



**HAL**  
open science

## Multilevel Surrogate-based Control Variates

Mohamed Reda El Amri, Paul Mycek, Sophie Ricci, Matthias De Lozzo

► **To cite this version:**

Mohamed Reda El Amri, Paul Mycek, Sophie Ricci, Matthias De Lozzo. Multilevel Surrogate-based Control Variates. 2023. hal-04129561v1

**HAL Id: hal-04129561**

**<https://hal.science/hal-04129561v1>**

Preprint submitted on 15 Jun 2023 (v1), last revised 18 Jun 2024 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Multilevel Surrogate-based Control Variates

Reda El Amri<sup>\*1</sup>, Paul Mycek<sup>1,2</sup>, Sophie Ricci<sup>1,2</sup>, and Matthias De Lozzo<sup>1</sup>

<sup>1</sup>IRT Saint Exupéry, Toulouse, France (matthias.delozzo@irt-saintexupery.com).

<sup>2</sup>CECI, CNRS - CERFACS, Toulouse, France (mycek@cerfacs.fr, ricci@cerfacs.fr)

## Abstract

Monte Carlo (MC) sampling is a popular method for estimating the statistics (e.g. expectation and variance) of a random variable. Its slow convergence has led to the emergence of advanced techniques to reduce the variance of the MC estimator for the outputs of computationally expensive solvers. The control variates (CV) method corrects the MC estimator with a term derived from auxiliary random variables that are highly correlated with the original random variable. These auxiliary variables may come from surrogate models. Such a surrogate-based CV strategy is extended here to the multilevel Monte Carlo (MLMC) framework, which relies on a sequence of levels corresponding to numerical simulators with increasing accuracy and computational cost. MLMC combines output samples obtained across levels, into a telescopic sum of differences between MC estimators for successive fidelities. In this paper, we introduce three multilevel variance reduction strategies that rely on surrogate-based CV and MLMC. MLCV is presented as an extension of CV where the correction terms devised from surrogate models for simulators of different levels add up. MLMC-CV improves the MLMC estimator by using a CV based on a surrogate of the correction term at each level. Further variance reduction is achieved by using the surrogate-based CVs of all the levels in the MLMC-MLCV strategy. Alternative solutions that reduce the subset of surrogates used for the multilevel estimation are also introduced. The proposed methods are tested on a test case from the literature consisting of a spectral discretization of an uncertain 1D heat equation, where the statistic of interest is the expected value of the integrated temperature along the domain at a given time. The results are assessed in terms of the accuracy and computational cost of the multilevel estimators, depending on whether the construction of the surrogates, and the associated computational cost, precede the evaluation of the estimator. It was shown that when the lower fidelity outputs are strongly correlated with the high-fidelity outputs, a significant variance reduction is obtained when using surrogate models for the coarser levels only. It was also shown that taking advantage of pre-existing surrogate models proves to be an even more efficient strategy.

*Keywords:* Multifidelity, multilevel Monte Carlo, control variates, surrogate models, polynomial chaos, variance reduction.

---

<sup>\*</sup>Current affiliation: IFP Énergies Nouvelles, Solaize, France (mohamed-reda.el-amri@ifpen.fr)

# 1 Introduction

In recent years, the propagation of uncertainties in numerical simulators has become an essential step in the study of physical phenomena. Therefore, uncertainty quantification (UQ) has emerged as an important element in scientific computing [50, 52]. For complex nonlinear systems, the task of quantifying the effect of uncertainties on the simulator behaviour can pose major challenges as closed-form solutions often do not exist. Sampling-based algorithms are considered the default approach when it comes to UQ for complex nonlinear simulators. The Monte Carlo (MC) sampling method is the most popular and flexible method in UQ. Here, statistical information is extracted from a  $n$ -sample of simulator responses. Due to its non-intrusive nature, MC is straightforward to implement. On the one hand, the convergence of an MC statistic is independent of the dimension, but on the other hand, it is very slow, namely at a rate of  $n^{-0.5}$ . For computationally expensive simulators, the cost of MC is often considered impractically high. In some cases, slight improvements can be obtained through the use of importance sampling [27], latin hypercube sampling [35] or quasi-Monte Carlo (QMC) [13].

Another well-known approach in UQ consists in replacing the simulator by a surrogate model. Fast-to-evaluate surrogates can be used to approximate the high-fidelity model response and therefore reduce drastically the computational cost of estimating statistics. Polynomial chaos (PC) [34, 37], Gaussian process models [45], radial basis functions [41] and (deep) neural networks [54] are commonly used surrogate models. The downside of the surrogate-based approach is that it introduces approximation error, which causes biases in the statistics estimators. Besides, this approximation error tends to increase in high dimension.

One recent MC sampling framework based on control variates (CV) [29, 31, 30, 33] has been extensively developed and used. In a sampling-based CV strategy, one seeks to reduce the variance of the MC estimator of a random variable, arising from the high-fidelity model, by exploiting its correlation with an auxiliary random variable that arises from low-fidelity models approximating the same input-output relationship. In classic CV theory, the mean of the auxiliary random variable is assumed to be known. Unfortunately, in many cases such an assumption is not valid. This creates the need to use another estimator for the auxiliary random variable [24, 44, 43, 42, 39], which involves an additional computational cost. Recently, the adoption of surrogate models as such auxiliary random variables, in particular PC models, has been explored in [16, 56, 17, 25]. The benefits are twofold: (1) the prediction of the surrogate models may be highly correlated with the output high-fidelity model, leading to a reduced variance of the CV estimator as compared to the MC estimation; (2) certain surrogate models provide exact statistics that are needed by the CV approach. Another variance reduction technique, called the multilevel MC (MLMC) method [21, 10, 22], uses a hierarchy of models. Originally devised for the estimation of expected values, MLMC has since been extended to the estimation of other statistics, see, e.g., [4, 5] for the estimation of variance and higher order central moments and [36] for the computation of Sobol' indices. Typically, multilevel methods are based on a sequence of levels which correspond to a hierarchy of simulators with increasing accuracy and cost. From a practical standpoint, the different levels often correspond to simulators with increasing mesh resolutions. This translates into a lower accuracy of the so-called *coarse* levels, whereas the *finer* levels correspond

to accurate simulators. By construction, MLMC results in an unbiased estimator. It relies on a telescopic sum of terms based on the differences between the successive simulators. It debiases the MC estimator associated with the lowest-level simulator. Many other unbiased multilevel estimators have been devised in recent years. The multi-index MC estimator [26] is an extension of the MLMC estimator such that the telescoping sum idea is used in multiple directions. In [23], the authors developed a variant of MLMC which uses QMC samples instead of independent MC samples for each level. Multi-fidelity MC (MFMC) estimators [19, 18] are another extension that is based on the CV approach. Recently, the authors in [48, 47] took a different approach by formulating a multilevel estimator as the result of a linear regression problem.

In this work, we combine multilevel sampling with surrogate-based CV to define new estimators with the advantages of both. These novel multifidelity variance reduction strategies allow us to quantify efficiently the output uncertainties of simulators with limited computational budgets. A first strategy, named *multilevel control variates* (MLCV), uses CVs based on surrogate models of simulators corresponding to different levels. Although this strategy does not build on the MLMC approach specifically, it still exploits multilevel information through the surrogates constructed at different levels. The second strategy, named *multilevel Monte Carlo with control variates* (MLMC-MLCV), utilizes CV in an information fusion framework to exploit synergies between the flexible MLMC sampling and the correlation shared between the high- and low-fidelity components. The numerical results show that the unbiased MLMC-MLCV estimators can converge faster than the existing estimators.

The paper is organized as follows. We introduce notations and the necessary mathematical background in section 2. In section 3, we briefly discuss the MLMC estimator and then define our proposed MLMC-MLCV estimator that combines multilevel sampling and surrogate-based CV. In section 4, we conduct numerical experiments to support the theoretical results. Section 5 proposes concluding remarks.

## 2 Background

In this section, we summarize the important results of statistical estimation using control variates and we show how surrogate models can be leveraged in this setting.

### 2.1 Notation

We first introduce a few notations. Throughout the rest of this paper, the high-fidelity numerical model is abstractly represented by the deterministic mapping

$$\begin{aligned} f : \Xi &\longrightarrow \mathbb{R} \\ \mathbf{x} &\longmapsto f(\mathbf{x}), \end{aligned} \tag{1}$$

where  $\mathbf{x} := (x_1 \ \cdots \ x_d)^\top$  is a vector of  $d$  uncertain input parameters evolving in a measurable space denoted by  $\Xi = \Xi_1 \times \cdots \times \Xi_d$ , with  $\Xi_i \subset \mathbb{R}$ . Following a probabilistic approach, the  $d$  inputs are assumed to be continuous random variables  $X_1, \dots, X_d$ , defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , with known probability density functions (PDFs)  $p_{X_1}, \dots, p_{X_d}$ . They are assumed to be independent, so that the PDF of the random vector

$\mathbf{X} := (X_1 \ \cdots \ X_d)^\top : \Omega \rightarrow \Xi$  is  $p_{\mathbf{X}} = \prod_{i=1}^d p_{X_i}$ . In this work, we seek an accurate estimator of some statistic  $\theta$  of the output random variable  $Y := f(\mathbf{X}) : \Omega \rightarrow \mathbb{R}$ , e.g., its expected value ( $\theta = \mathbb{E}[Y]$ ) or variance ( $\theta = \mathbb{V}[Y]$ ), at a reasonable computational cost  $\mathcal{C}$ .

## 2.2 Crude Monte Carlo

In practice, the MC estimator  $\hat{\theta}$  of the statistic  $\theta$  is based on  $n$  observations  $Y^{(1)}, \dots, Y^{(n)}$  defined as  $Y^{(i)} = f(\mathbf{X}^{(i)})$  where  $\mathcal{X} = \{\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)}\}$  is a  $n$ -sample of  $\mathbf{X}$ , that is, a collection of independent and identically distributed random variables with the same distribution as  $\mathbf{X}$ . For instance, the *sample mean* and (unbiased) *sample variance* estimators,

$$\hat{E}[Y] = n^{-1} \sum_{i=1}^n Y^{(i)} \quad \text{and} \quad \hat{V}[Y] = (n-1)^{-1} \sum_{i=1}^n (Y^{(i)} - \hat{E}[Y])^2 \quad (2)$$

are MC estimators of the expectation and variance of  $Y$ , respectively. It is well known that these estimators are *unbiased*, that is,  $\mathbb{E}[\hat{\theta}] = \theta$ , so that the mean square error (MSE) of  $\hat{\theta}$  reduces to the variance of the estimator:

$$\text{MSE}(\hat{\theta}, \theta) := \mathbb{E}[(\hat{\theta} - \theta)^2] = \mathbb{V}[\hat{\theta}] + (\mathbb{E}[\hat{\theta}] - \theta)^2 = \mathbb{V}[\hat{\theta}]. \quad (3)$$

The convergence of these MC estimators is known to be slow, so that variance reduction techniques are needed when dealing with computationally expensive simulators.

## 2.3 Control variates

In this section, we present a well-known variance reduction technique using auxiliary random variables  $Z_1, \dots, Z_M$  as control variates. We denote by  $\tau_1, \dots, \tau_M$  the statistics of the control variates that correspond to the statistic  $\theta$  that we seek to estimate. These control statistics  $\tau_1, \dots, \tau_M$  are assumed to be known exactly. Then, the CV estimator is defined as

$$\hat{\theta}^{\text{CV}}(\boldsymbol{\alpha}) = \hat{\theta} - \boldsymbol{\alpha}^\top (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}), \quad (4)$$

where  $\hat{\theta}$  and  $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1 \ \cdots \ \hat{\tau}_M)^\top$  are *unbiased* MC estimators of  $\theta$  and  $\boldsymbol{\tau} = (\tau_1 \ \cdots \ \tau_M)^\top$ , respectively, based on a *common* input  $n$ -sample  $\mathcal{X}$ , and where  $\boldsymbol{\alpha} \in \mathbb{R}^M$  is the control parameter. We note that the CV estimator is unbiased by construction, regardless of the value of the parameter  $\boldsymbol{\alpha}$ , and that its variance reads

$$\mathbb{V}[\hat{\theta}^{\text{CV}}(\boldsymbol{\alpha})] = \mathbb{V}[\hat{\theta}] + \boldsymbol{\alpha}^\top \boldsymbol{\Sigma} \boldsymbol{\alpha} - 2\boldsymbol{\alpha}^\top \mathbf{c}. \quad (5)$$

with  $\mathbf{c} := \mathbb{C}[\hat{\boldsymbol{\tau}}, \hat{\theta}] \in \mathbb{R}^M$  and  $\boldsymbol{\Sigma} := \mathbb{C}[\hat{\boldsymbol{\tau}}] \in \mathbb{R}^{M \times M}$ . To fully take advantage of the control variates, the parameter  $\boldsymbol{\alpha}$  is selected so as to minimize the variance (5) of the CV estimator. Assuming that the covariance matrix  $\boldsymbol{\Sigma}$  is non-singular (and thus symmetric positive definite, SPD), the first- and second-order optimality conditions of the minimization problem imply that there exists a unique optimal solution,

$$\boldsymbol{\alpha}^* = \boldsymbol{\Sigma}^{-1} \mathbf{c}. \quad (6)$$

The optimal CV estimator is thus  $\hat{\theta}^{\text{CV}}(\boldsymbol{\alpha}^*)$ , and its variance is given by

$$\mathbb{V}[\hat{\theta}^{\text{CV}}(\boldsymbol{\alpha}^*)] = (1 - R^2)\mathbb{V}[\hat{\theta}], \quad (7)$$

where  $R^2 = \mathbb{V}[\hat{\theta}]^{-1}\mathbf{c}^\top \boldsymbol{\Sigma}^{-1}\mathbf{c}$  is nonnegative, as  $\boldsymbol{\Sigma}^{-1}$  is SPD. Denoting by  $\mathbf{D} = \text{Diag}(\boldsymbol{\Sigma}) \in \mathbb{R}^{M \times M}$  the diagonal matrix consisting of the diagonal of  $\boldsymbol{\Sigma}$ , it can be shown further that  $R^2 = \mathbf{r}^\top \mathbf{R}^{-1}\mathbf{r}$ , where  $\mathbf{r} = (\mathbb{V}[\hat{\theta}]\mathbf{D})^{-1/2}\mathbf{c}$  is the vector of the Pearson correlation coefficients between  $\hat{\theta}$  and  $\hat{\boldsymbol{\tau}}$ , and  $\mathbf{R} = \mathbf{D}^{-1/2}\boldsymbol{\Sigma}\mathbf{D}^{-1/2}$  is the correlation matrix of  $\hat{\boldsymbol{\tau}}$ . Thus,  $R^2 \in [0, 1]$  corresponds to the squared coefficient of multiple correlation between  $\hat{\theta}$  and the elements of  $\hat{\boldsymbol{\tau}}$  [1, section 2.5.2]. Consequently,  $0 \leq \mathbb{V}[\hat{\theta}^{\text{CV}}(\boldsymbol{\alpha}^*)] \leq \mathbb{V}[\hat{\theta}]$ , and we refer to  $R^2$  as the variance reduction factor of the CV estimator. This shows that the variance of the optimal CV estimator  $\hat{\theta}^{\text{CV}}(\boldsymbol{\alpha}^*)$  is always reduced (or, rigorously speaking, not increased) compared to the MC estimator  $\hat{\theta}$ . Furthermore, the higher  $R^2$ , the greater the reduction in variance.

**Remark 1.** *The requirement that  $\boldsymbol{\Sigma}$  be non-singular is a reasonable one. Indeed, let us suppose that  $\boldsymbol{\Sigma}$  is singular. Then, because  $\boldsymbol{\Sigma}$  is positive semi-definite by construction, this implies that there exists a nonzero vector  $\boldsymbol{\eta} \in \mathbb{R}^M \setminus \{\mathbf{0}\}$  such that  $\boldsymbol{\eta}^\top \boldsymbol{\Sigma} \boldsymbol{\eta} = \mathbb{V}[\boldsymbol{\eta}^\top \hat{\boldsymbol{\tau}}] = 0$ , indicating that any one element of  $\hat{\boldsymbol{\tau}}$  can be expressed as an affine function of the others. As such, it does not bring any additional information to the CV estimator, so that at least one of the control variates can simply be discarded.*

A desirable property of the CV estimator is that increasing the number of control variates improves the CV estimator. Specifically, under mild assumptions, proposition 1 states that the variance of the CV estimator is reduced (or, rigorously speaking, not increased) when adding a new control variate.

**Proposition 1.** *Let  $\hat{\boldsymbol{\tau}}_+ := [\hat{\boldsymbol{\tau}}^\top \ \hat{\tau}_{M+1}]^\top$ ,  $\boldsymbol{\tau}_+ := [\boldsymbol{\tau}^\top \ \tau_{M+1}]^\top$ , and define  $\mathbf{c}_+ = \mathbb{C}[\hat{\boldsymbol{\tau}}_+, \hat{\theta}]$ , and  $\boldsymbol{\Sigma}_+ = \mathbb{C}[\hat{\boldsymbol{\tau}}_+]$ . We further assume that  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\Sigma}_+$  are non-singular and that  $\mathbb{V}[\hat{\tau}_{M+1}] > 0$ . Let  $\hat{\theta}^{\text{CV}}(\boldsymbol{\alpha}_+^*) := \hat{\theta} - \boldsymbol{\alpha}_+^{*\top}(\hat{\boldsymbol{\tau}}_+ - \boldsymbol{\tau}_+)$ , with  $\boldsymbol{\alpha}_+^* = \boldsymbol{\Sigma}_+^{-1}\mathbf{c}_+$ , be the optimal CV estimator based on  $M + 1$  control variates, and let  $R_+^2$  denote its variance reduction factor. Then  $R_+^2 \geq R^2$ .*

*Proof.* We have  $R_+^2 = \mathbf{r}_+^\top \mathbf{R}_+^{-1}\mathbf{r}_+$ , where  $\mathbf{r}_+ = [\mathbf{r}^\top \ \gamma]^\top$  and  $\mathbf{R}_+ = \begin{bmatrix} \mathbf{R} & \mathbf{u} \\ \mathbf{u}^\top & 1 \end{bmatrix}$ , and where  $\mathbf{u} = (\mathbb{V}[\hat{\tau}_{M+1}]\mathbf{D})^{-1/2}\mathbb{C}[\hat{\boldsymbol{\tau}}, \hat{\tau}_{M+1}]$  and  $\gamma = (\mathbb{V}[\hat{\theta}]\mathbb{V}[\hat{\tau}_{M+1}])^{-1/2}\mathbb{C}[\hat{\theta}, \hat{\tau}_{M+1}]$ . Note that  $\mathbf{R}$  and  $\mathbf{R}_+$  are both non-singular, because so are  $\boldsymbol{\Sigma}$  and  $\boldsymbol{\Sigma}_+$ . The augmented matrix  $\mathbf{R}_+$  may be inverted by block using the Schur complement  $s = 1 - \mathbf{u}^\top \mathbf{R}^{-1}\mathbf{u} \neq 0$  of the (1,1)-block  $\mathbf{R}$ . It follows that  $R_+^2 = R^2 + s^{-1}(\gamma - \mathbf{u}^\top \mathbf{R}^{-1}\mathbf{r})^2$ . Now, because  $\mathbf{R}_+$  is SPD (it is positive semidefinite by construction and non-singular by assumption),  $\mathbf{x}^\top \mathbf{R}_+ \mathbf{x} > 0$  for any choice of  $\mathbf{x} \neq \mathbf{0}$ . The particular choice of  $\mathbf{x} = [-\mathbf{u}^\top \mathbf{R}^{-1} \ 1]^\top$  implies that  $s > 0$ , which in turn implies that  $R_+^2 \geq R^2$ .  $\square$

**Corollary 1.** *Under the assumptions of proposition 1, the equality  $R_+^2 = R^2$  holds if and only if  $\gamma = \mathbf{u}^\top \mathbf{R}^{-1}\mathbf{r}$ .*

*Proof.* Straightforward from  $R_+^2 = R^2 + s^{-1}(\gamma - \mathbf{u}^\top \mathbf{R}^{-1}\mathbf{r})^2$  with  $s > 0$ .  $\square$

For the CV estimation of specific statistics, the expressions of  $\mathbf{c}$  and  $\boldsymbol{\Sigma}$  may be further reduced to involve statistics on  $Y$  and  $\mathbf{Z} := (Z_1 \ \cdots \ Z_M)^\top$  directly. These expressions, as

well as the consequences on eqs. (6) and (7), are given in appendix A for the CV estimators of  $\mathbb{E}[Y]$  and  $\mathbb{V}[Y]$ .

