DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen”

Multi-Level Monte Carlo Approximation of Distribution Functions and Densities

M. Giles, T. Nagapetyan, K. Ritter

Preprint 157

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”

Multi-Level Monte Carlo Approximation of Distribution Functions and Densities

M. Giles, T. Nagapetyan, K. Ritter

Preprint 157

MULTI-LEVEL MONTE CARLO APPROXIMATION OF DISTRIBUTION FUNCTIONS AND DENSITIES

MIKE GILES, TIGRAN NAGAPETYAN, AND KLAUS RITTER

Abstract. We construct and analyze multi-level Monte Carlo methods for the approximation of distribution functions and densities of univariate random variables. Since, by assumption, the target distribution is not known explicitly, approximations have to be used. We provide a general analysis under suitable assumptions on the weak and strong convergence. We apply the results to smooth path-independent and path-dependent functionals and to stopped exit times of SDEs.

1. Introduction

Let Y denote a real-valued random variable with distribution function F and density ρ. We study the approximation of F and ρ with respect to the supremum norm on a compact interval $[S_0, S_1]$, without assuming that the distribution of Y is explicitly known or that the simulation of Y is feasible. Instead, we suppose that a sequence of random variables $Y^{(\ell)}$ is at hand that converge to Y in a suitable way and that are suited to simulation.

We present a general approach, which is later on applied in the context of stochastic differential equations (SDEs). In this specific setting we aim at the distribution of Lipschitz continuous, path-independent or path-dependent functionals of the solution process, or the distribution of stopped exit times from bounded domains.

In the general setting a naive Monte Carlo algorithm for the approximation of ρ works as follows: Choose a level $\ell \in \mathbb{N}$ and a replication number $n \in \mathbb{N}$, generate n independent samples according to $Y^{(\ell)}$, and apply a kernel density estimator, say, to these samples. For the approximation of F one proceeds analogously, and here the empirical distribution function of the samples is the most elementary choice.

In this paper we develop the multi-level Monte Carlo approach, which relies on the coupled simulation of $Y^{(\ell)}$ and $Y^{(\ell-1)}$ on a finite range of levels ℓ. For the multi-level approach to work well for the approximation of distribution functions or densities, a smoothing step is necessary on every level. The smoothing is based on rescaled translates of a suitable function g, which is meant to approximate either the indicator function of $]-\infty, 0]$ or the Dirac functional at zero. In a first stage the multi-level algorithm provides an approximation to F or ρ at discrete points, which is then extended to a function on $[S_0, S_1]$ in a standard and purely deterministic way.

For the approximation of F and ρ on $[S_0, S_1]$ our assumptions are as follows:

(i) The density ρ of Y is r-times continuously differentiable,

(ii) The simulation of the joint distribution of $Y^{(\ell)}$ and $Y^{(\ell-1)}$ is possible at cost $O(M^\ell)$ for every $\ell \in \mathbb{N}$, where $M > 1$.

(iii) A weak error estimate

$$\sup_{s \in [S_0, S_1]} \left| \mathrm{E} \left(g((Y - s)/\delta) - g((Y^{(\ell)} - s)/\delta) \right) \right| \leq O\left(\min\left(\delta^{-\alpha_1} \cdot M^{-\ell \cdot \alpha_2}, M^{-\ell \cdot \alpha_3} \right) \right)$$

Date: February 2, 2014.

holds for all positive, sufficiently small δ and all $\ell \in \mathbb{N}_0$, where $\alpha_1 \geq 0$, $\alpha_2 > 0$, and $\alpha_2 \geq \alpha_3 \geq 0$.

(iv) A strong error estimate

$$\mathrm{E}\min((Y - Y^{(\ell)})^2/\delta^2, 1) \leq O\left(\delta^{-\beta_4} \cdot M^{-\ell\cdot\beta_5}\right)$$

holds for all positive, sufficiently small δ and all $\ell \in \mathbb{N}_0$, where $\beta_4 \geq 0$ and $\beta_5 > 0$.

We also study the approximation of the distribution function F at a single point $s \in [S_0, S_1]$, and here (iv) is replaced by the following assumption:

(v) A strong error estimate

$$\sup_{s\in[S_0,S_1]} \mathrm{E}\left(g((Y-s)/\delta) - g((Y^{(\ell)}-s)/\delta)\right)^2 \leq O\left(\min\left(\delta^{-\beta_1}\cdot M^{-\ell\cdot\beta_2}, M^{-\ell\cdot\beta_3}\right)\right)$$

holds for all positive, sufficiently small δ and all $\ell \in \mathbb{N}_0$, where $\beta_1 \geq 0$, $\beta_2 > 0$, and $\beta_2 \geq \beta_3 \geq 0$.

The parameters of a multi-level algorithm $\mathcal{A}$ are the minimal and maximal level, the replication numbers per level, the smoothing parameter δ, and the number of discrete points to be used in the first stage. We derive upper bounds for error$(\mathcal{A})$, the root mean square error, and cost$(\mathcal{A})$, the computational cost, in terms of these parameters and the values of r, α_i, and β_i, and we present the asymptotically optimal choice of the parameters with respect to our upper bounds. This leads to a final estimate of the form

$$\mathrm{cost}(\mathcal{A}) \leq O\left(\mathrm{error}(\mathcal{A})^{-\theta+\varepsilon}\right)$$

for every $\varepsilon > 0$, where $\theta > 0$. Roughly speaking, θ is the order of convergence of the multi-level algorithm. See Theorems 1–3 for the precise statements involving also powers of $\log \mathrm{error}(\mathcal{A})$.

Here we only present a particular application of these theorems for functionals

$$\varphi : C([0,T], \mathbb{R}^d) \to \mathbb{R}$$

of the solution process X of a d-dimensional system of SDEs, i.e., $Y = \varphi(X)$. For simplicity we take the Euler scheme with equidistant time-steps for the approximation of X in the construction of the multi-level algorithm, and we assume that $r \geq 1$ for the rest of the introduction. Table 1 contains the values of θ for the approximation of F and ρ on $[S_0, S_1]$ as well as for the approximation of F at a single point $s \in [S_0, S_1]$. In the first row φ is assumed to be Lipschitz continuous, and based on a well known upper bound for the strong error of the Euler scheme we show that (iii)–(v) are satisfied with

$$\alpha_1 = 0, \qquad \alpha_2 = 1/2 - \varepsilon, \qquad \alpha_3 = 1/2 - \varepsilon$$

and

$$\beta_1 = 1 + \varepsilon, \qquad \beta_2 = 1 - \varepsilon, \qquad \beta_3 = 1/2 - \varepsilon, \qquad \beta_4 = 2, \qquad \beta_5 = 1 - \varepsilon$$

for every $\varepsilon > 0$. In the second row

$$\varphi(x) = \inf\{t \geq 0 : x(t) \in \partial D\} \wedge T$$

is a stopped exit time from a bounded domain $D \subset \mathbb{R}^d$, and based on a recent result by Bouchard, Geiss, Gobet (2013) we obtain

$$\alpha_1 = 1, \qquad \alpha_2 = 1/2, \qquad \alpha_3 = 1/4$$

	F	ρ	$F(s)$
smooth functional	$2+\frac{2}{r+1}$	$2+\frac{4}{r}$	$2+\frac{1}{r+1}$
stopped exit time	$3+\frac{2}{r+1}$	$3+\frac{5}{r}$	$3+\frac{2}{r+1}$

Table 1. Orders of convergence of the multi-level algorithm

and

$$\beta_1 = 1, \qquad \beta_2 = 1/2, \qquad \beta_3 = 1/4, \qquad \beta_4 = 1, \qquad \beta_5 = 1/2.$$

We add that in every case represented in Table 1 proper multi-level algorithms turn out to be superior to single-level algorithms, as far as our upper bounds are concerned. We do not achieve better upper bounds if we restrict considerations to path-independent functionals, i.e., $Y = \varphi(X_T)$ with $\varphi : \mathbb{R}^d \to \mathbb{R}$ being Lipschitz continuous; here, however, the situation changes if the Euler scheme is replaced by the Milstein scheme (in dimension $d = 1$ because of assumption (ii)), which yields $\theta = 2 + 1/(r + 1)$, $\theta = 2 + 3/r$, and $\theta = 2$ for the approximation of F, ρ, and $F(s)$, respectively.

Corresponding results are available for the approximation of the expectation of $\varphi(X)$ by means of multi-level Euler algorithms. It is well known that $\theta = 2$, if φ is Lipschitz continuous, and $\theta = 3$ holds for stopped exit times φ, see Higham *et al.* (2013). In the limit $r \to \infty$ we achieve the same values of θ for the approximation of the distribution function or the density of $\varphi(X)$.

Multi-level algorithms, which have been introduced by Heinrich (1998) and Giles (2008a), see also Kebaier (2005) for the two-level construction, are meanwhile applied to rather different computational problems. The approximation of distribution functions and densities seems to be a new application, which exhibits, in particular, the following features: a singularity, which is due to the presence of the indicator function or the Dirac functional, and the fact that we approximate elements of function spaces instead of just real numbers. The first issue is also investigated, without smoothing, by Avikainen (2009) and Giles, Higham, Mao (2009), and with implicit smoothing through the use of conditional expectations by Giles (2008b) and Giles, Debrabant, Rößler (2013). Furthermore, Altmayer, Neuenkirch (2013) combine smoothing by Malliavin integration by parts with the multi-level approach to approximate expectations of discontinuous payoffs in the Heston model. The second issue has already been worked out by Heinrich (1998) in the general setting of algorithms taking values in Banach spaces.

We stress that a two-level construction for the approximation of densities in the SDE setting with $Y = X_T$ has already been proposed and analyzed by Kebaier, Kohatsu-Higa (2008) in the case $r = \infty$, and their analysis yields $\theta = 5/2$.

Optimality results, which do not just concern upper bounds for the error and cost of specific families of algorithms, seem to be unknown for the problems studied in the present paper. The situation is different for the approximation of expectations of Lipschitz continuous functionals, and here suitable multi-level algorithms are almost worst case optimal in the class of all randomized algorithms, see Creutzig *et al.* (2008).

This paper is organized as follows. In Sections 2–4 we provide the general analysis of the three approximation problem, namely, for distribution functions and densities on compact intervals and for distribution functions at a single point. The structure and the approach in each of these sections is similar: we discuss, in particular, the assumptions on the weak and the strong convergence, and we construct and analyze the respective multi-level algorithms. Section 5 contains, in particular, the application of the results from Sections 2–4 to functionals of the solutions of SDEs, which is complemented by numerical experiments for simple test cases in Section 6.

2. Approximation of Distribution Functions on Compact Intervals

We consider a random variable Y, and we study the approximation of its distribution function F on a compact interval $[S_0, S_1]$, with $S_0 < S_1$ being fixed throughout this section. We do not assume that the distribution of Y can be simulated exactly. Instead, we assume that the simulation is feasible for random variables $Y^{(\ell)}$ that converge to Y in a suitable way.

2.1. **Smoothing.** For the approximation of F a straight-forward application of the multi-level Monte Carlo approach based on

$$F(s) = \mathrm{E}(1_{]-\infty,s]}(Y))$$

could suffer from the discontinuity of $1_{]-\infty,s]}$, see Remark 8 below. This can be avoided by a smoothing step, provided that a density exists and is sufficiently smooth. Specifically, we assume that

(A1) the random variable Y has a density ρ on $\mathbb{R}$ that is r-times continuously differentiable on $[S_0 - \delta_0, S_1 + \delta_0]$ for some $r \in \mathbb{N}_0$ and $\delta_0 > 0$.

The smoothing is based on rescaled translates of a function $g : \mathbb{R} \to \mathbb{R}$ with the following properties:

(S1) The cost of computing $g(s)$ is bounded by a constant, uniformly in $s \in \mathbb{R}$.
(S2) g is Lipschitz continuous.
(S3) $g(s) = 1$ for $s < -1$ and $g(s) = 0$ for $s > 1$.
(S4) $\int_{-1}^{1} s^j \cdot (1_{]-\infty,0]}(s) - g(s))\, ds = 0$ for $j = 0, \dots, r-1$.

Obviously, g is bounded due to (S2) and (S3).

Remark 1. Such a function g is easily constructed as follows. There exists a uniquely determined polynomial p of degree at most $r+1$ such that

$$\int_{-1}^{1} s^j \cdot p(s)\, ds = (-1)^j/(j+1), \qquad j = 0, \dots, r-1,$$

as well as $p(1) = 0$ and $p(-1) = 1$. The extension g of p with $g(s) = 1$ for $s < -1$ and $g(s) = 0$ for $s > 1$ has the properties as claimed. Since $g - 1/2$ is an odd function, the same function g arises in this way for r and $r+1$, if r is even.

We have the following estimate for the bias that is induced by smoothing with parameter δ, i.e., by approximation of $1_{]-\infty,s]}$ by $g((\cdot - s)/\delta)$.

Lemma 1. *There exists a constant $c > 0$ such that*

$$\sup_{s \in [S_0, S_1]} |F(s) - \mathrm{E}(g((Y-s)/\delta))| \leq c \cdot \delta^{r+1}$$

holds for all $\delta \in]0, \delta_0]$.

Proof. Clearly

$$\begin{aligned} F(s) - \mathrm{E}(g((Y-s)/\delta)) &= \int_{-\infty}^{\infty} \rho(u) \cdot (1_{]-\infty,s]}(u) - g((u-s)/\delta))\, du \\ &= \delta \cdot \int_{-1}^{1} \rho(\delta u + s) \cdot (1_{]-\infty,0]}(u) - g(u))\, du, \end{aligned}$$

so that the statement follows in the case $r = 0$. For $r \geq 1$ the Taylor expansion

$$\rho(\delta u + s) = \sum_{j=0}^{r-1} \rho^{(j)}(s) \cdot (\delta u)^j / j! + R(\delta u, s)$$

yields

$$|F(s) - \mathrm{E}(g((Y-s)/\delta))| \leq \delta \cdot \int_{-1}^{1} |R(\delta u, s)| \cdot |1_{]-\infty,0]}(u) - g(u)|\, du \leq c \cdot \delta^{r+1}.$$

□

2.2. **Assumptions on Weak and Strong Convergence.** Our multi-level Monte Carlo construction is based on a sequence $(Y^{(\ell)})_{\ell \in \mathbb{N}_0}$ of random variables, defined on a common probability space together with Y, with the following properties for some constant $c > 0$:

(A2) There exists a constant $M > 1$ such that the simulation of the joint distribution of $Y^{(\ell)}$ and $Y^{(\ell-1)}$ is possible at cost at most $c \cdot M^{\ell}$ for every $\ell \in \mathbb{N}$.

(A3) There exist constants $\alpha_1 \geq 0$, $\alpha_2 > 0$, and $\alpha_2 \geq \alpha_3 \geq 0$ such that the weak error estimate

$$\sup_{s \in [S_0, S_1]} \left| \mathrm{E}\left(g((Y-s)/\delta) - g((Y^{(\ell)} - s)/\delta) \right) \right| \leq c \cdot \min\left(\delta^{-\alpha_1} \cdot M^{-\ell \cdot \alpha_2}, M^{-\ell \cdot \alpha_3} \right)$$

holds for all $\delta \in]0, \delta_0]$ and $\ell \in \mathbb{N}_0$.

(A4) There exist constants $\beta_4 \geq 0$ and $\beta_5 > 0$ such that the strong error estimate

$$\mathrm{E} \min((Y - Y^{(\ell)})^2 / \delta^2, 1) \leq c \cdot \delta^{-\beta_4} \cdot M^{-\ell \cdot \beta_5}$$

holds for all $\delta \in]0, \delta_0]$ and $\ell \in \mathbb{N}_0$.

For specific applications we present suitable approximations $Y^{(\ell)}$ and corresponding values of the parameters M, α_i and β_i in Section 5. Here we proceed with a general discussion of (A3) and (A4).

