

Diffeomorphic Markov Chain Monte Carlo: fast mixing for heavy-tailed distributions

Aleksandar Mijatović
Department of Statistics



Joint work with **Miha Brešar**

August 4, 2026

EPSRC

Engineering and Physical Sciences
Research Council

EP/W006227/1 & EP/V009478/1

Problem: sampling heavy-tailed distributions on $\mathbb{R}^d$

Let $\pi : \mathbb{R}^d \rightarrow \mathbb{R}_+$ proportional to density with polynomial tails $\nu > 0$:

$$\pi(x) \approx |x|^{-\nu-d} \text{ as } |x| \rightarrow \infty \iff \pi(|\cdot| \geq r) \approx r^{-\nu} \text{ as } r \uparrow \infty$$

Objective: Find a fast algorithm $X = (X_n)_{n \in \mathbb{N}}$ (i.e. Markov Chain with $\mathbb{P}_\pi(X_1 \in \cdot) = \pi$) for sampling from π .

Problem: sampling heavy-tailed distributions on $\mathbb{R}^d$

Let $\pi : \mathbb{R}^d \rightarrow \mathbb{R}_+$ proportional to density with polynomial tails $\nu > 0$:

$$\pi(x) \approx |x|^{-\nu-d} \text{ as } |x| \rightarrow \infty \iff \pi(|\cdot| \geq r) \approx r^{-\nu} \text{ as } r \uparrow \infty$$

Objective: Find a fast algorithm $X = (X_n)_{n \in \mathbb{N}}$ (i.e. Markov Chain with $\mathbb{P}_\pi(X_1 \in \cdot) = \pi$) for sampling from π .

Applications: Bayesian inference with heavy-tailed priors (e.g. logistic regression, Cauchy and Student-t priors) [Gel+08; GLM18].

Problem: sampling heavy-tailed distributions on $\mathbb{R}^d$

Let $\pi : \mathbb{R}^d \rightarrow \mathbb{R}_+$ proportional to density with polynomial tails $\nu > 0$:

$$\pi(x) \approx |x|^{-\nu-d} \text{ as } |x| \rightarrow \infty \iff \pi(|\cdot| \geq r) \approx r^{-\nu} \text{ as } r \uparrow \infty$$

Objective: Find a fast algorithm $X = (X_n)_{n \in \mathbb{N}}$ (i.e. Markov Chain with $\mathbb{P}_\pi(X_1 \in \cdot) = \pi$) for sampling from π .

Applications: Bayesian inference with heavy-tailed priors (e.g. logistic regression, Cauchy and Student-t priors) [Gel+08; GLM18].

For standard algorithms (RWM, ULA, MALA, HMC), the rate of convergence is polynomial and slow with sharp lower bounds [BMR25]:

$$c_x n^{-\nu/2} \leq \|\mathbb{P}_x(X_n \in \cdot) - \pi\|_{\text{TV}} \leq C_x n^{-\nu/2} \quad \text{for all } n \in \mathbb{N}$$

Problem: sampling heavy-tailed distributions on $\mathbb{R}^d$

Let $\pi : \mathbb{R}^d \rightarrow \mathbb{R}_+$ proportional to density with polynomial tails $\nu > 0$:

$$\pi(x) \approx |x|^{-\nu-d} \text{ as } |x| \rightarrow \infty \iff \pi(|\cdot| \geq r) \approx r^{-\nu} \text{ as } r \uparrow \infty$$

Objective: Find a **fast** algorithm $X = (X_n)_{n \in \mathbb{N}}$ (i.e. **Markov Chain with** $\mathbb{P}_\pi(X_1 \in \cdot) = \pi$) for sampling from π .

Applications: Bayesian inference with heavy-tailed priors (e.g. logistic regression, Cauchy and Student-t priors) [Gel+08; GLM18].

For standard algorithms (RWM, ULA, MALA, HMC), the rate of convergence is polynomial and slow with sharp **lower bounds** [BMR25]:

$$c_x n^{-\nu/2} \leq \|\mathbb{P}_x(X_n \in \cdot) - \pi\|_{\text{TV}} \leq C_x n^{-\nu/2} \quad \text{for all } n \in \mathbb{N}$$

State-of-the-art No-U-Turn Sampler (NUTS) exhibits the same polynomial slowdown in practice

CLT failure

Objective: estimate $\pi(g) = \int_{\mathbb{R}^d} g(x)\pi(x)dx$ for $g : \mathbb{R}^d \rightarrow \mathbb{R} \in L^1(\pi)$

Run **sampling algorithm** $X = (X_n)_{n \in \mathbb{N}}$ (i.e. **MC** with $\mathbb{P}_\pi(X_1 \in \cdot) = \pi$):

estimator $S_N(g) = \frac{1}{N} \sum_{k=1}^N g(X_k) \longrightarrow \pi(g)$ converges a.s. as $N \uparrow \infty$.

