Sub-Cauchy Sampling

Escaping the Dark Side of the Moon

Jun Yang

University of Copenhagen, Denmark

BAYSM 2026, June 26-27, Chiba

UNIVERSITY OF COPENHAGEN

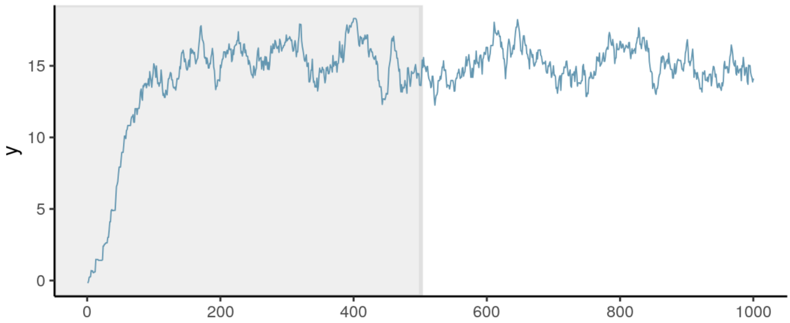
- *Sub-Cauchy Sampling: Escaping the Dark Side of the Moon*, arXiv:2601.11066.



Sebastiano Grazzi (Bocconi) Sifan Liu (Duke) Gareth Roberts (Warwick)

Sebastiano is on the faculty job market in 2026-2027.

Markov Chain Monte Carlo (MCMC)



Today's focus: sampling heavy-tailed distributions

Robust binary (robit) regression: replacing probit/logit by a heavier-tailed link

$$y_i \mid \beta \sim \text{Bernoulli}\{T_\nu(x_i^\top \beta)\}, \quad T_\nu \text{ is the Student-}t \text{ CDF}$$

- More robust to mislabeled or influential binary observations.
- Along some directions, the likelihood decays only polynomially.

Sparse logistic regression with horseshoe prior

$$y_i \mid \beta \sim \text{Bernoulli}\{\text{logit}^{-1}(x_i^\top \beta)\}, \qquad \beta_j \mid \lambda_j, \tau \sim N(0, \tau^2 \lambda_j^2), \quad \lambda_j \sim \text{Half-Cauchy}(0, 1).$$

- The horseshoe prior strongly shrinks noise variables, but has Cauchy-like tails.
- Under separation or weak identification, the likelihood does not kill these tails.

Statistical motivation: heavy-tailed posteriors are not exotic

Why heavy tails?

- Robust regression and outliers
- Binary regression under separation
- Extremes: finance, insurance, climate
- Weakly informative priors

What goes wrong computationally?

- Local proposals struggle in the tails
- HMC/MALA lose geometric ergodicity
- Gibbs samplers are model-specific and can mix slowly

Modeling has moved ahead of general-purpose computation.

In this talk: Sub-Cauchy Sampler (SCS)

SCS: a general-purpose random-walk-Metropolis-type sampler.

Definition

A density π on $\mathbb{R}^d$ is sub-Cauchy if

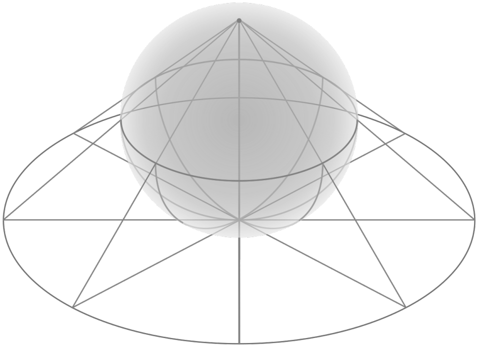
$$\sup_{x \in \mathbb{R}^d} \pi(x)\left(1 + \|x\|^2\right)^{(d+1)/2} < \infty.$$

Theorem

If π is positive, continuous, and sub-Cauchy, SCS is uniformly ergodic.

Previous idea: stereographic MCMC: map $\mathbb{R}^d$ to $\mathbb{S}^d$

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



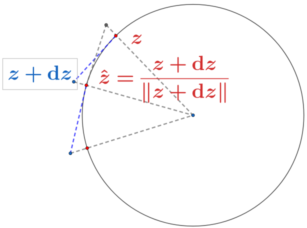
- Stereographic Projection (bijection: $\mathbb{R}^d \cup \{\infty\} \longleftrightarrow \mathbb{S}^d$).

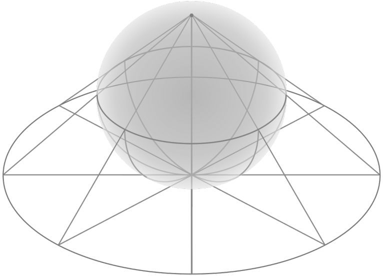
Stereographic Projection Sampler (SPS)

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

$$1 \wedge \frac{\pi(Y)(4+\|(Y-\mu)/R\|^2)^d}{\pi(x)(4+\|(x-\mu)/R\|^2)^d}$$

otherwise $X(t+1) = x$.





