# Metropolis Adjusted Microcanonical Hamiltonian Monte Carlo

### An Unbiased, Black-Box Alternative to NUTS

Jakob Robnik,
Reuben Cohn-Gordon,
Uroš Seljak

NEURAL INFORMATION PROCESSING SYSTEMS

## Motivation

- Sampling high-dimensional, complicated posteriors is a core bottleneck in science.
- HMC/NUTS [1] are strong defaults but can struggle in very high dimensions or with tricky geometry.
- MCLMC and MCHMC [1] show excellent speed but are biased.
- **Goal:** Keep microcanonical stability and speed, remove bias, and make it black-box.

## Background

**Canonical dynamics (HMC):**

$$\dot{\boldsymbol{x}} = \boldsymbol{u} \qquad \dot{\boldsymbol{u}} = \nabla \log p(\boldsymbol{x})$$

**Microcanonical dynamics (MCHMC):**

$$\dot{\boldsymbol{x}} = \boldsymbol{u} \qquad \dot{\boldsymbol{u}} = (I - \boldsymbol{u}\boldsymbol{u}^T)\nabla \log p(\boldsymbol{x})/(d-1)$$

Exact microcanonical flow preserves p(x) as the marginal distribution, making it a valid sampling strategy.

## Take-aways

- **Unbiased, black-box, fast.**
- **Simple to implement:** Minimal changes from HMC.
- **Better efficiency:** Substantial gains over NUTS across diverse problems.
- Promising as a new **default sampler** for high-dimensional, gradient-based inference.

## Algorithm

MAMS proposes trajectories using microcanonical dynamics and corrects with an MH step. The acceptance probability is

$$e^{-W(\boldsymbol{x}',\boldsymbol{x})} = \frac{p(\boldsymbol{x}')}{p(\boldsymbol{x})}\frac{q(\boldsymbol{x}|\boldsymbol{x}')}{q(\boldsymbol{x}'|\boldsymbol{x})}$$

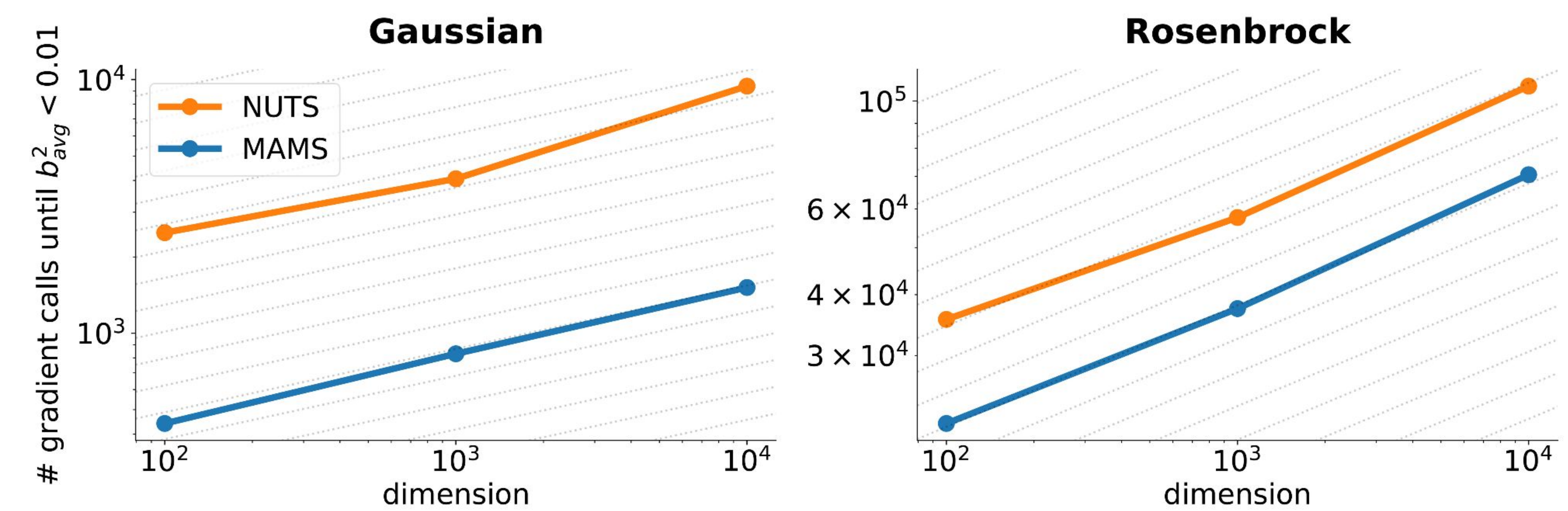
where W is the **total energy error** of the proposal.

