Contractive kinetic Langevin samplers beyond global Lipschitz continuity

Iosif Lytras[*,◇,c] and Panayotis Mertikopoulos[♯,*]

[*]Archimedes / Athena Research Centre, Athens, Greece
[◇]National Technical University of Athens, Athens, Greece
[♯]Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG, 38000 Grenoble, France
[c]Corresponding author

September 2025

Abstract

In this paper, we examine the problem of sampling from log-concave distributions with (possibly) superlinear gradient growth under kinetic (underdamped) Langevin algorithms. Using a carefully tailored taming scheme, we propose two novel discretizations of the kinetic Langevin SDE, and we show that they are both contractive and satisfy a log-Sobolev inequality. Building on this, we establish a series of non-asymptotic bounds in 2-Wasserstein distance between the law reached by each algorithm and the underlying target measure.

1 Introduction

Consider the problem of sampling from a distribution of the form $\pi \propto \exp(-u)$ where the potential function $u\colon \mathbb{R}^d \to \mathbb{R}$ is known but the normalizing constant isn't. This problem has a wide range of applications in, among others, machine learning, Bayesian inference, physics, and computer science, but the high dimensionality of the sampling space and the geometry of the potential function can make the design of efficient samplers a highly challenging endeavor.

A promising way to overcome (at least partially) the difficulties that arise in this setting is by simulating the associated overdamped Langevin SDE

$$\begin{aligned} \mathrm{d}L_t &= -\nabla u\,(L_t)\,\mathrm{d}t + \sqrt{\frac{2}{\beta}}\,\mathrm{d}B_t, \quad t \geq 0 \\ L_0 &= \theta_0 \end{aligned}$$

and exploiting the fact that this SDE admits the target measure $\exp(-\beta u)$ as its unique invariant measure. Inspired by this observation Langevin-based algorithms have become a popular choice for sampling from distributions in high dimensional spaces. There has been a lot of work in providing non-asymptotic results for Unadjusted Langevin algorithm (ULA) under various assumptions such as Lipschitz continuity of ∇u and convexity of

u. Under the assumption of convexity and gradient Lispchitz continuity important results are obtained in Dalalyan (2017); Durmus and Moulines (2017, 2019); Sabanis and Zhang (2019); Barkhagen et al. (2021), while in the non-convex case, under convexity at infinity or dissipativity assumptions, one may consult Cheng et al. (2018a) Majka et al. (2020), Erdogdu et al. (2022) for ULA while for the Stochastic Gradient variant (SGLD) important works are Raginsky et al. (2017), Chau et al. (2021), Zhang et al. (2023b).
More recently, starting with the work of Vempala and Wibisono (2019) important estimates have been obtained under the assumption that the target measure π satisfies an isoperimetric inequality and the gradient of u satisfies a global Lipschitz continuity, (Mou et al. (2022), Balasubramanian et al. (2022)). The latter assumption has also been relaxed to a weakly smooth assumption (with the gradient still satisfying a linear growth property), see Nguyen et al. (2021) and Erdogdu and Hosseinzadeh (2021).
Recent advances on sampling has led to a connection with optimization (Vempala and Wibisono (2019),Chewi and Stromme (2024)) where overdamped Langevin dynamics can be seen as an analogous to gradient descent. It is therefore natural to explore momentum-like algorithms (nes (1983)) for possible improvements. An example of such a method is the discretization of the underdamped/kinetic Langevin SDE given as

$$\begin{aligned} dV_t &= (-\gamma V_t - \nabla u(\theta_t))\, dt + \sqrt{2\gamma} dB_t \\ d\theta_t &= V_t \end{aligned} \tag{1}$$

where $\gamma > 0$ is the friction coefficient and B_t is a d-dimensional Brownian motion. Similarly to the overdamped Langevin SDE, the discretization of this diffusion can be used as both an MCMC sampler and non-convex optimizer (Gao et al. (2021)) since under appropriate conditions, the Markov process $\left\{ \left(\tilde{\theta}_t, \tilde{V}_t \right) \right\}_{t \geq 0}$ has a unique invariant measure given by

$$\Pi(\mathrm{d}\theta, \mathrm{d}v) \propto \exp\left(- \left(\frac{1}{2}|v|^2 + u(\theta) \right) \right) \mathrm{d}\theta \mathrm{d}v. \tag{2}$$

Consequently, the marginal distribution of (2) in θ is precisely the target measure π. This means that sampling from (2) in the extended space and then keeping the samples in the θ space defines a valid sampler for the density π. The underdamped Langevin MCMC has shown improved convergence rates in the case of convex and gradient Lipschitz smooth potential u, e.g see Cheng et al. (2018b), Dalalyan and Riou-Durand (2020),Monmarché (2021)Altschuler and Chewi (2024). In the non-convex setting important work has been done in Akyildiz and Sabanis (2020), Gao et al. (2018), Chau and Rasonyi (2019),Ma et al. (2021),Zhang et al. (2023a), while there has been a lot of effort in analyzing the contractive behaviour of these algorithms (Monmarché (2021), Leimkuhler et al. (2024)).
All these works assume that the gradient of the potential u grows at most linearly. The current article aims to extend the landscape and allow for super-linear gradient growth aiming to answer the following questions:

- Can we create contractive numerical schemes based on the underdamped Langevin SDE, when the gradient grows superlinearly?
- What are the convergence rates of these schemes?

The current article manages to answer these questions sufficiently by providing novel insights for the contraction rates of modified ULMC in this ill-conditioned scenario, but also provide state of the art convergence results. Our contributions can be summarized as follows:

1. Monotonicity-preserving taming scheme.

   We introduce a novel drift regularization h_λ which is both globally Lipschitz and strongly monotone, while converging to the true gradient ∇u as the stepsize $\lambda \to 0$. This construction extends taming approaches in the sampling literature and provides new technical tools for handling super-linear gradients in kinetic Langevin dynamics.

2. Contractivity of underdamped discretizations.

   Using a weighted Euclidean norm, we prove that both the tamed stochastic exponential scheme and the tamed OBABO scheme are contractive in an almost sure sense. This establishes, for the first time, explicit contraction rates for underdamped numerical schemes beyond the globally Lipschitz regime.

3. Log Sobolev inequality of the underdamped discretizations.
   We are able to prove that our discretizations satisfy uniform in the number of iterations Log-Soboelv inequalities. Establishing show the concentration properties of the algorithms and could possibly have other implications for the use of these algorithms in fields such as differential privacy.

4. Non-asymptotic convergence guarantees.

   We derive quantitative bounds in the 2 -Wasserstein distance between the law of the algorithms and the target distribution. Our results yield iteration complexity

$$O\left(\frac{\log(1/\varepsilon)}{\varepsilon^{2.5}}\right)$$

   to reach an ε-approximation, which improves upon existing bounds for tamed exponential Euler methods and is new for spitting methods in this regime.

2 Preliminaries

2.1 Blanket assumptions

Let $h := \nabla u$. The following assumptions will be in force throughout the article.

Assumption 1. *(strong convexity). There exists an $m > 0$ such that*

$$\langle h(x) - h(y), x - y \rangle \geq 2m|x - y|^2 \quad x, y \in \mathbb{R}^d.$$

Assumption 2. *(Local Lipschitz continuity). There exist $l > 0$ and $L > 0$ such that*

$$|h(x) - h(y)| \leq L(1 + |x| + |y|)^l |x - y| \quad x, y \in \mathbb{R}^d.$$

Assumption 3. *There holds*

$$\mathbb{E}|\theta_0|^{2k} + \mathbb{E}|V_0|^{2k} < \infty \quad \forall k \in \mathbb{N}.$$

2.2 Discretizations of the Kinetic Langevin SDE with general drift coefficient

We consider numerical schemes for the kinetic (underdamped) Langevin SDE

$$\begin{aligned} dV_t &= -\gamma V_t\,dt - \nabla u(\theta_t)\,dt + \sqrt{2\gamma}dB_t, \\ d\theta_t &= V_t\,dt, \end{aligned} \tag{3}$$

where $\gamma > 0$ is the friction coefficient and B_t is a d-dimensional Brownian motion.

In the standard case, the drif coeficient is $b = -\nabla u$. In our setting, we will introduce two schemes with modified drifts $b = h_\lambda$ depending on the step size λ, which is carefully designed to handle super-linear growth of the gradient.

2.2.1 Stochastic exponential scheme

In this section we propose a modified version of the stochastic exponential scheme, an algorithm which was first developed in Cheng et al. (2018b) and analysed in depth in Dalalyan and Riou-Durand (2020) and Gao et al. (2021) under the assumption of Lipschitz continuity for the gradient. The creation of this algorithm is motivated by the fact that (1), after applying Itô's formula to the product $e^{\gamma t}V_t$ and by following standard calculations, can be rewritten as

$$\begin{aligned} V_t &= e^{-\gamma t}V_0 - \int_0^t e^{-\gamma(t-s)}\nabla u(\theta_s)ds + \sqrt{2\gamma}\int_0^t e^{-\gamma(t-s)}dB_s \\ \theta_t &= \theta_0 + \int_0^t V_s ds. \end{aligned} \tag{4}$$

The stochastic exponential scheme is given by

$$\begin{aligned} \bar{Y}^\lambda_{n+1} &= \psi_0(\lambda)\bar{Y}^\lambda_n - \psi_1(\lambda)h_\lambda(\bar{x}^\lambda_n) + \sqrt{2\gamma}\Xi_{n+1} \\ \bar{x}^\lambda_{n+1} &= \bar{x}^\lambda_n + \psi_1(\lambda)v^\lambda_n - \psi_2(\lambda)h_\lambda(\bar{x}^\lambda_n) + \sqrt{2\gamma}\Xi'_{n+1} \end{aligned} \tag{5}$$

with initial condition $(\bar{Y}^\lambda_0, \bar{x}^\lambda_0) = (V_0, \theta_0)$ where $\psi_0(t) = e^{-\gamma t}$ and $\psi_{i+1} = \int_0^t \psi_i(s)ds$. More specifically,

$$\begin{aligned} \psi_0(\lambda) &= e^{-\gamma\lambda}, \\ \psi_1(\lambda) &= \frac{1}{\gamma}(1 - e^{-\gamma\lambda}), \\ \psi_2(\lambda) &= \frac{\lambda\gamma + e^{-\gamma\lambda} - 1}{\gamma^2}. \end{aligned} \tag{6}$$

$\left(\Xi_{k+1}, \Xi'_{k+1}\right)$ is a $2d$-dimensional centered Gaussian vector satisfying the following conditions:

- - $\left(\Xi_j, \Xi'_j\right)'$ s are iid and independent of the initial condition (V_0, θ_0),
- for any fixed j, the random vectors $\left(\Xi_j\right)_1, \left(\Xi'_j\right)_1, \left(\Xi_j\right)_2, \left(\Xi'_j\right)_2, \ldots, \left(\Xi_j\right)_d, \left(\Xi'_j\right)_d$ are iid with covariance matrix

$$\mathbf{C} = \int_0^\lambda [\psi_0(t), \psi_1(t)]^\top [\psi_0(t), \psi_1(t)]\,dt.$$

At this point and in view of (4), one claims that the continuous time interpolation of (5) is given by

$$\begin{aligned}\tilde{Y}_t &= e^{-\gamma(t-n\lambda)}\tilde{Y}_{n\lambda} - \int_{n\lambda}^{t} e^{-\gamma(t-s)} h_\lambda(\tilde{x}_{n\lambda})ds + \sqrt{2\gamma}\int_{n\lambda}^{t} e^{-\gamma(t-s)}dW_s \\ \tilde{x}_t &= \tilde{x}_{n\lambda} + \int_{n\lambda}^{t} \tilde{Y}_s ds,\end{aligned} \tag{7}$$

where $\{W\}_t$ is a d-dimensional Brownian motion. which naturally leads to the following Lemma.

Lemma 1. *Let $\lambda > 0$, $(v_n^\lambda, \bar{x}_n^\lambda)$ be given by (5) and $(\tilde{Y}_{n\lambda}, \tilde{x}_{n\lambda})$ be given by (7). Then,*

$$\mathcal{L}(\bar{Y}_n^\lambda, \bar{x}_n^\lambda) = \mathcal{L}(\tilde{Y}_{n\lambda}, \tilde{x}_{n\lambda}), \quad \forall n \in \mathbb{N}.$$

2.3 OBABO scheme

The kinetic Langevin SDE involves two coupled components: a position evolution (driven by velocity) and a velocity evolution (driven by gradient and friction). Direct discretization (Euler or exponential schemes) often leads to stiff systems and degraded contractive properties. Operator splitting, specifically the OBABO sequence, leverages the separable structure of the SDE, alternately updating momentum and position using analytically tractable sub-steps. This approach mirrors successful strategies in other fields, such as Hamiltonian Monte Carlo, for efficiently integrating stiff or high-dimensional dynamics.
Let $\lambda, \gamma > 0$ be respectively the stepsize and friction coefficient, and denote $\tilde{\eta} = e^{-\lambda\gamma/2}$. We consider the time-homogeneous Markov chain $(x_n, v_n)_{n\in\mathbb{N}}$ on $\mathbb{R}^d \times \mathbb{R}^d$ with transitions given by

$$\left\{\begin{array}{lcl} x_1 &=& x_0 + \lambda\left(\eta v_0 + \sqrt{1-\eta^2}G\right) - \frac{\lambda^2}{2}h_\lambda(x_0) \\ v_1 &=& \eta^2 v_0 - \frac{\lambda\tilde{\eta}}{2}\left(h_\lambda(x_0) + h_\lambda(x_1)\right) + \sqrt{1-\eta^2}\left(\tilde{\eta}G + G'\right), \end{array}\right. \tag{8}$$

where G and G' are two independent standard (mean 0 variance I_d) d-dimensional Gaussian random variables. The scheme can be decomposed as the following successive steps:

$$v_0' = \tilde{\eta}v_0 + \sqrt{1-\tilde{\eta}^2}G \tag{O}$$

$$v_{1/2} = v_0' - \frac{\lambda}{2}h_\lambda(x_0) \tag{B}$$

$$x_1 = x_0 + \lambda v_{1/2} \tag{A}$$

$$v_1' = v_{1/2} - \frac{\lambda}{2}h_\lambda(x_1) \tag{B}$$

$$v_1 = \tilde{\eta}v_1' + \sqrt{1-\tilde{\eta}^2}G'. \tag{O}$$

3 Monotonicity preserving taming scheme

In order to proceed with our analysis we want to create a drift coefficient that is Lipschitz and strongly monotone, and in order to make our scheme a logical candidate to sample from π, our drift should converge to ∇u as λ goes to zero.

