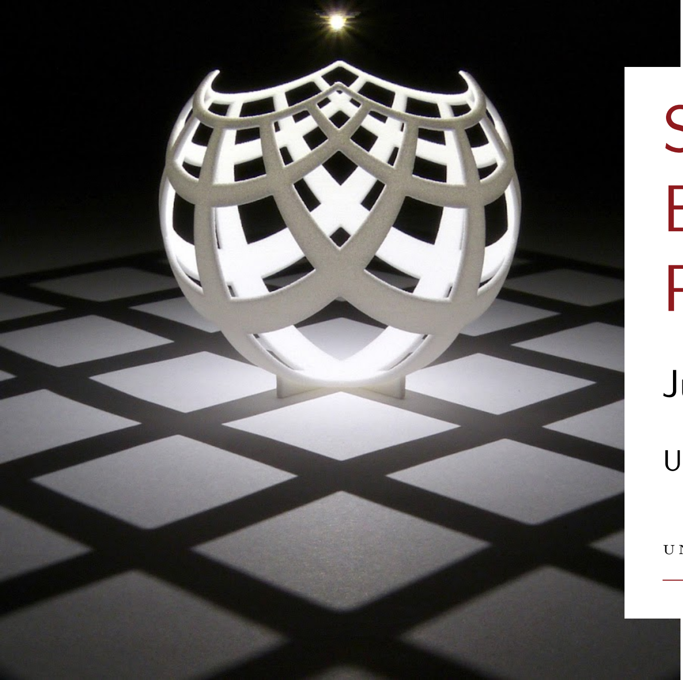


Stereographic Barker's MCMC Proposal

Jun Yang

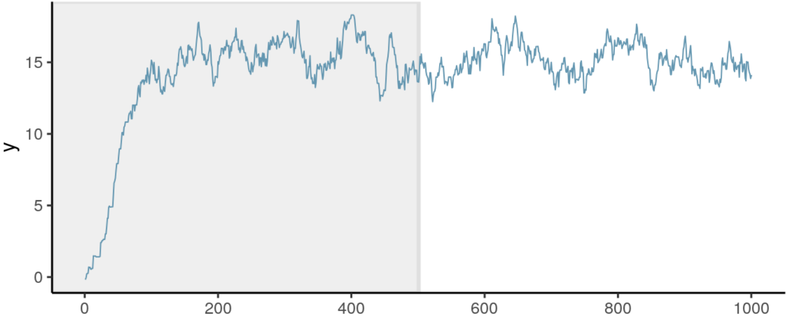
University of Copenhagen

UNIVERSITY OF COPENHAGEN

- *Stereographic Barker's MCMC Proposal: Efficiency and Robustness at Your Disposal*, ongoing work.

Figure: Cameron Bell, Krys Łatuszyński, Gareth Roberts, Jeffrey Rosenthal

Markov Chain Monte Carlo (MCMC)



- Ergodicity: Geometric Ergodicity (or Uniform Ergodicity)
- Scaling with d: Expected Squared Jumping Distance (ESJD)
- Robustness: to tuning parameters and to heavy/super-light tail

Outline

- Examples of gradient-based MCMC:
  - Unadjusted Langevin Algorithm (ULA or LMC)
  - Metropolis-adjusted Langevin Algorithm (MALA)
  - Barker's proposal (Livingstone and Zanella'22)
- Stereographic MCMC
  - Stereographic Projection Sampler (SPS)
  - Stereographic Barker's proposal (NEW!)
- Expected properties (ongoing work)
  - Ergodicity: uniform ergodicity for heavy-tail targets
  - Best scenario: "blessings of dimensionality"
  - Scaling with d: improve from SPS's $\mathcal{O}(d)$ to MALA's $\mathcal{O}(d^{1/3})$
  - Robustness: combine the robustness of SPS and Barker

Unadjusted Langevin Algorithm (ULA)

- Euler-Maruyama discretization of Langevin diffusion

$$X(t+1) = X(t) + \frac{h^2}{2}\nabla \log \pi(X(t)) + Z, \quad Z \sim \mathcal{N}(0, h^2 I_d)$$

- Popular in ML theory community

“bad” properties of ULA (Roberts and Tweedie’96)

