

The No-Underrun Sampler: a locally adaptive, gradient-free MCMC Method

Nawaf Bou-Rabee (Flatiron/Rutgers)

with Bob Carpenter (Flatiron), Sifan Liu (Flatiron/Duke), and Stefan Oberdörster (Bonn).

arXiv: 2501.18548 github.com/bob-carpenter/no-underrun-sampler

Markov Chain Monte Carlo¹

- ▶ **Goal:** Sample from a target probability measure μ on $\mathbb{R}^d$, with density also denoted by μ .
- ▶ **Approach:** Construct a Markov chain with transition kernel π such that $\mu = \mu\pi$.
- ▶ **Sufficient condition:** *Reversibility* (detailed balance):

$$\mu(dx) \pi(x, dx') = \mu(dx') \pi(x', dx)$$

guarantees that μ is invariant.

- ▶ **Two broad classes of MCMC methods:**
 - **Gradient-based:** transitions evaluate $\nabla \log \mu$ (e.g., MALA, HMC, NUTS).
 - **Gradient-free:** rely only on evaluations of μ (e.g., RWM, Goodman-Weare Sampler, pCN).

¹Diaconis, “The Markov Chain Monte Carlo Revolution,” *Bull. Amer. Math. Soc.*, 2009.

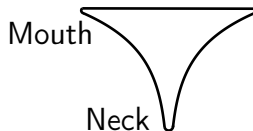
Motivation: Accelerated Mixing Without Gradients²

Problem: Multi-scale targets challenge MCMC methods.

State of the art: The *affine-invariant ensemble sampler* of (Goodman & Weare 2011) offers a gradient-free MCMC method via an interacting ensemble of walkers.

But: some practically relevant targets exhibit highly *local* geometry.

Example: Neal's funnel (Neal 2003), (Betancourt & Girolami 2013).

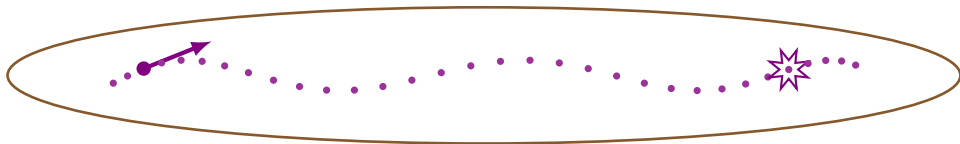


Wide mouth, narrow neck: extreme variation in scale.

²Diaconis, Neal, & Holmes, "Analysis of a nonreversible Markov chain sampler," *AAP*, 2000.

Motivation: Local Adaptivity without Gradients

Our idea: combine the coordinate-free direction sampling of **Hit-and-Run** with the path-length adaptivity of **NUTS** while remaining gradient-free.³



NUTS dynamically adapts path length to local scale — can we do this without evaluating $\nabla \log \mu$?

³NUTS requires local step-size adaptivity to handle multiscale targets (B.-R., Carpenter, Marsden 2024), (B.-R., Carpenter, Kleppe, Marsden 2024), (B.-R., Carpenter, Kleppe, Liu 2025)

Hit-and-Run

Key Idea: “Hit” a random direction, then “run” along it (Andersen & Diaconis 2007).

Origins: Introduced for general densities by (Turchin 1971), and for uniform sampling over convex bodies by (Smith 1984).

Theoretical Foundations:

- ▶ Proven diffusive for uniform sampling via conductance under a warm start (Lovász 1999).
- ▶ Warm start assumption removed via a warm-up phase (Lovász & Vempala 2006).

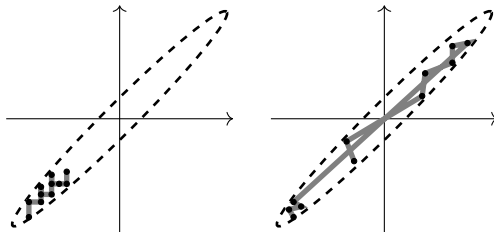
Recent Advances:

- ▶ Stochastic localization techniques yield improved mixing in isotropic bodies, replacing condition number with KLS constant (Chen & Eldan 2022).
- ▶ Proven diffusive for strongly log-concave targets (Ascolani, Lavenant & Zanella 2024).

