ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

MATHICSE

Mathematics Institute of Computational Science and Engineering

School of Basic Sciences - Section of Mathematics

MATHICSE Technical Report

Nr. 01.2015

January 2015

Reduced order methods for uncertainty quantification problems

Peng Chen, Alfio Quarteroni, Gianluigi Rozza

http://mathicse.epfl.ch

Phone: +41 21 69 37648

Address:
EPFL - SB - MATHICSE (Bâtiment MA)
Station 8 - CH-1015 - Lausanne - Switzerland

Fax: +41 21 69 32545

SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. xx, pp. x

Reduced order methods for uncertainty quantification problems

Peng Chen *, Alfio Quarteroni †, and Gianluigi Rozza ‡

Abstract. This work provides a review on reduced order methods in solving uncertainty quantification problems. A quick introduction of the reduced order methods, including proper orthogonal decomposition and greedy reduced basis methods, are presented along with the essential components of general greedy algorithm, a posteriori error estimation and Offline-Online decomposition. More advanced reduced order methods are then developed for solving typical uncertainty quantification problems involving pointwise evaluation and/or statistical integration, such as failure probability evaluation, Bayesian inverse problems and variational data assimilation. Three expository examples are provided to demonstrate the efficiency and accuracy of the reduced order methods, shedding the light on their potential for solving problems dealing with more general outputs, as well as time dependent, vectorial noncoercive parametrized problems.

Key words. uncertainty quantification, reduced basis method, proper orthogonal decomposition, greedy algorithm, a posteriori error estimate, failure probability evaluation, inverse problem, data assimilation

AMS subject classifications.

1. Introduction.

Various stochastic computational methods have been developed during the last decade for solving uncertainty quantification (UQ) problems, and can be classified as either intrusive or nonintrusive. Belonging to the former approach are stochastic Galerkin methods with multidimensional polynomial projection [34, 24, 65, 3], generalized polynomial chaos [73], modal reduction type generalized spectral decomposition [54], etc. These methods typically converge very fast provided the solution depends smoothly on the random variables. However, either they result in a global large-scale tensor system or they lead to many coupled stochastic and deterministic systems. Besides being computationally challenging, their solution cannot make advantage of complex legacy solvers. Among the nonintrusive approaches, Monte Carlo method [32] has been widely used because of its simplicity for implementation (reuse of legacy solvers) and its superior property that the convergence rate does not depend on the number of stochastic dimensions. Unfortunately, its low convergence rate, $O(N^{-1/2})$ with N samples, demands for a large number of samples and consequently an exorbitant number of PDEs need to be solved in order to achieve reasonable accuracy. Several techniques can be applied to accelerate this method, such as deterministic sampling known as quasi Monte Carlo [27], sampling over hierarchical spatial discretization known as multilevel Monte Carlo [5], sampling using ergodic property known as Markov chain Monte Carlo [35]. In the nonintrusive approach, another recently developed and widely used method is stochastic collocation based on multidimensional polynomial interpolation [2], which not only can achieve fast convergence as the stochastic Galerkin methods but also enjoys simple implementation

*Seminar for Applied Mathematics, Department of Mathematics, ETH Zürich, Rämistrasse 101, CH-8092 Zürich, Switzerland. (cpempire@gmail.com, peng.chen@sam.math.ethz.ch)

†Modelling and Scientific Computing, Mathematics Institute of Computational Science and Engineering, EPF Lausanne, Av. Piccard, Station 8, CH-1015 Lausanne, Switzerland. (alfio.quarteroni@epfl.ch)

‡SISSA, International School for Advanced Studies, Mathematics Area, mathLab, via Bonomea 265, I-34136 Trieste, Italy. (gianluigi.rozza@sissa.it).

that enable direct use of legacy solvers. By exploiting sparsity of the solution in the stochastic space, sparse grid [72, 53], anisotropic [52] and adaptive [33, 47] sparse grid have been developed to effectively alleviate the curse of dimensionality. Other nonintrusive methods such as regression [8], discrete L^2 projection [48] are also in active development based on polynomial approximation.

However, these polynomial-based (or more in general dictionary basis-based (e.g. wavelets, radial bases)) nonintrusive methods may still be too expensive to be affordable when the (high-fidelity) solution of the underpinning problem at a single realization of the random variables is already very expensive. This computational challenge is more stringent for problems with high-dimensional uncertainties. In the last few years, reduced order methods have been developed and demonstrated to be able to significantly reduce the computational expense as long as the solutions or outputs live in low-dimensional manifold [9, 21, 14], which is the typical situation for many UQ problems. Instead of stochastic Galerkin projection on polynomial bases and stochastic collocation using polynomial interpolation, reduced order methods employ (Petrov–)Galerkin projection on "snapshots" – i.e. solutions at some suitably chosen samples or principle components. As the high-fidelity "snapshots" can be computed by using legacy solvers (nonintrusive) while the reduced solutions are obtained by solving the same (Petrov–)Galerkin problems but in reduced basis spaces (intrusive), reduced order methods are typically regarded as semi-intrusive methods.

This work recalls and summarizes the basic reduced order methods, including proper orthogonal decomposition (POD) and greedy reduced basis (greedy-RB) methods, for the solution of several typical uncertainty quantification problems that involve both pointwise evaluation and statistical integration. The state of the art is detailed in [20, 14, 22]. We further develop reduced order modelling algorithms in UQ for general functional outputs, vectorial, time dependent and noncoercive problems with emphasis on some selected representative problems that are relevant in problems like structural mechanics, thermal analysis and flow simulation. These represent ongoing extensions of existing methods in this field.

The fundamental techniques for the construction of reduced basis spaces for well-posed linear partial differential equations are summarized in Sections 2 and 3. The uncertainties, which may arise from different kind of sources, such as computational geometry, external loading, boundary conditions, material properties, etc., can be represented by random fields that are further characterized or approximated by a finite number of random variables, leading to a parametric system with certain probability distribution prescribed on the parameters. A more advanced technique for the construction of reduced basis spaces to accommodate arbitrary probability density function, named weighted reduced basis (wRB) method, is proven to be very efficient for the evaluation of integrals, e.g. statistical moments as well as for dealing with inverse problems, as shown in Section 4. For pointwise evaluation, e.g. real-time risk analysis or local sensitivity analysis, *goal-oriented* reduced basis construction algorithms are more appropriate. Reduced order methods (formulated through either Bayesian approach or Lagrangian approach) are particularly effective for the solution of "backward" UQ problems, such as stochastic optimal control, statistical inversion and variational data assimilation, which commonly demand the solution of the underpinning system for many times.

In Section 5, we demonstrate the accuracy and efficiency of reduced order methods for three typical uncertainty quantification problems addressing failure probability evaluation in

structural mechanics (a crack propagation problem with a functional output generalization properly managed with dual problems), Bayesian inversion in time dependent heat conduction (thermal analysis in presence of flaws and/or delamination), and variational data assimilation in fluid dynamics. The results obtained on these three examples show the remarkable reduction of computational expense by Offline-Online decomposition and a posteriori error analysis, thus demonstrating the great potential of the reduced order methods in solving UQ problems when the solution manifold and/or the manifold of the output quantity of interest are low dimensional.

Further computational and mathematical challenges and perspective opportunities are outlined at the end in Section 6.

2. Problem setting: PDE with random inputs. Let $(\Omega, \mathcal{F}, P)$ denote a complete probability space with the set of outcome $\omega \in \Omega$, a σ-algebra $\mathcal{F}$ and a probability measure P. Let y denote a vector of random variables defined in $(\Omega, \mathcal{F}, P)$, i.e. $y = (y_1, \ldots, y_K) : \Omega \to \mathbb{R}^K$ with $K \in \mathbb{N}$. By Γ_k we denote the image of y_k in Ω, $1 \leq k \leq K$, and $\Gamma = \otimes_{k=1}^{K} \Gamma_k$. We associate the random vector y with a probability density function $\rho : \Gamma \to \mathbb{R}$. Let $D \subset \mathbb{R}^d$ $(d = 1, 2, 3)$ denote an open and bounded physical domain with Lipschitz boundary. By U and V we denote two Hilbert spaces defined in the domain D with duals U' and V'. We consider the following problem: given $y \in \Gamma$, find $u(y) \in U$ such that

$$A(u, v; y) = F(v; y) \quad \forall v \in V, \tag{2.1}$$

where $A : U \times V \to \mathbb{R}$ and $F : V \to \mathbb{R}$ are the continuous bilinear and linear forms that depend on the random vector y, respectively. The uncertainty represented by the random vector y may arise from external loading, boundary conditions, material properties and/or computational geometries. For the well-posedness of problem (2.1), besides the continuity of A and F, we assume that the bilinear form A satisfies the inf-sup condition, i.e. $\forall y \in \Gamma$

$$\inf_{0 \neq w \in U} \sup_{0 \neq v \in V} \frac{A(w, v; y)}{||w||_U ||v||_V} =: \beta(y) > 0; \tag{2.2}$$

moreover, $\sup_{w \in U} A(w, v; y) > 0 \; \forall 0 \neq v \in V$. Under these assumptions, there exists a unique solution for problem (2.1) that satisfies the a priori stability estimate

$$||u(y)||_U \leq \frac{||F(y)||_{V'}}{\beta(y)}. \tag{2.3}$$

In many practical applications [61, 59, 15], the quantity of interest is not (or not only) the solution but (also) some functional of the solution, i.e. $s(y) := s(u(y); y) : \Gamma \to \mathbb{R}$, for instance the solution at a given location $s(y) = u(x, y)$, $x \in D$, or a parameter-dependent linear functional

$$s(y) = L(u(y); y), \tag{2.4}$$

as well as their (of u and s) statistics such as the expectation, failure probability, etc., see Section 4.

Remark 2.1. *Thanks to the general formulation of* (2.1)*, we may consider a wide range of physical problems, including linear elasticity, convection-diffusion-reaction problem, Stokes equations for incompressible Newtonian fluid, Maxwell equations for electrodynamics, etc. For the solution of many nonlinear problems, linearization with proper hyper reduction of the nonlinear term [37, 13, 28] is mostly performed, resulting in a linear problem like* (2.1)*. For the solution of unsteady problems, temporal discretization, by e.g. backward or forward Euler scheme, also leads to steady problems similar to* (2.1) *[38, 68]. Examples of more general problems will be provided in Section 5.*

Suppose $u_N(y)$ is an approximation of the solution $u(y)$ corresponding to $y \in \Gamma$ in a subspace $U_N \subset U$, $N \in \mathbb{N}$. The main effort in solving various uncertainty quantification problems consists of finding a subspace U_N with N as small as possible to yield a required accuracy of the approximation. More rigorously, let $\mathcal{M} := \{u(y) \in U : y \in \Gamma\}$ denote the solution manifold; we look for $U_N \subset U$ such that the worst approximation error

$$\sigma_N(\mathcal{M})_U := \operatorname{dist}(U_N, \mathcal{M}) \equiv \sup_{u\in\mathcal{M}} \inf_{w\in U_N} ||u - w||_U \tag{2.5}$$

be as small as possible, in the sense that it can achieve the best approximation with error (known as Kolmogorov N-width)

$$d_N(\mathcal{M})_U := \inf_{U_N\subset U} \operatorname{dist}(U_N, \mathcal{M}) \equiv \inf_{U_N\subset U} \sup_{u\in\mathcal{M}} \inf_{w\in U_N} ||u - w||_U. \tag{2.6}$$

Meanwhile, the computational effort for finding such a subspace should be affordable. Many approximation methods have been developed in the last decade, among which the stochastic Galerkin [1] and stochastic collocation [2] methods are more widely applied. However, even featuring fast convergence when the solution is smooth with respect to y, they can hardly achieve an accuracy comparable with the best approximation error. Moreover, substantial computational challenges arise for these methods when the dimension of y is high, called curse-of-dimensionality, and/or when the solution of problem (2.1) at each realization $y \in \Gamma$ is very expensive, making only a few tens or hundreds of solutions affordable. In recent years, it was proven [67, 7, 26] that reduced order methods, in particular the greedy reduced basis method [56, 61], can achieve comparable convergence as the best approximation error. Moreover, the reduced order methods have been developed with great efficiency in solving high-dimensional and large-scale problems [9, 21, 20, 14].

For the sake of computational efficiency of reduced order methods, we assume that the bilinear form A, as well as the linear forms F and L, admit an affine expansion as

$$A(w, v; y) = \sum_{q=1}^{Q_a} \Theta_q^a(y) A_q(w, v), \quad F(v; y) = \sum_{q=1}^{Q_f} \Theta_q^f(y) F_q(v), \text{ and } L(w; y) = \sum_{q=1}^{Q_l} \Theta_q^l(y) L_q(w), \tag{2.7}$$

where A_q, F_q and L_q are suitable continuous bilinear and linear forms that are independent of the random vector y, and $\Theta_q^a(y)$, $\Theta_q^f(y)$ and $\Theta_q^l(y)$ represent the y-dependent coefficients.

Remark 2.2. *The affine expansion* (2.7) *is crucial in enabling efficient Offline-Online decomposition (which will be specified in a later section) of reduced order methods. Many problems do admit an affine expansion, for instance the Karhunen–Loève expansion [44, 64]*

that is widely applied in uncertainty quantification problems leads to an affine representation/approximation of a random field, which may further give rise to an affine expansion of the bilinear and linear forms A, F and L in (2.7). For more general non-affine problems, an empirical interpolation method [4] or its weighted variant for a random field [22] can be applied to obtain an approximate affine decomposition that leads to (2.7).

3. Reduced order methods – basic formulation. In this section, we present the key ingredients of reduced order methods in solving problem (2.1), including the reduced order approximation of a full order problem, greedy algorithm and proper orthogonal decomposition for the construction of reduced basis spaces, a posteriori error estimate and efficient Offline-Online computational decomposition.

From a computational perspective, the starting point for the development of reduced order methods is to rely on a high-fidelity approximation, which is also named full order approximation. Suppose a high-fidelity approximation of problem (2.1) is sought in the high-fidelity trial space $U^{\mathcal{N}}$, using as test space $V^{\mathcal{N}}$, with bases $(w_n^{\mathcal{N}})_{n=1}^{\mathcal{N}}$ and $(v_n^{\mathcal{N}})_{n=1}^{\mathcal{N}}$, respectively. Typically these may represent finite element or spectral approximations. The high-fidelity approximation problem corresponding to problem (2.1) reads: given $y \in \Gamma$, find the high-fidelity solution $u^{\mathcal{N}}(y) \in U^{\mathcal{N}}$ such that

$$A(u^{\mathcal{N}}(y), v^{\mathcal{N}}; y) = F(v^{\mathcal{N}}; y) \quad \forall v^{\mathcal{N}} \in V^{\mathcal{N}}. \tag{3.1}$$

Its well-posedness is guaranteed if the high-fidelity spaces $U^{\mathcal{N}}$ and $V^{\mathcal{N}}$ fulfill the discrete inf-sup condition

$$\inf_{0 \neq w^{\mathcal{N}} \in U^{\mathcal{N}}} \sup_{0 \neq v^{\mathcal{N}} \in V^{\mathcal{N}}} \frac{A(w^{\mathcal{N}}, v^{\mathcal{N}}; y)}{||w^{\mathcal{N}}||_U ||v^{\mathcal{N}}||_V} =: \beta^{\mathcal{N}}(y) > 0. \tag{3.2}$$

For an accurate approximation of the solution, a very large $\mathcal{N}$ (here representing the dimension of the solution space $U^{\mathcal{N}}$) is typically required. The large-scale system corresponding to (3.1) can therefore be solved only for a limited number of realizations $y \in \Gamma$. On the other hand, a large number of realizations, especially for high-dimensional random space, are necessary to achieve certain required accuracy in approximating some quantities of interest, such as failure probability or expectation. In order to tackle this computational challenge, we turn to the reduced order approximation.

3.1. Reduced order approximation. Suppose we have constructed the reduced trial space $U_N \subset U^{\mathcal{N}}$ and test space $V_N \subset V^{\mathcal{N}}$ with bases $(w_n^N)_{n=1}^N$ and $(v_n^N)_{n=1}^N$ for $N \ll \mathcal{N}$. Then the reduced order approximation problem associated with the high-fidelity approximation problem (3.1) reads: given $y \in \Gamma$, find the reduced order solution $u_N(y) \in U_N$ such that

$$A(u_N(y), v_N; y) = F(v_N; y) \quad \forall v_N \in V_N. \tag{3.3}$$

Let $\mathbf{u}_N(y) := (u_N^1(y), \ldots, u_N^N(y))^\top$, with $\top$ representing the transpose, denote the coefficient of the reduced order solution on the bases $(w_n^N)_{n=1}^N$, i.e.

$$u_N(y) = \sum_{n=1}^{N} u_N^n(y) w_n^N. \tag{3.4}$$

Then, thanks to the affine expansion of the bilinear and linear forms in (2.7), we can write the reduced order approximation problem (3.3) as

$$\sum_{n=1}^{N}\sum_{q=1}^{Q_a}\Theta_q^a(y)A_q(w_n^N,v_m^N)u_N^n(y)=\sum_{q=1}^{Q_f}\Theta_q^f(y)F_q(v_m^N)\quad m=1,\dots,N. \tag{3.5}$$

In order to solve the reduced order approximation problem (3.3), the y-independent quantities $A_q(w_n^N,v_m^N)$, $1\le m,n\le N$, $1\le q\le Q_a$ and $F_q(v_m^N)$, $1\le m\le N$, $1\le q\le Q_f$, can be assembled only once, whereas the y-dependent reduced order system (3.5) has to be assembled and solved for any given $y\in\Gamma$. As a result, the quantity of interest $s(u(y);y)$ can be approximated by $s_N(y):=s(u_N(y);y)$, whose computation requires a number of operations independent of $\mathcal{N}$, too. For instance, the linear output (2.4) can be evaluated as

$$s(y)\approx s_N(y)=L(u_N(y);y)=\sum_{n=1}^{N}\sum_{q=1}^{Q_l}\Theta_q^l(y)L_q(w_n^N)u_N^n(y), \tag{3.6}$$

where we can compute and store the quantities $L_q(w_n^N)$, $1\le n\le N$, $1\le q\le Q_l$, once and for all, and then compute $s_N(y)$ with $O(Q_lN)$ operations, for any given $y\in\Gamma$. Therefore, a considerable reduction of computational effort is achieved as long as $N\ll\mathcal{N}$. This saving is more evident when the output of interest has to be evaluated in correspondence with a very large number of samples, which is the case in many uncertainty quantification problems with high dimensional random input, like e.g. in reliability analysis, stochastic inverse problems, etc.

