DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

A Derandomization of the Euler Scheme for Scalar Stochastic Differential Equations

T. Müller-Gronbach, K. Ritter, L. Yaroslavtseva

Preprint 80

Edited by

AG Numerik/Optimierung
Fachbereich 12 - Mathematik und Informatik
Philipps-Universität Marburg
Hans-Meerwein-Str.
35032 Marburg

DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

A Derandomization of the Euler Scheme for Scalar Stochastic Differential Equations

T. Müller-Gronbach, K. Ritter, L. Yaroslavtseva

Preprint 80

A DERANDOMIZATION OF THE EULER SCHEME FOR SCALAR STOCHASTIC DIFFERENTIAL EQUATIONS

THOMAS MÜLLER-GRONBACH, KLAUS RITTER, AND LARISA YAROSLAVTSEVA

ABSTRACT. Consider a scalar stochastic differential equation with solution process X. We present a deterministic algorithm to approximate the marginal distribution of X at $t = 1$ by a discrete distribution, and hereby we get a deterministic quadrature rule for expectations $\mathbb{E}(f(X(1))$. The construction of the algorithm is based on a derandomization of the Euler scheme. We provide a worst case analysis for the computational cost and the error, assuming that the coefficients of the equation have bounded derivatives up to order four and that the derivatives of f are polynomially bounded up to order four. In terms of the computational cost the error is almost of the order $2/3$, if the diffusion coefficient is bounded away from zero, and in general we almost achieve the order $1/2$.

1. Introduction

Consider a scalar autonomous stochastic differential equation

$$
\begin{aligned}
dX(t) &= a(X(t))\,dt + b(X(t))\,dW(t), \quad t \in [0,1], \\
X(0) &= x,
\end{aligned} \tag{1}
$$

with drift coefficient $a : \mathbb{R} \to \mathbb{R}$ and diffusion coefficient $b : \mathbb{R} \to \mathbb{R}$, initial value $x \in \mathbb{R}$, and driving Brownian motion W, and let

$$
S(x,a,b) = \mathbb{P}_{X(1)}
$$

denote the distribution of the solution X at time $t = 1$. We present an algorithm $\widehat{S}$ that computes a discrete distribution

$$
\widehat{S}(x,a,b) = \sum_{i=1}^{N} c_i \cdot \delta_{y_i}
$$

as an approximation to $S(x,a,b)$, which obviously provides a quadrature formula

$$
\int_{\mathbb{R}} f \, d\widehat{S}(x,a,b) = \sum_{i=1}^{N} c_i \cdot f(y_i)
$$

for the integral

$$
\int_{\mathbb{R}} f \, dS(x,a,b) = \mathbb{E} f(X(1))
$$

of functions $f\colon \mathbb{R} \to \mathbb{R}$ w.r.t. $S(x,a,b)$.

We roughly explain the construction of $\widehat{S}(x,a,b)$ and discuss its properties in the case that b is bounded away from zero and, for simplicity, that $x=0$. Then $G = \{y_1,\dots,y_N\}$ is a set of equidistant nodes with center at zero and with spacing adjusted to N and to the minimum value of $|b|$. The corresponding weights $c_1,\dots,c_N$ only depend on the values of the coefficients a and b at the nodes y_i, and they are given by the distribution of a Markov chain with initial value $x=0$ and state space G after approximately $N^{2-\delta}$ steps, where $\delta > 0$ is a parameter of the algorithm. The transition probabilities of the Markov chain are obtained by applying a derandomization procedure to the respective Euler scheme with approximately $N^{2-\delta}$ equidistant steps in the interval $[0,1]$. Hereby, an Euler step is replaced by a step on the discrete set G to at most 6 possible positions. Therefore the resulting transition matrix is sparse, and the total computational cost $\operatorname{cost}(\widehat{S},(x,a,b))$ to provide the nodes and the weights is proportional to $N^{3-\delta}$.

To define the error of $\widehat{S}$ we consider the class

$$\mathcal{F}(\beta) = \{f \in C^4(\mathbb{R})\colon\ |f^{(\ell)}(u)| \le 1 + |u|^\beta, u \in \mathbb{R}, \ell = 1,\dots,4\} \tag{2}$$

of integrands with polynomially bounded derivatives up to order four, and we introduce a metric ρ on the set of all Borel probability measures on $\mathbb{R}$ with finite absolute moments of order $\beta+1$ by

$$\rho(\mu,\widehat{\mu}) = \sup_{f\in\mathcal{F}(\beta)} \left| \int_{\mathbb{R}} f\,d\mu - \int_{\mathbb{R}} f\,d\widehat{\mu} \right|. \tag{3}$$

For the coefficients of the equation (1) we also impose smoothness assumptions and we perform a worst case analysis, too. We show that

$$\sup_{x,a,b} \rho(S(x,a,b),\widehat{S}(x,a,b)) \le c \cdot \sup_{x,a,b} \operatorname{cost}(\widehat{S},(x,a,b))^{-2/3+\delta},$$

where the supremum is taken over all coefficients a and b that are four times continuously differentiable with bounded derivatives up to order four and over all initial values from a compact interval. The constant c only depends on β, the bounds for the derivatives of a and b, the bound on $|x|$, and on δ. Our algorithm $\widehat{S}$ thus almost achieves the order $2/3$ of convergence in terms of its total computational cost.

The algorithm $\widehat{S}$ is constructed in a similar way if the diffusion coefficient b is not bounded away from zero, but in this case we can only prove that the order of convergence in terms of the computational cost is almost $1/2$, up to now.

We conclude the introduction by a discussion of our result. At first we relate the approximation problem for $S(x,a,b)$ to integration on the real line or, more generally, on $\mathbb{R}^d$. Let μ denote a probability measure on $\mathbb{R}^d$ with a Lebesgue density that satisfies suitable decay properties, and let $\mathcal{F}^r(\beta)$ be defined analogously to (2) with polynomially bounded derivatives up to order $r \in \mathbb{N}$. Suppose that the metric ρ is defined via (3) with $\mathcal{F}^r(\beta)$ instead of $\mathcal{F}(\beta)$. Then there exists a sequence of N-point distributions $\widehat{\mu}_N$ on $\mathbb{R}^d$ such that

$$\rho(\mu, \widehat{\mu}_N) \leq c \cdot N^{-r/d}, \tag{4}$$

which follows from general results on weighted approximation and integration in [15, 16]. Moreover, if the density of μ is bounded away from zero on an open ball, then a matching lower bound holds for every N-point distribution on $\mathbb{R}^d$. In particular, for $r = 4$ and $d = 1$ we get the order 4 of convergence, which is substantially better than the order 2/3 or 1/2 as in our result. This gap is due to the following differences concerning the assumptions and the analysis. The construction of the weights that leads to (4) basically requires the density of μ to be explicitly known, while in our setting the distribution $\mu = S(x, a, b)$ is only given implicitly, and we only have access to function values of the coefficients a and b of the equation (1). Moreover, we fully take into account the computational cost to construct $\widehat{\mu}_N = \widehat{S}(x, a, b)$, while the estimate (4) only depends on the size N of the support of $\widehat{\mu}_N$.

In this sense we are not studying a quadrature problem but the construction of quadrature formulas. The latter is a non-linear problem, and standard techniques to derive lower bounds for the error in terms of the computational cost are less powerful in this setting. Actually, it seems to be challenging to close the gap between a lower bound of order 4 and our upper bound of order 2/3 or 1/2, respectively. In a different setting sharp upper and lower bounds for approximation of a marginal distribution of the solution of a stochastic differential equation have been obtained in [14].

In the particular case of $r = 1$ and $\beta = 0$ the class $\mathcal{F}^1(0)$ essentially is the class of Lipschitz continuous functions with Lipschitz constant bounded by one, and ρ essentially is a Wasserstein metric. Best approximation of a probability distribution μ on a separable metric space $\mathcal{M}$ by means of a discrete distribution $\widehat{\mu}_N$ w.r.t. a Wasserstein metric is called quantization, and we refer to the monograph [6] for quantization on finite-dimensional spaces $\mathcal{M}$ and to the surveys [1, 13] for quantization on infinite-dimensional spaces $\mathcal{M}$. We stress again that, in the finite-dimensional case, the known deterministic constructions of good approximations $\widehat{\mu}_N$ are not applicable in our setting, since the distribution of $X(1)$ is only given implicitly and the Lebesgue density, if it exists at all, is unknown in general. However, probabilistic methods for quantization of implicitly given distributions have recently been introduced in [2].

The situation is different if, instead of approximating a marginal distribution of a stochastic differential equation, we aim at the distribution on the path space, which constitutes an infinite-dimensional quantization problem. For scalar equations a fully constructive method for quantization is presented in [12], and it achieves strong asymptotic optimality in terms of the number N of points (i.e., of paths), while the computational cost is essentially given by N.

Finally, let us consider the weak Euler scheme for equation (1). Under the assumption of (polynomially) bounded derivatives up to order four of a, b and f, the bias of the Euler scheme is of the order 1 in terms of the number of equidistant time steps, see [7], and balancing the number of steps and the number of replications the Monte Carlo Euler algorithm yields the order 1/3 in terms of the computational cost. This can be substantially improved to the order 1/2 by the multi-level technique, see [5]. We achieve this order, too, by means of a deterministic algorithm, and we even achieve the order 2/3 if the diffusion coefficient b is bounded away from zero.

