ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

MATHICSE

Institute of Mathematics

School of Basic Sciences

MATHICSE Technical Report

Nr. 12.2017

June 2017

Multilevel Monte Carlo approximation of functions

Sebastian Krumscheid, Fabio Nobile

MULTILEVEL MONTE CARLO APPROXIMATION OF FUNCTIONS

S. KRUMSCHEID AND F. NOBILE

Abstract. Many applications across sciences and technologies require a careful quantification of non-deterministic 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 yet efficient 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 or 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 to construct multilevel Monte Carlo estimators for various robustness indicators, such as for quantiles (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 easily accessible usually. In fact, here we provide a framework that allows 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.

Keywords. multilevel Monte Carlo, parametric expectation, distribution function, quantiles, conditional value-at-risk, characteristic function, moments
AMS subject classifications. 65C05, 60H35, 65C20, 65Y20, 60E10

1. Introduction

Parametric expectations, such as $\Phi(\vartheta) = E\big(\phi(\vartheta, Q)\big)$ 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, say. Here, perhaps most notably are applications in which the cumulative distribution function $F_Q(\vartheta) = \mathbb{E}\big(I(Q \leq \vartheta)\big)$ 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

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

rely heavily on accurate and computationally affordable approximations of these parametric expectations. Consequently, an efficient approximation of parametric expectations, such as $\Phi(\vartheta) = E\big(\phi(\vartheta, Q)\big)$ 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 [7, 13, 14, 25] 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 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 [12]. In this work, we present a multilevel Monte Carlo methodology that allows a uniform approximation of general parametric expectations, in

Date: June 20, 2017.

other words of a function. Moreover, we carefully analyze the proposed methodology and provide a full complexity analysis.

Somewhat related topics have been addressed in [20]. Two of the main differences compared to the present work are the following. Firstly, the aforementioned work considers the case where exact sampling from the law of Q is possible and, secondly, there 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 instead. In fact, the present work is somewhat closer to the work presented in [15], where the authors discuss multilevel Monte Carlo ideas for the uniform approximation of a random variable's cumulative distribution function (CDF). Here, we build upon ideas presented in that work, but extend and generalize these further to approximate general parametric expectations. 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 value-at-risk) or for the conditional value-at-risk. It is noteworthy that these quantities cannot be expressed as moments. Consequently, they had 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 [9, 12]. 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. Moreover, to the best of our knowledge the conditional value-at-risk has still been inaccessible for an efficient treatment using multilevel Monte Carlo techniques so far.

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 the following Sections 3–4 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 $\ell = 0, 1, \ldots$, 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 $h_{\ell-1} = sh_\ell$ 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 $\bar{\Phi}_L^N\colon \mathbb{R}^n \to \mathbb{R}^n$ and is given by

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

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

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

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*

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

(ii) $\|\mathcal{I}_n(x)\|_{L^\infty(\Theta)} \le c_2 \|x\|_{\ell^\infty}$ *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*

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

(ii) $\|\mathcal{I}_n(x)\|_{L^\infty(\Theta)} \le c_2 \ln(n) \|x\|_{\ell^\infty}$ *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 $c_3 n \log(n)$,

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 [8]. 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 $\rho > 1$ [26].

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 so that $\xi \in L^2_{\mathbb{P}}(\Omega;\ell^\infty)$, the variance is defined as

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

It follows that

$$\mathbb{E}\big(\|x-\xi\|^2_{\ell^\infty}\big) \le 2\|x-\mathbb{E}(\xi)\|^2_{\ell^\infty} + 2\,\mathrm{Var}\,(\xi)\ , \tag{1}$$

for any deterministic $x \in \mathbb{R}^n$ and random variable $\xi \in \mathbb{R}^n$. Moreover, it holds that $\mathrm{Var}\,(\xi) \le 4\mathbb{E}\big(\|\xi\|^2_{\ell^\infty}\big)$. 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

$$\mathrm{Var}\,\Big(\sum_{i=1}^N \xi_i\Big) \le c\ln(n)\sum_{i=1}^N \mathrm{Var}\,(\xi_i)\ , \tag{2}$$

where $c>0$ is a generic constant [23]. See also [19, Lemma 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 $(Q_\ell, Q_{\ell-1})$, $\ell\in\mathbb{N}_0$. 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) $\sup_{\vartheta\in\Theta}\big|\mathbb{E}\big(\phi(\vartheta,Q)-\phi(\vartheta,Q_\ell)\big)\big| \le c_1 {h_l}^\alpha$,

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

(iii) $\mathfrak{c}_l \le c_3 {h_l}^{-\gamma}$,

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 exists 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

$$\mathbb{E}\Big(\big\|\hat\Phi_L^{N,n}-\Phi\big\|^2_{L^\infty(\Theta)}\Big) = \mathcal{O}(\varepsilon^2)\ . \tag{3}$$

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

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

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 (3). 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

$$\begin{aligned}\mathrm{MSE}\left(\hat{\Phi}_L^{N,n}\right) &\equiv \mathbb{E}\Big(\big\|\Phi - \mathcal{I}_n\big(\Phi(\theta)\big) + \mathcal{I}_n\big(\Phi(\theta) - \bar{\Phi}_L^N(\theta)\big)\big\|_{L^\infty(\Theta)}^2\Big) \\ &\leq 2\bigg(\big\|\Phi - \mathcal{I}_n\big(\Phi(\theta)\big)\big\|_{L^\infty(\Theta)}^2 + \mathbb{E}\Big(\big\|\mathcal{I}_n\big(\Phi(\theta) - \bar{\Phi}_L^N(\theta)\big)\big\|_{L^\infty(\Theta)}^2\Big)\bigg) .\end{aligned} \tag{4}$$

To bound the right-hand side of (4) 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

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

where we have used property (1). 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), we find that the bias is bounded by

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

Similarly, for the variance term we find that

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

where we have used property (2), the fact that $\mathrm{Var}\left(\xi\right) \leq 4\mathbb{E}\big(\|\xi\|_{\ell^\infty}^2\big)$ for any $\xi \in L^2_{\mathbb{P}}(\Omega; \ell^\infty)$, and hypothesis (ii). Combining these bounds, we eventually obtain

$$\mathrm{MSE}\left(\hat{\Phi}_L^{N,n}\right) \lesssim n^{-2(k+1)} + s^{-2\alpha L} + \ln(n)\sum_{\ell=0}^{L}\frac{s^{-\beta\ell}}{N_\ell} . \tag{5}$$

In view of this bound, we choose

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

to obtain a contribution of $\mathcal{O}(\varepsilon^2)$ originating from the first and the second term in (5), respectively. Notice that $n \geq 2$ and $L \geq 1$, since $\varepsilon < 1$. Moreover, it holds that

$$\varepsilon^{-\frac{\nu}{\alpha}} \leq s^{\nu L} < s^\nu \varepsilon^{-\frac{\nu}{\alpha}} , \tag{6}$$

for any $\nu > 0$. 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

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

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, subject to the constraint $\sum_{\ell=0}^{L} \frac{s^{-\beta\ell}}{N_\ell} = \varepsilon^2 \ln(n)^{-1}$, eventually implies

$$N_l = \left\lceil \varepsilon^{-2} \ln(n) \sqrt{\frac{s^{-l\beta}}{s^{l\gamma}+n}} \sum_{\ell=0}^{L} \sqrt{s^{-\ell\beta}\big(s^{\ell\gamma}+n\big)} \right\rceil \ , \quad 0 \le l \le L \ . \tag{7}$$

Consequently, the mean squared error is $\mathcal{O}(\varepsilon^2)$ as asserted. Bounding the number of samples per level by $N_\ell \le 1 + \varepsilon^{-2}\ln(n)\sqrt{s^{-l\beta}(s^{l\gamma}+n)^{-1}} \sum_{\ell=0}^{L} \sqrt{s^{-\ell\beta}\big(s^{\ell\gamma}+n\big)}$, which is a consequence of (7), 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 property (6). To quantify the sum, we introduce $L^* = \big\lceil \gamma^{-1}\log_s(n)\big\rceil - 1 \ge 0$, which implies that $s^{\ell\gamma} < n$ for all $\ell \le L^*$. Now, we distinguish two cases. Firstly, we consider $L \le L^*$, which, in the absence of rounding errors, implies $\alpha > \gamma(k+1)$. In that case it holds that

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

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

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

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

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

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 (4), the mean squared error can now be bounded by

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

where we have used the same steps that led to (5) before. Now, choosing

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

and then minimizing the cost bound

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

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

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

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

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

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

Remark 2.2. The condition $2\alpha \ge \min(\beta, \gamma)$ in the statement of Thm. 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 above. 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 an $\mathcal{O}(\varepsilon^{-2})$ term in absence of the $2\alpha \ge \min(\beta, \gamma)$ condition.

