A new algorithm on constrained sampling : MAPLA

Yassine Boukhateb, Mohamed Keteb & Imrane Zaakour

Paper : High-accuracy sampling from constrained spaces with the Metropolis-adjusted Preconditioned Langevin Algorithm, (SRINIVASAN; WIBISONO; WILSON, 2025)

May 5, 2025

Summary

Unconstrained Sampling

PROBLEM : Sampling from a density in the case of smooth log-concave density.

Definition

Given an integer $p \geq 2$ and a measurable function $f : \mathbb{R}^p \longrightarrow \mathbb{R}$ such that the integral $\int_{\mathbb{R}^p} \exp\{-f(\theta)\}\, d\theta < \infty$. If f is m-strongly convex with a M-Lipschitz continuous gradient then the density

$$\pi(\theta) \propto \exp\{-f(\theta)\}$$

is a smooth log-concave density.

with f satisfies for all $\theta, \bar{\theta} \in \mathbb{R}^p$

$$\begin{aligned} f(\theta) - f(\bar{\theta}) - \nabla f(\bar{\theta})^\top (\theta - \bar{\theta}) &\geq \frac{m}{2}\|\theta - \bar{\theta}\|_2^2, \\ \|\nabla f(\theta) - \nabla f(\bar{\theta})\|_2 &\leq M\|\theta - \bar{\theta}\|_2. \end{aligned} \tag{1}$$

Langevin Monte Carlo Algorithm (LMC)

- Given $\theta^{(0)} \in \mathbb{R}^p$ random or deterministic
- For $k \geq 0,\ \theta^{(k+1,h)} = \theta^{(k,h)} - h\nabla f(\theta^{(k,h)}) + \sqrt{2h}\xi^{k+1}$ with

Where $h > 0$ is the step-size and $\xi^{(k)} \underset{iid}{\sim} \mathcal{N}_p(0,\ I_p)$ independent of $\theta^{(0)}$

Notice that LMC is a gradient descent with an additional Gaussian random perturbation and also the Euler Discretization of the Langevin Dynamic (LD).

- the target distribution P_π with density $\pi \propto \exp\{-f\}$
- the distribution of $\theta^{(K,h)}$ when h is small and K is large such that $T = Kh$ is large.

An upper bound

Theorem

Let $\theta^\star \in \mathbb{R}^p$ be the global minimum of f and assume that for some $\alpha \geq 1$, we have $h \leq 1/(\alpha M)$ and $K \geq \alpha$. Then, for any time horizon $T = Kh$. The approximation νP_θ^K furnished by the LMC algorithm with the initial distribution $\nu = \mathcal{N}_p(\theta^*, M^{-1} I_p)$ satisfies

$$\|\nu P_\theta^K - P_\pi\|_{TV} \leq \frac{1}{2} \exp\left(\frac{p}{4} \log\left(\frac{M}{m}\right) - \frac{Tm}{2}\right) + \left(\frac{pM^2 Th\alpha}{4(2\alpha - 1)}\right)^{1/2}.$$

GOAL : Same result for constrained sampling

One approach for constrained sampling

See Sampling from a log-concave distribution with Projected Langevin Monte Carlo (BUBECK; ELDAN; LEHEC, 2015)

- Let $\mathcal{K} \subset \mathbb{R}^n$ be a convex body such that $0 \in \mathcal{K}$, $B(0, r) \subset \mathcal{K} \subset B(0, R)$.
- Let $\text{Proj}_{\mathcal{K}}$ denote the Euclidean projection onto K.
- Let $f : \mathcal{K} \to \mathbb{R}$ be an L-Lipschitz and β-smooth convex function, i.e., f is differentiable and satisfies the following conditions: for all $|\nabla f(x)| \leq L$, and $x, y \in K$, $|\nabla f(x) - \nabla f(y)| \leq \beta |x - y|$.