Quadrature formulas on the Wiener space, which are based on paths of bounded variation and are exact for iterated integrals up to a fixed degree m, are introduced and further developed in [8, 9, 10, 11]. Here, finite-dimensional stochastic differential equations with smooth coefficients a and b are considered, and an approximation to the marginal distribution $P_{X(1)}$ of the solution X is obtained by iteratively solving a collection of ODEs on k non-equidistant time intervals. For Lipschitz continuous integrands an error bound of order $k^{-(m-1)/2}$ is achieved. However, the number of ODEs to be solved grows polynomially in k, and the impact of this numerical task on the total computational cost of the method seems not to have been investigated in full detail so far.

We briefly outline the content of the paper. Our algorithm is presented in Section 2. In Section 3 we discuss the computational cost and the error of our method, and proofs are postponed to Section 4 and the Appendix.

2. The Algorithm

The algorithm depends on the parameters $\delta > 0, \varepsilon \in (0, 1]$, and $m \in \mathbb{N}$. Put

$$d = \varepsilon \cdot m^{-1/2}$$

as well as

$$J = \lceil m^{\delta} \cdot d^{-1} \rceil,$$

and let

$$G = \{i \cdot d : \, i = -J, \ldots, J\}.$$

For $x \in \mathbb{R}$ and $a, b : \mathbb{R} \to \mathbb{R}$ the algorithm yields a discrete distribution that is concentrated on the set $G \cup \{x\}$. The computation of the corresponding weights involves a transition matrix

$$Q = (q_{y,\widetilde{y}})_{y,\widetilde{y} \in G \cup \{x\}}$$

on the state space $G \cup \{x\}$, which is defined as follows. For $y \in \mathbb{R}$ we put

$$z_y = y + a(y) \cdot m^{-1}, \quad \bar{z}_y = \min\{i \cdot d : \, z_y \le i \cdot d, i \in \mathbb{Z}\},$$

as well as

$$u_y = d^{-1} \cdot (\bar{z}_y - z_y), \quad k_y = \lceil |b(y)| \cdot \varepsilon^{-1} \rceil.$$

Clearly, z_y corresponds to an Euler step of length m^{-1} for the deterministic counterpart of (1), and $\bar{z}_y$ essentially is a projection of z_y onto G. Note that $0 \le u_y < 1$ for every $y \in \mathbb{R}$. For the definition of $q_{y,\widetilde{y}}$ we distinguish three cases, given by

$$\begin{aligned} G_1 &= \{y \in G \cup \{x\} : \bar{z}_y \cdot d^{-1} \notin (-J + k_y, J - k_y]\}, \\ G_2 &= \{y \in G \cup \{x\} : \bar{z}_y \cdot d^{-1} \in (-J + k_y, J - k_y], \, |b(y)| \le \varepsilon\}, \\ G_3 &= \{y \in G \cup \{x\} : \bar{z}_y \cdot d^{-1} \in (-J + k_y, J - k_y], \, |b(y)| > \varepsilon\}. \end{aligned}$$

The points $y \in G_1$, where z_y is close to the extremal points $\pm J \cdot d$ of G, are absorbing states, i.e.,

$$q_{y,\widetilde{y}} = \begin{cases} 1 & \text{if } \widetilde{y} = y, \\ 0 & \text{otherwise.} \end{cases}$$

For $y \in G_2$ the diffusion is small and $\bar{z}_y, \bar{z}_y - d \in G$, and we define

$$q_{y,\widetilde{y}} = \begin{cases} 1 - u_y & \text{if } \widetilde{y} = \bar{z}_y, \\ u_y & \text{if } \widetilde{y} = \bar{z}_y - d, \\ 0 & \text{otherwise.} \end{cases}$$

Finally, we consider the case $y \in G_3$ of states with a large diffusion. We put

$$\vartheta_y^{(1)} = \frac{b^2(y) \cdot \varepsilon^{-2}}{2k_y^2} + \frac{u_y^2 - 2u_y}{6k_y^2}, \quad \vartheta_y^{(2)} = \frac{b^2(y) \cdot \varepsilon^{-2}}{2k_y^2} + \frac{u_y^2 - 1}{6k_y^2}.$$

Then

$$\vartheta_y^{(j)} \le \frac{b^2(y) \cdot \varepsilon^{-2}}{2k_y^2} \le \frac{1}{2}.$$

Moreover, $y \in G_3$ implies

$$\vartheta_y^{(j)} \ge \frac{b^2(y) \cdot \varepsilon^{-2}}{2k_y^2} - \frac{1}{6k_y^2} > 0$$

as well as $k_y \ge 2$ and $\bar{z}_y + i \cdot d \in G$ for $i = -k_y - 1, \ldots, k_y$. We define

$$q_{y,\widetilde{y}} = \begin{cases} (1 - u_y) \cdot (1 - 2\vartheta_y^{(1)}) & \text{if } \widetilde{y} = \bar{z}_y, \\ (1 - u_y) \cdot \vartheta_y^{(1)} & \text{if } \widetilde{y} = \bar{z}_y \pm k_y \cdot d, \\ u_y \cdot (1 - 2\vartheta_y^{(2)}) & \text{if } \widetilde{y} = \bar{z}_y - d, \\ u_y \cdot \vartheta_y^{(2)} & \text{if } \widetilde{y} = \bar{z}_y - d \pm k_y \cdot d, \\ 0 & \text{otherwise.} \end{cases}$$

We compute the probability vector $((Q^m)_{x,y})_{y \in G \cup \{x\}}$, which specifies the discrete distribution

$$\widehat{S}_{\delta,\varepsilon,m}(x, a, b) = \sum_{y \in G \cup \{x\}} (Q^m)_{x,y} \cdot \delta_y. \tag{5}$$

In different terms, the distribution $S(x, a, b)$ of the solution of (1) at time $t = 1$ is approximated by the m-step transition probability of a homogeneous Markov chain with state space $G \cup \{x\}$, initial value x, and transition matrix Q.

3. Analysis of Cost and Error

Throughout the following we use c, $c(K), \dots$ to denote unspecified positive constants, which only depend on the parameters specified inside the brackets.

We first discuss the computational cost of the method $\widehat{S}_{\delta,\varepsilon,m}$. For a given input (x, a, b) we consider

- the number $\#_{\mathrm{coeff}}$ of evaluations of the drift or diffusion coefficient a or b, respectively, and
- the number $\#_{\mathrm{op}}$ of arithmetical operations

needed to compute the approximation $\widehat{S}_{\delta,\varepsilon,m}(x, a, b)$, and we define the computational cost of $\widehat{S}_{\delta,\varepsilon,m}$ for (x, a, b) by

$$\mathrm{cost}(\widehat{S}_{\delta,\varepsilon,m}, (x, a, b)) = \#_{\mathrm{coeff}} + \#_{\mathrm{op}}.$$

Lemma 1. *For all $\delta > 0$, $\varepsilon \in (0, 1]$, $m \in \mathbb{N}$, every $x \in \mathbb{R}$ and all $a, b : \mathbb{R} \to \mathbb{R}$ we have*

$$\mathrm{cost}(\widehat{S}_{\delta,\varepsilon,m}, (x, a, b)) \le c \cdot \varepsilon^{-1} \cdot m^{3/2+\delta}.$$

Proof. At most $2J + 2$ evaluations of a and b and at most $c \cdot J$ arithmetical operations are needed to compute all non-zero transition probabilities $q_{y,\widetilde{y}}$ together with their respective positions in the transition matrix Q. Clearly, there are at most $6(2J+2)$ non-zero entries of Q and therefore at most $12(2J + 2)$ arithmetical operations are needed to compute $v^T \cdot Q$ for any vector v. Consequently, at most $m \cdot 12(2J + 2)$ arithmetical operations are needed to compute all m-step transition probabilities $(Q^m)_{x,y}$. Summing up, we obtain

$$\#_{\mathrm{coeff}} + \#_{\mathrm{op}} \le c \cdot m \cdot J \le c \cdot \varepsilon^{-1} \cdot m^{3/2+\delta}$$

as claimed. □

Note that the cost of $\widehat{S}_{\delta,\varepsilon,m}$ for (x, a, b) is much larger than the size of the support of $\widehat{S}_{\delta,\varepsilon,m}(x, a, b)$, which is bounded by $c \cdot \varepsilon^{-1} \cdot m^{1/2+\delta}$. The cost to actually apply the quadrature formula induced by $\widehat{S}_{\delta,\varepsilon,m}(x, a, b)$ is therefore dominated by the cost to compute the weights.

