A review of recent results on approximation of solutions of stochastic differential equations*

March 5, 2010

Abstract

In this article, we give a brief review of some recent results concerning the study of the Euler-Maruyama scheme and its high order extensions. These numerical schemes are used to approximate solutions of stochastic differential equations, which enables to approximate various important quantities including solutions of partial differential equations. Some have been implemented in Premia [56]. In this article we mainly consider results about weak approximation, the most important for financial applications.

1 Introduction

The Euler-Maruyama scheme is a simple and natural approximation method for the solution of various types of stochastic differential equations. It helps not only to simulate the solutions of stochastic equations but it also serves theoretical purposes (see e.g. the articles of E. Gobet [18] [19] on the local asymptotic mixed normality (LAMN) property in statistics).

To introduce this notion consider the stochastic differential equation

$$X(t) = x + \int_0^t b(X(s))ds + \sum_{j=1}^{r} \int_0^t \sigma_j(X(s))dZ^j(s), \tag{1}$$

where $b, \sigma_i : \mathbb{R}^d \to \mathbb{R}^d$, $i = 1, ..., r$, are Lipschitz coefficients and $Z = (Z^1, \ldots, Z^r)$ is a r dimensional Wiener process.

For a partition of the interval $[0, T]$ denoted as $\pi : 0 = t_0 < ... < t_n = T$, we define the norm of the partition as $\|\pi\| = \max\{t_{i+1} - t_i; i = 0, ..., n-1\}$ and $\eta(t) = \sup\{t_i; t_i \leq t\}$ the last discretization time before t. Then the Euler-Maruyama scheme is defined inductively by

$$\forall t \in [t_i, t_{i+1}],\ X^\pi(t) = X^\pi(t_i) + b(X^\pi(t_i))(t - t_i) + \sum_{j=1}^{r} \sigma_j(X^\pi(t_i))(Z^j(t) - Z^j(t_i)).$$

The simplicity and the generality of the possible applications are the main attractions of this scheme. In practice, one only needs to simulate the Brownian increments $(Z(t_{i+1}) - Z(t_i))_{0 \leq i \leq n-1}$ in order to compute $(X^\pi(t_1), X^\pi(t_2), \ldots, X^\pi(t_n))$. Let us first mention the strong convergence rate result.

Theorem 1 *Under the above assumptions*

$$\forall p \geq 1,\ E\left[\sup_{t \leq T} \|X(t) - X^\pi(t)\|^{2p}\right] \leq C \|\pi\|^p$$

where the constant C depends on p, T, x and the Lipschitz constants.

*Keywords: Euler-Maruyama scheme, Kusuoka scheme, Milshtein scheme, weak approximations, stochastic equations.

The proof of this result is standard and essentially goes through the same methodology used to prove existence of solutions to (1). This result can also be generalized to various equations without changing the essential ideas.

In this paper, we are mainly interested in analyzing different error terms which involve a test function. The corresponding results are called weak convergence results. Section two deals with weak convergence for the Euler scheme. In case the partition is uniform ($t_i = \frac{iT}{n}$), denoting the Euler scheme by X^n instead of X^π, we first state the convergence in law of the normalized error process $\sqrt{n}(X - X^n)$ to a process χ which writes as a stochastic integral with respect to a Brownian motion independent from Z. More precisely, for any continuous and bounded function F on the space of continuous paths, $E\left(F(\sqrt{n}(X - X^n))\right)$ converges to $E\left(F(\chi)\right)$ as $n \to +\infty$.
Analysis of the weak error $E\left(f(X(T))\right) - E\left(f(X^n(T))\right)$ turns out to be more important for applications : for instance, the price of a European option with payoff function f and maturity T written on an underlying evolving according to (1) under the risk-neutral measure writes $E(e^{-rT}f(X(T)))$ where r denotes the risk-free rate. In [60], Talay and Tubaro prove that this difference can be expanded in powers of $\frac{1}{n}$. This justifies the use of Romberg-Richardson extrapolations in order to speed-up the convergence : for instance $E\left(f(X(T))\right) - E\left(2f(X^{2n}(T)) - f(X^n(T))\right) = \mathcal{O}(\frac{1}{n^2})$ (see Pagès [53] for a recent study devoted to the numerical implementation of these extrapolations). The proof given by Talay and Tubaro relies on the Feynman-Kac partial differential equation associated with (1). Here, we present another methodology introduced in [33] and based on the integration by parts formula of Malliavin calculus. In [10], a general framework relying on the study of the linear stochastic equation satisfied by the error process $X - X^n$ is presented. This new methodology enables to deal with a great variety of equations including some which seem beyond the scope of the former PDE approach. We illustrate this latter point on the example of stochastic differential equations with delay.

The third section is devoted to a method of exact (in law) simulation of (1) recently introduced by [4] [5] in the one dimensional case $r = d = 1$. Then, for a smooth diffusion coefficient σ which does not vanish, one can make a change of variables which transforms (1) into a SDE with diffusion coefficient constant and equal to one. Under a new probability measure given by Girsanov theorem, the original Brownian motion Z solves the latter SDE. The exponential factor giving the change of probability measure is then simulated by a rejection/acceptation technique involving a Poisson point process.

It is possible to obtain schemes with convergence order higher than the one of the Euler scheme by keeping more terms in the stochastic Taylor expansion of the solution of the SDE. Section four deals with such schemes. We first introduce Stratonovitch stochastic integrals in order to write nice Taylor expansions. Then, on the example of the Milshtein scheme, we illustrate the difficulty to simulate the iterated Brownian integrals which appear in the expansions and therefore to implement schemes with high order of strong convergence. Recently, to overcome this difficulty, Kusuoka [40] [41] proposed to replace these iterated Brownian integrals by random variables with the same moments up to a given order. This leads to schemes with high order of weak convergence. These schemes and their application in finance are currently the subject of a consequent research activity : [42] [47] [48] [50] [17].

The last section adresses extensions of the results presented previously. The case where instead of Z we have a Levy process, Z, is first considered. Discretization of reflecting stochastic differential equations is also discussed.

2 Weak errors: from Jacod-Kurtz-Protter to Milshtein-Talay

If one is trying to approach the problem of weak convergence of the error process then the first natural approach is to study the weak convergence of the process

$$\sqrt{n}\left(X(t) - X^n(t)\right).$$

This is done in a series of articles by Jacod, Kurtz and Protter (e.g. see Section 5 in [39]). To simplify the ideas suppose that we are dealing with the Wiener case in one dimension ($r = 1$), $b \equiv 0$ and that the partition is uniform : $t_i = \frac{iT}{n}$. Since the process X^n solves

$$X^n(t) = x + \int_0^t \sigma(X^n(\eta(s)))dZ(s),$$

we have that

$$X(t) - X^n(t) = \int_0^t \sigma_1^n(s)\left(X(s) - X^n(s)\right)dZ(s) + \int_0^t \sigma_2^n(s)\left(Z(s) - Z(\eta(s))\right)dZ(s) \tag{2}$$

where

$$\begin{aligned}\sigma_1^n(s) &= \int_0^1 \sigma'(\alpha X(s) + (1-\alpha)X^n(s))d\alpha \\ \sigma_2^n(s) &= \int_0^1 \sigma'(\alpha X^n(s) + (1-\alpha)X^n(\eta(s)))d\alpha\sigma(X^n(\eta(s))).\end{aligned}$$

Given the strong convergence result and assuming smoothness of σ, one has that σ_1^n and σ_2^n converge in the $L^p(C[0,T],\mathbb{R})$-norm to

$$\begin{aligned}\sigma_1(s) &= \sigma'(X(s)) \\ \sigma_2(s) &= \sigma'\sigma(X(s)).\end{aligned}$$

