Balanced Milstein Methods for Ordinary SDEs

Christian Kahl
Department of Mathematics, University of Wuppertal,
Gaußstraße 20, Wuppertal, D-42119, Germany

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

Received April 12, 2005; Revised February 5, 2006

Abstract

Convergence, consistency, stability and pathwise positivity of balanced Milstein methods for numerical integration of ordinary stochastic differential equations (SDEs) are discussed. This family of numerical methods represents a class of highly efficient linear-implicit schemes which generate mean square converging numerical approximations with qualitative improvements and global rate 1.0 of mean square convergence, compared to commonly known numerical methods for SDEs.

1 Introduction

This paper deals with numerical methods for systems of ordinary stochastic differential equations (SDEs)

$$dX_t = a(t, X_t)dt + \sum_{j=1}^{m} b^j(t, X_t)dW_t^j \tag{1}$$

with Wiener processes $W^j = (W_t^j)_{0\le t\le T}$. The stochastic integration is interpreted in the Itô sense. Furthermore, we assume that $a \in C^{0,0}([0,T]\times\mathbb{D},\mathbb{D})$ and $b^j \in C^{0,1}([0,T]\times\mathbb{D},\mathbb{D})$ where $\mathbb{D}\subseteq\mathbb{R}^d$ is a nonrandom set such that, for all $0\le t\le T$, we have

$$\mathbb{P}(\{X_t \in \mathbb{D}\}) \;=\; 1. \tag{2}$$

In various areas we are confronted with the problem of invariance-preserving of certain subsets $\mathbb{D}$ of $\mathbb{R}^d$. A first approach to tackle this problem is the class of *balanced implicit methods* (BIMs) as introduced by Milstein, Platen and Schurz [14], studied by Kahl [8] and Schurz [17, 18, 22]. In the following we use the notation of multi-indices to describe the Itô integrals we have to use. Therefore we have

$$\begin{aligned} I_{(0)}^{s,t} &= (t-s), \\ I_{(j)}^{s,t} &= (W_t^j - W_s^j) \in \mathcal{N}(0, t-s), \\ I_{(i,j)}^{s,t} &= \int_s^t \int_s^u dW_v^i dW_u^j. \end{aligned}$$

One integration step of the BIMs is given by

$$Y_{n+1} = Y_n + \sum_{j=0}^{m} b^j(t_n, Y_n) I_{(j)}^{t_n,t_{n+1}} + \sum_{j=0}^{m} c^j(t_n, Y_n) |I_{(j)}^{t_n,t_{n+1}}| \,(Y_n - Y_{n+1}) \tag{3}$$

where $b^0(t_n, Y_n) = a(t_n, Y_n)$ and suitable weight functions $c^j \in C^{0,0}([0,T]\times\mathbb{D},\mathbb{D})$ along the discretization

$$0 \le t_0 \le t_1 \le \ldots \le t_n \le \ldots t_{n_T} \le T. \tag{4}$$

We assume the stepsizes to be variable and the maximum stepsize is given by

$$\Delta_{max} = \max_{n=0,\ldots,n_T-1} |t_{n+1} - t_n|. \tag{5}$$

We already know that the BIMs have a strong and mean square order of convergence $r_g = 0.5$. Improving the order of convergence requires more information about the underlying Wiener path. Using a stochastic Taylor expansion leads to the forward Milstein method with order $r_g = 1.0$ of mean square convergence. Unfortunately this scheme is neither very stable nor positive-preserving (hence not dynamically consistent). For details on these facts, see [18]. For improving the stability, we can derive the *balanced Milstein methods* (BMMs) based on the (explicit) Milstein method governed by

$$\begin{aligned} Y_{n+1} = Y_n \quad &+ \quad \sum_{j=0}^{m} b^j(t_n, Y_n) I_{(j)}^{t_n,t_{n+1}} + \sum_{i,j=1}^{m} L^i b^j(t_n, Y_n) I_{(i,j)}^{t_n,t_{n+1}} \\ &+ \quad \left(d^0(t_n, Y_n) I_{(0)}^{t_n,t_{n+1}} + \sum_{j=1}^{m} d^j(t_n, Y_n) I_{(j,j)}^{t_n,t_{n+1}} \right) (Y_n - Y_{n+1}) \end{aligned} \tag{6}$$

with

$$L^i = \sum_{k=1}^{m} b_k^i(t_n, Y_n) \frac{\partial^k}{\partial x^k} \quad \text{for} \quad x \in \mathbb{D}, 0 \le t \le T. \tag{7}$$

Here the functions $d^j \in C^{0,0}([0,T] \times \mathbb{D}, \mathbb{D})$ play the same role as the weights c^j in case of the BIMs (3) and will be specified later on (see section 5.2 and 5.3). The BMM (6) possesses also a one-step representation

$$Y_{s,y}(t) \quad = \quad y + M_{s,y}^{-1}(t) \left(\sum_{j=0}^{m} b^j(s,y) I_{(j)}^{s,t} + \sum_{i,j=1}^{m} L^i b^j(s,y) I_{(i,j)}^{s,t} \right) \tag{8}$$

where

$$M_{s,y}(t) \quad = \quad I_d + d^0(s,y) I_{(0)}^{s,t} + \sum_{j=1}^{m} d^j(s,y) I_{(j,j)}^{s,t}. \tag{9}$$

The paper is organized as follows. In Section 2 we state the basic concepts of consistency and main assumptions, and analyse the BMM with respect to conditional mean and mean-square consistency. Section 3 investigates the stability behavior in view of uniform boundedness of its moments. In Section 4 we shall prove that the BMM has a global strong and mean square order of convergence $r_g = 1.0$ which is the same as the underlying Milstein method. Finally, Section 5 discusses the almost sure positivity of BMM with diagonal noise and reports on numerical experiments related to some stochastic volatility models relevant to mathematical finance. Apart from our theoretical results, these experiments will support the obvious evidence of applicability and superiority of appropriately chosen BMMs compared to commonly used numerical methods.

2 Conditional mean and mean square consistency

Consider the following definitions. Throughout the paper, fix the time interval $[0,T]$ with finite and nonrandom terminal time T. Let $\|.\|_d$ be the Euclidean vector norm on $\mathbb{R}^d$ and $\mathcal{M}_p([s,t])$ the Banach space of $(\mathcal{F}_u)_{s \le u \le t}$-adapted, continuous, $\mathbb{R}^d$-valued stochastic processes X with finite norm $\|X\|_{\mathcal{M}_p} = (\sup_{s \le u \le t} \mathbb{E}\, ||X(s)||_d^p)^{1/p} < +\infty$ where $p \ge 1$, $\mathcal{M}([0,s])$ the space of $\mathcal{F}_s$-measurable stochastic processes and $\mathcal{B}(S)$ the σ-algebra of Borel sets of inscribed set S.

Definition 2.1. A numerical method Y with one-step representation $Y_{s,y}(t)$ is said to be **mean consistent with rate** r_0 on $[0,T]$ if $\exists$ Borel-measurable function $V : \mathbb{D} \to \mathbb{R}_+^1$ and $\exists$ real constants $K_0^C \ge 0, \delta_0 > 0$ such that $\forall (\mathcal{F}_s, \mathcal{B}(\mathbb{D}))$-measurable random variables $Z(s)$ with $Z \in \mathcal{M}_2([0,s])$ and $\forall s,t : 0 \le t - s \le \delta_0$

$$||\, \mathbb{E}\, [X_{s,Z(s)}(t) - Y_{s,Z(s)}(t) | \mathcal{F}_s]||_d \quad \le \; K_0^C \sqrt{V(Z(s))}\, (t-s)^{r_0}. \tag{10}$$

Remark. It is well-known from Milstein [13] and Kloeden and Platen [11] that the standard Milstein method is mean consistent with worst case rate $r_0 \ge 2.0$ and moment control function $V(x) = 1 + ||x||_d^2$ for SDEs (1) with global Lipschitz-continuous and linear growth-bounded coefficients $b^j \in F \subset C^{1,2}([0,T] \times \mathbb{D})$.

Definition 2.2. A numerical method Y with one-step representation $Y_{s,y}(t)$ is said to be **mean square consistent with rate** r_2 on $[0,T]$ if $\exists$ Borel-measurable function $V : \mathbb{D} \to \mathbb{R}_+^1$ and $\exists$ real constants $K_0^C \ge 0, \delta_0 > 0$ such that $\forall (\mathcal{F}_s, \mathcal{B}(\mathbb{D}))$-measurable random variables $Z(s)$ with $Z \in \mathcal{M}_2([0,s])$ and $\forall s,t : 0 \le t - s \le \delta_0$

$$\left(\mathbb{E}\, [||X_{s,Z(s)}(t) - Y_{s,Z(s)}(t)||_d^2 | \mathcal{F}_s] \right)^{1/2} \quad \le \; K_2^C \sqrt{V(Z(s))}\, (t-s)^{r_2}. \tag{11}$$

Remark. It is well-known from Milstein [13] and Kloeden and Platen [11] that the standard Milstein method is mean square consistent with worst case rate $r_2 \geq 1.5$ and moment control function $V(x) = 1 + ||x||_d^2$ for SDEs (1) with global Lipschitz-continuous and linear growth-bounded coefficients $b^j \in F \subset C^{1,2}([0,T] \times \mathbb{D})$.

2.1 The main assumptions

Let $||.||_{d\times d}$ denote a matrix norm on $\mathbb{R}^{d\times d}$ which is compatible to the Euclidian vector norm $||.||_d$ on $\mathbb{R}^d$, and $\langle .,. \rangle_d$ the Euclidean scalar product on $\mathbb{R}^d$. Furthermore we have to assume that the coefficients a and b^j are Caratheodory functions such that a strong, unique solution $X = (X_t)_{0\leq t\leq T}$ exists. Recall that $\mathbb{D} \subseteq \mathbb{R}^d$ is supposed to be a nonrandom set. To guarantee the convergence of the BMM the following conditions have to be satisfied:

(A0) $\forall s,t \in [0,T] \ : \ s < t \implies \mathbb{P}(\{X_t \in \mathbb{D} | X_s \in \mathbb{D}\}) = \mathbb{P}(\{Y_{s,y}(t) \in \mathbb{D} | y \in \mathbb{D}\}) = 1.$

(A1) $\exists$ constants $K_B = K_B(T), K_D = K_D(T), K_V = K_V(T) \geq 0$ such that

$$\forall t \in [0,T] \ \forall x \in \mathbb{D} : \sum_{j=0}^{m} ||b^j(t,x)||_d^2 \ \leq \ (K_B)^2 \, V(x) \tag{12}$$

$$\forall t \in [0,T] \ \forall x \in \mathbb{D} : \sum_{i,j=1}^{m} ||L^i b^j(t,x)||_d^2 \ \leq \ (K_D)^2 \, V(x) \tag{13}$$

$$\sup_{0\leq t\leq T} \mathbb{E}\, V(X_t) \leq K_V \, \mathbb{E}\, V(X_0) \ < \ +\infty \tag{14}$$

with appropriate Borel-measurable function $V : \mathbb{D} \to \mathbb{R}_+^1$.

(A2) The forward Milstein method Y^M applied to Itô SDE (1) is assumed to be mean consistent with rate $r_0^M \geq 2.0$ and mean square consistent with rate $r_2^M = 1.5$ with respect to V with real constants $K_0^M, K_2^M, 0 < \delta_0 \leq 1$.

