A filtered multilevel Monte Carlo method for estimating the expectation of cell-centered discretized random fields

Jérémy Briant, Paul Mycek, Mayeul Destouches, Olivier Goux, Serge Gratton, Selime Gürol, Ehouarn Simon, Anthony T. Weaver

A filtered multilevel Monte Carlo method for estimating the expectation of cell-centered discretized random fields*

Jérémy Briant[1,2], Paul Mycek[2], Mayeul Destouches[2,3], Olivier Goux[2], Serge Gratton[1,4], Selime Gürol[2], Ehouarn Simon[1], and Anthony T. Weaver[2]

[1]IRIT, UMR 5505 CNRS-INP, Toulouse, France (jeremy.briant@toulouse-inp.fr, serge.gratton@toulouse-inp.fr, ehouarn.simon@toulouse-inp.fr).
[2]CECI, UMR 5318 CNRS-CERFACS, Toulouse, France (mycek@cerfacs.fr, goux@cerfacs.fr, gurol@cerfacs.fr, weaver@cerfacs.fr).
[3]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France (mayeul.destouches@meteo.fr).
[4]ANITI, Toulouse, France.

Abstract

In this paper, we investigate the use of multilevel Monte Carlo (MLMC) methods for estimating the expectation of discretized random fields. Specifically, we consider a setting in which the input and output vectors of numerical simulators have inconsistent dimensions across the multilevel hierarchy. This requires the introduction of grid transfer operators borrowed from multigrid methods. By adapting mathematical tools from multigrid methods, we perform a theoretical spectral analysis of the MLMC estimator of the expectation of discretized random fields, in the specific case of linear, symmetric and circulant simulators. We then propose filtered MLMC (F-MLMC) estimators based on a filtering mechanism similar to the smoothing process of multigrid methods, and we show that the filtering operators improve the estimation of both the small- and large-scale components of the variance, resulting in a reduction of the total variance of the estimator. Next, the conclusions of the spectral analysis are experimentally verified with a one-dimensional illustration. Finally, the proposed F-MLMC estimator is applied to the problem of estimating the discretized variance field of a diffusion-based covariance operator, which amounts to estimating the expectation of a discretized random field. The numerical experiments support the conclusions of the theoretical analysis even with non-linear simulators, and demonstrate the improvements brought by the F-MLMC estimator compared to both a crude MC and an unfiltered MLMC estimator.

Keywords: Multilevel Monte Carlo, multigrid method, random field, spectral analysis, filtering, diffusion operator.

1 Introduction

Monte Carlo (MC) estimation (or integration) refers to a class of statistical methods that rely on the sampling of random quantities to construct statistical estimators. Such MC estimators are popular and widely-used owing to their simplicity and flexibility. However, their slow convergence in terms

 This project has received financial support from the CNRS (Centre National de la Recherche Scientifique) through the 80|Prime program and the French national program LEFE (Les Enveloppes Fluides et l'Environnement).

of their root mean square error (RMSE) with respect to the sample size makes them inefficient or even unaffordable for the estimation of statistics of outputs of computationally expensive numerical simulators. In recent years, multifidelity MC methods (see, e.g., [1]) have grown in popularity as a means to accelerate traditional MC methods by combining simulators of different fidelities to leverage the lower computational cost of lower-fidelity simulators while preserving the accuracy of the high-fidelity simulator. Such lower fidelity numerical simulators can be obtained, e.g., from a coarser spatial and/or temporal discretization as is the case in multigrid methods [2–5]. In such instances, the different fidelities are referred to as levels, and the multifidelity method is then referred to as a multilevel method.

Over the past few decades, numerous multifidelity estimation methods have been proposed [1, 6–9]. Among them, the multilevel Monte Carlo (MLMC) method [6, 7, 10], which is the focus of the present paper, combines samples (or ensembles) from different levels in such a way that, under certain assumptions, the variance of the resulting multilevel estimator is reduced, while leaving the bias unchanged. The closely-related approximate control variate estimators and multilevel best linear unbiased estimators (MLBLUEs) [9, 11] have since been proposed as general methodologies for optimally combining (in terms of variance reduction) samples of different fidelities. Originally designed for the estimation of the expected value of scalar, real-valued random variables, the MLMC methodology has since been extended to the estimation of higher-order statistical moments [12, 13] and variance-based global sensitivity measures [14]. The analysis of MLMC estimators has also been extended to the estimation of statistics of random variables with values in separable Hilbert spaces [12, 13]. Likewise, strategies for extending MLBLUEs to the estimation of statistical moments of multiple simulator outputs were proposed in [15–17].

The estimation of covariance matrices is a prominent example where MLMC approaches have started to emerge, especially in the field of data assimilation [15, 18–21]. Another example, which will constitute the motivating example of this paper, is embedded in the general approach of using a discretized linear differential operator to represent the application of a parametric form of a covariance matrix. In particular, we focus on an approach that uses a discretized diffusion operator to represent a covariance operator with a parametric kernel from the Matérn family [22]. Diffusion operators are commonly used for modelling spatial covariances in ocean data assimilation [23] and are closely related to other techniques for modelling spatial covariances in atmospheric data assimilation [24], geostatistical modelling [25], inverse problems [26] and uncertainty quantification [27]. Central to the approach is the need to extract the diagonal elements (*intrinsic* variances) of the diffusion-modelled covariance matrix so that they can be used to normalize the matrix, to transform it into a (unit-diagonal) correlation matrix. Once the covariance matrix is properly normalized, a desired variance field, different from the intrinsic variance field, can be imposed. The standard method for estimating the intrinsic variances of the diffusion-based covariance matrix is the randomization method [23, 28]. This method relies on the MC estimation of the expectation of a discretized random field, which, in a more abstract formulation, may be viewed as the output of a numerical simulator, whose input is also a discretized random field.

In this paper, we are interested in applying the MLMC methodology to improve the efficiency of estimating the expectation of cell-centered discretized random fields in the abstract setting described above. The main specificities of this setting are that, first, the considered simulators are based on a hierarchy of cell-centered discretizations on grids of different resolution; second, the input and output of the numerical simulators are discretized fields whose dimensions depend on the level; and third, we focus on the discrete representation (as opposed to a functional representation) of these discretized fields, arising from the discretization of the strong form of the differential operators at hand (e.g., in our example, diffusion operators). These specificities impose the introduction of grid transfer operators, namely restriction and prolongation operators, which may be borrowed from multigrid methods [2–5]. Using MLMC with simulators whose inputs and/or outputs are discretized fields has been studied before, typically in combination with the finite element (FE) method [29–33], whose functional framework allows the use of natural grid transfer operators, especially prolongation operators, which

are then usually only implicitly defined. Therefore, their impact on the quality of the resulting multilevel estimator is generally not studied. Nonetheless, the introduction of grid transfer operators raises crucial questions about their effect on the different spatial scales (or, equivalently, frequencies) of the discretized fields (or signals) that are transferred between grids of different resolution.

The question of the representation of small scales on coarse grids within the MLMC framework has recently been explored by [33] in an FE setting where the input of the considered simulator is a discretized random field. The authors then proposed the idea of smoothing highly oscillatory input signals based on a spectral truncation of their high-frequency components, which is controlled for each MLMC level so that the truncation error matches the discretization error. In this paper, we investigate these important questions in the abstract setting described above. Our objective is to gain a better understanding of the impact of grid transfer operators on the quality of the MLMC estimator in order to propose remedies for their negative effects. In multigrid methods [2–5, 34], studying the effects of grid transfer operators on the different frequencies of a discrete signal, which is typically achieved through a spectral analysis, is core to determining the effectiveness of the method. Along the same lines, we present a spectral analysis of the MLMC estimator in a simplified setting where the numerical simulators are assumed to be linear, symmetric and circulant operators. This leads us to propose a novel, filtered MLMC estimator that mitigates the negative effects of the grid transfer operators on the quality of the multilevel estimator.

The remainder of this paper is organized as follows. Section 2 introduces MLMC estimators of the expectation of random fields, in particular the adaptations of the framework in [12] to our specific setting, and positions our contribution with respect to related work. A spectral analysis of the MLMC estimator is then carried out in section 3 to investigate the effects of grid transfer on its variance, in the specific case of linear, symmetric and circulant numerical simulators. Following the conclusions drawn from the previous analysis, a filtered MLMC estimator, named F-MLMC, is introduced in section 4 and compared with its unfiltered counterpart through the same spectral analysis. In section 5, we apply the MLMC and F-MLMC estimators, first, to a one-dimensional illustration that comply with the assumptions of the spectral analysis, and, second, to our target problem of estimating the intrinsic variances of a two-dimensional, heterogeneous, diffusion-based covariance operator. General conclusions are drawn in section 6, along with prospective avenues for future work.

2 MLMC estimation of the expectation of discretized random fields

The MLMC method aims to improve the accuracy of MC estimators by combining samples of different fidelities. In favorable cases, large low-fidelity samples (i.e., with many members that are cheap to generate on coarse levels) are used to reduce the sampling error (variance), while smaller samples are required at finer and more expensive levels to correct the bias. Under certain assumptions, [29, Theorem 1] ensures that there exists a sample allocation on a finite number of levels such that the computational cost of the MLMC estimator decreases at a faster rate as a function of the mean square error (MSE), than that of the crude MC estimator. In practice, MLMC is typically implemented as a sequential algorithm, whose main idea is to start with a limited number of coarse fidelity levels, and add as many finer levels as needed to reach a target MSE, with a prescribed variance/bias balance. In the present work, however, we adopt a multilevel approach that is closer to multifidelity methods. Specifically, a fine, high-fidelity level is fixed (and thus so is the bias), while coarser, low-fidelity levels are considered to reduce the variance, for a prescribed computational budget.

In this paper, we focus on the multilevel estimation of the expectation of cell-centered discretized random fields. We assume that such discretized random fields can be represented as random variables defined on some probability space $(\Omega, \Sigma, \mathbb{P})$, with values in separable Hilbert spaces and having finite second-order moment. We thus rely on the mathematical framework proposed in [12] for the design

and analysis of MC and MLMC estimators in such a setting. Specifically, given a separable Hilbert space H equipped with the inner product $\langle\cdot,\cdot\rangle_H$ and induced norm $\|\cdot\|_H$, the space of second-order H-valued random variables on $(\Omega,\Sigma,\mathbb{P})$,

$$L^2(\Omega,H) := \{\xi\colon \Omega\to H \mid \textstyle\int_\Omega \|\xi(\omega)\|_H^2 \,\mathrm{d}\mathbb{P}(\omega) < +\infty\}, \tag{1}$$

is a Hilbert space when equipped with the inner product $\langle\cdot,\cdot\rangle_{L^2(\Omega,H)}$ defined as

$$\forall \xi,\eta\in L^2(\Omega,H),\quad \langle\xi,\eta\rangle_{L^2(\Omega,H)} := \int_\Omega \langle\xi(\omega),\eta(\omega)\rangle_H \,\mathrm{d}\mathbb{P}(\omega), \tag{2}$$

whose induced norm is denoted by $\|\cdot\|_{L^2(\Omega,H)}$. More specifically, we are interested in discretized random fields that correspond to the output of a numerical simulator whose input is also a discretized random field. Formally, we consider an abstract, deterministic numerical simulator as $f\colon H_{\text{in}}\to H_{\text{out}}$, where H_{in} and H_{out} are two separable Hilbert spaces, and we assume that f is such that $f(X)\in L^2(\Omega,H_{\text{out}})$ for any $X\in L^2(\Omega,H_{\text{in}})$. Then, given an input $X\in L^2(\Omega,H_{\text{in}})$, the Bochner integral $\mathbb{E}[f(X)] := \int_\Omega f(X(\omega))\mathrm{d}\mathbb{P}(\omega)$, which we shall refer to as the expectation of the output $f(X)$, is well-defined and we have $\mathbb{E}[f(X)]\in H_{\text{out}}$ [35, 36]. From here on, for the sake of exposition and without loss of generality, we restrict ourselves to the case where $H_{\text{in}} = H_{\text{out}} = H$.

We now turn to the multilevel estimation of $\mathbb{E}[f(X)]$, which relies on a hierarchy (sequence) of $L+1$ numerical simulators $(f_\ell)_{\ell=0}^L$. First, in section 2.1, we describe the setting where the input and ouput Hilbert space H is the same across the simulators and briefly recall the tools introduced in [12] for the analysis of the MLMC expectation estimator in this setting. Next, in section 2.2, we examine the more general case where the input and output spaces H_ℓ of the simulators vary across the levels of the hierarchy, and introduce transfer operators between these spaces, eventually leading to a multilevel estimator for which the tools of [12] remain relevant. We then we present in section 2.3 the particular choice of spaces H_ℓ that will be considered in the remainder of the paper, arising from the discrete approximation of differential operators. Finally, we position our setting with respect to a non-exhaustive, yet representative sample of related work in section 2.4.

2.1 MLMC with consistent input and output space

Let $X\in L^2(\Omega,H)$ be the H-valued random variable corresponding to a discretized random field. We consider a multilevel hierarchy of $L+1$ abstract numerical models $(f_\ell\colon H\to H)_{\ell=0}^L$ of increasing fidelity. We are interested in estimating the expectation of the output of the finest (highest-fidelity) model, $\mathbb{E}[f_L(X)]$. The MLMC estimator $\hat{\mu}_L^{\text{MLMC}}$ of $\mathbb{E}[f_L(X)]$, using a sequence $(\{X^{(\ell,i)}\}_{i=1}^{M_\ell})_{\ell=0}^L$ of $L+1$ independent samples of X, is defined as

$$\hat{\mu}_L^{\text{MLMC}} = \frac{1}{M_0}\sum_{i=1}^{M_0} f_0(X^{(0,i)}) + \sum_{\ell=1}^{L}\frac{1}{M_\ell}\sum_{i=1}^{M_\ell}\left[f_\ell(X^{(\ell,i)}) - f_{\ell-1}(X^{(\ell,i)})\right] \in L^2(\Omega,H). \tag{3}$$

The accuracy of this MLMC estimator may be quantified by its normwise MSE with respect to some $\mu\in H$, defined by (see, e.g., [12])

$$\mathrm{MSE}(\hat{\mu}_L^{\text{MLMC}},\mu) := \|\hat{\mu}_L^{\text{MLMC}} - \mu\|_{L^2(\Omega,H)}^2. \tag{4}$$

As shown in [12, Theorem 3.1], the MSE admits the decomposition

$$\mathrm{MSE}(\hat{\mu}_L^{\text{MLMC}},\mu) = \mathcal{V}(\hat{\mu}_L^{\text{MLMC}}) + \|\mathbb{E}[f_L(X)] - \mu\|_H^2, \tag{5}$$

where

$$\forall Y\in L^2(\Omega,H),\quad \mathcal{V}(Y) := \|Y-\mathbb{E}[Y]\|_{L^2(\Omega,H)}^2 = \mathbb{E}[\|Y\|_H^2] - \|\mathbb{E}[Y]\|_H^2, \tag{6}$$

with the shorthand notation $\mathbb{E}[\|\cdot\|_H^2]^{1/2} = \|\cdot\|_{L^2(\Omega,H)}$. The first term in eq. (5) is referred to as the variance of the multilevel estimator, while the second term corresponds to the squared bias. Furthermore, [12, Theorem 3.1] shows that the variance $\mathcal{V}(\hat{\mu}_L^{\mathrm{MLMC}})$ can be further decomposed level-wise into

$$\mathcal{V}(\hat{\mu}_L^{\mathrm{MLMC}}) = \frac{1}{M_0}\mathcal{V}(f_0(X)) + \sum_{\ell=1}^{L} \frac{1}{M_\ell}\mathcal{V}(f_\ell(X) - f_{\ell-1}(X)). \tag{7}$$

We note that the multilevel estimator is unbiased with respect to the expectation of the output at the finest level $\mu_L := \mathbb{E}[f_L(X)]$, so that $\mathrm{MSE}(\hat{\mu}_L^{\mathrm{MLMC}}, \mu_L) = \mathcal{V}(\hat{\mu}_L^{\mathrm{MLMC}})$.

2.2 MLMC with inconsistent input and output spaces

In typical applications, especially when the simulators are related to discretized partial differential equations (PDEs), and that their fidelity is dictated by the quality of the discretization, the input and output spaces H_ℓ are typically finite-dimensional and generally not the same across the simulators of the hierarchy. The multilevel hierarchy of numerical simulators then becomes $(\tilde{f}_\ell \colon H_\ell \to H_\ell)_{\ell=0}^L$, and we are interested in the expectation of the highest-fidelity output, $\mathbb{E}[\tilde{f}_L(X_L)] \in H_L$, for some input $X_L \in L^2(\Omega, H_L)$. As a consequence, the MLMC estimator defined in eq. (3) cannot be used directly, and transfer operators need to be introduced for the transfer of inputs and outputs between the highest-fidelity space H_L and lower-fidelity spaces H_ℓ with $\ell = 0, \dots, L-1$. Operators that transfer a discretized field from a level ℓ to a higher-fidelity level $\ell' > \ell$ will be referred to as prolongation operators. We denote by $P_\ell^{\ell'} \colon H_\ell \to H_{\ell'}$ the prolongation operator from level ℓ to a higher-fidelity level $\ell' > \ell$. In addition, restriction operators $R_{\ell'}^{\ell} \colon H_{\ell'} \to H_\ell$ are used to perform the "reverse" operation of transferring discretized fields defined on level ℓ' to a lower-fidelity level $\ell < \ell'$. We may now define f_ℓ from $\tilde{f}_\ell$ as

$$f_\ell = P_\ell^L \circ \tilde{f}_\ell \circ R_L^\ell \colon H_L \to H_L, \tag{8}$$

for $\ell = 0, \dots, L$. The numerical simulators in eq. (8) involve transfer operators between the highest-fidelity level L and lower-fidelity levels. In practice, grid transfer operators between arbitrary levels can be defined as the composition of transfer operators between successive levels, as

$$\begin{aligned} \forall \ell = 0, \dots, L-1, \quad \forall \ell' = \ell+1, \dots, L, \quad & P_\ell^{\ell'} := P_{\ell'-1}^{\ell'} \circ P_{\ell'-2}^{\ell'-1} \circ \cdots \circ P_{\ell+1}^{\ell+2} \circ P_\ell^{\ell+1}, \\ & R_{\ell'}^{\ell} := R_{\ell+1}^{\ell} \circ R_{\ell+2}^{\ell+1} \circ \cdots \circ R_{\ell'-1}^{\ell'-2} \circ R_{\ell'}^{\ell'-1}, \end{aligned} \tag{9}$$

and, for $\ell = 0, \dots, L$, $R_\ell^\ell = P_\ell^\ell = I_\ell$, where $I_\ell \colon H_\ell \to H_\ell$ is the identity operator on H_ℓ. We further assume that, for any $X \in L^2(\Omega, H_\ell)$,

- $\tilde{f}_\ell(X) \in L^2(\Omega, H_\ell)$, for $\ell = 0, \dots, L$;
- $P_\ell^{\ell+1}(X) \in L^2(\Omega, H_{\ell+1})$, for $\ell = 0, \dots, L-1$;
- $R_\ell^{\ell-1}(X) \in L^2(\Omega, H_{\ell-1})$, for $\ell = 1, \dots, L$.

It follows that, for any $X_L \in L^2(\Omega, H_L)$ and for $\ell = 0, \dots, L$, $f_\ell(X_L) \in L^2(\Omega, H_L)$. Consequently, the MLMC expression eq. (3) can now be used with $H = H_L$ and $X = X_L \in L^2(\Omega, H_L)$.

In this paper, we are particularly interested in the impact of the grid transfer operators on the quality of the resulting multilevel estimator. Specifically, we shall examine their impact on the representation of the different scales (or frequencies) of the multilevel correction fields. While it is clear that, owing to the telescoping correction mechanism, the bias of the estimator is unaffected, the impact on its variance, and hence on the MSE, remains to be investigated.

2.3 Discrete finite-dimensional spaces

In this paper, we focus on numerical simulators that arise from the discrete approximation of the strong form of elliptic PDEs, i.e., through the direct discretization of the differential operators involved in such PDEs. This is the case, for instance, in the finite difference (FD) or finite volume (FV) discretization of a PDE. Such discretization methods have to be distinguished from methods relying on a weak (or variational) formulation of the PDE, like the finite element (FE) method, as we shall see in section 2.4. This choice is motivated by the considered application, presented in section 5.1, and more importantly by the target ocean numerical model [37], which relies on an FD-like discretization. Consequently, input and ouput discretized fields of such simulators may only be evaluated at a finite number of discrete locations, and are thus represented as finite-dimensional real vectors, whose size depends on the number of discretization points in the spatial domain of interest $\mathcal{D}$. Formally, we consider a polytopal tessellation $\mathcal{T}$ of a polytopal domain $\mathcal{D} \subset \mathbb{R}^d$, with $d \in \{1, 2, 3\}$. We refer to elements of $\mathcal{T}$ as cells, and we denote by $|\mathcal{T}|$ the cardinality, or size, of $\mathcal{T}$, i.e., its number of cells. In what follows, we consider cell-centered discretizations, meaning that the discretization points are located at the center of the cells of $\mathcal{T}$. It should be noted that functional representations can be deduced *a posteriori* from such discretized fields. In particular, if $\mathbf{u} \in \mathbb{R}^n$ is a cell-centered discretized field on a tessellation $\mathcal{T} := \{T_i\}_{i=1}^n$, a natural functional representation is a piecewise constant field $u \in \mathbb{P}^0$, where $\mathbb{P}^0 := \mathrm{span}(\{\mathbb{1}_{T_i}\}_{i=1}^n)$ and $\mathbb{1}_T \colon \mathcal{D} \to \{0, 1\}$ denotes the indicator function associated with the subset $T \subseteq \mathcal{D}$. The piecewise constant representation can then be defined through $\mathbf{u} \in \mathbb{R}^n$ as $u = \mathbf{u}^{\mathrm{T}}\boldsymbol{\phi}$, where $\boldsymbol{\phi} := [\mathbb{1}_{T_1}, \dots, \mathbb{1}_{T_n}]^{\mathrm{T}}$.

We now describe the MLMC hierarchy of simulators (and corresponding Hilbert spaces) for this discretization setting. Let $(\mathcal{T}_\ell)_{\ell=0}^L$ be a sequence of $L+1$ tessellations of $\mathcal{D}$ of increasing size $n_\ell := |\mathcal{T}_\ell|$, for $\ell = 0, \dots, L$. We further assume that the tessellations are nested, i.e., for $\ell = 1, \dots, L$, for any $T \in \mathcal{T}_\ell$, there is a unique $T' \in \mathcal{T}_{\ell-1}$ such that $T \subset T'$. We note, however, that the sets of discretization points, i.e., the sets of cell centers, are not nested. These tessellations of different sizes define the geometric hierarchy of levels of the MLMC approach. Specifically, $\mathcal{T}_0$ is the coarsest tessellation corresponding to the lowest-fidelity level, while $\mathcal{T}_L$ is the finest tessellation corresponding to the highest-fidelity level. To each level $\ell = 0, \dots, L$, we associate the symmetric positive definite (SPD) Gram matrix $\mathbf{W}_\ell \in \mathbb{R}^{n_\ell \times n_\ell}$, which encodes the structural information related to the discrete approximation of scalar fields as vectors of $\mathbb{R}^{n_\ell}$ on the corresponding cell-centered discretization on $\mathcal{T}_\ell$. We then define the weighted inner product $\langle \cdot, \cdot \rangle_{\mathbf{W}_\ell}$ between elements of $\mathbb{R}^{n_\ell}$ as

$$\forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^{n_\ell}, \quad \langle \mathbf{u}, \mathbf{v} \rangle_{\mathbf{W}_\ell} = \mathbf{u}^{\mathrm{T}} \mathbf{W}_\ell \mathbf{v} = \langle \mathbf{V}_\ell^{\mathrm{T}} \mathbf{u}, \mathbf{V}_\ell^{\mathrm{T}} \mathbf{v} \rangle_{\mathbf{I}_{n_\ell}}, \tag{10}$$

where $\mathbf{V}_\ell \in \mathbb{R}^{n_\ell \times n_\ell}$ is an invertible matrix such that $\mathbf{W}_\ell = \mathbf{V}_\ell \mathbf{V}_\ell^{\mathrm{T}}$, which exists since $\mathbf{W}_\ell$ is SPD, and where $\langle \cdot, \cdot \rangle_{\mathbf{I}_{n_\ell}}$ denotes the canonical (Euclidean) dot product between vectors of $\mathbb{R}^{n_\ell}$. The inner product space $H_\ell := (\mathbb{R}^{n_\ell}, \langle \cdot, \cdot \rangle_{\mathbf{W}_\ell})$ is a separable Hilbert space, so that the multilevel hierarchy of simulators $(\tilde{f}_\ell \colon H_\ell \to H_\ell)_{\ell=0}^L$ fits in the framework described in section 2.2. For clarity, we shall from here on simply write $\mathbb{R}^{n_\ell}$ for $(\mathbb{R}^{n_\ell}, \langle \cdot, \cdot \rangle_{\mathbf{W}_\ell})$, and, unless explicitly stated otherwise, use $\langle \cdot, \cdot \rangle_{\mathbf{W}_\ell}$ as the default inner product between vectors of $\mathbb{R}^{n_\ell}$. Furthermore, we shall denote by $\| \cdot \|_{\mathbf{W}_\ell}$ the norm induced by $\langle \cdot, \cdot \rangle_{\mathbf{W}_\ell}$.

