Weierstraß-Institut

für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

Splitting methods for SPDEs: From robustness to financial engineering, optimal control and nonlinear filtering

Christian Bayer[1], Harald Oberhauser[2]

submitted: February 3, 2015

[1] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: christian.bayer@wias-berlin.de

[2] Oxford-Man Institute of Quantitative Finance
University of Oxford
Eagle House
Walton Well Road
Oxford OX2 6ED
United Kingdom
E-Mail:harald.oberhauser@oxford-man.ox.ac.uk

No. 2072

Berlin 2015

2010 *Mathematics Subject Classification.* 60H15, 60H35, 65C30.

Key words and phrases. Splitting methots, SPDEs, rough paths, Ninomiya-Victoir method.

Harald Oberhauser is grateful for the support of the ERC (grant agreement No.291244 Esig) and the Oxford-Man Institute of Quantitative finance.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

In this survey paper we give an overview of recent applications of the splitting method to stochastic (partial) differential equations, that is, differential equations that evolve under the influence of noise. We discuss weak and strong approximations schemes. The applications range from the management of risk, financial engineering, optimal control and nonlinear filtering to the viscosity theory of nonlinear SPDEs.

1 Introduction

The theory of (ordinary/partial) differential equations has been incredibly successful in modelling quantities that evolve over time. Many of these quantities can be profoundly affected by stochastic fluctuations, noise, and randomness. The theory of stochastic differential equations aims for a qualitative and quantitative understanding of the effects of such stochastic perturbations. This requires insights from pure mathematics and to deal with them in practice requires us to revisit and extend classic numerical techniques. Splitting methods turn out to be especially useful since they often allow to separate the problem into a deterministic and a stochastic part.

White noise and Brownian motion

The arguably simplest case of such a stochastic perturbation is an ODE driven by a vector field V that is affected by noise which is modelled below by a sequence of random variables $N = (N_t)_{t\geq 0}$ and picked up by a vector field W,

$$\frac{dy_t}{dt} = V(y_t) \underbrace{+W(y_t) N_t}_{\text{Noise}}.$$

Often a reasonable assumption is that $N = (N_t)_{t\geq 0}$ is *white noise*, that is

1. (independence) $\forall s \neq t$, N_t and N_s are independent,
2. (stationarity) $\forall t_1 \leq \cdots \leq t_n$ the law of $(N_{t_1+t}, \cdots, N_{t_n+t})$ does not depend on t,
3. (centered) $\mathbb{E}[N_t] = 0, \forall t \geq 0$.

Above properties imply that the trajectory $t \mapsto N_t$ cannot be continuous, and even worse if we assume that $\mathbb{E}[N_t^2] = 1$ then $(\omega, t) \mapsto N_t(\omega)$ is not even measurable (see [57, 38]). Putting mathematical rigour aside, let us rewrite the above differential equation as an integral equation, i.e. we work with $B_t = \int_0^t N_r dr$ and since integration smoothes out we expect $B = (B_t)_{t\geq 0}$ to have nicer trajectories than N. In this case the above becomes

$$\mathrm{d}y_t = V(y_t)\,\mathrm{d}t + W(y_t)\,\mathrm{d}B_t \text{ resp. } y_t = \int_0^t V(y_r)\,\mathrm{d}r + \int_0^t W(y_r)\,\mathrm{d}B_r. \tag{1}$$

It turns out that $B = (B_t)_{t\geq 0}$ can be rigorously defined as a stochastic process — i.e. a collection of (ω, t)-measurable random variables carried on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. This process B is the well-known *Brownian motion* (also called Wiener process)[1].

[1] Named after the botanist Robert Brown who observed in 1827 that pollen grains suspended in water execute continuous but jittery motions. The physical explanation was given by Albert Einstein in 1905 (his "annus mirabilis": small water molecules hit the pollen) and a little earlier Marian Smoluchowski had already emphasized the importance of this process for physics. Further important contributions are due to Louis Bachelier, Andrey Kolmogorov, Paul Lévy, Joseph Doob, Norbert Wiener and finally Kyoshi Ito

Definition 1.1. We call a real-valued stochastic process $B = (B_t)_{t\geq 0}$ defined on a probability space $(\Omega, \mathbb{P})$ a one-dimensional Brownian motion if

1 $B_0 = 0$ and $t \mapsto B_t$ is continuous (a.s.),

2 $\forall t_1 < \cdots < t_n$ and $n \in \mathbb{N}$, $B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$ are independent,

3 $\forall s, t, t - s \geq 0$, $B_t - B_s \sim \mathcal{N}(0, t - s)$.

The trajectories $t \mapsto B_t(\omega)$ are "degenerate" from the point of view of classical analysis: they are highly oscillatory, of infinite length, (statistically) self-similar and possess a rich fractal structure; see Figure 1. Developing a theory that can deal with such "degenerate" trajectories is what makes stochastic calculus such a fascinating and rich subject. Finally, let us note that while Brownian motion is probably the most important stochastic process, there are many other classes of noise that appear in the real-world and are not covered by the Brownian (e.g. so-called fractional Brownian motion [53]) and many of the methods we present here are not limited to the Brownian or even semimartingale setting.

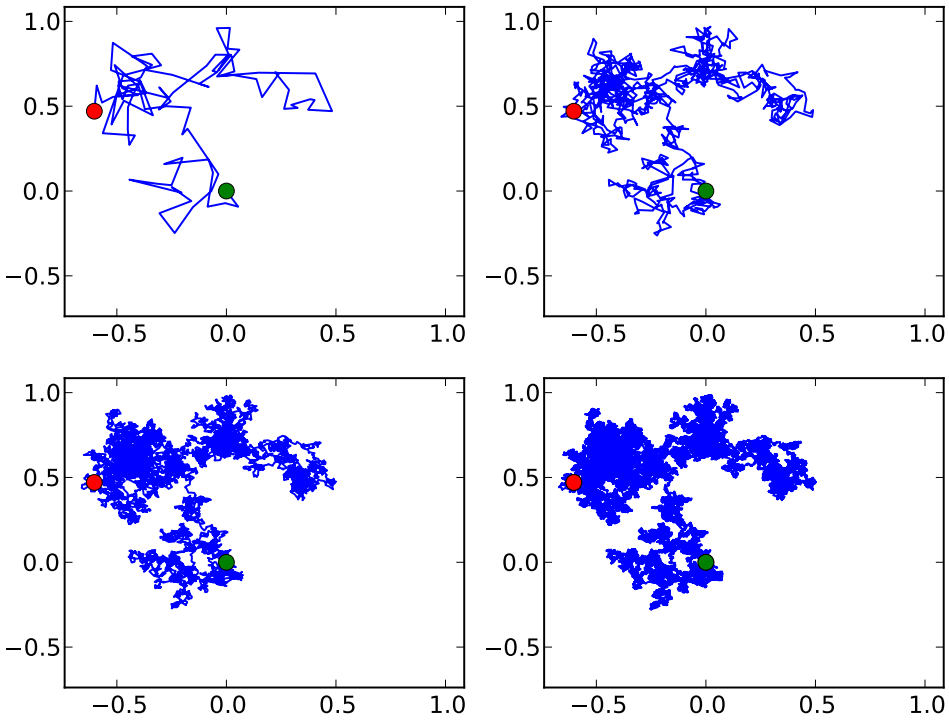

Figure 1: The piecewise linear interpolation between the points of a two dimensional Brownian motion started at $t = 0$ at $(0, 0)$ (green circle), stopped at $t = 1$ (red circle) and sampled at time steps of size $\{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$.

Stochastic integrals

The Gaussianity of Brownian increments implies $B_t - B_s \sim N\sqrt{t-s}$ for $N \sim \mathcal{N}(0, 1)$, hence we can at best expect at best a Hölder-modulus of $1/2$ and the problem of giving meaning to $\int_0^t W(y_r)\, dB_r$ appears. To see what goes wrong with Riemann–Stieltjes integrals consider integrating a one-dimensional Brownian trajectory against itself: with dyadic partitions of $[0, 1]$, $t_i^n =$

$i.2^{-n}$ a direct calculation shows that

$$\sum_i B_{\frac{t_{i+1}^n + t_i^n}{2}} (B_{t_{i+t}^n} - B_{t_i^n}) \text{ and } \sum_i B_{t_i^n} (B_{t_{i+1}^n} - B_{t_i^n}) \tag{2}$$

both converge (for a.e. Brownian trajectory $B(\omega)$) but to different limits. The difference of their limits equals $1/2$ times the $n \to \infty$ limit of

$$\sum_i (B_{t_{i+1}^n} - B_{t_i^n})^2 \tag{3}$$

and the quantity (3) is the so-called *quadratic variation process* $([B]_t)_{t \geq 0}$ of the Brownian motion B. Kyoshi Ito developed a powerful integration theory by generalizing the above limit construction (2). He gave meaning to $\int_0^t \phi_r dB_r$ for a large class of stochastic processes ϕ by taking the $L^2(\Omega)$-limit of $\sum_i \phi_{t_i^n} (B_{t_{i+1}^n} - B_{t_i^n})$ as $n \to \infty$. A crucial ingredient is that the integrand ϕ does not "look into the future evolution of B"[2]. For many applications like mathematical finance this is a desirable property. Instead of the right sum in in (2) one can also use the left sum, i.e. the mid-points $\phi_{(t_i^n + t_{i+1}^n)/2}$, to arrive at a different notion of stochastic integration called the Stratonovich integral, denoted $\int_0^t \phi_r \circ dB_r$. Above approaches to stochastic integration are not limited to Brownian motion and can be extended to the class of semi-martingales. Ito and his successors (especially the "Ecole de Strasbourg") a complete theory that gives existence and uniqueness for stochastic equations of the form (1); see [61, 41, 58, 57].

Ito's change of variable formula

Stochastic calculus is not a first order calculus: the change of variable formula, called "Ito's Lemma", reads as

$$f(t, B_t) = f(0, B_0) + \int_0^t \frac{\partial f}{\partial t}(r, B_r) dr + \int_0^t \frac{\partial f}{\partial x}(r, B_r) dB_r + \int_0^t \frac{\partial^2 f}{\partial x^2}(r, B_r) d[B]_r. \tag{4}$$

Replacing Ito with Stratonovich integration show a big advantage of the Stratonovich integral, namely that it follows a first order calculus,

$$f(t, B_t) = f(0, B_0) + \int_0^t \frac{\partial f}{\partial t}(r, B_r) dr + \int_0^t \frac{\partial f}{\partial x}(r, B_r) \circ dB_r. \tag{5}$$

This is only the starting point for one of the most exciting mathematical developments of the twentieth century and to make the above rigorous requires much more detail — we refer the interested reader to the many excellent introductory texts [59, 57, 39, 61].

A drawback: discontinuity of the solution map

While stochastic calculus had tremendous impact on theory and applications it has several shortcomings; two which are relevant for this article are that it is, firstly, limited to the class of semimartingales as noise (this for example excludes fractional Brownian motion) and secondly, that a very basic object, solution map associated with (1),

$$B \mapsto Y,$$

is not continuous. Over the last 20 years an approach to differential equations perturbed by irregular paths was developed and is usually referred to as the Over the last 20 years, Terry Lyons and collaborators [52, 50, 47, 30, 33] devolped a robust and completely analytic/algebraic approach to such differential equations; this is the so-called "theory of rough paths". It is not meant to replace stochastic calculus but it complements it where it runs into trouble; especially in view of splitting results this robustness becomes very useful and in the rest of this chapter we make heavy use of this theory.

[2] More precisely, the relevant property of the Brownian motion here is that B is a martingale. Geometrically, this is an orthogonality relation between the increments $B_t - B_s$ and the path up to time s. Hence, the construction works in a geometric $L^2(\Omega)$ sense which allows to take advantage of this structure.

Structure of this chapter

In Section 2 we introduce the main topic of this chapter, namely that splitting schemes can be derived from robustness of the solution map. In Section 3 we recall some key results from the the theory of rough paths which give a quantitative and qualitative understanding of the regularity of this solution map.
Splitting methods for S(P)DEs are naturally divided in strong and weak schemes. The goal of strong schemes is to approximate the solution Y of a S(P)DE (or a function of it, $f(Y)$) for a given realization of the noise. On the other hand, for many applications it is sufficient to only approximate the expected value $E[f(Y)]$. Strong approximations are discussed in Section 4 and applications to nonlinear filtering and optimal control are given in Section 5. In Sections 6 we discuss weak splitting schemes for S(P)DEs and their rate of convergence; we recall a popular weak approximation scheme called "cubature on Wiener space" and show that it has a natural interpretation as a splitting scheme. In Section 7 we present three applications of splitting schemes in financial engineering: efficient implementations for popular stochastic local volatility models [2]; a calibration of the Gatheral Double Mean Reverting model to market data [3]; and finally the Heath–Jarrow–Morton interest rate model [22].

Background

This chapter is inspired by a view on stochastic differential equations that emerged over the last 15 years, namely the theory of rough paths due to Terry Lyons and collaborators; for further developments and introduction see [52, 49, 50, 33, 30, 48]). This theory complements classic Ito-calculus and provides new, if not revolutionary insights, on how differential equations react to complex input signals. One of the earliest new applications was the so-called "cubature on Wiener space" of Kusuoka–Lyons–Victoir [51, 44]. Bayer, Dörsek, Teichmann among others [22, 64, 21, 5] then showed that these methods can be applied to the infinite-dimensional setting that is needed by SPDEs. More recently more applications were developed both in finite and in infinite dimensions (we survey some of these in Section 7). In a somewhat different direction, the work of Friz–Oberhauser [28] combined robustness from rough path theory with viscosity PDE methods to derive splitting schemes for strong approximations of (nonlinear) SPDEs.

Of course, splitting-up methods have appeared much earlier in stochastic calculus and we emphasize that these techniques remain highly relevant and form the basis of much of the recent developments that we present here. However, instead of giving a "horizontal" historical account we decided to give a "vertical" snapshot of what we believe are some exciting current developments in theory and applications. Unfortunately, this implies that we cannot do full justice to the existing rich literature. Nevertheless, we would like to point the reader to some classic articles as a starting point: one of the earliest motivation comes from the theory of nonlinear filtering we mention pars pro toto the work of Bensoussan and Glowinski [6] and Bensoussan, Glowinski and Răşcanu [7], Elliott and Glowinski [23], Florchinger and Le Gland [45, 26], Gyöngy and Krylov [34], Nagase [54], and Lototsky, Mikulevicius and Rozovskii [46]. The more general field of splitting is overwhelmingly large, so that we again cannot hope to give a balanced literature review. Some general works we want to mention are Jentzen and Kloeden [40], Debussche [19], Gyöngy and Krylov [35], Răşcanu and Tudor [60] Hausenblas [36] and, finally, Yan [65]. Let us finally stress that we consider partial differential equations driven by a temporal (possibly *also* spacial) noise, not partial differential equations with spacial noise, another very active research field in applied mathematics (see, for instance, Schwab and Gittelson [62]).

