MCMC Using Hamiltonian Dynamics

Presenter: Kairong Jiang

10/16/19

MCMC Using Hamiltonian Dynamics

- Hamiltonian Dynamics
  - Equations
  - Properties
  - Discrete integration — the leapfrog method
- HMM Algorithm
- Discussions on HMM — advantages, tuning, dimensionality, optimal acceptance rate

Hamiltonian Dynamics

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

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}$$

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

Hamiltonian Dynamics

- Reversibility
  - The mapping from t to t+s is one-to-one
  - Relation to detailed-balance
- Conservation of the Hamiltonian
- Volume preservation
  - We don’t need to compute the determinant of the Jacobian for the mapping

Hamiltonian Dynamics

- Discrete integration   Assuming $K(p) = \sum_{i=1}^{d} \frac{p_i^2}{2m_i}$
  - Euler's method

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

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

  - Modified Euler's method

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

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

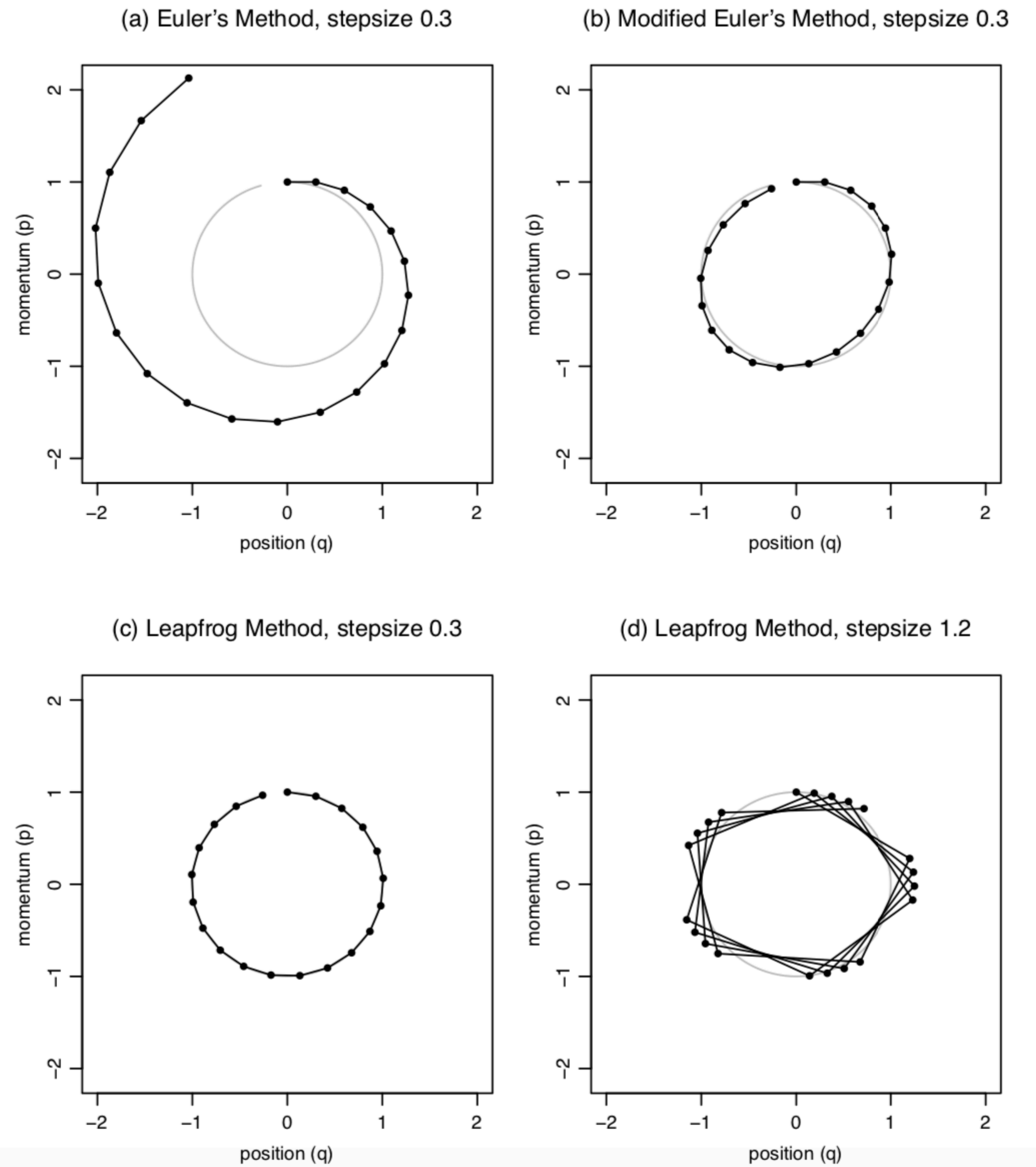
  - The leapfrog method

$$p_i(t+\varepsilon/2) = p_i(t) - (\varepsilon/2) \frac{\partial U}{\partial q_i}(q(t))$$

