A Malliavin calculus approach to score functions in diffusion generative models

Ehsan Mirafzali[1*], Frank Proske[2], Utkarsh Gupta[1], Daniele Venturi[3], Razvan Marinescu[1]

[1]Department of Computer Science, University of California Santa Cruz, Santa Cruz, California, USA.
[2]Department of Mathematics, University of Oslo, Oslo, Norway.
[3]Department of Applied Mathematics, University of California Santa Cruz, Santa Cruz, California, USA.

*Corresponding author(s). E-mail(s): smirafza@ucsc.edu;
Contributing authors: proske@math.uio.no; utgupta@ucsc.edu; venturi@ucsc.edu; ramarine@ucsc.edu;

Abstract

Score-based diffusion generative models have recently emerged as a powerful tool for modelling complex data distributions. These models aim at learning the score function, which defines a map from a known probability distribution to the target data distribution via deterministic or stochastic differential equations (SDEs). The score function is typically estimated from data using a variety of approximation techniques, such as denoising or sliced score matching, Hyvärien's method, or Schrödinger bridges. In this paper, we derive an exact, closed form, expression for the score function for a broad class of nonlinear diffusion generative models. Our approach combines modern stochastic analysis tools such as Malliavin derivatives and their adjoint operators (Skorokhod integrals or Malliavin Divergence) with a new Bismut-type formula. The resulting expression for the score function can be written entirely in terms of the first and second variation processes, with all Malliavin derivatives systematically eliminated, thereby enhancing its practical applicability. The theoretical framework presented in this work offers a principled foundation for advancing score estimation methods in generative modelling, enabling the design of new sampling algorithms for complex probability distributions. Our results can be extended to broader classes of stochastic differential equations, opening new directions for the development of score-based diffusion generative models.

Keywords: Score function, diffusion models, Bismut formula, Malliavin calculus, stochastic differential equations, generative modelling.

1 Introduction

Diffusion generative models [1] have recently emerged as a powerful tool for modelling complex data distributions in a variety of applications ranging from image synthesis [2, 3] to finance [4]. Score-based diffusion models [5], which study diffusion processes through score functions and SDEs, provide an equivalent formulation of these generative models. Rather than learning the data distribution directly, these models estimate the so-called score function $\nabla_y \log p_t(y)$, which governs the transformation of a simple reference distribution into the target distribution via a time-reversed nonlinear diffusion process.

Historically, score-based generative models have been built upon diffusion processes described by linear Itô's stochastic differential equations, i.e., SDEs with drift and diffusion coefficients that are linear in the state variables. Such a linearity assumption ensures analytical tractability and comes with other desirable properties. For example, the Fokker–Planck equation [6, 7] associated with the diffusion process admits a Gaussian stationary distribution and Gaussian transition densities, which have been leveraged in the development of techniques such as Denoising Diffusion Probabilistic Models (DDPM) [8]. While the linearity assumption in diffusion processes has been instrumental in developing the first generation of score-based generative models, it also represents an overly restrictive simplification. Many real-world systems exhibit nonlinear dynamics, where both the drift and diffusion coefficients depend nonlinearly on the state variables. Extending diffusion models to such settings poses substantial challenges. Specifically, unlike linear SDEs, nonlinear SDEs typically lack closed-form expressions for their transition probability densities, making the computation of the score function $\nabla_y \log p_t(y)$ extremely challenging.

The purpose of this paper is to bridge this gap by establishing a new rigorous link between score-based diffusion models and Malliavin calculus, a stochastic calculus of variations introduced by Paul Malliavin in the 1970s [9, 10]. Originally devised to investigate hypoelliptic operators and stochastic partial differential equations, Malliavin calculus provides a robust framework for analysing the smoothness of SDE solutions and their densities. By leveraging Malliavin derivatives and a novel Bismut-type formula we derive in this paper, which is distinct from the classical Bismut–Elworthy–Li formula [11–13] tailored to heat kernels, we derive an explicit representation of the score function for nonlinear diffusion processes. The resulting formula is expressed entirely in terms of first and second variation processes [14, 15], with all Malliavin derivatives systematically eliminated.

In particular, our framework generalises the state-independent results of [16] to fully nonlinear SDEs with state-dependent diffusion coefficients. For the numerical experiments of this framework, see the same work [16]. Very recently, other studies have applied Malliavin calculus to diffusion generative models for different ends: one focuses on robust numerical treatment of singular diffusion bridges and conditional control, and additionally derives an explicit yet distinct conditional-score formula via a generalised Tweedie identity [17], and another proposes an extension of diffusion models to infinite-dimensional diffusion settings using Malliavin calculus [18].

The implications of this new theoretical framework are profound. First, Malliavin calculus provides a powerful tool for overcoming the analytical challenges posed by nonlinear diffusion processes. Second, its rigorous foundations enable the development and analysis of new classes of generative models beyond linear diffusion. From a practical standpoint, this can potentially yield more accurate generative models for applications such as high-fidelity image synthesis, fluid dynamics [19], and protein folding simulations in molecular dynamics [20].

This paper is organised as follows. In Section 2 we state our main theorem expressing the score function $\nabla_y \log p_t(y)$ in terms of first and second variation processes via a Bismut-type formula. Unlike the Bismut-Elworthy-Li formula, our formula directly targets the score function, leveraging the first and second variation processes to yield a practical closed form expression. In Section 3 we briefly review basic notions of Malliavin calculus and and prove various theorems on covering vector fields. In Section 4 we discuss first and second variation processes in detail. In Section 5 we prove our main theorem, i.e., Theorem 1. The proof includes the full derivation of the score function for nonlinear SDEs, along with several useful lemmas that assist in the process. In Section 6 we provide a simplified expression of the score function for nonlinear SDEs with state-independent diffusion coefficients.

2 Problem statement and main result

Consider the m-dimensional stochastic differential equation

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dB_t, \qquad X_0 = x, \qquad 0 \le t \le T, \tag{1}$$

where $X_t \in \mathbb{R}^m$, $B_t \in \mathbb{R}^d$ is a standard Brownian motion, $b : [0, T] \times \mathbb{R}^m \to \mathbb{R}^m$ is the drift coefficient, and $\sigma : [0, T] \times \mathbb{R}^m \to \mathbb{R}^{m\times d}$ is the diffusion coefficient, with $\sigma^l(t, X_t)$ denoting its l-th column for $l = 1, \dots, d$. The initial condition $x \in \mathbb{R}^m$ is deterministic, and $T > 0$ is a fixed terminal time. We assume that b and σ are $\mathcal{C}^2$ functions with bounded derivatives, which ensures the existence of a unique strong solution to (1).

Let

$$\gamma_{X_T} := \big(\langle DX_T^i, DX_T^j\rangle_{L^2([0,T])}\big)_{i,j=1}^m$$

denote the Malliavin covariance matrix of X_T. If $X_T \in \mathbb{D}^\infty$ and γ_{X_T} is almost surely invertible with $\mathbb{E}\big[\|\gamma_{X_T}^{-1}\|^p\big] < \infty$ for every $p > 0$, then the terminal value X_T admits a smooth probability density $p(y)$ with respect to the Lebesgue measure on $\mathbb{R}^m$, (see [21], sec 7).

In the framework of score-based diffusion models, the SDE (1) represents the so-called *forward diffusion process*, which gradually transforms the initial data distribution into some other distribution over the time interval $[0, T]$. To enable generative sampling, we define the corresponding *reverse-time diffusion process*, which runs backwards in time from T to 0. Based on the general structure of reverse-time SDEs, this process is given by

$$dX_t = \left\{b(t, X_t) - \nabla \cdot \left[\sigma(t, X_t)\sigma(t, X_t)^\top\right] + \sigma(t, X_t)\nabla_{X_t} \log p_t(X_t)\right\} dt + \sigma(t, X_t)\, d\bar{B}_t, \tag{2}$$

for all $T \geq t \geq 0$, where $\nabla \cdot [\sigma(t, X_t)\sigma(t, X_t)^\top]$ denotes the divergence of the matrix $\sigma(t, X_t)\sigma(t, X_t)^\top$ (a correction term arising from the Itô interpretation of the SDE), $\nabla_{X_t} \log p_t(X_t)$ is the score function, and $d\bar{B}_t$ represents the increment of a reverse-time Brownian motion. The reverse-time SDE effectively reverses the forward diffusion by leveraging the score function to guide the process from a noise distribution back to the original data distribution. Our goal is to compute the score function

$$\nabla_y \log p_t(y) = \frac{\nabla_y p_t(y)}{p_t(y)}, \quad \partial_{y_k} \log p_t(y) = \frac{\partial_{y_k} p_t(y)}{p_t(y)}, \quad k = 1, \dots, m,$$

which is essential for implementing the reverse diffusion process in generative tasks. In particular, we are interested in $\nabla_y \log p_T(y)$[1]. To achieve this goal, we leverage Malliavin calculus and a new Bismut-type formula we develop in this paper. As we shall see hereafter, this allows us to express the score in terms of the first variation process $Y_t = \partial X_t/\partial x$ and the second variation process $Z_t = \partial^2 X_t/\partial x^2$, which capture the first- and the second-order sensitivity of the solution X_t relative to changes in the initial condition x. Our main result is the following theorem.

Theorem 1 (Skorokhod integral representation theorem for the score function) *Let X_t be the solution to the stochastic differential equation*

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t, \quad X_0 = x_0,$$

where $b : [0, T] \times \mathbb{R}^m \to \mathbb{R}^m$ and $\sigma : [0, T] \times \mathbb{R}^m \to \mathbb{R}^{m \times d}$ are sufficiently smooth functions, and B_t is a d-dimensional Brownian motion. Define the first variation process

$$dY_t = \partial_x b(t, X_t) Y_t\, dt + \sum_{l=1}^{d} \partial_x \sigma^l(t, X_t) Y_t\, dB_t^l, \quad Y_0 = I_m,$$

and the Malliavin covariance matrix of X_T as

$$\gamma_{X_T} = \int_0^T Y_T Y_s^{-1} \sigma(s, X_s)\sigma(s, X_s)^\top (Y_s^{-1})^\top Y_T^\top\, ds$$

which we assume to be invertible almost surely. Furthermore, define the random field $u_t(x)$ and the random variable F_k as follows[2]

$$u_t(x) = x^\top Y_t^{-1} \sigma(t, X_t), \quad x \in \mathbb{R}^m,$$
$$F_k = Y_T^\top \gamma_{X_T}^{-1} e_k, \quad \text{with} \quad F_k^j = e_j^\top Y_T^\top \gamma_{X_T}^{-1} e_k,$$

[1]For generative tasks, one requires the time-dependent score $\nabla_y \log p_t(y)$ for each intermediate time t, which is obtained by evaluating the integrals discussed in Theorem 1 at every such t.

[2]Note that $u_t(x) \in \mathbb{R}^d$ since $x^\top$ is a $1 \times m$ row vector, Y_t^{-1} is an $m \times m$ matrix, and $\sigma(t, X_t)$ is an $m \times d$ matrix, resulting in a $1 \times d$ row vector. Moreover, $u_t(x)$ is adapted to the natural Brownian filtration $\mathcal{F}_t$ for each fixed x, as it depends only on Y_t^{-1} and $\sigma(t, X_t)$, both of which are adapted processes.

where Y_t is the first variation process, Y_T is its value at time T, Y_t^{-1} its inverse at time t, and e_k is the k-th standard basis vector in $\mathbb{R}^m$. Then, following [21], for each $k = 1, \ldots, m$, the gradient of the log-density satisfies

$$\partial_{y_k} \log p_T(y) = -\mathbb{E}\left[\delta(u_k) \mid X_T = y\right], \tag{3}$$

where $\delta(u_k)$ denotes the Skorokhod integral of the process $u_k(t)$, i.e., the Malliavin divergence, whose explicit representation is given by

$$\begin{aligned} \delta(u_k) = \int_0^T u_t(x) \cdot dB_t \Big|_{x=F_k} &- \int_0^T \sum_{j=1}^m \left[Y_t^{-1}\sigma(t, X_t)\right]_j \cdot A_{jk}(t)\, dt \\ &+ \int_0^T \sum_{j=1}^m \left[Y_t^{-1}\sigma(t, X_t)\right]_j \cdot B_{jk}(t)\, dt + \int_0^T \sum_{j=1}^m \left[Y_t^{-1}\sigma(t, X_t)\right]_j \cdot C_{jk}(t)\, dt, \end{aligned} \tag{4}$$

where

$$\begin{aligned} A_{jk}(t) &= e_j^\top \Big[\sigma(t, X_t)^\top (Y_t^{-1})^\top Z_T^\top - \sigma(t, X_t)^\top (Y_t^{-1})^\top Z_t^\top (Y_t^{-1})^\top Y_T^\top \\ &\qquad + \left(Y_T Y_t^{-1} \partial_x \sigma(t, X_t) Y_t\right)^\top\Big] \gamma_{X_T}^{-1} e_k, \\ B_{jk}(t) &= e_j^\top Y_T^\top \gamma_{X_T}^{-1} \cdot \left[\int_0^t I_1(t, s)\, ds + \int_0^t I_2(t, s)\, ds\right] \gamma_{X_T}^{-1} e_k, \\ C_{jk}(t) &= e_j^\top Y_T^\top \gamma_{X_T}^{-1} \cdot \left[\int_t^T I_3(t, s)\, ds + \int_t^T I_4(t, s)\, ds\right] \gamma_{X_T}^{-1} e_k, \end{aligned}$$

with integrands (p and q below denote components)

$$\begin{aligned} I_1^{p,q}(t, s) &= \left[\Omega(t) Y_s^{-1} \sigma(s, X_s)\right]^p \cdot [Y_T Y_s^{-1} \sigma(s, X_s)]^q, \\ I_2^{p,q}(t, s) &= [Y_T Y_s^{-1} \sigma(s, X_s)]^p \cdot \left[\Omega(t) Y_s^{-1} \sigma(s, X_s)\right]^q, \\ I_3^{p,q}(t, s) &= \left[\Omega(t) Y_s^{-1} \sigma(s, X_s) + Y_T \Theta(t, s)\right]^p \cdot [Y_T Y_s^{-1} \sigma(s, X_s)]^q, \\ I_4^{p,q}(t, s) &= [Y_T Y_s^{-1} \sigma(s, X_s)]^p \cdot \left[\Omega(t) Y_s^{-1} \sigma(s, X_s) + Y_T \Theta(t, s)\right]^q, \end{aligned}$$

and

$$\begin{aligned} \Omega(t) &= Z_T Y_t^{-1} \sigma(t, X_t) - Y_T Y_t^{-1} Z_t Y_t^{-1} \sigma(t, X_t) + Y_T Y_t^{-1} \partial_x \sigma(t, X_t) Y_t, \\ \Theta(t, s) &= -Y_s^{-1} \Big[Z_s Y_t^{-1} \sigma(t, X_t) - Y_s Y_t^{-1} Z_t Y_t^{-1} \sigma(t, X_t) \\ &\qquad + Y_s Y_t^{-1} \partial_x \sigma(t, X_t) Y_t\Big] Y_s^{-1} \sigma(s, X_s) + Y_s^{-1} \partial_x \sigma(s, X_s) \left(Y_s Y_t^{-1} \sigma(t, X_t)\right). \end{aligned}$$

The second variation process appearing in $\Omega(t)$ and $\Theta(t, s)$ is defined as

$$dZ_t = [\partial_{xx} b(t, X_t)(Y_t \otimes Y_t) + \partial_x b(t, X_t) Z_t]\, dt$$

$$+ \sum_{l=1}^{d} \left[\partial_{xx}\sigma^l(t, X_t)(Y_t \otimes Y_t) + \partial_x \sigma^l(t, X_t) Z_t \right] dB_t^l,$$

with initial condition $Z_0 = 0$.

The formula (4) for the Skorokhod integral $\delta(u_k)$ consists of several components that capture the interaction between stochastic processes and their Malliavin derivatives. The first term $\int_0^T u_t(x) \cdot dB_t \Big|_{x=F_k}$ represents the Itô integral from Bismut's formula, evaluated at the specific point $x = F_k$. This term captures the direct influence of the Brownian motion on the process. The remaining components in the expansion of $\sum_{j=1}^m \partial_j u_t(F_k) \cdot D_t F_k^j$ are subtracted from the first component and are indexed by the triplet $(j, p, q) \in \{1, \ldots, m\}^3$. This indexing arises from extracting matrix components of the inverse Malliavin matrix $\gamma_{X_T}^{-1}$ and selecting vector components of $Y_t^{-1}\sigma(t, X_t)$. The formula incorporates integrals over different time intervals, i.e., integrals over $s \in [0, t]$ and integrals over $s \in [t, T]$. These appear in the $D_t \gamma_{X_T}^{p,q}$ expansions and represent how the Malliavin derivative propagates through the system depending on whether the perturbation time t occurs before or after the reference time s.

Note that Theorem 1 provides an explicit representation of $\nabla_y \log p_t(y)$ that avoids abstract Malliavin derivatives, and therefore it is computationally practical for applications in score-based generative modelling.

3 Malliavin calculus and covering vector fields

Malliavin calculus enables us to analyse the regularity of the random process X_T and compute derivatives of its density. It also allows us to represent score functions via the Bismut-type formula[3]

$$\partial_{y_k} \log p_T(y) = -\mathbb{E}\left[\delta(u_k) \mid X_T = y\right], \qquad k = 1, \ldots, m \tag{5}$$

where $\delta(u_k)$ is the Skorokhod integral and $u_k = \{u_k(t) : 0 \le t \le T\}$, with $u_k(t) \in \mathbb{R}^d$, is a process called the *covering vector field*, [21, 22]. The covering vector field $u_k(t)$ must satisfy

$$\langle DX_T^i, u_k \rangle_H = \delta_{ik}, \quad i, k = 1, \ldots, m, \tag{6}$$

where δ_{ik} is the Kronecker delta, $DX_T^i = \{D_t X_T^i : 0 \le t \le T\}$ is the Malliavin derivative of X_T^i, a process in the Hilbert space $H = L^2\left([0, T], \mathbb{R}^d\right)$,

$$\langle f, g \rangle_H = \int_0^T f(t) \cdot g(t)\, dt$$

is the inner product in H, and $\cdot$ denotes the standard dot product in $\mathbb{R}^d$. Condition (6) ensures that $u_k(t)$ "covers" the k-th direction in the Malliavin sense, which will

[3] The negative sign arises from the integration-by-parts formula in Malliavin calculus, and the expectation is conditioned on $X_T = y$, reflecting the evaluation of the density gradient at a specific point.

allow us to isolate $\partial_{y_k} p(y)$. Define the Malliavin covariance matrix $\gamma_{X_T} \in \mathbb{R}^{m\times m}$ with entries

$$\gamma^{i,j}_{X_T} = \langle DX^i_T, DX^j_T \rangle_H, \tag{7}$$

where $D_t X^i_T \in \mathbb{R}^d$ is the Malliavin derivative of X^i_T at time t. This matrix measures the "randomness" induced by the Brownian motion up to time T. We assume γ_{X_T} is invertible almost surely, so $\gamma^{-1}_{X_T}$ exists, which is necessary for $p(y)$ to be smooth. Hereafter we show that for each $k = 1, \dots, m$, the process

$$u_k(t) = \sum_{j=1}^{m} \left(\gamma^{-1}_{X_T}\right)_{k,j} D_t X^j_T, \tag{8}$$

where $\left(\gamma^{-1}_{X_T}\right)_{k,j}$ is the (k, j)-th element of the inverse Malliavin matrix $\gamma^{-1}_{X_T}$, satisfies the covering condition

$$\langle DX^i_T, u_k \rangle_H = \delta_{ik} \quad \text{for all} \quad i = 1, \dots, m.$$

In the remainder of this section, we state and prove several important properties of the chosen covering vector field (8). Although some of these results may already be known or established in prior works, we include proofs here for the reader's convenience and to ensure the exposition remains self-contained, for more details, see [23].

