Backward Itô-Ventzell and stochastic interpolation formulae

Pierre del Moral, Sumeetpal Sidhu Singh

Backward Itô-Ventzell and stochastic interpolation formulae

P. Del Moral*[1] and S. S. Singh[2]

[1]INRIA, Bordeaux Research Center & CMAP, Polytechnique Palaiseau, France
[2]Department of Engineering, University of Cambridge, United Kingdom.

Abstract

We present a novel backward Itô-Ventzell formula and an extension of the Alekseev-Gröbner interpolating formula to stochastic flows. We also present some natural spectral conditions that yield direct and simple proofs of time uniform estimates of the difference between the two stochastic flows when their drift and diffusion functions are not the same, yielding what seems to be the first results of this type for this class of anticipative models. We illustrate the impact of these results in the context of diffusion perturbation theory, interacting diffusions and discrete time approximations.

Keywords : Stochastic flows, variational equations, tangent and Hessian processes, perturbation semigroups, backward Itô-Ventzell formula, Alekseev-Gröbner lemma, Skorohod stochastic integral, two-sided stochastic integration, Malliavin differential, Bismut-Elworthy-Li formulae.

Mathematics Subject Classification : 47D07, 93E15, 60H07.

1 Introduction

Let $b_t(x)$ be a vector-valued function from $\mathbb{R}^d$ into $\mathbb{R}^d$ and $\sigma_t(x) = [\sigma_{t,1}(x), \ldots, \sigma_{t,r}(x)]$ be a matrix-valued function from $\mathbb{R}^d$ into $\mathbb{R}^{d\times r}$, for some parameters $d, r \geqslant 1$. Both functions will be assumed to be differentiable. Let W_t be an r-dimensional Brownian motion and denote by $\mathcal{W}_{s,t}$ the σ-field generated by the increments $(W_u - W_v)$ of the Brownian motion, with $u, v \in [s,t]$.

For any time horizon $s \geqslant 0$ we denote by $X_{s,t}(x)$ the stochastic flow defined for any $t \in [s, \infty[$ and any starting point $X_{s,s}(x) = x \in \mathbb{R}^d$ by the stochastic differential equation

$$dX_{s,t}(x) = b_t\left(X_{s,t}(x)\right)\ dt + \sigma_t\left(X_{s,t}(x)\right)\ dW_t \tag{1.1}$$

We assume that $x \mapsto b_t(x)$ and $x \mapsto \sigma_t(x)$ have continuous and uniformly bounded derivatives up to the third order. This condition is clearly met for linear Gaussian models as well as for the geometric Brownian motion. This condition ensures that the stochastic flow $x \mapsto X_{s,t}(x)$ is a twice differentiable function of the initialisation x. In addition, all absolute moments of the flow and the ones of its first and second order derivatives exists for any time horizon. As it is well known, dynamical systems and hence stochastic models involving drift functions with quadratic growth

*P. Del Moral was supported in part from the Chair Stress Test, RISK Management and Financial Steering, led by the French Ecole polytechnique and its Foundation and sponsored by BNP Paribas, and by the ANR Quamprocs on quantitative analysis of metastable processes.
Authors declaration of interests: none

require additional regularity conditions to ensure non explosion of the solution in finite time. It is also implicitly assumed that all functions (b_t, σ_t) are smooth functions w.r.t. the time parameter. The present article develop several constructive and stochastic analysis tools including Bismut-Elworthy-Li formulae, stochastic semigroup perturbation formulae, extended two-sided stochastic integration, Malliavin calculus, gradient and Hessian semigroup processes estimates. We are also looking for useful quantitative and time uniform estimates which are valid under a single set of easily checked conditions that only depend on the parameters of the model. Various techniques presented in the article and many results can be separately and readily extended to more general models with weaker and abstract custom assumptions that depend on the different quantities to handle.

Let $\overline{X}_{s,t}(x)$ be the stochastic flow associated with a stochastic differential equation defined as (1.1) by replacing (b_t, σ_t) by some drift and diffusion functions $(\overline{b}_t, \overline{\sigma}_t)$ with the same regularity properties. Constant diffusion functions $(\sigma_t, \overline{\sigma}_t)$ are defined by

$$\sigma_t(x) = \Sigma_t \quad \text{and} \quad \overline{\sigma}_t(x) = \overline{\Sigma}_t \quad \text{for some matrices } \Sigma_t \text{ and } \overline{\Sigma}_t. \tag{1.2}$$

In this context, we will assume that Σ_t and $\overline{\Sigma}_t$ are uniformly bounded w.r.t. the time horizon.

The Markov transition semigroups associated with the flows $X_{s,t}(x)$ and $\overline{X}_{s,t}(x)$ are defined for any measurable function f on $\mathbb{R}^d$ by the formula

$$P_{s,t}(f)(x) := \mathbb{E}\left(f(X_{s,t}(x))\right) \quad \text{and} \quad \overline{P}_{s,t}(f)(x) := \mathbb{E}\left(f(\overline{X}_{s,t}(x))\right)$$

In this paper we derive equations for the differences $(X_{s,t} - \overline{X}_{s,t})$ and $(P_{s,t} - \overline{P}_{s,t})$ in terms of the difference of their corresponding drifts and diffusion functions,

$$\Delta a_t := a_t - \overline{a}_t \qquad \Delta b_t := b_t - \overline{b}_t \quad \text{and} \quad \Delta\sigma_t = \sigma_t - \overline{\sigma}_t \tag{1.3}$$

where $a_t(x) := \sigma_t(x)\,\sigma_t(x)'$ and $\overline{a}_t(x) := \overline{\sigma}_t(x)\,\overline{\sigma}'_t(x)$. In some applications the functions $\overline{b}_t = b_t - \Delta b_t$ and $\overline{\sigma}_t = \sigma_t - \Delta\sigma_t$ can be interpreted as a local perturbation of the drift and the diffusion of the stochastic flow $X_{s,t}$.

We also address the problem of finding time-uniform estimates for the difference between the stochastic flows $X_{s,t}$ and $\overline{X}_{s,t}$ and their corresponding Markov transition kernels $P_{s,t}$ and $\overline{P}_{s,t}$.

These important questions arise in a variety of domains including stochastic perturbation theory as well as in the stability and the qualitative theory of stochastic systems. Classical analytic estimates on the difference between the stochastic flows driven by different drift and diffusion functions are often much too large for most diffusion processes of practical interest. In some instances none of the diffusion flows are stable. In this context, any local perturbation of the stochastic model propagates so that any global error estimate eventually tends to ∞ as the time horizon $t \to \infty$.

Whenever one of the stochastic flows is stable, classical perturbation bounds combining Lipschitz type inequalities with Gronwall lemma [8, 25] yield exceedingly pessimistic global estimates that grows exponentially fast w.r.t. the time horizon. Notice that an exponential type estimate of the form $e^{\lambda t}$ for some parameter $\lambda > 0$ and some time horizon t s.t. $\lambda t \geqslant 199$ would induce an error bound larger than the estimated number 10^{86} of elementary particles of matters in the visible universe. As mentioned in [29] in the context of Euler scheme type approximations of deterministic dynamical systems, one may encounter situations where $\lambda = 10^8$ and $t = 10^2$ and the resulting exponential bounds are clearly impractical from a numerical perspective.

The statement of the main results of the article are presented in section 1.1:

i. Section 1.1.1 presents a novel generalized backward Itô-Ventzell formula (cf. theorem 1.1). The Itô-Ventzell is a very important formula, arguably as useful as the Itô's change of variable, but surprisingly the backward Itô-Ventzell presented in this work has never been studied before. Theorem 1.1 can be seen as a new generalized backward version of the generalized Itô-Ventzell formula presented in [41].

ii. In section 1.1.2 we apply the backward Itô-Ventzell formula to derive a forward-backward stochastic perturbation formula that expresses the difference between the stochastic flows $X_{s,t}$ and $\overline{X}_{s,t}$ in terms of first and second order derivatives of the flows, which we call the tangent and Hessian processes respectively, with respect to the space parameter (cf. theorem 1.2).

iii. Section 1.1.2 also provides a novel forward-backward Itô type differential formula for interpolating stochastic diffusion flows (cf. the change of variable formula (1.9)).

iv. In the beginning of section 1.1.2 we present a discrete time approach based on the pivotal interpolating telescoping sum formula (4.2). This interpolating stochastic semigroup technique can be seen as an extension to stochastic flows of the stochastic perturbation analysis developed in [22, 18, 20, 21] and in [3, 5, 11] in the context of discrete time models, matrix and nonlinear interacting processes (see also [4, 5]). For a more thorough discussion on these models, we refer to section 1.2. This approach allows to derive a stochastic interpolation formula (1.10) with a fluctuation term (1.12) defined by an extended two-sided stochastic integral.

v. Section 1.1.3 presents some natural spectral conditions on the gradients of $b_t(x), \sigma_t(x), \overline{b}_t(x)$ and $\overline{\sigma}_t(x)$ that allows us to derive in a direct way a series of realistic uniform estimates with respect to the time horizon.

The rest of the article is organized as follows:

Section 3 provides some basic tools associated with the first and second variational equations associated with a diffusion flow. We also present some quantitative estimates of the tangent and the Hessian processes. For a more thorough discussion on stochastic flows and their differentiability properties we refer to [14, 32, 40].

Section 4 is mainly concerned with the forward-backward stochastic interpolation formula (1.10) stated in theorem 1.2. Two approaches are presented: The first one discussed in section 4.1 is based on an extension of the two-sided stochastic calculus introduced by Pardoux and Protter in [43] to stochastic interpolation flows. The second one discussed in section 4.2 is based on the generalized backward Itô-Ventzell formula. This section also discusses a multivariate Skorohod-Alekseev-Gröbner formula. Apart from more complex and sophisticated tensor notation, the quantitative stochastic analysis of these multivariate formulae follows the same arguments as the ones used in the proof of theorem 1.3. Thus, we have chosen to concentrate this introduction on stochastic flows.

Some extensions of the stochastic interpolation formula (1.10) are discussed in section 4.4.

Section 5 is dedicated to the analysis of the Skorohod fluctuation process introduced in (1.12).

Section 6 is dedicated to the analysis of an extended version of two-sided stochastic integrals and a generalized backward Itô-Ventzell formula.

Section 7 presents some illustrations of the forward-backward interpolation formulae discussed in the present article in the context of diffusion perturbation theory, interacting diffusions and discrete time approximations.

The technical proofs of some results are housed in the appendix.

1.1 Statement of some main results

1.1.1 A backward Itô-Ventzell formula

We represent the gradient of a real valued function of several variables as a column vector while the gradient and the Hessian of a (column) vector valued function as tensors of type $(1,1)$ and $(2,1)$, see for instance (2.2) and (2.3); in more layman terms a $(1,1)$ tensor is a matrix while the $(2,1)$ tensor can be visualized as a "row of matrices" $[A_1,\ldots,A_n]$ where the entries A_i are matrices of a common dimension. We also use the tensor product and the transpose operator defined in (2.1), see also (2.4).

We denote by D_t the Malliavin derivative from some dense domain $\mathbb{D}_{2,1}\subset\mathbb{L}_2(\Omega)$ into the space $\mathbb{L}_2(\Omega\times\mathbb{R}_+;\mathbb{R}^r)$. For multivariate d-column vector random variables F with entries F^j, we use the same rules as for the gradient and D_tF is the (r,p)-matrix with entries $(D_tF)_{i,j}:=D_t^iF^j$. For $(p\times q)$-matrices F with entries F_k^j we let D_tF be the tensor with entries $(D_tF)_{i,j,k}=D_t^iF_k^j$.

For a more thorough discussion on Malliavin derivatives and Skorohod integration we refer to section 2.3.

Let F be some function from $\mathbb{R}^p$ into $\mathbb{R}^q$, and let $y\in\mathbb{R}^p$ be some given state, for some $p,q\geqslant 1$. Suppose we are given a forward p-dimensional continuous semi-martingale $Y_{s,t}$ and a backward random field $F_{s,t}$ from $\mathbb{R}^p$ into $\mathbb{R}^q$ with a column-vector type canonical representation of the following form:

$$\left\{\begin{array}{rcl} Y_{s,t} &=& y+\displaystyle\int_s^t B_{s,u}\,du+\int_s^t \Sigma_{s,u}\,dW_u\\ F_{s,t}(x) &=& F(x)+\displaystyle\int_s^t G_{u,t}(x)\,du+\int_s^t H_{u,t}(x)\,dW_u\end{array}\right. \tag{1.4}$$

for some $\mathcal{W}_{s,t}$-adapted functions $B_{s,t},G_{s,t},H_{s,t},\Sigma_{s,t}$ with appropriate dimensions and satisfying the following conditions:

(H_1): *The functions $F_{s,t}$, $G_{u,t}$ and $H_{u,t}$ as well as $\nabla H_{u,t}$, $\nabla^2F_{u,t}$ and the derivatives $D_v\nabla F_{u,t}$ and $D_vH_{u,t}$ are continuous w.r.t. the state and the time variables for any given $\omega\in\Omega$.*

(H_2) *The function $G_{u,t},\nabla H_{u,t},\nabla^2F_{u,t}$, and the derivatives $D_vH_{u,t},D_v\nabla F_{u,t}$ have at most polynomial growth w.r.t. the state variable, uniformly with respect to $\omega\in\Omega$.*

(H_3) *The processes $B_{s,u},\Sigma_{s,u}$ as well as $D_v\Sigma_{s,u}$ are continuous and have moments of any order.*

In this notation, the first main result of this article is the following theorem.

Theorem 1.1. *Assume conditions $(H_i)_{i=1,2,3}$ are satisfied. In this situation, for any $s\leqslant u\leqslant v\leqslant t$ we have the generalized backward Itô-Ventzell formula*

$$\begin{aligned} F_{v,t}(Y_{s,v})-F_{u,t}(Y_{s,u}) &= \int_u^v \left(\nabla F_{r,t}(Y_{s,r})'\,B_{s,r}+\frac{1}{2}\,\nabla^2F_{r,t}(Y_{s,r})'\,\Sigma_{s,r}\Sigma'_{s,r}-G_{r,t}(Y_{s,r})\right)dr\\ &\quad+\int_u^v \left(\nabla F_{r,t}(Y_{s,r})'\,\Sigma_{s,r}-H_{r,t}(Y_{s,r})\right)\,dW_r \end{aligned} \tag{1.5}$$

The stochastic anticipating integral in the r.h.s. of 1.5 is understood as a Skorohod stochastic integral.

The above theorem can be seen as the backward version of the generalized Itô-Ventzell formula presented in [41, 42]. The proof of the above theorem is provided in section 6.2 (see theorem 6.3).

Conventional forward and backward Itô stochastic integrals are particular instances of the two-sided stochastic integrals introduced by Pardoux and Protter in [43]. The terminology " two-sided

" coined by the authors in [43] comes from the fact that the integrand of the Skorohod integral depend on the past as well as on the future of the history generated by the Brownian motion.

The stochastic anticipating integral in the r.h.s. of (1.5) involves a backward random field and a forward semimartingale, thus it is tempting to interpret this integral as a two sided integral. Unfortunately, this class of integrands are not considered in the construction of the two-sided stochastic integrals defined in [43]. In section 4.1 and section 6.1 we shall present an extended version of the two-sided stochastic integrals introduced in [43] that applies to integrands defined as a compositions of backward and forward stochastic flows. This extended version applies to backward stochastic flows but it doesn't encapsulate more general backward random fields. We believe more general extensions of the two-sided integrals can be developed but it is out of the scope of this article to develop a theory on generalized two-sided stochastic integrals. We finally mention that all two-sided stochastic integrals discussed in this article are particular instances of Skorohod integrals

1.1.2 A stochastic flow interpolation formula

The diffusion flow (1.1) is defined in term of a column vector with twice continuously differentiable entries. For $h \simeq 0$ we use the backward approximation:

$$\begin{aligned}
&X_{s,t}(x) - X_{s-h,t}(x) = X_{s,t}(x) - (X_{s,t} \circ X_{s-h,s})(x) \\
&\simeq X_{s,t}(x) - X_{s,t}\left(x + b_s(x)\ h + \sigma_s(x)\ \ (W_s - W_{s-h})\right) \\
&\simeq -\left[\left(\nabla X_{s,t}(x)'\ b_s(x) + \frac{1}{2}\ \nabla^2 X_{s,t}(x)'\ a_s(x)\right)\ h + \nabla X_{s,t}(x)'\sigma_s(x)\ (W_s - W_{s-h})\right]
\end{aligned} \tag{1.6}$$

In the above display, $X_{s,t} \circ X_{s-h,s}$ stands for the composition of the mappings $X_{s,t}$ and $X_{s-h,s}$.

The above approximations are rigorously justified in section 4.1 and lead to the backward stochastic flow evolution equation:

$$d_s X_{s,t}(x) = -\left[\left(\nabla X_{s,t}(x)'\ b_s(x) + \tfrac{1}{2}\ \nabla^2 X_{s,t}(x)'\ a_s(x)\right)\ ds + \nabla X_{s,t}(x)'\sigma_s(x)\ dW_s\right] \tag{1.7}$$

In the above display, $d_s X^i_{s,t}(x)$ represents the change in $X^i_{s,t}(x)$ w.r.t. the variable s.

In the same vein, for any $s < u < t$ we have the interpolating semigroup decompositions

$$\begin{aligned}
&X_{u+h,t} \circ \overline{X}_{s,u+h} - X_{u,t} \circ \overline{X}_{s,u} \\
&= (X_{u+h,t} - X_{u,t}) \circ \overline{X}_{s,u} + \left(X_{u+h,t} \circ \overline{X}_{s,u+h} - X_{u+h,t} \circ \overline{X}_{s,u}\right)
\end{aligned}$$

as well as the forward approximations

$$\begin{aligned}
&X_{u+h,t}\left(\overline{X}_{s,u}(x) + \left(\overline{X}_{s,u+h}(x) - \overline{X}_{s,u}(x)\right)\right) - X_{u+h,t}(\overline{X}_{s,u}(x)) \\
&\simeq (\nabla X_{u+h,t})\,(\overline{X}_{s,u}(x))'\ (\overline{X}_{s,u+h}(x) - \overline{X}_{s,u}(x)) + \frac{1}{2}\ \left(\nabla^2 X_{u+h,t}\right)(\overline{X}_{s,u}(x))'\ \overline{a}_u(\overline{X}_{s,u}(x))\ h
\end{aligned} \tag{1.8}$$

The above approximations are rigorously justified in section 4.1 and lead to the forward-backward stochastic interpolation equation

$$\begin{aligned}
&d_u\left(X_{u,t} \circ \overline{X}_{s,u}\right)(x) \\
&= (d_u X_{u,t})\,(\overline{X}_{s,u}(x)) + (\nabla X_{u,t})\,(\overline{X}_{s,u}(x))'\ d_u \overline{X}_{s,u}(x) + \frac{1}{2}\ \left(\nabla^2 X_{u,t}\right)(\overline{X}_{s,u}(x))'\ \overline{a}_u(\overline{X}_{s,u}(x))\ du
\end{aligned} \tag{1.9}$$

The discrete time version of the forward-backward stochastic formula in the above display reduces to the telescoping sum formula (4.2) and the second order Taylor expansions discussed in section 4.1. We already mention that (4.2) can be interpreted as a discrete time version of the Alekseev-Gröbner lemma [1, 24]. The terminology *forward-backward* comes from the forward and backward nature of (1.9) and the telescoping sum formula (4.2).

Also notice that (1.7) can also be deduced formally from (1.9) by replacing $\overline{X}_{s,u}$ by the stochastic flow $X_{s,u}$ in (1.9), and then letting $s = u$.

This yields the following interpolation theorem.

Theorem 1.2. *We have the forward-backward stochastic interpolation formula*

$$X_{s,t}(x) - \overline{X}_{s,t}(x) = T_{s,t}(\Delta a, \Delta b)(x) + S_{s,t}(\Delta\sigma)(x) \tag{1.10}$$

with the stochastic process

$$\begin{aligned} & T_{s,t}(\Delta a, \Delta b)(x) \\ & := \int_s^t \left[(\nabla X_{u,t}) \, (\overline{X}_{s,u}(x))' \; \Delta b_u(\overline{X}_{s,u}(x)) + \frac{1}{2} \; \left(\nabla^2 X_{u,t}\right) (\overline{X}_{s,u}(x))' \; \Delta a_u(\overline{X}_{s,u}(x)) \right] \, du \end{aligned} \tag{1.11}$$

and the fluctuation term given by the Skorohod stochastic integral

$$S_{s,t}(\Delta\sigma)(x) := \int_s^t \; (\nabla X_{u,t}) \, (\overline{X}_{s,u}(x))' \; \Delta\sigma_u(\overline{X}_{s,u}(x)) \; dW_u \tag{1.12}$$

The fluctuation term in the above display can also be seen as the extended two-sided stochastic integral defined in (4.3) (see also proposition 6.2).

These interpolation formulae combine the backward evolution (1.7) with the conventional forward evolution of the perturbed flow.

The proof of the interpolation formula (1.10) is provided in section 4.

We will present two different approaches: The first one presented in section 4.1 is rather elementary and very intuitive. It combines the conventional Itô-type discrete time approximations of stochastic integrals discussed above with the two-sided stochastic integration calculus introduced in [43]. Using this approximation technique the fluctuation term is defined by the extended two-sided stochastic integral defined in (4.3). In this interpretation, the equation (1.10) can be seen as an extended version of the Itô-type change rule formula stated in theorem 6.1 in the article [43] to the interpolating flow

$$Z^{s,t} \; : \; u \in [s,t] \; \mapsto \; Z^{s,t}_u := X_{u,t} \circ \overline{X}_{s,u} \quad \Longrightarrow \quad Z^{s,t}_s - Z^{s,t}_t = X_{s,t} - \overline{X}_{s,t} \tag{1.13}$$

Roughly speaking, the increments of the interpolating path are decomposed into two parts:

One comes from the backward increments of the flow $u \mapsto X_{u,t}$ given the past values of the stochastic flow $\overline{X}_{s,u}$. The other one comes from the conventional Itô increments of $u \mapsto \overline{X}_{s,u}$ given the future values of the stochastic flow $X_{u,t}$.

The second approach discussed in section 4.2 is based on the generalized backward Itô-Ventzell formula stated in theorem 1.1. More precisely we also recover (1.10) from (1.5) by choosing

$$\begin{aligned} (F_{s,t}(x), Y_{s,t}(y)) &= (X_{s,t}(x), \overline{X}_{s,t}(y)) \qquad (B_{s,t}, \Sigma_{s,t}) = \left(\overline{b}_t\left(\overline{X}_{s,t}(x)\right), \overline{\sigma}_t\left(\overline{X}_{s,t}(x)\right)\right) \\ G_{u,t}(x) &= \nabla F_{u,t}(x)' \; b_u(x) + \frac{1}{2} \; \nabla^2 F_{u,t}(x)' \; a_u(x) \quad \text{and} \quad H_{u,t}(x) = \nabla F_{u,t}(x)' \; \sigma_u(x) \end{aligned}$$

and letting $(u,v)=(s,t)$ in (1.5). The regularity conditions on the drift and the diffusion function ensure that conditions $(H_i)_i$ with $i=1,2,3$ stated in section 1.1.1 are satisfied.

We emphasize that the backward diffusion flow discussed in (1.7) and (4.1) is essential to apply theorem 1.1. Section 4.2 also provides a multivariate version of (1.10).

The interpolation formula (1.10) with a fluctuation term given by the Skorohod stochastic integral (1.12) can be seen as a Alekseev-Gröbner formula of Skorohod type.

In this context, the integrability of the fluctuation term and any quantitative type estimates require a refined analysis of the Malliavin derivatives of the integrand. Under our regularity conditions the stochastic flows $X_{s,t}(x)$ and $\overline{X}_{s,t}(x)$ are Holder-continuous w.r.t. the time parameters as well as twice differentiable w.r.t. the space variables, with almost sure uniformly bounded first and second order derivatives. In addition, for any $n\geqslant 1$ all the n-absolute moments of the stochastic flows are finite with at most linear growth w.r.t. the initial values. These properties ensure that the Skorohod stochastic integral (1.12) is well defined and they allow to derive several quantitative estimates. Section 5 provides a refined of the fluctuation term; see for instance theorem 5.2.

When $\sigma_t=0$ the flow $X_{s,t}(x)$ is deterministic so that the Skorohod fluctuation term (1.12) reduces to the traditional Itô stochastic integral. In this context, quantitative estimates of the fluctuation term are obtained combining Burkholder-Davis-Gundy inequalities with the generalized Minkowski inequality. The resulting interpolation formula (1.10) can be seen as a Alekseev-Gröbner formula of Itô-type.

To distinguish these two classes of models, the interpolation formulae (1.10) associated with the case $\sigma_t=0$ will be called an Itô-Alekseev-Gröbner formula; the one associated with the case $\Delta\sigma_t\neq 0$ will be called a Skorohod-Alekseev-Gröbner formula.

1.1.3 Uniform estimates w.r.t. the time horizon

The final objective of this article is to derive uniform estimates w.r.t. the time parameter. Our methodology is mainly based on two different types of regularity conditions to be defined and discussed in detail in section 2.2:

• The first is a technical condition that ensures that the n-absolute moments of the flows $X_{s,t}$ and $\overline{X}_{s,t}$ are uniformly bounded w.r.t. the time horizon; we call this condition $(M)_n$.

• The second is a spectral condition on the gradient of the drift and diffusion matrices of the stochastic flows, which we call condition $(T)_n$. Without going into details, we state one usual case of interest: for constant diffusion functions (1.2) the spectral condition $(T)_n$ is met for any $n\geqslant 2$ as soon as the following log-norm conditions are met

$$\nabla b_t+(\nabla b_t)'\leqslant -2\lambda\, I \quad\text{and}\quad \nabla\overline{b}_t+(\nabla\overline{b}_t)'\leqslant -2\overline{\lambda}\, I \quad\text{for some } \lambda\wedge\overline{\lambda}>0, \tag{1.14}$$

To motivate the above condition consider a linear drift function of the form $b_t(x)=B_t\,x$ and $\sigma=0$. In this case the tangent process $\nabla X_{s,t}(x)$ satisfies a time-varying deterministic linear dynamical system

$$\partial_t\,\nabla X_{s,t}(x)=\nabla X_{s,t}(x)\;B_t'$$

The asymptotic behavior of this process cannot be characterized by the statistical properties of the spectral abscissa of the matrices B_t. Indeed, unstable semigroups associated with time-varying (deterministic) matrices B_t with negative eigenvalues are exemplified in [15, 49]. Conversely, stable semigroups with B_t having positive eigenvalues are given by Wu in [49]. In contrast, the uniform log-norm condition (1.14) provides a readily verifiable condition.

To describe with some precision the second main result of the article, we need to introduce some additional terminology. When there is no ambiguity, we denote by $\|.\|$ any (equivalent) norm on some finite dimensional vector space. For some multivariate function $f_t(x)$, for $(t,x) \in [0,\infty) \times \mathbb{R}^d$, let $\|f(x)\| := \sup_t \|f_t(x)\|$ and the uniform norm be $\|f\| := \sup_{t,x} \|f_t(x)\|$. For any $n \geqslant 1$ we also set

$$\left\|\left\| f(x) \right\|\right\|_n := \sup_{s\geqslant 0} \sup_{t\geqslant s} \mathbb{E}\left(\|f_t(\overline{X}_{s,t}(x))\|^n\right)^{1/n} \tag{1.15}$$

We denote by κ_n and $\kappa_{\delta,n}$ some constants that depend on some parameters n and (δ, n) but do not depend on the time horizon, nor on the space variable.

