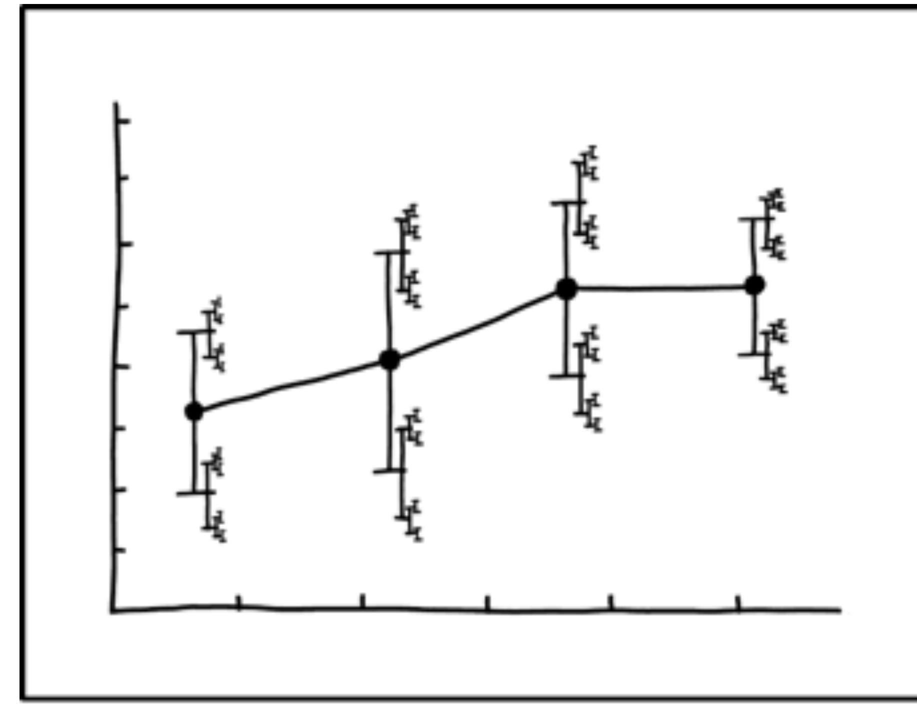
Additional Advanced Monte Carlo slides

Stopping criteria

- Many diagnostic exist
  - All have limitations
  - Some are dubious



I DON'T KNOW HOW TO PROPAGATE ERROR CORRECTLY, SO I JUST PUT ERROR BARS ON ALL MY ERROR BARS.

- Best approach is CLT (with same caveats as IS): for a 95% confidence interval, use

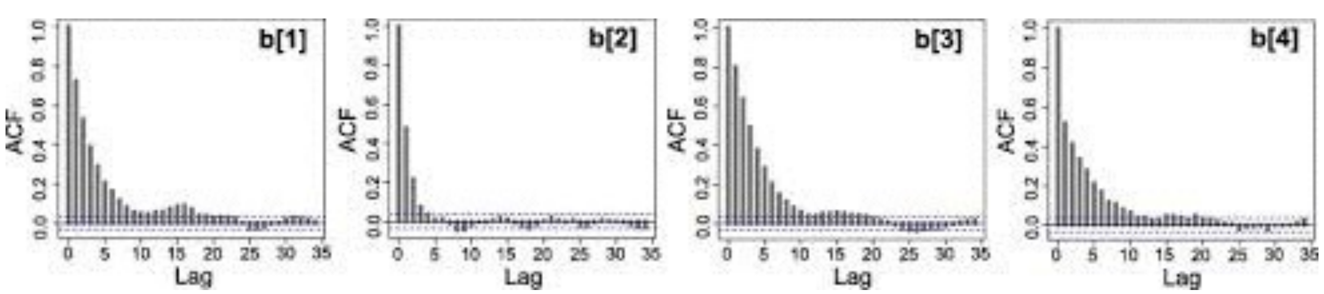
$$I_n \pm 1.96\sqrt{\sigma^2_{\text{asympt}}/n}$$

- The asymptotic variance is:

$$\sigma^2(\phi) = \mathbb{V}_\pi[\phi(X_1)] + 2\sum_{k=2}^{\infty} \mathbb{C}ov_\pi[\phi(X_1), \phi(X_k)].$$

Estimation of the asymptotic variance

- Direct method: estimate the auto-correlations (ACF)



  - Can be done quickly with FFT
  - But estimator has infinite variance! Need to truncate. Still unstable in many practical scenarios.