CLT failure

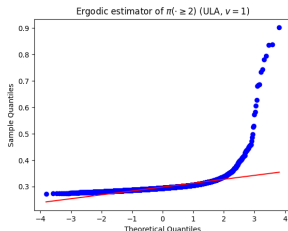
Objective: estimate $\pi(g) = \int_{\mathbb{R}^d} g(x)\pi(x)dx$ for $g : \mathbb{R}^d \rightarrow \mathbb{R} \in L^1(\pi)$

Run **sampling algorithm** $X = (X_n)_{n \in \mathbb{N}}$ (i.e. MC with $\mathbb{P}_\pi(X_1 \in \cdot) = \pi$):

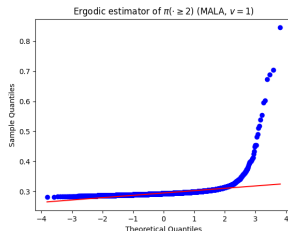
estimator
$$S_N(g) = \frac{1}{N} \sum_{k=1}^N g(X_k) \longrightarrow \pi(g) \quad \text{converges a.s. as } N \uparrow \infty.$$

QQ-plots estimating probability $\pi(\cdot \geq 2) = \pi(g)$ with $g(x) = 1_{|x| \geq 2}$:

$\sqrt{N}(S_N(g) - \pi(g))$ vs **Gaussian distribution**



(a) ULA (Euler scheme for Langevin diffusion)



(b) MALA (ULA with Metropolis correction)

Existing approaches for heavy-tailed targets

Previous attempts to solve the heavy-tailed sampling problem typically rely on **modification of existing methods**:

- ① **Infinite-Variance RWM [JR07; BMR25; He+24]** Replacing Gaussian with α -stable proposals in RWM (with $\alpha \in (0, 2)$)
 - **Result:** Provably faster mixing rate (from $n^{-\nu/2}$ to $n^{-\nu/\alpha}$) (**still polynomial!**)

Existing approaches for heavy-tailed targets

Previous attempts to solve the heavy-tailed sampling problem typically rely on **modification of existing methods**:

- ① **Infinite-Variance RWM [JR07; BMR25; He+24]** Replacing Gaussian with α -stable proposals in RWM (with $\alpha \in (0, 2)$)
 - **Result:** Provably faster mixing rate (from $n^{-\nu/2}$ to $n^{-\nu/\alpha}$) (**still polynomial!**)
- ② **Variable Transformation [JG12]:** Automorphism $\mathbb{R}^d \rightarrow \mathbb{R}^d$ yielding *super-exponential tails*.
 - **Result:** geometric ergodicity, but **fails** to achieve uniform ergodicity.
 - **Limited** to RWM (super-exponential tails destroy geometric ergodicity of MALA)

Existing approaches for heavy-tailed targets

Previous attempts to solve the heavy-tailed sampling problem typically rely on **modification of existing methods**:

- 1 **Infinite-Variance RWM [JR07; BMR25; He+24]** Replacing Gaussian with α -stable proposals in RWM (with $\alpha \in (0, 2)$)
 - **Result**: Provably faster mixing rate (from $n^{-\nu/2}$ to $n^{-\nu/\alpha}$) (**still polynomial!**)
- 2 **Variable Transformation [JG12]**: Automorphism $\mathbb{R}^d \rightarrow \mathbb{R}^d$ yielding *super-exponential tails*.
 - **Result**: geometric ergodicity, but **fails** to achieve uniform ergodicity.
 - **Limited** to RWM (super-exponential tails destroy geometric ergodicity of MALA)
- 3 **Stereographic Projections (SPS) [YLR24] & (SCS) [Gra+26]**: Maps $\mathbb{R}^d$ to punctured sphere $\mathbb{S}^d \setminus \{N\}$ or spherical cap in $\mathbb{R}^{d+1}$.
 - **Result (SPS)**: Uniformly ergodic, but **only if** target π has **d finite moments** [BMR25]
 - **Result (SCS)**: Uniformly ergodic if target has at least **one** finite moment. **Computationally expensive per sample** (numerical examples below)

Introducing: Diffeomorphic Contraction Sampler (DCS)

Core Idea: Transform the heavy-tailed target on $\mathbb{R}^d$ into a well-conditioned law on a **Euclidean ball** $B(R) \subset \mathbb{R}^d$ and sample there.