- Stationary distribution is not π
- Heavy tail: ULA is not geometric ergodic
- Super-light tail: ULA is transient (oscillation, over-corrects the tail)
- Sensitive to tuning:
  e.g., if π is standard Gaussian then ULA may be oscillating if $h^2 > 2$.

$$X(t+1) = (1 - h^2/2)X(t) + Z, \quad Z \sim \mathcal{N}(0, h^2 I_d)$$

Metropolis-adjusted Langevin Algorithm (MALA)

- ULA as proposal Y, then accept the proposal with prob

$$\alpha(x, y) = \min\left\{1, \frac{\pi(y)}{\pi(x)} \frac{q(y, x)}{q(x, y)}\right\}$$

- Optimal scaling: $\mathcal{O}(d^{1/3})$ (Roberts and Rosenthal'98)

"bad" properties of MALA (Roberts and Tweedie'96)

- Heavy tail: MALA is not geometric ergodic
- Super-light tail: not geometric ergodic if ULA is transient
- Sensitive to tuning (Livingstone and Zanella'22)

Sensitivity to tuning of MALA

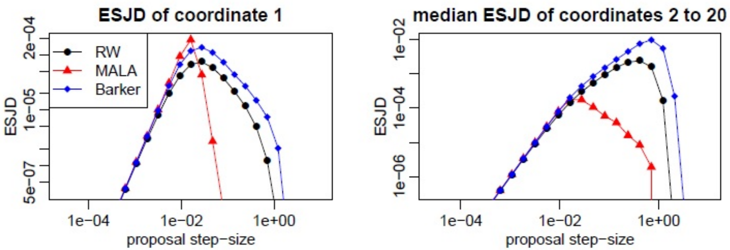


Figure: from (Livingstone and Zanella'22, supplemental material)

"Robust" MALA

MATLA (Roberts and Tweedie'96), tamed MALA (Brosse et.al.'18), Barker's proposal (Livingstone and Zanella'22)

One dimensional Barker's proposal (Livingstone and Zanella'22)

- Sample $Z \sim \mathcal{N}(0, h^2)$
- Proposal $Y = X + Z$ with probability $\frac{1}{1+\exp(-Z(\log \pi(x))')}$
- Proposal $Y = X - Z$ with residual probability.

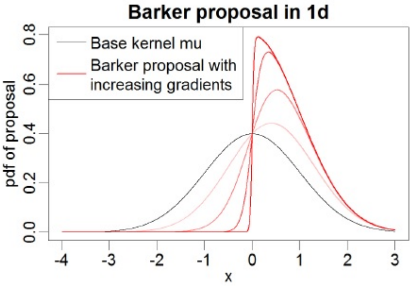
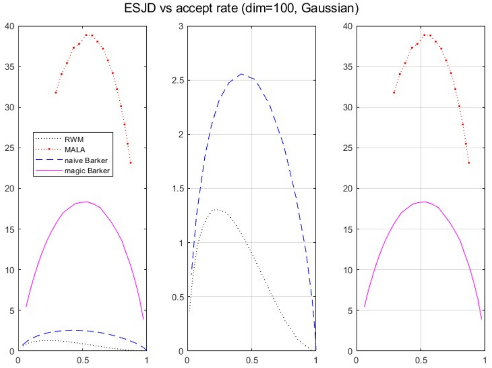


Figure: from (Livingstone and Zanella'22, supplemental material)

- Barker matches ULA "locally" (the first order).

High dimensions: naive Barker

- Sample $Z \sim \mathcal{N}(0, h^2 I_d)$;
- Proposal $Y = X + Z$ with probability $\frac{1}{1+\exp(-Z \cdot \nabla \log \pi(x))}$;
- Proposal $Y = X - Z$ with residual probability.



High dimensions: coordinate-wise Barker

- Sample $Z_i \sim \mathcal{N}(0, h^2)$, $i = 1, \ldots, d$;
- $Y_i = X_i + Z_i$ with probability $\frac{1}{1+\exp(-Z_i \frac{\partial \log \pi(x)}{\partial x_i})}$ for each i;
- $Y_i = X_i - Z_i$ with residual probability.

