Sampling from a Gaussian in high dimensions

Frederik Beaujean
C2PAP, Universe Cluster, LMU Munich

Data 2016

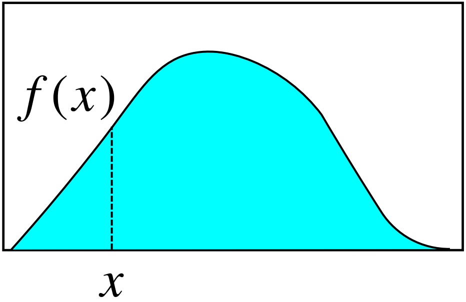
C2PAP
Excellence Cluster Universe

Motivation

- want to test samplers in $\geq 10^7$D
- HMC seems champion but worry that leapfrog integrator too unstable in very high dimensions
- can we take over characteristics of HMC: use gradient, unit acceptance
- *silly example* SKILLING: unit Gaussian
- difficult in high D if matrices too large to be inverted/stored. Wiener filter example

$$\text{posterior mean} = D^{-1} R N^{-1} \boldsymbol{d},\ D = S^{-1} + R N^{-1} R$$

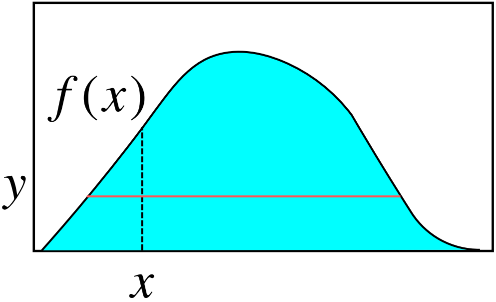
Slice sampling: Sample under the curve $f(x)$



Auxiliary variable $x \to (x, y)$
sample $P(x, y) = \text{const.}$ with Gibbs

1. vertical slice: $y \sim \mathcal{U}_{[0,f(x)]}$
2. horizontal slice: $x \sim \mathcal{U}_S, S = \{x \mid f(x) \geq y\}$

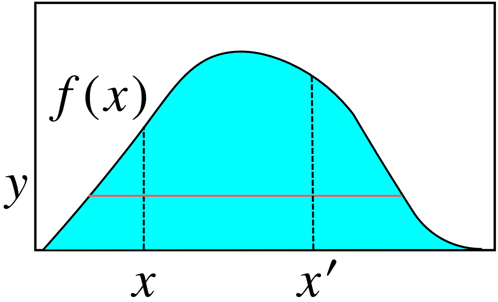
Slice sampling: Sample under the curve $f(x)$



Auxiliary variable $x \to (x, y)$
sample $P(x, y) = \text{const.}$ with Gibbs

1. vertical slice: $y \sim \mathcal{U}_{[0,f(x)]}$
2. horizontal slice: $x \sim \mathcal{U}_S, S = \{x \,|\, f(x) \geq y\}$

Slice sampling: Sample under the curve $f(x)$



Auxiliary variable $x \to (x, y)$
sample $P(x, y) = \text{const.}$ with Gibbs

1. vertical slice: $y \sim \mathcal{U}_{[0, f(x)]}$
2. horizontal slice: $x \sim \mathcal{U}_S, S = \{x \mid f(x) \geq y\}$

Slice sampling as a Metropolis algorithm

accept new x' if

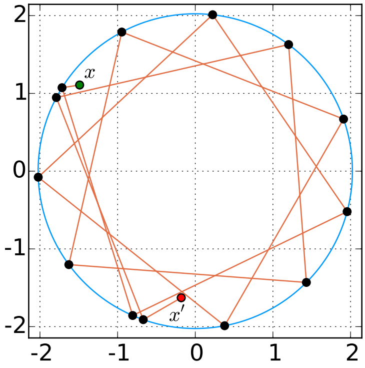
$$u \leq \frac{f(x')}{f(x)}, \; u \sim \mathcal{U}_{[0,1]}$$

if u known before x' proposed, we know all $x' : f(x') \geq y \equiv u\, f(x)$
$\Rightarrow$ no point on slice is rejected

