The Annals of Applied Probability
2020, Vol. 30, No. 6, 2613–2664
https://doi.org/10.1214/20-AAP1568

A SECOND ORDER ANALYSIS OF MCKEAN–VLASOV SEMIGROUPS

BY M. ARNAUDON[1] AND P. DEL MORAL[2,3]

[1]*Univ. Bordeaux, CNRS, Bordeaux INP, IMB, UMR 5251*
[2]*INRIA, Bordeaux Research Center, pierre.del-moral@inria.fr*
[3]*CMAP, Polytechnique Palaiseau*

We propose a second order differential calculus to analyze the regularity and the stability properties of the distribution semigroup associated with McKean–Vlasov diffusions. This methodology provides second order Taylor type expansions with remainder for both the evolution semigroup as well as the stochastic flow associated with this class of nonlinear diffusions. Bismut–Elworthy–Li formulae for the gradient and the Hessian of the integro-differential operators associated with these expansions are also presented.

The article also provides explicit Dyson–Phillips expansions and a refined analysis of the norm of these integro-differential operators. Under some natural and easily verifiable regularity conditions we derive a series of exponential decays inequalities with respect to the time horizon. We illustrate the impact of these results with a second order extension of the Alekseev–Gröbner lemma to nonlinear measure valued semigroups and interacting diffusion flows. This second order perturbation analysis provides direct proofs of several uniform propagation of chaos properties w.r.t. the time parameter, including bias, fluctuation error estimate as well as exponential concentration inequalities.

CONTENTS

Received June 2019; revised December 2019.

MSC2020 subject classifications. 65C35, 82C80, 58J65, 47J20.

Key words and phrases. Nonlinear diffusions, mean field particle systems, variational equations, logarithmic norms, gradient flows, Taylor expansions, contraction inequalities, Wasserstein distance, Bismut–Elworthy–Li formulae.

1. Introduction.

1.1. *Description of the models.* For any $n \geq 1$ we let $P_n(\mathbb{R}^d)$ be the convex set of probability measures η, μ on $\mathbb{R}^d$ with absolute nth moment and equipped with the Wasserstein distance of order n denoted by $\mathbb{W}_n(\eta,\mu)$. Also let $b_t(x_1,x_2)$ be some Lipschitz function from $\mathbb{R}^{2d}$ into $\mathbb{R}^d$ and let W_t be an d-dimensional Brownian motion defined on some filtered probability space $(\Omega,(\mathbb{F}_t)_{t\geq 0},\mathbb{P})$. We also consider the Hilbert space $\mathbb{H}_t(\mathbb{R}^d) := \mathbb{L}_2((\Omega,\mathbb{F}_t,\mathbb{P}),\mathbb{R}^d)$ equipped with the $\mathbb{L}_2$ inner product $\langle\cdot,\cdot\rangle_{\mathbb{H}_t(\mathbb{R}^d)}$. Up to a probability space enlargement there is no loss of generality to assume that $\mathbb{H}_t(\mathbb{R}^d)$ contains square integrable $\mathbb{R}^d$-valued variables independent of the Brownian motion.

For any $\mu \in P_2(\mathbb{R}^d)$ and any time horizon $s \geq 0$ we denote by $X^\mu_{s,t}(x)$ the stochastic flow defined for any $t \in [s,\infty[$ and any starting point $x \in \mathbb{R}^d$ by the McKean–Vlasov diffusion

$$dX^\mu_{s,t}(x) = b_t\big(X^\mu_{s,t}(x),\phi_{s,t}(\mu)\big)\,dt + dW_t \quad \text{with } b_t(x,\mu) := \int \mu(dy)b_t(x,y). \tag{1.1}$$

In the above display, $\phi_{s,t}$ stands for the evolution semigroup on $P_2(\mathbb{R}^d)$ defined by the formulae

$$\phi_{s,t}(\mu)(dy) = \mu P^\mu_{s,t}(dy) := \int \mu(dx)P^\mu_{s,t}(x,dy) \quad \text{with } P^\mu_{s,t}(x,dy) := \mathbb{P}\big(X^\mu_{s,t}(x) \in dy\big).$$

We denote by $L_{t,\phi_{s,t}(\mu)}$ the generator of the stochastic flow $X^\mu_{s,t}(x)$. The existence of the stochastic flow $X^\mu_{s,t}(x)$ is ensured by the Lipschitz property of the drift function; see, for instance, [41, 46]. To analyze the smoothness of the semigroup $\phi_{s,t}$ we need to strengthen this condition.

We shall assume that the function $b_t(x_1,x_2)$ is differentiable at any order with uniformly bounded derivatives. In addition, the partial differential matrices w.r.t. the first and the second coordinate are uniformly bounded; that is, for any $i = 1,2$ we have

$$\|b^{[i]}\|_2 := \sup_{t\geq 0}\sup_{(x_1,x_2)\in\mathbb{R}^{2d}} \|b^{[i]}_t(x_1,x_2)\|_2 < \infty \quad \text{with } b^{[i]}_t(x_1,x_2) := \nabla_{x_i}b_t(x_1,x_2). \tag{1.2}$$

In the above display, $\|A\|_2 := \lambda_{\max}(AA')^{1/2}$ stands for the spectral norm of some matrix A, where A' stands for the transpose of A, $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ the maximal and minimal eigenvalue. In the further development of the article, we shall also denote by $A_{\text{sym}} = (A+A')/2$ the symmetric part of a matrix A. In the further development of the article we represent the gradient of a real valued function as a column vector, or equivalently as the transpose of the differential-Jacobian operator which is, as any cotangent vector, represented by a row vector. The gradient and the Hessian of a column vector valued function as tensors of type $(1,1)$ and $(2,1)$; see for instance (3.1).

The mean field particle interpretation of the nonlinear diffusion (1.1) is described by a system of N-interacting diffusions $\xi_t = (\xi_t^i)_{1\leq i\leq N}$ defined by the stochastic differential equations

$$d\xi_t^i = b_t(\xi_t^i, m(\xi_t))\,dt + dW_t^i \quad \text{with } 1 \leq i \leq N \text{ and } m(\xi_t) := \frac{1}{N}\sum_{1\leq j\leq N} \delta_{\xi_t^j}. \tag{1.3}$$

In the above display, ξ_0^i stands for N independent random variables ξ_0^i with common distribution μ_0, and W_t^i are N independent copies of the Brownian motion W_t.

McKean–Vlasov diffusions and their mean field type particle interpretations arise in a variety of application domains, including in porous media and granular flows [7, 8, 18, 65], fluid mechanics [56, 57, 59, 66], data assimilation [10, 26, 37], and more recently in mean field game theory [9, 13–17, 43, 45], and many others.

The origins of this subject certainly go back to the beginning of the 1950s with the article by Harris and Kahn [44] using mean field type splitting techniques for estimating particle transmission energies. We also refer to the pioneering article by Kac [48, 49] on particle interpretations of Boltzmann and Vlasov equations, and the seminal articles by McKean [56, 57] on mean field particle interpretations of nonlinear parabolic equations arising in fluid mechanics. Since this period, the analysis of this class of mean field type nonlinear diffusions and their discrete time versions have been developed in various directions. For a survey on these developments we refer to [17, 26, 63], and the references therein.

The McKean–Vlasov diffusions discussed in this article belong to the class of nonlinear Markov processes. One of the most important and difficult research questions concerns the regularity analysis and more particularly the stability and the long time behavior of these stochastic models.

In contrast with conventional Markov processes, one of the main difficulties of these Markov processes comes from the fact that the evolution semigroup $\phi_{s,t}(\mu)$ is nonlinear w.r.t. the initial condition μ of the system. The additional complexity in the analysis of these models is that their state space is the convex set of probability measures, thus conventional functional analysis and differential calculus on Banach space cannot be directly applied.

The main contribution of this article is the development of a second order differential calculus to analyze the regularity and the stability properties of the distribution semigroup associated with McKean–Vlasov diffusions. This methodology provides second order Taylor type expansions with remainder for both the evolution semigroup as well as the stochastic flow associated with this class of nonlinear diffusions. We also provide a refined analysis of the norm of these integro-differential operators with a series of exponential decays inequalities with respect to the time horizon.

The article is organized as follows:

The main contributions of this article are briefly discussed in Section 1.2. The main theorems are stated in some detailed in Section 2. Section 3 provides some pivotal results on tensor integral operators and on integro-differential operators associated with the second order Taylor expansions of the semigroup $\phi_{s,t}(\mu)$. Section 4 is dedicated to the analysis of the tangent process associated with the nonlinear diffusion flow. We presents explicit Dyson–Phillips expansions as well as some spectral estimates. The last section, Section 4 is mainly concerned with the proofs of the first and second order Taylor expansions. The proof of some technical results are collected in the Appendix. Detailed comparisons with existing literature on this subject are also provided in Section 2.5.

1.2. *Statement of some main results.* One of the main contributions of the present article is the derivation of a second order Taylor expansion with the remainder of the semigroup

$\phi_{s,t}$ on probability spaces. For any pair of measures $\mu_0, \mu_1 \in P_2(\mathbb{R}^d)$, these expansions take basically the following form:

$$\phi_{s,t}(\mu_1) \simeq \phi_{s,t}(\mu_0) + (\mu_1 - \mu_0) D_{\mu_0}\phi_{s,t} + \frac{1}{2}(\mu_1 - \mu_0)^{\otimes 2} D^2_{\mu_0}\phi_{s,t}. \tag{1.4}$$

In the above display, $D^k_{\mu_0}\phi_{s,t}$ stands for some first and second order operators, with $k = 1, 2$. A more precise description of these expansions and the remainder terms is provided in Section 2.2.

Section 2.3.1, also provides an almost sure second order Taylor expansions with the remainder of the random state $X^{\mu}_{s,t}(x)$ of the McKean diffusion w.r.t. the initial distribution μ. These almost sure expansions take basically the following form:

$$\begin{aligned} X^{\mu_1}_{s,t}(x) - X^{\mu_0}_{s,t}(x) \simeq& \int (\mu_1 - \mu_0)(dy) D_{\mu_0} X^{\mu_0}_{s,t}(x, y) \\ &+ \frac{1}{2} \int (\mu_1 - \mu_0)^{\otimes 2}(dz) D^2_{\mu_0} X^{\mu_0}_{s,t}(x, z) \end{aligned} \tag{1.5}$$

for some random functions $D^k_{\mu_0} X^{\mu_0}_{s,t}$ from $\mathbb{R}^{(1+k)d}$ into $\mathbb{R}^d$, with $k = 1, 2$. A more precise description of these almost sure expansions is provided in Section 2.3.1 (see, for instance, (2.19) and Theorem 2.6).

Given some random variable $Y \in \mathbb{H}_s(\mathbb{R}^d)$ with distribution $\mu \in P_2(\mathbb{R}^d)$, observe that the stochastic flow $\psi_{s,t}(Y) := X^{\mu}_{s,t}(Y)$ satisfies the $\mathbb{H}_t(\mathbb{R}^d)$-valued stochastic differential equation

$$d\psi_{s,t}(Y) := B_t(\psi_{s,t}(Y))\, dt + dW_t. \tag{1.6}$$

In the above display, B_t stands for the drift function from $\mathbb{H}_t(\mathbb{R}^d)$ into itself defined by the formula

$$B_t(X) := \mathbb{E}(b_t(X, \overline{X})|X).$$

In the above display, $\overline{X}$ stands for an independent copy of X. The above Hilbert space valued representation of the McKean–Vlasov diffusion (1.1) readily implies that for any $Y_1, Y_0 \in \mathbb{H}_s(\mathbb{R}^d)$ we have the exponential contraction inequality

$$\|\psi_{s,t}(Y_1) - \psi_{s,t}(Y_0)\|_{\mathbb{H}_t(\mathbb{R}^d)} \leq e^{-\lambda(t-s)} \|Y_1 - Y_0\|_{\mathbb{H}_t(\mathbb{R}^d)}$$

for some $\lambda > 0$, as soon as the following condition is satisfied:

$$\langle X_1 - X_0, B_t(X_1) - B_t(X_0)\rangle_{\mathbb{H}_t(\mathbb{R}^d)} \leq -2\lambda \|X_1 - X_0\|^2_{\mathbb{H}_t(\mathbb{R}^d)} \tag{1.7}$$

for any $t \geq 0$ and any $X_1, X_0 \in \mathbb{H}_t(\mathbb{R}^d)$. In addition, in this framework the first order differential $\partial\psi_{s,t}(Y)$ of the stochastic flow coincides with the conventional Fréchet derivative of functions from an Hilbert space into another. In addition, we shall see that the gradient of first order operator $D_{\mu}\phi_{s,t}$ coincides with the dual of the tangent process associated with the Hilbert space-valued representation (1.6) of the McKean–Vlasov diffusion (1.1); that is, for any smooth function f we have that the dual tangent formula

$$\partial\psi_{s,t}(Y)^{\star} \cdot \nabla f(\psi_{s,t}(Y)) = \nabla D_{\mu}\phi_{s,t}(f)(Y). \tag{1.8}$$

A more precise description of the Fréchet differential $\partial\psi_{s,t}(Y)$ and the dual operator is provided in Section 2.1 and Section 4. A proof of the above formula is provided in Theorem 4.8.

The Taylor expansions discussed above are valid under fairly general and easily verifiable conditions on the drift function. For instance, the regularity condition (1.2) is clearly satisfied for linear drift functions. As is well known, dynamical systems and hence stochastic models involving drift functions with quadratic growth require additional regularity conditions to ensure nonexplosion of the solution in finite time.

Of course the expansions (1.4) and (1.5) will be of rather poor practical interest without a better understanding of the differential operators and the remainder terms. To get some useful approximations, we need to quantify with some precision the norm of these operators. A important part of the article is concerned with developing a series of quantitative estimates of the differential operators $D^k_{\mu_0}\phi_{s,t}$ and the remainder term; see, for instance, Theorem 2.3 and Theorem 2.4.

To avoid estimates that grow exponentially fast with respect to the time horizon, we need to estimate with some precision the operator norms of the differential operators in (1.4). To this end, we shall consider an additional regularity condition:

(H): There exists some $\lambda_0 > 0$ and $\lambda_1 > \|b^{[2]}\|_2$ such that for any $(x_1, x_2) \in \mathbb{R}^{2d}$ and any time horizon $t \geq 0$ we have

$$A_t(x_1, x_2)_{\mathrm{sym}} \leq -\lambda_0 I \quad \textit{and} \quad b^{[1]}_t(x_1, x_2)_{\mathrm{sym}} \leq -\lambda_1 I. \tag{1.9}$$

In the above display, I stands for the identity matrix and A_t the matrix-valued function defined by

$$A_t(x_1, x_2) := \begin{bmatrix} b^{[1]}_t(x_1, x_2) & b^{[2]}_t(x_2, x_1) \\ b^{[2]}_t(x_1, x_2) & b^{[1]}_t(x_2, x_1) \end{bmatrix} \quad \textit{and we set } \lambda_{1,2} := \lambda_1 - \|b^{[2]}\|_2. \tag{1.10}$$

Whenever (1.9) and (1.10) are met for some parameters λ_0 and $\lambda_1 \in \mathbb{R}$ all the exponential estimates stated in the article remains valid but they grow exponentially fast with respect to the time horizon. More detailed comments on the above regularity conditions, including illustrations for linear drift and gradient flow models, as well as comparisons with related conditions used in the literature on this subject are also provided in Section 2.4.

Under the above condition, we shall develop several exponential decays inequalities for the norm of the differential operators $D^k_{\mu_0}\phi_{s,t}$ as well as for the remainder terms in the Taylor expansions. The first order estimates are given in (2.6), the ones on the Bismut–Elworthy–Li gradient and Hessian extension formulae are provided in (2.7) and (2.8). Second and third order estimates can also be found in (2.12) and (2.15).

The second order differential calculus discussed above provides a natural theoretical basis to analyze the stability properties of the semigroup $\phi_{s,t}$ and the one of the mean field particle system discussed in (1.3).

For instance, a first order Taylor expansion of the form (1.4) already indicates that the sensitivity properties of the semigroup w.r.t. the initial condition μ are encapsulated in the first order differential operator $D_\mu\phi_{s,t}$. Roughly speaking, whenever (H) is satisfied, we show that there exists some parameter $\lambda > 0$ such that

$$\bigvee_{k=1,2} \left\|\left| D^k_{\mu_0}\phi_{s,t} \right|\right\| \simeq e^{-\lambda(t-s)} \quad \text{and therefore } \left\|\left| \phi_{s,t}(\mu_1) - \phi_{s,t}(\mu_0) \right|\right\| \simeq e^{-\lambda(t-s)} \tag{1.11}$$

for some operator norms $\||\cdot\||$. For a more precise statement we refer to Theorem 2.2 and the discussion following the theorem.

The second order expansion (1.4) also provides a natural basis to quantify the propagation of chaos properties of the mean field particle model (1.3). Combining these Taylor expansions with a backward semigroup analysis we derive a variety of uniform mean error estimates w.r.t. the time horizon. This backward second order analysis can be seen as a second order extension of the Alekseev–Gröbner lemma [1, 42] to nonlinear measure valued and stochastic semigroups. For a more precise statement we refer to Theorem 2.7. As in (1.11), one of the main feature of the expansion (1.4) is that it allows to enter the stability properties of the limiting semigroup $\phi_{s,t}$ into the analysis of the flow of empirical measures $m(\xi_t)$.

Roughly speaking, this backward perturbation analysis can be interpreted as a second order variation-of-constants technique applied to nonlinear equations in distribution spaces. As in the Ito's lemma, the second order term is essential to capture the quadratic variation of the processes; see, for instance, the recent articles [36, 47] in the context of conventional stochastic differential equation, as well as in [4, 32] in the context of interacting jump models.

The discrete time version of this backward perturbation semigroup methodology can also be found in Chapter 7 in [25], a well as in the articles [27, 28, 31] and [29, 35] for general classes of mean field particle systems.

The central idea is to consider the telescoping sum on some time mesh $t_n \leq t_{n+1}$ given by the interpolating formula

$$m_{t_n} - \phi_{t_0,t_n}(m_{t_0}) = \sum_{1\leq k\leq n} [\phi_{t_k,t_n}(m_{t_k}) - \phi_{t_k,t_n}(\phi_{t_{k-1},t_k}(m_{t_{k-1}}))] \quad \text{with } m_{t_k} := m(\xi_{t_k}).$$

Applying (1.4) and whenever $(t_k - t_{k-1}) \simeq 0$ we have the second order approximation

$$m_{t_n} - \phi_{t_0,t_n}(m_{t_0}) \simeq \frac{1}{\sqrt{N}} \sum_{1\leq k\leq n} \Delta M_{t_k} D_{\overline{m}_{t_k}} \phi_{t_k,t_n} + \frac{1}{2N} \sum_{1\leq k\leq n} (\Delta M_{t_k})^{\otimes 2} D^2_{\overline{m}_{t_k}} \phi_{t_k,t}$$

with the local fluctuation random fields

$$\Delta M_{t_k} := \sqrt{N}(m_{t_k} - \overline{m}_{t_k}) \quad \text{and} \quad \overline{m}_{t_k} := \phi_{t_{k-1},t_k}(m_{t_{k-1}}) \simeq m_{t_{k-1}}.$$

For discrete generation particle systems, $\xi^i_{t_k}$ are defined by N conditionally independent variables given the system $\xi_{t_{k-1}}$. For a more rigorous analysis we refer to Section 2.3.2.

The above decomposition shows that the first order operator $D_\mu \phi_{s,t}$ reflects the fluctuation errors of the particle measures, while the second order term encapsulates their bias. In other words, estimating the norm of second order operator $D^2_\mu \phi_{s,t}$ allows to quantify the bias induced by the interaction function, while the estimation of first order term is used to derive central limit theorems as well as $\mathbb{L}_p$-mean error estimates.

As in (1.11), these estimates take basically the following form. For $n \geq 1$ and any sufficiently regular function f we have

$$(1.12)\qquad \interleave D_{\mu_0}\phi_{s,t}\interleave \simeq e^{-\lambda(t-s)} \quad \Longrightarrow \quad \left|\mathbb{E}\left[\|m_t(f) - \phi_{0,t}(m_0)(f)\|^n\right]^{1/n}\right| \leq c_n/\sqrt{N}.$$

In addition, we have the uniform bias estimate w.r.t. the time horizon

$$(1.13)\qquad \interleave D^2_{\mu_0}\phi_{s,t}\interleave \simeq e^{-\lambda(t-s)} \quad \Longrightarrow \quad |\mathbb{E}[m_t(f) - \phi_{0,t}(m_0)(f)]| \leq c/N.$$

In the above display, $\interleave\cdot\interleave$ stands for some operator norm, and (c, c_n) stands for some finite constants whose values doesn't depend on the time horizon. We emphasize that the above results are direct consequence of a second order extension of the Alekseev–Gröbner type lemma for particle density profiles. For more precise statements we refer to Theorem 2.7 and the discussion following the theorem.

1.3. *Some basic notation.* Let $\mathrm{Lin}(\mathcal{B}_1, \mathcal{B}_2)$ be the set of bounded linear operators from a normed space $\mathcal{B}_1$ into a possibly different normed space $\mathcal{B}_2$ equipped with the operator norm $\interleave\cdot\interleave_{\mathcal{B}_1\to\mathcal{B}_2}$. When $\mathcal{B}_1 = \mathcal{B}_2$ we write $\mathrm{Lin}(\mathcal{B}_1)$ instead of $\mathrm{Lin}(\mathcal{B}_1, \mathcal{B}_1)$.

With a slight abuse of notation, we denote by I the identity $(d \times d)$-matrix, for any $d \geq 1$, as well as the identity operator in $\mathrm{Lin}(\mathcal{B}_1, \mathcal{B}_1)$. We also denote by $\|\cdot\|$ any (equivalent) norm on some finite dimensional vector space over $\mathbb{R}$.

We also use the conventional notation ∂_ϵ, ∂_{x_i}, ∂_s, ∂_t and so on for the partial derivatives w.r.t. some real valued parameters ϵ, x_i, s and t.

We let $\nabla f(x) = [\partial_{x_i} f(x)]_{1\leq i\leq d}$ be the gradient column vector associated with some smooth function $f(x)$ from $\mathbb{R}^d$ into $\mathbb{R}$. Given some smooth function $h(x)$ from $\mathbb{R}^d$ into $\mathbb{R}^d$ we denote by $\nabla h = [\nabla h^1, \dots, \nabla h^d]$ the gradient matrix associated with the column vector function $h = (h^i)_{1\leq i\leq d}$. We also let $(\nabla \otimes \nabla)$ be the second order differential operator defined for any twice differentiable function $g(x_1, x_2)$ on $\mathbb{R}^{2d}$ by the Hessian-type formula

$$((\nabla \otimes \nabla)g)_{i,j} = (\nabla_{x_1} \otimes \nabla_{x_2})(g)_{i,j} = (\nabla_{x_2} \otimes \nabla_{x_1})(g)_{j,i} = \partial_{x_1^i}\partial_{x_2^j} g. \tag{1.14}$$

We consider the space $\mathcal{C}^n(\mathbb{R}^d)$ of n-differentiable functions and we denote by $\mathcal{C}^n_m(\mathbb{R}^d)$ the subspace of functions f such that

$$\sup_{0\leq k\leq n} \|\nabla^k f(x)\| \leq c w_m(x) \quad \text{with the weight function } w_m(x) = (1+\|x\|)^m \text{ for some } m \geq 0.$$

We equip $\mathcal{C}^n_m(\mathbb{R}^d)$ with the norm

$$\|f\|_{\mathcal{C}^n_m(\mathbb{R}^d)} := \sum_{0\leq k\leq n} \|\nabla^k f/w_m\|_\infty \quad \text{with } \|\nabla^k f/w_m\|_\infty = \sup_{x\in\mathbb{R}^d} \|\nabla^k f(x)/w_m(x)\|.$$

When there are no confusions, we drop to lower symbol $\|\cdot\|_\infty$ and we write $\|f\|$ instead of $\|f\|_\infty$ the supremum norm of some real valued function. We let $e(x) := x$ be the identify function on $\mathbb{R}^d$ and for any $\mu \in P_n(\mathbb{R}^d)$ and $n \geq 1$ we set

$$\|e\|_{\mu,n} := \left[\int \|x\|^n \mu(dx)\right]^{1/n}.$$

For any $\mu_1, \mu_2 \in P_n(\mathbb{R}^d)$, we also denote by $\rho_n(\mu_1, \mu_2)$ some polynomial function of $\|e\|_{\mu_i,n}$ with $i = 1, 2$. When $\mu_1 = \mu_2$ we write $\rho_n(\mu_1)$ instead of $\rho_n(\mu_1, \mu_1)$.

Under our regularity conditions on the drift function, using elementary stochastic calculus for any $n \geq 2$ and $\mu \in P_n(\mathbb{R}^d)$ we check the following estimates:

$$\begin{gathered}\mathbb{E}(\|X^\mu_{s,t}(x)\|^n)^{1/n} \leq c_n(t)(\|x\| + \|e\|_{\mu,2}) \\ \text{which implies that } \phi_{s,t}(\mu)(\|e\|^n)^{1/n} \leq c_n(t)\|e\|_{\mu,n}.\end{gathered} \tag{1.15}$$

In the above display and throughout the rest of the article, we write $c(t)$, $c_\epsilon(t)$, $c_n(t)$, $c_{n,\epsilon}(t)$, $c_{\epsilon,n}(t)$ and $c_{m,n}(t)$ with $m, n \geq 0$ and $\epsilon \in [0, 1]$ some collection of nondecreasing and nonnegative functions of the time parameter t whose values may vary from line to line, but which only depend on the parameters m, n, ϵ, as well as on the drift function b_t. Importantly these contants do not depend on the probability measures μ. We also write c, c_ϵ, c_n, $c_{n,\epsilon}$, and $c_{m,n}$ when the constant does not depend on the time horizon.

2. Statement of the main theorems.

2.1. *First variational equation on Hilbert spaces.* As expected, the Fréchet differential $\partial\psi_{s,t}(Y)$ of the stochastic flow $\psi_{s,t}(Y)$ associated with the stochastic differential equation (1.6) satisfies an Hilbert space-valued linear equation (cf. (4.1)). The drift-matrix of this evolution equation is given by the Fréchet differential $\partial B_t(\psi_{s,t}(Y))$ of the drift function B_t evaluated along the solution of the flow. Mimicking the exponential notation of the solution of conventional homogeneous linear systems, the evolution semigroup (a.k.a. propagator) associated with the first variational equation is written as follows

$$\partial\psi_{s,t}(Y) = e^{\oint_s^t \partial B_u(\psi_{s,u}(Y))\,du} \in \mathrm{Lin}(\mathbb{H}_s(\mathbb{R}^d), \mathbb{H}_t(\mathbb{R}^d)).$$

The above exponential is understood as an operator valued Peano–Baker series [62]. A more detailed presentation of these models is provided in Section 4.

The $\mathbb{H}_t(\mathbb{R}^d)$-log-norm of an operator $T_t \in \mathrm{Lin}(\mathbb{H}_t(\mathbb{R}^d), \mathbb{H}_t(\mathbb{R}^d))$ is defined by

$$\gamma(T_t) := \sup_{\|Z\|_{\mathbb{H}_t(\mathbb{R}^d)}=1} \langle Z, (T_t + T_t^\star)/2 \cdot Z\rangle_{\mathbb{H}_t(\mathbb{R}^d)}.$$

Our first main result is an extension of an inequality of Coppel [22] to tangent processes associated with Hilbert-space valued stochastic flows.

THEOREM 2.1. *For any time horizon $t \geq s$ and any $Y \in \mathbb{H}_s(\mathbb{R}^d)$ we have the log-norm estimate*

$$\begin{aligned} -\int_s^t \gamma(-\partial B_u(\psi_{s,u}(Y)))\,du &\leq \frac{1}{t}\log \left|\!\left|\!\left| e^{\oint_s^t \partial B_u(\psi_{s,u}(Y))\,du} \right|\!\right|\!\right|_{\mathbb{H}_t(\mathbb{R}^d)\to\mathbb{H}_t(\mathbb{R}^d)} \\ &\leq \int_s^t \gamma(\partial B_u(\psi_{s,u}(Y)))\,du. \end{aligned} \tag{2.1}$$

In addition, we have

$$\begin{aligned} (H) \quad &\Longrightarrow \quad \partial B_t(X)_{\mathrm{sym}} \leq -\lambda_0 I \\ &\Longrightarrow \quad \frac{1}{t}\log \left|\!\left|\!\left| e^{\oint_s^t \partial B_u(\psi_{s,u}(Y))\,du} \right|\!\right|\!\right|_{\mathbb{H}_t(\mathbb{R}^d)\to\mathbb{H}_t(\mathbb{R}^d)} \leq -\lambda_0. \end{aligned} \tag{2.2}$$

The proof of the above theorem in provided in Section 4.1.

Let $Y_0, Y_1 \in \mathbb{H}_s(\mathbb{R}^d)$ be a pair of random variables with distributions $(\mu_0, \mu_1) \in P_2(\mathbb{R}^d)^2$. Also let μ_ϵ be the probability distribution of the random variable

$$Y_\epsilon := (1-\epsilon)Y_0 + \epsilon Y_1 \quad \Longrightarrow \quad \partial_\epsilon \psi_{s,t}(Y_\epsilon) = e^{\oint_s^t \partial B_u(\psi_{s,u}(Y))\,du} \cdot (Y_1 - Y_0). \tag{2.3}$$

This observation combined with the above theorem yields an alternative and more direct proof of an exponential Wasserstein contraction estimate obtained in [5]. Namely, using (2.2) we readily check the $\mathbb{W}_2$-exponential contraction inequality

$$\partial B_t(X)_{\mathrm{sym}} \leq -\lambda_0 I \quad \Longrightarrow \quad \mathbb{W}_2(\phi_{s,t}(\mu_1), \phi_{s,t}(\mu_0)) \leq e^{-\lambda_0(t-s)}\mathbb{W}_2(\mu_0, \mu_1). \tag{2.4}$$

For any function $f \in \mathcal{C}^1(\mathbb{R}^d)$ with bounded derivative we also quote the first order expansion

$$[\phi_{s,t}(\mu_1) - \phi_{s,t}(\mu_0)](f) = \int_0^1 \langle \partial\psi_{s,t}(Y_\epsilon)^\star \cdot \nabla f(\psi_{s,t}(Y_\epsilon)), (Y_1 - Y_0)\rangle_{\mathbb{H}_t(\mathbb{R}^d)}\,d\epsilon.$$

In the above display, $\langle \cdot, \cdot \rangle_{\mathbb{H}_t(\mathbb{R}^d)}$ stands for the conventional inner product on $\mathbb{L}_2((\Omega, \mathbb{F}_t, \mathbb{P}), \mathbb{R}^d)$. The above assertion is a direct consequence of Theorem 4.8.

2.2. *Taylor expansions with remainder.* The first expansion presented in this section is a first order linearization of the measure valued mapping $\phi_{s,t}$ in terms of a semigroup of linear integro-differential operators.

THEOREM 2.2. *For any $m, n \geq 1$ and $\mu_0, \mu_1 \in P_{m\vee 2}(\mathbb{R}^d)$, there exists a semigroup of linear operators $D_{\mu_1,\mu_0}\phi_{s,t}$ from $\mathcal{C}_m^n(\mathbb{R}^d)$ into itself such that*

$$\phi_{s,t}(\mu_1) = \phi_{s,t}(\mu_0) + (\mu_1 - \mu_0)D_{\mu_1,\mu_0}\phi_{s,t}. \tag{2.5}$$

In addition, when (H) is satisfied we have the gradient estimate

$$\|\nabla D_{\mu_1,\mu_0}\phi_{s,t}(f)\| \leq ce^{-\lambda(t-s)}\|\nabla f\| \quad \textit{for some } \lambda > 0. \tag{2.6}$$

The proof of the above theorem with a more explicit description of the first order operators $D_{\mu_1,\mu_0}\phi_{s,t}$ are provided in Section 4.3. In (2.6) we can choose $\lambda = \lambda_{1,2}$, with the parameter $\lambda_{1,2}$ introduced in (1.10). The semigroup property is a consequence of Theorem 4.5 and the gradient estimates is a reformulation of the operator norm estimate discussed in (4.13).

