Improving Markov Chain Sampling by Suppressing Random Walks

Radford M. Neal

Departments of Statistics and Computer Science
University of Toronto

http://www.cs.utoronto.ca/~radford/
radford@utstat.utoronto.ca

2 July 1996

The Problem to be Addressed

- Distributions we wish to sample from often have strong *dependencies*.

- Markov chain sampling must therefore proceed in *small steps*.

- In commonly-used schemes (eg, Gibbs sampling), these small steps will take the form of a *random walk*.

- Result: If the distribution spans a distance L, and we take steps of size d, we will need around $(L/d)^2$ steps to obtain an independent point using the Markov chain.

- Solution: Suppress the random walk. We then need only around (L/d) steps. This can be a very large improvement.

Two Ways to Suppress Random Walks

Hybrid Monte Carlo (Duane, Kennedy, Pendleton & Roweth 1987) can suppress random walks when sampling from many continuous distributions.

Hybrid Monte Carlo uses a dynamical simulation of a fictitious physical system, in which "momentum" keeps things going in the same direction. A number of variations are possible (eg, Horowitz 1991, Neal 1993).

Overrelaxation (Adler 1981, Barone & Frigessi 1990, Green & Han 1992, Neal 1995) appears less general than hybrid Monte Carlo, but may have advantages in some contexts.

An Illuminating Special Case: Sampling from a Uniform Distribution

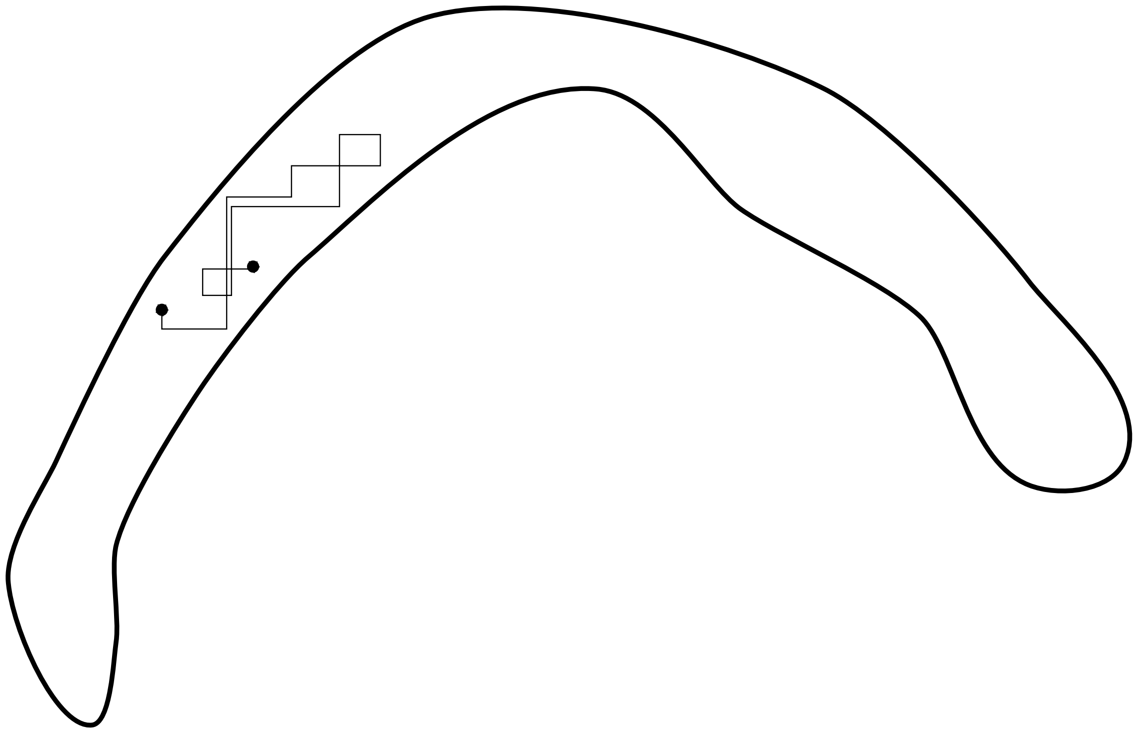
Consider sampling uniformly from some region of $\Re^n$. We could try several standard methods:

- Gibbs sampling: Successively sample from conditional distributions along the axes.
- The Hit & Run method: Sample from conditional distributions along randomly chosen directions.
- Single-variable Metropolis: Propose changes in one variable at a time.
- Simple multi-variable Metropolis: Propose changes in all variables at once, from some simple distribution.

