SIAM/ASA J. Uncertainty Quantification
Vol. 3, pp. 267–295

Multilevel Monte Carlo Approximation of Distribution Functions and Densities*

Michael B. Giles†, Tigran Nagapetyan‡, and Klaus Ritter§

Abstract. We construct and analyze multilevel 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 stochastic differential equations (SDEs).

Key words. multilevel Monte Carlo, approximation of distribution functions and densities, stochastic differential equations, path-(in)dependent functionals, stopped exit times, smoothing

AMS subject classifications. 65C30, 65C20, 65C05, 60H35

DOI. 10.1137/140960086

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 assume there exists a sequence of random variables $Y^{(\ell)}$ that converge to Y in a suitable way and that are suited to simulation.

We present a general approach which is later applied in the context of stochastic differential equations (SDEs). In this specific setting we aim for 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 multilevel 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 multilevel 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. At its first stage the multilevel algorithm provides an approximation to F or ρ at discrete

*Received by the editors March 10, 2014; accepted for publication (in revised form) February 9, 2015; published electronically April 21, 2015.
http://www.siam.org/journals/juq/3/96008.html

†Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK (mike.giles@maths.ox.ac.uk).
‡Fraunhofer ITWM, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany (nagapetyan@wias-berlin.de). The research of this author was supported by the BMBF within the project 03MS612D FROPT.
§Fachbereich Mathematik, TU Kaiserslautern, 67653 Kaiserslautern, Germany (ritter@mathematik.uni-kl.de). The research of this author was partially supported by the DFG within the Priority Program 1324.

Table 1
Orders of convergence of the multilevel algorithm.

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

points, which is extended to a function on $[S_0, S_1]$ in a standard and purely deterministic way. The parameters of our multilevel algorithm are the minimal and maximal levels, the replication numbers per level, the smoothing parameter, and the number of discrete points to be used in the first stage.

In general, the analysis of multilevel Monte Carlo algorithms is based on assumptions about the convergence rates of the weak and the strong error in terms of the computational cost. For the present problem, the standard assumptions that are used for the approximation of expectations have to be modified appropriately. We derive upper bounds for $\mathrm{error}(\mathcal{A})$, the root mean square error, and the computational cost, $\mathrm{cost}(\mathcal{A})$, in terms of the parameters of the algorithm $\mathcal{A}$, the strong and weak convergence rates, and the smoothness of ρ, 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 $\mathrm{cost}(\mathcal{A}) \le O\left(\mathrm{error}(\mathcal{A})^{-\theta+\varepsilon}\right)$ for every $\varepsilon > 0$, where $\theta > 0$. Roughly speaking, θ is the order of convergence of the multilevel algorithm. See Theorems 2.6 and 4.3 for the precise statements involving also powers of $\log \mathrm{error}(\mathcal{A})$.

Here we only present an 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 multilevel algorithm, and we assume that 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 $r \ge 1$ and $\delta_0 > 0$ 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 in the second row, $\varphi(x) = \inf\{t \ge 0 : x(t) \in \partial D\} \wedge T$ is a stopped exit time from a bounded domain $D \subset \mathbb{R}^d$. We note that in every case represented in Table 1, proper multilevel 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 $d = 1$), 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 multilevel 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. [15]. In the limit $r \to \infty$ we achieve the same values of θ for the approximation of the distribution function or the density of $\varphi(X)$. It is well known that the forward Kolmogorov equation provides an alternative way to approximate the density of the random variable Y in the case of a low-dimensional SDE and a path-independent functional φ. We refer the reader to Gobet and Menozzi [13] for a

corresponding representation result in the case of stopped exit times.

Multilevel algorithms, which have been introduced by Heinrich [14] and Giles [9] (see also Kebaier [17] for the two-level construction), are 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 [3] and Giles, Higham, and Mao [11], and with implicit smoothing through the use of conditional expectations by Giles [8] and Giles, Debrabant, and Rößler [10]. Furthermore, Altmayer and Neuenkirch [2] combine smoothing by Malliavin integration by parts with the multilevel approach to approximate expectations of discontinuous payoffs in the Heston model. The second issue has already been worked out by Heinrich [14] 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 and Kohatsu-Higa [18] in the case $r = \infty$, and their analysis yields $\theta = 5/2$.

This paper is organized as follows. In sections 2–4 we provide the general analysis of the three approximation problems, 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 are similar: We discuss, in particular, the assumptions on the weak and the strong convergence, and we construct and analyze the respective multilevel 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 straightforward application of the multilevel Monte Carlo approach based on $F(s) = \mathrm{E}(1_{]-\infty,s]}(Y))$ could suffer from the discontinuity of $1_{]-\infty,s]}$; see Remark 5.1 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 2.1. Such a function g is easily constructed as follows. There exists a uniquely determined polynomial p of degree at most $r+1$ such that $\int_{-1}^{1} s^j \cdot p(s)\,ds = (-1)^j/(j+1)$ for $j = 0, \dots, r-1$, 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 $g((\cdot - s)/\delta)$. We omit the proof, which is based on a Taylor expansion in a straightforward way.

Lemma 2.2. *There exists a constant $c > 0$ such that for all $\delta \in]0, \delta_0]$,*

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

2.2. Assumptions on weak and strong convergence. Our multilevel 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 a cost of at most $c \cdot M^\ell$ for every $\ell \in \mathbb{N}$.

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

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

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

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

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

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 5.2. 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 (stochastic partial differential equations), 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 $1 \le p < \infty$. 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

$$\sup_{s\in[S_0,S_1]} \left|\mathrm{E}\left(g((Y-s)/\delta) - g((Y^{(\ell)}-s)/\delta)\right)\right| \le c_L \cdot \delta^{-1} \cdot \|Y - Y^{(\ell)}\|_1, \tag{2.1}$$

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

$$\mathrm{E}\min((Y - Y^{(\ell)})^2, \delta^2) \le \min(\|Y - Y^{(\ell)}\|_2^2, \delta^2), \tag{2.2}$$

$$\mathrm{E}\min((Y - Y^{(\ell)})^2, \delta^2) \le \mathrm{E}(\delta \cdot \min(|Y - Y^{(\ell)}|, \delta)) \le \min(\delta \cdot \|Y - Y^{(\ell)}\|_1, \delta^2). \tag{2.3}$$

In the following case of equivalence of norms the upper bound in (2.2) 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 2.3. *Suppose that there exist $c_1 > 0$ and $p > 2$ such that*

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

for all $\ell \in \mathbb{N}_0$. Then there exists $c_2 > 0$ such that for all $\delta \in]0, \delta_0]$ and $\ell \in \mathbb{N}_0$,

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

Proof. Put $Z_\ell = (Y - Y^{(\ell)})^2 / \|Y - Y^{(\ell)}\|_2^2$. We show that there exists a constant $c_2 > 0$ such that $\mathrm{E} \min(Z_\ell, \delta) \ge c_2 \cdot \min(1, \delta)$ for all $\ell \in \mathbb{N}_0$ and $\delta > 0$.

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

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

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,

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

On the other hand, if $\|Y - Y^{(\ell)}\|_2^2$ and $\|Y - Y^{(\ell)}\|_1$ are asymptotically equivalent, then (2.3) is preferable to (2.2). See section 5.3 for examples.

