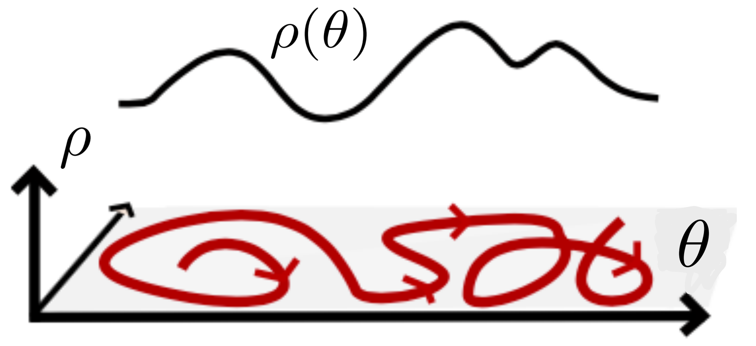
Magnetic Hamiltonian Monte Carlo

Nilesh Tripuraneni[1], Mark Rowland[2], Zoubin Ghahramani[23], Richard E. Turner[2]

[1]UC Berkeley, [2]University of Cambridge, [3]Uber AI Labs

Markov Chain Monte Carlo (MCMC)

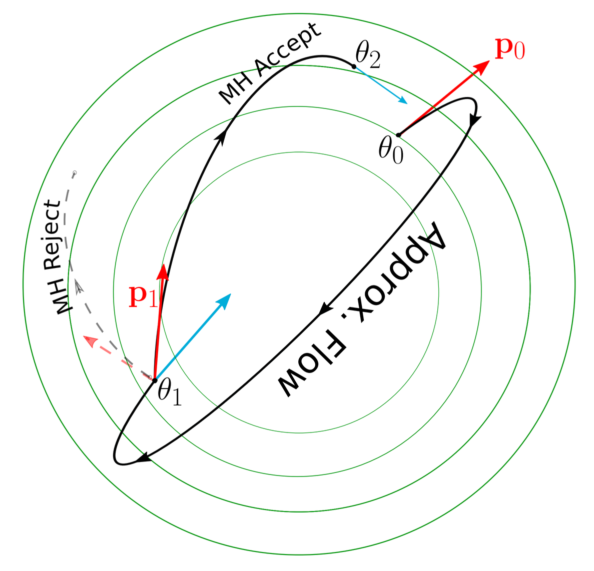
- Probabilistic inference in complex models requires *high-dimensional* integrals.
- The Metropolis-Hastings (MH) algorithm is an MCMC algorithm that constructs a Markov chain $(\Theta_n)_{n\in\mathbb{N}}$ yielding *asymptotically exact* but *correlated* samples from an unnormalized density $\rho(\theta)$.



Markov Chain Trajectory

- A *good* MH algorithm should have low inter-sample correlation while maintaining a high acceptance ratio.

Hamiltonian Monte Carlo (HMC)



HMC's exploration of an isotropic Gaussian target. Proposals inspired by Hamiltonian flows can span long distances in parameter space encouraging efficient exploration.

- $\rho(\theta)$ is augmented with independent, Gaussian "momentum" variables $\mathbf{p}$:

$$\rho(\theta, \mathbf{p}) \propto e^{-U(\theta) - \mathbf{p}^\top \mathbf{p}/2} \equiv e^{-H(\theta, \mathbf{p})}. \tag{1}$$

- From $(\theta_n, \mathbf{p}_n)$:
  - Resample momentum $\mathbf{p}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
  - Simulate the Hamiltonian flow over joint $(\theta, \mathbf{p})$ space defined by:

$$\frac{\mathrm{d}}{\mathrm{dt}}\begin{bmatrix}\theta(t)\\ \mathbf{p}(t)\end{bmatrix} = \underbrace{\begin{bmatrix}\mathbf{0} & \mathbf{I}\\ -\mathbf{I} & \mathbf{0}\end{bmatrix}}_{\mathbf{A}}\begin{bmatrix}\nabla_\theta H(\mathbf{p}(t), \theta(t))\\ \nabla_{\mathbf{p}} H(\mathbf{p}(t), \theta(t))\end{bmatrix} \equiv \begin{bmatrix}\mathbf{p}(t)\\ -\nabla_\theta U(\theta(t))\end{bmatrix}.$$