Note that (A4) implies (A3) with $\alpha_1 = \beta_4/2$, $\alpha_2 = \beta_5/2$, and $\alpha_3 = 0$, but often better estimates for the weak error are known, see Sections 4.2 and 5. The presence of α_1 and β_4 in these assumptions is motivated by weak and strong error estimates for SDEs or SPDEs, which often scale with some power of δ. See, however, Sections 5.1 and 5.2

Let $\|Z\|_p = (\mathrm{E}\,|Z|^p)^{1/p}$ for any random variable Z and $1 \leq p < \infty$. Typically, strong error estimates for $Y - Y^{(\ell)}$ instead of $\min(|Y - Y^{(\ell)}|, \delta)$ are available in the literature. Straightforward relations to (A3) and (A4) are provided by

$$\sup_{s \in [S_0, S_1]} \left| \mathrm{E}\left(g((Y-s)/\delta) - g((Y^{(\ell)} - s)/\delta) \right) \right| \leq c_L \cdot \delta^{-1} \cdot \|Y - Y^{(\ell)}\|_1, \tag{1}$$

where c_L denotes a Lipschitz constant for g, as well as

$$\mathrm{E} \min((Y - Y^{(\ell)})^2, \delta^2) \leq \min(\|Y - Y^{(\ell)}\|_2^2, \delta^2) \tag{2}$$

and

$$\mathrm{E} \min((Y - Y^{(\ell)})^2, \delta^2) \leq \mathrm{E}(\delta \cdot \min(|Y - Y^{(\ell)}|, \delta)) \leq \min(\delta \cdot \|Y - Y^{(\ell)}\|_1, \delta^2). \tag{3}$$

In the following case of equivalence of norms the upper bound in (2) is sharp, and then we have $\beta_4 = 2$ in (A4), while the optimal value of β_5 is determined by the asymptotic behavior of $\|Y - Y^{(\ell)}\|_2^2$. See Sections 5.1 and 5.2 for examples.

Lemma 2. *Suppose that there exist* $c_1 > 0$ *and* $p > 2$ *such that*

$$0 < \|Y - Y^{(\ell)}\|_p \leq c_1 \cdot \|Y - Y^{(\ell)}\|_2$$

for all $\ell \in \mathbb{N}_0$. *Then there exists* $c_2 > 0$ *such that*

$$\mathrm{E} \min((Y - Y^{(\ell)})^2, \delta^2) \geq c_2 \cdot \min(\|Y - Y^{(\ell)}\|_2^2, \delta^2)$$

for all $\delta \in \,]0, \delta_0]$ *and* $\ell \in \mathbb{N}_0$.

Proof. Put

$$Z_\ell = \frac{(Y - Y^{(\ell)})^2}{\|Y - Y^{(\ell)}\|_2^2}.$$

We show that there exists a constant $c_2 > 0$ such that

$$\mathrm{E} \min(Z_\ell, \delta) \geq c_2 \cdot \min(1, \delta)$$

for all $\ell \in \mathbb{N}_0$ and $\delta > 0$.

Clearly $\mathrm{E}(Z_\ell) = 1$ and $\mathrm{E}(Z_\ell^{p/2}) \leq c_1^p$. It follows that

$$P(\{Z_\ell > u\}) \leq \frac{c_1^p}{u^{p/2}}.$$

Put

$$d_\ell = P(\{Z_\ell > 1/2\}).$$

We claim that

$$d = \inf_{\ell \in \mathbb{N}_0} d_\ell > 0.$$

Assume that $d = 0$. Use

$$1 = \mathrm{E}(Z_\ell) = \int_0^\infty P(\{Z_\ell > u\})\, du \leq 1/2 + \int_{1/2}^\infty \min(d_\ell, c_1^p / u^{p/2})\, du$$

and dominated convergence to conclude that, for a minimizing subsequence,

$$\lim_{k \to \infty} \int_{1/2}^\infty \min(d_{\ell_k}, c_1^p / u^{p/2})\, du = 0,$$

which leads to a contradiction. Therefore

$$\mathrm{E} \min(Z_\ell, \delta) = \int_0^\delta P(\{Z_\ell > u\})\, du \geq \min(\delta, 1/2) \cdot d \geq d/2 \cdot \min(1, \delta).$$

□

On the other hand, if $\|Y - Y^{(\ell)}\|_2^2$ and $\|Y - Y^{(\ell)}\|_1$ are asymptotically equivalent, then (3) is preferable to (2). See Section 5.3 for examples.

Assumption (A4) and the Lipschitz continuity and boundedness of g immediately yield the following fact.

Lemma 3. *There exists a constant* $c > 0$ *such that*

$$\mathrm{E} \sup_{s \in [S_0, S_1]} \big(g((Y - s)/\delta) - g((Y^{(\ell)} - s)/\delta)\big)^2 \leq c \cdot \min(\delta^{-\beta_4} \cdot M^{-\ell \cdot \beta_5}, 1)$$

holds for all $\delta \in \,]0, \delta_0]$ *and* $\ell \in \mathbb{N}_0$.

2.3. **The Multi-level Algorithm.** The approximation of F on the interval $[S_0, S_1]$ is based on its approximation at finitely many points

$$S_0 \le s_1 < \cdots < s_k \le S_1, \tag{4}$$

followed by a suitable extension to $[S_0, S_1]$.

For notational convenience we put

$$g^{k,\delta}(t) = (g((t-s_1)/\delta), \dots, g((t-s_k)/\delta)) \in \mathbb{R}^k, \qquad t \in \mathbb{R},$$

as well as $Z_i^{(0)} = Y^{(-1)} = 0$.

We choose $L_0, L_1 \in \mathbb{N}_0$ with $L_0 \le L_1$ as the minimal and the maximal level, respectively, and we choose replication numbers $N_\ell \in \mathbb{N}$ for all levels $\ell = L_0, \dots, L_1$, as well as $k \in \mathbb{N}$ and $\delta \in \,]0, \delta_0]$. The corresponding multi-level algorithm for the approximation at the points s_i is defined by

$$\mathcal{M}^{k,\delta,L_0,L_1}_{N_{L_0},\dots,N_{L_1}} = \frac{1}{N_{L_0}} \cdot \sum_{i=1}^{N_{L_0}} g^{k,\delta}(Y_i^{(L_0)}) + \sum_{\ell=L_0+1}^{L_1} \frac{1}{N_\ell} \cdot \sum_{i=1}^{N_\ell} \Big(g^{k,\delta}(Y_i^{(\ell)}) - g^{k,\delta}(Z_i^{(\ell)}) \Big) \tag{5}$$

with an independent family of $\mathbb{R}^2$-valued random variables $(Y_i^{(\ell)}, Z_i^{(\ell)})$ for $\ell = L_0, \dots, L_1$ and $i = 1, \dots, N_\ell$ such that equality in distribution holds for $(Y_i^{(\ell)}, Z_i^{(\ell)})$ and $(Y^{(\ell)}, Y^{(\ell-1)})$.

Remark 2. In the particular case $L = L_0 = L_1$, i.e., in the single-level case, we actually have a classical Monte Carlo algorithm, based on independent copies of $Y^{(L)}$ only. In addition to

$$\mathcal{M}^{k,\delta,L,L}_{N} = \frac{1}{N} \cdot \sum_{i=1}^{N} g^{k,\delta}(Y_i^{(L)})$$

with $\delta > 0$, we also consider the single-level algorithm without smoothing. Hence we put

$$g^{k,0}(t) = \big(1_{]-,\infty,s_1]}(t), \dots, 1_{]-,\infty,s_k]}(t)\big) \in \mathbb{R}^k, \qquad t \in \mathbb{R},$$

to obtain

$$\mathcal{M}^{k,0,L,L}_{N} = \frac{1}{N} \cdot \sum_{i=1}^{N} g^{k,0}(Y_i^{(L)}).$$

Observe that $\mathcal{M}^{k,0,L,L}_{N}$ yields the values of the empirical distribution function, based on N independent copies of $Y^{(L)}$, at the points s_i.

For the analysis of the single-level algorithm it suffices to assume that the simulation of the distribution of $Y^{(\ell)}$ is possible at cost at most $c \cdot M^\ell$ for every $\ell \in \mathbb{N}$, cf. (A2). Furthermore, there is no need for a strong error estimate like (A4), and if we do not employ smoothing, then (A3) may be replaced by the following assumption. There exist a constant $\alpha > 0$ such that the weak error estimate

$$\sup_{s \in [S_0,S_1]} \left| \mathrm{E}\left(1_{]-\infty,s]}(Y) - 1_{]-\infty,s]}(Y^{(\ell)})\right) \right| \le c \cdot M^{-\ell \cdot \alpha} \tag{6}$$

holds for all $\ell \in \mathbb{N}_0$. It turns out that the analysis of single-level algorithms without smoothing is formally reduced to the case $\delta > 0$ if we take

$$\alpha_1 = 0, \qquad \alpha_2 = \alpha, \qquad \alpha_3 = \alpha. \tag{7}$$

In the sequel $\|\cdot\|_\infty$ denotes the supremum norm on $C([S_0,S_1])$ and $|\cdot|_\infty$ denotes the ℓ_∞-norm on $\mathbb{R}^k$.

For the extension we take a sequence of linear mappings $Q^r_k : \mathbb{R}^k \to C([S_0,S_1])$ with the following properties for some constant $c > 0$:

(E1) For all $k \in \mathbb{N}$ and $x \in \mathbb{R}^k$ the cost for computing $Q^r_k(x)$ is bounded by $c \cdot k$.

(E2) For all $k \in \mathbb{N}$ and $x \in \mathbb{R}^k$

$$\|Q^r_k(x)\|_\infty \leq c \cdot |x|_\infty.$$

(E3) For all $k \in \mathbb{N}$

$$\|F - Q^r_k(F(s_1),\dots,F(s_k))\|_\infty \leq c \cdot k^{-(r+1)}.$$

These properties are achieved, e.g., by piecewise polynomial interpolation with degree $\max(r,1)$ at equidistant points $s_i = S_0 + (i-1)\cdot(S_1-S_0)/(k-1)$ with $k \geq 2$.

We employ $Q^r_k(\mathcal{M})$ with $\mathcal{M} = \mathcal{M}^{k,\delta,L_0,L_1}_{N_{L_0},\dots,N_{L_1}}$ as a randomized algorithm for the approximation of F on $[S_0,S_1]$. Observe that $\mathcal{M}$ is square-integrable, since g is bounded, so that (E2) yields $\mathrm{E}\,\|Q^r_k(\mathcal{M})\|^2_\infty < \infty$. The error of $Q^r_k(\mathcal{M})$ is defined by

$$\mathrm{error}(Q^r_k(\mathcal{M})) = \left(\mathrm{E}\,\|F - Q^r_k(\mathcal{M})\|^2_\infty\right)^{1/2}.$$

Since the error is based on the supremum norm, $\mathrm{error}(Q^r_k(\mathcal{M}))$ does not increase if we replace $Q^r_k(x)$ by $s \mapsto \sup_{u\in[S_0,s]}(Q^r_k(x))(u)$ to get a non-decreasing approximation on $[S_0,S_1]$.

The variance of any square-integrable $\mathbb{R}^k$-valued random variable $\mathcal{M}$ is defined by

$$\mathrm{Var}(\mathcal{M}) = \mathrm{E}\,|\mathcal{M} - \mathrm{E}(\mathcal{M})|^2_\infty,$$

and

$$\mathrm{E}\,|x - \mathcal{M}|^2_\infty \leq 2\cdot(|x - \mathrm{E}(\mathcal{M})|^2_\infty + \mathrm{Var}(\mathcal{M}))$$

holds for $x \in \mathbb{R}^k$. Furthermore,

$$\mathrm{Var}(\mathcal{M}) \leq 4 \cdot \mathrm{E}(|\mathcal{M}|^2_\infty).$$

The Bienaymé formula for real-valued random variables turns into the inequality

$$\mathrm{Var}(\mathcal{M}) \leq c \cdot \log k \cdot \sum_{i=1}^n \mathrm{Var}(\mathcal{M}_i), \tag{8}$$

if $\mathcal{M} = \sum_{i=1}^n \mathcal{M}_i$ with independent square-integrable random variables $\mathcal{M}_i$ taking values in $\mathbb{R}^k$. Here c is a universal constant. In the context of multi-level algorithms this is exploited in Heinrich (1998).

We say that a sequence of randomized algorithms $\mathcal{A}_n$ converges with order $(\gamma,\eta) \in \,]0,\infty[\, \times \mathbb{R}$ if $\lim_{n\to\infty} \mathrm{error}(\mathcal{A}_n) = 0$ and if there exists a constant $c > 0$ such that

$$\mathrm{cost}(\mathcal{A}_n) \leq c \cdot (\mathrm{error}(\mathcal{A}_n))^{-\gamma} \cdot (-\log \mathrm{error}(\mathcal{A}_n))^{\eta}.$$

Moreover, we put

$$q = \min\left(\frac{r+1+\alpha_1}{\alpha_2}, \frac{r+1}{\alpha_3}\right). \tag{9}$$

Theorem 1. *The following order, with* $\eta = 1$*, is achieved by algorithms* $Q_k^r(\mathcal{M}^{k,\delta,L_0,L_1}_{N_{L_0},\dots,N_{L_1}})$ *with suitably chosen parameters:*

$$q \le \max(1, \beta_4/\beta_5) \quad \Rightarrow \quad \gamma = 2 + \frac{\max(1,q)}{r+1}, \tag{10}$$

$$q > \max(1, \beta_4/\beta_5) \ \wedge \ \beta_5 > 1 \quad \Rightarrow \quad \gamma = 2 + \frac{\max(1,\beta_4/\beta_5)}{r+1}, \tag{11}$$

$$q > 1 > \beta_4 \ \wedge \ \beta_5 = 1 \quad \Rightarrow \quad \gamma = 2 + \frac{1}{r+1}, \tag{12}$$

$$q > \max(1, \beta_4/\beta_5) \ \wedge \ \beta_5 < 1 \quad \Rightarrow \quad \gamma = 2 + \frac{\max(1, \beta_4 + (1-\beta_5)\cdot q)}{r+1}. \tag{13}$$

Moreover, with $\eta = 3$*,*

$$q > \beta_4 \ge 1 \ \wedge \ \beta_5 = 1 \quad \Rightarrow \quad \gamma = 2 + \frac{\beta_4}{r+1}. \tag{14}$$

Proof. Let $\mathcal{M}$ denote any square-integrable random variable with values in $\mathbb{R}^k$. For the error of $Q_k^r(\mathcal{M})$ we have

$$\begin{aligned}
\operatorname{error}(Q_k^r(\mathcal{M})) &\le \|F - Q_k^r(F(s_1),\dots,F(s_k))\|_\infty + \big(\mathrm{E}\,\|Q_k^r((F(s_1),\dots F(s_k)) - \mathcal{M})\|_\infty^2\big)^{1/2} \\
&\le c \cdot \Big(k^{-(r+1)} + \big(\mathrm{E}\,|(F(s_1),\dots F(s_k)) - \mathcal{M}|_\infty^2\big)^{1/2}\Big) \\
&\le 2c \cdot \big(k^{-2(r+1)} + |(F(s_1),\dots F(s_k)) - \mathrm{E}(\mathcal{M})|_\infty^2 + \operatorname{Var}(\mathcal{M})\big)^{1/2}
\end{aligned}$$

with a constant $c > 0$ according to (E2) and (E3).

Now we consider the algorithm $\mathcal{M} = \mathcal{M}^{k,\delta,L_0,L_1}_{N_{L_0},\dots,N_{L_1}}$ with $\delta > 0$. We write $a \preceq b$ if there exists a constant $c > 0$ that does not depend on the parameters k, δ, L_0, L_1, N_{L_0}, $\dots$, N_{L_1} such that $a \le c \cdot b$. Moreover, $a \succeq b$ means $b \preceq a$, and $a \asymp b$ stands for $a \preceq b$ and $a \succeq b$.

Note that $\mathrm{E}(\mathcal{M}) = \mathrm{E}(g^{k,\delta}(Y^{(L_1)}))$. Hence the bias term is estimated by

$$\begin{aligned}
|(F(s_1),\dots,F(s_k)) - \mathrm{E}(\mathcal{M})|_\infty &= \sup_{i=1,\dots,k} |F(s_i) - \mathrm{E}(g((Y^{(L_1)} - s_i)/\delta))| \\
&\preceq \delta^{r+1} + \min\left(\delta^{-\alpha_1} \cdot M^{-L_1\cdot\alpha_2}, M^{-L_1\cdot\alpha_3}\right),
\end{aligned}$$

see Lemma 1 and (A3).

