Examination of the likelihood function by simulation of a fictional Hamiltonian system

Joshua N. Pritikin

Department of Psychology
University of Virginia

16 October 2014

Likelihood

The likelihood of parameter vector θ given observed data x and some model is defined as

$$L(\theta) \equiv \Pr(x|\theta).$$

Recall Bayes' theorem,

$$\Pr(\theta|x) = \frac{\Pr(\theta)\Pr(x|\theta)}{\Pr(x)}.$$

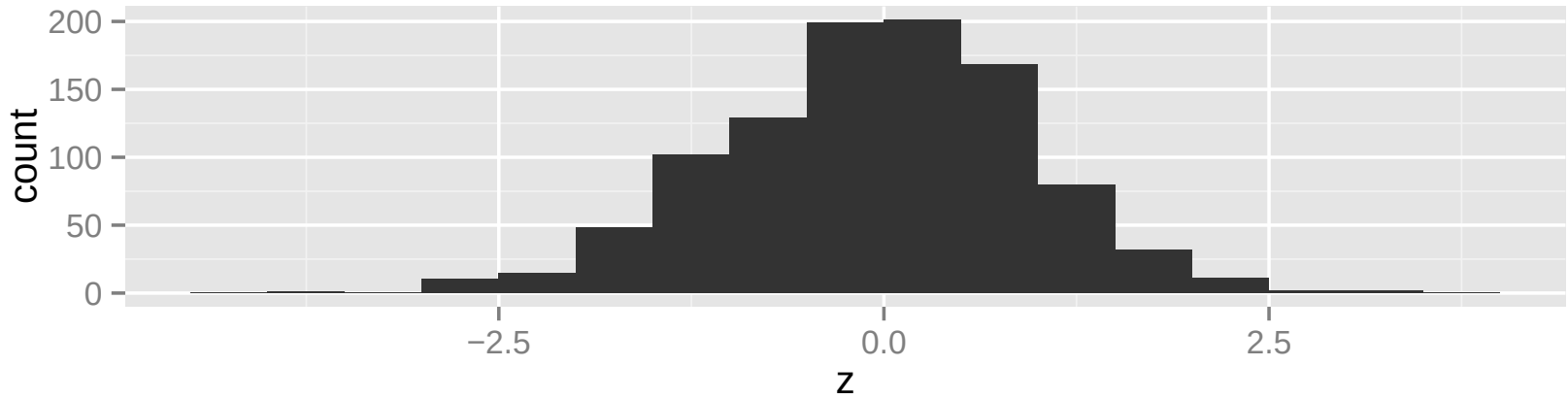
With a uniform prior $\Pr(\theta)$, the likelihood is the posterior density (up to a constant),

$$\Pr(\theta|x) \propto L(\theta)$$

Sampling, example

quantile	0.500	0.250	0.750	0.100	0.900	0.999
Z score	0.000	-0.674	0.674	-1.282	1.282	3.090

```
df <- data.frame(z=qnorm(runif(1000)))
ggplot(df, aes(z)) + geom_histogram(binwidth=.5)
```