For convenience, the transfer operators introduced in section 2.2 are typically chosen to be linear. Consequently, a linear prolongation operator $P_\ell^{\ell'}$ may be identified with the matrix $\mathbf{P}_\ell^{\ell'} \in \mathbb{R}^{n_{\ell'} \times n_\ell}$ such that, for any $\mathbf{x}_\ell \in \mathbb{R}^{n_\ell}$, $P_\ell^{\ell'}(\mathbf{x}_\ell) = \mathbf{P}_\ell^{\ell'} \mathbf{x}_\ell$, and, likewise, a linear restriction operator $R_{\ell'}^{\ell}$ may be identified with the appropriate matrix $\mathbf{R}_{\ell'}^{\ell} \in \mathbb{R}^{n_\ell \times n_{\ell'}}$. In the remainder of this paper, we will only consider linear transfer operators, and hence employ the associated matrix notation. The MLMC estimator eq. (3)

can then be re-written as

$$\hat{\boldsymbol{\mu}}_L^{\mathrm{MLMC}} = \frac{1}{M_0}\sum_{i=1}^{M_0} \mathbf{P}_0^L \tilde{f}_0(\mathbf{R}_L^0 \mathbf{X}_L^{(0,i)}) + \sum_{\ell=1}^{L} \frac{1}{M_\ell}\sum_{i=1}^{M_\ell} \left[\mathbf{P}_\ell^L \tilde{f}_\ell(\mathbf{R}_L^\ell \mathbf{X}_L^{(\ell,i)}) - \mathbf{P}_{\ell-1}^L \tilde{f}_{\ell-1}(\mathbf{R}_L^{\ell-1} \mathbf{X}_L^{(\ell,i)})\right], \tag{11}$$

$$= \frac{1}{M_0}\mathbf{P}_0^L \sum_{i=1}^{M_0} \tilde{f}_0(\mathbf{R}_L^0 \mathbf{X}_L^{(0,i)}) + \sum_{\ell=1}^{L} \frac{1}{M_\ell}\mathbf{P}_\ell^L \sum_{i=1}^{M_\ell} \left[\tilde{f}_\ell(\mathbf{R}_L^\ell \mathbf{X}_L^{(\ell,i)}) - \mathbf{P}_{\ell-1}^\ell \tilde{f}_{\ell-1}(\mathbf{R}_L^{\ell-1} \mathbf{X}_L^{(\ell,i)})\right], \tag{12}$$

where the second identity is obtained by linearity of the prolongation operators and the composition definition in eq. (9). The latter is to be preferred in practical implementations, as it involves fewer applications of the prolongation operators. Finally, we impose

$$\forall \ell = 0, \dots, L-1, \quad (\mathbf{P}_\ell^{\ell+1})^{\mathrm{T}} \mathbf{W}_{\ell+1} \mathbf{P}_\ell^{\ell+1} = \mathbf{W}_\ell. \tag{13}$$

A major consequence of eq. (13) is that

$$\forall \ell = 0, \dots, L-1,\ \forall \mathbf{x}_\ell \in \mathbb{R}^{n_\ell}, \quad \|\mathbf{P}_\ell^{\ell+1}\mathbf{x}_\ell\|_{\mathbf{W}_{\ell+1}} = \mathbf{x}_\ell^{\mathrm{T}} (\mathbf{P}_\ell^{\ell+1})^{\mathrm{T}} \mathbf{W}_{\ell+1} \mathbf{P}_\ell^{\ell+1} \mathbf{x}_\ell = \mathbf{x}_\ell^{\mathrm{T}} \mathbf{W}_\ell \mathbf{x}_\ell = \|\mathbf{x}_\ell\|_{\mathbf{W}_\ell}, \tag{14}$$

indicating that the prolongation operator is norm-preserving. We note that, by eq. (9), similar properties then hold for transfer operators between arbitrary levels.

2.4 Related work

Using MLMC in a setting where the numerical simulators are related to the discretization of a PDE and whose inputs and/or outputs are discretized random fields has been investigated before, mostly in combination with the FE discretization method. We briefly describe some of these earlier works in terms of input and output spaces and discretization method, and mention the differences with the setting described in the previous section, where $\tilde{f}_\ell \colon \mathbb{R}^{n_\ell} \to \mathbb{R}^{n_\ell}$, which is the focus of the present paper.

For instance, [29–33] consider an elliptic PDE with random (scalar or tensor) diffusivity fields, discretized using the FE method. In [29, 30], the numerical simulators are of the form $\tilde{f}_\ell \colon \mathbb{R}^K \to \mathbb{R}$. Specifically, the random diffusivity field is discretized through a truncated Karhunen-Loève expansion [38–40] of its logarithm, so that the field is parameterized by a finite collection of K (uncorrelated) real-valued random variables, with K fixed and independent of the discretization level ℓ. As a consequence, restriction operators are not needed. In addition, the output is a scalar quantity, and thus prolongation operators are not needed either. In [31, 32], the output of $\tilde{f}_\ell$ is defined as the discretized solution of the elliptic PDE in a finite-dimensional, FE function space V_ℓ on $\mathcal{D} \subset \mathbb{R}^d$ (d being the spatial dimension of the problem). The analysis is performed in a setting where the random diffusivity field is parameterized by a countable, possibly infinite collection of random variables, and the experiments are performed in a setting where the field can be exactly parameterized by a single random variable. We note that in [41], the authors consider a similar setting using an approximate parameterization with a finite (and level-dependent) number of random variables. The numerical simulators are thus of the general form $\tilde{f}_\ell \colon H_\ell \to V_\ell$, with $H_\ell = \mathbb{R}^{\mathbb{N}}$ or $H_\ell = \mathbb{R}^{K_\ell}$, where $K_\ell \in \mathbb{N}$ may be either level-dependent or constant across levels. Again, in such settings, restriction operators are not required. The considered output FE spaces being nested, i.e., $V_{\ell-1} \subset V_\ell$ for $\ell > 0$, the prolongation operators are implicitly defined by the canonical injection $\imath \colon V_{\ell-1} \hookrightarrow V_\ell$ between two successive spaces. In [33], the numerical simulators are of the form $\tilde{f}_\ell \colon \mathbb{R}^{n_\ell} \to \mathbb{R}$ so that prolongation operators are not needed. The discrete representation of the logarithm of the diffusivity field is sampled at the nodes of a Cartesian grid of size n_ℓ using the circulant embedding technique (see, e.g., [42, 43]). The inputs are here truly discrete by nature, and, as in our setting, their size is dictated by the grid size. Owing to the nestedness of the considered Cartesian grids and of their nodal discretization points, the restriction simply reduces to a selection operator, which selects the vector entries associated with the desired nodal locations on

the coarser grid. Nonetheless, this restriction operation deteriorates the representation of the input field, as small-scale (i.e., high-frequency) components cannot be captured on the coarse grid. This observation led the authors of [33] to introduce a filtering of the high frequencies of the input field prior to the restriction operation. In the circulant embedding framework, this can be achieved via spectral truncation. A similar high-frequency filtering (or smoothing) mechanism is also at the heart of the filtered MLMC estimator proposed in section 4, with the difference that the proposed filters are not spectral. In our setting, where the outputs of the simulators are discretized fields as well, we also introduce a post-prolongation high-frequency filtering mechanism, and we demonstrate numerically that both pre-restriction and post-prolongation filtering operations are necessary to damp the high-frequency pollution induced by the grid transfer operators.

In [44, 45], the authors consider cell-centered, FV discretizations of hyperbolic systems of conservation laws. In particular, [45] focuses on the shallow water equations discretized on a one- or two-dimensional Cartesian grid, with uncertain initial data and bed topography, which are both modelled as discretized random (vector) fields. Similar to [29–32, 41], these input random fields are parameterized by a finite collection of K random variables, so that the numerical simulators are of the form $\tilde{f}_\ell \colon \mathbb{R}^K \to \mathbb{R}^{mn_\ell}$, with $m \in \mathbb{N}$ corresponding to the number of components of the output vector field of interest. For the analysis, each of the m components of the discretized output field is interpreted as a piecewise constant field $u \in \mathbb{P}^0$. In such a setting, restriction operators are not needed, and, because the meshes of the multilevel hierarchy are assumed to be nested, the prolongation operators naturally correspond to the canonical injection.

In [46], the authors consider a problem closely related to the problem described in section 5.1, involving numerical simulators of the form $\tilde{f}_\ell \colon \mathbb{R}^{n_\ell} \to \mathbb{R}$, so that prolongation operators are not needed. Their MLMC construction reads

$$\hat{\mu}_L^{\mathrm{MLMC}} = \frac{1}{M_0} \sum_{i=1}^{M_0} \tilde{f}_0(\mathbf{X}_0^{(0,i)}) + \sum_{\ell=1}^{L} \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} \left[\tilde{f}_\ell(\mathbf{X}_\ell^{(\ell,i)}) - \tilde{f}_{\ell-1}(\mathbf{X}_{\ell-1}^{(\ell,i)}) \right], \tag{15}$$

which is different from eq. (11). In particular, we notice that eq. (15) does not require the use of restriction operators either. This construction stems from the fact that, for the specific problem considered in [46] in an FE setting, it is possible to jointly sample pairs $(\mathbf{X}_\ell, \mathbf{X}_{\ell-1})$ on successive levels in a way that ensures a stochastic coupling (correlation) between $\mathbf{X}_\ell$ and $\mathbf{X}_{\ell-1}$, for $\ell = 1, \dots, L$, which directly influences the variance of the correction terms, and hence the efficiency of the resulting multilevel estimator. Specifically, in [46], the fine (high-fidelity) random input vector is normally distributed with zero mean and covariance structure defined by the fine FE mass matrix $\mathbf{M}_L \in \mathbb{R}^{n_L \times n_L}$, i.e., $\mathbf{X}_L \sim \mathcal{N}(\mathbf{0}_{n_L}, \mathbf{M}_L)$. Then, the augmented vectors $\tilde{\mathbf{X}}_\ell := [\mathbf{X}_\ell^{\mathrm{T}} \quad \mathbf{X}_{\ell-1}^{\mathrm{T}}]^{\mathrm{T}}$ are defined for $\ell = 1, \dots, L$ as $\tilde{\mathbf{X}}_\ell \sim \mathcal{N}(\mathbf{0}_{n_\ell + n_{\ell-1}}, \tilde{\mathbf{M}}_\ell)$, with $\tilde{\mathbf{M}}_\ell$ given by (see [46, eq. (3.12)])

$$\tilde{\mathbf{M}}_\ell := \begin{bmatrix} \mathbf{M}_\ell & \mathbf{M}_{\ell,\ell-1} \\ \mathbf{M}_{\ell,\ell-1}^{\mathrm{T}} & \mathbf{M}_{\ell-1} \end{bmatrix} \in \mathbb{R}^{(n_\ell + n_{\ell-1}) \times (n_\ell + n_{\ell-1})}, \tag{16}$$

where $\mathbf{M}_\ell \in \mathbb{R}^{n_\ell \times n_\ell}$ denotes the FE mass matrix on level ℓ, while $\mathbf{M}_{\ell,\ell-1} \in \mathbb{R}^{n_\ell \times n_{\ell-1}}$ corresponds to a mixed mass matrix between levels ℓ and $\ell - 1$. On the coarsest level, $\mathbf{X}_0$ is simply defined as $\mathbf{X}_0 \sim \mathcal{N}(\mathbf{0}_{n_0}, \mathbf{M}_0)$. This construction ensures that $\mathbf{X}_\ell^{(\ell,i)}$ and $\mathbf{X}_\ell^{(\ell+1,j)}$ have the same distribution, for $\ell = 0, \dots, L-1$, $i = 1, \dots, M_\ell$ and $j = 1, \dots, M_{\ell+1}$. Hence, so do $\tilde{f}_\ell(\mathbf{X}_\ell^{(\ell,i)})$ and $\tilde{f}_\ell(\mathbf{X}_\ell^{(\ell+1,j)})$, thus ensuring that the multilevel estimator is unbiased, by the telescoping sum mechanism of eq. (15). This formulation requires the ability to sample Gaussian vectors with the covariance structure specified by $\tilde{\mathbf{M}}_\ell$. This can be achieved, e.g., using a Cholesky decomposition of $\tilde{\mathbf{M}}_\ell$, which can be computed efficiently in an FE framework [46]. In contrast, our multilevel estimator in eq. (11) circumvents this requirement by sampling all the input vectors on the finest level, before restricting them to the

appropriate, coarsest levels. As we shall see later in section 5, in our application, which also involves normally distributed input vectors, the multilevel estimator proposed in eq. (11) exhibits coupling properties similar to eq. (16).

In the above-mentioned references (except in [46]), we observe that whenever either of the input or the output is a discretized field, grid transfer operators are introduced. However, in the functional framework that is typically employed, those are only implicitly defined and usually not even mentioned. This is especially the case for the prolongation operator, since discretized fields are defined everywhere in the spatial domain through their functional representation. Nevertheless, the practical computation of the multilevel correction terms in eq. (3) implicitly requires the introduction of prolongation operators. By focusing on the discrete framework, we explicitly exhibit the grid transfer operators and study their effect on the quality of the multilevel estimator defined by eq. (11). We shall see that the operators that seem the most natural in the functional framework, such as the canonical injection for the prolongation operator in settings with nested approximation spaces, actually yield multilevel estimators with poor properties. This is particularly true for cell-centered discretizations, which are the focus of this paper, and which are also known in the multigrid community for causing similar difficulties [47–50]. In the next section, we perform a spectral analysis of the MLMC estimator defined by eq. (11) in a linear setting, inspired by typical spectral analyses of multigrid methods [3–5].

3 Spectral analysis

In this section, we conduct a spectral analysis of the MLMC estimator (11) to study more closely the effects of grid transfer operators on the variance of the multilevel estimator at different spatial scales. The analysis is performed in a one-dimensional spatial domain, with cell-centered discretized input and output fields on uniform grids. The setting is further simplified to the case where the numerical simulators are linear, and that their associated matrices are symmetric and circulant.

3.1 Linear simulators

We start by considering linear simulators of the form $\tilde{f}_\ell \colon \mathbf{x}_\ell \mapsto \tilde{\mathbf{F}}_\ell \mathbf{x}_\ell$, for $\ell = 0, \dots, L$, where $\tilde{\mathbf{F}}_\ell \in \mathbb{R}^{n_\ell \times n_\ell}$, and, consistently with eqs. (8) and (11), we define $\mathbf{F}_\ell := \mathbf{P}_\ell^L \tilde{\mathbf{F}}_\ell \mathbf{R}_L^\ell \in \mathbb{R}^{n_L \times n_L}$, with $\mathbf{P}_L^L = \mathbf{R}_L^L = \mathbf{I}_{n_L}$. For now, we make no further assumption on $\tilde{\mathbf{F}}_\ell$. Then, by linearity of the expectation operator, the variance $\mathcal{V}(\hat{\boldsymbol{\mu}}_L^{\mathrm{MLMC}})$ of the multilevel estimator (11), given by eq. (7), becomes

$$\mathcal{V}(\hat{\boldsymbol{\mu}}_L^{\mathrm{MLMC}}) = \frac{1}{M_0} \mathbb{E}[\|\mathbf{F}_0 \dot{\mathbf{X}}_L\|_{\mathbf{W}_L}^2] + \sum_{\ell=1}^{L} \frac{1}{M_\ell} \mathbb{E}[\|(\mathbf{F}_\ell - \mathbf{F}_{\ell-1}) \dot{\mathbf{X}}_L\|_{\mathbf{W}_L}^2], \tag{17}$$

where $\dot{\mathbf{X}}_L := \mathbf{X}_L - \mathbb{E}[\mathbf{X}_L]$. Now, for any matrix $\mathbf{F} \in \mathbb{R}^{n_L \times n_L}$,

$$\begin{aligned} \mathbb{E}[\|\mathbf{F}\dot{\mathbf{X}}_L\|_{\mathbf{W}_L}^2] &= \mathbb{E}[(\mathbf{F}\dot{\mathbf{X}}_L)^{\mathrm{T}} \mathbf{W}_L (\mathbf{F}\dot{\mathbf{X}}_L)] = \operatorname{tr} \mathbb{E}[\dot{\mathbf{X}}_L^{\mathrm{T}} \mathbf{F}^{\mathrm{T}} \mathbf{W}_L \mathbf{F} \dot{\mathbf{X}}_L] = \mathbb{E}[\operatorname{tr}(\dot{\mathbf{X}}_L^{\mathrm{T}} \mathbf{F}^{\mathrm{T}} \mathbf{W}_L \mathbf{F} \dot{\mathbf{X}}_L)] \\ &= \mathbb{E}[\operatorname{tr}(\mathbf{F}^{\mathrm{T}} \mathbf{W}_L \mathbf{F} \dot{\mathbf{X}}_L \dot{\mathbf{X}}_L^{\mathrm{T}})] = \operatorname{tr}(\mathbf{F}^{\mathrm{T}} \mathbf{W}_L \mathbf{F} \mathbb{E}[\dot{\mathbf{X}}_L \dot{\mathbf{X}}_L^{\mathrm{T}}]) = \operatorname{tr}(\mathbf{F}^{\mathrm{T}} \mathbf{W}_L \mathbf{F} \mathbf{G}) \\ &= \operatorname{tr}((\mathbf{F}\mathbf{G}^{1/2})^{\mathrm{T}} \mathbf{W}_L (\mathbf{F}\mathbf{G}^{1/2})) = \|\mathbf{F}\mathbf{G}^{1/2}\|_{F,\mathbf{W}_L}^2, \end{aligned} \tag{18}$$

where $\operatorname{tr}(\cdot)$ denotes the matrix trace operator, $\|\cdot\|_{F,\mathbf{W}} \colon \mathbf{F} \mapsto \operatorname{tr}(\mathbf{F}^{\mathrm{T}}\mathbf{W}\mathbf{F})^{1/2}$ denotes the $\mathbf{W}$-weighted Frobenius norm for any SPD weighting matrix $\mathbf{W}$, and $\mathbf{G} := \mathbb{E}[\dot{\mathbf{X}}_L \dot{\mathbf{X}}_L^{\mathrm{T}}] = \mathbf{G}^{1/2}(\mathbf{G}^{1/2})^{\mathrm{T}}$ denotes the covariance matrix of $\mathbf{X}_L$. Thus, eq. (17) reduces to

$$\mathcal{V}(\hat{\boldsymbol{\mu}}_L^{\mathrm{MLMC}}) = \frac{1}{M_0} \|\mathbf{F}_0 \mathbf{G}^{1/2}\|_{F,\mathbf{W}_L}^2 + \sum_{\ell=1}^{L} \frac{1}{M_\ell} \|(\mathbf{F}_\ell - \mathbf{F}_{\ell-1}) \mathbf{G}^{1/2}\|_{F,\mathbf{W}_L}^2, \tag{19}$$

which emphasizes that the variance reduction is closely related to the similarity of successive fidelity models. Furthermore, for any orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{n_L \times n_L}$ with respect to the weighted inner product $\langle \cdot, \cdot \rangle_{\mathbf{W}_L}$, i.e., such that $\mathbf{Q}\mathbf{W}_L\mathbf{Q}^{\mathrm{T}} = \mathbf{Q}^{\mathrm{T}}\mathbf{W}_L\mathbf{Q} = \mathbf{I}_{n_L}$, we have, for any $\mathbf{x} \in \mathbb{R}^{n_L}$,

$$\|\mathbf{Q}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}}\mathbf{x}\|_{\mathbf{W}_L} = \|\mathbf{Q}\mathbf{V}^{\mathrm{T}}\mathbf{x}\|_{\mathbf{W}_L} = \|\mathbf{x}\|_{\mathbf{W}_L}, \quad \text{and } \|\mathbf{Q}^{\mathrm{T}}\mathbf{x}\|_{\mathbf{W}_L} = \|\mathbf{Q}\mathbf{x}\|_{\mathbf{W}_L} = \|\mathbf{x}\|_{\mathbf{I}_{n_L}}, \tag{20}$$

and, similarly, for any $\mathbf{F} \in \mathbb{R}^{n_L \times n_L}$,

$$\|\mathbf{Q}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}}\mathbf{F}\|_{F,\mathbf{W}_L} = \|\mathbf{Q}\mathbf{V}^{\mathrm{T}}\mathbf{F}\|_{F,\mathbf{W}_L} = \|\mathbf{F}\|_{F,\mathbf{W}_L}, \quad \text{and } \|\mathbf{Q}^{\mathrm{T}}\mathbf{F}\|_{F,\mathbf{W}_L} = \|\mathbf{Q}\mathbf{F}\|_{F,\mathbf{W}_L} = \|\mathbf{F}\|_{F,\mathbf{I}_{n_L}}. \tag{21}$$

3.2 One-dimensional spatial setting

The analysis is conducted in a one-dimensional (1D) setting with periodic, cell-centered discretized input and output random fields defined on a unit spatial domain $\mathcal{D} := [0, 1) \subset \mathbb{R}$. Specifically, we consider a hierarchy of uniform grids $(\mathcal{T}_\ell)_{\ell=0}^L$, where $\mathcal{T}_\ell := \{T_{\ell,i}\}_{i=1}^{n_\ell}$ and $T_{\ell,i} := [(i-1)/n_\ell, i/n_\ell)$ on level ℓ, along with the associated Gram matrices $\mathbf{W}_\ell = n_\ell^{-1}\mathbf{I}_{n_\ell}$. It follows immediately that $\mathbf{V}_\ell = n_\ell^{-1/2}\mathbf{I}_{n_\ell} = \mathbf{W}_\ell^{1/2}$. For $\ell = 0, \dots, L-1$, we set a constant refinement factor $n_{\ell+1}/n_\ell = 2$, so that $n_\ell = 2^{\ell-L}n_L$. The inter-level transfer operators are defined as

$$\forall \ell = 0, \dots, L-1, \qquad \mathbf{P}_\ell^{\ell+1} := \begin{bmatrix} 1 & & \\ 1 & & \\ & 1 & \\ & 1 & \\ & & \ddots \end{bmatrix} \in \mathbb{R}^{n_{\ell+1} \times n_\ell}, \text{ and } \mathbf{R}_{\ell+1}^{\ell} = (\mathbf{P}_\ell^{\ell+1})^{\mathrm{T}} \in \mathbb{R}^{n_\ell \times n_{\ell+1}}. \tag{22}$$

Note that the prolongation operator in eq. (22) induces the canonical injection for the functional, piecewise constant representation mentioned in section 2.3. Indeed, letting $\mathbb{P}_\ell^0 := \mathrm{span}(\{\mathbb{1}_{T_{\ell,i}}\}_{i=1}^{n_\ell})$ denote the space of piecewise constant functions on $\mathcal{T}_\ell$ and letting $\imath_\ell^{\ell+1} \colon \mathbb{P}_\ell^0 \to \mathbb{P}_{\ell+1}^0$ denote the canonical injection from $\mathbb{P}_\ell^0$ to $\mathbb{P}_{\ell+1}^0$, then, for any $u_\ell = \mathbf{x}^{\mathrm{T}}\boldsymbol{\phi}_\ell \in \mathbb{P}_\ell^0$ with $\boldsymbol{\phi}_\ell := [\mathbb{1}_{T_{\ell,1}}, \dots, \mathbb{1}_{T_{\ell,n_\ell}}]^{\mathrm{T}}$, $\imath_\ell^{\ell+1}(u_\ell) = \mathbf{x}_{\ell+1}^{\mathrm{T}}\boldsymbol{\phi}_{\ell+1}$ with $\mathbf{x}_{\ell+1} = \mathbf{P}_\ell^{\ell+1}\mathbf{x}_\ell$. This prolongation operator satisfies eq. (13), which, in this setting, amounts to $(\mathbf{P}_\ell^{\ell+1})^{\mathrm{T}}\mathbf{P}_\ell^{\ell+1} = 2\mathbf{I}_{n_\ell}$, for $\ell = 0, \dots, L-1$. Then, grid transfer operators between arbitrary levels are defined by composition of successive inter-level operators as in eq. (9). We note that, in this particular setting, and assuming a piecewise constant representation of the discretized input fields, the covariance matrix in eq. (58) is equivalent to that in eq. (16).

3.3 The Hartley basis

The spectral analysis conducted below relies on the discrete Hartley basis [51–53], adapted here to cell-centered discretized fields. The Hartley basis is a Fourier-like basis, commonly used in circulant embedding techniques for generating stationary Gaussian random fields with prescribed covariance structure [33, 42, 43]. Its main advantage is that it consists of purely real basis vectors, as opposed to the Fourier basis, whose basis vectors are complex, thus easing interpretation and visualization. On level ℓ, the n_ℓ cell-centered Hartley basis vectors $\{\mathbf{h}_k^\ell\}_{k=0}^{n_\ell-1}$ correspond to the columns of the Hartley matrix $\mathbf{H}_\ell$ with entries

$$(\mathbf{H}_\ell)_{j,k} := \cos\frac{2(j+\frac{1}{2})k\pi}{n_\ell} + \sin\frac{2(j+\frac{1}{2})k\pi}{n_\ell}, \quad \forall j, k = 0, \dots, n_\ell - 1. \tag{23}$$

The matrices $\mathbf{H}_\ell$ are orthogonal with respect to their associated inner product $\langle \cdot, \cdot \rangle_{\mathbf{W}_\ell}$, i.e., $\mathbf{H}_\ell^{\mathrm{T}}\mathbf{H}_\ell = \mathbf{H}_\ell\mathbf{H}_\ell^{\mathrm{T}} = n_\ell\mathbf{I}_{n_\ell}$, for $\ell = 0, \dots, L$ (see appendix A). Figure 1 depicts the basis vectors $\mathbf{h}_k^1$ and $\mathbf{h}_k^0$ of fine-

and coarse-grid Hartley basis vectors, discretized on grids with $n_1 = 16$ and $n_0 = 8$ cells, respectively. These plots highlight that, because of aliasing, the basis vectors exhibit a discrete frequency that is different from their continuous counterpart. Specifically, for $\ell \in \{0, 1\}$, the vectors indexed by k close to 0 or $n_\ell - 1$ are discrete signals with low frequency, while their frequency increases as k tends to $n_\ell/2$.

