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

Convergence analysis of multifidelity Monte Carlo estimation

Benjamin Peherstorfer · Max Gunzburger · Karen Willcox

Received: date / Accepted: date

Abstract The multifidelity Monte Carlo method provides a general framework for combining cheap low-fidelity approximations of an expensive high-fidelity model to accelerate the Monte Carlo estimation of statistics of the high-fidelity model output. In this work, we investigate the properties of multifidelity Monte Carlo estimation in the setting where a hierarchy of approximations can be constructed with known error and cost bounds. Our main result is a convergence analysis of multifidelity Monte Carlo estimation, for which we prove a bound on the costs of the multifidelity Monte Carlo estimator under assumptions on the error and cost bounds of the low-fidelity approximations. The assumptions that we make are typical in the setting of similar Monte Carlo techniques. Numerical experiments illustrate the derived bounds.

Keywords multifidelity · multilevel · hierarchical methods · Monte Carlo · surrogates · coarse-grid approximations · partial differential equations with random coefficients · uncertainty quantification

B. Peherstorfer
Department of Mechanical Engineering and Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI 53706
E-mail: peherstorfer@wisc.edu

M. Gunzburger
Department of Scientific Computing, Florida State University, 400 Dirac Science Library, Tallahassee FL 32306-4120
E-mail: mgunzburger@fsu.edu

K. Willcox
Department of Aeronautics & Astronautics, MIT, Cambridge, MA 02139
E-mail: kwillcox@mit.edu

1 Introduction

Inputs to systems are often modeled as random variables to account for the uncertainties in the inputs due to inaccuracies and incomplete knowledge. Given the input random variable and a model of the system of interest, an important task is to estimate statistics of the model output random variable.

Monte Carlo estimation is one popular approach to estimate statistics. Basic Monte Carlo estimation generates samples of the input random variable, discretizes the model and then solves the discretized model—the high-fidelity model—up to the required accuracy at these samples, and averages over the corresponding outputs to estimate statistics of the model output random variable. This basic Monte Carlo estimation often requires many samples, and consequently many approximations of the model outputs, which can become too costly if the high-fidelity model solves are expensive. We note that other techniques than Monte Carlo estimation are available to estimate statistics of model outputs, see, e.g., [1,33,21,15,14,47,43,45].

Several variance reduction techniques have been presented to reduce the costs of Monte Carlo simulation compared to basic Monte Carlo estimators, e.g., antithetic variates [39,23,28] and importance sampling [39,27,36]. Our focus here is on the control variate framework that exploits the correlation between the model output random variable and an auxiliary random variable that is cheap to sample [30]. A major class of control variate methods derives the auxiliary random variable from cheap approximations of the outputs of the high-fidelity model. For example, in situations where the model is governed by (often elliptic) partial differential equations (PDEs), coarse-grid approximations of the PDE—low-fidelity models—can provide cheap approximations of the outputs obtained from a fine-grid high-fidelity discretization of the PDE; however, other types of low-fidelity models are possible in the context of PDEs, e.g., projection-based reduced models [41,40,20,3,37], data-fit interpolation and regression models [13,12], machine-learning-based models such as support vector machines [46,11], and other simplified models [29,32].

The multifidelity Monte Carlo (MFMC) method [38] uses a control variate approach to combine auxiliary random variables stemming from low-fidelity models into an estimator of the statistics of the high-fidelity model output. Key to the MFMC approach is the selection of how often each of the auxiliary random variables is sampled, and therefore how often each of the low-fidelity models is solved. The MFMC approach derives this selection from the correlation coefficients between the auxiliary random variables and the high-fidelity model output random variable. The selection of the MFMC approach is optimal in the sense that the variance of the MFMC estimator is minimized for given maximal costs of the estimation. We refer to the discussions in [38,31] for details on MFMC.

The work [38] discusses the properties of MFMC estimation in a setting where only mild assumptions on the high- and low-fidelity models are made. We consider here the setting where we can make further assumptions on the errors and costs of outputs obtained with a hierarchy of low- and high-fidelity models. Our contribution is to show that for an MFMC estimator with mean-squared error (MSE) below a threshold parameter $\epsilon > 0$, the costs of the estimation can be bounded by ϵ^{-1} up to a constant under certain conditions on the error and cost bounds of the models in the hierarchy.

We discuss that the conditions we require in the MFMC context are similar to the conditions exploited by the multilevel Monte Carlo method [9, Theorem 1]. Our analysis shows that MFMC estimation is as efficient in terms of error and costs as multilevel Monte Carlo estimation under certain conditions that we discuss below in detail. Multilevel Monte Carlo uses a hierarchy of low-fidelity models—typically coarse-grid approximations—to derive a hierarchy of auxiliary random variables, which are combined in a judicious way to reduce the runtime of Monte Carlo simulation. Multilevel Monte Carlo was introduced in [26] and extended and made popular by the work [18]. Since then, the properties of the multilevel Monte Carlo estimators have been studied extensively in different settings, see, e.g., [9,8,2,6,42]. Multilevel Monte Carlo and its variants have also been applied to density estimation [5], variance estimation [4], and rare event simulation [44]. We also mention the continuation multilevel Monte Carlo [10] and the extension multi-index Monte Carlo that allows different mesh widths in the dimensions [22]. In [34,35], a fault-tolerant multilevel Monte Carlo is introduced and analyzed, which is well suited for massively parallel computations. An integer optimization problem is solved to determine the optimal number of model evaluations depending on the rate of compute-node failures. The fault-tolerant approach thus takes into account node failure by adapting the number of model evaluations accordingly. The relationship between multilevel Monte Carlo and sparse grid quadrature [7, 16,17] is discussed in [24,25,19].

The outline of the presentation is as follows. Section 2 introduces the problem setup and basic, multilevel, and multifidelity Monte Carlo estimators. Section 3 derives the new convergence analysis of MFMC estimation. Numerical examples in Section 4 illustrate the derived bounds. Conclusions are drawn in Section 5.

2 Problem setup

This section introduces the problem setup and the various types of Monte Carlo estimators required throughout the presentation. Section 2.1 introduces the notation and Section 2.2 the basic Monte Carlo estimator. Multilevel Monte Carlo and the MFMC estimation are summarized in Section 2.3 and Section 2.4, respectively.

2.1 Preliminaries

The set of positive real numbers is denoted as $\mathbb{R}_+ = \{x \in \mathbb{R} : x > 0\}$. For two positive quantities a and b, we define $a \lesssim b$ to hold if a/b is bounded by a constant whose value is independent of any parameters on which a and b depend on.

Let $d \in \mathbb{N}$ be the dimension and define the Lipschitz domain $\mathcal{D} \subset \mathbb{R}^d$. Let $Z : \Omega \to \mathcal{D}$ be a random variable over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω denotes the set of outcomes, $\mathcal{F}$ the σ-algebra of events, and $\mathbb{P} : \mathcal{F} \to [0, 1]$ a probability measure. Let further $Q : \mathcal{D} \to \mathbb{R}$ be a function in a suitable function space and let $Q_\ell : \mathcal{D} \to \mathbb{R}$ be functions for $\ell \in \mathbb{N}$ that approximate Q in the sense of the following assumption. Note that we assume that $Q(Z)$ and $Q_\ell(Z)$ are integrable.

Assumption 1 *There exists $1 < s \in \mathbb{R}$ and rate $\alpha \in \mathbb{R}_+$ such that*

$$|\mathbb{E}[Q(Z) - Q_\ell(Z)]| \leq \kappa_1 s^{-\alpha \ell}\,, \qquad \ell \in \mathbb{N}\,,$$

where $\kappa_1 \in \mathbb{R}_+$ is a constant independent of ℓ.

The parameter $\ell \in \mathbb{N}$ is the level of Q_ℓ. Let further $w_\ell \in \mathbb{R}_+$ be the costs of evaluating Q_ℓ for $\ell \in \mathbb{N}$. The following assumption gives a bound on the costs with respect to the level ℓ.

Assumption 2 *There exists a rate $\gamma \in \mathbb{R}_+$ with*

$$w_\ell \leq \kappa_3 s^{\gamma \ell}\,,$$

where the constant s is given by Assumption 1 and $\kappa_3 \in \mathbb{R}_+$ is a constant independent of ℓ.

Note that in Assumption 2 the same constant s as in Assumption 1 is used.

The variance $\mathrm{Var}[Q_\ell(Z)]$ of the random variable $Q_\ell(Z)$ is denoted as

$$\sigma_\ell^2 = \mathrm{Var}[Q_\ell(Z)]\,, \qquad \ell \in \mathbb{N}\,.$$

We make the assumption that there exists a positive lower and upper bound for the variance σ_ℓ^2 with respect to level $\ell \in \mathbb{N}$.

Assumption 3 *There exist $\sigma_{low} \in \mathbb{R}_+$ and $\sigma_{up} \in \mathbb{R}_+$ such that $\sigma_{low} \leq \sigma_\ell \leq \sigma_{up}$ for $\ell \in \mathbb{N}$.*

The Pearson product-moment correlation coefficient of the random variables $Q_\ell(Z)$ and $Q_l(Z)$ is denoted as

$$\rho_{\ell,l} = \frac{\mathrm{Cov}[Q_\ell(Z), Q_l(Z)]}{\sigma_\ell \sigma_l}\,, \qquad \ell, l \in \mathbb{N}\,, \tag{1}$$

where $\mathrm{Cov}[Q_\ell(Z), Q_l(Z)]$ is the covariance of $Q_\ell(Z)$ and $Q_l(Z)$.