We begin with a theorem establishing the existence and uniqueness of a covering vector field $u_k(t)$ for an $\mathbb{R}^m$-valued random variable $X_T = (X^1_T, \dots, X^m_T)$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a d-dimensional Brownian motion B_t over $[0, T]$. The theorem assumes that X_T lies in $\mathbb{D}^{2,2}$, the space of twice Malliavin differentiable random variables relative to the Gaussian structure of B, and that its Malliavin covariance matrix γ_{X_T}, with entries $\gamma^{i,j}_{X_T} = \langle DX^i_T, DX^j_T \rangle_H$ in the Cameron-Martin space $H = L^2([0, T], \mathbb{R}^d)$, is invertible almost surely. Under these conditions, for each $k = 1, \dots, m$, it can be shown that there exists a unique process $u_k(t) \in L^2([0, T] \times \Omega, \mathbb{R}^d)$ in the domain of the Skorokhod integral Dom(δ), satisfying $\langle DX^i_T, u_k \rangle_H = \delta_{ik}$, where δ_{ik} is the Kronecker delta. This result is fundamental in stochastic analysis, particularly within the Malliavin calculus framework.

Theorem 2 (Existence and uniqueness of the covering vector field) *Let $X_T = (X^1_T, \dots, X^m_T)$ be an $\mathbb{R}^m$-valued random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, equipped with a Brownian motion $B = \{B_t\}_{t\in[0,T]}$ taking values in $\mathbb{R}^d$. Suppose X_T belongs to $\mathbb{D}^{2,2}$, the space of twice Malliavin differentiable random variables with respect to the Gaussian structure induced by B. Let γ_{X_T} be the Malliavin covariance matrix of X_T defined in (7). Assume that γ_{X_T} is invertible almost surely. Additionally, assume there exists $p > 2(m + 1)$ such that*

$$\mathbb{E}\left[\lambda_{\min}(\gamma_{X_T})^{-p}\right] < \infty,$$

where $\lambda_{\min}(\gamma_{X_T})$ denotes the smallest eigenvalue of γ_{X_T}.

Then, for each $k = 1, \dots, m$, there exists a unique process $u_k = \{u_k(t)\}_{t\in[0,T]}$ with $u_k(t) \in L^2([0, T] \times \Omega, \mathbb{R}^d)$, belonging to the domain of the Skorokhod integral δ, denoted Dom(δ), and minimising $\mathbb{E}\left[\int_0^T \|u(t)\|^2\, dt\right]$, such that

$$\langle DX^i_T, u_k \rangle_H = \delta_{ik}, \quad i = 1, \dots, m, \tag{9}$$

where δ_{ik} is the Kronecker delta.

Proof This construction leverages the invertibility of γ_{X_T} to define a process that aligns with the directions specified by the Malliavin derivatives. To verify (9), let us compute the inner product in H

$$\langle DX_T^i, u_k\rangle_H = \int_0^T D_t X_T^i \cdot u_k(t)\, dt.$$

Substitute (8) into the previous equation to obtain

$$\langle DX_T^i, u_k\rangle_H = \int_0^T D_t X_T^i \cdot \left(\sum_{j=1}^m (\gamma_{X_T}^{-1})_{jk} D_t X_T^j \right) dt.$$

Since $D_t X_T^i$ and $D_t X_T^j$ are $\mathbb{R}^d$-valued, and assuming integrability, interchange the sum and integral

$$\langle DX_T^i, u_k\rangle_H = \sum_{j=1}^m (\gamma_{X_T}^{-1})_{jk} \int_0^T D_t X_T^i \cdot D_t X_T^j\, dt.$$

Next, we notice that

$$\int_0^T D_t X_T^i \cdot D_t X_T^j\, dt = \langle DX_T^i, DX_T^j\rangle_H = \gamma_{X_T}^{i,j},$$

which implies

$$\langle DX_T^i, u_k\rangle_H = \sum_{j=1}^m (\gamma_{X_T}^{-1})_{jk} \gamma_{X_T}^{i,j}.$$

Since $\gamma_{X_T}\gamma_{X_T}^{-1} = I_m$, the (i,k)-th entry of the product is

$$\sum_{j=1}^m \gamma_{X_T}^{i,j} (\gamma_{X_T}^{-1})_{jk} = \delta_{ik},$$

confirming the covering condition holds.

To show that (8) is in $L^2([0,T]\times\Omega, \mathbb{R}^d)$ we must show that

$$\mathbb{E}\left[\int_0^T \|u_k(t)\|^2\, dt \right] < \infty.$$

To this end, we first compute the vector norm

$$\begin{aligned}
\|u_k(t)\|^2 =& u_k(t)\cdot u_k(t) \\
=& \left(\sum_{j=1}^m (\gamma_{X_T}^{-1})_{jk} D_t X_T^j \right) \cdot \left(\sum_{l=1}^m (\gamma_{X_T}^{-1})_{lk} D_t X_T^l \right) \\
=& \sum_{j=1}^m \sum_{l=1}^m (\gamma_{X_T}^{-1})_{jk} (\gamma_{X_T}^{-1})_{lk} (D_t X_T^j \cdot D_t X_T^l).
\end{aligned}$$

Thus,

$$\int_0^T \|u_k(t)\|^2\, dt = \sum_{j=1}^m \sum_{l=1}^m (\gamma_{X_T}^{-1})_{jk} (\gamma_{X_T}^{-1})_{lk} \int_0^T D_t X_T^j \cdot D_t X_T^l\, dt.$$

Next, we take the expectation

$$\mathbb{E}\left[\int_0^T \|u_k(t)\|^2\,dt\right] = \mathbb{E}\left[\sum_{j=1}^m \sum_{l=1}^m (\gamma_{X_T}^{-1})_{jk}(\gamma_{X_T}^{-1})_{lk}\gamma_{X_T}^{j,l}\right].$$

Upon definition of the vector $\mathbf{v}_k = \left((\gamma_{X_T}^{-1})_{1k}, \dots, (\gamma_{X_T}^{-1})_{mk}\right)^\top$, we can write

$$\sum_{j=1}^m \sum_{l=1}^m (\gamma_{X_T}^{-1})_{jk}(\gamma_{X_T}^{-1})_{lk}\gamma_{X_T}^{j,l} = \mathbf{v}_k^\top \gamma_{X_T}\mathbf{v}_k.$$

Since $\gamma_{X_T}\mathbf{v}_k = \mathbf{e}_k$ (the k-th standard basis vector),
 we have

$$\mathbf{v}_k^\top \gamma_{X_T}\mathbf{v}_k = \mathbf{v}_k^\top \mathbf{e}_k = (\gamma_{X_T}^{-1})_{kk}.$$

Thus,

$$\mathbb{E}\left[\int_0^T \|u_k(t)\|^2\,dt\right] = \mathbb{E}\left[(\gamma_{X_T}^{-1})_{kk}\right].$$

Since γ_{X_T} is symmetric and positive definite, we have $(\gamma_{X_T}^{-1})_{kk} \le \|\gamma_{X_T}^{-1}\|_{\text{op}} = \frac{1}{\lambda_{\min}(\gamma_{X_T})}$, where $\|\cdot\|_{\text{op}}$ denotes the operator norm (see [24], Sec. 4.2). Therefore,

$$\mathbb{E}\left[(\gamma_{X_T}^{-1})_{kk}\right] \le \mathbb{E}\left[\frac{1}{\lambda_{\min}(\gamma_{X_T})}\right].$$

By assumption, there exists $p > 2(m+1)$ such that $\mathbb{E}\left[\lambda_{\min}(\gamma_{X_T})^{-p}\right] < \infty$. Using Hölder's inequality with exponents p and $q = \frac{p}{p-1} > 1$, we have

$$\mathbb{E}\left[\frac{1}{\lambda_{\min}(\gamma_{X_T})}\right] \le \left(\mathbb{E}\left[\lambda_{\min}(\gamma_{X_T})^{-p}\right]\right)^{1/p} \left(\mathbb{E}[1^q]\right)^{1/q} < \infty,$$

since $\mathbb{E}[1^q] = 1$. Thus, $\mathbb{E}[(\gamma_{X_T}^{-1})_{kk}] < \infty$, ensuring $u_k(t) \in L^2([0,T]\times\Omega, \mathbb{R}^d)$.

To establish the uniqueness of the process $u_k(t)$, suppose there exists another process $v_k(t) \in L^2([0,T]\times\Omega,\mathbb{R}^d)$ that also satisfies the covering condition

$$\langle DX_T^i, v_k\rangle_H = \delta_{ik} \quad \text{for all} \quad i = 1,\dots,m,$$

and minimises $\mathbb{E}\left[\int_0^T \|v(t)\|^2\,dt\right]$. Since $u_k(t) \in L^2([0,T]\times\Omega,\mathbb{R}^d)$ satisfies the same condition and minimises the norm, we aim to prove that $u_k(t) = v_k(t)$ almost everywhere in $[0,T]\times\Omega$. To this end, define the difference $w(t) = u_k(t) - v_k(t)$. Then, for each $i = 1,\dots,m$, compute the inner product

$$\langle DX_T^i, w\rangle_H = \langle DX_T^i, u_k\rangle_H - \langle DX_T^i, v_k\rangle_H = \delta_{ik} - \delta_{ik} = 0.$$

Thus, $w(t)$ is orthogonal to each DX_T^i in H. Our goal is to show that this orthogonality, combined with the minimal-norm condition, implies $w(t) = 0$ almost everywhere.

To achieve this, consider the linear operator $A : H \to \mathbb{R}^m$ defined by

$$Ah = (\langle DX_T^1, h\rangle_H, \langle DX_T^2, h\rangle_H, \dots, \langle DX_T^m, h\rangle_H) \quad \text{for all} \quad h \in H.$$

This operator maps a function $h \in H$ to an m-dimensional vector whose components are the inner products of h with the Malliavin derivatives DX_T^i. Define the adjoint operator $A^* : \mathbb{R}^m \to H$, which satisfies the duality relation

$$\langle Ah, \mathbf{c}\rangle_{\mathbb{R}^m} = \langle h, A^*\mathbf{c}\rangle_H \quad \text{for all} \quad h \in H, \mathbf{c} \in \mathbb{R}^m.$$

To determine A^*, let $\mathbf{c} = (c_1, c_2, \dots, c_m) \in \mathbb{R}^m$. Then

$$\langle Ah, \mathbf{c} \rangle_{\mathbb{R}^m} = \sum_{i=1}^m c_i \langle DX_T^i, h \rangle_H = \left\langle h, \sum_{i=1}^m c_i DX_T^i \right\rangle_H .$$

This equality holds for all $h \in H$, so

$$A^* \mathbf{c} = \sum_{i=1}^m c_i DX_T^i.$$

Now, compute the operator $AA^* : \mathbb{R}^m \to \mathbb{R}^m$

$$AA^* \mathbf{c} = A \left(\sum_{i=1}^m c_i DX_T^i \right) = \left(\left\langle DX_T^j, \sum_{i=1}^m c_i DX_T^i \right\rangle_H \right)_{j=1}^m = \left(\sum_{i=1}^m c_i \langle DX_T^j, DX_T^i \rangle_H \right)_{j=1}^m .$$

Since $\gamma_{X_T}^{j,i} = \langle DX_T^j, DX_T^i \rangle_H$, this becomes

$$AA^* \mathbf{c} = \left(\sum_{i=1}^m c_i \gamma_{X_T}^{j,i} \right)_{j=1}^m = \gamma_{X_T} \mathbf{c}.$$

By assumption, γ_{X_T} is invertible almost surely, meaning AA^* is an invertible linear operator on $\mathbb{R}^m$. Since $\mathbb{R}^m$ is finite-dimensional, invertibility of AA^* implies that $A : H \to \mathbb{R}^m$ is surjective.

The invertibility of γ_{X_T} also ensures that the set $\{DX_T^1, DX_T^2, \dots, DX_T^m\}$ is linearly independent in H. To verify this, suppose that

$$\sum_{i=1}^m c_i DX_T^i = 0$$

for some coefficients $\mathbf{c} = (c_1, \dots, c_m)$. Taking the inner product with itself

$$\left\langle \sum_{i=1}^m c_i DX_T^i, \sum_{j=1}^m c_j DX_T^j \right\rangle_H = \sum_{i,j=1}^m c_i c_j \langle DX_T^i, DX_T^j \rangle_H = \mathbf{c}^\top \gamma_{X_T} \mathbf{c} = 0.$$

Since γ_{X_T} is positive-definite (a consequence of its invertibility), $\mathbf{c}^\top \gamma_{X_T} \mathbf{c} = 0$ if and only if $\mathbf{c} = 0$. Thus, the DX_T^i are linearly independent.

Next, consider the nullspace of the operator A, i.e.,

$$\ker A = \{h \in H \mid Ah = 0\} = \left\{ h \in H \mid \langle DX_T^i, h \rangle_H = 0 \text{ for all } i = 1, \dots, m \right\}.$$

Since we have shown that $w \in \ker A$, we need to determine the implications for w. Define $S = \operatorname{span}\{DX_T^1, \dots, DX_T^m\}$, the range of A^*, and note that $\ker A = S^\perp$. Since $u_k \in S$ by construction (as $u_k(t) = \sum_{j=1}^m (\gamma_{X_T}^{-1})_{jk} D_t X_T^j$), and $w \in S^\perp$, we have $\langle u_k, w \rangle_H = 0$. Compute the norm of v_k

$$\mathbb{E}\left[\int_0^T \|v_k(t)\|^2 \, dt \right] = \mathbb{E}\left[\int_0^T \|u_k(t) - w(t)\|^2 \, dt \right] = \mathbb{E}\left[\int_0^T \|u_k(t)\|^2 \, dt \right] + \mathbb{E}\left[\int_0^T \|w(t)\|^2 \, dt \right],$$

since the cross term vanishes due to orthogonality. Both terms are non-negative, so $\|v_k\|_{L^2}^2 \geq \|u_k\|_{L^2}^2$, with equality if and only if $w = 0$ almost everywhere. Since v_k also minimises the norm, $\|v_k\|_{L^2}^2 = \|u_k\|_{L^2}^2$, implying $w = 0$, hence $v_k = u_k$ almost everywhere. Therefore, u_k is the unique minimiser of $\mathbb{E}\left[\int_0^T \|u(t)\|^2 \, dt \right]$ among all processes satisfying (9). This establishes uniqueness without assuming $v_k \in S$.

Finally, we must show that $u_k \in \mathrm{Dom}(\delta)$, the domain of the Skorokhod integral. A process $u_k \in L^2([0,T] \times \Omega, \mathbb{R}^d)$ belongs to $\mathrm{Dom}(\delta)$ if there exists $\delta(u_k) \in L^2(\Omega)$ such that for all smooth random variables $F \in \mathbb{D}^\infty$

$$\mathbb{E}[F\delta(u_k)] = \mathbb{E}\left[\int_0^T D_t F \cdot u_k(t)\, dt\right].$$

A sufficient condition for this is that $u_k \in \mathbb{D}^{1,2}(H)$, meaning that u_k defined in (8) is Malliavin differentiable with $Du_k \in L^2([0,T] \times \Omega \times [0,T], \mathbb{R}^d \otimes \mathbb{R}^d)$.

To verify that $u_k \in \mathbb{D}^{1,2}(H)$, let us first consider the entries of the inverse Malliavin matrix $\gamma_{X_T}^{-1}$. Since $\gamma_{X_T}^{i,j} = \langle D_t X_T^i, D_t X_T^j \rangle_H$ and $X_T \in \mathbb{D}^{2,2}$, each entry $\gamma_{X_T}^{i,j}$ lies in $\mathbb{D}^{1,2}$.

Because $\mathbb{D}^{2,2}$ is stable under smooth polynomial operations, the cofactors $\mathrm{adj}(\gamma_{X_T})_{jk}$ also belong to $\mathbb{D}^{1,2}$. Writing Cramer's formula

$$(\gamma_{X_T}^{-1})_{jk} = \frac{\mathrm{adj}(\gamma_{X_T})_{jk}}{\det \gamma_{X_T}},$$

and using that $\det(\gamma_{X_T})$ is bounded away from zero on a set of full measure together with the moment condition $\mathbb{E}[\lambda_{\min}(\gamma_{X_T})^{-p}] < \infty$ for some $p > 2(m+1)$, we obtain $(\gamma_{X_T}^{-1})_{jk} \in \mathbb{D}^{1,2}$.

Although the map $A \mapsto A^{-1}$ is not $\mathcal{C}_b^1$, the Malliavin chain rule (Proposition 1.2.3 of [23]) applies to each smooth component of the adjugate, and the negative-moment bound guarantees the required square-integrability of the derivatives. (If $X_T \in \mathbb{D}^\infty$, one may invoke Lemma 2.1.6 of [23] directly to reach the same conclusion.) Applying the chain rule, the Malliavin derivative $D_s(\gamma_{X_T}^{-1})_{jk}$ involves terms bounded by $\lambda_{\min}(\gamma_{X_T})^{-(m+1)}$. Since $\mathbb{E}[\lambda_{\min}(\gamma_{X_T})^{-p}] < \infty$, and the derivatives $D_s(\gamma_{X_T}^{-1})_{jk}$ are expressible as sums of terms bounded by $C \cdot \lambda_{\min}(\gamma_{X_T})^{-(m+1)}$ (where C depends polynomially on the entries of γ_{X_T}), Hölder's inequality can be applied iteratively. Specifically, for any exponent $k > p$, set

$$r = \frac{k}{p}, \quad s = \frac{k}{k-p},$$

to obtain

$$\mathbb{E}\left[\lambda_{\min}(\gamma_{X_T})^{-k}\right] \leq \left(\mathbb{E}\left[\lambda_{\min}(\gamma_{X_T})^{-p}\right]\right)^{p/k} \cdot \left(\mathbb{E}[1^s]\right)^{(k-p)/k} < \infty.$$

This ensures all necessary moments exist, confirming $D_s(\gamma_{X_T}^{-1})_{jk} \in L^2(\Omega)$ and thus $(\gamma_{X_T}^{-1})_{jk} \in \mathbb{D}^{1,2}$. Additionally, since $X_T \in \mathbb{D}^{2,2}$, $D_t X_T^j \in \mathbb{D}^{1,2}(H)$, with finite second moments for its derivatives.

