High-dimensional MCMC with a standard splitting scheme for the underdamped Langevin diffusion.[*]

Pierre Monmarché

Sorbonne Université, 4 place Jussieu,
75005 Paris, France
e-mail: pierre.monmarche@sorbonne-universite.fr

Abstract: The efficiency of a Markov sampler based on the underdamped Langevin diffusion is studied for high dimensional targets with convex and smooth potentials. We consider a classical second-order integrator which requires only one gradient computation per iteration. Contrary to previous works on similar samplers, a dimension-free contraction of Wasserstein distances and convergence rate for the total variance distance are proven for the discrete time chain itself. Non-asymptotic Wasserstein and total variation efficiency bounds and concentration inequalities are obtained for both the Metropolis adjusted and unadjusted chains. In particular, for the unadjusted chain, in terms of the dimension d and the desired accuracy ε, the Wasserstein efficiency bounds are of order $\sqrt{d}/\varepsilon$ in the general case, $\sqrt{d/\varepsilon}$ if the Hessian of the potential is Lipschitz, and $d^{1/4}/\sqrt{\varepsilon}$ in the case of a separable target, in accordance with known results for other kinetic Langevin or HMC schemes.

1. Introduction

The Langevin diffusion (also called underdamped or kinetic Langevin diffusion) is the Markov process on $\mathbb{R}^d \times \mathbb{R}^d$ that solves the SDE

$$\begin{cases} \mathrm{d}X_t &= V_t \mathrm{d}t \\ \mathrm{d}V_t &= -\nabla U(X_t)\mathrm{d}t - \gamma V_t \mathrm{d}t + \sqrt{2\gamma}\mathrm{d}B_t \end{cases} \tag{1}$$

where B is a standard d-dimensional Brownian motion, $U \in \mathcal{C}^2(\mathbb{R}^d)$ is called the potential (or log-likelihood) and $\gamma > 0$ is a friction (or damping) parameter. Under general assumptions, it is known to be ergodic with respect to the Gibbs measure with Hamiltonian $H(x,v) = U(x) + |v|^2/2$, namely the probability measure π on $\mathbb{R}^{2d}$ with density proportional to $\exp(-H(x,v))$. It is a central subject of study in a wide variety of domains, at the intersection of probabilities,

PDE and statistical physics, modeling e.g. plasmas or stellar systems. In this work we focus on the use of the Langevin diffusion in Markov Chain Monte Carlo (MCMC) algorithms in order to estimate expectations with respect to π.

This process has been used for decades in molecular dynamics (MD), in particular due to physical motivations related to the motion of particles in classical physics (see [25, 38, 55, 30, 32, 31, 51, 8, 7, 6] and references within). In this context, the objective is not necessarily to compute only statistical averages, but also dynamical properties, like diffusion coefficients, transition times, reaction paths, etc., so that it is crucial to sample a trajectory of (a discrete-time approximation of) the Langevin diffusion, and not of another process with the same statistical equilibrium. More recently, the Langevin diffusion has gained some interest in the computational statistics and learning community [12, 11, 33, 14, 59], in which case there is no preferred underlying dynamics and only statistical averages with respect to π matter. The present work follows this viewpoint. Besides, in this context, the original question is to compute expectations with respect to $\pi_1 \propto \exp(-U)$ the first d-dimensional marginal of π, and the velocity V is added as an auxiliary variable in order to enhance the sampling. This is called a lifted MCMC method.

We will focus on the case where U is m-convex and L-smooth (see Assumptions (∇**Lip**) and (**Conv**) below). These are usual conditions in the statistics community to compare MCMC methods, the objective being then to obtain non-asymptotic explicit estimates in term of the dimension d, see e.g. [18, 12, 14]. These conditions are far from natural for MD applications where, except for very particular harmonic toy models, the potentials are highly non-convex and singular due to strong short-range repulsion of nuclei. On the other hand, and contrary to the recent works in computational statistics, we will study a discretization scheme of (1) obtained by a splitting procedure, as commonly used in MD [8, 30, 7, 6, 51] and implemented in all MD codes.

The study for MCMC purpose of a discretization of the Langevin diffusion for a convex and smooth potential has recently been conducted in [12, 14, 33, 59]. Concerning Metropolis adjusted chains, some non-asymptotic results have recently been established in [10, 20, 36, 37, 5] for HMC samplers or other Metropolis-adjusted algorithms. Let us give an informal summary of our contribution with respect to these works (a more detailed and quantitative discussion is provided in Section 3.4 after our results have been rigorously stated).

- As far as non-asymptotic efficiency bounds are concerned, standard splitting discretisation schemes (such as the OBABO chain studied in the present work, see Section 2.1) had not been studied yet. Contrary to the first order scheme of [12, 14, 11, 59], the OBABO sampler is a second order scheme (in the time-step). Like classical schemes, it only requires one computation of ∇U per step, contrary to the second order scheme of [14, 13] that also requires a computation of $\nabla^2 U$. It belongs to the class of splitting schemes considered in [30], but [30] only provides a formal expansion of the invariant measure in the time-step, and only for two particular splitting (and not the one studied here). A second order scheme

with the same computational complexity is also introduced in [33], and we obtain similar results in term of efficiency. Contrary to the schemes studied in [33, 30], the OBABO scheme is naturally connected to HMC type samplers, and in particular the discretization bias at equilibrium is only due to a deterministic approximation of the Hamiltonian dynamics by a Verlet (or leapfrog) scheme, which has some nice consequences in the analysis.

- Contrary to the previous works on other schemes based on the under-damped Langevin diffusion, instead of establishing long-time convergence estimates for the continuous-time process and then conducting an analysis of the discretization error, we directly prove a Wasserstein contraction for the discrete time Markov chain itself. This classically yields nice concentration results (and thus non-asymptotic confidence intervals) for empirical averages, see Theorem 6, which is not the case in previous works. Then, we study the bias at equilibrium separately. Notice that, ultimately, gathering these two parts yields efficiency bounds that are similar to those obtained with the method that relies only on the convergence rate of the continuous-time process, nevertheless this work was initially motivated by the theoretical question considered in [48] of obtaining convergence rates for (discrete-time) MCMC samplers in high dimension. This issue is not addressed with the basic method. Finally, with our method, we obtain that empirical averages of the discrete-time chain have a (biased) long-time limit (again, this is Theorem 6), which is a first step for using Romberg interpolation techniques, see e.g. [19, 1] (this issue is postponed to a future work, see [39]).
- We establish a Wasserstein/total variation regularization property. To our knowledge, this is the first result of this kind for a Markov chain based on an hypoelliptic non-elliptic diffusion. As a consequence, contrary to [12, 14, 59], we obtain results for the total variation distance. However, we acknowledge that schemes that are based on the exact solving of the SDE (1) but with $\nabla U(X_t)$ replaced by $\nabla U(X_{\delta \lfloor t/\delta \rceil})$ where $\delta > 0$ is the time-step (which is the case in [12, 14, 33, 59] and not of our scheme) have Gaussian transitions and thus enjoy a similar regularization property, although it hasn't been stated in previous works. In [33], bounds are obtained for the Kullback-Leibler divergence, which is stronger, but with hypocoercive PDE techniques, while we use an elementary coupling approach. Adapting their method to our case is not straightforward, as one iteration of our scheme is not based on solving (1) with fixed forces.
- In order to see that second order schemes are more efficient than first order ones, the gradient Lipschitz assumption is not sufficient, some information is required on higher order derivatives of U. For this reason, in [14, 33], the assumption that $\nabla^2 U$ is Lipschitz is enforced. We work under a weaker condition (see Assumption $(\nabla^2 \mathbf{pol}(\ell))$ below), that holds for instance if $\|\nabla^{(k)} U\|_\infty < +\infty$ for some $k > 2$. When $\nabla^2 U$ is Lipschitz, we obtain efficiency bounds similar to [14, 33].
- We establish non-asymptotic efficiency bounds and confidence intervals for

the Metropolis-adjusted OBABO scheme, which is an HMC sampler with a particular choice of parameters. This is the first results of this kind for a Metropolis-adjusted sampler based on the (kinetic) Langevin equation. By comparison, the recent works [10, 36, 37, 5] are concerned with the classical HMC sampler, with complete refreshment of velocities separated by long run of the Hamiltonian dynamics (and do not provide confidence intervals). Also, contrary to [10, 20, 36, 37], we do not study the case of so-called warm start for the initial distribution, and we obtain non-asymptotic bounds that involve only some moments of the initial distribution (similarly to the unadjusted case).

We are mainly interested in the dependency in the dimension d, the time-step δ and the accuracy ε of the explicit estimates. Concerning the dependency on other parameters like m or L, in practice, all our estimates could be improved by considering a preconditioned algorithm as in [59] and/or by taking into account explicitly a Gaussian part of U, i.e. decomposing $U(x) = x \cdot S^{-1}x/2 + \tilde{U}(x)$ with some covariance matrix S and potential $\tilde{U}$, see e.g. [47].

The article is organized as follows. The OBABO sampler and its Metropolis-adjusted counterpart are presented in Section 2, together with some general considerations on Wasserstein distances. Our main results are stated in Section 3, decomposed as: results on the long-time behavior of the OBABO chain (Section 3.1); on the equilibrium bias (Section 3.2); on the Metropolis-adjusted sampler (Section 3.3). The results are discussed and related to other works in Section 3.4. Section 4 contains the proofs for the long-time behaviour of the OBABO sampler, Section 5 the proofs for the equilibrium bias and Section 6 the proofs for the Metropolis-adjusted chain.

2. Definitions

2.1. A second-order scheme for the Langevin diffusion

We present the Markov chain which is the main object of study of this work, obtained as a time-discretization of the Langevin diffusion (already considered by Horowitz in [25], see also [8] and references within). First, we simply define the transition of the Markov chain, and the rest of the section is devoted to the motivation and discussion of this definition. Let $\delta, \gamma > 0$ be respectively the time-step and friction coefficient, and denote $\eta = e^{-\delta\gamma/2}$. 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

$$\begin{cases} x_1 & = & x_0 + \delta\left(\eta v_0 + \sqrt{1-\eta^2}G\right) - \frac{\delta^2}{2}\nabla U(x_0) \\ v_1 & = & \eta^2 v_0 - \frac{\delta\eta}{2}\left(\nabla U(x_0) + \nabla U(x_1)\right) + \sqrt{1-\eta^2}\left(\eta G + G'\right) \end{cases} \tag{2}$$

where G and G' are two independent d-dimensional standard (mean 0 variance I_d) Gaussian random variables. Denote by P the Markov transition operator associated to the chain, i.e. $P\varphi(x_0, v_0) = \mathbb{E}(\varphi(x_1, v_1))$ for all measurable bounded φ.

A first very important remark is that, although it seems that in one iteration, ∇U has to be computed for two different values of x, this is in fact only true for the first iteration, afterward $\nabla U(x_n)$ is already known from the previous iteration and only $\nabla U(x_{n+1})$ is computed. A second remark is that, contrary to the schemes considered in e.g. [12, 14, 33, 59] (where the transition is obtained by following for a time δ the continuous-time process (1) but with constant forces $\nabla U(x_t) = \nabla U(x_0)$), the law of (x_1, v_1) conditionally to (x_0, v_0) is not a Gaussian distribution, because of the term $\nabla U(x_1)$ in v_1 (except in the particular case where U is quadratic).

The scheme makes more sense when it is decomposed as the following successive steps:

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

$$v_{1/2} = v_0' - \delta/2 \nabla U(x_0) \tag{B}$$

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

$$v_1' = v_{1/2} - \delta/2 \nabla U(x_1) \tag{B}$$

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

Here, we use the notations O, B and A of [30], referring respectively to the damping (or friction/dissipation, or partial resampling), the acceleration and the free transport parts of the dynamics. So, in the rest of the article we will refer to the Markov chain $(x_n, v_n)_{n \in \mathbb{N}}$ as the OBABO sampler which, contrary to some of its other names like second order Langevin or midpoint Euler-Verlet-Midpoint Euler [8, 25, 31], has the advantage to give a compact yet explicit description of the specific scheme (in particular by comparison with other second order schemes for the underdamped Langevin diffusion).

Note that the parts BAB of the scheme is the classical velocity Verlet algorithm for the Hamiltonian dynamics. More generally, the Verlet, OBABO, or BAOAB (from [30]) algorithms all shares a common palindromic form since they are obtained from a Strang/Trotter splitting of the generator of the Langevin diffusion. We now give a brief and informal presentation of the latter, and refer to [30, 31] or [44, Section 5] for more details. The transition semigroup P_t associated to (1) is informally of the form $e^{t\mathcal{L}}$ where the infinitesimal generator $\mathcal{L}$ can be decomposed as $\mathcal{L} = \mathcal{L}_A + \mathcal{L}_B + \mathcal{L}_O$ with

$$\mathcal{L}_A = v \cdot \nabla_x \qquad \mathcal{L}_B = -\nabla U(x) \cdot \nabla_v \qquad \mathcal{L}_O = -\gamma v \cdot \nabla_v + \gamma \Delta_v .$$

Using that $e^{\delta(a+b)} = e^{\delta/2a} e^{\delta b} e^{\delta/2a} + o(\delta^2)$, we get formally a second order approximation of P_δ by

$$P_\delta = e^{\delta/2\mathcal{L}_O} e^{\delta/2\mathcal{L}_B} e^{\delta\mathcal{L}_A} e^{\delta/2\mathcal{L}_B} e^{\delta/2\mathcal{L}_O} + o(\delta^2) .$$

The operator on the right-hand side is exactly the transition kernel of the OBABO sampler. At least formally, the fact that the transition kernel of the chain approximates the transition semi-group of the continuous process up to second order in δ can be shown to yield a similar order for the error on the invariant

measure, see [30, 44] for general discussions and more details and Proposition 12 below for a rigorous and quantitative statement in the particular OBABO case. More precisely, in Proposition 12, we get an error term on the invariant measure of order δ^2 for the Wasserstein and total variation distances, to compare to e.g. [14, Theorem 2] or the proof of [12, Theorem 1] where the error is of order δ for the Wasserstein distance (see Section 3.4 for the consequence of this on the efficiency bounds). For convex and smooth potentials, for a suitable choice of γ, the continuous-time process (1) converges to equilibrium at a rate that is independent of the dimension, so that the error estimates obtained in [12, 14, 33, 59] depends on the dimension only because of the error on the invariant measure (which requires the time-step to be small enough). By replacing a first order discretization scheme by a second order one, we expect to improve the estimates accordingly, which is indeed what is observed in [14, 33] where other second-order schemes are considered, and in the present work (again, we refer to Section 3.4).

The reason we chose to study the OBABO algorithm rather than the BAOAB one or other similar scheme is the following. The OBABO scheme contains at its core a (deterministic) Verlet part BAB, which is time reversible (in the physicist sense, i.e. up to a reflection of the velocity). This fact classically leads to a natural Metropolized version of the algorithm, of Hamiltonian (or Hybrid) Monte Carlo (HMC) type, presented below. This Metropolis-adjusted algorithm is interesting in itself, and we use it as an auxiliary tool in the study of the bias of the OBABO chain. Besides, an unintended benefit of our choice of integrator is that, for the OBABO scheme, establishing a Wasserstein/total variation regularization property (Proposition 3) is rather straightforward because of the particular location of the noise in the scheme. That being said, it should be possible to adapt most (if not all) of our arguments to the BAOAB or other similar second order schemes like in [33]. Finally, remark that the BAOAB scheme is promoted in [30] because, in the overdamped limit $\gamma \to +\infty$, the invariant measure is correctly sampled at fourth order in the time-step. However, as $\gamma \to +\infty$, the convergence rate of the chain goes to zero, so it is not clear that this argument is decisive in the choice of the integrator (after a suitable scaling of time, the BAOAB converges as $\gamma \to +\infty$ to a discretization scheme of the overdamped Langevin diffusion which has a correct equilibrium up to second order).

Recall the goal of the algorithm is to compute averages with respect to π_1 the first d-dimensional marginal of π. As such, we are not constrained for the equilibrium of the velocities, and typically we could chose a symmetric Gaussian distribution with variance different from 1. Considering the Hamiltonian $H(x,v) = U(x) + |v|^2/(2\sigma^2)$, we would have a new parameter σ to optimize, which could improve the results. Nevertheless, in that case, we can always chose $\sigma = 1$ by the change of variables $v \leftarrow v/\sigma$, $\delta \leftarrow \delta\sigma$ and $\gamma \leftarrow \gamma/\sigma$. As a consequence, without loss of generality, we take $\sigma = 1$.

As mentioned above, the particular structure of the scheme has some nice implications for the theoretical study. Of course, in practice, if we are only interested in computing averages that only involve the position x_n, then the scheme is equivalent to alternate a Verlet step BAB and a partial refreshment

O^2 (which is simply O but with a full rather than half time-step, i.e. η is replaced by $e^{-\delta\gamma}$), since $(OBABO)^n = O(BABO^2)^{n-1}BABO$ and the first and last O have no effect on the position.

Finally, let us mention that our analysis can be straightforwardly adapted to get results on the $O(BAB)^k O$ scheme for some fixed $k \in \mathbb{N}_*$ independent from δ (i.e. k Verlet steps are performed between the partial refreshments). Indeed, as δ vanishes, the corresponding chain still converges toward the continuous time process (1), which is the main ingredient of our results concerning long-time convergence, and similarly there is no particular difficulty to adapt the results concerning discretization errors. For fixed parameters k and γ, this extension does not yield any improvement in terms of scaling in δ and d (also, note that one iteration of $O(BAB)^k O$ requires k computations of ∇U) and thus we don't consider it to avoid an unnecessary inflation of notations.

2.2. *Metropolis-adjusted algorithm*

The O parts of the scheme leaves invariant the standard Gaussian distribution for the velocities, hence the target π. This is not the case of the Verlet part BAB because of the numerical approximation. As a consequence, π is not invariant for the OBABO chain. Nevertheless, this is classically fixed by adding a Metropolis accept/reject step on this part of the dynamics, which gives an HMC algorithm [25, 45]. We now detail this. In the following we denote

$$\Phi_V(x,v) \;=\; \begin{pmatrix} x + \delta v - \frac{1}{2}\delta^2 \nabla U(x) \\ v - \frac{1}{2}\delta \left(\nabla U(x) + \nabla U \left(x + \delta v - \frac{1}{2}\delta^2 \nabla U(x) \right) \right) \end{pmatrix},$$

so that the steps BAB read $(x_1, v_1') = \Phi_V(x_0, v_0')$. Denote also $\Phi_R(x,v) = (x,-v)$. It is then straightforward to check that the Verlet algorithm is time-reversible, i.e. $\Phi_R \Phi_V \Phi_R \Phi_V(x,v) = (x,v)$ for all $x,v \in \mathbb{R}^d$. As a consequence, the deterministic proposal kernel

$$q\left((x,v), \mathrm{d}y\mathrm{d}w\right) \;=\; \delta_{\Phi_R \Phi_V(x,v)}(\mathrm{d}y\mathrm{d}w)$$

is symmetric. Denote P_{MH} the Markov transition operator of the Metropolis-Hastings algorithm with proposal kernel q. Considering the acceptance probability $\alpha(x_0, v_0) := \exp\left(-\left(H\left(\Phi_V(x_0,v_0)\right) - H(x_0,v_0)\right)_+\right)$, a transition of the corresponding chain is given by:

$$\begin{aligned} (x',v') &= \Phi_V(x_0,v_0) \\ (x_1,v_1) &= (x',-v') & \text{if } W \leqslant \alpha(x_0,v_0) \\ &= (x_0,v_0) & \text{otherwise,} \end{aligned}$$

where W is uniformly distributed over $[0,1]$. We used that $H(x',v') = H(x',-v')$. By construction of the Metropolis-Hastings algorithm, P_{MH} is reversible (in the probabilistic sense) with respect to π. Moreover, π is also invariant by the transformation $v \leftrightarrow -v$, namely $\pi P_R = \pi$ where $P_R \varphi(x,v) = \varphi(x,-v)$. Denote $P_{MV} = P_R P_{MH}$, which is similar to a Verlet transition BAB of the previous

section except for the accept/reject step. Remark that, in case of rejection, the velocity is reflected. This is invisible in the classical HMC algorithm for which the velocity is fully resampled after each accept/reject step (which will not be the case here).

