
Assignment 1: [Puchen Liu]

Puchen Liu
Boston University
pcliu@bu.edu

Abstract

We introduce a Tangential–Normal Mixture (TNM) sampler, a structured Metropolis–Hastings method designed to improve sampling efficiency in distributions with strong curvature and anisotropy. Unlike Random Walk Metropolis–Hastings (RWMH), which struggles with slow mixing in narrow, curved regions, and Hamiltonian Monte Carlo (HMC), which relies on carefully tuned integration dynamics, TNM decomposes proposals into tangential moves along high-density manifolds and normal moves across curvature directions. These structured updates are combined with occasional isotropic random-walk proposals to preserve ergodicity and robustness. We evaluate TNM on two challenging benchmarks, the Rosenbrock distribution and Neal’s Funnel, comparing acceptance rates, effective sample size (ESS), trace behavior, and autocorrelation against RWMH and HMC baselines. Results show that TNM improves geometric exploration relative to RWMH and provides competitive robustness in regions where HMC exhibits instability, highlighting the benefits of incorporating local geometric structure into proposal design.

1 Introduction

Markov Chain Monte Carlo (MCMC) methods are central tools for sampling from complex probability distributions in Bayesian inference and scientific computing. Among classical approaches, Random Walk Metropolis–Hastings (RWMH) is simple and broadly applicable but often exhibits slow mixing when the target distribution is highly correlated or curved. In such settings, isotropic proposals frequently step away from the narrow high-density region, leading to strong autocorrelation and inefficient exploration. Hamiltonian Monte Carlo (HMC) addresses this limitation by leveraging gradient information to generate long-distance proposals that follow energy-preserving trajectories. However, HMC introduces additional hyperparameters (e.g., step size and number of leapfrog steps) and can behave conservatively or unstably when the geometry of the distribution varies significantly across regions. This project proposes a Tangential–Normal Mixture (TNM) sampler that aims to bridge the gap between these two paradigms. The key idea is to incorporate local geometric structure into a Metropolis–Hastings framework without fully simulating Hamiltonian dynamics. At each step, TNM uses the gradient of the log-density to define a local directional basis, separating motion into a tangential component that explores along high-density manifolds and a normal component that allows controlled movement across curvature. These structured proposals are combined with occasional isotropic random-walk updates to ensure robustness and ergodicity. By aligning proposals with the local geometry while retaining the simplicity of MH correction, TNM seeks to improve exploration efficiency on challenging targets such as the Rosenbrock distribution and Neal’s Funnel.

2 Method

We propose a Tangential–Normal Mixture (TNM) sampler, a structured Metropolis–Hastings (MH) algorithm that adapts its proposal geometry using local gradient information. Let $\pi(x)$ denote the

target density and $g(x) = \nabla \log \pi(x)$ its score function. At each state $x \in \mathbb{R}^d$, TNM constructs a local unit direction

$$\hat{g}(x) = \frac{g(x)}{\|g(x)\| + \varepsilon},$$

which defines a one-dimensional “normal” direction aligned with the score and a $(d - 1)$ -dimensional orthogonal subspace interpreted as the “tangential” space. The proposal distribution is a rank-1 anisotropic Gaussian:

$$x' = x + \mu(x) + \eta,$$

where the drift term is

$$\mu(x) = \delta \hat{g}(x),$$

and $\eta \sim \mathcal{N}(0, \Sigma(x))$ with covariance

$$\Sigma(x) = \sigma_{\perp}^2 I + (\sigma_{\parallel}^2 - \sigma_{\perp}^2) \hat{g}(x) \hat{g}(x)^{\top}.$$

Here, σ_{\perp} controls exploration in the tangential subspace, while σ_{\parallel} controls movement along the normal (score-aligned) direction. Intuitively, large σ_{\perp} encourages motion along curved high-density manifolds, whereas smaller σ_{\parallel} stabilizes motion across sharp curvature (e.g., the funnel neck). To improve robustness, TNM employs a mixture proposal. With probability p_{mix} , we instead generate an isotropic random-walk proposal

$$x' = x + \sigma_{\text{rw}} z, \quad z \sim \mathcal{N}(0, I),$$

which helps prevent geometric misalignment from trapping the chain and ensures global ergodicity.

