# How and When to Use Hamiltonian Monte Carlo

Course : Stochastic Simulation
Students : Aude Maier & Tara Fjellman
Fall 2024

## 1  Introduction

## 2  Core Theory

### 2.1  RWMC

### 2.2  HMC

#### 2.2.1  The Algorithm

#### 2.2.2  Acceptance Rate

#### 2.2.3  Convergence to Target Distribution

## 3  (c)

### 3.1  Gibbs distribution invariance

Under the assumption that there is no numerical error, we want to prove that the Gibbs measure is invariant for the chain generated by the hamiltonian dynamics.

This is equivalent to saying that the Gibbs measure $\pi$ is the same before and after an evolution of $t$ seconds from the hamiltonian dynamics. To prove this we first introduce hamiltonian dynamics operators $\varphi, \Phi$ acting respectively on the phase space and the Gibbs measure : $\varphi_t(q_s, p_s) = (q_{s+t}, p_{s+t}); \Phi_t[\pi_s] = \pi_{t+s} \quad \forall t \in \mathbb{R}$. The statement we want to prove can then be expressed as

$$\Phi_t[\pi_s](D) = \pi_{s+t}(D) \quad \forall D \in \mathcal{B}(\Omega), \forall s, t \in \mathbb{R}, \tag{1}$$

with $\Omega$ the phase space.

We can now write the left hand side of the equation as

$$\Phi_t[\pi_s](D) = \int_D \pi_{s+t}(q, p) \ dqdp \tag{2}$$

$$= \int_{\varphi_{-t}(D)} \pi_s(q, p) \ dqdp \tag{3}$$

$$= \pi_s(\varphi_{-t}(D)). \tag{4}$$

The final result is obtained using the fact that volumes in phase space are preserved by the hamiltonian dynamics (in conservative systems). This result is known as Liouville's theorem, but is mentioned as theorems 2.3 in [cite].

This implies specifically that $q_k \sim \pi$ for all $k \in \mathbb{N}$ if $q_0 \sim \pi$.

If the dynamics is discretised with the Velocity Verlet algorithm, the volume in phase space is preserved up to a small error, which is why the algorithm is used in practice [cite wikipedia].
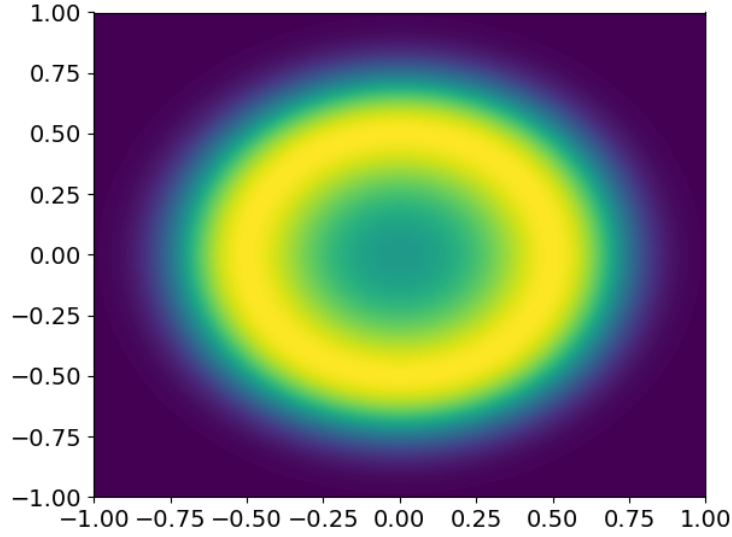
### 3.2  What happens for discrete evolution ?