Remark 3.1. *In the case of noncompliant output, i.e. $L\neq F$, $s(y)$ can be better approximated by adding a correction term at the expense of solving a dual problem of (3.3), which will be shown later in section 3.4. Note that if many, say M, output of interests are involved, we have to solve M dual problems, which brings even more computational effort.*

The accuracy and efficiency of the reduced order approximation crucially depend on the choice of the reduced trial and test spaces U_N and V_N. In order to construct an optimal reduced space $U_N\subset U^{\mathcal{N}}$ for the approximation of the solution manifold $\mathcal{M}^{\mathcal{N}}=\{u^{\mathcal{N}}(y)\in U^{\mathcal{N}}:y\in\Gamma\}$, the perhaps most natural choice for the bases of U_N is that made of a span of suitably chosen (independent) elements of this manifold. This idea has been exploited by both the proper orthogonal decomposition method [12, 70] (hereafter abbreviated as POD) and the greedy reduced basis method [56, 61] (greedy-RB), the two fundamental methods that have led to the development of various model order reduction techniques for many applications in scientific computing [58].

3.2. Proper orthogonal decomposition. POD, also known as Karhunen–Loève expansion in stochastic theory or principle component analysis in statistical analysis, was applied in the earlier days for simulation of turbulent flows in extracting the essential flow features, which provided computational evidence for the so-called coherent structures that were observed in experiments [6]. The starting point for constructing a sequence of POD basis is to compute the high-fidelity solution at n_t training samples $u^{\mathcal{N}}(y^n)$, $n=1,\dots,n_t$, typically $n_t\ll\mathcal{N}$. The the correlation matrix $\mathbb{C}\in\mathbb{R}^{n_t\times n_t}$ is formed with these entries

$$\mathbb{C}_{mn}=(u^{\mathcal{N}}(y^m),u^{\mathcal{N}}(y^n))_U,\quad 1\le m,n\le n_t, \tag{3.7}$$

where $(\cdot,\cdot)_U$ represents an inner product defined in U. Let $(\sigma_n^2, \boldsymbol{\eta}_n)$, $n = 1,\dots,r$, be the eigenpairs of $\mathbb{C}$ with rank r, ordered in such way that $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2$. Then the reduced space U_N is constructed by

$$U_N := \text{span}\{\zeta_1,\dots,\zeta_N\}, \tag{3.8}$$

where $\zeta_1,\dots,\zeta_N$, are the POD bases given by

$$\zeta_n = \sum_{m=1}^{n_t} \frac{1}{\sigma_n} \boldsymbol{\eta}_n^m u^{\mathcal{N}}(y^m), \quad 1 \leq n \leq N. \tag{3.9}$$

The reduced space U_N constructed by the POD bases solves the following constrained optimization problem

$$U_N := \underset{Z_N \subset \mathcal{M}_t^{\mathcal{N}}}{\text{argmin}} \sum_{n=1}^{n_t} ||u^{\mathcal{N}}(y^n) - P_{Z_N} u^{\mathcal{N}}(y^n)||_U^2 \text{ such that } (z_i, z_j)_U = \delta_{ij}, 1 \leq i,j \leq N, \tag{3.10}$$

where $\mathcal{M}_t^{\mathcal{N}} := \text{span}\{u^{\mathcal{N}}(y^n), n = 1,\dots,n_t\}$, $Z_N := \text{span}\{z_1,\dots,z_N\}$, being the basis $z_n \in \mathcal{M}_t^{\mathcal{N}}$, $1 \leq n \leq N$, and $P_{Z_N} : \mathcal{M}_t^{\mathcal{N}} \to Z_N$ is a projection operator defined as

$$P_{Z_N} v = \sum_{n=1}^{N} (v, z_n)_U z_n. \tag{3.11}$$

Furthermore, the approximation error defined in (3.10) of the reduced space U_N is given by [58]

$$\mathcal{E}_N^{POD} := \sum_{n=1}^{n_t} ||u^{\mathcal{N}}(y^n) - P_{U_N} u^{\mathcal{N}}(y^n)||_U^2 = \sum_{n=N+1}^{r} \sigma_n^2. \tag{3.12}$$

As a consequence, given any relative error tolerance $0 < \varepsilon < 1$, we may choose the number N of POD basis functions to be the smallest such that $\mathcal{E}_N^{POD}/\mathcal{E}_0^{POD} \leq \varepsilon$. When the eigenvalues σ_n^2, $1 \leq n \leq r$, decay very fast, only a small number of POD basis functions is needed.

Remark 3.2. *Instead of minimizing the worst approximation error as defined for the Kolmogorov N-width in (2.6), the POD basis minimize an "averaged" approximation error in "energy" norm over all the training samples. In this sense, the reduced space U_N constructed by POD is optimal as demonstrated in (3.10), in the sense that it extracts the largest energy among all the N-dimensional subspace $Z_N \subset \mathcal{M}_t^{\mathcal{N}}$. This is particularly relevant when the interest is not some quantity corresponding to worst case scenario but rather the averaged value, e.g. statistical moments.*

Remark 3.3. *The accuracy of the POD basis approximation depends not only on the number of bases but also on how well the training samples can represent the whole parameter space Γ. Ideally as many training samples as possible should be used to explore Γ. However a full solution of the high-fidelity problem (3.1) is needed at each training sample, rendering the use of many training samples impossible as long as the computation of the high-fidelity solution is expensive. Consequently, the purely POD based reduced order method is limited to problems typically with low-dimensional parameter space that can be well represented by a limited number of training samples.*

3.3. Greedy reduced basis method. The greedy reduced basis method aims to construct the reduced basis space in hierarchical manner by a greedy algorithm. To start, one would choose the first sample in the parameter space such that

$$y^1 = \operatorname*{argmax}_{y\in\Gamma} ||u^{\mathcal{N}}(y)||_U. \tag{3.13}$$

Note, however, that this choice is computationally very expensive since we have to solve an optimization problem involving possibly many high-fidelity solutions. To get ride of this difficulty, we may choose y^1 as the center of the parameter space Γ or randomly sample it according to its probability density [20]. Correspondingly, the reduced trial space is initialized as

$$U_1 = \operatorname{span}\{u^{\mathcal{N}}(y^1)\}, \tag{3.14}$$

where $u^{\mathcal{N}}(y^1) \in U^{\mathcal{N}}$, called "snapshot", is the high-fidelity solution at y^1. We defer the construction of the reduced test space V_1 to the end of this section. For $N = 1, 2, \ldots$, we seek the sample y^{N+1} at which the error between the high-fidelity solution and the reduced-order solution attains its maximum, i.e.

$$y^{N+1} = \operatorname*{argmax}_{y\in\Gamma} ||u^{\mathcal{N}}(y) - u_N(y)||_U, \tag{3.15}$$

where $u^{\mathcal{N}}(y)$ and $u_N(y)$ are the solutions of the high-fidelity problem (3.1) and reduced-order problem (3.3), respectively. Again in order to solve the optimization problem (3.15), one typically needs to solve a large number of high-fidelity problems, especially in the case of a high-dimensional parameter space. To avoid the considerable computational cost involved, we replace the true error $||u^{\mathcal{N}}(y)-u_N(y)||_U$ by an error estimator $\triangle_N(y)$, such that its evaluation cost is independent of $\mathcal{N}$, and

$$c_\triangle(y)||u^{\mathcal{N}}(y) - u_N(y)||_U \le \triangle_N(y) \le C_\triangle(y)||u^{\mathcal{N}}(y) - u_N(y)||_U, \tag{3.16}$$

where the constants $0 < c_\triangle(y) \le C_\triangle(y) < \infty$ measure the effectivity of the error estimator. When $c_\triangle(y) \ge 1$, the error estimator $\triangle_N(y)$ becomes an upper error bound for the true reduced solution error. When $C_\triangle(y)/c_\triangle(y)$ is close to one for every $y \in \Gamma$, we expect that the error estimator leads to an effective construction of the reduced space defined as

$$U_{N+1} = \operatorname{span}\big\{u^{\mathcal{N}}(y^1), \ldots, u^{\mathcal{N}}(y^{N+1})\big\}. \tag{3.17}$$

Further simplification of the optimization problem (3.15) entails the replacement of the whole parameter space Γ by a training sample set $\Xi_{\text{train}} \subset \Gamma$ of finite cardinality, i.e. $|\Xi_{\text{train}}| = n_{\text{train}} < \infty$. It is important that the training set is representative of the whole parameter space, meaning that it should be fine enough such that Γ can be well explored by Ξ_{train} and the reduced basis space constructed by the training set leads to comparable error decay as the one constructed from the whole parameter space. Meanwhile, this training set should be as small as possible such that less computational cost is demanded by the greedy algorithm. In the context of uncertainty quantification, various sampling techniques, such as Monte Carlo

[15] and sparse grid [16], can be effectively employed in choosing the training set, in alternative to algorithms based on adaptive refinement [41] and efficient partition of the parameter space [29]. As the number of solution snapshots becomes large, the reduced system (3.5) may become ill-conditioned. For stability consideration, Gram–Schmidt process [36] is performed for the snapshots, yielding a set of orthonormal basis functions $(\zeta_n)_{n=1}^{N+1}$ with respect to the norm U, i.e.

$$U_{N+1} = \text{span}\{\zeta_1, \dots, \zeta_{N+1}\}. \tag{3.18}$$

The reduced test space V_{N+1} is also enriched according to the the reduced trial space U_{N+1}.

For the construction of the reduced test space V_N, $N = 1, 2, \dots$, in the case that $U = V$ and the bilinear form A is coercive in U, i.e. $\exists \alpha(y) > 0$:

$$A(w, w; y) \geq \alpha(y)||w||_U^2 \quad \forall w \in U, \tag{3.19}$$

we can simply set $V_N = U_N$ so that the coercivity property (3.19) is preserved in the reduced space U_N; this corresponds to the most basic case of consideration for reduced basis method [56, 55, 61]. However, when the bilinear form A only satisfies the sufficient and necessary condition for well-posedness of the linear problem (2.1) in different spaces U and V, namely the continuity and stability (inf-sup condition) (2.2), we need to construct the reduced trial space $V_N \subset V^{\mathcal{N}}$ such that the reduced inf-sup condition holds

$$\inf_{0 \neq w \in U_N} \sup_{0 \neq v \in V_N} \frac{A(w, v; y)}{||w||_U ||v||_V} =: \beta_N(y) > 0. \tag{3.20}$$

An ideal case is to pick the optimal elements (with respect to approximation accuracy) of $V^{\mathcal{N}}$ that also guarantee the stability constraint (3.20), which are obtained via the "supremizer" operator $T_y : U^{\mathcal{N}} \to V^{\mathcal{N}}$ defined by

$$(T_y w, v)_V = A(w, v; y) \quad \forall v \in V^{\mathcal{N}}. \tag{3.21}$$

Correspondingly, for any $N = 1, 2, \dots$, the reduced test space can be defined as

$$V_N = \text{span}\{T_y\zeta_1, \dots, T_y\zeta_N\}. \tag{3.22}$$

By this definition and using (3.21), we have from (3.20)

$$\begin{aligned} \beta_N(y) &= \inf_{0 \neq w \in U_N} \frac{A(w, T_y w; y)}{||w||_U ||T_y w||_V} \\ &= \inf_{0 \neq w \in U_N} \frac{||T_y w||_V}{||w||_U} \\ &\geq \inf_{0 \neq w \in U^{\mathcal{N}}} \frac{||T_y w||_V}{||w||_U} \\ &= \inf_{0 \neq w \in U^{\mathcal{N}}} \sup_{0 \neq v \in V^{\mathcal{N}}} \frac{A(w, v; y)}{||w||_U ||v||_V} =: \beta^{\mathcal{N}}(y), \end{aligned} \tag{3.23}$$

so that the reduced spaces U_N and V_N preserve the stability condition. Moreover, it can be shown that, under the renormation of the trial space $||w||_{\hat{U}} = ||T_y w||_V$, the reduced solution $u_N(y)$ is the best approximation of $u(y)$ in U_N with respect to this norm [25, 71]. In fact,

$$A(w, v; y) = (T_y w, v)_V \leq ||w||_{\hat{U}} ||v||_V \quad \forall w \in U, \forall v \in V, \tag{3.24}$$

so that the continuity constant of A is $\gamma(y) = 1$. On the other hand

$$\inf_{0 \neq w \in U_N} \sup_{0 \neq v \in V_N} \frac{A(w, v; y)}{||w||_{\hat{U}} ||v||_V} = \inf_{0 \neq w \in U_N} \frac{A(w, T_y w; y)}{||w||_{\hat{U}} ||T_y w||_V} = \inf_{0 \neq w \in U_N} \frac{(T_y w, T_y w)_V}{||T_y w||_V ||T_y w||_V} = 1, \tag{3.25}$$

which implies that the stability constant of A in $U_N \times V_N$ is $\beta_N(y) = 1$. By Petrov–Galerkin orthogonality, we have

$$||u^{\mathcal{N}}(y) - u_N(y)||_{\hat{U}} \leq \frac{\gamma(y)}{\beta_N(y)} \inf_{w \in U_N} ||u^{\mathcal{N}}(y) - w||_{\hat{U}} = \inf_{w \in U_N} ||u^{\mathcal{N}}(y) - w||_{\hat{U}}, \tag{3.26}$$

which demonstrates that the test space (3.22) leads to optimal reduced order approximation. However, as indicated by the definition $T_y w$, the basis functions of the test basis depend on the parameter y, which would require a high-fidelity solve of (3.21) for each parameter value. In order to avoid this large computational cost, we take advantage of the assumption of affine structure (2.7) by solving the following high-fidelity problem only once

$$(T^q \zeta_n, v)_V = A_q(\zeta_n, v) \quad \forall v \in V^{\mathcal{N}}, n = 1, \ldots, N, q = 1, \ldots, Q_a, \tag{3.27}$$

and then assemble for each parameter $y \in \Gamma$ and $n = 1, \ldots, N$,

$$T_y \zeta_n = \sum_{q=1}^{Q_a} \Theta_q^a(y) T^q \zeta_n. \tag{3.28}$$

The construction of the reduced spaces is summarized in the Greedy Algorithm 3.3.

3.4. A posteriori error estimation. Effective error estimator plays a crucial role not only for efficient construction of reduced spaces by permitting sufficient training samples in the greedy algorithm but also for reliable quantification of the reduced order approximation error at each new parameter value. It must be *rigorous* such that it is valid for the whole parameter space and for each of the N-dimensional reduced basis approximations. It should also be relatively *sharp* or *tight*, i.e. with $c_\triangle$ and $C_\triangle$ close to one, such that proper number of reduced bases are constructed, neither too large (because $c_\triangle \gg 1$) to achieve efficiency nor too small ($C_\triangle \ll 1$) to attain accuracy of the reduced order approximation. Most importantly, evaluation of the error estimator at each given parameter should be very inexpensive, depending on N but independent of $\mathcal{N}$. In the following, a posteriori error estimator that exploits the error-residual relationship is presented, which will be demonstrated to fulfill the above requirements. Let $R : V^{\mathcal{N}} \times \Gamma \to \mathbb{R}$ denote the residual defined as

$$R(v; y) := F(v; y) - A(u_N(y), v; y) \quad \forall v \in V^{\mathcal{N}}. \tag{3.29}$$

Algorithm 1 Greedy algorithm

procedure INITIALIZATION
  Set the training set Ξ_{train}, tolerance ε_{tol}, $N = 1$;
  Choose the first sample $y^1 \in \Xi_{\text{train}}$, construct U_1 and V_1;
  Compute error estimator $\triangle_1(y)$ for each $y \in \Xi_{\text{train}}$;
end procedure
procedure CONSTRUCTION
  while $\max_{y\in\Xi_{\text{train}}} \triangle_N(y) \geq \varepsilon_{tol}$ **do**
    Pick $y^{N+1} = \text{argmax}_{y\in\Xi_{\text{train}}} \triangle_N(y)$;
    Compute $u^{\mathcal{N}}(y^{N+1})$ by solving (3.1);
    Construct $U_{N+1} = U_N \oplus \text{span}\{u^{\mathcal{N}}(y^{N+1})\}$ and V_{N+1};
    Set $N = N + 1$ and compute $\triangle_N(y)$ for each $y \in \Xi_{\text{train}}$;
  end while
  Set $N_{max} = N$;
end procedure

Let $e_N^{\mathcal{N}}(y) := u^{\mathcal{N}}(y) - u_N(y)$ for any $y \in \Gamma$. Then the discrete inf-sup condition (2.2) implies

$$||e_N^{\mathcal{N}}(y)||_U \leq \sup_{0\neq v\in V^{\mathcal{N}}} \frac{A(e_N(y), v; y)}{\beta^{\mathcal{N}}(y)||v||_V} = \sup_{0\neq v\in V^{\mathcal{N}}} \frac{R(v; y)}{\beta^{\mathcal{N}}(y)||v||_V} = \frac{||R(\cdot; y)||_{(V^{\mathcal{N}})'}}{\beta^{\mathcal{N}}(y)}. \tag{3.30}$$

Therefore, we can define the a posteriori error estimator as

$$\triangle_N(y) := \frac{||R(\cdot; y)||_{(V^{\mathcal{N}})'}}{\beta_{LB}^{\mathcal{N}}(y)}, \tag{3.31}$$

where $0 < \beta_{LB}^{\mathcal{N}}(y) \leq \beta^{\mathcal{N}}(y)$ is a lower bound for the discrete inf-sup constant, which can be evaluated by a successive constraint method [43] with computational cost independent of the high-fidelity degree of freedom $\mathcal{N}$ thanks to a Offline-Online decomposition procedure. In the context of uncertainty quantification with statistical quantity of interest, a uniform lower bound $0 < \beta_{LB}^{\mathcal{N}} \leq \beta^{\mathcal{N}}(y)$ for any $y \in \Gamma$ is feasible and computationally more efficient.