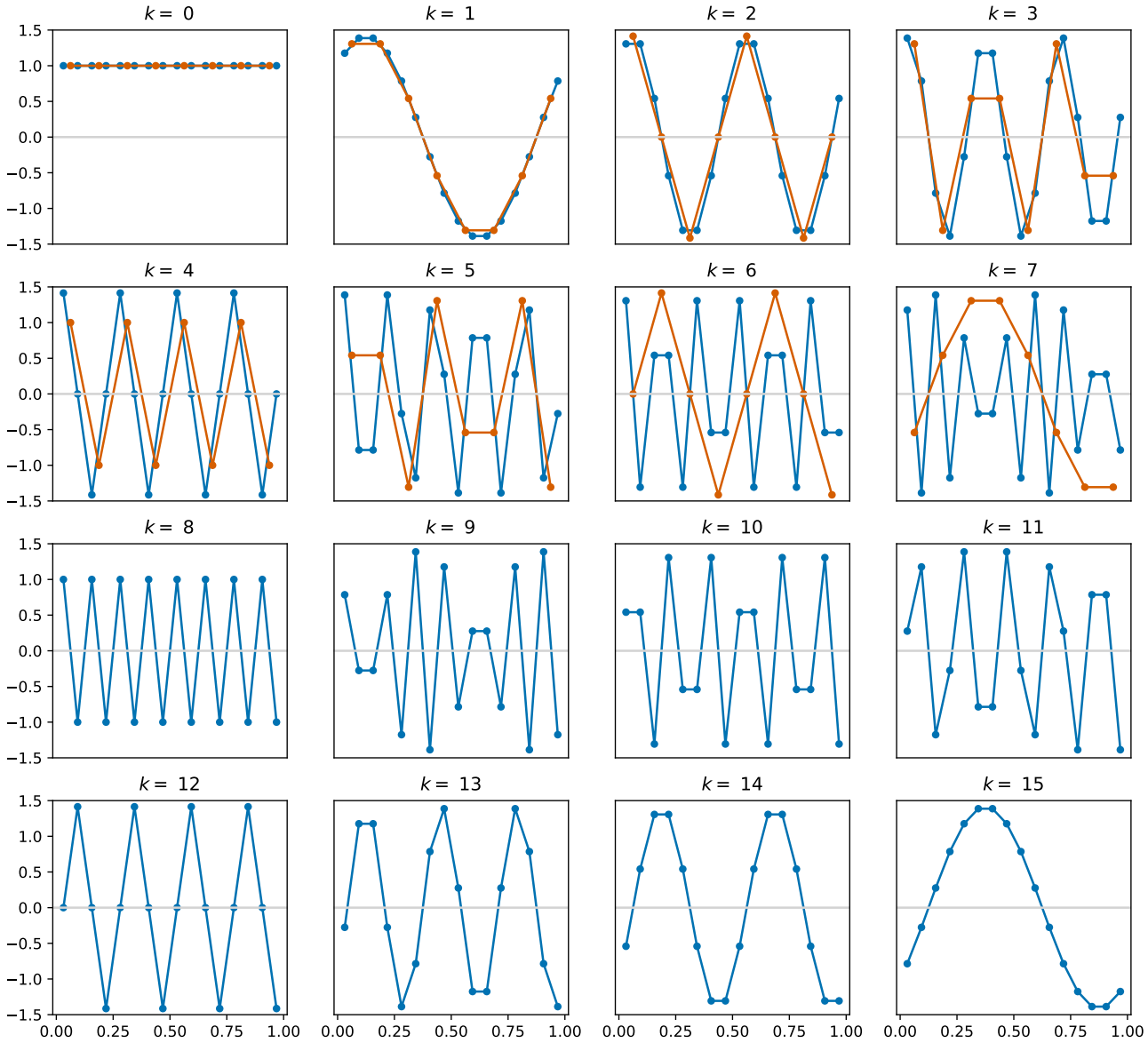


Figure 1: The vectors $\mathbf{h}_k^0$ (orange) of a coarse-grid Hartley basis on level $\ell = 0$, with $n_0 = 8$, and the vectors $\mathbf{h}_k^1$ (blue) of a fine-grid Hartley basis on level $\ell = 1$, with $n_1 = 16$. For $\ell \in \{0, 1\}$ and for each vector $k = 0, \ldots, n_\ell - 1$, the entries of the vector $[\mathbf{h}_k^\ell]_j$ are plotted against $x_j := (2j+1)/(2n_\ell) \in (0, 1)$, for $j = 0, \ldots, n_\ell - 1$.

We now examine the effects of the inter-level grid transfer operators between two successive levels on Hartley basis vectors. Such effects have been studied extensively for multigrid methods [4, 5, 34] using different bases. We succinctly present here results for the specific, cell-centered Hartley basis eq. (23). For the sake of using lighter notations, we denote by $(0, 1)$ the considered pair of successive levels, but the analysis and its conclusions naturally hold for any pair $(\ell, \ell+1)$. With the prolongation

operator $\mathbf{P} := \mathbf{P}_0^1$ defined in eq. (22), we have (see appendix B)

$$\mathbf{P}\mathbf{h}_k^0 = c_k \mathbf{h}_k^1 - c_{n_0+k}\mathbf{h}_{n_0+k}^1, \quad \forall k = 0, \dots, n_0 - 1, \tag{24}$$

where the coefficients $c_k := \cos(k\pi/n_1) = \cos(k\pi/(2n_0))$ are strictly decreasing with $k = 0, \dots, 2n_0 - 1$, i.e.,

$$1 = c_0 > c_1 > \cdots > (c_{n_0} = 0) > \cdots > c_{2n_0-1} > -1. \tag{25}$$

Prolongating a coarse-grid basis vector $\mathbf{h}_k^0$, with $k = 0, \dots, n_0 - 1$, produces a (fine-grid) vector consisting of a linear combination of two fine-grid basis vectors $\mathbf{h}_k^1$ and $\mathbf{h}_{n_0+k}^1$. Figure 1 shows that, for $k \leq n_0/2$, the fine-grid signal $\mathbf{h}_k^1$ has the same frequency as the original, coarse-grid signal $\mathbf{h}_k^0$, while $\mathbf{h}_{n_0+k}^1$ has higher frequency. Conversely, for $k > n_0/2$, the fine-grid signal $\mathbf{h}_{n_0+k}^1$ has the same frequency as the coarse-grid signal $\mathbf{h}_k^0$, while $\mathbf{h}_k^1$ has higher frequency. In both cases, the prolongation of a coarse-grid signal introduces spurious high-frequency (i.e., small-scale) components to the prolongated signal. Fortunately, both $\mathbf{h}_k^1$ and $\mathbf{h}_{n_0+k}^1$ are damped by a factor which is closer to zero for the spurious, high-frequency signals than for the consistent, low-frequency signals. Specifically, c_k tends to 1 and c_{n_0+k} tends to 0 as k tends to 0, thus damping more severely the spurious, high-frequency signals $\mathbf{h}_{n_0+k}^1$ than the consistent, low-frequency signals $\mathbf{h}_k^1$. Conversely, c_k tends to 0 and c_{n_0+k} tends to 1 as k tends to n_0, thus damping more severely the spurious, high-frequency signals $\mathbf{h}_k^1$ than the consistent, low-frequency signals $\mathbf{h}_{n_0+k}^1$.

For the restriction operator $\mathbf{R} := \mathbf{R}_1^0$ defined in eq. (22), we have (see appendix C)

$$\mathbf{R}\mathbf{h}_k^1 = 2c_k \mathbf{h}_k^0, \quad \text{and } \mathbf{R}\mathbf{h}_{n_0+k}^1 = -2c_{n_0+k}\mathbf{h}_k^0, \quad \forall k = 0, \dots, n_0 - 1. \tag{26}$$

Restricting a fine-grid basis vector $\mathbf{h}_k^1$, with $k = 0, \dots, n_0 - 1$, produces a (coarse-grid) vector proportional to the corresponding coarse-grid basis vector $\mathbf{h}_k^0$, specifically by a factor $2c_k \leq 2$. For $n_0/2 < k < n_0$, the restricted signal has lower frequency than the original, fine-grid signal, as illustrates fig. 1. Similarly, restricting a fine-grid basis vector $\mathbf{h}_k^1$, with $k = n_0, \dots, 2n_0 - 1$, produces a (coarse-grid) vector proportional to the complementary coarse-grid basis vector $\mathbf{h}_{k-n_0}^0$, specifically by a factor $-2c_k < 2$. Again, high-frequency fine-grid signals $\mathbf{h}_k^1$ corresponding to $n_0 \leq k < 3n_0/2$ are restricted to a signal with lower frequency. In conclusion, high-frequency fine-grid basis vectors that cannot be represented on the coarse grid are thus restricted to lower-frequency signals. Fortunately, such signals are the most damped, since they correspond to ranges of k where c_k is closer to 0.

Remark 1. *Denoting* $\mathbf{C} := \begin{bmatrix}\mathrm{Diag}(\{c_k\}_{k=0}^{n_0-1}) & \mathrm{Diag}(\{-c_{n_0+k}\}_{k=0}^{n_0-1})\end{bmatrix} \in \mathbb{R}^{n_0 \times 2n_0}$, *the identities in eqs.* (24) *and* (26) *may be compactly recast as*

$$\mathbf{P}\mathbf{H}_0 = \mathbf{H}_1\mathbf{C}^{\mathrm{T}}, \qquad \mathbf{R}\mathbf{H}_1 = 2\mathbf{H}_0\mathbf{C}, \qquad \mathbf{H}_1^{\mathrm{T}}\mathbf{P} = 2\mathbf{C}^{\mathrm{T}}\mathbf{H}_0^{\mathrm{T}} \qquad \mathbf{H}_0^{\mathrm{T}}\mathbf{R} = \mathbf{C}\mathbf{H}_1^{\mathrm{T}}, \tag{27}$$

where the last two identities follow from the first two by exploiting the orthogonality of $\mathbf{H}_0$ *and* $\mathbf{H}_1$.

3.4 Two-level MLMC with linear, symmetric, circulant simulators

We now assume that the operators $\tilde{\mathbf{F}}_\ell$ are symmetric, circulant matrices. Such matrices can be diagonalized in the Hartley basis [53] (see appendix D for the proof with the cell-centered basis), i.e.,

$$\tilde{\mathbf{F}}_\ell = \mathbf{H}_\ell \mathbf{\Lambda}_\ell \mathbf{H}_\ell^{\mathrm{T}}, \quad \mathbf{\Lambda}_\ell := \mathrm{Diag}(\{\lambda_k^\ell\}_{k=0}^{n_\ell-1}), \quad \text{for } \ell \in \{0, 1\}. \tag{28}$$

This property, along with the identities in eq. (27), allows $\mathbf{F}_0 := \mathbf{P}\tilde{\mathbf{F}}_0\mathbf{R} = \mathbf{P}\mathbf{H}_0\mathbf{\Lambda}_0\mathbf{H}_0^{\mathrm{T}}\mathbf{R}$ to be decomposed as $\mathbf{F}_0 = \mathbf{H}_1\mathbf{M}\mathbf{H}_1^{\mathrm{T}}$, where $\mathbf{M} := \mathbf{C}^{\mathrm{T}}\mathbf{\Lambda}_0\mathbf{C} = \begin{bmatrix}\mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22}\end{bmatrix}$, with

$$\mathbf{M}_{11} := \mathrm{Diag}(\{c_k^2 \lambda_k^0\}_{k=0}^{n_0-1}), \tag{29}$$

$$\mathbf{M}_{22} := \mathrm{Diag}(\{c_{n_0+k}^2 \lambda_k^0\}_{k=0}^{n_0-1}), \tag{30}$$

$$\mathbf{M}_{12} := \mathrm{Diag}(\{-c_k c_{n_0+k} \lambda_k^0\}_{k=0}^{n_0-1}) = \mathbf{M}_{21}. \tag{31}$$

As a consequence, eq. (19) becomes

$$\mathcal{V}(\hat{\boldsymbol{\mu}}_1^{\mathrm{MLMC}}) = \frac{1}{M_0}\|\mathbf{H}_1\mathbf{M}\mathbf{H}_1^{\mathrm{T}}\mathbf{G}^{1/2}\|_{F,\mathbf{W}_L}^2 + \frac{1}{M_1}\|\mathbf{H}_1(\boldsymbol{\Lambda}_1 - \mathbf{M})\mathbf{H}_1^{\mathrm{T}}\mathbf{G}^{1/2}\|_{F,\mathbf{W}_L}^2, \tag{32}$$

which, owing to the orthogonality of the Hartley matrix, and more particularly eq. (21), can be recast as

$$\mathcal{V}(\hat{\boldsymbol{\mu}}_1^{\mathrm{MLMC}}) = \frac{1}{M_0}\|\mathbf{M}\mathbf{H}_1^{\mathrm{T}}\mathbf{G}^{1/2}\|_{F,\mathbf{I}_{n_L}}^2 + \frac{1}{M_1}\|(\boldsymbol{\Lambda}_1 - \mathbf{M})\mathbf{H}_1^{\mathrm{T}}\mathbf{G}^{1/2}\|_{F,\mathbf{I}_{n_L}}^2. \tag{33}$$

To reduce the variance of the correction term in eq. (33), the difference between $\mathbf{M}$ and $\boldsymbol{\Lambda}_1$ needs to be as small as possible. First, we note that the two off-diagonal blocks $\mathbf{M}_{12}$ and $\mathbf{M}_{21}$, which are themselves diagonal matrices, contribute to increasing this difference. On the main diagonal, i.e., in the diagonal blocks $\mathbf{M}_{11}$ and $\mathbf{M}_{22}$, scaled eigenvalues of $\tilde{\mathbf{F}}_0$ appear twice. To compare these diagonal blocks to the eigenvalues of $\tilde{\mathbf{F}}_1$ in $\boldsymbol{\Lambda}_1$, further assumptions on $\tilde{\mathbf{F}}_0$ are required. We thus introduce the Galerkin coarse-grid operator, which is an algebraic way of constructing the coarse-grid operator $\tilde{\mathbf{F}}_0$ from the fine-grid operator $\tilde{\mathbf{F}}_1$, and which is widely used in multigrid methods and their analysis [4]. Specifically, the Galerkin operator is defined as

$$\tilde{\mathbf{F}}_0 := \frac{1}{4}\mathbf{R}\tilde{\mathbf{F}}_1\mathbf{P} \in \mathbb{R}^{n_0\times n_0}. \tag{34}$$

As shown in appendix E, this operator is the optimal operator in terms of minimizing $\|\tilde{\mathbf{F}}_1 - \mathbf{P}\tilde{\mathbf{F}}_0\mathbf{R}\|_{F,\mathbf{W}_1}^2$ for $\mathbf{W}_1 = n_1^{-1}\mathbf{I}_{n_1}$ and the grid transfer operators defined by eq. (22). It follows from eqs. (27) and (34) that the Galerkin operator $\tilde{\mathbf{F}}_0$ can be diagonalized in the Hartley basis as $\tilde{\mathbf{F}}_0 = \mathbf{H}_0\boldsymbol{\Lambda}_0\mathbf{H}_0^{\mathrm{T}}$ with $\boldsymbol{\Lambda}_0 = \mathbf{C}\boldsymbol{\Lambda}_1\mathbf{C}^{\mathrm{T}}$. In other words, the eigenvalues $\{\lambda_k^0\}_{k=0}^{n_0-1}$ of the Galerkin operator $\tilde{\mathbf{F}}_0$ can be expressed from the eigenvalues $\{\lambda_k^1\}_{k=0}^{n_1-1}$ of $\tilde{\mathbf{F}}_1$,

$$\lambda_k^0 = c_k^2\lambda_k^1 + c_{n_0+k}^2\lambda_{n_0+k}^1, \quad \forall k = 0,\dots,n_0-1. \tag{35}$$

The resulting blocks of $\mathbf{M}$ then read

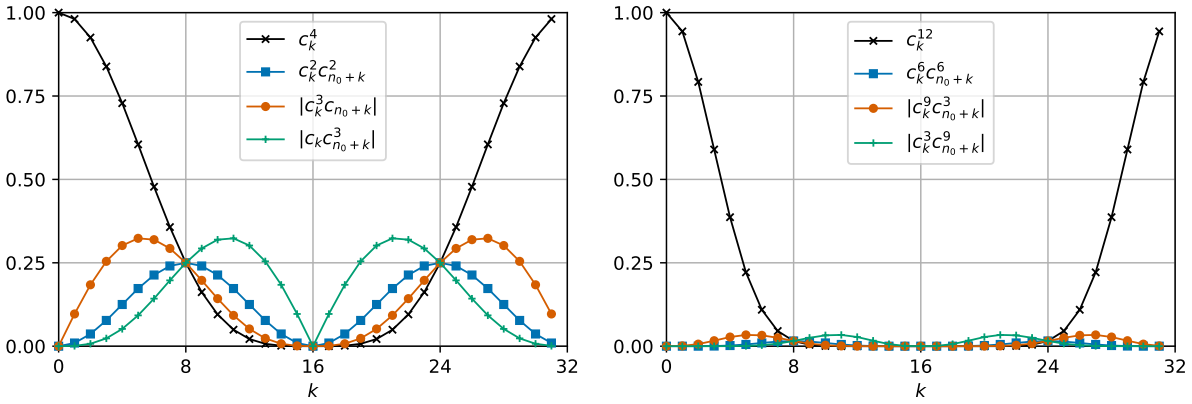
$$\mathbf{M}_{11} = \mathrm{Diag}(\{c_k^4\lambda_k^1 + c_k^2c_{n_0+k}^2\lambda_{n_0+k}^1\}_{k=0}^{n_0-1}), \tag{36}$$

$$\mathbf{M}_{22} = \mathrm{Diag}(\{c_{n_0+k}^4\lambda_{n_0+k}^1 + c_k^2c_{n_0+k}^2\lambda_k^1\}_{k=0}^{n_0-1}) = \mathrm{Diag}(\{c_k^4\lambda_k^1 + c_{k-n_0}^2c_k^2\lambda_{k-n_0}^1\}_{k=n_0}^{2n_0-1}), \tag{37}$$

$$\mathbf{M}_{12} = \mathrm{Diag}(\{-c_k^3c_{n_0+k}\lambda_k^1 - c_kc_{n_0+k}^3\lambda_{n_0+k}^1\}_{k=0}^{n_0-1}) = \mathbf{M}_{21}. \tag{38}$$

Using elementary trigonometric identities, we remark that $c_{k-n_0} = c_{n_0+k-2n_0} = -c_{n_0+k}$. Therefore, the main diagonal of $\mathbf{M}$ deviates from $\boldsymbol{\Lambda}_1$ by a multiplicative damping factor $c_k^4 \le 1$ for $k = 0,\dots,2n_0-1$, on the one hand, and by the addition of a spurious, complementary eigenvalue, though also damped by $c_k^2c_{n_0+k}^2 < 1$ for $k = 0,\dots,2n_0-1$. The off-diagonal blocks $\mathbf{M}_{12}$ and $\mathbf{M}_{21}$ introduce spurious terms that contribute to the difference $\boldsymbol{\Lambda}_1 - \mathbf{M}$ and thus increase the variance of the 2-level estimator. Note that, because these terms are to be compared with 0, the comparison with the eigenvalues of the fine-grid operator $\tilde{\mathbf{F}}_1$ is of little interest, which is why we simply consider the damping factors $-c_kc_{n_0+k}$ with respect to eigenvalues λ_k^0 of the coarse-grid operator $\tilde{\mathbf{F}}_0$, for $k = 0,\dots,n_0-1$, given by eq. (31).

The evolution with $k = 0,\dots,n_1-1$ of the four damping factors is presented in fig. 2(a). We observe that the eigenvalues λ_k^1 associated with low-frequency vectors of the fine Hartley basis, i.e., for k close to 0 and n_1-1 (see fig. 1), are well-represented on the main diagonal of $\mathbf{M}$, since $c_k^4 \approx 1$. At the same time, the damping factors $c_k^2c_{n_0+k}^2$ corresponding to spurious components are close to 0, resulting in small values of the first and last few diagonal entries of the difference $\boldsymbol{\Lambda}_1 - \mathbf{M}$. In particular, we note that the first entry of $\mathbf{M}_{11}$ exactly matches that of $\boldsymbol{\Lambda}_1$, i.e., λ_0^1, because $c_0 = 1$ and $c_{n_0} = 0$. On the other hand, eigenvalues λ_k^1 associated with medium- to high-frequency vectors of the fine Hartley basis are severely damped since c_k^4 quickly decreases to 0 as k approaches $n_0 = n_1/2$. Spurious diagonal components are also somewhat damped by a factor $c_k^2c_{n_0+k}^2$, which is maximal for $k \in \{n_1/4, 3n_1/4\}$. Finally, the spurious, off-diagonal components are damped by factors $-c_k^3c_{n_0+k}$ and $-c_kc_{n_0+k}^3$, whose magnitudes are maximal for $k \in \{n_1/8+1, 7n_1/8-1\}$ and $k \in \{3n_1/8-1, 5n_1/8+1\}$.



(a) Damping factors in $\mathbf{M}$.

(b) Damping factors in $\bar{\mathbf{M}}$.

Figure 2: Damping factors of the eigenvalues in $\mathbf{M}$ and $\bar{\mathbf{M}}$ as functions of $k = 0, \ldots, n_1 - 1$ with $n_1 = 2n_0 = 32$, when using the Galerkin operator eq. (34). The black curves correspond to the factors of the correct eigenvalues on the main diagonal, while blue curves represent the factors of the spurious eigenvalues. The orange and green curves correspond to the factors of the off-diagonal blocks.

4 The filtered MLMC method

The spectral variance of the MLMC estimator for the 1D variance estimation problem shows a significant deterioration of the variance in the high frequencies. Inspired by multigrid methods [4, 5, 34], we propose an improvement of the MLMC estimator for discretized random fields by adding pre- and post-filtering (or, in multigrid terminology, smoothing). The objective is to filter out the smaller scales, which cannot be represented on the coarse grids, before using a restriction operator and after using a prolongation operator.

4.1 Filtered grid transfer operators

Filtering the small-scale components out of a signal defined on level ℓ is achieved using a low-pass filtering operator $S_\ell \colon \mathbb{R}^{n_\ell} \to \mathbb{R}^{n_\ell}$. In what follows, we resort to linear filtering operators S_ℓ, which can thus be identified with matrices $\mathbf{S}_\ell$. Filtered grid transfer operators $\bar{P}_\ell^L$ and $\bar{R}_L^\ell$ are defined through their corresponding matrices $\bar{\mathbf{P}}_\ell^L$ and $\bar{\mathbf{R}}_L^\ell$ by

$$\bar{\mathbf{P}}_\ell^L := \bar{\mathbf{P}}_{L-1}^L \cdots \bar{\mathbf{P}}_{\ell+1}^{\ell+2} \bar{\mathbf{P}}_\ell^{\ell+1} \text{ and } \bar{\mathbf{R}}_L^\ell := \bar{\mathbf{R}}_{\ell+1}^\ell \cdots \bar{\mathbf{R}}_{L-1}^{L-2} \bar{\mathbf{R}}_L^{L-1}, \quad \forall \ell = 0, \ldots, L-1 \tag{39}$$

where

$$\bar{\mathbf{R}}_\ell^{\ell-1} := \mathbf{R}_\ell^{\ell-1} \mathbf{S}_\ell, \quad \text{and} \quad \bar{\mathbf{P}}_{\ell-1}^\ell := \mathbf{S}_\ell \mathbf{P}_{\ell-1}^\ell, \quad \forall \ell = 1, \ldots, L, \tag{40}$$

and $\bar{\mathbf{R}}_L^L = \bar{\mathbf{P}}_L^L = \mathbf{I}_{n_L}$. In general, these filtered operators no longer satisfy the norm-preserving property of eq. (13). However, the restriction operator remains the transpose of the prolongation operator, provided that the associated filtering operator is symmetric. The filtered MLMC estimator, hereafter referred to as the F-MLMC estimator, then reads

$$\hat{\boldsymbol{\mu}}_L^{\text{F-MLMC}} = \frac{1}{M_0} \sum_{i=1}^{M_0} \bar{f}_0(\mathbf{X}_L^{(0,i)}) + \sum_{\ell=1}^{L} \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} [\bar{f}_\ell(\mathbf{X}_L^{(\ell,i)}) - \bar{f}_{\ell-1}(\mathbf{X}_L^{(\ell,i)})], \tag{41}$$

where $\bar{f}_\ell = \bar{P}_\ell^L \circ \tilde{f}_\ell \circ \bar{R}_L^\ell$, for $\ell = 0, \ldots, L$.

4.2 Two-level F-MLMC with linear, symmetric, circulant simulators

We turn again to the spectral analysis in the 1D setting described in section 3.2 and consider the second-order Shapiro filter [54–56] defined as

$$\mathbf{S}_\ell := \frac{1}{4}\begin{bmatrix} 2 & 1 & & & 1 \\ 1 & 2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 2 & 1 \\ 1 & & & 1 & 2 \end{bmatrix} \in \mathbb{R}^{n_\ell \times n_\ell}. \tag{42}$$

With the specific grid transfer operators defined in eq. (22), the operator $\bar{\mathbf{P}}_{\ell-1}^{\ell}$ corresponds to the linear interpolation operator between the levels $\ell-1$ and ℓ. We again restrict ourselves to a two-level analysis and define the filtered grid transfer operators as $\bar{\mathbf{P}} = \mathbf{S}_1\mathbf{P}$ and $\bar{\mathbf{R}} = \mathbf{R}\mathbf{S}_1$, where $\mathbf{P} := \mathbf{P}_0^1$ and $\mathbf{R} := \mathbf{R}_1^0$ as in section 3, and where $\mathbf{S}_1$ denotes the second-order Shapiro filter defined in eq. (42). Similarly to eqs. (24) and (26), it is possible to study the effect the filtered transfer operators on the Hartley basis vectors (see appendix F). For the prolongation $\bar{\mathbf{P}}$, we have

$$\bar{\mathbf{P}}\mathbf{h}_k^0 = c_k^3\mathbf{h}_k^1 - c_{n_0+k}^3\mathbf{h}_{n_0+k}^1, \quad \forall k = 0, \dots, n_0 - 1. \tag{43}$$

The addition of the Shapiro filter raises the damping factors c_k to the power 3. The prolongated Hartley basis vectors are thus more severely damped than they were without filtering. Again, the most damped fine-grid basis vectors $\mathbf{h}_k^1$ are those corresponding to k close to $n_0 = n_1/2$, i.e., high-frequency signals. For the restriction operator we have

$$\bar{\mathbf{R}}\mathbf{h}_k^1 = 2c_k^3\mathbf{h}_k^0, \quad \text{and } \bar{\mathbf{R}}\mathbf{h}_{n_0+k}^1 = -2c_{n_0+k}^3\mathbf{h}_k^0, \quad \forall k = 0, \dots, n_0 - 1. \tag{44}$$

Similar conclusions can be drawn as for the unfiltered case, but with the damping factors raised to the power of 3, thus increasing their effect, which still affects more strongly the fine-grid, high-frequency signals that cannot be represented on the coarse grid. Similarly to the unfiltered case (see remark 1), identities eqs. (43) and (44) can be recast as

$$\bar{\mathbf{P}}\mathbf{H}_0 = \mathbf{H}_1\mathbf{C}_3^{\mathrm{T}}, \qquad \bar{\mathbf{R}}\mathbf{H}_1 = 2\mathbf{H}_0\mathbf{C}_3, \qquad \mathbf{H}_1^{\mathrm{T}}\bar{\mathbf{P}} = 2\mathbf{C}_3^{\mathrm{T}}\mathbf{H}_0^{\mathrm{T}} \qquad \mathbf{H}_0^{\mathrm{T}}\bar{\mathbf{R}} = \mathbf{C}_3\mathbf{H}_1^{\mathrm{T}}, \tag{45}$$

where $\mathbf{C}_3 := \begin{bmatrix}\mathrm{Diag}(\{c_k^3\}_{k=0}^{n_0-1}) & \mathrm{Diag}(\{-c_{n_0+k}^3\}_{k=0}^{n_0-1})\end{bmatrix} \in \mathbb{R}^{n_0 \times 2n_0}$.