We consider the situation where the random variable Z represents an input random variable and Q is a function that maps an input, i.e., a realization of Z, onto an output. In our situation, evaluating Q entails solving a PDE ("model"), but the solutions to the PDE are unavailable. We therefore revert to solving an approximate PDE ("discretized model"), where the approximation (e.g., the mesh width) is controlled by the level ℓ. The functions Q_ℓ map the input onto the output obtained by solving the approximate PDE on level ℓ. Assumption 1 specifies in which sense Q_ℓ converges to Q with $\ell \to \infty$. Solving the approximate PDE on level ℓ incurs costs w_ℓ. One task in this context is to derive estimators of $\mathbb{E}[Q(Z)]$ using the functions Q_ℓ. We assess the efficiency of an estimator $\widehat{Q}$ with its MSE

$$e(\widehat{Q}) = \mathbb{E}\left[\left(\widehat{Q} - \mathbb{E}[Q(Z)]\right)^2\right] ,$$

and its costs $c(\widehat{Q})$, which are the sum of the evaluation costs w_ℓ of the functions Q_ℓ used in the estimator $\widehat{Q}$. An estimator $\widehat{Q}$ with MSE $e(\widehat{Q}) \lesssim \epsilon$ below a threshold $\epsilon \in \mathbb{R}_+$ is efficient, if the costs $c(\widehat{Q}) \lesssim \epsilon^{-1}$ are bounded by ϵ^{-1} up to a constant. Note that ϵ bounds the MSE, in contrast to the root-mean-squared error (RMSE) as in, e.g., [9].

2.2 Basic Monte Carlo estimation

Let $\ell \in \mathbb{N}$ and define the basic Monte Carlo estimator $\widehat{Q}^{\mathrm{MC}}_{\ell,m}$ of $\mathbb{E}[Q_\ell(Z)]$ as

$$\widehat{Q}^{\mathrm{MC}}_{\ell,m} = \frac{1}{m}\sum_{i=1}^{m} Q_\ell(Z_i) ,$$

with $m \in \mathbb{N}$ independent and identically distributed (i.i.d.) samples $Z_1, \dots, Z_m$ of Z. The MSE of the Monte Carlo estimator $\widehat{Q}^{\mathrm{MC}}_{\ell,m}$ with respect to $\mathbb{E}[Q(Z)]$ is

$$e(\widehat{Q}^{\mathrm{MC}}_{\ell,m}) = m^{-1}\operatorname{Var}[Q_\ell(Z)] + \left(\mathbb{E}[Q(Z) - Q_\ell(Z)]\right)^2 . \tag{2}$$

The term $m^{-1}\operatorname{Var}[Q_\ell(Z)]$ is the variance term and term $(\mathbb{E}[Q_\ell(Z) - Q(Z)])^2$ is the bias term. The costs of the estimator $\widehat{Q}^{\mathrm{MC}}_{\ell,m}$ are

$$c(\widehat{Q}^{\mathrm{MC}}_{\ell,m}) = m w_\ell ,$$

because Q_ℓ is evaluated at m samples, with one evaluation having costs w_ℓ.

Let now $\epsilon \in \mathbb{R}_+$ be a threshold. One approach to obtain a basic Monte Carlo estimator $\widehat{Q}^{\mathrm{MC}}_{\ell,m}$ with $e(\widehat{Q}^{\mathrm{MC}}_{\ell,m}) \lesssim \epsilon$ is to derive a maximal level $L \in \mathbb{N}$ and a number of samples m such that the bias and the variance term are bounded by $\epsilon/2$ up to constants. Consider first the choice of the maximal level $L \in \mathbb{N}$. With Assumption 1, the maximal level L is given by

$$L = \left\lceil \alpha^{-1}\log_s\left(\sqrt{2}\kappa_1\epsilon^{-1/2}\right)\right\rceil , \tag{3}$$

where κ_1 is the constant in Assumption 1. Note that the maximal level L defines the high-fidelity model Q_L in the terminology of the introduction, see Section 1.

To achieve that the variance term is bounded by $\epsilon/2$ up to a constant, the number of samples m is selected such that $\epsilon^{-1} \lesssim m$. With Assumption 2, and assuming the variance σ_ℓ^2 is approximately constant with respect to the level ℓ, the costs of the basic Monte Carlo estimator $\widehat{Q}_{L,m}^{\mathrm{MC}}$ are

$$c(\widehat{Q}_{L,m}^{\mathrm{MC}}) \lesssim \epsilon^{-1-\gamma/(2\alpha)}\,,$$

see [9, Section 2.1] for a proof. The costs of the basic Monte Carlo estimator scale with the rates γ and α.

2.3 Multilevel Monte Carlo estimation

We follow [9] for the presentation of the multilevel Monte Carlo estimation. Consider the threshold $\epsilon \in \mathbb{R}_+$ and define the maximal level $L \in \mathbb{N}$ as in (3). Multilevel Monte Carlo exploits the linearity of the expected value to write

$$\mathbb{E}[Q_L(Z)] = \mathbb{E}[Q_1(Z)] + \sum_{\ell=2}^{L} \mathbb{E}[Q_\ell(Z) - Q_{\ell-1}(Z)] = \sum_{\ell=1}^{L} \mathbb{E}[\Delta_\ell(Z)]\,,$$

where $\Delta_\ell(Z) = Q_\ell(Z) - Q_{\ell-1}(Z)$ for $\ell > 1$ and $\Delta_1(Z) = Q_1(Z)$. The basic Monte Carlo estimator of $\Delta_\ell(Z)$ with $m_\ell \in \mathbb{N}$ samples $Z_1, \dots, Z_{m_\ell}$ is

$$\widehat{\Delta}_{\ell,m_\ell}^{\mathrm{MC}} = \frac{1}{m_\ell} \sum_{i=1}^{m_\ell} Q_\ell(Z_i) - Q_{\ell-1}(Z_i)\,.$$

The multilevel Monte Carlo estimator $\widehat{Q}_{L,\boldsymbol{m}}^{\mathrm{ML}}$ is then given by

$$\widehat{Q}_{L,\boldsymbol{m}}^{\mathrm{ML}} = \sum_{\ell=1}^{L} \widehat{\Delta}_{\ell,m_\ell}^{\mathrm{MC}}\,, \tag{4}$$

where the vector $\boldsymbol{m} = [m_1, \dots, m_L]^T \in \mathbb{N}^L$ is the vector of the number of samples at each level. Note that each basic Monte Carlo estimator $\widehat{\Delta}_{\ell,m_\ell}^{\mathrm{MC}}$ in (4) uses a separate, independent set of samples. Note further that the functions $Q_1, \dots, Q_{L-1}$ are low-fidelity models in the terminology of the introduction, see Section 1.

Under the following two assumptions, and with a judicious choice of the number of samples $\boldsymbol{m}$, the multilevel Monte Carlo estimator is efficient, which means that the estimator $\widehat{Q}_{L,\boldsymbol{m}}^{\mathrm{ML}}$ achieves an MSE of $e(\widehat{Q}_{L,\boldsymbol{m}}^{\mathrm{ML}}) \lesssim \epsilon$ with costs $c(\widehat{Q}_{L,\boldsymbol{m}}^{\mathrm{ML}}) \lesssim \epsilon^{-1}$. The first assumption states that the variance of Δ_ℓ decays with the level ℓ.

Assumption 4 *There exists a rate $\beta \in \mathbb{R}_+$ with*

$$\mathrm{Var}[Q_\ell(Z) - Q_{\ell-1}(Z)] \leq \kappa_2 s^{-\beta\ell}\,, \qquad \ell \in \mathbb{N}\,,$$

where s is the constant of Assumption 1 and $\kappa_2 \in \mathbb{R}_+$ is a constant independent of ℓ.

The following assumption sets the rate β of the decay of the variance $\mathrm{Var}[Q_\ell(Z)-Q_{\ell-1}(Z)]$ in relation to the rate γ of the increase of the costs with level ℓ.

Assumption 5 *For the rates γ of Assumption 2 and β of Assumption 4, we have $\beta > \gamma$.*

Set the number of samples $\boldsymbol{m}^{\mathrm{ML}} = [m_1^{\mathrm{ML}}, \dots, m_L^{\mathrm{ML}}]^T$ to

$$m_\ell^{\mathrm{ML}} = \left\lceil 2\epsilon^{-1}\kappa_2 \left(1 - s^{-(\beta-\gamma)/2}\right)^{-1} s^{-(\beta+\gamma)\ell/2} \right\rceil\,, \qquad \ell = 1, \dots, L\,, \tag{5}$$

where κ_2 is the constant in Assumption 4 and s is defined as in Assumption 1. Note that the components of $\boldsymbol{m}^{\mathrm{ML}}$ are rounded up. It is shown in [9] that if Assumptions 1–5 hold, then the multilevel Monte Carlo estimator $\widehat{Q}^{\mathrm{ML}}_{L,\boldsymbol{m}^{\mathrm{ML}}}$ with $\boldsymbol{m}^{\mathrm{ML}}$ defined in (5) achieves an MSE of $e(\widehat{Q}^{\mathrm{ML}}_{L,\boldsymbol{m}^{\mathrm{ML}}}) \lesssim \epsilon$ with costs $c(\widehat{Q}^{\mathrm{ML}}_{L,\boldsymbol{m}^{\mathrm{ML}}}) \lesssim \epsilon^{-1}$. Note that under Assumptions 1–5 it is sufficient to select the number of samples with the rates β and γ to achieve an efficient estimator. We refer to [26,18,9] for details on multilevel Monte Carlo estimation.

2.4 Multifidelity Monte Carlo estimation

The MFMC estimator [38] uses functions $Q_1, \dots, Q_L$ up to the maximal level L to derive an estimate of $\mathbb{E}[Q(Z)]$, similarly to the multilevel Monte Carlo estimator; however, the functions $Q_1, \dots, Q_L$ are combined in a different way than in the multilevel Monte Carlo estimator, and the number of samples $\boldsymbol{m}$ are selected by directly using correlation coefficients and costs instead of rates.