Taming schemes have appeared in numerous works in the sampling literature for example see Brosse et al. (2019), Lovas et al. (2023), Lytras and Sabanis (2025), Neufeld et al. (2022),Johnston et al. (2024) Lytras and Mertikopoulos (2025).
The drift coefficient will be designed based on a construction borrowed from the numerics of SDEs literature (Johnston and Sabanis (2024)) , and the preserved monotonicity and Lipschtiz assumption are a major technical contribution compared to these earlier works.

Definition 1. *We define the following quantities:*

$$g(x) = h(x) - mx.$$

$$r_\lambda = (L+m)(\frac{1}{\sqrt{\lambda}})^{\frac{1}{l+2}}.$$

$$R_\lambda = \sup_{B(0,r_\lambda)} |g(x)| \leq (L + |h(0)|) r_\lambda^{l+1}$$

We define

$$g_\lambda(x) = t(x)g(x) + R_\lambda s(x)x$$

where

$$t(x) := \begin{cases} 1, & |x| \leq r_\lambda - 1 \\ r_\lambda - |x|\,, & r_\lambda - 1 < |x| < r_\lambda, \\ 0, & |x| \geq r_\lambda \end{cases}$$

$$s(x) := \begin{cases} 0, & |x| \leq r_\lambda - 2 \\ |x| - r_\lambda + 2, & r_\lambda - 2 < |x| < r_\lambda - 1 \\ 1, & r_\lambda - 1 \leq |x| \leq r \\ \frac{r_\lambda}{|x|}, & |x| \geq r_\lambda \end{cases}$$

Then the tamed drift coefficient is given by $h_\lambda = g_\lambda + mx$

Lemma 2. *The tamed coefficient has the following properties*

- $$h_\lambda(x) = h(x) \quad \forall x \in B(0, r_\lambda - 2)$$
- $$|h_\lambda(x)| \leq \frac{2}{\sqrt{\lambda}}(1 + |x|) \quad \forall x \in \mathbb{R}^d$$
- $$|h_\lambda(x) - h_\lambda(y)| \leq \frac{1}{\sqrt{\lambda}}|x - y| \quad \forall x, y \in \mathbb{R}^d$$
- $$\langle h_\lambda(x) - h_\lambda(y), x - y \rangle \geq m|x - y|^2 \quad \forall x, y \in \mathbb{R}^d$$

Proof. See Appendix □

4 Main results

In order to prove a contraction for the scheme, we shall use a modified norm which is inspired by standard Lyapunov structures for the underdamped Langevin SDE.

Definition 2. *For $z = (x, v) \in \mathbb{R}^{2d}$ we introduce the weighted Euclidean norm*

$$|z|_{a,b}^2 = |x|^2 + 2b\langle x, v\rangle + a|v|^2$$

for $a, b > 0$, which is equivalent to the Euclidean norm on $\mathbb{R}^{2d}$ as long as $b^2 < a$. Under the condition $b^2 < a/4$, we have

$$\frac{1}{2}\min\{1, a\}|z|^2 \leq |z|_{a,b}^2 \leq \frac{3}{2}\max\{1, a\}|z|^2 \tag{9}$$

Let $a = \frac{1}{M_\lambda}$ where $M_\lambda = \mathcal{O}(\lambda^{-\frac{1}{2}})$ is the Lipschitz constant of h_λ and $b = \frac{1}{\gamma}$. For $\gamma \geq \frac{5}{\lambda^{1/4}}$ and $\lambda \leq \frac{1}{2\gamma}$.

Theorem 1. *Let $\gamma \geq 5\sqrt{M_\lambda} = \frac{5}{\lambda^{\frac{1}{4}}}$ and $\lambda \leq \frac{1}{2\gamma}$.*
Let $z_n = (x_n, v_n)$ the $n-th$ iterate of the stochastic exponential tamed scheme with initial condition z_0 and $\tilde{z}_n = (\tilde{x}_n, \tilde{v}_n)$ the n-th iterate of the same scheme with initial condition $\tilde{z}_0$. Then,

$$||z_{n+1} - \tilde{z}_{n+1}||_{a,b}^2 \leq (1 - f(\lambda))|z_n - \tilde{z}_n||_{a,b}^2$$

where $f(\lambda) = \frac{m\lambda}{4\gamma}$.

Theorem 2. *Let $M(x, v) = (x, x+\frac{2}{\gamma}v)$. Assume that for the initial conditions, $\mathcal{L}(M(\theta_0, v_0))$ satisfy a Log-Sobolev inequality with constant $C_{LSI,0}$. Then $\mathcal{L}(M(\theta_n, v_n)$ satisfies an LSI with constant*

$$C_{LSI,n} \leq (1 - \frac{c_0 m}{\gamma}\lambda)^n C_{LSI,0} + \frac{1}{c_0 m} + \mathcal{O}(\lambda\gamma)$$

Theorem 3. *Let (θ_n, V_n) the n-th iterate of the stocahstic exponential tamed scheme with stepsize λ. Then, there holds*

$$W_2(\mathcal{L}(\theta_n), \pi) \leq W_2(\mathcal{L}(\theta_n, V_n), \Pi) \leq 2\sqrt{M_\lambda}(1-f(\lambda))^{n/2}W_2(\mathcal{L}(\theta_0, V_0), \Pi)+C\lambda^2/f(\lambda)$$

where C depends polynomially on the dimension. As a result, to reach ϵ tolerance, one needs $\mathcal{O}(\frac{\log(\epsilon)}{\epsilon^{2.5}})$ iterations.

Theorem 4. *Let $\gamma \geqslant 2\sqrt{M_\lambda}$ and that $\lambda \leqslant m/(33\gamma^3)$. Set*

$$a = 1/M_\lambda, \qquad b = 1/\gamma, \qquad \kappa = m/(3\gamma).$$

Then, $b^2 \leqslant a/4$ and for all $(x_0, v_0), (y_0, w_0) \in \mathbb{R}^{2d}$, if (x_n, v_n) and (y_n, w_n) is the chain in (8) started at (x_0, v_0) and (y_0, w_0) respectively then, almost surely, for all $n \in \mathbb{N}$,

$$\|(x_n, v_n) - (y_n, w_n)\|_{a,b}^2 \leqslant (1 - \lambda\kappa)^n \|(x_0, v_0) - (y_0, w_0)\|_{a,b}^2 .$$

Theorem 5. *Let $S(x,v) = (x + bv, \sqrt{a-b^2}v)$. Let $z_0 = (x_0, v_0)$ be the initial conditions of the algorithm such that $\mathcal{L}(z_0)$ satisfies and LSI with constant $C_{LSI,0}$ Then $\mathcal{L}(S(x_n, v_n))$ satisfies an LSI with*

$$C_{LSI,n} \leq (1-\lambda\kappa)^n \sqrt{1+a} C_{LSI,0} + \frac{1-r^{2n}}{1-r^2} C_1$$

where $r := \sqrt{1-\lambda\kappa}$ and $C_1 = 8\sqrt{1-e^{-2\lambda\gamma}}$

Theorem 6. *Let (x_n, v_n) be the n-th iterate of the OBABO tamed scheme with stepsize λ. Then, we have*

$$\begin{aligned} W_2(\mathcal{L}(x_n), \pi) &\leq W_2(\mathcal{L}(x_n, v_n), \Pi) \\ &\leq 3\sqrt{M_\lambda}(1 - \frac{3m}{\gamma}\lambda)^{n/2} W_2(\mathcal{L}(x_0, v_0), \Pi) + C' M_\lambda \frac{\lambda\gamma}{m} \end{aligned} \tag{10}$$

where C' depends polynomially on the dimension, M_λ is the Lipschitz constant of h_λ $\gamma \geq \sqrt{M_\lambda}$ To reach ϵ needs $\mathcal{O}(\frac{\log(\epsilon)}{\epsilon^{2.5}})$ tolerance.

4.1 Comparison with related literature

The contribution of the current article is twofold. On the one hand, it provides contraction rates for underdamped numerical schemes beyond global Lipschitz continuity of the drift. This is quite novel, as the common way to study these schemes is to study the contraction properties of the continuous type dynamics, which may sometimes be very complicated. For example in Johnston et al. (2024) a very complicated proof was employed which relied on the use of the Moreau-Yosida transform to essentially first sample from the Moreau-Yosida measure and the compare to the original. On the other hand, this work provides the first convergence results for the Velvet-scheme operator in the super-linearly growing case while also considerably improves the rate of convergence for a tamed exponential euler scheme than the one obtained in Johnston et al. (2024). Given the the fact that in many applications the underdamped scheme has been performing better than the overdamped our algorithm can be employed as a viable alternative to other overdamped tamed Langevin algorithms (Brosse et al. (2019),Lytras and Mertikopoulos (2024),Lytras and Sabanis (2025),Lovas et al. (2023),Neufeld et al. (2022)). In addition, another novel contribution of our work is the proof of Log-Sobolev inequalities for the iterates of algorithm. There has been a lot of active research in this domain, as such inequalites are quite important to establish results in differential privacy.

5 Contraction rates for different schemes-Strategy

In order to prove contraction rates for different schemes we will adopt the following notation. Let $z_n = (x_n, v_n)$ the $n-th$ iterate of the scheme with initial condition z_0 and $\tilde{z}_n = (\tilde{x}_n, \tilde{v}_n)$ the n-th iterate of the same scheme with initial condition $\tilde{z}_0$. In order to prove a contraction of the form

$$||z_{n+1} - \tilde{z}_{n+1}||^2_{a,b} \leq (1 - f(\lambda))||z_n - \tilde{z}_n||^2_{a,b} \tag{11}$$

if one sets $G = \begin{pmatrix} I_d & bI_d \\ bI_d & aI_d \end{pmatrix}$ and P being the update rule $(z_{n+1} - \tilde{z}_{n+1}) = P(z_n - \tilde{z}_n)$ (11) is equivalent to showing

$$\mathcal{H} := (1 - f(\lambda))G - P^T GP \succ 0 \tag{12}$$

Since H is symmetric it can take the form

$$\mathcal{H} = \begin{pmatrix} A & B \\ B & C \end{pmatrix}, \tag{13}$$

a key component in our proofs will be the use of the following theorem:

Theorem 7 (Horn and Zhang (2005), Theorem 1.6)**.** *Theorem 1.12 Let A be a Hermitian matrix partitioned as*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{pmatrix}$$

in which A_{11} is square and nonsingular. Then
(a) $A > 0$ if and only if both $A_{11} > 0$ and $A_{22} - A_{12}^* A_{11}^{-1} A_{12} > 0$.
(b) $A \geq 0$ if and only if $A_{11} > 0$ and $A_{22} - A_{12}^* A_{11}^{-1} A_{12} \geq 0$.

Throughout our proof a prominent role plays the control of the quantity $h_\lambda(x) - h_\lambda(y) \quad \forall x, y \in \mathbb{R}^d$. A key observation is the fact that if h_λ is L-Lipschitz and m strongly monotone (with the constant L possibly depending on the stepsize, which will be discussed in the next section) then one obtains $h_\lambda(x) - h_\lambda(y) = Q(x - y)$ for some matrix, where $mI_d \leq Q \leq LI_d$. It is therefore crucial to have construct a drift coefficient which is global Lipschitz smooth but also strongly monotone.

