School of Mathematical
and Physical Sciences

University of Reading

Department of Mathematics and Statistics

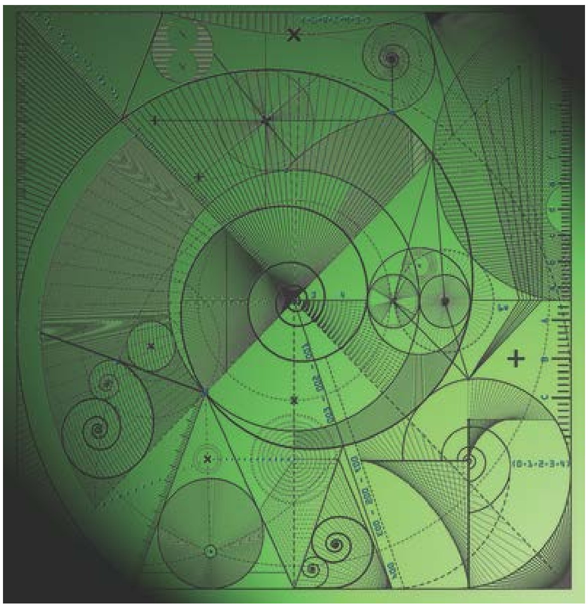
Preprint MPS-2014-13

6 May 2014

Estimation of arbitrary order central statistical moments by the Multilevel Monte Carlo Method

by

Claudio Bierig and **Alexey Chernov**



Noname manuscript No.
(will be inserted by the editor)

Estimation of arbitrary order central statistical moments by the Multilevel Monte Carlo Method

Claudio Bierig · Alexey Chernov

the date of receipt and acceptance should be inserted later

Abstract We extend the general framework of the Multilevel Monte Carlo method to multilevel estimation of arbitrary order central statistical moments. In particular, we prove that under certain assumptions, the total cost of a MLMC central moment estimator is asymptotically the same as the cost of the multilevel sample mean estimator and thereby is asymptotically the same as the cost of a single deterministic forward solve. The general convergence theory is applied to a class of obstacle problems with rough random obstacle profiles. Numerical experiments confirm theoretical findings.

Keywords Uncertainty quantification, multilevel Monte Carlo, stochastic partial differential equation, variational inequality, rough surface, random obstacle, statistical moments.

1 Introduction

Estimation of central statistical moments is important for many reasons. The variance (or the standard deviation) is one of the most important characteristics of a random variable, along with the mean. Higher order moments, particularly the third and the fourth moments (or the related skewness and kurtosis) are important in statistical applications, e.g. for tests whether a random variable is normally distributed [7]. Another example is [3], where skewness and kurtosis are utilized in a stopping criteria for a Monte Carlo method. Higher order moments inherit further characterization of a random variable; the problem of determining a probability distribution from its sequence of moments is widely known as *the problem of moments* [1]. This paper is dedicated to estimation of arbitrary order central statistical moments by means of the Multilevel Monte Carlo method, a non-intrusive sampling-based multiscale approach particularly suitable for uncertainty propagation in complex forward problems.

Claudio Bierig · Alexey Chernov
Department of Mathematics and Statistics, University of Reading, Whiteknights Campus, PO Box 220, Berkshire RG6 6AX, United Kingdom. E-mail: c.bierig@pgr.reading.ac.uk, a.chernov@reading.ac.uk

To facilitate the presentation, let us consider an abstract well posed forward problem

$$u = S(\alpha), \tag{1}$$

where α are model parameters (model input), u is the unique solution of the forward problem (model output) and S is the corresponding solution operator. As an illustrative example, we consider a contact problem between deformable bodies. The quantity of interest (observable) X might be either the solution itself, or some general, possibly nonlinear, continuous functional of the solution $X = F(u)$. Therefore the observable X might be either a spatially varying function (e.g. the displacement of a deformable body or the contact stress) or a scalar quantity (e.g. the size of the actual contact area). Typically, a model description contains probabilistic information about input parameters α and if it is possible to generate samples of α, then samples of X may be generated via the forward map $X = F \circ S(\alpha)$.

However, the solution operator S is usually given implicitly as an inverse of a certain differential or integral operator and the inverse can only be computed approximately by means of a numerical method. In view of this, only approximate samples of X are computable via

$$X_\ell = F_\ell \circ S_\ell(\alpha) \tag{2}$$

for a certain approximation F_ℓ of F. If the numerical method is convergent, we have $X_\ell \to X$ for $\ell \to \infty$ in a suitable sense. The numerical methods and computing resources available nowadays are often capable of providing a good quality approximation of the observable $X_\ell \approx X$. Given this, the mean of X can be approximated by the plain averaging of approximate samples X_ℓ, which is known as the Monte Carlo (MC) approximation. However, there is need to balance two sources of error, the statistical error and the discretization error, significantly increases the total computational cost, see e.g. [4, Section 3] for the rigorous discussion.

In this paper develop and analyze the approximation of the centered statistical moments or arbitrary order in a Multilevel Monte Carlo (MLMC) framework. The MLMC method was developed in the last few years as an improvement of the standard MC, see [10, 6, 2]. Instead of computing samples at a fixed resolution level ℓ, the MLMC uses a hierarchy of resolutions $\ell = 1, \dots, L$ and a level-dependent sampling strategy. Particularly, within MLMC one computes many (cheap) samples at the coarse resolution and only a few, e.g. several dozens, of (expensive) samples at fine resolution, thereby uniformly distributing the computational work over the level hierarchy. If built in a optimal way, MLMC allows to estimate the mean of the observable at the same asymptotic computational cost as a single forward solution of the deterministic problem, see e.g [6, 2] for application to the random diffusion equation. Of course, this analysis includes approximation of standard non-central (monomial) moments of arbitrary order, see e.g. [5]. Moreover, central moments can be computed as combinations of non-central moments in the post-processing phase. In this paper we are interested in direct MLMC approximation of central statistical moments for several reasons.

1. Frequently, central statistical moments are the final aim of the computation.
2. Evaluation of central moments via combination of non-centered moments computed by MLMC may suffer from numerical instabilities, particularly in regions where central moments are small.

3. Direct evaluation of central moments is no more demanding in terms of the overall computational cost than evaluation the non-centered moments. Particularly useful for stable and efficient numerical evaluation are one-pass update formulae from [15].

In this work we analyze simple and general form MLMC estimators for r-th order central statistical moments introduced in (25) and (29) below. The simple form of the estimators allowing for a unified analysis comes at the expense of a small systematic error which cannot be removed by a simple scaling when $r \geq 4$, cf. [13,8]. The rigorous control of this systematic error is presented below.

The paper is structured as follows. After preliminaries in Section 2, we give an overview of the general MLMC framework and the analysis strategy in Section 3. In Section 4 study in detail the MC estimator of arbitrary order central moments and particularly prove convergence of its bias and variance. In Section 5 we apply the developed theory to the MLMC estimator of arbitrary order central moments. Under additional assumptions we prove the same asymptotic work-error relation of the estimator for an arbitrary r-th moment as the same as for the estimation of the expectation value by the multilevel sample mean. In Section 6 we apply the general theory to a class of random obstacle problems (see also [9,14,4]). In Section 7 we report on the results of numerical experiments supporting the abstract theory.

2 Function spaces and statistical moments

Let $(\Omega, \Sigma, \mathbb{P})$ be a complete probability space with Ω being the set of random events, $\Sigma \subset 2^{\Omega}$ the σ-algebra and $\mathbb{P}$ the probability measure. Furthermore let Z be a complete metric space with Borel σ-algebra $\mathcal{B}(Z)$. A mapping $\alpha : \Omega \to Z$ is called a random field, if it is Σ-$\mathcal{B}(Z)$ measurable. We denote with

$$L^0(\Omega, Z) := \{\alpha : \Omega \to Z : \alpha \text{ is } \Sigma\text{-}\mathcal{B}(Z) \text{ measurable}\}$$

the set of all random fields $\alpha : \Omega \to Z$. For a Banach space Z we define

$$\|\alpha\|_{L^p(\Omega,Z)} := \begin{cases} \left(\int_{\Omega} \|\alpha(\omega)\|_Z^k \, \mathrm{d}\mathbb{P}(\omega) \right)^{1/p} & 1 \leq p < \infty, \\ \operatorname*{ess\,sup}_{\omega \in \Omega} \|\alpha(\omega)\|_Z & p = \infty. \end{cases} \tag{3}$$

The Bochner-Lebesgue spaces are defined by

$$L^p(\Omega, Z) := \left\{ \alpha \in L^0(\Omega, Z) : \|\alpha\|_{L^p(\Omega,Z)} < \infty \right\} / \mathcal{N}, \tag{4}$$

where $\mathcal{N} := \left\{ \alpha \in L^0(\Omega, Z) : \alpha = 0 \quad \mathbb{P}\text{-a.e.} \right\}$ and the norm of $L^p(\Omega, Z)$ is defined by (3). Throughout the paper the (Banach space-valued) elements of $L^p(\Omega, Z)$ are termed *random fields*. In this paper we investigate two particular cases particularly important in applications: i) the real-valued random variables, i.e. $Z = \mathbb{R}$, and ii) the Sobolev space-valued random fields, i.e. $Z = W^{s,p}(D)$ where D is a bounded Lipschitz domain, $s \geq 0$ is an integer and $1 \leq p \leq \infty$. The analysis for i) and ii) will be carried out in parallel and to cover both cases we will work with the family of Banach spaces B^p where

$$\text{i) } B^p = \mathbb{R} \quad \text{or} \quad \text{ii) } B^p = W^{s,p}(D), \quad 1 \leq p \leq \infty \tag{5}$$

and distinguish the special case $H = B^2$, i.e.

$$\text{i) } H = \mathbb{R} \quad \text{or} \quad \text{ii) } H = W^{s,2}(D) \equiv H^s(D). \tag{6}$$

The nonnegative integer s will be fixed throughout the paper therefore is omitted in the notation (5) and (6) for brevity. In both cases, H is a Hilbert space with an inner product $\langle \cdot, \cdot \rangle_H$ and B^p is a Banach space with the a norm $\| \cdot \|_{B^p}$ specified in Table 1 for definiteness.

H	$\langle f, g \rangle_H$	B^p	$\|f\|_{B^p}$
$\mathbb{R}$	fg	$\mathbb{R}$	$\lvert f \rvert$
$W^{s,2}(D)$	$\sum_{\lvert\alpha\rvert \le s} \int_D (\partial^\alpha f)(\partial^\alpha g)\,\mathrm{d}x$	$W^{s,p}(D)$	$\begin{cases} \left(\sum_{\lvert\alpha\rvert \le s} \int_D \lvert\partial^\alpha f\rvert^p \,\mathrm{d}x \right)^{\frac{1}{p}}, & p < \infty \\ \max_{\lvert\alpha\rvert \le s} \operatorname{ess\,sup}_D \lvert\partial^\alpha f\rvert, & p = \infty \end{cases}$

Table 1 Overview of the short-hand notations for Sobolev and Bochner spaces and their inner products and norms; $D \subset \mathbb{R}^d$, $\alpha = (\alpha_1, \dots, \alpha_d)$ is a multiindex with nonnegative integer components, $|\alpha| = \alpha_1 + \dots + \alpha_d$ and $\partial^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}$.

Under these notations a (generalized) Hölder inequality holds: Suppose $1 \le n, p, q \le \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$, then for $f \in B^{np}$ and $g \in B^{nq}$

$$\|fg\|_{B^n} \le c_H \|f\|_{B^{np}} \|g\|_{B^{nq}}. \tag{7}$$

We observe that $c_H = 1$ when $B^n = \mathbb{R}$ or $L^n(D)$ since (7) is the standard Hölder inequality in the latter case. Moreover, c_H depends only on the order of the derivative s when $B^n = W^{s,n}(D)$. In particular, $c_H = 3$ when $s = 1$. This estimate generalizes to the r-fold products with $r \ge 2$ as follows. For integers $1 \le n, p_1, \dots, p_r \le \infty$ satisfying $\frac{1}{p_1} + \dots + \frac{1}{p_r} = 1$ and functions $f_i \in B^{np_i}$, $i = 1, \dots, r$ it holds that

$$\left\| \prod_{i=1}^r f_i \right\|_{B^n} \le c_H^{r-1} \prod_{i=1}^r \|f_i\|_{B^{np_i}}. \tag{8}$$

Moreover for random variables $X_i \in L^{np_i}(\Omega, B^{np_i})$ there holds

$$\left\| \prod_{i=1}^r X_i \right\|_{L^n(\Omega, B^n)} \le c_H^{r-1} \prod_{i=1}^r \|X_i\|_{L^{np_i}(\Omega, B^{np_i})}. \tag{9}$$

Bochner spaces of the type $L^p(\Omega, B^p)$, will be frequently used throughout the paper and to simplify the notations we write

$$\| \cdot \|_p := \| \cdot \|_{L^p(\Omega, B^p)} \tag{10}$$

for the norm on these spaces. Notice that $L^2(\Omega, H)$ is a Hilbert space with inner product

$$\langle X, Y \rangle := \int_\Omega \langle X(\omega), Y(\omega) \rangle_H \, \mathrm{d}\mathbb{P}(\omega), \quad X, Y \in L^2(\Omega, H).$$

For $X \in L^1(\Omega, H)$ we define its mean

$$\mathbb{E}[X] = \int_\Omega X(\omega) \, \mathrm{d}\mathbb{P}(\omega)$$

and the M-sample mean Monte Carlo estimator

$$E_M[X] = \frac{1}{M} \sum_{i=1}^{M} X_i, \tag{11}$$

where the $X_i \in H$ are independent realizations of X. Notice that the sample mean estimator $E_M[X] \in L^1(\Omega, H)$ is in fact a random field whereas $\mathbb{E}[X] \in H$ is a deterministic quantity. We recall that a randomized estimator is called *unbiased* if it is exact in the mean, otherwise it is termed *biased*. In particular, the sample mean estimator (11) is unbiased, since $\mathbb{E}[E_M[X] - X] = 0$.