Comparisons:

- ▶ **Ball Walk** (Kannan, Lovász, & Simonovits 1997): simpler but for uniform sampling may take exponential time to escape corners.
- ▶ **In-and-Out Sampler** (Kook, Vempala, & Zhang 2024): proximal sampler with similar runtime as Ball Walk but stronger Rényi divergence guarantees.
- ▶ **Random Walk Metropolis** (Andrieu et al. 2024): diffusive for strongly log-concave μ .

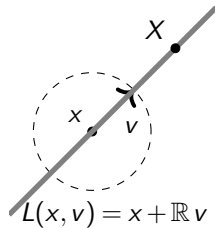
Hit-and-Run



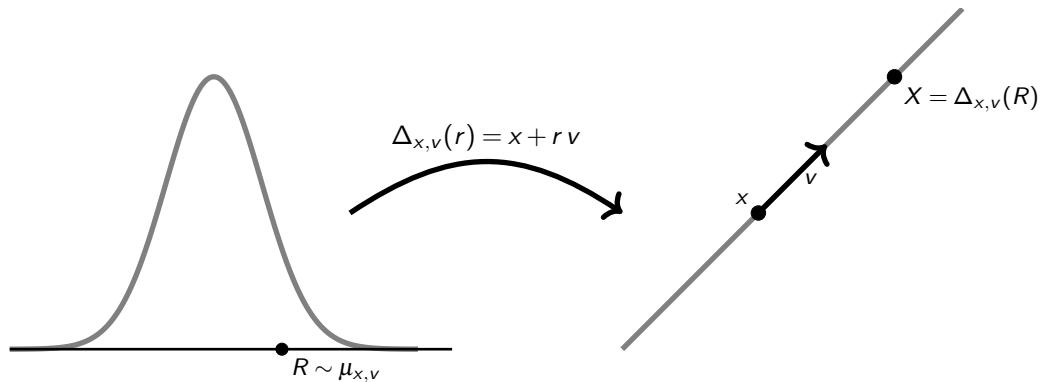
Coordinate-free alternative to Gibbs: allows accelerated mixing?

$$X \sim \pi_{\text{H\&R}}(x, \cdot)$$

- **Hit Step:** Sample $v \sim \text{Unif}(\mathbb{S}^{d-1})$.
- **Run Step:** Sample $X \sim \mu(\cdot | L(x, v))$.



Run Step: $X \sim \mu(\cdot | L(x, v)) = \mu_{x,v} \circ \Delta_{x,v}^{-1}$



Displacement Measure: $\mu_{x,v}(dr) \propto \mu(x + rv) dr$

Wasserstein Contraction of Hit-and-Run

Theorem (B.-R., Eberle, & Oberdörster 2025; EJP)

Consider $\mu = \mathcal{N}(0, C)$ on $\mathbb{R}^d$ and let $\rho = \frac{1}{2} \inf_{|\zeta|=1} \mathbb{E}_{v \sim \text{Unif}(\mathbb{S}^{d-1})} \left(\zeta \cdot \frac{C^{-1/2} v}{|C^{-1/2} v|} \right)^2$.

1. **Contraction:** For all $x, \tilde{x} \in \mathbb{R}^d$,

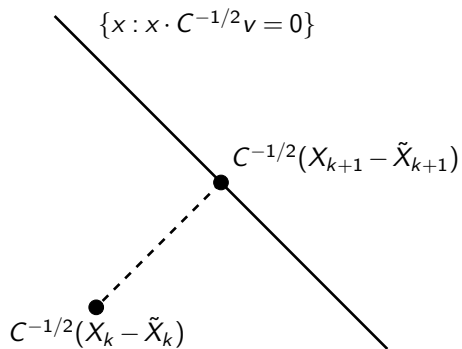
$$W_{C^{-1/2}}^2(\pi_{\text{H\&R}}(x, \cdot), \pi_{\text{H\&R}}(\tilde{x}, \cdot)) \leq (1 - \rho) |x - \tilde{x}|_{C^{-1/2}}.$$

2. **Sharpness:** This bound is sharp up to a constant factor: any $\rho' > 0$ such that

$$W_{C^{-1/2}}^2(\pi_{\text{H\&R}}(x, \cdot), \pi_{\text{H\&R}}(\tilde{x}, \cdot)) \leq (1 - \rho') |x - \tilde{x}|_{C^{-1/2}} \quad \text{for all } x, \tilde{x} \in \mathbb{R}^d,$$

must satisfy $\rho' \leq 6\rho$.