Introducing: Diffeomorphic Contraction Sampler (DCS)

Core Idea: Transform the heavy-tailed target on $\mathbb{R}^d$ into a well-conditioned law on a **Euclidean ball** $B(R) \subset \mathbb{R}^d$ and sample there.

We define a diffeomorphism $\text{RC}_{\text{total}} : B(R) \rightarrow \mathbb{R}^d$ in **two stages**:

- 1 **Radial Contraction** $\text{RC}_\beta : B(1) \rightarrow \mathbb{R}^d$: compresses arbitrarily heavy polynomial tails into the unit ball $B(1)$;

Introducing: Diffeomorphic Contraction Sampler (DCS)

Core Idea: Transform the heavy-tailed target on $\mathbb{R}^d$ into a well-conditioned law on a **Euclidean ball** $B(R) \subset \mathbb{R}^d$ and sample there.

We define a diffeomorphism $\text{RC}_{\text{total}} : B(R) \rightarrow \mathbb{R}^d$ in **two stages**:

- ① **Radial Contraction** $\text{RC}_\beta : B(1) \rightarrow \mathbb{R}^d$: compresses arbitrarily heavy polynomial tails into the unit ball $B(1)$;
- ② **Automorphism** $\mathcal{T}_\omega : \bar{B}(1) \rightarrow \bar{B}(1)$: geometry-correcting transformation inside the ball to centre and symmetrise the mass.

$$\text{RC}_{\text{total}}(z) := \mu + R \cdot \text{RC}_\beta(\mathcal{T}_\omega(z/R)), \quad z \in B(R)$$

Introducing: Diffeomorphic Contraction Sampler (DCS)

Core Idea: Transform the heavy-tailed target on $\mathbb{R}^d$ into a well-conditioned law on a **Euclidean ball** $B(R) \subset \mathbb{R}^d$ and sample there.

We define a diffeomorphism $\text{RC}_{\text{total}} : B(R) \rightarrow \mathbb{R}^d$ in **two stages**:

- 1 **Radial Contraction** $\text{RC}_\beta : B(1) \rightarrow \mathbb{R}^d$: compresses arbitrarily heavy polynomial tails into the unit ball $B(1)$;
- 2 **Automorphism** $\mathcal{T}_\omega : \bar{B}(1) \rightarrow \bar{B}(1)$: geometry-correcting transformation inside the ball to centre and symmetrise the mass.

$$\text{RC}_{\text{total}}(z) := \mu + R \cdot \text{RC}_\beta(\mathcal{T}_\omega(z/R)), \quad z \in B(R)$$

Why a Ball? Sampling on $B(R)$ has the **best mixing-time bounds** among all convex bodies in $\mathbb{R}^d$ [LV07] (**no corners** for algorithm to get stuck in).

Diffeomorphic Contraction Sampler (DCS): radial map

$$\text{RC}_\beta : B(1) \rightarrow \mathbb{R}^d$$

1. Breaking the Moment Barrier: tame arbitrary polynomial tails via

$$\text{RC}_\beta(y) = \frac{y}{(1 - |y|^\beta)^{1/\beta}}, \quad y \in B(1), \text{ for some } \beta > 0.$$

Intuition behind the radial map RC_β : singularity of its Jacobian at the boundary perfectly counteracts the tails as

$$\frac{J_{\text{RC}_\beta}(y)}{|\text{RC}_\beta(y)|^{d+\beta}} \rightarrow 1 \quad \text{as } |y| \nearrow 1.$$

Diffeomorphic Contraction Sampler (DCS): radial map $\text{RC}_\beta : B(1) \rightarrow \mathbb{R}^d$

1. Breaking the Moment Barrier: tame arbitrary polynomial tails via

$$\text{RC}_\beta(y) = \frac{y}{(1 - |y|^\beta)^{1/\beta}}, \quad y \in B(1), \text{ for some } \beta > 0.$$

Intuition behind the radial map RC_β : singularity of its Jacobian at the boundary perfectly counteracts the tails as

$$\frac{J_{\text{RC}_\beta}(y)}{|\text{RC}_\beta(y)|^{d+\beta}} \rightarrow 1 \quad \text{as } |y| \nearrow 1.$$

Pick parameter $\beta \in (0, \nu]$ smaller than the tail decay rate of the target $\pi(x) \approx |x|^{-d-\nu}$ as $|x| \rightarrow \infty$. Then the pull-back density

$$z \mapsto \pi(\text{RC}_\beta(z)) \cdot J_{\text{RC}_\beta}(z), \quad z \in B(1), \quad \text{is bounded!}$$

- **Result:** (DCS) can accommodate arbitrarily heavy polynomial tails!

