Unadjusted Langevin algorithm with multiplicative noise: Total variation and Wasserstein bounds

Gilles Pages, Fabien Panloup

Unadjusted Langevin algorithm with multiplicative noise: Total variation and Wasserstein bounds

Gilles Pagès * and Fabien Panloup †

September 14, 2022

Abstract

In this paper, we focus on non-asymptotic bounds related to the Euler scheme of an ergodic diffusion with a possibly multiplicative diffusion term (non-constant diffusion coefficient). More precisely, the objective of this paper is to control the distance of the standard Euler scheme with decreasing step (usually called Unadjusted Langevin Algorithm in the Monte Carlo literature) to the invariant distribution of such an ergodic diffusion. In an appropriate Lyapunov setting and under uniform ellipticity assumptions on the diffusion coefficient, we establish (or improve) such bounds for Total Variation and L^1-Wasserstein distances in both multiplicative and additive and frameworks. These bounds rely on weak error expansions using Stochastic Analysis adapted to decreasing step setting.

Mathematics Subject Classification: Primary 65C05-37M25-60F05-62L10 Secondary 65C40-93E3

Keywords: Unadjusted Langevin algorithm; Euler scheme with decreasing step; multiplicative noise; Malliavin calculus; weak error; ergodic diffusion; invariant distribution, total variation distance; L^1-Wasserstein distance.

1 Introduction

Let $(X_t)_{t\in[0,T]}$ be the unique strong solution to the stochastic differential equation (SDE)

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \tag{1.1}$$

starting at X_0 where W is a standard $\mathbb{R}^q$-valued standard Brownian motion, independent of X_0, both defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{M}(d, q, \mathbb{R})$ ($d \times q$-matrices with real entries) are Lipschitz continuous functions. The process $(X_t)_{t\geqslant 0}$ is a homogeneous Markov process, denoted $X^x = (X^x_t)_{t\geqslant 0}$ if $X_0 = x$, with transition semi-group $P_t(x, dy) = \mathbb{P}(X^x_t \in dy)$. We denote by $\mathbb{P}_\mu$ its distribution starting from $X_0 \sim \mu$ (and $\mathbb{P}_x$ when $\mu = \delta_x$). Let $\mathcal{L} = \mathcal{L}_X$ denote its infinitesimal generator, defined on twice differentiable functions $g : \mathbb{R}^d \to \mathbb{R}$ by

$$\mathcal{L}g = (b|\nabla g) + \frac{1}{2}\mathrm{Tr}\big(\sigma^* D^2 g\, \sigma\big),$$

where $(\,.|.\,)$ denotes the canonical inner product on $\mathbb{R}^d$, $D^2 g$ denotes the Hessian matrix of g and Tr denotes the Trace operator.

*Sorbonne Université, Laboratoire de Probabilités, Statistique et Modélisation, UMR 8001, case 158, 4, pl. Jussieu, F-75252 Paris Cedex 5, France. E-mail: gilles.pages@sorbonne-universite.fr

†LAREMA, Faculté des Sciences, 2 Boulevard Lavoisier, Université d'Angers, 49045 Angers, France. E-mail: fabien.panloup@univ-angers.fr

Let $(\gamma_n)_{n\geqslant 1}$ be a non-increasing sequence of positive steps. We consider the Euler scheme of the SDE with step $\gamma_n > 0$ starting from $\bar{X}_0 = X_0$ defined by

$$\bar{X}_{\Gamma_{n+1}} = \bar{X}_{\Gamma_n} + \gamma_{n+1} b(\bar{X}_{\Gamma_n}) + \sigma(\bar{X}_{\Gamma_n})(W_{\Gamma_{n+1}} - W_{\Gamma_n}), \quad n \geqslant 0. \tag{1.2}$$

where

$$\Gamma_0 = 0 \quad \text{and} \quad \Gamma_n = \gamma_1 + \cdots + \gamma_n.$$

with $(\gamma_n)_{n\geqslant 1}$ a sequence of varying time steps. We define the genuine (continuous time) Euler scheme by interpolation as follows: let $t \in [\Gamma_k, \Gamma_{k+1})$.

$$\bar{X}_t = \bar{X}_{\Gamma_k} + (t - \Gamma_k) b(\bar{X}_{\Gamma_k}) + \sigma(\bar{X}_{\Gamma_k})(W_t - W_{\Gamma_k}). \tag{1.3}$$

If we set $\underline{t} = \Gamma_k$ on the time interval $[\Gamma_k, \Gamma_{k+1})$, the genuine Euler scheme appears as an Itô process solution to the pseudo-SDE with frozen coefficients

$$d\bar{X}_t = b(\bar{X}_{\underline{t}})dt + \sigma(\bar{X}_{\underline{t}})dW_t. \tag{1.4}$$

It will be convenient in what follows to introduce

$$N(t) = \min\big\{k \geqslant 0 : \Gamma_{k+1} > t\big\} = \max\big\{k \geqslant 0 : \Gamma_k \leqslant t\big\}. \tag{1.5}$$

The Euler scheme is a discrete time non-homogeneous Markov process with transitions

$$\bar{P}_{\Gamma_n, \Gamma_{n+1}}(x, dy) = \bar{P}_{\gamma_{n+1}}(x, dy)$$

where the transition probability $\bar{P}_\gamma(x, dy)$ reads on Borel test functions

$$\bar{P}_\gamma g(x) = \mathbb{E}\, g\big(x + \gamma g(x) + \sqrt{\gamma}\sigma(x) Z\big), \qquad Z \sim \mathcal{N}(0; I_d). \tag{1.6}$$

We assume that the time step sequence $(\gamma_n)_{n\geqslant 1}$ satisfies Assumption (Γ) defined by:

$$(\Gamma): \qquad (\gamma_n)_{n\geqslant 1} \text{ non-increasing}, \quad \lim_n \gamma_n = 0 \qquad \text{and} \qquad \sum_{n\geqslant 1} \gamma_n = +\infty. \tag{1.7}$$

Then $\gamma_1 = \sup_{n\geqslant 1} \gamma_n$ and we will denote indifferently this quantity by $\|\boldsymbol{\gamma}\|$ or γ_1 depending on the context.

It is well-known that for a twice continuously differentiable function $V : \mathbb{R}^d \to \mathbb{R}_+$ such that $e^{-V} \in L^1_{\mathbb{R}_+}(\mathbb{R}^d, \lambda_d)$ (λ_d Lebesgue measure on $\mathbb{R}^d$), then for every $\sigma \in (0,1]$

$$\nu_\sigma(dx) = C_\sigma e^{-\frac{V(x)}{\sigma^2}} \lambda_d(dx) \quad \text{with} \quad C_\sigma = \Big(\int_{\mathbb{R}^d} e^{-\frac{V(x)}{\sigma^2}} \lambda_d(dx)\Big)^{-1}$$

is the unique invariant distribution of the Langevin (reversible) Brownian SDE

$$dX_t = -\nabla V(X_t)dt + \sqrt{2}\,\sigma dW_t \tag{1.8}$$

where $(W_t)_{t\geqslant 0}$ is d-dimensional standard Brownian motion.

A first application of this property is to devise an approximate simulation method of $\nu = \nu_1 = C_1^{-1} e^{-V} \cdot \lambda_d$ by introducing the above Euler scheme with decreasing step (1.2) with $b = -\nabla V$ and $\sigma(x) = \sqrt{2}$. Coupled with a Metropolis-Hasting speeding method, this simulation procedure is known as the Metropolis Adjusted Langevin algorithm whereas in absence of such an additional procedure it is known as the Unadjusted Langevin Algorithm (ULA) extensively investigated in the literature since the 1990's (see $e.g.$ [Pel96], [MP96]) and more recently in a series of papers, still in the additive setting, motivated by applications in machine learning (in particular in Bayesian or PAC-Bayesian statistics). Among others, we refer to [DM17, DM19, Dal17, MFWB19] and to the references therein.

A second application is to directly consider, σ being a fixed real number (or possibly a matrix of $\mathbb{M}(d,d,\mathbb{R})$), the Euler scheme

$$\bar{X}^{\sigma}_{\Gamma_{n+1}} = \bar{X}^{\sigma}_{\Gamma_n} - \gamma_{n+1}\nabla V(\bar{X}^{\sigma}_{\Gamma_n}) + \sqrt{2}\,\sigma\sqrt{\gamma_{n+1}}Z_{n+1} \tag{1.9}$$

where $(Z_k)_{k\geqslant 1}$ is an $\mathcal{N}(0,I_d)$-distributed i.i.d. sequence. It appears as a perturbation by a Gaussian white noise of the gradient descent

$$x_{n+1} = x_n - \gamma_{n+1}\nabla V(x_n)$$

aiming at minimizing the potential V. Then, using the notation $[Y]$ to denote the distribution of a random vector Y,

$$[\bar{X}^{\sigma}_{\Gamma_n}] \stackrel{TV}{\longrightarrow} \nu_\sigma \quad \text{and} \quad \nu_\sigma \stackrel{weakly}{\longrightarrow} \delta_{x*} \ \text{ as } \sigma \to 0$$

if $\mathrm{argmin}_{\mathbb{R}^d} V = \{x^*\}$ (or ν_σ is asymptotically supported by $\mathrm{argmin}_{\mathbb{R}^d} V$ when simply finite). So simulating (1.9) on the long run provides sharper and sharper information on the localization of $\mathrm{argmin}_{\mathbb{R}^d} V$. In fact making $\sigma = \sigma_n$ slowly vary in a decreasing way to 0 at rate $(\log n)^{-1/2}$ makes up a simulated annealing version of the above perturbed stochastic gradient procedure. This stochastic optimization procedure has been investigated in-depth in [GM93] with, as a main result, the convergence in probability of $\bar{X}^{\sigma_n}_{\Gamma_n}$ toward the (assumed) unique minimum x^* of V under various assumptions on the step γ_n and the invertibility of the Hessian of V at x^*.

For much more general multidimensional diffusions, say Brownian driven here for convenience, of the form (1.1) with infinitesimal generator $\mathcal{L}$ satisfying an appropriate mean-reverting drift (typically $\mathcal{L}V \leqslant \beta - \alpha V^a$, $a \in (0,1]$ for some Lyapunov function V), it is a natural problem of numerical probability to have numerical access to its invariant distribution ν (when unique). Taking full advantage of ergodicity, this can be achieved by introducing the weighted empirical measure

$$\bar{\nu}_n(\omega, d\xi) = \frac{1}{\Gamma_n}\sum_{k=1}^{n}\gamma_k\delta_{\bar{X}_{\Gamma_{k-1}}(\omega)}(d\xi), \quad n \geqslant 1, \tag{1.10}$$

where $(\bar{X}_{\Gamma_k})_{k\geqslant 0}$ is given by (1.2) (and the Brownian increments are simulated by a $\mathbb{R}^q$-valued white noise $(Z_k)_{k\geqslant 1}$ with $W_{\Gamma_{k+1}} - W_{\Gamma_k} = \sqrt{\gamma_{k+1}}Z_k$, $k \geqslant 1$. $A.s.$ weak convergence of $\bar{\nu}_n(d\xi)$ to ν, its convergence rate as well as deviation inequalities depending on the rate of decay of the time step γ_n have been extensively investigated in a series of papers in various settings, including the case of jump diffusion driven by Lévy processes (see [LP02], [LP03], [Pan08b], [Pan08a], [Lem05], [Lem07], [HMP20], etc). One specificity of interest of this method based on the simulation of the above weighted empirical measures $\bar{\nu}_n(\omega, d\xi)$ (see (1.10)) for applications is that no ellipticity is required to establish most of the main results. This turns out to be crucial for Hamiltonian systems or more generally for mean-reverting SDEs with more or less degenerate diffusion coefficients.

However it is a quite natural question to tackle the total variation (TV) and L^1-Wasserstein (rates of) convergence of $[\bar{X}_{\Gamma_n}]$ toward the (necessarily) unique invariant distribution ν when σ is not constant but uniformly elliptic. In particular, one aim of this paper is to check whether or not the VT and L^1-Wasserstein (or Monge-Kantorovich) rates of convergence remain unchanged in such a more general setting (in terms of $(\gamma_n)_{n\geqslant 1}$). Moreover, considering such diffusions with non constant σ will deeply impact the methods of proof. When σ is constant, the continuous-time Euler scheme $(\bar{X}^x_t)_{t\geqslant 0}$ and the diffusion $(X^x_t)_{t\geqslant 0}$ have the same diffusion component σW. Girsanov's theorem then implies that their distributions are equivalent and provides an explicit expression of the density of the distribution of $\bar{X}^x_\gamma$ with respect to the one of X^x_γ. This is the key to establish the estimates of $d_{TV}([X^x_{\gamma_n}], [\bar{X}^x_{\gamma_n}])$ through Pinsker's inequality (see [DM17] or Proposition 4.1 and Theorem 2.3 of our paper). In the multiplicative case, such an approach no longer works and will be replaced here by stochastic analysis arguments (see below for details).

Such investigations also have applied motivations since in the blossoming literature produced by the data science community to analyze and improve the performances of stochastic gradient procedures, non-constant matrix valued diffusion coefficients $\sigma(x)$ are introduced in such a way (see [MCF15a] and the references therein with in view Hamiltonian Monte Carlo, see [LCCC15a] among others) that the invariant distribution is unchanged but the exploration of the state space becomes non-isotropic, depending on the position of the algorithm or the value of potential function to be minimized with the hope to speed up its preliminary convergence phase. Note that a script of the [LCCC15a] version of Unadjusted Langevin Algorithm is made available in the API TensorFlowProbability ([1]).

As mentioned above we mainly focus on TV or L^1-Wasserstein bounds in the so-called multiplicative setting i.e. when the diffusion coefficient is state dependent, which is new in this field to our best knowledge. However we also show how to refine our methods of proof (see below) in order to derive improved rates in the additive setting (when σ is constant). These results improve those obtained $e.g.$ in [DM17] or in [Dal17] in terms of $(\gamma_n)_{n\geqslant 1}$ and seem quite consistent with more recent works (by very different methods) like [MFWB19] or [DM19]. In fact, we slightly improve the results of these papers by killing some logarithmic terms with the help of Malliavin calculus (see Remark 2.4 for details). However, compared with these papers, we do not tackle the problem related to the dependence of the bounds with respect to the dimension d, which would lead to very heavy technicalities, especially in the multiplicative setting which is the main goal of this paper.

Although this problem seems not to have been already tackled in the multiplicative case, we can yet connect our work with several other papers where non-asymptotic bounds between the Euler scheme and the invariant distribution have been established: in the recent paper [CDO21], the authors provide uniform in time bounds for the weak error (which in turn may be used to derive some bounds for the error with respect of the invariant distribution). Nevertheless, this paper only considers smooth functions which is clearly not adapted to TV or 1-Wasserstein bounds. We can also refer to [DMS20] where, with the help of a new *Backward Itô-Ventzell formula*, the authors interpolate the diffusion and its continuous-time Euler discretization to derive nice L^p-bounds under some pathwise contraction assumptions (close to Assumption $(\mathbf{C}_\alpha)$ below). These L^p-bounds lead in turn to 1-Wasserstein bounds but it is not clear that they may produce TV-bounds in an optimal way. The interesting fact is that our so-called *domino decomposition* described below can be seen as a discrete weak version of the pathwise interpolation proposed in [DMS20]. In particular, our approach is different from that in [DMS20] since we rely on the contraction of the semi-group of the diffusion instead of the pathwise assumptions required everywhere there (whereas contraction of the semi-group may hold in settings where pathwise contraction holds only outside a compact set, see $e.g.$ Corollary 2.5).

Now, let us be more specific about our results and methods. We start from some assumptions on the diffusion (1.1) itself: we mainly assume a classical Lyapunov mean-reverting assumption (denoted by $(\mathbf{S})$), an exponential contraction property (in 1-Wasserstein distance) of the distributions $[X_t^x]$ and $[X_t^y]$[2] and uniform ellipticity and boundedness assumptions on the diffusion coefficient σ (denoted by $(\mathcal{E}\ell)_{\underline{\sigma}_0^2}$, see Section 2.1 for details).

In the multiplicative setting, under these general assumptions (including uniform ellipticity), our main result (see Theorem 2.2) establishes that the Total Variation (TV) distance between the distribution of X_{Γ_n} and the invariant distribution ν (denoted by $\big\|[\bar{X}^x_{\Gamma_n}]-\nu\big\|_{TV}$, see below for notations) converges to 0 at rate

$$O(\gamma_n^{1-\varepsilon}), \text{ for every } \varepsilon\in(0,1), \text{ for the } TV\text{-distance (if } b \text{ and } \sigma \text{ are } C^6),$$

[1] see www.tensorflow.org/probability/api_docs/python/tfp/optimizer/StochasticGradientLangevinDynamics

[2] See Assumption $(\mathbf{H}_\mathfrak{d})$ and Remark 2.2 for details.

whereas its 1-Wasserstein counterpart (denoted by $\mathcal{W}_1([\bar{X}^x_{\Gamma_n}], \nu)$) converges to 0 at rate

$$O\big(\gamma_n \log(1/\gamma_n)\big) \text{ for the } \mathcal{W}_1\text{-distance (if } b \text{ and } \sigma \text{ are } C^4).$$

In the <u>additive case</u> (see Theorem 2.3 e.g. if b is C^3), we prove that the distance between the distribution of X_{Γ_n} and ν is :

$$O(\gamma_n) \text{ for both the } TV\text{-distance and the } \mathcal{W}_1\text{-distance.}$$

As mentioned before, these results are established under general contraction assumptions made on the dynamics of the underlying diffusion. Thus, in order to be more concrete, we recall and provide in Section 2.3 practical criterions which imply exponential contraction (and thus exponential convergence rate). Typically, such an assumption holds true if the drift coefficient is *strongly contracting* outside a compact set (see Corollary 2.5).

Our method of proof mostly relies on Numerical Probability and Stochastic Analysis techniques developed for diffusion processes since the 1980's, adapted to both decreasing step and long time behaviour. Namely, we carry out an in-depth analysis of the weak error of the one step Euler scheme (bounded) Borel and smooth functions, with a a special case in the latter case to the dependence of the resulting rate with respect to the regularity of the function. Then we rely on the regularizing properties of the semi-group of the underlying diffusion through an extensive use of Bismut-Elworthy-Li (<u>BEL</u>) identities and their resulting upper-bounds (see [Bis84, EL94]). To deal with (non-smooth) bounded Borel functions we call upon the Malliavin Calculus machinery adapted to the decreasing step setting relying, among others, on recent papers by Bally, Caramellino and Poly (see [BC19, BCP20]) which make these methods more accessible.

Our global strategy of proof (initiated by [TT90, BT96]) relies either on a partial (for <u>TV</u>-distance in the multiplicative case) or a full domino decomposition of the error to be controlled, formally reading in our long run behaviour as follows (here for the full one)

$$\begin{aligned}|\mathbb{E}\, f(\bar{X}^x_{\Gamma_n}) - \mathbb{E} f(X^x_{\Gamma_n})\big] &= \big|\bar{P}_{\gamma_1} \circ \cdots \bar{P}_{\gamma_n} f(x) - P_{\Gamma_n} f(x)\big| \\ &\leqslant \sum_{k=1}^{n} \Big|\bar{P}_{\gamma_1} \circ \cdots \bar{P}_{\gamma_{k-1}}\big(\bar{P}_{\gamma_k} - P_{\gamma_k}\big) P_{\Gamma_n - \Gamma_k} f(x)\Big|.\end{aligned}$$

Depending on the nature of the distance and σ we will subdivide the above sum in two or three partial sums and analyze them using the various tools briefly described above. The paper is organized as follows. Section 2 is devoted to the assumptions, the main results and the applications. In Section 3 we first provide some background on our main tools, especially on Stochastic Analysis (<u>BEL</u>, weak error by Malliavin calculus, having in mind that most background and proof are postponed in Appendices A and B and, in a second part of the section, we analyze in-depth the weak error of the one-step Euler scheme with in mind the strong specificity of our long run problem. In Section 4, we provide proofs for our main convergence results.

NOTATIONS. – The canonical Euclidean norm of a vector $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ is denoted by $|x| = (x_1^2 + \cdots + x_d^2)^{1/2}$.
– $\mathbb{N} = \{0, 1, \ldots\}$ and $\mathbb{N}^* = \{1, 2, 3, \ldots\}$.
– $\|A\|_F = [\mathrm{Tr}\,(AA^*)]^{1/2}$ denotes the Fröbenius (or Hilbert-Schmidt) norm of a matrix $A \in \mathbb{M}(d, q, \mathbb{R})$ where A^* stands for the transpose of A^* and Tr denotes the trace operator of a square matrix.
– $\mathcal{S}(d, \mathbb{R})$ denotes the set of symmetric $d \times d$ square matrices and $\mathcal{S}^+(d, \mathbb{R})$ the subset of non-negative symmetric matrices.
– $\|a\| = \sup_{n \geqslant 1} |a_n|$ denotes the sup-norm of a sequence $(a_n)_{n \geqslant 1}$.
– For $f : \mathbb{R}^d \to \mathbb{R}$, $[f]_{\mathrm{Lip}} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}$.
– For a transition $Q(x, dy)$ we define $[Q]_{\mathrm{Lip}} = \sup_{f,\, [f]_{\mathrm{Lip}} \leqslant 1} [Qf]_{\mathrm{Lip}}$.

– $[X]$ denotes the distribution of the random vector X.
– $a_n \asymp b_n$ means that there are positive real constants $c_1, c_2 > 0$ such that $c_1\, a_n \leqslant b_n \leqslant c_2\, a_n$.
– For every $x, y \in \mathbb{R}^d$, $(x, y) = \big\{ux + (1-u)y,\ u \in (0,1)\big\}$. One defines likewise $[x, y]$, etc.
– The space of probability distributions on $(\mathbb{R}^d, \mathcal{B}or(\mathbb{R}^d))$, endowed with the topology of weak convergence is denoted by $\mathcal{P}(\mathbb{R}^d)$.
– $\mathcal{W}_p(\mu, \mu') = \inf\Big\{\big(\int |x-y|^p \pi(dx, dy)\big)^{1/p},\ \pi \in \mathcal{P}_{\mu,\nu}(\mathbb{R}^d)\Big\}$ denotes the L^p-Wasserstein distance between the probability distributions μ and μ' where $\mathcal{P}_{\mu,\nu}(\mathbb{R}^d)$ stands for the set of probability distributions on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}or(\mathbb{R}^d)^{\otimes 2}$ with respective marginals μ and ν.
– $\|\mu\|_{TV} = \sup\big\{\int f d\mu,\ f : \mathbb{R}^d \to \mathbb{R}$, Borel, $\|f\|_{\sup} \leqslant 1\big\}$ where μ denotes a signed measure on $(\mathbb{R}^d, \mathcal{B}or(\mathbb{R}^d))$ and d_{TV} denotes the related distance: $d_{TV}(\mu, \nu) = \|\mu - \nu\|_{TV}$.

2 Main Results

2.1 Assumptions

In whole the paper, we assume that b and σ are Lipschitz continuous and satisfy the strong mean-reverting assumption

$(\mathbf{S})$: There exists a positive $\mathcal{C}^2$-function $V : \mathbb{R}^d \to (0, +\infty)$ such that

$$\lim_{|x|\to+\infty} V(x) = +\infty, \quad |\nabla V|^2 \leqslant CV \quad \text{and} \quad \sup_{x\in\mathbb{R}^d} \|D^2 V(x)\|_F < +\infty \tag{2.11}$$

(Frobenius norm) and there exist some real constants $C_b > 0$, $\alpha > 0$ and $\beta \geqslant 0$ such that:

(i) $|b|^2 \leqslant C_b V$ and σ is bounded (e.g. in Frobenius norm), *(ii)* $\big(\nabla V | b\big) \leqslant \beta - \alpha V$

Remark 2.1. • Note that $(\mathbf{S})$ implies that V attains a minimum value $\underline{v} > 0$ (possibly at several points in $\mathbb{R}^d$).
• Note that since σ is bounded, (ii) is equivalent to the existence of $\alpha > 0$ and $\beta \geqslant 0$ such that

$$\mathcal{L}V \leqslant \beta - \alpha V.$$

• Let us also remark that (2.11) implies that V is a subquadratic function, *i.e.* there exists a constant $C > 0$ such that $V \leqslant C(1 + |\,.\,|^2)$.

Under $(\mathbf{S})$, it is classical background (see e.g. [EK86, Theorem 9.3 and Lemma 9.7 with $\varphi = V$ and $\psi = \mathcal{L}V$] that the diffusion $(X_t)_{t\geqslant 0}$ (in fact its semi-group $(P_t)_{t\geqslant 0}$) has at least one invariant distribution ν i.e. such that $\nu P_t = \nu$, $t \geqslant 0$. Furthermore, Assumption $(\mathbf{S})$ implies *stability* of the diffusion and of its discretization scheme by involving long-time bounds on polynomial (and exponential) moments of $V(X_t)$ and $V(\bar{X}_{\Gamma_n})$. Such properties are recalled in Proposition A.1.

In all the main results of the paper, we will also assume that the diffusion coefficient σ satisfies the following *uniform ellipticity* assumption:

$$(\mathcal{E}\ell)_{\underline{\sigma}_0^2} \;\equiv\; \exists\, \underline{\sigma}_0 > 0 \;\text{ such that }\; \forall\, x \in \mathbb{R}^d, \qquad \sigma\sigma^*(x) \geqslant \underline{\sigma}_0^2 I_d \quad \text{in} \quad \mathcal{S}^+(d, \mathbb{R}). \tag{2.12}$$

This uniform ellipticity assumption implies that, when existing, the invariant distribution is unique (see e.g. [Pag01] among others).

Finally, we suppose that the semi-group $(P_t)_{t\geqslant 0}$ of the diffusion satisfies a contraction property at exponential rate in for a given distance $\mathfrak{d}$ on $\mathcal{P}(\mathbb{R}^d)$, namely

$(\mathbf{H}_{\mathfrak{d}})$: There exist $t_0 > 0$ and positive constants c and ρ such that for every $t \geqslant t_0$,

$$\forall\, x,\, y \in \mathbb{R}^d, \quad \mathfrak{d}([X_t^x], [X_t^y]) \leqslant c|x-y|e^{-\rho t}.$$

In the sequel we will use $(\mathbf{H}_{\mathcal{W}_1})$ and $(\mathbf{H_{TV}})$, *i.e.* the conditions related to $\mathfrak{d} = \mathcal{W}_1$ (1-Wasserstein) and to $\mathfrak{d} = d_{TV}$ (Total variation) respectively. Note that owing to the Monge-Kantorovich representation of $\mathcal{W}_1$, see e.g.[Vil09], (*resp.* the definition of d_{TV}), the condition $(\mathbf{H}_{\mathcal{W}_1})$ (*resp.* $(\mathbf{H_{TV}})$) also reads on Lipschitz continuous functions $f : \mathbb{R}^d \to \mathbb{R}$ (*resp.* on bounded Borel-measurable functions) $f : \mathbb{R}^d \to \mathbb{R}$

$$\forall\, t \geqslant t_0, \qquad [P_t f]_{\text{Lip}} \leqslant c e^{-\rho t}[f]_{\text{Lip}} \quad (\textit{resp.} \quad [P_t f]_{\text{Lip}} \leqslant c e^{-\rho t}[f]_{\infty}).$$

In fact, only $(\mathbf{H}_{\mathcal{W}_1})$ appears in the next theorems. Actually, by the regularizing effect of the elliptic semi-group, we have the following result (whose proof is postponed to Appendix C.2):

Proposition 2.1. *Suppose that b and σ are C^1 with bounded partial derivatives and that $(\mathcal{E}\ell)_{\underline{\sigma}_0^2}$ is in force. If $(\mathbf{H}_{\mathcal{W}_1})$ holds with some positive ρ and t_0, then $(\mathbf{H_{TV}})$ holds with the same ρ and t_0.*

Remark 2.2. • If b and σ are both Lipschitz continuous and $(\mathbf{H}_{\mathcal{W}_1})$ holds true, then it holds true from the origin, i.e. for $t_0 = 0$, with the same ρ, up to a change of the real constant c. Actually, if $f : \mathbb{R}^d \to \mathbb{R}$ is Lipschitz continuous, then, for every $t \in [0, t_0]$ and every $x,\, y \in \mathbb{R}^d$,

$$\big|\mathbb{E} f(X_t^x) - f(X_t^y)\big| \leqslant [f]_{\text{Lip}} \mathbb{E}|X_t^x - X_t^y| \leqslant C_{t_0, [b]_{\text{Lip}}, [\sigma]_{\text{Lip}}} [f]_{\text{Lip}} |x-y|$$

by standard arguments on the flow of the SDE (see e.g. [Pag18, Theorem 7.10]). One concludes by the Kantorovich-Rubinstein representation of $\mathcal{W}_1$. Note that for $(\mathbf{H_{TV}})$, the property does certainly not extend to $t_0 = 0$ since $\|\delta_x - \delta_y\|_{TV} = 2$ for any $x \neq y$.

• In Assumption $(\mathbf{H}_{\mathfrak{d}})$ (and especially in Assumption $(\mathbf{H}_{\mathcal{W}_1})$), we choose to base our main results on a contraction property of the semi-group, in order to avoid to mix up discretization problems and ergodic properties of the diffusion. However, we provide in Section 2.3 a large class of examples where this assumption is fulfilled: in the uniformly convex/dissipative setting as established in Corollary 2.4 later on but also, when b is only strongly contracting outside a compact set (see Corollary 2.5). When σ is constant, one can refer to [LW16] or [EGZ19] for bounds in Wasserstein distance for diffusions. For background on ergodicity properties of diffusions, we also refer to [BGL14, DKZ12] or to [CCDO21] for the degenerate setting.

2.2 Main results

To a non-increasing sequence of positive steps denoted $\boldsymbol{\gamma} = (\gamma_n)_{n \geqslant 1}$ we associate the index

$$\varpi := \overline{\lim_{n}} \frac{\gamma_n - \gamma_{n+1}}{\gamma_{n+1}^2} \in [0, +\infty].$$

This index is finite if and only if the convergence of γ_n to 0 is not too fast. To be more precise, if $\gamma_n = \frac{\gamma_1}{n^a}$ $(a > 0)$, $\varpi = 0$ if $0 < a < 1$ and $\varpi = \frac{1}{\gamma_1}$ if $a = 1$ and $\varpi = +\infty$ if $a > 1$. We are now in position to state our main result.

Theorem 2.2. *Assume $(\mathcal{E}\ell)_{\underline{\sigma}_0^2}$ and $(\mathbf{S})$ and $(\mathbf{H}_{\mathcal{W}_1})$ with $\rho > \varpi$. Let ν be the (unique) invariant distribution of $(X_t)_{t \geqslant 0}$. Suppose that the step sequence $(\gamma_n)_{n \geqslant 1}$ satisfies (Γ), that $\varpi \in [0, +\infty)$ and $\int |\xi| \nu(d\xi) < +\infty$.*

(a) If b and σ are $\mathcal{C}^4$ with bounded derivatives, then

$$\forall\, n \geqslant 1, \quad \mathcal{W}_1([\bar{X}_{\Gamma_n}^x], \nu) \leqslant C_{b,\sigma,\gamma,V} \cdot \gamma_n \big|\log(\gamma_n)\big| \vartheta(x)$$

where $C_{b,\sigma,\gamma}$ is a constant depending only on b, σ, γ and $\vartheta(x) = (|x|+1) \vee V^2(x)$.

(b) If b and σ are C^6 with bounded existing partial derivatives and if $\liminf_{|x|\to+\infty} V(x)/|x|^r > 0$ for some $r \in (0,2]$ (resp. $\liminf_{|x|\to+\infty} V(x)/\log(1+|x|) = +\infty$), then, for every small enough $\varepsilon > 0$, there exists a real constant $C_\varepsilon = C_{\varepsilon,b,\sigma,\gamma,V} > 0$ such that

$$\forall n \geqslant 1, \quad \big\|[\bar{X}^x_{\Gamma_n}] - \nu\big\|_{TV} \leqslant C_\varepsilon \cdot \gamma_n^{1-\varepsilon}\vartheta(x)$$

where $\vartheta(x) = V^{8/r}(x) \in L^1(\nu)$ (resp. $\vartheta(x) = e^{\lambda_0 V(x)} \in L^1(\nu)$ for some $\lambda_0 \in (0, \lambda_{\sup}/2)$ where $\lambda_{\sup}$ is defined in Proposition A.1(b)).