Reflective slice sampling: bouncing around

special case of HMC

- flat potential
- new momentum at start of trajectory
- particle moves in straight line
- kinetic energy conserved in reflection
- integrate for time T
- here: bounce point analytically known
- cost: 1 function + 1 gradient $\forall$ bounce



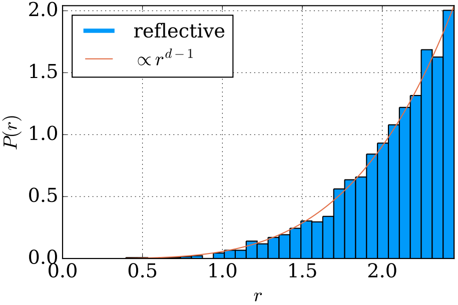
unit Gaussian in 2D: slice

$$S = \{x \mid f(x) \geq y\}$$
$$= \text{ball of radius } r_{\max}$$

Concentration of measure

uniform distribution on ball of radius $r_{\max}$ in d dimensions, volume grows exponentially with d at boundary. On dimensional grounds

$$P(r)\,\mathrm{d}r = \frac{d}{r_{\max}}\left(\frac{r}{r_{\max}}\right)^{d-1}\mathrm{d}r$$

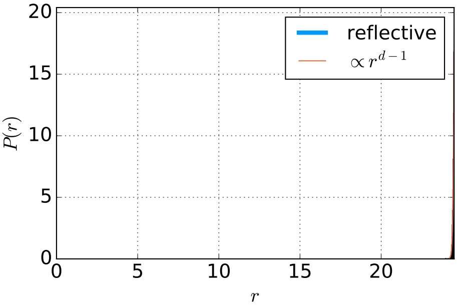


5D

Concentration of measure

uniform distribution on ball of radius $r_{\max}$ in d dimensions, volume grows exponentially with d at boundary. On dimensional grounds

$$P(r)\,\mathrm{d}r = \frac{d}{r_{\max}}\left(\frac{r}{r_{\max}}\right)^{d-1}\,\mathrm{d}r$$



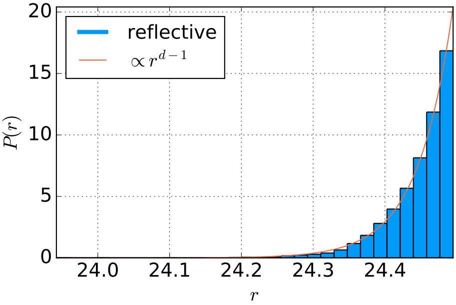
500D

Concentration of measure

uniform distribution on ball of radius $r_{\max}$ in d dimensions, volume grows exponentially with d at boundary. On dimensional grounds

$$P(r)\,\mathrm{d}r = \frac{d}{r_{\max}}\left(\frac{r}{r_{\max}}\right)^{d-1}\mathrm{d}r$$

all probability at surface, not inside



500D zoom

Effective sample size

Estimate expectation value of $g(x)$ with Monte Carlo

$$\bar{g} = \frac{1}{N} \sum_i g(x_i)$$

$$V[\bar{g}] = \frac{V[g]}{N} \times \tau$$

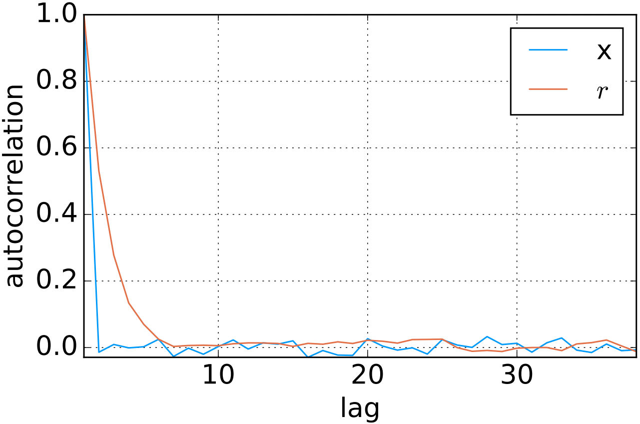
$$ESS = \frac{N}{\tau} \text{ or just } \frac{1}{\tau}$$