ESS for MC

- Idea is similar as for IS, but still tied to a test function:

$$\frac{\mathrm{ESS(N)}}{N} \to \frac{\sigma^2_{\mathrm{iid}}}{\sigma^2_{\mathrm{asymptotic}}}$$

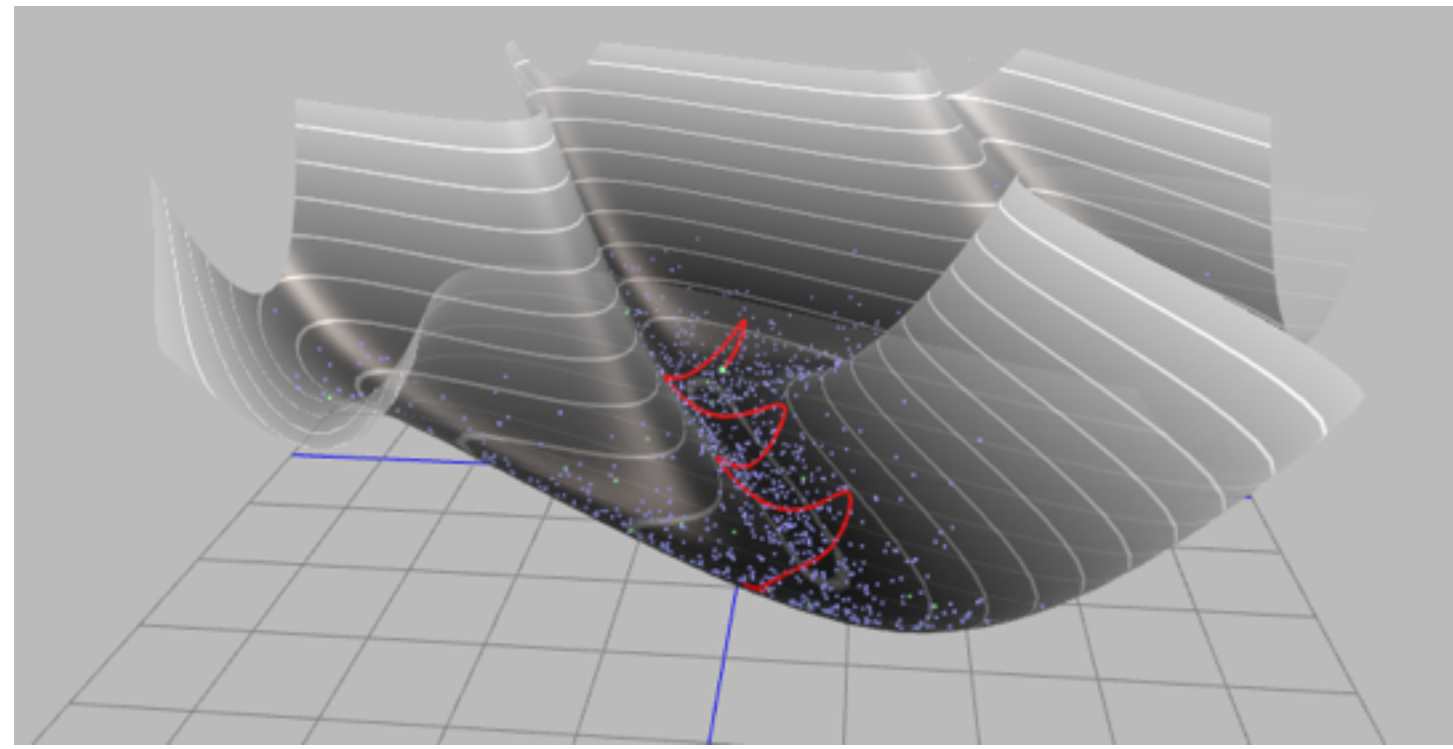
- Can estimate using ACF as in last slide
- Better method: batch estimators.
  - Segment the MCMC trace into chunks of length $\sqrt{n}$
  - Assume sampler is good enough so that behaviour across blocks is nearly iid
- Standard metric in MCMC literature to compare samplers: ESS per second or ESS per operation