MFMC imposes on the number of samples $\boldsymbol{m} = [m_1, \dots, m_L]^T$ that $m_1 \geq m_2 \geq \cdots \geq m_L > 0$. Let

$$Z_1, \dots, Z_{m_1} \in \mathcal{D} \tag{6}$$

be m_1 i.i.d. samples of the random variable Z. Let further

$$Q_\ell(Z_1), \dots, Q_\ell(Z_{m_\ell})\,, \tag{7}$$

be the evaluations of Q_ℓ at the first m_ℓ samples $Z_1, \dots, Z_{m_\ell}$, for $\ell = 1, \dots, L$. Consider now the basic Monte Carlo estimators

$$\widehat{Q}^{\mathrm{MC}}_{\ell,m_\ell} = \frac{1}{m_\ell}\sum_{i=1}^{m_\ell} Q_\ell(Z_i)\,, \qquad \ell = 1, \dots, L\,, \tag{8}$$

and

$$\widehat{Q}^{\mathrm{MC}}_{\ell,m_{\ell+1}} = \frac{1}{m_{\ell+1}}\sum_{i=1}^{m_{\ell+1}} Q_\ell(Z_i)\,, \qquad \ell = 1, \dots, L-1\,, \tag{9}$$

which use the samples (6) and the evaluations (7). Note that the estimators in (9) use the first $m_{\ell+1}$ samples of the samples (6). Thus, the estimators $\widehat{Q}^{\mathrm{MC}}_{\ell,m_\ell}$

and $\widehat{Q}^{\mathrm{MC}}_{\ell,m_{\ell+1}}$ are dependent for $\ell = 1, \ldots, L-1$. The MFMC estimator $\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}}$ is defined as

$$\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}} = \widehat{Q}^{\mathrm{MC}}_{L,m_L} + \sum_{\ell=1}^{L-1} a_\ell \left(\widehat{Q}^{\mathrm{MC}}_{\ell,m_\ell} - \widehat{Q}^{\mathrm{MC}}_{\ell,m_{\ell+1}} \right) ,$$

where $\boldsymbol{a} = [a_1, \ldots, a_{L-1}]^T \in \mathbb{R}^{L-1}$ are coefficients. The costs of the MFMC estimator $\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}}$ are

$$c(\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}}) = \boldsymbol{w}^T \boldsymbol{m} \,,$$

where $\boldsymbol{w} = [w_1, \ldots, w_L]^T$, see [38].

The MFMC method provides a framework to select the number of samples $\boldsymbol{m}$ and the coefficients $\boldsymbol{a}$ such that the variance $\mathrm{Var}[\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}}]$ of the MFMC estimator $\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}}$ with costs $c(\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}}) = p$ is minimized for a given computational budget $p \in \mathbb{R}_+$. The number of samples $\boldsymbol{m}$ and the coefficients $\boldsymbol{a}$ are derived under two assumptions on the correlation coefficients of $Q_1(Z), \ldots, Q_L(Z)$ and the costs $w_1, \ldots, w_L$. The first assumption specifies the ordering of the functions $Q_1(Z), \ldots, Q_L(Z)$.

Assumption 6 *The random variables $Q_1(Z), \ldots, Q_L(Z)$ are ordered ascending with respect to the absolute values of the correlation coefficients*

$$|\rho_{L,1}| < |\rho_{L,2}| < \cdots < |\rho_{L,L}| \,.$$

The second assumption describes inequalities of the correlation coefficients and the costs.

Assumption 7 *The costs $w_1, \ldots, w_L$ and correlation coefficients $\rho_{L,1}, \ldots, \rho_{L,L}$ satisfy*

$$\frac{w_{\ell+1}}{w_\ell} > \frac{\rho^2_{L,\ell+1} - \rho^2_{L,\ell}}{\rho^2_{L,\ell} - \rho^2_{L,\ell-1}}$$

for $\ell = 1, \ldots, L-1$.

Assumption 7 enforces that the cost savings associated with a model justify its decrease in accuracy (measured by correlation) relative to other models in the hierarchy. If a particular model violates the condition in Assumption 7, the MFMC method omits the model from the hierarchy. See [38] for more details.

Under Assumptions 6–7, the number of samples $\boldsymbol{m}$ and the coefficients $\boldsymbol{a}$, which minimize the variance of $\mathrm{Var}[\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}}]$ with costs $c(\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}}) = p$, are given as follows [38]. The coefficients $\boldsymbol{a}^{\mathrm{MF}} = [a_1^{\mathrm{MF}}, \ldots, a_{L-1}^{\mathrm{MF}}]^T$ are set to

$$a_\ell^{\mathrm{MF}} = \frac{\rho_{L,\ell}\sigma_L}{\sigma_\ell} \,, \qquad \ell = 1, \ldots, L-1 \,,$$

and the number of samples $\boldsymbol{m}^{\mathrm{MF}} = [m_1^{\mathrm{MF}}, \ldots, m_L^{\mathrm{MF}}]^T$ is set to

$$m_\ell^{\mathrm{MF}} = m_L^{\mathrm{MF}} r_\ell \,, \qquad \ell = 1, \ldots, L \,,$$

where

$$r_\ell = \sqrt{\frac{w_L(\rho_{L,\ell}^2 - \rho_{L,\ell-1}^2)}{w_\ell(1-\rho_{L,L-1}^2)}}\,, \qquad \ell = 1, \dots, L\,, \tag{10}$$

with $\rho_{L,0} = 0$. Note that the selection of $\boldsymbol{m}^{\text{MF}}$ and $\boldsymbol{a}^{\text{MF}}$ is independent of the rates α, β, γ, which means the approach is applicable also in situations where rates capture the behavior of the properties of the functions $Q_1, \dots, Q_L$ only poorly, see, e.g., [38] for examples. Note further that the components of the number of samples $\boldsymbol{m}^{\text{MF}}$ are rounded up to integer numbers as in the multilevel Monte Carlo method, see (5) in Section 2.3. We note that in [34] an integer optimization problem is solved to adapt the number of model evaluations in multilevel Monte Carlo for an increased processor-failure tolerance on massively-parallel compute platforms.

The MFMC estimator is unbiased with respect to $\mathbb{E}[Q_L(Z)]$, see [38, Lemma 3.1]. The variance of the MFMC estimator $\widehat{Q}^{\text{MF}}_{L,\boldsymbol{m}^{\text{MF}}}$ is [38]

$$\operatorname{Var}(\widehat{Q}^{\text{MF}}_{L,\boldsymbol{m}^{\text{MF}}}) = \frac{\sigma_L^2(1-\rho_{L,L-1}^2)}{\left(m_L^{\text{MF}}\right)^2 w_L} p\,.$$

The work [38] investigates the costs and the MSE of the MFMC estimator only in the context of Assumption 6 and Assumption 7, and does not give insights into the behavior of the MFMC estimator if additionally Assumptions 1–5 are made.

3 New properties of the multifidelity Monte Carlo estimator

We now discuss the error and costs behavior of the MFMC estimator in a typical setting of the multilevel Monte Carlo estimators where Assumption 4 on the rate of the variance decay and Assumption 5 on the relative costs hold. Our main result is Theorem 1 that states that the MFMC estimator is efficient under Assumptions 1–7, which means that the MFMC estimator achieves an MSE $e(\widehat{Q}^{\text{MF}}_{L,\boldsymbol{m}^{\text{MF}}}) \lesssim \epsilon$ with costs $c(\widehat{Q}^{\text{MF}}_{L,\boldsymbol{m}^{\text{MF}}}) \lesssim \epsilon^{-1}$, independent of the rates α and γ. We first state Theorem 1 and then prove two lemmata in Section 3.1 and provide the proof of Theorem 1 in Section 3.2. Corollary 1 discusses the convergence rates of MFMC if Assumption 5 is violated.

Theorem 1 *With Assumptions 1–5, as well as Assumption 6 and Assumption 7, set the maximum level L as in* (3) *and set the budget p to*

$$p = \kappa_4 \epsilon^{-1}\,, \tag{11}$$

with the constant

$$\kappa_4 = 2\frac{\sigma_{up}^2}{\sigma_{low}^2}\left(\frac{s^{\frac{\gamma-\beta}{2}}}{1 - s^{\frac{\gamma-\beta}{2}}}\right)^2 .$$

For the number of samples $\boldsymbol{m}^{MF}$ and the coefficients $\boldsymbol{a}^{MF} \in \mathbb{R}^{L-1}$ defined in Section 2.4, the MSE $e(\widehat{Q}^{MF}_{L,\boldsymbol{m}^{MF}})$ of the MFMC estimator with respect to the statistics $\mathbb{E}[Q(Z)]$ is bounded as

$$e(\widehat{Q}^{MF}_{L,\boldsymbol{m}^{MF}}) \lesssim \epsilon \,,$$

and the costs are bounded as $c(\widehat{Q}^{MF}_{L,\boldsymbol{m}^{MF}}) \lesssim \epsilon^{-1}$.

Note that the MLMC theory developed in [9, Theorem 1] and [18, Theorem 3.1] requires an additional assumption on the rate α because the rounding up of the numbers of samples to an integer is explicitly taken into account, see also [4, Theorem 3.2]. We ignore the rounding here and therefore can avoid that assumption; however, we emphasize that we expect that a similar assumption is necessary for MFMC as well if the rounding of the numbers of samples is taken into account explicitly.

3.1 Preliminary lemmata

This section proves two lemmata that we use in the proof of Theorem 1 in Section 3.2.

Lemma 1 *Let $L \in \mathbb{N}$ be the maximal level. From Assumption 4, it follows that*

$$\mathrm{Var}[Q_L(Z) - Q_{\ell-1}(Z)] \lesssim s^{-\beta\ell}\,, \tag{12}$$

for $\ell = 2, \dots, L-1$.