Solving (2), we obtain that

$$\begin{aligned}X(t) - X^n(t) = {} & \mathcal{E}^n(t)\int_0^t \mathcal{E}^n(s)^{-1}\sigma_2^n(s)\left(Z(s) - Z(\eta(s))\right)dZ(s) \\ & - \mathcal{E}^n(t)\int_0^t \mathcal{E}^n(s)^{-1}\sigma_1^n\sigma_2^n(s)\left(Z(s) - Z(\eta(s))\right)ds,\end{aligned}$$

where

$$\mathcal{E}^n(t) = \exp\left(\int_0^t \sigma_1^n(s)dZ(s) - \frac{1}{2}\int_0^t \left(\sigma_1^n(s)\right)^2 ds\right) \tag{3}$$

is the Doleans-Dade exponential, solution of the linear equation

$$\mathcal{E}^n(t) = 1 + \int_0^t \sigma_1^n(s)\mathcal{E}^n(s)dZ(s).$$

Now consider the process

$$\sqrt{n}\int_0^t \left(Z(s) - Z(\eta(s))\right)dZ(s) = \frac{\sqrt{n}}{2}\left(\sum_{i=0}^{j(t)-1}\left(Z(t_{i+1}) - Z(t_i)\right)^2 + \left(Z(t) - Z(\eta(t))\right)^2 - t\right),$$

where $t_{j(t)} = \eta(t)$. Then using Donsker's theorem (see e.g. Billingsley [6] p. 68) we have that

$$\sqrt{n}\int_0^{\cdot} \left(Z(s) - Z(\eta(s))\right)dZ(s) \Longrightarrow Z'$$

where $\sqrt{\frac{2}{T}}\,Z'$ is a Wiener process independent of Z. Furthermore if we consider

$$\left\langle Z, \sqrt{n}\int_0^{\cdot} \left(Z(s) - Z(\eta(s))\right)dZ(s)\right\rangle_t = \sqrt{n}\int_0^t \left(Z(s) - Z(\eta(s))\right)ds$$

we have that this quadratic covariation converges to 0 in L^2. This points to the following convergence

$$\left(Z, \sqrt{n}\int_0^{\cdot} (Z(s) - Z(\eta(s)))\, dZ(s)\right) \Longrightarrow (Z, Z')$$

where Z and $\sqrt{\frac{2}{T}}\, Z'$ are two independent Wiener processes. Therefore one can hint at the following result

$$\sqrt{n}\,(X(t) - X^n(t)) \Longrightarrow \mathcal{E}(t)\int_0^t \mathcal{E}(s)^{-1}\sigma_2(s)dZ'(s),$$

where

$$\mathcal{E}(t) = \exp\left(\int_0^t \sigma_1(s)dZ(s) - \frac{1}{2}\int_0^t (\sigma_1(s))^2\, ds\right)$$

and $(Z, \sqrt{\frac{2}{T}}\, Z')$ is a 2 dimensional Wiener process.

This result in a variety of forms and generalizations has been extensively proved by Jacod, Kurtz and Protter.

In particular, from this result one obtains that for any continuous bounded functional F in $C[0,T]$ one has that $E\left[F\left(\sqrt{n}\,(X - X^n)\right)\right]$ converges to $E\left[F\left(\mathcal{E}(\cdot)\int_0^{\cdot} \mathcal{E}(s)^{-1}\sigma_2(s)dZ'(s)\right)\right]$. On one hand these results give more detail about the limit law of the error process. Nevertheless, this does not give full information about the rate of convergence of various other functionals such as $E(X(t)^p) - E(X^n(t)^p)$, $p_{X(t)}(x) - p_{X^n(t)}(x)$ where p stands for the density function.

For this reason other efforts have been directed into extending the type of convergence into stronger topologies than the one given by weak convergence of processes. In [29], the authors prove that for any continuous bounded functional F and any bounded real variable Y we have that

$$E\left[YF\left(\sqrt{n}\,(X - X^n)\right)\right] \to E\left[YF\left(\mathcal{E}(\cdot)\int_0^{\cdot} \mathcal{E}(s)^{-1}\sigma_2(s)dZ'(s)\right)\right].$$

This type of convergence is called stable convergence in law. It is worth noting that if Y is restricted to a subfiltration this concept also allows the study of the convergence of the conditional expectation of the error process. This type of results are promising but still it does not allow the analysis of the convergence of quantities like the ones mentioned before.

In order to analyze this problem, there is another "parallel" theory called weak approximation that deals particularly with the error

$$E\left[f(X) - f(X^n)\right].$$

The state of the art of this problem is more advanced than the one given previously by the theory of Jacod-Kurtz-Protter. In fact one is able to deal with non bounded, non continuous and even Schwartz distribution functions f (see Guyon [24]). On the other hand one is not able to give precise information on the distribution of the limit error. Just to explain in simple terms the ideas behind this approach, let's explain in simple terms a complex result due to Bally and Talay [2] [3].

To clarify the methodology, we consider a real diffusion process (that is Z is a one dimensional Wiener process)

$$X(t) = x + \int_0^t \sigma(X(s))dZ(s),\ t \in [0,T],$$

and its Euler approximation

$$X^n(t) = x + \int_0^t \sigma(X^n(\eta(s)))dZ(s), t \in [0,T],$$

where $\eta(s) = kT/n$ for $kT/n \le s < (k+1)T/n$. The error process $Y = X - X^n$ solves

$$Y(t) = \int_0^t (\sigma(X(s)) - \sigma(X^n(\eta(s))))dZ(s) = \int_0^t \int_0^1 \sigma'(aX(s) + (1-a)X^n(\eta(s)))da(X(s) - X^n(\eta(s)))dZ(s),$$

this can be written

$$Y(t) = \int_0^t \sigma_1^n(s) Y(s) dZ(s) + G(t), \quad 0 \le t \le T,$$

with

$$\sigma_1^n(s) = \int_0^1 \sigma'(aX(s) + (1-a) X^n(\eta(s))) da$$

$$G(t) = \int_0^t \sigma_1^n(s)(X^n(s) - X^n(\eta(s))) dZ(s) = \int_0^t \sigma_1^n(s) \sigma(X^n(\eta(s)))(Z(s) - Z(\eta(s))) dZ(s).$$

In this simple case we have an explicit expression for Y_t,

$$Y(t) = \mathcal{E}^n(t) \int_0^t \mathcal{E}^n(s)^{-1} (dG(s) - \sigma_1^n(s) d < G, Z >_s)$$

where $\mathcal{E}^n(t)$ is given by (3). Finally we obtain

$$\begin{aligned} Y(t) =& \mathcal{E}^n(t) \int_0^t \mathcal{E}^n(s)^{-1} \sigma_1(s) \sigma(X^n(\eta(s)))(Z(s) - Z(\eta(s))) dZ(s) \\ & - \mathcal{E}^n(t) \int_0^t \mathcal{E}(s)^{-1} \sigma_1^n(s)^2 \sigma(X^n(\eta(s)))(Z(s) - Z(\eta(s))) ds. \end{aligned}$$

Now let f be a smooth function with possibly polynomial growth at infinity. We are interested in obtaining the rate of convergence of $Ef(X(T))$ to $Ef(X^n(T))$. We first write the difference

$$Ef(X(T)) - Ef(X^n(T)) = E\left[\int_0^1 f'(aX_T + (1-a) X^n(t)) da Y(T)\right].$$

Replacing $Y(T)$ by its expression, we obtain with the additional notation $F^n = \displaystyle\int_0^1 f'(aX(T) + (1 - a) X^n(T)) da$,

$$\begin{aligned} Ef(X(T)) - Ef(X^n(T)) =& E\left[F^n \mathcal{E}^n(T) \int_0^T \mathcal{E}^n(s)^{-1} \sigma_1^n(s) \sigma(X^n(\eta(s)))(Z(s) - Z(\eta(s))) dZ(s)\right] \\ & - E\left[F^n \mathcal{E}^n(T) \int_0^T \mathcal{E}^n(s)^{-1} \sigma_1^n(s)^2 \sigma(X^n(\eta(s)))(Z(s) - Z(\eta(s))) ds\right]. \end{aligned} \tag{4}$$

Applying the duality formula for stochastic integrals ($E[< DF, u >_{L^2[0,T]}] = E[F\delta(u)]$ see [51]) where D stands for the stochastic derivative and δ stands for the adjoint of the stochastic derivative, this gives

$$\begin{aligned} Ef(X(T)) - Ef(X^n(T)) \quad = \quad & E\left[\int_0^T D_s(F^n \mathcal{E}^n(T)) \mathcal{E}^n(s)^{-1} \sigma_1^n(s) \sigma(X^n(\eta(s)))(Z(s) - Z(\eta(s))) ds\right] \\ & - E\left[F^n \mathcal{E}^n(T) \int_0^T \mathcal{E}^n(s)^{-1} \sigma_1^n(s)^2 \sigma(X^n(\eta(s)))(Z(s) - Z(\eta(s)))\right]. \end{aligned}$$

Consequently, the difference$Ef(X(T)) - Ef(X^n(T))$ has the simple expression

$$Ef(X(T)) - Ef(X^n(T)) = E\left[\int_0^T U^n(s)(Z(s) - Z(\eta(s))) ds\right],$$

with

$$U^n(s) = (D_s(F^n \mathcal{E}^n(T)) - F^n \mathcal{E}^n(T) \sigma_1(s))(\mathcal{E}^n(s)^{-1} \sigma_1^n(s) \sigma(X^n(\eta(s)))).$$

We finally obtain the rate of convergence by applying once more the duality for stochastic integrals

$$Ef(X_T) - Ef(X_T^n) = E\left[\int_0^T \int_{\eta(s)}^s D_u U^n(s) du ds\right].$$

This last formula ensures that $|Ef(X(T)) - Ef(X^n(T))| \leq CT/n$ and leads to an expansion of $Ef(X(T)) - Ef(X^n(T))$ with some additional work. Furthermore the above argument extends easily in the case that f is an irregular function through the use of the integration by parts formula of Malliavin Calculus.

