ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

Sparse-Grid, Reduced-Basis Bayesian Inversion: Nonaffine-Parametric Nonlinear Equations

P. Chen and Ch. Schwab

Research Report No. 2015-21
June 2015

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Funding ERC: 247277 STAHDPDE

Sparse-Grid, Reduced-Basis Bayesian Inversion: Nonaffine-Parametric Nonlinear Equations

Peng Chen and Christoph Schwab

Seminar für Angewandte Mathematik, Eidgenössische Technische Hochschule, CH-8092 Zürich, Switzerland.

Abstract

We extend the reduced basis accelerated Bayesian inversion methods for affine-parametric, linear operator equations which are considered in [15, 16] to non-affine, nonlinear parametric operator equations. We generalize the analysis of sparsity of parametric forward solution maps in [18] and of Bayesian inversion in [41, 42] to the fully discrete setting, including Petrov-Galerkin high-fidelity ("HiFi") discretization of the forward maps. We develop adaptive, stochastic collocation based reduction methods for the efficient computation of reduced bases on the parametric solution manifold. The nonlinearity with respect to the distributed, uncertain parameters and the unknown solution is collocated; specifically, by the so-called Empiricial Interpolation Method (EIM). For the corresponding Bayesian inversion problems, computational efficiency is enhanced in two ways: first, expectations with respect to the posterior are computed by adaptive quadratures with dimension-independent convergence rates proposed in [42]; the present work generalizes [42] to account for the impact of the PG discretization in the forward maps on the expectation of the Quantities of Interest (QoI). Second, we propose to perform the Bayesian estimation only with respect to a parsimonious, reduced basis approximation of the posterior density. In [42], under general conditions on the forward map, the infinite-dimensional parametric, deterministic Bayesian posterior was shown to admit N-term approximations which converge at rates which depend only on the sparsity of the parametric forward map. We present dimension-adaptive collocation algorithms to build finite-dimensional parametric surrogates. In several numerical experiments, the proposed algorithms exhibit dimension-independent convergence rates which equal, at least, the currently known rate estimates for N-term approximation. We propose to accelerate Bayesian estimation by offline computation of reduced basis surrogates of the Bayesian posterior density. The parsimonious surrogates can be employed for online data assimilation and for Bayesian estimation. They also open a perspective for optimal experimental design.

Keywords:
Bayesian inversion, sparse grid, reduced basis, generalized empirical interpolation, greedy algorithm, high-fidelity, Petrov-Galerkin Finite Elements, goal-oriented a-posteriori error estimate, a-priori error estimate, uncertainty quantification

Email address: {peng.chen,christoph.schwab}@sam.math.ethz.ch (Peng Chen and Christoph Schwab)

Contents

1. Introduction

Efficient computational response prediction for complex systems in engineering and in the sciences, governed by partial differential or integral equations, subject to uncertain input parameters and conditional on noisy measurements, is a core task in computational Uncertainty Quantification (UQ), see e.g. [35, 45, 31, 27, 34, 5, 41, 21, 15, 16]. In the general framework of Bayesian estimation, this amounts to numerical evaluation of a mathematical expectation, with respect to the Bayesian posterior measure. Computationally, this boils down to "numerical integration", with respect to this posterior measure. Assuming additive gaussian observation noise with covariance $\Gamma > 0$, according to Bayes' theorem the posterior density depends on the negative log-likelihood of the (covariance weighted) quadratic observation-to-response mismatch. Therefore, given observation data, Bayesian estimation entails integration over all possible realizations of the uncertain input. If these inputs consist of a finite number of state variables, therefore, Bayesian estimation amounts to an integration problem over a finite-dimensional parameter space. Here, we are interested in the case of *"distributed uncertain input data"*, such as uncertain spatially heterogeneous permeabilities, constitutive parameters, loadings or uncertain domains of definition. Their exact, parametric description requires a *countable number of parameters*, such as for example a sequence of Fourier coefficients, or of spectral modes or wavelet coefficients. Bayesian estimation for such inputs then becomes a problem of *integration over function space*. Upon introduction of a basis in the space of uncertain input data results a (formally) infinite-dimensional, parametric integration problem; approximating parametric representations of uncertain inputs by truncated basis expansions renders these problems finite- but possibly very high-dimensional. Due to the curse of dimensionality, standard numerical integration methods then exhibit low convergence rates in terms of the number of "samples" (each "sample" corresponding to one so-called forward numerical solution of a differential or integral equation). To overcome the curse of dimensionality, *sampling methods* such as Markov Chain Monte-Carlo (MCMC) methods, have established themselves for the approximate computation of these expectations, in terms of sample averages, with samples drawn from the Bayesian posterior density. They have the advantage of dimension-independent convergence rate 1/2 in terms of N, the number of PDE "forward" solves, and the possibility of updating the sampler dynamically on datastreams. In [44], we proved sparsity results for the countably-parametric family of Bayesian posterior densities which implied that, in principle, dimension-independent N-term approximation rates $> 1/2$ were feasible provided that the parametric data-to-response maps have sufficient sparsity. Adaptive, deterministic quadrature algorithms which realized, in numerical experiments such high, dimension independent convergence rates were first presented in [41, 42]. Despite the drastic reduction in the number of forward solves afforded by these deterministic quadrature methods, for systems goverend by partial differential equations the high cost per solve renders these methods prohibitive, albeit superior to MCMC and other sampling methods.