We also provide Bismut–Elworthy–Li-type formulae that allow to extend the gradient and Hessian operators $\nabla^k D_{\mu_1,\mu_0}\phi_{s,t}$ with $k = 1, 2$ to measurable and bounded functions. When the condition (H) is satisfied we show the following exponential estimates:

$$\|\nabla D_{\mu_1,\mu_0}\phi_{s,t}(f)\| \le c(1 \vee 1/\sqrt{t-s})e^{-\lambda(t-s)}\|f\| \quad \text{for some } \lambda > 0 \tag{2.7}$$

In addition, we have the Hessian estimate

$$\|\nabla^2 D_{\mu_1,\mu_0}\phi_{s,t}(f)\| \le c\big(1 \vee 1/(t-s)\big)e^{-\lambda(t-s)}\|f\| \quad \text{for some } \lambda > 0 \tag{2.8}$$

The proof of the first assertion can be found in Remark 4.7 on page 2642. The proof of the Hessian estimates is a consequence of the decomposition of $\nabla^2 D_{\mu_0,\mu_1}\phi_{s,t}$ discussed in (5.1) and the Hessian estimates (3.17) and (3.32).

It is worth mentioning that the semigroup property is equivalent to the chain rule formula

$$D_{\mu_1,\mu_0}\phi_{s,t} = D_{\mu_1,\mu_0}\phi_{s,u} \circ D_{\phi_{s,u}(\mu_1),\phi_{s,u}(\mu_0)}\phi_{u,t} \tag{2.9}$$

which is valid for any $s \le u \le t$. Without further work, Theorem 2.2 also yields the exponential $\mathbb{W}_1$-contraction inequality

$$\mathbb{W}_1\big(\phi_{s,t}(\mu_1), \phi_{s,t}(\mu_0)\big) \le ce^{-\lambda(t-s)}\mathbb{W}_1(\mu_0, \mu_1) \tag{2.10}$$

with the same parameter λ a in (2.6). In the same vein, the estimate (2.7) yields the total variation estimate

$$\|\phi_{s,t}(\mu_1) - \phi_{s,t}(\mu_0)\|_{\mathrm{tv}} \le c(1 \vee 1/\sqrt{t-s})e^{-\lambda(t-s)}\|\mu_0 - \mu_1\|_{\mathrm{tv}}$$

with the same parameter λ a in (2.7). In all the inequalities discussed above we can choose any parameter $\lambda > 0$ such that $\lambda < \lambda_{1,2}$, with the parameter $\lambda_{1,2}$ introduced in (1.10). In the $\mathbb{W}_1$-contraction inequality (2.10) we can choose $\lambda = \lambda_{1,2}$. A more refined estimate is provided in Section 2.4.

Next theorem provides a first order Taylor expansion with remainder.

THEOREM 2.3. *For any $m, n \ge 0$ and $\mu_0, \mu_1 \in P_{m+2}(\mathbb{R}^d)$, there exists a linear operator $D^2_{\mu_1,\mu_0}\phi_{s,t}$ from $\mathcal{C}^{n+2}_m(\mathbb{R}^d)$ into $\mathcal{C}^n_{m+2}(\mathbb{R}^{2d})$ such that*

$$\phi_{s,t}(\mu_1) = \phi_{s,t}(\mu_0) + (\mu_1 - \mu_0)D_{\mu_0}\phi_{s,t} + \frac{1}{2}(\mu_1 - \mu_0)^{\otimes 2}D^2_{\mu_1,\mu_0}\phi_{s,t} \tag{2.11}$$

with the first order operator $D_{\mu_0}\phi_{s,t} := D_{\mu_0,\mu_0}\phi_{s,t}$ introduced in Theorem 2.2. In addition, when (H) is satisfied we also have the estimate

$$\|(\nabla \otimes \nabla)D^2_{\mu_1,\mu_0}\phi_{s,t}(f)\| \le ce^{-\lambda(t-s)} \sup_{i=1,2}\|\nabla^i f\| \quad \text{for some } \lambda > 0. \tag{2.12}$$

The proof of the above theorem in provided in Section 5.2. A more precise description of the second order operator $D^2_{\mu_1,\mu_0}\phi_{s,t}$ is provided in (5.9) and (5.13). Using (2.11) and arguing as in the proof of proposition 2.1 in [4], for any twice differentiable function f with bounded derivatives we check the backward evolution equation

$$\partial_s\phi_{s,t}(\mu)(f) = -\mu L_{s,\mu}\big(D_\mu\phi_{s,t}(f)\big) \tag{2.13}$$

with the first order operator $D_\mu\phi_{s,t}$ introduced in Theorem 2.3. The above equation is a central tool to derive an extended version of the Alekseev–Gröbner lemma [1, 42] to measure valued semigroups and interacting diffusions (cf. Theorem 2.7).

Next theorem provides a second order Taylor expansion with remainder.

THEOREM 2.4. *For any $m, n \geq 1$ and $\mu_0, \mu_1 \in P_{m+4}(\mathbb{R}^d)$, there exists a linear operator $D^3_{\mu_1,\mu_0}\phi_{s,t}$ from $\mathcal{C}^{n+3}_m(\mathbb{R}^d)$ into $\mathcal{C}^n_{m+4}(\mathbb{R}^{3d})$ such that*

$$\begin{aligned}&\phi_{s,t}(\mu_1) - \phi_{s,t}(\mu_0)\\ &\quad = (\mu_1 - \mu_0) D_{\mu_0}\phi_{s,t} + \frac{1}{2}(\mu_1 - \mu_0)^{\otimes 2} D^2_{\mu_0}\phi_{s,t} + (\mu_1 - \mu_0)^{\otimes 3} D^3_{\mu_0,\mu_1}\phi_{s,t}\end{aligned} \tag{2.14}$$

with the second order operator $D^2_{\mu_0}\phi_{s,t} := D^2_{\mu_0,\mu_0}\phi_{s,t}$ introduced in Theorem 2.3. In addition, when (H) is satisfied we have the third order estimate

$$\begin{aligned}&\left|(\mu_1 - \mu_0)^{\otimes 3} D^3_{\mu_0,\mu_1}\phi_{s,t}(f)\right|\\ &\quad \leq c e^{-\lambda(t-s)} \Big(\bigvee_{i=1,2,3} \|\nabla^i f\|\Big) \mathbb{W}_2(\mu_0, \mu_1)^3 \quad \text{for some } \lambda > 0.\end{aligned} \tag{2.15}$$

The proof of the first part of the above theorem is provided in Section 5.3. We can choose in (2.15) any parameter $\lambda > 0$ such that $\lambda < \lambda_{1,2}$, with the parameter $\lambda_{1,2}$ introduced in (1.10). The proof of the third order estimate (2.15) is rather technical, thus it is provided in the Appendix, on page 2654.

2.3. *Illustrations*. The first part of this section states with more details the almost sure expansions discussed in (1.5). Up to some differential calculus technicalities, this result is a more or less direct consequence of the Taylor expansions with remainder presented in Theorem 2.3 and Theorem 2.4 combining with a backward formula presented in [5].

The second part of this section is concerned with a second order extension of the Alekseev–Gröbner lemma to nonlinear measure valued semigroups and interacting diffusion flows. This second order stochastic perturbation analysis is also mainly based on the second order Taylor expansion with the remainder presented in Theorem 2.4.

In the further development of this section without further mention we shall assume that condition (H) is satisfied.

2.3.1. *Almost sure expansions*. We recall the backward formula

$$\begin{aligned}&X^{\mu_1}_{s,t}(x) - X^{\mu_0}_{s,t}(x)\\ &\quad = \int_s^t \big[\nabla X^{\phi_{s,u}(\mu_0)}_{u,t}\big]\big(X^{\mu_1}_{s,u}(x)\big)'\big[\phi_{s,u}(\mu_1) - \phi_{s,u}(\mu_0)\big]\big(b_u\big(X^{\mu_1}_{s,u}(x), \cdot\big)\big)\, du.\end{aligned} \tag{2.16}$$

The above formula combined with (2.4) and the tangent process estimates presented in Section 3.3 yields the uniform almost sure estimates

$$\|X^{\mu_1}_{s,t}(x) - X^{\mu_0}_{s,t}(x)\| \leq e^{-(\lambda_0 \wedge \lambda_1)(t-s)} \mathbb{W}_2(\mu_0, \mu_1). \tag{2.17}$$

The above estimate is a consequence of (2.4) and conventional exponential estimates of the tangent process $\nabla X^{\mu}_{s,t}$ (cf., for instance, (3.2)). A detailed proof of this claim and the backward formula (2.16) can be found in [5].

We extend the operators $D^k_{\mu}\phi_{s,t}$ introduced in Theorem 2.4 to tensor valued functions $f = (f_i)_{i \in [n]}$ with $i = (i_1, \dots, i_n) \in [n] := \{1, \dots, d\}^n$ by considering the same type tensor function with entries

$$D^k_{\mu}\phi_{s,t}(f)_i := D^k_{\mu}\phi_{s,t}(f_i) \quad \text{and we set } d^{\mu}_{s,t}(x, y) := D_{\mu}\phi_{s,t}\big(b_t(x, \cdot)\big)(y) \tag{2.18}$$

for any $(x, y) \in \mathbb{R}^{2d}$. A brief review on tensor spaces is provided in Section 3.1. We also consider the function

$$D_{\mu} X^{\mu}_{s,t}(x, y) := \int_s^t \big[\nabla X^{\phi_{s,u}(\mu)}_{u,t}\big]\big(X^{\mu}_{s,u}(x)\big)' d^{\mu}_{s,u}\big(X^{\mu}_{s,u}(x), y\big)\, du.$$

Combining the first order formulae stated in Theorem 2.3 with conventional Taylor expansions we check the following theorem.

THEOREM 2.5. *For any $x \in \mathbb{R}^d$, $\mu_0, \mu_1 \in P_2(\mathbb{R}^d)$ and $s \leq t$ we have the almost sure expansion*

$$X_{s,t}^{\mu_1}(x) - X_{s,t}^{\mu_0}(x) = \int (\mu_1 - \mu_0)(dy) D_{\mu_0} X_{s,t}^{\mu_0}(x, y) + \Delta_{s,t}^{[2],\mu_0,\mu_1}(x) \tag{2.19}$$

with the second order remainder function $\Delta_{s,t}^{[2],\mu_0,\mu_1}$ such that

$$\|\Delta_{s,t}^{[2],\mu_0,\mu_1}\| \leq c e^{-\lambda(t-s)} \mathbb{W}_2(\mu_0, \mu_1)^2 \quad \textit{for some } \lambda > 0.$$

The detailed proof of the above theorem is provided in the Appendix, on page 2659.

Second order expansions are expressed in terms of the functions defined for any $(x, y) \in \mathbb{R}^{2d}$ and for any $z \in \mathbb{R}^{2d}$ by the formulae

$$d_{s,t}^{[1,1],\mu}(x, y) := D_\mu \phi_{s,t}\big(b_t^{[1]}(x, \cdot)'\big)(y) \quad \text{and} \quad d_{s,t}^{[2],\mu}(x, z) := D_\mu^2 \phi_{s,t}\big(b_t(x, \cdot)\big)(z).$$

We associate with these objects the function $D_{\mu_0}^2 X_{s,t}^{\mu_0}$ defined by

$$\begin{aligned} D_\mu^2 X_{s,t}^\mu(x, z) := &\int_s^t [\nabla X_{u,t}^{\phi_{s,u}(\mu)}]\big(X_{s,u}^\mu(x)\big)' \big[d_{s,u}^{[2],\mu}\big(X_{s,u}^\mu(x), z\big) + D_\mu^{[1,1]} X_{s,u}^\mu(x, z)\big] du \\ &+ \int_s^t [\nabla^2 X_{u,t}^{\phi_{s,u}(\mu)}]\big(X_{s,u}^\mu(x)\big)' D_\mu^{[2,1]} X_{s,u}^\mu(x, z)\, du. \end{aligned}$$

In the above display, $D_\mu^{[i,1]} X_{s,u}^\mu$ stands for the functions given by

$$\begin{aligned} D_\mu^{[1,1]} X_{s,u}^\mu(x, z) := &\big[d_{s,u}^{[1,1],\mu}\big(X_{s,u}^\mu(x), z_2\big) D_\mu X_{s,u}^\mu(x, z_1) \\ &+ d_{s,u}^{[1,1],\mu}\big(X_{s,u}^{\mu_0}(x), z_1\big) D_\mu X_{s,u}^\mu(x, z_2)\big], \\ D_{\mu_0}^{[2,1]} X_{s,u}^\mu(x, z) := &\big[D_\mu X_{s,u}^\mu(x, z_1)\, d_{s,u}^\mu\big(X_{s,u}^\mu(x), z_2\big) + D_\mu X_{s,u}^\mu(x, z_2)\, d_{s,u}^\mu\big(X_{s,u}^\mu(x), z_1\big)\big]. \end{aligned}$$

We are now in position to state the main result of this section.

THEOREM 2.6. *For any $x \in \mathbb{R}^d$, $\mu_0, \mu_1 \in P_2(\mathbb{R}^d)$ and $s \leq t$ we have the almost sure expansion*

$$\begin{aligned} &X_{s,t}^{\mu_1}(x) - X_{s,t}^{\mu_0}(x) \\ &\quad = \int (\mu_1 - \mu_0)(dy) D_{\mu_0} X_{s,t}^{\mu_0}(x, y) \\ &\qquad + \frac{1}{2} \int (\mu_1 - \mu_0)^{\otimes 2}(dz) D_{\mu_0}^2 X_{s,t}^{\mu_0}(x, z) + \Delta_{s,t}^{[3],\mu_0,\mu_1}(x) \end{aligned} \tag{2.20}$$

with a third order remainder function $\Delta_{s,t}^{[3],\mu_1,\mu_0}$ such that

$$\|\Delta_{s,t}^{[3],\mu_0,\mu_1}\| \leq c e^{-\lambda(t-s)} \mathbb{W}_2(\mu_0, \mu_1)^3 \quad \textit{for some } \lambda > 0.$$

The proof of the above theorem is provided in the Appendix, on page 2659. In the remainder term estimates presented in the above theorems, we can choose any parameter $\lambda > 0$ such that $\lambda < \lambda_{1,2}$, with the parameter $\lambda_{1,2}$ introduced in (1.10).

2.3.2. *Interacting diffusions.* For any $N \geq 2$, the N-mean field particle interpretation associated with a collection of generators $L_{t,\eta}$ is defined by the Markov process $\xi_t = (\xi_t^i)_{1\leq i\leq N} \in (\mathbb{R}^d)^N$ with generators Λ_t given for any sufficiently smooth function F and any $x = (x^i)_{1\leq i\leq N} \in (\mathbb{R}^d)^N$ by

$$\Lambda_t(F)(x) = \sum_{1\leq i\leq N} L_{t,m(x)}(F_{x^{-i}})(x^i) \tag{2.21}$$

with the function

$$F_{x^{-i}}(y) := F(x^1, \dots, x^{i-1}, y, x^{i+1}, \dots, x^N) \quad \text{and the measure } m(x) = \frac{1}{N}\sum_{1\leq i\leq N} \delta_{x^i}.$$

We extend $L_{t,\mu}$ to symmetric functions $F(x^1, x^2)$ on $\mathbb{R}^{2d}$ by setting

$$L_{t,\mu}^{(2)}(F)(x^1, x^2) := L_{t,\mu}(F(x^1, \cdot))(x^2) + L_{t,\mu}(F(\cdot, x^2))(x^1).$$

In this notation, in our context we readily check that

$$\begin{aligned} \mathcal{F}(x) = m(x)(f) \quad &\Longrightarrow \quad \Lambda_t(\mathcal{F})(x) = m(x)L_{t,m(x)}(f), \\ \mathcal{F}(x) = m(x)^{\otimes 2}(F) \quad &\Longrightarrow \quad \Lambda_t(\mathcal{F})(x) = m(x)^{\otimes 2}L_{t,m(x)}^{(2)}(F) + \frac{1}{N}m(x)[\Gamma(F)] \end{aligned} \tag{2.22}$$

for any symmetric function $F(x^1, x^2) = F(x^2, x^1)$, with the function $\Gamma(F)$ on $\mathbb{R}^d$ defined for any $y \in \mathbb{R}^d$ by the formula

$$\Gamma(F)(y) := \mathrm{Tr}\big(([\nabla \otimes \nabla]F)(y, y)\big) = \sum_{1\leq i\leq d} (\partial_{x_1^i}\partial_{x_2^i}F)(y, y)$$

$$\Longrightarrow \quad \Gamma(f \otimes g)(y) = \sum_{1\leq k\leq d} \partial_{y_k}f(y)\partial_{y_k}g(y) = \mathrm{Tr}\big(\nabla f(y)\nabla g(y)'\big).$$

A proof of the above formula is provided in the Appendix, on page 2647. Applying Ito's formula, for any smooth function $g : t \in [0, \infty[\mapsto g_t \in \mathcal{C}_b^2(\mathbb{R}^d)$ we prove that

$$m_t := m(\xi_t) \quad \Longrightarrow \quad dm_t(g_t) = [m_t(\partial_t g_t) + m_t L_{t,m_t}(g_t)]\,dt + \frac{1}{\sqrt{N}}\,dM_t(g).$$

In the above display, $g \mapsto M_t(g)$ stands for a martingale random field with angle bracket

$$\partial_t\langle M(f), M(g)\rangle_t := m_t\big(\Gamma(f \otimes g)\big) \quad \Longrightarrow \quad \partial_t\langle M(g)\rangle_t = \int m_t(dx)\|\nabla g(x)\|^2.$$

The above evolution equation is rather standard in mean field type interacting particle system theory, a detailed proof can be found in [30] (see, for instance, Section 4.3). In the same vein, with some obvious abusive notation, using (2.22) we have

$$\begin{aligned} dm_s^{\otimes 2}(F) &= [m_s \otimes dm_s + dm_s \otimes m_s + (dm_s \otimes dm_s)](F) \\ &= \left[m_s^{\otimes 2}L_{s,m_s}^{(2)}(F) + \frac{1}{N}m_s[\Gamma(F)]\right]ds + \text{martingale increment} \\ \Longrightarrow \quad &[dm_s \otimes dm_s](F) = \frac{1}{N}m_s[\Gamma(F)]\,ds. \end{aligned}$$

We fix a final time horizon $t \geq 0$ and we denote by

$$s \in [0, t] \mapsto M_s\big(D_{m_\cdot}\phi_{\cdot,t}(f)\big)$$

the martingale associated with the predictable function

$$s \in [0, t] \mapsto g_s = D_{m_s}\phi_{s,t}(f).$$

Combining the Itô formula with the tensor product formula (2.22) and with the backward formula (2.13) we obtain

$$d\phi_{s,t}(m_s)(f) = -m_s L_{s,m_s}\big(D_{m_s}\phi_{s,t}(f)\big)\,ds + (dm_s)\big(D_{m_s}\phi_{s,t}(f)\big) + \frac{1}{2}(dm_s \otimes dm_s)\big(D^2_{m_s}\phi_{s,t}(f)\big)\,ds.$$

This implies that

$$d\phi_{s,t}(m_s)(f) = \frac{1}{2}(dm_s \otimes dm_s)\big(D^2_{m_s}\phi_{s,t}(f)\big)\,ds + \frac{1}{\sqrt{N}}\,dM_s\big(D_{m.}\phi_{.,t}(f)\big).$$

This yields the following theorem.

THEOREM 2.7. *For any time horizon $t \geq 0$, the interpolating semigroup $s \in [0,t] \mapsto \phi_{s,t}(m_s)$ satisfies for any $f \in \mathcal{C}^2(\mathbb{R}^d)$ with $\sup_{k=1,2}\|\nabla^k f\| \leq 1$ the evolution equation*

$$d\phi_{s,t}(m_s)(f) = \frac{1}{2N} m_s\big[\Gamma\big(D^2_{m_s}\phi_{s,t}(f)\big)\big]\,ds + \frac{1}{\sqrt{N}}\,dM_s\big(D_{m.}\phi_{.,t}(f)\big). \tag{2.23}$$

The above theorem can be seen as a second order extension of the Alekseev–Gröbner lemma [1, 42] to nonlinear measure valued and stochastic semigroups. This result also extends the perturbation theorem obtained in [4] (cf. Theorem 3.6) in the context of interacting jumps processes to McKean–Vlasov diffusions. The discrete time version of the backward perturbation analysis described above can also be found in [27, 28, 31] in the context of Feynman–Kac particle models (see also [25, 26, 32]).

We end this Section with some direct consequences of the above theorem. Firstly, using (2.6) and (2.12) we have the almost sure estimates

$$\big|\partial_s\langle M_{.,t}\big(D_{m.}\phi_{.,t}(f)\big)\rangle_s\big| \leq c e^{-2\lambda(t-s)}\|\nabla f\|^2$$
$$\text{and} \quad \big\|m_s\big[\Gamma\big(D^2_{m_s}\phi_{s,t}(f)\big)\big]\big\| \leq c e^{-\lambda(t-s)} \sup_{i=1,2}\|\nabla^i f\| \quad \text{for some } \lambda > 0.$$

Without further work, the above inequality yields the uniform bias estimate stated in the r.h.s. of (1.13), for any twice differentiable function f with bounded derivatives. Using well known martingale concentration inequalities (cf., for instance, Lemma 3.2 in [58]), there exists some finite parameter c such that for any $t \geq 0$ and any $\delta \geq 1$ the probability of the following event:

$$\left|m_t(f) - \phi_{0,t}(m_0)(f) - \frac{1}{2N}\int_0^t m_s\big[\Gamma\big(D^2_{m_s}\phi_{s,t}(f)\big)\big]\,ds\right| \leq c\sqrt{\frac{\delta}{N}}$$

is greater than $1 - e^{-\delta}$. In addition, using the Burkholder–Davis–Gundy inequality, for any $n \geq 1$ we obtain the time uniform estimates stated in the r.h.s. of (1.12). On the other hand, using (2.5) and (2.6) we have the almost sure exponential contraction inequality

$$\mathbb{W}_1\big(\phi_{0,t}(m_0), \phi_{0,t}(\mu_0)\big) \leq c e^{-\lambda t}\,\mathbb{W}_1(m_0,\mu_0) \quad \text{for some } \lambda > 0.$$

This yields the bias estimates

$$\big|\mathbb{E}\big[m_t(f) - \phi_{0,t}(\mu_0)(f)\big]\big| \leq \frac{c_1}{N} + \frac{c_2}{N^{1/d}}e^{-\lambda t}$$

for any twice differentiable function f with bounded derivatives. The r.h.s. estimate comes from well known estimates of the average of the Wassertein distance for occupation measures; see, for instance, [38] and the more recent studies [40, 54]. The above inequality yields the following uniform bias estimate:

$$\sup_{t \geq \frac{d-1}{d\lambda}\log N} \big|\mathbb{E}\big[m_t(f) - \phi_{0,t}(\mu_0)(f)\big]\big| \leq \frac{c}{N}.$$

2.4. *Comments on the regularity conditions.* We discuss in this section the regularity condition (H) introduced in (1.9). We illustrate these spectral conditions for linear-drift and gradient flow models. Comparisons with related conditions presented in other works are also provided.

First, we mention that the condition stated in (1.9) has been introduced in the article [5] to derive several Wasserstein exponential contraction inequalities as well as uniform propagation of chaos estimates w.r.t. the time horizon.

Using the log-norm triangle inequality and recalling that the log-norm is dominated by the spectral norm we check that

$$\lambda_{\max}(A_t(x_1,x_2)_{\mathrm{sym}}) \leq \lambda_{\max}(b_t^{[1]}(x_1,x_2)_{\mathrm{sym}}) + 2^{-1}\|b_t^{[2]}(x_2,x_1)+b_t^{[2]}(x_1,x_2)'\|_2.$$

Choosing λ_0 and λ_1 as the supremum of the maximal eigenvalue functional of the matrices $A_t(x_1,x_2)_{\mathrm{sym}}$ and $b_t^{[1]}(x_1,x_2)_{\mathrm{sym}}$, the Cauchy interlacing theorem (see, for instance, [53] on page 294) yields $\lambda_1 \geq \lambda_0 \geq \lambda_{1,2}$.

For linear drift functions

$$b_t(x_1,x_2) = B_1x_1 + B_2x_2 \tag{2.24}$$

the matrix $A_t(x_1,x_2)_{\mathrm{sym}}$ reduces to the two-by-two block partitioned matrix

$$\begin{aligned} A_t(x_1,x_2)_{\mathrm{sym}} &= \begin{bmatrix} (B_1)_{\mathrm{sym}} & (B_2)_{\mathrm{sym}} \\ (B_2)_{\mathrm{sym}} & (B_1)_{\mathrm{sym}} \end{bmatrix} \\ \Longrightarrow \quad \lambda_0 \geq \lambda_1 &= -\lambda_{\max}((B_1)_{\mathrm{sym}}) \quad \text{and} \quad \|b^{[2]}\|_2 = \|B_2\|_2. \end{aligned} \tag{2.25}$$

In this situation the diffusion flow $X_{s,t}^{\mu}(x) \in \mathbb{R}^d$ is given by the formula

$$X_{s,t}^{\mu}(x) = e^{(t-s)B_1}(x-\mu(e)) + e^{(t-s)[B_1+B_2]}\mu(e) + \int_s^t e^{B_1(t-u)}\,dW_u.$$

In the one-dimensional case we have

$$B_1 < 0 < B_2 \quad \Longrightarrow \quad B_1 = -\lambda_1 \leq B_1 + B_2 = -\lambda_{1,2} = -\lambda_0.$$

Nonlinear Langevin diffusions are associated with the drift function

$$\begin{aligned} &b(x_1,x_2) := -\nabla U(x_1) - \nabla V(x_1-x_2) \\ &\quad\Longrightarrow \quad b^{[1]}(x_1,x_2) = -\nabla^2 U(x_1) - \nabla^2 V(x_1-x_2) \quad \text{and} \quad b^{[2]}(x_1,x_2) = \nabla^2 V(x_1-x_2) \end{aligned}$$

some confinement type potential function U (a.k.a. the exterior potential) and some interaction potential function V. In this context we have

$$\begin{aligned} &-A_t(x_1,x_2)_{\mathrm{sym}} \\ &\quad= \begin{bmatrix} \nabla^2 U(x_1) & 0 \\ 0 & \nabla^2 U(x_2) \end{bmatrix} \\ &\qquad+ \begin{bmatrix} \nabla^2 V(x_1-x_2) & -(\nabla^2 V(x_2-x_1)+\nabla^2 V(x_1-x_2))/2 \\ -(\nabla^2 V(x_2-x_1)+\nabla^2 V(x_1-x_2))/2 & \nabla^2 V(x_2-x_1) \end{bmatrix}. \end{aligned}$$

When the potential function V is even and convex we have

$$A_t(x_1,x_2)_{\mathrm{sym}} \leq -\begin{bmatrix} \nabla^2 U(x_1) & 0 \\ 0 & \nabla^2 U(x_2) \end{bmatrix}.$$

In the reverse angle, when the function V is odd we have the formula

$$A_t(x_1,x_2)_{\mathrm{sym}} = -\begin{bmatrix} \nabla^2 U(x_1)+\nabla^2 V(x_1-x_2) & 0 \\ 0 & \nabla^2 U(x_2)+\nabla^2 V(x_2-x_1) \end{bmatrix}.$$

In both situations, condition (H) is satisfied when the strength of the confinement type potential dominates the one of the interaction potential; that is, when we have that

$$\nabla^2 U(x_1)+\nabla^2 V(x_2) \geq \lambda_1 > \|\nabla^2 V\|_2.$$

The decay rate λ_0 in the $\mathbb{W}_2$-contraction inequality (2.4) is larger than the decay rate $\lambda_{1,2}$ in the $\mathbb{W}_1$-contraction inequality (2.10). In addition, the $\mathbb{W}_1$-exponential stability requires that λ_0 dominates the spectral norm of the matrix $b^{[2]}$. Next we provide a more refined analysis based on the proof of the $\mathbb{W}_2$-contraction inequality presented in [5]. Using the interpolating paths $(Y_\epsilon,\mu_\epsilon)$ introduced in (2.3) we set

$$X^\epsilon_{s,t} := X^{\mu_\epsilon}_{s,t}(Y_\epsilon) \quad \text{and} \quad \overline{X}^\epsilon_{s,t} := \overline{X}^{\mu_\epsilon}_{s,t}(\overline{Y}_\epsilon). \tag{2.26}$$

In the above display $(\overline{X}^{\mu_\epsilon}_{s,t}(x),\overline{Y}_\epsilon)$ stands for an independent copy of $(X^{\mu_\epsilon}_{s,t}(x),Y_\epsilon)$. Arguing as in [5] we have

$$\begin{aligned}\partial_t\mathbb{E}(\|\partial_\epsilon X^\epsilon_{s,t}\|) = \mathbb{E}[\|\partial_\epsilon X^\epsilon_{s,t}\|^{-1}(&\langle\partial_\epsilon X^\epsilon_{s,t}, b^{[1]}(X^\epsilon_{s,t},\overline{X}^\epsilon_{s,t})\partial_\epsilon X^\epsilon_{s,t}\rangle \\ &+\langle\partial_\epsilon \overline{X}^\epsilon_{s,t}, b^{[2]}(X^\epsilon_{s,t},\overline{X}^\epsilon_{s,t})\partial_\epsilon X^\epsilon_{s,t}\rangle)].\end{aligned}$$

We consider the symmetric and anti-symmetric matrices

$$\begin{aligned} b^{[2]}_t(x_1,x_2)_{\mathrm{sym}} &:= \frac{1}{2}\big(b^{[2]}_t(x_1,x_2)+b^{[2]}_t(x_2,x_1)'\big), \\ b^{[2]}_t(x_1,x_2)_{\mathrm{asym}} &:= \frac{1}{2}\big(b^{[2]}_t(x_1,x_2)-b^{[2]}_t(x_2,x_1)'\big)\end{aligned}$$

and we set

$$(U^\epsilon_{s,t},\overline{U}^\epsilon_{s,t}) := \left(\frac{\partial_\epsilon X^\epsilon_{s,t}}{\sqrt{\|\partial_\epsilon X^\epsilon_{s,t}\|}}, \frac{\partial_\epsilon \overline{X}^\epsilon_{s,t}}{\sqrt{\|\partial_\epsilon \overline{X}^\epsilon_{s,t}\|}}\right) \quad \text{and} \quad (V^\epsilon_{s,t},\overline{V}^\epsilon_{s,t}) := \left(\frac{\partial_\epsilon X^\epsilon_{s,t}}{\|\partial_\epsilon X^\epsilon_{s,t}\|}, \frac{\partial_\epsilon \overline{X}^\epsilon_{s,t}}{\|\partial_\epsilon \overline{X}^\epsilon_{s,t}\|}\right).$$

By symmetry arguments and using some elementary manipulations we check the formula

$$\begin{aligned} 2\partial_t\mathbb{E}(\|\partial_\epsilon X^\epsilon_{s,t}\|) = \mathbb{E}\Bigg(&\left\langle \begin{pmatrix} U^\epsilon_{s,t} \\ \overline{U}^\epsilon_{s,t}\end{pmatrix}, A_t(X^\epsilon_{s,t},\overline{X}^\epsilon_{s,t})\begin{pmatrix} U^\epsilon_{s,t} \\ \overline{U}^\epsilon_{s,t}\end{pmatrix}\right\rangle \\ &+\left(\sqrt{\|\partial_\epsilon \overline{X}^\epsilon_{s,t}\|}-\sqrt{\|\partial_\epsilon X^\epsilon_{s,t}\|}\right)^2\langle \overline{V}^\epsilon_{s,t}, b^{[2]}_t(X^\epsilon_{s,t},\overline{X}^\epsilon_{s,t})_{\mathrm{sym}}V^\epsilon_{s,t}\rangle \\ &+(\|\partial_\epsilon \overline{X}^\epsilon_{s,t}\|-\|\partial_\epsilon X^\epsilon_{s,t}\|)\langle \overline{V}^\epsilon_{s,t}, b^{[2]}_t(X^\epsilon_{s,t},\overline{X}^\epsilon_{s,t})_{\mathrm{asym}}V^\epsilon_{s,t}\rangle\Bigg).\end{aligned}$$

This shows that

$$\partial_t\mathbb{E}(\|\partial_\epsilon X^\epsilon_{s,t}\|) \leq -\widehat{\lambda}_{1,2}\mathbb{E}(\|\partial_\epsilon X^\epsilon_{s,t}\|)$$

with the parameter $\widehat{\lambda}_{1,2}$ given by

$$-\widehat{\lambda}_{1,2} := \sup_{x_1,x_2}\big[\lambda_{\max}(A_t(x_1,x_2))+\|b^{[2]}_t(x_1,x_2)_{\mathrm{sym}}\|_2+\|b^{[2]}_t(x_1,x_2)_{\mathrm{asym}}\|_2\big] \leq -\lambda_{1,2}.$$

We conclude that the $\mathbb{W}_1$-contraction inequality (2.10) is met with $\lambda = \widehat{\lambda}_{1,2}$.

