Asymptotic Analysis of Multilevel Best Linear Unbiased Estimators *

Daniel Schaden† and Elisabeth Ullmann†

Abstract. We study the computational complexity and variance of multilevel best linear unbiased estimators introduced in [21]. We specialize the results in this work to PDE-based models which are parameterized by a discretization quantity e.g. the finite element mesh size. In particular, we analyze the asymptotic complexity of sample allocation optimal best linear unbiased estimators termed SAOBs which have a minimal variance given a fixed computational budget. Since the SAOBs are constructed adaptively and hence difficult to analyze, we study a class of auxiliary estimators based on the Richardson extrapolation of the parametric model family. We provide an upper bound for the complexity of the SAOB, showing that their complexity is optimal within a certain class of linear unbiased estimators. In particular, the complexity of the SAOB is not worse than the complexity of Multilevel Monte Carlo. The theoretical results are illustrated by numerical experiments with an elliptic PDE.

Key words. Uncertainty quantification, partial differential equation, Richardson extrapolation, Monte Carlo, Multilevel Monte Carlo

AMS subject classifications. 35R60, 62J05 62F12, 65N30, 65C05

1. Introduction.

A common model in the field of uncertainty quantification (UQ) is a partial differential equation (PDE) equipped with random coefficients, and other inputs whose uncertainty is modeled by a probability distribution on a suitable function space. An important building block in UQ is the estimation of expected values of output quantities of interest linked with such random PDEs. Estimators of Monte Carlo (MC) type are often infeasible in this situation due to the high costs per sample. In the last decade *multilevel* estimators have been developed to address this problem and provide estimates with much smaller computational costs. This is achieved by *variance reduction* and by working with a family of PDE models with different resolutions or fidelities. Typically, multilevel estimators couple an expensive, high resolution PDE with cheap, low resolution PDE models. Examples for multilevel estimators are multilevel Monte Carlo (MLMC) [8, 9], multifidelity Monte Carlo (MFMC) [16, 17] and approximate control variates (ACVs) [10]. In this work we revisit the multilevel best linear unbiased estimator (BLUE) introduced in [21]. The multilevel BLUE is a linear, unbiased combination of MC estimators associated with different PDE resolutions. Importantly, the multilevel BLUE selects the *optimal* linear combination with respect to the estimator variance.

The analysis in [21] is independent of the underlying model family used for the multilevel BLUE. Now we specialize the results in [21], and assume that we work with PDE-based models with random coefficients. For discretized PDEs it is often the case that the fidelity of a model

*Submitted to the editors DATE.

Funding: The authors gratefully acknowledge the support by the Deutsche Forschungsgemeinschaft (DFG) through the International Research Training Group IGDK 1754 "Optimization and Numerical Analysis for Partial Differential Equations with Nonsmooth Structures".

†Department of Mathematics, Technical University of Munich, Boltzmannstr. 3, 85748 Garching b. München, Germany, (schaden@ma.tum.de,eullmann@ma.tum.de).

can be linked with a discretization parameter, e.g., the mesh size of a finite element space. Thus, it is natural to assume that the models in the family are parameterized by the mesh size, where a small parameter value gives accurate results which might however be expensive to compute. We make the form of the parameterization precise in the following sections of the paper.

In [21] we also introduced a sample allocation optimal estimator termed SAOB which is a special case of the multilevel BLUE. The SAOB selects so called model groups and the number of samples for each group such that the estimator variance is minimal given a fixed computational budget. The SAOB is constructed adaptively, and is implicitly defined as the solution of an optimization problem. This complicates its analysis, even if we make more assumptions on the model family we work with. As an auxiliary tool we introduce and study a class of estimators termed RE estimators. These are based on Richardson extrapolation (RE) [4, 19], and on the telescoping sum approach in multilevel Monte Carlo [8, 9]. It turns out that the classical MLMC estimator in [8, 9] is a special case of an RE estimator.

RE is a well known technique in numerical analysis. It employs linear combinations of a family of approximations to improve the accuracy of the individual approximations within the family. Perhaps the most widely known application of RE is the Romberg method [20] for numerical quadrature. In addition, RE has been used for ordinary differential equations [5], stochastic differential equations (SDEs) [23], and partial differential equations (see e.g. [1, 2, 18]). By combining RE with the MLMC complexity theory [7, 9] we provide an upper bound on the computational complexity of the SAOB showing that its asymptotic complexity is not worse than the complexity of MLMC. In addition, we show that the coefficients in the linear combination of the multilevel BLUE can be approximated by RE coefficients.

The combination of RE and multilevel estimators has been discussed in a few places in the literature. However, these are typically restricted to bias errors of a specific form and SDE discretizations. Multilevel Richardson–Romberg extrapolation has been explored already by Giles in the pioneering MLMC paper [8] where one level of RE was used. In [14] the authors combine RE and MLMC for specific discretizations of the Langevin equation. Lemaire and Pagès [12] introduce a multilevel Richardson–Romberg estimator termed ML2R for SDE discretizations where the bias error w.r.t. the discretization parameter h has the form h^{β_k} with linearly growing exponents $\beta_k = k\alpha$. An antithetic extension of the ML2R estimator is studied in [13]. In our work we study RE estimators to analyze the complexity of the SAOB. In fact, this idea grew out of numerical experiments where we observed that the coefficients of the multilevel BLUE approached RE coefficients. This highlights a novel application of Richardson extrapolation in the study of multilevel estimators and uncertainty quantification.

The main contributions of this paper are as follows: (i) a general upper bound on the complexity of the SAOB, (ii) a specific upper bound in terms of Richardson extrapolation estimators for parametric model families, (iii) a lower variance bound and asymptotic characterization of the coefficients of the multilevel BLUE estimator for parametric model families, and (iv) a complete complexity and variance analysis of the Richardson extrapolation estimators.

The remainder of this work is structured as follows. In Section 2 we review the necessary definitions and results introduced in [21], in particular, the multilevel BLUE and the SAOB. In

Section 3 we formulate our main assumption on the parametric model family, and define and analyze a Richardson extrapolation for the family. Moreover, we study the spectral properties of the model covariance matrix which will later be linked to the variance and the coefficients in the multilevel BLUE. In Section 4 we analyze the computational complexity of the SAOB. To this end we study a class of Richardson extrapolation estimators. In Section 5 we analyze the minimal variance achieved by the multilevel BLUE in terms of the discretization parameter of the model family. We also give a lower bound on the variance of the Richardson extrapolation estimators. In Section 6 we verify the theoretical results by numerical experiments. Section 7 provides concluding remarks.

2. Problem formulation and main idea.

Let Z be a scalar-valued random variable whose expectation $\mathbb{E}[Z]$ we want to estimate. We assume that exact sampling from Z is not possible. Hence we work with a family of approximations $Z_1, \dots, Z_L$, $L \in \mathbb{N}$, depending on discretization parameters $h_1, \dots, h_L$, respectively, such that for an event ω

$$Z_\ell(\omega) := Z_\omega(h_\ell). \tag{2.1}$$

For convenience we often write $Z_\omega = Z_\omega(0)$ and $h = h_1$. Typical scenarios we have in mind are finite element based PDE discretizations where h_ℓ denotes the mesh size and $Z_\omega(h_\ell)$ is an output quantity of interest which requires solving the discretized PDE with random inputs depending on the event ω. In this case, for a small discretization parameter h_ℓ the approximation of Z_ω by $Z_\ell(\omega)$ is accurate yet computationally expensive.

Let us further define expectation and covariance for the random variables $Z_1, \dots, Z_L$ as

$$\mu_\ell := \mathbb{E}[Z_\ell], \quad c_{\ell,j} := \mathrm{Cov}(Z_\ell, Z_j), \quad \text{for all } \ell, j \in \{1, \dots, L\}.$$

We write $\mu := (\mu_1, \dots, \mu_L)^T$ and define the *model covariance matrix* $C := (c_{\ell,j})_{\ell,j=1}^L$. We assume throughout this paper that the expectations and variances of $Z, Z_1, \dots, Z_L$ exist and are finite. For every subset $Q \subseteq \{1, \dots, L\}$ we define the principal submatrix of C as

$$C_{Q,Q} := (c_{\ell,j})_{\ell,j \in Q} \in \mathbb{R}^{|Q| \times |Q|}.$$

We use a similar notation for vectors $\alpha \in \mathbb{R}^L$ with $\alpha_Q := (\alpha_\ell)_{\ell \in Q} \in \mathbb{R}^{|Q|}$.

Let us introduce the notation $\widehat{\mu}_\alpha$ for unbiased estimators of $\alpha^T \mu$, i.e., estimators which satisfy $\mathbb{E}[\widehat{\mu}_\alpha] = \alpha^T \mu$. In the special case of $\alpha = e_\ell$, that is, α is the ℓ-th unit vector, we write $\widehat{\mu}_\ell$ for an unbiased estimator of $e_\ell^T \mu = \mu_\ell$.

Finally, we introduce the notation $\simeq$ defined as

$$\varphi(h) \simeq \phi(h) \quad \iff \quad \exists c_0, c_1 > 0 : c_0 \phi(h) \le \varphi(h) \le c_1 \phi(h) \tag{2.2}$$

for deterministic functions ϕ, φ depending on h. We often abbreviate the constants c_0, c_1, which do not depend on h but may depend on other parameters, simply by c. Moreover, the value of this generic constant c may change from equation to equation.

2.1. Multilevel best linear unbiased estimator.

Our goal is to obtain a variance minimal, linear, unbiased estimator for $\alpha^T \mu$ for a given vector $\alpha \in \mathbb{R}^L$. In [21] we employed a block linear model of the form

$$Y = H\mu + \eta, \tag{2.3}$$

where Y is a vector of observations, H is the design matrix modeling a linear relationship between parameters and observations, and η is a mean zero noise vector with a certain covariance. Denote the $K := 2^L - 1$ non-empty subsets of the set $\{1, \dots, L\}$ by $S^1, \dots, S^K \in 2^{\{1,\dots,L\}} \setminus \{\emptyset\}$. In particular,

$$S^j \neq S^k, \text{ for } j \neq k.$$

We call each collection of indices S^j a *model group*. For an event ω we define the corresponding model group evaluation, the covariance, restriction and prolongation matrices, respectively, as follows:

$$Z^k(\omega) := (Z_\ell(\omega))_{\ell \in S^k} \in \mathbb{R}^{|S^k|}, \qquad C^k := C_{S^k,S^k} = \mathrm{Cov}(Z^k, Z^k) \in \mathbb{R}^{|S^k|\times|S^k|},$$
$$R^k v := v_{S^k} \text{ for all } v \in \mathbb{R}^L, \qquad P^k := (R^k)^T \in \mathbb{R}^{L\times|S^k|}.$$

Now consider $m_k \in \mathbb{N}_0$ independent samples for each model group S^k, denoted by ω_i^k, $i = 1, \dots, m_k$. We also assume that ω_i^k and ω_i^j are independent for $S^k \neq S^j$. Let

$$\Psi(m) := \sum_{k=1}^K m_k P^k (C^k)^{-1} R^k \in \mathbb{R}^{L\times L}, \quad y := \sum_{k=1}^K P^k (C^k)^{-1} \sum_{i=1}^{m_k} Z^k(\omega_i^k) \in \mathbb{R}^L,$$

where $m = (m_1, \dots, m_K)^T$. Then, the best linear unbiased estimator (BLUE) $\widehat{\mu}^{\mathrm{B}}(m)$ to estimate the vector μ associated with the sample numbers in m is the solution of the normal equation

$$\Psi(m)\widehat{\mu}^{\mathrm{B}}(m) = y. \tag{2.4}$$

Under certain assumptions the scalar value $\widehat{\mu}_\alpha^{\mathrm{B}}(m) := \alpha^T \widehat{\mu}^{\mathrm{B}}(m)$ is also the BLUE for $\alpha^T\mu$ and has the variance $\mathrm{Var}(\widehat{\mu}_\alpha^{\mathrm{B}}(m)) = \alpha^T \Psi(m)^{-1}\alpha$ (see [21, Theorem 2.7]).

2.2. Sample allocation optimal estimator. Observe that the estimator $\widehat{\mu}_\alpha^{\mathrm{B}}(m)$ depends on the number of samples m_k for each model group S^k. Now we want to select m in an optimal way given a fixed computational budget. We assume costs $w_\ell > 0$ to compute a sample of Z_ℓ, $\ell = 1, \dots, L$. This incurs the cost

$$W_k := \sum_{\ell \in S^k} w_\ell$$

to evaluate all models in the group S^k. For a fixed budget $p > 0$ we then solve the *sample allocation problem*:

$$\begin{cases} \min_{m\in\mathbb{N}_0^K} \ \mathrm{Var}(\widehat{\mu}_\alpha^{\mathrm{B}}(m)) & = \alpha^T\Psi(m)^{-1}\alpha \\ \text{s.t.} \quad \mathrm{Cost}(\widehat{\mu}_\alpha^{\mathrm{B}}(m)) & = \sum_{k=1}^K m_k W_k \leq p, \\ m_k & = 0, \quad \text{if } |S^k| > \kappa. \end{cases} \tag{2.5}$$

In (2.5), the coupling number $\kappa > 0$ defines the maximal number of models that are evaluated for the same input ω. We then define the SAOB, κ estimator

$$\widehat{\mu}_\alpha^{\mathrm{SAOB},\kappa} = \alpha^T \widehat{\mu}^{\mathrm{B}}(m^*), \tag{2.6}$$

where m^* is a minimizer of (2.5). If the coupling number $\kappa = +\infty$, we drop κ in the notation above. It can be proved that the SAOB is variance minimal in the class of linear unbiased estimators with costs bounded by p.