Following earlier work [12, 11] by the first author, in the present paper we extend our work [15, 16] on accelerating the sparse deterministic quadrature [41, 42] for Bayesian inversion by forward solutions through reduced basis surrogates. Specifically, we cover nonlinear as well as non-affine parametric forward problems, including Bayesian inversion for (smooth) nonlinear partial differential equations with uncertain loadings, uncertain coefficients, and in uncertain domains. While the main conclusions of this work are analogous to our findings in the case of affine-parametric, linear operator equations in [15], we introduce in the present paper additional algorithmic developments to deal with strong nonlinearities in parametric dependence as well as for governing equations.

Reduced basis (RB) [38, 14], and more general model order reduction techniques, were developed to accelerate large-scale computations in engineering by compressing the so-called "high-fidelity" solution manifold to a possibly small subspace, with certified approximation accuracy on a given set of input data. This implies obvious CPU time savings in case that many instances of the problem need to be solved. Recently, this concept has been extended to multi-parametric PDEs which are to be solved on possibly high-dimensional parameter spaces. Here, RB solution at any given parameter is sought by solving a RB compression problem formulated through either Galerkin or Petrov-Galerkin approximation. If the (HiFi) solution manifold is low dimensional, the RB compression can dramatically reduce the computational cost in the large number of approximate solutions of the HiFi problem. The quasi-optimal approximation property of the RB compression, as compared to the rates afforded by the Kolmogorov N-width, has been proved in [4, 2]. Dimension-independent bounds on N widths of solution manifolds for high-dimensional, parametric problems were recently proved for a wide range of nonlinear operator equations in [18], thereby establishing a foundation of RB algorithms for such problems. The scope of problems includes stochastic optimal control [10, 13] by multilevel RB construction, risk prediction [9] with hybrid evaluation, statistical moment evaluations for arbitrary probability distribution by a weighted version of RB [12], etc. Recently, RB compression for Bayesian inverse problems have also been considered in [36, 31, 27, 30, 21]. In [36], RB methods are presented for low dimensional Bayesian inverse problems, full tensor-grid numerical quadrature was used. Key to the efficient computational realization of RB methods are

efficient and low-cost a posteriori error estimates and greedy algorithms. In [27], a non-linear model reduction problem is solved by using masked projection of the discrete equations. A parameter and state model reduction method is introduced in [31] for the reduction in both the state variable dimension and the parameter dimension, where the parameter is taken as the discrete value of the coefficient field at each HiFi discrezation node. A combined computational and geometrical framework for model order reduction and estimation is presented in [30] in the context of hemodynamics. More recently, a data-driven model reduction is introduced in [21] in combination with Monte-Carlo sampling techniques. There, increased efficiency of sampling according to the posterior density instead of the prior density is demonstrated computationally. These methods are either limited to low dimensional parametric forward models or lack efficient algorithms to construct the reduced basis functions without taking advantage of the sparsity of the parametric forward solution and the posterior density function.

In order to deal with the nonaffinity and nonlearity in the RB compression, we adopt the (generalized) empirical interpolation method ((G)EIM) [32]. We carry out an a priori error analysis of the interpolation error, and show that it can achieve dimension-independent convergence rate under the assumption of ℓ^p sparsity and holomorphic dependence of the residual w.r.t. the parameters. A goal-oriented a posteriori error estimate for the interpolation error is also developed. In order to achieve better effectivity of the a posteriori error estimate, we compute surrogates by EIM of the nonaffine and nonlinear terms.

In the numerical experiments for demonstration of the proposed algorithms, we consider a nonlinear (semi-linear) elliptic problem with uncertainty from diffusion coefficient field, and a nonaffine and nonlinear problem with shape uncertainty. The sparse grid interpolation and integration error, the high-fidelity discretization error, the reduced basis compression error as well as the empirical interpolation error are reported for the two problems. Different dimensions, ranging from a few tens to a few hundreds, and different sparsity parameters are used to demonstrate the dependence of the convergence rates. The effectivity of the goal-oriented a posteriori error estimate is also reported for both reduced basis compression and empirical interpolation for the two problems.

The structure of the paper is as follows: in Section 2, we review the Bayesian approach to inverse problems, recently developed in [20, 22], for abstract operator equations with "distributed uncertainty", ie., with uncertain input data which may take values in infinite-dimensional function spaces. Upon *uncertainty parametrization* with an unconditional basis $\mathbf{\Psi}$ in the space X of uncertainties, the forward problem becomes a countably parametric operator equation. Generalizing our work [15, 16], we admit smooth, nonlinear operator equations and nonlinear parameter dependence including, for example, domain uncertainty, and semilinar elliptic and parabolic PDEs, as well as boundary integal equation formulations of forward problems. We also review in Section 2 results on *sparsity* of *holomorphic-parametric* forward maps, from [17, 18] and the references there, which entail, as we show by generalizing [41], corresponding sparsity of the Bayesian posterior densities. We include moreover also the error analysis for abstract Petrov-Galerkin discretizations of the parametric forward problems thereby accomodating all mixed Finite Element Methods for viscous, incompressible flows which arise, eg., in mathematical modelling of biofluid flow problems [26].