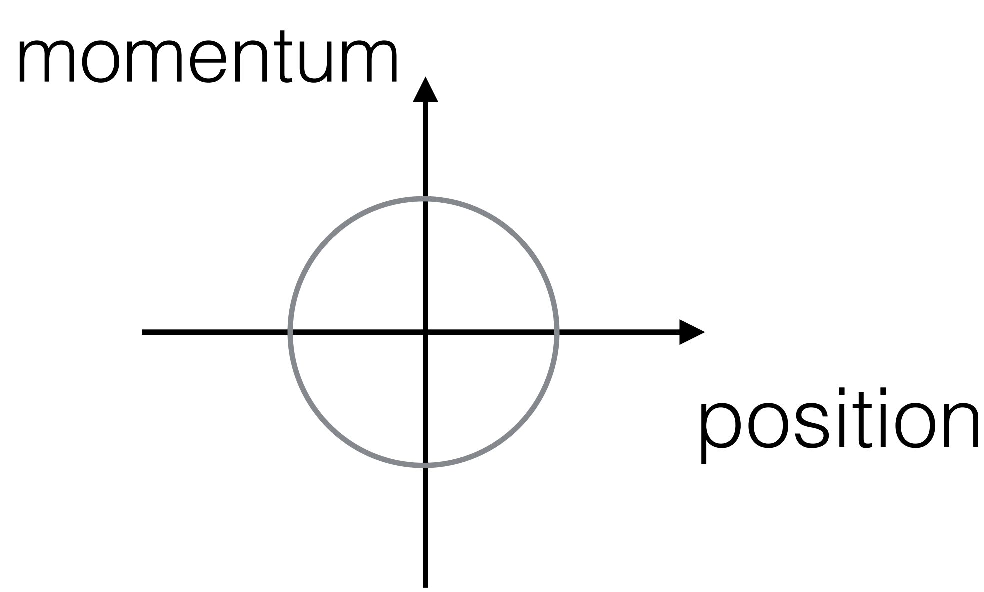
Asymptotic variance and ESS for MC

- References:
  - Honest exploration of intractable probability distributions (2001). Jones and Hobert.
  - Monte Carlo standard errors for MCMC (2008). Flegal.
- Multivariate confidence version:
  - Multivariate output analysis for Markov chain Monte Carlo (2015). Vats et al.

Hamiltonian Monte Carlo: intuition

- Physical ball rolling on the *energy*
  - U(x) = -log(p(x))
- Motion described by the *Hamiltonian flow*
- Phase space on a Gaussian target:





HMC: auxiliary variables

- Physics' notation: z = (q, p)
  - position q
  - Augment the state with a momentum random variable p
- Put an auxiliary distribution on p, with f(p) = exp(-K(p)) and s.t. K(p) = K(-p), e.g. normal.

$$H(q,p) = U(q) + K(p), \quad U(q) = q^2/2, \quad K(p) = p^2/2$$

- Can think of p as a velocity (when the mass matrix, i.e. covariance of f(x) is identity).
- Statistical notation would be then z = (x, v)

Exact HMC

- MCMC kernel is a non-reversible
- Given by a Dirac delta: k(z, dz’) = $\delta_{\Phi(z)}$(dz’)
- Φ is the Hamiltonian flow, i.e. solutions of the differential equations