We turn to the analysis of the error. Recall that the underlying metric ρ has already been defined by (2) and (3). To specify the smoothness assumption on the coefficients of the equation (1) we define

$$\mathcal{H}(K) = \{h \in C^4(\mathbb{R})\colon\ |h(0)|, \|h^{(\ell)}\|_\infty \le K, \ell = 1, \dots, 4\}$$

for $K > 0$, and we suppose that $a, b \in \mathcal{H}(K)$. Clearly, $h \in \mathcal{H}(K)$ if and only if

$$|h(u)| \le K \cdot (1 + |u|) \tag{6}$$

as well as

$$|h^{(\ell)}(u) - h^{(\ell)}(v)| \le K \cdot |u - v| \tag{7}$$

for $\ell = 0, \dots, 3$ and $u, v \in \mathbb{R}$. Finally, we require that the initial value x belongs to some compact interval $[-L, L]$. Altogether we consider the set

$$\mathcal{I} = [-L, L] \times \mathcal{H}(K) \times \mathcal{H}(K)$$

of inputs (x, a, b) as well as the subset

$$\mathcal{I}_\varepsilon(K) = [-L, L] \times \mathcal{H}(K) \times \mathcal{H}_\varepsilon(K)$$

with

$$\mathcal{H}_\varepsilon(K) = \{h \in \mathcal{H}(K) : |h| > \varepsilon\}$$

for $\varepsilon \in (0, 1]$, which corresponds to a non-degeneracy constraint on the coefficient b.

The worst case cost and the worst case error of $\widehat{S}_{\delta,\varepsilon,m}$ on $\mathcal{I}$ are defined by

$$\begin{aligned} \operatorname{cost}(\widehat{S}_{\delta,\varepsilon,m}, \mathcal{I}) &= \sup\bigl\{\operatorname{cost}(\widehat{S}_{\delta,\varepsilon,m}, (x, a, b)) : (x, a, b) \in \mathcal{I}\bigr\}, \\ e(\widehat{S}_{\delta,\varepsilon,m}, \mathcal{I}) &= \sup\bigl\{\rho(S(x, a, b), \widehat{S}_{\delta,\varepsilon,m}, (x, a, b)) : (x, a, b) \in \mathcal{I}\bigr\}, \end{aligned}$$

and the counterparts on $\mathcal{I}_\varepsilon$ are denoted by $\operatorname{cost}(\widehat{S}_{\delta,\varepsilon,m}, \mathcal{I}_\varepsilon)$ and $e(\widehat{S}_{\delta,\varepsilon,m}, \mathcal{I}_\varepsilon)$.

Theorem 1. *Let $L, K, \beta > 0$ and $\varepsilon \in (0, 1]$. Then*

$$e(\widehat{S}_{\delta,m^{-1/2},m}, \mathcal{I}) \le c(L, K, \beta, \delta) \cdot m^{-1}$$

and

$$e(\widehat{S}_{\delta,\varepsilon,m}, \mathcal{I}_\varepsilon) \le c(L, K, \beta, \delta) \cdot m^{-1}$$

for all $m \in \mathbb{N}$ and $\delta > 0$.

Lemma 1 and Theorem 1 imply the following result.

Theorem 2. *Let $L, K, \beta > 0$ and $\varepsilon \in (0, 1]$. Then*

$$e(\widehat{S}_{\delta,m^{-1/2},m}, \mathcal{I}) \le c(L, K, \beta, \delta) \cdot (\operatorname{cost}(\widehat{S}_{\delta,m^{-1/2},m}, \mathcal{I}))^{-\frac{1}{2+\delta}}$$

and

$$e(\widehat{S}_{\delta,\varepsilon,m}, \mathcal{I}_\varepsilon) \le c(L, K, \beta, \delta) \cdot (\operatorname{cost}(\widehat{S}_{\delta,\varepsilon,m}, \mathcal{I}_\varepsilon))^{-\frac{1}{3/2+\delta}}$$

for all $m \in \mathbb{N}$ and $\delta > 0$.

4. Proofs

Throughout this section we fix $\delta > 0$, $\varepsilon \in (0, 1]$, $m \in \mathbb{N}$, as well as $L, K > 0$, and we assume that

$$x \in [-L, L], \quad a, b \in \mathcal{H}(K).$$

We refer to Section 2 for the definition of the corresponding terms d, J, G, z_y, $\bar{z}_y$, u_y, k_y, G_i, $\vartheta_y^{(j)}$, $q_{y,\widetilde{y}}$, and Q. Furthermore, we write X^x instead of X for the solution of (1) to stress the dependence on the initial value x.

Let Z denote a standard normal variable. For every $y \in \mathbb{R}$ we put

$$Z_y = z_y + b(y) \cdot m^{-1/2} \cdot Z,$$

which corresponds to an Euler step of length m^{-1} for the equation (1), starting at y. We define

$$\Lambda_y^{(p)} = |\mathbb{E}(X^y(m^{-1}) - z_y)^p - \mathbb{E}(Z_y - z_y)^p|$$

for $p \in \mathbb{N}$ in order to compare moments of the solution and the Euler scheme.

Lemma 2. *We have*

$$\Lambda_y^{(p)} \le c(K) \cdot (1 + |y|^3) \cdot m^{-2}$$

for all $y \in \mathbb{R}$ and $p = 1, 2, 3$.

Proof. Put

$$R_1 = \int_0^{m^{-1}} (a(X^y(s)) - a(y))\, ds,$$

as well as

$$R_2 = \int_0^{m^{-1}} (b(X^y(s)) - b(y))\, dW(s), \quad R_3 = \int_0^{m^{-1}} b(X^y(s))\, dW(s),$$

and let $q \in \mathbb{N}$. Use property (7) of a and Lemma 8 from the Appendix to get

$$\mathbb{E}R_1^{2q} \le m^{-(2q-1)} \cdot \int_0^{m^{-1}} \mathbb{E}(a(X^y(s)) - a(y))^{2q}\, ds \le c(K, q) \cdot (1 + y^{2q}) \cdot m^{-3q}. \tag{8}$$

Employ the Burkholder-Davis-Gundy inequality, properties (6) and (7) of b as well as Lemma 8 to obtain

$$\mathbb{E}R_2^{2q} \le c(q) \cdot m^{-(q-1)} \cdot \int_0^{m^{-1}} \mathbb{E}(b(X^y(s)) - b(y))^{2q}\, ds \le c(K, q) \cdot (1 + y^{2q}) \cdot m^{-2q} \tag{9}$$

and

$$\mathbb{E}R_3^{2q} \le c(q) \cdot m^{-(q-1)} \cdot \int_0^{m^{-1}} \mathbb{E}b^{2q}(X^y(s))\, ds \le c(K, q) \cdot (1 + y^{2q}) \cdot m^{-q}. \tag{10}$$

We first treat the case $p = 1$. Clearly,

$$\Lambda_y^{(1)} = |\mathbb{E}R_1| \le \int_0^{m^{-1}} |\mathbb{E}(a(X^y(s)) - a(y))| \, ds.$$

Since $a \in \mathcal{H}(K)$ we have

$$|a(z) - a(y) - a'(y) \cdot (z - y)| \le K \cdot (z - y)^2$$

for every $z \in \mathbb{R}$. Therefore,

$$\begin{aligned} &|\mathbb{E}(a(X^y(s)) - a(y))| \\ &\qquad \le |\mathbb{E}(a(X^y(s)) - a(y) - a'(y) \cdot (X^y(s) - y))| + |a'(y) \cdot (\mathbb{E}X^y(s) - y)| \\ &\qquad \le K \cdot (\mathbb{E}(X^y(s) - y)^2 + |\mathbb{E}X^y(s) - y|) \end{aligned}$$

for every $s \ge 0$. By Lemma 8 and property (6) of a we have

$$\mathbb{E}(X^y(s) - y)^2 \le C(K) \cdot (1 + y^2) \cdot m^{-1}$$

and

$$|\mathbb{E}(X^y(s)) - y| \le \int_0^s \mathbb{E}|a(X^y(u))| \, du \le c(K) \cdot (1 + |y|) \cdot m^{-1} \tag{11}$$

for $s \in [0, m^{-1}]$, which yields

$$\Lambda_y^{(1)} \le c(K) \cdot (1 + y^2) \cdot m^{-2}.$$

Next, we consider the case $p = 2$. Employing the estimates (8) and (10) we obtain

$$\begin{aligned} \Lambda_y^{(2)} &= \left| \mathbb{E}R_1^2 - 2\mathbb{E}(R_1 \cdot R_3) + \int_0^{m^{-1}} \mathbb{E}(b^2(X^y(s)) - b^2(y)) \, ds \right| \\ &\le c(K) \cdot (1 + y^2) \cdot m^{-2} + \int_0^{m^{-1}} |\mathbb{E}(b^2(X^y(s)) - b^2(y))| \, ds. \end{aligned}$$