Denote by P_O the Markov transition kernel on $\mathbb{R}^d \times \mathbb{R}^d$ corresponding to the O transition, i.e.

$$(x_1, v_1) \;=\; \left(x_0, \eta v_0 + \sqrt{1 - \eta^2} G\right) \qquad \text{with} \qquad G \sim \mathcal{N}\left(0, I_d\right) .$$

In other words, P_O is given for any bounded observable φ by

$$P_O \varphi(x, v) \;=\; (2\pi)^{-d/2} \int_{\mathbb{R}^d} \varphi\left(x, \eta v + \sqrt{1 - \eta^2} w\right) e^{-|w|^2/2} \mathrm{d}w .$$

The Metropolis-adjusted version of the OBABO algorithm, introduced in [25], is the Markov chain with transition operator

$$P_M \;=\; P_O P_{MV} P_O .$$

By construction, π is invariant for P_M. Although π is reversible for P_O, P_R and P_{MH}, this property is not conserved by composition and π is not reversible with respect to P_M. Besides, using that $P_R P_O P_R = P_O$ and $P_R^2 = Id$, we see that that the adjoint operator $P_M^* = P_O P_{MH} P_R P_O$ in $L^2(\pi)$ is given by $P_M^* = P_R P_M P_R$, i.e. the process is reversible in the physics sense, up to a reflection of the velocity. We won't use this property in the article but we note that it allows to get an L^2 spectral gap from a contraction in an other distance, like a Wasserstein distance or the H^1 norm, as in the reversible case, see e.g. [15].

The operator P_M corresponds to the following transition:

$$\begin{aligned} v_0' &= \eta v_0 + \sqrt{1 - \eta^2} G \\ v_{1/2} &= v_0' - \delta/2 \nabla U(x_0) \\ \tilde{x}_1 &= x_0 + \delta v_{1/2} \\ \tilde{v}_1 &= v_{1/2} - \delta/2 \nabla U(\tilde{x}_1) \\ (x_1, v_1') &= (\tilde{x}_1, \tilde{v}_1) \qquad \text{if } W \leqslant \alpha(x_0, v_0') \\ &= (x_0, -v_0') \qquad \text{otherwise} \\ v_1 &= \eta v_1' + \sqrt{1 - \eta^2} G' . \end{aligned}$$

Remark that, in case of acceptation (namely if $W \leqslant \alpha(x_0, v_0')$) then this is exactly the OBABO transition. Similar Metropolis-adjusted discretizations of the Langevin diffusion have been studied in [51, 7, 6, 47], see also references within.

Due to the particular scaling of the partial refreshment mechanism, this HMC chain converges as δ vanishes to the Langevin diffusion rather than to the continuous-time Randomized HMC (see [15] and references within) where dissipativity in the velocity is ensured by a (possibly partial) resampling of the velocities at constant rate, rather than an Ornstein-Uhlenbeck diffusion.

Following motivations similar to the unadjusted case, we call this specific Metropolis-adjusted splitting scheme the OM(BAB)O chain (the M($\cdot$) standing for Metropolis).

2.3. Wasserstein distances

Let ρ be a distance on $\mathbb{R}^d$. For $p \in [1, \infty)$, denote by $\mathcal{P}_{p,\rho}(\mathbb{R}^d)$ the set of probability distributions on $\mathbb{R}^d$ having a finite p-th moment for ρ, i.e. $\nu \in \mathcal{P}_{p,\rho}(\mathbb{R}^d)$ if

$$\int_{\mathbb{R}^d} \rho^p(0, x)\nu(\mathrm{d}x) \;<\; +\infty\,.$$

For $p \in [1, \infty]$ the Wasserstein distance $\mathcal{W}_{p,\rho}$ on $\mathcal{P}_{p,\rho}(\mathbb{R}^d)$ is given for $\nu, \mu \in \mathcal{P}_{p,\rho}(\mathbb{R}^d)$ by

$$\mathcal{W}_{p,\rho}(\nu, \mu) \;=\; \inf_{\eta \in \Gamma(\nu, \mu)} \|\rho\|_{L^p(\eta)}\,,$$

where $\Gamma(\nu, \mu)$ is the set of transference plan of ν and μ, namely is the set of probability distributions on $\mathbb{R}^d \times \mathbb{R}^d$ with first marginal ν and second marginal μ. In other words,

$$\mathcal{W}_{p,\rho}^p(\nu, \mu) \;=\; \inf_{X \sim \nu, Y \sim \mu} \mathbb{E}\left(\rho^p(X, Y)\right)\,,$$

although this is only a formal definition since the infimum is not taken on a proper set. If (X, Y) is a random variable with law $\eta \in \Gamma(\nu, \mu)$, we say that (X, Y) is a coupling of ν and μ. Implicitly, $\mathcal{W}_\rho$ means $\mathcal{W}_{1,\rho}$, and $\mathcal{W}_p$ means $\mathcal{W}_{p,\rho}$ with $\rho(x, y) = |x - y|$. In fact, unless otherwise specified, in $\mathbb{R}^d$, the name *Wasserstein distances* usually refers to cases where ρ is equivalent to the Euclidean metric.

Besides, another interesting case is the discrete metric $\rho(x, y) = 2\mathbb{1}_{x \neq y}$, in which case the associated $\mathcal{W}_1$ distance is the total variation norm

$$\mathcal{W}_\rho(\nu, \mu) \;=\; \|\nu - \mu\|_{TV} \;=\; \inf_{X \sim \nu, Y \sim \mu} 2\mathbb{P}\left(X \neq Y\right)\,.$$

Given a Markov operator Q acting on $\mathcal{P}_p(\mathbb{R}^d)$, it is easily seen by conditioning on the initial condition that, for $C > 0$, the proposition

$$\forall \nu, \mu \in \mathcal{P}_p(\mathbb{R}^d)\,, \qquad \mathcal{W}_{p,\rho}(\nu Q, \mu Q) \;\leqslant\; C\mathcal{W}_{p,\rho}(\nu, \mu)$$

is equivalent to

$$\forall x, y \in \mathbb{R}^d\,, \qquad \mathcal{W}_{p,\rho}(Q(x, \cdot), Q(y, \cdot)) \;\leqslant\; C\rho(x, y)\,.$$

Besides, the Jensen's inequality implies that $\mathcal{W}_{p,\rho} \leqslant \mathcal{W}_{q,\rho}$ whenever $p \leqslant q$, which together with the previous remark implies that a contraction of the $\mathcal{W}_{q,\rho}$ distance for some q induces a contraction of $\mathcal{W}_{p,\rho}$ for all $p \leqslant q$.

For a distance ρ on $\mathbb{R}^d$ that is equivalent to the Euclidean metric, from [58, Corollary 5.22 and Theorem 6.18], there always exists an optimal coupling for $\mathcal{W}_{p,\rho}$, i.e. for $\nu, \mu \in \mathcal{P}_p(\mathbb{R}^d)$ there exists a coupling (X, Y) of μ on ν on some probability space Ω such that $\mathcal{W}_{p,\rho}^p(\nu, \mu) = \mathbb{E}\left(\rho^p(X, Y)\right)$, and moreover $(\mathcal{P}_p(\mathbb{R}^d), \mathcal{W}_{p,\rho})$ is a Banach space.

3. Main results

As already mentioned, we are principally interested in the smooth and convex case, corresponding to the following conditions.

Assumption (∇**Lip**). *There exists $L > 0$ such that for all $x, y \in \mathbb{R}^d$,*

$$|\nabla U(x) - \nabla U(y)| \leqslant L|x - y|.$$

Assumption (**Conv**). *There exists $m > 0$ such that for all $x, y \in \mathbb{R}^d$,*

$$(x - y) \cdot (\nabla U(x) - \nabla U(y)) \geqslant m|x - y|^2.$$

Under Assumptions (∇**Lip**) and (**Conv**), U admits a unique global minimum $x_\star$ and for all $x \in \mathbb{R}^d$,

$$|\nabla U(x)| \leqslant L|x - x_\star| \qquad \text{and} \qquad m|x - x_\star|^2 \leqslant 2U(x) \leqslant L|x - x_\star|^2. \qquad (3)$$

The two conditions don't have the same status. The condition (∇**Lip**) will be enforced in all the article. It classically implies the stability of the OBABO scheme for δ small enough (this will be a consequence of our results). Approximating diffusions with non-Lipschitz coefficient leads to various difficulties and can lead to truncated algorithm, see e.g. [7] and references within. The condition (**Conv**) gives a simple condition to obtain nice explicit convergence rates, and will not be used in most of the bias error analysis.

3.1. Dimension free convergence rates

Recall P is the Markov operator corresponding to the OBABO chain introduced in Section 2.1. In this section, we state our results obtained on the long-time behavior and regularization of this chain by itself, i.e. without referring to the continuous-time limit.

Under the the convex/smooth assumption, and for a suitable choice of the damping parameter, the Langevin diffusion is contractive, in the sense that its deterministic drift contracts the distances [4, 12]. Hence, starting from two different initial states, the parallel coupling (i.e. considering the same Brownian noise) of two processes yields a deterministic contraction. The problem is quite more involved in a non-convex case where one has to take advantage of the noise to overcome potential barriers, in which case a combination of parallel and mirror (using reflected Brownian noises) couplings and a suitable concave transformation of the distance can give a contraction in some modified $\mathcal{W}_1$ distance, see [21, 11, 13]. An alternative argument in the non-convex case to get a $\mathcal{W}_2$ hypocoercive contraction is to combine an entropic hypocoercive contraction with a Wasserstein/entropy regularization, as in [22, Theorem 2] (this gives sharper results than coupling arguments in the low temperature regime, see [42]).

In the present convex case, the OBABO chain being an approximation of the Langevin diffusion, a simple parallel coupling does the job, and we get the following.

Theorem 1. *Under Assumptions* (∇**Lip**) *and* (**Conv**), *suppose moreover that* $\gamma \geqslant 2\sqrt{L}$ *and that* $\delta \leqslant m/(33\gamma^3)$. *Then, for all* $p \geqslant 1$, $n \in \mathbb{N}$ *and all* $\nu, \mu \in \mathcal{P}_p(\mathbb{R}^{2d})$,

$$\mathcal{W}_p\left(\nu P^n, \mu P^n\right) \ \leqslant \ K_1 \left(1 - \delta\kappa\right)^{n/2} \mathcal{W}_p(\nu, \mu)$$

with

$$K_1 \ = \ \sqrt{3}\max\left(L^{1/2}, L^{-1/2}\right), \qquad \kappa \ = \ \frac{m}{3\gamma}.$$

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

This is proven in Section 4.1, see Corollary 20.

Remark 2. *The restriction on γ is consistent with all the works that studied the contraction of the parallel coupling for the continuous-time Langevin diffusion [4, 12, 14, 59] (see also [15] for the Randomized HMC algorithm, since this is exactly the same computations after taking expectations). In fact, as proven in [43, Proposition 4], for the continuous-time Langevin process, a Wasserstein contraction as in the proof of Theorem 1 holds for all potentials U satisfying* (∇**Lip**) *and* (**Conv**)*if and only if $L - m < \gamma(\sqrt{L} + \sqrt{m})$. When $m \ll L$ our condition $\gamma \geqslant 2\sqrt{L}$ is thus similar to the optimal one up to a factor 2. In fact, it is clear that combining the proof of [43, Proposition 4] and of Theorem 1, a Wasserstein contraction holds for the OBABO chain under the same optimal condition on γ as in the time-continuous case, but then κ and the condition on δ are less nice and this doesn't improve the convergence rate in terms of the dependency on m and L.*

The choice $\gamma = 2\sqrt{L}$ maximizes our bound on the contraction rate. In [14], the optimal rate for the continuous-time process is proven to be m/γ, obtained with the choice $\gamma = \sqrt{m + L}$, which means in the regime $m \ll L$ we are simply missing a factor 6 from the optimal bound. Again, this factor 6 can of course be reduced to get arbitrarily close to the optimal bound simply by following the proof and keeping sharper constants, to the cost of a stronger condition on δ, and without changing the order in terms of m and L.

Besides, as $\gamma \to +\infty$, we get a convergence rate of order $1/\gamma$, which is sharp and well-known in the case of the continuous-time Langevin diffusion.

It is clear that a condition on δ is needed, since the process is not even stable when δ is too large. Here the condition $\delta \leqslant m/(33\gamma^3)$ is not sharp, it is used to get a simple expression. Anyway, we have in mind the regime where m, L, γ are independent from the dimension and, as we will see in the next section, δ should vanish as $d \to +\infty$, so that the condition on δ is always satisfied in high dimension, which is why we chose simplicity over sharpness to state this condition.

As an hypoelliptic (non-elliptic) diffusion, the continuous-time Langevin diffusion (1) enjoys many nice regularization properties. In particular, under Assumption (∇**Lip**), [23, Corollary 4.7(2) and Remark 4.1] implies an instantaneous $\mathcal{W}_2$/relative entropy regularization, which together with the Pinsker

inequality implies that

$$\|\nu P_t - \mu P_t\|_{TV} \ \leqslant \ \frac{C}{\min(1, t^{3/2})} \mathcal{W}_2(\nu, \mu)$$

for some $C > 0$ for all $\nu, \mu \in \mathcal{P}_2(\mathbb{R}^{2d})$ and $t > 0$, where $(P_t)_{t \geqslant 0}$ is the transition semigroup associated to (1). For elliptic diffusions the result would typically hold with $\sqrt{t}$, but for the Langevin diffusion, there is no direct noise in the position and thus $t^{N+1/2}$ is to be expected where N is the number of times one has to consider Lie brackets to fulfill Hörmander's condition (here $N = 1$).

We now state a similar result, but at the level of the discrete time chain P. Convexity is not assumed. We are not aware of any similar result for a Markov chain based on the discretization of an hypoelliptic non-elliptic diffusion (although, as mentioned in the introduction, it holds indeed in many cases).

Proposition 3. *Under Assumption* (∇**Lip**), *for all* $\nu, \mu \in \mathcal{P}_1(\mathbb{R}^{2d})$,

$$\|\nu P - \mu P\|_{TV} \ \leqslant \ \frac{K_2}{\delta^{3/2}} \mathcal{W}_1(\nu, \mu) \,,$$

with

$$K_2 \ = \ \sqrt{\frac{\delta}{2\pi \left(1 - \eta^2\right)}} \max \left(2 + 3\delta^2 L/2, 3\delta\right) \ \xrightarrow[\delta \to 0]{} \ \sqrt{\frac{2}{\pi \gamma}} \,.$$

This is proven in Section 4.2, see Proposition 22.

Remark 4. *We retrieve for* $\delta \to 0$ *the expected* $\delta^{3/2}$. *In fact, more precisely, as can be seen in Proposition 22 below, we get the scaling* $\delta^{3/2}$ *for the positions and* $\sqrt{\delta}$ *for the velocities.*

Combining the two previous results naturally yields a convergence rate for the total variation distance. Following [50, 48], we consider the total variation convergence rate of P in the sense of:

$$r_*(P) \ := \ \exp \left(\sup_{z \in \mathbb{R}^{2d}} \limsup_{n \to +\infty} \frac{\ln \|\delta_z P^n - \pi_\delta\|_{TV}}{n} \right) \,.$$

An immediate consequence of Theorem 1 and of Proposition 3 is the following.

Corollary 5. *Under the conditions of Theorem 1, for all* $\nu, \mu \in \mathcal{P}_1(\mathbb{R}^{2d})$ *and all* $n \geqslant 1$,

$$\|\nu P^n - \mu P^n\|_{TV} \ \leqslant \ \frac{K_1 K_2}{\delta^{3/2}} \left(1 - \delta\kappa\right)^{(n-1)/2} \mathcal{W}_1(\nu, \mu) \,,$$

with K_1, κ *and* K_2 *given in Theorem 1 and Proposition 3. In particular,* $r_*(P) \leqslant 1 - \delta\kappa/2$.

The most classical settings to obtain bounds on the total variation convergence rate of a Markov chain is the combination of a Lyapunov (drift) and a local Doeblin (minoration) conditions. Nevertheless, the recent [48] rigorously

establishes that, in a variety of cases, no drift/minoration condition can yield a reasonable bound on $r_*(P)$ (here only one-step minoration conditions are concerned, of the form $\inf_{z \in \mathcal{C}} P(z, \cdot) \geqslant \varepsilon \nu$ for some probability ν and some $\varepsilon > 0$). More precisely, it is clear that the results of [48] apply to the OBABO sampler (essentially adapting the proof of [48, Proposition 15] for the MALA sampler), which proves that it is impossible to prove simply with a drift/minoration condition that, if δ scales polynomially with the dimension, then $r_*(P)$ is bounded away from 1 polynomially in the dimension. In Corollary 5, we have used the combination of a Wasserstein convergence and a Wasserstein/total variation regularization, called in [48, 49, 34] a one-shot coupling, which is indeed presented in [48] as an alternative to drift/minoration arguments.

We finish this section with a concentration result implied by the Wasserstein contraction of Theorem 1, or more precisely by the deterministic contraction at the core of the proof of this result (see Proposition 17 below). We say that a probability distribution ν on $\mathbb{R}^d$ satisfies a logarithmic Sobolev (or simply log-Sobolev) inequality with constant C if, for all bounded Lipschitz φ,

$$\int_{\mathbb{R}^d} \varphi^2 \ln\left(\varphi^2\right) \mathrm{d}\nu - \left(\int_{\mathbb{R}^d} \varphi^2 \mathrm{d}\nu\right) \ln\left(\int_{\mathbb{R}^d} \varphi^2 \mathrm{d}\nu\right) \leqslant C \int_{R^d} |\nabla\varphi|^2 \mathrm{d}\nu .$$

Such inequality yields many useful information, in particular related to concentration of measure, see e.g. [2, 28]. As a consequence of the Brascamp-Lieb inequality or of the Bakry-Emery curvature criterion [2], it is well known that, under the condition (**Conv**), the Gibbs measure π satisfies such an inequality with constant $2\max(1, 1/m)$. However, the invariant measure of P is not explicit. Nevertheless, the positive Wasserstein curvature (in the sense of [46, 26]) obtained in Theorem 1 (or more precisely in Proposition 17) yields a discrete-time Bakry-Emery condition, which in turn yields the following.

Theorem 6. *Under the conditions of Theorem 1, π_δ satisfies a log-Sobolev inequality with constant*

$$C_{lS} = \frac{6}{m} K_1^2 \gamma^2 \left(1 + \delta + \delta^2 L/2\right)^2 .$$

Moreover, if $\mu_0 \in \mathcal{P}(\mathbb{R}^{2d})$ satisfies a log-Sobolev inequality with constant $C' > 0$ and if $(Z_k)_{k \in \mathbb{N}}$ is an OBABO chain with $Z_0 \sim \mu_0$, then for all 1-Lipshitz functions φ on $\mathbb{R}^{2d}$, $n \geqslant 1$ and $u \geqslant 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^n \varphi(Z_k) - \pi_\delta(\varphi)\right| \geqslant u + \frac{2K_1}{n\kappa\delta}\mathcal{W}_1(\mu_0, \pi_\delta)\right)$$

$$\leqslant 2\exp\left(-\frac{nu^2\kappa\delta}{4C_{lS} + 4K_1^2 C'/(n\kappa\delta)}\right) .$$

Remark 7. *If μ_0 is a Dirac measure, it satisfies a log-Sobolev inequality with constant $C' = 0$. In any cases, we can bound $\mathcal{W}_1(\mu_0, \pi_\delta)$ by $\mathcal{W}_1(\mu_0, \pi) + \mathcal{W}_1(\pi, \pi_\delta)$, and then it is possible to chose $\mu_0 = \delta_{(\tilde{x}_\star, 0)}$ or $\delta_{\tilde{x}_\star} \otimes \mathcal{N}(0, I_d)$ (which satisfies a*

log-Sobolev inequality with constant 2*), where $\tilde{x}_\star$ is an approximation of $x_\star$ obtained through a deterministic optimization scheme. As can be seen for instance in Lemma [30], this yields an initial distance $\mathcal{W}_1(\pi, \mu_0)$ of order $\sqrt{d} + |x_\star - \tilde{x}_\star|$. The remaining term $\mathcal{W}_1(\pi, \pi_\delta)$ (that also controls $|\pi_\delta(\varphi) - \pi(\varphi)|$) is studied in the next section.*

Remark 8. *In fact, a $\mathcal{W}_1$ contraction (which does not necessarily implies a log-Sobolev inequality) is sufficient to get a similar result, see [16, Corollary 2.6] and [26].*

3.2. Equilibrium bias and efficiency

The invariant measure π_δ of P is not explicit. Since the OBABO sampler converges to the Langevin diffusion as δ vanishes, we expect π_δ to converge to π. There are many classical ways to quantify this; in view of the previous long-time convergence results, in this work we are naturally interested by an estimation of the Wasserstein and total variation distances between π_δ and π.

Recall the definition of the Verlet map Φ_V in Section 2.2, and consider P_V the corresponding transition operator, i.e. $P_V\varphi(x,v) = \varphi(\Phi_V(x,v))$ (equivalently, $\nu P_V = \Phi_V \sharp \nu$). From the contraction of Wasserstein distances and the Wasserstein/total variation distance regularization property, it is not difficult to get the following.