Theorem 2.1 ([21, Theorem 3.2]). *Let the model covariance matrix C be positive definite. Then any linear unbiased estimator $\widehat{\mu}_\alpha$ that uses only evaluations of $Z_1, \ldots, Z_L$ with cost* $\mathrm{Cost}(\widehat{\mu}_\alpha) \leq p$ *satisfies*

$$\mathrm{Var}(\widehat{\mu}_\alpha) \geq \mathrm{Var}(\widehat{\mu}_\alpha^{\mathrm{SAOB}}).$$

The SAOB is difficult to analyze since it depends on a minimizer m^* of (2.5) and is constructed *adaptively*. In this paper we derive an upper bound on the asymptotic complexity of the SAOB by constructing a simpler estimator $\widehat{\mu}_\alpha$ which we can actually analyze for particular values of $\alpha = \alpha^{L,q}$. This will be done in Section 4 with the help of Richardson extrapolation [4, 19].

2.3. Main idea. Recall that RE employs linear combinations of a family of approximations with different accuracies to improve the overall accuracy. In our context we see that the BLUE $\widehat{\mu}_L^B$ defined in (2.4) (see also [21, Equ. (2.9)]) relies on the exact same idea. Indeed, by combining the definitions in [21, Equ. (2.9)], [21, Equ. (2.7)], and rearranging, it is easy to see that for suitable coefficients β_ℓ^k it holds

$$\widehat{\mu}_L^B = \sum_{k=1}^{K} \sum_{\ell \in S^k} \beta_\ell^k \frac{1}{m_k} \sum_{i=1}^{m_k} Z_\ell(\omega_i^k). \tag{2.7}$$

That is, $\widehat{\mu}_L^B$ is a linear combination of correlated Monte Carlo estimators of $\mu_1, \ldots, \mu_L$ which approximate the expectation $\mathbb{E}[Z]$ with different fidelities. Defining $\beta_\ell^k = 0$ if $\ell \notin S^k$, and recalling that $Z_\ell(\omega_i^k)$ and $Z_\ell(\omega_j^m)$ are statistically independent for $i \neq j$ or $k \neq m$, we see that the variance of $\widehat{\mu}_L^B$ has the form

$$\mathrm{Var}(\widehat{\mu}_L^B) = \sum_{k=1}^{K} \frac{1}{m_k} \mathrm{Var}\left(\sum_{\ell \in S^k} \beta_\ell^k Z_\ell \right) = \sum_{k=1}^{K} \frac{1}{m_k} (\beta^k)^T C \beta^k.$$

Thus, if either $\sum_{\ell \in S^k} \beta_\ell^k Z_\ell$ or $(\beta^k)^T C \beta^k$ are asymptotically small, then a small number of samples m_k for the model group S^k is sufficient to achieve a small estimator variance $\mathrm{Var}(\widehat{\mu}_L^B)$.

In general, assume that some vectors $\alpha^1, \ldots, \alpha^K \in \mathbb{R}^L$ span the space $\mathbb{R}^L$, that $\alpha_\ell^k = 0$ if $\ell \notin S^k$ and that the variance $(\alpha^k)^T C \alpha^k \simeq \phi^k(h)$ for some function ϕ^k, $k = 1, \ldots, K$. Then there exist weights $g_1, \ldots, g_K \in \mathbb{R}$ such that $e_L = \sum_{k=1}^{K} g_k \alpha^k$. Moreover, the estimator

$$\widehat{\mu}_L = \sum_{k=1}^{K} \sum_{\ell \in S^k} g_k \alpha_\ell^k \frac{1}{m_k} \sum_{i=1}^{m_k} Z_\ell(\omega_i^k)$$

is an unbiased estimator for μ_L with variance

$$\mathrm{Var}(\widehat{\mu}_L) \simeq \sum_{k=1}^{K} \frac{g_k^2}{m_k} \phi^k(h).$$

The cost of $\widehat{\mu}_L$ can now be studied following the same steps as in the complexity analysis of MLMC [9] if the rates $g_k^2\phi^k(h)$ w.r.t. h are known. Finally, Theorem 4.5 in Subsection 4.2 shows that the SAOB has an equally large or smaller complexity compared to $\widehat{\mu}_L$.

We remark that we carry out this analysis for a parametric model of common type where RE yields asymptotically small quantities $g_k^2\phi^k(h)$. However, the outlined strategy may be used for other types of models as long as the variances of the linear combinations $\mathrm{Var}(\sum_{\ell\in S^k}\alpha_\ell^k Z_\ell)$ are asymptotically small.

2.4. Variance of linear unbiased estimators. We also study the asymptotic behavior of a lower bound for the variance of linear unbiased multilevel estimators. We specialize the results given in [21, Sec. 4] to the setting of parametric model families of the form (2.1). In [21, Corollary 4.4] it was proved that the variance of *any* linear unbiased estimator $\widehat{\mu}_L$ for $\mathbb{E}[Z_L]$ with at most M evaluations of the high fidelity model Z_L and using only further samples of $Z_1,\dots,Z_{L-1}$ is bounded from below as follows:

$$\mathrm{Var}(\widehat{\mu}_L) \geq \gamma_{\min} := \frac{1}{M} \min_{\substack{\overline{\alpha}\in\mathbb{R}^L,\\ \overline{\alpha}_L=1}} \overline{\alpha}^T C\overline{\alpha}. \tag{2.8}$$

Moreover, a classical result in [11] shows that the minimizing vector $\overline{\alpha}^*$ and minimum $\gamma_{\min}$ satisfy

$$\overline{\alpha}^* = (-C_{Q,Q}^{-1}C_{Q,L},1)^T, \qquad \gamma_{\min} = (C_{L,L} - C_{L,Q}C_{Q,Q}^{-1}C_{Q,L})/M, \tag{2.9}$$

respectively, where $Q=\{1,\dots,L-1\}$. We will prove the convergence and state convergence rates for $\gamma_{\min}$ and $\overline{\alpha}^*$ w.r.t. the discretization parameter $h=h_1$ in Section 5.

3. Analysis of parametric model families.

In this section we formulate the main assumption of our work, an asymptotic expansion linking the models in the family (2.1). This enables us to study the RE and the spectral properties of the model covariance matrix.

3.1. Asymptotic expansion. We assume that the discretization parameters can be written as elements of a geometric series, that is, they satisfy

$$h_\ell := s^{-\ell+1}h, \quad \text{for all } \ell=1,\dots,L, \tag{3.1}$$

where $s>1$ is a fixed constant. We denote $h_1=h$ and state an assumption for RE that is used throughout the remainder of this paper.

Assumption 3.1 (Expansion in h). Let $q\in\mathbb{N}$, $0=\beta_1<\beta_2<\dots<\beta_q$, and let q random variables $(c_k)_{k=2}^{q+1}$ be given. We formulate the following assumptions:

(i) There exists a remainder term $r^q(h)\in o(h^{\beta_q}; h\to 0)$ such that for almost all events ω the realization $Z_\omega(h)$ satisfies

$$Z_\omega(h) = Z_\omega(0) + \begin{cases} c_2(\omega)r^1(h), & q=1,\\ \sum_{k=2}^q c_k(\omega)h^{\beta_k} + c_{q+1}(\omega)r^q(h), & q\geq 2.\end{cases} \tag{3.2}$$

(ii) The random variables Z and $(c_k)_{k=2}^{q+1}$ have bounded second moments with covariance matrix $K^q \in \mathbb{R}^{(q+1)\times(q+1)}$,

(iii) There exists a constant $q_{low} \leq q$ such that for the truncated expansion

$$Z_\omega(h) = Z_\omega(0) + \sum_{k=2}^{q_{low}} c_k(\omega) h^{\beta_k} + c_{low}(\omega) r^{q_{low}}(h) \tag{3.3}$$

with $r^{q_{low}} \in o(h^{\beta_{q_{low}}}; h \to 0)$ the random variables Z and $(c_k)_{k=2}^{q_{low}+1}$ have a positive definite covariance matrix $K^{q_{low}} \in \mathbb{R}^{(q_{low}+1)\times(q_{low}+1)}$.

Assumption 3.1 (i) tells us that the approximation $Z_\omega(h)$ is close to the truth $Z_\omega(0)$ for sufficiently small h. If in addition $q \geq 2$ we know the convergence rate of the remainder $r^1(h)$ term as $h \to 0$. Assumption 3.1 (ii) ensures that the variance of Z is well defined. We will see that this allows us to derive upper bounds on $\gamma_{\min}$ in (2.8). Assumption 3.1 (iii) enables us to derive lower bounds on $\gamma_{\min}$.

3.2. Definition and properties of Richardson extrapolation. For convenience we introduce the notation r^ℓ for the remainder term on the right-hand side in (3.2) truncated after $\ell \leq q$ terms. Hence it holds $r^\ell \in o(h^{\beta_\ell}; h \to 0)$. The exact form of the remainder term is not relevant as long as its order w.r.t. h is preserved. Thus we may redefine the remainder term, e.g., to drop constants.

If Assumption 3.1 (i) is satisfied, we define the RE recursively for $j \in \{1, \ldots, q\}$ as follows:

$$R_\omega^j(h) := \begin{cases} Z_\omega(h), & \text{if } j = 1, \\ (s^{\beta_j} R_\omega^{j-1}(h/s) - R_\omega^{j-1}(h))/(s^{\beta_j} - 1), & \text{if } j > 1. \end{cases} \tag{3.4}$$

Note that $R_\omega^j(h)$ depends on the event $\omega \in \Omega$ only through the random variable $Z_\omega(h)$. The parameters s^{β_j} do not depend on $\omega \in \Omega$.

It is well known that RE increases the order of individual approximations within a parametric family. We now show that this is also the case in our setting.

Lemma 3.2 (Richardson extrapolation). *Let Assumption 3.1(i) be true. Then, for every $j \in \{1, \ldots, q\}$, the RE satisfies*

$$R_\omega^j(h) = Z_\omega(0) + \sum_{k=j+1}^{q} c_k^j(\omega) h^{\beta_k} + c_{q+1}(\omega) r^q(h), \tag{3.5}$$

where the random variables c_k^j are scalar multiples of c_k,

$$c_k^j(\omega) = a_k^j c_k(\omega), \tag{3.6}$$

for some coefficients $a_k^j \neq 0$.

Proof. We use induction to show (3.5) and (3.6). The statement clearly holds for R^1. For $j \in \{2, \dots, q\}$ we insert the definition and apply the induction hypothesis. It follows

$$
\begin{aligned}
R^j_\omega(h) &= \frac{s^{\beta_j} R^{j-1}_\omega(h/s) - R^{j-1}_\omega(h)}{s^{\beta_j} - 1} \\
&= Z_\omega(0) + \frac{s^{\beta_j} \sum_{k=j}^q c_k^{j-1}(\omega)(h/s)^{\beta_k} - \sum_{k=j}^q c_k^{j-1}(\omega) h^{\beta_k}}{s^{\beta_j} - 1} + c_{q+1}(\omega)\frac{s^{\beta_j} r^q(h/s) - r^q(h)}{s^{\beta_j} - 1} \\
&= Z_\omega(0) + \sum_{k=j+1}^q \frac{s^{\beta_j - \beta_k} - 1}{s^{\beta_j} - 1} c_k^{j-1}(\omega) h^{\beta_k} + c_{q+1}(\omega)\frac{s^{\beta_j} r^q(h/s) - r^q(h)}{s^{\beta_j} - 1}.
\end{aligned}
$$

Now we define $c_k^j(\omega)$ and redefine the remainder r^q accordingly and use the induction hypothesis for $j-1$ to show the statement for j. ■

Recall that RE linearly combines parametric approximations. Looking at (3.4) it is easy to see that the coefficient vectors $\alpha^{RE,q}$ in the linear combination are recursively given as

$$
\alpha^{RE,j} := \begin{cases} e_1, & \text{if } j = 1, \\ (s^{\beta_j} D \alpha^{RE,j-1} - \alpha^{RE,j-1})/(s^{\beta_j} - 1), & \text{if } j > 1, \end{cases} \tag{3.7}
$$

where D is a downshift matrix. D corresponds to $R^j_\omega(h/s)$ using finer models compared to $R^j_\omega(h)$, and is defined as

$$
D := \begin{pmatrix} 0 & 0 \\ I_{q-1,q-1} & 0 \end{pmatrix} \in \mathbb{R}^{q \times q}. \tag{3.8}
$$

Importantly, the coefficients $\alpha^{RE,j}$ in (3.7) do not depend on the event ω but only on $\beta_1, \dots, \beta_q$ and s, which do not depend on ω. Since these recursively defined coefficients are not easy to handle for our purposes, we now express the RE directly as linear combination of the models $Z_1(\omega), \dots, Z_L(\omega)$. For $q \le L$ we define the RE with finest model Z_L of size q as follows:

$$
R^{L,q}_\omega := R^q_\omega(s^{-L+q} h) = \sum_{k=1}^q \alpha_k^{RE,q} Z_\omega(s^{-k+1} s^{-L+q} h) = \sum_{\ell=1}^L \alpha_\ell^{L,q} Z_\ell(\omega), \tag{3.9}
$$

where the coefficient $\alpha_\ell^{L,q}$ in front of $Z_\ell(\omega)$ is given by

$$
\alpha_\ell^{L,q} := \begin{cases} 0, & \text{if } \ell \le L - q, \\ \alpha_{\ell-L+q}^{RE,q}, & \text{otherwise.} \end{cases} \tag{3.10}
$$

The coefficient $\alpha_L^{L,q}$ in front of $Z_L(\omega)$ satisfies

$$
\alpha_L^{L,q} = \prod_{k=2}^q \frac{s^{\beta_k}}{s^{\beta_k} - 1} \ge 1. \tag{3.11}
$$

Let us now define the difference of two consecutive REs as follows:

$$
\Delta R^{k,q}_\omega := R^{k,\min\{k,q\}}_\omega - R^{k-1,\min\{k-1,q\}}_\omega, \tag{3.12}
$$

where we define $R_\omega^{0,0} := 0$. Analogously, we define the associated coefficients

$$\Delta\alpha^{k,q} := \alpha^{k,\min\{k,q\}} - \alpha^{k-1,\min\{k-1,q\}}. \tag{3.13}$$

Next we bound the variance of the difference of two consecutive REs by an expresssion with a certain rate w.r.t. h. To this end we define

$$\gamma^{k,q}(h) := \begin{cases} h^{2\beta_k}, & k \le q, \\ (r^q(s^{-k+q}h))^2, & k > q. \end{cases} \tag{3.14}$$

We obtain the following result.