In a more recent article [67] the author presents some Wasserstein contraction inequalities of the same form as in (2.4) with λ_0 replaced by some parameter $\lambda_0^- = (\kappa_1 - \kappa_2)$, under the assumption

$$\langle x_1 - y_1, b_t(x_1, \mu_1) - b_t(y_1, \mu_2)\rangle \leq -\kappa_1 \|x_1 - y_1\|^2 + \kappa_2 \mathbb{W}_2(\mu_1, \mu_2)^2 \quad \text{for some } \kappa_1 > \kappa_2.$$

Taking Dirac measures $\mu_1 = \delta_{x_2}$ and $\mu_2 = \delta_{y_2}$ we check that the above condition is equivalent to the fact that

$$\langle x_1 - y_1, b_t(x_1, x_2) - b_t(y_1, y_2)\rangle \leq -\kappa_1 \|x_1 - y_1\|^2 + \kappa_2 \|x_2 - y_2\|^2.$$

By symmetry arguments this implies that

$$\begin{aligned} &\langle x_1 - y_1, b_t(x_1, x_2) - b_t(y_1, y_2)\rangle + \langle x_2 - y_2, b_t(x_2, x_1) - b_t(y_2, y_1)\rangle \\ &\quad \leq -\lambda_0^- [\|x_1 - y_1\|^2 + \|x_2 - y_2\|^2]. \end{aligned} \tag{2.27}$$

For the linear drift model discussed in (2.25) the above condition reads

$$\begin{bmatrix} (B_1)_{\text{sym}} & (B_2)_{\text{sym}} \\ (B_2)_{\text{sym}} & (B_1)_{\text{sym}} \end{bmatrix} \leq -\lambda_0^- I \quad \text{which is implies that} \quad \lambda_0 \geq \lambda_0^-.$$

We also have (2.27) $\Longrightarrow$ (1.7) with $\lambda = \lambda_0^-$.

2.5. *Comparisons with existing literature.* The perturbation analysis developed in the article differs from the Otto differential calculus on $(P_2(\mathbb{R}^d), \mathbb{W}_2)$ introduced in [59] and further developed by Ambrosio and his co-authors [2, 3] and Otto and Villani in [60]. These sophisticated gradient flow techniques in Wasserstein metric spaces are based on optimal transport theory.

The central idea is to interpret $P_2(\mathbb{R}^d)$ as an infinite-dimensional Riemannian manifold. In this context, the Benamou–Brenier formulation of the Wasserstein distance provides a natural way to define geodesics, gradients and Hessians w.r.t. the Wasserstein distance. The details of these gradient flow techniques are beyond the scope of the semigroup perturbation analysis considered herein.

This methodology is mainly used to quantify the entropy dissipation of Langevin-type nonlinear diffusions. Thus, it cannot be used to derive any Taylor expansion of the form (1.4) nor to analyze the stability properties of more general classes of McKean–Vlasov diffusions.

Besides some interesting contact points, the methodology developed in the present article doesn't rely on the more recent differential calculus on $(P_2(\mathbb{R}^d), \mathbb{W}_2)$ developed by P. L. Lions and his co-authors in the seminal works on mean field game theory [14, 43]. In this context, the first order Lions differential of a smooth function from $P_2(\mathbb{R}^d)$ into $\mathbb{R}$ is defined as the conventional derivative of lifted real valued function acting on the Hilbert space of square integrable random variables. In this interpretation, for a given test function, say f the gradient $\nabla D_\mu \phi_{s,t}(f)(Y)$ of the first order differential in (1.4) can be seen as the Lions derivative $(\delta u_{s,t}/\delta\mu)(Y)$ of the lifted scalar function $Y \mapsto u_{s,t}(Y) := \mathbb{E}(f(X_{s,t}^\mu(Y)))$, for some random variable Y with distribution μ.

In the recent book [17], to distinguish these two notions, the authors called the random variable $D_\mu \phi_{s,t}(f)(Y)$ the linear functional derivative. For a more thorough discussion on the origins and the recent developments in mean field game theory, we refer to the book [17] as well as the more recent articles [13, 19, 23] and the references therein.

To the best of our knowledge, most of the literature on Lions' derivatives is concerned with existence theorems without a refined analysis of the exponential decays of these differentials w.r.t. the time parameter. Last but not least, from the practical point of view all differential estimates we found in the literature are rather quite deceiving since after carefully checking, they grow exponentially fast with respect to the time horizon (cf., for instance, [13, 19, 20, 23]).

Taylor expansions of the form (1.4) have already been discussed in the book [26] and in the article [33] for discrete time nonlinear measure valued semigroups (cf., for instance, Chapters 3 and 10 in [26]). We also refer to the more recent articles [4, 34] in the context of continuous time Feynman–Kac semigroups. In this context, we emphasize that the semigroup $\phi_{s,t}(\mu)$ is explicitly given by a normalization of a linear semigroup of positive operators. Thus, a fairly simple Taylor expansion yields the second order formula (1.4). In contrast with Feynman–Kac models, McKean–Vlasov semigroups don't have any explicit form nor an analytical description. As a result, none of above methodologies cannot be used to analyze nonlinear diffusions.

The second order perturbation analysis discussed in this article has been used with success in [27, 28, 31] to analyze the stability properties of Feynman–Kac type particle models, as well as the fluctuations and the exponential concentration of this class of interacting jump processes; see also [29, 35] for general classes of discrete generation mean field particle systems, a well as Chapter 7 in [25] and [4, 32] for continuous time models.

These second order perturbation techniques have also been extended in the seminal book by V. N. Kolokoltsov [50] to general classes of nonlinear Markov processes and kinetic equations. Chapter 8 in [50] is dedicated to the analysis of the first and the second order derivatives of nonlinear semigroups with respect to initial data. The use of the first and the second order derivatives in the analysis of central limit theorems and propagation of chaos properties respectively is developed in Chapter 9 and Chapter 10 in [50]. We underline that these results are obtained for diffusion processes as well as for jump-type processes and their combinations, see also [51, 52].

Nevertheless none of these studies apply to derive nonasymptotic Taylor expansions (2.14) and (2.20) with exponential decay-type remainder estimates for McKean–Vlasov diffusions nor to estimate the stability properties of the associated semigroups. In addition, to the best of our knowledge the stochastic perturbation Theorem 2.7 is the first result of this type for mean field type interacting diffusions.

Last but not least, the idea of considering the flow of empirical measures $m(\xi_t)$ of a mean field particle model as a stochastic perturbation of the limiting flow $\phi_{0,t}(\mu_0)$ certainly goes back to the work by Dawson [24], itself based on the martingale approach developed by Papanicolaou, Stroock and Varadhan in [61], published in the end of the 1970's. These two works are mainly centered on fluctuation type limit theorems. They don't discuss any Taylor expansion on the limiting semigroup $\phi_{s,t}$ nor any question related to the stability properties of the underlying processes.

3. Some preliminary results. The first part of this section provides a review of tensor product theory and Fréchet differential on Hilbert spaces. Section 3.1 is concerned with conventional tensor products and Fréchet derivatives. Section 3.2 provides a short introduction to tensor integral operators.

In the second part of this section we review some basic tools of the theory of stochastic variational equations, including some differential properties of Markov semigroups. Section 3.3 is dedicated to variational equations. Section 3.5 discusses Bismut–Elworthy–Li extension formulae. We also provide some exponential inequalities for the gradient and the Hessian operators on bounded measurable functions.

The differential operator arising in the Taylor expansions (1.4) are defined in terms of tensor integral operators that depend on the gradient of the drift function $b_t(x_1, x_2)$ of the nonlinear diffusion. These integro-differential operators are described in Section 3.6. The last section, Section 3.7 provides some differential formulae as well as some exponential decays estimates of the norm of these operators w.r.t. the time horizon.

3.1. *Fréchet differential.* We let $[n]$ stands for the set of n multiple indexes $i = (i_1, \dots, i_n) \in \mathcal{I}^n$ over some finite set $\mathcal{I}$. Notice that $[n_1] \times [n_2] = [n_1 + n_2]$. We denote by $\mathcal{T}_{p,q}(\mathcal{I})$ the space of (p,q)-tensor X with real entries $(X_{i,j})_{(i,j)\in[p]\times[q]}$. Given a (p_1, q_1)-tensor X and a (p_2, q_2)-tensor Y we denote by $(X \otimes Y)$ the $((p_1 + q_1), (p_2 + q_2))$-tensor defined by

$$(X \otimes Y)_{(i,j),(k,l)} := X_{i,k} Y_{j,l}.$$

For a given (p_1, q)-tensor X and a given (q, p_2) tensor Y, the product XY and the transposition Y' are the (p_1, p_2) and (p_2, q) tensors with entries

$$\forall (i,j) \in [p_1] \times [p_2] \quad (XY)_{i,j} := \sum_{k\in[q]} X_{i,k} Y_{k,j} \quad \text{and} \quad Y'_{j,k} = Y_{k,j}.$$

We equip $\mathcal{T}_{p,q}(\mathcal{I})$ with the Frobenius inner product

$$\langle X, Y\rangle := \operatorname{Tr}(XY') := \sum_{i\in[p]} (XY')_{i,i} \quad \text{and the norm } \|X\|_{\mathrm{Frob}} := \sqrt{\operatorname{Tr}(XX')}.$$

Identifying $(1,0)$-tensors $\mathcal{T}_{1,0}(\mathcal{I}) = \mathbb{R}^{\mathcal{I}}$ with column vectors $(X_i)_{i\in\mathcal{I}} \in \mathbb{R}^{\mathcal{I}}$ the above quantities coincide with the conventional Euclidian inner product and norm on the product space $\mathbb{R}^{\mathcal{I}}$. When $\mathcal{I} = \{1, \dots, d\}$ we simplify notation and we set $\mathbb{R}^d$ instead of $\mathbb{R}^{\{1,\dots,d\}}$. For any tensors X and Y with appropriate dimensions, using Cauchy–Schwartz inequality we check that

$$\langle X, Y\rangle^2 \le \|X\|_{\mathrm{Frob}} \|Y\|_{\mathrm{Frob}} \quad \text{and} \quad \|XY\|_{\mathrm{Frob}} \le \|X\|_{\mathrm{Frob}} \|Y\|_{\mathrm{Frob}}.$$

Let $\mathbb{H}(\mathcal{T}_{p,q}(\mathcal{I})) := \mathbb{L}_2((\Omega, \mathbb{F}, \mathbb{P}), \mathcal{T}_{p,q}(\mathcal{I}))$ be the Hilbert space of $\mathcal{T}_{p,q}(\mathcal{I})$-valued random variables defined on some probability space $(\Omega, \mathbb{F}, \mathbb{P})$, equipped with the inner product

$$\langle X, Y\rangle_{\mathbb{H}(\mathcal{T}_{p,q}(\mathcal{I}))} = \mathbb{E}(\langle X, Y\rangle) \quad \text{and the norm } \|X\|_{\mathbb{H}(\mathcal{T}_{p,q}(\mathcal{I}))} := \langle X, X\rangle^{1/2}_{\mathbb{H}(\mathcal{T}_{p,q}(\mathcal{I}))}$$

induced by the inner product $\langle X, Y\rangle$ on $\mathcal{T}_{p,q}(\mathcal{I})$. We denote by $\mathbb{E}(X) = \mathbb{E}(X_{i,j})_{(i,j)\in[p]\times[q]}$ the entry-wise expected value of a (p,q)-tensor.

When $\mathcal{I} = \{1, \dots, d\}$ and $(p,q) = (1,0)$ the space $\mathbb{H}(\mathcal{T}_{p,q}(\mathcal{I}))$ coincides with be the Hilbert space $\mathbb{H}(\mathbb{R}^d) = \mathbb{L}_2((\Omega, \mathbb{F}, \mathbb{P}), \mathbb{R}^d)$ of square integrable $\mathbb{R}^d$-valued and $\mathbb{F}$-measurable random variables.

We denote by

$$\mathbb{H}_n(\mathcal{T}_{p,q}(\mathcal{I})) := \mathbb{L}_2((\Omega, \mathbb{F}_n, \mathbb{P}), \mathcal{T}_{p,q}(\mathcal{I}))$$

the nondecreasing sequence of Hilbert spaces associated with some increasing filtration $\mathbb{F}_n \subset \mathbb{F}_{n+1}$.

In Landau notation, we recall that a function

$$F : X \in \mathbb{H}_1(\mathcal{T}_{p_1,q_1}(\mathcal{I})) \mapsto F(X) \in \mathbb{H}_2(\mathcal{T}_{p_2,q_2}(\mathcal{J}))$$

is said to be Fréchet differentiable at X if there exists a continuous map

$$X \in \mathbb{H}_1(\mathcal{T}_{p,q}(\mathcal{I})) \mapsto \partial F(X) \in \operatorname{Lin}(\mathbb{H}_1(\mathcal{T}_{p_1,q_2}(\mathcal{I})), \mathbb{H}_2(\mathcal{T}_{p_2,q_2}(\mathcal{J})))$$

such that

$$F(X + Y) = F(X) + \partial F(X) \cdot Y + \mathrm{o}(Y).$$

3.2. *Tensor integral operators.* Let $\mathcal{B}(E, \mathcal{T}_{p,q}(\mathcal{I}))$ be the set of bounded measurable functions from a measurable space E into some tensor space $\mathcal{T}_{p,q}(\mathcal{I})$. Signed measures μ on E act on bounded measurable functions g from E into $\mathbb{R}$. We extend these integral operators to tensor valued functions $g = (g_{i,j})_{(i,j)\in[p]\times[q])} \in \mathcal{B}(E, \mathcal{T}_{p,q}(\mathcal{I}))$ by setting for any $(i, j) \in [p] \times [q]$

$$\mu(g)_{i,j} = \mu(g_{i,j}) := \int \mu(dx) g_{i,j}(x) \quad \text{and we set } \mu(g) := \int \mu(dx) g(x).$$

Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be some pair of measurable spaces. A (p, q)-tensor integral operator

$$\mathcal{Q} : g \in \mathcal{B}(F, \mathcal{T}_{q,r}(\mathcal{I})) \mapsto \mathcal{Q}(g) \in \mathcal{B}(E, \mathcal{T}_{p,r}(\mathcal{I}))$$

is defined for $r \geq 0$ and $g \in \mathcal{B}(\mathcal{F}, \mathcal{T}_{q,r}(\mathcal{I}))$ by the tensor valued and measurable function $\mathcal{Q}(g)$ with entries given $x \in E$ and $(i, j) \in ([p] \times [r])$ by the integral formula

$$\mathcal{Q}(g)_{i,j}(x) = \sum_{k\in[q]} \int_F \mathcal{Q}_{i,k}(x, d\overline{x}) g_{k,j}(\overline{x})$$

for some collection of integral operators $\mathcal{Q}_{i,k}(x_1, dx_2)$ from $\mathcal{B}(E, \mathbb{R})$ into $\mathcal{B}(F, \mathbb{R})$. We also consider the operator norm

$$|||\mathcal{Q}||| := \sup_{\|g\|\leq 1} \|\mathcal{Q}(g)\| \quad \text{for some tensor norm } \|\cdot\|.$$

The tensor product $(\mathcal{Q}^1 \otimes \mathcal{Q}^2)$ of a couple of (p_i, q_i)-tensor integral operators

$$\mathcal{Q}^i : g \in \mathcal{B}(F_i, \mathcal{T}_{q_i,r_i}(\mathcal{I})) \mapsto \mathcal{Q}(g) \in \mathcal{B}(E_i, \mathcal{T}_{p_i,r_i}(\mathcal{I})) \quad \text{with } i = 1, 2$$

is a (p, q)-tensor integral operator

$$\mathcal{Q}^1 \otimes \mathcal{Q}^2 : h \in \mathcal{B}(F, \mathcal{T}_{q,r}(\mathcal{I})) \mapsto \mathcal{Q}(g) \in \mathcal{B}(E, \mathcal{T}_{p,q}(\mathcal{I}))$$

with the product spaces

$$(E, F) := (E_1 \times E_2, F_1 \times F_2) \quad \text{and} \quad (p, q, r) = (p_1 + p_2, q_1 + q_2, r_1 + r_2).$$

The entries of $(\mathcal{Q}^1 \otimes \mathcal{Q}^2)(h)$ are given for any $x = (x_1, x_2)$ and any pair of multi-indices $i = (i_1, i_2) \in ([p_1] \times [p_2])$, $j = (j_1, j_2) \in ([r_1] \times [r_2])$ by the integral formula

$$(\mathcal{Q}^1 \otimes \mathcal{Q}^2)(h)_{i,j}(x) = \sum_{k\in([q_1]\times[q_2])} \int_{F_1\times F_2} (\mathcal{Q}^1 \otimes \mathcal{Q}^2)_{i,k}(x, dy) h_{k,j}(y)$$

with the tensor product measures defined for any $k = (k_1, k_2) \in ([q_1] \times [q_2])$ and any $y = (y_1, y_2)$ by

$$(\mathcal{Q}^1 \otimes \mathcal{Q}^2)_{(i_1,i_2),(k_1,k_2)}((x_1, x_2), d(y_1, y_2)) := \mathcal{Q}^1_{i_1,k_1}(x_1, dy_1) \mathcal{Q}^2_{i_2,k_2}(x_2, dy_2).$$

3.3. *Variational equations.* The gradient and the Hessian of a multivariate smooth function $h(x) = (h_i(x))_{i\in[p]}$ is defined by the $(1, p)$ and $(2, p)$ tensors $\nabla h(x)$ and $\nabla^2 h(x)$ with entries given for any $1 \leq k, l \leq d$ and $i \in [p]$ by the formula

$$\nabla h(x)_{k,i} = \partial_{x_k} h_i(x) \quad \text{and} \quad \nabla^2 h(x)_{(k,l),i} = \partial_{x_k} \partial_{x_l} h_i(x). \tag{3.1}$$

We consider the tensor valued functions $b_t^{[k_1,k_2]}$ and $b_t^{[k_1,k_2,k_3]}$ defined for any $k_1, k_2, k_3 = 1, 2$ by

$$b_t^{[k_1,k_2]} := (\nabla_{x_{k_1}} \otimes \nabla_{x_{k_2}}) b_t \quad \text{and} \quad b_t^{[k_1,k_2,k_3]} := (\nabla_{x_{k_1}} \otimes \nabla_{x_{k_2}} \otimes \nabla_{x_{k_3}}) b_t$$

with the (2, 1) and (3, 1)-tensor valued functions

$$(b_t^{[k_1,k_2]})_{(i_1,i_2),j} = \partial_{x_{k_1}^{i_1}} \partial_{x_{k_2}^{i_2}} b_t^j \quad \text{and} \quad (b_t^{[k_1,k_2,k_3]})_{(i_1,i_2,i_3),j} = \partial_{x_{k_1}^{i_1}} \partial_{x_{k_2}^{i_2}} \partial_{x_{k_3}^{i_3}} b_t^j.$$

In the above display, $\partial_{x_k^i} b_t^j(x_1, x_2)$ stands for the partial derivative of the scalar function $b_t^j(x_1, x_2)$ w.r.t. the coordinate x_k^i, with the drift function $b_t(x_1, x_2)$ from $\mathbb{R}^{2d}$ into $\mathbb{R}^d$ introduced in Section 1.1. In the same vein, $\partial_{x_{k_1}^{i_1}} \partial_{x_{k_2}^{i_2}} b_t^j(x_1, x_2)$ and $\partial_{x_{k_1}^{i_1}} \partial_{x_{k_2}^{i_2}} \partial_{x_{k_3}^{i_3}} b_t^j(x_1, x_2)$ stands for the second and third partial derivatives of $b_t^j(x_1, x_2)$ w.r.t. the coordinates $x_{k_1}^{i_1}$, $x_{k_2}^{i_2}$ and $x_{k_3}^{i_3}$ with $k_1, k_2, k_3 \in \{1, 2\}$.

For any $\mu \in P_2(\mathbb{R}^d)$ and $x_1 \in \mathbb{R}^d$ we also consider the tensor functions

$$b_t^{[1]}(x_1, \mu)_{i,j} := \int \mu(dx_2) \partial_{x_1^i} b_t^j(x_1, x_2),$$
$$b_t^{[1,1]}(x_1, \mu)_{(i_1,i_2),j} := \int \mu(dx_2) \partial_{x_1^{i_1}} \partial_{x_1^{i_2}} b_t^j(x_1, x_2).$$

Recalling that $b_t(x, \phi_{s,t}(\mu))$ has continuous and uniformly bounded derivatives up to the third order, the stochastic flow $x \mapsto X_{s,t}^\mu(x)$ is a twice differentiable function of the initial state x. In addition, when (H) holds the gradient $\nabla X_{s,t}^\mu(x)$ of the diffusion flow $X_{s,t}^\mu(x)$ satifies the $(d \times d)$-matrix valued stochastic diffusion equation

$$\partial_t \nabla X_{s,t}^\mu(x) = \nabla X_{s,t}^\mu(x) b_t^{[1]}(X_{s,t}^\mu(x), \phi_{s,t}(\mu)) \quad \Longrightarrow \quad \|\nabla X_{s,t}^\mu(x)\|_2 \leq e^{-\lambda_1(t-s)}. \tag{3.2}$$

The above estimate is a direct consequence of well-known log-norm estimates for exponential semigroups; see, for instance, [22] as well as Section 1.3 in the recent article [11].

We have the stochastic tensor evolution equation

$$\begin{aligned}
&\partial_t \nabla^2 X_{s,t}^\mu(x) \\
&\quad = \nabla^2 X_{s,t}^\mu(x) b_t^{[1]}(X_{s,t}^\mu(x), \phi_{s,t}(\mu)) + [\nabla X_{s,t}^\mu(x) \otimes \nabla X_{s,t}^\mu(x)] b_t^{[1,1]}(X_{s,t}^\mu(x), \phi_{s,t}(\mu)).
\end{aligned}$$

This implies that

$$\partial_t \|\nabla^2 X_{s,t}^\mu(x)\|_{\mathrm{Frob}}^2 \leq -2\lambda_1 \|\nabla^2 X_{s,t}^\mu(x)\|_{\mathrm{Frob}}^2 + 2\|b^{[1,1]}\|_{\mathrm{Frob}} \|\nabla X_{s,t}^\mu(x)\|_{\mathrm{Frob}}^2 \|\nabla^2 X_{s,t}^\mu(x)\|_{\mathrm{Frob}}$$

from which we check that

$$\partial_t \|\nabla^2 X_{s,t}^\mu(x)\|_{\mathrm{Frob}} \leq -\lambda_1 \|\nabla^2 X_{s,t}^\mu(x)\|_{\mathrm{Frob}} + \|b^{[1,1]}\|_{\mathrm{Frob}} \|\nabla X_{s,t}^\mu(x)\|_{\mathrm{Frob}}^2.$$

Using (3.2), this yields the estimate

$$\|\nabla^2 X_{s,t}^\mu(x)\|_{\mathrm{Frob}} \leq c_1 e^{-\lambda_1(t-s)} \int_s^t e^{\lambda_1(u-s)} \|\nabla X_{s,u}^\mu(x)\|_{\mathrm{Frob}}^2 \, du \leq c_2 e^{-\lambda_1(t-s)}. \tag{3.3}$$

More generally, using the multivariate version of the de Faà di Bruno derivation formula [21] (see also formula (A.1) in the Appendix), for any $n \geq 1$ we also check the uniform estimate

$$\|\nabla^n X_{s,t}^\mu(x)\|_{\mathrm{Frob}} \leq c_n e^{-\lambda_1(t-s)}. \tag{3.4}$$

A detailed proof is provided in the Appendix, on page 2648.

3.4. *Differential of Markov semigroups.* We have the commutation formula

$$\nabla \circ P^{\mu}_{s,t} = \mathcal{P}^{\mu}_{s,t} \circ \nabla \tag{3.5}$$

with the $(1,1)$-tensor integral operator $\mathcal{P}^{\mu}_{s,t}$ defined for any $x \in \mathbb{R}^d$ and any differentiable function f on $\mathbb{R}^d$ by the formula

$$\mathcal{P}^{\mu}_{s,t}(\nabla f)(x) := \mathbb{E}\big[\nabla X^{\mu}_{s,t}(x) \nabla f\big(X^{\mu}_{s,t}(x)\big)\big]. \tag{3.6}$$

The tensor product of $\mathcal{P}^{\mu}_{s,t}$ is also given by the $(2,2)$-tensor integral operator

$$\big(\mathcal{P}^{\mu}_{s,t}\big)^{\otimes 2}(h)(x_1,x_2) := \mathbb{E}\big[\big[\nabla X^{\mu}_{s,t}(x_1) \otimes \nabla \overline{X}^{\mu}_{s,t}(x_2)\big] h\big(X^{\mu}_{s,t}(x_1), \overline{X}^{\mu}_{s,t}(x_2)\big)\big].$$

In the above display, $\overline{X}^{\mu}_{s,t}(x)$ stands for an independent copy of $X^{\mu}_{s,t}(x)$ and $h = (\nabla \otimes \nabla) g$ stands for the matrix valued function defined in (1.14). We also have the commutation formula

$$\big(\mathcal{P}^{\mu}_{s,t}\big)^{\otimes 2} \circ (\nabla \otimes \nabla) = (\nabla \otimes \nabla) \circ \big(P^{\mu_0}_{s,t}\big)^{\otimes 2}.$$

In the same vein, we have the second order differential formula

$$\nabla^2 P^{\mu}_{s,t}(f) = \mathcal{P}^{[2,1],\mu}_{s,t}(\nabla f) + \mathcal{P}^{[2,2],\mu}_{s,t}(\nabla^2 f) \tag{3.7}$$

with the $(2,1)$ and $(2,2)$-tensor integral operators

$$\begin{aligned} &\mathcal{P}^{[2,1],\mu}_{s,t}(\nabla f)(x) := \mathbb{E}\big[\nabla^2 X^{\mu}_{s,t}(x) \nabla f\big(X^{\mu}_{s,t}(x)\big)\big], \\ &\mathcal{P}^{[2,2],\mu}_{s,t}(\nabla^2 f)(x) := \mathbb{E}\big[\big(\nabla X^{\mu}_{s,t}(x) \otimes \nabla X^{\mu}_{s,t}(x)\big) \nabla^2 f\big(X^{\mu}_{s,t}(x)\big)\big] \end{aligned}. \tag{3.8}$$

Iterating the above procedure, the nth differential of $P^{\mu}_{s,t}(f)$ at any order $n \geq 1$ takes the form

$$\nabla^n P^{\mu}_{s,t}(f) = \sum_{1 \leq k \leq n} \mathcal{P}^{[n,k],\mu}_{s,t}(\nabla^k f)$$

for some integral operators $\mathcal{P}^{[n,k],\mu}_{s,t}$. For instance, we have the third order differential formula

$$\nabla^3 P^{\mu}_{s,t}(\nabla f) = \mathcal{P}^{[3,1],\mu}_{s,t}(\nabla f) + \mathcal{P}^{[3,2],\mu}_{s,t}(\nabla^2 f) + \mathcal{P}^{[3,3],\mu}_{s,t}(\nabla^3 f) \tag{3.9}$$

with the $(2,1)$ and $(2,2)$-tensor integral operators

$$\begin{aligned} &\mathcal{P}^{[3,1],\mu}_{s,t}(\nabla f)(x) := \mathbb{E}\big[\nabla^3 X^{\mu}_{s,t}(x) \nabla f\big(X^{\mu}_{s,t}(x)\big)\big], \\ &\mathcal{P}^{[3,2],\mu}_{s,t}(\nabla^2 f)(x) := \mathbb{E}\big[\big(\nabla^2 X^{\mu}_{s,t}(x) \mathbin{\widehat{\otimes}} \nabla X^{\mu}_{s,t}(x)\big) \nabla^2 f\big(X^{\mu}_{s,t}(x)\big)\big], \\ &\mathcal{P}^{[3,3],\mu}_{s,t}(\nabla^3 f)(x) := \mathbb{E}\big[\big(\nabla X^{\mu}_{s,t}(x) \otimes \nabla X^{\mu}_{s,t}(x) \otimes \nabla X^{\mu}_{s,t}(x)\big) \nabla^3 f\big(X^{\mu}_{s,t}(x)\big)\big] \end{aligned} \tag{3.10}$$

with the $\widehat{\otimes}$-tensor product of type $(3,2)$ given for any $i = (i_1, i_2, i_3)$ and $l = (l_1, l_2)$ by

$$\begin{aligned} &\big(\nabla^2 X^{\mu}_{s,t}(x) \mathbin{\widehat{\otimes}} \nabla X^{\mu}_{s,t}(x)\big)_{i,l} \\ &\quad := \big(\nabla^2 X^{\mu}_{s,t}(x) \otimes \nabla X^{\mu}_{s,t}(x)\big)_{((i_1,i_2),i_3),l} \\ &\qquad + \big(\nabla^2 X^{\mu}_{s,t}(x) \otimes \nabla X^{\mu}_{s,t}(x)\big)_{((i_2,i_3),i_1),l} + \big(\nabla^2 X^{\mu}_{s,t}(x) \otimes \nabla X^{\mu}_{s,t}(x)\big)_{((i_3,i_1),i_2),l}. \end{aligned}$$

The above formulae remains valid for any column vector multivariate function $f = (f_i)_{1 \leq i \leq d}$. An explicit description of the integral operators $\mathcal{P}^{[n,k],\mu}_{s,t}$ for any $1 \leq k \leq n$ can be obtained using multivariate derivations and combinatorial manipulations; see, for instance the multivariate version of the de Faà di Bruno derivation formulae (A.1) and (A.2) in the Appendix. Following the proof of (3.4) we also check the uniform estimates

$$\sup_{1 \leq k \leq n} \big\|\big|\mathcal{P}^{[n,k],\mu}_{s,t}\big\|\big| \leq c_n e^{-\lambda_1(t-s)}. \tag{3.11}$$

Using the moment estimates (1.15) for any $\mu \in P_2(\mathbb{R}^d)$, $m, n \geq 0$, and any $s \leq t$, we also check the rather crude estimate

$$\|\!| P^{\mu}_{s,t} \|\!|_{\mathcal{C}^n_m(\mathbb{R}^d)\to\mathcal{C}^n_m(\mathbb{R}^d)} \vee \|\!| (P^{\mu}_{s,t})^{\otimes 2} \|\!|_{\mathcal{C}^n_m(\mathbb{R}^{2d})\to\mathcal{C}^n_m(\mathbb{R}^{2d})} \leq c_{m,n}(t)\big[1 + \|e\|_{\mu,2}\big]^m. \tag{3.12}$$

For instance, using the de Faà di Bruno derivation formula (A.2) for any function $f \in \mathcal{C}^n_m(\mathbb{R}^d)$ such that $\|f\|_{\mathcal{C}^n_m(\mathbb{R}^d)} \leq 1$ and for any $0 \leq k \leq n$ we check that

$$\|\nabla^k P^{\mu}_{s,t}(f)(x)\| = \|\mathbb{E}(\nabla^k(f \circ X^{\mu}_{s,t})(x))\| \leq c_{n,m}(t)\mathbb{E}\big((1 + \|X^{\mu}_{s,t}(x)\|)^m\big).$$

The estimates (1.15) implies that

$$\|\nabla^k P^{\mu}_{s,t}(f)(x)\| \leq c_{n,m}(t)\big(\|x\| + \|e\|_{\mu,2}\big)^m \leq c_{n,m}(t)\big(1 + \|x\|\big)^m\big(1 \vee \|e\|_{\mu,2}\big)^m$$

from which we conclude that

$$\|\!| P^{\mu}_{s,t} \|\!|_{\mathcal{C}^n_m(\mathbb{R}^d)\to\mathcal{C}^n_m(\mathbb{R}^d)} \leq c_{m,n}(t)\big[1 + \|e\|_{\mu,2}\big]^m.$$