6 Proofs

6.1 Proof of properties of taming scheme

Proof of 2. Proof of (1). Noticing that $t(x) = 1$ and $s(x) = 0 \quad \forall x \in B(0, r_\lambda - 2)$, it follows that

$$h_\lambda(x) = g_\lambda(x) + mx = g(x) + mx = h(x) \quad \forall x \in B(0, r_\lambda - 2).$$

□

Proof of (3). We define $A_1 = \bar{B}(0, r_\lambda - 2)$, $A_2 = \bar{B}(0, r_\lambda - 1)/B(0, r_\lambda - 2)$, $A_3 = \bar{B}(0, r_\lambda)/B_{r_\lambda - 1}$, and $A_4 = B(0, r_\lambda)^c$. Since $g_\lambda = g$, on A_1 we have:

$$|g_\lambda(x) - g_\lambda(y)| \leq Lip_{B_r}|x - y|.$$

For $x, y \in A_2$ $Lip(s(x)) = 1$ and $t(x) = 1$ on A_2 one obtains

$$|g_\lambda(x) - g_\lambda(y)| \leq (Lip_B)_r + R_{r_\lambda}\lambda)|x - y|.$$

For $x, y \in A_3$ since $Lip(t) = 1$ and

$$|g(x)| \leq R_\lambda \quad \forall x \in A_3$$

then,

$$|g_\lambda(x) - g_\lambda(y)| \leq (Lip(g)_{Br_\lambda} + 2R_\lambda)|x-y|$$

For $x, y \in A_4$, since $|x|, |y| \geq r_\lambda$ which implies that $\frac{r^2}{|x||y|} \leq 1$, there holds

$$\begin{aligned}
R_\lambda|x-y|^2 - |g_\lambda(x) - g_\lambda(y)|^2 &= R_\lambda^2|x-y|^2 - \left|R_\lambda r_\lambda \frac{x}{|x|} - R_\lambda r_\lambda \frac{y}{|y|}\right|^2 \\
&= R_\lambda^2\left(|x|^2 + |y|^2 - 2r_\lambda + 2\langle x, y\rangle(\frac{r_\lambda^2}{|x||y|} - 1)\right) \\
&\geq R_\lambda^2(|x|^2 + |y|^2 - 2r_\lambda^2 + 2r_\lambda^2 - 2|x||y|) \\
&\geq R_\lambda^2(|x| - |y|)^2 \\
&\geq 0
\end{aligned}$$

so for each set A_i there holds

$$|g_\lambda(x) - g_\lambda(y)| \leq (Lip(g)_{Br_\lambda} + r_\lambda R_\lambda)|x-y|$$

Let $x, y \in \mathbb{R}^d$ arbitrary. It is easy to see that every straight line l intersects the boundary of of any ball at most twice so that there exist at most $l \leq 6$ points where l intersects any ∂A_i. We denote $l_1 \dots l_{k-1}$ the points of intersection and $l_0 = x$ and $l_k = y$. For every line segment $[l_j, l_{j+1}]$ since A_i cover all $\mathbb{R}^d$, every point in $[l_j, l_{j+1}]$ belong at least in one A_i. If the two endpoints belong in different sets by the continuity of line segment there would exists a point $x_0 \in [l_j, l_{j+1}]$ that belongs in the intersection which is not possible. So both endpoints belong in the same A_i so since A_i is closed the whole segment belongs in an A_i. As a result, using a triangle and the Lipschitzness in each A_i one obtains

$$\begin{aligned}
|g_\lambda(x) - g_\lambda(y)| &= |g_\lambda(l_0) - g_\lambda(l_k)| \\
&\leq \sum_{i=1}^{k} |g(l_{i-1}) - g(l_i)| \\
&\leq (Lip_{B_r} + R_\lambda r_\lambda) sum_{i=1}^{k} |l_i - l_{i-1}| \\
&= (Lip_{B_r} + R_\lambda r_\lambda)|x-y|
\end{aligned}$$

where the last line holds since the segments are collinear. □

Proof of (4). In order to show that g_λ is monotone, we will first prove it is monotone in each A_i. Let $x, y \in A_1$. Then, it is trivial due to the monotonicity of g. Let $x, y \in A_2$. We assume that $|x| \geq |y|$ so $s(x) \geq s(y)$. Then,

$$\begin{aligned}
\langle g_\lambda(x) - g_\lambda(y), x-y\rangle &= \langle g(x) - g(y), x-y\rangle + R_\lambda\langle s(x)x - s(y)y, x-y\rangle \\
&\geq R_\lambda s(x)|x|^2 + s(y)|y|^2 - (s(x) + s(y))\langle x, x-y\rangle \\
&\geq R_\lambda(s(x)|x| - s(y)|y|)(|x| - |y|) \\
&\geq 0.
\end{aligned}$$

For $x, y \in A_3$,

$$\begin{aligned}
\langle g_\lambda(x) - g_\lambda(y), x-y\rangle &= R_\lambda|x-y|^2 + t(y)\langle g(x) - g(y), x-y\rangle \\
&\quad + (t(x) - t(y))\rangle g(x), x-y\rangle \\
&\geq R_\lambda|x-y|^2 - ||x| - |y|||g(x)||x-y| \\
&\geq (R - |g(x)|)|x-y|^2 \geq 0
\end{aligned}$$

For $x, y \in A_4$,

$$\langle g_\lambda(x) - g_\lambda(y), x - y\rangle = R_\lambda r_\lambda \langle \frac{x}{|x|} - \frac{y}{|y|}, x - y\rangle \geq 0$$

To complete the proof for arbitrary x, y using the l_i notation of the previous proof then l_i are collinear and two consecutive belong to the same A_i so

$$\langle g_\lambda(x) - g_\lambda(y), x - y\rangle = \sum_{i=1}^{k} \langle g_\lambda(l_i) - g_\lambda(li - 1), l_i - l_{i-1}\rangle \geq 0. \qquad \square$$

This proves our claim and completes our proof. $\square$

6.2 Proof of Contraction rates of the Monotonic polygonal Stochastic exponential scheme

In the following section we will prove that our scheme has a contractive behaviour a.s in the weighted euclidean norm with $a = \frac{1}{M_{\lambda,Lip}}$, $b = \frac{1}{\gamma}$.

Proposition 1. *Let $\gamma \geq 5\sqrt{M_\lambda} = \frac{5}{\lambda^{\frac{1}{4}}}$ and $\lambda \leq \frac{1}{2\gamma}$.*
Let $z_n = (x_n, v_n)$ the $n - th$ iterate of the stochastic exponential tamed scheme with initial condition z_0 and $\tilde{z}_n = (\tilde{x}_n, \tilde{v}_n)$ the n-th iterate of the same scheme with initial condition $\tilde{z}_0$. Then,

$$||z_{n+1} - \tilde{z}_{n+1}||^2_{a,b} \leq (1 - f(\lambda))|z_n - \tilde{z}_n||^2_{a,b}$$

where $f(\lambda) = \frac{m\lambda}{4\gamma}$.

Proof of contraction rates for exponential scheme. Let $\bar{x}_j := (\tilde{x}_j - x_j)$, $\bar{v}_j = (\tilde{v}_j - v_j)$ and $\bar{z}_j = (\bar{x}_j, \bar{v}_j)$. Let $\eta = e^{-\lambda\gamma}$.
Since h_λ is Lipschitz its Hessian exists almost everywhere and one can write :

$$h_\lambda(\tilde{x}_n) - h_\lambda(x_n) = Q(\bar{x} + \lambda\bar{v})$$

where $Q = \int_0^1 \nabla h_\lambda(\tilde{x}_n + t(x_n - \tilde{x}_n))\, dt$.
The process $\bar{z}_n$ evolves as following:

$$\bar{x}_{n+1} = \bar{x}_n + \frac{1-\eta}{\gamma}\bar{v}_n - \frac{\gamma\lambda + \eta - 1}{\gamma^2} Q\bar{x}_n, \quad \bar{v}_{n+1} = \eta\bar{v}_n - \frac{1-\eta}{\gamma} Q\bar{x}_n,$$

Following the notation of (13) the matrix has the following components:

$$\begin{aligned}
A &= -f(\lambda) I_d + 2\left(\frac{b(1-\eta)}{\gamma} + \frac{\eta - 1 + \gamma\lambda}{\gamma^2}\right) Q \\
&\quad - \left(\frac{a(1-\eta)^2}{\gamma^2} + \frac{2b(1-\eta)(-1+\eta+\gamma\lambda)}{\gamma^3} + \frac{(-1+\eta+\gamma\lambda)^2}{\gamma^4}\right) Q^2 \\
B &= \left(b(1-\eta) - \frac{(1-\eta)}{\gamma} - bf(\lambda)\right) I_d \\
&\quad + \left(\frac{a\eta(1-\eta)}{\gamma} + \frac{b(1-\eta)^2}{\gamma^2} + \frac{b\eta(-1+\eta+\gamma\lambda)}{\gamma^2} + \frac{(1-\eta)(-1+\eta+\gamma\lambda)}{\gamma^3}\right) Q \\
C &= \left(a\left(1-\eta^2\right) - af(\lambda) - \frac{2b\eta(1-\eta)}{\gamma} - \frac{(1-\eta)^2}{\gamma^2}\right) I_d
\end{aligned}$$

In order to use theorem 7 we need to show that A is positive definite.

Lemma 3. *A is positive definite.*

Let eig the eigenvalues of Q. Since $A = R(Q)$ for some polynomial R , the eigenvalues of A are given as:

$$\begin{aligned} \mathrm{eig}(A) &= -f(\lambda) + 2\left(\frac{b(1-\eta)}{\gamma} + \frac{\eta - 1 + \gamma\lambda}{\gamma^2}\right)\mathrm{eig}(Q) \\ &\quad - \left(\frac{a(1-\eta)^2}{\gamma^2} + \frac{2b(1-\eta)(-1+\eta+\gamma\lambda)}{\gamma^3} + \frac{(-1+\eta+\gamma\lambda)^2}{\gamma^4}\right)\mathrm{eig}(Q^2) \end{aligned} \tag{14}$$

By the definition of Q and the fact that h_λ is $m-$ strongly monotone and M_λ-Lipschitz one deduces

$$\begin{aligned} \mathrm{eig}(A) &= -f(\lambda) + 2\left(\frac{b(1-\eta)}{\gamma} + \frac{\eta - 1 + \gamma\lambda}{\gamma^2}\right)m \\ &\quad - \left(\frac{a(1-\eta)^2}{\gamma^2} + \frac{2b(1-\eta)(-1+\eta+\gamma\lambda)}{\gamma^3} + \frac{(-1+\eta+\gamma\lambda)^2}{\gamma^4}\right)M_\lambda^2 \end{aligned}$$

Using (37) and (38), and setting $a = \frac{1}{M_\lambda}$, $b = \frac{1}{\gamma}$, $f(\lambda) = \frac{m\lambda}{4\gamma}$. one notices that since $\mathrm{eig}(Q) > 0$,

$$-f(\lambda) + 2\left(\frac{b(1-\eta)}{\gamma} + \frac{\eta - 1 + \gamma\lambda}{\gamma^2}\right)\mathrm{eig}(Q) \geq \frac{7}{8}\frac{\lambda\,\mathrm{eig}(Q)}{\gamma}$$

$$-\frac{a(1-\eta)^2}{\gamma^2}\mathrm{eig}(Q^2) \geq -\frac{(1-\eta)^2}{\gamma^3}\mathrm{eig}(Q)^2 \geq -\frac{\lambda^2}{\gamma}\mathrm{eig}(Q)^2$$

$$\begin{aligned} -\left(\frac{2b(1-\eta)(-1+\eta+\gamma\lambda)}{\gamma^3} + \frac{(\eta+\gamma\lambda-1)^2}{\gamma^4}\right)\mathrm{eig}(Q^2) &\geq \left((1-\eta)^2 - \lambda^2\gamma^2\right)\frac{\mathrm{eig}(Q^2)}{\gamma^4} \\ &\geq -\frac{\lambda^2}{\gamma^2}\mathrm{eig}(Q)^2 \end{aligned}$$

Bringing all together

$$\mathrm{eig}(A) \geq \lambda\,\mathrm{eig}(Q)\left(\frac{7}{8\gamma} - \lambda\frac{M_\lambda}{\gamma} - \lambda\frac{M_\lambda^2}{\gamma^2}\right) > 0 \tag{15}$$

for sufficiently small stepsize. □

For the rest of the proof we need to show that $AC - B^2 > 0$. Again this a polynomial of Q so the expression of the eigenvalues depends on the eigenvalues of Q. Writing A in the form $A = c_{0,A}I_d + c_{1,A}Q + c_{2,A}Q^2$ and $B = c_{0,B}I_d + c_{1,B}Q$ and $C = c_{0,C}I_d$ then,

$$AC - B^2 = \left(c_{0,A}c_{0,C} - c_{0,B}^2\right)I_d + \left(c_{1,A}c_{0,C} - 2c_{0,B}c_{1,B}\right)Q + \left(c_{2,A}c_{0,C} - c_{1,B}^2\right)Q^2$$

The first term can be written as

$$
\begin{aligned}
c_0 &= f^2(\lambda)(a - \frac{1}{\gamma^2}) \\
&\quad - f(\lambda)(1-\eta)(1+\eta)(a - \frac{1}{\gamma^2}) \\
&\geq -\frac{1}{2}\lambda^2 m(a - \frac{1}{\gamma^2}) \\
&\geq -2\lambda^2 m \frac{1}{\gamma^2} \\
&\geq -2\frac{\lambda^2}{\gamma^2}\,\mathrm{eig}(Q)
\end{aligned}
$$

