SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 6, No. 3, pp. 1256–1293

Multilevel Monte Carlo Approximation of Functions*

S. Krumscheid† and F. Nobile†

Abstract. Many applications across sciences and technologies require a careful quantification of nondeterministic effects to a system output, for example, when evaluating the system's reliability or when gearing it towards more robust operation conditions. At the heart of these considerations lies an accurate characterization of uncertain system outputs. In this work we introduce and analyze novel multilevel Monte Carlo techniques for an efficient characterization of an uncertain system output's distribution. These techniques rely on accurately approximating general parametric expectations, i.e., expectations that depend on a parameter, uniformly on an interval. Applications of interest include, for example, the approximation of the characteristic function and of the cumulative distribution function of an uncertain system output. A further important consequence of the introduced approximation techniques for parametric expectations (i.e., for functions) is that they allow us to construct multilevel Monte Carlo estimators for various robustness indicators, such as for a quantile (also known as value-at-risk) and for the conditional value-at-risk. These robustness indicators cannot be expressed as moments and are thus not usually easily accessible. In fact, here we provide a framework that allows us to simultaneously estimate a cumulative distribution function, a quantile, and the associated conditional value-at-risk of an uncertain system output at the cost of a single multilevel Monte Carlo simulation, while each estimated quantity satisfies a prescribed tolerance goal.

Key words. multilevel Monte Carlo, parametric expectation, distribution function, quantiles, conditional value-at-risk, characteristic function, moments

AMS subject classifications. 65C05, 60H35, 65C20, 65Y20, 60E10

DOI. 10.1137/17M1135566

1. Introduction.

Parametric expectations, such as $\Phi(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big)$ for some random variable Q, are commonly used in a wide range of applications, for example, when characterizing the distribution of an uncertain system output Q. Here, perhaps most notably, are applications in which the cumulative distribution function $F_Q(\vartheta) = \mathbb{E}\big(I(Q \le \vartheta)\big)$ or the characteristic function $\varphi_Q(\vartheta) = \mathbb{E}\big(\exp(\mathrm{i}\vartheta Q)\big)$ of the random variable Q are sought after on some interval Θ. But also many problems arising in the field of stochastic optimization, such as

$$\min_{\vartheta\in\Theta}\left(\mathbb{E}\big(\phi(\vartheta,Q)\big) + c\sqrt{\mathbb{E}\Big(\phi(\vartheta,Q)^2\Big) - \mathbb{E}\big(\phi(\vartheta,Q)\big)^2}\right),$$

*Received by the editors June 21, 2017; accepted for publication (in revised form) June 27, 2018; published electronically September 27, 2018.

http://www.siam.org/journals/juq/6-3/M113556.html

Funding: This work has been supported by CADMOS (Center for Advanced Modeling Science). The second author also acknowledges funding from the EU-FP7 framework program within the project UMRIDA under grant agreement ACP3-GA-2013-605036.

†CSQI, Institute of Mathematics, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (sebastian.krumscheid@epfl.ch, fabio.nobile@epfl.ch).

rely heavily on accurate and computationally affordable approximations of these parametric expectations. Consequently, an efficient approximation of parametric expectations, such as $\Phi(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big)$ with $\vartheta \in \Theta$ uniformly on the interval Θ, is of fundamental interest.

In many situations sampling Q requires the solution of a complex problem (e.g., stochastic differential equation, stochastic/random partial differential equation, etc.), which inevitably involves a discretization step and can only be done up to a prescribed tolerance level. For the approximation of moments of a random variable Q based on such approximate samples, the multilevel Monte Carlo method [11, 17, 18, 33] has been established as a computationally efficient sampling method that is applicable to a wide range of applications. However, its applicability for general parametric expectations and quantities that cannot be expressed as moments is not straightforward and requires a special treatment. For example, recently multilevel Monte Carlo techniques have been incorporated into stochastic approximation algorithms used for stochastic optimization in the context of diffusion processes [16]. In this work, we present a multilevel Monte Carlo methodology that provides a uniform approximation of general parametric expectations, in other words, of a function. Moreover, we carefully analyze the proposed methodology and provide a full complexity analysis.

Somewhat related topics have already been addressed in [24, 25], which initiated the general multilevel Monte Carlo research. One of the main differences compared to the present work is that the work in [25], in particular, considers the case where exact sampling from the law of Q is possible and the construction of the multilevel hierarchy is based on different interpolation grids. Conversely, in this work we do not assume that sampling from the law of Q is possible, and we construct the multilevel hierarchy based on different approximations to the law of Q combined with different interpolation grids instead. In fact, the present work is somewhat closer to the work presented in [20], where the authors discuss multilevel Monte Carlo ideas for the uniform approximation of a random variable's cumulative distribution function (CDF) using a single interpolation grid. Here, we build upon ideas presented in that work, but we extend and generalize these further to approximate general parametric expectations and exploit a suitably chosen hierarchy of interpolation grids. A direct implication of the greater generality of our work is, for example, that it enables us to derive novel multilevel Monte Carlo estimators for the characteristic function in addition to CDF approximations. This is particularly useful when characterizing a random variable's distribution in the presence of atoms (i.e., for mixed distribution) or in cases when not all moments exist (e.g., Lévy distribution). A further important consequence of the results presented here is that they provide multilevel Monte Carlo estimators for derived quantities, such as for quantiles (also known as values-at-risk) or for the conditional value-at-risk. It is noteworthy that these quantities cannot be expressed as moments. Consequently, they have been out of reach for an efficient treatment via standard multilevel Monte Carlo methods until recently. In fact, first results in this direction, at least for quantiles, are available through the recent works on multilevel stochastic approximation algorithms [13, 16]. Moreover, the recent study [19] introduces an efficient estimator for the (conditional) value-at-risk in the context of a diffusion process by combining a stochastic approximation algorithm with multilevel Monte Carlo techniques and nested expectations. Although these first results are certainly insightful, this research direction is still in its infancy, so that efficient multilevel Monte Carlo quantile estimators applicable to a wide class of problems are still of major interest.

The rest of the paper is organized as follows. In section 2 we present the multilevel Monte Carlo estimator for parametric expectations in a general setting and provide the corresponding complexity analysis. In sections 3 and 4 we apply these abstract results to two different scenarios and illustrate the theoretical findings with numerical examples. Specifically, in section 3 we present a novel multilevel Monte Carlo estimator for the characteristic function of a random variable, while we present and analyze an approximation to a random variable's CDF in section 4. Moreover, in section 4 we also present multilevel Monte Carlo estimators for various derived estimators, such as a quantile and a conditional value-at-risk. Finally, section 5 offers a summary and a discussion of our results.

2. Approximation of parametric expectations on compact intervals. Throughout this work, we consider a real-valued random variable Q defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. As anticipated in the introduction, we assume that it is not possible to sample from the distribution of Q exactly. Instead, we assume that one can only draw samples $Q_\ell^{(i)}$ from a random variable Q_ℓ, which is a suitable approximation to Q and which is defined on the same probability space. Specifically, we consider a hierarchy of approximations on different levels $\ell = 0, 1, \dots$ in the sense that the level ℓ approximation Q_ℓ of Q corresponds to some discretization parameter h_ℓ, and the different approximation levels are related by $h_{\ell-1} = s h_\ell$ for some $s > 1$, so that $h_\ell = s^{-\ell} h_0$ for $\ell \geq 0$ and $Q_\ell \to Q$ in an appropriate sense (specified below) as $\ell \to \infty$. For example, Q_ℓ could be derived from an approximate solution to a stochastic/random (partial) differential equation that is obtained via a numerical scheme with discretization parameter h_ℓ. Based on this multilevel hierarchy we aim at approximating

$$\Phi(\vartheta) := \mathbb{E}\big(\phi(\vartheta, Q)\big)$$

uniformly on some compact interval $\Theta \subset \mathbb{R}$ for a given function $\phi\colon \Theta \times \mathbb{R} \to \mathbb{R}$. The multilevel Monte Carlo approximation of Φ on Θ is obtained by first evaluating ϕ in a set of nodes in Θ by a standard multilevel Monte Carlo estimator and then appropriately interpolating the collected values to obtain a function on Θ. More precisely, let

$$\theta := (\theta_1, \theta_2, \dots, \theta_n)^T \in \Theta^n$$

denote the set of $n \in \mathbb{N}$ deterministic nodes. With a slight abuse of notation, we denote by $f(\theta)$ the vector with components $f(\theta_j)$, $j = 1, \dots, n$, for any function $f\colon \mathbb{R} \to \mathbb{R}$. Furthermore, let $L \in \mathbb{N}_0$ and $N := (N_0, N_1, \dots, N_L)^T \in \mathbb{N}^{L+1}$. The collection of the pointwise multilevel Monte Carlo estimators is then denoted by $\bar{\Phi}_L^N\colon \mathbb{R}^n \to \mathbb{R}^n$, which is given by

$$\begin{aligned}\bar{\Phi}_L^N(\theta) &:= \left[\frac{1}{N_0}\sum_{i=1}^{N_0} \phi\Big(\theta_j, Q_0^{(i,0)}\Big) + \sum_{\ell=1}^{L} \frac{1}{N_\ell}\sum_{i=1}^{N_\ell} \phi\Big(\theta_j, Q_\ell^{(i,\ell)}\Big) - \phi\Big(\theta_j, Q_{\ell-1}^{(i,\ell)}\Big)\right]_{1\leq j\leq n} \\ &= \left[\sum_{\ell=0}^{L} \frac{1}{N_\ell}\sum_{i=1}^{N_\ell} \phi\Big(\theta_j, Q_\ell^{(i,\ell)}\Big) - \phi\Big(\theta_j, Q_{\ell-1}^{(i,\ell)}\Big)\right]_{1\leq j\leq n}\end{aligned}$$

upon setting $\phi(\cdot, Q_{-1}) \equiv 0$. Here, $\Big(Q_\ell^{(i,\ell)}, Q_{\ell-1}^{(i,\ell)}\Big)$, $i = 1, \dots, N_\ell$, denote the independently and identically distributed (i.i.d.) samples from $(Q_\ell, Q_{\ell-1})$ that are also mutually independent

across levels. The extension (e.g., by means of interpolation) of this collection of pointwise estimators to a function on Θ is eventually achieved by

$$\hat{\Phi}_L^{N,n} := \mathcal{I}_n\big(\bar{\Phi}_L^N(\theta)\big)\,, \tag{1}$$

where $\mathcal{I}_n$ denotes an appropriate extension operator. The accuracy of this extension depends, of course, on the regularity of Φ. In this work we will consider two different scenarios: $\Phi \in C^{k+1}(\Theta)$ for some $k \in \mathbb{N}_0$ or Φ being analytic. In the first case we will consider extension operators that satisfy the following assumptions.

Assumption A1 (extension operator). *Let $k \in \mathbb{N}_0$ be given. The sequence of linear extension operators $\mathcal{I}_n \colon \mathbb{R}^n \to L^\infty(\Theta)$ based on the set of nodes $\theta \in \Theta^n$ satisfies the following:*

(i) $\|f - \mathcal{I}_n\big(f(\theta)\big)\|_{L^\infty(\Theta)} \leq c_1 n^{-(k+1)}$ *for any $f \in C^{k+1}(\Theta)$ with $c_1 \equiv c_1(f)$;*

(ii) $\|\mathcal{I}_n(x)\|_{L^\infty(\Theta)} \leq c_2 \|x\|_{\ell^\infty}$ *for any $x \in \mathbb{R}^n$;*

(iii) *the cost for computing $\mathcal{I}_n(x)$ based on n fixed nodes is uniformly bounded by $c_3 n$*

for all $n \in \mathbb{N}$. Here, the constants $c_1, c_2, c_3 > 0$ are independent of n.

For an appropriate analytic function Φ, we will consider extension operators with the following properties.

Assumption A2 (extension operator for analytic functions). *The sequence of linear extension operators $\mathcal{I}_n \colon \mathbb{R}^n \to L^\infty(\Theta)$ based on the set of nodes $\theta \in \Theta^n$ satisfies the following:*

(i) $\|f - \mathcal{I}_n\big(f(\theta)\big)\|_{L^\infty(\Theta)} \leq c_1 \rho^{-n}$ *for any suitable function f that is analytic in Θ and for which $\rho > 1$ and $c_1 \equiv c_1(f)$;*

(ii) $\|\mathcal{I}_n(x)\|_{L^\infty(\Theta)} \leq c_2 \ln(n) \|x\|_{\ell^\infty}$ *for any $x \in \mathbb{R}^n$;*

(iii) *the cost for computing $\mathcal{I}_n(x)$ based on n fixed nodes is uniformly bounded by $c_3 n \ln(n)$*

for all $n \in \mathbb{N}$, with constants $c_1, c_2, c_3 > 0$ independent of n.

Remark 2.1. Assumption A1 is, for example, satisfied for an interpolation with piecewise polynomials (i.e., splines) of degree k on a uniform grid [12, Chap. XII]. An example for which Assumption A2 holds true is that of polynomial interpolation in Chebyshev points on $\Theta = [-1, 1]$, say, of a function f that is analytic on Θ and analytically continuable to the closed Bernstein ellipse with elliptical radius $\rho > 1$ [34, Chap. 8].

The error of the multilevel approximation $\hat{\Phi}_L^{N,n}$ is quantified through the mean squared error

$$\mathrm{MSE}\left(\hat{\Phi}_L^{N,n}\right) := \mathbb{E}\Big(\big\|\hat{\Phi}_L^{N,n} - \Phi\big\|_{L^\infty(\Theta)}^2\Big)\,.$$

The asymptotic analysis that will follow is partly based on considering an increasing number of nodes in Θ, in the sense that $\theta \in \mathbb{R}^n$ with $n \to \infty$ as the prescribed mean squared error tolerance tends to zero. It is therefore necessary (and natural) to consider elements of $\mathbb{R}^n$ as elements of an appropriate sequence space, which we will then use to investigate the statistical properties. Specifically, it is convenient to work in the Banach space ℓ^∞. For a random variable ξ with values in ℓ^∞ and finite second moment the variance is defined as

$$\mathrm{Var}\,(\xi) := \mathbb{E}\big(\|\xi - \mathbb{E}(\xi)\|_{\ell^\infty}^2\big)\,.$$

It follows that

$$\mathbb{E}\big(\|x-\xi\|_{\ell^\infty}^2\big) \le 2\|x-\mathbb{E}(\xi)\|_{\ell^\infty}^2 + 2\,\mathrm{Var}\,(\xi) \tag{2}$$

for any deterministic $x \in \mathbb{R}^n$ and random variable $\xi \in \mathbb{R}^n$. Moreover, it holds that $\mathrm{Var}\,(\xi) \le 4\mathbb{E}\big(\|\xi\|_{\ell^\infty}^2\big)$. In this Banach space setting, the formula for the variance of a sum of real-valued independent random variables becomes an inequality. In fact, for a sequence $\big(\xi_i\big)_{1\le i\le N}$ of mutually independent $\mathbb{R}^n$-valued random variables it holds that

$$\mathrm{Var}\left(\sum_{i=1}^N \xi_i\right) \le c\ln(n)\sum_{i=1}^N \mathrm{Var}\,(\xi_i)\ , \tag{3}$$

where $c>0$ is a generic constant [29] (alternatively [28, Chap. 9]). See also [24, Lem. 1], where this inequality has been used in the context of a multilevel method.

Before we can characterize the computational complexity of the multilevel Monte Carlo approximation $\hat\Phi_L^{N,n}$, we need to specify the cost model for evaluating the function ϕ. Throughout this work we assume that the computational cost of evaluating $\phi(\vartheta,q)$ is bounded by a constant for any $(\vartheta,q)\in\Theta\times\mathbb{R}$. Moreover, we denote by $\mathfrak{c}_\ell$ the computational cost for generating a sample $(Q_\ell,Q_{\ell-1})$, on level $\ell\in\mathbb{N}_0$. Finally, throughout this work we will use the notation $a\lesssim b$ if there exists a constant c so that $a\le cb$ as $a,b\to\infty$ (or $a,b\to 0$); analogously for $\gtrsim$. Then the following result holds.

Theorem 2.1. *Let $\Phi(\vartheta)=\mathbb{E}\big(\phi(\vartheta,Q)\big)$. Suppose there exist constants $\alpha,\beta,\gamma>0$ such that $2\alpha\ge\min(\beta,\gamma)$ and*

(i) $\sup_{\vartheta\in\Theta}\big|\mathbb{E}\big(\phi(\vartheta,Q)-\phi(\vartheta,Q_\ell)\big)\big| \le c_1 h_\ell{}^\alpha$;

(ii) $\mathbb{E}\Big(\sup_{\vartheta\in\Theta}\big|\phi(\vartheta,Q_\ell)-\phi(\vartheta,Q_{\ell-1})\big|^2\Big)\le c_2 h_\ell{}^\beta$;

(iii) $\mathfrak{c}_\ell \le c_3 h_\ell{}^{-\gamma}$

for all $\ell\in\mathbb{N}_0$ with positive constants c_1,c_2,c_3 independent of ℓ.

If $\Phi\in C^{k+1}(\Theta)$ for some $k\in\mathbb{N}_0$, and if Assumption A1 is satisfied for that k, then for any $\varepsilon>0$ there exist parameters $L\in\mathbb{N}_0$, $n\in\mathbb{N}$, and $N\in\mathbb{N}^{L+1}$ such that the corresponding multilevel Monte Carlo estimator $\hat\Phi_L^{N,n}$ satisfies

$$\mathbb{E}\Big(\big\|\hat\Phi_L^{N,n}-\Phi\big\|_{L^\infty(\Theta)}^2\Big) = \mathcal{O}(\varepsilon^2)\ . \tag{4}$$

Furthermore, for any $0<\varepsilon<e^{-1}$ the associated computational cost $\mathfrak{C}\big(\hat\Phi_L^{N,n}\big)$ is bounded by

$$\mathfrak{C}\big(\hat\Phi_L^{N,n}\big) \lesssim \varepsilon^{-\left(2+\frac{1}{k+1}\right)}\ln\big(\varepsilon^{-1}\big) + \ln\big(\varepsilon^{-1}\big)\begin{cases}\varepsilon^{-2}\ln\big(\varepsilon^{-1}\big)^2 & \text{if } \beta=\gamma\ ,\\ \varepsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)} & \text{if } \beta<\gamma\ ,\\ \varepsilon^{-2} & \text{if } \beta>\gamma\ .\end{cases}$$

If $\Phi\colon\Theta\to\mathbb{R}$ is analytic in Θ and if Assumption A2 holds for Φ with $\rho>1$, then for any $\varepsilon>0$ there exist parameters $L\in\mathbb{N}_0$, $n\in\mathbb{N}$, and $N\in\mathbb{N}^{L+1}$ such that the corresponding estimator $\hat\Phi_L^{N,n}$ satisfies (4). Moreover, for any $0<\varepsilon<\min(\rho^{-1},\ln(2))$ the required computational cost is bounded by

$$\mathfrak{C}\big(\hat{\Phi}_L^{N,n}\big) \lesssim \varepsilon^{-2}\ln\big(\varepsilon^{-1}\big)^4 + \ln\big(\varepsilon^{-1}\big)^3 \begin{cases} \varepsilon^{-2}\ln(\varepsilon^{-1})^2 & \text{if } \beta = \gamma\,, \\ \varepsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)}\ln(\varepsilon^{-1})^{\frac{\gamma-\beta}{\alpha}} & \text{if } \beta < \gamma\,, \\ \varepsilon^{-2} & \text{if } \beta > \gamma\,. \end{cases}$$

Proof. The mean squared error can be bounded by

$$\begin{aligned} \mathrm{MSE}\left(\hat{\Phi}_L^{N,n}\right) &\equiv \mathbb{E}\Big(\big\|\Phi - \mathcal{I}_n\big(\Phi(\theta)\big) + \mathcal{I}_n\big(\Phi(\theta) - \bar{\Phi}_L^N(\theta)\big)\big\|_{L^\infty(\Theta)}^2\Big) \\ &\le 2\bigg(\big\|\Phi - \mathcal{I}_n\big(\Phi(\theta)\big)\big\|_{L^\infty(\Theta)}^2 + \mathbb{E}\Big(\big\|\mathcal{I}_n\big(\Phi(\theta) - \bar{\Phi}_L^N(\theta)\big)\big\|_{L^\infty(\Theta)}^2\Big)\bigg)\,. \end{aligned} \tag{5}$$

To bound the right-hand side of (5) further, we will treat the two regularity classes for Φ separately.