3.5. *Bismut–Elworthy–Li extension formulae.* We have the Bismut–Elworthy–Li formula

$$\begin{aligned} &\nabla P^{\mu}_{s,t}(f)(x) = \mathbb{E}\big(f\big(X^{\mu}_{s,t}(x)\big)\tau^{\mu,\omega}_{s,t}(x)\big) \\ &\qquad \text{with } \tau^{\mu,\omega}_{s,t}(x) := \int_s^t \partial_u\omega_{s,t}(u)\nabla X^{\mu}_{s,u}(x)\,dW_u. \end{aligned} \tag{3.13}$$

The above formula is valid for any function $\omega_{s,t} : u \in [s,t] \mapsto \omega_{s,t}(u) \in \mathbb{R}$ of the following form:

$$\omega_{s,t}(u) = \varphi\big((u-s)/(t-s)\big) \quad \Longrightarrow \quad \partial_u\omega_{s,t}(u) = \frac{1}{t-s}\partial\varphi\big((u-s)/(t-s)\big) \tag{3.14}$$

for some nondecreasing differentiable function φ on $[0,1]$ with bounded continuous derivatives and such that

$$\big(\varphi(0), \varphi(1)\big) = (0,1) \quad \Longrightarrow \quad \omega_{s,t}(t) - \omega_{s,t}(s) = 1.$$

In the same vein, for any $s \leq u \leq t$ we have

$$\nabla^2 P^{\mu}_{s,t}(f)(x) = \mathbb{E}\big(f\big(X^{\mu}_{s,t}(x)\big)\big[\tau^{[2],\mu,\omega}_{s,u}(x) + \nabla X^{\mu}_{s,u}(x)\tau^{\phi_{s,u}(\mu),\omega}_{u,t}\big(X^{\mu}_{s,u}(x)\big)\tau^{\mu,\omega}_{s,u}(x)'\big]\big) \tag{3.15}$$

with the stochastic process

$$\tau^{[2],\mu,\omega}_{s,t}(x) := \int_s^t \partial_u\omega_{s,t}(u)\nabla^2 X^{\mu}_{s,u}(x)\,dW_u.$$

Besides the fact that $X^{\mu}_{s,t}(x)$ is a nonlinear diffusion, the proof of the above formula follows the same proof as the one provided in [6, 12, 39, 55, 64] in the context of diffusions on differentiable manifolds. For the convenience of the reader, a detailed proof is provided in the Appendix on page 2650. Using (3.13), for any f s.t. $\|f\| \leq 1$ we check that

$$\begin{aligned} \|\nabla P^{\mu}_{s,t}(f)\|^2 &\leq \mathbb{E}\big(\|\tau^{\mu,\omega}_{s,t}(x)\|^2\big) \\ &\leq \int_s^t e^{-2\lambda_1(u-s)}\|\partial_u\omega^{s,t}(u)\|^2\,du = \frac{1}{t-s}\int_0^1 e^{-2\lambda_1(t-s)v}\big(\partial\varphi(v)\big)^2\,dv. \end{aligned}$$

Let φ_ϵ with $\epsilon \in]0,1[$ be some differentiable function on $[0,1]$ null on $[0, 1-\epsilon]$ and such that $|\partial\varphi_\epsilon(u)| \leq c/\epsilon$ and $(\varphi_\epsilon(1-\epsilon), \varphi(1)) = (0,1)$, for instance, we can choose

$$\varphi(u) = \begin{cases} 0 & \text{if } u \in [0, 1-\epsilon], \\ 1 + \cos\left(\left(1 + \dfrac{1-u}{\epsilon}\right)\dfrac{\pi}{2}\right) & \text{if } u \in [1-\epsilon, 1]. \end{cases}$$

In this situation, we find the rather crude uniform estimate

$$\begin{aligned}&\|\nabla P^{\mu}_{s,t}(f)\|^2 \leq \left(\frac{c}{\epsilon}\right)^2 \frac{1}{t-s}\int_{1-\epsilon}^{1} e^{-2\lambda_1(t-s)v}\,dv\\ \Longrightarrow\quad &\|\nabla P^{\mu}_{s,t}(f)\| \leq \frac{c}{\epsilon}\frac{1}{\sqrt{t-s}}e^{-\lambda_1(1-\epsilon)(t-s)}.\end{aligned}\tag{3.16}$$

In the same vein, combining (3.15) with the estimate (3.3) for any $\epsilon \in]0,1[$ and $u \in]s,t[$ we also check the rather crude uniform estimate

$$\|\nabla^2 P^{\mu}_{s,t}(f)\| \leq \frac{c_1}{\epsilon}\frac{1}{\sqrt{u-s}}e^{-\lambda_1(u-s)(1-\epsilon)} + \frac{c_2}{\epsilon^2}\frac{1}{\sqrt{(t-u)(u-s)}}e^{-\lambda_1(u-s)}e^{-\lambda_1(t-s)(1-\epsilon)}.$$

Choosing $u = s + (1-\epsilon)(t-s)$ in the above display we readily check that

$$\begin{aligned}\|\nabla^2 P^{\mu}_{s,t}(f)\| \leq{}& \frac{c_1}{\epsilon\sqrt{1-\epsilon}}\frac{1}{\sqrt{t-s}}e^{-\lambda_1(1-\epsilon)^2(t-s)}\\ &+\frac{c_2}{\epsilon^2}\frac{1}{\sqrt{\epsilon(1-\epsilon)}}\frac{1}{t-s}e^{-2\lambda_1(t-s)(1-\epsilon)}.\end{aligned}\tag{3.17}$$

3.6. *Integro-differential operators.* Let $\mathbb{B}^{\mu}_{s,t}(x_0,x_1)$ be the matrix-valued function defined for any $(x_0,x_1) \in \mathbb{R}^{2d}$, $\mu \in P_2(\mathbb{R}^d)$ and any $s \leq t$ by the formulae

$$\mathbb{B}^{\mu}_{s,t}(x_0,x_1) := \nabla_{x_0} b^{\mu}_{s,t}(x_0,x_1) \quad \text{with } b^{\mu}_{s,t}(x_0,x_1) := \mathbb{E}\big[b_t\big(x_1, X^{\mu}_{s,t}(x_0)\big)\big].\tag{3.18}$$

For instance, for the linear model discussed in (2.24) we have

$$\begin{aligned}&\mathbb{B}^{\mu}_{s,t}(x_0,x_1)' = B_2 e^{(t-s)B_1} \quad \text{and}\\ &b^{\mu}_{s,t}(x_0,x_1) = B_1 x_1 + B_2\big[e^{(t-s)B_1}\big(x_0-\mu(e)\big) + e^{(t-s)[B_1+B_2]}\mu(e)\big].\end{aligned}$$

We also consider the collection Weyl chambers $[s,t]_n$ defined for any $n \geq 1$ by

$$[s,t]_n := \{u = (u_1,\ldots,u_n) \in [s,t]^n : s \leq u_1 \leq \cdots \leq u_n \leq t\} \quad \text{and set } du := du_1\cdots du_n.$$

We consider the space-time Weyl chambers

$$\Delta_{s,t} := \bigcup_{n\geq 1}\Delta^n_{s,t} \quad \text{with } \Delta^n_{s,t} := [s,t]_n \times \mathbb{R}^{nd}.\tag{3.19}$$

The coordinates of a generic point $(u,y) \in \Delta^n_{s,t}$ for some $n \geq 1$ are denoted by

$$u = (u_1,\ldots,u_n) \in [s,t]_n \quad \text{and} \quad y = (y_1,\ldots,y_n) \in \mathbb{R}^{nd}.$$

We also use the convention $u_0 = s$ and $u_{n+1} = t$. We consider the measures $\Phi_{s,u}(\mu)$ on $\Delta_{s,t}$ given on every set $\Delta^n_{s,t}$ and any $n \geq 1$ by

$$\Phi_{s,u}(\mu)\big(d(u,y)\big) = \phi_{s,u}(\mu)(dy)\,du$$

with the tensor product measures

$$\phi_{s,u}(\mu)(dy) := \phi_{s,u_1}(\mu)(dy_1)\ldots\phi_{s,u_n}(\mu)(dy_n).$$

DEFINITION 3.1. Let $b^{\mu}_{s,u}(x,y)$ be the function defined for any $\mu \in P_2(\mathbb{R}^d)$, $x \in \mathbb{R}^d$, and any $(u,y) \in \Delta^n_{s,t}$ and $n \geq 1$ by the formula

$$b^{\mu}_{s,u}(x,y)' := b^{\mu}_{s,u_1}(x,y_1)' \prod_{1\leq k<n}\mathbb{B}^{\phi_{s,u_k}(\mu)}_{u_k,u_{k+1}}(y_k,y_{k+1}).\tag{3.20}$$

In the above display the product of matrices is understood as a directed product from $k = 1$ to $k = (n-1)$. For instance, for the linear model discussed in (2.24) we have

$$b^{\mu}_{s,u}(x,y) = B_2 e^{(u_n - u_{n-1})B_1} \cdots B_2 e^{(u_2 - u_1)B_1} b^{\mu}_{s,u_1}(x, y_1).$$

For any $x \in \mathbb{R}^d$, and any $(u,y) \in \Delta^n_{s,t}$ and $n \geq 1$ we also set

$$\mathbb{B}^{\phi_{s,u}(\mu)}_{u,t}(y,x) := \mathbb{B}^{\phi_{s,u_n}(\mu)}_{u_n,t}(y_n,x) \quad \text{and} \quad \mathcal{P}^{\phi_{s,u}(\mu)}_{u,t}(\nabla f)(y) := \mathcal{P}^{\phi_{s,u_n}(\mu)}_{u_n,t}(\nabla f)(y_n). \tag{3.21}$$

DEFINITION 3.2. For any $\mu_0, \mu_1 \in P_2(\mathbb{R}^d)$ and $s \leq t$ we let $Q^{\mu_1,\mu_0}_{s,t}$ be the operator defined on differentiable functions f on $\mathbb{R}^d$ by

$$Q^{\mu_1,\mu_0}_{s,t}(f) := \mathcal{Q}^{\mu_1,\mu_0}_{s,t}(\nabla f) \tag{3.22}$$

with the (0, 1)-tensor integral operator $\mathcal{Q}^{\mu_1,\mu_0}_{s,t}$ defined by the integral formula

$$\mathcal{Q}^{\mu_1,\mu_0}_{s,t}(\nabla f)(x) := \int_{\Delta_{s,t}} \Phi_{s,u}(\mu_1)\big(d(u,y)\big) b^{\mu_0}_{s,u}(x,y)' \mathcal{P}^{\phi_{s,u}(\mu_0)}_{u,t}(\nabla f)(y).$$

Recall that $b_t(x,y)$ is differentiable at any order with uniformly bounded derivatives. Thus, using the estimates (1.15) and (3.4), for any $m, n \geq 0$, $\mu_0, \mu_1 \in P_{m\vee 2}(\mathbb{R}^d)$ we have

$$\| Q^{\mu_1,\mu_0}_{s,t} \|_{\mathcal{C}^1_m(\mathbb{R}^d) \to \mathcal{C}^n_1(\mathbb{R}^d)} \leq c_{m,n}(t) \rho_{m\vee 2}(\mu_0, \mu_1). \tag{3.23}$$

DEFINITION 3.3. Let $p^{\mu_1,\mu_0}_{s,t}$ be the function defined for any $s \leq t$ and $x, z \in \mathbb{R}^d$ by the formula

$$p^{\mu_1,\mu_0}_{s,t}(x,z)' = b^{\mu_0}_{s,t}(x,z)' + \int_{\Delta_{s,t}} \Phi_{s,u}(\mu_1)\big(d(u,y)\big) b^{\mu_0}_{s,u}(x,y)' \mathbb{B}^{\phi_{s,u}(\mu_0)}_{u,t}(y,z). \tag{3.24}$$

In this notation, we readily check the following proposition.

PROPOSITION 3.4. *The* (0, 1)*-tensor integral operator* $\mathcal{Q}^{\mu_1,\mu_0}_{s,t}$ *can be rewritten as follows*:

$$\mathcal{Q}^{\mu_1,\mu_0}_{s,t}(\nabla f)(x) = \int_{\Delta^1_{s,t}} \Phi_{s,u}(\mu_1)\big(d(u,y)\big) p^{\mu_1,\mu_0}_{s,u}(x,y)' \mathcal{P}^{\phi_{s,u}(\mu_0)}_{u,t}(\nabla f)(y).$$

For instance, for the linear model discussed in (2.24) the function $p^{\mu_1,\mu_0}_{s,t}(x,z)$ defined in (3.24) reduces to

$$\begin{aligned} p^{\mu_1,\mu_0}_{s,t}(x,z) &= B_1 z + B_2 e^{(t-s)(B_1+B_2)} x \\ &\quad + B_2 \bigg[\int_s^t e^{(t-u)(B_1+B_2)} B_1 e^{(u-s)(B_1+B_2)}\, du\, \mu_1(e) \\ &\quad + \int_s^t e^{(t-u)(B_1+B_2)} B_2 e^{(u-s)(B_1+B_2)}\, du\, \mu_0(e) \bigg]. \end{aligned} \tag{3.25}$$

We check this claim expanding in (3.24) the exponential series coming from the integration over the set $\Delta_{s,t}$. A detailed proof of the above formula is provided in the Appendix on page 2651.

3.7. *Some differential formulae.* The matrix $\nabla_{y_0} b^{\mu}_{s,t}(y_0, y_1)$ defined in (3.18) can alternatively be written as follows

$$\nabla_{y_0} b^{\mu}_{s,t}(y_0, y_1) = \mathcal{P}^{\mu}_{s,t}\big(b^{[2]}_t(y_1, \cdot)\big)(y_0) = \mathbb{E}\big[\nabla X^{\mu}_{s,t}(y_0) b^{[2]}_t\big(y_1, X^{\mu}_{s,t}(y_0)\big)\big].$$

We also have the (2, 1) and (3, 1)-tensor formulae

$$\nabla^2_{y_0} b^{\mu}_{s,t}(y_0, y_1) = \mathcal{P}^{[2,1],\mu}_{s,t}\big(b^{[2]}_t(y_1, \cdot)\big)(y_0) + \mathcal{P}^{[2,2],\mu}_{s,t}\big(b^{[2,2]}_t(y_1, \cdot)\big)(y_0),$$

$$\nabla^3_{y_0} b^{\mu}_{s,t}(y_0, y_1) = \mathcal{P}^{[3,1],\mu}_{s,t}\big(b^{[2]}_t(y_1, \cdot)\big)(y_0) + \mathcal{P}^{[3,2],\mu}_{s,t}\big(b^{[2,2]}_t(y_1, \cdot)\big)(y_0) + \mathcal{P}^{[3,3],\mu}_{s,t}\big(b^{[2,2,2]}_t(y_1, \cdot)\big)(y_0).$$

For any $(u, y) \in \Delta^n_{s,t}$ with $n \geq 1$ and for any $k \geq 1$ we have the $(k, 1)$-tensor formulae

$$\nabla^k_{y_0} b^{\mu}_{s,u}(y_0, y) = \mathbb{B}^{[k],\mu}_{s,u}(y_0, y) := \nabla^k_{y_0} b^{\mu}_{s,u_1}(y_0, y_1) \prod_{1 \leq k < n} \mathbb{B}^{\phi_{s,u_k}(\mu)}_{u_k,u_{k+1}}(y_k, y_{k+1}). \tag{3.26}$$

We consider the $(n, 1)$-tensor valued function

$$q^{[n],\mu_1,\mu_0}_{s,t}(x, z) := \mathbb{B}^{[n],\mu_0}_{s,t}(x, z) + \int_{\Delta_{s,t}} \Phi_{s,u}(\mu_1)\big(d(u, y)\big) \mathbb{B}^{[n],\mu_0}_{s,u}(x, y) \mathbb{B}^{\phi_{s,u}(\mu_0)}_{u,t}(y, z)$$

and we use the convention

$$\mathbb{B}^{[0],\mu_0}_{s,t}(x, z) = b^{\mu_0}_{s,t}(x, z)' \quad \text{so that } q^{[0],\mu_1,\mu_0}_{s,t}(x, z) = p^{\mu_1,\mu_0}_{s,t}(x, z)'.$$

For instance, for the linear model discussed in (2.24) and (3.25) the above objects reduce to

$$q^{[1],\mu_1,\mu_0}_{s,t}(x, y)' = B_2 e^{(B_1+B_2)(t-s)} \quad \text{and} \quad \forall n \geq 2 \quad q^{[n],\mu_1,\mu_0}_{s,t}(x, y) = 0.$$

In this notation, we have the following proposition.

PROPOSITION 3.5. *For any $n \geq 0$ the nth differential of the operator $Q^{\mu_1,\mu_0}_{s,t}$ is given by the formula*

$$\nabla^n Q^{\mu_1,\mu_0}_{s,t}(f) = \mathcal{Q}^{[n],\mu_1,\mu_0}_{s,t}(\nabla f)$$

with the $(n, 1)$-tensor integral operator given by

$$\mathcal{Q}^{[n],\mu_1,\mu_0}_{s,t}(\nabla f)(x) := \int_{\Delta^1_{s,t}} \Phi_{s,u}(\mu_1)\big(d(u, y)\big) q^{[n],\mu_1,\mu_0}_{s,u}(x, y) \mathcal{P}^{\mu_0}_{u,t}(\nabla f)(y). \tag{3.27}$$

In addition, when condition (H) is satisfied for any $n \geq 1$ we have the exponential estimates

$$\big\|\big| \mathcal{Q}^{[n],\mu_1,\mu_0}_{s,t} \big\|\big| \leq c_n e^{-\lambda(t-s)} \quad \textit{for some } \lambda > 0. \tag{3.28}$$

PROOF. The proof of the first assertion follows from (3.24). More precisely, using (3.24) we have

$$\nabla^n_x p^{\mu_1,\mu_0}_{s,t}(x, y) = q^{[n],\mu_1,\mu_0}_{s,t}(x, y).$$

On the other hand, by Proposition 3.4 we also have

$$\begin{aligned}
\nabla^n Q^{\mu_1,\mu_0}_{s,t}(f)(x) &= \nabla^n \mathcal{Q}^{\mu_1,\mu_0}_{s,t}(\nabla f)(x) \\
&= \int_{\Delta^1_{s,t}} \Phi_{s,u}(\mu_1)\big(d(u, y)\big) \nabla^n_x p^{\mu_1,\mu_0}_{s,u}(x, y) \mathcal{P}^{\phi_{s,u}(\mu_0)}_{u,t}(\nabla f)(y) \\
&= \mathcal{Q}^{[n],\mu_1,\mu_0}_{s,t}(\nabla f)(x).
\end{aligned}$$

This ends the proof of the first assertion. When condition (H) is satisfied, for any $x \in \mathbb{R}^d$ and $(u, y) \in \Delta^n_{s,t}$ we have

$$\|\mathbb{B}^{\mu}_{s,t}(y_0, y_1)\|_2 \le \|b^{[2]}\|_2 e^{-\lambda_1(t-s)} \quad \text{and} \quad \|\mathbb{B}^{\mu}_{s,u}(x, y)\|_2 \le \|b^{[2]}\|^n_2 e^{-\lambda_1(u_n-s)}. \tag{3.29}$$

Using (3.4) we also check the uniform estimate

$$\|q^{[n],\mu_1,\mu_0}_{s,t}(x, y)\| \le c_n e^{-\lambda_{1,2}(t-s)}. \tag{3.30}$$

The end of the proof is now a consequence of (3.2). □

PROPOSITION 3.6. *For any $n \ge 0$ any bounded function f on $\mathbb{R}^d$ and for any function ω of the form* (3.14) *we have the Bismut–Elworthy–Li formula*

$$\nabla^n Q^{\mu_1,\mu_0}_{s,t}(f) = \int_{\Delta^1_{s,t}} \Phi_{s,u}(\mu)(d(u, y)) q^{[n],\mu_1,\mu_0}_{s,u}(x, y)\mathbb{E}(f(X^{\mu_0}_{u,t}(y))\tau^{\mu_0,\omega}_{u,t}(y)). \tag{3.31}$$

In the above display, $\tau^{\mu,\omega}_{u,t}(y)$ stands for the stochastic process defined in (3.13). *In addition, when condition (H) is satisfied we have the exponential estimates*

$$\|\nabla^n Q^{\mu_1,\mu_0}_{s,t}(f)\| \le c_n e^{-\lambda(t-s)}\|f\| \quad \textit{for some } \lambda > 0. \tag{3.32}$$

PROOF. The proof of the first assertion is a direct application of the Bismut–Elworthy–Li formula (3.13). More precisely, using (3.13) we have

$$\mathcal{P}^{\mu_0}_{u,t}(\nabla f)(y) = \mathbb{E}(f(X^{\mu_0}_{u,t}(y))\tau^{\mu_0,\omega}_{u,t}(y)).$$

The formula (3.31) is now a direct consequence of (3.27).

We check (3.32) combining (3.16) with (3.30). This ends the proof of the proposition. □

When $n = 1$ we drop the upper index and we write $(\mathbb{B}^{\mu}_{s,u}, q^{\mu_1,\mu_0}_{s,t})$ instead of $(\mathbb{B}^{[1],\mu}_{s,u}, q^{[1],\mu_1,\mu_0}_{s,t})$.

The operators discussed above are indexed by a pair of measures (μ_0, μ_1). To simplify notation, when $\mu_1 = \mu_0 = \mu$ we suppress one of the indices and we write $(Q^{\mu}_{s,t}, \mathcal{Q}^{[n],\mu}_{s,t})$ and $(p^{\mu}_{s,t}, q^{[n],\mu}_{s,t})$ instead of $(Q^{\mu,\mu}_{s,t}, \mathcal{Q}^{[n],\mu,\mu}_{s,t})$ and $(p^{\mu,\mu}_{s,t}, q^{[n],\mu,\mu}_{s,t})$.

4. Tangent processes. The tangent process associated with the diffusion flow $\psi_{s,t}(Y)$ introduced in (1.6) is given for any $U \in \mathbb{H}_s(\mathbb{R}^d)$ by the evolution equation

$$\partial_t(\partial\psi_{s,t}(Y) \cdot U) = \partial B_t(\psi_{s,t}(Y)) \cdot (\partial\psi_{s,t}(Y) \cdot U). \tag{4.1}$$

In the above display, $\partial B_t(X) \in \mathrm{Lin}(\mathbb{H}_t(\mathbb{R}^d), \mathbb{H}_t(\mathbb{R}^d))$ stands for the Fréchet differential of the drift function B_t defined for any $Z \in \mathbb{H}_t(\mathbb{R}^d)$ by

$$\partial B_t(X) \cdot Z = \mathbb{E}(\nabla_{x_1} b_t(X, \overline{X})' Z + \nabla_{x_2} b_t(X, \overline{X})' \overline{Z} | \mathbb{F}_t),$$

where $(\overline{X}, \overline{Z})$ stands for an independent copy of (X, Z).

4.1. *Spectral estimate.* This section is mainly concerned with the proof of Theorem 2.1.

For any pair of random variables $Z_1, Z_2 \in \mathbb{H}_t(\mathbb{R}^d)$ we have the duality formula

$$\langle Z_1, \partial B_t(X) \cdot Z_2\rangle_{\mathbb{H}_t(\mathbb{R}^d)} = \langle \partial B_t(X)^\star \cdot Z_1, Z_2\rangle_{\mathbb{H}_t(\mathbb{R}^d)}$$

with the dual operator $\partial B_t(X)^\star$ defined by the formula

$$\partial B_t(X)^\star \cdot Z_1 := \mathbb{E}(b^{[1]}_t(X, \overline{X}) Z_1 + b^{[2]}_t(\overline{X}, X)\overline{Z}_1 | \mathbb{F}_t).$$

In the above display, $(\overline{X}, \overline{Z}_1)$ stands for an independent copy of (X, Z_1). The symmetric part of $\partial B_t(X)$ is given by the formula

$$\partial B_t(X)_{\text{sym}} := \frac{1}{2}\big[\partial B_t(X) + \partial B_t(X)^\star\big].$$

We are now in position to prove Theorem 2.1.

The first assertion is a direct consequence of the evolution equation

$$2^{-1}\partial_t \|\partial\psi_{s,t}(Y)\cdot U\|^2_{\mathbb{H}_t(\mathbb{R}^d)} = \big\langle (\partial\psi_{s,t}(Y)\cdot U), \partial B_t\big(\psi_{s,t}(Y)\big)_{\text{sym}} \cdot \big(\partial\psi_{s,t}(Y)\cdot U\big)\big\rangle_{\mathbb{H}_t(\mathbb{R}^d)}.$$

Whenever (H) is met we have $\partial B_t(X)_{\text{sym}} \le -\lambda_0 I$ for some $\lambda_0 > 0$. In this situation, the r.h.s. estimate in (2.2) is a direct consequence of (2.1). Given an independent copy $(\overline{X}, \overline{Z}_2)$ of (X, Z_2) we have

$$\begin{aligned} 2\langle Z_1, \partial B_t(X)^\star \cdot Z_2\rangle_{\mathbb{H}_t(\mathbb{R}^d)} &= \mathbb{E}\left(\left\langle \begin{bmatrix} Z_1 \\ \overline{Z}_1 \end{bmatrix}, A_t(X, \overline{X}) \begin{bmatrix} Z_2 \\ \overline{Z}_2 \end{bmatrix}\right\rangle\right) \\ &= 2\langle \partial B_t(X)\cdot Z_1, Z_2\rangle_{\mathbb{H}_t(\mathbb{R}^d)}. \end{aligned}$$

This yields the log-norm estimate

$$A_t(X, \overline{X})_{\text{sym}} \le -\lambda_0 I \quad \Longrightarrow \quad \partial B_t(X)_{\text{sym}} \le -\lambda_0 I.$$

The proof of Theorem 2.1 is now completed.

4.2. *Dyson–Phillips expansions*. In the further development of this section we shall denote by

$$(\overline{\psi}_{s,t}, \overline{U}, \overline{X}^\mu_{s,t}, \overline{Y}) \quad \text{and} \quad (\overline{\psi}^n_{s,t}, \overline{U}^n, \overline{X}^{\mu,n}_{s,t}, \overline{Y}^n)_{n\ge 0}$$

a collection of independent copies of the stochastic flows $(\psi_{s,t}, X^\mu_{s,t})$ and some given $U, Y \in \mathbb{H}_s(\mathbb{R}^d)$. To simplify notation, we also set

$$X_{s,t} := \psi_{s,t}(Y), \qquad \overline{X}_{s,t} := \overline{\psi}_{s,t}(\overline{Y}) \quad \text{and} \quad \overline{X}^n_{s,t} := \overline{\psi}^n_{s,t}(\overline{Y}^n).$$

We are now in position to state and prove the main result of this section.

THEOREM 4.1. *The tangent process $\partial\psi_{s,t}$ is given for any $U \in \mathbb{H}_s(\mathbb{R}^d)$ and any $Y \in \mathbb{H}_s(\mathbb{R}^d)$ with distribution $\mu \in P_2(\mathbb{R}^d)$ by the Dyson–Phillips series*

$$\begin{aligned} \partial\psi_{s,t}(Y)\cdot U &= \nabla X^\mu_{s,t}(Y)' U \\ &\quad + \sum_{n\ge 1} \int_{[s,t]_n} (\nabla X^{\phi_{s,u_n}(\mu)}_{u_n,t})(X_{s,u_n})' \\ &\quad \times \mathbb{E}\left(\left[\prod_{1\le k\le n} \mathbb{B}^{\phi_{s,u_{k-1}}(\mu)}_{u_{k-1},u_k}(\overline{X}^{k-1}_{s,u_{k-1}}, \overline{X}^k_{s,u_k})\right]' \overline{U} \,\middle|\, \mathbb{F}_{u_n}\right) du \end{aligned} \tag{4.2}$$

with the boundary conventions

$$u_0 = s, \qquad \overline{X}^0_{s,u_1} = \overline{X}_{s,u_1} \quad \textit{and} \quad \overline{X}^n_{s,u_n} = X_{s,u_n} \quad \textit{for any } n \ge 1.$$

PROOF. For any $s \le u \le t$ and $x \in \mathbb{R}^d$ we have

$$\partial_t \nabla X^\mu_{s,t}(x)^{-1} = -b^{[1]}_t\big(X^\mu_{s,t}(x), \phi_{s,t}(\mu)\big)\nabla X^\mu_{s,t}(x)^{-1}$$

and

$$\nabla X^\mu_{s,t}(x) = \nabla X^\mu_{s,u}(x)\big(\nabla X^{\phi_{s,u}(\mu)}_{u,t}\big)\big(X^\mu_{s,u}(x)\big).$$

In addition, for any $s \leq u \leq t$ and $x_0, x_1 \in \mathbb{R}^d$ we have

$$\nabla_{x_0} b_t\big(x_1, X_{u,t}^{\phi_{s,u}(\mu)}(x_0)\big) = \nabla X_{u,t}^{\phi_{s,u}(\mu)}(x) b_t^{[2]}\big(x_1, X_{u,t}^{\phi_{s,u}(\mu)}(x_0)\big).$$

Combining the above with (4.1) we check that

$$\begin{aligned}
&\partial_t\big((\nabla X_{s,t}^{\mu}(Y)^{-1})'(\partial\psi_{s,t}(Y)\cdot U)\big) \\
&\quad = (\nabla X_{s,t}^{\mu}(Y)^{-1})'\mathbb{E}\big(\nabla b_t\big(\psi_{s,t}(Y), \overline{X}_{s,t}^{\mu}(\cdot)\big)(\overline{Y})'(\nabla \overline{X}_{s,t}^{\mu}(\overline{Y})^{-1})'(\partial\overline{\psi}_{s,t}(\overline{Y})\cdot\overline{U})|\mathbb{F}_t\big).
\end{aligned}$$

In the above display, $\nabla b_t(\psi_{s,t}(Y), \overline{X}_{s,t}^{\mu}(\cdot))(\overline{Y}) = \nabla h(\overline{Y})$ stands for the gradient of the random function

$$h : x \mapsto h(x) = b_t\big(\psi_{s,t}(Y), \overline{X}_{s,t}^{\mu}(x)\big) \quad \text{evaluated at } x = \overline{Y}.$$

Equivalently, we have

$$\begin{aligned}
&\big(\nabla X_{s,t}^{\mu}(Y)^{-1}\big)'\big(\partial\psi_{s,t}(Y)\cdot U\big) \\
&\quad = U + \int_s^t \big(\nabla X_{s,u}^{\mu}(Y)^{-1}\big)' \\
&\qquad \times \mathbb{E}\big(\nabla b_u\big(\psi_{s,u}(Y), \overline{X}_{s,u}^{\mu}(\cdot)\big)(\overline{Y})'(\nabla \overline{X}_{s,u}^{\mu}(\overline{Y})^{-1})'(\partial\overline{\psi}_{s,u}(\overline{Y})\cdot\overline{U})|\mathbb{F}_u\big)\,du
\end{aligned}$$

and therefore

$$\begin{aligned}
\partial\psi_{s,t}(Y)\cdot U = {}&\big(\nabla X_{s,t}^{\mu}(Y)\big)'U + \int_s^t \big((\nabla X_{u,t}^{\phi_{s,u}(\mu)})(X_{s,u}^{\mu}(Y))\big)' \\
&\times \mathbb{E}\big(\nabla b_u\big(\psi_{s,u}(Y), \overline{X}_{s,u}^{\mu}(\cdot)\big)(\overline{Y})'(\nabla \overline{X}_{s,u}^{\mu}(\overline{Y})^{-1})'(\partial\overline{\psi}_{s,u}(\overline{Y})\cdot\overline{U})|\mathbb{F}_u\big)\,du.
\end{aligned}$$

Now, the end of the proof of (4.2) follows a simple induction, thus it is skipped. □

COROLLARY 4.2. *For any $V \in \mathbb{H}_t(\mathbb{R}^d)$ and for any $Y \in \mathbb{H}_s(\mathbb{R}^d)$ with distribution $\mu \in P_2(\mathbb{R}^d)$ we have*

$$\begin{aligned}
&\partial\psi_{s,t}(Y)^{\star}\cdot V \\
&\quad = \mathbb{E}\big(\nabla X_{s,t}^{\mu}(Y)V|\mathbb{F}_s\big) \\
&\qquad + \sum_{n\geq 1}\int_{[s,t]_n} \mathbb{E}\bigg(\bigg[\prod_{1\leq k\leq n} \mathbb{B}_{u_{k-1},u_k}^{\phi_{s,u_{k-1}}(\mu)}\big(\overline{X}_{s,u_{k-1}}^{k-1}, \overline{X}_{s,u_k}^{k}\big)\bigg](\nabla\overline{X}_{u_n,t}^{\phi_{s,u_n}(\mu)})(\overline{X}_{s,u_n})\overline{V}\bigg|\mathbb{F}_s\bigg)\,du
\end{aligned} \tag{4.3}$$

with the boundary conditions

$$u_0 = s \quad \textit{and} \quad \overline{X}_{s,u_1}^{0} = \psi_{s,u_1}(Y) \quad \textit{and} \quad \overline{X}_{s,u_n}^{n} = \overline{X}_{s,u_n}.$$

4.3. *Gradient semigroup analysis.* This section is concerned with a gradient semigroup description of the dual of the tangent process.