All these methods will explore the region by a random walk, and will be slow in difficult cases (high dimensionality, high dependencies).

Gibbs Sampling for a 2D Example

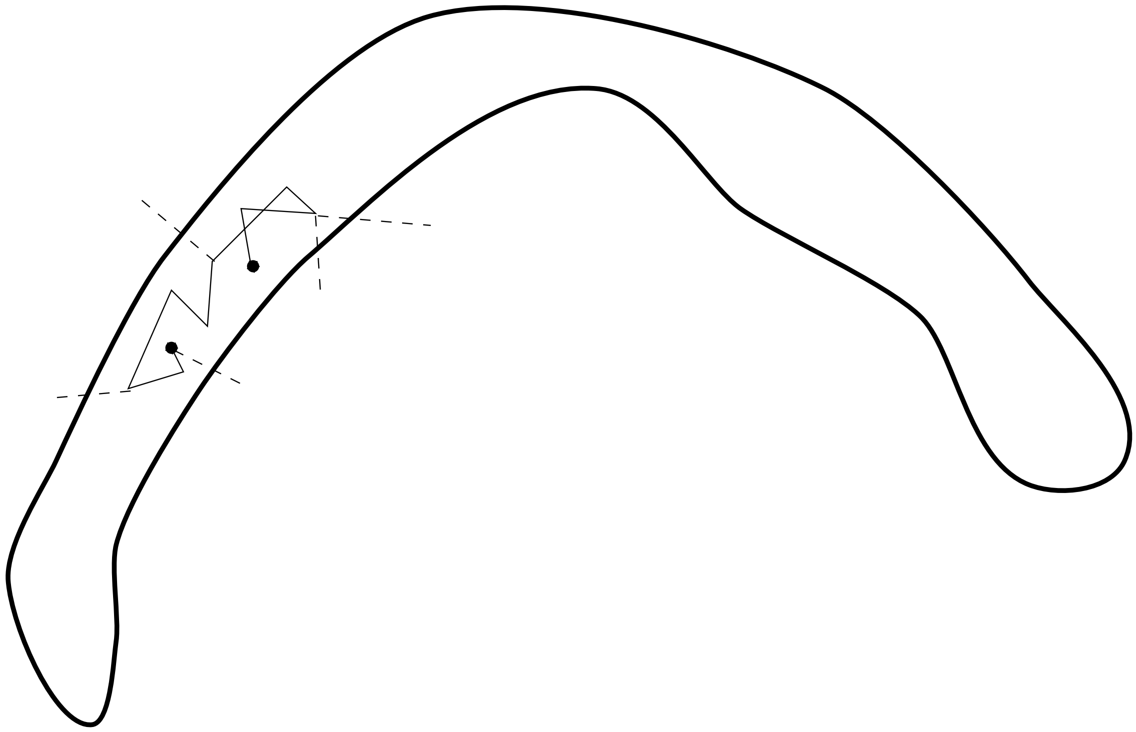
For the 2D example below, I have assumed that Gibbs sampling is local to one conditional mode, since jumping to a different mode would be unlikely in higher dimensions.



Single-variable Metropolis behaves similarly.

Simple Multi-Variable Metropolis for the 2D Example

I assume below that the proposal distribution is narrow enough to keep the acceptance rate high. For 2D problems, a very wide proposal distribution, with very low acceptance rate, is actually better, but this is not true in higher dimensions.



The hit & run method behaves similarly in high-dimensional problems.