It should be noted at this point that, in practice, the optimal parameter  $\boldsymbol{\alpha}^*$  needs to be approximated. Specifically, the statistics involved in  $\mathbf{c}$  and  $\boldsymbol{\Sigma}$  need to be estimated. This can be done either using the same sample  $\mathcal{X}$  as for the CV estimator itself, or using an independent, pilot sample  $\mathcal{X}'$ . The former strategy introduces a bias in the resulting CV estimator, while the latter guarantees unbiasedness but requires additional high-fidelity simulator evaluations, so that the former strategy is generally preferred. Besides, statistical remedies such as jackknifing, splitting or bootstrapping have been proposed to reduce or eliminate the bias introduced by the former strategy [38]. In both strategies, however, neither the theoretical variance reduction given by eq. (7) nor proposition 1 hold anymore.

**Remark 2.** *In some instances, because it only involves the control variates,  $\boldsymbol{\Sigma}$  may be known exactly (see remark 3 for an example in a multilevel setting involving PC-based control variates) or estimated accurately using many independent samples at negligible cost. Unfortunately, this is not the case for  $\mathbf{c}$ , which involves the output  $Y = f(\mathbf{X})$  of the high-fidelity simulator  $f$ .*

## 2.4 Surrogate-based control variates

The efficiency of the CV approach relies on the strong correlation between  $Y = f(\mathbf{X})$  and the control variates  $\mathbf{Z}$ . In a multifidelity framework, the control variates correspond to the output of low-fidelity versions of  $f$ , typically in the form of simulators with simplified physics and/or coarser discretization. In many applications, the exact statistical measures  $\boldsymbol{\tau}$  of such control variates  $\mathbf{Z}$  may not be available. One way to circumvent this limitation is to use additional samples to also estimate  $\boldsymbol{\tau}$ . This type of estimators led to an *approximate* CV class of methods [24, 39, 43], where different model management and sample allocation strategies may be used to find a suitable trade-off between the additional cost of sampling  $\mathbf{Z}$  and the resulting variance reduction. Recently, an optimal strategy was proposed [48, 49, 47] leading to the so-called multilevel best linear unbiased estimator (MLBLUE), as well as a unifying framework for a large class of multilevel and multifidelity estimators.

In this paper, we focus on the case where the low-fidelity models correspond to surrogate models of the high-fidelity simulator  $f$ . The main advantage is that the statistics  $\boldsymbol{\tau}$  may, in some instances, be directly available, or at least may be estimated arbitrarily accurately at negligible cost, so that the original CV approach described in section 2.3 can be used. Using surrogate models in a CV strategy has been explored previously in [16, 56], where the available computational budget is allocated both to the construction of the surrogates and to the actual CV estimations. In [56], the authors introduce an approach that optimally balances the computational effort needed to select the optimal degree of the polynomial chaos (PC) expansion used in a stochastic Galerkin CV approach. In our work, we focus on the situation in which the surrogate models are available, so that we do not consider any optimization strategy for the construction of the surrogates. Nevertheless, for the fairness of comparison, we still report the pre-processing cost of constructing the surrogate models.

We now briefly describe three different surrogate models commonly used in a UQ framework, and discuss the availability of statistics of their output.

### 2.4.1 Gaussian process modelling

We assume that the simulator  $f$  is a realization of a Gaussian process (GP)  $F$  indexed by  $\mathbf{x}$  and defined by its mean function  $m_F$  and covariance kernel  $k_F$ ,

$$\mathbb{E}[F(\mathbf{x})] = m_F(\mathbf{x}), \quad (8)$$

$$\mathbb{C}[F(\mathbf{x}), F(\mathbf{x}')] = k_F(\mathbf{x}, \mathbf{x}'), \quad \forall \mathbf{x}, \mathbf{x}' \in \Xi. \quad (9)$$

In practice, one can parametrize the forms of  $m_F$  and  $k_F$ . For example, in the widely used ordinary GP method, a stationary GP is assumed. In this case,  $m_F$  is set as a constant  $m_F(\mathbf{x}) = m$ . More importantly, it is assumed that  $k_F(\mathbf{x}, \mathbf{x}') = \bar{k}_F(\mathbf{x} - \mathbf{x}')$ , and  $k_F(\mathbf{x}, \mathbf{x}) = \bar{k}_F(0) = \sigma^2$  is a constant. Popular forms of kernels include polynomial, exponential, Gaussian, and Matérn functions. For example, the Gaussian kernel can be written as  $k_F(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp(-\frac{1}{2}\|\mathbf{x} - \mathbf{x}'\|_{\mathbf{h}}^2)$ , where the weighted norm is defined as  $\|\mathbf{x} - \mathbf{x}'\|_{\mathbf{h}} = \left(\sum_{i=1}^d \frac{(x_i - x'_i)^2}{h_i^2}\right)^{1/2}$  where  $h_1, \dots, h_d$  are correlation lengths. The hyperparameters  $\sigma$  and  $h_i$  can be obtained by maximum likelihood. Then, given  $n$  observations  $\mathcal{F} = (f(\mathbf{x}^{(1)}) \ \dots \ f(\mathbf{x}^{(n)}))^\top$  of  $F$  at  $\mathcal{X} = (\mathbf{x}^{(1)} \ \dots \ \mathbf{x}^{(n)})^\top$ , the posterior  $\tilde{F}$  of  $F$  can be defined as

$$\tilde{F} = [F \mid F(\mathcal{X}) = \mathcal{F}], \quad (10)$$

whose expectation and covariance are given by

$$m_{\tilde{F}}(\mathbf{x}) = \mathbb{E}[\tilde{F}(\mathbf{x})] = m_F(\mathbf{x}) + k_F(\mathbf{x}, \mathcal{X})^\top k_F(\mathcal{X}, \mathcal{X})^{-1}(\mathcal{F} - m_F(\mathcal{X})), \quad (11)$$

$$k_{\tilde{F}}(\mathbf{x}, \mathbf{x}') = \mathbb{C}[\tilde{F}(\mathbf{x}), \tilde{F}(\mathbf{x}')] = k_F(\mathbf{x}, \mathbf{x}') - k_F(\mathbf{x}, \mathcal{X})^\top k_F(\mathcal{X}, \mathcal{X})^{-1}k_F(\mathcal{X}, \mathbf{x}'). \quad (12)$$

Thereafter, the high-fidelity model  $f$  at  $\mathbf{x}$  will be approximated by the conditional expectation,

$$g^{\text{GP}}(\mathbf{x}) = m_{\tilde{F}}(\mathbf{x}). \quad (13)$$

Thus, the expectation and variance of  $g^{\text{GP}}(\mathbf{X})$  are defined as

$$\mathbb{E}[g^{\text{GP}}(\mathbf{X})] = \int_{\Xi} m_{\tilde{F}}(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}, \quad (14)$$

$$\mathbb{V}[g^{\text{GP}}(\mathbf{X})] = \int_{\Xi} (m_{\tilde{F}}(\mathbf{x}) - \mathbb{E}[g^{\text{GP}}(\mathbf{X})])^2 p_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}. \quad (15)$$

These two statistics can be approximated empirically by taking a large sample of  $\mathbf{X}$ , as the GP model is inexpensive. Analytical formulas exist for some pairs of distributions and covariance functions (see [8, Table 1] and [28, 15]).

### 2.4.2 Taylor polynomials

We assume that the numerical simulator  $f$  is infinitely differentiable and that the moments of  $\mathbf{X}$  are finite. Under this assumption, it is possible to expand the original simulator  $f$  around the input's expected value  $\boldsymbol{\mu}_{\mathbf{X}} := \mathbb{E}[\mathbf{X}]$  as the infinite polynomial series according to Taylor's theorem,

$$f(\mathbf{X}) = g^{\text{T}\infty}(\mathbf{X}) = \sum_{|\boldsymbol{\beta}| \leq p} \frac{(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})^{\boldsymbol{\beta}}}{\boldsymbol{\beta}!} D^{\boldsymbol{\beta}} f(\boldsymbol{\mu}_{\mathbf{X}}), \quad (16)$$



where  $\boldsymbol{\beta} \in \mathbb{N}^d$ ,  $|\boldsymbol{\beta}| = \sum_{i=1}^d \beta_i$ ,  $\boldsymbol{\beta}! = \prod_{i=1}^d \beta_i!$ ,  $x^\boldsymbol{\beta} = \prod_{i=1}^d x_i^{\beta_i}$  and  $D^\boldsymbol{\beta} f = \frac{\partial^{|\boldsymbol{\beta}|} f}{\partial^{\beta_1} x_1 \cdots \partial^{\beta_d} x_d}$ . Thus, the function  $f$  may be approximated by the first- and second-order Taylor polynomials,

$$g^{\text{T1}}(\mathbf{X}) = f(\boldsymbol{\mu}_{\mathbf{X}}) + \mathbf{J}_f(\boldsymbol{\mu}_{\mathbf{X}})(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}), \quad (17)$$

$$g^{\text{T2}}(\mathbf{X}) = f(\boldsymbol{\mu}_{\mathbf{X}}) + \mathbf{J}_f(\boldsymbol{\mu}_{\mathbf{X}})(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}) + \frac{1}{2}(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})^\top \mathbf{H}_f(\boldsymbol{\mu}_{\mathbf{X}})(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}), \quad (18)$$

where  $\mathbf{J}_f(\boldsymbol{\mu}_{\mathbf{X}}) = \nabla f(\boldsymbol{\mu}_{\mathbf{X}})^\top \in \mathbb{R}^{1 \times d}$  and  $\mathbf{H}_f(\boldsymbol{\mu}_{\mathbf{X}}) \in \mathbb{R}^{d \times d}$  are the Jacobian and Hessian matrices of  $f$  at  $\boldsymbol{\mu}_{\mathbf{X}}$ , respectively. In practical computations, the derivatives may be approximated by numerical differentiation if they are not provided. The expectation and variance of  $g^{\text{T1}}(\mathbf{X})$  and  $g^{\text{T2}}(\mathbf{X})$  are then defined as

$$\mathbb{E}[g^{\text{T1}}(\mathbf{X})] = f(\boldsymbol{\mu}_{\mathbf{X}}), \quad (19)$$

$$\mathbb{V}[g^{\text{T1}}(\mathbf{X})] = \mathbf{J}_f(\boldsymbol{\mu}_{\mathbf{X}})^{\odot 2} \boldsymbol{\sigma}_{\mathbf{X}}^2, \quad (20)$$

$$\mathbb{E}[g^{\text{T2}}(\mathbf{X})] = f(\boldsymbol{\mu}_{\mathbf{X}}) + \frac{1}{2} \text{Tr}(\mathbf{H}_f(\boldsymbol{\mu}_{\mathbf{X}}) \boldsymbol{\Sigma}_{\mathbf{X}}^2), \quad (21)$$

$$\mathbb{V}[g^{\text{T2}}(\mathbf{X})] = \text{Tr}(\mathbf{J}_f(\boldsymbol{\mu}_{\mathbf{X}})^\top \mathbf{J}_f(\boldsymbol{\mu}_{\mathbf{X}}) \boldsymbol{\Sigma}_{\mathbf{X}}^2) + \frac{1}{2} \text{Tr}(\mathbf{H}_f(\boldsymbol{\mu}_{\mathbf{X}})^{\odot 2} \boldsymbol{\sigma}_{\mathbf{X}}^2 (\boldsymbol{\sigma}_{\mathbf{X}}^2)^\top), \quad (22)$$

where for any matrix  $\mathbf{A}$ ,  $\mathbf{A}^{\odot 2}$  denotes the element-wise square of  $\mathbf{A}$ ,  $\text{Tr}(\mathbf{A})$  is the trace of  $\mathbf{A}$ ,  $\boldsymbol{\sigma}_{\mathbf{X}}^2 = [\sigma_{X_1}^2 \cdots \sigma_{X_d}^2]^\top \in \mathbb{R}^d$  is the element-wise variance of  $\mathbf{X}$ , and  $\boldsymbol{\Sigma}_{\mathbf{X}}^2 = \text{Diag}(\boldsymbol{\sigma}_{\mathbf{X}}^2) \in \mathbb{R}^{d \times d}$ .

### 2.4.3 Polynomial chaos expansion

We consider the truncated polynomial chaos (PC) expansion of  $f(\mathbf{X})$  [55, 9, 20, 32],

$$f(\mathbf{X}) \simeq g^{\text{PC}_P}(\mathbf{X}) = \sum_{k=0}^P \mathbf{g}_k \Psi_k(\mathbf{X}), \quad (23)$$

where the coefficients  $\mathbf{g}_k$  are real scalars and  $\Psi_k$  are orthonormal multivariate polynomials:

$$\forall i, j \geq 0, \quad \langle \Psi_i, \Psi_j \rangle_{p_{\mathbf{X}}} := \int_{\Xi} \Psi_i(\mathbf{x}) \Psi_j(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x} = \mathbb{E}[\Psi_i(\mathbf{X}) \Psi_j(\mathbf{X})] = \delta_{ij}, \quad (24)$$

where  $\delta_{ij}$  denotes the Kronecker delta. The coefficients  $\mathbf{g}_k$  may be approximated using non-intrusive techniques such as non-intrusive (pseudo)spectral projection [46, 12, 11], regression [3, 51, 6, 7], interpolation [2, 40], or from intrusive approaches such as the stochastic Galerkin method [20, 32]. Assuming  $\Psi_0 \equiv 1$ , the expectation and variance of  $g^{\text{PC}_P}(\mathbf{X})$  are then given by

$$\mathbb{E}[g^{\text{PC}_P}(\mathbf{X})] = \mathbf{g}_0 \quad \text{and} \quad \mathbb{V}[g^{\text{PC}_P}(\mathbf{X})] = \sum_{k=1}^P \mathbf{g}_k^2. \quad (25)$$

## 3 Multilevel estimators

In this section, we present so-called *multilevel* statistical estimation techniques based on a sequence of simulators  $(f_\ell)_{\ell=0}^L$ , with increasing accuracy and computational cost, where

$f_L \equiv f$  denotes the high-fidelity numerical simulator. The levels are ordered from the *coarsest* ( $\ell = 0$ ) to the *finest* ( $\ell = L$ ). We denote by  $Y_\ell$  the random variable  $Y_\ell = f_\ell(\mathbf{X})$  and  $\theta_\ell$  the statistic of  $f_\ell(\mathbf{X})$  increasingly close to  $\theta_L \equiv \theta$ .

### 3.1 Multilevel Monte Carlo

The statistic  $\theta_L$  can be expressed as the telescoping sum

$$\theta_L = \sum_{\ell=0}^L T_\ell, \quad (26)$$

where  $T_\ell = \theta_\ell - \theta_{\ell-1}$ , and, by convention,  $\theta_{-1} := 0$ . The MLMC estimator  $\hat{\theta}_L^{\text{MLMC}}$  of  $\theta_L$  is then defined as [21, 22, 53]

$$\hat{\theta}^{\text{MLMC}} = \sum_{\ell=0}^L \hat{T}_\ell^{(\ell)}, \quad (27)$$

where, at each level  $\ell$ ,  $\hat{T}_\ell^{(\ell)}$  is an unbiased MC estimator of  $T_\ell$ , based on an input sample  $\mathcal{X}^{(\ell)} = \{\mathbf{X}^{(\ell,i)}\}_{i=1}^{n_\ell}$  such that the members of  $\mathcal{X}^{(\ell)}$  and  $\mathcal{X}^{(\ell')}$  are mutually independent for  $\ell \neq \ell'$ . In many instances,  $\hat{T}_\ell^{(\ell)}$  is actually defined as  $\hat{T}_\ell^{(\ell)} = \hat{\theta}_\ell^{(\ell)} - \hat{\theta}_{\ell-1}^{(\ell)}$ , where  $\hat{\theta}_k^{(\ell)}$  denotes the unbiased MC estimator of  $\theta_k$  based on the simulator  $f_k$  using the  $n_\ell$ -sample  $\mathcal{X}^{(\ell)}$ .

For example, the MLMC estimator  $\hat{E}^{\text{MLMC}}[Y]$  of the expectation  $\mathbb{E}[Y]$  is given by

$$\hat{E}^{\text{MLMC}}[Y] = \hat{E}^{(0)}[Y_0] + \sum_{\ell=1}^L \hat{E}^{(\ell)}[Y_\ell] - \hat{E}^{(\ell)}[Y_{\ell-1}], \quad (28)$$

where  $\hat{E}^{(\ell)}[Y_k] = n_\ell^{-1} \sum_{i=1}^{n_\ell} f_k(\mathbf{X}^{(\ell,i)})$ , with  $k \in \{\ell, \ell-1\}$ . We stress that the correction terms at each level  $\ell$  are computed from the same input sample  $\mathcal{X}^{(\ell)}$ , but using two successive simulators,  $f_\ell$  and  $f_{\ell-1}$ . Similarly, the MLMC estimator  $\hat{V}^{\text{MLMC}}[Y]$  of the variance  $\mathbb{V}[Y]$  is defined as [4]

$$\hat{V}^{\text{MLMC}}[Y] = \hat{V}^{(0)}[Y_0] + \sum_{\ell=1}^L \hat{V}^{(\ell)}[Y_\ell] - \hat{V}^{(\ell)}[Y_{\ell-1}], \quad (29)$$

where  $\hat{V}^{(\ell)}[Y_k] = \frac{n_\ell}{n_\ell-1} (\hat{E}^{(\ell)}[Y_k^2] - \hat{E}^{(\ell)}[Y_k]^2)$  is the single-level unbiased MC variance estimator. Owing to the independence of the estimators  $\hat{T}_\ell^{(\ell)}$ , the variance of the MLMC estimator is

$$\mathbb{V}[\hat{\theta}^{\text{MLMC}}] = \sum_{\ell=0}^L \mathbb{V}[\hat{T}_\ell^{(\ell)}]. \quad (30)$$

In practice, the MLMC method relies on the allocation of the total (expected) computational cost of the MLMC estimator across the different levels, with

$$\text{cost}(\hat{\theta}^{\text{MLMC}}) = \sum_{\ell=0}^L n_\ell (\mathcal{C}_\ell + \mathcal{C}_{\ell-1}), \quad (31)$$

where  $\mathcal{C}_\ell$  is the (expected) cost of one evaluation of the simulator  $f_\ell$ , with  $\mathcal{C}_{-1} := 0$  by convention. Thus, a key aspect is played by the choice of the number of samples  $n_\ell$  allocated to each level  $\ell$ . The goal is to find the sample sizes  $n_0, \dots, n_L$  that minimize the variance of the estimator  $\mathbb{V}[\hat{\theta}^{\text{MLMC}}]$  given a computational budget  $\mathcal{C}$ . Thus, the sample allocation problem reads

$$\begin{aligned} & \underset{n_0, \dots, n_L \in \mathbb{N}^*}{\text{minimize}} && \mathbb{V}[\hat{\theta}^{\text{MLMC}}] \\ & \text{subject to} && \text{cost}(\hat{\theta}^{\text{MLMC}}) = \mathcal{C}. \end{aligned} \tag{32}$$

This minimization problem has a unique solution which can be computed analytically (see, e.g., [21, 10, 36]). In practice,  $\mathbb{V}[\hat{\theta}^{\text{MLMC}}]$  is not known, and we instead rely on the assumption that  $\mathbb{V}[\hat{T}_\ell^{(\ell)}] \lesssim n_\ell^{-1} \mathcal{V}_\ell$ , with  $\mathcal{V}_\ell$  independent of  $n_\ell$  [53, 36]. This is a reasonable assumption that holds for the MLMC estimation of the expectation, variance and covariance [36, Table 1]. Note that, for the estimation of the expectation, we have  $\mathbb{V}[\hat{T}_\ell^{(\ell)}] = n_\ell^{-1} \mathcal{V}_\ell$ , with  $\mathcal{V}_\ell = \mathbb{V}[Y_\ell - Y_{\ell-1}]$  and  $Y_{-1} \equiv 0$ . The sample allocation problem eq. (32) is then replaced with

$$\begin{aligned} & \underset{n_0, \dots, n_L \in \mathbb{N}^*}{\text{minimize}} && \sum_{\ell=0}^L n_\ell^{-1} \mathcal{V}_\ell \\ & \text{subject to} && \text{cost}(\hat{\theta}^{\text{MLMC}}) = \mathcal{C}, \end{aligned} \tag{33}$$

which is equivalent for the expectation, and an approximation for other statistics.

## 3.2 Multilevel surrogate-based control variate strategies

In this section, we introduce various surrogate-based control variate strategies in a multilevel framework where a hierarchy of simulators  $(f_\ell)_{\ell=0}^L$  is available. These strategies rely on using the random variables  $(g_\ell(\mathbf{X}))_{\ell=0}^L$  as control variates, where  $g_\ell$  is a surrogate model of the simulator  $f_\ell$ .