The eigenvalues of $AC - B^2$ are given by $c_0 + c_1 a + c_2 a^2$

$$
\begin{aligned}
&c_1 + c_2 a = \left(\eta^2 - 1\right) f(\lambda) + f(\lambda)^2 + 2\left(1-\eta^2\right)\frac{\lambda}{\gamma} eig(Q) \\
&+ \frac{2b(1-\eta)\eta f(\lambda) eig(Q)}{\gamma} - 2\frac{\lambda}{\gamma} f(\lambda) eig(Q) + \frac{(1-\eta)^4 eig(Q)^2}{\gamma^4} \\
&- \frac{2\eta(1-\eta)^2(-1+\eta+\gamma\lambda) eig(Q)^2}{\gamma^4} - \frac{2b(1-\eta)(-1+\eta+\gamma\lambda) eig(Q)^2}{\gamma^3} \\
&- \frac{\left(1-\eta^2\right)(-1+\eta+\gamma\lambda)^2 eig(Q)^2}{\gamma^4} + \frac{2b(1-\eta)(-1+\eta+\gamma\lambda) f(\lambda) eig(Q)^2}{\gamma^3} \\
&+ \frac{(-1+\eta+\gamma\lambda)^2 f(\lambda) eig(Q)^2}{\gamma^4} + a\left(-\frac{(1-\eta)^2 eig(Q)^2}{\gamma^2} + \frac{(1-\eta)^2 f(\lambda) eig(Q)^2}{\gamma^2}\right) \\
&\geq (1-\eta^2)\left(\frac{2\lambda}{\gamma} eig(Q) - f(\lambda) - \frac{1-\eta}{(1+\eta)\gamma^2}\,\mathrm{eig}(Q)\right) + 8(\lambda^3\gamma)\,\mathrm{eig}(Q) \\
&\geq \left(\frac{3\lambda^2}{4} - 8\lambda^3\gamma\right)\mathrm{eig}(Q)
\end{aligned}
$$

Multiplying by a and bringing all together,

$$
\mathrm{eig}(AC - B^2) \geq \left(\frac{\lambda^2}{2} a - 2\frac{\lambda^2}{\gamma^2}\right)\mathrm{eig}(Q) \geq \left(\frac{5\lambda^2}{2\gamma^2} - 2\frac{\lambda^2}{\gamma^2}\right)\mathrm{eig}(Q) > 0
$$

6.3 Proof of Contraction rates for Tamed OBABO scheme

Proposition 2. *Let $\gamma \geqslant 2\sqrt{M_\lambda}$ and that $\lambda \leqslant m/(33\gamma^3)$. Set*

$$
a \;=\; 1/M_\lambda\,, \qquad b \;=\; 1/\gamma\,, \qquad \kappa \;=\; m/(3\gamma)\,.
$$

Then, $b^2 \leqslant a/4$ and for all $(x_0, v_0), (y_0, w_0) \in \mathbb{R}^{2d}$, if (x_n, v_n) and (y_n, w_n) is the chain in (8) started at (x_0, v_0) and (y_0, w_0) respectively then, almost surely, for all $n \in \mathbb{N}$,

$$
\|(x_n, v_n) - (y_n, w_n)\|_{a,b}^2 \;\leqslant\; (1 - \lambda\kappa)^n \,\|(x_0, v_0) - (y_0, w_0)\|_{a,b}^2\,.
$$

Proof. Denote $\bar{x} = x - y$, $\bar{x}' = x' - y'$, $\bar{v} = v - w$, $\bar{v}' = v' - w'$ and $\Delta\bar{x} = \bar{x}' - \bar{x}$. Let $\xi, \xi' \in \mathbb{R}^d$ be such that $h_\lambda(x') - h_\lambda(y') = \nabla Q'\bar{x}'$ and $h_\lambda(x) - h_\lambda(y) = \nabla Q\bar{x}$, and let

$Q = \nabla h_\lambda u(\xi)$, $Q' = \nabla h_\lambda(\xi')$. Note that $mI_d \leqslant Q \leqslant M_\lambda I_d$ in the sense of symmetric matrices, and similarly for Q'. With these notations,

$$\begin{aligned}\begin{pmatrix}\bar{x}'\\ \bar{v}'\end{pmatrix} = \begin{pmatrix}\bar{x}\\ \bar{v}\end{pmatrix} + \begin{pmatrix} 0 & \lambda\tilde{\eta} I_d \\ -\frac{\lambda\tilde{\eta}}{2}(Q+Q') & (\tilde{\eta}^2-1)I_d\end{pmatrix}\begin{pmatrix}\bar{x}\\ \bar{v}\end{pmatrix} \\ - \frac{\lambda}{2}\begin{pmatrix}\lambda Q\bar{x}\\ \tilde{\eta}Q'\Delta\bar{x}\end{pmatrix} = (I_{2d}+\lambda A)\begin{pmatrix}\bar{x}\\ \bar{v}\end{pmatrix} + h\end{aligned}$$

with

$$A = \begin{pmatrix} 0 & I_d \\ -\frac{1}{2}(Q+Q') & -\gamma I_d\end{pmatrix}$$

and

$$h = \begin{pmatrix} 0 & \lambda(\tilde{\eta}-1)I_d \\ -\frac{\lambda(\tilde{\eta}-1)}{2}(Q+Q') & (\tilde{\eta}^2-1+\lambda\gamma)I_d\end{pmatrix}\begin{pmatrix}\bar{x}\\ \bar{v}\end{pmatrix} - \frac{\lambda}{2}\begin{pmatrix}\lambda Q\bar{x}\\ \tilde{\eta}Q'\Delta\bar{x}\end{pmatrix}.$$

Writing $\bar{z} = (\bar{x},\bar{v})$, $\bar{z}' = (\bar{x}',\bar{v}')$ and

$$M = \begin{pmatrix} I_d & bI_d \\ bI_d & aI_d\end{pmatrix},$$

we get

$$\begin{aligned}\|\bar{z}'\|_{a,b}^2 &= \|\bar{z}\|_{a,b}^2 + \lambda\bar{z}\cdot(MA+A^TM)\bar{z} + 2\bar{z}\cdot Mh + \|\lambda A\bar{z}+h\|_{a,b}^2 \\ &\leqslant \|\bar{z}\|_{a,b}^2 + \lambda\bar{z}\cdot(MA+A^TM)\bar{z} + 2\|\bar{z}\|_{a,b}\|h\|_{a,b} + 2\lambda^2\|A\bar{z}\|_{a,b}^2 + 2\|h\|_{a,b}^2.\end{aligned}$$

The choice $a = 1/M_\lambda$ and $b = 1/\gamma$ with the condition $\gamma \geqslant 2\sqrt{M_\lambda}$ ensures that $b^2 \leqslant a/4$ and thus for all $z \in \mathbb{R}^d$,

$$\frac{1}{2}\|z\|_{a,0}^2 \leqslant \|z\|_{a,b}^2 \leqslant \frac{3}{2}\|z\|_{a,0}^2. \tag{16}$$

Using (16) and the bounds $|Q|, |Q'| \leqslant M_\lambda$, $|1-\tilde{\eta}| \leqslant \lambda\gamma/2$, $|1-\tilde{\eta}^2-\lambda\gamma| \leqslant \lambda^2\gamma^2/2$ and

$$|\Delta\bar{x}| = |\lambda\tilde{\eta}\bar{v} - \lambda^2/2Q\bar{x}| \leqslant \lambda|\bar{v}| + \lambda^2 M_\lambda/2|\bar{x}|$$

we roughly bound

$$\begin{aligned}\|h\|_{a,b}^2 \leqslant \frac{3}{2}\left|\lambda(\tilde{\eta}-1)\bar{v} - \frac{\lambda^2}{2}Q\bar{x}\right|^2 \\ + \frac{3}{2}a\left|-\frac{\lambda(\tilde{\eta}-1)}{2}(Q+Q')\bar{x} + (\tilde{\eta}^2-1+\lambda\gamma)\bar{v} - \frac{\lambda\tilde{\eta}}{2}Q'\Delta\bar{x}\right|^2 \\ \leqslant 3\lambda^4\gamma^2/4|\bar{v}|^2 + 3\lambda^4 M_\lambda^2/4|\bar{x}|^2 + 3a\left(\lambda^2\gamma M_\lambda/2 + \lambda^3 M_\lambda^2/4\right)^2|\bar{x}|^2 \\ + 3a\left(\lambda^2\gamma^2/2 + \lambda^2 M_\lambda/2\right)^2|\bar{v}|^2 \\ \leqslant 6\lambda^4\max\left(\frac{\left(M_\lambda\gamma/2+\lambda M_\lambda^2/4\right)^2}{M_\lambda} + \frac{3M_\lambda^2}{4}, \frac{M_\lambda\gamma^2}{4} + \left(\frac{\gamma^2}{2}+\frac{M_\lambda}{2}\right)^2\right)\|\bar{z}\|_{a,b}^2 \\ \leqslant 3\lambda^4\gamma^4\|\bar{z}\|_{a,b}^2\end{aligned}$$

where we used that $M_\lambda \leqslant \gamma^2/4$ and $\lambda^2 M_\lambda \leqslant M_\lambda^3/\gamma^6 \leqslant 1$ to simply the expression, and similarly

$$\begin{aligned}\|A\bar z\|_{a,b}^2 &\leqslant \frac{3}{2}|\bar v|^2 + \frac{3}{2}a\left|-\frac{1}{2}(Q+Q')\bar x - \gamma \bar v\right|^2 \\ &\leqslant 3M_\lambda |\bar x|^2 + 3a\left(\frac{M_\lambda}{2}+\gamma^2\right)|\bar v|^2 \\ &\leqslant \frac{27\gamma^2}{8}\|\bar z\|_{a,b}^2\,,\end{aligned}$$

so that, using that $6\lambda^2\gamma^2 \leqslant 1/2$,

$$2\|\bar z\|_{a,b}\|h\|_{a,b} + 2\lambda^2\|A\bar z\|_{a,b}^2 + 2\|h\|_{a,b}^2 \;\leqslant\; 11\lambda^2\gamma^2\|\bar z\|_{a,b}^2\,.$$

On the other hand,

$$MA + A^T M = \begin{pmatrix} -b(Q+Q') & -a(Q+Q')/2 \\ -a(Q+Q')/2 & 2(b-a\gamma)I_d\end{pmatrix}.$$

Bounding $2\bar x\cdot(Q+Q')\bar v \leqslant \bar x \cdot (Q+Q')\bar x/\theta + \theta \bar v\cdot (Q+Q')\bar v$ with $\theta = \gamma/M_\lambda$ we obtain

$$\begin{aligned}\bar z\cdot(MA+A^TM)\bar z &\leqslant \bar z\begin{pmatrix}\left(-\frac{1}{\gamma}+\frac{1}{2M_\lambda\theta}\right)(Q+Q') & 0\\ 0 & 2(b-a\gamma)I_d + \frac{a\theta}{2}(Q+Q')\end{pmatrix}\bar z \\ &\leqslant -\frac{m}{\gamma}|\bar x|^2 - a\gamma\left(1-\frac{2M_\lambda}{\gamma^2}\right)|\bar v|^2 \\ &\leqslant -\frac{2m}{3\gamma}\|\bar z\|_{a,b}^2\,,\end{aligned}$$

where we used that $\gamma/2 \geqslant M_\lambda/\gamma \geqslant m/\gamma$ and (16). Finally, we have obtained that

$$\|\bar z'\|_{a,b}^2 \leqslant \left(1-\frac{2m}{3\gamma}\lambda + 11\lambda^2\gamma^2\right)\|\bar z\|_{a,b}^2 \;\leqslant\; \left(1-\frac{m}{3\gamma}\lambda\right)\|\bar z\|_{a,b}^2\,.$$

$\square$

Corollary 1. *There holds*

$$W_{a,b}\left(\mathcal{L}\left(z_{n+1}\right),\mathcal{L}\left(\tilde z_{n+1}\right)\right) \leq (1-f(\lambda))W_{a,b}\left(\mathcal{L}\left(z_n\right),\mathcal{L}\left(\tilde z_n\right)\right)$$

6.4 Proof of Log-Sobolev inequality for the OBABO scheme

Let $S(x,v) = ||(x,v)||_{a,b}$. We denote $P_{a,b}$ the corresponding Markov semigroup induced the transition operator $S(:) \to S(OBABO(:))$

Lemma 4. *There holds for all bounded Lispchitz functions f,*

$$|\nabla P_{a,b} f(z)| \leq r|P_{a,b}\nabla f(z)|$$

where $r = \sqrt{1-\lambda\kappa}$.

By the one-step a.s contraction of the OBABO scheme, denoting $r = \sqrt{1 - \lambda\kappa}$ one deduces that starting from different intilizations $z_0 = (x_0, v_0)$ and $z_0' = (x_0', v_0')$ there holds

$$|Sz_1 - Sz_1'| \le r|Sz_0 - Sz_0'|$$

a.s which implies the W_∞ contraction of the Markov operator $P_{a,b}$. By Theorem 2.2 in Kuwada (2010) one obtains the result.

Lemma 5. *Let $z \in \mathbb{R}^d$. Then, $P_{a,b}(z)$ satisfies a Log-Sobolev inequality with constant $C_1 := 8\sqrt{1-\eta^2}$*

Proof.