(A3) $\exists$ real constants $K_M = K_M(T) \geq 0$, and $K_{OB} = K_{OB}(T)$ such that, for the chosen weight matrices $d^j \in \mathbb{R}^{d\times d}$ of BMMs (6), we have

$$\begin{aligned} \forall s,t : 0 \leq t-s \leq \delta_0, \forall x \in \mathbb{D} \qquad & M_{s,x}^{-1}(t) \text{ exists} \\ \text{with} \qquad & ||M_{s,x}^{-1}(t)||_{d\times d} \ \leq \ K_M \\ & \langle x, M_{s,x}^{-1}(t) b^0(s,x) \rangle_d \ \leq \ K_{OB} V(x) \end{aligned} \tag{15}$$

(A4) $\exists$ real constants $K_I = K_I(T) \geq 0$ and $K_{II} = K_{II}(T) \geq 0$ such that, for the chosen weight matrices $d^j \in \mathbb{R}^{d\times d}$ of BMMs (6), we have

$$\forall t \in [0,T] \ \forall x \in \mathbb{D} \ : \ ||d^0(t,x) b^0(t,x)||_d^2 \ \leq \ K_I^2 V(x) \tag{16}$$

$$\forall t \in [0,T] \ \forall x \in \mathbb{D} \ : \ \sum_{j=1}^{m} ||d^j(t,x) L^j b^j(t,x)||_d^2 \ \leq \ K_{II}^2 V(x) \tag{17}$$

(A5) $\exists$ real constants $K_{III} = K_{III}(T) \geq 0$, $K_{IV} = K_{IV}(T) \geq 0$, $K_V = K_V(T) \geq 0$ and $K_{VI} = K_{VI}(T) \geq 0$ such that, for the chosen weight matrices $d^j \in \mathbb{R}^{d\times d}$ of BMMs (6), we have

$$\forall t \in [0,T]\ \forall x \in \mathbb{D} \quad : \quad \sum_{j=0}^{m} ||d^0(t,x)b^j(t,x)||_d^2 \ \leq \ K_{III}^2 V(x) \tag{18}$$

$$\forall t \in [0,T]\ \forall x \in \mathbb{D} \quad : \quad \sum_{i,j=1}^{m} ||d^0(t,x)L^i b^j(t,x)||_d^2 \ \leq \ K_{IV}^2 V(x) \tag{19}$$

$$\forall t \in [0,T]\ \forall x \in \mathbb{D} \quad : \quad \sum_{l=1}^{m}\sum_{j=0}^{m} ||d^l(t,x)b^j(t,x)||_d^2 \ \leq \ K_{V}^2 V(x) \tag{20}$$

$$\forall t \in [0,T]\ \forall x \in \mathbb{D} \quad : \quad \sum_{l=1}^{m}\sum_{i,j=1}^{m} ||d^l(t,x)L^i b^j(t,x)||_d^2 \ \leq \ K_{VI}^2 V(x) \tag{21}$$

Remark 2.1. *The assumption (A1) guarantees the existence of unique and continuous solutions for the system* (1) *with boundedness of moments along the function V. (A2) is neccessary to prove the mean and mean square consistency of the BMM while comparing it with the forward Milstein method. Assumption (A3) ensures that the BMM* (6) *is well defined and a first proposal for the parameters choice is given in Theorem 2.2 below. To prove the mean consistency of the BMM we have to restrict the weight functions by the conditions given in (A4) and (A5) is needed to guarantee the mean square consistency and, above all, to prevent any possible local explosions and to guarantee uniform boundedness of its increments and moments (i.e. these methods are indeed well-defined). If both the exact solution X and numerical approximation Y leave the metric space $\mathbb{D} \subset \mathbb{R}^d$ invariant with probability one as required by (A0), then the conditions (A1)-(A5) can be relaxed to $x \in \mathbb{D}$ instead of whole-space requirement $x \in \mathbb{R}^d$. Note, the constant K_{OB} of one-sided boundedness can be estimated from above by the positive constant $K_{OB} \leq (1 + K_M^2 K_B^2)/2$ if $V(x) \geq \|x\|_d^2$. However, for the later purpose of more efficient results on stability, by (A3) we allow it to be negative as well.*

Theorem 2.2. *The assumption (A3) under (A0) with $\|x\|_d^2 \leq \rho^2 V(x)$ and constant $\rho^2 > 0$ for all $x \in \mathbb{D}$ can be fulfilled by weight functions d^j satisfying*

(A6) $\forall t \in [0,T]\ \forall x \in \mathbb{D} : d^0(t,x) - \frac{1}{2}\sum_{j=1}^{m} d^j(t,x)$ is positive semi-definite and

(A7) $\forall j = 1,2,...,m\ \forall t \in [0,T]\ \forall x \in \mathbb{D} : d^j(t,x)$ is positive semi-definite.

Proof. Set $\Delta = t - s$. We can easily verify that $M_{s,x}(t)$ is positive definite

$$\begin{aligned} M_{s,x}(t) &= I_d + d^0(t,x)\Delta + \frac{1}{2}\sum_{j=1}^{m} d^j(t,x)\left(\left(\Delta W^j\right)^2 - \Delta\right) \\ &= I_d + \underbrace{\left(d^0(t,x) - \frac{1}{2}\sum_{j=1}^{m} d^j(t,x)\right)}_{\text{positive semi-definite}}\Delta + \frac{1}{2}\underbrace{\sum_{j=1}^{m} d^j(t,x)}_{\text{positive semi-definite}}\underbrace{\left(\Delta W^j\right)^2}_{\geq 0}. \end{aligned}$$

Therefore, $\exists K_M \leq 1$ such that $\|M_{s,x}^{-1}(t)\|_{d\times d} \leq K_M \leq 1$ for all $\Delta = (t-s) > 0$. Moreover, one finds that

$$\langle x, M_{s,x}^{-1}(t) b^0(s,x)\rangle \quad \leq \quad \frac{1}{2}\Big(\|x\|_d^2 + K_M^2 \|b^0(s,x)\|_d^2\Big) \leq \frac{1}{2}(\rho^2 + K_M^2 K_B^2)V(x)$$

since $2ab \leq a^2 + b^2$, i.e. $\exists K_{OB} \leq (\rho^2 + K_M^2 K_B^2)/2$. Thus, the proof is complete. □

Remark 2.3. *Conditions of Theorem 2.2 imply that the constant K_M involved in (A3) can be estimated by $K_M \leq 1$. Of course, a uniform estimation of K_M under (A6)-(A7) for all $t \geq s \geq 0$ leads to the computation of $K_M = 1$ since $\Delta = t - s$ can be chosen arbitrarily small. However, for fixed partitions of $[0,T]$, we can observe that the better the estimate on K_M is the greater the stabilizing effects on numerical dynamics of BMM* (6) *along the underlying partition (cf. also its use in later estimates with respect to consistency and stability). Note that the choice of K_M is strongly related to the estimation of constant K_{OB} resulting from (A3), which can be negative too.*

2.2 Mean consistency of BMMs (6)

Using the mean consistency of forward Milstein methods, we are able to prove the mean consistency of the BMM.

Theorem 2.4. *Assuming that the conditions (A0)-(A5) hold with a worst case rate $r_0^M \geq 2.0$, control functional V and consistency constants K_0^M and δ_0. Then the BMM (6) is also mean consistent with worst case rate $r_0 \geq 2.0$, control functional V with δ_0 and consistency constant*

$$K_0^C \leq K_0^M + K_M(K_I + \frac{1}{2}K_{II}\sqrt{m}) \tag{22}$$

Before we can prove this theorem we have to recall some facts about the expectation of multiple Itô integrals. This is done by the following Lemma. Let $\delta_{l,j}$ denote the Kronecker symbol.

Lemma 2.5. *The following properties of multiple Itô integrals hold:*

$$\begin{aligned}
\forall j = 0, ..., m \qquad \mathbb{E}\left[I_{(0)}^{s,t} I_{(j)}^{s,t}\right] &= \delta_{0,j}(t-s)^2, \\
\forall i, j = 1, ..., m \qquad \mathbb{E}\left[I_{(0)}^{s,t} I_{(i,j)}^{s,t}\right] &= 0, \\
\forall i, j = 1, ..., m \qquad \mathbb{E}\left[I_{(j)}^{s,t} I_{(i,i)}^{s,t}\right] &= 0, \\
\forall l, i, j = 1, ..., m \qquad \mathbb{E}\left[I_{(l,l)}^{s,t} I_{(i,j)}^{s,t}\right] &= \frac{1}{2}\delta_{l,i}\delta_{i,j}(t-s)^2.
\end{aligned}$$

Proof. Set $\Delta W^j = W_t^j - W_s^j$ and $\Delta = t - s$ for each $t \geq s$. Consider the following conclusions. First, for all $j = 0, 1, ..., m$, we have

$$\mathbb{E}\left[I_{(0)}^{s,t} I_{(j)}^{s,t}\right] = I_{(0)}^{s,t}\,\mathbb{E}\left[I_{(j)}^{s,t}\right] = (t-s)\,\mathbb{E}\left[\int_s^t dW_u^j\right] = \delta_{0,j}(t-s)^2$$

since $W_u^0 = u$ and $\mathbb{E}\left[W_t^j - W_s^j\right] = 0$ for $j \geq 1$. Next, for $i, j = 1, ..., m$, we arrive at

$$\mathbb{E}\left[I_{(0)}^{s,t} I_{(i,j)}^{s,t}\right] = I_{(0)}^{s,t}\,\mathbb{E}\left[I_{(i,j)}^{s,t}\right] = (t-s)\,\mathbb{E}\left[\int_s^t \int_s^u dW_v^i dW_u^j\right] = 0$$

since $I_{(i,j)}$ is a martingale starting at 0. Furthermore, for all $i, j = 1, ..., m$, we may conclude that

$$\begin{aligned}
\mathbb{E}\left[I_{(j)}^{s,t} I_{(i,i)}^{s,t}\right] &= (1-\delta_{j,i})\,\mathbb{E}\left[I_{(j)}^{s,t}\right]\,\mathbb{E}\left[I_{(i,i)}^{s,t}\right] + \delta_{j,i}\frac{1}{2}\,\mathbb{E}\left[\Delta W^j\Big((\Delta W^j)^2 - \Delta\Big)\right] \\
&= \delta_{j,i}\frac{1}{2}\,\mathbb{E}\left[(\Delta W^j)^3 - \Delta(\Delta W^j)\Big)\right] = 0
\end{aligned}$$

while using the fact $\mathbb{E}\left[\xi\right]^4 = 3 \cdot \sigma^4$ whenever $\xi \in N(0, \sigma^2)$ is Gaussian with variance $\sigma^2 = t - s$. Finally, we exploit orthogonality of Itô integrals as known from Lemma 5.7.2 of [11] (see page 191 and page 223). Thus, for all $l, i, j = 1, ..., m$, we have

$$\mathbb{E}\left[I_{(l,l)}^{s,t} I_{(i,j)}^{s,t}\right] = \delta_{l,i}\delta_{i,j}\,\mathbb{E}\left[(I_{(l,l)}^{s,t})^2\right] = \frac{1}{2}\delta_{l,i}\delta_{i,j}(t-s)^2$$

since independence of Wiener processes and Itô integrals holds for independent factors of them and, more precisely, one obtains

$$\begin{aligned}
\mathbb{E}\left[\left(I_{(l,l)}^{s,t}\right)^2\right] &= \mathbb{E}\left[\frac{(\Delta W^l)^2 - \Delta}{2}\right]^2 = \frac{1}{4}\,\mathbb{E}\left[(\Delta W^l)^4 - 2\Delta(\Delta W^l)^2 + \Delta^2\right] \\
&= \frac{1}{4}(t-s)^2[3 - 2 + 1] = \frac{1}{2}(t-s)^2.
\end{aligned}$$

This completes the proof of Lemma 2.5. □

Remark 2.6. *A more general discussion on relations between Itô-integrals can be found in [11].*

Proof of Theorem 2.4. Let $Z(s) \in \mathcal{M}([0,s])$ (i.e. $Z(s)$ is at least $\mathcal{F}_s$-measurable). Then, $\forall s,t : 0 \le t-s \le \delta_0$, we get

$$
\begin{aligned}
&||\,\mathbb{E}\left[X_{s,Z(s)}(t) - Y_{s,Z(s)}(t)|\mathcal{F}_s\right]||_d \\
&\le \quad ||\,\mathbb{E}\left[X_{s,Z(s)}(t) - Y^M_{s,Z(s)}(t)|\mathcal{F}_s\right]||_d + ||\,\mathbb{E}\left[Y^M_{s,Z(s)}(t) - Y_{s,Z(s)}(t)|\mathcal{F}_s\right]||_d \\
&\le \quad K_0^M\sqrt{V(Z(s))}(t-s)^2 \\
&\qquad +||\,\mathbb{E}\left[M^{-1}_{s,z}(t)\left(M_{s,z}(t) - I_d\right)\left(\sum_{j=0}^m b^j(s,z)I^{s,t}_{(j)} + \sum_{i,j=1}^m L^i b^j(s,z)I^{s,t}_{(i,j)}\right)\right]\Big|_{z=Z(s)}||_d \\
&\le \quad K_0^M\sqrt{V(Z(s))}(t-s)^2 \\
&\qquad +K_M||\,\mathbb{E}\Big[\sum_{j=0}^m d^0(s,z)b^j(s,z)I^{s,t}_{(0)}I^{s,t}_{(j)} + \sum_{i,j=1}^m d^0(s,z)L^i b^j(s,z)I^{s,t}_{(0)}I^{s,t}_{(i,j)} \\
&\qquad +\sum_{l=1}^m\sum_{j=0}^m d^l(s,z)b^j(s,z)I^{s,t}_{(l,l)}I^{s,t}_{(j)} + \sum_{l=1}^m\sum_{i,j=1}^m d^l(s,z)L^i b^j(s,z)I^{s,t}_{(l,l)}I^{s,t}_{(i,j)}\Big]\Big|_{z=Z(s)}||_d \\
&\le \quad K_0^M\sqrt{V(Z(s))}(t-s)^2 \\
&\qquad +K_M||\,\mathbb{E}\Big[d^0(s,z)b^0(s,z)I^{s,t}_{(0)}I^{s,t}_{(0)} + \sum_{j=1}^m d^j(s,z)L^j b^j(s,z)I^{s,t}_{(j,j)}I^{s,t}_{(j,j)}\Big]\Big|_{z=Z(s)}||_d \\
&\le \quad K_0^M\sqrt{V(Z(s))}(t-s)^2 + K_M\left(||d^0(s,Z(s))b^0(s,Z(s))||_d^2\right)^{1/2} \\
&\qquad +\frac{1}{2}K_M\sqrt{m}\left(\sum_{j=1}^m ||d^j(s,Z(s))L^j b^j(s,Z(s))||_d^2\right)^{1/2}(t-s)^2 \\
&\le \quad \left(K_0^M + K_M(K_I + \frac{1}{2}K_{II}\sqrt{m})\right)\sqrt{V(Z(s))}(t-s)^2
\end{aligned}
$$

where we have used the discrete Hölder inequality and Lemma 2.5. Thus, the verification of local rate 2.0 of mean consistency with related consistency constant K_0^C along the control function V is complete. □

2.3 Mean square consistency of BMMs (6)

In a similar way we can prove the worst case rate of mean square consistency of BMMs.