Section 3 then presents a class of dimension-adaptive sparse grid algorithms to sweep the parameter space of the uncertain inputs, which will be used for empirical interpolation in the reduced basis construction as well as for numerical integration in the Bayesian estimation.

Section 4 presents the error analysis of general Petrov-Galerkin discretizations of the parametric forward problems, generalizing our earlier work [41, 42] where this discretization error was not explicitly analyzed.

Section 5 finally develops all algorithms necessary for the RB compression of the forward maps in the present context, drawing on previous work [38, 15, 8]. The nonlinear dependence on the parameter sequence is handled by the empirical interpolation method (EIM) and its variants. We present, in particular, a novel greedy strategy for computing a *parsimonious computational RB compression of the Bayesian posterior density*. The existence of approximate, low-parametric surrogate Bayesian densities follows from the abstract sparsity results presented in Section 2, and in [41, 42]. Importantly, this approximation can be computed offline. The presently proposed algorithms therefore allow for *fast, online Bayesian estimation*. The RB posterior density can also be used for efficient computational data-to-response sensitivity analysis, thereby opening the perspective of *online experimental design*, an aspect of our work which is to be developed elsewhere.

In Section 6, we present several numerical experiments for nonlinear, elliptic problems with uncertain coefficients and in uncertain domains.

We include mathematical analysis of approximation errors, sparsity results and dimension-independent convergence rate estimates in an Appendix, Section 8.

2. Bayesian Inversion of Parametric Operator Equations

We consider Bayesian inversion of a class of abstract, smooth, nonlinear parametric operator equations. Admissible problems comprise elliptic and parabolic PDEs with uncertain coefficients, source terms and domains. We review classical results on well-posedness of these equations, with particular attention to the case when the "parameter" is related to an uncertain, "distributed" input u taking values in an infinite-dimensional, separable Banach space X. Assuming at hand an unconditional basis $\boldsymbol{\Psi} = \{\psi_j\}_{j\geq 1}$ of X, uncertain inputs are associated with infinite sequences of parameters. We review Baysian estimation of response functionals ("Quantities of Interest") as mathematical expectationsa over all admissible inputs from X, conditional on given data $\delta \in Y$, based on [22]. Upon uncertainty parametrization with the basis $\boldsymbol{\Psi}$, we convert the problem of Bayesian estimation with respect to a given Bayesian prior measure $\boldsymbol{\pi}_0$ on X into an integration problem over an infinite-dimensional coordinate set associated with the basis $\boldsymbol{\Psi}$. In Section 2.6, we present a generalization of results from [18, 41, 42] stating that for *holomorphic, parametric operator equations* the parametric forward response maps and the Bayesian posterior densities are sparse. This allows for dimension-independent N-term approximation rates limited only by sparsity. We also account for the effect of Petrov-Galerkin discretizations of the forward maps on the error in the Bayesian estimate, which effect was not considered in the mentioned references.

2.1. Nonlinear operator equations with uncertain input

Let $\mathcal{X}$ and $\mathcal{Y}$ denote separable Hilbert spaces. For a distributed, uncertain parameter $u \in X$, we consider a "forward" operator $\mathcal{R}(q;u)$ depending on u and acting on $q \in \mathcal{X}$. We assume known a *"nominal parameter instance" $\langle u\rangle \in X$ (such as, for example, the expectation of an X-valued random field u),* and consider, for $u \in B_X(\langle u\rangle; R)$, an open ball of sufficiently small radius $R > 0$ in X centered at a nominal input instance $\langle u\rangle \in X$, the nonlinear operator equation

$$\text{given } \ u \in B_X(\langle u\rangle; R)\ , \text{find } q \in \mathcal{X} \quad \text{s.t.} \quad {}_{\mathcal{Y}'}\langle \mathcal{R}(q;u), v\rangle_{\mathcal{Y}} = 0 \quad \forall v \in \mathcal{Y}\ . \tag{1}$$

Given $u \in B_X(\langle u\rangle; R)$, we call a solution q_0 of (1) *regular at u* if and only if $\mathcal{R}(\cdot;u)$ is differentiable with respect to q and if the differential $D_q\mathcal{R}(q_0;u) \in \mathcal{L}(\mathcal{X};\mathcal{Y}')$ is an isomorphism. For the well-posedness of operator equations involving $\mathcal{R}(q;u)$, we assume the map $\mathcal{R}(\cdot;u) : \mathcal{X} \mapsto \mathcal{Y}'$ admits a family of regular solutions *locally, in an open neighborhood of the nominal parameter instance $\langle u\rangle \in X$.*

Assumption 1. *We assume the structural conditions*

$$\mathcal{R}(q;u) = A(q;u) - F(u) \quad \textit{in} \quad \mathcal{Y}'\ , \tag{2}$$

and that for all u in a sufficiently small, closed neighborhood $\tilde{X} \subseteq X$ of $\langle u\rangle \in X$ the parametric forward problem: for every $u \in \tilde{X} \subseteq X$, given $F(u) \in \mathcal{Y}'$, find $q(u) \in \mathcal{X}$ such that the residual equation (1) *is well-posed. Ie., for every fixed $u \in \tilde{X} \subset X$, and for every $F(u) \in \mathcal{Y}'$, there exists a unique solution $q(u)$ of* (1) *which depends continuously on u.*

