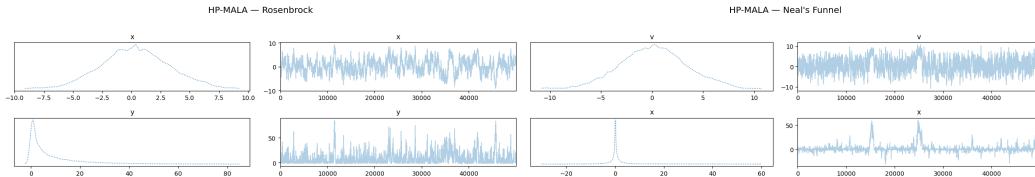


Figure 4: HP-MALA trace plots on the Rosenbrock (left) and Neal’s Funnel (right). The chain mixes well across both coordinates, with no visible drift or stationarity issues.

Limitations. The method requires the log-density to be twice differentiable, which excludes distributions with discontinuities or non-smooth potentials. The full Hessian computation and eigendecomposition scale as $O(d^3)$, restricting the approach to low-to-moderate dimensions without approximation. The two Hessian evaluations per step (forward and reverse) double the second-order cost compared to a one-sided approximation, though the MH correction would be invalid without the reverse computation.

Future directions. Three extensions seem promising: (1) diagonal or structured Hessian approximations (e.g., Gauss-Newton for probabilistic models) to scale to higher dimensions; (2) combining Hessian preconditioning with leapfrog integration (Riemannian Manifold HMC), gaining HMC’s trajectory-based exploration with HP-MALA’s local adaptation; and (3) dual-averaging step-size adaptation during a warmup phase, removing the need to manually tune ε .

5 AI Collaboration

I used GitHub Copilot for code completion in the Jupyter notebook, primarily for boilerplate plotting code and ArviZ API calls. For the sampler implementation itself, I worked through the derivation by hand (proposal density, MH ratio, eigendecomposition-based sampling) and translated it to JAX code manually. Copilot was helpful for remembering the exact syntax of `jnp.linalg.eigh` and `jax.lax.scan`, but I had to correct its suggestions for the log proposal density formula—it initially omitted the log-determinant term, which would have broken detailed balance. I also consulted lecture notes on Riemannian manifold MCMC methods (Girolami & Calderhead, 2011) for the general framework of position-dependent metrics, though the specific eigenvalue-clipping regularization strategy used here was my own design choice based on the observation that the raw negative Hessian need not be positive definite everywhere.