Thus, the a posteriori error estimator defined in (3.31) is an upper bound for the reduced solution error as a result of the estimate (3.30), establishing the first inequality of (3.16) with $c_{\triangle}(y) = 1$, i.e.

$$||e_N^{\mathcal{N}}(y)||_U \leq \triangle_N(y) \quad \forall N = 1, \ldots, N_{max}, \forall y \in \Gamma. \tag{3.32}$$

To see second inequality of (3.16), we define the Riesz representative $\hat{e}_N^{\mathcal{N}} \in V^{\mathcal{N}}$ of the residual $R(\cdot; y)$ that satisfies

$$(\hat{e}_N^{\mathcal{N}}(y), v)_V = R(v; y) \quad \forall v \in V^{\mathcal{N}}, \tag{3.33}$$

so that $||\hat{e}_N^{\mathcal{N}}(y)||_V = ||R(\cdot; y)||_{(V^{\mathcal{N}})'}$. By setting $v = \hat{e}_N^{\mathcal{N}}(y)$ in (3.33), we have

$$||\hat{e}_N^{\mathcal{N}}(y)||_V^2 = R(\hat{e}_N^{\mathcal{N}}(y); y) = A(e_N^{\mathcal{N}}(y), \hat{e}_N^{\mathcal{N}}(y); y) \leq \gamma(y)||e_N^{\mathcal{N}}(y)||_U||\hat{e}_N^{\mathcal{N}}(y)||_V, \tag{3.34}$$

where the inequality follows from the continuity of the bilinear form A with continuity constant $\gamma(y) < \infty$ for any $y \in \Gamma$. Therefore, by definition of the error estimator (3.31) and (3.34), we have

$$\triangle_N(y) \leq \frac{\gamma(y)}{\beta_{LB}^{\mathcal{N}}(y)} ||e_N^{\mathcal{N}}(y)||_U \quad \forall N = 1, \ldots, N_{max}, \forall y \in \Gamma, \tag{3.35}$$

which establishes the second inequality of (3.16) with $C_\triangle(y) = \gamma(y)/\beta_{LB}^{\mathcal{N}}(y) > 1$.

Remark 3.4. *The bounds* (3.32) *and* (3.35) *are valid for any* $N = 1, \ldots, N_{max}$ *and* $y \in \Gamma$, *so that the a posteriori error estimator* $\triangle_N$ *is rigorous. As* $c_\triangle(y) = 1$, *when* $C_\triangle(y) = \gamma(y)/\beta_{LB}^{\mathcal{N}}(y)$ *is close to one, this error estimator is also sharp. However, when this is violated, i.e.* $C_\triangle \gg 1$ *encountered in convection dominated problems, the error estimator could be overly conservative. In order to deal with this difficulty, we may choose a new norm for* U *or* V, *such that* $C_\triangle$ *is close or even equals to one with the renormation [71].*

In the case $U = V$, it can be shown for a compliant output $s^{\mathcal{N}}(y) = F(u^{\mathcal{N}}(y); y)$ that

$$|s^{\mathcal{N}}(y) - s_N(y)| \leq \triangle_N^s(y) := \beta_{LB}^{\mathcal{N}}(y)\triangle_N^2(y) \equiv \frac{||R(\cdot, y)||^2_{(V^{\mathcal{N}})'}}{\beta_{LB}^{\mathcal{N}}(y)} \leq \frac{\gamma(y)}{\beta_{LB}^{\mathcal{N}}(y)} |s^{\mathcal{N}}(y) - s_N(y)|, \tag{3.36}$$

so that we can use the output error estimator $\triangle_N^s$ in the Greedy Algorithm 3.3 to construct the goal-oriented bases. In fact, by linearity of F and the Galerkin orthogonality we have

$$\begin{aligned} |s^{\mathcal{N}}(y) - s_N(y)| &= |F(u^{\mathcal{N}}(y); y) - F(u_N(y); y)| \\ &= A(u^{\mathcal{N}}(y), u^{\mathcal{N}}(y) - u_N(y); y) \\ &= A(e_N^{\mathcal{N}}(y), e_N^{\mathcal{N}}(y); y). \end{aligned} \tag{3.37}$$

Therefore, the left inequality of (3.36) is obtained by the definition of $\hat{e}_N^{\mathcal{N}}(y)$ and (3.32) via

$$A(e_N^{\mathcal{N}}(y), e_N^{\mathcal{N}}(y); y) \leq ||\hat{e}_N^{\mathcal{N}}(y)||_V ||e_N^{\mathcal{N}}(y)||_U \leq ||\hat{e}_N^{\mathcal{N}}(y)||_V \triangle_N(y) = ||R(\cdot; y)||^2_{(V^{\mathcal{N}})'} / \beta_{LB}^{\mathcal{N}}(y). \tag{3.38}$$

To prove the right inequality of (3.36), we define the energy norm $||w||_y^2 := A(w, w; y)$, so that $||w||_y^2 \leq \gamma(y)||w||_V^2$ by continuity of A. Together with Cauchy-Schwarz inequality, we have

$$||\hat{e}_N^{\mathcal{N}}(y)||_V^2 = A(e_N^{\mathcal{N}}(y), \hat{e}_N^{\mathcal{N}}(y); y) \leq ||e_N^{\mathcal{N}}(y)||_y ||\hat{e}_N^{\mathcal{N}}(y)||_y \leq \sqrt{\gamma(y)} ||e_N^{\mathcal{N}}(y)||_y ||\hat{e}_N^{\mathcal{N}}(y)||_V. \tag{3.39}$$

Hence, the second inequality of (3.36) is established by noting $||\hat{e}_N^{\mathcal{N}}(y)||_V^2 = ||R(\cdot, y)||^2_{(V^{\mathcal{N}})'}$.

In the more general case when $U \neq V$ and the output is not compliant, $L \neq F$, a direct error estimate for the reduced output error is given by

$$|s^{\mathcal{N}}(y) - s_N(y)| = |L(u^{\mathcal{N}}(y); y) - L(u_N(y); y)| \leq ||L(\cdot; y)||_{(U^{\mathcal{N}})'} \triangle_N(y). \tag{3.40}$$

However, evaluating $||L(\cdot; y)||_{(U^{\mathcal{N}})'}$ for each $y \in \Gamma$ is unfeasible; moreover it is possible that

$$\lim_{N \to \infty} \frac{||L(\cdot; y)||_{(U^{\mathcal{N}})'} \triangle_N(y)}{|s^{\mathcal{N}}(y) - s_N(y)|} \propto \lim_{N \to \infty} \frac{1}{\triangle_N(y)} \to \infty, \tag{3.41}$$

(where $\propto$ represents propositional to) when the reduced output is "close" to compliant so that (3.36) holds, making the error estimate (3.40) very ineffective. To address this problem and retain the quadratic convergence effect as in (3.36), we modify the output by adding a correction/residual term as

$$s_{N,M}(y) = s_N(y) - R(\psi_M(y); y), \tag{3.42}$$

where $\psi_M(y) \in V_M^{du}$ is the solution of the following reduced order dual problem

$$A(w_M, \psi_M; y) = -L(w_M; y) \quad \forall w_M \in U_M^{du}, \tag{3.43}$$

being $U_M^{du} \subset U^{\mathcal{N}}$ and $V_M^{du} \subset V^{\mathcal{N}}$ the reduced test and trial spaces of dimension M, which can be constructed by the Greedy Algorithm 3.3, noting the change of the test and trial spaces.

With the modification of the reduced output s_N by $s_{N,M}$, we have

$$|s^{\mathcal{N}}(y) - s_{N,M}(y)| \le \triangle_{N,M}^s(y) := \frac{||R(\cdot; y)||_{(V^{\mathcal{N}})'} ||R^{du}(\cdot; y)||_{(U^{\mathcal{N}})'}}{\beta_{LB}^{\mathcal{N}}(y)}, \tag{3.44}$$

where R^{du} is the residual of the dual problem defined as

$$R^{du}(w; y) = -L(w; y) - A(w; \psi_M(y); y) \quad \forall w \in U^{\mathcal{N}}. \tag{3.45}$$

The bound (3.44) can be obtained using an argument similar to (3.38), by noting that

$$\begin{aligned} |s^{\mathcal{N}}(y) - s_{N,M}(y)| &= |L(u^{\mathcal{N}}(y)) - L(u_N(y); y) + F(\psi_M(y); y) - A(u_N(y), \psi_M(y); y)| \\ &= | - A(e_N^{\mathcal{N}}(y), \psi^{\mathcal{N}}(y)) + A(e_N^{\mathcal{N}}(y), \psi_M(y); y)| \\ &= |A(e_N^{\mathcal{N}}(y), \epsilon_M^{\mathcal{N}}(y); y)| \quad (\text{ with } \epsilon_M^{\mathcal{N}}(y) := \psi^{\mathcal{N}}(y) - \psi_M(y)). \end{aligned} \tag{3.46}$$

Numerical evidence also demonstrates that $\triangle_{N,M}^s(y)$ is upperly bounded by $C|s(y) - s_{N,M}(y)|$.

3.5. Offline-Online decomposition. Offline-Online decomposition plays a pivotal role in computational reduction of the reduced order methods, for both real-time evaluation of the quantity of interest and construction of the reduced spaces. With the construction of the reduced spaces in the last section, particularly the reduced test space explicitly constructed as (3.22) with the basis given by (3.28), the reduced order problem (3.3) can be written as: for any $m = 1, \ldots, N$

$$\sum_{q}^{Q_a} \sum_{q'}^{Q_a} \sum_{n=1}^{N} \Theta_q^a(y) \Theta_{q'}^a(y) A_q(\zeta_n, T^{q'} \zeta_m) u_N^n(y) = \sum_{q}^{Q_f} \sum_{q'}^{Q_a} \Theta_q^f(y) \Theta_{q'}^a(y) F_q(T^{q'} \zeta_m), \tag{3.47}$$

where $A_q(\zeta_n, T^{q'}\zeta_m)$ and $F_q(T^{q'}\zeta_m)$ can be computed in the Offline stage only once. In the Online stage, the reduced system (3.47) can be assembled and solved with $O(Q_a^2 N + Q_f Q_a + N^3)$ operations, a number independent of $\mathcal{N}$, which makes it affordable in real-time as long as Q_a, Q_f and N are small.

For the evaluation of the rigorous and sharp a posteriori error estimator $\triangle_N(y)$, or its component $||R(\cdot; y)||_{(V^{\mathcal{N}})'}$, by Riesz representation, it is equivalent to evaluate $||\hat{e}_N^{\mathcal{N}}(y)||_V$. For

$q = 1, \dots, Q_f$, let $\mathcal{C}_q$ denote the Riesz representative of the functional $F_q(\cdot)$ of (2.7) in $V^{\mathcal{N}}$, i.e.

$$(\mathcal{C}_q, v)_V = F_q(v) \quad \forall v \in V^{\mathcal{N}}, q = 1, \dots, Q_f; \tag{3.48}$$

analogously, for $q = 1, \dots, Q_a$, and $n = 1, \dots, N$, let $\mathcal{L}_q^n$ denote the Riesz representative of the functional $A_q(\zeta_n, \cdot; y)$ of (2.7) in $V^{\mathcal{N}}$, being ζ_n the n-th reduced basis, we have

$$(\mathcal{L}_q^n, v)_V = A_q(\zeta_n, v; y) \quad \forall v \in V^{\mathcal{N}}, q = 1, \dots, Q_a, n = 1, \dots, N. \tag{3.49}$$

By the definition of the residual (3.29) and the relation (3.33), we obtain

$$\hat{e}_N^{\mathcal{N}}(y) = \sum_{q=1}^{Q_f} \Theta_q^f(y) \mathcal{C}_q - \sum_{q=1}^{Q_a} \sum_{n=1}^{N} \Theta_q^a(y) u_N^n(y) \mathcal{L}_q^n. \tag{3.50}$$

Therefore,

$$\begin{aligned} ||\hat{e}_N^{\mathcal{N}}(y)||_V^2 = & \sum_{q=1}^{Q_f} \sum_{q'=1}^{Q_f} \Theta_q^f(y) \Theta_{q'}^f(y) (\mathcal{C}_q, \mathcal{C}_{q'})_V \\ & - 2 \sum_{q=1}^{Q_f} \sum_{q'=1}^{Q_a} \sum_{n=1}^{N} \Theta_q^f(y) \Theta_{q'}^f(y) u_N^n(y) (\mathcal{C}_q, \mathcal{L}_{q'}^n)_V \\ & + \sum_{q=1}^{Q_a} \sum_{q'=1}^{Q_a} \sum_{n=1}^{N} \sum_{n'=1}^{N} \Theta_q^a(y) \Theta_{q'}^a(y) u_N^n(y) u_N^{n'}(y) (\mathcal{L}_q^n, \mathcal{L}_{q'}^{n'})_V. \end{aligned} \tag{3.51}$$

The y-independent quantities $(\mathcal{C}_q, \mathcal{C}_{q'})_V$, $(\mathcal{C}_q, \mathcal{L}_{q'}^n)_V$ and $(\mathcal{L}_q^n, \mathcal{L}_{q'}^{n'})_V$ should only be computed once in the Offline stage, then we have to assemble the y-dependent terms to compute $||\hat{e}_N^{\mathcal{N}}||_V^2$ with $O((Q_f + NQ_a)^2)$ operations for each $y \in \Gamma$, a cost still independent of $\mathcal{N}$. Thus, the Offline-Online decomposition not only enables efficient solution of the reduced order problem (3.3) but also a very inexpensive evaluation of the error estimator, allowing a large number of training samples as long as $Q_f + Q_a N$ is small.

Both the inexpensive evaluation of the solution and the error estimator can be carried out by using the Offline-Online decomposition for the dual problem in the same manner.

4. Solving uncertainty quantification problems. The basic reduced order methods presented in the last section can be further developed aimed at the solution of UQ problems [14], in the context of both forward UQ problems (sensitivity analysis, risk prediction or reliability analysis, statistical moment evaluation) and "backward" or inverse UQ problems (optimal control/design, shape optimization or reconstruction, parameter estimation). The essential computational tasks for these UQ problems involve pointwise evaluation, i.e. evaluate some quantity of interest at a large number of samples (e.g. risk prediction or failure probability computation), as well as numerical integration, for the evaluation of statistical moments (e.g. variance based sensitivity analysis).

4.1. Pointwise evaluation – goal-oriented adaptive algorithm for risk prediction. A major UQ problem is the quantification of the reliability of a system, or otherwise said how to predict the risk of failure of the system, given uncertainties in some inputs, e.g. material fatigue under random external loading, overheating under uncertain thermal conductivity, etc. This requires evaluation of the failure probability defined as

$$P_0 := P(\omega \in \Omega : s(y(\omega)) > s_0) = \int_\Gamma \chi_{\Gamma_0}(y)\rho(y)dy, \tag{4.1}$$

where s is known as *limit state function* or *performance function* to measure the reliability of the system, e.g. the averaged temperature distribution for thermal conduction, while s_0 is a critical value above which the system fails. Correspondingly, the domain of *failure probability* is defined as

$$\Gamma_0 := \{y \in \Gamma : s(y) > s_0\}; \tag{4.2}$$

χ_{Γ_0} is the characteristic function of Γ_0, that is $\chi_{\Gamma_0}(y) = 1$ if $y \in \Gamma_0$ and vanishes otherwise. The failure probability can be evaluated by Monte Carlo sampling: sample a sequence of realizations, y^m, $m = 1, \ldots, M_{mc}$, of the random variables according to their joint probability density function $\rho : \Gamma \to \mathbb{R}$, solve the underpinning high-fidelity model at each sample, evaluate the limit state function and compute the Monte Carlo failure probability by

$$P_0^m = \frac{1}{M_{mc}} \sum_{m=1}^{M_{mc}} \chi_{\Gamma_0}(y^m). \tag{4.3}$$

In order to alleviate the heavy computational burden of this approach, several methods have been developed, for instance, the first and second order reliability method [60, 63], the response surface method [31, 10]. They all share the same paradigm: first constructing a surrogate for the limit state surface $S_0 = \{y \in \Gamma : s(y) = s_0\}$, then evaluating the failure probability by (4.1). However, when the limit state surface lacks smoothness or features possible discontinuities or singularities, these methods either possibly fail in reconstructing the right limit state surface, resulting in erroneous failure probability, or demand too much computational effort for more accurate reconstruction. An approach based on the polynomial chaos expansion and a hybrid algorithm for iterative evaluation of failure probability has been proposed in [45, 46]. However, reconstructing the limit state surface in high dimensional probability space by polynomial chaos approximation is rather challenging. Moreover, lack of smoothness brings an essential difficulty for the polynomial chaos approximation, due to the onset of Gibbs phenomenon [15].