$$q_i(t+\varepsilon) = q_i(t) + \varepsilon \frac{p_i(t+\varepsilon/2)}{m_i}$$

$$p_i(t+\varepsilon) = p_i(t+\varepsilon/2) - (\varepsilon/2) \frac{\partial U}{\partial q_i}(q(t+\varepsilon))$$

Hamiltonian Dynamics



HMC: Canonical distributions

Canonical distribution:

$$P(x) = \frac{1}{Z}\exp(-E(x)/T)$$

Plug in Hamiltonian:

$$P(q,p) = \frac{1}{Z}\exp(-H(q,p)/T)$$

$$P(q,p) = \frac{1}{Z}\exp(-U(q)/T)\exp(-K(p)/T)$$

Define potential with regard to the distribution we care:

$$U(q) = -\log\Big[\pi(q)L(q|D)\Big]$$

Define momentum (Kinetic energy) to be convenient:

$$K(p) = \sum_{i=1}^{d}\frac{p_i^2}{2m_i}$$

HMC: two steps

- Step 1: sample momentum from independently from the current position values.

- Step 2: Repeatedly perform Metropolis updates using Hamiltonian dynamics to propose new state. (negate momentum at last step to make the update symmetric.)

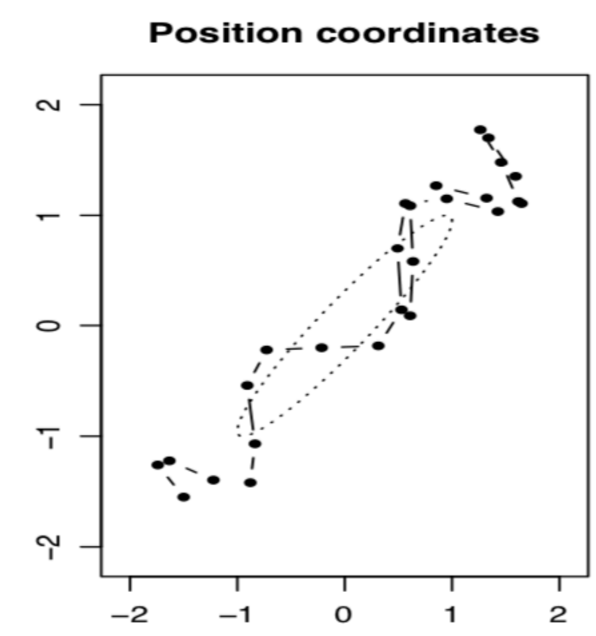
  - i.e. use the leapfrog method to move the state for various time steps.

  - Accept propability:

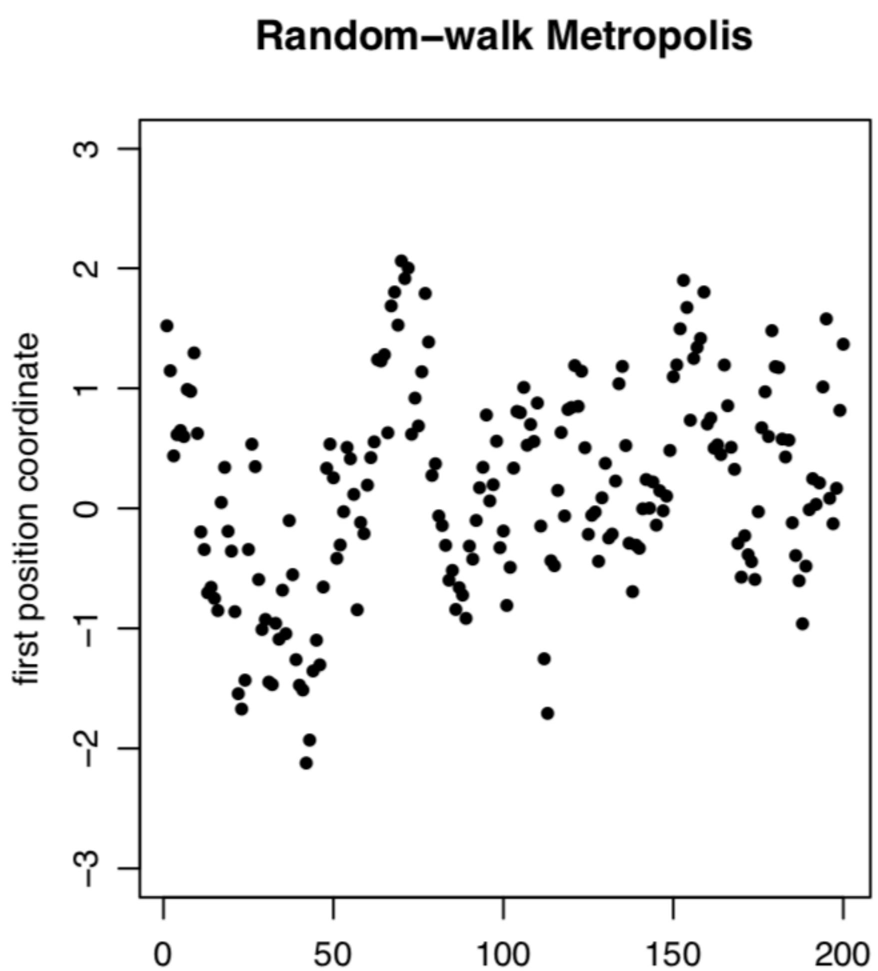
$$\min\left[1,\, \exp(-H(q^*, p^*) + H(q, p))\right] \;=\; \min\left[1,\, \exp(-U(q^*) + U(q) - K(p^*) + K(p))\right]$$

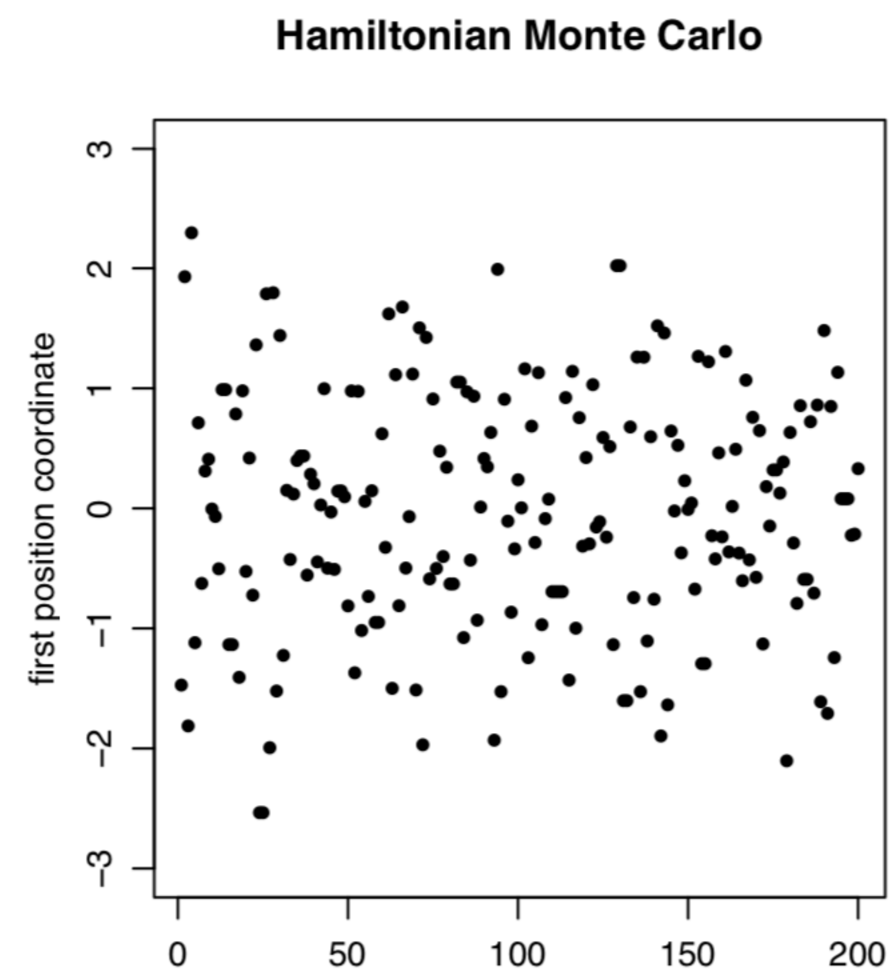
HMC: proof to work

- Detailed balance
- Ergodicity
    - Step 1 sampling can dramatically move the point.
    - It might fail if step 2 moves the point for a full cycle.