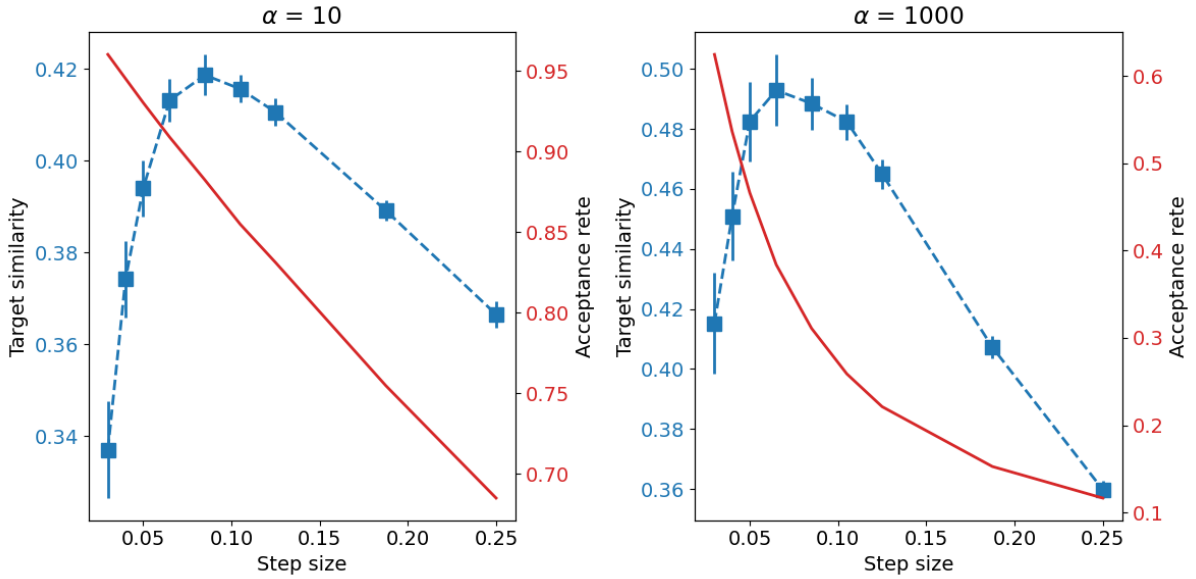## 4  Exploring a 2D example

### 4.1  Context

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy

eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.
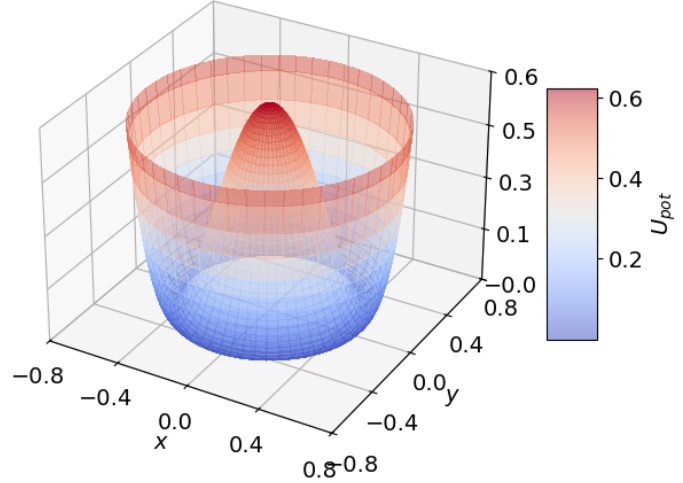


**Fig 1:** *Hello there.*

## 4.2 RWMC Solution



**Fig 2:** *Similarity as a function of RWMC step size for considered values of $\alpha$.*

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy

eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

### 4.3 HMC Solution

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices



**Fig 3:** *Potential energy landscape associated to HMC algorithm for $\alpha = 10$. Version for $\alpha = 1000$ is identical, except scales are scaled by a factor of 100.*

bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.
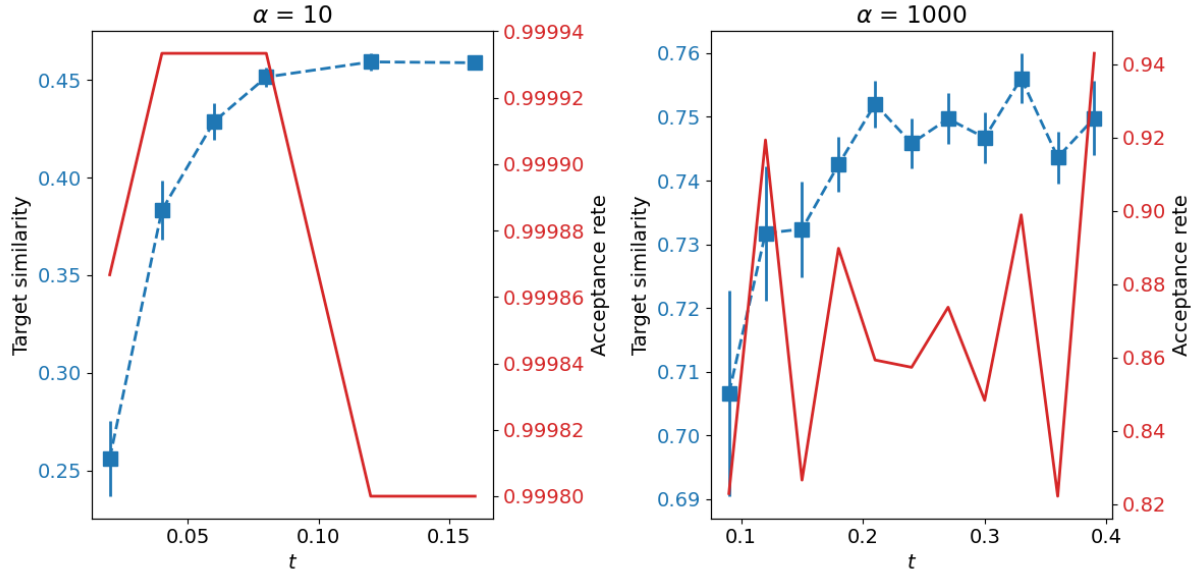
#### 4.3.1 Impact of Integration Time