In other stochastic equations, one cannot explicitly solve the stochastic linear equation satisfied by Y, but in a recent article [10], one can find a general framework that allows treating a great variety of equations. As example we have developed the case of delay equations. The idea explained above appeared for the first time at some workshop proceedings (in [33]) and later was used by various authors (see [20], [22] and [21]) to prove weak approximations errors in other contexts such as the Zakai equation or backward stochastic differential equations.

In fact, the first time this argument appeared in [33], it was just considered as an alternative argument to prove the classical results of weak approximation of Milshtein [43] which are usually obtained through a PDE method. Later it has been shown that in fact this new approach can go beyond the classical proof method. To explain this with a concrete example, we will briefly describe the problem with delay equations which is solved in [10]. Notice that such equations have been introduced in finance by Rogers and Hobson [25] in order to propose a complete model with stochastic volatility. In few words the problem with the Euler approximation for delay equations is that if one tries to use the Milshtein method one gets into infinite dimensional problems quite rapidly and therefore the degree of generalization is quite limited. In fact, consider (see the article of Buckwar and Shardlow [8]) the following one dimensional delay equation

$$dX(t) = \left(\int_{-\tau}^0 X(t+s)dm(s) + b(X(t))\right) dt + \sigma(X(t))dZ(t)$$

where m is a deterministic finite measure on the interval $[-\tau, 0]$ and the initial conditions are $X(s) = x(s)$ for $s \in [-\tau, 0]$.

Consider the integral operator A

$$Ax(t) = \int_{-\tau}^0 x(t+s)dm(s).$$

Then using classical theory of stochastic differential equations in infinite dimensions (an extension of the variation of constants method, see Da Prato-Zabczyk) one obtains

$$X(t) = S(t)x + \int_0^t S(t-s)b(X(s))ds + \int_0^t S(t-s)\sigma(X(s))dZ(s)$$

where S is the semigroup associated with the linear first term in the equation for X.

The natural definition of the Euler scheme is obviously obtained by discretization of the integral in the drift term. That is,

$$X^n(t_{i+1}) = X^n(t_i) + \sum_{j=0}^m X(t_i + s_j)m(s_j, s_{j+1}] + b(X^n(t_i))(t_{i+1} - t_i) + \sigma(X^n(t_i))(Z(t_{i+1}) - Z(t_i))$$

where s_j is a partition of the interval $[-\tau, 0]$ such that $t_i + s_j = t_l$ for some $l \leq i$.

Similarly, one finds that X^n is generated using instead of S the Yoshida approximations to this semigroup. That is the semigroup S^n associated with

$$A^n x(t) = \sum_{j=0}^m x(t+s_j)(m(s_{j+1}) - m(s_j)).$$

Then, as when n tends to infinity $A^n x \longrightarrow Ax$ and $S^n x \longrightarrow Sx$, one expects the strong convergence of the Euler scheme. In order to study the weak errors one has to go further and define the solution of the partial differential equation associated with this problem :

$$u_t(t,y) = \frac{1}{2} u_{x_0 x_0}(t,y)\sigma(x_0)^2 + u_x(t,y)Ax + u_{x_0}(t,y)b(x_0)$$

where $x(0) = x_0$ for $y = (x_0, x) \in \mathbb{R} \times L^2[-\tau, 0]$. The (non-trivial) argument is then similar to the classical Milshtein argument.

Nevertheless, it is also clear from the above set-up that this approach has its limitations. For example, one cannot suppose that there is also a continuous delay in the diffusion coefficient or that the delay term is non-linear.

In comparison, using the method explained previously, one obtains the following result:

Let $(X_t)_{t\in[0,T]}$ be the solution stochastic delay equation :

$$\begin{cases} dX_t &= \sigma\left(\int_{-\tau}^0 X_{t+s} d\nu(s)\right) dZ_t + b\left(\int_{-\tau}^0 X_{t+s} d\nu(s)\right) dt \\ X_s &= \xi_s, s \in [-\tau, 0], \end{cases}$$

where $\tau > 0$, $\xi \in \mathcal{C}([-\tau, 0], I\!R)$ and ν is a finite measure.

We consider the Euler approximation of (X_t) with step $h = \tau/n$

$$\begin{cases} dX_t^n &= \sigma\left(\int_{-\tau}^0 X^n_{\eta(t)+\eta(s)} d\nu(s)\right) dZ_t + b\left(\int_{-\tau}^0 X^n_{\eta(t)+\eta(s)} d\nu(s)\right) dt \\ X_s^n &= \xi_s, s \in [-\tau, 0], \end{cases}$$

with $\eta(s) = \frac{[ns/\tau]}{n/\tau}$, where $[t]$ stands for the entire part of t. We assume that the functions f, σ and b are $\mathcal{C}_b^3$. Then we obtain that

$$Ef(X_T) - Ef(X_T^n) = hC_f + I^h(f) + o(h) \tag{5}$$

where $C_f = C(U^0)$ and $I^h(f) = I^h(U^0)$ are defined in [10]. In particular $|I^h(f)| \leq Ch$ and

$$\begin{aligned} U_s^0 &= \sigma'\left(\int_{-r}^0 X_{s+u} d\nu(u)\right) D_s f'(X_T) + b'\left(\int_{-r}^0 X_{s+u} d\nu(u)\right) f'(X_T) \\ &\quad + \sigma'\left(\int_{-r}^0 X_{s+u} d\nu(u)\right) D_s\left(\int_0^T \theta_t dt\right) + b'\left(\int_{-r}^0 X_{s+u} d\nu(u)\right)\int_s^T \theta_t dt \end{aligned}$$

and θ is the unique solution of

$$\theta_t = \alpha^*\left(J\left(f'(X_T) + \int_0^T \theta_s ds\right)\right)(t) + \beta^*\left(E\left(f'(X_T) + \int_\cdot^T \theta_s ds | \mathcal{F}_\cdot\right)\right)(t)$$

with

$$\begin{aligned} \alpha^*(X)(t) &= E\left(\int_{\max(t-T,-r)}^0 \sigma'\left(\int_{-r}^0 X_{t-u+v} d\nu(v)\right) X_{t-u} d\nu(u) | \mathcal{F}_t\right) \\ \beta^*(X)(t) &= E\left(\int_{\max(t-T,-r)}^0 b'\left(\int_{-r}^0 X_{t-u+v} d\nu(v)\right) X_{t-u} d\nu(u) | \mathcal{F}_t\right). \end{aligned}$$

The above quoted result (5) is an expansion of the error. This result is used in order to increase the rate of convergence of the method using the Romberg extrapolation (see [60] for details in the diffusion case). That is, if $Ef(X_T) - Ef(X_T^n) = hC_f + o(h)$. Then if we define $Y = 2f(X_T^n) - f(X_T^{n/2})$ one obviously obtains that $Ef(X_T) - EY = o(h)$ therefore increasing the rate of convergence of the method. In order to know exactly what has been gained, it is also important to obtain the order of the term $o(h)$ in (5). This can be done using the same method but through tedious work obtaining that in fact, $Ef(X_T) - Ef(X_T^n) = hC_f + O(h^2)$ under enough smoothness conditions on σ and b. In the diffusion case, Pagès [53] addresses variance issues in the context of multi-step Romberg extrapolation.

3 An exact simulation method for one dimensional elliptic diffusions

Recently in two articles by Beskos et.al. [4] [5], an interesting exact method of simulation in dimension one has been introduced. Consider the one dimensional diffusion

$$X(t) = x + \int_0^t b(X(s))ds + \int_0^t \sigma(X(s))dZ(s)$$

where $\sigma(x) \geq c > 0$ for any $x \in \mathbb{R}$ and $\sigma \in C^1(\mathbb{R})$. Then perform the change of variables $Y_t = \eta(X_t)$ where $\eta(z) = \int_x^z \frac{1}{\sigma(u)}du$. By Ito's formula, Y satisfies the following sde:

$$Y(t) = \int_0^t \alpha(Y(s))ds + Z(t)$$

where $\alpha(y) = \frac{b}{\sigma}(\eta^{-1}(y)) - \frac{\sigma'}{2}(\eta^{-1}(y))$. Suppose that we want to compute $E(f(X_T))$. Then using Girsanov's Theorem we have that

$$E(f(X_T)) = E\left[f(B_T)\exp\left(\int_0^T \alpha(B_s)dB_s - \frac{1}{2}\int_0^T \alpha(B_t)^2 dt\right)\right] \tag{6}$$

where B is another Wiener process and here we assume that α is bounded. This idea is usually found when one proves existence of weak solutions for stochastic differential equations.