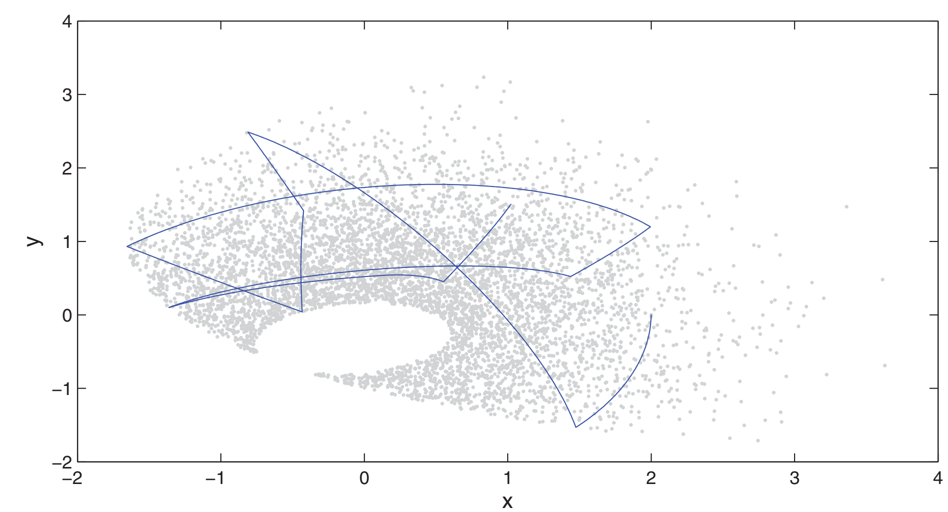
$$\begin{aligned} \frac{dq_i}{dt} &= \frac{\partial H}{\partial p_i} \\ \frac{dp_i}{dt} &= -\frac{\partial H}{\partial q_i} \end{aligned} \qquad \Longrightarrow \qquad \begin{aligned} \frac{dq_i}{dt} &= [M^{-1}p]_i \\ \frac{dp_i}{dt} &= -\frac{\partial U}{\partial q_i} \end{aligned}$$

- Exact HMC: Analytic solution only in special cases, e.g. for (truncated) normal target we get:

$$q(t) = r\cos(a+t), \quad p(t) = -r\sin(a+t)$$

Application: truncated normal distributions

- See Pakman and Paninski (2014)
- Truncated normal arise in many practical contexts:
  - Probit and tobit models
  - Bayesian splines for positive functions
  - Bayesian lasso



$$
\begin{aligned}
y_i &= \operatorname{sign}(w_i) \\
w_i &= -\mathbf{z}_i \cdot \boldsymbol{\beta} + \varepsilon_i \\
\varepsilon_i &\sim \mathcal{N}(0, 1)
\end{aligned}
$$

Exact HMC: invariance

- MCMC kernel is a non-reversible
- Given by a Dirac delta: $k(z, dz') = \delta_{\Phi(z)}(dz')$
- Invariance equivalent to:
  - given $Z \sim$ extended target π'
    $\pi'(x, v) = \pi(x) \times \text{normal}(v)$
  - Define $Y = \Phi(Z)$
  - Do we have $Y \sim \pi'$?

Exact HMC: invariance

- By change of variable formula, break into two factors:

$$f_Y(y) = f_Z(\Phi^{-1}(y))\,|\det J_{\Phi^{-1}}(y)|$$

hence ingredient to show Y ~ π' are:

- Φ invertible (yes, set v $\longleftarrow$ -v)
- *Conservation of Hamiltonian*: first factor is constant
- *Volume preservation*: second factor is constant

Conservation of Hamiltonian

- Want f(z) = f(Φ(z))
- Enough: no infinitesimal Hamiltonian changes, H' = 0
- Use total derivative identity

$$\frac{dH}{dt} = \sum_{i=1}^{d}\left[\frac{dq_i}{dt}\frac{\partial H}{\partial q_i} + \frac{dp_i}{dt}\frac{\partial H}{\partial p_i}\right]$$