Proposition 9. *Under the conditions of Theorem 1, for all $p \geqslant 1$,*

$$\mathcal{W}_p\left(\pi_\delta, \pi\right) \;\leqslant\; \frac{2K_1}{\delta\kappa}\mathcal{W}_p\left(\pi, \pi P_V\right)$$

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

$$\|\pi_\delta - \pi\|_{TV} \;\;\leqslant\;\; \frac{K_1 K_2}{\delta^{3/2}}(1 - \delta\kappa)^{(n-1)/2}\mathcal{W}_1(\pi_\delta, \pi) + n\|\pi P_V - \pi\|_{TV}$$

where K_2 is given in Proposition 3.

This is proven in Section 5.

Remark 10. *Using that $(x', v') = \Phi_V(x,v)$ is a unitary change of coordinates (see [56, Problem 1.5]) and that it is reversible up to a reflection of the velocity (which preserves the Hamiltonian H) we see that*

$$\int_{\mathbb{R}^{2d}} \varphi\left(\Phi_V(x,v)\right)e^{-H(x,v)}\mathrm{d}x\mathrm{d}v \;=\; \int_{\mathbb{R}^{2d}} \varphi(x,v)e^{-H(\Phi_V(x,-v))}\mathrm{d}x\mathrm{d}v\,.$$

In other words, πP_V is the probability law with density proportional to $\exp(-H \circ \Phi_V \circ \Phi_R)$.

This leads us to the study of the Verlet step. In order to control the distance from πP_V to π, a natural way is to consider a Markov operator Q that leaves π invariant and to control the distance between πP_V and $\pi Q = \pi$ by constructing

a coupling of a transition of P_V with a transition of Q, starting from the same initial condition distributed according to π. If Q is close to P_V then we will be able to do so while keeping the two chains close. Since P_V has been obtained as the discretization of the Hamiltonian dynamics, a natural candidate for Q is Q_δ where $(Q_t)_{t \geqslant 0}$ is the transition semi-group of this deterministic flow, which indeed leaves π invariant. Comparing the discrete-time Markov chain with its continuous-time limit is indeed the main argument in similar works such as [12, 11, 18, 14]. However, we remark that another candidate is given by the Metropolis-adjusted Verlet transition $P_{MV} = P_R P_{MH}$ introduced in Section 2.2. Indeed, π is invariant for P_{MV}, and P_V and P_{MV} only differ by an accept/reject step. Comparing P_V with Q_δ or with P_{MV} leads to two different kind of results. On the one hand, P_V and Q_δ are deterministic, so there is no real coupling here, only a deterministic numerical error, in particular it behaves well with $\mathcal{W}_p$ for all $p \geqslant 1$. On the other hand, it is not suitable to bound the total variation distance since the probably that the deterministic flow and the numerical integrator are equal after one step is in general zero. On the contrary, by design, a transition of P_V and P_{MV} gives exactly the same result provided the step is accepted in P_{MV}, which enables to get information on the total variation distance. Nevertheless, in case of rejection, because of the velocity reflection, although the initial condition is the same for the two chains, the distance instantaneously get large. This makes the comparison with P_{MV} less and less efficient for bounding $\mathcal{W}_p$ as p increases. In a word, when comparing P_V with Q_δ, we get two points that are very close (but distinct) with probability 1 while, when comparing P_V with P_{MV}, we get two points that are equal with large probability but distant otherwise.

This analysis is conducted in Section 5. We now present its results.

A first remark is that the main difference between the OBABO sampler and the Markov chain studied in [12] is that the former is based on a second-order integrator, at least formally. Nevertheless, a Taylor expansion of the Verlet algorithm to an order larger than one requires informations on derivatives of U higher than two. As a consequence, relying only on Assumption (∇**Lip**), we can only get results similar to [12] for the Wasserstein distances.

Proposition 11. *Under the conditions of Theorem 1, for all $p \geqslant 1$,*

$$
\begin{aligned}
\mathcal{W}_p\left(\pi, \pi_\delta\right) &\leqslant \delta\sqrt{d + (p-2)_+}\frac{2K_1 K_3}{\kappa} \\
\|\pi - \pi_\delta\|_{TV} &\leqslant \delta d\left(2 + |\ln\left(\delta^3 d\right)|\right)\left(\frac{2K_1^2 K_2 K_3}{\kappa\sqrt{1 - \delta\kappa}} + \frac{K_4}{\kappa}\right)
\end{aligned}
$$

where K_2, K_3, K_4 are given in Propositions 3 and 31 (in particular K_3 and K_4 converge to L as δ vanishes).

Proposition 11 is a corollary of Propositions 9 and 31, as proven at the end of Section 5. The factor $|\ln\left(\delta^3 d\right)|$ may seem surprising, but cannot be avoided with our results.

In order to see the benefits of the second-order approximation, some conditions on higher derivatives of U have to be enforced. Ideally, we would like

conditions that are general, simple, and under which the OBABO sampler proves to be efficient. Of course these three criteria are contradictory and, as a compromise, we study several conditions to get a partial picture of the behavior of the algorithm. We could state a technical ad hoc condition containing exactly what is used in the proof, but then such a condition would have to be checked on particular cases. We will rather focus on two simple and general conditions, which provides a simple proof that there exist situations for which the bias of the chain is of second order in δ with possibly a nice dependence in the dimension.

The simplest condition one can think of, considered in other works like [18, 14, 33], is that the Hessian of U is Lipschitz:

$$\exists L_2 > 0 \ \text{ such that } \ \forall x, y \in \mathbb{R}^d, \qquad |\nabla^2 U(y) - \nabla^2 U(x)| \ \leqslant \ L_2 |x - y| \,.$$

In fact, we will work under a possibly weaker generalization of this condition:

Assumption $(\nabla^2\mathbf{pol}(\ell))$. *There exist $\ell \geqslant 2$, $L_\ell > 0$ and $x_\star \in \mathbb{R}^d$ such that, for all $x, y \in \mathbb{R}^d$,*

$$|\nabla^2 U(y) - \nabla^2 U(x)| \ \leqslant \ L_\ell |x - y| \left(2 + |x - x_\star|^{\ell-2} + |y - x_\star|^{\ell-2} \right) / 4 \,. \qquad (4)$$

Note that, up to a change of the constant L_ℓ, if Assumption $(\nabla^2\mathbf{pol}(\ell))$ holds for some $x_\star$ then it holds with any other value of $x_\star$. As a consequence, implicitly, whenever Assumptions (**Conv**) and $(\nabla^2\mathbf{pol}(\ell))$ are simultaneously satisfied, then we chose $x_\star$ in $(\nabla^2\mathbf{pol}(\ell))$ to be the global minimum of U given by (**Conv**).

As can be seen through a Taylor expansion, $(\nabla^2\mathbf{pol}(\ell))$ holds in particular if $\|\nabla^{(k)}U\|_\infty < +\infty$ for some $k \geqslant 3$.

As we will see, under Assumption $(\nabla^2\mathbf{pol}(\ell))$ for any $\ell \geqslant 2$, we still get a second-order error in δ. Nevertheless, for $\ell > 2$, due to the presence of additional moments, the dependence in the dimension worsen. However, as illustrated by the next condition, there are cases where, even though the local Lipschitz constant of $\nabla^2 U$ is polynomial, the OBABO sampler is super efficient.

Assumption $(\perp\!\!\!\perp)$. *There exist $\ell \in \mathbb{N}$, $L_\ell > 0$, $x_\star = (x_{\star,1}, \ldots, x_{\star,d}) \in \mathbb{R}^d$, an orthogonal matrix $\mathcal{Q} \in \mathcal{O}(d)$ and, for $i \in [\![1, d]\!]$, potentials $U_i \in \mathcal{C}^2(\mathbb{R})$ such that $U(\mathcal{Q}x) = \sum_{i=1}^d U_i(x_i)$ and for all $i \in [\![1, d]\!]$ and $x, y \in \mathbb{R}$,*

$$|U_i''(y) - U_i''(x)| \ \leqslant \ L_\ell |x - y| \left(2 + |x - x_{\star,i}|^{\ell-2} + |y - x_{\star,i}|^{\ell-2} \right) / 4 \,.$$

Similarly to Assumption $(\nabla^2\mathbf{pol}(\ell))$, when Assumptions $(\perp\!\!\!\perp)$ and (**Conv**) are both satisfied, $x_\star$ is implicitly the minimum of U. Assumption $(\perp\!\!\!\perp)$ means that the target distribution is separable, namely, in a suitable orthonormal basis, the coordinates of a random variable distributed accorded to π are independent. It is important to note that we do not need to know $\mathcal{Q}$. Indeed, contrary to e.g. the Zig-Zag sampler [3] or more generally Gibbs algorithms, in the OBABO sampler, the basis does not play a particular role, i.e. if $(x_n, v_n)_{n\in\mathbb{N}}$ is an OBABO chain for some potential U then $(\mathcal{Q}^{-1}x_n, \mathcal{Q}^{-1}v_n)_{n\in\mathbb{N}}$ is an OBABO chain with potential $x \mapsto U(\mathcal{Q}x)$. As a particular case, Assumption $(\perp\!\!\!\perp)$ is satisfied for all Gaussian distributions (with $\ell = 2$ and $L_\ell = 0$).

Although the independence assumption is very restrictive, it is the classical condition under which scaling limits are established for MCMC algorithms, see e.g. [15] and references within. Non-asymptotic bounds for the HMC algorithm for a separable target have been established in [36], and under a similar but weaker condition in [37].

Proposition 12. *Under the conditions of Theorem 1, and considering K_5, K_6, K_7, K_8 as given by Proposition 31, (in particular*

$$K_5 \to \frac{L}{4\sqrt{m}} + \frac{L}{6}, \quad K_6 \to 2^{\ell-2}\left(1 + \frac{1}{m^{\ell/2}}\right), \quad K_8 \to 2^{\ell+1}\left(1 + \frac{1}{m^{(\ell+1)/2}}\right)$$

and $K_7 \to L$ as $\delta \to 0$), the following holds:

1. *If Assumption $(\nabla^2\mathbf{pol}(\ell))$ holds then, for all $p \geqslant 1$,*

$$
\begin{aligned}
\mathcal{W}_p\left(\pi_\delta, \pi\right) &\leqslant \delta^2\left(d + \ell p - 2\right)^{\ell/2} \frac{2K_1\left(K_5 + L_\ell K_6\right)}{\kappa} \\
\|\pi_\delta - \pi\|_{TV} &\leqslant \delta^2\left(2 + |\ln\left(\delta^3(d + \ell - 1)\right)|\right)(d + \ell - 1)^{(\ell+1)/2} \\
&\quad \times \frac{2K_1^2 K_2\left(K_5 + L_\ell K_6\right) + K_7 + L_\ell K_8}{\kappa}.
\end{aligned}
$$

2. *If Assumption $(\perp\!\!\!\perp)$ holds then, for all $p \geqslant 2$,*

$$
\begin{aligned}
\mathcal{W}_p\left(\pi_\delta, \pi\right) &\leqslant \delta^2\sqrt{d}\frac{2K_1\left(K_5 + L_\ell K_6\right)\left(\ell p - 1\right)^{\ell/2}}{\kappa} \\
\|\pi_\delta - \pi\|_{TV} &\leqslant \delta^2 d\left(2 + |\ln\left(\delta^3\ell\right)|\right)\ell^{(\ell+1)/2} \\
&\quad \times\frac{2K_1^2 K_2\left(K_5 + L_\ell K_6\right) + K_7 + L_\ell K_8}{\kappa}.
\end{aligned}
$$

This is proven at the end of Section 5.

Remark 13. *In the Gaussian case, π_δ is an explicit Gaussian measure and thus $\mathcal{W}_2(\pi, \pi_\delta)$ is known. More precisely, in that case, we get $\mathcal{W}_2(\pi, \pi_\delta) \simeq \delta^2\sqrt{dL}/4$ as $\delta \to 0$ (we refer to the upcoming work [39] for the detailed analysis of the Gaussian case), which means the $\delta^2\sqrt{d}$ dependency under Assumption $(\perp\!\!\!\perp)$ is sharp.*

On the contrary, for the dependency in d of the total variation bound, the results of [33] give a bound of order $\sqrt{d}$ under the condition $(\nabla^2\mathbf{pol}(\ell))$ with $\ell = 2$, while we only get $d^{3/4}$. We discuss this in more details in Section 3.4.

Although one can design academic examples where $(\nabla^2\mathbf{pol}(\ell))$ holds for some $\ell > 2$ but not for $\ell = 2$, while $(\nabla\mathbf{Lip})$ holds, this does not correspond to practical cases of interest. The reason why we decided to work under a framework with possibly $\ell > 2$ is the following. For potentials that behave like $U(x) = |x|^\alpha$ with $\alpha > 2$ at infinity, $(\nabla^2\mathbf{pol}(\ell))$ holds with $\ell > 2$ but not $\ell = 2$. On the other hands $(\nabla\mathbf{Lip})$ doesn't hold, but it can be replaced by a condition similar to $(\nabla^2\mathbf{pol}(\ell))$ but on ∇U rather than $\nabla^2 U$. Moreover the OBABO integrator

becomes unstable and should be replaced by an adaptive time-step/tamed version. Adapting our analysis to this case, similar but more involved, is out of the scope of the present work, but we wanted to highlight that working with such local Lipschitz-continuity conditions is possible and possibly leads to a higher dependency in the dimension of the bias estimates due to additional moments, but not necessarily, as illustrated by the separable case.

3.3. The Metropolis-adjusted chain

Since the continuous-time Langevin diffusion has, in the smooth and convex case, a convergence rate that is independent from the dimension, and as seen in Theorem 1, the OBABO chain has a convergence rate that depends on the dimension only, possibly, through the time-step δ. Hence, the number of steps of the chain required in order to sample the target measure is of order $1/\delta$ (up to some logarithmic terms). On the other hand, the rejection probability in the OM(BAB)O scheme is of order δ^2. This means that, as δ vanishes, there is a high probability that no rejection occurs during the whole simulation. Consequently, results obtained for the unadjusted OBABO chain are straightforwardly transfered to the OM(BAB)O chain. For simplicity we focus on the total variation distance and confidence intervals, but similar adaptations could be conducted for Wasserstein distances.

In the following, under assumption (**Conv**), we denote $z_\star = (x_\star, 0)$ (which is the minimum of $H(x, v) = U(x) + |v|^2/2$) and, for $p \geqslant 1$ and $\nu \in \mathcal{P}_p(\mathbb{R}^{2d})$, consider the moments

$$M_{\nu, p} \;=\; \mathbb{E}_\nu \left(|Z - z_\star|^p \right) .$$

Proposition 14. *Under the conditions of Theorem 1, for all $n \in \mathbb{N}_*$ and $\nu, \mu \in \mathcal{P}_2(\mathbb{R}^{2d})$,*

$$\|\nu P_M^n - \mu P_M^n\|_{TV} \leqslant \|\nu P^n - \mu P^n\|_{TV} + \mathcal{E}(\nu, \delta, n, d) + \mathcal{E}(\mu, \delta, n, d) ,$$

where $\mathcal{E}$ is such that:

1. *For all $\nu \in \mathcal{P}_2(\mathbb{R}^{2d})$, considering K_9 as in Corollary 35,*

$$\mathcal{E}(\nu, \delta, n, d) \;\leqslant\; \delta K_9 \left(M_{\nu, 2} + \delta n d \right) .$$

2. *If condition $(\nabla^2 \mathbf{pol}(\ell))$ holds and $\nu \in \mathcal{P}_{\ell+1}(\mathbb{R}^{2d})$, considering K_{10} as in Corollary 35,*

$$\mathcal{E}(\nu, \delta, n, d) \;\leqslant\; \delta^2 K_{10} \left(M_{\nu, \ell+1} + \delta n \left(d + \ell - 1 \right)^{(\ell+1)/2} \right) .$$

3. *If condition $(\perp\!\!\!\perp)$ holds, and $\nu \in \mathcal{P}_{\ell+1}(\mathbb{R}^{2d})$, considering K_{10} as in Corollary 35 and, for $i \in [\![1, d]\!]$, denoting by ν_i the law of Y_i where $Y = \mathcal{Q}^{-1} X$ with $X \sim \nu$,*

$$\mathcal{E}(\nu, \delta, n, d) \;\leqslant\; \delta^2 K_{10} \left(\sum_{i=1}^{d} M_{\nu_i, \ell+1} + \delta n d \ell^{(\ell+1)/2} \right) .$$

This is proven at the end of Section 6. To fix ideas, taking first the limit as $\delta \to 0$ and then the leading terms as $L, L_\ell \to +\infty$ and $m \to 0$ (with $L_\ell \ll L^2$), we get

$$K_9 \simeq \frac{18L^2\gamma}{m^2}, \qquad K_{10} \simeq \frac{9 \times 12^{(\ell-1)/2}L^{(\ell+5)/2}\gamma}{m^{(\ell+3)/2}}.$$

Applied with $\mu = \pi$ (so that $\mu P_M^n = \pi$) and in combination with Corollary 5, this gives a non-asymptotic bound on $\|\nu P_M^n - \pi\|_{TV}$.

Remark 15. *Contrary to Corollary 5 for the OBABO chain, Proposition 14 does not provide for the OM(BAB)O chain a bound on the total variation convergence rate. Nevertheless, combining Proposition 14 with Corollary 5 yields local coupling estimates, i.e. for all $R > 0$, a suitable choice of δ and n ensures that*

$$\forall z, z' \in \{y \in \mathbb{R}^{2d}, \ |y - z_\star| \leqslant R\}, \qquad \|P_M^n(z, \cdot) - P_M^n(z', \cdot)\|_{TV} \leqslant \frac{1}{2}.$$

Contrary to the one-step Doeblin condition considered in [48], these n-steps coupling bounds have a reasonably nice scaling with δ and d.

Similarly, the non-asymptotic confidence intervals of Theorem 6 can be transfered to the Metropolis-adjusted chain:

Theorem 16. *Under the conditions of Theorem 1, if $\mu_0 \in \mathcal{P}(\mathbb{R}^{2d})$ satisfies a log-Sobolev inequality with constant $C' > 0$ and if $(Z_k)_{k \in \mathbb{N}}$ is an OM(BAB)O chain with $Z_0 \sim \mu_0$, then for all 1-Lipschitz functions φ on $\mathbb{R}^{2d}$, $n \geqslant 1$ and $u \geqslant 0$,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n}\varphi(Z_k) - \pi(\varphi)\right| \geqslant u + \frac{2K_1}{n\kappa\delta}\mathcal{W}_1(\mu_0, \pi_\delta) + \mathcal{W}_1(\pi, \pi_\delta)\right)$$

$$\leqslant 2\exp\left(-\frac{nu^2\kappa\delta}{4C_{lS} + 4K_1^2C'/(n\kappa\delta)}\right) + \mathcal{E}(\mu_0, \delta, n, d).$$

where C_{lS} is given in Theorem 6 and $\mathcal{E}(\nu, \delta, n, d)$ is as in Proposition 14.

This is proven at then end of Section 6.

3.4. Conclusion and related works

Let us summarize the previous results, focusing on the behavior as $\delta \to 0$ and $d \to +\infty$ for fixed values of p, m, L, L_ℓ, ℓ and $\gamma = 2\sqrt{L}$. We suppose that the initial condition μ_0 is such that, for $p \geqslant 1$, $\mathcal{W}_p(\mu_0, \pi) + M_{\mu_0, p}^{1/p} \leqslant \tilde{K}_p\sqrt{d}$ for some $\tilde{K}_p$ that does not depend on d, which is feasible in practice (see Remark 7 and Lemma 30). For the dependency on L, m, L_ℓ, we only write the leading terms when $L, L_\ell \to +\infty$ and $m \to 0$.

For a desired accuracy $\varepsilon > 0$, we denote

$$n_{\varepsilon, p} = \inf\{n \in \mathbb{N}, \ \mathcal{W}_p(\mu_0 P^n, \pi) \leqslant \varepsilon\}$$

$$
\begin{aligned}
n_{\varepsilon,TV} &= \inf\{n \in \mathbb{N}, \ \|\mu_0 P^n - \pi\|_{TV} \leqslant \varepsilon\} \\
n_{\varepsilon,M} &= \inf\{n \in \mathbb{N}, \ \|\mu_0 P_M^n - \pi\|_{TV} \leqslant \varepsilon\}.
\end{aligned}
$$

Choosing those as non-asymptotic criteria for efficiency is debatable, see the discussion in [13] in particular for a different scaling for the Wasserstein distances, nevertheless we stick to these definitions for comparison with previous works. Table 1 gathers the results on the discretization bias and Table 2 the efficiency bounds that are obtained from our various results. For simplicity, in Table 2, concerning the dependency on L, m and L_ℓ, we only consider the case where the term with L_ℓ is negligible with respect to the other one (as in the Gaussian case where $L_\ell = 0$) and we don't write the logarithmic terms in these variables.

TABLE 1
Leading terms of the bounds on the equilibrium bias respectively obtained in Propositions 11 and 12.