Next, one defines the function $A(u) = \int_0^u \alpha(y)dy$. With this definition we have, applying Ito's formula, that

$$A(B_T) = \int_0^T \alpha(B_s)dB_s + \frac{1}{2}\int_0^T \alpha'(B_s)ds.$$

Therefore

$$Ef(X_T) = E\left[f(B_T)\exp\left(A(B_T) - \frac{1}{2}\int_0^T \left(\alpha(B_t)^2 + \alpha'(B_t)\right)dt\right)\right].$$

If one was to simulate the above quantity, one would need the whole path of the Wiener process B. In fact this is done in a series of papers by Detemple et. al. [14], [15] and [16] where the Doss-Sussman formula is used to improve the approximation and obtain a scheme which is of strong order one. Instead, Beskos et.al. [5] propose to use a Poisson process to simulate the exponential in the above expression. In fact, one assumes that $\phi(x) = \frac{1}{2}\left(\alpha(x)^2 + \alpha'(x)\right)$ is such that $\forall x \in R$, $0 \leq \phi(x) \leq M$ and introduces a Poisson point process N with intensity $ds \times du$ on $[0,T] \times [0,M]$, independent of B. For any Borel subset S of $[0,T] \times [0,M]$, $N(S)$ is a Poisson random variable with parameter the Lebesgue measure of S. Hence, the random variable $N_1 = N(\{(s,u) \in [0,T] \times [0,M] : 0 \leq u \leq \phi(B(s))\})$ is such that

$$P(N_1 = 0|B) = \exp\left(-\int_0^T \phi(B_s)ds\right).$$

The simulation scheme follows from the equality

$$E(f(X_T)) = E\left[f(B_T)\exp\left(A(B_T)\right)E(1_{\{N_1=0\}}|B)\right] = E\left[f(B_T)\exp\left(A(B_T)\right)1_{\{N_1=0\}}\right].$$

How is the simulation done? First one simulates independent exponential random variables with parameter M say $X_1, ..., X_\nu$ until $\sum_{i=1}^{\nu} X_i > T$. Then one simulates independent random variable $U_i, \ldots, U_{\nu-1}$ uniformly distributed on the interval $[0,M]$. The resulting point process $N = \sum_{i=1}^{\nu-1}\delta_{(X_1+\ldots+X_i, U_i)}$ on $[0,T] \times [0,M]$ is Poisson with intensity $ds \times du$. Now one simulates the independent increments $B(X_1), B(X_1+X_2) - B(X_1), ..., B(T) - B(\sum_{i=1}^{\nu-1} X_i)$ of the Brownian motion and computes $N_1 = \sum_{i=1}^{\nu-1} 1_{\{U_i \leq B(X_1+\ldots+X_i)\}}$.

Obviously there are various issues that have not been considered in this short introduction which rest as open problems or that had already been treated by the authors. Also as it was well known before, the one dimensional case always permits various reductions that do not happen in higher dimensions. For instance, in higher dimensions, the so-called Doss transformation which permits to obtain a SDE with a constant diffusion coefficient is only possible when σ satisfies a restrictive commutativity condition. Notice that under that condition, the discretization scheme obtained by applying the Euler scheme to the SDE with constant diffusion coefficient and making the inverse change of variables is of strong order one (see [14]). Moreover, in higher dimensions the replacement of the stochastic integral in (6) by a standard integral thanks to Itô's formula is only possible when α is a gradient function.

Nevertheless, the one dimensional case always remains as a testing ground for new methodology as it was proven by our recent development in Section 2. And an exact Monte Carlo method for the pricing of Asian options in the Black-Scholes model inspired by the above ideas will be implemented by Jourdain and Sbai [30] in the version 10 of Premia [56].

4 Schemes with high order of convergence

4.1 Stochastic Taylor expansions

In order to make such expansions, it is more convenient to rewrite (1) in Stratonovich form. The interest is that the chain rule holds for Stratonovich integrals. We recall that for a regular adapted one-dimensional process $(H(s))_{s\leq t}$ the Stratonovich integral $\int_0^t H(s)\circ dZ^j(s)$ is equal to the limit in probability of $\sum_i \frac{1}{2}(H(t_{i+1}\wedge t)+H(t_i\wedge t))(Z^j(t_{i+1}\wedge t)-Z^j(t_i\wedge t))$ as $\max_i |t_{i+1}-t_i|$ tends to 0. Hence

$$\int_0^t H(s)\circ dZ^j(s)=\int_0^t H(s)dZ^j(s)+\frac{1}{2}<H,Z^j>_t$$

and (1) writes

$$X(t)=x+\int_0^t \sigma_0(X(s))ds+\sum_{j=1}^r\int_0^t \sigma_j(X(s))\circ dZ^j(s) \tag{7}$$

where $\sigma_0=b-\frac{1}{2}\sum_{j=1}^r \partial\sigma_j\sigma_j$ with $\partial\sigma_j$ denoting the matrix $\left(\frac{\partial\sigma_{ij}}{\partial x_l}\right)_{1\leq i,l\leq d}$ for $\sigma_j=(\sigma_{1j},\ldots,\sigma_{dj})^*$. Let us introduce the differential operators $V_j=\sum_{i=1}^d \sigma_{ij}(x)\partial_{x_i}$ for $0\leq j\leq r$. Since the chain rule holds for Stratonovich integrals, for f a smooth function on $\mathbb{R}^d$,

$$f(X(t))=f(x)+\int_0^t V_0f(X(s))ds+\sum_{j=1}^r\int_0^t V_jf(X(s))\circ dZ^j(s)=f(x)+\sum_{j=0}^r\int_0^t V_jf(X(s))\circ dZ^j(s),$$

where for notational convenience we set $Z^0(s)=s$.

Now remarking that $V_jf(X(s))=V_jf(x)+\sum_{l=0}^r\int_0^s V_lV_jf(X(u))\circ dZ^l(u)$, one obtains

$$f(X(t))=f(x)+\sum_{j=0}^r V_jf(x)\int_0^t \circ dZ^j(s)+\sum_{j,l=0}^r\int_{0\leq u\leq s\leq t} V_lV_jf(X(u))\circ dZ^l(u)\circ dZ^j(s).$$

Iterating the reasoning, one obtains that for any positive integer m,

$$\begin{aligned} f(X(t))=&f(x)+\sum_{k=1}^m\sum_{j_1,\ldots,j_k=0}^r V_{j_1}V_{j_2}\ldots V_{j_k}f(x)Z^{(j_1,\ldots,j_k)}(t)\\ &+\sum_{j_1,\ldots,j_{m+1}=0}^r\int_{0\leq s_1\leq\ldots\leq s_{m+1}\leq t} V_{j_1}\ldots V_{j_{m+1}}f(X(s_1))\circ dZ^{j_1}(s_1)\circ\ldots\circ dZ^{j_{m+1}}(s_{m+1})\end{aligned}$$

where $Z^{(j_1,\ldots,j_k)}(t)=\int_{0\le s_1\le\ldots\le s_k\le t}\circ dZ^{j_1}(s_1)\circ\ldots\circ dZ^{j_k}(s_k)$.
For $1\le j\le r$, $Z^j(s)$ is of order $\sqrt{s}$ while $Z^0(s)=s$ or, in other words, by scaling, $Z^{(j_1,\ldots,j_k)}(t)$ has the same distribution as $t^{(k+\#\{1\le l\le k:j_l=0\})/2}Z^{(j_1,\ldots,j_k)}(1)$. Hence to obtain terms with the same order of magnitude in the above expansion, one has to count the integrals with respect to $Z^0(s)$ twice. That is why for $\alpha=(j_1,\ldots,j_k)\in\mathcal{A}=\bigcup_{l\in N^*}\{0,\ldots,r\}^l$, we set $|\alpha|=k$ and $\|\alpha\|=k+\#\{1\le l\le k:j_l=0\}$. Then we write

$$\begin{aligned}
f(X(t))=&f(x)+\sum_{\alpha:\|\alpha\|\le m}V_{j_1}\ldots V_{j_k}f(x)Z^\alpha(t)+R_{m,f}(t)\text{ where}\\
R_{m,f}(t)=&\sum_{\alpha:|\alpha|\le m,\|\alpha\|>m}V_{j_1}\ldots V_{j_k}f(x)Z^\alpha(t)\\
&+\sum_{j_1,\ldots,j_{m+1}=0}^{r}\int_{0\le s_1\le\ldots\le s_{m+1}\le t}V_{j_1}\ldots V_{j_{m+1}}f(X(s_1))\circ dZ^{j_1}(s_1)\circ\ldots\circ dZ^{j_{m+1}}(s_{m+1}).
\end{aligned}\tag{8}$$

Since the remainder $R_{m,f}(t)$ involves termes scaling like t to a power greater or equal to $(m+1)/2$, the following result (see Proposition 2.1 [42] for $p=1$) is not surprising.

