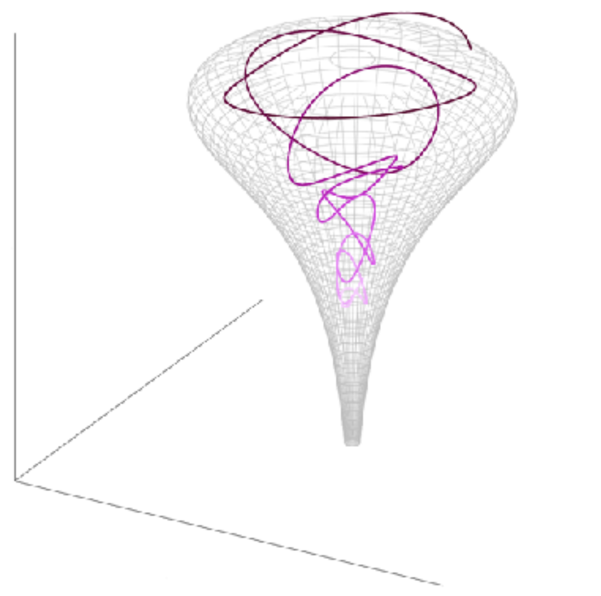
Towards the Ergodicity of Hamiltonian Monte Carlo

Michael Betancourt
University of Warwick

Simon Byrne
University College London

Hamiltonian Monte Carlo



Hamiltonian Monte Carlo uses techniques from differential geometry to efficiently explore probability distributions, admitting practical Bayesian inference that scales to the frontiers of applied statistics.

Here Riemannian Hamiltonian Monte Carlo generates a trajectory that sweeps through the probability mass of a high-dimensional hierarchical model.

Geometric Ergodicity

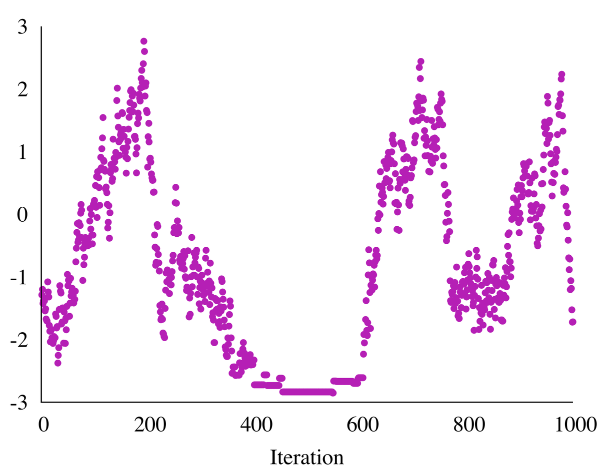
The *ergodicity* of a Markov chain ensures that it explores the entire target distribution as the number of transitions continues towards infinity. More critical to the empirical performance of a Markov chain, however, is *geometric ergodicity*, which ensures that the chain converges to the target distribution geometrically,

$$\| \mathcal{T}^n(q|q_0) - \varpi(q) \|_{\mathrm{TV}} \le R(q_0)\, \rho^n$$

An important consequence of geometric ergodicity is that Monte Carlo estimators from finite iterations are well-behaved, as formalized in the *Markov Chain Monte Carlo Central Limit Theorem*,

$$\frac{1}{N} \sum_{n=1}^{N} f(q_n) \to \mathcal{N}\left(\mathbb{E}_{\varpi}[f], \frac{\sigma^2}{N_{\mathrm{eff}}} \right).$$

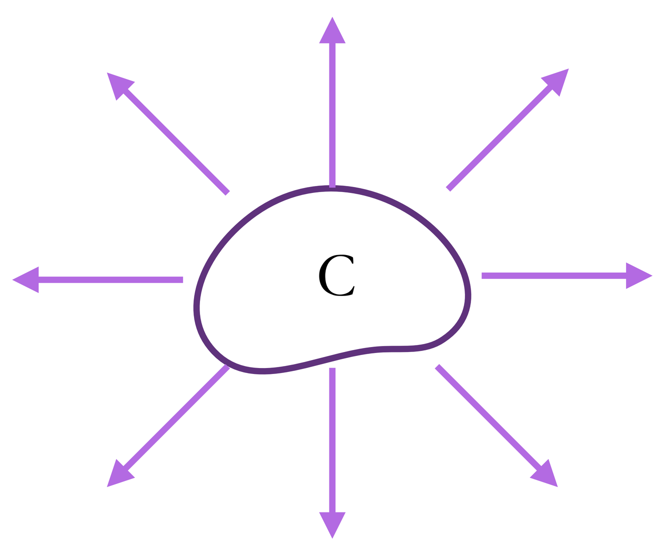
Geometric ergodicity is not just a theoretical concern. In practice, the lack of geometric ergodicity manifests as chains ignoring neighborhoods of high probability and "sticking" in corners of parameter space, both of which strongly bias subsequent Monte Carlo estimators.