SPS versus Random-walk Metropolis (RWM)

Ergodicity

- SPS is uniformly ergodic for a broad class of heavy-tailed targets
- RWM is not geometrically ergodic for any heavy-tailed target [Jarner & Hansen, 2000]

Convergence Bounds, Optimal Scaling, Robustness

- SPS: under warm start, $\mathcal{O}(d)$ for heavy-tailed targets
- Scaling $\mathcal{O}(d)$, max ESJD: SPS is never worse than RWM
- SPS is robust to tuning (stepsize σ, location μ and radius R of the sphere)

But there is a boundary

- Stereographic projection handles tails roughly no heavier than a Student t_d.
- It does not solve the multivariate Cauchy case in dimension d.

Limitation of SPS

Suppose π is positive and continuous on $\mathbb{R}^d$.

- SPS is uniformly ergodic if and only if

$$\sup_{x \in \mathbb{R}^d} \pi(x) \|x\|^{2d} < \infty.$$

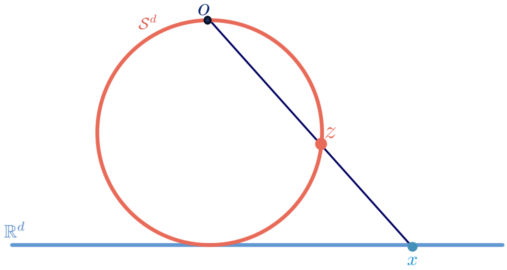
- SPS is not even geometrically ergodic if

$$\sup_{x \in \mathbb{R}^d} \pi(x) \|x\|^{2d} = \infty.$$

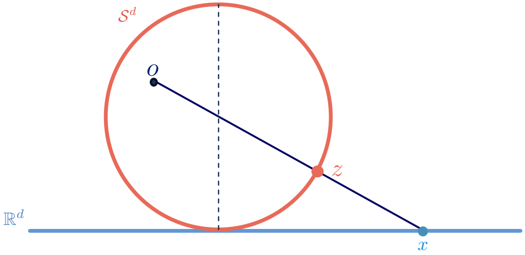
The missing cases: heavier tails than Student t_d

- The transformed density becomes unbounded near the north pole.
- Geometrically: the chain can get trapped close to the point representing infinity.

Key idea: Lower the Anchor



Stereographic projection



Sub-Cauchy projection

Sub-Cauchy Projection

Move the observer to the interior of the sphere. Infinity is mapped to the boundary of a cap.

Sub-Cauchy Projection: geometry

Construction

- Place an observer $o = (h_o, \ell_o)$ inside the sphere.
- The dark side of the moon is the cap above the observer.

The Jacobian is the key difference

$\ell_o = 2$: stereographic projection
dark side degenerates to a point

$$J(x) \propto \left(\|x\|^2 + 4\right)^d$$

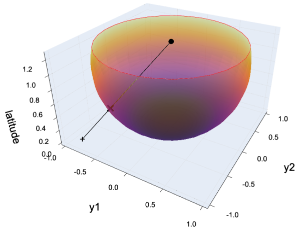
$J(x)^{-1}$ has Student t_d-type decay.

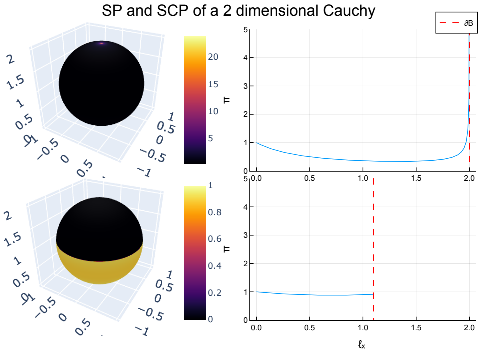
$\ell_o = 1,\ h_o = 0$: Cauchy projection
dark side is a half-sphere

$$J(x) \propto \left(\|x\|^2 + 1\right)^{\frac{d+1}{2}}$$

$J(x)^{-1}$ has Cauchy-type decay.