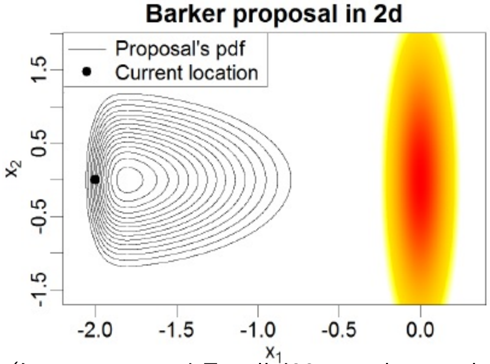


Figure: from (Livingstone and Zanella'22, supplemental material)

Properties of coordinate-wise Barker

- Optimal scaling $\mathcal{O}(d^{1/3})$
  ESJD smaller by a ratio $15^{1/3} \approx 2.47$ (Vogrinc et.al.'23, for Gaussian)
- Heavy tail: not geometrically ergodic
- Robustness to tuning and super-light tail

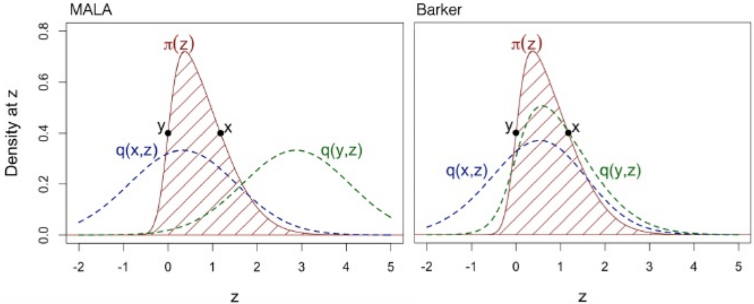
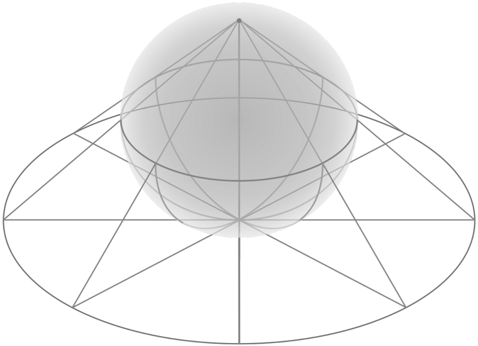


Figure: from (Hird, Livingstone, and Zanella'20)

Stereographic MCMC: map $\mathbb{R}^d$ to $\mathbb{S}^d$

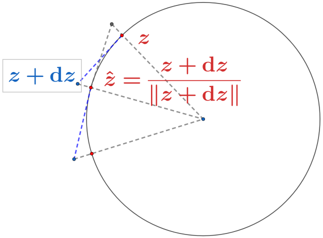
- Y., Łatuszyński and Roberts, *Stereographic Markov Chain Monte Carlo*, AoS 2024.



- Stereographic Projection (bijection: $\mathbb{R}^d \cup \{\infty\} \longleftrightarrow \mathbb{S}^d$).

Stereographic Projection Sampler (SPS)

- Let the current state be $X(t) = x$;
- Compute the proposal Y:
  - Let $z := \mathrm{SP}^{-1}(x)$;
  - Sample independently $\mathrm{d}\tilde{z} \sim \mathcal{N}(0, h^2 I_{d+1})$;
  - Let $\mathrm{d}z := \mathrm{d}\tilde{z} - \frac{(z^T \cdot \mathrm{d}\tilde{z})z}{\|z\|^2}$ and $\hat{z} := \frac{z+\mathrm{d}z}{\|z+\mathrm{d}z\|}$;



  - The proposal $Y := \mathrm{SP}(\hat{z})$.
- $X(t+1) = Y$ with prob. $1 \wedge \frac{\pi(Y)(R^2+\|Y\|^2)^d}{\pi(x)(R^2+\|x\|^2)^d}$; or $X(t+1) = x$.

SPS versus Random-walk Metropolis (RWM)

Ergodicity

- SPS is uniformly ergodic for sub-Cauchy-tail targets (with Grazzi)
- RWM is not geometrically ergodic for any heavy-tailed target.