integrated autocorrelation time from autocorrelation $\hat{C}$ at lag t

$$\tau = 1 + 2 \sum_{t=1}^{\infty} \hat{C}(t)$$

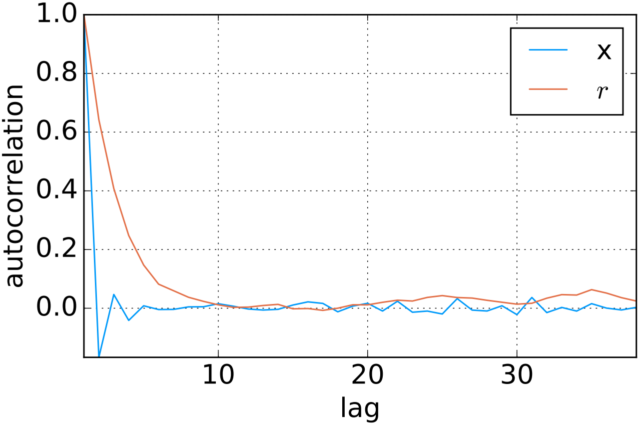
Independent samples $\Rightarrow \tau = 1$
estimating τ is nontrivial!

Sampling performance for reflections



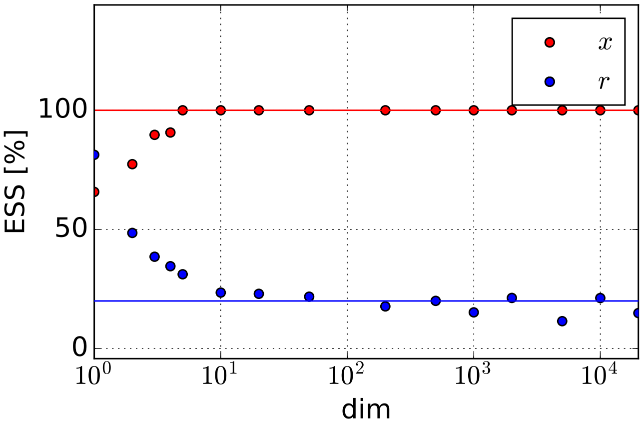
5D

Sampling performance for reflections



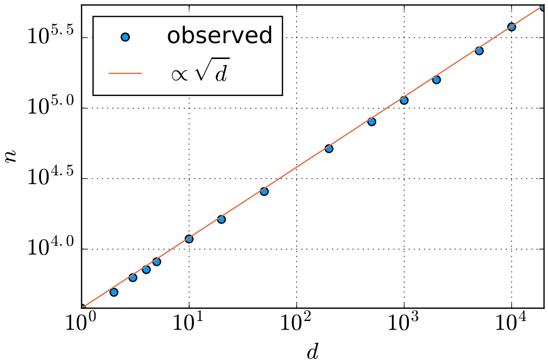
500D

Sampling performance for reflections



- fixed T
- x: small T, small ESS, capped at 100 %
- r: limiting ESS with scatter, independent of T

