INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 5, Number 1, Pages 55–72

HOW RATES OF L^p-CONVERGENCE CARRY OVER TO NUMERICAL APPROXIMATIONS OF SOME CONVEX, NON-SMOOTH FUNCTIONALS OF SDES

HENRI SCHURZ

Abstract. The relation between weak and p-th mean convergence of numerical methods for integration of some convex, non-smooth and path-dependent functionals of ordinary stochastic differential equations (SDEs) is discussed. In particular, we answer how rates of p-th mean convergence carry over to rates of weak convergence for such functionals of SDEs in general. Assertions of this type are important for the choice of approximation schemes for discounted price functionals in dynamic asset pricing as met in mathematical finance and other commonly met functionals such as passage times in engineering.

Key Words. stochastic differential equations, approximation of convex and path-dependent functionals, numerical methods, stability, L^p-convergence, weak convergence, rates of convergence, non-negativity, discounted price functionals, asset pricing, approximation of stochastic exponentials

1. Introduction

Suppose that the risky asset price $(X(t))_{t\geq 0}$ is governed by systems of Itô-type stochastic differential equations (SDEs) such as noisy ordinary differential equations

$$dX(t) = a(t, X(t))dt + \sum_{j=1}^{m} b^j(t, X(t))dW_j(t) \tag{1}$$

driven by Wiener processes or martingale-type noises W_j with respect to the forward filtration $(\mathcal{F}_t)_{t\geq 0}$ on the complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For some overview on the theory of SDEs, e.g. see Arnold [2], Gard [10], Oksendal [20] or Protter [21].

One obviously knows that the construction of efficient numerical approximations of path-dependent functionals F of X such as discounted price functionals

$$F_{r,p,X}(t,T) \;=\; \mathbb{E}\left[\exp\Big(-\int_t^T r(s)ds\Big)p((X(s))_{0\leq s\leq T})\Big|\mathcal{F}_t\right] \tag{2}$$

at exercise times $0 \leq t \leq T$ (T time of maturity) is important in the theory of dynamic asset pricing. Here $(r(t))_{t\geq 0} \geq 0$ is interpreted as an interest rate and p as a Borel-measurable functional on the risky asset price $(X(t))_{t\geq 0}$. The simplest and most cited example in finance is that of constant nonrandom interest rate r (or r satisfying SDEs such as (1)) and non-differentiable, but convex pricing functional

$$p((X(s))_{0\leq s\leq T}) = (X(T) - K)_+ \tag{3}$$

Received by the editors September 19, 2006 and, in revised form, December 12, 2006.
2000 *Mathematics Subject Classification.* 65C30, 65L20, 65D30, 34F05, 37H10, 60H10.

where K is the striking price, T time of maturity and $(.)_+$ denotes the nonnegative part of inscribed expression. This occurs in the European call and put options. Others are given by look-back, Russian and Asian options.

Unfortunately, there are only very few models which allow to compute the price functionals F analytically. So one has to resort to numerical techniques to approximate F and p in general. Many authors have dealt with methods for numerical integration of solutions of Itô-type SDEs (1) and its functions $F(X(T))$ at fixed terminal time T. For example, see Allen [1], Artemiev and Averina [3], Bouleau and Lépingle [6], Gard [10], Kloeden, Platen and Schurz [14], Milstein [17], Schurz [25, 27, 31], Talay [33, 34, 35], Wagner and Platen [36]. Almost all of their methods are based on the classic Taylor expansion and its Runge-Kutta-type substitutions. However, most of those methods sometimes lack of rigorous statements on stability, positivity and convergence when complex nonlinearities, convexity or path-dependence in F are present.

The aim of this paper is to show how one can have a "minimal guarantee" of convergence and qualitative justification of numerical integration techniques which are needed to approximate functionals F such as given by (2) or similar ones under non-smooth assumptions or path-dependence. For this purpose, we shall exploit known and more easily verifiable facts on L^p-convergence rates. There are several good reasons why we prefer to use nonstandard implicit, strongly converging methods as originally introduced in [25, 18], studied in [23, 30] and continued by [19], [11], or even for quasilinear random PDE by [5]. Their good stability, boundary and positivity behavior is one of them. We shall justify these methods by studying how the rates of p-th mean convergence carry over to the rates of weak convergence along some functionals F despite non-smoothness or path-dependence. In particular, some new proof techniques come up by using integral representations of convex functions involving positive Radon measures. Another advantage is seen by the fact that we do not need to suppose very restrictive assumptions on the smoothness and boundedness of the coefficients of underlying SDEs as commonly met in the literature on stochastic numerics. This paper exhibits supplemental remarks to the results presented in Kanagawa and Ogawa [13] and Talay [33, 34]. Moreover, we do not focus too much on fairly known results which are supposed to be known to the readership. See Allen [1], Schurz [27], Talay [34] or appendix A for a quick overview on basic facts related to stochastic-numerical analysis.

The paper is organized as follows. Section 2 discusses how convergence rates of L^p-approximations carry over to rates of weak approximations while dealing with functionals involving convex functions. These estimates are only advantageous when not so much smoothness can be imposed on the functional F and its ingredients r, p and X (in contrast to standard requirements such as $p \in C^\infty$, p continuously differentiable or Lipschitz-continuous drift and diffusion coefficients of r and X). See [17, 34] for approximation rates of very smooth functions $F(X(T))$ (actually they only consider functions $F(X(T))$, not real functionals) or those F with non-degenerate infinitesimal generator of price process X. Section 3 proves a general theorem to control the total L^1-approximation error of path-dependent functionals such as F in (2). We also state a theorem on convergence of Hölder-continuous functionals. Eventually, we list numerous examples of functionals involving convex structures in Section 4. An appendix (Sections A.1 - A.3) resumes basic facts on numerical methods for Itô SDEs and its concepts convergence to increase the understanding of a more general audience. All in all, this paper presents just a supplemental discussion on some more complex issues related to numerical approximation of not-so-smooth functionals of strong solutions of SDEs with not necessarily very smooth drift or diffusion coefficients.

2. Approximation of convex functionals of SDEs

An interesting question related to approximation of functionals is how the p-th mean convergence orders can be carried over to the weak convergence order during the approximation of functionals of SDE solutions. One important aim is to approximate the functional

$$F_\rho(t,X) \;=\; \mathbb{E}\,[f(t,X(t),\inf_{\rho\le s\le t}\|X(s)\|,\sup_{\rho\le s\le t}\|X(s)\|)|\mathcal{F}_s] \tag{4}$$

where $f = f(t,x,y,z)$ is a Borel-measurable function at t,x,y,z and $0 \le \rho \le t \le T$ (or even more general functionals). Let us suppose that T is nonrandom throughout this paper (despite of the fact that some of the herein presented results can be generalized to finite stopping time $\tau = T$ which are uniformly bounded).

2.1. A review on some properties of convex functions. Here we assemble some of the most useful properties of convex functions. First, recall this definition.

Definition 2.1. *A set $A \subset \mathbb{R}^d$ is called* **convex** *iff*

$$\forall x,y\in A\ \forall\lambda\in[0,1] \Rightarrow \lambda x+(1-\lambda)y\in A \quad .$$

A function $f:\mathbb{R}\to\mathbb{R}$ is said to be **convex** *on domain $\mathbb{D}\subset\mathbb{R}$ iff*

$$\forall x,y\in\mathbb{D}\ \forall\lambda\in[0,1] \Rightarrow f(\lambda x+(1-\lambda)y)\le\lambda f(x)+(1-\lambda)f(y) \quad .$$

For the statement of some properties, let f : $\mathbb{R}\to\mathbb{R}$ be a convex function on the open set $\mathbb{D}\subset\mathbb{R}$. Let $x,y\in\bar{\mathbb{D}}$. Then one can verify the following properties.