Hit-and-Run Coupling



In Hit-and-Run, the coupling difference contracts under random projection.

A General Lemma for Random Projections

Key idea: Random projections onto hyperplanes induce average contraction of distances.

Lemma. Let η be a probability measure on $\mathbb{R}^d$. For each $w \sim \eta$, define the projection

$$\Pi_w = I_d - \frac{w \otimes w}{|w|^2} \quad \text{onto } w^\perp.$$

Then for any $z \in \mathbb{R}^d$,

$$\mathbb{E}_{w \sim \eta} |\Pi_w z|^2 \leq \left(1 - \inf_{|\zeta|=1} \mathbb{E}_{w \sim \eta} \left(\zeta \cdot \frac{w}{|w|} \right)^2 \right) |z|^2.$$

► **Hit-and-Run:** $w = C^{-1/2}v$, with $v \sim \text{Unif}(\mathbb{S}^{d-1})$

Hit-and-Run vs. Randomized HMC vs. RWM

Complexity $\mathfrak{C}$ for $\mu = \mathcal{N}(0, \text{diag}(\kappa, 1))$, with $\kappa \gg 1$:

► **H&R:**

$$\mathfrak{C} \asymp t_{\text{mix}} \asymp \kappa^{1/2} \quad \text{for constant complexity per transition}$$

► **Randomized HMC:** mean time T , leapfrog step-size $h \asymp 1$

$$\mathfrak{C} \asymp t_{\text{mix}} \frac{T}{h} \asymp \begin{cases} \kappa & \text{for } T \asymp 1 \\ \kappa^{1/2} & \text{for } T \asymp \kappa^{1/2} \end{cases}$$

► **RWM:** proposal variance σ^2

$$\mathfrak{C} \asymp t_{\text{mix}} \asymp \kappa \quad \text{for } \sigma \asymp 1$$

Takeaway: Hit-and-Run achieves accelerated convergence, i.e., $\kappa \rightarrow \kappa^{1/2}$.

Accelerated Convergence of Hit-and-Run

Hit-and-Run⁴ in anisotropic Gaussians ($\kappa \gg 1$)

- For $\mu = \mathcal{N}(0, \text{diag}(\kappa, \kappa, 1))$

$$t_{\text{mix}} \asymp \kappa^{1/2}$$

- For $\mu = \mathcal{N}(0, \text{diag}(\kappa, 1, 1))$

$$t_{\text{mix}} \asymp \frac{\kappa}{\log \kappa}$$

- For $\mu = \mathcal{N}(0, \kappa I_{d_1}) \otimes \mathcal{N}(0, I_{d_2})$, $d_1, d_2 \gg 1$

$$t_{\text{mix}} \asymp \begin{cases} O(d_1 + d_2) & \text{if } d_2/d_1 = O(\kappa^{-1}) \\ O(\kappa^{1/2}(d_1 + d_2)) & \text{if } d_2/d_1 = O(\kappa^{-1/2}) \end{cases}$$

Prior work by (Lovász 1999), (Lovász & Vempala 2004), (Chen & Eldan 2022), (Ascolani, Lavenant, & Zanella 2024) show diffusive rates under more general conditions.

⁴(B.-R., Eberle, & Oberdörster 2025; EJP)

Why Hit-and-Run?

- ▶ **Tuning-free:** Involves no free parameters.
- ▶ **Coordinate-free:** Explores the space via random directions.
- ▶ **Theoretical guarantees:** Provable contraction in Wasserstein distance, conductance-based mixing bounds, entropy contraction, etc.
- ▶ **Accelerated convergence:** Achieves faster mixing if the number of high-variance directions is sufficiently large.
- ▶ **Main Challenge:** The Run step requires exact sampling from $\mu(\cdot \mid L(x, v))$, which is tractable only in special cases.

