A Repelling-Attracting Metropolis (RAM) Algorithm for Multimodality

Tak (Hyungsuk) Tak[1], Xiao-Li Meng[1], and David A. van Dyk[2]

Harvard University[1] and Imperial College London[2]

International CHASC Astro-Statistics Collaboration

9 June 2016

Motivation

Time delay estimation problem

Credit: NASA's Goddard Space Flight Center

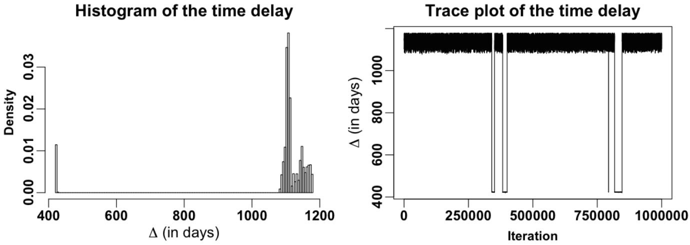
The strong gravitational field of a lensing galaxy splits light into two images.

- Light rays take different routes with different lengths.
- Difference between their arrival times $\rightarrow$ **Time delay (Δ)**
- **Time delay** is used to infer cosmological parameters including H_o.

Motivation (cont.)

Modeling time delay $\rightarrow$ MCMC for multimodality

- Full posterior density function: $\pi(\Delta, \boldsymbol{\theta} \mid \mathrm{Data})$ (Tak et al., arXiv).
- Metropolis within Gibbs sampler (Tierney, 1994)
- A multimodal (marginal) posterior of Δ for Quasar Q0957+561



- Just 8 jumps out of a million iterations!
- Could we improve Metropolis' ability to jump between modes without losing its simple-to-implement characteristic?

Idea

There is a RAM on top of the mountain.
How would this RAM move to the top of the other mountain?

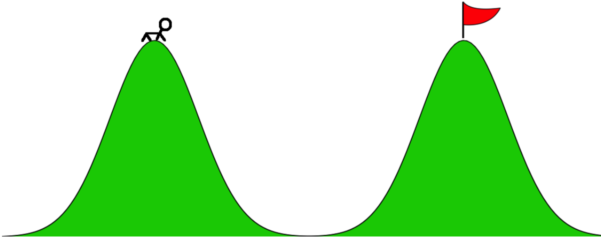


Image credit: www.launsteinimagery.com, www.cliparts.com, www.iconfinder.com

Idea (cont.)

1. Make a down-up movement in density to generate a proposal x''.

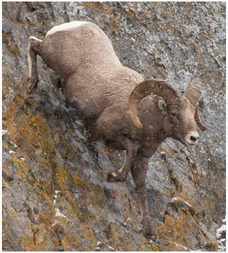
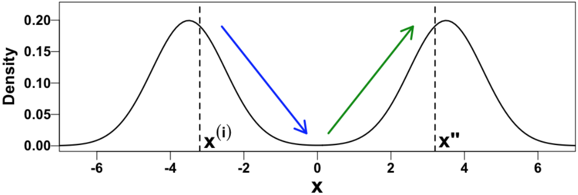


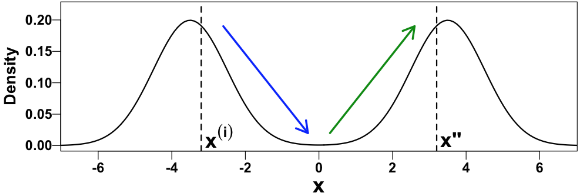
Image credit: http://www.bestofthetetons.com/, http://blog.showmenaturephotography.com/



2. Accept or reject x'' with probability $\min\left\{1, \frac{\pi(x'')q^{\mathrm{DU}}(x^{(i)}|x'')}{\pi(x^{(i)})q^{\mathrm{DU}}(x''|x^{(i)})}\right\}$.

Note: q^{DU} is a down-up (DU) proposal density.

RAM: Proposal



Two-step procedure

$x^{(i)}$: Current state $\searrow$ x': Intermediate proposal $\nearrow$ x'': Final proposal

1. (Downhill Metropolis) Generate $x' \sim q(\cdot \mid x^{(i)}) \sim N(x^{(i)}, \sigma^2)$ and accept x' with probability $\alpha_\epsilon^{\mathrm{D}}(x' \mid x^{(i)}) = \min\left\{1, \frac{\pi(x^{(i)})+\epsilon}{\pi(x')+\epsilon}\right\}$. Repeat this proposal step until one is accepted (forced Metropolis).

