A fast Monte-Carlo method with a Reduced Basis of Control Variates applied to Uncertainty Propagation and Bayesian Estimation

Sébastien Boyaval

▶ To cite this version:

HAL Id: hal-00666264

https://hal-enpc.archives-ouvertes.fr/hal-00666264

Submitted on 3 Feb 2012

A fast Monte-Carlo method with a Reduced Basis of Control Variates applied to Uncertainty Propagation and Bayesian Estimation

Sébastien Boyaval

Université Paris Est, Laboratoire d'hydraulique Saint-Venant (EDF R&D – Ecole des Ponts ParisTech – CETMEF), 78401 Chatou Cedex, France ; and INRIA, MICMAC team–project, Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay, France. E-mail: sebastien.boyaval@saint-venant.enpc.fr

Abstract

The Reduced-Basis Control-Variate Monte-Carlo method was introduced recently in [S. Boyaval and T. Lelièvre, CMS, 8 2010] as an improved Monte-Carlo method, for the fast estimation of many parametrized expected values at many parameter values. We provide here a more complete analysis of the method including precise error estimates and convergence results. We also numerically demonstrate that it can be useful to some parametrized frameworks in Uncertainty Quantification, in particular (i) the case where the parametrized expectation is a scalar output of the solution to a Partial Differential Equation (PDE) with stochastic coefficients (an Uncertainty Propagation problem), and (ii) the case where the parametrized expectation is the Bayesian estimator of a scalar output in a similar PDE context. Moreover, in each case, a PDE has to be solved many times for many values of its coefficients. This is costly and we also use a reduced basis of PDE solutions like in [S. Boyaval, C. Le Bris, Nguyen C., Y. Maday and T. Patera, CMAME, 198 2009]. This is the first combination of various Reduced-Basis ideas to our knowledge, here with a view to reducing as much as possible the computational cost of a simple approach to Uncertainty Quantification.

Keywords: Monte-Carlo method, Variance reduction, Control variate, Reduced Basis method, Partial Differential Equations with stochastic coefficients, Uncertainty Quantification, Bayesian estimation .

1. Introduction

The Reduced-Basis (RB) control-variate Monte-Carlo (MC) method was recently introduced in [1] to compute fast many expectations of scalar outputs of the solutions to parametrized ordinary Stochastic Differential Equations (SDEs) at many parameter values. But as a simple, generic MC method with reduced variance, the RB control-variate MC method can also be useful in other parametric contexts and the main goal of this article is to show that it can be useful to Uncertainty Quantification (UQ) too, possibly in combination with the standard RB method in a PDE context.

There is a huge literature on the UQ subject. Indeed, to be actually predictive in real-life situations [2], most numerical models require (i) to calibrate as much as possible the parameters and (ii) to quantify the remaining uncertainties propagated by the model. Besides, the latter two steps are complementary in an iterative procedure to improve numerical models using datas from experi-

ments: quantifying the variations of outputs generated by input parameters allows one to calibrate the input uncertainties with data and in turn reduces the epistemic uncertainty in outputs despite irreducible aleatoric uncertainty. Various numerical techniques have been developped to quantify uncertainties and have sometimes been used for years [3, 4]. But there are still a number of challenges [5, 6, 7].

For PDEs in particular, the coefficients are typical sources of uncertainties. One common modelling of these uncertainties endows the coefficients with a probability distribution that presumably belongs to some parametric family and the PDEs solutions inherit the random nature of the uncertainty sources. A Bayesian approach is often favoured to calibrate the parameters in the probability law using observations of the reality [8, 9]. But the accurate numerical simulation of the PDEs solutions as a function of parametrized uncertain coefficients is a computational challenge due to its complexity, and even more so is the numerical optimization of the parameters in uncertain models. That is why new/improved techniques are still being investigated [10, 11]. Our goal in this work is to develop a practically useful numerical approach that bases on the simple MC method to simulate the probability law of the uncertain coefficients. We suggest to use the RB control-variate MC method in some UQ frameworks to improve the computational cost of the naive MC method, in particular in contexts where some coefficients in a PDE are uncertain and other are controlled.

There exist various numerical approaches to UQ. The computational cost of MC methods is certainly not optimal when the random PDE solution is regular, see e.g. [12]. But we focus here on MC methods because they are (a) very robust, that is useful when the regularity of the solution with respect to the random variables degrades, and (b) very easy to implement (they are *non-intrusive* in the sense that they can use a PDE numerical solver as a black-box, with the values of the PDE coefficients as input and that of the discrete solution as output). Besides, note that even when the random PDE solution is very regular with respect to the random variables, it is not yet obvious how algorithms can take optimal profit of the regularity of random PDE solutions and remain practically efficient as the dimension of the (parametric) probability space increases, see e.g. [13]. So, focusing on a MC approach, our numerical challenge here is basically two-sided: (i) on the probabilistic side, one should sample fast the statistics of the random PDE solution (or of some random output that is the quantity of interest), and (ii) on the deterministic side, one has to compute fast the solution to a PDE for many realizations of the random coefficients. It was proposed in [14] to use the RB method in order to reduce the numerical complexity of (ii), but this does not fully answer the numerical challenge. In particular, although the RB method can improve naive MC approaches at no-cost (since the selection of the reduced basis for the PDE solutions at various coefficients values can be trained on the same large sample of coefficients values that is necessary to the MC sampling), the resulting MC approach might still be very costly, maybe prohibitively, due to the large number of realizations that is necessary to accurately sample the statistics of the PDE solution (side (i) of our challenge above). In this work, we thus tackle the question how to reduce the numerical complexity of (i). We have in mind the particular but useful case where one is interested in the expected value of a random scalar output of the random PDE solution as a function of a (deterministic) *control parameter*, typically another (deterministic) coeffi-

cient in the UQ problem which is "under control". (Think of the construction of response surfaces for a mean value as a function of control parameters.) A similar parametric context occurs in Bayesian estimation, sometimes by additionally varying the *hyper parameters* or the observations. In any case, our goal is to reduce the computational cost of a parametrized (scalar) MC estimation when the latter has to be done many times for many values of a parameter, and we illustrate it with examples meaningful in a UQ context.

To accelerate the convergence of MC methods as numerical quadratures for the expectation of a random variable, one idea is to tune the sampling for a given family of random variables like in the quasi-Monte-Carlo (qMC) methods [15, 16, 17]. Another common idea is to sample another random variable with same mean but with a smaller variance. Reducing the variance allows one to take smaller MC samples of realizations and yet get MC estimations with confidence intervals of similar (asymptotic) probability. Many techniques have been designed in order to reduce the variance in various contexts [18, 19]. Our RB control-variate MC method bases on the so-called control-variate technique. It has a specific scope of application in parametric contexts. But it suits very well to some computational problems in mathematical finance and molecular dynamics as shown in [1], and can be useful in UQ as we are going to see.

The paper is organized as follows. In Section 2, we recall the RB control-variate technique as a general variance reduction tool for the MC approximation of a parametrized expected value at many values of the parameter. The presentation is a bit different to that in [1], which was more focused on SDEs. Moreover, we also give new error estimates and convergence results. In Section 3, the RB control-variate MC method is applied to compute the mean of a random scalar output in a model PDE with stochastic coefficients (the input uncertainty) at many values of a control parameter. In Section 4, it is applied to Bayes estimation, first for a toy model where various parametric contexts are easily discussed, then for the same random PDE as in section 3.

We also note that this work does not only improve on the RB approach to UQ [14] but also on an RB approach to Bayesian estimation proposed in [20] with a deterministic quadrature formula to evaluate integrals. For both applications, to our knowledge, our work is the first attempt at optimally approximating the solution with a simple MC/FE method by combining RB ideas of two kinds, stochastic and deterministic ones [21]. Note that for convenience of the reader non-expert in RB methods, the standard RB method [22, 23] is briefly recalled in Section 3.3.

2. The RB Control-Variate MC Method

The RB control-variate technique is a generic variance reduction tool for the MC approximation of a parametrized expected value at many values of the parameter. In this section we recall the technique for the expectation $E(Z^\lambda)$ of a generic square-integrable random variable $Z^\lambda \in L_P^2$ parametrized by λ. The principle for the reduction of computations is based on the same paradigm as the standard RB method and allows one to accelerate the MC computations of many $E(Z^\lambda)$ at many values of λ. Our presentation is slightly different than the initial one in [1] and gives new elements of analysis (error estimates and convergence results).

2.1. Principles of RB control-variate MC method

Let P be a probability measure such that Z^λ is a random variable in L_P^2 for all parameter values λ

in a given fixed range Λ. Assume one has an algorithm to simulate the law of Z^λ whatever $\lambda \in \Lambda$. Then, at any $\lambda \in \Lambda$, one can define MC estimators $E_M(Z^\lambda)$ that provide useful approximations of $E(Z^\lambda)$, by virtue of the strong law of large numbers

$$E_M(Z^\lambda) := \frac{1}{M} \sum_{m=1}^{M} Z_m^\lambda \xrightarrow[M\to\infty]{P-a.s.} E(Z^\lambda), \qquad (2.1)$$

provided the number M of independent identically distributed (i.i.d.) random variables $Z_m^\lambda \sim Z^\lambda$, $m = 1 \dots M$, is sufficiently large. Here, the idea is: if $E(Z^{\lambda_i^I}) := \int Z^{\lambda_i^I} dP$ is already known with a good precision for I parameter values λ_i^I, $i = 1 \dots I$, ($I \in \mathbb{N}_{>0}$) and if the law of Z^λ depends smoothly on λ then, given I well-chosen real values α_i^λ, $i = 1 \dots I$, the standard MC estimator $E_M(Z^\lambda)$ could be efficiently replaced by a MC estimator for $E(Z^\lambda - \sum_{i=1}^{I} \alpha_i^\lambda Z^{\lambda_i^I}) + \sum_{i=1}^{I} \alpha_i^\lambda E(Z^{\lambda_i^I})$ that is as accurate and uses much less than M copies of Z^λ.

In other words, if the random variable

$$\hat{Y}^\lambda = \sum_{i=1}^{I} \alpha_i^\lambda (Z^{\lambda_i^I} - E(Z^{\lambda_i^I})) \qquad (2.2)$$

is correlated with Z^λ such that the control of Z^λ by $\hat{Y}^\lambda$ reduces the variance, that is if

$$V(Z^\lambda) := \int |Z^\lambda - E(Z^\lambda)|^2 dP \geq V(Z^\lambda - \hat{Y}^\lambda),$$

then the confidence intervals with asymptotic probability $\mathrm{erf}\left(\frac{a}{\sqrt{2}}\right)$ ($a > 0$) for MC estimations

$$E_M(Z^\lambda - \hat{Y}^\lambda) := \sum_{m=1}^{M} \frac{Z_m^\lambda - \hat{Y}_m^\lambda}{M} \xrightarrow[M\to\infty]{P-a.s.} E(Z^\lambda) \quad (2.3)$$

that are in the Central Limit Theorem (CLT)

$$P\left(\frac{\left| E_M(Z^\lambda - \hat{Y}^\lambda) - E(Z^\lambda) \right|}{\sqrt{V(Z^\lambda - \hat{Y}^\lambda)/M}} \leq a \right) \xrightarrow[M\to\infty]{} \mathrm{erf}\left(\frac{a}{\sqrt{2}}\right)$$
$$(2.4)$$

converge faster with respect to the number M of realizations than the confidence intervals for (2.1).

The unbiased estimator (2.3) ($E(\hat{Y}^\lambda) = 0$) is thus a better candidate than (2.1) for a fast MC method.

At a given $\lambda \in \Lambda$, we define α_i^λ, $i = 1 \dots I$, in (2.2) to obtain the optimal variance reduction minimizing $V(Z^\lambda - \hat{Y}^\lambda)$. Equivalently, the control variates of the form (2.2) are thus defined as best approximations of the ideal control variate $Y^\lambda := Z^\lambda - E(Z^\lambda) \in L_P^2$ that reduces the variance to zero

$$\inf_{(\alpha_i^\lambda)\in\mathbb{R}^I} E\left(\left\| Y^\lambda - \sum_{i=1}^{I} \alpha_i^\lambda Y^{\lambda_i^I} \right\|^2 \right). \qquad (2.5)$$

The α_i^λ, $i = 1 \dots I$, thus solve a least-squares problem with normal equations for $i = 1 \dots I$

$$\sum_{j=1}^{I} C\left(Z^{\lambda_i^I}, Z^{\lambda_j^I} \right) \alpha_j^\lambda = C\left(Z^{\lambda_i^I}, Z^\lambda \right), \qquad (2.6)$$

where $C(Y, Z) = E((Y - E(Y))(Z - E(Z)))$ denotes the covariance between $Y, Z \in L_P^2$. And they are unique as long as the matrix $\mathbf{C}$ with entries $\mathbf{C}_{i,j} = C\left(Z^{\lambda_i^I}, Z^{\lambda_j^I} \right)$ remains definite.

In general, the computation of a good control variate is difficult (the ideal one Y^λ requires the result $E(Z^\lambda)$ itself). But we proved in [1] that control variates of the form (2.2) actually make sense in some contexts, a result inspired by [24, 25]. Let us denote by $L_{P,0}^2$ the Hilbert subspace of random variables $Y \in L_P^2$ that are centered $E(Y) = 0$.

Proposition 1. *Consider a set of random variables*

$$Y^\lambda = \sum_{j=1}^{J} g_j(\lambda) Y_j, \ \forall \lambda \in \Lambda, \qquad (2.7)$$

where $Y_j \in L_{P,0}^2$, $j = 1 \dots J$, are uncorrelated and $(g_j)_{1\leq j\leq J}$ are $C_{\geq 0}^\infty(\mathbb{R})$ functions. If there exists a constant $C > 0$ and an interval $\widetilde{\Lambda}$ such that, for all parameter ranges $\Lambda = [\lambda_{\min}, \lambda_{\max}] \subset \mathbb{R}$, there exists a C^∞ diffeomorphism $\tau_\Lambda : \Lambda \to \widetilde{\Lambda}$ where

$$\sup_{1\leq j\leq J} \sup_{\tilde{\lambda}\in\tau_\Lambda(\Lambda)} (g_j \circ \tau_\Lambda^{-1})^{(M)}(\tilde{\lambda}) \leq M! C^M, \qquad (2.8)$$

for all M-derivatives $(g_j \circ \tau_\Lambda^{-1})^{(M)}$ of $g_j \circ \tau_\Lambda^{-1}$, then there exist constants $c_1, c_2 > 0$ independent of Λ and J such that, for all parameter ranges $\Lambda = [\lambda_{\min}, \lambda_{\max}] \subset \mathbb{R}$, for all $I \in \mathbb{N}$, $I \geq I_0 > 0$,

$$\inf_{Y \in \mathcal{Y}_I} V(Y^\lambda - Y) \leq e^{-\frac{c_2(I-1)}{I_0-1}} V(Y^\lambda), \; \forall \lambda \in \Lambda, \quad (2.9)$$

where $I_0 := 1 + c_1 (\tau_\Lambda(\lambda_{\max}) - \tau_\Lambda(\lambda_{\min}))$ and $\mathcal{Y}_I = \mathrm{span}(Y^{\lambda_i^I}, i = 1, \ldots, I)$ uses the I explicit values $\lambda_i^I = \tau_\Lambda^{-1} \left(\tau_\Lambda(\lambda_{\min}) + \frac{i-1}{I-1} (\tau_\Lambda(\lambda_{\max}) - \tau_\Lambda(\lambda_{\min})) \right)$.