Proof Let κ_2 be the constant in Assumption 4 so that we have

$$\mathrm{Var}[Q_\ell(Z) - Q_{\ell-1}(Z)] \leq \kappa_2 s^{-\beta\ell}\,,$$

for $\ell \in \mathbb{N}$. We obtain

$$\begin{aligned}\mathrm{Var}[Q_{\ell+1}(Z) - Q_{\ell-1}(Z)] \leq{}& \mathrm{Var}[Q_{\ell+1}(Z) - Q_\ell(Z)] + \mathrm{Var}[Q_\ell(Z) - Q_{\ell-1}(Z)]\\ &+ 2|\,\mathrm{Cov}[Q_{\ell+1}(Z) - Q_\ell(Z), Q_\ell(Z) - Q_{\ell-1}(Z)]|\,.\end{aligned} \tag{13}$$

With Assumption 4 and the Cauchy-Schwarz inequality, it follows that

$$\begin{aligned}\mathrm{Var}[Q_{\ell+1}(Z) - Q_{\ell-1}(Z)] \leq{}& \kappa_2 s^{-\beta(\ell+1)} + \kappa_2 s^{-\beta\ell}\\ &+ 2\sqrt{\mathrm{Var}[Q_{\ell+1}(Z) - Q_\ell(Z)]\,\mathrm{Var}[Q_\ell(Z) - Q_{\ell-1}(Z)]}\,,\end{aligned}$$

and therefore we have

$$\begin{aligned}\mathrm{Var}[Q_{\ell+1}(Z) - Q_{\ell-1}(Z)] \leq{}& \kappa_2 s^{-\beta(\ell+1)} + \kappa_2 s^{-\beta\ell} + 2\kappa_2 s^{-\beta(2\ell+1)/2}\\ \leq{}& \kappa_2 s^{-\beta\ell}(s^{-\beta} + 1 + 2s^{-\beta/2})\\ \leq{}& \kappa_2 s^{-\beta\ell}(s^{-\beta/2} + 1 + 2s^{-\beta/2})\,,\end{aligned} \tag{14}$$

where the last inequality holds because $s > 1$. Define now the sequence (b_j) with

$$b_0 = 1\,, \qquad b_j = s^{-\beta j/2} + b_{j-1}(1 + 2s^{-\beta j/2})\,, \qquad j \in \mathbb{N}\,.$$

From (14) and from the definition of the sequence (b_j), it follows with induction that

$$\begin{aligned}
\mathrm{Var}[Q_{\ell+j}(Z) - Q_{\ell-1}(Z)] \leq& \kappa_2 s^{-\beta(\ell+j)} + \kappa_2 b_{j-1} s^{-\beta\ell} + 2\kappa_2 s^{-\beta\ell}(b_{j-1}s^{-\beta j})^{1/2} \\
\leq& \kappa_2 s^{-\beta\ell}(s^{-\beta j} + b_{j-1} + 2(b_{j-1}s^{-\beta j})^{1/2}) \\
\leq& \kappa_2 s^{-\beta\ell}(s^{-\beta j} + b_{j-1} + 2b_{j-1}s^{-\beta j/2}) \\
\leq& \kappa_2 s^{-\beta\ell}(s^{-\beta j/2} + b_{j-1}(1 + 2s^{-\beta j/2})) \\
\leq& \kappa_2 s^{-\beta\ell} b_j\,,
\end{aligned}$$

because $b_j \geq 1$ (and therefore $b_j^{1/2} \leq b_j$) and $s > 1$ for $j \in \mathbb{N}$. To bound the sequence (b_j), rewrite

$$b_j = \sum_{i=0}^{j} s^{-\beta i/2} \prod_{r=i+1}^{j} (1 + 2s^{-\beta r/2})\,,$$

and observe that

$$\prod_{r=i+1}^{j} (1 + 2s^{-\beta r/2}) \leq \prod_{r=0}^{\infty}(1 + 2s^{-\beta r/2})\,.$$

The infinite product converges if and only if the series

$$\sum_{r=0}^{\infty} 2s^{-\beta r/2}$$

converges, which is the case because $s > 1$ and therefore $s^{-\beta/2} < 1$. Denote

$$\prod_{r=0}^{\infty}(1 + 2s^{-\beta r/2}) = \kappa_5\,,$$

with the constant $\kappa_5 \in \mathbb{R}$, and obtain the bound $\kappa_6 \in \mathbb{R}$

$$b_j \leq \kappa_5 \sum_{i=0}^{j} s^{-\beta i/2} \leq \kappa_6\,, \qquad j \in \mathbb{N}\,.$$

Using $b_j \leq \kappa_6$ and (14) shows the lemma.

Lemma 2 *From Assumption 4, Assumption 3, and Lemma 1, it follows that*

$$\rho_{L,\ell}^2 - \rho_{L,\ell-1}^2 \lesssim \frac{1}{\sigma_{low}^2} s^{-\beta\ell}\,.$$

Proof First, we ensure $\rho_{L,\ell} \geq 0$ for $\ell \in \mathbb{N}$ w.l.o.g. by redefining Q_ℓ to $-Q_\ell$ if necessary, and subsequently using $-Q_\ell$ in the estimators (8) and (9). With the definition of the correlation coefficient (1), we obtain

$$
\begin{aligned}
0 \leq \rho_{L,\ell} - \rho_{L,\ell-1} =& \rho_{L,\ell} - \frac{\operatorname{Cov}[Q_L(Z), Q_{\ell-1}(Z)]}{\sigma_L \sigma_{\ell-1}} \\
=& \rho_{L,\ell} - \frac{1}{\sigma_L \sigma_{\ell-1}} \operatorname{Cov}[Q_L(Z), Q_{\ell-1}(Z)] + \frac{1}{2} \frac{\sigma_L^2}{\sigma_L \sigma_{\ell-1}} - \frac{1}{2} \frac{\sigma_L^2}{\sigma_L \sigma_{\ell-1}} \\
&+ \frac{1}{2} \frac{\sigma_{\ell-1}^2}{\sigma_L \sigma_{\ell-1}} - \frac{1}{2} \frac{\sigma_{\ell-1}^2}{\sigma_L \sigma_{\ell-1}} \\
=& \rho_{L,\ell} + \frac{1}{2\sigma_L \sigma_{\ell-1}} \operatorname{Var}[Q_L(Z) - Q_{\ell-1}(Z)] - \frac{1}{2} \left(\frac{\sigma_L}{\sigma_{\ell-1}} + \frac{\sigma_{\ell-1}}{\sigma_L} \right) ,
\end{aligned}
\tag{15}
$$

where we used

$$
\operatorname{Var}[Q_L(Z) - Q_{\ell-1}(Z)] = \operatorname{Var}[Q_L(Z)] + \operatorname{Var}[Q_{\ell-1}(Z)] - 2 \operatorname{Cov}[Q_L(Z), Q_{\ell-1}(Z)] \,.
$$

With $x = \sigma_L / \sigma_{\ell-1}$, we can rewrite the last term in (15) as

$$
\frac{1}{2} \left(\frac{\sigma_L}{\sigma_{\ell-1}} + \frac{\sigma_{\ell-1}}{\sigma_L} \right) = \frac{1}{2} \left(x + \frac{1}{x} \right) .
$$

Because

$$
\frac{1}{2} \left(x + \frac{1}{x} \right) \geq 1
$$

holds for $x \in \mathbb{R}_+$, and because $0 \leq \rho_{L,\ell} \leq 1$ per definition, we obtain the following bound on $\rho_{L,\ell} - \rho_{L,\ell-1}$

$$
\begin{aligned}
0 \leq \rho_{L,\ell} - \rho_{L,\ell-1} =& \rho_{L,\ell} + \frac{1}{2\sigma_L \sigma_{\ell-1}} \operatorname{Var}[Q_L(Z) - Q_{\ell-1}(Z)] - \frac{1}{2} \left(\frac{\sigma_L}{\sigma_{\ell-1}} + \frac{\sigma_{\ell-1}}{\sigma_L} \right) \\
\leq& \frac{1}{2\sigma_L \sigma_{\ell-1}} \operatorname{Var}[Q_L(Z) - Q_{\ell-1}(Z)] \\
\lesssim& \frac{1}{\sigma_{\text{low}}^2} s^{-\beta\ell} ,
\end{aligned}
$$

where we used Lemma 1 to bound $\operatorname{Var}[Q_L(Z) - Q_{\ell-1}(Z)]$ and the lower bound σ_{low} of Assumption 3. Since $\rho_{L,\ell} + \rho_{L,\ell-1} \leq 2$, we obtain

$$
\rho_{L,\ell}^2 - \rho_{L,\ell-1}^2 = (\rho_{L,\ell} - \rho_{L,\ell-1})(\rho_{L,\ell} + \rho_{L,\ell-1}) \lesssim \frac{1}{\sigma_{\text{low}}^2} s^{-\beta\ell} \,.
$$

3.2 Proof of main theorem

With the Lemmata 1–2 discussed in Section 3.1, we now prove Theorem 1.

Proof (of Theorem 1) The MSE of the MFMC estimator $\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}^{\mathrm{MF}}}$ is split into the biasing and the variance term

$$e(\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}^{\mathrm{MF}}}) = \mathrm{Var}[\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}^{\mathrm{MF}}}] + (\mathbb{E}[Q(Z) - Q_L(Z)])^2 \,. \tag{16}$$