  with a $\approx$ *energy-conserving*, *symplectic*, *time-reversible* leapfrog integrator with L steps of size ϵ to give $(\Phi_{\epsilon,H})^L(\theta_n, \mathbf{p}_n) = (\theta'_n, \mathbf{p}'_n)$.
  - Flip the momentum component $(\theta', \mathbf{p}'') = (\theta'_n, -\mathbf{p}'_n)$ and apply a MH accept/reject step with probability $\min(1, \exp(H(\theta_n, \mathbf{p}_n) - H(\theta', \mathbf{p}'')))$.

Non-Canonical Hamiltonian Dynamics

The map $\Phi^{\mathbf{A}}_{\tau,H}(\theta, \mathbf{p})$ defined by integrating:

$$\frac{\mathrm{d}}{\mathrm{dt}}\begin{bmatrix}\theta(t)\\ \mathbf{p}(t)\end{bmatrix} = \underbrace{\begin{bmatrix}\mathbf{E} & \mathbf{F}\\ -\mathbf{F}^\top & \mathbf{G}\end{bmatrix}}_{\mathbf{A}}\begin{bmatrix}\nabla_\theta H(\mathbf{p}(t), \theta(t))\\ \nabla_{\mathbf{p}} H(\mathbf{p}(t), \theta(t))\end{bmatrix} \tag{2}$$

where $\mathbf{A} \in \mathcal{M}_{2n\times 2n}$ is *any* invertible, antisymmetric matrix induces a flow that is:

- *energy-conserving*: $\partial_\tau H(\Phi^{\mathbf{A}}_{\tau,H}(\theta, \mathbf{p})) = 0$.
- *symplectic*: $[\nabla_{\theta,\mathbf{p}}\Phi^{\mathbf{A}}_{\tau,H}(\theta, \mathbf{p})]^\top \mathbf{A}^{-1}[\nabla_{\theta,\mathbf{p}}\Phi^{\mathbf{A}}_{\tau,H}(\theta, \mathbf{p})] = \mathbf{A}^{-1}$ ($\implies$ *volume-preservation*).
- *(pseudo)-time-reversible*: If $(\theta(t), \mathbf{p}(t))$ is a solution to (2), then $(\widetilde{\theta}(t), \widetilde{\mathbf{p}}(t)) = (\theta(-t), -\mathbf{p}(-t))$ is a solution to the modified non-canonical dynamics:

$$\frac{\mathrm{d}}{\mathrm{dt}}\begin{bmatrix}\widetilde{\theta}(t)\\ \widetilde{\mathbf{p}}(t)\end{bmatrix} = \underbrace{\begin{bmatrix}-\mathbf{E} & \mathbf{F}\\ -\mathbf{F}^\top & -\mathbf{G}\end{bmatrix}}_{\mathbf{A}}\begin{bmatrix}\nabla_{\widetilde{\theta}} H(\widetilde{\theta}(t), \mathbf{p}(t))\\ \nabla_{\widetilde{\mathbf{p}}} H(\widetilde{\theta}(t), \widetilde{\mathbf{p}}(t))\end{bmatrix} \tag{3}$$

if $H(\theta, \mathbf{p}) = H(\theta, -\mathbf{p})$.

The Physics

Non-canonical dynamics with matrix

$$\mathbf{A} = \begin{bmatrix}\mathbf{0} & \mathbf{M}^{1/2}\\ -(\mathbf{M}^{1/2})^\top & \mathbf{0}\end{bmatrix}$$