In this notation, the second main result of the article takes basically the following form.

Theorem 1.3. *Assume conditions $(M)_{2n/\delta}$ and $(T)_{2n/(1-\delta)}$ are satisfied for some parameters $n \geqslant 2$ and $\delta \in]0,1[$. In this situation, we have the time-uniform estimates*

$$\begin{aligned} &\mathbb{E}\left[\|X_{s,t}(x) - \overline{X}_{s,t}(x)\|^n\right]^{1/n} \\ &\leqslant \kappa_{\delta,n} \left(\left\|\left\|\Delta a(x)\right\|\right\|_{2n/(1+\delta)} + \left\|\left\|\Delta b(x)\right\|\right\|_{2n/(1+\delta)} + \left\|\left\|\Delta\sigma(x)\right\|\right\|_{2n/\delta} (1 \vee \|x\|) \right) \end{aligned} \tag{1.16}$$

For constant diffusion functions (1.2), the estimate simplifies to

$$(1.14) \Longrightarrow \forall n \geqslant 2 \quad \mathbb{E}\left[\|X_{s,t}(x) - \overline{X}_{s,t}(x)\|^n\right]^{1/n} \leqslant \kappa_n \left(\left\|\left\|\Delta b(x)\right\|\right\|_n + \|\Sigma - \overline{\Sigma}\| \right) \tag{1.17}$$

The estimates (1.16) come from (7.5) and (5.9). A more detailed proof is provided in the appendix, on page 51. The estimates (1.17) are direct consequences of (2.17) and (5.11).

When $\sigma_t = \overline{\sigma}_t$ the Skorohod term is indeed absent and (1.10) reduces to

$$X_{s,t}(x) - \overline{X}_{s,t}(x) = \int_s^t (\nabla X_{u,t})(\overline{X}_{s,u}(x))' \, \Delta b_u(\overline{X}_{s,u}(x)) \, du \tag{1.18}$$

We recover the interpolation formula for nonlinear stochastic flows presented in section 3.1 in the article [3]. In this context the analysis of $\mathbb{L}_n$-errors will proceed via two-step procedure. In section 3.1 we will derive the exponential bound

$$\sup_x \mathbb{E}(\| (\nabla X_{u,t})(x)\|_2^n)^{1/n} \leqslant \kappa_n \exp(-\lambda(n)\ (t-u)) \quad \text{for some} \quad \lambda(n) > 0$$

Using the Minkowski integral inequality in (1.18) yields

$$\mathbb{E}\left[\|X_{s,t}(x) - \overline{X}_{s,t}(x)\|^n\right]^{1/n} \leqslant \int_s^t \mathbb{E}\left[\| (\nabla X_{u,t})(\overline{X}_{s,u}(x))\|^n \times \|\Delta b_u(\overline{X}_{s,u}(x))\|^n\right]^{1/n} \, du.$$

A further conditioning argument and the above exponential bound on the tangent process yields

$$\mathbb{E}\left[\|X_{s,t}(x) - \overline{X}_{s,t}(x)\|^n\right]^{1/n} \leqslant \kappa_n \int_s^t \exp(-\lambda(n)\ (t-u)) du \quad \sup_{s\leqslant u} \mathbb{E}[\|\Delta b_u(\overline{X}_{s,u}(x))\|^n]^{1/n}.$$

Replacing the term outside the time integral with $\left\|\left\|\Delta b(x)\right\|\right\|_n$ yields the stated result in (1.16) excluding the terms representing the difference in the diffusions.

We illustrate one use of theorem 1.2 in the context of analyzing the error in discretising the diffusion $X_{s,t}(x)$ for some initial time point $s \geqslant 0$. Let $h > 0$ denote the discretisation interval size and for any $t \in [s + kh, s + (k+1)h[$ let

$$dX^h_{s,t}(x) = Y^h_{s,t}(x)\ dt + \Sigma\ dW_t \quad \text{with} \quad Y^h_{s,t}(x) := b\left(X^h_{s,s+kh}(x)\right)$$

for a fixed diffusion matrix $\sigma_t(x) = \Sigma$. Here $X^h_{s,t}(x)$ is the discretisation of $X_{s,t}(x)$ with resolution h. Note that that the drift at time t is not a function of the instantaneous value of $X^h_{s,t}(x)$, at time t, but rather the value it took at the largest discrete time-point before t. In section 4.4 we discuss how the formula in (1.10) also applies in this context and establish that

$$X^h_{s,t}(x) - X_{s,t}(x) = \int_s^t \ (\nabla X_{u,t})\,(X^h_{s,u}(x))' \ \left[Y^h_{s,u}(x) - b(X^h_{s,u}(x))\right] \ du.$$

This comparison result when combined with the regularity assumptions (1.19) yields the moment bound below.

Proposition 1.4. *Assume that*

$$\nabla b + (\nabla b)' \leqslant -2\lambda\ I \qquad \|\nabla b\| := \sup_x \|\nabla b(x)\| < \infty \quad \text{and} \quad \langle x, b(x)\rangle \leqslant -\beta\ \|x\|^2 \tag{1.19}$$

for some $\lambda > 0$, $\beta > 0$. In this situation, for any $n \geqslant 1$ we have the uniform estimates

$$\mathbb{E}\left(\|X^h_{s,t}(x) - X_{s,t}(x)\|^n\right)^{1/n} \leqslant \|\nabla b\| \ \left([\|b(0)\| + \widehat{m}_n(x)\ \|\nabla b\|]\ \ h + \sigma\ \sqrt{h}\right)/\lambda$$

where $\widehat{m}_n(x) \leqslant \kappa_n\ (1 + \|x\|)$.

Proposition 1.4 is proved in section 7.3. To apply proposition 1.4 to a Langevin diffusion with a convex potential $U(x)$, the drift would be $b_t(x) = -\nabla U(x)$ and the corresponding assumptions on $U(x)$ are typical.

1.2 Comments and comparisons with existing literature

The interpolation formula (1.10) can be interpreted as an extension of Alekseev-Gröbner lemma [1, 24, 30] as well as an extended version of the variation-of-constant and related Gronwall type lemma [8, 25] to diffusion processes. In this connection we underline that the forward-backward formula (1.10) differs from the stochastic Gronwall lemma presented in [45] based on particular classes of stochastic linear inequalities that doesn't involve Skorohod type integrals.

The forward-backward interpolation formula (1.10) can also be seen as an extension of theorem 6.1 in [43] on two-sided stochastic integrals to diffusion flows. This interpolation formula can also be interpreted as a backward version of the generalized Itô-Ventzell formula presented in [41] (see also theorem 3.2.11 in [37]).

Stochastic interpolation formulae of the form (1.10) and their discrete time version discussed in (4.2) are not really new. To describe their origins, it is worth to mention that the stochastic perturbations may come from auxiliary random sources, uncertainty propagations, as well as time discretization schemes and mean field type particle fluctuations.

The pivotal interpolating telescoping sum formula (4.2) and the second order forward-backward perturbation semigroup methodology discussed in the present article can also be found in chapter 7 in [18] for discrete time models as well as in the series of articles [20, 21, 22] published at the beginning of the 2000s, see also chapter 10 in [19]. In this context, the random perturbations come from the fluctuations of a genetic type particle interpretation of nonlinear Feynman-Kac semigroups.

The more recent articles [9, 10, 11] also provide a series of backward-forward interpolation formulae of the same form as (1.10) for stochastic matrix Riccati diffusion flows arising in data assimilation theory (cf. for instance theorem 1.3 in [11] as well as section 2.2 in [10] and the proof of theorem 2.3 in [9]). In this context, the random perturbations come from the fluctuations of a mean field particle interpretation of a class of nonlinear diffusions equipped with an interacting sample covariance matrix functional.

We underline that the Itô-Alekseev-Gröbner formula (4.6) discussed in [11] is an extension of the interpolation formula (1.10) to stochastic diffusion flows in matrix spaces. In this context the unperturbed model is given by the flow of a deterministic matrix Riccati differential equation and the random perturbations are described by matrix-valued diffusion martingales. The corresponding Itô-Alekseev-Gröbner formulae can be seen as a matrix version of theorem 1.2 in the present article when $\sigma = 0$. These stochastic interpolation formulae were used in [11] to quantify the fluctuation of the stochastic flow around the limiting deterministic Riccati equation, at any order. We will briefly discuss the analog of these Taylor type expansions in section 7.1 in the context of Euclidian diffusions.

The forward-backward perturbation methodology discussed in the present article has also been used in [3, 5] in the context of nonlinear diffusions and their mean field type interacting particle interpretations, see for instance section 2.3 in [5]. In this context, the random perturbations come from the fluctuations of a mean field particle interpretation of a class of nonlinear diffusions. The extended version of the Itô-Alekseev-Gröbner formula (1.18) to nonlinear diffusions is also discussed in section 3.1 in the article [3]. In this situation, the time varying drift and diffusion functions of the stochastic flows depend on some possibly different nonlinear measure valued semigroups which may start from two possibly different initial distributions. For a more thorough discussion on this class of nonlinear diffusions, we refer to the Itô-Alekseev-Gröbner formula (3.2) and corollary 3.2 in the article [3]. These Itô-Alekseev-Gröbner formulae correspond to theorem 1.2 in the present article when $\sigma = 0$.

The interpolating stochastic semigroup techniques discussed in the present article are also applied to mean field particle systems and deterministic nonlinear measure valued semigroups. In this context, the process $X_{s,t}$ is given a deterministic measure-valued process and $\overline{X}_{s,t}$ represents the evolution of the particle density profiles associated with an approximating mean field particle interpretation of $X_{s,t}$. For instance, the article [4] is concerned with interacting jumps models on path spaces, the second article [5] discusses the propagation of chaos properties of mean field type interacting diffusions. The stochastic interpolation formulae discussed in [4, 5] correspond to the case (1.10) with $\sigma = 0$ and or $\overline{\sigma} \neq \sigma$ (see for instance the interpolation formula (3.5), theorem 2.6, theorem 2.7 and the interpolating telescoping sum in section 1.2 in [5])

In the series of articles discussed above, as in (1.9) the central common idea is to analyse the evolution of the interpolating process (1.13) between a given process $X_{s,t}$ and some stochastic flow $\overline{X}_{s,t}$ with an extra level of randomness. In discrete time settings, the differential interpolation formula (1.9) can also recasted in terms of a telescoping sum of the same form as (4.2) combined with a second order Taylor expansion reflecting the differences between a stochastic semigroup and its perturbations, see for instance chapter 7 in [18].

In most of the application domains discussed above, this second order stochastic perturbation methodology has been developed to quantify uniformly w.r.t. the time horizon the propagations of some stochastic perturbations entering in *some deterministic and stable reference or unperturbed process.* In the context of Euclidian diffusions, this corresponds to the situation where the diffusion function $\sigma = 0$ (the case $\overline{\sigma} = 0$ can be treated by symmetry arguments). The Itô-Alekseev-Gröbner type formulae discussed in section 3.1 in the article [3] correspond to theorem 1.2 in the present article when $\sigma = \overline{\sigma}$.

The present article can be seen as a natural extension of the second order perturbation methodology developed in the above referenced articles to diffusion type perturbed processes when $\sigma \neq \overline{\sigma}$.

To the best of our knowledge, the first article considering the case $\sigma \neq \overline{\sigma}$ with $\sigma \neq 0$ and $\overline{\sigma} \neq 0$ is the independent work of Hudde-Hutzenthaler-Jentzen-Mazzonetto [27]. In this article, the authors discuss an Itô-Alekseev-Gröbner formula for abstract diffusion perturbation models of the form (4.11). Here again, as in the list of referenced articles discussed above, the common central idea is to use discrete time approximations and combine the pivotal interpolating telescoping sum formulae (4.2) with a second order Taylor expansion. Besides this fact and in contrast with our analysis, the fluctuation term (1.12) discussed in [27] cannot be interpreted in terms of the extended two-sided stochastic integral defined in (4.3) (see also proposition 6.2) but only in terms of a Skorohod stochastic integral. The study [27] is also based on a series of particularly chosen and custom regularity conditions. For instance, the authors assume that the abstract diffusion perturbation models are chosen so that the Skorohod fluctuation term exists without providing any quantitative type estimate. This work is also not connected to the two-sided stochastic integration calculus developed by Pardoux and Protter in [43] nor to any type of backward Itô-Ventzell formula.

We feel that our approach is more direct and intuitive as it relies on an extended version of Itô's change rule formula (1.9) to interpolating stochastic flows. It also allows to interpret the fluctuation term (1.12) as an extended two-sided stochastic integral.

In section 5 in the present article, we will also see that any quantitative analysis requires to estimate the absolute moments of the Malliavin derivatives of the stochastic integrands of the Brownian motion arising in the Skorohod fluctuation term. In our framework, these Malliavin derivatives depend on the gradient of both of the diffusion functions $(\sigma, \overline{\sigma})$ as well as on the tangent process of the perturbed diffusion flow. The quantitative analysis developed in 5 can be extended without difficulties to abstract diffusion perturbation models satisfying appropriate differentiability and integrability conditions.

The article [27] also presents an application to tamed Euler type discrete time approximations of a stochastic van-der-Pol process introduced in [47], simplifying the analysis provided in an earlier work [28]. In this situation, we underline that the Skorohod fluctuation term is null so that the resulting Alekseev-Gröbner type formula resumes to the simple and elementary case discussed in (1.18) and in the article [3]. As expected for this class of "unstable processes", the authors recast a series of $\mathbb{L}_2$-estimates discussed in [28] into a series of estimates that grow exponentially fast with respect to the time horizon.

In contrast with the present work, the above article doesn't discuss any quantitative uniform estimates w.r.t. the time horizon. The analysis presented in [27] is mainly concerned with the proof of a Skorohod-Alekseev-Gröbner type formula for abstract diffusion perturbation models and it doesn't apply to derive any type of estimates to general diffusion perturbation models without adding regularity conditions.

Besides its elegance the forward-backward interpolation formula (1.10) is clearly of rather poor mathematical and numerical interest without a better understanding of the variational processes and the Skorohod fluctuation term (1.12). A crucial problem is to avoid exceedingly pessimistic exponential estimates that grow exponentially fast w.r.t. the time horizon.

One advantage of the second order perturbation methodology developed in the present article is that it takes advantage of the stability properties of the tangent and the Hessian flow in the estimation of Skorohod fluctuation term and this sharpen analysis of the difference between stochastic flows. Our main contribution is to develop a refined analysis of these variational processes and the Skorohod fluctuation terms. We also deduce several uniform perturbation propagation estimates with respect to the time horizon, yielding what seems to be the first results of this type for this class of models.

The forward-backward stochastic interpolation formula (1.10) can also be extended to more general classes of stochastic flows on abstract state spaces. For instance the recent article [30] provides a deterministic first order version of (1.10) on abstract Banach spaces. The stochastic perturbation analysis developed in the series of articles [4, 5, 9, 10, 11, 20, 21, 22] and the books [18, 19] is applied to matrix-valued diffusions and measure valued processes, including mean field type interacting diffusions and Feynman-Kac type interacting jumps models.

The stability properties of these abstract models discussed above depend on the problem at hand. To focus on the main ideas without clouding the article with unnecessary technical details and sophisticated mathematical tools based on abstract ad hoc regularity conditions we have chosen to concentrate the article on diffusion flows on Euclidian spaces with simple and easily checked regularity conditions.

2 Preliminary results

2.1 Some basic notation

With a slight abuse of notation, we denote by I the identity $(d \times d)$-matrix, for any $d \geqslant 1$. We also denote by $\|.\|$ any (equivalent) norm on a finite dimensional vector space over $\mathbb{R}$. All vectors are column vectors by default.

We introduce some matrix notation needed from the onset.

We denote by $\mathrm{Tr}(A)$, $\|A\|_2 := \lambda_{max}(AA')^{1/2} = \lambda_{max}(A'A)^{1/2}$, resp. $\|A\|_F = \mathrm{Tr}(AA')^{1/2}$ and $\rho(A) = \lambda_{max}((A+A')/2)$ the trace, the spectral norm, the Frobenius norm, and the logarithmic norm of some matrix A. A' is the transpose of A and $\lambda_{max}(.)$ the largest eigenvalue. The spectral norm is sub-multiplicative or $\|AB\|_2 \leqslant \|A\|_2\|B\|_2$ and compatible with the Euclidean norm for vectors, by that we mean for a vector x we have $\|Ax\| \leqslant \|A\|_2\|x\|$.

Let $[n]$ be the set of n multiple indexes $i = (i_1, \ldots, i_n) \in \mathcal{I}^n$ over some finite set $\mathcal{I}$. We denote by $(A_{i,j})_{(i,j)\in[p]\times[q]}$ the entries of a (p,q)-tensor A with index set $\mathcal{I}$ for $[p]$ and $\mathcal{J}$ for $[q]$. For the sake of brevity, the index sets will be implicitly defined through the context.

For a given (p_1, q)-tensor A and a given (q, p_2) tensor B, AB and B' is a (p_1, p_2)-tensor resp. a (p_2, q)-tensor with entries given by

$$\forall (i,j) \in [p_1] \times [p_2] \qquad (AB)_{i,j} = \sum_{k\in[q]} A_{i,k}\, B_{k,j} \quad \text{and} \quad B'_{j,k} := B_{k,j}. \tag{2.1}$$

The symmetric part A_{sym} of a (p,p)-tensor is the (p,p)-tensor A_{sym} with entries

$$\forall (i,j) \in [p] \times [p] \qquad (A_{sym})_{i,j} = (A_{i,j} + A_{j,i})/2$$

We consider the Frobenius inner product given for any (p,q)-tensors A and B by

$$\langle A, B\rangle_F = \mathrm{Tr}(AB') = \sum_i (AB')_{i,i} \quad \text{and the norm} \quad \|A\|_F = \sqrt{\mathrm{Tr}(AA')}$$

For any (p,q)-tensors A and B we also check the Cauchy-Schwartz inequality

$$\langle A, B\rangle_F^2 \leqslant \|A\|_F\ \|B\|_F \quad \text{and} \quad \|A\|_2 \leqslant \|A\|_F \leqslant \mathrm{Card}(\mathcal{I})^p\ \|A\|_2 \quad \text{with} \quad \|A\|_2 := \lambda_{max}(AA')^{1/2}$$

For any tensors A, B with appropriate dimensions we have the inequality

$$\|AB\|_F \leqslant \|A\|_F\ \|B\|_F$$

Given some tensor valued function $T : (t,x) \mapsto T_t(x)$ we also set

$$\|T\|_F := \sup_{t,x} \|T_t(x)\|_F \qquad \|T\|_2 := \sup_{t,x} \|T_t(x)\|_2 \qquad \text{and} \qquad \|T\| := \sup_{t,x} \|T_t(x)\|$$

Given some smooth function $h(x)$ from $\mathbb{R}^p$ into $\mathbb{R}^q$ we denote by

$$\nabla h = \left[\nabla h^1, \ldots, \nabla h^q\right] \quad \text{with} \quad \nabla h^i = \left[\begin{array}{c} \partial_{x_1} h^i \\ \vdots \\ \partial_{x_p} h^i \end{array}\right] \tag{2.2}$$

the gradient (p,q)-matrix associated with the column vector-valued function $h = (h^i)_{1\leqslant i\leqslant q}$. Building on this notation: let $b : \mathbb{R}^n \to \mathbb{R}^p$ and let the mapping $x \to G(x) = h(b(x))$. Then $\nabla G(x) = \nabla b(x) \times \nabla h(b(x))$. Let

$$\nabla^2 h = \left[\nabla^2 h^1, \ldots, \nabla^2 h^q\right] \quad \text{with} \quad \nabla^2 h^i = \left[\begin{array}{ccc} \partial_{x_1,x_1} h^i & \ldots & \partial_{x_1,x_p} h^i \\ \vdots & \ldots & \vdots \\ \partial_{x_p,x_1} h^i & \ldots & \partial_{x_p,x_p} h^i \end{array}\right] \tag{2.3}$$

The Hessian $H = \nabla^2 h$ associated with the function $h = (h^i)_{1\leqslant i\leqslant q}$ is a $(2,1)$-tensor where $H_{(i,j),k} = (\nabla^2 h^k)_{i,j} = \partial_{x_i,x_j} h^k$. In this notation we can compactly represent the second order term of the Taylor expansion of the the vector valued function h. For a vector $y = (y_1, \ldots, y_p)'$

$$\left[\begin{array}{c} y'\ \nabla^2 h^1(x)\ y \\ \vdots \\ y'\ \nabla^2 h^q(x)\ y \end{array}\right] = \nabla^2 h(x)'\ yy'$$

where we have regarded the matrix yy' as the $(2,1)$-tensor Y with $Y_{(i,j),1} = y_i y_j$.

In the same vein, in terms of the tensor product (2.1), for any pair of column vector-valued function $h = (h^k)_{1\leqslant k\leqslant q}$ and $b = (b^i)_{1\leqslant i\leqslant p}$ and any matrix function $a = (a^{i,j})_{1\leqslant i,j\leqslant p}$ from $\mathbb{R}^p$ into $\mathbb{R}^q$, for any parameter $1 \leqslant k \leqslant q$ we also have

$$\begin{aligned}
\left(\nabla h(x)'\ b(x)\right)^k &= \sum_{1\leqslant i\leqslant p} (\nabla h(x))'_{k,i}\ b^i(x) = \sum_{1\leqslant j\leqslant p} \partial_{x_i} h^k(x)\ b^i(x) = \langle \nabla h^k(x), b(x)\rangle \\
\left(\nabla^2 h(x)'\ a(x)\right)^k &= \sum_{1\leqslant i,j\leqslant p} (\nabla^2 h(x))'_{k,(i,j)}\ a^{i,j}(x) \\
&= \sum_{1\leqslant i,j\leqslant p} \partial_{x_i,x_j} h^k(x)\ a^{i,j}(x) = \langle \nabla^2 h^k(x), a(x)\rangle_F
\end{aligned}$$

In a more compact form, the above formula takes the form

$$\nabla h(x)'\ b(x) = \left[\begin{array}{c} \langle \nabla h^1(x), b(x)\rangle \\ \vdots \\ \langle \nabla h^q(x), b(x)\rangle \end{array}\right] \quad \text{and} \quad \nabla^2 h(x)'\ a(x) = \left[\begin{array}{c} \langle \nabla^2 h^1(x), a(x)\rangle_F \\ \vdots \\ \langle \nabla^2 h^q(x), a(x)\rangle_F \end{array}\right] \tag{2.4}$$

For any $n \geqslant 1$ we let $\mathcal{P}_n(\mathbb{R}^d)$ be the convex set of probability measures μ_1, μ_2 on $\mathbb{R}^d$ with absolute n-th moment and equipped with the Wasserstein distance of order n denoted by

$$\mathbb{W}_n(\mu_1, \mu_2) := \inf \mathbb{E}(\|X_1 - X_2\|^n)^{1/n}$$

In the above display the infimum is taken over all pair or random variables (X_1, X_2) with marginal distributions (μ_1, μ_2). The stochastic transition semigroups associated with the flows $X_{s,t}(x)$ and $\overline{X}_{s,t}(x)$ are defined for any measurable function f on $\mathbb{R}^d$ by the formulae

$$\mathbb{P}_{s,t}(f)(x) := f(X_{s,t}(x)) \quad \text{and} \quad \overline{\mathbb{P}}_{s,t}(f)(x) := f(\overline{X}_{s,t}(x))$$

Given some column vector-valued function $f = (f^i)_{1\leqslant i\leqslant p}$, let $\mathbb{P}_{s,t}(f)$ and $P_{s,t}(f)$ denote the column vector-valued functions with entries $\mathbb{P}_{s,t}(f^i)$ and $P_{s,t}(f^i)$. Building on the tensor notation, let $\mathbb{P}_{s,t}(\nabla f)$ and $\mathbb{P}_{s,t}(\nabla^2 f)$ respectively denote the $(1,1)$ and $(2,1)$-tensor valued functions with entries

$$\mathbb{P}_{s,t}(\nabla f)(x)_{i,k} := \mathbb{P}_{s,t}(\partial_{x_i} f^k)(x) \quad \text{and} \quad \mathbb{P}_{s,t}(\nabla^2 f)(x)_{(i,j),k} := \mathbb{P}_{s,t}(\partial_{x_i,x_j} f^k)(x)$$

We also consider the random $(2,1)$ and $(2,2)$-tensors given by

$$\begin{aligned} \nabla^2 X_{s,t}(x)_{(i,j),k} &= \partial_{x_i,x_j} X^k_{s,t}(x) = \left[\nabla^2 X_{s,t}(x)\right]'_{k,(i,j)} \\ \left[\nabla X_{s,t}(x) \otimes \nabla X_{s,t}(x)\right]_{(i,j),(k,l)} &= \nabla X_{s,t}(x)_{i,k} \nabla X_{s,t}(x)_{j,l} = \left[\nabla X_{s,t}(x) \otimes \nabla X_{s,t}(x)\right]'_{(k,l),(i,j)} \end{aligned}$$

Throughout the rest of the article, unless otherwise stated $\kappa, \kappa_\epsilon, \kappa_n, \kappa_{n,\epsilon}$ denote constants whose values may vary from line to line but only depend on the parameters in their subscripts, i.e. $n \geqslant 0$ and $\epsilon > 0$, as well as on the parameters of the model; that is, on the drift and diffusion functions. We also use the letters $c, c_\epsilon, c_n, c_{n,\epsilon}$ to denote universal constants. Importantly these contants do not depend on the time horizon. We also consider the uniform log-norm parameters

$$\rho(\nabla\sigma)^2 := \sum_{1\leqslant k\leqslant r} \rho(\nabla\sigma_k)^2 \quad \text{and} \quad \rho_\star(\nabla\sigma) := \sup_{1\leqslant k\leqslant r} \rho(\nabla\sigma_k) \quad \text{with} \quad \rho(\nabla\sigma_k) := \sup_{t,x} \rho(\nabla\sigma_{t,k}(x)) \tag{2.5}$$

and the parameters $\chi(b,\sigma)$ defined by

$$\chi(b,\sigma) := c + \|\nabla^2 b\| + \|\nabla^2\sigma\|^2 + \rho_\star(\nabla\sigma)^2 \tag{2.6}$$

2.2 Regularity conditions and some preliminary results

We consider two different types of regularity conditions $(\mathcal{M})_n$ and $(\mathcal{T})_n$, indexed by some parameter $n \in [2,\infty[$, for the diffusion (b_t,σ_t).

$(\mathcal{M})_n$ There exists some parameter $\kappa_n \geqslant 0$ such that for any $x \in \mathbb{R}^d$ we have

$$m_n(x) := \sup_{s\leqslant t} \mathbb{E}\left(\|X_{s,t}(x)\|^n\right)^{1/n} \leqslant \kappa_n \ (1 \vee \|x\|)$$

$(\mathcal{T})_n$ There exists some parameter $\lambda_A > 0$ such that

$$A_t := \nabla b_t + (\nabla b_t)' + \sum_{1\leqslant k\leqslant r} \nabla\sigma_{k,t}(\nabla\sigma_{k,t})' \leqslant -2\lambda_A \ I \tag{2.7}$$

where $\sigma_{k,t}$ denotes the k-th column of σ_t. In addition, the following condition is satisfied

$$\lambda_A(n) := \lambda_A - \frac{d(n-2)}{2} \ \rho_\star(\nabla\sigma)^2 > 0 \tag{2.8}$$

We now define the corresponding assumptions for the diffusion $(\overline{b}_t, \overline{\sigma}_t)$.

$(\overline{\mathcal{M}})_n$ The regularity condition defined as in $(\mathcal{M})_n$ for the diffusion $(\overline{b}_t, \overline{\sigma}_t)$.