MCMC methods

Popular Markov chain Monte Carlo algorithms include

- Metropolis-Hastings
- Gibbs
- Hamiltonian Monte Carlo

Acknowledgment

- Timo, Cynthia (Xin)
- UVa grad students
- colleagues who I forgot to mention

Metropolis-Hastings

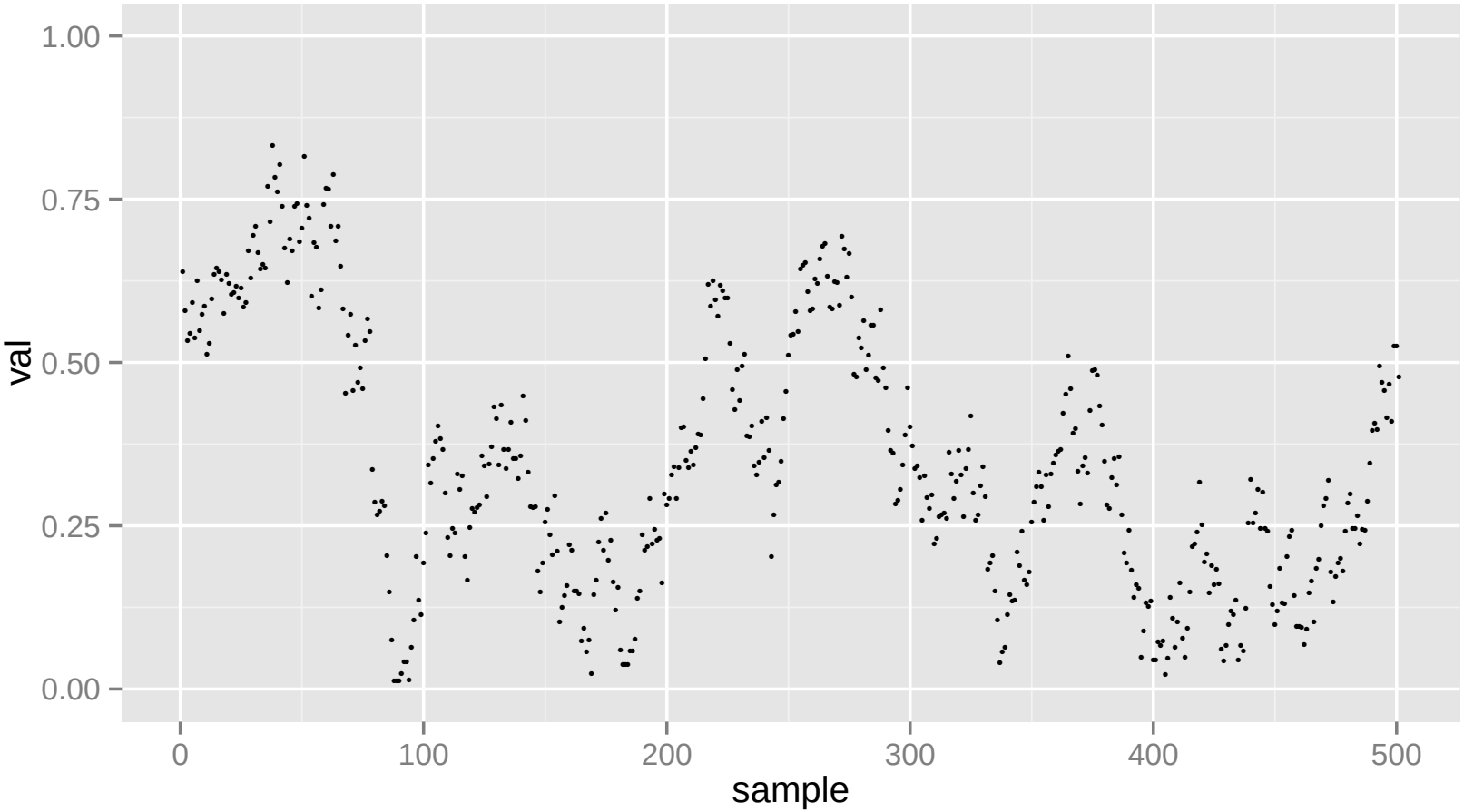
Choose an arbitrary point $y_{t=0}$ for the first sample and an arbitrary conditional distribution $Q(y_{t+1}|y_t)$.

Select the next candidate point $y' \sim Q(y_t)$ and compute the acceptance ratio,