Proposition 2 *When the functions f, b and σ_j are smooth, the remainder $R_{m,f}(t)$ is such that for $p\ge1$, $E(|R_{m,f}(t)|^{2p})^{1/(2p)}\le Ct^{\frac{m+1}{2}}$ where the constant C depends on p, f, b, σ_j and their derivatives.*

4.2 The Milshtein scheme

The Milsthein scheme consists in choosing $f(x)=x$ and $m=2$ in the above expansion (8) and removing the remainder :

$$\forall t\in[t_i,t_{i+1}],\ X^\pi(t)=X^\pi(t_i)+\sum_{j=0}^{r}\sigma_j(X^\pi_{t_i})(Z^j(t)-Z^j(t_i))+\sum_{j,l=1}^{r}\partial\sigma_j\sigma_l(X^\pi_{t_i})(Z^{(l,j)}(t)-Z^{(l,j)}(t_{i+1})).\tag{9}$$

The strong order of convergence of the Milshtein scheme is one (see for instance [32]) :

Theorem 3 *Assume that the functions σ_j and b are C^2 with bounded derivatives. Then for $p\ge1$,*

$$\sup_{t\le T}E\left[\|X(t)-X^\pi(t)\|^p\right]\le C\|\pi\|^p,$$

where the constant C does not depend on the partition π.

To implement the Milshtein scheme, one faces the difficulty usually encountered when trying to construct practical discretization schemes from Taylor expansions : the need to simulate increments of the multiple stochastic integrals which appear. On the one hand, by the fundamental theorem of calculus, for $1\le j\le r$, $Z^{(j,j)}(t)=\int_{0\le u\le s\le t}\circ dZ^j_u\circ dZ^j_s$ is equal to $\frac12(Z^j(t))^2$. But on the other hand, no such nice expression in terms of $Z^j(t)$, $Z^l(t)$ holds for $Z^{(l,j)}(t)$ when $j\ne l$. The generalization of equality $2Z^{(j,j)}(t)=(Z^j(t))^2$ writes $Z^{(l,j)}(t)+Z^{(j,l)}(t)=Z^lZ^j(t)$.
Hence, for the Milshtein scheme to be simulable, one needs the following commutativity condition

$$\mathrm{(C)}\ \ \forall 1\le \mathrm{l}<\mathrm{j}\le \mathrm{r},\ \partial\sigma_\mathrm{j}\sigma_\mathrm{l}=\partial\sigma_\mathrm{l}\sigma_\mathrm{j}$$

which always holds when $r=1$ (single Brownian motion case). Under (C), it is enough to simulate the Brownian increments since the Milshtein scheme writes

$$\begin{aligned}
X^\pi(t_{i+1})&=X^\pi(t_i)+b(X^\pi_{t_i})(t_{i+1}-t_i)+\sum_{j=1}^{r}\sigma_j(X^\pi_{t_i})(Z^j(t_{i+1})-Z^j(t_i))\\
&+\sum_{1\le l<j\le r}\partial\sigma_j\sigma_l(X^\pi_{t_i})(Z^lZ^j(t_{i+1})-Z^lZ^j(t_i))+\frac12\sum_{j=1}^{r}\partial\sigma_j\sigma_j[(Z^j(t_{i+1}))^2-(Z^j(t_i))^2-(t_{i+1}-t_i)].
\end{aligned}$$

In the elliptic case (when $\forall x \in \mathbb{R}^d$, $(\sigma_1(x), \ldots, \sigma_r(x))$ is a basis of $\mathbb{R}^d$), Cruzeiro, Malliavin and Thalmaier [11] have recently proposed a new version of the Milshtein scheme $\tilde{X}^\pi$ which does not involve iterated stochastic integrals of second order (even if commutativity (C) fails). It instead involves a process with values in orthogonal matrices which governs a dynamical rotation of the driving Brownian motion. The solution of $\tilde{X}$ of the equation obtained from (1) by this rotation has the same distribution as X. They prove that $E(\sup_{t\leq T} |\tilde{X}^\pi_t - \tilde{X}_t|^2) \leq C\|\pi\|^2$.

4.3 Schemes with high order of weak convergence

We just saw that the simulation of iterated Brownian integrals of order greater than one is a problem. To overcome this difficulty and obtain simulable schemes with high order of weak convergence, Kusuoka [40] [41] proposed to replace the iterated Brownian integrals which appear in the stochastic Taylor expansion (8) by random variables with the same moments up to order m. See also [47] and [48] where Ninomiya discusses the numerical implementation and efficiency of the resulting scheme and [42] in which Lyons and Victoir propose another scheme based on similar ideas.

Definition 4 *Let $m \in N^*$. A family $(\zeta^\alpha)_{\|\alpha\|\leq m}$ of random variables with finite moments of any order is called m-moment like if*

$$\forall \alpha_1, \ldots, \alpha_k \in \mathcal{A} \text{ such that } \|\alpha_1\| + \ldots + \|\alpha_k\| \leq m,\ E\left[\zeta^{\alpha_1} \ldots \zeta^{\alpha_k}\right] = E\left[Z^{\alpha_1}(1) \ldots Z^{\alpha_k}(1)\right]. \quad (10)$$

The following 5-moment like family in dimension $r = 1$ is given in [48] where other examples are also presented :

Example 5 *Let η be a random variable such that $P(\eta = 0) = \frac{2}{3}$ and $P(\eta = \pm\sqrt{3}) = \frac{1}{6}$. Then one obtains a 5-moment like family in dimension $r = 1$ by setting*

$$\zeta^0 = 1,\ \zeta^1 = \eta,\ \zeta^{(1,1)} = \frac{1}{2}\eta^2,\ \zeta^{(1,0)} = \zeta^{(0,1)} = \frac{1}{2}\eta,\ \zeta^{(1,1,1)} = \frac{1}{6}\eta^3,$$
$$\zeta^{(1,1,0)} = \zeta^{(0,1,1)} = \frac{1}{4},\ \zeta^{(0,0)} = \frac{1}{2},\ \zeta^{(1,1,1,1)} = \frac{1}{8} \text{ and } \zeta^\alpha = 0 \text{ otherwise.}$$

Now replacing the mutiple Brownian integrals by a m-moment like family in (8) written for $f(x) = I(x)$ where $I(x) = x$ denotes the idendity function on R^d, one approximates the law of $X(t)$ by the one of

$$Y^x_t = x + \sum_{\|\alpha\|\leq m} t^{\|\alpha\|/2} V_{j_1} \ldots V_{j_k} I(x) \zeta^\alpha.$$

Let $Q_t f(x) = E(f(Y^x_t))$ denote the corresponding approximation of $E(f(X(t)))$.

Theorem 6 *When the functions f, b and σ_j are smooth,*

$$\left| E(f(X(T))) - Q_{t_1} Q_{t_2 - t_1} \ldots Q_{T - t_{n-1}} f(x) \right| \leq C \sum_{i=0}^{n-1} (t_{i+1} - t_i)^{(m+1)/2}$$

where the constant C depends on f, b, σ_j and their derivatives.

Remark 7
- *For a regular grid $t_i = \frac{iT}{n}$, one has $\sum_{i=0}^{n-1} (t_{i+1} - t_i)^{(m+1)/2} = \frac{T^{(m+1)/2}}{n^{(m-1)/2}}$. Hence the order of weak convergence of the scheme is $(m-1)/2$.*
- *Setting $\zeta^j = Z^j(1)$ for $0 \leq j \leq r$, $\zeta^{(j,j)} = \frac{1}{2}$ for $1 \leq j \leq r$, $\zeta^{(j,l)} = 0$ for $0 \leq j \neq l \leq r$ and $\zeta^{(j,k,l)} = 0$ for $1 \leq j, k, l \leq r$, one obtains a $m = 3$-moment like family. According to the Markov property, this family is such that $Q_{t_1} Q_{t_2 - t_1} \ldots Q_{T - t_{n-1}} f(x) = E(f(X^\pi(T)))$ where X^π denotes the Euler-Maruyama scheme. The order of weak convergence of this scheme is $1 = (3-1)/2$.*

- *Of course, for the numerical approximation of $E(f(X(T)))$, m-moment like families which can be generated on finite probability spaces with as few elements as possible are preferable. For instance, the 5-moment like family given in example 5 can be generated on a probability space with 3 elements supporting the random variable η. For this choice, the exact computation of the approximation $Q_{t_1}Q_{t_2-t_1}\dots Q_{T-t_{n-1}}f(x)$ is possible using a non-recombining trinomial tree with n time-steps and therefore 3^n leaves. When m and r increase, so does the cardinality of the probability space necessary to support a m-moment like family. If exact computation of the approximation is no longer possible, one has to resort to a partial sampling technique such as the Monte-Carlo method (see [48] which is devoted to that issue). An important open problem is how to generate m-moment like families of order higher than 5.*