For each term $\left(\gamma_{X_T}^{-1}\right)_{jk} D_t X_T^j$ in (8), note that if $F \in \mathbb{D}^{1,2}$ and $U \in \mathbb{D}^{1,2}(H)$, then $FU \in \mathbb{D}^{1,2}(H)$ and

$$D_s(FU(t)) = (D_s F)U(t) + F D_s U(t).$$

Since $(\gamma_{X_T}^{-1})_{jk} \in \mathbb{D}^{1,2}$ and $D_t X_T^j \in \mathbb{D}^{1,2}(H)$, their product is in $\mathbb{D}^{1,2}(H)$. Thus, $u_k(t)$, a finite sum of such terms, is in $\mathbb{D}^{1,2}(H)$. In Malliavin calculus, $u \in \mathbb{D}^{1,2}(H)$ implies $u \in \mathrm{Dom}(\delta)$, and $\delta(u) \in L^2(\Omega)$. Hence, $u_k \in \mathrm{Dom}(\delta)$.

Thus, the covering vector field (8) exists, is unique under the minimal-norm condition, and satisfies all required properties. This completes the proof. □

The next theorem concerns the regularity of the covering vector field $u_k(t)$ for a stochastic process $\left(X_t\right)_{t\in[0,T]}$ adapted to a Brownian filtration. It assumes that the terminal value X_T lies in $\mathbb{D}^{2,2}$, that the Malliavin covariance matrix γ_{X_T} is almost surely invertible, and it defines $u_k(t)$ as in (8). Additional regularity conditions are: uniform ellipticity of γ_{X_T} (i.e. $\gamma_{X_T} \geq \lambda I$ for some $\lambda > 0$); and, when X_t solves the SDE $\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t$, the coefficients b and σ are $\mathcal{C}^2$ with all derivatives up to second order bounded. Furthermore, under the uniform ellipticity bound one automatically has $\mathbb{E}\big[\|\gamma_{X_T}^{-1}\|^q\big] < \infty$ for every $q > 0$. Under these hypotheses the theorem shows that each $u_k(t)$ belongs to $\mathrm{Dom}\,(\delta)$ and that its Skorokhod integral satisfies $\delta(u_k) \in L^p(\Omega)$ for some $p > 1$, thereby confirming the desired regularity of the vector field.

Theorem 3 (Regularity of the covering vector field) *Let $X = (X_t)_{t\in[0,T]}$ be the unique solution of*

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t,$$

where $b, \sigma \in \mathcal{C}^2$ with all mixed derivatives up to order two bounded. Suppose further that

(i) $X_T \in \mathbb{D}^{2,2}$, and its Malliavin covariance matrix $\gamma_{X_T} = \big(\langle DX_T^i, DX_T^j\rangle_{L^2([0,T])}\big)_{i,j}$ is a.s. invertible with $\gamma_{X_T} \geq \lambda I_m$ for some $\lambda > 0$.

(ii) Define, for each $k = 1, \ldots, m$, the covering vector field

$$u_k(t) = \sum_{j=1}^{m} (\gamma_{X_T}^{-1})_{jk}\; D_t X_T^j, \quad t \in [0, T].$$

Then each u_k belongs to $\mathrm{Dom}\,(\delta)$*, and its Skorokhod integral $\delta(u_k)$ is well-defined with $\delta(u_k) \in L^p(\Omega)$ for some $p > 1$.*

Proof The Skorokhod integral δ is the adjoint of the Malliavin derivative operator. For a process $u = \{u(t)\}_{t\in[0,T]}$ taking values in $\mathbb{R}^d$, the Skorokhod integral is defined componentwise as $\delta(u) = \sum_{\ell=1}^{d} \delta(u^\ell e_\ell)$, where e_ℓ are the standard basis vectors in $\mathbb{R}^d$. Because $u_k \in \mathbb{D}^{1,2}(\mathcal{H})$, each scalar component $u_k^\ell \in \mathbb{D}^{1,2}\big(L^2[0,T]\big)$, hence $\delta(u_k^\ell e_\ell)$ is well-defined and lies in $L^2(\Omega)$. The process u belongs to $\mathrm{Dom}(\delta)$ if

a) $u \in L^2([0,T] \times \Omega; \mathbb{R}^d)$,

b) There exists a random variable $\delta(u) \in L^2(\Omega)$ such that for all smooth test functions $F \in \mathbb{D}^\infty$

$$\mathbb{E}\left[F\delta(u)\right] = \mathbb{E}\left[\int_0^T D_t F \cdot u(t)\, dt\right].$$

A sufficient condition for $u \in \mathrm{Dom}(\delta)$ is that $u \in \mathbb{D}^{1,2}(L^2([0,T]; \mathbb{R}^d))$, and

$$\|u\|^2_{\mathbb{D}^{1,2}(L^2([0,T];\mathbb{R}^d))} = \mathbb{E}\left[\int_0^T |u(t)|^2\, dt\right] + \mathbb{E}\left[\int_0^T \int_0^T |D_s u(t)|^2\, ds\, dt\right] < \infty.$$

Hereafter we show that the process $u_k(t)$ defined in (ii) satisfies this condition. To prove $u_k \in \mathbb{D}^{1,2}(L^2([0,T]; \mathbb{R}^d))$, we need to compute its Malliavin derivative $D_s u_k(t)$. Using the product rule

$$D_s u_k(t) = \sum_{j=1}^{m} \left[D_s \left((\gamma_{X_T}^{-1})_{jk}\right) D_t X_T^j + (\gamma_{X_T}^{-1})_{jk} D_s D_t X_T^j\right].$$

Since $X_T \in \mathbb{D}^{2,2}$, the second Malliavin derivative $D_s D_t X_T^j$ exists and belongs to $L^2([0,T]^2 \times \Omega; \mathbb{R}^d)$. For $(\gamma_{X_T}^{-1})_{jk}$, we use the formula for the derivative of a matrix inverse

$$D_s(\gamma_{X_T}^{-1})_{jk} = -\sum_{p,q=1}^{m} (\gamma_{X_T}^{-1})_{jp}(D_s(\gamma_{X_T})_{pq})(\gamma_{X_T}^{-1})_{qk},$$

where

$$D_s(\gamma_{X_T})_{pq} = \int_0^T \left[D_s D_t X_T^p \cdot D_t X_T^q + D_t X_T^p \cdot D_s D_t X_T^q \right] dt.$$

Given the smoothness assumptions and $\gamma_{X_T} \geq \lambda I_m$, we have $\|\gamma_{X_T}^{-1}\| \leq \lambda^{-1}$, a constant. Thus, $|D_s(\gamma_{X_T}^{-1})_{jk}| \leq C \sum_{p,q} |D_s(\gamma_{X_T})_{pq}|$, and since $X_T \in \mathbb{D}^{2,2}$, $D_s(\gamma_{X_T})_{pq} \in L^2(\Omega \times [0,T])$, implying $D_s(\gamma_{X_T}^{-1})_{jk} \in L^2(\Omega \times [0,T])$.

Thus, $u_k(t)$ is Malliavin differentiable, with $D_s u_k(t)$ expressed in terms of known quantities.

We need to verify the following two conditions

a) $\mathbb{E}\left[\int_0^T |u_k(t)|^2\, dt\right] < \infty$,

b) $\mathbb{E}\left[\int_0^T \int_0^T |D_s u_k(t)|^2\, ds\, dt\right] < \infty$.

For a), let us begin from the Cauchy-Schwarz inequality

$$\left| \sum_{j=1}^{m} (\gamma_{X_T}^{-1})_{jk} D_t X_T^j \right|^2 \leq \left\| \gamma_{X_T}^{-1} \right\|^2 \sum_{j=1}^{m} |D_t X_T^j|^2,$$

where $\|\gamma_{X_T}^{-1}\|$ is a matrix norm compatible with the vector 2-norm. This yields

$$\int_0^T |u_k(t)|^2\, dt \leq \|\gamma_{X_T}^{-1}\|^2 \sum_{j=1}^{m} \int_0^T |D_t X_T^j|^2\, dt = \left\| \gamma_{X_T}^{-1} \right\|^2 \operatorname{Tr}(\gamma_{X_T}).$$

We notice that $\mathbb{E}[\operatorname{Tr}\gamma_{X_T}] = \sum_{j=1}^{m} \mathbb{E}\left[\int_0^T |D_t X_T^j|^2\, dt\right] < \infty$, because $X_T \in \mathbb{D}^{1,2}$. With uniform ellipticity $\gamma_{X_T} \geq \lambda I_m$, we have $\|\gamma_{X_T}^{-1}\| \leq \lambda^{-1}$, so

$$\mathbb{E}\left[\int_0^T |u_k(t)|^2\, dt\right] \leq \lambda^{-2} \sum_{j=1}^{m} \mathbb{E}\left[\int_0^T |D_t X_T^j|^2\, dt\right] < \infty,$$

since $X_T \in \mathbb{D}^{2,2}$ implies $D_t X_T^j \in L^2([0,T] \times \Omega; \mathbb{R}^d)$.

For b), we first notice that

$$\begin{aligned} |D_s u_k(t)|^2 &= \left| \sum_{j=1}^{m} \left[D_s(\gamma_{X_T}^{-1})_{jk} D_t X_T^j + (\gamma_{X_T}^{-1})_{jk} D_s D_t X_T^j \right] \right|^2, \\ &\leq 2 \left| \sum_{j=1}^{m} D_s(\gamma_{X_T}^{-1})_{jk} D_t X_T^j \right|^2 + 2 \left| \sum_{j=1}^{m} (\gamma_{X_T}^{-1})_{jk} D_s D_t X_T^j \right|^2. \end{aligned}$$

Each term at the right hand side of the previous equation can be bounded in a similar way as

$$\left|\sum_{j=1}^{m} D_s(\gamma_{X_T}^{-1})_{jk} D_t X_T^j\right|^2 \leq m \sum_{j=1}^{m} |D_s(\gamma_{X_T}^{-1})_{jk}|^2 |D_t X_T^j|^2,$$

and

$$\left|\sum_{j=1}^{m} (\gamma_{X_T}^{-1})_{jk} D_s D_t X_T^j\right|^2 \leq m \sum_{j=1}^{m} |(\gamma_{X_T}^{-1})_{jk}|^2 |D_s D_t X_T^j|^2.$$

Integrating over $[0,T] \times [0,T]$

$$\int_0^T \int_0^T |D_s u_k(t)|^2 \, ds \, dt \leq 2m \sum_{j=1}^{m} \int_0^T \int_0^T \Big[|D_s(\gamma_{X_T}^{-1})_{jk}|^2 |D_t X_T^j|^2 + |(\gamma_{X_T}^{-1})_{jk}|^2 |D_s D_t X_T^j|^2 \Big] \, ds \, dt.$$

Taking expectations and applying Fubini's theorem, we get

$$\mathbb{E}\left[\int_0^T \int_0^T |D_s u_k(t)|^2 \, ds \, dt\right] \leq 2m \sum_{j=1}^{m} \mathbb{E}\left[\int_0^T \int_0^T \Big[|D_s(\gamma_{X_T}^{-1})_{jk}|^2 |D_t X_T^j|^2 + |(\gamma_{X_T}^{-1})_{jk}|^2 |D_s D_t X_T^j|^2 \Big] \, ds \, dt\right].$$

For the first term, since $|D_s(\gamma_{X_T}^{-1})_{jk}| \leq C\|D_s \gamma_{X_T}\|$ and $\|\gamma_{X_T}^{-1}\| \leq \lambda^{-1}$, we have $|D_s(\gamma_{X_T}^{-1})_{jk}|^2 \leq C\|D_s \gamma_{X_T}\|^2$. Thus,

$$\mathbb{E}\left[\int_0^T \int_0^T |D_s(\gamma_{X_T}^{-1})_{jk}|^2 |D_t X_T^j|^2 \, ds \, dt\right] \leq C\mathbb{E}\left[\int_0^T \int_0^T \|D_s \gamma_{X_T}\|^2 |D_t X_T^j|^2 \, ds \, dt\right].$$

Since $\gamma_{X_T} \in \mathbb{D}^{1,2}$, $\|D_s \gamma_{X_T}\|^2 \in L^2(\Omega \times [0,T])$, and $|D_t X_T^j|^2 \in L^2(\Omega \times [0,T])$, their product is in $L^1(\Omega \times [0,T])$, so the expectation is finite. For the second term, since $|(\gamma_{X_T}^{-1})_{jk}|^2 \leq C$,

$$\mathbb{E}\left[\int_0^T \int_0^T |(\gamma_{X_T}^{-1})_{jk}|^2 |D_s D_t X_T^j|^2 \, ds \, dt\right] \leq C\mathbb{E}\left[\int_0^T \int_0^T |D_s D_t X_T^j|^2 \, ds \, dt\right] < \infty,$$

because $X_T \in \mathbb{D}^{2,2}$. Hence, the second condition holds. Thus, $u_k \in \mathbb{D}^{1,2}(L^2([0,T];\mathbb{R}^d))$, and consequently, $u_k \in \mathrm{Dom}(\delta)$, with $\delta(u_k) \in L^2(\Omega)$.

To show $\delta(u_k) \in L^p(\Omega)$ for some $p > 1$, we first establish that $u_k \in \mathbb{D}^{1,p}(H)$ where $H = L^2([0,T];\mathbb{R}^m)$. Choose $p = 2$, which satisfies $p > 1$. We now verify that $u_k \in \mathbb{D}^{1,p}(H)$.

For $\mathbb{E}\left[\left(\int_0^T |u_k(t)|^2 dt\right)^{p/2}\right]$ with $p = 2$, we have the bound $\int_0^T |u_k(t)|^2 dt \leq \|\gamma_{X_T}^{-1}\|^2 \operatorname{Tr}(\gamma_{X_T}) \leq \lambda^{-2} \operatorname{Tr}(\gamma_{X_T})$. Thus,

$$\mathbb{E}\left[\left(\int_0^T |u_k(t)|^2 dt\right)^{2/2}\right] = \mathbb{E}\left[\int_0^T |u_k(t)|^2 dt\right] \leq \lambda^{-2}\mathbb{E}\left[\operatorname{Tr}(\gamma_{X_T})\right] < \infty,$$

which is finite as shown earlier.

For $\mathbb{E}\left[\left(\int_0^T\int_0^T |D_s u_k(t)|^2 dsdt\right)^{p/2}\right]$ with $p=2$, we use the earlier bound

$$\int_0^T\int_0^T |D_s u_k(t)|^2 dsdt \leq 2m\sum_{j=1}^m \int_0^T\int_0^T \left[|D_s(\gamma_{X_T}^{-1})_{jk}|^2|D_t X_T^j|^2 + |(\gamma_{X_T}^{-1})_{jk}|^2|D_s D_t X_T^j|^2\right] ds\, dt,$$

so

$$\mathbb{E}\left[\left(\int_0^T\int_0^T |D_s u_k(t)|^2 dsdt\right)^{2/2}\right] = \mathbb{E}\left[\int_0^T\int_0^T |D_s u_k(t)|^2 dsdt\right] < \infty,$$

as established above.

This implies $u_k \in \mathbb{D}^{1,2}(H)$. The Skorokhod integral $\delta : \mathbb{D}^{1,2}(\mathcal{H}) \to L^2(\Omega)$ is continuous, and therefore $\delta(u_k) \in L^2(\Omega)$, which satisfies $\delta(u_k) \in L^p(\Omega)$ for $p = 2 > 1$. This completes the proof. □

In the next theorem we establish the L^2-continuity of the covering vector field associated with a stochastic process $(X_t)_{t\in[0,T]}$, adapted to a Brownian filtration. The process is assumed to have a terminal value X_T in the Malliavin space $\mathbb{D}^{1,2}$, with an invertible Malliavin covariance matrix γ_{X_T}. The covering vector field $u_k(t)$ is defined using $\gamma_{X_T}^{-1}$ and the Malliavin derivatives of X_T. Additionally, X_t satisfies an SDE with $\mathcal{C}^2$ coefficients b and σ having bounded derivatives up to second order. Under conditions of uniform ellipticity of γ_{X_T} and, the theorem proves that $t \mapsto u_k(t)$ is continuous into $L^2(\Omega)$, highlighting the regularity of the vector field in the stochastic setting.

Theorem 4 (L^2–continuity of the covering vector field) *Let $X = (X_t)_{t\in[0,T]}$ be the unique solution of*

$$\mathrm{d}X_t \;=\; b(t,X_t)\,\mathrm{d}t \;+\; \sigma(t,X_t)\,\mathrm{d}B_t, \quad X_0 = x_0,$$

where $b,\sigma \in \mathcal{C}^2\big([0,T]\times\mathbb{R}^m\big)$ have all first and second derivatives bounded, and σ satisfies the linear growth bound $\|\sigma(t,x)\| \leq C(1+\|x\|)$. Suppose additionally that

(i) $X_T \in \mathbb{D}^{2,2}$ and its Malliavin covariance $\gamma_{X_T} = \big(\int_0^T D_t X_T^i \cdot D_t X_T^j\, dt\big)_{i,j}$ is a.s. invertible with $\gamma_{X_T} \geq \lambda I_m$ for some $\lambda > 0$;

(ii) for each $t \in [0,T]$, the mapping $x \mapsto X_t^x$ is a random diffeomorphism of class $\mathcal{C}^2$ almost surely, and all derivatives up to order 2 are uniformly bounded in $L^2(\Omega)$.

Define for each $k = 1,\dots,m$ the covering field

$$u_k(t) \;=\; \sum_{j=1}^m (\gamma_{X_T}^{-1})_{jk}\, D_t X_T^j.$$

Then $t \mapsto u_k(t)$ is continuous as a map $[0,T] \to L^2(\Omega)$; equivalently

$$\lim_{s\to t} \mathbb{E}\big[|u_k(s) - u_k(t)|^2\big] \;=\; 0 \quad \forall\, t \in [0,T].$$

Proof and iii) combining these results to prove the continuity of $u_k(t)$ in $L^2(\Omega)$.