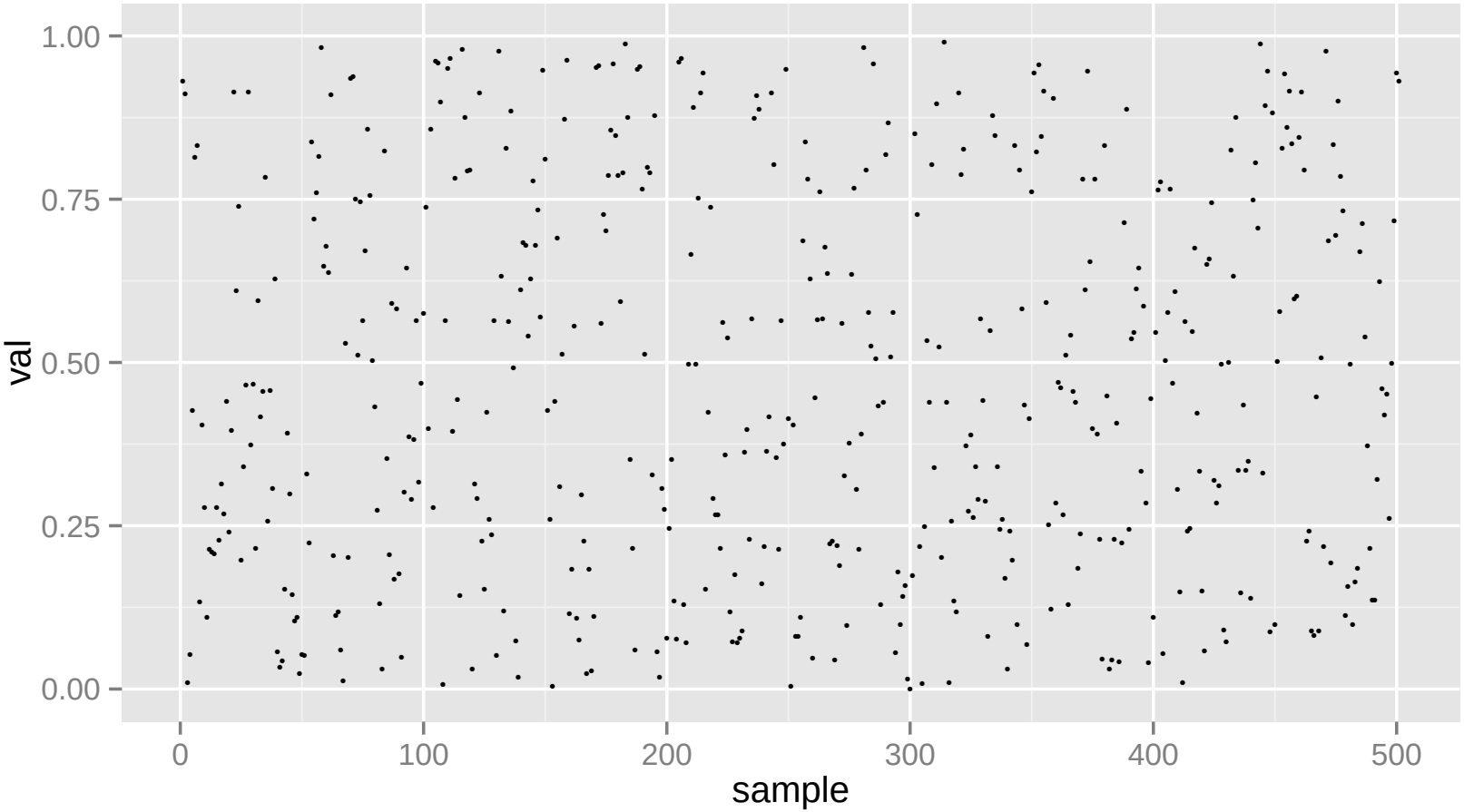
$$\alpha = \frac{L(y')}{L(y_t)}.$$

If $\alpha \geq 1$ then set $y_{t+1} = y'$. Otherwise accept the candidate y' with probability α. If the candidate is rejected, set $y_{t+1} = y_t$.

M-H sampling of a 1-dimensional uniform



M-H sampling of a 1-dimensional uniform



Markov Chain

A memoryless Markov Chain can be represented as a transition matrix,

	s_1	s_2	s_3
s_1	0.17	0.56	0.28
s_2	0.05	0.51	0.44
s_3	0.23	0.60	0.17

Markov Chain

To construct a Markov Chain that samples from a unique stationary distribution π, two conditions are required:

1. Ergodicity
2. Detailed balance (a.k.a. reversibility) means that

$$\pi_i \Pr(i \to j) = \pi_j \Pr(j \to i)$$

$$\ldots \text{ or alternately that}$$

$$\frac{\pi_i}{\pi_j} = \frac{\Pr(j \to i)}{\Pr(i \to j)}$$

where π_i and π_j are equalibrium probabilities of being in states i and j, respectively.

Gibbs sampling

Without loss of generality, suppose we have a model with 3 parameters, $L(\beta_1, \beta_2, \beta_3)$.

