The zig-zag

Gareth Roberts, University of Warwick

Joint work with Joris Bierkens

Talk outline

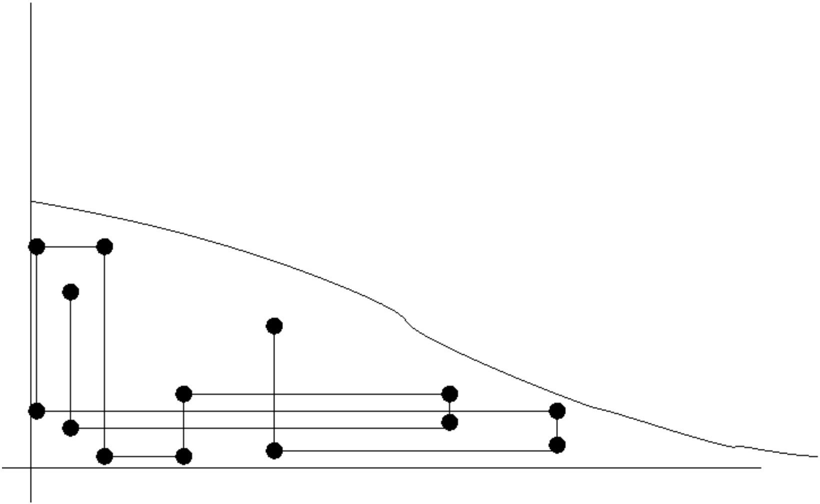
Plan

1. Consider why we might want non-reversible MCMC.

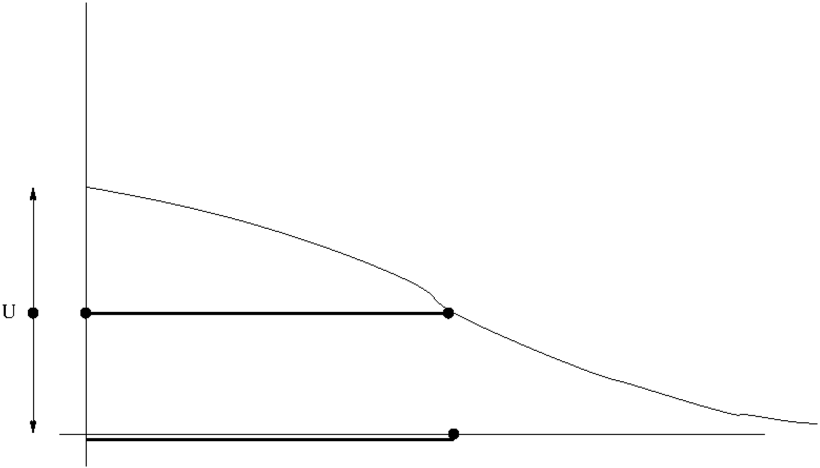
2. Describe schemes for carrying this out.

3. Introduce the zigzag

The slice sampler



An alternative



Metropolis-Hastings [Metropolis et al. 1953, Hastings 1970]

- S finite set (*state space*)
- $P(x, y)$ transition probabilities on S (*proposal chain*)
- $\pi(x)$ a probability distribution on S (*target distribution*)

Define acceptance probabilities

$$A(x, y) = \min\left(1, \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)}\right).$$

The Metropolis-Hastings (MH) chain is

$$Q(x, y) = \begin{cases} Q(x, y)A(x, y) & y \neq x, \\ 1 - \sum_{z \neq x} Q(x, z)A(x, z) & y = x. \end{cases}$$

The MH chain is reversible:

$$\pi(x)Q(x, y) = \pi(y)Q(y, x) \quad \forall x, y \in S.$$

In particular, π is invariant for P.

Non-Reversible Metropolis-Hastings [B., 2015]

- S finite set (*state space*)
- $P(x, y)$ transition probabilities on S (*proposal chain*)
- $\pi(x)$ a probability distribution on S (*target distribution*)
- $\Gamma \in \mathbb{R}^{S \times S}$: skew-symmetric matrix with zero row-sums (vorticity matrix)

Define acceptance probabilities

$$A(x, y) = \min \left(1, \frac{\pi(y)Q(y, x) + \Gamma(x, y)}{\pi(x)Q(x, y)} \right).$$

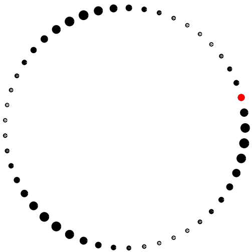
The non-reversible Metropolis-Hastings (MH) chain is

$$Q(x, y) = \begin{cases} Q(x, y)A(x, y) & y \neq x, \\ 1 - \sum_{z \neq x} Q(x, z)A(x, z) & y = x. \end{cases}$$

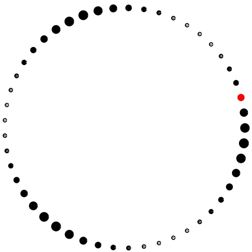
The NRMH chain is non-reversible:

$$\pi(x)Q(x, y) \neq \pi(y)Q(y, x) \quad \exists x, y \in S.$$

But π is invariant for P.