Remark 2.3. • The parameter ρ does not appear in the above constants since ρ can be in turn considered as a function of b and σ. But the constant clearly depends on it. For the sake of readability, we will sometimes omit the dependency in the next results. The main point is that these constants do not depend on x.

• The proofs of the above convergence rates certainly rely on ergodic arguments but also on refined bounds on the *one-step weak error* between the Euler scheme and the diffusion for non-smooth functions. In particular, one important tool for the total variation bound is a one-step control of the weak error for bounded Borel functions when the initial condition is an "almost" non-degenerated (in a Malliavin sense) random variable ([3]). More precisely, this random initial condition is precisely an Euler scheme at a given positive (non-small) time and $(\mathcal{E}\ell)_{\underline{\sigma}_0^2}$ guarantees that the related Malliavin matrix is non-degenerated with high probability but not almost surely (since the tangent process of the continuous-time Euler scheme does not almost surely map into $GL_d(\mathbb{R})$). This almost but not everywhere non-degeneracy induces a cost which mainly explains that the bound in Theorem 2.2(b) is proportional to $\gamma^{1-\varepsilon}$ and not to $\gamma|\log(\gamma)|$, as in the above claim (a). However, one could wonder about the optimality of this bound and on the opportunity to get a bound in γ. Such a result could perhaps follow from a sharper control of the probability of non-degeneracy of the Euler scheme but this appears as a non trivial task, not achieved in [BCP20]. An alternative (used for instance in [Guy06]) is to base the proof on parametrix-type expansions of the error between the density of the Euler scheme and that of the diffusion obtained in [KM02]. But relying on such an alternative would require to adapt their arguments to the decreasing step setting and to prove that the coefficients of the resulting expansion do not depend on the considered step sequence ([4]). Solving this problem would yield a TV bound in γ_n as can be checked from proof of the theorem.

Let us now turn to the so-called additive case, $\sigma(x) = \sigma$.

Theorem 2.3 (Additive case)**.** *Assume that b is C^3 with bounded existing partial derivatives and $\sigma(x) \equiv \sigma$ with $\sigma\sigma^*$ is definite positive. Assume $(\mathbf{S})$ holds and $\varpi \in (0,+\infty)$. If $(\mathbf{H}_{\mathcal{W}_1})$ holds with $\rho > \varpi$ and $\int |x|\nu(dx) < +\infty$, then there exists a real constant $C = C_{b,\sigma,\gamma,V} > 0$ such that for all $n \geqslant 1$,*

$$\mathcal{W}_1([\bar{X}^x_{\Gamma_n}], \nu) \leqslant C \cdot \gamma_n\, \vartheta(x) \quad \textit{and} \quad \big\|[\bar{X}^x_{\Gamma_n}] - \nu\big\|_{TV} \leqslant C \cdot \gamma_n \big|\log(\gamma_n)\big|\, \vartheta(x)$$

with $\vartheta(x) = (1+|x|) \vee V^a(x)$ with $a = 2$ for $\|\cdot\|_{TV}$ and $a = 3/2$ for $\mathcal{W}_1$.

If, furthermore, $\liminf_{|x|\to+\infty} V(x)/|x|^r > 0$ for some $r \in (0,2]$ (resp. $\liminf_{|x|\to+\infty} V(x)/\log(1+|x|) = +\infty$), then there exists a real constant $C = C_{b,\sigma,\gamma,V}$ such that for all $n \geqslant 1$,

$$\big\|[\bar{X}^x_{\Gamma_n}] - \nu\big\|_{TV} \leqslant C \cdot \gamma_n\, \vartheta(x)$$

[3] This result is established in Theorem 3.7. Among other arguments, the related proof relies on recent Malliavin bounds obtained in [BCP20].

[4] More precisely, the main result of [KM02] establishes existence of error expansions reading as polynomials (null at 0) of the step but, surprisingly, with coefficients still "slightly" varying with the step. Then the authors claim that such a dependence can be canceled by further (non-detailed) arguments.

where $\vartheta(x) = V^{2\vee\frac{1}{r}}(x) \in L^1(\nu)$ *(resp.* $\vartheta(x) = e^{\lambda_0 V(x)} \in L^1(\nu)$ *for some* $\lambda_0 \in (0, \lambda_{\sup}/2)$*).*

The Wasserstein bound is thus proportional to γ_n whereas the one in Total Variation is proportional to $\gamma_n \log(1/\gamma_n)$ or to γ_n under a very slight additional assumption. Note that in our proof, passing from $\gamma_n \log(1/\gamma_n)$ to γ_n, without adding smoothness assumptions on b, results from a sharp combination of Bismut-Elworthy-Li formula and Malliavin calculus (see end of Subsection 4.3). This bound in $O(\gamma_n)$ is optimal (in Wasserstein or in *TV*-distance). Actually, explicit computations can be done for the Ornstein-Uhlenbeck process which lead to lower-bounds proportional to γ_n. To be more precise, let us consider the α-confluent centered Ornstein-Uhlenbeck process defined by

$$dX_t = -\alpha X_t dt + \sigma dW_t, \quad X_0 = 0,$$

where α, $\sigma > 0$. Then, there exists $c_\alpha > 0$ such that, for large enough n (see Section 4.6 for a proof),

$$\left\|[\bar{X}_{\Gamma_n}] - \nu\right\|_{TV} \geqslant \frac{1}{200} \min\left(1, \left|1 - \frac{\sigma_n^2}{\sigma^2/(2\alpha)}\right|\right) \geqslant c_\alpha \gamma_n.$$

Remark 2.4. Although, this paper is mainly concerned with the multiplicative setting, it is interesting to compare our additive result in Theorem 2.2 with the literature. First, note that such bounds have been extensively investigated in the literature. For instance, one retrieves TV-bounds in a somewhat hidden way in works about recursive simulated annealing (see [GM91], [MP96]). But more recently, many papers tackled this question, in decreasing or constant step settings with a focus on the dependency of the constants in the dimension. Here, we consider the first setting and the dependency in γ_n. From this point of view, our *TV*-bounds improve those obtained in [DM17] or [Dal17] (in $O(\sqrt{\gamma_n})$) and are mostly comparable to the more recent [DM19, Theorem 14] or [MFWB19], up to logarithmic terms. More precisely, these two papers respectively lead (in a constant step setting) to bounds in $O(\gamma \log \gamma)$ or $O(\gamma\sqrt{|\log\gamma|})$, whereas in our work, we obtain a rate in $O(\gamma_n)$ with the help of a refinement of the proof based on Malliavin calculus techniques.

2.3 Applications

The assumptions of the above theorems hold under *contraction* assumptions of the semi-group of the diffusion. Here, we provide some standard settings where the result applies (proofs are postponed to Sections 4.4 and 4.5 respectively).

$\triangleright$ **Uniformly dissipative (or convex) setting**. A first classical assumption which ensures contraction properties is the following:

$$(\mathbf{C}_\alpha) \;\equiv\; \forall x, y \in \mathbb{R}^d, \quad \big(b(x) - b(y)\,|\,x - y\big) + \tfrac{1}{2}\|\sigma(x) - \sigma(y)\|_F^2 \leqslant -\alpha|x - y|^2. \tag{2.13}$$

In particular, if $b = -\nabla U$ where $U : \mathbb{R}^d \to \mathbb{R}$ is $\mathcal{C}^2$ and σ is constant, this assumption is satisfied as soon as $D^2U \geqslant \alpha I_d$ where $\alpha > 0$ i.e. U is α-convex. This leads to the following result which appears as a corollary of the above theorems (its proof is postponed in to Subsection 4.4).

Corollary 2.4. *Assume* $(\mathcal{E}\ell)_{\underline{\sigma}_0^2}$ *and* $(\mathbf{S})$*. Assume* $(\mathbf{C}_\alpha)$*. Then,* $(\mathbf{H}_{\mathcal{W}_1})$ *is satisfied with* $\rho = \alpha$*. As a consequence, the conclusions of Theorem 2.2 (resp. Theorem 2.3 when* σ *is constant) hold true.*

Remark 2.5. When σ is constant and $(\mathbf{C}_\alpha)$ holds true, a 2-Wasserstein bound can be directly deduced by some discrete Gronwall like arguments based on recursive estimates of $\mathbb{E}\,|X_{\Gamma_n} - \bar{X}_{\Gamma_n}|^2$ (with X_{Γ_n} and $\bar{X}_{\Gamma_n}$ built from the same Brownian motion) combined with expansions of the one step error similar to those which lead to the control of the L^p-error in finite horizon for the Milstein scheme (which coincides with the Euler-Maruyama scheme when σ is constant), see $e.g.$ [Pag18, Corollary 7.2].

$\triangleright$ **Non uniformly dissipative settings**. In fact, our main results are adapted to some settings where the contraction holds only outside a compact set. The following result is a fairly simple consequence of [Wan20] and of our main theorems (see Section 4.5 for a detailed proof).

Corollary 2.5. *Assume $(\mathcal{E}\ell)_{\underline{\sigma}_0^2}$ and $(\mathbf{S})$ (in particular σ is bounded). Assume that b is Lipschitz continuous and that some positive α and $R > 0$ exist such that for all*

$$\forall\, x, y \in B(0, R)^c, \quad \big(b(x) - b(y)\,|\, x - y\big) \leqslant -\alpha |x - y|^2.$$

Then, $(\mathbf{H}_{\mathcal{W}_1})$ is satisfied. Hence, the conclusions of Theorem 2.2 (resp. Theorem 2.3 when σ is constant) hold true.

Remark 2.6. It is clear that Assumption $(\mathbf{C}_\alpha)$ implies that $\big(b(x) - b(y)\,|\, x - y\big) \leqslant -\alpha|x - y|^2$ for all x, y, hence outside any compact set. Thus Corollary 2.5 contains Corollary 2.4. However, the first result emphasizes that the exponent ρ in Assumption $(\mathbf{H}_{\mathcal{W}_1})$ can be made explicit in the uniformly dissipative case, opening the way to more precise error bounds.

When σ is constant, one can also deduce $(\mathbf{H}_{\mathcal{W}_1})$ in the non-uniformly dissipative case from [LW16] or [EGZ19].

2.4 Langevin Monte Carlo and multiplicative (multi-dimensional) SDEs

A significant portion of the paper is devoted to the multiplicative case (in particular, a significant part of the proof of Theorem 2.2). However, in applications and in particular in the Langevin Monte-Carlo method (whose principle is recalled below), diffusions with constant σ are more frequently used. Below, we show that using multiplicative SDEs may be of interest for applications to the Langevin Monte-Carlo method. Let us recall that for a potential $V : \mathbb{R}^d \to \mathbb{R}$ and its related Gibbs distribution

$$\nu_V(dx) = C_V e^{-V(x)} \cdot \lambda_d(dx) \quad \text{with } C_V^{-1} = \int e^{-V(x)} \cdot \lambda_d(dx),$$

the Langevin Monte-Carlo usually refers to the numerical approximation of ν_V, viewed as the invariant distribution of the additive SDE

$$dX_t = -\sigma^2 \nabla V(X_t)dt + \sqrt{2}\sigma dW_t, \tag{2.14}$$

where σ is a positive constant (usually equal to 1). In fact, it is possible to exhibit a large class of multiplicative diffusions which also share with the same invariant distribution ν_V as shown in Proposition 2.6 below.

Proposition 2.6. *Let $V : \mathbb{R}^d \to \mathbb{R}_+$ be a C^2 function such that ∇V is Lipschitz continuous and $e^{-V} \in L^1(\lambda_d)$. Let $\sigma : \mathbb{R}^d \to \mathbb{M}(d, q, \mathbb{R})$ be a C^1, bounded matrix valued field with bounded partial derivatives and satisfying $(\mathcal{E}\ell)_{\underline{\sigma}_0^2}$. Let $(X_t^x)_{t\geqslant 0}$ be solution to the SDE*

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t,\ X_0 = x, \tag{2.15}$$

($W = (W_t)_{t\geqslant 0}$ standard Brownian motion defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$) with drift

$$b = -\tfrac{1}{2}\left((\sigma\sigma^*)\nabla V - \Big[\sum_{j=1}^d \partial_{x_j}(\sigma\sigma^*)_{ij}\Big]_{i=1:d}\right).$$

Then, the distribution

$$\nu_V(dx) = C_V e^{-V(x)} \cdot \lambda_d(dx)$$

is the unique invariant distribution of the above Brownian diffusion (2.15).

The proof of this proposition is postponed to Appendix C.3.

For a given Gibbs distribution ν_V, the existence of such a family of diffusions opens the opportunity to optimize the choice of the diffusion coefficient in view of the numerical approximation ν_V. In some cases, it is clearly of interest to introduce non constant diffusion coefficients. For instance, in the example below, we show that the weak mean-reverting of the Langevin diffusion (with constant σ) related to a particular Gibbs distribution ν_V can be dramatically strengthened by replacing it by a diffusion with non-constant diffusion coefficient (which is shown to be strongly reverting and exponentially contracting).

In the same direction, in [BJM16], the authors show that the optimal constant in one-dimensional weighted Poincaré inequalities can be obtained as the spectral gap of diffusion operators with non constant σ. This toy-example and the above reference emphasize the fact that considering non constant σ may help devising procedures whose rate of convergence can be more precisely controlled. Using non-constant σ, i.e. non-isotropic colored noises in stochastic gradient procedures frequently appears in the abundant literature on machine learning (see e.g. [MCF15b] or [LCCC15b] among many others). Nevertheless, investigating this problem in greater depth is beyond the scope of the paper and will be the object of future works.

Example. Let us consider the distribution on $\mathbb{R}^d$ with exponent $\kappa > 0$ defined by

$$\nu_\kappa(dx) = \frac{C_\kappa}{(1+|x|^2)^{d+\kappa}}\lambda_d(dx) = C_\kappa e^{-V(x)}\lambda_d(dx) \quad \text{with} \quad V(x) = (d+\kappa)\log(1+|x|^2)+1.$$

By (2.14) applied with $\sigma = I_d$, the distribution ν_κ is the invariant distribution of the one-dimensional Brownian diffusion,

$$dY_t = -(d+\kappa)\frac{Y_t}{1+|Y_t|^2}dt + dW_t.$$

Let $\mathcal{L}_Y$ denote the infinitesimal generator of this SDE. One has

$$\mathcal{L}_Y V(y) = -|\nabla V(y)|^2 + \tfrac{1}{2}\mathrm{Tr}(\nabla^2 V(y)) = -(d+\kappa)\frac{(2(d+\kappa)+1)|y|^2-1}{(1+|y|^2)^2} \sim -\frac{\big(2(d+\kappa)+1\big)(d+\kappa)}{|y|^2}$$

as $|y| \to +\infty$. Hence, the diffusion cannot be strongly mean-reverting since

$$\mathcal{L}_Y V(y) \to 0 \quad \text{as} \quad |y| \to +\infty.$$

On the other hand, applying now (2.14) applied with $\sigma(x) = (1+|x|^2)^{1/2}I_d$, the distribution ν_κ is also the invariant distribution of the Brownian diffusion

$$dX_t = -(d+\kappa-1)X_t dt + \sqrt{1+|X_t|^2}\,dW_t \tag{2.16}$$

whose infinitesimal generator $\mathcal{L}_X$ satisfies, when applied to the functions $W_\alpha(x) = (1+|x|^2)^\alpha$, $\alpha \in (0,1]$,

$$\mathcal{L}_X W_\alpha(x) \sim -\alpha\big(2(d+\kappa)-1-2\alpha\big)(|x|^2+1)^\alpha \quad \text{as} \quad |x| \to +\infty.$$

Hence, one can easily deduce that, strong mean-reversion $(\mathbf{S})$ holds for W_α iff $\alpha < d+\kappa-\frac{1}{2}$ and $\alpha \in (0,1]$ (in particular, this is always true for $\alpha = 1$ when $d \geqslant 2$). Furthermore, setting $b(x) = -(d+\kappa-1)x$ and using that $x \mapsto (1+|x|^2)^{\frac{1}{2}}$ is 1-Lipschitz, one also remarks that

$$\big(b(x)-b(y)\,|\,x-y\big) + \tfrac{1}{2}\|\sigma(x)-\sigma(y)\|_F^2 \leqslant \left(-(d+\kappa-1)+\frac{d}{2}\right)|x-y|^2$$

so that $(\mathbf{C}_\alpha)$ is satisfied as soon as $\kappa > 1-\frac{d}{2}$. Hence, for (2.16), $(\mathbf{H}_{\mathcal{W}_1})$ and $(\mathbf{S})$ hold true for any $\kappa > (1-\frac{d}{2})_+$ (true for any $\kappa > 0$ when $d \geqslant 2$).

2.5 Roadmap of the proof

The sequel of the paper is devoted to the proof of the above theorems. The aim of the next Section 3 is to recall or provide tools used to establish our main results: thus we recall in Section 3.1, basic confluence properties, the *Bismut-Elworthy-Li* formula (BEL in what follows), Then, in Subsection 3.2, we provide a series of strong and weak error bounds for a one-step Euler scheme which will play a key role to deduce the results (see also Appendix A). Finally, we state in Subsection 3.3 a general result on weak error expansions for non-smooth functions of the Euler scheme with decreasing step under an ellipticity assumption which relies on Malliavin calculus. The proofs of both Theorems 2.2 and 2.3 are divided in several steps and detailed in Section 4, some parts of the proofs are postponed in the Appendices A, B, C and D (to improve te readability).

3 Toolbox and preliminary results

Throughout the paper we will use the notations

$$S(x) = 1 + |b(x)| + \|\sigma(x)\| \quad \text{and} \quad S_{p,b,\sigma,\ldots}(x) = C_{p,b,\sigma,\ldots}.S(x) \tag{3.17}$$

where $C_{p,b,\sigma,\ldots}$ denotes a real constant depending on p, b, σ, etc, that may vary from line to line. These dependencies will sometimes be (partially) omitted.

3.1 BEL formula and differentiability of the diffusion semi-group

We now recall the classical Bismut-Elworthy-Li formula (see [Bis84, EL94, Cer00]), referred to as BEL formula in what follows.

Theorem 3.1 (Bismut-Elworthy-Li formula)**.** *Assume b and σ are $\mathcal{C}^1$ with bounded first order partial derivatives. Assume furthermore that $(\mathcal{E}\ell)_{\underline{\sigma}_0^2}$ holds. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a bounded Borel function. Then, denote by σ^{-1} the right-inverse matrix of σ. Then, for every $t > 0$, the mapping $x \mapsto P_t f(x) = \mathbb{E}\, f(X_t^x)$ is differentiable and*

$$\nabla_x P_t f(x) = \mathbb{E}\, \nabla_x f(X_t^x) = \nabla_x \mathbb{E}\Big[f(X_t^x)\frac{1}{t}\int_0^t \big(\sigma(X_s^x)^{-1} Y_s^{(x)}\big)^* dW_s\Big] \tag{3.18}$$

where $(Y_s^{(x)})_{s\geqslant 0}$ stands for the tangent process at x of the SDE (1.1) defined by $Y_t^{(x)} = \frac{dX_t^x}{dx}$, $t \geqslant 0$.

Moreover the above result remains true if f is a Borel function with polynomial growth.

The proof for unbounded f is postponed to Annex C.1.

Proposition 3.2. *(a) Let $f : \mathbb{R}^d \to \mathbb{R}$ be a bounded Borel function. Let $T > 0$. Then for every $k = 1, 2, 3$, there exist a real constant C_k depending on b and σ (and possibly on T) such that,*

$$\forall\, t \in (0, T], \quad |\partial_{x^k} P_t f(x)| \leqslant \frac{C_k}{\underline{\sigma}_0^k t^{\frac{k}{2}}} \|f\|_{\sup}. \tag{3.19}$$

(b) Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz continuous function. Let $T > 0$. Then for every $k = 1, 2, 3$, there exist a real constant C'_k depending on b and σ (and possibly on T) such that,

$$\forall\, t \in (0, T], \quad |\partial_{x^k} P_t f(x)| \leqslant \frac{C'_k}{\underline{\sigma}_0^k t^{\frac{k-1}{2}}} [f]_{\mathrm{Lip}} S(x) \tag{3.20}$$

The proof is postponed to Appendix C.4.

3.2 One step L^p-strong and weak error bounds for the Euler scheme

Strong error.

Lemma 3.3 (One step strong error I)**.** *Let $p\in[2,+\infty)$. Assume b and σ Lipschitz continuous so that $(X^x_t)_{t\geqslant 0}$ is well-defined as the unique strong solution of SDE starting from $x\in\mathbb{R}^d$. Let $(\bar X^{\gamma,x}_t)_{t\in[0,\gamma]}$ denote the (continuous) one step Euler scheme with step $\gamma>0$ starting from x at time 0.*

(a) For every $t\in[0,\gamma]$,

$$\|X^x_t-\bar X^{\gamma,x}_t\|_p\leqslant[b]_{\rm Lip}\int_0^t\|X^x_s-x\|_p ds+C_p[\sigma]_{\rm Lip}\left(\int_0^t\|X^x_s-x\|_p^2ds\right)^{1/2}.$$

where C_p is a positive real constant only depending on p.

(b) In particular, if $\sigma(x)=\sigma$ is a constant matrix,

$$\|X^x_t-\bar X^{\gamma,x}_t\|_p\leqslant[b]_{\rm Lip}\int_0^t\|X^x_s-x\|_p ds.$$

Lemma 3.4 (One step strong error II)**.** *Assume b and σ Lipschitz continuous. Let $\bar\gamma>0$.*

(a) $p\in[2,+\infty)$. The diffusion process $(X^x_t)_{t\geqslant 0}$ satisfies for every $t\in[0,\bar\gamma]$

$$\|X^x_t-x\|_p\leqslant S_{d,p,b,\sigma,\bar\gamma}(x)\sqrt{t}\tag{3.21}$$

where the underlying real constant $C_{d,p,b,\sigma,\bar\gamma}$ depends on b and σ only through $[b]_{\rm Lip}$, $[\sigma]_{\rm Lip}$. As for the one step Euler scheme $(\bar X^{\gamma,x}_t)_{t\geqslant 0}$ with step $\gamma\in(0,\bar\gamma]$, we have

$$\forall\, t\in[0,\gamma],\quad \|\bar X^{\gamma,x}_t-x\|_p\leqslant S_{d,p,b,\sigma,\bar\gamma}(x)\sqrt{t}.\tag{3.22}$$

(b) Let $p\in[1,+\infty)$. The one step strong error satisfies, for every $\gamma\in(0,\bar\gamma]$ and every $t\in[0,\gamma]$,

$$\|X^x_t-\bar X^{\gamma,x}_t\|_p\leqslant S_{d,p\vee 2,b,\sigma,\bar\gamma}(x)\left(\tfrac{2}{3}[b]_{\rm Lip}\sqrt{t}+\frac{[\sigma]_{\rm Lip}}{\sqrt{2}}\right)t.\tag{3.23}$$

(c) Let $p\in[1,+\infty)$. In particular, if $\sigma(x)=\sigma>0$ is constant, then, for every $\gamma>0$ and every $t\in[0,\gamma]$,

$$\|X^x_t-\bar X^{\gamma,x}_t\|_p\leqslant S_{d,p\vee 2,b,\sigma,\bar\gamma}(x)t^{3/2}.\tag{3.24}$$

Both proofs are postponed to the Appendix A.2.

Weak error. We first establish a weak error bound for smooth enough functions (C^3, see below) with a control by its fist three derivatives. Then we apply this to the semigroup P_tf where f is simply Lipschitz to take advantage of the regularizing effect of the semi-group.

Proposition 3.5 (Weak error for smooth functions)**.** *Assume b and σ are $\mathcal{C}^2$ with bounded first and second order derivatives. Let $\bar\gamma>0$. Let $g:\mathbb{R}^d\to\mathbb{R}$ be a three times differentiable function.*

(a) There exists a real constant $C_{d,b,\sigma,\bar\gamma}>0$ such that, for every $\gamma\in(0,\bar\gamma]$,

$$\left|\mathbb{E}\,[g(\bar X^x_\gamma)]-\mathbb{E}\,[g(X^x_\gamma)]\right|\leqslant S_{d,b,\sigma,\bar\gamma}(x)^3\gamma^2\Phi_{1,g}(x)\tag{3.25}$$

where $\Phi_{1,g}(x)=\max\Big(|\nabla g(x)|,\|D^2g(x)\|,\Big\|\sup_{\xi\in(X^x_\gamma,\bar X^x_\gamma)}\|D^2g(\xi)\|\Big\|_2,\Big\|\sup_{\xi\in(x,X^x_\gamma)}\|D^3g(\xi)\|\Big\|_4\Big)$ *and $(a,b)=\{\lambda a+(1-\lambda)b,\lambda\in(0,1)\}$ stands for the open geometric interval with endpoints a, b.*

(*b*) *If* $\sigma(x)=\sigma$ *is constant, the inequality can be refined for every* $\gamma\in(0,\bar\gamma]$ *as follows*

$$
\begin{aligned}
&|\mathbb{E}\,[g(\bar X^x_\gamma)]-\mathbb{E}\,[g(X^x_\gamma)]-\tfrac{\gamma^2}{2}\mathfrak{T}(g,b,\sigma)(x)|\\
&\qquad\leqslant \gamma^2 S_{d,b,\sigma,\bar\gamma}(x)^2|\nabla g(x)|+\gamma^{5/2}\Phi_{2,g}(x)S_{d,b,\sigma,\bar\gamma}(x)^3
\end{aligned}
\tag{3.26}
$$

where

$$
\mathfrak{T}(g,b,\sigma)(x)=\sum_{1\leqslant i,j\leqslant d}\partial^2_{x_ix_j}g(x)\big((\sigma\sigma^*)_{i\cdot}|\nabla b_j\big)(x),
\tag{3.27}
$$

and $\Phi_{2,g}(x)=\max\Big(\|D^2g(x)\|,\Big\|\sup_{\xi\in[x,X^x_\gamma)}\|D^3g(\xi)\|\Big\|_4\Big)$.

Proof. (*a*) By the second order Taylor formula, for every $y,z\in\mathbb{R}^d$,

$$
g(z)-g(y)=(\nabla g(y)|z-y)+\int_0^1(1-u)D^2g\big(uz+(1-u)y\big)du(z-y)^{\otimes 2}
$$

where, for a $d\times d$-matrix A and a vector $u\in\mathbb{R}^d$, $Au^{\otimes 2}=(Au|u)$. For a given $x\in\mathbb{R}^d$, it follows that

$$
\begin{aligned}
g(z)-g(y)&=(\nabla g(x)|z-y)+(\nabla g(y)-\nabla g(x)|z-y)+\int_0^1(1-u)D^2g\big(uz+(1-u)y\big)(z-y)^{\otimes 2}du\\
&=(\nabla g(x)|z-y)+\big(D^2g(x)(y-x)|z-y)\\
&\quad+\int_0^1(1-u)D^3g(uy+(1-u)x)(y-x)^{\otimes 2}(z-y)du\\
&\quad+\int_0^1(1-u)D^2g\big(uz+(1-u)y\big)du(z-y)^{\otimes 2}.
\end{aligned}
$$

Applying this expansion with $y=X^x_\gamma$ and $z=\bar X^x_\gamma$, this yields:

$$
\begin{aligned}
\mathbb{E}\,[g(\bar X^x_\gamma)-g(X^x_\gamma)]&=\underbrace{(\nabla g(x)|\mathbb{E}\,[\bar X^x_\gamma-X^x_\gamma])}_{=:A_1}+\underbrace{\mathbb{E}\,\big[(D^2g(x)(X^x_\gamma-x)|\bar X^x_\gamma-X^x_\gamma)\big]}_{=:A_2}\\
&+\underbrace{\mathbb{E}\left[\int_0^1(1-u)D^3g(uX^x_\gamma+(1-u)x)(X^x_\gamma-x)^{\otimes 2}(\bar X^x_\gamma-X^x_\gamma)du\right]}_{=:A_3}\\
&+\underbrace{\int_0^1(1-u)\mathbb{E}\,\big[D^2g\big(u\bar X^x_\gamma+(1-u)X^x_\gamma\big)(\bar X^x_\gamma-X^x_\gamma)^{\otimes 2}\big]du}_{=:A_4}.
\end{aligned}
$$

Let us inspect successively the four terms of the right-hand member.
Term A_1. First,

$$
\mathbb{E}\,[(\bar X^x_\gamma-X^x_\gamma)_i]=\mathbb{E}\Big[\int_0^\gamma\big(b(X_s)-b(x)\big)_i ds\Big]=\int_0^\gamma\int_0^s\mathbb{E}\,[\mathcal{L}b_i(X^x_u)]duds,
\tag{3.28}
$$

Since b has bounded partial derivatives, $|\mathcal{L}b_i(x)|\leqslant C_{b,\sigma}\big(|b(x)|+\|\sigma(x)\|^2\big)$ so that

$$
|(\nabla g(x)|\mathbb{E}\,[\bar X^x_\gamma-X^x_\gamma])|\leqslant|\nabla g(x)||\mathbb{E}\,[\bar X^x_\gamma-X^x_\gamma]|\leqslant C_{b,\sigma}\Psi(x)|\nabla g(x)|\gamma^2
$$

with

$$
\Psi(x)=\sup_{0\leqslant t\leqslant\bar\gamma}\mathbb{E}\,[|b(X^x_t)|+\|\sigma(X^x_t)\|^2].
\tag{3.29}
$$

Now note that

$$\begin{aligned}\Psi(x) &\leqslant \big(|b(x)| + 2\,\|\sigma(x)\|^2\big) + [b]_{\rm Lip} \sup_{0\leqslant t\leqslant\bar\gamma} \|X_t^x - x\|_1 + 2\,[\sigma]^2_{\rm Lip} \sup_{0\leqslant t\leqslant\bar\gamma} \|X_t^x - x\|_2^2\\ &\leqslant \big(|b(x)| + 2\|\sigma(x)\|^2\big) + [b]_{\rm Lip} C_{d,b,1,\sigma,\bar\gamma} S_1(x) + [\sigma]^2_{\rm Lip} C_{d,b,2,\sigma,\bar\gamma} S(x)^2\\ &\leqslant S_{d,b,\sigma,\bar\gamma}(x)^2 \end{aligned}\tag{3.30}$$

(where real constants $C_{d,b,p,\sigma,\bar\gamma}$ come from Lemma 3.4).

For the sake of simplicity, we omit the dependence in x in the notations of the sequel of the proof.