Prop. 1 tells us that using (2.2) as a control variate makes sense because the variance in (2.9) decays very fast with I for all $\lambda \in \Lambda$, and whatever Λ.

But the explicit construction of λ_i^I, $i = 1 \ldots I$, in Prop. 1, for random variables Z^λ that are analytic in a 1D parameter λ, does not generalize to higher-dimensions in an efficient way, a well-known manifestation of the *curse of dimensionality* [1]. Yet, our numerical results (here and in [1]) show our variance reduction principle can work well in contexts with higher-dimensional or less regular parametrization. Then, given a parametric family Z^λ, $\lambda \in \Lambda$, how to choose parameter values λ_i^I, $i = 1 \ldots I$, that can in practice efficiently reduce the MC computations at all $\lambda \in \Lambda$?

Given $I \in \mathbb{N}$, let us define λ_i^I, $i = 1 \ldots I$, as the parameter values with optimal approximation properties in $L_\Lambda^\infty(L_{P,0}^2)$: the linear space spanned by $Y^{\lambda_i^I} = Z^{\lambda_i^I} - E(Z^{\lambda_i^I})$ minimizes the variances over all $\lambda \in \Lambda$, or equivalently the λ_i^I, $i = 1 \ldots I$, minimize

$$\bar{d}_I := \inf_{\{\lambda_1^I, \ldots, \lambda_I^I\} \in \Lambda^I} \sup_{\lambda \in \Lambda} \inf_{(\alpha_i^\lambda) \in \mathbb{R}^I} V(Z^\lambda - \hat{Y}^\lambda). \quad (2.10)$$

The variates $Y^{\lambda_i^I}$, $i = 1 \ldots I$, define an optimal I-dimensional *reduced basis* in $L_{P,0}^2$ for the approximations $\hat{Y}^\lambda$ of the ideal controls Y^λ at all $\lambda \in \Lambda$. The values $\bar{d}_I, I \in \mathbb{N}$, in (2.10) coincide with a well-known notion in approximation theory [26]:

the Kolmogorov widths of $\mathcal{Y} := \{Z^\lambda - E(Z^\lambda), \lambda \in \Lambda\}$ in $\mathcal{Y} \subset L_{P,0}^2$, and are upper bounds for the more usual Kolmogorov widths of $\mathcal{Y}$ in $L_{P,0}^2$

$$d_I := \inf_{\{Y_i^I\} \in [L_{P,0}^2]^I} \sup_{\lambda \in \Lambda} \inf_{Y \in \mathrm{span}(Y_1^I, \ldots, Y_I^I)} \|Y^\lambda - Y\|_{L_P^2}.$$

Explicitly computing Kolmogorov widths is a difficult problem solved only for a few simple sets $\mathcal{Y}$, let alone the computation of (non-necessarily unique) minimizers. But in a *many-query* parametric framework where the MC computations are queried many times for many values of the parameter, one can take advantage of a presumed smooth dependence of the MC computations on the parameter to identify a posteriori some approximations for the optimal parameter values λ_i^I, $i = 1 \ldots I$, with prior computations at only a few parameter values. Moreover, in practice, we shall be content with approximations of λ_i^I, $i = 1 \ldots I$ that are not very good, provided they produce sufficiently fast-decaying upper-bounds of $\bar{d}_I$ as $I \to \infty$. To our knowledge, this reduction paradigm has been applied for the first time in [1] to minimize the variance of many parametrized MC estimations. It is inspired by the (by now) standard RB method to minimize the discretization error in many-query Boundary Value Problems (BVP) for parametrized PDEs (see [22, 23] e.g. and Section 3.3).

Let us define linear combinations like (2.2) when $I = 0$ by 0. We next define a (non-unique) sequence of parameter values $\lambda_i \in \Lambda$, $i \in \mathbb{N}_{>0}$ by the sucessive iterations $I \in \mathbb{N}$ of a *greedy* algorithm

$$\lambda_{I+1} \in \mathrm{argsup}_{\lambda \in \Lambda} V(Z^\lambda - \sum_{i=1}^I Z^{\lambda_i}). \quad (2.11)$$

For all $I \in \mathbb{N}$, the parameter values λ_i, $i = 1 \ldots I$, approximate λ_i^I, $i = 1 \ldots I$ in (2.10) and yield the following upper-bound

$$\sigma_I := \sup_{\lambda \in \Lambda} V(Z^\lambda - \sum_{i=1}^I \alpha_i^\lambda Z^{\lambda_i}) \geq \bar{d}_I \geq d_I. \quad (2.12)$$

---

[1] A tensor-product of (2.9) in dimension d yields a rate $-I/d$.

Using a finite discrete subset $\tilde{\Lambda} \subset \Lambda$ (possibly $\tilde{\Lambda} = \Lambda$ in case of finite cardinality $\text{card}(\Lambda) < \infty$) and a MC estimator for the variance, like

$$V_M(Z^\lambda - \sum_{i=1}^{I} Z^{\lambda_i}) = \frac{1}{M-1} \sum_{m=1}^{M} \left| (Z_m^\lambda - \sum_{i=1}^{I} \alpha_i^\lambda Z_m^{\lambda_i}) - \frac{m-1}{m} E_m(Z^\lambda - \sum_{i=1}^{I} \alpha_i^\lambda Z^{\lambda_i}) \right|^2, \quad (2.13)$$

we can finally define a computable sequence of parameter values $\tilde{\lambda}_i \in \Lambda$, $i \in \mathbb{N}_{>0}$ by sucessive iterations $I \in \mathbb{N}$ of a *weak greedy* algorithm at $\omega \in \Omega$

$$\tilde{\lambda}_{I+1} \in \underset{\lambda \in \tilde{\Lambda}}{\arg\sup} V_M(Z^\lambda - \sum_{i=1}^{I} Z^{\tilde{\lambda}_i})(\omega). \quad (2.14)$$

For all $I \in \mathbb{N}$, our computable approximations of the optimal parameter values λ_i^I, $i = 1 \dots I$ used hereafter in practical applications will be $\tilde{\lambda}_i \equiv \tilde{\lambda}_i(\omega)$, $i = 1 \dots I$, from a set like in (2.14). Note that the $\tilde{\lambda}_i$, $i \in \mathbb{N}_{>0}$, are constructed iteratively and thus do not depend on I but only on $\tilde{\lambda}_1$ (and possibly on $\tilde{\Lambda}$), which is useful in practice: the adequate value of I for a good variance reduction is typically not known in advanced and one simply stops the weak greedy algorithm at $I \in \mathbb{N}$ when, at a given $\omega \in \Omega$,

$$\tilde{\sigma}_I := \sup_{\lambda \in \tilde{\Lambda}} V_M(Z^\lambda - \sum_{i=1}^{I} \alpha_i^\lambda Z^{\tilde{\lambda}_i})(\omega) \le TOL \quad (2.15)$$

is smaller than a maximum tolerance $TOL > 0$.

Even if greedy algorithms might yield approximate minimizers of (2.10) that are far suboptimal, numerical results show that they can nevertheless be useful to computational reductions using the RB control-variate MC method, just as they already proved useful to computational reductions with the standard RB method in numerous practical examples. Recent theoretical results [27, 28] also support that viewpoint as concerns the standard RB method and can be straightforwardly adapted to our framework. Comparing directly σ_I with d_I at

same $I \in \mathbb{N}$ will not, in general, give estimates better than $\sigma_I \le \frac{2^{I+1}}{\sqrt{3}} d_I$. And this is pessimistic as regards the convergence of the greedy algorithm insofar as it predicts variance reduction only for sets with extremely fast decaying Kolmogorov widths d_I. Yet, the two sequences σ_I and d_I typically have comparable decay rates as $I \to \infty$, and it holds for adequate $\eta, \beta, c, C > 0$ given any $d_0, \alpha, a > 0$:

a) $d_I \le d_0 I^{-\alpha}$, $\forall I \Rightarrow \sigma_I \le C d_0 I^{-\alpha}$, $\forall I$,

b) $d_I \le d_0 e^{-aI^{-\alpha}}$, $\forall I \Rightarrow \sigma_I \le C d_0 e^{-cI^{-\beta}}$, $\forall I$,

c) $d_I \le d_0 e^{-aI^{-\alpha}}$, $\forall I \Rightarrow \sigma_I \le C I \eta^I e^{-aI^{-\alpha}}$, $\forall I$,

where c) is sharper than b) if, and only if, $\alpha > 1$ or $\alpha = 1$ and $a > \ln 2$. So, when the Kolmogorov widths d_I decay fast (variance reduction is a priori possible), greedy algorithms can be useful if used with I sufficiently many iterations [2].

The latter results can also be adapted to the weak greedy algorithm, like in [28]. For all $\epsilon \in (0, 1)$ and $I \in \mathbb{N}$, given any $\theta_I \in (0, 1)$, we define the smallest number of realizations $M_{\text{test}}^{\epsilon, I}$ such that

$$P\left(|V_{M_{\text{test}}^{\epsilon, I}}(Z^\lambda - \hat{Y}^\lambda) - V(Z^\lambda - \hat{Y}^\lambda)| \le \theta_I d_I \right) \ge 1 - \epsilon$$

holds for all λ and control variates $\hat{Y}^\lambda$ of dimension $0 \le i \le I$. Of course, we do not know exactly $M_{\text{test}}^{\epsilon, I}$ in practice, which could even not be finite if variations in λ are not smooth enough. But if, for all $\epsilon \in (0, 1)$, we assume that one can use more than $M_{\text{test}}^{\epsilon, I}$ realizations in the computations of step I of the weak greedy algorithm, then the following events occur with probability more than $1 - \epsilon$, for adequate $\eta, \beta, c, C > 0$ given any $d_0, \alpha, a > 0$:

a) $d_I \le d_0 I^{-\alpha}$, $\forall I \Rightarrow \tilde{\sigma}_I \le C d_0 I^{-\alpha}$, $\forall I$,

b) $d_I \le d_0 e^{-aI^{-\alpha}}$, $\forall I \Rightarrow \tilde{\sigma}_I \le C d_0 e^{-cI^{-\beta}}$, $\forall I$,

---

[2]Then computational reductions are possible, but only for sufficiently many queries in the parameter of course, as we shall see.

c) $d_I \le d_0 e^{-aI^{-\alpha}}, \forall I \Rightarrow \tilde{\sigma}_I \le CI\eta^I e^{-aI^{-\alpha}}, \forall I,$

see Appendix A for a proof. Of course, the constants C in the upper-bounds above can be overly pessimistic. But our numerical results show that one does not need to go to the asymptotics $I \to \infty$ in order to get good variance reduction results.

2.2. Implementation and analysis

At any step $I \in \mathbb{N}$ of a weak greedy algorithm, whether we are looking for a new parameter value $\tilde{\lambda}_{I+1}$ or not (TOL reached), the practical implementation of our RB control-variate MC method still requires a few discretization ingredients. In particular, one has to be able to actually compute approximations of $E(Z^\lambda) = E(Z^\lambda - \sum_{i=1}^{I} \alpha_i^\lambda Z^{\tilde{\lambda}_i}) + \sum_{i=1}^{I} \alpha_i^\lambda E(Z^{\tilde{\lambda}_i})$ and $V(Z^\lambda - \hat{Y}^\lambda)$ whatever $\lambda \in \Lambda$. We are now going to suggest a generic practical approach to that question. On the contrary, practically useful choices of $\tilde{\Lambda} \subset \Lambda$ specifically depend on the parametric context and will be discussed only for specific applications in the next sections.

First, we need to approximate the expectations $E(Z^{\tilde{\lambda}_i})$, $i = 1 \ldots I$. To obtain good useful approximations of $E(Z^{\tilde{\lambda}_i})$ in general, we suggest to use an expensive MC estimation $E_{M_{\text{large}}}(Z^{\tilde{\lambda}_i})$, independent from subsequent realizations of the random variable $Z^{\tilde{\lambda}_i}$. Although this is a priori as computationally expensive as approximating well $E(Z^\lambda)$ at any λ by a naive MC estimation, we do it only once at each of the few $\tilde{\lambda}_i$. Once a realization of $E_{M_{\text{large}}}(Z^{\tilde{\lambda}_i})$ has been computed at step i of the weak greedy algorithm ($1 \le i \le I$) with $M_{\text{large}} \gg 1$ realizations, the deterministic result can be stored in memory. We denote the computable approximations of $Y^{\tilde{\lambda}_i} \equiv Z^{\tilde{\lambda}_i} - E(Z^{\tilde{\lambda}_i})$ by $\tilde{Y}^{\tilde{\lambda}_i} \equiv Z^{\tilde{\lambda}_i} - E_{M_{\text{large}}}(Z^{\tilde{\lambda}_i})$, $i = 1 \ldots I$, so that the actual control variate $\hat{Y}^\lambda$ used in practice is in fact a linear combination in a linear space of dimension I spanned by $\tilde{Y}^{\tilde{\lambda}_i}$, $i = 1 \ldots I$. Note that given

$E_{M_{\text{large}}}(Z^{\tilde{\lambda}_i})$ we expect realizations of $\tilde{Y}^{\tilde{\lambda}_i}$ to be computed concurrently with realizations of Z^λ, using the same pseudo-random numbers generated by a computer, with approximately the same cost.

Second, for any λ, we need to replace the coefficients α_i^λ, $i = 1 \ldots I$, with a tractable solution $\tilde{\alpha}_i^\lambda$ of the minimization problem (2.5), so that practical control variates in fact read

$$\hat{Y}^\lambda = \sum_{i=1}^{I} \tilde{\alpha}_i^\lambda \tilde{Y}^{\tilde{\lambda}_i}. \qquad (2.16)$$

Now, the MC computations at any one specific λ should be fast, in particular the approximate solution to the minimization problem (2.5), whether we decide to stop the weak greedy algorithm at step I and use only I parameter values or still want to explore the parameter values in $\tilde{\Lambda}$ to select a new parameter value $\tilde{\lambda}_{I+1}$. So we invoke only a *small* number M_{small} of realizations of Z^λ and $\tilde{Y}^{\tilde{\lambda}_i}$ to compute the $\tilde{\alpha}_i^\lambda$. Note that there are different strategies for the numerical solution to a least-squares problem like (2.5) that use only M_{small} realizations of Z^λ and $\tilde{Y}^{\tilde{\lambda}_i}$. One can try to compute directly the solution to an approximation of the linear system (2.6) where variances and covariances are approximated with MC estimations similar to (2.13) using M_{small} i.i.d. realizations Z_m^λ and $\tilde{Y}_m^{\tilde{\lambda}_i}$, $m = 1 \ldots M_{\text{small}}$, of Z^λ and $\tilde{Y}^{\tilde{\lambda}_i}$. The MC estimations similar to (2.13) are indeed interesting insofar as they remain positive, see [29, 30] for a study of consistency. But there may still be difficulties when **C** is too ill-conditioned. One can thus also apply a QR decomposition approach to a small set of M_{small} realizations, using the Modified Gram-Schmidt (MGS) algorithm for instance [3]. In any case, realizations of Z^λ and $\tilde{Y}^{\tilde{\lambda}_i}$, $i = 1 \ldots I$,

---

[3] The matrix with entries $(\tilde{Y}_m^{\tilde{\lambda}_i})_{m=1\ldots M_{\text{small}}, i=1\ldots I}$ uniquely decomposes as QR where Q is a $M \times I$ matrix such that $Q^T Q$ is the $I \times I$ identity matrix and R is an upper triangular $I \times I$ ma-

should be correlated and thus computed with the same collection of random numbers in practice. Besides, one can use a single collection of random numbers (generated by initializing the random number generator at one fixed seed) for all λ, including $\tilde{\lambda}_i$, $i = 1 \ldots I$, and realizations of $\tilde{Y}^{\tilde{\lambda}_i}$ can thus be computed only once (at step i of the weak greedy algorithm) and next stored in memory.