Given $Y^t \equiv (\beta_1^t, \beta_2^t, \beta_3^t)$ at time t, we want to draw

$$Y^{(t+1)} \sim L(Y^t).$$

If simple conditional distributions are available, we can draw samples parameter-wise,

$$\begin{aligned}
\beta_1^{(t+1)} &\sim L(\beta_1 | \beta_2^t, \beta_3^t) \\
\beta_2^{(t+1)} &\sim L(\beta_2 | \beta_1^{(t+1)}, \beta_3^t) \\
\beta_3^{(t+1)} &\sim L(\beta_3 | \beta_1^{(t+1)}, \beta_2^{(t+1)})
\end{aligned}$$

Hamiltonian Monte Carlo (HMC)

Let q and p be d-dimensional position and momentum vectors, respectively. The system $H(q,p)$ is known as a *Hamiltonian*. The laws of motion are,

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = \frac{\partial H}{\partial p_i}$$
$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -\frac{\partial H}{\partial q_i}$$

for $i \in \{1 \dots d\}$.

$$U(q) = \text{potential energy}$$
$$K(p) = \text{kinetic energy}$$
$$H(q,p) = U(q) + K(p)$$

HMC, kinetic energy

Choose

$$K(p) = \frac{p^T M^{-1} p}{2}.$$

M is the mass of the puck.

HMC, nice properties

- Hamiltonian dynamics is reversible
- Energy is conserved

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \sum_{i=1}^{d} \left[\frac{\mathrm{d}q_i}{\mathrm{d}t} \frac{\partial H}{\partial q_i} + \frac{\mathrm{d}p_i}{\mathrm{d}t} \frac{\partial H}{\partial p_i} \right] = 0$$

- Volume is preserved (due to sympelecticness)

HMC, discrete simulation

For simplicity, assume M is diagonal with elements $m_1, \ldots, m_d$.

- Euler's method

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

- Euler's method (improved)