2.1. **A more refined complexity result.** The proof of Theorem 2.1 above reveals that the contributions from $(\mathfrak{c}_\ell + n)$ to the overall computational cost are essentially constant on every level ℓ for which $\ell \le L^*$, because the cost is dominated by the cost for evaluating ϕ there. That is, the cost of generating samples on these levels is negligible compared to n. Consequently, there is no complexity benefit in using the contributions from the levels $\ell \le L^*$. In fact, by not considering these coarse levels within the construction of the multilevel Monte Carlo estimator at all, we will be able to improve upon the complexity result presented in Theorem 2.1. That is, we will consider an estimator with a possibly positive lower level bound $\underline{L} \in \mathbb{N}_0$ and, as before, an upper level bound that is denoted by $\overline{L} \ge \underline{L}$. The refined multilevel estimator then reads

$$\hat{\Phi}_{\underline{L},\overline{L}}^{N,n} := \mathcal{I}_n\big(\bar{\Phi}_{\underline{L},\overline{L}}^{N}(\theta)\big) \, , \quad \bar{\Phi}_{\underline{L},\overline{L}}^{N}(\theta) := \left[\sum_{\ell=\underline{L}}^{\overline{L}} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \phi\Big(\theta_j, Q_\ell^{(i,\ell)}\Big) - \phi\Big(\theta_j, Q_{\ell-1}^{(i,\ell)}\Big) \right]_{1 \le j \le n} , \tag{8}$$

with $\phi(\cdot, Q_{\underline{L}-1}) \equiv 0$. Here, the number of levels is $L+1$, with $L := \overline{L} - \underline{L} \ge 0$. Notice that this approach can still degenerate to a single-level (i.e. classic) Monte Carlo method, namely if $L = 0$. Intuitively this will only happen, if either a very crude tolerance demand is required (e.g. ε being of order one), or if the generation of highly accurate samples is very cheap (e.g. when exact sampling is possible); see also Remark 2.3 below. In all other cases, in particular asymptotically as $\varepsilon \to 0$, the estimator will have a full multilevel structure. Its computational complexity is summarized in the following result.

Theorem 2.2. *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) $\sup_{\vartheta\in\Theta}\big|\mathbb{E}\big(\phi(\vartheta, Q) - \phi(\vartheta, Q_\ell)\big)\big| \le c_1 h_l{}^{\alpha}$,

(ii) $\mathbb{E}\Big(\sup_{\vartheta\in\Theta}\big|\phi(\vartheta, Q_\ell) - \phi(\vartheta, Q_{\ell-1})\big|^2\Big) \le c_2 h_l{}^{\beta}$, $\quad \mathbb{E}\Big(\sup_{\vartheta\in\Theta}\big|\phi(\vartheta, Q_\ell)\big|^2\Big) \le \bar{c}_2 < \infty$,

(iii) $\mathfrak{c}_l \le c_3 h_l{}^{-\gamma}$,

for all $\ell \in \mathbb{N}_0$ with positive constants $c_1, c_2, \bar{c}_2, c_3$ 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 $\underline{L}, \overline{L} \in \mathbb{N}_0$ with $L := \overline{L} - \underline{L} \geq 0$, $n \in \mathbb{N}$, and $N \in \mathbb{N}^{L+1}$ such that the corresponding multilevel Monte-Carlo estimator $\hat{\Phi}^{N,n}_{\underline{L},\overline{L}}$ satisfies

$$\mathbb{E}\Big(\big\|\hat{\Phi}^{N,n}_{\underline{L},\overline{L}} - \Phi\big\|^2_{L^\infty(\Theta)}\Big) = \mathcal{O}(\varepsilon^2)\,. \tag{9}$$

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

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

if $L > 0$, and by $\mathfrak{C}\big(\hat{\Phi}^{N,n}_{\underline{L},\underline{L}}\big) \lesssim \ln\big(\varepsilon^{-1}\big)\varepsilon^{-\left(2+\frac{\gamma}{\alpha}\right)}$ if $L = 0$.

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 $\hat{\Phi}^{N,n}_{\underline{L},\overline{L}}$ satisfies (9). Moreover, for any $0 < \varepsilon < \min(\rho^{-1}, \ln(2))$ the required computational cost is bounded by

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

if $L > 0$, and by $\mathfrak{C}\big(\hat{\Phi}^{N,n}_{\underline{L},\underline{L}}\big) \lesssim \ln\big(\varepsilon^{-1}\big)^{3+\frac{\gamma}{\alpha}}\varepsilon^{-\left(2+\frac{\gamma}{\alpha}\right)}$ if $L = 0$.

Proof. The proof is very similar to the proof of Thm. 2.1 and we only discuss the differences. Here, we use

$$n = \Big\lceil \varepsilon^{-\frac{1}{k+1}} \Big\rceil \quad \text{and} \quad \overline{L} = \big\lceil \alpha^{-1}\log_s\big(\varepsilon^{-1}\big) \big\rceil\,,$$

as before and set $\underline{L} = \min(L^*, \overline{L})$ with $L^* = \big\lceil \gamma^{-1}\log_s(n) \big\rceil - 1$. Selecting

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

results in the mean squared error of order $\mathcal{O}(\varepsilon^2)$. Recall that the computational cost is bounded by $\mathfrak{C}\big(\hat{\Phi}^{N,n}_{\underline{L},\overline{L}}\big) \lesssim n + \sum_{\ell=\underline{L}}^{\overline{L}} N_\ell\big(s^{\ell\gamma}+n\big)$, from which the claims then follow. In fact, for $L = 0$ (i.e. $\underline{L} = \overline{L}$) the assertion is obvious and for $L > 0$ (i.e. $\underline{L} = L^* < \overline{L}$) we note that

$$\mathfrak{C}\big(\hat{\Phi}^{N,n}_{\underline{L},\overline{L}}\big) \lesssim (L+2)n + \sum_{\ell=\underline{L}}^{\underline{L}+L} s^{\ell\gamma} + \varepsilon^{-2}\ln(n)\left(\sum_{\ell=\underline{L}}^{\underline{L}+L} s^{\ell\frac{\gamma-\beta}{2}}\right)^2\,.$$

The argument for the case of Φ being analytic is analogous. □

Remark 2.3. An explicit identification of the case $L = 0$ (i.e. of the case when the multilevel estimator degenerates to a classic Monte Carlo estimator) in terms of the rates α, β, and γ is not straightforward. This is mainly due to the dependence on several (unknown) constants as well as to integer rounding in practice. However, formal calculations provide some insight. In fact, treating the variables as continuous reveals the necessary condition $\gamma(k+1) \leq \alpha$ for having $L = 0$ (i.e. a single level method) in the case of $\Phi \in C^{k+1}(\Theta)$. This agrees well with our intuition that a single-level Monte Carlo methodology will be sufficient, whenever highly accurate samples are very cheap to obtain (e.g. exact sampling). Moreover, notice that the condition $\gamma(k+1) \leq \alpha$ also implies that the complexity result in Thm. 2.2 is superior to the one in Thm. 2.1 even for $L = 0$.

Due to the complexity improvement of the multilevel Monte Carlo estimator identified in the refined complexity Thm. 2.2 compared to the one of Thm. 2.1, we will only use the improved estimator (8) in what follows. It is moreover noteworthy that the results above illustrate that the multilevel Monte Carlo estimator offers a worse complexity for the case of an analytic function Φ, than for $\Phi \in C^{k+1}(\Theta)$. 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 below, where we will derive multilevel Monte Carlo approximations for derivatives of Φ. Finally, we mention that L^* is typically close or equal to the coarsest level considered for the application at hand. If, however, the sampling is indeed significantly cheaper than the evaluation of ϕ (e.g. if exact sampling is possible or if the evaluation of ϕ itself requires some additional numerical simulations), then one should telescope not only on the number of samples per level, but also on the number of nodes used per level. Such a construction would thus result in a multi-index Monte Carlo estimator [17], an approach that we will leave for future work.