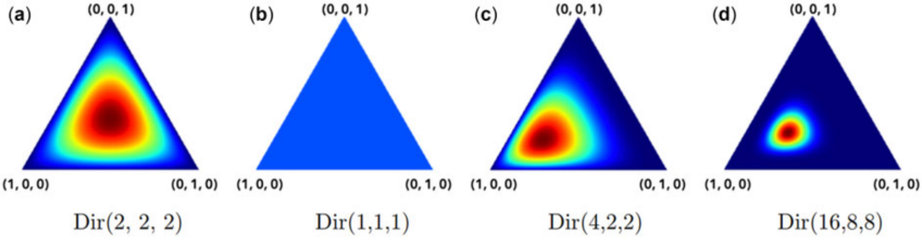
$$X_{k+1} = \text{Proj}_{\mathcal{K}} \left(X_k - \frac{\eta}{2} \nabla f(X_k) + \sqrt{\eta}\, \xi_k \right), \ \eta > 0$$

Dirichlet distribution $\mathcal{D}(a)$

Let $\mathcal{K} := \Delta_{d+1} = \{x \in \mathbb{R}_+^d : \mathbf{1}^\top x \leq 1\}$ and

$$\pi(x) \propto \exp(-f(x)); \quad f(x) = -\sum_{i=1}^{d} a_i \log x_i - a_{d+1} \log\left(1 - \sum_{i=1}^{d} x_i\right) .$$

where $\mathbf{a} \in \mathbb{R}^{d+1}$ and $a_i > 0$,



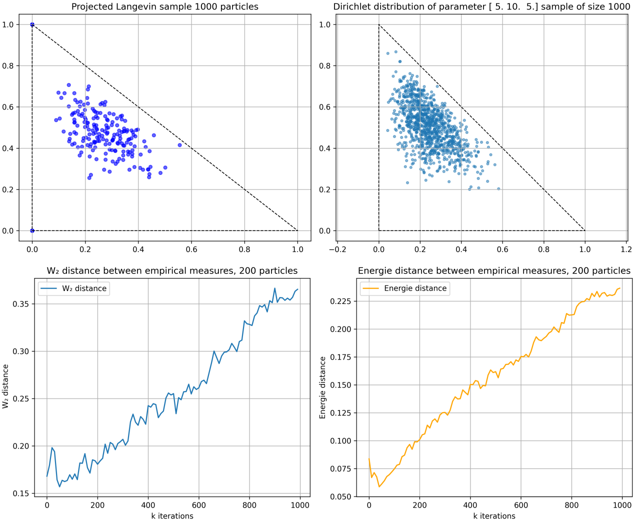
Metrics

$$W_2^2(P, Q) = \inf\{\mathbb{E}[\|X - Y\|^2], X \sim P \text{ and } Y \sim Q\}$$

$$ED(\hat{P}, \hat{Q}) = \frac{2}{N^2} \sum_{i,j=1}^{N} \|X_i - Y_j\| - \frac{1}{N^2} \sum_{i,j=1}^{N} \|X_i - X_j\| - \frac{1}{N^2} \sum_{i,j=1}^{N} \|Y_i - Y_j\|$$

where $\hat{P} = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i}$ and $\hat{Q} = \frac{1}{N} \sum_{i=1}^{N} \delta_{Y_i}$

First case : Corner issue, case of $\mathcal{D}([5, 10, 5])$



Solution : Removing the points on the boundary of the polytope

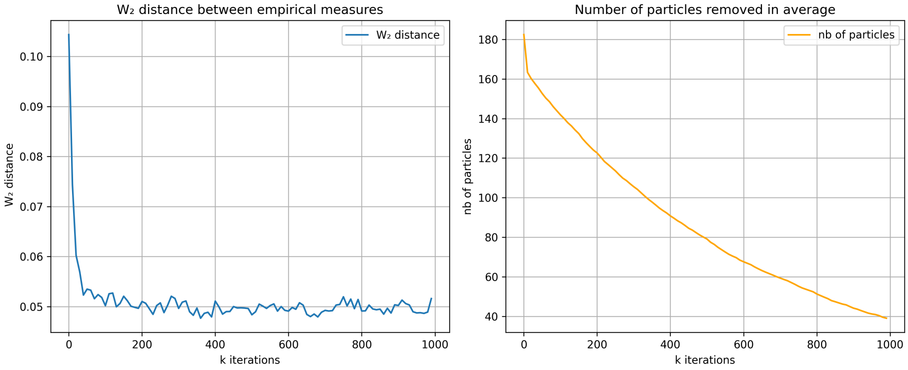


Figure: W_2 distance between the empirical measures (Left) and number of particle removed (Right)

