Hamiltonian Monte Carlo

MSc in AI — Probabilistic Methods

Ignacio Bayón Jimenez-Ugarte
Sergio Herreros Pérez
Adrián López-Lanchares Echezarreta

Universidad Pontificia Comillas, ICAI

08/29/2025

Table of Contents

COMILLAS UNIVERSIDAD PONTIFICIA ICAI

The Sampling Problem

Goal: draw samples from a target posterior distribution

$$p(\theta \mid \text{data}) = \frac{p(\text{data} \mid \theta)\, p(\theta)}{Z}, \quad Z = \int p(\text{data} \mid \theta)\, p(\theta)\, d\theta.$$

Issue: the normalising constant Z is often intractable.

Workaround in MCMC:

- We only need the unnormalised density

$$\tilde{\pi}(\theta) = p(\text{data} \mid \theta)\, p(\theta) \propto p(\theta \mid \text{data}).$$

- MCMC algorithms use ratios such as $\tilde{\pi}(\theta')/\tilde{\pi}(\theta)$, where Z cancels out.

Conclusion: We can sample from the posterior using only the *unnormalised* density $\tilde{\pi}(\theta)$.

The Metropolis-Hastings Algorithm

Goal: construct a Markov chain whose stationary distribution is the target posterior $\pi(\theta)$ (known up to a normalising constant).

Algorithm (one iteration):

1. At current state θ, draw a proposal

$$\theta' \sim q(\theta' \mid \theta) \quad (q\text{: proposal distribution}).$$

2. Compute the acceptance probability

$$\alpha = \min\left(1,\ \frac{\tilde{\pi}(\theta')\, q(\theta \mid \theta')}{\tilde{\pi}(\theta)\, q(\theta' \mid \theta)}\right).$$

3. With probability α, move to θ'; otherwise, stay at θ.

Key properties:

- Ensures $\pi(\theta)$ is the stationary distribution.
- Works with the *unnormalised* density $\tilde{\pi}(\theta)$.
- Very general: any proposal q is allowed (almost).

Limitations of Metropolis-Hastings

Random-walk proposals:

$$\theta' = \theta + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2 I).$$

Problems:

- Moves are local and uninformed.
- Slow exploration of the parameter space.
- High autocorrelation between samples.
- Very low acceptance rates for larger moves.
- Curse of dimensionality.

Result: MH becomes inefficient in high dimensions or for correlated posteriors.

Table of Contents

COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

Introduction to Hamiltonian Dynamics

Idea: Use physics to propose efficient moves in the parameter space.

- Treat parameters as a **position** vector q.
- Introduce an auxiliary **momentum** variable p.
- Define a **total energy** (Hamiltonian)

$$H(q, p) = U(q) + K(p).$$

- The particle moves through the energy landscape following smooth, long-range trajectories.

Why this helps: proposals follow the geometry of the target distribution, reducing random-walk behaviour.

Potential and Kinetic Energy

Potential energy:

$$U(q) = -\log \tilde{\pi}(q),$$

where $\tilde{\pi}(q)$ is the unnormalised target density.

- High-probability regions $\Rightarrow$ low potential energy.
- The energy landscape encodes the geometry of the posterior.

Kinetic energy:

$$K(p) = \tfrac{1}{2} p^\top M^{-1} p, \qquad p \sim \mathcal{N}(0, M).$$

- Introduces momentum to help the sampler move through q-space.
- Mass matrix M controls scaling and correlations.

Hamiltonian: $H(q, p) = U(q) + K(p).$

COMILLAS UNIVERSIDAD PONTIFICIA ICAI

Hamilton's Equations

Given $H(q, p) = U(q) + K(p)$, the dynamics are defined by:

$$\frac{dq}{dt} = \frac{\partial H}{\partial p}, \qquad \frac{dp}{dt} = -\frac{\partial H}{\partial q}.$$

Interpretation:

- q evolves according to the momentum.
- p evolves according to the gradient of the potential.
- Total energy H is approximately conserved.

These trajectories form the basis of efficient proposals in HMC

Table of Contents

COMILLAS UNIVERSIDAD PONTIFICIA ICAI

The Leapfrog Integrator

Goal: simulate Hamilton's equations numerically while preserving the key geometric properties of the dynamics.