In this context, reduced order methods can achieve both efficiency and accuracy thanks to their optimal approximation property with respect to the Kolmogorov width as well as their capability of producing certified results thanks to a rigorous, reliable and inexpensive a posteriori error estimators. As higher accuracy is needed to determine if a sample y near the state limit surface leads to a failure or not (because $s(y)$ is rather close to s_0), a goal-oriented adaptive algorithm can be developed to facilitate accurate reconstruction of the limit state surface and tolerate a coarse approximation of output far from it [15]. The goal-oriented

adaptive error estimator is defined as

$$\triangle_N^a(y) = \frac{\triangle_N^s(y)}{|s_N(y) - s_0|}, \tag{4.4}$$

where s_N is the reduced output by primal-dual approximation in reduced space of dimension N given by (3.42) (here we take $N = M$), $\triangle_N^s(y)$ is the error estimator for the error $|s^{\mathcal{N}}(y) - s_N(y)|$ defined in (3.44). The next sample for construction of reduced spaces is chosen as

$$y^{N+1} = \max_{y \in \Xi_{\text{train}}} \triangle_N^a(y). \tag{4.5}$$

For a more accurate evaluation of the failure probability, the number of samples could be very large, e.g. 10^6 or larger, due to the slow convergence of the Monte Carlo method. In order to efficiently construct the reduced spaces, we propose an adaptive approach to explore the Monte Carlo sample set by adaptively enrich the training set of the greedy algorithm.

Algorithm 2 Goal-oriented Adaptive Greedy Algorithm

procedure INITIALIZATION
  Set the initial training set set M_0, maximum steps S_{max}, choose scaling parameter θ;
  Set $N = 1$, choose $y^1 \in \Gamma$, construct U_1 and V_1 for both primal and dual problems;
end procedure
procedure CONSTRUCTION AND EVALUATION
  for $m = 0, \ldots, S_{max}$ **do**
    Sample the training set Ξ_{train}^m with cardinality $|\Xi_{\text{train}}^m| = M_0\theta^m$;
    Compute $s_N(y)$ and $\triangle_N^a(y)$ according to (4.4) for each $y \in \Xi_{\text{train}}^m$;
    while $\max_{y \in \Xi_{\text{train}}^m} \triangle_N^a(y) \geq 1$ **do**
      Pick $y^{N+1} = \operatorname{argmax}_{y \in \Xi_{\text{train}}^m} \triangle_N^a(y)$;
      Compute $u^{\mathcal{N}}(y^{N+1})$ by solving (3.1);
      Construct $U_{N+1} = U_N \oplus \operatorname{span}\{u^{\mathcal{N}}(y^{N+1})\}$ and V_{N+1};
      Compute $s_N(y)$ and $\triangle_N^a(y)$ for each $y \in \{y \in \Xi_{\text{train}}^m : \triangle_N^a(y) \geq 1\}$;
      Set $N = N + 1$;
    end while
    Evaluate the failure probability according to (4.3) with $M_{mc} = M_0(\theta^0 + \cdots + \theta^m)$.
  end for
end procedure

The a posteriori error estimator $\triangle_N^a$, related to $\triangle_N^s$, plays a critical role in accurate evaluation of the failure probability. Whenever the former is smaller than 1, the output is guaranteed to be either a failure (when $s_N(y) > s_0$) or not (when $s_N(y) \leq s_0$). This certification is a result that the reduced error is bounded by the error estimator as demonstrated both in (3.32) and (3.44). In fact, if $\triangle_N^a(y) < 1$, either $\triangle_N^s(y) < s_N(y) - s_0$, we have

$$s^{\mathcal{N}}(y) - s_0 = s^{\mathcal{N}}(y) - s_N(y) + s_N(y) - s_0 \geq -\triangle_N^s(y) + s_N(y) - s_0 > 0, \tag{4.6}$$

which verifies that y is in the failure domain, or $\triangle_N^s(y) < s_0 - s_N(y)$, so that

$$s^{\mathcal{N}}(y) - s_0 = s^{\mathcal{N}}(y) - s_N(y) + s_N(y) - s_0 \leq \triangle_N^s(y) + s_N(y) - s_0 < 0, \tag{4.7}$$

implying that y falls outside of the failure domain. By the goal-oriented adaptive greedy algorithm 4.1, the error estimator $\triangle_N^a$ is smaller than one for all the samples, so that the output at each sample can be explicitly determined whether it leads to a failure or not, resulting in accurate evaluation of the failure probability.

Remark 4.1. *In the presence of high-fidelity error $s(y) - s^{\mathcal{N}}(y)$, or errors arising from other sources, e.g. empirical interpolation of nonaffine random field, we also need to take them into account in determining if the sample is in the failure domain [15]. If the errors are difficult to quantify by a posteriori error estimator, or if the error estimator is not an upper bound for the error, one may apply an iterative procedure to verify if the failure probability is accurate, leading to solution of the high fidelity problem at more samples [15, 46].*

4.2. Evaluation of moments – weighted algorithm for arbitrary probability density. In many UQ applications, rather than the probability distribution of the output, we are interested in its statistical moments, e.g. expectation, variance, and their related quantities e.g. variance based global sensitivity. In these cases, numerical integration with respect to arbitrary probability density $\rho : \Gamma \to \mathbb{R}$ of the random variables $y : \Omega \to \mathbb{R}$ is of more interest than pointwise evaluation. To evaluate the integration under general probability density functions, the generalized polynomial chaos expansion that chooses the orthonormal polynomials according to the density functions can achieve the same accuracy [73] with less polynomial basis functions. In the framework of reduced order methods, a weighted algorithm can be developed by taking advantage of the density functions with the aim of using less reduced basis functions yet achieving the same accuracy of integration. In particular, for the evaluation of the expectation of an output $s : \Gamma \to \mathbb{R}$ by reduced basis approximation, we have the error

$$|\mathbb{E}[s] - \mathbb{E}[s_N]| \leq \int_\Gamma |s(y) - s_N(y)|\rho(y)dy \leq \int_\Gamma \triangle_N^s(y)\rho(y)dy. \tag{4.8}$$

In order to balance the approximation error of the expectation and the number of reduced basis functions to construct, we define a new weighted a posteriori error estimator $\triangle_N^\rho$ by weighting the error estimator $\triangle_N^s$ via the density function ρ, i.e. [20]

$$\triangle_N^\rho(y) = \rho(y)\triangle_N^s(y), \tag{4.9}$$

which implies that the error estimator $\triangle_N^s$ tends to be smaller at the sample whose density is larger. For instance, in the case of discrete distribution with probability $0 < p_n < 1$ at y^n for $n = 1, \ldots, N_p$ with $p_1 + \cdots + p_{N_p} = 1$, the expectation error is bounded by $\triangle_N^s(y^1)p_1 + \cdots + \triangle_N^s(y^{N_p})p_{N_p}$, which is minimized if $\triangle_N^s(y^1)p_1 = \cdots = \triangle_N^s(y^{N_p})p_{N_p}$. In the case of continuous distribution, by Monte Carlo quadrature, we have the approximation of the expectation as

$$|\mathbb{E}[s] - \mathbb{E}[s_N]| \leq \int_\Gamma \triangle_N^s(y)\rho(y)dy \approx \frac{1}{M}\sum_m^M \triangle_N^s(y^m), \tag{4.10}$$

where the realizations $y^1, \ldots, y^M$ are sampled according to the density function ρ, yielding relatively more samples where the density function is large, so that a relatively smaller error

estimator $\triangle_N^s$ at these samples gives rise to a smaller quadrature error. As for high order statistical moments, we have

$$\begin{aligned} |\mathbb{E}[s^k] - \mathbb{E}[s_N^k]| &\leq \int_\Gamma |s(y) - s_N(y)| \Big| \sum_{i=1}^k s^{i-1}(y) s_N^{k-i}(y) \Big| \rho(y) dy \\ &\leq \sup_{y\in\Gamma} \Big| \sum_{i=1}^k s^{i-1}(y) s_N^{k-i}(y) \Big| \int_\Gamma \triangle_N^s(y) \rho(y) dy, \end{aligned} \tag{4.11}$$

which make the weighted error estimator an effective choice to minimize the statistical error at the same number of reduced bases. For the error bound of the reduced order approximation of statistical moments, we refer to [9, 40]. For efficient evaluation of the statistical moments using tensor product quadrature formula based on sparse grid techniques, see [16].

4.3. Stochastic/statistical inverse problems – Bayesian and Lagrangian approaches. Reduced order methods are especially useful for stochastic and/or statistical inverse problems in the "many-query" context, including stochastic optimal control and design, data assimilation, optimization and parameter identification under uncertainties. These problems can be commonly addressed by the Bayesian inference approach (probability approach) and/or by Lagrangian variational formulation (deterministic approach). In this section, we present the two approaches with the application of reduced order methods in dealing with these inverse problems.

In the Bayesian approach, upon prescribing some knowledge of the unknown input characterized by some prior probability distribution, and given some data for the model output, we update the distribution of the input terms as posterior distribution (e.g. [66]). More precisely, suppose the unknown input can be represented by a sequence of (possibly infinite) random variables y obeying some prior distribution with density $\rho_0(y)$, which can be either constructed from some measurements or provided by expert opinion. Moreover, suppose some data of the model output $O : U \to \mathbb{R}^K$ is available, typically described by

$$\kappa = O(u) + \eta, \tag{4.12}$$

where $\kappa \in \mathbb{R}^K$ is a finite ($K < \infty$) dimensional vector representing observational data; O is a K dimensional output functional depending on the solution u, e.g. $O(u) = (s^{(1)}(u), \cdots, s^{(K)}(u))^\top$ with each $s^{(k)}(u)$ as a functional for $1 \leq k \leq K$; η is a (measurement) noise described by K dimensional random vector. A common choice for η is Gaussian random vector obeying distribution $\mathcal{N}(0, \Sigma_\eta)$ with covariance matrix $\Sigma_\eta \in \mathbb{R}_+^{K\times K}$. A practical assumption for the structure of this covariance is given as $\Sigma = \sigma_\eta \mathbb{I}$, being $0 < \sigma_\eta < \infty$ the standard deviation and $\mathbb{I}$ the $K \times K$ identity matrix. Under such setting, we define the likelihood function as

$$\rho_y(\kappa - O(u)) := \frac{1}{\sqrt{(2\pi)^K |\Sigma_\eta|}} \exp\left(-\frac{1}{2} ||\kappa - O(u(y))||^2_{\Sigma_\eta^{-1}} \right), \tag{4.13}$$

where $|\Sigma_\eta|$ is the determinant of Σ_η and the norm $||\xi||^2_{\Sigma_\eta^{-1}} = \xi^\top \Sigma_\eta^{-1} \xi$ for any $\xi \in \mathbb{R}^K$. Then by Bayes' theorem, the posterior density function is given by

$$\rho_\kappa(y) = \frac{1}{Z} \rho_0(y) \rho_y(\kappa - O(u)), \text{ with } Z = \int_\Gamma \rho_0(y) \rho_y(\kappa - O(u(y))) dy, \tag{4.14}$$

where Z is the normalization constant and Γ is the image of the random variables y. If the prescribed prior distribution is also a Gaussian distribution $\mathcal{N}(0, \Sigma_y)$, then the posterior distribution is a Gaussian distribution with the posterior density

$$\rho_\kappa(y) \propto \exp\left(-\frac{1}{2}||\kappa - O(u(y))||^2_{\Sigma_\eta^{-1}} - \frac{1}{2}||y||^2_{\Sigma_y^{-1}}\right). \tag{4.15}$$

In the computation of the posterior density $\rho_\kappa(y)$, we either need to evaluate the constant Z by some numerical integration technique or to sample from a dynamics that preserves the posterior measure, or converges to the posterior density, e.g. the Markov chain Monte Carlo sampling using Metropolis–Hastings algorithm. In both techniques, a large number of numerical solutions of the underpinning PDE has to be computed, leading to considerable computational cost if just a single solution of the PDE is already very expensive. This can be regarded as an ideal case for reduced order methods: actually, the expensive high-fidelity solution can be replaced by an inexpensive reduced solution in the numerical integration Z, yielding

$$Z = \int_\Gamma \rho_0(y)\rho_y((\kappa - O(u_N(y)))|y)dy, \tag{4.16}$$

where u_N is the reduced order approximate of the high fidelity solution $u^{\mathcal{N}}$.

The Bayesian approach incorporates the statistical data and the model with unknown random input to calibrate the information of the input, which further results in better prediction of solution and related quantities. This approach is naturally suited for parameter estimation, model calibration, data assimilation, optimization and design, etc. Another approach that does not require the measurement noise on the data, relies on a Lagrangian variational formulation. It aims to minimize a cost functional consisting of two terms,the former meaning the discrepancy between model prediction and the given data (a desirable state or experimental observation), the latter representing a regularization of the unknown input that makes the inverse problem well-posed:

$$J(u, \varsigma; y) = \frac{1}{2}||\kappa - O(u(y))||^2_D + \frac{\lambda}{2}||\varsigma(y)||^2_R. \tag{4.17}$$

Here κ could be a finite dimensional observation data as before, or a distributional field, e.g. a desirable solution field at a certain region, when O is a restriction operator that maps the solution to itself in this region; the norm $||\cdot||_D$ measures the discrepancy between the data κ and the model prediction $O(u)$; ς is called control function and represents the unknown input, which could be a forcing term, a boundary condition or a material coefficient, etc.; $||\cdot||_R$ is the Tikhonov regularization norm. The relative importance of the regularization term compared to the discrepancy term is characterized by the parameter $\lambda > 0$. A typical inverse problem, e.g. parameter estimation or optimal control, can be formulated as:

$$\min_{u \in U, \varsigma \in R} J(u, \varsigma; y) \text{ such that (2.1) holds.} \tag{4.18}$$

We remark that if the measure D is taken as Ξ_η^{-1} and the regularization term $\varsigma(y)$ coincides with y with the measure R taken such that $\lambda||\cdot||^2_R = ||\cdot||^2_{\Xi_y^{-1}}$, then the optimal solution of the

constrained optimization problem (4.18) coincides with the MAP estimator of the Bayesian approach, i.e. the point of maximal posterior density.

To solve the nonlinear constrained optimization problem (4.18), direct nonlinear programming based on Newton or quasi Newton methods can be used. Alternatively, we can use a Lagrangian approach [19, 17] with the Lagrangian functional defined as

$$\mathcal{L}(u, p, \varsigma; y) = J(u, \varsigma; y) + A(u, p; y) - F(p; y), \tag{4.19}$$

where $p \in U$ is the adjoint (or dual) variable, corresponding to the state (or primal) variable $u \in U$. A necessary condition is that the optimal solution $(u^*(y), p^*(y), \varsigma^*(y)) \in U \times V \times R$ is a critical point of the Lagrangian functional, i.e.

$$\begin{cases} A(u^*(y), v; y) & = F(v) & \forall v \in V, \\ A(q, p^*(y); y) & = (O(u^*(y)) - \kappa, O_u(u^*(y))(q))_D & \forall q \in U, \\ \mathcal{L}_\varsigma(u^*(y), p^*(y), \varsigma^*(y); y)(\vartheta) & = 0 & \forall \vartheta \in R. \end{cases} \tag{4.20}$$

This Karush–Kuhn–Tucker (KKT) or first order optimality system comprises the state equation, the adjoint equation and the optimality condition, with O_u and $\mathcal{L}_\varsigma$ standing for the Fréchet derivative of O with respect to u and $\mathcal{L}$ with respect to ς, respectively. The optimiality condition depends on the unknown input $\varsigma(y)$; for instance when the latter is the forcing term, we obtain

$$\mathcal{L}_\varsigma(u^*(y), p^*(y), \varsigma^*(y); y)(\vartheta) = (\lambda \varsigma^*(y) + p^*, \vartheta)_R. \tag{4.21}$$

As we need to solve this coupled KKT system (4.20) for many times, reduced order methods can be effectively applied for the reduction of the total computational cost. With this aim, (4.20) can be approximated by the following reduced KKT system: find $(u_N^*(y), p_N^*(y), \varsigma_N^*(y)) \in U_N \times V_N \times R_N$ (the reduced optimal solution) such that

$$\begin{cases} A(u_N^*(y), v_N; y) & = F(v_N) & \forall v_N \in V_N, \\ A(q_N, p_N^*(y); y) & = (O(u_N^*(y)) - \kappa, O'(u_N^*(y))(q_N))_D & \forall q_N \in U_N, \\ \mathcal{L}_\varsigma(u_N^*(y), p_N^*(y), \varsigma_N^*(y); y)(\vartheta_N) & = 0 & \forall \vartheta_N \in R_N. \end{cases} \tag{4.22}$$

Construction of the reduced spaces U_N, V_N and R_N, an effective a posteriori error estimation, and offline-online computational decomposition play a crucial role in the efficiency and accuracy of the reduced order approximation for the optimality system. We will exemplify this Lagrangian approach by solving a data assimilation problem in the next section.