For $x, v, g, g' \in \mathbb{R}^d$, denote

$$\Theta_1(x, v, g, g') = x + \lambda\left(\eta v + \sqrt{1-\eta^2} g\right) - \frac{\lambda^2}{2} h_\lambda(x)$$
$$\Theta_2(x, v, g, g') = \eta^2 v - \frac{\lambda\eta}{2}\left(h_\lambda(x) + h_\lambda(\Theta_1(x, v, g, g'))\right) + \sqrt{1-\eta^2}(\eta g + g')$$

$\Theta = (\Theta_1, \Theta_2)$. The transition of the OBABO chain is given by $\Theta(x_1, v_1) = \Theta(x_0, v_0, G, G')$ with independent $G, G' \sim \mathcal{N}(0, I_d)$. For fixed $z = (x, v)$ one notices that the OBABO chain is Lipschitz map of $N(0, I_{2d})$ since for G=(g,g')

$$|\Theta_1(z, g_1, g_1') - \Theta_1(z, g_2, g_2')| \le \lambda\sqrt{1-\eta^2}|g_1 - g_2|$$

and

$$\begin{aligned}|\Theta_2(z, g_1, g_1') - \Theta_2(z, g_2, g_2')| &\le \frac{\lambda\eta}{2} M_\lambda |\Theta_1(z, g_1, g_1') - \Theta_1(z, g_2, g_2')| + \sqrt{1-\eta^2}\left(\eta|g_1 - g_2| + |g_1' - \right.\\ &\le \left(1 + \lambda + \lambda^2 M_\lambda/2\right)\sqrt{1-\eta^2}|G_1 - G_2|\end{aligned}$$

which shows that Θ is Lispchtiz with constant $2\sqrt{1-\eta^2}$ Since $P_{a,b}(z) = S(S^{-1}z, \Theta(G))$ then it is a Lipschitz function of G with constant $|S|_{Lip}\left(1 + \lambda + \lambda^2 M_\lambda/2\right)\sqrt{1-\eta^2}$. Since the distribution of G satisfies LSI with constant 2 one obtains the result. □

Lemma 6. *Assume that μ satisfies an LSI with constant C_μ. Then, $\mu P_{a,b}$ also satisfies an LSI with constant $r^2 C_\mu + C_1$ where C_1 is given in the previous Lemma.*

Proof. See proof of Theorem 23, point 1, in Monmarché (2021). □

Theorem 8. *Let $z_0 = (x_0, v_0)$ be the initial conditions such that $\mathcal{L}(S(z_0))$ satisfies and LSI with constant $C_{LSI,0}$ Then $\mathcal{L}(S(x_n, v_n))$ satisfies an LSI with*

$$C_{LSI,n} \le r^{2n} C_{LSI,0} + \frac{1 - r^{2n}}{1 - r^2} C_1$$

where $r := \sqrt{1 - \lambda\kappa}$.

Proof. Since S is Lipschitz it is easy to see that $S(z_0)$ satisfies an LSI with constant $|S|C_{LSI,0}$. Consequently, using the statement of the previous Lemma inductively, the result follows easily. □

6.5 Proof of Log-Sobolev inequality for the exponential Euler scheme

We notice that the mean of the next iterate of the tamed exponential euler scheme started at (x, v) is given by

$$F(x,v) := \left(x + \frac{1-\exp(-\gamma\lambda)}{\gamma}v - \frac{h-\gamma^{-1}(1-\exp(-\gamma\lambda))}{\gamma}h_\lambda(x),\right.$$
$$\left.\exp(-\gamma h)v - \frac{1-\exp(-\gamma h)}{\gamma}h_\lambda(x)\right).$$

We make a convenient change of coordinates

$$(\varphi, \psi) = M(x,v) = (x, x + \frac{2}{\gamma}v).$$

In the new coordianates,the mean of the next iteration of the scheme is given by

$$\bar{F} = M \circ F \circ \mathcal{M}^{-1}.$$

In the new coordinates

$$\left(\varphi_{(k+1)h}, \psi_{(k+1)h}\right) \stackrel{d}{=} \bar{F}\left(\varphi_{kh}, \psi_{kh}\right) + \mathcal{N}\left(0, \mathcal{M}\Sigma\mathcal{M}^\top\right).$$

We will show that $\bar{F}$ is a contraction, so the law of the next iterates will be a pushforward contractive map convoluted with a normal random variable. For the Gaussian writing $\mathcal{M}\Sigma\mathcal{M}^\top = \Sigma \otimes I_d$, we can compute

$$\bar{\Sigma}_{1,1} = \frac{2\lambda}{\gamma} - \frac{3}{\gamma^2} + \frac{4\exp(-\gamma\lambda)}{\gamma^2} - \frac{\exp(-2\gamma\lambda)}{\gamma^2} = O(\gamma\lambda^3),$$
$$\bar{\Sigma}_{1,2} = \frac{2\lambda}{\gamma} - \frac{1}{\gamma^2} + \frac{\exp(-2\gamma\lambda)}{\gamma^2} = O(\lambda^2),$$
$$\bar{\Sigma}_{2,2} = \frac{2\lambda}{\gamma} + \frac{5}{\gamma^2} - \frac{8\exp(-\gamma\lambda)}{\gamma^2} + \frac{3\exp(-2\gamma\lambda)}{\gamma^2} = \frac{4\lambda}{\gamma^2} + O(\lambda^2).$$

We conclude that

$$\|\bar{\Sigma}\|_{op} \leq \frac{4\lambda}{\gamma} + O(\lambda^2)$$

As a result,

$$C_{LSI}(\mathcal{N}(0, \mathcal{M}\Sigma\mathcal{M}^\top\lambda)) \leq \frac{4\lambda}{\gamma} + O(\lambda^2) \tag{17}$$

.

Lemma 7. *The map $\bar{F}$ is Lipschitz with constant $||\bar{F}||_{lip} \leq 1 - c_0\frac{m}{\sqrt{\gamma}}\lambda$ for some $c_0 > 0$ absolute constant.*

Proof. The map $\bar{F}$ can be explicIty written as

$$\bar{F}(\varphi,\psi) = \left(\varphi + \frac{1-\exp(-\gamma\lambda)}{2}(\psi-\varphi) - \frac{\lambda-\gamma^{-1}(1-\exp(-\gamma\lambda))}{\gamma}h_\lambda(\varphi),\right.$$
$$\left.\varphi + \frac{1+\exp(-\gamma\lambda)}{2}(\psi-\varphi) - \frac{\lambda+\gamma^{-1}(1-\exp(-\gamma\lambda))}{\gamma}h_\lambda(\varphi)\right).$$

It is easy to see that since h_λ is Lipschitz which implies that it is almost everywhere differentiable, then $\bar{F}$ is also almost everywhere differentiable and continuous everywhere. For every φ that h_λ is differentiable, one calculates

$$\nabla \bar{F}(\varphi,\psi) = \begin{pmatrix} \partial_\varphi \bar{F}_\varphi & \partial_\varphi \bar{F}_\psi \\ \partial_\psi \bar{F}_\varphi & \partial_\psi \bar{F}_\psi \end{pmatrix}$$
$$= \begin{pmatrix} \frac{1+e^{-\gamma\lambda}}{2} I_d - \frac{\lambda-\gamma^{-1}(1-e^{-\gamma\lambda})}{\gamma} \nabla h_\lambda(\varphi) & \frac{1-e^{-\gamma\lambda}}{2} I_d - \frac{\lambda+\gamma^{-1}(1-e^{-\gamma\lambda})}{\gamma} \nabla h_\lambda(\varphi) \\ \frac{1-e^{-\gamma\lambda}}{2} I_d & \frac{1+e^{-\gamma\lambda}}{2} I_d \end{pmatrix}.$$

which can be further split as

$$\nabla \bar{F}(\varphi,\psi) = A + B$$
$$= \frac{1}{2}\begin{pmatrix} (1+\eta)I_d & (1-\eta)I_d - b\nabla h_\lambda(\varphi) \\ (1-\eta)I_d & (1+\eta)I_d \end{pmatrix} + \begin{pmatrix} \frac{\lambda-\gamma^{-1}(1-e^{-\gamma\lambda})}{\gamma} \nabla h_\lambda(\varphi) & 0_d \\ 0_d & 0_d \end{pmatrix}$$

where $\eta = e^{-\lambda\gamma}$ and $b = \frac{\lambda+\gamma^{-1}(1-e^{-\gamma\lambda})}{\gamma}$ Trivially the operator norm of B is $\mathcal{O}(\lambda^2 M_\lambda)$ so we turn to attention to the operator norm of A. Since

$$A^T A = \frac{1}{4}\left(\begin{array}{cc} 2\left(1+\eta^2\right) I_d - 2b(1-\eta)b^2 H^2 & 2\left(1-\eta^2\right) I_d - (1+\eta)bH \\ 2\left(1-\eta^2\right) I_d - (1+\eta)bH & 2\left(1+\eta^2\right) I_d \end{array}\right). \tag{18}$$

which is equal to

$$\begin{pmatrix} 2\left(1+\eta^2\right) I_d & 2\left(1-\eta^2\right) I_d - (1+\eta)bH \\ 2\left(1-\eta^2\right) I_d - (1+\eta)bH & 2\left(1+\eta^2\right) I_d \end{pmatrix} + \begin{pmatrix} -2b(1-\eta)b^2H^2 & 0_d \\ 0_d & 0_d \end{pmatrix}$$

The eignevalues of the first matrix are given by $\frac{1}{4}\left(2(1+\eta^2) + 2(1-\eta^2) - (1+\eta)b\,\mathrm{eig}(H)\right)$ or $\frac{1}{4}\left(2(1+\eta^2) - (2(1-\eta^2) - (1+\eta)b\,\mathrm{eig}(H))\right)$ so the operator norm of the first matrix can be bounded by

$$\frac{1}{4}\max\{4\eta^2 + (1+\eta)bM_\lambda, 4 - (1+\eta)bm\}$$

while the operator norm of the second matrix is trivially given by

$$2b^3(1-\eta)M_l^2 \leq \mathcal{O}(\lambda^3 M_\lambda).$$

As a result $||A||_2 = \sqrt{eig_{max}(A^T A)} \leq \sqrt{1-(1+\eta)bm}$ which yields that

$$||\nabla \bar{F}|| \leq \sqrt{1 - c_0\gamma^{-1}\lambda m} + \mathcal{O}(\lambda^2 M_\lambda) \leq 1 - \frac{c_0 m}{4}\gamma^{-1}\lambda$$

almost everywhere. Since $\bar{F}$ is continuous and differentiable almost everywhere then it is Lipschitz. $\square$

Corollary 2. *If the initial condition of the algorithm satisfies an LSI then $M(x_n, v_n)$ satisfy an LSI.*

Proof. Since in the new coordinates

$$(\varphi_{n+1}, \psi_{n+1}) = \bar{F}(\varphi_n, \psi_n) + N(0, \bar{\Sigma})$$

where the normal distribution is independent, by the behaviour of the log-Sobolev inequality under Lipschitz maps and convolutions using the fact that $\bar{F}$ is a contraction by the previous lemma one obtains

$$C_{LSI,n+1} \le (||\bar{F}||_{lip})^n C_{LSI,0} + \sum_{k=1}^{n} ||\bar{F}||_{lip}^k C_{LSI}(\mathcal{N}(0, \bar{\Sigma})).$$

Substitung the expressions from the previous lemma and (17) yields the result. □

Proof of Lemma 1. For simplicity we prove that $\mathcal{L}(\bar{Y}_1^\lambda, \bar{x}_1^\lambda) = \mathcal{L}(\tilde{Y}_\lambda, \tilde{x}_\lambda)$.
Expanding $\tilde{Y}_\lambda$ one notices that

$$\begin{aligned}
\tilde{Y}_\lambda &= e^{-\lambda\gamma}\tilde{Y}_0 - e^{-\lambda\gamma} h_\lambda(\tilde{x}_0) \int_0^\lambda e^{\gamma s} ds + \sqrt{2\gamma} e^{-\lambda\gamma} \int_0^\lambda e^{\gamma s} dW_s \\
&= \psi_0(\lambda)\tilde{Y}_0 - e^{-\lambda\gamma} h_\lambda(\tilde{x}_0) \frac{1}{\gamma}(e^{\lambda\gamma} - 1) + \sqrt{2\gamma} e^{-\lambda\gamma} \int_0^\lambda e^{\gamma s} dW_s \\
&= \psi_0(\lambda)\tilde{Y}_0 - h_\lambda(\tilde{x}_0) \int_0^\lambda \psi_0(s) ds + \sqrt{2\gamma} e^{-\lambda\gamma} \int_0^\lambda e^{\gamma s} dW_s \\
&= \psi_0(\lambda)\tilde{Y}_0 - \psi_1(\lambda) h_\lambda(\tilde{x}_0) + \sqrt{2\gamma} e^{-\lambda\gamma} \int_0^\lambda e^{\gamma s} dW_s
\end{aligned} \tag{19}$$

One notices that $e^{-\lambda\gamma} \int_0^\lambda e^{\gamma s} dW_s$ follows a zero mean Gaussian distribution and

$$\begin{aligned}
\mathbb{E}(e^{-\lambda\gamma} \int_0^\lambda e^{\gamma s} dW_s)(e^{-\lambda\gamma} \int_0^\lambda e^{\gamma s} dW_s)^T &= e^{-2\lambda\gamma} \int_0^\lambda e^{2\gamma s} ds I_d \\
&= \frac{1}{2\gamma} e^{-2\lambda\gamma}(e^{2\lambda\gamma} - 1) I_d \\
&= \frac{1}{2\gamma}(1 - e^{-2\lambda\gamma}) I_d \\
&= (\int_0^\lambda e^{-2\gamma s} ds) I_d \\
&= (\int_0^\lambda \psi_0^2(s) ds) I_d
\end{aligned}$$