Since $b \in \mathcal{H}(K)$ we have

$$\left| (b^2)'' \right| \le c(K) \cdot (1 + |b|),$$

and therefore,

$$\begin{aligned} |b^2(z) - b^2(y) - (b^2)'(y) \cdot (z - y)| &\le \sup_{|u - y| \le z} \left| (b^2)''(u) \right| \cdot (z - y)^2 \\ &\le c(K) \cdot (1 + |y| + |z|) \cdot (z - y)^2 \end{aligned}$$

for every $z \in \mathbb{R}$. Using (11) and Lemma 8, we conclude

$$\begin{aligned} &|\mathbb{E}(b^2(X^y(s)) - b^2(y))| \\ &\qquad \le |\mathbb{E}(b^2(X^y(s)) - b^2(y) - (b^2)'(y) \cdot (X^y(s) - y))| + |(b^2)'(y) \cdot (\mathbb{E}X^y(s) - y)| \\ &\qquad \le c(K) \cdot \mathbb{E}((1 + |y| + |X^y(s)|) \cdot (X^y(s) - y)^2) + 2K \cdot (1 + |y|) \cdot |\mathbb{E}(X^y(s)) - y| \\ &\qquad \le c(K) \cdot (1 + |y|^3) \cdot s \end{aligned}$$

for every $s \in [0, m^{-1}]$, which yields

$$\int_0^{m^{-1}} |\mathbb{E}(b^2(X^y(s)) - b^2(y))| \, ds \le c(K) \cdot (1 + |y|^3) \cdot m^{-2}.$$

Hence

$$\Lambda_y^{(2)} \le c(K) \cdot (1 + |y|^3) \cdot m^{-2}.$$

Finally, we consider the case $p = 3$. We have

$$\Lambda_y^{(3)} = |\mathbb{E}(R_1 + R_3)^3|.$$

Use (8) and (10) to derive

$$\begin{aligned} \Lambda_y^{(3)} &\le (\mathbb{E}R_1^4)^{3/4} + 3(\mathbb{E}R_1^4 \cdot \mathbb{E}R_3^2)^{1/2} + 3(\mathbb{E}R_1^2 \cdot \mathbb{E}R_3^4)^{1/2} + |\mathbb{E}R_3^3| \\ &\le c(K) \cdot (1 + |y|^3) \cdot m^{-5/2} + |\mathbb{E}R_3^3|. \end{aligned}$$

Note that $R_3 = R_2 + b(y) \cdot W(m^{-1})$. Hence, by (9),

$$\begin{aligned} |\mathbb{E}R_3^3| &= |\mathbb{E}R_2^3 + 3b(y) \cdot \mathbb{E}(R_2^2 \cdot W(m^{-1})) + 3b^2(y) \cdot \mathbb{E}(R_2 \cdot W^2(m^{-1}))| \\ &\le (\mathbb{E}R_2^4)^{3/4} + 3|b(y)| \cdot (\mathbb{E}R_2^4)^{1/2} \cdot m^{-1/2} + 9b^2(y) \cdot (\mathbb{E}R_2^2)^{1/2} \cdot m^{-1} \\ &\le c(K) \cdot (1 + |y|^3) \cdot m^{-2}, \end{aligned}$$

which implies

$$\Lambda_y^{(3)} \le c(K) \cdot (1 + |y|^3) \cdot m^{-2}$$

and completes the proof of the lemma. □

Consider a homogeneous Markov chain $Y = (Y_\ell)_{\ell \in \mathbb{N}_0}$ with state space $G \cup \{x\}$, initial value x and transition matrix Q, and let $\widetilde{Z}_y$ denote a real-valued random variable with

$$\mathbb{P}(\widetilde{Z}_y = \widetilde{y}) = q_{y,\widetilde{y}}, \quad \widetilde{y} \in G \cup \{x\},$$

for $y \in G \cup \{x\}$, which corresponds to a single step of the Markov chain Y starting from y. We define

$$\Delta_y^{(p)} = |\mathbb{E}(\widetilde{Z}_y - z_y)^p - \mathbb{E}(Z_y - z_y)^p|$$

for $p \in \mathbb{N}$ in order to compare moments of the Markov chain and the Euler scheme.

Lemma 3. *We have*

(i) $\Delta_y^{(p)} \le c(K, \delta, r) \cdot (1 + |y|^{p+(r-1)/\delta}) \cdot m^{-r}$ *for all* $y \in G_1$, $p = 1, 2, 3$ *and* $r \in \mathbb{N}$,

(ii) $\Delta_y^{(1)} = 0$ *and* $\Delta_y^{(p)} \le 2\,\varepsilon^p \cdot m^{-p/2}$ *for all* $y \in G_2$ *and* $p = 2, 3$,

(iii) $\Delta_y^{(p)} = 0$ *for all* $y \in G_3$ *and* $p = 1, 2, 3$.

Proof. Let $y \in G_1$ and $p \in \{1, 2, 3\}$. Then $\mathbb{P}(\widetilde{Z}_y = y) = 1$ and therefore

$$\begin{aligned} \Delta_y^{(p)} &= |(-a(y) \cdot m^{-1})^p - b^2(y) \cdot m^{-1} \cdot 1_{\{2\}}(p)| \\ &\le c(K) \cdot (1 + |y|)^p \cdot m^{-p} + c(K) \cdot (1 + |y|)^2 \cdot m^{-1} \cdot 1_{\{2\}}(p) \\ &\le c(K) \cdot (1 + |y|)^p \cdot m^{-1}. \qquad (12) \end{aligned}$$

We will show that

$$c(K) \cdot (1 + |y|) \ge m^\delta \qquad (13)$$

for every $y \in G_1$. Combining (12) with (13) yields

$$\Delta_y^{(p)} \le c(K, \delta, r) \cdot (1 + |y|)^{p+(r-1)/\delta} \cdot m^{-r}$$

for every $r \in \mathbb{N}$, which implies (i). It remains to derive (13). By definition of J and k_y we have

$$d \cdot (J - k_y) > m^\delta - (|b(y)| + 1) \ge m^\delta - c(K) \cdot (1 + |y|) \qquad (14)$$

for every $y \in G \cup \{x\}$. Clearly, (14) implies (13) for all y with $J - k_y \le 0$. On the other hand, $J - k_y \ge 0$ together with $y \in G_1$ imply

$$d \cdot (J - k_y) \le |\bar{z}_y| \le |y| + |a(y)| \le c(K) \cdot (1 + |y|),$$

and using (14) we obtain (13) again.

Next, assume $y \in G_2$. Consider a real-valued random variable U with

$$\mathbb{P}(U = d \cdot u_y) = 1 - u_y = 1 - \mathbb{P}(U = d \cdot (u_y - 1)). \qquad (15)$$

Then $\widetilde{Z}_y - z_y$ and U are identically distributed, and, consequently,

$$\Delta_y^{(p)} = |\mathbb{E}U^p - b^2(y) \cdot m^{-1} \cdot 1_{\{2\}}(p)|.$$

We have

$$\mathbb{E}(U) = 0, \quad |U| \le d, \qquad (16)$$

and therefore $\Delta_y^{(1)} = 0$ as well as

$$\Delta_y^{(p)} \le d^p + b^2(y) \cdot m^{-1} \cdot 1_{\{2\}}(p) \le \varepsilon^p \cdot m^{-p/2} + \varepsilon^2 \cdot m^{-1} \cdot 1_{\{2\}}(p),$$

which completes the proof of (ii).

Finally, we turn to the case $y \in G_3$. Consider a random vector (U, V), where U satisfies (15) and the distribution of V is specified by $\mathbb{P}(V \in \{-d \cdot k_y, 0, d \cdot k_y\}) = 1$ and

$$\mathbb{P}(V = d \cdot k_y | U = x) = \mathbb{P}(V = -d \cdot k_y | U = x) = \begin{cases} \vartheta_y^{(1)} & \text{if } x = d \cdot u_y, \\ \vartheta_y^{(2)} & \text{if } x = d \cdot (u_y - 1). \end{cases}$$

Then $\widetilde{Z}_y - z_y$ and $U + V$ are identically distributed, and therefore

$$\Delta_y^{(p)} = |\mathbb{E}(U+V)^p - b^2(y) \cdot m^{-1} \cdot 1_{\{2\}}(p)|.$$

Clearly, $\mathbb{E}(V|U) = \mathbb{E}(V^3|U) = 0$, which implies

$$\mathbb{E}V = \mathbb{E}V^3 = \mathbb{E}(U \cdot V) = \mathbb{E}(U^2 \cdot V) = 0. \tag{17}$$

Furthermore, straightforward calculations yield

$$\mathbb{E}U^p = \begin{cases} d^2 \cdot u_y \cdot (1-u_y) & \text{if } p = 2, \\ d^3 \cdot u_y \cdot (1-u_y) \cdot (2u_y - 1) & \text{if } p = 3, \end{cases} \tag{18}$$

as well as

$$\mathbb{E}(V^2) = 2(d \cdot k_y)^2 \cdot (\vartheta_y^{(1)} \cdot (1-u_y) + \vartheta_y^{(2)} \cdot u_y) = b^2(y) \cdot m^{-1} - d^2 \cdot u_y \cdot (1-u_y) \tag{19}$$

and

$$\mathbb{E}(U \cdot V^2) = 2d^3 \cdot k_y^2 \cdot u_y \cdot (1-u_y) \cdot (\vartheta_y^{(1)} - \vartheta_y^{(2)}) = d^3 \cdot u_y \cdot (1-u_y) \cdot (1-2u_y)/3. \tag{20}$$

Use (16) to (20) to conclude $\mathbb{E}(U+V) = 0$ as well as

$$\mathbb{E}(U+V)^2 = \mathbb{E}U^2 + \mathbb{E}V^2 = b^2(y) \cdot m^{-1}$$

and

$$\mathbb{E}(U+V)^3 = \mathbb{E}U^3 + 3\mathbb{E}(U \cdot V^2) = 0,$$

which finishes the proof of (iii). □

We estimate the length of a single step of the Markov chain.