2 From robustness to splitting schemes

On an abstract level, we have to understand how the output path (the solution of a differential equation) of a complex system (a differential equation) responds to an input path (e.g. time and noise). In this section we show that if a continuous dependence between output and input signal holds, then splitting results follow immediately.

A toy example

Let us consider the simple example of a quantity y whose evolution over time is described by the differential equation

$$\frac{\mathrm{d}y_t}{\mathrm{d}t} = V(y_t) + W(y_t), \quad y_0 \in \mathbb{R}^e$$

where V, W are Lipschitz vector fields on $\mathbb{R}^e$. We identify this differential equation as a special case of the integral equation

$$y_t = \int_0^t V(y_r)\, da_r + \int_0^t W(y_t)\, db_r, \quad y_0 \in \mathbb{R}^e \tag{6}$$

where a and b are continuous, real-valued paths that are regular enough that above integrals have meaning. While finite 1-variation as recalled in Definition 2.1 is sufficient for the Riemann–Stieltjes integrals, we will treat paths having much less regularity in later parts of this chapter. Equations of type (6) are often called *controlled (differential/integral) equations* and a, b are referred to as the *controls* or also as the *driving paths/signals*. Such equations arise naturally in the engineering sciences and have been very well studied (see the seminal work of work of Brockett, Sussmann, Fliess, et al. [10, 25, 63]). We henceforth use the shorthand/differential notation

$$\mathrm{d}y_t = V(y_t)\, da_t + W(y_t)\, db_t, \quad y_0 \in \mathbb{R}^e \tag{7}$$

to denote (6). A basic question is the regularity of the solution map

$$(a, b) \mapsto y. \tag{8}$$

Obviously, the answer depends on what norms we use to measure distances between paths. What might be somewhat surprising is that the above mapping, defined on smooth paths

$$C^1\left([0,T], \mathbb{R}^2\right) \to C^1\left([0,T], \mathbb{R}^e\right),$$

is not even continuous under the usual uniform norm $|a|_\infty = \sup_{t\in[0,T]} |a_t|$; we invite the reader to find an example for this discontinuity and come back to this issue in detail in Example 2.5. Motivated by this, we introduce a cascade of metrics that are stronger than the uniform norm.

Definition 2.1. Let x be a continuous path defined on $[0,T]$ that takes values in a complete metric space (E,d). For every $p \geq 1$ the p-variation norm of x is defined as

$$|x|_{p-var} = \sup_{\substack{n\in\mathbb{N},(t_1,\dots,t_n):\\ 0\leq t_1<\dots<t_n\leq T}} \left(\sum_{i=1}^{n} d(x_{t_{i+1}}, x_{t_i})^p\right)^{1/p}$$

We denote the subset of $C([0,T], E)$ of paths finite p-variation norm by $C^{p-var}([0,T], E)$.

Standard arguments show that $\left(C^{p-var}([0,T], E), |.|_{p-var}\right)$ is a Banach space. We now see that the p-variation norm resolves the non-continuity of the uniform norm.

Theorem 2.2 (Robustness [30].)**.** *Let $V, W : \mathbb{R}^e \to \mathbb{R}^e$ be Lipschitz continuous and $(a,b) \in C^{1-var}\left([0,T], \mathbb{R}^2\right)$. Then there exists a unique solution $y \in C^{1-var}([0,T], \mathbb{R}^e)$ to the controlled differential equation*

$$\mathrm{d}y_t = V(y_t)\, \mathrm{d}a_t + W(y_t)\, \mathrm{d}b_t, \quad y_0 \in \mathbb{R}^d$$

and the map $(a,b) \mapsto y$ is continuous in 1-variation norm $|.|_{1-var}$.

Lie and Strang splitting

The connection with splitting is now immediate: Fix $\Delta > 0$ and divide $[0, T]$ into intervals of size $\Delta > 0$; further, denote $t_\Delta = \left\lfloor \frac{t}{\Delta} \right\rfloor \Delta$, $t^\Delta = t_\Delta + \Delta$ and define two time-changes (real-valued increasing paths) $\phi^\Delta, \varphi^\Delta$,

$$\phi_t^\Delta = \begin{cases} t_\Delta + 2\,(t - t_\Delta) & \text{, if } t \in \left[t_\Delta, \frac{t_\Delta + t^\Delta}{2}\right] \\ t^\Delta & \text{, if } t \in \left[\frac{t_\Delta + t^\Delta}{2}, t^\Delta\right] \end{cases}, \varphi_t^\Delta = \phi_{t+\frac{\Delta}{2}}^\Delta. \tag{9}$$

In other words, we approximate $t \mapsto (t, t)$ with $t \mapsto \left(\phi_t^\Delta, \varphi_t^\Delta\right)$ as $\Delta \to 0$, as depicted in Figure 2. Basic arguments show that this convergence holds in p-variation norm for every $p > 1$ (for $p = 1$ it is not true).

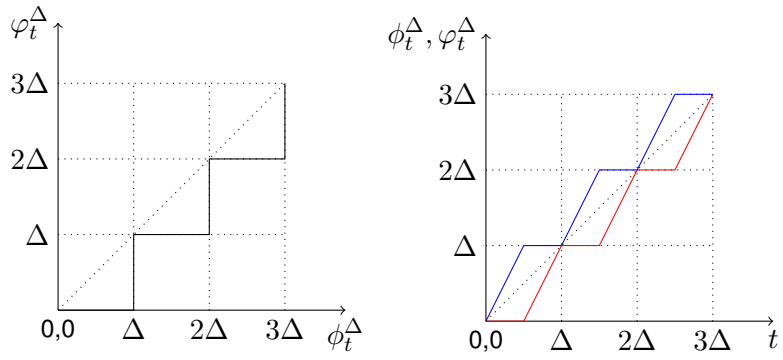

Figure 2: (Lie-Splitting) The two-dimensional path $t \mapsto \left(\phi^\Delta, \varphi^\Delta\right)$ approximates the identity $t \mapsto (t, t)$ and exactly one of $\frac{d\phi_t^\Delta}{dt}, \frac{d\varphi_t^\Delta}{dt}$ is 0 for any given time $t \geq 0$. This gives rise to the so-called Lie-splitting scheme.

This particular choice of control paths immediately implies a splitting result since by composition of a (resp. b) with ϕ^Δ (resp. φ^Δ) we flow at any moment in time either along V or along W. The above choice of a^Δ, b^Δ yields the classic Lie–splitting but obviously other choices are possible, for example we recover the Strang–Splitting scheme by using

$$\tilde{\phi}_t^\Delta = \phi_{t+\frac{\Delta}{4}}^\Delta \text{ and } \tilde{\varphi}_t^\Delta = \tilde{\phi}_{t+\frac{\Delta}{2}}^\Delta$$

To state it precisely, we introduce the notion of solution operators.

Definition 2.3. For every $y_0 \in \mathbb{R}^e$ denote by $\boldsymbol{P}_t^{\Delta,V} y_0$ the solution at time t of the controlled differential equation

$$\mathrm{d}y_t^\Delta = V\left(y_t^\Delta\right) \mathrm{d}(a \circ \phi^\Delta)_t$$

started at $y_0^\Delta = y_0$. Similarly denote by $\boldsymbol{Q}_t^{\Delta,W} y_0$ the solution at time t of the controlled differential equation

$$\mathrm{d}y_t^\Delta = W\left(y_t^\Delta\right) \mathrm{d}(b \circ \varphi^\Delta)_t$$

started at $y_0^\Delta = y_0$.

Corollary 2.4 (Splitting)**.** *We have* $\forall t > 0$

$$\lim_{\Delta \to 0} \left| \left(\boldsymbol{P}_\Delta^{\Delta,V} \boldsymbol{Q}_\Delta^{\Delta,W}\right)^{\lfloor t/\Delta \rfloor} y_0 - y_t \right| = 0$$

where y is the solution of the differential equation (6) started at time 0 with y_0. Moreover, the convergence even holds uniformly in t.

Proof. A simple calculation shows that the path $\left(a \circ \phi^\Delta, b \circ \varphi^\Delta\right)$ converges to the the path $t \mapsto (t, t)$ with uniform 1-variation bounds, i.e. $\sup_{\Delta>0} |a^\Delta|_{1-var} + \sup_{\Delta>0} |b^\Delta|_{1-var} < \infty$. The claim then follows from a slight variation of Theorem 2.2; see [28]. □

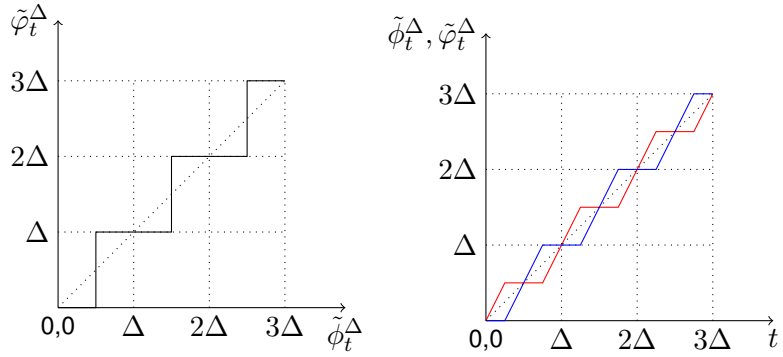

Figure 3: (Strang–Splitting) The two-dimensional path $t \mapsto \left(\tilde{\phi}^\Delta, \tilde{\varphi}^\Delta\right)$ approximates the identity $t \mapsto (t,t)$ better than $t \mapsto \left(\phi_t^\Delta, \varphi_t^\Delta\right)$ as depicted in Figure 2. Therefore it should be not surprising that Strang–splitting leads to better rates than Lie–splitting.

Highly oscillatory paths and the Lie brackets of vector fields

We now replace the path b in (6) by one with highly oscillatory trajectories (which are typically for many stochastic processes).

Example 2.5. Consider the sequence of paths $(a_t^n, b_t^n)_n = \left(\frac{1}{n}\cos 2\pi n^2 t, \frac{1}{n}\sin 2\pi n^2 t\right)$ and note that converges uniform norm as $n \to \infty$ to $\left(a^0, b^0\right) = (0,0)$. Define the vector fields $W(y^1,y^2,y^3) := (1,0,\frac{-y^2}{2})^T$ and $V(y^1,y^2,y^3) = (0,1,\frac{y^1}{2})^T$ and denote by y^Δ the solution of

$$\mathrm{d}y_t^\Delta = V\left(y_t^\Delta\right)\mathrm{d}a_t^\Delta + W\left(y_t^\Delta\right)\mathrm{d}b_t^\Delta \text{ with } y_0^\Delta = y_0 \in \mathbb{R}^3. \tag{10}$$

A simple calculation then shows that y_1^Δ does not converge as $\Delta \to 0$ to y_0 (the solution of (10) applied with $\Delta \equiv 0$).

In above Example 2.5,the hghly oscillatory motions of the driving signals affect the evolution of y not directly via V or W but via their Lie bracket $[V,W] = V\cdot W - W\cdot V$ which picks up the signed area (recall the Green/Stokes formula)

$$\frac{1}{2}\left(\int_0^\cdot a_r^n db_r^n - \int_0^\cdot b_r^n da_r^n\right)$$

swept out by (a^n, b^n). To sum up, the highly oscillatory behaviour of the driving signal leads to a subtle interplay between the iterated integrals of the driving signal and the Lie brackets of the involved vector fields that can destroy continuity of the solution map. However, was central to our derivation above of the Lie and Strang-splittings. In the next section we show how the theory of rough paths provide the needed continuity.

3 Rough path theory

We still have to give meaning to differential equations driven by non-smooth paths and study the properties of the associated solution map. Example 2.5 suggests that the iterated integrals

$$\int_{s<r_1<t} \mathrm{d}x_{r_1}, \int_{s<r_1<r_2<t} \mathrm{d}x_{r_1}\otimes \mathrm{d}x_{r_2}, \ldots, \int_{s<r_1<\cdots<r_n<t} \mathrm{d}x_{r_1}\otimes\cdots\otimes \mathrm{d}x_{r_n}.$$

of the driving signal x (resp. their linear combinations) play a special role when the path is highly oscillatory. These integrals will in general not make sense as Riemann–Stieltjes integrals if x is of unbounded variation. However, the theory of rough paths shows that it is enough to find a sequence of tensors that "behaves algebraically" like such a sequence of iterated integrals to derive the well-posedness of differential equations driven by this "iterated integrals".

The space of iterated integrals

The sequence of iterated integrals has a rich algebraic structure. Let us first give the definition for the case of a bounded variation path.

Definition 3.1. For every u, v such that $0 \leq u \leq v \leq T$ define $\Delta^1_{u,v} = \{(s,t) : u \leq s \leq t \leq v\}$. Let $x \in C^{1-var}\left([0,T], \mathbb{R}^d\right)$, $(s,t) \in \Delta_{0,T}$ and $k \in \mathbb{N}$. We define the iterated integrals $\int_{\Delta^k_{s,t}} \mathrm{d}x \otimes \cdots \otimes \mathrm{d}x \in \left(\mathbb{R}^d\right)^{\otimes k}$ recursively as

$$\int_{\Delta^1_{s,t}} \mathrm{d}x := x(t) - x(s) \text{ and } \int_{\Delta^k_{s,t}} \underbrace{\mathrm{d}x \otimes \cdots \otimes \mathrm{d}x}_{k \text{ times}} := \int_s^t \int_{\Delta^{k-1}_{s,r}} \underbrace{\mathrm{d}x \otimes \cdots \otimes \mathrm{d}x}_{(k-1) \text{ times}} \otimes \mathrm{d}x_r.$$

Recall that the space $\left(\mathbb{R}^d\right)^{\otimes k}$ used above is the space of k-tensors which has as basis $\left(e_{i_1} \otimes \cdots \otimes e_j\right)_{i,j \in \{1,\ldots,d\}}$.

Definition 3.2. Let $x \in C^{1-var}\left([0,T], \mathbb{R}^d\right)$ and $(s,t) \in \Delta^1_{0,T}$. The signature of x over $[s,t]$, denoted by $S(x)_{s,t}$, is the elementof $\bigoplus_{k=0}^{\infty} \left(\mathbb{R}^d\right)^{\otimes k}$ given as

$$S(x)_{s,t} = \left(1, \int_{\Delta^1_{s,t}} \mathrm{d}x, \int_{\Delta^2_{s,t}} \mathrm{d}x \otimes \mathrm{d}x, \ldots\right)$$

with the convention that $\left(\mathbb{R}^d\right)^{\otimes 0} = \{1\}$. Similarly, we define for $n \in \mathbb{N}$ the truncated signature of x over $[s,t]$, denoted $S^n(x)_{s,t}$, as the element of $\bigoplus_{k=0}^{n} \left(\mathbb{R}^d\right)^{\otimes k}$ given as

$$S^n(x)_{s,t} = \left(1, \int_{\Delta^1_{s,t}} \mathrm{d}x, \int_{\Delta^2_{s,t}} \mathrm{d}x \otimes \mathrm{d}x, \ldots, \int_{\Delta^n_{s,t}} \underbrace{\mathrm{d}x \otimes \cdots \otimes \mathrm{d}x}_{n \text{ times}}\right).$$

We call the path $t \mapsto S^n(x)_{0,t}$ the step-n lift of x.