We begin with the finite regularity case, that is, $\Phi \in C^{k+1}(\Theta)$ for some $k \in \mathbb{N}_0$. As a consequence of the properties of the extension operator $\mathcal{I}_n$ satisfying Assumption A1, it follows that

$$\begin{aligned} \mathrm{MSE}\left(\hat{\Phi}_L^{N,n}\right) &\le 2c\Big(n^{-2(k+1)} + \mathbb{E}\Big(\big\|\Phi(\theta) - \bar{\Phi}_L^N(\theta)\big\|_{\ell^\infty}^2\Big)\Big) \\ &\le 2c\Big(n^{-2(k+1)} + 2\big\|\Phi(\theta) - \mathbb{E}\big(\bar{\Phi}_L^N(\theta)\big)\big\|_{\ell^\infty}^2 + 2\,\mathrm{Var}\left(\bar{\Phi}_L^N(\theta)\right)\Big)\,, \end{aligned}$$

where we have used inequality (2). That is, the mean squared error can be decomposed into three terms: one controlling the interpolation error, one the bias, and one the variance. From hypothesis (i) of Theorem 2.1, we find that the bias is bounded by

$$\begin{aligned} \big\|\Phi(\theta) - \mathbb{E}\big(\bar{\Phi}_L^N(\theta)\big)\big\|_{\ell^\infty} &= \max_{1\le j\le n}\big|\mathbb{E}\big(\phi(\theta_j, Q) - \phi(\theta_j, Q_L)\big)\big| \\ &\le \sup_{\vartheta\in\Theta}\big|\mathbb{E}\big(\phi(\vartheta, Q) - \phi(\vartheta, Q_L)\big)\big| \le c_1 {h_0}^\alpha s^{-\alpha L}\,. \end{aligned}$$

Similarly, for the variance term we find that

$$\begin{aligned} \mathrm{Var}\left(\bar{\Phi}_L^N(\theta)\right) &\le c\ln(n)\sum_{\ell=0}^{L}\frac{\mathrm{Var}\Big(\phi(\theta, Q_\ell) - \phi(\theta, Q_{\ell-1})\Big)}{N_\ell} \\ &\le 4c\ln(n)\sum_{\ell=0}^{L}\frac{\mathbb{E}\Big(\sup_{\vartheta\in\Theta}\big|\phi(\vartheta, Q_\ell) - \phi(\vartheta, Q_{\ell-1})\big|^2\Big)}{N_\ell} \le 4cc_2{h_0}^\beta\ln(n)\sum_{\ell=0}^{L}\frac{s^{-\beta\ell}}{N_\ell}\,, \end{aligned}$$

where we have used inequality (3), the fact that $\mathrm{Var}\,(\xi) \le 4\mathbb{E}\big(\|\xi\|_{\ell^\infty}^2\big)$, and hypothesis (ii) of Theorem 2.1. Combining these bounds, we eventually obtain

$$\mathrm{MSE}\left(\hat{\Phi}_L^{N,n}\right) \lesssim n^{-2(k+1)} + s^{-2\alpha L} + \ln(n)\sum_{\ell=0}^{L}\frac{s^{-\beta\ell}}{N_\ell}\,. \tag{6}$$

In view of this bound, we choose

$$n = \left\lceil \varepsilon^{-\frac{1}{k+1}} \right\rceil \quad \text{and} \quad L = \left\lceil \alpha^{-1} \log_s\left(\varepsilon^{-1}\right) \right\rceil$$

to obtain a contribution of $\mathcal{O}(\varepsilon^2)$ originating from the first and the second term in (6), respectively. Notice that $n \geq 2$ and $L \geq 1$, since $\varepsilon < 1$. Moreover, it holds that

$$\varepsilon^{-\frac{\mu}{\alpha}} \leq s^{\mu L} < s^{\mu} \varepsilon^{-\frac{\mu}{\alpha}} \tag{7}$$

for any $\mu > 0$. The computational cost $\mathfrak{C}\big(\hat{\Phi}_L^{N,n}\big)$ of constructing the estimator $\hat{\Phi}_L^{N,n}$ is given by the sum of the cost for computing all pointwise estimators and the cost for computing the extension to a function. That is, for a generic constant c, the cost is bounded by

$$\mathfrak{C}\big(\hat{\Phi}_L^{N,n}\big) \leq c\left(\sum_{\ell=0}^{L} N_\ell(\mathfrak{c}_\ell + n) + n\right) \lesssim n + \sum_{\ell=0}^{L} N_\ell\big(s^{\ell\gamma} + n\big) \, .$$

Treating the variables $N = (N_0, N_1, \ldots, N_L)$ as continuous and minimizing the cost $\mathfrak{C}\big(\hat{\Phi}_L^{N,n}\big)$ with respect to N, subject to the constraint $\sum_{\ell=0}^{L} \frac{s^{-\beta\ell}}{N_\ell} = \varepsilon^2 \ln(n)^{-1}$, eventually implies

$$N_l = \left\lceil \varepsilon^{-2} \ln(n) \sqrt{\frac{s^{-l\beta}}{s^{l\gamma} + n}} \sum_{\ell=0}^{L} \sqrt{s^{-\ell\beta}\big(s^{\ell\gamma} + n\big)} \right\rceil \, , \quad 0 \leq l \leq L \, . \tag{8}$$

Consequently, the mean squared error is $\mathcal{O}(\varepsilon^2)$ as asserted. Bounding the number of samples per level by $N_\ell \leq 1 + \varepsilon^{-2} \ln(n) \sqrt{s^{-l\beta}\big(s^{l\gamma} + n\big)^{-1}} \sum_{\ell=0}^{L} \sqrt{s^{-\ell\beta}\big(s^{\ell\gamma} + n\big)}$, which is a consequence of (8), the corresponding computational cost is bounded by

$$\begin{aligned}
\mathfrak{C}\big(\hat{\Phi}_L^{N,n}\big) &\lesssim n(L+2) + \sum_{\ell=0}^{L} s^{\ell\gamma} + \varepsilon^{-2} \ln(n) \left(\sum_{\ell=0}^{L} \sqrt{s^{-\ell\beta}\big(s^{\ell\gamma} + n\big)}\right)^2 \\
&\lesssim n(L+2) + \frac{s^{\gamma L}}{1 - s^{-\gamma}} + \varepsilon^{-2} \ln(n) \left(\sum_{\ell=0}^{L} \sqrt{s^{-\ell\beta}\big(s^{\ell\gamma} + n\big)}\right)^2 \\
&\lesssim \varepsilon^{-\frac{1}{k+1}} \ln\big(\varepsilon^{-1}\big) + \varepsilon^{-\frac{\gamma}{\alpha}} + \varepsilon^{-2} \ln\big(\varepsilon^{-1}\big) \left(\sum_{\ell=0}^{L} \sqrt{s^{-\ell\beta}\big(s^{\ell\gamma} + n\big)}\right)^2 ,
\end{aligned}$$

where we have also used the bound in (7). To quantify the sum, we introduce $L^* = \left\lceil \gamma^{-1} \log_s(n) \right\rceil - 1 \geq 0$, which implies that $s^{\ell\gamma} < n$ for all $\ell \leq L^*$. Now, we distinguish two cases. Firstly, we consider $L \leq L^*$, which, in the absence of rounding errors, implies $\alpha > \gamma(k+1)$. In that case it holds that

$$\sum_{\ell=0}^{L} \sqrt{s^{-\ell\beta}\big(s^{\ell\gamma} + n\big)} \leq \sqrt{2n} \sum_{\ell=0}^{L} s^{-\ell\frac{\beta}{2}} = \sqrt{2n} \frac{s^{\frac{\beta}{2}} - s^{-L\frac{\beta}{2}}}{s^{\frac{\beta}{2}} - 1} \lesssim \sqrt{n} \, .$$

Secondly, for $L > L^*$ we find that

$$\sum_{\ell=0}^{L} \sqrt{s^{-\ell\beta}\left(s^{\ell\gamma}+n\right)} \le c\left(\sqrt{n} + \sum_{\ell=L^*+1}^{L} s^{\ell\frac{\gamma-\beta}{2}}\right) \lesssim \sqrt{n} + \begin{cases} L+1\,, & \beta=\gamma\,,\\ s^{\frac{\gamma-\beta}{2}L}\,, & \beta<\gamma\,,\\ s^{\frac{\gamma-\beta}{2}L^*}\,, & \beta>\gamma\,. \end{cases}$$

Notice that the right-hand side above is also dominating the sum in the case $L \le L^*$. Collecting all the parts together, we eventually find

$$\mathfrak{C}\left(\hat{\Phi}_L^{N,n}\right) \lesssim \varepsilon^{-\frac{1}{k+1}} \ln\left(\varepsilon^{-1}\right) + \varepsilon^{-\frac{\gamma}{\alpha}} + \varepsilon^{-\left(2+\frac{1}{k+1}\right)} \ln\left(\varepsilon^{-1}\right) + \varepsilon^{-2} \ln\left(\varepsilon^{-1}\right) \begin{cases} \ln\left(\varepsilon^{-1}\right)^2\,, & \beta=\gamma\,,\\ \varepsilon^{-\frac{\gamma-\beta}{\alpha}}\,, & \beta<\gamma\,,\\ \varepsilon^{\frac{\beta-\gamma}{\gamma(k+1)}}\,, & \beta>\gamma\,. \end{cases}$$

Using the hypothesis $2\alpha \ge \min(\beta,\gamma)$, the claim follows, and the proof for $\Phi \in C^{k+1}(\Theta)$ is complete.

Consider now $\Phi\colon \Theta \to \mathbb{R}$ analytic in Θ and such that Assumption A2 holds for Φ with $\rho > 1$. The proof of the claim in this case is very similar to the previous one. Indeed, starting from inequality (5), the mean squared error can now be bounded by

$$\begin{aligned} \mathrm{MSE}\left(\hat{\Phi}_L^{N,n}\right) &\le c\left(\rho^{-2n} + \ln(n)^2 \left\|\Phi(\theta) - \mathbb{E}\left(\bar{\Phi}_L^N(\theta)\right)\right\|_{\ell^\infty}^2 + \ln(n)^2\,\mathrm{Var}\left(\bar{\Phi}_L^N(\theta)\right)\right)\\ &\lesssim \rho^{-2n} + \ln(n)^2 s^{-2\alpha L} + \ln(n)^3 \sum_{\ell=0}^{L} \frac{s^{-\beta\ell}}{N_\ell}\,, \end{aligned}$$

where we have used the same steps that led to (6) before. Now, choosing

$$n = \left\lceil \log_\rho\left(\varepsilon^{-1}\right)\right\rceil \quad \text{and} \quad L = \left\lceil \alpha^{-1} \log_s\left(\varepsilon^{-1}\ln(n)\right)\right\rceil$$

and then minimizing the cost bound

$$\mathfrak{C}\left(\hat{\Phi}_L^{N,n}\right) \lesssim \sum_{\ell=0}^{L} N_\ell(\mathfrak{c}_\ell + n) + n\log(n)$$

subject to the constraint $\sum_{\ell=0}^{L} \frac{s^{-\beta\ell}}{N_\ell} = \varepsilon^2 \ln(n)^{-3}$ yields

$$N_l = \left\lceil \varepsilon^{-2}\ln(n)^3 \sqrt{\frac{s^{-l\beta}}{s^{l\gamma}+n}} \sum_{\ell=0}^{L} \sqrt{s^{-\ell\beta}\left(s^{\ell\gamma}+n\right)} \right\rceil\,, \quad 0 \le l \le L\,.$$

Therefore, the mean squared error is of order $\mathcal{O}(\varepsilon^2)$, and the corresponding cost is bounded by

$$\mathfrak{C}\left(\hat{\Phi}_L^{N,n}\right) \lesssim n\log(n) + n(L+1) + \varepsilon^{-\frac{\gamma}{\alpha}} \ln(n)^{\frac{\gamma}{\alpha}} + \varepsilon^{-2}\ln(n)^3 n + \varepsilon^{-2}\ln(n)^3 \begin{cases} L^2, & \beta=\gamma,\\ \varepsilon^{-\frac{\gamma-\beta}{\alpha}} \ln(n)^{\frac{\gamma-\beta}{\alpha}}, & \beta<\gamma,\\ n^{-\frac{\beta-\gamma}{\gamma}}, & \beta>\gamma. \end{cases}$$

Noticing that $2 \le n \lesssim \ln\left(\varepsilon^{-1}\right)$, $0 < \ln(2) \le \ln(n) \lesssim \ln\left(\ln(\varepsilon^{-1})\right) \le \ln\left(\varepsilon^{-1}\right)$, and $1 \le L \lesssim \ln\left(\varepsilon^{-1}\right)$ for $0 < \varepsilon < \min(\rho^{-1}, \ln(2))$ completes the proof. ∎

Remark 2.2. The condition $2\alpha \geq \min(\beta, \gamma)$ in the statement of Theorem 2.1 above is satisfied in many applications. It is nonetheless noteworthy that this condition is not essential for the proof of the complexity result. In fact, omitting this condition will not change the analysis; it will merely add an extra $\varepsilon^{-\gamma/\alpha} \ln(\varepsilon^{-1})$ term to the complexity. This term originates from the $\sum_{\ell=0}^{L} \mathfrak{c}_\ell$ contribution to the cost, which may no longer be dominated by the other terms in absence of the $2\alpha \geq \min(\beta, \gamma)$ condition. A similar observation for the complexity analysis of the standard multilevel Monte Carlo method has also been made in [6].

It is noteworthy that the results above illustrate that the multilevel Monte Carlo estimator may offer a worse complexity for the case of an analytic function Φ than for $\Phi \in C^{k+1}(\Theta)$ in some cases, for example, if $\frac{\gamma-\beta}{\alpha} > \frac{1}{k+1}$. This is due to the worse Lebesgue constant (cf. condition (ii) in Assumption A2), which is necessary for global Chebyshev polynomial interpolation, for example. The benefit of an analytic function Φ will, however, become apparent in section 2.2 below, where we will derive multilevel Monte Carlo approximations for derivatives of Φ.

2.1. A refined estimator. The proof of Theorem 2.1 reveals that the first term in the bound for the multilevel Monte Carlo cost $\mathfrak{C}\big(\hat{\Phi}_L^{N,n}\big)$ is due to levels $\ell \in \{0, \dots, L\}$, for which the cost $\mathfrak{c}_\ell$ for generating one sample $(Q_\ell, Q_{\ell-1})$ is cheaper than n evaluations of the function ϕ. In applications where the generation of accurate samples is rather expensive (e.g., for models involving partial differential equations), no such levels $\ell \in \{0, \dots, L\}$ may actually exist for a given mean squared error tolerance ε^2. In these cases, the bound stated in Theorem 2.1 for the computational cost corresponding to this tolerance does not have the first term and reduces to (ignoring logarithmic factors) the standard multilevel Monte Carlo cost bound for the expectation.

We emphasize that the practically driven motivation above is merely a fixed tolerance heuristic, since asymptotically as $\varepsilon \to 0$ there will always be levels $\ell \in \{0, \dots, L\}$, for which the cost of the n function evaluations is more costly than generating one sample of $(Q_\ell, Q_{\ell-1})$. However, the underlying principle of this heuristic can nonetheless be used to systematically derive a multilevel Monte Carlo estimator for functions that does not feature the first cost term and thus provides an improved computational complexity in many cases. Specifically, let $\mathfrak{c}_\phi$ denote the constant that bounds the cost of evaluating $\phi(\vartheta, q)$ for any $(\vartheta, q) \in \Theta \times \mathbb{R}$. The first term in the complexity bound of Theorem 2.1 is then due to levels $\ell \in \{0, \dots, L\}$ for which $\mathfrak{c}_\ell \leq n\mathfrak{c}_\phi$. The number of such levels could thus be decreased if n were not fixed across levels but varied instead. The idea is to use a functional extension where the set of nodes varies across levels, since it may pay off to use the functional extension with only a small number of nodes on coarser levels. That is, by considering both a hierarchy of interpolation grids and a hierarchy of discretization grids, the techniques developed here build upon combining ideas put forward in [25] and [20].

To make this idea concrete, let $n_\ell \in \mathbb{N}$ denote the number of nodes in Θ used for the function extension on level $\ell \in \mathbb{N}_0$. We will use the usual extension operator $\mathcal{I}_{n_\ell}$ in these n_ℓ nodes on level ℓ. For the sake of notation, however, we associate the nodes in Θ directly with the operator and view it as a map from continuous functions to some finite dimensional function space V_ℓ. Specifically, for $\ell \in \mathbb{N}_0$ we consider the sequence (i.e., hierarchy) of extension operators $I_\ell \colon C(\theta) \to V_\ell \subset L^\infty(\Theta)$ as a shorthand notation for

$$I_\ell(f) := \mathcal{I}_{n_\ell}\big[f\big(\theta^\ell\big)\big] \quad \text{with } \theta^\ell = \big(\theta_1^\ell, \dots, \theta_{n_\ell}^\ell\big) \in \Theta^{n_\ell} ,$$

where θ^ℓ denotes the deterministic nodes used on level $\ell \in \mathbb{N}_0$. For example, V_ℓ is the space of piecewise polynomials of a certain degree for a spline interpolation or the space of polynomials of degree n_ℓ for a (global) polynomial interpolation. Notice that Assumption A1 still holds for I_ℓ for any ℓ, provided that $\mathcal{I}_{n_\ell}$ satisfies it (analogously for Assumption A2). With slight abuse of notation, viewing n as the vector $n = (n_0, n_1, \dots, n_L)$, we then introduce the refined multilevel Monte Carlo estimator as

$$\hat{\Phi}_L^{N,n} := \sum_{\ell=0}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \left(I_\ell\Big[\phi\Big(\cdot, Q_\ell^{(i,\ell)}\Big)\Big] - I_{\ell-1}\Big[\phi\Big(\cdot, Q_{\ell-1}^{(i,\ell)}\Big)\Big] \right) \tag{9}$$

with the additional convention that $I_{-1}(\cdot) \equiv 0$. Notice that the definition of the refined estimator above contains the special case of using just one (global) extension operator, namely, if $n_\ell = n$ and $\theta^\ell = \big(\theta_1, \dots, \theta_n\big)$ for all levels $\ell \in \mathbb{N}_0$. The estimator's computational complexity for increasing sequences of nodes is summarized in the following result, where we denote by $(I_{\ell_1} \otimes I_{\ell_2})[f]$ the bivariate (tensor) interpolation of a function $f\colon \Theta \times \Theta \to \mathbb{R}$ with n_{ℓ_1} and n_{ℓ_2} grid points in the first and second variable, respectively.

Theorem 2.2. *Let $\Phi(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big)$. Suppose that the hypotheses (i)–(iii) of Theorem 2.1 are satisfied for ϕ. Furthermore, suppose that the sequences of finite dimensional spaces V_ℓ and extension operators $I_\ell\colon C(\theta) \to V_\ell \subset L^\infty(\Theta)$ satisfy*

(iv) $\|g\|_{L^\infty(\Theta)} \le c_4 n_\ell^{\nu_1} \|g\|_{L^2(\Theta)}$ $\forall\, g \in V_\ell$, *for some* $\nu_1 \ge 0$;

(v) $\int_\Theta \big((I_\ell - I_{\ell-1}) \otimes (I_\ell - I_{\ell-1})[\mathrm{Cov}_{\psi_{\ell-1}}]\big)(\vartheta, \vartheta)\, d\vartheta \le c_5 n_\ell^{-\nu_2}$ *for some* $\nu_2 > 0$, *where* $\psi_{\ell-1}(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q_{\ell-1})\big) - \phi(\vartheta, Q_{\ell-1})$ *and* $c_5 \equiv c_5(\psi_{\ell-1}) < \infty$

for all $\ell \in \mathbb{N}$, where c_4, c_5 are positive constants independent of n_ℓ.