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

Lemma 2.4. *There exists a constant $c > 0$ such that for all $\delta \in]0, \delta_0]$ and $\ell \in \mathbb{N}_0$,*

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

2.3. The multilevel algorithm. The approximation of F on the interval $[S_0, S_1]$ is based on its approximation at finitely many points

$$S_0 \le s_1 < \cdots < s_k \le S_1, \tag{2.4}$$

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

For notational convenience we put

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

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 multilevel algorithm for the approximation at the points s_i is defined by

$$\mathcal{M}^{k,\delta,L_0,L_1}_{N_{L_0},\dots,N_{L_1}} = \frac{1}{N_{L_0}} \cdot \sum_{i=1}^{N_{L_0}} g^{k,\delta}(Y_i^{(L_0)}) + \sum_{\ell=L_0+1}^{L_1} \frac{1}{N_\ell} \cdot \sum_{i=1}^{N_\ell} \left(g^{k,\delta}(Y_i^{(\ell)}) - g^{k,\delta}(Z_i^{(\ell)}) \right) \tag{2.5}$$

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 2.5. 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

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

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

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

to extend the definition of $\mathcal{M}_N^{k,\delta,L,L}$ to $\delta \ge 0$. Observe that $\mathcal{M}_N^{k,0,L,L}$ 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 a cost of 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 exists a constant $\alpha > 0$ such that the weak error estimate

$$\sup_{s \in [S_0,S_1]} \left| \mathrm{E}\left(1_{]-\infty,s]}(Y) - 1_{]-\infty,s]}(Y^{(\ell)}) \right) \right| \le c \cdot M^{-\ell\cdot\alpha} \tag{2.6}$$

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

$$\alpha_1 = 0, \qquad \alpha_2 = \alpha, \qquad \alpha_3 = \alpha. \tag{2.7}$$

In what follows, $\|\cdot\|_\infty$ 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 $Q_k^r : \mathbb{R}^k \to C([S_0, S_1])$ 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 $Q_k^r(x)$ is bounded by $c \cdot k$.
(E2) For all $k \in \mathbb{N}$ and $x \in \mathbb{R}^k$ we have $\|Q_k^r(x)\|_\infty \le c \cdot |x|_\infty$.
(E3) For all $k \in \mathbb{N}$ we have $\|F - Q_k^r(F(s_1), \dots, F(s_k))\|_\infty \le c \cdot k^{-(r+1)}$.

For instance, these properties are achieved by piecewise polynomial interpolation with a fixed degree $\max(r,1)$ at equidistant points $s_i = S_0 + (i-1)\cdot(S_1-S_0)/(k-1)$ with $k \geq 2$. As a particular example, cubic spline interpolation achieves these properties for $r = 3$ [7].

We employ $Q_k^r(\mathcal{M})$ with $\mathcal{M} = \mathcal{M}_{N_{L_0},\ldots,N_{L_1}}^{k,\delta,L_0,L_1}$ 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 $\mathrm{E}\,\|Q_k^r(\mathcal{M})\|_\infty^2 < \infty$. The error of $Q_k^r(\mathcal{M})$ is defined by

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

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

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

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

and

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

holds for $x \in \mathbb{R}^k$. Furthermore, $\mathrm{Var}(\mathcal{M}) \leq 4\cdot\mathrm{E}(|\mathcal{M}|_\infty^2)$. The Bienaymé formula for real-valued random variables turns into the inequality

$$\mathrm{Var}(\mathcal{M}) \leq c\cdot\log k\cdot\sum_{i=1}^n \mathrm{Var}(\mathcal{M}_i) \tag{2.8}$$

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 multilevel algorithms this is exploited in Heinrich [14].

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

$$q = \min\left(\frac{r+1+\alpha_1}{\alpha_2}, \frac{r+1}{\alpha_3}\right). \tag{2.9}$$

Theorem 2.6. *The following order, with $\eta = 1$, is achieved by algorithms $Q_k^r(\mathcal{M}_{N_{L_0},\ldots,N_{L_1}}^{k,\delta,L_0,L_1})$ with suitably chosen parameters:*

$$q \leq \max(1,\beta_4/\beta_5) \quad\Rightarrow\quad \gamma = 2 + \frac{\max(1,q)}{r+1}, \tag{2.10}$$

$$q > \max(1,\beta_4/\beta_5) \ \wedge\ \beta_5 > 1 \quad\Rightarrow\quad \gamma = 2 + \frac{\max(1,\beta_4/\beta_5)}{r+1}, \tag{2.11}$$

$$q > 1 > \beta_4 \ \wedge\ \beta_5 = 1 \quad\Rightarrow\quad \gamma = 2 + \frac{1}{r+1}, \tag{2.12}$$

$$q > \max(1,\beta_4/\beta_5) \ \wedge\ \beta_5 < 1 \quad\Rightarrow\quad \gamma = 2 + \frac{\max(1,\beta_4+(1-\beta_5)\cdot q)}{r+1}. \tag{2.13}$$

Moreover, with $\eta = 3$,

$$q > \beta_4 \geq 1 \ \wedge \ \beta_5 = 1 \quad \Rightarrow \quad \gamma = 2 + \frac{\beta_4}{r+1}. \tag{2.14}$$

Proof. Let $\mathcal{M}$ denote any square-integrable random variable with values in $\mathbb{R}^k$. For the error of $Q_k^r(\mathcal{M})$ the properties (E2) and (E3) yield

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

Now we consider the algorithm $\mathcal{M} = \mathcal{M}_{N_{L_0}, \dots, N_{L_1}}^{k,\delta,L_0,L_1}$ 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 $a \leq c \cdot b$. 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 Lemma 2.2 and (A3) imply a bias estimate

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

The variance of $\mathcal{M}$ is estimated as follows. From (2.8) 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,

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

for $\ell = L_0 + 1, \dots, L_1$ (see Lemma 2.4), 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

$$\begin{aligned}
\operatorname{error}^2(Q_k^r(\mathcal{M})) \preceq\ & k^{-2(r+1)} + \delta^{2(r+1)} + \min\left(\delta^{-2\alpha_1} \cdot M^{-L_1 \cdot 2\alpha_2}, M^{-L_1 \cdot 2\alpha_3}\right) \\
& + \log k \cdot \left(\frac{1}{N_{L_0}} + \sum_{\ell=L_0+1}^{L_1} \frac{\min(\delta^{-\beta_4} \cdot M^{-\ell \cdot \beta_5}, 1)}{N_\ell}\right).
\end{aligned} \tag{2.15}$$

Now we analyze the computational cost of the algorithm $\mathcal{M}$. For $\ell = L_0, \dots, L_1$ and $i = 1, \dots, N_\ell$, 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

$$\operatorname{cost}(Q_k^r(\mathcal{M})) \preceq c(k, L_0, L_1, N_{L_0}, \dots, N_{L_1}) = \sum_{\ell=L_0}^{L_1} N_\ell \cdot (M^\ell + k). \tag{2.16}$$

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

Observe that the estimates (2.15) and (2.16) are also valid 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 (2.7), 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 (2.16) for the cost, subject to the constraint that the upper bound (2.15) for the squared error is at most ϵ^2, up to multiplicative constants for both quantities.