Lemma 4. *With probability one,*

$$\max(|\widetilde{Z}_y - y|, |\widetilde{Z}_y - z_y|) \le c(K) \cdot (1 + |y|) \cdot m^{-\nu},$$

where $\nu = 1$ for $y \in G_1$ and $\nu = 1/2$ otherwise.

Proof. By definition of the transition probabilities we have

$$\max(|\widetilde{Z}_y - y|, |\widetilde{Z}_y - z_y|) = |y - z_y| = |a(y)| \cdot m^{-1}$$

almost surely, if $y \in G_1$, and

$$\begin{aligned} \max(|\widetilde{Z}_y - y|, |\widetilde{Z}_y - z_y|) &\le |\widetilde{Z}_y - z_y| + |y - z_y| \le (k_y + 1) \cdot d + |a(y)| \cdot m^{-1} \\ &\le (|b(y)| + 2 + |a(y)|) \cdot m^{-1/2} \end{aligned}$$

almost surely, if $y \in G_2 \cup G_3$. It remains to apply property (7) of a and b. □

We provide a uniform bound for the moments of the chain Y up to the m-th step.

Lemma 5. *For every $p \in \mathbb{N}$ we have*

$$\max_{\ell=0,\ldots,m} \mathbb{E}Y_\ell^{2p} \le c(K,p) \cdot (1 + x^{2p}).$$

Proof. By definition, $\mathbb{E}Y_0^{2p} = x^{2p}$. Let $\ell \in \{1, \dots, m\}$. We have

$$\mathbb{E}Y_\ell^{2p} = \sum_{y \in G \cup \{x\}} \mathbb{E}(Y_\ell^{2p} | Y_{\ell-1} = y) \cdot \mathbb{P}(Y_{\ell-1} = y) = \sum_{y \in G \cup \{x\}} \mathbb{E}\widetilde{Z}_y^{2p} \cdot \mathbb{P}(Y_{\ell-1} = y). \tag{21}$$

Assume we have shown that

$$\mathbb{E}\widetilde{Z}_y^{2p} \le y^{2p} + c(K,p) \cdot (1 + y^{2p}) \cdot m^{-1}. \tag{22}$$

Then

$$\mathbb{E}Y_\ell^{2p} \le \mathbb{E}Y_{\ell-1}^{2p} \cdot (1 + c(K,p) \cdot m^{-1}) + c(K,p) \cdot m^{-1}$$

follows from (21) and (22), and Gronwall's inequality yields the statement of the lemma.

It remains to prove the bound (22). If $y \in G_1$ then $\mathbb{E}\widetilde{Z}_y^{2p} = y^{2p}$. For $y \in G_2 \cup G_3$ we use the expansion

$$\mathbb{E}\widetilde{Z}_y^{2p} = \sum_{\ell=0}^{2p} A_\ell,$$

where

$$A_\ell = \binom{2p}{\ell} \cdot y^{2p-\ell} \cdot \mathbb{E}(\widetilde{Z}_y - y)^\ell.$$

Clearly, $A_0 = y^{2p}$. Moreover, $\mathbb{E}(\widetilde{Z}_y - y) = a(y) \cdot m^{-1}$ follows from Lemma 3, and therefore

$$|A_1| \le c(K,p) \cdot (1 + y^{2p}) \cdot m^{-1}.$$

By Lemma 4 we have

$$\mathbb{E}|\widetilde{Z}_y - y|^\ell \le c(K,\ell) \cdot (1 + |y|^\ell) \cdot m^{-\ell/2}$$

for every $\ell \in \mathbb{N}$. Hence

$$|A_\ell| \le c(K,p) \cdot (1 + y^{2p}) \cdot m^{-1}$$

for $\ell = 2, \dots, 2p$, which completes the proof of (22). □

Fix $\beta > 0$ in the sequel. Put

$$\mathcal{F}_M(\beta) = \{f \in C^4(\mathbb{R})\colon\ |f^{(\ell)}(u)| \le M \cdot (1 + |u|^\beta), u \in \mathbb{R}, \ell = 1, \dots, 4\}$$

as well as

$$\mathcal{F}_\infty(\beta) = \bigcup_{M>0} \mathcal{F}_M(\beta),$$

and define a semigroup of linear operators

$$P_t : \mathcal{F}_\infty(\beta) \to \mathcal{F}_\infty(\beta), \quad t \in [0, \infty),$$

by

$$P_t f(y) = \mathbb{E} f(X^y(t)), \quad y \in \mathbb{R},$$

see Lemma 11 in the Appendix. Thus $P_t P_s = P_{t+s}$ and

$$\int_{\mathbb{R}} f \, dS(x,a,b) = P_1 f(x) \tag{23}$$

for $f \in \mathcal{F}_\infty(\beta)$. In the sequel, we use

$$\overline{f} = f_{|G\cup\{x\}} \in \mathbb{R}^{G\cup\{x\}}$$

to denote the restriction of a function $f : \mathbb{R} \to \mathbb{R}$ to the state space $G \cup \{x\}$ of the Markov chain Y. Clearly,

$$(Q^{\ell_1+\ell_2}\overline{f})_x = \mathbb{E}(Q^{\ell_1}\overline{f})_{Y_{\ell_2}} \tag{24}$$

for all $\ell_1, \ell_2 \in \mathbb{N}_0$, and in particular,

$$\int_{\mathbb{R}} f \, d\widehat{S}_{\delta,\varepsilon,m}(x,a,b) = (Q^m \overline{f})_x. \tag{25}$$

Moreover, we have

$$(Q\overline{f})_y = \mathbb{E} f(\widetilde{Z}_y) \tag{26}$$

for every $y \in G \cup \{x\}$, and hereby we approximate $P_{m^{-1}} f$ on $G \cup \{x\}$.

Lemma 6. *Let $M > 0$. For every $f \in \mathcal{F}_M(\beta)$ and all $y \in G \cup \{x\}$ we have*

$$|P_{m^{-1}} f(y) - (Q\overline{f})_y| \le c(K,M,\beta) \cdot (1 + |y|^{4+\beta}) \cdot m^{-2} \cdot A(K,\delta,m,\varepsilon,y),$$

where

$$A(K,\delta,m,\varepsilon,y) = \begin{cases} c(K,\delta)\cdot(1+|y|^{1/\delta}) & \text{if } y \in G_1, \\ (1+\varepsilon^2 \cdot m) & \text{if } y \in G_2, \\ 1 & \text{if } y \in G_3. \end{cases}$$

Proof. By definition of P_t and (26),

$$P_{m^{-1}} f(y) - (Q\overline{f})_y = \int_{\mathbb{R}} f(z) \cdot (d\mathbb{P}_{X^y(m^{-1})}(z) - d\mathbb{P}_{\widetilde{Z}_y}(z)).$$

Since $f \in C^4(\mathbb{R})$ we have

$$f(z) = \sum_{j=0}^{3} f^{(p)}(z_y) \cdot \frac{(z-z_y)^p}{p!} + f^{(4)}(\theta_z) \cdot \frac{(z-z_y)^4}{4!}$$

with

$$|\theta_z| \le |z_y| + |z - z_y| \tag{27}$$

for every $z \in \mathbb{R}$. Hence

$$|P_{m^{-1}} f(y) - (Q\overline{f})_y| \le \kappa_1 + \kappa_2 + \kappa_3 + \rho,$$

where

$$\kappa_p = 1/p! \cdot |f^{(p)}(z_y)| \cdot |\mathbb{E}(X^y(m^{-1}) - z_y)^p - \mathbb{E}(\widetilde{Z}_y - z_y)^p|$$

for $p = 1, 2, 3$, and

$$\rho = \frac{1}{4!} \cdot \Big| \int_{\mathbb{R}} f^{(4)}(\theta_z) \cdot (z - z_y)^4 \, (d\mathbb{P}_{X^y(m^{-1})}(z) - d\mathbb{P}_{\widetilde{Z}_y}(z)) \Big|.$$

Let $p \in \{1, 2, 3\}$. Since $f \in \mathcal{F}_M(\beta)$ and $a \in \mathcal{H}(K)$ we have

$$|f^{(p)}(z_y)| \leq M \cdot (1 + |z_y|^\beta) \leq c(K, M, \beta) \cdot (1 + |y|^\beta).$$

Furthermore, using Lemma 2 and Lemma 3 with $r = 2$, we get

$$\begin{aligned} |\mathbb{E}(X^y(m^{-1}) - z_y)^p - \mathbb{E}(\widetilde{Z}_y - z_y)^p| &\leq \Lambda_y^{(p)} + \Delta_y^{(p)} \\ &\leq c(K) \cdot (1 + |y|^3) \cdot m^{-2} \cdot A(K, \delta, m, \varepsilon, y). \end{aligned}$$

Hence

$$\sum_{p=1}^{3} \kappa_p \leq c(K, M, \beta) \cdot (1 + |y|^{3+\beta}) \cdot m^{-2} \cdot A(K, \delta, m, \varepsilon, y).$$