The above definition is not efficient concerning the state space since it does not account for the recursive structure of $S^n(x)$ and we can hope to work with a much smaller subspace of $\bigoplus_{k=0}^{n} \left(\mathbb{R}^d\right)^{\otimes k}$. With slight abuse of notation denote with $\otimes : \bigoplus_{k=0}^{n} \left(\mathbb{R}^d\right)^{\otimes k} \to \bigoplus_{k=0}^{n} \left(\mathbb{R}^d\right)^{\otimes k}$ the natural extension of the tensor multiplication to the graded space $\bigoplus_{k=0}^{n} \left(\mathbb{R}^d\right)^{\otimes k}$, i.e. for

$$g = \sum_{k=0}^{n} \sum_{i_1,\ldots,i_k} g^{i_1 \ldots i_k} e_{i_1} \otimes \cdots \otimes e_{i_k}, h = \sum_{k=0}^{n} \sum_{i_1,\ldots,i_k} h^{i_1 \ldots i_k} e_{i_1} \otimes \cdots \otimes e_{i_k} \in \bigoplus_{k=0}^{n} \left(\mathbb{R}^d\right)^{\otimes k}$$

define

$$g \otimes h = \sum_{k=0}^{n} \sum_{l,m : l+m=k} g^{i_1 \ldots i_l} h^{i_1 \ldots i_m} e_{i_1} \otimes \cdots \otimes e_{i_k} \otimes e_{i_1} \otimes \cdots \otimes e_{i_m}.$$

We can now describe the algebraic structure of the subspace of $\bigoplus_{k=0}^{n} \left(\mathbb{R}^d\right)^{\otimes k}$ that contains the iterated integrals.

Theorem 3.3 ([30]). *For $n \geq 1$ and $d \geq 1$ define $G_{n,d} := \left\{S(x)_{0,1} : x \in C^{1-var}\left([0,T], \mathbb{R}^d\right)\right\}$. Then*

1. *$(G_{n,d}, \otimes)$ is a Lie group,*
2. *$G_{n,d} = \exp \mathfrak{g}_{n,d}$ where $(\mathfrak{g}_{n,d}, [\cdot,\cdot])$ is Lie algebra and*
3. *$\mathfrak{g}_{n,d} = \mathbb{R}^d \oplus \left[\mathbb{R}^d, \mathbb{R}^d\right] \oplus \cdots \oplus \left[\mathbb{R}^d, \left[\mathbb{R}^d, \left[\cdots, \left[\mathbb{R}^d, \mathbb{R}^d\right] \cdots\right]\right]\right]$*

We call $G_{n,d}$ the free step-n Lie group with d generators and $\mathfrak{g}_{n,d}$ the free step-n Lie algebra with d generators. The geodesic (so-called Carnot–Caratheodory) distance d_{CC} turns $(G_{n,d}, d_{CC})$ in a metric space.

(Weak) geometric rough paths

Since $(G_{n,d}, d_{CC})$ is a complete metric space, Definition 2.1 applies and we can speak of paths of bounded p-variation — this is exactly the definition of a weak geometric p-rough path.

Definition 3.4. Let $p \geq 1$ and $n = \lfloor p \rfloor$. We define the space of weak geometric p rough paths as

$$C^{p-var}([0,T], G_{n,d}) := \{\mathbf{x} \in C([0,T], G_{n,d}) : d_{p-var}(0, \mathbf{x}) < \infty\}$$

(here 0 denotes constant path that takes the value of the neutral element of the group $G_{n,d}$).

Example 3.5 (The Brownian rough path)**.** Let B be a two-dimensional Brownian motion. This gives rise to the $G_{2,2}$-valued path

$$\mathbf{B}_t = \underbrace{\left(1, B_t, \int_0^t dB \otimes dB\right)}_{\in G_{2,2}} = \exp \underbrace{\left(B_t^1 e_1 + B_t^2 e_2 + \frac{1}{2}\left(\int_0^t B_r^1 dB_r^2 - \int_0^t B_r^2 dB_r^1\right)(e_1 \otimes e_2 - e_2 \otimes e_1)\right)}_{\in \mathfrak{g}_{2,2}}$$

where the integrals are understood as (Stratonovich) stochastic integrals. One can show that $\mathbf{B} \in C^{p-var}([0,T], G_{2,2})$ for any $p > 2$, see [52, 30].

Differential equations driven by rough paths

Ito's approach to differential equations driven by highly oscillatory stochastic processes exploits the underlying probabilistic structure of the driving signal. Lyons [52, 49, 48] developed a different approach that relies only on analytic and algebraic methods; most important for us, it comes with a cascade of metrics which provide the needed continuity of the solution map.

Theorem 3.6 ("Universal Limit Theorem": Existence, uniqueness and continuity of RDEs; see [52, 30])**.** *Let $p \in (2,3)$, $d \geq 1$, $\mathbf{x} \in C^{p-var}([0,T], G_{2,d})$ and $V_i \in C_b^3(\mathbb{R}^e, \mathbb{R}^e)$. There exists a $y \in C^{p-var}([0,T], \mathbb{R}^e)$ such that for every sequence $(x^n)_n \subset C^{1-var}([0,T], \mathbb{R}^d)$ such that $d_{p-var}(S_2(x), \mathbf{x}) \to 0$, the solutions of the ODE*

$$dy_t^n = V(y_t^n) dx_t^n \equiv \sum_i V_i(y_t^n) d(x^n)_t^i$$

converge uniformly to y. We say that y is a solution of the RDE driven by $\mathbf{x}$ and write

$$dy_t = V(y_t) d\mathbf{x}_t.$$

The solution map is uniformly continuous on compact sets, that is for every $R > 0$ the map

$$\begin{array}{rcl} (y_0, \boldsymbol{x}) & \mapsto & y \\ \mathbb{R}^e \times \{d_{p-var}(\boldsymbol{x}, 0) < R\} & \to & C^{p-var}([0,T], G_{2,d}) \end{array}$$

is uniformly continuous in d_{p-var}-metric.

Summary

Rough path theory provides us with a machinery that allows to solve differential equations driven by non-smooth signals (like Brownian motion, semimartingales but also many other classes of noise that are not covered by stochastic calculus). As opposed to the classical Ito-theory it not only requires the trajectory as input to construct the solution but also the "iterated integrals" of the driving signal; to be precise, it requires a set of tensors that "behave like" classical Riemann–Stieltjes iterated integrals. Finding efficient state spaces for these "enhanced paths" required us to work with nonlinear spaces, i.e. Lie groups. In return we get a completely analytic/algebraic approach that provides the well-posedness of such differential equations, and the rough path theory comes with a cascade of metrics which makes the solution map continuous (the metric d_{p-var} for $p \geq 1$). This robustness is in stark contrast with Ito's theory and allows us to translate our splitting proof from the toy example in Section 2 to the case of S(P)DEs.

4 Strong splitting schemes for SPDEs

In this section we extend the splitting method to parabolic PDEs that evolve under the influence of noise. A large class of such stochastic partial differential equations (SPDEs) is of the form

$$\left\{ \begin{array}{rcll} \mathrm{d}u & = & F\left(t,x,u,Du,D^2u\right)\mathrm{d}t+\sum_{i=1}^{d}\Lambda_i\left(t,x,u,Du\right)\mathrm{d}z_t^i & \text{on } [0,T]\times\mathbb{R}^n \\ u\left(0,x\right) & = & u_0\left(x\right) & \text{on } \mathbb{R}^n \end{array} \right. \tag{11}$$

where $u=u\left(t,x\right)$ is scalar-valued, F denotes a nonlinear, (possibly degenerate) elliptic differential operator, Λ is affine linear in (u,Du) and $z\in C\left(\left[0,T\right],\mathbb{R}^d\right)$ is a multidimensional path with the same (or worse) regularity properties as Brownian trajectories.

Several issues appear: firstly, even if $\Lambda\equiv 0$, then the nonlinearity of F implies that we cannot hope for a smooth solution $u\in C^{1,2}\left([0,T]\times\mathbb{R}^n,\mathbb{R}\right)$. Therefore we have to work with a suitable concept of generalized solutions. Secondly, the path z is not differentiable and similar to our toy example, we have to give appropriate meaning to $\Lambda\left(t,x,u,Du\right)\circ\mathrm{d}z_t$. Put simply, we solve the first problem by working with the theory of viscosity solutions and the second problem with the theory of rough paths.

Approximating time

As in our toy example in Section 2, we now want to look at equation (11) as a special case of

$$\left\{ \begin{array}{rcl} \mathrm{d}u^{\Delta} & = & F\left(t,x,u^{\Delta},Du^{\Delta},D^2u^{\Delta}\right)\mathrm{d}(a\circ\phi^{\Delta})_t+\sum_{i=1}^{d}\Lambda\left(t,x,u^{\Delta},Du^{\Delta}\right)\circ\mathrm{d}(z^i\circ\varphi^{\Delta})_t, \\ u^{\Delta}\left(0,x\right) & = & u_0\left(x\right). \end{array} \right. \tag{12}$$

However, the situation is more subtle.

Example 4.1. Consider $n=1$, $F\left(t,x,u,Du,D^2u\right)=D^2u$ and $\Lambda\equiv 0$ in which the above reduces to the one-dimensional heat equation: $\mathrm{d}u^{\Delta}=D^2u^{\Delta}\mathrm{d}(a\circ\phi^{\Delta})_t$. Then one cannot hope for continuity of $(a,z)\mapsto u$ since this requires to give meaning to the heat equation when $\frac{\mathrm{d}(a\circ\phi^{\Delta})_t}{\mathrm{d}t}<0$, i.e. when time is run backwards which is in general not well-posed.

We simply resolve above issue by replacing C^{1-var} by a smaller class of paths.

Proposition 4.2 ([28])**.** *Define*

$$C_0^{1,+}\left([0,T],\mathbb{R}\right)=\left\{\xi\in C^1\left([0,T],\mathbb{R}\right):\xi_T=T,\dot{\xi}_t>0\ \forall t\right\}$$

and its closure $C_0^{1-var,+}\left([0,T],\mathbb{R}\right):=\overline{C_0^{1,+}\left([0,T],\mathbb{R}\right)}^{|\cdot|_\infty}$ *where* $|a|_\infty\equiv\sup_{t\in[0,T]}|a_t|$. *Then*

$$C_0^{1-var,+}\left([0,T],\mathbb{R}\right)=\Big\{\xi\in C_0\left([0,T],\mathbb{R}\right):\xi_T=T \text{ and } \exists\xi^{sing}\in C^{1-var}\left([0,T],\mathbb{R}_{\geq 0}\right),\xi^{sing}=0 \text{ a.s. and } \xi_t=\xi_t^{sing}+\int_0^t\xi_r^{cont}dr\Big\}$$

and $C_0^{1-var,+}\left([0,T],\mathbb{R}\right)\subsetneq C_0^{1-var}\left([0,T],\mathbb{R}\right)$.

Viscosity solutions of PDEs

Given a map

$$F:[0,T]\times\mathbb{R}^n\times\mathbb{R}\times\mathbb{R}^n\times\mathbb{S}^n\to\mathbb{R}$$

(with $\mathbb{S}^n$ denoting the set of symmetric $(n\times n)$-matrices) that is *proper* in the sense that

$$\begin{array}{rcl} F\left(t,x,r,p,A\right) & \leq & F\left(t,x,r,A+B\right)\ \forall A\in\mathbb{S}^n \text{ and } B\geq 0 \\ r & \mapsto & F\left(t,x,r,A\right) \text{ is increasing,} \end{array}$$

then the theory of viscosity solutions provides well-posedness for parabolic PDEs of the form

$$\begin{cases} \partial_t u - F\left(t,x,u,Du,D^2u\right) &= 0 & \text{on } [0,T]\times\mathbb{R}^n, \\ u\left(0,x\right) &= u_0\left(x\right) & \text{on } \mathbb{R}^n. \end{cases} \tag{13}$$

More precisely, if $u : [0,T]\times\mathbb{R}^n \to \mathbb{R}$ is bounded and uniformly continuous then we call u a *subsolution* of the PDE (13) if for every $\varphi \in C^{1,2}\left([0,T]\times\mathbb{R}^n,\mathbb{R}\right)$ it holds that whenever $\left(\hat{t},\hat{x}\right)$ is a local maximum of

$$(t,x) \mapsto u\left(t,x\right) - \varphi\left(t,x\right)$$

then

$$\partial_t\varphi\left(\hat{t},\hat{x}\right) - F\left(\hat{t},\hat{x},\varphi,D\varphi,D^2\varphi\right) \leq 0. \tag{14}$$

Similarly, we define *supersolutions* and call u a solution if it is a sub- and supersolution. Viscosity theory provides *comparison results*, that is given a subsolution v and a supersolution w of (13) this guarantees that

$$v \leq w$$

(note that this immediately implies uniqueness of solutions), see [14, 24].

Robustness for (nonlinear) SPDEs

We can now have an educated guess of a good solution concept for the nonlinear SPDE (11): let us approximate $t \mapsto (t, z_t)$ with a sequence $\left(\xi^n, z^n\right)_n \subset C^1\left([0,T],\mathbb{R}^d\right)$ of smooth paths. Then for every fixed $n \in \mathbb{N}$ we can speak of a viscosity solution $u^n \in \mathrm{BUC}\left([0,T]\times\mathbb{R}^n\right)$ — the space of real-valued, bounded and uniformly continuous functions — of

$$\begin{cases} \mathrm{d}u^n &= F\left(t,x,u^n,Du^n,D^2u^n\right)\mathrm{d}\xi^n + \sum_{i=1}^d \Lambda\left(t,x,u^n,Du^n\right)\mathrm{d}z_t^{n;i} & \text{on } [0,T]\times\mathbb{R}^n, \\ u^n\left(0,x\right) &= u_0\left(x\right) & \text{on } \mathbb{R}^n. \end{cases} \tag{15}$$

We expect that $\left(u^n\right)_n$ converges to a function $u \in \mathrm{BUC}\left([0,T]\times\mathbb{R}^n\right)$ as $n\to\infty$ and it is natural to identify this function u as the solution of the SPDE (11). It turns out that it is natural to define convergence of the sequence $\left(\xi^n,z^n\right)_n$ to $(t,\mathbf{z}_t)$ if

$$\begin{aligned} \sup_n \left\|S(z^n)\right\|_{p-var;[0,T]} + \sup_n \left|\xi^n\right|_{1-var} &< \infty \\ d_0\left(\boldsymbol{z},S(z^n)\right) + \left|\xi^n_t - t\right|_\infty &\to_n 0 \end{aligned} \tag{16}$$

holds. Here $d_0(\mathbf{x},\mathbf{y}) \equiv \sup\sum_{t_i} d_{CC}(\mathbf{x}_{t_i,t_{i+1}},\mathbf{y}_{t_i,t_{i+1}})$ where the $\sup$ is taken over all partitions (t_i) and we use the notation $\mathbf{x}_{t_i,t_{i+1}} \equiv \mathbf{x}_{t_i}^{-1}\mathbf{x}_{t_{i+1}}$ for increments in the group. Let us take this as definition of a solution.