Lemma 3.3. *Let Assumption 3.1 $(i)-(ii)$ be satisfied. Then for all $k \in \{1,\dots,L\}$, asymptotically as $h \to 0$, the variance of the difference of two consecutive REs is bounded by*

$$\mathrm{Var}(\Delta R_\omega^{k,q}) = (\Delta\alpha^{k,q})^T C \Delta\alpha^{k,q} \le c\gamma^{k,q}(h). \tag{3.15}$$

Proof. For $k = 1$ the bound in (3.15) follows immediately since $\beta_1 = 0$ and the variance of

$$\Delta R_\omega^{1,q} = R_\omega^{1,1} = R_\omega^1(s^{-1+1}h) = Z_\omega(0) + \sum_{k=2}^{q} c_k^1(\omega)(h)^{\beta_k} + c_{q+1}(\omega) r^q(h)$$

is bounded by Assumption 3.1(ii). For $k = 2,\dots,q$ we consider (3.5) and obtain

$$\mathrm{Var}(\Delta R_\omega^{k,q}) = \mathrm{Var}(R_\omega^{k,k} - R_\omega^{k-1,k-1}) = \mathrm{Var}(-c_k^{k-1}(\omega)h^{\beta_k} + o(h^{\beta_k}; h \to 0)),$$

from which we deduce (3.15). Finally, for $k > q$ we have

$$\begin{aligned}\mathrm{Var}(\Delta R_\omega^{k,q}) &= \mathrm{Var}(R_\omega^{k,q} - R_\omega^{k-1,q}) = \mathrm{Var}(c_{q+1}(\omega)[r^q(s^{-k+q}h) - r^q(s^{-k+q+1}h)]) \\ &\le c(r^q(s^{-k+q}h) - r^q(ss^{-k+q}h))^2,\end{aligned}$$

which has the same rate as $(r^q(s^{-k+q}h))^2$. This concludes the proof. ∎

3.3. Spectral properties of the model covariance matrix.

In this section we study the asymptotic behavior of the eigenvalues and eigenvectors of the model covariance matrix C in terms of the discretization parameter $h = h_1$. To this end we define the auxiliary vectors

$$b^k := (s^{(-\ell+1)\beta_k})_{\ell=1}^L \in \mathbb{R}^L \quad \text{for all } k = 1,\dots,q. \tag{3.16}$$

We will see that we can bound the inner products $v^T C v$ using the vectors in (3.16).

Lemma 3.4. *Let Assumption 3.1(i)–(ii) be true. Then the vectors $b^1,\dots,b^q$ are linearly independent. Furthermore, for every $v \in \mathbb{R}^L$ it holds*

$$v^T C v \le \lambda_{\max}(K^q) \sum_{k=1}^{q} h^{2\beta_k}(v^T b^k)^2 + o(\|v\|^2 (r^q(h))^2; h \to 0). \tag{3.17}$$

If in addition Assumption 3.1(iii) is satisfied, then it holds

$$v^T C v \ge \lambda_{\min}(K^{q_{low}}) \sum_{k=1}^{q_{low}} h^{2\beta_k}(v^T b^k)^2 + o(\|v\|^2 (r^{q_{low}}(h))^2; h \to 0). \tag{3.18}$$

Proof. The vectors b^k are linearly independent since the leading $q \times q$ principal submatrix of the matrix

$$\left(\frac{1}{s^{\beta_1}} b^1 \quad \frac{1}{s^{\beta_2}} b^2 \quad \cdots \quad \frac{1}{s^{\beta_q}} b^q \right) \in \mathbb{R}^{L \times q}$$

is a Vandermonde matrix. Now, observe that the discretization parameters satisfy

$$h_\ell^{\beta_k} = s^{(-\ell+1)\beta_k} h^{\beta_k} = b_\ell^k h^{\beta_k}. \tag{3.19}$$

We define b_j^{q+1} such that the remainder term satisfies $r^q(h_j) = h^{\beta_{q+1}} b_j^{q+1}$ for $\beta_{q+1} := \beta_q$. Combining (3.19) with Assumption 3.1 (i) and using the abbreviation $c_1 := Z$ we arrive at

$$\operatorname{Cov}(Z_i, Z_j) = \sum_{k,n=1}^{q+1} h^{\beta_k} b_i^k \operatorname{Cov}(c_k, c_n) h^{\beta_n} b_j^n.$$

For $v \in \mathbb{R}^L$ we obtain the upper bound (3.17)

$$\begin{aligned} v^T C v = \sum_{i,j=1}^{L} v_i \operatorname{Cov}(Z_i, Z_j) v_j &= \sum_{k,n=1}^{q+1} \left(h^{\beta_k} \sum_{i=1}^{L} v_i b_i^k \right) \operatorname{Cov}(c_k, c_n) \left(h^{\beta_n} \sum_{j=1}^{L} v_j b_j^n \right) \\ &\leq \lambda_{\max}(K^q) \sum_{k=1}^{q+1} h^{2\beta_k} (v^T b^k)^2. \end{aligned}$$

Now we use $h^{\beta_{q+1}} b_\ell^{q+1} = r^q(h_\ell)$ to show the upper bound

$$h^{2\beta_{q+1}} (v^T b^k)^2 \leq \|v\|^2 \sum_{\ell=1}^{L} (h^{\beta_{q+1}} b_\ell^{q+1})^2 = \|v\|^2 \sum_{\ell=1}^{L} (r^q(h_\ell))^2.$$

The lower bound (3.18) follows analogously. ∎

It is not difficult to show that, up to higher order terms, C is a sum of rank one updates of the vectors $b^1, \ldots, b^q$. That is,

$$C = \sum_{k,n=1}^{q} h^{\beta_k} h^{\beta_n} \operatorname{Cov}(c_k, c_n) b^k (b^n)^T + \text{h.o.t.}$$

We now construct an orthonormal basis (ONB) $u^1, \ldots, u^q$ using the vectors $b^1, \ldots, b^q$ by Gram–Schmidt orthogonalization, that is,

$$\begin{aligned} \operatorname{span}(b^1, \ldots, b^k) &= \operatorname{span}(u^1, \ldots, u^k), && \text{for all } k \in \{1, \ldots, q\}, \\ (u^k)^T u^\ell &= \delta_{k\ell}, && \text{for all } k, \ell \in \{1, \ldots, q\}, \end{aligned}$$

where $\delta_{k\ell}$ denotes the Kronecker Delta. We add suitable vectors $u^{q+1}, \ldots, u^L$ such that the vectors $u^1, \ldots, u^L$ form an ONB of $\mathbb{R}^L$. Then, Lemma 3.4 together with the orthogonality $b^k \perp u^\ell$ for $k < \ell$ implies

$$(u^\ell)^T C u^\ell \begin{cases} \simeq \gamma^{\ell,q}(h), & \text{if } \ell \leq q_{low}, \\ \leq c\gamma^{\ell,q}(h), & \text{otherwise}. \end{cases} \tag{3.20}$$

This allows us derive bounds on the eigenvalues of the model covariance matrix C. We denote and order the eigenvalues of C as follows:

$$\lambda_1(C) \geq \lambda_2(C) \geq \cdots \geq \lambda_L(C) \geq 0. \tag{3.21}$$

Theorem 3.5 (Eigenvalue bound for C). *Let Assumption 3.1(i)–(ii) be satisfied. Then, asymptotically as $h \to 0$, the k^{th} largest eigenvalue of C is bounded from above by*

$$\lambda_k(C) \leq c\gamma^{k,q}(h). \tag{3.22}$$

If in addition Assumption 3.1 (iii) is true, then for $k \leq q_{low}$ we have the lower bound

$$\lambda_k(C) \geq c\gamma^{k,q}(h), \tag{3.23}$$

asymptotically as $h \to 0$.

Proof. The proof is presented in Appendix A. ■

Theorem 3.5 shows that the model covariance matrix C is ill-conditioned in the limit $h \to 0$. This is an intuitive result, since for small values of h the random variables $Z_1, \ldots, Z_L$ have nearly identical realizations and are thus highly correlated. In particular, the condition number $\kappa_2(C)$ explodes if Z_L can be well approximated by linear combinations of $Z_1, \ldots, Z_{L-1}$. We make this statement precise in the next corollary.

Corollary 3.6. *Let Assumption 3.1 (i)–(ii) be satisfied. Then the condition number of the model covariance matrix C is bounded from below by*

$$\kappa_2(C) \geq c(\gamma^{L,q}(h))^{-1},$$

asymptotically as $h \to 0$. If in addition Assumption 3.1 (iii) holds with $q_{low} = q = L$, then

$$\kappa_2(C) \simeq c(\gamma^{L,q}(h))^{-1},$$

asymptotically as $h \to 0$.

We are further able to quantify the convergence of the eigenvectors of C. To this end we define the gap $\beta^{\text{gap}(k)}$ as follows:

$$\beta^{\text{gap}(k)} := \begin{cases} \beta_2 - \beta_1, & \text{if } k = 1, \\ \min\{\beta_{k+1} - \beta_k, \beta_k - \beta_{k-1}\}, & \text{if } 1 < k < q_{low}, \\ \beta_{q_{low}} - \beta_{q_{low}-1}, & \text{if } k \geq q_{low}. \end{cases}$$

Theorem 3.7 (Convergence of eigenvectors of C). *Let Assumption 3.1(i)–(iii) be satisfied and let $q_{low} = q = L$. Then the eigenvectors v^k corresponding to the eigenvalue $\lambda_k(C)$ satisfy*

$$\text{dist}(v^k, \text{span}(u^k)) \leq c\|v^k\|h^{\beta^{\text{gap}(k)}}, \tag{3.24}$$

asymptotically as $h \to 0$.

Proof. The proof is given in Appendix B. ■

Finally, we relate the vectors $u^1, \dots, u^L$ to the RE coefficients in (3.13).

Lemma 3.8. *Let Assumption 3.1 (i)–(iii) be true. Then it holds*

$$u^k \in \begin{cases} \mathrm{span}(\Delta\alpha^{k,q}, \dots, \Delta\alpha^{L,q}), & \text{if } k \le q_{low}, \\ \mathrm{span}(\Delta\alpha^{q_{low},q}, \dots, \Delta\alpha^{L,q}), & \text{if } k > q_{low}. \end{cases} \tag{3.25}$$

Proof. We apply (3.18) and (3.15) to conclude $b^k \perp \Delta\alpha^\ell$ for $k < \ell$ and $k \le q_{low}$. With $\mathrm{span}(b^1, \dots, b^{k-1}) = \mathrm{span}(u^1, \dots, u^{k-1})$ we conclude

$$\mathrm{span}(u^1, \dots, u^{k-1}) \perp \mathrm{span}(\Delta\alpha^{k,q}, \dots, \Delta\alpha^{L,q}).$$

This proves (3.25) for $k \le q_{low}$. The case $k > q_{low}$ also follows from these arguments. ■

Let us comment on the relation (3.25). The vector u^k corresponds to a linear combination of RE differences starting with index k. That is, there exist deterministic coefficients $(a_\ell)_{\ell=1}^L$, such that

$$\sum_{\ell=1}^{L} u_\ell^k Z_\ell(\omega) = \sum_{\ell=k}^{L} a_\ell \Delta R_\omega^{\ell,q}.$$

Furthermore, for $q_{low} = q = L$ the vectors $u^L, \dots, u^2$ can also be obtained by Gram–Schmidt orthogonalization of the difference vectors $\Delta\alpha^{L,q}, \dots, \Delta\alpha^{2,q}$.

4. Complexity analysis.

We return to the complexity analysis of the SAOB in (2.6). To begin, we define an estimator termed Richardson extrapolation estimator, and analyze its properties. We will then show that the RE estimator allows us to bound the complexity of the SAOB.