Theorem 2.7. *Assuming that the conditions (A0)-(A5) hold with a worst case rate $r_2^M \ge 1.5$, control functional V and consistency constants K_2^M and $\delta_0 \le 1$. Then the BMM (6) is also mean consistent with worst case rate $r_2 \ge 1.5$ control functional V with $\delta_0 \le 1$ and consistency constant*

$$
\begin{aligned}
K_2^C \quad \le \quad & \Big(K_2^M + 2K_M(\sqrt{m+1}K_{III}\sqrt{K_5^I} + mK_{IV}\sqrt{K_6^I} \\
& +\sqrt{m}\sqrt{m+1}K_V\sqrt{K_7^I} + m^{3/2}K_{VI}\sqrt{K_8^I})\Big)
\end{aligned} \tag{23}
$$

where $K_5^I \le 1, K_6^I \le 1/2, K_7^I \le 5/2$ and $K_8^I \le \sqrt{1350}$.

Before we can prove this theorem we have to recall some further facts about the moments of multiple Itô integrals. This is done in the following Lemma. Recall that $\delta_{l,j}$ denotes the Kronecker symbol.

Lemma 2.8. *Assume that $0 \le t-s \le 1$. The following properties of multiple Itô integrals hold: $\exists$ real constants K_5^I, K_6^I, K_7^I and K_8^I such that*

$$
\begin{aligned}
\forall j = 0,...,m \qquad \mathbb{E}\left[\left(I^{s,t}_{(0)}I^{s,t}_{(j)}\right)^2\right] &= K_5^I(t-s)^3 \le (t-s)^3, \quad (\textit{equal if } j = 1,...,m) \\
\forall i,j = 1,...,m \qquad \mathbb{E}\left[\left(I^{s,t}_{(0)}I^{s,t}_{(i,j)}\right)^2\right] &= K_6^I(t-s)^3 = \frac{1}{2}(t-s)^4, \\
\forall l,j = 1,...,m \qquad \mathbb{E}\left[\left(I^{s,t}_{(j)}I^{s,t}_{(l,l)}\right)^2\right] &= K_7^I(t-3)^3 = \left(\delta_{j,l}\frac{5}{2} + (1-\delta_{j,l})\frac{1}{2}\right)(t-s)^3 \le \frac{5}{2}(t-s)^3 \\
\forall l,i,j = 1,...,m \qquad \mathbb{E}\left[\left(I^{s,t}_{(l,l)}I^{s,t}_{(i,j)}\right)^2\right] &= K_8^I(t-s)^4 \le \sqrt{1350}(t-s)^4 \quad (K_8^I = \frac{15}{4} \textit{ if } l = i = j).
\end{aligned}
$$

Proof. The proof of this Lemma relies on elementary relationships between Itô integrals, the fact $\mathbb{E}[\xi]^{2n} = (2n-1)!! \cdot \sigma^{2n}$ whenever Gaussian $\xi \in N(0,\sigma^2)$ with $\sigma^2 = t-s$, the binomial theorem, Cauchy-Bunjakowskii-Schwarz inequality and above all the isometry property of Itô integrals. In details: Set $\Delta W^j = W_t^j - W_s^j$ and $\Delta = t-s \leq 1$ for each $t \geq s$. Recall that $I_{(i,j)}^{s,t}$ represents a square-integrable martingale. Let $< M >_s^t$ denote the quadratic variation of inscribed martingale M on $[s,t]$. Note, by isometry of Itô integrals, that

$$\mathbb{E}[< I_{(i,j)}^{s,t} >_s^t] = \frac{(t-s)^2}{2} = \mathbb{E}[(I_{(i,j)}^{s,t})^2] = \mathbb{E}[\int_s^t (W_u^i - W_s^i)^2 du]$$

for $i,j \geq 1$. Now, consider the following conclusions. First, for all $j = 0,1,...,m$, we have

$$\begin{aligned}\mathbb{E}\left[\left(I_{(0)}^{s,t} I_{(j)}^{s,t}\right)^2\right] &= \delta_{0,j}(t-s)^4 + (1-\delta_{0,j})(t-s)^2\,\mathbb{E}\left[\left(W_t^j - W_s^j\right)^2\right] \\ &\leq (t-s)^3.\end{aligned}$$

Next, for all $j = 1,...,m$, we may elementarily calculate that

$$\begin{aligned}\mathbb{E}\left[\left(I_{(0)}^{s,t} I_{(j,j)}^{s,t}\right)^2\right] &= (t-s)^2\,\mathbb{E}\left[\left(\frac{(\Delta W^j)^2 - \Delta}{2}\right)^2\right] \\ &= \frac{(t-s)^2}{4}\,\mathbb{E}\left[(\Delta W^j)^4 - 2\Delta(\Delta W^j)^2 + \Delta^2\right] = \frac{1}{2}(t-s)^4\end{aligned}$$

by binomial theorem and the fact $\mathbb{E}[(\Delta W^j)^4] = 3\Delta^2 = 3(t-s)^2$. More general, for all $i,j = 1,...,m$, we arrive at

$$\begin{aligned}\mathbb{E}\left[\left(I_{(0)}^{s,t} I_{(i,j)}^{s,t}\right)^2\right] &= (t-s)^2\,\mathbb{E}\left[\left(\int_s^t \int_s^u dW_v^i dW_u^j\right)^2\right] \\ &= (t-s)^2\,\mathbb{E}\left[< I_{(i,j)}^{s,t} >_s^t\right] = \frac{(t-s)^4}{2}\end{aligned}$$

by applying the isometry property of Itô integrals to the martingale $I_{(i,j)}^{s,t}$. Furthermore, for all $l,j = 1,...,m$, we can show that

$$\begin{aligned}&\mathbb{E}\left[\left(I_{(j)}^{s,t} I_{(l,l)}^{s,t}\right)^2\right] = \frac{1}{4}\,\mathbb{E}\left[\left(\Delta W^j\left((\Delta W^l)^2 - \Delta\right)\right)^2\right] \\ &= \frac{\delta_{j,l}}{4}\,\mathbb{E}\left[\left(\Delta W^j\left((\Delta W^j)^2 - \Delta\right)\right)^2\right] + \frac{1-\delta_{j,l}}{4}\,\mathbb{E}\left[(\Delta W^j)^2\right]\,\mathbb{E}\left[\left((\Delta W^l)^2 - \Delta\right)^2\right] \\ &= \frac{\delta_{j,l}}{4}\,\mathbb{E}\left[(\Delta W^j)^6 - 2\Delta(\Delta W^j)^4 + \Delta^2(\Delta W^j)^2\right] + \frac{1-\delta_{j,l}}{4}(t-s)\,\mathbb{E}\left[(\Delta W^l)^4 - 2\Delta(\Delta W^l)^2 + \Delta^2\right] \\ &= \frac{\delta_{j,l}}{4}(t-s)^3\left[5\cdot 3 - 2\cdot 3 + 1\right] + \frac{1-\delta_{j,l}}{4}(t-s)^3\,\mathbb{E}\left[3 - 2\cdot 1 + 1\right] \\ &= \left(\delta_{j,l}\frac{5}{2} + (1-\delta_{j,l})\frac{1}{2}\right)(t-s)^3 \leq \frac{5}{2}(t-s)^3\end{aligned}$$

since

$$\mathbb{E}[(\Delta W^j)^{2n}] = (2n-1)!!\Delta^n = (2n-1)\cdot(2n-3)\cdot ... \cdot 3\cdot 1\cdot (t-s)^n$$

for all $n \in \mathbb{N}$. Finally, for all $l,i,j = 1,...,m$ with $l = i = j$ or $l \neq i,j$, we find that

$$\begin{aligned}&\mathbb{E}\left[\left(I_{(l,l)}^{s,t} I_{(i,j)}^{s,t}\right)^2\right] = \delta_{l,i}\delta_{i,j}\,\mathbb{E}\left[\left(I_{(l,l)}^{s,t}\right)^4\right] + (1-\delta_{l,i})(1-\delta_{l,j})\,\mathbb{E}\left[(I_{(l,l)}^{s,t})^2\right]\cdot\,\mathbb{E}\left[(I_{(i,j)}^{s,t})^2\right] \\ &= \delta_{l,i}\delta_{i,j}\frac{1}{16}\,\mathbb{E}\left[\left((\Delta W^j)^2 - \Delta\right)^4\right] + (1-\delta_{l,i})(1-\delta_{l,j})\frac{1}{4}\,\mathbb{E}\left[\left((\Delta W^l)^2 - \Delta\right)^2\right]\cdot\,\mathbb{E}\left[< I_{(i,j)}^{s,t} >_s^t\right] \\ &= \delta_{l,i}\delta_{i,j}\frac{1}{16}\,\mathbb{E}\left[(\Delta W^j)^8 - 4\Delta(\Delta W^j)^6 + 6\Delta^2(\Delta W^j)^4 - 4\Delta^3(\Delta W^j)^2 + \Delta^4\right] \\ &\quad + (1-\delta_{l,i})(1-\delta_{l,j})\frac{1}{4}\,\mathbb{E}\left[(\Delta W^l)^4 - 2\Delta(\Delta W^j)^2 + \Delta\right]\cdot\frac{(t-s)^2}{2} \\ &= \delta_{l,i}\delta_{i,j}\frac{7\cdot 5\cdot 3 - 4\cdot 5\cdot 3 + 6\cdot 3 - 4\cdot 1 + 1}{16}(t-s)^4 + (1-\delta_{l,i})(1-\delta_{l,j})\frac{1}{4}(t-s)^2[3-2+1]\cdot\frac{(t-s)^2}{2} \\ &= \delta_{l,i}\delta_{i,j}\frac{15}{4}(t-s)^4 + (1-\delta_{l,i})(1-\delta_{l,j})\frac{(t-s)^4}{4} \leq \frac{15}{4}(t-s)^4\end{aligned}$$

by using the isometry-relation for Itô integrals and the binomial theorem. It remains to check the case for all $l, i, j = 1, ..., m$ with $l = i \neq j$ or $l = j \neq i$. Due to symmetry of its analysis, it remains to verify only one of those two cases. For example, for $l = i \neq j$, consider

$$\begin{aligned}
\mathbb{E}\left[\left(I^{s,t}_{(l,l)} I^{s,t}_{(l,j)}\right)^2\right] &\leq \left(\mathbb{E}\left[\left(I^{s,t}_{(l,l)}\right)^4\right]\right)^{1/2} \left(\mathbb{E}\left[\left(I^{s,t}_{(l,j)}\right)^4\right]\right)^{1/2} \\
&\leq \frac{\sqrt{15}}{2}(t-s)^2 \left(\mathbb{E}\left[\left(\int_s^t (W^l_u - W^l_s) dW^j_u\right)^4\right]\right)^{1/2} \\
&\leq \frac{\sqrt{15}}{2}(t-s)^2 (K_{BDG}(4))^{1/2} \left(\mathbb{E}\left[(< I^{s,t}_{(l,j)} >^t_s)^2\right]\right)^{1/2} \leq \frac{\sqrt{15}}{2}(t-s)^4 (K_{BDG}(4))^{1/2} < \sqrt{1350}(t-s)^4
\end{aligned}$$

by applying Burkholder-Davis-Gundy inequality (see [16]) and Cauchy-Bunjakowskii-Schwarz inequality twice. Notice also that

$$\begin{aligned}
\mathbb{E}\left[(< I^{s,t}_{(l,j)} >^t_s)^2\right] &= \mathbb{E}\left[\left(\int_s^t (W^l_u - W^l_s)^2 \, du\right)^2\right] \\
&\leq (t-s) \int_s^t \mathbb{E}\left[(W^l_u - W^l_s)^4\right] du = 3(t-s) \int_s^t (u-s)^2 du = (t-s)^4.
\end{aligned}$$

Thus, all constants K^I_p for $p = 5, ..., 8$ can be estimated as claimed by Lemma 2.8, hence its proof is complete. □

Remark 2.9. *The universal constant K_{BDG} resulting from the Burkholder-Davis-Gundy inequality is governed by*

$$K_{BDG}(p) \leq \left(\frac{p}{p-1}\right)^{p^2/2} \left(\frac{p(p-1)}{2}\right)^{p/2},$$

hence we may estimate $K_{BDG}(4) < 360$ or $K_{BDG}(2) \leq 4$. In fact, some of our estimates are essentially better than those stated in [11] (e.g. as in Lemma 5.7.5 at page 197 for the case $l = i = j$).