This paper is dedicated to approximation of central statistical moments

$$\mathbb{M}^r[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^r\right] \tag{12}$$

where the order $r \geq 2$ is an arbitrary fixed integer. We distinguish the second order central moment, the variance $\mathbb{V}[X]$, and the related covariance $\mathbb{C}[X, Y]$ defined by

$$\mathbb{V}[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right], \qquad \mathbb{C}[X, Y] := \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]$$

for two sufficiently regular random fields X, Y. From Jensen's inequality [18, Corollary V.5.1] and the Hölder inequality we obtain the upper bound

$$\|\mathbb{E}[X^r]\|_{B^n} \leq \mathbb{E}[\|X^r\|_{B^n}] \leq \mathbb{E}[\|X\|_{B^{nr}}^r] \leq \mathbb{E}[\|X\|_{B^{nr}}^{nr}]^{\frac{1}{n}} = \|X\|_{nr}^r. \tag{13}$$

for $1 \leq n \leq \infty$ and an integer r. Estimate (13) implies that $\mathbb{M}^r[X] \in H$ when $X \in L^{2r}(\Omega, B^{2r})$. In particular, $\mathbb{V}[X] \in H$ and $\mathbb{C}[X, Y] \in H$ when $X, Y \in L^4(\Omega, B^4)$.

3 Preliminary results on the single and multilevel randomized estimation

3.1 Abstract single and multilevel estimation

In the forthcoming error analysis we shall frequently use the quantities

$$\mathcal{V}(S) := \|S - \mathbb{E}[S]\|_{L^2(\Omega, H)}^2 \tag{14}$$

and

$$\mathcal{C}(S, T) := \mathbb{E}\left[\langle S - \mathbb{E}[S], T - \mathbb{E}[T] \rangle_H\right]$$

well defined for random fields $S, T \in L^2(\Omega, H)$. Notice that $\mathcal{V}(S)$ coincides with $\mathbb{V}[S]$ and $\mathcal{C}(S, T)$ with $\mathbb{C}(S, T)$ if $S, T \in L^2(\Omega, \mathbb{R})$ are *real-valued random variables*, which is not true for general *random fields* $S, T \in L^2(\Omega, H)$, $H \neq \mathbb{R}$. Nonetheless, if $S \in L^2(\Omega, H)$ is a randomized estimator of a deterministic quantity $\mathbb{M} \in H$, we occasionally call $\mathcal{V}(S)$ the *variance of the estimator* S. The reason for this slight abuse of terminology is that $\mathcal{V}(S) \in \mathbb{R}_+$ is indeed a measure of variation of the quantity S. Furthermore it is consistent with the existing literature on Monte Carlo Methods and, in particular, the well-known splitting of the mean-square error (MSE) of an estimator into the sum of its (squared) bias and the variance

$$\begin{aligned} \|\mathbb{M}-S\|^2_{L^2(\Omega,H)} &= \|\mathbb{M} - \mathbb{E}[S] + \mathbb{E}[S] - S\|^2_{L^2(\Omega,H)} \\ &= \|\mathbb{M} - \mathbb{E}[S]\|^2_{L^2(\Omega,H)} + 2\mathbb{E}\langle \mathbb{M} - \mathbb{E}[S], \mathbb{E}[S] - S\rangle_H + \|\mathbb{E}[S] - S\|^2_{L^2(\Omega,H)} \\ &= \|\mathbb{M} - \mathbb{E}[S]\|^2_H + \mathcal{V}(S). \end{aligned} \tag{15}$$

Indeed, the inner product is zero since $\mathbb{M} - \mathbb{E}[S]$ is deterministic and $\langle \mathbb{M} - \mathbb{E}[S], \cdot\rangle_H$ is a linear functional on H. In order to facilitate the further discussion we introduce the relative root mean-square error

$$\mathrm{Rel}(\mathbb{M}, S) := \frac{\|\mathbb{M} - S\|_{L^2(\Omega,H)}}{\|\mathbb{M}\|_H}.$$

Frequently, in practical applications the approximate evaluation of the quantity $\mathbb{M}$ involves some kind of *deterministic approximation procedure.* In this case the method can be interpreted as a two-stage approximation: there exists a sequence $\mathbb{M}_\ell \to \mathbb{M}$ converging strongly in H as $\ell \to \infty$, and a family of *single level* randomized estimators S_ℓ approximating $\mathbb{M}_\ell$. Then by the triangle inequality we have the upper bound (notice that if $\mathbb{M}_L = \mathbb{E}[S_L]$, the identity (15) provides a sharper result)

$$\|\mathbb{M} - S_L\|_{L^2(\Omega,H)} \le \|\mathbb{M} - \mathbb{M}_L\|_H + \|\mathbb{M}_L - S_L\|_{L^2(\Omega,H)}. \tag{16}$$

This estimate suggests that S_L should be chosen to balance the summands, namely

$$\|\mathbb{M} - \mathbb{M}_L\|_H \approx \|\mathbb{M}_L - S_L\|_{L^2(\Omega,H)} \tag{17}$$

implying

$$\|\mathbb{M} - S_L\|_{L^2(\Omega,H)} \lesssim \mathcal{E}, \qquad \mathcal{E} := \|\mathbb{M} - \mathbb{M}_L\|_H. \tag{18}$$

Typically, the evaluation cost of the estimator S_L, while keeping the balance (17), increases significantly with increasing L and might become unfeasibly large. As an alternative, we consider an abstract *multilevel estimator*

$$S_{\mathrm{ML}} := \sum_{\ell=1}^{L} T_\ell, \tag{19}$$

built of a sequence of estimators T_ℓ each approximating the individual differences

$$\Delta\mathbb{M}_\ell := \begin{cases} \mathbb{M}_\ell - \mathbb{M}_{\ell-1}, & \ell > 1, \\ \mathbb{M}_\ell, & \ell = 1. \end{cases} \tag{20}$$

These estimators may e.g. depend on the finest level $T_\ell = T_\ell(L)$; we omit such details for simplicity of the presentation in this section. We remark that the estimators T_ℓ are not required to be either unbiased or independent (in fact, independence would imply sharper upper bounds for the MSE, see Theorem 1, Theorem 2 and Remark 2 in Section 5). The approximation (19) is reasonable since

$\mathbb{M}_L = \sum_{\ell=1}^{L} \Delta\mathbb{M}_\ell$. Thus, similarly to (16) we obtain

$$\begin{aligned}\|\mathbb{M} - S_{\mathrm{ML}}\|_{L^2(\Omega,H)} &\le \|\mathbb{M} - \mathbb{M}_L\|_H + \|\mathbb{M}_L - S_{\mathrm{ML}}\|_{L^2(\Omega,H)} \\ &\le \|\mathbb{M} - \mathbb{M}_L\|_H + \sum_{\ell=1}^{L} \|\Delta\mathbb{M}_\ell - T_\ell\|_{L^2(\Omega,H)}.\end{aligned} \tag{21}$$

This upper bound is balanced when T_ℓ are chosen to guarantee

$$\mathcal{E} = \|\mathbb{M} - \mathbb{M}_L\|_H \approx L\|\Delta\mathbb{M}_\ell - T_\ell\|_{L^2(\Omega,H)}$$

for all $\ell = 1, \dots, L$ implying

$$\|\mathbb{M} - S_{\mathrm{ML}}\|_{L^2(\Omega,H)} \lesssim \mathcal{E}. \tag{22}$$

Comparing this estimate with (18) we observe that both single and multilevel estimators admit the same upper bound for the mean-square error. However, the multilevel estimator is (potentially) much faster to compute. Indeed, the computational cost of the estimators T_ℓ is typically determined by the magnitude of the *relative error* $\mathrm{Rel}(\Delta\mathbb{M}_\ell, T_\ell)$. Then the sum in the right-hand side of (21) takes the form

$$\mathcal{E} \approx \sum_{\ell=1}^{L} \|\Delta\mathbb{M}_\ell - T_\ell\|_{L^2(\Omega,H)} = \sum_{\ell=1}^{L} \mathrm{Rel}(\Delta\mathbb{M}_\ell, T_\ell) \times \|\Delta\mathbb{M}_\ell\|_H. \tag{23}$$

In view of this representation and the fact $\Delta\mathbb{M}_\ell \to 0$ in H we observe that in the asymptotic regime the estimators T_ℓ are allowed to have larger relative errors for higher levels ℓ than for lower levels, as long as they are balanced with the value $\|\Delta\mathbb{M}_\ell\|_H$. Whereas for the single level estimator we have

$$\mathcal{E} \approx \|\mathbb{M}_L - S_L\|_{L^2(\Omega,H)} = \mathrm{Rel}(\mathbb{M}_L, S_L) \times \|\mathbb{M}_L\|_H \tag{24}$$

where $\|\mathbb{M}_L\|_H$ is (asymptotically) bounded from below whenever $\mathbb{M} \neq 0$. In other words, the single level estimator have to achieve a small relative error to keep the absolute mean-square error of the estimator small.

Comparing (23) and (24) we observe that the structure of the estimator S_{ML} built of individual approximations of level corrections $\Delta\mathbb{M}_\ell$ has a significant advantage over the estimator S_L since it allows to transfer the time consuming computations on high levels $\ell \sim L$ to significantly less demanding computations on lower levels $\ell \sim 1$.

3.2 Single and Multilevel Monte Carlo estimators for central statistical moments

In this section we introduce Single and Multilevel Monte Carlo estimators for a central statistical moment $\mathbb{M}^r[X]$ of order $r \ge 2$, cf. (12), that will be studied in detail in the forthcoming sections. Recalling the definition (11) of the sample mean $E_M[X]$ we introduce a single level estimator

$$S_M^r[X] := \frac{1}{M} \sum_{i=1}^{M} (X_i - E_M[X])^r, \tag{25}$$

where the X_i are independent and identically distributed (iid) samples of X. The estimator (25) is possibly the most natural and intuitive computable sample approximation for $\mathbb{M}^r[X]$ with $r \geq 2$. However, as we will prove in Lemma 3 below, the estimator (25) is biased in general. In particular but important cases $r = 2$ and 3, the estimator (25) can be made unbiased with a minor modification. Indeed, the rescaled estimators

$$\tilde{S}^2_M[X] := \frac{M}{M-1} S^2_M[X], \qquad \tilde{S}^3_M[X] := \frac{M^2}{(M-1)(M-2)} S^3_M[X], \tag{26}$$

are unbiased, i.e. satisfy $\mathbb{E}[\tilde{S}^r_M[X]] = \mathbb{M}^r[X]$, for $r = 2, 3$. One might expect that a multiple of $S^r_M[X]$ is an unbiased estimator for higher order central moments as well, but already for $r = 4$ it holds that

$$\mathbb{E}[S^4_M[X]] = \frac{M-1}{M^3} \left((M^2 - 3M + 3)\mathbb{M}^4[X] + 3(2M-3)\mathbb{M}^2[X]^2 \right) \tag{27}$$

see e.g. [13,8] (in Lemma 3 below we derive a general representation for $\mathbb{E}[S^r_M[X]]$ with an arbitrary r). In this case the unbiased estimate for $\mathbb{M}^4[X]$ takes the form

$$\tilde{S}^4_M[X] := \frac{M^2}{(M-2)(M-3)} \left(\frac{M+1}{M-1} S^4_M[X] - 3 S^2_M[X]^2 \right). \tag{28}$$

A similar result holds for any r: an unbiased estimator $\tilde{S}^r_M[X]$ can be built as a weighted sum of $S^r_M[X]$ with a nonlinear combination of $S^2_M[X], \dots, S^{r-2}_M[X]$. Such representations for $\tilde{S}^r_M[X]$ can be obtained for an arbitrary r and used for a single level Monte Carlo estimation of $\mathbb{M}^r[X]$. However, an unbiased estimation for $r \geq 4$ may cause some technical difficulties, as we explain below.