4.1. Richardson extrapolation estimator.

Let us define the RE estimator with finest model index L and coupling number κ, $2 \le \kappa \le L$, as follows:

$$\widehat{\mu}^{\mathrm{RE},\kappa}_{\alpha^{L,\kappa-1}} := \sum_{\ell=1}^{L} \widehat{E}^{MC}\left(\Delta R_\omega^{\ell,\kappa-1}\right) = \sum_{\ell=1}^{L}\sum_{j=1}^{L} \Delta\alpha_j^{\ell,\kappa-1}\left(\frac{1}{N_\ell}\sum_{i=1}^{N_\ell} Z_j(\omega_i^\ell)\right). \tag{4.1}$$

In (4.1) we combine L statistically independent Monte Carlo estimators using N_ℓ samples each. Hence, two events ω_i^ℓ and ω_j^k are independent if the subscript or superscript differ. The terms $\Delta\alpha^{L,\kappa-1}$ in (4.1) are defined in (3.13) as the coefficients in the difference of two consecutive REs $\Delta R^{L,\kappa-1}$ on level L of size $\kappa - 1$. The coupling number κ denotes the maximum number of models that are coupled, that is, at most κ models out of $Z_1, \dots, Z_L$ are evaluated with the same input event ω. At the first glance the sum over j in (4.1) seems to require us to evaluate all models on all events $(\omega_i^\ell)_{i=1}^{N_\ell}$. However, from (3.13) we conclude that at most κ entries of the vector $\Delta\alpha^{\ell,\kappa-1}$ are non-zero. This justifies the name coupling number for κ.

For $\kappa = 2$ the RE estimator in (4.1) coincides with the Multilevel Monte Carlo (MLMC) estimator [8, 9]. To see this, consider the telescoping sum relation

$$R_\omega^{L,1} = \sum_{\ell=2}^{L} (R_\omega^{\ell,\min\{\ell,1\}} - R_\omega^{\ell-1,\min\{\ell-1,1\}}) + R_\omega^{1,1} = \sum_{\ell=1}^{L} \Delta R_\omega^{\ell,1} = \sum_{\ell=1}^{L}\sum_{j=1}^{L} \Delta\alpha_j^{\ell,1} Z_j(\omega). \tag{4.2}$$

It is now easy to show that $\alpha^{\ell,1} = e_\ell$ is the ℓ-th unit vector, $\Delta\alpha^{\ell,1} = e_\ell - e_{\ell-1}$, $\ell \geq 2$, and $\Delta\alpha^{1,1} = e_1$. Hence it follows

$$R_\omega^{L,1} = \sum_{\ell=2}^{L} (Z_\ell(\omega) - Z_{\ell-1}(\omega)) + Z_1(\omega) = Z_L(\omega),$$

and

$$\widehat{\mu}_{\alpha^{L,1}}^{\mathrm{RE},2} = \sum_{\ell=2}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} (Z_\ell(\omega_i^\ell) - Z_{\ell-1}(\omega_i^\ell)) + \frac{1}{N_1}\sum_{i=1}^{N_1} Z_1(\omega_i^1)$$

is indeed the MLMC estimator. We remark that for the coupling number $\kappa = 3$, an RE estimator as in (4.1) is discussed briefly in [8, p. 612] for the special case $s = M$ and exponents $\beta_1 = 0$, $\beta_2 = 1$, $\beta_3 = 2$ in Assumption 3.1. The goal in [8] is to eliminate the leading order bias term.

Note that a telescoping sum relation as in (4.2) holds also for $\kappa > 2$. Hence most results in this section follow straightforwardly from the MLMC theory [9]. First, we prove that the RE estimator is an unbiased estimator for $(\alpha^{L,\kappa-1})^T\mu$ and analyze its bias rate. Note that classical multilevel estimators such as MLMC are unbiased estimators for μ_L. In general, $\alpha^{L,\kappa-1} \neq e_L$ and thus the RE estimator is not unbiased for μ_L.

Lemma 4.1 (RE estimator bias). *Let Assumption 3.1 (i) and (ii) be true and let $\kappa \leq q+1$. Then, the estimator $\widehat{\mu}_{\alpha^{L,\kappa-1}}^{\mathrm{RE},\kappa}$ is an unbiased estimator for $\mathbb{E}[R^{L,\kappa-1}] = (\alpha^{L,\kappa-1})^T\mu$. The bias w.r.t. Z is bounded from above by*

$$|\mathbb{E}[\widehat{\mu}_{\alpha^{L,\kappa-1}}^{\mathrm{RE},\kappa}] - \mathbb{E}[Z]| \leq c r^{\kappa-1}(s^{-L+\kappa-1}h) \leq c \begin{cases} s^{-L\beta_{\kappa-1}} h^{\beta_{\kappa-1}}, & \kappa = q+1, \\ s^{-L\beta_\kappa} h^{\beta_\kappa}, & \kappa \leq q, \end{cases} \tag{4.3}$$

asymptotically as $h \to 0$.

Proof. The telescoping sum in the definition of $\widehat{\mu}_{\alpha^{L,\kappa-1}}^{\mathrm{RE},\kappa}$ in (4.1) and $R_\omega^{0,0} = 0$ show that $\widehat{\mu}_{\alpha^{L,\kappa-1}}^{\mathrm{RE},\kappa}$ is an unbiased estimator for $\mathbb{E}[R^{L,\kappa-1}]$. Next, we revisit Lemma 3.2 and set $j = \kappa - 1$ in (3.5) to obtain

$$R_\omega^{L,\kappa-1} - Z_\omega(0) = R_\omega^{\kappa-1}(s^{-L+\kappa-1}h) - Z_\omega(0) = c_\kappa(\omega) r^{\kappa-1}(s^{-L+\kappa-1}h),$$

where c_κ is a random variable with bounded first and second moment. Now we take the expectation and use the fact that $r^{\kappa-1}$ is deterministic to obtain

$$\mathbb{E}[\widehat{\mu}_{\alpha^{L,\kappa-1}}^{\mathrm{RE},\kappa}] - \mathbb{E}[Z] = \mathbb{E}[R_\omega^{L,\kappa-1}] - \mathbb{E}[Z_\omega(0)] \leq \mathbb{E}[c_\kappa] r^{\kappa-1}(s^{-L+\kappa-1}h).$$

The bound in (4.3) for $\kappa = q+1$ follows since $r^{\kappa-1} \in o(h^{\beta_{\kappa-1}}; h \to 0)$. If $\kappa \leq q$ we have the improved rate $r^{\kappa-1}(h) = h^{\beta_\kappa}$ in (3.2). ∎

Next we investigate the variance of the RE estimator.

Lemma 4.2 (RE estimator variance). *Let Assumption 3.1 (i) and (ii) be true and let $\kappa \le q+1$. Then it holds*

$$\mathrm{Var}(\Delta R_\omega^{\ell,\kappa-1}) \le c\gamma^{\ell,\kappa-1}(h) \le c \begin{cases} s^{-2\ell\beta_{\kappa-1}} h^{2\beta_{\min\{\ell,\kappa-1\}}}, & \kappa = q+1, \\ s^{-2\ell\beta_\kappa} h^{2\beta_{\min\{\ell,\kappa\}}}, & \kappa \le q. \end{cases} \quad (4.4)$$

asymptotically as $h \to 0$.

Proof. The first inequality in (4.4) follows from Lemma 3.3. For $\kappa = q+1$ and $\ell \le \kappa - 1$ we use $s^{2\ell\beta_\kappa} \le s^{2(\kappa-1)\beta_\kappa}$ to estimate

$$\gamma^{\ell,\kappa-1}(h) = s^{2\ell\beta_\kappa} s^{-2\ell\beta_\kappa} h^{2\beta_\ell} \le s^{2(\kappa-1)\beta_\kappa} s^{-2\ell\beta_\kappa} h^{2\beta_\ell}.$$

For $\ell \ge \kappa$ we use $r^{\kappa-1}(h) \in o(h^{\beta_{\kappa-1}}; h \to 0)$ to get

$$\gamma^{\ell,\kappa-1}(h) = (r^{\kappa-1}(s^{-\ell+\kappa-1}h))^2 \le cs^{2(\kappa-1)\beta_{\kappa-1}} s^{-2\ell\beta_{\kappa-1}} h^{2\beta_{\kappa-1}}.$$

In case of $\kappa \le q$ we look at the definition of $\gamma^{\ell,\kappa-1}$ and use $r^{\kappa-1}(h) = h^{\beta_\kappa}$ to get

$$\begin{aligned} \gamma^{\ell,\kappa-1}(h) &= \begin{cases} h^{2\beta_\ell}, & \ell \le \kappa-1, \\ (r^{\kappa-1}(s^{-\ell+\kappa-1}h))^2, & \ell > \kappa - 1. \end{cases} \\ &= \begin{cases} s^{-2\ell\beta_\kappa+2\ell\beta_\kappa} h^{2\beta_\ell}, & \ell \le \kappa-1, \\ s^{-2\ell\beta_\kappa+2(\kappa-1)\beta_\kappa} h^{2\beta_\kappa}, & \ell > \kappa-1. \end{cases} \\ &\le s^{2(\kappa-1)\beta_\kappa} s^{-2\ell\beta_\kappa} h^{2\beta_{\min\{\ell,\kappa\}}}. \end{aligned}$$

This concludes the proof. ∎

Finally, we study the computational complexity of the RE estimator for $\mathbb{E}[Z]$ up to a given tolerance ε^2 w.r.t. the mean square error (MSE). The next theorem is a variant of the MLMC Complexity Theorem (see [9, Theorem 2.1] or [7, Theorem 1]).

Theorem 4.3 (RE Complexity Theorem). *Assume that there exist positive constants β^{Bias}, β^{Var} and β^{Cost}, such that the following statements hold:*

(M1) $|\mathbb{E}[\widehat{\mu}^{\mathrm{RE},\kappa}_{\alpha^{L,\kappa-1}}] - \mathbb{E}[Z]| \le c_1 s^{-L\beta^{\mathrm{Bias}}}$,
(M2) $\mathrm{Var}(\Delta R^{\ell,\kappa-1}) \le c_2 s^{-\ell\beta^{\mathrm{Var}}}$,
(M3) $\mathrm{Cost}(\Delta R^{\ell,\kappa-1}) \le c_3 s^{\ell\beta^{\mathrm{Cost}}}$,

where the constants $c_1, c_2, c_3 > 0$ are independent of ℓ. Then there exists a final level $L > 0$ and a sequence of samples $\{N_\ell\}_{\ell=1}^L$ to achieve a MSE error bound

$$\mathbb{E}[|\widehat{\mu}^{\mathrm{RE},\kappa}_{\alpha^{L,\kappa-1}} - \mathbb{E}[Z]|^2] \le \varepsilon^2$$

with costs bounded by

$$\mathrm{Cost}(\widehat{\mu}^{\mathrm{RE},\kappa}_{\alpha^{L,\kappa-1}}) \le c\varepsilon^{-\beta^{\mathrm{Cost}}/\beta^{\mathrm{Bias}}} + c \begin{cases} \varepsilon^{-2}, & \text{if } \beta^{\mathrm{Cost}} < \beta^{\mathrm{Var}}, \\ \varepsilon^{-2}(\log\varepsilon)^2, & \text{if } \beta^{\mathrm{Cost}} = \beta^{\mathrm{Var}}, \\ \varepsilon^{-2-\frac{\beta^{\mathrm{Cost}}-\beta^{\mathrm{Var}}}{\beta^{\mathrm{Bias}}}}, & \text{if } \beta^{\mathrm{Cost}} > \beta^{\mathrm{Var}}. \end{cases} \quad (4.5)$$

Proof. The proof is analogous to the proof for [7, Theorem 1]. Note that we do not use the assumption $\beta^{\text{Bias}} \geq \min\{\beta^{\text{Var}}, \beta^{\text{Cost}}\}/2$. Thus we do not explicitly bound the term $\varepsilon^{-\beta^{\text{Cost}}/\beta^{\text{Bias}}}$. ∎

To finish the complexity analysis we state the rates for β^{Bias}, β^{Var}, and β^{Cost} in (4.5).

Corollary 4.4 (RE complexity bound). *Assume that the costs to obtain a sample Z_ℓ on level ℓ scale as*

$$\text{Cost}(Z_\ell) \leq c s^{\ell \beta^{\text{Cost}}}, \tag{4.6}$$

and that Assumption 3.1 (i)–(ii) are satisfied. Then Theorem 4.3 holds with rates

$$\beta^{\text{Bias}} = \begin{cases} \beta_{\kappa-1}, & \kappa = q+1, \\ \beta_\kappa, & \kappa \leq q, \end{cases} \quad \textit{and} \quad \beta^{\text{Var}} = \begin{cases} 2\beta_{\kappa-1}, & \kappa = q+1, \\ 2\beta_\kappa, & \kappa \leq q. \end{cases}$$

Proof. Assumption (M1) in Theorem 4.3 follows from Lemma 4.1, and Assumption (M2) follows from Lemma 4.2. Finally, the cost assumption (M3) can be proved using the geometric cost increase in (4.6), and by observing that ∎

$$\text{Cost}(\Delta R^{\ell,\kappa-1}) \leq \sum_{j=1}^{\ell} \text{Cost}(Z_j) \leq c\,\text{Cost}(Z_\ell).$$

4.2. SAOB. Now we return to the SAOB in (2.6). We prove the main result in this section: To achieve a MSE bounded by ε^2 where $\varepsilon > 0$ is a given tolerance, the estimator $\widehat{\mu}^{\text{SAOB}}_{\alpha^{L,\kappa-1}}$ requires a computational complexity which is not larger than the complexity of the RE estimator $\widehat{\mu}^{\text{RE},\kappa}_{\alpha^{L,\kappa-1}}$. To this end we make a general observation on the complexity of the SAOB.