We call the set $\{(q(u),u) : u \in \tilde{X}\} \subset \mathcal{X} \times X$ a *regular branch of solutions* of (2) if

$$\begin{aligned} &\tilde{X} \ni u \mapsto q(u) \text{ is continuous as mapping from } X \mapsto \mathcal{X}\ , \\ &\mathcal{R}(q(u);u) = 0 \quad \text{in} \quad \mathcal{Y}'\ . \end{aligned} \tag{3}$$

We call the solutions in the regular branch (3) *nonsingular* if, in addition, the differential

$$(D_q\mathcal{R})(q(u);u) \in \mathcal{L}(\mathcal{X},\mathcal{Y}') \text{ is an isomorphism from } \mathcal{X} \text{ onto } \mathcal{Y}', \text{ for all } u \in \tilde{X}\ . \tag{4}$$

The following proposition collects well-known sufficient conditions for well-posedness of (2). For regular branches of nonsingular solutions given by (2) - (4), the differential $D_q\mathcal{R}$ satisfies the so-called *inf-sup conditions.*

Proposition 2.1. *Assume that $\mathcal{Y}$ is reflexive and that, for some nominal value $\langle u\rangle \in X$ of the uncertain input data, the operator equation* (2) *admits a* regular branch of solutions (3). *Then the differential $D_q\mathcal{R}$ at $(\langle u\rangle, q_0)$ given by the bilinear map*

$$\mathcal{X} \times \mathcal{Y} \ni (\varphi,\psi) \mapsto {}_{\mathcal{Y}'}\langle D_q\mathcal{R}(q_0;\langle u\rangle)\varphi,\psi\rangle_{\mathcal{Y}}$$

is boundedly invertible, uniformly with respect to $u \in \tilde{X}$ *where* $\tilde{X} \subset X$ *is an open neighborhood of the nominal instance* $\langle u \rangle \in X$ *of the uncertain parameter. In particular, there exists a constant* $\beta > 0$ *such that there holds*

$$\forall u \in \tilde{X}: \qquad \begin{array}{c} \inf\limits_{0\neq\varphi\in\mathcal{X}} \sup\limits_{0\neq\psi\in\mathcal{Y}} \dfrac{{}_{\mathcal{Y}'}\langle (D_q\mathcal{R})(q_0;u)\varphi,\psi\rangle_{\mathcal{Y}}}{\|\varphi\|_{\mathcal{X}}\|\psi\|_{\mathcal{Y}}} \geq \beta > 0\,, \\ \inf\limits_{0\neq\psi\in\mathcal{Y}} \sup\limits_{0\neq\varphi\in\mathcal{X}} \dfrac{{}_{\mathcal{Y}'}\langle (D_q\mathcal{R})(q_0;u)\varphi,\psi\rangle_{\mathcal{Y}}}{\|\varphi\|_{\mathcal{X}}\|\psi\|_{\mathcal{Y}}} \geq \beta > 0 \end{array} \tag{5}$$

and

$$\forall u \in \tilde{X}: \quad \|(D_q\mathcal{R})(q_0,u)\|_{\mathcal{L}(\mathcal{X},\mathcal{Y}')} = \sup_{0\neq\varphi\in\mathcal{X}} \sup_{0\neq\psi\in\mathcal{Y}} \frac{{}_{\mathcal{Y}'}\langle (D_q\mathcal{R})(q_0;u)\varphi,\psi\rangle_{\mathcal{Y}}}{\|\varphi\|_{\mathcal{X}}\|\psi\|_{\mathcal{Y}}} \leq \beta^{-1}\,. \tag{6}$$

Under conditions (5) and (6), for every $u \in \tilde{X} \subseteq X$, there exists a unique, regular solution $q(u)$ of (2) which is uniformly bounded with respect to $u \in \tilde{X}$ in the sense that there exists a constant $C(F,\tilde{X}) > 0$ such that

$$\sup_{u\in\tilde{X}} \|q(u)\|_{\mathcal{X}} \leq C(F,\tilde{X})\,. \tag{7}$$

For (5) - (7) being valid, we shall say that the set $\{(q(u),u) : u \in \tilde{X}\} \subset \mathcal{X}\times\tilde{X}$ forms a *regular branch of nonsingular solutions.*

If the data-to-solution map $\tilde{X} \ni u \mapsto q(u)$ is also Fréchet differentiable with respect to u at every point of the regular branch $\{(q(u);u) : u \in \tilde{X}\} \subset \mathcal{X}\times\tilde{X}$, the dependence of the "forward map", i.e. the mapping relating u to $q(u)$ with the branch of nonsingular solutions, is locally Lipschitz on $\tilde{X}$: there exists a Lipschitz constant $L(F,\tilde{X})$ such that

$$\forall u,v \in \tilde{X}: \quad \|q(u)-q(v)\|_{\mathcal{X}} \leq L(F,\tilde{X})\|u-v\|_X\,. \tag{8}$$

This follows from the identity $(D_uq)(u) = -(D_q\mathcal{R})^{-1}(D_u\mathcal{R})$, and from the isomorphism property $(D_u\mathcal{R}_q)(q_0;\langle u\rangle) \in \mathcal{L}_{iso}(\mathcal{X},\mathcal{Y}')$ which is implied by (5) and (6), and from the continuity of the differential $D_q\mathcal{R}$ on the regular branch.