Sub-Cauchy Projection ($1 \leq \ell_o < 2$)





For any fixed $1 \leq \ell_0 < 2$, $J(y)^{-1}$ has Cauchy-type decay

Sub-Cauchy Projection (SCP): Summary

Parameters and interpretation

Parameter	Role	Practical effect
μ	location	captures the mean
R	scale	matches typical variance
ℓ_o	observer's latitude	controls heavy tail
h_o	horizontal observer	captures skewness

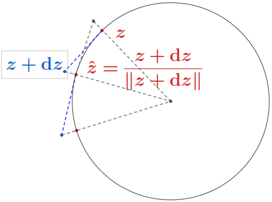
SCP maps $\mathbb{R}^d$ to a spherical cap of a unit sphere (bright side)

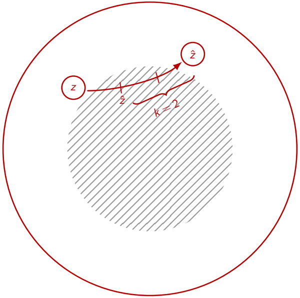
Sub-Cauchy Sampler (SCS)

- Let the current state be $X(t) = x$;
- Compute the proposal Y:
  - Let $z := \mathrm{SCP}^{-1}(x)$;
  - Propose $\hat{z}$ by SPS proposal on sphere
  - If $\hat{z}$ lands in the dark side, escape by stepping-out
  - The proposal $Y := \mathrm{SCP}(\hat{z})$.
- $X(t+1) = Y$ with probability

  $$1 \wedge \frac{\pi(Y)J(Y)}{\pi(x)J(x)}$$

  otherwise $X(t+1) = x$.





Tuning SCP by variational approximation

Fix the observer latitude ℓ_o. Then the bright side $\mathcal{C}_{\ell_o}$ is fixed.

$$z \sim \mathrm{Unif}(\mathcal{C}_{\ell_o}), \qquad X = \mathrm{SCP}_\theta(z), \qquad \theta = (\ell_o, \bar{\theta}), \qquad \bar{\theta} = (h_o, \mu, R)$$

$$q_{\bar{\theta}}(X) = q_{\bar{\theta}}(\mathrm{SCP}_\theta(z)) = \frac{1}{|\mathcal{C}_{\ell_o}| J_\theta(z)} \propto J_\theta(z)^{-1}.$$

Thus, minimizing $\mathrm{KL}(q_{\bar{\theta}} \| \pi)$ over $\bar{\theta}$ is equivalent to

$$\min_{\bar{\theta}} \mathbb{E}_{z \sim \mathrm{Unif}(\mathcal{C}_{\ell_o})} \left[-\log J_\theta(z) - \log \pi(\mathrm{SCP}_\theta(z)) \right].$$

- Variational inference: optimize $\bar{\theta}$ by Monte Carlo + automatic differentiation.
- Keep ℓ_o fixed because it determines the support $\mathcal{C}_{\ell_o}$.

Example: Robust Bayesian binary regression under separation

Model

$$\Pr(Y_i = 1 \mid x_i, \beta) = F(x_i^\top \beta),$$

with heavy-tailed priors on β, and separable data.

- Heavy-tailed priors stabilize inference under separation.

Summary of DA-Gibbs, HMC, and SCS

- DA-Gibbs can mix extremely slowly when the posterior is very heavy-tailed.
- HMC improves in milder cases, but struggles in the extreme tail case.
- SCS gives more reliable tail quantiles across logit and robit examples.

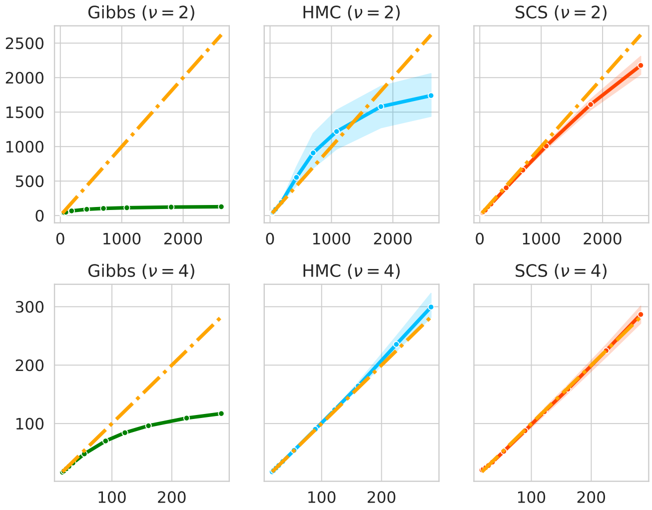


Figure: Q–Q plot of β_1 in logistic regression with independent student's t prior

More talks/posters during ISBA

BAYESM (Chiba), ISBA (Nagoya), ISBA Satellite (Tokyo)

- Zhihao Wang (Copenhagen): Couplings of stereographic MCMC

ISBA (Nagoya), ISBA Satellite (Tokyo)

- Federica Millinani (Northwestern): Rapid mixing of stereographic MCMC

ISBA Satellite Meeting (Tokyo)

- Sebastiano Grazzi (Bocconi): Sub-Cauchy Sampler

Thank you!