Suppressing Random Walks when Sampling from a Uniform Distribution

We can often avoid doing a random walk by either of two methods:

- Move some distance from the current point, starting off in some direction, and reflecting when we reach the boundaries.
- Do overrelaxed updates along each axis in succession.

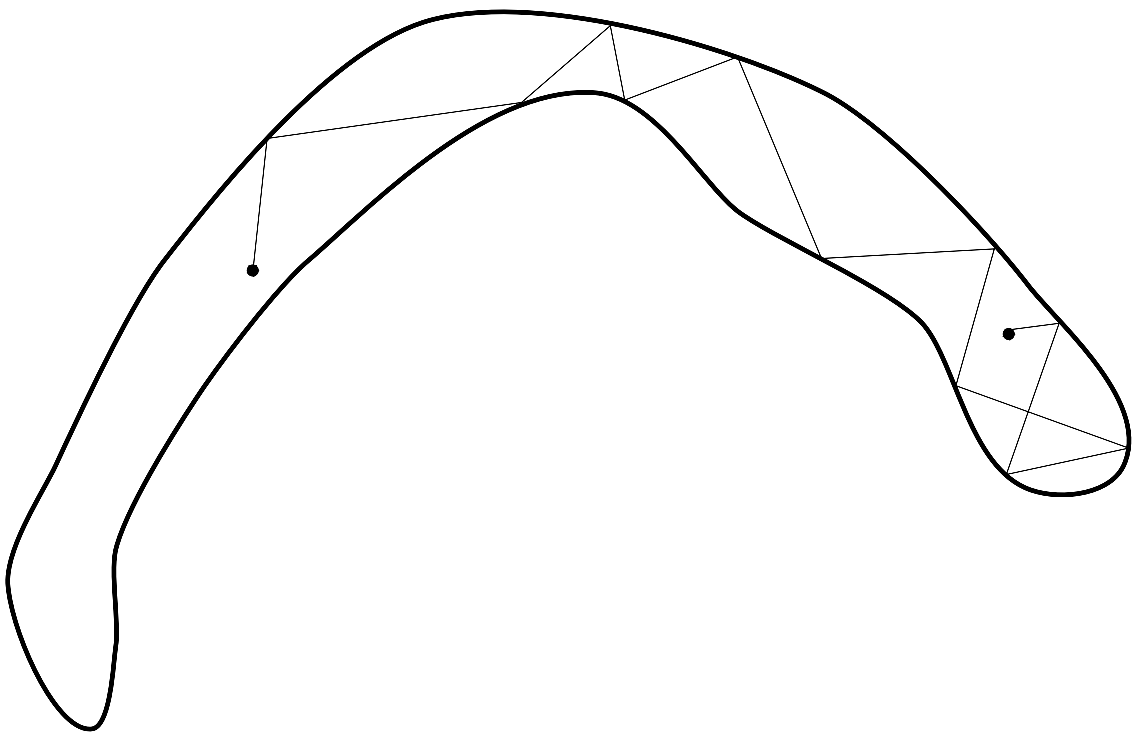
These are special cases of the more general Hybrid Monte Carlo and overrelaxation methods. We assume for the moment that all required computations can be done exactly.

To ensure ergodicity, we may mix in other updates (eg, Gibbs sampling or single-variable Metropolis) — but not too often, or we will not suppress random walks for long enough.

Sampling Using Exact Reflection for the 2D Example

Algorithm for sampling with reflection:

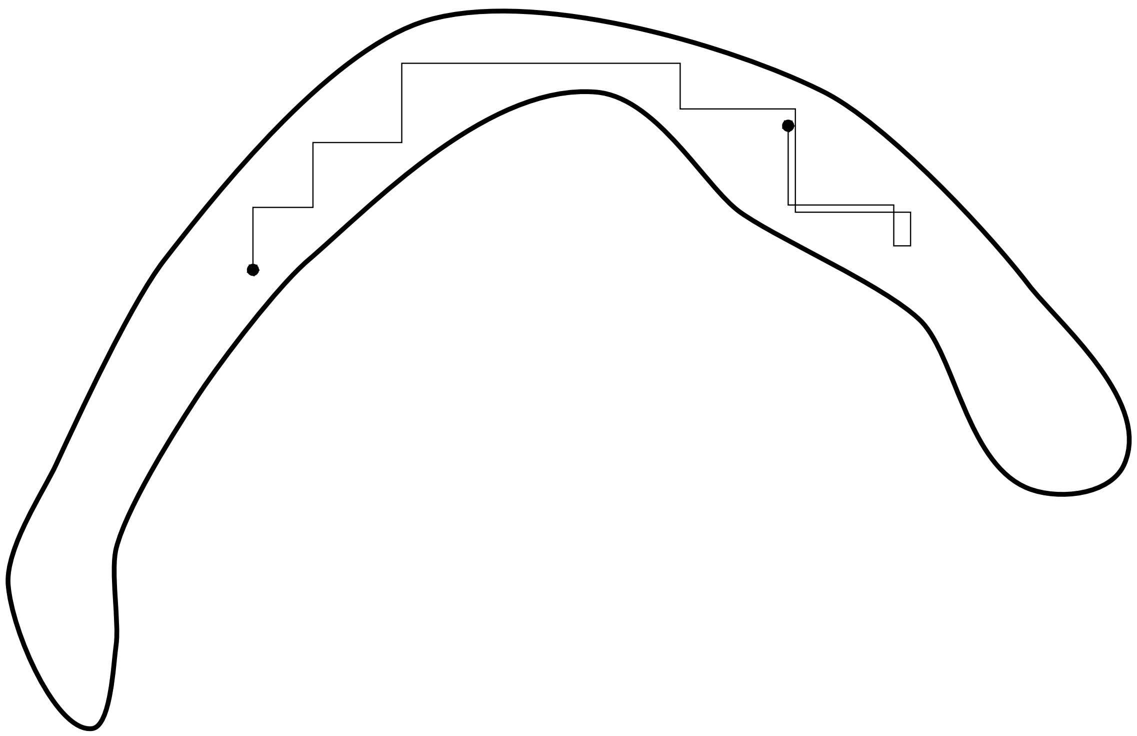
1. Choose a velocity vector from a spherical Gaussian distribution.
2. Travel for a pre-determined time with that velocity, except that when you hit a boundary, change direction by reflection.
3. Let the endpoint of this trajectory be the new state.



Sampling Using Exact Overrelaxation for the 2D Example

Algorithm: For each coordinate axis, in turn:

1. Compute the intersection points with the boundary, moving in the direction of this coordinate axis.
2. Find the mid-point of these intersections.
3. Move from the current state to the state the same distance from the mid-point, but on the opposite side.



Reflection vs. Overrelaxation

- Reflection requires computation of the normal vector at the boundary.

  Overrelaxation requires only computation of intersection points.

- Reflection is a rotationally-invariant algorithm; overrelaxation is not.