Notice that the above description fits into the abstract framework of Section 3.1. Indeed, suppose that X_ℓ is an approximation to X at level ℓ, then (16) holds with

$$\mathbb{M} := \mathbb{M}^r[X], \qquad \mathbb{M}_L := \mathbb{M}^r[X_L], \qquad S_L := S^r_M[X_L].$$

We introduce a multilevel estimator

$$\mathcal{S}^r_{\mathrm{ML}}[X] = \sum_{\ell=1}^{L} S^r_{M_\ell}[X_\ell] - S^r_{M_\ell}[X_{\ell-1}], \tag{29}$$

where in the summands $S^r_{M_\ell}[X_\ell] - S^r_{M_\ell}[X_{\ell-1}]$ are built from M_ℓ pairs of samples $(X_\ell, X_{\ell-1})_i$, both computed for the *same* realization of input parameters (the same random event $\omega_i \in \Omega$). This fits into the abstract framework of Section 3.1 with

$$S_{\mathrm{ML}} := \mathcal{S}^r_{\mathrm{ML}}[X], \qquad T_\ell := S^r_{M_\ell}[X_\ell] - S^r_{M_\ell}[X_{\ell-1}].$$

Evidently, since the estimator (25) is biased the multilevel estimator (29) is (in general) biased as well, whereas an unbiased estimator can be defined as

$$\tilde{\mathcal{S}}^r_{\mathrm{ML}}[X] = \sum_{\ell=1}^{L} \tilde{S}^r_{M_\ell}[X_\ell] - \tilde{S}^r_{M_\ell}[X_{\ell-1}]. \tag{30}$$

The unbiasedness is an important property in the framework of Monte Carlo estimation. In particular, unbiased estimators *i)* have no systematic statistical error (i.e. are exact in the mean, by definition) and *ii)* allow for a sharper and more elegant upper bound for the mean-square error (since unbiasedness enables the usage of orthogonality instead of the triangle inequality). Nonetheless, we consider the estimator (30) as inconvenient for Multilevel Monte Carlo estimation when $r \geq 4$ in view of the following difficulties with its numerical evaluation:

1. The multilevel estimator (29) allows for a general convergence theory for an arbitrary moment order r developed and presented in the this paper. The treatment of unbiased estimators (30) will require a separate analysis for every particular r. We refer to [4] for the analysis of the unbiased multilevel sample variance estimators, cf. (30) with $r = 2$.
2. Though the unbiased estimator (30) has no systematic statistical error, whereas the estimator (29) does not enjoy this property, it remains unclear which of the terms yields the smaller mean-square error.
3. A straight forward evaluation of $\tilde{S}^r_{M_\ell}[X_\ell]$ requires evaluation of $S^r_{M_\ell}[X_\ell]$ as an intermediate step. Possibly, no further post-processing is needed when the approximation $S^r_{M_\ell}[X_\ell]$ is good enough.

As we will see in Section 5, the estimate for the MSE of $\mathcal{S}^r_{\mathrm{ML}}[X]$ requires estimates for

$$\|\mathbb{M}^r[X] - \mathbb{M}^r[Y] - \mathbb{E}[S^r_M[X] - S^r_M[Y]]\|_H \tag{31}$$

and

$$\mathcal{V}(S^r_M[X] - S^r_M[Y]), \tag{32}$$

as an intermediate step, which is similar to the multilevel sample mean estimators, analyzed e.g. in [2,6]. The situation we face in the present work is more complicated in two ways. Firstly, as discussed above, the estimator (29) is biased and thus a special care should be paid to the control of the bias. Secondly, the summands in the sample mean estimator $E_M[X] = \frac{1}{M}\sum_{i=1}^M X_i$ are uncorrelated by construction, implying $\mathcal{C}[X_i, X_j] = 0$ whenever $i \neq j$. This is no longer true for the estimator (25), i.e. in general

$$\mathcal{C}\left[\left(X_i - \frac{1}{M}\sum_{k=1}^M X_k\right)^r, \left(X_j - \frac{1}{M}\sum_{k=1}^M X_k\right)^r\right] \neq 0, \quad i \neq j.$$

This leads to a more complicated estimation of the variance for the higher order moment case compared to sample mean estimation, see [4] for the estimation of the sample variance.

We remark that, as a by-product of our error analysis, by choosing $Y = 0$ in (31) and (32) we immediately obtain upper bounds for the Single Level MC estimator $S^r_M[X]$, cf. [13, p. 348-349] for a related estimate in the special case when $X : \Omega \to H$ is a real-valued random variable, i.e. $H = \mathbb{R}$. We refer to Corollary 1 below for the rigorous formulation of this result.

4 Single Level Monte Carlo Approximation

Throughout this section we work with random fields $X, X_L, Y : \Omega \to H$ with H the Hilbert space $\mathbb{R}$ or $W^{s,2}(D)$ and B^p be the Banach space $\mathbb{R}$ or $W^{s,p}(D)$ respectively. Throughout this section we assume that for a fixed integer $r \geq 2$

there holds $X, X_L, Y \in L^{2r}(\Omega, B^{2r})$. This condition guarantees that the first r central moments are well defined in H. We derive upper bounds for (31) and (32), and thereby, as indicated in Section 3, an upper bound for the MSE of the Single and the Multilevel Monte Carlo estimators.

4.1 Preliminary results

Let us denote by $\hat{\Lambda}^r_M := \{1, \dots, M\}^r$ the set of multiindices with r integer components ranging (independently from each other) from 1 to M and consider its subset $\Lambda^r_M \subset \hat{\Lambda}^r_M$ containing all multiindices with *no unique components*, i.e.

$$\Lambda^r_M := \left\{ \underline{j} \in \hat{\Lambda}^r_M \,|\, \forall k \; \exists \ell \neq k : j_k = j_\ell \right\}. \tag{33}$$

This abstract index set will allow for an easy representation and further handling of correlations of multiple centered iid random fields. In particular, suppose X is a sufficiently smooth random field and let $\{X_i\}_{i=1}^M$ be iid samples of X. Denote $\bar{X} := X - \mathbb{E}[X]$. For $\underline{j} = (j_1, \dots, j_r) \in \hat{\Lambda}^r_M$ we introduce the product

$$\bar{X}_{\underline{j}} := \prod_{k=1}^{r} \bar{X}_{j_k}. \tag{34}$$

Evidently, for the multiindex $\underline{j} \in \hat{\Lambda}^r_M \setminus \Lambda^r_M$ whose ν-th component has a unique value it holds that

$$\mathbb{E}[\bar{X}_{\underline{j}}] = \mathbb{E}[\bar{X}_\nu] \mathbb{E}\left[\prod_{k=1, k\neq\nu}^{r} \bar{X}_{j_k} \right] = 0 \tag{35}$$

since $\{\bar{X}_i\}_{i=1}^M$ are iid realizations of the centered random field $\bar{X}$. Moreover, any two r-multiindices $\underline{j}, \underline{j}' \in \hat{\Lambda}^r_M$ and their concatenation $(\underline{j}, \underline{j}') \in \hat{\Lambda}^{2r}_M$ it necessarily holds that

$$\mathcal{C}\left(\bar{X}_{\underline{j}}, \bar{X}_{\underline{j}'} \right) = 0, \tag{36}$$

once $(\underline{j}, \underline{j}') \in \hat{\Lambda}^{2r}_M \setminus \Lambda^{2r}_M$. This can be seen easily for $H = \mathbb{R}$, $H = L^2(D)$ and $H = H^1(D)$, and actually any other Hilbertian Sobolev space $W^{k,2}(D)$. Equation (36) also holds true if there exist no ν, μ such that $\underline{j}_\nu = \underline{j}'_\mu$, since in this case $\bar{X}_{\underline{j}}$ and $\bar{X}_{\underline{j}'}$ are independent random fields.

Lemma 1 *The cardinality of the set Λ^r_M defined in* (33) *admits the upper bound*

$$|\Lambda^r_M| \leq (r-1)^{r-1} M^{\lfloor \frac{r}{2} \rfloor}. \tag{37}$$

Proof We claim that for integer M and $r \geq 2$ there holds the representation

$$|\Lambda^r_M| = M \left(1 + \sum_{k=1}^{r-3} \binom{r-1}{k} |\Lambda^{r-k-1}_{M-1}| \right). \tag{38}$$

To show this, we count the number of multiindices $\underline{i} = (i_1, \dots, i_r) \in \Lambda^r_M$ involving combinatorial arguments. Assume that the first component i_1 is fixed (the multiplier M appears in (38) since i_1 may take M distinct values). Considering the sum in the parenthesis we observe that its k-th term represents the number of indices in Λ_M^r, for which there exist $2 \leq j_1 < \ldots < j_k \leq r$ with $i_1 = i_{j_1} = \ldots = i_{j_k}$ and all other indices are different from i_1. To see this, we notice that there are $\binom{r-1}{k}$ different possibilities to choose $j_1, \ldots, j_k$. Moreover, there are $r-k$ remaining components distinct from i_1, containing no unique values (otherwise there would be a contradiction to the definition of Λ_M^r). Therefore there are precisely $|\Lambda_{M-1}^{r-k-1}|$ admissible combinations for the remaining components. The summands with $k = 0$ and $k = r-2$ have no contribution to the sum since they both correspond to multiindices with at least one unique component value. The case $k = r-1$ implies that all components of $\underline{i}$ coincide with the fixed i_1 and there is only one such multiindex. Thus, (38) is proved.

We have $|\Lambda_M^1| = 0$ and $|\Lambda_1^r| = 1$ for $r > 1$. Let the statement of the lemma hold true for any $M-1$ and $1 \leq s < r$. Then we have

$$\begin{aligned}|\Lambda_M^r| =& M\left(1 + \sum_{k=1}^{r-3}\binom{r-1}{k}|\Lambda_{M-1}^{r-k-1}|\right)\\ \leq& M\left(1 + \sum_{k=1}^{r-3}\binom{r-1}{k}(r-k-2)^{r-k-2}(M-1)^{\lfloor\frac{r-k-1}{2}\rfloor}\right).\end{aligned}$$

The statement of the lemma follows by the estimate $\binom{r}{k} \leq r^k$ and

$$|\Lambda_M^r| \leq M\left(1 + \sum_{k=1}^{r-3}(r-1)^{r-2}M^{\lfloor\frac{r-k-1}{2}\rfloor}\right) \leq (r-1)^{r-1}M^{\lfloor\frac{r}{2}\rfloor}.$$

Lemma 2 *Let $X, Y : \Omega \to H$ be two random variables with H the Hilbert space $\mathbb{R}$ or $W^{s,2}(D)$ and B^p be the Banach space $\mathbb{R}$ or $W^{s,p}(D)$ respectively. Furthermore let $r \geq 2$ and $1 \leq p \leq q(r-1) \leq \infty$ be such that $\frac{1}{p} + \frac{1}{q} = 1$ and $X, Y \in L^{2q(r-1)}(\Omega, B^{2q(r-1)})$. Let $\underline{j} \in \mathbb{N}^r$ and $X_{\underline{j}}$, $Y_{\underline{j}}$ as defined in* (34)*. Then it holds that*

$$\|\mathbb{E}[X_{\underline{j}} - Y_{\underline{j}}]\|_H \leq \|X_{\underline{j}} - Y_{\underline{j}}\|_2 \leq \mathcal{K}(X, Y, r) \tag{39}$$

where

$$\begin{aligned}\mathcal{K}(X, Y, r) := \|X - Y\|_{2p} \times c_H^{r-1} \min\Bigg(& r \max\left\{\|X\|_{2q(r-1)}^{r-1}, \|Y\|_{2q(r-1),}^{r-1}\right\},\\ & (\|X\|_{2q(r-1)} + \|Y\|_{2q(r-1)})^{r-1}\Bigg),\end{aligned}$$

with c_H the Hölder constant from inequality (7)*.*

Proof The first inequality in (39) follows by Jensen's inequality. For the second inequality we need the following identity

$$\prod_{i=1}^{r} a_i - \prod_{i=1}^{r} b_i = \sum_{i=1}^{r}(a_i - b_i)\prod_{k=1}^{i-1} a_k \prod_{k=i+1}^{r} b_k. \tag{40}$$

Applying (40) and the Hölder inequality (9) gives us

$$\begin{aligned}\|X_{\underline{j}} - Y_{\underline{j}}\|_2 &= \left\| \sum_{i=1}^{r} (X_{j_i} - Y_{j_i}) \prod_{k=1}^{i-1} X_{j_k} \prod_{k=i+1}^{r} Y_{j_k} \right\|_2 \\ &\le \sum_{i=1}^{r} c_H^{r-1} \|X - Y\|_{2p} \|X\|_{2q(r-1)}^{i-1} \|Y\|_{2q(r-1)}^{r-i}.\end{aligned} \tag{41}$$

Estimating the sum with

$$\sum_{i=1}^{r} \|X\|_{2q(r-1)}^{i-1} \|Y\|_{2q(r-1)}^{r-i} \le \min \Big(r \max \left\{ \|X\|_{2q(r-1)}^{r-1}, \|Y\|_{2q(r-1)}^{r-1} \right\}, \\ (\|X\|_{2q(r-1)} + \|Y\|_{2q(r-1)})^{r-1} \Big)$$

completes the proof.