We first consider the biasing term of the MSE. With the maximal level L defined as in (3), we obtain with Assumption 1

$$(\mathbb{E}[Q(Z) - Q_L(Z)])^2 \lesssim \frac{\epsilon}{2}\,.$$

Consider now the variance term $\mathrm{Var}[\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}^{\mathrm{MF}}}]$. Assumption 3 means that $\sigma_\ell \leq \sigma_{\mathrm{up}}$ for $\ell = 1, \dots, L$. We therefore have

$$\mathrm{Var}[\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}^{\mathrm{MF}}}] \leq \frac{\sigma_{\mathrm{up}}^2 \left(1 - \rho_{L,L-1}^2\right)}{\left(m_L^{\mathrm{MF}}\right)^2 w_L} p = \frac{\sigma_{\mathrm{up}}^2 \left(1 - \rho_{L,L-1}^2\right)}{p w_L} \left(\sum_{\ell=1}^{L} w_\ell r_\ell\right)^2 ,$$

where we used $m_L^{\mathrm{MF}} = p/(\boldsymbol{w}^T \boldsymbol{r})$ and $\boldsymbol{r} = [r_1, \dots, r_L]^T$ defined in (10) in Section 2.4. Note that Assumptions 6–7 are required for $\boldsymbol{m}^{\mathrm{MF}}$ and $\boldsymbol{a}^{\mathrm{MF}}$ to be optimal in the sense defined in Section 2.4. We further have with the definition of $\boldsymbol{r}$ in (10) in Section 2.4 that

$$\frac{\sigma_{\mathrm{up}}^2 \left(1 - \rho_{L,L-1}^2\right)}{p w_L} \left(\sum_{\ell=1}^{L} w_\ell r_\ell\right)^2 = \frac{\sigma_{\mathrm{up}}^2}{p} \left(\sum_{\ell=1}^{L} \sqrt{w_\ell \left(\rho_{L,\ell}^2 - \rho_{L,\ell-1}^2\right)}\right)^2 , \tag{17}$$

see [38, Proof of Corollary 3.5] for the transformations. With Assumption 2 and Lemma 2, we obtain

$$\sum_{\ell=1}^{L} \sqrt{w_\ell \left(\rho_{L,\ell}^2 - \rho_{L,\ell-1}^2\right)} \lesssim \frac{1}{\sigma_{\mathrm{low}}} \sum_{\ell=1}^{L} \sqrt{s^{\gamma\ell} s^{-\beta\ell}} \lesssim \frac{1}{\sigma_{\mathrm{low}}} \sum_{\ell=1}^{L} \left(s^{\frac{\gamma-\beta}{2}}\right)^{\ell} . \tag{18}$$

Assumption 5 gives $\beta > \gamma$, and therefore $s^{\gamma-\beta} < 1$ (because $s > 1$). Therefore, we obtain with the geometric series that

$$\sum_{\ell=1}^{L} \sqrt{w_\ell \left(\rho_{L,\ell}^2 - \rho_{L,\ell-1}^2\right)} \lesssim \frac{1}{\sigma_{\mathrm{low}}} \frac{s^{\frac{\gamma-\beta}{2}}}{1 - s^{\frac{\gamma-\beta}{2}}}\,.$$

This means that we have

$$\mathrm{Var}[\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}^{\mathrm{MF}}}] \lesssim \frac{\sigma_{\mathrm{up}}^2}{p} \left(\frac{1}{\sigma_{\mathrm{low}}} \frac{s^{\frac{\gamma-\beta}{2}}}{1 - s^{\frac{\gamma-\beta}{2}}}\right)^2 = \frac{1}{2\epsilon^{-1}} = \frac{\epsilon}{2}\,.$$

This means that we bounded the variance and the biasing term by $\epsilon/2$ and therefore have that the MSE is bounded by ϵ. The choice of the budget p in (11) leads to $c(\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}^{\mathrm{MF}}}) \lesssim \epsilon^{-1}$.

The following corollary considers the case where Assumption 5 is violated, i.e., where $\beta \leq \gamma$.

Corollary 1 *Consider the same setup as in Theorem 1, except that Assumption 5 is violated and that $\epsilon < \mathrm{e}^{-1}$. We obtain the following bounds on the costs*

$$c(\widehat{Q}^{MF}_{L,\boldsymbol{m}^{MF}}) \lesssim \begin{cases} \epsilon^{-1}\ln(\epsilon)^2\,, & \gamma = \beta \\ \epsilon^{-1-\frac{\gamma-\beta}{2\alpha}}\,, & \gamma > \beta \end{cases}, \tag{19}$$

where ln *denotes the logarithm with base* e*.*

Proof Consider (18) in the proof of Theorem 1 and note that equation (18) holds even if Assumption 5 is violated. Note that the following proof closely follows [9, Theorem 1] and [18, Theorem 3.1].

We first consider the case $\gamma > \beta$ and obtain

$$\sum_{\ell=0}^{L}\left(s^{\frac{\gamma-\beta}{2}}\right)^{\ell} = \frac{1 - s^{\frac{\gamma-\beta}{2}(L+1)}}{1 - s^{\frac{\gamma-\beta}{2}}} = \frac{s^{-\frac{\gamma-\beta}{2}} - s^{\frac{\gamma-\beta}{2}L}}{s^{-\frac{\gamma-\beta}{2}} - 1} = \frac{s^{-\frac{\gamma-\beta}{2}}}{s^{-\frac{\gamma-\beta}{2}} - 1} - \frac{s^{\frac{\gamma-\beta}{2}L}}{s^{-\frac{\gamma-\beta}{2}} - 1}\,.$$

Because $\gamma > \beta$ and $s > 1$, we obtain for the first term

$$\frac{s^{-\frac{\gamma-\beta}{2}}}{s^{-\frac{\gamma-\beta}{2}} - 1} \leq 0\,,$$

and therefore

$$\sum_{\ell=0}^{L}\left(s^{\frac{\gamma-\beta}{2}}\right)^{\ell} \leq \frac{s^{\frac{\gamma-\beta}{2}L}}{1 - s^{-\frac{\gamma-\beta}{2}}}\,.$$

With the definition of L in (3) and $\lceil x \rceil \leq x + 1, x \in \mathbb{R}$, we obtain

$$\frac{s^{\frac{\gamma-\beta}{2}L}}{1 - s^{-\frac{\gamma-\beta}{2}}} \leq \frac{s^{\frac{\gamma-\beta}{2}}}{1 - s^{-\frac{\gamma-\beta}{2}}} s^{\frac{\gamma-\beta}{2\alpha}\log_s\left(\sqrt{2}\kappa_1\epsilon^{-1/2}\right)} = \frac{s^{\frac{\gamma-\beta}{2}}\left(\sqrt{2}\kappa_1\right)^{\frac{\gamma-\beta}{2\alpha}}}{1 - s^{-\frac{\gamma-\beta}{2}}}\epsilon^{-\frac{\gamma-\beta}{4\alpha}}\,.$$

With the constant

$$\kappa_7 = 2\frac{\sigma^2_{\text{up}}}{\sigma^2_{\text{low}}}\left(\frac{s^{\frac{\gamma-\beta}{2}}\left(\sqrt{2}\kappa_1\right)^{\frac{\gamma-\beta}{2\alpha}}}{1 - s^{-\frac{\gamma-\beta}{2}}}\right)^2$$

and (17), we obtain

$$\mathrm{Var}[\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}^{\mathrm{MF}}}] \lesssim \frac{k_7}{2p}\left(\epsilon^{-\frac{\gamma-\beta}{4\alpha}}\right)^2\,.$$

Thus, with $p = \kappa_7\epsilon^{-1-\frac{\gamma-\beta}{2\alpha}}$ follows the bound (19) for the case $\gamma > \beta$.

Consider now the case $\gamma = \beta$. We obtain

$$\sum_{\ell=0}^{L} \left(s^{\frac{\gamma-\beta}{2}}\right)^{\ell} = L + 1 \leq \alpha^{-1} \log_s(\sqrt{2}\kappa_1 \epsilon^{-1/2}) + 2$$
$$= \alpha^{-1} \log_s(\sqrt{2}\kappa_1) + \alpha^{-1} \frac{\ln(\epsilon^{-1})}{2\ln(s)} + 2\,.$$

With $\epsilon < \mathrm{e}^{-1}$ follows $1 \leq \ln(\epsilon^{-1})$, and therefore

$$L + 1 \leq \kappa_8 \ln(\epsilon^{-1})\,,$$

with

$$\kappa_8 = \alpha^{-1} \log_s(\sqrt{2}\kappa_1) + \alpha^{-1} \frac{1}{2\ln(s)} + 2\,.$$

Set

$$p = 2\frac{\sigma_{\text{up}}^2}{\sigma_{\text{low}}^2} \kappa_8^2 \epsilon^{-1} \ln(\epsilon)^2\,,$$

where we used that $\ln(\epsilon^{-1})^2 = \ln(\epsilon)^2$, to obtain the bound (19) for the case $\gamma = \beta$.

4 Numerical experiment

This section demonstrates Theorem 1 numerically on an elliptic PDE with random coefficients.

4.1 Problem setup

Let $\mathcal{G} = (0, 1)^2$ be a domain with boundary $\partial\mathcal{G}$. Consider the linear elliptic PDE with random coefficients

$$-\nabla \cdot (k(\omega, \boldsymbol{x})\nabla u(\omega, \boldsymbol{x})) = f(\boldsymbol{x})\,, \qquad \boldsymbol{x} \in \mathcal{G}\,, \tag{20}$$
$$u(\omega, \boldsymbol{x}) = 0\,, \qquad \boldsymbol{x} \in \partial\mathcal{G}\,, \tag{21}$$

where $u : \Omega \times \bar{\mathcal{G}} \to \mathbb{R}$ is the solution function defined on the set of outcomes Ω and the closure $\bar{\mathcal{G}}$ of $\mathcal{G}$. The coefficient k is given as

$$k(\omega, \boldsymbol{x}) = \sum_{i=1}^{d} z_i(\omega) \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{v}_i\|_2}{0.045}\right)\,,$$

where $d = 9$, $Z = [z_1, \ldots, z_d]^T$ is a random vector with components that are independent and distributed uniformly in $[10^{-4}, 10^{-1}]$, and the points in $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d] \in \mathbb{R}^{2\times d}$ are given by the matrix

$$\boldsymbol{V} = \begin{bmatrix} 0.5 & 0.2 & 0.8 & 0.8 & 0.2 & 0 & 0.5 & 1 & 0.5 \\ 0.5 & 0.2 & 0.2 & 0.8 & 0.8 & 0.5 & 0 & 0.5 & 1 \end{bmatrix}\,.$$

	rate	constant
Assumption 1	$\alpha \approx 1.0579$	$\kappa_1 \approx 4.0528 \times 10^1$
Assumption 2	$\gamma \approx 1.0551$	$\kappa_3 \approx 2.3615 \times 10^{-6}$
Assumption 4	$\beta \approx 1.9365$	$\kappa_2 \approx 1.3744 \times 10^3$

Table 1: The table reports the rates and constants of Assumptions 1,2,4 that we estimated for our problem (20)–(21).

