Cubature Methods and Applications

D. Crisan, K. Manolarakis, and C. Nee

Abstract We present an introduction to a new class of numerical methods for approximating distributions of solutions of stochastic differential equations. The convergence results for these methods are based on certain sharp gradient bounds established by Kusuoka and Stroock under non-Hörmader constraints on diffusion semigroups. These bounds and some other subsequent refinements are covered in these lectures. In addition to the description of the new class of methods and the corresponding convergence results, we include an application of these methods to the numerical solution of backward stochastic differential equations. As it is well-known, backward stochastic differential equations play a central role in pricing financial derivatives.

1 Introduction

Stochastic differential equations (SDEs) are ideal models for the evolution of randomly perturbed dynamical systems. Such systems pervade a variety of areas of human activity, including biology, communications, engineering, finance and physics.

The solution of an SDE is amenable to numerical approximations even in high dimensions. Classical methods such as the Euler method work well provided the distribution of the SDE and the function that we wish to integrate are sufficiently smooth. In particular, when the SDE is driven by non-singular noise, the convergence properties of classical numerical methods are well understood. However, in the 1980s, Kusuoka and Stroock [34] relaxedthe conditions under which some of

D. Crisan (✉) · K. Manolarakis · C. Nee
Department of Mathematics, Imperial College London, 180 Queen's Gate, London SW7 2AZ, UK
e-mail: d.crisan@imperial.ac.uk

F.E. Benth et al., *Paris-Princeton Lectures on Mathematical Finance 2013*, Lecture Notes in Mathematics 2081, DOI 10.1007/978-3-319-00413-6_4,

the smoothness properties of the semigroup associated to the solution of the SDE remain valid. They replaced the classical Hörmander condition requirement by a weaker condition: the so-called UFG condition. Essentially, this condition states that the Lie algebra generated by the vector fields appearing in the noise term of the equation is finite dimensional when viewed as a module over the space of bounded infinitely differentiable functions. Kusuoka and Stroock showed that the semigroup remains smooth in any direction belonging to the above algebra. This fundamental result forms the theoretical basis of a recently developed class of high accuracy numerical methods. In the last 10 years, Kusuoka, Lyons, Ninomiya and Victoir [29,36,49] developed several numerical algorithms based on Chen's iterated integrals expansion. These new algorithms generate approximations to the solution of the SDE in the form of the empirical distribution of a cloud of particles with deterministic trajectories. They work under a weaker condition (termed the UFG condition, see Sect. 2.3 for details) rather than the ellipticity/Hörmander condition and are faster than the corresponding classical methods. Let us describe briefly the framework and structure of these methods:

In the following, let $(\Omega, \mathcal{W}, \mathbb{P})$ be the standard (d-dimensional) Wiener space:

$$\Omega = \{\omega \in \mathcal{C}([0,\infty);\mathbb{R}^d), \omega(0) = 0\}, \qquad \mathcal{W} = \mathcal{B}(C([0,\infty);\mathbb{R}^d)),$$

where $\mathcal{C}([0,\infty);\mathbb{R}^d)$ is the set of $\mathbb{R}^d$-valued continuous paths endowed with the corresponding Borel σ-algebra $\mathcal{B}(C([0,\infty);\mathbb{R}^d))$ and $\mathbb{P}$ is the probability measure such that the coordinate mapping process:

$$B = \{B_t = (B_t^i)_{i=1}^d, t \in [0,\infty)\}, \qquad B_t(\omega) := \omega(t) := (\omega_i(t) : i = 1,\dots,d)$$

is a d-dimensional Brownian motion under $\mathbb{P}$. We define $B_t^0 := t$ for notational simplicity.

Let $V_0, V_1, \dots, V_d \in \mathcal{C}(\mathbb{R}^N;\mathbb{R}^N)$ be $d+1$ Lipschitz vector fields and

$$X = \{X_t^x, t \in [0,\infty), x \in \mathbb{R}^N\}$$

be the solution of the following stochastic differential equation

$$X_t^x = x + \sum_{i=0}^{d} \int_0^t V_i\left(X_s^x\right) dB_s^i. \tag{1}$$

Equation (1) has a unique solution (see, for example, Theorem 2.9 page 289 in[26]). To be more precise, there exists a unique stochastic process adapted with respect to the augmented filtration generated by the Brownian motion B for which identity (1) holds true. The measurability property of X_t is crucial. However, this condition is sometimes overlooked and treated as a rather meaningless theoretical requirement. In effect, the condition means that there is a $\mathcal{B}(C([0,t];\mathbb{R}^d))/\mathcal{B}(\mathbb{R}^N)$-measurable mapping $\alpha_{t,x} : \mathcal{C}([0,t];\mathbb{R}^d) \to \mathbb{R}^N$ such that

$$X_t^x = \alpha_{t,x}(B_{[0,t]}), \quad \mathbb{P} - \text{a.s.} \tag{2}$$

Hence X_t^x is determined by the driving noise $\{B_s, s \in [0,t]\}$. Put differently, if we know B then (theoretically) we will also know the value of X_t^x.[1]

Example 1. For the following equations, one can explicitly write the solution of the SDE as a function of the Brownian motion B:

$$X_t^x = x + \int_0^t aX_s^x dB_s^0 + \int_0^t bX_s^x dB_s^1,\ X_t^x = x\exp\left(bB_t^1 + (a - b^2/2)B_t^0\right) \tag{3}$$

$$X_t^x = x + \int_0^t aX_s^x dB_s^0 + \int_0^t bdB_s^1,\ X_t = xe^{aB_t^0} + b\int_0^t e^{a(B_t^0 - B_s^0)}dB_s^1 \tag{4}$$

$$\begin{pmatrix} X_t^{x,1} \\ X_t^{x,2} \end{pmatrix} = \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} + \int_0^t \begin{pmatrix} a \\ 0 \end{pmatrix} dB_s^1 + \int_0^t \begin{pmatrix} 0 \\ bX_s^1 \end{pmatrix} dB_s^2 \tag{5}$$

$$\begin{pmatrix} X_t^{x,1} \\ X_t^{x,2} \end{pmatrix} = \begin{pmatrix} x^1 + aB_t^1 \\ x^2 + \int_0^t b(x^1 + aB_s^1)dB_s^2 \end{pmatrix} \tag{6}$$

In general it is not possible to have explicit formulae for the solution of the stochastic differential equation, in other words the mapping $\alpha_{t,x}$ appearing in the representation (2) is not known. Hence accurate numerical approximations of X are highly desirable. In particular, we are interested in computing quantities of the form

$$\mathbb{E}[\varphi(X_{t,x})] = \mathbb{E}[\varphi \circ \alpha_{t,x}(B_{[0,t]})] = \int_\Omega \varphi \circ \alpha_{t,x}(\omega)\, \mathbb{P}(d\omega), \tag{7}$$

where φ is a given test function and $\mathbb{P}$ is the probability distribution of the Brownian motion (the Wiener measure). The computation of expectations of the form (7) has particular relevance in mathematical finance through the pricing of financial contracts. Indeed, calculating the expected value of functionals of the solution of a stochastic differential equation (which would be assumed as the model of the underlying price process) in a very short time is a standard problem in finance and is one which has ruled more exotic models out of practical implementation in industry.

Computing quantities of the form (7) is also relevant for the estimation of infinite dimensional random dynamical systems. The theory of infinite dimensional

[1]The process X is uniquely identified by (1) only up to a set of measure 0. Two processes X^1 and X^2 satisfying (1) are indistinguishable: the set $\{\omega \in \Omega | \exists t \in [0,\infty)$ such that $X_t^1(\omega) = X_t^2(\omega)\}$ is a $\mathbb{P}$-null set (has probability zero). Similarly, the identity (2) holds $\mathbb{P}$-almost surely, i.e. there can be a subset of Ω of probability zero where (2) does not hold.

random dynamical systems shares many of the concepts and results with their finite dimensional counter-parts. Many examples are determined by stochastic and deterministic partial differential equations. These partial differential equations have solutions $u(t,x)$ that admit certain representations, called Feynman–Kac representations, in terms of certain functionals integrated with respect to the law of a stochastic process:

$$u(t,x) = \mathbb{E}[\Lambda_{t,x}\left(X^x_{[0,t]}\right)]. \tag{8}$$

A large class of such PDEs exhibit the common feature that the process X appearing in (8) has a representation of the form (2) hence their solution $u(t,x)$ can be represented as

$$u(t,x) = \mathbb{E}[(\Lambda_{t,x}\circ\alpha_{t,x})\left(B_{[0,t]}\right)], \tag{9}$$

where the functional $\Lambda'_{t,x} \equiv \Lambda_{t,x}\circ\alpha_{t,x}$ is nonlinear and, possibly, implicitly defined. Examples, include linear PDEs, semilinear PDEs such as those appearing in the pricing of financial derivatives under trading constraints, McKean–Vlasov equations, Navier–Stokes equation, Burgers equation, Zakai equation, etc.

It follows that the computation of $u(t,x)$ requires the approximation of the law of the process X if the Feynman–Kac formula (8) is used, or the law of the Brownian motion B if one uses (9) instead. However this is not enough. The functionals $\Lambda_{t,x}$ respectively $\Lambda'_{t,x}$ do not have a closed form, in other words they cannot be explicitly described and, more importantly, integrated with respect to the approximating law. One needs to approximate them with versions whose integral with respect to the corresponding approximating law can be easily computed. Obviously, the error of the approximation of the solution of the PDE obtained in this manner will depend on both the error introduced when approximating the functional and that introduced when approximating the law of the process. Care must be taken so as not to compound the corresponding errors. In practice both approximations are performed simultaneously. Nevertheless, when it comes to estimating the approximation error it helps to separate them. The numerical methods discussed in the following entail the following three steps:

- Replacing the law of B with the law of a simpler process $\tilde{B}$. The process $\tilde{B}$ will have bounded variation paths and its so-called "signature" will approximate that of the original B. The support of the law of the process $\tilde{B}_{[0,t]}$ is chosen to have finite support. In other words, there are only a finite number of paths, $\omega_i : [0,t] \to \mathbb{R}^d$, $i = 1,\ldots,n_t$ such that

$$\lambda_{i,t} := \mathbb{P}(\tilde{B}_{[0,t]} = \omega_i) > 0.$$[2]

[2]Of course the sum of the weights $\lambda_{i,t}$ is 1, i.e., $\sum_{i=1}^{n_t}\lambda_{i,t} = 1$.

- Approximating $\Lambda'_{t,x}$ with an explicit/simple version $\tilde{\Lambda}'_{t,x}$. Here we will exploit the smoothness properties of the functional $\Lambda'_{t,x}$. Such properties will be analyzed in the next chapter by using Malliavin Calculus techniques.
- Integrate $\tilde{\Lambda}'_{t,x}$ with respect to the law of $\tilde{B}$. This step consists in computing the average of $\tilde{\Lambda}'_{t,x}$ estimated over the n_t realizations of $\tilde{B}_{[0,t]}$. In essence, we will have

$$u(t,x) \simeq \mathbb{E}\left[\tilde{\Lambda}'_{t,x}\left(\tilde{B}_{[0,t]}\right)\right] = \sum_{i=1}^{n_t} a_{i,t}\tilde{\Lambda}'_{t,x}(\omega_i).$$

If the number of paths n_t contained in the support of $\tilde{B}_{[0,t]}$ is above a threshold that depends on the capabilities of the hardware on which the algorithm is run, then an additional procedure is required to reduce n_t to a manageable size. One can employ a Monte Carlo procedure similar to that used in the classical schemes (e.g. Euler–Maruyama) or the so-called "tree based branching algorithm" [14], a minimal variance selection procedure analysed in Sect. 3.

To understand the choice of the simple process $\tilde{B}$, let us introduce briefly the classical Euler–Maruyama method.[3] For this we choose a partition Π of a generic interval, say, $[0,T]$

$$\Pi \ : \quad 0 = \tau_0 < \tau_1 < \ldots < \tau_n < \ldots \tau_N = T.$$

and we denote by δ the mesh of the partition $\delta = \max_{i=1,\ldots,N}(\tau_i - \tau_{i-1})$. We do not specify the choice of the partition. However if the partition is equidistant, then $\delta = T/N$ and it is also called the time step. Let $Y^x = \{Y_t^x,\ t \in [0,T]\}$ be the continuous time process satisfying the evolution equation

$$Y_t^x = Y_{\tau_n}^x + \int_{\tau_n}^t V_0\left(Y_s^x\right)ds + \sum_{i=1}^d \frac{1}{\sqrt{\tau_{n+1}-\tau_n}}\int_{\tau_n}^t V_i\left(Y_s^x\right)\xi_n^i ds, \quad t \in [\tau_n, \tau_{n+1}], \tag{10}$$

where $\{\xi_n^i,\ i = 1,\ldots,d,\ n = 0,\ldots,N-1\}$ are mutually independent random variables whose moments match the moments of a standard Gaussian random variable up to order 3 and with initial value $Y_0^x = x$. More precisely we require that the random variables ξ_n^i must be independent, with moments satisfying,

$$\mathbb{E}\left[\xi_n^i\right] = \mathbb{E}\left[\left(\xi_n^i\right)^3\right] = 0, \quad \mathbb{E}\left[\left(\xi_n^i\right)^2\right] = 1. \tag{11}$$

In particular, ξ_n^i can be chosen to have the Bernoulli distribution

---

[3] To be more precise, following the phraseology of [27], we describe here the *simplified weak* Euler scheme for a scalar SDE driven by a multi-dimensional noise.

$$P\left(\xi_n^i = \pm 1\right) = \frac{1}{2}. \tag{12}$$

We can recast the evolution equation of the process Y^x in a similar manner to that of X^x. Let $\tilde{B} = \{\tilde{B}_t, t \in [0, \infty)\}$ be the d-dimensional stochastic process[4]

$$\tilde{B}_t = \xi_{[Nt]}(t - \tau_{[NT]}) + \sum_{n=1}^{[Nt]} \xi_{n-1}\sqrt{\tau_n - \tau_{n-1}}, \tag{13}$$

where the last term is chosen to be 0 if $[NT] = 0$ and $\{\xi_n,\ n = 0, \ldots, N\}$ are d-dimensional random vectors with corresponding entries $\xi_n = \left(\xi_n^1, \ldots, \xi_n^d\right)$. Then $\tilde{B}$ has piecewise-linear trajectories and, if we use an equidistant partition, the support of $\tilde{B}_t$ has $n_t = 2^n$ paths for $t \in (\tau_{n-1}, \tau_n], n = 1, \ldots, [NT]$. Then Y is the solution of the following *ordinary* differential equation

$$Y_t^x = x + \sum_{i=0}^{d} \int_0^t V_i\left(Y_s^x\right) d\tilde{B}_s^i, \tag{14}$$

where, as in (1), we defined $\tilde{B}_t^0 := t$. Under suitable conditions, the process Y^x, is a first order approximation of the equation (1) associated with the partition Π (see, for example Theorem 14.1.5, page 460 in [27]). More precisely, we have

$$|\mathbb{E}[\varphi(X_t^x)] - \mathbb{E}[\varphi(Y_t^x)]| \leq C_\varphi \delta, \quad t \in [0, T].$$

The paths $\{\omega_1, \ldots, \omega_{n_t}\}$ in the support of $\tilde{B}$ are the realizations of a random walk (linearly interpolated between jumps). Then $\lambda_{i,t} := \mathbb{P}(\tilde{B}_{[0,t]} = \omega_i) = \frac{1}{n_t}$ and

$$\mathbb{E}[\varphi(Y_t^x)] = \sum_{i=1}^{n_t} \lambda_{i,t}\varphi(Y_t^{x,i}),$$

where $Y^{x,i}$ is the solution of the ordinary differential equation (14) corresponding to the path ω_i. That is

$$Y_t^{x,i} = x + \sum_{j=0}^{d} \int_0^t V_j\left(Y_s^{x,i}\right) d\omega_i^j\,(s).$$

If the ordinary differential equation (14) has no explicit solution, one can choose without loss of accuracy, a process Z^x which satisfies an explicit/implicit

[4]In (13) and subsequently, $[z]$ denotes the integer part of $z \in \mathbb{R}$.

discretization of (14). For example

$$Z^x_{\tau_{n+1}} = Z^x_{\tau_n} + V_0\left(Z^x_{\tau_n}\right)(\tau_{n+1} - \tau_n) + \sum_{i=1}^{d} V_i\left(Z^x_{\tau_n}\right)\xi^i_n\sqrt{\tau_{n+1} - \tau_n}. \tag{15}$$

The solution of (15) is customarily called the Euler–Maruyama approximation of X and has the same order of approximation as Y (order 1). The ODE (14) has solutions that evolve in the support of the original diffusion so it manifests good numerical stability conditions. Classical higher order approximations of (1) such as those described in Chaps. 14 and 15 in [27] no longer have this property. The question that arises is whether it would be possible to produce a high order approximation that still has this property. The answer is yes and this is exactly what a cubature method does. One can replace the process $\tilde{B}$ by a "better" approximation of B which, in turn, will lead to a high order approximation of the solution of (1). To understand in what sense $\tilde{B}$ is an approximation of B and how can it be improved we need to explain in brief the concept of a signature of a path. Let

$$T\left(\mathbb{R}^d\right) = \bigoplus_{i=0}^{\infty}\left(\mathbb{R}^d\right)^{\otimes i}, \quad T^{(m)}\left(\mathbb{R}^d\right) = \bigoplus_{i=0}^{m}(\mathbb{R}^d)^{\otimes i}$$

be the tensor algebra of all non-commutative polynomials over $\mathbb{R}^d$ and, respectively, the tensor algebra of all non-commutative polynomials of degree less than $m+1$. For a path $\omega : [0,\infty) \to \mathbb{R}^d$ with finite variation we define its signature $S_{s,t}(\omega) \in T\left(\mathbb{R}^d\right)$ to be the corresponding Chen's iterated integrals expansion:

$$S_{s,t}(\omega) = \sum_{k=0}^{\infty}\int_{s<t_1\ldots t_k<t} d\omega_{t_1}\otimes\ldots\otimes d\omega_{t_k},$$

where

$$\int_{0<t_1\ldots t_k<t} d\omega_{t_1}\otimes\ldots\otimes d\omega_{t_k} := \sum_{i_1,\ldots,i_k}\left(\int_{0<t_1\ldots t_k<t} d\omega^{i_1}_{t_1}\ldots d\omega^{i_k}_{t_k}\right)e_{i_1}\otimes\ldots\otimes e_{i_k},$$

and $(e_{i_1}\otimes\ldots\otimes e_{i_k}), i_1,\ldots,i_k \in \{1,\ldots,d\}$, is the canonical basis of $(\mathbb{R}^d)^{\otimes k}$. Similarly we define its truncated signature $S_{s,t}(\omega) \in T^{(m)}\left(\mathbb{R}^d\right)$ to be

$$S^m_{s,t}(\omega) = \sum_{k=0}^{m}\int_{s<t_1\ldots t_k<t} d\omega_{t_1}\otimes\ldots\otimes d\omega_{t_k}.$$

Similarly the (random) signature and, respectively, the truncated signature of the Brownian motion are

$$S_{s,t}(B) = \sum_{k=0}^{\infty} \int_{s<t_1 \ldots t_k<t} dB_{t_1} \otimes \ldots \otimes dB_{t_k}, \; S^m_{s,t}(B) = \sum_{k=0}^{m} \int_{s<t_1 \ldots t_k<t} dB_{t_1} \otimes \ldots \otimes dB_{t_k}. \tag{16}$$

In (16), the stochastic (iterated) integrals are of Stratonovitch type.

The expected value of $S_{s,t}(B)$ uniquely identifies the law of B, i.e., the Wiener measure.[5] Moreover, if $\hat{B}$ is a process such that

$$\mathbb{E}\left[S^m_{k\delta,(k+1)\delta}(B)\right] = \mathbb{E}[S^m_{k\delta,(k+1)\delta}(\hat{B})], k = 0, 1, \ldots, N-1, \tag{17}$$

then for certain classes of functionals Λ', $\mathbb{E}[\Lambda'(B')]$ is a high order approximation of $\mathbb{E}[\Lambda'(B)]$. In particular, if $\Lambda'_{t,x}$ is the functional that gives the solution of the SDE (1) for $t = N\delta$, i.e., $\Lambda'_{t,x}(B) = \varphi(X^x_t)$, then

$$|\mathbb{E}[\varphi(X^x_t)] - \mathbb{E}[\varphi(Y^x_t)]| \le C_\varphi \delta^{\frac{m-1}{2}}, \tag{18}$$

where Y^x_t is the solution of the ordinary differential equation (14) driven by $\tilde{B}$. We prove this result in Sect. 3 of the current lecture notes. In particular, the process $\tilde{B}$ as defined (13) satisfies (17) with $m = 3$.

The proof of (18), requires the smoothness of the (diffusion) semigroup $\{P_t, \;\; t \in [0, \infty)\}$ defined as

$$(P_t\varphi)(x) = \mathbb{E}[\varphi(X^x_t)], \quad x \in \mathbb{R}^d, \quad t \ge 0,$$

where φ is an appropriately chosen test function. If the vector fields satisfy the *ellipticity* or, more generally, the *uniform Hörmander* condition, $P_t\varphi$ is smooth for any bounded measurable function φ and $t > 0$. Many of the classical numerical schemes rely on this property and so Hörmander's paper [24] is a major contribution to this field. A probabilistic version of this result led Malliavin [40] to develop his celebrated stochastic calculus of variations through which one can prove, probabilistically, the sufficiency of Hörmander's condition.

The work of Kusuoka and Stroock [32–34] in the 1980s provided an extension of Malliavin's results. In it, they proved precise gradient bounds that are valid under a general condition termed the UFG condition, see Sect. 2.3 for details. The UFG condition imposed on the vector fields $\{V_i, i = 0, \ldots, d\}$ essentially states that the $C^\infty_b(\mathbb{R}^d)$-module $\mathcal{M}$ generated by the vector fields $\{V_i, i = 1, \ldots, d\}$ within the Lie algebra generated by $\{V_i, i = 0, \ldots, d\}$ is finite dimensional. The UFG condition implies Hörmander's hypoellipticity condition, but not viceversa. There are explicit examples for which Hörmander's condition fails to hold, but for which the UFG condition is satisfied (see Example 15). In particular, the condition does not require that the vector space $\{W(x)|W \in \mathcal{M}\}$ is homeomorphic to $\mathbb{R}^d$ for

[5] See Proposition 118 in [18].

any $x \in \mathbb{R}^d$. Moreover, under the UFG condition, the dimension of the space $\{W(x)|W \in \mathcal{M}\}$ is not required to be constant over $\mathbb{R}^d$. Such generality makes any Frobenius type approach to prove smoothness of the solution very difficult. Indeed the authors are not aware of any alternative proof of the smoothness results of the solution of $P_t\varphi$ (under the UFG condition) other than that given by Kusuoka and Stroock. Kusuoka and Stroock prove that, under the UFG condition, $P_t\varphi$ is differentiable in the direction of any vector field W belonging to $\mathcal{M}$ and deduce precise gradient bounds of the form:

$$\|W_1 \ldots W_k P_t\varphi\|_\infty \le \frac{C^k}{t^l}\|\varphi\|_p, \tag{19}$$

where l is a constant that depends explicitly on the vector fields $W_i \in \mathcal{M}$, $i = 1, \ldots, k$ and $\|\varphi\|_p$ is the standard L_p norm of the function φ.

Whilst the Kusuoka–Stroock result does not suffice to justify the convergence of classical numerical schemes, it is tailor-made for the cubature methods. The global error of numerical schemes depends intrinsically on the smoothness of $P_t\varphi$, but only in the direction of the vector fields W belonging to $\mathcal{M}$. As a result, the cubature methods are proved to work under the more general UFG condition, unlike the classical numerical methods.

The lecture notes are structured as follows: In the following section, we provide a “clean” treatment of the (sharp) gradient bounds of the type (19) deduced under the minimal smoothness requirements on imposed on the vector fields $\{V_i, i = 0, \ldots, d\}$. Such results are intrinsically related to the solution of the linear parabolic partial differential equation

$$\partial_t u(t,x) = \frac{1}{2}\sum_{i=1}^{d} V_i^2 u(t,x) + V_0 u(t,x), \quad (t,x) \in (0,\infty) \times \mathbb{R}^d. \tag{20}$$

We show how the Kusuoka–Stroock approach can be used to recover the smoothness of the solution of (20) under the Hörmander condition. In the Hörmander case, it is straightforward to show that $P_t\varphi$ is indeed the (unique) classical solution of (20) with φ being the initial condition of the PDE. In particular we show that u is differentiable in any direction including direction V_0. The situation is more delicate in the absence of the Hörmander condition. Under the UFG condition, (20) may not have a solution in the classical sense. As explained in [44], it turns out that $P_t\varphi$ remains differentiable in the direction $\mathcal{V}_0 := \partial_t - V_0$ when viewed as a function $(t,x) \to P_t\varphi(x)$ over the product space $(0,\infty) \times \mathbb{R}^d$. This together with the continuity at $t = 0$ implies that $P_t\varphi$ is the unique (classical) solution of the equation

$$\mathcal{V}_0 u(t,x) = \frac{1}{2}\sum_{i=1}^{d} V_i^2 u(t,x), \quad (t,x) \in (0,\infty) \times \mathbb{R}^d. \tag{21}$$

In Sect. 3, we incorporate cubature methods into a larger class of methods, and deduce their convergence rates under the UFG condition and an additional constraint called *the* V_0 *condition*. We also deduce the convergence rates of the cubature methods combined with an algorithm for controlling the computational effort—the tree based branching algorithm (or TBBA for short). The section is concluded with an application of the cubature and TBBA method to the approximation of a call option on a Heston model price process.

Section 4 is dedicated to the application of cubature methods to the numerical solution of backward stochastic differential equations.

The lecture notes are concluded with an appendix comprising a number of technical lemmas and a proof of the convergence of the cubature method in the absence of the V_0 condition.

2 Sharp Gradient Bounds

In this chapter we give a full and self-contained proof of Kusuoka's gradient bounds (cf. [30]). The main difference between what is done there and what is presented here, is that we relax the restrictive assumptions on the SDE coefficients (in [30] they are assumed to be smooth and uniformly bounded). In later chapters, we shall apply these results to prove convergence of the cubature method.

2.1 Framework

Recall that $(\Omega, \mathcal{W}, \mathbb{P})$ is the standard (d-dimensional) Wiener space:

$$\Omega = \{\omega \in \mathcal{C}([0,\infty);\mathbb{R}^d), \omega(0) = 0\}, \qquad \mathcal{W} = \mathcal{B}(\mathcal{C}([0,\infty);\mathbb{R}^d)),$$

where $\mathcal{C}([0,\infty);\mathbb{R}^d)$ is the set of $\mathbb{R}^d$-valued continuous paths endowed with the uniform norm topology, $\mathcal{W}$ is the corresponding Borel σ-algebra $\mathcal{B}(\mathcal{C}([0,\infty);\mathbb{R}^d))$ and $\mathbb{P}$ is the probability measure such that the coordinate mapping process:

$$B = \{B_t, t \in [0,\infty)\}, \qquad B_t(\omega) := \omega(t) := (\omega_i(t) : i = 1,\ldots,d)$$

is a d-dimensional Brownian motion under $\mathbb{P}$. We define $B_t^0 := t$ for notational simplicity.

Let k be a positive integer to be determined at a later stage. Assume that $V_1,\ldots,V_d \in \mathcal{C}_b^{k+1}(\mathbb{R}^N;\mathbb{R}^N)$[6] and $V_0 \in \mathcal{C}_b^k(\mathbb{R}^N;\mathbb{R}^N)$ are $d+1$ vector fields and let

[6]For any positive integer m, the set $\mathcal{C}_b^m(\mathbb{R}^a;\mathbb{R}^b)$ is the set of all bounded continuous functions $\varphi : \mathbb{R}^a \to \mathbb{R}^b$, m-times continuously differentiable with all derivatives bounded.

$X = \{X_t^x, t \in [0, \infty), x \in \mathbb{R}^N\}$ be the following stochastic flow

$$X_t^x = x + \sum_{i=0}^{d} \int_0^t V_i(X_s^x) \circ dB_s^i. \tag{22}$$

In (22) the stochastic integrals $\int_0^t V_i(X_s^x) \circ dB_s^i$, $i = 1, \dots, d$ are Stratonovitch integrals whereas $\int_0^t V_0(X_s^x) \circ dB_s^0$ is a standard Riemann integral.

Remark 2. In the following, we will view the vector fields $V_0, V_1, \dots, V_d$ as both vector-valued functions and first order differential operators defined as follows: for $V_i(x) = (V_i^1(x), \dots, V_i^N(x))^\top$ the corresponding first order differential operator will be

$$V_i = \sum_{j=1}^{N} V_i^j \partial_j, \qquad V_i f(x) = \nabla f(x) V_i(x), \quad \textit{where} \quad \nabla f(x) = (\partial_1 f(x), \dots, \partial_N f(x)).$$

Using this notation, from (22) we have the standard chain rule

$$f(X_t^x) = f(x) + \sum_{i=0}^{d} \int_0^t V_i f(X_s^x) \circ dB_s^i$$

for any $f \in \mathcal{C}_b^3(\mathbb{R}^N, \mathbb{R})$. We remark that the different levels of differentiability chosen for V_0 and $V_1, \dots, V_d$ ensure that the corresponding Itô equation has $\mathcal{C}_b^k(\mathbb{R}^N; \mathbb{R}^N)$ coefficients.

It is a classical result that the stochastic flow $X = \{X_t^x, t \in [0, \infty), x \in \mathbb{R}^N\}$ is differentiable in the space variable x. See for example Kunita [28] or Nualart [51, Theorem 2.2.1, p. 119]. We state the required result in the following:

Theorem 3. *Let $X = \{X_t^x, t \in [0, \infty), x \in \mathbb{R}^N\}$ be the solution of (22). Then X has a modification (again denoted by X) such that the mapping*

$$x \in \mathbb{R}^N \longrightarrow X_t^x \in \mathbb{R}^N$$

is k-times continuously differentiable, for each t, P-almost surely. Moreover the Jacobian of $X_t^{(\cdot)}$ at x, $J_t^{(\cdot)} := (\partial_j X_t^{i,(\cdot)})_{1 \le i,j \le N}$ satisfies the matrix stochastic differential equation[7]:

[7] In (23) and subsequently, ∂V_i is the matrix valued map $\partial V_i := (\partial_n V_i^m)_{1 \le n,m \le N}$.

$$\begin{cases} dJ_t^x = \sum_{i=0}^d \partial V_i(X_t^x) J_t^x \circ dB_t^i, \\ J_0^x = I. \end{cases} \tag{23}$$

The Jacobian is almost surely invertible (as a matrix) and its inverse, $(J_t^x)^{-1}$*, satisfies the SDE*

$$\begin{cases} d(J_t^x)^{-1} = -\sum_{i=0}^d (J_t^x)^{-1} \partial V_i(X_t^x) \circ dB_t^i, \\ (J_0^x)^{-1} = I. \end{cases} \tag{24}$$

In addition, the following integrability result holds

$$\sup_{t\in[0,T]} \mathbb{E}\left[\left|\frac{\partial^{|\gamma|} X_t^x}{\partial x^\gamma}\right|^p\right] < C_{T,p}, \quad \forall\, p \geq 1, T > 0, 0 < |\gamma| \leq k, \forall x \in \mathbb{R}^N. \tag{25}$$

2.2 Malliavin Differentiation

For an absolutely continuous path $h \in \mathcal{C}([0,\infty);\mathbb{R}^d)$, we denote by h' its derivative. Let H be the space

$$H = \{h \in \Omega,\ h \text{ absolutely continuous},\ h' \in L^2([0,\infty);\mathbb{R}^d)\} \subset \Omega.$$

H is endowed with a Hilbert structure under the inner product

$$\langle h, g\rangle_H := \langle h', g'\rangle_{L^2([0,\infty);\mathbb{R}^d)} := \int_0^\infty h'(u) \cdot g'(u) du$$

and is called the *Cameron–Martin* space. We use this space to define the Malliavin derivative.

Definition 4 (Malliavin Derivative). Let $f \in \mathcal{C}_b^\infty(\mathbb{R}^n, \mathbb{R})$, $h_1, \ldots, h_n \in H$ and $F : \Omega \to \mathbb{R}$ be the functional given by:

$$F(\omega) = f\left(\int_0^\infty h_1'(t) dB_t(\omega), \ldots, \int_0^\infty h_n'(t) dB_t(\omega)\right), \tag{26}$$

where, for any $h_i' = (h_{i,1}', \ldots, h_{i,d}')$,

$$\int_0^\infty h_i'(t) dB_t := \sum_{j=0}^d \int_0^\infty h_{i,j}'(t) dB_t^j.$$

Any functional of the form (26) is called *smooth* and we denote the class of all such functionals by $\mathcal{S}$. Then the Malliavin derivative of F, denoted by $DF \in L^2(\Omega; H)$ is given by:

$$DF = \sum_{i=1}^{n} \partial_i f \left(\int_0^\infty h_1'(u) dB_u, \ldots, \int_0^\infty h_n'(u) dB_u \right) h_i \tag{27}$$

We will often make use of the notation: $D_h F := \langle DF, h \rangle_H$ for $h \in H$. Observe that $D_h F$ is the directional derivative of F in the direction h as

$$\begin{aligned} D_h F(\omega) &= \sum_{i=1}^{d} \partial_i f \left(\int_0^\infty h_1'(u) dB_u(\omega), \ldots, \int_0^\infty h_n'(u) dB_u(\omega) \right) \langle h_i, h \rangle_H \\ &= \frac{d}{d\epsilon} f \left(\int_0^\infty h_1'(u) dB_u(\omega) + \epsilon \langle h_1', h' \rangle_{L^2([0,\infty))}, \right. \\ &\quad \left. \ldots, \int_0^\infty h_n'(u) dB_u(\omega) + \epsilon \langle h_n', h' \rangle_{L^2([0,\infty))} \right) \bigg|_{\epsilon=0}. \end{aligned}$$

and, since $B_t(\omega + \varepsilon h) = B_t(\omega) + \varepsilon h(t)$, this yields

$$dB_t(\omega + \varepsilon h) = dB_t(\omega) + \varepsilon h'(t)\, dt.$$

Hence

$$\begin{aligned} D_h F(\omega) &= \frac{d}{d\varepsilon} f \left(\int_0^\infty h_1'(u)\, dB_u(\omega + \varepsilon h), \ldots, \int_0^\infty h_n'(u)\, dB_u(\omega + \varepsilon h) \right) \bigg|_{\varepsilon=0}. \\ &= \frac{d}{d\epsilon} F(\omega + \epsilon h) \bigg|_{\epsilon=0}. \end{aligned} \tag{28}$$

If $F \in \mathcal{S}$ and $h \in H$, then the following basic integration by parts formula holds

$$\mathbb{E}\left[F \int_0^\infty h'(t) dB_t \right] = \mathbb{E}[\langle DF, h \rangle_H]. \tag{29}$$

The proof of this formula is very simple: It uses an integration by parts formula for the finite dimensional Gaussian density (see, e.g., Lemma 1.2.1 in Nualart [51]).

The set of smooth functionals (random variables) $\mathcal{S}$ is dense in $L^p(\Omega)$, for any $p \geq 1$. That is, for any $F \in L^p(\Omega)$ there exists $\{F_n\} \subset \mathcal{S}$ such that

$$\| F_n - F \|_{L^p(\Omega)} \to 0.$$

This result is available in, for example, Nualart [51]. Its proof relies on showing that a subset of $\mathcal{S}$ (the Wiener polynomials) is dense in $L^p(\Omega)$. This is done by using

Hermite polynomials and the Wiener–Itô chaos expansion. The density property of $\mathcal{S}$ is used to extend the definition of the Malliavin derivative to the set of all square integrable random variable for which there exist an approximating sequence of smooth random variables such that the corresponding Malliavin derivatives converge too. This approach works as the Malliavin derivatives of two convergent sequences of smooth random variables converging to the same $L^2(\Omega)$-limit have the same $L^2([0,\infty)\times\Omega)$-limit. To be more precise we have the following (see, e.g., Nualart [51]) :

Proposition 5 (Closability of the Malliavin Derivative operator). *The Malliavin derivative, a linear unbounded operator $D : \mathcal{S} \to L^2([0,\infty)\times\Omega;\mathbb{R}^d)$ is closable as an operator from $L^2(\Omega;\mathbb{R}^d)$ into $L^2([0,\infty)\times\Omega;\mathbb{R}^d)$. In other words if $\{F_n\}\subset\mathcal{S}$ is a sequence of smooth random variables such that: $\|F_n\|_{L^2(\Omega)}\to 0$ and $\|DF_n\|_{L^2([0,\infty)\times\Omega)}$ is convergent then it follows that*

$$\|DF_n\|_{L^2([0,\infty)\times\Omega)}\to 0.$$

More generally, the Malliavin derivative operator is closable as an operator from $L^p(\Omega;\mathbb{R}^d)$ into $L^p(\Omega;H)$ for any $p\geq 1$. For $p\neq 2$ we use with the norm:

$$\|DF\|^p_{L^p(\Omega;H)} := \mathbb{E}\left[\|DF\|^p_H\right]. \tag{30}$$

The proof of the closability of the Malliavin operator relies on the basic integration by parts formula (29).

We denote by $\mathbb{D}^{1,p}$ the domain of the Malliavin derivative operator as an operator from $L^p(\Omega;\mathbb{R}^d)$ into $L^p(\Omega;H)$ for any $p\geq 1$. More precisely, $\mathbb{D}^{1,p}$ is the closure of the set $\mathcal{S}$ within $L^p(\Omega;\mathbb{R}^d)$ with respect to the norm:

$$\|F\|_{\mathbb{D}^{1,p}} = \left(\mathbb{E}[|F|^p]+\mathbb{E}[\|DF\|^p_H]\right)^{\frac{1}{p}}.$$

The higher order Malliavin derivatives are defined in a similar manner. For smooth random variables, the iterated derivative D^kF, $k\geq 2$, is a random variable with values in $H^{\otimes k}$ defined as

$$D^kF := \sum_{i_1,\ldots,i_k=1}^{n}\partial_{i_1,\ldots,i_k}f\left(\int_0^\infty h_1'(u)dB_u,\ldots,\int_0^\infty h_n'(u)dB_u\right)h_{i_1}\otimes\ldots\otimes h_{i_k},$$

where $h_i(.) := \int_0^{\cdot} h_i'(s)ds$. The above expression for D^kF coincides with that obtained by iteratively applying the Malliavin differential operation. Indeed, for $h\in H$, $F\in\mathcal{S}$, it is easily seen that $D_hF\in\mathcal{S}$. As per (28), it can be shown that,

$$D_{h_k}D_{h_{k-1}}\ldots D_{h_1}F = \langle D^kF, h_1\otimes\ldots\otimes h_k\rangle_{H^{\otimes k}}.$$

In an analogous way, one can close the operator D^k from $L^p(\Omega)$ to $L^p(\Omega; H^{\otimes k})$. So, for any $p \geq 1$ and natural $k \geq 1$, we define $\mathbb{D}^{k,p}$ to be the closure of $\mathcal{S}$ with respect to the norm:

$$\|F\|^p_{\mathbb{D}^{k,p}} := \mathbb{E}[|F|^p] + \sum_{j=1}^{k} \mathbb{E}[\|D^j F\|^p_{H^{\otimes j}}].$$

Note that for $p = 2$ the following isometry holds $L^p(\Omega \times [0,\infty)^k; \mathbb{R}^d) \simeq L^2(\Omega; H^{\otimes k})$. Hence one may identify $D^k F$ as a process: $D^k_{t_1,\ldots,t_k} F$.

A random variable F is said to be *smooth in the Malliavin sense* if $F \in \mathbb{D}^{k,p}$ for all $p \geq 1$ and all $k \geq 1$. We denote by $\mathbb{D}^\infty$ the set of all smooth random variables in the Malliavin sense. For example, the solution X^x_t to (22) satisfies $X^i_t \in \mathbb{D}^{k,p}$ for all $t \in [0,\infty)$ and $p \geq 1$ provided $V_0, \ldots, V_d \in \mathcal{C}^\infty_b(\mathbb{R}^N; \mathbb{R}^N)$ (see Theorem 8 below).

Moreover, there is nothing which restricts consideration to $\mathbb{R}^d$-valued random variables. Indeed, one can consider more general Hilbert space-valued random variables, and the theory would extend in an appropriate way. To this end, denote $\mathbb{D}^{k,p}(E)$ to be the appropriate space of E-valued random variables, where E is some separable Hilbert space. For more details, see [51], where also the proof of the following chain rule formula can be found:

Proposition 6 (Chain Rule for the Malliavin Derivative). *If $\varphi : \mathbb{R}^m \to \mathbb{R}$ is a continuously differentiable function with bounded partial derivatives, and $F = (F_1, \ldots, F_m)$ is a random vector with components belonging to $\mathbb{D}^{1,p}$ for some $p \geq 1$. Then $\varphi(F) \in \mathbb{D}^{1,p}$, with*

$$D\varphi(F) = (\nabla\varphi)(F)DF = \sum_{i=1}^{m} \partial_i \varphi(F) DF_i,$$

where $\nabla\varphi$ is the row vector $(\partial_1\varphi, \ldots, \partial_n\varphi)$ and DF is the (column) vector $(DF_1, \ldots, DF_n)^\top$.

Lemma 7 (The Malliavin derivative and integration). *Assume that E is a separable real Hilbert space. Consider $f : [0,\infty) \times \Omega \to E$, and suppose that for each $t \in [0,T]$ we have $f(t) \in \mathbb{D}^{1,2}(E)$ and $t \to f(t)$ is adapted with respect to the natural filtration of B.*[8] *Moreover, suppose that:*

$$\mathbb{E}\int_0^T \|f(t)\|^2_E dt < \infty \qquad \mathbb{E}\int_0^T \|Df(t)\|^2_{E\otimes H} dt < \infty \tag{31}$$

[8] Although not used in the sequel, the result holds for general $f : [0,\infty) \times \Omega \to E$ such that $f(t) \in \mathbb{D}^{1,2}(E)$ for any $t \in [0,T]$, i.e., not necessarily adapted with respect to the natural filtration of B. In this case, the $F_i(T)$ is the Skorohod integral and not the Itô integral of f. See, for example, Proposition 1.38 page 43 in [51].

Then $F_i(T) := \int_0^T f(t)dB_t^i \in \mathbb{D}^{1,2}(E)$ *for all* $i = 0, 1, \ldots, d$*, with*

$$DF_0(T) = \int_0^T Df(t)dB_t^0$$
$$DF_i(T) = \int_0^T Df(t)dB_t^i + \int_0^{T\wedge\cdot} f(s)ds, \quad i = 1, \ldots, d.$$

Also

$$D_h F_0(T) = \int_0^T D_h f(t)dB_t^0$$
$$D_h F_i(T) = \int_0^T D_h f(t)dB_t^i + \int_0^T f(t)h_i'(t)dt, \quad i = 1, \ldots, d.$$

Moreover, assuming that

$$\mathbb{E}\int_0^T \|D^{k-1}f(t)\|^2_{E\otimes H^{\otimes(k-1)}}dt < \infty, \qquad \mathbb{E}\int_0^T \|D^k f(t)\|^2_{E\otimes H^{\otimes k}}dt < \infty$$

one has for the iterated Malliavin derivative operator D^k*:*

$$D^k F_0 = \int_0^T D^k f(t)dB_t^0$$
$$D^k F_i(T) = \int_0^T D^k f(t)dB_t^i + \int_0^{T\wedge\cdot} D^{k-1}f(s)ds, \quad i = 1, \ldots, d.$$

Proof. The proof is done using an induction argument. See Kusuoka and Stroock [32] for details. □

Theorem 8. *Assume* X *is the stochastic flow which solves* (22)*, where the coefficients* $V_1, \ldots, V_d \in \mathcal{C}_b^{k+1}(\mathbb{R}^N;\mathbb{R}^N)$ *and* $V_0 \in \mathcal{C}_b^k(\mathbb{R}^N;\mathbb{R}^N)$*. Then* $X_t^{x,i} \in \mathbb{D}^{k,p}$ *for all* $t \in [0,\infty)$*,* $i = 1, \ldots, N$ *and* $p \geq 1$*. Furthermore, the matrix valued process* $DX_t^x := (D^j X_t^{x,i})_{i=1,\ldots,N;\, j=1,\ldots,d}$ *satisfies the stochastic differential equation:*

$$DX_t^x = \sum_{i=0}^d \int_0^t \partial V_i(X_u^x)DX_u^x \circ dB_u^i + \left(\int_0^{t\wedge\cdot} V_j(X_u^x)du\right)_{j=1,\ldots,d}. \tag{32}$$

Hence,

$$D_h X_t^x = \sum_{i=0}^d \int_0^t \partial V_i(X_u^x)D_h X_u^x \circ dB_u^i + \sum_{k=1}^d \int_0^t V_k(X_u^x)h_i'(u)du. \tag{33}$$

If the vector fields $V_0, \ldots, V_d$ are uniformly bounded then the following bound on the norms of the derivatives can be shown to hold:

$$\sup_{\substack{t\in[0,T]\\ x\in\mathbb{R}^N}} \mathbb{E}\left[\| D^k X_t^x \|_{H^{\otimes k}}^p \right] < C_{k,p}, \quad \forall\ p \in [1,\infty),\ T>0. \tag{34}$$

If, however, the vector fields $V_0, \ldots, V_d$ are globally Lipschitz continuous but not necessarily bounded, then it may only be deduced that the following holds:

$$\sup_{t\in[0,T]} \mathbb{E}\left[\| D^k X_t^x \|_{H^{\otimes k}}^p \right] < C_{k,p}(1+|x|)^p, \quad \forall\ p \in [1,\infty),\ T>0. \tag{35}$$

Proof. See Nualart [51, pp. 119–124]. □

Corollary 9. *For any $(t,x) \in [0,\infty) \times \mathbb{R}^N$, we have that*

$$(J_t^x)^{-1} D X_t^x = \left(\int_0^{t\wedge .} (J_s^x)^{-1} V_j(X_s^x) ds \right)_{j=1,\ldots,d}. \tag{36}$$

Proof. This is a simple result of applying integration by parts to the product $(J_t^x)^{-1} D X_t^x$, using the SDEs from the respective processes. For a complete proof see, for example, Nualart [51, Sect. 2.3.1]. □

Definition 10 (Lie Bracket of Vector Fields). Let $V, W \in \mathcal{C}^1(\mathbb{R}^N;\mathbb{R}^N)$ be two vector fields. The Lie bracket of V and W is a third vector field, $[V,W]$, defined by:

$$[V,W] := \partial W.V - \partial V.W,$$

where $\partial V := (\partial_j V^i)_{1\le i,j\le N}$ and the multiplication is that of a matrix by a vector.