Term A_2. Temporary denoting by $u_1,\dots,u_d$ the components of a vector u of $\mathbb{R}^d$, we have for every $i,j \in \{1,\dots,d\}$,

$$|A_2| \leqslant \sum_{1\leqslant i,j\leqslant d} \big|\partial_{x_i x_j} g(x)\big| \big|\mathbb{E}\,[(X_\gamma - x)_i (X_\gamma - \bar X_\gamma)_j]\big|$$

with $\quad \mathbb{E}\,[(X_\gamma - x)_i (X_\gamma - \bar X_\gamma)_j] = -\mathbb{E}\,[(X_\gamma - \bar X_\gamma)_i (X_\gamma - \bar X_\gamma)_j] + \mathbb{E}\,[(\bar X_\gamma - x)_i (X_\gamma - \bar X_\gamma)_j].$

By Lemma 3.4(c), we deduce the existence of a positive constant $C_{b,\sigma,\bar\gamma}$ such that

$$|\mathbb{E}\,[(X_\gamma - \bar X_\gamma)_i (X_\gamma - \bar X_\gamma)_j]| \leqslant \mathbb{E}\,[|X_\gamma - \bar X_\gamma|^2] \leqslant S_{b,\sigma,\bar\gamma}(x)^2\gamma^2.$$

On the other hand,

$$(\bar X_\gamma - x)_i (X_\gamma - \bar X_\gamma)_j = (\gamma b(x) + \sigma(x) W_\gamma)_i \left(\int_0^\gamma \big(b(X_s) - b(x)\big) ds + \int_0^\gamma \big(\sigma(X_s) - \sigma(x)\big) dW_s\right)_j,$$

hence (using that the increments of the Brownian Motion are independent and centered),

$$\begin{aligned}\mathbb{E}\Big[(\bar X_\gamma - x)_i (X_\gamma - \bar X_\gamma)_j\Big] &= \gamma\, b_i(x) \mathbb{E}\Big[\int_0^\gamma \int_0^s \mathcal{L} b_j(X_u)\Big] du + \mathbb{E}\Big[\int_0^\gamma (\sigma(x) W_\gamma)_i (b(X_s) - b(x))_j ds\Big]\\ &\quad + \mathbb{E}\left[(\sigma(x) W_\gamma)_i \Big(\int_0^\gamma (\sigma(X_s) - \sigma(x)) dW_s\Big)_j\right].\end{aligned}\tag{3.31}$$

By the same argument used to upper-bound A_1, we first get

$$\gamma\Big| b_i(x) \mathbb{E}\Big[\int_0^\gamma \int_0^s \mathcal{L} b_j(X_u)\Big] du ds\Big| \leqslant C_{b,\sigma} \Psi(x) |b(x)| \gamma^3,$$

where Ψ is defined by (3.29). Then, it follows from Cauchy-Schwarz inequality and (3.21) that

$$\begin{aligned}\mathbb{E}\,[|(\sigma(x) W_\gamma)_i (b(X_s) - b(x))_j|] &\leqslant \Big\| \sum_{1\leqslant j\leqslant q} \sigma_{ij}(x) W_\gamma^j \Big\|_2 \|(b(X_s) - b(x))_j\|_2\\ &\leqslant |\sigma_{i\cdot}(x)| \sqrt{\gamma}\, [b]_{\rm Lip} \|X_s - x\|_2 \leqslant [b]_{\rm Lip} \|\sigma(x)\| S_{d,2,b,\sigma,\bar\gamma}(x) \sqrt{\gamma}\sqrt{s}.\end{aligned}$$

Hence, as $\int_0^\gamma \sqrt{s} ds = \frac{2}{3}\gamma^{3/2}$, one has

$$\left|\mathbb{E}\left[\int_0^\gamma (\sigma(x) W_\gamma)_i (b(X_s) - b(x))_j ds\right]\right| \leqslant C_{d,2,b,\sigma,\bar\gamma} [b]_{\rm Lip} \|\sigma(x)\| S(x) \gamma^2.$$

For the third term in the right hand side of (3.31), we deduce from Itô's isometry that

$$\begin{aligned}\mathbb{E}\left[(\sigma(x) W_\gamma)_i \Big(\int_0^\gamma (\sigma(X_s) - \sigma(x)) dW_s\Big)_j\right] &= \sum_{k=1}^d \int_0^\gamma \mathbb{E}\,[\sigma_{i,k}(x)(\sigma_{jk}(X_s) - \sigma_{jk}(x)] ds\\ &= \sum_{k=1}^d \sigma_{ik}(x) \int_0^\gamma \int_0^s \mathbb{E}\,[\mathcal{L}\sigma_{jk}(X_u)] du ds.\end{aligned}$$

Since the partial derivatives of σ are bounded, we again deduce that this term is bounded $C'_{b,\sigma}\|\sigma(x)\|\Psi(x)\gamma^2$. Finally, collecting the above bounds yields

$$|A_2| \leqslant C_{b,\sigma,\bar\gamma}\max\big(\|D^2g(x)\|, |\nabla g(x)|\big)\max\big(S(x),\Psi(x)\big)(1+\|\sigma(x)\|+\gamma|b(x)|)\gamma^2.$$

Now, we focus on A_3:

$$|A_3| \leqslant \tfrac{1}{2}\mathbb{E}\left[\sup_{\xi\in(x,X^x_\gamma)}\|D^3g(\xi)\||X^x_\gamma - x|^2|\bar X^{\gamma,x}_\gamma - X^x_\gamma|\right].$$

By (three fold) Cauchy-Schwarz inequality and Lemma 3.4(b)

$$\begin{aligned}|A_3| &\leqslant \tfrac{1}{2}\Big\|\sup_{\xi\in(x,X^x_\gamma)}\|D^3g(\xi)\|\Big\|_4\|X^x_\gamma - x\|_4^2\|\bar X^{\gamma,x}_\gamma - X^x_\gamma\|_4\\ &\leqslant \tfrac{1}{2}\Big\|\sup_{\xi\in(x,X^x_\gamma)}\|D^3g(\xi)\|\Big\|_4 C_{d,4,b,\sigma,\bar\gamma}S(x)^3\gamma^2. \qquad (3.32)\end{aligned}$$

Note that the power 3 in b (and σ) comes from this term. To conclude the proof, let consider A_4:

$$|A_4| \leqslant \tfrac{1}{2}\Big\|\sup_{\xi\in(X^{\gamma,x}_\gamma,\bar X^{\gamma,x}_\gamma)}\|D^2g(\xi)\|\Big\|_2\|\bar X^{\gamma,x}_\gamma - X^x_\gamma\|_4^2 \leqslant \frac{C'_{d,4,b,\sigma,\bar\gamma}}{2}\Big\|\sup_{\xi\in(X^x_\gamma,\bar X^x_\gamma)}\|D^2g(\xi)\|\Big\|_2 S(x)^2\gamma^2.$$

(b) First note that the third term in the right hand side of (3.31) vanishes since σ is constant. Secondly, note that using the improved bound for $\|\bar X^{\gamma,x}_\gamma - X^x_\gamma\|_4$ (in $\gamma^{3/2}$) from Lemma 3.4(c) in that setting, γ^2 can be replaced in the above bound for $|A_4|$ by $\gamma^{5/2}$.

Let us focus now on the second term in the right hand side of (3.31). We write

$$\begin{aligned}\int_0^\gamma(\sigma W_s)_i\big(b_j(X^x_s)-b_j(x)\big)ds &= \int_0^\gamma(\sigma W_s)_i\int_0^s\mathcal{L}b_j(X^x_u)du ds\\ &+ \int_0^\gamma(\sigma W_s)_i(\nabla b_j(x)|\sigma W_s)ds + \int_0^\gamma(\sigma W_s)_i\int_0^s\big(\nabla b_j(X^x_u)-\nabla b_j(x)|\sigma dW_u\big)ds.\end{aligned}$$

We inspect these three terms. Using that W has independent increments, we get

$$\mathbb{E}\Big[\int_0^\gamma(\sigma W_s)_i\int_0^s\mathcal{L}b_j(X^x_u)du ds\Big] = \int_0^\gamma\int_0^s\mathbb{E}\big[(\sigma W_u)_i\mathcal{L}b_j(X^x_u)\big]du ds$$

so that, by Cauchy-Schwarz inequality,

$$\begin{aligned}\Big|\int_0^\gamma\int_0^s\mathbb{E}\big[(\sigma W_u)_i\mathcal{L}b_j(X^x_u)\big]du ds\Big| &\leqslant \int_0^\gamma\int_0^s\|(\sigma W_u)_i\|_2\|\mathcal{L}b_j(X^x_u)\|_2 du ds\\ &\leqslant C_{\|\nabla b_j\|_{\sup},\|\sigma\|}\big(1+\sup_{u\in(0,\gamma)}\|b(X^x_u)\|_2\big)\gamma^{5/2}\\ &\leqslant C'_{b,\|\sigma\|}\big(1+|b(x)|\big)\gamma^{5/2}.\end{aligned}$$

On the other hand, noting $(\sigma\sigma)^*_{i\cdot} = [(\sigma\sigma)^*_{ik}]_{1\leqslant k\leqslant d}$,

$$\mathbb{E}\int_0^\gamma(\sigma W_s)_i(\nabla b_j(x)|\sigma W_s)ds = \frac{\gamma^2}{2}\big((\sigma\sigma^*)_{i\cdot}|\nabla b_j\big)$$

Finally, using Itô's isometry and the boundedness of second partial derivatives of b, we get

$$\begin{aligned}\Big|\mathbb{E}\int_0^\gamma(\sigma W_s)_i\int_0^s\big(\nabla b_j(X^x_u)-\nabla b_j(x)|\sigma dW_u\big)ds\Big| &= \Big|\int_0^\gamma\mathbb{E}\Big[(\sigma W_s)_i\int_0^s\big(\nabla b_j(X^x_u)-\nabla b_j(x)|\sigma dW_u\big)\Big]ds\Big|\\ &\leqslant C_{b,\sigma}\int_0^\gamma\int_0^s\|X^x_u - x\|_2\,du\,ds \leqslant C'_{b,\sigma}\gamma^{5/2}S(x)\end{aligned}$$

which completes the proof. $\square$

Combining the above results with Proposition 3.2(b) and Lemma A.2 yields the following precise error bound for the one step weak error.

Proposition 3.6 (One step weak error at time t). *Assume b is $\mathcal{C}^3$ and σ is $\mathcal{C}^4$ with bounded existing partial derivatives and $|b|^2 + \|\sigma\|^2 \leqslant C \cdot V$. Assume that $(\mathcal{E}\ell)_{\underline{\sigma}_0^2}$ holds. Let T, $\bar{\gamma} > 0$.*

Then, there exists a positive constant $C = C_{b,\sigma,\underline{\sigma}_0,T,\bar{\gamma},V}$ such that, for every Lipschitz continuous function f and every $t \in (0,T]$,

$$\forall \gamma \in (0,\bar{\gamma}], \quad \left|\mathbb{E}\left[P_t f(\bar{X}_\gamma^{\gamma,x})\right] - \mathbb{E}\left[P_t f(X_\gamma^x)\right]\right| \leqslant C[f]_{\rm Lip}\gamma^2 t^{-1} V^2(x).\big(1 + |b(x)|^3 + \|\sigma(x)\|^3\big).$$

Proof. We apply Proposition 3.5(a) to $g_t = P_t f(x)$ with $t > 0$. It follows from Proposition 3.2(b) (see (3.20)) that the function $\Phi_{1,g}$ in (3.25) satisfies

$$\begin{aligned}\Phi_{1,g_t}(x) &\leqslant C_{b,\sigma,\underline{\sigma}_0}\frac{[f]_{\rm Lip}}{t}\max\Big(S(x), \big\|\sup_{\xi\in(X_\gamma^x,\bar{X}_\gamma^x)} S(\xi)\big\|_2, \big\|\sup_{\xi\in(x,X_\gamma^x)} S(\xi)\big\|_4\Big)\\ &\leqslant C_{b,\sigma,\underline{\sigma}_0}\frac{[f]_{\rm Lip}}{t}V^{\frac{1}{2}}(x)\end{aligned}$$

owing to Lemma A.2 in Appendix A and where we used that $S \leqslant C_{b,\sigma}V^{\frac{1}{2}}$. Consequently

$$\begin{aligned}\left|\mathbb{E}\left[P_t f(\bar{X}_\gamma^{\gamma,x})\right] - \mathbb{E}\left[P_t f(X_\gamma^x)\right]\right| &\leqslant C[f]_{\rm Lip}\gamma^2\,\big(1 + |b(x)|^3 + \|\sigma(x)\|^3\big)V^{\frac{1}{2}}(x)t^{-1}\\ &\leqslant C[f]_{\rm Lip}\gamma^2 t^{-1}V^2(x).\end{aligned}$$

$\square$

3.3 Domino-Malliavin for non smooth functions

For the control in variation distance, we will need a weak error estimate for Borel functions of the one step Euler scheme starting from a "non-degenerate" random variable to produce a "regularization form the past". It mainly relies on a Malliavin calculus approach. In the theorem below $(h_n)_{n\geqslant 1}$ denotes a non-increasing step sequence. Set $t_n = \sum_{k=1}^n h_k$ (and $t_0 = 0$) in what follows.

Theorem 3.7 (Domino-Malliavin). *Assume that σ is bounded and satisfies $(\mathcal{E}\ell)_{\underline{\sigma}_0^2}$, that b has sublinear growth: $|b(x)| \leqslant C(1+|x|)$. Assume that b and σ are $\mathcal{C}^6$-functions with bounded partial derivatives. Then, for every $\varepsilon > 0$, $T > 0$ and $\bar{h} > 0$, there exists $C_{T,\bar{h},\varepsilon} > 0$ such that for any $h_1 \in (0,\bar{h})$ and any $n \geqslant 1$ satisfying $\frac{T}{2} \leqslant t_n \leqslant T$ and any bounded Borel function $f : \mathbb{R}^d \to \mathbb{R}$,*

$$|\bar{P}_{h_1} \circ \cdots \circ \bar{P}_{h_{n-1}} \circ (P_{h_n} - \bar{P}_{h_n}) \circ f(x)| \leqslant C_{T,\bar{h},\varepsilon}(1+|x|^8)\|f\|_{\sup} h_1^{2-\varepsilon}. \tag{3.33}$$

Remark 3.1. With further technicalities, it seems that we could obtain $1 + |x|^6$ instead of $1 + |x|^8$. Nevertheless, since the degree of the polynomial function involved in the result is not fundamental for our paper, we did not detail this point (more precisely, the improvement could be obtained by separating drift and diffusion components in the Taylor formula (B.48).

4 Proof of the main theorems

The starting point of the proofs of both claims of the main theorem is to decompose the error using a *domino strategy*. Let us provide the heuristic by only considering a given function $f : \mathbb{R}^d \to \mathbb{R}$ (typically, a bounded Borel function when dealing with the total variation distance or a 1-Lipschitz continuous function if dealing with the L^1- Wasserstein distance $\mathcal{W}_1$). In this case, we can write:

$$\big|\mathbb{E}\, f(X_{\Gamma_n}^x) - \mathbb{E}\, f(\bar{X}_{\Gamma_n}^x)\big| \leqslant \sum_{k=1}^n \big|\bar{P}_{\gamma_1} \circ \cdots \circ \bar{P}_{\gamma_{k-1}} \circ (P_{\gamma_k} - \bar{P}_{\gamma_k}) \circ P_{\Gamma_n - \Gamma_k} f(x)\big|.$$

4.1 Proof of Theorem 2.2(b) (Total variation distance)

Let $\bar\gamma = \|\boldsymbol{\gamma}\| = \sup_{n\geqslant 1}\gamma_n$. Let $T > 2\bar\gamma$ be fixed. We may assume without loss of generality (w.l.g.) that $\Gamma_n > 2T$ ([5]). Furthermore, under $(\mathcal{E}\ell)_{\underline{\sigma}_0^2}$, $(\mathbf{H_{TV}})$ holds for any $t_0 > 0$ owing to Proposition 2.1, so we may set $t_0 = \bar\gamma$ throughout the proof.

For the *TV* distance, the idea is then to separate this sum into two partial sums, namely,

$$\begin{aligned}\big|\mathbb{E}\, f(X^x_{\Gamma_n}) - \mathbb{E}\, f(\bar X^x_{\Gamma_n})\big| \leqslant & \sum_{k=1}^{N(\Gamma_n-T)} \big|\bar P_{\gamma_1}\circ\cdots\circ \bar P_{\gamma_{k-1}}\circ(P_{\gamma_k}-\bar P_{\gamma_k})\circ P_{\Gamma_n-\Gamma_k} f(x)\big| \\ & + \sum_{k=N(\Gamma_n-T)+1}^{n} \big|\bar P_{\gamma_1}\circ\cdots\circ \bar P_{\gamma_{k-1}}\circ(P_{\gamma_k}-\bar P_{\gamma_k})\circ P_{\Gamma_n-\Gamma_k} f(x)\big|.\end{aligned}$$

where $f:\mathbb{R}^d\to\mathbb{R}$ is bounded Borel function.

These two terms, say (A) and (B) respectively, correspond to two different types of *weak errors*: first the "ergodic term" where the exponential contraction of the semi-group can be exploited and weak error results for smooth functions (here $P_{\Gamma_n-\Gamma_k}f$ with $\Gamma_n-\Gamma_k\geqslant T$) can be used (see Proposition 3.6), then the second term where the smoothing effect of the operator $P_{\Gamma_n-\Gamma_k}$ ($\Gamma_n-\Gamma_k\in[0,T]$) is no longer smooth enough leading us to establish a one step weak error expansion for bounded Borel functions (see Theorem 3.7).

<u>Term (A).</u> Let $k\in\{1,\ldots,N(\Gamma_n-T)\}$. Then $\Gamma_n-\Gamma_k > T$ and

$$\begin{aligned}|P_{\gamma_k}\circ P_{\Gamma_n-\Gamma_k} f(x) - \bar P_{\gamma_k}\circ P_{\Gamma_n-\Gamma_k} f(x)| & \\ &= \big|P_{\gamma_k}\circ P_{\frac{T}{2}}\circ P_{\Gamma_n-\Gamma_k-T/2}f(x) - \bar P_{\gamma_k}\circ P_{\frac{T}{2}}\circ P_{\Gamma_n-\Gamma_k-T/2}f(x)\big| \qquad (4.34)\\ &= \big|\mathbb{E}\, P_{\Gamma_n-\Gamma_k-T/2}f\big(\Xi^x_k\big) - \mathbb{E}\, P_{\Gamma_n-\Gamma_k-T/2}f\big(\bar\Xi^x_k\big)\big| \\ &\leqslant c e^{-\rho(\Gamma_n-\Gamma_k-T/2)}\|f\|_{\sup}\mathbb{E}\,\big[|X^{\Xi^x_k}_{\frac{T}{2}} - X^{\bar\Xi^x_k}_{\frac{T}{2}}|\big] \qquad (4.35)\end{aligned}$$

where we applied $(\mathbf{H_{TV}})$ with $t_0=\bar\gamma$ at time $t=\Gamma_n-\Gamma_k-\frac{T}{2}\geqslant\frac{T}{2}\geqslant\bar\gamma = t_0$, the bounded function f and $\bar\Xi^x_k$ and Ξ^x_k are any random vectors such that $\Xi^x_k \overset{d}{=} X^{X^x_{\gamma_k}}_{\frac{T}{2}}$ and $\bar\Xi^x_k \overset{d}{=} X^{\bar X^x_{\gamma_k}}_{\frac{T}{2}}$ (having in mind that X^x_t denotes the solution of (<u>SDE</u>) (1.1) starting from x at time t).

Thus, it follows from the definition of the L^1-Wasserstein distance that

$$\big|P_{\gamma_k}\circ P_{\Gamma_n-\Gamma_k} f(x) - \bar P_{\gamma_k}\circ P_{\Gamma_n-\Gamma_k} f(x)\big| \leqslant C_{\rho,T}e^{-\rho(\Gamma_n-\Gamma_k)}\|f\|_{\sup}\mathcal{W}_1\big(P_{\gamma_k}\circ P_{\frac{T}{2}}(x,dy), \bar P_{\gamma_k}\circ P_{\frac{T}{2}}(x,dy)\big)$$

with $C_{\rho,T} = c_{t_0}e^{\rho T/2}$. On the one hand, the Kantorovich-Rubinstein (see [Vil09]) representation of the L^1-Wasserstein distance says that

$$\begin{aligned}\mathcal{W}_1\big(P_{\gamma_k}\circ P_{\frac{T}{2}}(x,dy), \bar P_{\gamma_k}\circ P_{\frac{T}{2}}(x,dy)\big) &= \sup_{[g]_{\rm Lip}\leqslant 1}\mathbb{E}\big[g\big(X^{X^x_{\gamma_k}}_{\frac{T}{2}}\big) - g\big(X^{\bar X^x_{\gamma_k}}_{\frac{T}{2}}\big)\big] \\ &= \sup_{[g]_{\rm Lip}\leqslant 1}\mathbb{E}\big[P_{\frac{T}{2}}g(X^x_{\gamma_k}) - P_{\frac{T}{2}}g(\bar X^x_{\gamma_k})\big]\end{aligned}$$

Now, it follows from Proposition 3.6 applied with $t = T/2$ that

$$\big|\mathbb{E}\big[P_{\frac{T}{2}}g(X^x_{\gamma_k}) - P_{\frac{T}{2}}g(\bar X^x_{\gamma_k})\big]\big| \leqslant [g]_{\rm Lip}\frac{2}{T}C_{b,\sigma,\underline{\sigma}_0,T}\gamma_k^2V^2(x) \leqslant C'_{b,\sigma,\underline{\sigma}_0,T,\|\boldsymbol{\gamma}\|}\gamma_k^2V^2(x)$$

[5] When $\Gamma_n\leqslant 2T$, we can artificially upper-bound $\big|\mathbb{E}\, f(X^x_{\Gamma_n}) - \mathbb{E}\, f(\bar X^x_{\Gamma_n})\big|$ by $2\,\|f\|_{\sup}\gamma^{-1}_{N(2T)}\gamma_n$.

so that $\mathcal{W}_1(P_{\gamma_k}\circ P_{\frac{T}{2}}, \bar P_{\gamma_k}\circ P_{\frac{T}{2}}) \leqslant C'_{b,\sigma,\underline{\sigma}_0,T,\|\boldsymbol{\gamma}\|}\gamma_k^2 V^2(x)$. Hence

$$|P_{\gamma_k}\circ P_{\Gamma_n-\Gamma_k}f(x) - \bar P_{\gamma_k}\circ P_{\Gamma_n-\Gamma_k}f(x)| \leqslant C_{b,\sigma,\underline{\sigma}_0,T,\|\boldsymbol{\gamma}\|}e^{-\rho(\Gamma_n-\Gamma_k)}\|f\|_{\sup}\gamma_k^2 V^2(x). \tag{4.36}$$

Finally, integrating with respect to $\bar P_{\gamma_1}\circ\cdots\circ\bar P_{\gamma_{k-1}}$ yields

$$\begin{aligned}\left|\bar P_{\gamma_1}\circ\cdots\circ\bar P_{\gamma_{k-1}}\circ(P_{\gamma_k}-\bar P_{\gamma_k})\circ P_{\Gamma_n-\Gamma_k}f(x)\right| &\leqslant C_{b,\sigma,\underline{\sigma}_0,T,\|\boldsymbol{\gamma}\|}e^{-\rho(\Gamma_n-\Gamma_k)}\|f\|_{\sup}\gamma_k^2\sup_{\ell\geqslant 0}\mathbb{E}\,V^2(\bar X^x_{\Gamma_\ell})\\ &\leqslant C_{b,\sigma,\underline{\sigma}_0,T,\boldsymbol{\gamma}}e^{-\rho(\Gamma_n-\Gamma_k)}\|f\|_{\sup}\gamma_k^2 V^2(x)\end{aligned}$$

owing to Proposition A.1(a) (and where the constant $C_{\dots}$ may vary from line to line). As $\varpi<\rho$, Lemma A.3(i) implies the existence of a constant $C_{\boldsymbol{\gamma}}>0$ such that

$$\sum_{k=1}^{N(\Gamma_n-T)}\gamma_k^2 e^{-\rho(\Gamma_n-\Gamma_k)} \leqslant C_{\boldsymbol{\gamma}}\cdot\gamma_n$$

so that $|(A)|\leqslant C^{(4)}_{b,\sigma,\sigma_0,T,\boldsymbol{\gamma}}\|f\|_{\sup}\gamma_n V^2(x)$.

Term (B). Let us deal now with the the second term, when $k\in\{N(\Gamma_n-T)+1,\dots,n\}$. We assume that n is large enough so that $\Gamma_n>2T$ and temporarily set $\varphi_k = P_{\Gamma_n-\Gamma_k-T/2}f$. We apply Theorem 3.7 with $t_\ell = \Gamma_{N(\Gamma_n-2T)+\ell}-\Gamma_{N(\Gamma_n-2T)+\ell}$, $\ell\geqslant 1$, $2T$ (instead of T), $\bar h=\bar\gamma$ and $\varepsilon\in(0,2)$. Owing to the very definition of $N(t)$ and the fact that $\gamma_\ell\leqslant\bar\gamma$ for every $\ell\geqslant 1$, one checks that $\Gamma_k-\Gamma_{N(\Gamma_n-2T)+1}\leqslant\Gamma_n-(\Gamma_n-2T)=2T$ and

$$\Gamma_k-\Gamma_{N(\Gamma_n-2T)+1}\geqslant\Gamma_n-T-(\Gamma_n-2T+\|\gamma\|)\geqslant T-\bar\gamma\geqslant T/2.$$

Hence, it follows form (3.33) that

$$\left|\bar P_{\gamma_{N(\Gamma_n-2T)+1}}\circ\cdots\circ\bar P_{\gamma_{k-1}}\circ(P_{\gamma_k}-\bar P_{\gamma_k})\varphi_k(x)\right|\leqslant C_\varepsilon(1+|x|^8)\gamma^{2-\varepsilon}_{N(\Gamma_n-2T)+1}\|\varphi_k\|_{\sup}.$$

As a consequence

$$\left|\bar P_{\gamma_1}\circ\cdots\circ\bar P_{\gamma_{k-1}}\circ(P_{\gamma_k}-\bar P_{\gamma_k})\varphi_k(x)\right|\leqslant C_\varepsilon\sup_{\ell\geqslant 1}\mathbb{E}(1+|\bar X^x_{\Gamma_\ell}|^8)\gamma^{2-\varepsilon}_{N(\Gamma_n-2T)+1}\|f\|_{\sup}.$$

Finally as the step sequence satisfies $\varpi<\rho<+\infty$, $\gamma_{N(\Gamma_n-2T)+1}=O(\gamma_n)$ (see Lemma A.3(ii)), one has

$$\left|\bar P_{\gamma_1}\circ\cdots\circ P_{\gamma_{k-1}}\circ(P_{\gamma_k}-\bar P_{\gamma_k})\varphi_k(x)\right|\leqslant C'_{\gamma,\varepsilon}\sup_{\ell\geqslant 1}\mathbb{E}(1+|\bar X^x_{\Gamma_\ell}|^8)\gamma_k^{2-\varepsilon}\|f\|_{\sup}.$$

If $c_{V,r}=\liminf_{|x|\to+\infty}\frac{V(x)}{|x|^r}>0$, it follows from Proposition A.1(a) that

$$\sup_{\ell\geqslant 1}\mathbb{E}\left(1+|\bar X^x_{\Gamma_\ell}|^8\right)\leqslant c'_{V,r}\sup_{\ell\geqslant 1}\mathbb{E}\left(1+V^{8/r}(\bar X^x_{\Gamma_\ell})\right)\leqslant C'_{V,r,\,\gamma}\left(1+V(x)^{8/r}\right). \tag{4.37}$$

Now, by the definition of $N(\Gamma_n-T)$ and using again that $\varpi<\rho$, one has

$$\sum_{k=N(\Gamma_n-T)+1}^{n}\gamma_k^{2-\varepsilon}\leqslant\gamma^{1-\varepsilon}_{N(\Gamma_n-T)+1}\sum_{k=N(\Gamma_n-T)+1}^{n}\gamma_k\leqslant\gamma^{1-\varepsilon}_{N(\Gamma_n-T)}T\leqslant C'_{\|\boldsymbol{\gamma}\|}T\cdot\gamma_n^{1-\varepsilon}.$$

Applying $(\mathbf{H_{TV}})$, Proposition 2.1 (which allows to choose $t_0=\gamma_1>0$) and using that ν has a finite first moment, we have for the diffusion and for every $n\geqslant 1$,

$$\begin{aligned}d_{TV}([X^x_{\Gamma_n}],\nu) &= \int\nu(dy)d_{TV}([X^x_{\Gamma_n}],[X^y_{\Gamma_n}])\leqslant c_{\|\boldsymbol{\gamma}\|}\,\nu(|x-\cdot|)e^{-\rho\Gamma_n}\\ &\leqslant c_{\|\boldsymbol{\gamma}\|}\,\nu(|x-\cdot|)\left(|x|+\nu(|\cdot|)\right)e^{-\rho\Gamma_n}\end{aligned}$$

where we used that ν is invariant. Collecting all what precedes, we get for large enough n,

$$d_{TV}([\bar{X}^x_{\Gamma_n}],\nu) \leqslant d_{TV}([X^x_{\Gamma_n}],\nu) + d_{TV}([\bar{X}^x_{\Gamma_n}],[X^x_{\Gamma_n}]) \leqslant C_{b,\sigma,\|\boldsymbol{\gamma}\|}\psi(x)\left(e^{-\rho\Gamma_n} + \gamma_n^{1-\varepsilon} + \gamma_n\right) \\ \leqslant C_{b,\sigma,\|\boldsymbol{\gamma}\|}\vartheta(x)\gamma_n^{1-\varepsilon}$$

with $\vartheta(x) = C_{b,\sigma,\|\boldsymbol{\gamma}\|}V^{8/r}(x)$ (since $V^{8/r}$ dominates both V^2 and $|x|$) and where we used Lemma A.3(iii) with $a = 1$ to control $e^{-\rho\Gamma_n}$ by γ_n. As d_{TV} is bounded by 2 this holds for every n by changing the constant $C_{b,\sigma,\|\boldsymbol{\gamma}\|}$ if necessary.

If $\liminf_{|x|\to+\infty} V(x)/\log(1+|x|) = +\infty$, it follows from Proposition A.1(b) that $1+|x|^8 \leqslant c_{V,\lambda_0}e^{\lambda_0 V(x)}$ for any fixed $\lambda_0 \in (0,\lambda_{\sup}]$ and that $\sup_{n\geqslant 1}\mathbb{E}\, e^{\lambda_0 V(\bar{X}^x_{\Gamma_n})} \leqslant C_{b,\sigma,\lambda_0,\boldsymbol{\gamma}}e^{\lambda_0 V(x)}$ so that one may set $\vartheta(x) = e^{\lambda_0 V(x)}$ since this function also dominates $V(x)$ and $|x|$.

4.2 Proof of Theorem 2.2(a) (Wasserstein distance)

Let $f:\mathbb{R}^d\to\mathbb{R}$ be a Lipschitz continuous function with coefficient $[f]_{\rm Lip}$. The idea is now to separate this sum into three parts, namely, for a given $T>0$ ([6]).

$$\begin{aligned}\left|\mathbb{E}\, f(X^x_{\Gamma_n}) - \mathbb{E}\, f(\bar{X}^x_{\Gamma_n})\right| \leqslant & \sum_{k=1}^{N(\Gamma_n-T)}\left|\bar{P}_{\gamma_1}\circ\cdots\circ\bar{P}_{\gamma_{k-1}}\circ(P_{\gamma_k}-\bar{P}_{\gamma_k})\circ P_{\Gamma_n-\Gamma_k}f(x)\right| \\ & + \sum_{k=N(\Gamma_n-T)+1}^{n-1}\left|\bar{P}_{\gamma_1}\circ\cdots\circ\bar{P}_{\gamma_{k-1}}\circ(P_{\gamma_k}-\bar{P}_{\gamma_k})\circ P_{\Gamma_n-\Gamma_k}f(x)\right| \\ & + \left|\bar{P}_{\gamma_1}\circ\cdots\circ\bar{P}_{\gamma_{n-1}}\circ(P_{\gamma_n}-\bar{P}_{\gamma_n})\circ f(x)\right|.\end{aligned}$$

The three terms on the right hand side of the inequality denoted from the left to the right (a), (b) and (c) respectively, contain three different types of *weak errors*: respectively, the "ergodic term" (a) where the exponential contraction of the semi-group can be exploited, the "semi-regular weak error term" (b), where the smoothing effect of the operator $P_{\Gamma_n-\Gamma_k}$ ($\Gamma_n-\Gamma_k\in[\gamma_n,T]$) helps us in controlling the weak error related to the function $x\mapsto P_{\Gamma_n-\Gamma_k}f(x)$ and finally, the "less smooth term" (c) where the weak error applies directly on f. The control of each term then relies on quite different arguments.