In order to ensure that algorithms like Hamiltonian Monte Carlo yield robust inference we have to understand for which target distributions geometric ergodicity will hold, and how a loss of geometric ergodicity manifests in the output of the Markov chain. This latter consideration is crucial for constructing the practical diagnostics necessary for resilient statistical tools.

Verifying Geometric Ergodicity with a Drift Condition

One way of verifying that a Markov chain will be geometrically ergodic with respect to a given target distribution is to show that the chain tends to drift towards some small set, C. Verification proceeds in two stages.

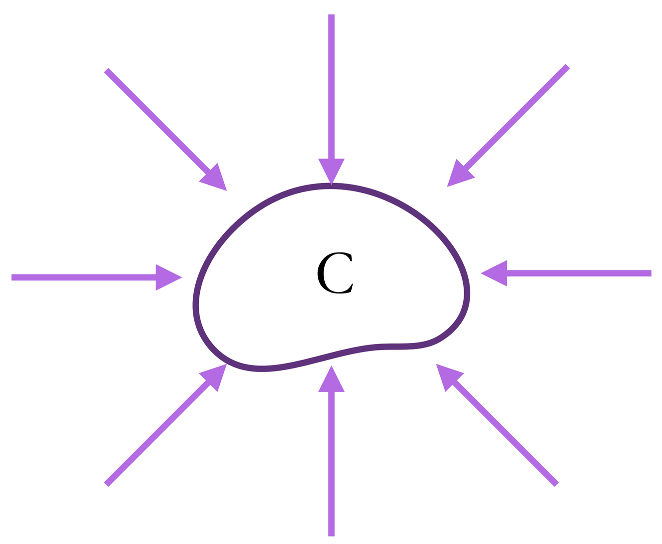


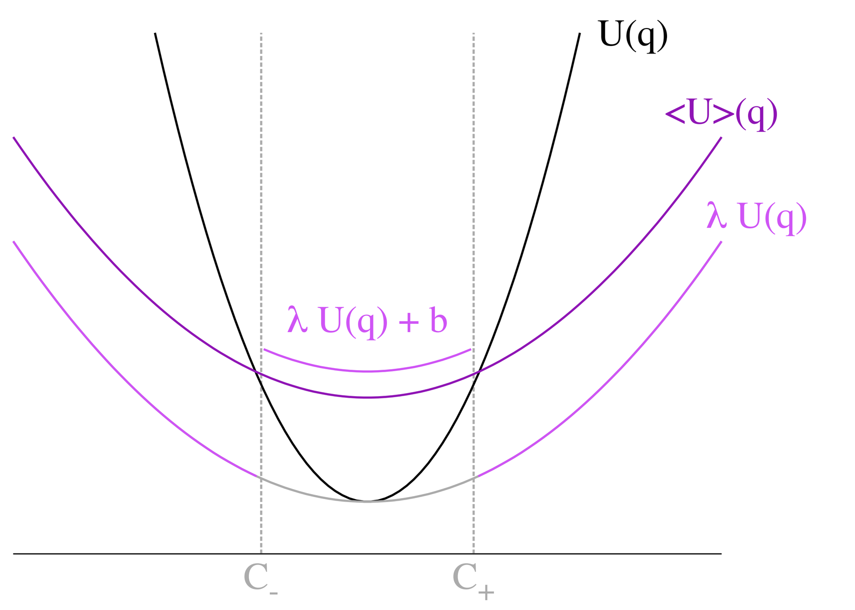
We first have to establish a *minorization condition* to ensure that once a chain reaches the small set it will explore the entire target distribution. Mathematically, we require that

$$\mathcal{T}(q|q_0) \ge \epsilon\, \nu(q)\,, \forall q_0 \in C.$$

Then we ensure that the chain will converge to the small set by establishing a *drift condition*,

$$\langle U \rangle(q) \le \lambda\, U(q) + b\, \mathbb{I}_C(q)\,.$$





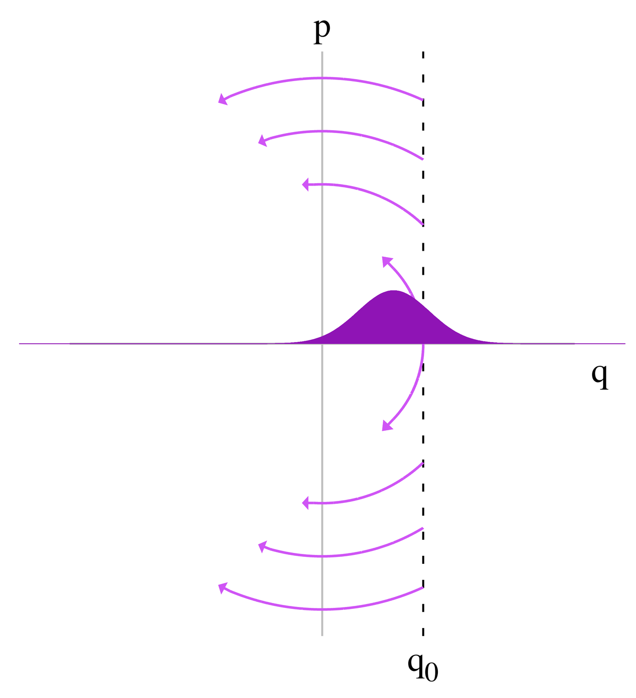
The drift condition requires a function that uniformly shrinks for any initial state outside of the small set, and uniformly shrinks given an arbitrary offset within the small set.