Let us first establish continuity of $D_t X_T$. Since X_t satisfies the SDE

$$dX_t = b(t,X_t)\,dt + \sigma(t,X_t)\,dB_t, \quad X_0 = x_0,$$

and the coefficients b and σ are $\mathcal{C}^2$ with bounded derivatives up to second order, it follows that $X_t \in \mathbb{D}^{2,2}$, meaning X_t is twice Malliavin differentiable with square-integrable derivatives. The Malliavin derivative $D_t X_T$ can be expressed using the first variation process $Y_t = \partial X_t/\partial x$, which satisfies

$$dY_t = \partial_x b(t, X_t) Y_t \, dt + \sum_{l=1}^{d} \partial_x \sigma_l(t, X_t) Y_t \, dB_t^l, \quad Y_0 = I_m,$$

where $\partial_x b$ and $\partial_x \sigma_l$ are the Jacobians with respect to x, and σ_l denotes the l-th column of σ. The process Y_t is well-defined, with a continuous version in t, and invertible almost surely because the coefficients' derivatives are bounded. For $t \leq T$, the Malliavin derivative is

$$D_t X_T = Y_T Y_t^{-1} \sigma(t, X_t) \mathbf{1}_{\{t \leq T\}}.$$

To prove continuity, we need to show that

$$\lim_{s \to t} \mathbb{E}\left[|D_s X_T - D_t X_T|^2\right] = 0.$$

To this end, we compute the difference

$$D_s X_T - D_t X_T = Y_T Y_s^{-1} \sigma(s, X_s) \mathbf{1}_{\{s \leq T\}} - Y_T Y_t^{-1} \sigma(t, X_t) \mathbf{1}_{\{t \leq T\}}.$$

For $s, t \leq T$, this simplifies to

$$D_s X_T - D_t X_T = Y_T \left(Y_s^{-1} \sigma(s, X_s) - Y_t^{-1} \sigma(t, X_t) \right).$$

Therefore

$$\mathbb{E}\left[|D_s X_T - D_t X_T|^2\right] \leq \mathbb{E}\left[\|Y_T\|^2 \left| Y_s^{-1} \sigma(s, X_s) - Y_t^{-1} \sigma(t, X_t) \right|^2 \right].$$

Since b and σ have bounded derivatives up to second-order, Y_t has finite moments of all orders (via standard SDE moment estimates), so $\mathbb{E}[\|Y_T\|^2] < \infty$. Also, the term $Y_t^{-1}\sigma(t, X_t)$ is continuous in t because X_t is continuous (as an SDE solution with Lipschitz coefficients), $\sigma(t, x)$ is $\mathcal{C}^2$, hence Lipschitz in t and x due to bounded first derivatives, Y_t^{-1} is continuous since Y_t is a continuous, invertible process (invertibility follows from the boundedness of coefficients' derivatives). Thus, $Y_s^{-1}\sigma(s, X_s) - Y_t^{-1}\sigma(t, X_t) \to 0$ almost surely as $s \to t$. At this point we notice that

$$\|Y_s^{-1} \sigma(s, X_s)\| \leq C(1 + \|X_s\|) \sup_{u \in [0,T]} \|Y_u^{-1}\|,$$

and define the random variable $Z := C^2(1 + \sup_{u\in[0,T]} \|X_u\|^2) \sup_{u\in[0,T]} \|Y_u^{-1}\|^2$, which dominates $\|Y_s^{-1}\sigma(s, X_s) - Y_t^{-1}\sigma(t, X_t)\|^2$ and satisfies $\mathbb{E}Z < \infty$ due to the moment bounds on X_t and Y_t^{-1}. Therefore,

$$\mathbb{E}\left[\left| Y_s^{-1} \sigma(s, X_s) - Y_t^{-1} \sigma(t, X_t) \right|^2\right] \to 0 \quad \text{as} \quad s \to t.$$

using dominated convergence. Hence,

$$\mathbb{E}\left[|D_s X_T - D_t X_T|^2\right] \to 0 \quad \text{as} \quad s \to t,$$

establishing the continuity of $D_t X_T$ in $L^2(\Omega)$.

Next we analyse the properties of the inverse Malliavin covariance matrix

$$(\gamma_{X_T})_{ij} = \int_0^T D_t X_T^i \cdot D_t X_T^j \, dt,$$

By the uniform ellipticity assumption, $\gamma_{X_T} \geq \lambda I$ almost surely, so $\gamma_{X_T}^{-1}$ exists and

$$\|\gamma_{X_T}^{-1}\| \leq \lambda^{-1},$$

implying $(\gamma_{X_T}^{-1})_{jk} \in L^\infty(\Omega)$. The additional condition $\mathbb{E}[\|\gamma_{X_T}^{-1}\|^2] < \infty$ ensures finite second moments, providing stronger integrability. Since γ_{X_T} does not depend on t, neither does $\gamma_{X_T}^{-1}$.

We now prove continuity of the covering vector field (8). We need to show

$$\mathbb{E}\left[|u_k(s) - u_k(t)|^2\right] \to 0 \quad \text{as} \quad s \to t.$$

Computing the difference

$$u_k(t) - u_k(s) = \sum_{j=1}^{m} (\gamma_{X_T}^{-1})_{jk} (D_t X_T^j - D_s X_T^j).$$

and taking the expectation yields

$$\mathbb{E}\left[|u_k(t) - u_k(s)|^2\right] = \mathbb{E}\left[\left|\sum_{j=1}^{m} (\gamma_{X_T}^{-1})_{jk} (D_t X_T^j - D_s X_T^j)\right|^2\right].$$

Using Cauchy-Schwarz inequality

$$\left|\sum_{j=1}^{m} a_j b_j\right|^2 \leq \left(\sum_{j=1}^{m} |a_j|^2\right) \left(\sum_{j=1}^{m} |b_j|^2\right),$$

with $a_j = (\gamma_{X_T}^{-1})_{jk}$ and $b_j = D_t X_T^j - D_s X_T^j$, we obtain

$$\mathbb{E}\left[|u_k(t) - u_k(s)|^2\right] \leq \mathbb{E}\left[\left(\sum_{j=1}^{m} |(\gamma_{X_T}^{-1})_{jk}|^2\right) \left(\sum_{j=1}^{m} |D_t X_T^j - D_s X_T^j|^2\right)\right].$$

Since $\gamma_{X_T}^{-1}$ is bounded

$$\sum_{j=1}^{m} |(\gamma_{X_T}^{-1})_{jk}|^2 = \|\gamma_{X_T}^{-1} e_k\|^2 \leq \lambda^{-2},$$

and therefore

$$\mathbb{E}\left[|u_k(t) - u_k(s)|^2\right] \leq \lambda^{-2} \mathbb{E}\left[\sum_{j=1}^{m} |D_t X_T^j - D_s X_T^j|^2\right].$$

To apply dominated convergence, note that

$$|u_k(t) - u_k(s)|^2 \leq \lambda^{-2} \sum_{j=1}^{m} (|D_t X_T^j|^2 + |D_s X_T^j|^2),$$

and since $\mathbb{E}\left[\sum_{j=1}^{m} |D_u X_T^j|^2\right] \leq \operatorname{Tr} \gamma_{X_T} < \infty$ for each u, the expectation is bounded by $2\lambda^{-2} \operatorname{Tr} \gamma_{X_T} < \infty$, providing an integrable dominating random variable independent of s. Now recall that for each j

$$\mathbb{E}\left[|D_t X_T^j - D_s X_T^j|^2\right] \to 0 \quad \text{as} \quad s \to t,$$

hence

$$\mathbb{E}\left[\sum_{j=1}^{m} |D_t X_T^j - D_s X_T^j|^2\right] \to 0,$$

i.e.,

$$\mathbb{E}\left[|u_k(t) - u_k(s)|^2\right] \to 0 \quad \text{as} \quad s \to t.$$

This establishes the continuity of $u_k(t)$ in $L^2(\Omega)$, completing the proof.

□

Next we present a stability theorem for covering vector fields associated with stochastic processes under controlled perturbations. The theorem examines two sequences of stochastic processes, $(X_t^n)_{t\in[0,T]}$ and $(X_t)_{t\in[0,T]}$, described by stochastic differential equations (SDEs) with coefficients b_n, σ_n and b, σ, respectively, where these coefficients converge uniformly and satisfy smoothness and boundedness conditions. Both processes are adapted to a Brownian filtration, belong to the Malliavin Sobolev space $\mathbb{D}^{1,2}$, and possess invertible Malliavin covariance matrices that are uniformly elliptic. The theorem proves that the covering vector fields $u_k^n(t)$, defined using the inverse Malliavin covariance matrices and Malliavin derivatives, converge to $u_k(t)$ in the $L^2([0,T] \times \Omega)$ norm as $n \to \infty$. This result highlights the robustness of the covering vector field construction against perturbations in the SDE dynamics.

Theorem 5 (Stability under perturbations) *Let X^n and X be the unique strong solutions on $[0,T]$ of*

$$\mathrm{d}X_t^n = b_n(t, X_t^n)\,\mathrm{d}t + \sigma_n(t, X_t^n)\,\mathrm{d}B_t, \qquad \mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t,$$

driven by the same d–dimensional Brownian motion B. Assume:

(i) $b_n, b, \sigma_n, \sigma \in C^2([0,T] \times \mathbb{R}^m)$ with all first and second derivatives bounded uniformly in n, and $(b_n, \sigma_n) \to (b, \sigma)$ locally uniformly.

(ii) For each n and for X, the terminal value lies in $\mathbb{D}^{1,2}$, and their Malliavin covariance matrices

$$\gamma^n \;=\; \int_0^T D_t X_T^n \,(D_t X_T^n)^\top \,dt, \quad \gamma \;=\; \int_0^T D_t X_T \,(D_t X_T)^\top \,dt$$

satisfy $\gamma^n, \gamma \geq \lambda I_m > 0$ a.s. for some $\lambda > 0$, together with $\sup_n \mathbb{E}[\|(\gamma^n)^{-1}\|^2] < \infty$ and $\mathbb{E}[\|\gamma^{-1}\|^2] < \infty$.

(iii) The first variation processes $Y_t^n = \partial_x X_t^n$ satisfy $\sup_{n,t} \mathbb{E}[\|Y_t^n\|^2 + \|(Y_t^n)^{-1}\|^2] < \infty$, and $\sup_{n,t} \mathbb{E}[\|X_t^n\|^2] < \infty$.

Define, for each $k = 1, \ldots, m$,

$$u_k^n(t) \;=\; \sum_{j=1}^m (\gamma^n)^{-1}_{jk}\, D_t X_T^{n,j}, \quad u_k(t) \;=\; \sum_{j=1}^m \gamma^{-1}_{jk}\, D_t X_T^j.$$

Then for every k,

$$\lim_{n\to\infty} \mathbb{E}\Big[\int_0^T \big|u_k^n(t) - u_k(t)\big|^2 \,dt\Big] \;=\; 0.$$

Proof To prove convergence of X_T^n to X_T in $L^2(\Omega)$, we aim to show $\mathbb{E}[|X_T^n - X_T|^2] \to 0$. Define the difference between the processes

$$X_t^n - X_t = \int_0^t \big[b_n(s, X_s^n) - b(s, X_s)\big]\,ds + \int_0^t \big[\sigma_n(s, X_s^n) - \sigma(s, X_s)\big]\,dW_s.$$

We bound the supremum over $t \in [0,T]$

$$\mathbb{E}\left[\sup_{t\in[0,T]} |X_t^n - X_t|^2\right].$$

Using the inequality $(a+b)^2 \leq 2a^2 + 2b^2$, we split the supremum as

$$\begin{aligned}\sup_{t\in[0,T]} |X_t^n - X_t|^2 \leq 2\left(\sup_{t\in[0,T]} \left|\int_0^t \big[b_n(s,X_s^n) - b(s,X_s)\big]\, ds\right|\right)^2 \\ + 2\left(\sup_{t\in[0,T]} \left|\int_0^t \big[\sigma_n(s,X_s^n) - \sigma(s,X_s)\big]\, dW_s\right|\right)^2.\end{aligned}$$

Taking expectations

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in[0,T]} |X_t^n - X_t|^2\right] \leq 2\mathbb{E}\left[\left(\sup_{t\in[0,T]} \left|\int_0^t \big[b_n(s,X_s^n) - b(s,X_s)\big]\, ds\right|\right)^2\right] \\ + 2\mathbb{E}\left[\sup_{t\in[0,T]} \left|\int_0^t \big[\sigma_n(s,X_s^n) - \sigma(s,X_s)\big]\, dW_s\right|^2\right].\end{aligned}$$

For the deterministic integral (drift term) we have

$$\sup_{t\in[0,T]} \left|\int_0^t \big[b_n(s,X_s^n) - b(s,X_s)\big]\, ds\right| \leq \int_0^T \big|b_n(s,X_s^n) - b(s,X_s)\big|\, ds.$$

Applying the Cauchy-Schwarz inequality

$$\begin{aligned}\left(\int_0^T \big|b_n(s,X_s^n) - b(s,X_s)\big|\, ds\right)^2 &\leq \left(\int_0^T \big|b_n(s,X_s^n) - b(s,X_s)\big|^2\, ds\right)\left(\int_0^T 1^2 ds\right) \\ &= T\int_0^T \big|b_n(s,X_s^n) - b(s,X_s)\big|^2\, ds.\end{aligned}$$

Thus

$$2\mathbb{E}\left[\left(\sup_{t\in[0,T]} \left|\int_0^t \big[b_n(s,X_s^n) - b(s,X_s)\big]\, ds\right|\right)^2\right] \leq 2T\mathbb{E}\left[\int_0^T \big|b_n(s,X_s^n) - b(s,X_s)\big|^2\, ds\right].$$

For the stochastic integral (diffusion term), we apply the Burkholder-Davis-Gundy inequality

$$\mathbb{E}\left[\sup_{t\in[0,T]} \left|\int_0^t \big[\sigma_n(s,X_s^n) - \sigma(s,X_s)\big]\, dW_s\right|^2\right] \leq C_{\text{BDG}}\mathbb{E}\left[\int_0^T \big\|\sigma_n(s,X_s^n) - \sigma(s,X_s)\big\|^2\, ds\right],$$

where $C_{\text{BDG}} > 0$ is a universal constant whose exact value is irrelevant to the proof, so

$$2\mathbb{E}\left[\sup_{t\in[0,T]} \left|\int_0^t \big[\sigma_n(s,X_s^n) - \sigma(s,X_s)\big]\, dW_s\right|^2\right] \leq 2C_{\text{BDG}}\mathbb{E}\left[\int_0^T \big\|\sigma_n(s,X_s^n) - \sigma(s,X_s)\big\|^2\, ds\right].$$

Combining these bounds

$$\mathbb{E}\left[\sup_{t\in[0,T]} |X_t^n - X_t|^2\right] \leq 2T\mathbb{E}\left[\int_0^T \big|b_n(s,X_s^n) - b(s,X_s)\big|^2\, ds\right]$$

$$+ 2C_{\text{BDG}}\mathbb{E}\left[\int_0^T \left\|\sigma_n(s,X_s^n) - \sigma(s,X_s)\right\|^2 ds\right].$$

Decompose the differences

$$b_n(s,X_s^n) - b(s,X_s) = \left[b_n(s,X_s^n) - b(s,X_s^n)\right] + \left[b(s,X_s^n) - b(s,X_s)\right],$$
$$\sigma_n(s,X_s^n) - \sigma(s,X_s) = \left[\sigma_n(s,X_s^n) - \sigma(s,X_s^n)\right] + \left[\sigma(s,X_s^n) - \sigma(s,X_s)\right].$$

Since $b_n \to b$ and $\sigma_n \to \sigma$ uniformly on compact sets, and b and σ are Lipschitz continuous (due to bounded first derivatives), we have

$$\left|b_n(s,X_s^n) - b(s,X_s^n)\right| \le \sup_{t,x} |b_n(t,x) - b(t,x)| \to 0,$$
$$\left|b(s,X_s^n) - b(s,X_s)\right| \le C|X_s^n - X_s|,$$
$$\left\|\sigma_n(s,X_s^n) - \sigma(s,X_s^n)\right\| \le \sup_{t,x} \|\sigma_n(t,x) - \sigma(t,x)\| \to 0,$$
$$\left\|\sigma(s,X_s^n) - \sigma(s,X_s)\right\| \le C|X_s^n - X_s|.$$

Bound the squared terms

$$\begin{aligned}\left|b_n(s,X_s^n) - b(s,X_s)\right|^2 &\le 2\left|b_n(s,X_s^n) - b(s,X_s^n)\right|^2 + 2\left|b(s,X_s^n) - b(s,X_s)\right|^2 \\ &\le 2\sup_{t,x}|b_n(t,x) - b(t,x)|^2 + 2C^2|X_s^n - X_s|^2,\end{aligned}$$

$$\left\|\sigma_n(s,X_s^n) - \sigma(s,X_s)\right\|^2 \le 2\sup_{t,x}\|\sigma_n(t,x) - \sigma(t,x)\|^2 + 2C^2|X_s^n - X_s|^2.$$

Substitute into the bound

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in[0,T]}|X_t^n - X_t|^2\right] \le 2T\mathbb{E}&\left[\int_0^T \left(2\sup_{t,x}|b_n(t,x) - b(t,x)|^2 + 2C^2|X_s^n - X_s|^2\right) ds\right] \\ &+ 2C_{\text{BDG}}\mathbb{E}\left[\int_0^T \left(2\sup_{t,x}\|\sigma_n(t,x) - \sigma(t,x)\|^2 + 2C^2|X_s^n - X_s|^2\right) ds\right],\end{aligned}$$

and simplify to obtain

$$\begin{aligned}&= 4T^2\sup_{t,x}|b_n(t,x) - b(t,x)|^2 + 4TC^2\mathbb{E}\left[\int_0^T |X_s^n - X_s|^2 ds\right] \\ &\quad + 4C_{\text{BDG}}T\sup_{t,x}\|\sigma_n(t,x) - \sigma(t,x)\|^2 + 4C_{\text{BDG}}C^2\mathbb{E}\left[\int_0^T |X_s^n - X_s|^2 ds\right].\end{aligned}$$

Let $A_n = 4T^2\sup_{t,x}|b_n(t,x) - b(t,x)|^2 + 4C_{\text{BDG}}T\sup_{t,x}\|\sigma_n(t,x) - \sigma(t,x)\|^2$ and $K = 4TC^2 + 4C_{\text{BDG}}C^2$, so

$$\mathbb{E}\left[\sup_{t\in[0,T]}|X_t^n - X_t|^2\right] \le A_n + K\int_0^T \mathbb{E}\left[\sup_{u\in[0,s]}|X_u^n - X_u|^2\right] ds.$$

Applying Gronwall's inequality

$$\phi(t) = \mathbb{E}\left[\sup_{u\in[0,t]}|X_u^n - X_u|^2\right] \qquad \Rightarrow \phi(T) \le A_n + K\int_0^T \phi(s)ds,$$

we obtain

$$\mathbb{E}\left[\sup_{t\in[0,T]}|X_t^n - X_t|^2\right] \le A_n e^{KT}.$$

Since $b_n \to b$ and $\sigma_n \to \sigma$ uniformly, $A_n \to 0$ as $n \to \infty$, we obtain

$$\mathbb{E}\left[\sup_{t\in[0,T]} |X_t^n - X_t|^2\right] \to 0.$$

Thus,

$$\mathbb{E}[|X_T^n - X_T|^2] \le \mathbb{E}\left[\sup_{t\in[0,T]} |X_t^n - X_t|^2\right] \to 0,$$

proving $X_T^n \to X_T$ in $L^2(\Omega)$.