2.2. **Approximation of derivatives.** In view of the multilevel estimator construction (8), one may think that an alternative extension of the pointwise estimates to a function is possible, provided the samples on all levels are saved. That is, it may be tempting to simply consider the mapping $\Theta \ni \vartheta \mapsto \bar{\Phi}^N_{\underline{L},\overline{L}}(\vartheta)$ instead of the function $\hat{\Phi}^{N,n}_{\underline{L},\overline{L}} = \mathcal{I}_n\big(\bar{\Phi}^N_{\underline{L},\overline{L}}(\theta)\big)$. However, this naive approach is not guaranteed to be accurate between the nodes θ, so that uniform error criterion may not be met. Moreover, considering an appropriate extension operator $\mathcal{I}_n$ is crucial when also approximations to derivatives of Φ are 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 $\hat{\Phi}^{N,n}_{\underline{L},\overline{L}}$ via an appropriate extension operator $\mathcal{I}_n$ 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 $\Phi(\vartheta) = \mathbb{E}\big(I(Q \leq \vartheta)\big)$ for example. Any finite sample size (single- or multilevel) Monte Carlo approximation 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 operator $\mathcal{I}_n$, 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*

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

(v) $\|\frac{d^m}{d\vartheta^m}\mathcal{I}_n(x)\|_{L^\infty(\Theta)} \leq c_5 n^m \|\mathcal{I}_n(x)\|_{L^\infty(\Theta)}$ *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) $\|f^{(m)} - \frac{d^m}{d\vartheta^m}\mathcal{I}_n\big(f(\theta)\big)\|_{L^\infty(\Theta)} \leq c_4 \rho^{-n}$,

(v) $\|\frac{d^m}{d\vartheta^m}\mathcal{I}_n(x)\|_{L^\infty(\Theta)} \leq c_5 n^{2m} \|\mathcal{I}_n(x)\|_{L^\infty(\Theta)}$ *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) [18]. As for Assumption A2, Assumption A4 is satisfied for polynomial interpolation on Chebyshev nodes as a result of the spectral convergence [26, Ch. 21] and the Markov brothers' inequality [16].

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 based on identical hypotheses on the function ϕ as is Thm. 2.2 for the approximation of Φ.

Theorem 2.3. *Let $\Phi(\vartheta) = \mathbb{E}\big(\phi(\vartheta, Q)\big)$ and $m \in \mathbb{N}$. Suppose that the hypotheses (i) – (iii) of Thm. 2.2 are satisfied for ϕ.*

If $\Phi \in C^{k+1}(\Theta)$ for some $k \in \mathbb{N}_0$ such that $m \leq k$ and if assumption A3 is satisfied for that k, then for any $0 < \varepsilon$ there exist parameters $\underline{L}, \overline{L} \in \mathbb{N}_0$ with $L := \overline{L} - \underline{L} \geq 0$, $n \in \mathbb{N}$ and $N \in \mathbb{N}^{L+1}$ such that the m-th derivative of corresponding multilevel Monte Carlo estimator $\hat{\Phi}^{N,n}_{\underline{L},\overline{L}}$ satisfies

$$\mathbb{E}\left(\left\|\Phi^{(m)} - \frac{d^m}{d\vartheta^m}\hat{\Phi}^{N,n}_{\underline{L},\overline{L}}\right\|^2_{L^\infty(\Theta)}\right) = \mathcal{O}(\varepsilon^2)\,. \tag{10}$$

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

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

if $L > 0$, and by $\mathfrak{C}\big(\frac{d^m}{d\vartheta^m}\hat{\Phi}^{N,n}_{\underline{L},\underline{L}}\big) \lesssim \varepsilon^{-\left(2\frac{k+1}{k+1-m}+\max\left(\frac{\gamma}{\alpha},\frac{1}{k+1-m}\right)\right)} \ln\big(\varepsilon^{-1}\big)$ if $L = 0$.

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 such that the m-th derivative of the corresponding multilevel Monte Carlo estimator $\hat{\Phi}^{N,n}_{\underline{L},\overline{L}}$ satisfies (10)*. Moreover, for any $0 < \varepsilon < \min(\rho^{-1}, \ln(2))$ the required computational cost is bounded by*

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

if $L = 0$, and by $\mathfrak{C}\big(\frac{d^m}{d\vartheta^m}\hat{\Phi}^{N,n}_{\underline{L},\underline{L}}\big) \lesssim \varepsilon^{-\left(2+\frac{\gamma}{\alpha}\right)}\ln\big(\varepsilon^{-1}\big)^{3+\frac{\gamma}{\alpha}+2m\left(2+\frac{\gamma}{\alpha}\right)}$ if $L = 0$.

Proof. The mean squared error can be decomposed into

$$\begin{aligned} \mathrm{MSE}\left(\frac{d^m}{d\vartheta^m}\hat{\Phi}^{N,n}_{\underline{L},\overline{L}}\right) &\equiv \mathbb{E}\left(\left\|\Phi^{(m)} - \frac{d^m}{d\vartheta^m}\mathcal{I}_n\big(\bar{\Phi}^N_{\underline{L},\overline{L}}(\theta)\big)\right\|^2_{L^\infty(\Theta)}\right) \\ &\leq 2\left\|\Phi^{(m)} - \frac{d^m}{d\vartheta^m}\mathcal{I}_n\big(\Phi(\theta)\big)\right\|^2_{L^\infty(\Theta)} + 2\mathbb{E}\left(\left\|\frac{d^m}{d\vartheta^m}\mathcal{I}_n\big(\Phi(\theta) - \bar{\Phi}^N_{\underline{L},\overline{L}}(\theta)\big)\right\|^2_{L^\infty(\Theta)}\right). \end{aligned}$$

Next, we proceed as in the proof of Thm. 2.2. For $\Phi \in C^{k+1}(\Theta)$ and assumption A3 being satisfied, we find that

$$\mathrm{MSE}\left(\frac{d^m}{d\vartheta^m}\hat{\Phi}^{N,n}_{\underline{L},\overline{L}}\right) \lesssim n^{-2(k+1-m)} + n^{2m}s^{-2\alpha\overline{L}} + n^{2m}\ln(n)\sum_{\ell=\underline{L}}^{\overline{L}}\frac{s^{-\beta\ell}}{N_\ell}\,. \tag{11}$$

Therefore, we chose

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

and, $\underline{L} = \min(L^*, \overline{L})$, with $L^* = \big\lceil \gamma^{-1} \log_s(n) \big\rceil - 1$. Next, minimizing the cost bound of the corresponding function approximation, that is

$$\mathfrak{C}\left(\frac{d^m}{d\vartheta^m} \hat{\Phi}^{N,n}_{\underline{L},\overline{L}} \right) \lesssim \sum_{\ell=\underline{L}}^{\overline{L}} N_\ell (\mathfrak{c}_\ell + n) + n \;,$$

subject to the constraint $\sum_{\ell=\underline{L}}^{L} \frac{s^{-\beta\ell}}{N_\ell} = \varepsilon^2 \ln(n)^{-1} n^{-2m}$, yields

$$N_{\underline{L}} = \big\lceil \varepsilon^{-2} \ln(n) n^{2m} \big\rceil \;, \quad N_l = \left\lceil \varepsilon^{-2} \ln(n) n^{2m} \sqrt{\frac{s^{-l\beta}}{s^{l\gamma}+n}} \sum_{\ell=\underline{L}}^{\overline{L}} \sqrt{s^{-\ell\beta}\big(s^{\ell\gamma}+n\big)} \right\rceil \;, \quad \underline{L} < l \leq L \;.$$

These choices eventually result in the asserted mean squared error of $\mathcal{O}(\varepsilon^2)$ and also provide the cost bound

$$\mathfrak{C}\left(\frac{d^m}{d\vartheta^m} \hat{\Phi}^{N,n}_{\underline{L},\overline{L}} \right) \lesssim nL + s^{\gamma \underline{L}} + \varepsilon^{-2} \ln(n) n^{2m} \begin{cases} L^2 \;, & \beta = \gamma \;, \\ \varepsilon^{-\frac{\gamma-\beta}{\alpha}} n^{m\frac{\gamma-\beta}{\alpha}} \;, & \beta < \gamma \;, \\ n^{-\frac{\beta-\gamma}{\gamma}} \;, & \beta > \gamma \;, \end{cases}$$

with $L = \overline{L} - \underline{L}$, from which the claim follows.