### 3.2.1 Multilevel control variates (MLCV)

The first strategy, referred to as *multilevel control variates* and hereafter abbreviated MLCV, consists of using the surrogate models at *all levels* to build the control variates in eq. (4). Thus,  $\tau_\ell$  corresponds to the statistical measure of the random variable  $Z_\ell = g_\ell(\mathbf{X})$ , and  $\hat{\tau}_\ell$  to its unbiased MC estimator. For instance, the MLCV estimator of the expectation based on an  $n_L$ -sample  $\mathcal{X}^{(L)} = \{\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n_L)}\}$  reads

$$\hat{E}^{\text{MLCV}}[Y](\boldsymbol{\alpha}) = \hat{E}^{(L)}[Y_L] - \sum_{\ell=0}^L \alpha_\ell \left( \hat{E}^{(L)}[Z_\ell] - \mu_{Z_\ell} \right), \tag{34}$$

where  $\hat{E}^{(L)}[Y_L] = n_L^{-1} \sum_{i=1}^{n_L} f_L(\mathbf{X}^{(i)})$ ,  $\hat{E}^{(L)}[Z_\ell] = n_L^{-1} \sum_{i=1}^{n_L} g_\ell(\mathbf{X}^{(i)})$ , and  $\mu_{Z_\ell} = \mathbb{E}[g_\ell(\mathbf{X})]$ . Note that this approach does not build on the MLMC methodology described in section 3.1, but still exploits multilevel information through the surrogates constructed at different levels.

### 3.2.2 Multilevel Monte Carlo with control variates (MLMC-CV)

The second strategy consists of improving the MLMC estimator eq. (27) using the surrogate-based control variates  $Z_0, \dots, Z_L$ . Specifically, the MLMC-CV estimator improves the MC estimation of each of the correction terms of the MLMC estimator by using a surrogate model of the corresponding correction term as a control variate. Note that the MLMF approach proposed in [18, 19] is based on a similar strategy, using arbitrary low-fidelity models at level in an *approximate* CV setting. The MLMC-CV estimator reads

$$\hat{\theta}^{\text{MLMC-CV}}(\alpha_1, \dots, \alpha_L) = \sum_{\ell=0}^L \hat{T}_\ell^{\text{CV}}(\alpha_\ell), \quad (35)$$

where  $\hat{T}_\ell^{\text{CV}}(\alpha_\ell)$  is the CV estimator of the multilevel correction term  $T_\ell$  (cf. eq. (26)),

$$\hat{T}_\ell^{\text{CV}}(\alpha_\ell) = \hat{T}_\ell^{(\ell)} - \alpha_\ell(\hat{U}_\ell^{(\ell)} - U_\ell), \quad (36)$$

with  $\hat{U}_k^{(\ell)}$  an unbiased MC estimator of the control variate statistic  $U_k = \tau_k - \tau_{k-1}$ , based on the same input sample  $\mathcal{X}^{(\ell)}$  as  $\hat{T}_\ell^{(\ell)}$ , again with members of  $\mathcal{X}^{(\ell)}$  and  $\mathcal{X}^{(\ell')}$  being independent for  $\ell \neq \ell'$ . Note that, in eq. (36), because a single control variate is used at each level, the definition of  $\hat{U}_k^{(\ell)}$  is used in the specific case where  $k = \ell$ . The more general definition when  $k$  and  $\ell$  are not necessarily equal will be useful later in section 3.2.3, where we consider multiple control variates per level. In practice,  $\hat{U}_k^{(\ell)}$  may be defined as  $\hat{U}_k^{(\ell)} = \hat{\tau}_k^{(\ell)} - \hat{\tau}_{k-1}^{(\ell)}$ , where  $\hat{\tau}_k^{(\ell)}$  is an unbiased estimator of  $\tau_k$  using the  $n_\ell$ -sample  $\mathcal{X}^{(\ell)}$ .

The optimal value  $\alpha_\ell^*$  for  $\alpha_\ell$  is obtained individually for each  $\ell = 0, \dots, L$  as the optimal (single) CV parameter for  $\hat{T}_\ell^{\text{CV}}(\alpha_\ell)$ ,

$$\alpha_\ell^* = \frac{\mathbb{C}[\hat{T}_\ell^{(\ell)}, \hat{U}_\ell^{(\ell)}]}{\mathbb{V}[\hat{U}_\ell^{(\ell)}]}, \quad (37)$$

and the resulting variance of the control variate estimator of the correction is

$$\mathbb{V}[\hat{T}_\ell^{\text{CV}}(\alpha_\ell^*)] = (1 - \rho_\ell^2)\mathbb{V}[\hat{T}_\ell^{(\ell)}], \quad \text{with } \rho_\ell = \frac{\mathbb{C}[\hat{T}_\ell^{(\ell)}, \hat{U}_\ell^{(\ell)}]}{\mathbb{V}[\hat{T}_\ell^{(\ell)}]^{1/2}\mathbb{V}[\hat{U}_\ell^{(\ell)}]^{1/2}} \in [-1, 1]. \quad (38)$$

The correction estimators  $(\hat{T}_\ell^{\text{CV}})_{\ell=0}^L$  being mutually independent, the variance of the optimal MLMC-CV estimator is

$$\mathbb{V}[\hat{\theta}^{\text{MLMC-CV}}(\alpha_0^*, \dots, \alpha_L^*)] = \sum_{\ell=0}^L \mathbb{V}[\hat{T}_\ell^{\text{CV}}(\alpha_\ell^*)] = \sum_{\ell=0}^L (1 - \rho_\ell^2)\mathbb{V}[\hat{T}_\ell^{(\ell)}] \leq \sum_{\ell=0}^L \mathbb{V}[\hat{T}_\ell^{(\ell)}], \quad (39)$$

indicating that the variance of the MLMC-CV estimator is smaller (as long as  $\rho_\ell^2 > 0$ ) than that of the MLMC estimator; see eq. (30). We remark that the variance reduction depends on the (squared) correlation between  $\hat{T}_\ell^{(\ell)}$  and  $\hat{U}_\ell^{(\ell)}$ , which, in turn, typically relates to some measure of similarity between high-fidelity corrections  $Y_\ell - Y_{\ell-1}$  and the corresponding CV corrections (see appendix A for the expectation and variance estimators in a single-level setting).

In our surrogate-based approach, we may define control variates as  $Z_\ell = g_\ell(\mathbf{X})$ , where  $g_\ell$  is a surrogate of  $f_\ell$ , so that  $\tau_\ell$  and  $\tau_{\ell-1}$  could be estimated using samples of  $Z_\ell$  and  $Z_{\ell-1}$ , respectively. It is then crucial to construct the surrogates such that their successive differences  $g_\ell - g_{\ell-1}$  are good approximations of  $f_\ell - f_{\ell-1}$ , to ensure a high similarity between  $Y_\ell - Y_{\ell-1}$  and  $Z_\ell - Z_{\ell-1}$ . However, constructing  $g_\ell$  as a surrogate of  $f_\ell$  does not give any guarantee on the quality of  $g_\ell - g_{\ell-1}$  as a surrogate of  $f_\ell - f_{\ell-1}$ . Instead, in addition to the surrogate models  $g_\ell$  of  $f_\ell$ , we construct surrogate models  $h_\ell$  of the differences  $f_\ell - f_{\ell-1}$ , and we define auxiliary surrogate models  $\tilde{g}_{\ell-1} = g_\ell - h_\ell$ , for  $\ell = 1, \dots, L$ . On each level  $\ell$ , we then use samples of  $Z_\ell = g_\ell(\mathbf{X})$  and  $\tilde{Z}_{\ell-1} = \tilde{g}_{\ell-1}(\mathbf{X})$  for the estimation of  $\tau_\ell$  and  $\tau_{\ell-1}$ , respectively. As a result, the variance reduction now depends on the similarity between  $Y_\ell - Y_{\ell-1}$  and  $W_\ell := Z_\ell - \tilde{Z}_{\ell-1} = g_\ell(\mathbf{X}) - \tilde{g}_{\ell-1}(\mathbf{X}) = h_\ell(\mathbf{X})$ , where  $h_\ell$  has been constructed to ensure such a similarity. Specifically, for the expectation, the MLMC-CV estimator reads

$$\hat{E}^{\text{MLMC-CV}}[Y](\alpha_0, \dots, \alpha_L) = \sum_{\ell=0}^L \left( \hat{E}^{(\ell)}[Y_\ell] - \hat{E}^{(\ell)}[Y_{\ell-1}] \right) - \alpha_\ell \left( \hat{E}^{(\ell)}[Z_\ell] - \hat{E}^{(\ell)}[\tilde{Z}_{\ell-1}] - (\mu_{Z_\ell} - \mu_{\tilde{Z}_{\ell-1}}) \right), \quad (40)$$

with optimal values of  $\alpha_\ell$  given by (see appendix A.1, with  $M = 1$ )

$$\alpha_\ell^* = \frac{\mathbb{C}[Y_\ell - Y_{\ell-1}, W_\ell]}{\mathbb{V}[Z_\ell - Z_{\ell-1}]}, \quad \forall \ell = 0, \dots, L, \quad (41)$$

resulting in level-dependent reduction factors  $1 - \rho_\ell^2$ , where

$$\rho_\ell^2 = \frac{\mathbb{C}[Y_\ell - Y_{\ell-1}, W_\ell]^2}{\mathbb{V}[Y_\ell - Y_{\ell-1}]\mathbb{V}[W_\ell]} \quad (42)$$

is the squared correlation coefficient between  $Y_\ell - Y_{\ell-1}$  and  $W_\ell = h_\ell(\mathbf{X})$ . It should be noted that, because of the linearity of the expectation operator and its MC estimator, the use of  $\tilde{g}_{\ell-1}$  is superfluous. Indeed, we may directly define the MLMC-CV estimator of the expectation as

$$\hat{E}^{\text{MLMC-CV}}[Y](\alpha_0, \dots, \alpha_L) = \sum_{\ell=0}^L \left( \hat{E}^{(\ell)}[Y_\ell] - \hat{E}^{(\ell)}[Y_{\ell-1}] \right) - \alpha_\ell \left( \hat{E}^{(\ell)}[W_\ell] - \mu_{W_\ell} \right), \quad (43)$$

with  $\mu_{W_\ell} = \mathbb{E}[W_\ell]$ . For the variance, the MLMC-CV estimator reads

$$\hat{V}^{\text{MLMC-CV}}[Y](\alpha_0, \dots, \alpha_L) = \sum_{\ell=0}^L \left( \hat{V}^{(\ell)}[Y_\ell] - \hat{V}^{(\ell)}[Y_{\ell-1}] \right) - \alpha_\ell \left( \hat{V}^{(\ell)}[Z_\ell] - \hat{V}^{(\ell)}[\tilde{Z}_{\ell-1}] - (\sigma_{Z_\ell}^2 - \sigma_{\tilde{Z}_{\ell-1}}^2) \right), \quad (44)$$

with  $\sigma_{Z_\ell}^2 = \mathbb{V}[Z_\ell]$  and  $\sigma_{\tilde{Z}_{\ell-1}}^2 = \mathbb{V}[\tilde{Z}_{\ell-1}]$ . The resulting level-dependent reduction factors are then related to the correlation between  $(Y_\ell - Y_{\ell-1} - \mathbb{E}[Y_\ell - Y_{\ell-1}])^2$  and  $(h_\ell(\mathbf{X}) - \mathbb{E}[h_\ell(\mathbf{X})])^2$  (see appendix A.2, with  $M = 1$ ). Further details on the construction of  $h_\ell$  will be given in section 4.

### 3.2.3 Multilevel Monte Carlo with multilevel control variates (MLMC-MLCV)

We now propose to further improve the MLMC-CV estimator eq. (35) by combining MLMC with the MLCV approach described in section 3.2.1, resulting in the MLMC-MLCV method. The approach consists in using, at each level  $\ell$ , the surrogate-based control variates of *all the levels*  $\ell' = 0, \dots, L$ , rather than only using those of level  $\ell$ , as was previously done with the MLMC-CV approach.

The MLMC-MLCV estimator then reads

$$\hat{\theta}^{\text{MLMC-MLCV}}(\boldsymbol{\alpha}_0, \dots, \boldsymbol{\alpha}_L) = \sum_{\ell=0}^L \hat{T}_\ell^{\text{MLCV}}(\boldsymbol{\alpha}_\ell), \quad (45)$$

where  $\boldsymbol{\alpha}_\ell$  denotes the CV parameter at level  $\ell$ , and  $\hat{T}_\ell^{\text{MLCV}}(\boldsymbol{\alpha}_\ell)$  is the MLCV estimator of  $T_\ell$ ,

$$\hat{T}_\ell^{\text{MLCV}}(\boldsymbol{\alpha}_\ell) = \hat{T}_\ell^{(\ell)} - \boldsymbol{\alpha}_\ell^\top (\hat{\mathbf{U}}_\ell^{(\ell)} - \mathbf{U}_\ell), \quad (46)$$

with

$$\mathbf{U}_0 = (\tau_k)_{k=0}^L \quad \hat{\mathbf{U}}_0^{(0)} = (\hat{\tau}_k^{(0)})_{k=0}^L \quad (47)$$

$$\mathbf{U}_\ell = (U_k)_{k=1}^L, \quad \text{for } \ell > 0, \quad \hat{\mathbf{U}}_\ell^{(\ell)} = (\hat{U}_k^{(\ell)})_{k=1}^L, \quad \text{for } \ell > 0, \quad (48)$$

and with  $U_k$  and  $\hat{U}_k^{(\ell)}$  defined as in section 3.2.2.

Because each term eq. (46) is an unbiased (multiple) CV estimator of  $T_\ell$ , the resulting estimator eq. (45) is also unbiased, and the optimal (variance minimizing) value  $\boldsymbol{\alpha}_\ell^*$  of  $\boldsymbol{\alpha}_\ell$  is given individually for each  $\ell = 0, \dots, L$  as the optimal (multiple) CV parameter for  $\hat{T}_\ell^{(\ell)}(\boldsymbol{\alpha}_\ell)$ ,

$$\boldsymbol{\alpha}_\ell^* = \mathbb{C}[\hat{\mathbf{U}}_\ell^{(\ell)}]^{-1} \mathbb{C}[\hat{\mathbf{U}}_\ell^{(\ell)}, \hat{T}_\ell^{(\ell)}]. \quad (49)$$

The resulting variance is given by

$$\mathbb{V}[\hat{\theta}^{\text{MLMC-MLCV}}(\boldsymbol{\alpha}_0^*, \dots, \boldsymbol{\alpha}_L^*)] = \sum_{\ell=0}^L \mathbb{V}[\hat{T}_\ell^{\text{MLCV}}(\boldsymbol{\alpha}_\ell^*)] = \sum_{\ell=0}^L (1 - R_\ell^2) \mathbb{V}[\hat{T}_\ell^{(\ell)}] \quad (50)$$

with

$$R_\ell^2 = \mathbb{V}[\hat{T}_\ell^{(\ell)}]^{-1} \mathbb{C}[\hat{\mathbf{U}}_\ell^{(\ell)}, \hat{T}_\ell^{(\ell)}]^\top \boldsymbol{\alpha}_\ell^* \in [0, 1]. \quad (51)$$

Again, owing to the fact that  $R_\ell^2 \leq 1$ , the variance of the MLMC-MLCV estimator is always less than or equal to the variance of the MLMC estimator given by  $\mathbb{V}[\hat{\theta}_L^{\text{MLMC}}] = \sum_{\ell=0}^L \mathbb{V}[\hat{T}_\ell^{(\ell)}]$  (see eq. (30)).

In our surrogate-based approach,  $L + 1$  surrogate-based control variates can be used for the coarsest level  $\ell = 0$ , namely  $g_0, \dots, g_L$ , so that  $\boldsymbol{\alpha}_0 \in \mathbb{R}^{L+1}$ . At correction levels  $\ell > 0$ , we can use  $L$  control variates based on  $g_1, \dots, g_L$  and  $\tilde{g}_0, \dots, \tilde{g}_{L-1}$ , as described in section 3.2.2, so that  $\boldsymbol{\alpha}_\ell \in \mathbb{R}^L$ , for  $\ell > 0$ .

The MLMC-MLCV estimator of the expectation  $\mathbb{E}[f(\mathbf{X})]$  is defined by eq. (45), with

$$\hat{T}_0^{\text{MLCV}}(\boldsymbol{\alpha}_0) = \hat{E}^{(0)}[Y_0] - \boldsymbol{\alpha}_0^\top (\hat{E}^{(0)}[\mathbf{Z}] - \boldsymbol{\mu}_Z), \quad (52)$$

$$\hat{T}_\ell^{\text{MLCV}}(\boldsymbol{\alpha}_\ell) = \hat{E}^{(\ell)}[Y_\ell] - \hat{E}^{(\ell)}[Y_{\ell-1}] - \boldsymbol{\alpha}_\ell^\top (\hat{E}^{(\ell)}[\mathbf{W}] - \boldsymbol{\mu}_W), \quad \text{for } \ell > 0, \quad (53)$$

where

$$\mathbf{Z} = (Z_0, \dots, Z_L) = (g_0(\mathbf{X}), \dots, g_L(\mathbf{X})), \quad \boldsymbol{\mu}_{\mathbf{Z}} = \mathbb{E}[\mathbf{Z}], \quad (54)$$

$$\mathbf{W} = (W_1, \dots, W_L) = (h_1(\mathbf{X}), \dots, h_L(\mathbf{X})), \quad \boldsymbol{\mu}_{\mathbf{W}} = \mathbb{E}[\mathbf{W}]. \quad (55)$$

The optimal values  $\boldsymbol{\alpha}_\ell^*$  of the CV parameters are given by

$$\boldsymbol{\alpha}_0^* = \boldsymbol{\Sigma}_0^{-1} \mathbf{c}_0, \quad \boldsymbol{\Sigma}_0 = \mathbb{C}[\mathbf{Z}], \quad \mathbf{c}_0 = \mathbb{C}[\mathbf{Z}, Y_0], \quad (56)$$

$$\boldsymbol{\alpha}_\ell^* = \boldsymbol{\Sigma}_\ell^{-1} \mathbf{c}_\ell, \quad \boldsymbol{\Sigma}_\ell = \mathbb{C}[\mathbf{W}], \quad \mathbf{c}_\ell = \mathbb{C}[\mathbf{W}, Y_\ell - Y_{\ell-1}], \quad \text{for } \ell > 0, \quad (57)$$

resulting in  $R_\ell^2 = \mathbb{V}[Y_\ell - Y_{\ell-1}]^{-1} \mathbf{c}_\ell^\top \boldsymbol{\Sigma}_\ell \mathbf{c}_\ell$ , for  $\ell = 0, \dots, L$ . Note that  $\boldsymbol{\Sigma}_\ell = \mathbb{C}[\mathbf{W}]$  is the same for all  $\ell > 0$ . The optimal MLMC-MLCV variance estimator is derived in appendix B.

**Remark 3.** *In practice,  $\boldsymbol{\Sigma}$  and  $\mathbf{c}_\ell$  may be estimated using either a pilot sample or the same sample as for the estimation of  $\hat{\mathbf{U}}_\ell^{(\ell)}$  and  $\hat{T}_\ell^{(\ell)}$ . Alternatively, in the specific context of PC-based control variates, for the estimation of the expectation, a closed-form expression for  $\boldsymbol{\Sigma}$  can be obtained. Letting  $Z_\ell = g_\ell(\mathbf{X}) = \sum_{k=0}^{P_g^\ell} \mathfrak{g}_{\ell,k} \Psi_k(\mathbf{X})$  and  $W_\ell = h_\ell(\mathbf{X}) = \sum_{k=0}^{P_h^\ell} \mathfrak{h}_{\ell,k} \Psi_k(\mathbf{X})$ , we have*

$$\forall m, m' = 0, \dots, L, \quad [\boldsymbol{\Sigma}_0]_{m,m'} = \mathbb{C}[Z_m, Z_{m'}] = \sum_{k=1}^{\min(P_g^m, P_g^{m'})} \mathfrak{g}_{m,k} \mathfrak{g}_{m',k}, \quad (58)$$

$$\forall m, m' = 1, \dots, L, \quad [\boldsymbol{\Sigma}_\ell]_{m,m'} = \mathbb{C}[W_m, W_{m'}] = \sum_{k=1}^{\min(P_h^m, P_h^{m'})} \mathfrak{h}_{m,k} \mathfrak{h}_{m',k}, \quad \text{for } \ell > 0. \quad (59)$$

### 3.3 Practical details

A summary of the methods is presented in table 1. We remark that all the MLMC-like estimators (including the MLMC estimator), hereafter abbreviated MLMC-\*, have variance  $\sum_{\ell=0}^L (1 - R_\ell^2) \mathbb{V}[\hat{T}_\ell^{(\ell)}]$ , with

- $R_\ell^2 = 0$  for plain MLMC;
- $R_\ell^2 = \rho_\ell^2$  as defined in eq. (38) for MLMC-CV; and
- $R_\ell^2$  defined by eq. (51) for MLMC-MLCV.