5. Applications.

In this section, we demonstrate the accuracy and efficiency of reduced order methods for the solution of three uncertainty quantification problems that are representative of the families of problems considered in the previous section: a problem of failure probability evaluation/risk prediction for crack propagation relevant to fracture mechanics; the Bayesian inversion of the material property via time dependent heat conduction; a stochastic data assimilation problem for velocity field of blood flow in a carotid artery. These problems are used as illustrative examples to extend the reduced order methodology from its state of the art to a more complex settings that is more suitable for UQ applications.

5.1. Structural mechanics – linear elasticity model for crack propagation. Accurate computational simulation and prediction of material failure in the form of crack propagation is difficult. We consider prediction of brittle failure or the risk of fatigue-induced crack growth, where the geometry of the crack can be parametrized to accommodate variation of crack length and the material is subject to loading with uncertainty. In particular, we consider the center-cracked tension speciman as in [42]: a plate contains an internal center crack under tension from top and bottom edge, whose geometry is symmetric with four equal regions as visualized in the left part of Figure 1. Because of the symmetry, we only need to consider a quarter of it, e.g. the top-right region D_1 as in the middle part of Figure 1. In the non-dimensional setting, the half width of the plate is 1 and the half height of the plate is $\mu_2 \in [0.5, 2]$, the half length of the center crack is $\mu_1 \in [0.2, 0.8]$. The domain $D_\mu = D_1$ can be split into subdomains $D_\mu = D_\mu^1 \cup D_\mu^2 \cup D_\mu^3$ as in Figure 1 (middle) in order to accommodate the large variation of displacement by refining the finite element discretization near the crack tip region around $(\mu_1, 0)$, see one sample mesh in Figure 1 (right).

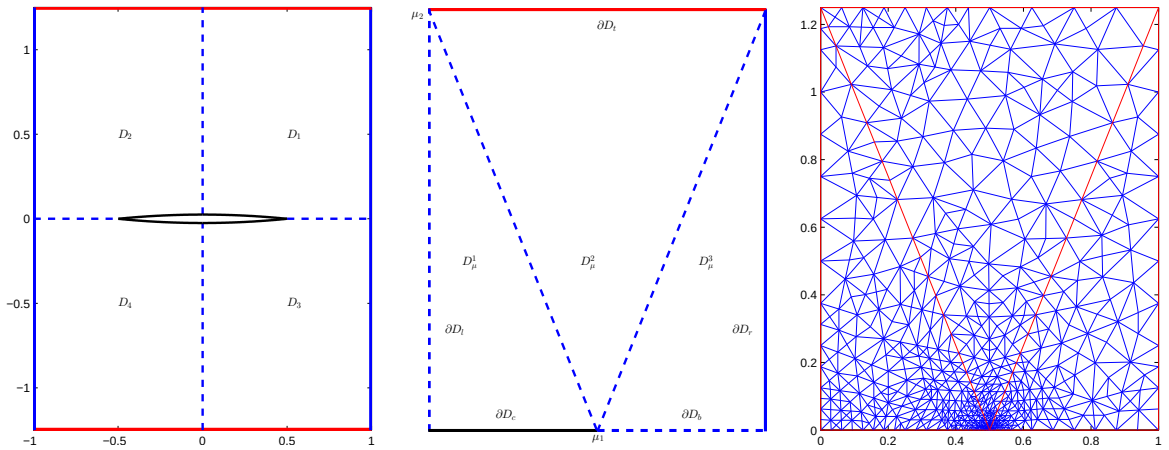

Figure 1. *Left: geometry of a plate containing an internal center crack under tension, with four symmetric regions D_1–D_4 ; middle: the top-right region D_1; right: an adapted/refined mesh near the crack tip region.*

A normal tension subject to uncertainty $\sigma(y) = \sigma_0(1+y)$ is imposed on the top boundary ∂D_t, where $\sigma_0 = 1$ and $y \in U(-\sqrt{3}, \sqrt{3})$ obeys uniform distribution with zero mean and unit variance. Zero traction is imposed on the right and on the crack of the plate. Symmetric boundary conditions are imposed on the remaining boundaries. The displacement of the plate is governed by a linear elasticity equation in the physical domain D_μ, where $\mu \in P_\mu = [0.2, 0.8] \times [0.5, 2]$, i.e. given $\mu \in P_\mu$ and realization $y \in \mathbb{R}$, find $u(\mu, y) = (u_1(\mu, y), u_2(\mu, y)) \in U_\mu = \{v \in (H^1(D_\mu))^2 : v_1|_{\partial D_l} = 0, v_2|_{\partial D_b} = 0\}$, such that

$$\sum_{i,j,k,l=1}^{2} \int_{D_\mu} \frac{\partial u_i}{\partial x_j} C_{ijkl} \frac{\partial v_k}{\partial x_l} dx_1 dx_2 = \int_{\partial D_t} \sigma(y) v_2 dx_1 \quad \forall v \in U_\mu. \tag{5.1}$$

where the $C_{ijkl} = c_1 \delta_{ij}\delta_{kl} + c_2(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$ is the constitutive tensor, being $c_1 = \frac{\nu}{(1+\nu)(1-2\nu)}$ and $c_2 = \frac{1}{2(1+\nu)}$ the Lamé constants for the plain strain with the Poisson ratio $\nu = 0.3$. Let

$D_{\bar{\mu}} = D^1_{\bar{\mu}} \cup D^2_{\bar{\mu}} \cup D^3_{\bar{\mu}} = (0,1) \times (0,1.25)$, where $\bar{\mu} = (\bar{\mu}_1, \bar{\mu}_2) = (0.5, 1.25)$ (the center of the parameter domain), denote the reference domain, we transform the equation (5.1) into the one expressed in the reference domain, which we write as: find $u \in U_{\bar{\mu}}$, such that

$$a(u, v; \mu) = f(v; \mu, y) \quad \forall v \in U_{\bar{\mu}}. \tag{5.2}$$

As the transformation is affine with respect to the parameter μ, we obtain that the bilinear and linear forms in (5.2) (a combination of integration in the three reference subdomains) are still affine with 10 and 1 terms, respectively, thus permitting efficient Offline-Online decomposition.

Our quantity of interest is energy release rate (ERR) (or its related quantity stress intensity factor (SIF)), one of the most important quantities for the prediction of crack propagation, that summarizes stress, strain and displacement field in the near crack tip region:

$$s(u; \mu, y) = -\frac{1}{2}\frac{\partial}{\partial \mu_1} a(u, u; \mu) + \frac{\partial}{\partial \mu_1} f(u; \mu, y) =: b(u, u; \mu) + \ell(u; \mu, y). \tag{5.3}$$

We consider the dual problem: find $\psi \in U_{\bar{\mu}}$ s.t.

$$a(v, \psi; \mu) = b(u, v; \mu) + \frac{1}{2}\ell(v; \mu, y) \quad \forall v \in U_{\bar{\mu}}, \tag{5.4}$$

and define the new variables $U = (U_+, U_-) \in U^2_{\bar{\mu}}$,

$$U_+(\mu, y) = \frac{1}{2}(u(\mu, y) + \psi(\mu, y)) \quad \text{and} \quad U_-(\mu, y) = \frac{1}{2}(u(\mu, y) - \psi(\mu, y)). \tag{5.5}$$

Then the system (5.2), (5.4) can be reformulated as: find $U(\mu, y) \in U^2_{\bar{\mu}}$, such that

$$A(U, V; \mu) = F(V; \mu, y) \qquad \forall V \in U^2_{\bar{\mu}}, \tag{5.6}$$

where the bilinear form $A : U^2_{\bar{\mu}} \times U^2_{\bar{\mu}} \to \mathbb{R}$ and the linear form $F : U^2_{\bar{\mu}} \to \mathbb{R}$ are given by

$$\begin{aligned} A(U, V; \mu) = &- b(U_+, V_+; \mu) - b(U_-, V_+; \mu) + 2a(U_+, V_+; \mu) \\ &- b(U_-, V_-; \mu) - b(U_+, V_-; \mu) - 2a(U_-, V_-; \mu), \end{aligned} \tag{5.7}$$

and

$$F(V; \mu, y) = f(V_+; \mu, y) + \frac{1}{2}\ell(V_+; \mu, y) - f(V_-; \mu, y) + \frac{1}{2}\ell(V_-; \mu, y). \tag{5.8}$$

Both are affine forms with respect to μ and y. Moreover, it can be shown by definition that

$$s(u; \mu, y) = F(U; \mu, y). \tag{5.9}$$

This renders s a compliant quantity in the new formulation (5.6), for which we obtain a reliable and sharp a posteriori error bound as given in (3.38) for its reduced basis approximation s_N. We remark that the bilinear form (5.7) does not necessarily satisfy the inf-sup condition (2.2) for arbitrary parameter μ if one uses L_2 norm for $U^2_{\bar{\mu}}$, but it does satisfy this condition in the parameter range P_μ when we rather take the energy norm $||U||^2_{U^2_{\bar{\mu}}} = A(U, U; \tilde{\mu})$ at

$\tilde{\mu} = (0.2, 2)$ (note that for the sake of well-posedness of (5.6), we have to take $\tilde{\mu}$ different from $\bar{\mu} = (0.5, 1.25)$, the parameter value for the reference geometry, in this particular example). The dependence of the inf-sup constant on the parameter μ is shown in Figure 2 (left), where the minimum value $\beta^{\mathcal{N}} \approx 0.0192$ is attained at $\mu = (0.5, 0.8)$. We use interpolation (based on Gaussian radial basis functions at 100 random samples [1]) to approximate $\beta^{\mathcal{N}}$ for its inexpensive Online evaluation, denoted as $\beta_i^{\mathcal{N}}$, whose relative error $(\beta_i^{\mathcal{N}} - \beta^{\mathcal{N}})/\beta^{\mathcal{N}}$ is concentrated at 0 and uniformly smaller than 4% as shown in Figure 2 (right).

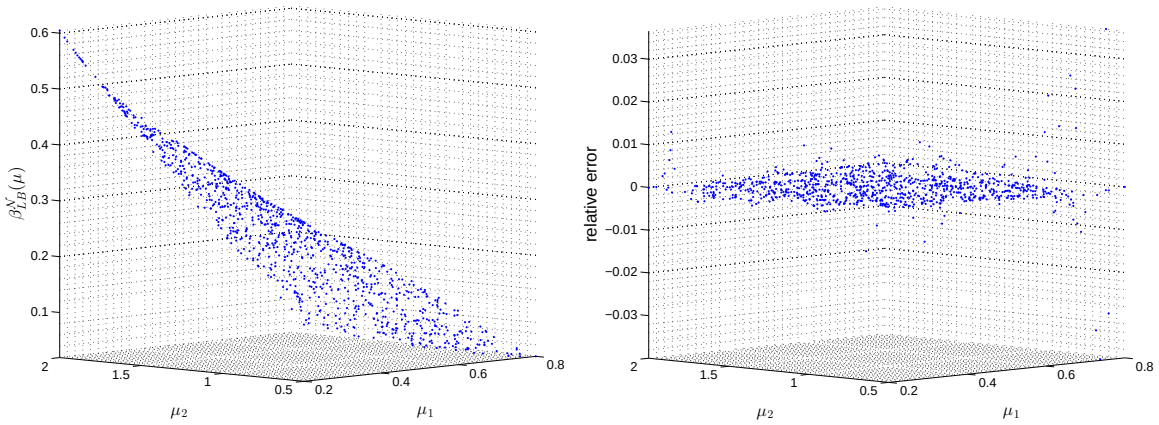

Figure 2. *Dependence of the high-fidelity inf-sup constant $\beta^{\mathcal{N}}(\mu)$ on $\mu = (mu_1, \mu_2)$ (left) and the relative approximation error of $\beta^{\mathcal{N}}(\mu)$ by Gaussian radial basis interpolation at 100 random samples (right).*

Figure 3 displays the displacement field $u = (u_1(\mu, y), u_2(\mu, y))$ at three parameter values μ as well as the dependence of the energy release rate $s(u; \mu, y)$ on the parameter μ at the realization $y = 0$. We can observe that the energy release rate depends nonlinearly on μ and becomes large (singular) near the region $\mu = (0.8, 0.5)$, i.e. when the length of the crack is large and the height of the plate is small. Therefore, in order to predict the risk of material failure (or center crack propagation), reduced basis approximation in this region should be reasonably more accurate than in the other regions. Figure 4 depicts the parameter samples selected by the greedy algorithm (left) and the corresponding maximum and mean values of the error and the a posteriori error bound of the reduced basis approximation s_N of the energy release rate s at 100 random parameter samples (right). We can observe that the samples are distributed over all the parameter domain and more samples are selected in the region of high risk, where the energy release rate varies/grows very fast. From the right part of Figure 4 we can see that the a posteriori error bound is reliable, i.e. always larger than the error, and relatively sharp with effectivity (bound/error) staying around several tens.

For the evaluation of the failure probability/risk prediction of the crack propagation, we set the critical value that indicates the crack propagation as $s_0 = 10$, and apply the goal-oriented adaptive algorithm in both the parameter domain $P_\mu \ni \mu$ and the probability domain $\Gamma \ni y$, with $M_0 = 100$ and $\theta = 10$, presented in section 4.1 for the construction of the reduced basis approximation. The reduced basis samples selected (order of selection indicated by marker

[1] Other methods such as Lagrange interpolation, sparse grid interpolation in high dimensions [21], and successive constraint method [43] are also applicable for Online evaluation of the inf-sup constant $\beta^{\mathcal{N}}$.

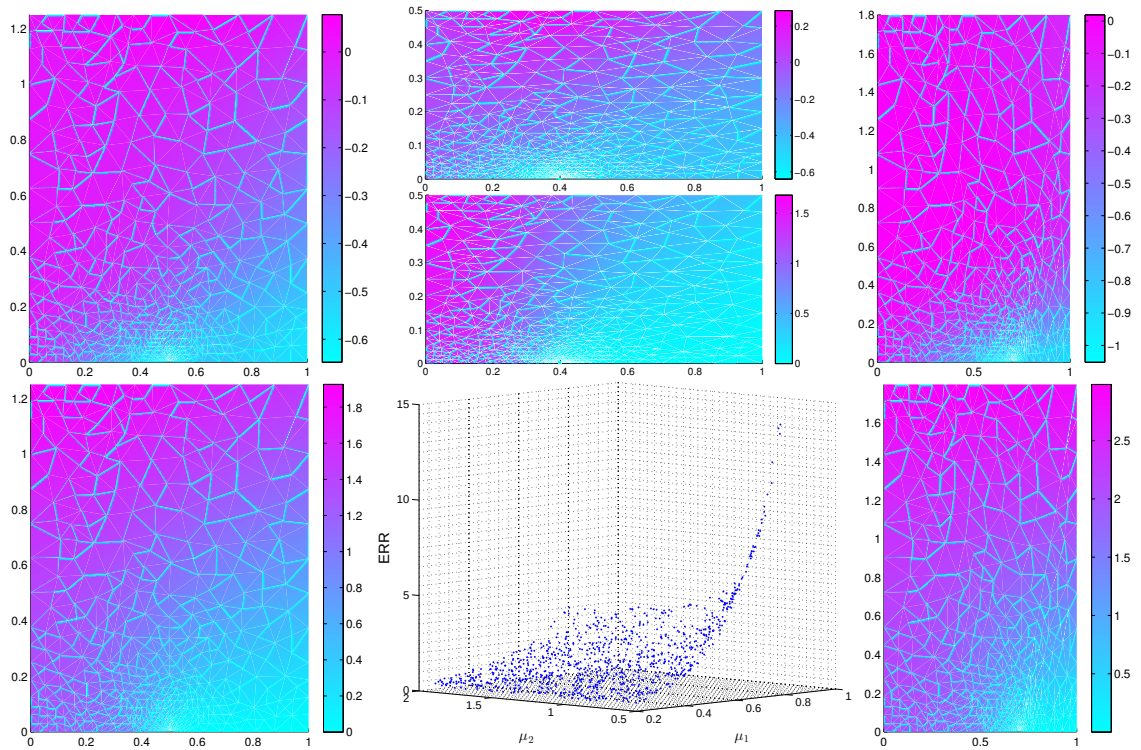

Figure 3. *Left: displacement field u_1 (top) and u_2 (bottom) at $\mu = (0.5, 1.25)$; middle: u_1 and u_2 (top) at $\mu = (0.4, 0.5)$ and energy release rate (ERR) (bottom); right: u_1 (top) and u_2 (bottom) at $\mu = (0.7, 1.8)$.*

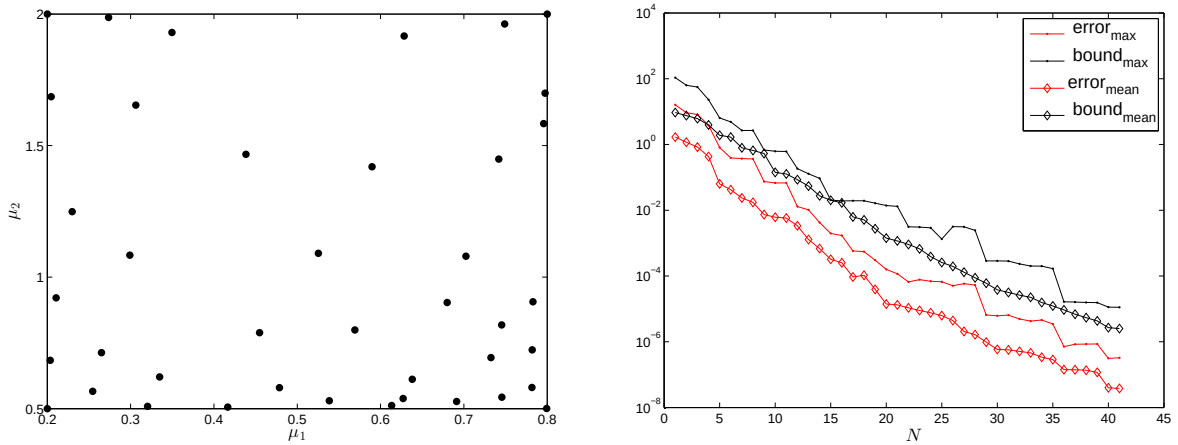

Figure 4. *Left: parameter samples for RB construction by greedy algorithm; right: the maximum and mean values of the RB error and the a posteriori error bound for 100 random parameter samples.*

size) by this algorithm is shown in Figure 5, from which we can see that all the samples (except the first one pre-determined at $\tilde{\mu} = (0.2, 2)$) are located in the region of high risk or displays large variation, particularly the (singular) region near $\mu = (0.8, 0.5)$, leading to