The variance of $\mathcal{M}$ is estimated as follows. From (8) we obtain

$$\operatorname{Var}(\mathcal{M}) \preceq \log k \cdot \left(\frac{1}{N_{L_0}} \cdot \operatorname{Var}(g^{k,\delta}(Y^{(L_0)})) + \sum_{\ell=L_0+1}^{L_1} \frac{1}{N_\ell} \cdot \operatorname{Var}\left(g^{k,\delta}(Y^{(\ell)}) - g^{k,\delta}(Y^{(\ell-1)})\right)\right).$$

Moreover,

$$\begin{aligned}
\operatorname{Var}\left(g^{k,\delta}(Y^{(\ell)}) - g^{k,\delta}(Y^{(\ell-1)})\right) &\le 4 \cdot \mathrm{E} \sup_{i=1,\dots,k} \left(g((Y^{(\ell)} - s_i)/\delta) - g((Y^{(\ell-1)} - s_i)/\delta)\right)^2 \\
&\preceq \min(\delta^{-\beta_4} \cdot M^{-\ell\cdot\beta_5}, 1)
\end{aligned}$$

for $\ell = L_0 + 1, \dots, L_1$, see Lemma 3, and

$$\operatorname{Var}(g^{k,\delta}(Y^{(L_0)})) \preceq 1,$$

since g is bounded. Therefore

$$\operatorname{Var}(\mathcal{M}) \preceq \log k \cdot \left(\frac{1}{N_{L_0}} + \sum_{\ell=L_0+1}^{L_1} \frac{\min(\delta^{-\beta_4} \cdot M^{-\ell \cdot \beta_5}, 1)}{N_\ell} \right).$$

Combining these estimates we finally get

$$\begin{aligned} \text{error}^2(Q_k^r(\mathcal{M})) \preceq{}& k^{-2(r+1)} + \delta^{2(r+1)} + \min\left(\delta^{-2\alpha_1} \cdot M^{-L_1 \cdot 2\alpha_2}, M^{-L_1 \cdot 2\alpha_3}\right) \\ &+ \log k \cdot \left(\frac{1}{N_{L_0}} + \sum_{\ell=L_0+1}^{L_1} \frac{\min(\delta^{-\beta_4} \cdot M^{-\ell \cdot \beta_5}, 1)}{N_\ell} \right). \end{aligned} \tag{15}$$

Now we analyze the computational cost of the algorithm $\mathcal{M}$. For $\ell = L_0, \ldots, L_1$ and $i = 1, \ldots, N_\ell$ the cost of computing $g^{k,\delta}(Y_i^{(\ell)})$ or $g^{k,\delta}(Y_i^{(\ell)}) - g^{k,\delta}(Z_i^{(\ell)})$ is bounded by $M^\ell + k$, up to a constant, see (S1) and (A2). Use (E1) to obtain

$$\text{cost}(Q_k^r(\mathcal{M})) \preceq c(k, L_0, L_1, N_{L_0}, \ldots, N_{L_1}) \tag{16}$$

with

$$c(k, L_0, L_1, N_{L_0}, \ldots, N_{L_1}) = \sum_{\ell=L_0}^{L_1} N_\ell \cdot (M^\ell + k). \tag{17}$$

Note that for every k the cost per replication is essentially constant on all levels $\ell \le L^*$, where

$$L^* = \log_M k. \tag{18}$$

Observe that the estimates (15) and (16) are valid, too, for single-level algorithms without smoothing, i.e., for $L_0 = L_1$ and $\delta = 0$, if we formally define the parameters α_i by (7), which leads to $q = (r+1)/\alpha$.

We determine parameters of the algorithm $Q_k^r(\mathcal{M})$ such that an error of about $\epsilon \in \left]0, \min(1, \delta_0^{r+1})\right[$ is achieved at a small cost. More precisely, we minimize the upper bound (16) for the cost, subject to the constraint that the upper bound (15) for the squared error is at most ϵ^2, up to multiplicative constants for both quantities.

First of all we consider the case $\delta > 0$, and we choose

$$\delta = \epsilon^{1/(r+1)} \tag{19}$$

and, up to integer rounding,

$$k = \epsilon^{-1/(r+1)} \tag{20}$$

and

$$N_{L_0} = \epsilon^{-2} \cdot \log_M \epsilon^{-1}. \tag{21}$$

This yields

$$\text{error}^2(Q_k^r(\mathcal{M})) \preceq \epsilon^2 + a^2(L_1) + \log \epsilon^{-1} \cdot \sum_{\ell=L_0+1}^{L_1} \frac{\min(\delta^{-\beta_4} \cdot M^{-\ell \cdot \beta_5}, 1)}{N_\ell}$$

with

$$a(L_1) = \min\left(\delta^{-\alpha_1} \cdot M^{-L_1 \cdot \alpha_2}, M^{-L_1 \cdot \alpha_3}\right). \tag{22}$$

Furthermore,

$$L^* = \frac{1}{r+1} \cdot \log_M \epsilon^{-1}. \tag{23}$$

Due to the dependence of (16) on k and the decay of $a(L_1)$ and $\min(\delta^{-\beta_4} \cdot M^{-\ell \cdot \beta_5}, 1)$ as functions of L_1 and ℓ, respectively, it suffices to study

$$L_0 \geq L^*. \tag{24}$$

Moreover, $a(L_1) \leq \epsilon$ requires $L_1 \geq q \cdot L^*$. Consequently, we choose

$$L_1 = \max(1, q) \cdot L^*, \tag{25}$$

up to integer rounding.

For a single-level algorithm with smoothing, i.e., for $L_0 = L_1$ and $\delta > 0$, all parameters have thus been determined, and we obtain $\mathrm{error}(Q_k^r(\mathcal{M})) \preceq \epsilon$ as well as

$$c(k, L_1, L_1, N_{L_1}) \asymp \epsilon^{-2} \cdot \log \epsilon^{-1} \cdot M^{L^*} = \epsilon^{-2-1/(r+1)} \cdot \log \epsilon^{-1} \tag{26}$$

if $q \leq 1$, and

$$c(k, L_1, L_1, N_{L_1}) \asymp \epsilon^{-2} \cdot \log \epsilon^{-1} \cdot M^{q \cdot L^*} = \epsilon^{-2-q/(r+1)} \cdot \log \epsilon^{-1}, \tag{27}$$

if $q > 1$. For a single-level algorithm without smoothing we obtain the same result.

For a proper multi-level algorithm with

$$L^* \leq L_0 < L_1$$

we obtain

$$\mathrm{error}^2(Q_k^r(\mathcal{M})) \preceq \epsilon^2 + \log \epsilon^{-1} \cdot \sum_{\ell=L_0+1}^{L_1} \frac{v_\ell}{N_\ell}$$

with

$$v_\ell = \min(M^{L^* \cdot \beta_4} \cdot M^{-\ell \cdot \beta_5}, 1)$$

as well as

$$c(k, L_0, L_1, N_{L_0}, \dots, N_{L_1}) \asymp \epsilon^{-2} \cdot \log \epsilon^{-1} \cdot M^{L_0} + \sum_{\ell=L_0+1}^{L_1} N_\ell \cdot M^\ell.$$

Observing

$$c(k, L_0, L_1, N_{L_0}, \dots, N_{L_1}) \succeq \epsilon^{-2} \cdot \log \epsilon^{-1} \cdot M^{L^*}$$

and (26), we get (10) in the case $q \leq 1$ already by single-level algorithms.

To establish the theorem in the case

$$q > 1$$

we fix L_0 for the moment, and we minimize

$$h(L_0, N_{L_0+1}, \dots, N_{L_1}) = \epsilon^{-2} \cdot \log \epsilon^{-1} \cdot M^{L_0} + \sum_{\ell=L_0+1}^{L_1} N_\ell \cdot M^\ell$$

subject to

$$\sum_{\ell=L_0+1}^{L_1} \frac{v_\ell}{N_\ell} \leq \epsilon^2 / \log \epsilon^{-1}.$$

A Lagrange multiplier leads to

$$N_\ell = \epsilon^{-2} \cdot \log \epsilon^{-1} \cdot G(L_0) \cdot \left(v_\ell \cdot M^{-\ell}\right)^{1/2}, \tag{28}$$

up to integer rounding, which satisfies the constraint with

$$G(L_0) = \sum_{\ell=L_0+1}^{L_1} \left(v_\ell \cdot M^\ell\right)^{1/2} = \sum_{\ell=L_0+1}^{L_1} \left(\min(M^{L^*\cdot\beta_4} \cdot M^{-\ell\cdot\beta_5}, 1) \cdot M^\ell\right)^{1/2}.$$

Moreover, this choice of $N_{L_0+1}, \dots, N_{L_1}$ yields

$$h(L_0, N_{L_0+1}, \dots, N_{L_1}) = \epsilon^{-2} \cdot \log \epsilon^{-1} \cdot \left(M^{L_0} + G^2(L_0)\right). \tag{29}$$

Put

$$L^\dagger = \frac{\beta_4}{\beta_5} \cdot L^*.$$

Consider the case

$$1 < q \leq \beta_4/\beta_5.$$

Then we have $L_1 \leq L^\dagger$, and therefore

$$M^{L_0} + G^2(L_0) = M^{L_0} + \left(\sum_{\ell=L_0+1}^{L_1} M^{\ell/2}\right)^2 \asymp M^{L_0} + M^{L_1} \asymp M^{L^*\cdot q}.$$

Observing (27) we get (10) in the present case already by single-level algorithms.

From now on we consider the case

$$q > \max(1, \beta_4/\beta_5).$$

Suppose that $L_0 < L^\dagger$, which requires $\beta_4/\beta_5 > 1$ to hold. Then we get

$$\begin{aligned} M^{L_0} + G^2(L_0) &\asymp M^{L_0} + \left(\sum_{\ell=L_0+1}^{L^\dagger} M^{\ell/2}\right)^2 + M^{L^*\cdot\beta_4} \cdot \left(\sum_{\ell=L^\dagger+1}^{L_1} M^{\ell\cdot(1-\beta_5)/2}\right)^2 \\ &\asymp M^{L^\dagger} + M^{L^*\cdot\beta_4} \cdot \left(\sum_{\ell=L^\dagger+1}^{L_1} M^{\ell\cdot(1-\beta_5)/2}\right)^2 \asymp M^{L^\dagger} + G^2(L^\dagger). \end{aligned}$$

It therefore suffices to study the case

$$L_0 \geq L^\dagger,$$

where we have

$$M^{L_0} + G^2(L_0) = M^{L_0} + M^{L^*\cdot\beta_4} \cdot \left(\sum_{\ell=L_0+1}^{L_1} M^{\ell\cdot(1-\beta_5)/2}\right)^2.$$

If $\beta_5 = 1$ then

$$M^{L_0} + G^2(L_0) \asymp M^{L_0} + M^{L^*\cdot\beta_4} \cdot (L_1 - L_0)^2.$$

If $\beta_5 > 1$ then

$$M^{L_0} + G^2(L_0) \asymp M^{L_0} + M^{L^*\cdot\beta_4} \cdot M^{L_0\cdot(1-\beta_5)} \asymp M^{L_0}.$$

If $\beta_5 < 1$ then

$$M^{L_0} + G^2(L_0) \asymp M^{L_0} + M^{L^*\cdot\beta_4} \cdot M^{L_1\cdot(1-\beta_5)}.$$

Hence we choose

$$L_0 = \max(1, \beta_4/\beta_5) \cdot L^* \tag{30}$$

in all these cases. Hereby we obtain

$$M^{L_0} + G^2(L_0) \asymp M^{L^* \cdot \max(1,\beta_4/\beta_5)} \cdot \begin{cases} (L^*)^2, & \text{if } \beta_5 = 1 \text{ and } \beta_4 \geq 1, \\ 1, & \text{if } \beta_5 > 1 \text{ or } \beta_5 = 1 \text{ and } \beta_4 < 1, \end{cases}$$

as well as

$$M^{L_0} + G^2(L_0) \asymp M^{\max(1,\beta_4/\beta_5,\beta_4+(1-\beta_5)\cdot q)\cdot L^*}$$

if $\beta_5 < 1$. In any case these estimates are superior to $M^{L^* \cdot q}$, cf. (27). Use (29) and $M^{L^*} = \epsilon^{-1/(r+1)}$ to derive (11)–(14). □

Remark 3. Theorem 1 is based on the upper bounds (15) and (16) for the error and the cost, respectively, of the algorithms $Q_k^r(\mathcal{M}_{N_{L_0},\dots,N_{L_1}}^{k,\delta,L_0,L_1})$. The parameters that we have determined in the proof of Theorem 1 are optimal in the following sense: they minimize the upper bound for the cost, subject to the constraint that the upper bound for the error is at most ϵ, up to multiplicative constants for both quantities.

Obviously, this optimality holds true for the choice of δ, k, N_{L_0}, and L_1 according to (19), (20), (21), and (25). Moreover, the constraint (24) is without loss of generality, so that the minimal level L_0 slowly increases with decreasing ϵ.

This completes, in particular, the optimization of the parameters of single-level algorithms, where $L_0 = L_1$. For proper multi-level algorithms the optimal values of N_ℓ for $\ell = L_0 + 1, \dots, L_1$ are presented in (28) and the optimal value of L_0 is presented in (30), if $q > \max(1, \beta_4/\beta_5)$. It is straightforward to verify

$$N_\ell = \epsilon^{-2-\beta_4/(r+1)} \cdot \log \epsilon^{-1} \cdot M^{-\ell\cdot(1+\beta_5)/2} \cdot \begin{cases} L^*, & \text{if } \beta_5 = 1, \\ M^{L^*\cdot\max(1,\beta_4/\beta_5)\cdot(1-\beta_5)/2} & \text{if } \beta_5 > 1, \\ M^{L^*\cdot q\cdot(1-\beta_5)/2}, & \text{if } \beta_5 < 1. \end{cases} \tag{31}$$

Furthermore, we have carried out the comparison of multi-level and single-level algorithms in the proof of Theorem 1. This comparison, too, is merely based on the upper bounds for the error and the cost, and on the assumption that $\alpha = \alpha_3$ in (6). In this sense we have a superiority of proper multi-level algorithms over single-level algorithms if and only if

$$q > \max(1, \beta_4/\beta_5), \tag{32}$$

which corresponds to (11)–(14) in Theorem 1. The lack of superiority, which is present in (10) in Theorem 1, is explained as follows. For $q \leq 1$ the maximal level can be chosen so small that the computational cost on all levels is dominated by the number k of discretization points that is needed to achieve a good approximation of F even from exact data $F(s_1), \dots, F(s_k)$. For $1 < q \leq \beta_4/\beta_5$ the negative impact of smoothing on the variances leads to variances $\min(\delta^{-\beta_4}\cdot M^{-\ell\cdot\beta_5}, 1)$ of order one on all levels $\ell = L_0+1, \dots, L_1$.

Single-level algorithms with smoothing are never inferior to single-level algorithms without smoothing, and they are superior if and only if

$$\frac{r+1}{\alpha_3} > \max(1, q). \tag{33}$$

For large values of r the latter holds true if and only if $\alpha_2 > \alpha_3$; see Section 5.3 for an example.

Remark 4. In the limit $r \to \infty$ we get

$$\gamma = 2 + \frac{\max(1-\beta_5, 0)}{\alpha_2}$$

in Theorem 1, which coincides with the order for the approximation of expectations by means of multi-level algorithms, see Giles (2008a, Thm. 3.1).

Consider the empirical distribution function $\hat{F}_n$ based on n independent copies of Y. The Dvoretzky-Kiefer-Wolfowitz inequality, with the optimal constant due to Massart (1990), yields

$$\left(\mathrm{E} \sup_{s \in \mathbb{R}} |F(s) - \hat{F}_n(s)|^2 \right)^{1/2} \leq n^{-1/2},$$

which corresponds to an order two of approximation in terms of the number of samples from the target distribution. In our analysis we do not assume that sampling from the target distribution is feasible, and we fully take into account the computational cost to generate samples from approximate distributions. Still, if β_5 is almost one and if r is large, a suitable multi-level algorithm almost achieves the order two. See Sections 5.1 and 5.2 for examples.

3. Approximation of Densities on Compact Intervals

In this section we study the approximation of the density ρ of Y on an interval $[S_0, S_1]$ for some fixed $S_0 < S_1$. The construction and analysis closely follows the approach from Section 2.