Definition 4.3. Let $\boldsymbol{z} \in C_0^{0,p-var}\left([0,T],G_{[p],d}\right)$,$\xi \in C_0^{1-var,+}\left([0,T],\mathbb{R}\right)$. Let

$$\left(z^n,\xi^n\right)_n \subset C_0^{0,p-var}\left([0,T],G_{[p],d}\right)\times C_0^{1-var,+}\left([0,T],\mathbb{R}\right)$$

be a sequence that converges to $(t,\boldsymbol{z}_t)$ in the sense of (16) and assume that there exists for every n a unique viscosity solution u^n of the PDE (15). We call every accumulation point (in the metric of uniform convergence on compacts) of $\left(u^n\right)$ a solution of the RPDE

$$\begin{cases} \mathrm{d}u &= F\left(t,x,u,Du,D^2u\right)\mathrm{d}\xi_t + \Lambda\left(t,x,u,Du\right)\circ\mathrm{d}\boldsymbol{z}_t & \text{on } [0,T]\times\mathbb{R}^n, \\ u\left(0,x\right) &= u_0\left(x\right) & \text{on } \mathbb{R}^n. \end{cases} \tag{17}$$

If this limit is unique and does not depend on the choice of the approximating sequence $\left(\xi^n,z^n\right)_n$ and the solution map

$$(\xi,\boldsymbol{z}) \mapsto u$$

is continuous then we say that (17) is robust in the rough path sense.

It is clear that the above robustness in rough path sense immediately gives a splitting result when use the time changes $(\phi^{1/n}, \varphi^{1/n})$ from Section 2 to define the approximating sequence

$$\left(\xi \circ \phi^{1/n}, \boldsymbol{z} \circ \varphi^{1/n}\right)_n .$$

In Section 5 below we show that large classes of SPDEs are robust in rough path sense as defined above.

5 Applications of strong schemes to nonlinear filtering and optimal control

Nonlinear filtering

In many areas of science, the quantities of interest are not available for direct measurement. Fortunately, we can make reasonable inferences about them by combining mathematical models that describe their evolution with partial observations of these quantities. These partial observations are typically corrupted by noise and we need to account for this. Applications range from cryptography, tracking and guidance, the study of the global climate, to the management of risk in a economic context (see for example [15, 9, 27, 32]). Consider a Markov process (X, Y) that takes its values in $\mathbb{R}^{d_{sig}+d_{obs}}$ with its dynamics given by

$$\left\{ \begin{array}{rcll} \mathrm{d}X_t & = & \mu\left(X_t\right)\mathrm{d}t + \sigma\left(X_t\right)\mathrm{d}B_t & \text{(signal)}, \\ \mathrm{d}Y_t & = & h\left(X_t\right)\mathrm{d}t + \mathrm{d}\tilde{B}_t & \text{(observation)}. \end{array} \right.$$

Here, B and $\tilde{B}$ are multidimensional Brownian motions that are defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The goal is to compute for a given real-valued function f the conditional expectation

$$\pi_t f \equiv \mathbb{E}\left[f\left(X_t\right)|\mathcal{Y}_t\right],$$

i.e. to find the best estimate for $f\left(X_t\right)$ given the observation σ-algebra[3] $\mathcal{Y}_t = \sigma\left(\{Y_r, r \in [0,t]\}\right) \vee \mathcal{N}$ with $\mathcal{N}$ denoting the $\mathbb{P}$-null-sets. From basic principles it follows that there exists a measurable map $\phi_t^f : C\left([0,t], \mathbb{R}^{d_Y}\right) \to \mathbb{R}$ such that

$$\phi_t^f\left(Y|_{[0,t]}\right) = \pi_t f \;\; \mathbb{P} - \text{a.s.} \tag{18}$$

and our problem reduces to effectively calculate this functional ϕ_t^f.

Clark's robustness problem

In practice, only a finite number of observations $(Y_{t_i})_i$ of Y is available and we evaluate ϕ_t^f along some continuous interpolation of these points, $Y^{\text{interpolated}}$. Of course we expect that

$$\phi_t^f\left(Y^{\text{interpolated}}|_{[0,t]}\right) \simeq \phi_t^f\left(Y|_{[0,t]}\right)$$

but this is not guaranteed by (18), as the interpolation is a path of bounded variation, hence a null-set under the Wiener measure or any equivalent measure, see [16] for a detailed discussion. Clark [12] sketched a proof (a rigorous argument was given later by Clark and Crisan [13]) that if B and $\tilde{B}$ are uncorrelated, then there exists a functional $\phi_t^{f,\text{robust}}$ that is continuous in supremum norm and fulfills (18). In the correlated case, there cannot exist such a functional but recently (see [16]), it was shown that also in the correlated case there exists a functional $\phi_t^{f,\text{robust}}$ defined on the space of rough paths such that

$$\phi_t^{f,\text{robust}}\left(\boldsymbol{Y}\right) = \pi_t f \;\; \mathbb{P} - \text{a.s and } \boldsymbol{Y} \mapsto \phi_t^{f,\text{robust}}\left(\boldsymbol{Y}\right)$$

is continuous in rough path metric. This solves Clark's robustness problem (for semimartingale piecewise linear approximations converge in the appropiate rough path metric).

[3] There are some subtle measure-theoretic issues which we gloss over but refer the reader to [1] for more details.

The Kallianpur–Striebel and Zakai equations

Theorem 5.1 (Kallianpur–Striebel). *There exists a probability measure $\tilde{\mathbb{P}}$ on $(\Omega, \mathcal{F})$ such that*

1. *$\tilde{\mathbb{P}}$ is equivalent to $\mathbb{P}$,*
2. *$\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}}|_{\mathcal{F}_t} = \exp\left(-\int_0^t h(X_s)\cdot \mathrm{d}B_s - \frac{1}{2}\int_0^t |h(X_s)|^2\,\mathrm{d}s\right)$,*
3. *the observation process Y is a Brownian motion under $\tilde{\mathbb{P}}$,*
4. *for every $f \in B\left(\mathbb{R}^{d_{sig}}\right)$—the space of real-valued, bounded, measurable functions on $\mathbb{R}^d$—and every fixed $t > 0$*

$$\pi_t f = \frac{\tilde{\mathbb{E}}\left[f(X_t)\exp\left(\int_0^t h(X_s)\cdot \mathrm{d}Y_s - \frac{1}{2}\int_0^t |h(X_s)|^2\,\mathrm{d}s\right)|\mathcal{Y}_t\right]}{\tilde{\mathbb{E}}\left[\exp\left(\int_0^t h(X_s)\cdot \mathrm{d}Y_s - \frac{1}{2}\int_0^t |h(X_s)|^2\,\mathrm{d}s\right)\right]} \quad \mathbb{P} \text{ and } \tilde{\mathbb{P}} \text{ a.s.}$$

Proof. This can be found in every text book on nonlinear filtering; see for example [1, 15]. □

It turns out that it is advantageous to work with an unnormalized version of the inferred probability measure π. Indeed, if we define for every $f \in B\left(\mathbb{R}^{d_{sig}}\right)$

$$\rho_t f = \pi_t f \cdot \tilde{\mathbb{E}}\left[\exp\left(\int_0^t h(X_s)\cdot \mathrm{d}Y_s - \frac{1}{2}\int_0^t |h(X_s)|^2\,\mathrm{d}s\right)\right]$$

then obviously $\pi_t f = \frac{\rho_t f}{\rho_t 1}$. The Fokker–Planck/Kolmogorov forward equation is a PDE given by the generator of X that describes the time evolution of the density of the diffusion X via a parabolic PDE with the elliptic differential operator

$$A = \sum_{i,j}(\sigma^T\cdot\sigma)^{i,j}\frac{\partial^2}{\partial x_i \partial x_j} + \sum_i \mu^i \frac{\partial}{\partial x_i}.$$

The Zakai equation can be seen an extensionthat incorporates the additional information we get from the observation process Y. Indeed, set $h \equiv 0$ in Theorem below to recover the Fokker–Planck equation.

Theorem 5.2 (The Zakai SDE; uncorrelated case). *Under standard assumptions[4] we have $\tilde{\mathbb{P}}$-a.s. for every $t \geq 0$ and every $f \in B\left(\mathbb{R}^{d_{sig}}\right)$ that*

$$(\rho_t f) = \pi_0 f + \int_0^t \rho_s(Af)\,\mathrm{d}s + \int_0^t \rho_s\left(fh^T\right)\mathrm{d}Y_s.$$

Proof. See for example [1, Chapter 3]. □

The above applies to the case when B and $\tilde{B}$ are uncorrelated. In the correlated case a slight variation of the above Zakai SDE holds (an extra differential operator appears in the stochastic integral against Y).

Splitting for the Zakai SPDE

It is advantageous to work with densities instead of measures. Indeed, under well-known conditions ρ has a density u and we can write

$$\rho_t(A) = \int_A u(t,x)\,\mathrm{d}x$$

[4] For example $E\left[\int_0^t |h(X_s)|^2 ds\right] < \infty$, $E\left[\int_0^t Z_s |h(X_s)|^2 ds\right] < \infty$ and $\tilde{\mathbb{P}}\left[\int_0^t [\rho_s(|h|)]^2 ds < \infty\right] = 1$ is sufficient where $Z_s = \exp\left(-\sum_i \int_0^s h^i(X_r) dB_r^i - \frac{1}{2}\int_0^s h^i(X_r)^2 dr\right)$; see [1, Chapter 3]

for some $u \in \mathrm{BUC}([0,T]\times\mathbb{R}^n)$. In this case we can rewrite the above (infinite-dimensional) Zakai SDE from Theorem 5.2 for the unnormalized measure ρ as a SPDE for the density u. Since the generator of the signal X is linear, second order parabolic it is not surprising that the resulting SPDE will be linear (with linear noise). In fact, our setup is more general than needed by the nonlinear filtering application and below we treat general semi-linear PDEs (of which the SPDE for the density u is a special case).

Assumption 1. Let

$$L(t,x,r,p,M) = \mathrm{Tr}\left[M(x)\cdot X\right] + b(x)\cdot p + f(x,r)$$

with $M(x) = \sigma(x)\sigma^T(x)$, $\sigma:\mathbb{R}^n\to\mathbb{R}^{n\times m}$ and $b:\mathbb{R}^e\to\mathbb{R}^e$ bounded, Lipschitz in x. Further, let $f:\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}$ be continuous, bounded whenever r is bounded and with a lower Lipschitz bound, i.e.

$$f(x,r) - f(x,s) \geq c(r-s) \;\; \forall r\geq s, x\in\mathbb{R}^n.$$

Assumption 2. Let

$$\Lambda(t,x,r,p) = p\cdot\sigma_k(t,x) + r\cdot\nu_k(t,x) + g_k(t,x)$$

where σ,ν and g are Lip^γ for $\gamma > p+2$.

Theorem 5.3 (Well-posedness of linear RPDEs)**.** *Let $\boldsymbol{z}\in C^{0,p-var}\left([0,T],\mathbb{R}^d\right)$ and let L and Λ fulfill assumption (1) resp. (2). Then*

$$\begin{cases} \mathrm{d}u &= L\left(t,x,u,Du,D^2u\right)\mathrm{d}t + \sum_{i=1}^d \Lambda(t,x,u,Du)\circ\mathrm{d}\boldsymbol{z}_t^i & \text{on } [0,T]\times\mathbb{R}^n,\\ u(0,x) &= u_0(x) & \text{on } \mathbb{R}^n. \end{cases} \tag{19}$$

is robust in rough path sense.

Proof. We only sketch the idea of the proof for the case

$$\begin{cases} \mathrm{d}u &= \sigma^2(t,x)D^2u\mathrm{d}t + \sum_{i=1}^d V_i(x)\,Du\circ\mathrm{d}\boldsymbol{z}_t^i & \text{on } [0,T]\times\mathbb{R}^n,\\ u(0,x) &= u_0(x) & \text{on } \mathbb{R}^n. \end{cases} \tag{20}$$

First assume that z is a smooth path and denote by ϕ the ODE flow

$$\mathrm{d}\phi(t,x) = V(\phi(t,x))\mathrm{d}z_t, \phi(0,x) = x\in\mathbb{R}^n. \tag{21}$$

Then (at least formally) we see that the function $v(t,x) := u(t,\phi(t,x))$ solves the standard parabolic heat equation

$$\begin{cases} \mathrm{d}v &= \sigma_\phi^2(t,x)D^2v\mathrm{d}t & \text{on } [0,T]\times\mathbb{R}^n,\\ v(0,x) &= u_0(x) & \text{on } \mathbb{R}^n, \end{cases} \quad \text{where } \sigma_\phi^2(t,x) := \sigma^2(t,\phi(t,x)). \tag{22}$$

An obvious idea for the case that z is no longer a smooth path is to approximate z by a sequence of smooth paths (z^n). For each fixed $n\in\mathbb{N}$ one can solve the ODE flow ϕ^n (the ODE (21) with z replaced by z^n) and subsequently the corresponding simple PDE (22) to arrive at the sequence of PDE solutions (v^n). Since the flow ϕ^n will be a diffeomorphism we also know that

$$u^n(t,x) = v^n(t,(\phi_t^n)^{-1}(x)) \tag{23}$$

where u^n denotes the solution of (20) where the driving signal z is replaced by z^n. Obviously we expect that $(v^n)_n$ as well as $(\phi^n)_n$ converge as $n\to\infty$: for $(v^n)_n$ this should follow from the robust approximations of operators from viscosity theory and for $(\phi^n)_n$ this should follow if we consider convergence in rough path metric—recall Section 2 and 3 on the problems caused by highly oscillatory driving signals z. If this holds, then (23) implies that $(u^n)_n$ converges to some u and this function is a natural candidate for a solution. Of course, all the above was completely formal and the convergence can go wrong. However, with more care it can be made rigorous even for fully nonlinear operators; for the detailed argument see [11, 29]. □

Corollary 5.4 (Splitting for the Zakai SPDE)**.** *Denote with* $\{\boldsymbol{P}_t, t \geq 0\}$ *the solution operator*

$$\varphi \mapsto v \text{ where } v \text{ is the viscosity solution of } \mathrm{d}v = L\left(t, x, Dv, D^2 v\right) \mathrm{d}t,\ v\left(0, \cdot\right) = \varphi\left(\cdot\right)$$

and with $\left\{\boldsymbol{Q}_{s,t}, 0 \leq s \leq t\right\}$ *the solution operator*

$$\varphi \mapsto v \text{ where } v \text{ is the SDE solution of } \mathrm{d}y = \Lambda\left(t, x, Dv\right) \circ \mathrm{d}B_t,\ y\left(0, \cdot\right) = \varphi\left(\cdot\right).$$

Then for a.e. ω

$$u^n\left(t, x\right) := \prod_{i=0}^{\lfloor t/n \rfloor - 1} \left[\boldsymbol{Q}_{i/n, i/n+1/n} \circ \boldsymbol{P}_{1/n}\right]\left(u_0\left(x\right)\right)$$

converges locally uniformly (in (t, x)*) as* $n \to \infty$ *to the unique solution* u *of (19) given by Theorem 5.3 with* $\boldsymbol{z}_t = \boldsymbol{B}_t\left(\omega\right) \equiv \left(1, B_t\left(\omega\right), \left(\int_0^t B \otimes \circ \mathrm{d}B\right)\left(\omega\right)\right)$.