The impact of filtering on the total variance of the MLMC estimator is now assessed, considering a 2-level MLMC estimator and assuming that $\tilde{\mathbf{F}}_1$ and $\tilde{\mathbf{F}}_0$ are symmetric, circulant matrices. From eqs. (28) and (45), we deduce the decomposition $\bar{\mathbf{F}}_0 := \bar{\mathbf{P}}\tilde{\mathbf{F}}_0\bar{\mathbf{R}} = \mathbf{H}_1\bar{\mathbf{M}}\mathbf{H}_1^{\mathrm{T}}$, where $\bar{\mathbf{M}} := \begin{bmatrix}\bar{\mathbf{M}}_{11} & \bar{\mathbf{M}}_{12} \\ \bar{\mathbf{M}}_{21} & \bar{\mathbf{M}}_{22}\end{bmatrix}$, with

$$\bar{\mathbf{M}}_{11} = \mathrm{Diag}(\{c_k^6\lambda_k^0\}_{k=0}^{n_0-1}), \tag{46}$$

$$\bar{\mathbf{M}}_{22} = \mathrm{Diag}(\{c_{n_0+k}^6\lambda_k^0\}_{k=0}^{n_0-1}), \tag{47}$$

$$\bar{\mathbf{M}}_{12} = \mathrm{Diag}(\{-c_k^3c_{n_0+k}^3\lambda_k^0\}_{k=0}^{n_0-1}) = \bar{\mathbf{M}}_{21}. \tag{48}$$

Upon replacing $\mathbf{F}_0$ with its filtered counterpart $\bar{\mathbf{F}}_0$, eq. (19) can be written as in eq. (33), but with $\mathbf{M}$ replaced by $\bar{\mathbf{M}}$. The sparsity pattern of $\bar{\mathbf{M}}$ is identical to that of $\mathbf{M}$, and its entries are similar, but with damping factors raised to increased powers. We remark that the off-diagonal blocks $\bar{\mathbf{M}}_{12}$ and $\bar{\mathbf{M}}_{21}$ have entries that are more strongly damped than those of their unfiltered counterparts, $\mathbf{M}_{12}$ and $\mathbf{M}_{21}$, as can be visualized in fig. 2 (orange plots), which contributes to reducing the off-diagonal entries of $\mathbf{\Lambda}_1 - \bar{\mathbf{M}}$ compared to those of $\mathbf{\Lambda}_1 - \mathbf{M}$.

To study more closely the diagonal entries of $\bar{\mathbf{M}}$ and compare them to the eigenvalues of $\tilde{\mathbf{F}}_1$ in $\mathbf{\Lambda}_1$ we resort to the Galerkin operator defined by $\tilde{\mathbf{F}}_0 := \frac{1}{4}\bar{\mathbf{R}}\tilde{\mathbf{F}}_1\bar{\mathbf{P}}$. This definition is inspired by the form of eq. (34), although there is no guarantee of its optimality. Then, from eqs. (28) and (45), it follows that the Galerkin operator $\tilde{\mathbf{F}}_0$ can be diagonalized in the Hartley basis as $\tilde{\mathbf{F}}_0 = \mathbf{H}_0\mathbf{\Lambda}_0\mathbf{H}_0^{\mathrm{T}}$ with $\mathbf{\Lambda}_0 = \mathbf{C}_3\mathbf{\Lambda}_1\mathbf{C}_3^{\mathrm{T}}$, or, equivalently,

$$\lambda_k^0 = c_k^6\lambda_k^1 + c_{n_0+k}^6\lambda_{n_0+k}^1, \quad \forall k = 0, \dots, n_0 - 1. \tag{49}$$

The diagonal blocks of $\bar{\mathbf{M}}$ thus become

$$\bar{\mathbf{M}}_{11} = \mathrm{Diag}(\{c_k^{12}\lambda_k^1 + c_{n_0+k}^6 c_k^6 \lambda_{n_0+k}^1\}_{k=0}^{n_0-1}), \tag{50}$$

$$\bar{\mathbf{M}}_{22} = \mathrm{Diag}(\{c_{n_0+k}^{12}\lambda_{n_0+k}^1 + c_{n_0+k}^6 c_k^6 \lambda_k^1\}_{k=0}^{n_0-1}) = \mathrm{Diag}(\{c_k^{12}\lambda_k^1 + c_{k-n_0}^6 c_k^6 \lambda_{k-n_0}^1\}_{k=n_0}^{2n_0-1}), \tag{51}$$

$$\bar{\mathbf{M}}_{12} = \mathrm{Diag}(\{-c_k^9 c_{n_0+k}^3\lambda_k^1 - c_k^3 c_{n_0+k}^9 \lambda_{n_0+k}^1\}_{k=0}^{n_0-1}) = \bar{\mathbf{M}}_{21}. \tag{52}$$

The conclusions are the same as for the diagonal blocks of $\mathbf{M}$, but with damping factors raised to higher powers. Consequently, the spurious eigenvalues on the main diagonal are damped more strongly than in the unfiltered case, as can be visualized in fig. 2 (blue plots). Moreover, the factors c_k^{12} also damp more strongly the consistent eigenvalues compared to the factors c_k^4 of $\mathbf{M}$, as can be visualized in fig. 2 (black plots). Although the addition of filters induces a certain loss of (mainly high-frequency) information, the spurious signals introduced by the grid transfer operators are significantly reduced. These results suggest that, for filtering to be beneficial, a tradeoff needs to be found between reducing the detrimental effects induced by the grid transfer operators and degrading of the information of the original signal. The second-order Shapiro filters considered in this paper seem to offer a good compromise. Further endeavors to study and improve the filtering process may be pursued in future work.

5 Numerical experiments

As discussed in the introduction, one area where the estimation of the expectation of discretized random fields arises is covariance modelling, specifically when estimating the intrinsic variances of a diffusion-based covariance operator using a randomization method [23, 28]. This particular problem, which is briefly outlined below, will be the focus of our numerical experiments.

5.1 Problem description

Let $u\colon \mathcal{D} \to \mathbb{R}$ and $b\colon \mathcal{D} \to \mathbb{R}$ be square-integrable functions on the domain $\mathcal{D} \subset \mathbb{R}^d$ where $d \in \{1, 2, 3\}$ is the spatial dimension. We consider numerical solutions of the following elliptic equation, subject to application-dependent boundary conditions (BCs):

$$(I - \nabla \cdot \mathbf{K}\nabla)^m u = b, \tag{53}$$

where m is a positive integer, I is the identity operator, and $\mathbf{K}\colon \mathcal{D} \to \mathbb{R}^{d\times d}$ is a symmetric, positive-definite (SPD) tensor field with entries $[K_{ij}]_{i,j=1,\dots,d}$. Equation (53) can be interpreted as a semi-discrete representation of a diffusion equation integrated over m time-steps where the temporal derivative is discretized with a backward Euler (implicit) scheme, the time-step is equal to unity, $\mathbf{K}$ is a diffusivity tensor, and the initial condition is b [57]. If $\mathbf{K}$ is constant then the integral solution on $\mathbb{R}^d$ defines a covariance operator whose kernel is a covariance function from the Matérn class [58, 59].

We assume that the operator in eq. (53) is discretized in space on a (not necessarily structured) grid of n cells. We can then deduce the covariance matrix associated with the numerical solution of eq. (53) as

$$\mathbf{L} := (\mathbf{I} - \mathbf{\Delta})^{-m}\mathbf{W}^{-1}, \tag{54}$$

where $\mathbf{\Delta}$ is the matrix representing a spatial discretization of the differential operator $\nabla \cdot \mathbf{K}\nabla$, and $\mathbf{W} \in \mathbb{R}^{n\times n}$ is an SPD Gram matrix that encodes the geometrical and structural information related to the discrete approximation of the diffusive term on the grid. Specifically, $\mathbf{W}$ is such that $\mathbf{\Delta}$ is self-adjoint with respect to $\langle\cdot,\cdot\rangle_{\mathbf{W}}$, i.e., $\mathbf{W}\mathbf{\Delta} = \mathbf{\Delta}^{\mathrm{T}}\mathbf{W}$. Consequently, the matrix $\mathbf{L}$ is self-adjoint (symmetric) with respect to the canonical inner product. In the experiments, we consider only a diagonal diffusivity tensor $K_{ij} = K_{ij}\delta_{ij}$ where δ_{ij} is the Kronecker delta. Specifically, we define the diagonal elements according to the relation $K_{ii}(\boldsymbol{x}) = (2m-d-2)^{-1}(D_{ii}(\boldsymbol{x}))^2$ where the elements $[D_{ii}(\boldsymbol{x})]_{i=1,\dots,d}$ correspond to the directional correlation length-scales at the spatial location $\boldsymbol{x}$, and $m > d/2+1$ [57, section 3].

The matrix $\mathbf{L}$ is SPD but does not define a covariance matrix with meaningful variances for applications like data assimilation. As such, $\mathbf{L}$ must be normalized by its diagonal so that the desired variances can be applied. Thus, we define the covariance matrix of interest as $\mathbf{B} = \mathbf{\Sigma\Gamma L\Gamma\Sigma}$, where $\mathbf{\Gamma} = \mathrm{Diag}(\mathrm{diag}(\mathbf{L}))^{-1/2}$ is a normalizing diagonal matrix such that $\mathrm{Diag}(\mathrm{diag}(\mathbf{\Gamma L\Gamma})) = \mathbf{I}_n$, and $\mathbf{\Sigma}^2 = \mathbf{\Sigma\Sigma}$ is the diagonal matrix with entries corresponding to the desired variances, i.e., $\mathrm{diag}(\mathbf{B}) = \mathrm{diag}(\mathbf{\Sigma}^2) = (\sigma_1^2,\dots,\sigma_n^2)$. In these expressions, the operator $\mathrm{diag}(\cdot)$ maps a matrix to the vector consisting of the diagonal elements of that matrix, while the operator $\mathrm{Diag}(\cdot)$ maps a vector to the diagonal matrix whose diagonal consists of the entries of that vector. For large-scale problems, the matrix $\mathbf{L}$ is not assembled, and only applications of $\mathbf{L}$ to vectors are accessible. Thus, its diagonal entries $\mathrm{diag}(\mathbf{L}) = \boldsymbol{\theta} = (\theta_1,\dots,\theta_n)$ are not explicitly stored and need to be determined differently [28]. A direct way would be to recover these by applying $\mathbf{L}$ to the canonical basis vectors of $\mathbb{R}^n$, i.e., $\theta_k = (\mathbf{L}\mathbf{e}_k)_k$, for $k = 1,\dots,n$. For large n, this approach is not computationally tractable.

An alternative strategy is to approximate $\boldsymbol{\theta}$ by randomization. Taking $m = 2q$, $\mathbf{L}$ can be subsequently factored (see appendix G) as $\mathbf{L} = \mathbf{A}\mathbf{W}\mathbf{A}^{\mathrm{T}}$, where

$$\mathbf{A} := (\mathbf{I} - \mathbf{\Delta})^{-q}\mathbf{W}^{-1}, \tag{55}$$

which, as for $\mathbf{L}$ itself, cannot be explicitly assembled in large-scale applications. This decomposition of $\mathbf{L}$ implies that, for any random vector $\mathbf{X}$ with $\mathbb{E}[\mathbf{X}] = \mathbf{0}_n$ and $\mathbb{E}[\mathbf{X}\mathbf{X}^{\mathrm{T}}] = \mathbf{W}$,

$$\boldsymbol{\theta} = \mathrm{diag}(\mathbf{L}) = \mathrm{diag}(\mathbf{A}\mathbb{E}[\mathbf{X}\mathbf{X}^{\mathrm{T}}]\mathbf{A}^{\mathrm{T}}) = \mathrm{diag}(\mathbb{C}[\mathbf{A}\mathbf{X}]) = \mathbb{V}[\mathbf{A}\mathbf{X}] = \mathbb{E}[\mathbf{A}\mathbf{X}\odot\mathbf{A}\mathbf{X}], \tag{56}$$

where $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$ denote the element-wise expectation and variance of a random vector, $\mathbb{C}[\cdot]$ denotes the covariance matrix of a random vector, and where $\odot$ denotes the Schur product (a.k.a. the Hadamard or element-wise product) between two vectors or matrices of the same size. One way to construct such a vector $\mathbf{X}$ is to define $\mathbf{X} = \mathbf{V}\mathbf{Z}$, where $\mathbf{V}$ arises from the factorization $\mathbf{W} = \mathbf{V}\mathbf{V}^{\mathrm{T}}$, and where $\mathbf{Z}$ is a random vector with $\mathbb{E}[\mathbf{Z}] = \mathbf{0}_n$ and $\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}] = \mathbf{I}_n$. In the rest of this paper, we consider normal random vectors $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{W})$, which can be constructed by $\mathbf{X} = \mathbf{V}\mathbf{Z}$, with $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I}_n)$. The expectation $\boldsymbol{\theta}$ of the $\mathbb{R}^n$-valued random vector $\mathbf{Y} := \mathbf{A}\mathbf{X}\odot\mathbf{A}\mathbf{X}$ can then be estimated using MC sampling. Given a random M-sample $\{\mathbf{X}^{(i)}\}_{i=0}^{M}$ of $\mathbf{X}$, an unbiased estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is the sample mean

$$\hat{\boldsymbol{\theta}} = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{A}\mathbf{X}^{(i)})\odot(\mathbf{A}\mathbf{X}^{(i)}). \tag{57}$$

We remark that the estimator $\hat{\boldsymbol{\theta}}$ only requires M applications of $\mathbf{A}$ to a vector, typically with $M \ll n$ in large-scale applications. Furthermore, by construction, the MC estimator defined by eq. (57) yields non-negative estimates, which is a fundamental requirement for the problem under consideration. Alternative MC estimators have been proposed in the literature for estimating the diagonal of a general (not necessarily symmetric) matrix [60, 61], which do not employ a factored form of the matrix and hence do not guarantee non-negative estimates. Here, we investigate the use of the MLMC and F-MLMC estimators described in sections 2 and 4 to improve (in terms of RMSE) the estimation of

$\boldsymbol{\theta}$ and hence the efficiency of determining accurate normalization coefficients $\boldsymbol{\Gamma}$. It should be noted that the proposed (F)-MLMC construction does not guarantee non-negative estimates either, although negative estimates were never encountered in our experiments.

In practice, the matrix-vector product $\mathbf{Ax}$ can be computed by solving the sequence of SPD systems of linear equations $(\mathbf{W}(\mathbf{I}-\boldsymbol{\Delta})\mathbf{y}^{(j)} = \mathbf{z}^{(j)})_{j=1}^{q}$, with $\mathbf{z}^{(1)} := \mathbf{x}$ and $\mathbf{z}^{(j)} := \mathbf{W}\mathbf{y}^{(j-1)}$ for $j = 2, \dots, q$, so that $\mathbf{y}^{(q)} = \mathbf{Ax}$. In our experiments, the numerical solving of these systems is achieved by precomputing a Cholesky decomposition of $\mathbf{W}(\mathbf{I}-\boldsymbol{\Delta})$. However, for large-scale problems, it is not reasonably possible to compute or store such a decomposition, and sparse iterative methods are typically used instead [62].

5.2 1D illustration

Before turning to the (F-)MLMC estimation of $\boldsymbol{\theta}$, we briefly consider a simpler experiment to verify numerically the conclusions of the spectral analysis of section 3. Specifically, we consider a hierarchy of simulators $(f_\ell)_{\ell=0}^{L}$ defined through eq. (8) by $\tilde{f}_\ell \colon \mathbf{x}_\ell \mapsto \mathbf{A}_\ell \mathbf{x}_\ell$, where $\mathbf{A}_\ell$ is the factor in eq. (55) arising from a cell-centered discretization on the 1D domain $\mathcal{D} := [0, 1]$, with periodic boundary conditions. Specifically, we consider the same setting and grid transfer operators as in section 3.2, and we define hierarchies of different depths (i.e., numbers of levels), corresponding to $L \in \{0, \dots, 5\}$, with a fixed finest discretization corresponding to $n_L = 512$. The diffusion tensor field $\mathbf{K}$, which reduces here to a scalar field, is taken to be constant by setting $D_{11}(\boldsymbol{x}) = D \in \mathbb{R}$ for all $\boldsymbol{x} \in \mathcal{D}$. The scalar value D will be referred to as the length-scale. Furthermore, we choose a fixed value of $m = 2q = 10$. We note that in this setting, the matrices $\mathbf{A}_\ell$ are symmetric and circulant [63].

The statistical parameter of interest is $\mathbb{E}[f_L(\mathbf{X}_L)] = \mathbb{E}[\mathbf{A}_L\mathbf{X}_L]$, with $\mathbf{X}_L \sim \mathcal{N}(\mathbf{0}_{n_L}, \mathbf{W}_{n_L})$. It trivially follows that $\mathbb{E}[f_L(\mathbf{X}_L)] = \mathbf{0}_{n_L}$. The MLMC estimator $\hat{\boldsymbol{\mu}}_L^{\mathrm{MLMC}}$ is defined as in eqs. (11) and (12), where the transfer operators are defined as in eq. (22). With this choice, it follows from eqs. (9) and (13) that the MLMC estimator $\hat{\boldsymbol{\mu}}_L^{\mathrm{MLMC}}$ involves, at each correction level, input random vectors that are such that $\mathbf{R}_L^0\mathbf{X}_L \sim \mathcal{N}(\mathbf{0}_{n_0}, \mathbf{W}_0)$ and $[(\mathbf{R}_L^\ell\mathbf{X}_L)^{\mathrm{T}} \quad (\mathbf{R}_L^{\ell-1}\mathbf{X}_L)^{\mathrm{T}}]^{\mathrm{T}} \sim \mathcal{N}(\mathbf{0}_{n_\ell+n_{\ell-1}}, \tilde{\mathbf{W}}_\ell)$ for $\ell = 1, \dots, L$, with

$$\tilde{\mathbf{W}}_\ell := \begin{bmatrix} \mathbf{W}_\ell & \mathbf{W}_\ell\mathbf{P}_{\ell-1}^\ell \\ \mathbf{R}_\ell^{\ell-1}\mathbf{W}_\ell & \mathbf{W}_{\ell-1} \end{bmatrix} \in \mathbb{R}^{(n_\ell+n_{\ell-1})\times(n_\ell+n_{\ell-1})}, \tag{58}$$

which has the same structure as eq. (16). In fact, eqs. (16) and (58) are equivalent under the functional, piecewise constant representation of the input fields. The filtered estimator $\hat{\boldsymbol{\mu}}_L^{\text{F-MLMC}}$ is defined as in eq. (41), with the same transfer operators and the 2nd-order Shapiro filters defined in eq. (42). For all the numerical experiments, we rely on the optimal sample allocation that minimizes, for a given computational budget $\mathcal{C}$, the variance of the multilevel estimator [14],

$$M_\ell = \left\lfloor \frac{\mathcal{C}}{\mathcal{S}_L} \sqrt{\frac{\mathcal{V}_\ell}{\mathcal{C}_\ell + \mathcal{C}_{\ell-1}}} \right\rfloor^+, \qquad \mathcal{V}_\ell := \mathcal{V}(\bar{f}_\ell(\mathbf{X}_L) - \bar{f}_{\ell-1}(\mathbf{X}_L)), \qquad \mathcal{S}_L := \sum_{\ell=0}^{L} \sqrt{\mathcal{V}_\ell(\mathcal{C}_\ell + \mathcal{C}_{\ell-1})}, \tag{59}$$

where $\mathcal{C}_\ell$ denotes the mean computational cost of evaluating $\tilde{f}_\ell(\mathbf{X}_L)$ and $\lfloor\cdot\rfloor^+ := \max(1, \lfloor\cdot\rfloor)$, where $\lfloor\cdot\rfloor$ denotes the floor function. It should be noted that, by convention, the quantities indexed by $\ell = -1$ vanish, so that $\mathcal{C}_{-1} = 0$ and $\mathcal{V}_0 = \mathcal{V}(\bar{f}_0(\mathbf{X}_L))$. For conciseness, we have used the same notation $\bar{f}_\ell$ to refer either to the filtered version of f_ℓ (i.e., $\bar{f}_\ell$ as defined in section 4.1) when considering F-MLMC, or to its unfiltered counterpart (i.e., simply f_ℓ) when dealing with unfiltered MLMC. In what follows, we assume a linear computational cost model, specifically $\mathcal{C}_\ell = \mathcal{O}(n_\ell)$, consistent with a fixed number of sparse (banded) matrix-vector applications required for the evaluation of $\tilde{f}_\ell$ (see, e.g., [62]). The variances $\mathcal{V}_\ell$ are estimated in a preprocessing stage with a pilot sample of size 1000, and the multilevel estimation is conducted with a computational budget $\mathcal{C} = 100\mathcal{C}_L$.

In what follows, $\hat{\boldsymbol{\mu}}_L$ denotes an unbiased estimator of $\boldsymbol{\mu}_L := \mathbb{E}[f_L(\mathbf{X}_L)]$, typically, the plain MC estimator $\hat{\boldsymbol{\mu}}_L^{\mathrm{MC}}$ on the finest level L, the MLMC estimator $\hat{\boldsymbol{\mu}}_L^{\mathrm{MLMC}}$ or the F-MLMC estimator $\hat{\boldsymbol{\mu}}_L^{\text{F-MLMC}}$. We recall that the unbiasedness of $\hat{\boldsymbol{\mu}}_L$ implies that $\mathrm{MSE}(\hat{\boldsymbol{\mu}}_L, \boldsymbol{\mu}_L) = \mathcal{V}(\hat{\boldsymbol{\mu}}_L)$. In order to examine the different scales of the proposed estimators, we decompose the variance $\mathcal{V}(\hat{\boldsymbol{\mu}}_L)$ into contributions of the individual Hartley modes. Specifically, exploiting the orthogonality of $\mathbf{H}_L$, it follows from eq. (20) that

$$\mathcal{V}(\hat{\boldsymbol{\mu}}_L) = \mathbb{E}[\|\hat{\boldsymbol{\mu}}_L - \boldsymbol{\mu}_L\|_{\mathbf{W}_L}^2] = \mathbb{E}[\|\mathbf{H}_L^{\mathrm{T}}\mathbf{V}_L^{\mathrm{T}}(\hat{\boldsymbol{\mu}}_L - \boldsymbol{\mu}_L)\|_{\mathbf{W}_L}^2], \tag{60}$$

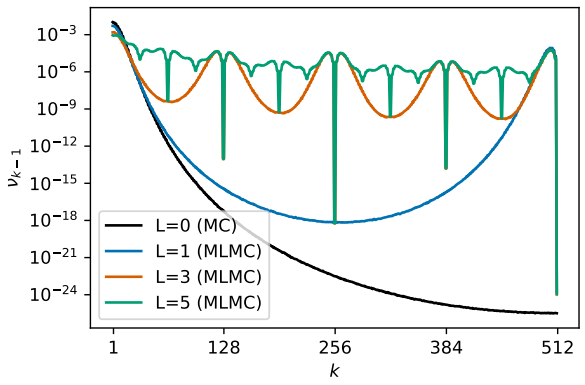
which, owing to the linearity of the expectation operator and the fact that $\mathbf{W}_L = n_L^{-1}\mathbf{I}_{n_L}$, may be recast as

$$\mathcal{V}(\hat{\boldsymbol{\mu}}_L) = \|\boldsymbol{\nu}\|_1 := \sum_{k=0}^{n_L-1} \nu_k, \qquad \nu_k := n_L^{-2}\mathbb{E}[((\mathbf{h}_k^L)^{\mathrm{T}}(\hat{\boldsymbol{\mu}}_L - \mathbb{E}[f_L(\mathbf{X}_L)]))^2] = n_L^{-2}\mathbb{V}[(\mathbf{h}_k^L)^{\mathrm{T}}\hat{\boldsymbol{\mu}}_L], \tag{61}$$

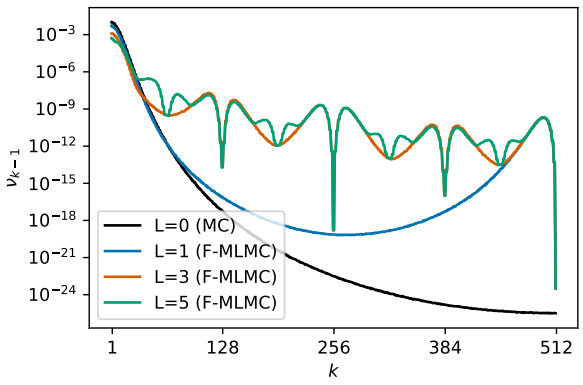
where $\boldsymbol{\nu} = (\nu_k)_{k=0}^{n_L-1}$ will be referred to as the spectral variance. Furthermore, we define the cumulative spectral variance $\boldsymbol{\nu}^{\mathrm{cml}} = (\nu_k^{\mathrm{cml}})_{k=0}^{n_L-1}$ such that $\nu_k^{\mathrm{cml}} = \sum_{k'=0}^{k} \nu_{k'}$, implying that the total variance is given by $\nu_{n_L-1}^{\mathrm{cml}} = \|\boldsymbol{\nu}\|_1$. For better visualization and interpretation, the columns of the 1D Hartley matrix $\mathbf{H}_L$ used in eq. (60) are actually reordered so that they are sorted by increasing representable frequency on the corresponding discrete grid (see fig. 1, which depicts the Hartley basis vectors without reordering). Specifically, the new matrix with reordered columns is obtained as $\mathbf{H}_L\boldsymbol{\Pi}$, where $\boldsymbol{\Pi} = (\Pi_{j,k})_{j,k=0}^{n_L-1}$ is the permutation matrix defined by $\Pi_{j,2k} = \delta_{j,k}$ and $\Pi_{j,2k+1} = \delta_{j,n_L-k-1}$, for $j = 0, \dots, n_L - 1$ and $k = 0, \dots, n_L/2 - 1$ (assuming n_L is even), and where $\delta_{i,j}$ denotes the Kronecker delta.