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

$$\delta = \epsilon^{1/(r+1)}, \tag{2.17}$$

$$k = \epsilon^{-1/(r+1)}, \tag{2.18}$$

$$N_{L_0} = \epsilon^{-2} \cdot \log_M \epsilon^{-1} \tag{2.19}$$

up to integer rounding for k and N_{L_0}. This yields

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

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

$$L^* = \frac{1}{r+1} \cdot \log_M \epsilon^{-1}. \tag{2.20}$$

Due to the dependence of (2.16) 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

$$L_0 \ge L^*. \tag{2.21}$$

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

$$L_1 = \max(1, q) \cdot L^* \tag{2.22}$$

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 $\operatorname{error}(Q_k^r(\mathcal{M})) \preceq \epsilon$ as well as

$$c(k, L_1, L_1, N_{L_1}) \asymp \epsilon^{-2} \cdot \log \epsilon^{-1} \cdot M^{\max(1,q) \cdot L^*} = \log \epsilon^{-1} \cdot \begin{cases} \epsilon^{-2-1/(r+1)} & \text{if } q \le 1, \\ \epsilon^{-2-q/(r+1)} & \text{if } q > 1. \end{cases} \tag{2.23}$$

For a single-level algorithm without smoothing we obtain the same result.

For a proper multilevel algorithm with $L^* \le L_0 < L_1$ we obtain

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

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

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

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 (2.23), we get (2.10) in the case $q \le 1$ by single-level algorithms, without any need for a multilevel treatment.

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 $\sum_{\ell=L_0+1}^{L_1} v_\ell/N_\ell \le \epsilon^2/\log \epsilon^{-1}$. A Lagrange multiplier leads to

$$N_\ell = \epsilon^{-2} \cdot \log \epsilon^{-1} \cdot G(L_0) \cdot \left(v_\ell \cdot M^{-\ell}\right)^{1/2}, \tag{2.24}$$

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

$$h(L_0, N_{L_0+1}, \dots, N_{L_1}) = \epsilon^{-2} \cdot \log \epsilon^{-1} \cdot \left(M^{L_0} + G^2(L_0)\right). \tag{2.25}$$

Put $L^\dagger = \beta_4/\beta_5 \cdot L^*$. Consider the case $1 < q \le \beta_4/\beta_5$. Then we have $L_1 \le L^\dagger$, 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 (2.23), we get (2.10) in the present case by single-level algorithms, without any need for a multilevel treatment.

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

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

It therefore suffices to study the case $L_0 \geq L^\dagger$, where we have

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

Note that

$$\begin{aligned}
\beta_5 = 1 \quad &\Rightarrow \quad M^{L_0} + G^2(L_0) \asymp M^{L_0} + M^{L^*\cdot\beta_4} \cdot (L_1 - L_0)^2, \\
\beta_5 > 1 \quad &\Rightarrow \quad M^{L_0} + G^2(L_0) \asymp M^{L_0} + M^{L^*\cdot\beta_4} \cdot M^{L_0\cdot(1-\beta_5)} \asymp M^{L_0}, \\
\beta_5 < 1 \quad &\Rightarrow \quad M^{L_0} + G^2(L_0) \asymp M^{L_0} + M^{L^*\cdot\beta_4} \cdot M^{L_1\cdot(1-\beta_5)}.
\end{aligned}$$

Hence, we choose

$$L_0 = \max(1, \beta_4/\beta_5) \cdot L^* \tag{2.26}$$

in all of 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. (2.23). Use (2.25) and $M^{L^*} = \epsilon^{-1/(r+1)}$ to derive (2.11)–(2.14). ∎

Remark 2.7. Theorem 2.6 is based on the upper bounds (2.15) and (2.16) for the error and the cost, respectively, of the algorithms $Q_k^r(\mathcal{M}^{k,\delta,L_0,L_1}_{N_{L_0},\dots,N_{L_1}})$. The parameters that we determined in the proof of Theorem 2.6 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 for the choice of δ, k, N_{L_0}, and L_1 according to (2.17), (2.18), (2.19), and (2.22). Moreover, the constraint (2.21) 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 multilevel algorithms the optimal values of N_ℓ for $\ell = L_0+1, \dots, L_1$ are presented in (2.24), and the optimal value of L_0 is presented in (2.26), if $q > \max(1, \beta_4/\beta_5)$. It is straightforward to verify

$$N_\ell = \epsilon^{-2-\beta_4/(r+1)} \cdot \log \epsilon^{-1} \cdot M^{-\ell\cdot(1+\beta_5)/2} \cdot \begin{cases} L^* & \text{if } \beta_5 = 1, \\ M^{L^*\cdot\max(1,\beta_4/\beta_5)\cdot(1-\beta_5)/2} & \text{if } \beta_5 > 1, \\ M^{L^*\cdot q\cdot(1-\beta_5)/2} & \text{if } \beta_5 < 1. \end{cases} \tag{2.27}$$

Furthermore, we have carried out the comparison of multilevel and single-level algorithms in the proof of Theorem 2.6. This comparison also is merely based on the upper bounds for

the error and the cost, and on the assumption that $\alpha = \alpha_3$ in (2.6). In this sense we have a superiority of proper multilevel algorithms over single-level algorithms if and only if

$$q > \max(1, \beta_4/\beta_5), \tag{2.28}$$

which corresponds to (2.11)–(2.14) in Theorem 2.6. The lack of superiority, which is present in (2.10) of Theorem 2.6, is explained as follows. For $q \le 1$, 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 $1 < q \le \beta_4/\beta_5$, 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

$$\frac{r+1}{\alpha_3} > \max(1, q). \tag{2.29}$$

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

Remark 2.8. In the limit $r \to \infty$ we get $\gamma = 2 + \max(1 - \beta_5, 0)/\alpha_2$ in Theorem 2.6, which coincides with the order for the approximation of expectations by means of multilevel algorithms; see Giles [9, 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 [19] yields

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

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 multilevel 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 follow the approach from section 2.

3.1. Smoothing.

We employ assumption (A1) with $r \ge 1$, 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 $j = 1, \dots, r-1$.

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 3.1. We modify the construction from Remark 2.1 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

$\int_{-1}^{1} s^j \cdot p(s)\,ds = 0$ for $j = 0, \dots, r-1$, 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 [24, Prop. 1.2].

Lemma 3.2. *There exists a constant $c > 0$ such that for all $\delta \in]0, \delta_0]$,*

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

3.2. Assumptions on weak and strong convergence. We employ 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 2.4, and we refer the reader to section 5 for specific examples with corresponding values of α_i.

3.3. The multilevel algorithm. The definition (2.5) 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

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

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

Hereby we obtain approximations to ρ at the points (2.4), 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:

(E4) For all $k \in \mathbb{N}$, we have $\|\rho - Q^r_k(\rho(s_1), \dots, \rho(s_k))\|_\infty \le c \cdot k^{-r}$.

As before, piecewise polynomial interpolation at equidistant points, now with a fixed 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 (2.9).