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

### 4.4 Impact of $\Delta t$

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus

**Fig 4:** *Similarity as a function of HMC integration time for considered values of* $\alpha$.
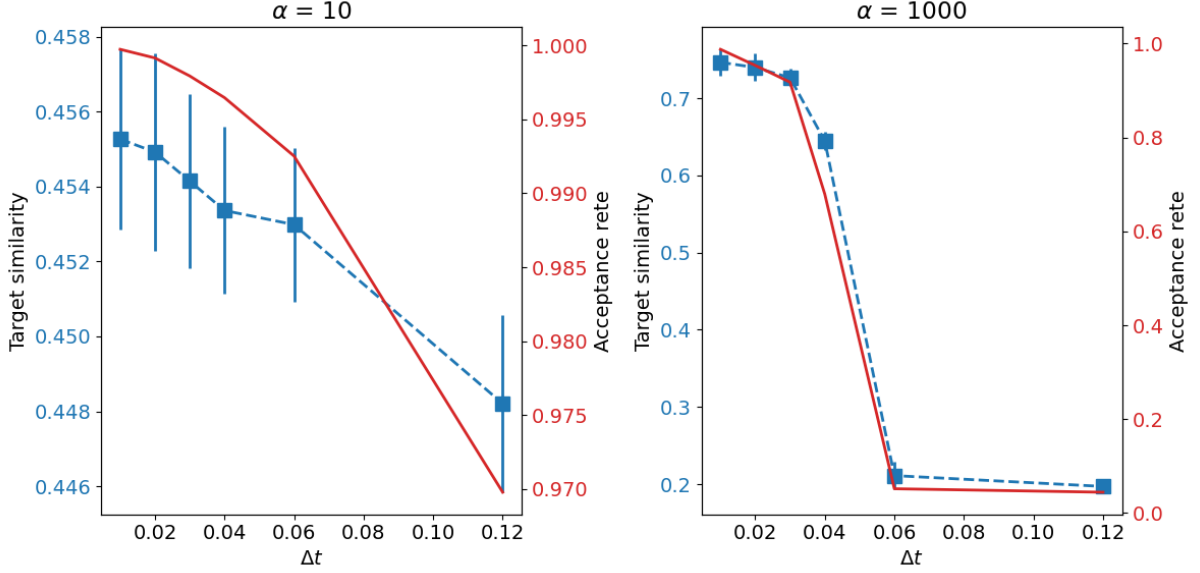
eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

### 4.5 Impact of Mass Scale

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

### 4.6 Impact of Mass Symmetry

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

**Fig 5:** *Similarity as a function of HMC time step for considered values of $\alpha$.*

## 4.7 Comparison and Sample-Size Evolution

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

# 5 US Birthweight Data

## 5.1 HMC Solution

## 5.2 RWMC Solution

# 6 (f)

As an alternative to HMC we consider rejection sampling. We therefore want to find a function $g(q)$ and a constant $C$ such that the following inequality holds for all $q$:

$$\tilde{f}(q) = e^{q^T X^T (y-1_n)} e^{-1_n^T \log[1+\exp(-x_i^T q)]_{n \times 1}} e^{-\frac{1}{2} q^T \Sigma^{-1} q} \leq Cg(q), \tag{5}$$
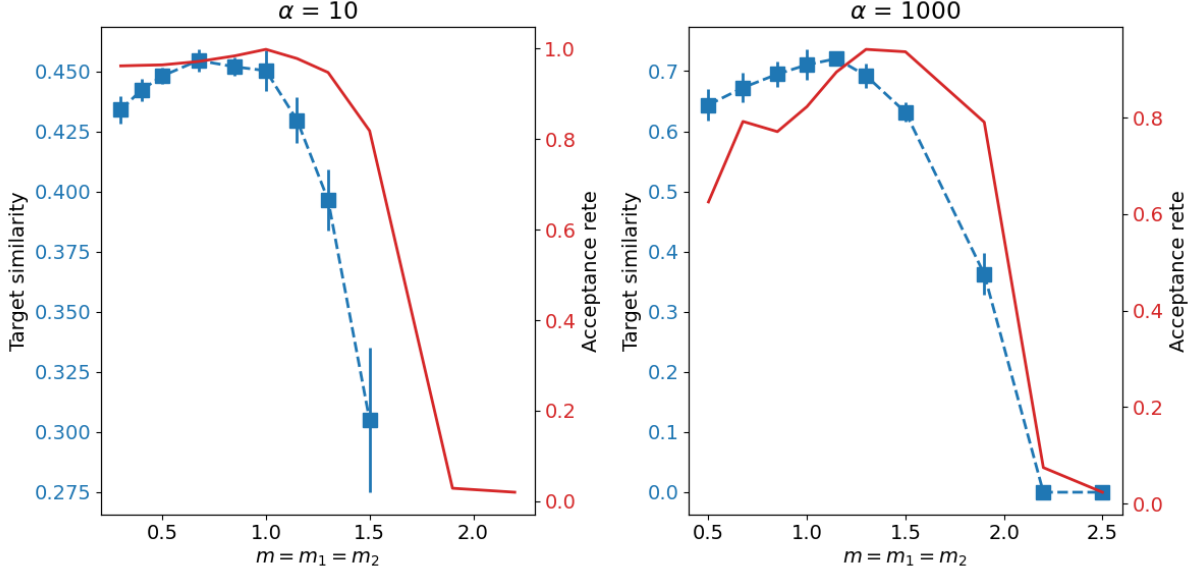
where we have denoted $\Sigma = \text{Diag}(\sigma_1^2, ..., \sigma_p^2)$. Given that

$$\log[1 + \exp(-x_i^T q)] \leq \log(2) - x_i^T q, \tag{6}$$
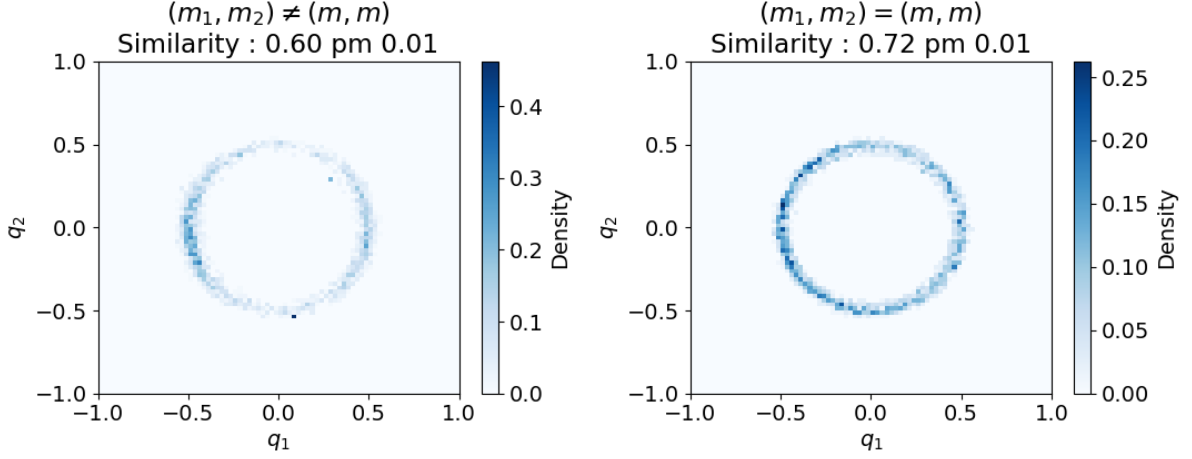
$$e^{-\sum_i \log[1+\exp(-x_i^T q)]} = \prod_i \frac{1}{1 + \exp(-x_i^T q)} < 1, \tag{7}$$

we can simplify the problem to finding a function $g(q)$ such that

$$\tilde{f}(q) \leq 2^{-n} e^{-q^T X^T 1_n} e^{q^T X^T (y-1_n)} e^{-\frac{1}{2} q^T \Sigma^{-1} q} = 2^{-n} e^{q^T b} e^{-\frac{1}{2} q^T \Sigma^{-1} q} =: Cg(q), \tag{8}$$

**Fig 6:** *Similarity as a function of HMC mass scale for considered values of $\alpha$.*



**Fig 7:** *Similarity for HMC samplers with asymmetric and symmetric masses for $\alpha = 1000$.*

with $b = X^T(y - 2_n)$.