If $\Phi \in C^{k+1}(\Theta)$ for some $k \in \mathbb{N}_0$, and if Assumption A1 is satisfied for that k, then for any $\varepsilon > 0$ there exist parameters $L \in \mathbb{N}_0$, $n \in \mathbb{N}^{L+1}$, and $N \in \mathbb{N}^{L+1}$ such that the corresponding multilevel Monte Carlo estimator $\hat{\Phi}_L^{N,n}$ satisfies

$$\mathbb{E}\Big(\big\|\hat{\Phi}_L^{N,n} - \Phi\big\|_{L^\infty(\Theta)}^2\Big) = \mathcal{O}(\varepsilon^2) . \tag{10}$$

Furthermore, for any $0 < \varepsilon < e^{-1}$ the associated computational cost $\mathfrak{C}\big(\hat{\Phi}_L^{N,n}\big)$ is bounded by

$$\mathfrak{C}\big(\hat{\Phi}_L^{N,n}\big) \lesssim \varepsilon^{-p_1} + \max\Big\{1, \varepsilon^{\frac{\beta}{\alpha\nu} - \frac{1}{k+1}}\Big\} \varepsilon^{-(2+p_2)} \ln\big(\varepsilon^{-1}\big)^{2+p_3} + \ln\big(\varepsilon^{-1}\big)^2 \begin{cases} \varepsilon^{-2} \ln\big(\varepsilon^{-1}\big)^2 & \text{if } \beta = \gamma , \\ \varepsilon^{-\left(2 + \frac{\gamma-\beta}{\alpha}\right)} & \text{if } \beta < \gamma , \\ \varepsilon^{-2} & \text{if } \beta > \gamma \end{cases}$$

with $\nu := \nu_2 - 2\nu_1 > 0$ as well as

$$p_1 = \begin{cases} 0 , & \nu \ge 1 , \\ \frac{\beta}{\alpha\nu} , & \nu < 1 , \end{cases} \quad p_2 = \begin{cases} 0 , & \nu \ge 1 , \\ \frac{\beta}{\alpha\nu}(1-\nu) , & \nu < 1 , \end{cases} \quad \text{and} \quad p_3 = \begin{cases} 0 , & \nu \ne 1 , \\ 2 , & \nu = 1 . \end{cases}$$

If $\Phi\colon \Theta \to \mathbb{R}$ is analytic in Θ, and if Assumption A2 holds for Φ with $\rho > 1$, then for any $\varepsilon > 0$ there exist parameters such that the corresponding estimator $\hat{\Phi}_L^{N,n}$ satisfies (10). Moreover, for any $0 < \varepsilon < \min(\rho^{-1}, \ln(2)/2)$ the required computational cost is bounded by

$$\mathfrak{C}\big(\hat{\Phi}_L^{N,n}\big) \lesssim \Big(\varepsilon^{-1}\ln\big(\varepsilon^{-1}\big)\Big)^{p_1} + \varepsilon^{-(2+p_2)}\ln\big(\varepsilon^{-1}\big)^{4+p_3} + \ln\big(\varepsilon^{-1}\big)^4 \begin{cases} \varepsilon^{-2}\ln(\varepsilon^{-1})^2 & \text{if } \beta = \gamma, \\ \varepsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)}\ln(\varepsilon^{-1})^{\frac{\gamma-\beta}{\alpha}} & \text{if } \beta < \gamma, \\ \varepsilon^{-2} & \text{if } \beta > \gamma \end{cases}$$

with p_1, p_2, and p_3 as before.

Proof. As in the proof of Theorem 2.1, we begin by decomposing the mean squared error:

$$\begin{aligned} \mathrm{MSE}\big(\hat{\Phi}_L^{N,n}\big) &\equiv \mathbb{E}\Big(\big\|\Phi - I_L(\Phi) + I_L(\Phi) - \hat{\Phi}_L^{N,n}\big\|_{L^\infty(\Theta)}^2\Big) \\ &\le 2\Big(\big\|\Phi - I_L(\Phi)\big\|_{L^\infty(\Theta)}^2 + \mathbb{E}\Big(\big\|I_L(\Phi) - \hat{\Phi}_L^{N,n}\big\|_{L^\infty(\Theta)}^2\Big)\Big)\,. \end{aligned} \tag{11}$$

First, we consider $\Phi \in C^{k+1}(\Theta)$. It follows from Assumption A1 that

$$\mathrm{MSE}\big(\hat{\Phi}_L^{N,n}\big) \lesssim 2c\Big({n_L}^{-2(k+1)} + 2\big\|\Phi(\theta^L) - \Phi_L(\theta^L)\big\|_{\ell^\infty}^2 + 2\mathbb{E}\Big(\big\|I_L(\Phi_L) - \hat{\Phi}_L^{N,n}\big\|_{L^\infty(\Theta)}^2\Big)\Big)\,,$$

where $\Phi_L(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q_L)\big)$. As in Theorem 2.1, the second term on the right-hand side above (i.e., the squared bias) is bounded as a consequence of the hypothesis:

$$\big\|\Phi(\theta^L) - \Phi_L(\theta^L)\big\|_{\ell^\infty} = \max_{1\le j\le n_L}\big|\mathbb{E}\big[\phi\big(\theta_j^L, Q\big) - \phi\big(\theta_j^L, Q_L\big)\big]\big| \le c_1 {h_0}^\alpha s^{-\alpha L}\,.$$

The remaining term in the mean squared error bound accounts for both the variance decay across the levels and the fact that the functional extensions use a different number of nodes on each level. To quantify this term, we first introduce the notation $\Delta_\ell f \equiv \Delta_\ell f(\vartheta) := f(\vartheta, Q_\ell) - f(\vartheta, Q_{\ell-1})$ for any suitable function $f\colon \Theta \times \mathbb{R} \to \mathbb{R}$, as well as the sample average $E_\ell(\xi) := \frac{1}{N_\ell}\sum_{i=1}^{N_\ell} \xi^{(i,\ell)}$ for any random variable ξ. We then have

$$I_L(\Phi_L) - \hat{\Phi}_L^{N,n} = \sum_{\ell=0}^{L}\big(I_\ell\big[\Delta_\ell\Phi - E_\ell(\Delta_\ell\phi)\big] + (I_\ell - I_{\ell-1})\big[\Phi_{\ell-1} - E_\ell(\phi_{\ell-1})\big]\big)\,,$$

where $\Delta_\ell\Phi = \mathbb{E}(\Delta_\ell\phi)$ and $\phi_{\ell-1}(\vartheta) := \phi(\vartheta, Q_{\ell-1})$. Consequently, we find that

$$\mathbb{E}\Big(\big\|I_L(\Phi_L) - \hat{\Phi}_L^{N,n}\big\|_{L^\infty(\Theta)}^2\Big) \le 2(L+1)\sum_{\ell=0}^{L}\big(T_{1,\ell} + T_{2,\ell}\big)\,,$$

where

$$T_{1,\ell} = \mathbb{E}\Big(\big\|I_\ell[\Delta_\ell\Phi - E_\ell(\Delta_\ell\phi)]\big\|_{L^\infty(\Theta)}^2\Big)\,, \quad T_{2,\ell} = \mathbb{E}\Big(\big\|(I_\ell - I_{\ell-1})[\Phi_{\ell-1} - E_\ell(\phi_{\ell-1})]\big\|_{L^\infty(\Theta)}^2\Big)\,.$$

The term $T_{1,\ell}$ can be bounded as in the proof of Theorem 2.1, yielding

$$T_{1,\ell} \le c\ln(n_\ell)\frac{\mathbb{E}\Big(\big\|\Delta_\ell\phi(\theta^\ell)\big\|_{\ell^\infty}^2\Big)}{N_\ell} \le cc_2{h_0}^\beta \ln(n_\ell)\frac{s^{-\beta\ell}}{N_\ell}\,.$$

To derive a bound for the term $T_{2,\ell}$, let $\psi_{\ell-1}(\vartheta) = \Phi_{\ell-1}(\vartheta) - \phi_{\ell-1}(\vartheta)$. It then follows from hypothesis (iv) that

$$\begin{aligned} T_{2,\ell} &\le c_4 {n_\ell}^{2\nu_1} \mathbb{E}\Big(\big\| (I_\ell - I_{\ell-1})[E_\ell(\psi_{\ell-1})] \big\|^2_{L^2(\Theta)} \Big) \\ &= \frac{c_4 {n_\ell}^{2\nu_1}}{N_\ell} \int_\Theta \big((I_\ell - I_{\ell-1}) \otimes (I_\ell - I_{\ell-1}) \operatorname{Cov}_{\psi_{\ell-1}} \big)(\vartheta,\vartheta)\, d\vartheta \,, \end{aligned}$$

where $\operatorname{Cov}_{\psi_{\ell-1}}(\vartheta_1,\vartheta_2) = \mathbb{E}\big(\psi_{\ell-1}(\vartheta_1)\psi_{\ell-1}(\vartheta_2)\big)$. In view of hypothesis (v), we eventually find

$$\operatorname{MSE}\big(\hat\Phi_L^{N,n}\big) \lesssim {n_L}^{-2(k+1)} + s^{-2\alpha L} + (L+1) \sum_{\ell=0}^L \frac{1}{N_\ell} \big(\ln(n_\ell) s^{-\beta\ell} + {n_\ell}^{-\nu} \big) \,.$$

To guarantee a mean squared error of order $\mathcal{O}(\varepsilon^2)$, this bound implies choosing

$$L = \big\lceil \alpha^{-1} \log_s\big(\varepsilon^{-1}\big) \big\rceil$$

as well as $n_L \gtrsim \varepsilon^{-\frac{1}{k+1}}$. Furthermore, the last term of the mean squared error bound suggests balancing the variance term and the interpolation term so that ${n_\ell}^{-\nu} \lesssim \ln(n_\ell) s^{-\beta\ell}$, which is implied by ${n_\ell}^{-\nu} \lesssim s^{-\beta\ell}$ if $n_\ell \ge 3$ for all $\ell \in \mathbb{N}_0$. We thus choose

$$n_\ell = \Big\lceil n_L s^{-\frac{\beta}{\nu}(L-\ell)} \Big\rceil \quad \text{with} \quad n_L = \Big\lceil \max\Big\{ c\varepsilon^{-\frac{\beta}{\alpha\nu}}, \varepsilon^{-\frac{1}{k+1}} \Big\} \Big\rceil \,,$$

with a generic positive constant c independent of $\varepsilon > 0$ ensuring that $n_0 \ge 3$. The number of samples N_ℓ on each level is again obtained by minimizing the computational cost subject to variance constraint. Here, the computational cost $\mathfrak{C}\big(\hat\Phi_L^{N,n}\big)$ of constructing the estimator $\hat\Phi_L^{N,n}$ is bounded by

$$\mathfrak{C}\big(\hat\Phi_L^{N,n}\big) \le c \sum_{\ell=0}^L \big(N_\ell(\mathfrak{c}_\ell + n_\ell) + n_\ell \big) \lesssim \sum_{\ell=0}^L N_\ell \big(s^{\ell\gamma} + n_\ell \big) + \sum_{\ell=0}^L n_\ell \,.$$

Treating the variables $N = (N_0, N_1, \dots, N_L)$ as continuous and minimizing the cost $\mathfrak{C}\big(\hat\Phi_L^{N,n}\big)$ with respect to $(N_0,\dots,N_L)$, subject to the constraint $\sum_{\ell=0}^L \frac{s^{-\beta\ell}\ln(n_\ell)}{N_\ell} = \varepsilon^2 L^{-1}$, eventually yields

$$N_l = \left\lceil \varepsilon^{-2} L \sqrt{\frac{s^{-l\beta}\ln(n_l)}{s^{l\gamma} + n_l}} \sum_{\ell=0}^L \sqrt{s^{-\ell\beta}\ln(n_\ell)\big(s^{\ell\gamma} + n_\ell\big)} \right\rceil \,, \quad 0 \le l \le L \,,$$

so that the mean squared error is $\mathcal{O}(\varepsilon^2)$ as required. The corresponding computational cost is bounded by

$$\begin{aligned} \mathfrak{C}\big(\hat\Phi_L^{N,n}\big) &\lesssim \sum_{\ell=0}^L s^{\gamma\ell} + 2\sum_{\ell=0}^L n_\ell + \varepsilon^{-2} L \left(\sum_{\ell=0}^L \sqrt{s^{-\ell\beta}\ln(n_\ell)\big(s^{\ell\gamma} + n_\ell\big)} \right)^2 \\ &\lesssim \varepsilon^{-\frac{\gamma}{\alpha}} + n_L + \varepsilon^{-2} L \left(\sum_{\ell=0}^L \sqrt{s^{-\ell\beta}\ln(n_\ell)\big(s^{\ell\gamma} + n_\ell\big)} \right)^2 . \end{aligned}$$

To quantify the sum in the cost bound above, we distinguish the levels for which the cost $\mathfrak{c}_\ell$ is cheaper than the n_ℓ evaluations of the function ϕ. Therefore, we introduce

$$L^* := \begin{cases} \min\left(\left\lceil \frac{\nu}{\nu\gamma-\beta}\log_s(n_0)\right\rceil - 1, L\right) , & \frac{\gamma}{\beta} > \frac{1}{\nu} , \\ L , & \text{otherwise,} \end{cases}$$

so that $s^{\ell\left(\gamma-\frac{\beta}{\nu}\right)} < n_L s^{-\frac{\beta}{\nu}L} = n_0$ for all $0 < \ell \le L^*$, which implies $\mathfrak{c}_\ell \lesssim n_\ell$ for these levels. We can thus decompose the remaining sum as

$$\begin{aligned}\sum_{\ell=0}^{L}\sqrt{s^{-\ell\beta}\ln(n_\ell)\left(s^{\ell\gamma}+n_\ell\right)} &= \sum_{\ell=0}^{L^*}\sqrt{s^{-\ell\beta}\ln(n_\ell)\left(s^{\ell\gamma}+n_\ell\right)} + \sum_{\ell=L^*+1}^{L}\sqrt{s^{-\ell\beta}\ln(n_\ell)\left(s^{\ell\gamma}+n_\ell\right)} \\ &=: S_1 + S_2 .\end{aligned}$$

For S_1 we know that $s^{\ell\gamma} \lesssim n_\ell$, so that

$$S_1 \lesssim s^{-\frac{\beta}{2\nu}L}\sqrt{n_L}\sum_{\ell=0}^{L^*}\sqrt{s^{-\ell\beta\left(1-\frac{1}{\nu}\right)}\ln(n_\ell)} \lesssim s^{-\frac{\beta}{2\nu}L}\sqrt{n_L\ln(n_L)}\sum_{\ell=0}^{L^*}s^{-\frac{\ell\beta}{2\nu}(\nu-1)} ,$$

where the last bound follows from the monotonic growth of n_ℓ across levels, which shows that

$$S_1 \lesssim \sqrt{\varepsilon^{\frac{\beta}{\nu\alpha}}n_L\ln(n_L)}\begin{cases} 1 , & \nu > 1 , \\ L^*+1 , & \nu = 1 , \\ s^{\frac{\beta}{2\nu}(1-\nu)L^*} , & \nu < 1 \end{cases}$$

and therefore yields

$$S_1^2 \lesssim \ln(\varepsilon^{-1})\max\left\{1, \varepsilon^{\frac{\beta}{\nu\alpha}-\frac{1}{k+1}}\right\}\begin{cases} 1 , & \nu > 1 , \\ \ln(\varepsilon^{-1})^2 , & \nu = 1 , \\ \varepsilon^{-\frac{\beta}{\alpha\nu}(1-\nu)} , & \nu < 1 . \end{cases}$$

Similarly, for S_2 we find

$$S_2 \lesssim \sqrt{\ln(n_L)}\sum_{\ell=L^*+1}^{L}s^{-\ell\frac{\beta-\gamma}{2}} \lesssim \sqrt{\ln(\varepsilon^{-1})}\begin{cases} L+1 , & \beta = \gamma , \\ s^{\frac{\gamma-\beta}{2}L} , & \beta < \gamma , \\ s^{\frac{\gamma-\beta}{2}L^*} , & \beta > \gamma , \end{cases}$$

from which the claim follows.

For $\Phi\colon \Theta \to \mathbb{R}$ analytic, the proof is very similar. In fact, it follows from Assumption A2 that the mean squared error can be bounded by

$$\mathrm{MSE}\left(\hat{\Phi}_L^{N,n}\right) \lesssim \rho^{-2n_L} + \ln(n_L)^2 s^{-2\alpha L} + (L+1)\sum_{\ell=0}^{L}\frac{1}{N_\ell}\left(\ln(n_\ell)^3 s^{-\beta\ell} + n_\ell^{-\nu}\right) ,$$

using the hypotheses and the same steps as before. Consequently, we chose

$$L = \left\lceil \alpha^{-1}\log_s\left(\varepsilon^{-1}\ln(n_L)\right)\right\rceil \quad \text{and} \quad n_\ell = \left\lceil n_L s^{-\frac{\beta}{\nu}(L-\ell)}\right\rceil$$

with

$$n_L = \left\lceil \max\left\{ \log_\rho\left(\varepsilon^{-1}\right), c_\rho\left(\varepsilon^{-1}\ln\left(\varepsilon^{-1}\right)\right)^{\frac{\beta}{\alpha\nu}} \right\} \right\rceil ,$$

where the generic constant c_ρ is independent of $\varepsilon < \ln(2)/2$, ensuring that $n_0 \geq 3$. Moreover, minimizing the cost bound

$$\mathfrak{C}\left(\hat{\Phi}_L^{N,n}\right) \leq c\sum_{\ell=0}^{L}\left(N_\ell(\mathfrak{c}_\ell + n_\ell) + n_\ell\ln(n_\ell)\right) \lesssim \sum_{\ell=0}^{L} N_\ell\left(s^{\ell\gamma} + n_\ell\right) + \sum_{\ell=0}^{L}\ln(n_\ell)n_\ell ,$$

subject to the constraint $\sum_{\ell=0}^{L}\frac{s^{-\beta\ell}\ln(n_\ell)^3}{N_\ell} = \varepsilon^2 L^{-1}$ yields

$$N_l = \left\lceil \varepsilon^{-2}L\sqrt{\frac{s^{-l\beta}\ln(n_l)^3}{s^{l\gamma}+n_l}}\sum_{\ell=0}^{L}\sqrt{s^{-\ell\beta}\ln(n_\ell)^3\left(s^{\ell\gamma}+n_\ell\right)} \right\rceil , \quad 0 \leq l \leq L ,$$

so that the mean squared error is $\mathcal{O}(\varepsilon^2)$. Moreover, we find that

$$\mathfrak{C}\left(\hat{\Phi}_L^{N,n}\right) \lesssim \varepsilon^{-\frac{\gamma}{\alpha}} + n_L\ln(n_L) + \varepsilon^{-2}L\left(\sum_{\ell=0}^{L}\sqrt{s^{-\ell\beta}\ln(n_\ell)^3\left(s^{\ell\gamma}+n_\ell\right)}\right)^2 .$$

Using L^* as above, only accounting for the modified value of n_0, as well as S_1 and S_2 as defined before, we have that $\sum_{\ell=0}^{L}\sqrt{s^{-\ell\beta}\ln(n_\ell)^3\left(s^{\ell\gamma}+n_\ell\right)} = S_1 + S_2$, which can be bounded by

$$S_1 \lesssim \sqrt{s^{-\frac{\beta}{\nu}L}n_L\ln(n_L)^3}\begin{cases} 1 , & \nu > 1 , \\ L^*+1 , & \nu = 1 , \\ s^{\frac{\beta}{2\nu}(1-\nu)L^*} , & \nu < 1 , \end{cases} \quad \text{and} \quad S_2 \lesssim \sqrt{\ln(n_L)^3}\begin{cases} L+1 , & \beta = \gamma , \\ s^{\frac{\gamma-\beta}{2}L} , & \beta < \gamma , \\ s^{\frac{\gamma-\beta}{2}L^*} , & \beta > \gamma , \end{cases}$$

so that the claim eventually follows. ∎

Remark 2.3. Hypothesis (iv) in Theorem 2.2 is a special case of what is known as an inverse inequality (or inverse estimate) in polynomial approximation theory. For example, hypothesis (iv) is satisfied with $\nu_1 = \frac{1}{2}$ for an interpolation with piecewise polynomials (i.e., splines) on a uniform grid [9, Thm. 4.5.11] and with $\nu_1 = 1$ for a global polynomial interpolation [10, Chap. 5]. Hypothesis (v), on the other hand, encodes a regularity condition for the covariance Cov_{ψ_ℓ}. A sufficient condition for hypothesis (v) to be satisfied is $\mathrm{Cov}_{\psi_\ell} \in C^{r,r}(\Theta)$ for any $\ell \in \mathbb{N}_0$, where

$$C^{r,r}(\Theta) = \left\{ f\colon \Theta\times\Theta \to \mathbb{R} \,\middle|\, \frac{\partial^{\alpha_1+\alpha_2}}{\partial\vartheta_1^{\alpha_1}\partial\vartheta_2^{\alpha_2}} f \text{ continuous in } \Theta\times\Theta \text{ for all } \alpha_1,\alpha_2 \in \{0,1,\dots,r\} \right\}$$

for some $r \in \mathbb{N}$, which is equipped with the norm

$$\|f\|_{C^{r,r}(\Theta)} = \max_{\alpha_1,\alpha_2\in\{0,1,\dots,r\}}\sup_{\vartheta_1,\vartheta_2\in\Theta}\left|\frac{\partial^{\alpha_1+\alpha_2}}{\partial\vartheta_1^{\alpha_1}\partial\vartheta_2^{\alpha_2}} f(\vartheta_1,\vartheta_2)\right| ;$$

see, e.g., [5]. Notice that $C^{r,r}(\Theta)$ contains, in particular, functions of class $C^{2r}(\Theta)$. For example, for a spline interpolation of degree k, we then find

$$\int_\Theta \big((I_\ell - I_{\ell-1}) \otimes (I_\ell - I_{\ell-1})[\mathrm{Cov}_{\psi_{\ell-1}}]\big)(\vartheta,\vartheta)\, d\vartheta \le c n_\ell^{-2s} \big\|\mathrm{Cov}_{\psi_{\ell-1}}\big\|_{C^{s,s}(\Theta)} ,$$

with $s = \min\{r, k+1\}$ and for a generic positive constant $c < \infty$. Consequently, we have $\nu_2 = 2s$ in that case. A similar result holds for the case of a (global) polynomial interpolation. Finally, we recall that $\phi(\cdot, Q_{\ell-1}) \in C^r(\Theta)$ with uniformly bounded derivatives with respect to the expectation $\mathbb{E}$ implies that $\Phi_{\ell-1} \in C^r(\Theta)$ and $\mathrm{Cov}_{\psi_{\ell-1}} \in C^{r,r}(\Theta)$. We emphasize, however, that this is merely a sufficient condition, which may not be necessary.