$(\overline{\mathcal{T}})_n$ Let $\overline{A}_t$ be the symmetric matrix defined as A_t in (2.7) when $(b_t,\sigma_t) = (\overline{b}_t, \overline{\sigma}_t)$. Assume there exists some $\lambda_{\overline{A}} > 0$ such that $\overline{A}_t \leqslant -2\lambda_{\overline{A}} \ I$. Furthermore, assume $\lambda_{\overline{A}}(n) > 0$ where $\lambda_{\overline{A}}(n)$ is defined as $\lambda_A(n)$ when $(\lambda_A, \sigma_t) = (\lambda_{\overline{A}}, \overline{\sigma}_t)$.

$(M)_n$ We write $(M)_n$ when both conditions $(\mathcal{M})_n$ and $(\overline{\mathcal{M}})_n$ are satisfied.

$(T)_n$ Both conditions $(\mathcal{T})_n$ and $(\overline{\mathcal{T}})_n$ are met, and let

$$\lambda_{A,\overline{A}}(n) := \lambda_A(n) \wedge \lambda_{\overline{A}}(n)$$

In practice, the uniform moment condition $(\mathcal{M})_n$ is often checked using Lyapunov techniques. For example we can use the following polynomial growth condition.

$(\mathcal{P})_n$ There exists some parameters $\alpha_i, \beta_i \geqslant 0$ with $i = 0,1,2$ such that for any $t \geqslant 0$ and any $x \in \mathbb{R}^d$ we have

$$\|\sigma_t(x)\|_F^2 \leqslant \alpha_0 + \alpha_1\|x\| + \alpha_2\|x\|^2 \quad \text{and} \quad \langle x, b_t(x)\rangle \leqslant \beta_0 + \beta_1\|x\| - \beta_2\|x\|^2 \tag{2.9}$$

for some norm $\|\sigma_t(x)\|$ of the matrix-valued diffusion function. In addition, we have

$$\beta_2(n) := \beta_2 - \frac{(n-1)}{2}\,\alpha_2 > 0$$

Lemma 2.1. *For any $n \geqslant 2$ we have*

$$(\mathcal{P})_n \quad \Longrightarrow \quad (\mathcal{M})_n \quad \text{with} \quad \kappa_n = 1 + \frac{(\gamma_1 + (n-2)\alpha_1) + (\gamma_0 + (n-2)\alpha_0)^{1/2}}{2\beta_2(n)^{1/2}} \tag{2.10}$$

The proof of the above assertion follows standard stochastic calculations, thus it is housed in the appendix, on page 47.

For one-dimensional geometric Brownian motions the condition $(\mathcal{P})_n$ is a sufficient and necessary condition for the existence of uniformly bounded absolute n-moments. In this case $(\mathcal{T})_n$ coincides with $(\mathcal{P})_n$ by setting

$$\lambda_A = \beta_2 - \alpha_2/2 \quad \text{and} \quad \alpha_2 = \rho_\star(\nabla\sigma)^2$$

Whenever condition $(M)_n$ is met for some $n \geqslant 2$, we also check the uniform estimates

$$\mathbb{E}\left(\|[X_{u,t} \circ \overline{X}_{s,u}](x)\|^n\right)^{1/n} \leqslant \kappa_n\ (1 + \|x\|) \tag{2.11}$$

with the same parameter κ_n as the one associated with the condition $(M)_n$.

Recalling that the functions $(b_t, \overline{b}_t)$ and $(\sigma_t, \overline{\sigma}_t)$ have at most linear growth, with the $\mathbb{L}_n$-norms $|\!|\!|.|\!|\!|_n$ introduced in (1.15) we also have that

$$|\!|\!|\Delta b(x)|\!|\!|_n \leqslant \kappa_{1,n}(1 \vee \|x\|) \quad \text{and} \quad |\!|\!|\Delta a(x)|\!|\!|_{n/2} \leqslant \kappa_{2,n}\ (1 \vee \|x\|)^2 \tag{2.12}$$

To give more insight where these assumptions will be used, we now briefly state the stability results that stem from them. Condition $(\mathcal{T})_n$ ensures that the exponential decays of the absolute and uniform n-moments of the tangent and the Hessian processes; that is, when $(\mathcal{T})_n$ is met for some $n \geqslant 2$ we have that

$$\mathbb{E}\left(\|\nabla X_{s,t}(x)\|^n\right)^{1/n} \vee \mathbb{E}\left(\|\nabla^2 X_{s,t}(x)\|^n\right)^{1/n} \leqslant \kappa_n\ e^{-\lambda(n)(t-s)} \quad \text{for some} \quad \lambda(n) > 0 \tag{2.13}$$

A more precise statement is provided in proposition 3.2 and proposition 3.10. These uniform estimates clearly imply, via a conditioning argument, that for any $n \geqslant 2$ and $s \leqslant u \leqslant t$ we have

$$\mathbb{E}\left(\|(\nabla X_{u,t})(\overline{X}_{s,u}(x))\|^n\right)^{1/n} \vee \mathbb{E}\left(\|(\nabla^2 X_{u,t})(\overline{X}_{s,u}(x))\|^n\right)^{1/n} \leqslant \kappa_n\ e^{-\lambda(n)(t-u)} \tag{2.14}$$

with the same parameters $(\kappa_n, \lambda(n))$ as in (2.13).

The case $\nabla\sigma = 0$ will also serve a useful purpose, for example in analysing the error of a numerical implementation as in proposition 1.4. For instance whenever $(\mathcal{T})_2$ is met we have the almost sure and uniform gradient estimates

$$\|\nabla X_{s,t}\|_2 := \sup_x \|\nabla X_{s,t}(x)\|_2 \leqslant e^{-\lambda_A(t-s)} \tag{2.15}$$

In addition, we have the almost sure and uniform Hessian estimates

$$\|\nabla^2 X_{s,t}\|_F := \sup_x \|\nabla^2 X_{s,t}(x)\|_F \leqslant \frac{d}{\lambda_A} \|\nabla^2 b\|_F \; e^{-\lambda_A(t-s)} \tag{2.16}$$

A proof of the above estimates is provided in the beginning of section 3.1 and section 3.2. In this situation, whenever $(\mathcal{T})_2$ is met we have

$$\mathbb{E}\left[\|T_{s,t}(\Delta a, \Delta b)(x)\|^n\right]^{1/n} \leqslant \kappa \; \left(|||\Delta b(x)|||_n + |||\Delta a(x)|||_n\right). \tag{2.17}$$

In the above display, $T_{s,t}(\Delta a, \Delta b)(x)$ stands for the stochastic process discussed in (1.11), and κ stands for some finite constant that doesn't depend on the parameter n. For instance, for a Langevin diffusion associated with some convex potential function U we have $b = -\nabla U$ and $\nabla\sigma = 0$. Then assuming

$$\begin{array}{l} \nabla^2 U \geqslant \lambda\, I \quad \Longrightarrow \quad (\mathcal{T})_2 \quad \text{is met} \\ \\ \Longrightarrow \quad \|\nabla X_{s,t}\|_2 \leqslant e^{-\lambda(t-s)} \quad \text{and} \quad \|\nabla^2 X_{s,t}\|_F \leqslant \dfrac{d}{\lambda} \|\nabla^3 U\|_F \; e^{-\lambda(t-s)} \end{array} \tag{2.18}$$

where the almost sure tangent and Hessian bounds follow from (2.15) and (2.16) respectively.

In practice, it is often easier to work with $a_t(x) = \sigma_t(x)\sigma_t(x)'$ than $\sigma_t(x)$ and we now discuss some ways of estimating $\Delta\sigma_t(x) = \sigma_t(x) - \overline{\sigma}_t(x)$ in terms of $\Delta a_t(x) = a_t(x) - \overline{a}_t(x)$ and in the reverse direction. The latter is straightforward:

$$\|\Delta a_t(x)\| \leqslant \|\Delta\sigma_t(x)\| \; \left[\|\sigma_t(x)\| + \|\overline{\sigma}_t(x)\|\right].$$

To estimate $\Delta\sigma_t$ in terms of Δa_t, assume the following ellipticity condition is satisfied

$$a_t(x) \geqslant \upsilon\, I \quad \text{and} \quad \overline{a}_t(x) \geqslant \upsilon\, I \quad \text{for some parameter } \upsilon > 0. \tag{2.19}$$

We recall the Ando-Hemmen inequality [2] for any symmetric positive definite matrices Q_1, Q_2

$$\|Q_1^{1/2} - Q_2^{1/2}\| \leqslant \left[\lambda_{min}^{1/2}(Q_1) + \lambda_{min}^{1/2}(Q_2)\right]^{-1} \; \|Q_1 - Q_2\| \tag{2.20}$$

for any unitary invariant matrix norm $\|.\|$. In the above display, $\lambda_{min}(.)$ stands for the minimal eigenvalue. We also have the square root inequality

$$Q_1 \geqslant Q_2 \Longrightarrow Q_1^{1/2} \geqslant Q_2^{1/2} \tag{2.21}$$

See for instance theorem 6.2 on page 135 in [26], as well as proposition 3.2 in [2]. A proof of (2.21) can be found in [7]. In this situation, using (2.20) and (2.21) we check that

$$\|\Delta\sigma_t(x)\| \leqslant \frac{1}{\sqrt{\upsilon}} \; \|\Delta a_t(x)\| \quad \text{and} \quad \|\sigma_t(x)\| \leqslant \|\sigma_t(0)\| + \frac{1}{\sqrt{\upsilon}} \; \left[\|a_t(x)\| + \|a_t(0)\|\right] \tag{2.22}$$

This provides a way to estimate the growth of $\sigma_t(x)$ in terms of the one of $a_t(x)$. For instance the estimate (1.16) combined with (2.22) implies that

$$\mathbb{E}\left[\|X_{s,t}(x)-\overline{X}_{s,t}(x)\|^n\right]^{1/n} \leqslant \kappa_{\delta,n} \left(|||\Delta b(x)|||_{2n/(1+\delta)} + |||\Delta a(x)|||_{2n/\delta}(1\vee\|x\|)\right)$$

• Assume that $(\overline{\mathcal{M}})_n$ is satisfied for some $n \geqslant 1$. Also let $f_t(x)$ be some multivariate function such that

$$\|f(0)\| := \sup_t \|f_t(0)\| < \infty \quad \text{and} \quad \|\nabla f\| := \sup_{t,x} \|\nabla f_t(x)\| < \infty$$

In this situation, we have the estimates

$$|||f(x)|||_n \leqslant \|f(0)\| + \|\nabla f\|\ \overline{m}_n(x) \quad \text{and therefore} \quad |||f(x)|||_n \leqslant \kappa_n\ (\|f(0)\| + \|\nabla f\|)\ (1\vee\|x\|)$$

2.3 Some results on anticipating stochastic calculus

In this section we review some results on Malliavin derivatives and Skorohod integration calculus which will be needed below. We restrict the presentation to unit time intervals. Let $(\Omega, \mathcal{W})$ be the canonical space equipped with the Wiener measure $\mathbb{P}$ associated with the r-dimensional Brownian motion W_t discussed in the introduction.

The Malliavin derivative D_t is a linear operator from some dense domain $\mathbb{D}_{2,1} \subset \mathbb{L}_2(\Omega)$ into the space $\mathbb{L}_2(\Omega\times[0,1];\mathbb{R}^r)$ of r-dimensional processes with square integrable states on the unit time interval. For multivariate d-column vector random variables F with entries F^i, we use the same rules as for the gradient and we set

$$D_tF = \left[D_tF^1,\ldots,D_tF^d\right] \quad \text{with} \quad D_tF^i = \left[\begin{array}{c} D_t^1F^i \\ \vdots \\ D^rF^i \end{array}\right]$$

For $(p\times q)$-matrices F with entries F_k^j we let D_tF be the tensor with entries

$$(D_tF)_{i,j,k} = D_t^iF_k^j$$

It is clearly out of the scope of this article to review the analytical construction of Malliavin differential calculus. For a more thorough discussion we refer the reader to the seminal book by Nualart [37], see also the more synthetic presentation in the articles [38, 41].

Formally, one can think the Malliavin derivatives D_t^iF of some $F \in \mathbb{D}_{2,1}$ as way to extract from the random variable F the integrand of Brownian increment dW_t^i. For instance, when $s \leqslant t$ we have

$$\begin{array}{rcl} D_t^iX_{s,t}(x) & = & \sigma_{t,i}(X_{s,t}(x)) \\ (D_t\,\nabla X_{s,t}(x))_{i,j,k} & = & D_t^i\,(\nabla X_{s,t}(x))_{j,k} := \left(\nabla X_{s,t}(x)\ \nabla\sigma_{t,i}(X_{s,t}(x))\right)_{j,k} \end{array} \tag{2.23}$$

As conventional differentials, for any smooth function G from $\mathbb{R}^d$ into $\mathbb{R}^{p\times q}$, Malliavin derivatives satisfy the chain rule properties

$$D_t^i(G_k^j\circ F) = \sum_{1\leqslant l\leqslant d} \left(\partial_{x_l}G_k^j\right)(F)\times D_t^iF^l \quad \Longleftrightarrow \quad D_t(G\circ F) = D_tF\ ((\nabla G)\circ F)$$

For instance, for any $s\leqslant u\leqslant v$ we have

$$D_u\left(X_{u,t}\circ X_{s,u}\right) = \left(D_uX_{s,u}\right)\ \left[(\nabla X_{u,t})\circ X_{s,u}\right] \text{ and } D_u\left(\varsigma_t\circ X_{s,t}\right) = \left(D_uX_{s,t}\right)\ \left[(\nabla\varsigma_t)\circ X_{s,t}\right] \tag{2.24}$$

In the same vein, we have

$$
\begin{aligned}
&D_u \left(\nabla X_{s,u} \ \left[(\nabla X_{u,t}) \circ X_{s,u} \right] \right) \\
&= (D_u \nabla X_{s,u}) \left[(\nabla X_{u,t}) \circ X_{s,u} \right] + (D_u X_{s,u} \otimes \nabla X_{s,u}) \left[\left(\nabla^2 X_{u,t} \right) \circ X_{s,u} \right]
\end{aligned} \tag{2.25}
$$

Let $\mathbb{L}_{2,1}(\mathbb{R}^r) \subset \mathbb{L}_2(\Omega \times [0,1]; \mathbb{R}^r)$ be the Hilbert space of r-dimensional process U_t with Malliavin differentiable entries $U_t^i \in \mathbb{D}_{2,1}$ equipped with the norm

$$
|||U||| := \mathbb{E}\left(\int_{[0,1]} \|U_t\|^2 \ dt \right)^{1/2} + \mathbb{E}\left(\int_{[0,1]^2} \|D_s U_t\|^2 \ ds\, dt \right)^{1/2}
$$

The Skorohod integral w.r.t. the Brownian motion W_t^i on the unit interval is defined a linear and continuous mapping from

$$
V \in \mathbb{L}_{2,1}(\mathbb{R}) \mapsto \int_0^1 V_t \ dW_t^i \in \mathbb{L}_2(\Omega)
$$

characterized by the two following properties

$$
\begin{aligned}
\mathbb{E}\left(\int_0^1 V_t \ dW_t^i \right) &= 0 \\
\mathbb{E}\left(\left(\int_0^1 V_t \ dW_t^i \right)^2 \right) &= \mathbb{E}\left(\int_{[0,1]} V_t^2 \ dt \right) + \mathbb{E}\left(\int_{[0,1]^2} D_s^i V_t \ D_t^i V_s \ ds\, dt \right)
\end{aligned} \tag{2.26}
$$

The above formula can be seen as an extended version of the Itô isometry to Skorohod integrals, for instance [39], as well as chapters 1.3 to 1.5 in the book by Nualart [37].

As for the Itô integral, the Skorohod integral w.r.t. the r-dimensional Brownian motion W_t of a matrix valued process with entries $V_k^i \in \mathbb{L}_{2,1}(\mathbb{R})$ is defined by the column vector with entries

$$
\left(\int_0^1 V_t \ dW_t \right)^i := \int_0^1 V_t^i \ dW_t := \sum_{1 \leqslant k \leqslant r} \int_0^1 V_{t,k}^i \ dW_t^k
$$

3 Variational equations

3.1 The tangent process

In terms of the tensor product (2.4), the gradient $\nabla X_{s,t}(x)$ of the diffusion flow $X_{s,t}(x)$ is given by the gradient $(d \times d)$-matrix

$$
d \nabla X_{s,t}(x) = \nabla X_{s,t}(x) \ \left[\nabla b_t \left(X_{s,t}(x) \right) \ dt + \sum_{1 \leqslant k \leqslant r} \nabla \sigma_{t,k} \left(X_{s,t}(x) \right) \ dW_t^k \right]
$$

where W_t^k is the k-th component of the Brownian motion. After some calculations we check that

$$
d \left[\nabla X_{s,t}(x) \, \nabla X_{s,t}(x)' \right] = \nabla X_{s,t}(x) \ A_t \left(X_{s,t}(x) \right) \ \nabla X_{s,t}(x)' \ dt + dM_{s,t}(x) \tag{3.1}
$$

with the matrix function $A_t(x)$ defined in (2.7) and the symmetric matrix valued martingale

$$
dM_{s,t}(x) := \sum_{1 \leqslant k \leqslant r} \nabla X_{s,t}(x) \left[\nabla \sigma_{t,k} \left(X_{s,t}(x) \right) + \nabla \sigma_{t,k} \left(X_{s,t}(x) \right)' \right] \nabla X_{s,t}(x)' \ dW_t^k
$$

These expansions, when combined with condition $(\mathcal{T})_2$, yield the following estimates of the difference between $X_{s,t}(x)$ and $X_{s,t}(y)$.

Proposition 3.1. *Assume* $(\mathcal{T})_2$ *is satisfied. Then*

$$\mathbb{E}\left(\|X_{s,t}(x)-X_{s,t}(y)\|^2\right)^{1/2} \leqslant \sqrt{d}\ e^{-\lambda_A(t-s)}\quad \|x-y\|. \tag{3.2}$$

In addition, we have the almost sure estimate

$$\nabla\sigma = 0 \Longrightarrow \|X_{s,t}(x)-X_{s,t}(y)\| \leqslant e^{-\lambda_A(t-s)}\quad \|x-y\| \tag{3.3}$$

Proof of Prop. 3.1. Whenever $(\mathcal{T})_2$ is met, we have the following uniform estimate from (3.1)

$$(\mathcal{T})_2 \Longrightarrow \mathbb{E}\left(\|\nabla X_{s,t}(x)\|_2^2\right)^{1/2} \leqslant \mathbb{E}\left(\|\nabla X_{s,t}(x)\|_F^2\right)^{1/2} \leqslant \sqrt{d}\ e^{-\lambda_A(t-s)} \tag{3.4}$$

where the $\sqrt{d}$ term arises from imposing the initial condition $\nabla X_{s,s}(x) = I$ on the resulting differential equation for $\partial_t \mathbb{E}\left(\|\nabla X_{s,t}(x)\|_F^2\right)^{1/2}$. In addition, when $\nabla\sigma = 0$ the martingale $M_{s,t}(x) = 0$ is null and as a consequence of (3.1) we have the following almost sure estimate

$$\|\nabla X_{s,t}\|_2 := \sup_x \|\nabla X_{s,t}(x)\|_2 \leqslant e^{-\lambda_A(t-s)} \tag{3.5}$$

The Taylor expansion

$$X_{s,t}(x) - X_{s,t}(y) = \int_0^1\ \nabla X_{s,t}(\epsilon x + (1-\epsilon)y)'(x-y)\ d\epsilon$$

$$\Longrightarrow \|X_{s,t}(x)-X_{s,t}(y)\|^2 \leqslant \left[\int_0^1\ \|\nabla X_{s,t}(\epsilon x + (1-\epsilon)y)\|_2^2\ d\epsilon\right]\ \|x-y\|^2$$

combined with (3.4) and (3.5) completes the proof. ◻

These contraction inequalities quantify the stability of the stochastic flow $X_{s,t}(x)$ w.r.t. the initial state x. For instance, the estimate (3.2) ensures that the Markov transition semigroup is exponentially stable; that is, we have that

$$\mathbb{W}_2\left(\mu_0 P_{s,t}, \mu_1 P_{s,t}\right) \leqslant c\ \exp\left[-\lambda_A(t-s)\right]\ \mathbb{W}_2\left(\mu_0,\mu_1\right) \tag{3.6}$$

For the Langevin diffusions discussed in (2.18) the stochastic flow is time homogeneous; that is we have that $X_{s,t} = X_{t-s} := X_{0,(t-s)}$ and $P_{s,t} = P_{t-s} := P_{0,(t-s)}$. In addition when $\sigma(x) = \sigma\ I$, the diffusion flow $X_t(x)$ has a single invariant measure on $\mathbb{R}^d$ given by the Boltzmann-Gibbs measure

$$\pi(dx) = \frac{1}{Z}\ \exp\left(-\frac{2}{\sigma^2}\,U(x)\right)\ dx \quad \text{with} \quad Z := \int\quad e^{-\frac{2}{\sigma^2}U(x)}\ dx \tag{3.7}$$

From (2.18), it follows that

$$\nabla^2 U \geqslant \lambda\ I \quad \Longrightarrow \quad \mathbb{W}_n\left(\mu P_{s,t},\pi\right) \leqslant \exp\left[-\lambda(t-s)\right]\ \mathbb{W}_n\left(\mu,\pi\right)$$

for all $n \geqslant 1$.

Taking the trace in (3.1) we also find that

$$d\,\|\nabla X_{s,t}(x)\|_F^2 = \mathrm{Tr}\left[\nabla X_{s,t}(x)\ A_t\left(X_{s,t}(x)\right)\ \nabla X_{s,t}(x)'\right]\ dt + dN_{s,t}(x)$$

with the martingale

$$dN_{s,t}(x) = \sum_{1\leqslant k\leqslant r} \mathrm{Tr}\left(\nabla X_{s,t}(x)\left[\nabla\sigma_{t,k}\left(X_{s,t}(x)\right) + \nabla\sigma_{t,k}\left(X_{s,t}(x)\right)'\right]\nabla X_{s,t}(x)'\right)\ dW_t^k$$

Observe that

$$\partial_t\langle N_{s,\cdot}(x)\rangle_t = \sum_k \mathrm{Tr}\left(\nabla X_{s,t}(x)\left[\nabla\sigma_{t,k}\left(X_{s,t}(x)\right) + \nabla\sigma_{t,k}\left(X_{s,t}(x)\right)'\right]\nabla X_{s,t}(x)'\right)^2$$

This implies that

$$\begin{aligned}\partial_t\mathbb{E}\left(\|\nabla X_{s,t}(x)\|_F^4\right) &= 2\ \mathbb{E}\left(\|\nabla X_{s,t}(x)\|_F^2\ \mathrm{Tr}\left[\nabla X_{s,t}(x)\ A_t\left(X_{s,t}(x)\right)\ \nabla X_{s,t}(x)'\right]\right)\\ &\quad + \sum_{1\leqslant k\leqslant r}\mathbb{E}\left(\mathrm{Tr}\left(\nabla X_{s,t}(x)\left[\nabla\sigma_{t,k}\left(X_{s,t}(x)\right) + \nabla\sigma_{t,k}\left(X_{s,t}(x)\right)'\right]\nabla X_{s,t}(x)'\right)^2\right)\end{aligned}$$

Whenever $(\mathcal{T})_2$ is met, we have the estimate

$$\partial_t\mathbb{E}\left(\|\nabla X_{s,t}(x)\|_F^4\right) \leqslant -4\left[\lambda_A - \rho(\nabla\sigma)^2\right]\ \mathbb{E}\left(\|\nabla X_{s,t}(x)\|_F^4\right)$$

with the uniform log-norm parameter $\rho(\nabla\sigma)$ defined in (2.5). This yields the estimate

$$\partial_t\mathbb{E}\left(\|\nabla X_{s,t}(x)\|_F^4\right)^{1/4} \leqslant -\left[\lambda_A - \rho(\nabla\sigma)^2\right]\ \mathbb{E}\left(\|\nabla X_{s,t}(x)\|_F^4\right)^{1/4}$$

More generally, we readily check the following result.

Proposition 3.2. *When condition $(\mathcal{T})_n$ is met we have the following time-uniform bounds,*

$$\mathbb{E}\left(\|\nabla X_{s,t}(x)\|_F^n\right)^{1/n} \leqslant \sqrt{d}\ e^{-\left[\lambda_A-(n-2)\rho(\nabla\sigma)^2/2\right](t-s)} \tag{3.8}$$

3.2 The Hessian process

In terms of the tensor product (2.1), we have the matrix diffusion equation

$$\begin{aligned}&d\,\nabla^2 X_{s,t}(x)\\ &= \left[\left[\nabla X_{s,t}(x)\otimes\nabla X_{s,t}(x)\right]\nabla^2 b_t(X_{s,t}(x)) + \nabla^2 X_{s,t}(x)\nabla b_t(X_{s,t}(x))\right]dt + d\mathcal{M}_{s,t}(x)\end{aligned}$$

with the null matrix initial condition $\nabla^2 X_{s,s}(x) = 0$ and the matrix-valued martingale

$$d\mathcal{M}_{s,t}(x) = \sum_{1\leqslant k\leqslant r}\left(\left[\nabla X_{s,t}(x)\otimes\nabla X_{s,t}(x)\right]\nabla^2\sigma_{t,k}(X_{s,t}(x)) + \nabla^2 X_{s,t}(x)\nabla\sigma_{t,k}(X_{s,t}(x))\right)\ dW_t^k$$

Consider the tensor functions

$$\upsilon_t := \sum_{1\leqslant k\leqslant d}\left(\nabla^2\sigma_{t,k}\right)\ \left(\nabla^2\sigma_{t,k}\right)' \quad\text{and}\quad \tau_t := \nabla^2 b_t + \sum_{1\leqslant k\leqslant d}\left(\nabla^2\sigma_{t,k}\right)\ \left(\nabla\sigma_{t,k}\right)' \tag{3.9}$$

After some computations, we check that

$$\begin{aligned}&d\left[\nabla^2 X_{s,t}(x)\nabla^2 X_{s,t}(x)'\right]\\ &= \Big\{\left[\nabla^2 X_{s,t}(x)\ A_t(X_{s,t}(x))\ \nabla^2 X_{s,t}(x)'\right] + 2\left[\left[\nabla X_{s,t}(x)\otimes\nabla X_{s,t}(x)\right]\ \tau_t(X_{s,t}(x))\ \nabla^2 X_{s,t}(x)'\right]_{sym}\\ &\qquad + \left[\left[\nabla X_{s,t}(x)\otimes\nabla X_{s,t}(x)\right]\upsilon_t(X_{s,t}(x))\left[\nabla X_{s,t}(x)\otimes\nabla X_{s,t}(x)\right]'\right]\Big\}\,dt + d\mathcal{N}_{s,t}(x)\end{aligned}$$

with the matrix function $A_t(x)$ defined in (2.7) and the tensor-valued martingale

$$d\mathcal{N}_{s,t}(x:) = 2 \sum_{1\leqslant k\leqslant r} \left\{[\nabla X_{s,t}(x)\otimes\nabla X_{s,t}(x)]\ \nabla^2\sigma_{t,k}(X_{s,t}(x))\ \nabla^2 X_{s,t}(x)' \right.$$

$$\left. +\nabla^2 X_{s,t}(x)\ \nabla\sigma_{t,k}(X_{s,t}(x))\ \nabla^2 X_{s,t}(x)'\right\}_{sym}\ dW_t^k$$

When $\nabla\sigma = 0$ the above equation reduces to

$$\begin{aligned}&\partial_t\ \left[\nabla^2 X_{s,t}(x)\nabla^2 X_{s,t}(x)'\right]\\ &= \left[\nabla^2 X_{s,t}(x)\ A_t(X_{s,t}(x))\ \nabla^2 X_{s,t}(x)'\right] + 2\left[[\nabla X_{s,t}(x)\otimes\nabla X_{s,t}(x)]\ \nabla^2 b_t(X_{s,t}(x))\ \nabla^2 X_{s,t}(x)'\right]_{sym}\end{aligned}$$

Whenever $(\mathcal{T})_2$ is met, taking the trace in the above display we check that

$$\partial_t\,\|\nabla^2 X_{s,t}(x)\|_F^2 \leqslant -2\lambda_A\ \|\nabla^2 X_{s,t}(x)\|_F^2 + 2\|\nabla^2 b\|_F\ \|\nabla X_{s,t}(x)\|_F^2\ \|\nabla^2 X_{s,t}(x)\|_F$$

This yields the estimate

$$\partial_t\,\|\nabla^2 X_{s,t}(x)\|_F \leqslant -\lambda_A\ \|\nabla^2 X_{s,t}(x)\|_F + \|\nabla^2 b\|_F\ \|\nabla X_{s,t}(x)\|_F^2$$

Using (2.15) this implies that

$$\|\nabla^2 X_{s,t}(x)\|_F \leqslant \|\nabla^2 b\|_F\ e^{-\lambda_A(t-s)}\ \int_s^t\ e^{\lambda_A(u-s)}\ \|\nabla X_{s,u}(x)\|_F^2\ du \leqslant \frac{d}{\lambda_A}\ \|\nabla^2 b\|_F\ e^{-\lambda_A(t-s)}$$

This ends the proof of the almost sure estimate (2.16).