Solution : Metropolis Hasting Adjustment

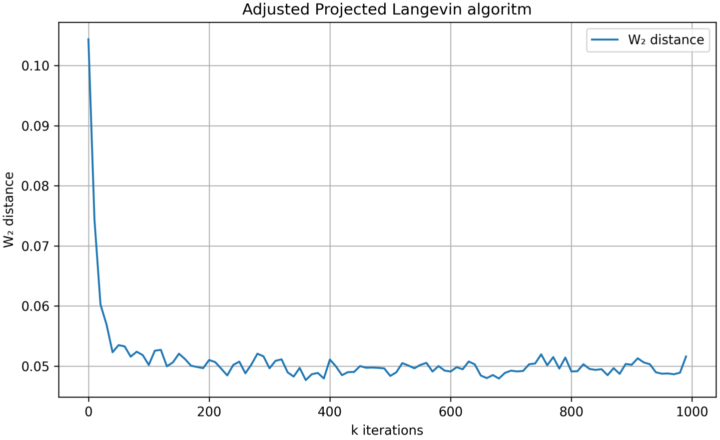
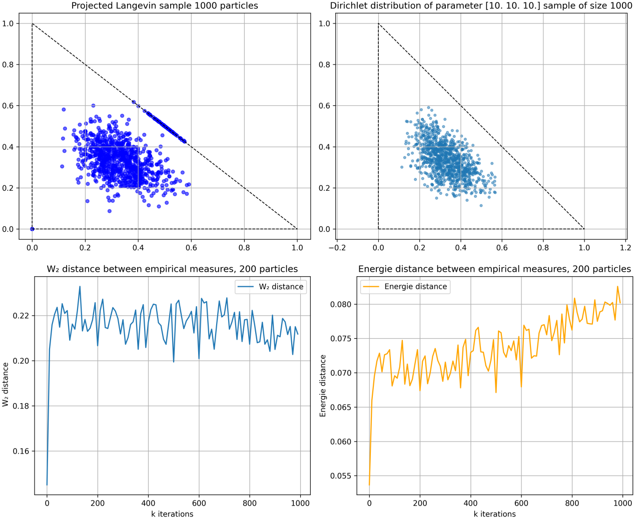


Figure: 200 particles filtered by Metropolis-Hastings and distance to target every 20 iteration

Second case : Diagonal issue, case of $\mathcal{D}([10, 10, 10])$



Solution : Removing the points on the boundary of the polytope

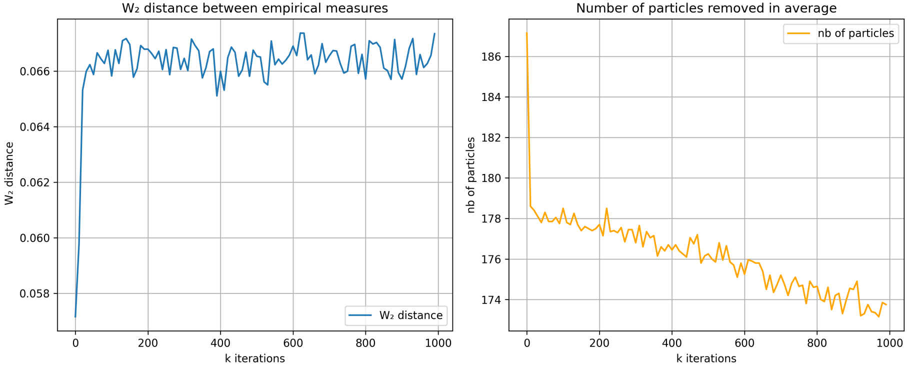


Figure: W_2 distance between the empirical measures (Left) and number of particle removed (Right)