(i) $\quad S(x,y)=S(y,x)=\dfrac{f(y)-f(x)}{y-x}$ is increasing in y, x fixed

(ii) $\quad \exists f'_-(x)=\lim\limits_{h\uparrow 0}\dfrac{f(x+h)-f(x)}{h}, \exists f'_+(x)=\lim\limits_{h\downarrow 0}\dfrac{f(x+h)-f(x)}{h}$

(iii) $\quad f(y)\ge f(x)+f'_+(x)(y-x)$ for all $x\in\overset{o}{\mathbb{D}}, y\in\bar{\mathbb{D}}$

(iv) $\quad f(x^*)\le f(x)+S(x,y)(x^*-y)$ for all $x^*\in[x,y]$

(v) $\quad f$ convex on $\mathbb{D}$ $\iff$ $\forall x,y,z\in\mathbb{D}; x<z<y : S(x,z)\le S(x,y)\le S(z,y)$

(vi) $\quad f(z)=\sup\limits_{x\in\mathbb{D}}[f(x)+f'_+(x)(z-x)]$ for all $z\in\mathbb{D}$

(vii) $\quad f(z)=\sup\limits_{x\in\mathbb{D}\cap\mathbb{Q}}[f(x)+f'_+(z)(z-x)]$ for all $z\in\mathbb{D}$ where $\mathbb{Q}$ is the set of rational numbers

$(viii)$ $\quad f'_+, f'_-$ are increasing, right-continuous with $f'_-(x)\le f'_+(x)$

(ix) $\quad f$ is differentiable at almost all points of $\mathbb{D}$, except for a countable set for which $f'_+\ne f'_-$

(x) $\quad f\in C^2(\mathbb{D})\Longrightarrow f''\ge 0\,.$

A further useful result can be derived while involving second derivatives assumed to be generalized functions. The proof is found in Revuz and Yor [22]. Let f be convex with its second derivative f'' in the distributional sense. Note that $\mu = f''$ represents a positive Radon measure on $\mathbb{R}$ (whole line) for convex functions f.

Theorem 2.1. *Assume $\mathbb{D}$ is open subset of $\mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ convex on $\mathbb{D}$. Then, there are constants α_0, α_1 such that*

$$f(x) \;=\; \frac{1}{2}\int_{\mathbb{D}} |x-a|\;\mu(da) + \alpha_1 x + \alpha_0, x \in \mathbb{D} \text{ and} \tag{5}$$

$$f'_-(x) \;=\; \frac{1}{2}\int_{\mathbb{D}} sgn(x-a)\;\mu(da) + \alpha_1, x \in \mathbb{D} \tag{6}$$

$$\left(= \frac{1}{2}\Big(\mu(a \in \mathbb{R} : a \le x) - \mu(a \in \mathbb{R} : a > x)\Big) + \alpha_1\right).$$

Hence we may exploit this integral representation of convex functions later. However, this representation is not unique (unique up to affine transformations). If $\int |x-a|\,\mu(da)$ is finite for all $x \in \mathbb{R}$, then one even has a global representation. A more specific result is the following one which can also be generalized to s–convexity (see Revuz and Yor [22]).

Theorem 2.2. *Let $\mathbb{D} = [a,b] \subset \mathbb{R}$ with $a < b$. Assume that $f(a) = f(b) = 0$. Set $G(x,y) = G(y,x) = \frac{(x-a)(b-y)}{b-a}$ for $x \le y (x,y \in \mathbb{R})$. Then we have*

$$f(x) \;=\; -\int_a^b G(x,y)\,\mu(dy) \tag{7}$$

where μ is the Radon measure generated by second derivative f'' of convex f.

2.2. Simple convex functionals. At first consider

$$F_0(t,X) \;=\; \mathbb{E}\, f(T,X(t)) \;=\; \mathbb{E}\, f_T(X(t)) \quad (t \in [0,T], T \text{ fixed}) \tag{8}$$

where $f : [0,T] \times \mathbb{D} \longrightarrow \mathbb{R}$ is convex at x with its second space derivative $\mu_T = f''_T$. Let Y_{n_t} be a right-continuous approximation as step function, $\mathcal{F}_t$–adapted numerical approximation of $X(t)$, based on a numerical method generating random values Y_n and $n_t = \sup\{n : t_n \le t\}$. The expression $p_X = p_X(t,x)$ denotes the probability density of process $X = (X(t))_{0 \le t \le T}$ at point $x \in \mathbb{D}$ at time t, with support $supp(p_X(t,x))$. Let $\tau^\Delta([0,T])$ denote the collection of $\mathcal{F}_t$-adapted time instants belonging to time discretization of $[0,T]$ with maximum step size Δ.

Theorem 2.3. *Let $\mathcal{I} = [0,T]$ or $\mathcal{I} = \tau^\Delta([0,T])$. Assume that*

- (0) *$\mathbb{D}$ is an open, deterministic subset of $\mathbb{R}^1$,*
- *(i)* *$f = f(t,x)$ is convex at $x \in \mathbb{D}$ with second (weak) derivative $\mu_T = f''_T$,*
- *(ii)* $\forall t \in \mathcal{I} \quad \int_{supp(p_X(t,x)) \cap \mathbb{D}} |a| \mu_T(da) < +\infty$,
- *(iii)* $\exists p \ge 1 (p \in \mathbb{R})\, \forall t \in \mathcal{I} \quad \Big(\mathbb{E}\,|X(t)|^p\Big)^{1/p} + \Big(\mathbb{E}\,|Y_{n_t}|^p\Big)^{1/p} \le K_0 < +\infty$,
- *(iv)* $\mathbb{P}(\{\omega \in \Omega : \forall t \in \mathcal{I}\, X(t)(\omega) \in \mathbb{D}\}) = \mathbb{P}(\{\omega \in \Omega : \forall t \in \mathcal{I}\, Y_{n_t}(\omega) \in \mathbb{D}\}) = 1$,
- *(v)* $\exists K_p = K_p(T) > 0\, \exists \gamma \ge 0 \quad \sup_{t \in \mathcal{I}} \Big(\mathbb{E}\,|X(t) - Y_{n_t}|^p\Big)^{1/p} \le K_p \cdot \Delta^\gamma$,
- *(vi)* *$supp(p_X = p_X(t,x)) \cap \mathbb{D}$ is compact.*

Then there is a real constant $K = K(p,T) > 0$ such that

$$\epsilon := \sup_{t \in \mathcal{I}} \Big| \mathbb{E}\, f(T,X(t)) - \mathbb{E}\, f(T,Y_{n_t}) \Big| \;\le\; K \cdot \Delta^\gamma. \tag{9}$$

Proof. First, one verifies that

$$\left|\mathbb{E}\int_{\mathbb{D}}|X(t)-a|\mu_T(da) \;+\; \mathbb{E}\int_{\mathbb{D}}|Y_{n_t}-a|\mu_T(da)\right| \quad < \quad +\infty \tag{10}$$

for all $t \in [0,T]$. For this purpose consider the estimate

$$\begin{aligned}\left|\mathbb{E}\int_{\mathbb{D}}|X(t)-a|\mu_T(da)\right| \quad &\leq \quad \mathbb{E}\int_{\mathbb{D}}|X(t)|\mu_T(da)+\int_{\mathbb{D}}|a|\mu_T(da)\\ &\leq \quad \left(\mathbb{E}\,|X(t)|^2\right)^{\frac{1}{2}}\mu_T(\mathbb{D}) \;+\; \int_{\mathbb{D}}|a|\mu_T(da)\\ &< \quad +\infty\end{aligned}$$

under hypotheses (ii) and (iii). Analogously one arrives at

$$\left|\mathbb{E}\int_{\mathbb{D}}|Y_{n_t}-a|\mu_T(da)\right| < +\infty\,.$$

Therefore, we have the right to apply Fubini's theorem in the later context. From Theorem 2.1, recall the existence of integral representation

$$f(x) \quad = \quad \frac{1}{2}\int_{\mathbb{D}}|x-a|\;\mu(da)+\alpha_1 x+\alpha_0, x\in\mathbb{D} \tag{11}$$

of any convex real-valued functions f with the positive Radon measure μ generated by its second derivative f'' (such that $\mu([a,b]) = f'(b) - f'(a)$ for all $a \leq b$), where α_1 and α_2 are constants and $\mathbb{D}$ is an open subset of $\mathbb{R}$. (Note this representation is not unique (unique up to affine transformations)! If $\int|x-a|\,\mu(da)$ is finite for all $x \in \mathbb{R}$, then one even has a global representation on $\mathbb{R}$.) Now, apply this representation to f_T. One encounters

$$\begin{aligned}|\,\mathbb{E}\,f(T,X(t)) - \,\mathbb{E}\,f(T,Y_{n_t})| \;&=\; |\,\mathbb{E}\,[f_T(X(t)) - f_T(Y_{n_t})]|\\ =\quad &\left|\mathbb{E}\,\frac{1}{2}\int_{\mathbb{D}}\Big\{|X(t)-a|-|Y_{n_t}-a|\Big\}\mu_T(da)+\alpha_1(T)(X(t)-Y_{n_t})\right|\\ \leq\quad &\frac{1}{2}\int_{\mathbb{D}}\,\mathbb{E}\,\big||X(t)-a|-|Y_{n_t}-a|\big|\mu_T(da)+|\alpha_1(T)|\,\mathbb{E}\,|X(t)-Y_{n_t}|\\ \leq\quad &\frac{1}{2}\int_{\mathbb{D}}\,\mathbb{E}\,|X(t)-Y_{n_t}|\mu_T(da)+|\alpha_1(T)|\,\mathbb{E}\,|X(t)-Y_{n_t}|\\ \leq\quad &\Big(\,\mathbb{E}\,|X(t)-Y_{n_t}|^2\Big)^{\frac{1}{2}}\Big(\frac{1}{2}\mu_T(\mathbb{D})+|\alpha_1(T)|\Big)\\ \leq\quad &K_1(\frac{1}{2}\mu_T(\mathbb{D})+|\alpha_1(T)|)\cdot\Delta^\gamma \;\;=:\;\; K\cdot\Delta^\gamma\end{aligned}$$

for all $t \in [0,T]$. Thereby, γ is the rate of convergence of approximate functionals towards the exact ones. □