The complexity results obtained in Theorem 2.2 confirm that a hierarchy of extension operators I_ℓ does indeed provide means of eliminating the first term in the cost bound found in Theorem 2.1. Moreover, it shows that the refined estimator (9) is more advantageous for $\Phi \in C^{k+1}(\Theta)$ than for Φ analytic and can offer the standard multilevel Monte Carlo complexity result up to a logarithmic factor in that case.

Corollary 2.3. *Suppose that the assumptions of Theorem 2.2 are satisfied and that $\Phi \in C^{k+1}(\Theta)$. Let $\hat{\Phi}_L^{N,n}$ be the refined multilevel Monte Carlo estimator (9) that is identified in Theorem 2.2. If $1 < \nu \le (k+1)\frac{\beta}{\alpha}$, then the computational cost $\mathfrak{C}\big(\hat{\Phi}_L^{N,n}\big)$ is bounded by*

$$\mathfrak{C}\big(\hat{\Phi}_L^{N,n}\big) \lesssim \ln\big(\varepsilon^{-1}\big)^2 \begin{cases} \varepsilon^{-2}\ln\big(\varepsilon^{-1}\big)^2 & \text{if } \beta = \gamma\,, \\ \varepsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)} & \text{if } \beta < \gamma\,, \\ \varepsilon^{-2} & \text{if } \beta > \gamma \end{cases}$$

for any $0 < \varepsilon < e^{-1}$.

Further inspection of the general complexity results in Theorem 2.2 reveals that the complexity of the multilevel Monte Carlo estimator for $\Phi \in C^{k+1}(\Theta)$ is actually improved for any $\nu > 1$ compared to the estimator analyzed in Theorem 2.1, which relies on the same function extension on every level. However, these gains are getting smaller as ν increases, unless $\nu \le (k+1)\frac{\beta}{\alpha}$. As one may expect that the function $\Phi \in C^{k+1}(\Theta)$ is also very smooth (i.e., $k \gg 1$) whenever the covariances are (i.e., $\nu \gg 1$), the loss of effectiveness is due to the fact that the complexity of the multilevel Monte Carlo estimator for very smooth functions is already almost optimal (i.e., close to the standard multilevel Monte Carlo complexity result).

Related to this is the observation that the complexity result obtained in Theorem 2.2 for Φ analytic is actually worse than the one found in Theorem 2.1, which is already close the the standard multilevel Monte Carlo complexity. This is due to the fact that eliminating the first term in the cost bound in Theorem 2.1 comes at the cost of an extra logarithmic factor, which may originate from a nonoptimal choice for the sequence n_ℓ and a possibly pessimistic complexity analysis. In fact, we suspect that this can be improved, in the sense that the actual complexity may be better, but we have not been able to prove it. We will leave this further complexity investigation for future work.

As the refined multilevel Monte Carlo estimator (9), which uses a hierarchy of extension operators, is most effective for $\Phi \in C^{k+1}(\Theta)$ (and $\nu > 1$), in what follows we will only use the refined estimator (9) in that case, provided that $\nu > 0$. Conversely, we will use the multilevel Monte Carlo estimator with one (global) extension operator studied in Theorem 2.1 if $\nu \le 0$ or when estimating Φ analytic. For notational convenience, we will denote the multilevel Monte

Carlo estimator by $\hat{\Phi}_L^{N,n}$ in both cases, with the understanding that if $n = (n_0, \dots, n_L) \in \mathbb{N}^{L+1}$, then the estimator is defined using a hierarchy of extension operators as in (9), while if $n \in \mathbb{N}$, then estimator is constructed using one (global) extension operator as in (1), which we recall is a special case of (9). Finally, we emphasize that the results we present in the following sections can, in principle, also be derived for the other versions of the presented estimators.

2.2. Approximation of derivatives. In view of the construction of the basic multilevel Monte Carlo estimator in (1), one may think that an alternative extension of the pointwise estimates to a function is possible, provided the samples on all levels are saved. Specifically, it may be tempting to simply consider the mapping $\Theta \ni \vartheta \mapsto \bar{\Phi}_L^N(\vartheta)$ instead of the function $\hat{\Phi}_L^{N,n} = \mathcal{I}_n\big(\bar{\Phi}_L^N(\theta)\big)$ for $n \in \mathbb{N}$. However, this naive approach is not guaranteed to be accurate between the nodes θ, so that the uniform error criterion may not be met. Moreover, considering an appropriate extension operator (or a hierarchy thereof) is crucial when approximations to derivatives of Φ are also sought after. As a matter of fact, this is desirable in many applications including the stochastic optimization problem mentioned in the introduction, but it is also essential in scenarios related to the characteristic function and the CDF as we will discuss in the following sections.

The advantage of extending the pointwise estimate to a function $\hat{\Phi}_L^{N,n}$ via an appropriate extension operator $\mathcal{I}_n$, or a hierarchy of extension operators $\mathcal{I}_{n_\ell}$, is that it provides a natural and efficient way of computing derivatives of the estimated function. Notice that this is not possible without the extension operator in general. To see this, consider the CDF $\Phi(\vartheta) = \mathbb{E}\big(I(Q \le \vartheta)\big)$, for example. Any finite sample size (single- or multilevel) Monte Carlo approximation $\bar{\Phi}_L^N(\vartheta)$ will, as a function of ϑ, only provide a piecewise constant approximation, regardless of the regularity of Φ. Consequently, derivatives of the estimated function will vanish almost everywhere, so that no further information concerning the derivatives of Φ can be gained. Conversely, by taking advantage of the extension operators it is possible to overcome this shortcoming.

To characterize the accuracy of derivative approximations, we have to strengthen the assumptions on the extension operators.

Assumption A3 (derivatives of the extension operator). *For $k, m \in \mathbb{N}_0$ given so that $m < k+1$, let the sequence of extension operators $\mathcal{I}_n \colon \mathbb{R}^n \to C^m(\Theta)$ based on the set of nodes $\theta \in \Theta^n$ satisfy Assumption A1. Furthermore, for all $n \in \mathbb{N}$ the operators $\mathcal{I}_n$ satisfy the following:*

(iv) $\|f^{(m)} - \frac{d^m}{d\vartheta^m}\mathcal{I}_n\big(f(\theta)\big)\|_{L^\infty(\Theta)} \le c_4 n^{-(k+1-m)}$ *for any $f \in C^{k+1}(\Theta)$;*

(v) $\|\frac{d^m}{d\vartheta^m}\mathcal{I}_n(x)\|_{L^\infty(\Theta)} \le c_5 n^m \|\mathcal{I}_n(x)\|_{L^\infty(\Theta)}$ *for any $x \in \mathbb{R}^n$;*

(vi) *the cost for computing $\frac{d^m}{d\vartheta^m}\mathcal{I}_n(x)$ is proportional to the cost of computing $\mathcal{I}_n(x)$.*

Here, the constants $c_4, c_5 > 0$ are independent of n but may depend on m.

Similarly, for analytic functions we will assume the following.

Assumption A4 (derivatives of the extension operator for analytic functions). *Let $f \in C^\infty(\Theta)$ and the sequence of extension operators $\mathcal{I}_n \colon \mathbb{R}^n \to C^\infty(\Theta)$ based on the set of nodes $\theta \in \Theta^n$ be such that Assumption A2 holds with $\rho > 1$. Moreover, for any $m \in \mathbb{N}_0$ it holds that*

(iv) $\|f^{(m)} - \frac{d^m}{d\vartheta^m}\mathcal{I}_n\big(f(\theta)\big)\|_{L^\infty(\Theta)} \le c_4 \rho^{-n}$;

(v) $\|\frac{d^m}{d\vartheta^m}\mathcal{I}_n(x)\|_{L^\infty(\Theta)} \le c_5 n^{2m} \|\mathcal{I}_n(x)\|_{L^\infty(\Theta)}$ *for any $x \in \mathbb{R}^n$;*

(vi) *the cost for computing* $\frac{d^m}{d\vartheta^m}\mathcal{I}_n(x)$ *is proportional to the cost of computing* $\mathcal{I}_n(x)$ *for all* $n \in \mathbb{N}$ *with the constants* $c_4, c_5 > 0$ *being independent of* n *but possibly depending on* m.

Remark 2.4. Assumption A3 holds true, for example, for C^m continuous piecewise polynomials (i.e., splines) of degree k on a uniform grid, provided that $k+1$ is even (so-called odd degree polynomials) [23]. As for Assumption A2, Assumption A4 is satisfied for polynomial interpolation on Chebyshev nodes as a result of the spectral convergence [34, Chap. 21] and the Markov brothers' inequality [22].

Now we are in the position to address the complexity result for the approximation of derivatives. It is noteworthy that the following result concerning derivatives of Φ is essentially based on identical hypotheses on the function ϕ as is Theorem 2.1 for the approximation of Φ itself.

Theorem 2.4. *Let* $\Phi(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big)$ *and* $m \in \mathbb{N}$. *Suppose that the hypotheses* (i)–(iii) *of Theorem* 2.1 *are satisfied for* ϕ.

For $\Phi \in C^{k+1}(\Theta)$ *with* $m \le k \in \mathbb{N}_0$ *such that Assumption* A3 *is satisfied for that* k, *suppose that hypotheses* (iv) *and* (v) *of Theorem* 2.2 *are satisfied with* $\nu := \nu_2 - 2\nu_1 > 2m$. *Then for any* $0 < \varepsilon$ *there exist parameters* $L \in \mathbb{N}_0$, $n = (n_0, \dots, n_L) \in \mathbb{N}^{L+1}$, *and* $N \in \mathbb{N}^{L+1}$ *such that the* m*th derivative of corresponding refined multilevel Monte Carlo estimator* $\hat{\Phi}_L^{N,n}$ *defined in* (9) *satisfies*

$$\mathbb{E}\left(\left\|\Phi^{(m)} - \frac{d^m}{d\vartheta^m}\hat{\Phi}_L^{N,n}\right\|^2_{L^\infty(\Theta)}\right) = \mathcal{O}(\varepsilon^2)\,. \tag{12}$$

Furthermore, for any $0 < \varepsilon < e^{-1}$ *the associated computational cost* $\mathfrak{C}\big(\frac{d^m}{d\vartheta^m}\hat{\Phi}_L^{N,n}\big)$ *is bounded by*

$$\mathfrak{C}\left(\frac{d^m}{d\vartheta^m}\hat{\Phi}_L^{N,n}\right) \lesssim \varepsilon^{-\frac{\gamma}{\alpha}p_0} + \varepsilon^{-p_1} + \max\left\{1, \varepsilon^{(2m+1)\frac{\beta(k+1)-\alpha\nu}{\alpha\nu(k+1-m)}}\right\}\varepsilon^{-(2+p_2)}\ln\big(\varepsilon^{-1}\big)^{2+p_3}$$
$$+ \max\left\{1, \varepsilon^{2m\frac{\beta(k+1)-\alpha\nu}{\alpha\nu(k+1-m)}}\right\}\ln\big(\varepsilon^{-1}\big)^2\varepsilon^{-2}\begin{cases}\ln\big(\varepsilon^{-1}\big)^2\,, & \beta\big(1-\frac{2m}{\nu}\big) = \gamma\,,\\ \varepsilon^{-\left(\frac{\gamma-\beta}{\alpha}+\frac{2\beta m}{\alpha\nu}\right)p_0}\,, & \beta\big(1-\frac{2m}{\nu}\big) < \gamma\,,\\ 1\,, & \beta\big(1-\frac{2m}{\nu}\big) > \gamma\,,\end{cases}$$

where $p_0 = \max\left\{\frac{\alpha\nu}{\alpha\nu-\beta m}, \frac{k+1}{k+1-m}\right\}$, $p_1 = \max\left\{\frac{\beta}{\alpha\nu-\beta m}, \frac{1}{k+1-m}\right\}$, *as well as*

$$p_2 = \begin{cases}0\,, & \nu \ge 2m+1\,,\\ \max\left\{\frac{\beta(2m+1-\nu)}{\alpha\nu-\beta m}, \frac{\beta(2m+1-\nu)(k+1)}{\alpha\nu(k+1-m)}\right\}\,, & \nu < 2m+1\,,\end{cases} \quad \text{and} \quad p_3 = \begin{cases}0\,, & \nu \ne 2m+1\,,\\ 2\,, & \nu = 2m+1\,.\end{cases}$$

On the other hand, if $\nu \le 2m$, *then there exist parameters* $L \in \mathbb{N}_0$, $n \in \mathbb{N}$, *and* $N \in \mathbb{N}^{L+1}$ *such that the* m*th derivative of multilevel Monte Carlo estimator* $\hat{\Phi}_L^{N,n}$ *defined by* (1) *satisfies* (12) *and whose computational cost is bounded by*

$$\mathfrak{C}\left(\frac{d^m}{d\vartheta^m}\hat{\Phi}_L^{N,n}\right) \lesssim \varepsilon^{-\frac{\gamma}{\alpha}\left(1+\frac{2m}{k+1-m}\right)} + \ln\left(\varepsilon^{-1}\right)\varepsilon^{-\left(2+\frac{1+2m}{k+1-m}\right)} + \ln\left(\varepsilon^{-1}\right)\begin{cases}\varepsilon^{-2\frac{k+1}{k+1-m}}\ln\left(\varepsilon^{-1}\right)^2, & \beta=\gamma,\\ \varepsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)\frac{k+1}{k+1-m}}, & \beta<\gamma,\\ \varepsilon^{-2\frac{k+1}{k+1-m}}, & \beta>\gamma,\end{cases}$$

for any $0<\varepsilon<e^{-1}$.

If $\Phi\colon\Theta\to\mathbb{R}$ *is analytic in* Θ *and if Assumption* A4 *holds for* Φ *with* $\rho>1$, *then for any* $0<\varepsilon$ *there exist parameters* $L\in\mathbb{N}_0$, $n\in\mathbb{N}$, *and* $N\in\mathbb{N}^{L+1}$ *such that the* m*th derivative of the corresponding multilevel Monte Carlo estimator* $\hat{\Phi}_L^{N,n}$ *defined by* (1) *satisfies* (12). *Moreover, for any* $0<\varepsilon<\min(\rho^{-1},\ln(2))$ *the required computational cost is bounded by*

$$\mathfrak{C}\left(\frac{d^m}{d\vartheta^m}\hat{\Phi}_L^{N,n}\right) \lesssim \varepsilon^{-2}\ln\left(\varepsilon^{-1}\right)^{4(1+m)} + \varepsilon^{-2}\ln\left(\varepsilon^{-1}\right)^{3+4m}\begin{cases}\ln\left(\varepsilon^{-1}\right)^2, & \beta=\gamma,\\ \varepsilon^{-\frac{\gamma-\beta}{\alpha}}\ln\left(\varepsilon^{-1}\right)^{\frac{\gamma-\beta}{\alpha}(1+2m)}, & \beta<\gamma,\\ 1, & \beta>\gamma.\end{cases}$$

Proof. The mean squared error can be decomposed into two terms,

$$\begin{aligned}\operatorname{MSE}\left(\frac{d^m}{d\vartheta^m}\hat{\Phi}_L^{N,n}\right) &\equiv \mathbb{E}\left(\left\|\Phi^{(m)}-\frac{d^m}{d\vartheta^m}\hat{\Phi}_L^{N,n}\right\|_{L^\infty(\Theta)}^2\right)\\ &\le 2\left\|\Phi^{(m)}-\frac{d^m}{d\vartheta^m}I_L(\Phi)\right\|_{L^\infty(\Theta)}^2 + 2\mathbb{E}\left(\left\|\frac{d^m}{d\vartheta^m}\left(I_L(\Phi)-\hat{\Phi}_L^{N,n}\right)\right\|_{L^\infty(\Theta)}^2\right),\end{aligned}$$

with the understanding that $I_L(\Phi)\equiv\mathcal{I}_{n_L}\left(\Phi(\theta)\right)=\mathcal{I}_n\left(\Phi(\theta)\right)$ for Φ analytic, since $n_\ell=n$ and $\theta^\ell=(\theta_1,\dots,\theta_n)$ for all $\ell\in\mathbb{N}_0$ in that case; cf. (9).

For $\Phi\in C^{k+1}(\Theta)$ with $\nu>2m$ we proceed as in the proof of Theorem 2.2. In fact, using Assumption A3 and the hypotheses we find that

$$\operatorname{MSE}\left(\frac{d^m}{d\vartheta^m}\hat{\Phi}_L^{N,n}\right) \lesssim {n_L}^{-2(k+1-m)} + {n_L}^{2m}s^{-2\alpha L} + L\sum_{\ell=1}^{L}\frac{{n_\ell}^{2m}}{N_\ell}\left(\ln(n_\ell)s^{-\beta\ell}+{n_\ell}^{-\nu}\right). \tag{13}$$

Following the same strategy as in the proof of Theorem 2.2, we choose

$$L=\left\lceil\alpha^{-1}\log_s\left(\varepsilon^{-1}{n_L}^m\right)\right\rceil,\quad n_\ell=\left\lceil n_L s^{-\frac{\beta}{\nu}(L-\ell)}\right\rceil,\quad n_L=\left\lceil\max\left\{c\varepsilon^{-\frac{\beta}{\alpha\nu-\beta m}},\varepsilon^{-\frac{1}{k+1-m}}\right\}\right\rceil$$

with a generic positive constant c independent of ε such that $n_0\ge 3$, as well as

$$N_l=\left\lceil\varepsilon^{-2}L\sqrt{\frac{s^{-l\beta}{n_l}^{2m}\ln(n_l)}{s^{l\gamma}+n_l}}\sum_{\ell=0}^{L}\sqrt{s^{-\ell\beta}{n_\ell}^{2m}\ln(n_\ell)\left(s^{\ell\gamma}+n_\ell\right)}\right\rceil,\quad 0\le l\le L,$$

to guarantee that the mean squared error is of order $\mathcal{O}(\varepsilon^2)$ at minimal cost. In fact, the corresponding computational cost $\mathfrak{C}\left(\hat{\Phi}_L^{N,n}\right)\lesssim\sum_{\ell=0}^{L}\left(N_\ell(\mathfrak{c}_\ell+n_\ell)+n_\ell\right)$ is bounded by

$$\mathfrak{C}\big(\hat{\Phi}_L^{N,n}\big) \lesssim \varepsilon^{-\frac{\gamma}{\alpha}} n_L^{\frac{\gamma m}{\alpha}} + n_L + \varepsilon^{-2} L \left(\sum_{\ell=0}^{L} \sqrt{s^{-\ell\beta} n_\ell^{2m} \ln(n_\ell)\big(s^{\ell\gamma} + n_\ell\big)} \right)^2 .$$

With L^* as in the proof of Theorem 2.2, only accounting for the modified value of n_L (and thus of n_0), we have

$$\sum_{\ell=0}^{L} \sqrt{s^{-\ell\beta} n_\ell^{2m} \ln(n_\ell)\big(s^{\ell\gamma} + n_\ell\big)} =: S_1 + S_2$$

with

$$\begin{aligned} S_1 &= \sum_{\ell=0}^{L^*} \sqrt{s^{-\ell\beta} n_\ell^{2m} \ln(n_\ell)\big(s^{\ell\gamma} + n_\ell\big)} \lesssim \sum_{\ell=0}^{L^*} \sqrt{s^{-\ell\beta} n_\ell^{2m+1} \ln(n_\ell)} \\ &\lesssim \sqrt{\big(n_L s^{-\frac{\beta}{\nu}L}\big)^{2m+1} \ln(n_{L^*})} \begin{cases} 1 , & \nu > 2m+1 , \\ L^* + 1 , & \nu = 2m+1 , \\ s^{\frac{\beta}{2\nu}(2m+1-\nu)L^*} , & \nu < 2m+1 , \end{cases} \end{aligned}$$

as well as

$$\begin{aligned} S_2 &= \sum_{\ell=L^*+1}^{L} \sqrt{s^{-\ell\beta} n_\ell^{2m} \ln(n_\ell)\big(s^{\ell\gamma} + n_\ell\big)} \lesssim \sum_{\ell=L^*+1}^{L} \sqrt{s^{-\ell(\beta-\gamma)} n_\ell^{2m} \ln(n_\ell)} \\ &\lesssim \sqrt{\big(n_L s^{-\frac{\beta}{\nu}L}\big)^{2m} \ln(n_L)} \begin{cases} L+1 , & \beta\big(1-\frac{2m}{\nu}\big) = \gamma , \\ s^{L\frac{\gamma-\beta(1-\frac{2m}{\nu})}{2}} , & \beta\big(1-\frac{2m}{\nu}\big) < \gamma , \\ s^{-L^*\frac{\beta(1-\frac{2m}{\nu})-\gamma}{2}} , & \beta\big(1-\frac{2m}{\nu}\big) > \gamma , \end{cases} \end{aligned}$$

since $\nu > 2m$. Collecting all error bounds, the claim eventually follows.