Figure 3 reports the spectral variance $\boldsymbol{\nu}$, as well as the corresponding cumulative variance $\boldsymbol{\nu}^{\mathrm{cml}}$, associated with the 2-, 4- and 6-level MLMC and F-MLMC estimators ($L \in \{1, 3, 5\}$) and with the single-level MC estimator ($L = 0$), for the estimation of $\mathbb{E}[f_L(\mathbf{X}_L)] = \mathbf{0}_{n_L}$, where the simulators correspond to length-scale $D = 0.06$. Figures 3(a) and 3(b) show that, for the single-level MC estimator, most of the error arises from the first Hartley modes, which are associated with the large scales, or low frequencies, of the discretized field. This is confirmed in figs. 3(c) and 3(d), where the cumulative variance of the MC estimator rapidly increases before reaching a plateau, showing that the variance is concentrated on the first few modes. For the 2-level MLMC estimator ($L = 1$), the spectral variance plotted in fig. 3(a) starts with a similar decay as that of the MC estimator in the low frequencies, before increasing again in the high frequencies. Nevertheless, the variance is still concentrated on the first few modes, and the low-frequency components of the spectral variance are lower than those of the single-level MC estimator. This translates into a lower plateau reached by the cumulative spectral variance, as shown in fig. 3(c). Moreover, the variance increase in the high-frequencies translates into a noticeable increase in the cumulative variance in the last few Hartley modes, which, in turn, results in a non-negligible increase in the total variance. As more levels are added, the spectral variance significantly deteriorates in the high-frequencies. While for the 4-level MLMC estimator ($L = 3$), this deterioration is compensated by a lower variance in the low frequencies, this is no longer the case for the 6-level MLMC estimator ($L = 5$), whose cumulative spectral variance eventually gets larger than that of the 4-level MLMC estimator, thus resulting in a larger total variance.

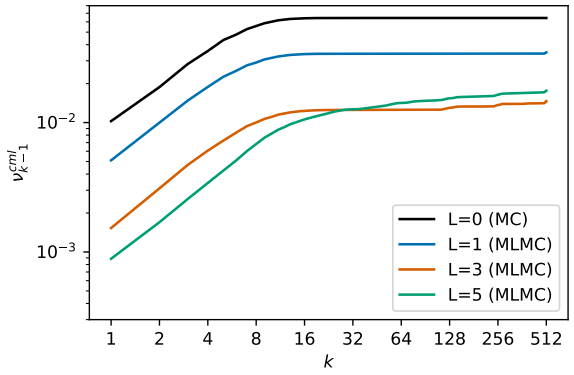
Figures 3(b) and 3(d) show the spectral variance $\boldsymbol{\nu}$ and the cumulative variance $\boldsymbol{\nu}^{\mathrm{cml}}$ of the F-MLMC estimators with different grid hierarchies, corresponding to $L \in \{1, 3, 5\}$. The effects of the filters are especially visible on the spectral variance plotted in fig. 3(b), which is significantly reduced, not only in the high frequencies, but also in the lowest ones (i.e., corresponding to the first few Hartley modes), compared to that of the unfiltered MLMC plotted in fig. 3(a). The reduced error in the high frequencies (i.e., small scales) prevents the cumulative spectral variance, plotted in fig. 3(d), from being significantly impacted in the last few Hartley modes, as opposed to that of the unfiltered MLMC,



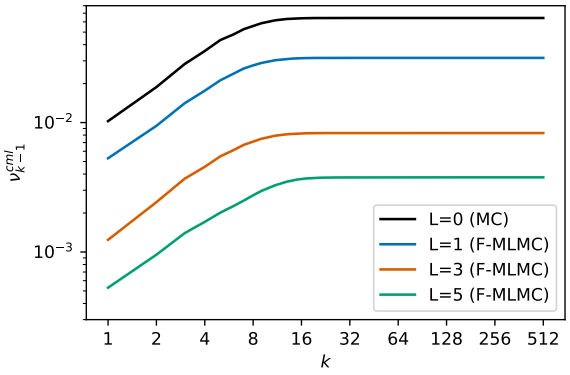
(a) Spectral variance $\boldsymbol{\nu}$ of MC and MLMC estimators.



(b) Spectral variance $\boldsymbol{\nu}$ of MC and F-MLMC estimators.



(c) Cumulative spectral variance $\boldsymbol{\nu}^{\text{cml}}$ of MC and MLMC estimators.

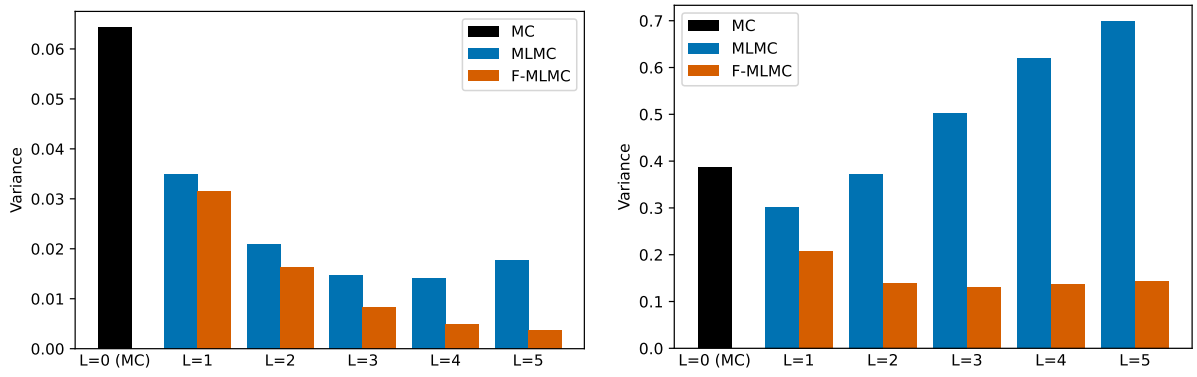


(d) Cumulative spectral variance $\boldsymbol{\nu}^{\text{cml}}$ of MC and F-MLMC estimators.

Figure 3: Spectral variance (top) and cumulative spectral variance (bottom) of the MC estimator ($L = 0$) and different MLMC (left) and F-MLMC (right) estimators ($L \in \{1, 3, 5\}$), for the illustrative 1D estimation problem described in section 5.2, with length-scale $D = 0.06$. The finest level L always corresponds to a discretization with $n_L = 512$ cells, and the total budget is set to $\mathcal{C} = 100\mathcal{C}_L$. The variance is estimated from 1000 estimators.

plotted in fig. 3(c). Furthermore, the reduced error in the low frequencies (i.e., large scales) translates into a lower plateau of the cumulative variance, hence a lower total variance, than for the unfiltered MLMC. This is well visible in fig. 4(a), which summarizes the total variance of the MC estimator and the different MLMC and F-MLMC estimators for $L \in \{1, \dots, 5\}$. Specifically, the addition of filters leads to a 95% reduction in total variance of the F-MLMC estimator compared to the single-level MC estimator, and to a 80% reduction compared to the best, 5-level unfiltered MLMC estimator.

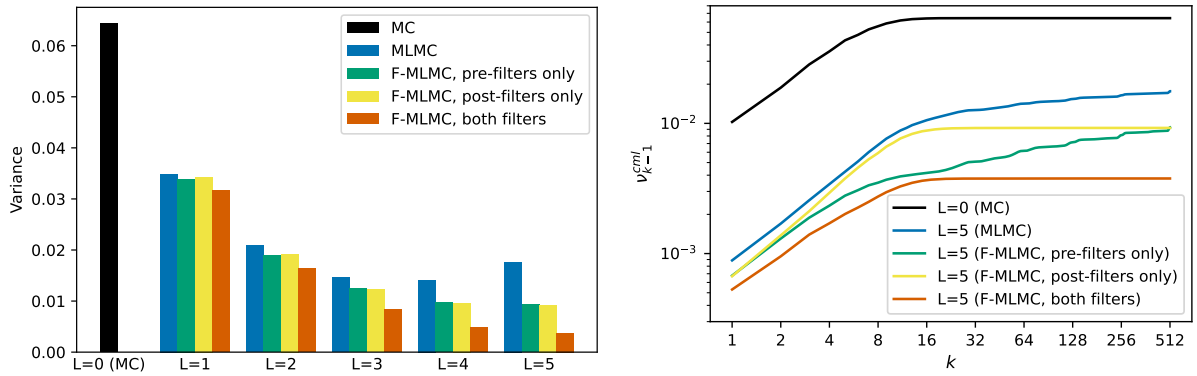
Figure 5 compares F-MLMC estimators with pre-filtering only, post-filtering only, and both pre- and post-filtering, along with the (unfiltered) MLMC and MC estimators, on the experiment with $D = 0.06$. The total variances reported on fig. 5(a) show that both the pre-filtering and the post-filtering operations contribute to reducing the variance of the estimator. The cumulative variance of the 6-level estimators ($L = 5$) presented in fig. 5(b) explicitly shows the effects of the pre- and post-filtering on the different frequencies of the variance. Using an estimator with post-filtering alone leads to a reduction of the variance in the high frequencies, since the spurious high-frequencies components introduced by the prolongation operator are damped. In addition, we observe that the post-filtering also improves the estimation of the low frequencies. On the other hand, the pre-filtering reduces the frequencies that cannot be represented on the coarse grids and that lead to spurious low-frequency



(a) $D = 0.06$. (b) $D = 0.01$.

Figure 4: Total variance of the MC estimator ($L = 0$) and different MLMC and F-MLMC estimators ($L \in \{1, \dots, 5\}$), for the illustrative 1D estimation problem described in section 5.2, with length-scale $D = 0.06$ (left) and $D = 0.01$ (right). The finest level L always corresponds to a discretization with $n_L = 512$ cells, and the total budget is set to $\mathcal{C} = 100\mathcal{C}_L$. The variance is estimated from 1000 estimators.

components, thus improving the variance in those low frequencies, corresponding to the very first entries of the cumulative variance vector $\boldsymbol{\nu}^{\text{cml}}$. The joint use of pre- and post-filtering combines both benefits, further reducing the variance over the entire range of frequencies.

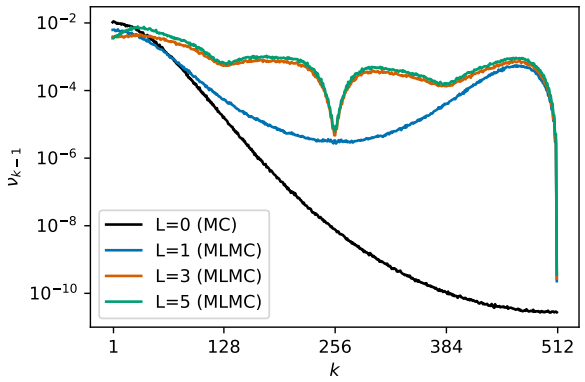


(a) Total variance. (b) Cumulative variance.

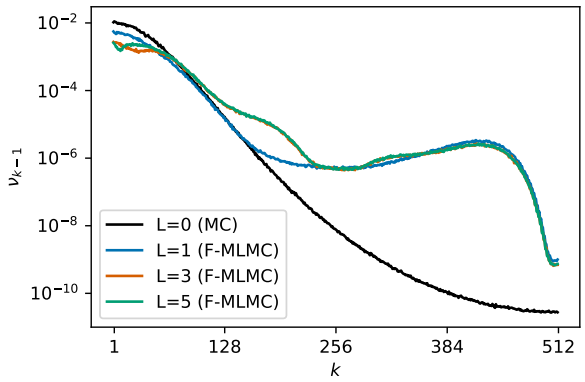
Figure 5: Total variance (left) of the MC estimator ($L = 0$) and different MLMC and F-MLMC estimators, ($L \in \{1, \dots, 5\}$), and cumulative spectral variance (right) of the different estimators with $L = 5$. Both figures are for the illustrative 1D estimation problem described in section 5.2, with length-scale $D = 0.06$. The finest level L always corresponds to a discretization with $n_L = 512$ cells, and the total budget is set to $\mathcal{C} = 100\mathcal{C}_L$. The variance is estimated from 1000 estimators.

It should be noted that the length-scale $D = 0.06$ used for the experiments presented above induces output fields mostly composed of large scales (low frequencies). Decreasing its value increases the frequencies of the output field, thus introducing smaller scales. Figure 6 shows the spectral and cumulative variance of the MC, MLMC and F-MLMC estimators for a smaller length-scale $D = 0.01 \approx 5n_L^{-1}$. We observe that the decay of the spectral variance of the MC estimator as the frequency increases is slower than for $D = 0.06$, so that the variance is concentrated on a wider low-frequency range. For the MLMC estimators, we see that the significant deterioration of the variance in the

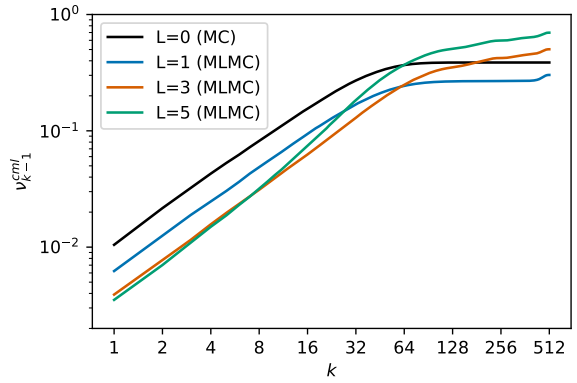
high-frequencies is no longer compensated by a better estimation in the low frequencies, except for the 2-level estimator, which remains slightly better than the single-level estimator in terms of total variance. Again, the addition of filters improves the multilevel estimation in both the low and the high frequencies, which in turn benefits the cumulative and thus the total variance. Filtering is here even more beneficial than for $D = 0.06$, in the sense that the F-MLMC estimator has a significantly lower variance than the single-level MC estimator, as shown in fig. 4(b), which the unfiltered MLMC estimator fails to achieve. Specifically, the variance of the 4-level F-MLMC estimator is reduced by about 60% compared to the single-level MC estimator, and by more than 40% compared to the best, 3-level unfiltered MLMC estimator.



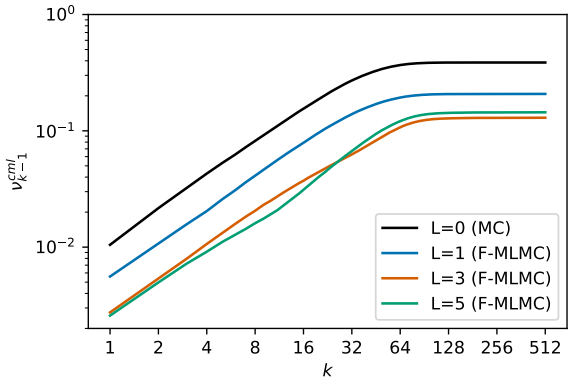
(a) Spectral variance $\boldsymbol{\nu}$ of MC and MLMC estimators.



(b) Spectral variance $\boldsymbol{\nu}$ of MC and F-MLMC estimators.



(c) Cumulative spectral variance $\boldsymbol{\nu}^{\text{cml}}$ of MC and MLMC estimators.



(d) Cumulative spectral variance $\boldsymbol{\nu}^{\text{cml}}$ of MC and F-MLMC estimators.

Figure 6: Same as fig. 3 but with length-scale $D = 0.01$.

These experiments highlight that the effect of spurious high frequencies caused by grid transfer operations is detrimental to the MLMC estimator. The addition of pre- and post-filtering operations is necessary to mitigate these effects, so that the multilevel estimator can reach its full potential.

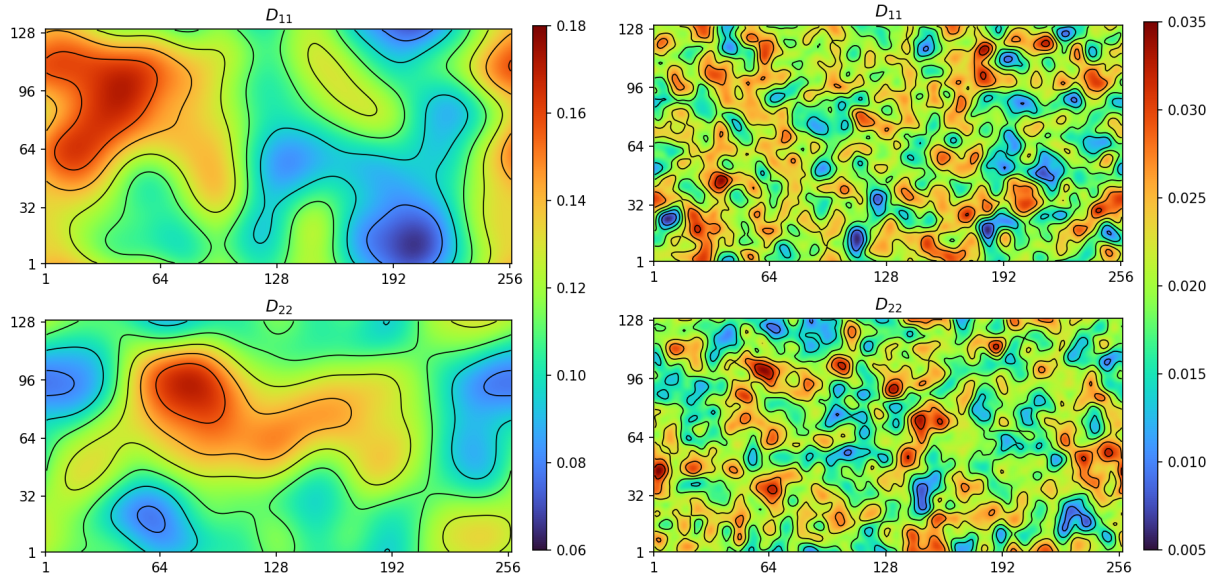
5.3 2D application

We now apply the MLMC and F-MLMC methodology to the variance estimation problem described in section 5.1, in a 2D setting with more complex diffusion operators than those used in the 1D illustration of the previous section. Specifically, the diffusivity field $\mathbf{K}$ is now specified to be non-uniform, making the variance field also non-uniform. The considered domain is $\mathcal{D} = (0, 2) \times (0, 1) \subset \mathbb{R}^2$. The boundary conditions are chosen periodic along both directions. The 2D diffusivity tensor field $\mathbf{K}$ is chosen to be

diagonal and heterogeneous

$$\mathbf{K} = \frac{1}{2m-4}\begin{bmatrix} D_{11}^2 & 0 \\ 0 & D_{22}^2 \end{bmatrix}, \tag{62}$$

where, for $i = 1, 2$, $D_{ii} \colon \mathcal{D} \to \mathbb{R}$ represents a length-scale field in the i-th direction. As previously, we let $m = 2q = 10$. We model $D_{11} = \zeta(\omega_1)$ and $D_{22} = \zeta(\omega_2)$ as two different realizations of a 2D, periodic Gaussian random field ζ over $\mathcal{D}$ of uniform mean μ_ζ, and of quasi-Gaussian covariance structure with uniform variance $\sigma_\zeta^2 = (\mu_\zeta/5)^2 = 0.04\mu_\zeta^2$ and uniform length-scale D_ζ. In the following experiments, two sets of parameters are considered for ζ, namely $(\mu_\zeta = 0.12, D_\zeta = 0.2)$ and $(\mu_\zeta = 0.02, D_\zeta = 0.04)$. The corresponding realizations used in the subsequent experiments are depicted in fig. 7.



(a) $\mu_\zeta = 0.12$ and $D_\zeta = 0.2$.

(b) $\mu_\zeta = 0.02$ and $D_\zeta = 0.04$.

Figure 7: Length-scale fields D_{11} (top) and D_{22} (bottom) used in the reported experiments for the definition of $\mathbf{K}$ in eq. (62), with $\mu_\zeta = 0.12$ and $D_\zeta = 0.2$ (left), and with $\mu_\zeta = 0.02$ and $D_\zeta = 0.04$ (right).

The diffusion operator in eq. (53) is discretized on Cartesian grids of size $n_\ell = n_\ell^x \times n_\ell^y$, with the length-scale fields D_{11} and D_{22} discretized at edge centers, and the solution discretized at cell centers, as depicted in fig. 8. The coarse operators $\mathbf{A}_\ell$ are based on restrictions of D_{11} and D_{22}, obtained by recursively averaging, at each point of a given coarse level, the values of the two nearest points of the immediately finer length-scale field, from the finest level L down to the desired level $\ell < L$. The finest grid considered here is composed of $n_L = 256 \times 128$ cells. Three coarser grids are used with a uniform coarsening factor of 4, i.e., $n_{L-1} = 128 \times 64$, $n_{L-2} = 64 \times 32$ and $n_{L-3} = 32 \times 16$. The resulting Gram matrix on level ℓ is given by $\mathbf{W}_\ell = 2n_\ell^{-1}\mathbf{I}_{n_\ell}$. The discrete operators $\mathbf{A}_\ell \in \mathbb{R}^{n_\ell \times n_\ell}$ are designed to apply to and return vectors of size n_ℓ whose entries are associated with cell centers that are sorted by increasing x-coordinate first, then by increasing y-coordinate. In other words, the entry indexed by $k = jn_\ell^x + i$ in such vectors is associated with a cell center located at $(x_i, y_j) \in \mathbb{R}^2$, where $x_i := 2(i + 1/2)/n_\ell^x$ and $y_j := (j + 1/2)/n_\ell^y$, for $i = 0, \ldots, n_\ell^x - 1$ and $j = 0, \ldots, n_\ell^y - 1$.

The hierarchy of simulators $(f_\ell)_{\ell=0}^L$ is defined through eq. (8) by $\tilde{f}_\ell \colon \mathbf{x}_\ell \mapsto (\mathbf{A}_\ell \mathbf{x}_\ell) \odot (\mathbf{A}_\ell \mathbf{x}_\ell)$, and we are interested in the multilevel estimation of the fine discretized field $\boldsymbol{\theta}_L := \mathbb{E}[f_L(\mathbf{X}_L)]$. With the Cartesian ordering of the unknowns described above, the 2D prolongation and restriction operators may be constructed as the Kronecker product of their 1D counterpart, defined in eq. (22), in the x

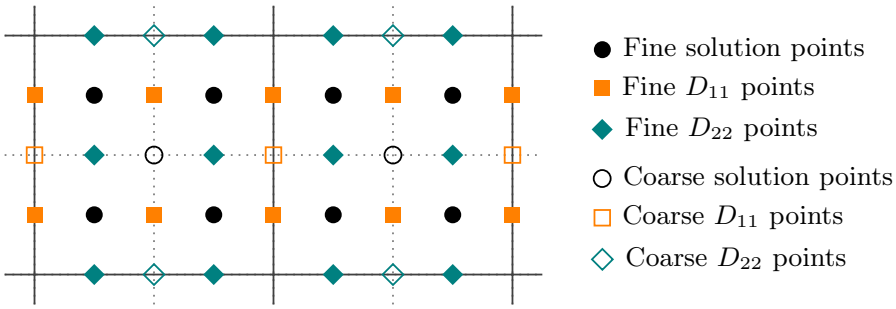


Figure 8: Illustration of the 2D discretization of the differential operator $\nabla \cdot \mathbf{K}\nabla$ in eq. (53) on a coarse uniform grid, represented by solid lines, whose sub-tessellation into a finer, geometrically nested, uniform grid is represented by dotted lines. The discretized solution points on the fine grid (respectively, on the coarse grid) are located at the center of the fine (respectively, coarse) cells and represented by ● (respectively, ○). The fine (respectively, coarse) discretized scalar field D_{11} is defined at points located on the fine (respectively, coarse) vertical edges and represented by ■ (respectively, □). The fine (respectively, coarse) discretized scalar field D_{22} is defined at points located on the fine (respectively, coarse) horizontal edges and represented by ◆ (respectively, ◇).

and y directions. The 2D grid transfer operators defined in this manner satisfy eq. (13). Likewise, the second-order 2D Shapiro filter is defined as the Kronecker product of two 1D, second-order Shapiro filters defined in eq. (42). The cost model is still considered linear in the number of cells, implying here that one simulator evaluation on level ℓ is as computationally expensive as 4 simulator evaluations on level $\ell - 1$. The baseline is a crude, single-level MC ($L = 0$) computed with a sample of size 100; hence the total budget is $\mathcal{C} = 100\mathcal{C}_L$.

As discussed in section 5.1, the k-th entry $(\boldsymbol{\theta}_L)_k$ of the exact discretized field $\boldsymbol{\theta}_L$ can be computed explicitly as $(\boldsymbol{\theta}_L)_k = (\mathbf{L}_L \mathbf{e}_k)_k$, for $k = 1, \ldots, n_L$, where $\mathbf{L}_L$ denotes the discrete diffusion operator defined in eq. (54) on the finest level L and $\mathbf{e}_k$ denotes the k-th canonical basis vector of $\mathbb{R}^{n_L}$ (i.e., the k-th column of $\mathbf{I}_{n_L}$). The top-left sub-figure of fig. 9 shows the variance field $\boldsymbol{\theta}_L$ for the tensor field $\mathbf{K}$ obtained by eq. (62) from the length-scale fields D_{11} and D_{22} with parameters $\mu_\zeta = 0.12$ and $D_\zeta = 0.2$ depicted in fig. 7(a). The other sub-figures represent the expectation of the compared estimators, namely the single-level MC estimator, and the 2- and 4-level (F-)MLMC estimators. The expectation is approximated from 500 estimators, each constructed with a computational budget $\mathcal{C} = 100\mathcal{C}_L$. These figures confirm a key property of the MC and (F-)MLMC estimators, namely that they are unbiased. Indeed, their expectations visually coincide (up to statistical error due to the estimation) with the reference, i.e., $\mathbb{E}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$. Consequently, the MSE solely consists of the variance of the estimators.