Pathwise optimal control

Consider the SDE

$$\mathrm{d}X_t = a\left(X_t, \alpha_t\right) \mathrm{d}t + b\left(X_t, \alpha_t\right) \circ \mathrm{d}B_t + c\left(X_t\right) \circ \mathrm{d}\tilde{B}_t$$

where $t \mapsto \alpha_t$ is a path, B and $\tilde{B}$ are multi-dimensional, independent Brownian motions and (a, b, c) are (sufficiently regular) vector fields. In applications (engineering, economics, etc.) one often faces the problem that one can influence the evolution of X by controlling the path α. The aim is then to minimize a cost function (consisting of a terminal cost g and a running cost f) of the form

$$v\left(t, x\right) = \inf_{\alpha} \mathbb{E}\left[g\left(X_T^{t,x}\right) + \int_t^T f\left(X_s^{t,x}, \alpha_s\right) \mathrm{d}s | \tilde{B}\right]. \tag{24}$$

It turns out that we can use the Bellman principle to describe the change in the cost function over time by a SPDE, the so-called Hamilton–Jacobi–Bellman (S)PDE. Indeed, a formal computation (see [20] for a rigorous derivation from basic principles) shows that after the time reversal $u\left(t, x\right) := v\left(T - t, x\right)$, we get a SPDE of the form

$$\begin{cases} \mathrm{d}u|_{t,x} + \inf_\alpha \left[b\left(x, \alpha_t\right) Du|_{t,x} + L_\alpha u|_{t,x} + f\left(x, \alpha_t\right)\right] \mathrm{d}t + Du|_{t,x} \cdot c\left(x\right) \circ \mathrm{d}\tilde{B}_t &= 0 & \text{on } [0,T] \times \mathbb{R}^n, \\ u\left(0, x\right) &= g\left(x\right) & \text{on } \mathbb{R}^n, \end{cases} \tag{25}$$

where L_α is the linear differential operator with (a, b). Using

Corollary 5.5. *Let* $\boldsymbol{z} \in C^{0,p-var}\left(\left[0,T\right], \mathbb{R}^d\right)$. *The SPDE* (25) *is robust in rough path sense.*

The proof is a slight modification of the proof of Theorem 5.3 since the usual comparison results from viscosity theory is stable under taking $\inf_\alpha$.

6 Weak splitting schemes for SPDEs

In the previous sections we have concentrated on strong approximation of (partial) differential equations driven by random signals, i.e., on the approximation of the solution $y_T = y_T(\omega)$ of the rough or stochastic (partial) differential equation as a random variable, ω-for-ω (resp. rough path by rough path). However, in many applications one is only interested in the law of the solution y_T of the equation. Indeed, if the quantity of interest is just the expectation of a functional of the solution, say

$$E\left[f(y_T)\right],$$

then it is sufficient to only approximate the law of y_T. This corresponds to the notion of *weak convergence* of random variables, and hence schemes for approximating the law of the solution of a

stochastic (partial) differential equations are referred to as *weak schemes*. More precisely, let us consider the solution y_T of a stochastic differential equation defined on the Banach space X (which is infinite-dimensional in the case of an SPDE) and let us consider a sequence of approximations $\overline{y}_N$ taking values in X indexed by $N \in \mathbb{N}$. Fix a space of sufficiently regular test functions $f : X \to \mathbb{R}$ (classically chosen to be $C_b(X)$ in theoretical probability theory, but more flexibility is needed in numerics). Then we say that $\overline{y}_N$ converges to y_T in the weak sense if for any test function f we have

$$E\left[f(\overline{y}_N)\right] \xrightarrow{N\to\infty} E\left[f(y_T)\right].$$

In particular, note that weak schemes — unlike strong ones — do not have to operate on the same probability space as the true solution.

Of course, if the space of test functions is a subspace of the Lipschitz continuous functions, then strong convergence (i.e., convergence in $L^1(\Omega; X)$) implies weak convergence, and the rate of weak convergence is at least as good as the rate of strong convergence. However, in many cases the weak rate of convergence is, in fact, much better than the strong one.

Cubature on Wiener space

For simplicity, let us concentrate on the finite-dimensional case first — we will come back to the infinite-dimensional (SPDE) setting at the end of this section. Consider the stochastic differential equation

$$dy_t = V_0(y_t)dt + \sum_{i=1}^{d} V_i(y_t) \circ dB_t^i, \tag{26}$$

with $y_0 \in \mathbb{R}^e$ fixed, B denoting a d-dimensional standard Brownian motion and "$\circ$" indicating that the stochastic integral is understood in the Stratonovich sense. We furthermore introduce the notation $B_t^0 \equiv t$ to simplify the presentation.

Assumption 3. We assume that the vector fields $V_0, \dots, V_d : \mathbb{R}^e \to \mathbb{R}^e$ are C^∞-bounded, i.e., they are smooth and all the derivatives are bounded (but not necessarily the functions themselves). Moreover, the test function f is smooth and bounded.

Remark 6.1. Of course, these assumptions can be relaxed. For instance, the boundedness requirements can be removed by working with properly weighted norms [21, 43]. Moreover, assuming a (hypo-)ellipticity condition for the vector fields, we can actually rely on the smoothing property of the diffusion equation and drop the smoothness assumption for the test function f – at the cost of possibly having to work with non-uniform grids, see [51].

In order to derive appropriate weak splitting schemes for the equation (26), we first recall the short time behaviour of the solution using the *stochastic Taylor expansion*, see for instance [42]. By iterating the Ito-formula for the Stratonovich-SDE (26) m times, we obtain

$$f(y_t) = f(y_0) + \sum_{k=1}^{m} \sum_{(i_1,\dots,i_k)\in\{0,\dots,d\}^k} V_{i_1} \cdots V_{i_k} f(y_0) \int_{0<t_1<\cdots<t_k<t} \circ dB_{t_1}^{i_1} \cdots \circ dB_{t_k}^{i_k} + \mathcal{O}\left(t^{(m+1)/2}\right), \tag{27}$$

where we iteratively use the geometrical notion $Vf(x) \equiv \nabla f(x) \cdot V(x)$ for a function f and a vector field V. We also denote

$$B_t^I = B_t^{(i_1,\dots,i_k)} \equiv \int_{0<t_1<\cdots<t_k<t} \circ dB_{t_1}^{i_1} \cdots \circ dB_{t_k}^{i_k}, \quad I = (i_1, \dots, i_k) \in \{0, \dots, d\}^k. \tag{28}$$

Remark 6.2. We once again see that the short-time behaviour of the solution y is controlled by the truncated signature.

Remark 6.3. As a matter of fact, sharper versions of (27) are possible, in so far that (27) ignores the different scaling of $t = B_t^0$ and $B_t^1, \dots, B_t^d$. Once again, we refer to [51].

Remark 6.4. Of course, analogous stochastic Taylor expansions can also be formulated in terms of the Ito integral, which would then lead to the Ito-signature. We prefer the geometrically more intuitive Stratonovich versions in this chapter.

This motivates the following methodology for constructing higher order weak approximation schemes termed the *ODE method* (originally introduced as *cubature on Wiener space* by [51] and, independently, [44]).

Theorem 6.5. *In the setting of Assumption 3, we are given a time-grid* $0 = t_0 < t_1 < \cdots < t_N = T$ *with corresponding increments* Δt_i, $i = 1, \dots, N$. *Let* $W_i : [0, \Delta t_i] \to \mathbb{R}^{d+1}$ *be a* $(d+1)$*-dimensional path of bounded variation satisfying*

$$\forall 0 \le k \le m,\ I \in \{0, \dots, d\}^k :\ E\left[B^I_{\Delta t_i}\right] = E\left[W^I_i(\Delta t_i)\right].$$

Moreover, let $W : [0, T] \to \mathbb{R}^{d+1}$ *be the bounded-variation process obtained by concatenating the processes* $W_1, \dots, W_N$. *Finally, let* $\overline{y}_N \equiv y_T(W)$ *be defined as the solution of the* ODE

$$\frac{dy(W)_t}{dt} = V_0(y(W)_t)\dot{W}^0_t + \sum_{i=1}^{d} V_i(y(W)_t)\dot{W}^i_t \tag{29}$$

formally obtained from (26) by replacing B *by* W. *Then there is a constant* $C > 0$ *such that*

$$|E[f(y_T)] - E[f(y(W)_T]| \le C\left(\max_{i=1,\dots,N} \Delta t_i\right)^{(m-1)/2}.$$

Proof. We do not give a detailed proof, as the underlying argument is quite standard in numerical analysis. Indeed, by (27) the local error of the approximation is of order $(\Delta t_i)^{(m+1)/2}$. Thus, by summing up the local errors we obtain that the global error is of order $\left(\max_i \Delta t_i\right)^{(m-1)/2}$ □

Remark 6.6. The above theorem is somewhat imprecise, as the constant C depends on T, f, the vector fields $V_0, \dots, V_d$ and the *method* of constructing the processes $W_1, \dots, W_N$, but not on the grid. E.g., in the case of the Ninomiya-Victoir method introduced below, C will only depend on $T, f, V_0, \dots, V_d$.

6.1 The Ninomiya–Victoir splitting

If liberally interpreted — e.g., for Euler schemes, when the path W is actually a step-function — Theorem 6.5 encompasses a large class of discretization schemes for the stochastic differential equation (26). In particular, it allows for a simple construction of *stochastic splitting schemes*, as we shall exemplify by the arguably most popular version, the *Ninomiya–Victoir scheme* [56]. In that case, the paths of the process W are *axis-paths*, i.e., the paths are continuous and piecewise-parallel to the axis in $\mathbb{R}^{d+1}$, similar to the construction used in Definition 2.3, see (9). More precisely, choose $1 \le i \le N$ and a sequence of independent (of all other sources of randomness) random variables Λ_i, $i = 1, \dots, N$ with $P(\Lambda_i = 1) = P(\Lambda_i = -1) = 1/2$. Construct a process W_i on $[0, \Delta t_i]$ in the following way: set $\delta_i \equiv \Delta t_i/(d+1)$ and when $\Lambda_i = +1$, set

$$\dot{W}_i(t) = \begin{cases} \Delta t_i/\delta_i e_0, & 0 \le t < 1/2\delta_i, \\ \Delta B^j_i/\delta_i e_j, & (1/2 + (j-1))\delta_i \le t < (1/2 + j)\delta_i,\ 1 \le j \le d, \\ \Delta t_i/\delta_i e_0, & \Delta t_i - 1/2\delta_i \le t \le \Delta t_i, \end{cases} \tag{30a}$$

where we recall that $\Delta B^j_i \equiv B^j_{t_i} - B^j_{t_{i-1}}$ and where we denote by $(e_0, e_1, \dots, e_d)$ the standard basis of $\mathbb{R}^{d+1}$. In the other case ($\Lambda_i = -1$), we define W_i by

$$\dot{W}_i(t) = \begin{cases} \Delta t_i/\delta_i e_0, & 0 \le t < 1/2\delta_i, \\ \Delta B^{d-j+1}_i/\delta_i e_{d-j+1}, & (1/2 + (j-1))\delta_i \le t < (1/2 + j)\delta_i,\ 1 \le j \le d, \\ \Delta t_i/\delta_i e_0, & \Delta t_i - 1/2\delta_i \le t \le \Delta t_i. \end{cases} \tag{30b}$$

As in the general construction, the independent processes $W_1, \dots, W_N$ are then concatenated to form the process W defined on $[0, T]$.

Inserting the process W just constructed into the general methodology (29), we see that the Ninomiya–Victoir method boils down to solving the ODEs driven by the individual vector fields $V_0, \dots, V_d$ on $\mathbb{R}^e$. Indeed, let $e^{sV_i}x$ denote the flow associated to the vector field V_i at time s, i.e., $e^{sV_i}x = z(s)$ solution to

$$\dot{z}(t) = V_i(z(t)), \quad z(0) = x \in \mathbb{R}^e,$$

then the solution $\overline{y}_l \equiv y(W)_{t_l}, l = 0, \dots, N$, of (29) for the Ninomiya–Victoir process W satisfies $\overline{y}_0 = y_0$ and

$$\overline{y}_l = \begin{cases} e^{\frac{\Delta t_l}{2}V_0} e^{\Delta B_l^d V_d} \cdots e^{\Delta B_l^1 V_1} e^{\frac{\Delta t_l}{2}V_0} \overline{y}_{l-1}, & \Lambda_l = +1, \\ e^{\frac{\Delta t_l}{2}V_0} e^{\Delta B_l^d V_1} \cdots e^{\Delta B_l^1 V_d} e^{\frac{\Delta t_l}{2}V_0} \overline{y}_{l-1}, & \Lambda_l = -1, \end{cases} \tag{31}$$

$l = 1, \dots, N$. This explains from the SDE side, why we consider the Ninomiya–Victoir scheme a stochastic splitting scheme for the SDE (26).

Theorem 6.7. *Under Assumption 3, the Ninomiya–Victoir scheme is a weak scheme of second order, i.e., there is a constant $C > 0$ (depending on $T, f, V_0, \dots, V_d$, but not on the grid) such that*

$$|E[f(y_T)] - E[f(y(W)_T]| \le C \left(\max_{i=1,\dots,N} \Delta t_i \right)^2.$$

Proof. We show that $E\left[S^5_{0,t}(B)\right] = E\left[S^5_{0,t}(W)\right]$ for $t = \Delta t_i$ and any $W = W_i$, which implies the conclusion by Theorem 6.5.