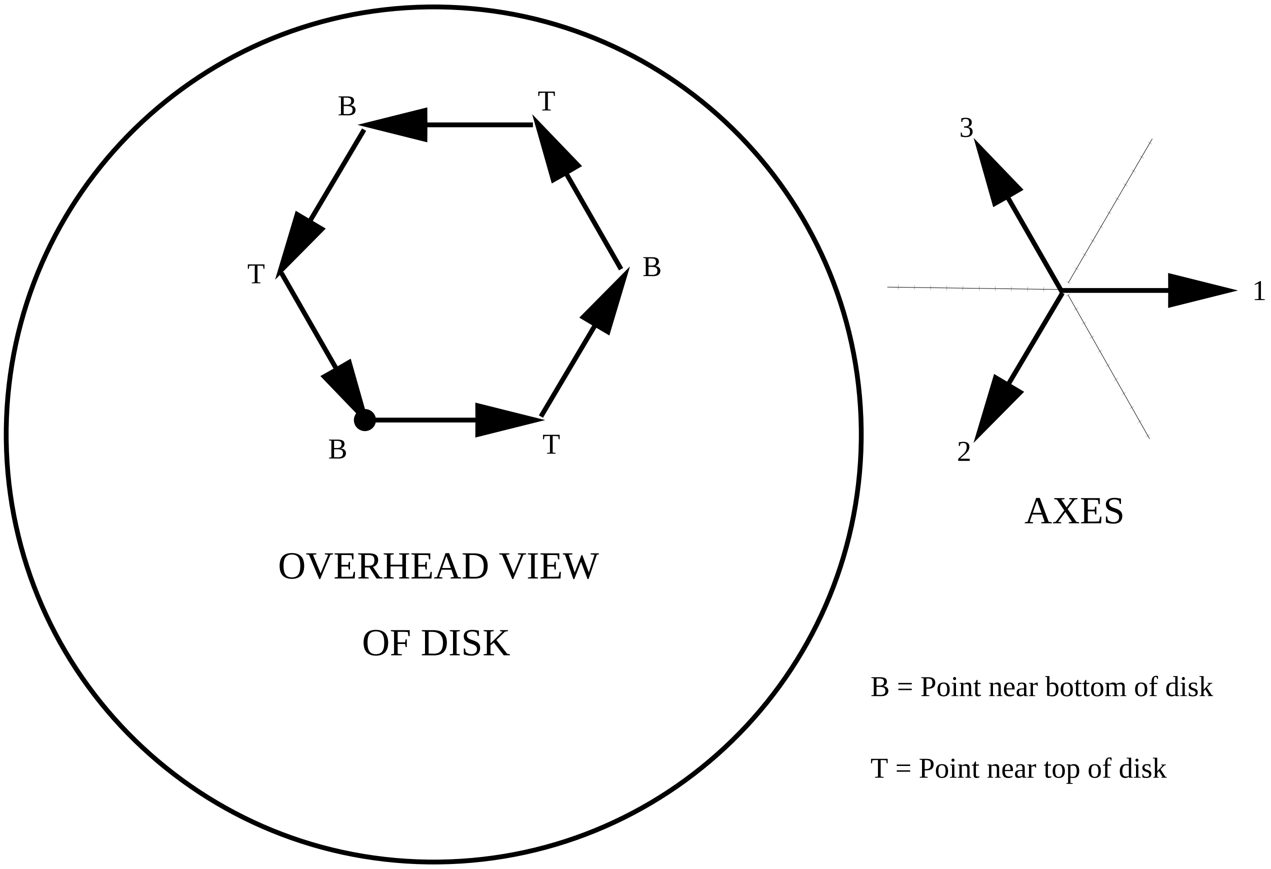
  Overrelaxation changes only one variable at a time in the original coordinate system. This sometimes allows computational savings.

- Reflection works well in some situations where overrelaxation works poorly.

  One way of looking at overrelaxation is as a cheap approximation to reflection.