4.2 Estimation of the building blocks (31) and (32)

In this section we obtain upper bounds for (31) and (32) required later on in convergence theorem for the multilevel estimator in the forthcoming Section 5. The following notation will be essential in the forthcoming analysis. Let $r \ge 2$ be an integer and $1 \le k \le r$. Denote $m := \min(k+1, r)$. For an m-multiindex $\underline{j} = (j_1, \dots, j_m) \in \Lambda_M^m$ we define its extension to an r-multiindex by

$$\mathcal{E}(\underline{j}) = \begin{cases} (j_1, \dots, j_k, \underbrace{j_{k+1}, \dots, j_{k+1}}_{r-k \text{ times}}), & k < r, \\ (j_1, \dots, j_r), & k = r, \end{cases} \tag{42}$$

Notice the alternative expression: $m = k + 1 - \delta_{k,r}$ where δ_{kr} is the Dirac delta. These definitions and notation (34) allow for a compact representation of the sample estimator (25) as a sum products. Indeed, opening the brackets in (25) we observe

$$S_M^r[X] = \sum_{k=0}^{r} \frac{(-1)^k}{M^{k+1-\delta_{k,r}}} \binom{r}{k} \sum_{\underline{j} \in \hat{\Lambda}_M^{k+1-\delta_{k,r}}} X_{\mathcal{E}(\underline{j})}. \tag{43}$$

The following lemma provides the quantitative structure for the bias of the estimator $S_M^r[X]$.

Lemma 3 *Suppose X is a sufficiently smooth random field so that its statistical moments of any order up to $r \ge 2$ exist. Then it holds that*

$$\begin{aligned}\mathbb{E}\left[S_M^r[X]\right] &= \mathbb{M}^r[X] + M^{-1} \left(\frac{r(r-1)}{2} \mathbb{M}^{r-2}[X] \mathbb{M}^2[X] - r \mathbb{M}^r[X] \right) + \mathcal{R} \\ \mathcal{R} &= \begin{cases} 0, & r = 2, \\ 2M^{-2} \mathbb{M}^3[X], & r = 3, \\ \sum_{\underline{j} \in \Lambda_M^r} c(\underline{j}, M, r) \mathbb{E}[\bar{X}_{\underline{j}}], & r \ge 4, \end{cases}\end{aligned} \tag{44}$$

where the constants $c(\underline{j}, M, r)$ are independent of X. This set of constants is non-unique, however, there exist $c(\underline{j}, M, r)$ such that for $r \geq 4$

$$\sum_{\underline{j} \in \Lambda_M^r} |c(\underline{j}, M, r)| \leq 2 \sum_{k=3}^{r-2} M^{-\lceil \frac{k}{2} \rceil} \binom{r}{k} (k-1)^{k-1} + (r-1)^r M^{-\lceil \frac{r}{2} \rceil}. \tag{45}$$

Proof We assume w.l.o.g. that $\mathbb{E}[X] \equiv \mathbb{M}^1[X] = 0$, since estimator $S_M^r[X]$ and central moments are independent of the value $\mathbb{E}[X]$. Notice that (44) is satisfied when $r = 2$. Indeed, (45) implies that the sum over $\underline{j} \in \Lambda_M^2$ vanish and therefore, since $\mathbb{M}^0[X] \equiv 1$, the identity (44) is equivalent to

$$\mathbb{E}\left[S_M^2[X]\right] = \mathbb{M}^2[X] + M^{-1}\left(\mathbb{M}^0[X]\mathbb{M}^2[X] - 2\mathbb{M}^2[X]\right) = \frac{M-1}{M}\mathbb{M}^2[X] \tag{46}$$

Analogously, for $r = 3$ the estimate (44) is equivalent to

$$\mathbb{E}\left[S_M^3[X]\right] = \mathbb{M}^3[X] - \frac{3}{M}\mathbb{M}^3[X] + \frac{2}{M^2}\mathbb{M}^3[X]. \tag{47}$$

and for $r = 4$ we have

$$\begin{aligned}\mathbb{E}\left[S_M^4[X]\right] =& \mathbb{M}^4[X] + \frac{1}{M}(6\mathbb{M}^2[X]^2 - 4\mathbb{M}^4[X]) + \frac{1}{M^2}(6\mathbb{M}^4[X] - 15\mathbb{M}^2[X]^2) \\ &+ \frac{1}{M^3}(9\mathbb{M}^2[X]^2 - 3\mathbb{M}^4[X]).\end{aligned} \tag{48}$$

Representations (46), (47) and (48) hold true in view of (26),(27) and the assertion of the lemma follows for $2 \leq r \leq 4$. It remains to prove (44) and (45) for the case $r \geq 5$. Taking the expectation value of (43) we obtain

$$\mathbb{E}[S_M^r[X]] = \sum_{k=0}^{r} A_k, \tag{49}$$

where for $m := k + 1 - \delta_{k,r}$

$$A_k := \frac{(-1)^k}{M^m} \binom{r}{k} \sum_{\underline{j} \in \hat{\Lambda}_M^m}^{M} \mathbb{E}[X_{\mathcal{E}(\underline{j})}], \tag{50}$$

By virtue of (35) we observe $\mathbb{E}[X_{\mathcal{E}(\underline{j})}] = 0$ for $\mathcal{E}(\underline{j}) \notin \Lambda_M^r$ and thus

$$\begin{aligned}A_0 &= \frac{1}{M} \sum_{j_1=1}^{M} \mathbb{E}[X_{j_1}^r] = \mathbb{M}^r[X], \\ A_1 &= \frac{-1}{M^2} \sum_{j_1, j_2=1}^{M} \binom{r}{1} \mathbb{E}[X_{j_1} X_{j_2}^{r-1}] = -\frac{r}{M}\mathbb{M}^r[X], \\ A_2 &= \cdots = \frac{1}{M^2}\binom{r}{2}\left(\mathbb{M}^r[X] + (M-1)\mathbb{M}^{r-2}[X]\mathbb{M}^2[X]\right).\end{aligned} \tag{51}$$

This yields

$$\begin{aligned} \mathbb{E}\left[S_M^r[X]\right] &= \mathbb{M}^r[X] + M^{-1}\left(\frac{r(r-1)}{2}\mathbb{M}^{r-2}[X]\mathbb{M}^2[X] - r\mathbb{M}^r[X]\right) + R, \\ R &= \frac{1}{M^2}\frac{r(r-1)}{2}\left(\mathbb{M}^r[X] - \mathbb{M}^{r-2}[X]\mathbb{M}^2[X]\right) + \sum_{k=3}^{r} A_k, \end{aligned} \tag{52}$$

where the remainder can be written as

$$R = \mathcal{R} = \sum_{\underline{j}\in\Lambda_M^r} c(\underline{j}, M, r)\mathbb{E}[X_{\underline{j}}]$$

for some $c(\underline{j}, M, r) \in \mathcal{O}(M^{-2})$ independent of X. This implies the representation (44). It remains to prove (45). For this we observe

$$\begin{aligned} A_r &= \left(\frac{-1}{M}\right)^r \sum_{\underline{j}\in\Lambda_M^r} \mathbb{E}[X_{\underline{j}}], \\ A_{r-1} &= \frac{(-1)^{r-1}}{M^r} r \sum_{\underline{j}\in\Lambda_M^r} \mathbb{E}[X_{\underline{j}}], \\ A_3 &= \frac{-1}{M^4}\binom{r}{3} \sum_{j_1,\dots,j_4=1}^{M} \mathbb{E}[X_{j_1}X_{j_2}X_{j_3}X_{j_4}^{r-3}] \end{aligned} \tag{53}$$

$$= \frac{-1}{M^3}\binom{r}{3}\left(\mathbb{M}^r[X] + (M-1)\mathbb{M}^{r-3}[X]\mathbb{M}^3[X] + 3(M-1)\mathbb{M}^{r-2}[X]\mathbb{M}^2[X]\right)$$

Defining $c(\underline{j}, M, r)$ naturally as given by the definition of A_k (50), neglecting the sign and using (53) we derive the upper bound

$$\begin{aligned} \sum_{\underline{j}\in\Lambda_M^r} |c(\underline{j}, M, r)| \leq & \frac{2}{M^2}\binom{r}{2} + \frac{4}{M^2}\binom{r}{3} \\ & + \sum_{k=4}^{r-2} M^{-k-1}\binom{r}{k} \sum_{\underline{j}\in\hat{\Lambda}_M^{k+1}:\, \mathcal{E}(\underline{j})\in\Lambda_M^r} 1 \\ & + (r-1)M^{-r}|\Lambda_M^r|. \end{aligned}$$

We claim that for $4 \leq k \leq r-2$ the cardinality of the index set $\{\underline{j} \in \hat{\Lambda}_M^{k+1} : \mathcal{E}(\underline{j}) \in \Lambda_M^r\}$ admits the representation

$$\sum_{\underline{j}\in\hat{\Lambda}_M^{k+1}:\, \mathcal{E}(\underline{j})\in\Lambda_M^r} 1 = M\left(|\Lambda_M^k| + k|\Lambda_{M-1}^{k-1}|\right). \tag{54}$$

To prove this we denote by $\underline{j}'$ the first k components of $\underline{j} = (\underline{j}', j_{k+1})$. If $\underline{j}' \in \Lambda_M^k$, then also $\mathcal{E}(\underline{j}) \in \Lambda_M^r$. In the case $\underline{j}' \notin \Lambda_M^k$ and $\mathcal{E}(\underline{j}) \in \Lambda_M^r$ there exists exactly one component j_ℓ, $1 \leq \ell \leq k$, such that $j_\ell \neq j_i$ for $1 \leq i \leq k$, $i \neq \ell$ and $j_\ell = j_{k+1}$. Thus the identity (54) follows by the counting argument similar to the proof of Lemma 1. Utilizing the estimate (37) we get the upper bound

$$\begin{aligned}\sum_{k=4}^{r-2}\frac{1}{M^k}\binom{r}{k}\Big(|\Lambda_M^k|+k|\Lambda_{M-1}^{k-1}|\Big)\\ &\leq\sum_{k=4}^{r-2}\frac{1}{M^k}\binom{r}{k}\Big((k-1)^{k-1}M^{\lfloor\frac{k}{2}\rfloor}+k(k-2)^{k-2}(M-1)^{\lfloor\frac{k-1}{2}\rfloor}\Big)\\ &\leq\sum_{k=4}^{r-2}2M^{-\lceil\frac{k}{2}\rceil}\binom{r}{k}(k-1)^{k-1}.\end{aligned}$$

Moreover, it holds that

$$\frac{2}{M^2}\binom{r}{2}+\frac{4}{M^2}\binom{r}{3}\leq 2M^{-2}\binom{r}{3}2^2,\quad (r-1)M^{-r}|\Lambda_M^r|\leq M^{-\lceil\frac{r}{2}\rceil}(r-1)^r$$

and thereby the proof is complete.

Lemma 4 *Let $X,Y:\Omega\to H$ be two sufficiently smooth random variables with H the Hilbert space $\mathbb{R}$ or $W^{s,2}(D)$. For $r\geq 2$ the estimate*

$$\|\mathbb{M}^r[X]-\mathbb{M}^r[Y]-\mathbb{E}[S_M^r[X]-S_M^r[Y]]\|_H\leq\frac{r(r+1)}{2M}(1+\varepsilon_b(M,r))\mathcal{K}(\bar{X},\bar{Y},r)\quad(55)$$

holds, where $\mathcal{K}(\bar{X},\bar{Y},r)$ is the upper bound in Lemma 2. We have $\varepsilon_b(M,r)\in\mathcal{O}(M^{-1})$ and for $r>3$ the estimate holds for

$$\varepsilon_b(M,r)=\frac{2}{r(r+1)}\left(2\sum_{k=3}^{r-2}M^{-\lceil\frac{k}{2}\rceil+1}\binom{r}{k}(k-1)^{k-1}+(r-1)^rM^{-\lceil\frac{r}{2}\rceil}\right)\quad(56)$$

and $\varepsilon_b(M,r)=0$ for $r=2,3$.

Proof Assume w.l.o.g. that $\mathbb{E}[X]=\mathbb{E}[Y]=0$. Then by Lemma 3 and the triangle inequality, we obtain

$$\begin{aligned}&\|\mathbb{M}^r[X]-\mathbb{M}^r[Y]-\mathbb{E}[S_M^r[X]-S_M^r[Y]]\|_H\\ &\leq M^{-1}\frac{r(r-1)}{2}\|\mathbb{M}^{r-2}[X]\mathbb{M}^2[X]-\mathbb{M}^{r-2}[Y]\mathbb{M}^2[Y]\|_H\\ &\quad+M^{-1}r\|\mathbb{M}^r[X]-\mathbb{M}^r[Y]\|_H+\sum_{\underline{j}\in\Lambda_M^r}|c(\underline{j},M,r)|\|\mathbb{E}[X_{\underline{j}}-Y_{\underline{j}}]\|_H.\end{aligned}$$

We first apply Lemma 2 to estimate the norms in H and then Lemma 3 to bound the sum over Λ_M^r and gain ε_b for $r>3$. For $r=2,3$ we have $\varepsilon_b(r)=0$ due to (46) and (47). Thus the Lemma is proved.