Theorem 3.3. *The following order, with $\eta = 1$, is achieved by algorithms $Q^r_k(\mathcal{M}^{k,\delta,L_0,L_1}_{N_{L_0},\dots,N_{L_1}})$ with suitably chosen parameters:*

$$q \le \max(1, \beta_4/\beta_5) \quad \Rightarrow \quad \gamma = 2 + \frac{\max(1, q) + 2}{r}, \tag{3.1}$$

$$q > \max(1, \beta_4/\beta_5) \ \wedge \ \beta_5 > 1 \quad \Rightarrow \quad \gamma = 2 + \frac{\max(1, \beta_4/\beta_5) + 2}{r}, \tag{3.2}$$

$$q > 1 > \beta_4 \ \wedge \ \beta_5 = 1 \quad \Rightarrow \quad \gamma = 2 + \frac{3}{r}, \tag{3.3}$$

$$q > \max(1, \beta_4/\beta_5) \ \wedge \ \beta_5 < 1 \quad \Rightarrow \quad \gamma = 2 + \frac{\max(1, \beta_4 + (1 - \beta_5) \cdot q) + 2}{r}. \tag{3.4}$$

Moreover, with $\eta = 3$,

$$q > \beta_4 \geq 1 \ \wedge \ \beta_5 = 1 \quad \Rightarrow \quad \gamma = 2 + \frac{\beta_4 + 2}{r}. \tag{3.5}$$

Proof. We mimic the proof of Theorem 2.6. We use (A3), (E2) and (E4), Lemmas 2.4 and 3.2, and the boundedness of g to obtain

$$\begin{aligned} \text{error}^2(Q_k^r(\mathcal{M})) \preceq\ & k^{-2r} + \delta^{2r} + 1/\delta^2 \cdot \min\left(\delta^{-2\alpha_1} \cdot M^{-L_1 \cdot 2\alpha_2}, M^{-L_1 \cdot 2\alpha_3}\right) \\ & + \log k/\delta^2 \cdot \left(\frac{1}{N_{L_0}} + \sum_{\ell=L_0+1}^{L_1} \frac{\min(\delta^{-\beta_4} \cdot M^{-\ell \cdot \beta_5}, 1)}{N_\ell} \right), \end{aligned} \tag{3.6}$$

where $\mathcal{M} = \mathcal{M}^{k,\delta,L_0,L_1}_{N_{L_0},\dots,N_{L_1}}$. The upper bound (2.16) for the computational cost is also valid in the present case. We minimize (2.16), subject to the constraint that the upper bound (3.6) for the squared error is at most ϵ^2, up to multiplicative constants for both quantities. The analysis literally follows the proof of Theorem 2.6, with ϵ being replaced by $\epsilon^{1+1/r}$. ∎

Remark 3.4. The following comments on optimality, etc., are meant in the sense of Remark 2.7. We have a superiority of proper multilevel algorithms over single-level algorithms if and only if (2.28) holds true. Moreover, the optimal values of δ, k, N_{L_0}, and L_1 are given by (2.17), (2.18), (2.19), and (2.22), with ϵ being replaced by $\epsilon^{1+1/r}$. 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 multilevel algorithms. The optimal value of L_0 is given by $L_0 = \max(1, \beta_4/\beta_5)/r \cdot \log_M \epsilon^{-1}$ (see (2.26)), and the optimal replication numbers N_ℓ for $\ell = L_0 + 1, \dots, L_1$ are given by (2.27), with ϵ being replaced by $\epsilon^{1+1/r}$.

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 assumptions (S1)–(S4). In particular, we make use of Lemma 2.2.

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 2.4 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

$$\beta_1 = \beta_4, \qquad \beta_2 = \beta_5, \qquad \beta_3 = 0, \tag{4.1}$$

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 $\alpha_i = \beta_i/2$ for $i = 1, 2, 3$.

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 4.1 (Avikainen [3]). *There exists a constant $c > 0$ such that*

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

and

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

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

Proof. See Avikainen [3, p. 387] for the proof of the first estimate and Avikainen [3, Lem. 3.4] for the second estimate. ∎

Lemma 4.2. *For every $1 \le q \le p < \infty$ there exists a constant $c > 0$ such that*

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

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 what follows, we adopt the notation $\preceq$ from the proof of Theorem 2.6, 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 4.1 we have $\mathrm{E}\,\Delta^q \preceq \|Y - Y^{(\ell)}\|_p^{p/(p+1)}$, so all that remains is to establish is

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

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

$$\|Y - Y^{(\ell)}\|_p \le \delta^{1+1/p}. \tag{4.2}$$

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

$$A_1 = \{|Y - s| \le 2\delta\}, \quad A_2 = \{|Y - s| > 2\delta\} \cap \{|Y - Y^{(\ell)}| \ge \delta\},$$

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 $p_1 = P(A_1) > 0$, then for $q < p$, Jensen's inequality and the Lipschitz continuity of g give

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

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 $P(\{|Y - Y^{(\ell)}| \ge \delta\}) \le \delta^{-p} \cdot \|Y - Y^{(\ell)}\|_p^p$, and hence, using the boundedness of g,

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

with the last step coming from (4.2). ∎

If $\|Y - Y^{(\ell)}\|_p$ and $\|Y - Y^{(\ell)}\|_1$ are asymptotically equivalent for every $1 \le p < \infty$, then Lemmas 4.1 and 4.2 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 multilevel algorithm. We study multilevel 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 $g^\delta(t) = g((t-s)/\delta)$ for $t \in \mathbb{R}$, which form a particular instance of (2.5). The error of $\mathcal{M} = \mathcal{M}^{\delta,L_0,L_1}_{N_{L_0},\dots,N_{L_1}}$ is defined by

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

and Remark 2.5 applies to single-level algorithms.

Put $\beta^\dagger = \beta_1/(\beta_2 - \beta_3)$, and recall the definition of q from (2.9).

Theorem 4.3. *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:*

$$q \le \beta^\dagger \ \wedge\ \beta_3 \ne 1 \quad \Rightarrow \quad \gamma = 2 + \frac{(1-\beta_3)_+ \cdot q}{r+1}, \tag{4.3}$$

$$q > \beta^\dagger \ \wedge\ \beta_3 \ne 1 \ \wedge\ \beta_2 > 1 \quad \Rightarrow \quad \gamma = 2 + \frac{(1-\beta_3)_+ \cdot \beta^\dagger}{r+1}, \tag{4.4}$$

$$q > \beta^\dagger \ \wedge\ \beta_2 < 1 \quad \Rightarrow \quad \gamma = 2 + \frac{\beta_1 + (1-\beta_2)\cdot q}{r+1}. \tag{4.5}$$

Moreover, with $\eta = 2$,

$$\beta_3 = 1 \quad \Rightarrow \quad \gamma = 2, \tag{4.6}$$

$$q > \beta^\dagger \ \wedge\ \beta_2 = 1 \quad \Rightarrow \quad \gamma = 2 + \frac{\beta_1}{r+1}. \tag{4.7}$$

Proof. We proceed analogously to the proof of Theorem 2.6. Use Lemma 2.2, the assumptions (A3) and (A5), and the boundedness of g to obtain

$$\begin{aligned}\operatorname{error}^2(\mathcal{M}) \preceq \delta^{2(r+1)} &+ \min\left(\delta^{-2\alpha_1}\cdot M^{-L_1\cdot 2\alpha_2}, M^{-L_1\cdot 2\alpha_3}\right)\\ &+ \frac{1}{N_{L_0}} + \sum_{\ell=L_0+1}^{L_1} \frac{\min\left(\delta^{-\beta_1}\cdot M^{-\ell\cdot\beta_2}, M^{-\ell\cdot\beta_3}\right)}{N_\ell\cdot\delta^2}\end{aligned} \tag{4.8}$$

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

$$\operatorname{cost}(\mathcal{M}) \preceq c(L_0, L_1, N_{L_0}, \dots, N_{L_1}), \tag{4.9}$$

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 (4.9) for the cost, subject to the constraint that the upper bound (4.8) for the squared error is at most ϵ^2, up to multiplicative constants for both quantities.