3.1. **Smoothing.**

We employ assumption (A1) with $r \geq 1$, and $g : \mathbb{R} \to \mathbb{R}$ is assumed to satisfy the properties (S1) and (S2), while (S3) and (S4) are replaced by:

- (S5) $g(s) = 0$ if $|s| > 1$.
- (S6) $\int_{-1}^{1} g(s)\, ds = 1$ and $\int_{-1}^{1} s^j \cdot g(s)\, ds = 0$ for $j = 1, \ldots, r-1$.

Obviously, g is bounded due to (S2) and (S5). Moreover, if $g \in C^1(\mathbb{R})$ satisfies (S3) and (S4) and g' is Lipschitz continuous, then $-g'$, instead of g, satisfies (S5) and (S6). In kernel density estimation, a function g with integral one and vanishing moments up to order $r-1$ is called a kernel of order (at least) r.

Remark 5. We modify the construction from Remark 1 as follows. There exists a uniquely determined polynomial p of degree at most $r+1$ such that

$$\int_{-1}^{1} p(s)\, ds = 1$$

and

$$\int_{-1}^{1} s^j \cdot p(s)\, ds = 0, \qquad j = 0, \ldots, r-1,$$

as well as $p(1) = p(-1) = 0$. Extend p by zero to obtain g with the properties as claimed. Since g is an even function, the same function g arises in this way for r and $r+1$, if r is odd.

We have the following estimate for the bias that is induced by smoothing with parameter δ, i.e., by approximation of the Dirac functional at s by $1/\delta \cdot g((\cdot - s)/\delta)$. See, e.g., Tsybakov (2009, Prop. 1.2).

Lemma 4. *There exists a constant $c > 0$ such that*

$$\sup_{s\in[S_0,S_1]} |\rho(s) - 1/\delta \cdot \mathrm{E}(g((Y-s)/\delta))| \leq c \cdot \delta^r$$

holds for all $\delta \in \,]0, \delta_0]$.

Proof. Clearly

$$\rho(s) - 1/\delta \cdot \mathrm{E}(g((Y-s)/\delta)) = \int_{-1}^{1} g(u) \cdot (\rho(s) - \rho(\delta u + s))\, du.$$

Use a Taylor expansion to derive

$$|\rho(s) - 1/\delta \cdot \mathrm{E}(g((Y-s)/\delta))| \leq c \cdot \delta^r.$$

□

3.2. **Assumptions on Weak and Strong Convergence.** We employ the assumptions (A2)–(A4) from Section 2.2 with possibly different values of α_i in the weak error estimate (A3). We make use of Lemma 3, and we refer to Section 5 for specific examples with corresponding values of α_i.

3.3. **The Multi-level Algorithm.** The definition (5) of the algorithms $\mathcal{M}^{k,\delta,L_0,L_1}_{N_{L_0},\dots,N_{L_1}}$ also applies for the approximation of densities, except for $g^{k,\delta}$, which is now defined by

$$g^{k,\delta}(t) = \frac{1}{\delta} \cdot (g((t-s_1)/\delta), \dots, g((t-s_k)/\delta)) \in \mathbb{R}^k, \qquad t \in \mathbb{R}.$$

In the present setting we have $\delta > 0$ also for single-level algorithms.

Hereby we obtain approximations to ρ at the points (4), which are extended to functions on $[S_0, S_1]$ by means of linear mappings $Q^r_k : \mathbb{R}^k \to C([S_0, S_1])$. We assume that (E1) and (E2) are satisfied, but instead of (E3) the following property is assumed to hold with some constant $c > 0$:

(E4) For all $k \in \mathbb{N}$

$$\|\rho - Q^r_k(\rho(s_1), \dots, \rho(s_k))\|_\infty \leq c \cdot k^{-r}.$$

As before, piecewise polynomial interpolation at equidistant points, now of degree $\max(r-1, 1)$, might be used for this purpose.

We employ $Q^r_k(\mathcal{M})$ with $\mathcal{M} = \mathcal{M}^{k,\delta,L_0,L_1}_{N_{L_0},\dots,N_{L_1}}$ as a randomized algorithm for the approximation of ρ on $[S_0, S_1]$, and the error of $Q^r_k(\mathcal{M})$ is defined by

$$\mathrm{error}(Q^r_k(\mathcal{M})) = \left(\mathrm{E}\, \|\rho - Q^r_k(\mathcal{M})\|^2_\infty\right)^{1/2}.$$

Clearly the error does not increase if we replace $Q^r_k(x)$ by $\max(Q^r_k(x), 0)$.

Recall the definition of q from (9).

Theorem 2. *The following order, with* $\eta = 1$*, is achieved by algorithms* $Q_k^r(\mathcal{M}^{k,\delta,L_0,L_1}_{N_{L_0},\dots,N_{L_1}})$ *with suitably chosen parameters:*

$$q \leq \max(1, \beta_4/\beta_5) \quad \Rightarrow \quad \gamma = 2 + \frac{\max(1,q)+2}{r}, \tag{34}$$

$$q > \max(1, \beta_4/\beta_5) \ \wedge \ \beta_5 > 1 \quad \Rightarrow \quad \gamma = 2 + \frac{\max(1,\beta_4/\beta_5)+2}{r}, \tag{35}$$

$$q > 1 > \beta_4 \ \wedge \ \beta_5 = 1 \quad \Rightarrow \quad \gamma = 2 + \frac{3}{r}, \tag{36}$$

$$q > \max(1, \beta_4/\beta_5) \ \wedge \ \beta_5 < 1 \quad \Rightarrow \quad \gamma = 2 + \frac{\max(1,\beta_4 + (1-\beta_5)\cdot q)+2}{r}. \tag{37}$$

Moreover, with $\eta = 3$*,*

$$q > \beta_4 \geq 1 \ \wedge \ \beta_5 = 1 \quad \Rightarrow \quad \gamma = 2 + \frac{\beta_4 + 2}{r}. \tag{38}$$

Proof. We mimic the proof of Theorem 1. We use (A3), (E2) and (E4), Lemma 3 and Lemma 4, and the boundedness of g to obtain

$$\begin{aligned} \operatorname{error}^2(Q_k^r(\mathcal{M})) \preceq k^{-2r} + \delta^{2r} + 1/\delta^2 \cdot \min\left(\delta^{-2\alpha_1} \cdot M^{-L_1 \cdot 2\alpha_2}, M^{-L_1 \cdot 2\alpha_3}\right) \\ + \log k/\delta^2 \cdot \left(\frac{1}{N_{L_0}} + \sum_{\ell=L_0+1}^{L_1} \frac{\min(\delta^{-\beta_4} \cdot M^{-\ell \cdot \beta_5}, 1)}{N_\ell} \right), \end{aligned} \tag{39}$$

where $\mathcal{M} = \mathcal{M}^{k,\delta,L_0,L_1}_{N_{L_0},\dots,N_{L_1}}$. The upper bound (16) for the computational cost is also valid in the present case. We minimize (16), subject to the constraint that the upper bound (39) for the squared error is at most ϵ^2, up to multiplicative constants for both quantities.

Put

$$\tilde{\epsilon} = \epsilon^{1+1/r}.$$

First of all we choose

$$\delta = \epsilon^{1/r} = \tilde{\epsilon}^{1/(r+1)} \tag{40}$$

and, up to integer rounding,

$$k = \epsilon^{-1/r} = \tilde{\epsilon}^{-1/(r+1)} \tag{41}$$

and

$$N_{L_0} = \epsilon^{-2-2/r} \cdot \log_M \epsilon^{-1} \asymp \tilde{\epsilon}^{-2} \cdot \log_M \tilde{\epsilon}^{-1}. \tag{42}$$

This yields

$$\operatorname{error}^2(Q_k^r(\mathcal{M})) \preceq \epsilon^2 + 1/\delta^2 \cdot \left(a^2(L_1) + \log \tilde{\epsilon}^{-1} \cdot \sum_{\ell=L_0+1}^{L_1} \frac{\min(\delta^{-\beta_4} \cdot M^{-\ell \cdot \beta_5}, 1)}{N_\ell} \right),$$

where $a(L_1)$ is given by (22). Furthermore,

$$L^* = \frac{1}{r} \cdot \log_M \epsilon^{-1} = \frac{1}{r+1} \cdot \log_M \tilde{\epsilon}^{-1}, \tag{43}$$

see (18), and it suffices to study $L_0 \geq L^*$.

Since $\delta \cdot \epsilon = \tilde{\epsilon}$, the proof of Theorem 1 is applicable with ϵ being replaced by $\tilde{\epsilon}$. We obtain the same values for η, but γ must be replaced by $\gamma \cdot (1 + 1/r)$. $\square$

Remark 6. The following comments on optimality etc. are meant in the sense of Remark 3. We have a superiority of proper multi-level algorithms over single-level algorithms if and only if (32) holds true. Moreover, the optimal values of δ, k, and N_{L_0}, and L_1 are given by (40), (41), (42), and

$$L_1 = \frac{\max(1,q)}{r} \cdot \log_M \epsilon^{-1},$$

see (25). In particular, this completes the optimization of the parameters of single-level algorithms, where $L_0 = L_1$.

Suppose that $q > \max(1, \beta_4/\beta_5)$, so that we consider proper multi-level algorithms. The optimal value of L_0 is given by

$$L_0 = \frac{\max(1,\beta_4/\beta_5)}{r} \cdot \log_M \epsilon^{-1},$$

see (30), The optimality of

$$N_\ell = \epsilon^{-2-(\beta_4+2)/r} \cdot \log \epsilon^{-1} \cdot M^{-\ell\cdot(1+\beta_5)/2} \cdot \begin{cases} L^*, & \text{if } \beta_5 = 1, \\ M^{L^*\cdot \max(1,\beta_4/\beta_5)\cdot(1-\beta_5)/2} & \text{if } \beta_5 > 1, \\ M^{L^*\cdot q \cdot (1-\beta_5)/2}, & \text{if } \beta_5 < 1. \end{cases}$$

for $\ell = L_0 + 1, \dots, L_1$, with L^* given by (43), is derived from (28) in a straightforward way.

4. Approximation of Distribution Functions at a Single Point

Now we study the approximation of the distribution function F of Y at a single fixed point $s \in [S_0, S_1]$.

4.1. **Smoothing.** We employ assumption (A1) and the smoothing approach from Section 2.1, which involves the assumptions (S1)–(S4). In particular, we make use of Lemma 1.

4.2. **Assumptions on Weak and Strong Convergence.** We consider the setting from Section 2.2, and we assume (A2) and (A3) while, instead of (A4), the following property is assumed to hold with a constant $c > 0$:

(A5) There exist constants $\beta_1 \geq 0$ and $\beta_2 > \beta_3 \geq 0$ such that the strong error estimate

$$\sup_{s\in[S_0,S_1]} \mathrm{E}\left(g((Y-s)/\delta) - g((Y^{(\ell)}-s)/\delta)\right)^2 \leq c \cdot \min\left(\delta^{-\beta_1} \cdot M^{-\ell\cdot\beta_2}, M^{-\ell\cdot\beta_3}\right)$$

holds for all $\delta \in\,]0, \delta_0]$ and $\ell \in \mathbb{N}_0$.

See Section 5 for specific applications and approximations $Y^{(\ell)}$ with corresponding values of the parameters β_i.

We use different assumptions on the strong error for approximation of F on compact intervals and at a single point, namely (A4) with Lemma 3 as an immediate consequence in the first case and (A5) in the second case. Clearly, (A4) implies (A5) for every bounded and Lipschitz continuous function g with

$$\beta_1 = \beta_4, \qquad \beta_2 = \beta_5, \qquad \beta_3 = 0, \tag{44}$$

which is used in Section 5.3, but better values of β_1, β_2, and β_3 may be available. See Section 5 for examples where $\beta_1 < \beta_4$ and $\beta_3 > 0$. Note that (A5) corresponds directly to the weak error estimate (A3), and it yields the latter for every bounded and measurable function g with

$$\alpha_i = \beta_i/2 \tag{45}$$

for $i = 1, 2, 3$. See Section 5 for applications.

Strong error estimates for $Y - Y^{(\ell)}$ or $1_{]-\infty,s]}(Y) - 1_{]-\infty,s]}(Y^{(\ell)})$ may be used to establish (A5) and (A3). From the Lipschitz continuity of g we immediately get (A5) with $\beta_1 = 2$ and $\beta_3 = 0$, while the value of β_2 is determined by the asymptotic behavior of $\|Y - Y^{(\ell)}\|_2^2$. A refined analysis, which merely requires Y to have a bounded density, yields the following results, which are applicable under the assumptions (S2) and (S3) or (S2) and (S5) on g.

Lemma 5 (Avikainen (2009))**.** *There exists a constant $c > 0$ such that*

$$\sup_{s \in [S_0, S_1]} \|g((Y - s)/\delta) - g((Y^{(\ell)} - s)/\delta)\|_q^q \leq c^q \cdot \sup_{s \in [S_0 - \delta_0, S_1 + \delta_0]} \|1_{]-\infty,s]}(Y) - 1_{]-\infty,s]}(Y^{(\ell)})\|_1$$

and

$$\sup_{s \in [S_0 - \delta, S_1 + \delta]} \|1_{]-\infty,s]}(Y) - 1_{]-\infty,s]}(Y^{(\ell)})\|_1 \leq c \cdot \|Y - Y^{(\ell)}\|_p^{p/(p+1)}$$

holds for all $p, q \geq 1$, $\delta \in]0, \delta_0]$, and $\ell \in \mathbb{N}_0$.

Proof. See Avikainen (2009, p. 387) for the proof of the first estimate and Avikainen (2009, Lemma 3.4) for the second estimate. ☐

Lemma 6. *For every $1 \leq q \leq p < \infty$ there exists a constant $c > 0$ such that*

$$\sup_{s \in [S_0, S_1]} \|g((Y - s)/\delta) - g((Y^{(\ell)} - s)/\delta)\|_q^q \leq c \cdot \delta^{1-q-q/p} \cdot \|Y - Y^{(\ell)}\|_p^q$$

holds for all $\delta \in]0, \delta_0/2]$ and $\ell \in \mathbb{N}_0$.

Proof. Put

$$\Delta = g((Y - s)/\delta) - g((Y^{(\ell)} - s)/\delta).$$

In the sequel, we adopt the notation $\preceq$ from the proof of Theorem 1, where now the hidden constant must not depend on δ, ℓ or s.

Because of assumption (A1), the density ρ of Y is bounded on $[S_0 - \delta_0, S_1 + \delta_0]$. By Lemma 5,

$$\mathrm{E}\,\Delta^q \preceq \|Y - Y^{(\ell)}\|_p^{p/(p+1)},$$

so all that remains is to establish

$$\mathrm{E}\,\Delta^q \preceq \delta^{1-q-q/p} \cdot \|Y - Y^{(\ell)}\|_p^q$$

in the case $\delta^{1-q-q/p} \cdot \|Y - Y^{(\ell)}\|_p^q \leq \|Y - Y^{(\ell)}\|_p^{p/(p+1)}$, i.e., for

$$\|Y - Y^{(\ell)}\|_p \leq \delta^{1+1/p}. \tag{46}$$

If $|Y - s| > 2\delta$ and $|Y - Y^{(\ell)}| < \delta$, then $|Y^{(\ell)} - s| > \delta$ and hence $\Delta = 0$ follows, since g is constant on $]-\infty, -1[$ as well as on $]1, \infty[$. Accordingly, we consider

$$\begin{aligned} A_1 &= \{|Y - s| \leq 2\delta\}, \\ A_2 &= \{|Y - s| > 2\delta\} \cap \left\{|Y - Y^{(\ell)}| \geq \delta\right\}, \\ A_3 &= \{|Y - s| > 2\delta\} \cap \left\{|Y - Y^{(\ell)}| < \delta\right\}, \end{aligned}$$

and we then have

$$\mathrm{E}\,\Delta^q = \mathrm{E}(\Delta^q \cdot 1_{A_1}) + \mathrm{E}(\Delta^q \cdot 1_{A_2}).$$

Provided that $p_1 = P(A_1) > 0$, Jensen's inequality and the Lipschitz continuity of g give

$$\mathrm{E}(\Delta^q \,|\, A_1) \leq (\mathrm{E}(\Delta^p \,|\, A_1))^{q/p} \preceq \delta^{-q} p_1^{-q/p} \cdot \|Y - Y^{(\ell)}\|_p^q.$$

Hence, using the boundedness of the density of Y,

$$\mathrm{E}(\Delta^q \cdot 1_{A_1}) \preceq \delta^{-q} p_1^{1-q/p} \cdot \|Y - Y^{(\ell)}\|_p^q \preceq \delta^{1-q-q/p} \cdot \|Y - Y^{(\ell)}\|_p^q.$$

Turning now to A_2, Markov's inequality gives

$$P(\{|Y - Y^{(\ell)}| \geq \delta\}) \leq \delta^{-p} \cdot \|Y - Y^{(\ell)}\|_p^p,$$

and hence, using the boundedness of g,

$$\mathrm{E}(\Delta^q \cdot 1_{A_2}) \preceq \delta^{-p} \cdot \|Y - Y^{(\ell)}\|_p^p \leq \delta^{1-q-q/p} \cdot \|Y - Y^{(\ell)}\|_p^q,$$

with the last step coming from (46). ☐

If $\|Y - Y^{(\ell)}\|_p$ and $\|Y - Y^{(\ell)}\|_1$ are asymptotically equivalent for every $1 \leq p < \infty$, then Lemma 5 and Lemma 6 should be applied with large values of p, and this yields (A5) with β_1 arbitrarily close to 1 and (A3) with α_1 arbitrarily close to 0. See Sections 5.1 and 5.2 for examples.