DCS: Automorphism of the ball & Variational Inference

2. Balancing the mass

A purely symmetric contraction $\text{RC}_\beta : B(1) \rightarrow \mathbb{R}^d$ does not perform well on its own if the target is highly skewed or off-center. Compose RC_β with automorphism $\mathcal{T}_\omega : \bar{B}(1) \rightarrow \bar{B}(1)$, scaling $R > 0$ & centering $\mu \in \mathbb{R}^d$:

$$\text{RC}_{\text{total}}(z) := \mu + R \cdot \text{RC}_\beta\left(\mathcal{T}_\omega(z/R)\right), \quad z \in B(R)$$

DCS: Automorphism of the ball & Variational Inference

2. Balancing the mass

A purely symmetric contraction $\text{RC}_\beta : B(1) \rightarrow \mathbb{R}^d$ does not perform well on its own if the target is highly skewed or off-center. Compose RC_β with **automorphism** $\mathcal{T}_\omega : \overline{B}(1) \rightarrow \overline{B}(1)$, **scaling** $R > 0$ & **centering** $\mu \in \mathbb{R}^d$:

$$\text{RC}_{\text{total}}(z) := \mu + R \cdot \text{RC}_\beta \left(\mathcal{T}_\omega(z/R) \right), \quad z \in B(R)$$

Variational Inference (VI) preconditioning:

Before sampling, tune $\theta = (\mu, R, \omega)$ by minimizing KL divergence between $\pi_B(z) := \pi(\text{RC}_{\text{total}}(z)) \cdot J_{\text{RC}_{\text{total}}}(z)$ and the **Uniform** on $B(R)$:

$$\arg \min_{\theta} \text{KL} \left(\text{Unif}_{B(R)} \parallel \pi_B \right)$$

This flattens **multi-modality/asymmetry** of π on the compacts in $\mathbb{R}^d$.

DCS Algorithm & Markov kernels on the ball $B(R)$

DCS Iteration ($X_n \rightarrow X_{n+1}$)

Step 1: Map current state to the ball: $\mathbf{z} := \text{RC}_{\text{total}}^{-1}(X_n)$

Step 2: Sample proposal $\hat{\mathbf{Z}}$ from a Markov kernel $Q(\mathbf{z}, \cdot)$ on $B(R)$

Step 3: Map back to $\mathbb{R}^d$: $X_{n+1} = \text{RC}_{\text{total}}(\hat{\mathbf{Z}})$

DCS Algorithm & Markov kernels on the ball $B(R)$

DCS Iteration ($X_n \rightarrow X_{n+1}$)

Step 1: Map current state to the ball: $\mathbf{z} := \text{RC}_{\text{total}}^{-1}(X_n)$

Step 2: Sample proposal $\hat{\mathbf{Z}}$ from a Markov kernel $Q(\mathbf{z}, \cdot)$ on $B(R)$

Step 3: Map back to $\mathbb{R}^d$: $X_{n+1} = \text{RC}_{\text{total}}(\hat{\mathbf{Z}})$

Available kernels Q :

- **DCS-BW (Ball Walk):** Propose $\hat{\mathbf{Y}}$ uniformly from a local ball $\mathbf{z} + \gamma B(1)$, for some $\gamma > 0$. Reject out-of-domain proposals.
Computationally cheap as it ignores flat gradients

DCS Algorithm & Markov kernels on the ball $B(R)$

DCS Iteration ($X_n \rightarrow X_{n+1}$)

Step 1: Map current state to the ball: $\mathbf{z} := \text{RC}_{\text{total}}^{-1}(X_n)$

Step 2: Sample proposal $\hat{\mathbf{Z}}$ from a Markov kernel $Q(\mathbf{z}, \cdot)$ on $B(R)$

Step 3: Map back to $\mathbb{R}^d$: $X_{n+1} = \text{RC}_{\text{total}}(\hat{\mathbf{Z}})$

Available kernels Q :