Remark 2.1. *This result is not so surprising since convex functions are quasi-linearizable and, on compact sets, even Lipschitz-continuous. However, it possesses an interesting proof. For any Lipschitz-continuous function f the p-th mean convergence rates γ_g carry over one to one to weak convergence rates $\beta = \gamma_g$. With this result in hand, one can justify using numerical approximation with some guarantee of a certain least rate of accuracy, depending on regularity of price process X, to estimate European call and put options.*

2.3. Corollaries for distance and extreme value functionals.

Corollary 2.1. *Assume conditions* $(0)-(v)$ *of Theorem 2.3, that* $supp(p_{\|X-c\|} = p_{\|X-c\|}(t,z)) \cap \mathbb{D}$ *is compact. Consider functionals of the form*

$$F_1(t) = f(t, \|X(t)-c\|), c = const, t \in \mathcal{I} \tag{12}$$

where $f(t,z)$ *is convex with respect to the space coordinate* $z \in R^1$*. Then, there is a real constant* $K = K(p,T)$ *such that for all* $t \in [0,T]$

$$\varepsilon(t) = |\,\mathbb{E}\, f(t, \|X(t)-c\|) - \mathbb{E}\, f(t, \|Y_{n_t} - c\|)| \ \leq\ K \cdot \Delta^{\gamma}. \tag{13}$$

For **concave** and some **path-dependent functionals**, similar results hold. More general integral representations of functionals with **signed measures** can be exploited. The proof of Corollary 2.1 follows straight forward that of Theorem 2.3, hence it can be omitted here.

Corollary 2.2. *Assume conditions* $(0)-(v)$ *of Theorem 2.3, that* $supp(p_{\sup\|X\|} = p_{\sup_{0\leq s\leq t}\|X(s)\|}(t,z)) \cap \mathbb{D}$ *is compact and* $X(t) - Y_{n_t}$ *is a right-continuous submartingale with respect to the natural filtration* $\mathcal{F}_t = \sigma\{W_s^j : 0 \leq s \leq t, j = 1,2,...,m\}$*. Consider functionals of the form*

$$F_2(T,t) = f_T(\sup_{0\leq s\leq t} \|X(s)\|) \tag{14}$$

$$or \qquad F_3(T,t) = f_T(\sup_{0\leq s\leq t} X(s)^i)(i \in 1,2,...,d \textit{ fixed}) \tag{15}$$

where $f_T(z)$ *is convex with respect to the space coordinate* $z \in R^1$*. Then, error estimate (9) is also valid for* F_2, F_3 *(with a constant* $K > 0$ *which may differ from that constant above, see (9)).*

Proof. For simplicity, consider only the case (14). Analogously to Theorem 2.3, one arrives at

$$\begin{aligned}
\varepsilon(t) &= \left| \mathbb{E}\left(f_T(\sup_{0\leq s\leq t}\|X(s)\|) - f_T(\sup_{0\leq s\leq t}\|Y_{n_s}\|)\right)\right| \\
&\leq \frac{1}{2}\int_{\mathbb{D}} \mathbb{E}\left| |\sup_{0\leq s\leq t}\|X(s)\| - a| - |\sup_{0\leq s\leq t}\|Y_{n_s}\| - a|\right| \mu(da) \\
&\qquad + |\alpha_1(T)|\,\mathbb{E}\left|\sup_{0\leq s\leq t}\|X(s)\| - \sup_{0\leq s\leq t}\|Y_{n_s}\|\right| \\
&\leq (\frac{1}{2}|\mu_T(\mathbb{D})| + |\alpha_1(T)|)\,\mathbb{E}\left|\sup_{0\leq s\leq t}\|X(s)\| - \sup_{0\leq s\leq t}\|Y_{n_s}\|\right| \\
&\overset{(*)}{\leq} (\frac{1}{2}\mu_T(\mathbb{D}) + |\alpha_1(T)|)\,\mathbb{E}\sup_{0\leq s\leq t}\|X(s) - Y_{n_s}\| \\
&\overset{(**)}{\leq} 2(\frac{1}{2}\mu_T(\mathbb{D}) + |\alpha_1(T)|)\sup_{0\leq s\leq t}\left(\mathbb{E}\,\|X(s) - Y_{n_s}\|^2\right)^{\frac{1}{2}} \ \leq\ K \cdot \Delta^{\gamma}
\end{aligned}$$

where $K = K(T) = 2K_1(T)(\frac{1}{2}\mu_T(\mathbb{D}) + |\alpha_1(T)|)$, hence the desired estimate has been obtained. Note that we have used Lemma 2.1 (see below) to estimate $(*)$ above, and Doob's maximal inequality for right–continuous submartingales to receive $(**)$. $\square$

Lemma 2.1. *Let $C_{rc}([0,t],\bar{\mathbb{D}})$ be the set of right–continuous functions mapping from $[0,t]$ to $\bar{\mathbb{D}}$, $\mathbb{D}$ some open domain of $\mathbb{R}^d$, $\|.\|$ a vector norm in $\mathbb{R}^d$. Then we have*

$$\left| \sup_{0\le s\le t} \|f(s)\| - \sup_{0\le s\le t} \|g(s)\| \right| \le \sup_{0\le s\le t} \|f(s) - g(s)\|$$

for all $f, g \in C_{rc}([0,t],\bar{\mathbb{D}})$.

Proof. First, it is not hard to verify that $|||f||| := \sup_{0\le s\le t} \|f(s)\|$ is a norm on $C_{rc}([0,t],\mathbb{R}^d)$. Then, as an immediate consequence of an application of inverse triangular inequality of norms, the assertion of Lemma 2.1 follows. □

Remark 2.2. *If one replaces the assumption (v) of Theorem 2.3 by*

$$(v') \qquad \exists K_1 = K_1(T) > 0 \, \exists \gamma > 0 \quad \mathbb{E} \sup_{t\in I} |X(t) - Y_{n_t}| \le K_1 \cdot \Delta^\gamma,$$

then analogous results as in Theorem 2.3, Corollaries 2.1 and 2.2 can be proved. Thus, for path-dependent convex functionals and problems of optimal stochastic control, clarification of the problem of practical construction of approximations with a $\mathcal{F}_t$-sub-martingale error process remains to be done. The latter problem seems to be solvable for the class of X-subharmonic functionals f (but in general it is an open question).