Lemma 5 *Let $X,Y:\Omega\to H$ be two sufficiently smooth random variables with H the Hilbert space $\mathbb{R}$ or $W^{s,2}(D)$. For $r\geq 2$ the estimate*

$$\mathcal{V}(S_M^r[X]-S_M^r[Y])\leq M^{-1}(r+1)^2(1+\varepsilon_v(M,r))\mathcal{K}(\bar{X},\bar{Y},r)^2\quad(57)$$

holds, where $\mathcal{K}(\bar{X},\bar{Y},r)$ is the upper bound in Lemma 2. We have $\varepsilon_v(M,r)\in\mathcal{O}(M^{-1})$ and for $r>3$ the estimate holds for

$$\begin{aligned}\varepsilon_v(M,r) =& \frac{3}{(r+1)^2}\sum_{\substack{k,\,k'\,=\,0,\\ k\,+\,k'\,>\,2}}^{r}(k+k'-1)^{k+k'-1}\binom{r}{k}\binom{r}{k'}M^{-\left\lceil\frac{k+k'}{2}\right\rceil+1}\\ &+\frac{5r^2-4r}{(r+1)^2}M^{-1}.\end{aligned} \tag{58}$$

Furthermore we have $\varepsilon_v(M,2)=0$ and

$$\varepsilon_v(M,3)=\frac{1}{16}\left(\frac{57}{M}+\frac{55}{M^2}+\frac{-120}{M^3}+\frac{28}{M^4}\right).$$

Proof Let us assume w.l.o.g. $\mathbb{E}[X]=\mathbb{E}[Y]=0$ and consider the case $r>3$. First, using (43) we obtain

$$\mathcal{V}(S_M^r[X]-S_M^r[Y])=\mathcal{C}(S_M^r[X]-S_M^r[Y],S_M^r[X]-S_M^r[Y])=\sum_{k,k'=0}^{r}s_{k,k'}$$

where

$$\begin{aligned}s_{k,k'} =& \sum_{k,k'=0}^{r}\frac{(-1)^{k+k'}}{M^{m+m'}}\binom{r}{k}\binom{r}{k'}\\ &\times\sum_{\underline{j}\in\Lambda_M^m}\sum_{\underline{j}'\in\Lambda_M^{m'}}\mathcal{C}(X_{\mathcal{E}(\underline{j})}-Y_{\mathcal{E}(\underline{j})},X_{\mathcal{E}(\underline{j}')}-Y_{\mathcal{E}(\underline{j}')})\end{aligned}$$

and $m=k+1-\delta_{kr}$ and $m'=k'+1-\delta_{k'r}$. We start by considering the summands of the order $\mathcal{O}(M^{-2})$, i.e. $s_{k,k'}$ with $k+k'>2$. Estimating the absolute value of each summand with equation (36), the Cauchy-Schwartz inequality and Lemma 2 leads to

$$|s_{k,k'}|\le\frac{1}{M^{m+m'}}\binom{r}{k}\binom{r}{k'}\mathcal{K}(X,Y,r)^2S(m,m'), \tag{59}$$

with

$$S(m,m'):=\sum_{\substack{\underline{j}\in\hat{\Lambda}_M^m,\underline{j}'\in\hat{\Lambda}_M^{m'}:\\ (\mathcal{E}(\underline{j}),\mathcal{E}(\underline{j}'))\in\Lambda_M^{2r}}}1. \tag{60}$$

We have

$$S(r,r)=|\Lambda_M^{2r}|\le(2r-1)^{2r-1}M^r \tag{61}$$

and for $k<r-1$

$$S(r,k+1)=S(k+1,r)=M(|\Lambda_M^{r+k}|+k|\Lambda_{M-1}^{r+k-1}|)\le 2(r+k-1)^{r+k-1}M^{\left\lfloor\frac{r+k}{2}\right\rfloor+1},$$

which can be proved similarly to (54). Finally, in the case $k,k'<r-1$ we have

$$\begin{aligned}S(k+1,k'+1) =& M^2|\Lambda_M^{k+k'}|+M(k+k')|\Lambda_{M-1}^{k+k'-1}|+2M(M-1)(k+k')|\Lambda_{M-1}^{k+k'-1}|\\ &+2M(M-1)\binom{k+k'}{2}|\Lambda_{M-2}^{k+k'-2}|.\end{aligned}$$

Again we use a similar argumentation to prove this identity. Let us consider $\underline{j} = (\hat{\underline{j}}, j_{k+1})$ and $\underline{j}' = (\hat{\underline{j}}', j'_{k'+1})$.

- When $(\hat{\underline{j}}, \hat{\underline{j}}') \in \Lambda_M^{k+k'}$, the components j_{k+1} and $j'_{k'+1}$ may take any value between 1 and M, resulting in $M^2|\Lambda_M^{k+k'}|$ possible combinations.
- If there exists exactly one component in the concatenated multiindex $(\hat{\underline{j}}, \hat{\underline{j}}')$ occurring only ones, it has to coincide with j_{k+1} or $j'_{k'+1}$. If $j_{k+1} = j'_{k'+1}$ there exist $M(k+k')|\Lambda_{M-1}^{k+k'-1}|$ admissible combinations, in the case $j_{k+1} \neq j'_{k'+1}$ there exist $2M(M-1)(k+k')|\Lambda_{M-1}^{k+k'-1}|$ combinations.
- For two indices occurring just once, one has to be equal to j_{k+1} and the other to $j'_{k'+1}$. Thus there are $2M(M-1)\binom{k+k'}{2}|\Lambda_{M-2}^{k+k'-2}|$ admissible combinations.

Thus the identity for $S(k+1, k'+1)$ is proved and by (37) we get the estimate

$$\begin{aligned} S(k+1,k'+1) \leq & (1 + M^{-1} + 2(1-M^{-1}) + M^{-1})(k+k'-1)^{k+k'-1} M^{\lfloor \frac{k+k'}{2} \rfloor + 2} \\ = & 3(k+k'-1)^{k+k'-1} M^{\lfloor \frac{k+k'}{2} \rfloor + 2}. \end{aligned}$$

This leads for all m, m' to the estimate

$$M^{-m-m'} S(m,m') \leq 3(k+k'-1)^{k+k'-1} M^{\lceil -\frac{k+k'}{2} \rceil}.$$

Thus we have

$$\begin{aligned} \sum_{\substack{k,k'=0 \\ k+k'>2}}^{r} s_{k,k'} \leq \sum_{\substack{k,k'=0 \\ k+k'>2}}^{r} & 3(k+k'-1)^{k+k'-1} \\ & \times \binom{r}{k}\binom{r}{k'} M^{-\lceil \frac{k+k'}{2} \rceil} \mathcal{K}(X,Y,r)^2. \end{aligned} \tag{62}$$

It remains to estimate the summands with $k + k' \leq 2$. For $k = k' = 0$ we have

$$s_{0,0} = \frac{1}{M^2} \sum_{j,j'=1}^{M} \mathcal{C}(X_j^r - Y_j^r, X_{j'}^r - Y_{j'}^r) = \frac{1}{M}\mathcal{V}(X^r - Y^r) \leq \frac{1}{M}\mathcal{K}(X,Y,r)^2,$$

since $\mathcal{C}(X_j^r - Y_j^r, X_{j'}^r - Y_{j'}^r) = 0$ for $j \neq j'$. In the case $k = 1$, $k' = 0$ we gain

$$s_{1,0} = \frac{-r}{M^3} \sum_{j_1,j_2,j'=1}^{M} \mathcal{C}(X_{j_1}X_{j_2}^{r-1} - Y_{j_1}Y_{j_2}^{r-1}, X_{j'}^r - Y_{j'}^r)$$

The summands vanish for $j_1 \neq j'$. Thus we have

$$|s_{1,0} + s_{0,1}| \leq \frac{2r}{M}\mathcal{K}(X,Y,r)^2.$$

For $k = 2$, $k' = 0$ we have

$$s_{2,0} = \frac{1}{M^4}\binom{r}{2} \sum_{j_1,j_2,j_3,j'=1}^{M} \mathcal{C}(X_{j_1}X_{j_2}X_{j_3}^{r-2} - Y_{j_1}Y_{j_2}Y_{j_3}^{r-2}, X_{j'}^r - Y_{j'}^r).$$

The summands are only nonzero when *i)* $j_3 = j'$ implying $j_1 = j_2$ or *ii)* $j_3 \neq j'$ falling apart in three sub-cases:

$$(j_1 = j_2 = j' \neq j_3) \quad \text{or} \quad (j_1 = j_3 \neq j_2 = j') \quad \text{or} \quad (j_2 = j_3 \neq j_1 = j').$$

Case *i)* results in M^2 admissible combinations, whereas the sub-cases of case *ii)* result in $M(M-1)$ admissible combinations each. Thus we have

$$|s_{2,0} + s_{0,2}| \leq \frac{r(r-1)(4M-3)}{M^3} \mathcal{K}(X,Y,r)^2.$$

The remaining term is

$$s_{1,1} = \frac{r^2}{M^4} \sum_{j_1,j_2,j_1',j_2'=1}^{M} \mathcal{C}(X_{j_1} X_{j_2}^{r-1} - Y_{j_1} Y_{j_2}^{r-1}, X_{j_1'} X_{j_2'}^{r-1} - Y_{j_1'} Y_{j_2'}^{r-1}).$$

Counting nontrivial summands we find the following admissible combinations. There are M terms, where all the components take the same value. The case of three identical components different from the fourth results in $2M(M-1)$ admissible combinations (the cases when either j_1 or j_2 are different from the others are not admissible). Furthermore, there are $2M(M-1)$ admissible combinations of two distinct pairs ($j_1 = j_1' \neq j_2 = j_2'$ or $j_1 = j_2' \neq j_2 = j_1'$). Finally there are $M(M-1)(M-2)$ terms where the components take 3 different values. The case of four distinct values is always an inadmissible combination. Thus we have the upper bound

$$|s_{1,1}| \leq \frac{r^2(M^3 + M^2 - M)}{M^4} \mathcal{K}(X,Y,r)^2.$$

Combining the obtained estimates we get

$$\sum_{k+k' \leq 2} s_{k,k'} \leq \Big(M^{-1}(r+1)^2 + M^{-2}(5r^2 - 4r) + M^{-3}(-4r^2 + 3r)\Big) \mathcal{K}(X,Y,r)^2, \tag{63}$$

where we can neglect the last term, since it is negative. A combination of (62) and (63) completes the proof for $r > 3$. The cases $r = 2$ and $r = 3$ are computed explicitly. For $r = 2$ we observe that $\varepsilon_v(2) = 0$ since

$$\begin{aligned} \mathcal{V}(S_M^2[X] - S_M^2[Y]) =& \mathcal{V}(X^2 - Y^2)\frac{(M-1)^2}{M^3} + \mathcal{V}(X_1 X_2 - Y_1 Y_2)\frac{M-1}{M^3} \\ \leq& \mathcal{K}(X,Y,2)^2 M^{-1}. \end{aligned}$$

For $r = 3$ we obtain the estimate

$$\mathcal{V}(S_M^3[X] - S_M^3[Y]) \leq \mathcal{K}(X,Y,3)^2 \left(\frac{16}{M} + \frac{57}{M^2} + \frac{55}{M^3} - \frac{120}{M^4} + \frac{28}{M^5}\right).$$

The proof is complete.

Remark 1 The estimates (55) and (57) are exact in the leading order terms but might be too pessimistic in the higher order terms, due to the use of the triangle inequality and the estimation of the sign-alternating sum by the sum of the absolute values. Moreover, sometimes (e.g. in (60)) we count some inadmissible combinations.

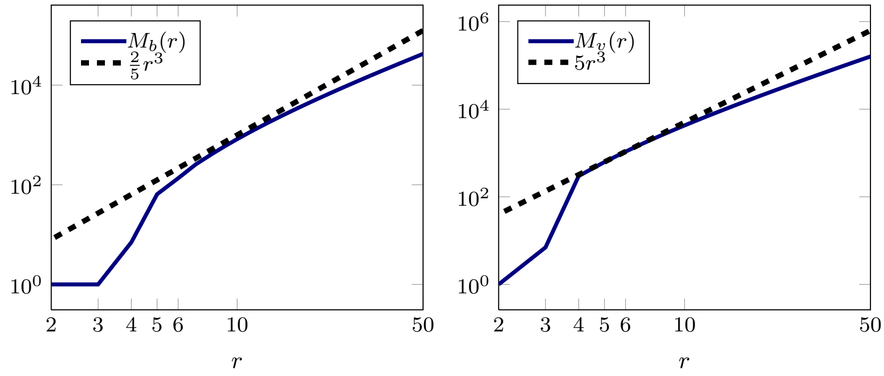


Fig. 1 The maximum number of samples up to which pre-asymptotic behavior can appear for the bias (left) and the variance (right). The black dashed lines are the upper bounds proposed in (66) for small r.

The high order terms decay quickly with increasing M when the moment order r is fixed. To better understand this behavior we consider two quantities

$$M_b(r) := \min\left\{M \in \mathbb{N} \,|\, \forall M' \geq M : \; \varepsilon_b(M', r) \leq 1\right\} \tag{64}$$

and

$$M_v(r) := \min\left\{M \in \mathbb{N} \,|\, \forall M' \geq M : \; \varepsilon_v(M', r) \leq 1\right\}. \tag{65}$$

In particular, there holds $M_b(2) = M_v(2) = 1$ and $M_b(3) = 1$, $M_v(3) = 7$. For higher values $4 \leq r \leq 50$ we solved equations $\varepsilon_b(M, r) = 1$ and $\varepsilon_v(M, r) = 1$ numerically and observe that the preasymptotic bounds

$$M_b(r) \leq \frac{2}{5}r^3, \qquad M_v(r) \leq 5r^3 \tag{66}$$

hold in this range, cf. Fig. 1. The slope r^3 in the above upper bounds appears pessimistic and is chosen to be a good fit for low r.