– Term (c): first, it follows from Lemma 3.4(b) with $p=2$ and $\bar{\gamma}=\|\boldsymbol{\gamma}\|$ that

$$\left|P_{\gamma_n}f(x)-\bar{P}_{\gamma_n}f(x)\right| \leqslant [f]_{\rm Lip}\|X^x_{\gamma_n}-\bar{X}^x_{\gamma_n}\|_2 \leqslant [f]_{\rm Lip}\gamma_n\Psi_1(x),$$

where $\Psi_1(x) = C_{d,b,\sigma,\|\boldsymbol{\gamma}\|}(1+|b(x)|+\|\sigma(x)\|) \leqslant C_{V,d,b,\sigma,\|\boldsymbol{\gamma}\|}\cdot V(x)$ with $C = C_{V,d,b,\sigma,\|\boldsymbol{\gamma}\|}>0$.

Consequently, it follows from Proposition A.1(a)

$$|(c)| \leqslant C\,[f]_{\rm Lip}\gamma_n\mathbb{E}\,V(\bar{X}^x_{\Gamma_{n-1}}) \leqslant C\,[f]_{\rm Lip}\gamma_n\sup_{k\geqslant 0}\mathbb{E}\,V(\bar{X}^x_{\Gamma_k}) \leqslant C\,[f]_{\rm Lip}\gamma_n V(x)$$

where $C_{V,d,b,\sigma,\boldsymbol{\gamma}}>0$ (may vary in the above inequalities).

– Term (b). Let $k\in\{N(\Gamma_n-T)+1, n-1\}$. It follows from Proposition 3.6 applied with $t=\Gamma_n-\Gamma_k$ and $\bar{\gamma}=\|\boldsymbol{\gamma}\|$ so that $\gamma_k\leqslant\bar{\gamma}$ that

$$\left|P_{\gamma_k}\circ P_{\Gamma_n-\Gamma_k}f(x)-\bar{P}_{\gamma_k}\circ P_{\Gamma_n-\Gamma_k}f(x)\right| \leqslant C_{b,\sigma,\|\boldsymbol{\gamma}\|}\,[f]_{\rm Lip}\frac{\gamma_k^2}{\Gamma_n-\Gamma_k}V^2(x)$$

[6]Once again, we assume w.l.g. that $\Gamma_n\geqslant T$ keeping in mind that if $n\in\{1,\dots,N(T)\}$, $|\mathbb{E}\,f(X^x_{\Gamma_n})-\mathbb{E}\,f(\bar{X}^x_{\Gamma_n})|$ can be artificially controlled (for instance) by $C[f]_{\rm Lip}\gamma^{-1}_{N(T)}\gamma_n$ with $C = 2\big(1+\sup_{n\geqslant1}\mathbb{E}[|X^x_{\Gamma_n}|]+\sup_{n\geqslant1}\mathbb{E}[|\bar{X}^x_{\Gamma_n}|]\big)$

which in turn implies (up to an update of the real constant $C_{b,\sigma,\|\boldsymbol{\gamma}\|}$)

$$
|(b)| \leqslant C_{b,\sigma,\|\boldsymbol{\gamma}\|}\, V^2(x) \sum_{k=N(\Gamma_n-T)+1}^{n-1} \frac{\gamma_k^2}{\Gamma_n-\Gamma_k}.
$$

– Term (a). We adopt a strategy very similar to that of the proof of Theorem 2.2(b), namely we get a variant of (4.35) where $\|f\|_{\sup}$ is replaced by $[f]_{\rm Lip}$ i.e., for n large enough,

$$
\left|P_{\gamma_n}\circ P_{\Gamma_n-\Gamma_k} f(x) - \bar{P}_{\gamma_n}\circ P_{\Gamma_n-\Gamma_k} f(x)\right| \leqslant c e^{-\rho(\Gamma_n-\Gamma_k-T/2)}[f]_{\rm Lip}\mathbb{E}\left|X^{\Xi^x_k}_{\frac{T}{2}} - X^{\bar{\Xi}^x_k}_{\frac{T}{2}}\right|
$$

owing to $(\mathbf{H}_{\mathcal{W}_1})$ applied at time $\Gamma_n-\Gamma_k-T/2$ where $\Xi^x_k \stackrel{d}{=} X^{X^x_{\gamma_k}}_{\frac{T}{2}}$ and $\bar{\Xi}^x_k \stackrel{d}{=} X^{\bar{X}^x_{\gamma_k}}_{\frac{T}{2}}$. Finally, still following the lines of the proof of Theorem 2.2(b), we obtain for a constant $C_{b,\sigma,\underline{\sigma}_0,T,\boldsymbol{\gamma}}>0$

$$
\left|\bar{P}_{\gamma_1}\circ\cdots\circ\bar{P}_{\gamma_{k-1}}\circ(P_{\gamma_k}-\bar{P}_{\gamma_k})\circ P_{\Gamma_n-\Gamma_k} f(x)\right| \leqslant C_{b,\sigma,\underline{\sigma}_0,T,\boldsymbol{\gamma}} e^{-\rho(\Gamma_n-\Gamma_k)}[f]_{\rm Lip}\gamma_k^2 V^2(x).
$$

On the other hand, applying $(\mathbf{H}_{\mathcal{W}_1})$, we have for the diffusion

$$
\begin{aligned}
\mathcal{W}_1\big([X^x_{\Gamma_n}],\nu\big) &= \int_{\mathbb{R}^d}\nu(dy)\mathcal{W}_1\big([X^x_{\Gamma_n}],[X^y_{\Gamma_n}]\big)\\
&\leqslant c\,\nu(|x-\cdot|)e^{-\rho\Gamma_n} \leqslant c\big(|x|+\nu(|\cdot|)\big)e^{-\rho\Gamma_n}
\end{aligned}
$$

so that we obtain:

$$
\mathcal{W}_1([\bar{X}^x_{\Gamma_n}],\nu) \leqslant C_{b,\sigma,V,T,\|\boldsymbol{\gamma}\|}\vartheta(x)\left(e^{-\rho\Gamma_n} + \sum_{k=1}^{N(\Gamma_n-T)}\gamma_k^2 e^{-\rho(\Gamma_n-\Gamma_k)} + \sum_{k=N(\Gamma_n-T)+1}^{n-1}\frac{\gamma_k^2}{\Gamma_n-\Gamma_k}+\gamma_n\right)
$$

with $\vartheta(x)=(|x|+1)\vee V^2(x)$. As $\varpi<\rho$, $e^{-\rho\Gamma_n}+\sum_{1\leqslant k\leqslant N(\Gamma_n-T)}\gamma_k^2 e^{-\rho(\Gamma_n-\Gamma_k)}\leqslant C\,\gamma_n$ like in the proof of claim (b), owing to Lemma A.3(ii)-(iii). As for the last sum, one proceeds as follows: still using $\varpi<+\infty$, one checks that $\sup_{n\geqslant 1}\frac{\gamma_n}{\gamma_{n+1}}<+\infty$ so that, for $k\leqslant n-1$,

$$
\frac{\Gamma_n-\Gamma_{k-1}}{\Gamma_n-\Gamma_k} = \frac{\Gamma_n-\Gamma_k+\gamma_k}{\Gamma_n-\Gamma_k} = 1+\frac{\gamma_k}{\Gamma_n-\Gamma_k}\leqslant 1+\frac{\gamma_k}{\gamma_{k+1}}\leqslant C_{\boldsymbol{\gamma}}.
$$

Consequently (still with $C_{\boldsymbol{\gamma}}>0$ a real constant that may vary from line to line),

$$
\begin{aligned}
\sum_{k=N(\Gamma_n-T)+1}^{n-1}\frac{\gamma_k^2}{\Gamma_n-\Gamma_k} &\leqslant C_{\boldsymbol{\gamma}}\sum_{k=N(\Gamma_n-T)+1}^{n-1}\frac{\gamma_k^2}{\Gamma_n-\Gamma_{k-1}}\\
&\leqslant C_{\boldsymbol{\gamma}}\cdot\gamma_{N(\Gamma_n-T)}\int_{\Gamma_{N(\Gamma_n-T)}}^{\Gamma_{n-1}}\frac{1}{\Gamma_n-t}dt\\
&\leqslant C_{\boldsymbol{\gamma}}\cdot\gamma_n\log\left(\tfrac{\Gamma_n-\Gamma_{N(\Gamma_n-T)}}{\gamma_n}\right)\leqslant C_{\boldsymbol{\gamma}}\gamma_n\log\left(\tfrac{T+\|\boldsymbol{\gamma}\|_\infty}{\gamma_n}\right)
\end{aligned}
\tag{4.38}
$$

where we used in the second line that $(\gamma_n)_{n\geqslant 1}$ is non-increasing and a classical comparison argument between sums and integrals and, in the third line, Lemma A.3(ii). This completes the proof.

4.3 Proof of Theorem 2.3

We will follow the global structure of the proof of Theorem 2.2(a) for both distances. However, taking advantage of the fact that when σ is constant the distributions of the diffusions and the Euler scheme on finite horizon T are equivalent, we will replace Theorem 3.7 by a more straightforward and less technical Pinsker's inequality, as developed in the next proposition.

Proposition 4.1. *If b is Lipschitz continuous, $\sigma(x) = \sigma \in GL(d,\mathbb{R})$ is constant (so that it satisfies $(\mathcal{E}\ell)_{\underline{\sigma}_0^2}$). Then there exists a real constant $\kappa_\sigma > 0$ solution to $ue^u = \frac{\underline{\sigma}_0}{\|\sigma\|}$ and a real constant $C = C_{b,\sigma}$ such that, for every $\gamma \in \big(0, \frac{\kappa_\sigma}{[b]_{\rm Lip}}\big)$ and every bounded Borel function $f : \mathbb{R}^d \to \mathbb{R}$,*

$$\big|\mathbb{E}_{\mathbb{P}}\, f(X_\gamma^x) - \mathbb{E}_{\mathbb{P}}\, f(\bar{X}_\gamma^{\gamma,x})\big| \leqslant \|f\|_{\sup} C \cdot V(x)^{1/2}\gamma.$$

Proof. Set

$$\mathbb{Q}_\gamma = \mathcal{E}\Big(-\int_0^{\cdot} \sigma^{-1}(b(X_s^x) - b(x))dW_s\Big)_\gamma \cdot \mathbb{P} = L_\gamma \cdot \mathbb{P}$$

where $\mathcal{E}$ denotes the Doléans exponential.
First we prove that $\mathbb{Q}_\gamma$ is a true probability measure.

$$\begin{aligned} |X_t^x - x| &\leqslant \int_0^t |b(X_s^x) - b(x)|ds + |b(x)t + \sigma W_t| \\ &\leqslant [b]_{\rm Lip} \int_0^t |X_s^x - x|ds + |b(x)|t + \sigma W_t^\star, \end{aligned}$$

where $W_t^\star = \sup_{0\leqslant s\leqslant t}|W_s|$. By Gronwall's lemma,

$$|X_t^x - x| \leqslant e^{[b]_{\rm Lip}t}\big(|b(x)|t + \sigma W_t^\star\big)$$

so that

$$\begin{aligned} \int_0^\gamma |X_t^x - x|^2 dt &\leqslant e^{2[b]_{\rm Lip}\gamma} \int_0^\gamma \big(|b(x)|t + \sigma W_t^\star\big)^2 dt \\ &\leqslant e^{2[b]_{\rm Lip}\gamma}\Big(|b(x)|^2(1+1/\eta)\frac{\gamma^3}{3} + \|\sigma\|^2(1+\eta)\gamma(W_\gamma^\star)^2\Big), \end{aligned}$$

where the second inequality holds for any $\eta > 0$. By Novikov's criterion (see $e.g.$ [RY99]), it easily follows that $\mathbb{Q}_\gamma$is a probability measure if for some small enough $\eta > 0$,

$$\mathbb{E}\,\exp\Big(\tfrac{1}{2}\tfrac{[b]_{\rm Lip}^2}{\underline{\sigma}_0^2} e^{2[b]_{\rm Lip}\gamma}\|\sigma\|^2(1+\eta)\gamma(W^\star)_\gamma^2\Big) < +\infty.$$

The Brownian motions $W^1, \cdots, W^d$ being independent and $(W_\gamma^\star)^2 \leqslant \big((W^1)_\gamma^\star\big)^2 + \cdots + \big((W^d)_\gamma^\star\big)^2$ it is suffices (in fact equivalent) to show that

$$\mathbb{E}\exp\Big(\tfrac{1}{2}[b]_{\rm Lip}^2 e^{2[b]_{\rm Lip}\gamma}\tfrac{\|\sigma\|^2}{\underline{\sigma}_0^2}(1+\eta)\gamma((W^1)_\gamma^\star)^2\Big) < +\infty.$$

Now, it is classical background that

$$\mathbb{E}\, e^{\lambda(W^\star)_t^2} \leqslant \mathbb{E}\, e^{\lambda(\overline{W}_t)^2} + \mathbb{E}\, e^{\lambda(\overline{-W})_t^2}$$

where $\overline{B}_t = \sup_{0\leqslant s\leqslant t} B_s$. As $-W$ is a standard Brownian motion and $\overline{W}_t \stackrel{\mathcal{L}}{\sim} \sqrt{t}|B_1|$, we derive that, if $\lambda t < \frac{1}{2}$, then

$$\mathbb{E}\, e^{\lambda(W^\star)_t^2} \leqslant 2\,\mathbb{E}\, e^{\lambda t B_1^2} = \frac{2}{\sqrt{1-2\lambda t}} < +\infty$$

Consequently, the above measure $\mathbb{Q}_\gamma$ is a probability if

$$[b]_{\rm Lip}^2\gamma^2 e^{2[b]_{\rm Lip}\gamma} < \Big(\frac{\underline{\sigma}_0}{\|\sigma\|}\Big)^2,$$

which is equivalent to

$$0 < \gamma < \frac{\kappa_\sigma}{[b]_{\rm Lip}},$$

where κ_σ is the unique solution to $u\, e^u = \frac{\underline{\sigma}_0}{\|\sigma\|}$. By Girsanov's Theorem

$$B_t = W_t + \int_0^t \sigma^{-1}\big(b(X_s^x) - b(x)\big)ds \quad \text{is a } \mathbb{Q}_\gamma\text{-M.B.S.}$$

so that, under $\mathbb{Q}_\gamma$,

$$X_t^x = b(x)t + \sigma B_t,\ t\in [0,\gamma].$$

Hence, for every bounded Borel function $f : \mathbb{R}^d \to \mathbb{R}$,

$$\mathbb{E}_{\mathbb{P}}\, f(X_\gamma^x) = \mathbb{E}_{\mathbb{Q}_\gamma} L_\gamma^{-1} f(x+\gamma b(x)+\sigma B_\gamma) \quad \text{and} \quad \mathbb{E}_{\mathbb{P}}\, f(\bar{X}_\gamma^{\gamma,x}) = \mathbb{E}_{\mathbb{Q}_\gamma} f(x+\gamma b(x)+\sigma B_\gamma).$$

It follows from Pinsker's inequality (see [CBL06]) that

$$\begin{aligned} d_{TV}(\mathbb{P},\mathbb{Q}_\gamma)^2 &\leqslant 2\int_\Omega \log\big(L_\gamma^{-1}\big)L_\gamma^{-1}d\mathbb{Q}_\gamma = -2\int_\Omega \log L_\gamma d\mathbb{P} \\ &= 2\mathbb{E}\left[\int_0^\gamma \big(\sigma^{-1}(b(X_s^x)-b(x))|dW_s\big) + \int_0^\gamma \big|\sigma^{-1}(b(X_s^x)-b(x))\big|^2 ds\right] \\ &\leqslant \frac{[b]_{\rm Lip}^2}{\underline{\sigma}_0^2}\int_0^\gamma \mathbb{E}_{\mathbb{P}}|X_s^x - x|^2 ds. \end{aligned}$$

It follows from Lemma 3.4 (a) (see (3.22)) and the fact that $S_2(x) = (1+|b(x)|+\|\sigma\|)$ that for $s\in (0,\kappa_\sigma/[b]_{\rm Lip})$

$$\mathbb{E}_{\mathbb{P}}\,|X_s^x - x|^2 \leqslant C'_{b,\|\sigma\|_{\sup}}\Big(|b(x)|^2+1\Big)s.$$

Hence

$$d_{TV}(\mathbb{P},\mathbb{Q}_\gamma)^2 \leqslant C'_{b,\|\sigma\|_{\sup}}\frac{[b]_{\rm Lip}^2}{\underline{\sigma}_0^2}\big(|b(x)|^2+1\big)\frac{\gamma^2}{2}$$

so that, for $\gamma\in (0,\kappa_\sigma/[b]_{\rm Lip})$,

$$d_{TV}(\mathbb{P},\mathbb{Q}_\gamma) \leqslant C_{\underline{\sigma}_0,b,\|\sigma\|_{\sup},V}V(x)^{1/2}\gamma$$

Finally, for a bounded Borel function f

$$\big|\mathbb{E}_{\mathbb{P}}\, f(X_\gamma^x) - \mathbb{E}_{\mathbb{P}}\, f(\bar{X}_\gamma^{\gamma,x})\big| \leqslant \|f\|_{\sup} d_{TV}(\mathbb{P},\mathbb{Q}_\gamma) \leqslant \|f\|_{\sup} C''_{\underline{\sigma}_0,b,\|\sigma\|_{\sup},V}V(x)^{1/2}\gamma. \qquad \square$$

Remark 4.1. In fact we could avoid to call upon Pinsker's inequality by noting that

$$\mathbb{E}_{\mathbb{Q}_\gamma}|L_\gamma^{-1}-1| = \mathbb{E}_{\mathbb{P}}|L_\gamma - 1| = \mathbb{E}\Big|\int_0^\gamma L_s\sigma^{-1}\big(b(X_s^x)-b(x)\big)dW_s\Big| \leqslant \frac{[b]_{\rm Lip}}{\underline{\sigma}_0}\left(\int_0^\gamma \|L_s\|_4^2\|X_s^x - x\|_4^2 ds\right)^{1/2}.$$

Then the conclusion follows from Lemma 3.4 applied with $p=4$ (after having classically controlled $\sup_{0\leqslant s\leqslant \frac{\kappa_\sigma}{[b]_{\rm Lip}}}\|L_s\|_4$). The resulting constants are (probably) less sharp.

Proof of Theorem 2.3. (Wasserstein distance). Let $T > 0$ be fixed and let n be such that $\Gamma_n > T$. Like in the proof of Theorem 2.2(a) (see the footnote), we may assume that n is large enough so that $\Gamma_n > T$. Then we write for a Lipschitz continuous function $f : \mathbb{R}^d \to \mathbb{R}$

$$\begin{aligned}\big|\mathbb{E}\, f(X^x_{\Gamma_n}) - \mathbb{E}\, f(\bar{X}^x_{\Gamma_n})\big| \leqslant \sum_{k=1}^{N(\Gamma_n - T)} \big|\bar{P}_{\gamma_1} \circ \cdots \circ \bar{P}_{\gamma_{k-1}} \circ (P_{\gamma_k} - \bar{P}_{\gamma_k}) \circ P_{\Gamma_n - \Gamma_k} f(x)\big| \\ + \big|\bar{P}_{\gamma_1} \circ \cdots \circ \bar{P}_{\gamma_{N(\Gamma_n-T)}} \circ \big(P_{\Gamma_n - \Gamma_{N(\Gamma_n - T)}} - \bar{P}_{\gamma_{N(\Gamma_n-T)+1}} \circ \cdots \circ \bar{P}_{\gamma_n}\big) f(x)\big)\big|.\end{aligned}$$

STEP 1. First we note that

$$\begin{aligned}&\Big|\bar{P}_{\gamma_1} \circ \cdots \circ \bar{P}_{\gamma_{N(\Gamma_n-T)}} \circ \big(P_{\Gamma_n - \Gamma_{N(\Gamma_n - T)}} - \bar{P}_{\gamma_{N(\Gamma_n-T)+1}} \circ \cdots \circ \bar{P}_{\Gamma_n - \Gamma_{n-1}}\big) f(x)\Big| \\ &\quad = \Big|\mathbb{E}\Big[f\Big(X^{\bar{X}^x_{\Gamma_{N(\Gamma_n-T)}}}_{\Gamma_n - \Gamma_{N(\Gamma_n-T)}}\Big) - f\Big(\bar{X}^{\bar{X}^x_{\Gamma_{N(\Gamma_n-T)}}}_{\Gamma_n - \Gamma_{N(\Gamma_n-T)}}\Big)\Big]\Big| \\ &\quad \leqslant [f]_{\rm Lip} \int \big|X^{\xi}_{\Gamma_n - \Gamma_{N(\Gamma_n-T)}} - \bar{X}^{\xi}_{\Gamma_n - \Gamma_{N(\Gamma_n-T)}}\big| \mathbb{P}_{\bar{X}^x_{\Gamma_n - \Gamma_{N(\Gamma_n-T)}}}(d\xi) \\ &\quad \leqslant [f]_{\rm Lip} C_{T+\|\gamma\|_{\sup}} \gamma_{N(\Gamma_n-T)+1} \int V^{1/2}(\xi) \mathbb{P}_{\bar{X}^x_{\Gamma_n - \Gamma_{N(\Gamma_n-T)}}}(d\xi) \\ &\quad \leqslant [f]_{\rm Lip} C_{T+\|\gamma\|_{\sup}} \gamma_{N(\Gamma_n-T)} \mathbb{E}\, V^{1/2}\big(\bar{X}^x_{\Gamma_{N(\Gamma_n-T)}}\big) \\ &\quad \leqslant [f]_{\rm Lip} C_{T+\|\gamma\|_{\sup}, \gamma} \cdot \gamma_{N(\Gamma_n-T)} V^{1/2}(x),\end{aligned}$$

where we used Proposition A.1 (a) in the last inequality and, in the second one, the fact that the Euler scheme with decreasing step is of order 1 when σ is constant. This expected result follows by mimicking the proof of the convergence rate of the Euler scheme with decreasing step from in [PP14] adapted by taking advantage of the one step strong error from Lemma 3.4(c) with $p = 2$ ([7]). We know from Lemma A.3(ii) that $\limsup_n \frac{\gamma_{\Gamma_{N(\Gamma_n-T)+1}}}{\gamma_n} \leqslant \limsup_n \frac{\gamma_{\Gamma_{N(\Gamma_n-T)}}}{\gamma_n} < +\infty$ so that finally

$$\big|\bar{P}_{\gamma_1} \circ \cdots \circ \bar{P}_{\gamma_{N(\Gamma_n-T)}} \circ \big(P_{\Gamma_n - \Gamma_{N(\Gamma_n-T)}} f(x) - \bar{P}_{\gamma_{N(\Gamma_n-T)+1}} \circ \cdots \circ \bar{P}_{\Gamma_n - \Gamma_{n-1}} f(x)\big)\big| \leqslant [f]_{\rm Lip} C_{T,\gamma} \gamma_n V^{1/2}(x).$$

STEP 2. Let $k \in \{1, \ldots, N(\Gamma_n - T)\}$. Using that $\Gamma_n - \Gamma_k \geqslant T$ and adapting the treatment of term (A) in the proof of Theorem 2.2(b), we have

$$\big|\bar{P}_{\gamma_1} \circ \cdots \circ \bar{P}_{\gamma_{k-1}} \circ (P_{\gamma_k} - \bar{P}_{\gamma_k}) \circ P_{\Gamma_n - \Gamma_k} f(x)\big| \leqslant C_{b,\sigma,\underline{\sigma}_0,T,\gamma} e^{-\rho(\Gamma_n - \Gamma_k)} [f]_{\rm Lip} \gamma_k^2 V^2(x).$$

Thus, it follows from the Kantorovich-Rubinstein representation of the $\mathcal{W}_1$-distance

$$\mathcal{W}_1([\bar{X}^x_{\Gamma_n}], \nu) \leqslant C_{b,\sigma,T,\gamma} \cdot \vartheta(x) \left(e^{-\rho\Gamma_n} + \gamma_n + \sum_{k=1}^{N(\Gamma_n-T)} \gamma_k^2 e^{-\rho(\Gamma_n - \Gamma_k)}\right)$$

with $\vartheta(x) = \big(V^2(x) \vee (|x| + 1)\big)$ and one concludes that, since $\rho < \varpi$,

$$\mathcal{W}_1([\bar{X}^x_{\Gamma_n}], \nu) \leqslant C_{b,\sigma,T,\gamma} \cdot \gamma_n \vartheta(x).$$

[7] Thus, one shows for the Euler scheme with decreasing step, say δ_n with $t_n := \delta_1 + \cdots + \delta_n \to +\infty$, that for every $T > 0$, there exists a real constant (not depending on (δ_n)) such that

$$\Big\| \max_{k : t_k \leqslant T} |X^x_{t_k} - \bar{X}^x_{t_k}| \Big\|_2 \leqslant C_{b,\sigma,T}(1 + |b(x)| + \|\sigma(x)\|)\delta_1 \leqslant C_{b,\sigma,T} V^{1/2}(x) \delta_1.$$

Proof of Theorem 2.3 (TV distance, first TV-bound). First note that $(\mathcal{E}\ell)_{\underline{\sigma}_0^2}$ is satisfied so that $(\mathbf{H_{TV}})$ holds by Proposition 2.1. Then, we will use (3.26) from Proposition 3.5 in its less sharp form

$$
\begin{aligned}
|\mathbb{E}\,[g(\bar{X}_\gamma^x)] - \mathbb{E}\,[g(X_\gamma^x)]| \leqslant \gamma^2 \max\big(\|\nabla g\|_\infty \vee \|D^2 g\|_\infty\big) S_{d,b,\sigma,\bar{\gamma}}(x)^2 \\
+ \gamma^{5/2} \max\big(\|D^2 g\|_{\sup}, \|D^3 g\|_{\sup}\big) S_{d,b,\sigma,\bar{\gamma}}(x)^3.
\end{aligned}
$$

We rely again on the three-fold decomposition used for the proof of Theorem 2.2(a), this time with $f : \mathbb{R}^d \to \mathbb{R}$ a bounded Borel function.

We still consider $T > \bar{\gamma}$ with $\bar{\gamma} = \|\boldsymbol{\gamma}\|$. First, we may assume w.l.g. that n is large enough so that $\Gamma_n > T$ and $\gamma_n \leqslant \frac{\kappa_\sigma}{2[b]_{\mathrm{Lip}}}$ (coming from the above Proposition 4.1) since for $n \leqslant N(T) \vee n_0$ (with $\gamma_{n_0+1} \leqslant \frac{\kappa_\sigma}{2[b]_{\mathrm{Lip}}} < \gamma_{n_0}$), we may artificially bound $|\mathbb{E}\, f(X_{\Gamma_n}^x) - \mathbb{E}\, f(\bar{X}_{\Gamma_n}^x)|$ by $2\|f\|_{\sup}\gamma_{N(t)\vee n_0}^{-1}\gamma_n$. Then we may apply Proposition 4.1 and Lemma 3.4 respectively with steps γ_n.

Term (a). Let $k \in \{1, \ldots, N(\Gamma_n - T\}$. The proof used in Theorem 2.2(b) with σ non constant for term (A) still works here without modification (see in particular (4.36)): it follows from $(\mathbf{H}_{\mathcal{W}_1})$ (which implies $(\mathbf{H_{TV}})$ with $t_0 = \gamma_1$ by Proposition 2.1) that

$$
\big|P_{\gamma_k} \circ P_{\Gamma_n - \Gamma_k} f(x) - \bar{P}_{\gamma_k} \circ P_{\Gamma_n - \Gamma_k} f(x)\big| \leqslant C_{b,\sigma,T} \|f\|_{\sup} \gamma_n^2\, e^{-\rho(\Gamma_n - \Gamma_k)} V^2(x)
$$

and, as $\varpi < \rho$, one still has $\sum_{1 \leqslant k \leqslant N(\Gamma_n - T)} \gamma_k^2 e^{-\rho(\Gamma_n - \Gamma_k)} \leqslant C_{\boldsymbol{\gamma}} \cdot \gamma_n$ which yields

$$
|(a)| \leqslant C_{b,\sigma,T,\boldsymbol{\gamma}} \gamma_n \|f\|_{\sup} V^2(x).
$$

Term (b). Let $k \in \{N(\Gamma_n - T) + 1, \ldots, n-1\}$. Applying Proposition 3.5(b) to $g = P_{\Gamma_n - \Gamma_k} f$ with the help of BEL identity and the resulting inequalities yields

$$
\big|P_{\gamma_k} \circ P_{\Gamma_n - \Gamma_k} f(x) - \bar{P}_{\gamma_k} \circ P_{\Gamma_n - \Gamma_k} f(x)\big| \leqslant C_{d,b,\sigma,\bar{\gamma}} \cdot \|f\|_{\sup} \Big(V(x) \frac{\gamma_k^2}{\Gamma_n - \Gamma_k} + V^{3/2}(x) \frac{\gamma_k^{5/2}}{(\Gamma_n - \Gamma_k)^{3/2}}\Big).
$$

Now, as in the proof of Theorem 2.2(a), still using that $\varpi < \rho$,

$$
\sum_{k=N(\Gamma_n - T)+1}^{n-1} \frac{\gamma_k^2}{\Gamma_n - \Gamma_k} \leqslant C_{\boldsymbol{\gamma}} \cdot \gamma_n \log\Big(\tfrac{T + \|\boldsymbol{\gamma}\|}{\gamma_n}\Big)
$$

and, proceeding likewise

$$
\sum_{k=N(\Gamma_n - T)+1}^{n-1} \frac{\gamma_k^{5/2}}{(\Gamma_n - \Gamma_k)^{3/2}} \leqslant C_{\boldsymbol{\gamma}} \cdot \gamma_{N(\Gamma_n - T)}^{3/2} \int_{\Gamma_{N(\Gamma_n - T)}}^{\Gamma_{n-1}} \frac{dt}{(\Gamma_n - t)^{3/2}} \leqslant C_{\boldsymbol{\gamma}} \gamma_{N(\Gamma_n - T)}^{3/2} \gamma_n^{-1/2}.
$$

It follows from Lemma A.3(ii) that $\gamma_{N(\Gamma_n - T)+1}^{3/2} \leqslant C_{\boldsymbol{\gamma},T} \cdot \gamma_n^{3/2}$ so that, still using Proposition A.1(a),

$$
|(b)| \leqslant C_{b,\sigma,\boldsymbol{\gamma},T} \cdot \gamma_n.
$$

Term (c). It follows from the former Proposition 4.1 that

$$
\big|P_{\gamma_n} f(x) - \bar{P}_{\gamma_n} f(x)\big| = \big|\mathbb{E}\, f(X_{\gamma_n}^x) - \mathbb{E}\, f(\bar{X}_{\gamma_n}^x)\big| \leqslant C_{b,\sigma} \|f\|_{\sup} \gamma_n V^{1/2}(x).
$$

One concludes as in the multiplicative setting.

Proof of Theorem 2.3 (TV distance, second TV-bound). Assume now that $T > 2\bar{\gamma}$ (still with $\bar{\gamma} = \|\boldsymbol{\gamma}\|$). In addition to the former constraints on γ_n, we may assume w.l.g. in this specific setting that $n \geqslant n_0$ where $\gamma_{N(\Gamma_{n_0}-2T)} < \frac{1}{2d\|\nabla b\|_\infty}$. We rely now on a four fold decomposition

$$\begin{aligned}
\big|\mathbb{E}\, f(X^x_{\Gamma_n}) - \mathbb{E}\, f(\bar{X}^x_{\Gamma_n})\big| \leqslant & \sum_{k=1}^{N(\Gamma_n-T)} \big|\bar{P}_{\gamma_1} \circ \cdots \circ \bar{P}_{\gamma_{k-1}} \circ (P_{\gamma_k} - \bar{P}_{\gamma_k}) \circ P_{\Gamma_n-\Gamma_k} f(x)\big| \\
& + \sum_{k=N(\Gamma_n-T)+1}^{n-1} \big|\bar{P}_{\gamma_1} \circ \cdots \circ \bar{P}_{\gamma_{k-1}} \circ \Big((P_{\gamma_k} - \bar{P}_{\gamma_k}) \circ P_{\Gamma_n-\Gamma_k} f(x) - \tfrac{\gamma_k^2}{2}\mathfrak{T}(P_{\Gamma_n-\Gamma_k} f, b, \sigma)\Big)\big| \\
& + \tfrac{1}{2}\left|\sum_{k=N(\Gamma_n-T)+1}^{n-1} \gamma_k^2 \bar{P}_{\gamma_1} \circ \cdots \circ \bar{P}_{\gamma_{k-1}} \mathfrak{T}(P_{\Gamma_n-\Gamma_k} f, b, \sigma)(x)\right| \\
& + \big|\bar{P}_{\gamma_1} \circ \cdots \circ \bar{P}_{\gamma_{n-1}} \circ (P_{\gamma_n} - \bar{P}_{\gamma_n}) \circ f(x)\big|.
\end{aligned}$$

Let us call the second and third term of the decomposition (b) and (b') respectively, the treatment of other terms being unchanged.