Theorem 4.5 (Complexity bound for SAOB). *Assume that the model covariance matrix C is positive definite. Let $\widehat{\mu}_\alpha$ be a linear unbiased estimator for $\alpha^T \mu$ using only the models $Z_1, \ldots, Z_L$ such that a MSE of ε^2 is achieved with a cost bounded by $\phi(\varepsilon^2)$ for some function ϕ. That is, it holds*

$$\mathbb{E}[|\widehat{\mu}_\alpha - \mathbb{E}[Z]|^2] \leq \varepsilon^2, \quad \textit{and} \quad \text{Cost}(\widehat{\mu}_\alpha) \leq \phi(\varepsilon^2).$$

Then, the estimator $\widehat{\mu}^{\text{SAOB}}_\alpha$ achieves a MSE bounded by ε^2 with the same cost bound. That is,

$$\mathbb{E}[|\widehat{\mu}^{\text{SAOB}}_\alpha - \mathbb{E}[Z]|^2] \leq \varepsilon^2, \quad \textit{and} \quad \text{Cost}(\widehat{\mu}^{\text{SAOB}}_\alpha) \leq \phi(\varepsilon^2).$$

Proof. We use the bias variance decomposition and the required bound on the MSE to show

$$\mathbb{E}[(\widehat{\mu}_\alpha - \mathbb{E}[Z])^2] = (\alpha^T \mu - \mathbb{E}[Z])^2 + \text{Var}(\widehat{\mu}_\alpha) \leq \varepsilon^2.$$

Now, the estimator $\widehat{\mu}^{\text{SAOB}}_\alpha$ has the same bias as $\widehat{\mu}_\alpha$. Therefore we only compare the variance of $\widehat{\mu}^{\text{SAOB}}_\alpha$ and $\widehat{\mu}_\alpha$, respectively, as unbiased estimators of $\alpha^T \mu$. If we choose $p = \text{Cost}(\widehat{\mu}_\alpha)$ in Theorem 2.1, then – by construction – the estimator $\widehat{\mu}^{\text{SAOB}}_\alpha$ has an equally large or smaller variance than $\widehat{\mu}_\alpha$. This shows the claim. ∎

Next, observe that Theorem 4.5 allows us to state a complexity bound for the SAOB in the case $\alpha = \alpha^{L,\kappa-1}$.

Corollary 4.6 (SAOB complexity bound). *Assume that the costs to obtain a sample of Z_ℓ on level ℓ scale as follows:*

$$\mathrm{Cost}(Z_\ell) \le c s^{\ell\beta^{\mathrm{Cost}}},$$

and that Assumption 3.1 (i)–(ii) are satisfied. Furthermore, assume that C is positive definite. Then the complexity bound in Theorem 4.3 holds with the rates

$$\beta^{\mathrm{Bias}} = \begin{cases} \beta_{\kappa-1}, & \kappa = q+1, \\ \beta_\kappa, & \kappa \le q, \end{cases}, \quad \beta^{\mathrm{Var}} = \begin{cases} 2\beta_{\kappa-1}, & \kappa = q+1, \\ 2\beta_\kappa, & \kappa \le q. \end{cases}$$

if we replace the RE estimator $\widehat{\mu}^{\mathrm{RE},\kappa}_{\alpha^{L,\kappa-1}}$ by the estimator $\widehat{\mu}^{\mathrm{SAOB}}_{\alpha^{L,\kappa-1}}$.

Proof. The assumptions of Corollary 4.4 are satisfied and thus Theorem 4.3 holds with the rates stated for $\widehat{\mu}^{\mathrm{RE},\kappa}_{\alpha^{L,\kappa-1}}$. Now Theorem 4.5 tells us that the bounds in (4.5) are upper bounds for the complexity of $\widehat{\mu}^{\mathrm{SAOB}}_{\alpha^{L,\kappa-1}}$. ∎

Remark 4.7. Recall that for $\kappa = 2$ it holds $\alpha^{L,\kappa-1} = e^L$, and that the RE estimator $\widehat{\mu}^{\mathrm{RE},2}_{\alpha^{L,\kappa-1}}$ coincides with the MLMC estimator $\widehat{\mu}^{\mathrm{MLMC}}_{\alpha^{L,\kappa-1}}$. Hence, Theorem 4.5 tells us that the complexity of the SAOB $\widehat{\mu}^{\mathrm{SAOB}}_L$ and $\widehat{\mu}^{\mathrm{SAOB},\kappa}_L$ ($\kappa \ge 2$) is asymptotically not worse than the complexity of the MLMC estimator $\widehat{\mu}^{\mathrm{MLMC}}_L$.

Remark 4.8. The treatment of the general case $\alpha \in \mathbb{R}^L$ in Corollary 4.6 is possible, however, it requires us to work with a *family* of SAOBs defined by a suitable *sequence* of vectors $(\alpha^n)_{n\in\mathbb{N}}$. The reason is that a bias of arbitrary order ε cannot be reached by working with a fixed vector $\alpha \in \mathbb{R}^L$; the vector α must depend on the given tolerance ε. Note that this detail is also present in the MLMC Complexity Theorem in [7, 9] where the vector $\alpha = e_L \in \mathbb{R}^L$ depends on the final level L and thus on the tolerance ε.

4.3. Weighted Richardson extrapolation estimator. Note that the complexity bound for the SAOB $\widehat{\mu}^{\mathrm{SAOB}}_{\alpha^{L,\kappa-1}}$ in Corollary 4.6 is based on the variance reduction rate $\beta^{\mathrm{Var}} = 2\beta_\kappa$ associated with the RE estimator $\widehat{\mu}^{\mathrm{RE},\kappa}_{\alpha^{L,\kappa-1}}$ as proved in Corollary 4.4. If we can construct an estimator with the same bias and cost but a larger variance reduction rate, this will also improve the complexity bound for the SAOB. We now explain briefly how a larger rate β^{Var} might be achieved.

Observe that the RE estimator in (4.1) is based on the expansion

$$\alpha^{L,\kappa-1} = \sum_{\ell=1}^{L} \Delta\alpha^{\ell,\kappa-1}.$$

The use of $\Delta\alpha^{L,\kappa-1}$ in this expansion yields a variance reduction rate $\beta^{\mathrm{Var}} = 2\beta_\kappa$. Alternatively, let $g_1, \dots, g_L$ be weights such that

$$\alpha^{L,\kappa-1} = \sum_{\ell=1}^{L} g_\ell \Delta\alpha^{\ell,q}, \tag{4.7}$$

where we use the fact that the vectors $\Delta\alpha^{1,q},\dots,\Delta\alpha^{L,q}$ form a basis of $\mathbb{R}^L$. This property follows from $\Delta\alpha_j^{\ell,q}=0$ for $j>\ell$ and $\Delta\alpha_\ell^{\ell,q}=\alpha_\ell^{\ell,q}\geq 1$. Importantly, the expansion in (4.7) uses the terms $\Delta\alpha^{L,q}$ which enables a larger rate $\beta^{\mathrm{Var}}=2\beta_q$.

Using the representation in (4.7) we define a *weighted* version of the RE estimator:

$$\widehat{\mu}_{\alpha^{L,\kappa-1}}=\sum_{\ell=1}^{L}g_\ell\widehat{E}^{MC}[\Delta R_\omega^{\ell,q-1}]=\sum_{\ell=1}^{L}\sum_{j=1}^{L}g_\ell\Delta\alpha_j^{\ell,q-1}\left(\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}Z_j(\omega_i^\ell)\right). \tag{4.8}$$

By construction, $\widehat{\mu}_{\alpha^{L,\kappa-1}}$ in (4.8) is an unbiased estimator for $\mu_{\alpha^{L,\kappa-1}}$. Lemma 4.2 then tells us that the variance on every level satisfies

$$\mathrm{Var}(g_\ell\Delta R^{\ell,q-1})\leq cg_\ell^2\,\mathrm{Var}(\Delta R^{\ell,q-1})\leq cg_\ell^2 s^{-2\ell\beta_q}.$$

Now, if the coefficients g_ℓ are uniformly bounded w.r.t. ℓ and independently of L, then Assumption (M2) in Theorem 4.3 holds with the rate $\beta^{\mathrm{Var}}=2\beta_q$ for the weighted RE estimator in (4.8).

5. Bounding the variance.

We now return to the lower bound $\gamma_{\min}$ in (2.8). The asymptotic behavior of $\gamma_{\min}$ can be characterized as follows.

Theorem 5.1 (Convergence rate of $\gamma_{\min}$). *Let Assumption 3.1(i)–(ii) be satisfied. Then, asymptotically as $h\to 0$, $\gamma_{\min}$ is bounded from above by*

$$\gamma_{\min}\leq c\gamma^{L,q}(h). \tag{5.1}$$

If in addition Assumption 3.1(iii) holds with $q_{low}=q=L$, then

$$\gamma_{\min}\simeq c\gamma^{L,q}(h). \tag{5.2}$$

In this case, the unique minimizer $\overline{\alpha}^*$ in (2.8) satisfies the estimate

$$\left\|\overline{\alpha}^*-u^L/u_L^L\right\|=\left\|\overline{\alpha}^*-\Delta\alpha^{L,q}/\Delta\alpha_L^{L,q}\right\|\leq ch^{\beta_q-\beta_{q-1}}. \tag{5.3}$$

Proof. The upper bound in (5.1) for $q\leq L$ follows from Lemma 3.3 by choosing

$$\overline{\alpha}=\Delta\alpha^{L,q}/\Delta\alpha_L^{L,q}$$

in the minimization problem (2.8). Note that $\Delta\alpha_L^{L,q}=\alpha_L^{L,q}\neq 0$ by (3.11).

Now we consider $q_{low}=q=L$ and rewrite (2.8) in terms of the eigenvectors $(v^\ell)_{\ell=1}^L$ of C:

$$\min_{w\in\mathbb{R}^L}\sum_{\ell=1}^{L}w_\ell^2\lambda_\ell,\quad\text{such that}\quad\overline{\alpha}_L=\sum_{\ell=1}^{L}w_\ell v_L^\ell=1.$$

Using standard techniques from convex optimization [3, Section 5.5.3.] the minimizer is

$$w_\ell=\left(\sum_{k=1}^{L}\frac{\lambda_\ell}{\lambda_k}(v_L^k)^2\right)^{-1}v_L^\ell\quad\text{for all }\ell=1,\dots,L.$$

We insert the minimizer into the objective function to arrive at (5.2). Indeed, it holds

$$\sum_{\ell=1}^{L} w_\ell^2 \lambda_\ell = \sum_{\ell=1}^{L} \left(\sum_{k=1}^{L} \frac{\sqrt{\lambda_\ell}}{\lambda_k} (v_L^k)^2 \right)^{-2} (v_L^\ell)^2 \simeq \lambda_L/(v_L^L)^2,$$

where we used Theorem 3.5 and Theorem 3.7 for $\lim_{h\to 0}(v_L^L)^2 = (u_L^L)^2$. Note that we have $u^L \in \operatorname{span}(\Delta\alpha^{L,q})$ by (3.25) and $\Delta\alpha_L^{L,q} = \alpha_L^{L,q} \neq 0$ by (3.11) showing that $(v_L^L)^2 > c > 0$ for h small enough. Finally, it holds

$$\operatorname{dist}(\overline{\alpha}^*, \operatorname{span}(v^L))^2 = \sum_{\ell=1}^{L-1} w_\ell^2 \leq c \sum_{\ell=1}^{L-1} h^{4\beta_q - 4\beta_\ell} \leq c h^{4\beta_q - 4\beta_{q-1}}.$$

Combining this with Theorem 3.7 gives the desired error bound (5.3). ■

The estimate in (5.3) tells us that the minimizing vector $\overline{\alpha}^*$ can be well approximated in terms of RE coefficients. This demonstrates again the key role that Richardson extrapolation plays in the analysis of the multilevel BLUE (2.4). Notably, the approximation $\Delta\alpha^{L,q}/\Delta\alpha_L^{L,q}$ is computable if the rates β_k in Assumption 3.1 are known and does not require solves with the ill-conditioned covariance matrix $C_{Q,Q}$ as in (2.9). Finally, observe that the approximation error of $\overline{\alpha}^*$ in (5.3) is of order $h^{\beta_q - \beta_{q-1}}$. Replacing the optimal value $\overline{\alpha}^*$ in (2.9) by the approximation u^L/u_L^L gives a variance bound of the exact same order w.r.t. h as for $\gamma_{\min}$. Indeed, asymptotically as $h \to 0$ it holds

$$(u^L)^T C u^L/(u_L^L)^2 \simeq h^{2\beta_q} \simeq \gamma_{\min}.$$

As a by-product we can also analyze the variance of the RE estimator defined in (4.1). We have the following lower bound.

Corollary 5.2 (Variance bound for the RE estimator). *Let Assumption 3.1 (i)–(ii) be true and let $L = q = \kappa - 1$. Then the variance of the RE estimator $\widehat{\mu}_{\alpha^{L,\kappa-1}}^{\mathrm{RE},\kappa}$ is bounded from below by*

$$\operatorname{Var}(\widehat{\mu}_{\alpha^{L,\kappa-1}}^{\mathrm{RE},\kappa}) \geq \frac{1}{N_L} (\Delta\alpha^{L,\kappa-1})^T C \Delta\alpha^{L,\kappa-1} \geq (\alpha_L^{L,\kappa-1})^2 \gamma_{\min}. \tag{5.4}$$

Proof. The proof is analogous to the proof of [21, Corollary 5.2]. The first inequality in (5.4) follows from the definition of $\widehat{\mu}_{\alpha^{L,\kappa-1}}^{\mathrm{RE},\kappa}$ in (4.1), and by simply dropping the variance terms for the levels $\ell = 1, \ldots, L-1$. For the second inequality observe that ■

$$\frac{1}{N_L} (\Delta\alpha^{L,\kappa-1})^T C \Delta\alpha^{L,\kappa-1} \geq \frac{1}{N_L} (\Delta\alpha_L^{L,\kappa-1})^2 \min_{\substack{\overline{\alpha} \in \mathbb{R}^L, \\ \overline{\alpha}_L = 1}} (\overline{\alpha})^T C \overline{\alpha} = (\alpha_L^{L,\kappa-1})^2 \gamma_{\min}.$$

Note that the lower bound on the right-hand side of (5.4) cannot be smaller than $\gamma_{\min}$. Indeed, in (3.11) we see that $\alpha_L^{L,\kappa-1} \geq 1$. This is consistent with the theory since the RE estimator $\widehat{\mu}_{\alpha^{L,\kappa-1}}^{\mathrm{RE},\kappa}$ is in general not variance minimal.