4.3. The Multi-level Algorithm.

We study multi-level algorithms

$$\mathcal{M}^{\delta, L_0, L_1}_{N_{L_0}, \dots, N_{L_1}} = \frac{1}{N_{L_0}} \cdot \sum_{i=1}^{N_{L_0}} g^\delta(Y_i^{(L_0)}) + \sum_{\ell=L_0+1}^{L_1} \frac{1}{N_\ell} \cdot \sum_{i=1}^{N_\ell} \Big(g^\delta(Y_i^{(\ell)}) - g^\delta(Z_i^{(\ell)}) \Big)$$

with

$$g^\delta(t) = g((t-s)/\delta), \qquad t \in \mathbb{R},$$

which form a particular instance of (5). The error of $\mathcal{M} = \mathcal{M}^{\delta, L_0, L_1}_{N_{L_0}, \dots, N_{L_1}}$ is defined by

$$\mathrm{error}(\mathcal{M}) = \left(\mathrm{E}\, |F(s) - \mathcal{M}|^2\right)^{1/2},$$

and Remark 2 applies to single-level algorithms.

Put

$$\beta^\dagger = \frac{\beta_1}{\beta_2 - \beta_3},$$

and recall the definition of q from (9).

Theorem 3. *The following order, with $\eta = 0$, is achieved by algorithms $\mathcal{M}^{\delta, L_0, L_1}_{N_{L_0}, \dots, N_{L_1}}$ with suitably chosen parameters:*

$$q \leq \beta^\dagger \;\wedge\; \beta_3 \neq 1 \quad \Rightarrow \quad \gamma = 2 + \frac{(1-\beta_3)_+ \cdot q}{r+1}, \tag{47}$$

$$q > \beta^\dagger \;\wedge\; \beta_3 \neq 1 \;\wedge\; \beta_2 > 1 \quad \Rightarrow \quad \gamma = 2 + \frac{(1-\beta_3)_+ \cdot \beta^\dagger}{r+1}, \tag{48}$$

$$q > \beta^\dagger \;\wedge\; \beta_2 < 1 \quad \Rightarrow \quad \gamma = 2 + \frac{\beta_1 + (1-\beta_2) \cdot q}{r+1}. \tag{49}$$

Moreover, with $\eta = 2$,

$$\beta_3 = 1 \quad \Rightarrow \quad \gamma = 2, \tag{50}$$

$$q > \beta^\dagger \;\wedge\; \beta_2 = 1 \quad \Rightarrow \quad \gamma = 2 + \frac{\beta_1}{r+1}. \tag{51}$$

Proof. We proceed analogously to the proof of Theorem 1. Use Lemma 1, the assumptions (A3) and (A5), and the boundedness of g to obtain

$$\begin{aligned}\text{error}^2(\mathcal{M}) \preceq \delta^{2(r+1)} &+ \min\left(\delta^{-2\alpha_1} \cdot M^{-L_1 \cdot 2\alpha_2}, M^{-L_1 \cdot 2\alpha_3}\right) \\ &+ \frac{1}{N_{L_0}} + \sum_{\ell=L_0+1}^{L_1} \frac{\min\left(\delta^{-\beta_1} \cdot M^{-\ell \cdot \beta_2}, M^{-\ell \cdot \beta_3}\right)}{N_\ell \cdot \delta^2}\end{aligned} \tag{52}$$

for $\mathcal{M} = \mathcal{M}^{\delta, L_0, L_1}_{N_{L_0}, \dots, N_{L_1}}$. Furthermore, by (S1) and (A2),

$$\text{cost}(\mathcal{M}) \preceq c(L_0, L_1, N_{L_0}, \dots, N_{L_1}) \tag{53}$$

with

$$c(L_0, L_1, N_{L_0}, \dots, N_{L_1}) = \sum_{\ell=L_0}^{L_1} N_\ell \cdot M^\ell.$$

We minimize the upper bound (53) for the cost, subject to the constraint that the upper bound (52) for the squared error is at most ϵ^2, up to multiplicative constants for both quantities.

To this end we choose δ according to (19), and, up to integer rounding,

$$N_{L_0} = \epsilon^{-2} \tag{54}$$

as well as

$$L_1 = q \cdot L^* \tag{55}$$

with L^* given by (23).

For a single-level algorithm, i.e., $L_0 = L_1$, this yields $\text{error}(\mathcal{M}) \preceq \epsilon$ and

$$c(L_1, L_1, N_{L_1}) \asymp \epsilon^{-2-q/(r+1)}. \tag{56}$$

For a proper multi-level algorithm, i.e., $L_0 < L_1$, we obtain

$$\text{error}^2(\mathcal{M}) \preceq \epsilon^2 + \sum_{\ell=L_0+1}^{L_1} \frac{v_\ell}{N_\ell}$$

with

$$v_\ell = \min\left(M^{L^* \cdot \beta_1} \cdot M^{-\ell \cdot \beta_2}, M^{-\ell \cdot \beta_3}\right)$$

as well as

$$c(L_0, L_1, N_{L_0}, \dots, N_{L_1}) \asymp \epsilon^{-2} \cdot M^{L_0} + \sum_{\ell=L_0+1}^{L_1} N_\ell \cdot M^\ell.$$

Fix L_0 for the moment. We minimize

$$h(L_0, N_{L_0+1}, \dots, N_{L_1}) = \epsilon^{-2} \cdot M^{L_0} + \sum_{\ell=L_0+1}^{L_1} N_\ell \cdot M^\ell$$

subject to

$$\sum_{\ell=L_0+1}^{L_1} \frac{v_\ell}{N_\ell} \le \epsilon^2.$$

A Lagrange multiplier leads to

$$N_\ell = \epsilon^{-2} \cdot G(L_0) \cdot \left(v_\ell \cdot M^{-\ell}\right)^{1/2}, \tag{57}$$

up to integer rounding, which satisfies the constraint with

$$G(L_0) = \sum_{\ell=L_0+1}^{L_1} \left(v_\ell \cdot M^\ell\right)^{1/2} = \sum_{\ell=L_0+1}^{L_1} \left(\min(M^{L^*\cdot\beta_1} \cdot M^{-\ell\cdot\beta_2}, M^{-\ell\cdot\beta_3}) \cdot M^\ell\right)^{1/2}.$$

Moreover, this choice of $N_{L_0+1}, \dots, N_{L_1}$ yields

$$h(L_0, N_{L_0+1}, \dots, N_{L_1}) = \epsilon^{-2} \cdot \left(M^{L_0} + G^2(L_0)\right).$$

Put

$$L^\dagger = \beta^\dagger \cdot L^*.$$

In the case $q \leq \beta^\dagger$ we have $L_1 \leq L^\dagger$, and therefore

$$M^{L_0} + G^2(L_0) = M^{L_0} + \left(\sum_{\ell=L_0+1}^{L_1} M^{\ell\cdot(1-\beta_3)/2}\right)^2.$$

In the case $q > \beta^\dagger$ we have $L^\dagger < L_1$, and therefore

$$M^{L_0} + G^2(L_0) = M^{L_0} + \left(\sum_{\ell=L_0+1}^{L^\dagger} M^{\ell\cdot(1-\beta_3)/2} + M^{L^*\cdot\beta_1/2} \cdot \sum_{\ell=L^\dagger+1}^{L_1} M^{\ell\cdot(1-\beta_2)/2}\right)^2.$$

Since

$$M^{L_0} + \left(\sum_{\ell=L_0+1}^{L} M^{\ell\cdot(1-\beta_3)/2}\right)^2 \asymp \begin{cases} M^{L_0}, & \text{if } \beta_3 > 1,\\ M^{L_0} + (L-L_0)^2, & \text{if } \beta_3 = 1,\\ M^{L_0} + M^{L\cdot(1-\beta_3)}, & \text{if } \beta_3 < 1, \end{cases}$$

for $L = L_1$ and $L = L^\dagger$, we take

$$L_0 = 0$$

in both cases.

This leads to

$$M^{L_0} + G^2(L_0) \asymp \begin{cases} 1, & \text{if } \beta_3 > 1,\\ L_1^2, & \text{if } \beta_3 = 1,\\ M^{L_1\cdot(1-\beta_3)}, & \text{if } \beta_3 < 1, \end{cases}$$

if $q \leq \beta^\dagger$. Moreover, it is straightforward to verify

$$M^{L_0} + G^2(L_0) \asymp \begin{cases} 1, & \text{if } \beta_3 > 1,\\ (L^\dagger)^2, & \text{if } \beta_3 = 1,\\ M^{L^\dagger\cdot(1-\beta_3)}, & \text{if } \beta_3 < 1 \text{ and } \beta_2 > 1,\\ M^{L^*\cdot\beta_1} \cdot (L_1 - L^\dagger)^2, & \text{if } \beta_2 = 1,\\ M^{L^*\cdot(\beta_1+q(1-\beta_2))}, & \text{if } \beta_2 < 1, \end{cases}$$

if $q > \beta^\dagger$. Except for the case $\beta_3 = 0$ and $q \leq \beta^\dagger$ these estimates are superior to M^{L_1}, which corresponds to (56). □

Remark 7. The following comments on optimality etc. are meant in the sense of Remark 3. The optimal values of δ, N_{L_0}, and L_1 are given by (19), (54), and (55), which completes the optimization of the parameters of single-level algorithms. For proper multilevel algorithms, $L_0 = 0$ is optimal, and the optimal replication numbers $N_{L_0+1}, \dots, N_{L_1}$ and L_0 can be easily derived from (57).

Proper multi-level algorithms are superior to single-level algorithms if and only if

$$\beta_3 \neq 0 \ \vee \ q > \beta_1/\beta_2.$$

In the case $\beta_3 = 0$ and $q \leq \beta_1/\beta_2$ the lack of superiority is caused by the negative impact of smoothing, which leads to variances of order one on all levels level $\ell = 0, \dots, L_1$.

Single-level algorithms with smoothing are superior to single-level algorithms without smoothing if and only if

$$\frac{r+1}{\alpha_3} > \frac{r+1+\alpha_1}{\alpha_2}. \tag{58}$$

5. Applications

At first we consider a general situation, where all we have at hand is (A1), (A2), and an upper bound on the order of the strong error of $Y - Y^{(\ell)}$, which does not depend on p. Specifically, we assume that there exists a constant

$$0 < \beta \leq 2$$

with the following property. For every $1 \leq p < \infty$ there exists a constant $c_p > 0$ such that

$$\|Y - Y^{(\ell)}\|_p \leq c_p \cdot M^{-\ell \cdot \beta/2} \tag{59}$$

for every $\ell \in \mathbb{N}$. In the sequel $\varepsilon > 0$ may be chosen arbitrarily small.

From (59) we obtain (A4) with

$$\beta_4 = 2, \qquad \beta_5 = \beta, \tag{60}$$

see (2), and Lemma 5 and Lemma 6 yield (A5) with

$$\beta_1 = 1 + \varepsilon, \qquad \beta_2 = \beta, \qquad \beta_3 = \beta/2 - \varepsilon \tag{61}$$

under the assumptions (S2) and (S3) or (S2) and (S5). Using Lemma 5 and Lemma 6 again we get (A3) under both sets of assumptions on g with

$$\alpha_1 = \varepsilon, \qquad \alpha_2 = \beta/2, \qquad \alpha_3 = \beta/2 - \varepsilon, \tag{62}$$

and (6) holds with

$$\alpha = \beta/2 - \varepsilon. \tag{63}$$

It follows that

$$q = \frac{2 \cdot (r+1)}{\beta} + \varepsilon$$

and

$$\max(1, \beta_4/\beta_5) = 2/\beta,$$

so that (11), (13), and (14) in Theorem 1 yield

$$1 \leq \beta \leq 2 \qquad \Rightarrow \qquad \gamma = 2 + \frac{2}{\beta \cdot (r+1)}, \tag{64}$$

$$0 < \beta < 1 \qquad \Rightarrow \qquad \gamma = \frac{2}{\beta} + \frac{2}{r+1} + \varepsilon \tag{65}$$

for the approximation of F on $[S_0, S_1]$. Likewise, (35), (37), and (38) in Theorem 2 yield

$$1 \le \beta \le 2 \quad \Rightarrow \quad \gamma = 2 + \frac{2 \cdot (1+\beta)}{\beta \cdot r}, \tag{66}$$

$$0 < \beta < 1 \quad \Rightarrow \quad \gamma = \frac{2}{\beta} + \frac{2 \cdot (1+\beta)}{\beta \cdot r} + \varepsilon, \tag{67}$$

for the approximation of ρ on $[S_0, S_1]$. Moreover,

$$\beta^\dagger = 2/\beta + \varepsilon,$$

so that (48), (49), and (51) in Theorem 3 yield

$$1 \le \beta \le 2 \quad \Rightarrow \quad \gamma = 2 + \frac{2-\beta}{\beta \cdot (r+1)} + \varepsilon, \tag{68}$$

$$0 < \beta < 1 \quad \Rightarrow \quad \gamma = \frac{2}{\beta} + \frac{1}{r+1} + \varepsilon \tag{69}$$

for the approximation of F at a single point $s \in [S_0, S_1]$. For all three problems we get $\gamma = \max(2, 2/\beta)$ in the limit $r \to \infty$, and proper multi-level algorithms are always superior to single-level algorithms, see Remarks 3, 6, and 7.

Remark 8. We compare the smoothing approach for the approximation of F at a single point with a direct approach, which is due to Avikainen (2009) and which only requires that Y has a bounded density ρ, see Lemma 5.

We study multi-level algorithms

$$\mathcal{M}^{L_0,L_1}_{N_{L_0},\dots,N_{L_1}} = \frac{1}{N_{L_0}} \cdot \sum_{i=1}^{N_{L_0}} 1_{]-\infty,s]}(Y_i^{(L_0)}) + \sum_{\ell=L_0+1}^{L_1} \frac{1}{N_\ell} \cdot \sum_{i=1}^{N_\ell} \left(1_{]-\infty,s]}(Y_i^{(\ell)}) - 1_{]-\infty,s]}(Z_i^{(\ell)}) \right)$$

for the approximation of $F(s)$. As previously, we assume that (59) with $0 < \beta \le 2$ is all we have at hand. The analysis from Theorem 3 directly applies, if we take

$$\beta_1 = 0, \qquad \beta_2 = \beta/2 - \varepsilon, \qquad \beta_3 = \beta/2 - \varepsilon,$$

and

$$\alpha_1 = 0, \qquad \alpha_2 = \beta/2 - \varepsilon, \qquad \alpha_3 = \beta/2 - \varepsilon.$$

We achieve the order (γ', η') with

$$\gamma' = \frac{2+\beta}{\beta} + \varepsilon,$$

so that the smoothing approach is superior to the direct approach iff $\beta < 2$ and $r \ge 1$.

In the sequel we consider three specific settings in the context of stochastic differential equations (SDEs). We let X denote the solution process of the SDE, which is supposed to take values in $\mathbb{R}^d$. For simplicity, we alway take the Euler scheme with equidistant time-steps for approximation of X, and we do not discuss results on the existence and smoothness of densities. As previously, $\varepsilon > 0$ may be chosen arbitrarily small.