Corollary 1 *Let $X, X_L : \Omega \to H$ be two random fields where H is the Hilbert space $\mathbb{R}$ or $W^{s,2}(D)$ and B^p is the Banach space $\mathbb{R}$ or $W^{s,p}(D)$ respectively. Furthermore let $r \geq 2$ and $X, X_L \in L^{2r}(\Omega, B^{2r})$. Then it holds that*

$$\|\mathbb{M}^r[X] - S^r_M[X_L]\|_2^2 \leq 2\|\mathbb{M}^r[X] - \mathbb{M}^r[X_L]\|_H^2 + c(\bar{X}_L, r)M^{-1}, \tag{67}$$

with

$$c(\bar{X}_L, r) = (r+1)^2 c_H^{2(r-1)} \|\bar{X}_L\|_{2r}^{2r}(1 + \varepsilon_e(M, r)),$$

where c_H is the Hölder constant from inequality (7) *and*

$$\varepsilon_e(M, r) = M^{-1}\frac{r^2}{2}(1 + \varepsilon_b(M, r))^2 + \varepsilon_v(M, r)$$

with $\varepsilon_e(M, r) \in \mathcal{O}(M^{-1})$. Furthermore we define

$$M_e(r) := \min\left\{M \in \mathbb{N} \,|\, \forall M' \geq M : \; \varepsilon_e(M', r) \leq 2\right\},$$

where we have the estimate

$$M_e(r) \le \max(M_b(r), M_v(r), 2r^2). \tag{68}$$

Proof Due to (15) we have

$$\begin{aligned}\|\mathbb{M}^r[X]-S^r_M[X_L]\|_2^2 &= \|\mathbb{M}^r[X] - \mathbb{E}[S^r_M[X_L]]\|_H^2 + \mathcal{V}(S^r_M[X_L]) \\ &\le 2\|\mathbb{M}^r[X] - \mathbb{M}^r[X_L]\|_H^2 + 2\|\mathbb{M}^r[X_L] - \mathbb{E}[S^r_M[X_L]]\|_H^2 + \mathcal{V}(S^r_M[X_L]).\end{aligned}$$

by applying the triangle inequality in the second step. Using Lemma 4 and Lemma 5 we gain

$$\begin{aligned}\|\mathbb{M}^r[X] - S^r_M[X_L]\|_2^2 \le& 2\|\mathbb{M}^r[X] - \mathbb{M}^r[X_L]\|_H^2 \\ &+ M^{-2}\frac{r^2(r+1)^2}{2} c_H^{2(r-1)} \|\bar{X}_L\|_{2r}^{2r}(1+\varepsilon_b(M,r))^2 \\ &+ M^{-1}(r+1)^2 c_H^{2(r-1)} \|\bar{X}_L\|_{2r}^{2r}(1+\varepsilon_v(M,r)).\end{aligned}$$

It remains to prove the estimate for $M_e(r)$. However the upper bound in (68) is chosen in such a way, that it follows directly by estimating $\varepsilon_b, \varepsilon_v \le 1$ and $M^{-1}2r^2 \le 1$. The proof is complete.

5 Multilevel Monte Carlo Approximation

Let X and $\{X_\ell\}_{\ell\ge 0}$ be random variables with values in the Hilbert space H being $\mathbb{R}$ or $W^{s,2}(D)$, and let B^p be the Banach space $\mathbb{R}$ or $W^{s,p}(D)$ respectively. For $\ell \to \infty$ we assume $X_\ell \to X$ in some sense specified later. Again we assume for some $r \ge 2$ $X, X_\ell \in L^{2r}(\Omega, B^{2r})$ to assure existence of the first r moments in H. We can write

$$\mathbb{M}^r[X_L] = \sum_{\ell=1}^{L} \Delta\mathbb{M}_\ell, \qquad \Delta\mathbb{M}_\ell := \mathbb{M}^r[X_\ell] - \mathbb{M}^r[X_{\ell-1}] \tag{69}$$

assuming $X_0 := 0$. To approximate the r-th central moment we define the Multilevel Monte Carlo estimator

$$\mathcal{S}^r_{\mathrm{ML}}[X] = \sum_{\ell=1}^{L} T_\ell, \qquad T_\ell := S^r_{M_\ell}[X_\ell] - S^r_{M_\ell}[X_{\ell-1}]. \tag{70}$$

for some sequence $\{M_\ell\}_{\ell=1}^L$. As mentioned in Section 3.2, the summands T_ℓ are built from M_ℓ pairs of samples $(X_\ell, X_{\ell-1})_i$, $i = 1, \dots, M_\ell$, both computed for the *same* realization of input parameters (the same random event $\omega_i \in \Omega$). Moreover, we assume that the level corrections on different levels T_j and T_ℓ, $j \ne \ell$ are built of independent realizations and therefore are statistically independent (this condition can be relaxed, see Remark 2 below).

Theorem 1 *Let X and $X_\ell : \Omega \to H$ be random fields with values in a Hilbert space H being either $\mathbb{R}$ or $W^{s,2}(D)$ and let the Banach space B^p be either $\mathbb{R}$ or $W^{s,p}(D)$ respectively. Furthermore let $r \ge 2$ and $1 \le p \le q(r-1) \le \infty$ be such that $\frac{1}{p} + \frac{1}{q} = 1$*

and $X \in L^{2r}(\Omega, B^{2r}), X_\ell \in L^{2q(r-1)}(\Omega, B^{2q(r-1)})$ *with uniformly bounded norms. Then it holds that*

$$\begin{aligned} \|\mathbb{M}^r[X] - \mathcal{S}^r_{\mathrm{ML}}[X]\|_2^2 \le \left(\left\|\mathbb{M}^r[X] - \mathbb{M}^r[X_L]\right\|_H + rQ_r \sum_{\ell=1}^{L} \frac{\|\bar{X}_\ell - \bar{X}_{\ell-1}\|_{2p}}{M_\ell} \right)^2 \\ +2Q_r^2 \sum_{\ell=1}^{L} \frac{\|\bar{X}_\ell - \bar{X}_{\ell-1}\|_{2p}^2}{M_\ell}, \end{aligned} \tag{71}$$

where $Q_r = r(r+1)c_H^{r-1} \max_\ell(\|\bar{X}_\ell\|_{2q(r-1)}^{r-1})$ *with* c_H *defined in* (7)*, and it holds that* $M_\ell \ge \max(M_b(r), M_v(r))$ *for all* $1 \le \ell \le L$.

Proof By (15) and the independence of the level corrections we have

$$\begin{aligned} \|\mathbb{M}^r[X] - \mathcal{S}^r_{\mathrm{ML}}[X]\|_2^2 =& \|\mathbb{M}^r[X] - \mathbb{E}[\mathcal{S}^r_{\mathrm{ML}}[X]]\|_H^2 + \mathcal{V}(\mathcal{S}^r_{\mathrm{ML}}[X]) \\ =& \|\mathbb{M}^r[X] - \mathbb{E}[\mathcal{S}^r_{\mathrm{ML}}[X]]\|_H^2 + \sum_{\ell=1}^{L} \mathcal{V}(S^r_{M_\ell}[X_\ell] - S^r_{M_\ell}[X_{\ell-1}]). \end{aligned}$$

Lemma 5 yields an upped bound for the variance being the last sum in (71). The triangle inequality implies

$$\|\mathbb{M}^r[X] - \mathbb{E}[\mathcal{S}^r_{\mathrm{ML}}[X]]\|_2 \le \|\mathbb{M}^r[X] - \mathbb{M}^r[X_L]\|_2^2 + \|\mathbb{M}^r[X_L] - \mathbb{E}[\mathcal{S}^r_{\mathrm{ML}}[X]]\|_H^2.$$

We claim that the last term admits the upper bound

$$\|\mathbb{M}^r[X_L] - \mathbb{E}[\mathcal{S}^r_{\mathrm{ML}}[X]]\|_H \le rQ_r \sum_{\ell=1}^{L} \frac{\|X_\ell - X_{\ell-1}\|_{2p}}{M_\ell}.$$

Indeed, this estimate follows directly from (69), the triangle inequality and Lemma 4. The proof is complete.

Notice that the first sum in (71) is caused by the statistical bias of the estimator (69) (cf. [4, (5.16)] where an unbiased estimator for the variance has been analyzed). It is dominated by the last sum (i.e. the variance of the estimator) when M_ℓ decays exponentially. Recall that for positive α_ℓ and $b > 1$ it holds that

$$\left(\sum_{\ell=1}^{L} \alpha_\ell\right)^2 \le \sum_{j\ell=1}^{L} \frac{1}{2}\left(b^{\ell-j}\alpha_j^2 + \frac{1}{b^{\ell-j}}\alpha_\ell^2\right) < \frac{b}{b-1} \sum_{\ell=1}^{L} b^{L-\ell}\alpha_\ell^2.$$

Thus, the condition $M_\ell \gtrsim b^{L-\ell}$ is sufficient for

$$\left(\sum_{\ell=1}^{L} \frac{\|\bar{X}_\ell - \bar{X}_{\ell-1}\|_{2p}}{M_\ell}\right)^2 \lesssim \sum_{\ell=1}^{L} b^{L-\ell} \frac{\|\bar{X}_\ell - \bar{X}_{\ell-1}\|_{2p}^2}{M_\ell^2} \lesssim \sum_{\ell=1}^{L} \frac{\|\bar{X}_\ell - \bar{X}_{\ell-1}\|_{2p}^2}{M_\ell}$$

which together with (71) implies

$$\|\mathbb{M}^r[X] - \mathcal{S}^r_{\mathrm{ML}}[X]\|_2^2 \le 2\left\|\mathbb{M}^r[X] - \mathbb{M}^r[X_L]\right\|_H^2 + C \sum_{\ell=1}^{L} \frac{\|\bar{X}_\ell - \bar{X}_{\ell-1}\|_{2p}^2}{M_\ell}, \tag{72}$$

cf. [4, (5.16)]. This estimate implies in particular the following result.

Theorem 2 *Suppose that assumptions of Theorem 1 hold true. Let* $\mathcal{C}_\ell = \mathrm{Work}(X_\ell^i)$ *be the cost of evaluation for a single sample of* X_ℓ *and* $\{N_\ell\}_{\ell=1}^\infty$ *be an exponentially increasing sequence of positive integers satisfying* $N_\ell/N_{\ell-1} \geq a$ *for some fixed* $a > 1$. *Moreover, suppose there are positive constants* $\alpha, \beta, \gamma > 0$ *such that*

$$1)\ \|\mathbb{M}^r[X] - \mathbb{M}^r[X_L]\|_H \lesssim N_\ell^{-\alpha}, \quad 2)\ \|\bar{X}_\ell - \bar{X}_{\ell-1}\|_{2p}^2 \lesssim N_\ell^{-\beta}, \quad 3)\ \mathcal{C}_\ell \lesssim N_\ell^{\gamma}. \quad (73)$$

Then for any $\varepsilon > 0$ *there exists an integer* L *and a sequence* $\{M_\ell\}_{\ell=1}^L$ *such that*

$$\|\mathbb{M}^r[X] - \mathcal{S}_{\mathrm{ML}}^r[X]\|_2 < \varepsilon \quad (74)$$

and

$$\mathrm{Work}(\mathcal{S}_{\mathrm{ML}}^r[X]) \lesssim \varepsilon^{-\frac{\gamma}{\alpha}} + \begin{cases} \varepsilon^{-2}, & \text{if } \beta > \gamma, \\ \varepsilon^{-2}\log(\varepsilon)^2, & \text{if } \beta = \gamma, \\ \varepsilon^{-2-\frac{\gamma-\beta}{\alpha}}, & \text{if } \beta < \gamma. \end{cases} \quad (75)$$

Proof The proof is an adapted version of the proof in [6] for the expectation value, see also [4, Theorem 3.2, Theorem 5.2]. We choose M_ℓ

$$M_\ell \sim \left\lceil N_\ell^{-\frac{\beta+\gamma}{2}} \begin{cases} N_L^{2\alpha}, & \text{if } \beta > \gamma, \\ LN_L^{2\alpha}, & \text{if } \beta = \gamma, \\ N_L^{2\alpha+\frac{\gamma-\beta}{2}}, & \text{if } \beta < \gamma \end{cases} \right\rceil \quad (76)$$

and L such that $N_L^{-\alpha} \sim \varepsilon$, i.e. $L \sim \log(\varepsilon)$. This implies in particularly

$$\frac{M_{\ell-1}}{M_\ell} \sim \left(\frac{N_\ell}{N_{\ell-1}}\right)^{\frac{\beta+\gamma}{2}} \geq a^{\frac{\beta+\gamma}{2}} =: b > 1.$$

This implies $M_\ell \gtrsim b^{L-\ell}$ and herewith (72). The estimate (72) and assumptions 1)–3) have a similar structure as in [6], see also [4, Theorem 3.2, Theorem 5.2]. Therefore, the estimates (74) and (75) follow with similar arguments.