The analysis for the case that $\Phi \colon \Theta \to \mathbb{R}$ is analytic and that A4 is valid is very similar and therefore omitted. For the sake of completeness, we merely report the parameter choices: $n = \big\lceil \log_\rho\big(\varepsilon^{-1}\big) \big\rceil$, $\overline{L} = \big\lceil \alpha^{-1} \log_s\big(\varepsilon^{-1} \ln(n) n^{2m}\big) \big\rceil$, as well as

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

□

Notice that Thm. 2.3 above is stated under weak assumptions on the function ϕ. Specifically, we only assume that hypotheses (i) – (iii) of Thm. 2.2 are satisfied for ϕ, but we do not make such assumptions for the derivatives of ϕ.

Remark 2.5. As mentioned in Remark 2.3 already, an explicit identification of the case $L = 0$ in terms of the rates α, β, and γ is not straightforward. Here, a formal calculation provides the necessary condition $\gamma(k+1-m) \leq \alpha - m$ for having $L = 0$ (i.e. a single level method) in the case of $\Phi \in C^{k+1}(\Theta)$. Notice that this necessary condition also implies that $\frac{\gamma}{\alpha} \leq \frac{1}{k+1-m}$, so that the computational complexity for the single level estimator simplifies to $\mathfrak{C}\big(\frac{d^m}{d\vartheta^m} \hat{\Phi}^{N,n}_{\underline{L},\overline{L}}\big) \lesssim \varepsilon^{-\left(2\frac{k+1}{k+1-m} + \frac{\gamma}{\alpha}\right)} \ln\big(\varepsilon^{-1}\big)$ in the absence of integer rounding.

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 cumulative distribution function, which may be cumbersome, for example due to 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 set $\Theta \subset \mathbb{R}$ is therefore to use the results from Sect. 2 by simultaneously constructing multilevel Monte Carlo approximations to both Φ_1 and Φ_2. That is, we consider an approximation of the form

$$\hat{\varphi}^{N,n}_{\underline{L},\overline{L}} := \mathcal{I}_n\big(\bar{\varphi}^N_{\underline{L},\overline{L}}(\theta)\big) \,, \quad \bar{\varphi}^N_{\underline{L},\overline{L}}(\theta) = \bar{\Phi}^N_{1|\underline{L},\overline{L}}(\theta) + \mathrm{i}\bar{\Phi}^N_{2|\underline{L},\overline{L}}(\theta) \,,$$

based on extending the collection of pointwise estimates

$$\bar{\Phi}^N_{r|\underline{L},\overline{L}}(\theta) := \left[\sum_{\ell=\underline{L}}^{\overline{L}} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \phi_r\Big(\theta_j, Q_\ell^{(i,\ell)}\Big) - \phi_r\Big(\theta_j, Q_{\ell-1}^{(i,\ell)}\Big)\right]_{1\le j\le n} .$$

We recall that $\phi_r(\vartheta, Q_{\underline{L}-1}) \equiv 0$ for notational convenience. Alternatively,one could, of course, directly approximate $\vartheta \mapsto \mathbb{E}\big(\exp(\mathrm{i}\vartheta Q)\big)$. This would, however, require to extend the framework presented in Sect. 2 to complex-valued functions. It is moreover noteworthy that the collection of pointwise estimators $\bar{\Phi}^N_{r|\underline{L},\overline{L}}(\theta)$ defined above is based on the same samples for both values of $r \in \{1,2\}$. 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.

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

The following result, which is essentially a spacial case of Thm. 2.2, then characterizes the computational complexity of the multilevel Monte Carlo approximation to φ_Q.

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 Thm. 2.2 are satisfied for both ϕ_1 and ϕ_2.*

If $\Phi_1, \Phi_2 \in C^{k+1}(\Theta)$ for some $k \in \mathbb{N}_0$ and if assumption A1 is satisfied for that k, then for any $\varepsilon > 0$ there exist parameters $\underline{L}, \overline{L} \in \mathbb{N}_0$ with $L := \overline{L} - \underline{L} \ge 0$, $n \in \mathbb{N}$, and $N \in \mathbb{N}^{L+1}$ such that the corresponding multilevel Monte Carlo estimator $\hat{\varphi}^{N,n}_{\underline{L},\overline{L}}$ satisfies

$$\mathbb{E}\Big(\big\|\hat{\varphi}^{N,n}_{\underline{L},\overline{L}} - \varphi_Q\big\|^2_{L^\infty(\Theta,\mathbb{C})}\Big) = \mathcal{O}(\varepsilon^2) \,. \tag{12}$$

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

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

if $L > 0$, and by $\mathfrak{C}\big(\hat{\varphi}^{N,n}_{\underline{L},\underline{L}}\big) \lesssim \ln\big(\varepsilon^{-1}\big)\varepsilon^{-\left(2+\frac{\gamma}{\alpha}\right)}$ if $L = 0$.

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 such that the corresponding estimator $\hat{\varphi}^{N,n}_{\underline{L},\overline{L}}$ satisfies (12). Moreover, for any $0 < \varepsilon < \min(\rho^{-1}, \ln(2))$ the required computational cost is bounded by

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

if $L > 0$, and by $\mathfrak{C}\big(\hat{\varphi}^{N,n}_{\underline{L},\underline{L}}\big) \lesssim \ln\big(\varepsilon^{-1}\big)^{3+\frac{\gamma}{\alpha}}\varepsilon^{-\left(2+\frac{\gamma}{\alpha}\right)}$ if $L = 0$.

Proof. The mean squared error can be decomposed as

$$\mathbb{E}\Big(\big\|\hat{\varphi}^{N,n}_{\underline{L},\overline{L}} - \varphi_Q\big\|^2_{L^\infty(\Theta,\mathbb{C})}\Big) \leq \mathbb{E}\Big(\big\|\Phi_1 - \mathcal{I}_n\big(\bar{\Phi}^N_{1|\underline{L},\overline{L}}(\theta)\big)\big\|^2_{L^\infty(\Theta)}\Big) + \mathbb{E}\Big(\big\|\Phi_2 - \mathcal{I}_n\big(\bar{\Phi}^N_{2|\underline{L},\overline{L}}(\theta)\big)\big\|^2_{L^\infty(\Theta)}\Big)\,,$$

due to the linearity of the extension operator. 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 then follow directly from Thm. 2.2. □

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 M-th order, then $\varphi_Q \in C^M(\mathbb{R})$ and it holds that

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

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 post-processing 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 Thm. 2.2 are satisfied for both ϕ_1 and ϕ_2.*

If $\Phi_1, \Phi_2 \in C^{k+1}(\Theta)$ for some $k \in \mathbb{N}_0$ such that $M \leq k$ and if assumption A3 is satisfied for that k, then for any $0 < \varepsilon$ there exist parameters $\underline{L}, \overline{L} \in \mathbb{N}_0$ with $L := \overline{L} - \underline{L} \geq 0$, $n \in \mathbb{N}$, and $N \in \mathbb{N}^{L+1}$ such that multilevel Monte Carlo estimator $\hat{\varphi}^{N,n}_{\underline{L},\overline{L}}$ satisfies

$$\max_{0\leq m\leq M} \mathbb{E}\left(\left|\mathbb{E}(Q^m) - (-\mathrm{i})^m \big(\hat{\varphi}^{N,n}_{\underline{L},\overline{L}}\big)^{(m)}(0)\right|^2\right) = \mathcal{O}(\varepsilon^2)\,. \tag{13}$$

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

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

if $L > 0$, and by $\mathfrak{C}_M \lesssim \varepsilon^{-\left(2\frac{k+1}{k+1-M} + \max\left(\frac{\gamma}{\alpha}, \frac{1}{k+1-M}\right)\right)} \ln\big(\varepsilon^{-1}\big)$ if $L = 0$.

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 such that the multilevel Monet Carlo estimator $\hat{\varphi}^{N,n}_{\underline{L},\overline{L}}$ satisfies (13). Moreover, for any $0 < \varepsilon < \min(\rho^{-1}, \ln(2))$ the required computational cost $\mathfrak{C}_M$ is bounded by

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

if $L = 0$, and by $\mathfrak{C}_M \lesssim \varepsilon^{-\left(2+\frac{\gamma}{\alpha}\right)} \ln\big(\varepsilon^{-1}\big)^{3+\frac{\gamma}{\alpha}+2M\left(2+\frac{\gamma}{\alpha}\right)}$ if $L = 0$.