Term (b). Now using the sharp form of (3.26) and using the same tools (inequalities derived from BEL identities), we can upper bound this "corrected " term by

$$C_{d,b,\sigma,\bar{\gamma}} \cdot \|f\|_{\sup} \left(V(x) \frac{\gamma_k^2}{(\Gamma_n - \Gamma_k)^{1/2}} + V^{3/2}(x) \frac{\gamma_k^{5/2}}{(\Gamma_n - \Gamma_k)^{3/2}} \right)$$

and we check that by the usual arguments that

$$\sum_{k=N(\Gamma_n-T)+1}^{n-1} \frac{\gamma_k^2}{(\Gamma_n - \Gamma_k)^{1/2}} \leqslant C_{\|\boldsymbol{\gamma}\|} \gamma_n \int_{\Gamma_{N(\Gamma_n-T)}}^{\Gamma_{n-1}} \frac{dt}{(\Gamma_n - t)^{1/2}} \leqslant C_{\|\boldsymbol{\gamma}\|} \gamma_n. \tag{4.39}$$

Term (b'). First, remark that, for every $k \in \{N(\Gamma_n - T) + 1, \ldots, n-1\}$,

$$\begin{aligned}
\bar{P}_{\gamma_1} \circ \cdots \circ \bar{P}_{\gamma_{k-1}} \mathfrak{T}(P_{\Gamma_n-\Gamma_k} f, b, \sigma)(x) &= \sum_{1 \leqslant i,j \leqslant d} \mathbb{E}\left[\partial^2_{x_i x_j} P_{\Gamma_n-\Gamma_k} f(\bar{X}^x_{\Gamma_{k-1}})\big((\sigma\sigma^*)_{i\cdot} | \nabla b_j(\bar{X}^x_{\Gamma_{k-1}})\big)\right] \\
&= \sum_{1 \leqslant i,j,\ell \leqslant d} (\sigma\sigma^*)_{i\ell} \mathbb{E}\left[\Upsilon_{i,j,\ell,k}(\bar{X}^x_{\Gamma_{N(\Gamma_n-2T)}})\right] \quad \text{with} \quad \Upsilon_{i,j,\ell,k}(x) = \mathbb{E}_x[\partial_{x_i} f_j(\bar{X}^{\tilde{\boldsymbol{\gamma}},x}_{t_{k-1}}) \partial_\ell b_j(\bar{X}^{\tilde{\boldsymbol{\gamma}},x}_{t_{k-1}})],
\end{aligned}$$

where $\tilde{\gamma}_\ell = \Gamma_{N(\Gamma_n-2T)+\ell} - \Gamma_{N(\Gamma_n-2T)+\ell-1}$, $\ell \geqslant 1$, $\bar{X}^{\tilde{\boldsymbol{\gamma}},x}$ is the Euler scheme with time step sequence $\tilde{\boldsymbol{\gamma}}$, $t_{k-1} = \Gamma_{k-1} - \Gamma_{N(\Gamma_n-2T)} = \widetilde{\Gamma}_{k-1-N(\Gamma_n-2T)}$ and $f_j = \partial_{x_j} P_{\Gamma_n-\Gamma_k} f$. The next step is to perform an integration by parts using Malliavin calculus for $\bar{X}^{\tilde{\boldsymbol{\gamma}},x}$ using the "toolbox" developed in Appendix B for the TV-convergence with varying σ, but taking into account that now the tangent process of the scheme is $GL_d(\mathbb{R})$-valued without any truncation. More precisely, with the notations of Proposition B.3, the tangent process $(\bar{Y}_t)_{t \geqslant 0}$ of the (continuous-time version of) $\bar{X}^{\tilde{\boldsymbol{\gamma}},x}$ reads $\tilde{Y}_0^{(x)} = I_d$ and $\tilde{Y}_t^{(x)} = (I_d + (t - \widetilde{\Gamma}_{\ell-1})\nabla b(\bar{X}^{\tilde{\boldsymbol{\gamma}},x}_{\widetilde{\Gamma}_{\ell-1}}))\tilde{Y}^{(x)}_{\widetilde{\Gamma}_{\ell-1}}$ for any $t \in [\widetilde{\Gamma}_{\ell-1}, \widetilde{\Gamma}_\ell]$. Hence, as $\tilde{\gamma}_1 \leqslant \gamma_{N(\Gamma_{n_0}-2T)} < \frac{1}{2d\|\nabla b\|_\infty}$, for any $\Theta > 0$, $\inf_{x \in \mathbb{R}^d, t \in [0,\Theta]} \det(\bar{Y}_t^{(x)})$ is lower-bounded by a positive deterministic constant. Applying this with $\Theta = 2T + \bar{\gamma}$ and noting that $T/2 \leqslant t_k \leqslant 2T + \bar{\gamma}$ for every $k \in \{N(\Gamma_n - T)+1, \ldots, n-1\}$ for large enough n, one checks that the (determinant of the) Malliavin covariance of $\bar{X}^{\tilde{\boldsymbol{\gamma}},x}_{t_{k-1}}$ (see Proposition B.3 for similar computations) is bounded from below by a positive constant $\kappa_{b,\sigma}$ only depending on $\|\nabla b\|_{\sup}$, $\underline{\sigma}_0^2$ and T. This allows us to apply (B.62) (which comes from Lemma 2.4(i) of [BCP20]) with $f = f_j$, $\bar{F} = \bar{X}^{\tilde{\boldsymbol{\gamma}},x}_{t_{k-1}}$, $G = \partial_\ell b_j(\bar{X}^{\tilde{\boldsymbol{\gamma}},x}_{t_{k-1}})$ and $|\alpha| = 1$. With the notation introduced in Section B.2, this leads to

$$\big|\mathbb{E}_x[\partial_{x_i} f_j(\bar{X}^{\tilde{\boldsymbol{\gamma}},x}_{t_{k-1}}) \partial_\ell b_j(\bar{X}^{\tilde{\boldsymbol{\gamma}},x}_{t_{k-1}})]\big| \leqslant C\|f_j\|_\infty \Big|\mathbb{E}\Big[(1 + |\bar{X}^{\tilde{\boldsymbol{\gamma}},x}_{t_{k-1}}|^{4d-2}_{1,2})(|\bar{X}^x_{t_k}|_{1,2} + |L\bar{X}^{\tilde{\boldsymbol{\gamma}},x}_{t_{k-1}}|_1)|\partial_\ell b_j(\bar{X}^{\tilde{\boldsymbol{\gamma}},x}_{t_{k-1}})|_1\Big]\Big|.$$

By Proposition 3.2(a), $\|f_j\|_\infty \leqslant C(\Gamma_n-\Gamma_k)^{-\frac{1}{2}}\|f\|_{\sup}$. By (B.60) and Proposition B.3(ii), $\mathbb{E}\,[|\bar{X}^{\tilde{\gamma},x}_{t_{k-1}}|^p_{1,2}] \leqslant C_{p,T}$ for any $p>0$, where $C_{p,T}$ does not depend on x and k. As well, using that $\partial_\ell b_j$ is bounded with bounded partial derivatives, $\|\partial_\ell b_j(\bar{X}^{\tilde{\gamma},x}_{t_{k-1}})|_{1,p} \leqslant C_{p,T}$ where $C_{p,T}$ is again a constant independent of x and k. Finally, by (B.60) and the fact that b is $\mathcal{C}^3$, one checks that for any $p>0$,

$$\|L\bar{X}^{\tilde{\gamma},x}_{t_{k-1}}\|_{1,p} \leqslant C_{p,T}(1+\mathbb{E}[|\bar{X}^{\tilde{\gamma},x}_{t_{k-1}}|^p]^{\frac{1}{p}}) \leqslant C_{p,T}(1+|x|),$$

where in the second line, we used a Gronwall argument. Finally, using Hölder inequality, we deduce that a constant $C_{p,T}$ exists such that

$$\left|\mathbb{E}_x[\partial_{x_i} f_j(\bar{X}^{\tilde{\gamma},x}_{t_{k-1}})\partial_\ell b_j(\bar{X}^{\tilde{\gamma},x}_{t_{k-1}})]\right| \leqslant \frac{C_{p,T}}{\sqrt{\Gamma_n-\Gamma_k}}(1+|x|).$$

If $\liminf_{|x|\to+\infty} V(x)/|x|^r > 0$, we deduce from Proposition A.1(a) and (4.39), that

$$\begin{aligned}|(b')| &\leqslant C_{p,T}\sum_{k=N(\Gamma_n-T)+1}^{n-1}\frac{\gamma_k^2}{\sqrt{\Gamma_n-\Gamma_k}}\sup_{k\geqslant 0}\mathbb{E}[V^{\frac{1}{r}}(\bar{X}^x_{\Gamma_k})]\\ &\leqslant C_{p,T}\sum_{k=N(\Gamma_n-T)+1}^{n-1}\frac{\gamma_k^2}{\sqrt{\Gamma_n-\Gamma_k}}V^{\frac{1}{r}}(x) \leqslant C_{p,T,\gamma,V}\,\gamma_n V^{\frac{1}{r}}(x).\end{aligned}$$

The alternative growth assumption on V can be treated likewise owing to Proposition A.1(b). □

4.4 Proof of Corollary 2.4

The result is a consequence of the following lemma.

Lemma 4.2. *Assumption* $(\mathbf{C}_\alpha)$ *implies that* $(\mathbf{H}_{\mathcal{W}_1})$ *holds with* $\rho=\alpha$. *To be more precise, one has*

$$\forall\, x,\, y \in \mathbb{R}^d,\ \forall\, t\geqslant 0,\quad \mathbb{E}|X^x_t - X^y_t|^2 \leqslant e^{-2\alpha t}|x-y|$$

so that
$$\mathcal{W}_1([X^x_t],[X^y_t]) \leqslant \mathcal{W}_2([X^x_t],[X^y_t]) \leqslant e^{-\alpha t}|x-y|.$$

Proof. It follows from Itô's formula applied to $e^{2\alpha t}|X^x_t - X^y_t|^2$ that this process is a supermartingale starting from $|x-y|^2$ owing to $(\mathbf{C}_\alpha)$. □

4.5 Proof of Corollary 2.5

By Proposition 2.1, it is enough to show that $(\mathbf{H}_{\mathcal{W}_1})$ holds true. When σ is constant, this is a direct consequence of [LW16]. In the multiplicative case, we rely on [Wan20, Theorem 2.6]. Since σ is bounded, we remark that Assumption (2.17) of [Wan20] is true as soon as there exist positive constants K_1, K_2 and R_0 such that for every $x, y \in \mathbb{R}^d$,

$$(b(x)-b(y)|x-y) \leqslant K_1\mathbf{1}_{\{|x-y|\leqslant R_0\}} - K_2|x-y|^2. \tag{4.40}$$

But it is easy to check that this assumption is equivalent to the existence of some $\alpha,\ R>0$ such that

$$\forall (x,y)\in B(0,R)^c,\quad (b(x)-b(y)|x-y) \leqslant -\alpha|x-y|^2. \tag{4.41}$$

Actually, the direct implication is obvious by setting $R=R_0$ and $\alpha=K_2$.

In order to prove the converse, set $R_0 = 4R\Big(1+\frac{[b]_{\rm Lip}}{\alpha}\Big)$. Let $x,\, y\in\mathbb{R}^d$ be such that $|x-y|\geqslant R_0$. If both x and y lie outside $B(0,R)$ (closed Euclidean ball centered at 0 with radius R), then $(b(x)-b(y)\,|\,x-y)\leqslant -\alpha|x-y|^2$. Otherwise, one may assume w.l.g. that $x\in B(0,R)$ and $y\notin B(0,R)$ since $R_0>2R$. Then let $\widetilde{x}=\lambda x+(1-\lambda)y$ be such that $|\tilde{x}|=R$ (i.e the point of the segment $[x,y]$ which intersects the boundary of the ball $B(0,R)$). It is clear that $\lambda\in(0,1]$ and that

$$x-y=\frac{\tilde{x}-y}{\lambda}\quad\text{ and }\quad 1-\lambda=\frac{|x-\tilde{x}|}{|x-y]}\leqslant\frac{2R}{R_0}=\frac{\alpha}{2(\alpha+[b]_{\rm Lip})}.$$

Consequently

$$\begin{aligned}\big(b(x)-b(y)\,|\,x-y\big)&\leqslant\big(b(x)-b(\tilde{x})\,|\,x-y\big)+\frac{\big(b(\tilde{x})-b(y)\,|\,\tilde{x}-y\big)}{\lambda}\\&\leqslant[b]_{\rm Lip}|x-\tilde{x}||x-y|-\tfrac{\alpha}{\lambda}|\tilde{x}-y|^2\\&=-\Big(\alpha\lambda-[b]_{\rm Lip}(1-\lambda)\Big)|x-y|^2\\&=-\Big(\alpha-(1-\lambda)(\alpha+[b]_{\rm Lip})\Big)|x-y|^2\leqslant-\tfrac{\alpha}{2}|x-y|^2.\end{aligned}$$

Finally, (4.40) holds with R_0 defined above, $K_1=[b]_{\rm Lip}R_0^2$ and $K_2=-\frac{\alpha}{2}$.

4.6 Explicit bounds for the Ornstein-Uhlenbeck process

Let us consider the α-confluent centered Ornstein-Uhlenbeck process defined by

$$dX_t=-\alpha X_t dt+\sigma dW_t,\quad X_0=0,$$

where $\sigma>0$. It satisfies $(\mathbf{H}_{\mathcal{W}_1})$ and $(\mathbf{S})$ with $\rho=\alpha$. As $X_t=e^{-\alpha t}\int_0^t e^{\alpha s}dW_s$, one checks that

$$\mathrm{Var}(X_t)=\sigma^2e^{-2\alpha t}\int_0^t e^{2\alpha s}ds=\frac{\sigma^2}{2\alpha}\big(1-e^{-2\alpha t}\big)$$

and its (unique) invariant distribution is given by $\nu=\mathcal{N}\Big(0,\frac{\sigma^2}{2\alpha}\Big)$.

Now, let us consider the Euler scheme with a decreasing step $(\gamma_n)_{n\geqslant1}$ such that $\varpi<\alpha$ and $\sum_n\gamma_n^2<+\infty$. It reads

$$\bar{X}_{\Gamma_{n+1}}=\bar{X}_{\Gamma_n}(1-\alpha\gamma_{n+1})+\sigma\big(W_{\Gamma_{n+1}}-W_{\Gamma_n}\big),\quad X_0=0.$$

The scheme is centered and its variance $\sigma_n^2=\mathrm{Var}(\bar{X}_{\Gamma_n})$ at time Γ_n satisfies $\sigma^2=0$ and

$$\sigma_{n+1}^2=\sigma_n^2(1-\alpha\gamma_{n+1})^2+\sigma^2\gamma_{n+1},\ n\geqslant0.$$

Elementary computations show that,

$$\begin{aligned}\sigma_n^2-\frac{\sigma^2}{2\alpha}&=\frac{\sigma^2}{2}\alpha\left[\prod_{k=1}^n(1-\alpha\gamma_k)^2\right]\sum_{k=1}^n\frac{\gamma_k^2}{\prod_{1\leqslant\ell\leqslant k}(1-\alpha\gamma_\ell)^2}\\&\asymp\int_0^{\Gamma_n}e^{-2\alpha(\Gamma_n-s)}\gamma_{N(s)}ds\geqslant\int_0^{\Gamma_n}e^{-2\alpha(\Gamma_n-s)}\,\gamma_{N(\Gamma_n)}=\frac{1-e^{-2\alpha\Gamma_n}}{2\alpha}\gamma_n\sim\frac{1}{2\alpha}\gamma_n\end{aligned}$$

where, for two sequences (a_n) and (b_n), $a_n\asymp b_n$ means $a_n=O(b_n)$ and $b_n=O(a_n)$ as $n\to+\infty$.

Hence, one checks that, as $\sigma_n\to\sigma$,

$$\mathcal{W}_1\big([\bar{X}_{\Gamma_n}],\nu\big)=|\sigma_n-\sigma/\sqrt{2\alpha}|\mathbb{E}\,|Z|\asymp\gamma_n.$$

As for the total variation distance we rely on the lower bound from [DMR18] for two one dimensional Gaussian distributions (sharing the same mean)

$$\left\|[\bar{X}_{\Gamma_n}] - \nu\right\|_{TV} \geqslant \frac{1}{200} \min\left(1, \left|1 - \frac{\sigma_n^2}{\sigma^2/(2\alpha)}\right|\right) \geqslant c_\alpha \gamma_n$$

for large enough n where $c_\alpha > 0$ so that $\left\|[\bar{X}_{\Gamma_n}] - \nu\right\|_{TV} \asymp \gamma_n$.

References

[BC19] Vlad Bally and Lucia Caramellino. Total variation distance between stochastic polynomials and invariance principles. Ann. Probab., 47(6):3762–3811, 2019.

[BCP20] Vlad Bally, Lucia Caramellino, and Guillaume Poly. Regularization lemmas and convergence in total variation. Pre-print, 2020.

[BGL14] 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.

[Bis84] Jean-Michel Bismut. Large deviations and the Malliavin calculus, volume 45 of Progress in Mathematics. Birkhäuser Boston, Inc., Boston, MA, 1984.

[BJM16] Michel Bonnefont, Aldéric Joulin, and Yutao Ma. Spectral gap for spherically symmetric log-concave probability measures, and beyond. J. Funct. Anal., 270(7):2456–2482, 2016.

[BT96] Vlad Bally and Denis Talay. The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function. Probab. Theory Related Fields, 104(1):43–60, 1996.

[CBL06] Nicolò Cesa-Bianchi and Gábor Lugosi. Prediction, learning, and games. Cambridge University Press, Cambridge, 2006.

[CCDO21] Thomas Cass, Dan Crisan, Paul Dobson, and Michela Ottobre. Long-time behaviour of degenerate diffusions: UFG-type SDEs and time-inhomogeneous hypoelliptic processes. Electron. J. Probab., 26:Paper No. 22, 72, 2021.

[CDO21] D. Crisan, P. Dobson, and M. Ottobre. Uniform in time estimates for the weak error of the Euler method for SDEs and a pathwise approach to derivative estimates for diffusion semigroups. Trans. Amer. Math. Soc., 374(5):3289–3330, 2021.

[Cer00] Sandra Cerrai. Analytic semigroups and degenerate elliptic operators with unbounded coefficients: a probabilistic approach. J. Differential Equations, 166(1):151–174, 2000.

[Dal17] Arnak S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. J. R. Stat. Soc. Ser. B. Stat. Methodol., 79(3):651–676, 2017.

[DKZ12] Federica Dragoni, Vasilis Kontis, and Bogusł aw Zegarliński. Ergodicity of Markov semigroups with Hörmander type generators in infinite dimensions. Potential Anal., 37(3):199–227, 2012.

[DM17] Alain Durmus and Éric Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. Ann. Appl. Probab., 27(3):1551–1587, 2017.

[DM19] Alain Durmus and Éric Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. Bernoulli, 25(4A):2854–2882, 2019.

[DMR18] Luc Devroye, Abbas Mehrabian, and Tommy Reddad. The total variation distance between high-dimensional Gaussians. arXiv e-prints, page arXiv:1810.08693, Oct 2018.

[DMS20] Pierre Del Moral and Sumeetpal S. Singh. A backward Itô-Ventzell formula with an application to stochastic interpolation. C. R. Math. Acad. Sci. Paris, 358(7):881–886, 2020.

[EGZ19] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Quantitative Harris-type theorems for diffusions and McKean-Vlasov processes. Trans. Amer. Math. Soc., 371(10):7135–7173, 2019.

[EK86] Stewart N. Ethier and Thomas G. Kurtz. Markov processes. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.

[EL94] K. David Elworthy and Xuei-Mei Li. Formulae for the derivatives of heat semigroups. J. Funct. Anal., 125(1):252–286, 1994.

[GM91] Saul B. Gelfand and Sanjoy K. Mitter. Recursive stochastic algorithms for global optimization in $\mathbf{R}^d$. SIAM J. Control Optim., 29(5):999–1018, 1991.

[GM93] Saul B. Gelfand and Sanjoy K. Mitter. Metropolis-type annealing algorithms for global optimization in $\mathbf{R}^d$. SIAM J. Control Optim., 31(1):111–131, 1993.

[GPP20] Sébastien Gadat, Fabien Panloup, and Clément Pellegrini. On the cost of bayesian posterior mean strategy for log-concave models. 2020.

[Guy06] Julien Guyon. Euler scheme and tempered distributions. Stochastic Process. Appl., 116(6):877–904, 2006.

[HMP20] Igor Honoré, Stéphane Menozzi, and Gilles Pagès. Non-asymptotic Gaussian estimates for the recursive approximation of the invariant distribution of a diffusion. Ann. Inst. Henri Poincaré Probab. Stat., 56(3):1559–1605, 2020.

[KM02] Valentin Konakov and Enno Mammen. Edgeworth type expansions for Euler schemes for stochastic differential equations. Monte Carlo Methods Appl., 8(3):271–285, 2002.

[Kun97] Hiroshi Kunita. Stochastic flows and stochastic differential equations, volume 24 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 1997. Reprint of the 1990 original.

[LCCC15a] Chunyuan Li, Changyou Chen, David Carlson, and Lawrence Carin. Preconditioned Stochastic Gradient Langevin Dynamics for Deep Neural Networks. arXiv e-prints, page arXiv:1512.07666, December 2015.

[LCCC15b] Chunyuan Li, Changyou Chen, David Carlson, and Lawrence Carin. Preconditioned stochastic gradient langevin dynamics for deep neural networks, 2015.

[Lem05] Vincent Lemaire. Estimation récursive de la mesure invariante d'un processus de diffusion. Thèse, Université de Marne la Vallée, December 2005.

[Lem07] Vincent Lemaire. An adaptive scheme for the approximation of dissipative systems. Stochastic Process. Appl., 117(10):1491–1518, 2007.

[LP02] Damien Lamberton and Gilles Pagès. Recursive computation of the invariant distribution of a diffusion. Bernoulli, 8(3):367–405, 2002.

[LP03] Damien Lamberton and Gilles Pagès. Recursive computation of the invariant distribution of a diffusion: the case of a weakly mean reverting drift. Stoch. Dyn., 3(4):435–451, 2003.

[LW16] Dejun Luo and Jian Wang. Exponential convergence in L^p-Wasserstein distance for diffusion processes without uniformly dissipative drift. Math. Nachr., 289(14-15):1909–1926, 2016.

[MCF15a] Yi-An Ma, Tianqi Chen, and Emily B. Fox. A Complete Recipe for Stochastic Gradient MCMC. arXiv e-prints, page arXiv:1506.04696, June 2015.

[MCF15b] Yi-An Ma, Tianqi Chen, and Emily B. Fox. A complete recipe for stochastic gradient mcmc, 2015.

[MFWB19] Wenlong Mou, Nicolas Flammarion, Martin J. Wainwright, and Peter L. Bartlett. Improved Bounds for Discretization of Langevin Diffusions: Near-Optimal Rates without Convexity. arXiv e-prints, page arXiv:1907.11331, July 2019.

[MP96] David Márquez and Mariane Pelletier. Sur la vitesse de convergence en loi du recuit simulé. C. R. Acad. Sci. Paris Sér. I Math., 322(4):391–394, 1996.

[Nua06] David Nualart. The Malliavin calculus and related topics. Probability and its Applications (New York). Springer-Verlag, Berlin, second edition, 2006.

[Pag01] Gilles Pagès. Sur quelques algorithmes récursifs pour les probabilités numériques. ESAIM Probab. Statist., 5:141–170, 2001.

[Pag18] Gilles Pagès. Numerical probability. Universitext. Springer, Cham, 2018. An introduction with applications to finance.

[Pan08a] Fabien Panloup. Computation of the invariant measure for a Lévy driven SDE: rate of convergence. Stochastic Process. Appl., 118(8):1351–1384, 2008.

[Pan08b] Fabien Panloup. Recursive computation of the invariant measure of a stochastic differential equation driven by a Lévy process. Ann. Appl. Probab., 18(2):379–426, 2008.

[Pel96] Mariane Pelletier. Sur le comportement asymptotique des algorithmes stochastiques. PhD thesis, Université Paris Sud, 1996.

[PP14] Gilles Pagès and Fabien Panloup. A mixed-step algorithm for the approximation of the stationary regime of a diffusion. Stochastic Process. Appl., 124(1):522–565, 2014.

[RY99] Daniel Revuz and Marc Yor. Continuous martingales and Brownian motion, volume 293 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, third edition, 1999.

[TT90] Denis Talay and Luciano Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. Stochastic Anal. Appl., 8(4):483–509 (1991), 1990.

[Vil09] Cédric Villani. Optimal transport, volume 338 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 2009. Old and new.

[Wan20] Feng-Yu Wang. Exponential contraction in Wasserstein distances for diffusion semigroups with negative curvature. Potential Anal., 53(3):1123–1144, 2020.

A Useful properties for the Euler scheme and its step sequence

A.1 Bounds on the moments of the Euler scheme with decreasing step

Proposition A.1. *Assume* $(\mathbf{S})$ *and* (Γ).

(a) *For every* $a>0$ *such that there exist a real constant* $\kappa_{b,\sigma,a}>0$ *such that, for any invariant distribution* ν, *one has*

$$\nu\big(V^a\big)\leqslant \kappa_{b,\sigma,a}$$

Furthermore, there exist real constants $C_{b,\sigma,a}>$ *and* $\bar{C}_{b,\sigma,a,\gamma}>0$ *such that, for every* $x\in\mathbb{R}^d$,

$$\sup_{t\geqslant 0}\mathbb{E}\,V^a(X^x_t)\leqslant C_{b,\sigma,a}V^a(x)\quad \textit{and}\quad \sup_{n\geqslant 0}\mathbb{E}\,V^a(\bar{X}^x_{\Gamma_n})\leqslant \bar{C}_{b,\sigma,a,\gamma}V^a(x). \tag{A.42}$$

(b) *There exists* $\lambda_{\sup}>0$ *such that, for any invariant distribution* ν, *for every* $x\in\mathbb{R}^d$ *and* $\forall\,\lambda\in(0,\lambda_{\sup})$, $\nu\big(e^{\lambda V}\big)<+\infty$.

Furthermore, there exists real constants $C_{b,\sigma,\lambda}>0$ *and* $C_{b,\sigma,\lambda,\gamma}>0$ *such that, for every* $x\in\mathbb{R}^d$,

$$\sup_{t\geqslant 0}\mathbb{E}\,e^{\lambda V(X^x_t)}\leqslant C_{b,\sigma,\lambda}e^{\lambda V(x)}\quad \textit{and}\quad \sup_{n\geqslant 0}\mathbb{E}\,e^{\lambda V(\bar{X}^x_{\Gamma_n})}\leqslant C_{b,\sigma,\lambda,\gamma}e^{\lambda V(x)}. \tag{A.43}$$

The bounds in (a) are straightforward consequences of (the proof of) Lemma 2 in [LP02] (established in more general setting where σ is possibly unbounded). The bounds in (b) are established in [Lem05, Theorem II.1] for the diffusion and [Lem05, Corollary III.1] for the Euler scheme (see also [GPP20, Lemma D.5 and D.6] for sharper exponential bounds in the additive setting).

A.2 Strong L^p-errors for the one-step Euler scheme (proofs of Lemmas 3.3 and 3.4)

Proof of Lemma 3.3. (a) It follows from the generalized Minkowski inequality and the B.D.G. inequality that

$$\begin{aligned}\|X_t^x - \bar{X}_t^{\gamma,x}\|_p &\leqslant \Big\|\int_0^t \big(b(X_s^x) - b(x)\big)ds\Big\|_p + \Big\|\int_0^t \big(\sigma(X_s^x) - \sigma(x)\big)dW_s\Big\|_p \\ &\leqslant [b]_{\rm Lip}\int_0^t \|X_s^x - x\|_p ds + C_p^{BDG}[\sigma]_{\rm Lip}\left(\Big\|\int_0^t |X_s^x - x|^2 ds\Big\|_{\frac{p}{2}}\right)^{1/2} \\ &\leqslant [b]_{\rm Lip}\int_0^t \|X_s^x - x\|_p ds + C_p^{BDG}[\sigma]_{\rm Lip}\left(\int_0^t \|X_s^x - x\|_p^2 ds\right)^{1/2}\end{aligned}$$

where $[\sigma]_{\rm Lip}$ should be understood with respect to the Frobenius norm. $\square$

Proof of Lemma 3.4 (a) One has by the general Minkowski inequality and BDG inequality

$$\begin{aligned}\|X_t^x - x\|_p &\leqslant \int_0^t \|b(X_s^x)\|_p ds + \left\|\int_0^t \sigma(X_s^x)dW_s\right\|_p \\ &\leqslant t|b(x)| + \sqrt{t}\|W_1\|_p\|\sigma(x)\| + \int_0^t \|b(X_s^x) - b(x)\|_p ds + \left\|\int_0^t (\sigma(X_s^x) - \sigma(x))dW_s\right\|_p \\ &\leqslant t|b(x)| + \sqrt{t}\|W_1\|_p\|\sigma(x)\| + [b]_{\rm Lip}\int_0^t \|X_s^x - x\|_p ds + C_{d,p}^{BDG}\left\|\int_0^t \|\sigma(X_s^x) - \sigma(x)\|^2 ds\right\|_{\frac{p}{2}}^{1/2} \\ &\leqslant t|b(x)| + \sqrt{t}\|W_1\|_p\|\sigma(x)\| + [b]_{\rm Lip}\int_0^t \|X_s^x - x\|_p ds + [\sigma]_{\rm Lip}C_{d,p}^{BDG}\left(\int_0^t \|X_s^x - x\|_p^2 ds\right)^{1/2}.\end{aligned}$$

Set $\varphi(t) = \sup_{0\leqslant s\leqslant t}\|X_s^x - x\|_p$ and $\psi(t) = t|b(x)| + \sqrt{t}\|W_1\|_p\|\sigma(x)\|$. Both functions are nondecreasing so that one derives from the above inequality that

$$\varphi(t) \leqslant \psi(t) + [b]_{\rm Lip}\int_0^t \varphi(s)ds + [\sigma]_{\rm Lip}C_{d,p}^{BDG}\left(\int_0^t \varphi(s)^2 ds\right)^{1/2}.$$

Now using that φ is non-decreasing, we derive for every $a > 0$,

$$\left(\int_0^t \varphi(s)^2 ds\right)^{1/2} \leqslant \sqrt{\varphi(t)}\sqrt{\int_0^t \varphi(s)ds} \leqslant \tfrac{a}{2}\varphi(t) + \tfrac{1}{2a}\int_0^t \varphi(s)ds.$$

As a consequence, setting $a = \frac{1}{[\sigma]_{\rm Lip}C_{d,p}^{BDG}}$, yields

$$\varphi(t) \leqslant 2\psi(t) + \Big(2[b]_{\rm Lip} + (C_{d,p}^{BDG}[\sigma]_{\rm Lip})^2\Big)\int_0^t \varphi(s)ds.$$

It follows from Gronwall's Lemma that, for every $t \in [0,\bar{\gamma}]$

$$\varphi(t) \leqslant 2\,e^{(2[b]_{\rm Lip} + [\sigma]_{\rm Lip}^2 C_{d,p}^{BDG}\bar{\gamma}}\psi(t)$$

which completes the proof.