- *In [40] [41], Kusuoka works under a uniformly non-degeneracy assumption weaker than the uniform Hörmander condition (called the UFG condition) for* (1) *which ensures, thanks to the Malliavin calculus, that for $t > 0$ $x \to E(f(X(t)))$ is smooth in the directions given by the fields generated by the Lie brackets of V_j even if f is not. For the non uniform grid refined near the maturity T*
$$t_i = T\left(1 - \left(\frac{n-i}{n}\right)^{\gamma}\right) \quad \textit{with } \gamma > m,$$
he obtains convergence of the approximation with order $(m-1)/2$ for functions f only C^1 :
$$\left|E(f(X(T))) - Q_{t_1}Q_{t_2-t_1}\dots Q_{T-t_{n-1}}f(x)\right| \le \frac{C\|\nabla f\|_\infty}{n^{(m-1)/2}}.$$

Proof. Setting $P_t f(x) = E(f(X(t)))$, one has the following decomposition of the error
$$\begin{aligned}|E(f(X(T)) - Q_{t_1}\dots Q_{T-t_{n-1}}f(x)| \le & |E(P_{T-t_{n-1}}f(X(t_{n-1}))) - Q_{t_1}\dots Q_{t_{n-1}-t_{n-2}}P_{T-t_{n-1}}f(x)| \\ & + |Q_{t_1}Q_{t_2-t_1}\dots Q_{t_{n-1}-t_{n-2}}(P_{T-t_{n-1}}f - Q_{T-t_{n-1}}f)(x)|\end{aligned} \tag{11}$$

We now prove the result by induction on n. For $n = 1$, one deals with $P_t f(z) - Q_t f(z)$ which by (8) is equal to
$$E\left[f\left(z + \sum_{\|\alpha\|\le m} V_{j_1}\dots V_{j_k}I(z)Z^\alpha(t) + R_{m,I}(t)\right) - f\left(z + \sum_{\|\alpha\|\le m} t^{\|\alpha\|/2}V_{j_1}\dots V_{j_k}I(z)\zeta^\alpha\right)\right].$$

Assuming for simplicity that $d = 1$ and making a standard Taylor expansion of the function f in the neighborhood of z, one deduces that $P_t f(z) - Q_t f(z)$ is equal to
$$\sum_{k=1}^{m}\frac{f^{(k)}(z)}{k!}E\left[\left(\sum_{\|\alpha\|\le m} V_{j_1}\dots V_{j_k}I(z)Z^\alpha(t) + R_{m,I}(t)\right)^k - \left(\sum_{\|\alpha\|\le m} t^{\|\alpha\|/2}V_{j_1}\dots V_{j_k}I(z)\zeta^\alpha\right)^k\right] + \mathcal{O}(t^{\frac{m+1}{2}}).$$

According to Proposition 2, $R_{m,I}(t)$ scales like $t^{(m+1)/2}$. Now developping the powers and using (10) one obtains that the expectation of all terms scaling like $t^{l/2}$ with $l \le m$ vanish. Therefore $|P_t f(z) - Q_t f(z)| \le Ct^{(m+1)/2}$ and the induction hypothesis holds for $n = 1$.
We now assume that the induction hypothesis holds at rank $n-1$. Since when f is smooth, so is $P_{T-t_{n-1}}f$, we deduce that the first term of the right-hand-side of (11) is smaller than $C\sum_{i=0}^{n-2}(t_{i+1} - t_i)^{(m+1)/2}$. By the result proved for $n = 1$, $\|P_{T-t_{n-1}}f - Q_{T-t_{n-1}}f\|_\infty \le C(T-t_{n-1})^{(m+1)/2}$. Since for all $t \ge 0$, $\|Q_t g\|_\infty \le \|g\|_\infty$, one deduces that the second term of the r.h.s. of (11) is smaller than $C(T - t_{n-1})^{(m+1)/2}$. This concludes the proof. ■

Let us briefly present the related approximation based on the notion of cubature proposed by Lyons and Victoir [42].

Definition 8 *Let $m \in N^*$ and $t > 0$. Continuous paths $\omega_{t,1}, \ldots, \omega_{t,N}$ with bounded variation from $[0,t]$ to R^r and positive weights $\lambda_1, \ldots, \lambda_N$ such that $\sum_{l=1}^N \lambda_l = 1$ define a cubature formula with degree m at time t if*

$$\forall \alpha = (j_1, \ldots, j_k) \in \mathcal{A} \text{ such that } \|\alpha\| \le m, \; E(Z^\alpha(t)) = \sum_{l=1}^N \lambda_l \int_{0 \le s_1 \le \ldots \le s_k \le t} d\omega_{t,l}^{j_1}(s_1) \ldots d\omega_{t,l}^{j_k}(s_k) \tag{12}$$

where $\omega_{t,l}^j(s)$ denotes the j-th coordinate of $\omega_{t,l}(s)$ when $1 \le j \le d$ and $\omega_{t,l}^0(s) = s$.

According to [42], there exists a cubature with degree m at time 1 such that N is smaller than the cardinality of $\{\alpha \in \mathcal{A} : \|\alpha\| \le m\}$. Moreover, one deduces a cubature of degree m at time t by scaling. For $l \in \{1, \ldots, N\}$ let $(y_{t,l}(s,x))_{s \le t}$ denote the solution of the ODE :

$$y_{t,l}(0,x) = x \text{ and } \forall s \in [0,t], \; dy_{t,l}(s,x) = \sum_{j=0}^r \sigma_j(y_{t,l}(s,x)) d\omega_{t,l}^j(s).$$

Lyons and Victoir propose to approximate $E(f(X(t))$ by $Q_t f(x) = \sum_{l=1}^N \lambda_l f(y_{t,l}(t,x))$. Theorem 6 still holds with this new definition of $Q_t f$. The proof is based on a similar decomposition of the error but the analysis of $E(f(X(t)) - Q_t f(x)$ is easier. Indeed the Taylor expansion (8) holds for Z replaced by $\omega_{t,l}$. Multiplying by λ_l, summing over l and substracting (8) then using Proposition 2 and (12) , one obtains that $|E(f(X(t)) - Q_t f(x)| \le C t^{(m+1)/2}$.

Let us finally mention, interesting schemes with high weak order of convergence recently proposed by Ninomiya and Victoir [50] and Fujiwara [17]. Even if the idea of these schemes also comes from stochastic Taylor expansions, their implementation is different from the previous ones. It requires a sequence of independent uniform random variables $(U_i)_{1 \le i \le n}$ independent from $(Z^1, \ldots, Z^r)$. For $\theta \in N^*$, to go from $\bar{X}_\theta^\pi(t_i)$ to $\bar{X}_\theta^\pi(t_{i+1})$, one repeats the following steps for $k \in \{1, \ldots, \theta\}$

1. integrate the ordinary differential equation $\frac{d}{dt}x(t) = \sigma_0(x(t))$ on an interval with length $(t_{i+1} - t_i)/2\theta$,
2. depending on whether $U_{i+1} \le \frac{1}{2}$ or not, integrate successively for j increasing from 1 to r or for j decreasing from r to 1 the ODE $\frac{d}{dt}x(t) = \sigma_j(x(t))$ on an interval with random length $Z^j(t_i^k) - Z^j(t_i^{k-1})$ where $t_i^k = t_i + k(t_{i+1} - t_i)/\theta$.
3. do the first step again.

Ninomiya and Victoir [50] prove that $\bar{X}_1^\pi(T)$ is an approximation of $X(T)$ with weak order of convergence 2. The idea of Fujiwara [17] is to make Romberg-like extrapolations in order to improve the weak rate of convergence. Indeed, he proves that $E(f(X(T)))$ is respectively approximated by $\frac{1}{3}E(4f(\bar{X}_2^n(T)) - f(\bar{X}_1^n(T)))$ and $\frac{1}{120}E(243f(\bar{X}_3^n(T)) - 128f(\bar{X}_2^n(T)) + 5f(\bar{X}_1^n(T)))$ with order of convergence 4 and 6. When, for some of the above ODEs, no analytical expression of the solution is available, one has to resort to discretization schemes. Those schemes have to be chosen carefully in order to preserve the weak order for the resulting scheme for (1). For instance, Fujiwara suggests a Runge-Kutta scheme with order 13 to preserve the weak order 6. More recently, Ninomiya and Ninomiya [49] have proposed a scheme with weak order 2 in which, for each time-step, only two ordinary differential equations have to be integrated on a random time-horizon. They have also analysed the effect in terms of weak error of the resort to Runge-Kutta schemes to integrate the ordinary differential equations. The schemes with weak order 2 proposed by Ninomiya and Victoir and by Ninomiya and Ninomiya are both implemented in Premia [56] for the pricing of Asian options under the Heston model of stochastic volatility.