Next, we estimate ρ. Use (27) to obtain

$$|f^{(4)}(\theta_z)| \leq M \cdot (1 + |\theta_z|^\beta) \leq c(M, \beta) \cdot (1 + |z_y|^\beta + |z - z_y|^\beta)$$

for every $z \in \mathbb{R}$. Hence,

$$\begin{aligned} \rho \leq\ & c(M, \beta) \cdot (1 + |z_y|^\beta) \cdot (\mathbb{E}(X^y(m^{-1}) - z_y)^4 + \mathbb{E}(\widetilde{Z}_y - z_y)^4) \\ & + c(M, \beta) \cdot (\mathbb{E}|X^y(m^{-1}) - z_y)|^{4+\beta} + \mathbb{E}|\widetilde{Z}_y - z_y|^{4+\beta}). \end{aligned}$$

Employing Lemma 4 as well as (29) and Lemma 5 in the Appendix we conclude

$$\begin{aligned} \rho &\leq c(K, M, \beta) \cdot (1 + |y|^\beta) \cdot (1 + y^4) \cdot m^{-2} + c(K, M, \beta) \cdot (1 + |y|^{4+\beta}) \cdot m^{-(2+\beta/2)} \\ &\leq c(K, M, \beta) \cdot (1 + |y|^{4+\beta}) \cdot m^{-2}, \end{aligned}$$

which finishes the proof of the lemma. □

Finally, we estimate the error of the quadrature rule provided by $\widehat{S}_{\delta,\varepsilon,m}(x, a, b)$ on the class $\mathcal{F}_M(\beta)$.

Lemma 7. *Let $M > 0$. For every $f \in \mathcal{F}_M(\beta)$ we have*

$$|P_1 f(x) - (Q^m \overline{f})_x| \leq c(L, K, M, \beta, \delta) \cdot m^{-1} \cdot (1 + \varepsilon^2 \cdot m \cdot \min(1, \#G_2)).$$

Proof. Let $f \in \mathcal{F}_M(\beta)$ and put

$$g_\ell = P_{(\ell-1)\cdot m^{-1}} f$$

for $\ell = 1, \ldots, m$. Note that

$$g_\ell \in \mathcal{F}_{\widetilde{M}}(\beta)$$

with $\widetilde{M} = c(M, K, \beta)$, due to Lemma 11 in the Appendix. Clearly,

$$P_1 f(x) - (Q^m \overline{f})_x = \sum_{\ell=1}^{m} E_\ell,$$

where

$$E_\ell = (Q^{m-\ell} \overline{g}_{\ell+1})_x - (Q^{m-\ell+1} \overline{g}_\ell)_x.$$

Let $\ell \in \{1, \dots, m\}$. By (26) we get

$$E_\ell = \mathbb{E}(g_{\ell+1}(Y_{m-\ell})) - \mathbb{E}((Q\overline{g}_\ell)_{Y_{m-\ell}}) = \sum_{y \in G \cup \{x\}} (P_{m^{-1}} g_\ell(y) - (Q\overline{g}_\ell)_y) \cdot \mathbb{P}(Y_{m-\ell} = y),$$

and, employing Lemma 6 as well as Lemma 5, we conclude that

$$\begin{aligned} |E_\ell| &\leq \sum_{y \in G \cup \{x\}} |P_{m^{-1}} g_\ell(y) - (Q\overline{g}_\ell)_y| \cdot \mathbb{P}(Y_{m-\ell} = y) \\ &\leq c(K, \widetilde{M}, \beta, \delta) \sum_{y \in G \cup \{x\}} (1 + |y|^{4+\beta+1/\delta}) \cdot m^{-2} \cdot (1 + \varepsilon^2 \cdot m \cdot 1_{G_2}(y)) \cdot \mathbb{P}(Y_{m-\ell} = y) \\ &\leq c(K, M, \beta, \delta) \cdot (1 + \mathbb{E}|Y_{m-\ell}|^{4+\beta+1/\delta}) \cdot m^{-2} \cdot (1 + \varepsilon^2 \cdot m \cdot \min(1, \#G_2)) \\ &\leq c(L, K, M, \beta, \delta) \cdot m^{-2} \cdot (1 + \varepsilon^2 \cdot m \cdot \min(1, \#G_2)), \end{aligned}$$

which implies the statement of the lemma. □

Observe (23) as well as (25) and apply Lemma 7 with $M = 1$ and $\varepsilon = m^{-1/2}$ to obtain part (i) of Theorem 1. Part (ii) follows from Lemma 7 with $M = 1$ and the fact that $G_2 = \emptyset$ if $b \in \mathcal{H}_\varepsilon(K)$.

Appendix

Let $r \in \mathbb{N}$ and $K, M, \beta > 0$, and put

$$\begin{aligned} \mathcal{H}^r(K) &= \{h \in C^r(\mathbb{R}) : |h(0)|, ||h^{(\ell)}||_\infty \leq K, \ell = 1, \dots, r\}, \\ \mathcal{F}^r_M(\beta) &= \{f \in C^r(\mathbb{R}) : |f^{(\ell)}(u)| \leq M \cdot (1 + |u|^\beta), u \in \mathbb{R}, \ell = 1, \dots, r\}. \end{aligned}$$

In this section we consider equation (1) with fixed coefficients

$$a, b \in \mathcal{H}^r(K), \tag{28}$$

and we collect some facts on the dependence of its solution X^x on the initial value x.

Lemma 8. *For every $t \in [0, 1]$, $x \in \mathbb{R}$, and $p \in \mathbb{N}$ we have*

$$\mathbb{E} \sup_{0 \leq s \leq t} (X^x(s) - x)^{2p} \leq c(K, p) \cdot (1 + x^{2p}) \cdot t^p.$$

See, e.g., [3, Chap. 5, Thm. 2.3]. Clearly, Lemma 8 together with (28) implies

$$\mathbb{E}(X^x(t) - x - a(x) \cdot t)^{2p} \le c(K,p) \cdot (1 + x^{2p}) \cdot t^p \tag{29}$$

for every $x \in \mathbb{R}$, $t \in [0,1]$, and $p \in \mathbb{N}$.

Assumption (28) also assures that the random field $(X^x(t))_{x\in\mathbb{R},t\in[0,1]}$ is r-times differentiable with respect to the parameter x in the p-th mean sense.

Lemma 9. *There exist processes*

$$\frac{\partial^\ell}{\partial x^\ell} X^x = \left(\frac{\partial^\ell}{\partial x^\ell} X^x(t) \right)_{t\in[0,1]}, \qquad x \in \mathbb{R}, \ell = 1, \ldots, r,$$

such that

$$\mathbb{E}\left| \frac{\partial^\ell}{\partial x^\ell} X^x(t) \right|^p \le c(K,p,r)$$

and

$$\lim_{h\to 0} \mathbb{E}\left| \frac{1}{h}\left(\frac{\partial^{\ell-1}}{\partial x^{\ell-1}} X^{x+h}(t) - \frac{\partial^{\ell-1}}{\partial x^{\ell-1}} X^x(t) \right) - \frac{\partial^\ell}{\partial x^\ell} X^x(t) \right|^p = 0$$

for every $x \in \mathbb{R}$, $t \in [0,1]$, $p > 0$, and $\ell = 1, \ldots, r$.

See, e.g., [4, Sec. 8, Thm. 1]. Furthermore, adapting the methods from the latter reference it is straightforward to show that

$$\mathbb{E}\left| \frac{\partial^\ell}{\partial x^\ell} X^x(t) - \frac{\partial^\ell}{\partial x^\ell} X^y(t) \right|^p \le c(K,p,r) \cdot |x-y|^p \tag{30}$$

for all $x, y \in \mathbb{R}$ with $|x - y| \le 1$ and $\ell = 0, \ldots, r-1$.

We turn to a Riordan's type formula for the p-th mean derivative of a function applied to the process X^x at time t. Put

$$S_\ell = \left\{ (j_1, \ldots, j_\ell) \in \mathbb{N}^\ell : \sum_{k=1}^{\ell} j_k = r \right\}$$

and define processes A_ℓ^x by

$$A_\ell^x(t) = \sum_{j\in S_\ell} \binom{r}{j_1, \ldots, j_\ell} \prod_{k=1}^{\ell} \frac{\partial^{j_k}}{\partial x^{j_k}} X^x(t)$$

for $\ell \in \{1, \ldots, r\}$.

Lemma 10. *Let $f \in \mathcal{F}_M^r(\beta)$. For every $p > 0$ the random field*

$$\eta^x(t) = f(X^x(t)), \quad t \in [0,1], x \in \mathbb{R},$$

is r-times differentiable w.r.t. x in the p-th mean with r-th derivative

$$\frac{\partial^r}{\partial x^r} \eta^x(t) = \sum_{\ell=1}^{r} \frac{f^{(\ell)}(X^x(t))}{\ell!} \cdot A_\ell^x(t).$$

See [4, Sec. 8, Cor. 1] for a proof of Lemma 10 in the case $r = 1$. The general case follows by induction on r, employing Lemma 9 and (30).

Using Lemmas 8 to 10 we immediately obtain the following result.