Bounces: universal behavior



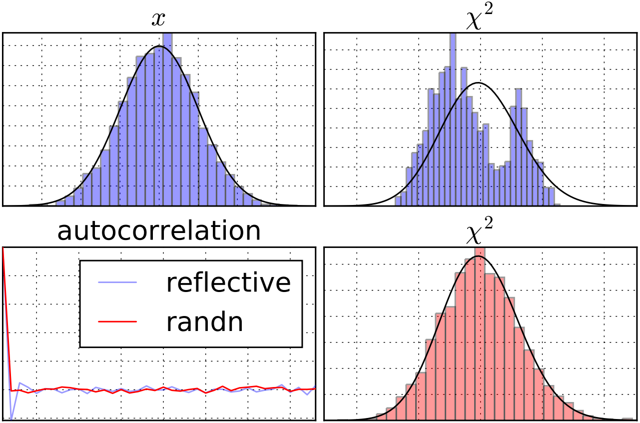
same runs as before, number of reflections (cost/iteration)

$$n \approx n(d_0)\sqrt{d/d_0}$$

simple geometric arguments?

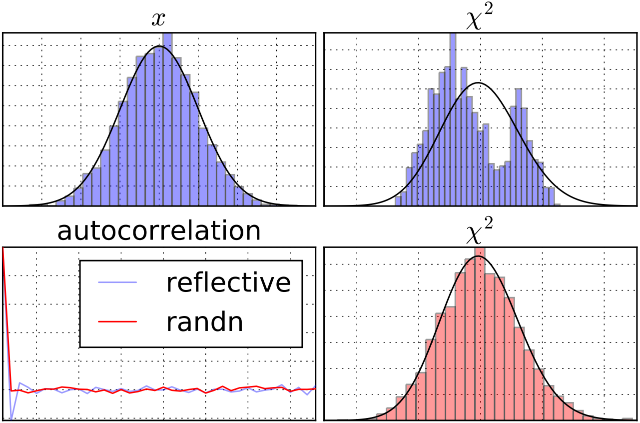
compare to unconstrained HMC: cost $\sim d^{1/4}$

Back to slice sampling



500 D

change $r_{\max}$, bounce for time T, update $r_{\max}$, repeat

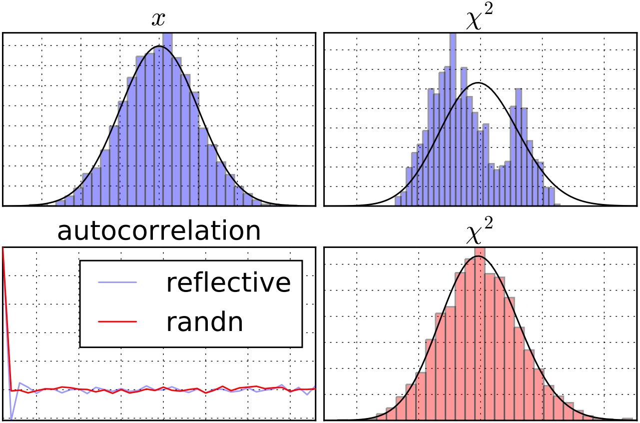


500 D

change $r_{\max}$, bounce for time T, update $r_{\max}$, repeat

Works very well in 1D subspace!

Back to slice sampling



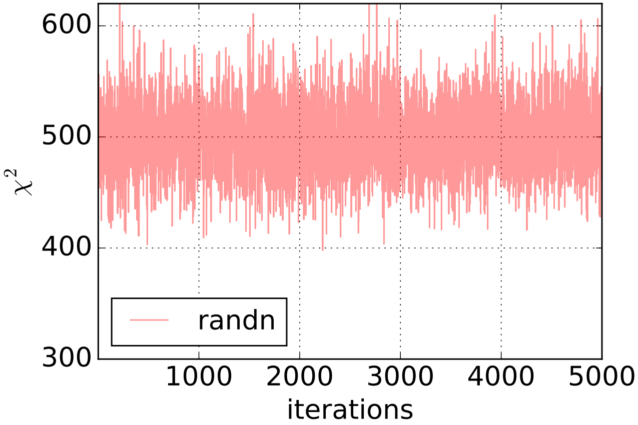
500 D

change r_{max}, bounce for time T, update r_{max}, repeat

Works very well in 1D subspace!