Solution : Metropolis Hasting Adjustment

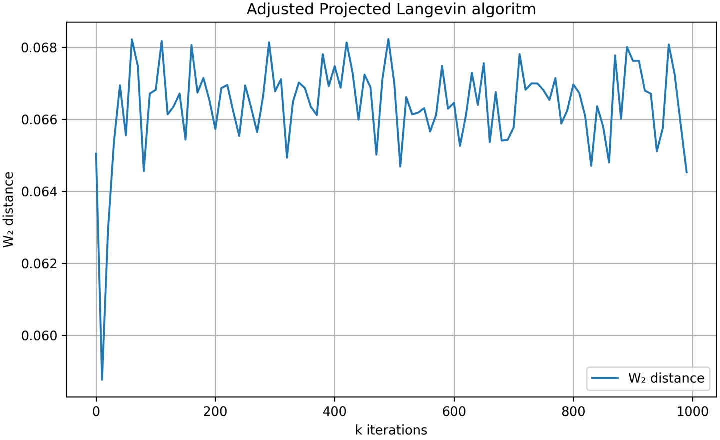


Figure: 200 particles filtered by Metropolis-Hastings and distance to target every 20 iteration

Metropolis-Adjusted Preconditioned Langevin Algorithm (MAPLA)

Idea : The procedure is adjusting the Markov chain of the discrete version (Euler) of the following SDE with Metropolis Hasting filter.
Given a function $\mathscr{G} : \mathrm{int}(\mathcal{K}) \to \mathbb{S}_+^d$ called a metric,

$$\mathrm{d}X_t = \mathfrak{m}(X_t)\mathrm{d}t + \sqrt{2} \cdot \mathscr{G}(X_t)^{-1/2}\,\mathrm{d}B_t \tag{2}$$

$$\mathfrak{m}(X_t) = -\mathscr{G}(X_t)^{-1}\nabla f(X_t) \tag{3}$$

The discretization gives MAPLA

$$X_{k+1} - X_k = -h \cdot \mathscr{G}(X_k)^{-1}\nabla f(X_k) + \sqrt{2h} \cdot \mathscr{G}(X_k)^{-1/2}\,\xi_k \;; \quad \xi_k \sim \mathcal{N}\left(0,\ \mathrm{I}_{d\times d}\right)$$

$(X_k)_{k\geq 0}$ is an homogeneous Markov chain defined by a collection of transition probability $\mathbf{P} = \{\mathcal{P}_x,\ x \in \mathcal{K}\}$, for any $x \in \mathcal{K}$

$$\mathcal{P}_x = \mathcal{N}\left(x - h \cdot \mathscr{G}(x)^{-1}\nabla f(x),\ 2h \cdot \mathscr{G}(x)^{-1}\right)$$

Take this kernel as a proposal in the Metropolis Hasting filter

Related Algorithms

Other proposal distributions $P_X \sim \mathcal{N}(\mathfrak{m}(X), 2h \cdot \mathscr{G}(X)^{-1})$

Algorithm	Mean $\mathfrak{m}(X)$
MAPLA	$X - h \cdot \mathscr{G}(X)^{-1} \nabla f(X)$
ManifoldMALA (*GIROLAMI*; *CALDERHEAD*, 2011)	$X + h \cdot \{(\nabla \cdot \mathscr{G}^{-1})(X) - \mathscr{G}(X)^{-1} \nabla f(X)\}$
Dikin (*KOOK*; *VEMPALA*, 2024)	X

Algorithm

Generate $X_0 \sim \Pi_0$ (initialization)

1. Generate a proposal Z from $\mathcal{P}_X$ where X is current the iterate
2. If $Z \notin \mathcal{K}$, reject Z and set X to be X' (next iterate).
3. If $Z \in \mathcal{K}$, compute the Metropolis-Hastings acceptance probability $p_{X \to Z}$.
4. With probability $p_{X \to Z}$, set $X' = Z$ (accept); otherwise set $X' = X$ (reject).