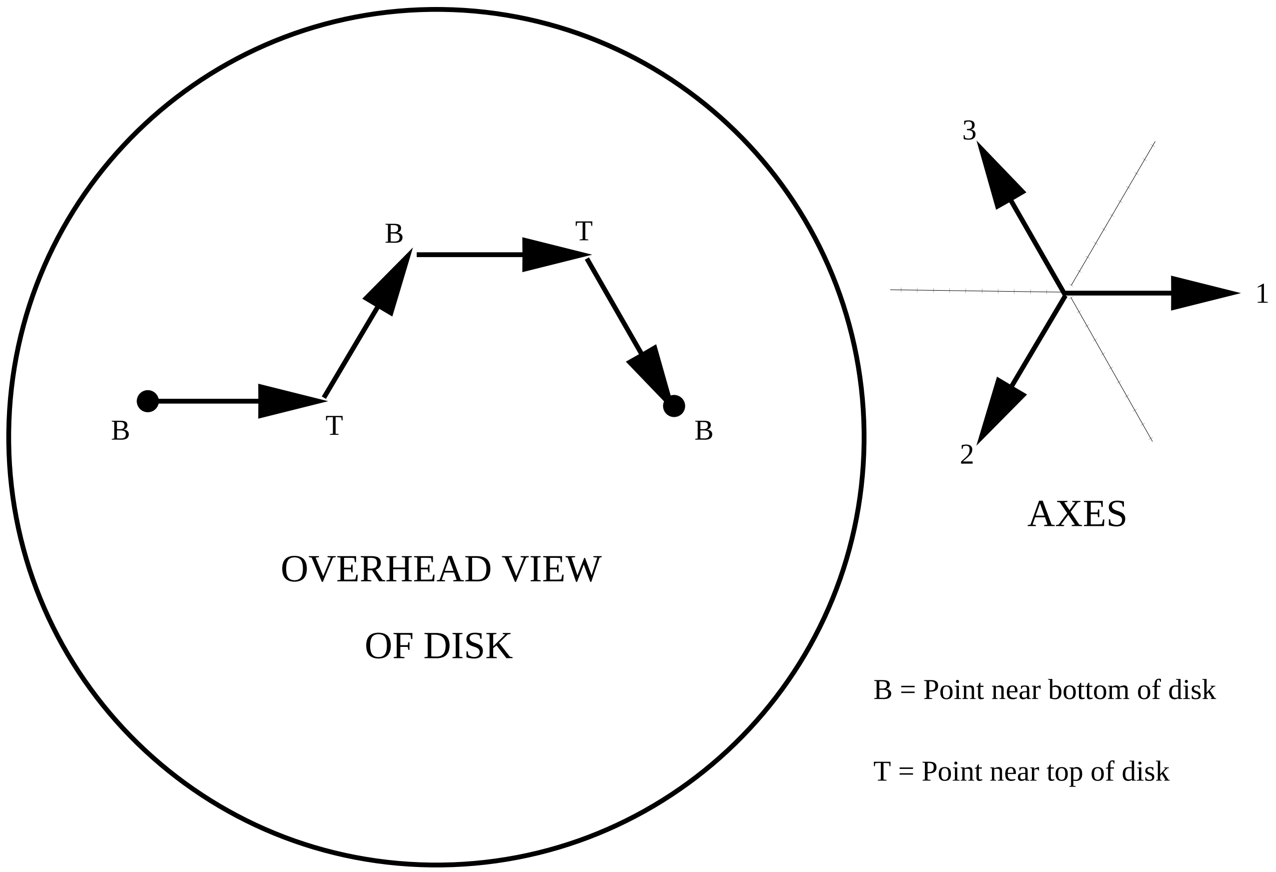
A 3D Example for Which Overrelaxation Does Not Work

Imagine a circular disk in three dimensions, drawn in projection below, with the coordinate axes pointing out of the page. As you can see, trying to sample using overrelaxation fails completely!



A Fix for Overrelaxation?

A possible fix for cases where overrelaxation does work: Instead of doing updates for each axis is turn (ie, 1 2 3) do a sequence of updates picked at random instead — eg, 1 2 1 3:



We persist with one such update sequence for a while, then pick another at random, which will take us in a new direction.

Methods for Non-Uniform Distributions

Reflection and overrelaxation for uniform distributions are related to corresponding methods for more general distributions.

Reflection is a special case of Hybrid Monte Carlo, in which motion occurs as in a physical system. For MCMC use, the dynamics is naturally described in the Hamiltonian formulation, which highlights the volume conservation property needed for validity.

Overrelaxation when conditional distributions are Gaussian is easily done by Adler's (1981) method. We update component x_i as follows:

$$x'_i \;=\; \mu_i \;+\; \alpha\,(x_i - \mu_i) \;+\; \sigma_i\,(1-\alpha^2)^{1/2}\,n$$

where μ_i and σ_i^2 are the mean and variance of the conditional distribution of x_i given $x_j : j \neq i$, and n is a Gaussian random variate with mean zero and variance one.

Problems with Existing Methods

Hybrid Monte Carlo works very well, but requires calculation of derivatives, and needs careful tuning of stepsize and trajectory length parameters.

Adler's overrelaxation method often works well, but does not apply when conditional distributions are non-Gaussian.