Finally we approximate $E(Z^\lambda)$ by one draw of

$$\frac{1}{M_{\text{test}}} \sum_{m=1}^{M_{\text{test}}} \left(Z_m^\lambda - \sum_{i=1}^{I} \tilde{\alpha}_i^\lambda \tilde{Y}_m^{\tilde{\lambda}_i} \right) \qquad (2.17)$$

whatever λ and compute a MC estimation (2.13) of $V(Z^\lambda - \hat{Y}^\lambda)$ with the same M_{test} realizations of Z^λ and $\tilde{Y}^{\tilde{\lambda}_i}$. Remember that at step $I \in \mathbb{N}$ of the weak greedy algorithm, (2.17) is useful either to inspect all parameter values $\lambda \in \tilde{\Lambda}$, check whether TOL is reached and next select a new parameter value $\tilde{\lambda}_{I+1}$ when it is not. Or if TOL has been reached before, then (2.17) is used for definitive estimations at any λ (still with companion variance estimations (2.13) to concurrently certify that TOL is maintained).

Let us compare our strategy with the cost of direct MC estimations that have same confidence levels and thus use M realizations such that $V_M(Z^\lambda)/M = V_{M_{\text{test}}}(Z^\lambda - \hat{Y}^\lambda)/M_{\text{test}}$. We denote by C the cost of one realization at one parameter value λ compared with that of one multiplication. The evaluation of $M_{\text{small}} + M_{\text{test}}$ realizations at each λ, plus $I^2 + IM_{\text{small}}$ multiplications for the QR solution to the least-squares problem, IM_{test} multiplications for the MC estimation of the output expectation, and M_{test}^2 for the variance, make our RB control-variate MC method interesting as soon as

$M\mathsf{C} + M^2 \geq (M_{\text{test}} + M_{\text{small}})(\mathsf{C} + I) + I^2 + M_{\text{test}}^2$, at least for "real-time" purposes. Then, the price of identifying I control variates with the weak greedy algorithm pays back if one needs to compute very fast $E(Z^\lambda)$ for any λ, provided the I control variates still provide good variance reduction in the case where they were constructed by a weak greedy algorithm trained on $\tilde{\lambda} \subsetneq \Lambda$. In addition, the RB control-variate MC method is also interesting in the many-query cases where one has to compute $E(Z^\lambda)$ for a sufficiently large number $\sharp\lambda$ of parameter values λ. At each greedy step $i = 1 \ldots I - 1$, in addition to variance estimations at each $\lambda \in \tilde{\Lambda}$, the selection of $\tilde{\lambda}_{i+1}$ requires a quicksort of the sample of estimated variances $\{ V_{M_{\text{test}}} \left(Z^\lambda - \hat{Y}^\lambda \right), \lambda \in \tilde{\Lambda} \}$, plus the computation of $E_{M_{\text{large}}}(Z^{\tilde{\lambda}_{i+1}})$ and M_{small} realizations of $Z^{\tilde{\lambda}_{i+1}}$ to be stored. The cost of the greedy algorithm $\mathsf{C}_{\text{greedy}} = \text{card}(\tilde{\Lambda}) \Big(I(M_{\text{test}} + M_{\text{small}})\mathsf{C} + I(I+1)(2I + 1)/6 + I(I+1)M_{\text{small}} + M_{\text{test}}/2 + IM_{\text{test}}^2 + \ln \text{card}(\tilde{\Lambda}) \Big) + I(M_{\text{large}} + M_{\text{small}})\mathsf{C}$ is then compensated by variance reductions as soon as $\sharp\lambda \geq \mathsf{C}_{\text{greedy}}/(M\mathsf{C} + M^2 - (M_{\text{test}} + M_{\text{small}})(\mathsf{C} + I) - I^2 - M_{\text{test}}^2)$.

We also have a posteriori error estimates. For values $\tilde{\alpha}_i^\lambda$ given by a fixed MC estimation with M_{small} realizations independent of M_{test} new realizations in (2.17), CLT (2.4) holds[4], $\forall a > 0$, for

$$P \left(\frac{\left| E_{M_{\text{test}}}(Z^\lambda - \hat{Y}^\lambda) - E(Z^\lambda) + \sum_{i=1}^{I} \tilde{\alpha}_i^\lambda E(\tilde{Y}^i)) \right|}{\sqrt{V(Z^\lambda - \hat{Y}^\lambda)/M_{\text{test}}}} \leq a \right) \qquad (2.18)$$

where the MC empirical estimator $E_{M_{\text{test}}}(Z^\lambda - \hat{Y}^\lambda)$ is defined in (2.17) and where one can replace $V(Z^\lambda - \hat{Y}^\lambda)$ by a MC estimator $V_{M_{\text{test}}}(Z^\lambda - \hat{Y}^\lambda)$ by Slutsky theorem. Furthermore, although (2.18) is not

---

trix. The I-dimensional vector $\tilde{\alpha}^\lambda$ with entries $\tilde{\alpha}_i^\lambda$, $i = 1 \ldots I$, useful in $\hat{Y}^\lambda$ is next defined as $R^{-1}Q^T Z$ (where Z is the M_{small}-dimensional vector with entries Z_m^λ). The $\tilde{\alpha}_i^\lambda$ are thus random variables depending on M_{small} realizations and an approximate minimum of (2.5) is $Z^T Z - Z^T Q Q^T Z$.

[4]If the M_{test} realizations in (2.17) are the same as the M_{small} ones used to compute $\tilde{\alpha}_i^\lambda$, then (2.17) is not a MC empirical estimator of the type $E_{M_{\text{test}}}(Z^\lambda - \hat{Y}^\lambda)$. It does not use independent realizations of the random variable $Z^\lambda - \hat{Y}^\lambda$, so the CLT does not hold and it is difficult to give a rigorous quantitative estimate of the statistical error.

a full error analysis because it does not take into account the bias $\sum_{i=1}^{I} \tilde{\alpha}_i^{\lambda} E(\tilde{Y}^i) = \sum_{i=1}^{I} \tilde{\alpha}_i^{\lambda}(E(Z^{\lambda_i}) - E_{M_{\text{large}}}(Z^{\lambda_i}))$, a function of M_{large} and M_{small} precomputed realizations, probabilities like (2.18) are conditionally to $E_{M_{\text{large}}}(Z^{\lambda_i})$ (and to $\tilde{\alpha}_i^{\lambda}$), so Bayes rule applies and it also holds, for all $a, a_i > 0$:

$$P\left(\left|E_{M_{\text{test}}}(Z^{\lambda} - \hat{Y}^{\lambda}) - E(Z^{\lambda})\right| + \sum_{i=1}^{I} |\tilde{\alpha}_i^{\lambda}| a_i \sqrt{\frac{V_{M_{\text{large}}}(Z^{\lambda_i})}{M_{\text{large}}}}\right)$$

$$\xrightarrow[M_{\text{large}} \geq M_{\text{test}} \to \infty]{} \text{erf}\left(\frac{a}{\sqrt{2}}\right) \prod_{i=1}^{I} \text{erf}\left(\frac{a_i}{\sqrt{2}}\right) \quad (2.19)$$

where variances have been replaced by estimators.

Last, to get a full convergence analysis that predicts an efficient variance reduction with the RB control-variate MC method in practice, at least when $\tilde{\Lambda} = \Lambda$, one should take into account all realizations in $\hat{Y}^{\lambda}$ which in fact reads $\hat{Y}_{M_{\text{small}}, M_{\text{large}}}^{\lambda} = \sum_{i=1}^{I} \tilde{\alpha}_{i, M_{\text{small}}}^{\lambda} \left(Z^{\tilde{\lambda}_i} - E_{M_{\text{large}}}(Z^{\tilde{\lambda}_i})\right)$. Yet, note first that in the weak greedy algorithm (2.14), only the M_{small} realizations introduce new statistical error. And second, one can again predict that, with a good probability, the greedy algorithm is robust to discretization (in the same sense as in [28]) provided the realizations of the least-squares problems are not too close to rank-deficient and their numerical solution is close to the solution. We do not state this more rigorously but instead refer to [28], whose results can be adapted in the same way as in Appendix A for the week greedy algorithm.

3. Application to Uncertainty Propagation

In this section, we numerically demonstrate the efficiency of the RB control-variate MC method for uncertainty propagation in a representative UQ framework. As example, we consider a PDE parametrized by stochastic coefficients and other non-stochastic coefficients which we term *control parameters*. The goal is to compute fast many expectations of a scalar output of the random PDE solution for many values of the control parameters.

For the numerical simulations, as usual in UQ frameworks [3], we use stochastic coefficients that are random fields with a Karhunen-Loève (KL) spectral decomposition. In practice, one can only use representations with a finite-rank K of course and the MC method allows one to simulate the law of the random field just by generating realizations of the K random numbers in the KL decomposition. We compute approximations to the realizations of the random PDE solution with a standard FE discretization method (the same one whatever the realizations). The expectation of the random scalar output can be computed for many values of a control parameter just by reiterating many times the MC procedure. But this is costly. So first, we do not actually compute the PDE solution with the FE method at each realization of the stochastic coefficients and for each control parameter value. In fact, we replace it with a cheap reliable surrogate built with the standard RB method like in [14]. Yet, even with a cheap surrogate model instead of the full FE, the MC method is still costly, because computing the expectation of the parametrized output under the input probability law still requires a very large number of realizations. This quickly becomes untractable as soon as one has to do it many times for many values of the control parameter. Then, we use the RB control-variate MC method to decrease the number of MC realizations needed at most of the control parameter values.

Compared with [14], one uses here an improved MC method to show how various RB ideas can be combined to efficiently tackle some uncertainty propagation problems for partial differential equations with stochastic coefficients. We thus especially discuss the practice of the RB control-

variate MC method here. Yet, although the example is exactly the same as in [14], the specific use of the standard RB method in the frame of UQ, and in combination with the RB control-variate MC method, requires special care. That is why at the end of this section we briefly expose our implementation of the standard RB method, without which we cannot use the MC method at a reasonable price, and next discuss our specific use of it.

3.1. An elliptic PDE with stochastic coefficients

Consider first the solution u to a scalar Robin Boundary Value Problem (BVP) in a regular domain $\mathcal{D} \subset \mathbb{R}^2$: u satisfies Laplace equation in $\mathcal{D}$

$$-\text{div}\,(\kappa \nabla u) = 0\,, \qquad (3.1)$$

and conditions of third type on the boundary $\partial \mathcal{D}$

$$\kappa(\boldsymbol{n} \cdot \nabla)u + b\,u = g\,, \qquad (3.2)$$

where div and ∇ denote the usual divergence and gradient operators in $\mathcal{D}$ equipped with a cartesian frame, $\boldsymbol{n}$ the outward unit normal on $\partial \mathcal{D}$, and k, b and g are scalar parameter functions. For the simulations we will choose in particular

$$\kappa = k_1 1_{\mathcal{D}_1} + k_2 1_{\mathcal{D}_2} \quad (k_1, k_2) \in \mathbb{R}^2_{>0} \qquad (3.3)$$

$$b = b 1_{\Gamma_B} \qquad\qquad b \in L^\infty(\partial \mathcal{D}, \mathbb{R}_{>0}) \quad (3.4)$$

$$g = g 1_{\Gamma_R} \qquad\qquad g \in L^2(\partial \mathcal{D}) \qquad (3.5)$$

where $1_{\mathcal{D}_i}$ is the characteristic function of $\mathcal{D}_i$,

$$\mathcal{D}_1 \cap \mathcal{D}_2 = \emptyset \quad \mathcal{D}_1 \cup \mathcal{D}_2 \subset \mathcal{D} \subset \overline{\mathcal{D}_1 \cup \mathcal{D}_2}\,,$$

and the boundary decomposes into subsets

$$\Gamma_B \cap \Gamma_R = \emptyset \quad \Gamma_B \cup \Gamma_R \subset \partial \mathcal{D} \quad \Gamma_N = \partial \mathcal{D} \backslash \overline{\Gamma_B \cup \Gamma_R}\,.$$

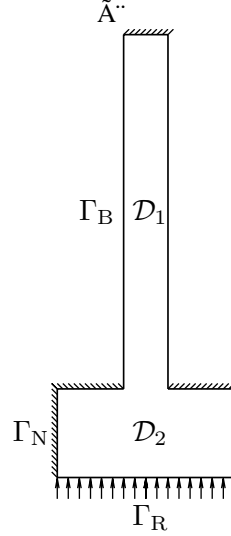There exists a unique weak solution $u \in H^1(\mathcal{D})$ to (3.1–3.2), which can also be defined as the

Figure 1: Thermal fin geometry $\mathcal{D}$ used in the numerical simulations ($\Gamma_N = \overline{\partial \mathcal{D} \backslash (\Gamma_R \cup \Gamma_B)}$) denotes the boundary subject to homogeneous pure Neumann conditions).

unique solution to the following variational problem [31]: Find $u \in H^1(\mathcal{D}) / \forall v \in H^1(\mathcal{D})$

$$\int_{\mathcal{D}_1} k_1 \nabla u \cdot \nabla v + \int_{\mathcal{D}_2} k_2 \nabla u \cdot \nabla v + \int_{\Gamma_B} b\,u\,v = \int_{\Gamma_R} g\,v\,.$$
$$(3.6)$$

As output, we consider the compliance [5]

$$s \equiv l(u) = \int_{\Gamma_R} g\,u\,. \qquad (3.7)$$

The accurate numerical discretization of (3.6) is standard, for instance using the Finite-Element (FE) method. In what follows, we shall invoke continuous piecewise linear approximations defined in conforming, regular FE spaces of $H^1(\mathcal{D})$.