$=$

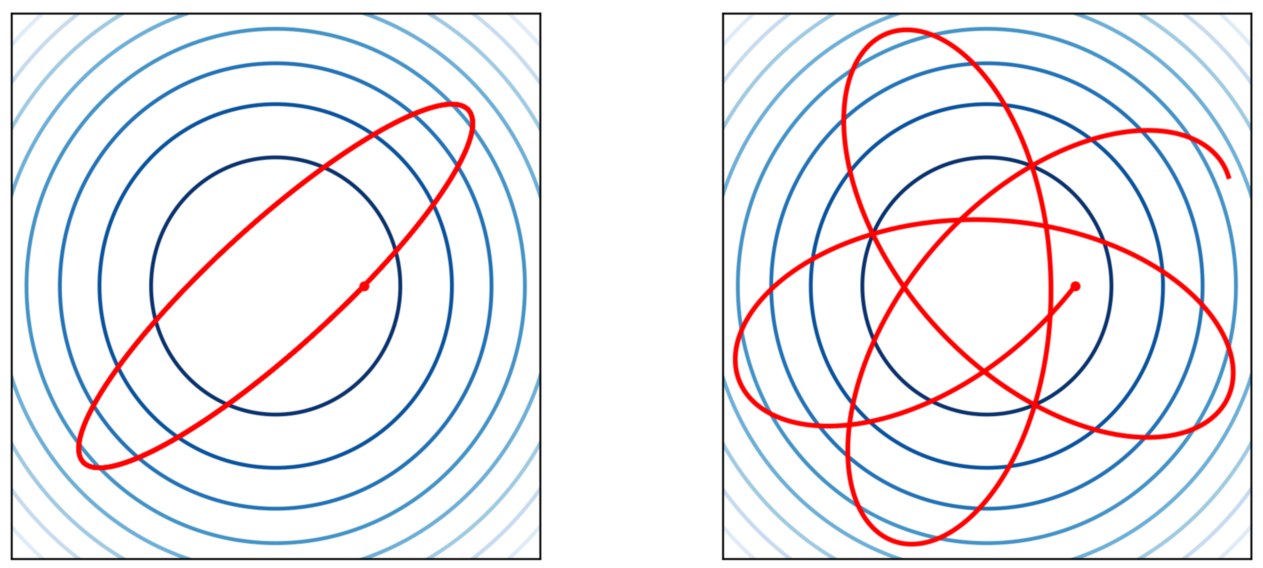
- "Preconditioned" HMC with Hamiltonian $H(\theta, \mathbf{p}) = \frac{1}{2}\mathbf{p}^\top \mathbf{M}^{-1}\mathbf{p} + U(\theta)$
- Achieved by redefining Hamiltonian and using "standard" $\mathbf{A}$ ☺

Non-canonical dynamics with matrix

$$\mathbf{A} = \begin{pmatrix}\mathbf{0} & \mathbf{I}\\ -\mathbf{I} & \mathbf{G}\end{pmatrix}$$

$=$

- Mechanics of charged particle coupled to magnetic field (in 3-dimensions).
- Cannot be achieved by defining *any* smooth Hamiltonian with "standard" $\mathbf{A}$ ☺



Example sample paths for standard HMC (left) and MHMC (right) for an isotropic Gaussian target distribution.

Symplectic Integration for MHMC

With the symmetric, leapfrog splitting:

$$H(\theta, \mathbf{p}) = \underbrace{U(\theta)/2}_{H_1(\theta)} + \underbrace{\mathbf{p}^\top \mathbf{p}/2}_{H_2(\mathbf{p})} + \underbrace{U(\theta)/2}_{H_1(\theta)} \tag{4}$$

the corresponding non-canonical dynamics for the sub-Hamiltonians $H_1(\theta)$ and $H_2(\mathbf{p})$ are:

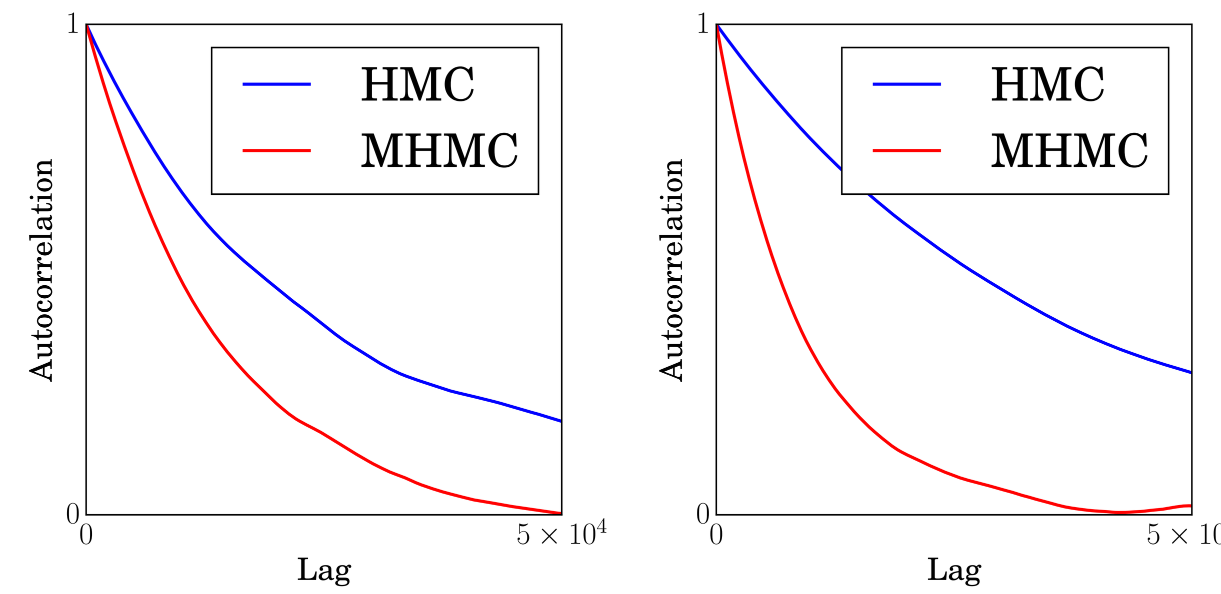
$$\frac{\mathrm{d}}{\mathrm{dt}}\begin{bmatrix}\theta\\ \mathbf{p}\end{bmatrix} = \underbrace{\begin{bmatrix}\mathbf{0} & \mathbf{I}\\ -\mathbf{I} & \mathbf{G}\end{bmatrix}}_{\mathbf{A}}\begin{bmatrix}\nabla_\theta U(\theta)/2\\ \mathbf{0}\end{bmatrix} = \begin{bmatrix}0\\ -\nabla_\theta U(\theta)/2\end{bmatrix}; \quad \frac{\mathrm{d}}{\mathrm{dt}}\begin{bmatrix}\theta\\ \mathbf{p}\end{bmatrix} = \underbrace{\begin{bmatrix}\mathbf{0} & \mathbf{I}\\ -\mathbf{I} & \mathbf{G}\end{bmatrix}}_{\mathbf{A}}\begin{bmatrix}\mathbf{0}\\ \mathbf{p}\end{bmatrix} = \begin{bmatrix}\mathbf{p}\\ \mathbf{G}\mathbf{p}\end{bmatrix}.$$

Algorithm 1 Magnetic HMC (MHMC)