The complexity analysis of the multilevel Monte Carlo estimator with one extension operator (i.e., $n \in \mathbb{N}$) is similar to the proof of Theorem 2.1. Therefore we will only present the proof for the case $\Phi\colon \Theta \to \mathbb{R}$ analytic here and omit the details for $\Phi \in C^{k+1}(\Theta)$ with $\nu \leq 2m$ to minimize repetitions. For the sake of completeness, though, we will report the parameter values for constructing the multilevel estimator. For $\Phi\colon \Theta \to \mathbb{R}$ analytic it follows from Assumption A4 that the mean squared error is bounded by

$$\text{MSE}\left(\frac{d^m}{d\vartheta^m} \hat{\Phi}_L^{N,n} \right) \lesssim \rho^{-2n} + n^{4m} \ln(n)^2 s^{-2\alpha L} + n^{4m} \ln(n)^3 \sum_{\ell=0}^{L} \frac{s^{-\beta\ell}}{N_\ell} . \tag{14}$$

Therefore, we chose $n = \lceil \log_\rho(\varepsilon^{-1}) \rceil$, $L = \lceil \alpha^{-1} \log_s(\varepsilon^{-1} \ln(n) n^{2m}) \rceil$, and

$$N_l = \left\lceil \varepsilon^{-2} \ln(n)^3 n^{4m} \sqrt{\frac{s^{-l\beta}}{s^{l\gamma} + n}} \sum_{\ell=0}^{L} \sqrt{s^{-\ell\beta}\big(s^{\ell\gamma} + n\big)} \right\rceil , \quad 0 \leq l \leq L ,$$

which is obtained by minimizing the cost bound $\mathfrak{C}\big(\frac{d^m}{d\vartheta^m}\hat{\Phi}_L^{N,n}\big) \lesssim \sum_{\ell=0}^{L} N_\ell(\mathfrak{c}_\ell + n) + n\ln(n)$, subject to the constraint $\frac{s^{-\beta\ell}}{N_\ell} = \varepsilon^2 n^{-4m}\ln(n)^{-3}$. These choices result in the asserted mean squared error of $\mathcal{O}(\varepsilon^2)$ and also provide the cost bound

$$\begin{aligned}\mathfrak{C}\left(\frac{d^m}{d\vartheta^m}\hat{\Phi}_L^{N,n}\right) &\lesssim n\big(L+\ln(n)\big) + s^{\gamma L} + \varepsilon^{-2}\ln(n)^3 n^{4m}\left(\sum_{\ell=0}^{L}\sqrt{s^{-\ell\beta}\big(s^{\ell\gamma}+n\big)}\right)^2 \\ &\lesssim \varepsilon^{-2}\ln\big(\varepsilon^{-1}\big)^{4(1+m)} + \varepsilon^{-2}\ln\big(\varepsilon^{-1}\big)^{3+4m}\begin{cases}\ln\big(\varepsilon^{-1}\big)^2\,, & \beta=\gamma\,,\\ \varepsilon^{-\frac{\gamma-\beta}{\alpha}}\ln\big(\varepsilon^{-1}\big)^{\frac{\gamma-\beta}{\alpha}(1+2m)}\,, & \beta<\gamma\,,\\ \varepsilon^{\frac{\beta-\gamma}{\alpha}}\ln\big(\varepsilon^{-1}\big)^{-\frac{\beta-\gamma}{\alpha}(1+2m)}\,, & \beta>\gamma\,,\end{cases}\end{aligned}$$

which yields the claim.

Finally, the claim for $\Phi \in C^{k+1}(\Theta)$ with $\nu \le 2m$ is realized upon selecting

$$n = \left\lceil \varepsilon^{-\frac{1}{k+1-m}} \right\rceil\,, \qquad L = \left\lceil \alpha^{-1}\log_s\big(\varepsilon^{-1}n^m\big)\right\rceil\,,$$

as well as

$$N_l = \left\lceil \varepsilon^{-2}\ln(n)n^{2m}\sqrt{\frac{s^{-l\beta}}{s^{l\gamma}+n}}\sum_{\ell=0}^{L}\sqrt{s^{-\ell\beta}\big(s^{\ell\gamma}+n\big)}\right\rceil\,, \quad 0\le l\le L\,. \qquad \blacksquare$$

The complexity results obtained in Theorem 2.4 are direct extensions of the results obtained in Theorems 2.1 and 2.2, respectively, in the sense that Theorem 2.4 with $m = 0$ yields the same complexity results that we already obtained previously.

Remark 2.5. Although the complexity result for the case $\Phi \in C^{k+1}(\Theta)$ appears quite involved, it significantly simplifies for $2m+1 < \nu \le \frac{\beta}{\alpha}(k+1)$, which straightforwardly generalizes the condition $1 < \nu \le \frac{\beta}{\alpha}(k+1)$ for which the refined function estimator is most effective (see Corollary 2.3). In fact, then the computational cost is bounded by

$$\mathfrak{C}\left(\frac{d^m}{d\vartheta^m}\hat{\Phi}_L^{N,n}\right) \lesssim \varepsilon^{-\frac{\gamma}{\alpha}p_0} + \ln\big(\varepsilon^{-1}\big)^2\varepsilon^{-2}\begin{cases}\ln\big(\varepsilon^{-1}\big)^2\,, & \beta\big(1-\frac{2m}{\nu}\big)=\gamma\,,\\ \varepsilon^{-\left(\frac{\gamma-\beta}{\alpha}+\frac{2\beta m}{\alpha\nu}\right)p_0}\,, & \beta\big(1-\frac{2m}{\nu}\big)<\gamma\,,\\ 1\,, & \beta\big(1-\frac{2m}{\nu}\big)>\gamma\,,\end{cases}$$

with $p_0 = 1 + \frac{\beta m}{\alpha\nu-\beta m}$. Since $p_0 \le 1 + \frac{2m}{\nu-2m}$, the refined multilevel Monte Carlo estimator for the mth derivative will be most effective for $\nu \gg m$, providing the standard multilevel Monte Carlo complexity (up to logarithmic factors) in that case. As a matter of fact, the term $\varepsilon^{-\frac{\gamma}{\alpha}p_0}$ may only contribute to the $\beta(1-\frac{2m}{\nu}) \ge \gamma$ cases, since the term $\varepsilon^{-\frac{\gamma}{\alpha}p_0}$ is dominated by the other term in the case $\beta(1-\frac{2m}{\nu}) \le \gamma$ already. For $\beta \ge \beta(1-\frac{2m}{\nu}) \ge \gamma$ we have $2\alpha \ge \min(\gamma,\beta) = \gamma$. However, if $2\alpha > \gamma$, then $\varepsilon^{-\frac{\gamma}{\alpha}p_0} = \mathcal{O}\big(\varepsilon^{-2}\big)$, provided that $\nu \ge \frac{2\beta m}{2\alpha-\gamma}$. Consequently, the refined estimator is most effective for $\nu > 2m+1$ such that $\frac{2\beta m}{2\alpha-\gamma} \le \nu \le \frac{\beta}{\alpha}(k+1)$. Such a ν exists, if and only if k and m are such that $1 \le (1-\frac{\gamma}{2\alpha})\frac{k+1}{m}$, which confirms the heuristic mentioned above, since we expect k to be large whenever ν is.

The observations made in the previous remark yield the following complexity result for $\Phi \in C^{k+1}(\Theta)$ under suitable conditions.

Corollary 2.5. *Suppose that the assumptions of Theorem 2.4 are satisfied and that* $\Phi \in C^{k+1}(\Theta)$. *Let* $\hat{\Phi}_L^{N,n}$ *be the refined multilevel Monte Carlo estimator* (9) *that is identified in Theorem 2.4. If* $2m+1 < \nu \leq (k+1)\frac{\beta}{\alpha}$, *then the computational cost is bounded by*

$$\mathfrak{C}\left(\frac{d^m}{d\vartheta^m}\hat{\Phi}_L^{N,n}\right) \lesssim \varepsilon^{-\frac{\gamma}{\alpha}p_0} + \ln\left(\varepsilon^{-1}\right)^2\varepsilon^{-2}\begin{cases}\ln\left(\varepsilon^{-1}\right)^2, & \beta\left(1-\frac{2m}{\nu}\right)=\gamma,\\ \varepsilon^{-\left(\frac{\gamma-\beta}{\alpha}+\frac{2\beta m}{\alpha\nu}\right)p_0}, & \beta\left(1-\frac{2m}{\nu}\right)<\gamma,\\ 1, & \beta\left(1-\frac{2m}{\nu}\right)>\gamma,\end{cases}$$

for any $0 < \varepsilon < e^{-1}$ *with* p_0 *as in Theorem 2.4. If, moreover,* $2\alpha > \gamma$ *and* $\max\{2m+2, \frac{2\beta m}{2\alpha-\gamma}\} \leq \nu \leq (k+1)\frac{\beta}{\alpha}$, *then the cost satisfies*

$$\mathfrak{C}\left(\frac{d^m}{d\vartheta^m}\hat{\Phi}_L^{N,n}\right) \lesssim \ln\left(\varepsilon^{-1}\right)^2\varepsilon^{-2}\begin{cases}\ln\left(\varepsilon^{-1}\right)^2, & \beta\left(1-\frac{2m}{\nu}\right)=\gamma,\\ \varepsilon^{-\left(\frac{\gamma-\beta}{\alpha}+\frac{2\beta m}{\alpha\nu}\right)\frac{\alpha\nu}{\alpha\nu-\beta m}}, & \beta\left(1-\frac{2m}{\nu}\right)<\gamma,\\ 1, & \beta\left(1-\frac{2m}{\nu}\right)>\gamma.\end{cases}$$

3. Approximation of the characteristic function.

The characteristic function $\varphi_Q\colon \mathbb{R} \to \mathbb{C}$ of a random variable Q is given by

$$\varphi_Q(\vartheta) := \mathbb{E}\big(\exp(\mathrm{i}\vartheta Q)\big),$$

and it completely defines the random variable's probability distribution. The characteristic function is thus a convenient and alternative tool to (analytically) characterize the distribution of the random variable Q compared to an approach based on the CDF, which may be cumbersome, for example, in the presence of atoms (i.e., for a mixed distribution). Notice that the characteristic function always exists, since $\exp(\mathrm{i}tQ)$ is bounded. Finally, it is noteworthy that the characteristic function can also be defined for vector-valued random variables.

3.1. The multilevel estimator.

In view of Euler's formula, the characteristic function of Q can be written as

$$\varphi_Q(\vartheta) = \mathbb{E}\big(\cos(\vartheta Q)\big) + \mathrm{i}\mathbb{E}\big(\sin(\vartheta Q)\big) \equiv \Phi_1(\vartheta) + \mathrm{i}\Phi_2(\vartheta)$$

with $\Phi_r(\vartheta) := \mathbb{E}\big(\phi_r(\vartheta, Q)\big)$ for $r \in \{1,2\}$, where $\phi_1(\vartheta, Q) = \cos(\vartheta Q)$ and $\phi_2(\vartheta, Q) = \sin(\vartheta Q)$. A natural approximation of the characteristic function φ_Q on the interval $\Theta \subset \mathbb{R}$ is therefore to use the results from section 2 by simultaneously constructing multilevel Monte Carlo approximations to both Φ_1 and Φ_2. That is, we consider an approximation of the form

$$\hat{\varphi}_L^{N,n} := \sum_{\ell=0}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\left(I_\ell\Big[\varphi\Big(\cdot, Q_\ell^{(i,\ell)}\Big)\Big] - I_{\ell-1}\Big[\varphi\Big(\cdot, Q_{\ell-1}^{(i,\ell)}\Big)\Big]\right), \tag{15}$$

where $\varphi(\vartheta, Q) = \phi_1(\vartheta, Q) + \mathrm{i}\phi_2(\vartheta, Q)$. Recall that this representation contains the basic multilevel Monte Carlo estimator of the form (1) (i.e., $n \in \mathbb{N}$) when $n_\ell = n$ and $\theta^\ell = (\theta_1, \dots, \theta_n)$ for all levels $\ell \in \mathbb{N}_0$. Alternatively, one could, of course, directly approximate $\vartheta \mapsto \mathbb{E}\big(\exp(\mathrm{i}\vartheta Q)\big)$.

This would, however, require one to extend the framework presented in section 2 to complex-valued functions. It is moreover noteworthy that the functions ϕ_r, $r \in \{1,2\}$, are evaluated in the same random samples for both values of r. Consequently, the additional effort for approximating two functions instead of one is negligible. As before, the accuracy of such an approximation to the characteristic function φ_Q is quantified through the mean squared error, viz.,

$$\mathrm{MSE}\left(\hat{\varphi}_L^{N,n}\right) := \mathbb{E}\Big(\big\|\hat{\varphi}_L^{N,n} - \varphi_Q\big\|^2_{L^\infty(\Theta,\mathbb{C})}\Big) .$$

The following result, which is essentially a special case of Theorems 2.1 and 2.2, then characterizes the computational complexity of the multilevel Monte Carlo approximation to φ_Q. Here, we limit the discussion to the case that $\nu > 0$ for $\Phi_1, \Phi_2 \in C^{k+1}(\Theta)$ for brevity; the result for $\nu \le 0$ follows accordingly.

Corollary 3.1. *Let $\varphi_Q = \Phi_1 + \mathrm{i}\Phi_2$ with $\Phi_r(\vartheta) := \mathbb{E}\big(\phi_r(\vartheta, Q)\big)$ for $r \in \{1,2\}$, where $\phi_1(\vartheta, Q) = \cos(\vartheta Q)$ and $\phi_2(\vartheta, Q) = \sin(\vartheta Q)$. Suppose that the hypotheses (i)–(iii) of Theorem 2.1 are satisfied for both ϕ_1 and ϕ_2.*

For $\Phi_1, \Phi_2 \in C^{k+1}(\Theta)$ with $k \in \mathbb{N}_0$ such that Assumption A1 is satisfied for that k, suppose that hypotheses (iv) and (v) of Theorem 2.2 are satisfied with $\nu := \nu_2 - 2\nu_1 > 0$. Then for any $0 < \varepsilon$ there exist parameters $L \in \mathbb{N}_0$, $n = (n_0, \dots, n_L) \in \mathbb{N}^{L+1}$, and $N \in \mathbb{N}^{L+1}$ such that the corresponding refined multilevel Monte Carlo estimator $\hat{\varphi}_L^{N,n}$ (15) satisfies

$$\mathbb{E}\Big(\big\|\hat{\varphi}_L^{N,n} - \varphi_Q\big\|^2_{L^\infty(\Theta,\mathbb{C})}\Big) = \mathcal{O}(\varepsilon^2) . \tag{16}$$

Furthermore, for any $0 < \varepsilon < e^{-1}$ the associated computational cost $\mathfrak{C}\big(\hat{\varphi}_L^{N,n}\big)$ is bounded by

$$\mathfrak{C}\big(\hat{\Phi}_L^{N,n}\big) \lesssim \varepsilon^{-p_1} + \max\Big\{1, \varepsilon^{\frac{\beta}{\alpha\nu} - \frac{1}{k+1}}\Big\}\varepsilon^{-(2+p_2)}\ln\big(\varepsilon^{-1}\big)^{2+p_3} + \ln\big(\varepsilon^{-1}\big)^2 \begin{cases} \varepsilon^{-2}\ln\big(\varepsilon^{-1}\big)^2 & \text{if } \beta = \gamma , \\ \varepsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)} & \text{if } \beta < \gamma , \\ \varepsilon^{-2} & \text{if } \beta > \gamma , \end{cases}$$

where p_1, p_2, and p_3 are as in Theorem 2.2.

If $\Phi_1, \Phi_2 \colon \Theta \to \mathbb{R}$ are analytic in Θ and if Assumption A2 holds for Φ_1, Φ_2 with $\rho > 1$, then for any $\varepsilon > 0$ there exist parameters $L \in \mathbb{N}_0$, $n \in \mathbb{N}$, and $N \in \mathbb{N}^{L+1}$ such that the corresponding estimator $\hat{\varphi}_L^{N,n}$ (15) with $n_\ell = n$ for all $\ell \in \mathbb{N}_0$ satisfies (16). Moreover, for any $0 < \varepsilon < \min(\rho^{-1}, \ln(2))$ the required computational cost is bounded by

$$\mathfrak{C}\big(\hat{\varphi}_L^{N,n}\big) \lesssim \varepsilon^{-2}\ln\big(\varepsilon^{-1}\big)^4 + \ln\big(\varepsilon^{-1}\big)^3 \begin{cases} \varepsilon^{-2}\ln\big(\varepsilon^{-1}\big)^2 & \text{if } \beta = \gamma , \\ \varepsilon^{-\left(2+\frac{\gamma-\beta}{\alpha}\right)}\ln\big(\varepsilon^{-1}\big)^{\frac{\gamma-\beta}{\alpha}} & \text{if } \beta < \gamma , \\ \varepsilon^{-2} & \text{if } \beta > \gamma . \end{cases}$$

Proof. For $r \in \{1,2\}$, let $\hat{\Phi}_{r|L}^{N,n}$ be the multilevel Monte Carlo estimator of Φ_r so that $\hat{\varphi}_L^{N,n} = \hat{\Phi}_{1|L}^{N,n} + \mathrm{i}\hat{\Phi}_{2|L}^{N,n}$ in view of the linearity of the extension operator; cf. (9). The mean squared error of $\hat{\varphi}_L^{N,n}$ can thus be decomposed as

$$\mathbb{E}\Big(\big\|\hat{\varphi}_L^{N,n} - \varphi_Q\big\|^2_{L^\infty(\Theta,\mathbb{C})}\Big) \le \mathbb{E}\Big(\big\|\Phi_1 - \hat{\Phi}_{1|L}^{N,n}\big\|^2_{L^\infty(\Theta)}\Big) + \mathbb{E}\Big(\big\|\Phi_2 - \hat{\Phi}_{2|L}^{N,n}\big\|^2_{L^\infty(\Theta)}\Big) .$$

As the terms on the right-hand side are nothing else but the mean squared errors corresponding to a multilevel Monte Carlo approximation of Φ_1 and Φ_2, respectively, the claims follow directly from Theorems 2.1 and 2.2, respectively. ∎

3.2. Moment estimation. As mentioned above, the characteristic function φ_Q of a random variable Q completely characterizes its probability distribution. As a consequence, moments of Q can be directly derived from φ_Q. In fact, if a random variable Q has moments up to Mth order, then $\varphi_Q \in C^M(\mathbb{R})$ and it holds that

$$\mathbb{E}\big(Q^m\big) = (-\mathrm{i})^m \varphi_Q^{(m)}(0)\,, \quad 0 \le m \le M\,.$$

In view of this identity, it appears natural to use an approximation to φ_Q to derive approximations to the first $M \in \mathbb{N}$ moments simultaneously. This task can be approached from two different perspectives. The first one is to tune the general multilevel formulation to the construction of moments, that is, to minimize the computational cost for constructing a multilevel Monte Carlo estimator for φ_Q subject to the constraint that the maximum mean squared moment error is of order $\mathcal{O}(\varepsilon^2)$. The second perspective is to view the computation of moments simply as a pure postprocessing step. That is, assuming one has access to an approximation to the characteristic function φ_Q with (uniform) mean squared error of order $\mathcal{O}(\varepsilon^2)$, it remains to quantify how big the corresponding maximum mean squared moment error is, if moment estimates are computed from said approximation.