The domain $\mathcal{D}$ is $\mathcal{D} = [10^{-4}, 10^{-1}]^9$. The right-hand side is set to $f(\boldsymbol{x}) = 10$. The function $Q : \mathcal{D} \to \mathbb{R}$ is

$$Q(Z(\omega)) = \left(\int_{\mathcal{G}} u(\omega, \boldsymbol{x})^2 \mathrm{d}\boldsymbol{x} \right)^{1/2} .$$

We are interested in estimating $\mathbb{E}[Q(Z)]$.

We discretize the problem (20)–(21) with piecewise bilinear finite elements on a rectangular grid in the domain $\mathcal{G}$. The level ℓ defines the mesh width $2^{-\ell}$ of the grid in one dimension. The solution of the discretized problem at level ℓ is denoted as u_ℓ, which gives rise to the functions

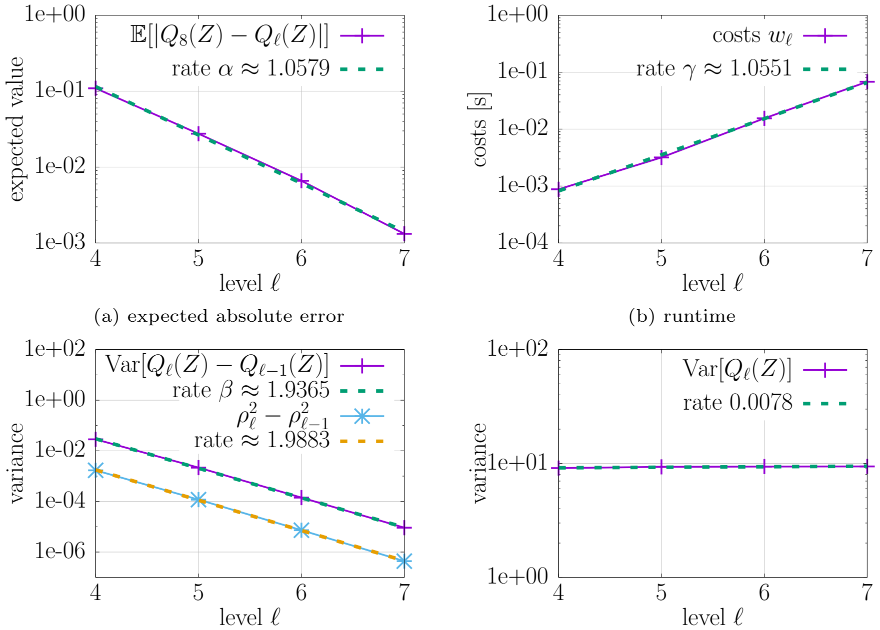
$$Q_\ell(Z(\omega)) = \left(\int_{\mathcal{G}} u_\ell(\omega, \boldsymbol{x})^2 \mathrm{d}\boldsymbol{x} \right)^{1/2}, \qquad \ell \in \mathbb{N}\,.$$

Our reference estimate $\widehat{Q}^{\mathrm{Ref}} \approx 10.54829$ of $\mathbb{E}[Q(Z)]$ is a basic Monte Carlo estimate obtained from 10^4 samples

4.2 Numerical illustration of the assumptions

Dirichlet problems such as (20)–(21) are well studied in the multilevel Monte Carlo literature. We therefore refer to the literature for theoretical considerations in the context of multilevel Monte Carlo of problem (20)–(21) and its variations [9,8].

We estimate the rates of Assumptions 1–4 numerically from $n = 10^4$ samples $Z_1, \dots, Z_n$ of the random variable Z and the corresponding evaluations of $Q_3, \dots, Q_8$. Consider first Assumption 1. We use basic Monte Carlo estimators with $n = 10^4$ samples to estimate $|\mathbb{E}[Q_8(Z) - Q_\ell(Z)]|$ for $\ell = 3, \dots, 7$ and then find $\kappa_1 \in \mathbb{R}_+$ and $\alpha \in \mathbb{R}_+$ that best fit the estimates in the L_2 norm. Since the domain $\mathcal{G}$ is in a two-dimensional space, we set $s = 2^2 = 4$. Note that we estimate the constant κ_1 and the rate α with respect to Q_8 instead of Q. We follow [9] and ignore levels that lead to too coarse grids. Note that a general discussion on which models to select for MFMC estimation is given in [38, Section 3.5]. The behavior of $|\mathbb{E}[Q_8(Z) - Q_\ell(Z)]|$ for $\ell = 3, \dots, 7$ is shown in Figure 1a. The constant κ_1 and the rate α are reported in Table 1. We repeat the same procedure to obtain the rates and constants of Assumptions 2–4, which are visualized in Figure 1 and reported in Table 1. Note that our estimated rates satisfy $\beta > \gamma$, cf. Assumption 5.



(a) expected absolute error

(b) runtime

(c) decay of variance

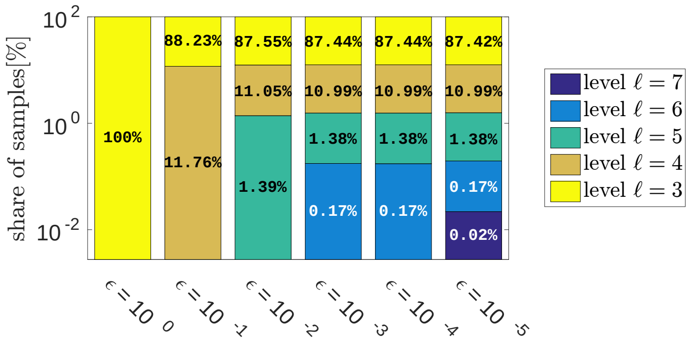
(d) variance

Fig. 1: The plot in (a) shows that the rate of the decay of the expected absolute error is $\alpha \approx 1$, see Assumption 1. The plot in (b) reports the rate $\gamma \approx 1$ of the increase of the runtime of the evaluations Q_ℓ for $\ell = 3, \ldots, 7$, see Assumption 2. The plots in (c) and (d) report the behavior of the variance with respect to Assumption 4 and Assumption 3, respectively. Note that $\beta > \gamma$ as required by Assumption 5.

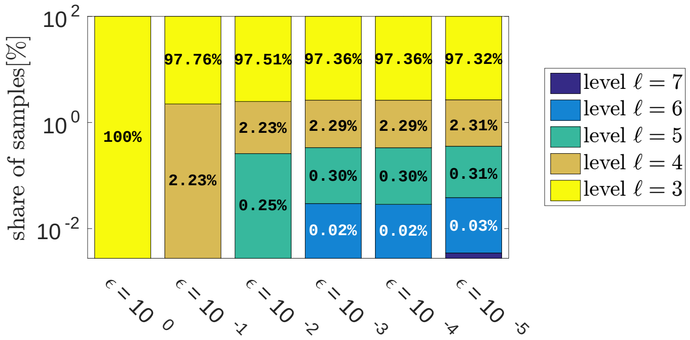
	costs [s]	variances	correlation coefficients
level $\ell = 3$	2.94×10^{-4}	9.41	$9.990894578 \times 10^{-1}$
level $\ell = 4$	8.77×10^{-4}	9.40	$9.999374083 \times 10^{-1}$
level $\ell = 5$	3.18×10^{-3}	9.34	$9.999961196 \times 10^{-1}$
level $\ell = 6$	1.54×10^{-2}	9.10	$9.999997721 \times 10^{-1}$
level $\ell = 7$	6.78×10^{-2}	8.27	$9.999999908 \times 10^{-1}$

Table 2: The table reports the costs $w_3, \ldots, w_7$ of functions $Q_3, \ldots, Q_7$, and the sample estimates of the variances $\sigma_3^2, \ldots, \sigma_7^2$ and the correlation coefficients $\rho_{8,3}, \ldots, \rho_{8,7}$ of the random variables $Q_3(Z), \ldots, Q_7(Z)$ estimated from 10^4 samples.

We measure the costs of evaluating the functions $Q_3, \ldots, Q_7$ by averaging the runtime over 10^4 runs. We use Matlab for the implementation and Matlab's backslash operator to solve systems of linear equations. The time measurements were performed on nodes with Xeon E5-1620 CPUs and 32GB RAM. The variances $\sigma_3^2, \ldots, \sigma_7^2$ and the correlation coefficients $\rho_{8,3}, \ldots, \rho_{8,7}$ are obtained from 10^4 samples, see [38]. The costs $w_3, \ldots, w_7$, the variances $\sigma_3^2, \ldots, \sigma_7^2$, and the correlation coefficients $\rho_{8,3}, \ldots, \rho_{8,7}$ are reported in Table 2.



(a) multilevel Monte Carlo



(b) MFMC

Fig. 2: The plots report the share of the number of samples of each level in the total number of samples. MFMC evaluates the coarsest model more often than multilevel Monte Carlo in this example.