The Lie bracket is a bilinear differential form $[.,.] : \mathcal{C}^{m_1} \times \mathcal{C}^{m_2} \to \mathcal{C}^{m_1\wedge m_2 - 1}$, where $1 \le m_1, m_2 \le \infty$, which satisfies the identities:

$$[V,W] = -[W,V] \quad \text{and} \quad [U,[V,W]] + [W,[U,V]] + [V,[W,U]] = 0.$$

The latter is known as the *Jacobi Identity*.

Corollary 11. *Let $W \in \mathcal{C}^3(\mathbb{R}^N;\mathbb{R}^N)$ then there holds:*

$$d\left[(J_t^x)^{-1} W(X_t^x) \right] = -\sum_{k=0}^{d} (J_t^x)^{-1} [W, V_k](X_t^x) \circ dB_t^k. \tag{37}$$

Proof. Note that

$$W(X_t^x) = W(x) + \sum_{k=0}^{d} \int_0^t \partial W(X_s^x) V_k(X_s^x) \circ dB_s^k.$$

Thus, by an analogous formula for matrix–vector SDEs we have:

$$\begin{aligned}
(J_t^x)^{-1} W(X_t^x) &= \int_0^t (J_s^x)^{-1} \circ dW(X_s^x) + \int_0^t d(J_s^x)^{-1} \circ W(X_s^x) \\
&= \sum_{k=0}^{d} \int_0^t (J_s^x)^{-1} \partial W(X_s^x) V_k(X_s^x) \circ dB_s^k \\
&\quad - \sum_{k=0}^{d} \int_0^t (J_s^x)^{-1} \partial V_k(X_s^x) W(X_s^x) \circ dB_s^k \\
&= - \sum_{k=0}^{d} \int_0^t (J_s^x)^{-1} [W, V_k](X_s^x) \circ dB_s^k.
\end{aligned}$$

□

The alternative representation (36) for $(J_t^x)^{-1} DX_t^x$ will be used in deriving the integration by parts formula and Lie brackets are a natural occurrence in this analysis. We may apply Corollary (11) iteratively to expand an expression for $(J_t^x)^{-1} V_i(X_t^x)$ for $i = 1, \ldots, d$, as far as the differentiability constraints on the vector fields permit. The divergence operator—which is the adjoint of the Malliavin derivative—plays a vital role in the construction of our integration by parts formula. This operator is also called the Skorohod integral. It coincides with a generalisation of the Itô integral to anticipating integrands. A detailed discussion of the divergence operator can be found in Nualart [51].

Definition 12 (Divergence operator). Denote by δ the adjoint of the operator D. That is, δ is an unbounded operator on $L^2(\Omega \times [0,\infty); \mathbb{R}^d)$ with values in $L^2(\Omega)$ such that:

1. $\mathrm{Dom}\,\delta = \{u \in L^2(\Omega \times [0,\infty); \mathbb{R}^d); |\mathbb{E}(\langle DF, u\rangle_H)| \le c\|F\|_{L^2(\Omega)}, \ \ \forall F \in \mathbb{D}^{1,2}\}$.
2. For every $u \in \mathrm{Dom}\,\delta$, then $\delta(u) \in L^2(\Omega)$ satisfies:

$$\mathbb{E}(F\delta(u)) = \mathbb{E}(\langle DF, u\rangle_H).$$

The following important results are shown in Sect. 1.5 of Nualart [51]:

1. D is continuous from $\mathbb{D}^{k,p}(E)$ into $\mathbb{D}^{k-1,p}(H \otimes E)$
2. $\langle DF, DF\rangle_H \in \mathbb{D}^\infty$ if $F, G \in \mathbb{D}^\infty$
3. δ is continuous from $\mathbb{D}^\infty(H)$ into $\mathbb{D}^\infty$.

Remark 13. If $u = (u^1, \ldots, u^d) \in \operatorname{Dom}\delta$ has the property that $t \to u(\cdot, t)$ is $\mathcal{F}_t$-adapted, then the adjoint $\delta(u)$, is nothing more than the Itô integral of u with respect to the d-dimensional Brownian motion $B_t = (B_t^1, \ldots, B_t^d)$. i.e.

$$\delta(u) = \sum_{i=1}^{d} \int_0^\infty u^i(\cdot, s) dB_s^i.$$

2.3 The UFG Condition

Define $\mathcal{A}$ to be the set of all n-tuples of natural numbers of any size n with the following form: $\mathcal{A} := \{1, \ldots, d\} \cup \bigcup_{k \in \mathbb{N}_0} \{0, 1, \ldots, d\}^k$. We endow $\mathcal{A}$ with the product:

$$\alpha * \beta := (\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_l), \text{ where } \alpha = (\alpha_1, \ldots, \alpha_k),\ \ \beta = (\beta_1, \ldots, \beta_l) \in \mathcal{A}.$$

Define $\mathcal{A}_{\emptyset,0} := \mathcal{A} \cup \{\emptyset, 0\}$. We consider the following n-tuples lengths:

$$|\,\alpha\,| := \begin{cases} k, & \text{if } \alpha = (\alpha_1, \ldots, \alpha_k), \\ 0, & \text{if } \alpha = \emptyset. \end{cases}$$

$$\|\alpha\| := |\,\alpha\,| + \operatorname{card}\{i : \alpha_i = 0, i = 1, \ldots, d\}.$$

We also introduce the sets

$$\mathcal{A}(m) = \{\alpha \in \mathcal{A} : \|\alpha\| \le m\} \quad \mathcal{A}_{\emptyset,0}(m) := \{\alpha \in \mathcal{A}_{\emptyset,0} : \|\alpha\| \le m\}.$$

We now define the vector field concatenation $V_{[\alpha]}$, $\alpha \in \mathcal{A}_{\emptyset,0}$ inductively, as follows:

$$\begin{aligned} V_{[\emptyset]} &:= 0, \\ V_{[i]} &:= V_i, & i &= 0, 1, \ldots, d, \\ V_{[\alpha * i]} &:= [V_{[\alpha]}, V_i], & i &= 0, 1, \ldots, d. \end{aligned}$$

In a similar vein to the above, we also define the Stratonovich integral concatenation,

$$\hat{B}_t^{\circ\alpha},\ t \in [0, \infty),\ \alpha \in \mathcal{A}_{\emptyset,0}$$

inductively:

$$\hat{B}_t^{\circ\emptyset} \equiv 1, \quad \hat{B}_t^{\circ i} := B_t^i, \quad \hat{B}_t^{\circ\alpha * i} := \int_0^t \hat{B}_s^{\circ\alpha} \circ dB_s^i, \quad i = 0, 1, \ldots, d.$$

We now introduce the main assumption for the gradient bounds analysis: the UFG condition. The purpose of the UFG condition, in its purest form, is to truncate the expansion obtained when considering the expression $(J_t^x)^{-1}V_i(X_t^x)$, for $i = 1, \ldots, d$. Recalling the work of the previous section, this appears when considering the product $(J_t^x)^{-1}DX_t^x$ between the Malliavin derivative and the inverse of the Jacobian of the stochastic flow. The UFG condition is a *"finite generation"* assumption, which helps to provide integration by parts formula.

Definition 14 (UFG Condition). Let $\{V_i : i = 0, \ldots, d\}$, be a system of vector fields such that $V_1, \ldots, V_d \in \mathcal{C}_b^{k+1}(\mathbb{R}^N;\mathbb{R}^N)$ and $V_0 \in \mathcal{C}_b^k(\mathbb{R}^N;\mathbb{R}^N)$. We say that $\{V_i : i = 0, \ldots, d\}$ satisfy the UFG condition if, there exists $m \in \mathbb{N}$, $m \leq k-1$, such for any $\alpha \in \mathcal{A}$, $\alpha = \alpha' * i$, $\alpha' \in \mathcal{A}(m)$ and $i = 0, \ldots, d$, there exist uniformly bounded functions $\varphi_{\alpha,\beta} \in \mathcal{C}_b^{k+1-|\alpha|}(\mathbb{R}^N, \mathbb{R})$, with $\beta \in \mathcal{A}(m)$ such that

$$V_{[\alpha]}(x) = \sum_{\beta \in \mathcal{A}(m)} \varphi_{\alpha,\beta}(x) V_{[\beta]}(x).$$

Heuristically, the UFG conditions states that higher order Lie brackets can be expressed as a linear combination of lower order Lie brackets, for some fixed order m. The uniform Hörmander condition implies the UFG condition, but not vice versa as we can see from the following example, taken from Kusuoka [30]:

Example 15. Assume $d = 1$ and $N = 2$. Let $V_0, V_1 \in \mathcal{C}_b^\infty(\mathbb{R}^2;\mathbb{R}^2)$ be given by

$$V_0(x_1, x_2) = \sin x_1 \frac{\partial}{\partial x_1} \qquad V_1(x_1, x_2) = \sin x_1 \frac{\partial}{\partial x_2}$$

Then $\{V_0, V_1\}$ do not satisfy the Hörmander condition. However the UFG condition is satisfied with $m = 4$.

Remark 16.

1. The UFG condition is defined in such a way (i.e. with $m \leq k-1$) that the elements $V_{[\alpha]}$ are well-defined and such that we may apply Corollary 11 to $V_{[\alpha]}$ for all $\alpha \in \mathcal{A}(m)$.
2. The regularity of the coefficients $\varphi_{\alpha,\beta}$ is chosen in accordance with what one would expect, given the regularity of $V_{[\alpha]}$.
3. We draw attention to the fact that we have assumed the coefficients are uniformly bounded. Although this assumption does not materially restrict the strength of the results, it does make them more presentable and reduces the complexity in the proof. Essentially the boundedness of the coefficients means there is a natural and elegant description for how the gradient bounds may increase as a function of $|x|$. We endeavor to draw attention to the effects of this assumption where appropriate. In many examples of interest, this assumption imposes no unnecessary restrictions.

2.4 The Central Representation Formula

From Corollary (11) and the UFG condition, we have, for each $\alpha \in \mathcal{A}(m)$,

$$\begin{aligned} d\left[(J_t^x)^{-1} V_{[\alpha]}(X_t^x)\right] &= -\sum_{i=0}^{d} (J_t^x)^{-1} [V_{[\alpha]}, V_i](X_t^x) \circ dB_t^i \\ &= -\sum_{i=0}^{d} (J_t^x)^{-1} V_{[\alpha * i]}(X_t^x) \circ dB_t^i \\ &= \sum_{i=0}^{d} \sum_{\beta \in \mathcal{A}(m)} c_{\alpha,\beta}^i (X_t^x)(J_t^x)^{-1} V_{[\beta]}(X_t^x) \circ dB_t^i \ , \end{aligned} \tag{38}$$

where the coefficients $c_{\alpha,\beta}^i$, $\alpha, \beta \in \mathcal{A}(m)$, $i = 0, \ldots, d$ are given by

$$c_{\alpha,\beta}^i(x) = \begin{cases} -1 & \text{if } \alpha * i \in \mathcal{A}(m) \text{ and } \beta = \alpha * i \\ 0 & \text{if } \alpha * i \in \mathcal{A}(m) \text{ and } \beta \neq \alpha * i \\ -\varphi_{\alpha * i, \beta} & \text{if } \alpha * i \notin \mathcal{A}(m) \end{cases} . \tag{39}$$

We note, in particular, that $c_{\alpha,\beta}^i \in C_b^{k+1-|\alpha|}(\mathbb{R}^N, \mathbb{R})$ are uniformly bounded. We obtained a representation of the vector fields $V_{[\alpha]}$, $\alpha \in \mathcal{A}(m)$ (estimated at $(X_\cdot^x)$) in terms of the Lie brackets $V_{[\alpha * i]} := [V_{[\alpha]}, V_i]$, $\alpha \in \mathcal{A}(m)$, $i = 0, \ldots, d$, which where then reverted back to the original set of vector fields $V_{[\alpha]}$, $\alpha \in \mathcal{A}(m)$ via the UFG condition. Without the UFG condition, the resulting representation would potentially be infinite. Indeed, the Hörmander approach relies on showing that, after a certain number of iterations (taking Lie brackets of the resulting vector fields), the remainder term arising from the expansion becomes very small. The UFG condition is more general than Hörmander's (see [24]) famous criterion for hypoellipticity of linear differential operators and it allows us to take a different approach. We can view (38) as a linear system of SDEs whose coefficients are of suitably chosen differentiability whose solutions are the processes $t \to (J_t^x)^{-1} V_{[\alpha]}(X_t^x)$, $\alpha \in \mathcal{A}(m)$. This enables us to represent these processes in terms of their initial values $V_{[\alpha]}(x)$, $\alpha \in \mathcal{A}(m)$ and the corresponding representation facilitates the integration by parts formula. Moreover, we shall see how the same representation leads to the classical non-degeneracy result: The gradient bounds obtained under the UFG condition shall implicitly recover Hörmander's result, see Theorem 70.

By considering the above as a closed linear system of equations, we are able to equivalently view it as the matrix SDE:

$$Y(t, x) = Y(0, x) + \sum_{i=0}^{d} \int_0^t C^i(X_s^x) Y(s, x) \circ dB_s^i, \tag{40}$$

where $Y(0,x) = V(x) := \left(V_{[\alpha]}(x)\right)_{\alpha\in\mathcal{A}(m)} \in \mathbb{R}^{N_m} \times \mathbb{R}^N$, ($N_m = card(\mathcal{A}_m)$) and $C^i : \mathbb{R}^N \to \mathbb{R}^{N_m} \otimes \mathbb{R}^{N_m}$ are given by

$$C^i(x) := \left(c^i_{\alpha,\beta}(x)\right)_{\alpha,\beta\in\mathcal{A}(m)}.$$

We are able to take advantage of the linear nature of this system of SDEs by considering in more generality the matrix which produces such vectors. Namely,

Lemma 17. *Assume that $A(t,x)$, $(t,x) \in [0,\infty) \times \mathbb{R}^N$ is the $N_m \times N_m$-matrix which is the unique solution to the matrix stochastic differential equation*

$$dA(t,x) = \sum_{i=0}^{d} C^i(X^x_t)A(t,x) \circ dB^i_t, \tag{41}$$

where $A(0,x) = I$. Then $Y(t,x) = A(t,x)Y(0,x)$.

Proof. We need only show that $A(t,x)Y(0,x)$ solves equation (40), then by the uniqueness of SDE solutions (see, for example Karatzas and Shreve [26]), the result follows. But,

$$\begin{aligned} d(A(t,x)Y(0,x)) &= A(t,x) \circ dY(0,x) + \circ dA(t,x)Y(0,x) = \circ dA(t,x)Y(0,x) \\ &= \sum_{i=0}^{d} C^i(X^x_t)A(t,x)Y(0,x) \circ dB^i_t \end{aligned}$$

and, clearly, $A(0,x)Y(0,x) = Y(0,x)$. □

The above results show that all the relevant information about the solution (40) is captured by the solution (41). We can apply classical results about solutions of SDEs to obtain the following proposition.

Proposition 18. *The matrix stochastic differential equation (41) has a unique solution, $A = (a_{\alpha,\beta})_{\alpha,\beta\in\mathcal{A}(m)}$ with components $a_{\alpha,\beta} : [0,\infty) \times \mathbb{R}^N \to \mathbb{R}$, $\alpha,\beta \in \mathcal{A}(m)$ that satisfy the mutually dependent SDEs:*

$$a_{\alpha,\beta}(t,x) = \delta_{\alpha,\beta} + \sum_{i=0}^{d} \sum_{\gamma\in\mathcal{A}(m)} \int_0^t c^i_{\alpha,\gamma}(X^x_u)a_{\gamma,\beta}(u,x) \circ dB^i_u.$$

Moreover $a_{\alpha,\beta}(t,.) : \mathbb{R}^N \to \mathbb{R}$ are a.s. $k-m$ times differentiable in x for fixed $t \in [0,\infty)$ and $a_{\alpha,\beta}(.,.)$ is jointly continuous in $[0,\infty) \times \mathbb{R}^N$ with probability one, for each $\alpha,\beta \in \mathcal{A}(m)$ and

$$\sup_{\substack{x\in\mathbb{R}^N\\ t\in[0,T]}} \mathbb{E}\left[\left|\frac{\partial^{|\gamma|}}{\partial x^{\gamma}} a_{\alpha,\beta}(t,x)\right|^p\right] < \infty, \quad \forall\, p \in [1,\infty),\ T > 0, \tag{42}$$

for any multi-index γ *with* $|\gamma| \le k-m$*. Finally, for any* $l \le k-m$

$$\sup_{t\in[0,T]} \mathbb{E}\left[\left\| D^l a_{\alpha,\beta}(t,x)\right\|^p_{H^{\otimes l}}\right] < C_{l,p}(1+|x|)^p \quad \forall\, p \in [1,\infty),\ T > 0, \tag{43}$$

Furthermore, the matrix $A = (a_{\alpha,\beta})_{\alpha,\beta\in\mathcal{A}(m)}$ *is invertible, and its inverse* $B = (b_{\alpha,\beta})_{\alpha,\beta\in\mathcal{A}(m)}$ *satisfies the matrix SDE:*

$$B(t,x) = I - \sum_{i=0}^{d} \int_0^t B(u,x) C^i(X_u^x) \circ dB_u^i.$$

Moreover, the components $b_{\alpha,\beta}$*,* $\alpha,\beta \in \mathcal{A}(m)$*, are a.s.* $k-m$ *times differentiable in* x *for fixed* $t \in [0,\infty)$*, jointly continuous in* (t,x) *and*

$$\sup_{\substack{t\in[0,T]\\ x\in\mathbb{R}^N}} \mathbb{E}\left[\left|\frac{\partial^{|\gamma|}}{\partial x^{\gamma}} b_{\alpha,\beta}(t,x)\right|^p\right] < C_{T,p}, \tag{44}$$

for each $p \in [1,\infty)$*,* $T > 0$*,* $|\gamma| \le k-m$ *and some constant* $C_{T,p}$*. Finally, for any* $l \le k-m$

$$\sup_{t\in[0,T]} \mathbb{E}\left[\left\| D^l b_{\alpha,\beta}(t,x)\right\|^p_{H^{\otimes l}}\right] < C_{l,p}(1+|x|)^p \quad \forall\, p \in [1,\infty),\ T > 0, \tag{45}$$

Proof. This is very similar to Theorem 8. The only difference here is that the bounds on the norms of the iterated Malliavin derivatives are now bounded only linearly in $|x|$. This is obvious once one considers Theorem 8 and, in particular, inequality (34). It is clear from this equation that the norm of the Malliavin derivatives inherits the linear growth of the vector fields. All higher order Malliavin derivatives inherit this linearity from the first order Malliavin derivative, but given the boundedness of the derivatives of the vector fields, have no worse than linear growth. □

Remark 19. (**a**) The above proposition highlights an idiosyncratic difference between the Malliavin derivative and the normal derivative for the solutions of such SDEs. It stems from the fact that the Malliavin derivative of X_t^x has an unbounded norm over $x \in \mathbb{R}^N$, as it has Lipschitz continuous coefficients. However, the same result for the norm of the classical derivative of X_t^x **is** bounded over $x \in \mathbb{R}^N$. Note this difference would not appear if we assumed the vector fields were uniformly bounded.

(**b**) Although not used in the sequel, identities (42)–(45) hold true with the supremum taken inside the expectation.

We now seek to study the solution to (41), whose elements will be absolutely fundamental to our analysis. We note initially, that although this matrix is potentially very large, with potentially significant mutual dependence, many of the terms which make up this mutual dependence are zero. This allows us to get a good handle on the matrix. Note that for fixed $\alpha, \beta \in \mathcal{A}(m)$ we have

$$a_{\alpha,\beta}(t,x) = \delta_{\alpha\beta} + \sum_{i=0}^{d} \sum_{\gamma \in \mathcal{A}(m)} \int_0^t c^i_{\alpha,\gamma}(X^x_s) a_{\gamma,\beta}(s,x) \circ dB^i_s. \tag{46}$$

The coefficients $c^i_{\alpha,\gamma}$ identified in (39) lead to the following:
For $\|\alpha\| \leq m-2$ there holds: $\|\alpha * i\| \leq m$ for all $i = 0, \dots, d$, so $c^i_{\alpha,\gamma} \neq 0$ only when $\gamma = \alpha * i$. In which case $c^i_{\alpha,\gamma} = -1$. i.e.

$$a_{\alpha,\beta}(t,x) = \delta_{\alpha\beta} - \sum_{i=0}^{d} \int_0^t a_{\alpha * i,\beta}(s,x) \circ dB^i_s.$$

For $\|\alpha\| = m-1$ there holds: $\|\alpha * i\| = m$ for $i = 1, \dots, d$, with $\|\alpha * 0\| = m+1$. Hence $\alpha * i \in \mathcal{A}(m)$ for $i = 1, \dots, d$, and $\alpha * 0 \notin \mathcal{A}(m)$. i.e.

$$a_{\alpha,\beta}(t,x) = \delta_{\alpha\beta} - \sum_{i=1}^{d} \int_0^t a_{\alpha * i,\beta}(s,x) \circ dB^i_s - \sum_{\gamma \in \mathcal{A}(m)} \int_0^t \varphi_{\alpha * 0,\gamma}(X^x_s) a_{\gamma,\beta}(s,x) ds.$$

For $\|\alpha\| = m$ there holds: $\|\alpha * i\| > m$ for $i = 0, \dots, d$. Hence $\alpha * i \notin \mathcal{A}(m)$ for $i = 0, \dots, d$. i.e.

$$a_{\alpha,\beta}(t,x) = \delta_{\alpha\beta} - \sum_{i=0}^{d} \sum_{\gamma \in \mathcal{A}(m)} \int_0^t \varphi_{\alpha * i,\gamma}(X^x_s) a_{\gamma,\beta}(s,x) \circ dB^i_s.$$

An explicit form for $a_{\alpha,\beta}$ is sought and is easy to identify from (46). In fact, each element of the matrix A can be split up into a sum of two terms: the term which arises from $\delta_{\alpha\beta}$—a iterated Stratonovich integral of a constant—and a remainder term. That is, for any $\alpha, \beta \in \mathcal{A}(m)$,

$$a_{\alpha,\beta}(t,x) = a^0_{\alpha,\beta}(t,x) + r_{\alpha,\beta}(t,x), \tag{47}$$

where

$$a^0_{\alpha,\beta}(t,x) = \begin{cases} (-1)^{|\gamma|} \hat{B}^{\circ\gamma}_t & \text{if } \beta = \alpha * \gamma \text{ for some } \gamma \in \mathcal{A}(m) \\ 0 & \text{otherwise} \end{cases}$$

and

$$r_{\alpha,\beta}(t,x) = \sum_{\substack{\gamma\in\mathcal{A}, j=0,\dots,d \\ \text{s.t. } \|\alpha*\gamma\|\le m \\ \|\gamma*j\|\ge m+1-\|\alpha\|}} \sum_{\delta\in\mathcal{A}(m)} \int_0^t \int_0^{s_k} \dots \int_0^{s_1} (-1)^{|\gamma|} c^j_{\alpha*\gamma,\delta}(X^x_s)$$
$$\times a_{\delta,\beta}(s,x) \circ dB^j_s \circ dB^{\gamma_1}_{s_1} \dots dB^{\gamma_k}_{s_k}$$

The following proposition is a good indicator of how we are able to identify the explicit short-time asymptotic rates in terms of time, t.

Proposition 20. *For any $T > 0$, $p \in [1,\infty)$, $\alpha,\beta \in \mathcal{A}(m)$ and $\gamma \in \mathcal{A}_0$, the following hold*

$$\sup_{t\in(0,T]} \mathbb{E}\left[\left(t^{-\|\gamma\|/2}\left|\hat{B}^{\circ\gamma}_t\right|\right)^p\right] < \infty, \tag{48}$$

$$\sup_{\substack{x\in\mathbb{R}^N \\ t\in(0,T]}} \mathbb{E}\left[\left(t^{-(m+1-\|\alpha\|)/2}\left|r_{\alpha,\beta}(t,x)\right|\right)^p\right] < \infty. \tag{49}$$

Proof. The proof of these result is left for the appendix. □

We are now ready to derive the integration by parts formula. Let $f \in \mathcal{C}^\infty_b(\mathbb{R}^N,\mathbb{R})$, then, using (36), we get

$$\begin{aligned} Df(X^x_t) &= \nabla f(X^x_t) DX^x_t \\ &= \nabla(f\circ X_t)(x)(J^x_t)^{-1} DX^x_t \\ &= \nabla(f\circ X_t)(x)\left(\int_0^{t\wedge\cdot} (J^x_s)^{-1} V_i(X^x_s) ds\right)_{i=1,\dots,d}. \end{aligned}$$

The idea is to develop the preceding equality to isolate terms involving $\nabla(f \circ X_t)(x)$. Once isolated, the operators of the Malliavin calculus will be used to derive an integration by parts formula. Now we note that, from Lemma 17:

$$(J^x_s)^{-1} V_i(X^x_s) = (A(s,x)V(x))_i = \sum_{\beta\in\mathcal{A}(m)} a_{i,\beta}(t,x) V_{[\beta]}(x).$$

Hence,

$$\begin{aligned} Df(X^x_t) &= \nabla(f\circ X_t)(x)\left(\int_0^{t\wedge\cdot} \sum_{\beta\in\mathcal{A}(m)} a_{i,\beta}(s,x) V_{[\beta]}(x) ds\right)_{i=1,\dots,d} \\ &= \nabla(f\circ X_t)(x) \sum_{\beta\in\mathcal{A}(m)} V_{[\beta]}(x)\left(\int_0^{t\wedge\cdot} a_{i,\beta}(s,x) ds\right)_{i=1,\dots,d} \\ &= \sum_{\beta\in\mathcal{A}(m)} V_{[\beta]}(f\circ X_t)(x) k_\beta(t,x), \end{aligned}$$

where

$$k_\beta(t,x) := \left(\int_0^{t\wedge\cdot} a_{i,\beta}(s,x)ds\right)_{i=1,\dots,d}.$$

We re-write the previous equation into a linear system of equations by taking the H inner product with $k_\alpha(t,x)$ for all $\alpha \in \mathcal{A}(m)$. i.e.

$$\left\langle Df(X_t^x), k_{(1)}(t,x)\right\rangle_H = \sum_{\beta\in\mathcal{A}(m)} V_{[\beta]}(f\circ X_t)(x)\left\langle k_\beta(t,x), k_{(1)}(t,x)\right\rangle_H$$

$$\vdots$$

$$\langle Df(X_t^x), k_\alpha(t,x)\rangle_H = \sum_{\beta\in\mathcal{A}(m)} V_{[\beta]}(f\circ X_t)(x)\left\langle k_\beta(t,x), k_\alpha(t,x)\right\rangle_H$$

$$\vdots$$

$$\left\langle Df(X_t^x), k_{\alpha(N_m)}(t,x)\right\rangle_H = \sum_{\beta\in\mathcal{A}(m)} V_{[\beta]}(f\circ X_t)(x)\left\langle k_\beta(t,x), k_{\alpha(N_m)}(t,x)\right\rangle_H$$

Define, for $\alpha \in \mathcal{A}(m)$:

$$\begin{aligned} D^{(\alpha)}f(X_t^x) &:= \langle Df(X_t^x), k_\alpha(t,x)\rangle_H \\ M_{\alpha,\beta}(t,x) &:= t^{-(\|\alpha\|+\|\beta\|)/2}\left\langle k_\alpha(t,x), k_\beta(t,x)\right\rangle_H \\ &= t^{-(\|\alpha\|+\|\beta\|)/2}\sum_{i=1}^d \int_0^t a_{i,\alpha}(s,x)a_{i,\beta}(s,x)ds. \end{aligned}$$

This leaves us with

$$D^{(\alpha)}f(X_t^x) = \sum_{\beta\in\mathcal{A}(m)} t^{(\|\alpha\|+\|\beta\|)/2} M_{\alpha,\beta}(t,x)V_{[\beta]}(f\circ X_t)(x).$$

The above can be seen as a linear system of equations driven by a random matrix

$$M(t,x) = (M_{\alpha,\beta}(t,x))_{\alpha,\beta}.$$

The invertibility of this matrix is a major step forward towards an integration by parts. For then there would hold, $\mathbb{P}$-a.s:

$$V_{[\alpha]}(f \circ X_t)(x) = t^{-\|\alpha\|/2} \sum_{\beta \in \mathcal{A}(m)} t^{-\|\beta\|/2} M^{-1}_{\alpha,\beta}(t,x) D^{(\beta)} f(X_t^x).$$

Proposition 21. *$M(t,x)$ is $\mathbb{P}$-a.s. invertible. Moreover, for $p \in [1,\infty)$ and $\alpha, \beta \in \mathcal{A}(m)$, there holds*

$$\sup_{t \in (0,1], x \in \mathbb{R}^N} \mathbb{E}\left[(M^{-1}_{\alpha,\beta}(t,x))^p \right] < \infty \tag{50}$$

Proof. The proof of invertibility is lengthy and is left to the appendix. □

2.5 Kusuoka–Stroock Functions

We introduce now a class of functions which we shall call Kusuoka–Stroock functions. Such functions play the central role in the deduction of the integration by parts formulae (IBPF) and the control of the derivatives of the semigroup P_t. In particular we will show that if $(t,x) \to \Phi(t,x)$ is a Kusuoka–Stroock function, then there exists another Kusuoka–Stroock function $(t,x) \to \Phi_\alpha(t,x)$, $\alpha \in \mathcal{A}(m)$ such that:

$$\mathbb{E}[\Phi(t,x) V_{[\alpha]}(f \circ X_t)(x)] = t^{-\|\alpha\|/2} \mathbb{E}[\Phi_\alpha(t,x) f(X_t^x)].$$

This class of functions is closed under the operations which are taken during the formation of the IBPF. As a result this space supports iterative applications of the above formula.

Definition 22 (Local Kusuoka–Stroock functions). Let E be a separable Hilbert space and let $r \in \mathbb{R}$, $n \in \mathbb{N}$. We denote by $\mathcal{K}_r^{\mathrm{loc}}(E,n)$ the set of functions: $f : (0,T] \times \mathbb{R}^N \to \mathbb{D}^{n,\infty}(E)$ satisfying the following:

1. $f(t,.)$ is n-times continuously differentiable and $\frac{\partial^\alpha f}{\partial x^\alpha}(.,.)$ is continuous in $(t,x) \in (0,T] \times \mathbb{R}^N$ a.s. for any $\alpha \in \mathcal{A}$ satisfying $|\alpha| \le n$
2. For any K compact subset of $\mathbb{R}^N$ and $k \in \mathbb{N}$, $p \in [1,\infty)$ with $k \le n - |\alpha|$, we have

$$\sup_{t \in (0,T], x \in K} t^{-r/2} \left\| \frac{\partial^\alpha f}{\partial x^\alpha} \right\|_{\mathbb{D}^{k,p}(E)} < \infty. \tag{51}$$

If (51) holds globally over $\mathbb{R}^N$, we write $f \in \mathcal{K}_r(E,n)$ and denote $\mathcal{K}_r^{loc}(n) := \mathcal{K}_r^{loc}(\mathbb{R},n)$ and, respectively, $\mathcal{K}_r(n) := \mathcal{K}_r(\mathbb{R},n)$.

The functions belonging to the set $\mathcal{K}_r^{loc}(E,n)$ satisfy the following properties which form the basis of the integration by parts formula.

Lemma 23 (Properties of local Kusuoka–Stroock functions). *The following hold*

1. *Suppose* $f \in \mathcal{K}_r^{loc}(E,n)$*, where* $r \geq 0$*. Then, for* $i = 1, \ldots, d$*,*

$$\int_0^{\cdot} f(s,x) dB_s^i \in \mathcal{K}_{r+1}^{loc}(E,n) \quad \textit{and} \quad \int_0^{\cdot} f(s,x) ds \in \mathcal{K}_{r+2}^{loc}(E,n).$$

2. $a_{\alpha,\beta}, b_{\alpha,\beta} \in \mathcal{K}^{loc}_{(\|\beta\|-\|\alpha\|)\vee 0}(k-m)$ *for any* $\alpha, \beta \in \mathcal{A}(m)$*.*
3. $k_\alpha \in \mathcal{K}^{loc}_{\|\alpha\|}(H, k-m)$ *for any* $\alpha \in \mathcal{A}(m)$*.*
4. $D^{(\alpha)}u := \langle Du(t,x), k_\alpha \rangle_H \in \mathcal{K}^{loc}_{r+\|\alpha\|}(n \wedge [k-m])$ *where* $u \in \mathcal{K}_r^{loc}(n)$ *and* $\alpha \in \mathcal{A}(m)$*.*
5. *If* $M^{-1}(t,x)$ *is the inverse matrix of* $M(t,x)$*, then* $M^{-1}_{\alpha,\beta} \in \mathcal{K}_0^{loc}(k-m)$*,* $\alpha, \beta \in \mathcal{A}(m)$*.*
6. *If* $f_i \in \mathcal{K}^{loc}_{r_i}(n_i)$ *for* $i = 1, \ldots, N$*, then*

$$\prod_{i=1}^N f_i \in \mathcal{K}^{loc}_{r_1+\ldots+r_N}(\min_i n_i) \quad \textit{and} \quad \sum_{i=1}^N f_i \in \mathcal{K}^{loc}_{\min_i r_i}(\min_i n_i).$$

Moreover, if we assume the vector fields $V_0, \ldots, V_d$ *are also uniformly bounded, then (2)–(5) hold with* $\mathcal{K}^{loc}$ *replaced by* $\mathcal{K}$*.*

Proof. This is proved in the appendix. □

2.6 Integration by Parts Formulae

In this section we synthesise the developed results to obtain various integration by parts formulae, in a way which should now be familiar. We note that some of the stated results are for iterated derivatives of the semigroup P_t (cf. Corollary 28) along vector fields of the Lie algebra. Seeing as the purpose of this section is to look at derivatives of the semigroup, we shall always assume that $V_{[\alpha_1]}, \ldots, V_{[\alpha_{N+M}]}$ have sufficient smoothness for this operation to be well-defined.

Theorem 24 (Integration by Parts formula I). *Under the UFG condition, for any* $\Phi \in \mathcal{K}_r^{loc}(n)$ *and for any* $\alpha \in \mathcal{A}(m)$*, there exists* $\Phi_\alpha \in \mathcal{K}_r^{loc}((n-1) \wedge (k-m-1))$ *such that:*

$$\mathbb{E}\left[\Phi(t,x) V_{[\alpha]}(f \circ X_t)(x)\right] = t^{-\|\alpha\|/2} \mathbb{E}\left[\Phi_\alpha(t,x) f(X_t^x)\right], \quad t > 0, x \in \mathbb{R}^N \tag{52}$$

for any $f \in \mathcal{C}_b^\infty(\mathbb{R}^N; \mathbb{R})$*. In addition, for any* $q > p$

$$\sup_{t \in (0,T]} \mathbb{E}\left[|\, \Phi_\alpha(t,x)\,|^p\right] \leq C_{p,q}(1 + |\, x\,|)^p \sup_{t \in (0,T]} \mathbb{E}[\|\Phi(t,x)\|^p_{\mathbb{D}^{2,q}}]. \tag{53}$$

Moreover, if $\Phi \in \mathcal{K}_r(n)$ *and the vector fields* V_i, $i = 0, 1, \ldots, d$ *are uniformly bounded, then* $\Phi_\alpha \in \mathcal{K}_r((n-1) \wedge (k-m-1))$. *In particular,*

$$\sup_{t\in(0,T]} \sup_{x\in\mathbb{R}^N} \mathbb{E}\left[|\ \Phi_\alpha(t,x)\ |^p\right] < \infty. \tag{54}$$

Proof. We showed in the previous section that

$$V_{[\alpha]}(f \circ X_t)(x) = t^{-\|\alpha\|/2} \sum_{\beta\in\mathcal{A}(m)} t^{-\|\beta\|/2} M^{-1}_{\alpha,\beta}(t,x) D^{(\beta)}(f(X_t^x))$$

holds $\mathbb{P}$-a.s. By the product rule for the Malliavin derivative:

$$\begin{aligned} D^{(\beta)}(\Phi(t,x)\ M^{-1}_{\alpha,\beta}(t,x)\ f(X_t^x)) = D^{(\beta)}&\Phi(t,x)\ M^{-1}_{\alpha,\beta}(t,x)\ f(X_t^x) \\ &+ \Phi(t,x)\ D^{(\beta)}M^{-1}_{\alpha,\beta}(t,x)\ f(X_t^x) \\ &+ \Phi(t,x)\ M^{-1}_{\alpha,\beta}(t,x)\ D^{(\beta)}f(X_t^x). \end{aligned}$$

Then

$$\begin{aligned} &\mathbb{E}\left[\Phi(t,x)V_{[\alpha]}(f \circ X_t)(x)\right] \\ &= t^{-\frac{\|\alpha\|}{2}} \sum_{\beta\in\mathcal{A}(m)} t^{-\frac{\|\beta\|}{2}} \mathbb{E}\left[\Phi(t,x)M^{-1}_{\alpha,\beta}(t,x)D^{(\beta)}(f(X_t^x))\right] \\ &= t^{-\frac{\|\beta\|}{2}} \mathbb{E}\left[\Phi_\alpha(t,x)f(X_t^x)\right], \end{aligned}$$

where

$$\begin{aligned} \Phi_\alpha(t,x) = \sum_{\beta\in\mathcal{A}(m)} t^{-\|\beta\|/2} \Big\{ &\Phi(t,x)\ M^{-1}_{\alpha,\beta}(t,x)\delta(k_\beta(t,x)) \\ &-\Phi(t,x)D^{(\beta)}M^{-1}_{\alpha,\beta}(t,x) - D^{(\beta)}\Phi(t,x)\ M^{-1}_{\alpha,\beta}(t,x)\Big\}. \end{aligned}$$

The claim $\Phi_\alpha \in \mathcal{K}_r^{\text{loc}}((n-1) \wedge (k-m-1))$ follows from a diligent application of Lemma 23, namely, parts 3–6. Note that the only term unbounded in x in the expression for Φ_α is $D^{(\beta)}M^{-1}_{\alpha,\beta}(t,x)$ which has linear growth in x. Finally, the bound (53) can be proved by observing that, due to (43)

$$\sup_{t\in(0,T]} \mathbb{E}\left|\ D^{(\gamma)}M^{-1}_{\beta,\gamma}(t,x)\ \right|^p \le C(1+|\ x\ |)^p. \tag{55}$$

Hence, the bound

$$\sup_{t\in(0,T]} \mathbb{E} \mid \Phi_\alpha(t,x) \mid^p \le C_p(1+\mid x \mid)^p \sup_{t\in(0,T]} \|\Phi(t,x)\|_{2,q}^p$$

follows by applying the following to the expression for $\Phi_\alpha(t,x)$: (55), Hölder's inequality, and the uniform boundedness of the L^r norm of M^{-1} and k_γ over $(t,x) \in (0,T] \times \mathbb{R}^N$ for each $r \ge 1$. □

Remark 25. Following from Remark 13, the adjoint $\delta(k_\gamma(t,x))$ is the Itô integral of $k_\gamma(t,x)$ with respect to the d-dimensional Brownian motion $B_t = (B_t^1, \ldots, B_t^d)$ as the process $s \to k_\gamma(t,x)(s)$ is $\mathcal{F}_s$-adapted for almost all $(t,x) \in (0,T] \times \mathbb{R}^N$. That is, we have that

$$\delta(k_\gamma(t,x)) = \sum_{i=1}^d \int_0^1 k_\gamma(t,x)^i(s) dB_s^i.$$

It follows that for processes with values in $K^r(E)$ which are a.e. adapted as stochastic processes in H, that $\delta(f) := \delta(f(.,.)) \in \mathcal{K}_{r+1}(E)$.