Next we prove convergence of Malliavin derivatives. For the SDE $dX_t^n = b_n(t, X_t^n)\,dt + \sigma_n(t, X_t^n)\,dB_t$, the Malliavin derivative is $D_t X_T^n = Y_T^n (Y_t^n)^{-1} \sigma_n(t, X_t^n)\mathbf{1}_{\{t\le T\}}$, where $Y_t^n = \partial X_t^n/\partial x$ satisfies

$$dY_t^n = \partial_x b_n(t, X_t^n) Y_t^n\,dt + \sum_{l=1}^d \partial_x \sigma_{n,l}(t, X_t^n) Y_t^n\,dB_t^l, \quad Y_0^n = I_m.$$

Similarly, $D_t X_T = Y_T Y_t^{-1} \sigma(t, X_t)\mathbf{1}_{\{t\le T\}}$, with

$$dY_t = \partial_x b(t, X_t) Y_t\,dt + \sum_{l=1}^d \partial_x \sigma_l(t, X_t) Y_t\,dB_t^l, \quad Y_0 = I_m.$$

Since b_n, σ_n are $\mathcal{C}^2$ with bounded derivatives up to second order, $\partial_x b_n \to \partial_x b$ and $\partial_x \sigma_n \to \partial_x \sigma$ uniformly, and all derivatives are bounded uniformly in n

$$\mathbb{E}\left[\sup_{t\in[0,T]} \|Y_t^n - Y_t\|^2\right] \to 0, \quad \mathbb{E}\left[\sup_{t\in[0,T]} \|(Y_t^n)^{-1} - Y_t^{-1}\|^2\right] \to 0.$$

We have

$$\begin{aligned} D_t X_T^n - D_t X_T =& Y_T^n \left[(Y_t^n)^{-1}\sigma_n(t, X_t^n) - Y_t^{-1}\sigma(t, X_t)\right]\mathbf{1}_{\{t\le T\}} \\ &+ (Y_T^n - Y_T) Y_t^{-1} \sigma(t, X_t)\mathbf{1}_{\{t\le T\}}, \end{aligned}$$

and

$$\begin{aligned} |D_t X_T^n - D_t X_T|^2 \le& 2\|Y_T^n\|^2 \left\|(Y_t^n)^{-1}\sigma_n(t, X_t^n) - Y_t^{-1}\sigma(t, X_t)\right\|^2 \mathbf{1}_{\{t\le T\}} \\ &+ 2\|(Y_T^n - Y_T)\|^2 \|Y_t^{-1}\sigma(t, X_t)\|^2 \mathbf{1}_{\{t\le T\}}. \end{aligned}$$

Integrate and take expectations

$$\mathbb{E}\left[\int_0^T |D_t X_T^n - D_t X_T|^2\,dt\right] \le I_1^n + I_2^n,$$

where

$$\begin{aligned} I_1^n &= 2\mathbb{E}\left[\int_0^T \|Y_T^n\|^2 \left\|(Y_t^n)^{-1}\sigma_n(t, X_t^n) - Y_t^{-1}\sigma(t, X_t)\right\|^2\,dt\right], \\ I_2^n &= 2\mathbb{E}\left[\int_0^T \|Y_T^n - Y_T\|^2 \|Y_t^{-1}\sigma(t, X_t)\|^2\,dt\right]. \end{aligned}$$

Using Cauchy-Schwarz, control I_2^n

$$I_2^n \le 2\left(\mathbb{E}[\|Y_T^n - Y_T\|^4]\right)^{1/2} \left(\mathbb{E}\left[\left(\int_0^T \|Y_t^{-1}\sigma(t, X_t)\|^2\,dt\right)^2\right]\right)^{1/2}.$$

Since $\sup_n \mathbb{E}[\|Y_T^n\|^4] < \infty$, $\mathbb{E}[\|Y_T^n - Y_T\|^4] \to 0$, and the second factor is finite, $I_2^n \to 0$.

Using Cauchy-Schwarz, control I_1^n

$$I_1^n \le 2\left(\mathbb{E}[\|Y_T^n\|^4]\right)^{1/2}\left(\mathbb{E}\left[\left(\int_0^T \|(Y_t^n)^{-1}\sigma_n(t,X_t^n) - Y_t^{-1}\sigma(t,X_t)\|^2\,dt\right)^2\right]\right)^{1/2}.$$

The first factor is bounded, and the integrand converges to zero, so $I_1^n \to 0$. Thus

$$\mathbb{E}\left[\int_0^T |D_tX_T^n - D_tX_T|^2\,dt\right] \to 0.$$

At this point we prove convergence of inverse Malliavin covariance matrices. Recall that

$$(\gamma_{X_T^n})_{ij} = \int_0^T D_tX_T^{n,i}\cdot D_tX_T^{n,j}\,dt.$$

$$\gamma_{ij}^n - \gamma_{ij} = \int_0^T (D_tX_T^{n,i}D_tX_T^{n,j} - D_tX_T^iD_tX_T^j)\,dt.$$

By Cauchy-Schwarz

$$\mathbb{E}[|\gamma_{ij}^n - \gamma_{ij}|^2] \le T\mathbb{E}\left[\int_0^T |D_tX_T^n - D_tX_T|^2\,dt\right] \to 0.$$

Thus, $\gamma_{X_T^n} \to \gamma_{X_T}$ in $L^2(\Omega)$. Given $\gamma^n, \gamma \ge \lambda I_m$,

$$\|(\gamma^n)^{-1} - \gamma^{-1}\| \le \lambda^{-2}\|\gamma^n - \gamma\|,$$

so:

$$\mathbb{E}[\|(\gamma_{X_T^n})^{-1} - \gamma_{X_T}^{-1}\|^2] \le \lambda^{-4}\mathbb{E}[\|\gamma^n - \gamma\|^2] \to 0.$$

Finally, we prove convergence of the covering vector fields

$$u_k^n(t) - u_k(t) = \sum_{j=1}^m (\gamma_{X_T^n}^{-1})_{jk}(D_tX_T^{n,j} - D_tX_T^j)\mathbf{1}_{\{t\le T\}} + \sum_{j=1}^m \left[(\gamma_{X_T^n}^{-1})_{jk} - (\gamma_{X_T}^{-1})_{jk}\right] D_tX_T^j\mathbf{1}_{\{t\le T\}}.$$

Square and bound

$$\begin{aligned} |u_k^n(t) - u_k(t)|^2 \le\ & 2\left(\sum_{j=1}^m |(\gamma_{X_T^n}^{-1})_{jk}|^2\right)\left(\sum_{j=1}^m |D_tX_T^{n,j} - D_tX_T^j|^2\right)\mathbf{1}_{\{t\le T\}} \\ & + 2\left(\sum_{j=1}^m |(\gamma_{X_T^n}^{-1})_{jk} - (\gamma_{X_T}^{-1})_{jk}|^2\right)\left(\sum_{j=1}^m |D_tX_T^j|^2\right)\mathbf{1}_{\{t\le T\}}. \end{aligned}$$

Since

$$\sum_j |(\gamma^n)_{jk}^{-1}|^2 \le \|(\gamma^n)^{-1}\|^2 \le \lambda^{-2} \quad \text{and} \quad \sum_j |(\gamma^n)_{jk}^{-1} - \gamma_{jk}^{-1}|^2 \le \|(\gamma^n)^{-1} - \gamma^{-1}\|^2,$$

the integrand is dominated by

$$2\lambda^{-2}\sum_j |D_tX_T^{n,j} - D_tX_T^j|^2 + 2\|(\gamma^n)^{-1} - \gamma^{-1}\|^2\sum_j |D_tX_T^j|^2,$$

which converges to 0 in $L^1([0,T]\times\Omega)$ by Steps 2 and 3, and is bounded by an integrable function (since $\mathrm{Tr}(\gamma) \in L^2(\Omega)$).

Integrate and take expectations

$$\mathbb{E}\left[\int_0^T |u_k^n(t) - u_k(t)|^2 \, dt\right] \leq J_1^n + J_2^n,$$

where

$$J_1^n = 2\mathbb{E}\left[\int_0^T \left(\sum_{j=1}^m |(\gamma_{X_T^n}^{-1})_{jk}|^2\right)\left(\sum_{j=1}^m |D_t X_T^{n,j} - D_t X_T^j|^2\right) dt\right],$$

$$J_2^n = 2\mathbb{E}\left[\int_0^T \left(\sum_{j=1}^m |(\gamma_{X_T^n}^{-1})_{jk} - (\gamma_{X_T}^{-1})_{jk}|^2\right)\left(\sum_{j=1}^m |D_t X_T^j|^2\right) dt\right].$$

By the dominated convergence theorem, since the integrand is dominated by an integrable function and converges pointwise to zero

$$\mathbb{E}\left[\int_0^T |u_k^n(t) - u_k(t)|^2 \, dt\right] \to 0.$$

This completes the proof. □

4 Variation processes

Given a stochastic process X_t evolving from the initial condition x in $\mathbb{R}^m$ we define the the first variation process as $Y_t = \partial X_t / \partial x \in \mathbb{R}^{m\times m}$. As is well known, Y_t represents the sensitivity of X_t relative to small variations in the initial condition x. Differentiating the SDE (1) with respect to x yields

$$dY_t = \partial_x b(t, X_t) Y_t \, dt + \sum_{l=1}^d \partial_x \sigma^l(t, X_t) Y_t \, dB_t^l, \quad Y_0 = I_m,$$

where $\partial_x b(t, X_t) \in \mathbb{R}^{m\times m}$, with $[\partial_x b(t, X_t)]_{i,j} = \partial b^i(t, X_t)/\partial x_j$, $\partial_x \sigma^l(t, X_t) \in \mathbb{R}^{m\times m}$, with $\left[\partial_x \sigma^l(t, X_t)\right]_{i,j} = \sigma^{i,l}(t, X_t)/\partial x_j$, and I_m is the $m \times m$ identity matrix. The Malliavin derivative $D_t X_T$ is the response of X_T to a perturbation in the Brownian motion at time t. For $t \leq T$

$$D_t X_T^j = \left[Y_T Y_t^{-1} \sigma(t, X_t)\right]^j.$$

For $t > T$, $D_t X_T = 0$ (future perturbations don't affect X_T). This follows because $D_t X_s = 0$ for $s < t$, and $D_t X_t = \sigma(t, X_t)$, with the perturbation propagating via $Y_{T,t} = Y_T Y_t^{-1}$.

The second variation process, denoted $Z_t = \partial^2 X_t / \partial x^2$, is a third-order tensor in $\mathbb{R}^{m\times m\times m}$. It represents the second-order sensitivities of the state process X_t with respect to the initial condition x. Each component of Z_t, written as $Z_t^{i,p,q}$, corresponds to the second partial derivative $\partial^2 X_t^i \partial x_p \partial x_q$, where $i, p, q = 1, \dots, m$. This process evolves according to the following SDE

$$dZ_t = [\partial_{xx} b(t, X_t)(Y_t \otimes Y_t) + \partial_x b(t, X_t) Z_t] \, dt$$

$$+ \sum_{l=1}^{d} \left[\partial_{xx}\sigma^l(t, X_t)(Y_t \otimes Y_t) + \partial_x \sigma^l(t, X_t) Z_t \right] dB_t^l,$$

With the initial condition $Z_0 = 0$, the stochastic differential equation for the second variation process Z_t is characterised by a collection of terms that describe the system's dynamics. The Hessian tensor of the drift coefficient b, denoted $\partial_{xx}b(t, X_t) \in \mathbb{R}^{m\times m\times m}$, has components defined as $[\partial_{xx}b]_{i,j,k} = \partial^2 b^i(t, X_t)/(\partial x_j \partial x_k)$, representing the second derivatives of the i-th component of b with respect to the spatial variables x_j and x_k. Likewise, the Hessian tensor of the l-th column of the diffusion coefficient σ, expressed as $\partial_{xx}\sigma^l(t, X_t) \in \mathbb{R}^{m\times m\times m}$, accounts for higher-order effects in the stochastic terms. The first variation process $Y_t \in \mathbb{R}^{m\times m}$ captures the first-order sensitivities of the state process X_t with respect to the initial condition x. This leads to the tensor product $Y_t \otimes Y_t$, a fourth-order tensor in $\mathbb{R}^{m\times m\times m\times m}$, which interacts with the Hessian tensors through contraction, enriching the structure of the second-variation SDE. The Jacobian matrices $\partial_x b(t, X_t) \in \mathbb{R}^{m\times m}$ and $\partial_x \sigma^l(t, X_t) \in \mathbb{R}^{m\times m}$ of b and σ^l, respectively, provide the first-order spatial dependencies of the drift and diffusion coefficients. Driving the stochastic nature of the system, B_t^l represents the l-th component of a d-dimensional Brownian motion, introducing randomness into the evolution of Z_t. To clarify the tensor contraction, the term $\partial_{xx}b(t, X_t)(Y_t \otimes Y_t)$ for each component $Z_t^{i,p,q}$ is

$$[\partial_{xx}b(t, X_t)(Y_t \otimes Y_t)]^{i,p,q} = \sum_{j,k=1}^{m} \frac{\partial^2 b^i(t, X_t)}{\partial x_j \partial x_k} Y_t^{j,p} Y_t^{k,q}.$$

Similarly, the term $\partial_x b(t, X_t) Z_t$ is

$$[\partial_x b(t, X_t) Z_t]^{i,p,q} = \sum_{r=1}^{m} \frac{\partial b^i(t, X_t)}{\partial x_r} Z_t^{r,p,q}.$$

The diffusion terms follow the same structure with σ^l replacing b.

By the classical existence-and-uniqueness theory for Itô–SDEs, together with the fact that all first and second-order derivatives of b and σ are uniformly bounded (so that the coefficient-processes in the linear equation are globally Lipschitz and square-integrable), one obtains a unique strong solution Z_t (see, e.g., Kunita [14], Ch. 4).

5 Proof of Theorem 1

In this section we prove our main result, i.e., Theorem 1. The proof relies on the results we obtained in Section 3 and Section 4 for covering vector fields and variation processes to define the gradient of the log density for general stochastic differential equations using a Bismut-type formula. This formulation enables the computation of score functions for general nonlinear diffusion processes governed by stochastic differential equations (SDEs). In this section, we first establish several preliminary results before proceeding to the proof of our main Theorem 1.

We begin by considering the representation of the Skorokhod integral $\delta(u_k)$. To this end, we first refer to Theorem 3.2.9 in [23], which provides a decomposition of the Skorokhod integral of a random field composed with a random variable. We also refer to the seminal works [25, 26], which study generalised stochastic integrals and anticipating integrals in Malliavin calculus. The result arises in the context of substitution formulas for stochastic integrals. Consider a random field $u = \{u_t(x) : 0 \le t \le T, x \in \mathbb{R}^m\}$ with $u_t(x) \in \mathbb{R}^d$, which is square integrable and adapted for each $x \in \mathbb{R}^m$. For each x, one can define the Itô integral $\int_0^T u_t(x) \cdot dB_t$. Given an m-dimensional random variable $F : \Omega \to \mathbb{R}^m$, Theorem 3.2.9 addresses the Skorokhod integrability of the nonadapted process $u(F) = \{u_t(F), 0 \le t \le T\}$ and provides a formula for its Skorokhod integral under the following conditions:

(h1) For each $x \in \mathbb{R}^m$ and $t \in [0, T]$, $u_t(x)$ is $\mathcal{F}_t$-measurable.

(h2) There exist constants $p \ge 2$ and $\alpha > m$ such that

$$E(|u_t(x) - u_t(y)|^p) \le C_{t,K}|x - y|^\alpha,$$

for all $|x|, |y| \le K$, $K > 0$, where $\int_0^T C_{t,K} dt < \infty$. Moreover,

$$\int_0^T E(|u_t(0)|^2) dt < \infty.$$

(h3) For each (t, ω), the mapping $x \mapsto u_t(x)$ is continuously differentiable, and for each $K > 0$,

$$\int_0^T E\left(\sup_{|x| \le K} |\nabla u_t(x)|^q\right) dt < \infty,$$

where $q \ge 4$ and $q > m$.

For the reader's convenience, we recall Theorem 3.2.9 from [23] below.

Theorem 6 (Theorem 3.2.9, [23]) *For a random field $u = \{u_t(x) : 0 \le t \le T, x \in \mathbb{R}^m\}$ with $u_t(x) \in \mathbb{R}^d$, and a random variable $F : \Omega \to \mathbb{R}^m$ such that $F^i \in \mathbb{D}^{1,4}_{loc}$ for $1 \le i \le m$, assume u satisfies the conditions (h1) and (h3) for Skorokhod integrability. Then, the composition $u(F) = \{u_t(F), 0 \le t \le T\}$ belongs to $(Dom(\delta))_{loc}$, and the Skorokhod integral of $u(F)$ is given by*

$$\delta(u(F)) = \int_0^T u_t(x) \cdot dB_t \bigg|_{x=F} - \sum_{j=1}^m \int_0^T \partial_j u_t(F) \cdot D_t F^j \, dt,$$

where B_t is a d-dimensional Brownian motion, $\partial_j u_t(x) = \partial u_t(x)/\partial x_j$ is the partial derivative of $u_t(x)$ with respect to the j-th component of x, $D_t F^j$ is the Malliavin derivative of the j-th component of F, and $\int_0^T u_t(x) \cdot dB_t\big|_{x=F}$ denotes the Itô integral $\int_0^T u_t(x) \cdot dB_t$ for fixed x, evaluated at $x = F$. We note that no smoothness in the sense of Malliavin calculus is required on the process $u_t(x)$ itself, but the above conditions ensure the integrability of $u(F)$ in the

Skorokhod sense. Furthermore, the operator δ is not known to be local in Dom(δ), and thus the value of $\delta(u(F))$ may depend on the particular localising sequence used in the definition of $(Dom(\delta))_{loc}$.

5.1 Useful lemmas

In this section, we state and prove some key results that are useful for deriving the score function formula (3)-(4) for nonlinear SDEs in terms of the first and second variation processes.