4.3 Numerical illustration of main theorem

For $\epsilon \in \{10^0, 10^{-1}, \ldots, 10^{-5}\}$, we derive multilevel Monte Carlo and MFMC estimates of $\mathbb{E}[Q(Z)]$ following Section 2.3 and Section 2.4, respectively. The number of samples for the multilevel Monte Carlo estimators are derived using the rates in Table 1. The number of samples and the coefficients for the MFMC estimators are obtained using the costs, variances, and correlation coefficients reported in Table 2. Figure 2 compares the number of samples obtained with multilevel Monte Carlo and MFMC. The absolute numbers of samples are reported in Table 3 for multilevel Monte Carlo and in Table 4 for MFMC. Both methods lead to similar numbers of samples. MFMC assigns more samples to level $\ell = 3$ than multilevel Monte Carlo. A detailed comparison is shown in Figure 3 for $\epsilon = 10^{-5}$, which illustrates that multilevel Monte Carlo distributes the number of samples logarithmically among the levels depending on the rates β and γ, see Section 2.4. MFMC directly uses the costs, variances, and correlation coefficients and derives a more fine-grained distribution among the

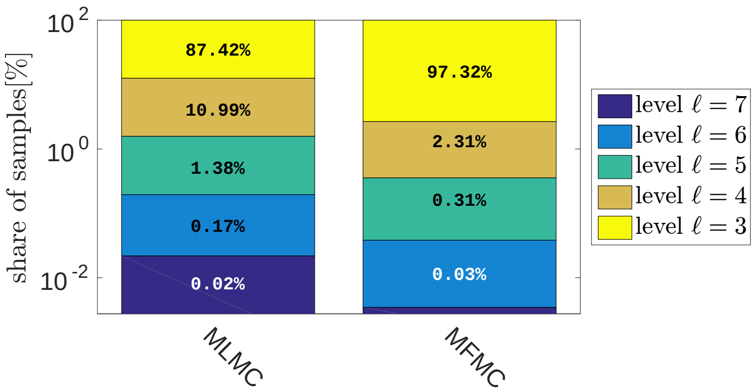


Fig. 3: The bar plot shows a detailed comparison of the share of the samples determined by multilevel Monte Carlo (MLMC) and MFMC for $\epsilon = 10^{-5}$. Multilevel Monte Carlo distributes the number of samples logarithmically among the levels, whereas MFMC determines a fine-grained distribution of the number of samples. Thus, the bars have the same size on a logarithmic scale for multilevel Monte Carlo but different sizes for MFMC. Note that the percent of the share of the total number of samples for each bar is shown in the plot.

levels than multilevel Monte Carlo. We refer to [38] for further investigations on the number of samples in the context of MFMC.

We repeat the multilevel Monte Carlo and the MFMC estimation 100 times and report in Figure 4 the estimated MSE

$$\hat{e}(\widehat{Q}) = \frac{1}{100}\sum_{i=1}^{100}\left(\widehat{Q}_i - \widehat{Q}^{\text{Ref}}\right)^2\,, \tag{22}$$

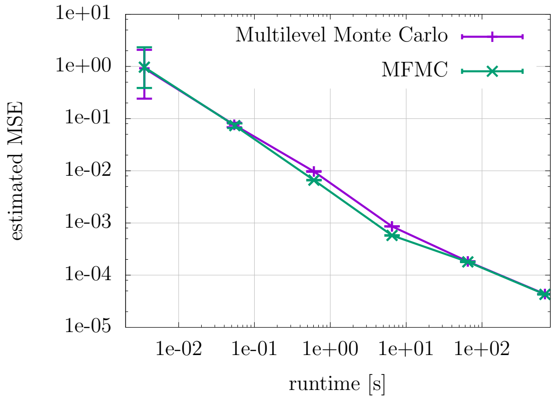
where $\widehat{Q}_i$ is either a multilevel Monte Carlo estimator or an MFMC estimator, and where $\widehat{Q}^{\text{Ref}}$ is the reference estimate, see Section 4.1. Figure 4 additionally shows error bars with length

$$\frac{1}{99}\sum_{i=1}^{100}\left(\hat{e}(\widehat{Q}) - \left(\widehat{Q}_i - \widehat{Q}^{\text{Ref}}\right)^2\right)^2\,, \tag{23}$$

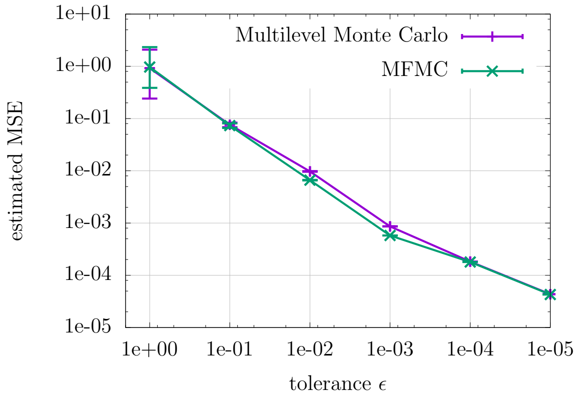
which is an estimate of the variance of the error $\hat{e}(\widehat{Q})$ if $\hat{e}(\widehat{Q})$ is considered as a random variable. The estimated MSE for the multilevel Monte Carlo and the MFMC estimators are reported in Figure 4. Both estimators lead to similar estimated MSEs, which is in agreement with Theorem 1. The runtime of the multilevel Monte Carlo and the MFMC estimator is reported in Table 3 and Table 4, respectively.

4.4 MFMC and coarse-grid (weakly-correlated) models

The random variables $Q_3(Z), \dots, Q_7(Z)$ corresponding to levels $\ell = 3, \dots, 7$ are highly correlated to the random variable $Q_8(Z)$, see Table 2. We now



(a) estimated MSE w.r.t. runtime in seconds



(b) estimated MSE w.r.t. tolerance ϵ

Fig. 4: The results are in agreement with Theorem 1, which states that the costs of the MFMC estimator with MSE $e(\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}^{\mathrm{MF}}}) \lesssim \epsilon$ are bounded by $c(\widehat{Q}^{\mathrm{MF}}_{L,\boldsymbol{m}^{\mathrm{MF}}}) \lesssim \epsilon^{-1}$ under Assumptions 1–7. The behavior of the MFMC estimator is similar to the behavior of the multilevel Monte Carlo estimator.

Table 3: The table reports the number of samples used in the multilevel Monte Carlo estimator and the runtime in seconds. The runtime is averaged over 100 runs.

ϵ	$\ell = 3$	$\ell = 4$	$\ell = 5$	$\ell = 6$	$\ell = 7$	total	time[s]
10^0	1.20×10^1	0	0	0	0	1.20×10^1	0.003
10^{-1}	1.20×10^2	1.60×10^1	0	0	0	1.36×10^2	0.054
10^{-2}	1.19×10^3	1.51×10^2	1.90×10^1	0	0	1.36×10^3	0.606
10^{-3}	1.19×10^4	1.50×10^3	1.89×10^2	2.40×10^1	0	1.36×10^4	6.499
10^{-4}	1.19×10^5	1.50×10^4	1.89×10^3	2.38×10^2	0	1.36×10^5	64.955
10^{-5}	1.19×10^6	1.50×10^5	1.88×10^4	2.37×10^3	2.99×10^2	1.36×10^6	674.360

Table 4: The tables reports the number of samples used in the MFMC estimator. While the total number of samples is higher than in multilevel Monte Carlo (see Table 3), the multilevel Monte Carlo method requires more samples than MFMC at higher levels (i.e., more expensive evaluations) and thus the runtimes are about the same for each $\epsilon \in \{1, \dots, 10^{-5}\}$.

ϵ	$\ell = 3$	$\ell = 4$	$\ell = 5$	$\ell = 6$	$\ell = 7$	total	time[s]
10^0	1.20×10^1	0	0	0	0	1.20×10^1	0.003
10^{-1}	1.75×10^2	4.00×10^0	0	0	0	1.79×10^2	0.051
10^{-2}	1.88×10^3	4.30×10^1	5.00×10^0	0	0	1.92×10^3	0.554
10^{-3}	1.97×10^4	4.65×10^2	6.20×10^1	6.00×10^0	0	2.02×10^4	6.505
10^{-4}	1.96×10^5	4.64×10^3	6.18×10^2	5.80×10^1	0	2.02×10^5	64.966
10^{-5}	2.01×10^6	4.80×10^4	6.60×10^3	7.24×10^2	7.20×10^1	2.07×10^6	674.380

	costs [s]	variances	correlation coefficients
level $\ell = 1$	1.12×10^{-4}	3.23	$7.761297293 \times 10^{-1}$
level $\ell = 2$	1.67×10^{-4}	6.09	$9.884862151 \times 10^{-1}$
level $\ell = 3$	2.94×10^{-4}	9.41	$9.990894578 \times 10^{-1}$
level $\ell = 4$	8.77×10^{-4}	9.40	$9.999374083 \times 10^{-1}$
level $\ell = 5$	3.18×10^{-3}	9.34	$9.999961196 \times 10^{-1}$

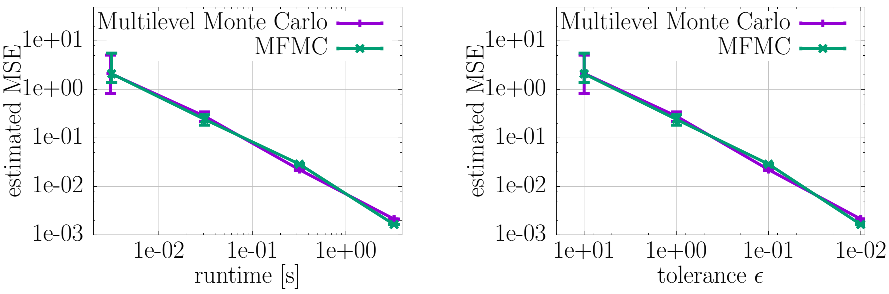
Table 5: The table reports the costs $w_1, \dots, w_5$ of functions $Q_1, \dots, Q_5$, and the sample estimates of the variances $\sigma_1^2, \dots, \sigma_5^2$ and the correlation coefficients $\rho_{8,1}, \dots, \rho_{8,5}$ of the random variables $Q_1(Z), \dots, Q_5(Z)$ estimated from 10^4 samples. Note that the costs, variances, and correlation coefficients for levels $\ell = 3, \dots, 7$ are reported in Table 2.

consider MFMC with $Q_1(Z), Q_2(Z), \dots, Q_5(Z)$, where we include levels $\ell = 1$ and $\ell = 2$. The estimated correlation coefficients, costs, and variances are reported in Table 5. The random variable $Q_1(Z)$ corresponding to level $\ell = 1$ is significantly weaker correlated to $Q_8(Z)$ than the random variables $Q_2(Z), \dots, Q_5(Z)$. As in Section 4.2, we measure the rates of Assumptions 1,2,4, and obtain $\alpha \approx 0.9255$, $\beta \approx 1.6202$ and $\gamma \approx 0.7160$. These rates are similar as the rates reported in Table 1. Note that $\beta > \gamma$.