By completing the square in the exponent of $Cg(q)$, we can write it in terms of a Multivariate Gaussian distribution with mean $\mu = \Sigma b$ and covariance $\Sigma$. Indeed :

$$e^{-\frac{1}{2}(q-\mu)^T\Sigma^{-1}(q-\mu)} = e^{-\frac{1}{2}\mu^T\Sigma^{-1}\mu}e^{q^T\Sigma^{-1}\mu}e^{-\frac{1}{2}q^T\Sigma^{-1}q} \tag{9}$$

$$\implies \tilde{f}(q) \leq 2^{-n}e^{\frac{1}{2}\mu^T\Sigma^{-1}\mu}e^{-\frac{1}{2}(q-\mu)^T\Sigma^{-1}(q-\mu)}. \tag{10}$$
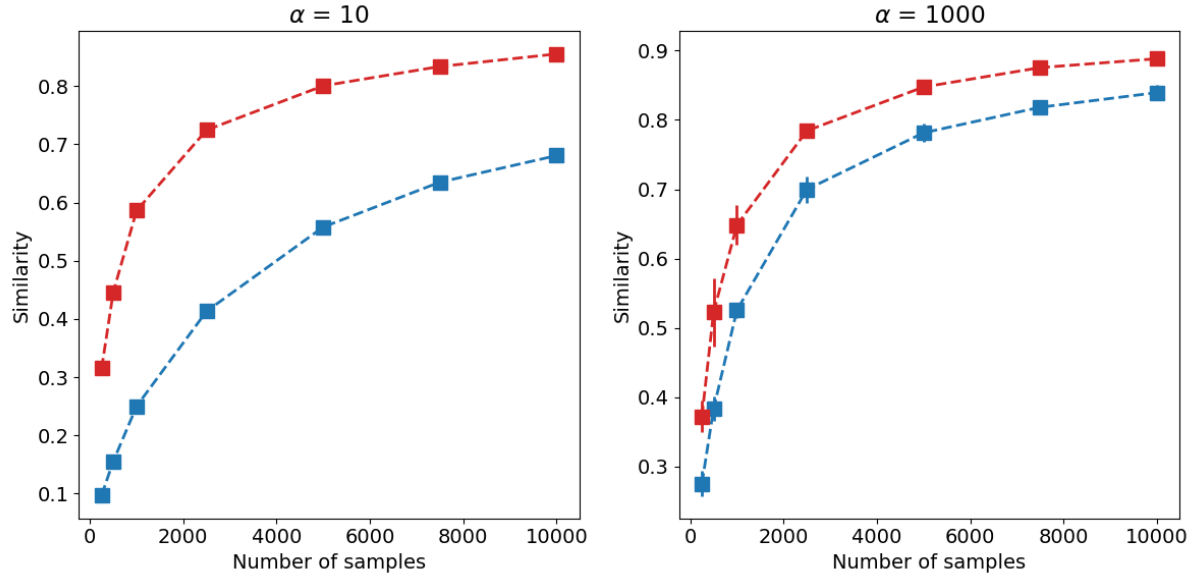
Using now the normalisation constant of the Multivariate Gaussian distribution

$$\sqrt{(2\pi)^p|\Sigma|} = \int_{\mathbb{R}^p} e^{-\frac{1}{2}(q-\mu)^T\Sigma^{-1}(q-\mu)} \, dq, \tag{11}$$

we can define $g$ and $C$ as

$$g(q) = \frac{1}{\sqrt{(2\pi)^p|\Sigma|}}e^{-\frac{1}{2}(q-\mu)^T\Sigma^{-1}(q-\mu)}, \tag{12}$$

$$C = 2^{-n}e^{\frac{1}{2}\mu^T\Sigma^{-1}\mu}\sqrt{(2\pi)^p|\Sigma|} = 2^{-n}\sqrt{(2\pi)^p|\Sigma|e^{\mu^T\Sigma^{-1}\mu}}. \tag{13}$$

**Fig 8:** *Similarity as a function of sample-size for RWMC and HMC samplers and considered values of $\alpha$. The sample-size represented on the x axis is that associated to the HMC sampler. The associated RWMC sample-size is obtained by matching the number of function evalutations of the HMC sampler.*

# 7 Section ...

# 8 Conclusion

# Aknowledgements

# References