	$(\nabla\mathbf{Lip})$, $(\mathbf{Conv})$	$(\nabla\mathbf{Lip})$, $(\mathbf{Conv})$, $(\nabla^2\mathbf{pol}(\ell))$	$(\nabla\mathbf{Lip})$, $(\mathbf{Conv})$, $(\perp\!\!\!\perp)$
$\mathcal{W}_p(\pi, \pi_\delta)$	$\delta\sqrt{d}\frac{L^2}{m}$	$\delta^2 d^{\ell/2}\left(\frac{L^2}{m^{3/2}} + \frac{\sqrt{L}L_\ell}{m^{1+\ell/2}}\right)$	$\delta^2\sqrt{d}\left(\frac{L^2}{m^{3/2}} + \frac{\sqrt{L}L_\ell}{m^{1+\ell/2}}\right)$
$\|\pi - \pi_\delta\|_{TV}$	$\delta d\|\ln\left(\delta^3 d\right)\| \times \frac{L^{9/4}}{m}$	$\delta^2 d^{(\ell+1)/2}\|\ln\left(\delta^3 d\right)\| \times \left(\frac{L^{9/4}}{m^{3/2}} + \frac{L^{5/4}L_\ell}{m^{(\ell+3)/2}}\right)$	$\delta^2 d\|\ln\delta\| \times \left(\frac{L^{9/4}}{m^{3/2}} + \frac{L^{5/4}L_\ell}{m^{(\ell+3)/2}}\right)$

Let us now compare our results to previous works.

First, we mention that, starting back at least to the work of Talay [54] (who was also already concerned with the discretized chain), there has been a substantial amount of works in the last two decades on obtaining quantitative convergence rates for the *continuous-time* (underdamped) Langevin diffusion (1), leading to the theory of hypocoercivity [57, 17, 24]. In the convex and smooth case, the dimension-free convergence is established by the author in [40] for mean-field particles using entropic hypocoercivity techniques (implying the Wasserstein convergence, see e.g. [22] for details), and later by [12, 14, 59] by direct coupling arguments in the context of MCMC. The Wasserstein contraction induced by the parallel coupling in the smooth and convex case was first established in [4] (but without the explicit dependency on the dimension). In the non-convex case, quantitative results are obtained through an explicit combination of reflection and parallel coupling in [21, 13, 11]. Using hypocoercive techniques, one of the first result on the underdamped Langevin algorithms motivated by stochastic algorithms (MCMC and simulated annealing) is established in [42], focusing on low-temperature rather than high-dimensional estimates. In the recent [33], a similar result is established, focusing on high-dimensional MCMC (in both [33] and [42], the convergence is obtained in term of the log-Sobolev constant of the target distribution; the main difference is thus how the dependency on the parameters of interest is presented).

TABLE 2
Leading terms of the efficiency bounds obtained with a suitable choice of δ, based on Propositions 11 and 12 (for the equilibrium bias in the unadjusted case), Theorem 1 and Corollary 5 (for the convergence of the OBABO chain to equilibrium) and Proposition 14 (for the OM(BAB)O chain).

	$(\nabla\mathbf{Lip})$, $(\mathbf{Conv})$	$(\nabla\mathbf{Lip})$, $(\mathbf{Conv})$, $(\nabla^2\mathbf{pol}(\ell))$	$(\nabla\mathbf{Lip})$, $(\mathbf{Conv})$, $(\perp\!\!\!\perp)$
$n_{\varepsilon,p}$	$\dfrac{\sqrt{d}}{\varepsilon}\ln\left(\dfrac{d}{\varepsilon}\right)\dfrac{L^{5/2}}{m^2}$	$\dfrac{d^{\ell/4}}{\sqrt{\varepsilon}}\ln\left(\dfrac{d}{\varepsilon}\right)\dfrac{L^{3/2}}{m^{7/4}}$	$\dfrac{d^{1/4}}{\sqrt{\varepsilon}}\ln\left(\dfrac{d}{\varepsilon}\right)\dfrac{L^{3/2}}{m^{7/4}}$
$n_{\varepsilon,TV}$	$\dfrac{d}{\varepsilon}\ln^2\left(\dfrac{d}{\varepsilon}\right)\dfrac{L^{11/4}}{m^2}$	$\dfrac{d^{(\ell+1)/4}}{\sqrt{\varepsilon}}\ln^2\left(\dfrac{d}{\varepsilon}\right)\dfrac{L^{13/8}}{m^{7/4}}$	$\dfrac{\sqrt{d}}{\sqrt{\varepsilon}}\ln^2\left(\dfrac{d}{\varepsilon}\right)\dfrac{L^{13/8}}{m^{7/4}}$
$n_{\varepsilon,M}$	$\dfrac{d}{\varepsilon}\ln^4\left(\dfrac{d}{\varepsilon}\right)\dfrac{L^{7/2}}{m^4}$	$\dfrac{d^{(\ell+1)/4}}{\sqrt{\varepsilon}}\ln^3\left(\dfrac{d}{\varepsilon}\right)\dfrac{L^{(\ell+9)/4}}{m^{(\ell+9)/4}}$	$\dfrac{\sqrt{d}}{\sqrt{\varepsilon}}\ln^3\left(\dfrac{d}{\varepsilon}\right)\dfrac{L^{(\ell+9)/4}}{m^{(\ell+9)/4}}$

By contrast, Theorem 1 and Corollary 5 give explicit quantitative convergence rates for a *discrete-time* Markov chain obtained as a discretization of (1). As mentioned in the introduction, this leads to the concentration inequalities of Theorem 6 and it was motivated by the work [48]. That being said, in [12, 14, 59, 13, 11, 33, 59], a discretization error analysis is conducted which, together with the convergence rate of the continuous-time process, provides *in fine* non-asymptotic bounds for the discrete-time algorithm. In the smooth and convex case, up to some logarithmic terms, $\mathcal{W}_2$ efficiency bounds of order $\sqrt{d}/\varepsilon$ are obtained in [12, 59] with a first-order approximation of (1) and of order $\sqrt{d/\varepsilon}$ in [14] with a second-order scheme (that requires the computation of $\nabla^2 U$ at each iteration) under the Lipschitz Hessian condition (that corresponds to $(\nabla^2\mathbf{pol}(\ell))$ with $\ell = 2$). These scalings are thus similar to our results on the OBABO chain under similar conditions. In [33], using a second order scheme with a similar complexity of the OBABO chain, efficiency bounds of order $\sqrt{d/\varepsilon}$ are obtained for the relative entropy (Kullback-Leibler divergence) under the Lipschitz Hessian condition. Through Pinsker's and Talagrand's inequalities, this yields $\mathcal{W}_2$ and total variation efficiency bounds of order $\sqrt{d}/\varepsilon$ (since the results have to be applied with ε replaced by ε^2), to compare to our results respectively of $\sqrt{d/\varepsilon}$ and $d^{3/4}/\sqrt{\varepsilon}$ for these distances under $(\nabla^2\mathbf{pol}(\ell))$ with $\ell = 2$. Thus, for the unadjusted process, under the same conditions as [33], we see that we have a better dependency in term of ε for both Wasserstein and total variation distances, the same dependency in the dimension for the Wasserstein distance but a worse dependency in the dimension ($d^{3/4}$ versus $\sqrt{d}$) for the total variation distance. In the separable case we recover the $\sqrt{d}$ scaling for the total variation distance, and improve the scaling to $d^{1/4}$ for the Wasserstein one. Our method is very different from the one of [33]. Comparing the discrete scheme with the continuous process, they get a bound on the relative entropy

that scales as d, which gives $\sqrt{d}$ for the total variation. We get a bound on the total variation by comparing the unadjusted process with the Metropolis-adjusted one. We can see that the bad dependency in d for the total variation in our case stems from the fact we bound the expected Metropolis rejection probability by a cubic term (Lemma 28 with $\ell = 2$) whose expectation is of order $d^{3/2}$. The bounds of Lemma 28 seems to be sharp, so it is the method of comparing the unadjusted and adjusted processes which leads to a non-optimal scaling. In some sense, by requiring the two processes to stay equal, we do not use the regularization properties of the process. By working with the relative entropy, the law of the unadjusted process can be compared with the law of the continuous time process: the processes (driven by the same Brownian motion) do not stay equal but remain close which, with the regularization properties of the process, is in fact sufficient to control the relative entropy of the laws (roughly speaking), hence the total variation.

Similarly to the results on the convergence rate of the Langevin diffusion, several works are concerned with the convergence rates of *idealized* HMC algorithms, i.e. algorithms that rely on the exact simulation of the continuous-time Hamiltonian dynamics, [52, 29, 15, 36, 5]. Non-asymptotic bounds on the unadjusted HMC are obtained in [36], of order $\sqrt{d}/\varepsilon$ for the $\mathcal{W}_1$ distance in the smooth and convex case with a first-order scheme, of order $d^{1/4}/\sqrt{\varepsilon}$ in the separable case (and later in [37] under a weaker condition), and similarly for the Metropolis-adjusted algorithm. Again, this is similar to our rates for the OBABO chain.

The works [20, 10] are concerned with Metropolis-adjusted algorithms (HMC and Metropolis-adjusted overdamped Langevin algorithm – MALA), and establish total variation efficiency bounds that are logarithmic in the accuracy ε. The method is quite different to ours, based on conductance bounds. In term of number of computations of gradients, the scalings are d for the MALA algorithm and $d^{11/12}$ for the HMC one in the smooth, convex, Lipschitz Hessian case. In term of dependency in d, for the Metropolis-adjusted case, we only get d in this case, and $\sqrt{d}$ in the separable case. It would be interesting to use the methods of [20, 10] for the OM(BAB)O chain (see also Remark 15).

For more considerations on the family of sampler based, like the HMC process and Langevin diffusion, on the Hamiltonian dynamics, we also refer to the recent preprint [53] and references therein.

4. Study of the OBABO chain

4.1. Wasserstein contraction in the convex case

For $z = (x, v) \in \mathbb{R}^{2d}$, and $a, b > 0$ with $b^2 < a$, we consider the Euclidean norm

$$\|z\|_{a,b}^2 = |x|^2 + 2bx \cdot v + a|v|^2 .$$

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

$$\Theta_1(x, v, g, g') = x + \delta \left(\eta v + \sqrt{1 - \eta^2} g \right) - \frac{\delta^2}{2} \nabla U(x)$$

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

and $\Theta = (\Theta_1, \Theta_2)$. The transition (2) of the OBABO chain is thus given by $(x_1, v_1) = \Theta(x_0, v_0, G, G')$ with independent $G, G' \sim \mathcal{N}(0, I_d)$. We consider two initial conditions $(x_0, v_0), (y_0, w_0) \in \mathbb{R}^d \times \mathbb{R}^d$, two independent sequences of independent standard Gaussian variables $(G_n)_{n \in \mathbb{N}}$ and $(G'_n)_{n \in \mathbb{N}}$, and the parallel coupling of two OBABO chains given by, for all $n \in \mathbb{N}$,

$$(x_{n+1}, v_{n+1}) = \Theta(x_n, v_n, G_n, G'_n), \qquad (y_{n+1}, w_{n+1}) = \Theta(y_n, w_n, G_n, G'_n). \tag{5}$$

Proposition 17. *Under Assumption* (∇**Lip**)*and* (**Conv**)*, suppose moreover that $\gamma \geqslant 2\sqrt{L}$ and that $\delta \leqslant m/(33\gamma^3)$. Set*

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

Then, $b^2 \leqslant a/4$ and for all $(x_0, v_0), (y_0, w_0) \in \mathbb{R}^{2d}$, the parallel coupling given by (5) *is such that almost surely, for all $n \in \mathbb{N}$,*

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

Remark 18. *It is possible to improve the condition on γ and the other estimates by assuming that $U(x) = x \cdot S^{-1}x + \tilde{U}(x)$ for some symmetric positive matrix S and then taking S into account in the definition of the modified norm with more care. In particular in the Gaussian case ($\tilde{U} = 0$) there is no condition on γ, it is always possible to design a Euclidean norm that is contracted by the coupling for sufficiently small δ, see e.g. [43]. Nevertheless, in this work we are mainly concerned with the fact that κ is independent from the dimension.*

Remark 19. *The contraction is almost sure, and not only in expectation. In fact, as will be clear in the proof, the particular law of (G, G') does not intervene, it is sufficient that it does not depend on the position (x, v).*

Proof. Let us show that, for this choice of parameters, for all $x, v, y, w, g, g' \in \mathbb{R}^d$, denoting $(x', v') = \Theta(x, v, g, g')$ and $(y', w') = \Theta(y, w, g, g')$,

$$\|(x', v') - (y', w')\|_{a,b}^2 \leqslant (1 - \delta\kappa) \|(x, v) - (y, w)\|_{a,b}^2,$$

which will immediately yield the result.

As a first step, we identify the leading terms (in term of δ) of the evolution of the norm. The contraction will be proven only thanks to these terms, the higher order ones being treated as a small perturbation. As a consequence, our arguments works for other Markov chains that are an approximation of order at least one (in δ) of the continuous-time Langevin diffusion. 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 $\nabla U(x') - \nabla U(y') = \nabla^2 U(\xi')\bar{x}'$ and $\nabla U(x) - \nabla U(y) = \nabla^2 U(\xi)\bar{x}$, and let $Q = \nabla^2 U(\xi)$, $Q' = \nabla^2 U(\xi')$. Under Assumptions (∇**Lip**) and (**Conv**), $mI_d \leqslant Q \leqslant LI_d$ in the sense of symmetric matrices, and similarly for Q'. With these notations,

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

$$= (I_{2d} + \delta A)\begin{pmatrix} \bar{x} \\ \bar{v} \end{pmatrix} + h$$

with

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

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

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

$$\|\bar{z}'\|_{a,b}^2 = \|\bar{z}\|_{a,b}^2 + \delta\bar{z} \cdot (MA + A^T M)\bar{z} + 2\bar{z} \cdot Mh + \|\delta A\bar{z} + h\|_{a,b}^2$$

$$\leqslant \|\bar{z}\|_{a,b}^2 + \delta\bar{z} \cdot (MA + A^T M)\bar{z} + 2\|\bar{z}\|_{a,b}\|h\|_{a,b} + 2\delta^2\|A\bar{z}\|_{a,b}^2 + 2\|h\|_{a,b}^2.$$

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

$$\frac{1}{2}\|z\|_{a,0}^2 \leqslant \|z\|_{a,b}^2 \leqslant \frac{3}{2}\|z\|_{a,0}^2. \tag{6}$$

Using (6) and the bounds $|Q|, |Q'| \leqslant L$, $|1 - \eta| \leqslant \delta\gamma/2$, $|1 - \eta^2 - \delta\gamma| \leqslant \delta^2\gamma^2/2$ and

$$|\Delta\bar{x}| = |\delta\eta\bar{v} - \delta^2/2 Q\bar{x}| \leqslant \delta|\bar{v}| + \delta^2 L/2|\bar{x}|$$

we roughly bound

$$\|h\|_{a,b}^2 \leqslant \frac{3}{2}\left| \delta(\eta - 1)\bar{v} - \frac{\delta^2}{2}Q\bar{x} \right|^2$$

$$+ \frac{3}{2}a\left| -\frac{\delta(\eta - 1)}{2}(Q + Q')\bar{x} + (\eta^2 - 1 + \delta\gamma)\bar{v} - \frac{\delta\eta}{2}Q'\Delta\bar{x} \right|^2$$

$$\leqslant 3\delta^4\gamma^2/4|\bar{v}|^2 + 3\delta^4 L^2/4|\bar{x}|^2 + 3a\left(\delta^2\gamma L/2 + \delta^3 L^2/4\right)^2 |\bar{x}|^2$$

$$+ 3a\left(\delta^2\gamma^2/2 + \delta^2 L/2\right)^2 |\bar{v}|^2$$

$$\leqslant 6\delta^4 \max\left(\frac{(L\gamma/2 + \delta L^2/4)^2}{L} + \frac{3L^2}{4}, \frac{L\gamma^2}{4} + \left(\frac{\gamma^2}{2} + \frac{L}{2}\right)^2 \right) \|\bar{z}\|_{a,b}^2$$

$$\leqslant 3\delta^4\gamma^4 \|\bar{z}\|_{a,b}^2$$

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

$$\|A\bar{z}\|_{a,b}^2 \leqslant \frac{3}{2}|\bar{v}|^2 + \frac{3}{2}a\left|-\frac{1}{2}(Q+Q')\bar{x}-\gamma\bar{v}\right|^2 \leqslant 3L|\bar{x}|^2 + 3a\left(\frac{L}{2}+\gamma^2\right)|\bar{v}|^2$$

$$\leqslant \frac{27\gamma^2}{8}\|\bar{z}\|_{a,b}^2,$$

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

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

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 $\theta = \gamma/L$ we obtain

$$\bar{z} \cdot (MA + A^T M)\bar{z} \leqslant \bar{z}\left(\begin{pmatrix} \left(-\frac{1}{\gamma}+\frac{1}{2L\theta}\right)(Q+Q') & 0 \\ 0 & 2(b-a\gamma)I_d + \frac{a\theta}{2}(Q+Q') \end{pmatrix}\right)\bar{z}$$

$$\leqslant -\frac{m}{\gamma}|\bar{x}|^2 - a\gamma\left(1-\frac{2L}{\gamma^2}\right)|\bar{v}|^2$$

$$\leqslant -\frac{2m}{3\gamma}\|\bar{z}\|_{a,b}^2,$$

where we used that $\gamma/2 \geqslant L/\gamma \geqslant m/\gamma$ and (6). Finally, we have obtained that

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

For $p \geqslant 1$ and $a,b > 0$ with $b^2 \leqslant a/4$, denote by $\mathcal{W}_{p,a,b}$ the $\mathcal{W}_p$ Wasserstein distance on $\mathcal{P}_p\left(\mathbb{R}^{2d}\right)$ associated to the distance $\|\cdot\|_{a,b}$, i.e.

$$\mathcal{W}_{p,a,b}(\nu_1,\nu_2) = \inf_{\pi \in \Gamma(\nu_1,\nu_2)}\left(\int_{\mathbb{R}^{2d}\times\mathbb{R}^{2d}}\|z-z'\|_{a,b}^p\pi(\mathrm{d}z\mathrm{d}z')\right)^{1/p}.$$

The equivalence of $\|\cdot\|_{a,b}$ with the Euclidean distance implies that

$$\frac{1}{2}\min(1,a)\mathcal{W}_p^2 \leqslant \frac{1}{2}\mathcal{W}_{p,a,0}^2 \leqslant \mathcal{W}_{p,a,b}^2 \leqslant \frac{3}{2}\mathcal{W}_{p,a,0}^2 \leqslant \frac{3}{2}\max(1,a)\mathcal{W}_p^2. \quad (7)$$

From this equivalence and Proposition 17, it is then straightforward to obtain the following, which proves Theorem 1.

Corollary 20. *Under the assumptions and with the notations of Proposition 17, for all $p \geqslant 1$ and all $\nu, \mu \in \mathcal{P}_p(\mathbb{R}^{2d})$,*

$$\mathcal{W}_{p,a,b}^2(\nu P, \mu P) \leqslant (1-\delta\kappa)\mathcal{W}_{p,a,b}^2(\nu,\mu), \quad (8)$$

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

$$\mathcal{W}_p^2 \left(\nu P^n, \mu P^n \right) \ \leqslant \ 3 \max \left(L, \frac{1}{L} \right) (1 - \delta \kappa)^n \mathcal{W}_p^2(\nu, \mu) \,.$$

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

Proof. Consider the parallel coupling (5) with $((x_0, v_0), (y_0, w_0))$ a $\mathcal{W}_p$-optimal coupling of ν and μ (which exists from [58, Corollary 5.22]). Then $z_n = (x_n, v_n) \sim \nu P^n$ and $z'_n = (y_n, w_n) \sim \mu P^n$, so that

$$
\begin{aligned}
\mathcal{W}_{p,a,b}^p \left(\nu P^n, \mu P^n \right) \ \leqslant \ \mathbb{E} \left(\| z_n - z'_n \|_{a,b}^p \right) \ &\leqslant \ (1 - \delta \kappa)^{np/2} \mathbb{E} \left(\| z_0 - z'_0 \|_{a,b}^p \right) \\
&= \ (1 - \delta \kappa)^{np/2} \mathcal{W}_{p,a,b}^p (\nu, \mu) \,.
\end{aligned}
$$

Conclusion follows from the distance equivalence (7) and the Banach fixed point theorem (recall $(\mathcal{P}_p(\mathbb{R}^{2d}), \mathcal{W}_p)$ is a Banach space, [58, Theorem 6.18]). $\qquad \square$

Remark 21. *The one-step contraction (8) implies that P has a positive Wasserstein curvature $\delta \kappa / 2$ in the sense of [46, 26] with respect to the metric $\| \cdot \|_{a,b}$.*

4.2. Wasserstein/total variance regularization

While the proof of [23, Corollary 4.7] for the continuous-time process relies on functional inequalities arguments, our proof will be a direct application of optimal coupling for Gaussian variables, in the spirit of [18, 34].