- Then substitute our choice of the differential equation:

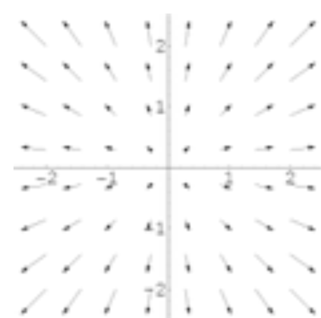
$$\begin{aligned}\frac{dq_i}{dt} &= \frac{\partial H}{\partial p_i}\\ \frac{dp_i}{dt} &= -\frac{\partial H}{\partial q_i}\end{aligned} \implies \sum_{i=1}^{d}\left[\frac{dq_i}{dt}\frac{\partial H}{\partial q_i} + \frac{dp_i}{dt}\frac{\partial H}{\partial p_i}\right] = \sum_{i=1}^{d}\left[\frac{\partial H}{\partial p_i}\frac{\partial H}{\partial q_i} - \frac{\partial H}{\partial q_i}\frac{\partial H}{\partial p_i}\right] = 0$$

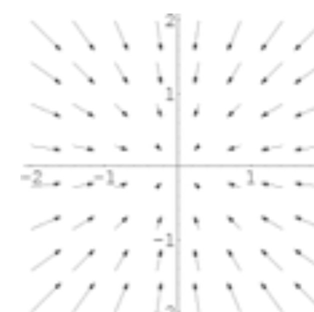
Volume preservation

The preservation of volume by Hamiltonian dynamics can be proved in several ways. One is to note that the divergence of the vector field defined by equations (2.1) and (2.2) is zero, which can be seen as follows:

$$\sum_{i=1}^{d}\left[\frac{\partial}{\partial q_i}\frac{dq_i}{dt}+\frac{\partial}{\partial p_i}\frac{dp_i}{dt}\right] = \sum_{i=1}^{d}\left[\frac{\partial}{\partial q_i}\frac{\partial H}{\partial p_i}-\frac{\partial}{\partial p_i}\frac{\partial H}{\partial q_i}\right] = \sum_{i=1}^{d}\left[\frac{\partial^2 H}{\partial q_i \partial p_i}-\frac{\partial^2 H}{\partial p_i \partial q_i}\right] = 0 \quad (2.13)$$

A vector field with zero divergence can be shown to preserve volume (Arnold, 1989).

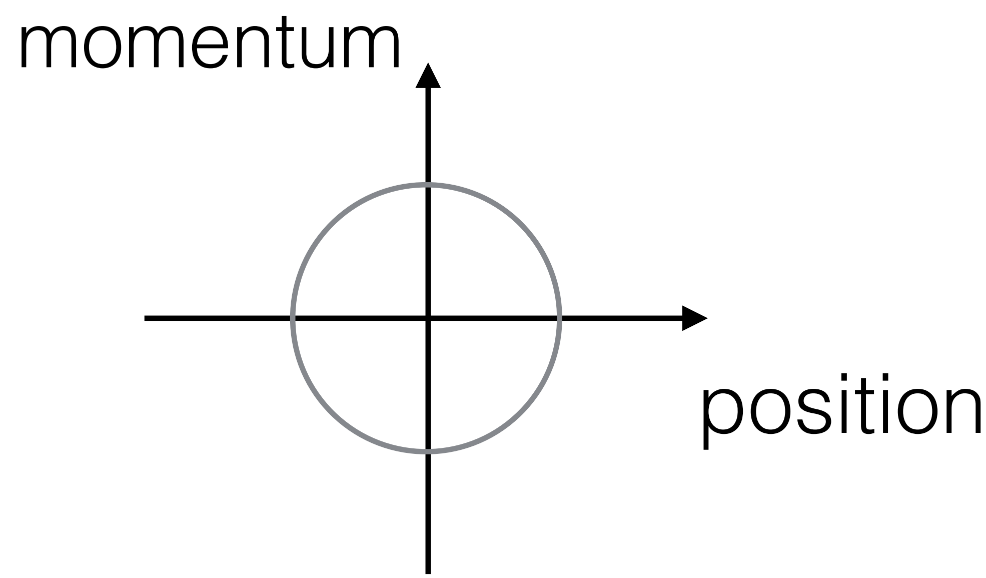




- See Neal (2012). MCMC using Hamiltonian dynamics for another, more direct argument