For all these methods, we will further assume that  $\mathbb{V}[\hat{T}_\ell^{(\ell)}] \lesssim n_\ell^{-1} \mathcal{V}_\ell$  (see section 3.1), which implies that  $\mathbb{V}[\hat{T}_\ell^{\text{MLMC-*}}(\boldsymbol{\alpha}_\ell^*)] = (1 - R_\ell^2) \mathbb{V}[\hat{T}_\ell^{(\ell)}] \lesssim n_\ell^{-1} \mathcal{V}_\ell^{\text{CV}}$ , with  $\mathcal{V}_\ell^{\text{CV}} := (1 - R_\ell^2) \mathcal{V}_\ell$ .

In the surrogate-based variants of MLMC, the cost of evaluating the surrogate models is assumed to be negligible compared to the costs of evaluating the simulators  $f_0, \dots, f_L$ . Therefore, the total computational cost of the MLMC-\* estimator reduces to the cost of the MLMC estimator,  $\text{cost}(\hat{\theta}^{\text{MLMC-MLCV}}) = \text{cost}(\hat{\theta}^{\text{MLMC}})$ , given by eq. (31). Similarly to the case of the MLMC estimator (see, e.g., [36]), the optimal sample sizes  $(n_\ell^*)_{\ell=0}^L$  such that  $\sum_{\ell=0}^L \mathcal{V}_\ell^{\text{CV}} / n_\ell^*$  is minimal under a constrained computational budget of  $\mathcal{C}$  are given by

$$n_\ell^* = \frac{\mathcal{C}}{\mathcal{S}_L} \sqrt{\frac{\mathcal{V}_\ell^{\text{CV}}}{\mathcal{C}_\ell + \mathcal{C}_{\ell-1}}}, \quad \text{with } \mathcal{S}_\ell := \sum_{\ell'=0}^{\ell} \sqrt{(\mathcal{C}_{\ell'} + \mathcal{C}_{\ell'-1}) \mathcal{V}_{\ell'}^{\text{CV}}}, \quad (60)$$

Table 1: Summary of the methods. The first three are state-of-the-art methods, the next three are the novel multilevel methods proposed in this paper, and the last two are variants of the proposed methods.

Method	Form of the estimator	Eq.
<b>Monte Carlo (MC)</b> .	$\hat{\theta}$	
<b>Control Variates (CV)</b> [29, 31, 30, 38, 33].	$\hat{\theta} - \sum_{m=1}^M \alpha_m (\hat{\tau}_m - \tau_m)$ .	4
<b>Multilevel Monte Carlo (MLMC)</b> [21, 22].	$\hat{\theta}_0^{(0)} + \sum_{\ell=1}^L \hat{T}_\ell^{(\ell)}$ .	27
<b>Multilevel Control Variates (MLCV)</b> . CV method where the CVs are based on surrogate models of simulators of different levels of fidelity.	$\hat{\theta} - \sum_{\ell=0}^L \alpha_\ell (\hat{\tau}_\ell - \tau_\ell)$ .	4
<b>MLMC-CV</b> . MLMC with one CV at each correction level based on surrogate models of the simulators on the corresponding level. Corresponds to a surrogate-based MLMF [18, 19] where the exact statistics of the CVs are known.	$\hat{\theta}_0^{(0)} + \alpha_0 (\hat{\tau}_0^{(0)} - \tau_0) + \sum_{\ell=1}^L \left( \hat{T}_\ell^{(\ell)} + \alpha_\ell (\hat{U}_\ell^{(\ell)} - U_\ell) \right)$ .	35, 36
<b>MLMC-MLCV</b> . MLMC-CV with the CVs based on the surrogate models $g_0, \dots, g_L$ on level 0 and on $h_1, \dots, h_L$ on levels $\ell > 0$ .	$\hat{\theta}_0^{(0)} + \alpha_0^\top \begin{bmatrix} \hat{\tau}_0^{(0)} - \tau_0 \\ \vdots \\ \hat{\tau}_L^{(0)} - \tau_L \end{bmatrix} + \sum_{\ell=1}^L \left( \hat{T}_\ell^{(\ell)} + \alpha_\ell^\top \begin{bmatrix} \hat{U}_1^{(\ell)} - U_1 \\ \vdots \\ \hat{U}_L^{(\ell)} - U_L \end{bmatrix} \right)$ .	45, 46
<b>MLMC-CV[0]</b> . MLMC-CV using only one CV based on the surrogate $g_0$ on level 0 and no CVs on levels $\ell > 0$	$\hat{\theta}_0^{(0)} + \alpha_0 (\hat{\tau}_0^{(0)} - \tau_0) + \sum_{\ell=1}^L \hat{T}_\ell^{(\ell)}$ .	
<b>MLMC-MLCV[0]</b> . MLMC-MLCV using only CVs based on the surrogates $g_0$ and $g_1$ on level 0, and $h_1$ on levels $\ell > 0$ .	$\hat{\theta}_0^{(0)} + \alpha_0^\top \begin{bmatrix} \hat{\tau}_0^{(0)} - \tau_0 \\ \hat{\tau}_1^{(0)} - \tau_1 \end{bmatrix} + \sum_{\ell=1}^L \left( \hat{T}_\ell^{(\ell)} + \alpha_\ell (\hat{U}_1^{(\ell)} - U_1) \right)$ .	



so that  $\sum_{\ell=0}^L \mathcal{V}_\ell^{\text{CV}}/n_\ell^* = S_L^2/\mathcal{C}$ . As a consequence,

$$\mathbb{V}[\hat{\theta}_L^{\text{MLMC-MLCV}}(\boldsymbol{\alpha}_\ell^*)] \lesssim \frac{S_L^2}{\mathcal{C}}, \quad (61)$$

with an equality between the left- and right-hand sides for the expectation estimators.

In practice,  $\mathcal{V}_\ell^{\text{CV}}$  is not known and must be estimated for each level. In this work, we consider the sequential algorithm proposed in [36, Algorithm 2] for the MLMC. The algorithm starts from an initial, small number of samples  $n_\ell^{\text{init}}$  on each level. Then, it selects the optimal level on which to increase the sample size by an inflation factor  $r_\ell > 1$ , i.e. the level  $\ell^*$  on which the reduction in total variance relative to the additional computational effort achieved by inflating the sample size by  $r_{\ell^*}$  is maximal.

---

**Algorithm 1** Simplified MLMC-\* algorithm inspired by [36].

---

**Require:**  $n_\ell^{\text{init}} > 1$ ,  $r_\ell > 1$ , surrogate models (depending on the method), and budget  $\mathcal{C}$ .

- 1: Set consumed budget to  $\tilde{\mathcal{C}} = 0$  and  $\delta n_\ell = n_\ell^{\text{init}}$  samples on levels  $\ell \leq L$ ;
  - 2: **while**  $\tilde{\mathcal{C}} \leq \mathcal{C}$  **do**
  - 3:   compute  $\delta n_\ell$  samples on each level by evaluating  $f_\ell$  and the appropriate surrogates;
  - 4:   update sample size on each level:  $n_\ell \leftarrow n_\ell + \delta n_\ell$ ;
  - 5:   update consumed budget:  $\tilde{\mathcal{C}} \leftarrow \tilde{\mathcal{C}} + \sum_{\ell=0}^L \delta n_\ell (\mathcal{C}_\ell + \mathcal{C}_{\ell-1})$ ;
  - 6:   estimate the optimal CV parameter(s) on each level;
  - 7:   compute/update CV estimates for  $\hat{T}_\ell^{(\ell)}$  and  $\mathcal{V}_\ell^{\text{CV}}$  from samples on levels  $\ell \leq L$ ;
  - 8:   select level  $\ell^* = \arg \max_{0 \leq \ell \leq L} \frac{\mathcal{V}_\ell^{\text{CV}}}{r_\ell n_\ell^2 (\mathcal{C}_\ell + \mathcal{C}_{\ell-1})}$ ;
  - 9:    $\delta n_{\ell^*} \leftarrow \lfloor (r_{\ell^*} - 1)n_{\ell^*} \rfloor$ ,  $\delta n_{\ell \neq \ell^*} \leftarrow 0$ ;
  - 10: **end while**
  - 11: **return**  $\hat{\theta}_L^{\text{MLMC-*}}$ , the MLMC-\* estimate of  $\theta_L$ .
- 

## 4 Numerical experiments

We demonstrate the value of our MLMC-MLCV method on the uncertain heat equation problem proposed in [18] and summarized in section 4.1. The surrogate models used for the control variates are described in section 4.2, and the results from numerical experiments are presented and discussed in section 4.3.

### 4.1 Problem description

We consider the partial differential equation describing the time-evolution of the temperature  $u(x, t; \mathbf{X})$  in a 1D rod of unit length over the time interval  $[0, T]$ , with uncertain (random)

initial data  $u_0$  and thermal diffusivity  $\nu$ ,

$$\begin{cases} \frac{\partial u(x, t; \mathbf{X})}{\partial t} = \nu(\mathbf{X}) \frac{\partial^2 u(x, t; \mathbf{X})}{\partial x^2}, & x \in \mathcal{D} := (0, 1), \quad t \in [0, T], \\ u(x, 0; \mathbf{X}) = u_0(x; \mathbf{X}), & x \in \mathcal{D}, \\ u(0, t; \mathbf{X}) = u(1, t; \mathbf{X}) = 0, & t \in [0, T], \end{cases} \quad (62)$$

where  $\mathbf{X}: \Omega \rightarrow \Xi$  is a random vector modelling the uncertainty in the input parameters, and where  $\nu(\mathbf{X}) > 0$  almost surely. The solution of eq. (62) may be expressed as

$$u(x, t; \mathbf{X}) = \sum_{k=1}^{\infty} a_k(\mathbf{X}) \exp(-\nu(\mathbf{X})k^2\pi^2t) \sin(k\pi x) \quad (63)$$

with

$$a_k(\mathbf{X}) = 2 \int_{\mathcal{D}} u_0(x; \mathbf{X}) \sin(k\pi x) dx. \quad (64)$$

The initial condition is chosen to have the same prescribed form as in [18]. Specifically, we consider  $u_0(x; \mathbf{X}) = \mathcal{G}(\mathbf{X})\mathcal{F}_1(x) + \mathcal{I}(\mathbf{X})\mathcal{F}_2(x)$  with

$$\mathcal{F}_1(x) = \sin(\pi x), \quad (65)$$

$$\mathcal{F}_2(x) = \sin(2\pi x) + \sin(3\pi x) + 50(\sin(9\pi x) + \sin(21\pi x)), \quad (66)$$

$$\mathcal{I}(\mathbf{X}) = \frac{7}{2} [\sin(X_1) + 7 \sin(X_2)^2 + 0.1X_3^4 \sin(X_1)], \quad (67)$$

$$\mathcal{G}(\mathbf{X}) = 50(4|X_5| - 1)(4|X_6| - 1)(4|X_7| - 1), \quad (68)$$

which allows to control the spectral content of the solution  $u$ . Furthermore, as in [18], the diffusion coefficient is modelled by  $\nu(\mathbf{X}) = X_4$ . The random output variable of interest is defined as the integral of the temperature along the rod at final time  $T$ ,

$$\mathcal{M}(\mathbf{X}) = \int_{\mathcal{D}} u(x, T; \mathbf{X}) dx \quad (69)$$

$$= \sum_{k=1}^{\infty} a_k(\mathbf{X}) \int_{\mathcal{D}} \exp(-\nu(\mathbf{X})k^2\pi^2T) \sin(k\pi x) dx \quad (70)$$

$$= \mathcal{G}(\mathbf{X})\mathcal{H}_1(\mathbf{X}) + \mathcal{I}(\mathbf{X}) [\mathcal{H}_3(\mathbf{X}) + 50\mathcal{H}_9(\mathbf{X}) + 50\mathcal{H}_{21}(\mathbf{X})], \quad (71)$$

where  $\mathcal{H}_k(\mathbf{X}) = \frac{2}{k\pi} \exp(-\nu(\mathbf{X})k^2\pi^2T)$ . In this experiment, we seek to estimate the expectation  $\mathbb{E}[\mathcal{M}(\mathbf{X})]$ , for a given uncertain setting. Consistently with [18], we consider the random variables  $X_1, \dots, X_7$  to be independent and distributed as

$$X_1, X_2, X_3 \sim \mathcal{U}[-\pi, \pi], \quad X_4 \sim \mathcal{U}[\nu_{\min}, \nu_{\max}], \quad X_5, X_6, X_7 \sim \mathcal{U}[-1, 1]. \quad (72)$$

The expected value  $\mathbb{E}[\mathcal{M}(\mathbf{X})]$  is then given by

$$\mathbb{E}[\mathcal{M}(\mathbf{X})] = 50H_1 + \frac{49}{4} (H_3 + 50H_9 + 50H_{21}), \quad (73)$$

where

$$H_k = \mathbb{E}[\mathcal{H}_k(\mathbf{X})] = \frac{2}{k^3 \pi^3 T} \frac{\exp(-\nu_{\min} k^2 \pi^2 T) - \exp(-\nu_{\max} k^2 \pi^2 T)}{\nu_{\max} - \nu_{\min}}. \quad (74)$$

Finally, we set  $T = 0.5$ ,  $\nu_{\min} = 0.001$  and  $\nu_{\max} = 0.009$ , resulting in  $\mathbb{E}[\mathcal{M}(\mathbf{X})] \approx 41.98$ .

Numerically,  $\mathcal{M}(\mathbf{X})$  is approximated by truncating the Fourier expansion in eq. (70) to  $K < \infty$  modes and by approximating the integrals in eqs. (64) and (70) by a trapezoidal quadrature rule with equispaced nodes in  $[0, 1]$ . The multilevel hierarchy of simulators  $\{f_\ell\}_{\ell=0}^L$  is then defined according to the number of quadrature nodes  $N_\ell$  used for the approximation at level  $\ell$ . Specifically,  $\mathcal{M}(\mathbf{X})$  is approximated at level  $\ell$  by

$$Y_\ell = f_\ell(\mathbf{X}) = \sum_{k=1}^K A_k^\ell(\mathbf{X}) B_k^\ell(\mathbf{X}), \quad (75)$$

with

$$A_k^\ell(\mathbf{X}) = 2 \sum_{i=1}^{N_\ell} w_i u_0(x_i; \mathbf{X}) \sin(k\pi x_i), \quad B_k^\ell(\mathbf{X}) = \exp(-\nu(\mathbf{X}) k^2 \pi^2 T) \sum_{i=1}^{N_\ell} w_i \sin(k\pi x_i), \quad (76)$$

where  $\{(x_i, w_i)\}_{i=1}^{N_\ell}$  are the pairs of quadrature nodes and associated weights on level  $\ell$ . It is then natural to assume that the computational cost  $\mathcal{C}_\ell$  of an evaluation of  $f_\ell$  is  $\mathcal{O}(KN_\ell)$ . The statistic of interest is thus  $\theta = \theta_L = \mathbb{E}[f_L(\mathbf{X})]$ , whose MC estimator  $\hat{\theta}_L = n_L^{-1} \sum_{i=1}^{n_L} f_L(\mathbf{X}^{(L,i)})$  will represent the baseline estimator for our experiments. Besides, the quality of all the presented estimators will be assessed in terms of their root mean square error (RMSE) w.r.t. the exact statistic  $\mathbb{E}[\mathcal{M}(\mathbf{X})]$  given by eq. (73).

In the following experiments, we set the number of quadrature nodes  $K = 21$  and the number of levels to 4 (i.e.  $L = 3$ ). Furthermore, we set  $N_\ell = 120 \times 2^{L-\ell}$  so that evaluating  $f_\ell$  is twice as expensive as evaluating  $f_{\ell-1}$ . Table 2 summarizes the number of quadrature nodes and the evaluation cost per level. Note that the costs are normalized so that  $\mathcal{C}_3 = 1$ .

Table 2: Number of quadrature nodes  $N_\ell$ , simulator evaluation cost  $\mathcal{C}_\ell$  and MLMC correction evaluation cost  $\mathcal{C}_\ell + \mathcal{C}_{\ell-1}$  per level.

$\ell$	0	1	2	3
$N_\ell$	15	30	60	120
$\mathcal{C}_\ell$	0.125	0.25	0.5	1
$\mathcal{C}_\ell + \mathcal{C}_{\ell-1}$	0.125	0.375	0.75	1.5

## 4.2 Surrogate models

We will mostly use PC models (see section 2.4.3) for the surrogate-based CV estimators. Constructing high-quality surrogates in 7 dimensions can be hard, especially in the presence of non-linearities and with a limited sample size. To avoid overfitting, we resort to the *least angle regression* (LARS) procedure [14], which is a model-selection regression method that

promotes sparsity. More precisely, we employ the basis-adaptive hybrid LARS algorithm proposed by [7, Fig. 5] for the selection of sparse PC bases. For a given design of experiment (DoE) for the PC surrogate construction, this algorithm applies the LARS procedure on candidate PC bases  $\mathcal{A}_p$  of increasing total polynomial degree  $p = 1, \dots, p_{\max}$ , resulting for each  $p$  in the selection of a limited number  $|\tilde{\mathcal{A}}_p| < |\mathcal{A}_p|$  of basis polynomial functions. For each  $p$ , a PC surrogate is constructed by classical least-squares regression on the corresponding reduced (or active) PC basis  $\tilde{\mathcal{A}}_p$ , and its quality is estimated using a corrected leave-one-out cross-validation procedure [7]. The best surrogate according to this quality measure is eventually retained, and the associated reduced PC basis is denoted by  $\tilde{\mathcal{A}}_{p^*}$ .

In our experiments, we set  $p_{\max} = 16$  and the construction budget to 400 times the evaluation cost of  $f_3$ , i.e.  $\mathcal{C}^{\text{DoE}} = 400$ . This budget is distributed equally among the different levels, so that the associated evaluation cost on each level corresponds to the cost of 100  $f_3$  evaluations, i.e.  $n_\ell^{\text{DoE}} \mathcal{C}_\ell = 100$ , as reported in tables 3 and 4. Once constructed, the quality of a surrogate  $g$  of  $f$  is assessed in terms of the  $Q^2$  measure,

$$Q^2(g, f) = 1 - \frac{\mathbb{E}[(g(\mathbf{X}) - f(\mathbf{X}))^2]}{\mathbb{V}[f(\mathbf{X})]}, \quad (77)$$

estimated using a test sample of size  $n_{\text{test}} = 10\,000$ , which is more robust than the corrected leave-one-out measure used for the model selection. The higher the  $Q^2$  value, the higher-quality the associated surrogate.

For the MLCV estimator, we learn a PC model  $g_\ell$  of  $f_\ell$  from a training DoE of size  $n_\ell^{\text{DoE}}$  generated by latin hypercube sampling (LHS) improved by simulated annealing. The training DoE samples for the surrogate construction are different for each level and independent. Table 3 summarizes the properties of the different PC models. Except for  $g_3$ , which is built with only 100 points, all the PC models have a good  $Q^2$ , greater than 0.8. We observe that the basis-adaptive LARS algorithm has selected a decreasing polynomial degree  $p^*$  with level  $\ell$ , while retaining only a limited number  $|\tilde{\mathcal{A}}_{p^*}|$  of polynomials in these reduced bases. Thus, although  $g_1$  and  $g_2$  have distinct degrees, the sizes of the associated reduced bases are similar.

Table 3: PC models for CV and MLCV estimators, with their sample size, degree and quality measure. These models are built with the basis-adaptive LARS algorithm of [7, Fig. 5].

PC models	$g_0$	$g_1$	$g_2$	$g_3$
$n_\ell^{\text{DoE}}$	800	400	200	100
$p^*$	10	8	6	4
$ \tilde{\mathcal{A}}_{p^*} $	211	77	74	14
$Q^2$	0.98	0.94	0.86	0.59

For the MLMC-based estimators, we learn a PC model  $g_\ell$  of  $f_\ell$  for  $\ell = 0, \dots, L$  and a PC model  $h_\ell$  of  $f_\ell - f_{\ell-1}$  for  $\ell = 1, \dots, L$ . In practice, training points used for the construction of  $g_0, \dots, g_L$  may be reused for the construction of  $h_1, \dots, h_L$  using nested DoEs. However, it should be noted that the generation of nested LHS DoEs is not as straightforward as for purely random DoEs. While using nested random DoEs is a perfectly valid strategy, we opt for an alternative choice based on LHS. Specifically, we first generate a DoE  $\mathcal{X}_{\text{DoE}}^{(0)}$  of size  $n_0^{\text{DoE}}$  using

LHS improved by simulated annealing. Then, for  $\ell = 1, \dots, L$ , we sequentially extract a DoE  $\mathcal{X}_{\text{DoE}}^{(\ell)} \subset \mathcal{X}_{\text{DoE}}^{(\ell-1)}$  of size  $n_\ell^{\text{DoE}}$ . The subset  $\mathcal{X}_{\text{DoE}}^{(\ell)}$  is selected such that it has minimal centered  $L^2$  discrepancy among a pool of 1 million random candidate subsets. Table 4 summarizes the properties of these different PC models. The PC models again have a quality measure higher than 0.8, except for  $g_3$  and  $h_3$ , which were constructed from only 100 evaluations. Note that the values of  $p^*$ ,  $|\tilde{\mathcal{A}}_{p^*}|$  and  $Q^2$  are of the same order of magnitude as those of the PC surrogate models reported in table 3.