For more general models, we have that

$$\begin{aligned}&d\,\|\nabla^2 X_{s,t}(x)\|_F^2\\ &= \left\{\mathrm{Tr}\left[\nabla^2 X_{s,t}(x)\ A_t(X_{s,t}(x))\ \nabla^2 X_{s,t}(x)'\right] + 2\ \mathrm{Tr}\left[[\nabla X_{s,t}(x)\otimes\nabla X_{s,t}(x)]\ \tau_t(X_{s,t}(x))\ \nabla^2 X_{s,t}(x)'\right]\right.\\ &\qquad \left.+\mathrm{Tr}\left[[\nabla X_{s,t}(x)\otimes\nabla X_{s,t}(x)]\,\upsilon_t(X_{s,t}(x))\,[\nabla X_{s,t}(x)\otimes\nabla X_{s,t}(x)]'\right]\right\} dt + dM_{s,t}(x)\end{aligned}$$

with a continuous martingale $M_{s,t}(x)$ with angle bracket

$$\begin{aligned}&\partial_t\langle M_{s,.}(x)\rangle_t\\ &= 4 \sum_{1\leqslant k\leqslant r} \mathrm{Tr}\left\{[\nabla X_{s,t}(x)\otimes\nabla X_{s,t}(x)]\ \nabla^2\sigma_{t,k}(X_{s,t}(x))\ \nabla^2 X_{s,t}(x)'\right.\\ &\qquad \left.+\nabla^2 X_{s,t}(x)\ \nabla\sigma_{t,k}(X_{s,t}(x))\ \nabla^2 X_{s,t}(x)'\right\}^2\end{aligned}$$

Proposition 3.3. *Assume $(\mathcal{T})_n$ is met. In this situation, for any $\epsilon > 0$ s.t. $\lambda_A(n) > \epsilon$ we have*

$$\mathbb{E}\left(\|\nabla^2 X_{s,t}(x)\|_F^n\right)^{1/n} \leqslant n\ \epsilon^{-1}\ \chi(b,\sigma)\ \exp\left(-\left[\lambda_A(n)-\epsilon\right](t-s)\right) \tag{3.10}$$

with the parameters $\chi(b,\sigma)$ and $\lambda_A(n)$ defined in (2.6) and (2.8).

In the above display, $\rho_\star(\nabla\sigma)$ is defined in (2.5). The proof of the above estimate is technical and thus housed in the appendix on page 48

3.3 Bismut-Elworthy-Li formulae

We further assume that ellipticity condition (2.19) is met. In this situation, we can extend gradient semigroup formulae to measurable functions using the Bismut-Elworthy-Li formula

$$\nabla P_{s,t}(f)(x) = \mathbb{E}\left(f(X_{s,t}(x))\ \tau^{\omega}_{s,t}(x)\right) \tag{3.11}$$

with the stochastic process

$$\tau^{\omega}_{s,t}(x) := \int_s^t \partial_u \omega_{s,t}(u)\ \nabla X_{s,u}(x)\ a_u(X_{s,u}(x))^{-1/2}\ dW_u$$

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

$$\omega_{s,t}(u) = \varphi\left((u-s)/(t-s)\right) \implies \partial_u \omega_{s,t}(u) = \frac{1}{t-s}\ \partial\varphi\left((u-s)/(t-s)\right) \tag{3.12}$$

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

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

Whenever $(\mathcal{T})_2$ is met, combining (3.4) with (3.11), for any f s.t. $\|f\| \leqslant 1$ we check that

$$\begin{aligned}
\|\nabla P_{s,t}(f)\|^2 &\leqslant \mathbb{E}\left(\|\tau^{\omega}_{s,t}(x)\|^2\right) \\
&\leqslant \kappa_1 \int_s^t e^{-2\lambda_A(u-s)}\ \|\partial_u \omega^{s,t}(u)\|^2\ du = \frac{\kappa_1}{t-s} \int_0^1 e^{-2\lambda_A(t-s)v}\ (\partial\varphi(v))^2\ dv
\end{aligned}$$

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

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

In this situation, we check that

$$\|\nabla P_{s,t}(f)\|^2 \leqslant \frac{\kappa_2}{\epsilon^2}\ \frac{1}{t-s} \int_{1-\epsilon}^1 e^{-2\lambda_A(t-s)v}\ dv$$

from which we find the rather crude uniform estimate

$$\|\nabla P_{s,t}(f)\| \leqslant \frac{\kappa}{\epsilon}\ \frac{1}{\sqrt{t-s}}\ e^{-\lambda_A(1-\epsilon)(t-s)} \tag{3.13}$$

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

$$\begin{aligned}
\nabla^2 P_{s,t}(f)(x) &= \mathbb{E}\left(f(X_{s,t}(x))\ \tau^{[2],\omega}_{s,t}(x) + \nabla X_{s,t}(x)\, \nabla f(X_{s,t}(x))\ \tau^{\omega}_{s,t}(x)'\right) &(3.14)\\
&= \mathbb{E}\left(f(X_{s,t}(x))\ \left[\tau^{[2],\omega}_{s,u}(x) + \nabla X_{s,u}(x)\ \tau^{\omega}_{u,t}(X_{s,u}(x))\, \tau^{\omega}_{s,u}(x)'\right]\right) &(3.15)
\end{aligned}$$

with the process

$$\tau^{[2],\omega}_{s,t}(x)$$

$$:= \int_s^t \partial_u \omega_{s,t}(u)\ \left[\nabla^2 X_{s,u}(x)\ a_u(X_{s,u}(x))^{-1/2} + (\nabla X_{s,u}(x) \otimes \nabla X_{s,u}(x))\ (\overline{\nabla} a_u^{-1/2})(X_{s,u}(x))\right]\ dW_u$$

In the above display $\overline{\nabla} a_u^{-1/2}$ stands for the tensor function

$$(\overline{\nabla} a_u^{-1/2}(x))_{(i,j),k} := \partial_{x_i} a_u^{-1/2}(x)_{j,k} = -\left(a_u^{-1/2}(x)\left[\partial_{x_i} a_u^{1/2}(x)\right] a_u^{-1/2}(x)\right)_{j,k}$$

A detailed proof of the formulae (3.14) and (3.15) in the context of nonlinear diffusion flows can be found in the appendix in [5].

Observe that

$$(2.19) \Longrightarrow \sup_i \|\partial_{x_i} a_u^{-1/2}(x)\| \leqslant c \ \|\nabla\sigma\|/\upsilon$$

Whenever $(\mathcal{T})_2$ is met, using the estimate (3.3) for any $\epsilon \in]0,1[$

$$\|\nabla^2 P_{s,t}(f)\| \leqslant \frac{\kappa}{\epsilon} \ \frac{1}{\sqrt{t-s}} \ e^{-\lambda_A(t-s)(1-\epsilon)} \ (\|f\| + \|\nabla f\|) \tag{3.16}$$

In the same vein, using (3.15) for any $u \in]s,t[$ and any bounded measurable function f s.t. $\|f\| \leqslant 1$ we also check the rather crude uniform estimate

$$\|\nabla^2 P_{s,t}(f)\| \leqslant \frac{\kappa_1}{\epsilon} \ \frac{1}{\sqrt{u-s}} \ e^{-\lambda_A(u-s)(1-\epsilon)} + \frac{\kappa_2}{\epsilon^2} \ \frac{1}{\sqrt{(t-u)(u-s)}} \ e^{-\lambda_A(u-s)} \ e^{-\lambda_A(t-s)(1-\epsilon)}$$

Choosing $u = s + (1-\epsilon)(t-s)$ in the above display we check that for any $\epsilon \in]0,1[$ we obtain the uniform estimate

$$\|\nabla^2 P_{s,t}(f)\| \leqslant \frac{c_1}{\epsilon\sqrt{1-\epsilon}} \ \frac{1}{\sqrt{t-s}} \ e^{-\lambda_A(1-\epsilon)^2(t-s)} + \frac{c_2}{\epsilon^2} \frac{1}{\sqrt{\epsilon(1-\epsilon)}} \ \frac{1}{t-s} \ \ e^{-2\lambda_A(1-\epsilon)(t-s)} \tag{3.17}$$

The extended versions of the above formulae in the context of diffusions on differentiable manifolds can be found in the series of articles [6, 13, 23, 36, 46].

4 Backward semigroup analysis

4.1 The two-sided stochastic integration

For any given time horizon $s \leqslant t$ we have the rather well known backward stochastic flow equation

$$X_{s,t}(x) = x + \int_s^t \left[\nabla X_{u,t}(x)' \ b_u(x) + \frac{1}{2} \quad \nabla^2 X_{u,t}(x)' \ a_u(x) \ \right] \ du + \int_s^t \nabla X_{u,t}(x)' \sigma_u(x) \ dW_u \tag{4.1}$$

The right hand side integral is understood as a conventional backward Itô-integral. In a more synthetic form, the above backward formula reduces to (1.7).

An elementary proof of the above formula based on Taylor expansions is presented in [17], different approaches can also be found in [31] and [33]. Extensions of the backward Itô formula (4.1) to jump type diffusion models as well as nonlinear diffusion flows can also be found in [16] and in the appendix of [3].

Consider the discrete time interval $[s,t]_h := \{u_0, \ldots, u_{n-1}\}$ associated with some refining time mesh $u_{i+1} = u_i + h$ from $u_0 = s$ to $u_n = t$, for some time step $h > 0$. In this notation, combining (1.6) with (1.8) for any $u \in [s,t]_h$ we have the Taylor type approximation

$$\begin{aligned}
&X_{u+h,t} \circ \overline{X}_{s,u+h} - X_{u,t} \circ \overline{X}_{s,u} \\
&\simeq -\left((\nabla X_{u+h,t}) \, (\overline{X}_{s,u}(x))' \ \Delta b_u(\overline{X}_{s,u}(x)) + \frac{1}{2} \ \left(\nabla^2 X_{u+h,t}\right) (\overline{X}_{s,u}(x))' \ \Delta a_u(\overline{X}_{s,u}(x)) \ \right) \ h \\
&\qquad - (\nabla X_{u+h,t}) \, (\overline{X}_{s,u}(x))' \ \Delta\sigma_u(\overline{X}_{s,u}(x)) \ (W_{u+h} - W_u)
\end{aligned}$$

This yields the interpolating forward-backward telescoping sum formula

$$
\begin{aligned}
&X_{s,t}(x) - \overline{X}_{s,t}(x)\\
&= - \sum_{u\in[s,t]_h} \left[X_{u+h,t}(\overline{X}_{s,u+h}(x)) - X_{u,t}\left(\overline{X}_{s,u}(x)\right)\right]\\
&\simeq \sum_{u\in[s,t]_h} \left(\left(\nabla X_{u+h,t}\right)\left(\overline{X}_{s,u}(x)\right)' \,\Delta b_u(\overline{X}_{s,u}(x)) + \frac{1}{2}\,\left(\nabla^2 X_{u+h,t}\right)\left(\overline{X}_{s,u}(x)\right)' \,\Delta a_u(\overline{X}_{s,u}(x)) \right)\, h\\
&\qquad\qquad + \sum_{u\in[s,t]_h} \left(\nabla X_{u+h,t}\right)\left(\overline{X}_{s,u}(x)\right)' \,\Delta\sigma_u(\overline{X}_{s,u}(x))\,\left(W_{u+h}-W_u\right)
\end{aligned}
\tag{4.2}
$$

We obtain formally (1.10) by summing the above terms and passing to the limit $h \downarrow 0$.

To be more precise, we follow the two-sided stochastic integration calculus introduced by Pardoux and Protter in [43]. As mentioned by the authors this methodology can be seen as a variation of Itô original construction of the stochastic integral. In this framework, the Skorohod stochastic integral (1.12) arising in (1.9) is defined by the $\mathbb{L}_2$-convergence

$$
\begin{aligned}
&S_{s,t}(\varsigma)(x)\\
&:= \lim_{h\to 0} \sum_{u\in[s,t]_h} \left(\nabla X_{u+h,t}\right)\left(\overline{X}_{s,u}(x)\right)' \,\varsigma_u(\overline{X}_{s,u}(x))\,\left(W_{u+h}-W_u\right) \quad \text{with} \quad \varsigma_u = \Delta\sigma_u
\end{aligned}
\tag{4.3}
$$

The proof of the above assertion is based on a slight extension of proposition 3.3 in [43] to Skorohod integrals of the form (1.12). For the convenience of the reader, a detailed proof of the above assertion for one dimensional models is provided in section 6.1.

Using (4.3), the complete proof of (1.9) now follows the same line of arguments as the ones used in the proof of Itô-type change rule formula stated in theorem 6.1 in [43], thus it is skipped.

4.2 A multivariate stochastic interpolation formulae

In terms of the tensor product (2.1), for any $p \geqslant 1$ and any twice differentiable function f from $\mathbb{R}^d$ into $\mathbb{R}^p$ with at most polynomial growth the function $F_{s,t} := \mathbb{P}_{s,t}(f)$ satisfies the backward formula (1.4) with the random fields

$$
G_{u,t}(x) := \nabla F_{u,t}(x)'\, b_u(x) + \frac{1}{2}\,\nabla^2 F_{u,t}(x)'\, a_u(x) \quad \text{and} \quad H_{u,t}(x) := \nabla F_{u,t}(x)'\,\sigma_u(x)
$$

Using the quantitative estimates presented in section 5.2, we checked that the regularity conditions (H_1), (H_2) and (H_3) stated in section 1.1.1 are satisfied. Rewritten in terms of the stochastic semigroups $\mathbb{P}_{s,t}$ and $\overline{\mathbb{P}}_{s,t}$ we obtain the forward-backward multivariate interpolation formula

$$
\mathbb{P}_{s,t}(f)(x) - \overline{\mathbb{P}}_{s,t}(f)(x) = \mathbb{T}_{s,t}(f,\Delta a,\Delta b)(x) + \mathbb{S}_{s,t}(f,\Delta\sigma)(x) \tag{4.4}
$$

with the stochastic integro-differential operator

$$
\begin{aligned}
&\mathbb{T}_{s,t}(f,\Delta a,\Delta b)(x)\\
&:= \int_s^t \left[\nabla\mathbb{P}_{u,t}(f)(\overline{X}_{s,u}(x))'\,\Delta b_u(\overline{X}_{s,u}(x)) + \frac{1}{2}\,\nabla^2\mathbb{P}_{u,t}(f)(\overline{X}_{s,u}(x))'\,\Delta a_u(\overline{X}_{s,u}(x))\right]\,du
\end{aligned}
\tag{4.5}
$$

and the two-sided stochastic integral term given by

$$\mathbb{S}_{s,t}(f,\Delta\sigma)(x) := \int_s^t \nabla\mathbb{P}_{u,t}(f)(\overline{X}_{s,u}(x))'\ \Delta\sigma_u(\overline{X}_{s,u}(x))\ dW_u \tag{4.6}$$

Using elementary differential calculus, for twice differentiable (column vector-valued) function f from $\mathbb{R}^d$ into $\mathbb{R}^p$ we readily check the gradient and the Hessian formulae

$$\begin{aligned} \nabla\,\mathbb{P}_{s,t}(f)(x) &= \nabla X_{s,t}(x)\ \mathbb{P}_{s,t}(\nabla f)(x) \\ \nabla^2\mathbb{P}_{s,t}(f)(x) &= [\nabla X_{s,t}(x)\otimes\nabla X_{s,t}(x)]\ \mathbb{P}_{s,t}(\nabla^2 f)(x) + \nabla^2 X_{s,t}(x)\ \mathbb{P}_{s,t}(\nabla f)(x) \end{aligned} \tag{4.7}$$

This shows that $\mathbb{T}_{s,t}(f,\Delta a,\Delta b)$ and $\mathbb{S}_{s,t}(f,\Delta\sigma)$ have the same form as the integrals $T_{s,t}(\Delta a,\Delta b)$ and $S_{s,t}(\Delta a,\Delta b)$ defined in (1.10) and (1.11) up to some terms involving the gradient and the Hessian of the function f. For instance, we have the two-sided stochastic integral formula

$$\mathbb{S}_{s,t}(f,\Delta\sigma)(x) = \int_s^t \mathbb{P}_{u,t}(\nabla f)(\overline{X}_{s,u}(x))'\ \nabla X_{u,t}(\overline{X}_{s,u}(x))'\ \Delta\sigma_u(\overline{X}_{s,u}(x))\ dW_u$$

Also observe that (4.4) coincides with (1.10) for the identity function; that is, we have that

$$f(x) = x \Longrightarrow \mathbb{T}_{s,t}(f,\Delta a,\Delta b) = T_{s,t}(\Delta a,\Delta b) \quad \text{and} \quad \mathbb{S}_{s,t}(f,\Delta\sigma) = S_{s,t}(\Delta\sigma)$$

The above discussion shows that the analysis of the differences of the stochastic semigroups $(\mathbb{P}_{s,t}-\overline{\mathbb{P}}_{s,t})$ in terms of the tangent and the Hessian processes is essentially the same as the one of the difference of the stochastic flows $(X_{s,t}-\overline{X}_{s,t})$. For instance using the discussion provided section 5.3, when the gradient and the Hessian of the function f are uniformly bounded the estimates stated in theorem 1.3 can be easily extended at the level of the stochastic semigroups.

The $\mathbb{L}_2$-norm of the two-sided stochastic integrals in (1.10) and (4.4) are uniformly estimated as soon as the pair of drift and diffusion functions (b_t,σ_t) and $(\overline{b},\overline{\sigma}_t)$ satisfy condition $(\mathcal{T})_2$. For a more thorough discussion we refer to section 5.1, see for instance the $\mathbb{L}_n$-norm estimates presented in theorem 5.2 applied to the difference function $\varsigma_t = \Delta\sigma_t$.

4.3 Semigroup perturbation formulae

Besides the fact that the Skorohod integral in the r.h.s. of (4.4) is not a martingale (w.r.t. the Brownian motion filtration) it is centered (see for instance (2.26) and the argument provided in the beginning of section 5.1). Thus, taking the expectation in the univariate version of (4.4) we obtain the following interpolation semigroup decomposition.

Corollary 4.1. *For any twice differentiable function f from $\mathbb{R}^d$ into $\mathbb{R}$ with bounded derivatives we have the forward-backward semigroup interpolation formula*

$$\begin{aligned} P_{s,t}(f)(x) - \overline{P}_{s,t}(f)(x) = & \int_s^t \mathbb{E}\left(\langle\nabla P_{u,t}(f)(\overline{X}_{s,u}(x)),\Delta b_u(\overline{X}_{s,u}(x))\rangle\right)\ du \\ & +\frac{1}{2}\int_s^t \mathbb{E}\left(\mathrm{Tr}\left[\nabla^2 P_{u,t}(f)(\overline{X}_{s,u}(x))\ \Delta a_u(\overline{X}_{s,u}(x))\right]\right)\ du \end{aligned} \tag{4.8}$$

In addition, under some appropriate regularity conditions for any differentiable function f such that $\|f\|\leqslant 1$ and $\|\nabla f\|\leqslant 1$ we have the uniform estimate

$$|P_{s,t}(f)(x) - \overline{P}_{s,t}(f)(x)| \leqslant \kappa\ \left[|||\Delta a(x)|||_1 + |||\Delta b(x)|||_1\right] \tag{4.9}$$

Rewritten in terms of the infinitesimal generators $(L_t, \overline{L}_t)$ of the stochastic flows $(X_{s,t}, \overline{X}_{s,t})$ we recover the rather well known semigroup perturbation formula

$$P_{s,t} = \overline{P}_{s,t} + \int_s^t \overline{P}_{s,u}(L_u - \overline{L}_u)P_{u,t}\, du \quad \Longleftrightarrow \quad (4.8)$$

The above formula can be readily checked using the interpolating formula given for any $s \leqslant u < t$ by the evolution equation

$$\partial_u(\overline{P}_{s,u}P_{u,t}) = (\partial_u \overline{P}_{s,u})P_{u,t} + \overline{P}_{s,u}(\partial_u P_{u,t}) = \overline{P}_{s,u}\overline{L}_u P_{u,t} - \overline{P}_{s,u}L_u P_{u,t}$$

Now we come to the proof of (4.9). Whenever $(\mathcal{T})_2$ is met, combining (3.13) with (3.16) for any differentiable function f s.t. $\|f\| \leqslant 1$ and $\|\nabla f\| \leqslant 1$ and for any $\epsilon \in]0,1[$ we check that

$$|P_{s,s+t}(f)(x) - \overline{P}_{s,s+t}(f)(x)| \leqslant \frac{\kappa}{\epsilon}\ [\vert\vert\vert \Delta a(x)\vert\vert\vert_1 + \vert\vert\vert \Delta b(x)\vert\vert\vert_1] \int_0^t \frac{1}{\sqrt{u}}\ e^{-\lambda_A(1-\epsilon)u}\ du$$

This ends the proof of (4.9). ∎

After some elementary manipulations the forward-backward interpolation formula (4.8) yields the following corollary.

Corollary 4.2. *Let X_t and $\overline{X}_t$ be some ergodic diffusions associated with some time homogeneous drift and diffusion functions (b,σ) and $(\overline{b},\overline{\sigma})$. The invariant probability measures π and $\overline{\pi}$ of X_t and $\overline{X}_t$ are connected for any twice differentiable function f from $\mathbb{R}^d$ into $\mathbb{R}$ with bounded derivatives by the following interpolation formula*

$$(\pi - \overline{\pi})(f) = \int_0^\infty \mathbb{E}\left(\left\langle \nabla P_t(f)(\overline{Y}), \Delta b(\overline{Y})\right\rangle + \frac{1}{2}\ \mathrm{Tr}\left[\nabla^2 P_t(f)(\overline{Y})\ \Delta a(\overline{Y})\right]\right)\ dt \qquad (4.10)$$

In the above display $\overline{Y}$ stands for a random variable with distribution $\overline{\pi}$ and P_t stands for the Markov transition semigroup of the process X_t.

The formula (4.10) can be used to estimate the invariant measure of a stochastic flow associated with some perturbations of the drift and the diffusion function.

For instance, for homogeneous Langevin diffusions X_t associated with some convex potential function U we have

$$b = -\nabla U \quad \text{and} \quad \sigma = I \quad \Longrightarrow \quad \pi(dx) \propto \exp\left(-2\,U(x)\right)\, dx$$

In the above display, dx stands for the Lebesgue measure on $\mathbb{R}^d$. In this situation, using (4.10), for any ergodic diffusion flow $\overline{X}_t$ with some drift $\overline{b}$ and an unit diffusion matrix we have

$$\overline{\pi}(f) = \pi(f) + \int_0^\infty \mathbb{E}\left(\left\langle(\overline{b} + \nabla U)(\overline{Y}), \nabla P_t(f)(\overline{Y})\right\rangle\right)\ dt$$

Notice that the above formula is implicit as the r.h.s. term depends on $\overline{\pi}$. By symmetry arguments, we also have the following more explicit perturbation formula

$$\overline{\pi}(f) = \pi(f) + \int_0^\infty \mathbb{E}\left(\left\langle(\overline{b} + \nabla U)(Y), \nabla \overline{P}_t(f)(Y)\right\rangle\right)\ dt$$

In the above display Y stands for a random variable with distribution π and $\overline{P}_t$ stands for the Markov transition semigroup of the process $\overline{X}_t$.

4.4 Some extensions

Several extensions of the forward-backward stochastic interpolation formula (1.10) to more general stochastic perturbation processes can be developed. For instance, suppose we are given some stochastic processes $\overline{Y}_{s,t}(x) \in \mathbb{R}^d$ and $\overline{Z}_{s,t}(x) \in \mathbb{R}^{d\times r}$ adapted to the filtration of the Brownian motion W_t, and let $\overline{X}_{s,t}(x)$ be the stochastic flow defined by the stochastic differential equation

$$d\overline{X}_{s,t}(x) = \overline{Y}_{s,t}(x)\ dt + \overline{Z}_{s,t}(x)\ dW_t \tag{4.11}$$

In this situation, the interpolation formula (1.9) remains valid when $\overline{a}_u(\overline{X}_{s,u}(x))$ is replaced by the stochastic matrices $\overline{Z}_{s,t}(x)\overline{Z}_{s,t}(x)'$. This yields without further work the forward-backward stochastic interpolation formula (1.10) with the local perturbations

$$\begin{array}{rcl}\Delta b_u(\overline{X}_{s,u}(x)) & := & b_u(\overline{X}_{s,u}(x)) - \overline{Y}_{s,u}(x)\\ \Delta\sigma_u(\overline{X}_{s,u}(x)) & := & \sigma_u(\overline{X}_{s,u}(x)) - \overline{Z}_{s,u}(x) \quad \text{and} \quad \Delta a_u(\overline{X}_{s,u}(x)) := a_u(\overline{X}_{s,u}(x)) - \overline{Z}_{s,u}(x)\overline{Z}_{s,u}(x)'\end{array}$$

The corresponding interpolation formula should be used with some caution as the $\mathbb{L}_2$-norm of the two-sided stochastic integral (1.12) depends on the Malliavin differential of the integrand process of the Brownian motion; see for instance the variance formula provided in lemma 5.1.

Assume that $\sigma = I$ and the regularity condition $(\mathcal{T})_2$ is met. Also suppose $\overline{X}_{s,t}(x)$ is given by a stochastic differential equation of the form (4.11) with $r = d$ and $\overline{Z}_{s,t}(x) = I$. Arguing as above, in terms of the tensor product (2.1) we have

$$X_{s,t}(x) - \overline{X}_{s,t}(x) = \int_s^t\ (\nabla X_{u,t})\,(\overline{X}_{s,u}(x))'\ (b_u(\overline{X}_{s,u}(x)) - \overline{Y}_{s,u}(x))\ du \tag{4.12}$$

Combining (2.15) with the generalized Minkowski inequality, we check the following proposition.