2.4. Remark on application to asset pricing. Asset- and option price processes X for **Randomly Exercised Exotic Options** (**American Look-back Call Option**) may cause the following payoff functionals

$$F_i(\tau, X) = \mathbb{E}\left[\exp\Big(- \int_\tau^T r(s)ds \Big) \Big(\sup_{\tau\le t\le T} |X_t^{(i)}| - K_i(T,\tau) \Big)_+ \right]$$

for calls of the i-th component of true observable price-process X (for puts respectively), where $K_i(T,\tau)$ represents the strike price at randomly stopped moment τ which is $\mathcal{F}_\tau$-adapted. Now, for example, there is the task of finding the optimal stopping strategy $0 \le \tau \le T < +\infty$ (i.e. random exercise time τ of the call option with bounded deterministic maximal terminal time) such that the expected discounted loss caused by the payoff at time τ is minimal under the amount of information $\mathcal{F}_t$ at current time t and discounted by $\mathcal{F}_s$-adapted random short interest rate $r(s)$, i.e. one wishes to approximate the optimal solution of the stochastic control problem

$$c(\tau^*) := \inf_{0\le\tau\le T} \mathbb{E}\left[\exp\Big(- \int_\tau^T r(s)ds \Big) \Big(\sup_{\tau\le t\le T} |X_t^{(i)}| - K_i(T,\tau) \Big)_+ \Big| \mathcal{F}_\tau \right],$$

where $c = c(\tau^*) = \mathbb{E}\,[F_i(\tau^*, X)|\mathcal{F}_{\tau^*}]$. This represents a composition of convex functionals, and to apply our results from before, we have to construct a p-th mean converging numerical approximation which is right-continuous and which has a $\mathcal{F}_t$-sub-martingale as its error process $X(t) - Y_{n_t}$. Then the convergence rate will be $\beta = \gamma_g$, and numerical approaches reported in the literature on mathematical finance can be justified by our convergence approach, even for convex, path-dependent functionals of X which can be non-continuously differentiable at some countable points. The practical construction is still a problem since the construction procedure which guarantees the sub-martingale error process may strongly depend on the structure of price process X. An alternative to control errors of numerical approximations of path-dependent functionals such as F is given by Theorem 3.2 in Section 3.

3. Convergence theorems for approximations of Hölder-continuous and path-dependent functionals F

For Hölder-continuous functionals one encounters the following result. Let $\mathbb{D}$ denote an open, deterministic domain of $\mathbb{R}^d$. Fix $d, k \in \mathbb{N}_+$. Define

$$C^0_{H(K_H,\alpha)}(\mathbb{D}) := \left\{ f : \mathbb{D} \subseteq \mathbb{R}^d \longrightarrow \mathbb{R}^k : \|f(x) - f(y)\|_k \leq K_H \|x - y\|_d^\alpha \right\}$$

with Hölder constant K_H and Hölder exponent $\alpha \in [0, 1]$. One arrives at

$$\begin{aligned} \|\,\mathbb{E}\, f(X(t)) - \mathbb{E}\, f(Y^\Delta_{n_t})\|_k &\leq \mathbb{E}\,\|f(X(t)) - f(Y^\Delta_{n_t})\|_k \\ &\leq K_H\, \mathbb{E}\,\|X(t) - Y^\Delta_{n_t}\|_d^\alpha \\ &\leq K_H(\,\mathbb{E}\,\|X(t) - Y^\Delta_t\|_d^p)^{\alpha/p} \\ &\leq K_H \cdot [K(p,T)]^\alpha \Delta^{\alpha\gamma} \end{aligned}$$

for $f \in C^0_{H(K_H,\alpha)}(\mathbb{D})$ and strongly converging approximation Y with rate γ. Taking the supremum leads to the following uniform estimation of convergence order determined by the Hölder exponent α, uniformly with respect to the class of Hölder-continuous mappings, exhibiting a natural loss of convergence speed with decreasing Hölder exponent α. Fix real constants $\alpha \in [0, 1]$ and $K_H \geq 0$.

Theorem 3.1. *Assume that $f \in C^0_{H(K_H,\alpha)}(\mathbb{D})$, processes $X = (X(t))_{0\leq t\leq T}$ and $Y = (Y^\Delta_{n_t})_{0\leq t\leq T}$ are two $\mathbb{D}$-invariant (a.s.), $\mathcal{F}_t$-adapted stochastic processes with respect to the same stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$ satisfying*

$$\sup_{t\in\mathcal{I}} \Big(\,\mathbb{E}\,\|X(t) - Y^\Delta_{n_t}\|_d^p\Big)^{1/p} \leq K(p,T)\Delta^\gamma.$$

with some $\gamma \in \mathbb{R}_+$. Then, there exists an appropriate deterministic constant $K_w(p, T, K_H, \alpha) = K_H \cdot [K(p, T)]^\alpha$ such that

$$\sup_{f\in C^0_{H(K,\alpha)}} \sup_{t\in\mathcal{I}} \|\,\mathbb{E}\, f(X(t)) - \mathbb{E}\, f(Y^\Delta_{n_t})\|_k \leq K_w(p,T,K_H,\alpha)\Delta^{\alpha\gamma}.$$

Its proof is fairly standard and follows from the argumentation above, hence more details can be omitted here.

The following theorem establishes a uniform estimate on the worst-case total approximation error of discounted functionals F as given initially by (2). This theorem also motivates some of our previous efforts to report on those results before by Section 2. Let

$$\frac{1}{p} + \frac{1}{q} \;=\; 1$$

for $p \geq 1$ (for $p = 1$, set $q = \infty$). Hence, $p = q/(q-1)$ and $q = p/(p-1)$ Recall that

$$\|f(t)\|_\infty = \sup_{0\leq t\leq T} |f(t)|$$

for any function $f : [0, T] \to \mathbb{R}$. Let $\mathcal{B}(S)$ denote the σ-algebra of Borel-sets of inscribed set S, and μ the Lebesgue measure.

Theorem 3.2. *Assume that there is a rate $\gamma \geq 0$ such that*

(i) $\forall t \geq 0 : r(t), \tilde{r}(t) \geq 0$ *with probability one,*

(ii) $\exists K_r \in L^p([0,T], \mathcal{B}([0,T]), \mu)\ \forall t \in [0,T]$:

$$\mathbb{E}\,|r(t) - \tilde{r}(t)|^p \leq (K_r(t))^p \Delta_{max}^{p\gamma},$$

(iii) $\exists K_p \in L^\infty([0,T], \mathcal{B}([0,T]), \mu)\ \forall t \in [0,T]$:

$$\mathbb{E}\,|p(X(t)) - p(\tilde{X}(t))| \leq K_p(t)\Delta_{max}^{\gamma},$$

(iv) $\exists K_b \in L^\infty([0,T], \mathcal{B}([0,T]), \mu)\ \forall t \in [0,T]$:

$$\mathbb{E}\,|p(X(t))|^{p\wedge q} + \mathbb{E}\,|p(\tilde{X}(t))|^p \leq (K_b(t))^q,$$

where $\tilde{r}$ and $\tilde{X}$ denote the approximations of r and X along partitions $0 = t_0 < t_1 < \ldots < t_{n_T}$ with maximum step size $\Delta_{max} = \max_{i=1,\ldots,n_T} |t_i - t_{i-1}|$. Then, the total error of related approximation satisfies

$$\varepsilon := \sup_{0\leq t\leq T} \mathbb{E}\,\left|F_{r,p,X}(t) - F_{\tilde{r},p,\tilde{X}}(t)\right| \quad\leq\quad K(T)\Delta_{max}^{\gamma} \tag{16}$$

where the constant $K = K(T)$ is bounded by

$$K(T) \leq T^{1/q}\|K_r\|_{L^p}\|K_b\|_\infty + \|K_p\|_\infty.$$

Proof. First, note that the expression

$$\begin{aligned}
&\Big|\exp(-x)p(u) - \exp(-y)p(v)\Big| \\
&\leq\quad \Big|(\exp(-x) - \exp(-y))p(u)\Big| + \Big|\exp(-y)(p(u) - p(v))\Big| \\
&\leq\quad |x-y||p(u)| + |p(u) - p(v)|
\end{aligned}$$

is bounded as stated for all nonnegative values $x, y \geq 0$. Set $x = \int_t^T r(s)ds$, $y = \int_t^T \tilde{r}(s)ds$, $u = (X(s))_{0\leq s\leq T}$ and $v = (\tilde{X}_s)_{0\leq s\leq T}$. Thus, by means of triangle and Hölder inequalities, we arrive at

$$\begin{aligned}
\varepsilon \quad&\leq\quad \mathbb{E}\left[|\int_s^T (r(s)-\tilde{r}(s))ds||p(X)|\right] + \mathbb{E}\left[|p(X)-p(\tilde{X})|\right] \\
&\leq\quad \left[\mathbb{E}\,|\int_s^T (r(s)-\tilde{r}(s))ds|^p\right]^{1/p} [\mathbb{E}\,|p(X)|^q]^{1/q} + \|K_p\|_\infty \Delta_{max}^{\gamma} \\
&\leq\quad \sup_{0\leq t\leq T}\left[(T-t)^{p/q}\int_t^T \mathbb{E}\,[|r(s)-\tilde{r}(s)|^p]ds\right]^{1/p} \|K_b\|_\infty + \|K_p\|_\infty \Delta_{max}^{\gamma} \\
&\leq\quad \Big(T^{1/q}\|K_r\|_{L^p}\|K_b\|_\infty + \|K_p\|_\infty\Big)\Delta_{max}^{\gamma} \quad ,
\end{aligned}$$

hence the assertion of Theorem 3.2 is confirmed. □

Remark 3.1. *Of course, under sufficient smoothness conditions one may also obtain uniform estimates on higher orders of errors of weak approximations as reported in Milstein* [17] *and Talay* [34] *for the case of deterministic r. Assumption (iii) of convexity of the price functional p can be relaxed too. We have just required assumption (iii) in order to be consistent with our previous results in this paper (see Section 2). Consequently, we can easily apply Theorem 3.2 to verify a least rate of L^1-convergence of nonstandard methods such as positivity-preserving balanced implicit methods (called BIMs, see* [24, 30]*) with $\gamma = 0.5$ and balanced Milstein methods* *(called BMMs, see* [11]*) with* $\gamma = 1.0$*, respectively, or other implicit methods for pricing functionals such as (2) involving SDEs (1). The occurring positive interest rates* r *are numerically integrated by positivity-preserving methods as presented in* [11, 19, 24, 25]. *Note that the positivity is important to guarantee the finiteness of discount factors and the control on the dynamical behavior of functionals* F *such as (2).*