Table 4: PC models for the estimators combining MLMC and CV techniques, with their sample size, degree and quality measure. More precisely,  $g_\ell$  is the surrogate of  $f_\ell$ ,  $h_\ell$  is the surrogate of  $f_\ell - f_{\ell-1}$  and the samples used to train  $h_\ell$  are also used to train  $g_\ell$  and  $g_{\ell-1}$ . These models are built with the basis-adaptive LARS algorithm of [7, Fig. 5].

PC models	$g_0$	$g_1$	$g_2$	$g_3$	$h_1$	$h_2$	$h_3$
$n_\ell^{\text{DoE}}$	800	400	200	100	400	200	100
$p^*$	10	8	8	2	9	7	5
$ \tilde{\mathcal{A}}_{p^*} $	211	83	57	6	112	71	30
$Q^2$	0.98	0.95	0.86	0.21	0.99	0.82	0.63

Hereafter, the CV and MLCV estimators use the PC models of table 3, while the MLMC-based estimators use the PC models presented in table 4.

### 4.3 Results

First, in section 4.3.1, we illustrate the use of one or several control variates to reduce the variance of a single-level MC estimator. Then, we compare the MLCV and MLMC-MLCV approaches with the MC and MLMC estimators in section 4.3.2, and we discuss variants of the MLMC-CV and MLMC-MLCV approaches, considering only a limited subset of the surrogate models, in section 4.3.3. We conclude the analysis by reporting the estimation budget allocation across levels resulting from the various MLMC-based methods in section 4.3.4. In practice, all the MLMC-based estimators are built using algorithm 1. Unless stated otherwise, the parameters are set to  $n_\ell^{\text{init}} = 30$  and  $r_\ell = 1.1$  for  $\ell = 0, \dots, 3$ . The quality of the various estimators will be assessed in terms of their RMSE w.r.t.  $\mathbb{E}[\mathcal{M}(\mathbf{X})]$ , estimated from 500 repetitions of the experiment.

#### 4.3.1 Single-level MC and control variates

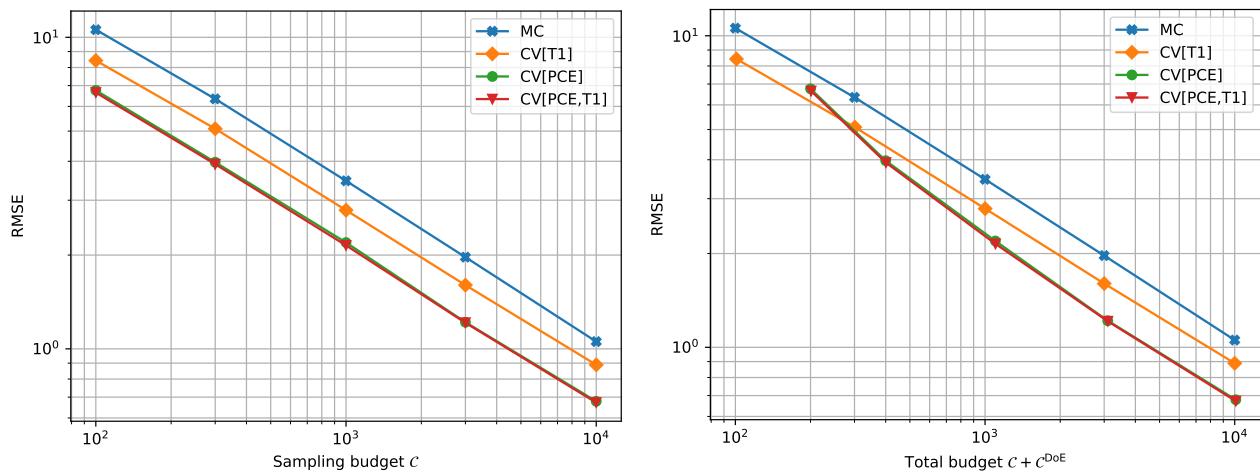
In this first part, we consider only the finest simulator  $f_3$  and try to reduce the variance of the MC estimator of  $\mathbb{E}[f_3(\mathbf{X})]$  by means of surrogate-based CVs. For that purpose, we consider a first-order Taylor polynomial expansion  $g_3^{\text{T1}}$  around  $\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}[\mathbf{X}]$  (see section 2.4.2 and appendix C) and the PC model  $g_3^{\text{PC}}$  described in table 3.

Table 5 shows that  $g_3^{\text{PC}}(\mathbf{X})$  is well correlated with  $Y_3 = f_3(\mathbf{X})$ , with a Pearson coefficient of 0.8, while  $g_3^{\text{T1}}$  is less so, with a coefficient of 0.57, which may be explained by the strong non-linearity of  $f_3$ . According to eq. (7), these correlation coefficients lead to a theoretical variance reduction factor  $R^2$  of 64% when using  $g_3^{\text{PC}}(\mathbf{X})$  as a single CV, and of about 32%

Table 5: Pearson correlation coefficients between the finest simulator  $f_3$  and the corresponding PC model (see table 3) and  $T_1$  model around  $\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}[\mathbf{X}]$ , estimated with a sample of size 1000.

	$Y_3$	$g_3^{\text{PC}}(\mathbf{X})$	$g_3^{\text{T}_1}(\mathbf{X})$
$Y_3$	1.00	0.80	0.57
$g_3^{\text{PC}}(\mathbf{X})$	0.80	1.00	0.64
$g_3^{\text{T}_1}(\mathbf{X})$	0.57	0.64	1.00

when using  $g_3^{\text{T}_1}(\mathbf{X})$  as a single CV. Using both CVs results in a reduction factor of 65%. This minor increase in  $R^2$  can be explained by the modest correlation coefficient of 0.64 between  $g_3^{\text{T}_1}(\mathbf{X})$  and  $g_3^{\text{PC}}(\mathbf{X})$ . Note that a reduction factor of  $R^2$  in the variance corresponds to a reduction factor of  $1 - \sqrt{1 - R^2}$  in the standard deviation.



(a) The evaluation cost does not include the cost of constructing the surrogate models.

(b) The evaluation cost includes the cost of constructing the surrogate models.

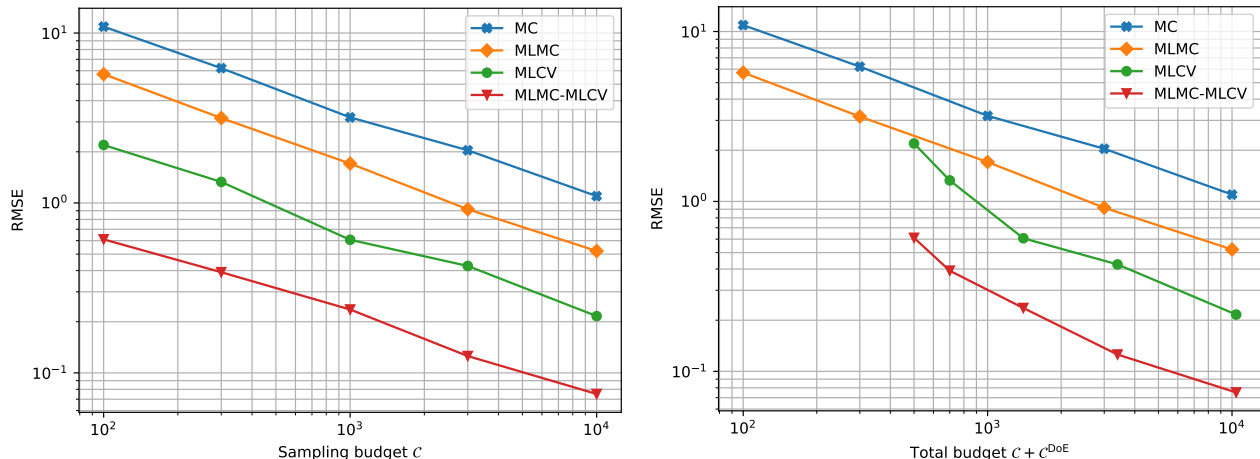
Figure 1: Convergence comparison of MC and CV estimators of  $\theta = \mathbb{E}[f_3(\mathbf{X})]$  (see eq. (75)) with respect to the number  $\mathcal{C} \in \{100; 300; 1000; 3000; 10\,000\}$  of  $f_3$  evaluations. The CV uses either a first-order Taylor approximation (T1) of  $f_3$  or a PC approximation of  $f_3$  trained from 100 evaluations ( $\mathcal{C}^{\text{DoE}} = 100$ ), or both. The RMSEs are computed using 500 replicates.

These theoretical expectations are reflected in fig. 1a with an RMSE reduction of about 20% when using  $g_3^{\text{T}_1}$  and 40% when using  $g_3^{\text{PC}}$  alone or jointly with  $g_3^{\text{T}_1}$ . This figure confirms that  $g_3^{\text{PC}}$  provides a better CV than  $g_3^{\text{T}_1}$ , reducing the RMSE of the MC estimator twice as much, regardless of the computational budget. However, the construction cost of the surrogate is not the same. While constructing  $g_3^{\text{T}_1}$  requires only one evaluation of  $f_3$  and its Jacobian matrix, namely at  $\boldsymbol{\mu}_{\mathbf{X}}$ , the construction of  $g_3^{\text{PC}}$  involved 100  $f_3$  evaluations. In the case where the surrogate is built specifically for the estimation of the statistic, the real estimation cost is higher as it includes this construction cost. Figure 1b illustrates this difference by including the surrogate construction cost in the total evaluation cost. As a result, a significant offset appears when using  $g_3^{\text{PC}}$ , since part of the computational budget (namely

100) is used for the surrogate construction and thus not for the estimation. Specifically, the RMSEs of the CV estimators using  $g_3^{\text{PC}}$  get below that of the CV estimator using only  $g_3^{\text{T1}}$  for a budget of 300  $f_3$  evaluations. For budgets under 200, using only  $g_3^{\text{T1}}$  is preferable, even without the analytical gradient available, which would then incur a construction cost of 8  $f_3$  evaluations to approximate the gradient using finite differences. This figure also illustrates proposition 1, that is, increasing the number of control variates improves (rigorously speaking, does not deteriorate) the variance of the CV estimator.

### 4.3.2 MLCV and MLMC-MLCV

Figure 2 compares the MLCV and MLMC-MLCV estimators proposed in eqs. (34) and (45) with the classical MC and MLMC estimators. This comparison is repeated for different budgets  $\mathcal{C}$ , expressed in terms of the equivalent number of  $f_3$  evaluations. Note that the MC and MLCV estimators only use evaluations of the finest simulator  $f_3$ , so that  $\mathcal{C} = n_3$ , while MLMC-based estimators use all the simulators, so that  $\mathcal{C}$  is given in terms of the MLMC cost eq. (31), indeed corresponding to the equivalent number of  $f_3$  evaluations, since the costs are normalized such that  $\mathcal{C}_3 = 1$ . From here on, only PC-based surrogates will be used, so that we omit the superscript “PC” in the notations of the surrogate models.



(a) The evaluation cost does not include the cost of construction of the surrogate models.

(b) The evaluation cost includes the cost of construction of the surrogate models.

Figure 2: Convergence comparison of MC, MLMC, MLCV and MLMC-MLCV estimators of  $\theta = \mathbb{E}[f_3(\mathbf{X})]$  (see eq. (75)) with respect to the sampling budget  $\mathcal{C} \in \{100; 300; 1000; 3000; 10\,000\}$  of  $f_3$  evaluations. The surrogates used for the MLCV and MLMC-MLCV estimators are described in tables 3 and 4, respectively, and their total construction cost is  $\mathcal{C}^{\text{DoE}} = 400$ . The RMSEs are estimated using 500 replicates.

Figure 2a shows that, for a given sampling budget  $\mathcal{C}$ , the MC estimator is the least accurate. This can be explained by the fact that it only has access to the finest simulator,  $f_3$ , whose cost only allows a limited number of evaluations. On the contrary, the MLMC estimator spreads this sampling budget over the four simulators. Ideally, the optimal sample allocation of MLMC eq. (60) results in many coarse, cheap evaluations and few fine, expensive evaluations. This is typically the case when the outputs of the simulators are highly correlated

and the associated computational cost grows exponentially. Table 7 shows that the first assumption holds. The second assumption also holds since  $\mathcal{C}_\ell = \mathcal{O}(KN_\ell) = \mathcal{O}(N_\ell)$ ,  $K$  being fixed, i.e., the evaluation cost grows linearly with the number of quadrature nodes. As a consequence, the MLMC estimator has a lower RMSE than the standard MC estimator.

Table 6: Pearson correlation coefficients between the finest output  $Y_3 = f_3(\mathbf{X})$  and the corresponding control variates based on the PC models of table 3 used for the MLCV method, estimated with a sample of size 1000.

	$Y_3$	$g_0(\mathbf{X})$	$g_1(\mathbf{X})$	$g_2(\mathbf{X})$	$g_3(\mathbf{X})$
$Y_3$	1.00	0.96	0.97	0.93	0.80
$g_0(\mathbf{X})$	0.96	1.00	0.96	0.92	0.80
$g_1(\mathbf{X})$	0.97	0.96	1.00	0.95	0.81
$g_2(\mathbf{X})$	0.93	0.92	0.95	1.00	0.80
$g_3(\mathbf{X})$	0.80	0.96	0.81	0.80	1.00

For this experiment, the MLCV estimator is more accurate than the MLMC estimator. Based on one control variate per level  $\ell$ , based on the PC model  $g_\ell$  of  $f_\ell$  from table 3, this estimator dedicates all the sampling budget to the finest simulator  $f_3$ , and uses these control variates based on  $g_0$ ,  $g_1$ ,  $g_2$  and  $g_3$  to reduce the variance of the MC estimator at no extra cost, as the evaluation cost of a PC model  $g_\ell$  is negligible compared to the evaluation cost of  $f_3$ . This MLCV technique works particularly well in this case because the control variates are highly correlated to  $f_3$ . Indeed, table 6 shows that their Pearson coefficients are at least 0.8, which guarantees a theoretical reduction of at least 94% in the variance of the MC estimator (i.e. a reduction of at least about 76% in standard deviation), corresponding to the variance reduction when using a single control variate based on  $g_3$ . In fact, the variance reduction factor  $R^2$  when using all the surrogates is only slightly higher, namely  $R^2 \approx 95\%$  corresponding to a standard reduction factor of around 78%, which is reflected in fig. 2a.

Combining the MLMC and MLCV techniques allows the resulting MLMC-MLCV estimator to reduce the variance even more significantly. This can be explained by the very high correlation between  $Y_0$ ,  $Y_1$ ,  $Y_2$  and  $Y_3$  on the one hand, which ensures the good performance of the MLMC approach, and by the strong correlation between the control variates on the other hand, ensuring their good performance in combination with the MLMC technique. In particular, table 7 shows that  $g_0(\mathbf{X})$ ,  $g_1(\mathbf{X})$  and  $g_2(\mathbf{X})$  are highly correlated with  $Y_3$ , with Pearson correlation coefficients greater than 0.9, while  $g_3(\mathbf{X})$  is poorly correlated with  $Y_3$ , with a correlation coefficient of 0.48. Besides,  $h_1(\mathbf{X})$ ,  $h_2(\mathbf{X})$  and  $h_3(\mathbf{X})$  are well-correlated with  $Y_1 - Y_0$ ,  $Y_2 - Y_1$  and  $Y_3 - Y_2$ , respectively, with correlation coefficients greater than 0.8.

Further insights regarding the expected variance reduction of the MLMC-MLCV estimator can be drawn from table 8, which reports the variance reduction factor  $R_\ell^2$  w.r.t. pure MLMC on each level defined in eqs. (50) and (51), as well as the quantity  $\mathcal{S}_\ell^2$  defined in eq. (60). In particular,  $\mathcal{S}_L^2/\mathcal{C}$  corresponds to the variance of the MLMC-MLCV estimator of the expectation with optimal sample allocation (see eq. (61)), so that the ratio between the variance of MLMC-MLCV estimator and that of the MC estimator is  $\mathcal{S}_L^2/(\mathcal{C}_L\mathbb{V}[Y_L])$  (here with  $\mathcal{C}_L = \mathcal{C}_3 = 1$ ). Furthermore, the variance of the different MLMC-based estimators



Table 7: Pearson correlation coefficients between the outputs  $Y_\ell = f_\ell(\mathbf{X})$  of the simulators and  $g_\ell(\mathbf{X})$  of the PC models, as well as the successive differences  $Y_\ell - Y_{\ell-1}$  and the outputs  $h_\ell(\mathbf{X})$  of the associated PC models, for the MLMC-based estimators. The coefficients are estimated from a sample of size 1000.

	$Y_0$	$Y_1$	$Y_2$	$Y_3$	$Y_1 - Y_0$	$Y_2 - Y_1$	$Y_3 - Y_2$	$g_0(\mathbf{X})$	$g_1(\mathbf{X})$	$g_2(\mathbf{X})$	$g_3(\mathbf{X})$	$h_1(\mathbf{X})$	$h_2(\mathbf{X})$	$h_3(\mathbf{X})$
$Y_0$	1.00	0.97	0.97	0.97	-0.11	-0.04	-0.04	0.99	0.96	0.93	0.47	-0.12	-0.13	-0.10
$Y_1$	0.97	1.00	1.00	1.00	0.13	0.19	0.19	0.97	0.97	0.93	0.48	0.12	0.09	0.09
$Y_2$	0.97	1.00	1.00	1.00	0.14	0.20	0.20	0.96	0.97	0.92	0.48	0.13	0.10	0.10
$Y_3$	0.97	1.00	1.00	1.00	0.14	0.20	0.20	0.96	0.97	0.92	0.48	0.13	0.10	0.10
$Y_1 - Y_0$	-0.11	0.13	0.14	0.14	1.00	0.96	0.96	-0.10	0.07	0.01	0.03	0.99	0.90	0.80
$Y_2 - Y_1$	-0.04	0.19	0.20	0.20	0.96	1.00	1.00	-0.03	0.12	0.06	0.06	0.94	0.90	0.81
$Y_3 - Y_2$	-0.04	0.19	0.20	0.20	0.96	1.00	1.00	-0.03	0.12	0.06	0.06	0.95	0.90	0.81
$g_0(\mathbf{X})$	0.99	0.97	0.96	0.96	-0.10	-0.03	-0.03	1.00	0.96	0.93	0.48	-0.11	-0.12	-0.09
$g_1(\mathbf{X})$	0.96	0.97	0.97	0.97	0.07	0.12	0.12	0.96	1.00	0.94	0.47	0.06	0.04	0.04
$g_2(\mathbf{X})$	0.93	0.93	0.92	0.92	0.01	0.06	0.06	0.93	0.94	1.00	0.47	-0.00	-0.03	-0.02
$g_3(\mathbf{X})$	0.47	0.48	0.48	0.48	0.03	0.06	0.06	0.48	0.47	0.47	1.00	0.03	0.01	0.02
$h_1(\mathbf{X})$	-0.12	0.12	0.13	0.13	0.99	0.94	0.95	-0.11	0.06	-0.00	0.03	1.00	0.91	0.80
$h_2(\mathbf{X})$	-0.13	0.09	0.10	0.10	0.90	0.90	0.90	-0.12	0.04	-0.03	0.01	0.91	1.00	0.80
$h_3(\mathbf{X})$	-0.10	0.09	0.10	0.10	0.80	0.81	0.81	-0.09	0.04	-0.02	0.02	0.80	0.80	1.00

per unit cost can be compared directly through  $S_L^2$ . We observe that the variance of the MLMC-MLCV estimator is significantly reduced compared to that of the MLMC estimator, by about 98.6%, resulting in a reduction in standard deviation of about 88%. Again, this is well reflected in fig. 2a.