- **DCS-BW (Ball Walk):** Propose $\hat{\mathbf{Y}}$ uniformly from a local ball $\mathbf{z} + \gamma B(1)$, for some $\gamma > 0$. Reject out-of-domain proposals. Computationally cheap as it ignores flat gradients
- **DCS-HnR (Hit-and-Run):** Sample a random direction $\theta \in \mathbb{S}^{d-1}$. Slice-sample strictly along the chord $\{\mathbf{z} + t\theta \in B(R) : t \in \mathbb{R}\}$ inside $B(R)$. Easily navigates funnel geometries (numerical example below)

DCS Algorithm & Markov kernels on the ball $B(R)$

DCS Iteration ($X_n \rightarrow X_{n+1}$)

Step 1: Map current state to the ball: $\mathbf{z} := \text{RC}_{\text{total}}^{-1}(X_n)$

Step 2: Sample proposal $\hat{\mathbf{Z}}$ from a Markov kernel $Q(\mathbf{z}, \cdot)$ on $B(R)$

Step 3: Map back to $\mathbb{R}^d$: $X_{n+1} = \text{RC}_{\text{total}}(\hat{\mathbf{Z}})$

Available kernels Q :

- **DCS-BW (Ball Walk):** Propose $\hat{\mathbf{Y}}$ uniformly from a local ball $\mathbf{z} + \gamma B(1)$, for some $\gamma > 0$. Reject out-of-domain proposals. Computationally cheap as it ignores flat gradients
- **DCS-HnR (Hit-and-Run):** Sample a random direction $\theta \in \mathbb{S}^{d-1}$. Slice-sample strictly along the chord $\{\mathbf{z} + t\theta \in B(R) : t \in \mathbb{R}\}$ inside $B(R)$. Easily navigates funnel geometries (numerical example below)
- **DCS-DW (Dikin Walk), DCS-RWM (Random Walk Metropolis) & DCS-LMC (Langevin Monte Carlo):** also produce fast-mixing algorithms

Theoretical results: uniform ergodicity

Theorem (uniform ergodicity of DCS)

Let cts. target π be proportional to density on $\mathbb{R}^d$ with tail decay:

$$\limsup_{|x| \rightarrow \infty} \pi(x) |x|^{d+\beta} < \infty \quad \text{for some } \beta > 0.$$

Then, for any choice of R, μ , and smooth automorphism $\mathcal{T}_\omega$, the DCS chain X , based on Hit-and-Run, Ball Walk, RWM, or Dikin Walk, is **uniformly ergodic**: $\exists$ constants $C > 0, \rho \in (0, 1)$ s.t.

$$\|\mathbb{P}_x(X_n \in \cdot) - \pi\|_{\text{TV}} \leq C\rho^n \quad \text{for all } n \in \mathbb{N} \text{ and } x \in \mathbb{R}^d.$$

Theoretical results: uniform ergodicity

Theorem (uniform ergodicity of DCS)

Let cts. target π be proportional to density on $\mathbb{R}^d$ with tail decay:

$$\limsup_{|x| \rightarrow \infty} \pi(x) |x|^{d+\beta} < \infty \quad \text{for some } \beta > 0.$$

Then, for any choice of R, μ , and smooth automorphism $\mathcal{T}_\omega$, the DCS chain X , based on Hit-and-Run, Ball Walk, RWM, or Dikin Walk, is **uniformly ergodic**: $\exists$ constants $C > 0, \rho \in (0, 1)$ s.t.

$$\|\mathbb{P}_x(X_n \in \cdot) - \pi\|_{\text{TV}} \leq C\rho^n \quad \text{for all } n \in \mathbb{N} \text{ and } x \in \mathbb{R}^d.$$

Uniform ergodicity for arbitrarily heavy polynomial tails:

if $\pi(x) \approx |x|^{-\nu-d}$ as $|x| \rightarrow \infty$ for some $\nu > 0$, pick $\beta \in (0, \nu]$.

Theoretical results: non-asymptotic assumptions

For $\varepsilon \in (0, 1)$, *mixing time* $\tau_X(\varepsilon, \varsigma)$ of DCS chain X , started at law ς , is

$$\tau_X(\varepsilon, \varsigma) := \inf \{n \geq 0 : \|\mathbb{P}_\varsigma(X_n \in \cdot) - \nu\|_{\text{TV}} \leq \varepsilon\}.$$

Theoretical results: non-asymptotic assumptions

For $\varepsilon \in (0, 1)$, *mixing time* $\tau_X(\varepsilon, \varsigma)$ of DCS chain X , started at law ς , is

$$\tau_X(\varepsilon, \varsigma) := \inf \{n \geq 0 : \|\mathbb{P}_\varsigma(X_n \in \cdot) - \nu\|_{\text{TV}} \leq \varepsilon\}.$$