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

$$\mathbb{E}\left(\left|\varphi_Q^{(m)}(0) - \big(\hat{\varphi}_{\underline{L},\overline{L}}^{N,n}\big)^{(m)}(0)\right|^2\right) \leq \mathbb{E}\left(\sup_{\vartheta\in\Theta}\left|\varphi_Q^{(m)}(\vartheta) - \big(\hat{\varphi}_{\underline{L},\overline{L}}^{N,n}\big)^{(m)}(\vartheta)\right|^2\right)$$
$$\leq \sum_{r=1}^{2} \mathbb{E}\left(\left\|\Phi_r^{(m)} - \frac{d^m}{d\vartheta^m}\mathcal{I}_n\Big(\bar{\Phi}^N_{r|\underline{L},\overline{L}}(\theta)\Big)\right\|^2_{L^\infty(\Theta)}\right).$$

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 Thm. 2.3. There it was also shown that these errors increase monotonically as m increases (ignoring constants), see relation (11), so that the claims follow. □

Next, we address the post-processing 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), 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 post-processing approach.

Corollary 3.3. *Let $\hat{\varphi}_{\underline{L},\overline{L}}^{N,n}$ be the approximation of φ_Q with mean squared error $\mathcal{O}(\varepsilon^2)$ (in the uniform norm) that is identified in Cor. 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*

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

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

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

This error can be bounded by

$$\mathrm{MSE}_m \leq 2\sum_{r=1}^{2} \mathbb{E}\left(\left\|\Phi_r^{(m)} - \frac{d^m}{d\vartheta^m}\mathcal{I}_n\big(\Phi_r(\theta)\big)\right\|^2_{L^\infty(\Theta)} + \left\|\frac{d^m}{d\vartheta^m}\mathcal{I}_n\Big(\Phi_r(\theta) - \bar{\Phi}^N_{r|\underline{L},\overline{L}}(\theta)\Big)\right\|^2_{L^\infty(\Theta)}\right).$$

As $\hat{\varphi}_{\underline{L},\overline{L}}^{N,n}$ is the $\mathcal{O}(\varepsilon^2)$ mean squared error approximation of φ_Q identified in Cor. 3.1, we have in particular that $\mathbb{E}\Big(\big\|\mathcal{I}_n\big(\Phi_r(\theta) - \bar{\Phi}^N_{r|\underline{L},\overline{L}}(\theta)\big)\big\|^2_{L^\infty(\Theta)}\Big) = \mathcal{O}(\varepsilon^2)$ for $r \in \{1,2\}$. In view of assumption A3 and assumption A4, respectively, we thus find

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

from which the claim follows. □

Related work on multilevel Monte Carlo estimators for higher order moments has recently been presented in [6]. There, the authors study a multilevel estimator for central moments, in contrast to the novel results for (raw) moments above. Further conceptual differences include the fact that the underlying assumptions for our framework (i.e. hypotheses (i) – (iii) of Thm. 2.2) are independent of M, while the setting of [6] requires uniformity of related assumptions with respect to central moment order smaller or equal to M. This uniformity assumption implies the need for considering the worst case scenarios, which may deteriorate the method's performance. On the other hand, the computational complexity of the moment estimators presented here (Cor. 3.2) depends on M, while the complexity result in the aforementioned work does not. We also note that the simultaneous moment estimators presented here are a byproduct of the

function estimators, while the work in reference [6] focuses on the estimation of central moments. Finally, we reiterate that the assumptions of Cor. 3.2 are weaker than the ones used in [6], as they are independent of M. It is worthwhile pointing out however, that one could, of course, strengthen the assumptions of our framework to be comparable to setting of [6] by imposing corresponding hypothesis on the first M derivatives of ϕ_1 and ϕ_2. As a consequence of Thm. 2.2 we would then also obtain standard multilevel Monte Carlo complexity results (up to logarithmic terms) 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

$$dS = rS\,dt + \sigma S\,dW\ , \quad S(0) = S_0\ , \tag{14}$$

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 $\mathbb{P}(Q = 0) = P\big(S(T) \leq K\big) > 0$, since $S(T)$ is log-normally distributed with mean $S_0 e^{rT}$ and variance ${S_0}^2 e^{2rT}\big(e^{\sigma^2 T} - 1\big)$.

The characteristic function of a log-normally distributed random variable is not analytic in the origin [21]. 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 density of $S(T)$, then the characteristic function of Q can be expressed as

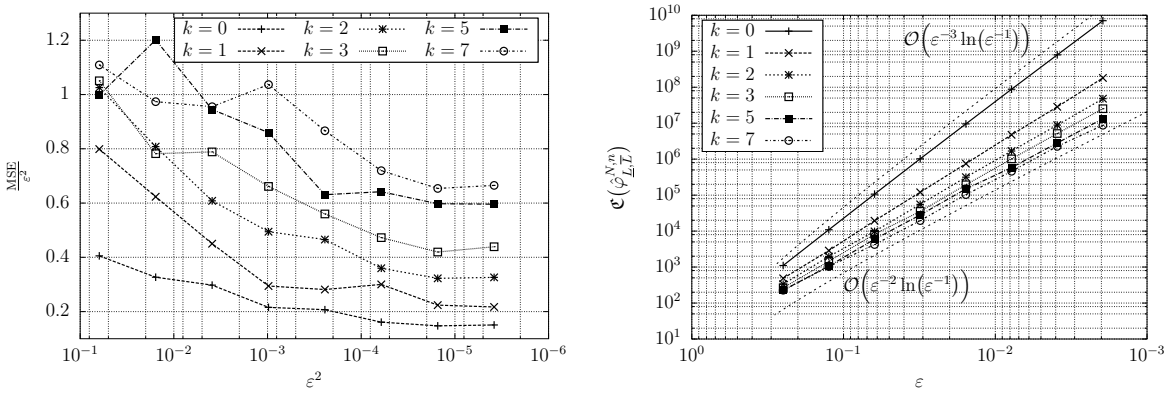
$$\varphi_Q(t) = \frac{1}{2} - \frac{1}{2}\operatorname{erf}\left(\frac{\big(r - \frac{\sigma^2}{2}\big)T + \ln\big(\frac{S_0}{K}\big)}{\sqrt{2T}\sigma}\right) + \int_0^\infty f_{S(T)}\big(K + qe^{rT}\big)e^{t(r+\mathrm{i}q)}\,dq\ ,$$

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 (14) via the Milstein scheme with uniform time step [22, Ch. 10.3], which reads

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

so that $\bar{S}_n \approx S(nh)$ for $h = T/n_T$ and $0 \leq n \leq n_T$. Here, $\big(\xi_n\big)_{n\geq 0}$ denotes a sequence of i.i.d. standard normally distributed random variables. The hierarchy of approximations is constructed based on using a time step $h_\ell = T2^{-\ell}$ 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 $S_0 = 10$. Furthermore, the multilevel Monte Carlo estimator is constructed to approximate the characteristic function φ_Q uniformly on the interval $\Theta = [-1, 1]$. Fig. 1(A) then shows the accuracy of the multilevel Monte Carlo estimator of φ_Q for various values of k. 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, Cor. 3.1 still holds. On the other, a very small value of k results in a poor computational complexity. 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 Fig. 1(A) the



(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.

mean squared error of each estimator (denoted by MSE in the figure and 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, Fig. 1(A) also indicates that the implementation is actually conservative for this example, in the sense that it produces estimates that are more accurate than required, provided the tolerance demand is sufficiently small. Fig. 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 Thm. 2.2 are $\alpha = 1$, $\beta = 2$, and $\gamma = 1$. Based on the formal calculation (i.e. ignoring integer rounding) described in Rem. 2.3, the necessary condition for the multilevel estimator to degenerate to a classic (i.e. single-level) Monte Carlo estimator thus turns out to be $k = 0$ here. This is confirmed by the results shown in Fig. 1(B). In fact, for $k = 0$ we observe a computational complexity of order $\mathcal{O}\big(\varepsilon^{-3}\ln(\varepsilon^{-1})\big)$. Conversely, for $k \geq 1$ the estimator is a proper multilevel estimator and we observe a computational complexity of order $\mathcal{O}\big(\varepsilon^{-2}\ln(\varepsilon^{-1})\big)$, as predicted by Cor. 3.1.