Corollary 26 (Integration by Parts formula II). *Under the UFG condition, for any $\Phi \in \mathcal{K}_r^{loc}(n)$ and for any $\alpha \in \mathcal{A}(m)$, there exists $\Phi'_\alpha \in \mathcal{K}_r^{loc}((n-1)\wedge(k-m-1))$ such that:*

$$\mathbb{E}[\Phi(t,x)(V_{[\alpha]}f)(X_t^x)] = t^{-\|\alpha\|/2}\mathbb{E}[\Phi'_\alpha(t,x) f(X_t^x)], \quad t > 0, x \in \mathbb{R}^N \tag{56}$$

for any $f \in C_b^\infty(\mathbb{R}^N;\mathbb{R})$. In addition, for any $q > p$

$$\sup_{t\in(0,T]} \mathbb{E}\left[\left| \Phi'_\alpha(t,x) \right|^p\right] \le C_{p,q}(1+\mid x \mid)^p \sup_{t\in(0,T]} \mathbb{E}\left[\|\Phi(t,x)\|_{\mathbb{D}^{2,q}}^p\right]. \tag{57}$$

Moreover, if $\Phi \in \mathcal{K}_r(n)$ and the vector fields V_i, $i = 0,1,\ldots,d$ are uniformly bounded, then $\Phi'_\alpha \in \mathcal{K}_r((n-1)\wedge(k-m-1))$. In particular,

$$\sup_{t\in(0,T]} \sup_{x\in\mathbb{R}^N} \mathbb{E}\left[\left| \Phi'_\alpha(t,x) \right|^p\right] < \infty. \tag{58}$$

Proof. The first observation is the following relationship:

$$\begin{aligned}(V_{[\alpha]}f)(X_t^x) &= \nabla f(X_t^x) V_{[\alpha]}(X_t^x)\\ &= (J_t^x)^{-T}\nabla(f \circ X_t)(x) V_{[\alpha]}(X_t^x)\\ &= \nabla(f \circ X_t)(x)(J_t^x)^{-1} V_{[\alpha]}(X_t^x),\end{aligned}$$

where $(J_t^x)^{-T} := ((J_t^x)^{-1})^T$. At this point refer back to the closed linear system of equations, which induced the expression:

$$(J_t^x)^{-1} V_{[\alpha]}(X_t^x) = \sum_{\beta \in \mathcal{A}(m)} a_{\alpha,\beta}(t,x) V_{[\beta]}(x).$$

Again, the central position of the UFG condition is emphasised, as

$$\begin{aligned}\nabla(f \circ X_t)(x)(J_t^x)^{-1} V_{[\alpha]}(X_t^x) &= \sum_{\beta \in \mathcal{A}(m)} a_{\alpha,\beta}(t,x) \nabla(f \circ X_t)(x) V_{[\beta]}(x) \\ &= \sum_{\beta \in \mathcal{A}(m)} a_{\alpha,\beta}(t,x) V_{[\beta]}(f \circ X_t)(x).\end{aligned}$$

From Lemma 23, $a_{\alpha,\beta} \in \mathcal{K}^{\text{loc}}_{(\|\beta\|-\|\alpha\|)\vee 0}(k-m)$. Hence, it has been shown that:

$$\mathbb{E}\left[\Phi(t,x) V_{[\alpha]} f(X_t^x)\right] = \sum_{\beta \in \mathcal{A}(m)} \mathbb{E}\left[\Phi(t,x) a_{\alpha,\beta}(t,x) V_{[\beta]}(f \circ X_t)(x)\right].$$

The integration by parts formula (52) can then be applied N_m times, after noting that the product $\Phi a_{\alpha,\beta} \in \mathcal{K}^{\text{loc}}_{r+[(\|\beta\|-\|\alpha\|)\vee 0]}((n-1) \wedge (k-m-1))$. And so,

$$\begin{aligned}\mathbb{E}\left[\Phi(t,x) V_{[\alpha]} f(X_t^x)\right] &= \sum_{\beta \in \mathcal{A}(m)} t^{-\frac{\|\beta\|}{2}} \mathbb{E}\left[\Psi_\beta(t,x) f(X_t^x)\right] \\ &= \sum_{\beta \in \mathcal{A}(m)} t^{-\frac{\|\beta\|}{2}} t^{-\frac{\|\alpha\|-\|\beta\|}{2}} \mathbb{E}\left[t^{\frac{\|\alpha\|-\|\beta\|}{2}} \Psi_\beta(t,x) f(X_t^x)\right] \\ &= t^{-\frac{\|\alpha\|}{2}} \mathbb{E}\left[\Phi'_\alpha(t,x) f(X_t^x)\right],\end{aligned}$$

where

$$\Phi'_\alpha = \sum_{\beta \in \mathcal{A}(m)} t^{\frac{\|\alpha\|-\|\beta\|}{2}} \Psi_\beta \in \mathcal{K}^{\text{loc}}_r((n-1) \wedge (k-m-1)).$$

The bounds (57), (58) can be deduced from the previous theorem. □

Corollary 27 (Integration by Parts formula III). *Under the same conditions as Theorem 24, the following integration by parts formula holds:*

$$V_{[\alpha]} \mathbb{E}\left[\Phi(t,x) f(X_t^x)\right] = t^{-\|\alpha\|/2} \mathbb{E}\left[\Phi''_\alpha(t,x) f(X_t^x)\right], \quad t > 0, x \in \mathbb{R}^N, \qquad (59)$$

where $\Phi''_\alpha \in \mathcal{K}_r^{loc}((n-1)\wedge(k-m-1))$. *In addition, for any* $q > p$*:*

$$\sup_{t\in(0,T]} \mathbb{E}\left[\left|\,\Phi''_\alpha(t,x)\,\right|^p\right] \le C_{p,q}(1+|\,x\,|)^p \sup_{t\in(0,T]} \|\Phi(t,x)\|^p_{\mathbb{D}^{2,q}}. \tag{60}$$

Moreover, if $\Phi \in \mathcal{K}_r(n)$ *and the vector fields* V_i*,* $i = 0,1,\dots,d$ *are uniformly bounded, then* $\Phi''_\alpha \in \mathcal{K}_r((n-1)\wedge(k-m-1))$. *In particular,*

$$\sup_{t\in(0,T]}\sup_{x\in\mathbb{R}^N} \mathbb{E}[\left|\,\Phi''_\alpha(t,x)\,\right|^p] < \infty. \tag{61}$$

Proof. Observe that

$$\begin{aligned} V_{[\alpha]}\mathbb{E}[\Phi(t,x)f(X_t^x)] &= \mathbb{E}\left[V_{[\alpha]}(\Phi(t,x))\ f(X_t^x) + \Phi(t,x)V_{[\alpha]}(f\circ X_t)(x)\right] \\ &= \mathbb{E}[V_{[\alpha]}(\Phi(t,x))\ f(X_t^x) + t^{-\|\alpha\|/2}\Phi_\alpha(t,x)f(X_t^x)] \\ &= t^{-\|\alpha\|/2}\mathbb{E}[\Phi''_\alpha(t,x)f(X_t^x)], \end{aligned}$$

where

$$\Phi''_\alpha(t,x) = t^{\|\alpha\|/2}V_{[\alpha]}(\Phi(t,x)) + \Phi_\alpha(t,x) \in \mathcal{K}_r^{\mathrm{loc}}((n-1)\wedge(k-m-1)).$$

It is also clear from the previous results that Φ''_α satisfies (60). □

Corollary 28 (Integration by Parts formula IV). *Under the same conditions as Theorem 24, the following integration by parts formula holds for* $m_1+m_2 \le k-m$ *and* $\alpha_1,\dots,\alpha_{m_1+m_2} \in \mathcal{A}(m)$*:*

$$\begin{aligned} &V_{[\alpha_1]}\dots V_{[\alpha_{m_1}]}P_t(V_{[\alpha_{m_1+1}]}\dots V_{[\alpha_{m_1+m_2}]}f)(x) \\ &\quad = t^{-(\|\alpha_1\|+\dots+\|\alpha_{m_1+m_2}\|)/2}\mathbb{E}\left[\Phi_{\alpha_1,\dots,\alpha_{m_1+m_2}}(t,x)f(X_t^x)\right], \end{aligned} \tag{62}$$

where $\Phi_{\alpha_1,\dots,\alpha_{m_1+m_2}} \in \mathcal{K}_0^{loc}((k-m-m_1-m_2))$. *Moreover,*

$$\sup_{t\in(0,T]} \mathbb{E}[\left|\,\Phi_{\alpha_1,\dots,\alpha_{m_1+m_2}}(t,x)\,\right|^p] \le C_p(1+|\,x\,|)^{(m_1+m_2)p}. \tag{63}$$

If the vector fields V_i*,* $i = 0,1,\dots,d$ *are uniformly bounded, then* $\Phi_{\alpha_1,\dots,\alpha_{m_1+m_2}} \in \mathcal{K}_0((k-m-m_1-m_2))$. *In particular,*

$$\sup_{t\in(0,T]}\sup_{x\in\mathbb{R}^N} \mathbb{E}[\left|\,\Phi_{\alpha_1,\dots,\alpha_{m_1+m_2}}(t,x)\,\right|^p] < \infty. \tag{64}$$

Proof. Once it is noted that constant functions are in $\mathcal{K}_0$, the proof follows from m_2 applications of Theorem 24 followed by m_1 applications of Corollary 26. The bounds (63), (64) follows likewise. □

Remark 29. Observe that we are able to quantify exactly how the derivatives explode (when t tends to 0)—as functions of x-based on an analysis of the integration by parts factors. In the next section, we shall use the above bounds to deduce sharp gradient bounds for the diffusion semigroup P_t.

2.7 Explicit Bounds

We discuss now how the integration by parts formulae allow the acquisition of several explicit gradient bounds. This section is by no means exhaustive, and for a more complete synopsis of obtainable gradient bounds, one should consult Nee [44]. We will use the following norms and semi-norms

$$\|f\|_\infty = \sup_{x\in\mathbb{R}^N} |f(x)|, \quad \|\nabla f\|_\infty = \max_{i\in\{1,\ldots,d\}} \left\|\frac{\partial f}{\partial x_i}\right\|_\infty, \quad f \in \mathcal{C}_b^\infty(\mathbb{R}^N,\mathbb{R})$$

$$\|f\|_{V,i} = \sum_{u=1}^{i} \sum_{\substack{\alpha_1,\ldots,\alpha_u\in\mathcal{A}_0 \\ \|\alpha_1*\ldots*\alpha_u\|=i}} \left\|V_{[\alpha_1]}\cdots V_{[\alpha_u]} f\right\|_\infty, \quad i\in\mathbb{N} \quad \mathcal{C}_b^{V,i}(\mathbb{R}^N) = \{f : \|f\|_{V,i} < \infty\}$$

$$\|f\|_p = \sum_{i=1}^{p} \left\|\nabla^i f\right\|_\infty, \quad f\in\mathcal{C}_b^p(\mathbb{R}^N,\mathbb{R}), \quad p\in\mathbb{N} \tag{65}$$

$$\left\|\nabla^i f\right\|_\infty = \max_{j_1,\ldots,j_i\in\{1,\ldots,d\}} \left\|\frac{\partial^i f}{\partial x_{j_1}\ldots\partial x_{j_i}}\right\|_\infty$$

$$\|f\|_{p,\infty} = \|f\|_\infty + \|f\|_p \quad f\in\mathcal{C}_b^\infty(\mathbb{R}^N,\mathbb{R}).$$

Remark 30. Note that $\|V_\alpha f\|_\infty \le C\,\|f\|_{|\alpha|}$ for any $\alpha\in\mathcal{A}$ and $f\in\mathcal{C}_b^{|\alpha|}(\mathbb{R}^N,\mathbb{R})$, hence

$$\|\varphi\|_{V,i} \le C\,\|\varphi\|_i \quad i\in\mathbb{N}.$$

Corollary 31. *Let $f\in\mathcal{C}_b^\infty(\mathbb{R}^N,\mathbb{R})$ and $\alpha_1,\ldots,\alpha_{m_1+m_2}\in\mathcal{A}(m)$ be such that $m_1+m_2\le k-m$. Then there is a constant C such that, for any $t\in[0,T]$,*

$$\begin{aligned}&|V_{[\alpha_1]}\ldots V_{[\alpha_{m_1}]}P_t(V_{[\alpha_{m_1+1}]}\ldots V_{[\alpha_{m_1+m_2}]}f)(x)| \\ &\quad = C\|f\|_\infty t^{-(\|\alpha_1\|+\ldots+\|\alpha_{m_1+m_2}\|)/2}(1+|x|)^{m_1+m_2}.\end{aligned} \tag{66}$$

Moreover, if the vector fields $V_0, \ldots, V_d$ are uniformly bounded, then there is a constant C such that, for any $t \in [0, T]$,

$$\| V_{[\alpha_1]} \ldots V_{[\alpha_{m_1}]} P_t (V_{[\alpha_{m_1+1}]} \ldots V_{[\alpha_{m_1+m_2}]} f) \|_\infty = C \| f \|_\infty t^{-(\|\alpha_1\| + \ldots + \|\alpha_{m_1+m_2}\|)/2}. \tag{67}$$

Proof. This now follows easily from Corollary 28. □

The following result will be of use in the next section:

Corollary 32. *Assume that $0 < p \le n \le k - m$, and let $f \in \mathcal{C}_b^\infty(\mathbb{R}^N, \mathbb{R})$. Then there is a constant $C < \infty$ such that for $\alpha_1, \ldots, \alpha_n \in \mathcal{A}(m)$ and any $t \in [0, T]$,*

$$|V_{[\alpha_1]} \ldots V_{[\alpha_n]} P_t f(x)| \le \frac{C\, t^{p/2}}{t^{(\|\alpha_1\| + \ldots + \|\alpha_n\|)/2}} \| f \|_p (1 + | x |)^n. \tag{68}$$

Moreover, if the vector fields $V_0, \ldots, V_d$ are uniformly bounded, then there is a constant C such that, for any $t \in [0, T]$,

$$\| V_{[\alpha_1]} \ldots V_{[\alpha_n]} P_t f) \|_\infty = \frac{C\, t^{p/2}}{t^{(\|\alpha_1\| + \ldots + \|\alpha_n\|)/2}} \| f \|_p. \tag{69}$$

Proof. We prove the result for $p = 1$ as the general case follows along the same lines. The idea behind this gradient bound is that one can "sacrifice" the derivative along $V_{[\alpha_n]}$ to obtain a new integration by parts formula involving the gradient of f. Observe,

$$\begin{aligned} V_{[\alpha_n]} P_t f(x) &= \sum_{i=1}^N V_{[\alpha_n]}^i (x) \partial_i \mathbb{E} [(f \circ X_t)(x)] \\ &= \sum_{j=1}^N \mathbb{E} \left[\partial_j f(X_t^x) \sum_{i=1}^N V_{[\alpha_n]}^i (x) (J_t^x)_{j,i} \right] \\ &= \sum_{j=1}^N \mathbb{E} \left[\partial_j f(X_t^x) \Phi^j (t, x) \right], \end{aligned}$$

where $\Phi^j(t, x) = \sum_{i=1}^N V_{[\alpha_k]}^i (x) (J_t^x)_{j,i} \in \mathcal{K}_0^{\text{loc}}(k - m)$. Hence, following $n - 1$ applications of Theorem 24 to the above expression, we see that:

$$V_{[\alpha_1]} \ldots V_{[\alpha_n]} P_t f(x) = t^{-(\|\alpha_1\| + \ldots \|\alpha_{n-1}\|)/2} \sum_{j=1}^N \mathbb{E} \left[\partial_j f(X_t^x) \Phi_{\alpha_1, \ldots, \alpha_{n-1}}^j (t, x) \right].$$

And therefore

$$|V_{[\alpha_1]}\dots V_{[\alpha_n]}P_t f(x)| \le Ct^{-(\|\alpha_1\|+\dots\|\alpha_{n-1}\|)/2}\|\nabla f\|(1+|x|)^n$$
$$\le \frac{Ct^{1/2}}{t^{(\|\alpha_1\|+\dots+\|\alpha_n\|)/2}}\|\nabla f\|(1+|x|)^n.$$

The last inequality follows because $t^{(1-\|\alpha_n\|)/2} \ge T^{(1-\|\alpha_n\|)/2}$. □

The gradient bounds presented above play the central role in determining the rates of convergence of the numerical approximations presented in the following chapters. In addition, we can use them to deduce the Hörmander's criterion in the particular case when the vector fields V_i, $i = 0, 1, \dots, d$ satisfy the uniform Hörmander condition.

2.8 Smoothness of the Diffusion Semigroup

In this section, we shall assume for simplicity that the vector fields V_i, $i = 0, 1, \dots, d$ are smooth and uniformly bounded. We prove that, under the assumption of Hörmander's criterion, $x \to P_t f(x)$ is a smooth function. This implies the existence and smoothness of the density of the law of the corresponding diffusion. To show this, we make use of the following proposition, provided by Malliavin in [40]:

Proposition 33. *Let μ be a finite measure defined on the Borel σ-algebra $\mathcal{B}(\mathbb{R}^N)$. Assume that for every multi-index α , there is a constant C_α, such that*

$$\left|\int \partial_\alpha f(x)\mu(dx)\right| \le C_\alpha \|f\|_\infty$$

for every smooth f with compact support. Then μ has a density with respect to the Lebesgue measure which is smooth on $\mathbb{R}^N$. In particular, if for every multi-index α, there is a constant C_α such that

$$|\mathbb{E}[(\partial_\alpha f)(X_t^x)]| \le C_\alpha \|f\|_\infty, \tag{70}$$

for every smooth f with compact support, then the law of X_t^x has a density, which is smooth on $\mathbb{R}^N$.

Remark 34. One can "localize" the result in Proposition 33 in the following standard way: Assume that for every $R > 0$ and every multi-index α there is a constant $C_{\alpha,R}$, such that

$$\left|\int \partial_\alpha f(x)\mu(dx)\right| \le C_{\alpha,R} \|f\|_\infty$$

for every smooth f with compact support in the ball $B(0, R)$. Then μ has a density with respect to the Lebesgue measure which is smooth on $\mathbb{R}^N$. To justify this, one uses Proposition 33 to show that for every $R > 0$, $\mu|_{B(0,R)}$ has a smooth density with respect to the Lebesgue measure.

In particular, if for every $R > 0$ and every multi-index α there is a constant $C_{\alpha,R}$, such that

$$|\,\mathbb{E}[(\partial_\alpha f)(X_t^x)]\,| \leq C_{\alpha,R}\,\|f\|_\infty\,. \tag{71}$$

for every smooth f with compact support in the ball $B(0, R)$ then the law of X_t^x has a smooth density with respect to the Lebesgue measure.

Gradient bounds such as (70), (71) may be deduced from the techniques of Kusuoka, provided some extra assumptions are made.

Theorem 35. *Assume that the following holds for all $x \in \mathbb{R}^N$:*

$$Span\{V_{[\alpha]}(x) : \alpha \in \mathcal{A}(m)\} = \mathbb{R}^N. \tag{72}$$

Then the law of X_t^x has a smooth density with respect to the Lebesgue measure.

Note that we may restate (72), as the property that there exists $\epsilon = \epsilon(x) > 0$ such that

$$\sum_{\alpha \in \mathcal{A}(m)} \left(V_{[\alpha]}(x), \xi\right)^2 \geq \epsilon\,|\,\xi\,|^2\,, \tag{73}$$

for all $\xi \in \mathbb{R}^N$, or equivalently: the matrix $(VV^T)(x)$ is invertible $\forall x \in \mathbb{R}^N$, where $V(x) := \left(V_{[\alpha]}^j\right)_{\substack{j=1,\ldots,N \\ \alpha \in \mathcal{A}(m)}}$. Note: upon taking the infimum over all $|\,\xi\,| = 1$, the LHS of (73) is the minimum eigenvalue of this matrix. The inverse must have smooth entries (by the inverse function theorem) and be bounded on compact sets.

Proof. Showing (71), amounts to deriving an integration by parts formula for the partial derivatives ∂_i. This can easily be iterated to obtain any combination of higher partial derivatives. We claim that there exist smooth functions C_α^i such that:

$$\partial_i = \sum_{\alpha \in \mathcal{A}(m)} C_\alpha^i(x) V_{[\alpha]}(x),$$

for all $x \in \mathbb{R}^m$. This can be re-written in matrix form as $\partial_i = VC^i$, where $V(x) := \left(V_{[\alpha]}^j(x)\right)_{\substack{j=1,\ldots,N \\ \alpha \in \mathcal{A}(m)}}$, and $C^i(x) = \left(C_\alpha^i(x)\right)_{\alpha \in \mathcal{A}(m)}$. But it holds that $(VV^T)(x)$ is invertible for all $x \in \mathbb{R}^N$. Therefore, we may choose

$$C^i = V^T(VV^T)^{-1}\partial_i,$$

that is, $C^i_\alpha(x) = (V^T(VV^T)^{-1}\partial_i,)_\alpha(x)$. Clearly, C^i_α is smooth by the inverse function theorem and it is also bounded on compacts. Let φ be a smooth function with compact support in the ball $B(0, R)$. Observe that

$$\mathbb{E}[(\partial_i\varphi)(X_t^x)] = \sum_{\alpha\in\mathcal{A}(m)} \mathbb{E}\left[(C^i_\alpha V_{[\alpha]}\varphi)(X_t^x)\right] = \sum_{\alpha\in\mathcal{A}(m)} \mathbb{E}\left[(\Lambda_R C^i_\alpha V_{[\alpha]}\varphi)(X_t^x)\right],$$

where $\Lambda_R : \mathbb{R}^N \to \mathbb{R}$ is a smooth "truncation" function such that $\Lambda_R(x) = 1$ if $x \in B(0, R)$ and $\Lambda_r(x) = 0$ if $x \notin B(0, 2R)$. We can therefore assume without loss of generality that both C^i_α and $V_{[\alpha]}$ are bounded. By Corollary 26 we deduce that there exists $\Phi'_{\alpha,R}$ such that:

$$\mathbb{E}\left[(\Lambda_R C^i_\alpha V_{[\alpha]}\varphi)(X_t^x)\right] = t^{-\|\alpha\|/2}\mathbb{E}[\Phi'_{\alpha,R}(t,x)\varphi(X_t^x)] \tag{74}$$

and

$$\sup_{t\in(0,T]}\sup_{x\in\mathbb{R}^N} \mathbb{E}[\left| \Phi'_{\alpha,R}(t,x) \right|^p] < \infty. \tag{75}$$

Hence there exists a constant $C_{i,R}$, such that

$$|\,\mathbb{E}[(\partial_i\varphi)(X_t^x)]\,| \le C_{i,R}\, \|\varphi\|_\infty \tag{76}$$

with

$$C_{i,R} = t^{-\|\alpha\|/2} \sum_{\alpha\in\mathcal{A}(m)} \sup_{x\in\mathbb{R}^N} \mathbb{E}\left[\left| \Phi'_{\alpha,R}(t,x) \right|\right] < \infty.$$

The same argument can be done for any partial derivative and the procedure can be iterated for any multi-index α. The result follows by Remark 34. □

2.9 The V0 Condition

Under the UFG condition alone, one cannot gauge any differentiability properties in the direction V_0. Even if we have differentiability in the direction V_0, the norm $\|V_0P_t\varphi\|_\infty$ may explode with arbitrary high rate. Kusuoka has given an explicit class of examples where, for arbitrary integers $l \ge 2$, it holds

$$ct^{-\frac{l}{2}}\, \|\varphi\|_\infty \le \|V_0P_t\varphi\|_\infty \le Ct^{-\frac{l}{2}}\, \|\varphi\|_\infty$$

for some constants $c, C > 0$ (see Propositions 14 and 16 in [30]). However the following condition allows us to have a suitable control in the direction V_0.

Definition 36 (The V0 Condition). Let $\{V_i : i = 0, \ldots, d\}$, be a system of vector fields such that $V_1, \ldots, V_d \in \mathcal{C}_b^{k+1}(\mathbb{R}^N; \mathbb{R}^N)$ and $V_0 \in \mathcal{C}_b^{k}(\mathbb{R}^N; \mathbb{R}^N)$. We say that $\{V_i : i = 0, \ldots, d\}$ satisfy the V0 condition if, there exist uniformly bounded functions $\varphi_\beta \in \mathcal{C}_b^k(\mathbb{R}^N, \mathbb{R})$, with $\beta \in \mathcal{A}(2)$ such that

$$V_0(x) = \sum_{\beta \in \mathcal{A}(2)} \varphi_\beta(x) V_{[\beta]}(x). \tag{77}$$

Condition V0 states that V_0 can be expressed as a linear combination of the vector fields $\{V_1, \ldots V_k\} \cup \left\{\left[V_i, V_j\right], 1 \le i < j \le k\right\}$. This premise is weaker than the ellipticity assumption and has been used, for example, by Jerison and Sánchez–Calle [25] to obtain estimates for the heat kernel. Under the V0 condition all results presented above extend to the differentiability in the direction V_0 as well. For example we have the following equivalent of the corollary 28:

Proposition 37. *Under the same conditions as Theorem 24 and the V0 condition, the following integration by parts formula holds for $m_1 + m_2 \le k - m$ and $\alpha_1, \ldots, \alpha_{m_1+m_2} \in \mathcal{A}(m) \cup \{(0)\}$:*

$$\begin{aligned} & V_{[\alpha_1]} \ldots V_{[\alpha_{m_1}]} P_t (V_{[\alpha_{m_1+1}]} \ldots V_{[\alpha_{m_1+m_2}]} f)(x) \\ & \quad = t^{-(\|\alpha_1\| + \ldots + \|\alpha_{m_1+m_2}\|)/2} \mathbb{E}\left[\Phi_{\alpha_1, \ldots, \alpha_{m_1+m_2}}(t, x) f(X_t^x)\right], \end{aligned} \tag{78}$$

where $\Phi_{\alpha_1,\ldots,\alpha_{m_1+m_2}} \in \mathcal{K}_0^{loc}((k - m - m_1 - m_2))$. Moreover,

$$\sup_{t \in (0,T]} \mathbb{E}\left[\left|\Phi_{\alpha_1,\ldots,\alpha_{m_1+m_2}}(t, x)\right|^p\right] \le C_p (1 + |x|)^{(m_1+m_2)p}. \tag{79}$$

Moreover, if the vector fields V_i, $i = 0, 1, \ldots, d$ are uniformly bounded, then $\Phi_{\alpha_1,\ldots,\alpha_{m_1+m_2}} \in \mathcal{K}_0((k - m - m_1 - m_2))$. In particular,

$$\sup_{t \in (0,T]} \sup_{x \in \mathbb{R}^N} \mathbb{E}\left|\Phi_{\alpha_1,\ldots,\alpha_{m_1+m_2}}(t, x)\right|^p < \infty. \tag{80}$$

From Proposition 37 one can deduce the following corollary similar to Corollary 32

Corollary 38. *Assume $n \le k - m$, and let $f \in \mathcal{C}_b^\infty(\mathbb{R}^N, \mathbb{R})$. Then, under the UFG+V0 conditions, there is a constant $C < \infty$ such that for $\alpha_1, \ldots, \alpha_n \in \mathcal{A}(m) \cup \{(0)\}$:*

$$|V_{[\alpha_1]} \ldots V_{[\alpha_n]} P_t f(x)| \le \frac{C\, t^{1/2}}{t^{(\|\alpha_1\| + \ldots + \|\alpha_n\|)/2}} \|\nabla f\|_\infty (1 + |x|)^N. \tag{81}$$

Moreover, if the vector fields $V_0, \ldots, V_d$ are uniformly bounded, then there is a constant C such that

$$\|V_{[\alpha_1]} \ldots V_{[\alpha_n]} P_t f)\|_\infty = \frac{C\, t^{1/2}}{t^{(\|\alpha_1\|+\ldots+\|\alpha_n\|)/2}} \|\nabla f\|. \tag{82}$$

and for any integer $p > 0$ there is a constant C_p such that

$$\|V_{[\alpha_1]} \ldots V_{[\alpha_n]} P_t f)\|_\infty = \frac{C_p}{t^{(\|\alpha_1\|+\ldots+\|\alpha_n\|)/2}} \|f\|_p. \tag{83}$$

3 Cubature Methods

3.1 Introduction

In this section we will be concerned with numerical approximations of solutions of stochastic differential equations (SDEs). There are two classes of numerical methods for approximating SDEs. The objective of the first is to produce a pathwise approximation of the solution (strong approximation). The second method involves approximating the distribution of the solution at a particular instance in time (weak approximation). For example when one is only interested in the expectation $\mathbb{E}[\varphi(X_t)]$ for some function φ, it is sufficient to have a good approximation of the distribution of the random variable X_t rather than of its sample paths. This observation was first made by Milstein [42] who showed that pathwise schemes and L^2 estimates of the corresponding errors are irrelevant in this context since the objective is to approximate the law of X_t. This section contains approximations that belong to this second class of methods.

Classical results in this area concentrate on solving numerically SDEs for which the so-called "ellipticity condition", or more generally the "Uniform Hörmander condition" (UH), is satisfied. For a survey of such schemes see, for example, Kloeden and Platen [27] or Burrage, Burrage and Tian [6]. Under this condition, for any bounded measurable function φ, $P_t\varphi$ is smooth for any $t > 0$. It is this property upon which the majority of these schemes rely.

For example, the classical Euler–Maruyama scheme requires $P_t\varphi$ to be four times differentiable in order to obtain the optimal rate of convergence. Talay [57, 58] and, independently, Milstein [43] introduced the appropriate methodology to analyse this scheme. They express the error as a difference including a sum of terms involving $P_t\varphi$. Their analysis also shows the relationship between the smoothness of φ and the corresponding error. Talay and Tubaro [59] prove an even more precise result showing that, under the same conditions, the errors corresponding to the Euler–Maruyama and many other schemes can be expanded in terms of powers of the discretization step. Furthermore, Bally and Talay [2] show the existence of such an expansion under a much weaker hypothesis on φ: that φ need only be measurable and bounded (even the boundedness condition can be relaxed). Higher order schemes require additional smoothness properties of $P_t\varphi$ (see for example, Platen and Wagner [52]).

As explained in the previous chapter, Kusuoka and Stroock [32, 33, 34] studied the properties of $P_t\varphi$ under the UFG condition which is weaker. A number of schemes have recently been developed to work under the UFG conditions rather than the ellipticity condition, their convergence depending intrinsically on the above estimates of $V_{[\alpha_1]}\dots V_{[\alpha_n]}P_t\varphi$. A further advantage of this new generation of schemes is a consequence of the classical result stating that the support of $X\,(x)$ is the closure of the set $S = \left\{x^{\varphi} : [0,\infty) \to \mathbb{R}^d\right\}$ where x^{φ} solves the ODE,

$$x_t^{\varphi} = x + \int_0^t V_0(x_s^{\varphi})ds + \sum_{j=1}^{d}\int_0^t V_j\,(x_s^{\varphi})\varphi\,(s)\,ds$$

and $\varphi : [0,\infty) \to \mathbb{R}^d$ is an arbitrary smooth function (see Stroock and Varadhan [54–56], Millet and Sanz-Sole[41]). These schemes attempt to keep the support of the approximating process on the set S. In this way, stability problems that are known to affect classical schemes can be avoided. For example, Ninomyia and Victoir [49] give an explicit example where the Euler–Maruyama approximation fails whilst their algorithm succeeds (see Example 43 below for their algorithm). Their example involves an SDE related to the Heston stochastic volatility model in finance.

In this chapter we give a general criterion for the convergence of a class of weak approximations incorporating this new category of schemes. The criterion is based upon the stochastic Stratonovich–Taylor expansion of $\varphi(X_t)$ and demonstrates how the rate of convergence depends on the smoothness of the test function φ.

For smooth test functions, an equidistant partition of the time interval on which the approximation is sought is optimal. For less smooth functions, this is no longer true. We emphasize that the UFG+V0 conditions are not required for smooth test functions.

3.2 M-Perfect Families

In this section we introduce the concept of an *m-perfect* family. Such families correspond to various weak approximations of SDEs, including the Lyons–Victoir and Ninomiya–Victoir schemes. The main result appears in Theorem 46 and Corollary 47.

For $\alpha = (i_1,\dots,i_r) \in \mathcal{A}$ and $\varphi \in C_b^r(\mathbb{R}^N)$, let $f_{\alpha,\phi}$ be defined as $f_{(i_1,\dots,i_r),\varphi} := V_{i_1}\dots V_{i_r}\varphi$ and $I_{f_{\alpha,\varphi}}(t)$ be the iterated Stratonovich integral

$$I_{f_{\alpha,\varphi}}(t) := \int_0^t\int_0^{s_0}\cdots\left(\int_0^{s_{r-2}} f_{\alpha,\varphi}(X_{s_{r-1}})\circ dW_{s_{r-1}}^{i_1}\right)\circ\cdots\circ dW_{s_1}^{i_{r-1}}\circ dW_{s_0}^{i_r},$$

for $t \geq 0$. If $i_1 = 0$ then $I_{f_{\alpha,\varphi}}(t)$ is well defined for $\varphi \in \mathcal{C}_b^r(\mathbb{R}^N)$. However, if $i_1 \neq 0$ then $I_{f_{\alpha,\varphi}}(t)$ is well defined provided $\varphi \in \mathcal{C}_b^{r+2}(\mathbb{R}^N)$, since the semimartingale property of $f_{\alpha,\varphi}(X)$ is required in the definition of the first Stratonovich integral $\int_0^{s_{r-2}} f_{\alpha,\varphi}(X_{s_{r-1}}) \circ dW_{s_{r-1}}^{i_1}$. Note that the Stratonovich integrals are evaluated innermost first. Finally let

$$I_\alpha(t) := \int_0^t \int_0^{s_0} \cdots \left(\int_0^{s_{r-2}} 1 \circ dW_{s_{r-1}}^{i_1} \right) \circ \cdots \circ dW_{s_1}^{i_{r-1}} \circ dW_{s_0}^{i_r}.$$

Let $\alpha = (i_1, \ldots, i_r) \in \mathcal{A}_0$ be an arbitrary multi-index such that $\|\alpha\| = m \in \mathbb{N}$ (and $|\alpha| = r \in \mathbb{N}$). If m is odd, then $\mathbb{E}[I_\alpha(t)] = 0$ and if m is even then

$$\mathbb{E}[I_\alpha(t)] = \begin{cases} \frac{t^{\frac{m}{2}}}{2^{r-\frac{m}{2}}(\frac{m}{2})!} & \text{if } \alpha \in \mathcal{A}_0^{m,r} \\ 0 & \text{otherwise} \end{cases}, \tag{84}$$

where $\mathcal{A}_0^{m,r}$ is the set of multi-indices $\alpha = \alpha_1 * \cdots * \alpha_{\frac{m}{2}} \in \mathcal{A}_0(m)$ such that each $\alpha_i = (0)$ or (j, j) for some $j \in \{1, \ldots, k\}$. Note that $r - \frac{m}{2}$ is equal to the number of pairs of indices (j, j) occurring in α. A proof of this result can be found in [19].

We state three further results in (85), (86) and (88). The proofs are all elementary and can be found in [19]. The first two give an upper bound on the L^2 norm of $I_{f_{\alpha,\varphi}}(t)$ for smooth φ. The third provides an explicit form for the remainder of $\varphi(X_t)$ when expanded in terms of iterated integrals.

For $\varphi \in \mathcal{C}_b^{\|\alpha\|+2}(\mathbb{R}^N)$ and any multi-index $\alpha = (i_1, \ldots, i_r) \in \mathcal{A}_0$ such that $i_1 \neq 0$, we have[9]

$$\left\| I_{f_{\alpha,\varphi}}(t) \right\|_2 \leq c \left\| f_{\alpha,\varphi} \right\|_\infty t^{\frac{\|\alpha\|}{2}} + c \sum_{i=1}^k \left\| V_i f_{\alpha,\varphi} \right\|_\infty t^{\frac{\|\alpha\|+1}{2}} \tag{85}$$

for some constant $c = c(\alpha) > 0$. For $\varphi \in \mathcal{C}_b^{\|\alpha\|}(\mathbb{R}^N)$ and any multi-index $\alpha = (i_1, \ldots, i_r) \in \mathcal{A}_0$ such that $i_1 = 0$ we have

$$\left\| I_{f_{\alpha,\varphi}}(t) \right\|_2 \leq c \left\| f_{\alpha,\varphi} \right\|_\infty t^{\frac{\|\alpha\|}{2}}. \tag{86}$$

For $m \in \mathbb{N}$, $\varphi \in \mathcal{C}_b^{m+3}(\mathbb{R}^N)$ and $x \in \mathbb{R}^N$, we define the truncation,

$$\varphi_t^m(x) := \varphi(x) + \sum_{\alpha \in \mathcal{A}_0(m)} f_{\alpha,\varphi}(x) I_\alpha(t). \tag{87}$$

[9] In the following, we allow for the constant c to take different values from one line to another.

Then for $t \geq 0$ the remainder is

$$R_{m,t,\varphi}(x) := \varphi(X_t) - \varphi_t^m(x) = \left(\sum_{\|\alpha\|=m+1} + \sum_{\|\alpha\|=m+2, \alpha=0*\beta, \|\beta\|=m} \right) I_{f_{\alpha,\varphi}}(t). \tag{88}$$

In the following, we define a class of approximations of X expressed in terms of certain families of stochastic processes, $\bar{X}(x) = \{\bar{X}_t(x)\}_{t \in [0,\infty)}$ for $x \in \mathbb{R}^N$, which are explicitly solvable. In particular, we can explicitly compute the operator,

$$(Q_t\varphi)(x) = \mathbb{E}[\varphi(\bar{X}_t(x))]. \tag{89}$$

The semigroup P_T will then be approximated by $Q_{h_n}^m Q_{h_{n-1}}^m \ldots Q_{h_1}^m$ where $\{h_j := t_j - t_{j-1}\}_{j=1}^n$ and $\pi_n = \{t_j := (\frac{j}{n})^\gamma T\}_{j=0}^n$ for $n \in \mathbb{N}$, is a sufficiently fine partition of the interval $[0, T]$. In particular $h_j \in [0, 1)$ for $j = 1, \ldots, n$. The underlying idea is that $Q_t\varphi$ will have the same truncation as $P_t\varphi$.

So let $\bar{X}(x) = \{\bar{X}_t(x)\}_{t \in [0,\infty)}$, where $x \in \mathbb{R}^N$, be a family of progressively measurable stochastic processes such that, $\lim_{y \to x_0} \bar{X}_t(y) = \bar{X}_t(x_0)$ $\mathbb{P}$-almost surely, for any $t \geq 0$ and $x_0 \in \mathbb{R}^N$. As a result, the operator Q_t defined in (89) has the property that $Q_t\varphi \in \mathcal{C}_b(\mathbb{R}^N)$ for any $\varphi \in \mathcal{C}_b(\mathbb{R}^N)$. In particular, $Q_t : \mathcal{C}_b(\mathbb{R}^N) \to \mathcal{C}_b(\mathbb{R}^N)$ is a Markov operator.

Definition 39. For $m \in \mathbb{N}$, the family $\bar{X}(x) = \{\bar{X}_t(x)\}_{t \in [0,\infty)}$ where $x \in \mathbb{R}^N$, is said to be **m-perfect** for the process X if there exist a constant $c > 0$ and an integer $M \geq m + 1$ such that for $\varphi \in \mathcal{C}_b^{V,M}(\mathbb{R}^N)$,

$$\sup_{x \in \mathbb{R}^N} |Q_t\varphi(x) - \mathbb{E}[\varphi_t^m(x)]| \leq c \sum_{i=m+1}^{M} t^{i/2} \|\varphi\|_{V,i}. \tag{90}$$

As we can see from (90), the quantity $\mathbb{E}[\varphi_t^m(x)]$ plays the same role as the classical truncation in the standard Taylor expansion of a function. Using (84) we deduce that,

$$\mathbb{E}[\varphi_t^0(x)] = \varphi(x)$$
$$\mathbb{E}[\varphi_t^2(x)] = \varphi(x) + L\varphi(x)t$$
$$\mathbb{E}[\varphi_t^4(x)] = \varphi(x) + L\varphi(x)t + L^2\varphi(x)\frac{t^2}{2},$$

where $L = V_0 + \frac{1}{2}\sum_{i=1}^d V_i^2$. Furthermore, since $\mathbb{E}[I_\alpha(t)] = 0$ for odd $\|\alpha\|$, it follows that $\mathbb{E}[\varphi_t^1(x)] = \mathbb{E}[\varphi_t^0(x)]$, $\mathbb{E}[\varphi_t^3(x)] = \mathbb{E}[\varphi_t^2(x)]$ and $\mathbb{E}[\varphi_t^5(x)] = \mathbb{E}[\varphi_t^4(x)]$.

3.3 Examples

There now follow some examples of m-perfect families corresponding to the semigroup $\{P_t\}_{t\in[0,\infty)}$, the Lyons–Victoir method and the Ninomiya–Victoir algorithm.

Example 40. The family of stochastic processes $\{X_t(x)\}_{t\in[0,\infty)}$, where $x \in \mathbb{R}^N$, is m-perfect. More precisely, there exists a constant $c > 0$ such that for $\varphi \in \mathcal{C}_b^{V,m+2}(\mathbb{R}^N)$,

$$\sup_{x\in\mathbb{R}^N} |P_t\varphi(x) - \mathbb{E}[\varphi_t^m(x)]| \leq c \sum_{i=m+1}^{m+2} t^{i/2} \|\varphi\|_{V,i}\,, \tag{91}$$

Proof. For $\varphi \in \mathcal{C}_b^{V,m+3}(\mathbb{R}^N)$,

$$\left|P_t\varphi(x) - \mathbb{E}[\varphi_t^m(x)]\right| = \left|\mathbb{E}[R_{m,t,\varphi}(x)]\right| = \left|\mathbb{E}[(\sum_{\|\alpha\|=m+1} + \sum_{\|\alpha\|=m+2,\alpha=0*\beta,\|\beta\|=m})I_{f_{\alpha,\varphi}}(t)]\right|$$

Applying inequality (85) to the first sum,

$$\begin{aligned}\sum_{\|\alpha\|=m+1} \left\|I_{f_{\alpha,\varphi}}(t)\right\|_2 &\leq \sum_{\|\alpha\|=m+1} \{c\left\|f_{\alpha,\varphi}\right\|_\infty t^{\frac{m+1}{2}} + c\sum_{i=1}^{k}\left\|V_i f_{\alpha,\varphi}\right\|_\infty t^{\frac{m+2}{2}}\} \\ &\leq c \sum_{i=m+1}^{m+2} t^{i/2}\|\varphi\|_{V,i}\end{aligned} \tag{92}$$

for some constant $\bar{c} > 0$. Applying result (86) to the second sum,

$$\begin{aligned}\sum_{\|\alpha\|=m+2,\alpha=0*\beta,\|\beta\|=m} \left\|I_{f_{\alpha,\varphi}}(t)\right\|_2 &\leq \sum_{\|\alpha\|=m+2,\alpha=0*\beta,\|\beta\|=m} c\left\|f_{\alpha,\psi}\right\|_\infty t \\ &\leq c\,\|\varphi\|_{V,m+2}\, t^{\frac{m+2}{2}}.\end{aligned} \tag{93}$$

The result for $\varphi \in \mathcal{C}_b^{V,m+3}(\mathbb{R}^N)$ follows from combining (92) and (93). Since none of the terms in (91) depend on partial derivatives of order $m+3$, the inequality is also valid for any $\varphi \in \mathcal{C}_b^{V,m+2}(\mathbb{R}^N)$ (a standard approximation method can be used). □

In the following example, the family of processes $\bar{X}(x) = \{\bar{X}_t(x)\}_{t\in[0,1]}$, where $x \in \mathbb{R}^N$, corresponds to the Lyons–Victoir approximation (see [36]). The example involves a set of l finite variation paths, $\omega_1, \ldots, \omega_l \in \mathcal{C}_0^0([0,1],\mathbb{R}^d)$, for some $l \in \mathbb{N}$, together with some weights $\lambda_1, \ldots, \lambda_l \in \mathbb{R}^+$ such that $\sum_{j=1}^{l} \lambda_j = 1$. These paths are said to define a **cubature formula on Wiener Space of degree m** if, for any $\alpha \in \mathcal{A}_0(m)$,

$$\mathbb{E}[I_\alpha(1)] = \sum_{j=1}^{l} \lambda_j I_\alpha^{\omega_j}(1),$$

where,

$$I_{(i_1,\ldots,i_r)}^{\omega_j}(1) := \int_0^1 \int_0^{s_0} \cdots \left(\int_0^{s_{r-2}} d\omega_j^{i_1}(s_{r-1})\right) \cdots d\omega_j^{i_{r-1}}(s_1) d\omega_j^{i_r}(s_0).$$

From the scaling properties of the Brownian motion we can deduce, for $t \geq 0$,

$$\mathbb{E}[I_\alpha(t)] = \sum_{j=1}^{l} \lambda_j I_\alpha^{\omega_{t,j}}(t),$$

where $\omega_{t,1}, \ldots, \omega_{t,l} \in \mathcal{C}_0^0([0,t], \mathbb{R}^d)$ is defined by $\omega_{t,j}(s) = \sqrt{t}\omega_j\left(\frac{s}{t}\right),\ s \in [0,t]$. In other words, the expectation of the iterated Stratonovich integrals $I_\alpha(t)$ is the same under the Wiener measure as it is under the measure,

$$\mathbb{Q}_t := \sum_{j=1}^{l} \lambda_j \delta_{\omega_{t,j}}.$$

Example 41. If we choose $\bar{X}$ to be the stochastic flow defined in (1), but with the driving Brownian motion replaced by the paths $\omega_{t,1}, \ldots, \omega_{t,l}$ defined above then the family of processes, $\{\overline{X}_t(x)\}_{t\in[0,1]}$, with corresponding operator $(\overline{Q}_t\varphi)(x) := \mathbb{E}_{\mathbb{Q}_t}[\varphi(\overline{X}_t(x))]$, is m-perfect. More precisely, there exists a constant $c > 0$ such that for $\varphi \in C_b^{V,m+2}(\mathbb{R}^N)$,

$$\sup_x \left|\overline{Q}_t\varphi(x) - \mathbb{E}[\varphi_t^m(x)]\right| \leq c \sum_{i=m+1}^{m+2} t^{i/2} \|\varphi\|_{V,i}$$

For example, if $(\lambda_j, \omega_{t,j})$ are chosen such that for $l = 2^d$ the paths are $\omega_{t,j} : t \mapsto t(1, z_j^1, .., z_j^d)$ for $j = 1, \ldots, 2^d$ with points $z_j \in \{-1,1\}^d$ and weights $\lambda_j = 2^{-d}$, we obtain a cubature formula of degree 3 and a corresponding 3-perfect family.