Lemma 11. *Consider the functions*

$$g_t(x) = \mathbb{E}(f(X^x(t))), \quad x \in \mathbb{R},$$

for $t \in [0, 1]$. *Then*

$$g_t \in \mathcal{F}^r_{c(K,M,r,\beta)}(\beta)$$

with

$$g_t^{(\ell)}(x) = \mathbb{E}\Big(\frac{\partial^\ell}{\partial x^\ell} \eta^x(t)\Big)$$

for $\ell = 1, \ldots, r$.

Acknowledgement

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) within the Priority Programme 1324. We are grateful to Steffen Dereich and Mike Giles for a number of stimulating discussions.

References

[1] Dereich, S. (2009), Asymptotic formulae for coding problems and intermediate optimization problems: a review, in: Trends in Stochastic Analysis, J. Blath, P. Moerters, M. Scheutzow, eds., pp. 187–232, Cambridge Univ. Press, Cambridge.

[2] Dereich, S., Scheutzow, M., Schottstedt, R. (2010), Constructive quantization: approximation by empirical measures, in progress.

[3] Friedman, A. (1975), *Stochastic Differential Equations and Applications, Volume 1*, Academic Press, New York.

[4] Gihman, I. I., Skorohod, A. V. (1972) *Stochastic Differential Equations* Springer, Berlin.

[5] Giles, M. (2008), Multilevel Monte Carlo path simulation, Oper. Res. **56**, 607–617.

[6] Graf, S., Luschgy, H. (2000), *Foundations of Quantization for Probability Distributions*, Lect. Notes in Math. **1730**, Springer-Verlag, Berlin.

[7] Kloeden, P. E., Platen, E. (1995), *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin.

[8] Kusuoka, S. (2001), Approximation of expectation of diffusion process and mathematical finance, Adv. Stud. Pure Math. **31**, 147–165.

[9] Kusuoka, S. (2004), Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus, Adv. Math. Econ. **6**, 69–83.

[10] Litterer, C., Lyons, T. (2010), High order recombination and an application to cubature on the Wiener space, Preprint. http://arxiv.org/pdf/1008.4942v1.

[11] Lyons, T., Victoir, N. (2004), Cubature on Wiener space, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **460**, 169–198.

[12] Müller-Gronbach, T., Ritter, K. (2011), A local refinement strategy for constructive quantization of scalar SDEs, Preprint **72**, DFG Priority Programme 1324.

[13] Pagès, G., Printems, J. (2008), Optimal quantization for finance: from random vectors to stochastic processes, in: Mathematical Modelling and Numerical Methods in Finance, Handbook of Numerical Analysis, Vol. XV, A. Bensoussan, Q. Zhang, eds., pp. 595–648, North-Holland, Amsterdam.

[14] Petras, K., Ritter, K. (2006), On the complexity of parabolic initial value problems with variable drift, J. Complexity **22**, 118–145.

[15] Wasilkowski, G. W., Woźniakowski, H. (2000), Complexity of weighted approximation over $\mathbb{R}$, J. Approx. Theory **103**, 223–251.

[16] Wasilkowski, G. W., Woźniakowski, H. (2001), Complexity of weighted approximation over $\mathbb{R}^d$, J. Complexity **17**, 722–740.

Fakultät für Informatik und Mathematik, Universität Passau, Innstr. 33, 94030 Passau, Germany

E-mail address: thomas.mueller-gronbach@uni-passau.de

Fachbereich Mathematik, Technische Universität Kaiserslautern, Postfach 3049, 67653 Kaiserslautern, Germany

E-mail address: ritter@mathematik.uni-kl.de

Fakultät für Informatik und Mathematik, Universität Passau, Innstr. 33, 94030 Passau, Germany

E-mail address: larisa.yaroslavtseva@uni-passau.de

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports

[1] R. Ramlau, G. Teschke, and M. Zhariy. A Compressive Landweber Iteration for Solving Ill-Posed Inverse Problems. Preprint 1, DFG-SPP 1324, September 2008.

[2] G. Plonka. The Easy Path Wavelet Transform: A New Adaptive Wavelet Transform for Sparse Representation of Two-dimensional Data. Preprint 2, DFG-SPP 1324, September 2008.

[3] E. Novak and H. Woźniakowski. Optimal Order of Convergence and (In-) Tractability of Multivariate Approximation of Smooth Functions. Preprint 3, DFG-SPP 1324, October 2008.

[4] M. Espig, L. Grasedyck, and W. Hackbusch. Black Box Low Tensor Rank Approximation Using Fibre-Crosses. Preprint 4, DFG-SPP 1324, October 2008.

[5] T. Bonesky, S. Dahlke, P. Maass, and T. Raasch. Adaptive Wavelet Methods and Sparsity Reconstruction for Inverse Heat Conduction Problems. Preprint 5, DFG-SPP 1324, January 2009.

[6] E. Novak and H. Woźniakowski. Approximation of Infinitely Differentiable Multivariate Functions Is Intractable. Preprint 6, DFG-SPP 1324, January 2009.

[7] J. Ma and G. Plonka. A Review of Curvelets and Recent Applications. Preprint 7, DFG-SPP 1324, February 2009.

[8] L. Denis, D. A. Lorenz, and D. Trede. Greedy Solution of Ill-Posed Problems: Error Bounds and Exact Inversion. Preprint 8, DFG-SPP 1324, April 2009.

[9] U. Friedrich. A Two Parameter Generalization of Lions' Nonoverlapping Domain Decomposition Method for Linear Elliptic PDEs. Preprint 9, DFG-SPP 1324, April 2009.

[10] K. Bredies and D. A. Lorenz. Minimization of Non-smooth, Non-convex Functionals by Iterative Thresholding. Preprint 10, DFG-SPP 1324, April 2009.

[11] K. Bredies and D. A. Lorenz. Regularization with Non-convex Separable Constraints. Preprint 11, DFG-SPP 1324, April 2009.

[12] M. Döhler, S. Kunis, and D. Potts. Nonequispaced Hyperbolic Cross Fast Fourier Transform. Preprint 12, DFG-SPP 1324, April 2009.

[13] C. Bender. Dual Pricing of Multi-Exercise Options under Volume Constraints. Preprint 13, DFG-SPP 1324, April 2009.

[14] T. Müller-Gronbach and K. Ritter. Variable Subspace Sampling and Multi-level Algorithms. Preprint 14, DFG-SPP 1324, May 2009.

[15] G. Plonka, S. Tenorth, and A. Iske. Optimally Sparse Image Representation by the Easy Path Wavelet Transform. Preprint 15, DFG-SPP 1324, May 2009.

[16] S. Dahlke, E. Novak, and W. Sickel. Optimal Approximation of Elliptic Problems by Linear and Nonlinear Mappings IV: Errors in L_2 and Other Norms. Preprint 16, DFG-SPP 1324, June 2009.

[17] B. Jin, T. Khan, P. Maass, and M. Pidcock. Function Spaces and Optimal Currents in Impedance Tomography. Preprint 17, DFG-SPP 1324, June 2009.

[18] G. Plonka and J. Ma. Curvelet-Wavelet Regularized Split Bregman Iteration for Compressed Sensing. Preprint 18, DFG-SPP 1324, June 2009.

[19] G. Teschke and C. Borries. Accelerated Projected Steepest Descent Method for Nonlinear Inverse Problems with Sparsity Constraints. Preprint 19, DFG-SPP 1324, July 2009.

[20] L. Grasedyck. Hierarchical Singular Value Decomposition of Tensors. Preprint 20, DFG-SPP 1324, July 2009.

[21] D. Rudolf. Error Bounds for Computing the Expectation by Markov Chain Monte Carlo. Preprint 21, DFG-SPP 1324, July 2009.

[22] M. Hansen and W. Sickel. Best m-term Approximation and Lizorkin-Triebel Spaces. Preprint 22, DFG-SPP 1324, August 2009.

[23] F.J. Hickernell, T. Müller-Gronbach, B. Niu, and K. Ritter. Multi-level Monte Carlo Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 23, DFG-SPP 1324, August 2009.

[24] S. Dereich and F. Heidenreich. A Multilevel Monte Carlo Algorithm for Lévy Driven Stochastic Differential Equations. Preprint 24, DFG-SPP 1324, August 2009.

[25] S. Dahlke, M. Fornasier, and T. Raasch. Multilevel Preconditioning for Adaptive Sparse Optimization. Preprint 25, DFG-SPP 1324, August 2009.

[26] S. Dereich. Multilevel Monte Carlo Algorithms for Lévy-driven SDEs with Gaussian Correction. Preprint 26, DFG-SPP 1324, August 2009.

[27] G. Plonka, S. Tenorth, and D. Roşca. A New Hybrid Method for Image Approximation using the Easy Path Wavelet Transform. Preprint 27, DFG-SPP 1324, October 2009.

[28] O. Koch and C. Lubich. Dynamical Low-rank Approximation of Tensors. Preprint 28, DFG-SPP 1324, November 2009.

[29] E. Faou, V. Gradinaru, and C. Lubich. Computing Semi-classical Quantum Dynamics with Hagedorn Wavepackets. Preprint 29, DFG-SPP 1324, November 2009.