Model	#Nodes	w_ℓ	$w_\ell/w_{\ell-1}$	$\mathrm{Var}(Z_\ell)$	$\mathrm{Bias}(Z_\ell)$
Z_1	81	0.0020s	–	$8.30 \cdot 10^0$	$1.90 \cdot 10^{-1}$
Z_2	289	0.0026s	1.33	$4.16 \cdot 10^1$	$7.05 \cdot 10^{-2}$
Z_3	1089	0.0052s	2.00	$7.10 \cdot 10^1$	$1.98 \cdot 10^{-2}$
Z_4	4225	0.0199s	3.81	$8.16 \cdot 10^1$	$4.88 \cdot 10^{-3}$
Z_5	16641	0.0820s	4.13	$8.46 \cdot 10^1$	$9.84 \cdot 10^{-4}$
Z_6	66049	0.3585s	4.37	$8.53 \cdot 10^1$	$2.71 \cdot 10^{-4}$

Table 1: PDE example ($\rho = 0.5$, $\sigma^2 = 3$): The number of finite element basis functions (#Nodes) characterizes the discretization. w_ℓ is the average time to compute a realization of Z_ℓ, and the cost increase factor is denoted by $w_\ell/w_{\ell-1}$. The last two columns contain the estimates for the variance and bias.

Finally, we remark that a result analogous to Corollary 5.2 can also be proved for the weighted RE estimator in (4.8) where we simply have to multiply the bounds in (5.4) by the square of the last weight g_L^2.

6. Numerical experiments.

In this section we verify the main results of this paper numerically. We revisit the example in [21, Sec. 6.3]. The random variable Z is the domain average

$$Z(\omega) := \frac{1}{|D_{obs}|} \int_{D_{obs}} y(x, \omega) dx, \tag{6.1}$$

where $D_{obs} := (3/4, 7/8) \times (7/8, 1)$ is a subset of the unit square domain $D := (0, 1)^2$. The function y solves the elliptic PDE such that for almost all ω it holds

$$\begin{aligned} -\operatorname{div}(a(x,\omega)\nabla y(x,\omega)) &= 1, \quad x \in D, \\ y(x,\omega) &= 0, \quad x \in \partial D. \end{aligned} \tag{6.2}$$

The diffusion coefficient is $a(x, \omega) := \exp(b(x, \omega))$, where b is a mean zero Gaussian random field with Whittle–Matérn covariance function [22] with smoothness parameter $\nu = 3/2$, variance $\sigma^2 = 3$, and correlation length $\rho = 0.5$. To discretize Z we use a standard piecewise linear finite element (FE) discretization of (6.2) with $L = 6$ different levels of accuracy. We obtain Z_ℓ by uniform refinement of the previous mesh, starting with a regular grid for Z_1. This gives $s = 2$ in the cost assumption (M3) in Theorem 4.3. We define the bias as follows:

$$\mathrm{Bias}(Z_\ell) := |\mathbb{E}[Z_\ell] - \mathbb{E}[Z]|. \tag{6.3}$$

The discretization data is summarized in Table 1. We use 10^5 pilot samples to estimate the Bias in (6.3), the model covariance matrix C, and the average cost w_ℓ to compute a sample of Z_ℓ. The cost for the pilot samples is not included in the subsequent analysis. The fourth column in Table 1 suggests the cost increase rate $\beta^{\mathrm{Cost}} = 2$ which corresponds to a four-fold cost increase as expected in 2D space.

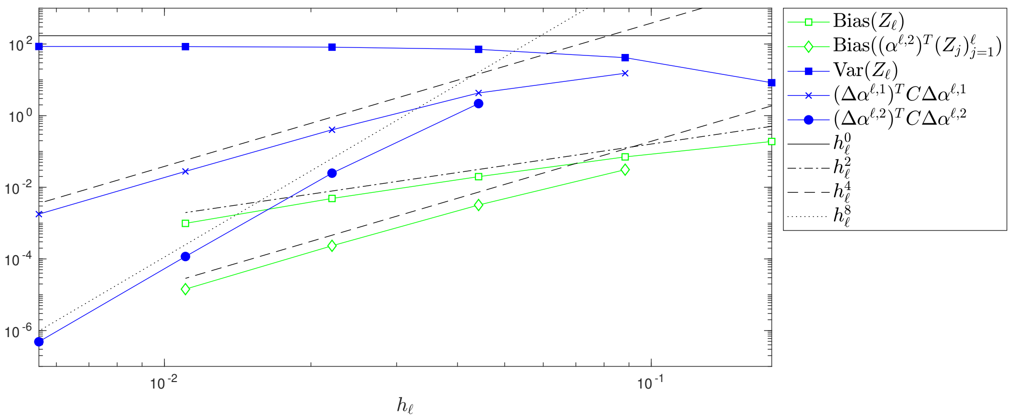

Figure 1: PDE example ($\rho = 0.5$, $\sigma^2 = 3$): Biases and variances for several quantities plotted against the mesh size h_ℓ.

Since the random field b is smooth and we integrate over y in (6.1), we expect a RE expansion of the form

$$Z_\omega(h) = Z_\omega(0) + c_2(\omega)h^2 + c_3(\omega)h^4 + o(h^4) \tag{6.4}$$

for $h \to 0$. The parameter h denotes the maximal diameter of an element in the FE mesh. If the expansion in (6.4) holds, we have $\beta_1 = 0$, $\beta_2 = 2$, $\beta_3 = 4$ and $q = 3$. Moreover, if the level ℓ is sufficiently large we expect to observe the following rates w.r.t. the mesh size h_ℓ according to the theoretical results in Subsection 4.1:

$$\begin{aligned} \mathrm{Var}(Z_\ell) &\simeq h_\ell^0, \quad \mathrm{Bias}(Z_\ell) \simeq h_\ell^2, \\ (\Delta\alpha^{\ell,1})^T C \Delta\alpha^{\ell,1} &\simeq h_\ell^4, \quad \mathrm{Bias}((\alpha^{\ell,2})^T (Z_j)_{j=1}^{\ell}) \simeq h_\ell^4, \\ (\Delta\alpha^{\ell,2})^T C \Delta\alpha^{\ell,2} &\simeq h_\ell^8. \end{aligned} \tag{6.5}$$

Note that Assumption 3.1 is not necessary to obtain Theorem 4.3. It is sufficient to provide bias and variance rates as in Corollary 4.4. Therefore we assume in the remainder of this section that the rates in (6.5) hold. The numerical verification of these rates is provided in Figure 1. A formal proof of the rates in (6.5) or the expansion in (6.4) is beyond the scope of this paper.

6.1. Complexity of estimators for μ_L. In this section we study the computational complexity of various unbiased estimators to approximate $\mu_L = \mathbb{E}[Z_L]$. These estimators have to be distinguished from the RE estimators which are in general *not* unbiased for μ_L. Precisely, we consider Monte Carlo $\widehat{\mu}_L^{\mathrm{MC}}$, Multilevel Monte Carlo $\widehat{\mu}_L^{\mathrm{MLMC}}$, Multifidelity Monte Carlo $\widehat{\mu}_L^{\mathrm{MFMC}}$, and the SAOB $\widehat{\mu}_L^{\mathrm{SAOB},2}$, $\widehat{\mu}_L^{\mathrm{SAOB},3}$, and $\widehat{\mu}_L^{\mathrm{SAOB}}$ with coupling numbers $\kappa = 2$, $\kappa = 3$ and $\kappa = +\infty$, respectively.

Consider an estimator $\widehat{\mu}_\ell$ which is an unbiased estimator for $\mu_\ell = \mathbb{E}[Z_\ell]$. We measure the accuracy of $\widehat{\mu}_\ell$ by the MSE

$$\mathbb{E}[(\widehat{\mu}_\ell - \mathbb{E}[Z])^2] = \mathrm{Bias}(Z_\ell)^2 + \mathrm{Var}(\widehat{\mu}_\ell). \tag{6.6}$$

For each level $\ell = 1, \dots, L$ we ensure that $\mathrm{Bias}(Z_\ell)^2 = \mathrm{Var}(\widehat{\mu}_\ell)$ resulting in a MSE equal to $2 \cdot \mathrm{Bias}(Z_\ell)^2$. Let $\varepsilon > 0$ denote a given tolerance. To achieve a MSE $\leq \varepsilon^2$ in (6.6) all estimators in this section require a computational cost bounded by a term of the form

$$\mathrm{Cost}(\widehat{\mu}_\ell) \leq \phi(\varepsilon) = \phi^{\mathrm{Round}}(\varepsilon) + \phi^{\mathrm{Var}}(\varepsilon), \tag{6.7}$$

where ϕ^{Round} is the complexity induced by rounding the (optimal) number of samples N_ℓ to the nearest integer, and ϕ^{Var} is the complexity due to the variance of the estimator. Given the bias, variance and cost rates β^{Bias}, β^{Var} and β^{Cost}, respectively, it can be shown that for all estimators in this section it holds

$$\phi^{\mathrm{Round}}(\varepsilon) = c\varepsilon^{-\beta^{\mathrm{Cost}}/\beta^{\mathrm{Bias}}} \tag{6.8}$$

and

$$\phi^{\mathrm{Var}}(\varepsilon) = c \begin{cases} \varepsilon^{-2}, & \text{if } \beta^{\mathrm{Cost}} < \beta^{\mathrm{Var}}, \\ \varepsilon^{-2}(\log \varepsilon)^2, & \text{if } \beta^{\mathrm{Cost}} = \beta^{\mathrm{Var}}, \\ \varepsilon^{-2-(\beta^{\mathrm{Cost}}-\beta^{\mathrm{Var}})/\beta^{\mathrm{Bias}}}, & \text{if } \beta^{\mathrm{Cost}} > \beta^{\mathrm{Var}}. \end{cases} \tag{6.9}$$

Indeed, for MLMC the complexity bounds in (6.7)–(6.9) can be proved by following the proof of the MLMC Complexity Theorem [7, Theorem 1]. Within this proof it is clear that the term $\varepsilon^{-\beta^{\mathrm{Cost}}/\beta^{\mathrm{Bias}}}$ is introduced by rounding up the (optimal) number of samples. For Multifidelity Monte Carlo it has been proved in [15] that the complexity bounds are the same as for MLMC. For Monte Carlo, observe that it always holds $\beta^{\mathrm{Cost}} > \beta^{\mathrm{Var}} = 0$. Hence we obtain the complexity

$$\mathrm{Cost}(\widehat{\mu}_L^{\mathrm{MC}}) \leq c\varepsilon^{-\beta^{\mathrm{Cost}}/\beta^{\mathrm{Bias}}} + c\varepsilon^{-2-\beta^{\mathrm{Cost}}/\beta^{\mathrm{Bias}}} \leq c\varepsilon^{-2-\beta^{\mathrm{Cost}}/\beta^{\mathrm{Bias}}}$$

which is consistent with the literature (see e.g. [7]). Finally, for the SAOB-type estimators the complexity bounds (6.7)–(6.9) follow from Theorem 4.5 and the complexity bounds for the RE estimator in Theorem 4.3. Note that for complexity bounds of the form (6.7)–(6.9) the minimal complexity is always of order ε^{-2} due to the term ϕ^{Var}. The cost due to rounding ϕ^{Round} is the same for all estimators.

We record the *theoretically proven* complexity of the estimators in Table 2. In the fourth column of Table 2 we work with the true cost rate $\beta^{\mathrm{Cost}} = 2$ as suggested in Table 1. We observe that in this case all estimators have the optimal complexity ε^{-2} except the standard Monte Carlo estimator with complexity ε^{-3}. These theoretical complexity bounds are verified numerically in the left panel of Figure 2 where we plot the *computed* costs. We generated this plot by first computing the optimal fractional number of samples and then ceiling the number of samples.