In what follows, we will place ourselves in the abstract setting (1) with uniformly continuously differentiable mapping $\mathcal{R}(q;u)$ in a product of neighborhoods $B_X(\langle u\rangle;R)\times B_{\mathcal{X}}(q(\langle u\rangle);R) \subset X\times\mathcal{X}$ of sufficiently small radius $R > 0$, satisfying the structural assumption (2). In Proposition 2.1 and throughout what follows, $q(\langle u\rangle) \in \mathcal{X}$ denotes the unique regular solution of (2) at the nominal input $\langle u\rangle \in X$.

2.2. Uncertainty parametrization

We shall be concerned with the particular case where $u \in X$ is a random variable taking values in (a subset $\tilde{X}$ of) the Banach space X. We assume that X is separable, infinite-dimensional, and admits an unconditional Schauder basis $\boldsymbol{\Psi}$: $X = \overline{\text{span}}\{\psi_j : j \geq 1\}$. Then, every $u \in \tilde{X} \subset X$ can be parametrized in this basis, i.e.

$$u = u(\boldsymbol{y}) := \langle u\rangle + \sum_{j\geq 1} y_j\psi_j \quad \text{for some } \boldsymbol{y} = (y_j)_{j\geq 1} \in U\,. \tag{9}$$

Examples of representations (9) are Karhunen–Loève expansions (see, e.g., [43, 45, 22]) or by unconditional Schauder bases (see, e.g., [19]). We point out that the representation (9) is not unique: rescaling y_j and ψ_j will not change u. We may and will assume, therefore, throughout what follows that *the sequence* $\{\psi_j\}_{j\geq 1}$ *is such that* $U = [-1,1]^{\mathbb{N}}$. Norm-convergence of the series (9) in X is then implied by the *summability condition*

$$\sum_{j\geq 1} \|\psi_j\|_X < \infty\,, \tag{10}$$

assumed throughout in what follows.

We assume further that uncertain inputs u with "higher regularity" (when measured in a smoothness scale $\{X_t\}_{t\geq 0}$ with $X = X_0 \supset X_1 \supset X_2 \supset \ldots$ on the admissible input data) correspond to stronger decay of ψ_j: for $u \in X_t \subset X$, in (9) the $\{\psi_j\}_{j\geq 1}$ are assumed scaled such that

$$\boldsymbol{b} := \{\|\psi_j\|_X\}_{j\geq 1} \in \ell^p(\mathbb{N}) \ \text{ for some } \ 0 < p(t) < 1\,, \tag{11}$$

where the sequence $\boldsymbol{b} = (b_j)_{j\geq 1}$ given by $b_j := \|\psi_j\|_X$. We also introduce the subset

$$U = \{\boldsymbol{y} \in [-1,1]^{\mathbb{N}} : u(\boldsymbol{y}) := \langle u\rangle + \sum_{j\geq 1} y_j\psi_j \in \tilde{X}\}\,. \tag{12}$$

Once an unconditional Schauder basis $\{\psi_j\}_{j\geq 1}$ has been chosen, every realization $u \in X$ can be identified in a one-to-one fashion with the pair $(\langle u\rangle, \boldsymbol{y})$ where $\langle u\rangle$ denotes the *nominal instance* of the uncertain datum u and $\boldsymbol{y}$ is the coordinate vector in representation (9). Inserting (9) into (1), we obtain the *equivalent, countably-parametric form*: given $F : U \to \mathcal{Y}'$,

$$\text{find } q(\boldsymbol{y}; F) \in \mathcal{X} : \ \forall \boldsymbol{y} \in U : \ \ \mathcal{R}(q; \boldsymbol{y}) := A(q; \boldsymbol{y}) - F(\boldsymbol{y}) = 0 \quad \text{in} \quad \mathcal{Y}' \ . \tag{13}$$

Remark 2.1. *In what follows, by a slight abuse of notation,* we identify the subset U in (12) with the countable set of parameters from the infinite-dimensional parameter domain $U \subseteq \mathbb{R}^{\mathbb{N}}$ *without explicitly writing so. The operator $A(q; u)$ in* (2) *then becomes, via the parametric dependence $u = u(\boldsymbol{y})$, a parametric operator family $A(q; u(\boldsymbol{y}))$ which we denote (with slight abuse of notation) by $\{A(q; \boldsymbol{y}) : \boldsymbol{y} \in U\}$, with the parameter set $U = [-1, 1]^{\mathbb{N}}$ (again, we use in what follows this definition in place of the set U as defined in* (12)*). In the particular case that the parametric operator family $A(q; \boldsymbol{y})$ in* (2) *is linear, we have $A(q; \boldsymbol{y}) = A(\boldsymbol{y})q$ with $A(\boldsymbol{y}) \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$. We do not assume, however, that the maps $q \mapsto A(q; \boldsymbol{y})$ are linear in what follows, unless explicitly stated.*