Proof of Theorem 2.7. Let $Z(s) \in \mathcal{M}([0,s])$ (i.e. $Z(s)$ is at least $\mathcal{F}_s$-measurable). Then, $\forall s, t : 0 \leq t - s \leq \delta_0$, we arrive at

$$\begin{aligned}
&\left(\mathbb{E}\left[||X_{s,Z(s)}(t) - Y_{s,Z(s)}(t)||^2_d | \mathcal{F}_s\right]\right)^{1/2} \\
&\leq \left(\mathbb{E}\left[||X_{s,Z(s)}(t) - Y^M_{s,Z(s)}(t)||^2_d | \mathcal{F}_s\right]\right)^{1/2} + \left(\mathbb{E}\left[||Y^M_{s,Z(s)}(t) - Y_{s,Z(s)}(t)||^2_d | \mathcal{F}_s\right]\right)^{1/2} \\
&\leq K^M_2 \sqrt{V(Z(s))}(t-s)^{3/2} \\
&\quad + \left(\mathbb{E}\left[||M^{-1}_{s,z}(t)\left(M_{s,z}(t) - I_d\right)\left(\sum_{j=0}^m b^j(s,z) I^{s,t}_{(j)} + \sum_{i,j=1}^m L^i b^j(s,z) I^{s,t}_{(i,j)}\right)||^2_d\right]\Big|_{z=Z(s)}\right)^{1/2} \\
&\leq K^M_2 \sqrt{V(Z(s))}(t-s)^{3/2} \\
&\quad + K_M \Big(\mathbb{E}\Big[||\sum_{j=0}^m d^0(s,z) b^j(s,z) I^{s,t}_{(0)} I^{s,t}_{(j)} + \sum_{i,j=1}^m d^0(s,z) L^i b^j(s,z) I^{s,t}_{(0)} I^{s,t}_{(i,j)} \\
&\quad + \sum_{l=1}^m \sum_{j=0}^m d^l(s,z) b^j(s,z) I^{s,t}_{(l,l)} I^{s,t}_{(j)} + \sum_{l=1}^m \sum_{i,j=1}^m d^l(s,z) L^i b^j(s,z) I^{s,t}_{(l,l)} I^{s,t}_{(i,j)}||^2_d\Big]\Big|_{z=Z(s)}\Big)^{1/2} \\
&\leq K^M_2 \sqrt{V(Z(s))}(t-s)^{3/2} + 2K_M \Big(\mathbb{E}\Big[||\sum_{j=0}^m d^0(s,z) b^j(s,z) I^{s,t}_{(0)} I^{s,t}_{(j)}||^2_d\Big]\Big|_{z=Z(s)} \\
&\quad + \mathbb{E}\Big[||\sum_{i,j=1}^m d^0(s,z) L^i b^j(s,z) I^{s,t}_{(0)} I^{s,t}_{(i,j)}||^2_d\Big]\Big|_{z=Z(s)} \\
&\quad + \mathbb{E}\Big[||\sum_{l=1}^m \sum_{j=0}^m d^l(s,z) b^j(s,z) I^{s,t}_{(l,l)} I^{s,t}_{(j)}||^2_d\Big]\Big|_{z=Z(s)} \\
&\quad + \mathbb{E}\Big[||\sum_{l=1}^m \sum_{i,j=1}^m d^l(s,z) L^i b^j(s,z) I^{s,t}_{(l,l)} I^{s,t}_{(i,j)}||^2_d\Big]\Big|_{z=Z(s)}\Big)^{1/2}.
\end{aligned}$$

Using the discrete Hölder inequality and moment properties of products of iterated stochastic integrals leads to

$$
\begin{aligned}
&\left(\mathbb{E}\left[||X_{s,Z(s)}(t)-Y_{s,Z(s)}(t)||_d^2|\mathcal{F}_s\right]\right)^{1/2} \\
\leq\ & K_2^M\sqrt{V(Z(s))}(t-s)^{3/2}+2K_M\Big((m+1)\sum_{j=0}^{m}\mathbb{E}\left[||d^0(s,z)b^j(s,z)||_d^2\Big(I_{(0)}^{s,t}I_{(j)}^{s,t}\Big)^2\right]\Big|_{z=Z(s)} \\
&+m^2\sum_{i,j=1}^{m}\mathbb{E}\left[||d^0(s,z)L^ib^j(s,z)||_d^2\Big(I_{(0)}^{s,t}I_{(i,j)}^{s,t}\Big)^2\right]\Big|_{z=Z(s)} \\
&+m(m+1)\sum_{l=1}^{m}\sum_{j=0}^{m}\mathbb{E}\left[||d^l(s,z)b^j(s,z)||_d^2\Big(I_{(l,l)}^{s,t}I_{(j)}^{s,t}\Big)^2\right]\Big|_{z=Z(s)} \\
&+m^3\sum_{l=1}^{m}\sum_{i,j=1}^{m}\mathbb{E}\left[||d^l(s,z)L^ib^j(s,z)||_d^2\Big(I_{(l,l)}^{s,t}I_{(i,j)}^{s,t}\Big)^2\right]\Big|_{z=Z(s)}\Big)^{1/2} \\
\leq\ & \Big(K_2^M+2K_M(\sqrt{m+1}K_{III}\sqrt{K_5^I}+mK_{IV}\sqrt{K_6^I} \\
&+\sqrt{m}\sqrt{m+1}K_V\sqrt{K_7^I}+m^{3/2}K_{VI}\sqrt{K_8^I})\Big)\sqrt{V(Z(s))}(t-s)^{3/2}
\end{aligned}
$$

where the constants K_l^I originating from the moment estimates of Lemma 2.8 are as defined in the statement of Theorem 2.7. Thus, the proof of its claim is complete. □

3 Stability (uniform boundedness) of 2nd moments

Consider the following definition as originally introduced in [22].

Definition 3.1. A numerical method Y with one-step representation $Y_{s,y}(t)$ is said to be **weakly V-stable** with real constant $K_S=K_S(T)$ on $[0,T]$ if $V:\mathbb{D}\to\mathbb{R}_+^1$ is Borel-measurable and $\exists$ real constant $\delta_0>0$ such that $\forall(\mathcal{F}_s,\mathcal{B}(\mathbb{D}))$-measurable random variables $Z(s)$ and $\forall s,t:0\leq t-s\leq\delta_0\leq 1$

$$\mathbb{E}\left[V(Y_{s,Z(s)}(t))|\mathcal{F}_s\right]\ \leq\ \exp(K_S(t-s))\,V(Z(s)). \tag{24}$$

Theorem 3.1. *Assume that (A0)-(A5) with $V(x)=\rho^2+||x||_d^2$ ($\rho\in\mathbb{R}^1$ some real constant) hold. Then the BMMs (6) with $\Delta_{max}\leq\delta_0\leq\min(1,T)$ are weakly V-stable with stability constant*

$$K_S^Y\ \leq\ 2K_{OB}+(m+1+m\cdot m)(K_B^2+K_D^2)K_M^2 \tag{25}$$

and they satisfy global weak V-stability estimates

$$\mathbb{E}\,V(Y_{0,Y_0}(t))\ \leq\ \exp(K_S^YT)\,\mathbb{E}\,V(Y_0), \tag{26}$$

$$\sup_{0\leq t\leq T}\mathbb{E}\,V(Y_{0,Y_0}(t))\ \leq\ \exp([K_S^Y]_+T)\,\mathbb{E}\,V(Y_0) \tag{27}$$

where $[.]_+$ denotes the positive part of the inscribed expression.

Proof. Recall that

$$\mathbb{E}\left[<I_{(i,j)}^{s,t}>_s^t\right]=\frac{(t-s)^2}{2}=\mathbb{E}\left[(I_{(i,j)}^{s,t})^2\right]=\mathbb{E}\left[\int_s^t(W_u^i-W_s^i)^2du\right]$$

for $i,j\geq 1$. Now, let $y\in\mathbb{D}$ be nonrandom. Calculate

$$
\begin{aligned}
\mathbb{E}\left[||Y_{s,y(t)}||_d^2\right]\ =\ & \mathbb{E}\left[\left\|y+M_{s,y}^{-1}(t)\left(\sum_{j=0}^{m}b^j(s,y)I_{(j)}^{s,t}+\sum_{i,j=1}^{m}L^ib^j(s,y)I_{(i,j)}^{s,t}\right)\right\|_d^2\right] \\
=\ & \|y\|_d^2+2\,\mathbb{E}\left[<y,M_{s,y}^{-1}(t)b^0(s,y)>_d\right](t-s)+\mathbb{E}\left[\left\|M_{s,y}^{-1}(t)\left(\sum_{j=0}^{m}b^j(s,y)I_{(j)}^{s,t}+\sum_{i,j=1}^{m}L^ib^j(s,y)I_{(i,j)}^{s,t}\right)\right\|_d^2\right] \\
\leq\ & \|y\|_d^2+2K_{OB}V(y)(t-s)+K_M^2(m+1+m\cdot m)(K_B^2+K_D^2)V(y)(t-s)^2 \\
\leq\ & \|y\|_d^2+[2K_{OB}+K_M^2(K_B^2+K_D^2)(m+1+m\cdot m)]V(y)(t-s) \\
\leq\ & \exp\Big(K_S^Y(t-s)\Big)V(y)
\end{aligned}
$$

using (A1), (A3), $0 \leq t-s \leq 1$, well-known martingale properties and Hölder inequality. Now, add ρ^2 to the derived inequality. Suppose that $V(y) = \rho^2 + \|y\|_d^2$. For nonrandom $y \in \mathbb{D}$ and $0 \leq t-s \leq \delta_0$, conclude that

$$\mathbb{E}\left[V(Y_{s,y}(t))\right] \quad \leq \quad \exp\Big(K_S^Y(t-s)\Big)V(y)$$

where

$$K_S^Y \leq 2K_{OB} + K_M^2(K_B^2 + K_D^2)(m+1+m\cdot m)$$

by using the elementary inequality $1+z \leq \exp(z)$ for all $z \in \mathbb{R}^1$. It remains to apply the fairly general Theorem 3.1 from [22] with constant $K_S^Y = 2K_{OB} + K_M^2(K_B^2 + K_D^2)(m+1+m\cdot m)$ along $V(y)$ (which is gained by exploiting the tower property of conditional expectations). This confirms the assertion of Theorem 3.1. □

4 Global mean square convergence

The concept of global mean square convergence is understood as follows.

Definition 4.1. A numerical method Y with one-step representation $Y_{s,y}(t)$ is said to be **(globally) mean square convergent with rate** r_g on $[0,T]$ if $\exists$ Borel-measurable function $V : \mathbb{D} \to \mathbb{R}_+^1$ and $\exists$ real constants $K_g = K_g(T) \geq 0, K_S^Y = K_S^Y(b^j), 0 < \Delta_{max} \leq \delta_0 \leq 1$ such that $\forall(\mathcal{F}_0, \mathcal{B}(\mathbb{D}))$-measurable random variables $Z(0)$ with $\mathbb{E}\left[\|Z(0)\|_d^2\right] < +\infty$ and $\forall t : 0 \leq t \leq T$

$$\left(\mathbb{E}\left[||X_{0,Z(0)}(t) - Y_{0,Z(0)}(t)||_d^2|\mathcal{F}_0\right]\right)^{1/2} \quad \leq \; K_g \exp\Big(K_S^Y t\Big)\sqrt{V(Z(0))}\,\Delta_{max}^{r_g} \tag{28}$$

along any nonrandom partitions $0 = t_0 \leq t_1 \leq ... \leq t_{n_T} = T$.

Using the results of the previous section, the following theorem is rather obvious in conjunction with standard L^2-convergence theorems following stochastic Lax-Richtmeyer principles as presented and proven in [19, 20, 21].