We derive multilevel Monte Carlo and MFMC estimates of $\mathbb{E}[Q(Z)]$ for $\epsilon \in \{10^1, 10^0, 10^{-1}, 10^{-2}\}$ and report the estimated MSE (22) in Figure 5. The error bars show the variance (23). The results illustrate that MFMC achieves an estimated MSE that is in agreement with Theorem 1 also in this case where the random variable $Q_1(Z)$ corresponding to the coarsest discretization is only weakly correlated to $Q_8(Z)$. Multilevel Monte Carlo and MFMC show a similar behavior. We refer to [38, Section 3.4, Section 4.3], where the performance of MFMC with weakly-correlated models is further investigated analytically and numerically.

5 Conclusions

The MFMC method provides a general framework for combining multiple approximations into an estimator of statistics of a random variable that is expensive (or impossible) to sample. We discussed MFMC in the special case where sampling the random variable requires solving a PDE, and where we can sample only approximations that correspond to a hierarchy of discretizations



(a) estimated MSE w.r.t. runtime in seconds (b) estimated MSE w.r.t. tolerance ϵ

Fig. 5: The plots report the estimated MSE of multilevel Monte Carlo and the MFMC estimators that combine $Q_1(Z), \dots, Q_5(Z)$ corresponding to levels $\ell = 1, \dots, 5$. The random variables $Q_1(Z)$ and $Q_2(Z)$ are only weakly correlated to $Q_8(Z)$. The MFMC estimator shows a similar behavior as the multilevel Monte Carlo estimator.

of the PDE. In this setting, and under standard assumptions on the discretizations of the PDE, the MFMC estimator is efficient, which means that the costs of the MFMC estimator with MSE below a threshold are bounded linearly in the threshold. Our numerical results illustrated the theory.

Acknowledgment

The first and the third author were supported in part by the AFOSR MURI on multi-information sources of multi-physics systems under Award Number FA9550-15-1-0038, program manager Jean-Luc Cambier, and by the United States Department of Energy Applied Mathematics Program, Awards DE-FG02-08ER2585 and DE-SC0009297, as part of the DiaMonD Multifaceted Mathematics Integrated Capability Center. The second author was supported by the US Department of Energy Office of Science grant DE-SC0009324 and the Air Force Office of Scientific Grant FA9550-15-1-0001. Some of the numerical examples were computed on the computer cluster of the Munich Centre of Advanced Computing.

References

1. I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 45(3):1005–1034, 2007.
2. A. Barth, C. Schwab, and N. Zollinger. Multi-level Monte Carlo Finite Element method for elliptic PDEs with stochastic coefficients. *Numerische Mathematik*, 119(1):123–161, Sept. 2011.
3. P. Benner, S. Gugercin, and K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57(4):483–531, 2015.
4. C. Bierig and A. Chernov. Convergence analysis of multilevel Monte Carlo variance estimators and application for random obstacle problems. *Numerische Mathematik*, 130(4):579–613, 2015.

5. C. Bierig and A. Chernov. Approximation of probability density functions by the multilevel Monte Carlo maximum entropy method. *Journal of Computational Physics*, 314:661 – 681, 2016.
6. C. Bierig and A. Chernov. Estimation of arbitrary order central statistical moments by the multilevel Monte Carlo method. *Stochastics and Partial Differential Equations Analysis and Computations*, 4(1):3–40, 2016.
7. H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:1–123, 2004.
8. J. Charrier, R. Scheichl, and A. L. Teckentrup. Finite element error analysis of elliptic PDEs with random coefficients and its application to multilevel Monte Carlo methods. *SIAM Journal on Numerical Analysis*, 51(1):322–352, 2013.
9. K. A. Cliffe, M. Giles, R. Scheichl, and A. L. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science*, 14(1):3–15, 2011.
10. N. Collier, A.-L. Haji-Ali, F. Nobile, E. von Schwerin, and R. Tempone. A continuation multilevel Monte Carlo algorithm. *BIT Numerical Mathematics*, 55(2):399–432, 2015.
11. C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.
12. A. I. J. Forrester and A. J. Keane. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1–3):50–79, Jan. 2009.
13. K. A. Forrester A., Sóbester A. *Engineering design via surrogate modelling: a practical guide*. Wiley, 2008.
14. F. Franzelin, P. Diehl, and D. Pflüger. Non-intrusive uncertainty quantification with sparse grids for multivariate peridynamic simulations. In M. Griebel and A. M. Schweitzer, editors, *Meshfree Methods for Partial Differential Equations VII*, pages 115–143, Cham, 2015. Springer International Publishing.
15. F. Franzelin and D. Pflüger. From data to uncertainty: An efficient integrated data-driven sparse grid approach to propagate uncertainty. In J. Garcke and D. Pflüger, editors, *Sparse Grids and Applications - Stuttgart 2014*, pages 29–49, Cham, 2016. Springer International Publishing.
16. T. Gerstner and M. Griebel. Numerical integration using sparse grids. *Numerical Algorithms*, 18(3):209–232, 1998.
17. T. Gerstner and M. Griebel. Dimension–adaptive tensor–product quadrature. *Computing*, 71(1):65–87, 2003.
18. M. Giles. Multi-level Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.
19. M. Griebel, H. Harbrecht, and M. Peters. Multilevel quadrature for elliptic parametric partial differential equations on non-nested meshes. *Stochastic Partial Differential Equations: Analysis and Computations*, 2015. submitted.
20. S. Gugercin, A. Antoulas, and C. Beattie. $\mathcal{H}_2$ Model Reduction for Large-Scale Linear Dynamical Systems. *SIAM Journal on Matrix Analysis and Applications*, 30(2):609–638, Jan. 2008.
21. M. D. Gunzburger, C. G. Webster, and G. Zhang. Stochastic finite element methods for partial differential equations with random input data. *Acta Numerica*, 23:521–650, 5 2014.
22. A.-L. Haji-Ali, F. Nobile, and R. Tempone. Multi-index Monte Carlo: when sparsity meets sampling. *Numerische Mathematik*, 132(4):767–806, June 2015.
23. J. M. Hammersley and D. C. Handscomb. *Monte Carlo Methods*. Methuen London, 1964.
24. H. Harbrecht, M. Peters, and M. Siebenmorgen. On multilevel quadrature for elliptic stochastic partial differential equations. In J. Garcke and M. Griebel, editors, *Sparse Grids and Applications*, pages 161–179, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
25. H. Harbrecht, M. Peters, and M. Siebenmorgen. Multilevel accelerated quadrature for PDEs with log-normally distributed diffusion coefficient. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):520–551, 2016.
26. S. Heinrich. Multilevel Monte Carlo Methods. In S. Margenov, J. Waśniewski, and P. Yalamov, editors, *Large-Scale Scientific Computing*, number 2179 in Lecture Notes in Computer Science, pages 58–67. Springer, 2001.

27. J. Li and D. Xiu. Evaluation of failure probability via surrogate models. *Journal of Computational Physics*, 229(23):8966–8980, 2010.
28. J. S. Liu. *Monte Carlo strategies in scientific computing*. Springer, 2008.
29. A. J. Majda and B. Gershgorin. Quantifying uncertainty in climate change science through empirical information theory. *Proceedings of the National Academy of Sciences of the United States of America*, 107(34):14958–14963, Aug. 2010.
30. B. L. Nelson. On control variate estimators. *Computers & Operations Research*, 14(3):219–225, 1987.
31. L. Ng and K. Willcox. Multifidelity approaches for optimization under uncertainty. *International Journal for Numerical Methods in Engineering*, 100(10):746–772, 2014.
32. L. Ng and K. Willcox. Monte-Carlo information-reuse approach to aircraft conceptual design optimization under uncertainty. *Journal of Aircraft*, pages 1–12, 2015.
33. F. Nobile, R. Tempone, and C. G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, 2008.
34. S. Pauli and P. Arbenz. Determining optimal multilevel Monte Carlo parameters with application to fault tolerance. *Computers & Mathematics with Applications*, 70(11):2638 – 2651, 2015.
35. S. Pauli, P. Arbenz, and C. Schwab. Intrinsic fault tolerance of multilevel Monte Carlo methods. *Journal of Parallel and Distributed Computing*, 84:24 – 36, 2015.
36. B. Peherstorfer, T. Cui, Y. Marzouk, and K. Willcox. Multifidelity importance sampling. *Computer Methods in Applied Mechanics and Engineering*, 300:490 – 509, 2016.
37. B. Peherstorfer and K. Willcox. Online adaptive model reduction for nonlinear systems via low-rank updates. *SIAM Journal on Scientific Computing*, 37(4):A2123–A2150, 2015.
38. B. Peherstorfer, K. Willcox, and M. Gunzburger. Optimal model management for multifidelity Monte Carlo estimation. *SIAM Journal on Scientific Computing*, 38(5):A3163–A3194, 2016.
39. C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2004.
40. G. Rozza, D. Huynh, and A. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Archives of Computational Methods in Engineering*, 15(3):1–47, 2007.
41. L. Sirovich. Turbulence and the dynamics of coherent structures. *Quarterly of Applied Mathematics*, 45:561–571, 1987.
42. A. L. Teckentrup, R. Scheichl, M. Giles, and E. Ullmann. Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients. *Numerische Mathematik*, 125(3):569–600, 2013.
43. E. Ullmann, H. C. Elman, and O. G. Ernst. Efficient iterative solvers for stochastic Galerkin discretizations of log-transformed random diffusion problems. *SIAM Journal on Scientific Computing*, 34(2):A659–A682, 2012.
44. E. Ullmann and I. Papaioannou. Multilevel estimation of rare events. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):922–953, 2015.
45. E. Ullmann and C. E. Powell. Solving log-transformed random diffusion problems by stochastic Galerkin mixed finite element methods. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):509–534, 2015.
46. V. Vapnik. *Statistical Learning Theory*. Wiley, 1998.
47. D. Xiu. Fast numerical methods for stochastic computations: A review. *Communications in computational physics*, 5:242–272, 2009.