With this understanding, and under the assumptions (7) and (8), the operator equation (2) will admit, for every $\boldsymbol{y} \in U$, a unique solution $q(\boldsymbol{y}; F)$ which is, due to (7) and (8), uniformly bounded and depends Lipschitz continuously on the parameter sequence $\boldsymbol{y} \in U$: there holds

$$\sup_{\boldsymbol{y}\in U} \|q(\boldsymbol{y}; F)\|_{\mathcal{X}} \leq C(F, U), \tag{14}$$

and, if the local Lipschitz condition (8) holds, there exists a Lipschitz constant $L > 0$ such that

$$\|q(\boldsymbol{y}; F) - q(\boldsymbol{y}'; F)\|_{\mathcal{X}} \leq L(F, U)\|u(\boldsymbol{y}) - u(\boldsymbol{y}')\|_X \ . \tag{15}$$

The Lipschitz constant $L > 0$ in (15) is not, in general, equal to $L(F, U)$ in (8): it depends on the nominal instance $\langle u\rangle \in X$ and on the choice of basis $\{\psi_j\}_{j\geq 1}$.

Unless explicitly stated otherwise, throughout what follows, we shall identify $q_0 = q(\mathbf{0}; F) \in \mathcal{X}$ in Proposition 2.1 with the solution of (1) at the nominal input $\langle u\rangle \in X$.

2.3. Bayesian Estimation

Following [45, 22, 44, 41, 42], we equip the space of uncertain inputs X and the space of solutions $\mathcal{X}$ of the forward maps with norms $\|\cdot\|_X$ and with $\|\cdot\|_{\mathcal{X}}$, respectively. We consider the abstract (possibly nonlinear) operator equation (2) where the uncertain operator $A(\cdot; u) \in C^1(\mathcal{X}, \mathcal{Y}')$ is assumed to be locally boundedly invertible, at least locally for the uncertain input u sufficiently close to a nominal input $\langle u\rangle \in X$, i.e. for $\|u - \langle u\rangle\|_X$ sufficiently small so that, for such u, the response of the forward problem (2) is uniquely defined. We define the *forward response map*, which maps a given uncertain input u and a given forcing F to the response q in (2) by

$$X \ni u \mapsto q(u) := G(u; F(u)) \ , \ \text{ where } \ G(u, F) : X \times \mathcal{Y}' \mapsto \mathcal{X} \ . \tag{16}$$

To ease notation, we do not list the dependence of the response on F and simply denote the dependence of the forward solution on the uncertain input as $q(u) = G(u)$. We assume given an observation functional $\mathcal{O}(\cdot) : \mathcal{X} \to Y$, which denotes a *bounded linear observation operator* on the space $\mathcal{X}$ of observed system responses in Y. Throughout the remainder of this paper, we assume that there is a finite number K of sensors, so that $Y = \mathbb{R}^K$ with $K < \infty$. Then $\mathcal{O} \in \mathcal{L}(\mathcal{X}; Y) \simeq (\mathcal{X}')^K$. We equip $Y = \mathbb{R}^K$ with the Euclidean norm, denoted by $|\cdot|$. For example, if $\mathcal{O}(\cdot)$ is a K-vector of observation functionals $\mathcal{O}(\cdot) = (o_k(\cdot))_{k=1}^K$.

In this setting, we wish to predict *computationally* an expected (under the Bayesian posterior) system response of the QoI, conditional on given, noisy measurement data δ. Specifically, we assume the data δ to consist of observations of system responses in the data space Y, corrupted by additive observation noise, e.g. by a realization of a random variable η taking values in Y with law $\mathbb{Q}_0$. We assume the following form of observed data, composed of the observed system response and the additive noise η

$$\delta = \mathcal{O}(G(u)) + \eta \ \in Y \ . \tag{17}$$

We assume that $Y = \mathbb{R}^K$ and that η in (17) is Gaussian, i.e. a random vector $\eta \sim \mathbb{Q}_0 \sim \mathcal{N}(0, \Gamma)$ with a positive definite covariance Γ on $Y = \mathbb{R}^K$ (i.e., a symmetric, positive definite covariance matrix $\Gamma \in \mathbb{R}^{K\times K}_{sym}$ which we assume to be known. The *uncertainty-to-observation map* of the system $\mathcal{G} : X \to Y = \mathbb{R}^K$ is $\mathcal{G} = \mathcal{O} \circ G$, so that

$$\delta = \mathcal{G}(u) + \eta = (\mathcal{O} \circ G)(u) + \eta \ \in Y \ ,$$

where $Y = L^2_\Gamma(\mathbb{R}^K)$ denotes random vectors taking values in $Y = \mathbb{R}^K$ which are square integrable with respect to the Gaussian measure on $Y = \mathbb{R}^K$. Bayes' formula [45, 22] yields a density of the Bayesian posterior with respect to the prior whose negative log-likelihood equals the observation noise covariance-weighted, least squares functional (also referred to as "potential" in what follows) $\Phi_\Gamma : X \times Y \to \mathbb{R}$ by $\Phi_\Gamma(u;\delta) = \frac{1}{2}|\delta - \mathcal{G}(u)|^2_\Gamma$, ie.