Proposition 4.3. *Assume that $(\mathcal{T})_2$ is met for some $\lambda_A > 0$. In this situation, for any $1 \leqslant n \leqslant \infty$ we have the estimates*

$$\mathbb{E}\left[\|X_{s,t}(x) - \overline{X}_{s,t}(x)\|^n\right]^{1/n} \leqslant \int_s^t\ e^{-\lambda_A(t-u)}\ \mathbb{E}\left[\|b_u(\overline{X}_{s,u}(x)) - \overline{Y}_{s,u}(x)\|\right]^{1/n}\ du \tag{4.13}$$

In the same vein, we have

$$P_{s,t}(f)(x) - \overline{P}_{s,t}(f)(x) = \int_s^t\ \mathbb{E}\left(\langle\nabla P_{u,t}(f)(\overline{X}_{s,u}(x)), b_u(\overline{X}_{s,u}(x)) - \overline{Y}_{s,u}(x)\rangle\right)\ du \tag{4.14}$$

For instance, for the Langevin diffusion discussed in (2.18) and (3.7) the weak expansion (4.14) implies that

$$[\pi\overline{P}_{s,t} - \pi](f) = \int_s^t\ \int\pi(dx)\ \mathbb{E}\left(\langle\nabla P_{t-u}(f)(\overline{X}_{s,u}(x)), \nabla U(\overline{X}_{s,u}(x)) + \overline{Y}_{s,u}(x)\rangle\right)\ du \tag{4.15}$$

This yields the $\mathbb{W}_1$-Wasserstein estimate

$$\mathbb{W}_1(\pi\overline{P}_{s,t}, \pi)| \leqslant \int_s^t\ e^{-\lambda_A(t-u)}\ \int\pi(dx)\ \mathbb{E}\left(\|\nabla U(\overline{X}_{s,u}(x)) + \overline{Y}_{s,u}(x)\|\right)\ du$$

Combining (3.13) with (4.15), for any $\epsilon \in]0,1[$ we also have the total variation norm estimate

$$\|\pi\overline{P}_{s,t} - \pi\|_{tv} \leqslant \frac{c}{\epsilon}\int_s^t\ \frac{1}{\sqrt{t-u}}\ e^{-\lambda_A(1-\epsilon)(t-u)}\ \left[\int\pi(dx)\ \mathbb{E}\left(\|\nabla U(\overline{X}_{s,u}(x)) + \overline{Y}_{s,u}(x)\|\right)\right]\ du \tag{4.16}$$

5 Skorohod fluctuation processes

5.1 A variance formula

Let $\varsigma_t(x)$ be some differentiable $(d \times r)$-matrix valued function on $\mathbb{R}^d$ such that

$$\|\nabla\varsigma\| < \infty \quad \text{and} \quad \|\varsigma(0)\| := \sup_t \|\varsigma_t(0)\| < \infty \tag{5.1}$$

Recalling that $(W_{u+h} - W_u)$ is independent of the flows $\overline{X}_{s,u}$ and $\nabla X_{u+h,t}$, the discrete time approximation (4.3) shows that Skorohod stochastic integral is centered; that is, we have that $\mathbb{E}(S_{s,t}(\varsigma)(x)) = 0$.

Following (4.3), the variance can be computed using the following approximation formula

$$\begin{aligned}\mathbb{E}\left[\|S_{s,t}(\varsigma)(x)\|^2\right] = \lim_{h\to 0} \sum_{u,v\in[s,t]_h} \sum_{1\leqslant i\leqslant d} \sum_{1\leqslant j,k\leqslant r} \\ \mathbb{E}\left\{\left[(\nabla X_{u+h,t})\,(\overline{X}_{s,u}(x))'\,\varsigma_u(\overline{X}_{s,u}(x))\right]_{i,j}\;\left[(\nabla X_{v+h,t})\,(\overline{X}_{s,v}(x))'\,\varsigma_v(\overline{X}_{s,v}(x))\right]_{i,k}\right. \\ \left.(W^j_{u+h} - W^j_u)(W^k_{v+h} - W^k_v)\right\}\end{aligned} \tag{5.2}$$

The proof of the above assertion is provided in section 6.1, see for instance proposition 6.2.

Consider the matrix valued function

$$\Sigma_{s,u,t}(x) := \left[(\nabla X_{u,t})' \circ \overline{X}_{s,u}\right](x)\;\varsigma_u(\overline{X}_{s,u}(x)) \tag{5.3}$$

In this notation, the limiting diagonal term $u = v$ in the r.h.s. of (5.2) is clearly equal to

$$\int_s^t \mathbb{E}\left[\sum_{i,j} \Sigma_{s,u,t}(x)_{i,j}\;\Sigma_{s,u,t}(x)_{i,j}\right] du = \int_s^t \mathbb{E}\left[\|\Sigma_{s,u,t}(x)\|_F^2\right]\,du$$

In addition, whenever condition $(\mathcal{T})_2$ is met and ς is bounded, (3.4) readily yields the estimate

$$\left[\int_s^t \mathbb{E}\left[\|\Sigma_{s,u,t}(x)\|_F^2\right]\,du\right]^{1/2} \leqslant \|\varsigma\|_2\;\sqrt{d/(2\lambda_A)} \tag{5.4}$$

More generally, using (3.8) whenever $(\overline{\mathcal{M}})_{2/\delta}$ and $(\mathcal{T})_{2/(1-\delta)}$ are met for some $\delta \in]0,1[$ we have the estimate

$$\mathbb{E}\left[\|\Sigma_{s,u,t}(x)\|^2\right] \leqslant c_{1,\delta}\;\left[\|\varsigma(0)\|^2 + \|\nabla\varsigma\|^2\;(1+\|x\|)^2\right]\;e^{-2\lambda_A(2/(1-\delta))(t-u)}$$

This implies that

$$\left[\int_s^t \mathbb{E}\left[\|\Sigma_{s,u,t}(x)\|_F^2\right]\,du\right]^{1/2} \leqslant\; c_{2,\delta}\;\left[\|\varsigma(0)\| + \|\nabla\varsigma\|\;(1+\|x\|)\right]/\sqrt{\lambda_A} \tag{5.5}$$

The non-diagonal term can be computed in a more direct way using Malliavin derivatives of the functions $\Sigma_{s,u,t}$. For any $s \leqslant u \leqslant v \leqslant t$ we have

$$D_v\left\{\left[(\nabla X_{u,t})' \circ \overline{X}_{s,u}\right]\;\left[\varsigma_u \circ \overline{X}_{s,u}\right]\right\} = \left[\left(D_v\,(\nabla X_{u,t})'\right) \circ \overline{X}_{s,u}\right]\;\left[\varsigma_u \circ \overline{X}_{s,u}\right] \tag{5.6}$$

As expected, observe that

$$\nabla\sigma = 0 \quad\Longrightarrow\quad D_v\Sigma_{s,u,t}(x) = 0$$

In the reverse angle, whenever $s \leqslant v \leqslant u \leqslant t$ we have the chain rule formula

$$\begin{aligned} & D_v\left(\left[\varsigma_u \circ \overline{X}_{s,u}\right]\ \left[(\nabla X_{u,t}) \circ \overline{X}_{s,u}\right]\right) \\ & := \left[D_v\left(\varsigma_u \circ \overline{X}_{s,u}\right)\right]\left[(\nabla X_{u,t}) \circ \overline{X}_{s,u}\right] + \left[D_v \overline{X}_{s,u} \otimes (\varsigma_u \circ \overline{X}_{s,u})\right]\left[\left(\nabla^2 X_{u,t}\right) \circ \overline{X}_{s,u}\right] \end{aligned} \tag{5.7}$$

As above, Malliavin differentials $D_v\left(\varsigma_u \circ \overline{X}_{s,u}\right)$ and $D_v \overline{X}_{s,u}$ can be computed using the chain rule formulae (2.24).

A more detailed analysis of the chain rules formulae (2.24), (2.25) and (5.7) for one dimensional models is provided in section 6.1 (cf. lemma 6.1).

Observe that

$$\nabla\varsigma = 0 \quad \Longrightarrow \quad D_v\left[\Sigma'_{s,u,t}\right] = \left[D_v \overline{X}_{s,u} \otimes (\varsigma_u \circ \overline{X}_{s,u})\right]\left[\left(\nabla^2 X_{u,t}\right) \circ \overline{X}_{s,u}\right]$$

We consider the inner product

$$\langle D_u \Sigma_{s,v,t}(x), D_v \Sigma_{s,u,t}(x)\rangle \quad := \quad \sum_{i,j,k} \left(D_v \Sigma_{s,u,t}(x)\right)_{k,i,j}\ \left(D_u \Sigma_{s,v,t}(x)\right)_{j,i,k}$$

In this notation, an explicit description of the $\mathbb{L}_2$-norm of the two-sided stochastic integral in terms of Malliavin derivatives is given below.

Lemma 5.1. *The $\mathbb{L}_2$-norm of the Skorohod integral $S_{s,t}(\varsigma)(x)$ introduced in (4.3) is given for any $x \in \mathbb{R}^d$ and $s \leqslant t$ by the formulae*

$$\mathbb{E}\left[\|S_{s,t}(\varsigma)(x)\|^2\right] = \int_{[s,t]} \mathbb{E}\left[\|\Sigma_{s,u,t}(x)\|_F^2\right]\ du + \int_{[s,t]^2} \mathbb{E}\left[\langle D_v \Sigma_{s,u,t}(x), D_u \Sigma_{s,v,t}(x)\rangle\right]\ du\ dv$$

with the random matrix function $\Sigma_{s,u,t}$ defined in (5.3) and the Malliavin derivative $D_v \Sigma_{s,u,t}$ given in formulae (5.6) and (5.7). In addition, we have

$$\nabla\sigma = 0 \quad \Longrightarrow \quad \mathbb{E}\left[\|S_{s,t}(\varsigma)(x)\|^2\right] = \int_{[s,t]} \mathbb{E}\left[\|\Sigma_{s,u,t}(x)\|_F^2\right]\ du$$

The above lemma can be interpreted as a matrix version of the isometry property (2.26). A proof of the above lemma based on the $\mathbb{L}_2$-approximation of two-sided stochastic integrals is provided in section 6.1 (see for instance proposition 6.2).

5.2 Quantitative estimates

For any $p > 1$ and any tensor norms we also quote the rather well known $\mathbb{L}_p$-norm estimates

$$\begin{aligned} & \mathbb{E}\left[\|S_{s,t}(\varsigma)(x)\|^p\right]^{2/p} \\ & \leqslant c_{1,p} \int_{[s,t]} \mathbb{E}\left[\|\Sigma_{s,u,t}(x)\|^2\right]\ du + c_{2,p}\ \mathbb{E}\left[\left(\int_{[s,t]^2} \|D_v \Sigma_{s,u,t}(x)\|^2\ du\ dv\right)^{p/2}\right]^{2/p} \end{aligned}$$

for some finite constants $c_{i,p}$ whose values only depend on p. A proof of these estimates can be found in [38, 48], see also [39] for multiple Skorohod integrals. By the generalized Minkowski inequality, for any $n \geqslant 2$ we also have the estimate

$$\begin{aligned}
&\mathbb{E}\left[\|S_{s,t}(\varsigma)(x)\|^n\right]^{2/n} \\
&\leqslant c_{1,n} \int_{[s,t]} \mathbb{E}\left[\|\Sigma_{s,u,t}(x)\|^2\right]\ du + c_{2,n} \int_{[s,t]^2} \mathbb{E}\left[\|D_v\Sigma_{s,u,t}(x)\|^n\right]^{2/n}\ du\ dv
\end{aligned} \tag{5.8}$$

Observe that for any $n \geqslant 2$ we have

$$(\overline{\mathcal{M}})_n \Longrightarrow |||\varsigma(x)|||_n \leqslant \kappa_n\ \left(\|\varsigma(0)\| + \|\nabla\varsigma\|\right)(1 \vee \|x\|)$$

The main objective of this section is to prove the following theorem.

Theorem 5.2. *Assume that* $(M)_{2n/\delta}$ *and* $(T)_{2n/(1-\delta)}$ *are satisfied for some parameter* $n \geqslant 2$ *and some* $\delta \in]0,1[$*. In this situation, we have the uniform estimate*

$$\mathbb{E}\left[\|S_{s,t}(\varsigma)(x)\|^n\right]^{1/n} \leqslant \kappa_{\delta,n}\ |||\varsigma(x)|||_{2n/\delta}\ (1 \vee \|x\|) \tag{5.9}$$

For uniformly bounded diffusion functions $(\varsigma, \sigma, \overline{\sigma})$ *whenever* $(T)_{2n}$ *is met for some* $n \geqslant 2$ *we have*

$$\mathbb{E}\left[\|S_{s,t}(\varsigma)(x)\|^n\right]^{1/n} \leqslant \kappa_n\ \left(\|\varsigma\| + \|\nabla\varsigma\|\right) \tag{5.10}$$

In addition, for constant diffusion functions $(\varsigma, \sigma, \overline{\sigma})$ *whenever* $(T)_2$ *is met, for any* $n \geqslant 2$ *we have the uniform estimate*

$$\mathbb{E}\left[\|S_{s,t}(\varsigma)(x)\|^n\right]^{1/n} \leqslant \kappa_n\ \|\varsigma\| \tag{5.11}$$

The proof of the above theorem, including a more detailed description of the parameters $\kappa_{\delta,n}$ and κ_n is provided below.

Next, we estimate the $\mathbb{L}_n$-norm of the Malliavin differential $D_v\Sigma_{s,u,t}(x)$ in the two cases $(s \leqslant u \leqslant v \leqslant t)$ and $(s \leqslant v \leqslant u \leqslant t)$.

Case $(s \leqslant u \leqslant v \leqslant t)$**:**

Using (5.6) we have

$$\|D_v\Sigma_{s,u,t}(x)\| \leqslant c\ \|\varsigma_u(\overline{X}_{s,u}(x))\|\ \|(D_v\nabla X_{u,t})(\overline{X}_{s,u}(x))\|$$

Using (2.24) and (2.25) this yields the estimate

$$\|D_v\Sigma_{s,u,t}(x)\| \leqslant c_1\ \mathbb{I}_{s,u,t}(x) + c_2\ \mathbb{J}_{s,u,t}(x)$$

with the functions

$$\mathbb{I}_{s,u,t}(x) := \|\nabla\sigma\|\ \|\varsigma_u(\overline{X}_{s,u}(x))\|\ \|(\nabla X_{u,v})(\overline{X}_{s,u}(x))\|\ \|(\nabla X_{v,t})(Z_u^{s,v}(x))\|$$

$$\mathbb{J}_{s,u,t}(x) := \|\sigma_v(Z_u^{s,v}(x))\|\ \|\varsigma_u(\overline{X}_{s,u}(x))\|\ \|(\nabla X_{u,v})(\overline{X}_{s,u}(x))\|\ \|(\nabla^2 X_{v,t})(Z_u^{s,v}(x))\|$$

In the above display, $Z_u^{s,v}(x)$ stands for the interpolating flow defined in (1.13).

- Firstly assume that $\|\varsigma\| \vee \|\sigma\| < \infty$ and $(\mathcal{T})_{2n}$ is satisfied for some parameter $n \geqslant 1$. In this situation, applying proposition 3.2 and proposition 3.3, for any $\epsilon \in]0,1[$ we have the uniform estimates

$$\mathbb{E}\left(\|D_v\Sigma_{s,u,t}(x)\|^n\right)^{1/n} \leqslant \|\varsigma\|\ \chi_{n,\epsilon}(b,\sigma)\ \exp\left(-(1-\epsilon)\lambda_A(2n)(t-u)\right)$$

with the parameter $\chi_{n,\epsilon}(b,\sigma)$ given by

$$\chi_{n,\epsilon}(b,\sigma) := c\ \left[\|\sigma\| \vee \|\nabla\sigma\|\right]\left[1 + \frac{1}{\epsilon}\ \frac{n}{\lambda_A(2n)}\ \chi(b,\sigma)\right] \qquad \text{with } \chi(b,\sigma) \text{ given in (2.6).}$$

- More generally, when $\|\nabla\varsigma\| \vee \|\nabla\sigma\| < \infty$ the functions $\varsigma_t(x)$ and $\sigma_t(x)$ may grow at the most linearly with respect to $\|x\|$. Assume that conditions $(M)_{2n/\delta}$ and condition $(\mathcal{T})_{2n/(1-\delta)}$ are satisfied for some parameters $n \geqslant 1$ and $\delta \in]0,1[$. In this situation, applying Hölder inequality we check that

$$\mathbb{E}\left(\|\mathbb{I}_{s,u,t}(x)\|^n\right)^{1/n} \leqslant c\ \|\nabla\sigma\|\ \mathbb{E}\left(\|\varsigma_u(\overline{X}_{s,u}(x))\|^{n/\delta}\right)^{\delta/n}$$

$$\times\mathbb{E}\left(\|(\nabla X_{u,v})(\overline{X}_{s,u}(x))\|^{2n/(1-\delta)}\right)^{(1-\delta)/(2n)}\ \mathbb{E}\left(\|(\nabla X_{v,t})(Z_u^{s,v}(x))\|^{2n/(1-\delta)}\right)^{(1-\delta)/(2n)}$$

Applying proposition 3.2 we check that

$$\mathbb{E}\left(\|\mathbb{I}_{s,u,t}(x)\|^n\right)^{1/n} \leqslant c_{n,\delta}\ \|\nabla\sigma\|\ |||\varsigma(x)|||_{n/\delta}\ e^{-\lambda_A(2n/(1-\delta))(t-u)}$$

In the same vein, combining proposition 3.2 and proposition 3.3 with the uniform moment estimates (2.11) we check that

$$\mathbb{E}\left(\|\mathbb{J}_{s,u,t}(x)\|^n\right)^{1/n} \leqslant c_{n,\delta}\ \left[\|\sigma(0)\| + \|\nabla\sigma\|\right]\ \frac{1}{\epsilon}\ \frac{\chi(b,\sigma)}{\lambda_A(2n/(1-\delta))}$$

$$\times\ |||\varsigma(x)|||_{2n/\delta}\ \ [1+\|x\|]\ \ e^{-(1-\epsilon)\lambda_A(2n/(1-\delta))(t-u)}$$

We conclude that

$$\mathbb{E}\left(\|D_v\Sigma_{s,u,t}(x)\|^n\right)^{1/n} \leqslant \chi_{n,\delta,\epsilon}(b,\sigma)\ |||\varsigma(x)|||_{2n/\delta}\ [1+\|x\|]\ \ e^{-(1-\epsilon)\lambda_A(2n/(1-\delta))(t-u)}$$

with the parameter

$$\chi_{n,\delta,\epsilon}(b,\sigma) := c_{n,\delta}\ \left[\|\sigma(0)\| + \|\nabla\sigma\|\right]\ \left(1 + \frac{1}{\epsilon}\ \frac{\chi(b,\sigma)}{\lambda_A(2n/(1-\delta))}\right)$$

Case $(s \leqslant v \leqslant u \leqslant t)$**:**

We use (5.7) to check that

$$\|D_v\Sigma_{s,u,t}(x)\| \leqslant \|[D_v\left(\varsigma_u \circ \overline{X}_{s,u}\right)](x)\|\ \|\left(\nabla X_{u,t}\right)\left(\overline{X}_{s,u}(x)\right)\|$$

$$+\|[D_v\overline{X}_{s,u}](x) \otimes \varsigma_u(\overline{X}_{s,u})(x)\|\ \|\left(\nabla^2 X_{u,t}\right)\left(\overline{X}_{s,u}(x)\right)\|$$

On the other hand, using the chain rules (2.24) we have

$$\begin{array}{rcl} D_v\overline{X}_{s,u} & := & \left(D_v\overline{X}_{s,v}\right)\ \left[\left(\nabla\overline{X}_{v,u}\right)\circ\overline{X}_{s,v}\right] \\ D_v\left(\varsigma_u\circ\overline{X}_{s,u}\right) & = & \left(D_v\overline{X}_{s,u}\right)\ \left[\left(\nabla\varsigma_u\right)\circ\overline{X}_{s,u}\right] \end{array}$$

This yields the estimate

$$\|D_v\Sigma_{s,u,t}(x)\| \leqslant c_1\ \|\overline{\sigma}_v(\overline{X}_{s,v}(x))\|\ \|\nabla\varsigma\|\ \|(\nabla\overline{X}_{v,u})(\overline{X}_{s,v}(x))\|\ \|\left(\nabla X_{u,t}\right)\left(\overline{X}_{s,u}(x)\right)\|$$

$$+c_2\ \|\overline{\sigma}_v(\overline{X}_{s,v}(x))\|\ \|\varsigma_u(\overline{X}_{s,u}(x))\|\ \|(\nabla\overline{X}_{v,u})(\overline{X}_{s,v}(x))\|\ \|\left(\nabla^2 X_{u,t}\right)\left(\overline{X}_{s,u}(x)\right)\|$$

- Firstly assume that $\|\varsigma\| \vee \|\overline{\sigma}\| < \infty$ and condition $(T)_{2n}$ is satisfied for some $n \geqslant 1$. In this situation, arguing as above for any $\epsilon \in]0,1[$ we have the uniform estimates

$$\mathbb{E}\left(\|D_v\Sigma_{s,u,t}(x)\|^n\right)^{1/n} \leqslant (\|\varsigma\| + \|\nabla\varsigma\|)\ \overline{\chi}_{n,\epsilon}(b,\sigma)\ \exp\left(-(1-\epsilon)\lambda_{A,\overline{A}}(2n)(t-v)\right)$$

  for some universal constant c and the parameter $\overline{\chi}_{n,\epsilon}(b,\sigma)$ given by

$$\overline{\chi}_{n,\epsilon}(b,\sigma) := c\ \|\overline{\sigma}\|\ \left[1 + \frac{1}{\epsilon}\ \frac{n}{\lambda_{A,\overline{A}}(2n)}\ \chi(b,\sigma)\right] \qquad \text{with } \chi(b,\sigma) \text{ given in (2.6).}$$

- More generally assume that $\|\nabla\varsigma\| \vee \|\nabla\overline{\sigma}\| < \infty$. Also assume that conditions $(M)_{2n/\delta}$ and $(T)_{2n/(1-\delta)}$ are satisfied for some parameters $n \geqslant 1$ and $\delta \in]0,1[$. In this situation, we have

$$\begin{aligned} &\mathbb{E}\left(\|D_v\Sigma_{s,u,t}(x)\|^n\right)^{1/n} \\ &\leqslant \chi_{n,\delta,\epsilon}(b,\sigma,\overline{\sigma})\ \left\|\|\varsigma(x)\|\right\|_{2n/\delta}\ [1+\|x\|]\ \ e^{-(1-\epsilon)\lambda_{A,\overline{A}}(2n/(1-\delta))(t-v)} \end{aligned}$$

  with the parameter

$$\chi_{n,\delta,\epsilon}(b,\sigma,\overline{\sigma}) := c_{n,\delta}\ \left[\|\overline{\sigma}(0)\| + \|\nabla\overline{\sigma}\|\right]\ \left(1 + \frac{1}{\epsilon}\ \frac{\chi(b,\sigma)}{\lambda_{A,\overline{A}}(2n/(1-\delta))}\right)$$

The end of the proof of theorem 5.2 is a direct consequence of the estimates discussed above combined with (5.8) and the diagonal estimates presented in (5.4). ■

5.3 Some extensions

This section is concerned with the two-sided stochastic integral (4.6). Using the gradient formula in (4.7) the Skorohod stochastic integral in (4.6) takes the form

$$\mathbb{S}_{s,t}(f,\Delta\sigma)(x) = \int_s^t\ \Sigma_{s,u,t}(f)(x)\ dW_u$$

with the integrands

$$\Sigma_{s,u,t}(f)(x) := \nabla f(Z_u^{s,t}(x))'\ \Sigma_{s,u,t}(x) \quad \text{and} \quad \Sigma_{s,u,t}(x) := \left[(\nabla X_{u,t})' \circ \overline{X}_{s,u}\right]\ \left[\Delta\sigma_u \circ \overline{X}_{s,u}\right]$$

As in (2.25), using the chain rule properties of Malliavin derivatives we check that

$$D_v^i\Sigma_{s,u,t}(f) = \left(D_v^i\nabla f(Z_u^{s,t})'\right)\ \Sigma_{s,u,t} + \nabla f(Z_u^{s,t})'\ D_v^i\Sigma_{s,u,t}$$

as well as

$$D_v^i\nabla f(Z_u^{s,t})' = \nabla^2 f(Z_u^{s,t})'\ D_v^i Z_u^{s,t}$$

This yields the differential formula

$$D_v^i\Sigma_{s,u,t}(f) = \nabla f(Z_u^{s,t})'\ D_v^i\Sigma_{s,u,t} + \nabla^2 f(Z_u^{s,t})'\ (D_v^i Z_u^{s,t})\ \Sigma_{s,u,t}$$

The Malliavin derivatives $D_v^i\Sigma_{s,u,t}$ are computed using formulae (5.6) and (5.7); thus, it remains to compute the Malliavin derivatives $D_v Z_u^{s,t}$ of the interpolating path.

- When $u \leqslant v$ we have

$$Z_u^{s,t} = (X_{v,t} \circ X_{u,v}) \circ \overline{X}_{s,u} = X_{v,t} \circ Z_u^{s,v}$$

In this situation, as in (2.24) using the chain rule properties of Malliavin derivatives we check that

$$D_v Z_u^{s,t} = D_v Z_u^{s,v} \ ((\nabla X_{v,t}) \circ Z_u^{s,v}) = ((D_v X_{u,v}) \circ \overline{X}_{s,u}) \ ((\nabla X_{v,t}) \circ Z_u^{s,v})$$

By (2.23) we conclude that

$$D_v Z_u^{s,t} = (\sigma_v \circ Z_u^{s,v}) \ ((\nabla X_{v,t}) \circ Z_u^{s,v})$$

- When $v \leqslant u$ we have

$$Z_u^{s,t} = X_{u,t} \circ (\overline{X}_{v,u} \circ \overline{X}_{s,v}) = Z_u^{v,t} \circ \overline{X}_{s,v}$$

In this situation, arguing as above we check that

$$D_v Z_u^{s,t} = D_v \overline{X}_{s,v} \ ((\nabla Z_u^{v,t}) \circ \overline{X}_{s,v}) = D_v \overline{X}_{s,v} \ ((\nabla \overline{X}_{v,u}) \circ \overline{X}_{s,v}) \ ((\nabla X_{u,t}) \circ \overline{X}_{s,u})$$

By (2.23) we conclude that

$$D_v Z_u^{s,t} = (\overline{\sigma}_v \circ \overline{X}_{s,v}) \ ((\nabla \overline{X}_{v,u}) \circ \overline{X}_{s,v}) \ ((\nabla X_{u,t}) \circ \overline{X}_{s,u})$$

6 Some anticipative calculus

For clarity and to avoid unnecessary sophisticated multi-index notation, we only consider one dimensional model. The proof of the results presented in this section in the general case can be reproduced word-for-word for multidimensional models.

To simplify the presentation, we write $\partial^n f$ the derivative of order $n \geqslant 1$ of a smooth function f. We also set $Y_{s,t}(x) := \overline{X}_{s,t}(x)$. We also reduce the analysis to the unit interval. In this context, for any $t \in [0,1]$ we set

$$Y_t := Y_{0,t} \quad \text{and} \quad X^t := X_{t,1} \tag{6.1}$$

6.1 Extended two-sided stochastic integrals

The aim of this section is to extend the two-sided stochastic integration introduced in [43] to Skorohod integrals of the form (4.3), for some time homogeneous function $\varsigma_u = \varsigma$ satisfying (5.1). For any $t \in [0,1]$ we set

$$\Phi(X^t, Y_t(x)) := \partial X^t(Y_t(x)) \ \varsigma(Y_t(x)) \tag{6.2}$$

In this notation the limiting integral in (4.3) takes formally the following form

$$S_{0,1}(\varsigma)(x) := \int_0^1 \ \Phi(X^t, Y_t(x)) \ dW_t$$

The existence of this two-sided stochastic integral is discussed below in (6.4).