Let us first consider the (Stratonovich) signature of the Brownian motion. By the construction of the Stratonovich integral in terms of the Ito integral, we have

$$B_t^{(i_1,\dots,i_k)} = int_{0<t_1<\cdots<t_k<t} \circ dB_{t_1}^{i_1} \cdots \circ dB_{t_k}^{i_k} = \begin{cases} \int_0^t B_{t_k}^{(i_1,\dots,i_{k-1})} dB_{t_k}^{i_k} + \frac{1}{2}\int_0^t B_s^{(i_1,\dots,i_{k-2})} ds \delta_{i_k i_{k-1}}, & i_k \neq 0, \\ \int_0^t B_{t_k}^{(i_1,\dots,i_{k-1})} dB_{t_k}^{i_k}, & i_k = 0. \end{cases}$$

Using the Ito isometry, the expectation of the iterated Stratonovich integral is iteratively given by

$$E\left[B_t^{(i_1,\dots,i_k)}\right] = \begin{cases} \frac{1}{2}\int_0^t E\left[B_s^{(i_1,\dots,i_{k-2})}\right] ds \delta_{i_k i_{k-1}}, & i_k \neq 0, \\ \int_0^t E\left[B_s^{(i_1,\dots,i_{k-1})}\right] ds, & i_k = 0. \end{cases}$$

As regards the Ninomiya–Victoir cubature formula defined above, we see that

$$\dot{W}_i^j(t) = \Delta B_i^j / \delta_i \mathbf{1}_{A_i^j}(t), \quad j = 0, \dots, d, \quad i = 1, \dots, N,$$

where we tacitly let $\Delta B_i^0 = \Delta t_i$ and define the set A_i^j by

$$A_i^0 = [0, 1/2\delta_i[\cup [\Delta t_i - 1/2\delta_i, \Delta t_i],$$
$$A_i^j = \begin{cases} [(j - 1/2)\delta_i, (j + 1/2)\delta_i], & \Lambda_i = +1, \\ [(d - j + 1/2)\delta_i, (d - j + 3/2)\delta_i], & \Lambda_i = -1, \end{cases} \quad j = 1, \dots, d.$$

So we have the general formula

$$E\left[W_i^{(i_1,\dots,i_k)}(\Delta t_i)\right] = E\left[\Delta B_i^{i_1} \cdots \Delta B_i^{i_k}\right] E\left[\int_{0<t_1<\cdots<t_k} \mathbf{1}_{A_i^{i_1}}(t_1) \cdots \mathbf{1}_{A_i^{i_k}}(t_k) dt_1 \cdots dt_k\right],$$

where the last expectation is necessary due to the random choice of intervals above, and, in fact, only involves the two alternatives $\Lambda_i = \pm 1$.

The verification of the theorem now boils down to a simple, albeit tedious exercise. For instance, for multi-indices of length 3, we see that the only non-zero components of the expectation of the signature restricted to multi-indices of length 3 for either B and W_i are

$$E\left[B_{\Delta t_i}^{(0,0,0)}\right] = \frac{\Delta t_i^3}{6} = E\left[W_i^{(0,0,0)}(\Delta t_i)\right],$$
$$E\left[B_{\Delta t_i}^{(0,j,j)}\right] = \frac{\Delta t_i^2}{4} = E\left[W_i^{(0,j,j)}(\Delta t_i)\right],$$
$$E\left[B_{\Delta t_i}^{(j,j,0)}\right] = \frac{\Delta t_i^2}{4} = E\left[W_i^{(j,j,0)}(\Delta t_i)\right],$$

$j = 1, \dots, d,$ □

A path-wise interpretation of the Ninomiya–Victoir splitting scheme

Interpreting the Ninomiya–Victoir scheme in the Lie/Strang splitting picture drawn in (9) and below, we define functions $a^{\Delta t_i}, b_1^{\Delta t_i}, \dots, b_d^{\Delta t_i}$ on the interval $[0, \Delta t_i]$ by

$$\dot{a}^{\Delta t_i}(t) = \frac{\Delta t_i}{\delta_i}\mathbf{1}_{[0,\delta_i/2[}(t),$$
$$\dot{b}_j^{\Delta t_i}(t) = \frac{\Delta B_i^j}{\delta_i}\mathbf{1}_{[(1/2+j-1)\delta_i,(1/2+j)\delta_i[}(t), \quad j = 1, \dots, d,$$
$$\dot{c}^{\Delta t_i}(t) = \frac{\Delta t_i}{\delta_i}\mathbf{1}_{[\Delta t_i-\delta_i/2,\Delta t_i]}(t).$$

After concatenating these paths, we could immediately construct a Lie-type splitting following Definition 2.3 (in fact, we would not need splitting the time component in the a and c paths) or a Strang-type splitting. However, taking the scaling of Brownian motion into account, we realize that we need to take care of Lie brackets of order up to 5 in order to obtain a high order scheme. Hence, we need even more "re-orderings" than in the ordinary Strang splitting. Thus, we further define paths

$$\dot{\widetilde{b}}_j^{\Delta t_i}(t) = \frac{\Delta B_i^j}{\delta_i}\mathbf{1}_{[(1/2+d-j)\delta_i,(1/2+d-j+1)\delta_i[}(t), \quad j = 1, \dots, d.$$

The two alternatives (30a) and (30b) of $W_i(t)$ are then given by

$$W_i(t) = (a^{\Delta t_i}(t) + c^{\Delta t_i}(t))e_0 + \sum_{j=1}^{d}\left(b_j^{\Delta t_i}(t)\mathbf{1}_{\Lambda_i=+1} + \widetilde{b}_j^{\Delta t_i}(t)\mathbf{1}_{\Lambda_i=-1}\right)e_j,$$

and the corresponding splitting scheme is indeed given by (31), taking into account that the solutions of the ODEs driven by $b_j^{\Delta t_i}$ and $\widetilde{b}_j^{\Delta t_i}$ eventually coincide at time Δt_i.

The Ninomiya–Victoir scheme as a splitting scheme for PDEs

It is well-known that the function $u(t, y) \equiv E[f(y_t)]$ with $y_0 = y$ satisfies the linear Cauchy problem

$$\frac{\partial}{\partial t}u(t, y) = Lu(t, y), \quad u(0, y) = f(y), \tag{32}$$

for $t > 0$ and $y \in \mathbb{R}^e$, respectively. Here, the partial differential operator L is defined by

$$Lg(y) = V_0 g(y) + \frac{1}{2}\sum_{i=1}^{d} V_i^2 g(y), \tag{33}$$

where we recall that for any vector field $V : \mathbb{R}^e \to \mathbb{R}^e$ and any smooth function $g : \mathbb{R}^e \to \mathbb{R}$ we set $Vg(y) \equiv \nabla g(y) \cdot V(y)$. Iterating this procedure also define $V^2 g$, with V^2 a second order differential operator. Hence, for any weak approximation $\overline{y}_N$ of y_T, we have that

$$E\left[f(\overline{y}_N)\right] \approx E\left[f(y_T)\right] = u(T, y_0), \tag{34}$$

and the order of the weak approximation is the order of the approximation in the solution of the PDE (32). In semi-group notation, we can denote the solution operator associated to L by $P_t \equiv \exp(tL)$, i.e.,

$$u(t, y) = P_t f(y).$$

Remark 6.8. Obviously, solving the SDE (26) is only one step for the solution of the PDE (32): in addition, one needs to approximate the expectation in (34). In principle, for the Ninomiya–Victoir method this is a numerical integral in dimension $d \times N$. As this dimension is typically quite high, one usually resorts to Monte Carlo or Quasi Monte Carlo methods for computing the integral. Numerically, the computational cost of the integration step is often much higher than the computational cost of the discretization of the SDE, as the rate of convergence of the integration schemes is only $\frac{1}{2}$ (for Monte Carlo) or (at best) 1 (for Quasi Monte Carlo). Nonetheless, higher order weak approximation methods can reduce the overall computational cost considerably as compared to low order methods, partly because they actually lead to a considerable reduction of the dimension of the integration problem in the second step. The advantages of using higher order schemes have been observed in many numerical studies, for instance [56, 2, 4, 43].

Remark 6.9. As compared to classical numerical solvers for the Cauchy problem (32), the stochastic approximation scheme presented here has some very different features. On the one hand, most standard numerical methods such as finite element or finite difference schemes produce approximate solutions $u(t, y)$ for all values of t and y simultaneously – within a certain region in time and space, and up to interpolation. On the other hand, using the stochastic representation (34), one only obtains an approximation of $u(t, y)$ for one particular t and one particular y. Moreover, the stochastic method crucially relies on the performance of the (Q)MC approximation for the expected values, and shares its strengths and weaknesses. Hence, for low-dimensional problems classical numerical PDE solvers are typically more efficient, whereas for high dimensions $e \gg 1$, the stochastic method is competitive or superior, as it does not suffer from the curse of dimensionality.

In light of (32), the question arises whether the Ninomiya–Victoir scheme can be naturally associated with a (PDE) splitting scheme for (32). To this end let us first consider the situation when there is only one time-step. Let Q_t^i, $i = 1, \dots, d$, be the semi-groups corresponding to the second order differential operators $\frac{1}{2}V_i^2$, $i = 1, \dots, d$, respectively. Formally, we write $Q_t^i = e^{\frac{t}{2}V_i^2}$. Moreover, we denote by Q_t^0 the semi-group associated with the operator V_0, i.e.,

$$Q_t^0 f(x) = f\left(e^{tV_0}x\right).$$

While the correspondence between e^{tV_0} and Q_t^0 is obvious, we note that Q_t^i is in some sense the expectation of $e^{B_t^i V_i}$. More precisely, stochastic Taylor expansion shows that for any C^∞-bounded test function f and any initial value $y \in \mathbb{R}^e$ we have

$$Q_t^i f(y) = E\left[f\left(e^{B_t^i V_i} y\right)\right].$$

Hence, in the case with only one time-step (with $\Delta t = T$, $\Delta B = B_T$) we obtain that

$$\begin{aligned} E\left[f(\overline{y}_1)\right] &= \frac{1}{2}E\left[f\left(e^{\frac{\Delta t}{2}V_0}e^{\Delta B^d V_d}\cdots e^{\Delta B^1 V_1}e^{\frac{\Delta t}{2}V_0}y\right)\right] + \frac{1}{2}E\left[f\left(e^{\frac{\Delta t}{2}V_0}e^{\Delta B^1 V_1}\cdots e^{\Delta B^d V_d}e^{\frac{\Delta t}{2}V_0}y\right)\right] \\ &= \frac{1}{2}Q_{\Delta t/2}^0 Q_{\Delta t}^d \cdots Q_{\Delta t}^1 Q_{\Delta t/2}^0 f(y) + \frac{1}{2}Q_{\Delta t/2}^0 Q_{\Delta t}^1 \cdots Q_{\Delta t}^d Q_{\Delta t/2}^0 f(y) \equiv Q_{\Delta t} f(y). \end{aligned}$$

Note that the weight "$\frac{1}{2}$" comes from the probability $\frac{1}{2}$ to choose either of the two alternatives in (31).

Iterating this construction along a discretization of the time interval $[0, T]$ as above, we recover a well-known splitting scheme from the PDE literature, sometimes referred to as "symmetrically

weighted sequential splitting" scheme, see [17]. In terms of the solution operator $P_t = \exp(tL)$, Theorem 6.7 thus says that

$$|P_T f(y) - Q_{\Delta t_N} \cdots Q_{\Delta t_1} f(y)| \leq C \left(\max_{i=1,\ldots,N} \Delta t_i \right)^2. \tag{35}$$

The Ninomiya–Victoir stochastic splitting for SPDEs

The stochastic splitting methodology introduced above can be directly generalized to the infinite-dimensional case, i.e., to the case of SPDEs instead of SDEs. This was first done by Bayer and Teichmann [5] for the abstract formulation of Theorem 6.5 under strong regularity conditions. Later on, Dörsek and Teichmann [21] have given a careful analysis of the Ninomiya–Victoir splitting and other splitting techniques for weak approximation of SPDEs under weaker assumptions. We will mainly follow their approach here.

Consider a stochastic partial differential equation of the form

$$dy_t = (Ay_t + V(y_t))dt + \sum_{i=1}^{d} V_i(y_t) dB_t^i, \quad y_0 \in X, \tag{36}$$

that is we assume that the stochastic fluctuation only depend on y_t, but not on derivatives of y_t. The state space X of the equation (36) is assumed to be a separable Hilbert space and the vector fields $V, V_1, \ldots, V_d : X \to X$ are Frechet-differentiable and Lipschitz continuous, whereas the operator $A : \mathcal{D}(A) \subset X \to X$ is the generator of a strongly continuous, pseudo-contractive semigroup on X — more regularity conditions on the coefficients are deferred until later. Then, a *mild* solution y_t of the SPDE exists. For details of the solution theory of this class of SPDEs we refer to the monograph [18].

As mild solutions to SPDEs of the form (36) are generally not semi-martingales, we cannot re-write (36) in Stratonovich form from the beginning, but have to work with the Ito formulation. Nonetheless, if we use the Ninomiya–Victoir stochastic splitting approach, all the resulting (simpler) SPDEs can, in fact, be written in Stratonovich form, hence we proceed just as above. Indeed, define

$$e^{sV_i} y = z_s, \text{ where } \dot{z}_t = V_i(z_t),\ z_0 = y \in X, \quad i = 1, \ldots, d.$$

Moreover, set $V_0(y) \equiv V(y) - \frac{1}{2} \sum_{i=1}^{d} DV_i(y) \cdot V_i(y)$, $y \in X$, which precisely corresponded to the Stratonovich drift if it actually was to exist, and define $e^{s(A+V_0)} y$ analogously, i.e., as solution z_s at time s of the Cauchy problem

$$\frac{\partial}{\partial t} z_t = Az_t + V_0(z_t), \quad z_0 = y \in X,$$

which may be represented in terms of the semi-group $\exp(tA)$ generated by A as

$$e^{s(A+V_0)} y = z_s = \exp(sA)y + \int_0^s \exp((s-t)A) V_0(z_t) dt.$$

Now we can define the Ninomiya–Victoir splitting essentially as in (31), i.e., we set $\overline{y}_0 = y_0$ and

$$\overline{y}_l = \begin{cases} e^{\frac{\Delta t_l}{2}(A+V_0)} e^{\Delta B_l^d V_d} \cdots e^{\Delta B_l^1 V_1} e^{\frac{\Delta t_l}{2}(A+V_0)} \overline{y}_{l-1}, & \text{with prob. } \frac{1}{2}, \\ e^{\frac{\Delta t_l}{2}(A+V_0)} e^{\Delta B_l^1 V_1} \cdots e^{\Delta B_l^d V_d} e^{\frac{\Delta t_l}{2}(A+V_0)} \overline{y}_{l-1}, & \text{else}, \end{cases} \tag{37}$$

$l = 1, \ldots, N$.

We formulate assumptions given in [21], which can, in fact, be weakened using suitably weighted spaces.

Assumption 4. Consider the coefficients $A, V, V_1, \ldots, V_d$ of the SPDE (36) and a function $f : X \to \mathbb{R}$. We assume that

- $A : \mathcal{D}(A) \subset X \to X$ generates a strongly continuous, pseudo-contractive semi-group on X and has a compact resolvent.
- $V, V_1, \dots, V_d \in C^6(X, X)$ and have bounded derivatives.
- $V, V_1, \dots, V_d$ are Lipschitz when considered as maps $\mathcal{D}(A^l) \to \mathcal{D}(A^l)$, $l = 1, \dots, 5$, where $\mathcal{D}(A^l)$ is equipped with the graph norm, i.e., the Hilbert norm given by $\|x\|^2_{\mathcal{D}(A^l)} \equiv \|x\|^2_X + \sum_{k=1}^l \left\|A^k x\right\|^2_X$.
- $f \in C^6_b(X)$.