4. Approximation of the cumulative distribution function

One of the most commonly used way to characterize the distribution of a random variable Q is via its cumulative distribution function (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 Sect. 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 rates α and β characterizing the hypotheses (i) and (ii) in Thm. 2.2, respectively, will deteriorate. Furthermore, the discontinuity also means that a numerical estimation of these rates (so-called screening procedure) 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 (i) and (ii) 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 [15]. 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 δ depends on the tolerance requirement ε. However, as the method presented in [15] 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 it is the case for the SDE based examples the authors considered in [15]. 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 regularization parameter turns screening the rates into a prohibitive task, as the screening would have to be carried out for every tolerance demand of interest. 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

$$\Phi' = F_Q \,. \tag{15}$$

One candidate to satisfies this identity is

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

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. Sect. 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 (15), 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 Sect. 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 an approximation of the form (8) for the function Φ such that also Φ' is approximated accurately with respect to the tolerance demand. We then define the CDF estimator via identity (15). Its computational complexity is thus a direct consequence of Thm. 2.3 with $m = 1$.

Corollary 4.1. *Suppose that the hypotheses (i) – (iii) of Thm. 2.2 are satisfied for $\phi(\vartheta, Q) = |Q-\vartheta| I(Q \leq \vartheta)$. If $F_Q \in C^{k+1}(\Theta)$ for some $k \in \mathbb{N}_0$ and if assumption A3 is satisfied for that k, then for any $0 < \varepsilon$ there exist parameters $\underline{L}, \overline{L} \in \mathbb{N}_0$ with $L := \overline{L} - \underline{L} \geq 0$, $n \in \mathbb{N}$, and $N \in \mathbb{N}^{L+1}$ such that the corresponding multilevel Monte Carlo CDF estimator $\hat{F}^{N,n}_{\underline{L},\overline{L}}$ satisfies*

$$\mathbb{E}\left(\left\| F_Q - \hat{F}^{N,n}_{\underline{L},\overline{L}} \right\|^2_{L^\infty(\Theta)}\right) = \mathcal{O}(\varepsilon^2) \,. \tag{16}$$

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

$$\mathfrak{C}\Big(\hat{F}^{N,n}_{\underline{L},\overline{L}}\Big) \lesssim \varepsilon^{-2\frac{k+2}{k+1}} \ln\big(\varepsilon^{-1}\big) \begin{cases} \ln(\varepsilon^{-1})^2 \,, & \text{if } \beta = \gamma \,, \\ \varepsilon^{-\frac{(\gamma-\beta)(k+2)}{\alpha(k+1)}} \,, & \text{if } \beta < \gamma \,, \\ 1 \,, & \text{if } \beta > \gamma \,, \end{cases}$$

if $L > 0$, and by $\mathfrak{C}\big(\hat{F}^{N,n}_{\underline{L},\underline{L}}\big) \lesssim \varepsilon^{-\left(2\frac{k+2}{k+1} + \max\left(\frac{\gamma}{\alpha}, \frac{1}{k+1}\right)\right)} \ln\big(\varepsilon^{-1}\big)$ if $L = 0$.

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 such that the corresponding multilevel Monte Carlo CDF estimator $\hat{F}^{N,n}_{\underline{L},\overline{L}}$ satisfies (16). Moreover, for any $0 < \varepsilon < \min(\rho^{-1}, \ln(2))$ the required computational cost is bounded by

$$\mathfrak{C}\Big(\hat{F}^{N,n}_{\underline{L},\overline{L}}\Big) \lesssim \varepsilon^{-2} \ln\big(\varepsilon^{-1}\big)^7 \begin{cases} \ln(\varepsilon^{-1})^6 \,, & \text{if } \beta = \gamma \,, \\ \varepsilon^{-\frac{\gamma-\beta}{\alpha}} \ln(\varepsilon^{-1})^{3\frac{\gamma-\beta}{\alpha}} \,, & \text{if } \beta < \gamma \,, \\ 1 \,, & \text{if } \beta > \gamma \,, \end{cases}$$

if $L = 0$, and by $\mathfrak{C}\big(\hat{F}^{N,n}_{\underline{L},\underline{L}}\big) \lesssim \varepsilon^{-\left(2+\frac{\gamma}{\alpha}\right)} \ln\big(\varepsilon^{-1}\big)^{3+\frac{\gamma}{\alpha}+2\left(2+\frac{\gamma}{\alpha}\right)}$ if $L = 0$.

It is noteworthy that one can, of course, approximate even higher order derivatives of Φ in the same way. In fact, using Thm. 2.3 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 [15], again making use of an explicit regularization procedure. As mentioned above already, the rate hypothesis in [15] are tolerance dependent, which could make estimating them numerically prohibitive. However, in cases where these hypotheses can be verified theoretically (which is for example the case for the geometric Brownian motion problem considered in Sect. 3.3), we point out that the multilevel procedure introduced in [15] and the one presented here provide the same complexity results for the estimation of both a cumulative distribution function and of a probability density function. Here, we also mention the work in [5], in which 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 $\hat{F}^{N,n}_{\underline{L},\overline{L}}$ constructed above, may not be a monotonically increasing function on Θ. This can, however, simply be achieved by as a post-processing step when replacing the estimated function $\hat{F}^{N,n}_{\underline{L},\overline{L}}$ by the function $\Theta \ni t \mapsto \sup_{\vartheta\in[\min(\Theta),t]} \hat{F}^{N,n}_{\underline{L},\overline{L}}(\vartheta)$, since this replacement does not increase the uniform mean squared error (16).

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 $q_\tau \equiv q$, 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 had been hampered thus far. As mentioned in the Sect. 1 already, first advances have been made in the context of multilevel stochastic approximation algorithms [9, 12]. We also mention the work [10] 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 post-processing step and assess their accuracy.

We do not treat here the problem of non-unique quantile estimation, i.e. the case when the CDF is constant in a region of support. This is a non trivial problem even in the classic Monte Carlo setting; see [11]. 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

$$\hat{F}^{N,n}_{\underline{L},\overline{L}}(\hat{q}) = \tau \ ,$$

with $\hat{F}^{N,n}_{\underline{L},\overline{L}}$ being a multilevel CDF estimator. Notice that the fact that F_Q is (locally) invertible does not imply that the multilevel Monte Carlo approximation $\hat{F}^{N,n}_{\underline{L},\overline{L}}$ 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 does the corresponding CDF estimator.

Proposition 4.1. *Let $\hat{F}^{N,n}_{\underline{L},\overline{L}}$ be the approximation of the CDF $F_Q \in C^1(\Theta)$ with mean squared error $\mathcal{O}(\varepsilon^2)$ that is identified in Cor. 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 $\hat{F}^{N,n}_{\underline{L},\overline{L}}(\hat{q}) = \tau$. 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 $\hat{F} := \hat{F}^{N,n}_{\underline{L},\overline{L}}$. 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

$$|q-\hat{q}| \leq \frac{1}{\inf_{\vartheta\in\Theta} F'_Q(\vartheta)} \sup_{\vartheta\in\Theta} |\hat{F}(\vartheta) - F(\vartheta)| \ , \tag{17}$$

from which the claim follows. □

In view of property (17), 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]. We will leave this aspect for future work. It is noteworthy, however, that this is not just a shortcoming of our techniques, but it is rather a general one for general purpose sampling techniques. In fact, a factor of the form $F'(q)^{-2}$ is also affecting the (asymptotic) mean squared error results for the classic Monte Carlo based quantile estimators; cf. [28].