Lemma 7 (SDE for the inverse first variation process) *Let Y_t be the first variation process. The inverse Y_t^{-1} satisfies the SDE*

$$dY_t^{-1} = -Y_t^{-1}\partial_x b(t, X_t)\, dt - \sum_{l=1}^{d} Y_t^{-1}\partial_x \sigma^l(t, X_t)\, dB_t^l + \sum_{l=1}^{d} Y_t^{-1}\left(\partial_x \sigma^l(t, X_t)\right)^2 dt,$$

with initial condition $Y_0^{-1} = I_m$, where $\left(\partial_x \sigma^l(t, X_t)\right)^2 = \partial_x \sigma^l(t, X_t)\partial_x \sigma^l(t, X_t)$.

Proof Since $Y_t Y_t^{-1} = I_m$ is constant, its differential is zero, i.e.,

$$d(Y_t Y_t^{-1}) = dY_t Y_t^{-1} + Y_t dY_t^{-1} + d[Y_t, Y_t^{-1}] = 0.$$

Given the first-variation SDE

$$dY_t = \partial_x b(t, X_t) Y_t\, dt + \sum_{l=1}^{d} \partial_x \sigma^l(t, X_t) Y_t\, dB_t^l, \tag{10}$$

we compute

$$\begin{aligned} dY_t Y_t^{-1} &= \left(\partial_x b(t, X_t) Y_t\, dt + \sum_{l=1}^{d} \partial_x \sigma^l(t, X_t) Y_t\, dB_t^l\right) Y_t^{-1} \\ &= \partial_x b(t, X_t)\, dt + \sum_{l=1}^{d} \partial_x \sigma^l(t, X_t)\, dB_t^l, \end{aligned}$$

where we used $Y_t Y_t^{-1} = I_m$. We assume that

$$dY_t^{-1} = \mu_t\, dt + \sum_{l=1}^{d} \nu_t^l\, dB_t^l.$$

We need to show Y_t^{-1} is an Itô process. To this end, consider the SDE (10) with $Y_0 = I_m$. The coefficients $\partial_x b(t, X_t)$ and $\partial_x \sigma^l(t, X_t)$ are assumed to be bounded and measurable. The SDE (10) is linear, and under standard conditions (e.g., Lipschitz continuity of the coefficients), the solution Y_t exists, is unique, and is a semimartingale — specifically, an Itô process — adapted to the filtration generated by the Brownian motions B_t^l. Moreover, since $Y_0 = I_m$ is invertible and the coefficients satisfy regularity conditions, Y_t remains invertible almost surely for all $t \geq 0$. Define the function $f : \mathrm{GL}(m, \mathbb{R}) \to \mathrm{GL}(m, \mathbb{R})$ by $f(A) = A^{-1}$, where $\mathrm{GL}(m, \mathbb{R})$ is the group of $m \times m$ invertible matrices. The map f is smooth (infinitely differentiable) on $\mathrm{GL}(m, \mathbb{R})$, with first derivative $Df(A)H = -A^{-1}HA^{-1}$ and second derivative terms involving higher-order products. Applying Itô' s formula to $Y_t^{-1} = f(Y_t)$, where Y_t is an Itô process, yields a stochastic differential of the form

$$dY_t^{-1} = \mu_t \, dt + \sum_{l=1}^{d} \nu_t^l \, dB_t^l,$$

where μ_t and ν_t^l are adapted processes derived from the drift and diffusion terms of Y_t. Thus, Y_t^{-1} is itself an Itô process, and we proceed to determine μ_t and ν_t^l. Continuing with the proof

$$Y_t dY_t^{-1} = Y_t \mu_t \, dt + \sum_{l=1}^{d} Y_t \nu_t^l \, dB_t^l.$$

The quadratic covariation is

$$d[Y_t, Y_t^{-1}] = \sum_{l=1}^{d} \left(\partial_x \sigma^l(t, X_t) Y_t \right) \nu_t^l \, dt.$$

Substituting into the differential

$$\begin{aligned} d(Y_t Y_t^{-1}) = &\left(\partial_x b(t, X_t) + Y_t \mu_t + \sum_{l=1}^{d} \partial_x \sigma^l(t, X_t) Y_t \nu_t^l \right) dt \\ &+ \sum_{l=1}^{d} \left(\partial_x \sigma^l(t, X_t) + Y_t \nu_t^l \right) dB_t^l \\ = &\, 0, \end{aligned}$$

and equating the coefficients yields

$$\begin{aligned} &dB_t^l\text{-term} : Y_t \nu_t^l + \partial_x \sigma^l(t, X_t) = 0, \quad \Rightarrow \quad \nu_t^l = -Y_t^{-1} \partial_x \sigma^l(t, X_t), \\ &dt\text{-term} : \quad Y_t \mu_t + \partial_x b(t, X_t) + \sum_{l=1}^{d} \partial_x \sigma^l(t, X_t) Y_t (-Y_t^{-1} \partial_x \sigma^l(t, X_t)) = 0. \end{aligned}$$

Simplifying

$$\sum_{l=1}^{d} \partial_x \sigma^l(t, X_t) Y_t (-Y_t^{-1} \partial_x \sigma^l(t, X_t)) = - \sum_{l=1}^{d} \partial_x \sigma^l(t, X_t) \partial_x \sigma^l(t, X_t),$$

$$Y_t \mu_t + \partial_x b(t, X_t) - \sum_{l=1}^{d} \partial_x \sigma^l(t, X_t) \partial_x \sigma^l(t, X_t) = 0,$$

$$\mu_t = Y_t^{-1} \left(-\partial_x b(t, X_t) + \sum_{l=1}^{d} \left(\partial_x \sigma^l(t, X_t) \right)^2 \right).$$

Thus

$$dY_t^{-1} = Y_t^{-1} \left(-\partial_x b(t, X_t) + \sum_{l=1}^{d} \left(\partial_x \sigma^l(t, X_t) \right)^2 \right) dt - \sum_{l=1}^{d} Y_t^{-1} \partial_x \sigma^l(t, X_t) \, dB_t^l.$$

The initial condition $Y_0^{-1} = I_m$ holds since $Y_0 = I_m$.

□

Lemma 8 (Malliavin derivative of inverse matrices) *Let A be a random $m \times m$ matrix that is invertible almost surely, and assume there exists $\lambda > 0$ such that $A \geq \lambda I_m$ almost surely (uniform ellipticity). Suppose further that A and A^{-1} are Malliavin differentiable. Then, for each $t \in [0, T]$,*

$$D_t(A^{-1}) = -A^{-1}(D_t A)A^{-1}.$$

Proof Since $AA^{-1} = I$, applying the Malliavin derivative yields

$$D_t(AA^{-1}) = (D_t A)A^{-1} + A(D_t A^{-1}) = D_t I = 0.$$

Consequently,

$$A(D_t A^{-1}) = -(D_t A)A^{-1}.$$

Multiplying on the left by A^{-1} gives the desired result. □

Remark 1 (The role of integrability and ellipticity conditions) *While Lemma 8 is purely algebraic, ensuring that A^{-1} (and the right-hand side) belongs to the Malliavin–Sobolev space $\mathbb{D}^{1,2}$ requires controlling the integrability of $\|A^{-1}\|$.*

Sufficient condition: uniform ellipticity. *If $A \geq \lambda I_m$ almost surely for some deterministic $\lambda > 0$, then $\|A^{-1}\| \leq \lambda^{-1}$ and all polynomial moments of A^{-1} are finite. Under this bound, every entry of A^{-1} lies in $\mathbb{D}^\infty$, and the componentwise identity*

$$D(A^{-1})^{ij} = -\sum_{k,\ell=1}^{m} (A^{-1})^{ik}(A^{-1})^{\ell j}\, DA^{k\ell},$$

established in [23, Lemma 2.1.6], is justified.

Necessity of moment conditions. *Without such a lower bound on the eigenvalues, the inverse may fail to be integrable. In dimension one, for instance, $A = \int_0^T B_s^2\, ds > 0$ satisfies $A^{-q} \in L^1(\Omega)$ only for $q < \frac{1}{2}$; thus $A^{-1} \notin \mathbb{D}^{1,2}$ and Lemma 2.1.6 cannot be applied. Uniform ellipticity (or, more generally, the moment hypothesis $|\det A|^{-1} \in L^p$ for all $p \geq 1$ required in [23, Lemma 2.1.6]) prevents this pathology.*

Lemma 9 (Commutativity of Malliavin and partial derivatives) *Let $X_t = X_t(x)$ be the solution to the stochastic differential equation*

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t, \quad X_0 = x, \quad 0 \leq t \leq T,$$

where B_t is a d-dimensional standard Brownian motion and $x \in \mathbb{R}^m$. Assume the coefficients satisfy $b, \sigma \in \mathcal{C}_b^2([0,T] \times \mathbb{R}^m)$ (all partial derivatives up to second order are continuous and bounded). Let $Y_t = \partial X_t / \partial x$ denote the first variation process. Then, for every $0 \leq t \leq T$, where D_t denotes the Malliavin derivative,

$$D_t\left(\frac{\partial X_T}{\partial x}\right) = \frac{\partial}{\partial x}(D_t X_T).$$

Proof We compute both sides explicitly to verify equality. The SDE for X_t can be rewritten in the integral form

$$X_t = x + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s.$$

The first variation process $Y_t = \partial X_t/\partial x$ satisfies

$$dY_t = \partial_x b(t, X_t) Y_t \, dt + \sum_{l=1}^{d} \partial_x \sigma^l(t, X_t) Y_t \, dB_t^l, \quad Y_0 = I_m.$$

We have seen that the inverse Y_t^{-1} satisfies the SDE

$$dY_t^{-1} = -Y_t^{-1} \partial_x b(t, X_t) \, dt - \sum_{l=1}^{d} Y_t^{-1} \partial_x \sigma^l(t, X_t) \, dB_t^l + \sum_{l=1}^{d} Y_t^{-1} \left(\partial_x \sigma^l(t, X_t) \right)^2 dt, \quad Y_0^{-1} = I_m$$

while the second variation process satisfies

$$\begin{aligned} dZ_t &= [\partial_{xx} b(t, X_t)(Y_t \otimes Y_t) + \partial_x b(t, X_t) Z_t] \, dt \\ &\quad + \sum_{l=1}^{d} \left[\partial_{xx} \sigma^l(t, X_t)(Y_t \otimes Y_t) + \partial_x \sigma^l(t, X_t) Z_t \right] dB_t^l, \quad Z_0 = 0. \end{aligned}$$

$$D_t X_T = Y_T Y_t^{-1} \sigma(t, X_t).$$

Then

$$\begin{aligned} \frac{\partial}{\partial x}(D_t X_T) &= \frac{\partial}{\partial x} \left(Y_T Y_t^{-1} \sigma(t, X_t) \right) \\ &= \frac{\partial Y_T}{\partial x} \left(Y_t^{-1} \sigma(t, X_t) \right) + Y_T \frac{\partial}{\partial x} \left(Y_t^{-1} \sigma(t, X_t) \right). \end{aligned}$$

Since $\partial Y_T/\partial x = Z_T$ we have

$$\frac{\partial Y_T}{\partial x} \left(Y_t^{-1} \sigma(t, X_t) \right) = Z_T Y_t^{-1} \sigma(t, X_t).$$

For the second term:

$$\frac{\partial}{\partial x} \left(Y_t^{-1} \sigma(t, X_t) \right) = \frac{\partial Y_t^{-1}}{\partial x} \sigma(t, X_t) + Y_t^{-1} \frac{\partial \sigma(t, X_t)}{\partial x},$$

where $\partial Y_t^{-1}/\partial x = -Y_t^{-1} Z_t Y_t^{-1}$ and $\partial \sigma(t, X_t)/\partial x = \partial_x \sigma(t, X_t) Y_t$. Thus

$$\frac{\partial}{\partial x} \left(Y_t^{-1} \sigma(t, X_t) \right) = -\left(Y_t^{-1} Z_t Y_t^{-1} \right) \sigma(t, X_t) + Y_t^{-1} \left(\partial_x \sigma(t, X_t) Y_t \right),$$

$$\frac{\partial}{\partial x}(D_t X_T) = Z_T Y_t^{-1} \sigma(t, X_t) + Y_T \left(-\left(Y_t^{-1} Z_t Y_t^{-1} \right) \sigma(t, X_t) + Y_t^{-1} \left(\partial_x \sigma(t, X_t) Y_t \right) \right).$$

Moreover, since $\partial X_T/\partial x = Y_T$ we have

$$Y_T = I_m + \int_0^T \partial_x b(s, X_s) Y_s \, ds + \int_0^T \partial_x \sigma(s, X_s) Y_s \, dB_s,$$

$$\begin{aligned} D_t Y_T = \partial_x \sigma(t, X_t) Y_t &+ \int_t^T \left[\partial_x b(s, X_s) D_t Y_s + \partial_{xx} b(s, X_s) \left(Y_s Y_t^{-1} \sigma(t, X_t) \right) Y_s \right] ds \\ &+ \int_t^T \left[\partial_x \sigma(s, X_s) D_t Y_s + \partial_{xx} \sigma(s, X_s) \left(Y_s Y_t^{-1} \sigma(t, X_t) \right) Y_s \right] dB_s. \end{aligned}$$

The solution is

$$D_t Y_T = Z_T Y_t^{-1} \sigma(t, X_t) - Y_T Y_t^{-1} Z_t Y_t^{-1} \sigma(t, X_t) + Y_T Y_t^{-1} \partial_x \sigma(t, X_t) Y_t.$$

Clearly the expressions are equal, confirming that

$$D_t \left(\frac{\partial X_T}{\partial x} \right) = \frac{\partial}{\partial x} \left(D_t X_T \right).$$

The regularity conditions ensure all derivatives and integrals are well-defined. □

Lemma 10 (Malliavin derivative for the first variation process) *For $t \le T$, the Malliavin derivative of the first variation process Y_T is given by*

$$\begin{aligned} D_t Y_T = &\, Z_T\, Y_t^{-1}\, \sigma(t, X_t) \\ &- Y_T\, Y_t^{-1}\, Z_t\, Y_t^{-1}\, \sigma(t, X_t) \\ &+ Y_T\, Y_t^{-1}\, \partial_x \sigma(t, X_t)\, Y_t. \end{aligned}$$

Proof Consider the SDE satisfied by the first variation process

$$dY_s = \partial_x b(s, X_s)\, Y_s\, ds \;+\; \partial_x \sigma(s, X_s)\, Y_s\, dB_s, \qquad Y_0 = I.$$

Recall that $Y_s = \partial X_s/\partial x$. The Malliavin derivative $D_t Y_T$ represents the sensitivity of Y_T to a perturbation in the Brownian motion at time t. Since $Y_T = \partial X_T/\partial x$ we have

$$D_t Y_T = D_t \left(\frac{\partial X_T}{\partial x} \right) = \frac{\partial}{\partial x} \left(D_t X_T \right).$$

This identity follows from interchanging the partial derivative w.r.t. x and the Malliavin derivative D_t. We use the known expression $D_t X_T = Y_T \left(Y_t^{-1}\, \sigma(t, X_t) \right)$ which is valid for all $t \le T$. Define $W_t = Y_t^{-1}\, \sigma(t, X_t)$. Thus,

$$D_t X_T = Y_T\, W_t.$$

Differentiate this equation with respect to the initial condition x to obtain

$$\frac{\partial}{\partial x}(D_t X_T) = \frac{\partial Y_T}{\partial x} W_t + \; Y_T\, \frac{\partial W_t}{\partial x}.$$

By definition of second variation process $\partial Y_T/\partial x = Z_T$. Hence

$$D_t Y_T = Z_T\, W_t + Y_T \frac{\partial W_t}{\partial x}. \tag{11}$$

Recalling that $W_t = Y_t^{-1}\, \sigma(t, X_t)$ we can write

$$\begin{aligned} \frac{\partial W_t}{\partial x} &= \frac{\partial}{\partial x} \left(Y_t^{-1}\, \sigma(t, X_t) \right) \\ &= \left(\frac{\partial Y_t^{-1}}{\partial x} \right) \sigma(t, X_t) + Y_t^{-1}\, \frac{\partial \sigma(t, X_t)}{\partial x}. \end{aligned}$$

Each of these two terms comes from the product rule (now for partial derivatives w.r.t. x). Since $Y_t\, Y_t^{-1} = I$, differentiating both sides w.r.t. x yields $\partial Y_t/\partial x\, Y_t^{-1} + Y_t\, \partial Y_t^{-1}/\partial x = 0$. Hence

$$\frac{\partial Y_t^{-1}}{\partial x} = -Y_t^{-1} \left(\frac{\partial Y_t}{\partial x} \right) Y_t^{-1} = -Y_t^{-1} Z_t Y_t^{-1}.$$

We also have $\partial X_t/\partial x = Y_t$. Thus by chain rule

$$\frac{\partial \sigma(t, X_t)}{\partial x} = \partial_x \sigma(t, X_t)\, Y_t.$$

Therefore,

$$\frac{\partial W_t}{\partial x} = -Y_t^{-1} Z_t Y_t^{-1} \sigma(t, X_t) + Y_t^{-1} \partial_x \sigma(t, X_t) Y_t.$$

This completes the computation of $\partial W_t/\partial x$. Putting this result back into (11) yields

$$D_t Y_T = Z_T\, Y_t^{-1}\, \sigma(t, X_t) \;+\; Y_T \Big[- Y_t^{-1}\, Z_t\, Y_t^{-1}\, \sigma(t, X_t) \;+\; Y_t^{-1}\, \partial_x \sigma(t, X_t)\, Y_t \Big].$$

Factor out common terms to rewrite it in the stated form

$$D_t Y_T = Z_T\, Y_t^{-1}\, \sigma(t, X_t) - \; Y_T\, Y_t^{-1}\, Z_t\, Y_t^{-1}\, \sigma(t, X_t) + \; Y_T\, Y_t^{-1}\, \partial_x \sigma(t, X_t)\, Y_t.$$

The second term $-Y_T\, Y_t^{-1}\, Z_t\, Y_t^{-1}\, \sigma(t, X_t)$ specifically accounts for the appearance of the second variation Z_t inside the inverse, while the final term $Y_T\, Y_t^{-1}\, \partial_x \sigma(t, X_t)\, Y_t$ encodes the effect of differentiating σ itself w.r.t. x. □

Lemma 11 (Malliavin derivative formula for the inverse first variation process) *For the inverse first variation process Y_s^{-1}, the Malliavin derivative is given by*

- *For $t \le s$:*

$$D_t Y_s^{-1} = -Y_s^{-1} \left[Z_s Y_t^{-1} \sigma(t, X_t) - Y_s Y_t^{-1} Z_t Y_t^{-1} \sigma(t, X_t) + Y_s Y_t^{-1} \partial_x \sigma(t, X_t) Y_t \right] Y_s^{-1}$$

- *For $t > s$:*

$$D_t Y_s^{-1} = 0$$

where $Z_s = \partial^2 X_s / \partial x^2$ is the second variation process.