To fix ideas, one could think of u as the temperature in a fin subject to a constant radiative flux

---

[5]For symmetric problems like (3.6), the output (3.7) is a particularly simple choice because it allows a simple accurate a posteriori error estimation without invoking a dual problem, but neither this choice nor the symmetric character of the problem are limitations to our approach, see e.g. [32].

g on Γ_R (in contact with a heat source) and to convective thermal exchanges on Γ_B, see Fig. 1. The function b determines the value of the Biot number along Γ_B, that is the intensity of local heat transfers on the part of the boundary in contact with a fluid. Now, the Biot number is typically uncertain. We shall model it as a random field in a probability space $(\Omega, \mathcal{F}, P)$ and one is typically interested in quantifying the uncertainty propagated in s. More precisely, we define an essentially bounded function $b \in L_P^\infty(\Omega, L^\infty(\Gamma_B, \mathbb{R}_{>0}))$, positive almost everywhere (a.e.) on Γ_B with probability 1 (P-almost surely or a.s. in abbrev.). Then, with a slight abuse of notation, u now denotes the solution in $L_P^\infty(\Omega, H^1(\mathcal{D}))$ to (3.1–3.2) when $b \in L_P^\infty(\Omega, L^\infty(\Gamma_B, \mathbb{R}_{>0}))$. And, for many values of the control parameter k_2, we consider computing expectations $E(s) = \int_\Omega s\, dP$ under P of the output s. To that aim, we will propagate the uncertainty (random "noise") from b to the scalar (random) output variable $s \in L_P^\infty(\Omega)$ by simulating the law of u. We recall that here we focus on MC discretizations of the noise. Compared with other discretizations [33, 34], it is very easy to implement, in particular because it is not intrusive with respect to existing discretizations of the deterministic problem, although the convergence is a priori much slower when the stochastic variations are smooth [12, 13].

We consider a bounded second-order stochastic process b as random input. We denote its constant mean value on Γ_B by $E(b|_{\Gamma_B}) \equiv \bar{E} > 0$, the spatial correlation length between random fluctuations by $\delta > 0$ and the relative intensity of the (bounded) random fluctuations around the mean by $\Upsilon > 0$. More precisely, like in [14], the random field b shall be the limit in $L_P^\infty(\Omega, L^\infty(\Gamma_B))$ of random fields b_K whose covariance defines a kernel operator of finite rank K in $L^2(\Gamma_B)$ and which can be easily simulated with a MC method using K inde-

pendent random scalar parameters in *a priori* given ranges. For the numerical simulations, we choose a gaussian kernel $c(x, y) = \exp\left(-|x - y|^2/\delta^2\right)/|\Gamma_B|$ to define a covariance operator in $L^2(\Gamma_B)$, where x, y are in the same connected component of the boundary Γ_B in Fig. 1. The covariance operator has eigencouples (λ_k, Φ_k), $k \in \mathbb{N}_{>0}$ and a normalized trace $\sum_{k=1}^\infty \lambda_k = 1$. Then, invoking uniformly distributed independent variates $Z_k \sim \mathcal{U}(-\sqrt{3}, \sqrt{3})$ for each spectral mode $k \in \mathbb{N}_{>0}$ (the same on each size of the boundary Γ_B in Fig. 1 for the sake of simplicity), we define b by its KL representation

$$b := \bar{E}\left(1 + \Upsilon \sum_{k \in \mathbb{N}_{>0}} \sqrt{\lambda_k}\, \Phi_k\, Z_k\right), \qquad (3.8)$$

which makes sense since truncated representations

$$b_K = \bar{E}\left(1 + \Upsilon \sum_{k=1}^K \sqrt{\lambda_k}\, \Phi_k\, Z_k\right) \qquad (3.9)$$

converge in $L_P^2(\Omega, L^2(\Gamma_B))$, and in $L_P^\infty(\Omega, L^\infty(\Gamma_B))$ here (see [35, 36] e.g.). We also require $b_K \geq b_{min} := \bar{E}/2 > 0$ a.s. $\forall K \in \mathbb{N}$ and fix $\Upsilon := .1$ for $\delta \in (.05, .5)$. Then one can define well-posed BVPs: Find $u_K \in L_P^\infty(\Omega, H^1(\mathcal{D}))$ such that P-a.s.

$$\kappa(\mathbf{n} \cdot \nabla)u_K + b_K\, u_K = g \text{ on } \partial\mathcal{D},$$
$$-\operatorname{div}(\kappa\nabla u_K) = 0 \text{ on } \mathcal{D}. \quad (3.10)$$

Approximations u_K converge to u in $L_P^\infty(\Omega, H^1(\mathcal{D}))$ as $K \to \infty$ and for a.e. $\omega \in \Omega$, realizations $u_K(\omega) \in H^1(\mathcal{D})$ of the solution to (3.10) solve a variational formulation with test function $v \in H^1(\mathcal{D})$

$$k_1 \int_{\mathcal{D}_1} \nabla u_K(\omega) \cdot \nabla v + k_2 \int_{\mathcal{D}_2} \nabla u_K(\omega) \cdot \nabla v$$
$$+ \bar{E} \sum_{k=1}^K (\Upsilon \sqrt{\lambda_k} Z_k(\omega)) \int_{\Gamma_B} \Phi_k\, u_K(\omega)\, v$$
$$+ \bar{E} \int_{\Gamma_B} e\, u_K(\omega)\, v = \int_{\Gamma_R} g\, v \quad (3.11)$$

parametrized by $k_1, k_2, \bar{E}$ and realizations $\Upsilon \sqrt{\lambda_k} Z_k(\omega)$, $k = 1 \dots K$. So finally, in practice, we shall compute approximations $u_{N,K}$ to u with the FE method defined previously, which converge in $L_P^\infty(\Omega, H^1(\mathcal{D}))$ as $\mathcal{N} \to \infty$, $K \to \infty$. We require $\mathcal{N}$ and K large enough such that, for a.e. values of the control parameter $k_1, k_2, \bar{E}$, a tolerance level $|s - s_{N,K}| \leq TOL|s|$ is reached a.s. for the approximate output $s_{N,K} := \int_{\Gamma_B} u_{N,K}$. Next one can easily simulate the law of $s_{N,K}$ with the MC method by mapping realizations $Z_k(\omega)$ of Z_k, $k = 1 \dots K$, to realizations $s_{N,K}(\omega)$ of $s_{N,K}$.

But a direct MC-FE method is very expensive computationally, since one has to compute many FE approximations $u_{N,K}(\omega)$, for many realizations of $Z_k(\omega)$, and next for many values of $\lambda = (k_2, \bar{E})$. That is why, like in [14], we replace the FE approximations $u_{N,K}$ by cheaper surrogates $u_{N,K,N}$ constructed with the standard RB method (wee will come back to this point in Section 3.3 and the non-expert reader can find a brief exposition of this technique there). At present, the RB method is one of the only few existing alternatives that can tackle some "high-dimensional" problems in more than 2 or 3 dimensions, and it proved useful[6] when the parameter is not too high-dimensional, for instance when each of the parameter components k_1, k_2, b and g are scalars. But when the parameter becomes high-dimensional, difficulties arise again at some point. To some extent, the RB method still showed efficient for MC simulations of a random field b with a moderately large number K of modes [14]. But there are situations where the approach is still computationally too expensive, in particular when one wants to explore variations of the MC simulation with respect to the control parameter k_2 and $\bar{E}$

---

[6]Note that it does not a priori require specific discretizations of the parameter space and speed-up can be increased for goal-oriented purposes like computing s accurately instead of u.

for instance. That is why, in this work, we would next like to show how to further improve the computational cost in cases where one is interested in computing many values of the expectation $E(s)$ of s under the law of the random field b for many values of the control parameters k_2 and $\bar{E}$. To this aim, we build on the standard RB ideas and use a MC method with a reduced basis of control variates. In particular, at a given value of λ, one still needs a large number of RB approximations $u_{N,K,N}$ to compute accurate MC estimations of $E(s_{N,K,N})$, which is too costly when it has to be done for many values of λ. We thus now try to reduce the computational cost by invoking still another computational reduction technique, the RB control-variate MC method. We will show that the number of realizations required by the MC method to reach a given statistical accuracy for $E(s_{N,K,N})$ at one parameter value λ can be efficiently decreased with our RB control-variate technique, in the limit where one has to compute sufficiently many expectations for sufficiently many values of λ.

3.2. RB control-variate MC estimations

We require a relative precision $TOL = 10^{-3}$ for the numerical approximations $E(s_{N,K,N})(\lambda)$ of the output expectations $E(s)(\lambda)$ at any $\lambda \in \Lambda := (.1, 10) \times (.1, 1)$. It is thus sufficient for the PDE solution u to be a.s. approximated with a relative precision $tol = 10^{-2}$ here. To this aim, we first define $\mathbb{P}_1$ FE approximations on a 2D simplicial mesh using FreeFem++ by F. Hecht and his collaborators. (Refining a coarse mesh, $\mathcal{N} = 3333$ nodes – 6272 triangles – seem sufficient.)

We first fix $\delta = 0.5$. By a standard "variational crime" analysis (see e.g. [14]), it is enough to use a truncated KL expansion b_K up to $K = 10$ (such that $\sum_{k>K} \sqrt{\lambda_k} \leq tol \sum_{k \in \mathbb{N}_{>0}} \sqrt{\lambda_k}$). We build "offline" (that is, before the MC simulation) a reduced basis
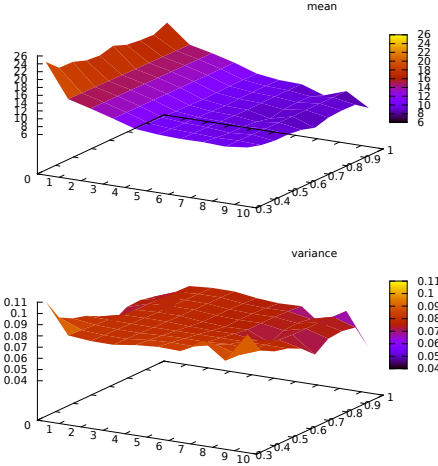
Figure 2: Maps of $E\left(s(k_2, \bar{E})\right)$ and $V\left(s(k_2, \bar{E})\right)$ (numerical approximations) for $(k_2, \bar{E}) \in (.1, 10) \times (.1, 1)$.

for a small linear space[7] of surrogates $u_{N,K,N}$ with dimension $N = 12$. Then we simulate the law of $Z^\lambda \equiv s_{N,K,N}(\lambda)$ at any $\lambda \in \Lambda$ with the MC method. This allows us to retrieve a MC estimation of $E(Z^\lambda)$ and to construct a response surface by piecewise linear interpolation of the MC estimations $E_M(Z^\lambda)$ at the 10×10 trial values of λ like in Fig. 2.

To compute the response surface of Fig. 2 with a good confidence level in the MC estimations, say with $\sqrt{V_M(Z^\lambda)/M} \leq TOL \times E_M(Z^\lambda)$ at each of the λ used for interpolation, one needs more than $M = 10^4$ realizations[8] at each λ. Now, the law of Z^λ with respect to λ is smooth. Let us use the RB

control-variate MC method with $\tilde{\Lambda}$ defined by the 10×10 trial values of λ needed to build the response surfaces in Fig. 2. Then, we can achieve MC estimation with confidence intervals of accuracy TOL using a large number of realizations $M_{\text{large}} = 10^4$ at only $I = 3$ values of λ, and much less realizations at the other values of λ to achieve satisfying confidence intervals. More precisely, $M_{\text{test}} = 10$ is enough, since one can reduce the variance to 10^{-4} with $I = 3$, and one already has a good MC estimation of the magnitude of the variance then. One also needs to compute coefficients α_i^λ at each λ with M_{small} additional realizations, but in practice one can use the same $M_{\text{test}} = M_{\text{small}}$ realizations[9]. So the marginal gain for each λ in the finite sample of queries $\tilde{\Lambda}$ is to divide the computational cost by $10^4/(10+I\times10^4/Card(\tilde{\Lambda})) \geq 30$ (the cost C of one realization is dominant here, even with RB surrogates).

Moreover, our algorithm is robust with respect to the specific choice of $\tilde{\Lambda}$ above. That is, the control variates identified previously can still be used to reduce the variance of $E_M(Z^\lambda)$ at other λ in Λ than those in the trial sample $\tilde{\Lambda}$ used by the greedy algorithm above. Let us evaluate the gain for real-time purposes or in the limit of many queries in λ when one can forget all the precomputations including the greedy construction. Then the computational cost per λ is approximately divided by $10^4/10 = 10^3$, provided the reduced variance is still approximately of magnitude 10^{-4} for all $\lambda \in \Lambda$. This observation even holds if we require a higher precision (hence smaller tolerances TOL, tol and correpons-

---

[7]A weak greedy algorithm needed $N = 12$ steps for the a posteriori error estimates to satisfy $\|u_{N,K} - u_{N,K,N}\|_{H^1} \leq \Delta_N \leq tol = 10^{-2}$ at all training parameter values in a cartesian grid using 2 trial values for each parameter component $\Upsilon \sqrt{\lambda_k}Z_k$, $k = 1 \dots K$, and 10 trial values for each parameter component $\lambda_1 = k_2$, $\lambda_2 = \bar{E}$, see Section 3.3 for some details about the RB method.

[8]Besides, one can check that $M = 10^4$ is sufficiently-many realizations here for the speed of convergence to be quite well evaluated by CLT, since the MC estimator $E_M(Z^\lambda)$ is already close to gaussian according to the Kolmogorov-Smirnov goodness-of-fit test. In particular, most often, one single draw

of the MC estimator already allows one to compute a good approximation to a confidence interval.

[9]Although introducing some dependency between the M_{test} and M_{small} realizations theoretically influences the possibilty to obtain a CLT for instance, it hardly changes siginificantly the numerical result for one MC estimation in practice.

dingly increase "offline" the dimension N of the RB space for the PDE surrogates). Then, the gain grows like TOL^{-2} in the infinitely-many-query asymptotics as long as one can reduce the variance by adding finitely-many control variates; and like $TOL^{-2}(1 - I \times TOL^{-2}/Card(\tilde{\Lambda}))$ if one takes into account the greedy construction for a finite sample of $Card(\tilde{\Lambda}) \gg I \times TOL^{-2}$ parameter values, where I is the minimal number of control variates required to achieve a reduced variance of magnitude TOL in $\tilde{\Lambda}$. (Note that the computational cost of one realization becomes even more dominant as N increases.) Here, Fig. 3 shows that that our greedy construction is robust: with control variates built using the 10×10 trial values of λ above, we did thorough MC estimations of the variance for 100 other parameter values. In our example, it roughly holds $I \leq -\log(TOL)$ for $TOL > 10^{-7}$ and the computational gain grows exponentially fast like $TOL^{-2}(1 + TOL^{-2}\log(TOL)/Card(\tilde{\Lambda}))$ with the number $-\log(TOL)$ of accurate digits required for sufficiently large $Card(\tilde{\Lambda})$.

In any case, if one cannot reduce the variance (or equivalently, if I needs to be huge for any useful variance reduction), there is – almost – no loss of effort in using the RB control-variate MC method compared with the naive MC method, since would simply observe the absence of variance reduction using M_{test} realizations and would next need to increase the number of realizations to the same number of realizations as the naive MC method. Only the computation of coefficients α_i^λ would have been unnecessary, which is not much compared with a large number of realizations.