These first results from fig. 2a highlight the interest of multilevel control variates, be it with the MLMC-MLCV estimator or simply with the MLCV one. These results suppose that the PC models are not built specifically for the study, so that the budget does not include the number of  $f_3$ -equivalent simulations required for their construction. Figure 2b illustrates the alternative case where the cost of the surrogate based CV estimators includes the cost of constructing the surrogates. As a result, an additional budget of  $\mathcal{C}^{\text{DoE}} = 400$  (see section 4.2), is allocated to the construction of the PC models. As was the case for the single-level PC-based CV estimators, this results in an offset of 400 in the total cost of the MLCV and MLMC-MLCV estimators. The effect is especially noticeable when the construction budget larger than the estimation budget, i.e. for  $\mathcal{C} \in \{100; 300\}$ . For a sampling cost of 100, i.e. a total evaluation cost of 500, the MLCV estimator is still slightly more accurate than the MLMC estimator, while the MLMC-MLCV estimator still largely outperforms both.

### 4.3.3 Variants of MLMC-based CV estimators

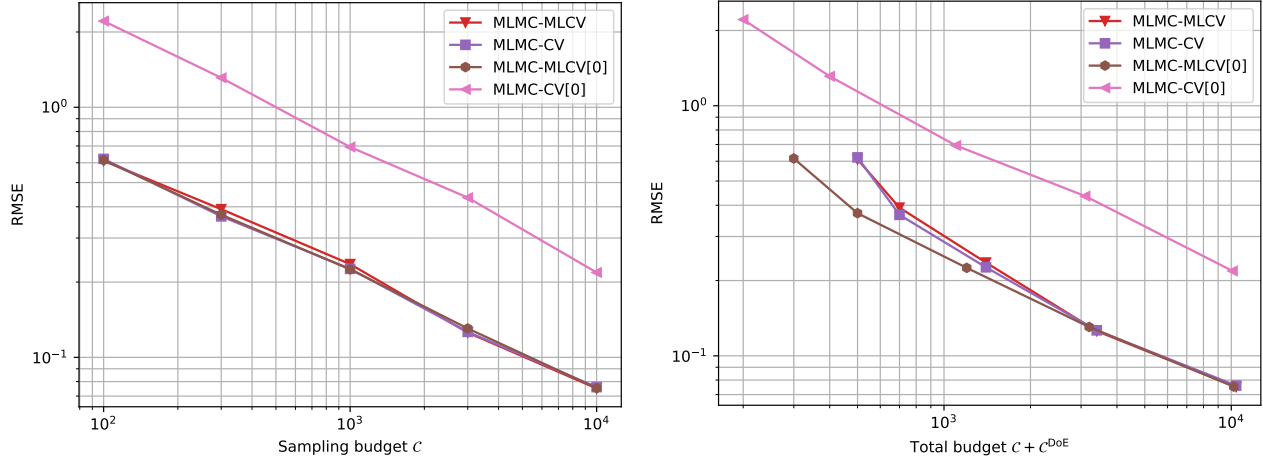
The discussion about the surrogate construction budget prompts us to investigate variants of MLMC-MLCV using fewer surrogates, in order to reduce the total evaluation cost for limited budgets. The MLMC-CV estimator is described in section 3.2.2, while the variants MLMC-CV[0] and MLMC-MLCV[0] of MLMC-CV and MLMC-MLCV are introduced in table 1. The MLMC-CV[0] estimator only uses a control variate based on  $g_0$ , so that the construction cost drops to  $\mathcal{C}^{\text{DoE}} = 100$ , while the MLMC-MLCV[0] estimator uses control variates based on

Table 8: Relevant quantities for the MLMC-based estimators, with  $n_\ell^*$  and  $\mathcal{S}_\ell^2$  as defined in eq. (60), and where  $\mathcal{V}_\ell$  and  $R_\ell^2$  have been estimated using an independent sample of size 10 000.

$\ell$	0	1	2	3	
$\mathcal{V}_\ell = \mathbb{V}[Y_\ell - Y_{\ell-1}]$	$1.0850 \times 10^4$	$5.9029 \times 10^2$	1.0590	$5.8160 \times 10^{-2}$	
$R_\ell^2$	MLMC	0	0	0	
	MLMC-CV	0.9838	0.9916	0.8224	0.6469
	MLMC-MLCV	0.9840	0.9920	0.9173	0.9202
	MLMC-CV[0]	0.9838	0	0	0
	MLMC-MLCV[0]	0.9840	0.9916	0.8992	0.9027
$n_\ell^* \frac{\mathcal{C}_\ell + \mathcal{C}_{\ell-1}}{\mathcal{C}}$	MLMC	69.63 %	28.13 %	1.68 %	0.56 %
	MLMC-CV	71.00 %	20.66 %	5.69 %	2.66 %
	MLMC-MLCV	73.66 %	20.97 %	4.05 %	1.32 %
	MLMC-CV[0]	22.59 %	71.69 %	4.29 %	1.42 %
	MLMC-MLCV[0]	72.84 %	21.30 %	4.42 %	1.44 %
$\mathcal{S}_\ell^2$	MLMC	1356.31	2673.54	2766.49	2797.65
	MLMC-CV	21.99	36.64	41.33	43.62
	MLMC-MLCV	21.72	35.85	38.99	40.04
	MLMC-CV[0]	21.99	382.87	418.54	430.71
	MLMC-MLCV[0]	21.76	36.35	39.84	41.01

$g_0$ ,  $g_1$  and  $h_1$ , so that the construction cost drops to  $\mathcal{C}^{\text{DoE}} = 200$ . The MLMC-CV estimator uses control variates based on surrogates at all levels, so that the construction cost remains  $\mathcal{C}^{\text{DoE}} = 400$ .

Figure 3a shows that MLMC-CV[0] has much higher RMSE than the other variants, resulting from the fact that it only reduces the variance associated with the coarsest level of the MLMC estimator. This behavior is consistent with the quantities of table 8. In particular, the value of  $\mathcal{S}_\ell^2$  is about 10 times higher than for the other variants, accounting for its RMSE being about 3 times higher than for the other variants. The remaining variants have similar performances regardless of the estimation budget, which is consistent with the  $\mathcal{S}_\ell^2$  values given in table 8. On the other hand, when considering the construction cost of the surrogates, fig. 3b shows that MLMC-MLCV[0] performs best, as it uses only surrogate models related to the two coarsest levels, namely  $g_0$ ,  $g_1$  and  $h_1$ , so that the construction cost is reduced. Furthermore, these surrogates have excellent  $Q^2$ , and they are such that  $g_0(\mathbf{X})$  is highly correlated with  $Y_0$ ,  $g_1(\mathbf{X})$  is highly correlated with  $Y_1$ , and  $h_1(\mathbf{X})$  is highly correlated with  $Y_1 - Y_0$ ,  $Y_2 - Y_1$  and  $Y_3 - Y_2$ . Namely, the associated Pearson correlation coefficients reported in table 7 are all at least 0.94. Therefore, should one have to build the surrogates specifically for the CV estimation of a statistic, it is more advantageous to adopt the MLMC-MLCV[0] variant over the others. In our case, the construction budget is divided by two compared to MLMC-MLCV and MLMC-CV, for a similar performance in terms of RMSE.



(a) The evaluation cost does not include the cost of construction of the surrogate models. (b) The evaluation cost includes the cost of construction of the surrogate models.

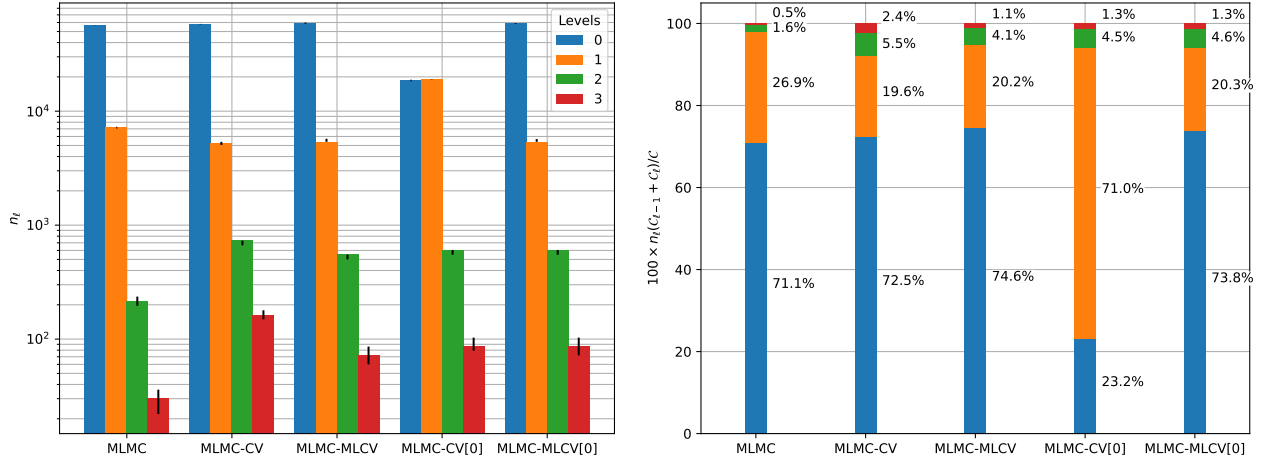
Figure 3: Convergence comparison of the MLMC-MLCV, MLMC-CV, MLMC-MLCV[0], and MLMC-CV[0] estimators of  $\theta = \mathbb{E}[f_3(\mathbf{X})]$  (see eq. (75)) with respect to the sampling budget  $\mathcal{C} \in \{100; 300; 1000; 3000; 10\,000\}$  of  $f_3$  evaluations. The surrogates used for the MLMC-based estimators are described in table 4. The RMSEs are computed using 500 replicates.

#### 4.3.4 Budget allocation

Lastly, fig. 4a shows the number  $n_\ell$  of evaluations for each of the  $L + 1$  correction levels of the MLMC-\* telescopic sum. Precisely,  $n_0$  is the number of evaluations of  $f_0$ , while  $n_\ell$  is the number of evaluations of  $f_\ell - f_{\ell-1}$ , for  $\ell > 0$ . We observe a typical sample allocation for MLMC-like estimators in ideal cases, that is, many coarse evaluations, and fewer and fewer fine evaluations. The MLMC-CV[0] estimator slightly deviates from this pattern, with  $n_1 \approx n_0$ , which can be explained by the fact that  $\mathcal{V}_1 \approx 590$  is of the same order of magnitude as  $(1 - R_0^2)\mathcal{V}_0 \approx 176$ . Figure 4b depicts the share of overall sampling cost associated with the different correction levels. Specifically,  $n_0\mathcal{C}_0\mathcal{C}^{-1}$  is the share of sampling budget dedicated to evaluating  $f_0$ , and  $n_\ell(\mathcal{C}_\ell + \mathcal{C}_{\ell-1})\mathcal{C}^{-1}$  is the share dedicated to evaluating  $f_\ell - f_{\ell-1}$ , for  $\ell > 0$ . We see that, except for MLMC-CV[0] for the reasons explained above, most of the sampling budget (around 70%) is allocated to the coarsest level, and most of the remaining budget is dedicated to correction level  $\ell = 1$ . We note that the CV-based MLMC estimators dedicate slightly more budget to levels 2 and 3 than for the standard MLMC estimator. The sampling budget shares of fig. 4b are consistent with the theoretical optimal shared reported in table 8, including those of the MLMC-CV[0] estimator. This suggests that the sample allocation resulting from algorithm 1 seems to converge to the theoretical optimal sample allocation eq. (60).

## 5 Conclusions

In this paper, we proposed multilevel variance reduction strategies relying on surrogate-based control variates. On the one hand, using specific surrogate models, such as polynomial



(a) Sample allocation. The bars represent the median sample size  $n_\ell$  associated with the corresponding correction level  $\ell$  of the MLMC-\* telescopic sum. The black vertical lines represent the 25% and 75% quantiles. (b) Budget allocation. The stacked bars represent the proportion of the sampling cost dedicated to each level of the MLMC-\* telescopic sum.

Figure 4: Comparison of the sample allocation and associate computational cost across the correction levels for the MLMC-based estimators of  $\theta = \mathbb{E}[f_3(\mathbf{X})]$  (see eq. (75)), with an estimation budget of  $\mathcal{C} = 10\,000$ .

chaos expansions or Taylor polynomial expansions, allows to directly access exact statistics (mean, variance) of the control variates. Even if these exactly statistics are not directly accessible (e.g., when using GPs), they can be estimated very accurately at negligible cost. This contrasts with typical control variates relying on lower-fidelity models based on models/simulators with degraded physics or coarser discretizations, for which approximate control variate strategies need to be devised, resulting in a lower variance reduction. On the other hand, when multiple levels of fidelities (e.g., based on the discretization) with a clear cost/accuracy hierarchy are available, the surrogate-based control variate approach can be efficiently combined with multilevel strategies. The first strategy, MLCV, simply consists in using multiple control variates based on surrogate models of the simulators corresponding to the different levels. The main advantage is that the surrogate models corresponding to coarse levels may be constructed using larger sample sizes than for the finest level, resulting in a more accurate surrogate model (i.e., with lower model error). This strategy thus leads to a greater variance reduction compared to only using one surrogate model based on the finest level. This is supported by the numerical experiments we conducted, as well as by the theoretical variance reduction provided by proposition 1. The second strategy, MLMC-MLCV, allows to further improve the variance reduction by combining the surrogate-based control variates with an MLMC strategy. The most appropriate way to construct and utilize the surrogate models is, however, not straightforward, and was discussed in detail in sections 3.2.2 and 3.2.3. The additional variance reduction as compared to plain MLMC is demonstrated in our numerical experiments and supported by the theoretical variance reduction factor eq. (51) derived in section 3.2.3.

The construction cost of the surrogate models was discussed from two perspectives. When

the surrogate models are constructed for the sole purpose of serving for the control variate estimation, then the cost of their construction must be taken into account for fair comparison with other approaches. In such a case, it may not be optimal, in terms of cost/accuracy tradeoff, to construct surrogate models on all levels, especially when only a limited budget is available. In particular, using a subset of surrogate models based on the coarser levels may already lead to considerable variance reduction, provided that the outputs of the coarse surrogate models are sufficiently correlated with that of the high-fidelity simulator. Then, considering additional surrogate models on finer levels might only result in marginal improvement, at the expense of a significant computational cost. On the contrary, if the surrogate models have already been constructed for other purposes, and are, in some sense, available “for free,” then their construction cost need not be considered, and the entire set of surrogate models may then be used.

From the former perspective, that is when the cost of the surrogate construction is considered as part of the estimation cost, one may devise more involved strategies seeking to optimize the tradeoff between the construction cost and the model error, directly impacting the projected variance reduction. For instance, in the context of polynomial chaos surrogate models, the truncation strategy may be controlled to this end, as proposed in a stochastic Galerkin framework in [56], where the total polynomial degree is optimized alongside the sample size and the CV parameter to minimize the PC-based CV estimator’s variance under a cost constraint. Another avenue to improve the proposed approach would be to replace the MLMC part of the MLMC-MLCV strategy by a more efficient multilevel approach, such as the multilevel best linear unbiased estimator (MLBLUE) [48, 49, 47]. In particular, an MLBLUE-MLCV strategy should be more efficient when a collection of low-fidelity simulators (e.g., with degraded physics) with no clear cost/accuracy hierarchy is available. Finally, although the proposed approaches apply to the estimation of arbitrary statistics, they were only tested here on the estimation of expected values. The theoretical and algorithmic ingredients for the estimation of variances are, however, described in this paper and may be tested in follow-up investigations and numerical experiments. Specifically, the multifidelity estimation of variance-based sensitivity indices is of particular interest to our team.

## 6 Acknowledgment

We wish to acknowledge the PIA framework (CGI, ANR) and the industrial members of the IRT Saint Exupéry project R-Evol: Airbus, Liebherr, Altran Technologies, Capgemini DEMS France, CENAERO and Cerfacs for their support, financial funding and own knowledge.

# A Optimal parameter for expectation and variance CV estimators

We derive here the expression of the optimal CV parameter  $\boldsymbol{\alpha}^*$  for the CV estimators of the expectation and of the variance. We define  $Y = f(\mathbf{X})$ ,  $Z_m = g_m(\mathbf{X})$  and  $\mathbf{Z} = (Z_m)_{m=1}^M$ . Then, given an input  $n$ -sample  $\{\mathbf{X}^{(i)}\}_{i=1}^n$ , we define  $Y^{(i)} = f(\mathbf{X}^{(i)})$ ,  $Z_m^{(i)} = g_m(\mathbf{X}^{(i)})$  and  $\mathbf{Z}^{(i)} = (Z_m^{(i)})_{m=1}^M$ .

## A.1 CV estimator of the expectation

For the expectation, we have  $\boldsymbol{\Sigma} = \mathbb{C}[\hat{E}[\mathbf{Z}]]$  and  $\mathbf{c} = \mathbb{C}[\hat{E}[Y], \hat{E}[\mathbf{Z}]]$ , i.e.

$$[\boldsymbol{\Sigma}]_{m,m'} = \mathbb{C}[\hat{E}[Z_m], \hat{E}[Z_{m'}]] = n^{-2} \sum_{i,j=1}^n \mathbb{C}[Z_m^{(i)}, Z_{m'}^{(j)}] = n^{-1} \mathbb{C}[Z_m, Z_{m'}], \quad (78)$$

$$[\mathbf{c}]_m = \mathbb{C}[\hat{E}[Y], \hat{E}[Z_m]] = n^{-2} \sum_{i,j=1}^n \mathbb{C}[Y^{(i)}, Z_m^{(j)}] = n^{-1} \mathbb{C}[Y, Z_m], \quad (79)$$

so that  $\boldsymbol{\Sigma} = n^{-1} \mathbb{C}[\mathbf{Z}]$ ,  $\mathbf{c} = n^{-1} \mathbb{C}[Y, \mathbf{Z}]$ , and, eventually,

$$\boldsymbol{\alpha}^* = \mathbb{C}[\mathbf{Z}]^{-1} \mathbb{C}[Y, \mathbf{Z}]. \quad (80)$$

Furthermore, we have  $\mathbb{V}[\hat{E}[Y]] = n^{-1} \mathbb{V}[Y]$ , so that

$$R^2 = \frac{\mathbb{C}[Y, \mathbf{Z}]^\top \mathbb{C}[\mathbf{Z}]^{-1} \mathbb{C}[Y, \mathbf{Z}]}{\mathbb{V}[Y]} = \mathbf{r}_{Y,\mathbf{Z}}^\top \mathbf{R}_{\mathbf{Z}}^{-1} \mathbf{r}_{Y,\mathbf{Z}}, \quad (81)$$

where

$$\mathbf{r}_{Y,\mathbf{Z}} = (\mathbb{V}[Y] \mathbf{D}_{\mathbf{Z}})^{-1/2} \mathbb{C}[Y, \mathbf{Z}], \quad \mathbf{R}_{\mathbf{Z}} = \mathbf{D}_{\mathbf{Z}}^{-1/2} \mathbb{C}[\mathbf{Z}] \mathbf{D}_{\mathbf{Z}}^{-1/2}, \quad \mathbf{D}_{\mathbf{Z}} = \text{Diag}(\mathbb{C}[\mathbf{Z}]). \quad (82)$$

Thus,  $R^2 \in [0, 1]$  corresponds to the squared coefficient of multiple correlation between  $Y$  and the control variates  $Z_1, \dots, Z_M$ .