Theorem 4.1. *Assume that the conditions (A0)-(A5) hold with a worst case rate $r_2^M \geq 1.5$, control functional $V(x) = \rho^2 + \|x\|_d^2$ with $\rho \in \mathbb{R}^1$, mean K_0^M and mean square consistency constants K_2^M, and $\delta_0 \leq 1$. Furthermore, let X be conditionally mean square contractive, i.e. $\exists$ real constant K_C^X such that, for all $0 \leq t-s \leq \delta_0 \leq \min(1,T)$ and all $(\mathcal{F}_s, \mathcal{B}(\mathbb{R}^d))$-measurable random variables $Y(s), Z(s)$ with $Y, Z \in \mathcal{M}_2([0,s])$, we have*

$$\Big(\mathbb{E}\left[\|X_{s,Y(s)}(t) - X_{s,Z(s)}(t)\|_d^2|\mathcal{F}_s\right]\Big)^{1/2} \quad \leq \quad \exp(K_C^X(t-s))\|Y(s) - Z(s)\|_d. \tag{29}$$

Assume that the coefficients $b^j (j = 1, 2, ..., m)$ are uniform Lipschitz-continuous with Lipschitz constant K_{sm} such that

$$\forall t \in [0,T]\; \forall x, y \in \mathbb{R}^d \quad \sum_{j=1}^m \|b^j(t,x) - b^j(t,y)\|_d^2 \;\leq\; K_{sm}^2\|x-y\|_d^2. \tag{30}$$

Then the BMMs (6) are also mean square converging with global worst case rate $r_g \geq 1.0$ along the control functional V with maximum stepsize $\Delta_{max} \leq \delta_0 \leq \max(1,T)$ and

$$K_g \quad \leq \quad \Big((K_0^C)^2 + (K_2^C)^2[1 + (K_{sm})^2]\Big)^{1/2} \exp\Big(([K_C^X]_- + [K_S^Y]_-)\Delta_{max}\Big)$$

where K_0^C and K_2^C are the constants as estimated by Theorems 2.4 and 2.7, respectively, K_C^X is the mean square contractivity constant of X and K_S^Y the mean square stability constant of BMM (6) as estimated by (25).

Proof. We may take $r_0 = 2.0$ from Theorem 2.4 and $r_0 = 1.5$ from Theorem 2.7. Furthermore, Theorem 3.1 guarantees us moment stability along $V(x) = \rho^2 + \|x\|_d^2$. It remains to apply Theorem 1.1 from Milstein [13] in the case $\mathbb{D} = \mathbb{R}^d$ and $V(x) = 1 + \|x\|_d^2$ and / or Theorem 2.1 from Schurz [20, 21] with $\mathbb{D} \subseteq \mathbb{R}^d$ and $V(x) = 1 + \|x\|_d^2$. For example, a little more detailed, recall the local rates of mean $r_0 \geq 2.0$ established by Theorem 2.4 and $r_2 \geq 1.5$ established by Theorem 2.7. The solutions of the underlying diffusion equations have Hölder-continuous paths with mean square Hölder-constant $r_{sm} = 0.5$. Therefore, Theorem 2.1 of [20] as well as the axiomatic approach described in [21] yield the global rate $r_g = r_2 + r_{sm} - 1.0 \geq 1.0$ of mean square convergence on the finite time-interval $[0,T]$. Furthermore, the error constant K_g can be estimated by

$$K_g \leq \Big((K_0^C)^2 + (K_2^C)^2[1 + (K_{sm})^2]\Big)^{1/2} \exp\Big(([K_C^X]_- + [K_S^Y]_-)\Delta_{max}\Big)$$

using Theorem 2.1 from [20]. So the proof is complete. □

5 Global almost sure positivity of BMMs with diagonal noise

The problem of positive invariance of BIMs (3) has already been studied by Kahl [8] and Schurz [17, 18]. In this paper, we illustrate that the class of BMMs provides an efficient alternative to generate positive-invariant numerical approximations.

Consider the following mean-reverting process

$$dX_t = \kappa(\theta - X_t)dt + \sigma X_t^p dW_t \tag{31}$$

with $\theta, \kappa, \sigma \geq 0$ which is of great importance in financial mathematics as well as in other areas of applied science. Focusing on the financial meaning of this equations we obtain the well known Cox-Ingersoll-Ross model [5] with exponent $p = 0.5$, describing the short-rate in the interest rate market. Furthermore this SDE can be used to model stochastic volatility as it is demonstrated in Andersen and Piterbarg [3]. In such a situation the mean-reverting process is only one part of a two-dimensional system of SDEs

$$\begin{aligned} dS_t &= \lambda(t)f(S_t)\sqrt{X_t}dW_1(t) \\ dX_t &= \kappa(\theta - X_t)dt + \sigma X_t^p dW_2(t) \end{aligned}$$

with suitable functions λ and f and correlated Brownian motions $dW_1(t)dW_2(t) = \rho dt$. Apart from the correlation of the Brownian motions the volatility is independent from the underlying S_t so that we can focus on the process X_t itself. This way of coupling is quite common in the modelling of financial markets, e.g. Andersen and Brotherton-Ratcliffe [2] used this mean-reverting process to simulate the stochastic volatility in the Libor market model. Nonetheless as the process X satisfies the positive-invariance property (2) with respect to the metric space $\mathbb{D} = (0, +\infty)$ combined with nonexploding higher moments it is a great task for numerical integration itself.

So this raises up the question whether BMMs (6) with appropriate weights d^j possess such a property? The classical BIMs (3) can preserve positivity almost surely too, however their global rate of mean square convergence is $r_g = 0.5$ (see relaxation of conditions by [22]) and their weights c^j must be chosen very carefully, e.g.

$$c^0(x) = a, c^1(x) = |\sigma||x|^{p-1}.$$

For more general SDEs (1), only a local ε-positivity can be verified for BIMs under the preservation of convergence properties as known from standard Euler-type methods on the whole axis $\mathbb{R}^1$, see [17, 18]. So it would be advantageous to construct a BMM with positive invariance of $\mathbb{D}$ and global rate 1.0 of mean square convergence to improve the numerical approximation qualitatively farther.

The standard explicit and the drift-implicit Milstein methods as discussed by Kahl [8] and Kahl, Günther and Roßberg [9] in the context of applications to interest rate modeling provide a possibility of preserving positivity too. However, certain stepsize restrictions apart from other conditions occur there. In a very natural way BMMs may inherit this invariance property for all stepsizes and they have the additional feature to control the stability of numerical integration by using the weight functions $d^0, d^1, \ldots$ in an appropriate manner.

5.1 Positivity preserving property of BMMs with diagonal noise

To simplify the notation we relate our discussion to stochastic differential equations driven exclusively by diagonal noise. Recall that $L^i b^j(t,x) = 0$ for all $i \neq j$ where $i, j = 1, ..., m$. In this case the **BMMs** (6) **with diagonal noise** follow the scheme

$$\begin{aligned} Y_{n+1} = Y_n \quad &+ \quad \sum_{j=0}^{m} b^j(t_n, Y_n) I_{(j)}^{t_n,t_{n+1}} + \sum_{j=1}^{m} L^j b^j(t_n, Y_n) I_{(j,j)}^{t_n,t_{n+1}} \\ &+ \quad \left(d^0(t_n, Y_n) I_{(0)}^{t_n,t_{n+1}} + \sum_{j=1}^{m} d^j(t_n, Y_n) I_{(j,j)}^{t_n,t_{n+1}} \right) (Y_n - Y_{n+1}) \,. \end{aligned} \tag{32}$$

Recall the following definition going back to [17, 18] to classify numerical methods with respect to the preservation of natural boundary conditions. Let the relation $x > c$ on $\mathbb{R}^d$ be defined by $x_i > c_i$ for all $i = 1, 2, ..., d$ in a componentwise manner.

Definition 5.1. Let $X = (X_t)_{t \geq 0}$ be the underlying real-valued stochastic process satisfying

$$\forall t > s \geq 0 \; : \; \mathbb{P}(\{X_t > c | X_s > c\}) = 1$$

with a fixed threshold $c \in \mathbb{R}^d$. Then, a numerical integration scheme $Y = (Y_n)_{n\in\mathbb{N}}$ is said to have **eternal life time** with respect to the threshold c with $|c| < +\infty$ if

$$\forall n \in \mathbb{N} \ : \ \mathbb{P}(\{Y_{n+1} > c | Y_n > c\}) = 1, \tag{33}$$

otherwise **finite life time**.

In particular, in another words, we are interested on the preservation of the natural boundary $c = 0$ by numerical approximations. Set $\mathbb{D} = (0, +\infty)$ for the remaining part of this paper.

Theorem 5.1. *The one-dimensional BMM* (32) *satisfying (A6)-(A7) along partitions*

$$t_0 < t_1 < ... < t_n < t_{n+1} < ...$$

has an eternal life time with respect to the threshold $c = 0$ if the following additional conditions hold:

(A8) $\forall j = 1, ..., m \ \forall t_n \in [0, T]$ *and* $\forall x \in \mathbb{D}$

$$b^j(t_n, x)\frac{\partial}{\partial x} b^j(t_n, x) + d^j(t_n, x)x \ > \ 0, \tag{34}$$

(A9) $\forall t_n \in [0, T]$ *and* $\forall x \in \mathbb{D}$

$$x - \sum_{j=1}^{m} \frac{(b^j(t_n, x))^2}{2b^j(t_n, x)\frac{\partial}{\partial x} b^j(t_n, x) + 2d^j(t_n, x)x} \ > \ 0, \tag{35}$$

(A10) If

$$D(t_n, x) = a(t_n, x) - \frac{1}{2}\sum_{j=1}^{m} b^j(t_n, x)\frac{\partial}{\partial x} b^j(t_n, x) + d^0(t_n, x)x - \frac{1}{2}\sum_{j=1}^{m} d^j(t_n, x)x < 0$$

for a value $x \in \mathbb{D}$ at time-instant $t_n \in [0, T]$ then the current stepsize Δ_n is chosen such that $\forall t_n \in [0, T]$, $\forall x \in \mathbb{D}$

$$\Delta_n \ < \ \frac{x + N(t_n, x)}{-D(t_n, x)} \tag{36}$$

where

$$N(t_n, x) \ = \ -\sum_{j=1}^{m} \frac{(b^j(t_n, x))^2}{2b^j(t_n, x)\frac{\partial}{\partial x} b^j(t_n, x) + 2d^j(t_n, x)x}.$$

Remark 5.2. *The first restriction (A8) guarantees that the BMM inherits the positivity preserving structure of the underlying Milstein method. Condition (A9) is more technical, but in many applications this is valid without the use of the weight function d^1. (A10) is only neccessary if $D(t_n, x) < 0$, otherwise we can drop this restriction for positivity. So, we obtain a first idea to apply BMMs as advanced Milstein-type methods to preserve positivity by choosing d^0 and d^1 in such a way that $D(t_n, x)$ is greater than zero and we do not have to restrict the stepsize through* (36) *in this case. Furthermore, for the application of our results on mean and mean-square consistency to guarantee global mean square convergence with worst case rate* 1.0 *and positivity at the same time, we need to require that $D(t, x) \geq 0$. Note that the adapted, but random stepsize selection depending on current random outcomes Y_n by condition (A10) in the case of $D(t, x) < 0$ would contradict to the exclusive use of nonrandom stepsizes as exploited in our major convergence proof-steps in previous sections. Moreover, a restricted step size selection as given by (A10) throws out the problem of proving that any terminal time T can be reached in a finite time with probability one. So it is advantageous to require $D(t, x) \geq 0$ for all $x \geq 0$ and $0 \leq t \leq T$ for meaningful and practically relevant approximations.*

Proof of Theorem 5.1. Set $x = Y_n$. Using the one-step representation of the BMM (8) we obtain

$$\begin{aligned}
&\left(1 + d^0(t_n,x) + \frac{1}{2}\sum_{j=1}^m d^j(t_n,x)\left((\Delta W_n^j)^2 - \Delta_n\right)\right) Y_{n+1} \\
&= \Big(x + a(t_n,x)\Delta_n + \sum_{j=1}^m b^j(t_n,x)\Delta W_n^j + \frac{1}{2}\sum_{j=1}^m b^j(t_n,x)\frac{\partial}{\partial x}b^j(t_n,x)\left((\Delta W_n^j)^2 - \Delta_n\right) \\
&\quad + d^0(t_n,x)x\Delta_n + \frac{1}{2}\sum_{j=1}^m d^j(t_n,x)x\left((\Delta W_n^j)^2 - \Delta_n\right)\Big) \\
&= R(t_n, Y_n).
\end{aligned}$$

The expression (...) infront of Y_{n+1} at the left hand side of this equation is positive due to (A6) and (A7). Rewriting the right hand side leads to

(37)
$$R(t_n,x) = x + \left(a(t_n,x) - \frac{1}{2}\sum_{j=1}^m b^j(t_n,x)\frac{\partial}{\partial x}b^j(t_n,x) + d^0(t_n,x)x - \frac{1}{2}\sum_{j=1}^m d^j(t_n,x)x\right)\Delta_n + g(\Delta W_n^1, ..., \Delta W_n^m)$$

with

$$g(\Delta W_n^1, ..., \Delta W_n^m) = \sum_{j=1}^m b^j(t_n,x)\Delta W_n^j + \frac{1}{2}\sum_{j=1}^m \left(b^j(t_n,x)\frac{\partial}{\partial x}b^j(t_n,x) + d^j(t_n,x)x\right)(\Delta W_n^j)^2. \tag{38}$$