To simplify the presentation, we fix the state variable x and we write Y_t and $\Phi(X^t, Y_t)$ instead of $Y_t(x)$ and $\Phi(X^t, Y_t(x))$. Next technical lemma provided a more explicit description of the Malliavin derivatives of the processes $\Phi(X^t, Y_t)$.

Lemma 6.1. *For any $s < t$ we have*

$$D_s\,\Phi(X^t, Y_t) = \left[\partial((\partial X^t)\circ Y_{s,t})(Y_s)\ (\varsigma\circ Y_{s,t})(Y_s) + ((\partial X^t)\circ Y_{s,t})(Y_s)\times \partial(\varsigma\circ Y_{s,t})(Y_s)\right]\ \overline{\sigma}(Y_s)$$

In addition, we have

$$D_t\,\Phi(X^s, Y_s) = \left[\partial((\partial X^t)\circ X_{s,t})(Y_s)\ (\sigma\circ X_{s,t})(Y_s) + \partial(X^t\circ X_{s,t})(Y_s)\ \partial\sigma(X_{s,t}(Y_s))\ \right]\varsigma(Y_s)$$

Proof. Using the chain rules properties, for any $s < t$ we have

$$D_s\,\Phi(X^t, Y_t) = D_s\left((\partial X^t)(Y_{s,t}(Y_s))\ (\varsigma\circ Y_{s,t})(Y_s)\right)$$

$$= D_s\left((\partial X^t)\circ Y_{s,t})(Y_s)\right)\ (\varsigma\circ Y_{s,t})(Y_s) + (\partial X^t)\circ Y_{s,t})(Y_s)\ D_s(\varsigma\circ Y_{s,t})(Y_s)$$

The end of the proof of the first assertion comes from the fact that

$$D_s\left((\partial X^t)\circ Y_{s,t})(Y_s)\right) = \partial((\partial X^t)\circ Y_{s,t})(Y_s)\ D_sY_s \quad \text{with} \quad D_sY_s = \overline{\sigma}(Y_s)$$

In the same vein, we have

$$D_s(\varsigma\circ Y_{s,t})(Y_s) = \partial(\varsigma\circ Y_{s,t})(Y_s)\ \overline{\sigma}(Y_s)$$

We also have that

$$D_t\,\Phi(X^s, Y_s) = D_t\left((\partial X^s)(Y_s)\ \varsigma(Y_s)\right)$$

$$= D_t\left(\partial(X^t\circ X_{s,t})(Y_s)\right)\ \varsigma(Y_s) = D_t\left(\left((\partial X^t)\circ X_{s,t}\right)(Y_s)\ (\partial X_{s,t})(Y_s)\right)\ \varsigma(Y_s)$$

The last assertion comes from the fact that

$$D_t\ \left(\left((\partial X^t)\circ X_{s,t}\right)(Y_s)\ (\partial X_{s,t})(Y_s)\right)$$

$$= D_t\left((\partial X^t)\circ X_{s,t}\right)(Y_s)\ (\partial X_{s,t})(Y_s) + \left((\partial X^t)\circ X_{s,t}\right)(Y_s)\ D_t(\partial X_{s,t})(Y_s)$$

The r.h.s. term in the above display can be rewritten as follows

$$D_t(\partial X_{s,t})(Y_s) = \partial\sigma(X_{s,t}(Y_s))\ (\partial X_{s,t})(Y_s)$$

$$\Longrightarrow ((\partial X^t)\circ X_{s,t})(Y_s)\ D_t(\partial X_{s,t})(Y_s) = \partial(X^t\circ X_{s,t})(Y_s)\ \partial\sigma(X_{s,t}(Y_s))$$

In the same vein, we have

$$D_t\left((\partial X^t)\circ X_{s,t}\right)(Y_s) = ((\partial^2 X^t)\circ X_{s,t})(Y_s)\ D_tX_{s,t}(Y_s) = ((\partial^2 X^t)\circ X_{s,t})(Y_s)\ \sigma(X_{s,t}(Y_s))$$

$$\Longrightarrow D_t\left((\partial X^t)\circ X_{s,t}\right)(Y_s)\ (\partial X_{s,t})(Y_s) = \partial((\partial X^t)\circ X_{s,t})(Y_s)\ \sigma(X_{s,t}(Y_s))$$

This ends the proof of the second assertion. The proof of the lemma is now completed. ■

From the above lemma, we also check that all the n-absolute moments of the Malliavin derivatives $D_s\,\Phi(X^t, Y_t)$ are finite with at most quadratic growth w.r.t. the initial values.

Next proposition extends proposition 3.3 in [43] to stochastic processes of the form (6.2).

Proposition 6.2. *Let* $[0,1]_h$ *be any refining sequence of partitions of the unit interval. For any* $h > 0$ *we define*

$$S^h(\Phi) := \sum_{t \in [0,1]_h} \Phi(X^{t+h}, Y_t)\ (W_{t+h} - W_t)$$

Then $S^h(\Phi)$ *is a Cauchy sequence in* $\mathbb{L}_2(\Omega)$*. In addition, for any decreasing sequence of time steps* $h_1 > h_2$ *we have the formula*

$$\lim_{h_1 \to 0} \mathbb{E}\left(S^{h_1}(\Phi)\ S^{h_2}(\Phi)\right) = \mathbb{E}\left(\int_0^1 \Phi(X^t, Y_t)^2\ dt + \int_{[0,1]^2} D_s\, \Phi(X^t, Y_t)\ D_t\, \Phi(X^s, Y_s)\ ds\ dt\right) \tag{6.3}$$

Before entering into the details of the proof of the proposition, we give a couple of comments. The hypothesis that $[0,1]_h$ is a refining sequence indexed by h is not essential but it simplifies the proof of the proposition, see for instance lemma 3.1.1 in [37]. Arguing as in the proof of theorem 3.3 and theorem 7.1 in [43] the above proposition ensures that the two-sided integral defined by the $\mathbb{L}_2(\Omega)$-limit coincides with the two-sided stochastic integral of the process $\Phi(X^t, Y_t)$ over the unit interval; that is, we have that

$$\int_0^1 \Phi(X^t, Y_t)\ dW_t := \mathbb{L}_2 - \lim_{h \to 0} \sum_{t \in [0,1]_h} \Phi(X^{t+h}, Y_t)\ (W_{t+h} - W_t) \tag{6.4}$$

In this context, proposition 6.2 can be interpreted as a version of the isometry property (2.26) for the generalized two-sided integral defined above.

Proof of proposition 6.2:

We fix $h_1 > h_2$ and we assume that $[0,1]_{h_2}$ is a refinement of $\in [0,1]_{h_1}$. For any $(s,t) \in ([0,1]_{h_1} \times [0,1]_{h_2})$ we also set

$$\Pi^{h_1,h_2}_{s,t} := \Phi(X^{s+h_1}, Y_s)\ \Phi(X^{t+h_2}, Y_t)\ (W_{s+h_1} - W_s)\ (W_{t+h_2} - W_t)$$

With a slight abuse of notation we set

$$\Delta W_s := (W_{s+h_1} - W_s) \quad \text{and} \quad \Delta W_t := (W_{t+h_2} - W_t)$$

- For any overlapping pair $s < t < t + h_2 < s + h_1$ using the decomposition

$$\Delta W_s = (W_{s+h_1} - W_{t+h_2}) + \Delta W_t + (W_t - W_s)$$

we have

$$\mathbb{E}\left(\ \Phi(X^{s+h_1}, Y_s)\ \Phi(X^{t+h_2}, Y_t)\ \Delta W_t\ \Delta W_s \mid \mathcal{W}_t \vee \mathcal{W}^{t+h_2}\right) = \Phi(X^{s+h_1}, Y_s)\ \Phi(X^{t+h_2}, Y_t)\ h_2$$

It follows from the continuity properties of the processes that

$$\mathbb{E}\left(\sum_{s<t<t+h_2<s+h_1} \Pi^{h_1,h_2}_{s,t}\right) \longrightarrow_{h_1 \to 0} \mathbb{E}\left(\int_0^1 \Phi(X^t, Y_t)^2\ \ dt\right)$$

- When $s + h_1 < t$ we have

$$\begin{aligned} \partial X^{s+h_1} &= \partial(X^t \circ X_{s+h_1,t}) \\ &= \partial(X^{t+h_2} \circ X_{s+h_1,t}) + \left(\left(\partial X^t - \partial X^{t+h_2}\right) \circ X_{s+h_1,t}\right)\ \times\ \partial X_{s+h_1,t} \end{aligned}$$

On the other hand, we have the decomposition

$$\begin{aligned}
\partial X^t - \partial X^{t+h_2} &= \Big((\partial X^{t+h_2}) \circ X_{t,t+h_2}\Big) \times \partial X_{t,t+h_2} - \partial X^{t+h_2} \\
&= \Big((\partial X^{t+h_2}) \circ (I + \Delta X_t) - \partial X^{t+h_2}\Big) + \partial X^{t+h_2} \times \Delta X'_t \\
&\qquad + \Big((\partial X^{t+h_2}) \circ (I + \Delta X_t) - \partial X^{t+h_2}\Big) \times \Delta X'_t
\end{aligned}$$

with the increment functions

$$\Delta X'_t := \partial X_{t,t+h_2} - 1 \quad \text{and} \quad \Delta X_t := X_{t,t+h_2} - I$$

With a slight abuse of notation, we shall denote by $\mathrm{O}(h^p)$ some possible random variable with any n-absolute moment of order h^p, for some $p > 0$ with $0 < h < 1$. In this notation, we have

$$\begin{aligned}
\Delta X'_t(x) &= \int_t^{t+h_2} \partial\sigma(X_{t,u}(x))\ \partial X_{t,u}(x)\ dW_u + \mathrm{O}(h_2) = \mathrm{O}(h_2^{1/2}) \\
\Delta X_t(x) &= \int_t^{t+h_2} \sigma(X_{t,u}(x))\ dW_u + \mathrm{O}(h_2) = \mathrm{O}(h_2^{1/2})
\end{aligned}$$

Given a smooth function θ we set

$$\partial^n\theta(x, y) := \int_0^1 \frac{(1-\epsilon)^{n-1}}{(n-1)!}\ \theta''(x + \epsilon y)\ d\epsilon$$

In this notation, we have the first and second order decompositions

$$\begin{aligned}
&\Big((\partial X^{t+h_2}) \circ (I + \Delta X_t) - \partial X^{t+h_2}\Big)(x) \\
&= (\partial^2 X^{t+h_2})(x, \Delta X_t(x))\ \Delta X_t(x) = (\partial^2 X^{t+h_2})(x)\ \Delta X_t(x) + (\partial^3 X^{t+h_2})(x, \Delta X_t(x))\ \Delta X_t(x)^2
\end{aligned}$$

This implies that

$$\begin{aligned}
&(\partial X^t - \partial X^{t+h_2})(x) \\
&= (\partial^2 X^{t+h_2})(x)\ \Delta X_t(x) + \partial X^{t+h_2}(x) \times \Delta X'_t(x) \\
&\quad + (\partial^3 X^{t+h_2})(x, \Delta X_t(x))\ \Delta X_t(x)^2 + (\partial^2 X^{t+h_2})(x, \Delta X_t(x))\ \Delta X_t(x) \times \Delta X'_t(x)
\end{aligned}$$

from which we conclude that

$$\begin{aligned}
&\partial X^{s+h_1} = \partial(X^{t+h_2} \circ X_{s+h_1,t}) \\
&+ \Big[\partial((\partial X^{t+h_2}) \circ X_{s+h_1,t})\ \times\ ((\Delta X_t) \circ X_{s+h_1,t}) + \partial(X^{t+h_2} \circ X_{s+h_1,t})\ \times\ ((\Delta X'_t) \circ X_{s+h_1,t})\Big] + \mathrm{O}(h_2)
\end{aligned}$$

This yields the first order decomposition

$$\begin{aligned}
&\Phi(X^{s+h_1}, Y_s) \\
&= \psi^0_{s,t}(Y_s) + \psi^1_{s,t}(Y_s)\ ((\Delta X_t) \circ X_{s+h_1,t})(Y_s) + \psi^2_{s,t}(Y_s)\ ((\Delta X'_t) \circ X_{s+h_1,t})(Y_s) + \mathrm{O}(h_2)
\end{aligned}$$

with the functions

$$\psi^0_{s,t}(Y_s) := \partial(X^{t+h_2} \circ X_{s+h_1,t})(Y_s)\ \varsigma(Y_s)$$

$$\psi^1_{s,t}(Y_s) := \partial((\partial X^{t+h_2}) \circ X_{s+h_1,t})(Y_s)\ \varsigma(Y_s) \quad \text{and} \quad \psi^2_{s,t}(Y_s) := \partial(X^{t+h_2} \circ X_{s+h_1,t})(Y_s)\ \varsigma(Y_s)$$

Notice that none of the functions but the increment functions (ΔX_t) and $(\Delta X'_t)$ depend on $\mathcal{W}_{t,t+h_2}$, nor on $\mathcal{W}_{s,s+h_1}$.

In the reverse angle, we have

$$\begin{aligned}&(\partial X^{t+h_2}) \circ Y_t\\ &= (\partial X^{t+h_2}) \circ (Y_{s+h_1,t} \circ Y_s) + \left[((\partial X^{t+h_2}) \circ Y_{s+h_1,t}) \circ (I + \Delta Y_s) - ((\partial X^{t+h_2}) \circ Y_{s+h_1,t})\right] \circ Y_s\end{aligned}$$

with

$$\Delta Y_s := (Y_{s,s+h_1} - I) \Longrightarrow Y_{s+h_1} = (I + \Delta Y_s) \circ Y_s$$

Arguing as above, we have

$$\begin{aligned}&\left[((\partial X^{t+h_2}) \circ Y_{s+h_1,t}) \circ (y + \Delta Y_s(y)) - ((\partial X^{t+h_2}) \circ Y_{s+h_1,t})(y)\right]\\ &= \partial((\partial X^{t+h_2}) \circ Y_{s+h_1,t})(y)\ \Delta Y_s(y) + \partial^2((\partial X^{t+h_2}) \circ Y_{s+h_1,t})(y, \Delta Y_s(y))\ (\Delta Y_s(y))^2\end{aligned}$$

We conclude that

$$\begin{aligned}&(\partial X^{t+h_2}) \circ Y_t\\ &= (\partial X^{t+h_2}) \circ (Y_{s+h_1,t} \circ Y_s) + \partial((\partial X^{t+h_2}) \circ Y_{s+h_1,t})(Y_s)\ ((\Delta Y_s) \circ Y_s) + \mathrm{O}(h_1)\end{aligned}$$

In the same vein, we have

$$\varsigma \circ Y_t = (\varsigma \circ Y_{s+h_1,t} \circ Y_s) + \partial(\varsigma \circ Y_{s+h_1,t})(Y_s)\ ((\Delta Y_s) \circ Y_s) + \mathrm{O}(h_1)$$

Multiplying these terms, we check that

$$\Phi(X^{t+h_2}, Y_t) = \Psi^0_{s,t}(Y_s) + \Psi^1_{s,t}(Y_s)\ ((\Delta Y_s) \circ Y_s) + \mathrm{O}(h_1)$$

with the functions

$$\begin{aligned}\Psi^0_{s,t}(Y_s) &:= ((\partial X^{t+h_2}) \circ Y_{s+h_1,t})(Y_s) \times (\varsigma \circ Y_{s+h_1,t})(Y_s)\\ \Psi^1_{s,t}(Y_s) &:= \big[\partial((\partial X^{t+h_2}) \circ Y_{s+h_1,t})(Y_s) \times (\varsigma \circ Y_{s+h_1,t})(Y_s)\\ &\qquad + ((\partial X^{t+h_2}) \circ Y_{s+h_1,t})(Y_s) \times \partial(\varsigma \circ Y_{s+h_1,t})(Y_s)\big]\end{aligned}$$

None of the functions but the increment ΔY_s depend on $\mathcal{W}_{s,s+h_1}$, nor on $\mathcal{W}_{t,t+h_2}$.

Recall that the functions $\Phi(X^{t+h_2}, Y_t)$ and $\psi^0_{s,t}(Y_s)$ don't depend on ΔW_t. In addition, the functions $\Phi(X^{s+h_1}, Y_s)$ and $\Psi^0_{s,t}(Y_s)$ don't depend on ΔW_s. This yields the formula

$$\begin{aligned}&\mathbb{E}\left(\Phi(X^{s+h_1}, Y_s)\ \Phi(X^{t+h_2}, Y_t)\ \ \Delta W_s\ \Delta W_t\right)\\ &= \mathbb{E}\left(\left[\Phi(X^{s+h_1}, Y_s) - \psi^0_{s,t}(Y_s)\right]\ \left[\Phi(X^{t+h_2}, Y_t) - \Psi^0_{s,t}(Y_s)\right]\ \Delta W_s\ \Delta W_t\right)\\ &= \mathbb{E}\left(\Psi^1_{s,t}(Y_s)\ \psi^1_{s,t}(Y_s)\ \ [((\Delta Y_s) \circ Y_s)\ \ \Delta W_s]\ \ [((\Delta X_t) \circ X_{s+h_1,t})(Y_s)\ \Delta W_t]\right)\\ &+ \mathbb{E}\left(\Psi^1_{s,t}(Y_s)\ \psi^2_{s,t}(Y_s)\ \ [((\Delta Y_s) \circ Y_s)\ \Delta W_s]\ \ [((\Delta X'_t) \circ X_{s+h_1,t})(Y_s)\ \Delta W_t]\right) + \mathrm{O}\left(h_1^{2+1/2}\right)\end{aligned}$$

To take the final step, observe that

$$
\begin{aligned}
&\mathbb{E}\left(\Delta Y_s(y)\ \Delta W_s\right)\\
&= \mathbb{E}\left(\int_s^{s+h_1}\ \overline{b}(Y_{s,u}(y))\ (W_{s+h_1}-W_u)\ du\right) + \mathbb{E}\left(\int_s^{s+h_1}\ \overline{\sigma}(Y_{s,u}(y))\ du\right)\\
&= \mathbb{E}\left(\int_s^{s+h_1}\ \overline{\sigma}(Y_{s,u}(y))\ du\right) + \mathrm{O}\left(h_1^{1+1/2}\right)
\end{aligned}
$$

In the same vein, we have

$$
\begin{aligned}
&\mathbb{E}\left((\Delta X_t)(X_{s+h_1,t}(y))\ \Delta W_t \mid \mathcal{W}_{s+h_1,t}\right)\\
&= \mathbb{E}\left(\int_t^{t+h_2}\ \sigma(X_{s+h_1,u}(y))\ du \mid \mathcal{W}_{s+h_1,t}\right) + \mathrm{O}\left(h_2^{1+1/2}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left((\Delta X_t')(X_{s+h_1,t}(y))\ \Delta W_t \mid \mathcal{W}_{s+h_1,t}\right)\\
&= \mathbb{E}\left(\int_t^{t+h_2}\ \partial\sigma(X_{s+h_1,u}(y))\ (\partial X_{t,u})(X_{s+h_1,t}(y))\ du\right) + \mathrm{O}\left(h_2^{1+1/2}\right)\\
&= \mathbb{E}\left(\int_t^{t+h_2}\ \partial\left(\sigma\circ X_{t,u}\right)\left(X_{s+h_1,t}(y)\right)\ du\right) + \mathrm{O}\left(h_2^{1+1/2}\right)
\end{aligned}
$$

This shows that

$$
\begin{aligned}
&h_1^{-1}\ h_2^{-1}\mathbb{E}\left(\Phi(X^{s+h_1},Y_s)\ \Phi(X^{t+h_2},Y_t)\ \ \Delta W_s\ \Delta W_t\right)\\
&= \mathbb{E}\left(\Psi^1_{s,t}(Y_s)\ \psi^1_{s,t}(Y_s)\ h_1^{-1}\left[\int_s^{s+h_1}\ \overline{\sigma}(Y_u)\ du\right]\ h_2^{-1}\left[\int_t^{t+h_2}\ \sigma(X_{s+h_1,u}(Y_s))\ du\right]\right)\\
&+ \mathbb{E}\left(\Psi^1_{s,t}(Y_s)\ \psi^2_{s,t}(Y_s)\ h_1^{-1}\left[\int_s^{s+h_1}\ \overline{\sigma}(Y_u)\ du\right]\ h_2^{-1}\left[\int_t^{t+h_2}\ \partial\left(\sigma\circ X_{t,u}\right)\left(X_{s+h_1,t}(Y_s)\right)\ du\right]\right) + \mathrm{O}\left(h_1^{1/2}\right)
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\lim_{h_1\to 0}\mathbb{E}\left(\sum_{s+h_1<t}\ \Pi^{h_1,h_2}_{s,t}\right)\\
&= \mathbb{E}\left(\int_0^1\int_0^t\left[\partial((\partial X^t)\circ Y_{s,t})(Y_s)\ (\varsigma\circ Y_{s,t})(Y_s) + ((\partial X^t)\circ Y_{s,t})(Y_s)\times\partial(\varsigma\circ Y_{s,t})(Y_s)\right]\ \overline{\sigma}(Y_s)\right.\\
&\qquad\left.\times\left[\partial((\partial X^t)\circ X_{s,t})(Y_s)\ \sigma(X_{s,t}(Y_s)) + \partial(X^t\circ X_{s,t})(Y_s)\ \partial\sigma(X_{s,t}(Y_s))\right]\varsigma(Y_s)\ \ ds\ dt\right)
\end{aligned}
$$

We end the proof of (6.3) using lemma 6.1 and symmetry arguments. This ends the proof of the proposition. ■

6.2 Generalized backward Itô-Ventzell formula

This section is mainly concerned with the proof of theorem 1.1. Before entering into the details of the proof we discuss how it applies to the process (X^t, Y_t) introduced in (6.1).

Consider the random fields

$$\begin{aligned} F_t(x) &:= X^t(x) \Longrightarrow \partial F_t = \partial X^t \quad \text{and} \quad \partial^2 F_t = \partial^2 X^t \\ G_t(x) &:= \partial X^t(x)\ b(x) + \frac{1}{2}\ \partial^2 X^t(x)\ a(x) \quad \text{and} \quad H_t(x) := \partial X^t(x)\ \sigma(x) \end{aligned} \tag{6.5}$$

In this notation, the backward random field formula (4.1) with $t \in [0,1]$ takes the form

$$F_t(x) := F_1(x) + \int_t^1 G_s(x)\ ds + \int_t^1 H_s(x)\ dW_s \quad \text{with} \quad F_1(x) = x \tag{6.6}$$

We fix some given $Y_0 = y \in \mathbb{R}$ and we write Y_t instead of $Y_t(y)$ and set

$$(A_u, B_u, \Sigma_u) := \left(\overline{a}(Y_u), \overline{b}(Y_u), \overline{\sigma}(Y_u)\right)$$

In this notation, we have

$$Y_t = y + \int_0^t B_u\ du + \int_0^t \Sigma_u\ dW_u \tag{6.7}$$

Observe that B_u, Σ_u as well as the Malliavin derivatives $D_v\Sigma_u = \partial\overline{\sigma}(Y_u)\ D_vY_u$ have moments of any order. Consider the processes

$$\begin{aligned} U_t &:= \partial F_t(Y_t)\ B_t + \frac{1}{2}\ \partial^2 F_t(Y_t)\ A_t - G_t(Y_t) = \partial X^t(Y_t)\ (\overline{b} - b)(Y_t) + \frac{1}{2}\ \partial^2 X^t(Y_t)\ (\overline{a} - a)(Y_t) \\ V_t &:= \partial F_t(Y_t)\ \Sigma_t - H_t(Y_t) = \partial X^t(Y_t)\ (\overline{\sigma} - \sigma)(Y_t) \quad \text{with} \quad A_t := \Sigma_t^2 \end{aligned}$$

In this notation, up to a change of sign and replacing x by Y_0 in (1.10) the stochastic interpolation formula stated in theorem 1.2 on the unit interval takes the following form

$$F_1(Y_1) - F_0(Y_0) = \int_0^1 U_s\ ds + \int_0^1 V_s\ dW_s$$

More generally, suppose we are given a forward real valued continuous semi-martingale Y_t of the form (6.7) for some $\mathcal{W}_{0,t}$-adapted functions B_t and Σ_t, and a backward random field models of the form (6.6) for some $\mathcal{W}_{t,1}$-adapted functions $F_t(x), G_t(x), H_t(x)$.

We consider the following conditions:

$(H_1)'$: The functions $F_t(x)$, $G_t(x)$ and $H_t(x)$ as well as the differentials $\partial H_t(x)$ and $\partial^2 F_t(x)$ are continuous w.r.t. (t,x) for any given $\omega \in \Omega$. In addition, for any $n \geqslant 1$ we have

$$\begin{aligned} \sup_{|y|\leqslant n} \left(|F_t(Y_t + y)| \vee |H_t(Y_t + y)| \vee |G_t(Y_t + y)|\right) &\leqslant g_n(t) \\ \sup_{|y|\leqslant n} \left(|\partial H_t(Y_t + y)| \vee |\partial F_t(Y_t + y)| \vee |\partial^2 F_t(Y_t + y)|\right) &\leqslant g_n(t) \quad \text{with} \quad \mathbb{E}\left(\int_0^1 g_n^4(t)\, dt\right) < \infty \end{aligned} \tag{6.8}$$

$(H_2)'$: *The Malliavin derivatives* $D_s\partial F_t(x)$ *and* $D_sH_t(x)$ *are continuous w.r.t.* x *and* (s,t) *for any given* $\omega\in\Omega$. *In addition, for any* $n\geqslant 1$ *we have*

$$\begin{array}{rcl}\sup_{|y|\leqslant n}\left(|(D_sF_t)(Y_t+y)|\vee|(D_sH_t)(Y_t+y)|\right) & \leqslant & h_n(s,t)\\ \sup_{|y|\leqslant n}\left(|(D_s\partial F_t)(Y_t+y)|\right) & \leqslant & h_n(s,t)\quad\text{with}\quad \mathbb{E}\left(\int_{[0,1]^2}h_n^4(s,t)\,dsdt\right)<\infty\end{array}\tag{6.9}$$

(H_3): *The random processes* B_u,Σ_u *as well as* $D_v\Sigma_u$ *are continuous w.r.t. the time parameter and they have moments of any order.*

The next theorem is a slight extension of theorem 1.1 applied to the semi-martingale and the random fields models discussed in (6.7) and (6.5).

Theorem 6.3. *Consider a backward random field models of the form (6.6) for some functions* $F_t(x),G_t(x),H_t(x)$ *satisfying* $(H_1)'$ *and* $(H_2)'$. *Also let* Y_t *be a continuous semi-martingale of the form (6.7) functions* B_t *and* Σ_t *satisfying* (H_3). *In this situation, for any* $t\in[0,1]$ *we have the generalized backward Itô-Ventzell formula*

$$\begin{aligned}&F_t(Y_t)-F_0(Y_0)\\&=\int_0^t\left(\partial F_s(Y_s)\;B_s+\frac{1}{2}\;\partial^2F_s(Y_s)\;A_s-G_s(Y_s)\right)\;ds+\int_0^t\;\left(\partial F_s(Y_s)\;\Sigma_s-H_s(Y_s)\right)\;dW_s\end{aligned}\tag{6.10}$$

The r.h.s. term in the above display is understood as a Skorohod integral.