Of course, the efficiency of our RB control-variate MC method is limited by the cost and accuracy of single realizations, like any MC method. This has nothing to do with its ability at reducing the variance (and thus controlling the number of
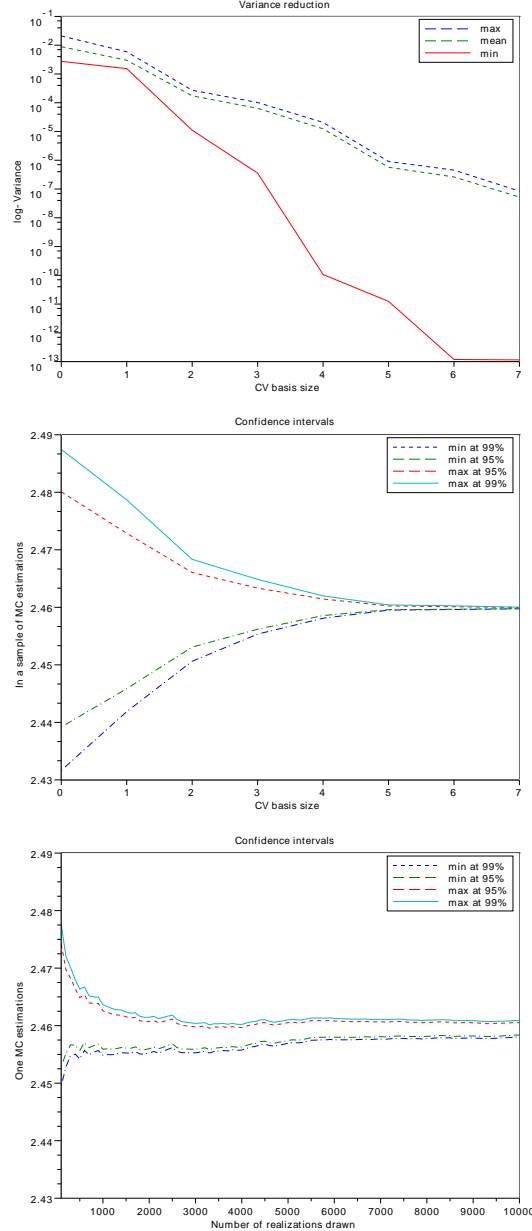

Figure 3: Top: maximum, mean and minimum in a sample of MC estimations $V_{M_{\text{test}}}(Z^\lambda - \sum_{i=1}^I \alpha_i^\lambda \tilde{Y}^i)$ at various λ, as a function of I. The decrease rate is less fast among parameter values λ not used by the greedy algorithm, but it is still reasonably good. Middle: confidence intervals with probability 99% and 95% for the MC estimation $E_{M_{\text{test}}}(Z^\lambda - \sum_{i=1}^I \alpha_i^\lambda \tilde{Y}^i)$ of $E(Z^\lambda)$ as a function of I at one λ. Bottom: confidence intervals with probability 99% and 95% for the MC estimation $E_M(Z^\lambda)$ of $E(Z^\lambda)$ as a function of M at same λ, with a much slower decrease rate than above at a similar computational cost.

realizations needed for accurate MC estimations), but it is a practical limitation everyone doing simulations is confronted with. For UQ in particular, if we use a smaller correlation length $\delta = 0.05$ but require the same tolerance TOL, the "variational crime" analysis above requires a KL truncation at a much higher-order $K = 70$. Now, it is practically impossible today (in 2012) to carefully inspect all $2 + K = 72$ directions with a fine grid of trial values for the "offline" construction of RB surrogates[10]. So, even though the RB control-variate MC method still works perfectly well with inaccurate RB surrogates $u_{N,K,N}$, there is little computational gain possible on the whole simply by reducing the variance when the statistical error is dominated by the KL truncation error (there is no point in getting accurate MC confidence intervals if the RB surrogates are not good enough). The RB control-variate MC method is not a definitive cure to the "high-dimensions" problem for PDEs with stochastic coefficients which we already mentionned previously.

All existing numerical approaches to UQ problems that invoke a KL decomposition of the input noise are limited in practice by the "curse of dimensionality" in K anyway, in so far as they require numerical approximations of realizations of the PDE solution in a space of high dimension K. (Most often in UQ people only use $K \leq 10$ in practice [35, 7, 37].) So is our approach. But in any case, when one can tackle the problem of simulating high-dimensional realizations, say only by expensive precomputations, then using our RB control-variate MC method and the PDE solution as a black-box still makes sense: the simulation

of one realization of Z^λ at one given λ is typically all the more expensive as the parameter size increases, for instance just by assembling the matrix of the discrete BVP here, and the gain obtained by variance reduction (if any) is in fact all the more interesting. Moreover, there exist other ways of simulating the law of b that do not require a KL representation (thus a limitation in K), even extend to other random fields than (3.8), and can be used with our RB control-variate MC method [38]. This might be a cure to some problems but we keep its investigation for future works of ours. Last, for the model problem here, we are in fact quite fortunate, as already noted in [14], that the KL spectrum has a fast decay. The first KL modes span all the solution space when the required precision is not too small. Then we can still build "offline" good enough surrogates $u_{N,K,N}$ with a very small training set of trial parameter values, as we are going to see in the next section. Furthermore, since the previous observation holds only for specific choices of the spatial correlations, before shifting to the numerical proof of efficiency for the RB control-variate MC method in another UQ framework, we are also going to mention one promising way of tackling the simulation problem with high-order KL truncations which is currently under study, and where the RB control-variate MC method can still be useful. Our conclusion of the first example is thus mostly concerned with an efficient combination of the standard RB method with the RB control-variate MC method in a UQ framework.

3.3. RB control-variate MC with RB surrogates

We briefly recall the "standard" RB method used above (see also [39, 40, 41] e.g.) before discussing its specific use in UQ with a MC method like our RB control-variate MC method. The reader already expert in the standard RB method may want to skip the next subsection.

---

[10]The computations presented in this work have been done in a very reasonable time, from a few minutes to one hour on a single processor unit, but inspecting 72 directions with 2 points per directions, that is $2^72 \approx 10^{21}$ points, would require yers of computations.

3.3.1. The standard RB method

The goal of the RB method is to reduce the computational cost of solutions $u(\mu)$ to a PDE parametrized by μ when $u(\mu)$ has to be computed many times for many values of μ. Here we set $\mu = (k_2, \bar{E}, \Upsilon \sqrt{\lambda_1} Z_1(\omega), \ldots, \Upsilon \sqrt{\lambda_K} Z_K(\omega))$ and given a Hilbert space $X = H^1(\mathcal{D})$ with inner-product

$$(u, v)_X = \int_{\mathcal{D}} \nabla u \cdot \nabla v + \int_{\Gamma_B} uv, \ \forall u, v \in X$$

and with energy norm $\|u\|_X = \sqrt{(u, u)_X}$, we denote

$$a(u, v; \mu) = k_1 \int_{\mathcal{D}_1} \nabla u \cdot \nabla v + k_2 \int_{\mathcal{D}_2} \nabla u \cdot \nabla v + \int_{\Gamma_B} buv$$

a bilinear form that is elliptic $\forall u, v \in X$. Then, for parameter values μ in a given range, we consider the solutions $u(\mu) \in X$ to (3.6), such that

$$a(u(\mu), v; \mu) = l(v), \ \forall v \in X, \qquad (3.12)$$

with outputs $s(\mu) = l(u(\mu))$. In practice, we assume that one can compute a good approximation of $u(\mu)$ in a linear subspace $X_{\mathcal{N}} \subset X$ with large dimension $\mathcal{N} \gg 1$ for any μ. Moreover, here we discretize $b \approx b_K$ so the fully computable approximations also involve a discretization parameter K and is denoted $u_{\mathcal{N},K}(\mu) \in X_{\mathcal{N}}$. Computing $u_{\mathcal{N},K}(\mu)$ for many μ is costly and the goal of the RB method is to construct a linear subspace $X_N \subset X_{\mathcal{N}}$ with small dimension $N \ll \mathcal{N}$ such that, for all μ in the given range, some approximations $u_{\mathcal{N},K,N}(\mu) \in X_N$ to $u(\mu) \in X$ can be computed faster. The RB method invokes linear approximations $u_{\mathcal{N},K,N}(\mu) := \sum_{n=1}^{N} \gamma_n(\mu) u(\mu_n)$ where the coefficients $\gamma_n(\mu)$, $n = 1 \ldots N$, are computed fast at any $\mu \in \Lambda$ by the Galerkin method in X_N: $m = 1 \ldots N$

$$\sum_{n=1}^{N} \gamma_n(\mu) a(u(\mu_n), u(\mu_m); \mu) = l(u(\mu_m)). \quad (3.13)$$

The Galerkin approximation error $\|u_{\mathcal{N},K}(\mu) - u_{\mathcal{N},K,N}(\mu)\|_X$ can be estimated *a posteriori* by a fully computable error estimator $\Delta_N(\mu)$

$$\|u_{\mathcal{N},K} - u_{\mathcal{N},K,N}\|_X \leq \sup_{\|v\|_X = 1} \frac{a(u_{\mathcal{N},K,N}(\mu), v; \mu) - l(v)}{\alpha_{LB}(\mu)}$$
$$(3.14)$$

using a lower bound $\alpha_{LB}(\mu)$ for the coercivity constant of the bilinear form, which is useful for the certifiability of the reduction method as well as for the selection of adequate parameter values μ_n, $n = 1 \ldots N$, in the range of interest. Here, given the choice of inner-product on $X = H^1(\mathcal{D})$, one can compute explicitly and at little expense quite a sharp lower bound $\alpha_{LB}(\mu) = \min(k_1, k_2, \inf b)$ that is uniformly good *with respect to the parameter values* and for any admissible parameter range.

A good *reduced basis* $u(\mu_1), \ldots, u(\mu_N)$ spanning X_N can be found in practice with a greedy algorithm like (2.11) [27, 28]. The standard procedure selects N parameter values $\mu_1, \ldots, \mu_N$ *incrementally* in a trial sample Θ for μ as follows: given any μ_1 we next define for $n = 1, \ldots, N - 1$

$$\mu_{n+1} \in \underset{\mu \in \Theta}{\mathrm{argsup}} \ \Delta_n(\mu). \qquad (3.15)$$

Note that the bilinear form a is affine with respect to functions of the parameters (viz. parametrized by μ only through coefficients in a linear combination of bilinear forms, recall (3.11)). This is useful to efficient RB implementations in practice since the parameter-independent matrices can be precomputed. It allows fast evaluations[11] of $u_{\mathcal{N},K,N}(\mu)$ and $\Delta_N(\mu)$ at any μ. Here, the quality of our RB approximations is similar to that in [14].

---

[11]Denoting by M the matrix of the inner-product $(\cdot, \cdot)_X$ computed for the discrete FE space where the parametrized BVP reads $A(\mu)U(\mu) = B$, the RB approximation error in the "energy" norm $\sqrt{(U(\mu) - U_n(\mu))^T M (U(\mu) - U_n(\mu))}$ is evaluated by the residual estimator $\sqrt{(A(\mu)U_n(\mu) - B)^T M^{-1}(A(\mu)U_n(\mu) - B)}/\alpha_{LB}(\mu)$. Now, $(A(\mu)U_n(\mu) - B)^T M^{-1}(A(\mu)U_n(\mu) - B)$ is a quadratic form in the small-dimensional vector $U_n(\mu)$ containing the coefficients of the approximation in the reduced basis. And its entries are themselves quadratic in (functions of) μ when the entries of

3.3.2. Certified RB control-variate MC

First, note that to get certified results in the frame of MC simulations, one should check the quality of those RB approximations when they are used, since in any case the precomputed basis for the RB surrogates invoked at each λ and for each realization has been trained "offline" only with a few trial parameter values in Θ (possibly forgotten afterwards). In the RB control-variate MC method, it is necessary to check the RB approximation error with (3.14) at least for the M_{large} realizations in $E_{M_{\text{large}}}(Z^{\lambda_i})$, $i = 1 \ldots I$, at the parameter values λ_i where control variates are constructed plus, for each λ, at the $(1+I)M_{\text{test}}$ realizations of Z^λ and Z^{λ_i}, $i = 1 \ldots I$, that are used to compute the final MC estimation with reduced variance. In practice here, this increase of computational cost per realization and λ was not much (that is, only a fraction of the total computational time without certification).

Moreover, for $\delta = 0.5$ and $K = 10$, we precomputed a reduced basis of dimension $N = 12$ for the PDE solutions using only a cartesian grid of $10 \times 10 \times 2^K$ values for μ. Although this is a coarse training sample, we never had to enrich the RB approximation space X_N "online", that is during the MC simulations. Thus, for UQ problems similar to our example, although our MC approach does not aim at exploiting optimally the regularity of the solution, in particular because the choice of trial parameter values for a greedy algorithm is "blind", the RB control-variate MC method is clearly interesting: it is simple to implement, without much a priori knowledge of the solution, and accurate at a reasonable computational cost ("offline" and "online" computations take a few minutes on a single processor unit here).

$A(\mu)$ depend linearly in (functions of) μ. The latter can thus be precomputed for a given reduced basis and next be assembled in $O(N^2 K^2)$ to yield fast error estimations at any μ.
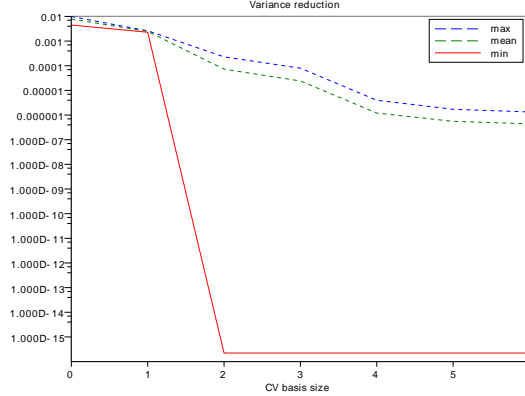

Figure 4: Maximum, mean and minimum in a sample of MC estimations $V_{M_{\text{test}}}(Z^\lambda - \sum_{i=1}^{I} \alpha_i^\lambda \tilde{Y}^i)$ at various λ, as a function of I. The decrease rate is less fast among parameter values λ not used by the greedy algorithm, and less fast than in Fig. 3.