As can be seen by conditioning with respect to the initial condition (distributed according to an optimal $\mathcal{W}_1$ coupling), Proposition 3 is an immediate corollary of the following.

Proposition 22. *Under Assumption (∇**Lip**), for all $(x_1, v_1), (x_2, v_2) \in \mathbb{R}^{2d}$,*

$$\| P \left((x_1, v_1), \cdot \right) - P \left((x_2, v_2), \cdot \right) \|_{TV} \ \leqslant \ \frac{3 \delta |v_1 - v_2| + \left(2 + 3 \delta^2 L/2 \right) |x_1 - x_2|}{\delta \sqrt{2 \pi (1 - \eta^2)}} \,.$$

Proof. As already mentioned, for a fixed (x, v), $(x', v') \sim P((x, v), \cdot)$ is not a Gaussian variable. Nevertheless, we are going to use successively two Gaussian couplings, coupling first the positions thanks to the random variable G and then the velocities with the random variable G'. In particular, if the first coupling is a success, then the non-Gaussian part $\nabla U(x') - \nabla U(y')$ vanishes. This two steps coupling (position first, then velocity) is natural in view of the hypoelliptic structure of the Langevin diffusion. Note that the choice of the OBABO sampler is particularly convenient here, since the damping part is splited in two half-steps, contrary to e.g. the BAOAB scheme [30] (although one would just have to consider two transitions of the Markov chain).

Recall the following result for the optimal coupling of Gaussian distributions. For $x, y \in \mathbb{R}^d$ and $\sigma > 0$,

$$\|\mathcal{N}(x, \sigma^2 I_d) - \mathcal{N}(y, \sigma^2 I_d)\|_{TV} \ \leqslant \ \frac{|x - y|}{\sqrt{2\pi\sigma^2}} \, .$$

Indeed, orthogonal coordinates being independent, we just have to couple the projections on $x - y$, so that the d dimensional case boils down to the one dimensional case, which is [34, Lemma 15].

Let $(x_1, v_1), (x_2, v_2) \in \mathbb{R}^d \times \mathbb{R}^d$. Denote

$$b_i = x_i + \delta\eta v_i - \delta^2/2\nabla U(x_i)$$

for $i = 1, 2$, and let (x_1', x_2') be an optimal coupling of $\mathcal{N}(b_i, \delta^2(1 - \eta^2)I_d)$, $i = 1, 2$. Set $G_i = (x_i' - b_i)/(\delta\sqrt{1 - \eta^2})$ for $i = 1, 2$. In other words, G_1 and G_2 are both standard Gaussian random variables and such that

$$x_i' \ = \ b_i + \delta\sqrt{1 - \eta^2}G_i \, , \qquad i = 1, 2 \tag{9}$$

and

$$2\mathbb{P}\left(x_1' \neq x_2'\right) \ \leqslant \ \frac{|b_1 - b_2|}{\delta\sqrt{2\pi(1 - \eta^2)}} \, . \tag{10}$$

Conditionally to (x_1', x_2'), denote

$$c_i \ = \ \eta^2 v_i - \frac{\delta\eta}{2}\left(\nabla U(x_i) + \nabla U(x_i')\right) + \sqrt{1 - \eta^2}\eta G_i$$

and let (v_1', v_2') be an optimal coupling of $\mathcal{N}(c_i, (1 - \eta^2)I_d)$, $i = 1, 2$, so that

$$2\mathbb{P}\left(v_1' \neq v_2' \mid (x_1', x_2')\right) \ \leqslant \ \frac{|c_1 - c_2|}{\sqrt{2\pi(1 - \eta^2)}} \, . \tag{11}$$

Set $G_i' = (v_i' - c_i)/\sqrt{1 - \eta^2}$ for $i = 1, 2$, so that G_1' and G_2' are both standard Gaussian random variables such that

$$v_i' \ = \ c_i + \sqrt{1 - \eta^2}G_i' \, , \qquad i = 1, 2. \tag{12}$$

The fact that, conditionally to (x_1', v_1'), G_1' is a standard Gaussian random variable, implies that G_1' is independent from (G_1, G_2) (and similarly for G_2'). As a consequence, (9) and (12) with G_1 and G_1' (resp. G_2 and G_2') that are two independent standard Gaussian variables proves that $(x_i', v_i') \sim P\left((x_i, v_i), \cdot\right)$ (recall the transition (2) corresponding to P). In particular,

$$\|\delta_{(x_1, v_1)}P - \delta_{(x_2, v_2)}P\|_{TV} \ \leqslant \ 2\mathbb{P}\left((x_1', v_1') \neq (x_2', v_2')\right) \, .$$

We now bound this probability. Remark that

$$x_1' = x_2' \qquad \Leftrightarrow \qquad \sqrt{1 - \eta^2}(G_1 - G_2) = \frac{b_2 - b_1}{\delta} \, .$$

Hence, under the event $\{x_1' = x_2'\}$,

$$c_1 - c_2 = \eta^2(v_1 - v_2) - \frac{\delta\eta}{2}\left(\nabla U(x_1) - \nabla U(x_2)\right) + \frac{\eta}{\delta}(b_2 - b_1).$$

Since

$$|b_1 - b_2| \leqslant \left(1 + \delta^2 L/2\right)|x_1 - x_2| + \delta|v_1 - v_2|,$$

then, under $\{x_1' = x_2'\}$,

$$|c_1 - c_2| \leqslant 2|v_1 - v_2| + (1/\delta + \delta L)|x_1 - x_2|.$$

As a conclusion, using (10) and (11),

$$2\mathbb{P}\left((x_1', v_1') \neq (x_2', v_2')\right) \leqslant 2\mathbb{P}\left(x_1' \neq x_2'\right) + 2\mathbb{P}\left(v_1' \neq v_2' \mid x_1' = x_2'\right)$$

$$\leqslant \frac{3\delta|v_1 - v_2| + \left(2 + 3\delta^2 L/2\right)|x_1 - x_2|}{\delta\sqrt{2\pi(1 - \eta^2)}}. \qquad \square$$

4.3. Concentration inequality

We first state the following general result (proven in [35, Theorem 1] and [9, Corollary 3.5] in particular cases, see also [41, Lemma 3.2]).

Theorem 23. *Let R be a Markov operator on $\mathbb{R}^d$ with invariant measure ν. Suppose that there exists $r \geqslant 0$ such that, for all bounded Lipschitz function φ and all $z \in \mathbb{R}^d$,*

$$|\nabla R\varphi(z)| \leqslant rR|\nabla\varphi|(z).$$

Suppose moreover that there exists $C > 0$ such $R(z, \cdot)$ satisfies a log Sobolev inequality with constant C for all $z \in \mathbb{R}^d$. Then:

1. *If $\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfies a log Sobolev with constant C', then μR satisfies a log Sobolev inequality with constant $r^2C' + C$.*
2. *If $r < 1$, then for all $n \in \mathbb{N}$ and all $z \in \mathbb{R}^d$, $R^n(z, \cdot)$ satisfies a log Sobolev inequality with constant $C(1 - r^{2n})/(1 - r^2)$, and ν satisfies a log Sobolev inequality with constant $C/(1 - r^2)$.*
3. *If $r < 1$, $\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfies a log Sobolev with constant C' and $(Z_n)_{n\in\mathbb{N}}$ is a Markov chain associated to R with initial condition $Z_0 \sim \mu$, then for all $n \in \mathbb{N}_*$, all $u \geqslant 0$, and all 1-Lipschitz functions φ,*

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}\left(\varphi(Z_k) - \mathbb{E}\left(\varphi(Z_k)\right)\right) \geqslant u\right) \leqslant \exp\left(-\frac{nu^2(1 - r)^2}{C + rC'/n}\right)$$

and

$$\left|\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left(\varphi(Z_k)\right) - \nu(\varphi)\right| \leqslant \frac{r}{n(1 - r)}\mathcal{W}_1(\nu, \mu).$$

Remark 24. *The first point generalizes the fact that log-Sobolev inequalities are transported by Lipschitz transformations, which corresponds to the case $R\varphi = \varphi \circ \Psi$ with Ψ an r-Lipschitz function (in that case, $C = 0$).*

Proof. By density, we can consider a smooth positive Lipschitz f. Denoting $g = \sqrt{Rf^2}$, we decompose

$$
\begin{aligned}
\mu R \left(f^2 \ln f^2 \right) - \mu R(f^2) \ln \mu R(f^2) &= \mu \left(R(f^2 \ln f^2) - R(f^2) \ln R(f^2) \right) \\
&\quad + \mu(g^2 \ln g^2) - \mu(g^2) \ln \mu(g^2) \\
&\leqslant C\mu R \left(|\nabla f|^2 \right) + C'\mu \left(|\nabla g|^2 \right) .
\end{aligned}
$$

Now, by the Cauchy-Schwarz inequality,

$$
|\nabla g|^2 = \frac{|\nabla R(f^2)|^2}{4R(f^2)} \leqslant \frac{r^2 \left(R|\nabla(f^2)| \right)^2}{4R(f^2)} \leqslant \frac{r^2}{4} R \left(\frac{|\nabla(f^2)|^2}{f^2} \right) = r^2 R \left(|\nabla f|^2 \right) ,
$$

which concludes the proof of the first claim. The second claim immediately follows by induction and the weak convergence of R^n toward ν, which follows from the Wasserstein contraction implied by the gradient/operator subcommutation, see [27, Theorem 2.2].

Concerning the third claim, if $\mu \in \mathcal{P}(\mathbb{R}^d)$ satisfies a log-Sobolev inequality with constant $C' > 0$ then, for all β-Lipschitz φ and all λ,

$$
\mu \left(e^{\lambda\varphi} \right) \leqslant e^{C'\beta^2\lambda^2/4} e^{\mu(\varphi)} , \tag{13}
$$

see e.g. [28]. Using successively Chernov's inequality, the Markov property and (13) together with the local log-Sobolev inequality satisfied by R, given any 1-Lipschitz φ, $\lambda > 0$ and $t \in \mathbb{R}$,

$$
\begin{aligned}
\mathbb{P} \left(\frac{1}{n} \sum_{k=1}^{n} \varphi(Z_k) \geqslant t \right) &\leqslant e^{-nt\lambda} \mathbb{E} \left(e^{\lambda \sum_{k=1}^{n} \varphi(Z_k)} \right) \\
&= e^{-nt\lambda} \mathbb{E} \left(e^{\lambda \sum_{k=1}^{n-1} \varphi(Z_k)} R \left(e^{\lambda\varphi} \right) (Z_{n-1}) \right) \\
&\leqslant e^{-nt\lambda + C\lambda^2/4} \mathbb{E} \left(e^{\lambda \sum_{k=1}^{n-2} \varphi(Z_k)} e^{\lambda(\varphi + R\varphi)(Z_{n-1})} \right) \\
&= e^{-nt\lambda + C\lambda^2/4} \mathbb{E} \left(e^{\lambda \sum_{k=1}^{n-2} \varphi(Z_k)} R \left(e^{\lambda(\varphi + R\varphi)} \right) (Z_{n-2}) \right)
\end{aligned}
$$

Using that $\varphi + R\varphi$ is $(1+r)$-Lipschitz, (13) and then a straightforward induction, we get

$$
\begin{aligned}
\mathbb{P} &\left(\frac{1}{n} \sum_{k=1}^{n} \varphi(Z_k) \geqslant t \right) \\
&\leqslant e^{-nr\lambda + C\lambda^2 \left[1 + (1+r)^2 \right]/4} \mathbb{E} \left(e^{\lambda \sum_{k=1}^{n-2} \varphi(Z_k)} e^{\lambda(R\varphi + R^2\varphi)(Z_{n-2})} \right) \\
&\leqslant e^{-nt\lambda + nC\lambda^2/(4(1-r)^2)} \mathbb{E} \left(e^{\lambda \sum_{k=1}^{n} R^k \varphi(Z_0)} \right)
\end{aligned}
$$

$$\leqslant \quad e^{-nt\lambda + (nC + rC')\lambda^2/(4(1-r)^2)} e^{\lambda \sum_{k=1}^{n} \mu R^k \varphi}.$$

Applying this with $t = \sum_{k=1}^{n} \mu R^k \varphi + u$ with $u \geqslant 0$ reads

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}\varphi(Z_k) - \frac{1}{n}\sum_{k=1}^{n}\mu R^k\varphi \geqslant u\right) \quad \leqslant \quad \inf_{\lambda > 0} \exp\left(-nu\lambda + \frac{nC + rC'}{4(1-r)^2}\lambda^2\right)$$

$$= \quad \exp\left(-\frac{nu^2(1-r)^2}{C + rC'/n}\right).$$

Using that ν is invariant by R, that $\sum_{k=1}^{n} R^k\varphi$ is a $r/(1-r)$-Lipschitz function for all $n \geqslant 1$ and that the dual representation of $\mathcal{W}_1$ (see e.g. [58]),

$$\left|\frac{1}{n}\sum_{k=1}^{n}\mu R^k\varphi - \nu\varphi\right| \quad = \quad \frac{1}{n}\left|(\mu - \nu)\left(\sum_{k=1}^{n} R^k\varphi\right)\right| \quad \leqslant \quad \frac{r}{n(1-r)}\mathcal{W}_1(\nu, \mu). \quad \square$$

In the rest of this section, considering a, b given by Proposition 17, we consider the matrix S such that $S(x, v) = (x + bv, \sqrt{a - b^2}v)$. Hence, $\|(x, v)\|_{a,b} = |S(x, v)|$. We also denote $P_{a,b}$ the transition operator of $(S(x_n, v_n))_{n \in \mathbb{N}}$ where $(x_n, v_n)_{n \in \mathbb{N}}$ is an OBABO chain. This change of variable is meant for simplicity, in order to work with the classical Euclidean metric and gradient in the following.

Lemma 25. *Under the conditions of Proposition 17, for all Lipschitz functions φ on $\mathbb{R}^{2d}$ and all $z \in \mathbb{R}^{2d}$,*

$$|\nabla P_{a,b}\varphi(z)| \quad \leqslant \quad \sqrt{1 - \delta\kappa} P_{a,b}|\nabla\varphi|(z).$$

Moreover, for all $z \in \mathbb{R}^{2d}$, $P_{a,b}(z, \cdot)$ satisfies a log-Sobolev inequality with constant

$$C \quad = \quad 2|S|^2(1 - \eta^2)\left(1 + \delta + \delta^2 L/2\right)^2.$$

Proof. Using the notations of Proposition 17, the latter means that, setting $z_n = (x_n, v_n)$ and $z'_n = (y_n, w_n)$,

$$|Sz_1 - Sz'_1| \quad \leqslant \quad \sqrt{1 - \delta\kappa}|Sz_0 - Sz'_0|$$

almost surely for all z_0, z'_0. This implies that $P_{a,b}$ contracts the $\mathcal{W}_\infty$ distances (for the classical Euclidean metric) by a factor $\sqrt{1 - \delta\kappa}$, which implies the claimed gradient/Markov operator subcommuation as proven in [27, Theorem 2.2].

For $z = (x, v) \in \mathbb{R}^{2d}$, $P_{a,b}(z, \cdot)$ is the image of $\mathcal{N}(0, I_{2d})$ by the transformation $\Psi : (G, G') \mapsto S\Theta(S^{-1}z, G, G')$. For $(g, g'), (h, h') \in \mathbb{R}^{2d}$,

$$|\Theta_1(z, g, g') - \Theta_1(z, h, h')| \quad = \quad \delta\sqrt{1 - \eta^2}|g - h|$$

and thus

$$|\Theta_2(z, g, g') - \Theta_2(z, h, h')|$$

$$
\leqslant \ \frac{\delta\eta}{2}L|\Theta_1(z,g,g') - \Theta_1(z,h,h')| + \sqrt{2(1-\eta^2)}\,(\eta|g-h| + |g'-h'|)
$$

$$
\leqslant \ \frac{\delta\eta}{2}L\delta\sqrt{1-\eta^2}|g-h| + \sqrt{1-\eta^2}\,(\eta|g-h| + |g'-h'|)\,.
$$

As a consequence, Ψ is Lipschitz with constant $|S|\left(1 + \delta + \delta^2 L/2\right)\sqrt{1-\eta^2}$. Since the standard Gaussian distribution satisfies a log-Sobolev inequality with constant 2, conclusion follows from Remark 24. $\square$

Proof of Theorem 6. From Theorem 1, $\tilde{\pi} = S\sharp\pi_\delta$ is the unique invariant measure of $P_{a,b}$. From Lemma 25 and Theorem 23, $\tilde{\pi}$ satisfies a log-Sobolev inequality with constant $C/(\delta\kappa)$. From Remark 24, $\pi_\delta = S^{-1}\sharp\tilde{\pi}$ then satisfies a log-Sobolev inequality with constant $|S^{-1}|^2 C/(\delta\kappa)$. Besides, $|S||S^{-1}| = K_1$ and $(1-\eta^2)/\delta \leqslant \gamma$, which gives the claimed expression for the log-Sobolev constant of π_δ.

For φ a 1-Lipschitz function, $\tilde{\varphi} := (\varphi \circ S^{-1})/|S^{-1}|$ is 1-Lipschitz. If $\mu \in \mathcal{P}(\mathbb{R}^{2d})$ satisfies a log-Sobolev inequality with constant C' then $\mu \circ S$ satisfies an inequality with constant $|S|^2 C'$. Hence, with $Z_0 \sim \mu$, Theorem 23 applied to $P_{a,b}$ reads

$$
\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}\left(\varphi(Z_k) - \mathbb{E}\left(\varphi(Z_k)\right)\right) \geqslant u\right)
$$

$$
= \ \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}\left(\tilde{\varphi}(SZ_k) - \mathbb{E}\left(\tilde{\varphi}(SZ_k)\right)\right) \geqslant \frac{u}{|S^{-1}|}\right)
$$

$$
\leqslant \ \exp\left(-\frac{nu^2(1-r)^2}{|S^{-1}|^2\left(C + r|S|^2 C'/n\right)}\right),
$$

with $r = \sqrt{1 - \delta\kappa} \leqslant 1 - \delta\kappa/2$. Applying this to φ and $-\varphi$ gives

$$
\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n}\left(\varphi(Z_k) - \mathbb{E}\left(\varphi(Z_k)\right)\right)\right| \geqslant u\right) \ \leqslant \ 2\exp\left(-\frac{nu^2\delta^2\kappa^2}{4|S^{-1}|^2\left(C + |S|^2 C'/n\right)}\right).
$$

Applying the last proposition of Theorem 23 to $P_{a,b}$, conclusion follows from

$$
\left|\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left(\varphi(Z_k)\right) - \pi_\delta(\varphi)\right| = |S^{-1}|\left|\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left(\tilde{\varphi}(SZ_k)\right) - \pi_\delta(\tilde{\varphi} \circ S)\right|
$$

$$
\leqslant \frac{2|S^{-1}|}{n\kappa\delta}\mathcal{W}_1(S\sharp\mu, S\sharp\pi_\delta) \ \leqslant \ \frac{2|S^{-1}||S|}{n\kappa\delta}\mathcal{W}_1(\mu, \pi_\delta)\,. \ \square
$$

5. Bias analysis

First, we prove Proposition 9, which relates the equilibrium error between π and π_δ to the one-step error between π and πP_V.