To this end we choose δ according to (2.17), and, up to integer rounding,

$$N_{L_0} = \epsilon^{-2}, \tag{4.10}$$

$$L_1 = q\cdot L^*, \tag{4.11}$$

with L^* given by (2.20).

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

$$c(L_1, L_1, N_{L_1}) \asymp \epsilon^{-2-q/(r+1)}. \tag{4.12}$$

For a proper multilevel algorithm, i.e., $L_0 < L_1$, we obtain

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

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

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

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 $\sum_{\ell=L_0+1}^{L_1} v_\ell/N_\ell \le \epsilon^2$. A Lagrange multiplier leads to

$$N_\ell = \epsilon^{-2}\cdot G(L_0)\cdot\left(v_\ell\cdot M^{-\ell}\right)^{1/2}, \tag{4.13}$$

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

Put $L^\dagger = \beta^\dagger \cdot L^*$. In the case $q \le \beta^\dagger$ we have $L_1 \le L^\dagger$, 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 $q \le \beta^\dagger$. 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 where $\beta_3 = 0$ and $q \le \beta^\dagger$, these estimates are superior to M^{L_1}, which corresponds to (4.12). ∎

Remark 4.4. The following comments on optimality, etc., are meant in the sense of Remark 2.7. The optimal values of δ, N_{L_0}, and L_1 are given by (2.17), (4.10), and (4.11), 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 (4.13).

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

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

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

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

$$\frac{r+1}{\alpha_3} > \frac{r+1+\alpha_1}{\alpha_2}. \tag{4.14}$$

5. Applications. 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

$$\|Y - Y^{(\ell)}\|_p \leq c_p \cdot M^{-\ell \cdot \beta/2} \tag{5.1}$$

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

From (5.1) we obtain (A4) with

$$\beta_4 = 2, \qquad \beta_5 = \beta \tag{5.2}$$

(see (2.2)), and Lemmas 4.1 and 4.2 yield (A5) with

$$\beta_1 = 1 + \varepsilon, \qquad \beta_2 = \beta, \qquad \beta_3 = \beta/2 - \varepsilon \tag{5.3}$$

under the assumptions (S2) and (S3) or (S2) and (S5). Using Lemmas 4.1 and 4.2 again we get (A3) under both sets of assumptions on g, with

$$\alpha_1 = \varepsilon, \qquad \alpha_2 = \beta/2, \qquad \alpha_3 = \beta/2 - \varepsilon, \tag{5.4}$$

and (2.6) holds with

$$\alpha = \beta/2 - \varepsilon. \tag{5.5}$$

It follows that $q = 2 \cdot (r+1)/\beta + \varepsilon$ and $\max(1, \beta_4/\beta_5) = 2/\beta$, so that (2.11), (2.13), and (2.14) in Theorem 2.6 yield

$$1 \leq \beta \leq 2 \;\Rightarrow\; \gamma = 2 + \frac{2}{\beta \cdot (r+1)} \quad \text{and} \quad 0 < \beta < 1 \;\Rightarrow\; \gamma = \frac{2}{\beta} + \frac{2}{r+1} + \varepsilon \tag{5.6}$$

for the approximation of F on $[S_0, S_1]$. Likewise, (3.2), (3.4), and (3.5) in Theorem 3.3 yield

$$1 \leq \beta \leq 2 \;\Rightarrow\; \gamma = 2 + \frac{2 \cdot (1+\beta)}{\beta \cdot r} \quad \text{and} \quad 0 < \beta < 1 \;\Rightarrow\; \gamma = \frac{2}{\beta} + \frac{2 \cdot (1+\beta)}{\beta \cdot r} + \varepsilon, \tag{5.7}$$

for the approximation of ρ on $[S_0, S_1]$. Moreover, $\beta^\dagger = 2/\beta + \varepsilon$, so that (4.4), (4.5), and (4.7) in Theorem 4.3 yield

$$1 \leq \beta \leq 2 \;\Rightarrow\; \gamma = 2 + \frac{2-\beta}{\beta \cdot (r+1)} + \varepsilon \quad \text{and} \quad 0 < \beta < 1 \;\Rightarrow\; \gamma = \frac{2}{\beta} + \frac{1}{r+1} + \varepsilon \tag{5.8}$$

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 multilevel algorithms are always superior to single-level algorithms; see Remarks 2.7, 3.4, and 4.4.

Remark 5.1. We compare the smoothing approach for the approximation of F at a single point with a direct approach, which is due to Avikainen [3] and which only requires that Y has a bounded density ρ; see Lemma 4.1.

We study multilevel algorithms

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

for the approximation of $F(s)$. As previously, we assume that (5.1) with $0 < \beta \le 2$ is all we have at hand. The analysis from Theorem 4.3 directly applies if we take $\beta_1 = 0$, $\beta_2 = \beta/2 - \varepsilon$, and $\beta_3 = \beta/2 - \varepsilon$ as well as $\alpha_1 = 0$, $\alpha_2 = \beta/2 - \varepsilon$, and $\alpha_3 = \beta/2 - \varepsilon$. We achieve the order (γ', η') with

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

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

In what follows, we consider three specific settings in the context of a stochastic differential equation (SDE)

$$\begin{aligned} dX_t &= \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t, \qquad t \in [0,T], \\ X_0 &= x_0. \end{aligned}$$

Here $\mu : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d \times m}$, $x_0 \in \mathbb{R}^d$, and $(W_t)_{t \in [0,T]}$ is an m-dimensional Brownian motion. We impose at least the following assumptions on the drift coefficient $f = \mu$ and the diffusion coefficient $f = \sigma$: There exists a constant $K > 0$ such that

$$\begin{aligned} |f(t,x) - f(t,y)| &\le K \cdot |x - y|, \\ |f(s,x) - f(t,x)| &\le K \cdot (1 + |x|) \cdot |s - t|^{1/2} \end{aligned} \tag{5.9}$$

for all $s, t \in [0,T]$ and $x, y \in \mathbb{R}^d$. Here $|\cdot|$ is used to denote arbitrary norms on the corresponding spaces. For simplicity, we always take the Euler scheme with equidistant time-steps for approximation of X, and generally 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 $M = 2$,

and (5.9) implies (5.1) with $\beta = 1$; see, e.g., Müller-Gronbach [21, Thm. 4]. Hence we get (A4) with

$$\beta_4 = 2, \qquad \beta_5 = 1 \tag{5.10}$$

(see (5.2)), and (A5) with $\beta_1 = 1 + \varepsilon$, $\beta_2 = 1$, and $\beta_3 = 1/2 - \varepsilon$ under assumptions (S2) and (S3) or (S2) and (S5) on g; see (5.3). We apply (5.4) to obtain (A3) with

$$\alpha_1 = \varepsilon, \qquad \alpha_2 = 1/2, \qquad \alpha_3 = 1/2 - \varepsilon; \tag{5.11}$$

see also Kebaier [17, sect. 2.2].