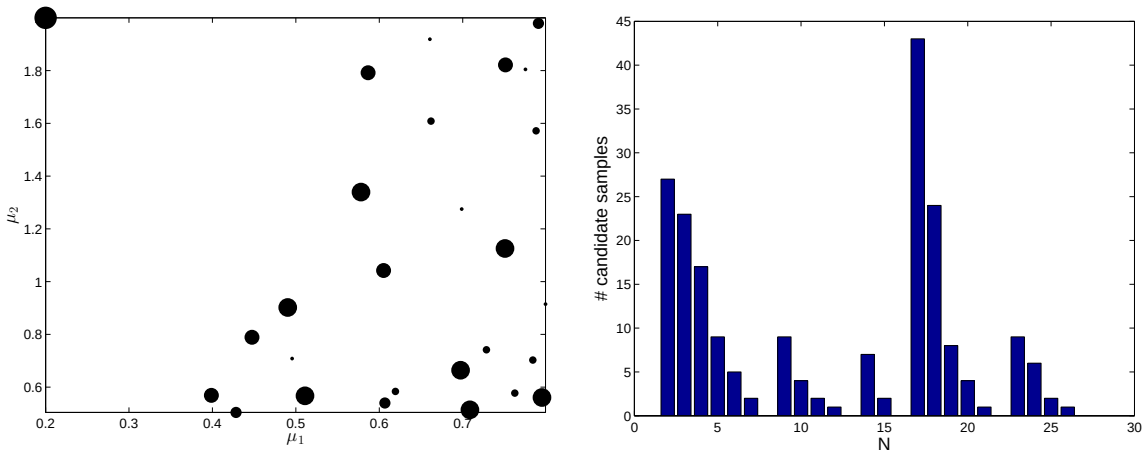

Figure 5. *Left: parameter samples for RB construction by goal-oriented adaptive greedy algorithm; right: number of potential samples* $\{y \in \Xi^m_{train}, \triangle^a_N(y) \geq 1\}$ *to be selected at each iteration step* $m = 0, \ldots, 4$.

smaller number of samples thus more efficient Online evaluation than the uniform greedy algorithm. Moreover, from Figure 5 (right) we can see that only a small number of samples at each iteration is the potential samples for the construction of reduced basis, though the number of Monte Carlo samples is rather large (10^{m+2}, $m = 0, \ldots, 4$) for the evaluation of failure probability. The failure probability at the parameter value $\mu = (0.4, 0.5)$ is $P_0^m = 0.067$ computed by formula (4.3) with reduced basis evaluation at 111100 Monte Carlo samples in total, which is free from the reduced basis approximation error of the energy release rate due to the property (4.6) and (4.7). More importantly, only 27 high-fidelity problems are solved instead of 111100, leading to considerable computational reduction.

5.2. Heat conduction – time-dependent Bayesian inversion of material flaw. In this example, we consider a transient thermal analysis for detection of flaws/defects/cracks in a composite material bonded to a concrete [51]. This example provides new insights for time dependent problems. The non-dimensional geometry is visualized in Figure 6, where a delamination crack is present on the interface in the center, whose length $2 \times y_1$ is unknown as it can not be observed; we assume that y_1 follows a beta distribution beta(α, α) supported on $[0.2, 0.8]$ with mean 0.5; the bottom part $D^1_{y_1}$ is made of the material of concrete and the top layer $D^2_{y_1}$ is made of the material of composite, which both depend on y_1. Moreover, the ratio of the conductivity of the composite over that of the concrete y_2, is unknown and obeys a beta distribution beta(α, α) supported on $[0.5, 2]$ with mean 1.25. A mesh sample is shown in the left part of Figure 6, where the mesh is refined near the the crack tip region in order to accommodate the large variations of temperature.

In order to calibrate the two unknowns, we impose time-dependent surface heat flux on the top edge and measure the temperature distribution on D^1_m and D^2_m, which are two squares of size 0.2×0.2 with centers at $(1.5, 1.2)$ and $(2.5, 1.2)$, respectively. We consider the time period of $t \in [0, 5]$ and prescribe zero initial condition at $t = 0$. A homogeneous Dirichlet boundary condition, i.e. zero temperature, is imposed on the bottom boundary;

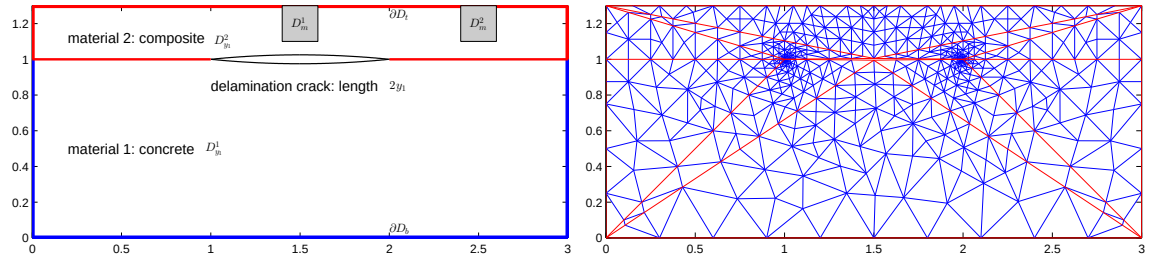

Figure 6. *Left: the geometry of the material of composite and concrete with delamination crack in the interface and with two measurement sites; right: an adapted/refined mesh near the crack tip region.*

a homogeneous Neumann boundary condition, i.e. zero heat flux, is prescribed on the left and right boundaries, and the crack interfaces from above and below. The heat flux on the whole top boundary is prescribed uniformly as one, $g(t) = 1$, during the first half period $t \in [0, 2.5)$ and zero, $g(t) = 0$, during the second half period $t \in [2.5, 5]$. The heat conduction problem reads: at any given time $t \in (0, 5]$, at any realization of the random variables $y \in \Gamma = [0.2, 0.8] \times [0.5, 2]$, find the temperature distribution field $u(t, y) \in U_{y_1} = \{v \in H^1(D_{y_1}) : v|_{\partial D_b} = 0\}$, being $D_{y_1} = D^1_{y_1} \cup D^2_{y_1}$, such that

$$\int_{D_{y_1}} \partial_t uv dx + \int_{D^1_{y_1}} \nabla u \cdot \nabla v dx + \int_{D^2_{y_1}} y_2 \nabla u \cdot \nabla v dx = \int_{\partial D_t} g(t) v dx_1 \; \forall v \in U_{y_1}. \tag{5.10}$$

After an affine transformation of the geometry to the reference geometry with $\bar{y}_1 = 1$, i.e. the reference length of the crack being 1, we obtain the following problem: at any $t \in (0, 5]$ and $y \in \Gamma$, find $u(t, y) \in U_{\bar{y}_1}$, such that

$$m(\partial_t u, v; y) + a(u, v; y) = f(v; t) \quad \forall v \in U_{\bar{y}_1}, \tag{5.11}$$

where m, a and f are affine with respect to y, with 4, 12 and 1 affine terms, respectively. After temporal discretization using backward Euler scheme with $\delta t = 0.05$ (thus leading to $5/0.05 = 100$ time steps), we have

$$m(u^l, v; y) + \delta t a(u^l, v; y) = \delta t f(v; t^l) + m(u^{l-1}, v; y) \quad \forall v \in U_{\bar{y}_1}, l = 1, \ldots, 100, \tag{5.12}$$

where $u^l = u(t^l)$ and $u^0 = 0$ (zero initial condition), being $t^l = l\delta t$, $l = 0, 1, \ldots, 100$. To solve (5.12), we first apply finite element for high-fidelity approximation of this problem, based on which we construct the reduced basis approximation by POD-greedy algorithm [39], that is POD (section 3.2) in temporal variable and greedy algorithm (section 3.3) in random variables. Note that (5.12) can be written in the general form (2.1), for the greedy algorithm we propose, following the same steps in section 3.4, a *weighted* a posteriori error estimator (introduced in section 4.2, formula (4.9))

$$\triangle_N^{\rho_0}(t^l, y) = \left(\frac{\rho_0 \delta t}{\beta^{\mathcal{N}}_{LB}} \sum_{l'=1}^{l} ||\hat{e}^{\mathcal{N}}_N(t^{l'}, y)||^2_V \right)^{1/2}, \quad l = 1, \ldots, 100, \tag{5.13}$$

where ρ_0 is the prior density of y, $\hat{e}_N^{\mathcal{N}}(t^{l'}, y)$ is the Riesz representative of the residual of (5.12) at the time $t^{l'}$. When the new reduced basis sample y is chosen according to this error estimator, we solve (5.12) by finite element approximation at each time step and apply POD to the finite element solutions at all the time steps to get one new basis. We remark that a few new basis functions are also feasible, whose number depends on the tolerance for the POD error (3.12).

In order to estimate the crack length y_1 and the conductivity ratio y_2, we consider the two measurements of the averaged temperature distribution at several time instances, defined as

$$s_i^{(k)}(u) = \int_{D_m^i} u(t^{l_k}, x, y)dx \quad i = 1, 2,\ k = 1, \ldots, K/2. \tag{5.14}$$

Then we denote the output as $O(u) = (s_1^{(1)}, \ldots, s_1^{(K/2)}, s_2^{(1)}, \ldots, s_2^{(K/2)})^\top \in \mathbb{R}^K$. The corresponding measurement data κ, defined in (4.12) as $\kappa = O(u) + \eta$, are provided with measurement noise $\eta \sim \mathcal{N}(0, \Sigma_\eta)$, where the covariance matrix $\Sigma_\eta = \mathrm{diag}(\sigma^2, \ldots, \sigma^2)$ with $\sigma = 0.1$. The measurement data are taken every 5 time steps, leading to $K = 2 \times (100/5) = 40$. Then by Bayes' theorem we obtain the posterior density of the random variable y in (4.14), where we need to evaluate the normalization constant Z by integration with respect to the prior (beta) distribution of y, for which we apply a two-dimensional tensor-product Gauss-Jacobi quadrature formula. In particular, we need to evaluate the output $O(u)$ for every integration nodes y. Note that this quantity is not compliant, so we can also apply the primal-dual approach introduced in section 3.4 and correct the output by the residual of the primal problem evaluated at the dual solution, as in (3.42). More explicitly, the dual problems (corresponding to measurement) read [2]: find $\psi_i(t, y) \in U_{\bar{y}_1}$ such that

$$m(v, \partial_t \psi; y) + a(v, \psi; y) = \ell_i(v) \quad \forall v \in U_{\bar{y}_1}, \qquad \text{being } \ell_i(v) = \int_{D_m^i} v dx, \quad i = 1, 2. \tag{5.15}$$

Then, similarly to the definition (3.44) for steady problems, the weighted a posteriori error estimator for the output $s_i^{(k)}$ defined in (5.14) is given by

$$\triangle_{N,M}^{s,\rho_0}(t^{l_k}, y) = \left(\frac{\rho_0 \delta t}{\beta_{LB}^{\mathcal{N}}} \sum_{l'=1}^{l_k} ||\hat{e}_N^{\mathcal{N}}(t^{l'}, y)||_V^2 \right)^{1/2} \left(\frac{\rho_0 \delta t}{\beta_{LB}^{\mathcal{N}}} \sum_{l'=l_k}^{100} ||\tilde{e}_M^{\mathcal{N}}(t^{l'}, y)||_U^2 \right)^{1/2}, \tag{5.16}$$

where $\tilde{e}_M^{\mathcal{N}}(t^{l'}, y)$ is the Riesz representative of the residual of the dual problem evaluated at the reduced basis dual solution approximated by M basis at time $t^{l'}$.

Figure 7 displays the temperature distribution at several time instances at the reference sample $y = (1, 1)$, from which we can see that the temperature is relatively higher in the local region above the crack than in the other region, so that the location of two measurement sites (featuring different temperature and variation) are reasonable. It increases (from $t = 0$ to $t = 2.5$) as heat flux is imposed and decreases (from $t = 2.5$ to $t = 5$) as heat flux is set to

[2] Instead of taking $O(u)$ as the output, we can also direct take the likelyhood function $\rho_0(y)\rho_y(\kappa - O(u))$ in (4.14) as the output for the evaluation of Z, which is however nonlinear with respect to the solution and nonaffine with respect to the parameter. This difficulty has been addressed in [23].

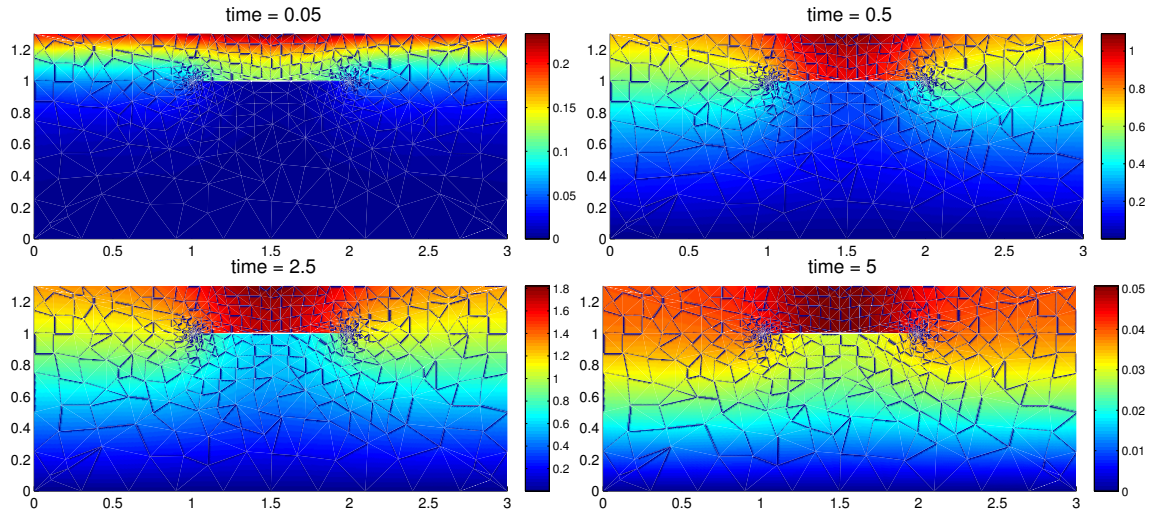

Figure 7. *Temperature distribution of the heat conduction at four time instances* $t = 0.05, 0.5, 2.5, 5$.

zero, and shows a large variation around the crack tip region so that the locally refined mesh is important to capture this variation.

The left part of Figure 8 shows the reduced basis samples that are constructed by the weighted error estimator (5.13) and by the non-weighted one, i.e. without weighting by the prior density function. From this figure we can observe that the weighted error estimator obeys the prior distribution (beta distribution beta$(10, 10)$) and selects the samples that lead to high density, while the non-weighted error estimator produces samples that are most located along the boundary. The right part of Figure 8 depicts the errors and the error bounds of reduced basis approximation of the average of the output $\ell_1(u(t^{50}))$ at one time instance by the two different error estimators indexed with RB (non-weighted type) and wRB (weighted type), from which we can see that in order to achieve the same errors, we need much less reduced bases by the weighted scheme, thus achieving more efficiency. Moreover, we can observe that the average of the weighted a posteriori error estimator, $\triangle_{N,M}^{s,\rho_0}(t^{50}, y)$, is relatively sharp. Here we take the number of dual bases as $M = 1, 2, \ldots, M_{max} - 1, M_{max}$ and N linearly spanned between 1 and N_{max} with M_{max} values, where $M_{max} \leq N_{max}$.

We create the observation data at the value $y = (0.6, 1)$ with noise as specified before and compute the posterior density of y. The left part of Figure 9 shows the prior and the posterior density functions of half of the crack length y_1, from which we can observe that the posterior density function concentrates in a narrow region (depending on the measurement noise) at the preset value $y_1 = 0.6$. The convergence of the weighted reduced basis approximation error of the expectation of y_1 with respect to the finite element approximation is displayed in the right part of Figure 9, which confirms that the weighted reduced basis approximation is efficient (with a small number of degrees of freedom) in achieving high accuracy for posterior expectation. We remark that for the computation of the expectation, we used tensor-product grid in this two-dimensional inverse problems. For high-dimensional problems, more advanced techniques should be applied together with the weighted reduced basis method [23].