5 Comments on some extensions

We discuss first the case when Z is a Lévy process. That is, process with independent and stationary increments with characteristic function given by

$$E\left[\exp\left(i\left\langle\theta, Z(t)\right\rangle\right)\right] = \exp\left(-\frac{1}{2}\left\langle\theta, \Gamma\theta\right\rangle t + i\left\langle b, \theta\right\rangle t + \int_{\mathbb{R}^r}\left(\exp(i\left\langle\theta, x\right\rangle) - 1 - i\theta x 1\left\{x \leq 1\right\}\right)\nu(dx)\right)$$

where $\theta \in \mathbb{R}^r$, $\Gamma \in \mathbb{R}^{r\times r}$ is a symmetric non-negative matrix and ν is a measure satisfying $\int_{\mathbb{R}^r}\left(1 \wedge |x|^2\right)\nu(dx) < \infty$. When $b = \nu = 0$ and Γ is the identity matrix, then Z is a standard r-dimensional Wiener process. The constant b denotes the drift of the process and ν is the Lévy measure associated to the process Z. We note that in comparison with the Wiener process case not all moments of Z are finite. In fact the moment of order k of Z is finite if $\int_{\mathbb{R}^r}|x|^k 1_{\{x\geq 1\}}\nu(dx) < \infty$.

The existence and uniqueness of the above equation (1) is ensured by standard theorems that can be found in e.g. Protter [57] under Lipschitz assumptions on the coefficients b and σ. Nevertheless it is not clear under which conditions the moments of the solution are finite if Z is a Lévy process, except for the case of bounded coefficients.

In particular, we do not know how the finite moment property transfers from Z into X when the coefficients are Lipschitz. These properties are important in order to determine the convergence properties of the Euler scheme. The situation in the case that σ is constant is already difficult enough. Nevertheless, this is an interesting problem.

We quote here some results of the article Kohatsu-Yamazato [34] who study this problem in the particular case that σ is constant.

For example, consider for simplicity the one dimensional case $r = d = 1$ with $\Gamma = 0$, $b = 0$ and ν a measure concentrated on $(0, \infty)$. The moment $E(X(t)^\beta)$ is finite or not depending on whether the integral with respect to ν in the last column is finite or not.

$b(y) = y^\alpha$	β	Criterion for finiteness of $E(X(t)^\beta)$
$0 \leq \alpha \leq 1$	$\beta > 0$	$\int_1^{+\infty} y^\beta \nu(dy)$
$\alpha > 1$	$0 < \beta < \alpha - 1$	always finite
$\alpha > 1$	$\beta = \alpha - 1$	$\int_1^{+\infty} \log(y)\,\nu(dy)$
$\alpha > 1$	$\beta > \alpha - 1$	$\int_1^{+\infty} y^{\beta-\alpha+1}\nu(dy)$

In the same lines of the above table, but in another set up, Grigoriu-Samorodnistsky [23] studied the tail behavior of $X(t)$. In either case the conclusions are similar.

The rule seems to be that if the drift coefficient is sublinear then the drift does not influence the finite moment property of Z and it transfers directly to X. If the drift is superlinear then the situation is different. That is, the finite moment property depends on the difference of power between the drift and the moment to be evaluated. Therefore, it can be conjectured that this is the situation in the Lipschitz cases.

Currently, as far as our knowledge goes, it is not known if X has finite moments even if the exponential moments of Z are bounded unless one imposes a series of stringent conditions. In most papers found in the literature, besides this assumption, one also has to make the assumption that the moments of X are bounded which is an unaccomplished feature of this problem. For example one has that

Theorem 9 *Suppose that Z has exponential moments and that X has finite moments. Then*

$$E\left[\sup_{t\leq T}\left\|X(t) - X^\pi(t)\right\|^{2p}\right] \leq C\left\|\pi\right\|^p$$

where the constant C depends on T, x and the Lipschitz constants.

One remarkable different case from the discussion in this paper is the situation of reflecting stochastic differential equations. In general, if the domain is closed and convex and the reflection is normal, then the results can be usually obtained as generalizations of the non-reflecting case. The main difference lies in how the inequalities are obtained. In fact, instead of using strong type inequalities directly on the error process $X(t) - X^\pi(t)$, one has to use Ito's formula and the fact that contribution of the reflecting processes brings X^n closer to X. If the domain is more general then the results are no longer valid. In fact, as proven by Pettersson [54] (later refined by Slominski [59]) the rates can decay slightly depending on the properties of the domain.

The latest refined results on this can be found in a recent thesis by S. Menozzi [44]. Nevertheless there is no parallel theory in the style of Jacod-Kurtz-Protter.

In finance, one-dimensional processes that remain non-negative are of particular interest. This non-negativity property comes from the choice of the coefficients in the SDE rather than from reflection. The typical example is the Cox-Ingersoll-Ross process :

$$X(t) = x + \int_0^t (a - kX(s))ds + \sigma \int_0^t \sqrt{X(s)}dZ(s), \quad \text{with } x, a, \sigma \geq 0 \text{ and } k \in \mathbb{R},$$

which is used to model short interest rates but also stochastic volatility in the Heston model. It is not possible to discretize this SDE by the standard Euler scheme : indeed, $X^\pi(t_1)$ is negative with positive probability and then it is not possible to compute the square root in the diffusion coefficient in order to define $X^\pi(t_2)$. To overcome this problem, Deelstra and Delbaen [12], propose to take the positive part before the square root and define recursively :

$$X^\pi(t_{i+1}) = X^\pi(t_i)(1 - k(t_{i+1} - t_i)) + a(t_{i+1} - t_i) + \sigma\sqrt{(X^\pi(t_i))^+}\,(Z(t_{i+1}) - Z(t_i)).$$

In her thesis [13], Diop studies the symmetrized Euler scheme defined by

$$X^\pi(t_{i+1}) = \left| X^\pi(t_i)(1 - k(t_{i+1} - t_i)) + a(t_{i+1} - t_i) + \sigma\sqrt{X^\pi(t_i)}\,(Z(t_{i+1}) - Z(t_i)) \right|.$$

In [1], Alfonsi compares those schemes with some new ones that he proposes. He concludes that the following explicit scheme combines the best features when $a \geq \frac{\sigma^2}{4}$:

$$X^\pi(t_{i+1}) = \left((1 - \frac{k}{2}(t_{i+1} - t_i))\sqrt{X^\pi(t_i)} + \frac{\sigma(Z(t_{i+1}) - Z(t_i))}{2(1 - \frac{k}{2}(t_{i+1} - t_i))} \right)^2 + (a - \frac{\sigma^2}{4})(t_{i+1} - t_i).$$

This scheme has been implemented in Premia [56] in order to discretize the SSRD model of credit risk [7].

Another interesting issue is the discussion about adaptive methods. That is, how to choose the partition as to improve the first term in the expansion of the strong error. For this, we refer the reader to [9] and subsequent articles [26] and [46].

References

[1] Alfonsi, A. On the discretization schemes for the CIR (and Bessel squared) processes. Monte Carlo Methods Appl. 11(4):355-384, 2005.

[2] Bally, V. and Talay, D. The law of the Euler scheme for stochastic differential equations (II) : convergence rate of the density. Monte Carlo Methods and Applications, 2:93-128, 1996.

[3] Bally, V. and Talay, D. The law of the Euler scheme for stochastic differential equations (I) : convergence rate of the distribution function. Probability Theory and Related Fields, 104:43-60, 1995.

[4] Beskos, A. and Roberts G.O. Exact Simulation of Diffusions, Ann. Appl. Probab., 15(4):2422-2444, 2005.

[5] Beskos, A., Papaspiliopoulos, O. and Roberts G. O. Retrospective exact simulation of diffusion sample paths with applications. Bernoulli 12(6):1077-1098, 2006.

[6] Billingsley, P. Convergence of Probability Measures. John Wiley and Sons, 1968.

[7] Brigo, D. and Alfonsi, A. Credit default swap calibration and derivatives pricing with the SSRD stochastic intensity model. Finance Stoch. 9(1):29-42, 2005.

[8] Buckwar, E. and Shardlow, T. Weak approximation of stochastic differential delay equations. IMA J. Numerical Analysis, 25:57-86, 2005.

[9] Cambanis, S. and Hu, Y. Exact convergence rate of the Euler-Maruyama scheme, with application to sampling design. Stochastics Stochastics Rep. 59(3-4):211-240, 1996.

[10] Clement, E., Kohatsu-Higa, A. and Lamberton, D. A duality approach for the weak approximation of stochastic differential equations. Ann. Appl. Probab. 16(3):1124-1154, 2006.