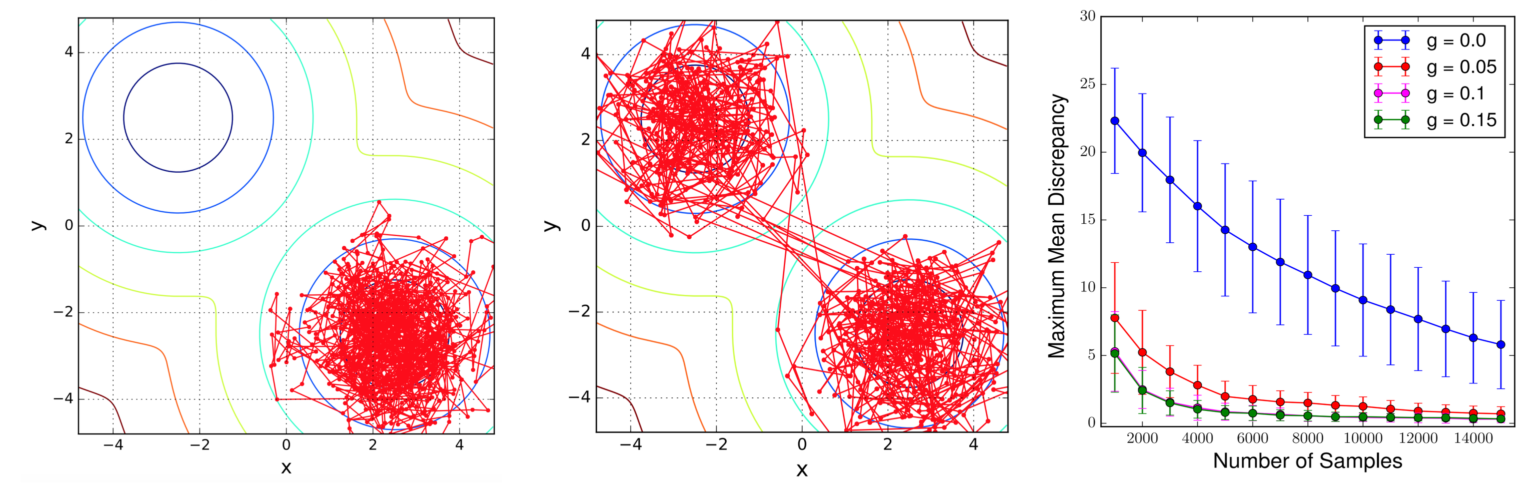
Input: H, $\mathbf{G}$, L, ϵ
  Initialize $(\theta_0, \mathbf{p}_0)$, and set $\mathbf{G}_0 \leftarrow \mathbf{G}$
  for $n = 1, \ldots, N$ **do**
    Resample $\mathbf{p}_{n-1} \sim N(\mathbf{0}, \mathbf{I})$
    Set $(\widetilde{\theta}_n, \widetilde{\mathbf{p}}_n) \leftarrow \mathrm{LF}(H, L, \epsilon, (\theta_{n-1}, \mathbf{p}_{n-1}, \mathbf{G}_{n-1}))$
    Flip momentum $(\widetilde{\theta}_n, \widetilde{\mathbf{p}}_n) \leftarrow (\widetilde{\theta}_n, -\widetilde{\mathbf{p}}_n)$ and set $\widetilde{\mathbf{G}}_n \leftarrow -\mathbf{G}_{n-1}$
    if $\mathrm{Unif}([0,1]) < \min(1, \exp(H(\theta_{n-1}, \mathbf{p}_{n-1}) - H(\widetilde{\theta}_n, \widetilde{\mathbf{p}}_n)))$ **then**
      Set $(\theta_n, \mathbf{p}_n, \mathbf{G}_n) \leftarrow (\widetilde{\theta}_n, \widetilde{\mathbf{p}}_n, \widetilde{\mathbf{G}}_n)$
    else
      Set $(\theta_n, \mathbf{p}_n, \mathbf{G}_n) \leftarrow (\theta_{n-1}, \mathbf{p}_{n-1}, \mathbf{G}_{n-1})$
    end if
    Flip momentum $\mathbf{p}_n \leftarrow -\mathbf{p}_n$ and flip $\mathbf{G}_n \leftarrow -\mathbf{G}_n$
  end for
Output: $(\theta_n)_{n=0}^N$

- "sub"-dynamics are *linear* $\implies$ analytic solutions ☺
- "sub"-dynamics are
  - *symplectic*
  - *(pseudo)-time-reversible*
  - $\approx$ *energy-conserving* with error $\mathcal{O}(\epsilon^3)$ **just like standard HMC.**

Experiments



Averaged Autocorrelation of HMC vs MHMC on a 10D ill-conditioned Gaussian (left) and Averaged Autocorrelation of HMC vs MHMC on a 2D ill-conditioned Gaussian.



Left: 500 samples from HMC; Middle: 500 samples from MHMC; Right: MMD between HMC/MHMC samples for various magnitudes of the non-zero component of the magnetic field – denoted g. Note $g = 0$ corresponds to standard HMC.