Proof We derive $D_t Y_s^{-1}$ by applying the Malliavin derivative to the identity $Y_s Y_s^{-1} = I_m$ and using the product rule. The proof splits into two cases based on the relationship between t and s, and we leverage the expression for $D_t Y_s$ derived similarly to $D_t Y_T$.

Let us begin with the case $t \le s$. Since $Y_s Y_s^{-1} = I_m$ (the $m \times m$ identity matrix), we apply the Malliavin derivative D_t to both sides

$$D_t(Y_s Y_s^{-1}) = D_t(I_m) = 0.$$

Using the product rule for Malliavin derivatives

$$D_t(Y_s Y_s^{-1}) = (D_t Y_s) Y_s^{-1} + Y_s (D_t Y_s^{-1}) = 0.$$

Rearranging to isolate $D_t Y_s^{-1}$

$$\begin{aligned} Y_s (D_t Y_s^{-1}) &= -(D_t Y_s) Y_s^{-1}, \\ D_t Y_s^{-1} &= -Y_s^{-1} (D_t Y_s) Y_s^{-1}. \end{aligned}$$

To proceed, we need $D_t Y_s$. Since $Y_s = \partial X_s / \partial x$ and $t \le s$, we adapt the derivation from the previous Lemma 10 for $D_t Y_T$, adjusting the endpoint from T to s

$$dY_u = \partial_x b(u, X_u) Y_u\, du + \partial_x \sigma(u, X_u) Y_u\, dB_u, \quad 0 \le u \le s, \qquad Y_0 = I.$$

We have

$$D_t Y_s = D_t \left(\frac{\partial X_s}{\partial x} \right) = \frac{\partial}{\partial x} (D_t X_s).$$

For $t \le s$, the Malliavin derivative of X_s is

$$D_t X_s = Y_s Y_t^{-1} \sigma(t, X_t).$$

Define $W_t = Y_t^{-1} \sigma(t, X_t)$, so:

$$D_t X_s = Y_s W_t.$$

$$\frac{\partial}{\partial x} (D_t X_s) = \frac{\partial}{\partial x} (Y_s W_t) = \frac{\partial Y_s}{\partial x} W_t + Y_s \frac{\partial W_t}{\partial x}.$$

Since $\partial Y_s/\partial x = Z_s$ (second variation process)

$$D_t Y_s = Z_s W_t + Y_s \frac{\partial W_t}{\partial x}.$$

Recalling that $W_t = Y_t^{-1}\sigma(t, X_t)$,

$$\frac{\partial W_t}{\partial x} = \frac{\partial Y_t^{-1}}{\partial x}\sigma(t, X_t) + Y_t^{-1}\frac{\partial \sigma(t, X_t)}{\partial x}.$$

For $\partial Y_t^{-1}/\partial x$ differentiate $Y_t Y_t^{-1} = I$ to obtain

$$\frac{\partial Y_t^{-1}}{\partial x} = -Y_t^{-1}\frac{\partial Y_t}{\partial x}Y_t^{-1} = -Y_t^{-1} Z_t Y_t^{-1}.$$

For $\partial\sigma(t, X_t)/\partial x$, since $\partial X_t/\partial x = Y_t$, we have

$$\frac{\partial \sigma(t, X_t)}{\partial x} = \partial_x \sigma(t, X_t) Y_t.$$

Therefore

$$\frac{\partial W_t}{\partial x} = -Y_t^{-1} Z_t Y_t^{-1}\sigma(t, X_t) + Y_t^{-1}\partial_x\sigma(t, X_t) Y_t$$

and

$$\begin{aligned} D_t Y_s &= Z_s Y_t^{-1}\sigma(t, X_t) + Y_s \left(-Y_t^{-1} Z_t Y_t^{-1}\sigma(t, X_t) + Y_t^{-1}\partial_x\sigma(t, X_t) Y_t\right), \\ &= Z_s Y_t^{-1}\sigma(t, X_t) - Y_s Y_t^{-1} Z_t Y_t^{-1}\sigma(t, X_t) + Y_s Y_t^{-1}\partial_x\sigma(t, X_t) Y_t. \end{aligned}$$

Finally, substitute $D_t Y_s$ into $D_t Y_s^{-1}$ to obtain

$$D_t Y_s^{-1} = -Y_s^{-1}\left[Z_s Y_t^{-1}\sigma(t, X_t) - Y_s Y_t^{-1} Z_t Y_t^{-1}\sigma(t, X_t) + Y_s Y_t^{-1}\partial_x\sigma(t, X_t) Y_t\right] Y_s^{-1}.$$

The case $t \le s$ is simpler. Since Y_s^{-1} is adapted to the filtration up to time s, and $t > s$, a perturbation in the Brownian motion at time t does not affect Y_s^{-1} (which depends only on information up to s). Thus

$$D_t Y_s^{-1} = 0.$$

This completes the proof, with the expression for $t \le s$ matching the Lemma statement, and the zero result for $t > s$ reflecting the causality of the stochastic process.

□

5.2 Proof of the main Theorem

Having established all necessary preliminary results, we now proceed to prove our Theorem 1.

Proof Applying Theorem 6, the Skorokhod integral $\delta(u_k) = \delta(u(F_k))$ is

$$\delta(u_k) = \int_0^T u_t(x)\cdot dB_t\Big|_{x=F_k} - \sum_{j=1}^m \int_0^T \partial_j u_t(F_k)\cdot D_t F_k^j\, dt. \tag{12}$$

This expression combines an Itô integral evaluated at $x = F_k$ with a correction term involving the partial derivatives of $u_t(x)$ evaluated at $x = F_k$ and the Malliavin derivatives of F_k^j. The first term in the expression for $\delta(u(F_k))$ is the Itô integral evaluated at $x = F_k$

$$\int_0^T u_t(x)\cdot dB_t\Big|_{x=F_k},$$

where for each fixed x, $u_t(x) = x^\top Y_t^{-1}\sigma(t, X_t)$ is an adapted process, so $\int_0^T u_t(x) \cdot dB_t$ is a well-defined Itô integral, and after computing this integral, we evaluate it at $x = F_k = Y_T^\top \gamma_{X_T}^{-1} e_k$, which is $\mathcal{F}_T$-measurable. This term is computationally manageable because the integration is performed with respect to an adapted integrand for fixed x, and the randomness of F_k is introduced only after the integration. With the redefined random field

$$u_t(x) = x^\top Y_t^{-1}\sigma(t, X_t) = \sum_{i=1}^{m} x_i \left[Y_t^{-1}\sigma(t, X_t)\right]_i,$$

the partial derivative with respect to x_j is

$$\partial_j u_t(x) = \frac{\partial}{\partial x_j} u_t(x) = \left[Y_t^{-1}\sigma(t, X_t)\right]_j,$$

since only the term involving x_j depends on x_j. Therefore, evaluating at $x = F_k$

$$\partial_j u_t(F_k) = \left[Y_t^{-1}\sigma(t, X_t)\right]_j.$$

This term will appear in the correction term of the Skorokhod integral decomposition. Before proceeding further, we state a general result from Lemma 8 for the Malliavin derivative of the inverse of a random matrix. To compute the Malliavin derivative $D_t F_k^j$, consider the new definition

$$F_k = Y_T^\top \gamma_{X_T}^{-1} e_k, \qquad F_k^j = e_j^\top Y_T^\top \gamma_{X_T}^{-1} e_k,$$

where Y_T is the first variation process at time T, γ_{X_T} is the Malliavin covariance matrix, and e_j, e_k are standard basis vectors. Applying the Malliavin derivative

$$D_t F_k^j = e_j^\top D_t(Y_T^\top \gamma_{X_T}^{-1}) e_k.$$

Using the product rule, we obtain

$$\begin{aligned} D_t(Y_T^\top \gamma_{X_T}^{-1}) &= (D_t Y_T^\top)\gamma_{X_T}^{-1} + Y_T^\top D_t(\gamma_{X_T}^{-1}), \\ D_t F_k^j &= e_j^\top (D_t Y_T^\top)\gamma_{X_T}^{-1} e_k + e_j^\top Y_T^\top D_t(\gamma_{X_T}^{-1}) e_k. \end{aligned}$$

For $t \leq T$, the Malliavin derivative $D_t Y_T$ is given in Lemma 10

$$D_t Y_T = Z_T Y_t^{-1}\sigma(t, X_t) - Y_T Y_t^{-1} Z_t Y_t^{-1}\sigma(t, X_t) + Y_T Y_t^{-1}\partial_x\sigma(t, X_t) Y_t,$$

where Z_t is the second variation process. Taking the transpose

$$D_t Y_T^\top = \left[Z_T Y_t^{-1}\sigma(t, X_t) - Y_T Y_t^{-1} Z_t Y_t^{-1}\sigma(t, X_t) + Y_T Y_t^{-1}\partial_x\sigma(t, X_t) Y_t\right]^\top.$$

From Lemma 8 we have that for an invertible random matrix A

$$D_t(A^{-1}) = -A^{-1}(D_t A)A^{-1}.$$

For the specific case of $A = \gamma_{X_T}$ this yields

$$D_t(\gamma_{X_T}^{-1}) = -\gamma_{X_T}^{-1}(D_t\gamma_{X_T})\gamma_{X_T}^{-1},$$

where

$$\gamma_{X_T} = \int_0^T Y_T Y_s^{-1}\sigma(s, X_s)\sigma(s, X_s)^\top (Y_s^{-1})^\top Y_T^\top \, ds,$$

and $D_t\gamma_{X_T}$ requires computing the Malliavin derivative of the integrand. Thus,

$$\begin{aligned} D_t F_k^j &= e_j^\top (D_t Y_T^\top)\gamma_{X_T}^{-1} e_k + e_j^\top Y_T^\top D_t(\gamma_{X_T}^{-1}) e_k, \\ D_t F_k^j &= e_j^\top (D_t Y_T^\top)\gamma_{X_T}^{-1} e_k - e_j^\top Y_T^\top \gamma_{X_T}^{-1}(D_t\gamma_{X_T})\gamma_{X_T}^{-1} e_k, \end{aligned}$$

with $D_t Y_T^\top$ and $D_t \gamma_{X_T}$ as derived. We now proceed by computing the Malliavin derivative $D_t\Big[(Y_T\, Y_s^{-1}\,\sigma(s,X_s))^p\Big]$. To this end, we first recall Lemma 10 which gives the precise form of $D_t Y_T$. Afterwards, we use the product rule for Malliavin derivatives on the product $Y_T\, Y_s^{-1}\,\sigma(s,X_s)$, distinguishing between the cases $t \le s$ and $t > s$. Finally, we assemble these pieces to obtain the expression for $D_t(Y_T Y_s^{-1}\sigma(s,X_s))^p$. We provide reasoning for each step to clarify why each term appears and how the partial derivatives interact with the inverse processes. Let

$$W_s^p \;=\; \Big(Y_T\, Y_s^{-1}\,\sigma(s,X_s)\Big)^p,$$

i.e. the p-th component of the vector $Y_T\, Y_s^{-1}\,\sigma(s,X_s)$. We want to find $D_t\big(W_s^p\big)$. Since

$$W_s^p = \big(Y_T\, Y_s^{-1}\,\sigma(s,X_s)\big)^p,$$

we begin with the Malliavin derivative of the product $Y_T\, Y_s^{-1}\,\sigma(s,X_s)$.

$$D_t W_s^p = D_t\Big(\big(Y_T\, Y_s^{-1}\,\sigma(s,X_s)\big)^p\Big).$$

- **Case 1: $t \le s$.** In this scenario, a “kick” in the Brownian motion at time t does affect X_s (and hence Y_s). Thus

  $$D_t\big(Y_T\, Y_s^{-1}\,\sigma(s,X_s)\big) = D_t Y_T \;\cdot\; \big(Y_s^{-1}\,\sigma(s,X_s)\big) \;+\; Y_T\;\underbrace{D_t\big(Y_s^{-1}\,\sigma(s,X_s)\big)}_{\text{chain rule}}.$$

  Using Lemma 10 we write $D_t Y_T$ as

  $$D_t Y_T \;=\; Z_T\, Y_t^{-1}\,\sigma(t,X_t) \;-\; Y_T\, Y_t^{-1}\, Z_t\, Y_t^{-1}\,\sigma(t,X_t) \;+\; Y_T\, Y_t^{-1}\,\partial_x\sigma(t,X_t)\, Y_t.$$

  Note that

  $$D_t\big(Y_s^{-1}\,\sigma(s,X_s)\big) = \big(D_t Y_s^{-1}\big)\,\sigma(s,X_s) \;+\; Y_s^{-1}\,\big(D_t\sigma(s,X_s)\big). \tag{13}$$

  For $t \le s$, the Malliavin derivative of the inverse is

  $$\begin{aligned} D_t Y_s^{-1} = -Y_s^{-1}\big[&Z_s Y_t^{-1}\sigma(t,X_t) \\ &- Y_s Y_t^{-1} Z_t Y_t^{-1}\sigma(t,X_t) \\ &+ Y_s Y_t^{-1}\partial_x\sigma(t,X_t) Y_t\big] Y_s^{-1}. \end{aligned}$$

  This expression accounts for the second variation processes Z_s and Z_t, as well as the derivative of the diffusion coefficient. Applying the chain rule

  $$D_t\sigma(s,X_s) \;=\; \partial_x\sigma(s,X_s)\,\big(Y_s\, Y_t^{-1}\,\sigma(t,X_t)\big).$$

  Substituting into (13) yields

  $$\begin{aligned} D_t\big(Y_s^{-1}\,\sigma(s,X_s)\big) = -Y_s^{-1}\Big[&Z_s Y_t^{-1}\sigma(t,X_t) \\ &- Y_s Y_t^{-1} Z_t Y_t^{-1}\sigma(t,X_t) \\ &+ Y_s Y_t^{-1}\partial_x\sigma(t,X_t) Y_t\Big] Y_s^{-1}\sigma(s,X_s) \\ &+ Y_s^{-1}\partial_x\sigma(s,X_s)\big(Y_s Y_t^{-1}\sigma(t,X_t)\big). \end{aligned}$$

Thus, for $t \le s$

$$
\begin{aligned}
D_t W_s^p = \Bigg[\Big(& Z_T\, Y_t^{-1}\, \sigma(t, X_t) - Y_T\, Y_t^{-1}\, Z_t\, Y_t^{-1}\, \sigma(t, X_t) \\
& + Y_T\, Y_t^{-1}\, \partial_x \sigma(t, X_t)\, Y_t \Big)\, Y_s^{-1}\, \sigma(s, X_s) \\
& + Y_T \Bigg(- Y_s^{-1} \Big[Z_s Y_t^{-1}\, \sigma(t, X_t) - Y_s Y_t^{-1}\, Z_t Y_t^{-1}\, \sigma(t, X_t) \\
& \quad + Y_s Y_t^{-1}\, \partial_x \sigma(t, X_t) Y_t \Big]\, Y_s^{-1}\, \sigma(s, X_s) \\
& \quad + Y_s^{-1}\, \partial_x \sigma(s, X_s) \Big(Y_s Y_t^{-1}\, \sigma(t, X_t) \Big) \Bigg) \Bigg]^p .
\end{aligned}
$$

We place the entire sum inside brackets $[\dots]^p$ because we are taking the p-th component of the resulting vector.

- **Case 2:** $t > s$**.** In this case, a Brownian perturbation at time t does *not* affect X_s (nor Y_s) because $s < t$. Hence

  $$D_t\big(Y_s^{-1}\, \sigma(s, X_s)\big) = 0,$$

  and the only contribution is from $D_t Y_T$. Therefore,

  $$
  \begin{aligned}
  D_t W_s^p = \Bigg[\Big(& Z_T\, Y_t^{-1}\, \sigma(t, X_t) \\
  & - Y_T\, Y_t^{-1}\, Z_t\, Y_t^{-1}\, \sigma(t, X_t) + Y_T\, Y_t^{-1}\, \partial_x \sigma(t, X_t)\, Y_t \Big)\, Y_s^{-1}\, \sigma(s, X_s) \Bigg]^p .
  \end{aligned}
  $$

Next, we recall that

$$\gamma_{X_T}^{p,q} = \int_0^T [Y_T Y_s^{-1} \sigma(s, X_s)]^p \cdot [Y_T Y_s^{-1} \sigma(s, X_s)]^q\, ds.$$

Hence, $D_t \gamma_{X_T}^{p,q}$ involves $\int_0^T [D_t([Y_T Y_s^{-1} \sigma(s, X_s)]^p) \cdot [Y_T Y_s^{-1} \sigma(s, X_s)]^q + [Y_T Y_s^{-1} \sigma(s, X_s)]^p \cdot D_t([Y_T Y_s^{-1} \sigma(s, X_s)]^q)]\, ds$. We split the integration region into $[0, t]$ and $[t, T]$ to reflect the piecewise definitions.

$$
\begin{aligned}
D_t \gamma_{X_T}^{p,q} = \int_0^t \Bigg[\Big(& Z_T Y_t^{-1} \sigma(t, X_t) \\
& - Y_T Y_t^{-1} Z_t Y_t^{-1} \sigma(t, X_t) \\
& + Y_T Y_t^{-1} \partial_x \sigma(t, X_t) Y_t \Big) Y_s^{-1} \sigma(s, X_s) \Bigg]^p \cdot [Y_T Y_s^{-1} \sigma(s, X_s)]^q\, ds \\
+ \int_0^t [Y_T Y_s^{-1} \sigma(s, X_s)]^p \cdot \Bigg[\Big(& Z_T Y_t^{-1} \sigma(t, X_t) \\
& - Y_T Y_t^{-1} Z_t Y_t^{-1} \sigma(t, X_t) \\
& + Y_T Y_t^{-1} \partial_x \sigma(t, X_t) Y_t \Big) Y_s^{-1} \sigma(s, X_s) \Bigg]^q ds
\end{aligned}
$$

$$
\begin{aligned}
&+ \int_t^T \Bigg[\Big(Z_T Y_t^{-1} \sigma(t, X_t) \\
&\qquad - Y_T Y_t^{-1} Z_t Y_t^{-1} \sigma(t, X_t) \\
&\qquad + Y_T Y_t^{-1} \partial_x \sigma(t, X_t) Y_t \Big) Y_s^{-1} \sigma(s, X_s) \\
&\qquad + Y_T \Big(-Y_s^{-1} \Big[Z_s Y_t^{-1} \sigma(t, X_t) \\
&\qquad\quad - Y_s Y_t^{-1} Z_t Y_t^{-1} \sigma(t, X_t) \\
&\qquad\quad + Y_s Y_t^{-1} \partial_x \sigma(t, X_t) Y_t \Big] Y_s^{-1} \sigma(s, X_s) \\
&\qquad\quad + Y_s^{-1} \partial_x \sigma(s, X_s) \big(Y_s Y_t^{-1} \sigma(t, X_t) \big) \Big) \Bigg]^p \cdot [Y_T Y_s^{-1} \sigma(s, X_s)]^q \, ds \\
&+ \int_t^T [Y_T Y_s^{-1} \sigma(s, X_s)]^p \cdot \Bigg[\Big(Z_T Y_t^{-1} \sigma(t, X_t) \\
&\qquad - Y_T Y_t^{-1} Z_t Y_t^{-1} \sigma(t, X_t) \\
&\qquad + Y_T Y_t^{-1} \partial_x \sigma(t, X_t) Y_t \Big) Y_s^{-1} \sigma(s, X_s) \\
&\qquad + Y_T \Big(-Y_s^{-1} \Big[Z_s Y_t^{-1} \sigma(t, X_t) \\
&\qquad\quad - Y_s Y_t^{-1} Z_t Y_t^{-1} \sigma(t, X_t) \\
&\qquad\quad + Y_s Y_t^{-1} \partial_x \sigma(t, X_t) Y_t \Big] Y_s^{-1} \sigma(s, X_s) \\
&\qquad\quad + Y_s^{-1} \partial_x \sigma(s, X_s) \big(Y_s Y_t^{-1} \sigma(t, X_t) \big) \Big) \Bigg]^q \, ds.
\end{aligned}
$$

