# Optimal scaling for the Metropolis-Hastings random-walk on elliptically symmetric targets

Chris Sherlock and Gareth Roberts
Lancaster University

## Synopsis

- For all unimodal spherically symmetric targets $\mathbf{X}^{(d)}$ the limiting optimal acceptance rate for a Metropolis-Hastings random walk is $\leq 0.234$.
- The (in)equality is characterised by the limiting behaviour of $\left|\mathbf{X}^{(d)}\right|$.
- This extends to elliptically symmetric targets subject to them not being too anisotropic.

## The Metropolis-Hastings random walk

Given a $d$-dimensional target distribution $\pi(\boldsymbol{x})$ calculatable at any point up to a constant of proportionality, the symmetric Metropolis-Hasting random-walk algorithm constructs a Markov chain with stationary distribution $\pi(.)$ as follows: starting at a point $\boldsymbol{x}$ from the support of $\pi(.)$ and using symmetric proposal density $q(\boldsymbol{y};\lambda) = \lambda^{-d}\, r(\boldsymbol{y}/\lambda)$

1. Propose a jump $\boldsymbol{y}^*$ sampled from $q(.)$.
2. Accept this proposal with probability $\alpha = \min(1, \pi(\boldsymbol{x}+\boldsymbol{y}^*)/\pi(\boldsymbol{x}))$. *If* the proposal is accepted write $\boldsymbol{x} \leftarrow \boldsymbol{x}+\boldsymbol{y}^*$ *else* leave $\boldsymbol{x}$ unchanged.
3. Store $\boldsymbol{x}$ and *goto* 1.

## The Goldilocks Principle

Suppose that the target distribution changes noticeably on a length scale $l$

- If $\lambda << l$ then proposed jumps will usually be accepted but it will take many jumps to explore the main support of the target.
- If $\lambda >> l$ then many proposed updates will lie outside the main support of the target and (probably) be rejected.
- This suggests there may be an optimal scale parameter $\lambda^*$ for the proposal distribution, such that the target is explored as efficiently as possible.

This optimal scaling has been explored for certain classes of $d$ dimensional target as $d \to \infty$. A single component of the random walk is examined with time sped up as dimension increases and the speed of the resulting diffusion is maximised (see for example Roberts et al. (1997) and Roberts and Rosenthal (2002)). For many classes of target (as $d \to \infty$) the scaling is optimised when $23.4\%$ of the proposed jumps are accepted.

## Exact forms in finite dimension

We examine limiting behaviour for a class of targets via exact forms for expected acceptance rate and squared jumping distance at stationarity *in finite dimension* $d$. In the definitions below, $\boldsymbol{X}$ and $\boldsymbol{X}'$ are two consequetive realisations from the chain at equilibrium, and the first expectation is with respect to their joint law. The second expectation is with repect to the joint law of $\boldsymbol{X}$ and the *proposed* next value $\boldsymbol{X}^*$.

$$\begin{aligned} \overline{S}_d &:= \frac{1}{d} E\left[\left|\boldsymbol{X}' - \boldsymbol{X}\right|^2\right] \\ \overline{\alpha}_d &:= E\left[\alpha(\boldsymbol{X}, \boldsymbol{X}^*)\right] \end{aligned}$$

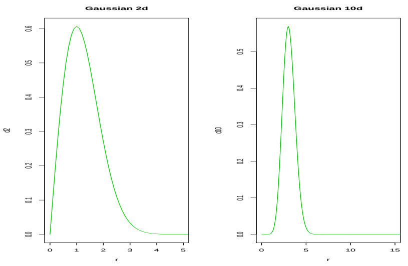
If the target is unimodal and spherically symmetric then

$$\begin{aligned} \overline{\alpha}_d(\lambda) &= 2E_q\left[F_{1|d}\left(-\frac{1}{2}\lambda\left|\mathbf{Y}\right|\right)\right] \\ \overline{S}_d(\lambda) &= \frac{2\lambda^2}{d} E_q\left[\left|\mathbf{Y}\right|^2 F_{1|d}\left(-\frac{1}{2}\lambda\left|\mathbf{Y}\right|\right)\right] \end{aligned}$$

where $F_{1|d}(.)$ is the one dimensional marginal distribution function of the $d$-dimensional target $\mathbf{X}^{(d)}$.