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

HMC, discrete simulation

- The leapfrog method

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

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

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

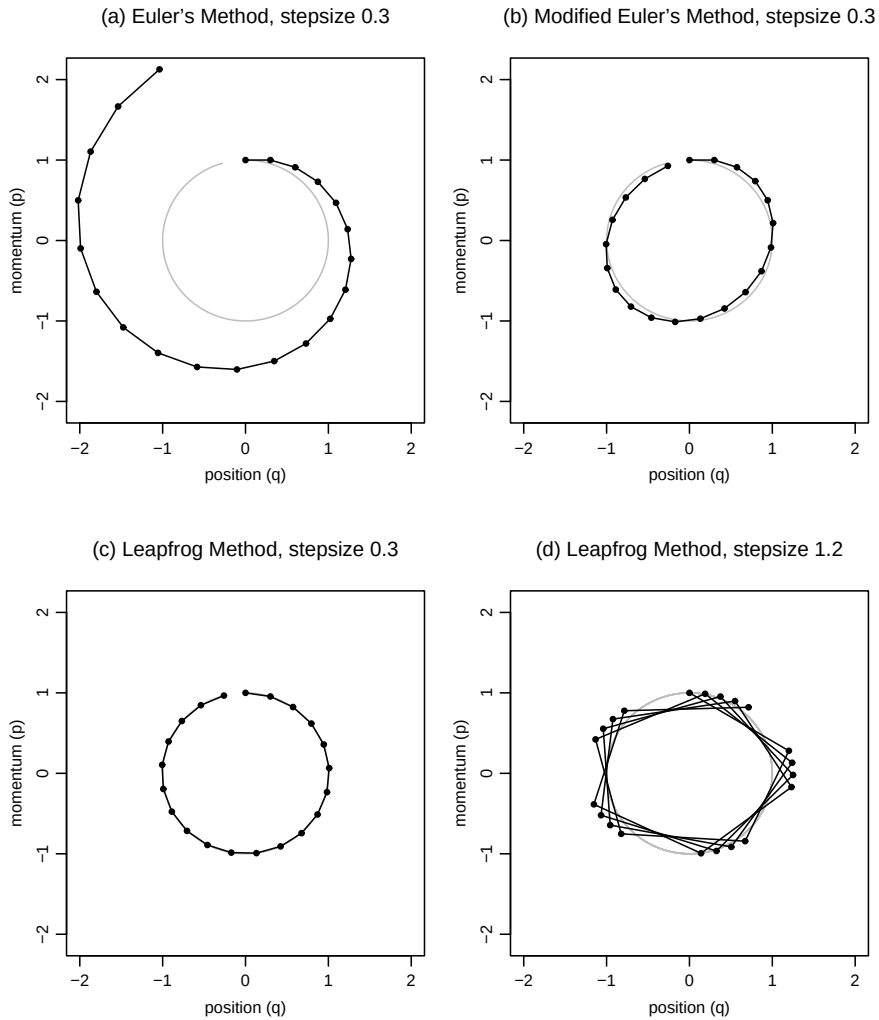


Figure 1: Results using three methods for approximating Hamiltonian dynamics, when $H(q,p) = q^2/2 + p^2/2$. The initial state was $q = 0$, $p = 1$. The stepsize was $\varepsilon = 0.3$ for (a), (b), and (c), and $\varepsilon = 1.2$ for (d). Twenty steps of the simulated trajectory are shown for each method, along with the true trajectory (in gray).

Construction of HMC

$$H(q,p) = \{U(q), K(P)\}$$

$$U(q) = -\log\left[\Pr(\theta)L(\theta)\right]$$

$$K(p) = \frac{p^T M^{-1} p}{2}$$

1. Draw a new momentum
2. Simulate Hamiltonian dynamics for S steps with step size ϵ. Uniformly sample a candidate from the newly simulated states. Accept the candidate with probability,

$$\min\left[1, \exp\left(-H(q',p') + H(q,p)\right)\right]$$

Does HMC satisfy M-H conditions?

$$H(q,p) = \{U(q), K(P)\}$$

$$U(q) = -\log\left[\Pr(\theta)L(\theta)\right]$$

$$K(p) = \frac{p^T M^{-1} p}{2}$$

1. Detailed balance (a.k.a. reversibility)
2. Ergodicity

Sounds cool, but does it work?

- 100-dimensional multivariate Gaussian distribution
- all dimensions independent
- standard deviations of 0.01, 0.02,. . . , 0.99, 1.00
- HMC S parameter set to 150
- For each HMC iteration, ϵ was drawn uniformly from $0.013 \pm 20\%$

Reject rate was 0.13 for HMC and 0.75 for random-walk Metropolis.