Proof: We use the same approximation technique as in [12, 41] and [42] (see also the proof of theorem 3.2.11 in [37]). Consider a mollifier type approximation of the identify given for any $\epsilon>0$ by the function

$$\varphi_\epsilon(x):=\varphi(x/\epsilon)/\epsilon\quad\text{for some smooth compactly supported function }\varphi\text{ s.t. }\int_{-\infty}^{\infty}\varphi(x)dx=1.$$

For any x, applying the Itô-type change rule formula stated in proposition 8.2 in [38] to the product function

$$\Gamma(X^t(x),\varphi_\epsilon(Y_t-x)):=X^t(x)\;\varphi_\epsilon(Y_t-x)$$

we check that

$$(F_t(x)\;\varphi_\epsilon(Y_t-x))-(F_0(x)\;\varphi_\epsilon(Y_0-x))=\int_0^t\;u_s^\epsilon(x)\;ds+\int_0^t\;v_s^\epsilon(x)\;dW_s\tag{6.11}$$

with

$$\begin{array}{rcl}u_s^\epsilon(x)&:=&F_s(x)\;\partial\varphi_\epsilon(Y_s-x)\;B_s+\frac{1}{2}\;F_s(x)\;\partial^2\varphi_\epsilon(Y_s-x)\;A_s-\varphi_\epsilon(Y_s-x)\;G_s(x)\\ v_s^\epsilon(x)&:=&F_s(x)\;\partial\varphi_\epsilon(Y_s-x)\;\Sigma_s-\varphi_\epsilon(Y_s-x)\;H_s(x)\end{array}$$

The stochastic integral in the r.h.s. of (6.11) can be interpreted as a two-sided stochastic integral. Recalling that

$$D_t\varphi_\epsilon(Y_s-x)=D_tY_s\;\partial\varphi_\epsilon(Y_s-x)$$

we check that

$$\begin{array}{rcl}D_tv_s^\epsilon(x)&=&D_tF_s(x)\;\partial\varphi_\epsilon(Y_s-x)\;\Sigma_s+F_s(x)\;D_tY_s\;\partial^2\varphi_\epsilon(Y_s-x)\;\Sigma_s\\&&+F_s(x)\;\partial\varphi_\epsilon(Y_s-x)\;D_t\Sigma_s-D_tY_s\;\partial\varphi_\epsilon(Y_s-x)\;H_s(x)-\varphi_\epsilon(Y_s-x)\;D_tH_s(x)\end{array}$$

Condition (H_3) ensures that the processes Y_t and D_tY_s have moments of any order. In addition, under the regularity conditions $(H_1)'$ and $(H_2)'$ we check that

$$\int \mathbb{E}\left(\int_0^t u_s^\epsilon(x)^2\ ds\right)\ dx < \infty \quad \text{and} \quad \int \mathbb{E}\left(\left[\int_0^t v_s^\epsilon(x)\ dW_s\right]^2\right)\ dx < \infty$$

Applying the Fubini theorem for Skorohod and measure theory integrals (see for instance [34, 37, 44] and the work by Leon [35]) we check that

$$F_t^\epsilon(Y_t) := \int\ F_t(x)\ \varphi_\epsilon(Y_t - x)\ dx = \int\ F_0(x)\ \varphi_\epsilon(Y_0 - x)\ dx + \int_0^t\ U_s^\epsilon\ ds = \int_0^t\ V_s^\epsilon\ dW_s$$

with

$$U_s^\epsilon := \int u_s^\epsilon(x)\ dx \quad \text{and} \quad V_s^\epsilon := \int v_s^\epsilon(x)\ dx$$

Integrating by parts where derivatives of φ_ϵ appear we check that

$$\begin{aligned} U_s^\epsilon &:= \int\ \left(\partial F_s(x)\ B_s + \frac{1}{2}\ \partial^2 F_s(x)\ A_s - G_s(x)\right)\ \varphi_\epsilon(Y_s - x)\ dx \\ V_s^\epsilon &:= \int\ (\partial F_s(x)\ \Sigma_s -\ H_s(x))\ \varphi_\epsilon(Y_s - x)\ dx \end{aligned}$$

From the a.s. continuity of $F_t(x)$ in x for each $t \geqslant 0$, we have

$$F_t^\epsilon(Y_t) - F_t(Y_t) = \int\ (F_t(Y_t - \epsilon\ x) - F_t(Y_t))\ \varphi(x)\ dx\ \longrightarrow_{\epsilon\to 0}\ 0$$

The functions $\partial F_t(x)$, $\partial^2 F_t(x)$ and $G_t(x)$ are almost surely continuous w.r.t. x and uniformly locally bounded. In addition, the random variables A_t and B_t are integrable at any order. Moreover, under $(H_1)'$ there exists some parameter $n \geqslant 0$ depending on the support of φ such that for any $\epsilon > 0$ we have the estimate

$$\begin{aligned} |U_s^\epsilon| &\leqslant \sup_{|y|\leqslant n} |\partial F_s(Y_s + y)|\ |B_s| + \frac{1}{2}\ \sup_{|y|\leqslant n} |\partial^2 F_s(Y_s + y)|\ |A_s| + \sup_{|y|\leqslant n} |G_s(Y_s + y)| \\ &\leqslant g_n(t)\ (1 + |A_s| + |B_s|) \end{aligned}$$

Thus, by the dominated convergence theorem on $(\Omega\times[0,1])$ equipped with the measure $(\mathbb{P}(d\omega)\otimes dt)$ we have

$$\int_0^t\ U_s^\epsilon\ ds \longrightarrow_{\epsilon\to 0}\ \int_0^t\ U_s\ ds \quad \text{as well as} \quad F_t^\epsilon(Y_t) \longrightarrow_{\epsilon\to 0} F_t(Y_t)$$

It remains to check that

$$\mathbb{E}(\int_0^t (V_s^\epsilon - V_s)^2\ ds) + \mathbb{E}\left(\int_{[0,t]^2} (D_rV_s^\epsilon - D_rV_s)\ (D_sV_r^\epsilon - D_sV_r)\ dr\ ds\right) \longrightarrow_{\epsilon\to 0}\ 0 \qquad (6.12)$$

Observe that

$$\int_0^t (V_s^\epsilon - V_s)^2\ ds$$

$$\leqslant 2\ \int_0^t \int\ (\partial F_s(x) - \partial F_s(Y_s))^2\ \Sigma_s^2\ \varphi_\epsilon(Y_s - x)\ dx\ ds + 2\ \int_0^t\ (H_s(x) - H_s(Y_s))^2\ \varphi_\epsilon(Y_s - x)\ dx\ ds$$

Using the chain rule property we have

$$D_tV_s^\epsilon$$

$$:= \int D_t\left(\partial F_s(x)\ \Sigma_s - H_s(x)\right)\ \varphi_\epsilon(Y_s-x)\ dx + \int\ \left(\partial F_s(x)\ \Sigma_s - H_s(x)\right)\ D_t\varphi_\epsilon(Y_s-x)\ dx$$

Integrating by parts, we check that

$$D_tV_s^\epsilon = \int\ \left[D_t\left(\partial F_s(x)\ \Sigma_s - H_s(x)\right)\ + \left(\partial^2 F_s(x)\ \Sigma_s - \partial H_s(x)\right)\ D_tY_s\right]\ \varphi_\epsilon(Y_s-x)\ dx$$

Observe that

$$\begin{aligned} & D_t\left(\partial F_s(x)\ \Sigma_s - H_s(x)\right)\ + \left(\partial^2 F_s(x)\ \Sigma_s - \partial H_s(x)\right)\ D_tY_s \\ &= \left((D_t\partial F_s)(x) + \partial^2 F_s(x)\ D_tY_s\right)\ \Sigma_s + \partial F_s(x)\ D_t\Sigma_s - \left((D_tH_s)(x) + \partial H_s(x)\ D_tY_s\right) \end{aligned}$$

On the other hand, we have

$$\begin{aligned} D_tV_s &= D_t(\partial F_s(Y_s))\ \Sigma_s + \partial F_s(Y_s)\ D_t\Sigma_s - D_t(H_s(Y_s)) \\ &= \left((D_t\partial F_s)(Y_s) + \partial^2 F_s(Y_s)\ D_tY_s\right)\Sigma_s + \partial F_s(Y_s)\ D_t\Sigma_s - \left((D_tH_s)(Y_s) + \partial H_s(Y_s)\ D_tY_s\right) \end{aligned}$$

Arguing as above, we have the estimate

$$\begin{aligned} & \mathbb{E}\left(\int_{[0,1]^2} (D_rV_s^\epsilon - D_rV_s)\ (D_sV_r^\epsilon - D_sV_r)\ dr\ ds\right) \\ & \leqslant 2\ \mathbb{E}\left(\int_{[0,1]^2} (D_tV_s^\epsilon - D_tV_s)^2\ ds\ dt\right) \leqslant 2^4 \sum_{1\leqslant i\leqslant 5} J_i(\epsilon) \end{aligned}$$

In the above display, $J_i(\epsilon)$ stands for the sequences

$$\begin{aligned} J_1(\epsilon) &:= \mathbb{E}\left(\int_{[0,1]^2\times\mathbb{R}} \left(\partial F_s(x) - \partial F_s(Y_s)\right)^2\ \left(D_t\Sigma_s\right)^2\ \varphi_\epsilon(Y_s-x)\ ds\ dt\ dx\right) \\ J_2(\epsilon) &:= \mathbb{E}\left(\int_{[0,1]^2\times\mathbb{R}} \left(\partial H_s(x) - \partial H_s(Y_s)\right)^2 \left(D_tY_s\right)^2\ \varphi_\epsilon(Y_s-x)\ ds\ dt\ dx\right) \\ J_3(\epsilon) &:= \mathbb{E}\left(\int_{[0,1]^2\times\mathbb{R}} \left(\partial^2 F_s(x) - \partial^2 F_s(Y_s)\right)^2\ \left(D_tY_s\right)^2\ A_s\ \varphi_\epsilon(Y_s-x)\ ds\ dt\ dx\right) \end{aligned}$$

The last two terms depend on the Malliavin derivatives of ∂F_s and H_s are they are given by

$$\begin{aligned} J_4(\epsilon) &:= \mathbb{E}\left(\int_{[0,1]^2\times\mathbb{R}} \left((D_t\partial F_s)(x) - (D_t\partial F_s)(Y_s)\right)^2\ A_s\ \varphi_\epsilon(Y_s-x)\ ds\ dt\ dx\right) \\ J_5(\epsilon) &:= \mathbb{E}\left(\int_{[0,1]^2\times\mathbb{R}} \left((D_tH_s)(x) - (D_tH_s)(Y_s)\right)^2\ \varphi_\epsilon(Y_s-x)\ ds\ dt\ dx\right) \end{aligned}$$

Arguing as above, by the dominated convergence theorem we conclude that the Skorohod integral

$$\int_0^t V_s^\epsilon\, dW_s \quad \text{converges in } \mathbb{L}_2(\Omega) \text{ as } \epsilon \to 0 \text{ to the Skorohod integral } \int_0^t V_s\, dW_s$$

This ends the proof of (6.12), and the proof of the theorem is now easily completed. ∎

We end this section with some comments.

Remark 6.4. *Recalling that the diffusion flow Y_t introduced in (6.1) has finite absolute moments of any order, the integrability conditions stated in (6.8) and (6.9) are satisfied as soon as the functions F_t, G_t, H_t, the differentials $\partial F_t, \partial^2 F_t, \partial H_t$, and the Malliavin derivatives $D_s H_t, D_s \partial F_t$ have at most polynomial growth w.r.t. the state variable.*

It is now readily check that $(H_1)'$ and $(H_2)'$ are met for the random fields introduced in (6.5).

The proof can be also be extended without difficulties to multivariate models. Following the proof of proposition 3.1 in [41], an alternative proof of theorem 6.3 based on Itô formula for Hilbert space valued processes can be developed. This elegant functional approach requires to introduce a custom Hilbert-space valued processes framework but this approach avoids to do explicitly the interchange of integration using the Fubini theorem for Skorohod and measure theory integrals. As the statement of proposition 3.1 in [41], the assumptions of theorem 6.3 can also be weaken when expressed in terms of this generalized stochastic calculus for Hilbert-space valued processes.

7 Illustrations

7.1 Perturbation analysis

Assume that $\overline{\sigma} = \sigma$ and the drift function $\overline{b}_t$ is given by a first order expansion

$$\overline{b}_t(x) = b_{\delta,t}(x) := b_t(x) + \delta\, b^{(1)}_{\delta,t}(x) \quad \text{with} \quad b^{(1)}_{\delta,t}(x) = b^{(1)}_t(x) + \frac{\delta}{2}\, b^{(2)}_{\delta,t}(x)$$

for some perturbation parameter $\delta \in [0,1]$ and some functions $b^{(i)}_{\delta,t}(x)$ with $i = 1,2$.

In this context, the stochastic flow $\overline{X}_{s,t}(x) := X^\delta_{s,t}(x)$ can be seen as a δ-perturbation of $X_{s,t}(x) := X^0_{s,t}(x)$.

We further assume that the unperturbed diffusion satisfies condition $(\mathcal{T})_2$.

To avoid unnecessary technical discussions on the existence of absolute moments of the flows we also assume that $b^{(i)}_{\delta,t}(x)$ are uniformly bounded w.r.t. the parameters (δ, t, x). In addition, $b^{(1)}_t(x)$ is differentiable w.r.t. the coordinate x and it has uniformly bounded gradients. In this situation, we set

$$\|b^{(i)}\| := \sup_{\delta,t,x} \|b^{(i)}_{\delta,t}(x)\| \quad \text{and} \quad \|\nabla b^{(1)}\| := \sup_{t,x} \|\nabla b^{(1)}_t(x)\|$$

With some additional work to estimate the absolute moments of the flows, the perturbation analysis presented below allows to handle more general models. The methodology described in this section can also be extended to expand the flow $X^\delta_{s,t}(x)$ at any order as soon as $\delta \mapsto b_{\delta,t}(x)$ is sufficiently smooth.

The first order approximation is given by the following theorem.

Theorem 7.1. *For any $s \leqslant t$, $x \in \mathbb{R}^d$ and $\delta \geqslant 0$ we have the first order expansion*

$$X^\delta_{s,t}(x) = X_{s,t}(x) + \delta\, \partial X_{s,t}(x) + \frac{\delta^2}{2}\, \partial^2_\delta X_{s,t}(x) \tag{7.1}$$

with the first order stochastic flow

$$\partial X_{s,t}(x) := \int_s^t \, (\nabla X_{u,t})\,(X_{s,u}(x))' \; b_u^{(1)}(X_{s,u}(x))\; du$$

The remainder second order term $\partial_\delta^2 X_{s,t}(x)$ in the above display is such that for any $n \geqslant 2$ s.t. $\lambda_A(n) > 0$ we have the uniform estimate

$$\sup_{s,t,x} \mathbb{E}[\|\partial_\delta^2 X_{s,t}(x)\|^n]^{1/n} \leqslant c_n$$

Proof. Using (4.12) we readily check that

$$DX^\delta_{s,t}(x) \;\; := \;\; \delta^{-1}[X^\delta_{s,t}(x) - X_{s,t}(x)] = \int_s^t \, (\nabla X_{u,t})\,(X^\delta_{s,u}(x))' \; b^{(1)}_{\delta,u}(X^\delta_{s,u}(x))\; du$$

By proposition 3.2 for any $n \geqslant 2$ we have

$$\lambda_A^+(n) := \lambda_A - (n-2)\rho(\nabla\sigma)^2/2 > 0 \Longrightarrow \mathbb{E}\left(\|DX^\delta_{s,t}(x)\|^n\right)^{1/n} \leqslant c\;\|b^{(1)}\|/\lambda_A^+(n) \qquad (7.2)$$

This yields the first order Taylor expansion (7.1) with

$$\partial_\delta^2 X_{s,t}(x) := \partial_\delta^{(2,1)} X_{s,t}(x) + \partial_\delta^{(2,2)} X_{s,t}(x)$$

and the second order remainder terms

$$\begin{aligned}
\partial_\delta^{(2,2)} X_{s,t}(x) \;\; &:= \;\; \int_s^t \, (\nabla X_{u,t})\,(X^\delta_{s,u}(x))' \; b^{(2)}_{\delta,t}(X^\delta_{s,u}(x))\; du \\
\partial_\delta^{(2,1)} X_{s,t}(x) \;\; &:= \;\; 2\delta^{-1}\int_s^t \, \Big[(\nabla X_{u,t})\,(X^\delta_{s,u}(x)) - (\nabla X_{u,t})\,(X_{s,u}(x))\Big]' \; b^{(1)}_u(X^\delta_{s,u}(x))\; du \\
&\qquad + 2\delta^{-1}\int_s^t \, (\nabla X_{u,t})\,(X_{s,u}(x))' \; [b^{(1)}_u(X^\delta_{s,u}(x)) - b^{(1)}_u(X_{s,u}(x))]\; du
\end{aligned}$$

Arguing as above, for any $n \geqslant 2$ s.t. $\lambda_A^+(n) > 0$ we have the uniform estimate

$$\mathbb{E}\left(\|\partial_\delta^{(2,2)} X_{s,t}(x)\|^n\right)^{1/n} \leqslant c\;\|b^{(2)}\|/\lambda_A^+(n)$$

To estimate $\partial_\delta^{(2,1)} X_{s,t}(x)$ we need to consider the second order decompositions

$$\begin{aligned}
&2^{-1}\;\partial_\delta^{(2,1)} X_{s,t}(x) \\
&= \int_0^1 \int_s^t \left[\nabla^2 X_{u,t}\right] \Big(X_{s,u}(x) + \epsilon(X^\delta_{s,u}(y) - X_{s,u}(x))\Big)' \;\; \Big[b^{(1)}_u(X^\delta_{s,u}(x)) \otimes DX^\delta_{s,u}(x)\Big] \;\; du\; d\epsilon \\
&+ \int_0^1 \int_s^t \, (\nabla X_{u,t})\,(X_{s,u}(x))' \; \nabla b^{(1)}_u \Big(X_{s,u}(x) + \epsilon(X^\delta_{s,u}(x) - X_{s,u}(x)), y\Big)' \; DX^\delta_{s,u}(x)\; du\; d\epsilon
\end{aligned}$$

Combining proposition 3.3 with the estimate (7.2) for any $n \geqslant 2$ s.t. $\lambda_A(n) > 0$ we check that

$$\mathbb{E}[\|\partial_\delta^{(2,1)} X_{s,t}(x)\|^n]^{1/n} \leqslant c\;\; (1 + n\;\chi(b,\sigma)/\lambda_A(n)) \;\; \Big(\|b^{(1)}\|/\lambda_A(n)\Big)^2$$

for some universal constant $c < \infty$ and the parameter $\chi(b,\sigma)$ introduced in (2.6). This ends the proof of (7.1). The proof of the theorem is completed. ■

7.2 Interacting diffusions

Consider a system of N interacting and $\mathbb{R}^d$-valued diffusion flows $X^i_{s,t}(x)$, with $1 \leqslant i \leqslant N$ given by a stochastic differential equation of the form

$$dX^i_{s,t}(x) = B_t\left(X^i_{s,t}(x), \frac{1}{N}\sum_{1\leqslant i\leqslant N} X^j_{s,t}(x)\right)\, dt + \sigma_t\left(\frac{1}{N}\sum_{1\leqslant i\leqslant N} X^j_{s,t}(x)\right)\, dW^i_t$$

for some Lipschitz functions $B_t(x,y)$ and $\sigma_t(y)$ with appropriate dimensions. In the above display, W^i_t stands for a collection of independent copies of d-dimensional Brownian motion W_t. Assume that $B_t(x,y)$ linear w.r.t. the first coordinate.

In this situation, up to a change of probability space, the empirical mean of the process

$$\overline{X}_{s,t}(x) := \frac{1}{N}\sum_{1\leqslant i\leqslant N} X^j_{s,t}(x)$$

satisfies the stochastic differential equation

$$d\overline{X}_{s,t}(x) = b_t\left(\overline{X}_{s,t}(x)\right)\, dt + \frac{1}{\sqrt{N}}\,\sigma_t\left(\overline{X}_{s,t}(x)\right)\, dW_t \quad \text{with} \quad b_t(x) := B_t(x,x)$$

Formally, the above diffusion converges as $N \to \infty$ to the flow $X_{s,t}(x)$ of the dynamical system defined by

$$\partial_t X_{s,t}(x) := b_t\left(X_{s,t}(x)\right)$$

More rigorously and without further work, the forward-backward interpolation formula (1.10) yields directly the bias-variance error decomposition

$$\overline{X}_{s,t}(x) - X_{s,t}(x) = \frac{1}{2N}\int_s^t \left(\nabla^2 X_{u,t}\right)(\overline{X}_{s,u}(x))'\; a_u(\overline{X}_{s,u}(x))\; du$$

$$+\frac{1}{\sqrt{N}}\int_s^t\; (\nabla X_{u,t})\,(\overline{X}_{s,u}(x))'\;\sigma_u(\overline{X}_{s,u}(x))\; dW_u$$

This readily implies the a.s. convergence

$$\overline{X}_{s,t}(x) \longrightarrow_{N\to\infty} X_{s,t}(x)$$

After some elementary manipulations we check the bias formula

$$\lim_{N\to\infty} N\;\left[\mathbb{E}(\overline{X}_{s,t}(x)) - X_{s,t}(x)\right] = \frac{1}{2}\int_s^t \left(\nabla^2 X_{u,t}\right)(X_{s,u}(x))'\; a_u(X_{s,u}(x))\; du$$

We also have the almost sure fluctuation theorem

$$\lim_{N\to\infty} \sqrt{N}\;\left[\overline{X}_{s,t}(x) - X_{s,t}(x)\right] = \int_s^t\; (\nabla X_{u,t})\,(X_{s,u}(x))'\;\sigma_u(X_{s,u}(x))\; dW_u$$

7.3 Time discretization schemes

This section is mainly concerned with the proof of proposition 1.4. We fix some parameter $h > 0$ and some $s \geqslant 0$ and for any $t \in [s+kh, s+(k+1)h[$ we set

$$dX^h_{s,t}(x) = Y^h_{s,t}(x)\; dt + \sigma\; dW_t \quad \text{with} \quad Y^h_{s,t}(x) := b\left(X^h_{s,s+kh}(x)\right)$$

for some fluctuation parameter $\sigma \geqslant 0$. For any $s+kh \leqslant u < s+(k+1)h$ we have

$$X^h_{s,u}(x) - X^h_{s,s+kh}(x) = Y^h_{s,u}(x)\ (u-(s+kh)) + \sigma\ (W_u - W_{s+kh})$$

Using (4.12), in terms of the tensor product (2.1) we readily check that

$$X^h_{s,t}(x) - X_{s,t}(x) = \int_s^t\ (\nabla X_{u,t})\,(X^h_{s,u}(x))'\ \ \left[Y^h_{s,u}(x) - b(X^h_{s,u}(x))\right]\ du$$

Combining (3.5) with the Minkowski integral inequality we check that

$$\begin{aligned}\mathbb{E}\left(\|X^h_{s,t}(x) - X_{s,t}(x)\|^n\right)^{1/n} &= \int_s^t\ \mathbb{E}\left(\|\,(\nabla X_{u,t})\,(X^h_{s,u}(x))'\ \ \left[Y^h_{s,u}(x) - b(X^h_{s,u}(x))\right]\|^n\right)^{1/n}\ du\\ &= \int_s^t e^{-\lambda(t-u)}\ \mathbb{E}\left(\|Y^h_{s,u}(x) - b(X^h_{s,u}(x))\|^n\right)^{1/n}\ du\end{aligned}$$

where the second line follows from the exponential estimate of the tangent process from proposition 3.1. The integrand will be bounded as follows: for any $s+kh \leqslant u < s+(k+1)h$ and any $n \geqslant 1$ we have

$$\mathbb{E}\left(\|b(X^h_{s,u}(x)) - Y^h_{s,u}(x))\|^n\right)^{1/n} \leqslant \|\nabla b\|\ \ \left([\|b(0)\| + \widehat{m}_n(x)\ \|\nabla b\|]\ h + \sigma\ \sqrt{h}\right)$$

which then yields the stated result of the proposition. We now prove the stated bound on the difference of the drift processes. For any $s+kh \leqslant u < s+(k+1)h$ we have

$$\begin{aligned}&b(X^h_{s,u}(x)) - Y^h_{s,u}(x)\\ &= \left[\int_0^1 \nabla b\left(X^h_{s,s+kh}(x) + \epsilon(X^h_{s,u}(x) - X^h_{s,s+kh}(x))\right)'\ b\left(X^h_{s,s+kh}(x)\right)\ d\epsilon\right]\ (u-(s+kh))\\ &\qquad + \left[\int_0^1 \nabla b\left(X^h_{s,s+kh}(x) + \epsilon(X^h_{s,u}(x) - X^h_{s,s+kh}(x))\right)'\ d\epsilon\right]\ \sigma\ \ (W_u - W_{s+kh})\end{aligned} \tag{7.3}$$

The $\mathbb{L}_n$-norm of the second integral term is bounded by $\|\nabla b\|\sigma\sqrt{h}$.

The assumption $\langle x, b(x)\rangle \leqslant -\beta\ \|x\|^2$, for some $\beta > 0$, implies the stochastic flows $X_{s,t}(x)$ has uniform absolute moments of any order $n \geqslant 1$ w.r.t. the time horizon, that is, we have that

$$m_n(x) \leqslant \kappa_n\ (1 + \|x\|) \quad \text{with } m_n(x) \text{ defined in (2.10).}$$

The stochastic flows $X^h_{s,t}(x)$ also obey a similar moment bound: observe that for any $t \in [s+kh, s+(k+1)h[$ we have

$$\begin{aligned}&d\|X^h_{s,t}(x)\|^2\\ &\leqslant \left[-2\lambda_0\ \|X^h_{s,t}(x)\|^2 + 2\ \langle X^h_{s,t}(x), b(X^h_{s,s+kh}(x)) - b(X^h_{s,t}(x))\rangle + \sigma^2 d\right]\ dt + 2\sigma\ X^h_{s,t}(x)'dW_t\end{aligned}$$

Thus, for any $\epsilon > 0$ we have

$$d\|X^h_{s,t}(x)\|^2 \leqslant \left[(-2\lambda_0 + \epsilon)\|X^h_{s,t}(x)\|^2 + \epsilon^{-1}\|\nabla b\| + \sigma^2 d\right]\ dt + 2\sigma\ X^h_{s,t}(x)'dW_t$$

We can check that the stochastic flows $X^h_{s,t}(x)$ also have uniform moments w.r.t. the time horizon; that is, for any $n \geqslant 1$ we have that

$$\widehat{m}_n(x) := \sup_{h\geqslant 0}\sup_{t\geqslant s}\mathbb{E}\left[\|X^h_{s,t}(x)\|^n\right]^{1/n} \leqslant c_n\ (1 + \|x\|)$$

Using this bounds, we check that

$$\mathbb{E}(\|b(X^h_{s,s+kh}(x))\|^n)^{1/n} = \|b(0)\| + \hat{m}_n(x)\|\nabla b\|$$

The end of the proof now follows elementary manipulations, thus it is skipped. The proof of proposition 1.4 is now completed. ■

Appendix

In this appendix we prove the estimates (1.16) and (2.10) and proposition 3.3.