Because the proposal is state-dependent and generally asymmetric, we apply the standard Metropolis–Hastings correction:

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

where $q(\cdot | \cdot)$ denotes the mixture proposal density. This guarantees detailed balance with respect to π and preserves the correct stationary distribution.

Algorithmically, each TNM iteration proceeds as follows:

1. Compute $g(x)$ and $\hat{g}(x)$.
2. Sample either a structured tangential–normal proposal or an isotropic random-walk proposal.
3. Compute the MH acceptance probability $\alpha(x, x')$.
4. Accept x' with probability α ; otherwise remain at x .

TNM can be interpreted as a geometry-aware random-walk method that leverages first-order information without simulating Hamiltonian dynamics. It retains the simplicity and generality of MH while partially adapting to local curvature through anisotropic proposals.

3 Experiments

3.1 Benchmarks

We evaluate the proposed TNM sampler on two standard stress-test distributions:

Rosenbrock (Banana). This two-dimensional distribution exhibits a narrow, curved ridge defined by $y \approx x^2$. The high-density region is thin and strongly correlated, making isotropic proposals inefficient. This benchmark tests whether a sampler can follow curved geometry without repeatedly stepping off the ridge.

Neal’s Funnel. The funnel distribution introduces hierarchical scale variation:

$$v \sim \mathcal{N}(0, 9), \quad x | v \sim \mathcal{N}(0, e^v).$$

When v is small, the conditional variance of x becomes extremely narrow, producing a “neck” that is notoriously difficult for MCMC methods. This benchmark tests robustness under strong anisotropy and varying local curvature.

3.2 Baselines and Setup

We compare TNM against:

- Random Walk Metropolis–Hastings (RWMH) with isotropic Gaussian proposals.
- Hamiltonian Monte Carlo (HMC) with identity mass matrix.

All samplers were run for 50,000 iterations from the same initialization. For TNM, we tuned $(\sigma_{\perp}, \sigma_{\parallel}, \delta, p_{\text{mix}})$ separately for each benchmark. We report acceptance rates and effective sample sizes (ESS) computed using ArviZ.

3.3 Rosenbrock Results

On the Rosenbrock distribution, RWMH achieved an acceptance rate of approximately 50%, but exhibited strong autocorrelation and slow traversal along the curved ridge. HMC achieved a higher acceptance rate (approximately 78%) and produced longer-range moves aligned with the geometry.

TNM achieved an acceptance rate of approximately 69%. Visual inspection of samples shows that TNM follows the curved manifold more effectively than RWMH, producing better ridge coverage with fewer rejected off-manifold proposals. However, ESS values indicate that while TNM improves geometric alignment relative to RWMH, it does not consistently surpass HMC in effective sample size.

Trace plots show reduced stickiness compared to RWMH but still noticeable local correlation due to the random-walk nature of the proposal.

3.4 Funnel Results

On Neal’s Funnel, RWMH achieved approximately 65% acceptance but struggled to move through the narrow neck region. HMC achieved very high acceptance (approximately 91%), though trajectory behavior depended sensitively on tuning.

TNM achieved an acceptance rate of approximately 50%. The structured anisotropic proposal improves stability near the funnel neck compared to pure isotropic proposals, particularly when σ_{\parallel} is small. The mixture component proved critical: removing the isotropic proposal significantly reduced ESS, indicating that occasional global moves are necessary to avoid geometric trapping.

Ablation experiments reveal that:

- Removing the mixture component drastically reduces ESS.
- Excessively large σ_{\parallel} harms stability in the neck region.
- Moderate σ_{\perp} values improve exploration along wide regions.

Overall, TNM demonstrates improved robustness over RWMH in anisotropic regions but does not fully match HMC’s efficiency when gradients are well-behaved.