Next we discuss the theoretical basis for the complexity bounds in the fourth column of Table 2 in detail. The bias rate $\beta^{\mathrm{Bias}} = 2$ follows from $\mathrm{Bias}(Z_\ell) \simeq h_\ell^2$ in (6.5) and is

Estimator			True ($\beta^{\text{Cost}} = 2$)	Manufactured ($\beta^{\text{Cost}} = 6$)	
	β^{Bias}	β^{Var}	$\phi(\varepsilon)$	$\phi^{\text{Var}}(\varepsilon)$	$\phi(\varepsilon)$
$\widehat{\mu}_L^{\text{MC}}$	2	0	ε^{-3}	ε^{-5}	ε^{-5}
$\widehat{\mu}_L^{\text{MLMC}}$	2	4	ε^{-2}	ε^{-3}	ε^{-3}
$\widehat{\mu}_L^{\text{MFMC}}$	2	4	ε^{-2}	ε^{-3}	ε^{-3}
$\widehat{\mu}_L^{\text{SAOB},2}$	2	4	ε^{-2}	ε^{-3}	ε^{-3}
$\widehat{\mu}_L^{\text{SAOB},3}$	2	4	ε^{-2}	ε^{-3}	ε^{-3}
$\widehat{\mu}_L^{\text{SAOB}}$	2	4	ε^{-2}	ε^{-3}	ε^{-3}

Table 2: PDE example ($\rho = 0.5$, $\sigma^2 = 3$): Bias, variance, and cost rates together with the resulting *theoretical* complexity bounds in (6.7)–(6.9) for various estimators of μ_L. The given complexity is needed to achieve a RMSE of order ε.

the same for all estimators. The important difference is the variance rate β^{Var}. The rate $(\Delta\alpha^{\ell,1})^T C \Delta\alpha^{\ell,1} \simeq h_\ell^4$ in (6.5) lets us conclude $\beta^{\text{Var}} = 4$ for the RE estimator $\widehat{\mu}_{\alpha^{\ell,\kappa-1}}^{\text{RE},\kappa}$ with coupling number $\kappa = 2$. Since for $\kappa = 2$ we have $\alpha^{L,\kappa-1} = e_L$, the complexity of the estimator $\widehat{\mu}_L^{\text{SAOB},2}$ is bounded by the complexity of the RE estimator $\widehat{\mu}_{\alpha^{\ell,1}}^{\text{RE},2} = \widehat{\mu}_L^{\text{RE},2}$ (see Corollary 4.6). Also, for $\kappa = 2$ the RE estimator $\widehat{\mu}_L^{\text{RE},2}$ coincides with the MLMC estimator as proved in Subsection 4.1. Hence $\widehat{\mu}_L^{\text{MLMC}}$ (and $\widehat{\mu}_L^{\text{MFMC}}$) have the same complexity bound as the estimator $\widehat{\mu}_L^{\text{SAOB},2}$. Next, we observe that the rate $(\Delta\alpha^{\ell,2})^T C \Delta\alpha^{\ell,2} \simeq h_\ell^8$ in (6.5) gives $\beta^{\text{Var}} = 8$ for the RE estimator $\widehat{\mu}_{\alpha^{\ell,2}}^{\text{RE},3}$ with $\kappa = 3$. Unfortunately, for $\kappa = 3$ we cannot invoke Corollary 4.6 since $\alpha^{L,2} \neq e_L$. We can however argue that the estimator $\widehat{\mu}_L^{\text{SAOB},3}$ is the variance minimal estimator that couples at most $\kappa = 3$ models with the same bias as $\widehat{\mu}_L^{\text{SAOB},2}$. Hence the complexity of $\widehat{\mu}_L^{\text{SAOB},3}$ is equal to or smaller than the cost of $\widehat{\mu}_L^{\text{SAOB},2}$. Since $\widehat{\mu}_L^{\text{SAOB},2}$ has already the minimal complexity of order ε^{-2} the asymptotic complexity of $\widehat{\mu}_L^{\text{SAOB},3}$ is the same as the asymptotic complexity of $\widehat{\mu}_L^{\text{SAOB},2}$. The analogous argument can be used to show the complexity of the estimator $\widehat{\mu}_L^{\text{SAOB}}$. Next we want to verify the worst case complexity stated in the last line of (6.9). Thus we assume increased *manufactured* costs of the form

$$\text{Cost}(Z_\ell) = 10^{-6} \cdot 2^{6\ell}, \quad \beta^{\text{Cost}} = 6. \tag{6.10}$$

Again, the *theoretically proven* complexity $\phi^{\text{Var}}(\varepsilon)$ to achieve a RMSE of order ε is recorded in the fifth column in Table 2. In the sixth column we give the total theoretical complexity bound including the cost of rounding. The *computed* complexity rates are shown in the right panel of Figure 2. We observe that the theoretical cost bound of order ε^{-3} in Table 2 is not achieved for the SAOB $\widehat{\mu}_L^{\text{SAOB},3}$ and $\widehat{\mu}_L^{\text{SAOB}}$. In fact, for these estimators we observe the better (and optimal) complexity of order ε^{-2}. Note that in theory the rate ε^{-2} can only be observed if the cost for the rounding is neglected, since $\phi^{\text{Round}}(\varepsilon) = c\varepsilon^{-3}$. However, in practice an improved rate of ε^{-2} may still be observed on coarse levels if the variance is relatively large compared to the bias. We further examine this rounding effect for a similar problem in Subsection 6.2. Finally, we remark that the complexity $\phi^{\text{Var}}(\varepsilon) = c\varepsilon^{-2}$ for $\widehat{\mu}_L^{\text{SAOB},3}$ and $\widehat{\mu}_L^{\text{SAOB}}$ might be

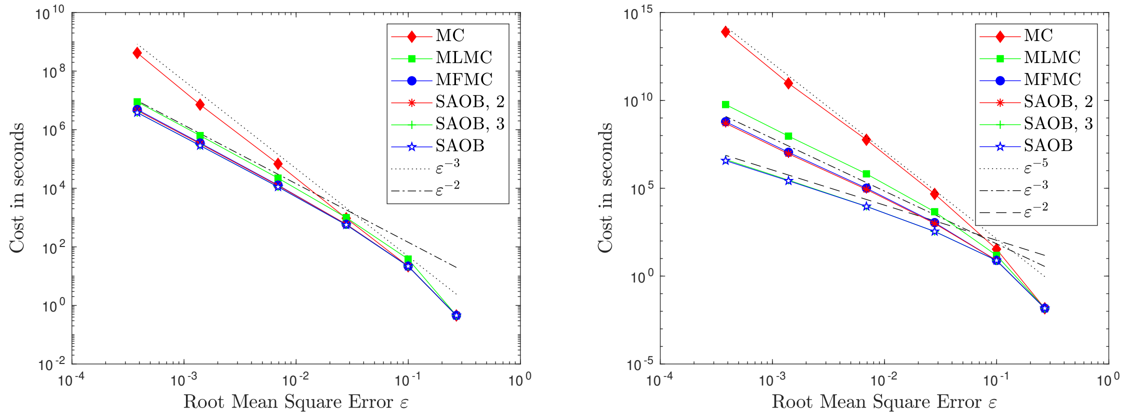

Figure 2: PDE example ($\rho = 0.5$, $\sigma^2 = 3$) with true cost rate $\beta^{\text{Cost}} = 2$ (left image) and manufactured cost rate $\beta^{\text{Cost}} = 6$ (right image). We show the computed complexity (including ceiling the number of samples) for different estimators.

proved by using the weighted RE estimator in Subsection 4.3. This weighted estimator has the same bias as the RE estimator $\widehat{\mu}_L^{\text{RE},2}$ for $\kappa = 2$ and a possibly improved variance rate $\beta^{\text{Var}} = 2\beta_q = 2\beta_3 = 8$ in our setting. The proof is beyond the scope of this paper.

6.2. The effect of rounding. For this experiment we want to showcase the effect of rounding the number of samples. Note that we did not yet observe an increased complexity by rounding in the right panel of Figure 2. To see the effect, we decrease the variance of the QoI by decreasing the variance of b to $\sigma^2 = 1$. This ensures that we require fewer samples of the high fidelity model. We work again with the manufactured cost rate $\beta^{\text{Cost}} = 6$ in (6.10). Looking at the fifth and sixth column in Table 2 we expect increased costs

$$\text{Cost}(\widehat{\mu}_L^{\text{SAOB},3}) \simeq \varepsilon^{-3}, \quad \text{Cost}(\widehat{\mu}_L^{\text{SAOB}}) \simeq \varepsilon^{-3},$$

if we round the number samples. The computed costs with rounding are now displayed in the right panel of Figure 3 and verify this claim. Moreover, we see in the left panel of Figure 3 that the $\widehat{\mu}_L^{\text{SAOB},3}$ and $\widehat{\mu}_L^{\text{SAOB}}$ have the optimal complexity of order ε^{-2} without rounding. Again, this rate is better than the theoretically proven rate of order ε^{-3} recorded in Table 2. We observe that the cost of the estimators $\widehat{\mu}_L^{\text{SAOB},3}$ and $\widehat{\mu}_L^{\text{SAOB}}$ increases by ≈ 6 times simply due to rounding. Thus, in scenarios where the complexity is dominated by the costs for the rounding, the complexity advantage of the SAOB-type estimators is reduced.

6.3. Complexity of RE estimators. Finally, we study the complexity of the RE estimators. Recall that these are not unbiased estimators for μ_L and have in general a larger bias rate (see Subsection 4.1). Let $\alpha = \alpha^{L,2}$ and consider the estimators $\widehat{\mu}_{\alpha^{L,2}}^{\text{SAOB}}$, $\widehat{\mu}_{\alpha^{L,2}}^{\text{SAOB},3}$, $\widehat{\mu}_{\alpha^{L,2}}^{\text{SAOB},2}$ and $\widehat{\mu}_{\alpha^{L,2}}^{\text{RE},3}$. We return to the setting of Subsection 6.1 with variance $\sigma^2 = 3$ for the log-diffusion coefficient b.

The rate $(\Delta\alpha^{\ell,2})^T C \Delta\alpha^{\ell,2} \simeq h_\ell^8$ In (6.5) implies $\beta^{\text{Var}} = 8$ for the RE estimator $\widehat{\mu}_{\alpha^{\ell,2}}^{\text{RE},3}$ with

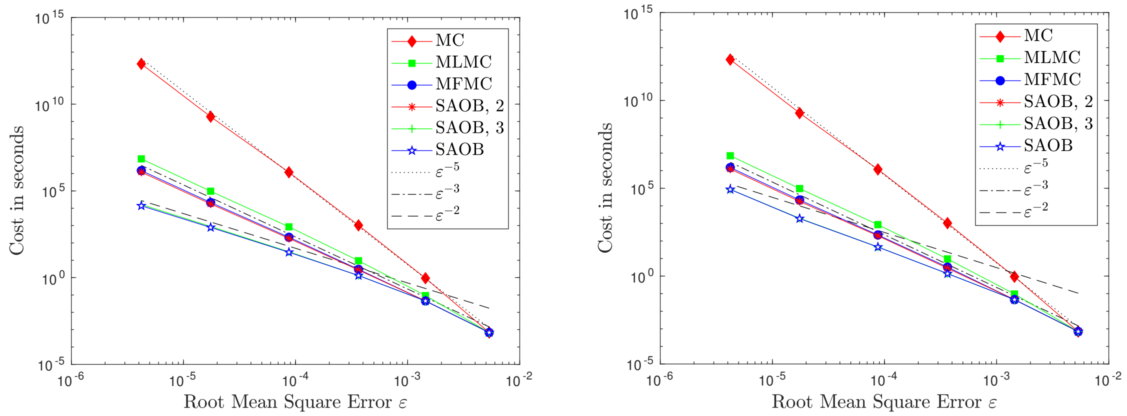

Figure 3: PDE example ($\rho = 0.5$, $\sigma^2 = 1$) with manufactured cost rate $\beta^{\text{cost}} = 6$. The left image shows the complexity $\phi^{\text{Var}}(\varepsilon)$ without the cost for rounding the number of samples. The right images shows the complexity $\phi(\varepsilon) = \phi^{\text{Var}}(\varepsilon) + \phi^{\text{Round}}(\varepsilon)$ if we ceil the number of samples.

$\kappa = 3$. Moreover, from $\text{Bias}((\alpha^{\ell,2})^T(Z_j)_{j=1}^{\ell}) \simeq h_\ell^4$ we conclude $\beta^{\text{Bias}} = 4$ for the RE estimator with $\kappa = 3$. Finally, the rate $(\Delta\alpha^{\ell,1})^T C \Delta\alpha^{\ell,1} \simeq h_\ell^4$ in (6.5) implies $\beta^{\text{Var}} = 4$ for the RE estimator $\widehat{\mu}^{\text{RE},2}_{\alpha^{\ell,1}}$ with coupling number $\kappa = 2$.

We work again with the manufactured cost rate $\beta^{\text{Cost}} = 6$. For $\kappa = 3$ with $\beta^{\text{Var}} = 8$ it holds $\beta^{\text{Cost}} < \beta^{\text{Var}}$. Thus we expect the optimal complexity of order ε^{-2} for $\widehat{\mu}^{\text{SAOB}}_{\alpha^{L,2}}$, $\widehat{\mu}^{\text{SAOB},3}_{\alpha^{L,2}}$, and $\widehat{\mu}^{\text{RE},3}_{\alpha^{L,2}}$. For $\kappa = 2$ with $\beta^{\text{Var}} = 4$ it holds $\beta^{\text{Cost}} > \beta^{\text{Var}}$. Hence for the estimator $\widehat{\mu}^{\text{SAOB},2}_{\alpha^{L,2}}$ we expect a complexity bound of the order $\varepsilon^{-2.5}$. Note that this improves the complexity bound ε^{-3} of the estimator $\widehat{\mu}^{\text{SAOB},2}_{L}$ due to the improved bias rate. All theoretical complexity bounds are confirmed numerically in Figure 4.