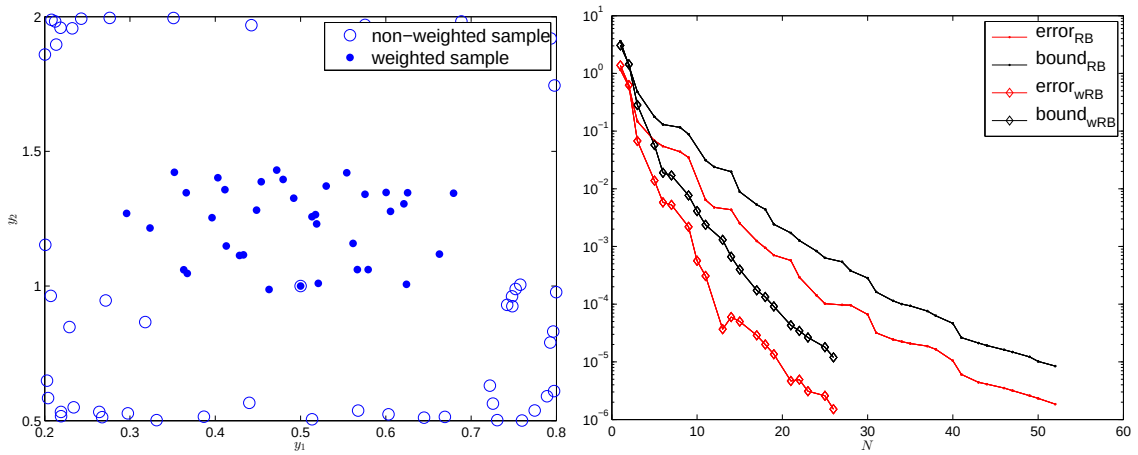

Figure 8. *Left: reduced basis samples constructed by the non-weighted error estimator (RB) and the weighted error estimator (wRB); right: decay of the error and error bound for the output* $\ell_1(u(t^{50}))$.

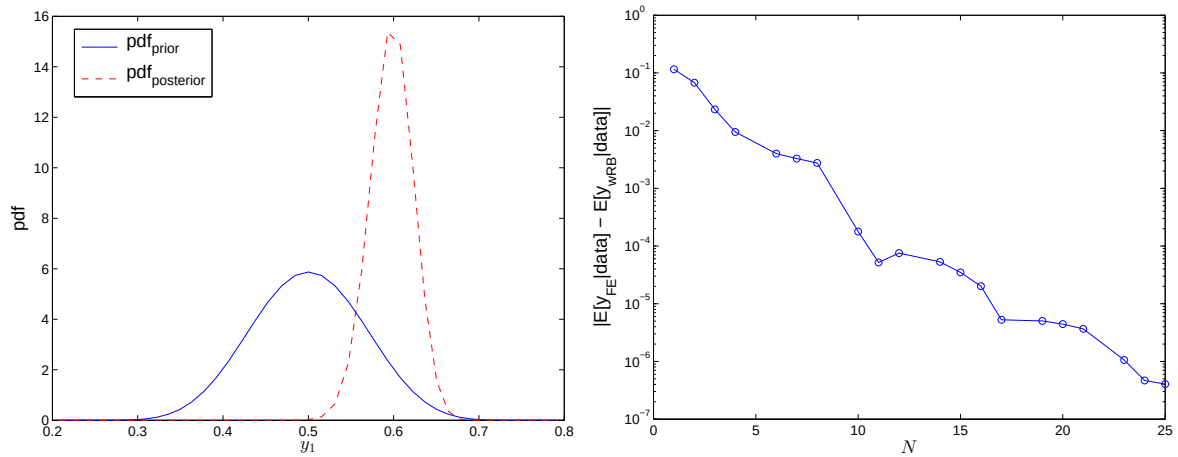

Figure 9. *Left: the prior and the posterior density functions (pdf) of* y_1*; right: decay of the weighted reduced basis approximation error for the expectation of* y_1 *conditioned on the measurement data.*

5.3. Fluid dynamics – variational data assimilation for blood flow. We consider a simplified blood flow model in the bifurcating human common carotid artery, with a typical geometry sketched in the left part of Figure (10) and a sample mesh in the right part.[3]

The blood flow pattern through this part of artery plays an important role, e.g. in investigating the development of vascular diseases, which may depend on many factors with uncertainties. Suppose the velocity on the cross-section can be measured along the dashed line, e.g. by MRI, then we are interested in using these data in the blood flow model in order to better characterize the velocity field. This process is known as data assimilation.

[3] A geometrical parametrization of a simplified carotid artery combined with a reduced order modelling techniques has been proposed in [62].

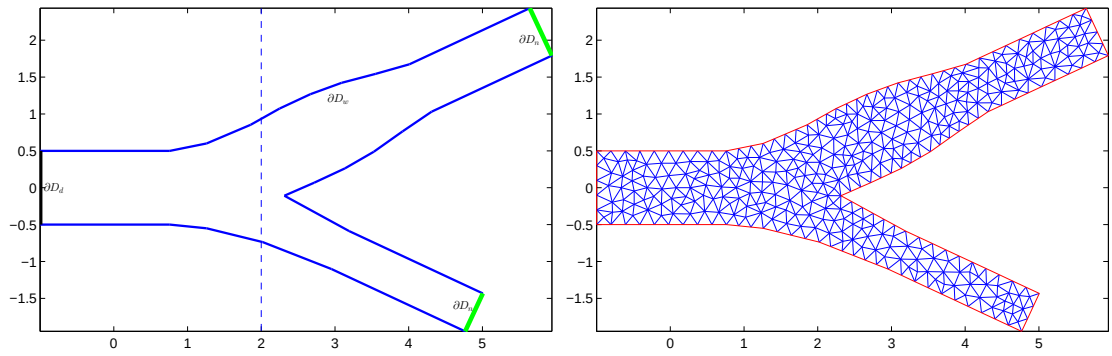

Figure 10. *Left: sketch of a simplified geometry of the common carotid artery bifurcation (in 2d) and the cross section of the observation data (along the dashed line); right: a sample mesh of the geometry.*

For simplicity, we consider Stokes equations [4] to describe the blood flow, written as: find $(u,p) \in U \times Q$, being $U = \{v \in (H^1(D))^2, v|_{\partial D_w} = 0\}$ and $Q = \{q \in L^2(D), \int_D q dx = 0\}$, such that

$$\begin{cases} \int_D \nu \nabla u \otimes \nabla v dx - \int_D (\nabla \cdot v) p dx &= \int_D f \cdot v dx \quad \forall v \in U, \\ \int_D (\nabla \cdot u) q dx &= 0 \quad \forall q \in Q, \end{cases} \tag{5.17}$$

where ν is the constant viscosity and set as 1, $u = (u_1, u_2)$ is the velocity field, p is the pressure field, f is a body force. We prescribe a parabolic velocity profile (Dirichelet boundary condition) $u = ((0.5 + x_2)(0.5 - x_2), 0)$ on the inflow boundary ∂D_d, assume homogeneous Neumann boundary condition on the outflow boundary ∂D_n and a homogeneous Dirichlet boundary condition on the arterial wall ∂D_w, i.e. on the rest of the boundary. Different realizations of the velocity field are possibly determined by many different factors under uncertainties, e.g. boundary conditions, external forces, wall roughness, etc. Here we assume for simplicity that the effect of these combined factors is expressed by a rather synthetic random body force f. More explicitly, we assume that the random force is given by the truncated Karhunen-Loéve expansion

$$f(x,y) = y_1 f_1(x) + \sum_{k=2}^{K} \frac{1}{\lambda_k} y_k f_k(x), \tag{5.18}$$

where λ_k, $2 \le k \le K$, are the eigenvalues of a covariance $\exp(-(x_1 - x_1')^2/L^2)$ with correlation length $L = 1/3$; the random variations y_k, $1 \le k \le K$, are assumed to be independent and uniformly distributed with zero mean and unit variance, i.e. $y_k \sim U(-\sqrt{3}, \sqrt{3})$; $f_1 = ((\sqrt{\pi}L/2)^{1/2}, 0)$ and $f_k = (\sin(k\pi x_1/2), 0)$ for k even and $f_k = (\cos((k-1)\pi x_1/2), 0)$ for k odd. Then a sample of the observation data is given by $u_o = (1 + \eta)u_y$, where u_y is the solution restricted to the observation location (the dashed line in Figure 10) of the Stokes equations (5.17) at a realization of the random force, $\eta \sim \mathcal{N}(0, \sigma^2)$ is a random variable of normal distribution with standard deviation $\sigma = 0.1$, i.e. a signal to noise ratio SNR = 10.

[4] We only consider the data assimilation at one time instance. In practical applications, more complex fluid equations such as time-dependent Navier-Stokes equations are more feasible to model the blood flow.

We employ the Lagrangian approach introduced in section 4.3 to solve the data assimilation problem. In particular, we write the weak formulation of the Stokes equations in the form of (2.1), where we define the bilinear form as

$$A((u,p),(v,q);y) = \int_D \nu \nabla u(y) \otimes \nabla v dx - \int_D (\nabla \cdot v) p(y) dx + \int_D (\nabla \cdot u) q dx \tag{5.19}$$

which satisfies the inf-sup condition (2.2) in the space $(U \times D) \times (U \times D)$, and the linear form

$$F((v,q);y) = \int_D \varsigma(y) \cdot v dx, \tag{5.20}$$

where $\varsigma(y)$ represents the distributed control function, which depends on the random variable y. Moreover, we specify the quantities in the cost functional (4.17) as $k = u_0$, $O(u(y)) = u_y$, $\lambda = 10^{-4}$ and $||\cdot||_D = |\cdot|$ is the Euclidean norm, while $||\cdot||_R = ||\cdot||_{(L^2(D)^2)}$. We build the reduced basis approximation of the high-dimensional parametric KKT problem (4.20) by greedy algorithm based on an adaptive sparse grid method with $K = 20$. More details for such construction, in particular for the stabilization of reduced Stokes problems and the reduced variational optimal control problems, i.e. the problems in reduced basis spaces, can be found in [18, 49] based on [**?**].

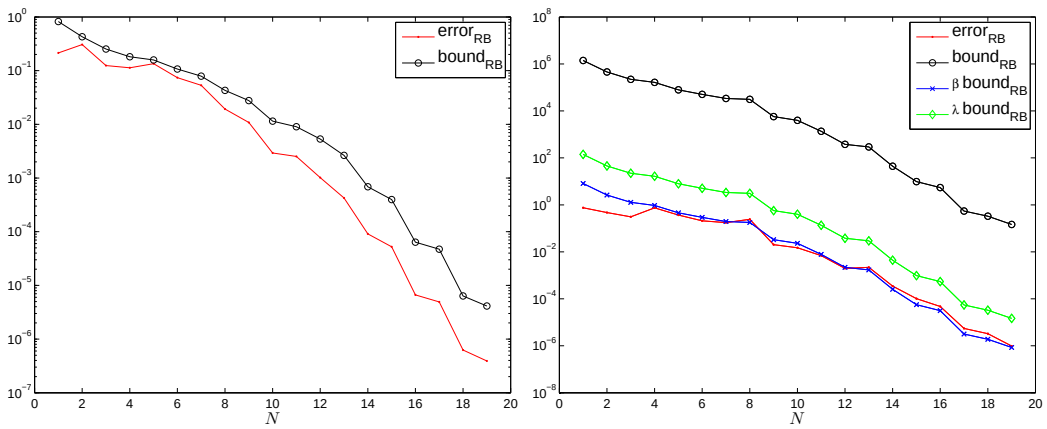

Figure 11. *The error and error bound of reduced basis approximation of the solution of the Stokes problem (left) and for the Stokes optimal control problem (right). β: inf-sup constant; λ: regularization constant.*

The Stokes equations is discretized by the finite element method with Taylor–Hood P2 elements (polynomials of degree 2) for velocity and P1 elements (polynomials of degree 1) for pressure, resulting in stable high-fidelity approximation [57]. For the reduced basis approximation of the parametric Stokes problems, we apply the "supremizer" enrichment of the reduced basis space for velocity, as introduced in (3.21), resulting in stable reduced basis problem. The decay of the approximation error and the a posteriori error bound for the reduced basis solution (u_N, p_N) at 100 random samples measured in $\sqrt{||\cdot||_U^2 + ||\cdot||_Q^2}$ norm is shown in the left part of Figure 11, where only the maximum error bound and the corresponding error is displayed. The error bound is reliable and also very sharp. We solve the

reduced basis approximation of the Lagrangian (KKT) system (4.20) by the reduced system (4.22). The approximation error and the a posteriori error bound for the optimal solution $(u_N^*, p_N^*, \varsigma_N^*, \tilde{u}_N^*, \tilde{p}_N^*)$ (being $(\tilde{u}_N^*, \tilde{p}_N^*)$ the adjoint variables of (u_N^*, p_N^*)) measured in the norm $\sqrt{||\cdot||_U^2 + ||\cdot||_Q^2 + \lambda||\cdot||_R^2 + ||\cdot||_U^2 + ||\cdot||_Q^2}$ is shown in the right part of Figure 11. The error bound is reliable but not very sharp compared to that for the approximation of the solution of the Stokes problem. This is because the stability constant β of the KKT system (4.20), depending on the regularization parameter λ, is rather small. The re-scaled error bound by the stability constant β and regularization constant λ are also displayed in the right part of Figure 11, from which we can observe that the rescaling with β leads to very sharp error bound and the rescaling with λ also results in much sharper error bound than that without rescaling. In order to improve the effectivity of the a posteriori error bound, the regularization and/or the stability constants should be taken into consideration.

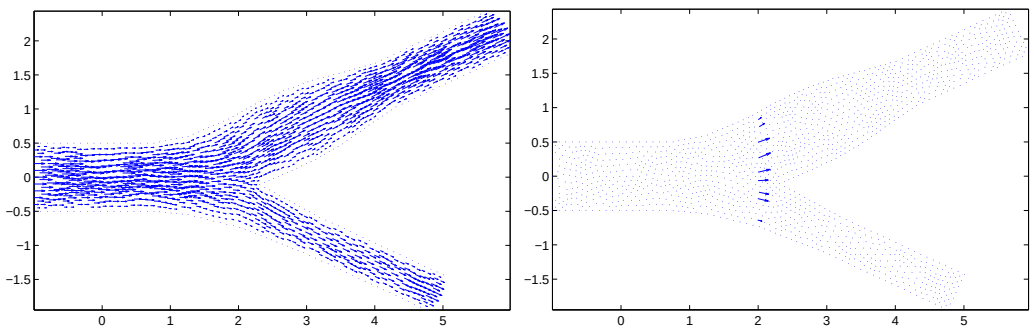

Figure 12. *Velocity field (left) at a random sample and the noisy observation with SNR = 10 (right).*

To demonstrate the effectivity of the data assimilation for blood flow, we first provide the velocity field by solving the high-fidelity Stokes problem at a random sample, as seen in the left part of Figure 12. Then we create the observation data as the velocity around $x_1 = 2$ with 10% noise, seen in the right part of Figure 12. With this observation data, we solve the reduced data assimilation problem, i.e. the reduced Lagrangian (KKT) system (4.22) using 19 bases. The assimilated velocity and pressure as well as the corresponding approximation errors are displayed in Figure 13, from which we can see that the errors for both velocity and pressure are small (approximately 1.93% for velocity in $||\cdot||_U$-norm and 0.41% for pressure in $||\cdot||_Q$-norm) and concentrate around the observation location due to the 10% noise.

We remark that the reduced basis approximation with 19 bases results in a very small system that can be solved in *real time*. In practice, besides the expository (linear) random body force as considered here (which is used to create the uncertainties and build reduced basis approximation), uncertainties may arise with explicit representation from many other different sources, in which cases the reduced data assimilation framework can be employed analogously. See, for example, [**?**] for ROM based on POD for time-dependent problems.

6. Conclusions and perspectives. We have demonstrated the efficiency and accuracy of the reduced order methods and extended suitable algorithms for some typical uncertainty quantification problems through three examples, concerning failure probability evaluation in structural mechanics, Bayesian inversion in heat conduction, and variational data assimilation

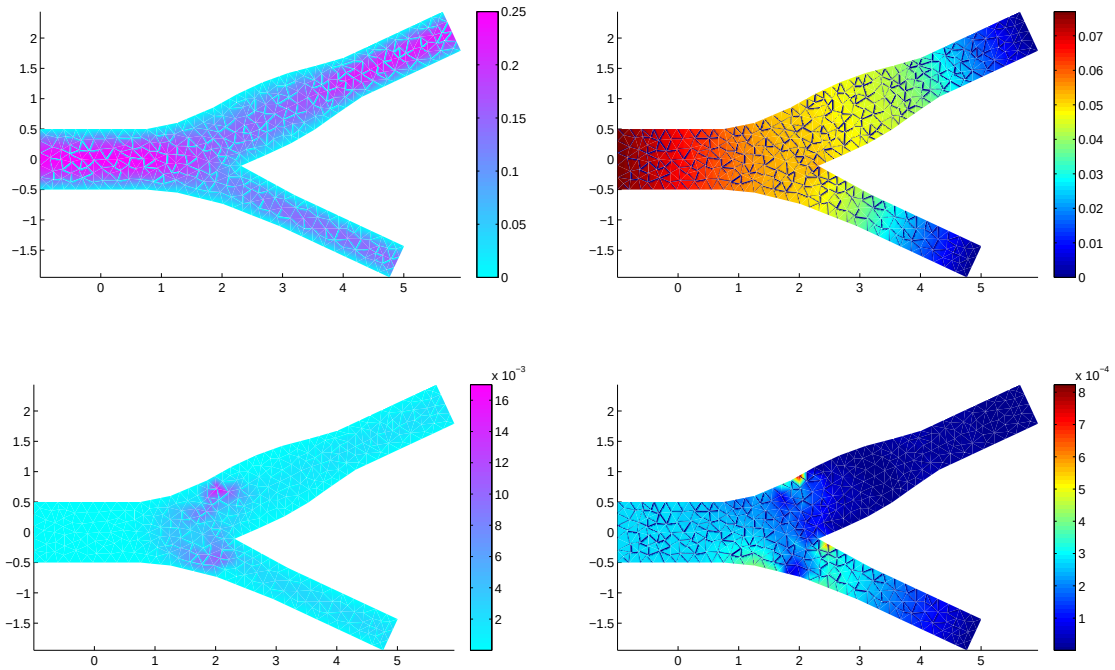

Figure 13. *Top: assimilated velocity field (left) and pressure (right); bottom: the corresponding errors.*

in fluid dynamics. Although rather academic, these examples have highlighted the remarkable reduction of computational costs, demonstrating the potential of reduced order methods in solving uncertainty quantification problems when the solution manifold and/or the manifold of the quantity of interest are low dimensional.