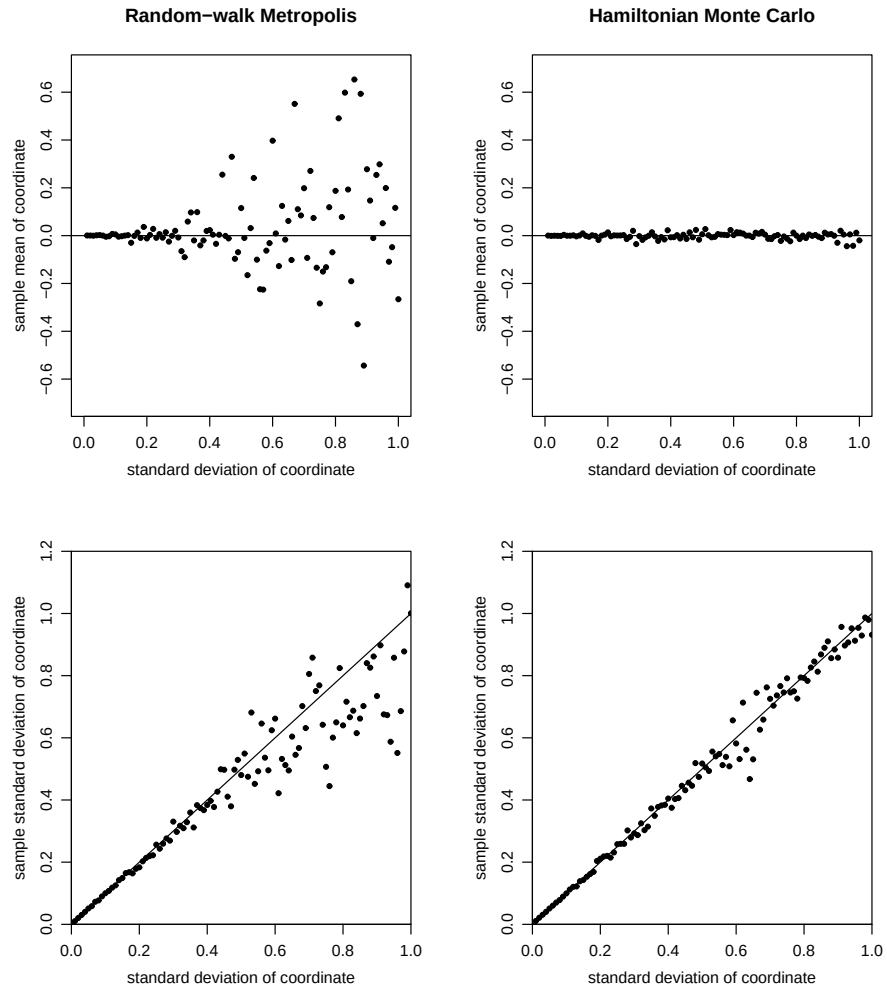


Figure 7: Estimates of means (top) and standard deviations (bottom) for the 100-dimensional example, using random-walk Metropolis (left) and HMC (right). The 100 variables are labelled on the horizontal axes by the true standard deviaton of that variable. Estimates are on the vertical axes.

No-U-Turn sampler

Problem: For HMC, we need to select the number of leapfrog steps S and step size ϵ.

Solution: NUTS (Hoffman & Gelman, 2011)

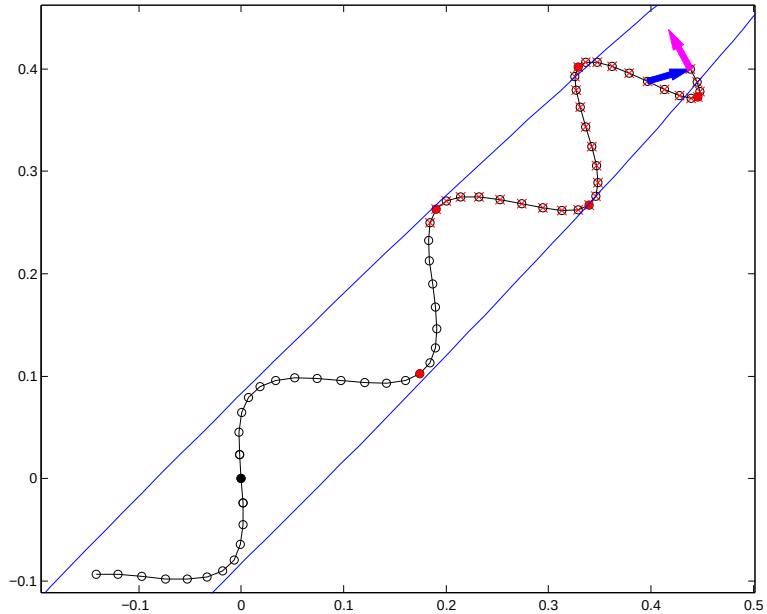


Figure 2: Example of a trajectory generated during one iteration of NUTS. The blue ellipse is a contour of the target distribution, the black open circles are the positions θ traced out by the leapfrog integrator and associated with elements of the set of visited states $\mathcal{B}$, the black solid circle is the starting position, the red solid circles are positions associated with states that must be excluded from the set $\mathcal{C}$ of possible next samples because their joint probability is below the slice variable u, and the positions with a red "x" through them correspond to states that must be excluded from $\mathcal{C}$ to satisfy detailed balance. The blue arrow is the vector from the positions associated with the leftmost to the rightmost leaf nodes in the rightmost height-3 subtree, and the magenta arrow is the (normalized) momentum vector at the final state in the trajectory. The doubling process stops here, since the blue and magenta arrows make an angle of more than 90 degrees. The crossed-out nodes with a red "x" are in the right half-tree, and must be ignored when choosing the next sample.

being more complicated, the analogous algorithm that eliminates the slice variable seems empirically to be slightly less efficient than the algorithm presented in this paper.

Dealing with curvature

Consider the 2-dimensional *banana* density on manifold R,

$$V(R) = \frac{1}{2}\left[\frac{q_1^2(R)}{\sigma_1^2} + \frac{(q_2(R) + \beta q_1^2(R) - 100\beta)^2}{\sigma_2^2}\right]$$

with $\beta = 0.03$, $\sigma_1 = 0.01$, and $\sigma_2 = 1$ (Betancourt, 2013).