For the second process, using the definition $\tilde{x}$ and the recurrent relationship for ψ_i,

$$\begin{aligned}\tilde{x}_\lambda &= \tilde{x}_0 + \int_0^\lambda \psi_0(s)ds\tilde{Y}_0 - h_\lambda(\tilde{x}_0)\int_0^\lambda\int_0^s e^{-\gamma(s-u)}duds \\ &\quad + \sqrt{2\gamma}\int_0^\lambda e^{-\gamma t}\int_0^t e^{\gamma s}dW_s dt \\ &= \tilde{x}_0 + \psi_1(\lambda)\tilde{Y}_0 - h_\lambda(\tilde{x}_0)\frac{1}{\gamma}\int_0^\lambda e^{-\gamma s}(e^{\gamma s}-1)ds + \sqrt{2\gamma}\int_0^\lambda e^{-\gamma t}\int_0^t e^{\gamma s}dW_s dt \\ &= \tilde{x}_0 + \psi_1(\lambda)\tilde{Y}_0 - h_\lambda(\tilde{x}_0)\int_0^\lambda \psi_1(s)ds + \sqrt{2\gamma}\int_0^\lambda e^{-\gamma t}\int_0^t e^{\gamma s}dW_s dt \\ &= \tilde{x}_0 + \psi_1(\lambda)\tilde{Y}_0 - \psi_2(\lambda)h_\lambda(\tilde{x}_0) + \sqrt{2\gamma}\int_0^\lambda e^{-\gamma t}M_t dt\end{aligned} \tag{20}$$

where $M_t := \int_0^t e^{\gamma s}dW_s$. Since

$$d(e^{-\gamma t}M_t) = -\gamma e^{-\gamma t}M_t dt + e^{-\gamma t}dM_t = -\gamma e^{-\gamma t}M_t dt + dW_t$$

or directly by Ito's formula,

$$e^{-\gamma t}M_t = -\gamma\int_0^t e^{-\gamma s}M_s ds + W_t$$

setting $t = \lambda$, one deduces that

$$\int_0^\lambda e^{-\gamma t}M_t dt = \frac{1}{\gamma}\int_0^\lambda (1 - e^{\gamma(t-\lambda)})dW_t \tag{21}$$

which is a zero-mean Gaussian distribution with

$$\mathbb{E}(\int_0^\lambda e^{-\gamma t}M_t dt)(\int_0^\lambda e^{-\gamma t}M_t dt)^T = \frac{1}{\gamma^2}\left(\int_0^\lambda (1-e^{-\gamma t})^2 dt\right) I_d = \left(\int_0^\lambda \psi_1^2(t)dt\right) I_d.$$

Finally,

$$\begin{aligned}&\mathbb{E}\left(e^{-\lambda\gamma}\int_0^\lambda e^{\gamma s}dW_s\right)\left(\frac{1}{\gamma}\int_0 (1-e^{\gamma(s-\lambda)})dW_s\right)^T \\ &\qquad = \frac{1}{\gamma}\left(\int_0^\lambda e^{\gamma(s-\lambda)}(1-e^{\gamma(s-\lambda)})ds\right) I_d = \int_0^\lambda \psi_0(s)\psi_1(s)ds \quad I_d\end{aligned}$$

We have proved that the terms $\int_0^\lambda e^{-\gamma t}M_t dt$ appearing in (20) and $e^{-\lambda\gamma}\int_0^\lambda e^{\gamma s}dW_s$ in (19) follow Gaussian distributions with the required cross-covariance matrix.
The result immediately follows. □

7 Convergence of the Monotonic polygonal Stochastic exponential scheme

Definition 3. *Let $n \in \mathbb{N}$. Let a random variable X. We define for $t \in [n\lambda, (n+1)\lambda]$*

$$\tilde{z}_t^{X,n,\lambda} = (\tilde{x}_t^{X,n,\lambda}, \tilde{V}_t^{X,n,\lambda})$$

the one step movement of an algorithm started at a random variable X.

Lemma 8. *Let $p \geq 2$ and let Y be a random variable such that $\mathcal{L}(Y) = \pi$. Then*

$$\mathbb{E}|Y|^p \leq 2^{p-1}((\frac{d}{m})^{\frac{p}{2}}(1+p/d)^{\frac{p}{2}-1} + \left(\frac{2}{m}(u(0)+d)\right)^{p/2} := C_{\pi,p}.$$

Proof. Since $e^{-u(x)} = e^{-(u-\frac{m}{2}|x|^2)}e^{-\frac{m}{2}|x|^2}$ and since the function $u - \frac{m}{2}|x|^2$ is convex (due to strong convexity of u) the assumptions of Theorem 21 are valid for X which has distribution $\mathcal{N}(0, \frac{1}{m}I_d)$, and for Y with density π. Since the function $g : x \to |x|^p$ is convex, applying the results of Theorem 21 leads to

$$\mathbb{E}|Y - \mathbb{E}Y|^p \leq \mathbb{E}|X|^p = (\frac{d}{m})^{\frac{p}{2}}\frac{\Gamma((d+p)/2)}{\Gamma(d/2)(d/2)^{\frac{p}{2}}} \leq (\frac{d}{m})^{\frac{p}{2}}(1+p/d)^{\frac{p}{2}-1}. \tag{22}$$

Combining (22) and the result of Lemma 20 yields

$$\mathbb{E}|Y|^p \leq 2^{p-1}\left(\mathbb{E}|Y-\mathbb{E}Y|^p + (\mathbb{E}|Y|)^p\right) \leq 2^{p-1}((\frac{d}{m})^{\frac{p}{2}}(1+p/d)^{\frac{p}{2}-1} + \left(\frac{2}{m}(u(0)+\frac{d}{-})\right)^{p/2}, \tag{23}$$

which completes the proof. □

Since Π is the product of π and a standard Gaussian the following result follows easily.

Lemma 9. *Let X such that $\mathcal{L}(X) = \Pi$. Let $p > 0$. Then,*

$$\mathbb{E}|X|^{2p} \leq C_{\pi,p}$$

Lemma 10. *Let X such that $\mathcal{L}(X) = \Pi$. Then, for $\tilde{z}_t^{X,n,\lambda}$ defined in Definition 3 there holds*

$$\sup_{t\in[n\lambda,(n+1)\lambda]} \mathbb{E}|\tilde{z}_t^{X,n,\lambda}|^{2p} \leq C_{z,p}$$

Lemma 11. *Let $n \in \mathbb{N}$ and X such that $\mathcal{L}(X) = \Pi$. Then,*

$$\mathbb{E}|\tilde{x}_t^{X,n,\lambda} - \tilde{x}_{n\lambda}^{X,n,\lambda}|^2 \leq C_{os,2}\lambda^2$$

and

$$\mathbb{E}|\tilde{x}_t^{X,n,\lambda} - \tilde{x}_{n\lambda}^{X,n,\lambda}|^4 \leq C_{os,4}\lambda^4$$

where $C_{os,p} \leq \mathcal{O}(d^p)$

Proof.

$$\mathbb{E}|\tilde{V}_t^{X,n,\lambda}|^2 \leq 3(1-e^{-\lambda(t-\lambda n)})^2\mathbb{E}|\tilde{V}_{n\lambda}^{X,n,\lambda}|^2 + 6\lambda\gamma d + 3\lambda|h_\lambda(x_{n\lambda}|^2 \leq C(\mathbb{E}|V_{n\lambda}|^2 + \mathbb{E}|x_{n\lambda}|^2 + \lambda\gamma d + 4)$$

where the last step is obtained by using the fact that the drift coefficient has linear growth. Thus,

$$\mathbb{E}|\tilde{x}_t^{X,n,\lambda} - \tilde{x}^{X,n,\lambda}n\lambda|^2 \leq \mathbb{E}\left(\int_{n\lambda}^t \tilde{V}_s^{X,n,\lambda}ds\right)^2 \leq \lambda\int \mathbb{E}|\tilde{V}_s^{X,n,\lambda}|^2 ds$$

which completes the first proof. With the same arguments,

$$\mathbb{E}|V_t|^4 \leq C'(\mathbb{E}|\tilde{V}_{n\lambda}^{X,n,\lambda}|^4 + \mathbb{E}|\tilde{x}_{n\lambda}^{X,n,\lambda}|^4 + \lambda^2\gamma^2 d^2 + 1)$$

and since

$$\mathbb{E}|\tilde{x}_t^{X,n,\lambda} - \tilde{x}_{n\lambda}^{X,n,\lambda}|^4 \leq \mathbb{E}\left(\int_{n\lambda}^t \tilde{V}_s^{X,n,\lambda} ds\right)^4 \leq \lambda^3 \int \mathbb{E}|\tilde{V}_s^{X,n,\lambda}|^4 ds$$

the result immediately follows. □

Lemma 12. *Let* $\tilde{z}_t^{X,n,\lambda} = (\tilde{x}_t^{X,n,\lambda}, \tilde{V}_t^{X,n,\lambda})$ *as in Lemma 10. Then,*

$$\mathbb{E}|h_\lambda(\tilde{x}_{n\lambda}^{X,n,\lambda}) - h(\tilde{x}_{n\lambda}^{X,n,\lambda})|^2 \leq C_{tam}\lambda^2$$

Lemma 13. *Let* X *such that* $\mathcal{L}(X) = \Pi$. *Let the Langevin SDE* (p_t, Q_t) *as in* (4) *started at time* $n\lambda$ *at* X *and* $\tilde{z}_t^{X,n,\lambda} = (\tilde{x}_t^{X,n,\lambda}, \tilde{V}_t^{X,n,\lambda})$ *the one-step movement of the algorithm as in Definition 3. There holds,*

$$W_2\left(\mathcal{L}(\tilde{z}_t^{X,n,\lambda}), \Pi\right) \leq C\lambda^2$$

As a result,

$$W_{a,b}\left(\mathcal{L}(\tilde{z}_t^{X,n,\lambda}), \Pi\right) \leq \sqrt{\frac{3}{2}}C\lambda^2$$

Proof. Since Π is the invariant measure of the kinetic Langevin SDE (4), $\mathcal{L}((p_t, Q_t)) = \Pi$ so it suffices to bound $\mathbb{E}|\tilde{z}_t^{X,n,\lambda} - (p_t, Q_t)|^2$. Since the process have the same initial conditions, one notices (using Jensen's inequality).

$$\begin{aligned}\mathbb{E}|\tilde{x}_t^{X,n,\lambda} - p_t|^2 \leq \mathbb{E}|\int_{n\lambda}^t \tilde{V}_s^{X,n,\lambda} - Q_s ds|^2 &\leq (t - n\lambda)\int \mathbb{E}|\tilde{V}_s^{X,n,\lambda} - Q_s|^2 ds \\ &\leq \lambda \int_{n\lambda}^t \mathbb{E}|\tilde{V}_s^{X,n,\lambda} - Q_s|^2 ds\end{aligned} \tag{24}$$

With the same argument,

$$
\begin{aligned}
\mathbb{E}|\tilde{V}_t^{X,n,\lambda} - Q_t|^2 &\leq \mathbb{E}|\int_{n\lambda}^{t} e^{-\gamma(t-s)}(h(p_s) - h_\lambda(\tilde{x}_{n\lambda}^{X,n,\lambda}))ds|^2 ds \\
&\leq \lambda \int_{n\lambda}^{t} e^{-2\gamma(t-s)}\mathbb{E}|h(p_s) - h_\lambda(\tilde{x}_{n\lambda}^{X,n,\lambda})|^2 ds \\
&\leq \lambda \int_{n\lambda}^{t} \mathbb{E}|h(p_s) - h_\lambda(\tilde{x}_{n\lambda}^{X,n,\lambda})|^2 ds \\
&\leq 3\lambda \int_{n\lambda}^{t} \mathbb{E}|h(\tilde{x}_s^{X,n,\lambda}) - h(\tilde{x}_{n\lambda}^{X,n,\lambda})|^2 ds \\
&+ 3\lambda \int_{n\lambda}^{t} \mathbb{E}|h(\tilde{x}_{n\lambda}^{X,n,\lambda}) - h_\lambda(\tilde{x}_{n\lambda}^{X,n,\lambda})|^2 ds \\
&+ 3\lambda \int \mathbb{E}|h(p_s) - h(\tilde{x}_s^{X,n,\lambda})|^2 ds \\
&\leq 3\lambda \int_{n\lambda}^{t} \mathbb{E}(1 + |\tilde{x}_{n\lambda}^{X,n,\lambda}| + |\tilde{x}_s^{X,n,\lambda}|)^{2l}|\tilde{x}_s^{X,n,\lambda} - \tilde{x}_{n\lambda}^{X,n,\lambda}|^2 ds \\
&+ 3\lambda \int_{n\lambda}^{t} \mathbb{E}|h(\tilde{x}_{n\lambda}^{X,n,\lambda}) - h_\lambda(\tilde{x}_{n\lambda}^{X,n,\lambda})|^2 ds \\
&+ 3\lambda \int_{n\lambda}^{t} \mathbb{E}(1 + |\tilde{x}_s^{X,n,\lambda}| + |p_s|)^{2l+1}|\tilde{x}_s^{X,n,\lambda} - p_s| ds \\
&\leq 3\lambda \int \sqrt{\mathbb{E}(1 + |\tilde{x}_{n\lambda}^{X,n,\lambda}| + |\tilde{x}_s^{X,n,\lambda}|)^{4l}}\sqrt{\mathbb{E}|\tilde{x}_s^{X,n,\lambda} - \tilde{x}_{n\lambda}^{X,n,\lambda}|^4} ds \\
&+ 3\lambda \int_{n\lambda}^{t} \mathbb{E}|h(\tilde{x}_{n\lambda}^{X,n,\lambda}) - h_\lambda(\tilde{x}_{n\lambda}^{X,n,\lambda})|^2 ds \\
&+ 9\lambda^3 \int_{n\lambda}^{t} \mathbb{E}(1 + |\tilde{x}_s^{X,n,\lambda}| + |p_s|)^{4l+2} ds \\
&+ \frac{1}{\lambda^2} \int_{n\lambda}^{t} \mathbb{E}|\tilde{x}_s^{X,n,\lambda} - p_s|^2 ds
\end{aligned}
\tag{25}
$$