Mixing Time and Warm start

Definition

We introduce the important indicator of convergence

$$\tau_{\mathrm{mix}}(\delta; \mathbf{P}, \mathcal{C}) \stackrel{\mathrm{def}}{=} \sup_{\pi_0 \in \mathcal{C}} \inf\{k \geq 0 : \mathrm{d}_{\mathrm{TV}}(\mathbb{T}_{\mathbf{P}}^{k} \pi_0, \nu) \leq \delta\}$$

with the operator $\mathbb{T}_{\mathbf{P}}$, $(\mathbb{T}_{\mathbf{P}}\mu)(S) = \int_{\mathcal{K}} \mathcal{P}_y(S) \cdot \mathrm{d}\mu(y)$

How to choose π_0? we choose it be *warm* relative to the target π

- $\mathrm{Warm}(L_\infty, M, \pi) = \{\pi_0 \in \mathcal{P}(\mathcal{K}), \sup_{A \in \mathcal{F}(\mathcal{K})} \frac{\pi_0(A)}{\pi(A)} \leq M\}$
- $\mathrm{Warm}(L_1, M, \pi) = \{\pi_0 \in \mathcal{P}(\mathcal{K}), \left\| \frac{\mu_0}{\nu} \right\|_{L^1(\mu_0)} = M\}$

Now one can wonder what kind of assumption to do on $\mathcal{K}$,$\mathscr{G}$ and f to derive a bound

Assumptions on $\mathscr{G}$: *Self-concordance*

- Properties introduced by Nestrerov in optimization.

The metric $\mathscr{G} : \mathrm{int}(\mathcal{K}) \to \mathbb{S}_+^d$ is ***self-concordant*** if

$$\forall\, x \in \mathrm{int}(\mathcal{K}),\ v \in \mathbb{R}^d, \quad \|\mathscr{G}(x)^{-1/2}\mathrm{D}\mathscr{G}(x)[v]\mathscr{G}(x)^{-1/2}\|_{op} \leq 2 \cdot \|v\|_{\mathscr{G}(x)} \, .$$

The metric is ***strongly self-concordant*** if for all $x \in \mathrm{int}(\mathcal{K})$ and $v \in \mathbb{R}^d$,

$$\|\mathscr{G}(x)^{-1/2}\mathrm{D}\mathscr{G}(x)[v]\mathscr{G}(x)^{-1/2}\|_{\mathrm{F}} \leq 2 \cdot \|v\|_{\mathscr{G}(x)} \, .$$

Assumptions on f : Curvature and Lipschitzness

The potential $f : \mathrm{int}(\mathcal{K}) \to \mathbb{R}$ satisfies a $(\mu, \mathscr{G})$-***curvature lower bound*** where $\mu \geq 0$ if for any $x \in \mathrm{int}(\mathcal{K})$,

$$\nabla^2 f(x) \succeq \mu \cdot \mathscr{G}(x)$$

The potential f satisfies a $(\lambda, \mathscr{G})$-***curvature upper bound*** where $\lambda \geq 0$ if for any $x \in \mathrm{int}(\mathcal{K})$,

$$\nabla^2 f(x) \preceq \lambda \cdot \mathscr{G}(x) \ .$$

the potential f satisfies a $(\beta, \mathscr{G})$-***gradient upper bound*** where $\beta \geq 0$ if for any $x \in \mathrm{int}(\mathcal{K})$,

$$\|\nabla f(x)\|_{\mathscr{G}(x)^{-1}} \leq \beta \ .$$

Assumption on K : symmetry

The metric $\mathscr{G}$ is said to be ***symmetric*** with parameter $\nu \geq 1$ if for any $x \in \mathrm{int}(\mathcal{K})$,

$$\mathcal{E}_x^{\mathscr{G}}(1) \subseteq \mathcal{K} \cap (2x - \mathcal{K}) \subseteq \mathcal{E}_x^{\mathscr{G}}(\sqrt{\nu})$$