4. A(n) (incomplete) list of interesting examples of functionals

Some of the following examples might look somehow artificial. However, they possess a countable set of points in the domain of definition of these functionals where they are not infinitely continuously differentiable. They are also of some practical interest in finance and engineering. Moreover, a considerable number has the built-in property of convexity.

4.1. Absolute p–th mean process. One of the simplest non-continuously differentiable functionals is the absolute norm of stochastic processes, i.e.

$$F(t, X(t)) \; = \; \|X(t)\|, \tag{17}$$

provided that $X(t)$ can not be bounded away from zero (as e.g. for oscillators with degenerate diffusion). More general, functionals of p–th absolute mean $\|X(t)\|^p$ have no 'infinite smoothness' everywhere (p not even). This functional occurs in engineering and physics where one measures the distance of oscillations from its rest point.

4.2. Occupation probabilities and probabilities of residence. The probabilities of occupation (residence) of stochastic processes in a given real set are of great interest in several disciplines. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ represent the underlying stochastic basis. The associated functional is representable as expectations

$$r_A(s) \; := \; \mathbb{P}(\{\omega \in \Omega \, : \, X(s)(w) \in A\}) \; = \; \mathbb{E} \; I\!I_A(X(s)) \tag{18}$$

where $I\!I$ denotes the indicator function of Borel-measurable set $A \subset \mathbb{R}^d$.

4.3. Exit probabilities and failure probabilities. Conversely, one has interest in

$$f_A(s) \; = \; \mathbb{P}(\{X(s) \notin A\}) \quad \text{or} \quad f_A(s) \; = \; \mathbb{P}(\{\|X(s)\| > a\}) \tag{19}$$

for Borel-measurable sets $A \subset \mathbb{R}^d$ and $a > 0$ some positive, significant, risky level in earthquake engineering.

4.4. (Mean) passage times. In engineering sciences and physics one is interested in the computation of

$$\int_0^t I\!I_A(X(s))\, ds \quad \text{or} \quad \mathbb{E} \int_0^t I\!I_A(X(s))\, ds \tag{20}$$

which represent the functionals of almost sure passage time and mean passage time of the corresponding stochastic process X up to time t, respectively. Subscript $A \subset \mathbb{R}^d$ in (20) denotes a certain Borel-measurable nonempty subset of interest in $\mathbb{R}^d$.

4.5. Maximum and minimum process. These processes are defined to be

$$\begin{aligned} &\mathbb{E}\,\max(0, X(t)) \text{ in } \mathbb{R}^1 \quad \text{or} \quad \mathbb{E}\,\max(0, X^i(t)), \, \mathbb{E}\,\max(a, \|X(t)\|) \text{ in } \mathbb{R}^d, \\ &\mathbb{E}\,\min(0, X(t)) \text{ in } \mathbb{R}^1 \quad \text{or} \quad \mathbb{E}\,\min(0, X^i(t)), \, \mathbb{E}\,\min(a, \|X(t)\|) \text{ in } \mathbb{R}^d. \end{aligned} \tag{21}$$

Obviously, they describe the extreme value behavior of processes X.

4.6. Cost and gain functionals in finance and investments. In option pricing (European put and call options) it arises an interest to compute the expected gain of the form

$$g(s) \;:=\; \mathbb{E}\,(X(s)-K)_+ \;=\; \mathbb{E}\,\max(0, X(s)-K) \tag{22}$$

where $K > 0$ is the striking price at maturity time T and $X(s)$ describes the underlying price process of an asset, market, product, etc. at time s.

4.7. Mean first entrance time. This functional takes the form

$$p_A(t) \;=\; \mathbb{E}\,\inf\{0 \le s \le t : X(s) \in A | X(0) = X_0 \notin A\} \tag{23}$$

where one assumes that $\mathbb{P}(\{X_0 \in A\}) = 0$, otherwise we define $p_A(t) = +\infty$ for a fixed Borel-measurable nonempty set A.

4.8. Mean first exit time. A mathematical description is given by

$$e_A(t) \;=\; \mathbb{E}\,\inf\{s \ge 0 : X(s) \notin A\} \tag{24}$$

for fixed (deterministic) $X_0 \in A$ and Borel-measurable nonempty set A.

4.9. Total maximum and minimum process. In $\mathbb{R}^1$ we set

$$X_{\max}(t) \;=\; \sup_{0\le s\le t}\{X(s)\}\,,\; X_{\min}(t) \;=\; \inf_{0\le s\le t}\{X(s)\}\,. \tag{25}$$

These amounts (magnitudes) are useful for the computation of first entrance and exit times. Furthermore, in $\mathbb{R}^d$ one is interested in the approximation of extreme value processes defined by

$$X^*(t) \;=\; \sup_{0\le s\le t}\|X(s)\|\,,\; X_*(t) \;=\; \inf_{0\le s\le t}\|X(s)\|\,. \tag{26}$$

Remark 4.1. *This list of most interesting functionals is not complete, but it reflects somehow the importance of research on approximations of functionals $F(X)$ involving convex, path-dependent and non-smooth structures (instead of functions $f(X(T))$ depending only on values of X at terminal time T) as met in applications. Besides, we can easily recognize a number of challenging open problems from this list in view of qualitative aspects of their numerical approximations, despite of numerous results such as known from Bally and Talay* [4] *on the rate of approximations of measurable functions $f(X(T))$ (instead of more complex functionals) where non-degeneracy conditions of Hörmander-type on the infinitesimal generator of process $(X(t))_{t\ge0}$ are imposed additionally (which is violated in case of multiplicative noise). We know that the presented results do not have to be most efficient in specific situations, but certainly it justifies the use of several approximation methods for functionals with some convex, path-dependent or not-so-smooth structure.*

Appendix A. Common numerical methods and convergence for Itô SDEs

A.1. Most used numerical methods. By truncation of Itô-Taylor expansions [36] and locally implicit or explicit substitutions of occurring differential operators, one arrives at numerical methods for (1). We state a selected list of the most common numerical methods which are indeed in use. Let $(Y_n)_{n\in\mathbb{N}}$ denote a sequence of approximation values for the solution $X(t_n)$ at time t_n along the *time-discretization*

$$0 = t_0 \le t_1 \le t_2 \le ... \le t_{n_T} = T \tag{27}$$

(for simplicity, we suppose that $t_0 = 0$ and $t_{n_T} = T$). The time-discretization is said to be *equidistant* if there is a number $\Delta \in \mathbb{R}_+$ (called the *step size*) such that $\Delta = t_{i+1} - t_i$ for all $i = 0, 1, ..., n_T - 1$. In general, we define

$$\Delta = \max_{i=0,1,\ldots,n_T-1} |t_{i+1} - t_i|$$

as the *step size*, and $\Delta_i = t_{i+1} - t_i$ as the *local step size*. Consider $\Delta W_n^j = W_j(t_{n+1}) - W_j(t_n)$ as the current increment of the Wiener process component W_j.