Given step size ε, one leapfrog step:

$$p \leftarrow p - \frac{\varepsilon}{2} \nabla U(q) \qquad \text{(half momentum update)}$$

$$q \leftarrow q + \varepsilon M^{-1} p \qquad \text{(full position update)}$$

$$p \leftarrow p - \frac{\varepsilon}{2} \nabla U(q) \qquad \text{(half momentum update)}$$

Why leapfrog?

- **Reversible:** can run forwards or backwards.
- **Volume-preserving:** Jacobian determinant $= 1$.
- **Nearly energy-conserving:** for moderate step sizes.

Result: long, stable trajectories = high-acceptance proposals.

The HMC Algorithm

Goal: propose distant states with high acceptance using Hamiltonian trajectories.

One HMC iteration:

1. **Sample momentum**
$$p \sim \mathcal{N}(0, M).$$
2. **Simulate a trajectory** Run L leapfrog steps with step size ε to obtain (q^*, p^*) from the initial (q, p).
3. **Metropolis acceptance step**
$$\alpha = \min\left(1,\ \exp\left[-H(q^*, p^*) + H(q, p)\right]\right).$$
Accept q^* with probability α, else keep q.

Key idea: leapfrog produces proposals with *almost-conserved energy*, so acceptance rates are high.

Why HMC Works Well

Key advantages:

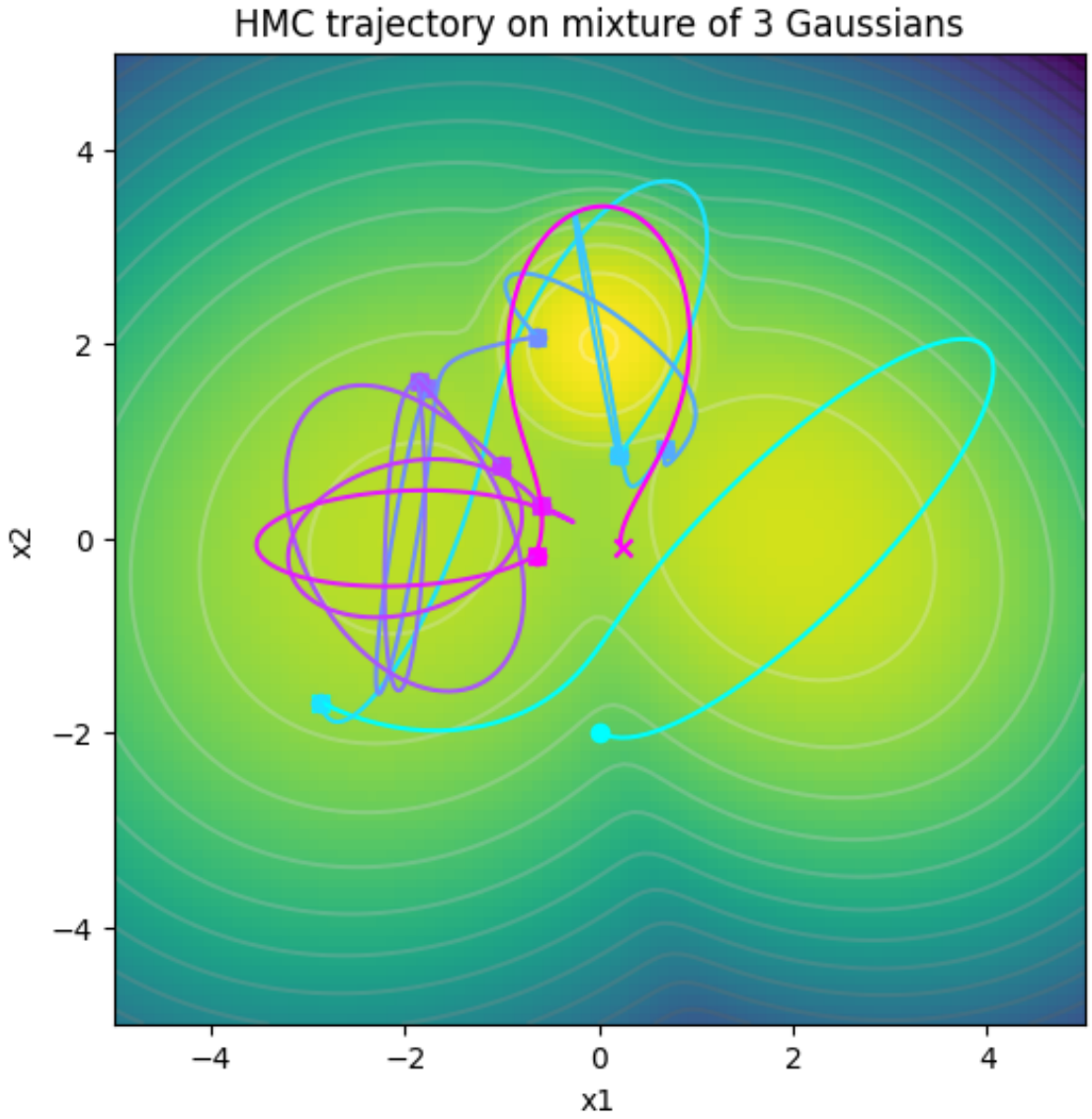
- **Informed, gradient-driven proposals:** by following the geometry of $\log \tilde{\pi}(q)$, HMC avoids random walks and therefore scales much better to high-dimensional spaces
- **Statistical Independence:** leapfrog integration produces coherent paths through the parameter space.
- **High acceptance rates:** energy is almost conserved, so HMC rejections are rare.
- **Low autocorrelation:** successive samples are less dependent, improving effective sample size.