With $\mathcal{E}_x^{\mathscr{G}}(r)$ the Dikin ellipsoid of radius r defined as

$$\mathcal{E}_x^{\mathscr{G}}(r) = \{y : \|y - x\|_{\mathscr{G}(x)} < r\}$$

First case : Self-concordant version

Define the quantity $b_{\text{SC}}(d, \lambda, \beta)$

$$b_{\text{SC}}(d, \lambda, \beta) \overset{\text{def}}{=} c_1 \cdot \min\left\{\frac{1}{d^3}, \frac{1}{d \cdot \lambda}, \frac{1}{\beta^2}, \frac{1}{\beta^{2/3}}, \frac{1}{(\beta \cdot \lambda)^{2/3}}\right\}$$

For precision $\delta \in (0, 1/2)$ and warmness parameter $M \geq 1$, if the step size h is bounded as $0 < h \leq b_{\text{SC}}(d, \lambda, \beta)$, $\mathcal{C} \in \{\text{Warm}(L_\infty, M, \pi), \text{Warm}(L_1, M, \pi)\}$ that

$$\tau_{\text{mix}}(\delta; \mathbf{T}, \mathcal{C}) = \frac{c_2}{h} \cdot \max\left\{1, \min\left\{\frac{1}{\widetilde{\mu}^2}, \nu\right\}\right\} \cdot \log\left(\frac{\mathfrak{m}_{\mathcal{C}}}{\delta}\right) ,$$

$$\mathrm{m}_{\mathcal{C}} = \begin{cases} M^{1/2}, & \mathcal{C} = \text{Warm}(L_\infty, M, \pi) \\ M^{1/3}, & \mathcal{C} = \text{Warm}(L_1, M, \pi) \end{cases}, \quad \widetilde{\mu} = \frac{\mu}{8 + 4\sqrt{\mu}}$$

where c_1, c_2 are universal positive constants.

Assumptions on $\mathscr{G}$: *Self-concordance*$_{++}$

The metric is said to be ***lower trace self-concordant*** with parameter $\alpha \geq 0$ if for all $x \in \mathrm{int}(\mathcal{K})$ and $v \in \mathbb{R}^d$,

$$\mathrm{tr}(\mathscr{G}(x)^{-1}\mathrm{D}^2\mathscr{G}(x)[v,v]) \geq -\alpha \cdot \|v\|^2_{\mathscr{G}(x)} .$$

The metric is said to be ***average self-concordant*** if for any $x \in \mathrm{int}(\mathcal{K})$ and $\varepsilon > 0$, there exists $r_\varepsilon > 0$ such that for any $h \in (0, \frac{r_\varepsilon^2}{2d}]$,

$$\mathbb{P}_{\xi \sim \mathcal{N}(x, 2h\cdot\mathscr{G}(x)^{-1})}\left(\|\xi - x\|^2_{\mathscr{G}(\xi)} - \|\xi - x\|^2_{\mathscr{G}(x)} \leq 4h \cdot \varepsilon\right) \geq 1 - \varepsilon$$

self-concordant$_{++}$ = ***strong, lower-trace and average*** *self-concordance*

Summary and main results

Conditions		Mixing time scaling
f is $\begin{cases} (\mu, \mathscr{G})\text{-curv. lower-bdd.} \\ (\lambda, \mathscr{G})\text{-curv. upper-bdd.} \\ (\beta, \mathscr{G})\text{-grad. upper-bdd.} \end{cases}$ &	$\mathscr{G}$ is $\begin{cases} \text{self-concordant} \\ \nu\text{-symmetric} \end{cases}$	$\min\left\{\nu, \frac{1}{\mu}\right\} \cdot \max\{d^3, d\lambda, \beta^2\}$ (Theorem 4.1)
	$\mathscr{G}$ is $\begin{cases} \text{self-concordant}_{++} \\ \nu\text{-symmetric} \end{cases}$	$\min\left\{\nu, \frac{1}{\mu}\right\} \cdot \max\{d\beta, d\lambda, \beta^2\}$ (Theorem 4.2)
f is linear &	$\mathscr{G}$ is $\begin{cases} \text{self-concordant}_{++} \\ \nu\text{-symmetric} \end{cases}$	$\nu \cdot d^2$ (Theorem 4.3)