Figure 10 shows the variance of the considered estimators. Although the variance of the MLMC estimators (middle row) visually is lower than that of the MC estimator (top row), we clearly observe high-frequency fluctuations in the variance field, both for the 2-level (left) and the 4-level (right) unfiltered MLMC estimators. These high-frequency components of the variance may have significant consequences on individual estimations of $\boldsymbol{\theta}$. Indeed, while the MLMC estimations will, on average, match the desired field (owing to the unbiasedness of the estimators), individual estimations will be polluted by high-frequency error components because of the large high-frequency components of the variance. The bottom row of fig. 10 demonstrates, at least visually, that the addition of filtering effectively damps these high-frequency components and reduced the variance. These observations are confirmed by fig. 11, which shows the spectral decomposition of the variance in the Hartley space, allowing for the visualization of the contribution of each scale (or frequency) to the variance. The Hartley matrix $\mathbf{H}_L$ used to project the 2D variance (discretized) fields onto the Hartley spectral space is defined as the Kronecker product of two 1D Hartley matrices defined in eq. (23). The 2D spectral

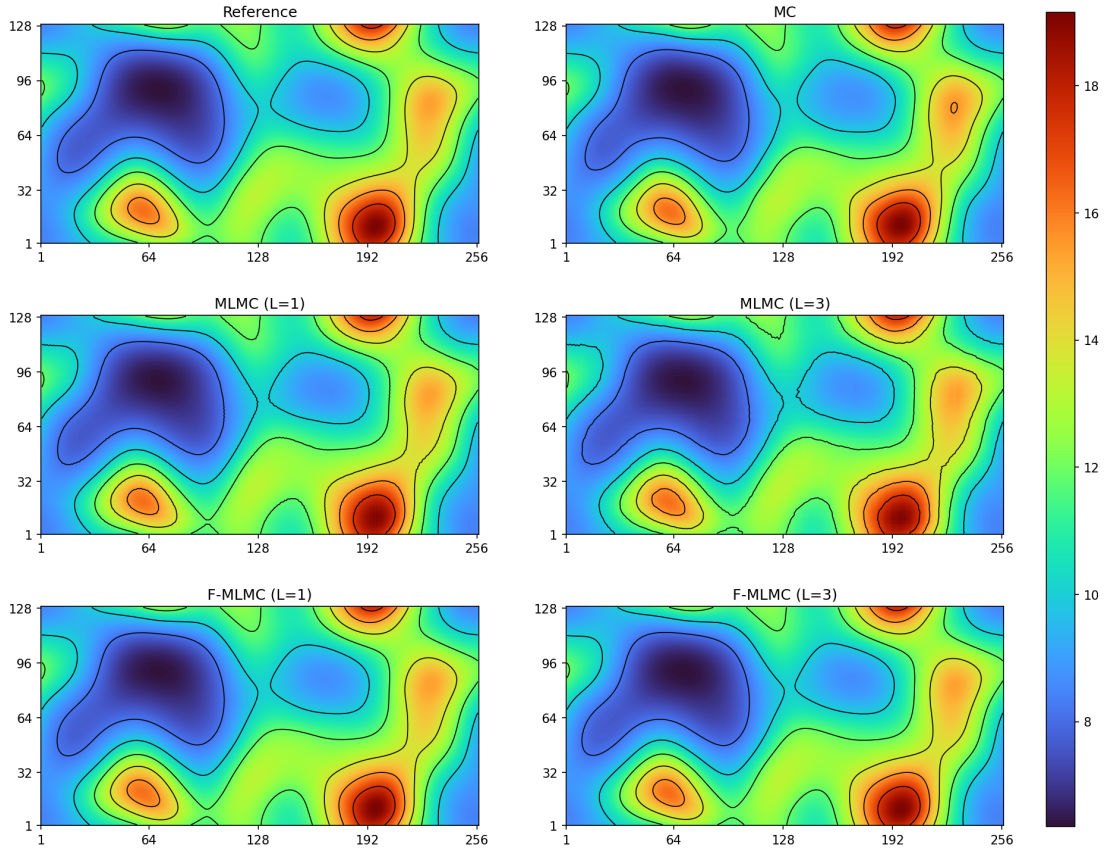


Figure 9: The top-left sub-figure shows the exact discretized field $\boldsymbol{\theta}$ on level L, while the other sub-figures depict the expectation of the single-level MC estimator (top-right), the expectation of the 2- and 4-level MLMC estimators (middle-left and middle-right), and the expectation of the 2- and 4-level F-MLMC estimators (bottom-left and bottom-right). The tensor field $\mathbf{K}$ corresponds to $\mu_\zeta = 0.12$ and $D_\zeta = 0.2$ (fig. 7(a)). The expectation is approximated from 500 estimators, each constructed with a computational budget $\mathcal{C} = 100\mathcal{C}_L$.

variance of an unbiased estimator $\hat{\boldsymbol{\theta}}_L$ of $\boldsymbol{\theta}_L$, representing either the MC estimator on the finest level L, or the MLMC or F-MLMC estimator, is defined as $\boldsymbol{\nu} = (\nu_k)_{k=0}^{n_L-1} \in \mathbb{R}^{n_L}$, with

$$\forall k = 0, \dots, n_L - 1, \quad \nu_k := 4n_L^{-2}\mathbb{E}[((\mathbf{h}_k^L)^{\mathrm{T}}(\hat{\boldsymbol{\theta}}_L - \boldsymbol{\theta}_L))^2] = 4n_L^{-2}\mathbb{V}[(\mathbf{h}_k^L)^{\mathrm{T}}\hat{\boldsymbol{\theta}}_L], \tag{63}$$

where $\mathbf{h}_k^L$ denotes the k-th column of $\mathbf{H}_L$. Again, the columns of $\mathbf{H}_L$ are re-ordered so that, in fig. 11, the frequencies in the x and y directions increase along the associated axes, starting from the lower-left corner, corresponding to low frequencies (large scales). Specifically, the 2D Hartley matrix with reordered columns is constructed as the Kronecker product of two 1D Hartley matrices, each with reordered columns. We observe that the variance of the unfiltered MLMC estimators (middle row) clearly exhibits larger high-frequency components than the single-level MC estimator (top row), with values of the same order of magnitude as the low-frequency components. This not only translates into noticeable high-frequency fluctuations of the variance field, as evidenced in fig. 10, but it may also deteriorate the overall variance, as was the case in the 1D illustration (see section 5.2).

With the Cartesian indexing and the re-ordering of the Hartley basis vectors described earlier, the 2D cumulative variance $\boldsymbol{\nu}^{\mathrm{cml}} = (\nu_k^{\mathrm{cml}})_{k=0}^{n_L^y-1} \in \mathbb{R}^{n_L^y}$ is computed by adding the components of the 2D spectral variance eq. (63) shown in fig. 11 in rectangle patterns starting with the bottom-left corner.

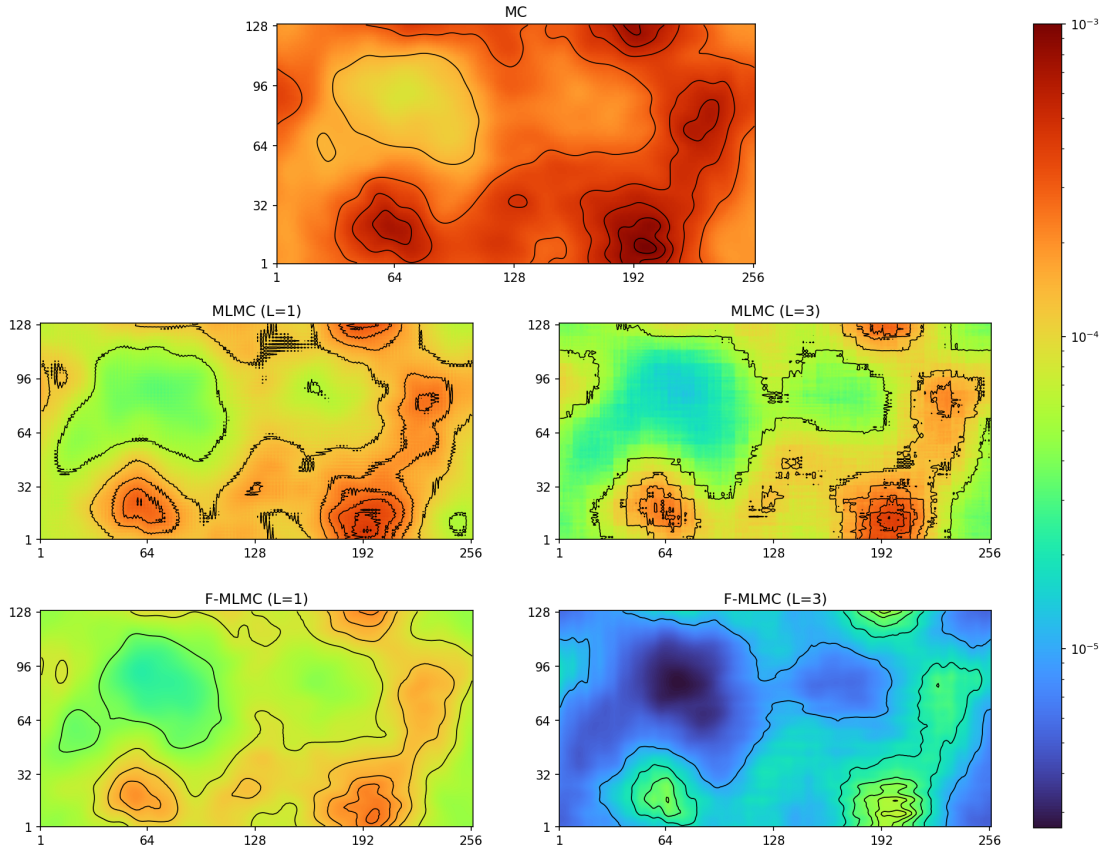


Figure 10: Variance of the MC estimator (top), of the 2- and 4-level MLMC estimators (middle-left and middle-right), and of the 2- and 4-level F-MLMC estimators (bottom-left and bottom-right). The tensor field $\mathbf{K}$ corresponds to $\mu_\zeta = 0.12$ and $D_\zeta = 0.2$ (fig. 7(a)). The variance is approximated from 500 estimators, each constructed with a computational budget $\mathcal{C} = 100\mathcal{C}_L$.

Specifically,

$$\nu_k^{\text{cml}} = \sum_{j=0}^{k} \sum_{i=0}^{2k+1} \nu_{jn_L^x+i}, \quad k = 0, \dots, n_L^y - 1. \tag{64}$$

As such, the first components of $\boldsymbol{\nu}^{\text{cml}}$ represent the cumulative variance associated with the low frequencies (large scales), while the last component $\nu^{\text{cml}}_{n_L^y-1}$ coincides with the total variance of the estimator. Figure 12 presents the cumulative variance of the MLMC and F-MLMC estimators for $\mathbf{K}$ corresponding to $(\mu_\zeta = 0.12, D_\zeta = 0.2)$ and $(\mu_\zeta = 0.02, D_\zeta = 0.04)$. In the first case (fig. 12(a)), for both the MC, the MLMC and the F-MLMC estimators, most of the variance is concentrated in the lower frequencies. Although the low-frequency contribution to the variance is significantly reduced by the MLMC estimator, a non-negligible contribution of the high frequencies to the variance is noticeable. Adding filters reduces the error in the high frequencies, as was observed previously in the spectral decomposition of the variance (fig. 11), but it also reduces the error on the lower frequencies, thus leading to lower total variance. In the second case (fig. 12(b)), corresponding to a tensor field $\mathbf{K}$ with smaller scales (see fig. 7(b)), the MLMC estimators deteriorate the variance compared to the crude MC estimator. The addition of a coarser grid ($L = 1$) and the corresponding grid transfer operators induce significant variance in the high-frequency components, leading to an increased total variance. The degradation is all the more pronounced as coarser grids are added ($L = 2$ and $L = 3$). We observe that the addition of filters mitigates these effects and allows the F-MLMC estimators to produce an effective reduction of

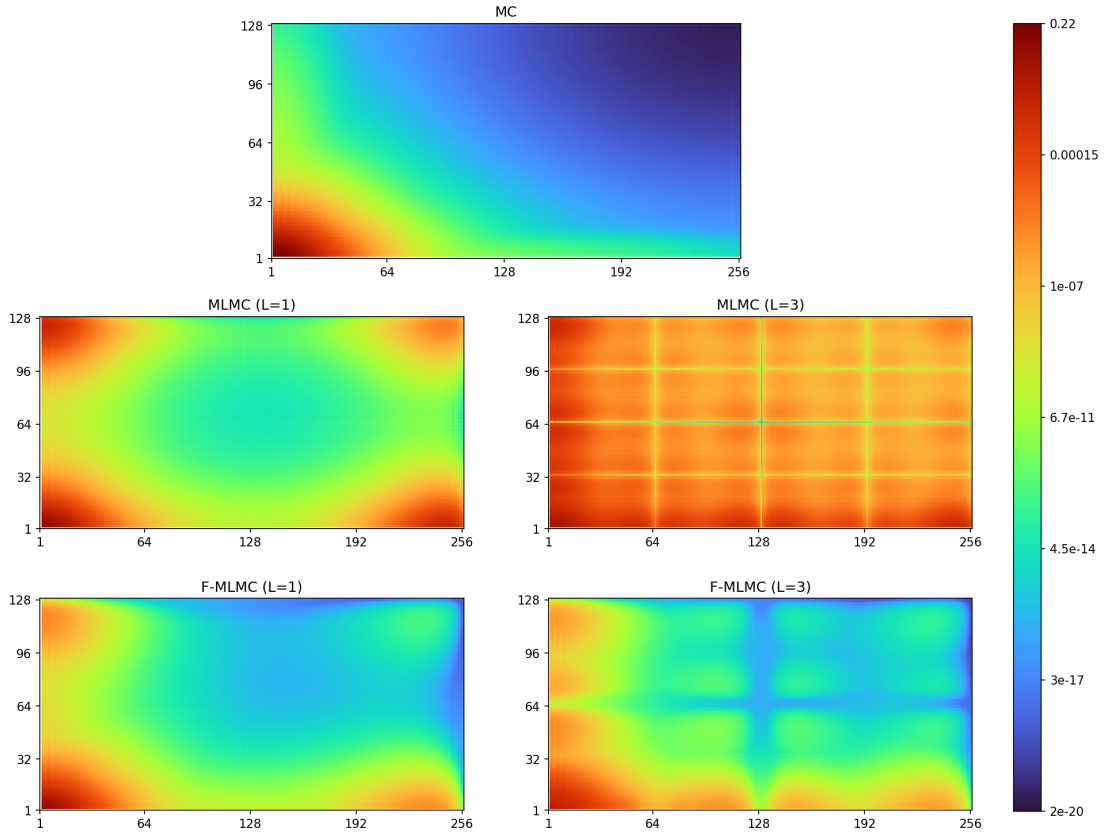
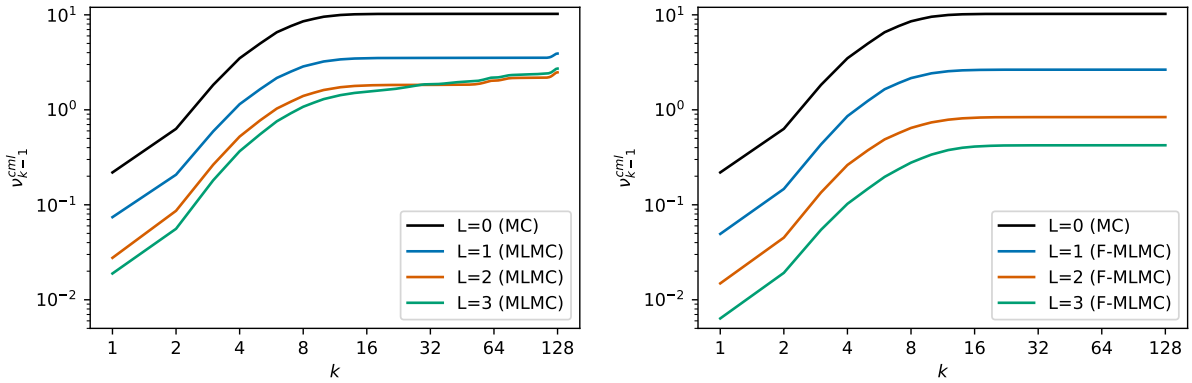


Figure 11: Spectral variance $\boldsymbol{\nu}$ of the MC estimator (top), of the 2- and 4-level MLMC estimators (middle-left and middle-right), and of the 2- and 4-level F-MLMC estimators (bottom-left and bottom-right). The tensor field $\mathbf{K}$ corresponds to $\mu_\zeta = 0.12$ and $D_\zeta = 0.2$ (fig. 7(a)). The variance is approximated from 500 estimators, each constructed with a computational budget $\mathcal{C} = 100\mathcal{C}_L$.

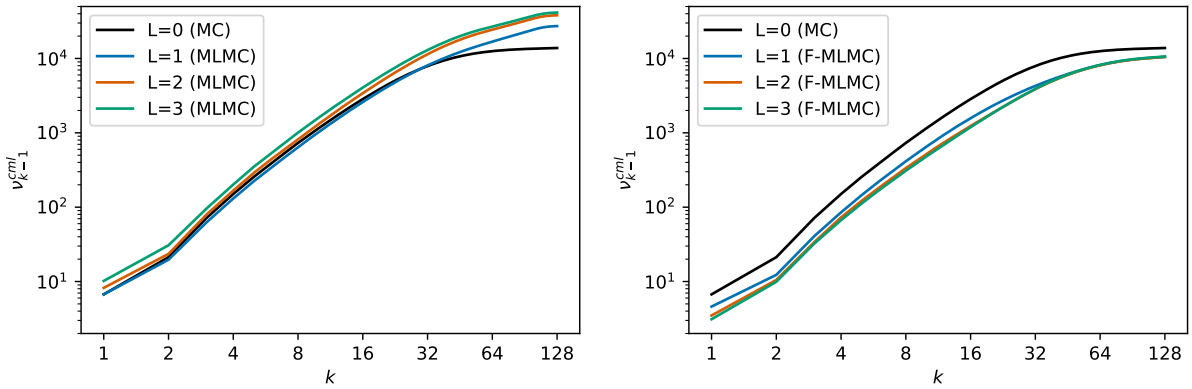
the variance. The addition of coarser grids ($L = 2$ and $L = 3$) does not bring further improvement, but does not deteriorate the total variance either. The low-frequency (large-scale) components are already well-captured by the two finer grids. This can be explained by the fact that the spectral content of the samples is shifted to higher frequencies. In such a case, a different grid hierarchy, starting from a finer grid, would be more appropriate, although finer discretizations may not always be available or affordable in an operational context. These results are summarized in terms of total variance in fig. 13.

6 Conclusion

In this paper, we focused on the estimation of the expectation of a discretized field using a multilevel, MLMC-like estimator. The different fidelity levels considered are grids of different resolutions, which requires the use of grid transfer operators in the estimator. The resulting MLMC estimator can then be used to reduce the variance of the estimation compared to a crude MC estimator, as confirmed with an idealized 1D problem of estimating the discretized intrinsic variance field of a diffusion-based covariance operator. However, projecting the variance of the MLMC estimator onto a spectral space revealed some discrepancy in the estimation of the different scales of the discretized field. In our experiments, the MLMC estimator was still able to achieve a lower total variance by improving the estimation of the low-frequency (large-scale) components compared to the MC estimator, but at the expense of degrading the estimation in the higher-frequency (smaller-scale) components.

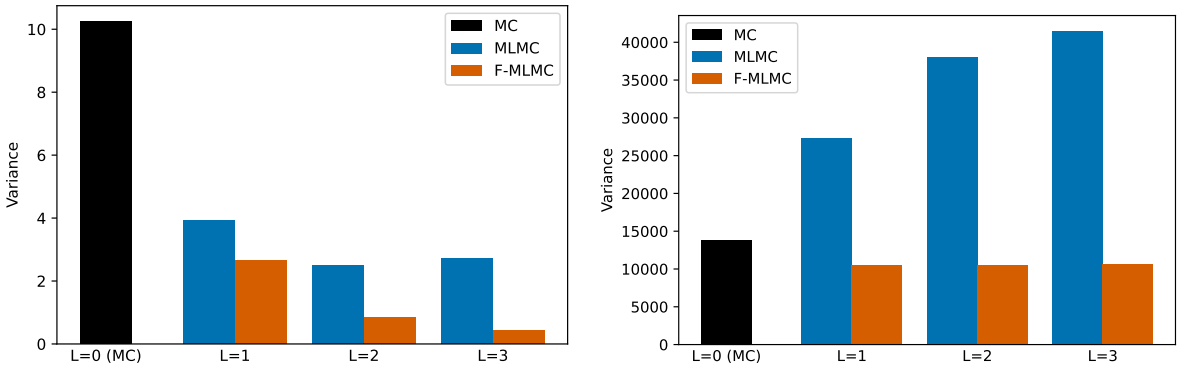


(a) $\mathbf{K}$ with $\mu_\zeta = 0.12$ and $D_\zeta = 0.2$ (fig. 7(a)).



(b) $\mathbf{K}$ with $\mu_\zeta = 0.02$ and $D_\zeta = 0.04$ (fig. 7(b)).

Figure 12: Cumulative variance of the MC estimator ($L = 0$) and of different MLMC estimators (left) and F-MLMC estimators (right) with $L \in \{1, 2, 3\}$. The tensor field $\mathbf{K}$ corresponds to $\mu_\zeta = 0.12$ and $D_\zeta = 0.2$ (top), and $\mu_\zeta = 0.02$ and $D_\zeta = 0.04$ (bottom). The variance is approximated from 500 estimators, each constructed with a computational budget $\eta = 100\mathcal{C}_L$.



(a) $\mathbf{K}$ with $\mu_\zeta = 0.12$ and $D_\zeta = 0.2$ (fig. 7(a)).

(b) $\mathbf{K}$ with $\mu_\zeta = 0.02$ and $D_\zeta = 0.04$ (fig. 7(b)).

Figure 13: Total variance of the MC estimator ($L = 0$) and of different MLMC estimators and F-MLMC estimators with $L \in \{1, 2, 3\}$. The tensor field $\mathbf{K}$ corresponds to $\mu_\zeta = 0.12$ and $D_\zeta = 0.2$ (left), and $\mu_\zeta = 0.02$ and $D_\zeta = 0.04$ (right). The variance is approximated from 500 estimators, each constructed with a computational budget $\mathcal{C} = 100\mathcal{C}_L$.

Inspired by multigrid methods, we proposed an improvement of the MLMC estimator by adding filtering operators, resulting in the F-MLMC estimator. Filtering out the high-frequency components

of a discretized field before restriction and after prolongation removes spurious features, thus yielding a better estimation of both the small- and large-scale components. These improvements significantly impact the total variance of the F-MLMC estimators, which is also reduced compared to the MLMC estimators in our experiments. In the specific case of linear, symmetric, circulant simulators, we quantified the effects of grid transfer and filtering operators on the total variance of (F-)MLMC estimators, which allowed us to improve our understanding of the influence of each ingredient. The proposed F-MLMC estimators were applied to the problem of estimating the discretized intrinsic variance field of a 2D diffusion-based covariance operator with a non-uniform diffusivity field, which relies on non-linear simulators $\tilde{f}_\ell \colon \mathbf{x}_\ell \mapsto (\mathbf{A}_\ell \mathbf{x}_\ell) \odot (\mathbf{A}_\ell \mathbf{x}_\ell)$. The conclusions of these experiments were consistent with the theoretical results derived in the spectral analysis for linear, symmetric, circulant simulators. Specifically, F-MLMC estimators do reduce the variance in both the low and high frequencies compared to their unfiltered counterparts, thus improving the total variance. Even in experiments where MLMC estimators were not able to reduce (but actually deteriorated) the variance compared to a crude MC estimator, F-MLMC estimators still achieved lower total variance.

It should be noted that, for the particular problem of estimating discretized variance fields considered in this paper, the proposed MLMC and F-MLMC estimators are not guaranteed to be almost surely non-negative. Although negative estimates were not encountered in our experiments, this is nonetheless a serious limitation of the multilevel estimators, as already pointed out for the MLMC estimation of the variance of random variables (possibly with values in Hilbert spaces) [12]. A similar issue exists for the estimation of covariance matrices, for which multilevel estimators are not guaranteed to be almost surely SPD [18], although advanced (but computationally expensive) approaches have been proposed to design multilevel estimators that are SPD by construction [19, 20]. These crucial issues still constitute an open research area.

Nonetheless, the investigations conducted in this paper and the proposed F-MLMC estimator expand the range of use of MLMC-like methods to discretized fields. In our study, the use of second-order Shapiro filters demonstrated the benefits of applying pre- and post-smoothing at each level of the MLMC estimators. A potential next step would be to investigate whether conditions can be derived for selecting the grid transfer and filtering operators, similar to the conditions stated in [3, 64] for multigrid methods. Furthermore, extensions of the MLBLUE techniques [9, 11] to discretized fields may allow for the derivation of optimal spectral weights for each scale component of the discretized field to act as a post-prolongation filter [15, sections 4.3–4.4], possibly in combination with the spectral pre-restriction smoothing technique of [33]. Finally, an interesting research avenue would consist of exploiting further the hierarchical nature of MLMC techniques by combining them with multigrid iterative methods for solving the systems of linear equations involved in the normalization problem of section 5, using the same grid hierarchy, possibly along the lines of [65, 66] or [67, 68].

Acknowledgments

This project has received financial support from the CNRS (Centre National de la Recherche Scientifique) through the 80|Prime program and the French national program LEFE (Les Enveloppes Fluides et l'Environnement).

A Orthogonality of the Hartley matrix

We focus here on a Hartley matrix $\mathbf{H}$ of arbitrary size n,

$$(\mathbf{H})_{j,k} := \cos\alpha_{jk} + \sin\alpha_{jk}, \quad \alpha_{jk} := \frac{(2j+1)k\pi}{n}, \quad \forall j,k = 0,\ldots,n-1. \tag{65}$$

We have, for $i,j = 0,\ldots,n-1$,

$$(\mathbf{H}\mathbf{H}^{\mathrm{T}})_{i,j} = \sum_{k=0}^{n-1}(\cos\alpha_{ik} + \sin\alpha_{ik})(\cos\alpha_{jk} + \sin\alpha_{jk}) = \sum_{k=0}^{n-1}\big[\cos(\alpha_{ik}-\alpha_{jk}) + \sin(\alpha_{ik}+\alpha_{jk})\big] \tag{66}$$

$$= \sum_{k=0}^{n-1}\cos\frac{2(i-j)k\pi}{n} + \sum_{k=0}^{n-1}\sin\frac{2(i+j+1)k\pi}{n}. \tag{67}$$

From [69, 1.342 (1&2)], we deduce, for $i,j = 0,\ldots,n-1$,

$$\sum_{k=0}^{n-1}\cos\frac{2(i-j)k\pi}{n} = n\delta_{ij}, \qquad \sum_{k=0}^{n-1}\sin\frac{2(i+j+1)k\pi}{n} = 0, \tag{68}$$

where δ_{ij} denotes the Kronecker delta, thus proving that $\mathbf{H}\mathbf{H}^{\mathrm{T}} = n\mathbf{I}_n$. Similarly,

$$(\mathbf{H}^{\mathrm{T}}\mathbf{H})_{i,j} = \sum_{k=0}^{n-1}\cos\frac{(2k+1)(i-j)\pi}{n} + \sum_{k=0}^{n-1}\sin\frac{(2k+1)(i+j)\pi}{n}, \tag{69}$$

and $\mathbf{H}^{\mathrm{T}}\mathbf{H} = n\mathbf{I}_n$ follows from [69, 1.342 (3&4)].