DEFINITION 4.3. For any $\mu_0, \mu_1 \in P_2(\mathbb{R}^d)$ and $s \leq t$ we let $D_{\mu_1,\mu_0}\phi_{s,t}$ be the operator defined on differentiable functions f on $\mathbb{R}^d$ by

$$D_{\mu_1,\mu_0}\phi_{s,t} := P_{s,t}^{\mu_0} + Q_{s,t}^{\mu_1,\mu_0}. \tag{4.4}$$

In the above display, $Q_{s,t}^{\mu_1,\mu_0}$ stands for the operator defined in (3.22).

Rewritten in terms of expectation operators we have

$$D_{\mu_1,\mu_0}\phi_{s,t}(f)(x)$$
$$=\mathbb{E}[(f\circ X^{\mu_0}_{s,t})(x)]+\sum_{n\geq 1}\int_{\Delta^n_{s,t}}\Phi_{s,u}(\mu_1)(d(u,y))\mathbb{E}[b^{\mu_0}_{s,u}(x,y)'\nabla(f\circ X^{\phi_{s,u_n}(\mu_0)}_{u_n,t})(y_n)].$$

Recall that $b_t(x,y)$ is differentiable at any order with uniformly bounded derivatives. Thus, arguing as in the proof of (3.12) and (3.23) for any $m,n\geq 1$, $\mu_0,\mu_1\in P_{m\vee 2}(\mathbb{R}^d)$ we have

$$\|D_{\mu_1,\mu_0}\phi_{s,t}\|_{\mathcal{C}^n_m(\mathbb{R}^d)\to\mathcal{C}^n_m(\mathbb{R}^d)}\leq c_{m,n}(t)\rho_{m\vee 2}(\mu_0,\mu_1). \tag{4.5}$$

In the same vein, we check that

$$\|(D_{\mu_1,\mu_0}\phi_{s,t})^{\otimes 2}\|_{\mathcal{C}^{n+1}_m(\mathbb{R}^{2d})\to\mathcal{C}^n_{m+1}(\mathbb{R}^{2d})}\leq c_{m,n}(t)\rho_{m\vee 2}(\mu_0,\mu_1). \tag{4.6}$$

The proof of the above estimate is rather technical, thus it is housed in the Appendix on page 2651.

REMARK 4.4. Using the Bismut–Elworthy–Li formula (3.31), we extend the operators $D_{\mu_1,\mu_0}\phi_{s,t}$ with $s<t$ to nonnecessarily differentiable and bounded functions.

We also extend the operator $D_{\mu_1,\mu_0}\phi_{s,t}$ to tensor functions $f=(f_i)_{i\in[n]}$ by considering the tensor function with entries

$$D_{\mu_1,\mu_0}\phi_{s,t}(f)_i=D_{\mu_1,\mu_0}\phi_{s,t}(f_i). \tag{4.7}$$

In this situation, the function $p^{\mu_1,\mu_0}_{s,t}$ introduced in (3.24) takes the form

$$p^{\mu_1,\mu_0}_{s,t}(x,z)=D_{\mu_1,\mu_0}\phi_{s,t}(b_t(z,\cdot))(x).$$

Let G_{t,μ_1} be the collection of integro-differential operators indexed by $\mu_1\in P_2(\mathbb{R}^d)$ defined by

$$G_{t,\mu_1}(f)(x_2):=\int\mu_1(dx_1)b_t(x_1,x_2)'\nabla f(x_1).$$

We also set

$$H_{t,\mu_0,\mu_1}:=L_{t,\mu_0}+G_{t,\mu_1}\quad\text{and}\quad H_{t,\mu_0}:=L_{t,\mu_0}+G_{t,\mu_0}.$$

In this notation, we have the first order expansion

$$\mu_1L_{t,\mu_1}-\mu_0L_{t,\mu_0}=(\mu_1-\mu_0)L_{t,\mu_0}+(\mu_1-\mu_0)G_{t,\mu_1}=(\mu_1-\mu_0)H_{t,\mu_0,\mu_1}. \tag{4.8}$$

THEOREM 4.5. *For any $m,n\geq 1$ and any $\mu_0,\mu_1\in P_{m\vee 2}(\mathbb{R}^d)$ the operator $D_{\mu_1,\mu_0}\phi_{s,t}$ coincides with the evolution semigroup of the integro-differential operator $H_{t,\phi_{s,t}(\mu_0),\phi_{s,t}(\mu_1)}$; that is, we have the forward evolution equation*

$$\partial_t D_{\mu_1,\mu_0}\phi_{s,t}=D_{\mu_1,\mu_0}\phi_{s,t}\circ H_{t,\phi_{s,t}(\mu_0),\phi_{s,t}(\mu_1)}\quad\text{on } \mathcal{C}^{n\vee 2}_m(\mathbb{R}^d). \tag{4.9}$$

In addition, for any $s\leq u<t$ we have the backward evolution equation

$$\partial_u D_{\phi_{s,u}(\mu_1),\phi_{s,u}(\mu_0)}\phi_{u,t}=-H_{u,\phi_{s,u}(\mu_0),\phi_{s,u}(\mu_1)}\circ D_{\phi_{s,u}(\mu_1),\phi_{s,u}(\mu_0)}\phi_{u,t}\quad\text{on } \mathcal{C}^n_m(\mathbb{R}^d). \tag{4.10}$$

PROOF. The proof of the forward equation (4.9) is a direct consequence of the forward evolution equation

$$\partial_t P^{\mu_0}_{s,t}=P^{\mu_0}_{s,t}L_{t,\phi_{s,t}(\mu_0)}$$

associated with the Markov semigroup $P^{\mu_0}_{s,t}$, thus it is skipped. The semigroup property (2.9) yields

$$\partial_u (D_{\mu_1,\mu_0}\phi_{s,u} \circ D_{\phi_{s,u}(\mu_1),\phi_{s,u}(\mu_0)}\phi_{u,t}) = 0.$$

Combining the above with the forward equation (4.9) we check that

$$\begin{aligned} & D_{\mu_1,\mu_0}\phi_{s,u} \circ \partial_u D_{\phi_{s,u}(\mu_1),\phi_{s,u}(\mu_0)}\phi_{u,t} \\ & \quad = -D_{\mu_1,\mu_0}\phi_{s,u} \circ H_{u,\phi_{s,u}(\mu_0),\phi_{s,u}(\mu_1)} \circ D_{\phi_{s,u}(\mu_1),\phi_{s,u}(\mu_0)}\phi_{u,t}. \end{aligned}$$

This implies that

$$[\partial_u D_{\phi_{s,u}(\mu_1),\phi_{s,u}(\mu_0)}\phi_{u,t}]_{u=s} = -H_{s,\mu_0,\mu_1} D_{\mu_1,\mu_0}\phi_{u,t}$$

from which we conclude that

$$\begin{aligned} & [\partial_u D_{\phi_{s,u}(\mu_1),\phi_{s,u}(\mu_0)}\phi_{u,t}]_{u=v} \\ & \quad = [\partial_u D_{\phi_{v,u}(\phi_{s,v}(\mu_1)),\phi_{v,u}(\phi_{s,v}(\mu_0))}\phi_{u,t}]_{u=v} = -H_{s,\phi_{s,v}(\mu_0),\phi_{s,v}(\mu_1)} D_{\phi_{s,v}(\mu_1),\phi_{s,v}(\mu_0)}\phi_{v,t}. \end{aligned}$$

This yields the backward evolution equation (4.10). This ends the proof of the theorem. □

Next proposition is a direct consequence of (4.4) combined with the formulae (3.5) and (3.27).

PROPOSITION 4.6. *We have the commutation formula*

$$\nabla \circ D_{\mu_1,\mu_0}\phi_{s,t} = \mathcal{D}_{\mu_1,\mu_0}\phi_{s,t} \circ \nabla \tag{4.11}$$

with the $(1,1)$-*tensor integral operator given by the column vector function*

$$\begin{aligned} \mathcal{D}_{\mu_1,\mu_0}\phi_{s,t}(\nabla f)(x) &:= \mathcal{P}^{\mu_0}_{s,t}(\nabla f)(x) \\ & \quad + \int_{\Delta^1_{s,t}} \Phi_{s,v}(\mu_1)\big(d(v,y)\big) q^{\mu_1,\mu_0}_{s,v}(x,y) \mathcal{P}^{\phi_{s,v}(\mu_0)}_{v,t}(\nabla f)(y). \end{aligned} \tag{4.12}$$

In addition, when condition (H) *is satisfied we have*

$$|||\mathcal{D}_{\mu_1,\mu_0}\phi_{s,t}||| \le c e^{-\lambda(t-s)} \quad \textit{for some } \lambda > 0. \tag{4.13}$$

REMARK 4.7. Following Remark 4.4, using the Bismut–Elworthy–Li formula (3.31), we extend the gradient operators $\nabla D_{\mu_1,\mu_0}\phi_{s,t}$ with $s < t$ to measurable and bounded functions. The exponential estimate stated in (3.32) are a direct consequence of the estimates presented in (3.32).

By (4.7) the commutation formula (4.11) is also satisfied for multivariate column functions f. In this situation $\mathcal{D}_{\mu_1,\mu_0}\phi_{s,t}(\nabla f)$ is a $(d \times d)$-matrix valued function.

The proof of Theorem 2.2 is now a consequence of the estimate (4.13) and the fact that

$$\partial_t[\phi_{s,t}(\mu_1) - \phi_{s,t}(\mu_0)] = [\phi_{s,t}(\mu_1) - \phi_{s,t}(\mu_0)] \circ H_{t,\phi_{s,t}(\mu_0),\phi_{s,t}(\mu_1)}.$$

More precisely, using (4.8) the above formula implies that

$$\begin{aligned} & \partial_u\big([\phi_{s,u}(\mu_1) - \phi_{s,u}(\mu_0)] D_{\phi_{s,u}(\mu_1),\phi_{s,u}(\mu_0)}\phi_{u,t}\big) = 0 \\ & \quad \Longrightarrow \quad \phi_{s,t}(\mu_1) - \phi_{s,t}(\mu_0) = (\mu_1 - \mu_0) D_{\mu_1,\mu_0}\phi_{s,t}. \end{aligned}$$

The operators discussed above are indexed by a pair of measures (μ_0, μ_1). To simplify notation, when $\mu_1 = \mu_0 = \mu$ we suppress one of the parameters and we write $(D_\mu\phi_{s,t}, \mathcal{D}_\mu\phi_{s,t})$ instead of $(D_{\mu,\mu}\phi_{s,t}, \mathcal{D}_{\mu,\mu}\phi_{s,t})$.

THEOREM 4.8. *For any $m, n \geq 1$, any function $f \in \mathcal{C}_m^n(\mathbb{R}^d)$ and any $Y \in \mathbb{H}_s(\mathbb{R}^d)$ with distribution $\mu \in P_2(\mathbb{R}^d)$ we have the gradient formula*

$$\partial \psi_{s,t}(Y)^\star \cdot \nabla f(\psi_{s,t}(Y)) = \nabla D_\mu \phi_{s,t}(f)(Y) = \mathcal{D}_\mu \phi_{s,t}(\nabla f)(Y).$$

PROOF. Given a smooth function f on $\mathbb{R}^d$ we have

$$\langle \nabla f(\psi_{s,t}(Y)), \partial \psi_{s,t}(Y) \cdot U \rangle_{\mathbb{H}_t(\mathbb{R}^d)} = \langle \partial \psi_{s,t}(Y)^\star \cdot \nabla f(\psi_{s,t}(Y)), U \rangle_{\mathbb{H}_s(\mathbb{R}^d)}.$$

Replacing V by $\nabla f(\psi_{s,t}(Y))$ in (4.3) we check that

$$\begin{aligned}
&\partial \psi_{s,t}(Y)^\star \cdot \nabla f(\psi_{s,t}(Y)) \\
&\quad = \nabla D_\mu \phi_{s,t}(f)(Y) \\
&\quad = \mathbb{E}(\nabla(f \circ X_{s,t}^\mu)(Y) | Y) \\
&\qquad + \sum_{n \geq 1} \int_{[s,t]_n} \overline{\mathbb{E}}\left(\left[\prod_{0 \leq k < n} \mathbb{B}_{u_k,u_{k+1}}^{\phi_{s,u_k}(\mu)}(\overline{X}_{s,u_k}^k, \overline{X}_{s,u_{k+1}}^{k+1})\right] \nabla(f \circ X_{u_n,t}^{\phi_{s,u_n}(\mu)})(\overline{X}_{s,u_n}) \Big| Y\right) du.
\end{aligned}$$

This ends the proof of the theorem □

5. Taylor expansions. This section is mainly concerned with the proof of the first and second order Taylor expansions stated in Theorem 2.3 and Theorem 2.4. Section 5.1 presents some preliminary differential formulae used in the proof of the theorems.

5.1. *Some differential formulae.* The commutation formula (4.11) takes the form

$$\nabla D_{\mu_1,\mu_0} \phi_{s,t}(f) = \mathcal{D}_{\mu_1,\mu_0} \phi_{s,t}(\nabla f).$$

Combining (4.4) with Proposition 3.5 and the second order formula (3.7) we also have

$$\begin{aligned}
\nabla^2 D_{\mu_1,\mu_0} \phi_{s,t}(f) &= \nabla^2 P_{s,t}^{\mu_0}(f) + \mathcal{Q}_{s,t}^{[2],\mu_1,\mu_0}(\nabla f) \\
&= \mathcal{P}_{s,t}^{[2,1],\mu}(\nabla f) + \mathcal{P}_{s,t}^{[2,2],\mu}(\nabla^2 f) + \mathcal{Q}_{s,t}^{[2],\mu_1,\mu_0}(\nabla f).
\end{aligned}$$

In summary, we have the first and second order differential formulae

$$\begin{aligned}
\nabla D_\mu \phi_{s,t}(f) &= \mathcal{D}_\mu \phi_{s,t}(\nabla f), \\
\nabla^2 D_\mu \phi_{s,t}(f) &= \mathcal{D}_\mu \phi_{s,t}^{[2,1]}(\nabla f) + \mathcal{P}_{s,t}^{[2,2],\mu}(\nabla^2 f) \\
&\quad \text{with } \mathcal{D}_\mu \phi_{s,t}^{[2,1]} = \mathcal{P}_{s,t}^{[2,1],\mu} + \mathcal{Q}_{s,t}^{[2],\mu}.
\end{aligned} \tag{5.1}$$

Similar formulae for $\nabla D_{\mu_0,\mu_1} \phi_{s,t}$ and $\nabla^2 D_{\mu_0,\mu_1} \phi_{s,t}$ can easily be found. In the same vein, using (3.9) we check the third order differential formula

$$\begin{aligned}
&\nabla^3 D_\mu \phi_{s,t}(f) \\
&\quad = \mathcal{D}_\mu \phi_{s,t}^{[3,1]}(\nabla f) + \mathcal{P}_{s,t}^{[3,2],\mu}(\nabla^2 f) + \mathcal{P}_{s,t}^{[3,3],\mu}(\nabla^3 f) \\
&\quad \text{with } \mathcal{D}_\mu \phi_{s,t}^{[3,1]} := \mathcal{P}_{s,t}^{[3,1],\mu} + \mathcal{Q}_{s,t}^{[3],\mu}.
\end{aligned} \tag{5.2}$$

In addition, when condition (H) is satisfied we have the exponential estimates

$$\|\!|\mathcal{D}_\mu \phi_{s,t}\|\!| \vee \|\!|\mathcal{D}_\mu \phi_{s,t}^{[2,1]}\|\!| \vee \|\!|\mathcal{D}_\mu \phi_{s,t}^{[3,1]}\|\!| \leq c e^{-\lambda(t-s)} \quad \text{for some } \lambda > 0. \tag{5.3}$$

DEFINITION 5.1. We let $S^{\mu}_{s,t}$ be the operator defined for any differentiable function f on $\mathbb{R}^d$ by

$$S^{\mu}_{s,t}(f) = \mathcal{S}^{\mu}_{s,t}(\nabla f)$$

with the (0, 1)-tensor integral operator $\mathcal{S}^{\mu}_{s,t}$ defined by the formula

$$\mathcal{S}^{\mu}_{s,t}(\nabla f)(x_1, x_2) := b_s(x_1, x_2)'\mathcal{D}_\mu \phi_{s,t}(\nabla f)(x_1) + b_s(x_2, x_1)'\mathcal{D}_\mu \phi_{s,t}(\nabla f)(x_2). \tag{5.4}$$

Using (4.5) and (4.12) for any $m, n \geq 0$ and $\mu \in P_{m\vee 2}(\mathbb{R}^d)$ we check that

$$\| S^{\mu}_{s,t} \|_{\mathcal{C}^{n+1}_m(\mathbb{R}^d) \to \mathcal{C}^n_{m+1}(\mathbb{R}^{2d})} \leq c_{m,n}(t) \rho_{m\vee 2}(\mu). \tag{5.5}$$

We also have the differential formula

$$(\nabla \otimes \nabla)\big(S^{\mu}_{s,t}(f)\big) = \mathbb{S}^{[2,1],\mu}_{s,t}(\nabla f) + \mathbb{S}^{[2,2],\mu}_{s,t}(\nabla^2 f) \tag{5.6}$$

with the matrix valued functions

$$\begin{aligned}
\mathbb{S}^{[2,1],\mu}_{s,t}(\nabla f)(x_1, x_2) &= b^{[1,2]}_s(x_1, x_2)\mathcal{D}_\mu \phi_{s,t}(\nabla f)(x_1) + b^{[2,1]}_s(x_2, x_1)\mathcal{D}_\mu \phi_{s,t}(\nabla f)(x_2) \\
&\quad + b^{[2]}_s(x_2, x_1)\mathcal{D}_\mu \phi^{[2,1]}_{s,t}(\nabla f)(x_2)' + \mathcal{D}_\mu \phi^{[2,1]}_{s,t}(\nabla f)(x_1) b^{[2]}_s(x_1, x_2)', \\
\mathbb{S}^{[2,2],\mu}_{s,t}(\nabla^2 f)(x_1, x_2) &:= b^{[2]}_s(x_2, x_1)\mathcal{P}^{[2,2],\mu}_{s,t}(\nabla^2 f)(x_2)' + \mathcal{P}^{[2,2],\mu}_{s,t}(\nabla^2 f)(x_1) b^{[2]}_s(x_1, x_2)'.
\end{aligned}$$

When condition (H) is satisfied we also have the exponential estimates

$$\left\|\mathbb{S}^{[2,1],\mu}_{s,t}\right\| \vee \left\|\mathbb{S}^{[2,2],\mu}_{s,t}\right\| \leq c e^{-\lambda(t-s)} \quad \text{for some } \lambda > 0. \tag{5.7}$$

In addition, using the Bismut–Elworthy–Li extension formulae and the estimates (2.7) and (2.8), or any bounded measurable function f on $\mathbb{R}^d$ we check that

$$\|(\nabla \otimes \nabla)\big(S^{\mu}_{s,t}(f)\big)\| \leq c\big(1 \vee 1/(t-s)\big) e^{-\lambda(t-s)} \|f\| \quad \text{for some } \lambda > 0.$$

5.2. *A first order expansion.* This section is mainly concerned with the proof of Theorem 2.3. The next technical lemma is pivotal.

LEMMA 5.2. *For any $m \geq 1$ for any $\mu_0, \mu_1 \in P_{m+1}(\mathbb{R}^d)$ we have the second order expansion*

$$\begin{aligned}
&\phi_{s,t}(\mu_1) - \phi_{s,t}(\mu_0) \\
&\quad = (\mu_1 - \mu_0) D_{\mu_0} \phi_{s,t} \\
&\qquad + \frac{1}{2} \int_s^t [\phi_{s,u}(\mu_1) - \phi_{s,u}(\mu_0)]^{\otimes 2} \circ S^{\phi_{s,u}(\mu_0)}_{u,t} \, du \quad \text{on } \mathcal{C}^{n+1}_m(\mathbb{R}^d).
\end{aligned} \tag{5.8}$$

PROOF. Combining (4.8) with the backward evolution equation (4.10) we check that

$$\begin{aligned}
&\partial_u \{[\phi_{s,u}(\mu_1) - \phi_{s,u}(\mu_0)] \circ D_{\phi_{s,u}(\mu_0)} \phi_{u,t}\} \\
&\quad = [\phi_{s,u}(\mu_1) - \phi_{s,u}(\mu_0)] \circ [H_{u,\phi_{s,u}(\mu_0),\phi_{s,u}(\mu_1)} - H_{u,\phi_{s,u}(\mu_0)}] \circ D_{\phi_{s,u}(\mu_0)} \phi_{u,t} \\
&\quad = [\phi_{s,u}(\mu_1) - \phi_{s,u}(\mu_0)] \circ [G_{u,\phi_{s,u}(\mu_1)} - G_{u,\phi_{s,u}(\mu_0)}] \circ D_{\phi_{s,u}(\mu)} \phi_{u,t}.
\end{aligned}$$

On the other hand, we have

$$[G_{u,\phi_{s,u}(\mu_1)} - G_{u,\phi_{s,u}(\mu_0)}](x_2) := \int \big(\phi_{s,u}(\mu_1) - \phi_{s,u}(\mu_0)\big)(dx_1) b_u(x_1, x_2)' \nabla f(x_1).$$

Integrating u from $u = s$ to $u = t$ we obtain the formula

$$\begin{aligned}&\big[\phi_{s,t}(\mu_1) - \phi_{s,t}(\mu_0) - (\mu_1 - \mu_0) D_{\mu_0}\phi_{s,t}\big](f)\\&\quad = \frac{1}{2}\int_s^t \int \big[\phi_{s,u}(\mu_1) - \phi_{s,u}(\mu_0)\big]^{\otimes 2}\big(d(x_1, x_2)\big)\\&\qquad \times \big[b_u(x_1, x_2)' \nabla D_{\phi_{s,u}(\mu_0)}\phi_{u,t}(f)(x_1) + b_u(x_2, x_1)' \nabla D_{\phi_{s,u}(\mu_0)}\phi_{u,t}(f)(x_2)\big]\,du.\end{aligned}$$

The end of the lemma is now completed. □

Combining the above lemma with (4.6) and (5.5) we check (2.11) with the operator $D^2_{\mu_1,\mu_0}\phi_{s,t}$ defined for any $m, n \geq 0$ and $\mu_0, \mu_1 \in P_{m+2}(\mathbb{R}^d)$ by

$$D^2_{\mu_1,\mu_0}\phi_{s,t} := \int_s^t (D_{\mu_1,\mu_0}\phi_{s,u})^{\otimes 2} \circ S^{\phi_{s,u}(\mu_0)}_{u,t}\, du \in \mathrm{Lin}\big(\mathcal{C}^{n+2}_m(\mathbb{R}^d), \mathcal{C}^n_{m+2}(\mathbb{R}^{2d})\big). \tag{5.9}$$

REMARK 5.3. The second order term in (2.11) can alternatively be expressed in terms of the Hessian of the semigroup $D^2_{\mu_1,\mu_0}\phi_{s,t}$; that is, we have that

$$\begin{aligned}&(\mu_1 - \mu_0)^{\otimes 2} D^2_{\mu_1,\mu_0}\phi_{s,t}(f)\\&\quad = \int_{[0,1]^2} \mathbb{E}\big(\big\langle\big[(\nabla \otimes \nabla) D^2_{\mu_1,\mu_0}\phi_{s,t}(f)\big](Y_{\epsilon,\overline{\epsilon}}), (Y_1 - Y_0) \otimes (\overline{Y}_1 - \overline{Y}_0)\big\rangle\big)\, d\epsilon\, d\overline{\epsilon}\end{aligned} \tag{5.10}$$

with the interpolating path

$$Y_{\epsilon,\overline{\epsilon}} := \big(Y_0 + \epsilon(Y_1 - Y_0), \overline{Y}_0 + \overline{\epsilon}(\overline{Y}_1 - \overline{Y}_0)\big).$$

In the above display, $(\overline{Y}_1, \overline{Y}_0)$ stands for an independent copy of a pair of random variables (Y_0, Y_1) with distribution (μ_0, μ_1). Also observe that

$$(\mu_1 - \mu_0)^{\otimes 2} D^2_{\mu_1,\mu_0}\phi_{s,t} = (\mu_1 - \mu_0)^{\otimes 2} \overline{D}^2_{\mu_1,\mu_0}\phi_{s,t}$$

with the centered second order operator

$$\begin{aligned}&\overline{D}^2_{\mu_1,\mu_0}\phi_{s,t}(f)(x_1, x_2)\\&\quad := \big[(\delta_{x_1} - \mu_0) \otimes (\delta_{x_2} - \mu_0)\big] D^2_{\mu_0}\phi_{s,t}(f)\\&\quad = \int_{[0,1]^2} \mathbb{E}\big(\big\langle\big[(\nabla \otimes \nabla) D^2_{\mu_1,\mu_0}\phi_{s,t}(f)\big]\big(Y_{\epsilon,\overline{\epsilon}}(x_1, x_2)\big), (x_1 - Y_0) \otimes (x_2 - \overline{Y}_0)\big\rangle\big)\, d\epsilon\, d\overline{\epsilon}.\end{aligned}$$

In the above display, $Y_{\epsilon,\overline{\epsilon}}(x_1, x_2)$ stands for the interpolating path

$$Y_{\epsilon,\overline{\epsilon}}(x_1, x_2) := \big(Y_0 + \epsilon(x_1 - Y_0), \overline{Y}_0 + \overline{\epsilon}(x_2 - \overline{Y}_0)\big).$$

PROPOSITION 5.4. *We have commutation formula*

$$(\nabla \otimes \nabla) \circ (D_{\mu_1,\mu_0}\phi_{s,t})^{\otimes 2} = (\mathcal{D}_{\mu_1,\mu_0}\phi_{s,t})^{\otimes 2} \circ (\nabla \otimes \nabla). \tag{5.11}$$

In addition, we have the estimate

$$\big|\!\big|\!\big|(\mathcal{D}_{\mu_1,\mu_0}\phi_{s,t})^{\otimes 2}\big|\!\big|\!\big| \leq c e^{-\lambda(t-s)} \quad \textit{for some } \lambda > 0. \tag{5.12}$$

PROOF. The proof of the first assertion is a consequence of the commutation formula (4.11). Letting $h = (\nabla \otimes \nabla) g$ we have

$$\begin{aligned}
&(\mathcal{D}_{\mu_1,\mu_0}\phi_{s,t})^{\otimes 2}(h)(x_1,x_2)\\
&\quad = (\mathcal{P}^{\mu_0}_{s,t})^{\otimes 2}(h)(x_1,x_2)\\
&\qquad + \int_{\Delta^1_{s,t}} \Phi_{s,v}(\mu_1)(d(u,y)) q^{\mu_1,\mu_0}_{s,u}(x_2,y)(\mathcal{P}^{\mu_0}_{s,t} \otimes \mathcal{P}^{\phi_{s,u}(\mu_0)}_{u,t})(h)(x_1,y)\\
&\qquad + \int_{\Delta^1_{s,t}} \Phi_{s,v}(\mu_1)(d(u,y)) q^{\mu_1,\mu_0}_{s,u}(x_1,y)(\mathcal{P}^{\phi_{s,u}(\mu_0)}_{u,t} \otimes \mathcal{P}^{\mu_0}_{s,t})(h)(y,x_2)\\
&\qquad + \int_{\Delta^1_{s,t}\times\Delta^1_{s,t}} \Phi_{s,u}(\mu_1)(d(u,y))\Phi_{s,v}(\mu_1)(d(v,z))\\
&\qquad \times [q^{\mu_1,\mu_0}_{s,u}(x_1,y) \otimes q^{\mu_1,\mu_0}_{s,v}(x_2,z)](\mathcal{P}^{\phi_{s,u}(\mu_0)}_{u,t} \otimes \mathcal{P}^{\phi_{s,v}(\mu_0)}_{v,t})(h)(y,z).
\end{aligned}$$

The proof of (5.12) now follows the same arguments as the ones we used in the proof of (4.13), thus it is skipped. This ends the proof of the proposition. □

Combining (5.6) with the commutation formula (5.11), for any twice differentiable function f and any $s \le t$ and $\mu_0, \mu_1 \in P_2(\mathbb{R}^d)$ we check that

$$\begin{aligned}
&(\nabla \otimes \nabla) D^2_{\mu_0,\mu_1}\phi_{s,t}(f)\\
&\quad := \int_s^t (\mathcal{D}_{\mu_0,\mu_1}\phi_{s,u})^{\otimes 2}(\mathbb{S}^{[2,1],\phi_{s,u}(\mu_0)}_{u,t}(\nabla f) + \mathbb{S}^{[2,2],\phi_{s,u}(\mu_0)}_{u,t}(\nabla^2 f))\,du
\end{aligned} \tag{5.13}$$

with the operators $\mathbb{S}^{[2,k],\mu}_{s,t}$ discussed in (5.6). The proof of (2.12) is a direct consequence of (5.7) and (5.12). The proof of Theorem 2.3 is now completed.

5.3. *Second order analysis.* This short section is mainly concerned with the proof of the first part of Theorem 2.4.

LEMMA 5.5. *For any $m \ge 1$ and $\mu_0, \mu_1 \in P_{m+3}(\mathbb{R}^d)$ and $s \le t$ we have the tensor product formula*

$$\begin{aligned}
&(\phi_{s,t}(\mu_1) - \phi_{s,t}(\mu_0))^{\otimes 2}\\
&\quad = (\mu_1-\mu_0)^{\otimes 2}(D_{\mu_0}\phi_{s,t})^{\otimes 2} + (\mu_1-\mu_0)^{\otimes 3}\mathcal{R}_{\mu_1,\mu_0}\phi_{s,t} \quad \text{on } \mathcal{C}^{n+2}_m(\mathbb{R}^{2d})
\end{aligned}$$

for some third order linear operator $\mathcal{R}_{\mu_1,\mu_0}\phi_{s,t}$ such that

$$\|\mathcal{R}_{\mu_1,\mu_0}\phi_{s,t}\|_{\mathcal{C}^{n+2}_m(\mathbb{R}^{2d})\to\mathcal{C}^n_{m+3}(\mathbb{R}^{3d})} \le c_{m,n}(t)\rho_{m+2}(\mu_0,\mu_1).$$

The proof of the above lemma is rather technical, thus it is housed in the Appendix, on page 2652.

Combining the above lemma with (5.8) we readily check the second order decomposition (2.14) with a the remainder linear operator $D^3_{\mu_0,\mu_1}\phi_{s,t}$ such that

$$\|D^3_{\mu_0,\mu_1}\phi_{s,t}\|_{\mathcal{C}^{n+3}_m(\mathbb{R}^d)\to\mathcal{C}^n_{m+4}(\mathbb{R}^{3d})} \le c_{m,n}(t)\rho_{m+3}(\mu_0,\mu_1).$$

This ends the proof of the first part of Theorem 2.4. The proof of the second part of the theorem is provided in the Appendix, on page 2654.

APPENDIX

PROOF OF (2.22). It is easy to check that this first assertion is true for any collection of generators $L_{t,\mu}$, thus we skip the details. The proof of the second assertion is a also a direct consequence of a more general result which is valid for any collection of generators and nonnecessarily symmetric functions.