Above, each bracketed term depends on whether $s < t$ or $s \geq t$, and the difference is precisely the extra terms involving the second variation processes and diffusion derivatives from the corrected $D_t(Y_T Y_s^{-1} \sigma(s, X_s))$. This final expression concisely encodes all contributions to $D_t \gamma_{X_T}^{p,q}$ from the single-time "kick" in the Brownian path at time t. We now handle the correction term $\sum_{j=1}^{m} \partial_j u_t(F_k) \cdot D_t F_k^j$ that appears in the Skorokhod integral decomposition $\delta(u_k) = \int_0^T u_t(x) \cdot dB_t \Big|_{x=F_k} - \sum_j \int_0^T \partial_j u_t(F_k) \cdot D_t F_k^j \, dt$. Hereafter we show how each step follows from the chain rule in Malliavin calculus, the use of $D_t(\gamma_{X_T}^{-1}) = -\gamma_{X_T}^{-1} (D_t \gamma_{X_T}) \gamma_{X_T}^{-1}$, and the expression for $D_t \gamma_{X_T}^{p,q}$. We also show how to integrate the resulting expression over t. Let us begin by recalling the general formula for the Skorokhod integral

$$
\delta(u_k) = \int_0^T u_t(x) \cdot dB_t \Big|_{x=F_k} - \int_0^T \sum_{j=1}^{m} \partial_j u_t(F_k) \cdot D_t\big(F_k^j\big) \, dt.
$$

The term $\sum_{j=1}^{m} \partial_j u_t(F_k) \cdot D_t F_k^j$ is often called the "correction term." We have already found that $\partial_j u_t(F_k) = \big[Y_t^{-1} \sigma(t, X_t)\big]_j$, and $F_k^j = e_j^\top Y_T^\top \gamma_{X_T}^{-1} e_k$. Since $D_t(\gamma_{X_T}^{-1}) = -\gamma_{X_T}^{-1} (D_t \gamma_{X_T}) \gamma_{X_T}^{-1}$, we obtain

$$
D_t F_k^j = D_t(e_j^\top Y_T^\top \gamma_{X_T}^{-1} e_k) = e_j^\top (D_t Y_T^\top) \gamma_{X_T}^{-1} e_k - e_j^\top Y_T^\top \gamma_{X_T}^{-1} (D_t \gamma_{X_T}) \gamma_{X_T}^{-1} e_k.
$$

Hence,

$$\sum_{j=1}^{m} \partial_j u_t(F_k) \cdot D_t F_k^j = \sum_{j=1}^{m} \big[Y_t^{-1}\sigma(t,X_t)\big]_j \left(e_j^\top (D_t Y_T^\top) \gamma_{X_T}^{-1} e_k - e_j^\top Y_T^\top \gamma_{X_T}^{-1} (D_t \gamma_{X_T}) \gamma_{X_T}^{-1} e_k \right). \tag{14}$$

Now, recall that $D_t \gamma_{X_T}^{p,q}$ splits into integrals over $[0,t]$ and $[t,T]$, and contains contributions from terms like $Z_T\, Y_t^{-1}\, \sigma(t,X_t)$, $Y_T\, Y_t^{-1}\, Z_t\, Y_t^{-1}\, \sigma(t,X_t)$, $Y_T\, Y_t^{-1}\, \partial_x \sigma(t,X_t)\, Y_t$, and so forth. Substituting this equation above, we get a sum of integrals split by whether $s < t$ or $s > t$. To complete the calculation of the correction term contribution in the integral $\int_0^T \ldots dt$, we integrate the expression (14) from $t = 0$ to $t = T$. This yields,

$$\begin{aligned}
\int_0^T \sum_{j=1}^{m} \partial_j u_t(F_k) \,\cdot\, D_t F_k^j \, dt &= \int_0^T \sum_{j=1}^{m} \big[Y_t^{-1}\sigma(t,X_t)\big]_j \cdot A_{jk}(t)\, dt \\
&\quad - \int_0^T \sum_{j=1}^{m} \big[Y_t^{-1}\sigma(t,X_t)\big]_j \cdot B_{jk}(t)\, dt \\
&\quad - \int_0^T \sum_{j=1}^{m} \big[Y_t^{-1}\sigma(t,X_t)\big]_j \cdot C_{jk}(t)\, dt
\end{aligned}$$

where

$$\begin{aligned}
A_{jk}(t) &= e_j^\top \Big[\sigma(t,X_t)^\top (Y_t^{-1})^\top Z_T^\top - \sigma(t,X_t)^\top (Y_t^{-1})^\top Z_t^\top (Y_t^{-1})^\top Y_T^\top \\
&\qquad + \Big(Y_T Y_t^{-1} \partial_x \sigma(t,X_t) Y_t \Big)^\top \Big] \gamma_{X_T}^{-1} e_k \\
B_{jk}(t) &= e_j^\top Y_T^\top \gamma_{X_T}^{-1} \cdot \left[\int_0^t I_1(t,s)\, ds + \int_0^t I_2(t,s)\, ds \right] \gamma_{X_T}^{-1} e_k \\
C_{jk}(t) &= e_j^\top Y_T^\top \gamma_{X_T}^{-1} \cdot \left[\int_t^T I_3(t,s)\, ds + \int_t^T I_4(t,s)\, ds \right] \gamma_{X_T}^{-1} e_k
\end{aligned}$$

with the integrands having components

$$\begin{aligned}
I_1^{p,q}(t,s) &= \Big[\Omega(t) Y_s^{-1} \sigma(s,X_s) \Big]^p \cdot [Y_T Y_s^{-1} \sigma(s,X_s)]^q \\
I_2^{p,q}(t,s) &= [Y_T Y_s^{-1} \sigma(s,X_s)]^p \cdot \Big[\Omega(t) Y_s^{-1} \sigma(s,X_s) \Big]^q \\
I_3^{p,q}(t,s) &= \Big[\Omega(t) Y_s^{-1} \sigma(s,X_s) + Y_T \Theta(t,s) \Big]^p \cdot [Y_T Y_s^{-1} \sigma(s,X_s)]^q \\
I_4^{p,q}(t,s) &= [Y_T Y_s^{-1} \sigma(s,X_s)]^p \cdot \Big[\Omega(t) Y_s^{-1} \sigma(s,X_s) + Y_T \Theta(t,s) \Big]^q,
\end{aligned}$$

where

$$\begin{aligned}
\Omega(t) &= Z_T Y_t^{-1} \sigma(t,X_t) - Y_T Y_t^{-1} Z_t Y_t^{-1} \sigma(t,X_t) + Y_T Y_t^{-1} \partial_x \sigma(t,X_t) Y_t \\
\Theta(t,s) &= -Y_s^{-1} \Big[Z_s Y_t^{-1} \sigma(t,X_t) - Y_s Y_t^{-1} Z_t Y_t^{-1} \sigma(t,X_t)
\end{aligned}$$

$$
+ Y_s Y_t^{-1} \partial_x \sigma(t, X_t) Y_t \Big] Y_s^{-1} \sigma(s, X_s)
$$

$$
+ Y_s^{-1} \partial_x \sigma(s, X_s) \Big(Y_s Y_t^{-1} \sigma(t, X_t) \Big)
$$

These expansions show how the correction term $\sum_{j=1}^{m} \partial_j u_t(F_k) \cdot D_t F_k^j$ expands in terms of $D_t \gamma_{X_T}^{p,q}$. This yields the final formula for the Skorokhod integral is given in (4).

□

6 State-independent diffusion processes

In this section, we derive a simplified expression for the Skorokhod integral $\delta(u_k)$ for SDEs with state-independent diffusion coefficients. Such SDEs are of interest in diffusion generative modelling because the noise amplitude remains identical across all sample paths, and can be chosen a priori via an appropriate noise scheduler. From a numerical standpoint, stochastic differential equations with state-independent diffusion coefficients exhibit reduced stiffness compared to those with multiplicative noise. This allows for larger time steps during sampling and offers direct control over the evolution of the marginal variance. This significantly simplifies the computation of likelihood weighting and score-matching losses. As a result, most practical architectures for diffusion-based generative modelling are based on state-independent diffusion coefficients – such as the VP (Variance-Preserving), VE (Variance-Exploding), and sub-VP SDEs introduced in [5]. Hence, consider SDEs of the form

$$
dX_t \;=\; b(t, X_t)\, dt \;+\; \sigma(t)\, dB_t,
$$

where the diffusion coefficient $\sigma : [0,T] \to \mathbb{R}^{m \times d}$ depends on time t only. This property will enable the simplifications of $\delta(u_k)$ stated in the forthcoming corollary.

Corollary 12 (Skorokhod integral for state-independent diffusion) *Consider the stochastic differential equation*

$$
dX_t = b(t, X_t)\, dt + \sigma(t)\, dB_t,
$$

where the diffusion coefficient $\sigma : [0,T] \to \mathbb{R}^{m \times d}$ depends exclusively on time t. Under the assumptions of Theorem 1, the Skorokhod integral $\delta(u_k)$ simplifies to

$$
\begin{aligned}
\delta(u_k) = & \int_0^T u_t(x) \cdot dB_t \bigg|_{x = F_k} \\
& - \int_0^T \sum_{j=1}^{m} \big[Y_t^{-1} \sigma(t) \big]_j \cdot A_{jk}(t)\, dt \\
& + \int_0^T \sum_{j=1}^{m} \big[Y_t^{-1} \sigma(t) \big]_j \cdot B_{jk}(t)\, dt \\
& + \int_0^T \sum_{j=1}^{m} \big[Y_t^{-1} \sigma(t) \big]_j \cdot C_{jk}(t)\, dt
\end{aligned}
$$

where

$$A_{jk}(t) = e_j^\top \left[\sigma(t)^\top (Y_t^{-1})^\top Z_T^\top - \sigma(t)^\top (Y_t^{-1})^\top Z_t^\top (Y_t^{-1})^\top Y_T^\top \right] \gamma_{X_T}^{-1} e_k$$

$$B_{jk}(t) = e_j^\top Y_T^\top \gamma_{X_T}^{-1} \cdot \left[\int_0^t I_1(t,s)\, ds \right] \gamma_{X_T}^{-1} e_k$$

$$C_{jk}(t) = e_j^\top Y_T^\top \gamma_{X_T}^{-1} \cdot \left[\int_t^T I_2(t,s)\, ds \right] \gamma_{X_T}^{-1} e_k$$

with integrands

$$I_1(t,s) = \left[\Omega(t) Y_s^{-1} \sigma(s) \right] W_s^\top + W_s \left[\Omega(t) Y_s^{-1} \sigma(s) \right]^\top$$

$$I_2(t,s) = \left[\Phi(t,s) \right] W_s^\top + W_s \left[\Phi(t,s) \right]^\top$$

and auxiliary processes

$$\Omega(t) = Z_T Y_t^{-1} \sigma(t) - Y_T Y_t^{-1} Z_t Y_t^{-1} \sigma(t)$$

$$\Phi(t,s) = \Omega(t) Y_s^{-1} \sigma(s) - Y_T Y_s^{-1} \left[Z_s Y_t^{-1} \sigma(t) - Y_s Y_t^{-1} Z_t Y_t^{-1} \sigma(t) \right] Y_s^{-1} \sigma(s)$$

$$W_s = Y_T Y_s^{-1} \sigma(s)$$

where $u_t(x) = x^\top Y_t^{-1} \sigma(t)$ *and* $F_k = Y_T^\top \gamma_{X_T}^{-1} e_k$. *The state-independence of* $\sigma(t)$ *eliminates all terms involving* $\partial_x \sigma(t, X_t) = 0$, *significantly simplifying the expression compared to the general case.*

Acknowledgements. Daniele Venturi was supported by the U.S. Department of Energy (DOE) grant DE-SC0024563.

References

[1] Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: Bach, F., Blei, D. (eds.) Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 37, pp. 2256–2265. PMLR, Lille, France (2015). https://proceedings.mlr.press/v37/sohl-dickstein15.html

[2] Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. In: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P.S., Vaughan, J.W. (eds.) Advances in Neural Information Processing Systems, vol. 34, pp. 8780–8794. Curran Associates, Inc., Red Hook, NY (2021). https://doi.org/10.48550/arXiv.2105.05233

[3] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-Resolution Image Synthesis with Latent Diffusion Models . In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10674–10685. IEEE Computer Society, Los Alamitos, CA, USA (2022). https://doi.org/10.1109/CVPR52688.2022.01042

[4] Sattarov, T., Schreyer, M., Borth, D.: Findiff: Diffusion models for financial tabular data generation. In: Proceedings of the Fourth ACM International Conference on AI in Finance. ICAIF '23, pp. 64–72. Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3604237.3626876

[5] Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: International Conference on Learning Representations (2021). https://doi.org/10.48550/arXiv.2011.13456

[6] Fokker, A.D.: Die mittlere energie rotierender elektrischer dipole im strahlungsfeld. Annalen der Physik **348**(5), 810–820 (1914) https://doi.org/10.1002/andp.19143480507

[7] Planck, M.: Über einen satz der statistischen dynamik und seine erweiterung in der quantentheorie. Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften, 324–341 (1917)

[8] Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (eds.) Advances in Neural Information Processing Systems, vol. 33, pp. 6840–6851. Curran Associates, Inc., Red Hook, NY, USA (2020). https://doi.org/10.48550/arXiv.2006.11239

[9] Malliavin, P.: Stochastic calculus of variation and hypoelliptic operators. In: Proceedings of the International Symposium on Stochastic Differential Equations (Res. Inst. Math. Sci., Kyoto Univ., Kyoto, 1976), pp. 195–263. Wiley, New York (1978)

[10] Malliavin, P.: Stochastic Analysis. Springer, Berlin, Heidelberg (1997). https://doi.org/10.1007/978-3-642-15074-6

[11] Bismut, J.-M.: Large Deviations and the Malliavin Calculus. Progress in mathematics ; v. 45. Birkhäuser, Boston (1984)

[12] Elworthy, K.D., Li, X.M.: Formulae for the derivatives of heat semigroups. Journal of Functional Analysis **125**(1), 252–286 (1994) https://doi.org/10.1006/jfan.1994.1124

[13] Elworthy, K.D.: Stochastic Differential Equations on Manifolds. London Mathematical Society Lecture Note Series. Cambridge University Press, Cambridge

(1982). https://doi.org/10.1017/CBO9781107325609

[14] Kunita, H.: Stochastic Flows and Stochastic Differential Equations. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge (1997)

[15] Bismut, J.-M.: Martingales, the Malliavin calculus and hypoellipticity under general hörmander's conditions. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **56**, 469–505 (1981) https://doi.org/10.1007/BF00531428

[16] Mirafzali, E., Gupta, U., Wyrod, P., Proske, F., Venturi, D., Marinescu, R.: Malliavin Calculus for Score-based Diffusion Models. arXiv:2503.16917 (2025). https://doi.org/10.48550/arXiv.2503.16917

[17] Pidstrigach, J., Baker, E.L., Domingo-Enrich, C., Deligiannidis, G., Nüsken, N.: Conditioning diffusions using malliavin calculus. In: Forty-second International Conference on Machine Learning (2025)

[18] Greco, G.: A Malliavin-Gamma calculus approach to Score Based Diffusion Generative models for random fields. arXiv:2505.13189 (2025). https://doi.org/10.48550/arXiv.2505.13189

[19] Venturi, D., Wan, X., Mikulevicius, R., Rozovskii, B.L.: Wick-Malliavin approximation to nonlinear stochastic PDEs: Analysis and simulation. Proc. R. Soc. A **469**(2158), 1–20 (2013) https://doi.org/10.1098/rspa.2013.0001

[20] Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., Bodenstein, S.W., Evans, D.A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A.I., Cowie, A., Figurnov, M., Fuchs, F.B., Gladman, H., Jain, R., Khan, Y.A., Low, C.M.R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E.D., Zielinski, M., Žídek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., Jumper, J.M.: Accurate structure prediction of biomolecular interactions with alphafold 3. Nature **630**(8016), 493–500 (2024) https://doi.org/10.1038/s41586-024-07487-w

[21] Nualart, D., Nualart, E.: Introduction to Malliavin Calculus. Institute of Mathematical Statistics Textbooks. Cambridge University Press, Cambridge (2018). https://doi.org/10.1017/9781139856485

[22] Malliavin, P., Thalmaier, A.: Stochastic Calculus of Variations in Mathematical Finance. Springer Finance. Springer, Berlin (2006). https://doi.org/10.1007/3-540-30799-0

[23] Nualart, D.: The Malliavin Calculus and Related Topics. Probability and Its Applications. Springer, Berlin (2006). https://doi.org/10.1007/3-540-28329-3

[24] Horn, R.A., Johnson, C.R.: Matrix Analysis, 2nd edn. Cambridge University Press, Cambridge (2012). https://doi.org/10.1017/CBO9780511810817

[25] Nualart, D., Pardoux, E.: Stochastic calculus with anticipating integrands. Probability Theory and Related Fields **78**(4), 535–581 (1988) https://doi.org/10.1007/BF00353876

[26] Nualart, D., Zakai, M.: Generalized stochastic integrals and the Malliavin calculus. Probability Theory and Related Fields **73**, 255–280 (1986) https://doi.org/10.1007/BF00339940