Non-asymptotic mixing bounds on $\tau(\varepsilon, \varsigma)$ require preconditioned target

$$\pi_B = \pi \circ \text{RC}_{\text{total}} \cdot J_{\text{RC}_{\text{total}}} \quad \text{on the ball } B(R)$$

(with potential by $U_B = -\log \pi_B$) to satisfy a subset of Assumptions:

- **(Conv)** Convexity: U_B is convex ($\nabla^2 U_B \succeq 0$).
- **(Iso)** C -Isotropy: second-moment matrix of π_B is well-conditioned.
- **(Smooth)** M -Smoothness: gradient ∇U_B is Lipschitz.
- **(Lip)** L -Lipschitz: U_B is Lipschitz.
- **(Warm)** Warm Start: initial law is bounded distance from π_B .

Theoretical results: non-asymptotic assumptions

For $\varepsilon \in (0, 1)$, *mixing time* $\tau_X(\varepsilon, \varsigma)$ of DCS chain X , started at law ς , is

$$\tau_X(\varepsilon, \varsigma) := \inf \{n \geq 0 : \|\mathbb{P}_\varsigma(X_n \in \cdot) - \nu\|_{\text{TV}} \leq \varepsilon\}.$$

Non-asymptotic mixing bounds on $\tau(\varepsilon, \varsigma)$ require preconditioned target

$$\pi_B = \pi \circ \text{RC}_{\text{total}} \cdot J_{\text{RC}_{\text{total}}} \quad \text{on the ball } B(R)$$

(with potential by $U_B = -\log \pi_B$) to satisfy a subset of Assumptions:

- **(Conv)** Convexity: U_B is convex ($\nabla^2 U_B \succeq 0$).
- **(Iso)** C -Isotropy: second-moment matrix of π_B is well-conditioned.
- **(Smooth)** M -Smoothness: gradient ∇U_B is Lipschitz.
- **(Lip)** L -Lipschitz: U_B is Lipschitz.
- **(Warm)** Warm Start: initial law is bounded distance from π_B .

Radial contraction RC_β guarantees asymptotic convexity.

Variational Inference aims to enforce global convexity.

Theoretical results: non-asymptotic mixing-time bounds

DCS inherits mixing-time bounds for convex bodies [LV07; LV06; BEL18].

Theorem (non-asymptotic mixing-time bounds for DCS)

The mixing time of DCS targeting $\pi(x) \approx |x|^{-d-\nu}$ ($|x| \rightarrow \infty$) is $\tau(\varepsilon, \varsigma) = \inf \{n \geq 0 : \|\mathbb{P}_\varsigma(X_n \in \cdot) - \nu\|_{TV} \leq \varepsilon\} = \mathcal{O}^*(d^\alpha), \alpha \in \mathbb{N}$:

Algorithm (on Ball)	Mixing Time	Assumptions on π_B
DCS-BW	$\mathcal{O}^*(d^2)$	(Conv), (Iso), (Warm)
DCS-HnR (Cold Start)	$\mathcal{O}^*(d^3)$	(Conv), (Iso)
DCS-LMC	$\mathcal{O}^*(d^{12})$	(Conv), (Smooth), (Lip)
DCS-Dikin Walk	$\mathcal{O}^*(d^2)$	(Conv), (Smooth), (Warm)

Theoretical results: non-asymptotic mixing-time bounds

DCS inherits mixing-time bounds for convex bodies [LV07; LV06; BEL18].

Theorem (non-asymptotic mixing-time bounds for DCS)

The mixing time of DCS targeting $\pi(x) \approx |x|^{-d-\nu}$ ($|x| \rightarrow \infty$) is $\tau(\varepsilon, \varsigma) = \inf \{n \geq 0 : \|\mathbb{P}_\varsigma(X_n \in \cdot) - \nu\|_{TV} \leq \varepsilon\} = \mathcal{O}^*(d^\alpha), \alpha \in \mathbb{N}$:

Algorithm (on Ball)	Mixing Time	Assumptions on π_B
DCS-BW	$\mathcal{O}^*(d^2)$	(Conv), (Iso), (Warm)
DCS-HnR (Cold Start)	$\mathcal{O}^*(d^3)$	(Conv), (Iso)
DCS-LMC	$\mathcal{O}^*(d^{12})$	(Conv), (Smooth), (Lip)
DCS-Dikin Walk	$\mathcal{O}^*(d^2)$	(Conv), (Smooth), (Warm)