5.1. **Smooth Path-independent Functionals for SDEs.** Let

$$Y = \varphi(X_T),$$

where $\varphi : \mathbb{R}^d \to \mathbb{R}$ is Lipschitz continuous. We assume that the cost of computing $\varphi(x)$ is uniformly bounded for $x \in \mathbb{R}^d$, and for approximation of Y we use $Y^{(\ell)} = \varphi(X_T^{(\ell)})$, where $X^{(\ell)}$ denotes the Euler scheme with 2^ℓ equidistant time-steps. Obviously, (A2) holds with $M = 2$. For weak error estimates we refer to Bally, Talay (1996a). Hereby we obtain (A3) with

$$\alpha_1 = 0, \qquad \alpha_2 = 1, \qquad \alpha_3 = 1 \tag{70}$$

under the assumptions (S2) and (S3) or (S2) and (S5) on g and the smoothness and non-degeneracy assumptions (C) and (UH) on the coefficients of the SDE. Furthermore, (6) holds with

$$\alpha = 1.$$

It is well-known that (59) holds with

$$\beta = 1$$

already under standard assumptions on the coefficients of the SDE. Hence we get (A4) with

$$\beta_4 = 2, \qquad \beta_5 = 1, \tag{71}$$

see (60), and (A5) with

$$\beta_1 = 1 + \varepsilon, \qquad \beta_2 = 1, \qquad \beta_3 = 1/2 - \varepsilon,$$

see (61).

We therefore have $q = r + 1$ and $\max(1, \beta_4/\beta_5) = 2$, and (10) and (14) in Theorem 1 yield

$$(\gamma, \eta) = \begin{cases} (3, 1), & \text{if } r \leq 1, \\ (2 + 2/(r+1), 3), & \text{if } r \geq 2, \end{cases}$$

for the approximation of F on $[S_0, S_1]$. Likewise, (34) and (38) in Theorem 2 yield

$$(\gamma, \eta) = \begin{cases} (6, 1), & \text{if } r = 1, \\ (2 + 4/r, 3), & \text{if } r \geq 2, \end{cases}$$

for the approximation of ρ on $[S_0, S_1]$. For both problems, proper multi-level algorithms are superior to single-level algorithms if and only if $r \geq 2$, see Remarks 3 and 6. Moreover, $\beta^\dagger = 2 + \varepsilon$, so that (47) and (51) in Theorem 3 yield

$$\gamma = \begin{cases} 5/2 + \varepsilon, & \text{if } r = 0, \\ 2 + 1/(r+1) + \varepsilon, & \text{if } r \geq 1, \end{cases}$$

for the approximation of F at a single point $s \in [S_0, S_1]$. For this problem, proper multi-level algorithms are superior to single-level algorithms for every $r \in \mathbb{N}_0$, see Remark 7. For all three problems we get $\gamma = 2$ in the limit $r \to \infty$.

If the coefficients of the SDE merely satisfy the standard assumptions, instead of (C) and (UH) from Bally, Talay (1996a), we may apply (62) to obtain $\alpha_1 = \varepsilon$, $\alpha_2 = 1/2$, and $\alpha_3 = 1/2 - \varepsilon$, see also Kebaier (2005, Sec. 2.2). While the latter is inferior to (70), it leads to essentially the same orders of convergence for approximation of densities or distribution functions if $r \geq 1$, see (64), (66), and (68).

Remark 9. A two-level construction of the form

$$\mathcal{M}^{\delta,L_0,L_1}_{N_{L_0},N_{L_1}} = \frac{1}{N_{L_0}} \cdot \sum_{i=1}^{N_{L_0}} g^\delta(Y_i^{(L_0)}) + \frac{1}{N_{L_1}} \cdot \sum_{i=1}^{N_{L_1}} \Big(g^\delta(Y_i^{(L_1)}) - g^\delta(Z_i^{(L_1)}) \Big),$$

which is the counterpart of the two-level construction from Kebaier (2005) for the approximation of $\mathrm{E}(\varphi(X_T))$, is employed in Kebaier, Kohatsu-Higa (2008) for the approximation of the density ρ of $Y = X_T$ at a single point s. Here the sequence $(Y^{(\ell)})_{\ell\in\mathbb{N}}$ consists of suitably regularized Euler schemes with ℓ equidistant time-steps. By assumption, $\rho \in C_{\mathrm{b}}^\infty(\mathbb{R}^d, \mathbb{R})$, i.e., the multi-dimensional counterpart to (A1) is satisfied for every $r \in \mathbb{N}_0$. Using Malliavin calculus techniques, the authors derive a central limit theorem for $L_1 \cdot (\mathcal{M}^{\delta,L_0,L_1}_{N_{L_0},N_{L_1}} - \rho(x))$ with properly chosen parameters L_0, N_{L_0}, N_{L_1}, and δ as L_1 tends to infinity. For every dimension d the order $\gamma = 5/2 + \varepsilon$ is achieved in this way, while the multi-level approach achieves the order $\gamma = 2 + \varepsilon$ (at least for $d = 1$).

Remark 10. Consider the problem of approximating a quantile of Y, which is studied in Talay, Zheng (2004) in the particular case of a projection $\varphi(x) = x_i$. By assumption, $\rho \in C_{\mathrm{b}}^\infty(\mathbb{R}, \mathbb{R})$. The authors employ a single-level algorithm that is based on a suitably regularized Euler scheme, cf. Remark 9. The approximation to the quantile is given as the corresponding empirical quantile, and an error of order $\gamma = 3$ is achieved, if ρ is bounded away from zero in a neighborhood of the quantile.

Under the latter assumption, the order of approximation to F in the supremum norm and to the quantile coincide, and given (A1) for every $r \in \mathbb{N}_0$ we expect our multi-level algorithm to achieve the order $\gamma = 2 + \varepsilon$ also for quantile approximation and every Lipschitz continuous function φ. Furthermore, the multi-level algorithm may be used to approximate the distribution function F and the density ρ in parallel, which allows to control the impact of inverting the approximation to F.

Remark 11. We comment on the optimality of the parameters α_i and β_i according to (70) and (71) in (A3) and (A4). Due to Bally, Talay (1996a), the estimate (A3) with (70) is sharp under the assumptions (C) and (UH). Under standard assumptions, $2^{\ell/2} \cdot (X - X^{(\ell)})$ converges in distribution to a stochastic process U with U_T being non-degenerate in general, see Jacod, Protter (1998). In the latter case we have a projection $\varphi(x) = x_i$ such that (59) with $M = 2$ and $p = 1$ does not hold for any $\beta > 1$. A slight generalization of Lemma 2 shows that (A4) does not hold for any $\beta_4 < 2$ or $\beta_5 > 1$. Hence the estimate (A4) with (71) cannot be improved in general for the Euler scheme.

The approximation of marginal densities of SDE in studied in a number of papers under different aspects. The convergence rate of the density of the Euler approximation $X_T^{(\ell)}$ towards ρ is studied in, e.g., Bally, Talay (1996b) and Gobet, Labart (2008). Milstein, Schoenmakers, Spokoiny (2004) construct a forward-reverse kernel estimator and provide an upper bound for its variance.

5.2. **Smooth Path-dependent Functionals for SDEs.** Let

$$Y = \varphi(X)$$

with $\varphi : C([0,T], \mathbb{R}^d) \to \mathbb{R}$ being Lipschitz continuous. We assume that the cost of computing $\varphi(x)$ for a piecewise linear path $x \in C([0,T], \mathbb{R}^d)$ with m breakpoints is bounded by a constant times m, and for approximation of Y we use $Y^{(\ell)} = \varphi(X^{(\ell)})$, where $X^{(\ell)}$

denotes the Euler scheme with 2^ℓ equidistant time-steps and piecewise linear interpolation. Then (A2) holds with $M = 2$, and the following fact is well-known under standard assumptions on the coefficients of the SDE. For every $1 \leq p < \infty$ there exists a constant $c_p > 0$ such that

$$\|Y - Y^{(\ell)}\|_p \leq c_p \cdot \left(\ell \cdot M^{-\ell}\right)^{1/2}$$

for every $\ell \in \mathbb{N}$. Consequently, (59) holds with

$$\beta = 1 - \varepsilon,$$

and we get (A4) with

$$\beta_4 = 2, \qquad \beta_5 = 1 - \varepsilon \tag{72}$$

see (60), (A5) with

$$\beta_1 = 1 + \varepsilon, \qquad \beta_2 = 1 - \varepsilon, \qquad \beta_3 = 1/2 - \varepsilon$$

under the assumptions (S2) and (S3) or (S2) and (S5), see (61), as well as (A3) with

$$\alpha_1 = 0, \qquad \alpha_2 = 1/2 - \varepsilon, \qquad \alpha_3 = 1/2 - \varepsilon, \tag{73}$$

see (62). Furthermore, (6) holds with

$$\alpha = 1/2 - \varepsilon,$$

see (63).

We therefore have $q = 2 \cdot (r + 1) + \varepsilon$ and $\max(1, \beta_4/\beta_5) = 2 + \varepsilon$, and (13) in Theorem 1 yields

$$\gamma = 2 + 2/(r + 1) + \varepsilon$$

for the approximation of F on $[S_0, S_1]$. Likewise, (37) in Theorem 2 yields

$$\gamma = 2 + 4/r + \varepsilon$$

for the approximation of ρ on $[S_0, S_1]$. Moreover, $\beta^\dagger = 2 + \varepsilon$, so that (49) in Theorem 3 yields

$$\gamma = 2 + 1/(r + 1) + \varepsilon$$

for the approximation of F at a single point $s \in [S_0, S_1]$. For all three problems proper multi-level algorithms are always superior to single-level algorithms, see Remarks 3, 6, and 7.

Note that Section 5.1 is dealing with a particular instance of the functionals studied here. We achieve essentially the same order of convergence for the problems studied in Sections 5.1 and 5.2, if $r \geq 1$, and we always get $\gamma = 2$ in the limit $r \to \infty$.

Remark 12. We comment on the optimality of the parameters α_i and β_i according to (73) and (72) in (A3) and (A4). Due to Remark 11 the estimate (A4) with (72) cannot be improved in general for the Euler scheme. Concerning (A3) we are not aware of an optimality result. We refer, however, to Alfonsi, Jourdain, Kohatsu-Higa (2013), who study processes $Y^{(\ell)}$ that coincide with the Euler scheme $X^{(\ell)}$ at the discretization points, but instead of 2^ℓ Brownian increments the whole trajectory of the Brownian motion is employed. They provide an upper bound of the order $2/3 - \varepsilon$ for Wasserstein distance of X and $Y^{(\ell)}$ in the case $d = 1$.

5.3. **Stopped Exit Times for SDEs.** Consider a bounded domain $D \subset \mathbb{R}^d$ such that $X_0 \in D$, and let

$$Y = \varphi(X)$$

be the corresponding exit time, stopped at $T > 0$, i.e.,

$$\varphi(x) = \inf\{t \geq 0 : x(t) \in \partial D\} \wedge T$$

for $x \in C([0,T], \mathbb{R}^d)$. We assume that the cost of computing $\varphi(x)$ for a piecewise linear path $x \in C([0,T], \mathbb{R}^d)$ with m breakpoints is bounded by a constant times m, and as in the previous section $Y^{(\ell)}$ is the Euler scheme $X^{(\ell)}$ composed with φ. Then (A2) holds with $M = 2$. For every $1 \leq p < \infty$ there exists a constant $c_p > 0$ such that

$$\|Y - Y^{(\ell)}\|_p \leq c_p \cdot M^{-\ell/(2p)} \tag{74}$$

for every $\ell \in \mathbb{N}$, see Bouchard, Geiss, Gobet (2013). From (3) we get (A4) with

$$\beta_4 = 1, \qquad \beta_5 = 1/2,$$

and (44) and Lemma 5 yield (A5) with

$$\beta_1 = 1, \qquad \beta_2 = 1/2, \qquad \beta_3 = 1/4.$$

Furthermore, (1) and Lemma 5 yield (A3) with

$$\alpha_1 = 1, \qquad \alpha_2 = 1/2, \qquad \alpha_3 = 1/4$$

under the assumptions (S2) and (S3) or (S2) and (S5), while (6) holds with

$$\alpha = 1/4.$$

We therefore have $q = 2r + 4$ and $\max(1, \beta_4/\beta_5) = 2$, and (13) in Theorem 1 yields

$$(\gamma, \eta) = (3 + 2/(r+1), 1)$$

for the approximation of F on $[S_0, S_1]$. Likewise, (37) in Theorem 2 yields

$$(\gamma, \eta) = (3 + 5/r, 1)$$

for the approximation of ρ on $[S_0, S_1]$. Moreover, $\beta^\dagger = 3$, so that (49) in Theorem 3 yields

$$(\gamma, \eta) = (3 + 2/(r+1), 0) \tag{75}$$

for the approximation of F at a single point $s \in [S_0, S_1]$. For all three problems, proper multi-level algorithms are superior to single-level algorithms for every $r \in \mathbb{N}_0$, see Remarks 3, 6, and 7, but we only get $\gamma = 3$ in the limit $r \to \infty$. The latter is in contrast to the results from Sections 5.1 and 5.2, and it is basically due to the fact that the upper bound (74) for strong approximation of Y by $Y^{(\ell)}$ depends on p in the most unfavorable way. We add that numerical experiments suggest that the upper bound (74) cannot be improved, in general. Furthermore, observe that for stopped exit times the same order γ is achieved for the approximation of F on a compact interval and at a single point.

We add that (33) and (58) are satisfied for every $r \geq 1$, and therefore smoothing already help for the single-level algorithm to approximate the distribution function of the stopped exit time.

Remark 13. For the approximation of the mean $\mathrm{E}(Y)$ of the stopped exit time a multi-level Euler algorithm has been constructed and analyzed in Higham *et al.* (2013). It is shown that the order $\gamma = 3 + \varepsilon$ is achieved under standard smoothness assumptions on the coefficients of the SDE and on the domain D.

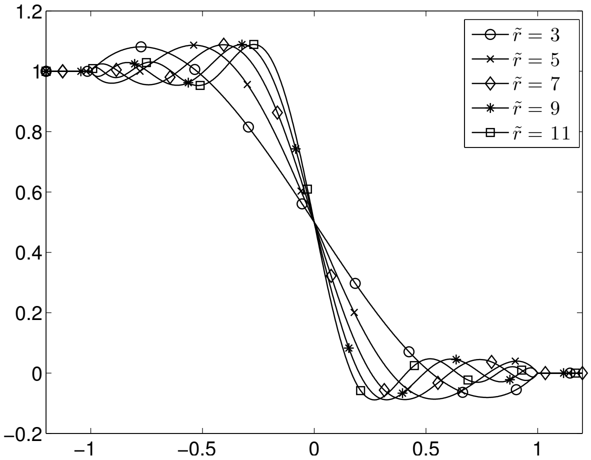

FIGURE 1. Smoothing polynomials g.

6. NUMERICAL EXPERIMENTS

The main goal of our numerical experiments is to demonstrate the potential of the new multi-level algorithm. We consider three benchmark problems according to Sections 5.1–5.3 for a simple, scalar SDE, where the solutions are known analytically. We present results only for the approximation of distribution functions on a compact interval $[S_0, S_1]$, as the main numerical difference to the other two problems studied in this paper is in the deterministic interpolation part. Our numerical experiments show the computational gain in terms of upper bounds, achieved by the multi-level Monte Carlo approach with smoothing in comparison to the single-level Monte Carlo approach without smoothing. Furthermore, we compare the error of the multi-level algorithm with the accuracy demand ϵ, which serves as an input to the algorithm. An extensive numerical study of our algorithm and the adaptive choice of its parameters is out of the scope of the current paper and will be presented in a subsequent paper.

Consider a geometric Brownian motion X, given by

$$\begin{aligned} dX_t &= \mu \cdot X_t \, dt + \sigma \cdot X_t \, dW_t, \qquad t \in [0,T], \\ X_0 &= 1, \end{aligned}$$

where W denotes a scalar Brownian motion. For the approximation of X we use the Euler scheme with equidistant time-steps, so that $M = 2$. The corresponding values of the parameters α_i and β_i are presented in Sections 5.1–5.3.

In the examples from this section, the assumption (A1) holds for every $r \in \mathbb{N}$, but typically we think of r being unknown. Hence we choose $\tilde{r} \in \mathbb{N}_0$, instead, and a particular purpose of the numerical experiments is to illustrate the impact of $\tilde{r}$. In all our experiments we take

$$\tilde{r} = 3,\ 5,\ 7,\ 9,\ 11,$$

and the corresponding smoothing polynomials g according to Remark 1 can be seen in Figure 1.