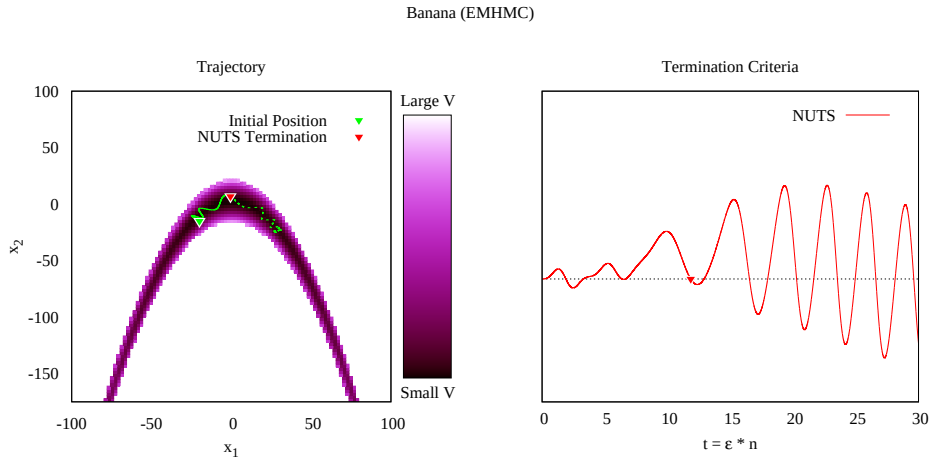


FIG. 4. On more complicated potentials, such as that arising from the banana distribution (4), the NUTS criterion terminates prematurely even when ignoring the transient behavior.

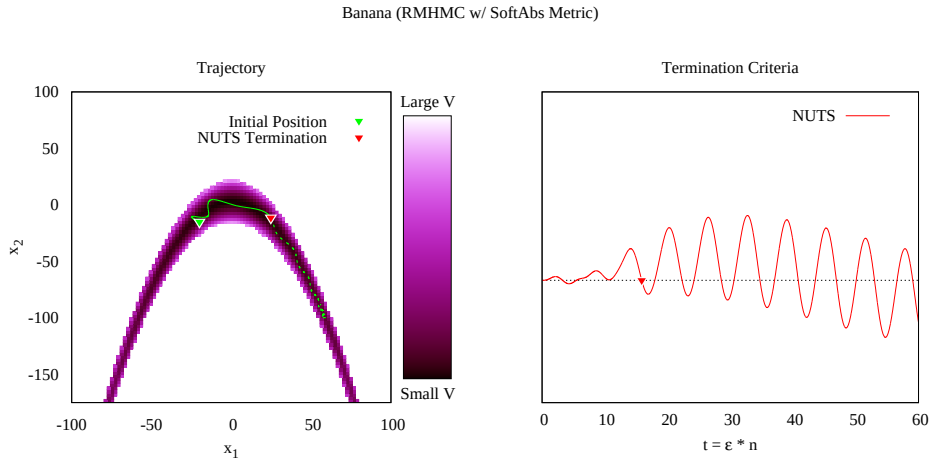


FIG. 5. RMHMC trajectories, here using the SoftAbs metric [8] with $\alpha = 1$, yield much smoother trajectories in the banana potential (4), but the ill-defined NUTS criterion still terminates prematurely.

Implementation status

- Metropolis-Hastings
- Gibbs
- Hamiltonian Monte Carlo

Vanilla NUTS implemented in stan, `http://mc-stan.org/`

Betancourt, M. J. (2013). Generalizing the No-U-Turn sampler to Riemannian manifolds. *arXiv preprint arXiv:1304.1920*.

Gubernatis, J. E. (2005). Marshall Rosenbluth and the Metropolis algorithm. *Physics of Plasmas (1994-present)*, *12*(5). doi: http://dx.doi.org/10.1063/1.1887186

Hoffman, M. D., & Gelman, A. (2011). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *arXiv preprint arXiv:1111.4246*.

Neil, R. M. (2010). MCMC using Hamiltonian dynamics. In Steve Brooks, Andrew Gelman, Galin Jones, & Xiao-Li Meng (Eds.), *Handbook of Markov Chain Monte Carlo* (pp. 113–162). Chapman & Hall / CRC Press.