B Prolongation of Hartley vectors

For $j,k = 0,\ldots,n_0-1$,

$$(\mathbf{P}\mathbf{h}^0_k)_{2j} = (\mathbf{P}\mathbf{h}^0_k)_{2j+1} = (\mathbf{h}^0_k)_j = \cos\frac{(4j+2)k\pi}{n_1} + \sin\frac{(4j+2)k\pi}{n_1}. \tag{70}$$

Upon writing $(4j+2)k\pi = (4j+1)k\pi + k\pi$ and applying elementary trigonometric identitites,

$$(\mathbf{P}\mathbf{h}^0_k)_{2j} = c_k(\mathbf{h}^1_k)_{2j} + \sin\frac{k\pi}{n_1}\left(\cos\frac{(4j+1)k\pi}{n_1} - \sin\frac{(4j+1)k\pi}{n_1}\right). \tag{71}$$

Noticing that

$$\sin\frac{k\pi}{n_1} = \sin\left(\frac{(n_0+k)\pi}{2n_0} - \frac{\pi}{2}\right) = -c_{n_0+k}, \tag{72}$$

$$\cos\frac{(4j+1)k\pi}{n_1} = \cos\left(\frac{(4j+1)(n_0+k)\pi}{n_1} - \frac{\pi}{2}\right) = \sin\frac{(4j+1)(n_0+k)\pi}{n_1}, \tag{73}$$

$$\sin\frac{(4j+1)k\pi}{n_1} = \sin\left(\frac{(4j+1)(n_0+k)\pi}{n_1} - \frac{\pi}{2}\right) = -\cos\frac{(4j+1)(n_0+k)\pi}{n_1}, \tag{74}$$

we conclude that $(\mathbf{P}\mathbf{h}^0_k)_{2j} = c_k(\mathbf{h}^1_k)_{2j} - c_{n_0+k}(\mathbf{h}^1_{n_0+k})_{2j}$. Then, upon writing $(4j+2)k\pi = (4j+3)k\pi - k\pi$, a similar derivation leads to $(\mathbf{P}\mathbf{h}^0_k)_{2j+1} = c_k(\mathbf{h}^1_k)_{2j+1} - c_{n_0+k}(\mathbf{h}^1_{n_0+k})_{2j+1}$, thus proving eq. (24).

C Restriction of Hartley vectors

For $j = 0, \dots, n_0 - 1$ and $k = 0, \dots, 2n_0 - 1$,

$$(\mathbf{R}\mathbf{h}^1_k)_j = (\mathbf{h}^1_k)_{2j} + (\mathbf{h}^1_k)_{2j+1} \tag{75}$$

$$= \cos\frac{(4j+1)k\pi}{n_1} + \cos\frac{(4j+3)k\pi}{n_1} + \sin\frac{(4j+1)k\pi}{n_1} + \sin\frac{(4j+3)k\pi}{n_1} \tag{76}$$

$$= 2\cos\frac{(8j+4)k\pi}{2n_1}\cos\frac{2k\pi}{2n_1} + 2\sin\frac{(8j+4)k\pi}{2n_1}\cos\frac{2k\pi}{2n_1} \tag{77}$$

$$= 2c_k\left[\cos\frac{(2j+1)k\pi}{n_0} + \sin\frac{(2j+1)k\pi}{n_0}\right]. \tag{78}$$

Hence, for $j, k = 0, \dots, n_0 - 1$, it follows immediately that $\mathbf{R}\mathbf{h}^1_k = 2c_k\mathbf{h}^0_k$. Furthermore,

$$(\mathbf{R}\mathbf{h}^1_{n_0+k})_j = 2c_{n_0+k}\left[\cos\left(\frac{(2j+1)k\pi}{n_0} + \pi\right) + \sin\left(\frac{(2j+1)k\pi}{n_0} + \pi\right)\right] = -2c_{n_0+k}\mathbf{h}^0_k, \tag{79}$$

thus proving eq. (26).

D Symmetric circulant matrices are diagonalizable in the Hartley basis

In this appendix, we prove theorem 2 below, which states that symmetric, circulant matrices can be diagonalized in the cell-centered Hartley basis $\mathbf{H}$ defined by eq. (65). To do so, we start by recalling or proving results on the node-centered Fourier basis $\check{\mathbf{F}} \in \mathbb{C}^{n\times n}$ and Hartley bases $\check{\mathbf{H}}^{\pm} \in \mathbb{R}^{n\times n}$ defined by

$$\check{\mathbf{F}} := \check{\mathbf{H}}_c + i\check{\mathbf{H}}_s, \quad \check{\mathbf{H}}^{\pm} := \check{\mathbf{H}}_c \pm \check{\mathbf{H}}_s, \quad (\check{\mathbf{H}}_c)_{j,k} := \frac{1}{\sqrt{n}}\cos\frac{2jk\pi}{n}, \quad (\check{\mathbf{H}}_s)_{j,k} := \frac{1}{\sqrt{n}}\sin\frac{2jk\pi}{n}, \tag{80}$$

where $i \in \mathbb{C}$ denotes the unit imaginary number such that $i^2 = -1$. It is clear that $\check{\mathbf{H}}_c$ and $\check{\mathbf{H}}_s$ are real, symmetric matrices, and hence so are $\check{\mathbf{H}}^{\pm}$, while $\check{\mathbf{F}}$ is a complex, symmetric (but not Hermitian) matrix. Furthermore, $\check{\mathbf{F}}$ is unitary (see, e.g., [70]), i.e., $\check{\mathbf{F}}^*\check{\mathbf{F}} = \check{\mathbf{F}}\check{\mathbf{F}}^* = \mathbf{I}_n$, where $\check{\mathbf{F}}^* = \check{\mathbf{H}}_c - i\check{\mathbf{H}}_s$ is the Hermitian transpose of $\check{\mathbf{F}}$.

Lemma 1. *$\check{\mathbf{H}}_c$ and $\check{\mathbf{H}}_s$ are such that*

$$\check{\mathbf{H}}_c^2 + \check{\mathbf{H}}_s^2 = \mathbf{I}_n, \qquad \check{\mathbf{H}}_c\check{\mathbf{H}}_s = \check{\mathbf{H}}_s\check{\mathbf{H}}_c = \mathbf{0}_n, \qquad \check{\mathbf{H}}_c\mathbf{1}_n = \sqrt{n}\mathbf{e}_1, \qquad \check{\mathbf{H}}_s\mathbf{1}_n = \mathbf{0}_n, \tag{81}$$

where $\mathbf{0}_n := (0, \dots, 0)^{\mathrm{T}} \in \mathbb{R}^n$, $\mathbf{1}_n := (1, \dots, 1)^{\mathrm{T}} \in \mathbb{R}^n$, and $\mathbf{e}_1$ denotes the first column of $\mathbf{I}_n$.

Proof. The first two identities follow from [53, Lemma 1], while the last two identities follow from [69, 1.342 (1&2)]. □

Corollary 1. $\check{\mathbf{H}}^{\pm}\mathbf{1}_n = \sqrt{n}\mathbf{e}_1$.

Corollary 2. *$\check{\mathbf{H}}^{\pm}$ are orthogonal, i.e., $(\check{\mathbf{H}}^{\pm})^2 = \mathbf{I}_n$. Furthermore, $\check{\mathbf{H}}^+\check{\mathbf{H}}^- = \check{\mathbf{H}}^-\check{\mathbf{H}}^+ = \check{\mathbf{F}}^2$.*

Proof. By the definitions eq. (80) and lemma 1, we have

$$\check{\mathbf{H}}^{\pm}\mathbf{1}_n = \check{\mathbf{H}}_c\mathbf{1}_n \pm \check{\mathbf{H}}_s\mathbf{1}_n = \check{\mathbf{H}}_c\mathbf{1}_n = \sqrt{n}\mathbf{e}_1, \tag{82}$$

$$(\check{\mathbf{H}}^{\pm})^2 = (\check{\mathbf{H}}_c^2 + \check{\mathbf{H}}_s^2) \pm (\check{\mathbf{H}}_c\check{\mathbf{H}}_s + \check{\mathbf{H}}_s\check{\mathbf{H}}_c) = \check{\mathbf{H}}_c^2 + \check{\mathbf{H}}_s^2 = \mathbf{I}_n, \tag{83}$$

$$\check{\mathbf{F}}^2 = \check{\mathbf{H}}_c^2 - \check{\mathbf{H}}_s^2 + i(\check{\mathbf{H}}_c\check{\mathbf{H}}_s + \check{\mathbf{H}}_s\check{\mathbf{H}}_c) = \check{\mathbf{H}}_c^2 - \check{\mathbf{H}}_s^2, \tag{84}$$

$$\check{\mathbf{H}}^-\check{\mathbf{H}}^+ = \check{\mathbf{H}}_c^2 - \check{\mathbf{H}}_s^2 + (\check{\mathbf{H}}_c\check{\mathbf{H}}_s - \check{\mathbf{H}}_s\check{\mathbf{H}}_c) = \check{\mathbf{F}}^2 = \check{\mathbf{H}}^+\check{\mathbf{H}}^-. \tag{85}$$

□

Lemma 2. *Let* $\mathbf{A} = \mathrm{Circ}(\mathbf{a}) \in \mathbb{R}^{n\times n}$ *be a symmetric, circulant matrix whose first column is* $\mathbf{a} = (a_k)_{k=0}^{n-1} \in \mathbb{R}^n$*, with* $a_{n-i} = a_i$*, for* $i = 1, \dots, n-1$*. Then* $\check{\mathbf{H}}_c\mathbf{A}\check{\mathbf{H}}_s + \check{\mathbf{H}}_s\mathbf{A}\check{\mathbf{H}}_c = \check{\mathbf{H}}_c\mathbf{A}\check{\mathbf{H}}_s - \check{\mathbf{H}}_s\mathbf{A}\check{\mathbf{H}}_c = \mathbf{0}_n$.

Proof. Proven in the intermediary steps of the proof of [53, Theorem 1]. □

Theorem 1. *Let* $\mathbf{A} = \mathrm{Circ}(\mathbf{a}) \in \mathbb{R}^{n\times n}$ *be a symmetric, circulant matrix whose first column is* $\mathbf{a} = (a_k)_{k=0}^{n-1} \in \mathbb{R}^n$*, with* $a_{n-i} = a_i$*, for* $i = 1, \dots, n-1$*. Then* $\check{\mathbf{H}}^+\mathbf{A}\check{\mathbf{H}}^+ = \check{\mathbf{H}}^-\mathbf{A}\check{\mathbf{H}}^- = \mathbf{\Lambda}$*, where* $\mathbf{\Lambda} = \sqrt{n}\,\mathrm{Diag}(\check{\mathbf{H}}^+\mathbf{a}) = \sqrt{n}\,\mathrm{Diag}(\check{\mathbf{H}}^-\mathbf{a}) = \sqrt{n}\,\mathrm{Diag}(\check{\mathbf{H}}_c\mathbf{a})$.

Proof. The fact that $\check{\mathbf{H}}^+\mathbf{A}\check{\mathbf{H}}^+ = \sqrt{n}\,\mathrm{Diag}(\check{\mathbf{H}}^+\mathbf{a}) = \mathbf{\Lambda} = \check{\mathbf{H}}_c\mathbf{A}\check{\mathbf{H}}_c + \check{\mathbf{H}}_s\mathbf{A}\check{\mathbf{H}}_s$ follows from [53, Theorem 1], since symmetric, circulant matrices belong to the larger class of matrices considered in [53]. Then, from the definition of $\check{\mathbf{H}}^-$ and lemma 2, we have

$$\check{\mathbf{H}}^-\mathbf{A}\check{\mathbf{H}}^- = (\check{\mathbf{H}}_c\mathbf{A}\check{\mathbf{H}}_c + \check{\mathbf{H}}_s\mathbf{A}\check{\mathbf{H}}_s) - (\check{\mathbf{H}}_c\mathbf{A}\check{\mathbf{H}}_s + \check{\mathbf{H}}_s\mathbf{A}\check{\mathbf{H}}_c) = \check{\mathbf{H}}_c\mathbf{A}\check{\mathbf{H}}_c + \check{\mathbf{H}}_s\mathbf{A}\check{\mathbf{H}}_s = \mathbf{\Lambda}. \tag{86}$$

Finally, from lemma 1 and corollary 1,

$$\mathrm{diag}(\mathbf{\Lambda}) = \mathbf{\Lambda}\mathbf{1}_n = \check{\mathbf{H}}^-\mathbf{A}\check{\mathbf{H}}^-\mathbf{1}_n = \check{\mathbf{H}}_c\mathbf{A}\check{\mathbf{H}}_c\mathbf{1}_n + \check{\mathbf{H}}_s\mathbf{A}\check{\mathbf{H}}_s\mathbf{1}_n = \sqrt{n}\check{\mathbf{H}}^-\mathbf{A}\mathbf{e}_1 = \sqrt{n}\check{\mathbf{H}}_c\mathbf{A}\mathbf{e}_1, \tag{87}$$

which concludes the proof, since $\mathbf{A}\mathbf{e}_1 = \mathbf{a}$. □

Corollary 3. *Let* $\mathbf{A} = \mathrm{Circ}(\mathbf{a}) \in \mathbb{R}^{n\times n}$ *be a symmetric, circulant matrix and let* $\mathbf{\Lambda}$ *be defined as in theorem 1. Then* $\check{\mathbf{H}}^+\mathbf{A}\check{\mathbf{H}}^- = \check{\mathbf{H}}^-\mathbf{A}\check{\mathbf{H}}^+ = \mathbf{\Lambda}\check{\mathbf{F}}^2 = \check{\mathbf{F}}^2\mathbf{\Lambda}$.

Proof. From theorem 1 and corollary 2, we have

$$\check{\mathbf{H}}^+\mathbf{A}\check{\mathbf{H}}^- = \check{\mathbf{H}}^+\mathbf{A}\check{\mathbf{H}}^+\check{\mathbf{H}}^+\check{\mathbf{H}}^- = \check{\mathbf{H}}^+\mathbf{A}\check{\mathbf{H}}^+\check{\mathbf{F}}^2 = \mathbf{\Lambda}\check{\mathbf{F}}^2, \tag{88}$$

$$\check{\mathbf{H}}^-\mathbf{A}\check{\mathbf{H}}^+ = \check{\mathbf{H}}^-\mathbf{A}\check{\mathbf{H}}^-\check{\mathbf{H}}^-\check{\mathbf{H}}^+ = \check{\mathbf{H}}^-\mathbf{A}\check{\mathbf{H}}^-\check{\mathbf{F}}^2 = \mathbf{\Lambda}\check{\mathbf{F}}^2, \tag{89}$$

$$(\check{\mathbf{H}}^+\mathbf{A}\check{\mathbf{H}}^-)^{\mathrm{T}} = (\mathbf{\Lambda}\check{\mathbf{F}}^2)^{\mathrm{T}} = \check{\mathbf{F}}^2\mathbf{\Lambda} = \check{\mathbf{H}}^-\mathbf{A}\check{\mathbf{H}}^+ = \mathbf{\Lambda}\check{\mathbf{F}}^2, \tag{90}$$

where the last row follows from the symmetry of $\mathbf{A}$, $\check{\mathbf{H}}^\pm$ and $\check{\mathbf{F}}^2$. □

Lemma 3. $\mathbf{H} = \sqrt{n}(\check{\mathbf{H}}^+\mathbf{C} + \check{\mathbf{H}}^-\mathbf{S})$*, with* $\mathbf{C} := \mathrm{Diag}(\{\cos\frac{k\pi}{n}\}_{k=0}^{n-1})$ *and* $\mathbf{S} := \mathrm{Diag}(\{\sin\frac{k\pi}{n}\}_{k=0}^{n-1})$.

Proof. From elementary trigonometric identities, we have

$$\cos\frac{(2j+1)k\pi}{n} = \cos\frac{2jk\pi}{n}\cos\frac{k\pi}{n} - \sin\frac{2jk\pi}{n}\sin\frac{k\pi}{n}, \tag{91}$$

$$\sin\frac{(2j+1)k\pi}{n} = \sin\frac{2jk\pi}{n}\cos\frac{k\pi}{n} + \cos\frac{2jk\pi}{n}\sin\frac{k\pi}{n}, \tag{92}$$

then lemma 3 follows. □

Lemma 4. $\mathbf{C}\check{\mathbf{F}}^2\mathbf{S} + \mathbf{S}\check{\mathbf{F}}^2\mathbf{C} = \mathbf{0}_n$*, where* $\mathbf{C}$ *and* $\mathbf{S}$ *are defined as in lemma 3.*

Proof. From [53, Lemma 1], we have that, for $j, k = 0, \dots, n-1$, $(\check{\mathbf{F}}^2)_{j,k} = 1$ if $j = k = 0$ or $j + k = n$, and $(\check{\mathbf{F}}^2)_{j,k} = 0$ otherwise. Moreover, $(\mathbf{C}\check{\mathbf{F}}^2\mathbf{S})_{j,k} = (\check{\mathbf{F}}^2)_{j,k}\cos\frac{j\pi}{n}\sin\frac{k\pi}{n}$, and, by symmetry, $\mathbf{S}\check{\mathbf{F}}^2\mathbf{C} = (\mathbf{C}\check{\mathbf{F}}^2\mathbf{S})^{\mathrm{T}}$. Now, if $j = k = 0$, $(\mathbf{C}\check{\mathbf{F}}^2\mathbf{S})_{j,k} = (\mathbf{S}\check{\mathbf{F}}^2\mathbf{C})_{j,k} = 0$ trivially. Otherwise, if $j \neq 0$, $k \neq 0$, and $j + k \neq n$, $(\mathbf{C}\check{\mathbf{F}}^2\mathbf{S})_{j,k} = (\mathbf{S}\check{\mathbf{F}}^2\mathbf{C})_{j,k} = 0$. Finally, if $j + k = n$, elementary trigonometric identities induce

$$(\mathbf{C}\check{\mathbf{F}}^2\mathbf{S})_{j,k} = \cos\frac{j\pi}{n}\sin\frac{(n-j)\pi}{n} = \cos\frac{j\pi}{n}\sin\left(\pi - \frac{j\pi}{n}\right) = \cos\frac{j\pi}{n}\sin\frac{j\pi}{n}, \tag{93}$$

$$(\mathbf{C}\check{\mathbf{F}}^2\mathbf{S})_{k,j} = \sin\frac{j\pi}{n}\cos\frac{(n-j)\pi}{n} = \sin\frac{j\pi}{n}\cos\left(\pi - \frac{j\pi}{n}\right) = -\cos\frac{j\pi}{n}\sin\frac{j\pi}{n}, \tag{94}$$

so that $(\mathbf{C}\check{\mathbf{F}}^2\mathbf{S} + \mathbf{S}\check{\mathbf{F}}^2\mathbf{C})_{j,k} = (\mathbf{C}\check{\mathbf{F}}^2\mathbf{S})_{j,k} + (\mathbf{C}\check{\mathbf{F}}^2\mathbf{S})_{k,j} = 0$, thus completing the proof. □

We are now ready to state and prove theorem 2.

Theorem 2. *Let* $\mathbf{A} = \mathrm{Circ}(\mathbf{a}) \in \mathbb{R}^{n\times n}$ *be a symmetric, circulant matrix and let* $\mathbf{\Lambda}$ *be defined as in theorem 1. Then* $\mathbf{H}^{\mathrm{T}}\mathbf{A}\mathbf{H} = n\mathbf{\Lambda}$.

Proof. Starting from the expression of lemma 3, and applying theorem 1, corollary 3 and lemma 4, we obtain

$$n^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{A}\mathbf{H} = \mathbf{C}\check{\mathbf{H}}^{+}\mathbf{A}\check{\mathbf{H}}^{+}\mathbf{C} + \mathbf{S}\check{\mathbf{H}}^{-}\mathbf{A}\check{\mathbf{H}}^{-}\mathbf{S} + \mathbf{C}\check{\mathbf{H}}^{+}\mathbf{A}\check{\mathbf{H}}^{-}\mathbf{S} + \mathbf{S}\check{\mathbf{H}}^{-}\mathbf{A}\check{\mathbf{H}}^{+}\mathbf{C} \tag{95}$$

$$= \mathbf{C}\mathbf{\Lambda}\mathbf{C} + \mathbf{S}\mathbf{\Lambda}\mathbf{S} + \mathbf{C}\mathbf{\Lambda}\check{\mathbf{F}}^{2}\mathbf{S} + \mathbf{S}\mathbf{\Lambda}\check{\mathbf{F}}^{2}\mathbf{C} = \mathbf{\Lambda}(\mathbf{C}^2 + \mathbf{S}^2 + \mathbf{C}\check{\mathbf{F}}^{2}\mathbf{S} + \mathbf{S}\check{\mathbf{F}}^{2}\mathbf{C}), \tag{96}$$

so that $n^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{A}\mathbf{H} = \mathbf{\Lambda}(\mathbf{C}^2 + \mathbf{S}^2) = \mathbf{\Lambda}$ since $\mathbf{C}^2 + \mathbf{S}^2 = \mathbf{I}_n$ by elementary trigonometry. □

E Optimality of the Galerkin operator

Let $n_1, n_0 \in \mathbb{N}$ be such that $n_1 > n_0 > 0$, and let $\mathbf{A}_1 \in \mathbb{R}^{n_1\times n_1}$ and $\mathbf{A}_0 \in \mathbb{R}^{n_0\times n_0}$ be two linear operators. Let $\mathbf{R} \in \mathbb{R}^{n_0\times n_1}$ and $\mathbf{P} \in \mathbb{R}^{n_1\times n_0}$ be a restriction operator and a prolongation operator, respectively, such that $\mathbf{R} = \mathbf{P}^{\mathrm{T}}$ and $\mathbf{P}^{\mathrm{T}}\mathbf{W}_1\mathbf{P} = \mathbf{W}_0$ (see eq. (13)). We seek the operator $\mathbf{A}_0^*$ that minimizes $\|\mathbf{A}_1 - \mathbf{P}\mathbf{A}_0\mathbf{R}\|^2_{\mathrm{F},\mathbf{W}_1}$ among the linear operators $\mathbf{A}_0 \in \mathbb{R}^{n_0\times n_0}$. We restrict ourselves to the case where, for $\ell \in \{0, 1\}$, $\mathbf{W}_\ell = n_\ell^{-1}\mathbf{I}_{n_\ell}$, so that $\mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{R}\mathbf{R}^{\mathrm{T}} = (n_1/n_0)\mathbf{I}_{n_0}$ and

$$\mathbf{A}_0^* := \operatorname*{arg\,min}_{\mathbf{A}_0\in\mathbb{R}^{n_0\times n_0}} \|\mathbf{A}_1 - \mathbf{P}\mathbf{A}_0\mathbf{R}\|^2_{\mathrm{F},\mathbf{W}_1} = \operatorname*{arg\,min}_{\mathbf{A}_0\in\mathbb{R}^{n_0\times n_0}} \|\mathbf{A}_1 - \mathbf{P}\mathbf{A}_0\mathbf{R}\|^2_{\mathrm{F}}, \tag{97}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the classical (unweighted) Frobenius norm. We start by writing

$$\|\mathbf{A}_1 - \mathbf{P}\mathbf{A}_0\mathbf{R}\|^2_{\mathrm{F}} = \|\mathbf{A}_1 - \mathbf{P}\mathbf{A}_0\mathbf{R}\|^2_{\mathrm{F}} = \|\operatorname{vec}(\mathbf{A}_1 - \mathbf{P}\mathbf{A}_0\mathbf{R})\|_2^2 \tag{98}$$

$$= \|\operatorname{vec}(\mathbf{A}_1) - \operatorname{vec}(\mathbf{P}\mathbf{A}_0\mathbf{R})\|_2^2 = \|\operatorname{vec}(\mathbf{A}_1) - (\mathbf{R}^{\mathrm{T}}\otimes\mathbf{P})\operatorname{vec}(\mathbf{A}_0)\|_2^2, \tag{99}$$

so that eq. (97) is recast as an ordinary linear least squares problem whose solution is

$$\operatorname{vec}(\mathbf{A}_0^*) = \left((\mathbf{R}^{\mathrm{T}}\otimes\mathbf{P})^{\mathrm{T}}(\mathbf{R}^{\mathrm{T}}\otimes\mathbf{P})\right)^{-1}(\mathbf{R}^{\mathrm{T}}\otimes\mathbf{P})^{\mathrm{T}}\operatorname{vec}(\mathbf{A}_1) \tag{100}$$

$$= \left((\mathbf{R}\otimes\mathbf{P}^{\mathrm{T}})(\mathbf{R}^{\mathrm{T}}\otimes\mathbf{P})\right)^{-1}(\mathbf{R}\otimes\mathbf{P}^{\mathrm{T}})\operatorname{vec}(\mathbf{A}_1) \tag{101}$$

$$= (\mathbf{R}\mathbf{R}^{\mathrm{T}}\otimes\mathbf{P}^{\mathrm{T}}\mathbf{P})^{-1}\operatorname{vec}(\mathbf{P}^{\mathrm{T}}\mathbf{A}_1\mathbf{R}^{\mathrm{T}}) = ((n_1/n_0)\mathbf{I}_{n_0}\otimes(n_1/n_0)\mathbf{I}_{n_0})^{-1}\operatorname{vec}(\mathbf{R}\mathbf{A}_1\mathbf{P}) \tag{102}$$

$$= (n_0/n_1)^2(\mathbf{I}_{n_0}\otimes\mathbf{I}_{n_0})\operatorname{vec}(\mathbf{R}\mathbf{A}_1\mathbf{P}) = (n_0/n_1)^2\operatorname{vec}(\mathbf{R}\mathbf{A}_1\mathbf{P}), \tag{103}$$

thus proving that $\mathbf{A}_0^* = (n_0/n_1)^2\mathbf{R}\mathbf{A}_1\mathbf{P}$.