Proof. Let us first observe that $I_{\alpha}^{\omega_{t,j}}(t) = t^{\frac{|\alpha|}{2}} I_{\alpha}^{\omega_j}(1)$ Hence, for $\varphi \in C_b^{V,m+2}(\mathbb{R}^N)$,

$$\begin{aligned}
\left|\overline{Q}_t\varphi(x) - \mathbb{E}[\varphi_t^m(x)]\right| &= \left|\mathbb{E}_{\mathbb{Q}_t}[R_{m,t,\varphi}(x)]\right| \\
&\leq \Big(\sum_{\|\alpha\|=m+1} + \sum_{\|\alpha\|=m+2,\alpha=0*\beta,\|\beta\|=m}\Big) \left\| f_{\alpha,\varphi}\right\|_\infty \left\|\mathbb{E}_{\mathbb{Q}_t}[I_\alpha(t)]\right\|_2 \\
&\leq \Big(\sum_{\|\alpha\|=m+1} + \sum_{\|\alpha\|=m+2,\alpha=0*\beta,\|\beta\|=m}\Big) \left\| f_{\alpha,\varphi}\right\|_\infty \sum_{j=1}^{l}\lambda_j \left\| I_\alpha^{\omega_{t,j}}(t)]\right\|_2 \\
&\leq \Big(\sum_{\|\alpha\|=m+1} + \sum_{\|\alpha\|=m+2,\alpha=0*\beta,\|\beta\|=m}\Big) k_\alpha \left\| f_{\alpha,\varphi}\right\|_\infty t^{\frac{|\alpha|}{2}}.
\end{aligned}$$

where $k_\alpha = \sum_{j=1}^{l}\lambda_j \left\| I_\alpha^{\omega_j}(1)\right\|_2$. □

Remark 42. (i) There has been no change to the underlying measure in the example above. Merely a representation in terms of the measure $\mathbb{Q}_t$ has been introduced to ease the computation of $\overline{Q}_t$. More precisely, the family of processes $\{\overline{X}_t(x)\}_{t\in[0,1]}$ where $x \in \mathbb{R}^N$ is constructed as follows. We take,

$$\overline{X}_0(x) = x$$

and then randomly choose a path $\omega_{t,r}$ from the set $\{\omega_{t,1},\ldots,\omega_{t,l}\}$ with corresponding probabilities $(\lambda_1,\ldots,\lambda_l)$. Each process then follows a deterministic trajectory driven by the solution of the ordinary differential equation,

$$d\overline{X}_t = V_0(\overline{X}_t)dt + \sum_{j=1}^{d} V_j(\overline{X}_t)d\omega_{t,k}^j.$$

We can therefore compute the expected value of a functional of $X_t(x)$ as integrals on the path space with respect to the Radon measure $\mathbb{Q}_t$. Hence the identities,

$$\overline{Q}_t\varphi(x) = \mathbb{E}\left[\varphi(\overline{X}_t(x))\right] = \mathbb{E}_{\mathbb{Q}_t}\left[\varphi(\overline{X}_t(x))\right]$$

(ii) The approach adopted by Lyons and Victoir to construct the above approximation resembles the ideas developed by Clark and Newton in a series of papers [10, 11, 45, 46]. Heuristically, Clark and Newton constructed strong approximations of SDEs using flows driven by vector fields which were measurable with respect to the filtration generated by the driving Wiener process. In a similar vein, Castell and Gaines [8] provide a method of strongly approximating the solution of an SDE by means of exponential Lie series.

(iii) The family of processes $\bar{X}(x) = \{\bar{X}_t(x)\}_{\{t\in[0,1],x\in\mathbb{R}^N\}}$ corresponding to the Lyons–Victoir approximation (see [36]) have the fundamental property that they match the expectation of the truncated signature as sketched in the Introduction. In [36], Lyons and Victoir constructed degree 3 and degree 5 approximations in general dimensions. More recently, Gyurko and Lyons developed in [22] higher degree approximation (degree 7, 9 and 11) in low dimensions and show how to extend the cubature method to piece-wise smooth test functions.

For the following example, we will denote by $\exp(Vt)f$ the value at time t of the solution of the ODE $y' = V(y)$, $y(0) = f$ where $V \in \mathcal{C}_b^\infty(\mathbb{R}^N, \mathbb{R}^N)$. In particular, $\exp(Vt)(x)$ is $\exp(Vt)f$ for f being the identity function. The family of processes $Y(x) = \{Y_t(x)\}_{t\in[0,1]}$ below corresponds to the Ninomiya–Victoir approximation (see [49]).

Example 43. Let Λ and Z be two independent random variables such that Λ is Bernoulli distributed $\mathbb{P}(\Lambda = 1) = \mathbb{P}(\Lambda = -1) = \frac{1}{2}$ and $Z = (Z^i)_{i=1}^k$ is a standard normal k-dimensional random variable. Consider the family of processes $Y(x) = \{Y_t(x)\}_{t\in[0,1]}$ defined by

$$Y_t(x) = \begin{cases} \exp(\frac{V_0}{2}t)\prod_{i=1}^k \exp(Z^i V_i t^{1/2})\exp(\frac{V_0}{2}t)(x) & \text{if } \Lambda = 1 \\ \exp(\frac{V_0}{2}t)\prod_{i=1}^k \exp(Z^{k+1-i} V_{k+1-i} t^{1/2})\exp(\frac{V_0}{2}t)(x) & \text{if } \Lambda = -1 \end{cases}$$

with the corresponding operator $(Q_t\varphi)(x) := \mathbb{E}[\varphi(Y_t(x))]$. Then there exists a constant $c > 0$ such that for $\varphi \in C_b^{V,8}(\mathbb{R}^N)$

$$\sup_x \left|\overline{Q}_t\varphi(x) - \mathbb{E}[\varphi_t^5(x)]\right| \le ct^3 \|\varphi\|_{V,6}$$

Hence $\{Y_t(x)\}_{t\in[0,1]}$ is 5-perfect.

Proof. See [13]. ☐

The following lemma is required to prove the main theorem below.

Lemma 44. *For $0 < s \le t \le 1$ and any m-perfect family $\{\overline{X}_t(x)\}_{t\in(0,1]}$ with corresponding operator $Q = \{Q_t\}_{t\in(0,1]}$ we have,*

$$\|P_t(P_s\varphi) - Q_t(P_s\varphi)\|_\infty \le c\|\varphi\|_p \sum_{j=m+1}^M \frac{t^{j/2}}{s^{\frac{j-p}{2}}}, \tag{94}$$

where $\varphi \in \mathcal{C}_b^p(\mathbb{R}^N)$ for $0 \le p < \infty$ and some constant $c > 0$. In particular, for $\varphi \in \mathcal{C}_b^M(\mathbb{R}^d)$,

$$\|P_t(P_s\varphi) - Q_t(P_s\varphi)\|_\infty \le c\|\varphi\|_p t^{\frac{m+1}{2}}. \tag{95}$$

Proof. Since $\mathcal{C}_b^\infty(\mathbb{R}^N)$ is dense in $\mathcal{C}_b^p(\mathbb{R}^d)$ in the topology generated by the norm $||\cdot||_{p,\infty}$ it suffices to prove (94) and (95) only for a function $\varphi \in \mathcal{C}_b^\infty(\mathbb{R}^N)$. By Corollary 32, we have

$$\|P_t\varphi\|_{V,j} = \sum_{i=1}^{j} \sum_{\substack{\alpha_1,\ldots,\alpha_i\in\mathcal{A}_0 \\ \|\alpha_1*\ldots*\alpha_i\|=j}} \left\|V_{[\alpha_1]}\cdots V_{[\alpha_i]}P_t\varphi\right\|_\infty$$

$$\leq \sum_{i=1}^{j} \sum_{\substack{\alpha_1,\ldots,\alpha_i\in\mathcal{A}_0 \\ \|\alpha_1*\ldots*\alpha_i\|=j}} \frac{c\,\|\varphi\|_p}{t^{(\|\alpha_1*\ldots*\alpha_i\|-p)/2}} \leq \frac{c\,\|\varphi\|_p}{t^{\frac{j-p}{2}}}$$

Then (94) and (95) follow from the definition of an m-perfect family. □

The family of processes $\bar{X}(x) = \{\bar{X}_t(x)\}_{t\in[0,\infty)}$ below corresponds to the Kusuoka approximation. We recall that Kusuoka's result requires only the UFG condition.

Example 45. A family of random variables $\{Z_\alpha : \alpha \in \mathcal{A}_0\}$ is said to be **m-moment** similar if $\mathbb{E}[|Z_\alpha|^r] < \infty$ for any $r \in \mathbb{N}$, $\alpha \in \mathcal{A}_0$ and $Z_{(0)} = 1$ with,

$$\mathbb{E}[Z_{\alpha_1}\ldots Z_{\alpha_j}] = \mathbb{E}[I_{\alpha_1}\ldots I_{\alpha_j}]$$

for any $j = 1,\ldots,m$ and $\alpha_1,\ldots,\alpha_j \in \mathcal{A}_0$ such that $\|\alpha_1\|+\cdots+\|\alpha_j\| \leq m$ where I_α is defined as above.
Let $\{Z_\alpha : \alpha \in \mathcal{A}_0\}$ be a family of m-moment similar random variables and let $\bar{X}(x) = \{\bar{X}_t(x)\}_{t\in[0,\infty)}$ be the family of processes,

$$\bar{X}_t(x) = \sum_{j=0}^{m} \frac{1}{j!} \sum_{\substack{\alpha_1,\ldots,\alpha_j\in\mathcal{A}_0, \\ \|\alpha_1\|+\cdots+\|\alpha_j\|\leq m}} t^{\frac{\|\alpha_1\|+\cdots+\|\alpha_j\|}{2}} (P^0_{\alpha_1}\ldots P^0_{\alpha_j})(V_{[\alpha_1]}\ldots V_{[\alpha_j]}H)(x) \tag{96}$$

where $H : \mathbb{R}^N \to \mathbb{R}^N$ is defined $H(x) = x$ and

$$P^0_\alpha := |\alpha|^{-1} \sum_{j=0}^{|\alpha|} \frac{(-1)^{j+1}}{j} \sum_{\beta_1*\ldots*\beta_j=\alpha} Z_{\beta_1}\ldots Z_{\beta_j}$$

with the corresponding operator $Q = \{Q_t\}_{t\in(0,1]}$ defined by,

$$Q_t\varphi(x) = \mathbb{E}[\varphi(\bar{X}_t(x))]$$

for $\varphi \in \mathcal{C}_b(\mathbb{R}^N)$ Then,

$$\|P_{t+s}\varphi - Q_t P_s\varphi\|_\infty \le c\,\|\nabla\varphi\|_\infty \sum_{j=m+1}^{m^{m+1}} \frac{t^{j/2}}{s^{\frac{j-1}{2}}} \tag{97}$$

for some constant $c > 0$.

Proof. See Definition 1, Theorem 3 and Lemma 18 in Kusuoka[29] for (97). □

The family $\bar{X}(x)$, $x \in \mathbb{R}^N$ as defined in (96) is not m-perfect. However, inequality (97) is a particular case of (94) where $p = 1$ and $M = m^{m+1}$. Since (94) is the only result required to obtain (98), we deduce from the proof of Theorem 46 that (98), with $p = 1$, holds for Kusuoka's method as well. Similarly part (ii) of Corollary 47 holds for Kusuoka's method. For numerical algorithms related to the family $\bar{X}(x)$, $x \in \mathbb{R}^N$ as defined in (96) see [31, 47, 48]. In particular, paper [48] uses a control on the computational effort based on the same algorithm (the TBBA) as the one employed in Sect. 3.5.

The set of vector fields appearing in (96) belong to the Lie algebra generated by the original vector fields $\{V_0, V_1, \ldots, V_d\}$. Ben Arous [1] and Burrage and Burrage [5] employ the same set of vector fields to produce *strong* approximations of solutions of SDEs. Notably, the same ideas appear much earlier in Magnus [39], in the context of approximations of the solution of linear (deterministic) differential equations. Castell [7] also gives an explicit formula for the solution of an SDE in terms of Lie brackets and iterated Stratonovich integrals.

3.4 Rates of Convergence

We now prove our main result on m-perfect families, the gist of which can be conveyed by the concept of local and global order of an approximation. Local order measures how close an approximation is to the exact solution on a sub-interval of the integration, given an exact initial condition at the start of that subinterval. The global order of an approximation looks at the build up of errors over the entire integration range. The theorem below states that, in the best possible case, the global order of an approximation obtained using an m-perfect family is one less than the local order. More precisely, for a suitable partition, the global error is of order $\frac{m-1}{2}$ whilst the local error is of order $\frac{m+1}{2}$.

Let us define the function,

$$\Upsilon^p(n) = \begin{cases} n^{-\frac{1}{2}\min(\gamma p,(m-1))} & \text{if } \gamma p \neq m-1 \\ n^{-(m-1)/2}\ln n & \text{for } \gamma p = m-1 \end{cases}.$$

In the following,

$$\mathcal{E}^{\gamma,n}(\varphi) := \left\| P_T\varphi - Q^m_{h_n} Q^m_{h_{n-1}} \cdots Q^m_{h_1}\varphi \right\|_\infty$$

for $\gamma \in \mathbb{R}$, $n \in \mathbb{N}$.

Theorem 46. *Let $T, \gamma > 0$ and $\pi_n = \{t_j = (\frac{j}{n})^\gamma T\}_{j=0}^n$ be a partition of the interval $[0, T]$ where $n \in \mathbb{N}$ is such that $\{h_j = t_j - t_{j-1}\}_{j=1}^n \subseteq (0, 1]$. Then for any m-perfect family $\{\overline{X}_t(x)\}_{t\in[0,T]}$ with corresponding operator $Q = \{Q_t\}_{t\in(0,1]}$ we have, for $\varphi \in \mathcal{C}_b^p(\mathbb{R}^N)$ where $p = 1, \ldots, m$,*

$$\mathcal{E}^{\gamma,n}(\varphi) \le c\Upsilon^p(n)\|\varphi\|_p + \left\|P_{h_1}\varphi - Q_{h_1}^m\varphi\right\|_\infty \tag{98}$$

for some constant $c \equiv c(\gamma, M, T) > 0$ where $M \ge m+1$, as in Definition 39. In particular, if $\gamma \ge \frac{m-1}{p}$ then,

$$\mathcal{E}^{\gamma,n}(\varphi) \le \frac{c}{n^{\frac{m-1}{2}}}\|\varphi\|_p + \left\|P_{h_1}\varphi - Q_{h_1}^m\varphi\right\|_\infty.$$

Proof. We have,

$$\begin{aligned}
\mathcal{E}^{\gamma,n}(\varphi) &= P_{h_n}(P_{T-h_n}\varphi) - Q_{h_n}^m(P_{T-h_n}\varphi) \\
&\quad + \sum_{j=1}^{n-1} Q_{h_n}^m \ldots Q_{h_{j+1}}^m (P_{T-h_{j+1}-\cdots-h_n}\varphi - Q_{h_j}^m P_{T-h_j-\cdots-h_n}\varphi) \\
&= P_{h_n}(P_{t_{n-1}}\varphi) - Q_{h_n}^m(P_{t_{n-1}}\varphi) \\
&\quad + \sum_{j=1}^{n-1} Q_{h_n}^m \ldots Q_{h_{j+1}}^m (P_{h_j}\left(P_{t_{j-1}}\varphi\right) - Q_{h_j}^m\left(P_{t_{j-1}}\varphi\right)).
\end{aligned}$$

By Lemma 44, there exists a constant $c > 0$ such that,

$$\left\|P_{h_n}(P_{t_{n-1}}\varphi) - Q_{h_n}^m(P_{t_{n-1}}\varphi)\right\|_\infty \le c\|\varphi\|_p \sum_{l=m+1}^{M} \frac{h_n^{l/2}}{t_{n-1}^{\frac{l-p}{2}}}$$

Since P is a semigroup and $Q_{h_j}^m$ is a Markov operator for $j = 2, \ldots, n-1$,

$$\begin{aligned}
\left\|Q_{h_n}^m \ldots Q_{h_{j+1}}^m (P_{h_j}\left(P_{t_{j-1}}\varphi\right) - Q_{h_j}^m\left(P_{t_{j-1}}\varphi\right))\right\|_\infty &\le \left\|P_{h_j}\left(P_{t_{j-1}}\varphi\right) - Q_{h_j}^m\left(P_{t_{j-1}}\varphi\right)\right\|_\infty \\
&\le c\|\varphi\|_p \sum_{l=m+1}^{M} \frac{h_j^{l/2}}{t_{j-1}^{\frac{l-p}{2}}}
\end{aligned}$$

for some $c > 0$. Finally, since $Q_{h_j}^m$ is a Markov operator, it follows from (102) that,

$$\left\|Q_{h_n}^m \ldots Q_{h_2}^m(P_{h_1}\varphi - Q_{h_1}^m\varphi)\right\|_\infty \le \left\|P_{h_1}\varphi - Q_{h_1}^m\varphi\right\|_\infty$$

Combining these last four results gives,

$$\mathcal{E}^{\gamma,n}(\varphi) = \left\| P_T\varphi - Q_{h_n}^m \dots Q_{h_1}^m \varphi \right\|_\infty \le \left\| P_{h_1}\varphi - Q_{h_1}^m \varphi \right\|_\infty + c\,||\varphi||_p \sum_{j=2}^{n} \sum_{l=m+1}^{M} \frac{h_j^{l/2}}{t_{j-1}^{\frac{l-p}{2}}}.$$

It follows, almost immediately from the definition of h_j that,

$$h_j = \frac{\gamma T (j-1)^{\gamma-1}}{n^\gamma} \int_{j-1}^{j} \left(\frac{u}{j-1}\right)^{\gamma-1} du,$$

but for $j \in \{2,\dots,n\}$, $(\frac{u}{j-1})^{\gamma-1} \le \max[(\frac{j}{j-1})^{\gamma-1}, 1] \le \max[2^{\gamma-1}, 1]$. Hence for $l = m+1, \dots, M$,

$$\frac{h_j^{l/2}}{t_{j-1}^{(l-p)/2}} \le \frac{(\frac{\gamma T (j-1)^{\gamma-1}}{n^\gamma} \max[2^{\gamma-1}, 1])^{l/2}}{\left(\left(\frac{j-1}{n}\right)^\gamma T\right)^{(l-p)/2}}$$

$$\le c(\frac{T}{n^\gamma})^{\frac{l}{2}-\frac{(l-p)}{2}} (j-1)^{\frac{(\gamma-1)l}{2}-\frac{\gamma(l-p)}{2}} = c(\frac{T}{n^\gamma})^{\frac{p}{2}} (j-1)^{\frac{\gamma p - l}{2}}$$

where $c = \max[1, (\gamma \max[2^{\gamma-1}, 1])^{M/2}]$. It follows that,

$$\sum_{l=m+1}^{M} \frac{h_j^{l/2}}{t_{j-1}^{(l-p)/2}} \le c \left(\frac{1}{n}\right)^{\frac{\gamma p}{2}} \sum_{l=m+1}^{M} (j-1)^{\frac{\gamma p - l}{2}}.$$

Since $\sum_{l=m+1}^{M} (j-1)^{\frac{\gamma p - l}{2}} = (j-1)^{\frac{\gamma p - (m+1)}{2}} \sum_{l=0}^{M-(m+1)} (j-1)^{-\frac{l}{2}} \le (j-1)^{\frac{\gamma p-(m+1)}{2}} M$ we have,

$$\sum_{l=m+1}^{M} \frac{h_j^{l/2}}{t_{j-1}^{(l-p)/2}} \le M \left(\frac{1}{n}\right)^{\frac{\gamma p}{2}} (j-1)^{\frac{\gamma p - (m+1)}{2}} \tag{99}$$

We now consider (99) for three different ranges of γ.
For $\gamma \in \left(0, \frac{m-1}{p}\right)$, $\sum_{j=2}^{n} (j-1)^{\frac{\gamma p-(m+1)}{2}} \le \sum_{j=2}^{\infty} (j-1)^{\frac{\gamma p-(m+1)}{2}}$ and since the series on the right hand side is convergent, we have,

$$n^{-\frac{\gamma p}{2}} \sum_{j=2}^{n} (j-1)^{\frac{\gamma p-(m+1)}{2}} \le c n^{-\frac{\gamma p}{2}}$$

for some constant $c = c(\gamma, M) > 0$.

For $\gamma = \frac{m-1}{p}$, $\sum_{j=2}^{n}(j-1)^{-1} \leq c \ln n$ for some constant $c \equiv c(\gamma, M) > 0$ so we have,

$$n^{-\frac{\gamma p}{2}} \sum_{j=2}^{n} (j-1)^{\frac{\gamma p-(m+1)}{2}} \leq c n^{-\frac{(m-1)}{2}} \ln n.$$

For $\gamma > \frac{m-1}{p}$, we have

$$\sum_{j=2}^{n} (\frac{j-1}{n})^{\frac{\gamma p-(m+1)}{2}} \frac{1}{n} \leq c \int_0^1 x^{\frac{\gamma p-(m+1)}{2}} dx = c \int_0^1 x^{-1+\frac{\gamma p-(m-1)}{2}} dx < \infty$$

so,

$$n^{-\frac{\gamma p}{2}} \sum_{j=2}^{n} (j-1)^{\frac{\gamma p-(m+1)}{2}} = n^{-\frac{m-1}{2}} \sum_{j=2}^{n} \left(\frac{j-1}{n}\right)^{\frac{\gamma p-(m+1)}{2}} \frac{1}{n} \leq c n^{-\frac{m-1}{2}}. \qquad \square$$

We observe that the rate of convergence is the controlled by the maximum between $\Upsilon(n)$ and the rate at which $\left\| P_{h_1}\varphi - Q_{h_1}^m \varphi \right\|_\infty$ converges to 0. We define $\bar{\Upsilon}^{k_1,k_2}(n) := \Upsilon^{k_1}(n) + n^{-\frac{\gamma k_2}{2}}$. We have the following corollary:

Corollary 47. *(i) For any $\varphi \in \mathcal{C}_b^M(\mathbb{R}^N)$,*

$$\mathcal{E}^{\gamma,n}(\varphi) \leq c \bar{\Upsilon}^{m+1,m+1}(n) \|\varphi\|_M .$$

for some constant $c > 0$. In particular, if $\gamma \geq 1$, then $\mathcal{E}^{\gamma,n}(\varphi) \leq \frac{c}{n^{\frac{m-1}{2}}} \|\varphi\|_M$.

(ii) If there exists a constant $c > 0$ independent of t such that,

$$\sup_{x \in \mathbb{R}^N} \left| \bar{X}_t(x) - x \right| \leq c\sqrt{t}, \tag{100}$$

then, for any $\varphi \in \mathcal{C}_b^1(\mathbb{R}^N)$,

$$\mathcal{E}^{\gamma,n}(\varphi) \leq c \bar{\Upsilon}^{1,1}(n) \|\varphi\|_1$$

for some constant $c > 0$. In particular, if $\gamma \geq m-1$, then $\mathcal{E}^{\gamma,n}(\varphi) \leq \frac{c}{n^{\frac{m-1}{2}}} \|\varphi\|_1$.

(iii) if there exist constants $c, \bar{c} > 0$ independent of t such that,

$$\| P_t\varphi - Q_t^m \varphi \|_\infty \leq c t^{\frac{\bar{c}}{2}} \|\varphi\|_l , \tag{101}$$

then, for any $\varphi \in \mathcal{C}_b^l(\mathbb{R}^N)$ *where* $1 < l < M$*, we have*

$$\mathcal{E}^{\gamma,n}(\varphi) \leq c\bar{\Upsilon}^{l,\bar{c}}(n)\,\|\varphi\|_l$$

for some constant $c > 0$*, In particular, if* $\gamma \geq m-1$*, then* $\mathcal{E}^{\gamma,n}(\varphi) \leq \frac{c}{n^{\frac{m-1}{2}}}\|\varphi\|_l$ *.*

Proof. (i) The result follows from Theorem 46 and the definition of an m-perfect family.

(ii) If $\varphi \in \mathcal{C}_b(\mathbb{R}^N)$ is Lipschitz then,

$$|Q_t\varphi(x) - \varphi(x)| \leq c\,||\nabla\varphi||_\infty\sqrt{t} \tag{102}$$

hence,

$$\left\| P_{h_1}\varphi - Q_{h_1}^m\varphi \right\|_\infty \leq c\,||\varphi||_1\sqrt{t}.$$

(iii) The result follows from Theorem 46 and (101). □

Finally we define μ_t to be the law of X_t, that is $\mu_t(\varphi) = E[\varphi(X_t)]$ for $\varphi \in \mathcal{C}_b(\mathbb{R}^N)$. We also define μ_t^N to be the probability measure defined by,

$$\mu_t^N(\varphi) = \mathbb{E}\left[Q_{h_n}^m Q_{h_{n-1}}^m \ldots Q_{h_1}^m\varphi(X_0)\right] = \int_{\mathbb{R}^N} Q_{h_n}^m Q_{h_{n-1}}^m \ldots Q_{h_1}^m\varphi(x)\,\mu_0(dx)$$

for $\varphi \in \mathcal{C}_b(\mathbb{R}^N)$ and introduce the family of norms on the set of signed measures:

$$|\mu|_l = \sup\left\{|\mu(\varphi)|, \varphi \in \mathcal{C}_b^l(\mathbb{R}^N), \|\varphi\|_{l,\infty} \leq 1\right\}, \quad l \geq 1.$$

Obviously, $|\mu|_l \leq |\mu|_{l'}$ if $l \leq l'$. In other words, the higher the value of l, the coarser the norm. We have the following:

Corollary 48. *(i) For* $l \geq M$*, we have* $\left|\mu_t - \mu_t^N\right|_l \leq c\bar{\Upsilon}^{m+1,m+1}(n)$*. In particular, if* $\gamma \geq 1$*, then* $\left|\mu_t - \mu_t^N\right|_l \leq \frac{c}{n^{\frac{m-1}{2}}}$*.*

(ii) If (100) is satisfied then $\left|\mu_t - \mu_t^N\right|_l \leq c\bar{\Upsilon}^{1,1}(n)$*. In particular, if* $\gamma \geq m-1$*, then* $\left|\mu_t - \mu_t^N\right|_l \leq \frac{c}{n^{\frac{m-1}{2}}}$*.*

(iii) If (101) is satisfied then $\left|\mu_t - \mu_t^N\right|_l \leq c\bar{\Upsilon}^{l,c_{23}}(n)$*. In particular, if* $\gamma \geq m-1$*, then* $\left|\mu_t - \mu_t^N\right|_l \leq \frac{c}{n^{\frac{m-1}{2}}}$*.*

Remark 49. We deduce that there is a payoff between the rate of convergence and the coarseness of the norm employed: the finer the norm the slower the rate of convergence. Hence intermediate results such as part *(iii)* of Corollaries 47 and 48 may prove useful in subsequent applications. The additional constraint (101) holds, for example, for the Lyons–Victoir method, as a cubature formula of degree m is also a cubature formula of degree m' for $m' \leq m$. Similarly, it holds for Kusuoka's approximation since an m-similar family is also m'-similar for any $m' \leq m$.

3.5 Cubature and TBBA

In this section we discuss an algorithm that is used to control the computational effort required for the implementation of the Lyons–Victoir cubature method. This method suffers from the usual drawback of any tree based method, namely an exponentially increasing support. This is not an issue in low dimensional problems or when only a sparse partition is used. However, the exponential growth is a major hurdle in more complex and/or high-dimensional problems. To the best of our knowledge, currently, there exist two methods that may be applied to control this growth: The recombination method of Litterer and Lyons [35] and the tree based branching algorithm (TBBA) of Crisan and Lyons [14]. The application of the former to the cubature method has been extensively discussed in [35], where as here, we focus on the TBBA.[10]

The idea behind the TBBA is to construct a finite random measure with a support of size less than a pre-determined value that is an unbiased, minimal variance estimator of the original measure. The method insures that every point in the support of the original measure remains in the support of the resulting measure with a probability (approximately) proportional to its original weight. To fix ideas, let us consider the cubature measure $\mathbb{Q}_1^m$ of degree $m \geq 3$ supported on the paths $\omega_1, \ldots, \omega_{c_d^m}$ with corresponding weights $\lambda_1, \ldots, \lambda_{c_d^m}$, $c_d^m \in \mathbb{N}_+$. As usual we may consider by scaling, cubature measures $\mathbb{Q}_t^m$ on any interval $[0, t]$. Let $\Xi_{t,x}(\omega)$, $\omega \in C_{0,bv}\left([0,t]; \mathbb{R}^d\right)$ denote the solution at time t of the ODE

$$\begin{cases} dy_{t,x} = \sum_{j=0}^{d} V_j(y_{t,x}) d\omega^j(t) \\ y_{0,x} = x \end{cases}. \tag{103}$$

Consider also a partition $\pi := \{0 = t_0 < t_1 < \ldots < t_n = t\}$ of $[0, t]$. By iterating the cubature measure along this partition and solving the successive ODEs (see Remark 42), we generate a collection of discrete measures $\{\mathbb{Q}_{t_k}^m\}_{k \leq n}$, where the cardinality of the support of the measure $\mathbb{Q}_{t_k}^m$ is $\left(c_d^m\right)^k$.

We wish to replace the measure $\mathbb{Q}_{t_k}^m$ by a random measure $\tilde{\mathbb{Q}}_{t_k}^m$ whose support is included in the support of the measure $\mathbb{Q}_{t_k}^m$ and whose cardinality is at most N (with $N < \left(c_d^m\right)^k$). Moreover we want $\tilde{\mathbb{Q}}_{t_k}^m$ to be an unbiased minimal variance estimator of $\mathbb{Q}_{t_k}^m$ in a sense that we will make explicit below. To handle the additional

[10] The TBBA has also been used to control on the computational effort for a class of numerical algorithm using the family $\bar{X}(x)$, $x \in \mathbb{R}^d$ as defined in (96), see [47] for details.

randomness we introduce an additional probability space $\left(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}\right)$ which supports the random probability measure $\tilde{\mathbb{Q}}_{t_k}^m$. We will require that $\tilde{\mathbb{E}}\left[\tilde{\mathbb{Q}}_{t_k}^m\right] = \mathbb{Q}_{t_k}^m$, where $\tilde{\mathbb{E}}$ denotes integration with respect to $\tilde{\mathbb{P}}$. Let

$$\mathbb{Q}_{t_k}^m = \sum_{j=1}^{c_d^m} \lambda_j \delta_{\gamma_j}, \quad \gamma_j = \omega_{\delta_1, i_1} \otimes \ldots \otimes \omega_{\delta_k, i_k}, \text{ for some } i_1, \ldots, i_k = 1, \ldots, c_d^m,$$

where $\omega_i \otimes \omega_j$ denotes the concatenation of two paths. We will construct a random *probability* measure $\tilde{\mathbb{Q}}_{t_k}^m$ such that

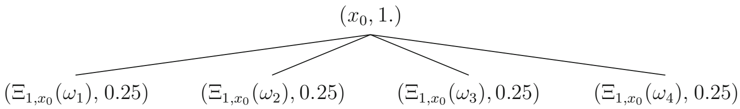
$$\tilde{\mathbb{Q}}_{t_k}^m(\gamma) = \begin{cases} \frac{\lfloor N\mathbb{Q}_{t_k}^m(\gamma)\rfloor}{N} & \text{with probability } 1 - \{N\mathbb{Q}_k^m(\gamma)\} \\ \frac{\lfloor N\mathbb{Q}_{t_k}^m(\gamma)\rfloor + 1}{N} & \text{with probability } \{N\mathbb{Q}_{t_k}^m(\gamma)\} \end{cases}, \gamma \in \text{supp}(\mathbb{Q}_{t_k}^m), \tag{104}$$

where for any real number y, $\lfloor y\rfloor$ denotes the lower integer part and $\{y\}$ the fractional part, $\{y\} = y - \lfloor y\rfloor$. As a result each point in the support of $\mathbb{Q}_k^m(\gamma)$ has either mass 0 (i.e. it does not appear in the support of $\tilde{\mathbb{Q}}_{t_k}^m$ or its mass is an integer multiple of $1/N$. Since $\tilde{\mathbb{Q}}_{t_k}^m$ is a probability measure, its support cannot therefore have cardinality larger than N and is included in the support of $\mathbb{Q}_{t_k}^m$. If $\tilde{\mathbb{Q}}_{t_k}^m(\gamma)$ has distribution described by (104) for any $\gamma \in \text{supp}(\mathbb{Q}_{t_k}^m)$, it is clearly an unbiased estimator of $\mathbb{Q}_{t_k}^m$, that is $\tilde{\mathbb{E}}\left[\tilde{\mathbb{Q}}_{t_k}^m\right] = \mathbb{Q}_{t_k}^m$. Moreover it has minimal variance amongst all unbiased estimators of $\mathbb{Q}_{t_k}^m$ for which the mass associated to any element in the support of the original measure takes values in the set $\{0, \frac{1}{N}, \frac{2}{N}, \ldots, 1\}$. See [14] pg. 344–345 for further optimality properties of $\tilde{\mathbb{Q}}_{t_k}^m$.

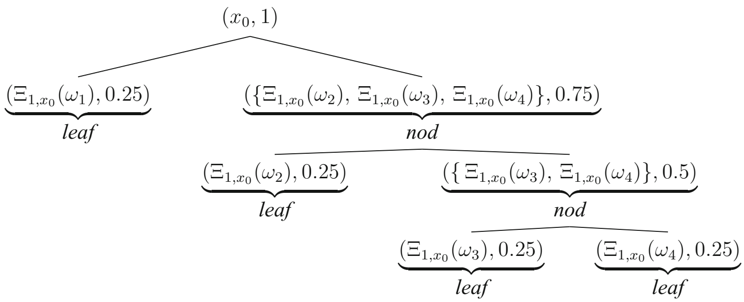
The algorithm that produces the random probability measure $\hat{\mathbb{Q}}_k^m$ from $\mathbb{Q}_k^m$ such that (104) is satisfied for every element in the support of the cubature measure is the subject of Theorem 2.6 of [14]. The idea is to embed the support of the cubature measure into a binary tree and distribute the weights recursively, targeting distribution (104) at every nod/leaf.

Each element in the support of $\mathbb{Q}_{t_k}^m$ is associated to the end nodes or leaves of the binary tree. We associate a weight to each leaf equal to the mass of the corresponding element in the support of $\mathbb{Q}_{t_k}^m$. We then recursively associate a weight to each of the (intermediate) nodes, equal to the sum of the weights of its offspring nodes. Eventually we associate weight 1 to the root node (the sum of the masses of all the elements in the support of $\mathbb{Q}_{t_k}^m$). We note that any tree (not necessarily a binary one) can be embedded into a binary tree as follows: For each intermediate node, we separate the set of its offsprings nodes in two sets. On the left, we take a singleton consisting of the first of the offsprings nodes and on the right we add a new intermediate node with offsprings corresponding to the rest of the offsprings of the original node. We then apply the same procedure to the intermediate node and repeat the process until we are left with only two offspring nodes which we keep as part of the new tree. The next example explain this procedure further through a concrete example.

Example 50. Let us consider the cubature method of order 3 in dimension 2. Starting from x_0 (the initial condition for the forward diffusion) we take on step forward say at time 1. This produces a measure with four elements in its support (see Example 41), as there are four paths that define the cubature formula,[11] all carrying equal weight. Schematically, $\mathbb{Q}_1^m$ looks as in the figure below



We embed the above tree into the following (by no means unique) binary tree:



Notice how every node carries the total weight of all of its offspring leaves.

We will next describe how one distributes the mass according to TBBA *per family*, hence achieving the distribution (104) at every point in the support of the measure. The reasoning relies of course on the structure of the binary tree.

Any path $\gamma \in \operatorname{supp}\left(\mathbb{Q}_k^m\right)$ carries the weight $\lambda_x = \mathbb{Q}_k^m(\gamma)$ which is the product of the cubature weights that correspond to the ODEs we solve to arrive at $x \equiv \Xi_{t_k,x_0}(\gamma)$ along the path γ. Assume that we have assigned to x the random weight $\hat{\lambda}_x = \tilde{\mathbb{Q}}_{t_k}^m(\gamma)$ distributed according to (104). The following algorithm shows how one assigns the corresponding weights to any of the offsprings of x:

[11] These are the straight lines connecting the origin with $(-1,-1)$, $(1,-1)$, $(-1,1)$, $(1,1)$.

Algorithm 1 TBBA(x, λ_x, $\hat{\lambda}_x$)

Require: $\lambda_1, \ldots, \lambda_{c_d^m}$ {The cubature weights and c_d^m is the cubature dimension.}
Define $\lambda_{i:n} = \sum_{j=i}^{c_d^m} \lambda_j$
Declare $\hat{\lambda}_1, \ldots, \hat{\lambda}_{c_d^m}$ and $\hat{\lambda}_{1:c_d^m}, \ldots, \hat{\lambda}_{c_d-1:c_d^m}$
{$\hat{\lambda}_1, \ldots, \hat{\lambda}_{c_d^m}$ store the TBBA weights at every leaf}
{ whereas the $\hat{\lambda}_{1:c_d^m}, \ldots, \hat{\lambda}_{c_d^m-1:c_d^m}$ store the TBBA weights at every nod.}
Set $\hat{\lambda}_{1:c_d^m} = \hat{\lambda}_x$.
for $i = 1$ **to** $N_m - 1$ **do**
  $u_i(x) \sim U[0,1]$, {Draw uniform}
  if $\left(\{N\lambda_x\lambda_{i:c_d^m}\} = \{N\lambda_x\lambda_i\} + \{N\lambda_x\lambda_{i+1:c_d^m}\}\right)$ **then**
    if $\left(u_i(x) < \frac{\{N\lambda_x\lambda_i\}}{\{N\lambda_x\lambda_{i:c_d^m}\}}\right)$ **then**
      $\hat{\lambda}_i = \frac{\lfloor N\lambda_x\lambda_i \rfloor}{N} + \hat{\lambda}_{i:c_d^m} - \frac{\lfloor N\lambda_x\lambda_{i:c_d^m} \rfloor}{N}$
    else
      $\hat{\lambda}_i = \frac{\lfloor N\lambda_x\lambda_i \rfloor}{N}$
    end if
  else
    if $\left(u_i(x) < \frac{1-\{N\lambda_x\lambda_i\}}{1-\{N\lambda_x\lambda_{i:c_d^m}\}}\right)$ **then**
      $\hat{\lambda}_i = \frac{\lfloor N\lambda_x\lambda_i \rfloor + 1}{N} + \hat{\lambda}_{i:c_d^m} - \frac{\lfloor N\lambda_x\lambda_{i:c_d^m} \rfloor + 1}{N}$
    else
      $\hat{\lambda}_i = \frac{\lfloor N\lambda_x\lambda_i \rfloor + 1}{N}$
    end if
  end if
  if ($\hat{\lambda}_i > 0$) **then**
    Solve the ODE (103) in the direction of path ω_i
    Store offspring $(x_i, \lambda_{x_i}, \hat{\lambda}_i)$, $\lambda_{x_i} = \lambda_x\lambda_i$
  end if
  Set $\hat{\lambda}_{i+1:c_d^m} = \hat{\lambda}_{i:c_d^m} - \hat{\lambda}_i$.
end for

Remark 51. All uniform random variables used in **Algorithm 1** are drawn independent of each other.

We apply **Algorithm 1** recursively until all nodes in the support of the cubature measure are assigned their corresponding random weights $\tilde{\lambda}_x$. We continue in this way until we reach the leaves of the tree. Recall that there can be at most N elements in the support of the original measure that get assigned a positive weight by the TBBA, hence indeed the new measure $\hat{\mathbb{Q}}_k^m$ will have a support of cardinality at most N.

We denote the set of all nodes corresponding to the (original) cubature tree at time t_k by

$$\mathcal{C}_k \equiv \bigcup_{i_1,\ldots,i_k=1}^{c_d^m} \left\{\Xi_{x_0,t_k}\left(\omega_{i_1} \otimes \ldots \otimes \omega_{i_k}\right)\right\}, \quad k = 1\ldots,n.$$

We denote by $\hat{\mathcal{C}}_k$ the set of remaining nodes after the TBBA is applied

$$\hat{\mathcal{C}}_k := \{x \in \mathcal{C}_k, \quad \hat{\lambda}_x > 0\},$$

where $\hat{\lambda}_x$ is the random weight computed by **Algorithm 1**. In other words, $\hat{\mathcal{C}}_k$ is the set of all nodes to which the TBBA assigns a positive weight. Finally, we shall also use the notation

$$\hat{\mathcal{C}}_k^x := \hat{\mathcal{C}}_k \bigcap \{\text{ children of x}\},\ x \in \hat{\mathcal{C}}_{k-1}, \quad k = 1, \ldots, n.$$

We collect in the following Lemma some properties of the random weights constructed via the **Algorithm 1**. In particular, the algorithm produces an unbiased estimator of the pure cubature measure and the random weights are sampled with minimal variance.

Lemma 52. *For any point $x \in \bigcup_{i=1}^n \hat{\mathcal{C}}_i$, algorithm 1 produces a random weight $\hat{\lambda}_x$, that is distributed according to (104), i.e.,*

$$\hat{\lambda}_x = \begin{cases} \frac{\lfloor N\lambda_x \rfloor}{N} & \text{with probability } 1 - \{N\lambda_x\} \\ \frac{\lfloor N\lambda_x \rfloor + 1}{N} & \text{with probability } \{N\lambda_x\} \end{cases}, \tag{105}$$

where λ_x is the original cubature weight. Moreover

$$\tilde{\mathbb{E}}\left[\hat{\lambda}_x\right] = \lambda_x, \quad \tilde{\mathbb{E}}\left[\left(\hat{\lambda}_x - \lambda_x\right)^2\right] = \frac{\{N\lambda_x\}\left(1 - \{N\lambda_x\}\right)}{N^2}.$$

Finally, the random weights that correspond to different leaves are negatively correlated, i.e.

$$\tilde{\mathbb{E}}\left[\left(\hat{\lambda}_x - \lambda_x\right)\left(\hat{\lambda}_y - \lambda_y\right)\right] \leq 0, \quad x \neq y, \quad x, y \in \hat{\mathcal{C}}_i,\ i = 1, \ldots, n.$$

A proof of the previous Lemma can be found in the appendix of [15].