The most well-known numerical method is given by the explicit *Euler method* which has been introduced by Maruyama [16]. That is why it is sometimes called *Euler-Maruyama method*. The scheme of explicit Euler method is defined by

$$Y_{n+1} \ = \ Y_n + a(t_n, Y_n)\,\Delta_n + \sum_{j=1}^{m} b^j(t_n, Y_n)\,\Delta W_n^j\,. \tag{28}$$

Its convergence in L^2 has been proved by Gikhman and Skorochod [9], and in L^p by Kanagawa [12]. It represents the most-studied, best-understood and simplest-implementable numerical method. Nowadays, it is even used to understand existence and uniqueness proofs of solutions of SDEs (Krylov [15]). A drawback of method (28) can be seen in the lack of numerical stability (in fact "sub-stable" behavior), the low convergence order, incorrect stationary laws and some problems with the geometrical invariance properties (e.g. non-simplectic integrator). Despite these facts it is a very popular and easily implemented, hence practical method. It is natural to ask for a counterpart to the deterministic *implicit Euler method*. It is given by

$$Y_{n+1} \ = \ Y_n + a(t_{n+1}, Y_{n+1})\,\Delta_n + \sum_{j=1}^{m} b^j(t_n, Y_n)\,\Delta W_n^j\,. \tag{29}$$

The use of this *drift-implicit Euler method* can control numerical stability of certain moments, boundary value replication and reduce variance-destabilizing effects. However, there are the drawbacks of "super-stability" [23], asymptotic non-exactness of stationary laws to be replicated [25], and more computational effort due to additional implementation of resolution algorithms of nonlinear algebraic equations.

A first natural generalization of explicit and implicit Euler methods is presented by *stochastic Theta methods*. They are convex linear combinations of explicit and implicit Euler increment functions of the drift part, whereas the diffusion part is explicitly treated due to the problem of adequate integration within one and the same stochastic calculus. The scheme of an *exterior drift-implicit Theta method* is written as

$$Y_{n+1} \ = \ Y_n + (\Theta_n a(t_{n+1}, Y_{n+1}) + (I - \Theta_n) a(t_n, Y_n))\,\Delta_n + \sum_{j=1}^{m} b^j(t_n, Y_n)\,\Delta W_n^j, \tag{30}$$

where I represents the $d \times d$ real unit matrix, and Θ_n is a uniformly bounded parameter matrix in $\mathbb{R}^{d\times d}$, which is also called the *matrix of implicitness parameters*. This family has been introduced by Schurz [25] as a generalization of deterministic Theta methods. An important special sub-class is $\Theta_n = \theta_n I$ with $\theta_n \in \mathbb{R}^1$. If all $\theta_n = 0$ then its scheme reduces to classical (forward) *Euler method*, if all $\theta_n = 1$ to the *backward Euler* or often called *implicit Euler method*, and if all $\theta_n = 0.5$ to the *implicit trapezoidal method*. A detailed study of the qualitative behavior of these methods can be found in Schurz [25] in stochastics. Another generalization is given by the *interior drift-implicit Theta method* following

$$(31)\quad Y_{n+1} \;=\; Y_n + a\left(t_n + \theta_n\Delta_n, \Theta_n Y_{n+1} + (I-\Theta_n)Y_n\right)\Delta_n + \sum_{j=1}^{m} b^j(t_n, Y_n)\,\Delta W_n^j$$

where $\theta_n \in \mathbb{R}$, $\Theta_n \in \mathbb{R}^{d\times d}$ such that local algebraic resolution can be guaranteed.

For the integration of conservation laws and Hamiltonian systems, it is recommended to take derivates of the *implicit midpoint method*

$$(32)\qquad Y_{n+1} \;=\; Y_n + a\Big(\frac{t_{n+1}+t_n}{2}, \frac{Y_{n+1}+Y_n}{2}\Big)\,\Delta_n + \sum_{j=1}^{m} b^j(t_n, Y_n)\,\Delta W_n^j$$

which is a special case of interior drift-implicit Theta method with $\Theta_n = \frac{1}{2}I$. This method seems to be very promising for the control of numerical stability, area-preservation and boundary laws in stochastics as well. The drawback can be the local resolution of nonlinear algebraic equations, which can be circumvented by predictor-corrector methods (PCMs) or linear-implicit implementations (see Belinskiy and Schurz [5] for a complex example). A natural extension of trapezoidal integration techniques is represented by the *implicit trapezoidal method* governed by

$$(33)\qquad Y_{n+1} \;=\; Y_n + \frac{1}{2}\left(a(t_{n+1}, Y_{n+1}) + a(t_n, Y_n)\right)\Delta_n + \sum_{j=1}^{m} b^j(t_n, Y_n)\,\Delta W_n^j$$

which is a special case of exterior drift-implicit Theta method with $\Theta_n = \frac{1}{2}I$. Both the trapezoidal and midpoint method have an improved local mean consistency behavior (they are of mean convergence order 2, locally considered of mean order 3, under enough smoothness of $a \in C_b^{3,3}([0,T]\times\mathbb{R}^d)$), compared to the explicit and implicit Euler methods. The trapezoidal method has problems when one integrates high-dimensional systems with boundary conditions, as reported by numerous deterministic numerical analysts. However, it is the only numerical method from the class of Theta methods with $\Theta_n = \theta I, \theta \in \mathbb{R}^1$ which asymptotically integrates linear stochastic systems without bias in stationary laws (i.e. asymptotically exact method with respect to stationary laws), see Schurz [25], [26].

For the control on the almost sure path-behavior, the incremental growth and on the error propagation, Milstein, Platen and Schurz [18] have introduced the class of *balanced implicit methods* (BIMs) determined by

$$(34)\qquad Y_{n+1} \;=\; Y_n + a(t_n, Y_n)\,\Delta_n + \sum_{j=1}^{m} b^j(t_n, Y_n)\,\Delta W_n^j + \sum_{j=0}^{m} C^j(t_n, Y_n)(Y_{n+1} - Y_n)|\Delta W_n^j|$$

with appropriate weight matrices $C^j(t,x)$ such that the inverse of $d\times d$ matrix

$$M(t,x) = I + \sum_{j=0}^{m} \theta_j C^j(t,x)$$

exists and is uniformly bounded for all values $\theta_j \in \mathbb{R}_+^1$, $0 \le \theta_0 \le \hat{\theta}_0 < +\infty$ and $(t,x) \in [0,T]\times\mathbb{R}^d$. This class has been studied in Schurz [24], [25], [29], [30]. It represents a linear-implicit integration technique, and hence local resolution can be guaranteed and made very simple as well. However, the choice of the matrix weights $C^j(t,x)$ is still a challenge for future research and exhibits a very problematic and practically oriented question (basically C^j has to be chosen according to the desired qualitative properties of discussed discretization, and thanks to Schurz [23], [25], [26], [29], [30], it is proved that the coefficients C^j with $j = 1, 2, ..., m$ are not really needed to have asymptotically exact control on the moments of approximation Y. However, all these coefficients C^j are needed in context of (almost sure) path-wise control, see [11], [24], [25], [30]).

The simplest higher order method is due to Milstein [17]. It has the scheme

$$Y_{n+1} \;=\; Y_n + a(t_n, Y_n)\,\Delta_n + \sum_{j=1}^{m} b^j(t_n, Y_n)\,\Delta W_n^j + \tag{35}$$
$$+ \sum_{j,k=1}^{m} \mathcal{L}^k b^j(t_n, Y_n) \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} dW_u^k dW_s^j.$$

This method has limited use when numerical stability is an important issue and multidimensional Wiener processes ($m > 1$) drive the dynamics (except for very restricted condition of commutative noise). The generation of multiple integrals $I_{(i,j)} = \int \int dW^k dW^j$ is described in Kloeden, Platen and Schurz [14] by using Karhunen-Loeve expansion. There is an idea to make the Milstein method implicit (see [14]). This idea is realized by the *family of drift-implicit Milstein methods* following the scheme

$$Y_{n+1} \;=\; Y_n + (\theta_n a(t_{n+1}, Y_{n+1}) + (I - \theta_n) a(t_n, Y_n))\,\Delta_n + \sum_{j=1}^{m} b^j(t_n, Y_n)\,\Delta W_n^j \tag{36}$$
$$+ \sum_{j,k=1}^{m} \mathcal{L}^k b^j(t_n, Y_n) \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} dW_u^k dW_s^j$$

where $\theta_n \in \mathbb{R}$ is a sequence of implicitness parameters to be chosen carefully. The convergence orders are as that of explicit Milstein method. However, the numerical stability behavior cannot be improved compared to corresponding Theta methods with the same θ. For more details in this respect, see Schurz [23], [25]. Thus, the balance between convergence and stability requirements is already a problem here with growing order of convergence. More generally, one might think of the usage of *exterior drift-implicit Theta–Milstein methods* governed by