We therefore have $q = 2 \cdot (r + 1) + \varepsilon$ and $\max(1, \beta_4/\beta_5) = 2$, and (2.14) in Theorem 2.6 yields

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

for the approximation of F on $[S_0, S_1]$. Likewise, (3.5) in Theorem 3.3 yields

$$(\gamma, \eta) = (2 + 4/r, 3)$$

for the approximation of ρ on $[S_0, S_1]$. Moreover, $\beta^\dagger = 2 + \varepsilon$, so that (4.7) in Theorem 4.3 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, we get $\gamma = 2$ in the limit $r \to \infty$, and proper multilevel algorithms are always superior to single-level algorithms; see Remarks 2.7, 3.4, and 4.4.

Remark 5.2. We refer the reader to Bally and Talay [4] for improved weak error estimates, which hold under stronger smoothness and nondegeneracy assumptions on the coefficients of the SDE; see conditions (UH) and (C) in [4]. In this case we obtain (A3) with

$$\alpha_1 = 0, \qquad \alpha_2 = 1, \qquad \alpha_3 = 1, \tag{5.12}$$

and furthermore, (2.6) holds with $\alpha = 1$. While (5.12) is superior to (5.11), it leads to essentially the same orders of convergence for approximation of densities or distribution functions if $r \geq 1$. For $r = 0$, we derive $\gamma = 3$ instead of $\gamma = 4$ and $\gamma = 5/2 + \varepsilon$ instead of $\gamma = 3 + \varepsilon$ for approximation of F on $[S_0, S_1]$ and at a single point, respectively, given (5.12) instead of (5.11).

In Remark 5.1 we compared the smoothing approach for the approximation of F at a single point with a direct approach in a general setting. Specifically, in the present setting and given (5.12) instead of (5.11), the direct approach achieves $\gamma' = 5/2 + \varepsilon$ instead of $\gamma' = 3 + \varepsilon$, so that the smoothing approach is superior to the direct approach if and only if $r \geq 2$. Let us note that conditions (UH) and (C) from [4] imply the existence of a density of X_T in $C_b^\infty(\mathbb{R}^d, \mathbb{R})$, so that r just depends on the properties of φ.

Remark 5.3. A two-level construction of the form

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

which is the counterpart of the two-level construction from Kebaier [17] for the approximation of $\mathrm{E}(\varphi(X_T))$, is employed by Kebaier and Kohatsu-Higa [18] 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 multidimensional counterpart to (A1) is satisfied for every $r \in \mathbb{N}_0$. Using Malliavin calculus techniques, they 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 multilevel approach achieves the order $\gamma = 2 + \varepsilon$ (at least for $d = 1$).

Remark 5.4. Consider the problem of approximating a quantile of Y, which is studied by Talay and Zheng [23] in the particular case of a projection $\varphi(x) = x_i$. By assumption, $\rho \in C_{\mathrm{b}}^{\infty}(\mathbb{R}, \mathbb{R})$. They employ a single-level algorithm that is based on a suitably regularized Euler scheme; cf. Remark 5.3. 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 multilevel algorithm to achieve the order $\gamma = 2 + \varepsilon$ also for quantile approximation and every Lipschitz continuous function φ. Furthermore, the multilevel algorithm may be used to approximate the distribution function F and the density ρ in parallel, which allows us to control the impact of inverting the approximation to F.

Remark 5.5. We comment on the optimality of the parameters α_i and β_i according to (5.12) and (5.10) in (A3) and (A4). Due to Bally and Talay [4], under their assumptions (UH) and (C) on the coefficients of the SDE, the estimate (A3) with (5.12) is sharp. Under standard assumptions, $2^{\ell/2} \cdot (X - X^{(\ell)})$ converges in distribution to a stochastic process U, with U_T being nondegenerate in general; see Jacod and Protter [16]. In the latter case we have a projection $\varphi(x) = x_i$ such that (5.1) with $M = 2$ and $p = 1$ does not hold for any $\beta > 1$. A slight generalization of Lemma 2.3 shows that (A4) does not hold for any $\beta_4 < 2$ or $\beta_5 > 1$. Hence, estimate (A4) with (5.10) cannot be improved in general for the Euler scheme.

The approximation of marginal densities of SDEs is 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 and Talay [5] and Gobet and Labart [12]. Milstein, Schoenmakers, and Spokoiny [20] 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 (5.9) implies the following upper bound; see, e.g., Müller-Gronbach

[21, Thm. 2]. For every $1 \le p < \infty$ there exists a constant $c_p > 0$ such that

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

for every $\ell \in \mathbb{N}$. Consequently, (5.1) holds with $\beta = 1 - \varepsilon$, and we get (A4) with

$$\beta_4 = 2, \qquad \beta_5 = 1 - \varepsilon \tag{5.13}$$

(see (5.2)), (A5) with $\beta_1 = 1 + \varepsilon$, $\beta_2 = 1 - \varepsilon$, and $\beta_3 = 1/2 - \varepsilon$ under the assumptions (S2) and (S3) or (S2) and (S5) (see (5.3)), as well as (A3) with

$$\alpha_1 = 0, \qquad \alpha_2 = 1/2 - \varepsilon, \qquad \alpha_3 = 1/2 - \varepsilon; \tag{5.14}$$

see (5.4). Furthermore, (2.6) holds with $\alpha = 1/2 - \varepsilon$; see (5.5).

We therefore have $q = 2 \cdot (r + 1) + \varepsilon$ and $\max(1, \beta_4/\beta_5) = 2 + \varepsilon$, and (2.13) in Theorem 2.6 yields

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

for the approximation of F on $[S_0, S_1]$. Likewise, (3.4) in Theorem 3.3 yields

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

for the approximation of ρ on $[S_0, S_1]$. Moreover, $\beta^\dagger = 2 + \varepsilon$, so that (4.5) in Theorem 4.3 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 multilevel algorithms are always superior to single-level algorithms; see Remarks 2.7, 3.4, and 4.4.

Note that section 5.1 deals 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 $r \ge 1$, and we always get $\gamma = 2$ in the limit $r \to \infty$.

Remark 5.6. We comment on the optimality of the parameters α_i and β_i according to (5.14) and (5.13) in (A3) and (A4). Due to Remark 5.5, estimate (A4) with (5.13) cannot be improved in general for the Euler scheme. Concerning (A3), we are not aware of an optimality result. However, we refer the reader to [1], which studies 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. The authors of [1] 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.,

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

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 $M = 2$. Furthermore, we suppose that we have an autonomous SDE with bounded and Lipschitz continuous drift and diffusion coefficients. The following upper bound is established by Bouchard, Geiss, and Gobet [6, Thm. 3.9] under suitable assumptions on the domain D. For every $1 \le p < \infty$ there exists a constant $c_p > 0$ such that

$$\|Y - Y^{(\ell)}\|_p \le c_p \cdot M^{-\ell/(2p)} \tag{5.15}$$

for every $\ell \in \mathbb{N}$. From (2.3) we get (A4) with

$$\beta_4 = 1, \qquad \beta_5 = 1/2, \tag{5.16}$$

and (4.1) and Lemma 4.1 yield (A5) with $\beta_1 = 1$, $\beta_2 = 1/2$, and $\beta_3 = 1/4$. Furthermore, (2.1) and Lemma 4.1 yield (A3) with

$$\alpha_1 = 1, \qquad \alpha_2 = 1/2, \qquad \alpha_3 = 1/4 \tag{5.17}$$

under assumptions (S2) and (S3) or (S2) and (S5), while (2.6) holds with $\alpha = 1/4$.