Remark 2 An important part of the construction of the multilevel estimate (70) was the assumption that the level corrections T_j and T_ℓ on different levels $j \neq \ell$ are statistically independent. This assumption can be removed so that even the same samples can be used within T_ℓ and $T_{\ell+1}$. In this case, the orthogonality of the level corrections cannot be used anymore (cf. (15)). But the estimation still can be realized by means of the triangle inequality, see Section 3 and in particular (21). The main convergence result in Theorem 2, namely (74) and (75), still holds true for $\beta \neq \gamma$. The estimate

$$\mathrm{Work}(\mathcal{S}_{\mathrm{ML}}^r[X]) \lesssim \varepsilon^{-\frac{\gamma}{\alpha}} + \varepsilon^{-2}\log(\varepsilon)^3, \qquad \text{if } \beta = \gamma \quad (77)$$

holds true instead. To show it, we choose $M_\ell \sim N_\ell^{-\frac{\beta+\gamma}{2}} L^2 N_L^{2\alpha}$ for $\beta = \gamma$. The proof is straight forward and is left to the reader.

6 A random obstacle problem

We apply the abstract framework developed above to a class of obstacle problems with rough random obstacles. In this section we briefly introduce the mathematical framework (see [4, Sect. 6, 8] for further details) and present results of numerical experiments in Section 7.

Let $D \subset \mathbb{R}^d$ be a bounded convex domain, $\psi \in C(D)$, $\psi \leq 0$ on ∂D a continuous function and $f \in L^2(D)$. The deterministic obstacle problem can be formulated as finding $u : D \to \mathbb{R}$ such that

$$\begin{cases} -\Delta u \geq f \text{ in } D, \\ u \geq \psi \text{ in } D, \\ (-\Delta u - f)(u-\psi) = 0 \text{ in } D, \\ u = 0 \text{ on } \partial D, \end{cases} \tag{78}$$

where ψ is called obstacle and f volume force. Formulation (78) has a unique solution $u \in H_0^1(D)$, which can be written as $u = S(\psi, f)$ in the framework of (1). For simplicity we consider the case $d = 2$ in what follows. If ψ and/or f are random, i.e. $\psi(\omega)$ and $f(\omega)$, we gain due to (1) also a random solution $u(\omega)$

$$u(\omega) = S(\psi(\omega), f(\omega)). \tag{79}$$

Indeed for almost all $\omega \in \Omega$ (79) is fulfilled by the unique weak solution u [11] of

$$\forall v \in \mathcal{K} : \ \mathbb{E}[a(u, v-u)] \geq \mathbb{E}[L(v-u)], \tag{80}$$

where

$$a(u,v) = \int_D \nabla_x u \cdot \nabla_x v \, \mathrm{d}x, \qquad L(v) = \int_D f v \, \mathrm{d}x, \tag{81}$$

∇_x the gradient operator (in spatial coordinates only) and

$$\mathcal{K} = \{v \in L^2(\Omega, H_0^1(D)) \, : \, v \geq \psi \text{ for almost all } (x,\omega) \in D \times \Omega\}. \tag{82}$$

To approximate a sample of $u(\omega)$ numerically, we use a standard piecewise affine globally continuous finite element spaces

$$V_\ell := \left\{ v \in H_0^1(D) \, : \, v|_T \in \mathcal{P}^1(T), \, \forall T \in \mathcal{T}_\ell \right\}, \tag{83}$$

where $\mathcal{T}_\ell$ is a family of quasiuniform conforming triangulation with $\mathcal{T}_\ell$ a refinement of $\mathcal{T}_{\ell-1}$ and $\mathcal{T}_1$ is some start triangulation on D. We denote with $\mathcal{N}_\ell$ the set of interior notes of $\mathcal{T}_\ell$ and with $N_\ell := |\mathcal{N}_\ell| = \dim(V_\ell)$ the number of interior nodes. The discrete obstacle problem for a fixed sample ω can now be formulated as: Find $u_\ell(\omega) \in \mathcal{K}_\ell(\omega)$, s.t.

$$\forall v_\ell \in \mathcal{K}_\ell(\omega) : \ \ a(u_\ell(\omega), v_\ell - u_\ell(\omega)) \geq L_\omega(v_\ell - u_\ell(\omega)),$$

where

$$\mathcal{K}_\ell(\omega) := \{v_\ell \in V_\ell \,|\, \forall n \in \mathcal{N}_\ell : v_\ell(n) \geq \psi(n,\omega)\}$$

and

$$L_\omega(v) = \int_D f(\omega) v \, \mathrm{d}x.$$

The solution $u_\ell(\omega)$ is unique, cf. [11]. Moreover, according to [4] it holds that

$$\|u - u_\ell\|_{L^{2p}(\Omega, W^{1,2p}(D))} \lesssim h_\ell^{\frac{1}{p}} \left(\|f\|_{L^{2p}(\Omega, L^2(D))} + \|\psi\|_{L^{2p}(\Omega, H^2(D))}\right) \tag{84}$$

for $1 \leq p \leq \infty$ provided $f \in L^{2p}(\Omega, L^2(D))$ and $\psi \in L^{2p}(\Omega, H^2(D))$, see [4] for details.

Instead of only having a random obstacle and a random volume force, one might also have to model random material parameters. For such formulations and related convergence results see [9, 14].

In this paper we are particularly interested in the case of rough random obstacles representing e.g. the uneven structure of asphalt read surfaces. We utilize a rough obstacle model from [16], representing of the obstacle $\psi(x)$ as a Fourier series

$$\psi(x) = \sum_q B_q(H) \cos(q \cdot x + \phi_q), \tag{85}$$

where $x \in [0, L]^2$ for simplicity. The sum is over all $q \in \frac{2\pi}{L}\mathbb{Z}^2$, the amplitudes $B_q(H)$ depend on the frequency q and the so called Hurst coefficient $H \in [0, 1]$. The ϕ_q are independent random variables, uniformly distributed in $[0, 2\pi)$. Isotropic self-affine obstacles obey the law

$$B_q(H) \sim \begin{cases} |q|^{-(H+1)}, \; q_\ell \leq |q| \leq q_s \\ \quad 0, \qquad \text{otherwise.} \end{cases} \tag{86}$$

As obstacle for the numerical experiments in the following section we use the surface (85) with particular parameters

$$\begin{aligned} B_q(H) = \frac{\pi}{25}(2\pi \max(|q|, q_l))^{-(H+1)}, \qquad q_0 \leq |q| \leq q_s, \\ q_0 = 1, \quad q_l = 10, \quad q_s = 26, \end{aligned} \tag{87}$$

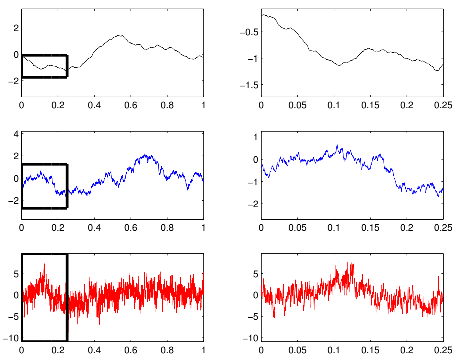


Fig. 2 Self affine surfaces $\psi(x)$ in 1d with $H = 1, 0.5$ and 0 (left column, from top to bottom). The right column shows the magnification of the box on the left.

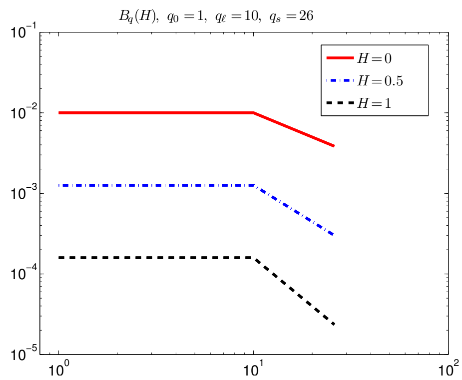


Fig. 3 The values of $B_q(H)$ for ψ defined in (87) in double logarithmic scale. H indicates the height of the plateau and the slope for $q_\ell \leq q \leq q_s$.

where the sum in (85) now runs over $\mathbb{Z}^2$ (see Figure 3). To gain a randomly rough obstacle we model H as a random variable as well as all phase shifts φ_q:

$$H \sim \mathcal{U}(0,1), \quad \varphi_q \sim \mathcal{U}(0, 2\pi), \quad q_0 \leq |q| \leq q_s. \tag{88}$$

Those random variables are assumed to be mutually independent. Two realizations of this obstacle are plotted in the next section (Figures 4 and 5). We refer to [16] and our previous work [4] for further details on this model.

7 Numerical Experiments

In this section we report on results of numerical experiments for the model obstacle problem described above with $D = [-1, 1]^2$, $f = -5$, $u|_{\partial D} = \frac{1}{2}$ and random obstacle parametrized by $\psi(x)$ as described in Section 6. In Fig. 4 and 5 we show two realizations of the obstacle profile and the corresponding solutions for the case of high and low roughness respectively. The computations involve the hierarchy of the finite element spaces V_ℓ defined in (83) with meshes $\mathcal{T}_\ell$. The coarsest triangulation $\mathcal{T}_{-1}$ consists of four congruent triangles sharing $(0, 0)$ as a vertex. Finer meshes $\mathcal{T}_{\ell+1}$ are defined recursively as the uniform red refinement of coarser meshes $\mathcal{T}_\ell$ by halving the edge of each element so that $h_\ell \sim N_\ell^{-\frac{1}{2}}$.

As a solver we implemented different variants of the Monotone Multigrid Method described in [12]; the *Multilevel Subset Decomposition Algorithm* appeared to be the best for our model problem. For this algorithm a log-linear cost has been proved, cf. [17], [14, Section 4.5]. In our experiments we observe almost linear complexity, see Fig. 8 indicating that $\gamma \approx 1$ in (73).

We mention that efficient updating of a single level estimator of higher moments is no trivial task. In our experiments approximation of higher order moments we use stable one-pass update formulae from [15], which are suitable for parallelization. Another technical difficulty emerges if the quantity of interest is a spatially varying function, for example $X_\ell = u_\ell$. Clearly, $u_\ell \in V_\ell$ implies

$$\mathcal{S}^r_{\mathrm{ML}}[u] \in V^r_\ell, \quad \text{where} \quad V^r_\ell := \left\{ v \in H^1_0(D) \,:\, v|_T \in \mathcal{P}^r(T), \, \forall T \in \mathcal{T}_\ell \right\}$$

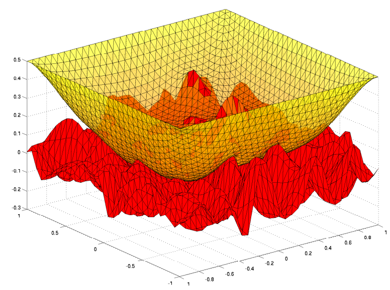


Fig. 4 A realization of the obstacle obstacle (red) and the corresponding solution (yellow) with $H = 0$.

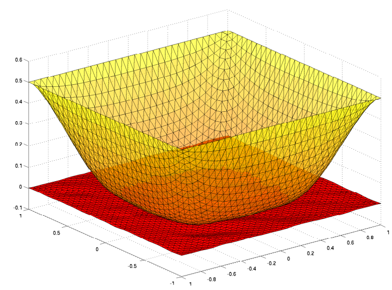


Fig. 5 A realization of the obstacle obstacle (red) and the corresponding solution (yellow) with $H = 1$.

is a space of polynomials of degree r. Notice, however, that V_ℓ^r is not required for solving the discrete formulation, but only for evaluating the estimator $\mathcal{S}^r_{\mathrm{ML}}[u]$ (cf. [4] for estimation of the variance). Therefore the computational cost associated with the use of the high order space V_ℓ^r is negligible.

Obviously, this issue does not appear when estimating higher order statistical moments of scalar quantities. In this paper we report on convergence results for the r-th central statistical moments of the size of the coincidence set

$$\Lambda(\omega) = \{x : u(x,\omega) = \psi(x,\omega)\},$$

which is a real number. Let $X(\omega) = |\Lambda(\omega)|$ and consider its approximation

$$X_\ell(\omega) := |D| \times \frac{\text{number of nodes } n \text{ of } \mathcal{T}_\ell \text{ s.t. } u_\ell(n,\omega) = \psi(n,\omega)}{\text{number of nodes of } \mathcal{T}_\ell}.$$

The authors are not aware of any analytical estimates for the parameters α and β from (73) for this quantity of interest. Therefore, to verify the convergence result from Theorem 2 we estimate α and β numerically. Observe, however, that $0 \le X(\omega) \le |D|$ and therefore Theorem 1 and Theorem 2 hold with $q = \infty$ and $p = 1$ for moments of arbitrary order $r \ge 1$. Since an analytic expression for the true contact area is not known, we approximate it by the fine grid estimate $X \approx X_9$ at the 9th refinement level.