$$Y_{n+1} \;=\; Y_n + (\Theta_n a(t_{n+1}, Y_{n+1}) + (I - \Theta_n) a(t_n, Y_n))\,\Delta_n + \sum_{j=1}^{m} b^j(t_n, Y_n)\,\Delta W_n^j \tag{37}$$
$$+ \sum_{j,k=1}^{m} \mathcal{L}^k b^j(t_n, Y_n) \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} dW_u^k dW_s^j$$

where $\Theta_n \in \mathbb{R}^{d \times d}$ is a certain matrix of implicitness parameters, and the usage of *interior drift-implicit Theta–Milstein methods*

$$Y_{n+1} \;=\; Y_n + a\left(t_n + \theta_n \Delta_n, \Theta_n Y_{n+1} + (I - \Theta_n) Y_n\right)\Delta_n + \sum_{j=1}^{m} b^j(t_n, Y_n)\,\Delta W_n^j \tag{38}$$
$$+ \sum_{j,k=1}^{m} \mathcal{L}^k b^j(t_n, Y_n) \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} dW_u^k dW_s^j$$

where $\theta_n \in \mathbb{R}, \Theta_n \in \mathbb{R}^{d \times d}$, as before, are such that the local resolution of implicit algebraic equations can be guaranteed. But the meaningfulness of the last two methods (37) and (38) is still in question.

There are other methods such as Taylor- or Runge-Kutta methods of local higher order of convergence. However, their efficient use is not quite clear in general since there is a lack of mathematically rigorous studies of their qualitative behavior apart from the knowledge on their local consistency under very restrictive assumptions and the problem of efficient generation of iterated stochastic integrals. We call these methods as **standard** ones (such as Euler-, Milstein-, Taylor-methods as listed in the books [14] and [17]) and others as **nonstandard** ones. The construction of nonstandard stochastic methods is still in its infancy, however one can consult [25] as its origin. The classes of balanced implicit methods (BIMs, see [30]), balanced Milstein methods (BMMs, see [11]) or linear-implicit methods (LIMs, see [5]) count as nonstandard methods.

A.2. Most common convergence concepts. In the statements below, let $||\cdot||$ be a vector norm of $\mathbb{R}^d$ and $K_0, K_p (p \in [1, +\infty])$ be deterministic, real constants which may depend on smoothness and boundedness parameters of the explicit solution, as well as initial values, the length of time-interval $[0, T]$, the dimensions d, m and some parameter of the corresponding numerical method. Recall

$$\Delta \;=\; \sup\{|t_{n+1} - t_n| : n = 0, 1, 2, ..., n_T - 1\}.$$

Fix the finite deterministic start instant $t_0 \in [0, T]$ with fixed terminal time $T > t_0$ where $T \in \mathbb{R}^1$. Let $Y = (Y_t)_{0\le t\le T}$ denote a right-continuous time approximation of process $X = (X(t))_{0\le t\le T}$ based on values Y_n at instants t_n along partitions (27).

Definition A.1. *A stochastic process $Y = (Y_t^\Delta)_{0\le t\le T}$ (method, scheme, etc.) is called a* **p-th mean approximation of** $X = (X(t))_{t\in[t_0,T]}$ **with order (rate)** $\gamma \ge 0$ *if*

$$\sup_{0\le t\le T} \left(\mathbb{E}\; ||X(t) - Y_t^\Delta\; ||^p\right)^{1/p} \;\le\; K_p \cdot \Delta^\gamma, \tag{39}$$

a **mean square approximation of** $X = (X(t))_{t\in[t_0,T]}$ **with order (rate)** $\gamma \ge 0$ *if*

$$\sup_{0\le t\le T} \left(\mathbb{E}\; ||X(t) - Y_t^\Delta\; ||^2\right)^{1/2} \;\le\; K_2 \cdot \Delta^\gamma, \tag{40}$$

a **strong approximation of** $X = (X(t))_{t\in[t_0,T]}$ **with order (rate)** $\gamma \ge 0$ *if*

$$\sup_{0\le t\le T} \mathbb{E}\; ||X(t) - Y_t^\Delta\; || \;\le\; K_1 \cdot \Delta^\gamma \tag{41}$$

a **strong mean square approximation of** $X = (X(t))_{t\in[t_0,T]}$ **with order (rate)** $\gamma \ge 0$ *if*

$$\left(\mathbb{E}\sup_{0\le t\le T} ||X(t) - Y_t^\Delta\; ||^2\right)^{1/2} \;\le\; K_2 \cdot \Delta^\gamma, \tag{42}$$

a **strong p-th mean approximation of** $X = (X(t))_{t\in[t_0,T]}$ **with order (rate)** $\gamma_p \ge 0$ *if*

$$\left(\mathbb{E}\sup_{0\le t\le T} ||X(t) - Y_t^\Delta\; ||^p\right)^{1/p} \;\le\; K_p \cdot \Delta^\gamma, \tag{43}$$

a **double L^p-approximation of** $(X(t))_{t\in[t_0,T]}$ **with order (rate)** $\gamma \ge 0$ *if*

$$\left(\mathbb{E}\int_0^T K(t)||X(t) - Y_t^\Delta\; ||^p\, \mu(dt)\right)^{1/p} \;\le\; K_p \cdot \Delta^\gamma, \tag{44}$$

with positive, μ-integrable kernel $K(t)$ where μ is a positive, finite measure on $([0,T],\mathcal{B}([0,T]))$ *($\mathcal{B}([0,T])$ denotes the σ-field of Borel sets of $[0,T]$), a* **weak approximation of** $X = (X(t))_{t\in[t_0,T]}$ **with order (rate)** $\beta \geq 0$ *if*

$$\sup_{g\in F}\sup_{0\leq t\leq T} || \mathbb{E}\ g(X(t)) - \mathbb{E}\ g(Y_t^{\Delta})\ || \ \leq\ K_0 \cdot \Delta^{\beta} \tag{45}$$

and a **weak τ-convergent approximation of** $X = (X(t))_{t\in[t_0,T]}$ **with order (rate)** $\beta \geq 0$ *if*

$$\sup_{g\in F}\sup_{0\leq \tau\leq T} || \mathbb{E}\ g(X(\tau)) - \mathbb{E}\ g(Y_\tau^{\Delta})\ || \ \leq\ K_0 \cdot \Delta^{\beta} \tag{46}$$

for all time–discretizations of $[t_0,T]$ with $\Delta < \delta_0 < +\infty$, where the supremum is taken over all finite stopping times τ and F is an appropriate class of functions.

Remarks. One also speaks of **p-th mean, mean square, strong, strong p-th mean, double L^p, and weak orders (rates)** $\gamma, \beta \in \mathbb{R}_+$ of convergence. The function class is frequently chosen to be

$$F_r = \Big\{ f : \mathbb{R}^d \longrightarrow \mathbb{R}^k, f \in C^\infty(\mathbb{D}), \exists K\, \forall x \in \mathbb{D}\, \|f(x)\| \leq K(1+\|x\|^r) \Big\}$$

where $r \in \mathbb{R}_+, r \geq 1$, and $d,k \in \mathbb{N}$ are fixed, but there are also attempts to relax conditions in F to certain classes of Lebesgue-measurable functions. The weak τ-convergence is introduced for the delicate problem of convergence and convergence rates for functionals involving random stopping times instead of deterministic terminal times. This concept is of great use and very reasonable in optimal stochastic control problems related to diffusions X. Note also that, for p-th mean convergence, it suffices to evaluate the error expressions at discretization points t_n under the commonly met assumptions on SDE coefficients and on approximating integrands arising by the related numerical method. This becomes clear from looking at the continuous time behavior of remainder terms of stochastic Taylor expansions and natural continuous continuation of discrete time approximations. Recall that p-th mean convergence analysis has importance for estimation of non-continuously differentiable or path-dependent functionals of SDEs as seen in previous sections.