Several computational and mathematical challenges still need to be tackled in order to pave the way for more general and practical applications. The first is curse of dimensionality, where a larger number of random variables are present in the underpinning system, more advanced sampling techniques are required in order to build the efficient and accurate reduced basis space. Dimension-adaptive sparse grid sampling [16], importance sampling, multilevel construction, are all in active development for reduced order methods in high-dimensional uncertainty quantification problems. Secondly, highly nonlinear problems expose reduced order methods to additional substantial difficulty for computational reduction, for which several techniques have been recently developed such as empirical interpolation and its discrete version [4, 37, 13, 22], best point interpolation [50], gappy POD [30, 69], Gauss Newton with approximated tensor [11], etc. However, these techniques do not necessarily guarantee the well-posedness of the approximated problem, in particular its stability, which is necessary to enable reliable computational reduction. Furthermore, in the construction of reduced basis spaces, it is crucial to balance the errors arising from high-fidelity and reduced basis approximation, which lead to the total error of the computational approximation of the underpinning parametric/stochastic problems. Other challenges, such as long-time integration behaviour,

nonlinear conservation laws and multiscale and multiphysics coupling are also important for the development of reduced order methods at large, and particularly for their application to uncertainty quantification problems.

Acknowledgements. Peng Chen acknowledges the support at CMCS, MATHICSE, EPFL for most of this work and the support by ERC project AdG 247277 at ETH Zurich when this work is finished. G. Rozza acknowledges NOFYSAS excellence grant initiative of SISSA and INDAM-GNCS activity group. The MATLAB library rbMIT is acknowledged.

REFERENCES

[1] I. Babuška, K.M. Liu, and R. Tempone. Solving stochastic partial differential equations based on the experimental data. *Mathematical Models and Methods in Applied Sciences*, 13(3):415–444, 2003.

[2] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Review*, 52(3):317, 2010.

[3] I. Babuška, R. Tempone, and G.E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM Journal on Numerical Analysis*, 42(2):800–825, 2004.

[4] M. Barrault, Y. Maday, N.C. Nguyen, and A.T. Patera. An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique, Analyse Numérique*, 339(9):667–672, 2004.

[5] A. Barth, C. Schwab, and N. Zollinger. Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients. *Numerische Mathematik*, 119(1):123–161, 2011.

[6] Gal Berkooz, Philip Holmes, and John L Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual review of fluid mechanics*, 25(1):539–575, 1993.

[7] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence rates for greedy algorithms in reduced basis methods. *SIAM Journal of Mathematical Analysis*, 43(3):1457–1472, 2011.

[8] G. Blatman and B. Sudret. Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach. *Comptes Rendus Mécanique*, 336(6):518–523, 2008.

[9] S. Boyaval, C. Le Bris, T. Lelièvre, Y. Maday, N.C. Nguyen, and A.T. Patera. Reduced basis techniques for stochastic problems. *Archives of Computational Methods in Engineering*, 17:435–454, 2010.

[10] C.G. Bucher and U. Bourgund. A fast and efficient response surface approach for structural reliability problems. *Structural Safety*, 7(1):57–66, 1990.

[11] K. Carlberg, C. Bou-Mosleh, and C. Farhat. Efficient non-linear model reduction via a least-squares Petrov–Galerkin projection and compressive tensor approximations. *International Journal for Numerical Methods in Engineering*, 86(2):155–181, 2011.

[12] A. Chatterjee. An introduction to the proper orthogonal decomposition. *Current science*, 78(7):808–817, 2000.

[13] S. Chaturantabut and D.C. Sorensen. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764, 2010.

[14] P. Chen. *Model order reduction techniques for uncertainty quantification problems*. PhD thesis, EPFL, 2014.

[15] P. Chen and A. Quarteroni. Accurate and efficient evaluation of failure probability for partial differential equations with random input data. *Computer Methods in Applied Mechanics and Engineering*, 267(0):233–260, 2013.

[16] P. Chen and A. Quarteroni. A new algorithm for high-dimensional uncertainty problems based on dimension-adaptive and reduced basis methods. *EPFL, MATHICSE Report 09, submitted*, 2014.

[17] P. Chen and A. Quarteroni. Weighted reduced basis method for stochastic optimal control problems with elliptic PDE constraints. *SIAM/ASA J. Uncertainty Quantification*, 2(1):364–396, 2014.

[18] P. Chen, A. Quarteroni, and G. Rozza. Multilevel and weighted reduced basis method for stochastic optimal control problems constrained by Stokes equations. *Submitted*, 2013.

[19] P. Chen, A. Quarteroni, and G. Rozza. Stochastic optimal Robin boundary control problems of advection-dominated elliptic equations. *SIAM Journal on Numerical Analysis*, 51(5):2700 – 2722, 2013.

[20] P. Chen, A. Quarteroni, and G. Rozza. A weighted reduced basis method for elliptic partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 51(6):3163 – 3185, 2013.

[21] P. Chen, A. Quarteroni, and G. Rozza. Comparison of reduced basis and stochastic collocation methods for elliptic problems. *Journal of Scientific Computing*, 59:187–216, 2014.

[22] P. Chen, A. Quarteroni, and G. Rozza. A weighted empirical interpolation method: a priori convergence analysis and applications. *ESAIM: Mathematical Modelling and Numerical Analysis*, 48:943–953, 7 2014.

[23] P. Chen and C. Schwab. Sparse grid, reduced basis Bayesian inversion. *ETH Zurich, SAM Report 36, submitted*, 2014.

[24] M.K. Deb, I.M. Babuska, and J.T. Oden. Solution of stochastic partial differential equations using Galerkin finite element techniques. *Computer Methods in Applied Mechanics and Engineering*, 190(48):6359–6372, 2001.

[25] L. Demkowicz and J. Gopalakrishnan. A class of discontinuous petrov–galerkin methods. ii. optimal test functions. *Numerical Methods for Partial Differential Equations*, 27(1):70–105, 2011.

[26] R. DeVore, G. Petrova, and P. Wojtaszczyk. Greedy algorithms for reduced bases in banach spaces. *Constructive Approximation*, 37(3):455–466, 2013.

[27] J. Dick, F.Y. Kuo, and I.H. Sloan. High-dimensional integration–the Quasi-Monte Carlo way. *Acta Numerica*, 22:133–288, 2013.

[28] M. Drohmann, B. Haasdonk, and M. Ohlberger. Reduced basis approximation for nonlinear parametrized evolution equations based on empirical operator interpolation. *SIAM Journal on Scientific Computing*, 34(2):A937–A969, 2012.

[29] J.L. Eftang, A.T. Patera, and E.M. Rønquist. An "hp" certified reduced basis method for parametrized elliptic partial differential equations. *SIAM Journal on Scientific Computing*, 32(6):3170–3200, 2010.

[30] R. Everson and L. Sirovich. Karhunen–Loève procedure for gappy data. *Journal of the Optical Society of America A*, 12(8):1657–1664, 1995.

[31] L. Faravelli. Response-surface approach for reliability analysis. *Journal of Engineering Mechanics*, 115(12):2763–2781, 1989.

[32] G.S. Fishman. *Monte Carlo: Concepts, Algorithms, and Applications*. Springer, New York, 1996.

[33] T. Gerstner and M. Griebel. Dimension–adaptive tensor–product quadrature. *Computing*, 71(1):65–87, 2003.

[34] R.G. Ghanem and P.D. Spanos. *Stochastic Finite Elements: a Spectral Approach*. Dover Civil and Mechanical Engineering, Courier Dover Publications, Springer-Verlag, New York, 1991.

[35] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman & Hall/CRC, 1996.

[36] G.H. Golub and C.F. Van Loan. *Matrix computations*, volume 3. JHU Press, 2012.

[37] M.A. Grepl, Y. Maday, N.C. Nguyen, and A.T. Patera. Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 41(03):575–605, 2007.

[38] M.A. Grepl and A.T. Patera. A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 39(01):157–181, 2005.

[39] Max D. Gunzburger and P. B. Bochev. Finite element methods for optimization and control problems for the Stokes equations. *Comp. Math. Appl.*, 48:1035–1057, 2004.

[40] Max D. Gunzburger, Janet S. Peterson, and John N. Shadid. Reduced-order modeling of time-dependent PDEs with multiple parameters in the boundary data. *Computer Methods in Applied Mechanics and Engineering*, 196(46):1030 – 1047, 2007.

[41] B. Haasdonk and M. Ohlberger. Reduced basis method for finite volume approximations of parametrized linear evolution equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 42(02):277–302, 2008.

[42] B. Haasdonk, K. Urban, and B. Wieland. Reduced basis methods for parameterized partial differential equations with stochastic influences using the Karhunen–Loève expansion. *SIAM/ASA J. Uncertainty Quantification*, 1(1):79–105, 2013.

[43] J. Hesthaven, B. Stamm, and S. Zhang. Efficient greedy algorithms for high-dimensional parameter spaces with applications to empirical interpolation and reduced basis methods. *ESAIM: Mathematical Modelling and Numerical Analysis*, 48(1):259–283, 2011.

[44] D.B.P. Huynh and A.T. Patera. Reduced basis approximation and a posteriori error estimation for stress intensity factors. *International Journal for Numerical Methods in Engineering*, 72(10):1219–1259, 2007.

[45] D.B.P Huynh, G. Rozza, S. Sen, and A.T. Patera. A successive constraint linear optimization method for lower bounds of parametric coercivity and inf-sup stability constants. *Comptes Rendus Mathematique, Analyse Numérique*, 345(8):473–478, 2007.

[46] K. Karhunen. Über lineare methoden in der wahrscheinlichkeitsrechnung. *Annales Academiae Scientiarum Fennicae, Series A.I. Mathematica-Phys.*, 37:1–79, 1947.

[47] J. Li, J. Li, and D. Xiu. An efficient surrogate-based method for computing rare failure probability. *Journal of Computational Physics*, 230(24):8683–8697, 2011.

[48] J. Li and D. Xiu. Evaluation of failure probability via surrogate models. *Journal of Computational Physics*, 229(23):8966–8980, 2010.

[49] X. Ma and N. Zabaras. An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations. *Journal of Computational Physics*, 228(8):3084–3113, 2009.

[50] G. Migliorati, F. Nobile, E. Von Schwerin, and R. Tempone. Approximation of quantities of interest in stochastic PDEs by the random discrete L2 projection on polynomial spaces. *SIAM Journal on Scientific Computing*, 35(3):A1440–A1460, 2013.

[51] F. Negri, G. Rozza, and A. Manzoni. Certified reduced basis method for parametrized optimal control problems governed by Stokes equations. *submitted*, 2013.

[52] N.C. Nguyen, A.T. Patera, and J. Peraire. A best points interpolation method for efficient approximation of parametrized functions. *International Journal for Numerical Methods in Engineering*, 73(4):521–543, 2008.

[53] N.C. Nguyen, G. Rozza, D.B.P. Huynh, and A.T. Patera. Reduced basis approximation and a posteriori error estimation for parametrized parabolic PDEs; application to real-time Bayesian parameter estimation. *Biegler, Biros, Ghattas, Heinkenschloss, Keyes, Mallick, Tenorio, van Bloemen Waanders, and Willcox, editors, Computational Methods for Large Scale Inverse Problems and Uncertainty Quantification, John Wiley & Sons, UK*, 2009.

[54] F. Nobile, R. Tempone, and C.G. Webster. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2411–2442, 2008.

[55] F. Nobile, R. Tempone, and C.G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, 2008.

[56] 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(45-48):4521–4537, 2007.

[57] A.T. Patera and G. Rozza. Reduced basis approximation and a posteriori error estimation for parametrized partial differential equations. *Copyright MIT, http://augustine.mit.edu*, 2007.

[58] C. Prudhomme, Y. Maday, A.T. Patera, G. Turinici, D.V. Rovas, K. Veroy, and L. Machiels. Reliable real-time solution of parametrized partial differential equations: reduced-basis output bound methods. *Journal of Fluids Engineering*, 124(1):70–80, 2002.

[59] A. Quarteroni. *Numerical Models for Differential Problems.* Springer Milano, MS & A, vol 8, 2nd ed., 2014.

[60] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced Basis Methods for Partial Differential Equations. An Introduction.* Springer, UNITEXT Series, to appear, 2015.

[61] A. Quarteroni, G. Rozza, and A. Manzoni. Certified reduced basis approximation for parametrized partial differential equations and applications. *Journal of Mathematics in Industry*, 1(1):1–49, 2011.

[62] R. Rackwitz. Reliability analysis – a review and some perspectives. *Structural Safety*, 23(4):365–395, 2001.

[63] G. Rozza, D.B.P. Huynh, and A.T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Archives of Computational*

Methods in Engineering, 15(3):229–275, 2008.

[64] G. Rozza, A. Manzoni, and F. Negri. Reduced strategies for pde-constrained optimization problems in haemodynamics. In *Proceedings of the 6th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS), Vienna, Austria*, 2012.

[65] G.I. Schuëller, H.J. Pradlwarter, and P.S. Koutsourelakis. A critical appraisal of reliability estimation procedures for high dimensions. *Probabilistic Engineering Mechanics*, 19(4):463–474, 2004.

[66] C. Schwab and R. A. Todor. Karhunen–Loève approximation of random fields by generalized fast multipole methods. *Journal of Computational Physics*, 217(1):100–122, 2006.

[67] C. Schwab and R.A. Todor. Sparse finite elements for elliptic problems with stochastic loading. *Numerische Mathematik*, 95(4):707–734, 2003.

[68] A.M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numerica*, 19(1):451–559, 2010.

[69] G. Turinici, C. Prud'Homme, A.T. Patera, Y. Maday, and A. Buffa. A priori convergence of the greedy algorithm for the parametrized reduced basis method. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46(3):595, 2012.

[70] K. Urban and A.T. Patera. A new error bound for reduced basis approximation of parabolic partial differential equations. *Comptes Rendus Mathematique*, 350(3):203–207, 2012.

[71] K. Willcox. Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition. *Computers & fluids*, 35(2):208–226, 2006.

[72] K. Willcox and J. Peraire. Balanced model reduction via the proper orthogonal decomposition. *AIAA journal*, 40(11):2323–2330, 2002.

[73] D. Wolfgang, P. Christian, and W. Gerrit. Double greedy algorithms: reduced basis methods for transport dominated problems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 48(03):623–663, 2014.

[74] D. Xiu and J.S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM Journal on Scientific Computing*, 27(3):1118–1139, 2005.

[75] D. Xiu and G.E. Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.

Recent publications:

MATHEMATICS INSTITUTE OF COMPUTATIONAL SCIENCE AND ENGINEERING
Section of Mathematics
Ecole Polytechnique Fédérale
CH-1015 Lausanne

37.2014 LARS KARLSSON, DANIEL KRESSNER, ANDRÉ USCHMAJEW:
Parallel algorithms for tensor completion in the CP format

38.2014 PAOLO TRICERRI, LUCA DEDÈ, SIMONE DEPARIS, ALFIO QUARTERONI, ANNE M. ROBERTSON, ADÉLIA SEQUEIRA:
Fluid-structure interaction siumlations of cerebral arteries modeled by isotropic and anistropic constitutive laws

39.2014 FRANCESCA BONIZZONI, FABIO NOBILE, DANIEL KRESSNER:
Tensor train approximation of moment equations for the log-normal Darcy problem

40.2014 DIANE GUIGNARD, FABIO NOBILE, MARCO PICASSO:
A posteriori error estimations for elliptic partial differential equations with small uncertainties

41.2014 FABIO NOBILE, LORENZO TAMELLINI, RAÚL TEMPONE:
Comparison of Clenshaw-Curtis and Leja quasi-optimal sparse grids for the approximation of random PDEs

42.2014 ASSYR ABDULLE, PATRICK HENNING:
A reduced basis localized orthogonal decomposition

43.2014 PAOLA F. ANTONIETTI, ILARIO MAZZIERI, ALFIO QUARTERONI:
Improving seismic risk protection through mathemactical modeling

44.2014 LUCA DEDÈ, ALFIO QUARTERONI, SEHNGFENG ZHU:
Isogeometric analysis and proper orthogonal decomposition for parabolic problems

45.2014 ZVONIMIR BUJANOVIC, DANIEL KRESSNER:
A block algorithm for computing antitriangular factorizations of symmetric matrices

46.2014 ASSYR ABDULLE:
The role of numerical integration in numerical in numerical homogenization

47.2014 ANDREA MANZONI, STEFANO PAGANI, TONI LASSILA:
Accurate solution of Bayesian inverse uncertainty quantification problems using model and error reduction methods

48.2014 MARCO PICASSO:
From the free surface flow of a viscoelastic fluid towards the elastic deformation of a solid

49.2014 FABIO NOBILE, FRANCESCO TESEI:
A multi level Monte Carlo method with control variate for elliptic PDEs with log-normal coefficients

***

01.2015 PENG CHEN, ALFIO QUARTERONI, GIANLUIGI ROZZA:
Reduced order methods for uncertainty quantification problems