Exact HMC: irreducibility

- Easy to see non irreducible in phase space

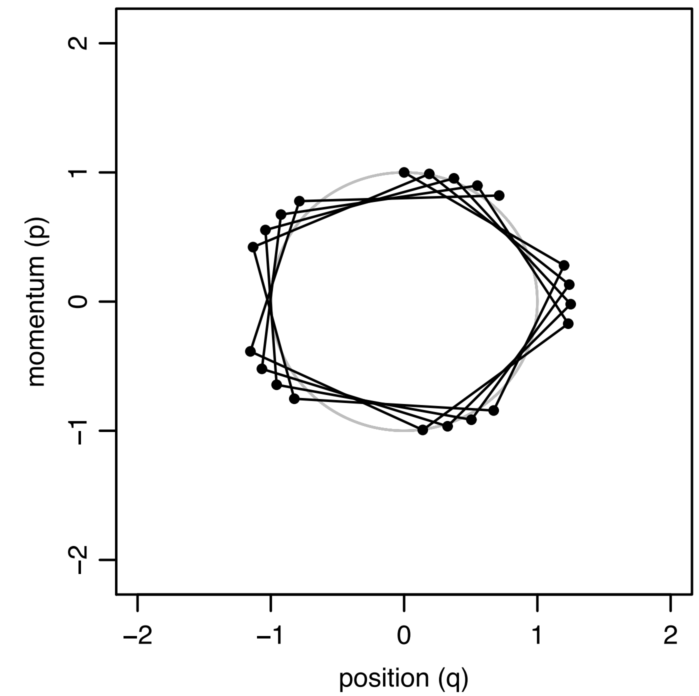


- Solution: refresh momentum

Symplectic HMC

- We can't simulate the exact Hamiltonian flow for most targets of interest.
- Idea:
  - solve the differential equation using numerical methods and initial condition given by current point
    - can be done so that volume still preserved (e.g. with leap-frog integrator)
  - Hamiltonian no longer exactly preserved, so use MH to accept-reject

Symplectic HMC



- Numerical solution example:
- Algorithm: numerically follow the evolution of diff. equation
  - Simplest version: Euler method

$$\frac{dp_i}{dt} = -\frac{\partial U}{\partial q_i}$$

$$\frac{dq_i}{dt} = [M^{-1}p]_i$$

$$p_i(t+\varepsilon) = p_i(t) + \varepsilon \frac{dp_i}{dt}(t) = p_i(t) - \varepsilon \frac{\partial U}{\partial q_i}(q(t))$$

$$q_i(t+\varepsilon) = q_i(t) + \varepsilon \frac{dq_i}{dt}(t) = q_i(t) + \varepsilon \frac{p_i(t)}{m_i}$$

- Need something better: leap-frog integrator (will see why soon when going over invariance)

Rough idea

- Use accept-reject
- Proposal: deterministic, given by numerical solution of DE followed for a fixed number of steps
- Accept-reject to take into account numerical error
- Why is this not quite correct?