Let us briefly summarize the rates of convergence of standard numerical methods such as Euler-, Milstein-type and balanced methods. The families of drift-implicit Euler-, Theta-methods and BIMs have the rate $\gamma = 0.5$ of p-th mean convergence in general. For systems with additive noise, this rate rises to $\gamma = 1.0$. The rate of p-th mean convergence can be improved by Milstein-type methods in general. The families of drift-implicit Theta-Milstein methods and balanced Milstein methods (called BMMs, see [11]) possess the rate $\gamma = 1.0$ of p-th mean convergence. The above mentioned methods have the rate $\beta = 1.0$ of weak convergence with respect to sufficiently smooth classes F of test functions. Only the drift-implicit midpoint-type Theta-methods from forementioned methods can achieve the rate $\beta = 2.0$ of weak convergence for some appropriate test functions. In general, there are maximum bounds on the rate of convergence. For example, Clark and Cameron [8] showed that the maximum possible L^2-rate along classical nonrandom partitions (such as along discrete time filtrations of increments of Wiener processes) is 1.0 for systems with additive noise in general. So to speak of higher order of convergence than 1.0 in L^p-sense does not make a lot of sense in general. For more details, see also Allen [1], Artemiev and Averina [3], Burrage, Burrage and Mitsui [7], Kanagawa and Ogawa [13], Schurz [25], [27], [28], [31], and Talay [34], [35].

A.3. On key relations between convergence concepts. As a consequence of the Lyapunov inequality and fast L^p-convergence (Borel-Cantelli Theorem), for $p \geq 1$, we may establish the following relationship.

Theorem A.1. *Assume that* $F = C^1_{Lip}(\mathbb{R}^d, \mathbb{R}^k)$, $\sup_{0\leq t\leq T} \mathbb{E}[||X(t)||^p] < +\infty$. *Then the following implications hold*

$$\begin{array}{rcl} \textit{Strong p-th mean} & \Longrightarrow & \textit{p-th mean conv.} \Longrightarrow \textit{strong conv.} \Longrightarrow \textit{weak conv.} \\ \textit{Strong p-th mean} & \Longrightarrow & \textit{double } L^p \\ \textit{Strong p-th mean} & \Longrightarrow & \textit{a.s. convergence} \\ \textit{Weak } \tau\textit{-convergence} & \Longrightarrow & \textit{weak convergence} \end{array}$$

where the related convergence orders are carried over one to one.

The detailed proof is left to the reader as an exercise.

Acknowledgments

The author thanks the anonymous referees for their comments to improve this paper. The appendix is added after the request of referees.

References

[1] E. Allen, Modeling with Itô Stochastic Differential Equations, Book-Manuscript, Texas Tech University, Lubbock, 2006.
[2] L. Arnold, Stochastic Differential Equations: Theory and Applications, Wiley & Sons, Inc., New York, 1974.
[3] S.S. Artemiev and T.A. Averina, Numerical Analysis of Systems of Ordinary and Stochastic Differential Equations, VSP, Utrecht, 1997.
[4] V. Bally and D. Talay, The Euler scheme for stochastic differential equations: error analysis with Malliavin calculus, Math. Comput. Simulation 38 (1995), No. 1-3, 35-41.
[5] B. Belinskiy and H. Schurz, Undamped nonlinear beam excited by additive L^2-regular noise, Revised Preprint M-04-004, p. 1-17, Department of Mathematics, Southern Illinois University, Carbondale, 2006.
[6] N. Bouleau and D. Lépingle, Numerical Methods for Stochastic Processes, Wiley & Sons, Inc., New York, 1993.
[7] K. Burrage, P.M. Burrage and T. Mitsui, Numerical solutions of stochastic differential equations – implementation and stability issues, J. Comput. Appl. Math. 125 (2000), No. 1-2, 171-182.
[8] J.M.C. Clark and R.J. Cameron, The maximum rate of convergence of discrete approximations for stochastic differential equations, Lecture Notes in Control and Information Sci. 25, p. 162-171, Springer-Verlag, Berlin, 1980.
[9] I.I. Gikhman and A.V. Skorochod, Stochastische Differentialgleichungen, Akademie-Verlag, Berlin, 1971.
[10] T.C. Gard, Introduction to Stochastic Differential Equations, Marcel Dekker, Basel, 1988.
[11] C. Kahl and H. Schurz, Balanced Milstein methods for SDEs, Monte Carlo Methods Appl. 12 (2006), No. 2, 143-170.
[12] S. Kanagawa: On the rate of convergence for Maruyama's approximation solutions of stochastic differential equations, Yokohama Math. J. 36 (1988), No. 1, 79-86.
[13] S. Kanagawa and S. Ogawa, Numerical solution of stochastic differential equations and their applications, Sugaku Expositions 18 (2005), No. 1, 75-99.
[14] P.E. Kloeden, E. Platen and H. Schurz, Numerical Solution of SDEs through Computer Experiments, Universitext, Springer, Berlin, 1994.
[15] N.V. Krylov: Introduction to the Theory of Diffusion Processes, Translations of Mathematical Monographs **142**, AMS, Providence, 1995.
[16] G. Maruyama, Continuous Markov processes and stochastic equations, Rend. Circ. Mat. Palermo 4 (1955), 48-90.
[17] G.N. Milstein, Numerical Integration of Stochastic Differential Equations, Kluwer, Dordrecht, 1995.

[18] G. N. Milstein, E. Platen, and H. Schurz, Balanced implicit methods for stiff stochastic systems, SIAM J. Numer. Anal. 35 (1998), No. 3, 1010-1019.

[19] E. Moro and. H. Schurz, Boundary preserving semi-analytic numerical algorithms for stochastic differential equations, Revised Preprint M-05-006, p. 1-24, Department of Mathematics, Southern Illinois University, Carbondale, 2006.

[20] B. Oksendal, Stochastic Differential Equations, Springer, New York, 1985.

[21] P. Protter, Stochastic Integration and Differential Equations, Springer, Berlin, 1990.

[22] D. Revuz and M. Yor, Continuous Martingales and Brownian Motion, 2nd edition, Springer, New York, 1994.

[23] H. Schurz, Asymptotical mean square stability of an equilibrium point of some linear numerical solutions, Stochastic Anal. Appl. 14 (1996), No. 3, 313-354.

[24] H. Schurz, Numerical regularization for SDEs: Construction of nonnegative solutions, Dynam. Systems Appl. 5 (1996), 323-352.

[25] H. Schurz, Stability, Stationarity, and Boundedness of Some Implicit Numerical Methods for SDEs and Applications, Logos-Verlag, Berlin, 1997.

[26] H. Schurz, The invariance of asymptotic laws of linear stochastic systems under discretization, Z. Angew. Math. Mech. 79 (1999), No. 6, 375–382.

[27] H. Schurz, *Numerical analysis of SDEs without tears*, In Handbook of Stochastic Analysis (Ed. D. Kannan and V. Lakshmikantham), p. 237-359, Marcel Dekker, Basel, 2002.

[28] H. Schurz, General theorems for numerical approximation of stochastic processes on the Hilbert Space $H_2([0,T],\mu,\mathbb{R}^d)$, Electr. Trans. Numer. Anal. 16 (2003), 50-69.

[29] H. Schurz, Stability of numerical methods for ordinary SDEs along Lyapunov-type and other functions with variable step sizes, Electr. Trans. Numer. Anal. 20 (2005), 27-49.

[30] H. Schurz, Convergence and stability of balanced implicit methods for SDEs with variable step sizes, Int. J. Numer. Anal. Model. 2 (2005), No. 2, 197-220.

[31] H. Schurz, An axiomatic approach to numerical approximations of stochastic processes, Int. J. Numer. Anal. Model. 3 (2006), No. 4., 459-480.

[32] A.N. Shiryaev, Probability, 2nd edition, Springer, New York, 1996.

[33] D. Talay, *Efficient numerical schemes for the approximation of expectations of functionals of the solution of a SDE and applications*, In Filtering and Control of Random Processes (Paris, 1983), Lecture Notes in Control and Inform. Sci. 61, p. 294-313, Springer, Berlin, 1984.

[34] D. Talay, *Simulation of stochastic differential systems*, In Probabilistic Methods in Applied Physics, Springer Lecture Notes in Physics 451 (Ed. P. Krée and W. Wedig), p. 54-96, Springer-Verlag, Berlin, 1995.

[35] D. Talay, *Simulation of stochastic processes and applications*, In Foundations of Computational Mathematics (Oxford, 1999), London Math. Soc. Lecture Note Ser. 284, p. 345-359, Cambridge Univ. Press, Cambridge, 2001.

[36] W. Wagner and E. Platen, Approximation of Itô integral equations, February Report at ZIMM of Academy of Sciences of GDR, Berlin, 1978.

Department of Mathematics, Southern Illinois University, 1245 Lincoln Drive, Carbondale, IL 62901-4408, USA

E-mail: `hschurz@math.siu.edu`

URL: `http://www.math.siu.edu/schurz/personal.html`