- In HMC, W is the Hamiltonian energy error.
- In MAMS, the same principle applies, despite the different dynamics.

**Partial refresh (Langevin variant):** Mixes the velocity with Gaussian noise before/after deterministic steps to reduce cycling. The MH step still depends only on the deterministic energy error.

1. Inputs
   - Log-density log p(x) with gradients (or a PPL model)
   - Initial point x.
2. Propose
   - Sample velocity u uniformly on the unit sphere.
   - (Optional) partial refresh: lightly mix with noise, renormalize.
   - Run a leapfrog-style trajectory, accumulate energy error W.
3. Accept/Reject
   - Accept new state with probability min(1, $e^{-W}$)
   - Otherwise, keep the current state.
4. Automatic tuning
   - Step size: dual averaging to ~90% acceptance.
   - Preconditioner: diagonal, from parameter variances.
   - Trajectory length: proportional to autocorrelation length of the chain

## Experiments




***Scaling of sampling cost with dimension.***
Number of gradient evaluations to reach low squared error vs. dimensionality. Left: standard Gaussian; right: Rosenbrock target. Both MAMS and NUTS scale as $d^{1/4}$, but MAMS achieves consistently fewer gradients.

| | NUTS | MALA | MAMS | MAMS (Langevin) | MAMS (Grid Search) |
|---|---|---|---|---|---|
| Gaussian | 19,652 | 11,010 | 3,249 | **3,172** | 3,121 |
| Banana | 95,519 | 140,524 | **14,078** | 14,818 | 15,288 |
| Bimodal | 210,758 | $> 10^6$ | 139,418 | **136,770** | 123,295 |
| Rosenbrock | 161,359 | $> 10^6$ | **94,184** | 103,545 | 93,782 |
| Cauchy | 171,373 | 824,429 | **110,404** | 155,963 | 87900 |
| Brownian | 29,816 | 597,119 | **13,528** | 15,232 | 14,015 |
| German Credit | 88,975 | $> 10^6$ | 55,748 | **49,979** | 52,265 |
| ItemResp | 76,043 | 249,470 | **45,371** | 56,902 | 45,640 |
| StochVol | 843,768 | $> 10^6$ | **430,088** | 510,190 | 431,957 |
| Funnel | $> 10^8$ | $> 10^7$ | 2,346,899 | **1,765,311** | 1,013,048 |

***Gradients to low error across benchmarks (lower is better).***
For each target, we report the gradient evaluations needed to reduce the worst standardized second-moment error below 0.01. MAMS (with/without partial refresh) consistently outperforms **NUTS**, often by large factors. Automatic tuning achieves nearly the same performance as grid-searched hyperparameters.

**Scaling:** MAMS matches HMC/NUTS scaling laws $\kappa^{1/2}$ with condition number, $d^{1/4}$ with dimension, but with a **better constant**.

**Benchmarks:** Across models, MAMS outperforms NUTS by factors of 2–7.

**Difficult geometry (Neal's funnel):** MAMS converges, while NUTS fails.

### References

[1] Matthew D. Hoffman and Andrew Gelman (2014). "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." In: *Journal of Machine Learning Research* 15, pp. 1593–1623.

[2] Jakob Robnik and Uroš Seljak (2024). "Fluctuation without dissipation: Microcanonical Langevin Monte Carlo." In: Proceedings of the 6th Symposium on Advances in Approximate Bayesian Inference, pp. 111–126. PMLR. ISSN: 2640-3498.

[3] Jakob Robnik et al. (2024). "Microcanonical Hamiltonian Monte Carlo." In: *Journal of Machine Learning Research 24(1), 311:14696–311:14729.*