F Filtered restriction of Hartley vectors

We start by looking at the effect of the Shapiro filter on the fine Hartley basis vectors,

$$\forall j, k = 0, \ldots, n_1 - 1, \quad (\mathbf{S}_1\mathbf{h}_k^1)_j = \frac{1}{4}\left[(\mathbf{h}_k^1)_{j-1} + 2(\mathbf{h}_k^1)_j + (\mathbf{h}_k^1)_{j+1}\right]. \tag{104}$$

Elementary trigonometric identities imply that, for $j, k = 0, \ldots, n_1 - 1$,

$$\cos\frac{(2j-1)k\pi}{n_1} + \cos\frac{(2j+1)k\pi}{n_1} = 2c_k\cos\frac{2jk\pi}{n_1}, \tag{105}$$

$$\cos\frac{(2j+1)k\pi}{n_1} + \cos\frac{(2j+3)k\pi}{n_1} = 2c_k\cos\frac{(2j+2)k\pi}{n_1}, \tag{106}$$

$$\sin\frac{(2j-1)k\pi}{n_1} + \sin\frac{(2j+1)k\pi}{n_1} = 2c_k\sin\frac{2jk\pi}{n_1}, \tag{107}$$

$$\sin\frac{(2j+1)k\pi}{n_1} + \sin\frac{(2j+3)k\pi}{n_1} = 2c_k\sin\frac{(2j+2)k\pi}{n_1}, \tag{108}$$

leading to

$$(\mathbf{S}_1\mathbf{h}_k^1)_j = \frac{c_k}{2}\left[\cos\frac{2jk\pi}{n_1} + \cos\frac{(2j+2)k\pi}{n_1} + \sin\frac{2jk\pi}{n_1} + \sin\frac{(2j+2)k\pi}{n_1}\right] = c_k^2(\mathbf{h}_k^1)_j, \tag{109}$$

where the last identity is obtained by applying the same trigonometric identities as for eqs. (105) to (108), and thus showing that $\mathbf{S}_1\mathbf{h}_k^1 = c_k^2\mathbf{h}_k^1$, for $k = 0, \dots, n_1 - 1$. Then, eq. (24) implies that, for $k = 0, \dots, n_0 - 1$, $\bar{\mathbf{P}}\mathbf{h}_k^0 = \mathbf{S}_1\mathbf{P}\mathbf{h}_k^0 = c_k\mathbf{S}_1\mathbf{h}_k^1 - c_{n_0+k}\mathbf{S}_1\mathbf{h}_{n_0+k}^1$, from which eq. (43) follows. Furthermore, eq. (44) follows immediately from the direct application of eq. (26) to $\bar{\mathbf{R}}\mathbf{h}_k^1 = \mathbf{R}\mathbf{S}_1\mathbf{h}_k^1 = c_k^2\mathbf{R}\mathbf{h}_k^1$.

G Factorization of the diffusion-based covariance matrix

Let $\boldsymbol{\Delta}$ and $\mathbf{W}$ as defined in section 5. Because $\boldsymbol{\Delta}$ is self-adjoint with respect to $\langle\cdot,\cdot\rangle_{\mathbf{W}}$, we have

$$\mathbf{W}(\mathbf{I} - \boldsymbol{\Delta}) = \mathbf{W} - \mathbf{W}\boldsymbol{\Delta} = \mathbf{W} - \boldsymbol{\Delta}^{\mathrm{T}}\mathbf{W} = (\mathbf{I} - \boldsymbol{\Delta})^{\mathrm{T}}\mathbf{W}, \tag{110}$$

and thus, for $q \in \mathbb{N}$,

$$\mathbf{W}(\mathbf{I} - \boldsymbol{\Delta})^q = \mathbf{W}(\mathbf{I} - \boldsymbol{\Delta})(\mathbf{I} - \boldsymbol{\Delta})^{q-1} = (\mathbf{I} - \boldsymbol{\Delta})^{\mathrm{T}}\mathbf{W}(\mathbf{I} - \boldsymbol{\Delta})^{q-1} = \cdots = [(\mathbf{I} - \boldsymbol{\Delta})^{\mathrm{T}}]^q\mathbf{W}. \tag{111}$$

It follows that

$$\mathbf{A} := (\mathbf{I} - \boldsymbol{\Delta})^{-q}\mathbf{W}^{-1} = (\mathbf{W}(\mathbf{I} - \boldsymbol{\Delta})^q)^{-1} = ([(\mathbf{I} - \boldsymbol{\Delta})^q]^{\mathrm{T}}\mathbf{W})^{-1} = \mathbf{W}^{-1}[(\mathbf{I} - \boldsymbol{\Delta})^{-q}]^{\mathrm{T}}, \tag{112}$$

and thus

$$\mathbf{L} := (\mathbf{I} - \boldsymbol{\Delta})^{-2q}\mathbf{W}^{-1} = (\mathbf{I} - \boldsymbol{\Delta})^{-q}\mathbf{A} = (\mathbf{I} - \boldsymbol{\Delta})^{-q}\mathbf{W}^{-1}[(\mathbf{I} - \boldsymbol{\Delta})^{-q}]^{\mathrm{T}} = \mathbf{A}\mathbf{W}\mathbf{A}^{\mathrm{T}}. \tag{113}$$

References

[1] B. Peherstorfer, K. Willcox, and M. Gunzburger. "Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization." In: *SIAM Review* 60.3 (2018), pp. 550–591. ISSN: 0036-1445. DOI: 10.1137/16M1082469.

[2] A. Brandt. "Guide to Multigrid Development." In: *Multigrid Methods*. Ed. by W. Hackbusch and U. Trottenberg. Lecture Notes in Mathematics. Berlin, Heidelberg: Springer, 1982, pp. 220–312. ISBN: 978-3-540-39544-7. DOI: 10.1007/BFb0069930.

[3] W. Hackbusch. *Multi-Grid Methods and Applications*. Vol. 4. Springer Series in Computational Mathematics. Berlin, Heidelberg: Springer, 1985. ISBN: 978-3-662-02427-0. DOI: 10.1007/978-3-662-02427-0.

[4] U. Trottenberg, C. W. Oosterlee, and A. Schuller. *Multigrid*. Elsevier, 2000. ISBN: 978-0-08-047956-9. URL: https://www.elsevier.com/books/multigrid/trottenberg/978-0-08-047956-9.

[5] W. L. Briggs, V. E. Henson, and S. F. McCormick. *A Multigrid Tutorial, Second Edition*. Second. Society for Industrial and Applied Mathematics, 2000. ISBN: 978-0-89871-462-3. DOI: 10.1137/1.9780898719505.

[6] S. Heinrich. "Multilevel Monte Carlo Methods." In: *Large-Scale Scientific Computing*. Ed. by S. Margenov, J. Waśniewski, and P. Yalamov. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2001, pp. 58–67. ISBN: 978-3-540-45346-8. DOI: 10.1007/3-540-45346-6_5.

[7] M. B. Giles. "Multilevel Monte Carlo Path Simulation." In: *Operations Research* 56.3 (2008), pp. 607–617. ISSN: 0030-364X. DOI: 10.1287/opre.1070.0496.

[8] A. A. Gorodetsky, G. Geraci, M. S. Eldred, and J. D. Jakeman. "A Generalized Approximate Control Variate Framework for Multifidelity Uncertainty Quantification." In: *Journal of Computational Physics* 408 (2020), p. 109257. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2020.109257.

[9] D. Schaden and E. Ullmann. "On Multilevel Best Linear Unbiased Estimators." In: *SIAM/ASA Journal on Uncertainty Quantification* 8.2 (2020), pp. 601–635. DOI: 10.1137/19M1263534.

[10] M. B. Giles. "Multilevel Monte Carlo methods." In: *Acta Numerica* 24 (2015), pp. 259–328. ISSN: 0962-4929, 1474-0508. DOI: 10.1017/S096249291500001X.

[11] D. Schaden and E. Ullmann. "Asymptotic Analysis of Multilevel Best Linear Unbiased Estimators." In: *SIAM/ASA Journal on Uncertainty Quantification* 9.3 (2021), pp. 953–978. DOI: 10.1137/20M1321607.

[12] C. Bierig and A. Chernov. "Convergence analysis of multilevel Monte Carlo variance estimators and application for random obstacle problems." In: *Numerische Mathematik* 130.4 (2015), pp. 579–613. ISSN: 0945-3245. DOI: 10.1007/s00211-014-0676-3.

[13] C. Bierig and A. Chernov. "Estimation of arbitrary order central statistical moments by the multilevel Monte Carlo method." In: *Stochastics and Partial Differential Equations Analysis and Computations* 4.1 (2016), pp. 3–40. ISSN: 2194-041X. DOI: 10.1007/s40072-015-0063-9.

[14] P. Mycek and M. De Lozzo. "Multilevel Monte Carlo Covariance Estimation for the Computation of Sobol' Indices." In: *SIAM/ASA Journal on Uncertainty Quantification* 7.4 (2019), pp. 1323–1348. DOI: 10.1137/18M1216389.

[15] M. Destouches, P. Mycek, and S. Gürol. *Multivariate Extensions of the Multilevel Best Linear Unbiased Estimator for Ensemble-Variational Data Assimilation*. 2023. DOI: 10.48550/arXiv.2306.07017.

[16] M. Croci, K. E. Willcox, and S. J. Wright. "Multi-Output Multilevel Best Linear Unbiased Estimators via Semidefinite Programming." In: *Computer Methods in Applied Mechanics and Engineering* 413 (2023), p. 116130. ISSN: 00457825. DOI: 10.1016/j.cma.2023.116130.

[17] T. O. Dixon, J. E. Warner, G. F. Bomarito, and A. A. Gorodetsky. *Covariance Expressions for Multi-Fidelity Sampling with Multi-Output, Multi-Statistic Estimators: Application to Approximate Control Variates*. 2023. DOI: 10.48550/arXiv.2310.00125.

[18] H. Hoel, K. J. H. Law, and R. Tempone. "Multilevel ensemble Kalman filtering." In: *SIAM Journal on Numerical Analysis* 54.3 (2016), pp. 1813–1839. ISSN: 0036-1429. DOI: 10.1137/15M100955X.

[19] A. Maurais, T. Alsup, B. Peherstorfer, and Y. Marzouk. "Multi-Fidelity Covariance Estimation in the Log-Euclidean Geometry." In: *Proceedings of the 40th International Conference on Machine Learning*. PMLR, 2023, pp. 24214–24235.

[20] A. Maurais, T. Alsup, B. Peherstorfer, and Y. Marzouk. *Multifidelity Covariance Estimation via Regression on the Manifold of Symmetric Positive Definite Matrices*. 2023. DOI: 10.48550/arXiv.2307.12438. arXiv: 2307.12438 [cs, math, stat].

[21] M. Destouches, P. Mycek, S. Gürol, A. T. Weaver, S. Gratton, and E. Simon. "Multilevel Monte Carlo Methods for Ensemble Variational Data Assimilation." In: *EGUsphere* (2024), pp. 1–33. DOI: 10.5194/egusphere-2024-3628.

[22] L. M. Stein. *Interpolation of Spatial Data*. Springer-Verlag, New York, NY, 1999.

[23] A. Weaver and P. Courtier. "Correlation modelling on the sphere using a generalized diffusion equation." In: *Quarterly Journal of the Royal Meteorological Society* 127.575 (2001), pp. 1815–1846. ISSN: 1477-870X. DOI: 10.1002/qj.49712757518.

[24] R. J. Purser, W.-S. Wu, D. F. Parrish, and N. M. Roberts. "Numerical Aspects of the Application of Recursive Filters to Variational Statistical Analysis. Part I: Spatially Homogeneous and Isotropic Gaussian Covariances." In: *Monthly Weather Review* 131.8 (2003), pp. 1524–1535. ISSN: 1520-0493, 0027-0644. DOI: 10.1175//1520-0493(2003)131<1524:NAOTAO>2.0.CO;2.

[25] F. Lindgren, H. Rue, and J. Lindström. "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach." In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73.4 (2011), pp. 423–498. ISSN: 1467-9868. DOI: 10.1111/j.1467-9868.2011.00777.x.

[26] T. Bui-Thanh, O. Ghattas, J. Martin, and G. Stadler. "A Computational Framework for Infinite-Dimensional Bayesian Inverse Problems Part I: The Linearized Case, with Application to Global Seismic Inversion." In: *SIAM Journal on Scientific Computing* 35.6 (2013), A2494–A2523. ISSN: 1064-8275, 1095-7197. DOI: 10.1137/12089586X.

[27] D. Drzisga, B. Gmeiner, U. Rüde, R. Scheichl, and B. Wohlmuth. "Scheduling Massively Parallel Multigrid for Multilevel Monte Carlo Methods." In: *SIAM Journal on Scientific Computing* 39.5 (2017), S873–S897. ISSN: 1064-8275. DOI: 10.1137/16M1083591.

[28] A. T. Weaver, M. Chrust, B. Ménétrier, and A. Piacentini. "An evaluation of methods for normalizing diffusion-based covariance operators in variational data assimilation." In: *Quarterly Journal of the Royal Meteorological Society* 147.734 (2021), pp. 289–320. ISSN: 1477-870X. DOI: 10.1002/qj.3918.

[29] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup. "Multilevel Monte Carlo Methods and Applications to Elliptic PDEs with Random Coefficients." In: *Computing and Visualization in Science* 14.1 (2011), pp. 3–15. ISSN: 1432-9360, 1433-0369. DOI: 10.1007/s00791-011-0160-x.

[30] J. Charrier, R. Scheichl, and A. L. Teckentrup. "Finite Element Error Analysis of Elliptic PDEs with Random Coefficients and Its Application to Multilevel Monte Carlo Methods." In: *SIAM Journal on Numerical Analysis* 51.1 (2013), pp. 322–352. ISSN: 0036-1429, 1095-7170. DOI: 10.1137/110853054. URL: http://epubs.siam.org/doi/10.1137/110853054 (visited on 04/15/2020).

[31] A. Barth, C. Schwab, and N. Zollinger. "Multi-Level Monte Carlo Finite Element Method for Elliptic PDEs with Stochastic Coefficients." In: *Numerische Mathematik* 119.1 (2011), pp. 123–161. ISSN: 0029-599X, 0945-3245. DOI: 10.1007/s00211-011-0377-0. URL: http://link.springer.com/10.1007/s00211-011-0377-0 (visited on 04/15/2020).

[32] A. Abdulle, A. Barth, and C. Schwab. "Multilevel Monte Carlo Methods for Stochastic Elliptic Multiscale PDEs." In: *Multiscale Modeling & Simulation* 11.4 (2013), pp. 1033–1070. ISSN: 1540-3459, 1540-3467. DOI: 10.1137/120894725. URL: http://epubs.siam.org/doi/10.1137/120894725 (visited on 04/15/2020).

[33] A. Istratuca and A. Teckentrup. *Smoothed Circulant Embedding with Applications to Multilevel Monte Carlo Methods for PDEs with Random Coefficients*. 2024. DOI: 10.48550/arXiv.2306.13493. arXiv: 2306.13493 [cs, math].

[34] R. Wienands. *Practical Fourier Analysis for Multigrid Methods*. Chapman and Hall/CRC, 2004. ISBN: 978-0-429-14328-1. DOI: 10.1201/9781420034998.

[35] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. 2nd ed. Encyclopedia of Mathematics and Its Applications. Cambridge: Cambridge University Press, 2014. ISBN: 978-1-107-05584-1. DOI: 10.1017/CBO9781107295513. URL: https://www.cambridge.org/core/books/stochastic-equations-in-infinite-dimensions/6218FF6506BE364F82E3CF534FAC2FC5 (visited on 10/01/2024).

[36] T. Hytönen, J. Van Neerven, M. Veraar, and L. Weis. *Analysis in Banach Spaces*. Cham: Springer International Publishing, 2016. ISBN: 978-3-319-48520-1. DOI: 10.1007/978-3-319-48520-1.

[37] G. Madec, M. Bell, A. Blaker, C. Bricaud, D. Bruciaferri, M. Castrillo, D. Calvert, J. Chanut, E. Clementi, A. Coward, I. Epicoco, C. Éthé, J. Ganderton, J. Harle, K. Hutchinson, D. Iovino, D. Lea, T. Lovato, M. Martin, N. Martin, F. Mele, D. Martins, S. Masson, P. Mathiot, F. Mele, S. Mocavero, S. Müller, A. J. G. Nurser, S. Paronuzzi, M. Peltier, R. Person, C. Rousset, S. Rynders, G. Samson, S. Téchené, M. Vancoppenolle, and C. Wilson. *NEMO Ocean Engine Reference Manual*. 27. Reference Manual. Zenodo, 2023. DOI: 10.5281/zenodo.8167700. URL: https://zenodo.org/records/8167700.

[38] M. Kac and A. J. F. Siegert. "An Explicit Representation of a Stationary Gaussian Process." In: *The Annals of Mathematical Statistics* 18.3 (1947), pp. 438–442. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177730391.

[39] O. P. Le Maître and O. M. Knio. *Spectral Methods for Uncertainty Quantification*. Scientific Computation. Dordrecht: Springer Netherlands, 2010. ISBN: 978-90-481-3519-6. DOI: 10.1007/978-90-481-3520-2.

[40] W. Betz, I. Papaioannou, and D. Straub. "Numerical Methods for the Discretization of Random Fields by Means of the Karhunen–Loève Expansion." In: *Computer Methods in Applied Mechanics and Engineering* 271 (2014), pp. 109–129. ISSN: 0045-7825. DOI: 10.1016/j.cma.2013.12.010. URL: https://www.sciencedirect.com/science/article/pii/S0045782513003502.

[41] C. J. Gittelson, J. Könnö, C. Schwab, and R. Stenberg. "The Multi-Level Monte Carlo Finite Element Method for a Stochastic Brinkman Problem." In: *Numerische Mathematik* 125.2 (2013), pp. 347–386. ISSN: 0945-3245. DOI: 10.1007/s00211-013-0537-5. URL: https://doi.org/10.1007/s00211-013-0537-5.

[42] I. Graham, F. Kuo, D. Nuyens, R. Scheichl, and I. Sloan. "Quasi-Monte Carlo Methods for Elliptic PDEs with Random Coefficients and Applications." In: *Journal of Computational Physics* 230.10 (2011), pp. 3668–3694. ISSN: 00219991. DOI: 10.1016/j.jcp.2011.01.023.

[43] I. G. Graham, F. Y. Kuo, D. Nuyens, R. Scheichl, and I. H. Sloan. "Analysis of Circulant Embedding Methods for Sampling Stationary Random Fields." In: *SIAM Journal on Numerical Analysis* 56.3 (2018), pp. 1871–1895. ISSN: 0036-1429, 1095-7170. DOI: 10.1137/17M1149730.

[44] S. Mishra and C. Schwab. "Sparse Tensor Multi-Level Monte Carlo Finite Volume Methods for Hyperbolic Conservation Laws with Random Initial Data." In: *Mathematics of Computation* 81.280 (2012), pp. 1979–2018. ISSN: 0025-5718, 1088-6842. DOI: 10.1090/S0025-5718-2012-02574-9. URL: http://www.ams.org/jourcgi/jour-getitem?pii=S0025-5718-2012-02574-9.

[45] S. Mishra, C. Schwab, and J. Šukys. "Multilevel Monte Carlo Finite Volume Methods for Shallow Water Equations with Uncertain Topography in Multi-dimensions." In: *SIAM Journal on Scientific Computing* 34.6 (2012), B761–B784. ISSN: 1064-8275, 1095-7197. DOI: 10.1137/110857295. URL: http://epubs.siam.org/doi/10.1137/110857295.

[46] M. Croci, M. B. Giles, M. E. Rognes, and P. E. Farrell. "Efficient White Noise Sampling and Coupling for Multilevel Monte Carlo with Nonnested Meshes." In: *SIAM/ASA Journal on Uncertainty Quantification* 6.4 (2018), pp. 1630–1655. DOI: 10.1137/18M1175239. URL: https://epubs-siam-org.proxy.lib.duke.edu/doi/abs/10.1137/18M1175239.

[47] P. Wesseling. "Cell-Centered Multigrid for Interface Problems." In: *Journal of Computational Physics* 79.1 (1988), pp. 85–91. ISSN: 0021-9991. DOI: 10.1016/0021-9991(88)90005-8. URL: https://www.sciencedirect.com/science/article/pii/0021999188900058 (visited on 12/01/2021).

[48] P. Hemker. "On the Order of Prolongations and Restrictions in Multigrid Procedures." In: *Journal of Computational and Applied Mathematics* 32.3 (1990), pp. 423–429. ISSN: 03770427. DOI: 10.1016/0377-0427(90)90047-4. URL: https://linkinghub.elsevier.com/retrieve/pii/0377042790900474 (visited on 12/07/2021).

[49] M. Khalil and P. Wesseling. "Vertex-centered and cell-centered multigrid for interface problems." In: *Journal of Computational Physics* 98.1 (1992), pp. 1–10. ISSN: 0021-9991. DOI: 10.1016/0021-9991(92)90168-X.

[50] M. Mohr and R. Wienands. "Cell-Centred Multigrid Revisited." In: *Computing and Visualization in Science* 7.3 (2004), pp. 129–140. ISSN: 1433-0369. DOI: 10.1007/s00791-004-0137-0. URL: https://doi.org/10.1007/s00791-004-0137-0 (visited on 12/01/2021).

[51] R. Hartley. "A More Symmetrical Fourier Analysis Applied to Transmission Problems." In: *Proceedings of the IRE* 30.3 (1942), pp. 144–150. ISSN: 2162-6634. DOI: 10.1109/JRPROC.1942.234333.

[52] R. N. Bracewell. "Discrete Hartley Transform." In: *JOSA* 73.12 (1983), pp. 1832–1835. DOI: 10.1364/JOSA.73.001832.

[53] D. Bini and P. Favati. "On a Matrix Algebra Related to the Discrete Hartley Transform." In: *SIAM Journal on Matrix Analysis and Applications* 14.2 (1993), pp. 500–507. ISSN: 0895-4798. DOI: 10.1137/0614035.

[54] R. Shapiro. "Smoothing, filtering, and boundary effects." In: *Reviews of Geophysics* 8.2 (1970), pp. 359–387. ISSN: 1944-9208. DOI: 10.1029/RG008i002p00359.

[55] F. Falissard. "Genuinely multi-dimensional explicit and implicit generalized Shapiro filters for weather forecasting, computational fluid dynamics and aeroacoustics." In: *Journal of Computational Physics* 253 (2013), pp. 344–367. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2013.07.001.

[56] F. Falissard. "Uneven-order decentered Shapiro filters for boundary filtering." In: *Journal of Computational Physics* 292 (2015), pp. 168–175. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2015.03.003.

[57] A. T. Weaver and I. Mirouze. "On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation." In: *Quarterly Journal of the Royal Meteorological Society* 139.670 (2013), pp. 242–260. ISSN: 1477-870X. DOI: 10.1002/qj.1955.

[58] P. Whittle. "Stochastic processes in several dimensions." In: *Bull. Inst. Internat. Statist.* 40 (1963), pp. 974–994.

[59] P. Guttorp and T. Gneiting. "Studies in the history of probability and statistics XLIX: On the Matérn correlation family." In: *Biometrika* 93 (2006), pp. 989–995. DOI: 10.1093/biomet/93.4.989.

[60] C. Bekas, E. Kokiopoulou, and Y. Saad. "An Estimator for the Diagonal of a Matrix." In: *Applied Numerical Mathematics* 57.11-12 (2007), pp. 1214–1229. ISSN: 01689274. DOI: 10.1016/j.apnum.2007.01.003.

[61] E. Hallman, I. C. F. Ipsen, and A. K. Saibaba. "Monte Carlo Methods for Estimating the Diagonal of a Real Symmetric Matrix." In: *SIAM Journal on Matrix Analysis and Applications* 44.1 (2023), pp. 240–269. ISSN: 0895-4798. DOI: 10.1137/22M1476277.

[62] A. T. Weaver, J. Tshimanga, and A. Piacentini. "Correlation Operators Based on an Implicitly Formulated Diffusion Equation Solved with the Chebyshev Iteration." In: *Quarterly Journal of the Royal Meteorological Society* 142.694 (2016), pp. 455–471. ISSN: 1477-870X. DOI: 10.1002/qj.2664.

[63] O. Goux, S. Gürol, A. T. Weaver, Y. Diouane, and O. Guillet. "Impact of Correlated Observation Errors on the Conditioning of Variational Data Assimilation Problems." In: *Numerical Linear Algebra with Applications* 31.1 (2024), e2529. ISSN: 1099-1506. DOI: 10.1002/nla.2529. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/nla.2529.

[64] L. Debreu, E. Neveu, E. Simon, F.-X. Le Dimet, and A. Vidard. "Multigrid solvers and multigrid preconditioners for the solution of variational data assimilation problems." In: *Quarterly Journal of the Royal Meteorological Society* 142.694 (2016), pp. 515–528. ISSN: 1477-870X. DOI: 10.1002/qj.2676.

[65] P. Kumar, C. W. Oosterlee, and R. P. Dwight. "A multigrid multilevel Monte Carlo method using high-order finite-volume scheme for lognormal diffusion problems." In: *International Journal for Uncertainty Quantification* 7.1 (2017). ISSN: 2152-5080, 2152-5099. DOI: 10.1615/Int.J.UncertaintyQuantification.2016018677.

[66] P. Kumar, C. Rodrigo, F. J. Gaspar, and C. W. Oosterlee. "On Local Fourier Analysis of Multigrid Methods for PDEs with Jumping and Random Coefficients." In: *SIAM Journal on Scientific Computing* 41.3 (2019), A1385–A1413. ISSN: 1064-8275. DOI: 10.1137/18M1173769.

[67] J. Goodman and A. D. Sokal. "Multigrid Monte Carlo Method. Conceptual Foundations." In: *Physical Review D* 40.6 (1989), pp. 2035–2071. DOI: 10.1103/PhysRevD.40.2035. URL: https://link.aps.org/doi/10.1103/PhysRevD.40.2035.

[68] Y. Kazashi, E. H. Müller, and R. Scheichl. *Multigrid Monte Carlo Revisited: Theory and Bayesian Inference*. 2024. DOI: 10.48550/arXiv.2407.12149. arXiv: 2407.12149 [cs, math, stat]. URL: http://arxiv.org/abs/2407.12149.

[69] I. S. Gradshteyn and I. M. Ryshik. *Table of Integrals, Series, and Products*. 8th. Elsevier, 2015. ISBN: 978-0-12-384933-5. URL: https://linkinghub.elsevier.com/retrieve/pii/C20100648395.

[70] P. J. Davis. *Circulant Matrices: Second Edition*. 2nd. American Mathematical Society, 2012. ISBN: 978-0-8218-9165-0. URL: https://bookstore.ams.org/chel-338.