Convergence Bounds and Optimal Scaling

- SPS: $\mathcal{O}(1) \sim \mathcal{O}(d^2)$ for heavy-tailed targets (with Milinanni)
- Maximum ESJD $\mathcal{O}(d)$: SPS is never worse than RWM

Robustness

- Both RWM and SPS are robust to super-light-tailed targets
- RWM is robust to tuning (stepsize h)
- SPS is even more robust to tuning than RWM (h, radius R, and location of the sphere)

Robustness to tuning of SPS

Theorem (Y., Łatuszyński, and Roberts)

$$\frac{\max_h \mathrm{ESJD}_{\mathrm{SPS}}}{\max_h \mathrm{ESJD}_{\mathrm{RWM}}} = \frac{1}{1-\alpha\cdot\beta\cdot\gamma}, \quad \alpha,\beta,\gamma \in [0,1]$$

- $\alpha = \frac{4\lambda}{(1+\lambda)^2}$ (penalty for misspecified radius $R = \sqrt{\lambda \mathbb{E}_\pi[\|X\|^2]}$);
- $\beta = \frac{\mathrm{Var}(X)}{\mathbb{E}[X^2]}$ (penalty for mislocating the sphere);
- γ distribution-specific penalty ($\gamma = 1$ for isotropy).

The optimal acceptance rate for SPS is also 0.234.

Motivation of Stereographic (coord-wise) Barker

Efficiency and Robustness				
	scaling	heavy tail	super-light tail	robust to tuning
RWM	$\mathcal{O}(d)$	✗	✓	✓
SPS	$\mathcal{O}(d)$	✓	✓	✓
ULA	✗	✗	✗	✗
MALA	$\mathcal{O}(d^{1/3})$	✗	✗	✗
naive Barker	$\mathcal{O}(d)$	✗	✓	✓
coord Barker	$\mathcal{O}(d^{1/3})$	✗	✓	✓

Stereographic gradient-based MCMC

- Stereographic MALA or naive Barker is trivial (and not desirable)

Stereographic gradient-based MCMC

- Stereographic MALA or naive Barker is trivial (and not desirable)

Roadblocks for Stereographic coordinate-wise Barker

- no global coordinate system on sphere
- how to choose a “good” local coordinate system?
- avoid computing basis vectors (matrix multiplication/inversion)

Stereographic gradient-based MCMC

- Stereographic MALA or naive Barker is trivial (and not desirable)

Roadblocks for Stereographic coordinate-wise Barker

- no global coordinate system on sphere
- how to choose a "good" local coordinate system?
- avoid computing basis vectors (matrix multiplication/inversion)

Solution: Givens/Rodrigues' rotation formula

Given n_1 and n_2 are orthonormal, the rotation matrix in d-dimension, which right-hand-rotates by an angle θ in the space spanned by n_1 and n_2, is given by

$$I_d + (n_2 n_1^T - n_1 n_2^T)\sin(\theta) + (n_1 n_1^T + n_2 n_2^T)[\cos(\theta) - 1]$$