Proof of Proposition 9. We use the notations of Proposition 17. Using the invariance of π_δ by P and Corollary 20,

$$
\begin{aligned}
\mathcal{W}_{p,a,b}\left(\pi_\delta,\pi\right) &\leqslant \mathcal{W}_{p,a,b}\left(\pi_\delta P,\pi P\right)+\mathcal{W}_{p,a,b}\left(\pi P,\pi\right) \\
&\leqslant (1-\delta\kappa/2)\mathcal{W}_{p,a,b}\left(\pi_\delta,\pi\right)+\mathcal{W}_{p,a,b}\left(\pi P,\pi\right)
\end{aligned}
$$

where we used that $\sqrt{1-x}\leqslant 1-x/2$ for $x\in[0,1]$. As a consequence,

$$
\mathcal{W}_{p,a,b}\left(\pi_\delta,\pi\right) \leqslant \frac{2}{\delta\kappa}\mathcal{W}_{p,a,b}\left(\pi P,\pi\right)
$$

which together with the equivalence of $|\cdot|$ and $\|\cdot\|_{a,b}$ (see (7)) gives

$$
\mathcal{W}_p\left(\pi_\delta,\pi\right) \leqslant \frac{2K_1}{\delta\kappa}\mathcal{W}_p\left(\pi P,\pi\right)
$$

where K_1 is given in Theorem 1. Now, decomposing $P=P_O P_V P_O$, where P_O has been introduced in Section 2.2, we remark that P_O does not increase $\mathcal{W}_p$, as can be seen with a parallel coupling. Indeed, 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 $(x_i',v_i')\sim\delta_{(x_i,v_i)}P_O$ 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 $\nu_i\in\mathcal{P}_p(\mathbb{R}^d)$ 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

$$
\mathcal{W}_p(\nu_1 P_O,\nu_2 P_O) \leqslant \mathcal{W}_p(\nu_1,\nu_2)\,.
$$

Moreover, π is invariant by P_O, so that

$$
\mathcal{W}_p\left(\pi P,\pi\right) = \mathcal{W}_p\left(\pi P_O P_V P_O,\pi P_O\right) \leqslant \mathcal{W}_p\left(\pi P_V,\pi\right)\,.
$$

For the total variation distance, similarly, for $n\geqslant 1$,

$$
\begin{aligned}
\|\pi_\delta-\pi\|_{TV} &\leqslant \|\pi_\delta P^n-\pi P^n\|_{TV}+\sum_{k=1}^{n}\|(\pi P-\pi)P^{k-1}\|_{TV} \\
&\leqslant \frac{K_1 K_2}{\delta^{3/2}}(1-\delta\kappa)^{(n-1)/2}\mathcal{W}_1(\pi_\delta,\pi)+n\|\pi P-\pi\|_{TV}\,,
\end{aligned}
$$

where we used Corollary 5 and the fact the total variation is decreased by any Markov operator. Using again that π is invariant by P_O,

$$
\|\pi P-\pi\|_{TV} = \|\pi P_O P_V P_O-\pi P_O\|_{TV} \leqslant \|\pi P_V-\pi\|_{TV}\,,
$$

which concludes. $\qquad\square$

As mentioned in Section 3.2, in the following, we compare P_V to two reference transitions: the first one is P_{MV}, the Metropolis-adjusted Verlet transition introduced in Section 2.2, with acceptance probability α. The second is Q_δ where

$(Q_t)_{t \geqslant 0}$ is the deterministic transition operator 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

$$\hat{x}'_t = \hat{v}_t, \qquad \hat{v}'_t = -\nabla U(\hat{x}_t), \qquad \hat{x}_0 = x, \qquad \hat{v}_0 = v, \qquad (14)$$

denote $\Phi_{HD}(x, v) = (x_\delta, v_\delta)$ and $Q_\delta \varphi(x, v) = \varphi(x_\delta, v_\delta)$.

Lemma 26. *For $\nu \in \mathcal{P}(\mathbb{R}^{2d})$ and $(X, V) \sim \nu$,*

$$\|\nu P_V - \nu P_{MV}\|_{TV} \leqslant 2\mathbb{E}\left(1 - \alpha(X, V)\right).$$

If, moreover, $\nu \in \mathcal{P}_p(\mathbb{R}^{2d})$ for some $p \geqslant 1$, then

$$\mathcal{W}_p\left(\nu P_V, \nu P_{MV}\right) \leqslant \left(\mathbb{E}\left[|(X, -V) - \Phi_V(X, V)|^p \left(1 - \alpha(X, V)\right)\right]\right)^{1/p}$$

and

$$\mathcal{W}_p\left(\nu P_V, \nu Q_\delta\right) \leqslant \left(\mathbb{E}\left[|\Phi_V(X, V) - \Phi_{HD}(X, V)|^p\right]\right)^{1/p}$$

Proof. The only difference between P_V and P_{MV} is the accept/reject step. Considering $(X, V) \sim \nu$ and, independently, W a random variable uniformly distributed over $[0, 1]$, set

$$(X', V') = (X, -V)\mathbb{1}_{W > \alpha(X,V)} + \Phi_V(X, V)\mathbb{1}_{W \leqslant \alpha(X,V)}$$

Then, $(X', V') \sim \nu P_{MV}$, while $\Phi_V(X, V) \sim \nu P_V$, so that

$$\|\nu P_V - \nu P_{MV}\|_{TV} \leqslant 2\mathbb{P}\left((X', V') \neq \Phi_V(X, V)\right) = 2\mathbb{E}\left(1 - \alpha(X, V)\right).$$

Similarly,

$$\begin{aligned} \mathcal{W}_p^p\left(\nu P_V, \nu P_{MV}\right) &\leqslant \mathbb{E}\left(|(X', V') - \Phi_V(X, V)|^p\right) \\ &= 2^p \mathbb{E}\left(|(X, -V) - \Phi_V(X, V)|^p \left(1 - \alpha(X, V)\right)\right). \end{aligned}$$

Finally, the last claim follows from the fact $(\Phi_V(X, V), \Phi_{HD}(X, V))$ is a coupling of νP_V and νQ_δ. $\qquad \square$

Remark 27. *As $\delta \to 0$, $|(x, -v) - \Phi_V(x, v)| \to 2|v|$. When comparing a Metropolis adjusted Verlet step and an unadjusted one, in case of rejection, the distance is not small with δ.*

Lemma 28.

1. *Under Assumption $(\nabla \mathbf{Lip})$, for all $x, v \in \mathbb{R}^d$, writing $Z = |v| + \delta/2|\nabla U(x)|$,*

$$1 - \alpha(x, v) \leqslant \delta^3 \frac{1}{8}|\nabla U(x)|^2 + \delta^2 L \left(\frac{1}{2} + \delta\frac{1}{8} + \delta^2 \frac{L}{8}\right) Z^2.$$

2. *Under Assumptions $(\nabla \mathbf{Lip})$ and $(\nabla^2 \mathbf{pol}(\ell))$, for all $(x, v) \in \mathbb{R}^{2d}$, writing $Z = |v| + \delta/2|\nabla U(x)|$,*

$$\begin{aligned} 1 - \alpha(x, v) \leqslant& \\ &\frac{\delta^3}{8}\left[|\nabla U(x)|^2 + (L + \delta L^2)Z^2 + L_\ell Z^3 \left(1 + (|x - x_\star| + \delta Z)^{\ell - 2}\right)\right]. \end{aligned}$$

Proof. Since $1 - e^{-s} \leqslant s$ for all $s > 0$,

$$1 - \alpha \;\leqslant\; (H \circ \Phi_V - H)_+ \;\leqslant\; |H \circ \Phi_V - H|.$$

For $x, v \in \mathbb{R}^d$, denoting $\xi = \delta v - \delta^2/2\nabla U(x)$ and $\overline{\nabla} U = (\nabla U(x) + \nabla U(x+\xi))/2$,

$$
\begin{aligned}
H(\Phi_V(x,v)) &- H(x,v) \\
&= U(x+\xi) - U(x) + \frac{1}{2}|v - \delta\overline{\nabla} U|^2 - \frac{1}{2}|v|^2 \\
&= \xi \cdot \int_0^1 \nabla U(x+s\xi)\mathrm{d}s - \delta v \cdot \overline{\nabla} U + \delta^2/2|\overline{\nabla} U|^2 \\
&= \xi \cdot \left(\int_0^1 \nabla U(x+s\xi)\mathrm{d}s - \overline{\nabla} U \right) + \delta^2/2\overline{\nabla} U \cdot \left(\overline{\nabla} U - \nabla U(x) \right) \\
&= \xi \cdot \int_0^1 (2g(s) - g(0) - g(1))\mathrm{d}s + \delta^2/2 g(1) \cdot (g(1) - g(0))
\end{aligned}
$$

with $g(s) = (\nabla U(x) + \nabla U(x+s\xi))/2$ for $s \in [0,1]$. Under Assumption (∇**Lip**),

$$
\begin{aligned}
|g(1) \cdot (g(1) - g(0))| &\leqslant \frac{1}{2}L|g(0)||\xi| + \frac{1}{4}L^2|\xi|^2 \\
&\leqslant \delta\frac{L}{4}\left(|\nabla U(x)|^2 + (1+L\delta)|\xi/\delta|^2 \right),
\end{aligned}
$$

and similarly, for all $s \in [0,1]$,

$$|2g(s) - g(0) - g(1)| \;\leqslant\; \frac{1}{2}L|\xi|.$$

We have finally obtained, in the case where Assumption (∇**Lip**) alone is assumed,

$$|H(\Phi_V(x,v)) - H(x,v)| \;\leqslant\; \delta^3\frac{L}{8}\left(|\nabla U(x)|^2 + (1+L\delta)|\xi/\delta|^2 \right) + \frac{1}{2}L|\xi|^2$$

and bounding $|\xi| \leqslant \delta|v| + \delta^2/2|\nabla U(x)|$ concludes the proof of the first claim.

However, under additional assumptions, we can treat more carefully the term

$$
\begin{aligned}
2\int_0^1 g(s)\mathrm{d}s - g(0) - g(1) &= 2\int_0^1 \int_0^s g'(u)\mathrm{d}u\mathrm{d}s - \int_0^1 g'(s)\mathrm{d}s \\
&= \int_0^1 (1-2s)g'(s)\mathrm{d}s \\
&= \int_0^1 (1-2s)\left(g'(s) - g'(0)\right)\mathrm{d}s. \qquad (15)
\end{aligned}
$$

Under Assumption (∇^2**pol**(ℓ)), as $\int_0^1 |1-2s|s\mathrm{d}s \leqslant 1/2$,

$$\left| \int_0^1 (1-2s)\left(g'(s) - g'(0)\right)\mathrm{d}s \right|$$

$$\leq \frac{1}{16}L_\ell|\xi|^2\left(2 + |x - x_\star|^{\ell-2} + (|x - x_\star| + |\xi|)^{\ell-2}\right)$$

$$\leq \frac{1}{8}L_\ell|\xi|^2\left(1 + (|x - x_\star| + |\xi|)^{\ell-2}\right). \qquad \Box$$

Lemma 29.

1. *Under Assumption* (∇**Lip**)*, for all* $(x, v) \in \mathbb{R}^{2d}$*, writing* $Z = |v| + \delta/2|\nabla U(x)|$,

$$|\Phi_V(x, v) - \Phi_{HD}(x, v)| \leq L\delta^2\left[1 + e^{L\delta^2}\left(\delta/6 + \delta^2 L/24\right)\right]Z.$$

2. *Under Assumptions* (∇**Lip**) *and* (∇^2**pol**(ℓ))*, for all* $(x, v) \in \mathbb{R}^{2d}$*, writing* $Z = |v| + \delta/2|\nabla U(x)|$,

$$|\Phi_V(x, v) - \Phi_{HD}(x, v)| \leq \frac{\delta^3}{4}|\nabla U(x)| + \frac{L\delta^3}{6}\left[1 + \delta\frac{L+1}{2} + \delta^3\frac{L}{40}\right]e^{L\delta^2}Z$$

$$+ \delta^3\frac{L_\ell}{8}\left(1 + e^{L\delta^2}\frac{\delta^2}{12}\right)Z^2\left(1 + \left[|x - x_\star| + \left(\delta + e^{L\delta^2}\frac{\delta^3}{6}\right)Z\right]^{\ell-2}\right).$$

Proof. Fix $x, v \in \mathbb{R}^{2d}$. Assumption (∇**Lip**) holds in the whole proof. Consider $(\hat{x}_t, \hat{v}_t)_{t\geq 0}$ given by (14) and $(y_t, w_t)_{t\geq 0}$ the solution of

$$y'_t = w_t, \qquad w'_t = -\nabla U(x), \qquad y_0 = x, \qquad w_0 = v.$$

In other words,

$$y_t = x + tv - \frac{t^2}{2}\nabla U(x), \qquad w_t = v - t\nabla U(x),$$

so that $\Phi_V(x, v) = (y_\delta, \tilde{w}_\delta)$ with $\tilde{w}_\delta = w_\delta + \delta/2(\nabla U(y_\delta) - \nabla U(x))$. Then,

$$|\hat{x}_t - y_t| = \left|\int_0^t\int_0^s(\nabla U(\hat{x}_u) - \nabla U(x))\mathrm{d}u\mathrm{d}s\right|$$

$$\leq L\int_0^t\int_0^s\left(|\hat{x}_u - y_u| + |y_u - x|\right)\mathrm{d}u\mathrm{d}s$$

$$\leq L\int_0^t\int_u^t\left(|\hat{x}_u - y_u| + u|v| + u^2/2|\nabla U(x)|\right)\mathrm{d}s\mathrm{d}u$$

$$\leq L\int_0^t(t - u)|\hat{x}_u - y_u|\mathrm{d}u + \frac{Lt^3}{6}|v| + \frac{Lt^4}{24}|\nabla U(x)|.$$

Bounding $(t - u) \leq t$ and applying Gronwall's Lemma, we get for all $t \in [0, \delta]$

$$|\hat{x}_t - y_t| \leq e^{Lt^2}\left(\frac{Lt^3}{6}|v| + \frac{Lt^4}{24}|\nabla U(x)|\right) \leq e^{Lt^2}\frac{Lt^3}{6}Z.$$

Similarly, for all $t \in [0, \delta]$,

$$|\hat{v}_t - w_t| = \left|\int_0^t(\nabla U(\hat{x}_s) - \nabla U(x))\mathrm{d}u\right|$$

$$\leqslant \quad L \int_0^t \left(|\hat{x}_s - y_s| + |y_s - x| \right) \mathrm{d}s$$

$$\leqslant \quad e^{Lt^2} \frac{L^2 t^4}{24} Z + \frac{Lt^2}{2} Z \,.$$

As a consequence,

$$|\hat{v}_\delta - \tilde{w}_\delta| \quad \leqslant \quad |\hat{v}_\delta - w_\delta| + \frac{\delta L}{2} |y_\delta - x|$$

$$\leqslant \quad e^{L\delta^2} \frac{L^2 \delta^4}{24} Z + L\delta^2 Z \,.$$

The conclusion of the proof of the first claim follows from the previous estimates and

$$|\Phi_V(x, v) - \Phi_{HD}(x, v)| \quad \leqslant \quad |\hat{x}_\delta - y_\delta| + |\hat{v}_\delta - \tilde{w}_\delta| \,.$$

Note that $|\hat{v}_\delta - w_\delta|$ yields a first order error (in δ). Now, if an additional assumption is available, in order to see that the Verlet scheme is in fact a second order integrator, we should rather directly bound

$$|\hat{v}_\delta - \tilde{w}_\delta| \quad = \quad \left| \int_0^\delta \nabla U(\hat{x}_s) \mathrm{d}s - \delta \frac{\nabla U(x) + \nabla U(y_\delta)}{2} \right|$$

$$\leqslant \quad \left| \int_0^\delta \nabla U(\hat{x}_s) \mathrm{d}s - \delta \frac{\nabla U(x) + \nabla U(\hat{x}_\delta)}{2} \right| + \frac{L\delta}{2} |y_\delta - \hat{x}_\delta| \,.$$

As in the proof of Lemma 28 (more precisely, (15)), denoting $g(s) = \nabla U(\hat{x}_s)$,

$$\left| \int_0^\delta g(s) \mathrm{d}s - \delta \frac{g(0) + g(\delta)}{2} \right| \quad = \quad \left| \int_0^\delta \left(\frac{\delta}{2} - u \right) (g'(s) - g'(0)) \, \mathrm{d}s \right| \,,$$

and

$$|g'(s) - g'(0)| \quad \leqslant \quad \left| \left(\nabla^2 U(\hat{x}_s) - \nabla^2 U(x) \right) v \right| + \left| \nabla^2 U(x) \left(\hat{v}_s - v \right) \right|$$

$$\leqslant \quad |\nabla^2 U(\hat{x}_s) - \nabla^2 U(x)||v| + L \left(|\hat{v}_s - w_s| + |w_s - v| \right) \,.$$

Under Assumption $(\nabla^2 \mathbf{pol}(\ell))$,

$$|\nabla^2 U(\hat{x}_s) - \nabla^2 U(x)| \quad \leqslant \quad \frac{L_\ell}{2} \left(|y_s - x| + |\hat{x}_s - y_s| \right)$$

$$\times \left(1 + \left(|x - x_\star| + |y_s - x| + |\hat{x}_s - y_s| \right)^{\ell - 2} \right)$$

$$\leqslant \quad \frac{L_\ell}{2} \left(s + e^{L\delta^2} \frac{s^3}{6} \right) Z A_\ell \,,$$

with

$$A_\ell \quad = \quad 1 + \left[|x - x_\star| + \left(\delta + e^{L\delta^2} \frac{\delta^3}{6} \right) Z \right]^{\ell - 2} \,,$$

so that

$$\frac{\delta}{2}\int_0^\delta |v||\nabla^2 U(\hat{x}_s) - \nabla^2 U(x)|\mathrm{d}s \quad \leqslant \quad \delta^3 \frac{L_\ell}{8}\left(1 + e^{L\delta^2}\frac{\delta^2}{12}\right)Z^2 A_\ell\,.$$

Gathering the previous bounds,

$$
\begin{aligned}
|y_\delta - \hat{x}_\delta| + |\hat{v}_\delta - \tilde{w}_\delta| \quad \leqslant\quad & \frac{\delta}{2}\int_0^\delta |g'(s) - g'(0)|\mathrm{d}s + \left(1 + \frac{L\delta}{2}\right)|y_\delta - \hat{x}_\delta| \\
\leqslant\quad & \delta^3 \frac{L_\ell}{8}\left(1 + e^{L\delta^2}\frac{\delta^2}{12}\right)Z^2 A_\ell \\
& + \frac{\delta}{2}\int_0^\delta \left(e^{Ls^2}\frac{L^2 s^4}{24}Z + \frac{Ls^2}{2}Z + s|\nabla U(x)|\right)\mathrm{d}t \\
& + \left(1 + \frac{L\delta}{2}\right)e^{L\delta^2}\frac{L\delta^3}{6}Z \\
\leqslant\quad & \delta^3 \frac{L_\ell}{8}\left(1 + e^{L\delta^2}\frac{\delta^2}{12}\right)Z^2 A_\ell + \frac{\delta^3}{4}|\nabla U(x)| \\
& + \left[1 + \delta\frac{L+1}{2} + \delta^3\frac{L}{40}\right]e^{L\delta^2}\frac{L\delta^3}{6}Z\,. \qquad \square
\end{aligned}
$$

Recall that π_1 denotes the first marginal of π, namely the probability law on $\mathbb{R}^d$ with density $\propto \exp(-U)$. Similarly, denote by π_2 the second marginal of π, which is simply the standard Gaussian distribution on $\mathbb{R}^d$.

Lemma 30. *Under Assumption* (∇**Lip**) *and* (**Conv**), *for all* $\beta \geqslant 0$,

$$\int_{\mathbb{R}^{2d}} |\nabla U(x)|^\beta \pi_1(\mathrm{d}x) \leqslant L^\beta \int_{\mathbb{R}^{2d}} |x - x_\star|^\beta \pi_1(\mathrm{d}x) \leqslant L^\beta \left(\frac{d + (\beta-2)_+}{m}\right)^{\beta/2}$$

and

$$\int_{\mathbb{R}^{2d}} |v|^\beta \pi_2(\mathrm{d}v) \leqslant (d + (\beta-2)_+)^{\beta/2}\,.$$

Proof. Consider $\mathcal{L} = -\nabla U + \Delta$ the generator of the overdamped Langevin diffusion, which leaves invariant π_1. For $\zeta \geqslant 1$ and $\varphi(x) = |x|^{2\zeta}$,

$$
\begin{aligned}
\mathcal{L}\varphi(x) \quad =\quad & -2\zeta|x|^{2(\zeta-1)}\nabla U(x)\cdot x + |x|^{2(\zeta-1)}\left(2\zeta d + 4\zeta(\zeta-1)\right) \\
\leqslant\quad & -2m\zeta|x|^{2\zeta} + |x|^{2(\zeta-1)}\left(2\zeta d + 4\zeta(\zeta-1)\right)\,.
\end{aligned}
$$

The invariance of π_1 implies that $\mathbb{E}_{\pi_1}(\mathcal{L}\varphi) = 0$, so that

$$
\begin{aligned}
0 \quad \leqslant\quad & -2m\zeta\mathbb{E}_{\pi_1}(\varphi) + (2\zeta d + 4\zeta(\zeta-1))\mathbb{E}_{\pi_1}\left(\varphi^{(\zeta-1)/\zeta}\right) \\
\leqslant\quad & -2m\zeta\mathbb{E}_{\pi_1}(\varphi) + (2\zeta d + 4\zeta(\zeta-1))\left(\mathbb{E}_{\pi_1}(\varphi)\right)^{(\zeta-1)/\zeta}
\end{aligned}
$$

from Jensen's inequality, and thus

$$\left(\mathbb{E}_{\pi_1}(\varphi)\right)^{1/\zeta} \quad \leqslant \quad \frac{d + 2(\zeta-1)}{m}\,.$$

This concludes the case of $\beta = 2\zeta \geqslant 2$. The case $\beta \in [0,2)$ is obtained by Jensen's inequality. Applying this result to the potential $U(x) = |x|^2/2$ (so that $m = 1$) bounds the moments for the velocity. $\square$

We can now gather the results of the last four lemmas to get the following.

Proposition 31.