Result: HMC explores complex posteriors much more efficiently than standard Metropolis-Hastings.

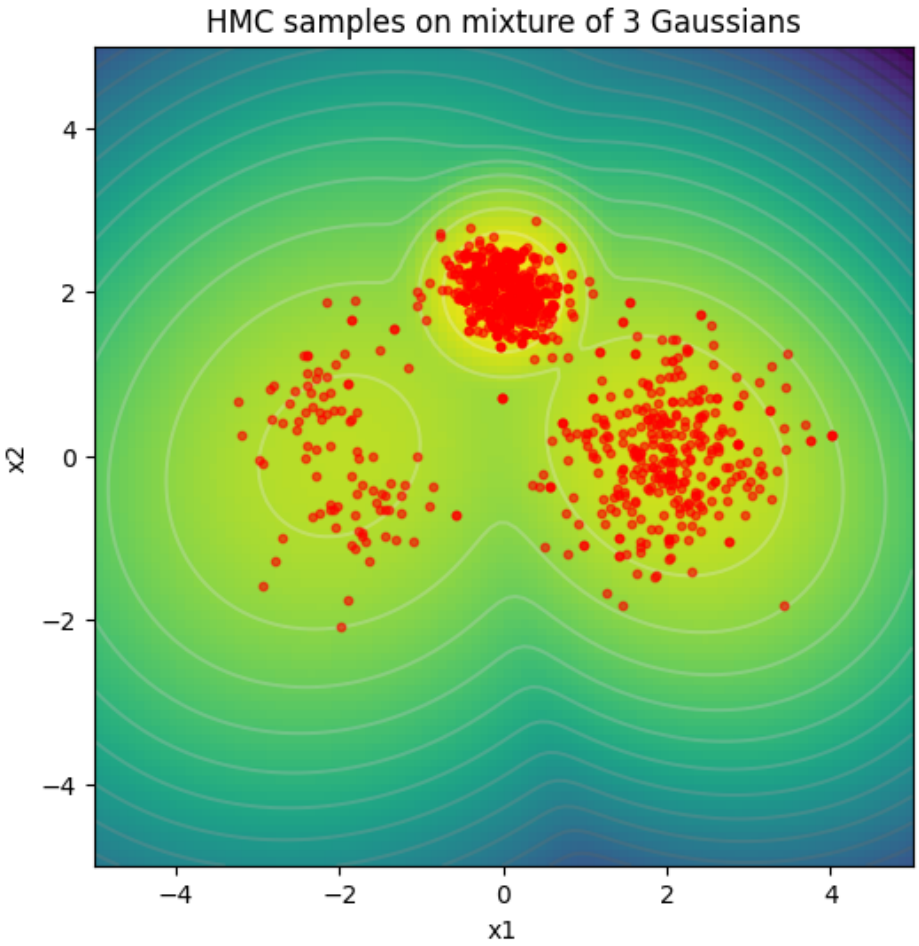
Table of Contents

COMILLAS UNIVERSIDAD PONTIFICIA ICAI

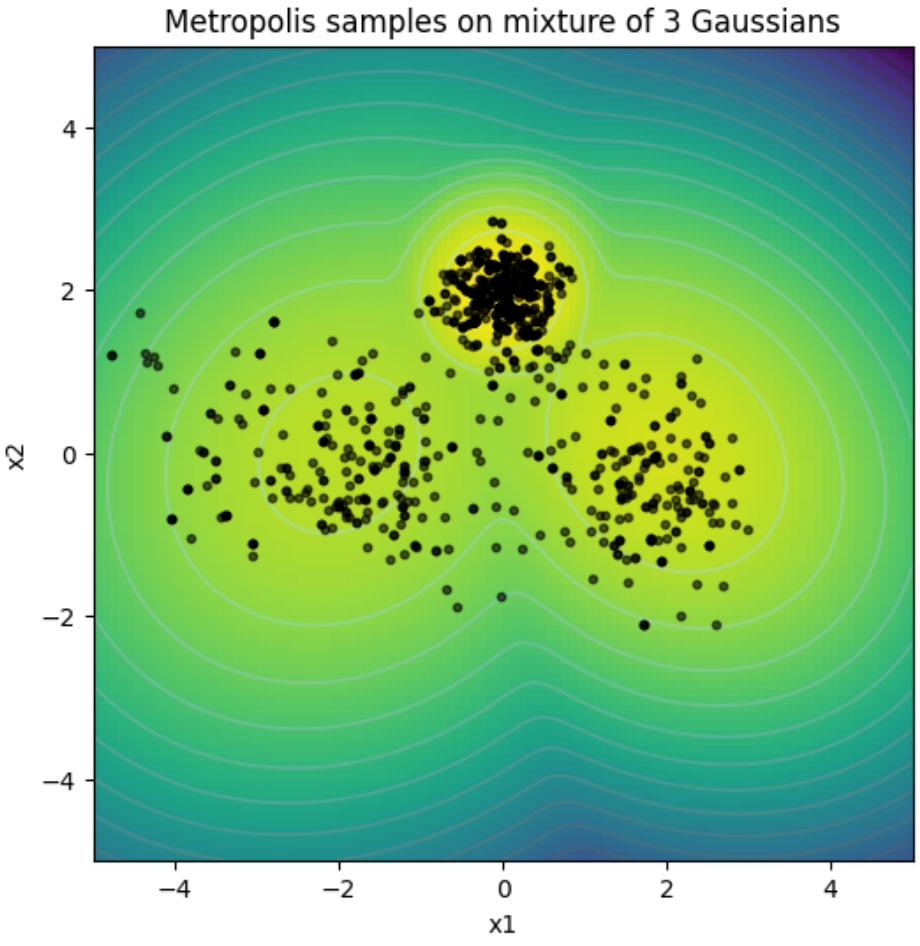
HMC Trajectories in a Gaussian Mixture Model