For any $N \geq 2$ and $x = (x^i)_{1\leq i\leq N} \in (\mathbb{R}^d)^N$ we set

$$m(x)^{\odot 2} := \frac{1}{N(N-1)} \sum_{1\leq i\neq j\leq N} \delta_{(x^i,x^j)} \quad \text{and} \quad \mathcal{F}(x) = m(x)^{\otimes 2}(F).$$

We extend $L_{t,\mu}$ to functions $F(x^1, x^2)$ on $\mathbb{R}^{2d}$ by setting

$$\begin{aligned} &L^{(2)}_{t,\mu}(F)(x^1, x^2) \\ &\quad = \frac{1}{2}\big(L_{t,\mu}\big(F(x^1, \cdot)\big)(x^2) + L_{t,\mu}\big(F(\cdot, x^2)\big)(x^1) \\ &\qquad + L_{t,\mu}\big(F(x^2, \cdot)\big)(x^1) + L_{t,\mu}\big(F(\cdot, x^1)\big)(x^2)\big). \end{aligned}$$

For any function $F(x^1, x^2)$ on $\mathbb{R}^{2d}$ we have

$$m(x)^{\otimes 2}(F) = \left(1 - \frac{1}{N}\right) m(x)^{\odot 2}(F) + \frac{1}{N} m(x)\big(C(F)\big) = m(x)^{\odot 2}\left(\left(1 - \frac{1}{N}\right) F + \frac{1}{N} C^{(2)}(F)\right)$$

with

$$C(F)(x) = F(x, x) \quad \text{and} \quad C^{(2)}(F)(x^1, x^2) = \frac{1}{2}\big(C(F)(x^1) + C(F)(x^2)\big).$$

This implies that

$$\Lambda_t(\mathcal{F})(x) = \left(1 - \frac{1}{N}\right) m(x)^{\odot 2}\big(L^{(2)}_{t,m(x)}(F)\big) + \frac{1}{N} m(x)^{\odot 2}\big(L^{(2)}_{t,m(x)}\big(C^{(2)}(F)\big)\big).$$

Recalling that

$$m(x)^{\odot 2}(F) = \frac{N}{N-1} m(x)^{\otimes 2}(F) - \frac{1}{N-1} m(x)^{\odot 2}\big(C^{(2)}(F)\big)$$

we conclude that

$$\Lambda_t(\mathcal{F})(x) = m(x)^{\otimes 2}\big(L^{(2)}_{t,m(x)}(F)\big) + \frac{1}{N} m(x)^{\odot 2}\big(\Gamma^{(2)}_{L_{t,m(x)}}(F)\big)$$

with the operator

$$\Gamma^{(2)}_{L_{t,m(x)}} = L^{(2)}_{t,m(x)} \circ C^{(2)} - C^{(2)} \circ L^{(2)}_{t,m(x)}.$$

Observe that

$$\begin{aligned} &\Gamma^{(2)}_{L_{t,m(x)}}(F)(x^1, x^2) \\ &\quad = \frac{1}{2}\big(L_{t,m(x)}\big(C(F)\big)(x^1) + L_{t,m(x)}\big(C(F)\big)(x^2)\big) \\ &\qquad - \frac{1}{2}\big(L_{t,m(x)}\big(F(x^1, \cdot)\big)(x^1) + L_{t,m(x)}\big(F(\cdot, x^1)\big)(x^1) \\ &\qquad + L_{t,m(x)}\big(F(x^2, \cdot)\big)(x^2) + L_{t,m(x)}\big(F(\cdot, x^2)\big)(x^2)\big). \end{aligned}$$

This yields the formula

$$\Gamma^{(2)}_{L_{t,m(x)}}(F)(x^1,x^2)=\frac{1}{2}\big(C\big(\Gamma^{(2)}_{L_{t,m(x)}}(F)\big)(x^1)+C\big(\Gamma^{(2)}_{L_{t,m(x)}}(F)\big)(x^2)\big)$$

from which we conclude that

$$\Lambda_t(\mathcal{F})(x)=m(x)^{\otimes 2}\big(L^{(2)}_{t,m(x)}(F)\big)+\frac{1}{N}m(x)\big(\Gamma_{L_{t,m(x)}}(F)\big)$$

with the function $\Gamma_{L_{t,m(x)}}(F)$ defined for any $y\in\mathbb{R}^d$ by

$$\begin{aligned}\Gamma_{L_{t,m(x)}}(F)(y)&=C\big(\Gamma^{(2)}_{L_{t,m(x)}}(F)\big)(y,y)\\&=L_{t,m(x)}\big(C(F)\big)(y)-L_{t,m(x)}\big(F(y,\cdot)\big)(y)-L_{t,m(x)}\big(F(\cdot,y)\big)(y).\end{aligned}$$

The above formula readily implies (2.22) as soon as $L_{t,\mu}$ is the collection of generators associated with the stochastic flow defined in (1.1). This ends the proof of (2.22). □

PROOF OF (3.4). For any given $1\le m\le n$, we denote by $\Pi_{n,m}$ the set of partitions $\pi=\{\pi_1,\dots,\pi_m\}$ of the set $\{1,\dots,n\}$ with m blocks π_i of size $|\pi_i|$, with $i\in\{1,\dots,m\}$. We also let Π_n the set of partitions of the set $\{1,\dots,n\}$ and $\flat(\pi)$ the number of blocks in a given partition π, and Π_n^+ the subset of partitions π s.t. $\flat(\pi)>1$.

Let $[n]$ be the set of m multiple indexes $i=(i_1,\dots,i_n)\in\{1,\dots,d\}^n$. For any given $i\in[n]$ and any subset $S=\{j_1,\dots,j_s\}\subset\{1,\dots,n\}$ we set

$$i_S=(i_{j_1},\dots,i_{j_s}).$$

For any $x=(x^1,\dots,x^d)\in\mathbb{R}^d$ and any multiple index $i\in[n]$ we write ∂_i instead of $\partial_{x^{i_1},\dots x^{i_n}}=\partial_{x^{i_1}}\cdots\partial_{x^{i_n}}$ the nth partial derivatives w.r.t. the coordinates $(x^{i_1},\dots,x^{i_n})$.

Let f and X be a couple of smooth functions from $\mathbb{R}^d$ into itself. In this notation for any $i\in[n]$ and $1\le j\le d$ we have the multivariate Faà di Bruno derivation formula

$$\partial_i(f^j\circ X)=\sum_{1\le m\le n}\sum_{k\in[m]}\partial_k f^j(X)\sum_{\pi\in\Pi_{n,m}}(\nabla^\pi X)_{i,k}$$

with the π-gradient tensor

$$(\nabla^\pi X)_{i,k}:=\big(\nabla^{|\pi_1|}X\big)_{i_{\pi_1},k_1}\cdots\big(\nabla^{|\pi_m|}X\big)_{i_{\pi_m},k_m}.$$

We check the above formula by induction w.r.t. the parameter n. In a more compact form we have checked the following lemma.

LEMMA A.1. *For any $n\ge 1$ we have the Faà di Bruno derivation formula*

$$\nabla^n(f\circ X)=\sum_{\pi\in\Pi_n}(\nabla^\pi X)\big(\nabla^{\flat(\pi)}f\big)(X).\tag{A.1}$$

Whenever $X(x)$ is a random function we have

$$P(f)(x):=\mathbb{E}\big((f\circ X)(x)\big)\quad\Longrightarrow\quad\nabla^n P(f)=\sum_{1\le m\le n}\mathcal{P}^{[n,m]}(\nabla^m f)\tag{A.2}$$

with the collection of integral operators

$$\mathcal{P}^{[n,m]}(\nabla^m f)(x):=\sum_{\pi\in\Pi_{n,m}}\mathbb{E}\big(\big(\nabla^\pi X(x)\big)\nabla^m f\big(X(x)\big)\big).$$

Using the above lemma we also check the stochastic tensor evolution equation

$$\begin{aligned}&\partial_t\big(\nabla^n X^{\mu}_{s,t}(x)\big)_{i,j}\\&\quad=\big(\nabla^n X^{\mu}_{s,t}(x)b^{[1]}_t\big(X^{\mu}_{s,t}(x),\phi_{s,t}(\mu)\big)\big)_{i,j}\\&\qquad+\sum_{1<m\le n}\sum_{k\in[m]}\sum_{\pi\in\Pi_{n,m}}\big(\nabla^{\pi}X^{\mu}_{s,t}(x)\big)_{i,k}b_t^{[1]_{\flat(\pi)}}\big(x,\phi_{s,t}(\mu)\big)_{k,j}:\end{aligned}$$

with

$$b_t^{[1]_m}(x,\mu)_{(k_1,\dots,k_m),j}:=\partial_{k_1,\dots,k_m}b^j_t(x,\mu).$$

In a more compact form we have

$$\partial_t\nabla^n X^{\mu}_{s,t}(x)=\nabla^n X^{\mu}_{s,t}(x)b^{[1]}_t\big(X^{\mu}_{s,t}(x),\phi_{s,t}(\mu)\big)+\sum_{\pi\in\Pi_n^+}\nabla^{\pi}X^{\mu}_{s,t}(x)b_t^{[1]_{\flat(\pi)}}\big(x,\phi_{s,t}(\mu)\big).$$

This implies that

$$\begin{aligned}&\partial_t\nabla^n X^{\mu}_{s,t}(x)\nabla^n X^{\mu}_{s,t}(x)'\\&\quad=\nabla^n X^{\mu}_{s,t}(x)\big(b^{[1]}_t\big(X^{\mu}_{s,t}(x),\phi_{s,t}(\mu)\big)+b^{[1]}_t\big(X^{\mu}_{s,t}(x),\phi_{s,t}(\mu)\big)'\big)\nabla^n X^{\mu}_{s,t}(x)'\\&\qquad+\sum_{\pi\in\Pi_n^+}\nabla^{\pi}X^{\mu}_{s,t}(x)\big(b_t^{[1]_{\flat(\pi)}}\big(x,\phi_{s,t}(\mu)\big)+b_t^{[1]_{\flat(\pi)}}\big(x,\phi_{s,t}(\mu)\big)'\big)\nabla^n X^{\mu}_{s,t}(x)'.\end{aligned}$$

Taking the trace in the above display, we check that

$$\begin{aligned}\partial_t\|\nabla^n X^{\mu}_{s,t}(x)\|^2_{\mathrm{Frob}}&\le-2\lambda_1\|\nabla^n X^{\mu}_{s,t}(x)\|^2_{\mathrm{Frob}}\\&\quad+2\|\nabla^n X^{\mu}_{s,t}(x)\|_{\mathrm{Frob}}\sum_{\pi\in\Pi_n^+}\|b^{[1]_{\flat(\pi)}}\|_{\mathrm{Frob}}\|\nabla^{\pi}X^{\mu}_{s,t}(x)\|_{\mathrm{Frob}}.\end{aligned}$$

This yields the rather crude estimate

$$\begin{aligned}&\partial_t\|\nabla^n X^{\mu}_{s,t}(x)\|^2_{\mathrm{Frob}}\\&\quad\le-2\lambda_1\|\nabla^n X^{\mu}_{s,t}(x)\|^2_{\mathrm{Frob}}\\&\qquad+c_n\|\nabla^n X^{\mu}_{s,t}(x)\|_{\mathrm{Frob}}\sum_{\pi\in\Pi_n^+}\|\nabla^{|\pi_1|}X^{\mu}_{s,t}(x)\|_{\mathrm{Frob}}\cdots\|\nabla^{|\pi_{\flat(\pi)}|}X^{\mu}_{s,t}(x)\|_{\mathrm{Frob}}\end{aligned}$$

from which we check that

$$\begin{aligned}&\partial_t\|\nabla^n X^{\mu}_{s,t}(x)\|_{\mathrm{Frob}}\\&\quad\le-\lambda_1\|\nabla^n X^{\mu}_{s,t}(x)\|_{\mathrm{Frob}}+c_n\sum\|\nabla X^{\mu}_{s,t}(x)\|^{l_1}_{\mathrm{Frob}}\|\nabla^2 X^{\mu}_{s,t}(x)\|^{l_2}_{\mathrm{Frob}}\cdots\|\nabla^{n-1}X^{\mu}_{s,t}(x)\|^{l_{n-1}}_{\mathrm{Frob}}.\end{aligned}$$

The summation in the above display is taken over all indices $l_1,\dots,l_{n-1}$ such that $l_1+\cdots+l_{n-1}=m$ and $l_1+2l_2+\cdots+(n-1)l_{n-1}=n$ and $1<m\le n$. Assume that (3.4) has been checked up to rank $(n-1)$. In this case, we have

$$\|\nabla^n X^{\mu}_{s,t}(x)\|_{\mathrm{Frob}}\le c_{n,1}e^{-\lambda_1(t-s)}\int_s^t e^{\lambda_1(u-s)}e^{-2\lambda_1(u-s)}\,du\le c_{n,2}e^{-\lambda_1(t-s)}.$$

This ends the proof of (3.4). □

PROOF OF (3.13) AND (3.15). We recall the backward formula

$$P^{\mu}_{s,t}(f)(x) = f(x) + \int_s^t L_{u,\phi_{s,u}(\mu)}\big(P^{\phi_{s,u}(\mu)}_{u,t}(f)\big)(x)\,du.$$

A detailed proof of the above formula based on backward stochastic flows can be found in Theorem 3.1 in the article [5]. This implies that

$$d\big(P^{\phi_{s,u}(\mu)}_{u,t}(f)\big(X^{\mu}_{u,t}(x)\big)\big) = \big(\nabla P^{\phi_{s,u}(\mu)}_{u,t}(f)\big)\big(X^{\mu}_{u,t}(x)\big)'\,dW_u$$

from which we check that

$$f\big(X^{\mu}_{s,t}(x)\big) = P^{\mu}_{s,t}(f)(x) + \int_s^t \big(\nabla P^{\phi_{s,u}(\mu)}_{u,t}(f)\big)\big(X^{\mu}_{u,t}(x)\big)'\,dW_u.$$

This yields the formula

$$\begin{aligned}
&\mathbb{E}\big(f\big(X^{\mu}_{s,t}(x)\big)\tau^{\mu,\omega}_{s,t}(x)\big)\\
&\quad = \mathbb{E}\left(\left(\int_s^t \big(\nabla P^{\phi_{s,u}(\mu)}_{u,t}(f)\big)\big(X^{\mu}_{u,t}(x)\big)'\,dW_u\right)\left(\int_s^t \partial_u\omega_{s,t}(u)\nabla X^{\mu}_{s,u}(x)\,dW_u\right)\right)\\
&\quad = \mathbb{E}\left(\int_s^t \nabla\big(P^{\phi_{s,u}(\mu)}_{u,t}(f)\circ X^{\mu}_{u,t}\big)(x)\partial_u\omega_{s,t}(u)\,du\right).
\end{aligned}$$

We conclude that

$$\begin{aligned}
&\mathbb{E}\big(f\big(X^{\mu}_{s,t}(x)\big)\tau^{\mu,\omega}_{s,t}(x)\big)\\
&\quad = \nabla P^{\mu}_{s,t}(f)(x)\mathbb{E}\left(\int_s^t \partial_u\omega_{s,t}(u)\,du\right) = \nabla P^{\mu}_{s,t}(f)(x)\big(\omega_{s,t}(t)-\omega_{s,t}(s)\big) = \nabla P^{\mu}_{s,t}(f)(x).
\end{aligned}$$

This ends the proof of (3.13). For any $s \le u \le t$ applying (3.13) to the function $P^{\phi_{s,u}(\mu)}_{u,t}(f)$ we have

$$\begin{aligned}
\nabla P^{\mu}_{s,t}(f)(x) &= \nabla P^{\mu}_{s,u}\big(P^{\phi_{s,u}(\mu)}_{u,t}(f)\big)(x)\\
&= \mathbb{E}\left(P^{\phi_{s,u}(\mu)}_{u,t}(f)\big(X^{\mu}_{s,u}(x)\big)\int_s^u \partial_v\omega_{s,u}(v)\nabla X^{\mu}_{s,v}(x)\,dW_v\right).
\end{aligned}$$

This implies that

$$\begin{aligned}
&\partial_{x_j,x_i}P^{\mu}_{s,t}(f)(x)\\
&\quad = \sum_{1\le l\le d}\mathbb{E}\left(\partial_{x_j}X^{\mu,l}_{s,u}(x)\partial_{x_l}\big(P^{\phi_{s,u}(\mu)}_{u,t}(f)\big)\big(X^{\mu}_{s,u}(x)\big)\int_s^u \partial_v\omega_{s,u}(v)\partial_{x_i}X^{\mu,k}_{s,v}(x)\,dW^k_v\right)\\
&\qquad + \mathbb{E}\left(P^{\phi_{s,u}(\mu)}_{u,t}(f)\big(X^{\mu}_{s,u}(x)\big)\int_s^u \partial_v\omega_{s,u}(v)\partial_{x_j,x_i}X^{\mu,k}_{s,v}(x)\,dW^k_v\right).
\end{aligned}$$

Applying (3.13) to the first term we check that

$$\begin{aligned}
&\mathbb{E}\big(\partial_{x_j}X^{\mu,l}_{s,u}(x)\partial_{x_l}\big(P^{\phi_{s,u}(\mu)}_{u,t}(f)\big)\big(X^{\mu}_{s,u}(x)\big)\tau^{\mu,\omega}_{s,u}(x)_i\big)\\
&\quad = \mathbb{E}\Bigg(f\big(X^{\mu}_{s,t}(x)\big)\partial_{x_j}X^{\mu,l}_{s,u}(x)\tau^{\mu,\omega}_{s,u}(x)_i\\
&\qquad\quad \times\left(\sum_{1\le m\le d}\int_u^t \partial_v\omega_{u,t}(v)\big(\partial_{x_l}X^{\phi_{s,u}(\mu),m}_{u,v}\big)\big(X^{\mu}_{s,u}(x)\big)\,dW^m_v\right)\Bigg)\\
&\quad = \mathbb{E}\big(f\big(X^{\mu}_{s,t}(x)\big)\partial_{x_j}X^{\mu,l}_{s,u}(x)\tau^{\phi_{s,u}(\mu),\omega}_{u,t}\big(X^{\mu}_{s,u}(x)\big)_l\tau^{\mu,\omega}_{s,u}(x)_i\big).
\end{aligned}$$

We conclude that

$$\begin{aligned}\nabla^2 P^{\mu}_{s,t}(f)(x)_{i,j} &= \nabla^2 P^{\mu}_{s,t}(f)(x)_{j,i} \\ &= \mathbb{E}\big(f\big(X^{\mu}_{s,t}(x)\big)\nabla X^{\mu}_{s,u}(x)_{j,l}\tau^{\phi_{s,u}(\mu),\omega}_{u,t}\big(X^{\mu}_{s,u}(x)\big)_l \tau^{\mu,\omega}_{s,u}(x)_i\big) \\ &\quad + \mathbb{E}\bigg(P^{\phi_{s,u}(\mu)}_{u,t}(f)\big(X^{\mu}_{s,u}(x)\big)\int_s^u \partial_v \omega_{s,u}(v)\nabla^2 X^{\mu}_{s,v}(x)_{(i,j),k}\, dW^k_v\bigg).\end{aligned}$$

This ends the proof of (3.15). □

PROOF OF (3.25). We have

$$\begin{aligned}p^{\mu_1,\mu_0}_{s,t}(x,z) = b^{\mu_0}_{s,t}(x,z) + \sum_{n\geq 1}\int_{[s,t]_n} & B_2 e^{(t-u_n)B_1} B_2 e^{(u_n-u_{n-1})B_1}\cdots B_2 e^{(u_2-u_1)B_1} \\ & \times \big(B_1\phi_{s,u_1}(\mu_1)(e) + B_2\big(e^{(u_1-s)B_1}\big(x-\mu_0(e)\big) \\ & + e^{(u_1-s)(B_1+B_2)}\mu_0(e)\big)\big)\, du_1\cdots du_n.\end{aligned}$$

Recalling that

$$\begin{aligned}\phi_{s,u_1}(\mu_1)(e) &= e^{(u_1-s)[B_1+B_2]}\mu_1(e), \\ b^{\mu_0}_{s,t}(x,z) &= B_1 z + B_2\big[e^{(t-s)B_1}\big(x-\mu_0(e)\big) + e^{(t-s)[B_1+B_2]}\mu_0(e)\big]\end{aligned}$$

and using the rather well known exponential formulae

$$\begin{aligned}e^{(t-s)(B_1+B_2)} &= e^{(t-s)B_1} + \int_s^t e^{(t-u)B_1} B_2 e^{(u-s)(B_1+B_2)}\, du \\ &= e^{(t-s)B_1} + \int_s^t e^{(t-u)(B_1+B_2)} B_2 e^{(u-s)B_1}\, du\end{aligned}$$

we check that

$$\begin{aligned}p^{\mu_1,\mu_0}_{s,t}(x,z) &= b^{\mu_0}_{s,t}(x,z) \\ &\quad + B_2\int_s^t e^{(t-u_1)(B_1+B_2)}\big(B_1\phi_{s,u_1}(\mu_1)(e) \\ &\quad + B_2\big(e^{(u_1-s)B_1}\big(x-\mu_0(e)\big) + e^{(u_1-s)(B_1+B_2)}\mu_0(e)\big)\big)\, du_1\end{aligned}$$

from which we find that

$$\begin{aligned}p^{\mu_1,\mu_0}_{s,t}(x,z) &= B_1 z + B_2\big[e^{(t-s)B_1}\big(x-\mu_0(e)\big) + e^{(t-s)[B_1+B_2]}\mu_0(e)\big] \\ &\quad + B_2\bigg[\int_s^t e^{(t-u_1)(B_1+B_2)} B_1 e^{(u_1-s)[B_1+B_2]}\, du_1\bigg]\mu_1(e) \\ &\quad + B_2\bigg[\int_s^t e^{(t-u_1)(B_1+B_2)} B_2 e^{(u_1-s)(B_1+B_2)}\, du_1\bigg]\mu_0(e) \\ &\quad + B_2\big[e^{(t-s)(B_1+B_2)} - e^{(t-s)B_1}\big]\big(x-\mu_0(e)\big).\end{aligned}$$

This ends the proof of (3.25). □

PROOF OF (4.6). We have the tensor product formula

$$(D_{\mu_1,\mu_0}\phi_{s,t})^{\otimes 2} := \big(P^{\mu_0}_{s,t}\big)^{\otimes 2} + \big(Q^{\mu_1,\mu_0}_{s,t}\big)^{\otimes 2} + Q^{\mu_1,\mu_0}_{s,t}\otimes P^{\mu_0}_{s,t} + P^{\mu_0}_{s,t}\otimes Q^{\mu_1,\mu_0}_{s,t}.$$

We also have

$$\begin{aligned}&\left(Q_{s,t}^{\mu_1,\mu_0}\otimes P_{s,t}^{\mu_0}\right)(g)(x,\overline{x})\\&\quad=\int_{\Delta_{s,t}}\Phi_{s,u}(\mu_1)\big(d(u,y)\big)\big(b_{s,u}^{\mu_0}(x,y)'\otimes I\big)\big(\mathcal{P}_{u,t}^{\phi_{s,u}(\mu_0)}\otimes P_{s,t}^{\mu_0}\big)(\nabla_{x_1}g)(y,\overline{x}).\end{aligned}$$

Recall that $b_t(x,y)$ is differentiable at any order with uniformly bounded derivatives. Thus all differentials of the above function w.r.t. the coordinate x have uniformly bounded derivatives. On the other hand, the mapping $x\mapsto b_t(x,y)$ has at most linear growth. Thus, using the estimates (1.15) and (3.4), for any $m\geq 0$ we check that

$$\left\|Q_{s,t}^{\mu_1,\mu_0}\otimes P_{s,t}^{\mu_0}\right\|_{\mathcal{C}_m^{n+1}(\mathbb{R}^{2d})\to\mathcal{C}_{m+1}^{n}(\mathbb{R}^{2d})}\leq c_{m,n}(t)\rho_{m\vee 2}(\mu_0,\mu_1).$$

In the same vein, we have the tensor product formula

$$\begin{aligned}\left(Q_{s,t}^{\mu_1,\mu_0}\right)^{\otimes 2}(g)(x,\overline{x})&=\left(\mathcal{Q}_{s,t}^{\mu_1,\mu_0}\right)^{\otimes 2}\big((\nabla\otimes\nabla)g\big)(x,\overline{x})\\&:=\int_{\Delta_{s,t}\times\Delta_{s,t}}\left[\Phi_{s,u}(\mu_1)\otimes\Phi_{s,u}(\mu_1)\right]\big(d\big((u,y),(\overline{u},\overline{y})\big)\big)\\&\qquad\times\widehat{b}_{s,u,\overline{u}}^{\mu_0}\big((x,\overline{x}),(y,\overline{y})\big)'\widehat{\mathcal{P}}_{u,\overline{u},t}^{\phi_{s,u,\overline{u}}(\mu_0)}\big((\nabla\otimes\nabla)g\big)(y,\overline{y})\end{aligned}$$

with

$$\widehat{b}_{s,u,\overline{u}}^{\mu_0}\big((x,\overline{x}),(y,\overline{y})\big)':=b_{s,u}^{\mu_0}(x,y)'\otimes b_{s,\overline{u}}^{\mu_0}(\overline{x},\overline{y})'\quad\text{and}\quad\widehat{\mathcal{P}}_{u,\overline{u},t}^{\phi_{s,u,\overline{u}}(\mu_0)}:=\mathcal{P}_{u,t}^{\phi_{s,u}(\mu_0)}\otimes\mathcal{P}_{\overline{u},t}^{\phi_{s,\overline{u}}(\mu_0)}.$$

Arguing as above and using the estimates (1.15) and (3.4) for any $m\geq 0$ we check that

$$\left\|\left(Q_{s,t}^{\mu_1,\mu_0}\right)^{\otimes 2}\right\|_{\mathcal{C}_m^{2}(\mathbb{R}^{2d})\to\mathcal{C}_{2}^{n}(\mathbb{R}^{2d})}\leq c_{m,n}(t)\rho_{m\vee 2}(\mu_0,\mu_1).\qquad\square$$

PROOF OF LEMMA 5.5. Using the decomposition

$$\begin{aligned}&\phi_{s,u}(\mu_1)-\phi_{s,u}(\mu_0)\\&\quad=\sum_{1\leq l\leq n}\left[\phi_{s,u_1}(\mu_1)\otimes\cdots\otimes\phi_{s,u_{l-1}}(\mu_1)\right]\otimes\left[\phi_{s,u_l}(\mu_1)-\phi_{s,u_l}(\mu_0)\right]\\&\qquad\otimes\left[\phi_{s,u_{l+1}}(\mu_0)\otimes\cdots\otimes\phi_{s,u_n}(\mu_0)\right]\end{aligned}$$

which is valid for any $\mu_0,\mu_1\in P_2(\mathbb{R}^d)$ and any $u=(u_1,\dots,u_n)\in[s,t]_n$ with $n\geq 1$, for any function

$$(u,y)\in\Delta_{s,t}\mapsto h_u(y)\in\mathbb{R}$$

we check that

$$\begin{aligned}&\int_{\Delta_{s,t}}\left[\Phi_{s,u}(\mu_1)-\Phi_{s,u}(\mu_0)\right]\big(d(u,y)\big)h_u(y)\\&\quad=\int_{\Delta_{s,t}^1}\left[\Phi_{s,v}(\mu_1)-\Phi_{s,v}(\mu_0)\right]\big(d(v,z)\big)\overline{h}_v(z)\end{aligned}\tag{A.3}$$

with the function

$$\begin{aligned}\overline{h}_v(z):=h_v(z)&+\int_{\Delta_{s,v}}\Phi_{s,u}(\mu_1)\big(d(u,y)\big)h_{u,v}(y,z)\\&+\int_{\Delta_{v,t}}\Phi_{v,u}\big(\phi_{s,v}(\mu_0)\big)\big(d(u,y)\big)h_{v,u}(z,y)\\&+\int_{\Delta_{s,v}\times\Delta_{v,t}}\Upsilon_{s,t}^{\mu_1,\mu_0}\big((v,z),d\big((u,y),(\overline{u},\overline{y})\big)\big)h_{(u,v,\overline{u})}(y,z,\overline{y}).\end{aligned}$$

In the above display, $\Upsilon_{s,t}^{\mu_1,\mu_0}$ stands for the tensor product measures

$$\Upsilon_{s,t}^{\mu_1,\mu_0}\big((v,z),d\big((u,y),(\overline{u},\overline{y})\big)\big)=\Phi_{s,u}(\mu_1)\big(d(u,y)\big)\Phi_{v,\overline{u}}\big(\phi_{s,v}(\mu_0)\big)\big(d(\overline{u},\overline{y})\big).$$

We also have the tensor product formula

$$\begin{aligned}&(D_{\mu_1,\mu_0}\phi_{s,t})^{\otimes 2}-(D_{\mu_0}\phi_{s,t})^{\otimes 2}\\&\quad=(Q_{s,t}^{\mu_1,\mu_0})^{\otimes 2}-(Q_{s,t}^{\mu_0})^{\otimes 2}+(Q_{s,t}^{\mu_1,\mu_0}-Q_{s,t}^{\mu_0})\otimes P_{s,t}^{\mu_0}+P_{s,t}^{\mu_0}\otimes(Q_{s,t}^{\mu_1,\mu_0}-Q_{s,t}^{\mu_0}).\end{aligned}$$

This yields the decomposition

$$\begin{aligned}&\big([Q_{s,t}^{\mu_1,\mu_0}-Q_{s,t}^{\mu_0}]\otimes P_{s,t}^{\mu_0}\big)(g)(x,\overline{x})\\&\quad:=\int_s^t\int[\phi_{s,v}(\mu_1)-\phi_{s,v}(\mu_0)](d\widehat{x})\mathcal{I}_{s,v,t}^{\mu_0,\mu_1}(g)(x,\overline{x},\widehat{x})\,dv\end{aligned}$$

with the integral operator

$$\begin{aligned}&\mathcal{I}_{s,v,t}^{\mu_0,\mu_1}(g)(x,\overline{x},\widehat{x})\\&\quad:=b_{s,v}^{\mu_0}(x,\widehat{x})'\big(\mathcal{P}_{v,t}^{\phi_{s,v}(\mu_0)}\otimes P_{s,t}^{\mu_0}\big)(\nabla_{x_1}g)(\widehat{x},\overline{x})\\&\qquad+\int_{\Delta_{s,v}}\Phi_{s,u}(\mu_1)\big(d(u,y)\big)b_{s,u,v}^{\mu_0}(x,y,\widehat{x})'\big(\mathcal{P}_{v,t}^{\phi_{s,v}(\mu_0)}\otimes P_{s,t}^{\mu_0}\big)(\nabla_{x_1}g)(\widehat{x},\overline{x})\\&\qquad+\int_{\Delta_{v,t}}\Phi_{v,u}\big(\phi_{s,v}(\mu_0)\big)\big(d(u,y)\big)b_{s,v,u}^{\mu_0}(x,\widehat{x},y)'\big(\mathcal{P}_{u,t}^{\phi_{s,u}(\mu_0)}\otimes P_{s,t}^{\mu_0}\big)(\nabla_{x_1}g)(y,\overline{x})\\&\qquad+\int_{\Delta_{s,v}\times\Delta_{v,t}}\Upsilon_{s,t}^{\mu_1,\mu_0}\big((v,z),d\big((u,y),(\overline{u},\overline{y})\big)\big)\\&\qquad\times b_{s,u,v,\overline{u}}^{\mu_0}(x,y,\widehat{x},\overline{y})'\big(\mathcal{P}_{\overline{u},t}^{\phi_{s,\overline{u}}(\mu_0)}\otimes P_{s,t}^{\mu_0}\big)(\nabla_{x_1}g)(\overline{y},\overline{x}).\end{aligned}$$

Arguing as in the proof of (3.12) and (4.5) we check that

$$\big\|\mathcal{I}_{s,v,t}^{\mu_0,\mu_1}\big\|_{\mathcal{C}_m^{n+1}(\mathbb{R}^{2d})\to\mathcal{C}_{m+2}^{n}(\mathbb{R}^{3d})}\le c_{m,n}(t)\rho_{m\vee 2}(\mu_0,\mu_1).$$