## A.2 CV estimator of the variance

Similarly, for the variance, we have  $\boldsymbol{\Sigma} = \mathbb{C}[\hat{V}[\mathbf{Z}]]$  and  $\mathbf{c} = \mathbb{C}[\hat{V}[Y], \hat{V}[\mathbf{Z}]]$ . We start by deriving useful identities. In what follows, for any random variable  $A$ , we denote the corresponding centered variable by  $\bar{A} := A - \mathbb{E}[A]$ . First, we remark that  $\hat{V}[A] = \hat{V}[\bar{A}]$ , so that, for any two random variables  $Y$  and  $Z$ ,

$$\mathbb{C}[\hat{V}[Y], \hat{V}[Z]] = \mathbb{E}[\hat{V}[\bar{Y}] \hat{V}[\bar{Z}]] - \mathbb{V}[Y] \mathbb{V}[Z]. \quad (83)$$

Furthermore, it can be shown that

$$\mathbb{E}[\hat{V}[\bar{Y}] \hat{V}[\bar{Z}]] = \left( \frac{n}{n-1} \right)^2 (a_n(\bar{Y}, \bar{Z}) + b_n(\bar{Y}, \bar{Z}) - c_n(\bar{Y}, \bar{Z}) - c_n(\bar{Z}, \bar{Y})), \quad (84)$$

with (see proof below)

$$a_n(\bar{Y}, \bar{Z}) := \mathbb{E}[\hat{E}[\bar{Y}^2]\hat{E}[\bar{Z}^2]] = \frac{1}{n}\mathbb{C}[\bar{Y}^2, \bar{Z}^2] + \mathbb{V}[Y]\mathbb{V}[Z] = a_n(\bar{Z}, \bar{Y}), \quad (85)$$

$$b_n(\bar{Y}, \bar{Z}) := \mathbb{E}[\hat{E}[\bar{Y}^2]^2\hat{E}[\bar{Z}^2]^2] = \frac{a_n(\bar{Y}, \bar{Z})}{n^2} + 2\frac{n-1}{n^3}\mathbb{C}[Y, Z]^2 = b_n(\bar{Z}, \bar{Y}), \quad (86)$$

$$c_n(\bar{Y}, \bar{Z}) := \mathbb{E}[\hat{E}[\bar{Y}^2]\hat{E}[\bar{Z}^2]^2] = \frac{a_n(\bar{Y}, \bar{Z})}{n} = c_n(\bar{Z}, \bar{Y}). \quad (87)$$

We thus have

$$\mathbb{C}[\hat{V}[Y], \hat{V}[Z]] = \frac{1}{n}\mathbb{C}[\bar{Y}^2, \bar{Z}^2] + \frac{2}{n(n-1)}\mathbb{C}[Y, Z]^2, \quad (88)$$

$$\mathbb{V}[\hat{V}[Y]] = \mathbb{C}[\hat{V}[Y], \hat{V}[Y]] = \frac{1}{n}\mathbb{V}[\bar{Y}^2] + \frac{2}{n(n-1)}\mathbb{V}[Y]^2, \quad (89)$$

eventually leading to

$$\boldsymbol{\Sigma} = \mathbb{C}[\hat{V}[\mathbf{Z}]] = \frac{1}{n} \left( \mathbb{C}[\bar{\mathbf{Z}}^{\odot 2}] + \frac{2}{n-1}\mathbb{C}[\mathbf{Z}]^{\odot 2} \right), \quad (90)$$

$$\mathbf{c} = \mathbb{C}[\hat{V}[Y], \hat{V}[\mathbf{Z}]] = \frac{1}{n} \left( \mathbb{C}[\bar{Y}^2, \bar{\mathbf{Z}}^{\odot 2}] + \frac{2}{n-1}\mathbb{C}[Y, \mathbf{Z}]^{\odot 2} \right), \quad (91)$$

so that

$$\boldsymbol{\alpha}^* = \left[ \mathbb{C}[\bar{\mathbf{Z}}^{\odot 2}] + \frac{2}{n-1}\mathbb{C}[\mathbf{Z}]^{\odot 2} \right]^{-1} \left[ \mathbb{C}[\bar{Y}^2, \bar{\mathbf{Z}}^{\odot 2}] + \frac{2}{n-1}\mathbb{C}[Y, \mathbf{Z}]^{\odot 2} \right], \quad (92)$$

$$R^2 = \frac{\left[ \mathbb{C}[\bar{Y}^2, \bar{\mathbf{Z}}^{\odot 2}] + \frac{2}{n-1}\mathbb{C}[Y, \mathbf{Z}]^{\odot 2} \right]^\top \boldsymbol{\alpha}^*}{\mathbb{V}[\bar{Y}^2] + \frac{2}{n-1}\mathbb{V}[Y]^2}. \quad (93)$$

As  $n \rightarrow \infty$ , we see that

$$\boldsymbol{\alpha}^* \rightarrow \mathbb{C}[\bar{\mathbf{Z}}^{\odot 2}]^{-1}\mathbb{C}[\bar{Y}^2, \bar{\mathbf{Z}}^{\odot 2}] \quad (94)$$

$$R^2 \rightarrow \mathbb{V}[\bar{Y}^2]^{-1}\mathbb{C}[\bar{Y}^2, \bar{\mathbf{Z}}^{\odot 2}]^\top \mathbb{C}[\bar{\mathbf{Z}}^{\odot 2}]^{-1}\mathbb{C}[\bar{Y}^2, \bar{\mathbf{Z}}^{\odot 2}] = \mathbf{r}_{Y, \mathbf{Z}}^\top \mathbf{R}_{\mathbf{Z}}^{-1} \mathbf{r}_{Y, \mathbf{Z}} =: R_{\text{lim}}^2, \quad (95)$$

where

$$\mathbf{r}_{Y, \mathbf{Z}} = (\mathbb{V}[\bar{Y}^2]\mathbf{D}_{\mathbf{Z}})^{-1/2}\mathbb{C}[\bar{Y}^2, \bar{\mathbf{Z}}^{\odot 2}], \quad \mathbf{R}_{\mathbf{Z}} = \mathbf{D}_{\mathbf{Z}}^{-1/2}\mathbb{C}[\bar{\mathbf{Z}}^{\odot 2}]\mathbf{D}_{\mathbf{Z}}^{-1/2}, \quad (96)$$

with  $\mathbf{D}_{\mathbf{Z}} = \text{Diag}(\mathbb{C}[\bar{\mathbf{Z}}^{\odot 2}])$ . Thus,  $R_{\text{lim}}^2 \in [0, 1]$  corresponds to the squared coefficient of multiple correlation between  $\bar{Y}^2$  and  $\bar{Z}_1^2, \dots, \bar{Z}_M^2$ .

We now proceed to the proof of identities eqs. (85) to (87). First, for eq. (85), by definition

$$a_n(\bar{Y}, \bar{Z}) = \mathbb{E}[\sum_{i=1}^n (\bar{Y}^{(i)})^2 \sum_{i=1}^n (\bar{Z}^{(i)})^2] = \frac{1}{n^2} \sum_{i,j=1}^n \mathbb{E}[(\bar{Y}^{(i)})^2 (\bar{Z}^{(j)})^2]. \quad (97)$$

We distinguish two (disjoint) cases:

1.  $i = j$ :  $\mathbb{E}[(\bar{Y}^{(i)})^2(\bar{Z}^{(j)})^2] = \mathbb{E}[(\bar{Y}^{(i)})^2(\bar{Z}^{(i)})^2] = \mathbb{E}[\bar{Y}^2\bar{Z}^2] = \mathbb{C}[\bar{Y}^2, \bar{Z}^2] + \mathbb{V}[Y]\mathbb{V}[Z]$ . There are  $n$  such terms in the sum.
2.  $i \neq j$ :  $\mathbb{E}[(\bar{Y}^{(i)})^2(\bar{Z}^{(j)})^2] = \mathbb{E}[(\bar{Y}^{(i)})^2]\mathbb{E}[(\bar{Z}^{(j)})^2] = \mathbb{E}[\bar{Y}^2]\mathbb{E}[\bar{Z}^2] = \mathbb{V}[Y]\mathbb{V}[Z]$ . There are  $n(n-1)$  such terms in the sum.

Then eq. (85) follows. For eq. (86),

$$b_n(\bar{Y}, \bar{Z}) = \mathbb{E}[(\sum_{i=1}^n \bar{Y}^{(i)})^2(\sum_{i=1}^n \bar{Z}^{(i)})^2] = \frac{1}{n^4} \sum_{i,j,k,\ell=1}^n \mathbb{E}[\bar{Y}^{(i)}\bar{Y}^{(j)}\bar{Z}^{(k)}\bar{Z}^{(\ell)}]. \quad (98)$$

We distinguish five (disjoint) cases:

1.  $i = j = k = \ell$ :  $\mathbb{E}[\bar{Y}^{(i)}\bar{Y}^{(j)}\bar{Z}^{(k)}\bar{Z}^{(\ell)}] = \mathbb{E}[(\bar{Y}^{(i)})^2(\bar{Z}^{(i)})^2] = \mathbb{E}[\bar{Y}^2\bar{Z}^2] = \mathbb{C}[\bar{Y}^2, \bar{Z}^2] + \mathbb{V}[Y]\mathbb{V}[Z]$ . There are  $n$  such terms in the sum.
2.  $i = j \neq k = \ell$ :  $\mathbb{E}[\bar{Y}^{(i)}\bar{Y}^{(j)}\bar{Z}^{(k)}\bar{Z}^{(\ell)}] = \mathbb{E}[(\bar{Y}^{(i)})^2]\mathbb{E}[(\bar{Z}^{(k)})^2] = \mathbb{E}[\bar{Y}^2]\mathbb{E}[\bar{Z}^2] = \mathbb{V}[Y]\mathbb{V}[Z]$ . There are  $n(n-1)$  such terms in the sum.
3.  $i = k \neq j = \ell$ :  $\mathbb{E}[\bar{Y}^{(i)}\bar{Y}^{(j)}\bar{Z}^{(k)}\bar{Z}^{(\ell)}] = \mathbb{E}[\bar{Y}^{(i)}\bar{Z}^{(i)}]\mathbb{E}[\bar{Y}^{(j)}\bar{Z}^{(j)}] = \mathbb{E}[\bar{Y}\bar{Z}]^2 = \mathbb{C}[Y, Z]^2$ . There are  $n(n-1)$  such terms in the sum.
4.  $i = \ell \neq j = k$ :  $\mathbb{E}[\bar{Y}^{(i)}\bar{Y}^{(j)}\bar{Z}^{(k)}\bar{Z}^{(\ell)}] = \mathbb{E}[\bar{Y}^{(i)}\bar{Z}^{(i)}]\mathbb{E}[\bar{Y}^{(j)}\bar{Z}^{(j)}] = \mathbb{E}[\bar{Y}\bar{Z}]^2 = \mathbb{C}[Y, Z]^2$ . There are  $n(n-1)$  such terms in the sum.
5. All remaining cases (at least one of the indices  $i, j, k, \ell$  is different from all the others):  $\mathbb{E}[\bar{Y}^{(i)}\bar{Y}^{(j)}\bar{Z}^{(k)}\bar{Z}^{(\ell)}] = 0$ .

Then eq. (86) follows. Finally, for eq. (87),

$$c_n(\bar{Y}, \bar{Z})\mathbb{E}[(\sum_{i=1}^n (\bar{Y}^{(i)})^2)(\sum_{i=1}^n \bar{Z}^{(i)})^2] = \frac{1}{n^3} \sum_{i,j,k=1}^n \mathbb{E}[(\bar{Y}^{(i)})^2\bar{Z}^{(j)}\bar{Z}^{(k)}]. \quad (99)$$

We distinguish three (disjoint) cases:

1.  $i = j = k$ :  $\mathbb{E}[(\bar{Y}^{(i)})^2\bar{Z}^{(j)}\bar{Z}^{(k)}] = \mathbb{E}[(\bar{Y}^{(i)})^2(\bar{Z}^{(i)})^2] = \mathbb{E}[\bar{Y}^2\bar{Z}^2] = \mathbb{M}^4[Y, Z]$ . There are  $n$  such terms in the sum.
2.  $i \neq j = k$ :  $\mathbb{E}[(\bar{Y}^{(i)})^2\bar{Z}^{(j)}\bar{Z}^{(k)}] = \mathbb{E}[(\bar{Y}^{(i)})^2]\mathbb{E}[(\bar{Z}^{(j)})^2] = \mathbb{E}[\bar{Y}^2]\mathbb{E}[\bar{Z}^2] = \mathbb{V}[Y]\mathbb{V}[Z]$ . There are  $n(n-1)$  such terms in the sum.
3. All remaining cases (at least one of the indices  $j, k$  is different from all the others):  $\mathbb{E}[(\bar{Y}^{(i)})^2\bar{Z}^{(j)}\bar{Z}^{(k)}] = 0$ .

Then eq. (87) follows.

**Remark 4.** When the expected value  $\boldsymbol{\mu}_{\mathbf{Z}}$  of  $\mathbf{Z}$  is known, as is the case when defining  $\mathbf{Z}$  from the prediction of certain surrogate models, such as PC expansion and Taylor polynomials (and, in some instances, GPs, see section 2.4.1), it is possible to replace  $\hat{\mathbf{V}}[\mathbf{Z}]$  with  $\hat{E}[\bar{\mathbf{Z}}^{\odot 2}]$ . The



derivation of the optimal CV parameter is somewhat easier and leads to similar results as in the unknown expectation case. Specifically,

$$[\boldsymbol{\Sigma}]_{m,m'} = \mathbb{C}[\hat{E}[\bar{Z}_m^2], \hat{E}[\bar{Z}_{m'}^2]] = \mathbb{C}[n^{-1} \sum_{i=1}^n (\bar{Z}_m^{(i)})^2, n^{-1} \sum_{i=1}^n (\bar{Z}_{m'}^{(i)})^2] \quad (100)$$

$$= n^{-2} \sum_{i,j=1}^n \mathbb{C}[(\bar{Z}_m^{(i)})^2, (\bar{Z}_{m'}^{(j)})^2] = n^{-2} \sum_{i=1}^n \mathbb{C}[(\bar{Z}_m^{(i)})^2, (\bar{Z}_{m'}^{(i)})^2] \quad (101)$$

$$= n^{-1} \mathbb{C}[\bar{Z}_m^2, \bar{Z}_{m'}^2], \quad (102)$$

i.e.  $\boldsymbol{\Sigma} = n^{-1} \mathbb{C}[\bar{\mathbf{Z}}^{\odot 2}]$ . Regarding the vector of covariances  $\mathbf{c}$ ,

$$[\mathbf{c}]_m = \mathbb{C}[\hat{V}[\bar{Y}], \hat{E}[\bar{Z}_m^2]] = \mathbb{E}[\hat{V}[\bar{Y}] \hat{E}[\bar{Z}_m^2]] - \mathbb{V}[Y] \mathbb{V}[Z_m] \quad (103)$$

$$= \frac{n}{n-1} (\mathbb{E}[\hat{E}[\bar{Y}^2] \hat{E}[\bar{Z}_m^2]] - \mathbb{E}[\hat{E}[\bar{Y}^2] \hat{E}[\bar{Z}_m^2]]) - \mathbb{V}[Y] \mathbb{V}[Z_m] \quad (104)$$

$$= \frac{n}{n-1} (a_n(\bar{Y}, \bar{Z}_m) - c_n(\bar{Z}_m, \bar{Y})) - \mathbb{V}[Y] \mathbb{V}[Z_m] = a_n(\bar{Y}, \bar{Z}_m) - \mathbb{V}[Y] \mathbb{V}[Z_m] \quad (105)$$

$$= n^{-1} \mathbb{C}[\bar{Y}^2, \bar{Z}_m^2], \quad (106)$$

i.e.  $\mathbf{c} = n^{-1} \mathbb{C}[\bar{\mathbf{Y}}^2, \bar{\mathbf{Z}}^{\odot 2}]$ , so that

$$\boldsymbol{\alpha}^* = \mathbb{C}[\bar{\mathbf{Z}}^{\odot 2}]^{-1} \mathbb{C}[\bar{\mathbf{Y}}^2, \bar{\mathbf{Z}}^{\odot 2}], \quad (107)$$

$$R^2 = \frac{\mathbb{C}[\bar{\mathbf{Y}}^2, \bar{\mathbf{Z}}^{\odot 2}]^\top \mathbb{C}[\bar{\mathbf{Z}}^{\odot 2}]^{-1} \mathbb{C}[\bar{\mathbf{Y}}^2, \bar{\mathbf{Z}}^{\odot 2}]}{\mathbb{V}[\bar{\mathbf{Y}}^2] + \frac{2}{n-1} \mathbb{V}[Y]^2} \xrightarrow{n \rightarrow \infty} R_{\text{lim}}^2, \quad (108)$$

with the same definitions as in eq. (95).

## B Optimal MLMC-MLCV variance estimator

The MLMC-MLCV variance estimator is given by eq. (45), with

$$\hat{T}_0^{\text{MLCV}}(\boldsymbol{\alpha}_0) = \hat{V}^{(0)}[Y_0] - \boldsymbol{\alpha}_0^\top (\hat{V}^{(0)}[\mathbf{Z}] - \boldsymbol{\sigma}_{\mathbf{Z}}^2), \quad (109)$$

$$\hat{T}_\ell^{\text{MLCV}}(\boldsymbol{\alpha}_\ell) = \hat{V}^{(\ell)}[Y_\ell] - \hat{V}^{(\ell)}[Y_{\ell-1}] - \boldsymbol{\alpha}_\ell^\top [\hat{V}^{(\ell)}[\mathbf{Z}_{1:}] - \hat{V}^{(\ell)}[\tilde{\mathbf{Z}}] - (\boldsymbol{\sigma}_{\mathbf{Z}_{1:}}^2 - \boldsymbol{\sigma}_{\tilde{\mathbf{Z}}}^2)], \text{ for } \ell > 1, \quad (110)$$

where

$$\mathbf{Z} = (Z_0, \dots, Z_L) = (g_0(\mathbf{X}), \dots, g_L(\mathbf{X})), \quad \boldsymbol{\sigma}_{\mathbf{Z}}^2 = \mathbb{V}[\mathbf{Z}], \quad (111)$$

$$\mathbf{Z}_{1:} = (Z_1, \dots, Z_L) = (g_1(\mathbf{X}), \dots, g_L(\mathbf{X})), \quad \boldsymbol{\sigma}_{\mathbf{Z}_{1:}}^2 = \mathbb{V}[\mathbf{Z}_{1:}], \quad (112)$$

$$\tilde{\mathbf{Z}} = (\tilde{Z}_0, \dots, \tilde{Z}_{L-1}) = (\tilde{g}_0(\mathbf{X}), \dots, \tilde{g}_{L-1}(\mathbf{X})), \quad \boldsymbol{\sigma}_{\tilde{\mathbf{Z}}}^2 = \mathbb{V}[\tilde{\mathbf{Z}}], \quad (113)$$

$$\tilde{g}_{\ell-1} = g_\ell - h_\ell, \quad \text{for } \ell > 0. \quad (114)$$

Furthermore, we assume that the expected values of the control variates are known, so that we use the following variance estimators:

$$\hat{V}^{(0)}[\mathbf{Z}] = \hat{E}^{(0)}[\bar{\mathbf{Z}}^2], \quad \hat{V}^{(\ell)}[\mathbf{Z}_{1:}] - \hat{V}^{(\ell)}[\tilde{\mathbf{Z}}] = \hat{E}^{(\ell)}[\bar{\mathbf{Z}}_{1:}^{\odot 2}] - \hat{E}^{(\ell)}[\tilde{\mathbf{Z}}^{\odot 2}] = \hat{E}^{(\ell)}[\bar{\mathbf{Z}}_{1:}^{\odot 2} - \tilde{\mathbf{Z}}^{\odot 2}]. \quad (115)$$

The optimal values  $\boldsymbol{\alpha}_\ell^*$  of the CV parameters are given by

$$\boldsymbol{\alpha}_0^* = \boldsymbol{\Sigma}_0^{-1} \mathbf{c}_0, \quad \boldsymbol{\Sigma}_0 = \mathbb{C}[\bar{\mathbf{Z}}^{\odot 2}], \quad \mathbf{c}_0 = \mathbb{C}[\bar{\mathbf{Z}}^{\odot 2}, \bar{Y}_0^2], \quad (116)$$

$$\boldsymbol{\alpha}_\ell^* = \boldsymbol{\Sigma}_\ell^{-1} \mathbf{c}_\ell, \quad \boldsymbol{\Sigma}_\ell = \mathbb{C}[\bar{\mathbf{Z}}_{1:}^{\odot 2} - \tilde{\mathbf{Z}}^{\odot 2}], \quad \mathbf{c}_\ell = \mathbb{C}[\bar{\mathbf{Z}}_{1:}^{\odot 2} - \tilde{\mathbf{Z}}^{\odot 2}, \bar{Y}_\ell^2 - \bar{Y}_{\ell-1}^2], \quad \text{for } \ell > 0. \quad (117)$$

Note that  $\boldsymbol{\Sigma}_\ell$  is the same for all  $\ell > 0$ . For PC-based control variates,  $g_\ell(\mathbf{X}) = \sum_{k=0}^{P_g^\ell} \mathfrak{g}_{\ell,k} \Psi_k(\mathbf{X})$  and  $h_\ell(\mathbf{X}) = \sum_{k=0}^{P_h^\ell} \mathfrak{h}_{\ell,k} \Psi_k(\mathbf{X})$ , letting  $P_{\tilde{g}}^\ell := \max(P_g^\ell, P_h^\ell)$ , we have

$$\tilde{Z}_{\ell-1} = \sum_{k=0}^{P_{\tilde{g}}^\ell} \tilde{\mathfrak{g}}_{\ell-1,k} \Psi_k(\mathbf{X}), \quad \text{with } \tilde{\mathfrak{g}}_{\ell-1,k} := \mathfrak{g}_{\ell,k} - \mathfrak{h}_{\ell,k}, \quad \text{for } \ell = 1, \dots, L, \quad (118)$$

where  $\mathfrak{g}_{\ell,k} := 0$  for  $k > P_g^\ell$  and  $\mathfrak{h}_{\ell,k} := 0$  for  $k > P_h^\ell$ . Consequently,

$$\mathbb{E}[Z_\ell] = \mathfrak{g}_{\ell,0}, \quad \mathbb{E}[\tilde{Z}_{\ell-1}] = \tilde{\mathfrak{g}}_{\ell-1,0}, \quad (119)$$

$$\mathbb{V}[Z_\ell] = \sum_{k=1}^{P_g^\ell} \mathfrak{g}_{\ell,k}^2, \quad \mathbb{V}[\tilde{Z}_{\ell-1}] = \sum_{k=1}^{P_{\tilde{g}}^\ell} \tilde{\mathfrak{g}}_{\ell-1,k}^2. \quad (120)$$