In Fig. 6 we show convergence of the quantities $|\mathbb{E}[X] - \mathbb{E}[X_L]|$ and $|\mathbb{M}^r[X] - \mathbb{M}^r[X_L]|$ for $r = 2, \dots, 6$, which are parts of the bias of the corresponding MLMC estimators, that can't be bounded by their variance, cf. (71). The dashed line has the slope of the least square fit to the curve for the first moment $|\mathbb{E}[X] - \mathbb{E}[X_\ell]|$. The convergence curves for higher order moments appear to achieve the same slope after a pre-asymptotic plateau. This indicates that the parameter α can be roughly estimated by the slope of the dashed line, i.e. $\alpha \approx \frac{2}{3}$. In Figure 7 we show the variance of the level corrections $Y_\ell := X_\ell - X_{\ell-1}$. The slope of the dashed line comes from the least square fit in (the asymptotic regime) and indicates that $\beta \approx 1.1$. As mentioned before, $\gamma \approx 1$, as seen from Fig. 8. For such values of parameters α, β and γ Theorem 2 implies

$$\mathrm{MSE} := \|\mathbb{M}^r[X] - \mathcal{S}^r_{\mathrm{ML}}[X]\|^2_{L^2(\Omega)} < \epsilon^2 \quad \text{whereas} \quad \mathrm{Work}(\mathcal{S}^r_{\mathrm{ML}}[X]) \lesssim \epsilon^{-2}. \tag{89}$$

In order to check this result numerically we plot an approximate MSE with respect to the runtime t in Figure 9. For the first moment the figure shows the MSE of the approximation of the mean $\mathbb{E}[X]$ by the multilevel sample mean $E_{\mathrm{ML}}[X] = \sum_{\ell=1}^{L} E_{M_\ell}[X_\ell - X_{\ell-1}]$. Since $t \sim \mathrm{Work}(\mathcal{S}^r_{\mathrm{ML}}[X])$, this is an ultimate test of performance for the method. To approximate the $L^2(\Omega)$-norm we averaged the error over 30 runs. We observe that MSE curves for the first six moments achieve the same slope as the curve $\mathrm{MSE} = Ct^{-1}$, thereby confirming (89).

While the expectation value in Fig. 9 seems to enter the asymptotic regime nearly from the beginning, the higher order moments again converge slowly at low levels, but achieve the asymptotic rate later on. This effect can be explained by the orthogonal decomposition (15) of MSE into the (squared) bias and the variance

$$\mathrm{MSE} = |\mathbb{M}^r[X] - \mathbb{E}[\mathcal{S}_{\mathrm{ML}}[X]]|^2 + \mathcal{V}(\mathcal{S}_{\mathrm{ML}}[X]]). \tag{90}$$

The ratio of the (squared) bias and the variance is plotted on Fig. 10 for the first six statistical moments. We observe that the bias part dominates over the variance

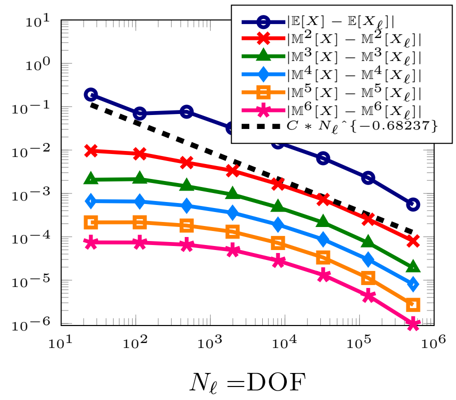


Fig. 6 Convergence of the bias part of the estimator for the first six moments.

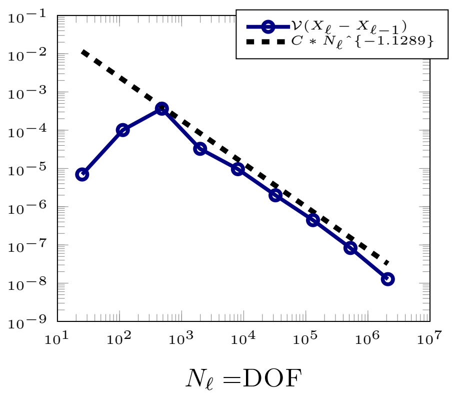


Fig. 7 Convergence of the variance of the level corrections (the variance part of the estimator)

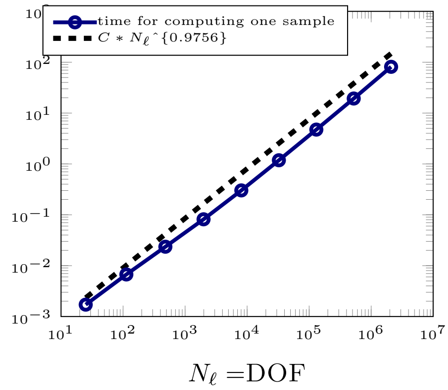


Fig. 8 Average time to compute one discrete sample with N_ℓ degrees of freedom.

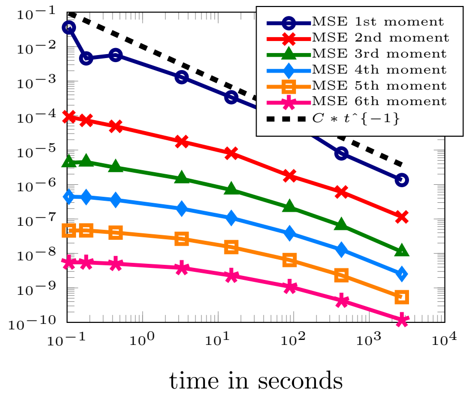


Fig. 9 Mean square error of the Multilevel Monte Carlo approximation of the first six moments versus the runtime.

at low levels. As the refinement progresses, the magnitudes of bias and variance come closer and achieve comparable values when $4 \leq L \leq 6$. For $L \geq 7$ the variance part is dominant. This transition phenomenon is quite natural due to the relation $2\alpha > \beta$ in our example and the choice $M_L = \mathcal{O}(1)$ in our computations. A better balancing of the bias and variance parts can be achieved by increasing M_L in accordance with (76).

In our computations we have utilized the adaptive choice of M_ℓ as proposed in [10] for estimation of the mean, namely $M_\ell \sim \sqrt{\frac{\mathcal{V}(X_\ell - X_{\ell-1})}{\mathcal{C}_\ell}}$, but using a fixed number of samples on the finest grid $M_L = \mathcal{O}(1)$, see also [4]. Using a fixed number of samples on the finest grid would not change the work-error relation in the case $\beta > \gamma$. This is also a reasonable choice for higher order moments, since we can choose the smallest possible Hölder parameter $p = 1$ in (72) and thus this upper bound leads to the same optimal choice of M_ℓ. This fact is observed in Figure 11,

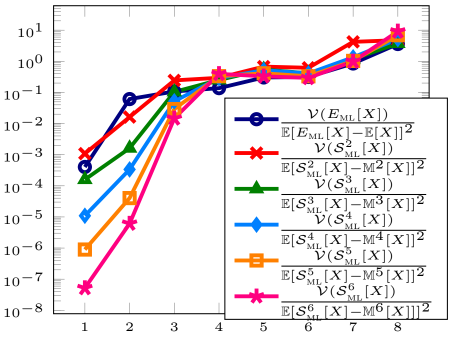


Fig. 10 Relation between the variance and the squared bias of the Multilevel Monte Carlo Estimators for the first six moments.

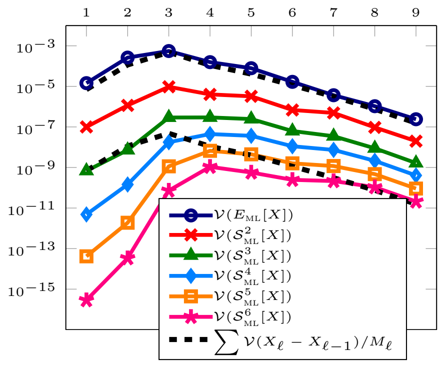


Fig. 11 The variance of the Multilevel Monte Carlo estimators for the first six moments. The slope of the dashed lines represents the decay of the aysmptotic upper bound.

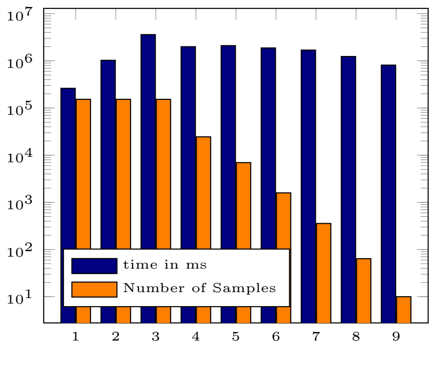


Fig. 12 Time (in ms) spent per discretization level and number of samples per discretization level.

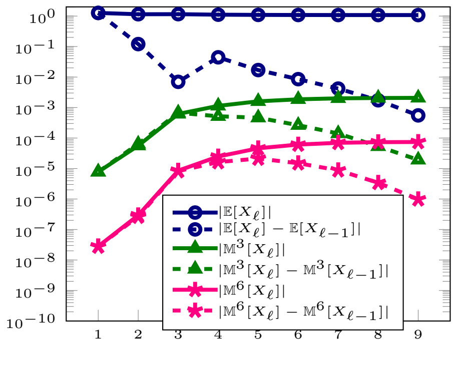


Fig. 13 Absolute value of the first, third and sixth moment and the absolute value of the level corrections.

where we see that the variance of the estimators for moments of the different order behave similar in the asymptotic regime.

Notice, however, that the value $\mathcal{V}(X_\ell - X_{\ell-1})$ is not a priori known and therefore should be estimated, as suggested in [10,6,4]. This estimation might suffer from a hidden difficulty: If the original algorithm accidentally estimates $\mathcal{V}(X_\ell - X_{\ell-1})$ by 0, it will never compute additional samples on this level. If in this case the true variance is nonzero $\mathcal{V}(X_\ell - X_{\ell-1}) \neq 0$ (this typically happens in our example at low levels), it would lead to an $\mathcal{O}(1)$ error, which would never be reduced by the original algorithm. We avoid this difficulty by demanding $M_\ell \geq M_{\ell+1}$ which leads to an additional computational cost. However, this cost is asymptotically negligible. In Fig. 12 we present the runtime and the number of samples per level. Observe that the same number of samples is used at the three coarsest levels, which is a consequence of the above mentioned condition. Starting from the third level we observe an almost uniform distribution of the runtime over the refinement levels whereas the number of samples M_ℓ decays exponentially with increasing level index.

Finally we show the convergence of two consecutive approximation of the first, third and sixth moment in Fig. 13 (other moments behave similarly and therefore are nor presented here). Again one observes a clear pre-asymptotic behavior for higher order moments. The preasymptotic region is larger for higher than for the lower order moments.

References

1. N. I. Akhiezer. *The classical moment problem and some related questions in analysis.* Translated by N. Kemmer. Hafner Publishing Co., New York, 1965.
2. A. Barth, C. Schwab, and N. Zollinger. Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients. *Numer. Math.*, 119(1):123–161, 2011.
3. C. Bayer, H. Hoel, E. von Schwerin, and R. Tempone. On non-asymptotic optimal stopping criteria in monte carlo simulations. 2012.
4. C. Bierig and A. Chernov. Convergence analysis of multilevel variance estimators in multilevel monte carlo methods and application for random obstacle problems. Preprint 1309, Institute for Numerical Simulation, University of Bonn, 2013. Submitted.
5. J. Charrier, R. Scheichl, and A. Teckentrup. Finite element error analysis of elliptic PDEs with random coefficients and its application to multilevel Monte Carlo methods. *SIAM Journal on Numerical Analysis*, 11, 2013.
6. 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(1):3–15, 2011.
7. R. B. D'Agostino, A. Belanger, and R. B. D'Agostino Jr. A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44(4):316–321, 1990.
8. Y. Dodge and V. Rousson. The complications of the fourth central moment. *Amer. Statist.*, 53(3):267–269, 1999.
9. R. Forster and R. Kornhuber. A polynomial chaos approach to stochastic variational inequalities. *J. Numer. Math.*, 18(4):235–255, 2010.
10. M. B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.
11. R. Glowinski. *Numerical methods for nonlinear variational problems.* Springer Series in Computational Physics. Springer-Verlag, New York, 1984.
12. C. Gräser and R. Kornhuber. Multigrid methods for obstacle problems. *J. Comput. Math.*, 27(1):1–44, 2009.
13. M. G. Kendall and A. Stuart. *The advanced theory of statistics. Vol. 6. Distribution theory.* Halsted Press, New York, 1994.
14. R. Kornhuber, C. Schwab, and M.-W. Wolf. Multi-Level Monte-Carlo Finite Element Methods for stochastic elliptic variational inequalities. Technical Report 2013-12, SAM, ETH Zürich, Switzerland, 2013.
15. P. Pébay. Formulas for robust, one-pass parallel computation of covariances and arbitrary-order statistical moments. *Sandia Report SAND2008-6212, Sandia National Laboraties*, (September), 2008.
16. B. N. J. Persson, O. Albohr, U. Tartaglino, A. I. Volokitin, and E. Tosatti. On the nature of surface roughness with application to contact mechanics, sealing, rubber friction and adhesion. *J. Phys.: Condens. Matter*, 17:R1–R62, 2005.
17. X.-C. Tai. Rate of convergence for some constraint decomposition methods for nonlinear variational inequalities. *Numer. Math.*, 93(4):755–786, 2003.
18. K. Yosida. *Functional analysis.* Die Grundlehren der Mathematischen Wissenschaften, Band 123. Academic Press Inc., New York, 1965.