HMC vs Metropolis-Hastings: Sample Comparison



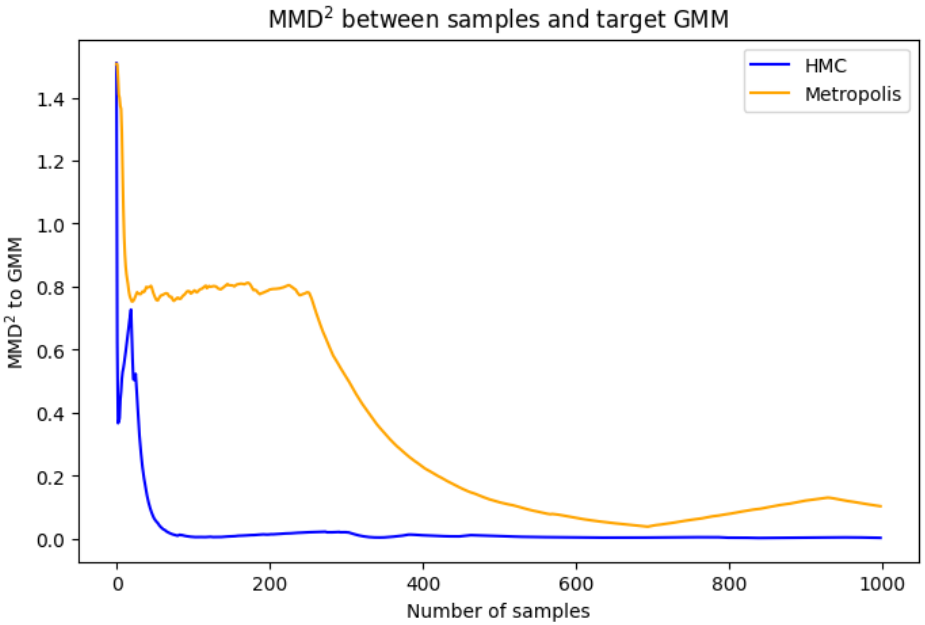
Hamiltonian Monte Carlo



Metropolis Hastings

COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

MMD^2 Comparison



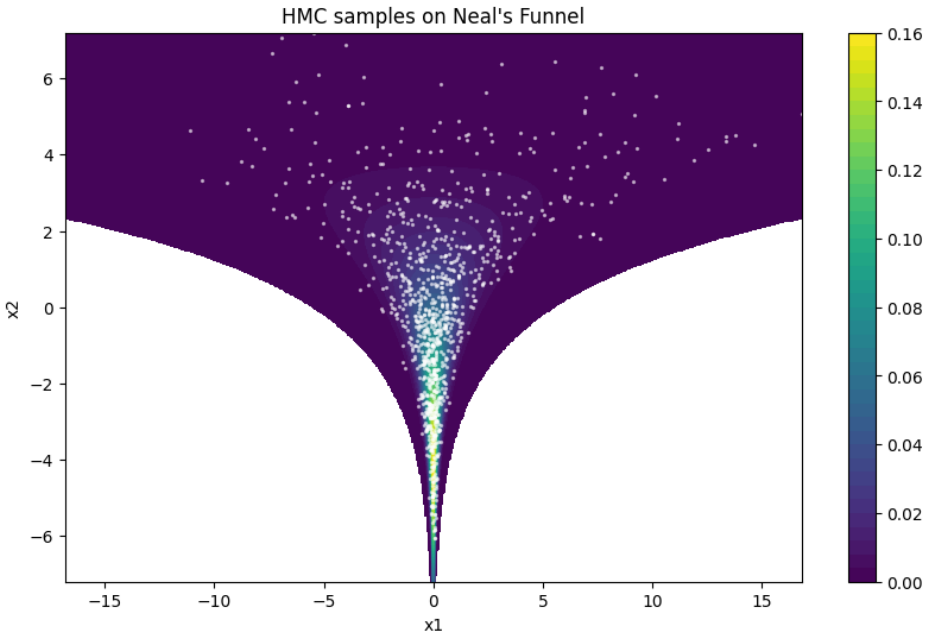
Key observations:

- **HMC converges quickly:** near-zero MMD^2 after $\sim$100 samples.