1. *Under Assumptions* (∇**Lip**) *and* (**Conv**), *for all* $p \geqslant 1$,

$$\begin{aligned} \mathcal{W}_p(\pi, \pi P_V) &\leqslant \delta^2 K_3 \sqrt{d + (p-2)_+} \\ \|\pi - \pi P_V\|_{TV} &\leqslant \delta^2 K_4 d \end{aligned}$$

   with

$$K_3 = L\left[1 + e^{L\delta^2}\left(\frac{\delta}{6} + \frac{\delta^2 L}{24}\right)\right]\left(1 + \frac{\delta L}{2\sqrt{m}}\right)$$

$$K_4 = L\left(1 + \delta\frac{1}{4} + \delta^2\frac{L}{4}\right)\left(1 + \delta\frac{L}{\sqrt{m}} + \frac{\delta^2 L^2}{4m}\right) + \delta\frac{L^2}{m}.$$

2. *Under Assumptions* (∇**Lip**), (**Conv**) *and* (∇^2**pol**(ℓ)), *for all* $p \geqslant 1$,

$$\mathcal{W}_p(\pi, \pi P_V) \leqslant \delta^3\left(K_5\sqrt{d + (p-2)_+} + L_\ell K_6(d + \ell p - 2)^{\ell/2}\right)$$

$$\|\pi - \pi P_V\|_{TV} \leqslant \delta^3\left(K_7 d + L_\ell K_8(d + \ell - 1)^{(\ell+1)/2}\right)$$

   with

$$K_5 = \frac{L}{4\sqrt{m}} + \frac{L}{6}\left[1 + \delta\frac{L+1}{2} + \delta^3\frac{L}{40}\right]e^{L\delta^2}\left(1 + \delta\frac{L}{2\sqrt{m}}\right)$$

$$K_6 = 2^{\ell-2}\left(1 + e^{L\delta^2}\frac{\delta^2}{12}\right)\left(1 + \delta + e^{L\delta^2}\frac{\delta^3}{6}\right)^{\ell-2}\left(1 + \frac{\delta L}{2}\right)^\ell\left(1 + \frac{1}{m^{\ell/2}}\right)$$

$$K_7 = L(1 + \delta L)\left(1 + \frac{\delta^2 L^2}{4m}\right)$$

$$K_8 = 2^{\ell+1}\left(1 + \delta + \frac{\delta^2 L}{2}\right)^{\ell-2}\left(1 + \frac{\delta L}{2}\right)^3\left(1 + \frac{1}{m^{(\ell+1)/2}}\right).$$

Proof. Both cases are similar so we only detail the first one. For $h \geqslant 1$, denote

$$m_{x,h} = \int_{\mathbb{R}^{2d}} |\nabla U(x)|^h \pi_1(\mathrm{d}x) \qquad \text{and} \qquad m_{v,h} = \int_{\mathbb{R}^{2d}} |v|^h \pi_2(\mathrm{d}v).$$

For $p \geqslant 1$, using Lemmas 26 and 29 and the invariance of π by Q_δ,

$$\begin{aligned} \mathcal{W}_p(\pi, \pi P_V) &= \mathcal{W}_p(\pi Q_\delta, \pi P_V) \\ &\leqslant L\delta^2\left[1 + e^{L\delta^2}\left(\delta/6 + \delta^2 L/24\right)\right]\left(\mathbb{E}_\pi\left[(|V| + \delta/2|\nabla U(X)|)^p\right]\right)^{1/p} \\ &\leqslant L\delta^2\left[1 + e^{L\delta^2}\left(\delta/6 + \delta^2 L/24\right)\right]\left[(m_{v,p})^{1/p} + \delta/2(m_{x,p})^{1/p}\right] \end{aligned}$$

$$\leqslant \delta^2 K_3 \sqrt{d + (p-2)_+}\,,$$

where we used Lemma 30. Similarly, from Lemmas 26, 28 and 30,

$$
\begin{aligned}
\|\pi - \pi P_V\|_{TV} & \\
\leqslant \quad & \delta^3 \frac{1}{4} \mathbb{E}_\pi \left(|\nabla U(X)|^2\right) + \delta^2 L \left(1 + \delta \frac{1}{4} + \delta^2 \frac{L}{4}\right) \mathbb{E}_\pi \left((|V| + \delta/2|\nabla U(X)|)^2\right) \\
= \quad & \delta^3 \frac{1}{4} m_{x,2} + \delta^2 L \left(1 + \delta \frac{1}{4} + \delta^2 \frac{L}{4}\right) \left(m_{v,2} + \delta m_{v,1} m_{x,1} + \frac{\delta^2}{4} m_{x,2}\right) \\
\leqslant \quad & \delta^2 d K_4\,.
\end{aligned}
$$

Remark that in the computations of the second case, to simplify the last term involved in the bound on $|\Phi_V - \Phi_{HD}|$ in Lemma 29, we use that

$$
\begin{aligned}
A'_\ell \;:=\; & 1 + \left[|X - x_\star| + \left(\delta + e^{L\delta^2} \frac{\delta^3}{6}\right) \left(|V| + \frac{\delta}{2}|\nabla U(X)|\right)\right]^{\ell-2} \\
\leqslant \; & 1 + \left(1 + \delta + e^{L\delta^2} \frac{\delta^3}{6}\right)^{\ell-2} \left[|X - x_\star| \left(1 + \frac{\delta L}{2}\right) + |V|\right]^{\ell-2} \\
\leqslant \; & \left(1 + \delta + e^{L\delta^2} \frac{\delta^3}{6}\right)^{\ell-2} \left(1 + \frac{\delta L}{2}\right)^{\ell-2} 2^{\ell-3} \left(2 + |X - x_\star|^{\ell-2} + |V|^{\ell-2}\right)
\end{aligned}
$$

and then

$$
\left(|V| + \frac{\delta}{2}|\nabla U(X)|\right)^2 A'_\ell \;\leqslant\; 2 \left(1 + \frac{\delta L}{2}\right)^2 \left(|V|^2 + |X - x_\star|^2\right) A'_\ell\,.
$$

Using then Lemma 30 to bound

$$
\begin{aligned}
\left(\mathbb{E}_\pi \left[\left(|V|^2 + |X - x_\star|^2\right)^p \left(2 + |X - x_\star|^{\ell-2} + |V|^{\ell-2})^p\right)\right]\right)^{1/p} & \\
\leqslant \; 2 \left(\mathbb{E}\left(|V|^{2p}\right)\right)^{1/p} + 2 \left(\mathbb{E}\left(|X - x_\star|^{2p}\right)\right)^{1/p} & \\
+ \; 3 \left(\mathbb{E}\left(|V|^{\ell p}\right)\right)^{1/p} + 3 \left(\mathbb{E}\left(|X - x_\star|^{\ell p}\right)\right)^{1/p} & \\
\leqslant \; 5 \left(d + \ell p - 2\right)^{\ell/2} \left(1 + \frac{1}{m^{\ell/2}}\right) &
\end{aligned}
$$

yields the claimed expression for K_6. Similarly, to bound the last term given in Lemma 28, we use that

$$
\begin{aligned}
1 + \left(|X - x_\star| + \delta \left(|V| + \frac{\delta}{2}|\nabla U(X)|\right)\right)^{\ell-2} & \\
\leqslant \; \left(1 + \delta + \frac{\delta^2 L}{2}\right)^{\ell-2} \left(1 + (|V| + |X - x_\star|)^{\ell-2}\right) & \\
\leqslant \; 2^{\ell-3} \left(1 + \delta + \frac{\delta^2 L}{2}\right)^{\ell-2} \left(2 + |V|^{\ell-2} + |X - x_\star|^{\ell-2}\right)\,, &
\end{aligned}
$$

then

$$\left(|V| + \frac{\delta}{2} |\nabla U(X)| \right)^3 \;\leqslant\; 4 \left(1 + \frac{\delta L}{2} \right)^3 \left(|V|^3 + |X - x_\star|^3 \right),$$

and proceed similarly to K_6 to get the expression of K_8.   □

The case of a separable target ensues from the following result.

Lemma 32. *Let* $p \geqslant 2$ *and, for* $i \in [\![1, d]\!]$, *let* $\nu_i, \mu_i \in \mathcal{P}_p(\mathbb{R})$ *and* $\nu = \otimes_{i=1}^d \nu_i$ *and* $\mu = \otimes_{i=1}^d \mu_i$. *Then*

$$\mathcal{W}_p^2(\nu, \mu) \;\leqslant\; \sum_{i=1}^d \mathcal{W}_p^2(\nu_i, \mu_i)$$

$$\|\nu - \mu\|_{TV} \;\leqslant\; \sum_{i=1}^d \|\nu_i - \mu_i\|_{TV}\,.$$

Proof. For all $i \in [\![1, d]\!]$, let (X_i, Y_i) optimal $\mathcal{W}_p$ coupling of ν_i and μ_i, with (X_i, Y_i) independent from (X_j, Y_j) for $j \neq i$. Then $X = (X_i, \ldots, X_d)$ and $Y = (Y_i, \ldots, Y_d)$ are respectively distributed according to ν and μ so that, for $p \geqslant 2$,

$$\mathcal{W}_p^2(\nu, \mu) \;\leqslant\; \left(\mathbb{E} \left[\left(\sum_{i=1}^d |X_i - Y_i|^2 \right)^{p/2} \right] \right)^{2/p}$$

$$\leqslant\; \sum_{i=1}^d \left(\mathbb{E} \left[|X_i - Y_i|^p \right] \right)^{2/p} \;=\; \sum_{i=1}^d \mathcal{W}_p^2(\nu_i, \mu_i)\,.$$

Alternatively, if the (X_i, Y_i)'s are optimal couplings of ν_i and μ_i for the total variation distance, then

$$\|\nu - \mu\|_{TV} \;\leqslant\; 2\mathbb{P}\left(\exists i \in [\![1, d]\!],\; X_i \neq Y_i \right) \;\leqslant\; 2 \sum_{i=1}^d \mathbb{P}\left(X_i \neq Y_i \right)$$

$$=\; \sum_{i=1}^d \|\nu_i - \mu_i\|_{TV}\,.   \quad\square$$

As a conclusion of the analysis of the equilibrium bias:

Proof of Propositions 11 and 12. Simply combine Propositions 9 and 31. More precisely, for the total variation, to prove Proposition 11, apply Proposition 9 with

$$n \;:=\; 1 + \left\lceil \frac{\left| \ln \left(\delta^3 d \right) \right|}{\left| \ln(1 - \delta\kappa) \right|} \right\rceil \;\leqslant\; 2 + \frac{\left| \ln \left(\delta^3 d \right) \right|}{\delta\kappa}\,,$$

so that

$$(1 - \delta\kappa)^{(n-1)/2} \;\leqslant\; \delta^{3/2} \sqrt{d}\,.$$

This choice of $n \geqslant 1$ has been motivated by the fact we want to minimize a bound of the order of

$$(1 - \delta\kappa)^n \sqrt{d/\delta} + n\delta^2 d = \sqrt{d/\delta} \left((1 - \delta\kappa)^n + (n\delta)\sqrt{\delta^3 d} \right) .$$

Remark that there is no way to deduce from this a bound that goes to zero if $e^{-1/(d\delta)}\sqrt{d/\delta}$ doesn't go to zero (which is not implied by the fact $d\delta \to 0$, since even in that case $e^{-1/(d\delta)}\sqrt{d\delta}$ may not be negligible with respect to δ). We chose n so that the terms $(1 - \delta\kappa)^n$ and $(n\delta)\sqrt{\delta^3 d}$ are equal; in fact taking into account the $n\delta$ factor in the second term only leads to a $\ln \ln$ correction so we simply take $(1 - \delta\kappa)^n = \sqrt{\delta^3 d}$.

Similarly, to prove Proposition 12, apply Proposition 9 with

$$n := 1 + \left\lceil \frac{\left| \ln \left(\delta^{3/2}\sqrt{d + \ell - 1} \right) \right|}{|\ln(1 - \delta\kappa)|} \right\rceil \leqslant 2 + \frac{\left| \ln \left(\delta^{3/2}\sqrt{d + \ell - 1} \right) \right|}{\delta\kappa} ,$$

so that

$$(1 - \delta\kappa)^{(n-1)/2} \leqslant \delta^{3/2}\sqrt{d + \ell - 1} .$$

Finally, under the additional condition ($\perp\!\!\!\perp$), notice that necessarily each of the potentials U_i for $i \in [\![1, d]\!]$ satisfies the conditions (∇**Lip**) and (**Conv**) with the same constants L and m than U. Moreover, the $\mathcal{W}_p$ and total variation distances are unchanged by an orthonormal change of coordinates, so without loss of generality we suppose that $U(x) = \sum_{i=1}^{d} U_i(x_i)$. In particular, $\pi = \otimes_{i=1}^{d}\nu_i$ where ν_i is the law of (X_i, V_i) where $(X, V) \sim \pi$. Moreover, the coordinates of the OBABO chain are independent one-dimensional OBABO chains associated to the potentials U_i, $i \in [\![1, d]\!]$, and thus similarly $\pi_\delta = \otimes_{i=1}^{d}\nu_{i,\delta}$ where $\nu_{i,\delta}$ is the equilibrium of the one-dimensional OBABO chain with potential U_i. The first part of Proposition 12 applies to $\nu_{i,\delta}$ (with $d = 1$) for all $i \in [\![1, d]\!]$, which together with Lemma 32 concludes. $\qquad\square$

6. Analysis of the Metropolis-adjusted algorithm

The results on the OBABO transition P are transfered to the OM(BAB)O transition P_M thanks to the following lemma.

Lemma 33. *For all $\nu, \mu \in \mathcal{P}(\mathbb{R}^{2d})$ and $n \in \mathbb{N}$,*

$$\|\nu P_M^n - \mu P_M^n\|_{TV} \leqslant \|\nu P^n - \mu P^n\|_{TV} + 2\sum_{k=0}^{n-1}(\nu + \mu)P^k P_O(1 - \alpha) .$$

Moreover, for $\nu \in \mathcal{P}(\mathbb{R}^{2d})$, let $(Z_k)_{k\in\mathbb{N}}$ and $(\tilde{Z}_k)_{k\in\mathbb{N}}$ be respectively an OBABO and OM(BAB)O chain with initial condition $Z_0 = \tilde{Z}_0 \sim \nu$. Then, for all $u \geqslant 0$ and all $\varphi \in L^1(\pi)$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n}\varphi(\tilde{Z}_k)-\pi(\varphi)\right|\geqslant u\right)\quad\leqslant\quad\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n}\varphi(Z_k)-\pi(\varphi)\right|\geqslant u\right)$$
$$+\sum_{k=0}^{n-1}\nu P^k P_O(1-\alpha)\,.$$

Proof. For the first claim, simply bound

$$\|\nu P_M^n-\mu P_M^n\|_{TV}\leqslant\|\nu P^n-\mu P^n\|_{TV}+\|\nu P_M^n-\nu P^n\|_{TV}+\|\mu P^n-\mu P_M^n\|_{TV}\,.$$

The two last terms are similar, we bound

$$\|\nu P_M^n-\nu P^n\|_{TV}\quad\leqslant\quad\sum_{k=1}^{n}\|\nu P^k P_M^{n-k}-\nu P^{k-1}P_M^{n-k+1}\|_{TV}$$
$$\leqslant\quad\sum_{k=1}^{n}\|\nu P^{k-1}(P-P_M)\|_{TV}$$
$$\leqslant\quad\sum_{k=1}^{n}\|\nu P^{k-1}P_O(P_V-P_{MV})\|_{TV}$$

Conclusion follows from Lemma 26.

The second claim is obtained simply by saying that an OM(BAB)O and a OBABO chains starting at the same point and sampled with the same random variables remain equal up to the first rejection of the OM(BAB)O chain. More precisely, let $(G_n,G_n',W_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence of random variables such that, for all $n\in\mathbb{N}$, G_n, G_n' and W_n are independent, $G_n,G_n'\sim\mathcal{N}(0,I_d)$ and W_n is uniformly distributed over $[0,1]$. Let $Z_0\sim\nu$ be independent from these variables. We construct $(Z_k)_{k\in\mathbb{N}}$ and $(\tilde{Z}_k)_{k\in\mathbb{N}}$ as the OBABO and OM(BAB)O chains with initial condition Z_0 and whose transitions at step n are given respectively in Section 2.1 with variables G_n,G_n' or in Section 2.2 with variables G_n,G_n',W_n. In other words, the damping parts O in both chains use the same Gaussian variables, and the only difference between $Z=(X,V)$ and $\tilde{Z}=(\tilde{X},\tilde{V})$ is that the proposal given by the Verlet step BAB is always accepted for Z and is accepted iff $W_n\leqslant\alpha(\tilde{Z}_n')$ for $\tilde{Z}$, where $\tilde{Z}_n'=(\tilde{X}_n,\eta\tilde{V}_n+\sqrt{1-\eta^2}G_n)$. In particular, the two chains are equal up to the first rejection, in other words, for all $n\geqslant 1$,

$$\mathbb{P}\left(Z_k=\tilde{Z}_k\ \forall k\in[\![1,n]\!]\right)$$
$$=\mathbb{P}\left(W_k\leqslant\alpha\left(X_k,\eta V_k+\sqrt{1-\eta^2}G_k\right)\ \forall k\in[\![0,n-1]\!]\right).$$

Bounding the probability of the union by the sum of the probabilities and using that $(X_k,\eta V_k+\sqrt{1-\eta^2}G_k)\sim\nu P^k P_O$ for all $k\in\mathbb{N}$, we get that

$$A\ :=\ \mathbb{P}\left(\exists k\in[\![1,n]\!],(Z_k)\neq(\tilde{Z}_k)\right)\quad\leqslant\quad\sum_{k=0}^{n-1}\nu P^k P_O(1-\alpha)\,,$$

and for $u \geqslant 0$ we bound

$$
\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n}\varphi(\tilde{Z}_k)-\pi(\varphi)\right|\geqslant u\right) \quad\leqslant\quad A+\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n}\varphi(Z_k)-\pi(\varphi)\right|\geqslant u\right). \quad\square
$$

In view of this result and of Lemma 28, in order to prove Proposition 14 and Theorem 16, it only remains to establish moment estimates for $\nu P^k P_O$ for $\nu \in \mathcal{P}_p(\mathbb{R}^{2d})$, which is the subject of the following lemma. In the rest of this section, denote $z_\star = (x_\star, 0)$ and $\varphi_\star(z) = |z - z_\star|$ for $z \in \mathbb{R}^{2d}$.