We therefore have $q = 2r + 4$ and $\max(1, \beta_4/\beta_5) = 2$, and (2.13) in Theorem 2.6 yields

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

for the approximation of F on $[S_0, S_1]$. Likewise, (3.4) in Theorem 3.3 yields

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

for the approximation of ρ on $[S_0, S_1]$. Moreover, $\beta^\dagger = 3$, so that (4.5) in Theorem 4.3 yields

$$(\gamma, \eta) = (3 + 2/(r+1), 0)$$

for the approximation of F at a single point $s \in [S_0, S_1]$. For all three problems, proper multilevel algorithms are superior to single-level algorithms for every $r \in \mathbb{N}_0$ (see Remarks 2.7, 3.4, and 4.4), 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 (5.15) 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 (5.15) 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 (2.29) and (4.14) are satisfied for every $r \ge 1$, and therefore smoothing already helps for the single-level algorithm to approximate the distribution function of the stopped exit time. We refer the reader to Gobet and Menozzi [13] for conditions that ensure the existence of a density of the exit time in $C^\infty([0,T[\,,\mathbb{R})$.

Remark 5.7. For the approximation of the mean $\mathrm{E}(Y)$ of the stopped exit time, a multilevel Euler algorithm has been constructed and analyzed in Higham et al. [15]. 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.

6. Numerical experiments. The main goal of our numerical experiments is to demonstrate the potential of the new multilevel 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 in 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 multilevel Monte Carlo approach with smoothing in comparison to the single-level Monte Carlo approach without smoothing. Furthermore, we compare the error of the multilevel 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 beyond the scope of the current paper and will be presented in a subsequent paper.

Consider a geometric Brownian motion X, given by

$$\begin{aligned} dX_t &= 0.05 \cdot X_t \, dt + 0.2 \cdot X_t \, dW_t, \qquad t \in [0, T], \\ X_0 &= 1, \end{aligned}$$

where W denotes a scalar Brownian motion. For the approximation of X, we use the Euler scheme with equidistant time-steps, so that $M = 2$. In the examples from this section, assumption (A1) holds for every $r \in \mathbb{N}$. In what follows, we employ (A1) with $r = 3, 5, 7, 9, 11$, and a particular purpose of the numerical experiments is to illustrate the impact of the different values of r. The values of the parameters α_i and β_i are presented later in each of the examples.

Given ϵ as well as the values of α_i, β_i, and r, we basically choose the remaining parameters of the multilevel (single-level) algorithm such that all four (three) terms in the upper bound (2.15) are of the order ϵ^2. For the multilevel algorithm with smoothing we choose the parameters L_0, L_1, and N_ℓ according to (2.26), (2.22), (2.19), and (2.27), while $\delta = 2^{-1/(r+1)} \cdot \epsilon^{1/(r+1)}$; cf. (2.17). For the single-level algorithm without smoothing (see Remark 2.5), we choose $L = L_0 = L_1$ and N_L according to (2.22) and (2.19), and note that (2.7) leads to $q = (r + 1)/\alpha$.

The second stage of the algorithm differs, however, from what we have presented and analyzed in sections 2–5. In the numerical experiments, we employ piecewise polynomial interpolation Q_k^3 of degree 3 with equidistant knots for any r. Due to the Lebesgue constants involved, this is preferable to Q_k^r with a large value of 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 single-level and multilevel cases, we therefore take $k = 3 \cdot \lceil 5 \cdot \epsilon^{-1/4} \cdot (S_1 - S_0)/3 \rceil + 1$; cf. (2.18).

To specify the computational gain, we compare the upper bound (2.16) for the cost of the multilevel Monte Carlo algorithm with smoothing to the corresponding upper bound $c(k, L, N) = N \cdot (2^L + k)$ for the cost of the single-level algorithm. Accordingly, the ratio $c(k, L_0, L_1, N_{L_0}, \ldots, N_{L_1})/c(k, L, N)$, which is a function of the desired accuracy ϵ, is used to describe the computational gain.

To assess the accuracy of the multilevel algorithm, $\mathrm{error}(Q_k^3(\mathcal{M}))$, which depends on ϵ and r, should be compared with the desired accuracy ϵ. Since $\mathrm{error}(Q_k^3(\mathcal{M}))$ is not known

exactly, we employ a simple Monte Carlo experiment with 25 independent replications for each of the values of r and each of the values $\epsilon = 2^{-i}$ for $i = 3, \dots, 11$. The estimate is denoted by $\mathrm{RMSE}(\epsilon, r)$. In the present approach we do not have an exact control of the error of the multilevel (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 aim only at $\mathrm{RMSE}(\epsilon, r)$ being reasonably close to ϵ.

6.1. Smooth path-independent functionals for SDEs. We set $T = 1$, and we approximate the distribution function

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

of X_T on the interval $[S_0, S_1] = [0, 2]$. Note that X_T is lognormally distributed with parameters $\mu - \sigma^2/2$ and σ^2. The parameters β_i are chosen according to (5.10). For the weak error estimates the result from [4] is not applicable, and we know only that (A3) holds with (5.11). Still, we choose the parameters according to (5.12), which is justified by the numerical experiments.

The computational gain as well as the replication numbers N_ℓ for the multilevel algorithm with $\epsilon = 2^{-11}$ are presented in Figure 1. The maximal level L_1 of the multilevel algorithm coincides with the level chosen by the single-level algorithm, and this level does not depend on r. For smaller values of r the multilevel algorithm start on a higher level L_0, and therefore the computational gain in the case $r = 3$ is only about a factor of 2. For large values of r, we observe a reasonable computational gain for moderate values of ϵ. We add that the computational gain would be much larger if (5.11) instead of (5.12) were employed. In Figure 1 we compare the estimate $\mathrm{RMSE}(\epsilon, r)$ for the error of the multilevel algorithm to the accuracy demand ϵ. Note that $\mathrm{RMSE}(\epsilon, r)$ is in the range of ϵ; actually, it is less than ϵ in almost all cases.

6.2. Smooth path-dependent functionals for SDEs. For this test case we choose the same value of T and the same interval $[S_0, S_1]$ as in section 6.1. We approximate the function

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

(see Shreve [22, 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 multilevel smoothing approach is applicable, and we choose the parameters α_i and β_i according to (5.14) and (5.13).

See Figure 2, with replication numbers for $\epsilon = 2^{-10}$, for the results. The main difference, compared to the previous section, is that 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 coincides with the level chosen by the single-level algorithm, is essentially twice 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 multilevel approach.