Elements of proof : One-step overlap technique, Lovász (1999)

Let $s \in (0, 1/2)$. The s-conductance of $\mathbf{T} = \{\mathcal{P}_x : x \in \mathcal{K}\}$ (MAPLA) with stationary distribution π (thanks to MF) supported on $\mathcal{K}$ is defined as

$$\Phi_{\mathbf{T}}^{s} = \inf_{\substack{A \in \mathcal{F}(\mathcal{K}) \\ \nu(A) \in (s, 1-s)}} \frac{1}{\min\{\nu(A) - s,\ 1 - \nu(A) - s\}} \cdot \int_A \mathcal{P}_x(\mathcal{K} \setminus A) \cdot \mathrm{d}\nu(x) \ .$$

The (ordinary) conductance $\Phi_{\mathbf{P}}$ of $\mathbf{P}$ is the limit of $\Phi_{\mathbf{P}}^{s}$ as $s \to 0$.

(see Vempala 2005) If $\pi_0 \in \mathrm{Warm}(L_\infty, M, \pi)$, then

$$\mathrm{d}_{\mathrm{TV}}(\mathbb{T}_{\mathbf{T}}^{k} \pi_0, \pi) \leq \sqrt{M} \cdot \left(1 - \frac{{\Phi_{\mathbf{T}}}^2}{2}\right)^k \leq \sqrt{M} \cdot \exp\left(-\frac{k \cdot {\Phi_{\mathbf{T}}}^2}{2}\right)$$

By setting the right hand side to be less than δ for $\delta \in (0, 1/2)$

$$\tau_{\text{mix}}(\delta; \mathbf{T}, \mathcal{C}) = \frac{2}{{\Phi_{\mathbf{T}}}^2} \cdot \log\left(\frac{\sqrt{M}}{\delta}\right) \leq \frac{C'}{h} \cdot \max\left\{1, \min\left\{\frac{1}{\widetilde{\mu}^2}, \ \nu\right\}\right\} \cdot \log\left(\frac{\sqrt{M}}{\delta}\right)$$

The most important part is to show that

$$\Phi_{\mathbf{T}} \geq C \cdot \sqrt{h} \cdot \min\left\{1, \max\left\{\widetilde{\mu}, \ \frac{1}{\sqrt{\nu}}\right\}\right\} .$$

Sampling from a Dirichlet distribution

Let $\mathcal{K} := \Delta_{d+1} = \{x \in \mathbb{R}_+^d : \mathbf{1}^\top x \leq 1\}$ and

$$\pi(x) \propto \exp(-f(x)); \quad f(x) = -\sum_{i=1}^{d} a_i \log x_i - a_{d+1} \log\left(1 - \sum_{i=1}^{d} x_i\right) .$$

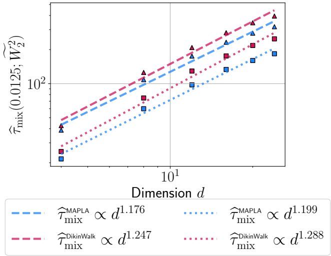
where $\mathbf{a} \in \mathbb{R}^{d+1}$ and $a_i > -1$, f is satisfies $(a_{\min}, \mathscr{G})$-curvature lower bound, $(a_{\max}, \mathscr{G})$-curvature upper bound, and $(\|a\|, \mathscr{G})$-gradient upper bound.

Which metrics $\mathscr{G}$? $\mathscr{G}$ is the norm induced by the Hessian of the log-barrier of the polytope Δ_{d+1} $K = \left\{x \in \mathbb{R}^d : Ax \leq b\right\}$, $\varphi(x) = -\sum_{i=1}^{m} \log\left(b_i - a_i^\top x\right)$

and $\nabla^2 \varphi(x) = \sum_{i=1}^{m} \frac{a_i a_i^\top}{(b_i - a_i^\top x)^2}$

DikinWalk vs MAPLA : Mixing Time

DikinWalk Algorithm (KOOK; VEMPALA, 2024) is the discretization of the process with the drift equal to 0_d



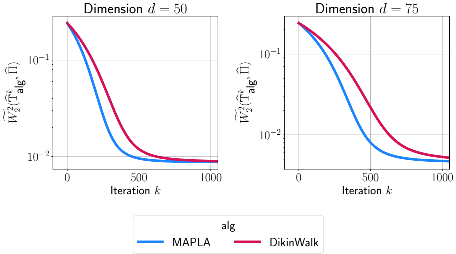


Figure: The dashed and dotted lines correspond to $C_h = 0.1$ and 0.2 respectively. The markers indicate the average over 20 simulations.