2. (Uphill Metropolis) Generate $x'' \sim q(\cdot \mid x')$ and accept x'' with probability $\alpha_\epsilon^{\mathrm{U}}(x'' \mid x') = \min\left\{1, \frac{\pi(x'')+\epsilon}{\pi(x')+\epsilon}\right\}$ Repeat this proposal step until one is accepted (forced Metropolis).

Note: $\epsilon = 10^{-308}$ to prevent a ratio of zeros (0/0).

RAM: Acceptance/Rejection

Accept x'' with a Metropolis-Hastings acceptance probability

$$\begin{aligned}\alpha^{\mathrm{DU}}(x'' \mid x^{(i)}) &= \min\left\{1,\ \frac{\pi(x'')q^{\mathrm{DU}}(x^{(i)} \mid x'')}{\pi(x^{(i)})q^{\mathrm{DU}}(x'' \mid x^{(i)})}\right\} \\ &= \min\left\{1,\ \frac{\pi(x'')\int q(x \mid x^{(i)})\alpha_\epsilon^{\mathrm{D}}(x \mid x^{(i)})dx}{\pi(x^{(i)})\int q(x \mid x'')\alpha_\epsilon^{\mathrm{U}}(x \mid x'')dx}\right\}.\end{aligned}$$

Is there a way to avoid calculating this intractable ratio?

If we explore an extended space with a correct marginal $\pi(x)$, then there can be a way to cancel this intractable ratio (Møller et al., 2006).

RAM: Auxiliary variable approach

An auxiliary variable z with $\pi^{\mathrm{C}}(z \mid x)$ well-defined.

- Joint target density: $\pi^{\mathrm{J}}(z, x) = \pi(x)\pi^{\mathrm{C}}(z \mid x) = \pi(x)q(z^{(i)} \mid x^{(i)})$
- Joint proposal density:

$$\begin{aligned} q^{\mathrm{J}}(z'', x'' \mid z^{(i)}, x^{(i)}) &= q_1(x'' \mid z^{(i)}, x^{(i)})q_2(z'' \mid x'', z^{(i)}, x^{(i)}) \\ &= q^{\mathrm{DU}}(x'' \mid x^{(i)})q^{\mathrm{D}}(z'' \mid x'') \end{aligned}$$

  Note: q^{D} is a forced downhill kernel density.

- Joint acceptance probability:

$$\begin{aligned} \alpha^{\mathrm{J}}(z'', x'' \mid z^{(i)}, x^{(i)}) &= \min\left[1,\ \frac{\pi^{\mathrm{J}}(z'', x'')q^{\mathrm{J}}(z^{(i)}, x^{(i)} \mid z'', x'')}{\pi^{\mathrm{J}}(z^{(i)}, x^{(i)})q^{\mathrm{J}}(z'', x'' \mid z^{(i)}, x^{(i)})}\right] \\ &= \min\left[1,\ \frac{\pi(x'')\min\{1, \frac{\pi(x^{(i)})+\epsilon}{\pi(z^{(i)})+\epsilon}\}}{\pi(x^{(i)})\min\{1, \frac{\pi(x'')+\epsilon}{\pi(z'')+\epsilon}\}}\right] \end{aligned}$$

RAM: Overall algorithm

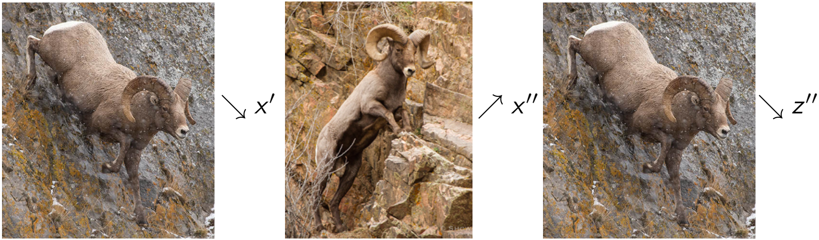


Image credit: http://www.bestofthetetons.com/, http://blog.showmenaturephotography.com/

A RAM is composed of four steps in each iteration.