By the one step approximation, it follows that

$$
3\lambda \int \sqrt{\mathbb{E}(1 + |\tilde{x}_{n\lambda}^{X,n,\lambda}| + |\tilde{x}_s^{X,n,\lambda}|)^{4l}}\sqrt{\mathbb{E}|\tilde{x}_s^{X,n,\lambda} - \tilde{x}_{n\lambda}^{X,n,\lambda}|^4} ds \leq C_1\lambda^4 \tag{26}
$$

By the taming error, the second term can be bounded as follows

$$
3\lambda \int_{n\lambda}^{t} \mathbb{E}|h(\tilde{x}_{n\lambda}^{X,n,\lambda}) - h_\lambda(\tilde{x}_{n\lambda}^{X,n,\lambda})|^2 ds \leq 3C_{tam}\lambda^4 \tag{27}
$$

By the uniform moment bounds the third term is $\mathcal{O}(\lambda^4)$ so one obtains

$$
\mathbb{E}|\tilde{V}_t^{X,n,\lambda} - Q_t|^2 \leq \hat{C}\lambda^4 + \frac{1}{\lambda^2}\int_{n\lambda}^{t} \mathbb{E}|\tilde{x}_s^{X,n,\lambda} - p_s|^2 ds \tag{28}
$$

Combining this with (24) leads to

$$
\mathbb{E}|\tilde{x}_t^{X,n,\lambda} - p_t|^2 \leq \hat{C}\lambda^5 + \frac{1}{\lambda}\int_{n\lambda}^{t}\int_{n\lambda}^{s} \mathbb{E}|\tilde{x}_u^{X,n,\lambda} - p_u|^2 du ds \quad \forall t \in [n\lambda, (n+1)\lambda]
$$

As a result,

$$\sup_{n\lambda\leq z\leq t}\mathbb{E}|\tilde{x}_z^{X,n,\lambda}-p_z|^2\leq \hat{C}\lambda^5+\int_{n\lambda}^t\sup_{n\lambda\leq u\leq s}\mathbb{E}|\tilde{x}_u^{X,n,\lambda}-p_u^{\lambda,n}|^2ds$$

Since by the uniform moment bounds the right hand side is finite, by Grownwall's inequality one obtains

$$\mathbb{E}|\tilde{x}_t^{X,n,\lambda}-p_t|^2\leq\sup_{n\lambda\leq z\leq t}\mathbb{E}|\tilde{x}_z^{X,n,\lambda}-p_z|^2\leq e^{\lambda}\hat{C}\lambda^5\leq 2\hat{C}\lambda^5 \tag{29}$$

Plugging this into (28) yields

$$\mathbb{E}|\tilde{V}_t^{X,n,\lambda}-Q_t|^2\leq\hat{C}\lambda^4$$

□

Theorem 9. *There holds*

$$W_2\left(\mathcal{L}(\theta_n,V_n),\Pi\right)\leq(1-f(\lambda))^nW_2\left(\mathcal{L}(\theta_0,V_0),\Pi\right)+C\lambda^2f(\lambda)$$

Using the connection between $W_{a,b}$ and W_2 one obtains

$$W_2\left(\mathcal{L}(\theta_n,V_n),\Pi\right)\leq 2\sqrt{M_\lambda}(1-f(\lambda))^nW_{a,b}\left(\mathcal{L}(\theta_0,V_0),\Pi\right)+3C\lambda^2\sqrt{M_\lambda}/f(\lambda)$$

Proof.

$$\begin{aligned}&W_{a,b}\left(\mathcal{L}(\theta_{n+1},V_{n+1}),\Pi\right)\\&\leq W_{a,b}\left(\mathcal{L}(\theta_{n+1},V_{n+1}),\mathcal{L}(\tilde{x}_{n\lambda}^{X,n,\lambda},\tilde{V}_{n\lambda}^{X,n,\lambda})\right)+W_{a,b}\left(\mathcal{L}(\tilde{x}_{n\lambda}^{X,n,\lambda},\tilde{V}_{n\lambda}^{X,n,\lambda}),\Pi\right)\\&\leq(1-f(\lambda))W_{a,b}(\mathcal{L}(\theta_n,V_n),\Pi)+\sqrt{\frac{3}{2}}C\lambda^2\end{aligned}\tag{30}$$

By induction, one deduces that

$$W_{a,b}\left(\mathcal{L}(\theta_{n+1},V_{n+1}),\Pi\right)\leq(1-f(\lambda))^nW_{a,b}\left(\mathcal{L}(\theta_0,V_0),\Pi\right)+C\lambda^2/f(\lambda)$$

As a result,

$$W_2\left(\mathcal{L}(\theta_{n+1},V_{n+1}),\Pi\right)\leq\sqrt{M_\lambda}(1-f(\lambda))^nW_{a,b}\left(\mathcal{L}(\theta_0,V_0),\Pi\right)+C\lambda^2/f(\lambda)\sqrt{M_\lambda}$$

□

8 Convergence of the OBABO scheme

Corollary 3. *For all $p\geqslant 1$ and all $\nu,\mu\in\mathcal{P}_p(\mathbb{R}^{2d})$,*

$$\mathcal{W}_{p,a,b}^2\left(\nu P,\mu P\right)\quad\leqslant\quad(1-\lambda\kappa)\mathcal{W}_{p,a,b}^2\left(\nu,\mu\right)\,, \tag{31}$$

and for all $n\in\mathbb{N}$,

$$\mathcal{W}_p^2\left(\nu P^n,\mu P^n\right)\;\leqslant 3M_\lambda(1-\lambda\kappa)^n\mathcal{W}_p^2(\nu,\mu)\,.$$

Moreover, P admits a unique invariant probability distribution π_λ, and $\pi_\lambda\in\mathcal{P}_p(\mathbb{R}^{2d})$ for all $p\geqslant 1$.

Proof. (31) is a direct consequence of the previous result.
Using the connection of the $|\cdot|_{2,a,b}$ and the Euclidean distance distance in (9) one easily obtains the result for the Wasserstein distance. The existence and uniqueness of the invariant measure is a consequence of Banach fixed point theorem. □

Let P_O the Markov kernel associated with O-step in (O). The Markov kernel P associated with the whole chain can be written as

$$P = P_O P_V P_O \tag{32}$$

Lemma 14. *Π is invariant under P_O i.e*

$$\Pi P_0 = \Pi$$

in distribution.

Proof. Writing $\Pi = (\Pi_1, \Pi_2)$ the first part follows π while the second follows $N(0, I_d)$. The O-step in (O), only acts on Π_2. Since the Gaussian random variable G added is independent from Π_2, the resulting distribution will still be a Gaussian with mean 0 and covariance $(\tilde{\eta}^2 + (1-\tilde{\eta}^2))I_d = I_d$ □

Lemma 15. *For any measures μ, $\nu \in P_p(\mathbb{R}^d)$ there holds*

$$W_p(\mu P_O, \nu P_O) \leq W_p(\mu, \nu)$$

Proof. Let $(x_i, v_i) \in \mathbb{R}^{2d}, i = 1, 2$, and $G \sim \mathcal{N}(0, I_d)$. Set $x_i' = x_i$ and $v_i' = \eta v_i + \sqrt{1-\eta^2} G$ for $i = 1, 2$. Then $(x_i', v_i') \sim \lambda_{(x_i, v_i)} P_O$ and

$$|x_1' - x_2'|^2 + |v_1' - v_2'|^2 \leqslant |x_1 - x_2|^2 + |v_1 - v_2|^2$$

almost surely.
Considering $\nu_i \in \mathcal{P}_p\left(\mathbb{R}^d\right)$ for $i = 1, 2$, sampling $(x_i, v_i) \sim \nu_i$, using the previous coupling. Taking the power $p/2$ and the expectation yields

$$\mathbb{E}|(x_1', v_1') - (x_2', v_2')|^p \leq W_p^p(\nu_1, \nu_2)$$

which leads to the result. □

Proposition 3. *Let Π_λ be the invariant measure of the tamed $OBABO$ algorithm. Then,*

$$\mathcal{W}_{p,a,b}\left(\Pi_\lambda, \Pi\right) \leqslant \frac{2}{\lambda\kappa} \mathcal{W}_{p,a,b}(\Pi P, \Pi) \leq K_1 \frac{2}{\lambda\kappa} \mathcal{W}_p\left(\Pi P_V, \Pi\right)$$

Proof. Writing

$$\begin{aligned}
W_{p,a,b}(\Pi_\lambda, \Pi) = W_{p,a,b}(\Pi_\lambda dP, \Pi) &\leq W_{p,a,b}(\Pi_\lambda P, \Pi P) + W_{p,a,b}(\Pi P, \Pi) \\
&\leq \sqrt{(1-\lambda\kappa)} W_{p,a,b}(\Pi_\lambda, \Pi) + W_{p,a,b}(\Pi P, \Pi) \\
&\leq (1 - \frac{\lambda\kappa}{2}) W_{p,a,b}(\Pi_\lambda, \Pi) + W_{p,a,b}(\Pi P, \Pi)
\end{aligned}$$

leads to

$$W_{p,a,b}(\Pi_\lambda, \Pi) \leq \frac{2}{\lambda\kappa} W_{p,a,b}(\Pi P, \Pi) \leq \sqrt{3M_\lambda} \frac{2}{\kappa\lambda} W_p(\Pi P, \Pi) \tag{33}$$

where the last step is obtained by (9). In addition, noting that since Π is invariant under the O-step due to Lemma 14, then

$$W_p(\Pi P, \Pi) = W_p(\Pi P_O P_V P_O, \Pi) = W_p(\Pi P_V P_0, \Pi P_0)$$

and applying the result of Lemma 15 one obtains

$$W_p(\Pi P, \Pi) \leq W_p(\Pi P_V, \Pi). \tag{34}$$

Combining (34) and (33) completes the proof. $\square$

8.1 Comparing the BAB steps to a Hamiltonian system

In order to complete our proof of the bound of the Wasserstein distance of Π_λ and Π one need to control $W_p(\Pi, \Pi P_V)$. We introduce the transition operator Q_t of the Hamiltonian dynamics: given $(x, v) \in \mathbb{R}^{2d}$ and considering $(\hat{x}_t, \hat{v}_t)_{t \geqslant 0}$ the solution of

$$\hat{x}_t' = \hat{v}_t, \quad \hat{v}_t' = -\nabla\left(\hat{x}_t\right), \quad \hat{x}_0 = x, \quad \hat{v}_0 = v, \tag{35}$$

denote $\Phi_{HD}(x, v) = (\hat{x}_\lambda, \hat{v}_\lambda)$ and $Q_\lambda \varphi(x, v) = \varphi\left(x_\lambda, v_\lambda\right)$.