What's wrong in full space?

Back to slice sampling



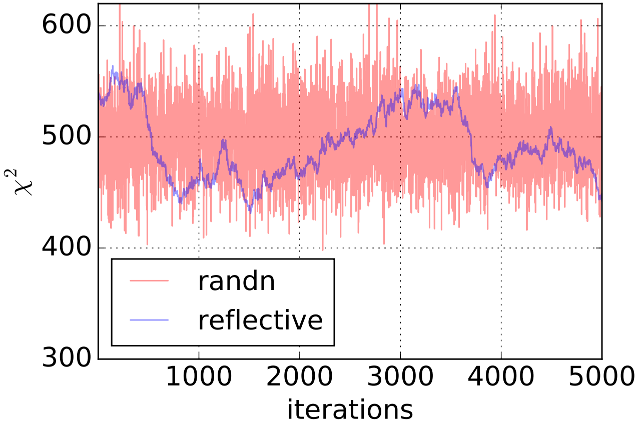
500 D

change r_{max}, bounce for time T, update r_{max}, repeat

Works very well in 1D subspace!

What's wrong in full space?

Back to slice sampling



500 D

change r_{max}, bounce for time T, update r_{max}, repeat

Works very well in 1D subspace!

What's wrong in full space?

Limitations of Metropolis Caracciolo, Polisetto, Sokal (1993)

$$u \leq \frac{f(x')}{f(x)} \Rightarrow \log f(x') \geq \log f(x) + \log u$$

$$u \sim \mathcal{U}_{[0,1]} \Rightarrow \log u \sim \text{Exp}(1) \in \mathcal{O}(1)$$

Change to $\log f$ (*energy*) indep. of d!

$$f(x) = \prod_{i=1}^{d} \mathcal{N}(x_i \mid 0, 1) \Rightarrow \log f(x) \sim -\frac{1}{2}\chi^2 = -\frac{1}{2}\sum_{i=1}^{d} x_i^2 \in \mathcal{O}(d)$$

independent sampler

$$E[\chi^2] = d, \;\; V[\chi^2] = 2d$$

so fluctuations are $\mathcal{O}\left(\sqrt{d}\right)$. Random walk in 1D needs d steps of $\mathcal{O}(1)$ to travel a distance $\sqrt{d}$

Limitations of Metropolis CARACCIOLO, POLISETTO, SOKAL (1993)

$$u \leq \frac{f(x')}{f(x)} \Rightarrow \log f(x') \geq \log f(x) + \log u$$

$$u \sim \mathcal{U}_{[0,1]} \Rightarrow \log u \sim \text{Exp}(1) \in \mathcal{O}(1)$$

Change to $\log f$ (*energy*) indep. of d!

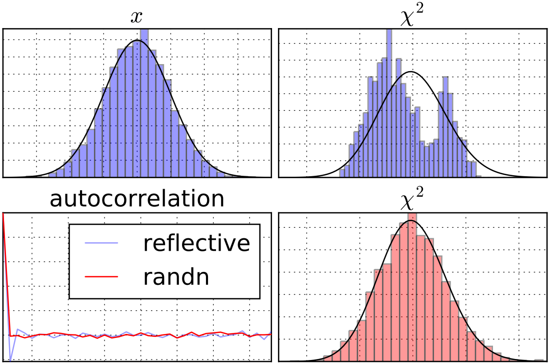
$$f(x) = \prod_{i=1}^{d} \mathcal{N}(x_i \mid 0, 1) \Rightarrow \log f(x) \sim -\frac{1}{2}\chi^2 = -\frac{1}{2}\sum_{i=1}^{d} x_i^2 \in \mathcal{O}(d)$$

independent sampler

$$E[\chi^2] = d, \;\; V[\chi^2] = 2d$$

so fluctuations are $\mathcal{O}\left(\sqrt{d}\right)$. Random walk in 1D needs d steps of $\mathcal{O}(1)$ to travel a distance $\sqrt{d}$

What can we do?