Implementing Hit-and-Run

Ansatz 1: Replace line by infinite h -lattice.

$$X \sim \pi_\infty(x, \cdot)$$

- **Hit Step:** Sample $v \sim \text{Unif}(\mathbb{S}^{d-1})$ and define lattice $L_h(x, v) = x + h\mathbb{Z}v$.
- **Run Step:** Sample $X \sim \text{Categorical}(L_h(x, v), \mu)$.

Ansatz 2: Truncate to a finite lattice.

- **Challenge:** Maintain reversibility.
- **Solution:** Build a finite lattice using a random doubling procedure (à la NUTS) stopping when the lattice $\mathfrak{L}_h = [x^{\text{left}}, \dots, x^{\text{right}}]$ satisfies the **No-Underrun condition**

$$\max(\mu(x^{\text{left}}), \mu(x^{\text{right}})) \leq \varepsilon h \sum_{\tilde{x} \in \mathfrak{L}_h} \mu(\tilde{x}) \quad \text{for } 0 < \varepsilon \ll 1.$$

Implementing Hit-and-Run

Ansatz 2: Replace line by finite h -lattice.

$$X \sim \pi_M(x, \cdot)$$

- **Hit Step:** Sample direction $v \sim \text{Unif}(\mathbb{S}^{d-1})$.
- **Run Step:**
 - Build $\mathcal{L}_h \subset L_h(x, v)$ via random doubling procedure + No-Underrun.
 - Sample $X \sim \text{Categorical}(\mathcal{L}_h, \mu)$.

► **Problem:** Not ergodic in dimension 1 (confined to lattice $L_h(x, 1)$).

► **Solution:** Randomly shift lattice.

The No-Underrun Sampler

Ansatz 3: Replace line by a randomly shifted, finite h -lattice.

$$X \sim \pi_{\text{NURS}}(x, \cdot)$$

- **Hit Step:** Sample direction $v \sim \text{Unif}(\mathbb{S}^{d-1})$.
- **Run Step:**
 - Random shift: propose $x_p \sim \text{Unif}(x + [-h/2, h/2] \cdot v)$; accept with Metropolis step.
 - Build $\mathcal{L}_h \subset L_h(x_s, v)$ via random doubling procedure + No-Underrun.
 - Sample $X \sim \text{Categorical}(\mathcal{L}_h, \mu)$.

► Parameters:

- Lattice spacing $h > 0$,
- No-Underrun threshold $\varepsilon > 0$,
- max number of doublings $M \in \mathbb{N}$.

► Tune h as in RWM (Andrieu et al. 2024).

Reversibility of NURS

Theorem (B., Carpenter, Liu, & Oberdörster 2025)

π_{NURS} is reversible with respect to μ .

Key idea: View NURS as the marginal of a reversible chain on an extended space.⁵

- ▶ Extended state: $z = (x, v, \ell, R, i)$ in $\mathcal{Z} = \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{N} \times \mathbb{N} \times \mathbb{Z}$.
- ▶ Joint density: $p_{\text{joint}}(z) \propto \mu(x) \tau(v) p_{\text{Orbit}}(\ell, R \mid x, v) \mu(x + hiv)$.
- ▶ Involution: define $\Psi(z) = (x + hiv, v, \ell, R - i, -i)$.
 - Ψ preserves reference measure and satisfies $p_{\text{joint}} \circ \Psi = p_{\text{joint}}$
 - $\Psi \circ \Psi = \text{id}$
- ▶ $\Rightarrow$ The lifted chain is reversible $\Rightarrow$ NURS is reversible w.r.t. μ

⁵(Andersen & Diaconis 2007), (B.-R., Carpenter, Marsden 2024)

Wasserstein Contraction of NURS

Theorem (B., Carpenter, Liu, & Oberdörster 2025)