Steps 1–3: Generating a joint proposal (z'', x'')

1. $(\searrow)$ Redraw $x' \sim q(\cdot \mid x^{(i)})$ until $u_1 \sim \mathrm{Unif}(0,1) < \alpha_\epsilon^{\mathrm{D}}(x' \mid x^{(i)})$
2. $(\nearrow)$ Redraw $x'' \sim q(\cdot \mid x')$ until $u_2 \sim \mathrm{Unif}(0,1) < \alpha_\epsilon^{\mathrm{U}}(x'' \mid x')$
3. $(\searrow)$ Redraw $z'' \sim q(\cdot \mid x'')$ until $u_3 \sim \mathrm{Unif}(0,1) < \alpha_\epsilon^{\mathrm{D}}(z'' \mid x'')$

Step 4: Accept or reject the joint proposal (z'', x'')

4. Set $(z^{(i+1)}, x^{(i+1)}) = (z'', x'')$ if $u_4 < \alpha^{\mathrm{J}}(z'', x'' \mid z^{(i)}, x^{(i)})$, where $u_4 \sim \mathrm{Unif}(0,1)$, and set $(z^{(i+1)}, x^{(i+1)}) = (z^{(i)}, x^{(i)})$ otherwise.

Example: Quasar Q0957+561 (Hainline et al, 2012)

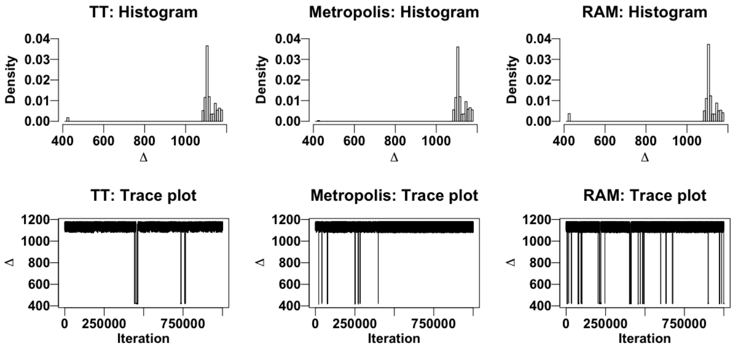
Metropolis within Gibbs sampler for $p(\Delta, \boldsymbol{\beta}, \boldsymbol{\theta} \mid \text{Data})$

- At iteration i,

$$\text{Step 1: Sample } \Delta^{(i)} \sim p(\Delta \mid \boldsymbol{\theta}^{(i-1)}, \text{Data})$$
$$\text{Step 2: Sample } \boldsymbol{\theta}^{(i)} \sim p(\boldsymbol{\theta} \mid \Delta^{(i)}, \text{Data})$$

- We use tempered transitions (TT) (Neal, 1996), Metropolis or RAM to draw Δ in Step 1.
- We run 10 chains each of length 150,000 with 50,000 burn-in, spreading 10 initial values of Δ across its space $[-1100,\ 1100]$.
- We set an arbitrarily large proposal scale ($\sigma = 400$), assuming modal locations are unknown.
- We consider the CPU time required by each algorithm, running and thinning longer chains so that each chain has 100,000 samples.

Example: Quasar Q0957+561 (cont.)



	Total # of iterations of each chain	Avg. CPU time (seconds)	# of jumps between modes
TT	150,000	3,083	16
Metropolis	1,992,032	3,085	52
RAM	488,267	3,080	120

Conclusion

Take-home messages

- Easy to implement (Keep it in your MCMC tool box!)
- Always possible to replace Metropolis with RAM for multimodality.
- Our extensive simulation studies will be reported in the final paper (some of them are on arXiv).

Future directions

- Theoretical convergence rate
- Possibly many (and better) down-up schemes, e.g., anti-Langevin (Christian P. Robert) or negative temperature (Art B. Owen)
- A global optimizer based on the down-up idea (analog to annealing)

Reference

1. Hainline, L. J. et al. (2012) "A New Microlensing Event in the Doubly Imaged Quasar Q0957+ 561" *The Astrophysical Journal*, **744**, 2, 104.
2. Møller, J., Pettitt, A. N., Reeves, R., Berthelsen, K. K. (2006) "An Efficient Markov Chain Monte Carlo Method for Distributions with Intractable Normalising Constants" *Biometrika*, **93**, 2, 451–458.
3. Neal, R. M. (1996) "Sampling From Multimodal Distributions Using Tempered Transitions" *Statistics and Computing*, **6**, 4, 353–366.
4. Tak, H., Meng, X.-L., van Dyk, D. A. (in progress), "A Repulsive-Attractive Metropolis Algorithm for Multimodality." arXiv 1601.05633
5. Tak, H., Mandel, K., van Dyk, D. A., Kashyap, V. L., Meng, X.-L., Siemiginowska, A. (in progress), "Bayesian Estimates of Astronomical Time Delays between Gravitationally Lensed Stochastic Light Curves." arXiv 1602.01462
6. Tierney, L. (2009) "Markov Chains for Exploring Posterior Distributions" *The Annals of Statistics*, **22**, 4, 1701–1728.