First, we present the result based on the classic multilevel Monte Carlo perspective, i.e., using the constraint minimization approach for the maximum mean squared moment error.

Corollary 3.2. *Let $M \in \mathbb{N}_0$ and $\varphi_Q = \Phi_1 + \mathrm{i}\Phi_2$ with $\Phi_r(\vartheta) := \mathbb{E}\big(\phi_r(\vartheta, Q)\big)$ for $r \in \{1,2\}$, where $\phi_1(\vartheta, Q) = \cos(\vartheta Q)$ and $\phi_2(\vartheta, Q) = \sin(\vartheta Q)$. Furthermore, let the interval $\Theta \subset \mathbb{R}$ be such that $0 \in \mathring{\Theta}$. Suppose that the hypotheses (i)–(iii) of Theorem 2.1 are satisfied for both ϕ_1 and ϕ_2.*

For $\Phi \in C^{k+1}(\Theta)$ with $M \le k \in \mathbb{N}$ such that Assumption A3 is satisfied for that k, suppose that hypotheses (iv) and (v) of Theorem 2.2 are satisfied with $\nu := \nu_2 - 2\nu_1 > 2M$. Then for any $0 < \varepsilon$ there exist parameters $L \in \mathbb{N}_0$, $n = (n_0, \dots, n_L) \in \mathbb{N}^{L+1}$, and $N \in \mathbb{N}^{L+1}$ such that the mth derivative of corresponding refined multilevel Monte Carlo estimator $\hat{\varphi}_L^{N,n}$ (15) satisfies

$$\max_{0\le m\le M} \mathbb{E}\left(\left|\mathbb{E}(Q^m) - (-\mathrm{i})^m \big(\hat{\varphi}_L^{N,n}\big)^{(m)}(0)\right|^2\right) = \mathcal{O}(\varepsilon^2)\,. \tag{17}$$

Furthermore, for any $0 < \varepsilon < e^{-1}$ the associated computational cost $\mathfrak{C}_M$ is bounded by

$$\begin{aligned}\mathfrak{C}_M \lesssim\ & \varepsilon^{-\frac{\gamma}{\alpha}p_0} + \varepsilon^{-p1} + \max\Big\{1, \varepsilon^{(2M+1)\frac{\beta(k+1)-\alpha\nu}{\alpha\nu(k+1-M)}}\Big\}\varepsilon^{-(2+p_2)}\ln\big(\varepsilon^{-1}\big)^{2+p_3} \\ & + \max\Big\{1, \varepsilon^{2M\frac{\beta(k+1)-\alpha\nu}{\alpha\nu(k+1-M)}}\Big\}\ln\big(\varepsilon^{-1}\big)^2\varepsilon^{-2}\begin{cases}\ln\big(\varepsilon^{-1}\big)^2\,, & \beta\big(1-\frac{2M}{\nu}\big) = \gamma\,,\\ \varepsilon^{-\left(\frac{\gamma-\beta}{\alpha}+\frac{2\beta M}{\alpha\nu}\right)p_0}\,, & \beta\big(1-\frac{2M}{\nu}\big) < \gamma\,,\\ 1\,, & \beta\big(1-\frac{2M}{\nu}\big) > \gamma\,,\end{cases}\end{aligned}$$

where p_0, p_1, p_2, and p_3 are as in Theorem 2.4 upon replacing m by M.

If $\Phi_1, \Phi_2 \colon \Theta \to \mathbb{R}$ are analytic in Θ and if Assumption A4 holds for Φ_1, Φ_2 with $\rho > 1$, then for any $0 < \varepsilon$ there exist parameters $L \in \mathbb{N}_0$, $n \in \mathbb{N}$, and $N \in \mathbb{N}^{L+1}$ such that the multilevel Monte Carlo estimator $\hat{\varphi}_L^{N,n}$ in (15) with $n_\ell = n$ for all $\ell \in \mathbb{N}_0$ satisfies (17). Moreover, for any $0 < \varepsilon < \min(\rho^{-1}, \ln(2))$ the required computational cost $\mathfrak{C}_M$ is bounded by

$$\mathfrak{C}_M \lesssim \varepsilon^{-2}\ln\big(\varepsilon^{-1}\big)^{4(1+M)} + \varepsilon^{-2}\ln\big(\varepsilon^{-1}\big)^{3+4M} \begin{cases} \ln\big(\varepsilon^{-1}\big)^2 , & \beta = \gamma , \\ \varepsilon^{-\frac{\gamma-\beta}{\alpha}} \ln\big(\varepsilon^{-1}\big)^{\frac{\gamma-\beta}{\alpha}(1+2M)} , & \beta < \gamma , \\ 1 , & \beta > \gamma . \end{cases}$$

Proof. The mean squared error of the estimated mth moment, $0 \le m \le M$, based on the characteristic function approximation is bounded by

$$\begin{aligned} \mathbb{E}\left(\left| \varphi_Q^{(m)}(0) - \big(\hat{\varphi}_L^{N,n}\big)^{(m)}(0) \right|^2 \right) &\le \mathbb{E}\left(\sup_{\vartheta \in \Theta} \left| \varphi_Q^{(m)}(\vartheta) - \big(\hat{\varphi}_L^{N,n}\big)^{(m)}(\vartheta) \right|^2 \right) \\ &\le \sum_{r=1}^{2} \mathbb{E}\left(\left\| \Phi_r^{(m)} - \frac{d^m}{d\vartheta^m} \hat{\Phi}_{r|L}^{N,n} \right\|_{L^\infty(\Theta)}^2 \right) , \end{aligned}$$

where $\hat{\Phi}_{r|L}^{N,n}$ denotes the multilevel Monte Carlo estimator of Φ_r. Each term of the sum above is a mean squared error of a multilevel Monte Carlo derivative estimator as it has been analyzed in the proof of Theorem 2.4. There, it was also shown that these errors increase monotonically as m increases (ignoring constants) (see relations (13) and (14), resp.), so that the claims follow. ∎

Remark 3.1. We emphasize that although we present here only the general, rather involved, complexity result for the case $\Phi_r \in C^{k+1}(\Theta)$, an improved complexity result closer to the usual multilevel Monte Carlo complexity can be derived under certain conditions. In fact, since the Corollary above is a direct consequence of Theorem 2.4, replacing m by M in Corollary 2.5 shows that a complexity of $\mathcal{O}\big(\varepsilon^{-2}\ln(\varepsilon^{-1})^2\big)$ is possible. We do not state the complete specialized result here for the sake of minimizing repetitions. The same motivation holds for not presenting the results for $\nu \le 2M$ for $C^{k+1}(\Theta)$ functions, which are as in Theorem 2.4.

Next, we address the postprocessing scenario, i.e., the case when moment approximations are computed by differentiating a previously obtained multilevel Monte Carlo approximation of φ_Q. Given this function estimator (e.g., in terms of a spline or a global polynomial), derivatives are straightforwardly available and the computational cost to obtain them is negligible compared to the cost of constructing the functional multilevel estimator, i.e., moment approximations come for (almost) free. However, as the function estimator has not been tuned for the estimation of moments, these approximations may not satisfy the same tolerance request. The result below quantifies the loss of accuracy of this postprocessing approach.

Corollary 3.3. *Let $\hat{\varphi}_L^{N,n}$ be the approximation of φ_Q with (uniform) mean squared error $\mathcal{O}(\varepsilon^2)$ that is identified in Corollary 3.1. Furthermore, let the interval $\Theta \subset \mathbb{R}$ be such that $0 \in \mathring{\Theta}$. Then for any $m \in \mathbb{N}_0$ it holds that*

$$\mathbb{E}\left(\left| \mathbb{E}(Q^m) - (-\mathrm{i})^m \big(\hat{\varphi}_L^{N,n}\big)^{(m)}(0) \right|^2 \right) \lesssim \varepsilon^2 \begin{cases} \varepsilon^{-\max\left\{\frac{\beta}{\alpha\nu}, \frac{1}{k+1}\right\} 2m} , & \Phi_1, \Phi_2 \in C^{k+1}(\Theta) \, , m \le k , \\ \ln\big(\varepsilon^{-1}\big)^{4m} , & \Phi_1, \Phi_2 \text{ analytic} . \end{cases}$$

Proof. For $m \in \mathbb{N}_0$, denote by MSE_m the mean squared error of the mth moment, that is,

$$\mathrm{MSE}_m := \mathbb{E}\left(\left| \mathbb{E}(Q^m) - (-\mathrm{i})^m \big(\hat{\varphi}_L^{N,n}\big)^{(m)}(0) \right|^2 \right) .$$

This error can be bounded by

$$\mathrm{MSE}_m \leq 2 \sum_{r=1}^{2} \mathbb{E}\left(\left\| \Phi_r^{(m)} - \frac{d^m}{d\vartheta^m} I_L(\Phi_r) \right\|_{L^\infty(\Theta)}^2 + \left\| \frac{d^m}{d\vartheta^m} \big(I_L(\Phi_r) - \hat{\Phi}_{r|L}^{N,n} \big) \right\|_{L^\infty(\Theta)}^2 \right) ,$$

with $\hat{\Phi}_{r|L}^{N,n}$ being the multilevel Monte Carlo estimator of Φ_r. As $\hat{\varphi}_L^{N,n}$ is the $\mathcal{O}(\varepsilon^2)$ mean squared error approximation of φ_Q identified in Corollary 3.1, we have in particular that $\mathbb{E}\Big(\big\| I_L(\Phi_r) - \hat{\Phi}_{r|L}^{N,n} \big\|_{L^\infty(\Theta)}^2 \Big) = \mathcal{O}(\varepsilon^2)$ for $r \in \{1, 2\}$. In view of Assumption A3 and Assumption A4, respectively, we thus find

$$\mathbb{E}\left(\left| \mathbb{E}(Q^m) - (-\mathrm{i})^m \big(\hat{\varphi}_L^{N,n}\big)^{(m)}(0) \right|^2 \right) \lesssim \varepsilon^2 \begin{cases} {n_L}^{2m} , & \Phi_1, \Phi_2 \in C^{k+1}(\Theta) \, , m \leq k \, , \\ n^{4m} , & \Phi_1, \Phi_2 \text{ analytic} \, , \end{cases}$$

from which the claim follows. ∎

Related work on multilevel Monte Carlo estimators for higher order moments has recently been presented in [8, 30]. There, the authors study multilevel estimators for central moments, in contrast to the novel results for (raw) moments above. Further conceptual differences include the fact that the main underlying assumptions for our framework (i.e., hypotheses (i)–(iii) of Theorem 2.1) are independent of M, while the settings in [8, 30] require uniformity of related assumptions with respect to the central moment order smaller or equal to M. This uniformity assumption implies the need for considering the worst case scenarios. On the other hand, the computational complexity of the moment estimators presented here (Corollary 3.2) depends on M, while the complexity result in the aforementioned works does not and offers the standard multilevel Monte Carlo complexity. We also note that the simultaneous moment estimators presented here are a byproduct of the function estimators, while references [8, 30] focus on the estimation of central moments. Finally, we reiterate that the assumptions of Corollary 3.2 are conceptually different than the ones used in [8, 30] as they are independent of M, which makes Corollary 3.2 more amenable for some applications, such as for those where no theoretical results for the underlying rate assumptions are available. It is worthwhile pointing out, however, that one could, of course, strengthen the assumptions of our framework to be comparable to settings of [8, 30] by imposing corresponding hypotheses on the first M derivatives of ϕ_1 and ϕ_2, respectively. As a consequence of Theorem 2.2 one would then also obtain standard multilevel Monte Carlo complexity results (up to logarithmic factors) that are independent of M.

3.3. Numerical example. To illustrate the performance of the multilevel Monte Carlo estimators presented above, we consider a stochastic differential equation (SDE) model that is used to describe a financial (European) call option. Specifically, we consider one asset that follows a geometric Brownian motion

$$dS = rS\,dt + \sigma S\,dW \, , \quad S(0) = S_0 \, , \tag{18}$$

and the quantity of interest Q is the corresponding discounted "payoff," which is given by

$$Q := e^{-rT} \max\big(S(T) - K, 0\big) \,, \quad T > 0 \,,$$

where $K > 0$ denotes the strike price. It is interesting to note that the random variable Q does not have a continuous distribution. In fact, it has an atom at the origin, in the sense that $\mathbb{P}(Q = 0) = P\big(S(T) \le K\big) > 0$, since $S(T)$ is log-normally distributed with mean $S_0 e^{rT}$ and variance ${S_0}^2 e^{2rT}\big(e^{\sigma^2 T} - 1\big)$ for any $T > 0$.

The characteristic function of a log-normally distributed random variable is not analytic in the origin; see [26]. Consequently, the characteristic function of the derived quantity Q is also not analytic in the origin. In what follows, we will therefore focus only on the finite regularity versions of the presented approximation techniques. Furthermore, no closed-form expression for the characteristic function of a log-normally distributed random variable is known. However, various approximating formulas, mainly based on an asymptotic expansion, exist in the literature; see, e.g., [2]. In any case, we are not aware of a closed form expression for the characteristic function of the derived quantity $Q = e^{-rT} \max\big(S(T) - K, 0\big)$. To compute a reference solution for the numerical experiments that follow, we proceed as follows. Let $f_{S(T)}$ denote the probability density function of the asset $S(T)$; then the characteristic function of Q can be expressed as

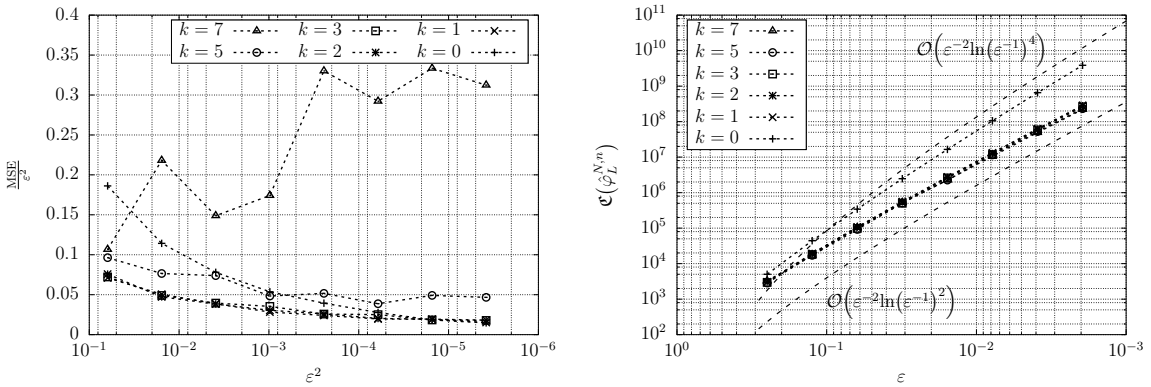
$$\varphi_Q(\vartheta) = \frac{1}{2} - \frac{1}{2} \operatorname{erf}\left(\frac{\big(r - \frac{\sigma^2}{2}\big)T + \ln\big(\frac{S_0}{K}\big)}{\sqrt{2T}\sigma} \right) + e^{rT} \int_0^\infty f_{S(T)}\big(K + q e^{rT}\big) e^{\mathrm{i}\vartheta q} \, dq \,,$$

where $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-s^2} \, ds$. A highly accurate numerical reference solution can then be obtained by using a symbolic software package such as Maple.

For the numerical experiment, we discretize the SDE (18) via the Milstein scheme with uniform time step [27, Chap. 10.3], which reads

$$\bar{S}_{m+1} = \bar{S}_m \Big(1 + rh + \sigma\sqrt{h}\xi_m + \frac{\sigma^2}{2} h\big({\xi_m}^2 - 1\big)\Big) \,, \quad \bar{S}_0 = S_0 \,,$$

so that $\bar{S}_m \approx S(mh)$ for $h = T/m_T$ and $0 \le m \le m_T$. Here, $\big(\xi_m\big)_{m \ge 0}$ denotes a sequence of i.i.d. standard normally distributed random variables. The hierarchy of approximations is constructed based on using a time step $h_\ell = T 2^{-\ell}$ on level ℓ. Figure 1 illustrates the results corresponding to the parameter values $r = \frac{1}{20}$, $\sigma = \frac{1}{5}$, $T = 1$, $K = 10$, and $S_0 = 10$. Specifically, the multilevel Monte Carlo estimator is constructed to approximate the characteristic function φ_Q uniformly on the interval $\Theta = [-1, 1]$, using the parameters L, $n = (n_0, \dots, n_L)$, and $N = (N_0, \dots, N_L)$ as identified in Theorem 2.2 with a spline interpolation of degree k in uniform nodes. Figure 1(A) then shows the accuracy of the multilevel Monte Carlo estimator of φ_Q for various values of k, where we have used $\nu = 2(k+1) - 1$ in view of Remark 2.3. Recall that the parameter k corresponds to the regularity of the real and imaginary parts of the characteristic function. In practice this value is usually unknown. On the one hand, as long as the used value of k does not overestimate the true regularity, Corollary 3.1 still holds. On the other, a very small value of k may result in a poorer computational complexity in general (depending on α, β, and ν), which is, however, not the case here as we will see below.



(A) error vs. tolerance demand

(B) comp. complexity

Figure 1. *Performance of the multilevel Monte Carlo approximation of the characteristic function φ_Q of $Q = e^{-rT}\max\big(S(T)-K,0\big)$ for various values of k.*

For the current example we know that the real and imaginary parts of φ_Q are smooth, and we use $k \in \{0,1,2,3,5,7\}$ to illustrate the performance of the multilevel Monte Carlo estimator with respect to the regularity parameter k. In fact, in Figure 1(A) the mean squared error of each estimator (denoted by MSE in the figure; it is approximated by repeating the experiment 100 times) is compared to the squared tolerance demand ε^2. We observe that the MSE is clearly in the range of $\mathcal{O}(\varepsilon^2)$ for all cases of k. Moreover, Figure 1(A) also indicates that the implementation based on L, n_ℓ, and N_ℓ as in the proof of Theorem 2.2 is actually conservative for this example, in the sense that it produces estimates that are more accurate than required, since the true regularity is drastically underestimated for small values of k. Figure 1(B) illustrates the computational complexity for the same values of k. For this example, one finds that the rates characterizing the hypotheses (i)–(iii) of Theorem 2.1 are $\alpha = 1$, $\beta = 2$, and $\gamma = 1$. Corollary 3.1 thus states that the computational complexity of computing the multilevel Monte Carlo estimator $\hat{\varphi}_L^{N,n}$ of φ_Q is bounded by

$$\mathfrak{C}\big(\hat{\varphi}_L^{N,n}\big) \lesssim \varepsilon^{-2}\ln\big(\varepsilon^{-1}\big)^2 \begin{cases} \ln\big(\varepsilon^{-1}\big)^2\,, & k=0\,,\\ 1\,, & k\geq 1\,, \end{cases}$$

since $\nu = 2k+1$. This is confirmed by the results shown in Figure 1(B). In fact, for $k=0$ we observe a computational complexity of order $\mathcal{O}\big(\varepsilon^{-2}\ln(\varepsilon^{-1})^4\big)$ when plotting the theorized cost model $\sum_{\ell=0}^{L} N_\ell\big(s^{\ell\gamma}+n_\ell\big)+\sum_{\ell=0}^{L} n_\ell$ as $\mathfrak{C}\big(\hat{\varphi}_L^{N,n}\big)$, while the corresponding curves for $k\geq 1$ follow the $\mathcal{O}\big(\varepsilon^{-2}\ln(\varepsilon^{-1})^2\big)$ order line. As a matter of fact, the computational complexities for the values of $k\geq 1$ considered here seem to approximately collapse to the same line, indicating that the proportionality constant may not depend on k, or that the dependence is only very mild at least. Moreover, these examples indicate that the computational cost increases more slowly than the order line for $k\geq 1$, which suggests that the theoretical cost bound may not be strict but rather conservative here.

4. Approximation of the CDF. One of the most commonly used way to characterize the distribution of a random variable Q is via its CDF $F_Q\colon \mathbb{R} \to [0,1]$, which is given by

$$F_Q(\vartheta) := \mathbb{P}(Q \leq \vartheta) \, .$$

In view of the identity $\mathbb{P}(Q \leq \vartheta) = \mathbb{E}\big(I(Q \leq \vartheta)\big)$ it may appear that the CDF F_Q is straightforwardly amenable to a multilevel Monte Carlo approximation via the techniques developed in section 2. This is, however, not the case. In fact, a naive approach based on $F_Q(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big)$ with $\phi(\vartheta, Q) = I(Q - \vartheta \leq 0)$ will suffer due to the discontinuity of the function $\phi(\vartheta, \cdot)$, in the sense that the rate β characterizing the hypothesis (ii) in Theorem 2.1 will deteriorate. Furthermore, the discontinuity also means that a numerical estimation (so-called screening procedure) of both rates α (characterizing the hypothesis (i)) and β will become prohibitive, because many samples on fine levels may be required to adequately resolve the effects of the discontinuity. Although it is possible, e.g., using the results in [4], to derive sufficient conditions for the hypotheses that are also amenable for a numerical treatment, the resulting rates are, however, not optimal and deteriorate in fact, making this approach disadvantageous.