DikinWalk vs MAPLA : Acceptance rate

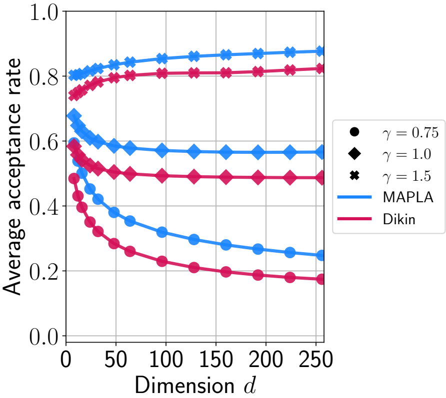


Figure: Variation of $\widehat{R}_{\mathrm{accept}}$ with dimension d when stepsize $h \propto d^{-\gamma}$.

END !

PLMC bound

Assume that $r = 1$ and let $\epsilon > 0$. Then one has

$$d_{\mathrm{TV}}(X_N, \mu) \leq \epsilon$$

provided that $\eta = \widetilde{\mathcal{O}}\left(\frac{R^2}{N}\right)$ and that N satisfies the following: if μ is uniform, then

$$N = \widetilde{\Omega}\left(\frac{R^6 n^7}{\epsilon^8}\right),$$

and otherwise

$$N = \widetilde{\Omega}\left(\frac{R^6 \max(n, RL, R\beta)^{12}}{\epsilon^{12}}\right).$$

MAPLA Algorithm

Algorithm 1: Metropolis-adjusted Preconditioned Langevin Algorithm (MAPLA)

Input : Potential f of Π, convex support $\mathcal{K} \subset \mathbb{R}^d$, metric $\mathscr{G} : \mathrm{int}(\mathcal{K}) \to \mathbb{S}_+^d$, step size $h > 0$, iterations K, initial distribution Π_0

```
Sample x_0 ~ Π_0.
for k ← 0 to K − 1 do
    Sample a random vector ξ_k ~ N(0, I_{d×d}).
    Generate proposal z = x_k − h · G(x_k)^{-1} ∇f(x_k) + sqrt(2h) · G(x_k)^{-1/2} ξ_k.
    if z ∉ K then
        Set x_{k+1} = x_k.
    else
        Compute acceptance ratio p_accept(z; x_k) defined in Eq. (1), where p_y is the density
         of N(y − h · G(y)^{-1} ∇f(y), 2h · G(y)^{-1}).
        Obtain U ~ Unif([0, 1]).
        if U ≤ p_accept(z; x_k) then
            Set x_{k+1} = z.
        else
            Set x_{k+1} = x_k.
        end
    end
end
Output: x_K
```

Références

BUBECK, S.; ELDAN, R.; LEHEC, J. *Sampling from a log-concave distribution with Projected Langevin Monte Carlo*. 2015.

GIROLAMI, M.; CALDERHEAD, B. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley Online Library, v. 73, n. 2, p. 123–214, 2011.

KOOK, Y.; VEMPALA, S. S. Gaussian Cooling and Dikin Walks: The Interior-Point Method for Logconcave Sampling. In: AGRAWAL, S.; ROTH, A. (Ed.). *Proceedings of Thirty Seventh Conference on Learning Theory*. [S.l.]: PMLR, 2024. (Proceedings of Machine Learning Research, v. 247), p. 3137–3240.

SRINIVASAN, V.; WIBISONO, A.; WILSON, A. *High-accuracy sampling from constrained spaces with the Metropolis-adjusted Preconditioned Langevin Algorithm*. 2025.