What if we add one or more auxiliary variables to increase fluctuations?

$$f(x) \to f(x)g(y)$$

In HMC: $x \to (x, p)$

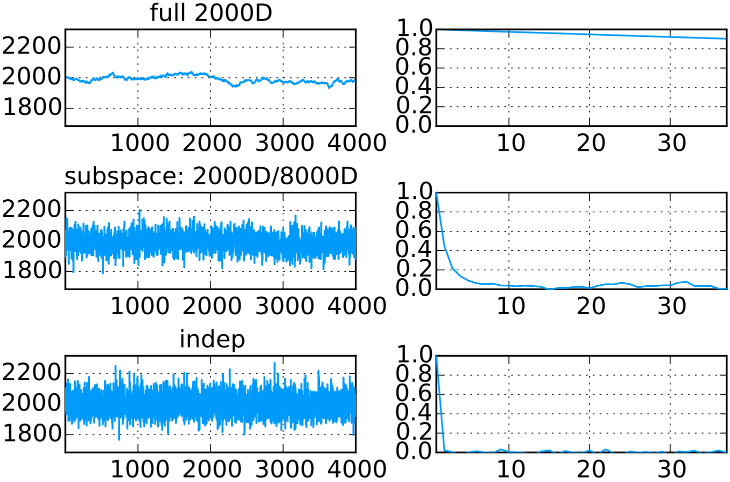
Scaling of densities

Can we get a scalar auxiliary variable $y \sim g(y)$ to scale like $d\ \mathcal{N}(0, 1)$ variables?
I tried invain. Seem prohibited by normalization

$$1 = \int \mathrm{d}y\, g(y),\ \ g(y) \geq 0$$

volume $\propto$ exp(1)
More rigorous derivation?

Lots of auxiliary variables



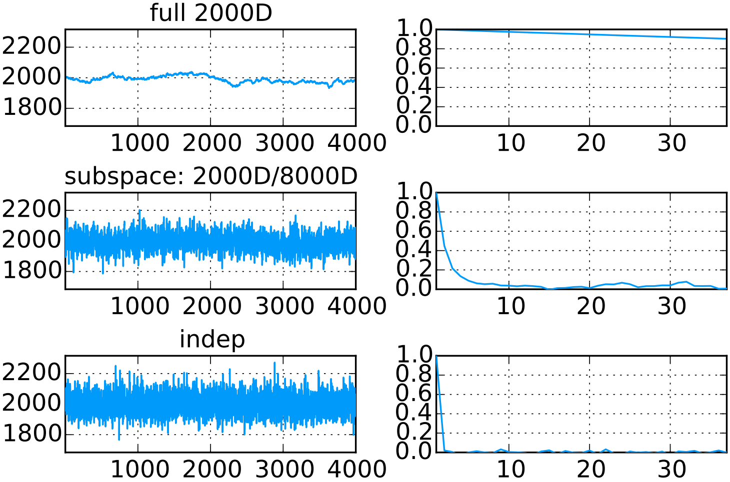
$|\chi^2_n - \chi^2_{n-1}|$: 1 $\rightarrow$ 43 $\rightarrow$ 70

ESS: 0.2% $\rightarrow$ 26% $\rightarrow$ 100%

#bounces: $72k \rightarrow 142k$

For 8x cost, at least 100x more indep. samples

Lots of auxiliary variables



$|\chi^2_n - \chi^2_{n-1}|$: 1 $\rightarrow$ 43 $\rightarrow$ 70

ESS: 0.2% $\rightarrow$ 26% $\rightarrow$ 100%

#bounces: 72k $\rightarrow$ 142k

For 8x cost, at least 100x more indep. samples

Conclusion

- standard Metropolis fails in high dimensions
- auxiliary variables can boost ESS/cost reflective slice sampling dramatically