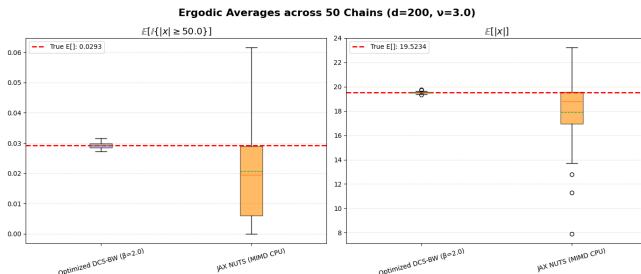
Contrast with standard RWM: Mixing time for standard RWM on heavy-tailed targets exhibits exponential $\exp(d)$ dependence from a cold start [And+26]. DCS-HnR reduces this to a polynomial $\mathcal{O}^*(d^3)$.

Simulations

1. DCS vs No-U-Turn Sampler (NUTS)
2. PosteriorsDB benchmarks
3. DCS vs Stereographic projection samplers

Skewed Student- t : DCS vs NUTS

Target: 200-dimensional skewed Student- t distribution ($\nu = 3$).
Heavy-tailed, skewed target.

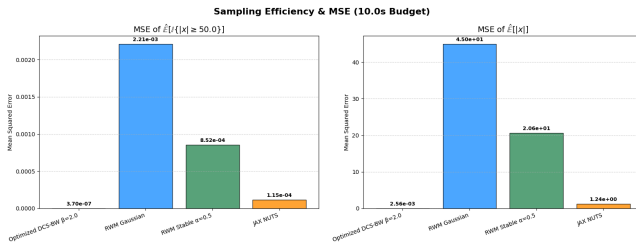


50 independent chains (20s CPU budget).

DCS-BW dramatically outperforms JAX-accelerated NUTS (red dashed lines = ground truth).

Skewed Student- t : NUTS vs RWM vs DCS

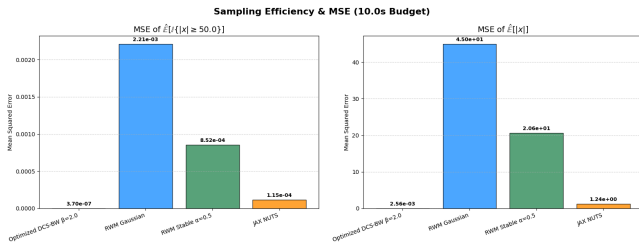
Adding RWM to the mix: NUTS remains state of the art in its category.



Even a naive implementation of DCS-BW **significantly outperforms** the highly optimised JAX NUTS.

Skewed Student- t : NUTS vs RWM vs DCS

Adding RWM to the mix: NUTS remains state of the art in its category.



Even a naive implementation of DCS-BW **significantly outperforms** the highly optimised JAX NUTS.

DCS achieves more than $100\times$ lower MSE compared to NUTS.

Simulations: PosteriorDB (Real-World Models)

Testing on known difficult real-world distributions from **PosteriorDB**:

- **Centred Eight Schools:** Heavy tails (Cauchy prior) + funnel geometry.
- **GARCH(1,1):** Time-varying volatility with a strict hard constraint boundary and a heavy-tailed ridge.

Simulations: PosteriorDB (Real-World Models)

Testing on known difficult real-world distributions from **PosteriorDB**:

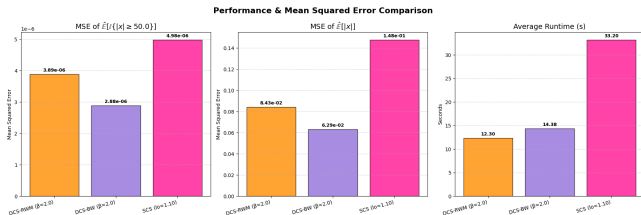
- **Centred Eight Schools:** Heavy tails (Cauchy prior) + funnel geometry.
- **GARCH(1,1):** Time-varying volatility with a strict hard constraint boundary and a heavy-tailed ridge.

Model	1st Moment RMSE		2nd Moment RMSE	
	NUTS	DCS-HnR	NUTS	DCS-HnR
Centred Eight Schools	0.117	0.038	1.814	0.354
GARCH(1,1)	0.0428	0.0125	0.1069	0.0348

The MSE of DCS-HnR is around $3\times$ smaller than NUTS.

Comparison: DCS vs Stereographic (SCS)

Replicating experiments from [Gra+26]. Target: $d = 100$ skewed Student- t $\nu = 3$. We run 20 chains, 500k iterations each.