(b)-(c) Having in mind that $\|\cdot\|_p \leqslant \|\cdot\|_{p\vee 2}$, it follows from Lemma 3.3 that

$$\|X_t^x - \bar{X}_t^{\gamma,x}\|_p \leqslant S_{p\vee 2}(x)\left([b]_{\rm Lip}\int_0^t \sqrt{s}ds + [\sigma]_{\rm Lip}\Big(\int_0^t s ds\Big)^{1/2}\right) = S_{p\vee 2}(x)\left(\tfrac{2}{3}[b]_{\rm Lip}\sqrt{t} + \frac{[\sigma]_{\rm Lip}}{\sqrt{2}}\right)t. \quad \square$$

Lemma A.2. (a) *Let $\Phi : \mathbb{R}^d \to (E, |\cdot|)$ be a Borel function with values in a normed vector space E and let $V : \mathbb{R}^d \to (0,+\infty)$ be a function such that $\sqrt{V}$ is Lipschitz continuous. If*

$$|\Phi| \leqslant C\cdot V^r \quad \textit{for some } C,\ r > 0,$$

then, for any $L^p(\mathbb{P})$-integrable $\mathbb{R}^d$-valued random vectors Y, Z, $p\in[1,+\infty)$,

$$\Big\|\sup_{\xi\in(Y,Z)}|\Phi(\xi)|\Big\|_p \leqslant C_{\Phi,V,r}\Big(\Big\|V(Y)\wedge V(Z)\Big\|_{rp}^r + \|Y-Z\|_{2rp}^{2r}\Big).$$

(b) Assume that the diffusion coefficients b and σ are Lipschitz continuous and satisfy $|b|^2+\|\sigma\|^2 \leqslant C.V$ where $\sqrt{V}$ is Lipschitz. Then, there exists a real constant for every $C_{\Phi,V,b,\sigma,p,\bar\gamma}$ such that, for every $\gamma\in(0,\bar\gamma)$,

$$\Big\|\sup_{\xi\in(x,X_\gamma^x)}|b(\xi)|\Big\|_p + \Big\|\sup_{\xi\in(X_\gamma^x,\bar X_\gamma^x)}|\sigma(\xi)|\Big\|_p \leqslant C_{\Phi,V,b,\sigma,p,\bar\gamma}V^{1/2}(x). \tag{A.44}$$

Proof. (a) This follows from the fact that $\sqrt{V}$ is Lipschitz continuous owing to assumption $(\mathbf{S})$ so that, for every $\xi\in(Y,Z)$,

$$\sqrt{V}(\xi)-\sqrt{V}(Z) \leqslant [\sqrt{V}]_{\rm Lip}|\xi-Z| \leqslant [\sqrt{V}]_{\rm Lip}|Y-Z|$$

and in turn

$$V(\xi)^r \leqslant 2^{(2r-1)_+}\big(V(Z)^r + [\sqrt{V}]_{\rm Lip}^{2r}|Y-Z|^{2r}\big).$$

One concludes using L^p-Minkowski's inequality.

(b) Note that by Lemma 3.4(a), $\|X_\gamma^x - x\|_{rp} \leqslant \bar\gamma^{\frac12}S_{rp,b,\sigma}(x) \leqslant \bar\gamma^{\frac12}V^{1/2}(x)$ which yields the bound for the first term on the left hand side. As for the second term, one proceeds likewise using Lemma 3.4(b). □

A.3 Technical lemmas on the steps

Lemma A.3. *Let $(\gamma_n)_{n\geqslant1}$ be a non-increasing positive sequence such that*

$$\varpi = \limsup_n \frac{\gamma_n-\gamma_{n+1}}{\gamma_{n+1}^2} < +\infty.$$

(i) Let $\rho>\varpi$ and let $(u_n)_{n\geqslant0}$ be the sequence defined by $u_0=0$ and, for every $n\geqslant1$, by

$$u_n = \sum_{k=1}^n \gamma_k^2 e^{-\rho(\Gamma_n-\Gamma_k)}.$$

Then,

$$\limsup_n \frac{u_n}{\gamma_n} < +\infty.$$

(ii) For every $T>0$, we have

$$\limsup_n \frac{\gamma_{N(\Gamma_n-T)}}{\gamma_n} < +\infty$$

(where $N(t)$ is defined in (1.5)).

(iii) Assume $\rho>\varpi$. Then for any $a\in\big(0,\frac{\rho}{\varpi}\big)$,

$$e^{-\rho\Gamma_n} = o(\gamma_n^a) \quad \text{as } n\to+\infty.$$

Proof. (i) Set $v_n=\frac{u_n}{\gamma_n}$, $n\geqslant1$. We have:

$$v_{n+1} = v_n\theta_n + \gamma_{n+1} \quad \text{with} \quad \theta_n = \frac{\gamma_n}{\gamma_{n+1}}e^{-\rho\gamma_{n+1}}. \tag{A.45}$$

Under the assumption, there exists $c\in(\varpi,\rho)$ and $n_0\in\mathbb{N}$ such that for all $n\geqslant n_0$,

$$\frac{\gamma_n}{\gamma_{n+1}} \leqslant 1+c\gamma_{n+1} \leqslant e^{c\gamma_{n+1}}. \tag{A.46}$$

Thus, for $n\geqslant n_0$, $\theta_n\leqslant e^{(c-\rho)\gamma_{n+1}}$ so that plugging this inequality into (A.45), we deduce

$$v_{n+1} \leqslant v_n e^{(c-\rho)\gamma_{n+1}} + \gamma_{n+1}$$

or, equivalently,

$$e^{(\rho-c)\Gamma_{n+1}}v_{n+1} \leqslant e^{(\rho-c)\Gamma_n}v_n + C'e^{(\rho-c)\Gamma_n}\gamma_{n+1}$$

where $C' = \sup_{k\geqslant 1} e^{(\rho-c)\gamma_k}$. Hence, by induction

$$e^{(\rho-c)\Gamma_n}v_n \leqslant e^{(\rho-c)\Gamma_{n_0}}v_{n_0} + C'\int_{\Gamma_{n_0}}^{\Gamma_n} e^{(\rho-c)u}du \leqslant e^{(\rho-c)\Gamma_{n_0}}v_{n_0} + \frac{C'}{\rho-c}e^{(\rho-c)\Gamma_n}$$

which clearly implies the announced boundedness.

(ii) By (A.46), for large enough n,

$$\frac{\gamma_{N(\Gamma_n-T)}}{\gamma_n} = \prod_{k=N(\Gamma_n-T)}^{n-1}\frac{\gamma_k}{\gamma_{k+1}} \leqslant e^{c(\Gamma_n-(\Gamma_{N(\Gamma_n-T)}))} \leqslant e^{c(T+\|\boldsymbol{\gamma}\|)}.$$

(iii) Set $w_n = e^{-\rho\Gamma_n}/\gamma_n^a$. Let $\varepsilon > 0$ be such that $a(\varpi+\varepsilon) < \rho$. Note that, for $n \geqslant n_0$, such that $\frac{\gamma_n-\gamma_{n+1}}{\gamma_{n+1}^2} \leqslant \varpi+\varepsilon$ for every $n \geqslant n_0$,

$$\begin{aligned} w_{n+1} = w_n e^{-\rho\gamma_{n+1}}\Big(\frac{\gamma_n}{\gamma_{n+1}}\Big)^a &= w_n e^{-\rho\gamma_{n+1}}e^{a\log(1+\frac{\gamma_n-\gamma_{n+1}}{\gamma_{n+1}})} \\ &\leqslant w_n e^{(a(\varpi+\varepsilon)-\rho)\gamma_{n+1}} \leqslant w_{n_0}e^{(a(\varpi+\varepsilon)-\rho)(\Gamma_{n+1}-\Gamma_{n_0})}. \end{aligned}$$

Hence, $\lim_n w_n = 0$ since $a(\varpi+\varepsilon)-\rho < 0$ and $\sum_{k\geqslant 1}\gamma_k = +\infty$. □

B Proof of Domino-Malliavin Theorem

The aim of this section is to prove Theorem 3.7. The proof is achieved in Subsection B.1 but strongly relies on a series of Malliavin bounds established in Subsection B.2. Note that w.l.g., we may only prove the result for $\bar{h}$ small enough. Actually, since the left-hand side of the inequality is bounded by 2, we can always extend to $\tilde{h}$ larger than $\bar{h}$ by artificially bounding the left-hand side by $2\bar{h}^{\varepsilon-2}h_1^{2-\varepsilon}$ for any h_1 greater than $\bar{h}$.

B.1 Proof of Theorem 3.7

By classical density arguments, it is enough to prove the result for a smooth function $f:\mathbb{R}^d\to\mathbb{R}$ with bounded derivatives as soon as the constant C of Inequality (3.33) only depends on $\|f\|_\infty$. Throughout the proof, $f:\mathbb{R}^d\to\mathbb{R}$ is thus assumed to be $\mathcal{C}^\infty$, bounded with bounded derivatives.

Step 1 (Expansion of $(P_h-\bar{P}_h)f(\xi)$*).* Let $\xi\in\mathbb{R}^d$ and let $h>0$. We have

$$P_hf(\xi) = \mathbb{E}f(X_h^\xi) = f(\xi) + \int_0^h \mathbb{E}[(\nabla f(X_s^\xi)|b(X_s^\xi))]ds + \frac{1}{2}\int_0^h \mathbb{E}\left[\mathrm{Tr}(D^2f(X_s^\xi)\sigma\sigma^*(X_s^\xi))\right]ds.$$

Again by Itô formula, for every $i\in\{1,\ldots,d\}$,

$$\mathbb{E}[\partial_i f(X_s^\xi)b_i(X_s^\xi))] = \partial_i f(\xi)b_i(\xi) + \int_0^s \mathbb{E}\left[(\nabla(\partial_i fb_i)(X_u^\xi)|b(X_u^\xi)) + \frac{1}{2}\mathrm{Tr}(D^2(\partial_i fb_i)\sigma\sigma^*)(X_u^\xi)\right]du,$$

and for every $i,j\in\{1,\ldots,d\}$,

$$\mathbb{E}\left[(D^2f(X_s^\xi)\sigma\sigma*)_{ii}(X_s^\xi)\right] = (D^2f\sigma\sigma^*)_{ii}(\xi) + \int_0^s \mathbb{E}\left[\mathcal{L}((D^2f\sigma\sigma^*)_{ii})(X_u^\xi)\right]du.$$

Thus,

$$P_hf(\xi) = \mathbb{E}f(X_h^\xi) = f(\xi) + h\mathcal{L}f(\xi) + \int_0^h\int_0^s\sum_{k=1}^4\sum_{|\alpha|=k}\mathbb{E}\left[\partial_\alpha f(X_u^\xi)\phi_\alpha(X_u^\xi))\right]ds, \tag{B.47}$$

where for any k, the functions ϕ_α are polynomial functions (which may be made explicit) of b, σ and their partial derivatives up, respectively, to order 2. Now, for the Euler scheme, let us introduce, for a positive M, a smooth and radial function $\mathfrak{T}_M : \mathbb{R}_d \to \mathbb{R}_+$ equal to 1 on $B(0,M)$ and 0 on $B(0,2M)^c$ and such that the derivatives of $\mathfrak{T}_M$ are uniformly bounded. Then,

$$\bar{P}_h f(\xi) = \mathbb{E}\,[f(\bar{X}_h^\xi)\mathfrak{T}_M(W_h)] + r_{h,M}(f), \quad \text{with} \quad |r_{h,M}(f)| \leqslant \|f\|_\infty \mathbb{P}(|W_h| > M).$$

Note that the rotation-invariance combined with the independence of the coordinates of the Brownian motion implies that for any $(a_1,\dots,a_d) \in \mathbb{N}^d$ with at least one odd integer, $\mathbb{E}\,[(W_h^1)^{a_1}\dots(W_h^d)^{a_d}\mathfrak{T}_M(W_h)] = 0$. Setting $By^{\otimes\ell} = \sum_{i_1,\dots,i_\ell}(B_{i_1,\dots,i_\ell})y_{i_1}\dots y_\ell$ for a an element B of $(\mathbb{R}^d)^\ell$, we deduce that

$$\mathbb{E}\,[D^2 f(\xi)(\sigma(\xi)W_h)^{\otimes 2}\mathfrak{T}_M(W_h)] = a(M,h)\mathrm{Tr}(D^2 f\sigma\sigma^*)(\xi) \quad \text{and} \quad \mathbb{E}\,[D^3 f(\xi)(\sigma(\xi)W_h)^{\otimes 3}\mathfrak{T}_M(W_h)] = 0,$$

where

$$a(M,h) = \mathbb{E}\,[(W_h^1)^2\mathfrak{T}_M(W_h)].$$

Then, it follows from the Taylor formula applied to $f(\bar{X}_h^\xi)$ that

$$\begin{aligned}
\mathbb{E}\,[f(\bar{X}_h^\xi)\mathfrak{T}_M(W_h)] &= \mathbb{E}\,\mathfrak{T}_M(W_h)\big(f(\xi) + h(\nabla f(\xi)|b(\xi))\big) + a(M,h)\mathrm{Tr}(D^2 f\sigma\sigma^*)(\xi) \\
&\quad + h^2\mathbb{E}\,[\mathfrak{T}_M(W_h)]\overbrace{\Big(\frac{1}{2}(D^2 f(\xi)b(\xi)|b(\xi)) + h\frac{1}{6}\sum_{i,j,k}^{3}\partial^3_{i,j,k}f(\xi)\,(b_i(b_jb_k + (\sigma\sigma^*)_{jk}))\,(\xi)\Big)}^{\varphi_h^{(1)}(\xi)} \\
&\quad + \frac{1}{24}\int_0^1 \mathbb{E}\,\big[\underbrace{D^4 f\,(\xi + \theta(hb(\xi) + \sigma(\xi)W_h))\,(hb(\xi) + \sigma(\xi)W_h)^{\otimes 4}\mathfrak{T}_M(W_h)}_{\varphi^{(2)}_{h,M}(\xi,\theta,W_h)}\big]d\theta.
\end{aligned} \tag{B.48}$$

Thus, noting that $1 - \mathbb{E}\,[\mathfrak{T}_M(W_h)] \leqslant \mathbb{P}(|W_h| > M)$, we deduce from what precedes and from (B.47), we get

$$\begin{aligned}
&\mathbb{E}\,[f(X_h^\xi)] - \mathbb{E}\,[f(\bar{X}_h^\xi)] = \varphi_{h,M}(\xi) \\
\text{where} \quad &\varphi_{h,M}(\xi) = r_{h,M}(f) + O(h\mathbb{P}(|W_h| > M))(\nabla f(\xi)|b(\xi)) \\
&\qquad + \frac{1}{2}(h - a(M,h))\mathrm{Tr}(D^2 f\sigma\sigma^*)(\xi) - h^2\mathbb{E}\,[\mathfrak{T}_M(W_h)]\varphi_h^{(1)}(\xi) \\
&\qquad + \int_0^h\int_0^s\sum_{k=1}^4\sum_{|\alpha|=k}\mathbb{E}\,[\partial_\alpha^k f(X_u^\xi)\phi_\alpha(X_u^\xi))]ds - \frac{1}{24}\int_0^1\mathbb{E}\,[\varphi^{(2)}_{h,M}(\xi,\theta,W_h)]d\theta.
\end{aligned}$$

Step 2: Assume now that $\xi = \bar{X}_{t_{n-1}}$, the Euler scheme at time t_{n-1} related to the step sequence $(h_n := t_n - t_{n-1})_{n\geqslant 1}$ starting from $x \in \mathbb{R}^d$. Let $\sigma_{\bar{X}_{t_{n-1}}}$ denote the Malliavin matrix of $\bar{X}_{t_{n-1}}$ (whose definition is recalled in Equation (B.56)). For $\eta \in (0,1]$, let Ψ_η denote a smooth function on $\mathbb{R}$ such that $\Psi_\eta(x) = 0$ on $(-\infty,\eta/2)$ and 1 on $(\eta,+\infty)$. We can furthermore assume that for every integer ℓ, $\|\Psi_\eta^{(\ell)}\|_\infty \leqslant C\eta^{-\ell}$ where C is a universal constant. Using that $W_{t_n} - W_{t_{n-1}}$ is independent from $\bar{X}_{t_{n-1}}$ and that $0 \leqslant 1 - \Psi_\eta(u) \leqslant 1_{\{u\leqslant\eta\}}$,

$$|\bar{P}_{h_1}\circ\dots\circ\bar{P}_{h_{n-1}}\circ(P_{h_n} - \bar{P}_{h_n})\circ f(x)| \leqslant 2\|f\|_\infty\mathbb{P}\big(\det\sigma_{\bar{X}_{t_{n-1}}} \leqslant \eta\big) + \Big|\mathbb{E}\Big[\varphi_{h_n,M}(\bar{X}_{t_{n-1}})\Psi_\eta(\det\sigma_{\bar{X}_{t_{n-1}}})\Big]\Big|.$$

Let us denote the unique solution at time u starting from x of (1.1) by $\mathfrak{X}(u,x)$ ($(u,x) \mapsto \mathfrak{X}(u,x)$ is the stochastic flow related to (1.1)). Note that $\mathbb{P}(|W_h| > M) = O(e^{-\frac{M^2}{4h}})$ and that

$$0 \leqslant h - a(M,h) = \mathbb{E}\,[(W_h^1)^2(1 - \mathfrak{T}_M(W_h)] \leqslant \mathbb{E}\,[(W_h^1)^2 1_{|W_h|>M}] \leqslant Che^{-\frac{M^2}{8h}},$$

by Cauchy-Schwarz and (exponential) Markov inequalities. Then, using the expansion of $\varphi_{h,M}$ obtained at the end of Step 1, we get,

$$\left|\bar{P}_{h_1}\circ\cdots\circ\bar{P}_{h_{n-1}}\circ(P_{h_n}-\bar{P}_{h_n})\circ f(x)\right| \leqslant 2\|f\|_\infty\left(e^{-\frac{M^2}{4h_n}}+\mathbb{P}(\det\sigma_{\bar{X}_{t_{n-1}}}\leqslant\eta)\right) \tag{B.49}$$

$$+O\big(h_n e^{-\frac{M^2}{4h_n}}\big)\left|\mathbb{E}\left[(\nabla f|b)(\bar{X}_{t_{n-1}})\Psi_\eta(\det\sigma_{\bar{X}_{t_{n-1}}})\right]\right| \tag{B.50}$$

$$+O\big(h_n e^{-\frac{M^2}{8h_n}}\big)\left|\mathbb{E}\left[\mathrm{Tr}(D^2 f\sigma\sigma^*)(\bar{X}_{t_{n-1}})\Psi_\eta(\det\sigma_{\bar{X}_{t_{n-1}}})\right]\right| \tag{B.51}$$

$$+O(h_n^2)\left|\mathbb{E}\left[\varphi^{(1)}_{h_n}(\bar{X}_{t_{n-1}})\Psi_\eta(\det\sigma_{\bar{X}_{t_{n-1}}})\right]\right| \tag{B.52}$$

$$+\int_0^h\int_0^s\sum_{k=1}^4\sum_{|\alpha|=k}\left|\mathbb{E}\left[\partial_\alpha f(\mathfrak{X}(u,\bar{X}_{t_{n-1}}))\phi_\alpha(\mathfrak{X}(u,\bar{X}_{t_{n-1}}))\Psi_\eta(\det\sigma_{\bar{X}_{t_{n-1}}})\right]\right|ds \tag{B.53}$$

$$+\frac{1}{24}\int_0^1\left|\mathbb{E}\left[\varphi^{(2)}_{h_n,M}(\bar{X}_{t_{n-1}},\theta,W_{t_n}-W_{t_{n-1}})\Psi_\eta(\det\sigma_{\bar{X}_{t_{n-1}}})\right]\right|d\theta. \tag{B.54}$$

Let us now consider all the above terms separately. We begin by the first term related to the probability of "degeneracy" of $\sigma_{\bar{X}_{t_{n-1}}}$. By Proposition B.3($i$) applied with $r=2$ a given positive T, we know that if $T/2\leqslant t_{n-1}\leqslant T$, we have for every $p>0$,

$$|(\text{B.49})|\leqslant C\|f\|_\infty\left(e^{-\frac{M^2}{4h_n}}+h_1^2+\eta^p\right)\leqslant C\|f\|_\infty\left(h_1^2+\eta^p\right),$$

where in the second inequality, we used that $e^{-\frac{M^2}{x}}\leqslant C_M x^2$ for $x\in(0,1]$. For (B.50) and (B.51), we use Lemma B.2(i) with $F=\bar{X}_{t_{n-1}}$. First, note that, owing to Proposition B.3(ii) and to the fact that b and σ are $\mathcal{C}^6$, Assumption (B.59) of this lemma holds true with $k\leqslant 4$. Then, one remarks that it is enough to apply Lemma B.2(i) with $|\alpha|=1$ and $G=b_i(F)$ $(i=1,\ldots,d)$ for (B.50), and, $|\alpha|=2$ and $G=\sigma_{i,j}\sigma_{k,i}(F)$, $(i,k)\in\{1,\ldots,d\}$ for (B.51). Since b_i has linear growth and bounded derivatives, it follows from Proposition B.3(ii) that $\|b_i(\bar{X}_{t_{n-1}})\|_{1,3}\leqslant C(1+\mathbb{E}\left[|\bar{X}_{t_{n-1}}|^3\right]^{\frac{1}{3}})$ whereas, since σ and its derivatives are bounded, $\|\sigma_{i,j}\sigma_{k,i}(\bar{X}_{t_{n-1}})\|_{2,3}\leqslant C$, where C does not depend on n. By Lemma B.2(i) and a Gronwall argument, it follows that a constant C exists (depending on T) such that

$$\begin{aligned}|(\text{B.50})|+|(\text{B.51})|&\leqslant Ch_n e^{-\frac{M^2}{8h_n}}\|f\|_\infty\eta^{-4}(1+\mathbb{E}_x[|\bar{X}_{t_{n-1}}|^6]^{\frac{1}{3}})(1+\mathbb{E}_x[|\bar{X}_{t_{n-1}}|^3]^{\frac{1}{3}}).\\&\leqslant Ch_1^2\|f\|_\infty\eta^{-4}(1+|x|^3).\end{aligned}$$

For (B.52), this is a direct application of Lemma B.2(ii) combined with Proposition B.3 (ii). This leads to

$$|(\text{B.52})|\leqslant Ch_n^2\eta^{-6}(1+\mathbb{E}_x[|\bar{X}_{t_{n-1}}|^9]^{\frac{2}{3}})\leqslant Ch_1^2\eta^{-6}(1+|x|^6).$$

For any α involved in (B.53), we can apply Lemma B.2(iii) with $F=\bar{X}_{t_{n-1}}$ and $\phi=\phi_\alpha$. Looking carefully into the definition of ϕ_α, one can check that for any α, for any $\ell\in\{0,\ldots,|\alpha|\}$, $|\phi_\alpha^{(\ell)}(x)|\leqslant C(1+|x|^2)$. Thus, taking the worst case $|\alpha|=4$ in Lemma B.2(iii), we get:

$$|(\text{B.53})|\leqslant Ch_n^2\|f\|_\infty\eta^{-12}\big(1+\mathbb{E}\left[|\bar{X}_{t_{n-1}}|^{24}\right]\big)^{\frac{1}{6}}\big(1+\mathbb{E}\left[|\bar{X}_{t_{n-1}}|^{24}\right]\big)^{\frac{1}{12}}\leqslant Ch_1^2\|f\|_\infty\eta^{-12}(1+|x|^6).$$

Finally, the control of (B.54) relies on Lemma B.2(iv) with $F=\bar{X}_{t_{n-1}}$. Once again, this statement holds true by Proposition B.3(ii). We have

$$|(\text{B.54})|\leqslant C\|f\|_\infty h_n^2\eta^{-12}(1+\mathbb{E}_x[|\bar{X}_{t_{n-1}}|^{24}])^{\frac{1}{3}}\leqslant C\|f\|_\infty h_1^2\eta^{-12}(1+|x|^8)$$

by using again that $\mathbb{E}\left[|\bar{X}_{t_{n-1}}|^p\right]\leqslant C(1+|x|^p)$.

Combining all the above controls, we deduce that there exists $\bar{h}>0$ and $T>0$ such that if $T/2\leqslant t_{n-1}\leqslant T$, then,

$$\left|\bar{P}_{h_1}\circ\cdots\circ\bar{P}_{h_{n-1}}\circ(P_{h_n}-\bar{P}_{h_n})\circ f(x)\right|\leqslant C\|f\|_\infty(\eta^p+h_1^2\eta^{-12}(1+|x|^8)).$$

For a given $\varepsilon>0$, it is now enough to fix $\eta=h_1^{\frac{\varepsilon}{12}}$ and $p=24\varepsilon^{-1}$ to conclude the proof.

B.2 Malliavin bounds

In this section, we detail the arguments which lead to the controls of the terms (B.49) to (B.54) involved in the decomposition of $\left|\bar{P}_{h_1} \circ \cdots \circ \bar{P}_{h_{n-1}} \circ (P_{h_n} - \bar{P}_{h_n}) \circ f(x)\right|$. All these terms are managed with the help of Malliavin-type arguments.

Without going into the details (for this, see $e.g.$ [Nua06]), let us recall some basic notations of Malliavin calculus on Wiener space. We set $\mathcal{H} = L^2(\mathbb{R}_+, \mathbb{R}^d)$ and denote by $W = \{W(h), h \in \mathcal{H}\}$, an isonormal Gaussian process on $\mathcal{H}$ which is assumed to be defined on a complete filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and that $\mathcal{F}$ is generated by W. We also denote by $(\mathcal{F}_t)_{t\geqslant 0}$ the completed natural filtration of $(W_t)_{t\geqslant 0}$.

The Malliavin operator is denoted by D and its domain by $\mathbb{D}^{1,p}$ for a given $p > 1$ (closure of $\mathcal{S}$, space of *smooth random variables*, in $L^p(\Omega)$ for the norm $\|\,.\,\|_{1,p}$ defined in (B.55)). For a ($\mathcal{F}$-measurable) random variable F in $\mathbb{D}^{p,1}$, DF is a random variable with values in $\mathcal{H}$ such that $\mathbb{E}\left[\|DF\|_{\mathcal{H}}^p\right] < +\infty$. For every multi-index $\alpha \in \{1, \dots, d\}^k\}$, the iterated derivative $D^\alpha F$ is defined on $\mathcal{H}^{\otimes k}$. The space $\mathbb{D}^{k,p}$ denotes the closure of $\mathcal{S}$ in $L^p(\Omega)$ for the norm $\|\,.\,\|_{k,p}$ defined for a given real-valued random variable F by

$$\|F\|_{k,p} = \mathbb{E}\left[|F|_k^p\right]^{\frac{1}{p}} \quad \text{with} \quad |F|_k = |F| + |F|_{k\backslash 0}, \quad \text{where} \quad |F|_{k\backslash 0} = \sum_{\ell=1}^{k} \|D^{(\ell)} F\|_{\mathcal{H}^{\otimes \ell}}, \tag{B.55}$$

and for every $\ell \geqslant 1$,

$$\|D^{(\ell)} F\|_{\mathcal{H}^{\otimes \ell}}^2 := \sum_{|\alpha|=\ell} \int_{[0,+\infty)^\ell} |D^\alpha_{s_1 \dots s_\ell} F|^2 ds_1 \dots ds_\ell.$$

For a random variable $F = (F^1, \dots, F^m)$, $|F|_{k\backslash 0} = \sum_{i=1}^m |F_i|_{k\backslash 0}$, $F|_k = \sum_{i=1}^m |F_i|$ and $\|F\|_{k,p}^p = \sum_{i=1}^m \mathbb{E}\left[|F_i|_k^p\right]$. Furthermore, for such $\mathbb{R}^m$-valued Malliavin-differentiable random variable F, the Malliavin matrix, denoted by σ_F, is defined by

$$\sigma_F = (\langle DF^i, DF^j \rangle_{\mathcal{H}})_{1\leqslant i,j\leqslant m}. \tag{B.56}$$

For any element A of $\mathbb{R}^{m^{\ell+1}}$, we will denote by $\|\,.\,\|$ the L^2-norm defined by

$$\|A\| = \sqrt{\sum_{1\leqslant i_1,\dots i_{\ell+1}\leqslant d} |A_{i_1,\dots,i_{\ell+1}}|^2}. \tag{B.57}$$

Note that when $\ell = 1$, this corresponds to the Frobenius norm on the space of $m \times m$ matrices.

B.2.1 Bounds for a general random variable F

In this subsection, we consider an $\mathcal{F}_{\mathfrak{t}}$-measurable random variable F and establish some useful bounds under appropriate Malliavin assumptions. Then, since in the proof of Theorem 3.7, we will use them with $F = \bar{X}_{t_{n-1}}$, we will prove in the next subsection that the assumptions of the results of this section hold true.

In the following lemma, we recall that $\mathfrak{X}(u, x)$ is the unique solution at time u starting from x (more precisely, $(u, x) \mapsto \mathfrak{X}(u, x)$ is the stochastic flow related to (1.1)). Furthermore, we implicitly assume that if F is an $\mathcal{F}_{\mathfrak{t}}$-measurable random variable, $\mathfrak{X}(u, x)$ is built with the increments of $W_{\mathfrak{t}+.} - W_{\mathfrak{t}}$. In particular, $\mathfrak{X}(u, x)$ is viewed as an $\mathcal{F}_{\mathfrak{t}+u}$ random variable.

Lemma B.1. *Let* $\mathfrak{t} > 0$ *Let* F *denote an* $\mathbb{R}^d$*-valued* $\mathcal{F}_{\mathfrak{t}}$*-measurable Malliavin-differentiable random variable. Assume* b *and* σ *have bounded first partial derivatives. Then,*

(i) *For every* $p \geqslant 1$ *and* $\eta > 0$*, a constant* $\mathfrak{C}$ *exists (which does not depend on* F*) such that*

$$\sup_{u\in[0,1]} \mathbb{E}\left[|\det(\sigma_{\mathfrak{X}(u,F)})|^{-p} \mathbf{1}_{\{\det \sigma_F \geqslant \eta\}}\right] \leqslant \mathfrak{C}\eta^{-p}.$$

(ii) *Set* $\bar{\mathfrak{X}}_\theta(u, x) = x + \theta\big(ub(x) + \sigma(x)(W_{\mathfrak{t}+u} - W_{\mathfrak{t}})\big)$*. Then, some positive* M*,* $\bar{h}$ *and* $\mathfrak{C}$ *exist such that for every* $u \in (0, \bar{h}]$ *and* $\theta \in (0, 1]$*,*

$$\mathbb{E}\left[|\det(\sigma_{\bar{\mathfrak{X}}_\theta(u,F)})|^{-p} \mathbf{1}_{\{\det \sigma_F \geqslant \eta, |W_{\mathfrak{t}+u} - W_{\mathfrak{t}}| \leqslant M\}}\right] \leqslant \mathfrak{C}\eta^{-p}.$$

Remark B.1. In (i), we state that on the set where σ_F is not degenerated, nor is $\mathfrak{X}(u,F)$ (with a non-degeneracy which is quantified along the parameter η). In (ii), we show that for the Euler scheme, this property is still true but up to a truncation of the Brownian increments (by M). Here, one retrieves that unfortunately, the Malliavin matrix of the Euler scheme is not invertible everywhere (see Proposition B.3(i) for a control of the lack of invertibility of $\sigma_{\bar{X}_{t_n}}$).