[30] D. Conte and C. Lubich. An Error Analysis of the Multi-configuration Time-dependent Hartree Method of Quantum Dynamics. Preprint 30, DFG-SPP 1324, November 2009.

[31] C. E. Powell and E. Ullmann. Preconditioning Stochastic Galerkin Saddle Point Problems. Preprint 31, DFG-SPP 1324, November 2009.

[32] O. G. Ernst and E. Ullmann. Stochastic Galerkin Matrices. Preprint 32, DFG-SPP 1324, November 2009.

[33] F. Lindner and R. L. Schilling. Weak Order for the Discretization of the Stochastic Heat Equation Driven by Impulsive Noise. Preprint 33, DFG-SPP 1324, November 2009.

[34] L. Kämmerer and S. Kunis. On the Stability of the Hyperbolic Cross Discrete Fourier Transform. Preprint 34, DFG-SPP 1324, December 2009.

[35] P. Cerejeiras, M. Ferreira, U. Kähler, and G. Teschke. Inversion of the noisy Radon transform on $SO(3)$ by Gabor frames and sparse recovery principles. Preprint 35, DFG-SPP 1324, January 2010.

[36] T. Jahnke and T. Udrescu. Solving Chemical Master Equations by Adaptive Wavelet Compression. Preprint 36, DFG-SPP 1324, January 2010.

[37] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Irregular Shearlet Frames: Geometry and Approximation Properties. Preprint 37, DFG-SPP 1324, February 2010.

[38] G. Kutyniok and W.-Q Lim. Compactly Supported Shearlets are Optimally Sparse. Preprint 38, DFG-SPP 1324, February 2010.

[39] M. Hansen and W. Sickel. Best m-Term Approximation and Tensor Products of Sobolev and Besov Spaces – the Case of Non-compact Embeddings. Preprint 39, DFG-SPP 1324, March 2010.

[40] B. Niu, F.J. Hickernell, T. Müller-Gronbach, and K. Ritter. Deterministic Multi-level Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 40, DFG-SPP 1324, March 2010.

[41] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Construction of Compactly Supported Shearlet Frames. Preprint 41, DFG-SPP 1324, March 2010.

[42] C. Bender and J. Steiner. Error Criteria for Numerical Solutions of Backward SDEs. Preprint 42, DFG-SPP 1324, April 2010.

[43] L. Grasedyck. Polynomial Approximation in Hierarchical Tucker Format by Vector-Tensorization. Preprint 43, DFG-SPP 1324, April 2010.

[44] M. Hansen und W. Sickel. Best m-Term Approximation and Sobolev-Besov Spaces of Dominating Mixed Smoothness - the Case of Compact Embeddings. Preprint 44, DFG-SPP 1324, April 2010.

[45] P. Binev, W. Dahmen, and P. Lamby. Fast High-Dimensional Approximation with Sparse Occupancy Trees. Preprint 45, DFG-SPP 1324, May 2010.

[46] J. Ballani and L. Grasedyck. A Projection Method to Solve Linear Systems in Tensor Format. Preprint 46, DFG-SPP 1324, May 2010.

[47] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence Rates for Greedy Algorithms in Reduced Basis Methods. Preprint 47, DFG-SPP 1324, May 2010.

[48] S. Kestler and K. Urban. Adaptive Wavelet Methods on Unbounded Domains. Preprint 48, DFG-SPP 1324, June 2010.

[49] H. Yserentant. The Mixed Regularity of Electronic Wave Functions Multiplied by Explicit Correlation Factors. Preprint 49, DFG-SPP 1324, June 2010.

[50] H. Yserentant. On the Complexity of the Electronic Schrödinger Equation. Preprint 50, DFG-SPP 1324, June 2010.

[51] M. Guillemard and A. Iske. Curvature Analysis of Frequency Modulated Manifolds in Dimensionality Reduction. Preprint 51, DFG-SPP 1324, June 2010.

[52] E. Herrholz and G. Teschke. Compressive Sensing Principles and Iterative Sparse Recovery for Inverse and Ill-Posed Problems. Preprint 52, DFG-SPP 1324, July 2010.

[53] L. Kämmerer, S. Kunis, and D. Potts. Interpolation Lattices for Hyperbolic Cross Trigonometric Polynomials. Preprint 53, DFG-SPP 1324, July 2010.

[54] G. Kutyniok and W.-Q Lim. Shearlets on Bounded Domains. Preprint 54, DFG-SPP 1324, July 2010.

[55] A. Zeiser. Wavelet Approximation in Weighted Sobolev Spaces of Mixed Order with Applications to the Electronic Schrödinger Equation. Preprint 55, DFG-SPP 1324, July 2010.

[56] G. Kutyniok, J. Lemvig, and W.-Q Lim. Compactly Supported Shearlets. Preprint 56, DFG-SPP 1324, July 2010.

[57] A. Zeiser. On the Optimality of the Inexact Inverse Iteration Coupled with Adaptive Finite Element Methods. Preprint 57, DFG-SPP 1324, July 2010.

[58] S. Jokar. Sparse Recovery and Kronecker Products. Preprint 58, DFG-SPP 1324, August 2010.

[59] T. Aboiyar, E. H. Georgoulis, and A. Iske. Adaptive ADER Methods Using Kernel-Based Polyharmonic Spline WENO Reconstruction. Preprint 59, DFG-SPP 1324, August 2010.

[60] O. G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On the Convergence of Generalized Polynomial Chaos Expansions. Preprint 60, DFG-SPP 1324, August 2010.

[61] S. Holtz, T. Rohwedder, and R. Schneider. On Manifolds of Tensors of Fixed TT-Rank. Preprint 61, DFG-SPP 1324, September 2010.

[62] J. Ballani, L. Grasedyck, and M. Kluge. Black Box Approximation of Tensors in Hierarchical Tucker Format. Preprint 62, DFG-SPP 1324, October 2010.

[63] M. Hansen. On Tensor Products of Quasi-Banach Spaces. Preprint 63, DFG-SPP 1324, October 2010.

[64] S. Dahlke, G. Steidl, and G. Teschke. Shearlet Coorbit Spaces: Compactly Supported Analyzing Shearlets, Traces and Embeddings. Preprint 64, DFG-SPP 1324, October 2010.

[65] W. Hackbusch. Tensorisation of Vectors and their Efficient Convolution. Preprint 65, DFG-SPP 1324, November 2010.

[66] P. A. Cioica, S. Dahlke, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Spatial Besov Regularity for Stochastic Partial Differential Equations on Lipschitz Domains. Preprint 66, DFG-SPP 1324, November 2010.

[67] E. Novak and H. Woźniakowski. On the Power of Function Values for the Approximation Problem in Various Settings. Preprint 67, DFG-SPP 1324, November 2010.

[68] A. Hinrichs, E. Novak, and H. Woźniakowski. The Curse of Dimensionality for Monotone and Convex Functions of Many Variables. Preprint 68, DFG-SPP 1324, November 2010.

[69] G. Kutyniok and W.-Q Lim. Image Separation Using Shearlets. Preprint 69, DFG-SPP 1324, November 2010.

[70] B. Jin and P. Maass. An Analysis of Electrical Impedance Tomography with Applications to Tikhonov Regularization. Preprint 70, DFG-SPP 1324, December 2010.

[71] S. Holtz, T. Rohwedder, and R. Schneider. The Alternating Linear Scheme for Tensor Optimisation in the TT Format. Preprint 71, DFG-SPP 1324, December 2010.

[72] T. Müller-Gronbach and K. Ritter. A Local Refinement Strategy for Constructive Quantization of Scalar SDEs. Preprint 72, DFG-SPP 1324, December 2010.

[73] T. Rohwedder and R. Schneider. An Analysis for the DIIS Acceleration Method used in Quantum Chemistry Calculations. Preprint 73, DFG-SPP 1324, December 2010.

[74] C. Bender and J. Steiner. Least-Squares Monte Carlo for Backward SDEs. Preprint 74, DFG-SPP 1324, December 2010.

[75] C. Bender. Primal and Dual Pricing of Multiple Exercise Options in Continuous Time. Preprint 75, DFG-SPP 1324, December 2010.

[76] H. Harbrecht, M. Peters, and R. Schneider. On the Low-rank Approximation by the Pivoted Cholesky Decomposition. Preprint 76, DFG-SPP 1324, December 2010.

[77] P. A. Cioica, S. Dahlke, N. Döhring, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Adaptive Wavelet Methods for Elliptic Stochastic Partial Differential Equations. Preprint 77, DFG-SPP 1324, January 2011.

[78] G. Plonka, S. Tenorth, and A. Iske. Optimal Representation of Piecewise Hölder Smooth Bivariate Functions by the Easy Path Wavelet Transform. Preprint 78, DFG-SPP 1324, January 2011.

[79] A. Mugler and H.-J. Starkloff. On Elliptic Partial Differential Equations with Random Coefficients. Preprint 79, DFG-SPP 1324, January 2011.

[80] T. Müller-Gronbach, K. Ritter, and L. Yaroslavtseva. A Derandomization of the Euler Scheme for Scalar Stochastic Differential Equations. Preprint 80, DFG-SPP 1324, January 2011.