7. Conclusions. In this paper we studied the asymptotic properties of multilevel best linear unbiased estimators for the expectation of scalar-valued, PDE-based random outputs. The main tool of our analysis is an expansion of the random output in terms of a discretization parameter, linking a family of models associated with the output. By using Richardson extrapolation we obtained convergence results for the eigenvalues and eigenvectors of the model covariance matrix. We further proved the convergence of the minimal estimator variance in the infinite low fidelity data limit. The model family is characterized by finite element mesh sizes. However, our analysis can be adapted to other model families, e.g. spectral elements or reduced basis models, if we are able to derive the asymptotic spectral properties of the model covariance matrix. In addition, we analyzed the complexity of a certain Richardson extrapolation estimator using multilevel Monte Carlo theory. This allowed us to bound the asymptotic complexity of the SAOB which is not worse than the complexity of MLMC. Numerical experiments with a PDE-based output in 2D space suggest that a Richardson extrapolation type of expansion for the quantity of interest holds both in terms of the bias and variance. If we use the true sample cost, then the SAOB, MLMC and MFMC estimators give the opti-

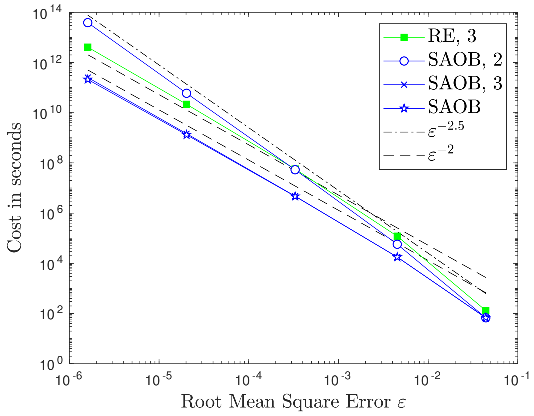

Figure 4: PDE example ($\rho = 0.5$, $\sigma^2 = 3$) with manufactured cost rate $\beta^{\text{Cost}} = 6$. We plot the computed complexity (including ceiling) for the RE estimator and SAOB-type estimators which are unbiased estimators for $(\alpha^{\ell,2})^T\mu$.

mal complexity with respect to the mean square error. For a manufactured, increased cost, the SAOB and Richardson extrapolation estimators have a smaller complexity compared to MLMC and MFMC. However, it remains an open research question whether the SAOB has an asymptotically smaller complexity than the Richardson extrapolation estimator.

Acknowledgements. The authors thank Michael Ulbrich for the finite element code that was used to solve the PDE and to sample from the mean zero Gaussian random field with Matern 3/2 covariance in Section 6.

Appendix A. Proof of Theorem 3.5. The Courant-Fischer min-max representation [6, Theorem 1.24] states

$$\lambda_k(C) = \min_V \left\{ \max \left\{ v^T C v \,|\, v \in V,\ \|v\| = 1 \right\} \,|\, \dim(V) = L - k + 1 \right\}.$$

We choose $V = \operatorname{span}(u^k, \dots, u^L)$ to obtain an upper bound on the outer minimization. Moreover, we use Lemma 3.4 to bound the inner maximization. We arrive at

$$\lambda_k(C) \le \max \left\{ v^T C v \,|\, v \in V,\ \|v\| = 1 \right\} \le c \sum_{j=k}^{q} h^{2\beta_j} ((v^*)^T b^j)^2 + o(\|v^*\|^2 (r^q(h))^2; h \to 0),$$

where we used $v^T b^j = 0$ for $j < k$ and the fact that $v^* \in V$ is a maximizer. The upper bound (3.22) for $k \le q$ now follows from $\|v^*\| = 1$. The result for $k > q$ can be obtained by restriction of the models to $Q = \{k - q + 1, \dots, L\}$, since

$$v^T C v = v_Q^T C_{Q,Q} v_Q, \quad \text{for all } v \text{ such that } v_\ell = 0 \text{ for } \ell \notin Q.$$

Therefore, we use Lemma 3.4 with $C_{Q,Q}$ instead of C and obtain (3.22) with $s^{-k+q}h$ instead of h for the discretization constant. We conclude the improved upper bound $(r^q(s^{-k+q}h))^2 =$

$\gamma^{k,q}(h)$. For the lower bound (3.23) we follow the proof of [6, Theorem 1.24] and deduce

$$\lambda_k(C) = \max_V \left\{ \min \left\{ v^T C v \,|\, v \in V,\ \|v\| = 1 \right\} \mid \dim(V) = k \right\}.$$

Now let $v \in V = \operatorname{span}(u^1, \ldots, u^k) = \operatorname{span}(b^1, \ldots, b^k)$ to obtain from Lemma 3.4

$$v^T C v \geq c \sum_{j=1}^{k} h^{2\beta_j} a_j^2 + o(\|v\|^2 h^{2\beta_k}; h \to 0)$$

for some vector a with $a_j = v^T b^j$. A change of coordinates $Bv = a$ shows

$$\min_{\|v\| \geq 1} \sum_{j=1}^{k} h^{2\beta_j} (v^T b^j)^2 = \min_{\|B^{-1} a\| \geq 1} \sum_{j=1}^{k} h^{2\beta_j} a_j^2 \geq \min_{\|B^{-1}\| \|a\| \geq 1} \sum_{j=1}^{k} h^{2\beta_j} a_j^2.$$

However, the right-hand side satisfies

$$\sum_{j=1}^{k} h^{2\beta_j} a_j^2 = a^T \operatorname{diag}(h^{2\beta_1}, \ldots, h^{2\beta_k}) a,$$

where the smallest eigenvalue of the matrix is equal to $h^{2\beta_k}$. This shows (3.23), since $\gamma^{k,q}(h) = h^{2\beta_k}$ for $k \leq q_{low} \leq q$.

Appendix B. Proof of Theorem 3.7.

Let v^k be the eigenvectors of C corresponding to λ_k with $\|v^k\| = 1$. Since $v^1, \ldots, v^L$ form an ONB we get

$$\operatorname{dist}(v^k, \operatorname{span}(u^k))^2 = \sum_{\ell=1, \ell \neq k}^{L} (a_\ell^k)^2 = 1 - (a_k^k)^2 \tag{B.1}$$

with $v^k = \sum_{\ell=1}^{L} a_\ell^k u^\ell$. We now use (3.20) to show that for any k and ℓ

$$(a_\ell^k)^2 = |(v^k)^T u^\ell|^2 \leq \lambda_k^{-1} \|C^{1/2} u^\ell\|^2 \leq c h^{2\beta_\ell - 2\beta_k}. \tag{B.2}$$

We use the orthogonality of v^k and v^ℓ for $k \neq \ell$ to get

$$(a_\ell^\ell)^2 (a_\ell^k)^2 = |(v^k)^T (v^\ell - a_\ell^\ell u^\ell)|^2 \leq \operatorname{dist}(v^\ell, \operatorname{span}(u^\ell))^2. \tag{B.3}$$

Inserting both (B.2) and (B.3) into (B.1) shows

$$1 - (a_k^k)^2 = \operatorname{dist}(v^k, \operatorname{span}(u^k))^2 \leq \sum_{\ell=1}^{k-1} \frac{1}{(a_\ell^\ell)^2} \operatorname{dist}(v^\ell, \operatorname{span}(u^\ell))^2 + c \sum_{\ell=k+1}^{L} h^{2\beta_\ell - 2\beta_k}.$$

Now use induction over $k \in \{1, \ldots, L\}$ to show $(a_k^k)^2 \to 1$ and $\operatorname{dist}(v^\ell, \operatorname{span}(u^\ell)) \to 0$ for $h \to 0$, thus we may assume $(a_k^k)^2 > c > 0$. We now strengthen (B.3) and use (B.2)

$$\begin{aligned} |(v^k)^T (v^\ell - a_\ell^\ell u^\ell)|^2 &\leq c |(v^k - a_k^k u^k)^T (v^\ell - a_\ell^\ell u^\ell)|^2 + c (a_k^k)^2 |(u^k)^T v^\ell|^2 \\ &\leq c \operatorname{dist}(v^k, \operatorname{span}(u^k))^2 \operatorname{dist}(v^\ell, \operatorname{span}(u^\ell))^2 + c h^{2\beta_k - 2\beta_\ell}. \end{aligned} \tag{B.4}$$

We combine this with (B.1) and the result can be deduced from induction over k and from

$$\left(1 - c\sum_{\ell=1}^{k-1} \mathrm{dist}(v^\ell, \mathrm{span}(u^\ell))^2\right) \mathrm{dist}(v^k, \mathrm{span}(u^k))^2 \le \sum_{\ell=1}^{k-1} \frac{1}{(a_\ell^\ell)^2} c h^{2\beta_k - 2\beta_\ell} + c \sum_{\ell=k+1}^{L} h^{2\beta_\ell - 2\beta_k}.$$

REFERENCES

[1] M. Asadzadeh, A. H. Schatz, and W. Wendland, *A new approach to Richardson extrapolation in the finite element method for second order elliptic problems*, Mathematics of Computation, 78 (2009), pp. 1951–1973, https://doi.org/10.1090/S0025-5718-09-02241-8.

[2] H. Blum, Q. Lin, and R. Rannacher, *Asymptotic error expansion and Richardson extrapolation for linear finite elements*, Numer. Math., 49 (1986), pp. 11–37, https://doi.org/10.1007/BF01389427.

[3] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004, https://doi.org/10.1017/CBO9780511804441.

[4] C. Brezinski, *Convergence acceleration during the 20th century*, vol. 122, 2000, pp. 1–21, https://doi.org/10.1016/S0377-0427(00)00360-5. Numerical analysis 2000, Vol. II: Interpolation and extrapolation.

[5] R. Bulirsch and J. Stoer, *Numerical treatment of ordinary differential equations by extrapolation methods*, Numerische Mathematik, 8 (1966), pp. 1–13, https://doi.org/10.1007/BF02165234.

[6] F. Chatelin, *Spectral approximation of linear operators*, Academic Press, New York, 1983.

[7] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup, *Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients*, Comput. Vis. Sci., 14 (2011), pp. 3–15, https://doi.org/10.1007/s00791-011-0160-x.

[8] M. B. Giles, *Multi-level Monte Carlo path simulation*, Operations Research, 56 (2008), pp. 607–617, https://doi.org/10.1287/opre.1070.0496.

[9] M. B. Giles, *Multilevel Monte Carlo methods*, Acta Numerica, 24 (2015), pp. 259–328, https://doi.org/10.1017/S096249291500001X.

[10] A. A. Gorodetsky, G. Geraci, M. Eldred, and J. D. Jakeman, *A Generalized Framework for Approximate Control Variates*, Preprint, (2018), https://arxiv.org/pdf/1811.04988.pdf.

[11] S. S. Lavenberg and P. Welch, *A Perspective on the Use of Control Variables to Increase the Efficiency of Monte Carlo Simulations*, Management Science, 27 (1981), pp. 322–335, https://doi.org/10.1287/mnsc.27.3.322.

[12] V. Lemaire and G. Pagès, *Multilevel Richardson-Romberg extrapolation*, Bernoulli, 20 (2017), pp. 1029–1067, https://doi.org/10.3150/16-BEJ822.

[13] C. Mbaye, G. Pagès, and F. Vrins, *An antithetic approach of multilevel Richardson-Romberg extrapolation estimator for multidimensional SDEs*, in Numerical analysis and its applications, vol. 10187 of Lecture Notes in Comput. Sci., Springer, Cham, 2017, pp. 482–491.

[14] E. H. Müller, R. Scheichl, and T. Shardlow, *Improving multilevel Monte Carlo for stochastic differential equations with application to the Langevin equation*, Proc. A., 471 (2015), pp. 20140679, 20, https://doi.org/10.1098/rspa.2014.0679.

[15] B. Peherstorfer, M. Gunzburger, and K. Willcox, *Convergence analysis of multifidelity Monte Carlo estimation*, Numerische Mathematik, 139 (2018), pp. 683–707, https://doi.org/10.1007/s00211-018-0945-7.

[16] B. Peherstorfer, K. Willcox, and M. Gunzburger, *Optimal Model Management for Multifidelity Monte Carlo Estimation*, SIAM Journal on Scientific Computing, 38 (2016), pp. A3163–A3194, https://doi.org/10.1137/15M1046472.

[17] B. Peherstorfer, K. Willcox, and M. Gunzburger, *Survey of multifidelity methods in uncertainty propagation, inference, and optimization*, SIAM Rev., 60 (2018), pp. 550–591, https://doi.org/10.1137/16M1082469.

[18] R. Rannacher, *Extrapolation techniques in the finite element method (a survey)*, no. MATC7 in Proc. Summer School on Numerical Analysis, Helsinki, Univ. of Tech., 1988, pp. 80–113.

[19] L. F. Richardson, *The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam*, Philosophical Transactions of the Royal Society of London. Series A, 210 (1911), pp. 307–357, https://doi.org/10.1098/rsta.1911.0009.

[20] W. Romberg, *Vereinfachte numerische Integration*, Det Kongelige Norske Videnskabers Selskab Forhandlinger, 28 (1955), pp. 30–36.

[21] D. Schaden and E. Ullmann, *On multilevel best linear unbiased estimators*, Preprint, No. IGDK-2019-11 (2019), https://www.igdk.eu/foswiki/pub/IGDK1754/Preprints/1905_SchadenUllmann.pdf.

[22] M. L. Stein, *Interpolation of spatial data*, Springer Series in Statistics, Springer-Verlag, New York, 1999, https://doi.org/10.1007/978-1-4612-1494-6. Some theory for Kriging.

[23] D. Talay and L. Tubaro, *Expansion of the global error for numerical schemes solving stochastic differential equations*, Stochastic Anal. Appl., 8 (1990), pp. 483–509 (1991), https://doi.org/10.1080/07362999008809220.