The function $g : \mathbb{R}^m \to \mathbb{R}^1$ possesses a global minimum due to (A8). More precisely, an obvious calculation shows that

$$\min_{z\in\mathbb{R}^m} g(z) = -\sum_{j=1}^m \frac{(b^j(t_n,x))^2}{2\left(b^j(t_n,x)\frac{\partial}{\partial x}b^j(t_n,x) + d^j(t_n,x)x\right)}. \tag{39}$$

This enables us to estimate R from below by replacing the value of $g(\Delta W_n^1, ..., \Delta W_n^m)$ by its minimum. So we arrive at

$$\begin{aligned}
R(t_n,x) \;\geq\; x \;+\; &\left(a(t_n,x) - \frac{1}{2}\sum_{j=1}^m b^j(t_n,x)\frac{\partial}{\partial x}b^j(t_n,x) + d^0(t_n,x)x - \frac{1}{2}\sum_{j=1}^m d^j(t_n,x)x\right)\Delta_n \\
&- \sum_{j=1}^m \frac{(b^j(t_n,x))^2}{2\left(b^j(t_n,x)\frac{\partial}{\partial x}b^j(t_n,x) + d^j(t_n,x)x\right)} \;=\; x + N(t_n,x) + D(t_n,x)\Delta_n
\end{aligned}$$

We can clearly see that (A9)-(A10) under (A8) are needed to get positive values $Y_{n+1} > 0$ whenever $Y_n > 0$ for all $n \in \mathbb{N}$. More precisely, if

$$D(t_n,x) = \left(a(t_n,x) - \frac{1}{2}\sum_{j=1}^m b^j(t_n,x)\frac{\partial}{\partial x}b^j(t_n,x) + d^0(t_n,x)x - \frac{1}{2}\sum_{j=1}^m d^j(t_n,x)x\right) \;\geq\; 0$$

then $R(t_n,x) > 0$ and we do not need any restriction of the stepsize Δ_n by (A10) at all. If $D(t_n,x) < 0$ then $x + N(t_n,x) + D(t_n,x)\Delta_n \geq 0$ guarantees that $R(t_n,x) > 0$, hence condition (A10) is needed in this case. Therefore, assumptions (A8)-(A10) imply the property of eternal life time of related BMMs with respect to the threshold $c = 0$. □

Remark 5.3. *The proof of Theorem 5.1 shows that the condition (A9) can be relaxed to*

$$x - \sum_{j=1}^m \frac{(b^j(t_n,x))^2}{2b^j(t_n,x)\frac{\partial}{\partial x}b^j(t_n,x) + 2d^j(t_n,x)x} \geq 0 \tag{40}$$

if $D(t_n,x) > 0$.

Moreover, in some cases it is more efficient to verify the following conditions instead of restrictions (A9) and (A10) known from Theorem 5.1.

Corollary 5.4. *The one-dimensional BMM* (32) *satisfying (A6)-(A7) along partitions*

$$t_0 < t_1 < \cdots < t_n < t_{n+1} < \cdots$$

has an eternal life time with respect to the threshold $c = 0$ *if condition (A8) is fulfilled and*

(A11) $\forall j = 1, ..., m \; \forall t_n \in [0, T]$ *and* $\forall x \in \mathbb{D}$

$$x + N(t_n, x) + D(t_n, x)\Delta_n > 0. \tag{41}$$

Where the functions N *and* D *are defined in Theorem 5.1.*

5.2 First applications of Theorem 5.1

To demonstrate the practial use of this result we present some examples with analytical positivity where the class of BMMs (6) proves to have preferrable positive-invariant integration schemes with higher order of accuracy.

Example 5.5. *Consider the one-dimensional geometric Brownian motion*

$$dX_t = \sum_{j=1}^{m} \sigma_j X_t \, dW_t^j, \qquad X_0 = x_0 > 0 \tag{42}$$

without any drift which is a standard example for stability analysis for numerical integration schemes for stochastic differential equations (e.g. see [14], [18], [22]) where the standard numerical methods possess serious stepsize restrictions or even fail to preserve stability and positivity. Kahl [8] showed that the Milstein scheme has an eternal life time if $m = 1$ *and*

$$\Delta_n < \frac{1}{\sigma_1^2}. \tag{43}$$

Obviously, for $|\sigma_1| \gg 1$*, this condition is very restrictive. Using any BMM with*

$$\begin{aligned} d^0(t_n, x) &\geq \frac{m}{2} \sum_{j=1}^{m} \sigma_j^2, \\ \forall l = 1, ..., m \qquad d^l(t_n, x) &= (m-1)\sigma_l^2 \end{aligned}$$

can solve this problem with higher order of accuracy very easy since

$$\begin{aligned} D(t_n, x) &= -\frac{1}{2} \sum_{j=1}^{m} \sigma_j^2 x + d^0(t_n, x)x - \frac{m-1}{2} \sum_{j=1}^{m} \sigma_j^2 x \\ &\geq -\frac{m}{2} \sum_{j=1}^{m} \sigma_j^2 x + \frac{m}{2} \sum_{j=1}^{m} \sigma_j^2 x \; = \; 0 \quad \textit{and} \\ N(t_n, x) &= -\sum_{j=1}^{m} \frac{\sigma_j^2 x^2}{2\sigma_j^2 x + 2(m-1)\sigma_j^2 x} \; = \; -\frac{x}{2} < 0 \end{aligned}$$

for $x \in \mathbb{D} = (0, +\infty)$*. Hence, the restriction (A10) on the stepsize* Δ_n *is not relevant here. However, notice that a restriction of the form* $D(t_n, x) \geq -K =$ *constant is important for the finiteness of the related numerical algorithm (i.e. in particular in order to reach any desired terminal time* $T > 0$ *with probability one). Moreover, all assumptions (A0)-(A9) are satisfied. Consequently, the related BMMs provide positive-invariant, consistent, stable and mean square converging numerical approximations to test SDE* (42) *with eternal life-time on* $\mathbb{D} = (0, +\infty)$*.*

Remark 5.6. *A simple choice to simulate SDE* (42) *on* $\mathbb{D} = (0, +\infty)$ *is given by BMMs* (6) *with weight functions* $d^0(t, x) = \frac{m}{2} \sum_{j=1}^{m} \sigma_j^2$ *and* $d^j(t, x) = 0$ *for all* $j = 1, 2, ..., m$*. To verify its positivity, one may apply Corollary 5.4 under the basic assumption that* $\sum_{j=1}^{m} \sigma_j^2 > 0$*. Also, note that the latter example can be even generalized to SDEs with diagonal noise on* $\mathbb{D} = (0, +\infty)^d$ *where all weights* d^j *in related diagonal BMMs are positive diagonal matrices, just by componentwise treatment.*

Next we want to focus our discussion on the mean-reverting process (31) and we will see that appropriately chosen BMMs possess an eternal life time with a suitable choice of the weight functions d^0 and d^1. According to the different behaviour of the mean-reverting process with respect to the parameter p we split up this discussion into several subclasses.

Corollary 5.7. *The one-dimensional BMM* (32) *satisfying (A6)-(A7) with* $\sigma^2 > 0$ *along partitions*

$$t_0 < t_1 < \cdots < t_n < t_{n+1} < \cdots$$

has an eternal life time with respect to the boundary $c = 0$ *for the mean-reverting process* (31) *with diffusion exponent* $p = 0.5$ *with the following choice of weight functions*

$$d^0(x) \;=\; \kappa, \qquad d^1(x) \;=\; 0. \tag{44}$$

Proof. We just have to check the requirements (A8)-(A10) of Theorem 5.1 with $a(x) = \kappa(\theta - x)$ and $b(x) = \sigma\sqrt{x}$ on $\mathbb{D} = (0, +\infty)$ since the conditions (A6)-(A7) are trivially valid with nonnegative weights d^0 and d^1. (A8) is satisfied while $\sigma^2 > 0$ since

$$\forall x \in \mathbb{D} \;:\; b(x)b'(x) + d^1(x) \;=\; p\sigma^2 x^{2p-1} \;=\; \frac{\sigma^2}{2} \;>\; 0$$

with $p = 1/2$. Next step is to calculate

$$D(t_n, x) \;=\; \kappa(\theta - x) - \frac{1}{4}\sigma^2 + \kappa x \;=\; \kappa\theta - \frac{1}{4}\sigma^2 \;>\; 0 \tag{45}$$

which is due to the requirement $\kappa\theta > \frac{1}{2}\sigma^2$ needed to guarantee strict positivity of exact solution. Hence we do not have to restrict the stepsize and (A10) is fulfilled too. The last step is to verify (A9)

$$x + N(t_n, x) \;=\; x - \frac{\sigma^2 x}{\sigma^2} \;=\; 0 \;\geq\; 0. \tag{46}$$

Hence, in view of Remark 5.3, the proof is completed. □

Remark 5.8. *The BMM with the forementioned choice of weight functions* d^j *coincides with the implicit Milstein method as discussed by Kahl [8] who has already proved the eternal life time of the implicit Milstein method in this case.*

Theorem 5.9. *The one-dimensional BMM* (32) *satisfying (A6)-(A7) with* $\sigma^2 > 0$ *along partitions*

$$t_0 < t_1 < \cdots < t_n < t_{n+1} < \cdots$$

has an eternal life time with respect to the boundary $c = 0$ *for the mean-reverting process* (31) *with diffusion exponent* $p \in (0.5, 1]$ *with the following choice of the weight functions*

$$d^0(x) = \alpha\kappa + \frac{1}{2}\sigma^2 p|x|^{(2p-2)}, \qquad d^1(x) \;=\; 0. \tag{47}$$

with relaxation parameter $\alpha \in [0, 1]$ *such that*

$$\Delta_n < \frac{2p-1}{2p\kappa(1-\alpha)}. \tag{48}$$

Remark 5.10. *The relaxation parameter* α *is similiar to the implicitness paramter* θ *in the class of stochastic* θ *methods and gives more space to adjust the BMM to the specific problem. The fully implicit case* $\alpha = 1$ *is a safe choice as we do not have to restrict the stepsize in that case. On the other hand numerical tests show that a reduced level of implicitness leads to better approximation results (see picture 2-4 and 5-7). Therefore we would recommend to use* $\alpha = 0.5$ *whenever the parameter configuration allows this choice, also supported by results from [18].*

Proof. With $a(x) = \kappa(\theta - x)$ and $b(x) = \sigma x^p$ on $\mathbb{D} = (0, +\infty)$ it is easy to see that (A8) is satisfied. Hence it is enough to check restriction (A11) of Corollary 5.4. Calculating

$$D(t_n, x) \;\geq\; \kappa(\theta - x) - \frac{1}{2}\sigma^2 p x^{(2p-1)} + \left(\alpha\kappa + \frac{1}{2}\sigma^2 p|x|^{(2p-2)}\right) x \;=\; \kappa\theta - \kappa(1-\alpha)x \;\geq\; -\kappa(1-\alpha)x \tag{49}$$

as well as

$$x + N(t_n, x) \;=\; x - \frac{\sigma^2 x^{(2p)}}{2p\sigma^2 x^{(2p-1)}} \;=\; x\left(1 - \frac{1}{2p}\right) \tag{50}$$

leads to

$$x + N(t_n, x) + D(t_n, x)\Delta_n \geq x\left(1 - \frac{1}{2p} - \Delta_n\left(\kappa(1-\alpha)\right)\right). \tag{51}$$

Eventually, the requirement $x + N(t_n, x) + D(t_n, x)\Delta_n > 0$ verifies the restriction (48) for Δ_n. □

Remark 5.11. *The choice of weights d^j has to be done with care in order to not to destroy the convergence rates as predicted in previous sections. This problem is strongly connected with the verification of the existence of control function $V(x)$ as met in (A1)-(A5) to guarantee the global rate 1.0 of strong and mean square convergence. In the case $p = 0.5$ or $p = 1$ one may take the standard function $V(x) = 1 + |x|^2$ to verify both convergence and positivity through applying our previous theorems. However, the case $p \in (0.5, 1)$ is more complicated.*

Remark 5.12. *The Cox-Ingersoll-Ross model*

$$dX_t = \kappa(\theta - X_t)dt + \sigma\sqrt{X_t}dW_t$$

with $\theta, \kappa, \sigma \geq 0$ is also treated in the paper of Alfonsi [1]. There one can find an alternative approach to numerical approximations of this model and further references on previous works. The main difficulty in this type of SDE results from the fact that the diffusion (volatility) coefficient is not Lipschitz. Our assumptions (A0), (A1), (A3) - (A9) for a detailed qualitative analysis of BMMs are clearly fulfilled. However, Theorem 4.1 works only under the assumption (A2) of mean and mean square consistency of related standard Milstein methods. To the best of our knowledge, mean square consistency for Milstein methods under assumptions other than of Lipschitz-type ones has not been proven in the literature so far (whereas the mean consistency can be established due to the Lipschitz-continuous drift coefficient $a(x) = \kappa(\theta - x)$). Thus, the question of mean square consistency and mean square convergence remains unsolved (it is left to future research due its complexity). The same observation applies to models with σX_t^p coefficients with $p \in (0.5, 1)$. An alternative numerical method is given by splitting techniques as introduced by Moro and Schurz [15] recently. There a proof of convergence of their split-step algorithm under Hölder-continuity assumptions of related dynamics can be found (which covers the case of more general model (31) with $p \in [0.5, 1)$ too).