Let $\mu = \mathcal{N}(0, C)$. For any $x, \tilde{x} \in \mathbb{R}^d$,

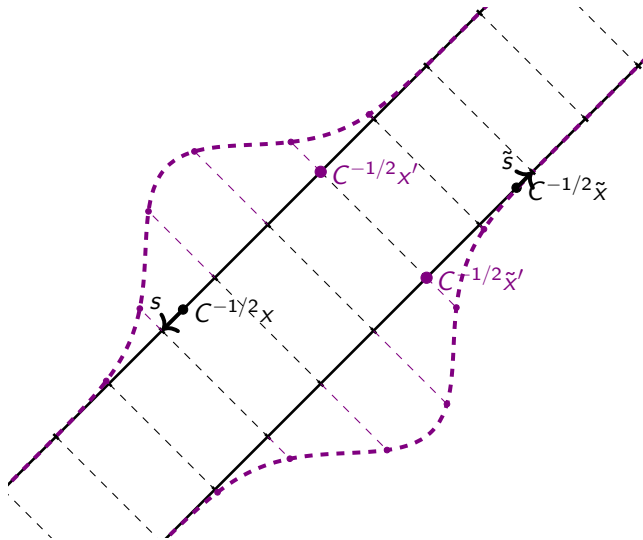
$$W_{C^{-1/2}}^2(\pi_{\infty, u}(x, \cdot), \pi_{\infty, u}(\tilde{x}, \cdot)) \leq (1 - \rho) |x - \tilde{x}|_{C^{-1/2}}$$

with the Hit-and-Run contraction rate ρ .

Key idea. Mimic H&R coupling:

- ▶ Sample shared direction $v \sim \text{Unif}(\mathbb{S}^{d-1})$
- ▶ Couple shifts $s, \tilde{s}$ to align scalar displacements on the same lattice
- ▶ Coupling ensures: $C^{-1/2}(x' - \tilde{x}') = \Pi_{C^{-1/2}v} C^{-1/2}(x - \tilde{x})$
- ▶ Apply bound for random projections

Coupling Illustration



Overlap with Hit-and-Run

Theorem (B., Carpenter, Liu, & Oberdörster 2025)

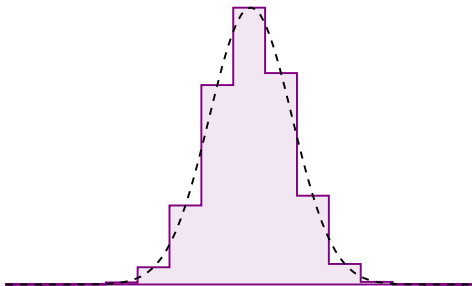
Let μ be differentiable on $\mathbb{R}^d$. For any $x \in \mathbb{R}^d$,

$$\mathrm{TV}(\pi_{\mathrm{H\&R}}(x, \cdot), \pi_{\infty, u}(x, \cdot)) \leq 2h \int_{\mathbb{R}^d} |\nabla \log \mu(x')| \pi_{\mathrm{H\&R}}(x, dx').$$

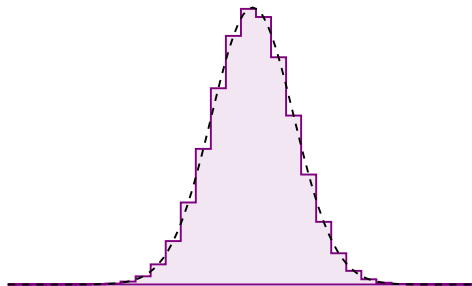
Key Idea.

- ▶ Compare H&R continuous line sampling with NURS discretized line sampling.
- ▶ Approximate displacement density using a piecewise-uniform density on lattice $h\mathbb{Z}$.
- ▶ Bound the TV distance using the smoothness of $\log \mu$ and a coupling argument.
- ▶ **Result:** NURS closely tracks H&R when h is small and μ is weakly differentiable.

Visualizing the Overlap Bound

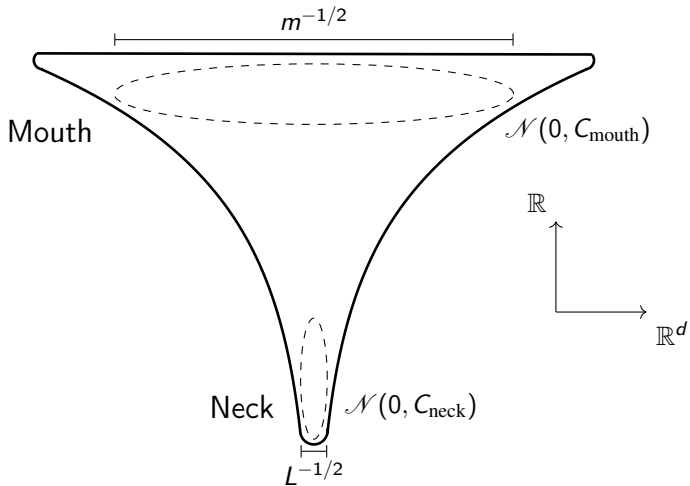


(a) *Lattice spacing h*



(b) *Lattice spacing $h/2$*

NURS in Neal's Funnel



Wide mouth, narrow neck: extreme variation in scale.