Together the minorization and drift conditions admit an explicitly geometric, if loose, bound on the convergence of the Markov chain,

$$\begin{aligned} &\| \mathcal{T}^n(q|q_0) - \varpi(q) \|_{\mathrm{TV}} \\ &\quad \le (1-\epsilon)^{rn} + f^n(\lambda, b, C)\, g^{rn}(\lambda, b, C) \left(1 + \frac{b}{1-\lambda} + U(q_0) \right). \end{aligned}$$

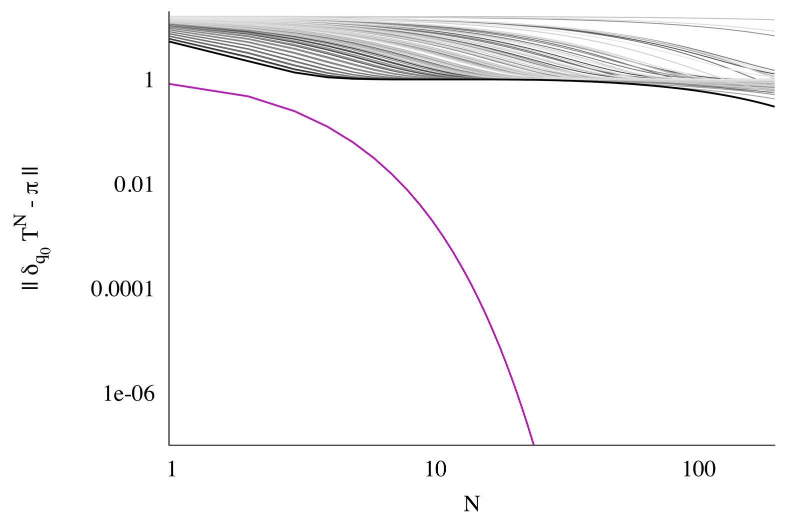
Hamiltonian Drift Conditions

Hamiltonian Monte Carlo generates a transition by sampling an auxiliary momentum, simulating a trajectory in (q, p) space, and then projecting back down to q. These trajectories naturally drift towards a small set around zero suggesting that any function monotonic in |q| serves as a drift function.

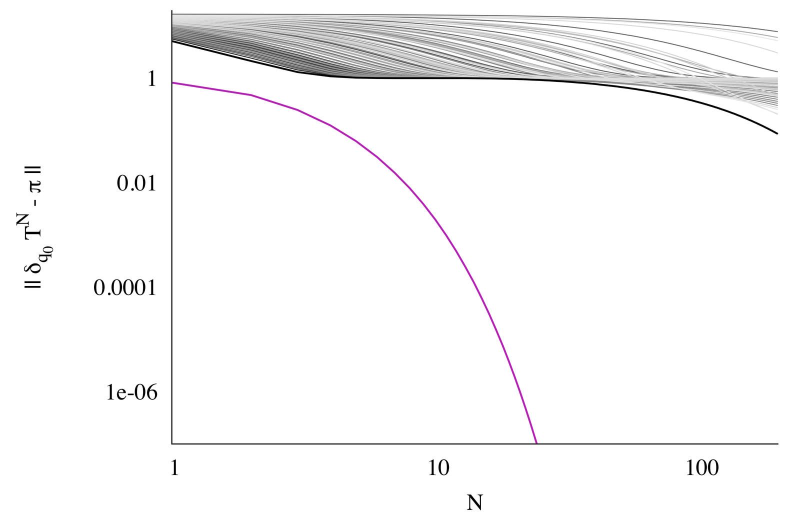


Here we consider the log density as a drift function for a univariate Gaussian target distribution with both an exact and numerical simulation of the Hamiltonian trajectory.

Gaussian with Exact Integrator



Gaussian with Numerical Integrator



Under certain stability conditions the numerical result is only a perturbation of the true result, and the failure of those stability conditions manifests in numerical divergences that can be monitored as a geometric ergodicity diagnostic!

Future Directions

Generalizing these results to a broader class of univariate target distributions and, ultimately, multivariate target distributions will require leveraging the underlying geometric structure of Hamiltonian Monte Carlo and interfacing with deep results from Hamiltonian chaos and dynamical systems.

One difficulty with drift conditions is that even if we can establish a drift condition the resulting bounds tend to be extremely weak, as evident even in the simple Gaussian case. Tighter bounds may be found by considering Hamiltonian Monte Carlo as a second-order stochastic process and using related techniques to constrain the relaxation towards the mass of the target distribution.