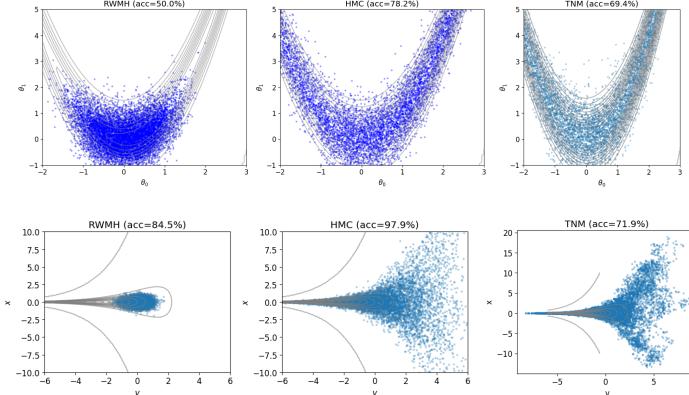


Figure 1: Comparison of sampling behavior on the Rosenbrock (top row) and Neal’s Funnel (bottom row) distributions. Left column: RWMH and HMC baselines. Right column: Proposed TNM sampler. TNM improves geometric alignment relative to isotropic proposals, though HMC remains superior in long-range decorrelation when well tuned.

3.5 Summary

Across both benchmarks, TNM provides a geometry-aware alternative to isotropic random walks while avoiding the full dynamical simulation of HMC. It improves manifold alignment relative to RWMH and remains stable in strongly anisotropic settings such as the funnel neck. However, its local random-walk character limits long-range decorrelation, leaving HMC superior in effective sample size when well tuned.

4 Discussion

TNM performs best in distributions where local geometric structure is informative but full Hamiltonian dynamics may be unnecessary or difficult to tune. On the Rosenbrock benchmark, TNM clearly improves over isotropic RWMH by aligning proposals with the curved ridge, reducing the frequency of off-manifold rejections. The anisotropic covariance allows the chain to move more freely along high-density directions while remaining conservative across sharp curvature. This demonstrates that even simple first-order geometric adaptation can substantially improve random-walk behavior.

On Neal’s Funnel, TNM exhibits improved robustness relative to pure isotropic proposals, particularly near the narrow neck region where scale differences become extreme. The ablation results show that the mixture component is essential: without occasional isotropic moves, the chain becomes locally aligned but globally trapped. This highlights an important design principle—local geometry awareness must be complemented by global exploration mechanisms.

However, TNM struggles to match HMC in effective sample size when gradients are smooth and well-behaved. Because TNM remains fundamentally a single-step Metropolis proposal, it lacks the long-range deterministic trajectories that enable HMC to decorrelate samples efficiently. As a result, autocorrelation remains visible even when acceptance rates are high. Increasing proposal scales to encourage longer jumps typically reduces acceptance, illustrating the inherent trade-off of random-walk-style methods.

Several extensions could improve TNM. First, incorporating a position-dependent scaling strategy (e.g., adapting σ_{\parallel} based on local curvature magnitude) could better handle strongly varying geometries such as the funnel. Second, a low-cost multi-step proposal (e.g., two or three geometry-aligned updates before MH correction) could approximate short Hamiltonian trajectories while retaining conceptual simplicity. Finally, extending TNM to higher-dimensional problems would require careful subspace construction, potentially using multiple gradient-informed directions rather than a single rank-1 structure.

Overall, TNM illustrates that meaningful gains can be achieved by modest geometric adaptation within the Metropolis–Hastings framework, but it also reinforces the strength of HMC when full gradient dynamics are feasible.

5 AI Collaboration

This assignment was primarily completed independently, with limited use of AI coding tools (ChatGPT) as a supplementary assistant. The core algorithmic design, experimental setup, ablation studies, and interpretation of results were developed and validated by the author.

AI assistance was used mainly for implementation support and formatting tasks, including:

- Debugging JAX-related issues such as random key handling and tensor shape mismatches.
- Clarifying library usage (e.g., ArviZ diagnostics and plotting functions).
- Minor LaTeX formatting adjustments within the NeurIPS template.

The conceptual design of the Tangential–Normal Mixture (TNM) sampler—including the rank-1 covariance structure, mixture strategy, drift mechanism, and hyperparameter selection—was independently proposed and empirically evaluated. All experimental conclusions were drawn from direct inspection of results and diagnostic metrics.

AI-generated suggestions were reviewed critically before integration, and in several cases were revised or discarded after testing. The AI tool was therefore used as a technical aid for implementation efficiency rather than as a source of core methodological ideas.

Overall, while AI assistance helped streamline debugging and formatting, the mathematical reasoning, sampler design, and experimental validation were conducted independently.