For $\delta = 0.05$, we mentionned previously that the problem is not to reduce the variance of MC estimations, but to get certified MC estimations when one uses a KL representation of the input noise at each of the MC realizations, because the parametrization of the latter needs approximately $K = 70$ dimensions, which cannot be explored finely "offline" in a reasonable time (one manifestation of the well-known "curse of dimensionality"). One way out is to simulate differently the noise, in a way that could invoke RB surrogates in a low-dimensional parameter space only [38]. We shall study this approach in future works. Another way if we want to use KL representations is, first, to observe that since the error per realization is averaged in $E(s_{N,K,N}(\lambda))$ one can still afford some realizations with large error as shown in [14]. Moreover, since the solution space very much depends on the first KL modes only as observed in [14], one can build here "offline" surrogates $u_{N,K,N}$ in a space of low dimension. Here, we could use RB surrogates from a small linear space X_N of dimension $N = 20$ without having

to enrich X_N "online", although X_N was built "of-fline" with a greedy algorithm exploring only a very coarse training sample of trial parameter values (using 10 trial values for each parameter component $\lambda_1 = k_2, \lambda_2 = \bar{E}$ and 2 trial values for each parameter component $\Upsilon \sqrt{\lambda_k} Z_k$, $k = 1 \ldots 10$, while for $k = 11 \ldots 70$ the parameter components were fixed). Then, compared with a direct "naive" MC simulation of the law of $Z^\lambda \equiv s_{N,K,N}(\lambda)$ at any $\lambda \in \Lambda$, the RB control-variate MC method provides computational reduction by reducing the variance of MC estimations like in the case where $\delta = 0.5$, see Fig. 4 in comparison with Fig. 3. Note yet that the variance $V(Z^\lambda)$ is approximately one order of magnitude smaller when $\delta = 0.05$ than in the previous example when $\delta = 0.5$, and the variance decreases a bit more slowly than before, so the gain at same accuracy is a bit less than one order of magnitude smaller (and non zero only if we require a minimum precision on the confidence interval of the MC estimations of a bit more than one order of magnitude higher than when $\delta = 0.05$).

Last, we would like to mention a way to reach higher KL truncation order K which is not specific to our choice of the correlation length and where the RB control-variate MC method can still be useful. Note first that, quite often, surrogates $u_{N,K,N}(\lambda)$ built for a coarse training sample (say with only 1 trial parameter value in most directions) are already not too bad because the KL spectrum decays anyway. So, at any λ, one can expect $V(s_{N,K,N'}(\lambda) - s_{N,K,N}(\lambda))$ to be significantly smaller[12] than $V(s_{N,K,N}(\lambda))$ and next improve the MC estimations of $E(s_{N,K,N}(\lambda))$ "on-line" at a very reasonable cost, by the combination of the RB control-variate MC method and the Multi-Level Monte-Carlo (MLMC) method [42] in two steps: estimate (i) $E(s_{N,K,N}(\lambda))$ with the RB control-variate MC method and (ii) $E(s_{N,K,N'}(\lambda) - s_{N,K,N}(\lambda))$ with a small number M'_{test} of realizations. The dimension $N' \geq N$ of the enriched basis at λ can be chosen a posteriori in (ii) from the few M_{test} realizations in the MC estimation $E_{M_{test}}(s_{N,K,N}(\lambda) - \hat{Y}^\lambda)$ of (i). In turn, the actual I in (i) can be re-adjusted after (ii) to balance the statistical error in $E_{M'_{test}}(s_{N,K,N'}(\lambda) - s_{N,K,N}(\lambda))$ such that $V_{M_{test}}(s_{N,K,N}(\lambda) - \hat{Y}^\lambda)/M_{test} \approx V_{M'_{test}}(s_{N,K,N'}(\lambda) - s_{N,K,N}(\lambda))/M'_{test}$ if possible. Of course, the latter new strategy deserves a specific study, which we also keep for future works.

4. Application to Bayesian estimation

Let us now numerically demonstrate the efficiency of the RB control-variate MC method in another context useful for UQ, with parameterizations of higher dimension than in the previous section. As example, we consider the computation of a Minimum-Mean-Square-Error (MMSE) Bayes estimator in various parametric settings. To identify various possible cases with many queries in the parameter, we first use a toy-model. Second, we use the same model PDE problem (3.1–3.2) as in the previous section. In the latter case, we also combine the RB control-variate MC method with standard RB surrogates of the PDE solutions, like in the previous section. But of course, the use of surrogate models in Bayesian estimation is not new [38, 20].

4.1. A toy model for various Bayesian estimations

The Bayesian estimation of some unknown input parameter θ in a model consists in improving a prior guess of its probability distribution π^ξ

---

[12]In our numerical example above, $V_{M'_{test}}(s_{N,K,N'}(\lambda) - s_{N,K,N}(\lambda))$ is in fact always close to zero machine by the property mentionned above: at a moderately high precision level, the solution space is almost entirely spanned by variations in the first K modes only. But of course this would not be true for random fields with thicker distribution tails.

(the probabilistic counterpart of the deterministic Tikhonov regularization in inverse problems, see [9] e.g.) by observations of an output $s^\lambda(\theta)$ of the model that are viewed as realizations of a random variable [43]. We denote by λ a control parameter entering the model, and by ξ a so-called *hyper-parameter* in the Bayesian frame. The random variations of $s^\lambda(\theta)$ can be generated by *aleatoric noise* (for a given fixed value of θ, the truth is stochastic and induces a distribution of value $s^\lambda(\theta)$) as well as the *epistemic* uncertainty about the parameter θ (our observations of the truth are noised).

When one knows the value of a control parameter λ in a model, as well as J observations s^λ_j, $j = 1 \ldots J$ (J i.i.d. realizations of $s^\lambda(\theta)$) with likelihood $f^{\lambda,\zeta}(s^\lambda_j|\theta)$, the probability density of s^λ knowing θ), then Bayes formula allows one to compute the posterior distribution of θ

$$\pi^{\lambda,\zeta,\xi}(\theta|\{s^\lambda_j\}) \propto \pi^\xi(\theta) \prod_{j=1}^{J} f^{\lambda,\zeta}(s^\lambda_j|\theta) \qquad (4.1)$$

where ζ is another hyperparameter (entering the likelihood and not the prior like ξ). The posterior is used to compute the probability distribution of quantities that depend on θ. It is also used to compute a deterministic approximation $\hat{\theta}^{\lambda,\zeta,\xi}(\{s^\lambda_j\})$ of θ given $\{s^\lambda_j, j = 1 \ldots J\}$, for instance the Minimum-Mean-Square-Error (MMSE) estimator [44, p.349], which in decision theory[13]

is interpreted as the minimum of the expected quadratic loss $E(|\hat{\theta}^{\lambda,\zeta,\xi}(\{s^\lambda_j\}) - \theta|^2)$, averaged over both distributions of θ and $\{s^\lambda_j, j = 1 \ldots J\}$

$$\hat{\theta}^{\lambda,\zeta,\xi}(\{s^\lambda_j\}) := \int \theta \, \pi^{\lambda,\zeta,\xi}\left(\theta|\{s^\lambda_j\}\right) d\theta. \qquad (4.2)$$

In practice, expectations using (4.1) like (4.2) are typically computed many times:

- for many values of the control parameter λ entering the model for the "truth",

- using many different sets of observations $\{s^{\lambda_0}_j, j = 1 \ldots J\}$, typically at $\lambda_0 = \lambda$ but not necessarily, and

- for many values of the hyperparameters ξ, ζ entering the Bayesian frame (prior and likelihood are often chosen in – conjugate – parametric families of distributions).

Think of the various pieces produced in a factory, or of real-time estimation procedures, where one may want to vary all these parameters ! We would thus like here to accelerate the computation of many parametrized MMSEs in many-query frameworks with our RB control-variate MC method and numerically explore the various parametrized settings with a toy-model first.

Let us choose a Bayesian frame for supposedly Gaussian observations of θ with known mean-square deviation λ, typically

$$f^{\lambda,\zeta}(s|\theta) \propto e^{-\frac{|s-\theta|^2}{2\lambda^2}} \qquad \pi^{\xi=(\mu,\sigma^2)} = \mathcal{N}(\mu, \sigma^2).$$

Gaussian prior and likelihood are conjugate one-another and the MMSE is analytically computable

$$\hat{\theta}^{\lambda,\zeta,\xi}(\{s^\lambda_j\}) = \frac{\sigma^2}{\sigma^2 + \lambda^2/J}\left(\frac{1}{J}\sum_{j=0}^{J} s^\lambda_j\right) + \frac{\lambda^2/J}{\sigma^2 + \lambda^2/J}\mu$$

with variance $\frac{\sigma^2 \lambda^2/J}{\sigma^2 + \lambda^2/J}$ (see [44, p.353] e.g.). Note that for this simple model, there is no degree of
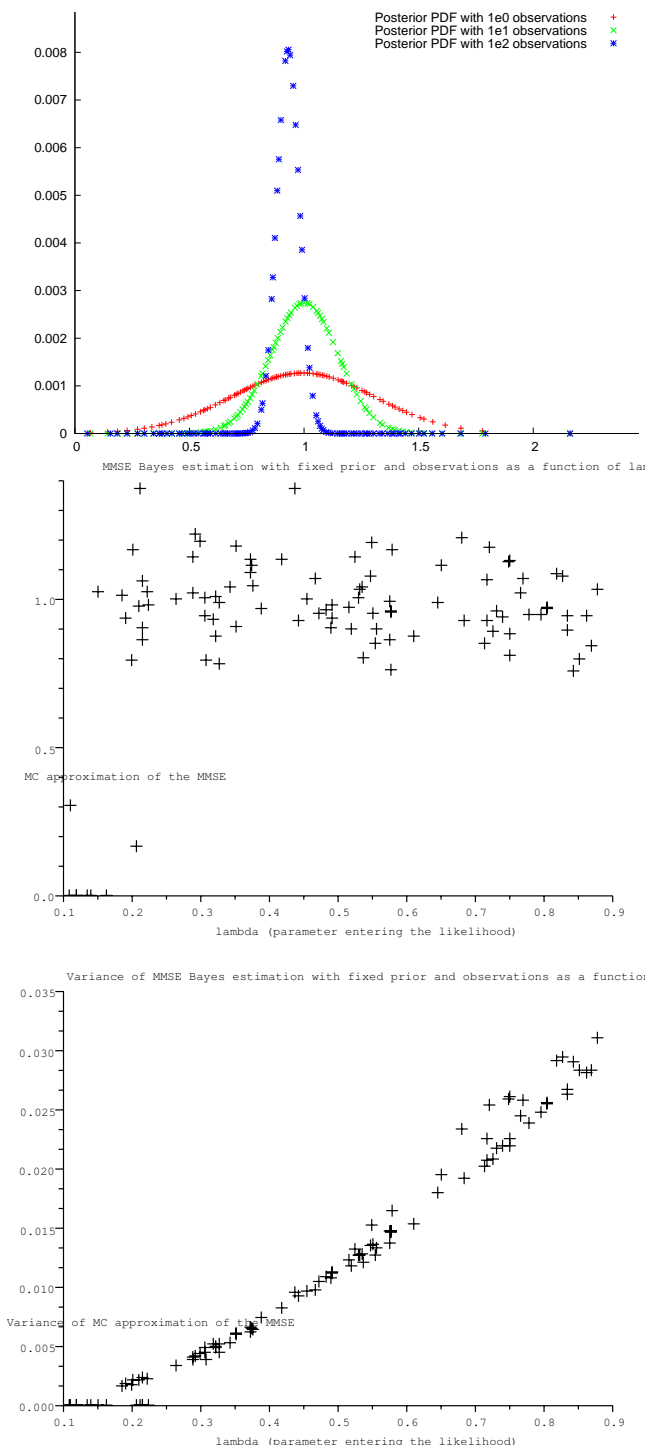
Figure 5: Posterior distributions computed for $J = 1, 10, 100$ i.i.d observations $s_j \sim \mathcal{N}(\theta_0 = 1, \lambda^2 = .25)$ (top) ; and variations of the MC MMSE Bayes estimator (middle) and its variance (bottom) as a function of λ for 10^3 realizations.

freedom for a hyperparameter ζ, but one already sees why one may want to use various values of the hyperparameter ξ. Although the MMSE Bayes estimation is asymptotically consistent and efficient (in fact, normally distributed) as $J \to \infty$, the quality of the estimation is clearly impacted by the choice of the hyperparameter at finite J. See the posteriors computed with $J = 1, 10, 100$ observations and $\theta_0 = 1, \lambda = .5, \mu = .9, \sigma = .4$ in Fig. 5.

More generally, one can show under regularity assumptions on the likelihood $f^{\lambda,\zeta}(\cdot|\theta)$ (see [45, p.490] *e.g.*) that, if the observations s_j^λ are indeed realizations of a random variable $s^\lambda(\theta)$ with density $f^{\lambda,\zeta}(\cdot|\theta)$ for one fixed value $\theta = \theta_0$ (hence distributed only due to the aleatoric noise of the model), then the MMSE converges in distribution when the number J of observations increases

$$\sqrt{J}(\hat{\theta}^{\lambda,\zeta,\xi} - \theta_0) \xrightarrow{d} \mathcal{N}(0, I(\theta_0)^{-1}),$$

with variance given by the *Fisher information* $I(\theta_0) = E\left(\left(\partial_\theta \log f(\{s_j^\lambda\}|\theta)\right)^2 |\theta = \theta_0\right)$. The MMSE estimator is thus *asymptotically consistent* and *asymptotically efficient* and the choice of the prior becomes irrelevant as $J \to \infty$. Yet, for finite sample of observations, a good MMSE Bayes estimation strongly depends on the choice of the prior, in particular because MMSE Bayes estimators are biased. So one may try to optimize the hyperparameters given training samples of observations, possibly for various values of the control parameter: a natural many-query problem for the MC computation of parametrized MMSEs ! Let us see how the RB control-variate MC method performs in various parametrized settings.

We do many MC estimation of the MMSE with $J = 10$ i.i.d. synthetic observations generated with distribution law $\mathcal{N}(\theta_0, \lambda_0^2)$ and $\theta_0 = 1, \mu = .9, \sigma = .4$. We tested the RB control-variate MC method for various meaningful parametric variations:

20

- $\lambda \in (.1, .9)$ with one set of observations generated at $\lambda_0 = .5$ fixed or at $\lambda_0 = \lambda$,

- $\lambda \in (.1, .9)$ and many sets of observations generated at $\lambda_0 = .5$ fixed or at $\lambda_0 = \lambda$,

- $\xi \equiv (\mu, \sigma) \in (.5, 1.5) \times (.1, .9)$, $\lambda \in (.1, .9)$ and many sets $\{s_j^{\lambda_0}, j = 1 \ldots J\}$ of observations generated at $\lambda_0 = .5$ fixed or at $\lambda_0 = \lambda$.

We always obtained meaningful estimations, and good variance reductions (more than 10 orders of magnitude) with only a few variates (less than 10). We show in Fig. 6 the "most difficult" case, which is still of course quite easy because of the very smooth dependence of our toy-model on the parameter. But this shows at least that even in potentially high-dimensional parameter contexts, there is some hope for the RB control-variate MC method to be useful.

4.2. A PDE model with uncertain coefficients

We now consider the PDE problem (3.1–3.2) as a parametrized numerical model for Bayesian estimation. It is discretized like in the previous section: given δ, one fixes the discretization parameters $\mathcal{N}$, K and N. Note that we invoke the standard RB method [40, 23] to compute fast numerous solutions u to (3.1–3.2) at each of the numerous values of the PDE coefficients invoked, which has already been used in a Bayesian estimation context with a deterministic quadrature formula for the output expectation [20] rather than MC.