Theorem 6.10 ([21, Cor. 7.11, Th. 7.20])**.** *Under Assumption 4, there is a constant C such that*

$$\left|E\left[f(y_T)\right] - E\left[f\left(\overline{y}_N\right)\right]\right| \leq C \left(\max_{i=1,\dots,N} \Delta t_i\right)^2 .$$

Remark 6.11. The theorem can also be re-formulated completely deterministically in the fashion of (35), i.e., as a deterministic splitting method for a PDE on an infinite-dimensional state space. This is the version actually given in [21].

7 Applications of weak schemes in financial engineering

Given a financial model of the form (26), where y_t could be a (one- or multi-dimensional) vector of asset (forward) prices, or a vector of asset prices and stochastic volatilities, or the individual factors of a multi-factor model, ... Disregarding financial technicalities (discounting, change to the pricing measure), we are concerned in computing a European option price with payoff function $f : \mathbb{R}^e \to \mathbb{R}$ and maturity T, i.e., our quantity of interest is

$$E\left[f\left(y_T\right)\right].$$

This simple option pricing problem is mostly relevant for *calibration* purposes, i.e., for identifying the model parameters which provide the best fit to the observed market prices. Hence, the speed of the pricing algorithm is extremely important for this application — more important than accuracy, due to the usually non negligible model error.[5] In this section we give an overview of some successful applications of weak stochastic splitting methods in this context. We begin with two numerical studies on the performance of the Ninomiya–Victoir scheme for two popular (finite-dimensional) models often used in financial engineering — the SABR model and the CEV model, a special case of the former [2, 4]. Then we report the performance of the Ninomiya–Victoir method in an actual calibration routine for yet another related model, the double-mean-reverting model, [3]. Finally, we present the performance of the weak stochastic splitting method for SPDEs, again in the context of a calibration problem, this time for an interest rate model, the Heath–Jarrow–Morton model, see [22].

Option pricing in high dimensions

The SABR model is a prominent example of a stochastic interest rate model. We consider the following generalizations of the classical SABR model (cf. [2]).

$$\begin{aligned} dy_1(t) =& a y_2(t)^\alpha y_1(t)^\beta dB^1_t, \\ dy_2(t) =& \kappa(\theta - y_2(t))dt + b y_2(t)\left(\rho dB^1_t + \sqrt{1-\rho^2} dB^2_t\right), \end{aligned} \tag{38}$$

with $X_1(0) = x_1$ and $X_2(0) = x_2$. We assume that the parameters satisfy $\frac{1}{2} \leq \beta \leq 1$, $\theta, \kappa \geq 0$, $\alpha > 0$, $a, b > 0$, $-1 < \rho < 1$. Here, the first component $y_1(t)$ models the (discounted) price of a stock, and $y_2(t)$ can be interpreted as some kind of stochastic volatility. In fact, the dynamics of y_1

[5] That is, even without any numerical error, it is generally not possible to obtain a perfect fit to market prices, due to the model limitations.

depend in a nonlinear way on y_1 (*local volatility*) and on a second stochastic process y_2 (*stochastic volatility*). Hence, models of this kind are known as stochastic local volatility models. Moreover, note that this model can be easily generalized to the multi-asset case by just adding new processes y with the same kind of dynamics, but driven by correlated Brownian motions for every new asset to be included in the model, see [2] for more details. We concentrate on the one-asset case for ease of presentation.

For the SABR model (38), the Stratonovich drift vector field and the diffusion vector fields are given by

$$V_0(y) = \begin{pmatrix} -\frac{1}{2}a^2\beta y_2^{2\alpha} y_1^{2\beta-1} - \frac{1}{2}\alpha a b \rho y_2^{\alpha} y_1^{\beta} \\ \kappa\theta - \left(\kappa + \frac{1}{2}b^2\right) y_2 \end{pmatrix}, \quad V_1(y) = \begin{pmatrix} a y_2^{\alpha} y_1^{\beta} \\ b\rho y_2 \end{pmatrix}, \quad V_2(y) = \begin{pmatrix} 0 \\ b\sqrt{1-\rho^2} y_2 \end{pmatrix}. \tag{39}$$

Quite typically for models in financial engineering, the Stratonovich drift vector field is considerably more complicated than the Ito drift vector field or the diffusion vector fields which can be seen as a consequence of the Stratonovich correction $V_0 = V - \frac{1}{2}\sum_{i=1}^{d} DV_i \cdot V_i$, noting that models in financial engineering are typically formulated in Ito form. As a consequence, it is not surprising that we have explicit formulas for the flow of the diffusion vector field, but not for the flow of the Stratonovich vector field V_0. In fact, we have

$$\mathrm{e}^{sV_1} y = \begin{pmatrix} g_1(s) \\ y_2 \exp\left(b\rho s\right) \end{pmatrix}, \quad \mathrm{e}^{sV_2} y = \begin{pmatrix} y_1 \\ y_2 \exp\left(b\sqrt{1-\rho^2} s\right) \end{pmatrix},$$

where

$$g_1(s) = \begin{cases} \left[(1-\beta) a y_2^{\alpha} \frac{\mathrm{e}^{\alpha b \rho s} - 1}{\alpha b \rho} + y_1^{1-\beta}\right]_+^{1/(1-\beta)}, & \frac{1}{2} \le \beta < 1, \\ y_1 \exp\left(a x_2^{\alpha} \frac{\mathrm{e}^{\alpha b \rho s} - 1}{\alpha b \rho}\right), & \beta = 1. \end{cases}$$

Of course, it is possible to compute $e^{sV_0} y$ numerically, but for efficiency (and often also for geometrical reasons) it is preferable to have explicit solutions whenever possible.[6] We therefore propose to slightly adjust the Ninomiya–Victoir splitting formula, taking the Stratonovich drift correction into account. That means, we replace the splitting $L = V_0 + \frac{1}{2}\sum_{i=1}^{d} V_i^2$ by the splitting

$$L = V_0^{(\gamma)} + \frac{1}{2}\sum_{i=1}^{d} \left(V_i^2 + 2\gamma_i V_i\right), \text{ where } V_0^{(\gamma)} = V_0 - \sum_{i=1}^{d} \gamma_i V_i,$$

and $\gamma \in \mathbb{R}^d$ is chosen such that the flow of $V_0^{(\gamma)}$ has an explicit solution. Note that $\frac{1}{2}V_i^2 + \gamma_i V_i$ corresponds to the stochastic equation

$$dz_t = \gamma_i V_i(z_t) dt + V_i(z_t) \circ dB_t^i = V_i(z_t) \circ d\left(B_t^i + \gamma_i t\right),$$

i.e., the stochastic weak splitting scheme actually looks exactly like the standard Ninomiya–Victoir scheme (31), but with ΔB_l^i replaced by $\Delta B_l^i + \gamma_i \Delta t_l$ and V_0 replaced by $V_0^{(\gamma)}$:

$$\overline{y}_l = \begin{cases} e^{\frac{\Delta t_l}{2} V_0^{(\gamma)}} e^{(\Delta B_l^d + \gamma_d \Delta t_l) V_d} \cdots e^{(\Delta B_l^1 + \gamma_1 \Delta t_l) V_1} e^{\frac{\Delta t_l}{2} V_0^{(\gamma)}} \overline{y}_{l-1}, & \text{with prob. } \frac{1}{2}, \\ e^{\frac{\Delta t_l}{2} V_0^{(\gamma)}} e^{(\Delta B_l^1 + \gamma_1 \Delta t_l) V_1} \cdots e^{(\Delta B_l^d + \gamma_d \Delta t_l) V_d} e^{\frac{\Delta t_l}{2} V_0^{(\gamma)}} \overline{y}_{l-1}, & \text{else}, \end{cases} \tag{40}$$

see [2]. As a matter of fact, we can easily find such a choice of γ for the generalized SABR model given by

$$\gamma_1 = -\frac{1}{2}\alpha b \rho, \quad \gamma_2 = \frac{\alpha b \rho^2 - 2\kappa/b - b}{2\sqrt{1-\rho^2}},$$

leading to

$$V_0^{(\gamma)}(y) = \begin{pmatrix} -\frac{1}{2}a^2\beta y_2^{2\alpha} y_1^{2\beta-1} \\ \kappa\theta \end{pmatrix}, \quad e^{sV_0^{(\gamma)}} y = \begin{pmatrix} g_0(s; y) \\ \kappa\theta s + y_2 \end{pmatrix},$$

[6] By which we do not mean difficult-to-evaluate series expansions, Bessel functions or similar solutions. Instead, we mean formulas with comparable complexity to the vector fields themselves.

with

$$g_0(s;y) = \begin{cases} \left(-a^2\beta(1-\beta)P(s;y) + y_1^{2(1-\beta)}\right)_+^{\frac{1}{2(1-\beta)}}, & \frac{1}{2} < \beta < 1, \\ y_1 \exp\left(-\frac{1}{2}a^2 P(s;y)\right), & \beta = 1, \\ -\frac{1}{4}a^2 P(s;y) + y_1, & \beta = \frac{1}{2}, \end{cases}$$

and

$$P(s;y) = \frac{1}{(2\alpha+1)\kappa\theta}\left((\kappa\theta s + y_2)^{2\alpha+1} - y_2^{2\alpha+1}\right).$$

Finally, let us present the results from one of the numerical experiments in [2] with real-world data. The parameters there were chosen to be $\beta = 1.0$, $\theta = 0.3$, $\kappa = 2.0$, $\alpha = 0.5$, $a = 1.0$, $b = 0.5$, $\rho = -0.7$, $y_1 = 1.0$ and $y_2 = 0.2$. The option has strike price $K = 1.05$ and time to maturity $T = 1.0$ years. The estimated "true result" is 0.1767505855. Note that the expectation is computed by quasi Monte Carlo based on Sobol numbers.

Method	K	M	Rel. Error	Time
Euler	32	8192000	0.00174	91.94 sec
Ninomiya–Victoir	4	2048000	0.00204	13.93 sec
NV with drift	4	1024000	0.00104	2.88 sec

Table 1: Computational time for the generalized SABR model

In Table 1 the computational times are reported for the generalized SABR model. Here, the computational parameter N (the number of uniform time-steps) is chosen such that the weak error is of order 10^{-3}. We see that the computational time needed for the adjusted splitting method (40) is indeed considerably smaller than the time for the classical one (31). The other numerical parameter (the number M of samples for the quasi Monte Carlo integration, restricted to be a power of 2) was chosen such that the integration error (i.e., the error in the computation of the expected value) is of order 10^{-5}. Indeed, we focus on the discretization problem here, an we do not want our results to be overshadowed by the integration error. Note that in the case of the Euler scheme, one has to compute a 64-dimensional integral, whereas in the case of the Ninomiya–Victoir scheme (with or without drift), the integration only needs to be performed on an eight-dimensional space. This explains why the M can be chosen smaller for the Ninomiya–Victoir splitting as compared to the Euler scheme, as quasi Monte Carlo is known to work better when the dimension is smaller – despite not suffering from the curse of dimensionality.

In [2] similar results were reported for the multi-asset case. More precisely, the authors of [2] also applied it to the case of four assets, meaning an eight-dimensional model. But, in fact, the stochastic splitting method can be used in even higher dimensions. For instance, we used it in [4] in order to obtain reference solutions for options depending on up to 100 assets for a pure local volatility model, coupled with quasi Monte Carlo or Monte Carlo methods for the integration step. In that case, it is difficult if not impossible to obtain reliable reference values, but the method seems to perform very well.

Calibration of the double mean reverting model

The double mean reverting model goes back to Jim Gatheral [31]. Its main advantage is that it allows joint calibration to market data for option prices on an index like the S & P 500 index (SPX) and a corresponding implied volatility index (like the VIX). The model is given by

$$dS_t = \sqrt{v_t} S_t dW_t^1, \tag{41a}$$

$$dv_t = \kappa_1(v_t' - v_t)dt + \xi_1 v_t^{\alpha_1} dW_t^2, \tag{41b}$$

$$dv_t' = \kappa_2(\theta - v_t')dt + \xi_2 v_t'^{\alpha_2} dW_t^3, \tag{41c}$$

where the Brownian motions W_i are all correlated with $E[dW_t^i dW_t^j] = \rho_{ij}dt$. Again, it is natural to interpret v_t as the (stochastic) volatility of the (discounted) asset price process S_t – or rather, as the

stochastic variance. Conforming to stylized facts about the volatility, v_t is a mean-reverting process due to the form of the drift, but unlike traditional stochastic volatility models, the long-term mean v'_t is itself a (mean-reverting) stochastic process, hence the name "double mean reverting" model.

Typically, one of the most numerically challenging tasks in financial engineering is the calibration of a model such as (41), i.e., the fitting of the model parameters (κ_1, κ_2, ξ_1, ξ_2, α_1, α_2, ρ_{12}, ρ_{13}, ρ_{23}, but also v_0 and v'_0 which, unlike S_0, are not directly observable) to market data, in particular to market option prices. Indeed, even though the model itself assumes these parameters to be constant, in reality they typically change frequently, which means that the model has to be *re-calibrated* on a regular bases, say daily.

In the case of the double mean reverting model, practical experience seems to show that θ, κ_1, κ_2, ρ_{23} α_1 and α_2 are fairly constant in time, implying that they can be excluded from the daily re-calibration. In fact, [3] found that the data available did not suffice to successfully estimate α_2. Hence, it was *assumed* to have the same value as α_1, which was calibrated to $\alpha_1 = 0.94$. Hence, for the purpose of their numerical study, [3] assumed θ, κ_1, κ_2, ρ_{23} α_1 and α_2 to be given (by parameters which where themselves, of course, calibrated to the market data) – leaving us with the task of fast calibration of the parameters ξ_1, ξ_2, ρ_{12} and ρ_{13}. In the context of [3], "market data" mean the prices of vanilla (i.e., European put and call options) on SPX and on VIX. The general calibration procedure proposed was the following:

1 Given a time series of VIX data, linear regression allows to construct time series for the processes v_t and v'_t. Out of these, a least-square optimization is used to estimate θ, κ_1 and κ_2. Moreover, the correlation between v_t and v'_t gives ρ_{23}. A similar regression on VIX time series data gives an estimate for α_1.

2 Note that options on VIX depend only on v_t and v'_t now, but not on S_t. Hence, one can calibrate ξ_1 and ξ_2 directly to VIX options, without needing to simulate the S_t component, i.e., without adding any constraints to ρ_{12} and ρ_{13}. The calibration boils down to a least-squares minimization of misfits of VIX-option prices from the model to the quoted market prices. The minimization was done using a Levenberg-Marquardt algorithm, for the option pricing algorithm the authors of [3] tested the Euler scheme and a variant of the Ninomiya–Victoir scheme for the discretization of the SDE and both classical and quasi Monte Carlo for the computation of the expected value.