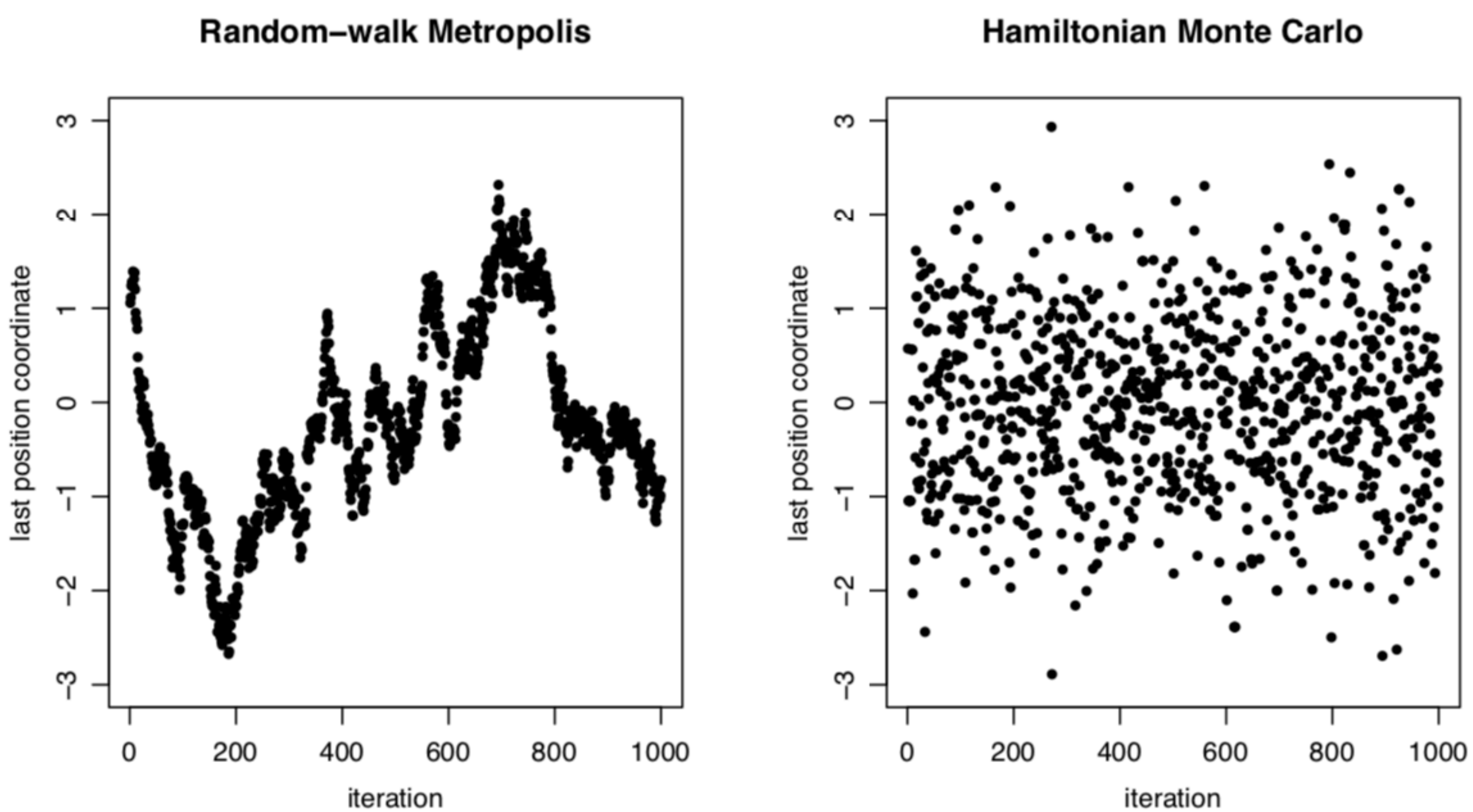


Benefits of HMC

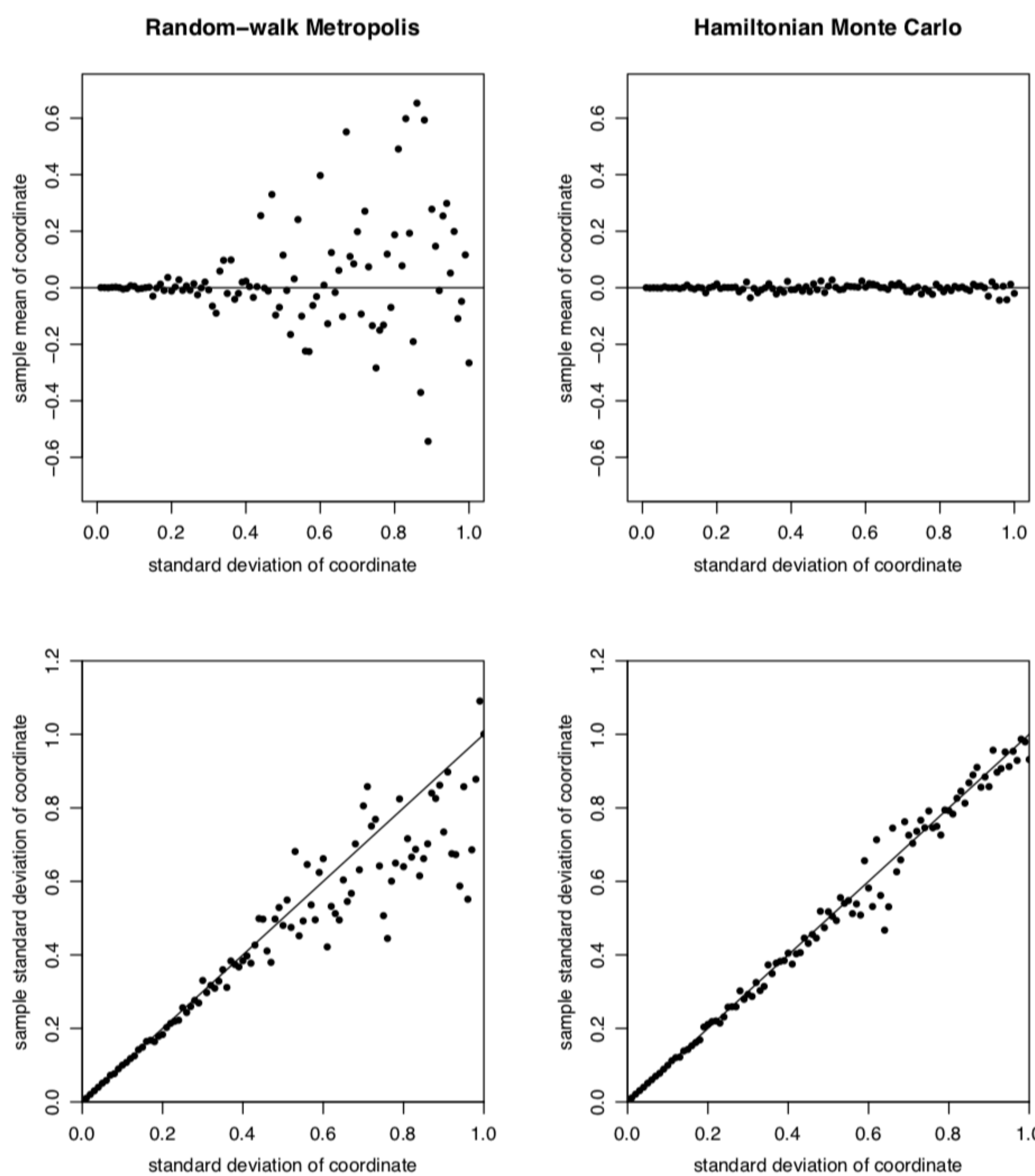




Benefits of HMC



Benefits of HMC



Practical Discussions

- Linear Transformations
  - Apply linear transformation to the variables
  - Useful to transform the covariance to nearly identity

Practical Discussions

- Tuning HMC
  - Stepsize
    - Too large — not stable
    - Too small — waste computation, and potentially cause random walk
  - Trajectory length
- Both can benefit from randomly pick from a range

Practical Discussions

- Varying step size for individual dimension
- Combining HMC with other MCMC updates
  - Fixed hyperparameters for low-level parameter updates
- Optimal acceptance rate: 65%