The issues related to using the discontinuous function $I(\cdot \leq 0)$ when approximating the CDF have also been noted in [20]. There, the authors remedy this defect by introducing a carefully constructed regularized version g_δ, say, of $I(\cdot \leq 0)$, where the tuning of the regularization parameter δ is part of the presented complexity analysis. Consequently, the regularization parameter $\delta \equiv \delta(\varepsilon)$ depends on the tolerance requirement ε. However, as the method presented in [20] relies on rates α and β that depend on said regularization parameter, this means that the hypotheses on the rates are ε dependent. This is not problematic when theoretical results for the rates are available, as is the case for the SDE based examples the authors considered in [20]. But this ε dependency gets critical when the rates need to be estimated numerically (e.g., in an offline screening step). In fact, the hypotheses' dependency on the tolerance parameter ε may turn screening the rates into a prohibitive task, as the screening would have to be carried out for every tolerance demand of interest if the user is not yet certain which tolerance will be most appropriate for the problem at hand. However, it is noteworthy that the authors of [20] have recently presented an adaptive algorithm to overcome this shortcoming in the context of SDE based examples [21]. Finally, we mention that this tolerance dependence can be removed by, e.g., taking advantage of the Lipschitz property of the regularized function g_δ, which would, however, also result in a deterioration of the rates, as a weak error condition was replaced by a strong error condition.

Here, we introduce an alternative approach that does not explicitly require a regularization of the indicator function. Moreover, it relies on easily verifiable rate hypotheses, making this strategy amenable for a wide range of applications. Instead of seeking a CDF approximation directly, our approach is based on finding an appropriate function $\Phi\colon \Theta \to \mathbb{R}$ such that

$$\Phi' = F_Q \, . \tag{19}$$

One candidate to satisfy this identity is

$$\Phi(\vartheta) = \mathbb{E}\big(\phi(Q - \vartheta)\big) \quad \text{with} \quad \phi(z) = |z| I(z \leq 0) \equiv z^- \, ,$$

provided that $F_Q \in C^1(\Theta)$, so that there are no atoms present in Θ. We reiterate that even in the presence of atoms (i.e., if $F_Q \notin C^1(\Theta)$) a characterization of the quantity of interest Q is nonetheless possible using the techniques developed in this work, for example, via the characteristic function (cf. section 3). Notice that the presented antiderivative based approach yields a function ϕ that is continuous (and even Lipschitz continuous), in contrast to the indicator function. Furthermore, as a consequence of identity (19), it follows that the function Φ is even more regular than the CDF F_Q. Consequently, the idea is to use the results presented in section 2 to construct a multilevel Monte Carlo approximation of the CDF with rate hypotheses that are amenable for both theoretical and numerical verification. Specifically, we will first construct a multilevel Monte Carlo approximation for the function Φ such that Φ' is also approximated accurately with respect to a tolerance demand. We then define the multilevel CDF estimator mimicking identity (19). However, for the particular form of the function ϕ, hypothesis (v) of Theorem 2.2 may not hold with $\nu > 2$, and the refined estimator (9) using a hierarchy of extension operators may not be effective. Indeed, the sufficient condition discussed in Remark 2.3 only implies $\nu < 2$, and, as a matter of fact, the covariance function will only be Lipschitz continuous on the diagonal in general. Although the refined estimator may be effective for approximating the function Φ itself (hypothesis (v) of Theorem 2.2 with $m = 0$ requires $\nu > 0$), it may not provide an effective approximation of Φ' that satisfies the tolerance demand. As the estimation of the CDF outlined above is based on estimating the first derivative, we will therefore resort to the multilevel Monte Carlo estimator using one (global) extension operator as in (1) for both cases $\Phi \in C^{k+1}(\Theta)$ and Φ being analytic. The computational complexity is therefore a direct consequence of Theorem 2.4 with $m = 1$ assuming that $\nu \leq 2m$.

Corollary 4.1. *Suppose that the hypotheses (i)–(iii) of Theorem 2.1 are satisfied for* $\phi(\vartheta, Q) = |Q - \vartheta| I(Q \leq \vartheta)$.

If $F_Q \in C^{k+1}(\Theta)$ *with* $k \in \mathbb{N}_0$ *such that Assumption A3 is satisfied for* $k+1$*, then for any* $0 < \varepsilon$ *there exist parameters* $L \in \mathbb{N}_0$, $n \in \mathbb{N}$, *and* $N \in \mathbb{N}^{L+1}$ *such that the corresponding multilevel Monte Carlo CDF estimator* $\hat{F}_L^{N,n} := \frac{d}{d\vartheta}\hat{\Phi}_L^{N,n}$ *satisfies*

$$\mathbb{E}\left(\left\| F_Q - \hat{F}_L^{N,n} \right\|^2_{L^\infty(\Theta)}\right) = \mathcal{O}(\varepsilon^2)\,. \tag{20}$$

Furthermore, for any $0 < \varepsilon < e^{-1}$ *the associated computational cost* $\mathfrak{C}\big(\hat{F}_L^{N,n}\big)$ *is bounded by*

$$\mathfrak{C}\big(\hat{F}_L^{N,n}\big) \lesssim \varepsilon^{-\frac{\gamma}{\alpha}\left(1+\frac{2}{k+1}\right)} + \ln\big(\varepsilon^{-1}\big)\varepsilon^{-\left(2+\frac{3}{k+1}\right)} + \ln\big(\varepsilon^{-1}\big)\varepsilon^{-\left(2+\frac{2}{k+1}\right)} \begin{cases} \ln\big(\varepsilon^{-1}\big)^2\,, & \beta = \gamma\,, \\ \varepsilon^{-\frac{\gamma-\beta}{\alpha}\left(1+\frac{1}{k+1}\right)}\,, & \beta < \gamma\,, \\ 1\,, & \beta > \gamma\,. \end{cases}$$

If $F_Q \colon \Theta \to \mathbb{R}$ *is analytic in* Θ *and if Assumption A4 holds for* F_Q *with* $\rho > 1$*, then for any* $0 < \varepsilon$ *there exist parameters* $L \in \mathbb{N}_0$, $n \in \mathbb{N}$, *and* $N \in \mathbb{N}^{L+1}$ *such that the corresponding multilevel Monte Carlo CDF estimator* $\hat{F}_L^{N,n}$ *satisfies* (20). *Moreover, for any* $0 < \varepsilon < \min(\rho^{-1}, \ln(2))$ *the required computational cost is bounded by*

$$\mathfrak{C}\Big(\hat{F}_L^{N,n}\Big) \lesssim \varepsilon^{-2}\ln\big(\varepsilon^{-1}\big)^8 + \varepsilon^{-2}\ln\big(\varepsilon^{-1}\big)^7 \begin{cases} \ln\big(\varepsilon^{-1}\big)^2\,, & \beta = \gamma\,, \\ \varepsilon^{-\frac{\gamma-\beta}{\alpha}}\ln\big(\varepsilon^{-1}\big)^{\frac{3(\gamma-\beta)}{\alpha}}\,, & \beta < \gamma\,, \\ 1\,, & \beta > \gamma\,. \end{cases}$$

One can, of course, approximate even higher order derivatives of Φ the same way it was done for Φ' above. In fact, using Theorem 2.4 or the analogous version when using only one (global) extension operator as in (1) with $m = 2$ yields the computational complexity for an approximation of the probability density function F_Q'. Related results regarding multilevel Monte Carlo approximations to probability density functions have also been presented in [20], again making use of an explicit regularization procedure. We also mention the work in [7], where the authors construct a multilevel Monte Carlo approximation of the probability density function based on an appropriate moment matching procedure within the maximum entropy framework.

Finally, we remark that the multilevel Monte Carlo CDF approximation $\hat{F}_L^{N,n}$ constructed above may not be a monotonically increasing function on Θ. This can, however, simply be achieved as a postprocessing step when replacing the estimated function $\hat{F}_L^{N,n}$ by the function $\Theta \ni t \mapsto \sup_{\vartheta\in[\min(\Theta),t]} \hat{F}_L^{N,n}(\vartheta)$, since this replacement does not increase the uniform mean squared error (20).

4.1. Estimating quantiles. In addition to characterizing a random variable's distribution via its CDF, another class of important statistical quantities are quantiles. In fact, quantiles are commonly used in various applications, ranging from financial mathematics (often called value-at-risk) to robust design optimization, for example, when specifying failure probabilities or when constructing hypothesis tests. However, quantiles, such as the median, can in general not be expressed in terms of moments of the random variable Q. In fact, the τ-quantile $q \equiv q_\tau$, say, is given by

$$q_\tau = F_Q^{-1}(\tau) \equiv \inf\big\{\vartheta \in \mathbb{R}\colon F_Q(\vartheta) \geq \tau\big\}\,, \quad \tau \in (0,1)\,.$$

Consequently, an efficient quantile estimation using standard multilevel Monte Carlo techniques has been hampered thus far. As mentioned in the section 1 already, first advances have been made in the context of multilevel stochastic approximation algorithms [13, 16]. We also mention the work [14] that addresses the computational complexity of quantile estimation based on the empirical single level CDF estimator, as well as an approximation scheme, in the context of physical models with uncertain inputs. Unlike the aforementioned works, here we will use an available multilevel Monte Carlo CDF approximation to estimate quantiles as a postprocessing step and assess their accuracy.

We do not treat the problem of nonunique quantile estimation, i.e., the case when the CDF is constant in a region of the support. This is a nontrivial problem even in the classic Monte Carlo setting; see [15]. That is, here we only consider the case that there exists a unique root of the function $\vartheta \to F_Q(\vartheta) - \tau$, for $\tau \in (0,1)$ given, so that $F_Q(q) = \tau$. As anticipated above, we therefore define the quantile estimator $\hat{q}$ as any value (if not unique) $\hat{q}$, such that

$$\hat{F}_L^{N,n}(\hat{q}) = \tau\,,$$

with $\hat{F}_L^{N,n}$ being a multilevel Monte Carlo CDF estimator. Notice that the fact that F_Q is (locally) invertible does not imply that the multilevel Monte Carlo approximation $\hat{F}_L^{N,n}$ is so as well. However, this is not a problem, since any $\hat{q}$ satisfying the identity above will be a sufficiently accurate approximation of q. In fact, the result below shows that this post-processed quantile estimator satisfies the same prescribed tolerance goal as the corresponding CDF estimator.

Proposition 4.2. *Let $\hat{F}_L^{N,n}$ be the approximation of the (strictly monotone) CDF $F_Q \in C^1(\Theta)$ with mean squared error $\mathcal{O}(\varepsilon^2)$ that is identified in Corollary 4.1. For $\tau \in (0,1)$, let q be the τ-quantile, in the sense that $q \in F^{-1}(\tau)$, and let $\hat{q}$ be any approximate τ-quantile satisfying $\hat{F}_L^{N,n}(\hat{q}) = \tau$. Furthermore, suppose that the interval $\Theta \subset \mathbb{R}$ is such that $q, \hat{q} \in \Theta$. If $\inf_{\vartheta\in\Theta} F_Q'(\vartheta) > 0$, then*

$$\mathbb{E}\Big(\big(q-\hat{q}\big)^2\Big) = \mathcal{O}(\varepsilon^2)\,.$$

Proof. Set $\hat{F} := \hat{F}_L^{N,n}$. It then follows from the hypotheses and Taylor's theorem that

$$|q-\hat{q}|\,F_Q'(\xi) = |F_Q(q) - F_Q(\hat{q})| = \big|\hat{F}(\hat{q}) - F_Q(\hat{q})\big|$$

for some ξ between $\hat{q}$ and q. We thus find that

$$|q-\hat{q}| \le \frac{1}{\inf_{\vartheta\in\Theta} F_Q'(\vartheta)} \sup_{\vartheta\in\Theta} |\hat{F}(\vartheta) - F(\vartheta)|\,, \tag{21}$$

from which the claim follows. ∎

In view of inequality (21), the proof above also reveals that estimating "rare" quantiles, i.e., quantiles for which $\inf_{\vartheta\in\Theta} F_Q'(\vartheta)$ is small, may not be very accurate. In order to use the developed multilevel Monte Carlo techniques to estimate such quantiles effectively, one will thus have to combine them with specialized techniques, such as rare-event simulation techniques [1, Chap. VI]. We will leave this aspect for future work. It is noteworthy, however, that this is not just a shortcoming of the techniques presented here but is rather a principal one for general purpose sampling techniques. In fact, a factor of the form $F_Q'(q)^{-2}$ is also affecting the (asymptotic) mean squared error results for the classic Monte Carlo based quantile estimators; cf. [36, Chap. 21].

4.2. Simultaneous approximation of the CDF, quantiles, and the conditional value-at-risk. While the previous section assesses the accuracy of a multilevel Monte Carlo quantile estimator via postprocessing a CDF estimator, here we will describe an approach that enables us to extract even more characteristic information about a random variable's distribution via appropriate postprocessing steps. The approach is motivated by the remedy that we already used above to overcome the lack of regularity of the indicator function when constructing a CDF approximation. Specifically, mimicking (19), we can consider a function $\Psi\colon \Theta \to \mathbb{R}$ such that

$$\Psi' = F_Q - \tau \tag{22}$$

for some value $\tau \in (0,1)$ given. A natural candidate is

$$\Psi(\vartheta) = \mathbb{E}\big(\psi(Q-\vartheta)\big) \quad \text{with} \quad \psi(z) = \tau z + |z| I(z \le 0)\,,$$

which also has all the advantageous properties (increased regularity, Lipschitz continuous ψ, etc.) that the function $\vartheta \to \mathbb{E}\big(|Q-\vartheta| I(Q \le \vartheta)\big)$ offered for the CDF estimation. Furthermore, by construction, the function Ψ has the property that if $q \in \Theta$ is such that $F_Q(q) = \tau$, then $q \in \arg\min_{\vartheta\in\Theta} \Psi(\vartheta)$. However, using this minimization approach to approximate the τ-quantile directly based on an order $\mathcal{O}(\varepsilon^2)$ mean squared error approximation of Ψ may not be advisable, due to the minimizer's sensitivity with respect to perturbations of the objective function. Specifically, an order $\mathcal{O}(\varepsilon^2)$ mean squared error approximation of Ψ may only yield an order $\mathcal{O}(\varepsilon)$ mean squared error approximation of the minimizer (i.e., of the quantile), unlike the approach in Proposition 4.2, for which we ensured that the derivative approximation is sufficiently accurate as well. Despite the fact that an approach based on Ψ directly is not advisable in general, the function Ψ nonetheless enables us to derive an effective estimator, for which identity (22) will be essential. In fact, upon noticing that

$$\psi(z) = \tau z + |z| I(z \le 0) = \tau z^+ + (1-\tau) z^- = z^+ - (1-\tau) z \,,$$

we find that

$$\Psi(\vartheta) = (1-\tau)\Big(\vartheta + \frac{1}{1-\tau}\mathbb{E}\big((Q-\vartheta)^+\big) - \mathbb{E}(Q)\Big) \equiv (1-\tau)\big(\Phi(\vartheta) - \mathbb{E}(Q)\big) \,, \tag{23}$$

where

$$\Phi(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big) \,, \quad \text{with} \quad \phi(\vartheta, Q) = \vartheta + \frac{1}{1-\tau}(Q-\vartheta)^+ \,.$$

The function Φ is such that its minimum value is the so-called conditional value-at-risk (see [32]), which is an important robustness indicator in various applications, ranging from robust design optimization to mathematical finance. By definition, (23) shows that

$$\Psi'(\vartheta) = (1-\tau)\Phi'(\vartheta) \,, \tag{24}$$

implying that both Φ and Ψ have the same minimizer, namely, the τ-quantile q. This motivates us to use an appropriate multilevel Monte Carlo approximation to the function Φ with $\tau \in (0,1)$ given, which will then provide approximations to the CDF F_Q, the τ-quantile, and the conditional value-at-risk, respectively, each via appropriate postprocessing steps. In other words, the following procedure offers a simultaneous approximation of these quantities, all of which characterize the distribution of a random variable Q provided that $q \in \Theta$:

1. Use Theorem 2.4 with $m = 1$ to construct a multilevel Monte Carlo approximation $\hat{\Phi}_L^{N,n} \equiv \hat{\Phi}$ of Φ that provides a (uniform) mean squared error of order $\mathcal{O}(\varepsilon^2)$ for both the function and its derivative, that is, $\mathrm{MSE}\big(\hat{\Phi}\big) + \mathrm{MSE}\big(\hat{\Phi}'\big) = \mathcal{O}(\varepsilon^2)$.
2. Construct a CDF approximation $\hat{F}_L^{N,n} \equiv \hat{F}$ of F_Q via postprocessing. In view of equations (22) and (24) it is natural to use $F_Q \approx \hat{F} := (1-\tau)\hat{\Phi}' + \tau$, whose mean-squared error is guaranteed to be of order $\mathcal{O}(\varepsilon^2)$ due to
$$\|F_Q - \hat{F}\|_{L^\infty(\Theta)} = (1-\tau)\|\Phi' - \hat{\Phi}'\|_{L^\infty(\Theta)} < \|\Phi' - \hat{\Phi}'\|_{L^\infty(\Theta)} \,.$$
3. An approximation of the τ-quantile $q = \arg\min_{\vartheta\in\Theta} \Phi(\vartheta)$ is then available via minimization of $\hat{\Phi}$ (or equivalently via root finding of $\vartheta \to \hat{F}(\vartheta) - \tau$), whose mean-squared error is also guaranteed to be of order $\mathcal{O}(\varepsilon^2)$ by Proposition 4.2.

4. An approximation to the conditional value-at-risk $\min_{\vartheta\in\Theta}\Phi(\vartheta)=\Phi(q)$ can be obtained via postprocessing through the approximation $\hat{\Phi}(\hat{q})\approx\Phi(q)$, which is also accurate with respect to the required $\mathcal{O}(\varepsilon^2)$ tolerance in view of

$$\begin{aligned}\big|\Phi(q)-\hat{\Phi}(\hat{q})\big| &\le \big|\Phi(q)-\Phi(\hat{q})\big|+\big|\Phi(\hat{q})-\hat{\Phi}(\hat{q})\big|\\ &\le \frac{1+\tau}{(1-\tau)\inf_{\vartheta\in\Theta}F_Q'(\vartheta)}\|F_Q-\hat{F}\|_{L^\infty(\Theta)}+\|\Phi-\hat{\Phi}\|_{L^\infty(\Theta)}\\ &= \frac{1+\tau}{\inf_{\vartheta\in\Theta}F_Q'(\vartheta)}\|\Phi'-\hat{\Phi}'\|_{L^\infty(\Theta)}+\|\Phi-\hat{\Phi}\|_{L^\infty(\Theta)}\,,\end{aligned}$$

where we have used inequality (21) again, as well as $\|F_Q-\tau\|_{L^\infty(\Theta)}\le 1+\tau$.

Since the quantities in steps 2.–4. are obtained via appropriate postprocessing, the computational complexity of this simultaneous estimation procedure is given by the complexity of constructing the function estimator $\hat{\Phi}_L^{N,n}$ in step 1. which can be quantified as in Corollary 4.1 but for the function $\phi(\vartheta,Q)=\vartheta+\frac{1}{1-\tau}(Q-\vartheta)^+$.