4.2. **Simultaneous approximation the CDF, quantiles, and the conditional value-at-risk.** While the previous section assesses the accuracy of a multilevel Monte Carlo quantile estimator via post-processing a CDF estimator, here we will describe an approach that enables extracting even more characteristic information about a random variable's distribution via appropriate post-processing 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 (15), we can consider a function $\Psi\colon \Theta \to \mathbb{R}$ such that

$$\Psi' = F_Q - \tau \ , \tag{18}$$

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

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

which also has all the advantageous properties (increased regularity, Lipschitz continuous ψ, etc.) that the function $\vartheta \to \mathbb{E}\big(|Q-\vartheta| I(Q \leq \vartheta)\big)$ offered for the CDF estimation. Furthermore, by construction, this 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 Prop. 4.1 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 will nonetheless enable us to derive an effective estimator, for which identity (18) will be essential. In fact, upon noticing that

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

we find that

$$\Psi(\vartheta) = (1-\tau)\Big(\vartheta + \frac{1}{1-\tau}\mathbb{E}\big((Q-\vartheta)^+\big) - \mathbb{E}(Q)\Big) \equiv (1-\tau)\big(\Phi(\vartheta) - \mathbb{E}(Q)\big) \ , \tag{19}$$

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 [24], which is an important robustness indicator in various applications, ranging from robust design optimization to mathematical finance. By definition, equation (19) shows that

$$\Psi'(\vartheta) = (1-\tau)\Phi'(\vartheta) \,, \tag{20}$$

implying that both Φ and Ψ have the same minimizer, namely the τ-quantile q, say. This motivates 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 post-processing steps. In other words, the following procedure enables a simultaneous approximation of these quantities, all of which characterize the distribution of a random variable Q:

1. Use Thm. 2.3 with $m = 1$ to construct a multilevel Monte Carlo approximation $\hat{\Phi}^{N,n}_{\underline{L},\overline{L}} \equiv \hat{\Phi}$ of Φ that provides a (uniform) mean squared error of order $\mathcal{O}(\varepsilon^2)$ for both the function and its derivative, that is $\mathrm{MSE}\left(\hat{\Phi}\right) + \mathrm{MSE}\left(\hat{\Phi}'\right) = \mathcal{O}(\varepsilon^2)$.
2. Construct a CDF approximation $\hat{F}^{N,n}_{\underline{L},\overline{L}} \equiv \hat{F}$ of F_Q via post-processing. In view of equations (18) and (20) 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 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 Prop. 4.1.
4. An approximation to the conditional value-at-risk can also be obtained via post-processing 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
$$\begin{aligned}\left|\Phi(q) - \hat{\Phi}(\hat{q})\right| &\leq \left|\Phi(q) - \Phi(\hat{q})\right| + \left|\Phi(\hat{q}) - \hat{\Phi}(\hat{q})\right| \\ &\leq \frac{1+\tau}{(1-\tau)\inf_{\vartheta\in\Theta} F_Q'(\vartheta)}\|F_Q - \hat{F}\|_{L^\infty(\Theta)} + \|\Phi - \hat{\Phi}\|_{L^\infty(\Theta)} \\ &= \frac{1+\tau}{\inf_{\vartheta\in\Theta} F_Q'(\vartheta)}\|\Phi' - \hat{\Phi}'\|_{L^\infty(\Theta)} + \|\Phi - \hat{\Phi}\|_{L^\infty(\Theta)} \,,\end{aligned}$$
where we have used property (17) again, as well as the term $(1+\tau)$ as a bound for $\|F_Q - \tau\|_{L^\infty(\Theta)}$.

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

$$-\Delta u = f \,, \quad \text{in } D = (0,1)^2 \,, \tag{21}$$

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

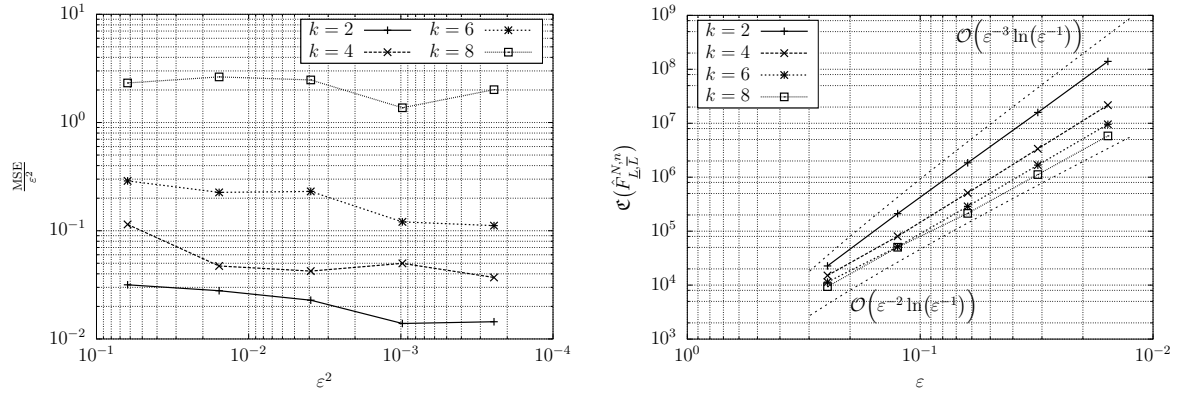
$$f(x) = -72\xi({x_1}^2 + {x_2}^2 - x_1 - x_2) \,,$$

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 (denoted by $\xi \sim \chi^2_1$), so that quantity of interest's CDF is

$$F_Q(\vartheta) = \begin{cases} \mathrm{erf}\left(\sqrt{\frac{\vartheta}{2}}\right) \,, & \vartheta \geq 0 \,, \\ 0 \,, & \text{else,} \end{cases}$$



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

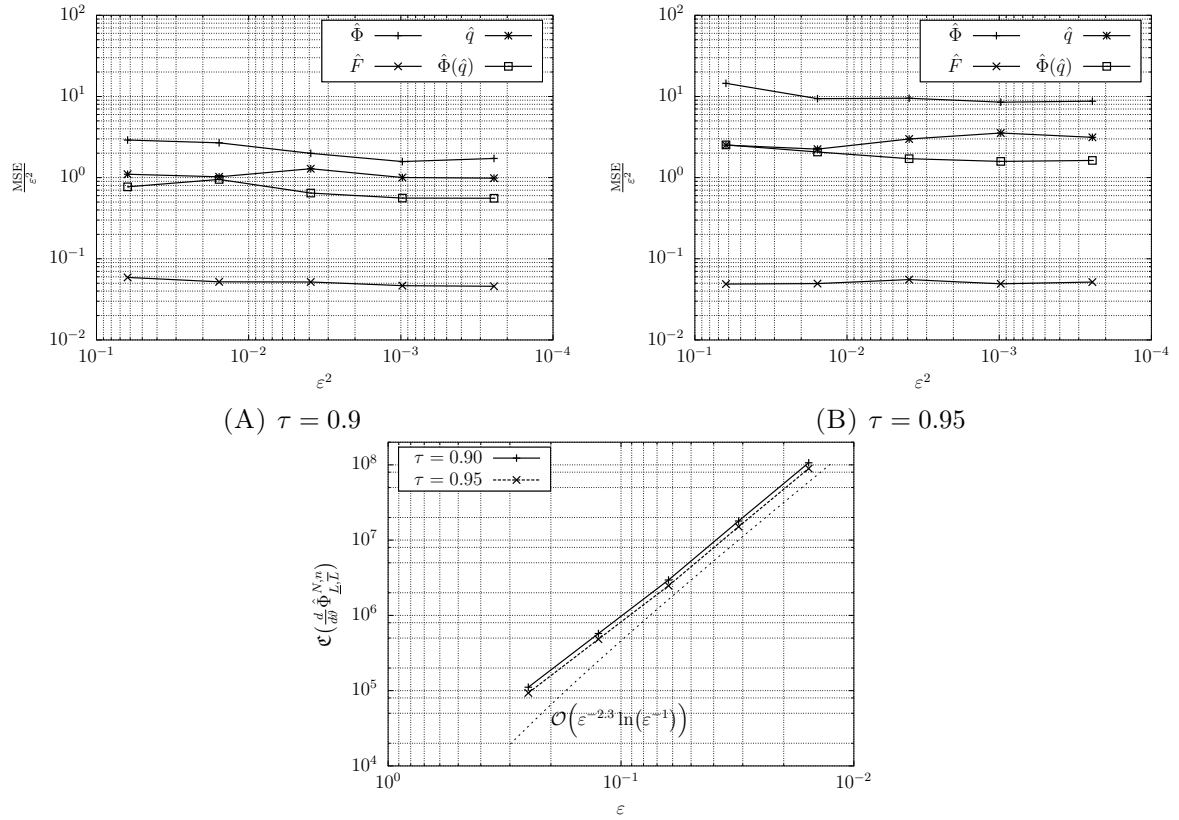
FIGURE 2. Performance of the multilevel Monte Carlo approximation of the CDF of the quantity of interest Q for various values of k.

where $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-s^2}\, ds$ as before. The boundary value problem (21) 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 $\Theta = [0, 10]$. Specifically, Figure 2(A) shows the accuracy of the multilevel Monte Carlo CDF approximation for various values of the regularity parameter k. Here, the mean squared error of the estimator is approximated by repeating 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 various values of k. For this example, we find that the rates characterizing the hypotheses (i) – (iii) of Thm. 2.2 are $\alpha = 2$, $\beta = 4$, and $\gamma \approx 2.4$, so that we expect a computational complexity of $\mathcal{O}\big(\varepsilon^{-2\frac{k+2}{k+1}} \ln(\varepsilon^{-1})\big)$ in view of Cor. 4.1. This complexity behavior is confirmed by the results illustrated in Fig. 2(B).