Given ϵ and $\tilde{r}$, we basically choose the remaining parameters of the multi-level (single-level) algorithm such that all four (three) terms in the upper bound (15) are of the order ϵ^2. For the multi-level algorithm with smoothing we choose the parameters L_0, L_1, and

N_ℓ according to (30), (25), (21), and (31), with r replaced by $\tilde{r}$, while

$$\delta = 2^{-1/(\tilde{r}+1)} \cdot \epsilon^{1/(\tilde{r}+1)},$$

cf. (19). For the single-level algorithm without smoothing, see Remark 2, we choose $L = L_0 = L_1$ and N_L according to (25) and (21), too, however observing (7), which leads to $q = (\tilde{r}+1)/\alpha$.

In the second stage of the algorithm we employ piecewise polynomial interpolation Q_k^3 of degree 3 with equidistant knots for any $\tilde{r}$. Due to the Lebesgue constants involved, this is preferable to $Q_k^{\tilde{r}}$ with a large value of $\tilde{r}$ if the overall number k of interpolation points is comparatively small. Furthermore, it is convenient if $k-1$ is a multiple of 3 and proportional to the length of the interval $[S_0, S_1]$. In both cases, single-level and multi-level, we therefore take

$$k = 3 \cdot \left\lceil 5 \cdot \epsilon^{-1/(\tilde{r}+1)} \cdot (S_1 - S_0)/3 \right\rceil + 1, \tag{76}$$

cf. (20).

To specify the computational gain we compare the upper bound (17) for the cost of the multi-level Monte Carlo algorithm with smoothing and the corresponding upper bound

$$c(k, L, N) = N \cdot (2^L + k),$$

for the cost of the single-level algorithm. The ratio $c(k, L_0, L_1, N_{L_0}, \dots, N_{L_1})/c(k, L, N)$, which is a function of the desired accuracy ϵ, is used to describe the computational gain.

To assess the accuracy of the multi-level algorithm, $\text{error}(Q_k^3(\mathcal{M}))$, which depends on ϵ and $\tilde{r}$, should be compared with the desired accuracy ϵ. Since $\text{error}(Q_k^3(\mathcal{M}))$ is not known exactly, we employ a simple Monte Carlo experiment with 25 independent replications for each of the values of $\tilde{r}$ and each of the values $\epsilon = 2^{-i}$ for $i = 3, \dots, 11$. The estimate is denoted by $\text{RMSE}(\epsilon, \tilde{r})$. In the present approach we do not have an exact control of the error of the multi-level (single-level) algorithm for a given ϵ, since the parameters of the algorithm are chosen on the basis of the asymptotic analysis from Section 2. Therefore we only aim at $\text{RMSE}(\epsilon, \tilde{r})$ being reasonably close to ϵ.

6.1. **Smooth Path-independent Functionals for SDEs.**

In this section we set

$$\mu = 0.05,\ \sigma = 0.2,\ T = 1,$$

and we approximate the distribution function $F(s) = \mathrm{E}(1_{]\infty,s]}(Y))$ of

$$Y = X_T$$

on the interval

$$[S_0, S_1] = [0, 2].$$

Note that Y is lognormally distributed with parameters $\mu - \sigma^2/2$ and σ^2.

The computational gain as well as the replication numbers N_ℓ for the multi-level algorithm with $\epsilon = 2^{-11}$ are presented in Figure 2. The maximal level L_1 of the multi-level algorithm coincides with the level chosen by the single-level algorithm, and this level does not depend on $\tilde{r}$. For smaller values of $\tilde{r}$ the multi-level algorithm start on a higher level L_0, and therefore the computational gain in the case $\tilde{r} = 3$ is only about a factor two. For large values of $\tilde{r}$ we observe a reasonable computational gain already for moderate values of ϵ. In Figure 3 we compare the estimate $\text{RMSE}(\epsilon, \tilde{r})$ for the error of the multi-level algorithm and the accuracy demand ϵ. Note that $\text{RMSE}(\epsilon, \tilde{r})$ is in the range of ϵ; actually, it is less that ϵ in almost all cases.

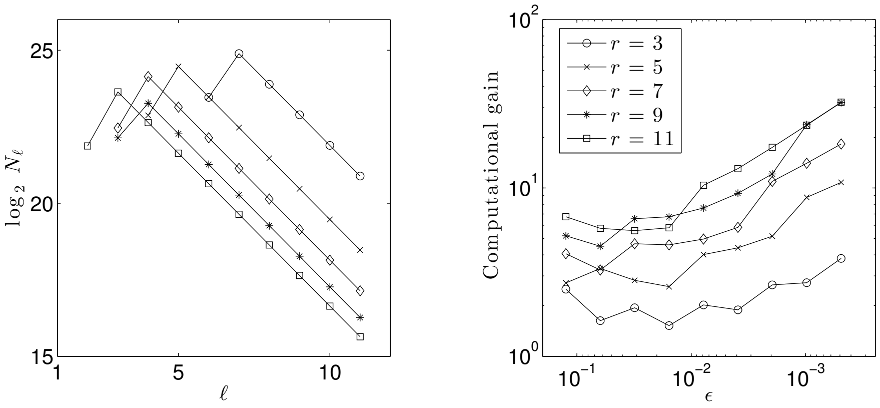

FIGURE 2. Path-independent functional: replication numbers (left) and computational gain (right).

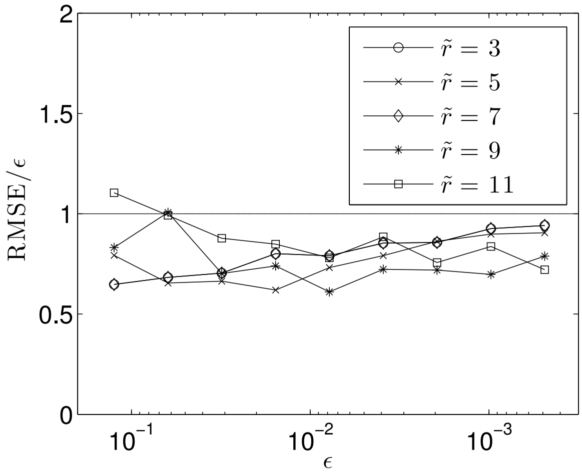

FIGURE 3. Path-independent functional: error vs. accuracy demand ϵ.

6.2. **Smooth Path-dependent Functionals for SDEs.** For this test case we use the same parameters for the SDE and the same interval $[S_0, S_1]$ as in Section 6.1. We approximate the function

$$F(s) = \mathrm{E}\left(e^{-\mu\cdot T} \cdot \max(X_T - X_0, 0) \cdot 1_{]-\infty,s]}(Y)\right),$$

where

$$Y = \max_{t\in[0,T]} X_t.$$

See Shreve (2008, p. 307) for the analytical solution. Note that this problem does not exactly fit into our framework, due the the presence of $\max(X_T - X_0, 0)$ in the definition of the functional. Still, the multi-level smoothing approach is applicable.

See Figures 4, with replication numbers for $\epsilon = 2^{-10}$, and 5 for the results. As the main difference, compared to the previous section, the computational gain is substantially larger for the path-dependent functional. This is due to the following facts. The orders of strong convergence are essentially the same for both problems. However, the maximal level, which once more coincide with the level chosen by the single-level algorithm, is essentially twice

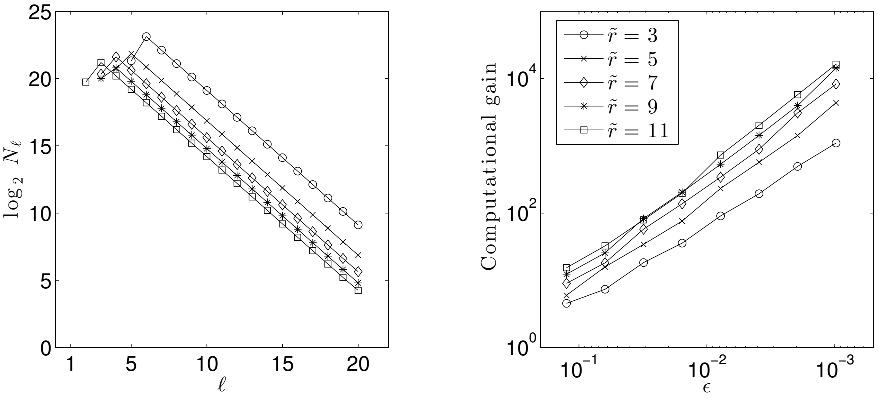

FIGURE 4. Path-dependent functional: replication numbers (left) and computational gain (right).

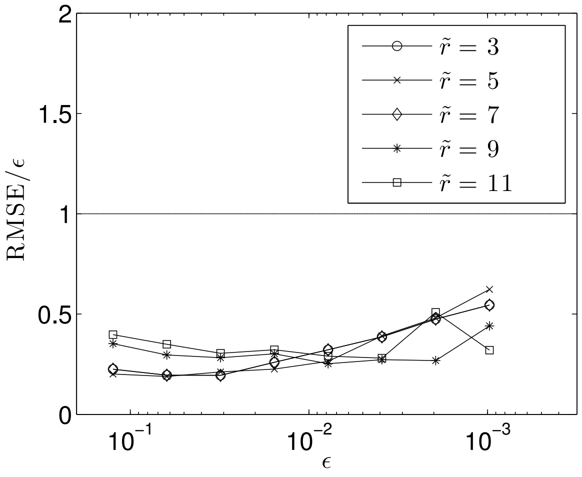

FIGURE 5. Path-dependent functional: error vs. accuracy demand ϵ.

as large as in the previous case, due to the slower decay of the bias. This results in a larger value of $L_1 - L_0$, which provides an advantage to the multi-level approach.

6.3. **Stopped Exit Times for SDEs.** In this section we set

$$\mu = 0.01,\ \sigma = 0.2,\ T = 2,$$

and we approximate the distribution function $F(s) = \mathrm{E}(1_{]\infty,s]}(Y))$ of

$$Y = \inf\{t \geq 0 : X_t = b\} \wedge T$$

with $b = 0.8$ on the interval

$$[S_0, S_1] = [0, 1].$$

The distribution of $\inf\{t \geq 0 : X_t = b\}$ is an inverse Gaussian distribution with parameters $\ln b/(\mu - \sigma^2/2)$ and $(\ln b)^2/\sigma^2$, and this yields an explicit formula for F since $T > S_1$.

See Figures 6, with replication numbers for $\epsilon = 2^{-9}$, and 7 for the results. Observe that the computational gain is even larger than in the previous section. This difference is due to the fact that smoothing already yields an improved weak error estimate for the present

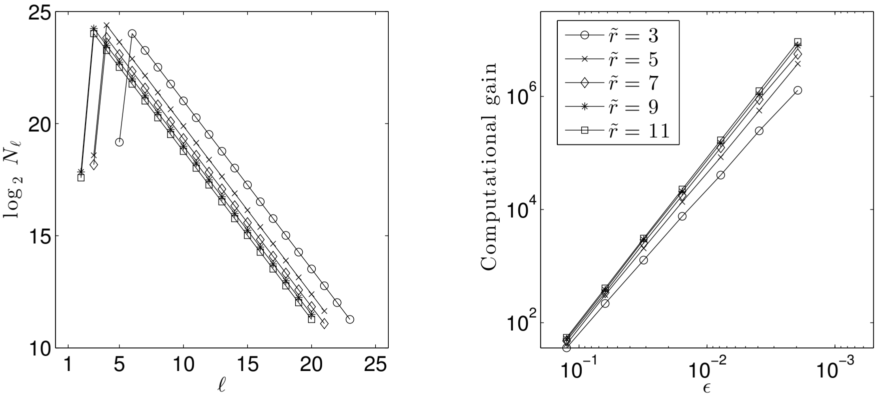

FIGURE 6. Stopped exit time: replication numbers per level (left) and computational gain (right).

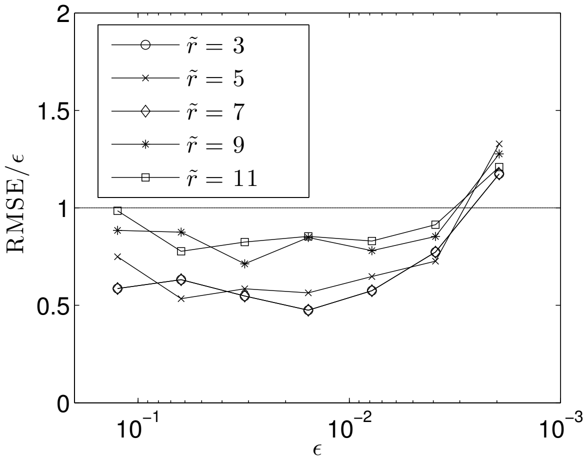

FIGURE 7. Stopped exit time: error vs. accuracy demand ϵ.

problem. Consequently,

$$L_1 = \left(2 + \frac{2}{\tilde r + 1}\right) \cdot \log_2 \epsilon^{-1}$$

is the maximal level for the multi-level algorithm, up to integer rounding, but for the single-level algorithm without smoothing we have to take

$$L = 4 \cdot \log_2 \epsilon^{-1}.$$

Acknowledgments. This work is inspired by a joint project with Oleg Iliev. The authors thank Oleg Iliev and Winfried Sickel for helpful discussions.

Tigran Nagapetyan was supported by the BMBF within the project 03MS612D FROPT, and Klaus Ritter was partially supported by the DFG within the Priority Program 1324.

REFERENCES

Alfonsi, A., Jourdain, B., Kohatsu-Higa, A. (2013), Pathwise optimal transport bounds between a one-dimensional diffusion and its Euler scheme, Preprint, arXiv:1209.0576.

Altmayer, M., Neuenkirch, A. (2013), Multilevel Monte Carlo quadrature of discontinuous payoffs in the generalized Heston model using Malliavin integration by parts, Preprint 144, DFG SPP 1324.

Avikainen, R. (2009), On irregular functionals of SDEs and the Euler scheme, Finance Stoch. **13**, 381–401.

Bally, V., Talay, D. (1996a), The law of the Euler scheme for stochastic differential equations, I. Convergence rate of the distribution function, Probab. Theory Relat. Fields **104**, 43–60.

Bally, V., Talay, D. (1996b), The law of the Euler scheme for stochastic differential equations, II. Convergence rate of the density, Monte Carlo Meth. Appl. **2**, 93–128

Bouchard, B., Geiss, S., Gobet, E. (2013), First time to exit of a continuous Itô process: general moment estimates and L_1-convergence rate for discrete time approximations, Preprint, arXiv:1307.4247.

Creutzig, J., Dereich, S., Müller-Gronbach, T., Ritter, K. (2009), Infinite-dimensional quadrature and approximation of distributions, Found. Comput. Math. **9**, 391–429.

Giles, M. B. (2008a), Multilevel Monte Carlo path simulation, Oper. Res. **56**, 607–617.

Giles, M. B. (2008b), Improved multilevel Monte Carlo convergence using the Milstein scheme, in Monte Carlo and Quasi-Monte Carlo Methods 2006, Keller, A., Heinrich, S., Niederreiter, H., eds., Springer, Heidelberg, pp. 343–358,

Giles, M. B., Debrabant, K., Rößler, A. (2013), Numerical analysis of multilevel Monte Carlo path simulation using the Milstein discretisation, Preprint, arXiv:1302.4676.

Giles, M. B., Higham, D. J., Mao, X. (2009), Analyzing multi-level Monte Carlo for options with non-globally Lipschitz payoff, Finance Stoch. **13**, 403–413.

Gobet, E., Labart, C. (2008), Sharp estimates for the convergence of the density of the Euler scheme in small time, Elect. Comm. in Probab. **13**, 352–363.

Heinrich, S. (1998), Monte Carlo complexity of global solution of integral equations, J. Complexity **14**, 151–175.

Higham, D. J., Mao. X., Roj, M., Song, Q., Yin, G. (2013), Mean exit times and the multilevel Monte Carlo method, SIAM/ASA J. Uncert. Quant. **1**, 2–18.

Jacod, J., Protter, P. (1998), Asymptotic error distributions for the Euler method for stochastic differential equations, Ann. Probab. **26**, 267–307.