5.3 Numerical tests

Finally we want to present some numerical results which underlines the necessity of construction of invariance-preserving numerical methods for the extended Cox-Ingersoll-Ross model. We compare the numerical approximations of standard Euler and Milstein methods versus the balanced methods BIM and BMM. In the following we consider the test equation

$$dX_t = (1 - X_t)dt + 1.4X_t^p dW_t, \qquad X(0) = 1, \tag{52}$$

with three different values of p ($p = 0.5$, $p = 3/4$ or $p = 1.0$). This parameter configuration guarantees strict analytical positivity even for $p = 0.5$ as $1 \cdot 1 > \frac{1.4^2}{2} = 0.98$. To get a first impression of this kind of mean-reverting processes Figure 1 illustrates the simulation of one path generated by the mentioned methods. For the plotting of the exact solution, we have used an essentially smaller step size $dt = 0.00125$ to generate the underlying path of Wiener process, whereas the stepsize Δ_n of numerical approximations is labeled by dT.

Due to the theoretical results, BIMs as well as BMMs are able to preserve the analytical positivity of this stochastic differential equation. Furthermore this parameter configuration forces the trajectories of Euler and Milstein schemes to become negative with positive probability.

To obtain a better feeling about the problems arising with negative paths we present in Table 1 the percentage of negative paths simulating the mean-reverting process (52). As one would expect, an increasing of the integration interval increases the percentage of negative paths in case of the Euler and Milstein methods. Vice versa, a decreasing of the integration stepsize dt leads to a decreasing number of negative paths. Both balanced methods BIM and BMM are not affected from this problem as we obtain positivity for all stepsizes dT independent of the length of integration interval $[0, T]$.

Next, an important aspect is the convergence speed of different methods where the most interesting aspect is a comparison of the two balanced methods BIM and BMM. The plots of the strong error of numerical approximations versus its equidistant stepsize $\Delta = dT$ are depicted in Figure 2-4. There we can clearly see that

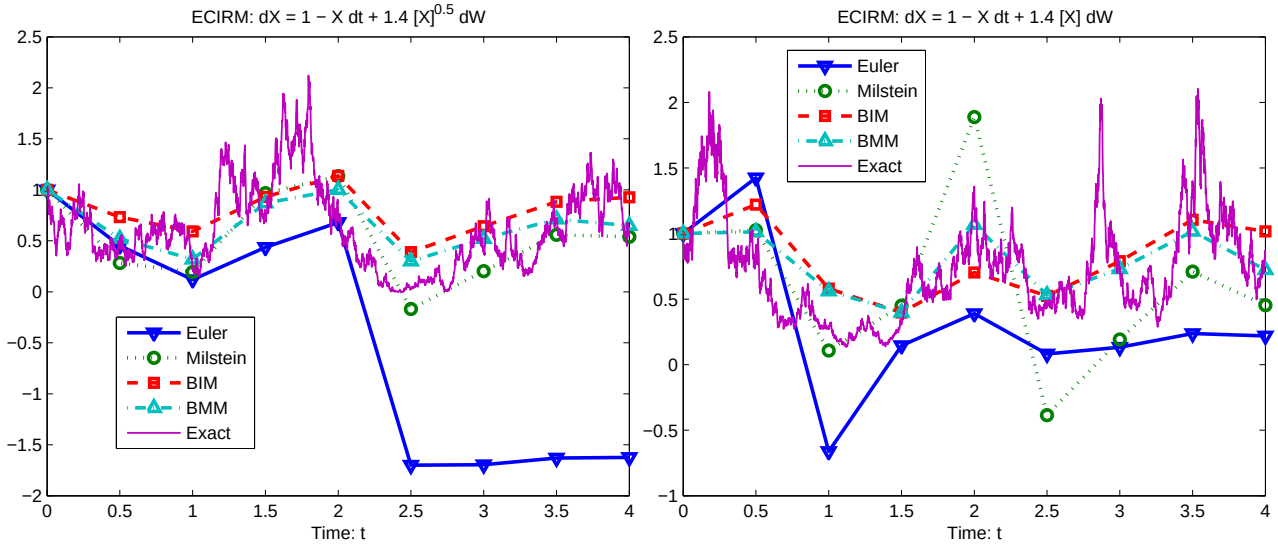

Figure 1: Left figure: $p = 0.5$, right figure: $p = 1.0$, time-interval: $[0, 4]$, stepsize of exact solution: $dt = 0.00125$, stepsize of numerical approximations: $dT = 0.5$, $\kappa = \theta = 1$, $\sigma = 1.4$, weight functions of BIM: $c^0(x) = 1$,$c^1(x) = \sigma|x|^{p-1}$, weight functions of BMM: for $(p = 0.5)$ - $d^0(x) = \kappa$, $d^1(x) = 0$ for $(p = 1.0)$ - $d^0(x) = \kappa + 0.5\sigma^2$

Time	Stepsize	Euler	Milstein	BIM	BMM
T = 1	$dT = \frac{1}{2}$	26.84%	22.68%	0%	0%
	$dT = \frac{1}{4}$	26.38%	7.92%	0%	0%
	$dT = \frac{1}{8}$	21.88%	0.7%	0%	0%
T = 4	$dT = \frac{1}{2}$	69.50%	53.50%	0%	0%
	$dT = \frac{1}{4}$	66.42%	19.90%	0%	0%
	$dT = \frac{1}{8}$	62.7%	1.92%	0%	0%
T = 16	$dT = \frac{1}{2}$	99.08%	93.22%	0%	0%
	$dT = \frac{1}{4}$	98.58%	56.62%	0%	0%
	$dT = \frac{1}{8}$	98.28%	8.24%	0%	0%

Table 1: Test results for $dX_t = (1 - X_t)dt + 1.4\sqrt{X_t}dW_t$; time-interval: $[0, T]$, stepsize of numerical approximations: dT, weight functions of BIM: $c^0(x) = 1$,$c^1(x) = \sigma/\sqrt{x}$, weight functions of BMM: $d^0(x) = \kappa$, $d^1(x) = 0$, number of simulated paths: 10000.

the balanced Milstein method is the best approximation method for relatively large stepsizes $\Delta = dT$ which is due to the positivity preserving behaviour. Hence we obtain the appearance that a small initial error of the numerical approximation does not result into an essentially larger becoming error at later integration times - a kind of fact which exhibits a stabilization effect on commonly known numerical dynamics and which is very important for simulation in computational finance where we are interested in longterm numerical approximations with large stepsizes. Nonetheless the better convergence property of the BMM would be useless if one integration step needs much more time than the Milstein method. Figure 5-7 shows that this is not the case and the BMM is indeed a superior scheme for large stepsizes.

Remark 5.13. *The comparison of the efficiency requires a slight modification of the Milstein method to prevent its trajectories from becoming negative as this would result in complex numbers. A simple modification is the* absorbing Milstein *method which sets negative values to zero in an additional step*

$$Y_{n+1} = \max\left(Y_n + \sum_{j=0}^{m} b^j(t_n, Y_n) I_{(j)}^{t_n,t_{n+1}} + \sum_{i,j=1}^{m} L^i b^j(t_n, Y_n) I_{(i,j)}^{t_n,t_{n+1}}, 0\right). \tag{53}$$

Furthermore Theorem 4.1 shows that the BMM is globally mean square convergent with order $r_g = 1.0$ and this numerical test indeed verifies all previously proven properties of BMMs by a practically oriented application.

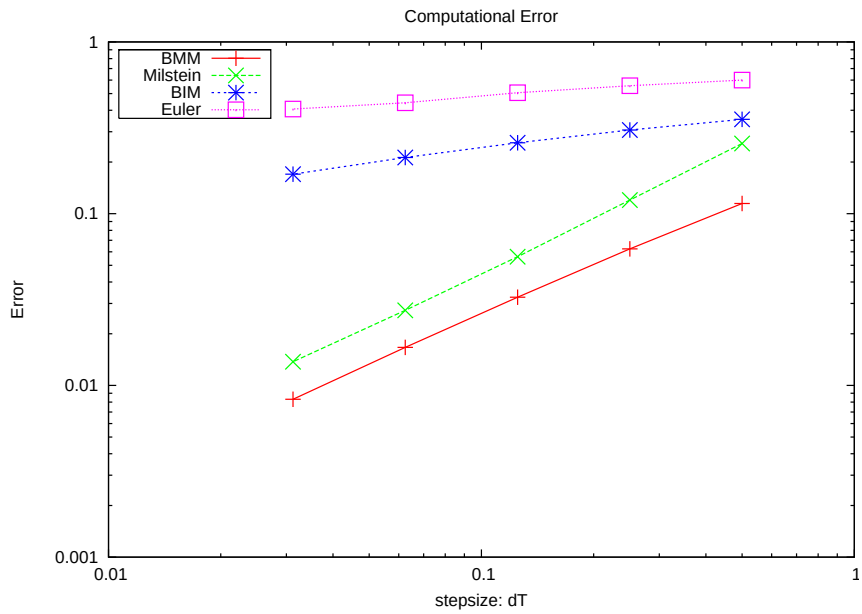

Figure 2: $dX_t = (1 - X_t)dt + 1.4\sqrt{X_t}dW_t$, time-interval: $[0, 1]$, stepsize of exact solution: $dt = 2^{-10}$, number of simulated paths: 100000, number generator: Sobol, L^2-Error $\left(\mathbb{E}\left[(X_{0,Z(0)}(T) - Y_{0,Z(0)}(T))^2|\mathcal{F}_0\right]\right)^{1/2}$ versus stepsize dT, $\kappa = \theta = 1$, $\sigma = 1.4$, weight functions of BIM: $c^0(x) = 1/2$,$c^1(x) = \sigma/\sqrt{x}$, weight functions of BMM: $d^0(x) = \kappa$, $d^1(x) = 0$

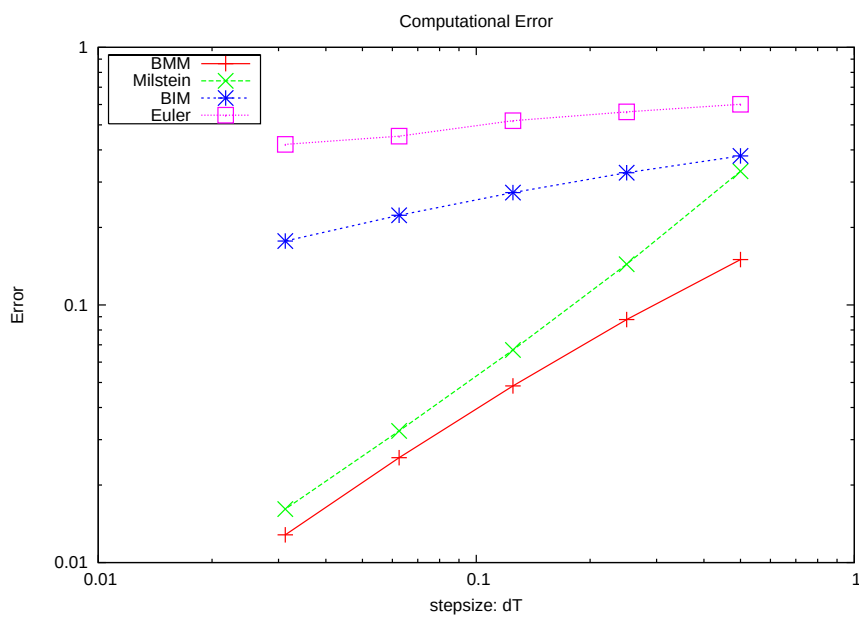

Figure 3: $dX_t = (1 - X_t)dt + 1.4X_t^{3/4}dW_t$, time-interval: $[0, 1]$, stepsize of exact solution: $dt = 2^{-10}$, number of simulated paths: 100000, number generator: Sobol, L^2-Error $\left(\mathbb{E}\left[(X_{0,Z(0)}(T) - Y_{0,Z(0)}(T))^2|\mathcal{F}_0\right]\right)^{1/2}$ versus stepsize dT, $\kappa = \theta = 1$, $\sigma = 1.4$, weight functions of BIM: $c^0(x) = 1/2$,$c^1(x) = \sigma|x|^{-1/4}$, weight functions of BMM: $d^0(x) = 0.5\kappa + 0.5\sigma^2 p/\sqrt{x}$, $d^1(x) = 0$

So a significant qualitative superiority of BMMs compared to commonly known numerical methods has been established by this paper.