Furthermore,

$$[\boldsymbol{\Sigma}_0]_{m,m'} = \sum_{i,j=1}^{P_g^m} \sum_{q,r=1}^{P_g^{m'}} \mathfrak{g}_{m,i} \mathfrak{g}_{m,j} \mathfrak{g}_{m',q} \mathfrak{g}_{m',r} (\Phi_{ijqr} - \delta_{ij} \delta_{qr}), \quad \text{for } m, m' = 0, \dots, L, \quad (121)$$

where  $\Phi_{ijqr} := \mathbb{E}[\Psi_i(\mathbf{X}) \Psi_j(\mathbf{X}) \Psi_q(\mathbf{X}) \Psi_r(\mathbf{X})]$  are the entries of the fourth-order Galerkin product tensor, which is a well-known object in stochastic Galerkin methods [20, 32], and  $\delta_{ij}$  denotes the Kronecker delta. Besides, noticing that  $\bar{\mathbf{Z}}_{1:}^{\odot 2} - \tilde{\mathbf{Z}}^{\odot 2} = (\bar{\mathbf{Z}}_{1:} - \tilde{\mathbf{Z}}) \odot (\bar{\mathbf{Z}}_{1:} + \tilde{\mathbf{Z}}) = \bar{\mathbf{W}} \odot (\bar{\mathbf{Z}}_{1:} + \tilde{\mathbf{Z}})$ , with  $\bar{\mathbf{W}} = \mathbf{W} - \boldsymbol{\mu}_{\mathbf{W}}$  and the definitions in eq. (55), we have

$$[\boldsymbol{\Sigma}_\ell]_{m,m'} = A_{m,m'} + B_{m,m'} + C_{m,m'} + C_{m',m}, \quad \text{for } \ell, m, m' = 1, \dots, L, \quad (122)$$

where

$$A_{m,m'} = \mathbb{C}[\bar{W}_m \bar{Z}_m, \bar{W}_{m'} \bar{Z}_{m'}] \quad (123)$$

$$= \sum_{i=1}^{P_h^m} \sum_{j=1}^{P_g^m} \sum_{q=1}^{P_h^{m'}} \sum_{r=1}^{P_g^{m'}} \mathfrak{h}_{m,i} \mathfrak{g}_{m,j} \mathfrak{h}_{m',q} \mathfrak{g}_{m',r} (\Phi_{ijqr} - \delta_{ij} \delta_{qr}), \quad (124)$$

$$B_{m,m'} = \mathbb{C}[\bar{W}_m \bar{Z}_{m-1}, \bar{W}_{m'} \bar{Z}_{m'-1}] \quad (125)$$

$$= \sum_{i=1}^{P_h^m} \sum_{j=1}^{P_g^m} \sum_{q=1}^{P_h^{m'}} \sum_{r=1}^{P_g^{m'}} \mathfrak{h}_{m,i} \tilde{\mathfrak{g}}_{m-1,j} \mathfrak{h}_{m',q} \tilde{\mathfrak{g}}_{m'-1,r} (\Phi_{ijqr} - \delta_{ij} \delta_{qr}), \quad (126)$$

$$C_{m,m'} = \mathbb{C}[\bar{W}_m \bar{Z}_{m-1}, \bar{W}_{m'} \bar{Z}_{m'}] \quad (127)$$

$$= \sum_{i=1}^{P_h^m} \sum_{j=1}^{P_g^m} \sum_{q=1}^{P_h^{m'}} \sum_{r=1}^{P_g^{m'}} \mathfrak{h}_{m,i} \tilde{\mathfrak{g}}_{m-1,j} \mathfrak{h}_{m',q} \mathfrak{g}_{m',r} (\Phi_{ijqr} - \delta_{ij} \delta_{qr}). \quad (128)$$

## C Taylor surrogate for the numerical test case

In our example,  $f_\ell$  is only differentiable in  $[-\pi, \pi]^3 \times [\nu_{\min}, \nu_{\max}] \times ([-1, 0) \cup (0, 1])^3$ . Consequently, the Taylor polynomial surrogate cannot be used directly as defined in (17) and (18), because the Jacobian and Hessian matrices are not defined at  $\boldsymbol{\mu}_{\mathbf{X}}$ . Instead, we define the first-order Taylor surrogate as

$$f_\ell(\mathbf{X}) \simeq g_\ell^{\text{T}_1}(\mathbf{X}) = f_\ell(\boldsymbol{\mu}_{\mathbf{X}}) + \sum_{i=1}^7 g_{\ell,i}^{\text{T}_1}(\boldsymbol{\mu}_{\mathbf{X}}; \mathbf{X}), \quad (129)$$

where, for  $i = 1, \dots, 4$ ,  $g_{\ell,i}^{\text{T}_1}(\boldsymbol{\mu}_{\mathbf{X}}; \mathbf{X}) = (X_i - \mu_{X_i}) \frac{\partial f_\ell}{\partial X_i}(\boldsymbol{\mu}_{\mathbf{X}})$ , and, for  $i = 5, \dots, 7$ ,

$$g_{\ell,i}^{\text{T}_1}(\boldsymbol{\mu}_{\mathbf{X}}; \mathbf{X}) = (X_i - \mu_{X_i}) \times \begin{cases} \frac{\partial f_\ell}{\partial X_i}(\boldsymbol{\mu}_{\mathbf{X}}), & \mu_{X_i} \neq 0, \\ \lim_{\boldsymbol{\mu}'_{\mathbf{X}} \rightarrow \boldsymbol{\mu}_{\mathbf{X}}^{i,0^-}} \frac{\partial f_\ell}{\partial X_i}(\boldsymbol{\mu}'_{\mathbf{X}}), & \mu_{X_i} = 0, X_i < 0, \\ \lim_{\boldsymbol{\mu}'_{\mathbf{X}} \rightarrow \boldsymbol{\mu}_{\mathbf{X}}^{i,0^+}} \frac{\partial f_\ell}{\partial X_i}(\boldsymbol{\mu}'_{\mathbf{X}}), & \mu_{X_i} = 0, X_i > 0, \\ 0 & \mu_{X_i} = X_i = 0, \end{cases} \quad (130)$$

where  $\boldsymbol{\mu}_{\mathbf{X}}^{i,0^\pm} = (\mu_{X_1}, \dots, \mu_{X_{i-1}}, 0^\pm, \mu_{X_{i+1}}, \dots, \mu_{X_7})$ , which is now well-defined.

With our choice of distributions for  $\mathbf{X}$  given in (72), we have  $\boldsymbol{\mu}_{\mathbf{X}} = (0, 0, 0, 0.005, 0, 0, 0)$ ,

and the first-order Taylor surrogate is defined by (129), with

$$g_{\ell,1}^{\text{T}_1}(\boldsymbol{\mu}_{\mathbf{X}}; \mathbf{X}) = 7X_1 \sum_{k=1}^K B_k^\ell(T; \boldsymbol{\mu}_{\mathbf{X}}) \sum_{j=1}^{N_\ell} w_j \sin(k\pi x_j) \mathcal{F}_2(x_j), \quad (131)$$

$$g_{\ell,2}^{\text{T}_1}(\boldsymbol{\mu}_{\mathbf{X}}; \mathbf{X}) = 0, \quad (132)$$

$$g_{\ell,3}^{\text{T}_1}(\boldsymbol{\mu}_{\mathbf{X}}; \mathbf{X}) = 0, \quad (133)$$

$$g_{\ell,4}^{\text{T}_1}(\boldsymbol{\mu}_{\mathbf{X}}; \mathbf{X}) = -(X_4 - 0.005)\pi^2 T \sum_{k=1}^K k^2 \exp(-0.005k^2\pi^2 T) A_k^\ell(\boldsymbol{\mu}_{\mathbf{X}}) \sum_{j=1}^{N_\ell} w_j \sin(k\pi x_j), \quad (134)$$

$$g_{\ell,5}^{\text{T}_1}(\boldsymbol{\mu}_{\mathbf{X}}; \mathbf{X}) = 400|X_5| \sum_{k=1}^K B_k^\ell(T; \boldsymbol{\mu}_{\mathbf{X}}) \sum_{j=1}^{N_\ell} w_j \sin(k\pi x_j) \mathcal{F}_1(x_j), \quad (135)$$

$$g_{\ell,6}^{\text{T}_1}(\boldsymbol{\mu}_{\mathbf{X}}; \mathbf{X}) = 400|X_6| \sum_{k=1}^K B_k^\ell(T; \boldsymbol{\mu}_{\mathbf{X}}) \sum_{j=1}^{N_\ell} w_j \sin(k\pi x_j) \mathcal{F}_1(x_j), \quad (136)$$

$$g_{\ell,7}^{\text{T}_1}(\boldsymbol{\mu}_{\mathbf{X}}; \mathbf{X}) = 400|X_7| \sum_{k=1}^K B_k^\ell(T; \boldsymbol{\mu}_{\mathbf{X}}) \sum_{j=1}^{N_\ell} w_j \sin(k\pi x_j) \mathcal{F}_1(x_j). \quad (137)$$

Because of the piecewise definition of  $g_\ell^{\text{T}_1}$ , the identity  $\mathbb{E}[g_\ell^{\text{T}_1}(\mathbf{X})] = f_\ell(\boldsymbol{\mu}_{\mathbf{X}})$  for a regular first-order Taylor surrogate no longer holds. Instead, we have

$$\mathbb{E}[g_\ell^{\text{T}_1}(\mathbf{X})] = f_\ell(\boldsymbol{\mu}_{\mathbf{X}}) + 600 \sum_{k=1}^K B_k^\ell(T; \boldsymbol{\mu}_{\mathbf{X}}) \sum_{j=1}^{N_\ell} w_j \sin(k\pi x_j) \mathcal{F}_1(x_j). \quad (138)$$

## References

- [1] T. W. ANDERSON, *An Introduction to Multivariate Statistical Analysis*, Wiley Series in Probability and Statistics, Wiley-Interscience, Hoboken, N.J, 3rd ed ed., 2003.
- [2] 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 (2007), pp. 1005–1034.
- [3] M. BERVEILLER, B. SUDRET, AND M. LEMAIRE, *Stochastic finite element: A non intrusive approach by regression*, European Journal of Computational Mechanics, 15 (2006), pp. 81–92.
- [4] C. BIERIG AND A. CHERNOV, *Convergence analysis of multilevel Monte Carlo variance estimators and application for random obstacle problems*, Numerische Mathematik, 130 (2015), pp. 579–613.
- [5] —, *Estimation of arbitrary order central statistical moments by the multilevel Monte Carlo method*, Stochastics and Partial Differential Equations Analysis and Computations, 4 (2016), pp. 3–40.

- [6] G. BLATMAN AND B. SUDRET, *Efficient computation of global sensitivity indices using sparse polynomial chaos expansions*, Reliability Engineering & System Safety, 95 (2010), pp. 1216–1229.
- [7] ———, *Adaptive sparse polynomial chaos expansion based on least angle regression*, Journal of Computational Physics, 230 (2011), pp. 2345–2367.
- [8] F.-X. BRIOL, C. J. OATES, M. GIROLAMI, M. A. OSBORNE, AND D. SEJDINOVIC, *Probabilistic Integration: A Role in Statistical Computation?*, Statistical Science, 34 (2019), pp. 1–22.
- [9] R. H. CAMERON AND W. T. MARTIN, *The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals*, Annals of Mathematics, 48 (1947), pp. 385–392.
- [10] K. A. CLIFFE, M. B. 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 (2011), pp. 3–15.
- [11] P. R. CONRAD AND Y. M. MARZOUK, *Adaptive Smolyak Pseudospectral Approximations*, SIAM Journal on Scientific Computing, 35 (2013), pp. A2643–A2670.
- [12] P. G. CONSTANTINE, M. S. ELDRED, AND E. T. PHIPPS, *Sparse pseudospectral approximation method*, Computer Methods in Applied Mechanics and Engineering, 229–232 (2012), pp. 1–12.
- [13] J. DICK, F. Y. KUO, AND I. H. SLOAN, *High-dimensional integration: The quasi-monte carlo way*, Acta Numerica, 22 (2013), p. 133–288.
- [14] B. EFRON, T. HASTIE, I. JOHNSTONE, AND R. TIBSHIRANI, *Least Angle Regression*, The Annals of Statistics, 32 (2004), pp. 407–451.
- [15] R. EL AMRI, R. LE RICHE, C. HELBERT, C. BLANCHET-SCALLIET, AND S. DA VEIGA, *A sampling criterion for constrained bayesian optimization with uncertainties*, arXiv preprint arXiv:2103.05706, (2021).
- [16] J. FOX AND G. ÖKTEN, *Polynomial chaos as a control variate method*, SIAM Journal on Scientific Computing, 43 (2021), pp. A2268–A2294.
- [17] S. GARG, N. SOOD, AND C. D. SARRIS, *Uncertainty quantification of ray-tracing based wireless propagation models with a control variate-polynomial chaos expansion method*, in 2015 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting, 2015, pp. 1776–1777.
- [18] G. GERACI, M. ELDRED, AND G. IACCARINO, *A multifidelity control variate approach for the multilevel Monte Carlo technique*, tech. rep., Stanford University, 2015.

- [19] G. GERACI, M. S. ELDRED, AND G. IACCARINO, *A multifidelity multilevel Monte Carlo method for uncertainty propagation in aerospace applications*, in 19th AIAA Non-Deterministic Approaches Conference, Grapevine, Texas, 2017, American Institute of Aeronautics and Astronautics.
- [20] R. G. GHANEM AND P. D. SPANOS, *Stochastic Finite Elements: A Spectral Approach*, Courier Corporation, 2003.
- [21] M. B. GILES, *Multilevel Monte Carlo Path Simulation*, Operations Research, 56 (2008), pp. 607–617.
- [22] —, *Multilevel Monte Carlo methods*, Acta Numerica, 24 (2015), pp. 259–328.
- [23] M. B. GILES AND B. J. WATERHOUSE, *Multilevel quasi-Monte Carlo path simulation*, Advanced Financial Modelling, Radon Series on Computational and Applied Mathematics, 8 (2009), pp. 165–181.
- [24] A. A. GORODETSKY, G. GERACI, M. S. ELDRED, AND J. D. JAKEMAN, *A generalized approximate control variate framework for multifidelity uncertainty quantification*, Journal of Computational Physics, 408 (2020), p. 109257.
- [25] Z. GU AND C. D. SARRIS, *Multi-parametric uncertainty quantification with a hybrid monte-carlo / polynomial chaos expansion ftdt method*, in 2015 IEEE MTT-S International Microwave Symposium, 2015, pp. 1–3.
- [26] A.-L. HAJI-ALI, F. NOBILE, AND R. TEMPONE, *Multi-index Monte Carlo: When sparsity meets sampling*, Numerische Mathematik, 132 (2016), pp. 767–806.
- [27] T. HESTERBERG, *Weighted average importance sampling and defensive mixture distributions*, Technometrics, 37 (1995), pp. 185–194.
- [28] J. JANUSEVSKIS AND R. LE RICHE, *Simultaneous kriging-based estimation and optimization of mean response*, Journal of Global Optimization, 55 (2013), pp. 313–336.
- [29] S. S. LAVENBERG, T. L. MOELLER, AND P. D. WELCH, *The application of control variables to the simulation of closed queueing networks*, in Proceedings of the 9th Conference on Winter Simulation - Volume 1, WSC '77, Gaithersburg, Maryland, USA, 1977, Winter Simulation Conference, pp. 152–154.
- [30] S. S. LAVENBERG, T. L. MOELLER, AND P. D. WELCH, *Statistical Results on Control Variables with Application to Queueing Network Simulation*, Operations Research, 30 (1982), pp. 182–202.
- [31] S. S. LAVENBERG AND P. D. WELCH, *A Perspective on the Use of Control Variables to Increase the Efficiency of Monte Carlo Simulations*, Management Science, 27 (1981), pp. 322–335.
- [32] O. P. LE MAÎTRE AND O. M. KNIO, *Spectral Methods for Uncertainty Quantification*, Scientific Computation, Springer Netherlands, Dordrecht, 2010.

- [33] C. LEMIEUX, *Control Variates*, in Wiley StatsRef: Statistics Reference Online, American Cancer Society, 2017, pp. 1–8.
- [34] N. LÜTHEN, S. MARELLI, AND B. SUDRET, *Sparse polynomial chaos expansions: Literature survey and benchmark*, SIAM/ASA Journal on Uncertainty Quantification, 9 (2021), pp. 593–649.
- [35] M. MCKAY, R. BECKMAN, AND W. CONOVER, *A comparison of three methods for selecting values of input variables in the analysis of output from a computer code*, Technometrics, 21 (1979), pp. 239–245.
- [36] P. MYCEK AND M. DE LOZZO, *Multilevel Monte Carlo Covariance Estimation for the Computation of Sobol’ Indices*, SIAM/ASA Journal on Uncertainty Quantification, 7 (2019), pp. 1323–1348.
- [37] H. N. NAJM, *Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics*, Annual Review of Fluid Mechanics, 41 (2009), pp. 35–52.
- [38] B. L. NELSON, *Control Variate Remedies*, Operations Research, 38 (1990), pp. 974–992.
- [39] L. W. T. NG AND K. E. WILLCOX, *Multifidelity approaches for optimization under uncertainty*, International Journal for Numerical Methods in Engineering, 100 (2014), pp. 746–772.
- [40] F. NOBILE, R. TEMPONE, AND C. G. WEBSTER, *An Anisotropic Sparse Grid Stochastic Collocation Method for Partial Differential Equations with Random Input Data*, SIAM Journal on Numerical Analysis, 46 (2008), pp. 2411–2442.
- [41] J. PARK AND I. W. SANDBERG, *Universal Approximation Using Radial-Basis-Function Networks*, Neural Computation, 3 (1991), pp. 246–257.
- [42] R. PASUPATHY, B. W. SCHMEISER, M. R. TAAFFE, AND J. WANG, *Control-variate estimation using estimated control means*, IIE Transactions, 44 (2012), pp. 381–385.
- [43] B. PEHERSTORFER, K. WILLCOX, AND M. GUNZBURGER, *Optimal Model Management for Multifidelity Monte Carlo Estimation*, SIAM Journal on Scientific Computing, 38 (2016), pp. A3163–A3194.
- [44] ———, *Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization*, SIAM Review, 60 (2018), pp. 550–591.
- [45] C. E. RASMUSSEN, *Gaussian Processes in Machine Learning*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 63–71.
- [46] M. T. REAGAN, H. N. NAJM, R. G. GHANEM, AND O. M. KNIO, *Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection*, Combustion and Flame, 132 (2003), pp. 545–555.
- [47] D. SCHADEN, *Variance Reduction with Multilevel Estimators*, PhD thesis, Technische Universität München, Munich, 2021.

- [48] D. SCHADEN AND E. ULLMANN, *On Multilevel Best Linear Unbiased Estimators*, SIAM/ASA Journal on Uncertainty Quantification, 8 (2020), pp. 601–635.
- [49] —, *Asymptotic Analysis of Multilevel Best Linear Unbiased Estimators*, SIAM/ASA Journal on Uncertainty Quantification, 9 (2021), pp. 953–978.
- [50] R. C. SMITH, *Uncertainty Quantification: Theory, Implementation, and Applications*, Society for Industrial and Applied Mathematics, USA, 2013.
- [51] B. SUDRET, *Global sensitivity analysis using polynomial chaos expansions*, Reliability Engineering & System Safety, 93 (2008), pp. 964–979.
- [52] T. J. SULLIVAN, *Introduction to Uncertainty Quantification*, vol. 63, Springer Cham, 2015.
- [53] A. L. TECKENTRUP, *Multilevel Monte Carlo Methods and Uncertainty Quantification*, PhD thesis, University of Bath, 2013.
- [54] R. K. TRIPATHY AND I. BILIONIS, *Deep uq: Learning deep neural network surrogate models for high dimensional uncertainty quantification*, Journal of Computational Physics, 375 (2018), pp. 565–588.
- [55] N. WIENER, *The Homogeneous Chaos*, American Journal of Mathematics, 60 (1938), pp. 897–936.
- [56] H. YANG, Y. FUJII, K. W. WANG, AND A. A. GORODETSKY, *Control Variate Polynomial Chaos: Optimal Fusion of Sampling and Surrogates for Multifidelity Uncertainty Quantification*, arXiv:2201.10745 [stat], (2022).