## Spherically symmetric R.V's

The figure shows the marginal radial density (i.e. the density of $|\boldsymbol{X}|$) for spherically symmetric Gaussians in 2 and 10 dimensions.



Many sequences of distributions increasing in dimension show a narrowing of the peak relative to its distance from the origin. We formalise this behaviour via a sequence of rescaling factors $k_x(d)$ such that $\left|\mathbf{X}^{(d)}\right|/k_x(d) \xrightarrow{p} 1$.

e.g. if $\mathbf{X}^{(d)}$ has density $\propto \exp(-\frac{1}{2}\left|\boldsymbol{x}\right|^2)$ then $\left|\mathbf{X}^{(d)}\right|^2 \sim \chi^2_d$ and $\left|\mathbf{X}^{(d)}\right|/d^{1/2} \xrightarrow{p} 1$.

It is straightforward to show that if $\left|\mathbf{X}^{(d)}\right|/k_x(d) \xrightarrow{p} 1$ then any single component of $\mathbf{X}^{(d)}$ satisfies $\frac{d^{1/2}}{k_x(d)} X^{(1|d)} \Rightarrow N(0,1)$.

Thus as $d \to \infty$, $F_{1|d}\left(-\frac{1}{2}\lambda\left|\mathbf{Y}\right|\right)$ becomes $\Phi\left(-\mu\frac{\left|\mathbf{Y}\right|}{k_y(d)}\right)$ where $\mu = \frac{d^{1/2} k_y(d)}{2\ k_x(d)}\lambda$.

Here and in the theorems that follow, $\Phi(.)$ denotes the distribution function of a $N(0,1)$ variable, $\mathbf{X}^{(d)}$ is an element in a sequence of spherically symmetric unimodal targets and $\mathbf{Y}^{(d)}$ is an element in a sequence of proposals.

## Limit results for optimal scaling

**Theorem 1** If $\left|\mathbf{X}^{(d)}\right|/k_x(d) \xrightarrow{p} 1$ and $\left|\mathbf{Y}^{(d)}\right|/k_y(d) \xrightarrow{m.s.} 1$ then

$$\begin{aligned} \overline{\alpha}_d(\mu) &\to 2\Phi(-\mu) & (1) \\ \frac{d^2}{4k_x(d)}\overline{S}_d(\mu) &\to 2\mu^2\Phi(-\mu) & (2) \end{aligned}$$

Differentiating (2) gives the optimum $\mu^* = 1.19$ from which $\lambda^* \sim \frac{2.38}{d^{1/2}}\frac{k_y(d)}{k_x(d)}$.

Substituting into (1) gives the limiting acceptance rate at the optimal scaling as $\overline{\alpha}_\infty := \lim_{d\to\infty}\overline{\alpha}_d(\mu^*) = 0.234$

More generally, an application of Jensen's inequality gives

**Theorem 2** If $\left|\mathbf{X}^{(d)}\right|/k_x(d) \Rightarrow W$ then $\overline{\alpha}_\infty(\mu^*) \leq 0.234$.

## Elliptically symmetric targets

**Theorem 3** For an elliptical target Theorems 1 and 2 still hold provided

$$\frac{\nu^2_{max}}{\sum \nu_i^2} \to 0$$

where $\nu_i$ is the inverse of the scale parameter along the $i^{th}$ axis.

If the elliptical shape of the target were known it would be natural to explore it using a similar elliptically symmetric proposal. Scaling can be optimised for both spherical and elliptical proposals with the elliptical proposal more efficient.

**Theorem 4** As $d \to \infty$ the ratio of the optimal efficiencies is

$$\text{eff}_{ell}/\text{eff}_{sph} = \overline{\nu^2}/\tilde{\nu^2}$$

the ratio of the arithmetic mean of the squares of the $\nu_i$ to their harmonic mean.