Important, overlooked condition on proposal *q*

- Mutual absolute continuity condition:

$$\int_A \pi(\mathrm{d}x) q(x, B) > 0 \Leftrightarrow \int_B \pi(\mathrm{d}x) q(x, A) > 0$$

- For example, in a discrete state space where the target has full support, this means:

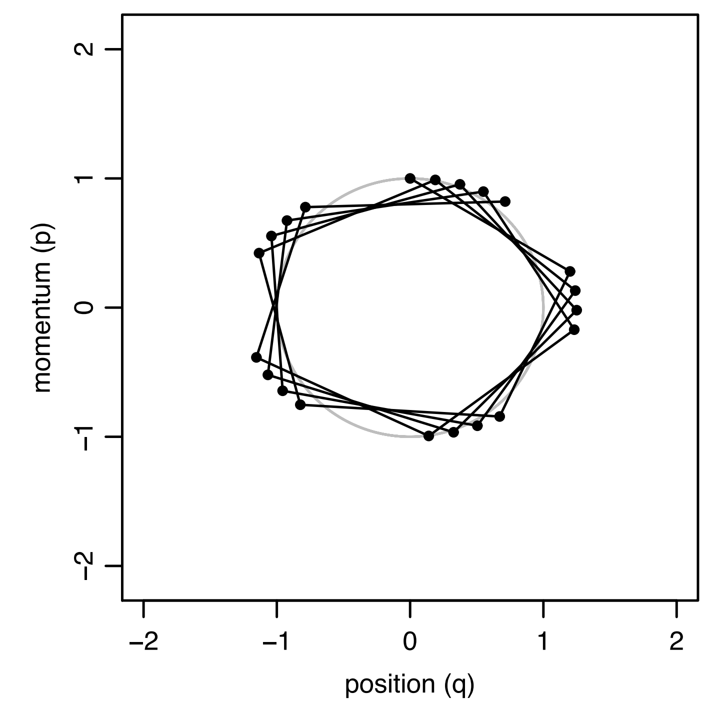
$$q(x, y) > 0 \Leftrightarrow q(y, x) > 0$$

- This can be tricky in combinatorial spaces (more on that soon)

Symplectic HMC

- 3 moves, which have to be deterministically cycled in the following order
  - Φ: an MH move with proposal given by:
    - follow the ~~exact~~ discretized trajectory
    - flip the momentum, R(q,p) = R(q, -p)
  - Flip again!
  - Momentum refreshment
- What properties do we need for invariance?

Symplectic HMC



- Numerical solution example:
- Algorithm: numerically follow the evolution of diff. equation
  - Replace Euler by leaf-frog