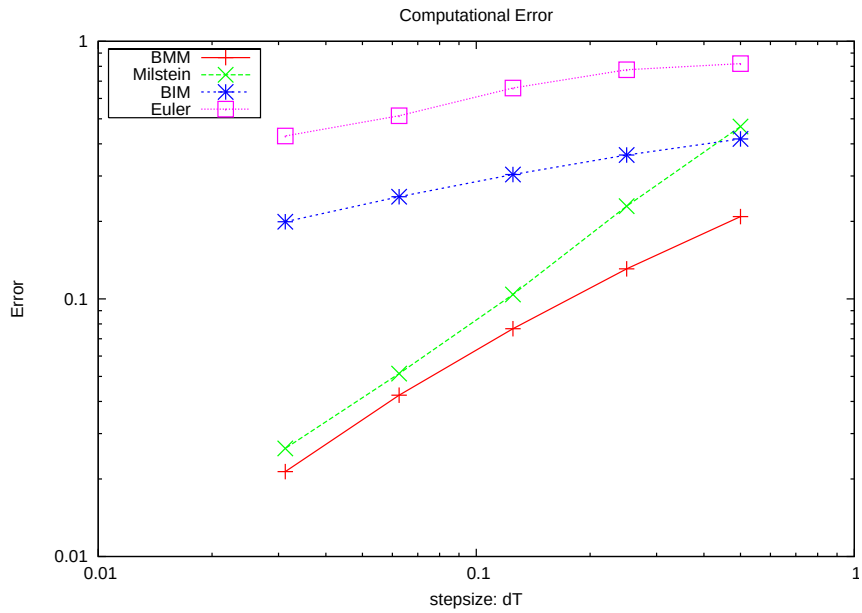

Figure 4: $dX_t = (1 - X_t)dt + 1.4X_t dW_t$, time-interval: $[0, 1]$, stepsize of exact solution: $dt = 2^{-10}$, number of simulated paths: 100000, number generator: Sobol, L^2-Error $\left(\mathbb{E}\left[(X_{0,Z(0)}(T) - Y_{0,Z(0)}(T))^2|\mathcal{F}_0\right]\right)^{1/2}$ versus stepsize dT, $\kappa = \theta = 1$, $\sigma = 1.4$, weight functions of BIM: $c^0(x) = 1/2$,$c^1(x) = \sigma$, weight functions of BMM: $d^0(x) = 0.5\kappa + 0.5\sigma^2$, $d^1(x) = 0$

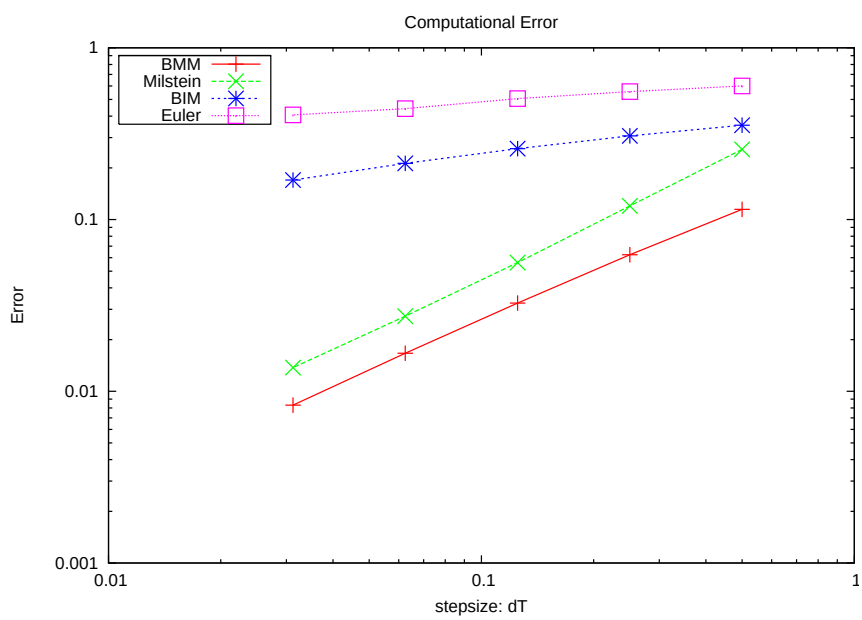

Figure 5: $dX_t = (1 - X_t)dt + 1.4\sqrt{X_t}dW_t$, time-interval: $[0, 1]$, stepsize of exact solution: $dt = 2^{-10}$, number of simulated paths: 100000, number generator: Sobol, L^2-Error $\left(\mathbb{E}\left[(X_{0,Z(0)}(T) - Y_{0,Z(0)}(T))^2|\mathcal{F}_0\right]\right)^{1/2}$ versus elapsed time, $\kappa = \theta = 1$, $\sigma = 1.4$, weight functions of BIM: $c^0(x) = 1/2$,$c^1(x) = \sigma/\sqrt{x}$, weight functions of BMM: $d^0(x) = \kappa$, $d^1(x) = 0$,

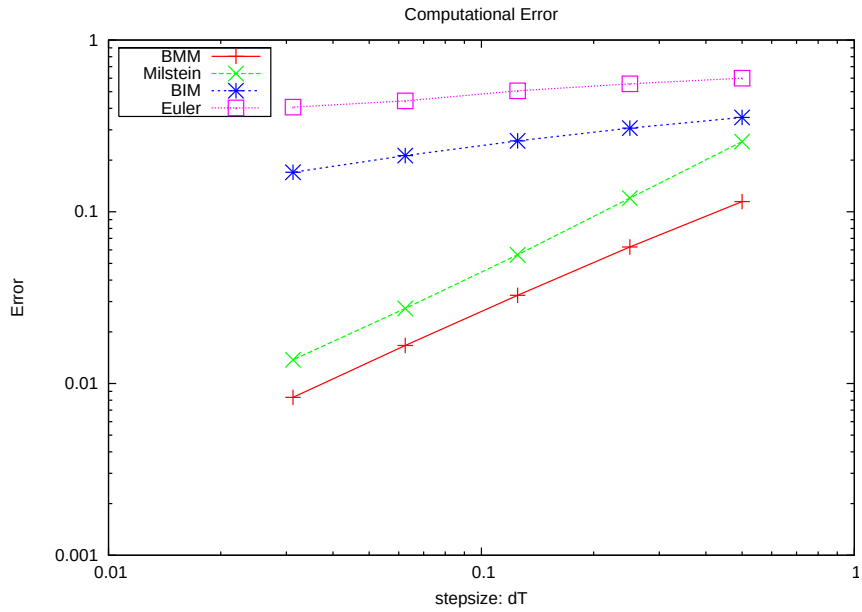

Figure 6: $dX_t = (1 - X_t)dt + 1.4X_t^{3/4}dW_t$, time-interval: $[0, 1]$, stepsize of exact solution: $dt = 2^{-10}$, number of simulated paths: 100000, number generator: Sobol, L^2-Error $\left(\mathbb{E}\left[(X_{0,Z(0)}(T) - Y_{0,Z(0)}(T))^2|\mathcal{F}_0\right]\right)^{1/2}$ versus elapsed time, $\kappa = \theta = 1$, $\sigma = 1.4$, weight functions of BIM: $c^0(x) = 1/2$,$c^1(x) = \sigma|x|^{-1/4}$, weight functions of BMM: $d^0(x) = 0.5\kappa + 0.5\sigma^2 p/\sqrt{x}$. $d^1(x) = 0$

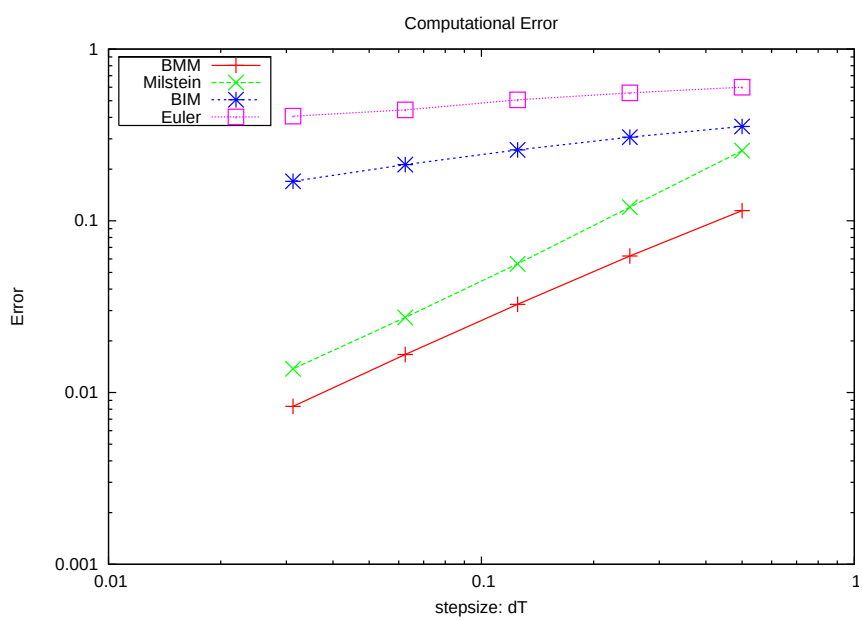

Figure 7: $dX_t = (1 - X_t dt) + 1.4X_t dW_t$, time-interval: $[0, 1]$, stepsize of exact solution: $dt = 2^{-10}$, number of simulated paths: 100000, number generator: Sobol, L^2-Error $\left(\mathbb{E}\left[(X_{0,Z(0)}(T) - Y_{0,Z(0)}(T))^2|\mathcal{F}_0\right]\right)^{1/2}$ versus elapsed time, $\kappa = \theta = 1$, $\sigma = 1.4$, weight functions of BIM: $c^0(x) = 1/2$,$c^1(x) = \sigma$, weight functions of BMM: $d^0(x) = 0.5\kappa + 0.5\sigma^2$, $d^1(x) = 0$

Acknowledgement

The authors are grateful for the valuable comments of the anonymous referees.

References

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

[2] L.B.G. Andersen and R. Brotherton-Ratcliffe, Extended Libor Market Models with Stochastic Volatility, Working Paper, Social Science Research Network, December 2001. `http://ssrn.com/abstract=294853`

[3] L.B.G. Andersen and V. Piterbarg, Moment Explosions in Stochastic Volatility Models, Working Paper, Social Science Research Network, April 2004. `http://ssrn.com/abstract=559481`

[4] L. Arnold, Stochastic Differential Equations: Theory and Applications, Wiley, New York, 1974.

[5] J. Cox, J. Ingersoll and S. Ross, A Theory of the Term Structures of Interest Rates, Econometrica 3 (1985), 229–263.

[6] K. Itô, Stochastic integral, Proc. Imperial Acad. Tokyo 20 (1944), 519–524.

[7] K. Itô, On a formula concerning stochastic differentials, Nagoya Math. J. 3 (1951), 55–65.

[8] C. Kahl, Positive numerical integration of stochastic differential equations, Diplomarbeit, University of Wuppertal, Wuppertal, Germany, 2004.

[9] C. Kahl, M. Günther and T. Roßberg, Structure preserving stochastic integration schemes in interest rate derivative modeling, Preprint BUW-AMNA 04/10, Bergische Universität, Wuppertal, 2004.

[10] R.Z. Khas'minskij, Stochastic Stability of Differential Equations, Sijthoff & Noordhoff, Alphen aan den Rijn, 1980.

[11] P.E. Kloeden and E. Platen, Numerical Solution of Stochastic Differential Equations, Springer-Verlag, Berlin, 1991.

[12] P.E. Kloeden, E. Platen and H. Schurz, Numerical Solution of SDEs through Computer Experiments, 2nd edition, Universitext, Springer-Verlag, Berlin, 1997.

[13] G.N. Milstein, Numerical Integration of Stochastic Differential Equations, Sverdlovsk, 1988 (English translation by Kluwer, Dordrecht, 1995.

[14] 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.

[15] E. Moro and H. Schurz, Nonnegativity preserving numerical algorithms for stochastic differential equations, Preprint M-05-006, pp. 23, Department of Mathematics, Carbondale, 2005.

[16] P. Protter, Stochastic Integration and Differential Equations, Springer-Verlag, Berlin, 1990.

[17] H. Schurz, Numerical regularization for SDEs: Construction of nonnegative solutions, Dynam. Syst. Appl. 5 (1996), No. 1, 323–352.

[18] H. Schurz, Stability, Stationarity, and Boundedness of some Implicit Numerical Methods for Stochastic Differential Equations and Applications, Logos-Verlag, Berlin, 1997.

[19] H. Schurz, *Numerical analysis of (ordinary) stochastic differential equations without tears.* In Handbook of Stochastic Analysis (Eds. D. Kannan and V. Lakshmikantham), p. 237-359, Marcel Dekker, Basel, 2002.

[20] H. Schurz, General theorems for numerical approximation of stochastic processes on the Hilbert space H_2, Electr. Trans. Numer. Anal. 16 (2003) 50–69.

[21] H. Schurz, An axiomatic approach to the approximation of some Hilbert-space-valued stochastic processes, pp. 18, submitted (also Preprint M-02-007, Department of Mathematics, Southern Illinois University, Carbondale, IL, 2002).

[22] 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.

[23] 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