[11] Cruzeiro, A.B., Malliavin, P. and Thalmaier, A.: Geometrization of Monte-Carlo numerical analysis of an elliptic operator: strong approximation. C. R. Math. Acad. Sci. Paris 338(6):481-486, 2004.

[12] Deelstra, G. and Delbaen, F. Convergence of discretized stochastic (interest rate) processes with stochastic drift term. Appl. Stochastic Models Data Anal. 14(1):77-84, 1998.

[13] Diop, A. Sur la discrétisation et le comportement à petit bruit d'EDS unidimensionnelles dont les coefficients sont à dérivées singulières. Thèse université de Nice Sophia-Antipolis, 2003.

[14] Detemple, J., Garcia, R., M. Rindisbacher. Asymptotic properties of Monte Carlo estimators of diffusion processes, Journal of Econometrics: 134(1):349-367, 2005.

[15] Detemple, J., Garcia, R., and M. Rindisbacher, "Representation Formulas for Malliavin Derivatives of Diffusion Processes", Finance and Stochastics, 9(3):349-367, 2005.

[16] Detemple, J., Garcia, R., and M. Rindisbacher, "Asymptotic Properties of Monte Carlo Estimators of Derivatives", Management Science, 51:1657-1675, 2005.

[17] Fujiwara, T. Sixth order method of Kusuoka approximation. Preprint, 2006.

[18] Gobet, E. LAMN property for elliptic diffusion: a Malliavin Calculus approach. Bernoulli 7:899-912, 2001.

[19] Gobet E. LAN property for ergodic diffusions with discrete observations. Ann. Inst. H. Poincare Probab. Statist. 38:711-737, 2002.

[20] Gobet E. and Munos, R. Sensitivity analysis using Itô-Malliavin calculus and martingales. Application to stochastic optimal control. SIAM Journal on Control and Optimization, 43(5):1676-1713, 2005.

[21] Gobet, E., Pages, G., Pham, H. and Printemps, J. Discretization and simulation of the Zakai equation. SIAM J. Numer. Anal. 44(6):2505-2538, 2006.

[22] Gobet E. and Labart C. Error expansion for the discretization of Backward Stochastic Differential Equations. Stochastic Process. Appl. 117(7):803–829, 2007.

[23] Grigoriu M. and Samorodnitsky G. Tails of solutions of certain nonlinear stochastic differential equations driven by heavy tailed Levy processes. Stochastic Proc. Appl. 105:69-97, 2003.

[24] Guyon, J. Euler scheme and tempered distributions. Stochastic Process. Appl. 116(6)877-904, 2006.

[25] Hobson, D.G. and Rogers L.C.G. Complete models with stochastic volatility. Math. Finance 8(1):27-48, 1998.

[26] Hofmann, N., Müller-Gronbach, T. and Ritter, K. Optimal approximation of stochastic differential equations by adaptive step-size control. Math. Comp. 69(231):1017-1034, 2000.

[27] Ikeda, N. and Watanabe, S.: Stochastic Differential Equations and Diffusion Processes. Amsterdam Oxford New York: North-Holland/Kodansha 1989.

[28] Jacod, J., Kurtz, T., Méléard, S. and Protter, P. The approximate Euler method for Levy driven stochastic differential equations. Ann. Inst. H. Poincare, Probab. Statist. 41(3):523-558, 2005.

[29] Jacod, J. and Protter, P. Asymptotic error distributions for the Euler method for stochastic differential equations. Ann. Probab. 26(1):267-307, 1998.

[30] Jourdain, B. and Sbai, M. Exact retrospective Monte Carlo computation of arithmetic average Asian options. Monte Carlo Methods Appl. 13(2):135-171, 2007.

[31] Kebaier, A. Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing. Ann. Appl. Probab. 15(4):2681-2705, 2005.

[32] Kloeden, P. E. and Platen, E. Numerical Solution of Stochastic Differential Equations. Springer Berlin Heidelberg New York, 1992.

[33] Kohatsu-Higa, A. and Pettersson, R. On the simulation of some functionals of diffusion processes. Proceedings of the “Fourth Workshop on Stochastic Numerics”. Research Institute for the Mathematical Sciences, Kyoto 2000.

[34] Kohatsu-Higa, A. and Yamazato, M. On moments and tail behaviors of storage processes. Journal of Applied Probability, 20:1069-1086, 2003.

[35] Kurtz, T. and Protter, P. Weak convergence of stochastic integrals and differential equations. II. Infinite-dimensional case. Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995), Lecture Notes in Math. 1627:197-285, Springer, Berlin, 1996.

[36] Kurtz, T. and Protter, P. Weak convergence of stochastic integrals and differential equations. Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995), Lecture Notes in Math. 1627:1-41, Springer, Berlin, 1996.

[37] Kurtz, T. and Protter, P. Characterizing the weak convergence of stochastic integrals. Stochastic analysis (Durham, 1990), , London Math. Soc. Lecture Note Ser. 167:255-259, Cambridge Univ. Press, Cambridge, 1991.

[38] Kurtz, T. and Protter, P. Wong-Zakai corrections, random evolutions, and simulation schemes for SDEs. Stochastic analysis, 331-346, Academic Press, Boston, MA, 1991.

[39] Kurtz, T. and Protter, P. Weak limit theorems for stochastic integrals and stochastic differential equations. Ann. Probab. 19(3): 1035-1070, 1991.

[40] Kusuoka, S. Approximation of expectation of diffusion process and mathematical finance. Taniguchi Conference on Mathematics Nara '98, Adv. Stud. Pure Math. 31:147-165, Math. Soc. Japan, Tokyo, 2001.

[41] Kusuoka, S.: Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus. Adv. Math. Econ., 6:69-83, Springer, Tokyo, 2004.

[42] Lyons, T. and Victoir, N.: Cubature on Wiener space. Stochastic analysis with applications to mathematical finance. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 460:169-198, 2004.

[43] Milshtein, G.N.: A method of second order accuracy integration of stochastic differential equations. Theory Probab. Appl. 23:396-401, 1978.

[44] Menozzi, S. Discrétisations associées à un processus dans un domaine et schémas numériques probabilistes pour les EDP paraboliques quasilinéaires. Thèse université Pierre et Marie Curie - Paris VI, 2004.

[45] Mikulevicius, R. and Platen, E. Time discrete Taylor approximations for Ito processes with jump component. Math. Nachr. 138(9):93-104, 1988.

[46] Müller-Gronbach, T. Optimal pointwise approximation of SDEs based on Brownian motion at discrete points. Ann. Appl. Probab. 14(4):1605-1642, 2004.

[47] Ninomiya, S. A new simulation scheme of diffusion processes : application of the Kusuoka approximation to finance problems. Math. Comput. Simulation 62:479-486, 2003.

[48] Ninomiya, S. A partial sampling method applied to the Kusuoka approximation. Monte Carlo Methods Appl. 9(1):27-38, 2003.

[49] Ninomiya, M. and Ninomiya, S.: A new higher-order weak approximation scheme for stochastic differential equations and the Runge-Kutta method. Finance and Stochastics 13:415-443, 2009.

[50] Ninomiya, S. and Victoir, N. Weak approximation of stochastic differential equations and application to derivative pricing. Applied Mathematical Finance, 15(2):107-121, 2008.

[51] Nualart, D. The Malliavin Calculus and Related Topics. Springer Berlin Heidelberg New York, 1995.

[52] Nualart, D. Analysis on Wiener space and anticipating stochastic calculus. Lecture Notes in Math., 1690:123-227, 1998.

[53] Pagès, G. Multi-step Richardson-Romberg Extrapolation: Remarks on Variance Control and complexity. Monte Carlo Methods Appl. 13(1):37-70, 2007.

[54] Pettersson, R. Approximations for stochastic differential equations with reflecting convex boundaries. Stochastic Proc. Appl. 59:295-308, 1995.

[55] Platen, E. An approximation method for a class of Ito processes with jump component. Litovsk. Mat. Sb. 22(2):124-136, 1982.

[56] PREMIA : An Option Pricer, Mathfi Project (INRIA, CERMICS, UMLV). *http://www-rocq.inria.fr/mathfi/Premia/*

[57] Protter, P. E. Stochastic integration and differential equations. Applications of Mathematics (New York) 21. Stochastic Modelling and Applied Probability. Springer-Verlag, 2004.

[58] Słomiński, L. On approximation of solutions of multidimensional SDEs with reflecting boundary condition. Stochastic Proc. Appl. 50:197-219, 1994.

[59] Słomiński, L. Euler approximation of solutions of SDEs with reflecting boundary. Stochastic Proc. Appl. 94:317-337, 2001.

[60] Talay, D. and Tubaro, L. Expansion of the global error for numerical schemes solving stochastic differential equations. Stochastic Anal. Appl. 8(4):483-509, 1990.