Theorem 53. *Let $\pi := \{0 = t_0 < t_1 < \ldots < t_n = T\}$ be a partition of the time interval $[0, T]$ on which we use a cubature formula of degree m to construct cubature measures along the partition π, $\{\mathbb{Q}_{h_i}^m\}_{i=1}^n$, $h_i = t_i - t_{i-1}$. Let $N \in \mathbb{N}_+$ be a given parameter which we use to define the cubature+TBBA measures $\{\tilde{\mathbb{Q}}_{h_i}^m\}_{i=1}^n$ supported on an additional probability space $\left(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}\right)$. Then for any function $\phi \in C_b^1\left(\mathbb{R}^d\right)$ we have*

$$\tilde{\mathbb{E}}\left[\left|P_T\phi - \tilde{\mathbb{Q}}_T^m\phi\right|^2\right]^{1/2} \leq C\left(\frac{1}{n^{\frac{m-1}{2}}} + \frac{n}{\sqrt{N}}\right).$$

Proof. The first term in the control of the error is explained via Corollary 47 as the error between the diffusion semigroup operator and the cubature measure.

$$\tilde{\mathbb{E}}\left[\left|P_T\phi - \tilde{\mathbb{Q}}_T^m\phi\right|^2\right]^{1/2} \le 2\tilde{\mathbb{E}}\left[\mid P_T\phi - \mathbb{Q}_T^m\phi\mid^2\right]^{1/2} + 2\tilde{\mathbb{E}}\left[\left|\mathbb{Q}_T^m\phi - \tilde{\mathbb{Q}}_T^m\phi\right|^2\right]^{1/2}$$

and hence we can focus on the second term. We proceed with the usual telescopic sum expansion

$$\tilde{\mathbb{Q}}_T^m\phi - \mathbb{Q}_T^m\phi = \sum_{i=1}^{n-1}\tilde{\mathbb{Q}}_{h_1}^m\dots\tilde{\mathbb{Q}}_{h_i}^m\mathbb{Q}_{h_{i+1}}^m\dots\mathbb{Q}_{h_n}^m\phi - \tilde{\mathbb{Q}}_{h_1}^m\dots\tilde{\mathbb{Q}}_{h_{i+1}}^m\mathbb{Q}_{h_{i+2}}^m\dots\mathbb{Q}_{h_n}^m\phi$$

From the Markov property of the cubature method and TBBA algorithm we understand that taking expectations under the family $\{\tilde{\mathbb{Q}}_{h_i}^m\}$ composes in the obvious manner, i.e.

$$\tilde{\mathbb{Q}}_{h_1}^m\dots\tilde{\mathbb{Q}}_{h_i}^m\phi = \sum_{x_1\in\hat{\mathcal{C}}_1}\hat{\lambda}_{x_1}\sum_{x_2\in\mathcal{C}_2^{x_1}}\frac{\hat{\lambda}_{x_2}}{\hat{\lambda}_{x_1}}\dots\sum_{x_i\in\mathcal{C}_i^{x_{i-1}}}\frac{\hat{\lambda}_{x_i}}{\hat{\lambda}_{x_{i-1}}}\phi(x_i) = \sum_{x_i\in\hat{\mathcal{C}}_i}\hat{\lambda}_{x_i}\phi(x_i)$$

In this way, we see that every term in the telescopic sum, may be written as

$$\begin{aligned}
&\tilde{\mathbb{Q}}_{h_1}^m\dots\tilde{\mathbb{Q}}_{h_i}^m\mathbb{Q}_{h_{i+1}}^m\dots\mathbb{Q}_{h_n}^m\phi - \tilde{\mathbb{Q}}_{h_1}^m\dots\tilde{\mathbb{Q}}_{h_{i+1}}^m\mathbb{Q}_{h_{i+2}}^m\dots\mathbb{Q}_{h_n}^m\phi\\
&= \tilde{\mathbb{Q}}_{h_1}^m\dots\tilde{\mathbb{Q}}_{h_i}^m\left(\tilde{\mathbb{Q}}_{h_{i+1}}^m - \mathbb{Q}_{h_{i+1}}^m\right)\mathbb{Q}_{T-t_{i+1}}^m\phi\\
&= \sum_{x_i\in\hat{\mathcal{C}}_i}\hat{\lambda}_{x_i}\sum_{x_{i+1}\in\hat{\mathcal{C}}_{i+1}^{x_i}}\left(\frac{\hat{\lambda}_{x_{i+1}}}{\hat{\lambda}_{x_i}} - \frac{\lambda_{x_{i+1}}}{\lambda_{x_i}}\right)\mathbb{Q}_{T-t_{i+1}}^m\phi(x_{i+1})
\end{aligned}$$

Next, by using the identity

$$\frac{\hat{a}}{\hat{b}} - \frac{a}{b} = \frac{\hat{a}-a}{\hat{b}} + \frac{a}{b\hat{b}}(b-\hat{b}),$$

we can re-write the above generic term of the telescopic sum as

$$\begin{aligned}
&\sum_{x_{i+1}\in\hat{\mathcal{C}}_{i+1}}(\hat{\lambda}_{x_{i+1}} - \lambda_{x_{i+1}})\mathbb{Q}_{T-t_{i+1}}^m\phi(x_{i+1})\\
&\quad + \sum_{x_i\in\hat{\mathcal{C}}_i}(\hat{\lambda}_{x_i} - \lambda_{x_i})\sum_{x_{i+1}\in\hat{\mathcal{C}}_{i+1}^{x_i}}\frac{\lambda_{x_{i+1}}}{\lambda_{x_i}}\mathbb{Q}_{T-t_{i+1}}^m\phi(x_{i+1})
\end{aligned}$$

We can then take squares in the above, and using the fact that the random TBBA weights are negatively correlated as well as the expression on the variance of the error (see Lemma 52) and the fact that $\phi(X_T)$ has a finite second moment under the pure cubature measure (this is quite trivial to show), we have that

$$\tilde{\mathbb{E}}\left[\left|\tilde{\mathbb{Q}}^m_{h_1}\dots\tilde{\mathbb{Q}}^m_{h_i}\mathbb{Q}^m_{h_{i+1}}\dots\mathbb{Q}^m_{h_n}\phi-\tilde{\mathbb{Q}}^m_{h_1}\dots\tilde{\mathbb{Q}}^m_{h_{i+1}}\mathbb{Q}^m_{h_{i+2}}\dots\mathbb{Q}^m_{h_n}\phi\right|^2\right]^{1/2}\leq C/\sqrt{N}$$

and the result follows. □

3.6 Numerical Simulations Under the Heston Model

In this section we present the application of the cubature and TBBA method for the approximation of a call option on a Heston model price process. This is a favorable set up since the Heston model is well known for capturing the volatility dynamics in various asset classes and hence has received a lot of attention by practioners and academics. On the other hand, pricing call options under the Heston model admits semi closed solutions (see [23]) against which we can compare the efficiency of our method. Let us recall briefly the Heston model. In the following, we consider $X=\{(X^1_t(x),X^1_t(x)),t\geq 0,x\in\mathbb{R}^2\}$ satisfying

$$X^1_t(x)=x^1+\int_0^t rX^1_s(x)\,ds+\int_0^t X^1_s(x)\sqrt{X^2_s(x)}dB^1_s \tag{106}$$

$$X^2_t(x)=x^2+\int_0^t \alpha\left(\theta-X^2_s(x)\right)ds+\int_0^t \beta\sqrt{X^2_s(x)}\left(\rho dB^1_s+\sqrt{1-\rho^2}dB^2_s\right), \tag{107}$$

where $x^1,x^2>0$ are positive values, $\left(B^1,B^2\right)$ is a standard two dimensional Brownian motion and α,θ,μ are positive constants satisfying

$$2\alpha\theta-\beta^2>0$$

to ensure the existence and uniqueness of a solution of the SDE (107) which never hits 0. This is a two factor stochastic volatility model with ρ being the correlation between the two random noises, $|\rho|\leq 1$. The payoff of a vanilla call option with maturity $T>0$ and strike price $K>0$ is given

$$C(T,K)=E\left[\left(X^1_T-K\right)_+\right].$$

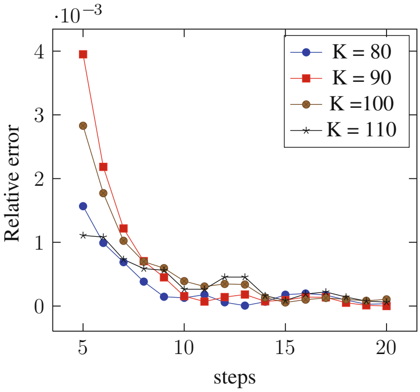
In the numerical example below we consider the following values for the various parameters

X_0	r	α	θ	β	ρ	T
100.	0.05	1	0.4	0.2	0	1.

We price a call option with strikes varying between $K = 80, 90, 100$ and 110. We keep the (maximal) number of particles that the TBBA allows to survive fixed at $N = 200000$. For every strike price and any number of steps, we launch the algorithm 10 times and average out the results. In other words if $\hat{c}(K, X_0, N, n)$ denotes the value computed by our algorithm when N particles and n steps for the discretization of time are used for a call option with strike K and spot at X_0 at time 0, we report on

$$\sum_{i=1}^{10} \left| \left(\hat{c}_i(K, X_0, N, n) - c(K, X_0) \right) / c(K, X_0) \right|,$$

where $\hat{c}_i(K, X_0, N, n)$ is the result of the i-th run of our algorithm and $c(K, X_0)$ is the value of the call option in the Heston model. We plot the results for the various strikes and varying number of steps in the figure below:



In all different strikes, we see that the algorithm behaves satisfactorily. It achieves an accuracy between 10^{-3} and 10^{-4} in the relevant error when 15 or more steps are used to discretize time. Recall that an at-the-money call is in general more difficult to approximate than in or out of the money calls, as its derivatives oscillate more as we approach maturity. However our algorithm does not seem affected by this.

4 Backward SDEs

In this section we present a brief overview to the theory of backward stochastic differential equations (BSDEs). These objects have received considerable attention over the last 20 years as they are intrinsically connected with three areas of stochastic analysis where research is very active: Non linear pricing, stochastic control and probabilistic representations of (viscosity) solutions of nonlinear PDEs and associated numerical methods. We will not go deep into the subject of BSDEs. Rather, we present some key points, mostly for ready reference, as in the following section we discuss an algorithm designed for the numerical solution of a BSDE (equivalently of a non linear PDE) based on the cubature and TBBA method.

4.1 The General Framework for Backward SDEs

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be complete probability space endowed with a filtration that satisfies the usual conditions $\{\mathcal{F}_t\}_{t\geq 0}$. Let W be a d-dimensional, $\{\mathcal{F}_t\}$-adapted Brownian motion and let $(X, Y, Z) = \{(X_t, Y_t, Z_t), t \in [0, T]\}$ be the solution of the (decoupled) system, called a Forward–Backward SDE:

$$X_t = X_0 + \int_0^t V_0(X_s)ds + \sum_{i=1}^d \int_0^t V_i(X_s) \circ dW_s^i, \quad \text{forward component} \tag{108}$$

$$Y_t = \Phi(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)ds - \sum_{i=1}^d \int_t^T Z_s^i dW_s^i, \quad \text{backward component.} \tag{109}$$

In (108)+(109), the process X is d-dimensional, Y is one dimensional and Z is d-dimensional. The coefficients $V_i : \mathbb{R}^d \to \mathbb{R}$ are smooth vector fields with $V_i \in \mathcal{C}_b^\infty(\mathbb{R}^d)$, $i = 0, 1, \ldots, d$. The stochastic integrals in (108) are *Stratonovitch* type integral. The quantity $\Phi(X_T)$ is called *the final condition*, whilst $f : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is a Lipschitz function called "the driver".

Initially, existence and uniqueness for solution of equations of the form (108)+(109) was shown under a general Lipschitz assumption on the coefficients. This has since been relaxed considerably but here, we will only consider systems whose coefficients satisfy at least the following Lipschitz assumptions:

(A) The coefficients of the forward SDE $V_i : \mathbb{R}^d \to \mathbb{R}^d$, $i = 0, 1, \ldots, d$ and the driver f are globally Lipschitz with respect to the spatial variables. Further on, the driver is $1/2$-Hölder continuous with respect to t.

(B) The coefficients of the forward SDE $V_i : \mathbb{R}^d \to \mathbb{R}^d$, $i = 0, 1, \ldots, d$ have all entries belonging to $\mathcal{C}_b^m(\mathbb{R}^d)$, the space of bounded m times differentiable functions with all partial derivatives bounded. The value of the parameter m shall be determined further on.

(C) The final condition Φ is Lipschitz continuous.

We denote by K the bound associated with all assumptions **(A), (B), (C)**.

Theorem 54 (Pardoux and Peng (1990)). *Under assumptions (A),(C) there exists a unique $\mathcal{F}_t$-adapted solution (X, Y, Z) of the system (108) + (109).*

Let us consider the simplest form of BSDE

$$Y_t = \Phi(X_T) + \int_t^T f(s, X_s)ds - \sum_{i=1}^{d} \int_t^T Z_s^i dW_s^i. \tag{110}$$

By the Martingale Representation Theorem, for

$$\xi \equiv \Phi(X_T) + \int_0^T f(s, X_s)ds$$

there exists a unique $\mathcal{F}_t$-adapted process Z such that the martingale $M = \{M_t, t \in [0, T]\}$ defined as $M_t = \mathbb{E}[\xi|\mathcal{F}_t], t \in [0, T]$ has the following representation

$$M_t = \mathbb{E}[\xi] + \sum_{i=1}^{d} \int_0^t Z_s^i dW_s^i$$

Define $Y = (Y_t, t \in [0, T])$ to be the $\mathcal{F}_t$-adapted process

$$Y_t \equiv M_t - \int_0^t f(s, X_s)ds. \tag{111}$$

It is the straightforward to show that the pair (Y, Z) are the unique solution of (110). Indeed from (111) we deduce that

$$\begin{aligned} Y_T &= M_T - \int_0^T f(s, X_s)ds \\ &= \mathbb{E}\left[\overbrace{\Phi(X_T) + \int_0^T f(s, X_s)ds}^{\xi} |\mathcal{F}_T\right] - \int_0^T f(s, X_s)ds = \Phi(X_T), \end{aligned}$$

hence

$$\begin{aligned} Y_t - \overbrace{\Phi(X_T)}^{Y_T} &= \overbrace{(M_t - \int_0^t f(s, X_s)ds)}^{Y_t} - \overbrace{(M_T - \int_0^T f(s, X_s)ds)}^{Y_T} \\ &= -\sum_{i=1}^{d} \int_t^T Z_s^i dW_s + \int_t^T f(s, X_s)ds. \end{aligned}$$

Thus (Y, Z) satisfies (110). The martingale representation theorem, as applied above, lies at the heart of the Picard iteration style argument for the proof of Theorem 54.

A celebrated result in the theory of BSDEs, is a theorem due to Pardoux and Peng that links their solution to the (viscosity) solution of semilinear PDEs. This is achieved by a Feynman–Kac type representation and it has since been extended to obstacle problems [16], quasi-linear PDEs [38] and indeed recently to fully non-linear PDEs [9,53]. Here we restrict ourselves to the simplest possible case, namely the one corresponding to semilinear PDEs (equivalently decoupled FBSDEs). Let us consider the following semilinear PDE,

$$\begin{cases} (\partial_t + L)u = -f\left(t, x, u, (\nabla u V)(x)\right), & t \in [0,T),\ x \in \mathbb{R}^d \\ u(T,x) = \Phi(x), \quad x \in \mathbb{R}^d \end{cases}. \tag{112}$$

In (112), L is the second order differential operator

$$Lv = V_0 + \frac{1}{2}\sum_{i=1}^{d} V_i^2, \tag{113}$$

V is the matrix valued function with columns $V_i(x)$, $i = 1, \ldots, d$, $V^*(x)$ is the transpose of $V(x)$ and u has final condition $u(T,x) = \Phi(x)$.

Theorem 55 (Pardoux and Peng 1992). *Under additional smoothness assumptions on its coefficients, the unique solution of the Cauchy problem* (112) *admits the following* ***Feynman–Kac representation***

$$u(t,x) = Y_t^{t,x} = \mathbb{E}\left[\Phi(X_T^{t,x}) + \int_t^T f(s, X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})\right], \tag{114}$$

where $(X^{t,x}, Y^{t,x}, Z^{t,x})$ *is the stochastic flow associated FBSDE* (108) + (109)*, i.e.,*

$$X_s^{t,x} = x + \int_t^s V_0(X_u^{t,x})du + \sum_{i=1}^{d}\int_t^s V_i(X_u^{t,x}) \circ dW_u^i, \quad s \in [t,T], \tag{115}$$

$$Y_s^{t,x} = \Phi(X_T^{t,x}) + \int_s^T f(u, X_u^{t,x}, Y_u^{t,x}, Z_u^{t,x})du - \sum_{i=1}^{d}\int_s^T (Z_u^{t,x})^i dW_u^i. \tag{116}$$

Moreover $Z_s^{t,x} = \nabla u(s, X_s^{t,x}) V(X_s^{t,x})$ *for* $s \in [t,T]$.

The representation for Y is true even if u exists only in the viscosity sense. Given such a viscosity solution, Ma and Zhang [37] show that the representation for Z holds as well, provided that the driver and the terminal condition are continuously differentiable. Numerical algorithms that are designed for the approximation of solutions of BSDEs are, in effect, probabilistic methods for solving semilinear PDEs.

4.2 Discretization of Backward SDEs

The Feynman–Kac representation (114) is instructive as it implies that the solution to a BSDE can be expressed as an integral against the law of the forward diffusion. Indeed taking expectations in a BSDE and substituting for Z, we have,

$$Y_t^{t,x} = \mathbb{E}\left[\Phi(X_T^{t,x}) + \int_t^T f(s, X_s^{t,x}, u(s, X_s^{t,x}), \nabla u(s, X_s^{t,x}) V(X_s^{t,x}))ds\right]$$

As $Y_s^{t,x}$ is adapted to $\{\mathcal{F}_s^{t,x}\}_{t \le s \le T}$, the filtration associated to $X^{t,x}$, it is almost surely deterministic and there exists a functional $\Lambda_t : C\,[t, T] \to \mathbb{R}$ such that

$$Y_t^{t,x} = \mathbb{E}[\Lambda_t\left(X_\cdot^{t,x}\right)],$$

where $C\,[t, T]$ is the space of continuous functions $\alpha : [t, T] \to \mathbb{R}^d$ and $X_\cdot^{t,x}$ is the path valued random map

$$\omega \in \Omega \longrightarrow \left\{X_s^{t,x}(\omega), s \in [t, T]\right\}.$$

Obviously the functional Λ_t is only implicitly defined by the dynamics of the backward equation. Hence, a numerical method for the approximation of $Y_t^{t,x}$ should rely on two components : A method that substitutes Λ_t with an explicitly computable functional and an approximation of the law of the forward diffusion to integrate against.

We approximate Λ_t in the following manner: Consider a partition $\pi = \{0 = t_0 < \ldots < t_{n-1} < t_n = T\}$ of $[0, T]$ with $h_i := t_i - t_{i-1}, i = 1, \ldots, n$. Assume that we know the values of Y, Z at time t_{i+1}, Y_{i+1}, Z_{i+1}. Consider the BSDE between times t_i, t_{i+1}

$$Y_{t_i} = Y_{t_{i+1}} + \int_{t_i}^{t_{i+1}} f(X_s, Y_s, Z_s)ds - \int_{t_i}^{t_{i+1}} Z_s \cdot dB_s$$

and discretize the Riemann integral using the left hand side point (the so called implicit Euler scheme of Bouchard–Touzi [3]), thus leading to an implicit equation for Y_{t_i} and the stochastic part in the usual way, to obtain

$$Y_{t_i} \simeq Y_{t_{i+1}} + h_{i+1} f(X_{t_i}, Y_{t_i}, Z_{t_i}) - Z_{t_i} \cdot \Delta W_{i+1}. \tag{117}$$

By conditioning (117) with respect to $\mathcal{F}_{t_i}$ we obtain a first order approximation for Y_{t_i}

$$Y_{t_i} \simeq \mathbb{E}\left[\, Y_{t_{i+1}} \middle| \mathcal{F}_{t_i} \right] + h_{i+1} f(X_{t_i}, Y_{t_i}, Z_{t_i}), \tag{118}$$

but for the presence of Z_{t_i}. To treat the Z_{t_i}, we can multiply both sides of (117) by $\Delta W^l_{i+1},\ l = 1,\ldots,d$ and condition with respect to $\mathcal{F}_{t_i}$, to obtain

$$Z^l_{t_i} \simeq \mathbb{E}\left[Y_{t_{i+1}} \frac{\Delta W^l_{i+1}}{h_{i+1}} \middle| \mathcal{F}_{t_i} \right], \quad l = 1,\ldots,d. \tag{119}$$

Inspired by (118), (119) we define the family $R_i : C_{Lip}\left(\mathbb{R}^d\right) \to C_{Lip}\left(\mathbb{R}^d\right)$ $i = 0, 1,\ldots,n-1$ of operators defined on the set of Lipschitz continuous functions $C_{Lip}\left(\mathbb{R}^d\right)$:

$$\begin{aligned} R_i g\left(x\right) = \mathbb{E}\left[g\left(X^{t_i,x}_{t_{i+1}}\right)\right] \\ +h_{i+1} f\left(t_i, x, R_i g\left(x\right), \frac{1}{h_{i+1}}\mathbb{E}\left[g\left(X^{t_i,x}_{t_{i+1}}\right)\left(W_{t_{i+1}} - W_{t_i}\right)\right]\right). \end{aligned} \tag{120}$$

The iteration of this family of operators $R_{i:n-1} := R_i \ldots R_{n-1}$ gives rise to an explicitly defined functional $\Lambda^{\pi}_{t_i}$, $i = 0,\ldots,n-1$,

$$\mathbb{E}[\Lambda^{\pi}_{t_i}(X^{t_i,x}_{\cdot})] = R_{i:n-1}\Phi(x).$$

The operator $R_{i:n-1}$ applied to the boundary data $\Phi(\cdot)$ and evaluated at $x = X^{t_i,x}_{t_i}$, can be viewed as a discretized version (corresponding to the partition π) of $Y^{t_i,x}_{t_i}$. In fact the above discretization is merely the Euler scheme for BSDEs (it should be clear that the Riemann integral is discretized in an Euler fashion). Relative to this, we have the following convergence result due to Bouchard and Touzi [3] and independently to Zhang [60]. This results were further refined by Gobet and Labart [20], where an error expansion, under additional smoothness assumptions, was obtained.

Theorem 56 (Bouchard and Touzi, Zhang). *Set $Y^{\pi,x}_0 = R_{0:n-1}\Phi(x)$. Under assumptions **(A), (C)***

$$|Y^{\pi,x}_0 - Y^x_0| \le C\sqrt{\|\pi\|},$$

where $\|\pi\|$ is the size of the partition mesh.

Remark 57. Originally, the proof of the convergence of the Euler scheme for BSDEs required an ellipticity assumption on the diffusion matrix of the forward component. However, the proof can be redone without this at least in the Markovian case. All one needs to show is that the value functions describing Y_t, Z_t as functions of time and X_t are smooth enough for the relevant stochastic Taylor expansions to be applied.

Theorem 58 (Gobet and Labart). *Let assumption **(B)** hold true with $m \ge 3$ and assume also that the partial derivatives of the driver with respect to space are*

Hölder continuous. Assume also that the terminal condition is twice continuously differentiable with bounded partial derivatives. Then

$$|Y_0^{\pi,x} - Y_0^x| \le C\|\pi\|.$$

To obtain a fully implementable scheme, a method of computation for the expectations appearing in (120) involved needs to be introduced. We will present next an algorithm that uses the cubature method to approximate the law of the forward diffusion and the TBBA algorithm to control the computational effort. Both of these when combined with the Euler style discretization (118), (119) provide a fully implementable scheme for BSDEs.

4.3 Cubature on BSDEs

We will use a cubature formula of degree m, supported on paths $\omega_1, \ldots, \omega_{c_d^m} : [0,1] \to \mathbb{R}^d$. We also fix throughout a parameter N to be used in the application of the TBBA. Using this cubature formula and TBBA we build (see Sect. 3.5) the sequence of explicit measures $\{\tilde{\mathbb{Q}}_{t_i}^m\}_{i=1}^n$. Substituting integration against the Wiener measure, with integration against the explicit measures $\{\tilde{\mathbb{Q}}_{t_i}^m\}_{i=1}^n$ in (120), we can define the following family of operators:

$$\begin{aligned}\hat{R}_i g\,(x) = \mathbb{E}_{\tilde{\mathbb{Q}}_{t_i}^m}[g(X_{t_{i+1}}^{t_i,x})] \\ +h_{i+1} f\left(t_i, x, \hat{R}_i g\,(x)\,, \frac{1}{h_{i+1}}\mathbb{E}_{\tilde{\mathbb{Q}}_{t_i}^m}\left[g(X_{t_{i+1}}^{t_i,x})(W_{t_{i+1}} - W_{t_i})\right]\right)\end{aligned} \tag{121}$$

where $g : \mathbb{R}^d \to \mathbb{R}$. Computations of the involved expectations in (121) are done in the obvious way, namely we work our way backwards along the cubature+TBBA tree.

Recall from the Sect. 3.5 the sets $\hat{\mathcal{C}}_i$, $i = 1, \ldots, n$ and for every $x \in \mathcal{C}_i$ the subset of its children $\hat{\mathcal{C}}^x$. Given that we are standing at depth i (equivalently, at time t_i), we need to evaluate the operator $\hat{R}_i$, when applied to $\hat{R}_{i+1:n-1}\Phi$, at all points $x \in \hat{\mathcal{C}}_i$. We have

$$\begin{aligned}&\mathbb{E}_{\tilde{\mathbb{Q}}_{t_i}^m}[g(X_{t_{i+1}}^{t_i,x})] \equiv \mathbb{E}_{\tilde{\mathbb{Q}}_{t_i}^m}\left[g(X_{t_{i+1}})|X_{t_i} = x\right] := \sum_{\bar{x}\in\mathcal{C}_{i+1}^x} \frac{\hat{\lambda}_{\bar{x}}}{\lambda_{\bar{x}}} g(\bar{x}), \quad x \in \hat{\mathcal{C}}_i \\ &\mathbb{E}_{\tilde{\mathbb{Q}}_{t_i}^m}\left[g(X_{t_{i+1}})\Delta W_{i+1}^l|X_{t_i} = x\right] := \sum_{\bar{x}\in\mathcal{C}_{i+1}^x} \frac{\hat{\lambda}_{\bar{x}}}{\lambda_{\bar{x}}} g(\bar{x})\delta\omega_{h_{i+1},\bar{x}}^l, \quad x \in \hat{\mathcal{C}}_i,\, l = 1, \ldots, d,\end{aligned} \tag{122}$$

where $\omega^l_{h_k,\bar{x}}$ is the l-th coordinate of the path $\omega_{h_{i+1},\bar{x}}$ in the cubature formula, that was used in the ODE that lead to the point $\bar{x} \in \hat{\mathcal{C}}^x_{i+1}$, scaled over the time interval $[t_i, t_{i+1})$. It should then be clear how one computes $\hat{R}_{i:n-1}\Phi(x)$ for $x \in \hat{\mathcal{C}}_i$.

Estimating the global error $\hat{R}_{0:n-1}\Phi(x_0) - Y_0^{0,x_0}$ requires standard numerical analysis arguments as well as some knowledge of the behavior of the solution to PDE (112). As estimating the errors of cubature formulas is done with the help of Taylor expansions, the derivatives of the involved functions need to be estimated. In other words, we need gradient bounds, in the spirit of Sect. 2 but here for the semilinear PDEs. For elliptic PDEs, such bounds are of course well known for a long time. But when one wishes to step into the realm of degenerate PDEs/SDEs the subject becomes quite technical and difficult. Recently, these issues were addressed in Crisan and Delarue [12] and we are able to report here on this gradient bounds for semi linear PDEs without discussing its proof.

Theorem 59 (Crisan and Delarue [12]). *Let assumption **(B)** hold true and consider an $m \geq 3$. Assume further that the vector fields $\{V_i : i = 0, \ldots, d\}$ satisfy the UFG condition. Assume also $\Phi \in C_b^m(\mathbb{R}^d)$. Define $u(t,x) = Y_t^{t,x}$. Then u is differentiable in all the direction that appear in (112). Moreover, for any multi-index $\alpha \in \mathcal{A}^1_m$, there exist increasing function $c_\alpha, \bar{c}_\alpha : [0,\infty) \to [0,\infty)$ such that for any $\Phi \in C_b^m(\mathbb{R}^d)$, we have*

$$\|V_\alpha u(t,\cdot)\|_\infty \leq c_\alpha \left(\sum_{\alpha \in \mathcal{A}_m} \|V_\alpha \Phi\| \right), \tag{123}$$

$$\|V_\alpha u(t,\cdot)\|_\infty \leq \frac{\bar{c}_\alpha(\|\Phi\|_{Lip})}{(T-t)^{(\|\alpha\|-1)/2}}, \quad t \in [0,T), \tag{124}$$

In analyzing the error we split it into two parts: The error between the solution of the BSDE and the Euler scheme and the error between the Euler scheme and its cubature and TBBA realization. The first part of the error is treated by Theorem 56. The second part of the error is split to the error due to cubature method and the error due to TBBA. Let us define the family of intermediate operators

$$\begin{aligned}\bar{R}_i g(x) &= \mathbb{E}_{\mathbb{Q}^m_{t_i}}[g(X^{t_i,x}_{t_{i+1}})] \\ &\quad + h_{i+1} f\left(t_i, x, \bar{R}_i g(x), \frac{1}{h_{i+1}} \mathbb{E}_{\mathbb{Q}^m_{t_i}}\left[g(X^{t_i,x}_{t_{i+1}})(W_{t_{i+1}} - W_{t_i}) \right] \right),\end{aligned} \tag{125}$$

which is merely the equivalent definition to the family $\{\hat{R}_i\}_{1\leq i\leq n}$ but using the pure cubature measures. It is obvious that in quantifying the error between $R_{i:n-1}\Phi$, $\bar{R}_{i:n-1}\Phi$, $i = 0, \ldots, n-1$ we need to quantify the errors

$$\mathbb{E}_{\mathbb{Q}^m_{t_i}}[g(X^{t_i,x}_{t_{i+1}})] - \mathbb{E}[g(X^{t_i,x}_{t_{i+1}})],\ \mathbb{E}_{\mathbb{Q}^m_{t_i}}[g(X^{t_i,x}_{t_{i+1}})\Delta W_{i+1}] - \mathbb{E}[g(X^{t_i,x}_{t_{i+1}})\Delta W_{i+1}],\ i = 0, \ldots, n-1.$$

We have already seen in Sect. 3 that

$$\sup_x \left|\mathbb{E}\left[g(X^{t_i,x}_{t_{i+1}})\right] - \mathbb{E}_{\mathbb{Q}^m_{t_{i+1}}}\left[g(X^{t_i,x}_{t_{i+1}}))\right]\right| \le C \sum_{j=m+1}^{m+2} h^{j/2}_{i+1} \sup_{I\in\mathcal{A}(j)\setminus\mathcal{A}(j-1)} \|V_I g\|_\infty. \tag{126}$$

For the second term, we also have

$$\begin{aligned} &\sup_x \left|\mathbb{E}\left[g(X^{t_i,x}_{t_{i+1}})\Delta W^l_{i+1}\right] - \mathbb{E}_{\mathbb{Q}^m_{h_{i+1}}}\left[g(X^{t_i,x}_{t_{i+1}})\Delta W^l_{i+1}\right]\right| \\ &\le C \sum_{j=m}^{m+2} h^{(j+1)/2}_{i+1} \sup_{I\in\mathcal{A}(j)\setminus\mathcal{A}(j-1)} \|V_I g\|_\infty. \end{aligned} \tag{127}$$

Proof of (127) Let us fix a value $l \in \{1, \ldots, d\}$. Since the function g is smooth it admits the Stratonovich–Taylor expansion. An easy application of Itô's formula, shows that the product of an iterated Stratonovich integral and a Brownian motion can be expressed as a sum of higher order iterated integrals (see for example Proposition 5.2.10 of [27]).

$$\left(\int_{\Delta^k[0,t]} \circ dW^I\right) W^l_t = \sum_{j=0}^{k} \int_{\Delta^{k+1}[0,t]} \circ dW^{(i_1,\ldots,i_j,l,i_{j+1},\ldots,i_k)},$$

where for any multi index $\alpha = (i_1, \ldots, i_k)$ we denote

$$\int_{\Delta^k[0,t]} \circ dW^\alpha := \int_{0<t_1<\ldots<t_k<t} \circ d\,W^{i_1}_{t_1} \ldots \circ dW^{i_k}_{t_k}.$$

Hence, we have that

$$\begin{aligned} g(X_t(0,x))W^l_t = &\sum_{(i_1,\ldots,i_k)\in\mathcal{A}_m} V_{(i_1,\ldots,i_k)} g(x) \sum_{j=0}^{k} \int_{\Delta^{k+1}[0,t]} \circ dW^{(i_1,\ldots,i_j,l,i_{j+1},\ldots,i_k)} \\ &+ R_m(t,x,g)W^l_t. \end{aligned}$$

Using this formula the error is

$$\begin{aligned} &\left|\mathbb{E}\left[g(X_t(0,x))W^l_t\right] - \mathbb{E}_{\mathbb{Q}^m_t}\left[g(X_t(0,x))W^l_t\right]\right| \\ &\quad\le \left|\mathbb{E}\left[R_m(t,x,g)W^l_t\right]\right| + \left|\mathbb{E}_{\mathbb{Q}^m_t}\left[R_m(t,x,g)W^l_t\right]\right| \\ &\qquad + \left|\left(\mathbb{E} - \mathbb{E}_{\mathbb{Q}^m_t}\right)\left[\sum_{(i_1,\ldots,i_k)\in\mathcal{A}_m} \sum_{j=0}^{k} V_{(i_1,\ldots,i_k)} g(x) \int_{\Delta^{k+1}[0,t]} \circ dW^{(i_1,\ldots,l,\ldots,i_k)}\right]\right|. \end{aligned} \tag{128}$$

According to estimates of Lemma 8 in [36] and (88), we have that

$$\left.\begin{array}{l}\sup_x \mathbb{E}\left[R_m(t,x,g)^2\right]^{1/2}\\ \sup_x \mathbb{E}_{\mathbb{Q}_t^m}\left[|R_{m,t,g}|^2\right]^{1/2}\end{array}\right\} \le C\sum_{j=m+1}^{m+2} t^{j/2} \sup_{I\in\mathcal{A}(j)\setminus\mathcal{A}(j-1)} \|V_\alpha g\|_\infty .$$

An application of Hölder's inequality gives us

$$\sup_x \left|\mathbb{E}\left[R_{m,t,g}W_t\right]\right| \le \sum_{j=m+1}^{m+2} t^{(j+1)/2} \sup_{\alpha\in\mathcal{A}(j)\setminus\mathcal{A}(j-1)} \|V_\alpha g\|_\infty .$$

To estimate the term $\mathbb{E}_{\mathbb{Q}_t^m}\left[R_{m,t,g}W_t\right]$ observe that

$$R_{m,t,g} = \sum_{\substack{(i_2,\ldots,i_k)\in\mathcal{A}_m\\(i_1,\ldots,i_k)\notin\mathcal{A}_m}} \int_{\Delta^k[0,t]} V_{i_1}\ldots V_{i_k} g(X_{t_1}(0,x))\circ dW_{t_1}^{i_1}\circ\ldots\circ dW_{t_k}^{i_k}.$$

So that, with $l\in\{1,\ldots,d\}$ fixed,

$$\begin{aligned}&\left|\mathbb{E}_{\mathbb{Q}_t^m}\left[R_{m,t,g}W_t^l\right]\right|\\ &\quad\le \sum_{j=1}^{N}\lambda_j \times \sum_{\substack{(i_2,\ldots,i_k)\in\mathcal{A}_m\\(i_1,\ldots,i_k)\notin\mathcal{A}_m}} \left|\int_{\Delta^k[0,t]} V_{i_1}\ldots V_{i_k} g\left(X_{t_1}(0,x)(\omega_{t_1,j})\right)\right.\\ &\qquad\left. d\omega_{t,j}^{i_1}(t_1)\ldots d\omega_{t,j}^{i_k}(t_k)\omega_{t,j}^{l}(t)\right|.\end{aligned}$$

Performing a change of variables to the paths $\omega_{t,j}$ to pass back to the paths that define the cubature formula on [0, 1] we obtain the estimate

$$\sup_x \left|\mathbb{E}_{\mathbb{Q}_t^m}\left[R_{m,t,g}W_t\right]\right| \le C\sum_{j=m+1}^{m+2} t^{(j+1)/2} \sup_{\alpha\in\mathcal{A}(j)\setminus\mathcal{A}(j-1)} \|V_\alpha g\|_\infty , \tag{129}$$

where the constant C depends on the bounds on the total variation of the paths $\omega_1,\ldots,\omega_N$. We now focus on the last term of (128).

$$\begin{aligned}&|\left(\mathbb{E}-\mathbb{E}_{\mathbb{Q}_t^m}\right)\Big[\sum_{\alpha\in\mathcal{A}(m)} V_\alpha g(x)\int_{\Delta^k[0,t]}\circ dW^\alpha\, W_t^l\Big]|\\ &=\left|\sum_{\alpha\in\mathcal{A}(m)} V_\alpha\, g(x)\left(\mathbb{E}-\mathbb{E}_{\mathbb{Q}_t^m}\right)\left[\sum_{j=0}^{k}\int_{\Delta^{k+1}[0,t]}\circ dW^{(i_1,\ldots,l,\ldots,i_k)}\right]\right|\\ &=\left|\sum_{\alpha\in\mathcal{A}(m)\setminus\mathcal{A}(m-1)} V_\alpha g(x)\left(\mathbb{E}-\mathbb{E}_{\mathbb{Q}_t^m}\right)\left[\sum_{j=0}^{k}\int_{\Delta^{k+1}[0,t]}\circ dW^{(i_1,\ldots,l,\ldots,i_k)}\right]\right|\end{aligned}$$

since the terms corresponding to $\alpha \in \mathcal{A}(m-1)$ are 0 by definition of the measure $\mathbb{Q}_t^m$.

Hence, to obtain the estimate, observe that, for any $\alpha \in \mathcal{A}(m)\backslash\mathcal{A}(m-1)$ the terms under the cubature measure satisfy

$$\left|\mathbb{E}_{\mathbb{Q}_t^m}\left[\int_{\Delta^{k+1}[0,t]} \circ dW^{(i_1,\dots,l,\dots,i_k)}\right]\right| \leq C t^{(m+1)/2}$$

since they are iterated integrals along paths of bounded variation and hence, with similar arguments to the ones we used to derive (129), we may show that they are of order $t^{(m+1)/2}$. As for the ones under the Wiener measure, they are either 0 or of order $t^{(m+1)/2}$ according to (84). The bounds on the derivatives of the vector fields complete the proof. □

We can now report on the main cubature for BSDEs error estimate

Theorem 60. *Consider a fixed $m \geq 3$ and assume that the system (115) + (116) satisfies assumption **(B)** and **(C)**. Given a partition π we consider the family of operators $\{\bar{R}_i\}_{0\leq i\leq n-1}$ along it and consider a $p > 1$. Then, there exists a constant C independent of the partition, such that*

$$\begin{aligned} |Y_0 - \bar{Y}_0^\pi| \leq C \sum_{i=0}^{n-2} \Bigg(& \sum_{j=3}^{4} h_{i+1}^{(j+1)/2} \sup_{\|I\|=j} \|V_I u(t_i,\cdot)\|_\infty \\ & + \sum_{j=m+1}^{m+2} h_{i+1}^{j/2} \sup_{\|I\|=j,j-1} \|V_I u(t_i,\cdot)\|_\infty \Bigg) \\ & + \mathbb{E}_{\mathbb{Q}_{t_{n-1}}^m}\left[\left| Y_{t_{n-1}} - \bar{R}_{n-1}\Phi\left(X_{t_{n-1}}\right)\right|^p\right]^{1/p} \end{aligned} \quad (130)$$

where $\bar{Y}_0^\pi = \bar{R}_{0:n-1}\Phi(x_0)$, $X_0 = x_0$.

The proof of the theorem requires the following lemma:

Lemma 61. *Consider two measurable functions $g_1, g_2 : \mathbb{R}^d \to \mathbb{R}$. The operators $\{\bar{R}_i\}_{i=1}^n \quad i = 0,\dots,n$ enjoy the following property*

$$|\bar{R}_i g_1 - \bar{R}_i g_2|(x) \leq \frac{1 + Ch_{i+1}}{1 - Kh_{i+1}} \mathbb{E}_{\mathbb{Q}^m}[|g_1 - g_2|^p (g(X_{t_{i+1}}^{t_i,x}))]^{1/p} \quad (131)$$

for any $p > 1$, where C is a constant which depends on the bounded variation constants of the paths ω_j, $j = 1,\dots,N$ that define cubature on the Wiener space and K is the Lipschitz constant of the driver f.

Proof. The Lipschitz property of f tells us that there exists bounded deterministic functions $\nu(x)$, $\zeta(x)$ such that

$$(1 - h_{i+1}\nu(x))(\bar{R}_i g_1(x) - \bar{R}_i g_2(x))$$
$$= \mathbb{E}_{\mathbb{Q}^m}\left[(g_1 - g_2)(X^{t_i,x}_{t_{i+1}})\right] + \zeta(x) \cdot \mathbb{E}_{\mathbb{Q}^m}\left[(g_1 - g_2)(X^{t_i,x}_{t_{i+1}})\Delta W_{i+1}\right].$$

Hence, for h_{i+1} small enough,

$$(1 - Kh_{i+1})|\bar{R}_i g_1(x) - \bar{R}_i g_2(x)|$$
$$\leq \mathbb{E}_{\mathbb{Q}^m}\left[|g_1 - g_2|\,(X^{t_i,x}_{t_{i+1}})|\Delta W_{i+1} \cdot \zeta(x) + 1|\right]$$
$$\leq \mathbb{E}_{\mathbb{Q}^m}\left[(|g_1 - g_2|^p\,(X^{t_i,x}_{t_{i+1}})\right]^{1/p} \mathbb{E}_{\mathbb{Q}^m}\left[(\Delta W_{i+1} \cdot \zeta(x) + 1)^{2k}\right]^{1/2k},$$

where $k > q/2$ and q is the conjugate of p. Observe that $\mathbb{E}_{\mathbb{Q}^m}[\Delta W_{i+1}] = \mathbb{E}[\Delta W_{i+1}] = 0$, since ΔW_{i+1} can be written as a stochastic integral of length 1. For any higher powers of the Brownian increment, it holds that

$$\mathbb{E}_{\mathbb{Q}^m}\left[\left(\Delta W^l_{i+1}\right)^r\right] \leq C h^{r/2}_{i+1},\ \forall l = 1, \ldots, d.$$

To see this, observe that for any $r \leq m$ we may express the increment $(\Delta W^l_{i+1})^r$ as a linear combination of iterated integrals of length less than m. The estimate then follows from the definition of the measure $\mathbb{Q}^m$. Hence,

$$\mathbb{E}_{\mathbb{Q}^m}\left[(\Delta W_{i+1} \cdot \zeta(x) + 1)^{2k}\right]^{1/2k} \leq (1 + C h_{i+1})$$

and this completes the proof. □

Proof of Theorem 60. To begin with, set

$$\epsilon_i = \sum_{j=3}^{4} h^{(j+1)/2}_{i+1} \sup_{\|I\|=j} \|V_I u(t_i, \cdot)\|_\infty, \quad i = 0, \ldots, n-1.$$

We expand the error as a telescopic sum

$$Y_0 - \bar{R}_{0:n-1}\Phi(x) = \sum_{i=0}^{n-1} \bar{R}_{0:i-1} Y^{0,x}_{t_i} - \bar{R}_{0:i} Y^{0,x}_{t_{i+1}}. \tag{132}$$

The size of each of the terms $\bar{R}_{0:i-1} Y^{0,x}_{t_i} - \bar{R}_{0:i} Y^{0,x}_{t_{i+1}}$ is then controlled using Lemma 61. We have, with for $p > 1$,

$$|Y_0 - \bar{Y}^\pi_0| \leq C \sum_{i=0}^{n-1} \mathbb{E}_{\mathbb{Q}^m_{t_i}}\left[\left|Y^{t_i,X_{t_i}}_{t_i} - \bar{R}_i Y^{t_i,X_{t_i}}_{t_{i+1}}\right|^p\right]^{1/p}. \tag{133}$$

Observe that for any $i \in \{0, \dots, n-1\}$ and $x \in \mathbb{R}^d$, by taking expectations on the backward part of (109), we have

$$Y_{t_i}^{t_i,x} = \mathbb{E}\left[Y_{t_{i+1}}^{t_i,x} + \int_{t_i}^{t_{i+1}} f(s, X_s^{t_i,x}, Y_s^{t_i,x}, Z_s^{t_i,x}) ds \right].$$

The above together with the definition of $\bar{R}_i$ tells us

$$Y_{t_i}^{t_i,x} - \bar{R}_i Y_{t_{i+1}}^{t_i,x} = \left(\mathbb{E} - \mathbb{E}_{\mathbb{Q}_{t_{i+1}}^m}\right)\left[Y_{t_{i+1}}^{t_i,x}\right] + \int_{t_i}^{t_{i+1}} \mathbb{E} f(s, X_s^{t_i,x}, Y_s^{t_i,x}, Z_s^{t_i,x}) ds$$
$$-h_{i+1} f\left(t_i, x, \bar{R}_i Y_{t_{i+1}}^{t_i,x}, \frac{1}{h_{i+1}} \mathbb{E}_{\mathbb{Q}_{t_{i+1}}^m}\left[\bar{R}_i Y_{t_{i+1}}^{t_i,x} \Delta W_{i+1}\right]\right)$$

We now fix a value for $i = 0, \dots, n-2$. To compare the drivers we need to add and subtract the right terms:

$$\begin{aligned} Y_{t_i}^{t_i,x} - \bar{R}_i Y_{t_{i+1}}^{t_i,x} &= \left(\mathbb{E} - \mathbb{E}_{\mathbb{Q}_{t_{i+1}}^m}\right)\left[Y_{t_{i+1}}^{t_i,x}\right] \\ &\quad + \int_{t_i}^{t_{i+1}} \mathbb{E}\left[f(s, X_s^{t_i,x}, Y_s^{t_i,x}, Z_s^{t_i,x}) - f(t_i, x, Y_{t_i}^{t_i,x}, Z_{t_i}^{t_i,x})\right] ds \\ &\quad + h_{i+1}\left(f(t_i, x, Y_{t_i}^{t_i,x}, Z_{t_i}^{t_i,x})\right. \\ &\qquad \left. - f(t_i, x, \bar{R}_i Y_{t_{i+1}}^{t_i,x}, \frac{1}{h_{i+1}} \mathbb{E}_{\mathbb{Q}_{t_{i+1}}^m}\left[Y_{t_{i+1}}^{t_i,x} \Delta W_{i+1}\right]\right) \\ &=: I_1^{t_i,x} + I_2^{t_i,x} + I_3^{t_i,x} \end{aligned} \tag{134}$$

with the obvious definition for the $I_k^{t_i,x}$'s. To estimate each of these terms the non linear Feynman–Kac formula for BSDE's plays a central role.