(a) Metropolis-Hastings



(b) Non-reversible Metropolis-Hastings

Cycles and lifting

Recall Γ skew-symmetric with zero row sums.

Also want acceptance probability

$$A(x, y) = \min\left(1, \frac{\pi(y)P(y, x) + \Gamma(x, y)}{\pi(x)P(x, y)}\right)$$

to be non-negative.

A 3-state example illustrates: No cycles $\Rightarrow$ no non-reversible Markov chain.

Cycles and lifting

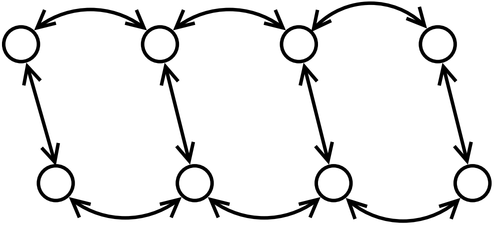
Recall Γ skew-symmetric with zero row sums.

Also want acceptance probability

$$A(x, y) = \min\left(1, \frac{\pi(y)P(y,x) + \Gamma(x,y)}{\pi(x)P(x,y)}\right)$$

to be non-negative.

A 3-state example illustrates: No cycles $\Rightarrow$ no non-reversible Markov chain.

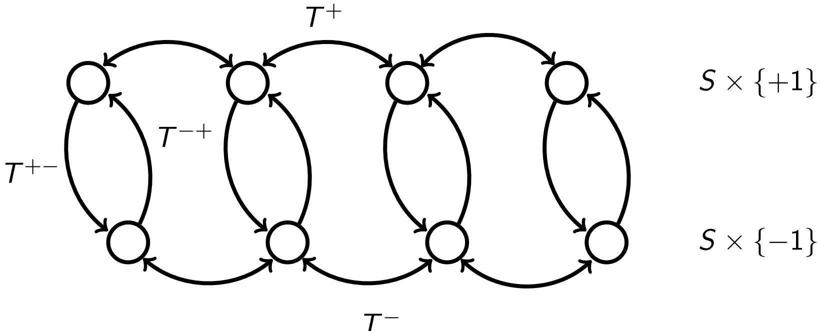


How to construct lifted MCMC algorithms?

A general lifted Markov chain [Turitsyn, Chertkov, Vucelja, 2011]

- State space S augmented to $S^\sharp = S \times \{-1, +1\}$.
- T^+, T^- are sub-Markov transition matrices on S.
- $T^\pm$ satisfy skew-detailed balance: for all $x, y \in S$, $\pi(x) T^+(x,y) = \pi(y) T^-(y,x)$.
- T^{-+}, T^{+-} transitions between replicas, e.g.

$$T^{-+}(x) = \max\left(0, \sum_{y \neq x} (T^+(x,y) - T^-(x,y))\right).$$



Lifted Metropolis-Hastings [Turitsyn, Chertkov, Vucelja, 2011]

How to choose T^+ and T^-?

Introduce a quantity of interest: $\eta : S \to \mathbb{R}$

Take (Q, π) reversible, e.g. Metropolis-Hastings chain.

Define

$$T^+(x,y) := \begin{cases} Q(x,y) & \text{if } \eta(y) \geq \eta(x) \\ 0 & \text{if } \eta(y) < \eta(x). \end{cases}$$

$$T^-(x,y) := \begin{cases} Q(x,y) & \text{if } \eta(y) \leq \eta(x) \\ 0 & \text{if } \eta(y) > \eta(x). \end{cases}$$

Then skew-detailed balance is satisfied:

$$\pi(x) T^+(x,y) = \pi(y) T^-(y,x) \quad \text{for all } x, y.$$

In practice, Lifted Metropolis-Hastings algorithm:

- Propose according to proposal chain P
- If move is allowed, accept with MH acceptance probability
- If move is not allowed, possibly switch replica.

Does lifting solve the non-reversible MCMC problem?