Intermediate algorithm: rotate Barker in $\mathbb{R}^d$

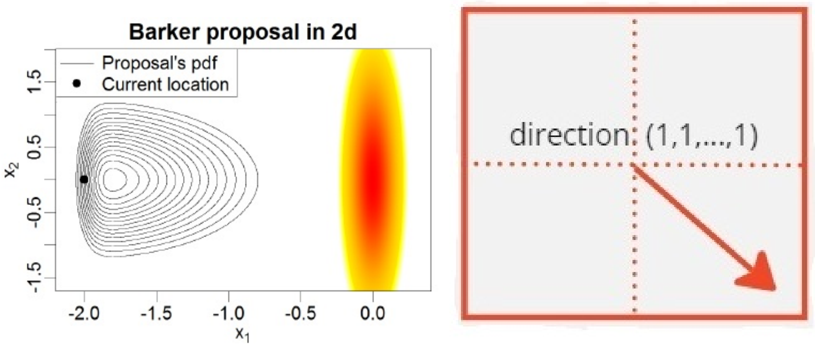


Figure: rotate Barker in $\mathbb{R}^d$ by one Givens rotation

Intermediate algorithm: rotate Barker in $\mathbb{R}^d$

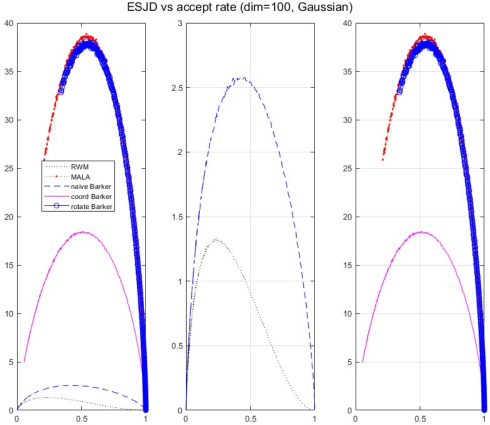


Figure: rotate Barker recovers the maximum ESJD of MALA

Final algorithm: stereo Barker

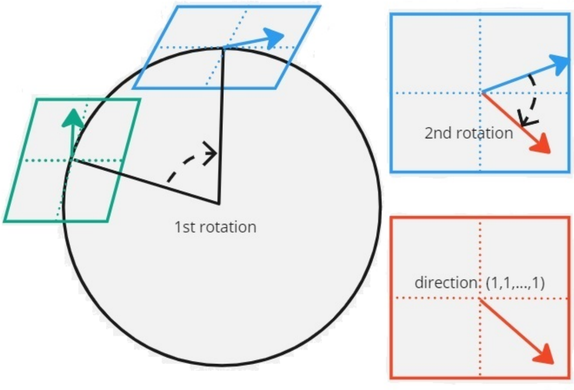


Figure: Stereo Barker by two Givens rotations

Simulation of Stereograhic Barkers and MALA

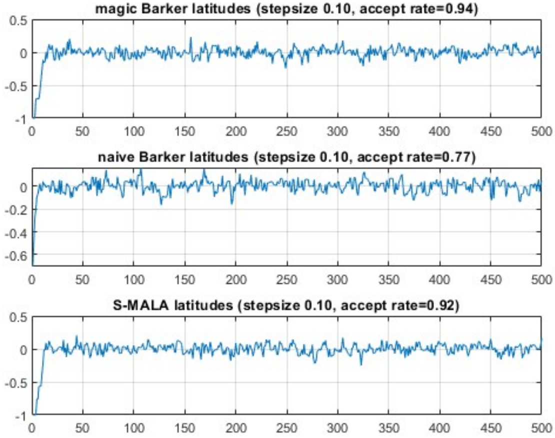


Figure: Starting from South Pole, Gaussian target in 100 dimensions.

Simulation of Stereograhic Barkers and MALA

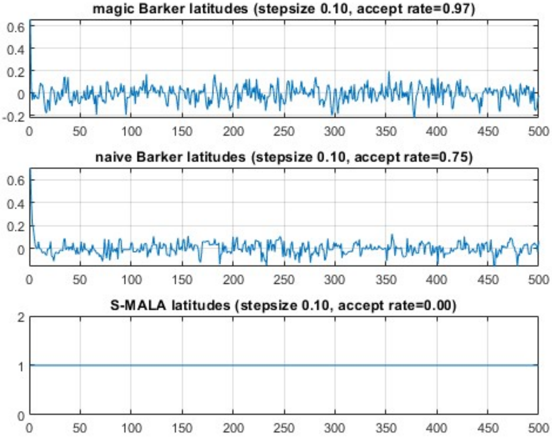


Figure: Starting from nbhd of North Pole, same target.

Summary (including ongoing work)

Efficiency and Robustness				
	scaling	heavy tail	super-light tail	robust to tuning
RWM	$\mathcal{O}(d)$	✗	✓	✓
SPS	$\mathcal{O}(d)$	✓	✓	✓
ULA	✗	✗	✗	✗
MALA	$\mathcal{O}(d^{1/3})$	✗	✗	✗
naive Barker	$\mathcal{O}(d)$	✗	✓	✓
coord Barker	$\mathcal{O}(d^{1/3})$	✗	✓	✓
rotate Barker	$\mathcal{O}(d^{1/3})$	✗	✓	✓
stereo Barker	$\mathcal{O}(d^{1/3})$	✓	✓	✓

Thank you!