Kebaier, A., (2005), Statistical Romberg extrapolation: a new variance reduction method and applications to option pricing, Ann. Appl. Prob. **15**, 2681–2705.

Kebaier, A., Kohatsu-Higa, A. (2008), An optimal control variance reduction method for density estimation, Stochastic Processes Appl. **118**, 2143–2180.

Massart, P. (1990), The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality, Ann. Probab. **18**, 1269–1283.

Milstein, G. N., Schoenmakers, J. G. M., Spokoiny, V. (2004), Transition density estimation for stochastic differential equations via forward-reverse representations, Bernoulli **10**, 281–312.

Shreve, E. S. (2008), Stochastic Calculus for Finance II. Continuous-Time Models. Springer, New York.

Talay, D., Zheng Z. (2004), Approximation of quantiles of components of diffusion processes, Stochastic Processes Appl. **109**, 23–46.

Tsybakov, A. B. (2009), Introduction to Nonparametric Estimation, Springer, New York.

MATHEMATICAL INSTITUTE, 24–29 ST GILES', OXFORD OX1 3LB, ENGLAND
E-mail address: `mike.giles@maths.ox.ac.uk`

DEPARTMENT OF FLOW AND MATERIAL SIMULATION, FRAUNHOFER ITWM, FRAUNHOFER-PLATZ 1, 67663 KAISERSLAUTERN, GERMANY
E-mail address: `nagapetyan@itwm.fraunhofer.de`

FACHBEREICH MATHEMATIK, TECHNISCHE UNIVERSITÄT KAISERSLAUTERN, POSTFACH 3049, 67653 KAISERSLAUTERN, GERMANY
E-mail address: `ritter@mathematik.uni-kl.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.

[81] W. Dahmen, C. Huang, C. Schwab, and G. Welper. Adaptive Petrov-Galerkin methods for first order transport equations. Preprint 81, DFG-SPP 1324, January 2011.

[82] K. Grella and C. Schwab. Sparse Tensor Spherical Harmonics Approximation in Radiative Transfer. Preprint 82, DFG-SPP 1324, January 2011.

[83] D.A. Lorenz, S. Schiffler, and D. Trede. Beyond Convergence Rates: Exact Inversion With Tikhonov Regularization With Sparsity Constraints. Preprint 83, DFG-SPP 1324, January 2011.

[84] S. Dereich, M. Scheutzow, and R. Schottstedt. Constructive quantization: Approximation by empirical measures. Preprint 84, DFG-SPP 1324, January 2011.

[85] S. Dahlke and W. Sickel. On Besov Regularity of Solutions to Nonlinear Elliptic Partial Differential Equations. Preprint 85, DFG-SPP 1324, January 2011.

[86] S. Dahlke, U. Friedrich, P. Maass, T. Raasch, and R.A. Ressel. An adaptive wavelet method for parameter identification problems in parabolic partial differential equations. Preprint 86, DFG-SPP 1324, January 2011.

[87] A. Cohen, W. Dahmen, and G. Welper. Adaptivity and Variational Stabilization for Convection-Diffusion Equations. Preprint 87, DFG-SPP 1324, January 2011.

[88] T. Jahnke. On Reduced Models for the Chemical Master Equation. Preprint 88, DFG-SPP 1324, January 2011.

[89] P. Binev, W. Dahmen, R. DeVore, P. Lamby, D. Savu, and R. Sharpley. Compressed Sensing and Electron Microscopy. Preprint 89, DFG-SPP 1324, March 2011.

[90] P. Binev, F. Blanco-Silva, D. Blom, W. Dahmen, P. Lamby, R. Sharpley, and T. Vogt. High Quality Image Formation by Nonlocal Means Applied to High-Angle Annular Dark Field Scanning Transmission Electron Microscopy (HAADF-STEM). Preprint 90, DFG-SPP 1324, March 2011.

[91] R. A. Ressel. A Parameter Identification Problem for a Nonlinear Parabolic Differential Equation. Preprint 91, DFG-SPP 1324, May 2011.

[92] G. Kutyniok. Data Separation by Sparse Representations. Preprint 92, DFG-SPP 1324, May 2011.

[93] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok. Introduction to Compressed Sensing. Preprint 93, DFG-SPP 1324, May 2011.

[94] H.-C. Kreusler and H. Yserentant. The Mixed Regularity of Electronic Wave Functions in Fractional Order and Weighted Sobolev Spaces. Preprint 94, DFG-SPP 1324, June 2011.

[95] E. Ullmann, H. C. Elman, and O. G. Ernst. Efficient Iterative Solvers for Stochastic Galerkin Discretizations of Log-Transformed Random Diffusion Problems. Preprint 95, DFG-SPP 1324, June 2011.

[96] S. Kunis and I. Melzer. On the Butterfly Sparse Fourier Transform. Preprint 96, DFG-SPP 1324, June 2011.

[97] T. Rohwedder. The Continuous Coupled Cluster Formulation for the Electronic Schrödinger Equation. Preprint 97, DFG-SPP 1324, June 2011.

[98] T. Rohwedder and R. Schneider. Error Estimates for the Coupled Cluster Method. Preprint 98, DFG-SPP 1324, June 2011.

[99] P. A. Cioica and S. Dahlke. Spatial Besov Regularity for Semilinear Stochastic Partial Differential Equations on Bounded Lipschitz Domains. Preprint 99, DFG-SPP 1324, July 2011.

[100] L. Grasedyck and W. Hackbusch. An Introduction to Hierarchical (H-) Rank and TT-Rank of Tensors with Examples. Preprint 100, DFG-SPP 1324, August 2011.

[101] N. Chegini, S. Dahlke, U. Friedrich, and R. Stevenson. Piecewise Tensor Product Wavelet Bases by Extensions and Approximation Rates. Preprint 101, DFG-SPP 1324, September 2011.

[102] S. Dahlke, P. Oswald, and T. Raasch. A Note on Quarkonial Systems and Multilevel Partition of Unity Methods. Preprint 102, DFG-SPP 1324, September 2011.

[103] A. Uschmajew. Local Convergence of the Alternating Least Squares Algorithm For Canonical Tensor Approximation. Preprint 103, DFG-SPP 1324, September 2011.

[104] S. Kvaal. Multiconfigurational time-dependent Hartree method for describing particle loss due to absorbing boundary conditions. Preprint 104, DFG-SPP 1324, September 2011.

[105] M. Guillemard and A. Iske. On Groupoid C*-Algebras, Persistent Homology and Time-Frequency Analysis. Preprint 105, DFG-SPP 1324, September 2011.

[106] A. Hinrichs, E. Novak, and H. Woźniakowski. Discontinuous information in the worst case and randomized settings. Preprint 106, DFG-SPP 1324, September 2011.

[107] M. Espig, W. Hackbusch, A. Litvinenko, H. Matthies, and E. Zander. Efficient Analysis of High Dimensional Data in Tensor Formats. Preprint 107, DFG-SPP 1324, September 2011.

[108] M. Espig, W. Hackbusch, S. Handschuh, and R. Schneider. Optimization Problems in Contracted Tensor Networks. Preprint 108, DFG-SPP 1324, October 2011.

[109] S. Dereich, T. Müller-Gronbach, and K. Ritter. On the Complexity of Computing Quadrature Formulas for SDEs. Preprint 109, DFG-SPP 1324, October 2011.

[110] D. Belomestny. Solving optimal stopping problems by empirical dual optimization and penalization. Preprint 110, DFG-SPP 1324, November 2011.

[111] D. Belomestny and J. Schoenmakers. Multilevel dual approach for pricing American style derivatives. Preprint 111, DFG-SPP 1324, November 2011.

[112] T. Rohwedder and A. Uschmajew. Local convergence of alternating schemes for optimization of convex problems in the TT format. Preprint 112, DFG-SPP 1324, December 2011.

[113] T. Görner, R. Hielscher, and S. Kunis. Efficient and accurate computation of spherical mean values at scattered center points. Preprint 113, DFG-SPP 1324, December 2011.

[114] Y. Dong, T. Görner, and S. Kunis. An iterative reconstruction scheme for photoacoustic imaging. Preprint 114, DFG-SPP 1324, December 2011.

[115] L. Kämmerer. Reconstructing hyperbolic cross trigonometric polynomials by sampling along generated sets. Preprint 115, DFG-SPP 1324, February 2012.

[116] H. Chen and R. Schneider. Numerical analysis of augmented plane waves methods for full-potential electronic structure calculations. Preprint 116, DFG-SPP 1324, February 2012.

[117] J. Ma, G. Plonka, and M.Y. Hussaini. Compressive Video Sampling with Approximate Message Passing Decoding. Preprint 117, DFG-SPP 1324, February 2012.

[118] D. Heinen and G. Plonka. Wavelet shrinkage on paths for scattered data denoising. Preprint 118, DFG-SPP 1324, February 2012.

[119] T. Jahnke and M. Kreim. Error bound for piecewise deterministic processes modeling stochastic reaction systems. Preprint 119, DFG-SPP 1324, March 2012.

[120] C. Bender and J. Steiner. A-posteriori estimates for backward SDEs. Preprint 120, DFG-SPP 1324, April 2012.

[121] M. Espig, W. Hackbusch, A. Litvinenkoy, H.G. Matthiesy, and P. Wähnert. Efficient low-rank approximation of the stochastic Galerkin matrix in tensor formats. Preprint 121, DFG-SPP 1324, May 2012.

[122] O. Bokanowski, J. Garcke, M. Griebel, and I. Klompmaker. An adaptive sparse grid semi-Lagrangian scheme for first order Hamilton-Jacobi Bellman equations. Preprint 122, DFG-SPP 1324, June 2012.

[123] A. Mugler and H.-J. Starkloff. On the convergence of the stochastic Galerkin method for random elliptic partial differential equations. Preprint 123, DFG-SPP 1324, June 2012.

[124] P.A. Cioica, S. Dahlke, N. Döhring, U. Friedrich, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R.L. Schilling. On the convergence analysis of Rothe's method. Preprint 124, DFG-SPP 1324, July 2012.

[125] P. Binev, A. Cohen, W. Dahmen, and R. DeVore. Classification Algorithms using Adaptive Partitioning. Preprint 125, DFG-SPP 1324, July 2012.

[126] C. Lubich, T. Rohwedder, R. Schneider, and B. Vandereycken. Dynamical approximation of hierarchical Tucker and Tensor-Train tensors. Preprint 126, DFG-SPP 1324, July 2012.

[127] M. Kovács, S. Larsson, and K. Urban. On Wavelet-Galerkin methods for semilinear parabolic equations with additive noise. Preprint 127, DFG-SPP 1324, August 2012.

[128] M. Bachmayr, H. Chen, and R. Schneider. Numerical analysis of Gaussian approximations in quantum chemistry. Preprint 128, DFG-SPP 1324, August 2012.

[129] D. Rudolf. Explicit error bounds for Markov chain Monte Carlo. Preprint 129, DFG-SPP 1324, August 2012.

[130] P.A. Cioica, K.-H. Kim, K. Lee, and F. Lindner. On the $L_q(L_p)$-regularity and Besov smoothness of stochastic parabolic equations on bounded Lipschitz domains. Preprint 130, DFG-SPP 1324, December 2012.

[131] M. Hansen. $n-$term Approximation Rates and Besov Regularity for Elliptic PDEs on Polyhedral Domains. Preprint 131, DFG-SPP 1324, December 2012.

[132] R. E. Bank and H. Yserentant. On the H^1 -stability of the L_2 -projection onto finite element spaces. Preprint 132, DFG-SPP 1324, December 2012.

[133] M. Gnewuch, S. Mayer, and K. Ritter. On Weighted Hilbert Spaces and Integration of Functions of Infinitely Many Variables. Preprint 133, DFG-SPP 1324, December 2012.

[134] D. Crisan, J. Diehl, P.K. Friz, and H. Oberhauser. Robust Filtering: Correlated Noise and Multidimensional Observation. Preprint 134, DFG-SPP 1324, January 2013.

[135] Wolfgang Dahmen, Christian Plesken, and Gerrit Welper. Double Greedy Algorithms: Reduced Basis Methods for Transport Dominated Problems. Preprint 135, DFG-SPP 1324, February 2013.

[136] Aicke Hinrichs, Erich Novak, Mario Ullrich, and Henryk Wozniakowski. The Curse of Dimensionality for Numerical Integration of Smooth Functions. Preprint 136, DFG-SPP 1324, February 2013.

[137] Markus Bachmayr, Wolfgang Dahmen, Ronald DeVore, and Lars Grasedyck. Approximation of High-Dimensional Rank One Tensors. Preprint 137, DFG-SPP 1324, March 2013.

[138] Markus Bachmayr and Wolfgang Dahmen. Adaptive Near-Optimal Rank Tensor Approximation for High-Dimensional Operator Equations. Preprint 138, DFG-SPP 1324, April 2013.

[139] Felix Lindner. Singular Behavior of the Solution to the Stochastic Heat Equation on a Polygonal Domain. Preprint 139, DFG-SPP 1324, May 2013.

[140] Stephan Dahlke, Dominik Lellek, Shiu Hong Lui, and Rob Stevenson. Adaptive Wavelet Schwarz Methods for the Navier-Stokes Equation. Preprint 140, DFG-SPP 1324, May 2013.

[141] Jonas Ballani and Lars Grasedyck. Tree Adaptive Approximation in the Hierarchical Tensor Format. Preprint 141, DFG-SPP 1324, June 2013.

[142] Harry Yserentant. A short theory of the Rayleigh-Ritz method. Preprint 142, DFG-SPP 1324, July 2013.

[143] M. Hefter and K. Ritter. On Embeddings of Weighted Tensor Product Hilbert Spaces. Preprint 143, DFG-SPP 1324, August 2013.

[144] M. Altmayer and A. Neuenkirch. Multilevel Monte Carlo Quadrature of Discontinuous Payoffs in the Generalized Heston Model using Malliavin Integration by Parts. Preprint 144, DFG-SPP 1324, August 2013.

[145] L. Kämmerer, D. Potts, and T. Volkmer. Approximation of multivariate functions by trigonometric polynomials based on rank-1 lattice sampling. Preprint 145, DFG-SPP 1324, September 2013.

[146] C. Bender, N. Schweizer, and J. Zhuo. A primal-dual algorithm for BSDEs. Preprint 146, DFG-SPP 1324, October 2013.

[147] D. Rudolf. Hit-and-run for numerical integration. Preprint 147, DFG-SPP 1324, October 2013.

[148] D. Rudolf and M. Ullrich. Positivity of hit-and-run and related algorithms. Preprint 148, DFG-SPP 1324, October 2013.

[149] L. Grasedyck, M. Kluge, and S. Krämer. Alternating Directions Fitting (ADF) of Hierarchical Low Rank Tensors. Preprint 149, DFG-SPP 1324, October 2013.

[150] F. Filbir, S. Kunis, and R. Seyfried. Effective discretization of direct reconstruction schemes for photoacoustic imaging in spherical geometries. Preprint 150, DFG-SPP 1324, November 2013.

[151] E. Novak, M. Ullrich, and H. Woźniakowski. Complexity of Oscillatory Integration for Univariate Sobolev Spaces. Preprint 151, DFG-SPP 1324, November 2013.

[152] A. Hinrichs, E. Novak, and M. Ullrich. A Tractability Result for the Clenshaw Curtis Smolyak Algorithm. Preprint 152, DFG-SPP 1324, November 2013.

[153] M. Hein, S. Setzer, L. Jost, and S. Rangapuram. The Total Variation on Hypergraphs - Learning on Hypergraphs Revisited. Preprint 153, DFG-SPP 1324, November 2013.

[154] M. Kovács, S. Larsson, and F. Lindgren. On the Backward Euler Approximation of the Stochastic Allen-Chan Equation. Preprint 154, DFG-SPP 1324, November 2013.

[155] S. Dahlke, M. Fornasier, U. Friedrich, and T. Raasch. Multilevel preconditioning for sparse optimization of functionals with nonconvex fidelity terms. Preprint 155, DFG-SPP 1324, December 2013.

[156] T. Müller-Gronbach, K. Ritter, and L. Yaroslavtseva. On the complexity of computing quadrature formulas for marginal distributions of SDEs. Preprint 156, DFG-SPP 1324, January 2014.

[157] M. Giles, T. Nagapetyan, and K. Ritter. Multi-Level Monte Carlo Approximation of Distribution Functions and Densities. Preprint 157, DFG-SPP 1324, February 2014.