Proof. (i) As mentioned before the lemma, we implicitly assume that $\mathfrak{X}(u,x)$ is built with the increments of $W_{\mathfrak{t}+.} - W_{\mathfrak{t}}$. Thus $\mathfrak{X}(u,F)$ is a functional of $(W_s, 0 \leqslant s \leqslant T+u)$. Then, owing to the chain rule for Malliavin calculus, we remark that for any $s \in [0,T]$, for any i and $j \in \{1,\dots,d\}$,

$$D_s^j \mathfrak{X}_i(u,F) = \sum_{\ell=1}^{d} Y_u^{i\ell,F} D_s^j F^\ell, \tag{B.58}$$

where we recall that $Y_u^{i\ell,x} = \partial_{x_\ell} X_u^{i,x}$ (where $X^{i,x}$ stands for the ith coordinate of X_u^x). It follows that

$$\sigma_{\mathfrak{X}(u,F)} = \int_0^{t_n} Y_u D_s F (Y_u D_s F)^* ds + \int_{t_n}^{u} D_s \mathfrak{X}(u,F)(D_s\mathfrak{X}(u,F))^* ds.$$

Since for two symmetric positive matrices A and B, $\det(A+B) \geqslant \max(\det A, \det B)$, we deduce that

$$\det \sigma_{\mathfrak{X}(u,F)} \geqslant |\det(Y_u^F)|^2 \det\left(\int_0^{t_n} D_s F (D_s F)^* ds\right) = |\det(Y_u^F)|^2 \det(\sigma_F).$$

Thus,

$$\mathbb{E}\left[\det \sigma_{\mathfrak{X}(u,F)} \mathbf{1}_{\{\det \sigma_F \geqslant \eta\}}\right] \leqslant \eta^{-p} \mathbb{E}\left[|\det(Y_u^F)|^{-2p}\right] \leqslant C_p \eta^{-p},$$

where $C_p = \sup_{x\in\mathbb{R}^d, u\in[0,1]} \mathbb{E}\left[|\det(Y_u^x)|^{-2p}\right] < +\infty$ (the fact that $Y_0^x = I_d$ and that ∇b and $\nabla\sigma$ are bounded implies that C_p is finite, with the help of a Gronwall argument, similar to the one used in (iii)).

(iii) The map $x \mapsto \bar{\mathfrak{X}}_\theta(u,x)$ is differentiable on $\mathbb{R}^d$. Then, owing to the chain rule for Malliavin calculus, for every $j \in \{1,\dots,d\}$,

$$D_s^j \bar{\mathfrak{X}}_\theta(u,F) = \nabla_x \bar{\mathfrak{X}}_\theta(u,F) \circ D_s^j F,$$

and with the same arguments as in (i),

$$\det \sigma_{\bar{\mathfrak{X}}_\theta(u,F)} \geqslant |\det(\nabla_x \bar{\mathfrak{X}}_\theta(u,F))|^2 (\det \sigma_F).$$

Now,

$$\nabla_x \bar{\mathfrak{X}}_\theta(u,x) = \mathrm{I_d} + \theta(u\nabla b(x) + \nabla\sigma(x)(W_{\mathfrak{t}+u} - W_{\mathfrak{t}})),$$

and one checks that

$$\|\theta(u\nabla b(x) + \nabla\sigma(x)(W_{\mathfrak{t}+u} - W_{\mathfrak{t}}))\|_F \leqslant ud\|\nabla b\|_\infty + d\|\nabla\sigma\|_\infty |W_{\mathfrak{t}+u} - W_{\mathfrak{t}}|.$$

Thus, setting

$$M = \frac{1}{4d(\|\nabla\sigma\|_\infty \wedge 1)} \quad \text{and} \quad \bar{h} = \frac{1}{4d\|\nabla b\|_\infty},$$

we conclude the proof by noting that, on the event $\{\det \sigma_F \geqslant \eta\}$,

$$\inf_{h\in(0,\bar{h}]} \det \sigma_{\bar{\mathfrak{X}}_\theta(u,F)} \geqslant 2^{-2d}\eta.$$

Lemma B.2. *Let k be a positive integer. Assume that $|b(x)| \leqslant C(1+|x|)$ and that σ is bounded. Let $\mathfrak{t} > 0$. Let F be an $\mathbb{R}^d$-valued $\mathcal{F}_{\mathfrak{t}}$-measurable random variable, Malliavin-differentiable up to order $k+2$, such that for every $p \geqslant 1$,*

$$\sup_{1\leqslant\ell\leqslant k+2} \sup_{|\alpha|=\ell} \sup_{s_1,\dots,s_\ell\in[0,\mathfrak{t}]} \left(\mathbb{E}\left[\|D^\alpha_{s_1,\dots,s_\ell} F\|^p\right]\right)^{\frac{1}{p}} =: \mathfrak{d}_{p,\mathfrak{t}}^{(k+2)} < +\infty. \tag{B.59}$$

Let Ψ_η denote a smooth function on $\mathbb{R}$ such that $\Psi_\eta(x) = 0$ on $(-\infty, \eta/2)$ and 1 on $(\eta, +\infty)$. Then, some positive C and M exist such that for any $\eta > 0$

(i) For any $\alpha \in \{1,\dots,d\}^k$ with $|\alpha| = k$, for any G in $\mathbb{D}^{k,3}$,

$$|\mathbb{E}\left[\partial_\alpha f(F) G \Psi_\eta(\det \sigma_F)\right]| \leqslant C\|f\|_\infty \eta^{-2k}\big(1 + \mathbb{E}\left[|F|^{3k}\right]^{\frac{1}{3}}\big)\|G\|_{k,3}.$$

(ii) Assume that b and σ are $\mathcal{C}^3$ with bounded existing partial derivatives.

$$\left|\mathbb{E}\left[\varphi_h^{(1)}(F)\Psi_\eta(\det \sigma_F)\right]\right| \leqslant C\eta^{-6}\big(1 + \mathbb{E}\left[|F|^9\right]^{\frac{2}{3}}\big)$$

where C depends on b, σ, T and $\mathfrak{d}_{p,\mathfrak{t}}^{(3)}$ for a given p (which could be made explicit).

(iii) Assume that b and σ are $\mathcal{C}^{k+2}$ with bounded existing partial derivatives. Let $\phi : \mathbb{R}^d \mapsto \mathbb{R}$ denote a $\mathcal{C}^k$-function such that $|\phi(x)| + \sum_{\ell=1}^k \|\nabla^{(\ell)}\phi(x)\| \leqslant C(1+|x|^2)$. Then, for any $\alpha \in \{1,\dots,d\}^k$,

$$\sup_{u\in[0,\tau]} |\mathbb{E}\left[\partial_\alpha f(\mathfrak{X}(u,F))\phi(\mathfrak{X}(u,F))\Psi_\eta(\det \sigma_F)\right]| \leqslant C\|f\|_\infty \eta^{-3k}\big(1 + \mathbb{E}\left[|F|^{6k}\right]^{\frac{1}{6}}\big)\big(1 + \mathbb{E}\left[|F|^{24}\right]\big)^{\frac{1}{12}}.$$

(iv) Let Z be a $\mathcal{N}(0, I_d)$-random variable independent of F. For any $\theta \in [0,1]$,

$$\left|\mathbb{E}\left[\varphi_{h,M}^{(2)}(F,\theta,\sqrt{h}Z)\Psi_\eta(\det \sigma_F)\right]\right| \leqslant C\|f\|_\infty h^2 \eta^{-12}\big(1 + \mathbb{E}\left[|F|^{24}\right]\big)^{\frac{1}{3}}.$$

Proof. The proof strongly relies on [BCP20, Lemmas 2.3, 2.4].
(i) Let α denote a multi-index. By Lemmas 2.3 and 2.4(ii) of [BCP20] (applied with $k = 0$ and $n = |\alpha|$),

$$\mathbb{E}\left[\partial_\alpha f(F) G \Psi_\eta(\det\sigma_F)\right] = \mathbb{E}\left[f(F) H_\alpha(F, G\Psi_\eta(\det\sigma_F))\right],$$

where for some random variables F and G in $\mathbb{D}^{|\alpha|,p}$,

$$|H_\alpha(F, G\Psi_\eta(\det\sigma_F))| \leqslant C\eta^{-2|\alpha|}\left(|F|_{(|\alpha|+1)\backslash 0} + |LF|_{|\alpha|}\right)^{|\alpha|}(1 + |F|_{(|\alpha|+1)\backslash 0})^{4d|\alpha|}|G|_{|\alpha|}.$$

where $|F|_{k\backslash 0}$ is defined in (B.55) and L denotes the Ornstein-Uhlenbeck operator. Thus, using Hölder inequality, we deduce that

$$\mathbb{E}\left[|H_\alpha(F, G\Psi_\eta(\det\sigma_F))|\right] \leqslant C\eta^{-2|\alpha|}\mathbb{E}\left[\big(|F|_{(|\alpha|+1)\backslash 0} + |LF|_{|\alpha|}\big)^{3|\alpha|}\right]^{\frac{1}{3}}\mathbb{E}\left[(1 + |F|_{(|\alpha|+1)\backslash 0})^{12d|\alpha|}\right]^{\frac{1}{3}}\|G\|_{|\alpha|,3}.$$

Now, on the one hand, by the definition of $|F|_{k\backslash 0}$ and Assumption (B.59), one easily checks that for every positive integer k and positive $p \geqslant 1$,

$$\mathbb{E}\left[|F|_{k\backslash 0}^p\right]^{\frac{1}{p}} \leqslant C_{p,\mathfrak{t}}\mathfrak{d}_{p,\mathfrak{t}}^{(k)}. \tag{B.60}$$

On the other hand, the term involving the Ornstein-Uhlenbeck operator L can be classically controlled by Meyer inequalities (see $e.g.$ [Nua06, Theorem 1.5.1] or [BC19, Section 2.4]), which ensures for every integer m and positive p, the existence of a constant $C_{m,p}$ such that

$$\|LF\|_{m,p} \leqslant C_{m,p}\|F\|_{m+2,p} \leqslant C_{m,p,\mathfrak{t}}\left(\mathbb{E}\left[|F|^p\right]^{\frac{1}{p}} + \mathfrak{d}_{p,\mathfrak{t}}^{(m+2)}\right). \tag{B.61}$$

where $\|\,,\,\|_{k,p}$ is defined by (B.55). Thus, by the Minkowski inequality, we deduce that a constant C exists depending on T, $|\alpha|$ and $\mathfrak{d}_{12d|\alpha|,\mathfrak{t}}^{(|\alpha|+2)}$ such that,

$$\mathbb{E}\left[|H_\alpha(F, G\Psi_\eta(\det\sigma_F))|\right] \leqslant C\eta^{-2|\alpha|}\big(1 + \mathbb{E}\left[|F|^{3|\alpha|}\right]^{\frac{1}{3}}\big)\|G\|_{|\alpha|,3}.$$

(ii) We have to apply (i) for some multi-indices α with $|\alpha| = 2$ or $|\alpha| = 3$. More precisely, on the one hand, the first term of $\varphi_h^{(1)}(F)$ can be written as follows:

$$(D^2 f(F) b(F) | b(F)) = \sum_{i,j} \partial_{i,j}^2 f(F) G_{i,j} \quad \text{with} \quad G_{i,j} = b_i(F) b_j(F).$$

Thus, since $|b(x)| \leqslant C(1+|x|)$ and b has bounded derivatives, one checks (using the chain rule for Malliavin calculus and (B.60)) that,

$$\|G_{i,j}\|_{2,3} \leqslant C(1 + \mathbb{E}\left[|F|^6\right]^{\frac{1}{3}}),$$

where C depends on T and $\mathfrak{d}_{p,\mathfrak{t}}^{(4)}$ with $p = 6$. Thus, it follows from (i) (applied with $|\alpha| = 2$) that

$$\left|\mathbb{E}\left[(D^2 f(F) b(F)|b(F))\Psi_\eta(\det\sigma_F)\right]\right| \leqslant C\|f\|_\infty \eta^{-4}(1 + \mathbb{E}\left[|F|^6\right]^{\frac{2}{3}})$$

where C depends on T and $\mathfrak{d}_{p,\mathfrak{t}}^{(|\alpha|+2)}$ with $p = 24d$. On the other hand, the second term of $\varphi_h^{(1)}(F)$ has the following form:

$$\frac{1}{6}\sum_{i,j,k}^{3} \partial^3_{i,j,k} f(F) G_{i,j,k} \quad \text{with} \quad G_{i,j,k} = \left(b_i(b_j b_k + (\sigma\sigma^*)_{jk})\right)(F).$$

Using the assumptions on b and σ, one checks (using the chain rule for Malliavin calculus and (B.60)) that,

$$\|G_{i,j,k}\|_{3,3} \leqslant C(1 + \mathbb{E}\left[|F|^9\right]^{\frac{1}{3}}),$$

where C depends on T and $\mathfrak{d}_{p,\mathfrak{t}}^{(5)}$ with $p = 9$. Thus, it follows from (ii) (applied with $|\alpha| = 3$) that for every $(i,j,k) \in \{1,\ldots,d\}^3$,

$$\mathbb{E}\left[\partial^3_{i,j,k} f(F) G_{i,j,k}\Psi_\eta(\det\sigma_F)\right] \leqslant C\eta^{-6}(1 + \mathbb{E}\left[|F|^9\right]^{\frac{2}{3}}),$$

where, once again, C depends on T and $\mathfrak{d}_{p,\mathfrak{t}}^{(|\alpha|+2)}$ for a given value of p ($p = 36d$). The result follows.

(iii) For this statement and the following, we use Lemma 2.4(i) of [BCP20], which states that for some random variables $\bar{F}$ and G in $\mathbb{D}^{|\alpha|,p}$,

$$\mathbb{E}\left[\partial_\alpha f(\bar{F}) G\right] = \mathbb{E}\left[f(\bar{F}) H_\alpha(\bar{F}, G)\right], \tag{B.62}$$

where, on the set $\det\sigma_{\bar{F}} > 0$,

$$|H_\alpha(\bar{F},G)| \leqslant C\left(\frac{|\bar{F}|^{2(d-1)}_{(|\alpha|+1)\backslash 0}(|\bar{F}|_{(|\alpha|+1)\backslash 0} + |L\bar{F}|_{|\alpha|})}{\det\sigma_{\bar{F}}}\right)^{|\alpha|} \times \sum_{p_1+p_2\leqslant|\alpha|} |G|_{p_2}\left(1 + \frac{|\bar{F}|^{2d}_{(|\alpha|+1)\backslash 0}}{\det\sigma_{\bar{F}}}\right)^{p_1}.$$

It follows that on the set $\{\det\sigma_{\bar{F}} > 0\}$,

$$|H_\alpha(\bar{F},G)| \leqslant C\overbrace{\left(1 + \det\sigma_{\bar{F}} + |\bar{F}|^{2d}_{(|\alpha|+1)\backslash 0}\right)^{|\alpha|}\left((|\bar{F}|_{(|\alpha|+1)\backslash 0} + |L\bar{F}|_{|\alpha|})\right)^{|\alpha|}}^{\Upsilon_\alpha(\bar{F})} |G|_{|\alpha|}\left(1 + (\det\sigma_{\bar{F}})^{-2|\alpha|}\right).$$

By Hölder inequality, we deduce that

$$|\mathbb{E}\left[\partial_\alpha f(\bar{F}) G\right]| \leqslant \mathbb{E}\left[\Upsilon_\alpha(\bar{F})^3\right]^{\frac{1}{3}}\|G\|_{|\alpha|,3}\mathbb{E}\left[\left(1 + (\det\sigma_{\bar{F}})^{-2|\alpha|}\right)^3 1_{|G|>0}\right]^{\frac{1}{3}}. \tag{B.63}$$

Let us upper-bound $\mathbb{E}\left[\Upsilon_\alpha(\bar{F})^3\right]^{\frac{1}{3}}$ by a simpler quantity. First, denoting the largest eigenvalue of a symmetric matrix A by $\bar{\lambda}_A$, we remark that

$$|\det\sigma_{\bar{F}}| \leqslant \bar{\lambda}^d_{\sigma_{\bar{F}}} \leqslant C\|\sigma_{\bar{F}}\|^d$$

where $\|\,.\,\|$ stands for the Frobenius norm and where the second inequality follows from the equivalence of norms in finite dimension. But, one easily checks that

$$\|\sigma_{\bar{F}}\|^2 \leqslant C|\bar{F}|^2_{1,1}$$

so that

$$1 + \det\sigma_{\bar{F}} + |\bar{F}|^{2d}_{(|\alpha|+1)\backslash 0} \leqslant C(1 + |\bar{F}|^{2d}_{(|\alpha|+1)\backslash 0}).$$

Thus, by the elementary inequality $|u+v|^{|\alpha|} \leqslant 2^{|\alpha|-1}(|u|^{|\alpha|} + |v|^{|\alpha|})$, we get:

$$\Upsilon_\alpha(\bar{F}) \leqslant C\left[(1 + |\bar{F}|^{(2d+1)|\alpha|}_{(|\alpha|+1)\backslash 0}) + (1 + |\bar{F}|^{2d}_{(|\alpha|+1)\backslash 0})|L\bar{F}|^{|\alpha|}_{|\alpha|}\right].$$

Thus, using (B.61) (Meyer inequality) and Cauchy-Schwarz inequality, we deduce that

$$\mathbb{E}\,[\Upsilon_\alpha(\bar F)^3] \leqslant C\left(\mathbb{E}\,[1+|\bar F|^{3(2d+1)|\alpha|}_{(|\alpha|+1)\backslash 0}] + \mathbb{E}\,[1+|\bar F|^{12d|\alpha|}_{(|\alpha|+1)\backslash 0}]^{\frac12}\,\mathbb{E}\,[|\bar F|^{6|\alpha|}_{|\alpha|+2}]^{\frac12}\right)$$

and, hence, if

$$\mathbb{E}\,[\Upsilon_\alpha(\bar F)^3]^{\frac13} \leqslant C\mathbb{E}\,[(1+|\bar F|_{(|\alpha|+1)\backslash 0})^{12d|\alpha|}]^{\frac13}(1+\mathbb{E}\,[|\bar F|^{6|\alpha|}_{|\alpha|+2}]^{\frac16}).$$

Thus, we get the following inequality (where as usual, C denotes a constant which may change from line to line):

$$|\mathbb{E}\,[\partial_\alpha f(\bar F)G]| \leqslant C(1+\mathbb{E}\,[|\bar F|^{12d|\alpha|}_{(|\alpha|+1)\backslash 0}]^{\frac13})(1+\mathbb{E}\,[|\bar F|^{6|\alpha|}_{|\alpha|+2}]^{\frac16})\|G\|_{|\alpha|,3}(1+\mathbb{E}\,[(\det\sigma_{\bar F})^{-6|\alpha|}1_{|G|>0}]^{\frac13}). \quad \text{(B.64)}$$

We now want to apply Inequality (B.64) with $\bar F = \mathfrak{X}(u,F)$ and $G = \phi(\mathfrak{X}(u,F))\Psi_\eta(\det\sigma_F)$. Note that as in Lemma B.1, we implicitly assume that $\mathfrak{X}(u,F)$ is an $\mathcal{F}_{T+u}$-measurable random variable. On the one hand,

$$\mathbb{E}\,[(\det\sigma_{\bar F})^{-6|\alpha|}1_{|G|>0}]^{\frac13} \leqslant \mathbb{E}\,[\big(\det\sigma_{\mathfrak{X}(u,F)}\big)^{-6|\alpha|}1_{\det\sigma_F\geqslant\frac{\eta}{2}}]^{\frac13} \leqslant C\eta^{-2|\alpha|}, \quad \text{(B.65)}$$

by Lemma B.1(i). Now, since $x\mapsto\mathfrak{X}(u,x)$ is $\mathcal{C}^{k+2}$ (since b and σ are $\mathcal{C}^{k+2}$), one derives that if F is Malliavin-differentiable up to order k, then $\mathfrak{X}(u,F)$ so is. Furthermore, since b and σ have bounded derivatives, one can check that for every multi-index β such that $1\leqslant|\beta|\leqslant|\alpha|$, for every $p>0$, for every $\tau>0$,

$$\sup_{x\in\mathbb{R}^d}\sup_{u\in[0,\tau]}\mathbb{E}\,[\|\partial_x^\beta\mathfrak{X}(u,x)\|^p] < +\infty. \quad \text{(B.66)}$$

Then, using Assumption (B.59), the boundedness of σ and the Hölder inequality, a tedious computation of the Malliavin derivatives of $\mathfrak{X}(u,F)$ shows that for every $p>0$,

$$\sup_{u\in[0,\tau]}\mathbb{E}\,[|\mathfrak{X}(u,F)|^p_{(|\alpha|+2)\backslash 0}] \leqslant C_p < +\infty. \quad \text{(B.67)}$$

Thus, in view of (B.64), we deduce that a constant C exists (which does only depend on τ) such that

$$\mathbb{E}\,[|\mathfrak{X}(u,F)|^{12d|\alpha|}_{(|\alpha|+1)\backslash 0}] \leqslant C \quad\text{and}\quad \mathbb{E}\,[|\mathfrak{X}(u,F)|^{6|\alpha|}_{|\alpha|+2}] \leqslant C(1+\mathbb{E}\,[|\mathfrak{X}(u,F)|^{6|\alpha|}]).$$

Now, by a classical Gronwall argument, for every $\tau>0$, for every $p>0$,

$$\mathbb{E}\,[|\mathfrak{X}(u,x)|^p] \leqslant C(1+|x|^p), \quad \text{(B.68)}$$

so that

$$\mathbb{E}\,[|\mathfrak{X}(u,F)|^{6|\alpha|}] \leqslant C(1+\mathbb{E}\,[|F|^{6|\alpha|}]). \quad \text{(B.69)}$$

At this stage, we thus deduce from (B.64),

$$|\mathbb{E}\,[\partial_\alpha f(\mathfrak{X}(u,F))\phi(\mathfrak{X}(u,F))\Psi_\eta(\det\sigma_F)]| \leqslant C\eta^{-2|\alpha|}(1+\mathbb{E}\,[|F|^{6|\alpha|}]^{\frac16})\|\phi(\mathfrak{X}(u,F))\Psi_\eta(\det\sigma_F)\|_{|\alpha|,3}. \quad \text{(B.70)}$$

It thus remains to bound the last right-hand term. We again use chain rule for Malliavin calculus. In view of the application of the Leibniz formula (for the derivative of the product of functions), we study the Malliavin derivatives of $\phi(\mathfrak{X}(u,F))$ and $\Psi_\eta(\det\sigma_F)$ separately. For $\phi(\mathfrak{X}(u,F))$, we choose to write the arguments in the one-dimensional case (the extension to multidimensional case involves technicalities but leads to the same conclusion (B.72) below). In this case, $D^{(\ell)}\phi(\mathfrak{X}(u,F))$ takes the form:

$$D^{(\ell)}\phi(\mathfrak{X}(u,F)) = \sum_{r=1}^{\ell}\phi^{(r)}(\mathfrak{X}(u,F))Q^{(r)}(D\mathfrak{X}(u,F),\ldots,D^{(\ell)}\mathfrak{X}(u,F)), \quad \text{(B.71)}$$

where $Q^{(r)}$ denotes a multivariate polynomial function (with degree lower than r). Since $|\phi^{(r)}(x)|\leqslant C(1+|x|^2)$, it follows from a Gronwall argument that for every $p>0$, for every $\tau>0$, a constant C exists such that

$$\sup_{u\in[0,\tau]}\mathbb{E}\,[|\phi^{(r)}(\mathfrak{X}(u,F))|^p] \leqslant C(1+\mathbb{E}\,[|F|^{2p}]).$$

On the other hand, by (B.67) and Assumption (B.59), one deduces that for every positive p and τ , a constant C exists such that

$$\sup_{u\in[0,\tau]}\sup_{s_1,s_2,\ldots,s_\ell\in[0,\mathfrak{t}+\tau]}\mathbb{E}\,[|Q^{(r)}_{s_1,\ldots,s_\ell}(D\mathfrak{X}(u,F),\ldots,D^{(\ell)}\mathfrak{X}(u,F))|^p]<+\infty.$$

By Cauchy-Schwarz inequality, one deduces that for every positive $\mathfrak{t}$, τ and p

$$\sup_{u\in[0,\tau]}\sup_{(s_1,s_2,\ldots,s_\ell)\in[0,\mathfrak{t}+\tau]}\mathbb{E}\,[\|D^{(\ell)}_{s_1,\ldots,s_\ell}\phi(\mathfrak{X}(u,F))\|^p]^{\frac{1}{p}}\leqslant C(1+\mathbb{E}\,[|F|^{4p}]^{\frac{1}{2p}}). \tag{B.72}$$

Let us now consider $\Psi_\eta(\det\sigma_F)$. We have $\|\Psi^{(\ell)}_\eta\|_\infty\leqslant C\eta^{-\ell}$. Then, using that $\det$ is a polynomial function and Assumption (B.59), one can deduce that

$$\sup_{(s_1,s_2,\ldots,s_\ell)\in[0,\mathfrak{t}]}\mathbb{E}\,[\|D^{(\ell)}_{s_1,\ldots,s_\ell}\Psi_\eta(\det\sigma_F)\|^p]^{\frac{1}{p}}\leqslant C\eta^{-\ell}. \tag{B.73}$$

Then, by Leibniz formula and Cauchy-Schwarz inequality, we deduce from (B.72) and (B.73) (applied with $p=6$) that for every $\ell\in\{1,\ldots,|\alpha|\}$,

$$\sup_{u\in[0,\tau]}\sup_{(s_1,s_2,\ldots,s_\ell)\in[0,\mathfrak{t}+\tau]}\mathbb{E}\,[\|D^{(\ell)}_{s_1,\ldots,s_\ell}\left(\phi(\mathfrak{X}(u,F))\Psi_\eta(\det\sigma_F)\right)\|^3]^{\frac{1}{3}}\leqslant C\eta^{-\ell}(1+\mathbb{E}\,[|F|^{24}]^{\frac{1}{12}}).$$

Now, since Ψ_η is bounded, one easily checks that

$$\sup_{u\in[0,\tau]}\mathbb{E}\,[|\phi(\mathfrak{X}(u,F))\Psi_\eta(\det\sigma_F)|^p]\leqslant C(1+\mathbb{E}\,[|F|^{2p}]).$$

It follows from the two previous inequalities that

$$\sup_{u\in[0,\tau]}\|\phi(\mathfrak{X}(u,F))\Psi_\eta(\det\sigma_F)\|_{|\alpha|,3}\leqslant C\eta^{-|\alpha|}(1+\mathbb{E}\,[|F|^{24}]^{\frac{1}{12}}).$$

Plugging this inequality into (B.71), the result follows.

(iv) We have:

$$\mathbb{E}\,[\varphi^{(2)}_{h,M}(F,\theta,\sqrt{h}Z)\Psi_\eta(\det\sigma_F)]=\sum_{\alpha,|\alpha|=4}\mathbb{E}\,[\partial_\alpha f(\bar{\mathfrak{X}}_\theta(h,F))G_\alpha]$$

where for a given $\alpha=(\alpha_1,\ldots,\alpha_4)$

$$G_\alpha=\prod_{i=1}^{4}(hb_{\alpha_i}(F)+\sigma_{\alpha_i,.}(F)(W_{\mathfrak{t}+h}-W_{\mathfrak{t}}))\,\mathfrak{T}_M(W_{\mathfrak{t}+h}-W_{\mathfrak{t}})\Psi_\eta(\det\sigma_F).$$

The strategy is then quite similar to (iii). More precisely, for any $\alpha=(\alpha_1,\ldots,\alpha_4)$, we start by applying (B.62) with $\bar{F}=\bar{\mathfrak{X}}_\theta(h,F)$ and $G=G_\alpha$, which leads to the inequality (B.64). Then, as in (iii), it remains to control each term of the right-hand side of (B.64). Let us begin by the last one. Noting that (with the definition of $\mathfrak{T}_M$),

$$\{|G_\alpha|>0\}\subset\{\det\sigma_F\geqslant\frac{\eta}{2},|W_{\mathfrak{t}+h}-W_{\mathfrak{t}}|\leqslant 2M\},$$

we deduce from Lemma B.1(ii) that

$$\mathbb{E}\,[\det\sigma^{-6|\alpha|}_{\bar{\mathfrak{X}}_\theta(h,F)}1_{|G_\alpha|>0}]^{\frac{1}{3}}\leqslant C\eta^{-6|\alpha|}.$$

Then, since $x\mapsto\bar{\mathfrak{X}}_\theta(h,x)$ admits similar bounds as $x\mapsto\mathfrak{X}(u,x)$ (in particular (B.66) and (B.68)), some arguments similar to (iii) lead to an inequality similar to (B.70) (with $|\alpha|=4$): $\forall\,\alpha=(\alpha_1,\ldots,\alpha_4)$,

$$|\mathbb{E}\,[\partial_\alpha f(\bar{\mathfrak{X}}_\theta(h,F))G_\alpha]|\leqslant C\eta^{-8}(1+\mathbb{E}\,[|F|^{24}])^{\frac{1}{6}}\|G_\alpha\|_{4,3}.$$

It remains to control $\|G_\alpha\|_{4,3}$. The strategy of proof follows the lines of the ones for the control of $\|G\|_{|\alpha|,3}$ in (iii). Once again, a tedious computation using that b has sublinear growth and the fact $\mathfrak{T}_M$ is smooth with bounded derivatives leads to:

$$\|G_\alpha\|_{4,3}\leqslant Ch^2\eta^{-4}(1+\mathbb{E}\,[|F|^{24}])^{\frac{1}{6}}.$$

The result follows. $\square$

B.2.2 Bounds of Malliavin derivatives and Semi-nondegeneracy for the Euler scheme

Proposition B.3. *Let $(\bar{X}_{t_n})$ denote a Euler scheme starting from x with non-increasing step sequence $(h_n := t_n - t_{n-1})_{n\geqslant 1}$. Let $\ell \geqslant 1$. Assume that b and σ are $\mathcal{C}^\ell$ with bounded partial derivatives and σ satisfies $(\mathcal{E}\ell)_{\underline{\sigma}_0^2}$.*

(i) Let the smoothness assumption hold with $\ell = 1$. Then, for any $p > 0$, there exists a real constant $\bar{h} > 0$ such that, for any $T, r > 0$, there is areal constant $\mathfrak{C} = C(T, p, \bar{h}, r) > 0$ satisfying: if $h_1 \leqslant \bar{h}$, for any $\eta > 0$ and any n such that $T/2 \leqslant t_n \leqslant T$,

$$\mathbb{P}(\det \sigma_{\bar{X}_{t_n}} \leqslant \eta) \leqslant \mathfrak{C}(h_1^r + \eta^p).$$

(ii) Furthermore, if smoothness assumption holds for $\ell \geqslant 1$,

$$\sup_{t_n \in [0,T], (s_1, \ldots, s_\ell) \in [0, t_n]^\ell} \mathbb{E}\,[\|D^{(\ell)}_{s_1 \ldots s_\ell} \bar{X}_{t_n}\|^{2p}] \leqslant \mathfrak{d}_{p,T,\ell} < +\infty,$$

where $\mathfrak{d}_{p,T,\ell}$ is a finite positive constant.