6.3. Stopped exit times for SDEs. We set $T = 2$, and we approximate the distribution function

$$F(s) = \mathrm{E}(1_{]\infty,s]}(\inf\{t \geq 0 : X_t = b\} \wedge T)),$$

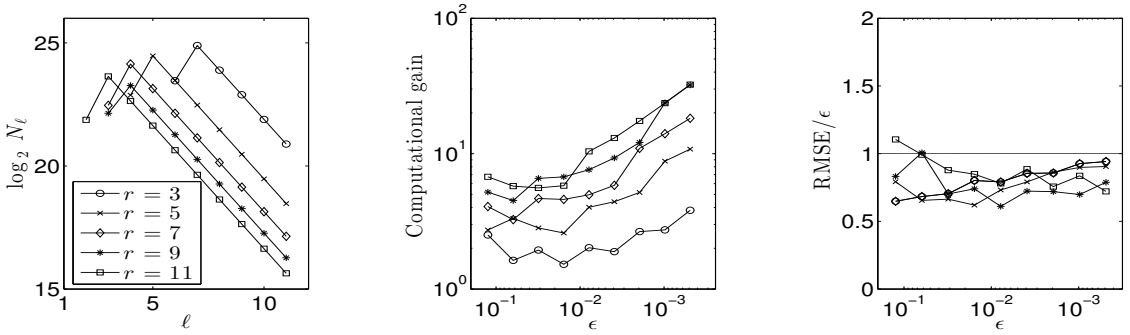

Figure 1. *Path-independent functional: Replication numbers, computational gain, and error vs. accuracy demand ϵ.*

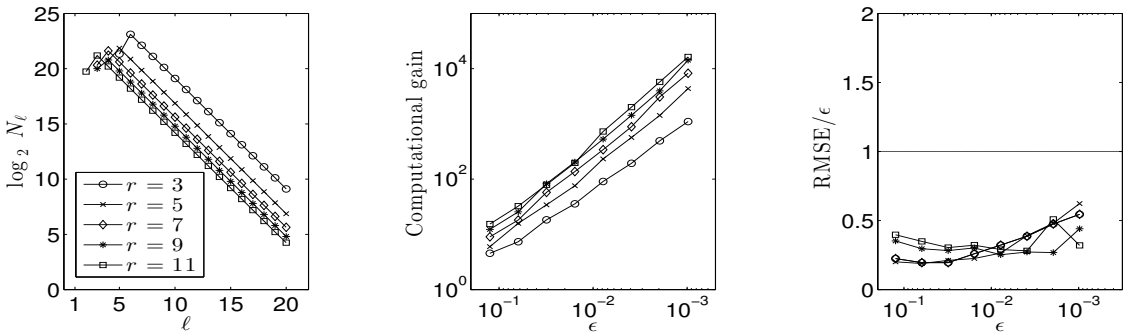

Figure 2. *Path-dependent functional: Replication numbers, computational gain, and error vs. accuracy demand ϵ.*

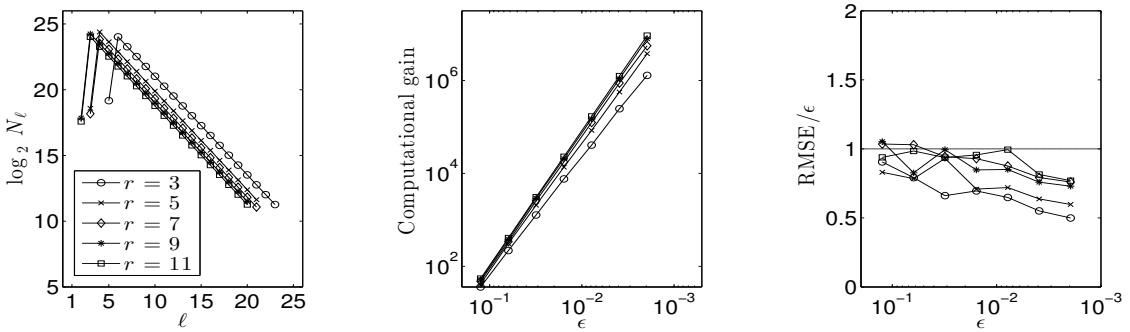

Figure 3. *Stopped exit time: Replication numbers, computational gain, and error vs. accuracy demand ϵ.*

with $b = 1.3$ 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 $T > S_1$. Since $b > X_0$, the drift and the diffusion coefficient may be redefined to achieve boundedness without any impact on the results. The parameters α_i and β_i are chosen according to (5.17) and (5.16).

See Figure 3, with replication numbers for $\epsilon = 2^{-9}$, 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 problem. Consequently, $L_1 = (2 + 2/(r + 1)) \cdot \log_2 \epsilon^{-1}$ is the maximal level for the multilevel 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 was inspired by a joint project with Oleg Iliev. The authors thank Oleg Iliev and Winfried Sickel for helpful discussions.

REFERENCES

[1] A. Alfonsi, B. Jourdain, and A. Kohatsu-Higa, *Pathwise optimal transport bounds between a one-dimensional diffusion and its Euler scheme*, Ann. Appl. Probab., 24 (2014), pp. 1049–1080.

[2] M. Altmayer and A. Neuenkirch, *Multilevel Monte Carlo quadrature of discontinuous payoffs in the generalized Heston model using Malliavin integration by parts*, SIAM J. Financial Math., 6 (2015), pp. 22–52.

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

[4] V. Bally and D. Talay, *The law of the Euler scheme for stochastic differential equations:* I. *Convergence rate of the distribution function*, Probab. Theory Related Fields, 104 (1996), pp. 43–60.

[5] V. Bally and D. Talay, *The law of the Euler scheme for stochastic differential equations:* II. *Convergence rate of the density*, Monte Carlo Methods Appl., 2 (1996), pp. 93–128.

[6] B. Bouchard, S. Geiss, and E. Gobet, *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, 2013.

[7] C. de Boor, *A Practical Guide to Splines*, Appl. Math. Sci. 27, Springer, Berlin, 1978.

[8] M. B. Giles, *Improved multilevel Monte Carlo convergence using the Milstein scheme*, in Monte Carlo and Quasi-Monte Carlo Methods 2006, A. Keller, S. Heinrich, and H. Niederreiter, eds., Springer, Berlin, 2008, pp. 343–358.

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

[10] M. B. Giles, K. Debrabant, and A. Rössler, *Numerical Analysis of Multilevel Monte Carlo Path Simulation Using the Milstein Discretisation*, preprint, arXiv:1302.4676, 2013.

[11] M. B. Giles, D. J. Higham, and X. Mao, *Analysing multilevel Monte Carlo for options with non-globally Lipschitz payoff*, Finance Stoch., 13 (2009), pp. 403–413.

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

[13] E. Gobet and S. Menozzi, *Exact approximation rate of killed hypoelliptic diffusions using the discrete Euler scheme*, Stochastic Process. Appl., 112 (2004), pp. 201–223.

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

[15] D. J. Higham, X. Mao, M. Roj, Q. Song, and G. Yin, *Mean exit times and the multi-level Monte Carlo method*, SIAM/ASA J. Uncertainty Quantification, 1 (2013), pp. 2–18.

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

[17] A. Kebaier, *Statistical Romberg extrapolation: A new variance reduction method and applications to options pricing*, Ann. Appl. Probab., 14 (2005), pp. 2681–2705.

[18] A. Kebaier and A. Kohatsu-Higa, *An optimal control variance reduction method for density estimation*, Stochastic Process. Appl., 118 (2008), pp. 2143–2180.

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

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

[21] T. Müller-Gronbach, *The optimal uniform approximation of systems of stochastic differential equations*, Ann. Appl. Probab., 12 (2002), pp. 664–690.
[22] S. E. Shreve, *Stochastic Calculus for Finance* II: *Continuous-Time Models*, Springer-Verlag, New York, 2004.
[23] D. Talay and Z. Zheng, *Approximation of quantiles of components of diffusion processes*, Stochastic Process. Appl., 109 (2004), pp. 23–46.
[24] A. B. Tsybakov, *Introduction to Nonparametric Estimation*, Springer, Berlin, 2009.