$$\Phi_\Gamma(u;\delta) = \frac{1}{2}|\delta - \mathcal{G}(u)|^2_\Gamma := \frac{1}{2}\left((\delta - \mathcal{G}(u))^\top \Gamma^{-1}(\delta - \mathcal{G}(u))\right) . \tag{18}$$

In [45, 22], an infinite-dimensional version of Bayes' rule was shown to hold in the present setting. In particular, the local Lipschitz assumption (8) on the solutions' dependence on the data implies a corresponding Lipschitz dependence of the Bayesian Potential (18) on $u \in X$. Specifically, there holds the following version of Bayes' theorem. Bayes' Theorem states that, under appropriate continuity conditions on the uncertainty-to-observation map $\mathcal{G} = (\mathcal{O} \circ G)(\cdot)$ and on the prior measure $\boldsymbol{\pi}_0$ on $u \in X$, *for positive observation noise covariance* Γ in (18), the posterior $\boldsymbol{\pi}^\delta$ of $u \in X$ given data $\delta \in Y$ is absolutely continuous with respect to the prior $\boldsymbol{\pi}_0$.

Theorem 2.2. *([22, Thm. 3.3]) Assume that the potential $\Phi_\Gamma : X \times Y \mapsto \mathbb{R}$ is, for given data $\delta \in Y$, $\boldsymbol{\pi}_0$ measurable on $(X, \mathcal{B}(X))$ and that, for $\mathbb{Q}_0$-a.e. data $\delta \in Y$ there holds*

$$Z := \int_X \exp\left(-\Phi(u;\delta)\right) \boldsymbol{\pi}_0(\mathrm{d}u) > 0 .$$

Then the conditional distribution of $u|\delta$ exists and is denoted by $\boldsymbol{\pi}^\delta$. It is absolutely continuous with respect to $\boldsymbol{\pi}_0$ and there holds

$$\frac{d\boldsymbol{\pi}^\delta}{d\boldsymbol{\pi}_0}(u) = \frac{1}{Z}\exp\left(-\Phi(u;\delta)\right) . \tag{19}$$

In particular, then, the Radon-Nikodym derivative of the Bayesian posterior w.r.t. the prior measure admits a bounded density w.r.t. the prior $\boldsymbol{\pi}_0$ which we denote by Θ, and which is given by (19).

2.4. Parametric Bayesian posterior

We parametrize the uncertain datum u in the forward equation (2) as in (9). Motivated by [41, 42], the basis for the presently proposed deterministic quadrature approaches for Bayesian estimation via the computational realization of Bayes' formula is a *parametric, deterministic representation* of the derivative of the posterior measure $\boldsymbol{\pi}^\delta$ with respect to the *uniform prior measure $\boldsymbol{\pi}_0$ on the set U of coordinates in the uncertainty parametrization* (12). The prior measure $\boldsymbol{\pi}_0$ being uniform, we admit in (9) sequences $\boldsymbol{y}$ which take values in the parameter domain $U = [-1,1]^{\mathbb{J}}$, with an index set $\mathbb{J} \subset \mathbb{N}$. We consider the countably-parametric, deterministic forward problem in the probability space

$$(U, \mathcal{B}, \boldsymbol{\pi}_0) . \tag{20}$$

We assume throughout what follows that the prior measure $\boldsymbol{\pi}_0$ on the uncertain input $u \in X$, parametrized in the form (9)*, is the uniform measure.* Being π_0 a countable product probability measure, this assumption implies the statistical independence of the coordinates y_j in the parametrization (9). With the parameter domain U as in (20) the parametric uncertainty-to-observation map $\Xi : U \to Y = \mathbb{R}^K$ is given by

$$\Xi(\boldsymbol{y}) = \mathcal{G}(u)\Big|_{u=\langle u\rangle + \sum_{j\in\mathbb{J}} y_j \psi_j} . \tag{21}$$

Our reduced basis approach is based on a parametric version of Bayes' Theorem 2.2, in terms of the uncertainty parametrization (9). To present it, we view U as the unit ball in $\ell^\infty(\mathbb{J})$, the Banach space of bounded sequences taking values in U.

Theorem 2.3. *Assume that $\Xi : \bar{U} \to Y = \mathbb{R}^K$ is bounded and continuous. Then $\boldsymbol{\pi}^\delta(d\boldsymbol{y})$, the distribution of $\boldsymbol{y} \in U$ given data $\delta \in Y$, is absolutely continuous with respect to $\boldsymbol{\pi}_0(d\boldsymbol{y})$, i.e. there exists a parametric density $\Theta(\boldsymbol{y})$ such that*

$$\frac{d\boldsymbol{\pi}^\delta}{d\boldsymbol{\pi}_0}(\boldsymbol{y}) = \frac{1}{Z}\Theta(\boldsymbol{y}) \tag{22}$$

with $\Theta(\boldsymbol{y})$ given by

$$\Theta(\boldsymbol{y}) = \exp\big(-\Phi_\Gamma(u;\delta)\big)\Big|_{u=\langle u\rangle + \sum_{j\in\mathbb{J}} y_j \psi_j}, \tag{23}$$

with Bayesian potential Φ_Γ *as in* (18) *and with normalization constant* Z *given by*

$$Z = \mathbb{E}^{\boldsymbol{\pi}_0}[\Theta] = \int_U \Theta(\boldsymbol{y})\, \mathrm{d}\boldsymbol{\pi}_0(\boldsymbol{y}) > 0\ . \tag{24}$$