Proof. (i) Let $s \in [0, T)$. Using the chain rule for Malliavin derivatives, one checks that $D_s \bar{X}_{s+.}$ formally satisfies for any $u \geqslant 0$:

$$D_s \bar{X}_{s+u} = \begin{cases} \sigma(\bar{X}_{\underline{s}}) & \text{if } u \leqslant \bar{s} - s \\ \sigma(\bar{X}_{\underline{s}}) + \int_{\bar{s}}^{s+u} \nabla b(\bar{X}_{\underline{v}}) D_s \bar{X}_{\underline{v}} dv + \int_{\bar{s}}^{s+u} \nabla \sigma(\bar{X}_{\underline{v}}) D_s \bar{X}_{\underline{v}} dW_v & \text{if } u > \bar{s} - s. \end{cases} \tag{B.74}$$

By "formally", we mean that we do not detail the rules for the operations between tensors. With some more precise notations, this yields in the case $u > \bar{s} - s$: for every ℓ and $i \in \{1, \ldots, d\}$,

$$D_s^\ell \bar{X}^i_{s+u} = \sigma_{i,\ell}(\bar{X}_{\underline{s}}) + \sum_{k=1}^d \int_{\bar{s}}^{s+u} \partial_k b_i(\bar{X}_{\underline{v}}) D_s^\ell \bar{X}^k_{\underline{v}} dv + \sum_{k,j=1}^d \int_{\bar{s}}^{s+u} \partial_k \sigma_{i,j}(\bar{X}_{\underline{v}}) D_s^\ell \bar{X}^k_{\underline{v}} dW_v^j.$$

For the sake of readability, we keep such formal notations in the sequel of the proof. Let us denote by $(\bar{Y}_t)_{t\geqslant 0}$ the "pseudo-tangent" process: $\bar{Y}_t = (\partial_{x_j} \bar{X}^i_t)_{1\leqslant i,j\leqslant d}$ where $(\bar{X}_t)_{t\geqslant 0}$ is the continuous-time Euler scheme defined by (1.3). One checks that $(\bar{Y}_t)_{t\geqslant 0}$ is recursively defined by: $\bar{Y}_0 = \mathrm{I_d}$ and for any $t \geqslant 0$:

$$\bar{Y}_t = (I_d + A_{\underline{t}t}) \bar{Y}_{\underline{t}},$$

where

$$A_{\underline{t}t} = (t - \underline{t}) \nabla b(\bar{X}_{\underline{t}}) + \nabla \sigma(\bar{X}_{\underline{t}})(W_t - W_{\underline{t}}).$$

Set

$$\Omega_\zeta := \bigcap_{k=0}^{n-1} \Big\{ \sup_{u \in [t_k, t_{k+1})} \|A_{\underline{u}u}\| < \zeta \Big\}, \quad \zeta \in (0, 1], \tag{B.75}$$

where $\|\cdot\|$ stands for the Fröbenius norm and ζ will be specified later (see (B.78)) as a constant only depending on d. On Ω_ζ, $\bar{Y}_s$ is invertible (as a product of invertible matrices), for every $s \in [0, t_n]$, and one checks that

$$D_s \bar{X}_t = \bar{Y}_{t \vee \bar{s}} \bar{Y}_{\bar{s}}^{-1} \sigma(\bar{X}_{\underline{s}}).$$

Let $t_n \in (0, T]$ and let $F_n = \bar{X}_{t_n}$. The Malliavin matrix σ_{F_n} of F_n is given by:

$$\sigma_{F_n} = \int_0^{t_n} D_s F_n D_s F_n^* ds,$$

which, after classical computations yields:

$$\sigma_{F_n} = \bar{Y}_{t_n} \bar{U}_{t_n} \bar{Y}^*_{t_n} \quad \text{with} \quad \bar{U}_{t_n} = \int_0^{t_n} \bar{Y}_{\bar{s}}^{-1} (\sigma\sigma^*)(\bar{X}_{\underline{s}}) (\bar{Y}_{\bar{s}}^{-1})^* ds.$$

For any $\eta > 0$ and $p > 0$,

$$\mathbb{P}(\det \sigma_{F_n} \leqslant \eta) \leqslant \mathbb{P}(\Omega_\zeta^c) + \eta^p \mathbb{E}\,[\det \sigma_{F_n}^{-p} \mathbf{1}_{\Omega_\zeta}]. \tag{B.76}$$

On the one hand, using that the partial derivatives of b and σ are bounded, we remark that

$$\sup_{t\in[t_{k-1},t_k)} \|A_{\underline{t}t}\| \leqslant C\big(h_k + \sup_{t\in[t_{k-1},t_k]} |W_t - W_{t_{k-1}}|\big),$$

where $C = C_{b,\sigma,d} > 0$ is a real constant depending on d, $\|\nabla b\|_\infty$ and $\|\nabla\sigma\|_\infty$. Hence, owing to the independence and the stationarity of the increments of the Brownian motion, we get

$$\mathbb{P}(\Omega_\zeta) \geqslant \prod_{k=1}^{n} \Big(1 - \mathbb{P}\big(\sup_{t\in[0,h_k]} |W_t| \geqslant \zeta C^{-1} - h_k\big)\Big).$$

Moreover, if B denotes a standard one-dimensional Brownian motion, one has for every h and $u > 0$,

$$\mathbb{P}\Big(\sup_{t\in[0,h]} |W_t| \geqslant u\Big) \leqslant q\,\mathbb{P}\Big(\sup_{t\in[0,h]} |B_t| \geqslant \tfrac{u}{q}\Big) \leqslant 2q\,\mathbb{P}\Big(\sup_{t\in[0,h]} B_t \geqslant \tfrac{u}{q}\Big) = 2q\,\mathbb{P}\Big(\sup_{t\in[0,1]} B_t \geqslant \tfrac{u}{q\sqrt{h}}\Big),$$

where we used that $B \overset{d}{=} -B$, $|x| = \max(x,-x)$ in the second inequality and the scaling property in the equality. Now $\sup_{t\in[0,1]} B_t \overset{d}{=} |Z|$ with $Z \overset{d}{=} \mathcal{N}(0;1)$ and $\mathbb{P}(|Z| \geqslant z) \leqslant e^{-\frac{z^2}{2}}$ for every $z \geqslant 0$ and we deduce that

$$\mathbb{P}(\Omega_\zeta) \geqslant \prod_{k=1}^{n} \Big(1 - 2qe^{-\frac{(\zeta C^{-1}-h_k)^2}{2q^2h_k}}\Big) \geqslant \prod_{k=1}^{n} \Big(1 - \kappa_0 e^{-\frac{\zeta^2}{2C^2q^2h_k}}\Big)$$

where $\kappa_0 = 2qe^{-\frac{\zeta}{Cq^2}}$ and $\kappa_1 = \frac{\zeta^2}{2C^2q^2}$ only depend on q, d, b and σ. For $h_1 \in (0,\bar{h}]$ with $\bar{h}$ small enough (and $\leqslant T$) so that that $\kappa_0 e^{-\frac{\kappa_1}{h_1}} \leqslant \frac{1}{2}$, we have $\kappa_0 e^{-\frac{\kappa_1}{h_k}} \leqslant \frac{1}{2}$ for every $k \geqslant 1$ since $(h_k)_k$ is non-increasing. Thus, combining this with the elementary inequalities $\log(1+u) \geqslant 2u$ on $[-1/2, 0]$ and $1 - e^{-u} \leqslant u$ on $[0,+\infty)$, we deduce that

$$\mathbb{P}(\Omega_\zeta^c) \leqslant 1 - \exp\Big(-2\kappa_0 \sum_{k=1}^{n} e^{-\frac{\kappa_1}{h_k}}\Big) \leqslant 2\kappa_0 \sum_{k=1}^{n} e^{-\frac{\kappa_1}{h_k}}.$$

Now, for any $r > 0$, there exists a constant C such that $e^{-\frac{\kappa_1}{x}} \leqslant C_r x^{r+1}$ for any $x \in [0,\bar{h}]$. Thus,

$$\mathbb{P}(\Omega_\zeta^c) \leqslant 2\kappa_0\, C_r \sum_{k=1}^{n} h_k^{r+1} \leqslant 2\kappa_0\, C_r t_n h_1^r \leqslant C_{T,r,\zeta,b,\sigma,d,q} \cdot h_1^r.$$

Let us now turn to the second term of (B.76). Recall that $\det$ is log-concave on $\mathcal{S}^{++}(d,\mathbb{R})$, hence $M \mapsto \det^{-p}(M)$ is convex on $\mathcal{S}^{++}(d,\mathbb{R})$ for any $p > 0$. Thus, by Jensen's inequality, we get

$$\mathbb{E}\big[\det \sigma_{F_n}^{-p} \mathbf{1}_\Omega\big] \leqslant t_n^{-pd-1} \int_0^{t_n} \mathbb{E}\big[\det{}^{-p}\big(\bar{Y}_{t_n}\bar{Y}_{\bar{s}}^{-1}(\sigma\sigma^*)(\bar{X}_{\underline{s}})(\bar{Y}_{\bar{s}}^{-1})^*\bar{Y}_{t_n}^*\big)\mathbf{1}_\Omega\big]ds. \tag{B.77}$$

Using that $\sigma\sigma^\star \geqslant \underline{\sigma}_0^2 I_d$ and det is also non-decreasing on $\mathcal{S}^{++}(d,\mathbb{R})$, we get

$$\begin{aligned}
\mathbb{E}\big[\det \sigma_{F_n}^{-p} \mathbf{1}_\Omega\big] &\leqslant \underline{\sigma}_0^{-2pd} t_n^{-pd} \sup_{s\in[0,t_n]} \mathbb{E}\,[|\det(\bar{Y}_{t_n}\bar{Y}_{\bar{s}}^{-1})|^{-2p}\mathbf{1}_\Omega] \\
&\leqslant \underline{\sigma}_0^{-2pd} t_n^{-pd} \mathbb{E}\,[|\det(\bar{Y}_{t_n}^{-1})|^{4p}\mathbf{1}_\Omega]^{\frac{1}{2}} \sup_{s\in[0,T]} \mathbb{E}\,[|\det(\bar{Y}_{\bar{s}})|^{4p}]^{\frac{1}{2}} \\
&\leqslant C_T \underline{\sigma}_0^{-2pd} t_n^{-pd} \mathbb{E}\,[|\det(\bar{Y}_{t_n})|^{-4p}\mathbf{1}_\Omega]^{\frac{1}{2}},
\end{aligned}$$

where in the last line we used that $\sup_{s\in[0,T]} \mathbb{E}\,[|\det(\bar{Y}_{\bar{s}})|^{4p}] \leqslant C_T$ (using that the moments of $\bar{Y}_t$ can be uniformly bounded on $[0,T]$ with the help of a Gronwall argument). Then, having in mind that the Trace operator is the differential of the determinant at I_d, yields a constant C such that, for $M \in \mathcal{M} = \{M \in \mathcal{M}(d,\mathbb{R}) : \|M\| \leqslant 1/2\}$,

$$\det(I_d + M) \geqslant 1 + \mathrm{Tr}(M) - C\|M\|^2.$$

Furthermore, one can choose $\zeta = \zeta_d > 0$ small enough in such a way that

$$\|M\| \leqslant \zeta_d \Longrightarrow 1 + \mathrm{Tr}(M) - C\|M\|^2 \geqslant 1/2. \tag{B.78}$$

Thus, taking such a ζ in (B.75) and using again that $\log(1+x) \geqslant 2x$ on $[-1/2, 0]$, we obtain:

$$
\begin{aligned}
\mathbb{E}\left[\det \sigma_{F_n}^{-p}\mathbf{1}_\Omega\right] &\leqslant C_T \underline{\sigma}_0^{-2pd} t_n^{-pd} \mathbb{E} \prod_{k=1}^n (1 + \mathrm{Tr}(A_{t_{k-1}t_k}) - C\|A_{t_{k-1}t_k}\|^2)^{-p} \\
&\leqslant C_T \underline{\sigma}_0^{-2pd} t_n^{-pd} \mathbb{E} \exp\Big(-2p \sum_{k=1}^n \Big(\mathrm{Tr}(A_{t_{k-1}t_k}) - C\|A_{t_{k-1}t_k}\|^2\Big)\Big) \\
&\leqslant C_T \underline{\sigma}_0^{-2pd} t_n^{-pd} \mathbb{E} \exp\Big(Cp\Big(1 - \mathrm{Tr}\Big(\int_0^{t_n} \nabla\sigma(\bar{X}_{\underline{t}}) dW_t\Big) + \|\nabla\sigma\|_\infty \sum_{k=1}^n |W_{t_k} - W_{t_{k-1}}|^2\Big)\Big),
\end{aligned}
$$

where in the last line, we used that ∇b is a bounded function. Now, using that for all $(u, v) \in \mathbb{R}^2$ such that $v < 1/2$, $\mathbb{E}_{Z\sim\mathcal{N}(0,1)}\big[e^{uZ+vZ^2}\big] = (1-2v)^{-\frac{1}{2}} e^{\frac{u^2}{2(1-2v)}}$, we deduce from a chain rule of towered expectations that if $Cp\|\nabla\sigma\|_\infty h_1 \leqslant 1/4$,

$$
\begin{aligned}
\mathbb{E}\left[\det \sigma_{F_n}^{-p}\mathbf{1}_\Omega\right] &\leqslant C_T \underline{\sigma}_0^{-2pd} t_n^{-pd} \exp\Big[\tilde{C}_{p,T}\|\nabla\sigma\|_\infty^2 t_n\Big]\Big(\prod_{k=1}^n \big(1 - 2Cp\|\nabla\sigma\|_\infty (t_k - t_{k-1})\big)\Big)^{-q/2} \\
&\leqslant C_{p,T} \underline{\sigma}_0^{-2pd} t_n^{-pd} \exp\big(2Cpq\|\nabla\sigma\|_\infty \sum_{k=1}^n h_k\big) \leqslant C_{p,T} \underline{\sigma}_0^{-2pd} t_n^{-pd} \exp\big(2Cpq\|\nabla\sigma\|_\infty T\big),
\end{aligned}
$$

where in the penultimate inequality, we again used that $\log(1+x) \geqslant 2x$ on $[-1/2, 0]$ (and where as usual, the constants may have changed from line to line). Hence, taking $t_n \in [T/2, T]$ yields

$$
\mathbb{E}\left[\det \sigma_{F_n}^{-p}\mathbf{1}_\Omega\right] \leqslant C_{p,q,T} \underline{\varepsilon}_0^{-2pd} (T/2)^{-pd},
$$

where $C_{p,T}$ stands for a finite constant depending on p, q and T. The statement follows.

(ii) For the sake of simplicity, we only prove the result in the one-dimensional case. For $\ell = 1$, we start with the formula (B.74) which implies, for any $s \in [0, t_n]$ and any $p > 0$,

$$
|D_s \bar{X}_{t_n}|^p = \Big(\sigma^2(\bar{X}_{\underline{s}}) \prod_{k=N(s)+1}^{n-1} \big(1 + (t_{k+1} - t_k) b'(\bar{X}_{\underline{v}}) + \sigma'(\bar{X}_{t_k})(W_{t_{k+1}} - W_{t_k})\big)^2\Big)^{\frac{p}{2}}
$$

with $N(s) = \max\{k, t_k \leqslant s\}$ and the convention $\prod_\varnothing = 1$. Thus, using the elementary inequality $\log\big((1+x)^2\big) = \log(1 + 2x + x^2) \leqslant 2x + x^2$ on $\mathbb{R}$ (with the convention $\log(0) = -\infty$), we get for any $p > 0$,

$$
\mathbb{E}\left[|D_s \bar{X}_{t_n}|^p\right] \leqslant \|\sigma\|_\infty^p \mathbb{E} \exp\left(\frac{p}{2} \sum_{k=N(s)+1}^{n-1} (2A_{t_k t_{k+1}} + |A_{t_k t_{k+1}}|^2)\right)
$$

with $A_{t_k t_{k+1}} = (t_{k+1} - t_k) b'(\bar{X}_{\underline{v}}) + \sigma'(\bar{X}_{t_k})(W_{t_{k+1}} - W_{t_k})$. Then,

$$
\mathbb{E}\left[|D_s \bar{X}_{t_n}|^p\right] \leqslant \|\sigma\|_\infty^p \exp\left(\frac{pT}{2}\|b'\|_\infty + \frac{p^2 T}{4}\|\sigma'\|_\infty^2\right)
$$

owing to standard estimates for exponential of stochastic integrals.

When $\ell \geqslant 1$, the idea is to iterate the Malliavin differentiation in (B.74). We give the main ideas when $\ell = 2$ but do not detail the general case. When $\ell = 2$, one can deduce from the chain rule and (B.74) that for any $t \geqslant 0$ and $(s, v) \in [0, t]^2$,

$$
D_{vs}^2 \bar{X}_t = D_v(D_s \bar{X}_t) = \begin{cases} \sigma'(\bar{X}_{\underline{t}}) D_v \bar{X}_{\underline{t}} & \text{if } v < \underline{t} \leqslant s < t, \\ \sigma'(\bar{X}_{\underline{s}}) D_v \bar{X}_{\underline{s}} + \int_{\bar{s}}^t D_v \bar{X}_{\underline{u}} D_s \bar{X}_{\underline{u}} (b''(\bar{X}_{\underline{u}}) du + \sigma''(\bar{X}_{\underline{u}}) dW_u) & \\ \qquad + \int_{\bar{s}}^t D_{vs}^2 \bar{X}_{\underline{u}} (b'(\bar{X}_{\underline{u}}) du + \sigma'(\bar{X}_{\underline{u}}) dW_u) & \text{if } 0 \leqslant s, v < \underline{t}. \end{cases}
$$

Thus, applying Itô formula to $|D_{vs}^2 \bar{X}_t|^p$ with $p \geqslant 2$, we easily deduce from martingale arguments and the boundedness of the derivatives that if $s, v < \underline{t}$

$$
\begin{aligned}
\mathbb{E}\left[|D_{vs}^2 \bar{X}_t|^p\right] &\leqslant \mathbb{E}\left[|\sigma'(\bar{X}_{\underline{s}}) D_v \bar{X}_{\underline{s}}|^p\right] + c_p \int_{\bar{s}}^t \mathbb{E}\left[|D_{vs}^2 \bar{X}_u|^{p-1} |D_v \bar{X}_{\underline{u}} D_s \bar{X}_{\underline{u}}|\right] du \\
&\quad + c_p \int_{\bar{s}}^t \mathbb{E}\left[|D_{vs}^2 \bar{X}_u|^{p-1} |D_{vs}^2 \bar{X}_{\underline{u}}|\right] du + c_p \int_{\bar{s}}^t \mathbb{E}\left[|D_{vs}^2 \bar{X}_u|^{p-2} \big(1 + |D_v \bar{X}_{\underline{u}} D_s \bar{X}_{\underline{u}}|^2\right] du\big).
\end{aligned}
$$

Then, by the Young inequality and the control of the moments of $D_{.}\bar{X}_{\underline{u}}$ previously established, we get by setting $S_t = \sup_{v\in[0,t]} |D^2_{vs}\bar{X}_v|^p$,

$$\mathbb{E}\,[S_t] \leqslant C_p + \int_{\bar{s}}^{t}(1 + \mathbb{E}\,[S_u])du.$$

The result then follows from Gronwall's inequality. □

Before proving Theorems 2.2(b) and 2.3(b), let us make the connection between $(\mathbf{H}_{\mathcal{W}_1})$ and its $\underline{\text{TV}}$-counterpart for uniformly elliptic diffusions.

C Proof of Theorem 3.1 and of Propositions 2.1, 2.6 and 3.2

C.1 Proof of extended BEL identity (Theorem 3.1)

Let $M > 0$ and $f_M(x) = f(x)1_{\{|f(x)|\leqslant M\}}$. Set $\phi_M(x) = P_t f_M(x)$. First, since $f(X_t^x)$ belongs to L^1 for any x since f has polynomial growth and b and σ are Lipschitz continuous. We deduce from the dominated convergence theorem that ϕ_M converges simply to $\phi = P_t f$. Furthermore, as f_M is bounded,

$$\phi'_M(x) = \mathbb{E}\Big[f_M(X_t^x)\frac{1}{t}\int_0^t (\sigma(X_s)^{-1}Y_s^{(x)})^* dW_s\Big].$$

We wish to prove that ϕ'_M converge uniformly on compact set K i.e.

$$\sup_{x\in K}\mathbb{E}\Big[\big|f(X_t^x)\big|\mathbf{1}_{\{|f(X_t^x)|>M\}}\left|\int_0^t (\sigma(X_s)^{-1}Y_s^{(x)})^* dW_s\right|\Big] \xrightarrow{M\to+\infty} 0.$$

It follows from Cauchy-Schwarz inequality that

$$\begin{aligned}
\mathbb{E}\left[\big|f(X_t^x)\big|\mathbf{1}_{\{|f(X_t^x)|>M\}}\left|\int_0^t (\sigma(X_s^x)^{-1}Y_s^{(x)})^* dW_s\right|\right] &\leqslant \Big[\mathbb{E}\,|f(X_t^x)|^2\mathbf{1}_{\{|f(X_t^x)|>M\}}\Big]^{\frac{1}{2}}\Big[\mathbb{E}\int_0^t \Big|\sigma(X_s^x)^{-1}Y_s^{(x)}\Big|^2 ds\Big]^{\frac{1}{2}}\\
&\leqslant \frac{1}{\sqrt{M}}\Big[\mathbb{E}\,|f(X_t^x)|^3\Big]^{\frac{1}{2}}\Big[\int_0^t \mathbb{E}\,\Big|\sigma(X_s^x)^{-1}Y_s^{(x)}\Big|^2 ds\Big]^{\frac{1}{2}}\\
&\leqslant \frac{1}{\sigma_0\sqrt{M}}\Big[1 + C_f\mathbb{E}\,|X_t^x|^{3r}\Big]^{\frac{1}{2}}\Big[\int_0^t \mathbb{E}\big|Y_s^{(x)}\big|^2 ds\Big]^{\frac{1}{2}}
\end{aligned}$$

where $|f(\xi)|^3 \leqslant C_f(1+|\xi|^{3r})$. Now, as b and σ have bounded partial derivatives (hence Lipschitz continuous), it is classical background that $\mathbb{E}\,|X_t^x|^{3r} \leqslant C_{r,t}(1+|x|^3)$ and $\sup_{x\in\mathbb{R}^d, s\in[0,t]}\mathbb{E}\,\big|Y_s^{(x)}\big|^2 < +\infty$. This shows that the right hand side goes to 0 as $M \to +\infty$ uniformly on compact sets of $\mathbb{R}^d$.

C.2 Proof of Proposition 2.1

Owing to Remark 2.2, we may assume that $(\mathbf{H}_{\mathcal{W}_1})$ holds starting from $t = 0$. Let $t_0 > 0$ being fixed. Let $t \geqslant t_0$ and let $f : \mathbb{R}^d \to \mathbb{R}$ be a bounded Borel function. By the Markov property,

$$\mathbb{E}\,[f(X_t^x) - f(X_t^y)] = \mathbb{E}\,[P_{t_0} f(X_{t-t_0}^x) - P_{t_0} f(X_{t-t_0}^y)].$$

By $\underline{\text{BEL}}$ identity (see Proposition 3.1), for any z_1 and $z_2 \in \mathbb{R}^d$,

$$P_{t_0} f(z_2) - P_{t_0} f(z_1) = \big(\nabla P_{t_0} f(\xi)\,|\, z_2 - z_1\big) = \frac{1}{t_0}\mathbb{E}\left[f(X_t)\Big(\int_0^{t_0} (\sigma^{-1}(X_s^\xi)Y_s^{(\xi)})^* dW_s \,|\, z_2 - z_1\Big)\right],$$

where $\xi \in (z_1, z_2)$ (geometric interval) and $(Y_s^{(\xi)})_{s\geqslant 0}$ denotes the tangent process of (X_s^ξ). But since b and σ have bounded derivatives and $(Y_s^{(x)})_{s\geqslant 0}$ starts from I_d, a Gronwall argument (see [Kun97]) shows that

$$\sup_{\xi\in\mathbb{R}^d,\, s\in[0,t_0]}\mathbb{E}\,\big\|Y_s^{(\xi)}\big\|^2 < +\infty.$$

By a standard martingale argument and the ellipticity condition $(\mathcal{E}\ell)_{\underline{\sigma}_0^2}$, we deduce that $x \mapsto P_{t_0}f(x)$ is Lipschitz continuous and that

$$[P_{t_0}f]_{\rm Lip} \leqslant C_0\|f\|_\infty.$$

Then, it follows from the Kantorovich-Rubinstein representation of the L^1-Wasserstein distance and the definition of total variation distance that

$$\|X_t^x - X_t^y\|_{TV} \leqslant C_0\mathcal{W}_1(X_{t-t_0}^x, X_{t-t_0}^y).$$

But under $(\mathbf{H}_{\mathcal{W}_1})$ it follows from what precedes, for every $t \geqslant t_0$,

$$\mathcal{W}_1(X_{t-t_0}^x, X_{t-t_0}^y) \leqslant ce^{-\rho(t-t_0)},$$

for some real constant $c > 0$. Hence, there exists a constant $C > 0$ such that, for every $t \geqslant t_0$,

$$\|X_t^x - X_t^y\|_{TV} \leqslant C|x-y|e^{-\rho t}. \qquad \square$$

C.3 Proof of Proposition 2.6

We need to check that $g = e^{-V}$ satisfies the stationary Fokker-Planck equation $\mathcal{L}^*g = 0$ where $\mathcal{L}^*$ denotes the adjoint operator of $\mathcal{L} = \mathcal{L}_X$ reading on C^2 test functions g

$$\mathcal{L}^*g = -\sum_{i=1}^d \partial_{x_i}(b_i g) + \tfrac{1}{2}\sum_{i,j=1}^d \partial^2_{x_ix_j}\big((\sigma\sigma^*)_{ij}g\big).$$

Temporarily set $a = \sigma\sigma^*$. For every $i, j \in \{1, \ldots, d\}$, elementary computations show that

$$\partial_{x_j}(b_i g) = \frac{e^{-V}}{2}\left[\sum_{j=1}^d a_{ij}(\partial_{x_i}V)(\partial_{x_j}V) - (\partial_{x_i}a_{ij})(\partial_{x_j}V) - (\partial_{x_j}a_{ij})(\partial_{x_i}V) - a_{ij}(\partial^2_{x_ix_j}V) + \partial^2_{x_ix_j}a_{ij}\right]$$

$$\partial^2_{x_ix_j}(a_{ij}g) = e^{-V}\left[\partial^2_{x_ix_j}a_{ij} - (\partial_{x_j}a_{ij})(\partial_{x_i}V) - (\partial_{x_i}a_{ij})(\partial_{x_j}V) + a_{ij}(\partial_{x_i}V)(\partial_{x_j}V) - a_{ij}(\partial^2_{x_ix_j}V)\right].$$

One checks from these identities that $\mathcal{L}^*g = 0$. Hence $\nu_V = C_V e^{-V(x)}\cdot\lambda_d(dx)$ is an invariant distribution for SDE (2.15). Uniqueness of the invariant distribution follows from uniform ellipticity.

C.4 Proof of Proposition 3.2

(a) Start from

$$\begin{aligned}\partial_x P_t f(x) &= \frac{1}{t}\mathbb{E}\left[f(X_t^x)\int_0^t (\sigma^{-1}(X_u^x)Y_u^{(x)})^* dW_u\right]\\ &= \partial_x P_{t-s}P_s f(x) = \frac{1}{t-s}\mathbb{E}\left[P_s f(X_{t-s}^x)\int_0^{t-s}(\sigma^{-1}(X_u^x)Y_u^{(x)})^* dW_u\right]\end{aligned}$$

so that, using that $\sup_{x\in\mathbb{R}^d}\sup_{s\in[0,T]}\mathbb{E}\big[|Y_s^{(x)}|^2\big] \leqslant C < +\infty$ since b and σ have bounded first partial derivatives,

$$|\partial_x P_t f(x)| \leqslant \frac{C_1}{\underline{\sigma}_0\sqrt{t}}\|f\|_{\sup}$$

and (with $s = \frac{t}{2}$)

$$\begin{aligned}\partial^2_{x^2}P_t f(x) &= \frac{2}{t}\partial_x\mathbb{E}\left[P_{\frac{t}{2}}f(X_{\frac{t}{2}}^x)\int_0^{\frac{t}{2}}(\sigma^{-1}(X_u^x)Y_u^{(x)})^* dW_u\right] \qquad \text{(C.79)}\\ &= \frac{2}{t}\mathbb{E}\left[\partial_x P_{\frac{t}{2}}f(X_{\frac{t}{2}}^x)\int_0^{\frac{t}{2}}(\sigma^{-1}(X_u^x)Y_u^{(x)})^* dW_u\right] + \frac{2}{t}\mathbb{E}\left[P_{\frac{t}{2}}f(X_{\frac{t}{2}}^x)\int_0^{\frac{t}{2}}\partial_x(\sigma^{-1}(X_u^x)Y_u^{(x)})^* dW_u\right].\end{aligned}$$

Let us denote (A) and (B) the two terms on the right hand side of the above equation. Using the above upper-bound for the first derivative, we obtain (with real constants varying from line to line denoted by capital letter C depending on b and σ and T)

$$\big|(A)\big| \leqslant \frac{2}{t}\frac{C}{\underline{\sigma}_0^2\sqrt{t}}\|f\|_{\sup} C'\sqrt{t} = \frac{C'}{\underline{\sigma}_0^2 t}\|f\|_{\sup}.$$

As for the second term

$$\big|(B)\big| \leqslant \frac{2}{t}\|f\|_{\sup}\left[\int_0^{\frac{t}{2}}\mathbb{E}\,\big|\partial_x(\sigma^{-1}(X_u^x)Y_u^{(x)})\big|^2 du\right]^{\frac{1}{2}}.$$

Using that b an σ have bounded existing partial derivatives, we derive by standard computations that $\sup_{x\in\mathbb{R}^d}\mathbb{E}\big[\sup_{s\in[0,T]}|\partial_x Y_s^{(x)}|^2\big] \leqslant C < +\infty$ so that (still using the $\underline{\sigma}_0^2$-ellipticity of $\sigma\sigma^*$)

$$\int_0^{\frac{t}{2}}\mathbb{E}\,\big|\partial_x(\sigma^{-1}(X_u^x)Y_u^{(x)})\big|^2 du \leqslant \frac{C''t}{\underline{\sigma}_0^4}$$

which finally implies that

$$|\partial^2_{x^2}P_t f(x)| \leqslant \frac{C_2}{\underline{\sigma}_0^2 t}\|f\|_{\sup}.$$

One shows likewise with similar arguments that

$$|\partial^3_{x^3}P_t f(x)| \leqslant \frac{C_3}{\underline{\sigma}_0^3 t^{\frac{3}{2}}}\|f\|_{\sup}.$$

(b) If f is Lipschitz continuous, note that

$$\partial_x P_t f(x) = \frac{1}{t}\mathbb{E}\left[\Big(f(X_t^x) - f(x)\Big)\int_0^t (\sigma^{-1}(X_u^x)Y_u^{(x)})^* dW_u\right]$$

so that

$$|\partial_x P_t f(x)| \leqslant \frac{C}{\underline{\sigma}_0\sqrt{t}}[f]_{\rm Lip}\|X_t^x - x\|_2 \leqslant \frac{C_1'}{\underline{\sigma}_0}[f]_{\rm Lip}S_2(x).$$

For the second differentiation, we still rely on (C.79) and its decomposition into two terms (A) and (B). Using the above bound for the first derivative, we derive like above that

$$\big|(A)\big| \leqslant \frac{C'}{\underline{\sigma}_0^2\sqrt{t}}S_2(x).$$

As for (B) we first note that

$$(B) = \frac{2}{t}\mathbb{E}\left[\Big(P_{\frac{t}{2}}f(X^x_{\frac{t}{2}}) - f(x)\Big)\int_0^{\frac{t}{2}}\partial_x(\sigma^{-1}(X_u^x)Y_u^{(x)})^* dW_u\right].$$

Now

$$\big|P_{\frac{t}{2}}f(X^x_{\frac{t}{2}}) - f(x)\big| \leqslant \big|\mathbb{E}\big[f(X_t^x) - f(x)\,|\,X^x_{\frac{t}{2}}\big]\big| \leqslant [f]_{\rm Lip}\mathbb{E}\big[|X_t^x - x|\,|\,X^x_{\frac{t}{2}}\big]$$

so that, using Cauchy-Schwarz inequality, the L^2-contraction property of conditional expectation and the above bound for the stochastic integral yields

$$\begin{aligned}\big|(B)\big| &\leqslant \frac{2}{t}[f]_{\rm Lip}\big\|X^x_{\frac{t}{2}} - x\big\|_2\left[\int_0^{\frac{t}{2}}\mathbb{E}\,\big|\partial_x(\sigma^{-1}(X_u^x)Y_u^{(x)})\big|^2 du\right]^{\frac{1}{2}}\\ &\leqslant \frac{2}{t}[f]_{\rm Lip}S_2(x)\sqrt{t}\sqrt{\frac{C''t}{\underline{\sigma}_0^4}} \leqslant C[f]_{\rm Lip}S_2(x).\end{aligned}$$

More generally, if $k = 1, 2, 3$, there exist real constants C_k' such that

$$|\partial^k_{x^k}P_t f(x)| \leqslant \frac{C_k'}{\underline{\sigma}_0^k t^{\frac{k-1}{2}}}[f]_{\rm Lip}S_2(x).$$