The problem is that we need to know the switching probabilities, eg

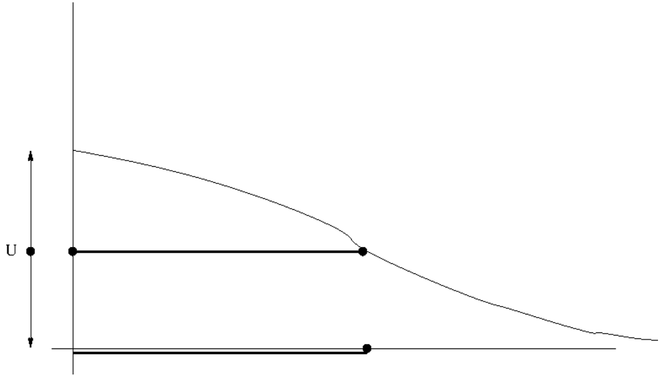
$$T^{-+}(x) = \max\left(0, \sum_{y \neq x} (T^{+}(x, y) - T^{-}(x, y))\right).$$

This will typically be difficult to calculate, usually impossible in continuous state spaces.

But, mathematically we can take a limit of smaller proposed moves and speed up the process to obtain a continuous time limit.

We initially did this for the Curie-Weiss model in statistical physics (http://arxiv.org/abs/1509.00302).

Another look at our initial example ...



Instead of apriori drawing the uniform random variable, change direction with hazard rate

$$\max\{0, -(\log \pi)'(x)\}$$

One-dimensional zig zag process

Langevin diffusion generator (for comparison):

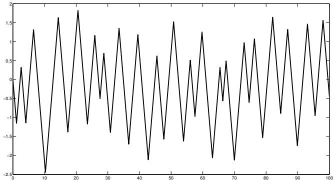
$$Lf(y) = -\Psi'(y)\frac{df}{dy} + \frac{d^2 f}{dy^2}, \quad y \in \mathbb{R}.$$

Invariant density $\pi(y) \propto \exp(-\Psi(y))$

Zig zag process generator:

$$Lf(y,j) = aj\frac{df}{dy} + \lambda(y,j)(f(y,-j) - f(y,j)), \quad y \in \mathbb{R}, j \in \{-1,+1\}.$$

Here speed $a > 0$ and switching rate $\lambda(y,j) \geq 0$.



The stochastic process dates back to the telegraph process as

Implementation

How do we simulate continuous time stochastic process like this?

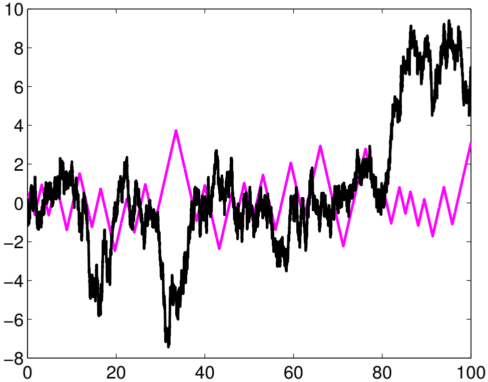
By using thinned poisson processes

For example, if $\|\Psi'(x)| < c$, simulate a Poisson process of rate c (by simulating the exponential inter-arrival times). Then at each poisson time, we accept as a direction change with probability $\max(-\Psi'(x), 0)/c$.

This makes the algorithm inexpensive to implement as we only need to calculate $\Psi'(x)$ occasionally.

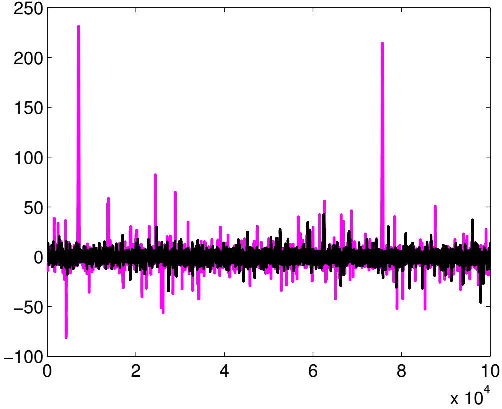
There are many other details .... some not completely worked out yet ...

Zig zag process for sampling the Cauchy distribution



$T = 100$

Zig zag process for sampling the Cauchy distribution



$T = 10,000$

Work in progress

- Multi-dimensional zig zag process: here we have a multi-dimensional binary velocity, eg $(1, -1, -1, 1, 1, -1, 1, 1)$.
- Efficient sampling (currently for potentials with Lipschitz gradients in multiple dimensions)
- Convergence properties, both theoretically and experimentally (and how to compare to other methods)
- Can this be a competitor to Hamiltonian MCMC?

[Bierkens + R, *A Piecewise Deterministic Scaling Limit of Lifted Metropolis-Hastings in the Curie-Weiss model*, 2015]