Next, we consider the procedure described in Sect. 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 this simultaneous multilevel Monte Carlo estimation procedure is showcased in Figure 3 for two probability levels, where we have fixed the CDF regularity parameter to be $k = 6$ (i.e. $F_Q \in C^7(\Theta)$; cf. also Fig. 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 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 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



(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 $\tau = 0.95$.

extreme quantile (same for the conditional value-at-risk), which are less accurately estimated in view of relation (17).

As discussed in Sect. 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 post-processing steps. A direct consequence of Thm. 2.3 with $m = 1$ thus is that the simultaneous multilevel Monte Carlo procedure has a computational complexity of $\mathcal{O}\big(\varepsilon^{-2.29}\ln(\varepsilon^{-1})\big)$, which is of course identical to the multilevel Monte Carlo CDF estimator discussed above. This complexity result is confirmed by the results shown in Fig. 3(C). In fact, we observe that the computational complexity is almost identical for both values of τ.

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 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 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 is for example the extension of the results presented in this work to 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 dominating the overall complexity in high dimension (c.f. "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; c.f. (17). 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, 27], 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 and will be presented elsewhere.

Acknowledgments

This work has been supported by CADMOS (Center for Advanced Modeling Science). F. N. also acknowledges funding from the EU-FP7 framework program within the project UMRIDA under grant agreement no. ACP3-GA-2013-605036.

References

[1] S. Asmussen and P. W. Glynn, *Stochastic Simulation: Algorithms and Analysis*, Springer, 2007.
[2] S. Asmussen, J. L. Jensen, and L. Rojas-Nandayapa, *On the Laplace transform of the lognormal distribution*, Methodol. Comput. Appl. Probab. **18** (2016), no. 2, 441–458.
[3] S.-K. Au and J. L. Beck, *Estimation of small failure probabilities in high dimensions by subset simulation*, Probabilist. Eng. Mech. **16** (2001), no. 4, 263–277.
[4] R. Avikainen, *On irregular functionals of SDEs and the Euler scheme*, Finance Stoch. **13** (2009), no. 3, 381–401.
[5] C. Bierig and A. Chernov, *Approximation of probability density functions by the multilevel Monte Carlo maximum entropy method*, J. Comput. Phys. **314** (2016), 661–681.
[6] 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), no. 1, 3–40.
[7] 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), no. 1, 3–15.
[8] C. de Boor, *A practical guide to splines*, Springer, 2001.
[9] S. Dereich and T. Mueller-Gronbach, *General multilevel adaptations for stochastic approximation algorithms*, 2015. arXiv:1506.05482.
[10] 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), no. 1, 826–850.
[11] D. Feldman and H. G. Tucker, *Estimation of non-unique quantiles*, Ann. Math. Statist. **37** (1966), no. 2, 451–457.
[12] N. Frikha, *Multi-level stochastic approximation algorithms*, Ann. Appl. Probab. **26** (2016), no. 2, 933–985.
[13] M. B. Giles, *Multilevel Monte Carlo path simulation*, Oper. Res. **56** (2008), no. 3, 607–617.
[14] M. B. Giles, *Multilevel Monte Carlo methods*, Acta Numer. **24** (2015), 259–328.
[15] M. B. Giles, T. Nagapetyan, and K. Ritter, *Multilevel Monte Carlo approximation of distribution functions and densities*, SIAM/ASA J. Uncertain. Quantif. **3** (2015), no. 1, 267–295.
[16] N. K. Govil and R. N. Mohapatra, *Markov and Bernstein type inequalities for polynomials*, J. Inequal. Appl. **3** (1999), no. 4, 349–387.
[17] A.-L. Haji-Ali, F. Nobile, and R. Tempone, *Multi-index Monte Carlo: when sparsity meets sampling*, Numer. Math. **132** (2016), no. 4, 767–806.
[18] C. A. Hall and W. W. Meyer, *Optimal error bounds for cubic spline interpolation*, J. Approximation Theory **16** (1976), no. 2, 105–122.
[19] S. Heinrich, *Monte Carlo complexity of global solution of integral equations*, J. Complexity **14** (1998), no. 2, 151–175.
[20] S. Heinrich, *Multilevel Monte Carlo methods*, Large-scale scientific computing, 2001, pp. 58–67.
[21] P. Holgate, *The lognormal characteristic function*, Comm. Statist. Theory Methods **18** (1989), no. 12, 4539–4548.

[22] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*, Applications of Mathematics, vol. 23, Springer, 1992.
[23] M. Ledoux and M. Talagrand, *Probability in Banach spaces*, Springer, 1991.
[24] R. T. Rockafellar and S. Uryasev, *Conditional value-at-risk for general loss distributions*, J. Bank. Financ. **26** (2002), no. 7, 1443–1471.
[25] 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), no. 3, 569–600.
[26] L. N. Trefethen, *Approximation Theory and Approximation Practice*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2013.
[27] E. Ullmann and I. Papaioannou, *Multilevel estimation of rare events*, SIAM/ASA J. Uncertain. Quantif. **3** (2015), no. 1, 922–953.
[28] A. W. van der Vaart, *Asymptotic statistics*, Cambridge University Press, 2000.

CSQI, Institute of Mathematics, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

E-mail address, S. Krumscheid: `sebastian.krumscheid@epfl.ch`

E-mail address, F. Nobile: `fabio.nobile@epfl.ch`

Recent publications:

INSTITUTE of MATHEMATICS
MATHICSE Group

Ecole Polytechnique Fédérale (EPFL)

CH-1015 Lausanne

2017

01.2017 LUCA DEDÈ, HARALD GARCKE, KEI FONG LAM:
A Hele-Shaw-Cahn-Hiliard model for incompressible two-phase flows with different densities

02.2017 ROBERT LUCE, OLIVIER SÈTE:
The index of singular zeros of harmonic mappings of anti-analytic degree one

03.2017 ELEONORA MUSHARBASH, FABIO NOBILE:
Dual Dynamically orthogonal approximation of incompressible Navier Stokes equations with random boundary conditions

04.2017 ANNA TAGLIABUE, LUCA DEDÈ, ALFIO QUARTERONI:
A Complex blood flow patterns in an idealized left ventricle: a numerical study

05.2017 DANIEL KRESSNER, ROBERT LUCE, FRANCESCO STATTI:
Incremental computation of block triangular matrix exponentials with application with option pricing

06.2017 ANDREA BARTEZZAGHI, LUCA DEDÈ, ALFIO QUARTERONI:
Biomembrane modeling with Isogeometric Analysis

07.2017 ANNALISA BUFFA, EDUARDO M. GARAU:
A posteriori error estimators for hierarchical B-spline discretizations

08.2017 ANNALISA BUFFA, CARLOTTA GIANNELLI:
Adaptive isometric methods with hierarchical splines: optimality and convergence rates

09.2017 PABLO ANTOLIN, ANNALISA BUFFA, MATHIEU FABRE:
A priori error for unilateral contact problems with Lagrange multipliers and IsoGeometric Analysis

10.2017 MICHELE PISARONI, FABIO NOBILE, PENELOPE LEYLAND:
Continuation Multi-A priori error for unilateral contact problems with Lagrange multipliers and IsoGeometric Analysis

11.2017 ANDREAS FROMMER, CLAUDIA SCHIMMEL, MARCEL SCHWEITZER:
Bounds for the decay of the entries in inverses and Cauchy-Stieltjes functions of sparse, normal matrices

12.2017 SEBASTIAN KRUMSCHEID, FABIO NOBILE:
Multilevel Monte Carlo approximation of functions,

*****