Since (112) has a classical solution on $[0, T) \times \mathbb{R}^d$, it holds that

$$Y_s^{t,x} = u(s, X_s^{t,x}), \qquad Z_s^{t,x} = \nabla u(s, X_s^{t,x}) V(X_s^{t,x}).$$

We can apply Itô's formula to the function $\bar{f} : (t, x) \to f(t, x, u(t, x), \nabla u(t, x) V(x))$ to control I_2,

$$|I_2^{t_i,x}| = \left| \mathbb{E}\left[\int_{t_i}^{t_{i+1}} \int_{t_i}^{s} \partial_t \bar{f}(r, X_r^{t_i,x}) + V_0 \bar{f}(r, X_r^{t_i,x}) dr ds \right.\right.$$
$$\left.\left. + \int_{t_i}^{s} \left(\frac{1}{2} \sum_{i=1}^{d} V_i^2 \bar{f}(r, X_r^{t_i,x}) dr + \sum_{i=1}^{d} V_i \bar{f}(r, X_r^{t_i,x}) dW_r^i \right) ds \right] \right| \tag{135}$$

Hence

$$\sup_{x\in\mathbb{R}^d} \left| I_2^{t_i,x} \right| \le C h_{i+1}^2 \left(\|V_0 \bar{f} + \partial_t \bar{f}\|_\infty + \max_{i,=1,\dots,d} \left\| V_i^2 \bar{f} \right\|_\infty \right) \\ \le C h_{i+1}^2 \sup_{\|I\|=3} \|V_I u\|_\infty , \tag{136}$$

where, the latter estimate follows by the chain rule. To estimate $I_3^{t_i,x}$ we use the mean value theorem, so that we can find two points $\theta_1 \in \mathbb{R},\ \theta_2 \in \mathbb{R}^d$ such that

$$I_3^{t_i,x} = h_{i+1} \Big(f_y(t_i, x, \theta_1, \theta_2)(Y_{t_i}^{t_i,x} - \bar{R}_i Y_{t_{i+1}}^{t_i,x}) \\ + f_z(t_i, x, \theta_1, \theta_2) \cdot \left(Z_{t_i}^{t_i,x} - \frac{1}{h_{i+1}} \mathbb{E}_{\mathbb{Q}_{t_{i+1}}^m} \left[Y_{t_{i+1}}^{t_i,x} \Delta W_{i+1} \right] \right) \Big)$$

$$|I_3^{t_i,x}| \le K h_{i+1} \left(\left| Y_{t_i}^{t_i,x} - \bar{R}_i Y_{t_{i+1}}^{t_i,x} \right| + \left| Z_{t_i}^{t_i,x} - \frac{1}{h_{i+1}} \mathbb{E}_{\mathbb{Q}_{t_{i+1}}^m} \left[Y_{t_{i+1}}^{t_i,x} \Delta W_{i+1} \right] \right| \right), \tag{137}$$

since the partial derivatives of f are bounded by K. As a next step observe that

$$h_{i+1} \left| Z_{t_i}^{t_i,x} - \frac{1}{h_{i+1}} \mathbb{E}_{\mathbb{Q}_{t_{i+1}}^m} \left[Y_{t_{i+1}}^{t_i,x} \Delta W_{i+1} \right] \right| \tag{138}$$

$$\le \left| h_{i+1} Z_{t_i}^{t_i,x} - \mathbb{E}\left[Y_{t_{i+1}}^{t_i,x} \Delta W_{i+1} \right] \right| + \left| \mathbb{E}\left[Y_{t_{i+1}}^{t_i,x} \Delta W_{i+1} \right] - \mathbb{E}_{\mathbb{Q}_{t_{i+1}}^m} \left[Y_{t_{i+1}}^{t_i,x} \Delta W_{i+1} \right] \right|$$

As before, $Y_{t_{i+1}}^{t_i,x} = u(t_{i+1}, X_{t_{i+1}}^{t_i,x})$ and we may apply the stochastic Taylor expansion to the latter, to treat the first term above. In particular, we do so using the hierarchical set $\mathcal{A}_2$. Let us fix an integer value $l = 1, \dots, d$ and denote by $Z_{t_i}^{t_i,x,l}$ the l-th entry of the vector $Z_{t_i}^{t_i,x}$. We then have,

$$\left| h_{i+1} Z_{t_i}^{t_i,x,l} - \mathbb{E}\left[u(t_{i+1}, X_{t_{i+1}}^{t_i,x}) \Delta W_{i+1}^l \right] \right| \\ = \left| h_{i+1} Z_{t_i}^{t_i,x,l} - \mathbb{E}\left[\left(u(t_i, x) + \sum_{i=0}^d V_i u(t_i, x) \int_{t_i}^{t_{i+1}} \circ dW_s^i \right. \right. \right. \\ \left. \left. \left. + \sum_{i,j=1}^d V_i V_j u(t_i, x) \int_{t_i}^{t_{i+1}} \int_{t_i}^{t} \circ dW_s^i \circ dW_t^j + R_2(h_{i+1}, x, u) \right) \Delta W_{i+1}^l \right] \right| . \tag{139}$$

Observe that

$$\mathbb{E}\left[\sum_{i=1}^d V_i u(t_i, x) \int_{t_i}^{t_{i+1}} \circ dW_s^i \Delta W_{i+1}^l \right] = h_{i+1} V_l u(t_i, x).$$

Moreover, according to Proposition 5.2.10 of Kloeden and Platen [27] we have that for any $k, r = 1, \ldots, d$,

$$\int_{t_i}^{t_{i+1}} \int_{t_i}^{t} \circ dW_s^k \circ dW_t^r \Delta W_{i+1}^l = J_{(k,r,l)}[1]_{t_i,t_{i+1}} + J_{(k,r,j)}[1]_{t_i,t_{i+1}} + J_{(l,k,r)}[1]_{t_i,t_{i+1}}$$

and the three terms on the right hand side will have expectation 0 according to (84). Due to the non linear Feynman- Kac formula we have for $i = 0, \ldots, n-2$ that $Z_{t_i}^{t_i,x,l} = \nabla u(t_i, x)^* \cdot V_l(x) = V_l u(t_i, x)$. Hence, (139) together with (138) and the estimate on the remainder process, give us

$$\begin{aligned} & h_{i+1} \left| Z_{t_i}^{t_i,x} - \frac{1}{h_{i+1}} \mathbb{E}_{\mathbb{Q}^m_{t_{i+1}}} \left[Y_{t_{i+1}}^{t_i,x} \Delta W_{i+1} \right] \right| && (140) \\ & \quad \le C\mathbb{E}\left[|R_2(t_i, x, u) \Delta W_{i+1}| \right] + \frac{1}{h_{i+1}} \left| \left(\mathbb{E} - \mathbb{E}_{\mathbb{Q}^m_{t_{i+1}}} \right) \left[u(t_{i+1}, X_{t_{i+1}}^{t_i,x}) \Delta W_{i+1} \right] \right| \end{aligned}$$

Equations (140) and (139) are plugged in (138) and the resulting estimate (137). The latter together with (136) and (134) gives us

$$\begin{aligned} & (1 - h_{i+1} K)\, \mathbb{E}_{\mathbb{Q}^m_{t_i}} \left[\left| Y_{t_i}^{t_i,X_{t_i}} - \bar{R}_i Y_{t_{i+1}}^{t_i,X_{t_i}} \right|^p \right]^{1/p} \\ & \quad \le \mathbb{E}_{\mathbb{Q}^m_{t_i}} \left[\left| \left(\mathbb{E} - \mathbb{E}_{\mathbb{Q}^m_{t_{i+1}}} \right) \left[Y_{t_{i+1}}^{t_i,X_{t_i}} \right] \right|^p \right]^{1/p} \\ & \qquad + \mathbb{E}_{\mathbb{Q}^m_{t_i}} \left[\left| \left(\mathbb{E} - \mathbb{E}_{\mathbb{Q}^m_{t_{i+1}}} \right) \left[Y_{t_{i+1}}^{t_i,X_{t_i}} \Delta W_{i+1} \right] \right|^p \right]^{1/p} + \epsilon_i && (141) \\ & \quad \le \epsilon_i + C \sum_{j=m+1}^{m+2} h_{i+1}^{j/2} \sup_{I \in \mathcal{A}_j \setminus \mathcal{A}_{j-1}} \|V_I g\|_\infty, \quad i = 0, \ldots, n-2. \end{aligned}$$

where we have used the estimates (126) and (127). This completes the proof. □

We have already discussed how a non even partition can compensate for the explosion in the gradient bounds, in the linear case. In view of Theorem 59, we have a similar result in the semilinear case. In the more interesting case where the terminal condition is only Lipschitz continuous, we have to appeal to the derivative bounds (124). In this case the control on the derivatives of u explodes as t approaches T. To compensate for this negative impact of the derivative bounds on the error estimate we shall use a non equidistant partition that becomes denser as we approach T.

Corollary 62. *Let* ***(A)*** *and* ***(B)*** *hold true, fix and* $m \ge 3$ *and assume further that the vector fields* $\{V_i : i = 0, \ldots, d\}$ *satisfy the UFG condition and that the final condition* Φ *is Lipschitz. We consider the family* $\{\bar{R}_i\}_{0 \le i \le n-1}$ *along the partition* π*:*

$$t_i = T\left(1-\left(1-\frac{i}{n}\right)^{\beta}\right), \quad i=0,\ldots,n, \quad \beta \geq 2.$$

Then, there exists an increasing function $c : [0,\infty) \to [0,\infty)$ independent of the partition such that

$$\left| Y_0 - \bar{R}_{0:n-1}\Phi(x_0) \right| \leq \frac{c(\|\Phi\|_{Lip})}{n}$$

Proof. Let us assume first that $\Phi \in C_b^m\left(\mathbb{R}^d\right)$. In the following, the functions $c_i : [0,\infty) \to [0,\infty)$ are all strictly increasing. Given the estimates of Theorems 130, 59, it is straightforward to see that the dominating term in our error bound is $h_i^2 \sup_{\|\alpha\|=3} \|V_\alpha u\|_\infty$. On the above partition we have, for a given multi index α with $\|\alpha\| = 3$,

$$\begin{aligned}(t_i - t_{i-1})^2 \|V_\alpha u\|_\infty &\leq c_1(\|\Phi\|_{Lip}) T^2 \left(\int_{1-\frac{i}{n}}^{1-\frac{i-1}{n}} \beta s^{\beta-1} ds\right)^2 \frac{1}{T\,(1-i/n\,)^{\beta}} \\ &\leq \frac{c_2(\|\Phi\|_{Lip})}{n^2}\end{aligned}$$

On the other hand, for the term corresponding to t_{n-1} we may argue, using the mean value theorem, that,

$$\begin{aligned}&\mathbb{E}_{\mathbb{Q}^m_{t_{n-1}}}\left[\,\left| Y_{t_{n-1}} - \bar{R}_{n-1}\Phi(X_{t_{n-1}}) | X_{t_{n-1}} \right|^p\,\right]^{1/p} \\ &\quad\leq C\sum_{l=0}^{d} \mathbb{E}_{\mathbb{Q}^m_{t_{n-1}}}\left[\,\left|\left(\mathbb{E} - \mathbb{E}_{\mathbb{Q}^m_{t_n}}\right)\left[\,\Phi(X_{t_n})\Delta W_n^l | X_{t_{n-1}}\,\right]\right|^p\,\right]^{1/p}\end{aligned}$$

from elementary properties of the Wiener and cubature measure, it is clear that

$$\left(\mathbb{E} - \mathbb{E}_{\mathbb{Q}^m_{t_n}}\right)\left[\,\Phi(X_{t_n})\Delta W_n^l | X_{t_{n-1}}\,\right] = \left(\mathbb{E} - \mathbb{E}_{\mathbb{Q}^m_{t_n}}\right)\left[\,(\,\Phi(X_{t_n}) - \Phi(X_{t_{n-1}})\,)\,\Delta W_n^l | X_{t_{n-1}}\,\right]$$

and hence, standard estimates on the increments of the forward diffusion together with the Lipschitz property of Φ, lead to

$$\mathbb{E}_{\mathbb{Q}^m_{t_{n-1}}}\left[\,\left| Y_{t_{n-1}} - \bar{R}_{n-1}\Phi(X_{t_{n-1}}) | X_{t_{n-1}} \right|^p\,\right]^{1/p} \leq \frac{c_3(\|\Phi\|_{Lip})}{n^{\beta/2}}$$

which concludes the proof for the case of smooth terminal conditions. Assume next that Φ is Lipschitz. Via a standard mollification result, one can construct a sequence of smooth functions $\{\Phi_m\}_{m\geq 0}$ that converge uniformly to Φ and such that $\|\Phi_m\|_{Lip} \leq \|\Phi\|_{Lip}$ for all $m \geq 0$. Using the continuity properties of both Y_0 and $\bar{R}_{0:n}$ as functions of the final condition, it follows that

$$\left| Y_0 - \bar{R}_{0:n-1}\Phi(x_0) \right| = \lim_{m\to\infty} \left| Y_0^m - \bar{R}_{0:n-1}\Phi^m(x_0) \right| \le \frac{c(\|\Phi\|_{Lip})}{n},$$

where Y_0^m is the solution of the BSDE corresponding to the final condition Φ_m. Crucially in the above inequality the function c is independent of m. The proof is complete. □

It remains to estimate the error $\bar{R}_{0:n-1}\Phi(x_0) - \bar{R}_{0:n-1}\Phi(x_0)$, i.e. the error due to the application of the TBBA. In this case one needs only to combine the arguments of the previous proof with the arguments that were presented in the proof of Theorem 53. Such analysis can be found in [15]. We report here on the this estimate.

Theorem 63. *Let assumptions **(A)** and **(B)** hold true and assume that Φ is Lipschitz continuous. Consider the family $\{\hat{R}_i\}_{0\le i\le n}$ defined with N particles. With the usual notation, on the iteration of operators, there exists a constant C independent of the partition, such that*

$$\tilde{\mathbb{E}}\left[\left| \bar{R}_{0:n}\Phi(x_0) - \hat{R}_{0:n}\Phi(x_0) \right|^2 \right]^{1/2} \le \frac{Cn}{\sqrt{N}}. \tag{142}$$

4.4 Numerical Simulations

In this section, we apply our numerical scheme for BSDEs in one and multidimensional problems where the involved coefficients can be smooth or non smooth. This empirical study helps us to validate the method described above.

One Dimensional Numerical Examples

Firstly, we consider the following popular non-linear example from finance, the problem of pricing with differential interest rates. In this set up, one is able to invest money in the money account at an interest rate r and borrow at an interest rate R with $R > r$. The underlying asset price evolves as a geometric Brownian motion under the objective probability measure:

$$X_t^{0,x_0} = \int_0^t \mu\, X_s^{0,x_0} ds + \int_0^t \sigma\, X_s^{0,x_0} dW_s.$$

It is shown in El Karoui et al. [17] that a self-financing trading strategy of portfolio Z and wealth process Y solves a BSDE with driver

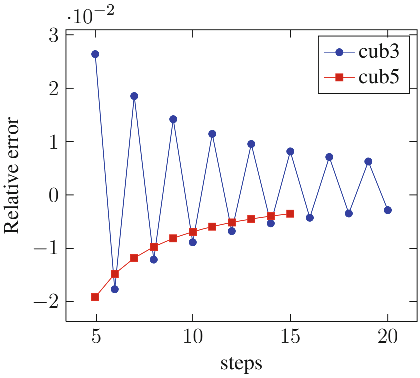
$$f(t,x,y,z) = -(ry + z\theta - (R-r)(y - z/\sigma)_-)$$

where $(x)_-$ denotes the negative part of x and $\theta = (\mu - r)/\sigma$. The problem of pricing a call option corresponds to a terminal condition of the form $\Phi(x) = (x - K)_+$.

We test our algorithm with parameters

μ	r	R	σ	X_0	K
0.03	0.06	0.08	0.2	10	10

As explained in Gobet et al. [21], in such an economy the issuer of the call option keeps borrowing money to hedge the call option so that the price of the option is the Black–Scholes with interest rate R. Hence we have the favorable set up of a non linear driver, but yet we know Y_0. Moreover we see that, even though the driver is *not differentiable* our algorithm still produces very good estimates. In the figure below, we plot the ratio of the computed value over the Black Scholes price against the number of steps.



Since this is only a one dimensional set up, we manage to achieve an accuracy of 10^{-3} with only a few time discretization steps and hence the application of TBBA to control the computational effort is not necessary here.

Since pure cubature can be applied successfully in one dimensional examples, we can next try to monitor the effect that TBBA has on the overall error. We do so in a smooth example. We consider a FBSDE system with smooth coefficients and a non linear driver for the backward part:

$$X_t^{0,x_0} = x_0 + \int_0^t \mu X_s ds + \int_0^t \sqrt{1 + X_t^2} dW_t, \quad 0 \le t \le T$$

$$Y_t^{0,x_0} = \arctan(X_T^{0,x_0}) - \int_t^T rY_s + e^{r(T-s)} (\mu - 1) X_s^{0,x_0} \left(Z_s^{0,x_0} \right)^2 ds$$
$$- \int_t^T Z_s^{0,x_0} dW_s. \tag{143}$$

It is easy to check , by means of Itô's lemma, that the solution to the above system is given by

$$Y_t^{0,x_0} = e^{-r(T-t)} \arctan(X_t^{0,x_0}), \quad Z_t^{0,x_0} = \frac{e^{-r(T-t)}}{\sqrt{1 + \left(X_t^{0,x_0}\right)^2}}.$$

We test our example with parameters

$$\frac{T \;\; \mu \quad\; r \quad x_0}{1. \; 0.02 \; 0.1 \; 2}.$$

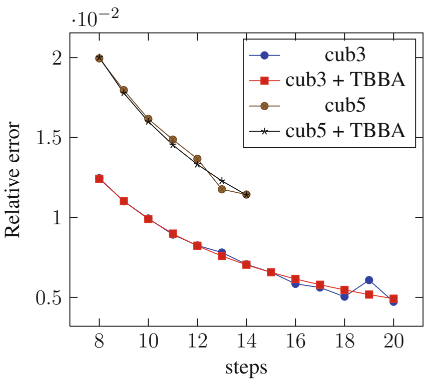
We denote by N the (maximal) number of paths that the support of the "pruned" cubature measure is allowed to hold, at every point on the partition. Let $Y_0 = e^{-rT} \arctan(X_0)$ denote the solution of (143) at time 0. We denote by $\hat{y}_0^N \equiv \hat{y}_0^N(\omega)$ the result we get at time 0 by solving the BSDE along the tree produced by one launch of the algorithm. In other words

$$\hat{y}_0^N = \hat{R}_{0:n-1}\Phi(x_0).$$

We also fix a further parameter M that counts the number of times the algorithm is launched. Obviously all the launches of the algorithm are independent of each other. Let $\hat{y}_0^{N,m}$ denote the result on the m-th run of the algorithm, $m = 1, \ldots, M$. Our approximation is then

$$\hat{y}_0^{N,M} = \frac{1}{M} \sum_{m=1}^{M} \hat{y}_0^{N,m}.$$

The figure below, monitors the error (we plot $\frac{\hat{y}_0^{N,M} - \bar{y}_0}{\bar{y}_0}$) on example (143), when using cubature of order 3, 5 with and without sampling, against the number of steps. In this case the parameters N, M are fixed as $N = 100000$, $M = 10$.

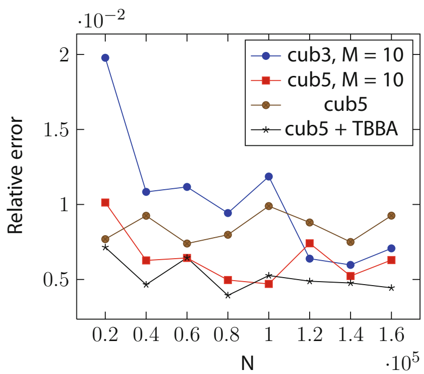


In particular we see that no accuracy is lost when applying the TBBA. Next we turn to a multidimensional example. The goal here is to show that the method produces good estimates but also to compare its performance with existing methods for solving BSDE. In a recent publication of Bouchard and Warin [4] the authors study the application of three other numerical methods (quantization, Malliavin integration by parts and regression on function basis) for BSDEs on the pricing of American/Bermudan options. In particular, we consider the case where the underlying is a Geometric Brownian motion and the payoff is a call or a put written on geometric/arithmetic averages. Here we shall consider the equivalent European pricing problem. In terms of computational complexity (on which the authors of [4] report), there is no significant difference. Indeed, the pricing of the Bermudan counterpart amounts to checking for optimal exercise on every point in the support of the underlying measure which would be negligible given the overall complexity of the algorithm.

We look at a five-dimensional example:

$$
\begin{aligned}
X_t^i &= x_0^i + \int_0^t \mu_i X_s^i ds + \sigma_i X_s^i dW_s^i, \quad i = 1, \dots, 5 \\
Y_t &= \left(\prod_{i=1}^5 X_1^i - K \right)_+ - \int_t^1 rY_s + \theta Z_s ds - \int_t^1 Z_s \cdot dW_s
\end{aligned}
\tag{144}
$$

where $\theta_i = (\mu_i - r)/\sigma_i$, $i = 1, \dots, 5$ is the market price of risk. The theoretical value for Y_0 can be produced with the usual Black Scholes methodology. Again we fix the number of steps to 10 and we do a plot the usual relative error. Of course we normalize against the Black Scholes price.



As far as the computational time is concerned, we report on the following values (the computational time is measured here in seconds)[12]:

N	40000	100000	160000
cub3(M=10)	3.8	8.6	13.2
cub5(M=10)	6.9	16.9	26.3
cub3(M=20)	7.4	17.4	26.5
cub5(M=20)	13.7	33.5	53

Comparing these performance results, in conjunction with the information on the errors, with Fig. 7(e), Fig. 8 of [4] we see that the cubature+TBBA algorithm can achieve similar accuracy in lesser time. On the other hand, we see that there is a small bias (relative error of order 0.5 %) that the algorithm does not treat with the increase in N. This bias is due to the discretization error (recall that we are normalizing against the theoretical Black Scholes value).

Appendix

In this section we provide various proofs of results left outstanding from the main body.

Proposition 20. *For any $T > 0$, $p \in [1, \infty)$, $\alpha, \beta \in \mathcal{A}(m)$ and $\gamma \in \mathcal{A}$, the following hold*

$$\sup_{t\in(0,T]} \mathbb{E}\left[t^{-\|\gamma\|/2} \left| \hat{B}_t^{\circ\gamma} \right| \right]^p < \infty, \tag{48}$$

$$\sup_{\substack{x\in\mathbb{R}^N \\ t\in(0,T]}} \mathbb{E}\left[t^{-(m+1-\|\alpha\|)/2} \left| r_{\alpha,\beta}(t,x) \right| \right]^p < \infty. \tag{49}$$

Proof. The proof is done as follows: we first show an intermediate result that holds for a general semimartingale. We then prove (48) and (49) via an inductive argument. Assume that W is a one dimensional $\mathcal{F}_t$-adapted Brownian motion and $t \to u(t,x)$, respectively $t \to u(t,x)$ are $\mathcal{F}_t$-adapted processes such that

$$\sup_{\substack{x\in\mathbb{R}^N \\ t\in[0,T]}} \mathbb{E}\left(t^{-r_u} \left| u(t,x) \right|\right)^p < \infty, \quad \text{and} \quad \sup_{\substack{x\in\mathbb{R}^N \\ t\in[0,T]}} \mathbb{E}\left(t^{-r_v} \left| v(t,x) \right|\right)^p < \infty,$$

[12]Numerical experiments were performed with single-threaded code on a Intel i7 processor at 2.8 GHz.

for some constants, $r_u, r_v \in [0, \infty)$. Next let ξ^x be the process defined as

$$\xi_t^x = \int_0^t u(s,x)dW_s + \int_0^t v(s,x)ds,$$

Then, for $p \geq 1$:

$$\begin{aligned}
\mathbb{E}\left[|\, \xi_t^x \,|^p\right] &= \mathbb{E}\left[\left| \int_0^t u(s,x)dW_s + \int_0^t v(s,x)ds \right|^p\right] \\
&\overset{(1)}{\leq} 2^{p-1}\left\{\mathbb{E}\left[\left|\int_0^t u(s,x)dW_s\right|^p\right] + \mathbb{E}\left[\left|\int_0^t v(s,x)ds\right|^p\right]\right\} \\
&\overset{(2)}{\leq} 2^{p-1}\left\{C_p\mathbb{E}\left[\left(\int_0^t |\, u(s,x)\,|^2\, ds\right)^{\frac{p}{2}}\right] + t^{p-1}\mathbb{E}\left[\int_0^t |\, v(s,x)\,|^p\, ds\right]\right\} \\
&\overset{(3)}{\leq} 2^{p-1}\left\{C_p\, t^{\frac{p}{2}-1}\mathbb{E}\left[\int_0^t |\, u(s,x)\,|^p\, ds\right] + t^{p-1}\mathbb{E}\left[\int_0^t |\, v(s,x)\,|^p\, ds\right]\right\} \\
&\overset{(4)}{\leq} 2^{p-1}\left\{C_p t^{\frac{1}{2}(p-1)}\int_0^t \mathbb{E}\left[|\, u(s,x)\,|^p\right] ds + t^{p-1}\int_0^t \mathbb{E}\left[|\, v(s,x)\,|^p\right] ds\right\},
\end{aligned}$$

where we used the following: Hölder's inequality for finite sums for (1), Burkholder's inequality, Jensen's inequality respectively, for (2), Jensen's inequality for definite integrals for (3), Fubini's theorem for (4).

Now we observe that

$$\mathbb{E}\,|\, u(s,x)\,|^p \leq \left(\sup_{\substack{x\in\mathbb{R}^N \\ s\in[0,T]}} \mathbb{E}\,[s^{-r_u}\,|\, u(s,x)\,|]^p\right)s^{pr_u}$$

$$\mathbb{E}\,|\, v(s,x)\,|^p \leq \left(\sup_{\substack{x\in\mathbb{R}^N \\ s\in[0,T]}} \mathbb{E}\,[s^{-r_v}\,|\, v(s,x)\,|]^p\right)s^{pr_v}.$$

And so,

$$\begin{aligned}
\mathbb{E}\,|\, \xi_t^x\,|^p \leq \tilde{C}_p\Bigg\{& t^{\frac{1}{2}p-1}\left(\sup_{\substack{x\in\mathbb{R}^N \\ s\in[0,T]}} \mathbb{E}\,[s^{-r_u}\,|\, u(s,x)\,|]^p\right)\left(\int_0^t s^{pr_u}ds\right) \\
&+ t^{p-1}\left(\sup_{\substack{x\in\mathbb{R}^N \\ s\in[0,T]}} \mathbb{E}\,[s^{-r_v}\,|\, v(s,x)\,|]^p\right)\left(\int_0^t s^{pr_v}ds\right)\Bigg\}
\end{aligned}$$

$$\leq \hat{C}_p \Bigg\{ t^{\frac{1}{2}p-1} t^{pr_u+1} \Bigg(\sup_{\substack{x\in\mathbb{R}^N \\ s\in[0,T]}} \mathbb{E}\left[s^{-r_u} \mid u(s,x) \mid\right]^p \Bigg)$$

$$+ t^{p-1} t^{pr_v+1} \Bigg(\sup_{\substack{x\in\mathbb{R}^N \\ s\in[0,T]}} \mathbb{E}\big(s^{-r_v} \mid v(s,x) \mid\big)^p \Bigg) \Bigg\}$$

$$\leq \hat{C}_{p,u,v} \Big\{ t^{p(r_u+\frac{1}{2})} + t^{p(r_v+1)} \Big\}.$$

That is, if we take $r_\xi = \min\{r_u + \frac{1}{2}, r_v + 1\}$, then for all $p \in [1, \infty)$,

$$\sup_{\substack{x\in\mathbb{R}^N \\ t\in(0,T]}} \mathbb{E}\left[\left(t^{-r_\xi} \mid \xi_t^x \mid\right)^p\right] < \infty. \tag{145}$$

Proof of (48): We prove by induction on $|\,\gamma\,|$. Observe that for $|\,\gamma\,| = 1$, we have that:

$$\hat{B}_t^{\circ\gamma} = \begin{cases} B_t^\gamma & \text{if } \gamma \in \{1,\ldots,d\} \\ t & \text{if } \gamma = 0 \end{cases}, \tag{146}$$

in which case, we split $|\,\gamma\,| = 1$ into $\gamma \in \{1,\ldots,d\}$ and $\gamma = 0$. For the former we apply the inductive step with $u \equiv 1$ and $v \equiv 0$. Then we may choose $0 = r_u \ll r_v$ to obtain:

$$\sup_{t\in[0,T]} \mathbb{E}\left[\left(t^{-1/2} \mid B_t^\gamma \mid\right)^p\right] < \infty.$$

In the latter case we obviously have:

$$\sup_{t\in[0,T]} \mathbb{E}\left[t^{-1} \mid B_t^\gamma \mid\right]^p < \infty.$$

We now assume that the result holds for some $k \in \mathbb{N}$, i.e. we have the following for all $\gamma \in \mathcal{A}$ satisfying $|\,\gamma\,| = k$:

$$\sup_{t\in(0,T]} \mathbb{E}\left[t^{-\|\gamma\|/2} \left| \hat{B}_t^{\circ\gamma} \right|\right]^p. \tag{147}$$

Observe, that for $i \in \{1,\ldots,d\}$

$$\hat{B}_t^{\circ(\gamma*i)} = \int_0^t \hat{B}_s^{\circ\gamma} \circ dB_s^i \tag{148}$$

$$= \int_0^t \hat{B}_s^{\circ\gamma} dB_s^i + \frac{1}{2}\left\langle \hat{B}^{\circ\gamma}, B^i \right\rangle_t, \tag{149}$$

and noting that

$$\hat{B}^{\gamma}_t = \int_0^t \hat{B}^{\circ\gamma'}_s dB^{\gamma_k}_s + \frac{1}{2}\left\langle \hat{B}^{\circ\gamma'}, B^{\gamma_k}\right\rangle_t .$$

It is clear that

$$\hat{B}^{\circ(\gamma * i)}_t = \int_0^t \hat{B}^{\circ\gamma}_s dB^{i}_s + \frac{1}{2}\delta_{\gamma_k, i}\hat{B}^{\circ(\gamma' * 0)}.$$

Now $|\gamma' * 0| = k$, so $\hat{B}^{\circ(\gamma' * 0)}$ satisfies (147) with $\|\gamma' * 0\| = \|\gamma'\| + 2$. Moreover, we can control $\int_0^t \hat{B}^{\circ\gamma}_s dB^i_s$ by using the inductive step with $u(t,x) = \hat{B}^{\circ\gamma}_t$ and $v \equiv 0$, so that $\frac{1}{2}\|\gamma\| = r_u \ll r_v$, by the inductive hypothesis, and we have:

$$\sup_{t\in(0,T]} \mathbb{E}\left[t^{-r_{\gamma * i}}\left|\hat{B}^{\circ(\gamma * i)}_t\right|\right]^p < \infty,$$

where $r_{\gamma * i} = \min\{(\|\gamma\| + 1)/2, (\|\gamma'\| + 2)/2\} = \|\gamma * i\|/2$.

If $i = 0$, then we may apply the inductive step with $u \equiv 0$ and $v(t,x) = \hat{B}^{\circ\gamma}_t$, so that $\frac{1}{2}\|\gamma\| = r_v \ll r_u$, by the inductive hypothesis. In this case,

$$\sup_{t\in(0,T]} \mathbb{E}\left[t^{-r_{\gamma * i}}\left|\hat{B}^{\circ(\gamma * i)}_t\right|\right]^p < \infty,$$

with, again, $r_{\gamma * i} = \|\gamma * i\|/2$. Hence the result is proved.

Proof of (49): The proof of this result is similar to the induction carried out above. We notice that the remainder term, as defined, is the sum of numerous iterated Stratonovich integrals. We prove that the result holds for each element of the sum. This may then be easily extended to the sum of multiple such objects. We have already seen (cf. Proposition 42) that, for any $\alpha, \beta \in \mathcal{A}(m)$, $p \in [1,\infty)$, $T > 0$:

$$\sup_{\substack{x\in\mathbb{R}^N \\ t\in[0,T]}} \mathbb{E}\left|a_{\alpha,\beta}(t,x)\right|^p < \infty. \tag{150}$$

Moreover, since $c^i_{\alpha,\beta} \in \mathcal{C}^{k+1-|\alpha|}_b(\mathbb{R}^N)$ is uniformly bounded, it follows that

$$\sup_{\substack{x\in\mathbb{R}^N \\ t\in[0,T]}} \mathbb{E}\left|c^j_{\alpha*\gamma,\beta}(X^x_t)\right|^p < \infty. \tag{151}$$

We again prove the result by induction on $|\gamma|$. Assume $|\gamma| = 1$. Using the fact that $c^j_{\alpha*\gamma,\delta}(X^x_t)$ and $a_{\delta,\beta}(t,x)$ satisfy (150) and (151) respectively, the product must satisfy an analogous inequality (by Hörmander's inequality). Note that this semimartingale will be comprised of integrands which are sums and products of objects like those in (150), (151), and hence if $\gamma \in \{1, \dots, d\}$ it has been demonstrated already that (cf. the first part of the proof, i.e. $r_u = r_v = 0$),

$$\mathbb{E}\left[t^{-r_\gamma} \int_0^t c^j_{\alpha*\gamma,\delta}(X^x_s) a_{\delta,\beta}(s,x) \circ dB^\gamma_t\right]^p < \infty, \tag{152}$$

where $r_\gamma = \min\{\frac{1}{2}, 1\} = \frac{1}{2}$. Now if $\gamma = 0$, then we apply the step with $u \equiv 0$ and $v(t,x) = c^j_{\alpha*\gamma,\delta}(X^x_t) a_{\delta,\beta}(t,x)$. That is, $0 = r_v \ll r_u$, to obtain

$$\mathbb{E}\left[t^{-r_\gamma} \int_0^t c^j_{\alpha*\gamma,\delta}(X^x_s) a_{\delta,\beta}(s,x) ds\right]^p < \infty, \tag{153}$$

where $r_\gamma = 1$. We now assume the result holds for some $k \in \mathbb{N}$. i.e. we have the following for all $\gamma \in \mathcal{A}$ satisfying $|\gamma| = k$:

$$\sup_{\substack{x\in\mathbb{R}^N \\ t\in(0,T]}} \mathbb{E}\left[t^{-\|\gamma\|/2} \left|\int_0^t \int_0^{s_k} \dots \int_0^{s_2} (-1)^{|\gamma|} c^j_{\alpha*\gamma,\delta}(X^x_{s_1}) a_{\delta,\beta}(s_1,x) \circ dB^{\gamma_1}_{s_1} \dots \circ dB^{\gamma_k}_{s_k}\right|\right]^p < \infty. \tag{154}$$

To ease the notational burden, we write,

$$Z(t,x,\gamma) := \int_0^t \int_0^{s_k} \dots \int_0^{s_2} (-1)^{|\gamma|} c^j_{\alpha*\gamma,\delta}(X^x_{s_1}) a_{\delta,\beta}(s_1,x) \circ dB^{\gamma_1}_{s_1} \dots \circ dB^{\gamma_k}_{s_k},$$

for $\gamma = (\gamma_1, \dots, \gamma_k)$. Observe, that for $i \in \{1, \dots, d\}$

$$\begin{aligned}
Z(t,x,\gamma*i) &= \int_0^t Z(s,x,\gamma) \circ dB^i_s \\
&= \int_0^t Z(s,x,\gamma) dB^i_s + \frac{1}{2}\left\langle Z(.,x,\gamma), B^i\right\rangle_t \\
&= \int_0^t Z(s,x,\gamma) dB^i_s + \frac{1}{2}\delta_{\gamma_{k-1},\gamma_k} \int_0^t Z(s,x,\gamma') dt \\
&= \int_0^t Z(s,x,\gamma) dB^i_s + \frac{1}{2}\delta_{\gamma_{k-1},\gamma_k} Z(t,x,\gamma'*0).
\end{aligned}$$

By the inductive hypothesis, $Z(t,x,\gamma'*0)$ satisfies (147) with $r_{\gamma'*0} = (\|\gamma'\|+2)/2$, and we also use the inductive step on the right-hand term with $u(t,x) = Z(t,x,\gamma)$ and $v \equiv 0$, so that $r_v \gg r_u = \|\gamma\|/2$, with

$$\sup_{\substack{x\in\mathbb{R}^N \\ t\in(0,T]}} \mathbb{E}\left[t^{-r_{\gamma*i}} \,|\, Z(t,x,\gamma*i)\,|\right]^p < \infty,$$

where $r_{\gamma * i} = \min\left\{\frac{\|\gamma\|+1}{2}, \frac{\|\gamma'\|+2}{2}\right\} = \frac{\|\gamma\|+1}{2}$. If $i = 0$ then we may apply the inductive step with $u \equiv 0$ and $v(t,x) = Z(t,x,\gamma)$, so that with $\|\gamma\|/2 = r_v \ll r_u$ we get

$$\sup_{\substack{x\in\mathbb{R}^N \\ t\in[0,T]}} \mathbb{E}\left[t^{-r_{\gamma*0}} \mid Z(t,x,\gamma*0) \mid\right]^p < \infty,$$

where $r_{\gamma*0} = \frac{\|\gamma\|+2}{2}$. Hence the result is proved.

Finally, note that a finite sum of these would also satisfy a similar inequality with $r_{sum} = \min\{r_k; r_k$ is optimal (i.e. (145) holds) for k-th sum member$\}$. i.e.

$$\sup_{\substack{x\in\mathbb{R}^N \\ t\in(0,T]}} \mathbb{E}\left[t^{-(m+1-\|\alpha\|)/2} \left| r_{\alpha,\beta}(t,x) \right|\right]^p < \infty,$$

as required. □

A.1 Invertibility of the Malliavin Covariance Matrix

The aim of this section is to prove the following proposition from the main body. The proof is demanding, but fundamental to the results, and so it is given its own subsection.

Proposition 21. *$M(t,x)$ is $\mathbb{P}$-a.s. invertible. Moreover, for $p \in [1,\infty)$, $\alpha,\beta \in \mathcal{A}(m)$,*

$$\sup_{t\in(0,1],\, x\in\mathbb{R}^N} \mathbb{E}\left[M^{-1}_{\alpha,\beta}(t,x)\right]^p < \infty. \tag{155}$$

For real-symmetric matrices such as $M(t,x)$ there is an elegant representation of the minimal eigenvalue. The following lemma utilises this to simplify the requirements for invertibility.

Lemma 64. *The statement of the previous proposition holds, providing the following can be shown for each $p \in [1,\infty)$: there exists $C > 0$ s.t.*

$$\mathbb{P}\left(\inf_{|\xi|=1} (\xi, M(t,x)\xi) < \frac{1}{n}\right) < Cn^{-p},$$

for all $n \geq 1,\ t \in (0,1],\ $ and $x \in \mathbb{R}^N$.

Proof. We sketch this proof. It is obvious from what has gone before that elements of the matrix M (rather than those of the inverse) satisfy (155). As the inverse matrix is comprised of the inverse of the determinant multiplied with multilinear combinations of elements of M, it suffices to show that the inverse of the determinant satisfies (155). The element $\inf_{|\xi|=1}(\xi, M(t,x)\xi)$ represents the smallest eigenvalue of M and hence its $-N$th power (where N is $\dim(M)$) provides an upper bound for the inverse of the determinant. Finally the expression in Lemma 64 may be used to deduce the L^p integrability (uniform over $t \in (0,1]$, $x \in \mathbb{R}$) of this upper bound, as it provides the required tail decay. □

In view of these results, consider $(\xi, M(t,x)\xi)$. The determinant of $M(t,x)$ is non-negative and increasing with t. This means that if $M(t,x)$ is a.s. invertible for some $t > 0$, then it must be invertible thereafter. Let $y \geq 1$.

$$
\begin{aligned}
(\xi, M(t,x)\xi) &= \sum_{\alpha,\beta \in \mathcal{A}(m)} \xi_\alpha \xi_\beta M_{\alpha,\beta}(t,x) \\
&= \sum_{\alpha,\beta \in \mathcal{A}(m)} \xi_\alpha \xi_\beta t^{-(\frac{\|\alpha\|}{2}+\frac{\|\beta\|}{2})} \left\langle k_\alpha(t,x), k_\beta(t,x) \right\rangle_H \\
&= \left\| \sum_{\alpha \in \mathcal{A}(m)} \xi_\alpha t^{-\frac{\|\alpha\|}{2}} k_\alpha(t,x) \right\|_H^2 \\
&= \left\| \sum_{\alpha \in \mathcal{A}(m)} \xi_\alpha t^{-\frac{\|\alpha\|}{2}} \int_0^{t\wedge .} (a^0_{.,\alpha} + r_{.,\alpha})(u,x)du \right\|_H^2 \\
&\geq \left\| \sum_{\alpha \in \mathcal{A}(m)} \xi_\alpha t^{-\frac{\|\alpha\|}{2}} \int_0^{t/y\wedge .} (a^0_{.,\alpha} + r_{.,\alpha})(u,x)du \right\|_H^2 .
\end{aligned}
\tag{156}
$$

Observe that, since $y \geq 1$, using the notation: $S^n := \{x \in \mathbb{R}^{n+1} : \|x\| = 1\}$,

$$
\begin{aligned}
&\inf_{\xi \in S^{N_m - 1}} \left\| \sum_{\alpha \in \mathcal{A}(m)} \xi_\alpha t^{-\frac{\|\alpha\|}{2}} \int_0^{t/y\,\wedge .} a_{.,\alpha}(u,x)du \right\|_H^2 \\
&\geq \inf_{\xi \in S^{N_m - 1}} \left\| \sum_{\alpha \in \mathcal{A}(m)} \xi_\alpha \left[\frac{t}{y}\right]^{-\frac{\|\alpha\|}{2}} \int_0^{t/y\,\wedge .} a_{.,\alpha}(u,x)du \right\|_H^2 .
\end{aligned}
$$

Now focus on the term appearing on the RHS:

$$\left\| \sum_{\alpha \in \mathcal{A}(m)} \xi_\alpha \left[\frac{t}{y}\right]^{-\frac{\|\alpha\|}{2}} \int_0^{t/y \wedge \cdot} a_{\cdot,\alpha}(u,x)du \right\|_H^2$$
$$\geq \frac{1}{2} \left\| \sum_{\alpha \in \mathcal{A}(m)} \xi_\alpha \left[\frac{t}{y}\right]^{-\frac{\|\alpha\|}{2}} \int_0^{t/y \wedge \cdot} a^0_{\cdot,\alpha}(u,x)du \right\|_H^2$$
$$- \left\| \sum_{\alpha \in \mathcal{A}(m)} \xi_\alpha \left[\frac{t}{y}\right]^{-\frac{\|\alpha\|}{2}} \int_0^{t/y \wedge \cdot} r_{\cdot,\alpha}(u,x)du \right\|_H^2$$
$$\geq \frac{1}{2} \left\| \sum_{\alpha \in \mathcal{A}(m)} \xi_\alpha \left[\frac{t}{y}\right]^{-\frac{\|\alpha\|}{2}} \int_0^{t/y \wedge \cdot} a^0_{\cdot,\alpha}(u,x)du \right\|_H^2$$
$$- \left(\sum_{\alpha \in \mathcal{A}(m)} \xi_\alpha^2 \right) \left(\int_0^{t/y} \sum_{\alpha \in \mathcal{A}(m)} \sum_{i=1}^{d} \left[\frac{t}{y}\right]^{-\|\alpha\|} r_{i,\alpha}(u,x)^2 du \right).$$

Recall that $a^0_{i,\alpha}(u,x) = 0$ whenever $\alpha \neq i * \gamma$ for all multiindices γ. Moreover, $a^0_{i,\alpha}(u,x) = \hat{B}^{\circ\gamma}_u$ when $i * \gamma = \alpha$. That is, as each multindex $\alpha \in \mathcal{A}(m)$ satisfies $\alpha_1 \in \{1, \ldots, d\}$:

$$a^0_{\cdot,\alpha}(u,x) = \left(0, \ldots, \hat{B}^{\circ\gamma}_u, \ldots, 0\right).$$

It is now necessary to briefly discuss the first term on the RHS. The following result is taken from Kusuoka and Stroock [33], but a comprehensive proof is provided in the next section.