In the same vein, we have

$$\begin{aligned}&\big[(Q_{s,t}^{\mu_1,\mu_0})^{\otimes 2}-(Q_{s,t}^{\mu_0})^{\otimes 2}\big](g)(x,\overline{x})\\&\quad=\int_{\Delta_{s,t}}[\Phi_{s,u}(\mu_1)-\Phi_{s,u}(\mu_0)]\big(d(u,y)\big)\big[\Theta_{s,u,t}^{\mu_1,\mu_0}+\overline{\Theta}_{s,u,t}^{\mu_1,\mu_0}\big](g)(x,\overline{x},y)\,dv\end{aligned}$$

with

$$\begin{aligned}&\overline{\Theta}_{s,u,t}^{\mu_1,\mu_0}(g)(x,\overline{x},y)\\&\quad:=\int_{\Delta_{s,t}}\Phi_{s,\overline{u}}(\mu_1)\big(d(\overline{u},\overline{y})\big)\widehat{b}_{s,u,\overline{u}}^{\mu_0}\big((x,\overline{x}),(y,\overline{y})\big)'\widehat{\mathcal{P}}_{u,\overline{u},t}^{\phi_{s,u,\overline{u}}(\mu_0)}\big((\nabla\otimes\nabla)g\big)(y,\overline{y})\end{aligned}$$

and

$$\begin{aligned}&\Theta_{s,u,t}^{\mu_1,\mu_0}(g)(x,\overline{x},y)\\&\quad:=\int_{\Delta_{s,t}}\Phi_{s,\overline{u}}(\mu_0)\big(d(\overline{u},\overline{y})\big)\widehat{b}_{s,\overline{u},u}^{\mu_0}\big((x,\overline{x}),(\overline{y},y)\big)'\widehat{\mathcal{P}}_{\overline{u},u,t}^{\phi_{s,\overline{u},u}(\mu_0)}\big((\nabla\otimes\nabla)g\big)(\overline{y},y).\end{aligned}$$

This yields the formula

$$\begin{aligned}&\big[(Q_{s,t}^{\mu_1,\mu_0})^{\otimes 2}-(Q_{s,t}^{\mu_0})^{\otimes 2}\big](g)(x,\overline{x})\\&\quad=\int_s^t[\phi_{s,v}(\mu_1)-\phi_{s,v}(\mu_0)](d\widehat{x})\mathcal{J}_{s,v,t}^{\mu_0,\mu_1}(g)(x,\overline{x},\widehat{x})\,dv\end{aligned}$$

with the integral operator

$$
\begin{aligned}
&\mathcal{J}^{\mu_0,\mu_1}_{s,v,t}(g)(x,\overline{x},\widehat{x})\\
&\quad:=\big[\Theta^{\mu_1,\mu_0}_{s,v,t}+\overline{\Theta}^{\mu_1,\mu_0}_{s,v,t}\big](g)(x,\overline{x},\widehat{x})\\
&\qquad+\int_{\Delta_{s,v}}\Phi_{s,u}(\mu_1)\big(d(u,y)\big)\big[\Theta^{\mu_1,\mu_0}_{s,(u,v),t}+\overline{\Theta}^{\mu_1,\mu_0}_{s,(u,v),t}\big](g)\big(x,\overline{x},(y,\widehat{x})\big)\\
&\qquad+\int_{\Delta_{v,t}}\Phi_{v,u}\big(\phi_{s,v}(\mu_0)\big)\big(d(u,y)\big)\big[\Theta^{\mu_1,\mu_0}_{s,(v,u),t}+\overline{\Theta}^{\mu_1,\mu_0}_{s,(v,u),t}\big](g)\big(x,\overline{x},(\widehat{x},y)\big)\\
&\qquad+\int_{\Delta_{s,v}\times\Delta_{v,t}}\Upsilon^{\mu_1,\mu_0}_{s,t}\big((v,z),d\big((u,y),(\overline{u},\overline{y})\big)\big)\\
&\qquad\times\big[\Theta^{\mu_1,\mu_0}_{s,(u,v,\overline{u}),t}+\overline{\Theta}^{\mu_1,\mu_0}_{s,(u,v,\overline{u}),t}\big](g)\big(x,\overline{x},(y,\widehat{x},\overline{y})\big).
\end{aligned}
$$

Arguing as above, we check that

$$
\|\mathcal{J}^{\mu_0,\mu_1}_{s,v,t}\|_{\mathcal{C}^2_m(\mathbb{R}^{2d})\to\mathcal{C}^n_2(\mathbb{R}^{3d})}\le c_{m,n}(t)\rho_{m\vee 2}(\mu_0,\mu_1).
$$

Combining the above decompositions we find that

$$
\begin{aligned}
&\big[(D_{\mu_1,\mu_0}\phi_{s,t})^{\otimes 2}-(D_{\mu_0}\phi_{s,t})^{\otimes 2}\big](g)(x,\overline{x})\\
&\quad=\int_s^t\big[\phi_{s,v}(\mu_1)-\phi_{s,v}(\mu_0)\big](d\widehat{x})\mathcal{K}^{\mu_0,\mu_1}_{s,v,t}(g)(x,\overline{x},\widehat{x})\,dv\\
&\quad\text{with } \mathcal{K}^{\mu_0,\mu_1}_{s,v,t}:=2\mathcal{I}^{\mu_0,\mu_1}_{s,v,t}+\mathcal{J}^{\mu_0,\mu_1}_{s,v,t}.
\end{aligned}
$$

For any $n\ge 2$ and $m\ge 0$ we have

$$
\|\mathcal{K}^{\mu_0,\mu_1}_{s,v,t}\|_{\mathcal{C}^{n+1}_m(\mathbb{R}^{2d})\to\mathcal{C}^n_{m+2}(\mathbb{R}^{3d})}\le c_{m,n}(t)\rho_{m\vee 2}(\mu_0,\mu_1).
$$

We conclude that

$$
\big(\phi_{s,t}(\mu_1)-\phi_{s,t}(\mu_0)\big)^{\otimes 2}=(\mu_1-\mu_0)^{\otimes 2}(D_{\mu_0}\phi_{s,t})^{\otimes 2}+(\mu_1-\mu_0)^{\otimes 3}\mathcal{R}_{\mu_1,\mu_0}\phi_{s,t}
$$

with the operator

$$
\begin{aligned}
&\mathcal{R}_{\mu_1,\mu_0}\phi_{s,t}(g)(x,\overline{x},\widehat{x})\\
&\quad:=\int_s^t\bigg[\int P^{\mu_0}_{s,v}(\widehat{x},dz)\mathcal{K}^{\mu_0,\mu_1}_{s,v,t}(g)(x,\overline{x},z)\\
&\qquad+\int_{\Delta_{s,v}}\Phi_{s,u}(\mu_1)\big(d(u,y)\big)b^{\mu_0}_{s,u}(\widehat{x},y)'\mathcal{L}^{\mu_0,\mu_1}_{s,u,v,t}(g)(x,\overline{x},y)\bigg]\,dv.
\end{aligned}
$$

In the above display, $\mathcal{L}^{\mu_0,\mu_1}_{s,u,v,t}$ stands for the integral operator operator

$$
\mathcal{L}^{\mu_0,\mu_1}_{s,u,v,t}(g)(x,\overline{x},y)=\mathcal{P}^{\phi_{s,u}(\mu_0)}_{u,t}\big(\nabla_{x_3}\mathcal{K}^{\mu_0,\mu_1}_{s,v,t}(g)(x,\overline{x},\cdot)\big)(y).
$$

We also check that

$$
\|\mathcal{R}_{\mu_1,\mu_0}\phi_{s,t}\|_{\mathcal{C}^{n+2}_m(\mathbb{R}^{2d})\to\mathcal{C}^n_{m+3}(\mathbb{R}^{3d})}\le c_{m,n}(t)\rho_{m+2}(\mu_1,\mu_2).
$$

This ends the proof of the lemma. □

PROOF OF THE ESTIMATE (2.15). For any $x=(x_1,x_2)\in\mathbb{R}^{2d}$ we set $\sigma(x_1,x_2):=\sigma(x_2,x_1)$. In this notation, for any matrix valued function $h(x)=(h_{i,j}(x))_{1\le i,j\le d}$ we have the tensor product formula

$$
\begin{aligned}
&(\mathcal{D}_{\mu_1,\mu_0}\phi_{s,t})^{\otimes 2}(h)(x)\\
&\quad= (\mathcal{P}^{\mu_0}_{s,t})^{\otimes 2}(h)(x) + \int_{\Delta_{s,t}} \Phi_{s,v}(\mu_1)\big(d(u,y)\big)\big[\mathbb{I}^{\mu_0}_{s,u,t}(h)(x,y) + \mathbb{I}^{\mu_0}_{s,u,t}(h)\big(\sigma(x),y\big)\big]\\
&\qquad+ \int_{\Delta_{s,t}\times\Delta_{s,t}} \Phi_{s,u}(\mu_1)\big(d(u,y)\big)\Phi_{s,v}(\mu_1)\big(d(v,z)\big)\mathbb{J}^{\mu_0}_{s,u,v,t}(h)(x,y,z)
\end{aligned}
$$

with the matrix valued functions $\mathbb{I}^{\mu_0}_{s,u,t}(h)$ and $\mathbb{J}^{\mu_0}_{s,u,v,t}(h)$ given for any $(u,y)\in\Delta^n_{s,t}$ and $(v,z)\in\Delta^m_{s,t}$ by the formula

$$
\begin{aligned}
\mathbb{I}^{\mu_0}_{s,u,t}(h)(x,y) &:= \mathbb{B}^{\mu_0}_{s,u}(x_1,y)\big(\mathcal{P}^{\phi_{s,u_n}(\mu_0)}_{u_n,t}\otimes\mathcal{P}^{\mu_0}_{s,t}\big)(h)(y_n,x_2),\\
\mathbb{J}^{\mu_0}_{s,u,v,t}(h)(x,y,z) &:= \big[\mathbb{B}^{\mu_0}_{s,u}(x_1,y)\otimes\mathbb{B}^{\mu_0}_{s,v}(x_2,z)\big]\big(\mathcal{P}^{\phi_{s,u_n}(\mu_0)}_{u_n,t}\otimes\mathcal{P}^{\phi_{s,v_m}(\mu_0)}_{v_m,t}\big)(h)(y_n,z_m).
\end{aligned}
$$

Using (3.7) we have

$$
\nabla\mathcal{P}^{\mu}_{s,t}(g) = \mathcal{P}^{[2,1],\mu}_{s,t}(g) + \mathcal{P}^{[2,2],\mu}_{s,t}(\nabla g)
$$

from which we check the formula

$$
\begin{aligned}
&\nabla_{y_n}\big(\mathcal{P}^{\phi_{s,u_n}(\mu_0)}_{u_n,t}\otimes\mathcal{P}^{\phi_{s,v_m}(\mu_0)}_{v_m,t}\big)(h)(y_n,z_m)\\
&\quad= \big[\mathcal{P}^{[2,1],\phi_{s,u_n}(\mu_0)}_{u_n,t}\otimes\mathcal{P}^{\phi_{s,v_n}(\mu_0)}_{v_n,t}\big](h)(y_n,z_m) + \big[\mathcal{P}^{[2,2]\phi_{s,u}(\mu_0)}_{u_n,t}\otimes\mathcal{P}^{\phi_{s,v_m}(\mu_0)}_{v_m,t}\big](\nabla_{x_1}h)(y_n,z_m).
\end{aligned}
$$

By symmetry arguments, we also have

$$
\begin{aligned}
&\nabla_{z_m}\big(\mathcal{P}^{\phi_{s,u_n}(\mu_0)}_{u_n,t}\otimes\mathcal{P}^{\phi_{s,v_m}(\mu_0)}_{v_m,t}\big)(h)(y_n,z_m)\\
&\quad= \big[\mathcal{P}^{[2,1],\phi_{s,v_m}(\mu_0)}_{v_m,t}\otimes\mathcal{P}^{\phi_{s,u_n}(\mu_0)}_{u_n,t}\big](h)(z_m,y_n)\\
&\qquad+ \big[\mathcal{P}^{[2,2],\phi_{s,v_m}(\mu_0)}_{v_m,t}\otimes\mathcal{P}^{\phi_{s,u_n}(\mu_0)}_{u_n,t}\big](\nabla_{x_1}h)(z_m,y_n).
\end{aligned}
$$

Using (3.11) for any differentiable matrix valued function $h(x_1,x_2)$ such that $\|h\|\vee\|\nabla_{x_1}h\|\le 1$ we have the uniform estimate

$$
\big\|\nabla_{y_n}\big(\mathcal{P}^{\phi_{s,u_n}(\mu_0)}_{u_n,t}\otimes\mathcal{P}^{\phi_{s,v_m}(\mu_0)}_{v_m,t}\big)(h)(y_n,z_m)\big\| \le c_1 e^{-\lambda_1[(t-u_n)+(t-v_m)]}.
$$

In the same vein, we have

$$
\begin{aligned}
\mathbb{B}^{\mu_0}_{s,u}(x,y) &= \mathbb{B}^{[1],\mu_0}_{s,u}(x,y)\\
&= \mathbb{E}\big[\nabla X^{\mu_0}_{s,u_1}(x) b^{[2]}_{u_1}\big(y_1, X^{\mu_0}_{s,u_1}(x)\big)\big]\\
&\quad\times\prod_{1\le l<n}\mathbb{E}\big[\nabla X^{\phi_{s,u_l}(\mu_0)}_{u_l,u_{l+1}}(y_l) b^{[2]}_{u_{l+1}}\big(y_{l+1}, X^{\phi_{s,u_l}(\mu_0)}_{u_l,u_{l+1}}(y_l)\big)\big].
\end{aligned}
$$

Using the gradient and the Hessian estimates (3.2) and (3.3) for any $1\le k\le n$ we check that

$$
\big\|\nabla_{y_k}\mathbb{B}^{\mu_0}_{s,u}(x_1,y)\big\| \le c_2\big\|b^{[2]}\big\|^n_2 e^{-\lambda_1(u_n-s)}.
$$

Combining the above estimates with (3.29) we check that

$$
\begin{aligned}
&\big\|\nabla_{y_n}\mathbb{I}^{\mu_0}_{s,u,t}(h)(x,y)\big\|\\
&\quad\le c_3\big\|b^{[2]}\big\|^n_2\big[e^{-\lambda_1(u_n-s)}e^{-\lambda_1[(t-u_n)+(t-s)]} + e^{-\lambda_1[(u_n-s)]}e^{-\lambda_1[(t-u_n)+(t-s)]}\big]\\
&\quad\le c_4\big\|b^{[2]}\big\|^n_2 e^{-2\lambda_1(t-s)}.
\end{aligned}
$$

In addition, for any $1 \le k < n$ we have

$$\begin{aligned}
&\left\|\nabla_{y_k} \mathbb{I}^{\mu_0}_{s,u,t}(h)(x,y)\right\| \\
&\quad \le c_5 \left\|b^{[2]}\right\|_2^n e^{-\lambda_1(u_n-s)} e^{-\lambda_1[(t-u_n)+(t-s)]} \le c_5 \left\|b^{[2]}\right\|_2^n e^{-2\lambda_1(t-s)}.
\end{aligned}$$

We conclude that

$$\sup_{1\le k\le n} \left\|\nabla_{y_k} \mathbb{I}^{\mu_0}_{s,u,t}(h)(x,y)\right\| \le c \left\|b^{[2]}\right\|_2^n e^{-2\lambda_1(t-s)}. \tag{A.4}$$

Arguing as above, for any $1 \le k < n$ we have

$$\begin{aligned}
&\left\|\nabla_{y_k} \mathbb{J}^{\mu_0}_{s,u,v,t}(h)(x,y,z)\right\| \\
&\quad \le c_1 \left\|b^{[2]}\right\|_2^{m+n} e^{-\lambda_1(u_n-s)} e^{-\lambda_1(v_m-s)} e^{-\lambda_1[(t-u_n)+(t-v_m)]} \le c_2 \left\|b^{[2]}\right\|_2^{m+n} e^{-2\lambda_1(t-s)}.
\end{aligned}$$

In addition, for $k = n$ we have

$$\begin{aligned}
&\left\|\nabla_{y_n} \mathbb{J}^{\mu_0}_{s,u,v,t}(h)(x,y,z)\right\| \\
&\quad \le c_3 \|\nabla_{x_2} b\|_2^{m+n} \big[e^{-\lambda_1(u_n-s)} e^{-\lambda_1(v_m-s)} e^{-\lambda_1[(t-u_n)+(t-v_m)]} \\
&\qquad + e^{-\lambda_1(u_n-s)} e^{-\lambda_1(v_m-s)} e^{-\lambda_1[(t-u_n)+(t-v_m)]}\big].
\end{aligned}$$

This implies that

$$\sup_{1\le k\le n} \left\|\nabla_{y_k} \mathbb{J}^{\mu_0}_{s,u,v,t}(h)(x,y,z)\right\| \le c \left\|b^{[2]}\right\|_2^{m+n} e^{-2\lambda_1(t-s)}. \tag{A.5}$$

On the other hand, we have the decomposition

$$\begin{aligned}
&\big[(\mathcal{D}_{\mu_1,\mu_0}\phi_{s,t})^{\otimes 2} - (\mathcal{D}_{\mu_0}\phi_{s,t})^{\otimes 2}\big](h)(x) \\
&\quad = \int_{\Delta_{s,t}} \big[\Phi_{s,v}(\mu_1) - \Phi_{s,v}(\mu_0)\big]\big(d(u,y)\big) \mathbb{K}^{\mu_0,\mu_1}_{s,u,t}(h)(x,y)
\end{aligned}$$

with the matrix valued function

$$\begin{aligned}
\mathbb{K}^{\mu_0,\mu_1}_{s,u,t}(h)(x,y) := {}& \mathbb{I}^{\mu_0}_{s,u,t}(h)(x,y) + \mathbb{I}^{\mu_0}_{s,u,t}(h)\big(\sigma(x),y\big) \\
&+ \int_{\Delta_{s,t}} \Phi_{s,v}(\mu_1)\big(d(v,z)\big) \mathbb{J}^{\mu_0}_{s,u,v,t}(h)(x,y,z) \\
&+ \int_{\Delta_{s,t}} \Phi_{s,v}(\mu_0)\big(d(v,z)\big) \mathbb{J}^{\mu_0}_{s,v,u,t}(h)(x,z,y).
\end{aligned}$$

Using the estimates (A.4) and (A.5), for any $(u,y) \in \Delta^n_{s,t}$ we check that

$$\begin{aligned}
&\sup_{1\le k\le n} \left\|\nabla_{y_k} \mathbb{K}^{\mu_0,\mu_1}_{s,u,t}(h)(x,y)\right\| \\
&\quad \le c_1 \left\|b^{[2]}\right\|_2^n e^{-\lambda_1(t-s)} \big[e^{-\lambda_1(t-s)} + \big(e^{\|b^{[2]}\|_2(t-s)} - 1\big) e^{-\lambda_1(t-s)}\big] \\
&\quad \le c_2 \left\|b^{[2]}\right\|_2^n e^{-\lambda_1(t-s)} e^{-\lambda_{1,2}(t-s)}.
\end{aligned} \tag{A.6}$$

Using the decomposition (A.3) we also check that

$$\big[(\mathcal{D}_{\mu_1,\mu_0}\phi_{s,t})^{\otimes 2} - (\mathcal{D}_{\mu_0}\phi_{s,t})^{\otimes 2}\big](h)(x) = \int_s^t \big[\phi_{s,v}(\mu_1) - \phi_{s,v}(\mu_0)\big](dz) \overline{\mathbb{K}}^{\mu_0,\mu_1}_{s,v,t}(h)(x,z)\, dv$$

with the matrix valued function

$$
\begin{aligned}
&\overline{\mathbb{K}}^{\mu_0,\mu_1}_{s,v,t}(h)(x_1,x_2,x_3)\\
&\quad=\mathbb{K}^{\mu_0,\mu_1}_{s,v,t}(h)(x_1,x_2,x_3)+\int_{\Delta_{s,v}}\Phi_{s,u}(\mu_1)\big(d(u,y)\big)\mathbb{K}^{\mu_0,\mu_1}_{s,u,v,t}(h)\big(x_1,x_2,(y,x_3)\big)\\
&\qquad+\int_{\Delta_{v,t}}\Phi_{v,u}\big(\phi_{s,v}(\mu_0)\big)\big(d(u,y)\big)\mathbb{K}^{\mu_0,\mu_1}_{s,v,u,t}(h)(x_1,x_2,x_3,y)\\
&\qquad+\int_{\Delta_{s,v}\times\Delta_{v,t}}\Upsilon^{\mu_1,\mu_0}_{s,t}\big((v,z),d\big((u,y),(\overline{u},\overline{y})\big)\big)\mathbb{K}^{\mu_0,\mu_1}_{s,u,v,\overline{u},t}(h)\big(x_1,x_2,(y,x_3,\overline{y})\big).
\end{aligned}
$$

Using (A.6) we find the uniform estimates

$$
\begin{aligned}
&\big\|\nabla_{x_3}\overline{\mathbb{K}}^{\mu_0,\mu_1}_{s,v,t}(h)(x_1,x_2,x_3)\big\|\\
&\quad\leq c_1\big[e^{-2\lambda_{1,2}(t-s)}+\big(e^{\|b^{[2]}\|_2(t-s)}-1\big)e^{-\lambda_1(t-s)}e^{-\lambda_{1,2}(t-s)}\big]\leq c_2e^{-2\lambda_{1,2}(t-s)}.
\end{aligned}
\tag{A.7}
$$

On the other hand, using (4.4) and (2.5) we have

$$
\big[\phi_{s,t}(\mu_1)-\phi_{s,t}(\mu_0)\big](f)=(\mu_1-\mu_0)P^{\mu_0}_{s,t}(f)+(\mu_1-\mu_0)\mathcal{Q}^{\mu_1,\mu_0}_{s,t}(\nabla f).
$$

Thus, recalling that

$$
\mathcal{Q}^{\mu_1,\mu_0}_{s,t}(\nabla f)(z):=\int_{\Delta_{s,t}}\Phi_{s,u}(\mu_1)\big(d(u,y)\big)b^{\mu_0}_{s,u}(z,y)'\mathcal{P}^{\phi_{s,u}(\mu_0)}_{u,t}(\nabla f)(y)
$$

we check that

$$
\begin{aligned}
&\big[(\mathcal{D}_{\mu_1,\mu_0}\phi_{s,t})^{\otimes 2}-(\mathcal{D}_{\mu_0}\phi_{s,t})^{\otimes 2}\big](h)(x)\\
&\quad=\int(\mu_1-\mu_0)(dz)\int_s^t P^{\mu_0}_{s,v}\big(\overline{\mathbb{K}}^{\mu_0,\mu_1}_{s,v,t}(h)(x,\cdot)\big)(z)\,dv\\
&\qquad+\int(\mu_1-\mu_0)(dz)\int_s^t\int_{\Delta_{s,v}}\Phi_{s,u}(\mu_1)\big(d(u,y)\big)\\
&\qquad\times b^{\mu_0}_{s,u}(z,y)'\mathcal{P}^{\phi_{s,u}(\mu_0)}_{u,v}\big(\nabla_{x_3}\overline{\mathbb{K}}^{\mu_0,\mu_1}_{s,v,t}(h)(x,\cdot)\big)(y)\,dv
\end{aligned}
$$

This implies that

$$
\begin{aligned}
&(\nabla\otimes\nabla)D^2_{\mu_0,\mu_1}\phi_{s,t}(f)(x_1,x_2)-(\nabla\otimes\nabla)D^2_{\mu_0}\phi_{s,t}(f)(x_1,x_2)\\
&\quad=\int(\mu_1-\mu_0)(dx_3)\int_s^t\mathbb{L}^{\mu_1,\mu_0}_{s,u}\big(\mathbb{S}^{[2,1],\phi_{s,u}(\mu_0)}_{u,t}(\nabla f)\\
&\qquad+\mathbb{S}^{[2,2],\phi_{s,u}(\mu_0)}_{u,t}(\nabla^2 f)\big)(x_1,x_2,x_3)\,du
\end{aligned}
$$

with the tensor integral operator

$$
\begin{aligned}
&\mathbb{L}^{\mu_1,\mu_0}_{s,t}(h)(x_1,x_2,x_3)\\
&\quad:=\int_s^t P^{\mu_0}_{s,v}\big(\overline{\mathbb{K}}^{\mu_0,\mu_1}_{s,v,t}(h)(x_1,x_2,\cdot)\big)(x_3)\,dv\\
&\qquad+\int_s^t\int_{\Delta_{s,v}}\Phi_{s,u}(\mu_1)\big(d(u,y)\big)b^{\mu_0}_{s,u}(x_3,y)'\mathcal{P}^{\phi_{s,u}(\mu_0)}_{u,v}\big(\nabla_{x_3}\overline{\mathbb{K}}^{\mu_0,\mu_1}_{s,v,t}(h)(x_1,x_2,\cdot)\big)(y)\,dv.
\end{aligned}
$$

On the other hand, using (5.10)

$$
\begin{aligned}
&(\mu_1-\mu_0)^{\otimes 2}D^2_{\mu_1,\mu_0}\phi_{s,t}(f)-(\mu_1-\mu_0)^{\otimes 2}D^2_{\mu_0}\phi_{s,t}(f)\\
&\quad=\int_{[0,1]^3}\int_s^t\mathbb{E}\big(\big\langle\nabla_{x_3}\mathbb{L}^{\mu_1,\mu_0}_{s,u}\big(\mathbb{S}^{[2,1],\phi_{s,u}(\mu_0)}_{u,t}(\nabla f)\\
&\qquad+\mathbb{S}^{[2,2],\phi_{s,u}(\mu_0)}_{u,t}(\nabla^2 f)\big)(\mathcal{Y}_\epsilon),(\mathcal{Y}_1-\mathcal{Y}_0)^{\otimes 3}\big\rangle\big)\,du\,d\epsilon
\end{aligned}
$$

with the interpolating path

$$\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3) \mapsto \mathcal{Y}_\epsilon := \big(\overline{Y}_0^1 + \epsilon_1(\overline{Y}_1^1 - \overline{Y}_0^1), \overline{Y}_0^2 + \epsilon_2(\overline{Y}_1^2 - \overline{Y}_0^2), \overline{Y}_0^3 + \epsilon_3(\overline{Y}_1^3 - \overline{Y}_0^3)\big)$$

and

$$(\mathcal{Y}_1 - \mathcal{Y}_0)^{\otimes 3} := (\overline{Y}_1^1 - \overline{Y}_0^1) \otimes (\overline{Y}_1^2 - \overline{Y}_0^2) \otimes (\overline{Y}_1^3 - \overline{Y}_0^2).$$

In the above display, $(\overline{Y}_1^i, \overline{Y}_0^i)_{i=1,2,3}$ stands for independent copies of a pair of random variables (Y_0, Y_1) with distribution (μ_0, μ_1).

Using the commutation formula (3.5) we check that

$$\begin{aligned}
&\nabla_{x_3} \mathbb{L}_{s,t}^{\mu_1,\mu_0}(h)(x_1, x_2, x_3) \\
&\quad := \int_s^t \mathcal{P}_{s,v}^{\mu_0}\big(\nabla_{x_3} \overline{\mathbb{K}}_{s,v,t}^{\mu_0,\mu_1}(h)(x_1, x_2, \cdot)\big)(x_3)\, dv \\
&\qquad + \int_s^t \int_{\Delta_{s,v}} \Phi_{s,u}(\mu_1)\big(d(u,y)\big) \mathbb{B}_{s,u}^{\mu_0}(x_3, y) \mathcal{P}_{u,v}^{\phi_{s,u}(\mu_0)}\big(\nabla_{x_3} \overline{\mathbb{K}}_{s,v,t}^{\mu_0,\mu_1}(h)(x_1, x_2, \cdot)\big)(y)\, dv.
\end{aligned}$$

Using (A.7) for any differentiable matrix valued function $h(x_1, x_2)$ such that $\|h\| \vee \|\nabla_{x_1} h\| \leq 1$ and for any $\epsilon \in]0, 1[$ we check that

$$\begin{aligned}
&\|\nabla_{x_3} \mathbb{L}_{s,t}^{\mu_1,\mu_0}(h)(x_1, x_2, x_3)\| \\
&\quad \leq c_1 e^{-2\lambda_{1,2}(t-s)} \left[\int_s^t e^{-\lambda_1(v-s)}\, dv + \int_s^t \big(e^{\|b^{[2]}\|_2(v-s)} - 1\big) e^{-\lambda_1(v-s)}\, dv\right] \leq c_2 e^{-2\lambda_{1,2}(t-s)}.
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
&\nabla_{x_1}\big[\mathbb{S}_{s,t}^{[2,1],\mu}(\nabla f) + \nabla_{x_1} \mathbb{S}_{s,t}^{[2,2],\mu}(\nabla^2 f)\big](x_1, x_2) \\
&\quad = b_s^{[1,1,2]}(x_1, x_2) \nabla D_\mu \phi_{s,t}(f)(x_1) + b_s^{[2,2,1]}(x_2, x_1) \nabla D_\mu \phi_{s,t}(f)(x_2) \\
&\qquad + \nabla^3 D_\mu \phi_{s,t}(f)(x_1) b_s^{[2]}(x_1, x_2)' + b_s^{[2,2]}(x_2, x_1) \nabla^2 D_\mu \phi_{s,t}(f)(x_2) \\
&\qquad + \nabla^2 D_\mu \phi_{s,t}(f)(x_1) \star b_s^{[1,2]}(x_1, x_2)
\end{aligned}$$

with the $\star$-tensor product

$$\begin{aligned}
&\big[\nabla^2 D_\mu \phi_{s,t}(f)(x_1) \star b_s^{[1,2]}(x_1, x_2)\big]_{k,i,j} \\
&\quad = \sum_{1 \leq l \leq d} \big[\nabla^2 D_\mu \phi_{s,t}(f)(x_1)_{k,l} b_s^{[1,2]}(x_1, x_2)'_{l,i,j} + b_s^{[1,2]}(x_1, x_2)'_{k,j,l} \nabla^2 D_\mu \phi_{s,t}(f)(x_1)'_{l,j}\big].
\end{aligned}$$

Using (5.3) we check that

$$\big\|\nabla_{x_1}\big[\mathbb{S}_{s,t}^{[2,1],\mu}(\nabla f) + \nabla_{x_1} \mathbb{S}_{s,t}^{[2,2],\mu}(\nabla^2 f)\big]\big\| \leq c e^{-\lambda(t-s)} \sup_{k=1,2,3} \|\nabla^k f\| \quad \text{for some } \lambda > 0.$$

We conclude that for any function $f \in \mathcal{C}^3(\mathbb{R}^d)$ s.t. $\sup_{k=1,2,3} \|\nabla^k f\| \leq 1$.

$$\begin{aligned}
&\big|(\mu_1 - \mu_0)^{\otimes 2} D_{\mu_1,\mu_0}^2 \phi_{s,t}(f) - (\mu_1 - \mu_0)^{\otimes 2} D_{\mu_0}^2 \phi_{s,t}(f)\big| \\
&\quad \leq c e^{-\lambda(t-s)} \mathbb{W}_2(\mu_0, \mu_1)^3 \quad \text{for some } \lambda > 0.
\end{aligned}$$

The last assertion comes from the formula

$$\frac{1}{2}(\mu_1 - \mu_0)^{\otimes 2} D_{\mu_1,\mu_0}^2 \phi_{s,t} = \frac{1}{2}(\mu_1 - \mu_0)^{\otimes 2} D_{\mu_0}^2 \phi_{s,t} + (\mu_1 - \mu_0)^{\otimes 3} D_{\mu_0,\mu_1}^3 \phi_{s,t}. \qquad \square$$

PROOF OF THEOREM 2.6. We extend the operators $D^k_{\mu_1,\mu_0}\phi_{s,t}$ introduced in Theorem 2.4 to tensor functions $f=(f_i)_{i\in[n]}$ by considering the tensor function with entries

$$D^k_{\mu_1,\mu_0}\phi_{s,t}(f)_i = D^k_{\mu_1,\mu_0}\phi_{s,t}(f_i). \tag{A.8}$$