We still denote $\lambda = (k_2, \bar{E})$ the parameter in the model. In addition to s^λ defined in (3.7), we also consider the temperature averaged at the top of the fin $o^\lambda = \int_{\Gamma_N \cap \overline{\mathcal{D}_1}} u$ as output. Clearly, there are many situations where the uncertain coefficient b can a priori be modelled using only some rough representation as a random field (3.8). Then, it often needs improving in a real setting using data from
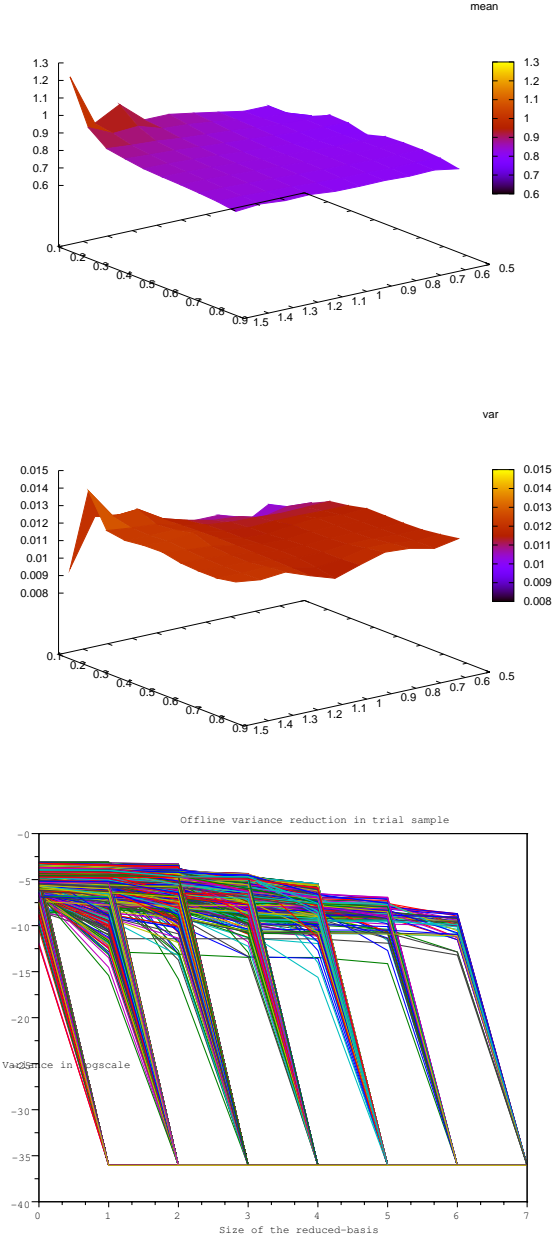

Figure 6: Variations of MC MMSE Bayes estimation (top) and their non-reduced empirical variance (middle) as functions of (μ, σ), using one set of observations and one λ. Bottom: empirical MC estimations of the reduced variances as functions of the number I of variates for 20^3 values of the parameter $(\{s_j^{\lambda_0}, j = 1 \ldots J\}, \lambda, \zeta \equiv (\mu, \sigma))$. Each colored line is the log-variance for one parameter value.

21

observations. This is the case of the permeability field for the Darcy equations in [38] e.g.

Here, we shall try to improve i.i.d. *Gaussian* priors[14] $Z_k \sim \pi^\xi \equiv \mathcal{N}(0, \xi_k)$, $\xi_k \equiv \sigma_k^2$, for the random variables Z_k, $k = 1 \ldots K$, using J i.i.d. observations $o_j^{\lambda_0}$, $j = 1 \ldots J$, of the ouput o^{λ_0} at some value λ_0 of the control parameter. Assuming one knows that the likelihood $f^{\lambda, \zeta}(o_j^{\lambda_0} | \{Z_k\})$ of the observations $o_j^{\lambda_0}$ given a realization $\{Z_k\}$ of the set of random variables Z_k, $k = 1 \ldots K$, are also Gaussian $f^{\lambda, \zeta}(o_j | \{Z_k\}) \propto exp(-|o_j - o^\lambda(\{Z_k\})|^2 / \zeta)$, $\zeta \equiv \sigma_O^2$, then the posterior distributions of the independent random variables Z_k can be used to describe the specific context associated with the observations $o_j^{\lambda_0}$, $j = 1 \ldots J$. (They are a priori better than the prior guesses and should become all the better as the number of observations increases.) In particular, the model posterior, given by the Bayes formula $\prod_{k=1}^K \pi_k^{\lambda, \zeta, \xi}(Z_k | \{o_j^{\lambda_0}\}) \propto \prod_{j=1}^J f^{\lambda, \zeta}(o_j^{\lambda_0} | \{Z_k\}) \prod_{k=1}^K \pi_k^\xi(Z_k)$, can be used to compute various quantities of interest, like the expectation of s^λ. Let us use the RB control-variate MC method to evaluate fast $E_{\pi^{\lambda, \zeta, \xi}}(s^\lambda | \{o_j^{\lambda_0}\})$ for various values of the hyper parameters ξ, ζ, various values of the control parameter λ in the model and possibly various sets of observations (at $\lambda_0 = \lambda$ possibly, but not necessarily). Quite importantly, note that here we assumed that we know explicitly the likelihood function $f^{\lambda, \zeta}(\cdot | \{Z_k\})$ by experience. This is a specific Bayesian framework which is often used, although the exact likelihood function is often not known in practice (see Rem. 1).

We use the RB control-variate MC method to compute the MMSE of s^λ with the parametrized

---

[14] For well-posedness of the PDE we will need to numerically truncate the Gaussian realizations by excluding those realizations where $b \leq 0$. This is a poor man's "truncated Gaussian", but in practice the standard deviations σ_k are chosen sufficiently small for most Z_k realizations to be in the range of admissibility where the surrogate model is valid.

posterior above in different settings for $\delta = .5$, $K = 10$, $N = 12$. For instance: (i) with $4 \times 4 \times 2^K = 2^1 2$ parameter values $\left(k_2, \bar{E}, \sigma_k^2\right) \in [.1, 10] \times [.1, 1] \times [10^{-4}, 10^{-3}]$ and one fixed $J = 1$ observation $o_1^{\lambda_0}$ generated synthetically by uncertainty propagation at $\lambda_0 = (2, 0.5)$ or (ii) with 10×10 parameter values $\left(k_2, \bar{E}\right) \in [.1, 10] \times [.1, 1]$ and 10 values for each of the $J = 3$ observations $o_j^{\lambda_0}, j = 1 \ldots J$, generated synthetically by uncertainty propagation at $\lambda_0 = (2, 0.5)$. In both cases, the RB control-variate technique can bring computational reductions to a direct MC method as one can see by confronting the numerical results in Fig. 7 with the reasoning detailed in Section 3.

Remark 1. *Sometimes we do not know how to choose the likelihood function and we need to precompute it numerically first. Let us illustrate this, in another Bayesian context for the sake of simplicity in the computations. For instance, assume that we want to esimate probabilities for k_2 from sufficiently many observations of the output s^λ whose random fluctuations are generated by those of b. To estimate k_2 from J i.i.d. observations $s_j^{\bar{E}_0}$, $j = 1 \ldots J$, with likelihood $f^{\bar{E}_0}(s_j^{\bar{E}_0} | k_2^0)$ depending only on the law of b at the unknown value $k_2 = k_2^0$, we could use a Bayesian approach with a prior distribution $k_2 \sim \pi$. The posterior distribution*

$$\pi^{\bar{E}_0}(k_2 | \{s_j^{\bar{E}_0}\}) \propto \pi(k_2) \prod_{j=1}^J f^{\bar{E}_0}(s_j^{\bar{E}_0} | k_2) \qquad (4.3)$$

allows one to compute quantities that depend on k_2 more accurately than with the prior distribution. (At least if the assumptions of the Bayesian model are satisfied by the "truth" and if the number of observations J is large enough.) But first we need to precompute the likelihood $f^{\bar{E}_0}(\cdot | k_2)$.

A "kernel" approach for instance allows one to precompute the parametrized Probability Density Functional (PDF) from a sample of realizations at
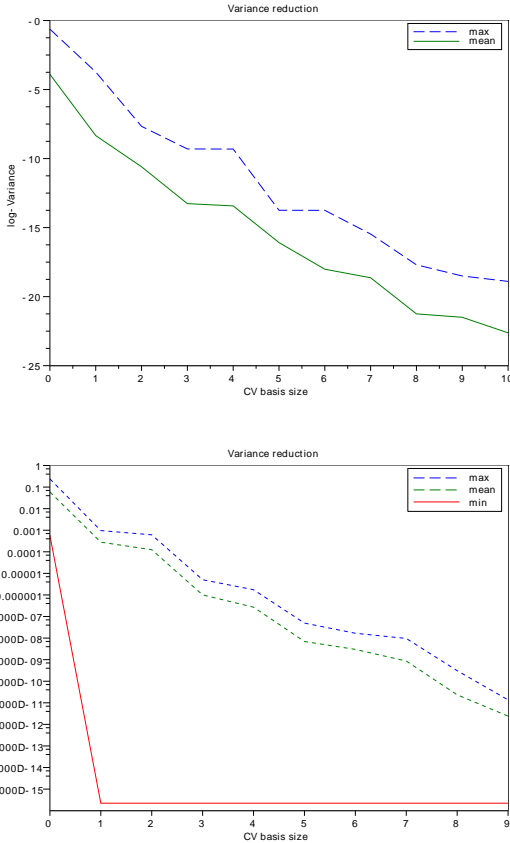
specific values of the controlled parameter $\bar{E}_0$. It amounts to give to each realization of $s^{\bar{E}_0}$ a weight. In Fig. 8, we show the approximate likelihood for observations $s_j^{\bar{E}_0}$ at one parameter value $\bar{E}_0$, after reconstruction of the PDF from a sample of realizations of $s^{(k_2,\bar{E}_0)}$ at various values of k_2, synthetically generated with a numerical MC simulation of our model and weighted by a truncated Gaussian weight with hyperparameters α_1, α_2, M

$$\sum_{m=1}^{M} 1_{|s-s_m^{(k_2,\bar{E}_0)}|<\alpha_2} e^{-\frac{|s-s_m^{(k_2,E_0)}|^2}{2\alpha_1^2}} . \qquad (4.4)$$

We numerically checked that the RB control-variate MC method was still efficient[15] to compute many expectations of a scalar functional of the uncertain parameter k_2 using (4.4) as likelihood in the posterior (4.3). We used it for many values of the control parameter $\bar{E}_0$, of the set of observations $\{s_j^{\bar{E}_0}\}$ and of the hyper parameters α_1, α_2, M. But notice that actually using (4.4) in the posterior (4.3) is very time-consuming, especially with large M. Moreover, one needs to interpolate (4.4) at the values of $\bar{E}_0$ where it has not been precomputed. So the computation of probability densities, for instance with a kernel approach, remains a challenging numerical problem in Bayesian estimation. And the computational reductions proposed in the present work do not tackle the latter point. But note that clearly, to improve the computation of (4.4) at many parameter values $\bar{E}_0$ in particular, RB ideas could still be useful. Yet this is still another topic for another work.

Figure 7: Top: maximum and mean in a sample of MC estimations $V_{\pi^{\lambda,\xi,\zeta}}(s^\lambda - \sum_{i=1}^{I} \alpha_i^\lambda \tilde{Y}^i | \{o_j^{\lambda_0}\})$ for various training values of λ, ξ, ζ and $\{o_j^{\lambda_0}\}$, as a function of I during the greedy search ("offline" stage) when $\delta = .5$, $K = 10$, $N = 12$ in case (ii) (case (i) is similar, with a decrease rate in fact a bit faster). Bottom: maximum, mean and minimum of the same estimators for various trial values of λ, ξ, ζ and $\{o_j^{\lambda_0}\}$ (average using many realizations) during the "online" stage.

---

[15]Efficiency is meant with respect to variance reduction. Of course k_2 is only a scalar in the example here and one can argue that most often the MC method is not the best integration method then. Our example is nevertheless an illustration (admittedly simple) of more general situations with high-dimensional k_2 when one uses a robust MC-like method.
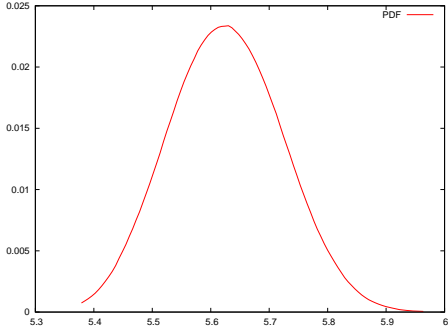
Figure 8: Empirical PDF (4.4) of k_2 at fixed $\bar{E}_0$ for $M = 10^4$ realizations with truncated Gaussian weight $\alpha_1 = .5, \alpha_2 = .01$.

5. Conclusion and Perspectives

In this article, we have presented new elements of analysis of the RB control-variate MC method (error estimates and convergence results) as well as new useful applications (to uncertainty quantification of PDEs with stochastic coefficients). But a number of questions remains, about theory and practice, which may be the object of future works.

First, our convergence results are very much inspired from those in [27, 28] about the standard RB method for PDEs and the same limitations apply: (i) they rely on assumptions that are difficult to check in practice (the fast decay of Kolmogorov widths), (ii) they may still be very pessimistic in comparison with practical possibilities (the constant in the upper-bound may be far suboptimal and a theoretical gain could be seen only for uninterestingly large dimensions I of the reduced basis), and (iii) they do not help choosing a good training sample $\tilde{\Lambda}$ of trial parameter values for the greedy algorithm (especially in the cases where Λ is large). Besides, (iii) is sometimes not only due to a lack of answer to the question how to het the optimal reduction, but also to the question how to get *any* reduction, when Λ is large. Indeed, although the maximal gain (in the infinitely-many-

query limit or, equivalently, for real-time purposes) reads only like the ratio of the reduced vs. non-reduced marginal computational time per parameter value λ, it is limited in practice by the possibility to inspect all trial parameter values $\lambda \in \tilde{\Lambda}$ and should take into account that "offline" part of the work. That is why, in absence of theory for choosing $\tilde{\Lambda}$ when Λ has a high-cardinality (possibly infinite) *and* high dimension, one still needs numerical tests to check the capabilities of the reduction method using specific (heuristic) choices of $\tilde{\Lambda}$. Fortunately, we have been able to show numerically that some gain is possible in practical situations.

In some UQ frameworks, we could actually prove numerically that some gains were possible by applying the RB control-variate MC method. But as any numerical proof, this was achieved on one specific model problem, for which some limitations were also clearly seen. So one would of course still need characterizing precisely the limitations for another specific problem. More precisely, one would need to specify numerically the efficiency regime in a given UQ context where a parametrized expectation has to be computed many time for many values of the parameter, which depends on the number of queries in the parameter values, but also on the decrease rate of the variance with respect to the dimension of the reduced basis of control variates and on the required accuracy. Furthermore, this efficiency regime will be reachable only for those problems whose "offline" work is actually do-able (in a human lifetime) and is thus limited in the dimension of the parameter λ by the size of $\tilde{\Lambda}$, as well as in the numerical complexity of the underlying model for random realizations by the computation time of one single realization. That latter limit also sets new challenges for future works in UQ: the simulation

24

of high-dimensional noise (with high-dimensional KL truncation order K) in the case of PDEs with stochastic coefficients, and the fast computation of empirical likelihoods (say by kernel approach) in Bayesian estimation. We already mentionned some tracks for future works about these issues in the course of this work.