Proposition 65. *Given $m \in \mathbb{N}$, there exist constants $C_m, \mu_m \in (0,\infty)$ such that for all $T > 0$*

$$\mathbb{P}\left(\inf_{a \in S^{N^{0,\emptyset}_{m-1}-1}} \int_0^T \left[\sum_{\gamma \in \mathcal{A}_{0,\emptyset}(m-1)} T^{-\frac{\|\gamma\|}{2}-\frac{1}{2}} a_\gamma \hat{B}^{\circ\gamma}_t \right]^2 dt \leq \frac{1}{n} \right) \leq C_m \exp\{-n^{\mu_m}\}. \tag{157}$$

Proof. The proof of this result requires a detour. For a detailed proof, consult the next section of the appendix. □

As a result of this strong bound, which is incidentally much stronger than that which is required for invertibility, it is very easy to deduce the following two equivalent properties:

Corollary 66. *For any $m \in \mathbb{N}$, and $p \in [1,\infty)$, there holds:*

$$\mathbb{E}\left(\inf_{a\in S^{N^{0,\emptyset}_{m-1}-1}} \int_0^T \left[\sum_{\gamma\in\mathcal{A}_{0,\emptyset}(m-1)} T^{-\frac{\|\gamma\|}{2}-\frac{1}{2}} a_\gamma \hat{B}_t^{\circ\gamma}\right]^2 dt\right)^{-p} < \infty. \tag{158}$$

And, equivalently, for all $q \in [1,\infty)$

$$\mathbb{P}\left(\inf_{a\in S^{N^{0,\emptyset}_{m-1}-1}} \int_0^t \left[\sum_{\gamma\in\mathcal{A}_{0,\emptyset}(m-1)} T^{-\frac{\|\gamma\|}{2}-\frac{1}{2}} a_\gamma \hat{B}_t^{\circ\gamma}\right]^2 dt \le \frac{1}{n}\right) < C_{m,q} n^{-q}. \tag{159}$$

The usefulness of the above might not be immediately clear, so turn attention back to the lower bound obtained for $(\xi, M(t,x)\xi)$. The fact that any $\alpha \in \mathcal{A}(m)$ can be expressed as $\alpha = j * \gamma$ for some $1 \le j \le d$ and $\gamma \in \mathcal{A}_{0,\emptyset}(m-1)$ is used. This allows the effective utilisation of the structure of $a^0_{\cdot,\alpha}(t,x)$.

$$\begin{aligned}
&\left\| \sum_{\alpha\in\mathcal{A}(m)} \xi_\alpha \left[\frac{t}{y}\right]^{-\frac{\|\alpha\|}{2}} \int_0^{t/y\,\wedge\,\cdot} a^0_{\cdot,\alpha}(u,x)du \right\|_H^2 \\
&= \left\| \sum_{j=1}^d \sum_{\gamma\in\mathcal{A}_{0,\emptyset}(m-1)} \xi_{j*\gamma} \left[\frac{t}{y}\right]^{-\frac{\|\gamma\|+1}{2}} \int_0^{t/y\,\wedge\,\cdot} a^0_{\cdot,j*\gamma}(u,x)du \right\|_H^2 \\
&= \left\| \sum_{\gamma\in\mathcal{A}_{0,\emptyset}(m-1)} \left[\frac{t}{y}\right]^{-\frac{\|\gamma\|+1}{2}} \int_0^{t/y\,\wedge\,\cdot} (\xi_{j*\gamma} \hat{B}_u^{\circ\gamma})_{j=1,\dots,d}\, du \right\|_H^2 \\
&= \left\| \int_0^{t/y\,\wedge\,\cdot} \left(\sum_{\gamma\in\mathcal{A}_{0,\emptyset}(m-1)} \left[\frac{t}{y}\right]^{-\frac{\|\gamma\|+1}{2}} \xi_{j*\gamma} \hat{B}_u^{\circ\gamma} \right)_{j=1,\dots,d} du \right\|_H^2 \\
&= \sum_{j=1}^d \int_0^{t/y\,\wedge\,\cdot} \left[\sum_{\gamma\in\mathcal{A}_{0,\emptyset}(m-1)} \left[\frac{t}{y}\right]^{-\frac{\|\gamma\|+1}{2}} \xi_{j*\gamma} \hat{B}_u^{\circ\gamma} \right]^2 du.
\end{aligned}$$

It can also easily be shown that by taking $\inf_{\xi \in S_{N_m - 1}}$ of both sides:

$$
\begin{aligned}
&\inf_{\xi \in S_{N_m-1}} \left\| \sum_{\alpha \in \mathcal{A}(m)} \xi_\alpha \left[\frac{t}{y}\right]^{-\frac{\|\alpha\|}{2}} \int_0^{t/y\wedge .} a^0_{.,\alpha}(u,x)du \right\|_H^2 \\
&= \inf_{\xi \in S_{N_m-1}} \sum_{j=1}^d \int_0^{t/y\wedge .} \left[\sum_{\gamma \in \mathcal{A}_{0,\emptyset}(m-1)} \left[\frac{t}{y}\right]^{-\frac{\|\gamma\|+1}{2}} \xi_{j*\gamma} \hat{B}^{\circ\gamma}_{s,u} \right]^2 du \\
&= \inf_{a \in S_{N^{0,\emptyset}_{m-1}-1}} \int_0^{t/y\wedge .} \left[\sum_{\gamma \in \mathcal{A}_{0,\emptyset}(m-1)} \left[\frac{t}{y}\right]^{-\frac{\|\gamma\|+1}{2}} a_\gamma \hat{B}^{\circ\gamma}_{u} \right]^2 du,
\end{aligned}
$$

recalling that $N^{0,\emptyset}_{m-1} = N^{0,\emptyset}_{m-1} + 2$ This is precisely why the upper bound derived in Proposition 65 was introduced. It enables a precise control over the tail behaviour of $(\xi, M(t,x)\xi)$. The various pieces of analysis are now synthesised. In what follows, note that

$$
\begin{aligned}
\mathbb{P}\left(\frac{1}{2}X - Y \le \frac{1}{n}\right) &= \mathbb{P}\left(\frac{1}{2}X - Y \le \frac{1}{n}, Y < \frac{1}{n}\right) + \mathbb{P}\left(\frac{1}{2}X - Y \le \frac{1}{n},\ Y \ge \frac{1}{n}\right) \\
&\le \mathbb{P}\left(Y \ge \frac{1}{n}\right) + \mathbb{P}\left(X \le \frac{4}{n}\right).
\end{aligned}
$$

This gives:

$$
\begin{aligned}
&\mathbb{P}\left(\inf_{\xi \in S^{N_m-1}} (\xi, M(t,x)\xi) < \frac{1}{n}\right) \\
&\quad \le \mathbb{P}\left(\inf_{a \in S_{N^{0,\emptyset}_{m-1}+1-1}} \int_0^{t/y} \left[\sum_{\gamma \in \mathcal{A}_{0,\emptyset}(m-1)} \left[\frac{t}{y}\right]^{-\frac{\|\gamma\|+1}{2}} a_\gamma \hat{B}^{\circ\gamma}_u \right]^2 du < \frac{4}{n} \right) \\
&\qquad + \mathbb{P}\left(\inf_{\xi \in S^{N_m-1}} \left[\sum_{\alpha \in \mathcal{A}(m)} \xi_\alpha^2 \right] \left[\int_0^{t/y} \sum_{\alpha \in \mathcal{A}(m)} \sum_{i=1}^d \left[\frac{t}{y}\right]^{-\|\alpha\|} r_{i,\alpha}(u,x)^2 du \right] \ge \frac{1}{n} \right) \\
&\quad = \mathbb{P}\left(\inf_{a \in S_{N^{0,\emptyset}_{m-1}+1-1}} \int_0^{t/y} \left[\sum_{\gamma \in \mathcal{A}_{0,\emptyset}(m-1)} \left[\frac{t}{y}\right]^{-\frac{\|\gamma\|+1}{2}} a_\gamma \hat{B}^{\circ\gamma}_u \right]^2 du < \frac{4}{n} \right) \\
&\qquad + \mathbb{P}\left(\int_0^{t/y} \sum_{\alpha \in \mathcal{A}(m)} \sum_{i=1}^d \left[\frac{t}{y}\right]^{-\|\alpha\|} r_{i,\alpha}(u,x)^2 du \ge \frac{1}{n} \right).
\end{aligned}
$$

The program is almost complete. The following is deduced from Proposition 20,

Lemma 67. *There holds, for all $p \in [1,\infty)$,*

$$\sup_{\substack{x\in\mathbb{R}^N\\ t\in(0,1]}} \mathbb{E}\left(\int_0^t \sum_{\alpha\in\mathcal{A}(m)} \sum_{i=1}^d t^{-\|\alpha\|-1} r_{i,\alpha}(u,x)^2 du\right)^p < \infty.$$

Proof. We may apply the semimartingale rate bound obtained in the proof of Proposition 20. Indeed, we observe that:

$$\begin{aligned}\xi_t &:= \int_0^t u(s,x)dB_s + \int_0^t v(s,x)ds,\\ u(s,x) &\equiv 0,\\ v(s,x) &= \sum_{\alpha\in\mathcal{A}(m)} \sum_{i=1}^d t^{-\|\alpha\|} r_{i,\alpha}(s,x)^2.\end{aligned}$$

Observe from Proposition 49, noting $\|\alpha\| \le m$,

$$\sup_{\substack{x\in\mathbb{R}^N\\ t\in(0,T]}} \mathbb{E}\left(t^{-r_u} \mid u(t,x) \mid\right)^p < \infty, \qquad \sup_{\substack{x\in\mathbb{R}^N\\ t\in(0,T]}} \mathbb{E}\left(t^{-r_v} \mid v(t,x) \mid\right)^p < \infty,$$

where $r_v = 0$ and r_u is arbitrarily large. Hence it follows that:

$$\sup_{\substack{x\in\mathbb{R}^N\\ t\in(0,T]}} \mathbb{E}\left(t^{-r} \mid \xi_t^x \mid\right)^p < \infty,$$

where $r_\xi = r_v + 1 = 1$, as required. □

The proof can now be completed.

$$\begin{aligned}&\mathbb{P}\left(\inf_{a\in S^{N_{m-1}^{0,\emptyset}-1}} \left\{\int_0^{t/y} \left[\sum_{\gamma\in\mathcal{A}_{0,\emptyset}(m-1)} \left[\frac{t}{y}\right]^{-\frac{\|\gamma\|+1}{2}} a_\gamma \hat{B}_u^{\circ\gamma}\right]^2 du\right\} < \frac{4}{n}\right)\\ &\quad + \mathbb{P}\left(\int_0^{t/y} \sum_{\alpha\in\mathcal{A}(m)} \sum_{i=1}^d \left[\frac{t}{y}\right]^{-\|\alpha\|} r_{i,\alpha}(u,x)^2 du \ge \frac{1}{n}\right)\\ &\qquad = \mathbb{P}\left(\inf_{a\in S_{N_{m-1}^{0,\emptyset}+1-1}} \left\{\int_0^{t/y} \left[\sum_{\gamma\in\mathcal{A}_{0,\emptyset}(m-1)} \left[\frac{t}{y}\right]^{-\frac{\|\gamma\|+1}{2}} a_\gamma \hat{B}_u^{\circ\gamma}\right]^2 du\right\} < \frac{4}{n}\right)\end{aligned}$$

$$+ \mathbb{P}\left(\int_0^{t/y} \sum_{\alpha \in \mathcal{A}(m)} \sum_{i=1}^{d} \left[\frac{t}{y} \right]^{-\|\alpha\|-1} r_{i,\alpha}(u,x)^2 du \geq \frac{y}{nt} \right)$$

$$< C_{m,q}\left(\frac{4}{n}\right)^q + \tilde{C}_{m,q}\left(\frac{nt}{y}\right)^q \leq C_{m,q}\left(\frac{4}{n}\right)^q + \tilde{C}_{m,q}\left(\frac{n}{y}\right)^q.$$

It is important to note that the above bounds hold $\forall t \in (0,1]$ and $\forall x \in \mathbb{R}^N$. The decision to introduce $y \geq 1$ should become clear. Without it, the analysis would fail. Indeed, there is a clever choice of y such that Lemma 64 holds. Set

$$y = \frac{n^2}{4},$$

so that

$$\frac{n}{y} = \frac{4}{n}.$$

And finally, combining this with the above we obtain:

$$\mathbb{P}\left(\inf_{\xi \in S^{N_m - 1}} (\xi, M(t,x)\xi) < \frac{1}{n} \right) < \tilde{\tilde{C}}_{m,p} \frac{1}{n^q},$$

as required.

In the next section regularity results about the inverse of the matrix are proved. These results shall be fundamental to the integration by parts formula.

A.2 *Diffuseness of Iterated Stratonovich Integrals*

It was seen in the last section that invertibility of the Malliavin covariance matrix can be achieved if Proposition 65 holds. Its statement is recalled and it is sought to prove this result using the work of Kusuoka/Stroock in [33] as a guide.

Proposition 68. *For any $m \in \mathbb{N}$, and $p \in [1,\infty)$, there holds:*

$$\mathbb{E}\left(\inf_{a \in S^{N_{m-1}^{0,\emptyset}-1}} \int_0^T \left[\sum_{\gamma \in \mathcal{A}_{0,\emptyset}(m-1)} T^{-\frac{\|\gamma\|}{2}-\frac{1}{2}} a_\gamma \hat{B}_t^{\circ\gamma} \right]^2 dt \right)^{-p} = C_{m,p} < \infty. \quad (160)$$

And, equivalently, for all $q \in [1,\infty)$

$$\mathbb{P}\left(\inf_{a \in S^{N_{m-1}^{0,\emptyset}-1}} \int_0^T \left[\sum_{\gamma \in \mathcal{A}_{0,\emptyset}(m-1)} T^{-\frac{\|\gamma\|}{2}-\frac{1}{2}} a_\gamma \hat{B}_t^{\circ\gamma} \right]^2 dt \leq \frac{1}{n} \right) < C_{m,q} n^{-q}. \quad (161)$$

Proof. The proof of this important result is begun through simplification of the problem. By considering the distribution of the iterated Stratonovich integrals one is able to make a change of variable to the integral. Indeed, note that:

$$\hat{B}_{st}^{\circ\gamma} \stackrel{\mathcal{D}}{=} s^{\frac{\|\gamma\|}{2}} \hat{B}_t^{\circ\gamma},$$

Hence it may be deduced:

$$\int_0^T \left[\sum_{\gamma\in\mathcal{A}_{0,\emptyset}(m-1)} T^{-\frac{\|\gamma\|}{2}-\frac{1}{2}} a_\gamma \hat{B}_t^{\circ\gamma} \right]^2 dt$$

$$\stackrel{\mathcal{D}}{=} \int_0^T \left[\sum_{\gamma\in\mathcal{A}_{0,\emptyset}(m-1)} T^{-\frac{1}{2}} a_\gamma \hat{B}_{\frac{t}{T}}^{\circ\gamma} \right]^2 dt$$

$$\stackrel{u=t/T}{=} \int_0^1 \left[\sum_{\gamma\in\mathcal{A}_{0,\emptyset}(m-1)} a_\gamma \hat{B}_u^{\circ\gamma} \right]^2 du.$$

Hence, the problem is reduced to showing that for each $p \geq 1$, there exists $C > 0$ s.t.

$$\mathbb{P}\left(\inf_{a\in S^{N_{m-1}^{0,\emptyset}-1}} \int_0^1 \left[\sum_{\gamma\in\mathcal{A}_{0,\emptyset}(m-1)} a_\gamma \hat{B}_u^{\circ\gamma} \right]^2 du < \frac{1}{n} \right) < Cn^{-p}, \tag{162}$$

for all $n \geq 1$.

Iterated Stratonovich integrals arise in a very natural way from the geometry of this problem. That said, one must often turn to the more established results in stochastic integration to do an accurate analysis of them. These results are almost always phrased in terms of Itô integration and the semimartingales resulting therefrom. Hence, attention is switched to iterated Itô integrals via the following proposition. The moral of the story is that, although undoubtedly different objects, iterated Itô and Stratonovich integrals are equally as diffuse.

Proposition 69. *Define* $\hat{B}_t^{\circ L} := (\hat{B}_t^{\circ\alpha})_{\|\alpha\|\leq L}$ *and* $\hat{B}_t^L := (\hat{B}_t^\alpha)_{\|\alpha\|\leq L}$. *Then, for all* $L \in \mathbb{N}$ *there exist constant matrices* $A_L, \tilde{A}_L \in \mathbb{R}^{N_L\times N_L}$ *such that*

$$\text{(i):} \quad \hat{B}_t^{\circ L} = A_L \hat{B}_t^L \qquad \text{and} \qquad \text{(ii):} \quad \hat{B}_t^L = \tilde{A}_L \hat{B}_t^{\circ L}.$$

i.e. A_L *is invertible with* $A_L^{-1} = \tilde{A}_L$.
Moreover, it follows that the existence of constants $C_m, \mu_m \in (0,\infty)$

$$\mathbb{P}\left(\inf_{\sum a_\gamma^2=1} \int_0^1 \left[\sum_{\gamma\in\mathcal{A}_{0,\emptyset}(m-1)} a_\gamma \hat{B}_t^{\circ\gamma} \right]^2 dt \leq \frac{1}{n} \right) \leq C_m \exp\{-n^{\mu_m}\},$$

is equivalent to the existence of constants $\tilde{C}_m, \tilde{\mu}_m \in (0, \infty)$ *such that*

$$\mathbb{P}\left(\inf_{\sum a_\gamma^2 = 1} \int_0^1 \left[\sum_{\gamma \in \mathcal{A}_{0,\emptyset}(m-1)} a_\gamma \hat{B}_t^\gamma \right]^2 dt \leq \frac{1}{n} \right) \leq \tilde{C}_m \exp\{-n^{\tilde{\mu}_m}\}.$$

Proof of (i) (adapted from the proof of Lemma A.12 in Kusuoka and Stroock [33]). (i) is approached by using an induction argument on L. Clearly if $L = 1$ then there is little to prove as $\hat{B}_t^{\circ L} = \hat{B}_t^L$. Hence, as $A_L = I_{d\times d} = \tilde{A}_L$. Now assume that the result holds for $L \leq k$. i.e. for all α such that $\|\alpha\| \leq k$ there holds, for some deterministic constants: $a_{\alpha,\beta}^k$, $\|\beta\| \leq k$.

$$\hat{B}_t^{\circ\alpha} = \sum_{\|\beta\| \leq k} a_{\alpha,\beta}^k \hat{B}_t^\beta.$$

It is clear one need only prove, for suitable constants $a_{\alpha,\beta}^{k+1}$, $\|\beta\| \leq k+1$ for $\|\alpha\| = k+1$

$$\hat{B}_t^{\circ\alpha} = \sum_{\|\beta\| \leq k+1} a_{\alpha,\beta}^{k+1} \hat{B}_t^\beta.$$

Let $\alpha = (\alpha', \alpha^*)$ where $\|\alpha'\| = k-1$ if $\alpha^* = 0$, and $\|\alpha'\| = k$ if $\alpha^* \in \{1, \ldots, d\}$. The cases $\alpha* = 0$ and $\alpha* \in \{1, \ldots, d\}$ are treated separately. Assume first that $\alpha* = 0$. Then

$$\begin{aligned}
\hat{B}_t^{\circ\alpha} = \int_0^t \hat{B}_s^{\circ\alpha'} ds &= \int_0^t \sum_{\|\beta\| \leq k} a_{\alpha',\beta}^k \hat{B}_s^{\circ\beta} ds \\
&= \sum_{\|\beta\| \leq k} a_{\alpha',\beta}^k \hat{B}_t^{\beta * 0} \\
&= \sum_{\substack{\|\beta\| \leq k+1 \\ \beta * = 0}} a_{\alpha',\beta'}^k \hat{B}_t^{\beta * 0} \\
&= \sum_{\|\beta\| \leq k+1} a_{\alpha,\beta}^{k+1} \hat{B}_t^\beta,
\end{aligned}$$

where $a_{\alpha,\beta}^{k+1} = \begin{cases} a_{\alpha',\beta'}^k & \text{if } \beta* = 0 \\ 0 & \text{if } \beta* \neq 0. \end{cases}$

Now assume $\alpha* \in \{1, \dots, d\}$:

$$\begin{aligned}
\hat{B}_t^{\circ\alpha} = \int_0^t \hat{B}_s^{\circ\alpha'} \circ dB_s^{\alpha*} &= \int_0^t \hat{B}_s^{\circ\alpha'} dB_s^{\alpha*} + \frac{1}{2}\int_0^t \hat{B}_s^{\circ\alpha''} ds\, 1_{\{\alpha*=(\alpha')*\}} \\
&= \int_0^t \sum_{\|\beta\|\le k} a_{\alpha',\beta}^k \hat{B}_s^\beta dB_s^{\alpha*} \\
&\quad + \frac{1}{2}\int_0^t \sum_{\|\beta\|\le k} a_{\alpha'',\beta}^k \hat{B}_s^\beta ds\, 1_{\{\alpha*=(\alpha')*\}} \\
&= \sum_{\substack{\|\beta\|\le k+1 \\ \beta*=\alpha*}} a_{\alpha',\beta'}^k \hat{B}_t^\beta + \frac{1_{\{\alpha*=(\alpha')*\}}}{2} \sum_{\substack{\|\beta\|\le k+1 \\ \beta*=0}} a_{\alpha'',\beta'}^k \hat{B}_t^\beta \\
&= \sum_{\|\beta\|\le k+1} a_{\alpha,\beta}^{k+1} \hat{B}_t^\beta,
\end{aligned}$$

where $a_{\alpha,\beta}^{k+1} = \begin{cases} a_{\alpha',\beta'}^k & \text{if } \alpha* = \beta*, \\ \frac{1}{2} a_{\alpha'',\beta'}^{k-1} & \text{if } \beta* = 0, \alpha* = (\alpha')*, \\ 0 & \text{otherwise.} \end{cases}$

This completes the argument. As (ii) can be proved in an analogous manner, its proof is omitted. It is now shown how $(i), (ii)$ imply the remaining equivalence result. Note that if A_L is invertible, then A_L^T is also invertible with $(A_L^T)^{-1} = (A_L^{-1})^T$. Moreover, from invertibility

$$0 < c_{\min} := \min_{|\xi|=1} \left| A_L^T \xi \right|.$$

Adopting the shorthand notation $\hat{B}_t^{\circ L}$, $\hat{B}_t^L$ employed above, there holds:

$$\begin{aligned}
\inf_{|\xi|=1} \int_0^1 \left(\xi, \hat{B}_t^{\circ L}\right)^2 dt &= \inf_{|\xi|=1} \int_0^1 \left(\xi, A_L \hat{B}_t^L\right)^2 dt \\
&= \inf_{|\xi|=1} \int_0^1 \left(A_L^T \xi, \hat{B}_t^L\right)^2 dt \\
&\ge \inf_{|\nu|=1} \int_0^1 \left(\nu, \hat{B}_t^L\right)^2 dt\ c_{\min}^2.
\end{aligned}$$

A similar estimate can be made from (ii). These estimates prove the remaining claim of the proposition. □

Before tackling Proposition 65 in earnest, some supplementary results about iterated Itô integrals are required.

Lemma 70. *Fix $l \in \mathbb{N}$. There exists $C_l < \infty$ and $\nu_l > 0$ such that for all $\alpha \in \mathcal{A}$ with $\|\alpha\| = l$, there holds:*

$$\mathbb{P}\left(\sup_{t\in(0,1]}\left|\hat{B}_t^{\alpha}\right| \geq n\right) \leq C_l \exp\left(-\frac{1}{2}n^{\nu_l}\right), \tag{163}$$

for all $n \geq 1$.

Proof (adapted from the proof of Lemma A.7 in Kusuoka and Stroock [33]). Fundamental use of the following martingale inequality is made. For $K_1, K_2 \geq 0$

$$\mathbb{P}\left(\sup_{t\in(0,T]} | M_T | \geq K_1,\ \langle M\rangle_T \leq K_2\right) \leq 2\exp\left\{-\frac{K_1^2}{2K_2}\right\}.$$

This result is proved by expressing the above martingale as time-changed Brownian motion (run at the "speed" of its quadratic variation, see Karatzas and Shreve [26, Theorem 3.4.6]), and then using the following two inequalities:

$$\mathbb{P}(\sup_{t\in(0,T]} | B_t | \geq K) \leq 2\mathbb{P}(B_T \geq K),$$

$$\int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \leq e^{-x^2/2}, \qquad x \geq 0.$$

The latter is seen by splitting consideration into two cases: $x \in [0, 1)$ and $x \geq 1$.

The relation in question can be obtained by iterative applications of this martingale inequality. Define $\nu_N \equiv 2$, and in what follows allow ν_i to be chosen optimally afterwards. First assume that $\alpha \in \{1, \ldots, d\}^N$.

$$\begin{aligned}
&\mathbb{P}\Big[\sup_{t\in(0,1]}\left|\hat{B}_t^{\alpha}\right| \geq K\Big]\\
&\leq \mathbb{P}\Big[\sup_{t\in(0,1]}\left|\hat{B}_t^{\alpha}\right| \geq K,\ \langle \hat{B}^{\alpha'}\rangle_1 < K^{\nu_N}\Big] + \mathbb{P}\Big[\langle \hat{B}^{\alpha}\rangle_1 \geq K^{\nu_N}\Big]\\
&= \mathbb{P}\Big[\sup_{t\in(0,1]}\left|\hat{B}_t^{\alpha}\right| \geq K,\ \langle \hat{B}^{\alpha'}\rangle_1 < K^{\nu_N}\Big] + \mathbb{P}\Big[\int_0^1 \left|\hat{B}_t^{\alpha'}\right|^2 dt \geq K^{\nu_N}\Big]\\
&\leq \mathbb{P}\Big[\sup_{t\in(0,1]}\left|\hat{B}_t^{\alpha}\right| \geq K,\ \langle \hat{B}^{\alpha'}\rangle_1 < K^{\nu_N}\Big] + \mathbb{P}\Big[\sup_{t\in(0,1]}\left|\hat{B}_t^{\alpha'}\right| \geq K^{\nu_N/2}\Big]\\
&\leq \sum_{i=1}^{N} \mathbb{P}\Big[\sup_{t\in(0,1]}\left|\hat{B}_t^{\alpha(N+1-i)}\right| \geq K^{\nu_i/2},\ \langle \hat{B}^{\alpha(i-1)}\rangle_1 < K^{\nu_{i-1}}\Big]\\
&\leq \sum_{i=1}^{N} 2\exp\left(-\frac{1}{2}K^{\nu_i-\nu_{i-1}}\right),
\end{aligned}$$

where $\alpha^{(i)}$ denotes that shortening of the multi-index $\alpha = (\alpha_1, \ldots, \alpha_N)$ by i. i.e. $\alpha^{(i)} = (\alpha_1, \ldots, \alpha_{N-i})$ (additionally: $\alpha^{(0)} = \alpha$).

Now choose ν_i for $i = 1, \ldots, N$ given that $\nu_N = 2$ and $\nu_0 \geq 0$. In fact, ν_0 can be chosen arbitrarily for $K \geq 1$. If it is assumed that $\nu_i - \nu_{i-1} \equiv \delta > 0$ for $i = 1, \ldots, N$, then

$$\sum_{i=1}^{N} \nu_i - \nu_{i-1} = N\delta \quad \Rightarrow \quad \delta = \frac{2}{N},$$

and $\nu_i = \frac{2i}{N}$. Assembling these facts gives:

$$\mathbb{P}\Big(\sup_{t\in(0,1]} \Big| \hat{B}_t^{\alpha} \Big| \geq K \Big) \leq 2N \exp\left(-\frac{1}{2} K^{\frac{2}{N}} \right), \tag{164}$$

for arbitrary $|\,\alpha\,| = N$. Assuming instead that $\|\alpha\| = l$ and noting that $|\,\alpha\,| \leq \|\alpha\|$ so that $\frac{l}{2} \leq |\,\alpha\,| \leq l$ gives the same upper bound with N replaced by l. i.e. $C_l = 2l$ and $\nu_l = 2/l$.

Now observe that if $\alpha_i = 0$ for some $i = 1, \ldots, N$ the situation is even simpler:

$$\mathbb{P}(\sup_{t\in(0,1]} \Big| \hat{B}_t^{(\alpha_1,\ldots,\alpha_i)} \Big| \geq K) \leq \mathbb{P}(\sup_{t\in(0,1]} \Big| \hat{B}_t^{(\alpha_1,\ldots,\alpha_{i-1})} \Big| \geq K),$$

as $\sup_{t\in(0,T]} \Big| \int_0^t \hat{B}_s^{\alpha} dt \Big| \leq T \sup_{t\in(0,T]} \Big| \hat{B}_t^{\alpha} \Big|$. Therefore, one needs only apply the martingale inequality Card $\{i : \alpha_i \neq 0\}$ times. i.e. $(2\,|\,\alpha\,| - \|\alpha\|)$ times. Hence, for a general α,

$$\mathbb{P}\Big(\sup_{t\in(0,1]} \Big| \hat{B}_t^{\alpha} \Big| \geq K \Big) \leq 2(2\,|\,\alpha\,| - \|\alpha\|) \exp\left(-\frac{1}{2} K^{\frac{2}{2|\alpha| - \|\alpha\|}} \right).$$

However, for any α such that $\|\alpha\| = l$ the identified constants in (164) are still appropriate, as $\sup_{\|\alpha\|=l} (2\,|\,\alpha\,| - \|\alpha\|) = l$. □

The main consequence of the above lemma is the following:

Proposition 71. *It suffices to show the existence of C_m, μ_m such that for all $n \geq 1$, there holds*

$$\sup_{a\in S^{N_{m-1}^{0,\emptyset}-1}} \mathbb{P}\left(\int_0^1 \Big[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)} a_\alpha \hat{B}_t^{\alpha} \Big]^2 dt \leq \frac{1}{n} \right) \leq C_m \exp\{-n^{\mu_m}\}. \tag{165}$$

Adapted from the proof of Lemma 2.3.1 in Nualart [51]. There is some constant M_m such that for all $n \geq 1$, $S^{N_{m-1}^{0,\emptyset}-1}$ contains some finite set $\Sigma(n)$ with

$$| \Sigma(n) | \leq M_m n^{2N_m} \qquad \text{and} \qquad S^{N_{m-1}^{0,\emptyset}-1} \subset \bigcup_{c\in\Sigma(n)} B_{1/\sqrt{2}n}.$$

Observe, for fixed $a^c \in S^{N_{m-1}^{0,\emptyset}-1} \cap B_{1/\sqrt{2}n}(c)$, there holds

$$\begin{aligned}
&\min_{c\in\Sigma(n)} \int_0^1 \left[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)} c_\alpha \hat{B}_t^\alpha \right]^2 dt \\
&= \min_{c\in\Sigma(n)} \int_0^1 \left[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)} (c_\alpha - a_\alpha^c + a_\alpha^c) \hat{B}_t^\alpha \right]^2 dt \\
&\leq 2 \min_{c\in\Sigma(n)} \left(\int_0^1 \left[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)} (c_\alpha - a_\alpha^c) \hat{B}_t^\alpha \right]^2 dt + \int_0^1 \left[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)} a_\alpha^c \hat{B}_t^\alpha \right]^2 dt \right) \\
&\leq 2 \min_{c\in\Sigma(n)} \int_0^1 \left[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)} (c_\alpha - a_\alpha^c) \hat{B}_t^\alpha \right]^2 dt + 2 \min_{c\in\Sigma(n)} \int_0^1 \left[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)} a_\alpha^c \hat{B}_t^\alpha \right]^2 dt \\
&\leq 2 \min_{c\in\Sigma(n)} |c - a^c|^2 \int_0^1 \sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)} \left| \hat{B}_t^\alpha \right|^2 dt + 2 \min_{c\in\Sigma(n)} \int_0^1 \left[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)} a_\alpha^c \hat{B}_t^\alpha \right]^2 dt \\
&\leq 2 \frac{1}{2n^2} \int_0^1 \sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)} \left| \hat{B}_t^\alpha \right|^2 dt + 2 \min_{c\in\Sigma(n)} \int_0^1 \left[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)} a_\alpha^c \hat{B}_t^\alpha \right]^2 dt.
\end{aligned}$$

Now, the above upper bound holds for any $a^c \in S^{N_{m-1}^{0,\emptyset}-1} \cap B_{1/\sqrt{2}n}(c)$, in particular, it must hold upon taking the infimum over all $a \in S^{N_{m-1}^{0,\emptyset}-1}$, as $S^{N_{m-1}^{0,\emptyset}-1} = \bigcup_{c\in\Sigma(n)} S^{N_{m-1}^{0,\emptyset}-1} \cap B_{1/2n^2}(c)$. This gives:

$$\begin{aligned}
\min_{c\in\Sigma(n)} \int_0^1 \left[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)} c_\alpha \hat{B}_t^\alpha \right]^2 dt \leq 2 \frac{1}{2n^2} \int_0^1 \sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)} \left| \hat{B}_t^\alpha \right|^2 dt \\
+ 2 \inf_{a\in S^{N_{m-1}^{0,\emptyset}-1}} \int_0^1 \left[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)} a_\alpha \hat{B}_t^\alpha \right]^2 dt.
\end{aligned} \tag{166}$$

Furthermore, it is evident that:

$$\mathbb{P}\Bigg[\inf_{a\in S^{N^{0,\emptyset}_{m-1}-1}}\int_0^1\Bigg[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)}a_\alpha\hat{B}^\alpha_t\,dt\Bigg]^2\leq\frac{1}{n}\Bigg]$$
$$\leq\mathbb{P}\Bigg(\min_{c\in\Sigma(n)}\int_0^1\Bigg[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)}a^c_\alpha\hat{B}^\alpha_t\,dt\Bigg]^2\leq\frac{3}{n}\Bigg)$$
$$+\mathbb{P}\Bigg(\int_0^1\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)}\left|\hat{B}^\alpha_t\right|^2dt\geq n\Bigg).$$

Using (166) to proceed, it is seen that:

$$\mathbb{P}\Bigg[\inf_{\xi\in S^{N^{0,\emptyset}_{m-1}-1}}\int_0^1\Bigg[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)}a_\alpha\hat{B}^\alpha_t\,dt\Bigg]^2\leq\frac{1}{n}\Bigg]$$
$$\leq\mathbb{P}\Bigg[\min_{c\in\Sigma(n)}\int_0^1\Bigg[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)}c_\alpha\hat{B}^\alpha_t\,dt\Bigg]^2\leq\frac{3}{n}\Bigg]+\mathbb{P}\Bigg[\int_0^1\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)}\left|\hat{B}^\alpha_t\right|^2dt\geq n\Bigg]$$
$$\leq\sum_{c\in\Sigma(n)}\mathbb{P}\Bigg[\int_0^1\Bigg[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)}c_\alpha\hat{B}^\alpha_t\,dt\Bigg]^2\leq\frac{3}{n}\Bigg]+\mathbb{P}\Bigg[\sup_{t\in(0,1]}\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)}\left|\hat{B}^\alpha_t\right|^2\geq n\Bigg]$$
$$\leq M_mK^{2N^{0,\emptyset}_{m-1}}\sup_{\xi\in S^{N^{0,\emptyset}_{m-1}-1}}\mathbb{P}\Bigg[\int_0^1\Bigg[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)}a_\alpha\hat{B}^\alpha_t\,dt\Bigg]^2\leq\frac{3}{n}\Bigg]$$
$$+N^{0,\emptyset}_{m-1}\max_{\alpha\in\mathcal{A}_{0,\emptyset}(m-1)}\mathbb{P}\Bigg[\sup_{t\in(0,1]}\left|\hat{B}^\alpha_t\right|^2\geq\frac{n}{N^{0,\emptyset}_{m-1}}\Bigg]$$
$$\leq M_mn^{2N^{0,\emptyset}_{m-1}}B_m\exp\left(-\left[\frac{n}{3}\right]^{\mu_m}\right)+N^{0,\emptyset}_{m-1}\max_{\substack{k=0,\dots,m-1\\ \|\alpha\|=k}}\mathbb{P}\Bigg[\sup_{t\in(0,1]}\left|\hat{B}^\alpha_t\right|\geq\sqrt{\frac{n}{N^{0,\emptyset}_{m-1}}}\Bigg]$$
$$\leq M_mn^{2N^{0,\emptyset}_{m-1}}B_m\exp\left(-\left[\frac{n}{3}\right]^{\mu_m}\right)+N^{0,\emptyset}_{m-1}\max_{k=0,\dots,m-1}C_m\exp\left(-\frac{1}{2}\left[\frac{n}{N^{0,\emptyset}_{m-1}}\right]^{\frac{\nu_m}{2}}\right)$$
$$\leq A_m\exp\left(-n^{\lambda_m}\right),$$

for some (large) constant A_m and (small) $\lambda_m > 0$, for all $n \geq 1$. Both (163) and (165) have been used. □

The goal is now reasonably clear. If inequality (165) can be proved, then the claim will have been justified. Before turning to this proof in earnest, another supporting result is proved. Note that the rest of the proof is, unless otherwise stated, taken from the appendix (p. 73 and onwards) of Kusuoka and Stroock [33].

Lemma 72. *Assume* $a \in S^{N^{0,\emptyset}_{m-1}-1}$ *such that* $|\, a_\emptyset \,| < 1$.[13] *Then there are constants* $Q_m < \infty$ *and* $\nu_m > 0$ *such that:*

$$\mathbb{P}\left(\int_0^1 \left[\sum_{\alpha \in \mathcal{A}_{0,\emptyset}(m-1)} a_\alpha \hat{B}_t^\alpha \right]^2 dt \le \frac{1}{n} \right) \le Q_m \exp\left\{ -\frac{1}{2} \left(\frac{\left[|\, a_\emptyset \,| - \frac{1}{\sqrt{n}} \right] \vee 0}{\sqrt{1 - a_\emptyset^2}} \right)^{2\nu_m} \right\}. \tag{167}$$

Proof. The starting point is noting that:

$$\begin{aligned}
\left(\int_0^1 \left[\sum_{\alpha \in \mathcal{A}_{0,\emptyset}(m-1)} a_\alpha \hat{B}_t^\alpha \right]^2 dt \right)^{\frac{1}{2}} &\ge |\, a_\emptyset \,| - \left(\int_0^1 \left[\sum_{1 \le \|\alpha\| \le m-1} a_\alpha \hat{B}_t^\alpha \right]^2 dt \right)^{\frac{1}{2}} \\
&\ge |\, a_\emptyset \,| - \sqrt{1 - a_\emptyset^2} \int_0^1 \left[\sum_{1 \le \|\alpha\| \le m-1} \left| \hat{B}_t^\alpha \right|^2 dt \right]^{\frac{1}{2}} \\
&\ge |\, a_\emptyset \,| - \sqrt{1 - a_\emptyset^2} \sup_{t \in (0,1]} \left[\sum_{1 \le \|\alpha\| \le m-1} \left| \hat{B}_t^\alpha \right|^2 \right]^{\frac{1}{2}}.
\end{aligned}$$

Consequently,

$$\sup_{t \in (0,1]} \sum_{1 \le \|\alpha\| \le m-1} \left| \hat{B}_t^\alpha \right|^2 \ge \left(\frac{\left[|\, a_\emptyset \,| - \left(\int_0^1 \left[\sum_{\alpha \in \mathcal{A}_{0,\emptyset}(m-1)} a_\alpha \hat{B}_t^\alpha \right]^2 dt \right)^{\frac{1}{2}} \right] \vee 0}{\sqrt{1 - a_\emptyset^2}} \right)^2.$$

In particular,

$$\begin{aligned}
&\mathbb{P}\left(\int_0^1 \left[\sum_{\alpha \in \mathcal{A}_{0,\emptyset}(m-1)} a_\alpha \hat{B}_t^\alpha \right]^2 dt \le \frac{1}{n} \right) \\
&\le \mathbb{P}\left(\sup_{t \in (0,1]} \sum_{1 \le \|\alpha\| \le m-1} \left| \hat{B}_t^\alpha \right|^2 \ge \left(\frac{\left(|\, a_\emptyset \,| - \frac{1}{\sqrt{n}} \right) \vee 0}{\sqrt{1 - a_\emptyset^2}} \right)^2 \right) \\
&\le Q_m \quad \exp\left\{ -\frac{1}{2} \left(\frac{\left[(|\, a_\emptyset \,| - \frac{1}{\sqrt{n}}) \vee 0 \right]}{\sqrt{1 - a_\emptyset^2}} \right)^{2\nu_m} \right\},
\end{aligned}$$

for some Q_m, ν_m, where (163) has been used. □

[13] Indeed, the consideration is trivial if this condition is violated.

A semimartingale inequality from Norris [50] is now recalled, which plays an identical role to a similar martingale inequality in Kusuoka and Stroock [33].

Lemma 73. *Assume $a, y \in \mathbb{R}$. Let $\beta = (\beta)_{t\geq 0}$ be a one-dimensional previsible process, and let $\gamma = (\gamma_t := (\gamma_t^1, \ldots, \gamma_t^d))_{t\geq 0}$, $u = (u_t := (u_t^1, \ldots, u_t^d))_{t\geq 0}$ be d-dimensional previsble processes. Moreover, assume $B = (B_t)_{t\geq 0}$ is a d-dimensional Brownian motion. Define,*

$$b_t = b + \int_0^t \beta_s ds + \int_0^t \gamma_s^i dB_s^i,$$

$$Y_t = y + \int_0^t b_s ds + \int_0^t u_s^i dB_s^i.$$

Then for any $q > 8$ and some $\nu < (q-8)/9$, there is a constant $C = C(q, \nu)$ (independent of K) such that

$$\mathbb{P}\left[\int_0^1 Y_t^2 dt < \frac{1}{n},\ \int_0^1 |b_t|^2 + |u_t|^2 dt \geq \frac{1}{n^{1/q}},\ \sup_{t\in(0,1]} |\beta_t| \vee |\gamma_t| \vee |b_t| \vee |u_t| \leq n\right]$$
$$\leq C\ \exp\{-n^{\nu}\}. \tag{168}$$

Remark 74. Upon checking the above result in Norris [50], the keen reader would observe that the result is stated in a different fashion. Namely, the bound

$$\sup_{t\in(0,T]} |\beta_t| \vee |\gamma_t| \vee |b_t| \vee |u_t| \leq M,$$

is assumed up to some bounded stopping time T, as an extra condition. The resulting statement is then phrased in terms of some constant, which depends on M. i.e. $C = C(q, \nu, M)$ in (168). This constraint has been circumvented by letting the constant M depend also on n (indeed: $M = n$). The observation that C is then of the form $C = \hat{C}(q, \nu)n^l$ for some $l \in \mathbb{N}$, is then made. This observation is a result of tracking the constant in the proof of the lemma. This does not affect (168) as there is some larger constant $\tilde{C}$ and smaller $\tilde{\nu}$, which can be chosen such that $\hat{C}(q, \nu)n^l \exp\{-n^{\nu}\} \leq \tilde{C}(q, \nu)\exp\{-n^{\tilde{\nu}}\}$, for all $n \geq 1$.

The proof of the bound in Proposition 71 is done via a strong induction argument. The base case $m - 1 = 0$ is trivial. Assume therefore, that (165) holds for $0 \leq m - 1 \leq k - 1$. Let $a \in S^{N_k^{0,\emptyset}-1}$. Define, using the notation of Lemma 73, the following:

$$Y_t := \sum_{\|\alpha\|\leq k} a_\alpha \hat{B}_t^\alpha,$$

$$
\begin{aligned}
b_t &:= \sum_{\substack{1\le\|\alpha\|\le k\\ \alpha^*=0}} a_\alpha \hat{B}_t^{\alpha'},\\
u_t^i &:= \sum_{\substack{1\le\|\alpha\|\le k\\ \alpha^*=i}} a_\alpha \hat{B}_t^{\alpha'},\\
\beta_t &:= \sum_{\substack{1\le\|\alpha\|\le k\\ \alpha^*=0,(\alpha')^*=0}} a_\alpha \hat{B}_t^{\alpha''}, \quad \text{for } |\alpha| \ge 2,\\
\gamma_t^i &:= \sum_{\substack{1\le\|\alpha\|\le k\\ \alpha^*=0,(\alpha')^*=i}} a_\alpha \hat{B}_t^{\alpha''}, \quad \text{for } |\alpha| \ge 2,\\
y &:= a_\emptyset,\\
b &:= 0.
\end{aligned}
$$

With these definitions it is easy to see

$$
\begin{aligned}
b_t &= b + \int_0^t \beta_s ds + \int_0^t \gamma_s^i dB_s^i,\\
Y_t &= y + \int_0^t a_s ds + \int_0^t u_s^i dB_s^i.
\end{aligned}
$$

Using Lemma (72) consideration may be split into two separate cases. Assume first that $1 - a_\emptyset^2 \le n^{-1/2q}$, where $q \ge 1$. So that

$$\sqrt{1 - a_\emptyset^2} \le n^{-1/4q},$$

and

$$
\begin{aligned}
&|\, a_\emptyset \,| \ge \{(1 - n^{-1/2q}) \vee 0\}^{1/2}\\
\Rightarrow &\left[\, |\, a_\emptyset \,| - \frac{1}{\sqrt{n}} \right] \vee 0 \ge (1 - 2n^{-1/2q}) \vee 0.
\end{aligned}
$$

Then, by (167):

$$
\begin{aligned}
\mathbb{P}\Big(\int_0^1 \Big[\sum_{\alpha\in\mathcal{A}_{0,\emptyset}(k)} a_\alpha \hat{B}_t^\alpha \Big]^2 dt \le \frac{1}{n} \Big) &\le Q_k \exp\left\{ -\frac{1}{2} n^{\nu_k/2q} \Big(\big[1 - \frac{2}{n^{1/2q}}\big] \vee 0 \Big)^{2\nu_k} \right\}\\
&\le P_k \exp\left\{ -n^{\lambda_k} \right\},
\end{aligned}
$$

for some (large) constant P_k and (small) λ_k, as required. Suppose now that $1-a_{\emptyset}^2 \geq 1/n^{1/2q}$. Then it is clear that

$$\left\{\int_0^1 \Big[\sum_{\|\alpha\|\leq k} a_\alpha \hat{B}_t^\alpha\Big]^2 dt \leq \frac{1}{n}\right\} \subset E_1 \cup E_2 \cup E_3,$$

where

$$\begin{aligned}
E_1 &= \Big\{ \int_0^1 Y_t^2 \leq \frac{1}{n}, \int_0^1 |\, b_t\,|^2 + |\, u_t\,|^2\, dt \geq \frac{1}{n^{1/q}}, \\
&\qquad \sup_{t\in(0,1]} |\,\beta_t\,| \vee |\,\gamma_t\,| \vee |\, b_t\,| \vee |\, u_t\,| \leq n \Big\}, \\
E_2 &= \Big\{ \sup_{t\in(0,1]} |\,\beta_t\,| \vee |\,\gamma_t\,| \vee |\, b_t\,| \vee |\, u_t\,| > n \Big\}, \\
E_3 &= \Big\{ \int_0^1 |\, b_t\,|^2 + |\, u_t\,|^2\, dt < \frac{1}{n^{1/q}} \Big\}.
\end{aligned}$$

It is now shown that $\mathbb{P}(E_i) \leq C_i \exp\{-n^{\nu_i}\}$ for $i = 1, 2, 3$. For $i = 1, 2$, Lemma 73 and Lemma 70 imply respectively, the required bounds (i.e. independent of $a \in S^{N_{m-1}^{0,\emptyset}-1}$). The case $i = 3$ is handled using the inductive hypothesis.