By Theorem 2.4 we have

$$\begin{aligned}
&[\phi_{s,u}(\mu_1)-\phi_{s,u}(\mu_0)]\big(b_u\big(X^{\mu_0}_{s,u}(x),\cdot\big)\big)\\
&\quad=\int(\mu_1-\mu_0)(dy)d^{[1],\mu_1,\mu_0}_{s,u}\big(X^{\mu_0}_{s,u}(x),y\big)\\
&\quad=\int(\mu_1-\mu_0)(dy)d^{[1],\mu_0}_{s,u}\big(X^{\mu_0}_{s,u}(x),y\big)\\
&\qquad+\frac{1}{2}\int(\mu_1-\mu_0)^{\otimes 2}(dz)d^{[2],\mu_1,\mu_0}_{s,u}\big(X^{\mu_0}_{s,u}(x),z\big)\\
&\quad=\int(\mu_1-\mu_0)(dy)d^{[1],\mu_0}_{s,u}\big(X^{\mu_0}_{s,u}(x),y\big)\\
&\qquad+\frac{1}{2}\int(\mu_1-\mu_0)^{\otimes 2}(dz)d^{[2],\mu_0}_{s,u}\big(X^{\mu_0}_{s,u}(x),z\big)\\
&\qquad+\int(\mu_1-\mu_0)^{\otimes 3}(dz)d^{[3],\mu_1,\mu_0}_{s,u}\big(X^{\mu_0}_{s,u}(x),z\big)
\end{aligned}\tag{A.9}$$

with the functions

$$\begin{aligned}
d^{[1],\mu_1,\mu_0}_{s,t}\big(X^{\mu_0}_{s,t}(x),y\big)&:=D_{\mu_1,\mu_0}\phi_{s,t}\big(b_t\big(X^{\mu_0}_{s,t}(x),\cdot\big)\big)(y),\\
d^{[2],\mu_1,\mu_0}_{s,t}\big(X^{\mu_0}_{s,t}(x),(z_1,z_2)\big)&:=D^2_{\mu_1,\mu_0}\phi_{s,t}\big(b_t\big(X^{\mu_0}_{s,t}(x),\cdot\big)\big)(z_1,z_2),\\
d^{[3],\mu_1,\mu_0}_{s,u}\big(X^{\mu_0}_{s,u}(x),(z_1,z_2,z_3)\big)&:=D^3_{\mu_1,\mu_0}\phi_{s,t}\big(b_t\big(X^{\mu_0}_{s,t}(x),\cdot\big)\big)(z_1,z_2,z_3).
\end{aligned}$$

We also write $d^{[k],\mu}_{s,t}$ instead of $d^{[k],\mu,\mu}_{s,t}$. Using (2.12) and (4.13) we check that

$$\big\|\nabla_y d^{[1],\mu_1,\mu_0}_{s,t}\big(X^{\mu_0}_{s,t}(x),y\big)\big\|\le c_1 e^{-\lambda(t-s)}$$

as well as

$$\big\|(\nabla_{z_1}\otimes\nabla_{z_2})d^{[2],\mu_1,\mu_0}_{s,t}\big(X^{\mu_0}_{s,t}(x),z_1,z_2\big)\big\|\le c_2 e^{-\lambda(t-s)}\quad\text{for some }\lambda>0. \tag{A.10}$$

Using (2.15) we also have

$$\begin{aligned}
&\left|\int(\mu_1-\mu_0)^{\otimes 3}(dz)d^{[3],\mu_1,\mu_0}_{s,t}\big(X^{\mu_0}_{s,t}(x),z\big)\right|\\
&\quad\le c_3 e^{-\lambda(t-s)}\mathbb{W}_2(\mu_0,\mu_1)^3\quad\text{for some }\lambda>0.
\end{aligned}\tag{A.11}$$

On the other hand, we have the second order expansions

$$\begin{aligned}
&[\nabla X^{\phi_{s,u}(\mu_0)}_{u,t}]\big(X^{\mu_1}_{s,u}(x)\big)'-[\nabla X^{\phi_{s,u}(\mu_0)}_{u,t}]\big(X^{\mu_0}_{s,u}(x)\big)'\\
&\quad=\int_0^1[\nabla^2 X^{\phi_{s,u}(\mu_0)}_{u,t}]\big(X^{\mu_0}_{s,u}(x)+\epsilon\big(X^{\mu_1}_{s,u}(y)-X^{\mu_0}_{s,u}(x)\big)\big)'[X^{\mu_1}_{s,u}(x)-X^{\mu_0}_{s,u}(x)]\,d\epsilon\\
&\quad=[\nabla^2 X^{\phi_{s,u}(\mu_0)}_{u,t}]\big(X^{\mu_0}_{s,u}(x)\big)'[X^{\mu_1}_{s,u}(x)-X^{\mu_0}_{s,u}(x)]\\
&\qquad+\int_0^1(1-\epsilon)[\nabla^3 X^{\phi_{s,u}(\mu_0)}_{u,t}]\big(X^{\mu_0}_{s,u}(x)\\
&\qquad+\epsilon\big(X^{\mu_1}_{s,u}(y)-X^{\mu_0}_{s,u}(x)\big)\big)'[X^{\mu_1}_{s,u}(x)-X^{\mu_0}_{s,u}(x)]^{\otimes 2}\,d\epsilon.
\end{aligned}$$

In the same vein, we have

$$
\begin{aligned}
& b_u(X^{\mu_1}_{s,u}(x), y) - b_u(X^{\mu_0}_{s,u}(x), y) \\
&\quad = \int_0^1 b^{[1]}_u(X^{\mu_0}_{s,u}(x) + \epsilon(X^{\mu_1}_{s,u}(x) - X^{\mu_0}_{s,u}(x)), y)'[X^{\mu_1}_{s,u}(x) - X^{\mu_0}_{s,u}(x)]\, d\epsilon \\
&\quad = b^{[1]}_u(X^{\mu_0}_{s,u}(x), y)'[X^{\mu_1}_{s,u}(x) - X^{\mu_0}_{s,u}(x)] \\
&\qquad + \int_0^1 (1-\epsilon) b^{[1,1]}_u(X^{\mu_0}_{s,u}(x) + \epsilon(X^{\mu_1}_{s,u}(x) - X^{\mu_0}_{s,u}(x)), y)'[X^{\mu_1}_{s,u}(x) - X^{\mu_0}_{s,u}(x)]^{\otimes 2}\, d\epsilon.
\end{aligned}
$$

This implies that

$$
\begin{aligned}
& X^{\mu_1}_{s,t}(x) - X^{\mu_0}_{s,t}(x) \\
&\quad = \int_s^t [\nabla X^{\phi_{s,u}(\mu_0)}_{u,t}](X^{\mu_0}_{s,u}(x))'[\phi_{s,u}(\mu_1) - \phi_{s,u}(\mu_0)](b_u(X^{\mu_0}_{s,u}(x), \cdot))\, du \\
&\qquad + \sum_{k=2,3} R^{[k],\mu_0,\mu_1}_{s,t}(x)
\end{aligned}
$$

with the second order remainder term

$$
\begin{aligned}
& R^{[2],\mu_0,\mu_1}_{s,t}(x) \\
&\quad := \int_s^t [\nabla^2 X^{\phi_{s,u}(\mu_0)}_{u,t}](X^{\mu_0}_{s,u}(x))'[X^{\mu_1}_{s,u}(x) - X^{\mu_0}_{s,u}(x)] \\
&\qquad \times [\phi_{s,u}(\mu_1) - \phi_{s,u}(\mu_0)](b_u(X^{\mu_0}_{s,u}(x), \cdot))\, du \\
&\qquad + \int_s^t [\nabla X^{\phi_{s,u}(\mu_0)}_{u,t}](X^{\mu_0}_{s,u}(x))'[\phi_{s,u}(\mu_1) - \phi_{s,u}(\mu_0)] \\
&\qquad \times (b^{[1]}_u(X^{\mu_0}_{s,u}(x), \cdot)')[X^{\mu_1}_{s,u}(x) - X^{\mu_0}_{s,u}(x)]\, du
\end{aligned}
$$

and the third order remainder term

$$
\begin{aligned}
& R^{[3],\mu_0,\mu_1}_{s,t}(x) \\
&\quad := \int_0^1 \int_s^t [\nabla^2 X^{\phi_{s,u}(\mu_0)}_{u,t}](X^{\mu_0}_{s,u}(x))'[X^{\mu_1}_{s,u}(x) - X^{\mu_0}_{s,u}(x)] \\
&\qquad \times [\phi_{s,u}(\mu_1) - \phi_{s,u}(\mu_0)](b^{[1]}_u(X^{\mu_0}_{s,u}(x) + \epsilon(X^{\mu_1}_{s,u}(x) - X^{\mu_1}_{s,u}(x)), \cdot)') \\
&\qquad \times [X^{\mu_1}_{s,u}(x) - X^{\mu_0}_{s,u}(x)]\, d\epsilon\, du \\
&\qquad + \int_0^1 (1-\epsilon) \int_s^t [\nabla X^{\phi_{s,u}(\mu_0)}_{u,t}](X^{\mu_0}_{s,u}(x))' \\
&\qquad \times [\phi_{s,u}(\mu_1) - \phi_{s,u}(\mu_0)](b^{[1,1]}_u(X^{\mu_0}_{s,u}(x) + \epsilon(X^{\mu_1}_{s,u}(x) - X^{\mu_1}_{s,u}(x)), \cdot)') \\
&\qquad \times [X^{\mu_1}_{s,u}(x) - X^{\mu_0}_{s,u}(x)]^{\otimes 2}\, du\, d\epsilon \\
&\qquad + \int_0^1 (1-\epsilon) \int_s^t [\nabla^3 X^{\phi_{s,u}(\mu_0)}_{u,t}](X^{\mu_0}_{s,u}(x) + \epsilon(X^{\mu_1}_{s,u}(y) - X^{\mu_1}_{s,u}(x)))' \\
&\qquad \times [X^{\mu_1}_{s,u}(x) - X^{\mu_0}_{s,u}(x)]^{\otimes 2}[\phi_{s,u}(\mu_1) - \phi_{s,u}(\mu_0)](b_u(X^{\mu_1}_{s,u}(x), \cdot))\, du\, d\epsilon.
\end{aligned}
$$

Combining (3.4) with (2.4) and (2.17) for any $k = 1, 2$ we check the uniform estimate

$$
\| R^{[k],\mu_0,\mu_1}_{s,t}(x) \| \le c e^{-\lambda(t-s)} \mathbb{W}_2(\mu_0, \mu_1)^k \quad \text{for some } \lambda > 0. \tag{A.12}
$$

We check (2.19) using (A.10) and (A.9).

Using (5.3) we also have the estimate

$$\|\nabla_y D_{\mu_0} X^{\mu_0}_{s,t}(x,y)\| \le c_3 e^{-\lambda(t-s)} \quad \text{for some } \lambda > 0.$$

Observe that

$$\begin{aligned}
&[\phi_{s,u}(\mu_1)-\phi_{s,u}(\mu_0)]\big(b^{[1]}_u\big(X^{\mu_0}_{s,u}(x),\cdot\big)'\big)\\
&\quad = \int (\mu_1-\mu_0)(dy)\, d^{[1,1],\mu_1,\mu_0}_{s,u}\big(X^{\mu_0}_{s,u}(x),y\big)\\
&\quad = \int (\mu_1-\mu_0)(dy)\, d^{[1,1],\mu_0}_{s,u}\big(X^{\mu_0}_{s,u}(x),y\big) + \frac{1}{2}\int (\mu_1-\mu_0)^{\otimes 2}(dz)\, d^{[2,1],\mu_1,\mu_0}_{s,u}\big(X^{\mu_0}_{s,u}(x),z\big)
\end{aligned}$$

with the matrix valued functions

$$\begin{aligned}
d^{[1,1],\mu_1,\mu_0}_{s,t}\big(X^{\mu_0}_{s,t}(x),y\big) &:= D_{\mu_1,\mu_0}\phi_{s,t}\big(b^{[1]}_t\big(X^{\mu_0}_{s,t}(x),\cdot\big)'\big)(y),\\
d^{[2,1],\mu_1,\mu_0}_{s,t}\big(X^{\mu_0}_{s,t}(x),z_1,z_2\big) &:= D^2_{\mu_1,\mu_0}\phi_{s,t}\big(b^{[1]}_t\big(X^{\mu_0}_{s,t}(x),\cdot\big)'\big)(z_1,z_2).
\end{aligned}$$

We also write $d^{[1,1],\mu}_{s,t}$ instead of $d^{[1,1],\mu,\mu}_{s,t}$. Observe that

$$\begin{aligned}
&R^{[2],\mu_0,\mu_1}_{s,t}(x)\\
&\quad = \frac{1}{2}\int (\mu_1-\mu_0)^{\otimes 2}(dz)\int_s^t \big[\nabla^2 X^{\phi_{s,u}(\mu_0)}_{u,t}\big]\big(X^{\mu_0}_{s,u}(x)\big)' D^{[2,1]}_{\mu_0} X^{\mu_0}_{s,u}(x,z)\,du\\
&\qquad + \frac{1}{2}\int (\mu_1-\mu_0)^{\otimes 2}(dy)\int_s^t \big[\nabla X^{\phi_{s,u}(\mu_0)}_{u,t}\big]\big(X^{\mu_0}_{s,u}(x)\big)' D^{[1,1]}_{\mu_0} X^{\mu_0}_{s,u}(x,z)\,du\\
&\qquad + R^{[3,2],\mu_0,\mu_1}_{s,t}(x)
\end{aligned}$$

with

$$\begin{aligned}
&R^{[3,2],\mu_0,\mu_1}_{s,t}(x)\\
&\quad = \frac{1}{2}\int (\mu_1-\mu_0)^{\otimes 3}(dy)\\
&\qquad \times \int_s^t \big[\nabla^2 X^{\phi_{s,u}(\mu_0)}_{u,t}\big]\big(X^{\mu_0}_{s,u}(x)\big)' D_{\mu_0} X^{\mu_0}_{s,u}(x,y_1)\, d^{[2],\mu_1,\mu_0}_{s,u}\big(X^{\mu_0}_{s,u}(x),(y_2,y_3)\big)\,du\\
&\qquad + \frac{1}{2}\int (\mu_1-\mu_0)^{\otimes 3}(dy)\\
&\qquad \times \int_s^t \big[\nabla X^{\phi_{s,u}(\mu_0)}_{u,t}\big]\big(X^{\mu_0}_{s,u}(x)\big)' d^{[2,1],\mu_1,\mu_0}_{s,u}\big(X^{\mu_0}_{s,u}(x),(y_2,y_3)\big) D_{\mu_0} X^{\mu_0}_{s,u}(x,y_1)\,du\\
&\qquad + \int_s^t \big[\nabla^2 X^{\phi_{s,u}(\mu_0)}_{u,t}\big]\big(X^{\mu_0}_{s,u}(x)\big)' \mathcal{R}^{[2],\mu_0,\mu_1}_{s,u}(x)\big[\phi_{s,u}(\mu_1)-\phi_{s,u}(\mu_0)\big]\big(b_u\big(X^{\mu_0}_{s,u}(x),\cdot\big)\big)\,du\\
&\qquad + \int_s^t \big[\nabla X^{\phi_{s,u}(\mu_0)}_{u,t}\big]\big(X^{\mu_0}_{s,u}(x)\big)' \big[\phi_{s,u}(\mu_1)-\phi_{s,u}(\mu_0)\big]\big(b^{[1]}_u\big(X^{\mu_0}_{s,u}(x),\cdot\big)'\big)\mathcal{R}^{[2],\mu_0,\mu_1}_{s,u}(x)\,du.
\end{aligned}$$

Observe that

$$\big\|R^{[3,2],\mu_0,\mu_1}_{s,t}(x)\big\| \le c e^{-\lambda(t-s)}\mathbb{W}_2(\mu_0,\mu_1)^3 \quad \text{for some } \lambda > 0. \tag{A.13}$$

This yields the second order decompositionn (2.20) with the remainder term

$$\begin{aligned}
\mathcal{R}^{\mu_1,\mu_0}_{s,t}(x) &:= R^{[3],\mu_0,\mu_1}_{s,t}(x) + R^{[3,2],\mu_0,\mu_1}_{s,t}(x)\\
&\qquad + \int (\mu_1-\mu_0)^{\otimes 3}(dz)\int_s^t \big[\nabla X^{\phi_{s,u}(\mu_0)}_{u,t}\big]\big(X^{\mu_0}_{s,u}(x)\big)' d^{[3],\mu_1,\mu_0}_{s,u}\big(X^{\mu_0}_{s,u}(x),z\big)\,du.
\end{aligned}$$

The end of the proof of is now a consequence of the estimates (A.11), (A.12) and (A.13). The proof of the theorem is completed. □

Acknowledgements. The authors are supported by the ANR Quamprocs on quantitative analysis of metastable processes. P. Del Moral is also supported in part from the Chair Stress Test, RISK Management and Financial Steering, led by the French École polytechnique and its Foundation and sponsored by BNP Paribas.

We also thank two anonymous reviewers for their excellent suggestions for improving the paper. Their detailed comments greatly improved the presentation of the article.

REFERENCES

[1] ALEKSEEV, V. M. (1961). An estimate for the perturbations of the solutions of ordinary differential equations. II. *Vestnik Moskov. Univ. Ser. I Mat. Meh.* **1961** 3–10. MR0133536

[2] AMBROSIO, L. and GIGLI, N. (2008). Construction of the parallel transport in the Wasserstein space. *Methods Appl. Anal.* **15** 1–29. MR2482206 https://doi.org/10.4310/MAA.2008.v15.n1.a3

[3] AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2005). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. *Lectures in Mathematics ETH Zürich*. Birkhäuser, Basel. MR2129498

[4] ARNAUDON, M. and DEL MORAL, P. (2018). A duality formula and a particle Gibbs sampler for continuous time Feynman–Kac measures on path spaces. arXiv:1805.05044.

[5] ARNAUDON, M. and DEL MORAL, P. (2019). A variational approach to nonlinear and interacting diffusions. *Stoch. Anal. Appl.* **37** 717–748. arXiv preprint arXiv:1812.04269. https://doi.org/10.1080/07362994.2019.1609985

[6] ARNAUDON, M., PLANK, H. and THALMAIER, A. (2003). A Bismut type formula for the Hessian of heat semigroups. *C. R. Math. Acad. Sci. Paris* **336** 661–666. MR1988128 https://doi.org/10.1016/S1631-073X(03)00123-7

[7] BENEDETTO, D., CAGLIOTI, E., CARRILLO, J. A. and PULVIRENTI, M. (1998). A non-Maxwellian steady distribution for one-dimensional granular media. *J. Stat. Phys.* **91** 979–990. MR1637274 https://doi.org/10.1023/A:1023032000560

[8] BENEDETTO, D., CAGLIOTI, E. and PULVIRENTI, M. (1997). A kinetic equation for granular media. *RAIRO Modél. Math. Anal. Numér.* **31** 615–641. MR1471181 https://doi.org/10.1051/m2an/1997310506151

[9] BENSOUSSAN, A., FREHSE, J. and YAM, S. C. P. (2015). The master equation in mean field theory. *J. Math. Pures Appl.* (9) **103** 1441–1474. MR3343705 https://doi.org/10.1016/j.matpur.2014.11.005

[10] BISHOP, A. N. and DEL MORAL, P. (2017). On the stability of Kalman–Bucy diffusion processes. *SIAM J. Control Optim.* **55** 4015–4047. arxiv e-print, updated, arXiv:1610.04686.

[11] BISHOP, A. N. and DEL MORAL, P. (2019). Stability properties of systems of linear stochastic differential equations with random coefficients. *SIAM J. Control Optim.* **57** 1023–1042. MR3924607 https://doi.org/10.1137/18M1182759

[12] BISMUT, J.-M. (1984). *Large Deviations and the Malliavin Calculus*. *Progress in Mathematics* **45**. Birkhäuser, Inc., Boston, MA. MR0755001

[13] BUCKDAHN, R., LI, J., PENG, S. and RAINER, C. (2017). Mean-field stochastic differential equations and associated PDEs. *Ann. Probab.* **45** 824–878. MR3630288 https://doi.org/10.1214/15-AOP1076

[14] CARDALIAGUET, P., DELARUE, F., LASRY, J.-M. and LIONS, P.-L. (2019). *The Master Equation and the Convergence Problem in Mean Field Games*. *Annals of Mathematics Studies* **201**. Princeton Univ. Press, Princeton, NJ. MR3967062 https://doi.org/10.2307/j.ctvckq7qf

[15] CARMONA, R. and DELARUE, F. (2013). Mean field forward-backward stochastic differential equations. *Electron. Commun. Probab.* **18** no. 68, 15. MR3091726 https://doi.org/10.1214/ECP.v18-2446

[16] CARMONA, R. and DELARUE, F. (2013). Probabilistic analysis of mean-field games. *SIAM J. Control Optim.* **51** 2705–2734. MR3072222 https://doi.org/10.1137/120883499

[17] CARMONA, R. and DELARUE, F. (2018). *Probabilistic Theory of Mean Field Games with Applications I–II*. *Probability Theory and Stochastic Modelling* **83–84**. Springer. MR3753660

[18] CATTIAUX, P., GUILLIN, A. and MALRIEU, F. (2008). Probabilistic approach for granular media equations in the non-uniformly convex case. *Probab. Theory Related Fields* **140** 19–40. MR2357669 https://doi.org/10.1007/s00440-007-0056-3

[19] CHASSAGNEUX, J. F., CRISAN, D. and DELARUE, F. (2014). A probabilistic approach to classical solutions of the master equation for large population equilibria. ArXiv preprint arXiv:1411.3009.

[20] CHAUDRU DE RAYNAL, P. E. and FRIKHA, N. (2018). Well-posedness for some non-linear diffusion processes and related PDE on the Wasserstein space. ArXiv preprint, arXiv:1811.06904.

[21] CONSTANTINE, G. M. and SAVITS, T. H. (1996). A multivariate Faà di Bruno formula with applications. *Trans. Amer. Math. Soc.* **348** 503–520. MR1325915 https://doi.org/10.1090/S0002-9947-96-01501-2

[22] COPPEL, W. A. (1965). *Stability and Asymptotic Behavior of Differential Equations*. D. C. Heath and Co., Boston, MA. MR0190463

[23] CRISAN, D. and MCMURRAY, E. (2018). Smoothing properties of McKean–Vlasov SDEs. *Probab. Theory Related Fields* **171** 97–148. MR3800831 https://doi.org/10.1007/s00440-017-0774-0

[24] DAWSON, D. A. (1983). Critical dynamics and fluctuations for a mean-field model of cooperative behavior. *J. Stat. Phys.* **31** 29–85. MR0711469 https://doi.org/10.1007/BF01010922

[25] DEL MORAL, P. (2004). *Feynman–Kac Formulae. Genealogical and Interacting Particle Systems with Applications. Probability and Its Applications* (*New York*). Springer, New York. MR2044973 https://doi.org/10.1007/978-1-4684-9393-1

[26] DEL MORAL, P. (2013). *Mean Field Simulation for Monte Carlo Integration. Monographs on Statistics and Applied Probability* **126**. CRC Press, Boca Raton, FL. MR3060209

[27] DEL MORAL, P. and GUIONNET, A. (1999). On the stability of measure valued processes with applications to filtering. *C. R. Acad. Sci. Paris Sér. I Math.* **329** 429–434. MR1710091 https://doi.org/10.1016/S0764-4442(00)88619-X

[28] DEL MORAL, P. and GUIONNET, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. Henri Poincaré Probab. Stat.* **37** 155–194. MR1819122 https://doi.org/10.1016/S0246-0203(00)01064-5

[29] DEL MORAL, P., HU, S. and WU, L. (2015). Moderate deviations for interacting processes. *Statist. Sinica* **25** 921–951. MR3409731

[30] DEL MORAL, P., JACOB, P. E., LEE, A., MURRAY, L. and PETERS, G. W. (2013). Feynman–Kac particle integration with geometric interacting jumps. *Stoch. Anal. Appl.* **31** 830–871. MR3175799 https://doi.org/10.1080/07362994.2013.817247

[31] DEL MORAL, P. and MICLO, L. (2000). Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to non-linear filtering. In *Séminaire de Probabilités, XXXIV. Lecture Notes in Math.* **1729** 1–145. Springer, Berlin. MR1768060 https://doi.org/10.1007/BFb0103798

[32] DEL MORAL, P. and MICLO, L. (2003). Particle approximations of Lyapunov exponents connected to Schrödinger operators and Feynman–Kac semigroups. *ESAIM Probab. Stat.* **7** 171–208. MR1956078 https://doi.org/10.1051/ps:2003001

[33] DEL MORAL, P., PATRAS, F. and RUBENTHALER, S. (2009). Coalescent tree based functional representations for some Feynman–Kac particle models. *Ann. Appl. Probab.* **19** 1–50.

[34] DEL MORAL, P., PATRAS, F. and RUBENTHALER, S. (2011). Convergence of U-statistics for interacting particle systems. *J. Theoret. Probab.* **24** 1002–1027. MR2851242 https://doi.org/10.1007/s10959-011-0355-6

[35] DEL MORAL, P. and RIO, E. (2009). Concentration inequalities for mean field particle models. *Ann. Appl. Probab.* **21** 1017–1052. HAL INRIA RR-6901, 2011.

[36] DEL MORAL, P. and SINGH, S. S. (2019). A forward-backward stochastic analysis of diffusion flows. arXiv preprint, arXiv:1906.09145.

[37] DEL MORAL, P. and TUGAUT, J. (2018). On the stability and the uniform propagation of chaos properties of ensemble Kalman–Bucy filters. *Ann. Appl. Probab.* **28** 790–850. MR3784489 https://doi.org/10.1214/17-AAP1317

[38] DUDLEY, R. M. (1968). The speed of mean Glivenko–Cantelli convergence. *Ann. Math. Stat.* **40** 40–50. MR0236977 https://doi.org/10.1214/aoms/1177697802

[39] ELWORTHY, K. D. and LI, X.-M. (1994). Formulae for the derivatives of heat semigroups. *J. Funct. Anal.* **125** 252–286. MR1297021 https://doi.org/10.1006/jfan.1994.1124

[40] FOURNIER, N. and GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** 707–738. MR3383341 https://doi.org/10.1007/s00440-014-0583-7

[41] GRAHAM, C. (1992). McKean–Vlasov Itô–Skorohod equations, and nonlinear diffusions with discrete jump sets. *Stochastic Process. Appl.* **40** 69–82. MR1145460 https://doi.org/10.1016/0304-4149(92)90138-G

[42] GRÖBNER, W. (1967). *Die Lie-Reihen und Ihre Anwendungen*, Zweite, überarbeitete und Erweiterte Auflage. *Mathematische Monographien* **3**. VEB Deutscher Verlag der Wissenschaften, Berlin. MR0217392

[43] GUÉANT, O., LASRY, J.-M. and LIONS, P.-L. (2011). Mean field games and applications. In *Paris-Princeton Lectures on Mathematical Finance* 2010. *Lecture Notes in Math.* **2003** 205–266. Springer, Berlin. MR2762362 https://doi.org/10.1007/978-3-642-14660-2_3

[44] HARRIS, T. E. and KAHN, H. (1951). Estimation of particle transmission by random sampling. *Natl. Bur. Stand., Appl. Math. Ser.* **12** 27–30.

[45] HUANG, M., MALHAMÉ, R. P. and CAINES, P. E. (2006). Large population stochastic dynamic games: Closed-loop McKean–Vlasov systems and the Nash certainty equivalence principle. *Commun. Inf. Syst.* **6** 221–251. MR2346927

[46] HUANG, X., RÖCKNER, M. and WANG, F.-Y. (2019). Nonlinear Fokker–Planck equations for probability measures on path space and path-distribution dependent SDEs. *Discrete Contin. Dyn. Syst.* **39** 3017–3035. MR3959419 https://doi.org/10.3934/dcds.2019125

[47] HUDDE, A., HUTZENTHALER, M., JENTZEN, A. and MAZZONETTO, S. (2018). On the Itô–Alekseev–Gröbner formula for stochastic differential equations. ArXiv preprint, arXiv:1812.09857.

[48] KAC, M. (1956). Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. III* 171–197. Univ. California Press, Berkeley and Los Angeles. MR0084985

[49] KAC, M. (1959). *Probability and Related Topics in Physical Sciences. Proceedings of the Summer Seminar, Boulder, Colo.* 1957. *Lectures in Applied Mathematics* Interscience Publishers, London. MR0102849

[50] KOLOKOLTSOV, V. N. (2010). *Nonlinear Markov Processes and Kinetic Equations. Cambridge Tracts in Mathematics* **182**. Cambridge Univ. Press, Cambridge. MR2680971 https://doi.org/10.1017/CBO9780511760303

[51] KOLOKOLTSOV, V. N., TROEVA, M. and YANG, W. (2014). On the rate of convergence for the mean-field approximation of controlled diffusions with large number of players. *Dyn. Games Appl.* **4** 208–230. MR3195847 https://doi.org/10.1007/s13235-013-0095-6

[52] KOLOKOLTSOV, V. N., TROEVA, M. S. and YANG, W. (2016). Mean field games based on stable-like processes. *Autom. Remote Control* **77** 2044–2064. MR3580196 https://doi.org/10.1134/S0005117916110138

[53] LANCASTER, P. and TISMENETSKY, M. (1985). *The Theory of Matrices*, 2nd ed. *Computer Science and Applied Mathematics*. Academic Press, Orlando, FL. MR0792300

[54] LEI, J. (2020). Convergence and concentration of empirical measures under Wasserstein distance in unbounded functional spaces. *Bernoulli* **26** 767–798. MR4036051 https://doi.org/10.3150/19-BEJ1151

[55] LI, X.-M. (2018). Doubly damped stochastic parallel translations and Hessian formulas. In *Stochastic Partial Differential Equations and Related Fields. Springer Proc. Math. Stat.* **229** 345–357. Springer, Cham. MR3828180 https://doi.org/10.1007/978-3-319-74929-7_2

[56] MCKEAN, H. P. JR. (1966). A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci. USA* **56** 1907–1911. MR0221595 https://doi.org/10.1073/pnas.56.6.1907

[57] MCKEAN, H. P. JR. (1967). Propagation of chaos for a class of non-linear parabolic equations. In *Stochastic Differential Equations* (*Lecture Series in Differential Equations, Session* 7, *Catholic Univ.*, 1967) 41–57. Air Force Office Sci. Res., Arlington, VA. MR0233437

[58] NISHIYAMA, Y. (1997). Some central limit theorems for l^{∞}-valued semimartingales and their applications. *Probab. Theory Related Fields* **108** 459–494. MR1465638 https://doi.org/10.1007/s004400050117

[59] OTTO, F. (2001). The geometry of dissipative evolution equations: The porous medium equation. *Comm. Partial Differential Equations* **26** 101–174. MR1842429 https://doi.org/10.1081/PDE-100002243

[60] OTTO, F. and VILLANI, C. (2000). Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** 361–400. MR1760620 https://doi.org/10.1006/jfan.1999.3557

[61] PAPANICOLAOU, G. C., STROOCK, D. and VARADHAN, S. R. S. (1977). Martingale approach to some limit theorems. In *Papers from the Duke Turbulence Conference* (*Duke Univ., Durham, N.C.*, 1976) *Duke Univ. Math. Ser.* **III** Paper No. 6. MR0461684

[62] PEANO, G. (1888). Intégration par séries des équations différentielles linéaires. *Math. Ann.* **32** 450–456. MR1510521 https://doi.org/10.1007/BF01443609

[63] SZNITMAN, A.-S. (1991). Topics in propagation of chaos. In *École D'Été de Probabilités de Saint-Flour XIX*—1989. *Lecture Notes in Math.* **1464** 165–251. Springer, Berlin. MR1108185 https://doi.org/10.1007/BFb0085169

[64] THOMPSON, J. (2019). Derivatives of Feynman–Kac semigroups. *J. Theoret. Probab.* **32** 950–973. MR3959634 https://doi.org/10.1007/s10959-018-0824-2

[65] TOSCANI, G. (2000). One-dimensional kinetic models of granular flows. *ESAIM Math. Model. Numer. Anal.* **34** 1277–1291. MR1812737 https://doi.org/10.1051/m2an:2000127

[66] VILLANI, C. (2002). A review of mathematical topics in collisional kinetic theory. In *Handbook of Mathematical Fluid Dynamics, Vol. I* 71–305. North-Holland, Amsterdam. MR1942465 https://doi.org/10.1016/S1874-5792(02)80004-0

[67] WANG, F. Y. (2019). Diffusions and PDEs on Wasserstein space. arXiv:1903.02148.