Brown and Woch (1987), Creutz (1987), Green and Han (1992), and Fodor and Jansen (1994) all propose more general overrelaxation methods, but all these have an uncontrolled probability of rejection, which can undermine suppression of random walks.

I proposed overrelaxation method based on order statistics that never rejects (Neal 1995), but it is not always easy to apply.

From Non-Uniform to Uniform: The Slice Sampling Approach

We can sample any continuous distribution by sampling uniformly from the region under its density function, or a multiple thereof, $f(x)$.

In *Slice Sampling* (Neal, in preparation), we sample under the density function as follows:

1. Sample uniformly from the vertical interval $[0, f(x)]$, giving a "slice point", s.

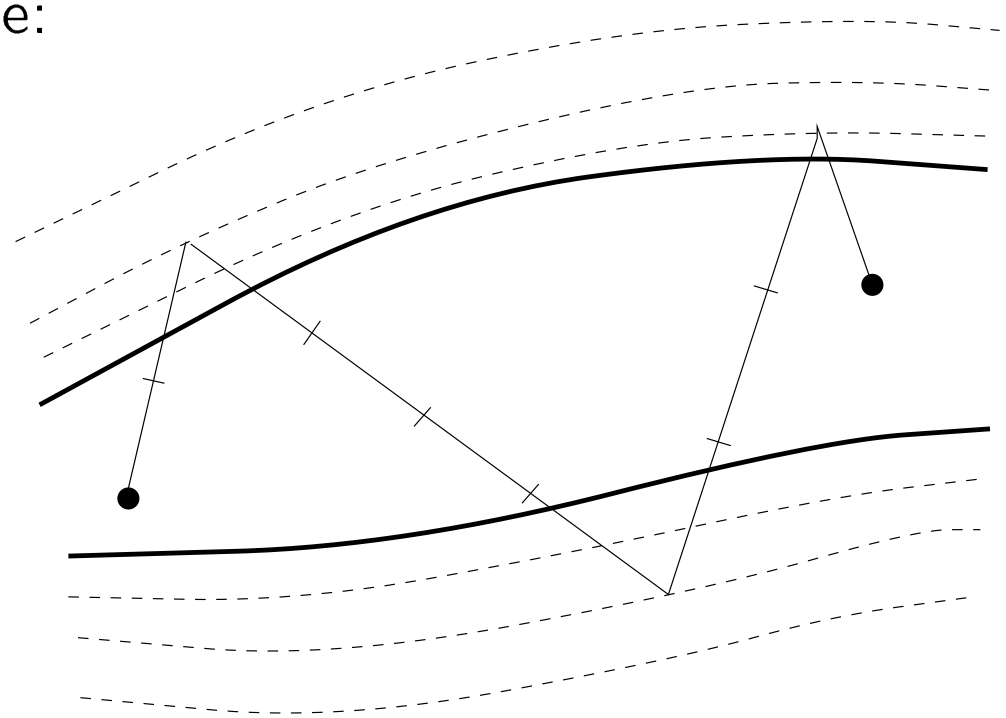
2. Do some update that leaves invariant the uniform distribution over the "slice", $\{x : f(x) = s\}$, producing a new state x'.

Applied to each variable in turn, slice sampling can be used instead of Gibbs sampling, when the latter is hard to implement, with step (2) sampling uniformly (or approximately so).

Step (2) can instead involve reflection or overrelaxation, as described previously.

Outside Reflection with Fixed Steps

In practice, computing exact intersection points where reflection occurs may be difficult. We can instead take fixed-sized steps, and reflect when we find ourselves outside the slice:

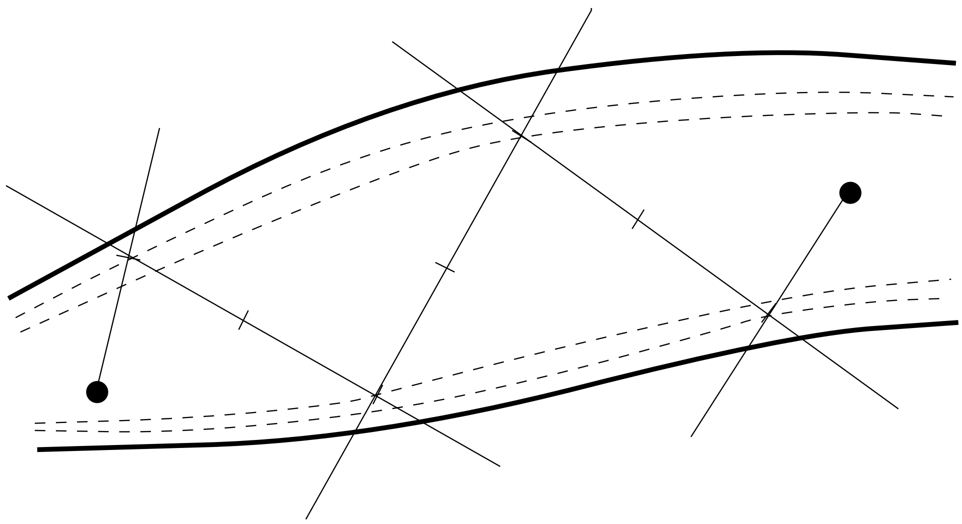


We could take many steps, then accept/reject based on whether we're outside at the end.

Or we could step only to the next inside point, rejecting if this takes too many steps (the velocity is retained if we accept, negated if we reject).

Inside Reflection with Fixed Steps

Alternatively, we can take fixed-sized steps and reflect from the inside point when the next step takes us outside the slice:



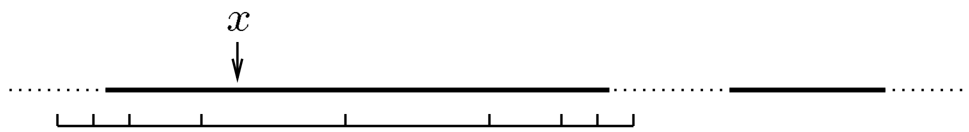
We must reject if the reverse trajectory would not take us outside at the same spot.

We could do many steps, rejecting them all if necessary, or take one step at a time. In the latter case, we retain the velocity if we accept, negate it if we reject.

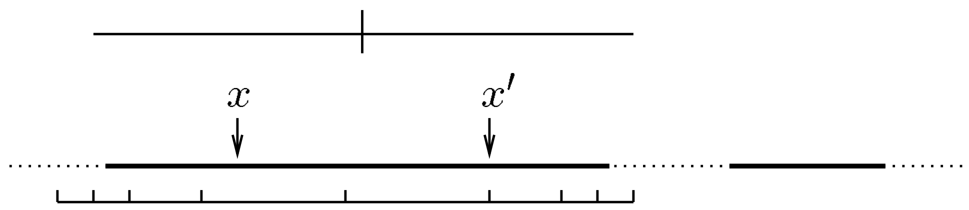
Overrelaxed Slice Sampling

To do an overrelaxed update using slice sampling, we first find an interval containing part of the slice:

We then bisect the two outermost steps to narrow down the edge of the slice as much as desired:



Finally, we flip to the opposite side of the middle:



We will have to reject if the new point is outside the slice, but the probability of rejection can be made arbitrarily small for unimodal distributions.

Issues in Comparing Methods

- Convenience of use — eg, are there parameters that have to be set carefully?
- Ability to exploit special features of the distribution — eg, fast re-calculation when just one variable is changed.
- Behaviour as the degree of dependence increases — eg, increasing correlation for Gaussian distributions.
- Behaviour as dimensionality increases — eg, by replication of independent copies of the system.

Which Methods are Most Promising?

- Hybrid Monte Carlo works very well, if you take the time to tune it carefully.

- Some form of overrelaxation seems attractive as a more convenient alternative to Hybrid Monte Carlo, which might also be more efficient in some circumstances. Overrelaxed slice sampling is perhaps the most generally applicable form.

- I've only begun to play with multivariate slice sampling using inside and outside reflection. Optimistically, one could hope for the advantages of Hybrid Monte Carlo, but with less tuning required.