DCS-BW and DCS-RWM achieve roughly $2\times$ lower MSE while reducing runtime by $3\times$.

Conclusion

Diffeomorphic Contraction Sampler

- Uniform ergodicity for arbitrarily heavy polynomial tails.
- Non-asymptotic mixing-time guarantees: $\mathcal{O}^*(d^3)$ from a cold start.
- Low computational cost: standard samplers on a Euclidean ball.
- Numerically outperforms JAX NUTS and stereographic samplers.

Thank you for your attention!

References

- [And+26] Christophe Andrieu, Anthony Lee, Sam Power, and Andi Q. Wang. "Weak Poincaré inequalities for Markov chains: theory and applications". *Ann. Appl. Probab.* 36.1 (2026), pp. 46–107. ISSN: 1050-5164,2168-8737. DOI: 10.1214/25-AAP2185. URL: <https://doi.org/10.1214/25-AAP2185>.
- [BEL18] Sébastien Bubeck, Ronen Eldan, and Joseph Lehec. "Sampling from a log-concave distribution with projected Langevin Monte Carlo". *Discrete Comput. Geom.* 59.4 (2018), pp. 757–783. ISSN: 0179-5376,1432-0444. DOI: 10.1007/s00454-018-9992-1. URL: <https://doi.org/10.1007/s00454-018-9992-1>.
- [BMR25] Miha Brešar, Aleksandar Mijatović, and Gareth Roberts. "Central Limit Theorem for ergodic averages of Markov chains & the comparison of sampling algorithms for heavy-tailed distributions". *arXiv preprint arXiv:2512.18255* (2025).
- [Gel+08] Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. "A weakly informative default prior distribution for logistic and other regression models". *Ann. Appl. Stat.* 2.4 (2008), pp. 1360–1383. ISSN: 1932-6157. DOI: 10.1214/08-AOAS191. URL: <https://doi.org/10.1214/08-AOAS191>.
- [GLM18] Joyee Ghosh, Yingbo Li, and Robin Mitra. "On the use of Cauchy prior distributions for Bayesian logistic regression". *Bayesian Anal.* 13.2 (2018), pp. 359–383. ISSN: 1936-0975. DOI: 10.1214/17-BA1051. URL: <https://doi.org/10.1214/17-BA1051>.
- [Gra+26] Sebastiano Grazi, Sifan Liu, Gareth O Roberts, and Jun Yang. "Sub-Cauchy Sampling: Escaping the Dark Side of the Moon". *arXiv preprint arXiv:2601.11066* (2026).
- [He+24] Ye He, Alireza Mousavi-Hosseini, Krishna Balasubramanian, and Murat A Erdogdu. "A Separation in Heavy-Tailed Sampling: Gaussian vs. Stable Oracles for Proximal Samplers". *The Thirty-eighth Annual Conference on Neural Information Processing Systems*. 2024.
- [JG12] Leif T. Johnson and Charles J. Geyer. "Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm". *Ann. Statist.* 40.6 (2012), pp. 3050–3076. ISSN: 0090-5364. DOI: 10.1214/12-AOS1048. URL: <https://doi.org/10.1214/12-AOS1048>.
- [JR07] Søren F. Jarner and Gareth O. Roberts. "Convergence of heavy-tailed Monte Carlo Markov chain algorithms". *Scand. J. Statist.* 34.4 (2007), pp. 781–815. ISSN: 0303-6898. DOI: 10.1111/j.1467-9469.2007.00557.x. URL: <https://doi.org/10.1111/j.1467-9469.2007.00557.x>.
- [LV06] László Lovász and Santosh Vempala. "Hit-and-run from a corner". *SIAM J. Comput.* 35.4 (2006), pp. 985–1005. ISSN: 0097-5397,1095-7111. DOI: 10.1137/S009753970544727X. URL: <https://doi.org/10.1137/S009753970544727X>.
- [LV07] László Lovász and Santosh Vempala. "The geometry of logconcave functions and sampling algorithms". *Random Structures Algorithms* 30.3 (2007), pp. 307–358. ISSN: 1042-9832,1098-2418. DOI: 10.1002/rsa.20135. URL: <https://doi.org/10.1002/rsa.20135>.
- [YLR24] Jun Yang, Krzysztof Łatuszyński, and Gareth O. Roberts. "Stereographic Markov chain Monte Carlo". *Ann. Statist.* 52.6 (2024), pp. 2692–2713. ISSN: 0090-5364. DOI: 10.1214/24-aos2426. URL: <https://doi.org/10.1214/24-aos2426>.