$$\frac{dp_i}{dt} = -\frac{\partial U}{\partial q_i}$$

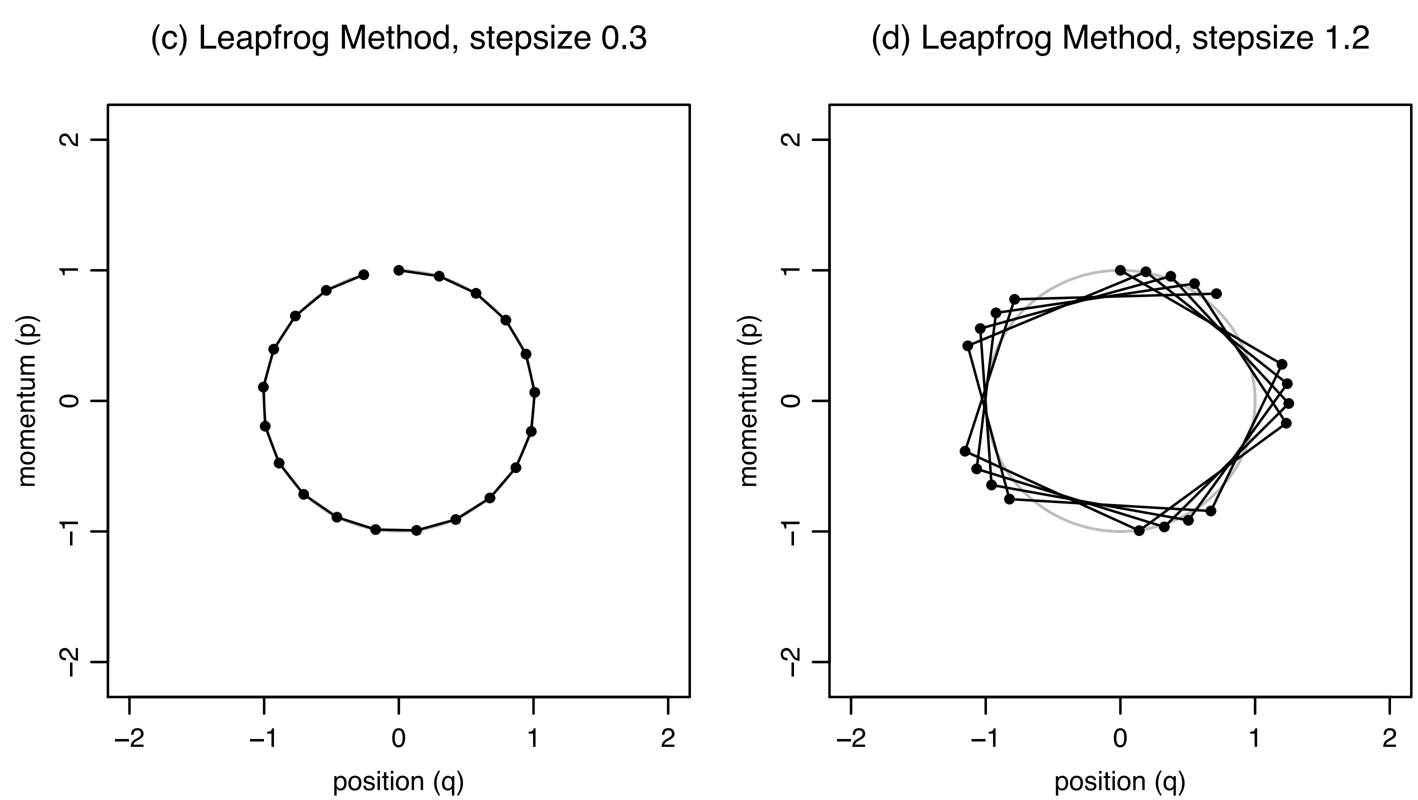
$$\frac{dq_i}{dt} = [M^{-1}p]_i$$

$$
\begin{aligned}
p_i(t+\varepsilon/2) &= p_i(t) - (\varepsilon/2)\,\frac{\partial U}{\partial q_i}(q(t)) \\
q_i(t+\varepsilon) &= q_i(t) + \varepsilon\,\frac{p_i(t+\varepsilon/2)}{m_i} \\
p_i(t+\varepsilon) &= p_i(t+\varepsilon/2) - (\varepsilon/2)\,\frac{\partial U}{\partial q_i}(q(t+\varepsilon))
\end{aligned}
$$

- Properties: let R(q,p) = (q,-p) (flip)
  - involution: R(Φ(R(Φ(z)))) = z
  - hence, volume preservation

Practical considerations

- Two critical parameters to tune:
  - L: number of leap-frog steps
  - epsilon: step size
- For L: Hoffman 2011, Sohl-Dickstein 2016
- For epsilon: mostly heuristics/adaptation



Special case: Metropolis-Adjusted Langevin (MALA)

- Use one leap frog step, and use the following order for the kernels
  - Refresh velocity first
  - Then do one leap frog, which simplifies into:

$$
\begin{aligned}
q_i^* &= q_i - \frac{\varepsilon^2}{2}\frac{\partial U}{\partial q_i}(q) + \varepsilon p_i \\
p_i^* &= p_i - \frac{\varepsilon}{2}\frac{\partial U}{\partial q_i}(q) - \frac{\varepsilon}{2}\frac{\partial U}{\partial q_i}(q^*)
\end{aligned}
$$

Dimensionality scaling

running time = number of samples needed to get a tolerance (with probability 95%) x compute cost per sample

	number of samples needed to get a tolerance (with probability 95%)	compute cost per sample
HMC	$d^{1/4}$	d^1
MALA	$d^{1/3}$	d^1
Random walk MH	d^1	d^1
Local BPS/Gibbs in weakly dep. sparse field	$\boldsymbol{d^1}$	***constant***