4.3. Numerical example. As a numerical illustration of the multilevel techniques developed above, we consider the random partial differential equation

$$-\Delta u = f \quad \text{in } D=(0,1)^2 \tag{25}$$

with homogeneous Dirichlet boundary conditions and as quantity of interest Q the solution's spatial average. Specifically, we consider the random forcing term f given by

$$f(x)=-72\xi({x_1}^2+{x_2}^2-x_1-x_2)\,,$$

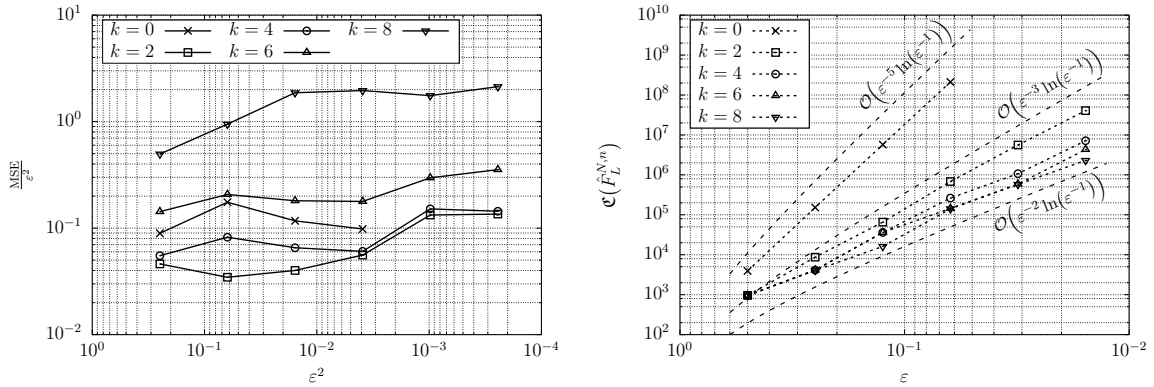
so that the exact solution is $u(x)=36\xi x_1(1-x_1)x_2(1-x_2)$ and the quantity of interest reads

$$Q:=\int_D u\,dx=\xi\,.$$

Here, ξ is a random variable that represents the model uncertainty. Specifically, we consider the case where ξ follows a chi-squared distribution with one degree of freedom, so that quantity of interest's CDF is

$$F_Q(\vartheta)=\begin{cases}\operatorname{erf}\Big(\sqrt{\frac{\vartheta}{2}}\Big)\,, & \vartheta\ge 0\,,\\ 0\,, & \text{else,}\end{cases}$$

where $\operatorname{erf}(z)=\frac{2}{\sqrt{\pi}}\int_0^z e^{-s^2}\,ds$ as before. The boundary value problem (25) is solved numerically via a second order finite difference scheme with a hierarchy of uniform square meshes of $D=(0,1)^2$ corresponding to the mesh sizes $h_\ell=2^{-(\ell+1)}$. Figure 2 showcases the results of the developed multilevel Monte Carlo method on the interval $\Theta=[0,10]$ using the parameters L, n, and $N=(N_0,\dots,N_L)$ as identified in Theorem 2.4 (see also Corollary 4.1) with a spline interpolation of degree $k+1$ in uniform nodes. Specifically, Figure 2(A) shows the accuracy of the multilevel Monte Carlo CDF approximation for various values of the regularity parameter $k\in\{2,4,6,8\}$. Here, the mean squared error of the estimator is approximated by repeating



(A) error vs. tolerance demand (B) comp. complexity

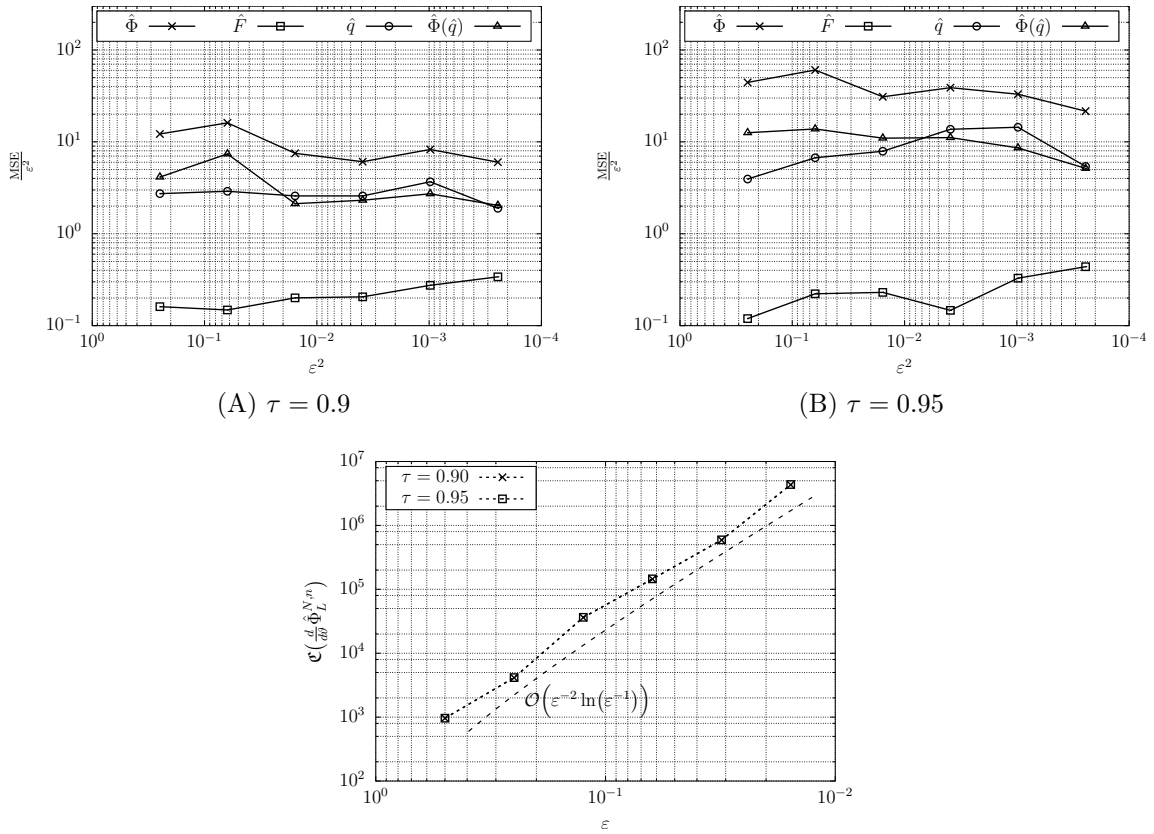
Figure 2. *Performance of the multilevel Monte Carlo approximation of the CDF of the quantity of interest Q for various values of k.*

the experiment 40 times. The accuracy results presented in Figure 2(A) verify that the MSE is clearly in the range of $\mathcal{O}(\varepsilon^2)$ for all cases of k. Figure 2(B) illustrates the computational complexity of the multilevel Monte Carlo CDF approximation for these values of k by plotting the cost model $n+\sum_\ell^L N_\ell(s^{\ell\gamma}+n)$ as $\mathfrak{C}\big(\hat{F}_L^{N,n}\big)$ for each case. For this example, we find that the rates characterizing the hypotheses (i)–(iii) of Theorem 2.1 are $\alpha = 2$, $\beta = 4$, and $\gamma \approx 2.4$, so that we expect a computational complexity of $\mathcal{O}\big(\varepsilon^{-\left(2+\frac{3}{k+1}\right)}\ln\big(\varepsilon^{-1}\big)\big)$ in view of Corollary 4.1. This complexity behavior is confirmed by the results illustrated in Figure 2(B). In fact, as for the numerical examples for the characteristic function discussed in section 3.3, we observe that the measured computational cost increases more slowly than the theoretical upper bound. In particular, the curves seem to follow the $\mathcal{O}\big(\varepsilon^{-2}\ln\big(\varepsilon^{-1}\big)\big)$ order lines for $k \geq 4$, at least for the considered tolerances.

Next, we consider the procedure described in section 4.2 to simultaneously approximate the CDF, a quantile, and the corresponding conditional value-at-risk. As described above, the procedure's starting point is an appropriate multilevel Monte Carlo approximation of the function $\Phi(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big)$, where $\phi(\vartheta, Q) = \vartheta + \frac{1}{1-\tau}(Q-\vartheta)^+$. For the current example it is straightforward to compute the exact function Φ, namely,

$$\Phi(\vartheta) = \frac{1}{(1-\tau)\sqrt{\pi}} \begin{cases} \left((\vartheta-1)\,\mathrm{erf}\left(\sqrt{\vartheta/2}\right)\sqrt{\pi} + (1-\vartheta\tau)\sqrt{\pi} + e^{-\vartheta/2}\sqrt{2\vartheta}\right), & \vartheta \geq 0\,, \\ (1-\tau\vartheta)\sqrt{\pi}\,, & \vartheta < 0\,, \end{cases}$$

which is used to verify the accuracy of the numerical experiments below. The performance of the simultaneous multilevel Monte Carlo estimation procedure is showcased in Figure 3 for two probability levels, where we have fixed the regularity parameter to be $k = 6$ (i.e., $F_Q \in C^7(\Theta)$; cf. also Figure 2). Specifically, Figures 3(A) and 3(B) show the accuracy of the estimated quantities for $\tau = 0.9$ and $\tau = 0.95$, respectively. We observe that the mean squared error (MSE) is clearly in the range of $\mathcal{O}(\varepsilon^2)$ for both values of τ. Here, the MSE is with respect to the uniform norm for both $\hat{\Phi}$ and $\hat{F}$, and with respect to the absolute value for the quantile estimator $\hat{q}$ and the estimated conditional value-at-risk $\hat{\Phi}(\hat{q})$. An interesting feature for both



(A) $\tau = 0.9$ (B) $\tau = 0.95$

(C) comp. complexity

Figure 3. *Performance of the simultaneous multilevel Monte Carlo approximation ($k = 6$) for the CDF, the quantile, and the conditional value-at-risk associated with the quantity of interest Q for two probability levels $\tau = 0.9$ and $\tau = 0.95$.*

values of τ is that the CDF estimator $\hat{F}$ appears to be significantly more accurate (about two orders of magnitude) than the estimator $\hat{\Phi}$. However, this is simply a consequence of the fact that the mean squared error tolerance ε^2 is an absolute error criterion. In fact, the suprema of the CDF F_Q and of the function Φ are also different by one order of magnitude, in the sense that $\|F_Q\|_{L^\infty(\Theta)} \approx 1$, while $\|\Phi\|_{L^\infty(\Theta)} \approx 10$ for $\tau = 0.9$ and $\|\Phi\|_{L^\infty(\Theta)} \approx 20$ for $\tau = 0.95$. A noticeable difference between the quantile and conditional value-at-risk results for $\tau = 0.9$ and the corresponding results for $\tau = 0.95$ is that the error constants are bigger in the $\tau = 0.95$ case. Essentially, the bigger constant is due to the fact that the $\tau = 0.95$ case corresponds to a more extreme quantile (same for the conditional value-at-risk), which may be estimated less accurately in view of (21).

As discussed in section 4.2, the computational cost for the simultaneous estimation procedure is given by the cost of computing one multilevel Monte Carlo approximation of Φ that satisfies a (uniform) mean squared error of order $\mathcal{O}(\varepsilon^2)$ for its derivative, because the other quantities are simply derived from this approximation during postprocessing steps. A direct consequence of Theorem 2.4 and the analogous version when using only one (global) extension

operator as in (1) with $m = 1$ thus is that the simultaneous multilevel Monte Carlo procedure has a computational complexity of $\mathcal{O}\big(\varepsilon^{-\left(2+\frac{3}{k+1}\right)}\ln(\varepsilon^{-1})\big)$, which is of course identical to the multilevel Monte Carlo CDF estimator discussed above. For $k = 6$ this complexity result is confirmed by the results shown in Figure 3(C). In fact, we observe that the computational complexity is almost identical for both values of τ. Moreover, these examples suggest that also the computational complexity of the simultaneous estimation procedure may be better than predicted by the theoretical upper bound, since both curves follow the $\mathcal{O}\big(\varepsilon^{-2}\ln(\varepsilon^{-1})\big)$ line closely.

5. Conclusion. In this work, we have introduced and analyzed a multilevel Monte Carlo framework for the estimation of parametric expectations, that is, of functions, uniformly on an interval. Specifically, we have constructed estimators based on appropriately interpolating pointwise estimators on a collection of points to derive function estimators. Direct applications of this framework include the estimation of a random variable's characteristic function and of its CDF. Furthermore, we have presented an antiderivative based formulation that allows us to construct accurate estimators for both the quantile and the conditional value-at-risk by post-processing suitable multilevel Monte Carlo function approximations. In fact, the procedure introduced here allows us to simultaneously estimate the CDF, the quantile, and the conditional value-at-risk subject to a prescribed mean squared error tolerance. These theoretical findings are illustrated by means of numerical examples.

There are still many interesting questions and extensions left open. One extension of practical importance is to tune the method to provide a mean squared error that does not exceed ε^2, instead of only guaranteeing the error to be of order $\mathcal{O}\big(\varepsilon^2\big)$. Since some of the constants affecting the mean squared error depend on the unknown function that we want to estimate, a method that will meet this strict error criterion is not immediate and will require some new approaches. As a matter of fact, such a fully practical algorithm, based on the theoretical framework provided here, is currently a work in progress [31] and will be presented elsewhere. Related ideas for an adaptive multilevel Monte Carlo algorithm for estimating the CDF have recently also been presented in [21].

An example of a conceptual extension of the results presented in this work is the case of a vector-valued quantity of interest. This will require constructing a function (e.g., via appropriate multivariate polynomials) based on pointwise estimates distributed in a multidimensional set Θ. In order to not dominate the overall complexity in high dimension (cf. "curse of dimensionality"), this function extension has to be tuned carefully. Another important direction will be to incorporate specialized methods for estimating a quantile (also for the conditional value-at-risk) when $\inf_{\vartheta\in\Theta} F_Q'(\vartheta)$ is very small; cf. (21). This may happen, for example, when estimating "rare" quantiles that are located in the tails of a probability distribution. Here, recent works on (multilevel) subset simulation techniques (see, e.g., [3, 35]) appear promising. Another interesting aspect that will require specialized methods is when the CDF F_Q is not very regular, say, only continuous. The interpolation based results presented here are then no longer applicable. These and related topics are part of ongoing work.

Acknowledgment. We are grateful to the anonymous referees for their insightful comments and suggestions, which led to substantial improvements of the paper.

REFERENCES

[1] S. Asmussen and P. W. Glynn, *Stochastic Simulation: Algorithms and Analysis*, Springer, New York, 2007, https://doi.org/10.1007/978-0-387-69033-9.

[2] S. Asmussen, J. L. Jensen, and L. Rojas-Nandayapa, *On the Laplace transform of the lognormal distribution*, Methodol. Comput. Appl. Probab., 18 (2016), pp. 441–458, https://doi.org/10.1007/s11009-014-9430-7.

[3] S.-K. Au and J. L. Beck, *Estimation of small failure probabilities in high dimensions by subset simulation*, Probabilist. Eng. Mech., 16 (2001), pp. 263–277, https://doi.org/10.1016/S0266-8920(01)00019-4.

[4] R. Avikainen, *On irregular functionals of SDEs and the Euler scheme*, Finance Stoch., 13 (2009), pp. 381–401, https://doi.org/10.1007/s00780-009-0099-7.

[5] V. Barthelmann, E. Novak, and K. Ritter, *High dimensional polynomial interpolation on sparse grids*, Adv. Comput. Math., 12 (2000), pp. 273–288, https://doi.org/10.1023/A:1018977404843.

[6] C. Bierig and A. Chernov, *Convergence analysis of multilevel Monte Carlo variance estimators and application for random obstacle problems*, Numer. Math., 130 (2015), pp. 579–613, https://doi.org/10.1007/s00211-014-0676-3.

[7] C. Bierig and A. Chernov, *Approximation of probability density functions by the multilevel Monte Carlo maximum entropy method*, J. Comput. Phys., 314 (2016), pp. 661–681, https://doi.org/10.1016/j.jcp.2016.03.027.

[8] C. Bierig and A. Chernov, *Estimation of arbitrary order central statistical moments by the multilevel Monte Carlo method*, Stoch. Partial Differ. Equ. Anal. Comput., 4 (2016), pp. 3–40, https://doi.org/10.1007/s40072-015-0063-9.

[9] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, 3rd ed., Springer, New York, 2008, https://doi.org/10.1007/978-0-387-75934-0.

[10] C. Canuto, M. Y. Hussainif, A. Quarteroni, and T. A. Zang, *Spectral Methods: Fundamentals in Single Domains*, Springer, New York, 2011.

[11] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup, *Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients*, Comput. Vis. Sci., 14 (2011), pp. 3–15, https://doi.org/10.1007/s00791-011-0160-x.

[12] C. de Boor, *A Practical Guide to Splines*, Springer, New York, 2001.

[13] S. Dereich and T. Mueller-Gronbach, *General Multilevel Adaptations for Stochastic Approximation Algorithms*, Preprint, arXiv:1506.05482 [math.PR], 2015, 1506.05482.

[14] D. Elfverson, D. J. Estep, F. Hellman, and A. Målqvist, *Uncertainty quantification for approximate p-quantiles for physical models with stochastic inputs*, SIAM/ASA J. Uncertain. Quantif., 2 (2014), pp. 826–850, https://doi.org/10.1137/140967039.

[15] D. Feldman and H. G. Tucker, *Estimation of non-unique quantiles*, Ann. Math. Statist., 37 (1966), pp. 451–457, https://doi.org/10.1214/aoms/1177699527.

[16] N. Frikha, *Multi-level stochastic approximation algorithms*, Ann. Appl. Probab., 26 (2016), pp. 933–985, https://doi.org/10.1214/15-AAP1109.

[17] M. B. Giles, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617, https://doi.org/10.1287/opre.1070.0496.

[18] M. B. Giles, *Multilevel Monte Carlo methods*, Acta Numer., 24 (2015), pp. 259–328, https://doi.org/10.1017/S096249291500001X.

[19] M. B. Giles and A.-L. Haji-Ali, *Multilevel Nested Simulation for Efficient Risk Estimation*, Preprint, arXiv:1802.05016 [Q-Fin.CP], 2018, https://arxiv.org/abs/1802.05016.

[20] M. B. Giles, T. Nagapetyan, and K. Ritter, *Multilevel Monte Carlo approximation of distribution functions and densities*, SIAM/ASA J. Uncertain. Quantif., 3 (2015), pp. 267–295, https://doi.org/10.1137/140960086.

[21] M. B. Giles, T. Nagapetyan, and K. Ritter, *Adaptive Multilevel Monte Carlo Approximation of Distribution Functions*, Preprint, arXiv:1706.06869 [math.PR], 2017, https://arxiv.org/abs/1706.06869.

[22] N. K. Govil and R. N. Mohapatra, *Markov and Bernstein type inequalities for polynomials*, J. Inequal. Appl., 3 (1999), pp. 349–387, https://doi.org/10.1155/S1025583499000247.

[23] C. A. Hall and W. W. Meyer, *Optimal error bounds for cubic spline interpolation*, J. Approx. Theory, 16 (1976), pp. 105–122, https://doi.org/10.1016/0021-9045(76)90040-X.
[24] S. Heinrich, *Monte Carlo complexity of global solution of integral equations*, J. Complexity, 14 (1998), pp. 151–175, https://doi.org/10.1006/jcom.1998.0471.
[25] S. Heinrich, *Multilevel Monte Carlo methods*, in Large-Scale Scientific Computing, Springer, Berlin, Heidelberg, 2001, pp. 58–67, https://doi.org/10.1007/3-540-45346-6_5.
[26] P. Holgate, *The lognormal characteristic function*, Comm. Statist. Theory Methods, 18 (1989), pp. 4539–4548, https://doi.org/10.1080/03610928908830173.
[27] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Applications of Mathematics 23, Springer, New York, 1992, https://doi.org/10.1007/978-3-662-12616-5.
[28] M. Ledoux and M. Talagrand, *Probability in Banach Spaces*, Springer, New York, 1991, https://doi.org/10.1007/978-3-642-20212-4.
[29] V. Linde and A. Pietsch, *Mappings of Gaussian cylindrical measures in Banach spaces*, Theory Probab. Appl., 19 (1974), pp. 445–460, https://doi.org/10.1137/1119054.
[30] M. Pisaroni, S. Krumscheid, and F. Nobile, *Quantifying Uncertain System Outputs via the Multilevel Monte Carlo Method - Part* I*: Central Moment Estimation*, MATHICSE Technical report 23.2017, École Polytechnique Fédérale de Lausanne, Switzerland, 2017.
[31] M. Pisaroni, S. Krumscheid, and F. Nobile, *Quantifying uncertain system outputs via the multilevel Monte Carlo method - Part* II*: Distribution and robustness measures*, manuscript, 2018.
[32] R. T. Rockafellar and S. Uryasev, *Conditional value-at-risk for general loss distributions*, J. Bank. Financ., 26 (2002), pp. 1443–1471, https://doi.org/10.1016/S0378-4266(02)00271-6.
[33] A. L. Teckentrup, R. Scheichl, M. B. Giles, and E. Ullmann, *Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients*, Numer. Math., 125 (2013), pp. 569–600, https://doi.org/10.1007/s00211-013-0546-4.
[34] L. N. Trefethen, *Approximation Theory and Approximation Practice*, Society for Industrial and Applied Mathematics, Philadelphia, 2013.
[35] E. Ullmann and I. Papaioannou, *Multilevel estimation of rare events*, SIAM/ASA J. Uncertain. Quantif., 3 (2015), pp. 922–953, https://doi.org/10.1137/140992953.
[36] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, Cambridge, UK, 2000.