In addition, the Verlet integrator BAB can be presented as

$$\Phi_V(x, v) = \left(x + \lambda v - \frac{1}{2}\lambda^2 h_\lambda(x), v - \frac{1}{2}\lambda\left(h_\lambda(x) + h_\lambda\left(x + \lambda v - \frac{1}{2}\lambda^2 h_\lambda(x)\right)\right)\right)$$

so that the steps BAB are given as $\left(x_1, v_1'\right) = \Phi_V\left(x_0, v_0'\right)$

Lemma 16.

$$\sup_{t \in [0,\lambda]} |\hat{x}_t|^p + |\hat{v}_t|^p \leq C(1 + |u(x)|^p + |v|^p) \quad \forall p \in \mathbb{N}.$$

Suppose that the intial condition is the measure Π then,

$$\mathbb{E}|\hat{x}_t|^p + \mathbb{E}|\hat{v}_t|^p \leq \hat{C}_p \quad \forall t \in [0, \lambda]$$

Proof. Since for large values, u is superlinear and the fact that in a Hamiltonian system the energy is preserved then,

$$|\hat{x}_t|^p + |\hat{v}_t|^p \leq C(1 + u(\hat{x}_t) + |\hat{v}_t|)^p \leq C(1 + u(x)^p + |v|^p).$$

The second claim follows by taking expectations. $\square$

Lemma 17. *There holds*

$$W_2(\Pi, \Pi P_V) = W_2(\Pi Q_\lambda, \Pi P_V) \leq \mathcal{O}\left(M_\lambda \lambda^2\right)$$

Proof. Let (x_t, w_t) the solution of

$$x_t' = w_t, \quad w_t' = -h_\lambda(x), \quad x_0 = x, \quad w_0 = v$$

In other words,

$$x_t = x + tv - \frac{t^2}{2}h_\lambda(x), \quad w_t = v - th_\lambda(x),$$

where $\mathcal{L}(x,v) = \Pi$ Thus,

$$\Phi_V(x,v) = (x_\lambda, \tilde{w}_\lambda) = \left(x_\lambda, w_\lambda + \frac{\lambda}{2}(h_\lambda(x) - h_\lambda(x_\lambda))\right)$$

For all $t \in [0,\lambda]$,

$$\begin{aligned}
\mathbb{E}|\hat{x}_t - x_t|^2 &\le \lambda \mathbb{E}|\int_0^t v_s - v ds + \int_0^t \int_0^s h_\lambda(x) duds|^2 \\
&\le \lambda \int_0^t \int_0^s \mathbb{E}|\nabla u(\hat{x}_u) - h_\lambda(x)|^2 duds \\
&\le 2\lambda \int_0^t \int_0^s \mathbb{E}|\nabla u(\hat{x}_u) - h_\lambda(\hat{x}_u)|^2 duds + 2\lambda \int_0^t \int_0^s \mathbb{E}|h_\lambda(x) - h_\lambda(\hat{x}_u)|^2 duds \\
&\le C\lambda^4 + \lambda M_\lambda^2 \int_0^t \int_0^s \mathbb{E}|x - \hat{x}_u|^2 duds \\
&\le C\lambda^4 + \lambda M_\lambda^2 \int_0^t \int_0^s (\int_0^u \mathbb{E}|\hat{v}_z| dz)^2 duds \\
&\le C\lambda^4 + \lambda M_\lambda^2 \int_0^t \int_0^s u \int_0^u \mathbb{E}|\hat{v}_z|^2 dzduds \\
&\le C\lambda^4 + C'\lambda^5 M_\lambda^2
\end{aligned} \tag{36}$$

In addition, with the same arguments as before

$$\mathbb{E}|w_t - \hat{v}_t|^2 \le \lambda \int_0^t \mathbb{E}|\nabla u(\hat{x}_s) - h_\lambda(x)|^2 ds \le \hat{C}\lambda^4 M_\lambda^2$$

Bringing all together,

$$\begin{aligned}
&\mathbb{E}|\hat{v}_\lambda - \tilde{w}_\lambda|^2 \le 2\mathbb{E}|\hat{v}_\lambda - w_\lambda|^2 + 2\frac{\lambda^2}{4}\mathbb{E}|h_\lambda(x_\lambda) - h_\lambda(x)|^2 \\
&\le 2\hat{C}\lambda^4 M_\lambda^2 + \frac{\lambda^2}{2} M_\lambda^2 \mathbb{E}|x_\lambda - x|^2 \\
&\le 2\hat{C}\lambda^4 M_\lambda^2 + \lambda^2 M_\lambda^2 (\lambda^2 |v|^2 + \frac{\lambda^4}{4}|h_\lambda(x)|^2) \\
&\le \mathcal{O}(\lambda^4 M_\lambda^2)
\end{aligned}$$

$\square$

9 Acknowledgements

This work has been partially supported by project MIS 5154714 of the National Recovery and Resilience Plan Greece 2.0 funded by the European Union under the NextGenerationEU Program.

A Appendix

Lemma 18. *Recall that $\eta = e^{-\lambda\gamma}$. Under our restrictions for γ and λ, There holds*

$$\frac{\lambda\gamma}{2} \leq 1 - \eta \leq \lambda\gamma \tag{37}$$

$$0 \leq -1 + \eta + \lambda\gamma \leq \frac{\lambda^2\gamma^2}{2} \tag{38}$$

Lemma 19 (Raginsky et al. (2017), Lemma 3, adapted)**.** *Let L_t be given by the overdamped Langevin SDE in* (1) *with initial condition L_0. Then, under the strong convexity assumotuin there holds*

$$\mathbb{E}|L_t|^2 \leq \mathbb{E}|L_0|^2 e^{-mt} + 2\frac{u(0)+d}{m}(1 - e^{-mt}). \tag{39}$$

Proof. Notice that in the proof of Lemma 3 in Raginsky et al. (2017) only the dissipativity condition is used which is a trivial consequence of strong convexity. □

Lemma 20. *If Y is a random variable such that $\mathcal{L}(Y) = \pi$, then*

$$\mathbb{E}|Y|^2 \leq \frac{2}{m}(u(0) + \frac{d}{\beta}).$$

As a result,

$$\mathbb{E}|Y| \leq \sqrt{\frac{2}{m}(u(0) + d)}. \tag{40}$$

Proof. Since the solution of the Langevin SDE (1) converges in W_2 distance to π and $\sup_{t\geq 0} \mathbb{E}|\bar{L}_t|^2 < \infty$ due to (39), this also implies the convergence of second moments. Therefore, if Y is random variable such that $\mathcal{L}(W) = \pi$ by Lemma 19 there holds that

$$\mathbb{E}|Y|^2 = \lim_{t\to\infty} \mathbb{E}|\bar{L}_t|^2 \leq \frac{2}{m}(u(0) + d).$$

□

Lemma 21. *[Hargé (2004), Theorem 1.1] Let X follow $\mathcal{N}(m_0, \Sigma)$ with density φ, and let Y have density φf where f is a log-concave function. Then for any convex map g there holds:*

$$\mathbb{E}[g\,(Y - \mathbb{E}Y)] \leq \mathbb{E}[g\,(X - \mathbb{E}X)].$$

References

, 1983. A method for unconstrained convex minimization problem with the rate of convergence o (1/k2), in: Dokl. Akad. Nauk. SSSR, p. 543.

Akyildiz, Ö.D., Sabanis, S., 2020. Nonasymptotic analysis of stochastic gradient hamiltonian monte carlo under local conditions for nonconvex optimization. arXiv preprint arXiv:2002.05465 .

Altschuler, J.M., Chewi, S., 2024. Faster high-accuracy log-concave sampling via algorithmic warm starts. Journal of the ACM 71, 1–55.

Balasubramanian, K., Chewi, S., Erdogdu, M.A., Salim, A., Zhang, S., 2022. Towards a theory of non-log-concave sampling: first-order stationarity guarantees for langevin monte carlo, in: Conference on Learning Theory, PMLR. pp. 2896–2923.

Barkhagen, M., Chau, N.H., Moulines, É., Rásonyi, M., Sabanis, S., Zhang, Y., 2021. On stochastic gradient langevin dynamics with dependent data streams in the logconcave case. Bernoulli 27, 1–33.

Brosse, N., Durmus, A., Moulines, É., Sabanis, S., 2019. The tamed unadjusted Langevin algorithm. Stochastic Processes and their Applications 129, 3638–3663.

Chau, H.N., Rasonyi, M., 2019. Stochastic Gradient Hamiltonian Monte Carlo for Non-Convex Learning. arXiv preprint arXiv:1903.10328 .

Chau, N.H., Moulines, É., Rásonyi, M., Sabanis, S., Zhang, Y., 2021. On stochastic gradient langevin dynamics with dependent data streams: The fully nonconvex case. SIAM Journal on Mathematics of Data Science 3, 959–986.

Cheng, X., Chatterji, N.S., Abbasi-Yadkori, Y., Bartlett, P.L., Jordan, M.I., 2018a. Sharp convergence rates for Langevin dynamics in the nonconvex setting. arXiv preprint arXiv:1805.01648 .

Cheng, X., Chatterji, N.S., Bartlett, P.L., Jordan, M.I., 2018b. Underdamped Langevin MCMC: A non-asymptotic analysis, in: Conference On Learning Theory, pp. 300–323.

Chewi, S., Stromme, A.J., 2024. The ballistic limit of the log-sobolev constant equals the polyak-{\L} ojasiewicz constant. arXiv preprint arXiv:2411.11415 .

Dalalyan, A.S., 2017. Theoretical guarantees for approximate sampling from smooth and log-concave densities. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79, 651–676.

Dalalyan, A.S., Riou-Durand, L., 2020. On sampling from a log-concave density using kinetic langevin diffusions. Bernoulli 26, 1956–1988.

Durmus, A., Moulines, E., 2017. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. The Annals of Applied Probability 27, 1551–1587.

Durmus, A., Moulines, E., 2019. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. Bernoulli 25, 2854–2882.

Erdogdu, M.A., Hosseinzadeh, R., 2021. On the convergence of langevin monte carlo: The interplay between tail growth and smoothness, in: Conference on Learning Theory, PMLR. pp. 1776–1822.

Erdogdu, M.A., Hosseinzadeh, R., Zhang, S., 2022. Convergence of langevin monte carlo in chi-squared and rényi divergence, in: International Conference on Artificial Intelligence and Statistics, PMLR. pp. 8151–8175.

Gao, X., Gurbuzbalaban, M., Zhu, L., 2018. Breaking reversibility accelerates langevin dynamics for global non-convex optimization. arXiv preprint arXiv:1812.07725 .

Gao, X., Gürbüzbalaban, M., Zhu, L., 2021. Global convergence of stochastic gradient hamiltonian monte carlo for nonconvex stochastic optimization: Nonasymptotic performance bounds and momentum-based acceleration. Operations Research .

Hargé, G., 2004. A convex/log-concave correlation inequality for gaussian measure and an application to abstract wiener spaces. Probability theory and related fields 130, 415–440.

Horn, R.A., Zhang, F., 2005. Basic properties of the schur complement, in: The Schur Complement and Its Applications. Springer, pp. 17–46.

Johnston, T., Lytras, I., Sabanis, S., 2024. Kinetic langevin mcmc sampling without gradient lipschitz continuity—the strongly convex case. J. Complex. 85.

Johnston, T., Sabanis, S., 2024. A strongly monotonic polygonal euler scheme. Journal of Complexity 80, 101801.

Kuwada, K., 2010. Duality on gradient estimates and wasserstein controls. Journal of Functional Analysis 258, 3758–3774.

Leimkuhler, B.J., Paulin, D., Whalley, P.A., 2024. Contraction and convergence rates for discretized kinetic langevin dynamics. SIAM Journal on Numerical Analysis 62, 1226–1258.

Lovas, A., Lytras, I., Rásonyi, M., Sabanis, S., 2023. Taming neural networks with tusla: Nonconvex learning via adaptive stochastic gradient langevin algorithms. SIAM Journal on Mathematics of Data Science 5, 323–345.

Lytras, I., Mertikopoulos, P., 2024. Tamed langevin sampling under weaker conditions. arXiv preprint arXiv:2405.17693 .

Lytras, I., Mertikopoulos, P., 2025. Tamed langevin sampling under weaker conditions, in: Proceedings of The 28th International Conference on Artificial Intelligence and Statistics, PMLR. pp. 847–855.

Lytras, I., Sabanis, S., 2025. Taming under isoperimetry. Stochastic Process. Appl. 188.

Ma, Y.A., Chatterji, N.S., Cheng, X., Flammarion, N., Bartlett, P.L., Jordan, M.I., 2021. Is there an analog of Nesterov acceleration for gradient-based MCMC? Bernoulli 27, 1942 – 1992. URL: https://doi.org/10.3150/20-BEJ1297, doi:10.3150/20-BEJ1297.

Majka, M.B., Mijatović, A., Szpruch, Ł., 2020. Nonasymptotic bounds for sampling algorithms without log-concavity. The Annals of Applied Probability 30, 1534–1581.

Monmarché, P., 2021. High-dimensional mcmc with a standard splitting scheme for the underdamped langevin diffusion. Electronic Journal of Statistics 15, 4117–4166.

Mou, W., Flammarion, N., Wainwright, M.J., Bartlett, P.L., 2022. Improved bounds for discretization of langevin diffusions: Near-optimal rates without convexity. Bernoulli 28, 1577–1601.

Neufeld, A., En, M.N.C., Zhang, Y., 2022. Non-asymptotic convergence bounds for modified tamed unadjusted langevin algorithm in non-convex setting. arXiv preprint arXiv:2207.02600 .

Nguyen, D., Dang, X., Chen, Y., 2021. Unadjusted langevin algorithm for non-convex weakly smooth potentials. arXiv preprint arXiv:2101.06369 .

Raginsky, M., Rakhlin, A., Telgarsky, M., 2017. Non-convex learning via Stochastic Gradient Langevin Dynamics: a nonasymptotic analysis, in: Conference on Learning Theory, pp. 1674–1703.

Sabanis, S., Zhang, Y., 2019. Higher order Langevin Monte Carlo algorithm. Electronic Journal of Statistics 13, 3805–3850.

Vempala, S., Wibisono, A., 2019. Rapid convergence of the unadjusted langevin algorithm: Isoperimetry suffices. Advances in neural information processing systems 32.

Zhang, M., Chewi, S., Li, M.B., Balasubramanian, K., Erdogdu, M.A., 2023a. Improved discretization analysis for underdamped langevin monte carlo. arXiv preprint arXiv:2302.08049 .

Zhang, Y., Akyildiz, Ö.D., Damoulas, T., Sabanis, S., 2023b. Nonasymptotic estimates for stochastic gradient langevin dynamics under local conditions in nonconvex optimization. Applied Mathematics & Optimization 87, 25.