Proposition 34. *Under the conditions of Proposition 17, for all $p \geqslant 2$, $\nu \in \mathcal{P}_p(\mathbb{R}^{2d})$ and $n \in \mathbb{N}$,*

$$
(\nu P^n P_O\,(\varphi_\star^p))^{1/p} \leqslant K_1(1-\delta\kappa)^{n/2}\,(\nu(\varphi_\star^p))^{1/p}
$$
$$
+(K_1+3)\left(1+\frac{1}{\sqrt{m}}+\frac{2\delta K_1 K_3}{\kappa}\right)\sqrt{d+p-2}\,.
$$

Proof. Using that P_O does not increase the $\mathcal{W}_p$ distance (see the proof of Proposition 9) and Theorem 1

$$
\begin{aligned}
(\nu P^n P_O(\varphi_\star^p))^{1/p} &\leqslant \mathcal{W}_p(\nu P^n P_O, \pi_\delta P^n P_O)+(\pi_\delta P^n P_O(\varphi_\star^p))^{1/p}\\
&\leqslant K_1(1-\delta\kappa)^{n/2}\mathcal{W}_p(\nu,\pi_\delta)+(\pi_\delta P_O(\varphi_\star^p))^{1/p}\\
&\leqslant K_1(1-\delta\kappa)^{n/2}\,(\nu(\varphi_\star^p))^{1/p}+K_1\,(\pi_\delta(\varphi_\star^p))^{1/p}+(\pi_\delta P_O(\varphi_\star^p))^{1/p}\,.
\end{aligned}
$$

Let $(X,V)\sim\pi_\delta$ and, independently, $G\sim\mathcal{N}(0,I_d)$, so that $(X,\eta V+\sqrt{1-\eta^2}G)\sim\pi_\delta P_O$. For $p\geqslant 1$,

$$
\begin{aligned}
(\pi_\delta P_O(\varphi_\star^p))^{1/p} &= \left(\mathbb{E}\left(|(X,\eta V+\sqrt{1-\eta^2}G)-z_\star|^p\right)\right)^{1/p}\\
&\leqslant (\mathbb{E}\,(|(X,\eta V)-z_\star|^p))^{1/p}+\sqrt{1-\eta^2}\,(\mathbb{E}\,(|G|^p))^{1/p}\\
&\leqslant (\mathbb{E}\,(|(X,V)-z_\star|^p))^{1/p}+\sqrt{d+p-2}\,,
\end{aligned}
$$

so that

$$
(\nu P^n P_O(\varphi_\star^p))^{1/p} \leqslant K_1(1-\delta\kappa)^{n/2}\,(\nu(\varphi_\star^p))^{1/p}
$$
$$
+(K_1+1)\,(\pi_\delta(\varphi_\star^p))^{1/p}+\sqrt{d+p-2}\,.
$$

The moments of π_δ are bounded by combining Lemma 30 and Proposition 11, as

$$
\begin{aligned}
(\pi_\delta(\varphi_\star^p))^{1/p} &\leqslant (\pi(\varphi_\star^p))^{1/p}+\mathcal{W}_p\,(\pi,\pi_\delta)\\
&\leqslant \left(1+\frac{1}{\sqrt{m}}+\frac{2\delta K_1 K_3}{\kappa}\right)\sqrt{d+p-2}\,. \qquad\square
\end{aligned}
$$

Corollary 35. *Under the conditions of Proposition 17:*

1. *For all $\nu \in \mathcal{P}_2(\mathbb{R}^{2d})$ and $n \in \mathbb{N}$,*

$$\sum_{k=0}^{n-1} \nu P^k P_O(1-\alpha) \leqslant K_9 \left(\delta \nu(\varphi_\star^2) + \delta^2 nd \right)$$

   with, considering K_1 and K_3 as in Theorem 1 and Proposition 31,

$$K_9 = 2L(K_1+3)^2 \left(1 + \frac{1}{\sqrt{m}} + \frac{2\delta K_1 K_3}{\kappa} \right)^2 \left(1 + \frac{1}{\kappa} \right)$$
$$\times \left[\frac{\delta L}{8} + \left(1 + \frac{\delta}{4} + \frac{\delta^2 L}{4} \right) \left(1 + \delta^2 \frac{L^2}{4} \right) \right] .$$

2. *If $(\nabla^2\mathbf{pol}(\ell))$ holds then for all $\nu \in \mathcal{P}_{\ell+1}(\mathbb{R}^{2d})$ and $n \in \mathbb{N}$,*

$$\sum_{k=0}^{n-1} \nu P^k P_O(1-\alpha) \leqslant K_{10} \left(\delta^2 \nu(\varphi_\star^{\ell+1}) + \delta^3 n(d+\ell-1)^{(\ell+1)/2} \right)$$

   with, considering K_1 and K_3 as in Theorem 1 and Proposition 31,

$$K_{10} = \left[1 + 2^\ell(K_1+3)^{\ell+1} \left(1 + \frac{1}{\sqrt{m}} + \frac{2\delta K_1 K_3}{\kappa} \right)^{\ell+1} \left(\frac{1}{\kappa} + 1 \right) \right]$$
$$\times \left[\frac{L+2L^2}{4} \left(1 + \delta^2 \frac{L^2}{4} \right) + 2^{(\ell-3)/2} L_\ell \left(1 + \frac{L\delta}{2} \right)^3 \left(1 + \delta + \frac{\delta^2 L}{2} \right)^{\ell-2} \right] .$$

3. *If $(\perp\!\!\!\perp)$ holds then for all $\nu \in \mathcal{P}_{\ell+1}(\mathbb{R}^{2d})$ and $n \in \mathbb{N}$, denoting $\varphi_{\star,i}(z) = |\mathcal{Q}^{-1}(z_i - z_{\star,i})|$ for $z \in \mathbb{R}^{2d}$,*

$$\sum_{k=0}^{n-1} \nu P^k P_O(1-\alpha) \leqslant K_{10} \left(\delta^2 \sum_{i=1}^d \nu(\varphi_{\star,i}^{\ell+1}) + \delta^3 nd\ell^{(\ell+1)/2} \right)$$

   with K_{10} given above.

Proof. The first two claims are straightforward consequences of the control of $1-\alpha$ given by Lemma 28 and of the moment estimates of Proposition 34. Indeed, under Assumptions $(\nabla\mathbf{Lip})$ and $(\mathbf{Conv})$, $|\nabla U(x)| \leqslant L|x - x_\star|$ for all $x \in \mathbb{R}^d$, so that Lemma 28 implies

$$1 - \alpha(x,v) \leqslant \delta^2 L \left[\frac{\delta L}{8} + \left(1 + \frac{\delta}{4} + \frac{\delta^2 L}{4} \right) \left(1 + \delta^2 \frac{L^2}{4} \right) \right] \left(|v|^2 + |x - x_\star|^2 \right) . \tag{16}$$

Similarly, under the additional condition $(\nabla^2\mathbf{pol}(\ell))$, using that

$$1 + \left(|x - x_\star| + \delta \left(|v| + \frac{\delta}{2}|\nabla U(x)| \right) \right)^{\ell-2}$$

$$\leqslant \left(1 + \delta + \frac{\delta^2 L}{2}\right)^{\ell-2} \left(1 + (|v| + |x - x_\star|)^{\ell-2}\right)$$

and that

$$\left(|v| + \frac{\delta}{2}|\nabla U(x)|\right)^3 \leqslant \left(1 + \frac{\delta L}{2}\right)^3 (|v| + |x - x_\star|)^3 ,$$

we see that Lemma 28 implies

$$1 - \alpha(x, v) \leqslant \delta^3 L \left[\frac{L}{8} + \frac{1+L}{4}\left(1 + \delta^2 \frac{L^2}{4}\right)\right] \left(|v|^2 + |x - x_\star|^2\right)$$

$$+ \frac{\delta^3}{8} L_\ell \left(1 + \frac{L\delta}{2}\right)^3 \left(1 + \delta + \frac{\delta^2 L}{2}\right)^{\ell-2} \left((|x - x_\star| + |v|)^3 + (|x - x_\star| + |v|)^{\ell+1}\right)$$

$$\leqslant \delta^3 \left[\frac{L + 2L^2}{4}\left(1 + \delta^2 \frac{L^2}{4}\right)\right.$$

$$\left.+ \frac{L_\ell}{8}\left(1 + \frac{L\delta}{2}\right)^3 \left(1 + \delta + \frac{\delta^2 L}{2}\right)^{\ell-2} \left(2^{3/2} + 2^{(\ell+1)/2}\right)\right] \left(1 + |z - z_\star|^{\ell+1}\right) \tag{17}$$

where we bounded $(|x - x_\star| + |v|)^k$ by $2^{k/2}|z - z_\star|^k$ for $k = 3$ and $\ell+1$ and $|z - z_\star|^k$ by $1 + |z - z_\star|^{\ell+1}$ for $k = 2$ and 3. Moreover, using that $(s + t)^p \leqslant 2^{p-1}(s^p + t^p)$ for $s, t > 0$, we see that Proposition 34 implies that for all $p \geqslant 2$,

$$\nu P^n P_O (\varphi_\star^p) \leqslant 2^{p-1} K_1^p (1 - \delta\kappa)^{np/2} \nu(\varphi_\star^p)$$

$$+ 2^{p-1}(K_1 + 3)^p \left(1 + \frac{1}{\sqrt{m}} + \frac{2\delta K_1 K_3}{\kappa}\right)^p (d + p - 2)^{p/2} ,$$

so that

$$\sum_{k=0}^{n-1} \nu P^k P_O (\varphi_\star^p) \leqslant 2^{p-1}(K_1 + 3)^p \left(1 + \frac{1}{\sqrt{m}} + \frac{2\delta K_1 K_3}{\kappa}\right)^p$$

$$\times \left(\frac{1}{\delta\kappa}\nu(\varphi_\star^p) + n(d + p - 2)^{p/2}\right)$$

Combining this with (16) (with $p = 2$) gives the expression of K_9, and with (17) (with $p = \ell + 1$) gives K_{10}.

In the separable case (condition ($\perp\!\!\!\perp$)), denoting $(y, v) = (\mathcal{Q}^{-1}x, \mathcal{Q}^{-1}v)$ and $H_i(y_i, w_i) = U_i(y_i) + |w_i|^2/2$ for $i \in [\![1, d]\!]$, using that $H(x, v) = \sum_{i=1}^d H_i(y_i, w_i)$ and that $\mathcal{Q}^{-1}\Phi_V = \Phi_V \mathcal{Q}^{-1}$, we get that

$$1 - \alpha(x, v) \leqslant 1 - e^{-\sum_{i=1}^d (H_i \circ \Phi_V(y_i, w_i) - H_i(y_i, w_i))_+} := 1 - \prod_{i=1}^d \alpha_i(y_i, w_i) ,$$

where α_i is the acceptance probability of a one-dimensional OM(BAB)O chain with potential U_i. Bounding the probability of the union by the sum of the probability we get that

$$1 - \alpha(x,v) \;\leqslant\; \sum_{i=1}^{d} \left(1 - \alpha_i(y_i, w_i)\right) .$$

Moreover, the OBABO transition P is such that, up to the orthonormal change of variables $(y,w) = (\mathcal{Q}^{-1}x, \mathcal{Q}^{-1}v)$, the coordinates (position/velocities) of the chain are independent one-dimensional OBABO chains. Similarly the transition P_O preserves this independence. Applying the second claim of Corollary 35 (with $d = 1$) to the ith one-dimensional chain gives

$$\sum_{k=0}^{n-1} \nu P^k P_O (1 - \alpha_i) \;\leqslant\; K_{10} \left(\delta^2 \nu(\varphi_{\star,i}^{\ell+1}) + \delta^3 n \ell^{(\ell+1)/2} \right)$$

for $i \in [\![1, d]\!]$. Summing these d inequalities concludes. $\qquad\square$

Proof of Proposition 14. Simply combine Lemma 33 and Corollary 35. $\qquad\square$

Proof of Theorem 16. This follows from the concentration bound for the unadjusted chain (Theorem 6), the bound between the adjusted and unadjusted process obtained by combining Lemma 33 and Corollary 35, and the dual representation of the $\mathcal{W}_1$ distance (see e.g. [58]) that implies that $|\pi(\varphi) - \pi_\delta(\varphi)| \leqslant \mathcal{W}_1(\pi, \pi_\delta)$ for all 1-Lipschitz function φ. $\qquad\square$

Acknowledgments

References

[1] Francis Bach. On the Effectiveness of Richardson Extrapolation in Machine Learning. *arXiv e-prints*, page arXiv:2002.02835, February 2020.

[2] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014. MR3155209

[3] Joris Bierkens, Paul Fearnhead, and Gareth Roberts. The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.*, 47(3):1288–1320, 2019. MR3911113

[4] François Bolley, Arnaud Guillin, and Florent Malrieu. Trend to equilibrium and particle approximation for a weakly selfconsistent vlasov-fokker-planck equation. *ESAIM: Mathematical Modelling and Numerical Analysis – Modélisation Mathématique et Analyse Numérique*, 44(5):867–884, 2010. MR2731396

[5] Nawaf Bou-Rabee, Andreas Eberle, and Raphael Zimmer. Coupling and convergence for Hamiltonian Monte Carlo. *The Annals of Applied Probability*, 30(3):1209 – 1250, 2020. MR4133372

[6] Nawaf Bou-Rabee and Eric Vanden-Eijnden. Pathwise accuracy and ergodicity of metropolized integrators for sdes. *Communications on Pure and Applied Mathematics*, 63(5):655–696, 2010. MR2583309

[7] Nawaf Bou-Rabee and Eric Vanden-Eijnden. A patch that imparts unconditional stability to explicit integrators for langevin-like equations. *Journal of Computational Physics*, 231(6):2565 – 2580, 2012. MR2881033

[8] Giovanni Bussi and Michele Parrinello. Accurate sampling using langevin dynamics. *Phys. Rev. E*, 75:056707, May 2007.

[9] Djalil Chafaï, Florent Malrieu, and Katy Paroux. On the long time behavior of the TCP window size process. *Stochastic Process. Appl.*, 120(8):1518–1534, 2010. MR2653264

[10] Yuansi Chen, Raaz Dwivedi, Martin J. Wainwright, and Bin Yu. Fast mixing of metropolized hamiltonian monte carlo: Benefits of multi-step gradients. *Journal of Machine Learning Research*, 21(92):1–72, 2020. MR4119160

[11] Xiang Cheng, Niladri S. Chatterji, Yasin Abbasi-Yadkori, Peter L. Bartlett, and Michael I. Jordan. Sharp Convergence Rates for Langevin Dynamics in the Nonconvex Setting. *arXiv e-prints*, page arXiv:1805.01648, May 2018.

[12] Xiang Cheng, Niladri S. Chatterji, Peter L. Bartlett, and Michael I. Jordan. Underdamped langevin MCMC: A non-asymptotic analysis. In Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet, editors, *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 300–323. PMLR, 06–09 Jul 2018.

[13] A. Dalalyan, Lionel Riou-Durand, and Avetik G. Karagulyan. Bounding the error of discretized langevin algorithms for non-strongly log-concave targets. *ArXiv*, abs/1906.08530, 2019.

[14] Arnak S. Dalalyan and Lionel Riou-Durand. On sampling from a log-concave density using kinetic langevin diffusions. *Bernoulli*, 26(3):1956–1988, 08 2020. MR4091098

[15] George Deligiannidis, Daniel Paulin, Alexandre Bouchard-Côté, and Arnaud Doucet. Randomized Hamiltonian Monte Carlo as Scaling Limit of the Bouncy Particle Sampler and Dimension-Free Convergence Rates. *arXiv e-prints*, page arXiv:1808.04299, August 2018. MR3911112

[16] Hacène. Djellout, Arnaud Guillin, and Li-Ming Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.*, 32(3B):2702–2732, 07 2004. MR2078555

[17] Jean Dolbeault, Clément Mouhot, and Christian Schmeiser. Hypocoercivity for kinetic equations with linear relaxation terms. *C. R. Math. Acad. Sci. Paris*, 347(9-10):511–516, 2009. MR2576899

[18] Alain Durmus and Éric Moulines. High-dimensional bayesian inference via the unadjusted langevin algorithm. *Bernoulli*, 25(4A):2854–2882, 11 2019. MR4003567

[19] Alain Durmus, Umut Simsekli, Éric Moulines, Roland Badeau, and Gael

Richard. Stochastic Gradient Richardson-Romberg Markov Chain Monte Carlo. In *Thirtieth Annual Conference on Neural Information Processing Systems (NIPS)*, Barcelone, Spain, December 2016.

[20] Raaz Dwivedi, Yuansi Chen, Martin J. Wainwright, and Bin Yu. Log-concave sampling: Metropolis-hastings algorithms are fast. *Journal of Machine Learning Research*, 20(183):1–42, 2019. MR4048994

[21] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Couplings and quantitative contraction rates for langevin dynamics. *Ann. Probab.*, 47(4):1982–2010, 07 2019. MR3980913

[22] Arnaud Guillin and Pierre Monmarché. Uniform long-time and propagation of chaos estimates for mean field kinetic particles in non-convex landscapes. *arXiv e-prints*, page arXiv:2003.00735, March 2020. MR3646428

[23] Arnaud Guillin and Feng-Yu Wang. Degenerate fokker–planck equations: Bismut formula, gradient estimate and harnack inequality. *Journal of Differential Equations*, 253(1):20 – 40, 2012. MR2917400

[24] Frederic Hérau. Short and long time behavior of the Fokker-Planck equation in a confining potential and applications. *J. Funct. Anal.*, 244(1):95–118, 2007. MR2294477

[25] Alan M. Horowitz. A generalized guided monte carlo algorithm. *Physics Letters B*, 268(2):247 – 252, 1991.

[26] Aldéric Joulin and Yann Ollivier. Curvature, concentration and error estimates for markov chain monte carlo. *Ann. Probab.*, 38(6):2418–2442, 11 2010. MR2683634

[27] Kazumasa Kuwada. Duality on gradient estimates and Wasserstein controls. *J. Funct. Anal.*, 258(11):3758–3774, 2010. MR2606871

[28] Michel Ledoux. The concentration of measure phenomenon. *Mathematical Surveys and Monographs*, 89, 2001. MR1849347

[29] Yin Tat Lee and Santosh S. Vempala. Convergence rate of riemannian hamiltonian monte carlo and faster polytope volume computation. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, page 1115–1121, New York, NY, USA, 2018. Association for Computing Machinery. MR3826321

[30] Benedict Leimkuhler and Charles Matthews. Rational Construction of Stochastic Numerical Methods for Molecular Sampling. *Applied Mathematics Research eXpress*, 2013(1):34–56, 06 2012. MR3040887

[31] Tony Lelièvre, Mathias Rousset, and Gabriel Stoltz. *Free energy computations: A mathematical perspective*. Imperial College Press, 2010. MR2681239

[32] Tony Lelièvre and Gabriel Stoltz. Partial differential equations and stochastic methods in molecular dynamics. *Acta Numerica*, 25:681–880, 2016. MR3509213

[33] Yi-An Ma, Niladri S. Chatterji, Xiang Cheng, Nicolas Flammarion, Peter L. Bartlett, and Michael I. Jordan. Is there an analog of Nesterov acceleration for gradient-based MCMC? *Bernoulli*, 27(3):1942 – 1992, 2021. MR4278799

[34] Neal Madras and Deniz Sezer. Quantitative bounds for markov chain convergence: Wasserstein and total variation distances. *Bernoulli*, 16(3):882–

908, 08 2010. MR2730652

[35] Florent Malrieu and Denis Talay. Concentration inequalities for euler schemes. *Monte Carlo and Quasi-Monte Carlo Methods*, page 355, 2004. MR2208718

[36] Oren Mangoubi and Aaron Smith. Rapid Mixing of Hamiltonian Monte Carlo on Strongly Log-Concave Distributions. *arXiv e-prints*, page arXiv:1708.07114, August 2017.

[37] Oren Mangoubi and Nisheeth K. Vishnoi. Dimensionally Tight Bounds for Second-Order Hamiltonian Monte Carlo. *arXiv e-prints*, page arXiv:1802.08898, February 2018.

[38] Bimal Mishra and Tamar Schlick. The notion of error in langevin dynamics. i. linear analysis. *The Journal of Chemical Physics*, 105(1):299–318, 1996.

[39] P. Monmarché. Efficiency bounds for stochastic gradient MCMC with Romberg interpolation. *in preparation*.

[40] P. Monmarché. Long-time behaviour and propagation of chaos for mean field kinetic particles. *Stochastic Process. Appl.*, 127(6):1721–1737, 2017. MR3646428

[41] Pierre Monmarché. On $\mathcal{H}^1$ and entropic convergence for contractive PDMP. *Electronic Journal of Probability*, 20, December 2015. MR3438742

[42] Pierre Monmarché. Hypocoercivity in metastable settings and kinetic simulated annealing. *Probability Theory and Related Fields*, Jan 2018. MR3877555

[43] Pierre Monmarché. Almost sure contraction for diffusions on $\mathbb{R}^d$. Application to generalised Langevin diffusions. *arXiv e-prints*, page arXiv:2009.10828, September 2020.

[44] Pierre Monmarché. Kinetic walks for sampling. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 17:491, 2020. MR4105928

[45] Radford M. Neal. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 54:113–162, 2010. MR2858447

[46] Yann Ollivier. Ricci curvature of markov chains on metric spaces. *Journal of Functional Analysis*, 256(3):810 – 864, 2009. MR2484937

[47] Michela Ottobre, Natesh S. Pillai, Frank J. Pinski, and Andrew M. Stuart. A function space hmc algorithm with second order langevin diffusion limit. *Bernoulli*, 22(1):60–106, 02 2016. MR3449777

[48] Qian Qin and J. Hobert. On the limitations of single-step drift and minorization in markov chain convergence analysis, 2020.

[49] Gareth O. Roberts and Jeffrey S. Rosenthal. One-shot coupling for certain stochastic recursive sequences. *Stochastic Processes and their Applications*, 99(2):195 – 208, 2002. MR1901153

[50] Gareth O. Roberts and Richard L. Tweedie. Geometric l 2 and l 1 convergence are equivalent for reversible markov chains. *Journal of Applied Probability*, 38(A):37–41, 2001. MR1915532

[51] Anthony Scemama, Tony Lelièvre, Gabriel Stoltz, Eric Cancès, and Michel Caffarel. An efficient sampling algorithm for variational monte carlo. *J Chem Phys.*, 125(11), 2006.

[52] Christof Seiler, Simon Rubinstein-Salzedo, and Susan Holmes. Positive

curvature and hamiltonian monte carlo. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 586–594. Curran Associates, Inc., 2014.

[53] Zexi Song and Zhiqiang Tan. Hamiltonian Assisted Metropolis Sampling. *arXiv e-prints*, page arXiv:2005.08159, May 2020.

[54] Denis Talay. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields*, 8(2):163–198, 2002. MR1924934

[55] Grigori N. MilsteinMichael V. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Springer, Berlin, Heidelberg, 2004.

[56] Mark E. Tuckerman. *Statistical mechanics theory and molecular simulation*. Oxford University Press, 2010. MR2723222

[57] Cédric Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009. MR2562709

[58] Cédric Villani. *Optimal transport, old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. MR2459454

[59] Tim Zajic. Non-asymptotic error bounds for scaled underdamped Langevin MCMC. *arXiv e-prints*, page arXiv:1912.03154, December 2019.