NURS in Neal's Funnel

Illustrates issues in Bayesian hierarchical models (Neal 2003), (Betancourt & Girolami 2013).

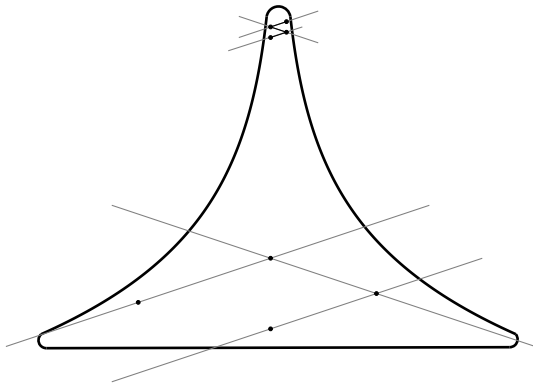
- ▶ **Geometry:** Narrow neck (high curvature), wide mouth (low curvature).
- ▶ **Setup:** Approximate extreme regions by Gaussians:

$$C_{\text{neck}} = \text{diag}(1, L^{-1}, \dots), \quad C_{\text{mouth}} = \text{diag}(1, m^{-1}, \dots), \quad m \ll L$$

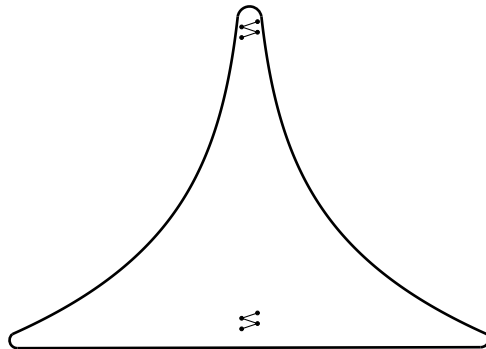
- ▶ **NURS:** adapts orbit length to local scale via No-Underrun condition.
- ▶ **Mixing:** Hit-and-Run achieves $O(1)$ mixing in the mouth, while NURS has complexity $O(\kappa^{1/2})$ due to its orbit cost, which is a significant improvement over RWM's $O(\kappa)$.

Takeaway: NURS adapts to local geometry, outperforming RWM.

NURS in Neal's Funnel



(a) *NURS*



(b) *RWM*

Recap and Outlook

Recap

- ▶ Introduced **NURS**: a locally adaptive, gradient-free MCMC method.
- ▶ Combines coordinate-free directions, orbit-based proposals, and a No-Underrun condition.
- ▶ Proven properties: **reversibility**, **Wasserstein contraction**, and **overlap with H&R**.
- ▶ Empirical results (Neal's funnel): **diffusive in narrow regions**, **ballistic in wide regions**.

Outlook

- ▶ **Lattice adaptivity**: locally tune lattice size via GIST (B.-R., Carpenter, & Marsden 2024).
- ▶ **Parallel implementation**: Exploit inherent parallelizability of NURS.
- ▶ **Beyond Gaussians**: extend convergence theory via conductance (Andrieu et al. 2024), stochastic localization (Chen & Eldan 2022) or coupling (B.-R. & Oberdörster 2024).
- ▶ **Direction adaptivity**: via, e.g., ensemble samplers (Goodman & Weare 2010).

Further Reading

- ▶ Companion papers: [arXiv:2412.07643](#) and [arXiv:2501.18548](#)
- ▶ Companion repo: github.com/bob-carpenter/no-underrun-sampler

Temporary page!

$\text{\LaTeX}$ was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it. If you rerun the document (without altering it) this surplus page will go away, because $\text{\LaTeX}$ now knows how many pages to expect for this document.