Define

$$N_j := \sum_{\substack{1\leq\|\alpha\|\leq k-\|(j)\| \\ \alpha^*=j}} a_\alpha^2$$

As $\sum_{j=0}^d N_j = 1 - a_\emptyset^2 \geq 1/n^{1/2q}$, there exists $j_0 \in \{0, \ldots, d\}$ such that $N_{j_0} \geq 1/(d+1)n^{1/2q}$. Moreover, $|\, b_t\,|^2 + |\, u_t\,|^2 \geq \Big|\sum_{\substack{1\leq\|\alpha\|\leq k-\|(j_0)\| \\ \alpha^*=j_0}} a_\alpha \hat{B}_t^{\alpha'}\Big|^2$. Thus, using the inductive hypothesis,

$$\begin{aligned}
\mathbb{P}(E_3) &\leq \mathbb{P}\Bigg(\int_0^1 \Big| \sum_{\substack{1\leq\|\alpha\|\leq k-\|(k_0)\| \\ \alpha^*=j_0}} a_\alpha \hat{B}_t^{\alpha'} \Big|^2 dt \leq \frac{1}{n^{1/q}} \Bigg) \\
&= \mathbb{P}\Bigg(\frac{1}{N_{j_0}} \int_0^1 \Big| \sum_{\substack{1\leq\|\alpha\|\leq k-\|(k_0)\| \\ \alpha^*=j_0}} a_\alpha \hat{B}_t^{\alpha'} \Big|^2 dt \leq \frac{1}{N_{j_0} n^{1/q}} \Bigg) \\
&\leq C_{k-1} \exp\{-(N_{j_0} n^{1/q})^{\nu_{k-1}}\} \\
&\leq C_{k-1} \exp\{-(n^{1/2q}/(d+1))^{\nu_{k-1}}\} \\
&\leq C_k \exp\{-n^{\nu_k}\},
\end{aligned}$$

for some C_k, ν_k. In applying the inductive hypothesis, care has been taken to check that $\left(\sum_{\substack{1 \le \|\alpha\| \le k - \|(k_0)\| \\ \alpha^* = k_0}} a_\alpha^2 \right) / N_{k_0} = 1$. This finishes the proof. □

We now move on the prove Lemma 23 which was fundamental to establishing relationships between the elements of our integration by parts formula. This is done in two stages: in the first stage we focus on demonstrating the result for $\mathcal{K}_r$, that is, those elements which are *smooth* processes. We then supplement this for the non-smooth case with additional comments/proofs where appropriate.

Lemma 75 (Properties of Kusuoka–Stroock Smooth Processes). *The following hold*

1. *Suppose* $f \in \mathcal{K}_r(E)$, *where* $r \ge 0$. *Then, for* $i = 1, \dots, d$,

$$\int_0^\cdot f(s,x) dB_s^i \in \mathcal{K}_{r+1}(E) \quad \textit{and} \quad \int_0^\cdot f(s,x) ds \in \mathcal{K}_{r+2}(E).$$

2. $a_{\alpha,\beta}, b_{\alpha,\beta} \in \mathcal{K}_{(\|\beta\| - \|\alpha\|) \vee 0}$ *where* $\alpha, \beta \in \mathcal{A}(m)$.
3. $k_\alpha \in \mathcal{K}_{\|\alpha\|}(H)$, *where* $\alpha \in \mathcal{A}(m)$.
4. $D^{(\alpha)} u := \langle Du(t,x), k_\alpha \rangle_H \in \mathcal{K}_{r+\|\alpha\|}$ *where* $u \in \mathcal{K}_r$ *and* $\alpha \in \mathcal{A}(m)$.
5. *If* $M^{-1}(t,x)$ *is the inverse matrix of* $M(t,x)$, *then* $M^{-1}_{\alpha,\beta} \in \mathcal{K}_0$, $\alpha, \beta \in \mathcal{A}(m)$.
6. *If* $f_i \in \mathcal{K}_{r_i}$ *for* $i = 1, \dots, N$, *then*

$$\prod_{i=1}^N f_i \in \mathcal{K}_{r_1 + \ldots + r_N} \quad \textit{and} \quad \sum_{i=1}^N f_i \in \mathcal{K}_{\min(r_1, \ldots, r_N)}.$$

Proof. (**1**) It is clear that if $f(t,.)$ is smooth and $\partial_\alpha f(.,.)$ is continuous then the same is true of $\int_0^\cdot f(s,x) dB_s^i$ for $i = 0, \dots, d$, with

$$\partial_\alpha \int_0^\cdot f(s,x) dB_s^i = \int_0^\cdot \partial_\alpha f(s,x) dB_s^i.$$

For $k \ge 1$, $p \in [1, \infty)$, $i = 1, \dots, d$, we have (note that the dependence of the norms on the Hilbert space E has been suppressed):

$$\left\| \int_0^t \partial_\alpha f(s,x) dB_s^i \right\|_{k,p}^p = \mathbb{E} \left\| \int_0^t \partial_\alpha f(s,x) dB_s^i \right\|^p + \sum_{j=1}^k \mathbb{E} \left\| D^j \int_0^t \partial_\alpha f(s,x) dB_s^i \right\|_{H^{\otimes j}}^p. \tag{169}$$

Focussing for a moment of the LHS, and assuming w.l.o.g. $p \ge 2$ (as there holds monotonicity of norms in p), we see that for $j = 0, \dots, k$, there holds

$$
\begin{aligned}
&\mathbb{E}\left\| D^j\left[\int_0^t \partial_\alpha f(s,x)dB_s^i\right]\right\|_{H^{\otimes j}}^p \\
&\quad = \mathbb{E}\left\|\int_0^t D^j\partial_\alpha f(s,x)dB_s^i + \int_0^t D^{j-1}\partial_\alpha f(s,x)\otimes e_i ds\right\|_{H^{\otimes j}\otimes E}^p \\
&\quad \le 2^{p-1}\left[\mathbb{E}\left\|\int_0^t D^j\partial_\alpha f(s,x)dB_s^i\right\|_{H^{\otimes j}}^p + \mathbb{E}\left\|\int_0^t D^{j-1}\partial_\alpha f(s,x)\otimes e_i ds\right\|_{H^{\otimes j}}^p\right] \\
&\quad \le \tilde{C}_p\left[\mathbb{E}\int_0^t t^{\frac{1}{2}(p-1)}\left\|D^j\partial_\alpha f(s,x)\right\|_{H^{\otimes j}}^p + t^{p-1}\left\|D^{j-1}\partial_\alpha f(s,x)\right\|_{H^{\otimes(j-1)}}^p ds\right] \\
&\quad \le \tilde{C}_p t^{\frac{1}{2}(p-1)}\left[\int_0^t \mathbb{E}\left\|D^j\partial_\alpha f(s,x)\right\|_{H^{\otimes j}}^p ds + \int_0^t \mathbb{E}\left\|D^{j-1}\partial_\alpha f(s,x)\right\|_{H^{\otimes(j-1)}}^p ds\right] \\
&\quad \le \tilde{C}_p t^{\frac{1}{2}(p-1)}\int_0^t \|\partial_\alpha f(s,x)\|_{k,p}^p ds \\
&\quad \le \tilde{C}_p t^{\frac{1}{2}(p-1)}\int_0^t s^{rp/2}\sup_{\substack{x\in\mathbb{R}^N\\ v\in(0,1]}} v^{-rp/2}\|\partial_\alpha f(v,x)\|_{k,p}^p ds \\
&\quad \le \tilde{\tilde{C}}_p t^{\frac{1}{2}(p[r+1])},
\end{aligned}
$$

where we have used Burkholder–Davis–Gundy inequality, Jensen's inequality and Hölder's inequality for finite sums. Note that the above holds for $j = 0$ by taking D^{j-1} to be the zero map. The upper bound is independent of $x \in \mathbb{R}^N$ and by a simple rearrangement, and combining with (169), the result follows. Note that the result for $\int_0^\cdot f(s,x)ds$ is proved similarly.

(**2**) The fact that $a_{\alpha,\beta}(t,.), b_{\alpha,\beta}(t,.)$ are smooth with partial derivatives which are jointly continuous in $(t,x)\in(0,1]\times\mathbb{R}^N$ and that $a_{\alpha,\beta}, b_{\alpha,\beta} : [0,T]\times\mathbb{R}^N \to \mathbb{D}^\infty$ follows from Proposition 18. The fact that the appropriate bound holds for $a_{\alpha,\beta}$ with rate $r = (\|\beta\| - \|\alpha\|)\wedge 0$ follows from applying the expression for $a_{\alpha,\beta}$, given in (47), and Proposition 20. The corresponding result for $b_{\alpha,\beta}$ is derived in an analogous way to $a_{\alpha,\beta}$.

(**3**) This follows easily from (**1**) and (**2**).

(**4**) From Nualart [51][Proposition 1.3.3] we have the following:

$$\langle Du, k_\alpha\rangle_H = u\,\delta(k_\alpha) - \delta(u\,k_\alpha)$$

Moreover, we know that $u, k_\alpha \in \mathbb{D}^\infty$, and that $\delta : \mathbb{D}^\infty \to \mathbb{D}^\infty$[14], hence it is clear that $\langle Du, k_\alpha\rangle_H \in \mathbb{D}^\infty$. The existence of regular derivatives of all orders follows from direct differentiation. The required bounds follows easily from 6.

(**5**) Our first observation is that if $f \in \mathcal{K}_r(E)$, where $r \ge 0$, then $g(t,x) := t^{-s/2} f(t,x)$ satisfies $g \in \mathcal{K}_{r-s}(E)$. This is obvious, and from this basic observation we note that $M_{\alpha,\beta}(t,x) := t^{-(\|\alpha\|+\|\beta\|)/2}\left\langle k_\alpha(t,x), k_\beta(t,x)\right\rangle_H$ must

[14] cf. Proposition 1.5.4 in Nualart [51]

satisfy $M_{\alpha,\beta} \in \mathcal{K}_0$. This comes from applying the above observation, along with (**3**) and (**4**) of this Lemma. To prove the same about elements of the inverse of $M(t,x)$ we first note that smoothness (in x) and joint continuity (in (t,x)) follows from the inverse function theorem. To prove Malliavin differentiability and the corresponding bounds, we use the ideas of the proof of Nualart [51, Lemma 2.1.6]. That is, we seek to prove the following:

Lemma 76. *Let $A(.,.)$ be a square random matrix, which is invertible almost surely and such that $|\,detA(t,x)\,|^{-1} \in L^p$ for all $p \geq 1$. Assume further that the elements of $A_{\alpha,\beta}(t,x) \in \mathbb{D}^\infty$ and satisfy:*

$$\sup_{\substack{x\in\mathbb{R}^N \\ s\in[0,t]}} \left\| A_{\alpha,\beta}(s,x) \right\|_{k,p} < \infty.$$

Then $A^{-1}_{\alpha,\beta}(t,x) \in \mathbb{D}^\infty$ and the elements satisfy:

$$\sup_{\substack{x\in\mathbb{R}^N \\ s\in[0,t]}} \left\| A^{-1}_{\alpha,\beta}(s,x) \right\|_{k,p} < \infty. \tag{170}$$

The proof of this lemma is almost identical to the proof of Nualart [51, Lemma 2.1.6]. One merely needs to take care in showing (170). This is done easily by using a Hölder-type inequality for the seminorms $\|.\|_{k,p}$ (cf. Nualart [51, Proposition 1.5.6].

Remark 77. If we hadn't chosen to mutliply and divide the elements of the matrix $\hat{M}(t,x) := (\langle k_\alpha(t,x), k_\beta(t,x)\rangle)$ by $t^{\frac{\|\alpha\|+\|\beta\|}{2}}$, when forming the matrix M, then more care would have been required to ensure that the rate of decay of the inverse (as a Kusuoka Stroock process) is independent of the dimension of the matrix. Indeed, it can be shown the inverse of the determinant of $\hat{M}$ is bounded above by a rate which **is** dimension dependent. However, this dimensionality dependence disappears when one considers the product with the adjugate matrix, which has the equal and opposite dimensionality dependence.

(**6**) It is clear that smoothness, joint continuity and Malliavin differentiability are inherited from the constituent functions. The second property remains to be shown. Consider $\prod_{i=1}^N f_i$. It may be shown that for the kth Malliavin derivative the following Leibniz-type rule holds:

$$D^k \prod_{i=1}^N f_i = \sum_{i_1+\ldots+i_N=k} \binom{k}{i_1,\ldots,i_N} D^{i_1} f_1 \otimes \ldots \otimes D^{i_N} f_N .$$

Now noting that, if $i_1 + \ldots + i_N = k$, we have

$$\left\| D^{i_1} f_1 \otimes \ldots \otimes D^{i_N} f_N \right\|_{H^{\otimes k}} = \prod_{j=1}^N \left\| D^{i_j} f_j \right\|_{H^{\otimes i_j}} ,$$

so that

$$\left\|\prod_{i=1}^{N} f_i(t,x)\right\|_{k,p}^{p}$$
$$= \mathbb{E}\left|\prod_{i=1}^{N} f_i(t,x)\right|^{p} + \sum_{j=1}^{k}\left\|D^j \prod_{i=1}^{N} f_i(t,x)\right\|_{H^{\otimes j}}^{p}$$
$$= \mathbb{E}\left|\prod_{i=1}^{N} f_i(t,x)\right|^{p} + \sum_{j=1}^{k}\left\|\sum_{i_1+\ldots+i_N=j}\binom{j}{i_1,\ldots,i_N}\bigotimes_{m=1}^{N} D^{i_m} f_m(t,x)\right\|_{H^{\otimes j}}^{p}$$
$$\leq \mathbb{E}\left|\prod_{i=1}^{N} f_i(t,x)\right|^{p} + \sum{}' C(p,j)\left\|\bigotimes_{m=1}^{N} D^{i_m} f_m(t,x)\right\|_{H^{\otimes j}}^{p}$$
$$\leq \prod_{i=1}^{N} \| f_i(t,x)\|_{L^{p_i}(\Omega)}^{p} + \sum{}' C(p,j)\prod_{m=1}^{N} \left\| D^{i_m} f_m(t,x)\right\|_{H^{\otimes i_m}}^{p},$$

where $p^{-1} = p_1^{-1} + \ldots p_N^{-1}$, applying Hölder's Generalised Inequality. Whence, letting $r = \sum_{i=1}^{N} r(i)$ we see that

$$\sup_{t\in(0,1],x\in\mathbb{R}^N} t^{-r/2}\left\|\prod_{i=1}^{N} f_i(t,x)\right\|_{k,p}^{p} \leq \prod_{i=1}^{N}\sup_{\substack{t\in(0,1],\\ x\in\mathbb{R}^N}} t^{-\frac{r_i}{2}}\,\| f_i(t,x)\|_{L^{p_i}(\Omega)}^{p}$$
$$+ \sum{}' C(p,j)\prod_{m=1}^{N}\sup_{\substack{t\in(0,1],\\ x\in\mathbb{R}^N}} t^{-\frac{r_i}{2}}\,\left\| D^{i_m} f_m(t,x)\right\|_{H^{\otimes i_m}}^{p}$$
$$< \infty.$$

To see that $\sum_{i=1}^{N} f_i \in \mathcal{K}_{\min(r_1,\ldots,r_N)}$. We note that $\mathcal{K}_r \subset \mathcal{K}_s$ for $r \leq s$. Hence, it should clear that the sum is contained in that $\mathcal{K}_r$ in which **all** of its terms are contained. Namely, $\mathcal{K}_{\min(r_1,\ldots,r_N)}$. A full proof is omitted. □

We now extend the result to coincide with the stated one

Lemma 23 (Properties of Kusuoka–Stroock processes).

1. *Suppose* $f \in \mathcal{K}_r^{loc}(E,n)$, *where* $r \geq 0$. *Then, for* $i = 1,\ldots,d$,

$$\int_0^{\cdot} f(s,x)dB_s^i \in \mathcal{K}_{r+1}^{loc}(E,n) \quad \textit{and} \quad \int_0^{\cdot} f(s,x)ds \in \mathcal{K}_{r+2}^{loc}(E,n).$$

2. $a_{\alpha,\beta}, b_{\alpha,\beta} \in \mathcal{K}^{loc}_{(\|\beta\|-\|\alpha\|)\vee 0}(k-m)$ *where* $\alpha,\beta \in \mathcal{A}(m)$.
3. $k_\alpha \in \mathcal{K}^{loc}_{\|\alpha\|}(H, k-m)$, *where* $\alpha \in \mathcal{A}(m)$.

4. *$D^{(\alpha)}u := \langle Du(t,x), k_\alpha\rangle_H \in \mathcal{K}^{loc}_{r+\|\alpha\|}(n \wedge [k-m])$ where $u \in \mathcal{K}^{loc}_r(n)$ and $\alpha \in \mathcal{A}(m)$.*
5. *If $M^{-1}(t,x)$ is the inverse matrix of $M(t,x)$, then $M^{-1}_{\alpha,\beta} \in \mathcal{K}^{loc}_0(k-m)$, $\alpha, \beta \in \mathcal{A}(m)$.*
6. *If $f_i \in \mathcal{K}^{loc}_{r_i}(n_i)$ for $i = 1, \ldots, N$, then*

$$\prod_{i=1}^{N} f_i \in \mathcal{K}^{loc}_{r_1+\ldots+r_N}(\min_i n_i) \quad \text{and} \quad \sum_{i=1}^{N} f_i \in \mathcal{K}^{loc}_{\min_i r_i}(\min_i n_i).$$

Moreover, if we assume the vector fields $V_0, \ldots, V_d$ are also uniformly bounded, then (2)–(5) hold with $\mathcal{K}^{loc}$ replaced by $\mathcal{K}$.

Proof. The proof of this lemma is very similar to the corresponding lemma in the second chapter. Notes are made on where the proof differs, rather than providing a full and extensive reproof, to avoid repetition.

Proof of 1. It is clear that if $f(t,.)$ n-times differentiable and $\partial_\alpha f(.,.)$ is continuous then the same is true of $\int_s^{\cdot} f(u,x)dB^i_u$ for $i = 0, \ldots, d$, with

$$\partial_\alpha \int_s^{\cdot} f(u,x)dB^i_u = \int_s^{\cdot} \partial_\alpha f(u,x)dB^i_u.$$

The remainder of the proof is analogous.

Proof of 2. The fact that $a_{\alpha,\beta}(t,.), b_{\alpha,\beta}(t,.)$ are k-times differentiable with partial derivatives of order $|\gamma|$, which are jointly continuous in (t,x), and which are in $\mathbb{D}^{k-|\gamma|,p}$ for all $p \geq 1$ follows from Proposition 18. The appropriate bounds can be seen to hold by observing the expression for $a_{\alpha,\beta}$ and applying Proposition 20. The corresponding result for $b_{\alpha,\beta}$ is derived in an analogous way.
Proof of 3. This follows easily from (**1**) and (**2**).
Proof of 4. From Nualart [51][Proposition 1.3.3] we have the following:

$$\langle Du, k_\alpha\rangle_H = u\,\delta(k_\alpha) - \delta(u\,k_\alpha)$$

Moreover, we know that for each $p \geq 1$, there holds $u \in \mathbb{D}^{n,p}$, $k_\alpha \in \mathbb{D}^{(k-m-1),p}$, and that $\delta : \mathbb{D}^{k,p} \to \mathbb{D}^{k-1,p}$ (see, e.g. Proposition 1.5.4 in Nualart [51]), hence it is clear that $\langle Du, k_\alpha\rangle_H \in \mathbb{D}^{n\wedge(k-m-1),q}$ for any $q \geq 1$. The existence of regular derivatives of orders less that $n \wedge (k-m-1)$ follows from direct differentiation, and the required bounds follow from 6.
Proof of 5. The k-times differentiability of the inverse (in x) and joint continuity (in (t,x)) is a result of the inverse function theorem. The Malliavin differentiability of the matrix inverse can be deduced by extending Lemma 76, for square matrices with elements of general Malliavin differentiability.
Proof of 6. It is clear and straightforward to demonstrate that the differentiability and joint continuity are inherited from the constituent functions. The level of differentiability is a result of the product rule for differentiation. The second

property of a K-S-process can be shown in a similar way, making sure to take care of the finite level of differentiability.

A.3 Convergence of the Cubature Method in the Absence of the V_0-Condition

The proof of the convergence of the cubature methods hinges on the control of the L_2-norms of the iterated integrals $I_{f_{\alpha,\varphi}}(t)$, $\alpha = (i_1, \ldots, i_r) \in \mathcal{A}$ in terms of the supremum norm of the gradient bounds of $f_{\alpha,\varphi} := V_{i_1} \ldots V_{i_r}\varphi$ and $V_i f_{\alpha,\varphi}$, $i = 1, \ldots, d$ with the function φ being replaced by $P_t\varphi$. In particular we need to be able to control $V_0 P_t \varphi$ (and higher derivatives involving $V_0 P_t \varphi$). However, under the UFG condition, gradient bounds are available only for derivatives in the directions $V_{[\alpha]}$, $\alpha \in \mathcal{A}$ which explicitly excludes V_0 (see Sect. 2.3 for the definition of $\mathcal{A}$ and Corollaries 31, 32 and respectively 38 for the corresponding bounds). We need to find a way to "hide" V_0. We succeed to do this by employing the Stratonovitch expansion not of $P_t\varphi$, but of $t \to P_{T-t}\varphi(X_t)$, $t \in [0, T]$. Assume $g \in C_b^\infty([0, T] \times \mathbb{R}^N)$. Then, by applying Itô's lemma for Stratonovich integrals, we see that

$$g(T-t, X_t^x) = g(T, x) + \sum_{i=0}^{d} \int_0^t \tilde{V}_i g(T-s, X_s^x) \circ dB_s^i, \tag{171}$$

where $\tilde{V}_i$, $i = 0, \ldots, d$ are the vector fields on $[0, T] \times \mathbb{R}^N$ defined as:

$$\tilde{V}_0 := V_0 - \partial_t, \quad \tilde{V}_i := V_i, \quad i = 1, \ldots, d.$$

Equation (171) may be iterated to obtain the following expansion for $g(T-t, X_t^x)$:

$$g(T-t, X_t^x) = \sum_{\{\alpha,\ \|\alpha\| \le m\}} (\tilde{V}_\alpha g)(T, x)\hat{B}_t^{\circ\alpha} + R_m(t, x, g), \quad m = 2, 3, \ldots, \tag{172}$$

where $\tilde{V}_\alpha = \tilde{V}_{i_1} \ldots \tilde{V}_{i_r}$ for $\alpha = (i_1, \ldots, i_r)$ and

$$R_m(t, x, g) = \sum_{\substack{\|\alpha\|=m+1 \\ \|\alpha\|=m+2, \alpha=0*\beta, \|\beta\|=m}} \tilde{I}_{\tilde{V}_\alpha g}(t) \tag{173}$$

In (173), $\tilde{I}_{\tilde{V}_\alpha g}(t)$, $\alpha = (i_1, \ldots, i_r)$ is defined as

$$\tilde{I}_{\tilde{V}_\alpha g}(t) := \int_0^t \int_0^{s_0} \cdots \left(\int_0^{s_{r-2}} \tilde{V}_\alpha g(T - s_{r-1}, X_{s_{r-1}}^x) \circ dW_{s_{r-1}}^{i_1} \right) \circ \cdots \circ dW_{s_1}^{i_{r-1}} \circ dW_{s_0}^{i_r}.$$

This Taylor expansion will in due course be applied to the diffusion semigroup $P_t f$. Note, in particular, that the cubature measure of order l, where $l \geq m$, agrees with the Wiener measure on the iterated Stratonovich integrals of (172). Therefore the convergence rate will be given by their difference on the "remainder term" R_m. Indeed, it is a simple exercise to show that:

$$\sqrt{\mathbb{E}[R_m(t,x,g)^2]} \leq C \sum_{\substack{\|\alpha\|=m+1 \\ \|\alpha\|=m+2, \alpha=0*\beta, \|\beta\|=m}} t^{\|\alpha\|/2} \|\tilde{V}_\alpha g\|, \tag{174}$$

where

$$\|\tilde{V}_\alpha g\| = \sup_{s\in[0,t], x\in\mathbb{R}^N} |\tilde{V}_\alpha g(T-s,x)|.$$

The expectation $\mathbb{E}[R_m(t,x,g)^2]$ in (174) can be exchanged with the expectation with respect to the cubature measure $\mathbb{Q}_t$, that is, $\mathbb{E}_{\mathbb{Q}_t}[R_m(t,x,g)^2]$ with the result still holding. The following inequality is therefore immediate:

$$\left|\mathbb{E}\left[g(T-t,X_t^x)\right] - \mathbb{E}_{\mathbb{Q}_t}\left[g(T-t,X_t^x)\right]\right| \leq C \sum_{j=m+1}^{m+2} t^{j/2} \sup_{\|\alpha\|=j} \|\tilde{V}_\alpha g\|. \tag{175}$$

The above is an upper bound for the error of a finite measure based on a single application of the cubature formula. Iterated applications of the cubature over the partition $\mathcal{D}$ will give us the correct rate. The Markovian property of the cubature method and the semigroup property of the diffusion allow us to deduce the required uppers bounds based on (175). Again, we emphasize that the difference between what is done here and the earlier proof is that the control on $V_0 P_t \varphi$ is no longer necessary. Instead we need a control along the vector field $\tilde{V}_0 = \partial_t - V_0$ which is available as $P_t f$ is smooth along the vector fields $V_1, \ldots, V_d$ and also for each $(t,x) \in (0,T] \times \mathbb{R}^N$

$$(\partial_t - V_0) P_t f(x) = \sum_{i=1}^{d} V_i^2 P_t f(x) = \frac{1}{t}\mathbb{E}\left[f(X_t^x)\Phi_1(t,x)\right] \tag{176}$$

for a suitably chosen Kusuoka function $\Phi_1(t,x)$ (see Corollary 28). This result may be iterated to prove a corollary similar to Corollary 32.

Corollary 78. *Let $f \in \mathcal{C}_b^\infty(\mathbb{R}^N, \mathbb{R})$. If the vector fields $V_0, \ldots, V_d$ are uniformly bounded then, under the UFG condition, there exists a constant $C_\alpha < \infty$ such that:*

$$\|\tilde{V}_\alpha P_t f\|_\infty \leq \frac{C_\alpha}{t^{(\|\alpha\|-1)/2}} \|\nabla f\|_\infty. \tag{177}$$

Proof. The proof hinges on the observation that $\tilde{V}_\alpha P_t f$ satisfies the following convenient identity

$$\tilde{V}_\alpha P_t f = \sum_{i=1}^{\|\alpha\|} \sum_{\substack{\beta_1,\ldots,\beta_i \in \mathcal{A}, \\ \|\beta_1\|+\ldots+\|\beta_i\|=\|\alpha\|}} c_{\alpha,\beta_1,\ldots,\beta_i} V_{[\beta_1]} \ldots V_{[\beta_i]} P_t f, \tag{178}$$

where $c_{\alpha,\beta_1,\ldots,\beta_i} \in \mathbb{R}$. This is proved by induction over the "length" m of the multi-index α, $m = \|\alpha\|$. The case $\|\alpha\| = 1$ is trivial and $\|\alpha\| = 2$ follows from the first identity in (176). We outline next the inductive step. If $\alpha = (i_1, \ldots, i_r)$, $\|\alpha\| = m + 1$, $m > 1$ and $i_1 \neq 0$, then by the inductive hypothesis

$$\tilde{V}_\alpha P_t f = \sum_{i=1}^{\|\alpha\|-1} \sum_{\substack{\beta_1,\ldots,\beta_i \in \mathcal{A}, \\ \|\beta_1\|+\ldots+\|\beta_i\|=\|\alpha\|-1}} c_{(\alpha_2,\ldots\alpha_r),\beta_1,\ldots,\beta_i} V_{[i_1]} V_{[\beta_1]} \ldots V_{[\beta_i]} P_t f,$$

as, by definition $V_{[i_1]} = V_{i_1}$. If $i_1 = 0$, note that

$$\begin{aligned}(\partial_t - V_0) V_{[\beta_1]} \ldots V_{[\beta_i]} &= [(\partial_t - V_0), V_{[\beta_1]}] V_{[\beta_2]} \ldots V_{[\beta_i]} + V_{[\beta_1]} (\partial_t - V_0) V_{[\beta_2]} \ldots V_{[\beta_i]} \\ &= V_{[(\beta_1,0)]} V_{[\beta_2]} \ldots V_{[\beta_i]} + V_{[\beta_1]} (\partial_t - V_0) V_{[\beta_2]} \ldots V_{[\beta_i]},\end{aligned}$$

since, as ∂_t commutes with $V_{[\beta_1]}$, we have

$$[(\partial_t - V_0), V_{[\beta_1]}] = -[V_0, V_{[\beta_1]}] = V_{[(\beta_1,0)]}$$

By applying the same procedure to the second term and iterating, we obtain eventually that

$$\begin{aligned}&(\partial_t - V_0) V_{[\beta_1]} \ldots V_{[\beta_i]} P_t f \\ &= V_{[(\beta_1,0)]} V_{[\beta_2]} \ldots V_{[\beta_i]} P_t f + \ldots + V_{[\beta_1]} V_{[\beta_2]} \ldots V_{[(\beta_i,0)]} P_t f \\ &\quad + \sum_{j=1}^{d} V_{[\beta_1]} \ldots V_{[\beta_i]} V_{[j]} V_{[j]} P_t f(x).\end{aligned}$$

The last identity together with the induction hypothesis gives us (178) also for the case $i_1 = 0$. From (178) we deduce (177) by using Corrolary 32. □

It is important to note that derivatives along $\tilde{V}_0 := \partial_t - V_0$ add 1 to the rate as a power of t. Let Q_t be the Markov operator defined in (89) corresponding to the m-perfect family of stochastic processes, $\bar{X}(x) = \{\bar{X}_t(x)\}_{t\in[0,\infty)}$ for $x \in \mathbb{R}^d$, constructed by the cubature method as described in Example 41. The following result simply tells us that Lemma 44 holds true also in the absence of the V_0 condition.

Lemma 79. *Under the UFG condition the exists a constant $C = C_T > 0$ independent of $s, t \in [0, T]$ such that*

$$\|P_t(P_s\varphi) - Q_t(P_s\varphi)\|_\infty \leq C\, \|\nabla\varphi\|_\infty \sum_{j=m+1}^{m+2} \frac{t^{\frac{j}{2}}}{s^{\frac{j-1}{2}}}, \tag{179}$$

where $\varphi \in \mathcal{C}_b^1(\mathbb{R}^N, \mathbb{R})$.

Proof. Immediate from (175) and Corollary 78. □

Following Lemma 79, it is now immediate that the same rates of convergence such as those described in Sect. 3.4 are valid for the approximation given by the cubature method in the absence of the cubature measure. Let $T, \gamma > 0$ and $\pi_n = \{t_j = (\frac{j}{n})^\gamma T\}_{j=0}^n$ be a partition of the interval $[0, T]$ where $n \in \mathbb{N}$ is such that $\{h_j = t_j - t_{j-1}\}_{j=1}^n \subseteq (0, 1]$. Just as in the Sect. 3.4, let us define the function,

$$\Upsilon^1(n) = \begin{cases} n^{-\frac{1}{2}\min(\gamma,(m-1))} & \text{if } \gamma \neq m-1 \\ n^{-(m-1)/2}\ln n & \text{for } \gamma = m-1 \end{cases}$$

and let $\mathcal{E}^{\gamma,n}(\varphi)$ be the cubature error In the following,

$$\mathcal{E}^{\gamma,n}(\varphi) := \left\| P_T\varphi - Q_{h_n}^m Q_{h_{n-1}}^m \cdots Q_{h_1}^m \varphi \right\|_\infty$$

for $\gamma \in \mathbb{R}$, $n \in \mathbb{N}$. The proof of the following theorem is identical with that of Theorem 46 and Corollary 47.[15]

Theorem 80. *Under the UFG condition, there exists a constant $C = C(\gamma, T) > 0$ such that, for any $\varphi \in \mathcal{C}_b^1(\mathbb{R}^N, \mathbb{R})$,*

$$\mathcal{E}^{\gamma,n}(\varphi) \leq C\,\Upsilon^1(n)\, \|\nabla\varphi\|_\infty + \left\| P_{h_1}\varphi - Q_{h_1}^m\varphi \right\|_\infty \tag{180}$$

In particular, if $\gamma \geq m-1$ there exists a constant $C' = C'(\gamma, T) > 0$ then,

$$\mathcal{E}^{\gamma,n}(\varphi) \leq \frac{C'}{n^{\frac{m-1}{2}}} \|\nabla\varphi\|_\infty .$$

Acknowledgements We should like to thank the editors of the volume, Vicky Henderson and Ronnie Sircar for their understanding and patience whilst handling our submission. One of the authors (D.C.) would like to thank Nizar Touzi for the invitation to give a series of lectures on the subject in February–April 2010 at the Institute Henri Poincaré. This work grew out of the notes of the IHP lectures. The material in the second section is based on the corresponding chapter in the PhD thesis of C. Nee [44]. The material in Sects. 3.1–3.3 is based on the paper [13].

[15]Theorem 80 can be extended to cover the rate of convergence for test functions $\varphi \in \mathcal{C}_b^p(\mathbb{R}^N, \mathbb{R})$, in the same manner as the corresponding results in Theorem 46 and Corollary 47.

The work of D. Crisan and K. Manolarakis was partially supported by the EPSRC Grant No: EP/H0005500/1. C. Nee's work was supported by an EPSRC Doctoral Training Award.

Special thanks are due to Salvator Ortiz–Latorre who read large portions of the first draft and suggested many corrections and improvements. The authors would also like to thank the anonymous referee for the careful reading of the manuscript and the constructive remarks.

References

1. G.B. Arous, Flots et series de Taylor stochastiques. Probab. Theory Relat. Fields **81**, 29–77 (1989)
2. V. Bally, D. Talay, The law of the Euler scheme for stochastic differential equations I. Convergence rate of the distribution function. Probab. Theory Relat. Fields **104**, 43–60 (1996)
3. B. Bouchard, N. Touzi, Discrete time approximation and Monte Carlo simulation for backward stochastic differential equations. Stoch. Process. Their Appl. **111**, 175–206 (2004)
4. B. Bouchard, X. Warin, Valuation of American options—new algorithm to improve on existing methods. Working paper, 2010
5. K. Burrage, P.M. Burrage, High strong order methods for non-commutative stochastic ordinary differential equation systems and the Magnus formula. Physica D **133**(1–4), 34–48 (1999)
6. K. Burrage, P.M. Burrage, T. Tian, Numerical methods for strong solutions of stochastic differential equations: An overview. Proc. R. Soc. Lond. A **460**, 373–402 (2004)
7. F. Castell, Asymptotic expansion of stochastic flows. Probab. Theory Relat. Fields **96**, 225–239 (1993)
8. F. Castell, J. Gaines, An efficient approximation method for stochastic differential equations by means of exponential lie series. Math. Comput. Simul. **38**(1), 13–19 (1995)
9. P. Cheridito, M. Soner, N. Touzi, N. Victoir, Second-order backward stochastic differential equations and fully non linear parabolic PDEs. Commun. Pure Appl. Math. **60**(7), 1081–1110 (2007)
10. M. Clark, An efficient approximation for a class of stochastic differential equations, in *Advances in Filtering and Optimal Stochastic Control*, vol. 42. Lecture Notes in Control and Information Sciences (Springer, Berlin, 1982), pp. 69–78
11. M. Clark, Asymptotically optimal quadrature formulae for stochastic integrals, in *Proceedings of the 23rd Conference on Decision and Control*, Las Vegas, Nevada (IEEE, 1984), pp. 712–715
12. D. Crisan, F. Delarue, Sharp derivative bounds for solutions of degenerate semi-linear partial differential equations. J. Funct. Anal. **263**(10), 3024–3101 (2012)
13. D. Crisan, S. Ghazali, *On the Convergence Rate of a General Class of SDEs*, (World Scientific, New Jersey, 2007), pp. 221–248
14. D. Crisan, T. Lyons, Minimal entropy approximations and optimal algorithms. Monte Carlo Methods Appl. **8**(4), 343–355 (2002)
15. D. Crisan, K. Manolarakis, Numerical solution for a BSDE using the Cubature method. Application to non linear pricing. SIAM J. Financ. Math. **3**, 534–571 (2012)
16. N. El Karoui, C. Kapoudjan, E. Pardoux, M. Quenez, Reflected solutions of backward SDEs and related obstacle problems. Ann. Probab. **25**(2), 702–737 (1997)
17. N. El Karoui, S. Peng, M. Quenez, Backward stochastic differential equations in finance. Math. Financ. **7**(1), 1–71 (1997)
18. T. Fawcett, Problems in Stochastic Analysis: Connections between Rough Paths and Non-Commutative Harmonic Analysis. Ph.D. thesis, University of Oxford, 2003
19. S. Ghazali, The Global Error in Weak Approximations of Stochastic Differential Equations. Ph.D. thesis, Imperial College London, 2006
20. E. Gobet, C. Labart, Error expansion for the discretization of backward stochastic differential equations. Stoch. Process. Their Appl. **117**(7), 803–829 (2007)

21. E. Gobet, J.P. Lemor, X. Warin, A regression based Monte Carlo method to solve backward stochastic differential equations. Ann. Appl. Probab. **15**(3), 2172–2002 (2005)
22. L.G. Gyurkó, T.J. Lyons, Efficient and practical implementations of cubature on Wiener space, in *Stochastic Analysis 2010* (Springer, Heidelberg, 2011), pp. 73–111
23. S.L. Heston, A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financ. Stud. **6**(2), 327–343 (1993)
24. L. Hörmander, Hypoelliptic second order differential equations. Acta Math. **119**, 147–171 (1967)
25. D. Jerison, Sánchez-Calle, Estimates for the heat kernel for a sum of squares of vector fields. Indiana Univ. Math. **35**(4), 835–854 (1986)
26. I. Karatzas, S. Shreve, *Brownian Motion and Stochastic Calculus* (Springer, New York, 1991)
27. P. Kloeden, E. Platen, *Numerical Solutions of Stochastic Differential Equations* (Springer, Berlin, 1999)
28. H. Kunita, Stochastic differential equations and stochastic flows of diffeomorphisms. Lect. Notes Math. **1097**, 143–303 (1984)
29. S. Kusuoka, Approximation of expectations of diffusion processes based on Lie algebra and Malliavin calculus. Adv. Math. Econ. (Springer, Tokyo) **6**, 69–83 (2004)
30. S. Kusuoka, Malliavin Calculus revisited. J. Math. Sci. Univ. Tokyo **10**, 261–277 (2003)
31. S. Kusuoka, S. Ninomiya, A new simulation method of diffusion processes applied to finance, in *Stochastic Processes and Applications to Mathematical Finance* (World Scientific, River Edge, 2004), pp. 233–253
32. S. Kusuoka, D. Stroock, Applications of the Malliavin calculus. I. Stochastic analysis (Katata/Kyoto, 1982). in *North-Holland Math. Library*, vol. 32 (North-Holland, Amsterdam, 1984), pp. 271–306
33. S. Kusuoka, D, Stroock, Applications of the Malliavin calculus - II. J. Faculty Sci. Univ. Tokyo **1**, 1–76 (1985)
34. S. Kusuoka, D. Stroock, Applications of the Malliavin calculus III. J. Faculty Sci. Univ. Tokyo **34**(1A), 391–422 (1987)
35. C. Litterer, T. Lyons, High order recombination and an application to cubature on Wiener space. Ann. Appl. Probab. **22**(4), 1301–1327 (2012)
36. T. Lyons, N. Victoir, Cubature on Wiener space. Proc. R. Soc. Lond. **468**, 169–198 (2004)
37. J. Ma, J. Zhang, Representation theorems for backward stochastic differential equations. Ann. Appl. Probab. **12**(4), 1390–1418 (2002)
38. J. Ma, P. Protter, J. Yong, Solving forward–backward SDEs expicitly—a four step scheme. Probab. Theory Relat. fields **122**(2), 163–190 (1994)
39. W. Magnus, On the exponential solution of differential equations for a linear operator. Commun. Pure Appl. Math. **7**, 649–673 (1954)
40. P. Malliavin, Stochastic calculus of variations and hypoelliptic operators, in *Proceedings of the International Symposium on Stochastic Differential Equations*, Kyoto, 1976
41. A. Millet, M. Sanz-Sole, A simple proof of the support theorem for diffusion processes, Séminaire de probabilitiés, XXVIII **1583**, 36–48 (1994)
42. G. Milstein, A method of second-order accuracy integration of stochastic differential equations. Theory Probab. Appl. **23**, 396–401 (1976)
43. G. Milstein, Weak approximation of solutions of systems of stochastic differential equations. Theory Probab. Appl. **30**, 750–766 (1985)
44. C. Nee, Sharp Gradient Bounds for the Diffusion Semigroup. Ph.D. thesis, Imperial College London, 2011
45. N. Newton, An asymptotically efficient difference formula for solving stochastic differential equations. Stochastics **19**(3), 175–206 (1986)
46. N. Newton, An efficient approximation for stochastic differential equations on the partition of symmetrical first passage times. Stoch. Stoch. Rep. **29**(2), 227–258 (1990)
47. S. Ninomiya, A partial sampling method applied to the Kusuoka approximation. Mt. Carlo Methods Appl. **9**(1), 27–38, (2003)

48. M. Ninomiya, S. Ninomiya, A new higher-order weak approximation scheme for stochastic differential equations and the Runge-Kutta method. Financ. Stoch. **13**(3), 415–443 (2009)
49. S. Ninomyia, N. Victoir, Weak approximation scheme of stochastic differential equations and applications to derivatives pricing. Appl. Math. Financ. **15**(2), 107–121 (2008)
50. J.R. Norris, Simplified Malliavin calculus. Séminaire de probabilités **20**, 101–130 (1986)
51. D. Nualart, in *The Malliavin Calculus and Related Topics*, 2nd edn. Probability and Its Applications, New York (Springer, Berlin, 2006)
52. E. Platen, W. Wagner, On a Taylor formula for a class of Itô processes. Probab. Math. Stat. **3**, 37–51 (1982)
53. M. Soner, N. Touzi, Well posedness of second order backward SDEs. Stoch. Process. Their Appli. **121**, 265–287 (2011)
54. D.W. Stroock, S.R.S. Varadhan, Diffusion processes with continuous coefficients. Commun. Pure Appl. Math **22**, 345–400 (1969)
55. D.W. Stroock, S.R.S. Varadhan, On the support of diffusion processes with applications to the strong maximum principle, in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 3. (1970), pp. 333–359, Probability Theory
56. D.W. Stroock, S.R.S. Varadhan, On degenerate elliptic-parabolic operators of second order and their associated diffusions. Commun. Pure. Appl. Math. **25**, 651–713 (1972)
57. D. Talay, Efficient numerical schemes for the approximation of expectations of functionals of the solution of an SDE and applications. Filter. Control Random Process. **61**, 294–313 (1984)
58. D. Talay, Discrétisation d'une e.d.s. et calcul approché d'ésperances de fontionelles de la solution,. Math. Model. Numer. Anal. **20**, 141–179 (1986)
59. D. Talay, L. Tubaro, Expansion of the global error for numerical schemes solving stochastic differential equations. Stoch. Anal. Appl. **8**(4), 483–509 (1990)
60. J. Zhang, A numerical scheme for BSDEs. Ann. Appl. Probab. **14**(1), 459–488 (2004)