- **MH converges slowly:** needs several hundred samples.

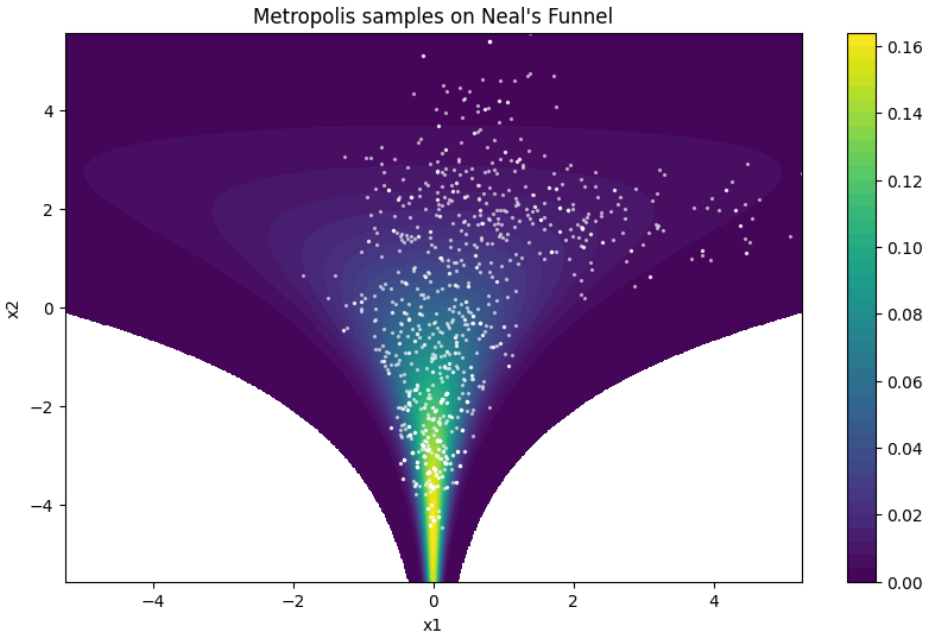
- **Practical implication:** in many applications, the goal is to obtain plausible samples rather than to perfectly reconstruct the full target distribution.
- **Result:** HMC produces high-quality samples with far fewer iterations, making it a better choice than standard Metropolis-Hastings.