To conclude, this work is a direct improvement on [14] which we believe to be a useful numerical approach to some UQ problems, in particular because it can bring huge computational gains at almost no cost to the simple and widely-used "naive" MC method. Notice that in any case we can certify the approximation error in all steps of the computations (of course in a probabilistic sense as for the MC method) and there is thus no loss of rigor in our approach compared with the naive MC method. Moreover, we hope that our suggestion, after [20], to use RB ideas in such fields as Bayesian estimations, which are known to have computationally demanding applications, will bring some new thoughts in that community, who might for instance want to identify new "many-query" opportunities of computational reductions using greedy algorithms.

Appendix A. Weak greedy convergence

An important tool for the proof is Lemma 1, which is a straightforward variation of a result proved in [28, 2.2].

Lemma 1. *If for some $I_C \in \mathbb{N}$ there exists $0 \leq I \leq I_C$, $q, m \in \mathbb{N}$ and $\theta \in (0, 1)$ such that $\theta\sigma_I \leq \sigma_{I+qm}$ then it also holds $\sigma_I \leq d_m/|\theta - q^{-\frac{1}{2}}|$. Moreover, if the same hypothesis holds for $\theta\tilde{\sigma}_I \leq \tilde{\sigma}_{I+qm}$, then it holds $\tilde{\sigma}_I \leq (d_m + \theta_{I_C}d_{I_C})/|\theta - q^{-\frac{1}{2}}|$ with a probability more than $1 - \epsilon$, more precisely in the events where it also holds $\tilde{\sigma}_I - \theta_I d_I \leq V(Z^{\tilde{\lambda}_{I+1}} - \hat{Y}^{\tilde{\lambda}_{I+1}}) \leq \tilde{\sigma}_I$.*

Let us treat the case in the weak greedy algorithm where for $\alpha > 0$ it holds $d_I \leq d_0 I^{-\alpha}$, $\forall 0 \leq I \leq I_M$. Assume we use the MC estimator $M_{\text{test}}^{\epsilon, I_M}$ whatever $\epsilon \in (0, 1)$. For $I \leq I_M$, the outcome $(\tilde{\sigma})_{0 \leq i \leq I_M}$ of the weak greedy algorithm is as follows with probability more than $1 - \epsilon$. For $q \in \mathbb{N}_{>0}$, and $\beta, \gamma \in \mathbb{R}_{>0}$ to be determined later, we define $I_0 := \lceil \gamma q \rceil > 0$ and assume

$$\forall C > 0, \exists I_0 \leq \hat{I} \leq I_M/\tilde{\sigma}_{\hat{I}} > Cd_0\hat{I}^{-\alpha}. \quad (A.1)$$

Given $C > 0$, we denote the smallest $\hat{I}$ satisfying A.1 by I_C. If for $0 < I \leq I_C$ we can define $m_I \in \mathbb{N}$ by $I + qm_I \leq I_C \leq I + q(m_I + 1)$ then, since $(\sigma_i)_i$ is a decreasing sequence, it holds by (A.1),

$$\tilde{\sigma}_I \leq \tilde{\sigma}_{I_C}(I_C/I)^\alpha \leq \tilde{\sigma}_{I_{I+qm_I}}(I_C/I)^\alpha. \quad (A.2)$$

Applying Lemma 1 with $m = m_I$ and $\theta = (I/I_C)^\alpha$

$$\tilde{\sigma}_{I_C} \leq \tilde{\sigma}_I \leq d_{m_I}(1 + \theta_{I_C})/|(I/I_C)^\alpha - q^{-\frac{1}{2}}| \quad (A.3)$$

(observe $d_{m_I} \geq d_{I_C}$) contradicts (A.1) as soon as one chooses $C \geq (I_C/m_I)^\alpha(1+\theta_{I_C})/|(I/I_C)^\alpha - q^{-\frac{1}{2}}| > 0$. In particular, if we set $I = \Theta I_C$ with some well-chosen $\Theta \in (0, 1 - 1/\beta - 1/\gamma)$, then

$$I_C \geq I_0 \geq \gamma q \geq \frac{q}{1 - \Theta - \frac{1}{\beta}}$$

and $\beta q m_I \geq \beta(I_C - I - q) \geq \beta((1-\Theta)I_C - q) \geq I_C$. In fact, it is always possible to find $I \in \mathbb{N}_{>0}$ with $m_I \in \mathbb{N}_{>0}$ ($m_I \geq \lfloor \gamma/\beta \rfloor > 0$ in particular) provided one takes $\gamma > 1$ large enough and sufficiently smaller $\beta > 1$. This requires $C \geq (\beta q)^\alpha(1+\theta_{I_C})/|\Theta^\alpha - q^{-\frac{1}{2}}|$. And finally, it also holds $\forall 0 \leq I < I_0, \tilde{\sigma}_I < Cd_0 I^{-\alpha}$ provided one chooses C such that

$$C \geq \max\{(\beta q)^\alpha(1+\theta_{I_C})/|\Theta^\alpha - q^{-\frac{1}{2}}|, (\tilde{\sigma}_0/d_0)I_0^\alpha\} > 0,$$

which can still be optimized as a function of $q > 0$. So the outcome of the weak greedy algorithm satisfies $\forall 0 \leq I < I_M, \tilde{\sigma}_I < Cd_0 I^{-\alpha}$ with probability more than $1 - \epsilon$. A similar reasoning can be adapted from [28] for the other cases.

Acknowledgement

Part of this work was completed while the author was an academic host at MATHICSE-ASN Chair, EPFL (Lausanne, Switzerland). Thanks to Professors J. Rappaz and M. Picasso for their hospitality.

References

[1] S. Boyaval, T. Lelièvre, A variance reduction method for parametrized stochastic differential equations using the reduced basis paradigm, Commun. Math. Sci. 8 (2010) 735–762. In special issue "Mathematical Issues on Complex Fluids". Preprint ARXIV 0906.3600 available at http://arxiv.org/pdf/0906.3600.

[2] F. Liu, M. Bayarri, J. Berger, R. Paulo, J. Sacks, A bayesian analysis of the thermal challenge problem, Computer Methods in Applied Mechanics and Engineering 197 (2008) 2457 – 2466. Validation Challenge Workshop.

[3] R. G. Ghanem, P. D. Spanos, Stochastic finite elements: a spectral approach, Springer-Verlag, New York, 1991. Revised First Dover Edition, 2003.

[4] B. F. S. L. A. Bergman, S. F. Wojtkiewicz, A state-of-the-art report on computational stochastic mechanics, Probabilistic Engrg. Mech. 12 (1997) 197–321.

[5] B. Debusschere, H. Najm, P. Pebay, O. Knio, R. Ghanem, O. L. Maître, Numerical challenges in the use of polynomial chaos representations for stochastic processes, SIAM Journal on Scientific Computing 26 (2004) 698–719.

[6] J. Burkardt, M. Gunzburger, C. Webster, Reduced order modeling of some nonlinear stochastic partial differential equations, International Journal of Numerical Analysis and Modeling 4 (2007) 368–391.

[7] A. Doostan, R. G. Ghanem, J. Red-Horse, Stochastic model reduction for chaos representations, Comput. Methods Appl. Mech. Engrg. 196 (2007) 3951–3966.

[8] J. Wang, N. Zabaras, Using Bayesian statistics in the estimation of heat source in radiation, International Journal of Heat and Mass Transfer 48 (2005) 15–29.

[9] Y. M. Marzouk, H. N. Najm, Dimensionality reduction and polynomial chaos acceleration of bayesian inference in inverse problems, Journal of Computational Physics 228 (2009) 1862 – 1902.

[10] R. G. Ghanem, A. Doostan, On the construction and analysis of stochastic models: characterization and propagation of the errors associated with limited data, J. Comput. Phys. 217 (2006) 63–81.

[11] C. Soize, Construction of probability distributions in high dimension using the maximum entropy principle: Applications to stochastic processes, random fields and random matrices, International Journal for Numerical Methods in Engineering 76 (2008) 1583–1611.

[12] A. Cohen, R. DeVore, C. Schwab, Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs, Found. Comput. Math. 10 (2010) 615–646.

[13] A. Cohen, R. Devore, C. Schwab, Analytic regularity and polynomial approximation of parametric stochastic elliptic pdes, Analysis and Applications 9 (2011) 11–47.

[14] S. Boyaval, C. Le Bris, Y. Maday, N. Nguyen, A. Patera, A reduced basis approach for variational problems with stochastic parameters: Application to heat conduction with variable robin coefficient, Computer Methods in Applied Mechanics and Engineering 198 (2009) 3187–3206.

[15] I. H. Sloan, H. Woźniakowski, When are quasi-Monte Carlo algorithms efficient for high-dimensional integrals?, J. Complexity 14 (1998) 1–33.

[16] G. Venkiteswaran, M. Junk, Quasi-monte carlo algorithms for diffusion equations in high dimensions, Math. Comput. Simul. 68 (2005) 23–41.

[17] F. Kuo, C. Schwab, I. Sloan, Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients, Technical Report 52, ETH SAM Report, 2011.

[18] J. M. Hammersley, D. C. Handscomb, Monte Carlo Methods, Methuen & Co. Ltd., London, 1965.

[19] B. Jourdain, Advanced Financial Modelling, Radon Series Comp. Appl. Math 8., De Gruyter, pp. 205–222.

[20] N. C. Nguyen, G. Rozza, D. B. P. Huynh, A. T. Patera, Reduced basis approximation and a posteriori error estimation for parametrized parabolic pdes; application to real-time bayesian parameter estimation, in: Biegler, Biros, Ghattas, Heinkenschloss, Keyes, Mallick, Tenorio, van Bloemen Waanders, Willcox (Eds.), Computational Methods for Large Scale Inverse Problems and Uncertainty Quantification, John Wiley & Sons, UK.

[21] S. Boyaval, C. Le Bris, T. Lelièvre, Y. Maday, N. Nguyen, A. T. Patera, Reduced basis techniques for stochastic problems, Archives of Computational Methods in Engineering 17 (2010) 435–454. 10.1007/s11831-010-9056-z, special issue E. Cueto, F. Chinesta, P. Ladeveze and A. Nouy ed.

[22] Y. Maday, Reduced–basis method for the rapid and reliable solution of partial differential equations, in: Proceedings of International Conference of Mathematicians,

Madrid, European Mathematical Society Eds., 2006.

[23] A. T. Patera, G. Rozza, Reduced Basis Approximation and A Posteriori Error Estimation for Parametrized Partial Differential Equations, Copyright MIT, 2006–2007. To appear in MIT Pappalardo Monographs in Mechanical Engineering.

[24] Y. Maday, A. T. Patera, G. Turinici, *A Priori* convergence theory for reduced-basis approximations of single-parameter elliptic partial differential equations, Journal of Scientific Computing 17 (2002) 437–446.

[25] Y. Maday, A. Patera, G. Turinici, Global a priori convergence theory for reduced-basis approximations of single-parameter symmetric coercive elliptic partial differential equations, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 289–294.

[26] A. Pinkus, *n*-widths in approximation theory, volume 7 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*, Springer-Verlag, Berlin, 1985.

[27] A. Buffa, Y. Maday, A. T. Patera, C. Prud'homme, G. Turinici, A priori convergence of the greedy algorithm for the parametrized reduced basis, Mathematical Modelling and Numerical Analysis (2009). Submitted.

[28] P. Binev, W. Cohen, Albert Dahmen, R. DeVore, G. Petrova, P. Wojtaszczyk, Convergence rates for greedy algorithms in reduced basis methods, SIAM J. Math. Anal 43 (2011) 1457–1472.

[29] B. P. Welford, Note on a method for calculating corrected sums of squares and products, TECHNOMETRIC 4 (1962) 419–420.

[30] D. West, Updating mean and variance estimates: An improved method, Communications of the ACM, Scientific Applications 22 (1979) 532–535.

[31] A. Quarteroni, A. Valli, Numerical Approximation of Partial Differential Equations, Springer, 2nd edition, 1997.

[32] N. C. Nguyen, K. Veroy, A. T. Patera, Certified real-time solution of parametrized partial differential equations, in: S. Yip (Ed.), Handbook of Materials Modeling, Springer, 2005, pp. 1523–1558.

[33] H. Matthies, A. Keese, Galerkin methods for linear and nonlinear elliptic stochastic pdes, Comput. Meth. Appl. Mech. Eng. 194 (2005) 1295–1331.

[34] I. Babuška, F. Nobile, R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data, SIAM Journal on Numerical Analysis 45 (2007) 1005–1034.

[35] P. Frauenfelder, C. Schwab, R. Todor, Deterministic FEM for elliptic problems with stochastic coefficients, Comput. Meth. Appl. Mech. Eng. 194 (2005) 205–228.

[36] C. Schwab, R. Todor, Karhunen-loève approximation of random fields by generalized fast multipole methods, Journal of Computational Physics 217 (2006) 100–122.

[37] A. Nouy, A generalized spectral decomposition technique to solve a class of linear stochastic partial differential equations, Computer Methods in Applied Mechanics and Engineering 196 (2007) 4521–4537.

[38] Y. Efendiev, T. Hou, W. Luo, Preconditioning markov chain monte carlo simulations using coarse-scale models, SIAM Journal on Scientific Computing 28 (2006) 776–803.

[39] L. Machiels, Y. Maday, I. B. Oliveira, A. T. Patera, D. Rovas, Output bounds for reduced-basis approximations of symmetric positive definite eigenvalue problems, C. R. Acad. Sci. Paris, Série I 331 (2000) 153–158.

[40] C. Prud'homme, D. Rovas, K. Veroy, Y. Maday, A. T. Patera, G. Turinici, Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bounds methods, Journal of Fluids Engineering 124 (2002) 70–80.

[41] G. Rozza, D. Huynh, A. T. Patera, Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations — application to transport and continuum mechanics, Arch. Comput. Methods Eng. 15 (2008) 229–275.

[42] R. S. K.A. Cliffe, M.B. Giles, A. Teckentrup, Multilevel monte carlo methods and applications to elliptic pdes with random coefficients, Computing and Visualization in Science 14 (2011) 3–15.

[43] C. P. Robert, The Bayesian choice: from decision-theoretic foundations to computational implementation, Springer Texts in Statistics, Springer Verlag, New York, 2007. Second edition translated and revised from the French original by the author.

[44] G. Casella, R. L. Berger, Statistical inference, The Wadsworth & Brooks/Cole Statistics/Probability Series, Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1990.

[45] E. L. Lehmann, G. Casella, Theory of point estimation, Springer Texts in Statistics, Springer-Verlag, New York, second edition, 1998.