Proof of (2.10)

Whenever $(\mathcal{M})_n$ is satisfied, we have

$$2\langle x, b_t(x)\rangle + \|\sigma_t(x)\|_F^2 \leqslant \gamma_0 + \gamma_1\|x\| - \gamma_2\|x\|^2$$

with the parameters

$$\gamma_0 = \alpha_0 + 2\beta_0 \qquad \gamma_1 = \alpha_1 + 2\beta_1 \quad \text{and} \quad \gamma_2 = 2\beta_2 - \alpha_2$$

Observe that

$$\begin{aligned} & d\|X_{s,t}(x)\|^2 \\ & = \left[2\left\langle X_{s,t}(x), b_t(X_{s,t}(x))\right\rangle + \|\sigma_t(X_{s,t}(x))\|_F^2\right]\ dt + 2\textstyle\sum_k \langle X_{s,t}(x), \sigma_{k,t}(X_{s,t}(x))\rangle\ dW_t^k \end{aligned}$$

After some elementary computations, for any $n \geqslant 1$ we check that

$$\begin{aligned} n^{-1}\partial_t\mathbb{E}\left[\|X_{s,t}(x)\|^{2n}\right] &\leqslant -\left[\gamma_2 - 2(n-1)\alpha_2\right]\ \mathbb{E}\left[\|X_{s,t}(x)\|^{2n}\right] \\ &\quad + \left[\gamma_1 + 2(n-1)\alpha_1\right]\ \mathbb{E}\left[\|X_{s,t}(x)\|^{2n-1}\right] + \left[\gamma_0 + 2(n-1)\alpha_0\right]\ \mathbb{E}\left[\|X_{s,t}(x)\|^{2(n-1)}\right] \end{aligned}$$

This implies that

$$\begin{aligned} \partial_t\mathbb{E}\left[\|X_{s,t}(x)\|^{2n}\right]^{1/n} &\leqslant -\left[\gamma_2 - 2(n-1)\alpha_2\right]\ \mathbb{E}\left[\|X_{s,t}(x)\|^{2n}\right]^{1/n} \\ &\quad + \left[\gamma_1 + 2(n-1)\alpha_1\right]\ \mathbb{E}\left[\|X_{s,t}(x)\|^{2n}\right]^{1/(2n)} + \left[\gamma_0 + 2(n-1)\alpha_0\right] \end{aligned}$$

from which we check that for any $\epsilon > 0$ we have

$$\begin{aligned} & \partial_t\mathbb{E}\left[\|X_{s,t}(x)\|^{2n}\right]^{1/n} \\ & \leqslant -\left[\gamma_2 - 2(n-1)\alpha_2 - 2\epsilon\right]\ \mathbb{E}\left[\|X_{s,t}(x)\|^{2n}\right]^{1/n} + \frac{1}{8\epsilon}\left[\gamma_1 + 2(n-1)\alpha_1\right]^2 + \left[\gamma_0 + 2(n-1)\alpha_0\right] \end{aligned}$$

This implies that

$$\begin{aligned} & \partial_t\mathbb{E}\left[\|X_{s,t}(x)\|^{2n}\right]^{1/n} \\ & \leqslant -2\left[\beta_2 - (n-1/2)\alpha_2 - \epsilon\right]\ \mathbb{E}\left[\|X_{s,t}(x)\|^{2n}\right]^{1/n} + \frac{1}{8\epsilon}\left[\gamma_1 + 2(n-1)\alpha_1\right]^2 + \left[\gamma_0 + 2(n-1)\alpha_0\right] \end{aligned}$$

from which we check that

$$\mathbb{E}\left[\|X_{s,t}(x)\|^{2n}\right]^{1/n} \leqslant e^{-2[\beta_2-(n-1/2)\alpha_2-\epsilon](t-s)}\ \|x\|^2 + \frac{1}{8\epsilon}\ \frac{\left[\gamma_1+2(n-1)\alpha_1\right]^2+\left[\gamma_0+2(n-1)\alpha_0\right]}{2\left[\beta_2-(n-1/2)\alpha_2-\epsilon\right]}$$

as soon as $\epsilon < \beta_2-(n-1/2)\alpha_2$ and $n \geqslant 1$. Replacing ϵ by $\epsilon(\beta_2-(n-1/2)\alpha_2)$ and then $(2n)$ by n we check that

$$\begin{aligned}
&\mathbb{E}\left[\|X_{s,t}(x)\|^{n}\right]^{1/n}\\
&\leqslant e^{-(1-\epsilon)\beta_2(n)(t-s)}\ \|x\| + \frac{1}{4\sqrt{\epsilon(1-\epsilon)}}\ \frac{\gamma_1(n)+\gamma_0(n)^{1/2}}{\beta_2(n)^{1/2}} \quad \text{with} \quad \gamma_i(n) := \gamma_i+(n-2)\alpha_i
\end{aligned}$$

This ends the proof of (2.10). ■

Proof of proposition 3.3

The proof of the estimate (3.10) is mainly based on the following technical lemma of its own interest.

Lemma 7.2. *Let Z_t be a non negative diffusion process satisfying in integral sense an inequality of the following form*

$$dZ_t \leqslant (-\lambda Z_t + \alpha_t\ \sqrt{Z_t}+\beta_t)\ dt + dM_t \quad \text{with} \quad \partial_t\langle M\rangle_t \leqslant (u_t\sqrt{Z_t}+v_tZ_t)^2$$

for some parameters $\lambda > 0$ and $v_t \geqslant 0$, and some non negative processes (α_t,β_t,u_t). In this situation, for any $\epsilon > 0$ we have

$$\mathbb{E}(Z_t^n)^{1/n} \leqslant e^{\int_0^t \lambda_{n,s}(\epsilon)ds}\ \mathbb{E}(Z_0^n)^{1/n} + \int_0^t e^{\int_s^t \lambda_{n,u}(\epsilon)du}\ z_s^n(\epsilon)\ ds \tag{7.4}$$

with the parameters

$$\begin{aligned}
\lambda_{n,t}(\epsilon) &:= -\lambda + \frac{n-1}{2}\ v_t^2 + \frac{\epsilon}{2}\\
z_t^n(\epsilon) &:= \mathbb{E}\left[\beta_t^n\right]^{1/n} + \frac{n-1}{2}\ \mathbb{E}\left[u_t^{2n}\right]^{1/n} + \frac{1}{\epsilon}\ \left(\mathbb{E}\left[\alpha_t^{2n}\right]^{1/n} + (n-1)^2\ \mathbb{E}\left[(u_tv_t)^{2n}\right]^{1/n}\right)
\end{aligned}$$

Proof. Applying Itô's formula, for any $n \geqslant 2$, we have

$$\begin{aligned}
&n^{-1}\partial_t\mathbb{E}(Z_t^n)\\
&\leqslant \mathbb{E}\left[Z_t^{n-1}(-\lambda Z_t+\alpha_t\ \sqrt{Z_t}+\beta_t) + \frac{n-1}{2}\ (u_t\sqrt{Z_t}+v_tZ_t)^2\ Z_t^{n-2}\right]\\
&= \left(-\lambda+\frac{n-1}{2}\ v_t^2\right)\ \mathbb{E}(Z_t^n) + \mathbb{E}\left[\left(\beta_t+\frac{n-1}{2}\ u_t^2\right)\ Z_t^{n-1}\right] + \mathbb{E}\left(\left[\alpha_t+(n-1)u_tv_t\right]\ Z_t^{n-1/2}\right)
\end{aligned}$$

On the other hand, for any $\epsilon > 0$ we have the almost sure inequality

$$\left[\alpha_t+(n-1)u_tv_t\right]\ Z_t^{(n-1)/2}\ Z_t^{n/2} \leqslant \frac{1}{2\epsilon}\ \left[\alpha_t+(n-1)u_tv_t\right]^2\ Z_t^{n-1} + \frac{\epsilon}{2}\ Z_t^n$$

This implies that

$$n^{-1}\partial_t\mathbb{E}(Z_t^n)$$

$$\leqslant \lambda_{n,t}(\epsilon)\ \mathbb{E}(Z_t^n)+\mathbb{E}\left[\left(\beta_t+\frac{n-1}{2}\ u_t^2+\frac{1}{2\epsilon}\ [\alpha_t+(n-1)u_tv_t]^2\right)\ Z_t^{n-1}\right]$$

Applying Hölder inequality we check that

$$\mathbb{E}\left[\left(\beta_t+\frac{n-1}{2}\ u_t^2+\frac{1}{2\epsilon}\ [\alpha_t+(n-1)u_tv_t]^2\right)\ Z_t^{n-1}\right]$$

$$\leqslant \mathbb{E}\left[\left(\beta_t+\frac{n-1}{2}\ u_t^2+\frac{1}{2\epsilon}\ [\alpha_t+(n-1)u_tv_t]^2\right)^n\right]^{1/n}\ \mathbb{E}(Z_t^n)^{1-1/n}\leqslant z_t^n\ \mathbb{E}(Z_t^n)^{1-1/n}$$

This yields the estimate

$$\partial_t\mathbb{E}(Z_t^n)^{1/n}=\mathbb{E}(Z_t^n)^{-(1-1/n)}\ n^{-1}\partial_t\mathbb{E}(Z_t^n)\leqslant \lambda_{n,t}(\epsilon)\ \mathbb{E}(Z_t^n)^{1/n}+z_t^n$$

This ends the proof of the lemma. ■

We set

$$Y_{s,t}(x):=\|\nabla^2X_{s,t}(x)\|_F^2\quad\text{and}\quad T_{s,t}(x):=\|\nabla X_{s,t}(x)\|_F$$

and we also consider the collection of parameters

$$\|\tau\|_F\ :=\ \sup_{t,x}\|\tau_t(x)\|_F\qquad \rho(\upsilon)\ :=\ \sup_{t,x}\lambda_{max}(\upsilon_t(x))$$

with the tensor functions (τ_t,υ_t) introduced in (3.9). Observe that

$$\|\tau\|_F\leqslant\|\nabla^2 b\|_F+d\ \|\nabla^2\sigma\|_F^2\quad\text{and}\quad\rho(\upsilon)\leqslant d\ \|\nabla^2\sigma\|_2^2$$

Whenever $(\mathcal{T})_2$ is met we have

$$\mathrm{Tr}\left[\nabla^2X_{s,t}(x)\ A_t(X_{s,t}(x))\ \nabla^2X_{s,t}(x)'\right]\leqslant -2\lambda_A\ Y_{s,t}(x)$$

Also observe that

$$|\mathrm{Tr}\left[[\nabla X_{s,t}(x)\otimes\nabla X_{s,t}(x)]\ \ \tau_t(X_{s,t}(x))\ \nabla^2X_{s,t}(x)'\right]|\leqslant\|\tau\|_F\ Y_{s,t}(x)^{1/2}\ T_{s,t}(x)^2$$

and

$$\mathrm{Tr}\left[[\nabla X_{s,t}(x)\otimes\nabla X_{s,t}(x)]\,\upsilon_t(X_{s,t}(x))\,[\nabla X_{s,t}(x)\otimes\nabla X_{s,t}(x)]'\right]\leqslant\rho(\upsilon)\ T_{s,t}(x)^4$$

In the same vein, we have

$$\begin{aligned}&|\mathrm{Tr}\left\{[\nabla X_{s,t}(x)\otimes\nabla X_{s,t}(x)]\ \ \nabla^2\sigma_{t,k}(X_{s,t}(x))\ \nabla^2X_{s,t}(x)'\right.\\&\qquad\qquad\left.+\nabla^2X_{s,t}(x)\ \nabla\sigma_{t,k}(X_{s,t}(x))\ \nabla^2X_{s,t}(x)'\right\}|\\&\leqslant\|\nabla^2\sigma_k\|_F\ T_{s,t}(x)^2\ Y_{s,t}(x)^{1/2}+\rho(\nabla\sigma_k)\ Y_{s,t}(x)\end{aligned}$$

We are now in position to prove proposition 3.3.

Proof of proposition 3.3:

Applying the above lemma to the processes

$$Z_t = Y_{s,t}(x) \qquad \lambda = 2\lambda_A \qquad \alpha_t = 2\|\tau\|_F\ T_{s,t}(x)^2 \qquad \beta_t = \rho(\upsilon)\ T_{s,t}(x)^4$$

and the parameters

$$u_t = 2\sqrt{d}\ \|\nabla^2\sigma\|_F\ T_{s,t}(x)^2 \quad \text{and} \quad v_t = 2\sqrt{d}\ \rho_\star(\nabla\sigma)$$

we obtain the estimate (7.4) with the parameters

$$\begin{aligned}
\lambda_{n,t}(\epsilon) &:= -2\left[\lambda_A - d(n-1)\rho_\star(\nabla\sigma)^2 - \frac{\epsilon}{4}\right] \\
z_t^n(\epsilon) &:= \left\{\rho(\upsilon)\ + 2d(n-1)\ \|\nabla^2\sigma\|_F^2\ \ + \frac{4}{\epsilon}\ \left(\|\tau\|_F^2\ + 4\ d^2(n-1)^2\ \rho_\star(\nabla\sigma)^2\ \|\nabla^2\sigma\|_F^2\ \right)\right\} \\
&\qquad\qquad \times \mathbb{E}\left[\|\nabla X_{s,t}(x)\|_F^{4n}\right]^{1/n}
\end{aligned}$$

Observe that

$$z_t^n(\epsilon) \leqslant cn^2\ (1 \vee \epsilon^{-1})\ \chi(b,\sigma)^2\ \mathbb{E}\left[\|\nabla X_{s,t}(x)\|_F^{4n}\right]^{1/n}$$

for some universal constant $c < \infty$ and the parameter $\chi(b,\sigma)$ defined in (2.6). Using (3.8) we check that

$$\begin{aligned}
&\mathbb{E}\left(\|\nabla^2 X_{s,t}(x)\|_F^{2n}\right)^{1/n} \\
&\leqslant cn^2\ (1 \vee \epsilon^{-1})\ \chi(b,\sigma)^2\ \int_s^t e^{-2\left[\lambda_A - d(n-1)\rho_\star(\nabla\sigma)^2 - \frac{\epsilon}{4}\right](t-u)}\ e^{-4\left[\lambda_A - (n-1)\rho(\nabla\sigma)^2\right](u-s)}\ du \\
&= cn^2\ (1 \vee \epsilon^{-1})\ \chi(b,\sigma)^2\ e^{-2\left[\lambda_A - d(n-1)\rho_\star(\nabla\sigma)^2 - \frac{\epsilon}{4}\right](t-s)} \\
&\qquad \int_s^t e^{-2\left[\lambda_A - (n-1)\rho(\nabla\sigma)^2 + (n-1)[d\rho_\star(\nabla\sigma)^2 - \rho(\nabla\sigma)^2] + \frac{\epsilon}{4}\right](u-s)}\ du
\end{aligned}$$

Assume that

$$\lambda_A > d(n-1)\rho_\star(\nabla\sigma)^2$$

In this case there exists some $0 < \epsilon_n \leqslant 1$ such that for any $0 < \epsilon \leqslant \epsilon_n$ we have

$$\lambda_A - d(n-1)\rho_\star(\nabla\sigma)^2 > \epsilon$$

and therefore

$$\mathbb{E}\left(\|\nabla^2 X_{s,t}(x)\|_F^{2n}\right)^{1/(2n)} \leqslant c\ n\ \epsilon^{-1}\ \chi(b,\sigma)\ \exp\left(-\left[\lambda_A - d(n-1)\rho_\star(\nabla\sigma)^2 - \epsilon\right](t-s)\right)$$

This ends the proof of the proposition. ∎

Proof of (1.16)

Using (2.14), the generalized Minkowski inequality applied to (1.10) whenever $(\mathcal{T})_{n/\delta}$ is met for some $\delta \in]0,1[$ and $n \geqslant 2$ gives

$$\begin{aligned}&\mathbb{E}\left[\|T_{s,t}(\Delta a,\Delta b)(x)\|^n\right]^{1/n}\\ &\leqslant \frac{\kappa_{n/\delta}}{\lambda(n/\delta)}\ \left(\vert\!\vert\!\vert \Delta b(x)\vert\!\vert\!\vert_{n/(1-\delta)}+\vert\!\vert\!\vert \Delta a(x)\vert\!\vert\!\vert_{n/(1-\delta)}\right)\quad \text{with } (\kappa_n,\lambda(n)) \text{ given in (2.13).}\end{aligned} \tag{7.5}$$

The Skorohod integral $S_{s,t}(\Delta\sigma)(x)$ is estimated using theorem 5.2. Using (7.5) and (5.9) we check that

$$\begin{aligned}&\mathbb{E}\left[\|X_{s,t}(x)-\overline{X}_{s,t}(x)\|^n\right]^{1/n}\\ &\leqslant \kappa_{(\delta_1,\delta_2),n}\ \left(\vert\!\vert\!\vert \Delta a(x)\vert\!\vert\!\vert_{n/(1-\delta_1)}+\vert\!\vert\!\vert \Delta b(x)\vert\!\vert\!\vert_{n/(1-\delta_1)}+\vert\!\vert\!\vert \Delta \sigma(x)\vert\!\vert\!\vert_{2n/\delta_2}\ (1\vee\|x\|)\right)\end{aligned}$$

as soon as the regularity conditions $(\mathcal{T})_{n/\delta_1}$, $(M)_{2n/\delta_2}$ and $(T)_{2n/(1-\delta_2)}$ are satisfied for some parameter $n \geqslant 2$ and some $\delta_1,\delta_2 \in]0,1[$. Choosing $\delta_1 = (1-\delta_2)/2$ and setting $\delta=\delta_2$ we check that

$$\begin{aligned}&\mathbb{E}\left[\|X_{s,t}(x)-\overline{X}_{s,t}(x)\|^n\right]^{1/n}\\ &\leqslant \kappa_{\delta,n}\ \left(\vert\!\vert\!\vert \Delta a(x)\vert\!\vert\!\vert_{2n/(1+\delta)}+\vert\!\vert\!\vert \Delta b(x)\vert\!\vert\!\vert_{2n/(1+\delta)}+\vert\!\vert\!\vert \Delta \sigma(x)\vert\!\vert\!\vert_{2n/\delta}\ (1\vee\|x\|)\right)\end{aligned}$$

as soon as $(M)_{2n/\delta}$ and $(T)_{2n/(1-\delta)}$ are satisfied for some parameter $n \geqslant 2$ and some $\delta \in]0,1[$. For instance, $(\mathcal{M})_{2n/\delta}$ and $(\mathcal{T})_{2n/(1-\delta)}$ are satisfied as soon as

$$\beta_2-\alpha_2/2 > (n/\delta-1)\ \alpha_2 \quad \text{and} \quad \lambda_A > d(n/(1-\delta)-1)\ \rho_\star(\nabla\sigma)^2$$

This ends the proof of (1.16). ■

References

[1] V. Alekseev. An estimate for the perturbations of the solution of ordinary differential equations. Vestn. Mosk.Univ., Ser. I, Math. Meh. vol. 2, (1961).

[2] T. Ando and J. L. van Hemmen. An inequality for trace ideals. Commun. Math. Phys., vol. 76, pp. 143–148 (1980).

[3] M. Arnaudon, P. Del Moral. A variational approach to nonlinear and interacting diffusions. ArXiv:1812.04269 (2018). Stochastic Analysis and Applications DOI: 10.1080/07362994.2019.1609985 (2019).

[4] M. Arnaudon, P. Del Moral. A duality formula and a particle Gibbs sampler for continuous time Feynman-Kac measures on path spaces. ArXiv 1805.05044 (2018). Electronic Journal of Probability 25 (2020).

[5] M. Arnaudon, P. Del Moral. A second order analysis of McKean-Vlasov semigroups. `ArXiv:1906.05140 (2019)`, Annals of Applied Probability, vol. 30, no. 6, pp. 2613–2664. (2020).

[6] M. Arnaudon, H. Plank, A. Thalmaier. A Bismut type formula for the Hessian of heat semigroups. C. R. Math. Acad. Sci. Paris, vol. 336, no. 8, pp. 661–666 (2003).

[7] R. Bellman. Some inequalities for the square Root of a Positive Definite Matrix. Linear Algebra and its applications, vol. 1, no. 3, pp. 321–324 (1968).

[8] R. Bellman, Stability Theory of Differential Equations, McGraw Hill, New York, (1953).

[9] A. N. Bishop, P. Del Moral. On the Stability of Matrix-Valued Riccati Diffusions. ArXiv:1808.00235 (2018).Electron. J. Probab., vol. 24, no. 84, 40 pp. (2019).

[10] A.N. Bishop, P. Del Moral, S.D. Pathiraja. Perturbations and projections of Kalman-Bucy semigroups. Stochastic Processes and their Applications, vol. 9, no.128, pp. 2857–2904 (2018).

[11] A. N. Bishop, P. Del Moral, A. Niclas. A perturbation analysis of stochastic matrix Riccati diffusions. `Arxiv 1709.05071 (2017)`. Ann. Inst. H. Poincaré Probab. Statist., vol. 56, no. 2, pp. 884–916.(2020).

[12] J.M. Bismut. A generalized formula of Itô and some other properties of stochastic flows, Z. Wahrschein. Werw. Geb., vol. 55, pp. 331–350 (1981).

[13] J.M. Bismut. Large deviations and the Malliavin calculus. Birkhauser Prog. Math. 45 (1984).

[14] A.P. Carverhill and K.D. Elworthy. Flows of Stochastic Dynamical Systems: The Functional Analytic Approach. Z. Wahrs 65, pp. 245–267 (1983).

[15] W.A. Coppel. Dichotomies in Stability Theory. Springer (1978).

[16] G. Da Prato, J.L. Menaldi, L. Tubaro. Some results of backward Itô formula. Stochastic analysis and applications, vol. 25, no. 3, pp. 679–703 (2007).

[17] G. Da Prato. Some remarks about backward Itô formula and applications. Stochastic analysis and applications, vol. 16, no. 6, pp. 993–1003 (1998).

[18] P. Del Moral. Feynman-Kac formulae. Genealogical and interacting particle systems with applications. Probability and its Applications (New York). (573p.) Springer-Verlag, New York (2004).

[19] Del Moral, Pierre. Mean field simulation for Monte Carlo integration. Chapman and Hall, CRC press. Monographs on Statistics and Applied Probability (2013).

[20] P. Del Moral, A. Guionnet. On the stability of measure valued processes with applications to filtering. Comptes Rendus de l'Académie des Sciences-Series I-Mathematics, vol. 329, no. 5, pp. 429–434 (1999).

[21] P. Del Moral and A. Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. Ann. Inst. Henri Poincaré, vol. 37, no. 2, pp. 155–194 (2001).

[22] P. Del Moral and L. Miclo. Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In *Séminaire de Probabilités, XXXIV*, volume 1729, *Lecture Notes in Math.*, pages 1–145. Springer, Berlin (2000).

[23] K.D. Elworthy, X.M. Li. Formulae for the Derivative of Heat Eemigroups. Journal of Functional Analysis 125, pp. 252–286 (1994).

[24] Gröbner, W. Die Lie-Reihen und ihre Anwendungen. VEB Deutscher Verlag der Wiss., Berlin (1960).

[25] T.H. Gronwall, Note on the derivatives with respect to a parameter of the solutions of a system of differential equations, Ann. Math., vol. 20, no. 2 , pp. 293–296 (1919).

[26] N. J. Higham. Functions of Matrices : Theory and Computation, SIAM, Philadelphia, PA (2008).

[27] A. Hudde, M. Hutzenthaler, A. Jentzen, S. Mazzonetto. On the Itô-Alekseev-Gröbner formula for stochastic differential equations. arXiv preprint arXiv:1812.09857 (2018).

[28] M. Hutzenthaler and A. Jentzen. On a perturbation theory and on strong convergence rates for stochastic ordinary and partial differential equations with non-globally monotone coefficients. `Arxiv 1401.0295 (2014)`. To appear in the Annals of Probability (2019)

[29] A. Iserles, G. Söderlind. Global bounds on numerical error for ordinary differential equations. Journal of Complexity, vol. 9, no. 1, pp. 97–112 (1993).

[30] A. Jentzen, F. Lindner, P. Pusnik. On the Alekseev-Gröbner formula in Banach spaces. arXiv preprint arXiv:1810.10030 (2018).

[31] H. Kunita. First order stochastic partial differential equations. Stochastic Analysis (K. Itøed.). Kinokunniya. Tokyo, pp. 249-269 (1984).

[32] H. Kunita, and M. K. Ghosh. Lectures on stochastic flows and applications. Bombay: Tata Institute of Fundamental Research (1986).

[33] N.V. Krylov. B.L. Rozowskii. On the first integrals and Liouville equations for diffusion processes. Stochastic Differential Systems. Proc. 3rd SFSP-WG 7/1, Visegrad, Hungary 1980. Lecture Notes in Control and Information Sciences, vol. 36, pp. 117-125 (1981).

[34] I. Kruk, F. Russo, C. A. Tudor. Wiener integrals, Malliavin calculus and covariance measure structure. Journal of Functional Analysis, vol. 249, no. 1, pp. 92–142 (2007).

[35] J.A. Leon. Fubini theorem for anticipating stochastic integrals in Hilbert space. Applied Mathematics and Optimization, vol. 27, no. 3, pp. 313–327 (1993).

[36] X. M. Li. Doubly Damped Stochastic Parallel Translations and Hessian Formulas. International Conference on Stochastic Partial Differential Equations and Related Fields. Springer, Cham (2016).

[37] D. Nualart. The Malliavin calculus and related topics. Vol. 1995. Berlin: Springer (2006).

[38] D. Nualart, E. Pardoux. Stochastic calculus with anticipating integrands. Probability Theory and Related Fields, vol. 78, no. 4, pp. 535–581 (1988).

[39] D. Nualart, M. Zakai. Generalized multiple stochastic integrals and the representation of wiener functionals. Stochastics, vol. 23, no. 3, pp. 311–330 (1988).

[40] J. R. Norris. Simplified Malliavin calculus. Séminaire de probabilités (Strasbourg), tome 20, p. 101–130 (1986).

[41] D. Ocone, E. Pardoux. A generalized Itô-Ventzell formula. Application to a class of anticipating stochastic differential equations. In Annales de l'IHP Probabilités et statistiques, vol. 25, no. 1, pp. 39–71 (1989).

[42] E. Pardoux. Applications of anticipating stochastic calculus to stochastic differential equations. In Stochastic Analysis and Related Topics II (pp. 63-105). Springer, Berlin, Heidelberg (1990).

[43] E. Pardoux, P. Protter. A two-sided stochastic integral and its calculus. Probability Theory and Related Fields, vol. 76, no. 1, pp. 15–49 (1987).

[44] O. Purtukhia. Fubini type theorems for ordinary and stochastic integrals. Proceedings of A. Razmadze Mathematical Institute, vol. 130, pp. 101–114 (2002).

[45] M. Scheutzow. A stochastic Gronwall lemma. Infinite Dimensional Analysis, Quantum Probability and Related Topics, vol. 16, no. 2, p. 1350019 (2013).

[46] J. Thompson. Derivatives of Feynman-Kac semigroups. Journal of Theoretical Probability, vol. 32, no. 2, pp. 950–973 (2019).

[47] J. Timmer, S. Haussler, M. Lauk, and C.-H. Lucking, Pathological tremors: Deterministic chaos or nonlinear stochastic oscillators. Chaos: An Interdisciplinary Journal of Nonlinear Science, vol. 10 no. 1, pp. 278–288 (2000).

[48] S. Watanabe. Lectures on stochastic differential equations and Malliavin Calculus. Tata Institute of Fundamental Research. Springer-Verlag (1984).

[49] M. Wu. A note on stability of linear time-varying systems. IEEE Transactions on Automatic Control. vol. 19, no. 2. pp. 162–162 (1974).

Acknowledgments

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

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