COMILLAS UNIVERSIDAD PONTIFICIA ICAI

HMC vs Metropolis-Hastings: Sample Comparison



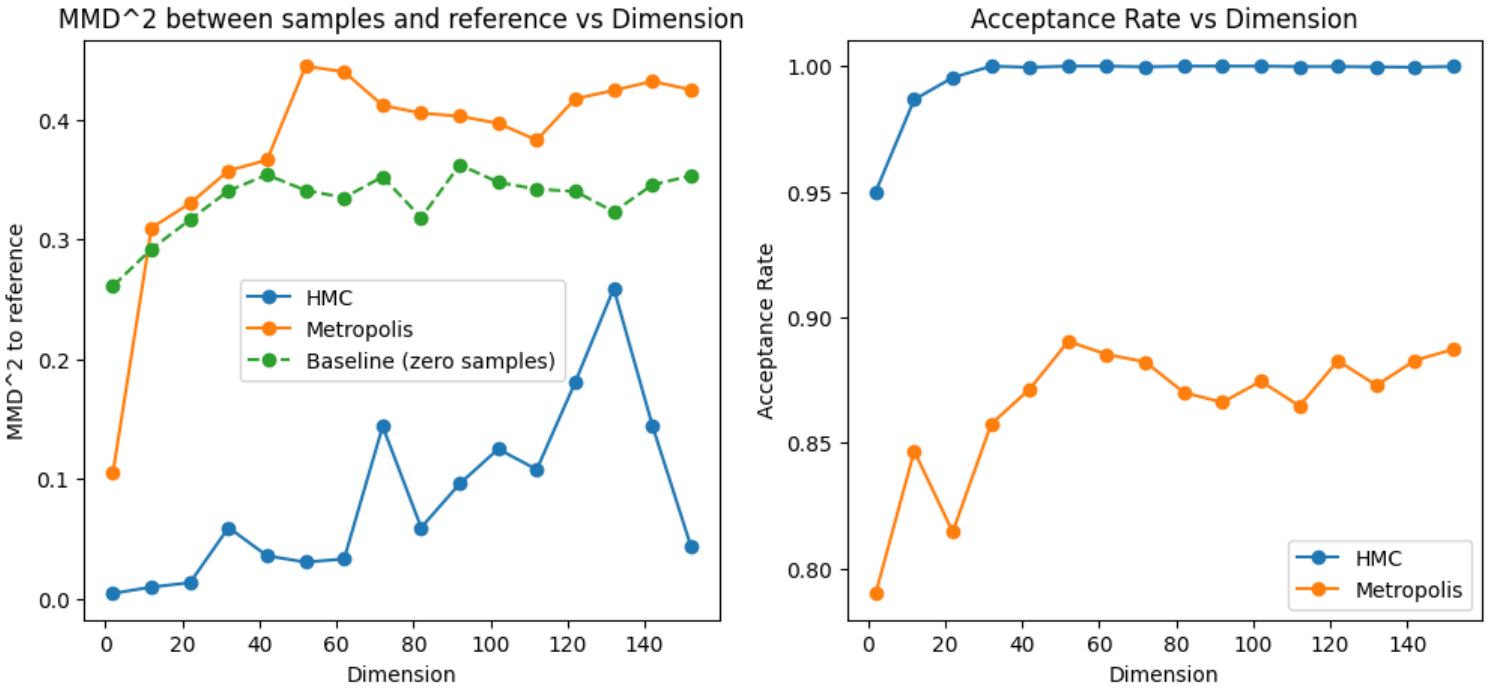
Hamiltonian Monte Carlo



Metropolis Hastings

COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

MMD^2 Comparison



- **Practical implication:** Hamiltonian Monte Carlo sampling can be much more effective in high-dimensional distributions.
- **Result:** HMC bypasses the curse of dimensionality and its acceptance rarte is much higher, making it a faster sampling algorithm.

COMILLAS UNIVERSIDAD PONTIFICIA ICAI

Thank you for your attention!

Questions?

Code:
`https://github.com/IgnacioBayon/hamilton-monte-carlo`