3 Having obtained ξ_1 and ξ_2 from the previous step, they used options on the SPX to calibrate the remaining parameters ρ_{12} and ρ_{13}. The procedure is very similar to the calibration of ξ_1 and ξ_2, except that now the full three-dimensional SDE needs to be solved.

In [3] the calibration was done for two particular days, namely April 3, 2007 (before the financial crisis) and September 15, 2011 (after the financial crisis). The fits to SPX options are quite good, especially for maturities which are not too small. The fit to VIX options is slightly worse, but in that time VIX options were also less liquid than today. Regarding the numerical algorithms, the Ninomiya–Victoir splitting method (with an additional splitting in the drift, not unlike the one presented in [2]) performs much better for the calculation of VIX options, where the classical Euler method would requires 500 time-steps as compared to 30 time-steps for the splitting method in order to achieve the required accuracy. Thereby, for this example, the Ninomiya–Victoir scheme reduces the calibration time for the VIX-step by a factor of around five. For the SPX options, the Euler scheme surprisingly gave sufficiently accurate results already for 30 time-steps, which implies that for this case the Euler scheme turned out to be seemingly In total, the authors of [3] report that their implementation can do the re-calibration to market data in about 5 seconds using the Ninomiya–Victoir splitting scheme.

Calibration of the Heath–Jarrow–Morton model

Finally, we want to present an application of the Ninomiya–Victoir weak splitting method to a true SPDE given by Dörsek and Teichmann [22], namely the fast calibration of a general, infinite-dimensional Heath–Jarrow–Morton model for interest rate dynamics, see [37]. We should also note alternative numerical treatments of the full, infinite dimensional HJM model in [8] and [55].

We start with a short description of the model. Let $P(t,T)$ denote the price of a zero-coupon bond with maturity T at time t. Of course, this entity only makes sense if $t \leq T$. The (instantaneous) *forward rate* at time t for the maturity T is defined by

$$f(t,T) = -\frac{\partial}{\partial T} \log P(t,T),$$

implying the natural inverse relation $P(t,T) = \exp\left(-\int_t^T f(t,u)du\right)$, which explains why we call f an interest rate. Unlike many other models, in which only the short rate $f(t,t)$ or only the rates $f(t,T_1),\ldots,f(t,T_n)$ for finitely many maturities are modeled, [37] propose an infinite-dimensional model for the whole forward rate curve $(f(t,T))_{T\in[t,\infty[}$. In order to avoid working with time-dependent state spaces, we introduce the parametrization $r_t(x) = f(t,t+x)$ in *time to maturity* $x = T - t \geq 0$. Then the HJM model corresponds to the SPDE

$$dr_t = \left(\frac{\partial}{\partial x} r_t + \alpha_{HJM}(r_t)\right) dt + \sum_{i=1}^{d} \sigma_i(r_t) dB_t^i. \tag{42}$$

Here, we restrict ourselves to a finite number d of driving Brownian motions, which can be justified empirically, but is not strictly necessary for the HJM model. Moreover, we note that there are inherent restrictions on the vector fields imposed by no-arbitrage constraints, which boil down to the relation

$$\alpha_{HJM}(h)(x) = \sum_{i=1}^{d} \sigma_i(h)(x) \int_0^x \sigma_i(h)(y)dy,$$

where $x \geq 0$ and h takes values in the state space H, a suitably weighted Sobolev space, see [22] for details on H and on further regularity requirements on σ_i.

Next, we describe the splitting. Note that the diffusion vector fields do not pose any additional complications as compared to the finite dimensional case, as they are (assumed to be) continuous vector fields on H. This is evidently not the case for the Stratonovich drift vector field $\sigma_0(h) = \frac{\partial}{\partial x}h + \alpha_{HJM}(h) - \frac{1}{2}D\sigma_i(h)\cdot\sigma_i(h)$, where the unbounded operator $\frac{\partial}{\partial x}$ appears. (Recall that the solution r_t cannot, in fact, be written in Stratonovich form.) Hence, they additionally split $\sigma_0 = \sigma_{0,1} + \sigma_{0,2}$ with $\sigma_{0,1} = \frac{\partial}{\partial x}$ and $\sigma_{0,2} = \alpha_{HJM} - \frac{1}{2}D\sigma_i \cdot \sigma_i$. Here we note that the flow of $\sigma_{0,1}$ is obviously given by the shift operator $S_t(h)(x) = h(x+t)$, so that the $e^{s\sigma_{0,1}} = S_s$ is given in closed form. Regarding the diffusion vector fields, the authors suggest to use the parametric form

$$\sigma_j(h,v)(x) = (\alpha_{j,0} + \alpha_{j,1}x)\, e^{-\beta x} \tanh\left(c_j e^v \int_0^{t_j} h(s)ds\right),$$

which includes a stochastic volatility component v, and $\alpha_{j,i}$, β, c_j and t_j are parameters, which need to be estimated. Moreover, they choose $d = 3$.

The authors of [22] calibrate against 120 market prices of caplets, again using a Levenberg-Marquardt type algorithm for the optimization. In total, they report that it takes about half a second to compute these 120 option prices to the required accuracy (on a workstation with 16 cores), and the total calibration can be done in 14.5 minutes.

References

[1] A. Bain and D. Crişan. *Fundamentals of stochastic filtering*. Applications of mathematics. Springer, 2008.

[2] Christian Bayer, Peter Friz, and Ronnie Loeffen. Semi-closed form cubature and applications to financial diffusion models. *Quant. Finance*, 13(5):769–782, 2013.

[3] Christian Bayer, Jim Gatheral, and Morten Karlsmark. Fast Ninomiya-Victoir calibration of the double-mean-reverting model. *Quantitative Finance*, 2013. Available online.

[4] Christian Bayer and Peter Laurence. Asymptotics beats Monte Carlo: The case of correlated CEV baskets. *Comm. Pure Appl. Math.*, 2013. Available online.

[5] Christian Bayer and Josef Teichmann. Cubature on Wiener space in infinite dimension. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 464(2097):2493–2516, 2008.

[6] A. Bensoussan and R. Glowinski. Approximation of Zakai equation by the splitting up method. In *Stochastic systems and optimization (Warsaw, 1988)*, volume 136 of *Lecture Notes in Control and Inform. Sci.*, pages 257–265. Springer, Berlin, 1989.

[7] A. Bensoussan, R. Glowinski, and A. Răşcanu. Approximation of the Zakai equation by the splitting up method. *SIAM J. Control Optim.*, 28(6):1420–1431, 1990.

[8] T. Björk, A. Szepessy, R. Tempone, and G. E. Zouraris. Monte Carlo Euler approximations of HJM term structure financial models. *BIT*, 53(2):341–383, 2013.

[9] Damiano Brigo and Bernard Hanzon. On some filtering problems arising in mathematical finance. *Insurance: Mathematics and Economics*, 22(1):53–64, 1998.

[10] Roger W. Brockett. Volterra series and geometric control theory. *Automatica*, 12(2):167 – 176, 1976.

[11] Michael Caruana, Peter Friz, and Harald Oberhauser. A (rough) pathwise approach to a class of non-linear stochastic partial differential equations. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 28(1):27–46, 2011.

[12] J.M.C. Clark. The design of robust approximations to stochastic differential equations in nonlinear filtering. In J.K. Skwirzynsky, editor, *Communication Systems in Random Processes Theory*. Sijthoff,Nordhoff, 1978.

[13] JMC Clark and D. Crisan. On a robust version of the integral representation formula of nonlinear filtering. *Probability theory and related fields*, 133(1):43–56, 2005.

[14] Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)*, 27(1):1–67, 1992.

[15] D. Crisan and B. Rozovskii. *The Oxford Handbook of Nonlinear Filtering*. Oxford Handbooks. Oxford University Press, 2011.

[16] Dan Crisan, Joscha Diehl, Peter Friz, and Harald Oberhauser. Robust filtering: multdimensional noise and multidimensional observation. *Annals of Applied Probability (in press)*, 2012.

[17] P. Csomós, I. Faragó, and Á. Havasi. Weighted sequential splittings and their analysis. *Comput. Math. Appl.*, 50(7):1017–1031, 2005.

[18] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.

[19] Arnaud Debussche. Weak approximation of stochastic partial differential equations: the nonlinear case. *Math. Comp.*, 80(273):89–117, 2011.

[20] J. Diehl, P. Friz, and P. Gassiat. Stochastic control with rough paths. *ArXiv e-prints*, March 2013.

[21] Philipp Dörsek and Josef Teichmann. A semi-group point of view on splitting schemes for stochastic (partial) differential equations. Preprint, 2010.

[22] Philipp Dörsek and Josef Teichmann. Efficient simulation and calibration of general HJM models by splitting schemes. *SIAM J. Financial Math.*, 4(1):575–598, 2013.

[23] Robert J Elliott and Roland Glowinski. Approximations to solutions of the Zakai filtering equation. *Stochastic Analysis and Applications*, 7(2):145–168, 1989.

[24] Wendell H. Fleming and H. Mete Soner. *Controlled Markov processes and viscosity solutions*, volume 25 of *Stochastic Modelling and Applied Probability*. Springer, New York, second edition, 2006.

[25] M. Fliess. Fonctionnelles causales non linéaires et indéterminées non commutatives. *Bull. Soc. Math. France*, 109(1):3–40, 1981.

[26] Patrick Florchinger and François Le Gland. Time-discretization of the Zakai equation for diffusion processes observed in correlated noise. *Stochastics Stochastics Rep.*, 35(4):233–256, 1991.

[27] Rüdiger Frey and Wolfgang J Runggaldier. A nonlinear filtering approach to volatility estimation with a view towards high frequency data. *International Journal of Theoretical and Applied Finance*, 4(02):199–210, 2001.

[28] Peter Friz and Harald Oberhauser. On the splitting-up method for rough (partial) differential equations. *Journal of Differential Equations*, 251(2):316 – 338, 2011.

[29] Peter Friz and Harald Oberhauser. Rough path stability of (semi-)linear SPDEs. *Probability Theory and Related Fields*, pages 1–34, 2013.

[30] Peter K. Friz and Nicolas B. Victoir. *Multidimensional stochastic processes as rough paths: theory and applications*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2010.

[31] Jim Gatheral. Consistent modeling of SPX and VIX options. Presentation at Bachelier Congress: London, 2008.

[32] Ramazan Gençay, Faruk Selçuk, and Brandon J Whitcher. *An introduction to wavelets and other filtering methods in finance and economics*. Access Online via Elsevier, 2001.

[33] M. Gubinelli. Controlling rough paths. *J. Funct. Anal.*, 216(1):86–140, 2004.

[34] István Gyöngy and Nicolai Krylov. On the splitting-up method and stochastic partial differential equations. *Ann. Probab.*, 31(2):564–591, 2003.

[35] István Gyöngy and Nicolai Krylov. Accelerated numerical schemes for PDEs and SPDEs. In *Stochastic analysis 2010*, pages 131–168. Springer, Heidelberg, 2011.

[36] Erika Hausenblas. Approximation for semilinear stochastic evolution equations. *Potential Anal.*, 18(2):141–186, 2003.

[37] David Heath, Robert Jarrow, and Andrew Morton. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, 60(1):77–105, 1992.

[38] Takeyuki Hida. *Brownian motion*. Springer, 1980.

[39] Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes*. North-Holland Publishing Co., Amsterdam, second edition, 1989.

[40] A. Jentzen and P. E. Kloeden. The numerical approximation of stochastic partial differential equations. *Milan J. Math.*, 77:205–244, 2009.

[41] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[42] Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[43] Arturo Kohatsu-Higa and Peter Tankov. Jump-adapted discretization schemes for Lévy-driven SDEs. *Stochastic Process. Appl.*, 120(11):2258–2285, 2010.

[44] Shigeo Kusuoka. Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus. In *Advances in mathematical economics. Vol. 6*, volume 6 of *Adv. Math. Econ.*, pages 69–83. Springer, Tokyo, 2004.

[45] François Le Gland. Splitting-up approximation for SPDEs and SDEs with application to nonlinear filtering. In *Stochastic partial differential equations and their applications (Charlotte, NC, 1991)*, volume 176 of *Lecture Notes in Control and Inform. Sci.*, pages 177–187. Springer, Berlin, 1992.

[46] Sergey Lototsky, Remigijus Mikulevicius, and Boris L. Rozovskii. Nonlinear filtering revisited: a spectral approach. *SIAM J. Control Optim.*, 35(2):435–461, 1997.

[47] Terry Lyons. Differential equations driven by rough signals. I. An extension of an inequality of L. C. Young. *Math. Res. Lett.*, 1(4):451–464, 1994.

[48] Terry Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana*, 14(2):215–310, 1998.

[49] Terry Lyons. Systems controlled by rough paths. In *European Congress of Mathematics*, pages 269–281. Eur. Math. Soc., Zürich, 2005.

[50] Terry Lyons and Zhongmin Qian. *System Control and Rough Paths*. Oxford University Press, 2002. Oxford Mathematical Monographs.

[51] Terry Lyons and Nicolas Victoir. Cubature on Wiener space. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 460(2041):169–198, 2004.

[52] Terry J. Lyons, Michael Caruana, and Thierry Lévy. *Differential equations driven by rough paths*, volume 1908 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, 2004, With an introduction concerning the Summer School by Jean Picard.

[53] Benoit B Mandelbrot and John W Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968.

[54] Noriaki Nagase. Remarks on nonlinear stochastic partial differential equations: an application of the splitting-up method. *SIAM J. Control Optim.*, 33(6):1716–1730, 1995.

[55] Mariko Ninomiya. Application of the Kusuoka approximation with a tree-based branching algorithm to the pricing of interest-rate derivatives under the HJM model. *LMS J. Comput. Math.*, 13:208–221, 2010.

[56] Syoiti Ninomiya and Nicolas Victoir. Weak approximation of stochastic differential equations and application to derivative pricing. *Appl. Math. Finance*, 15(1-2):107–121, 2008.

[57] Bernt Øksendal. *Stochastic differential equations*. Springer, 2003.

[58] Philip E. Protter. *Stochastic integration and differential equations*, volume 21 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, second edition, 2004. Stochastic Modelling and Applied Probability.

[59] Philip E. Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2005. Second edition. Version 2.1, Corrected third printing.

[60] A. Răşcanu and C. Tudor. Approximation of stochastic equations by the splitting up method. In *Qualitative problems for differential equations and control theory*, pages 277–287. World Sci. Publ., River Edge, NJ, 1995.

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

[62] Christoph Schwab and Claude Jeffrey Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numer.*, 20:291–467, 2011.

[63] H.J. Sussmann. Semigroup representations, bilinear approximation of input-output maps, and generalized inputs. *Mathematical systems theory*, 131, 1975.

[64] J. Teichmann. Another approach to some rough and stochastic partial differential equations. *Stochastics and Dynamics*, 11(02n03):535–550, 2011.

[65] Yubin Yan. Galerkin finite element methods for stochastic parabolic partial differential equations. *SIAM J. Numer. Anal.*, 43(4):1363–1384 (electronic), 2005.