Bayesian inversion is concerned with the approximation of a "most likely" *system response* $\phi : X \to \mathcal{Z}$ (sometimes also referred to as *Quantity of Interest (QoI)* which may take values in a Banach space $\mathcal{Z}$) of the QoI ϕ, conditional on given (noisy) observation data $\delta \in Y$. In particular the choice $\phi(u) = G(u)$ (with $\mathcal{Z} = \mathcal{X}$) facilitates computation of the "most likely" (as expectation under the posterior, given data δ) system response. With the QoI ϕ we associate the deterministic, infinite-dimensional, parametric map

$$\begin{aligned} \Psi(\boldsymbol{y}) &= \Theta(\boldsymbol{y})\phi(u)\,|_{u=\langle u\rangle+\sum_{j\in\mathbb{J}} y_j\psi_j} \\ &= \exp\big(-\Phi_\Gamma(u;\delta)\big)\phi(u)\Big|_{u=\langle u\rangle+\sum_{j\in\mathbb{J}} y_j\psi_j} : U \to \mathcal{Z}\ . \end{aligned} \tag{25}$$

Then the Bayesian estimate of the QoI ϕ, given noisy observation data δ, reads

$$\mathbb{E}^{\boldsymbol{\pi}^\delta}[\phi] = Z'/Z,\ \ Z' := \int_{\boldsymbol{y}\in U} \Psi(\boldsymbol{y})\, \boldsymbol{\pi}_0(\mathrm{d}\boldsymbol{y}),\ \ Z := \int_{\boldsymbol{y}\in U} \Theta(\boldsymbol{y})\, \boldsymbol{\pi}_0(\mathrm{d}\boldsymbol{y})\ . \tag{26}$$

The task in computational Bayesian estimation is therefore to approximate the ratio $Z'/Z \in \mathcal{Z}$ in (26). In the parametrization with respect to $\boldsymbol{y} \in U$, Z and Z' take the form of infinite-dimensional, iterated integrals with respect to the prior $\boldsymbol{\pi}_0(d\boldsymbol{y})$.

2.5. Well-posedness and approximation

For the computational viability of Bayesian inversion the quantity $\mathbb{E}^{\boldsymbol{\pi}^\delta}[\phi]$ should be stable under perturbations of the data δ and under changes in the forward problem stemming, for example, from discretizations as considered in Section 4.

Unlike deterministic inverse problems where the data-to-solution maps can be severely ill-posed, for $\Gamma > 0$ the expectations (26) are Lipschitz continuous with respect to the data δ, *provided that the potential* Φ_Γ *in* (18) *is locally Lipschitz with respect to the data* δ in the following sense.

Assumption 2. *Let* $\tilde{X} \subseteq X$ *and assume* $\Phi_\Gamma \in C(\tilde{X} \times Y; \mathbb{R})$ *is Lipschitz on bounded sets. Assume also that there exist functions* $M_i : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ *(depending on* $\Gamma > 0$*) which are monotone, non-decreasing separately in each argument, such that for all* $u \in \tilde{X}$*, and for all* $\delta, \delta_1, \delta_2 \in B_Y(0, r)$

$$\Phi(u;\delta) \geq -M_1(r, \|u\|_X), \tag{27}$$

and

$$|\Phi_\Gamma(u;\delta_1) - \Phi_\Gamma(u;\delta_2)| \leq M_2(r, \|u\|_X)\|\delta_1 - \delta_2\|_Y\ . \tag{28}$$

Under Assumption 2, the expectation (26) depends Lipschitz on δ (see [22, Sec. 4.1] for a proof):

$$\forall \phi \in L^2(\boldsymbol{\pi}^{\delta_1}, X; \mathbb{R}) \cap L^2(\boldsymbol{\pi}^{\delta_2}, X; \mathbb{R}) \quad \|\mathbb{E}^{\boldsymbol{\pi}^{\delta_1}}[\phi] - \mathbb{E}^{\boldsymbol{\pi}^{\delta_2}}[\phi]\|_{\mathcal{Z}} \leq C(\Gamma, r)\|\delta_1 - \delta_2\|_Y\ . \tag{29}$$

Below, we shall be interested in the impact of approximation errors in the forward response of the system (e.g. due to discretization and approximate numerical solution of system responses) on the Bayesian predictions (26). For continuity of the expectations (26) w.r.t. changes in the potential, we impose the following assumption.

Assumption 3. *Let* $\tilde{X} \subseteq X$ *and assume* $\Phi \in C(\tilde{X} \times Y; \mathbb{R})$ *is Lipschitz on bounded sets. Assume also that there exist functions* $M_i : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ *which are monotonically non-decreasing separately in each argument, such that for all* $u \in \tilde{X}$*, and all* $\delta \in B_Y(0, r)$*, Equation* (27) *is satisfied and*

$$|\Phi_\Gamma(u;\delta) - \Phi_\Gamma^N(u;\delta)| \leq M_2(r, \|u\|_X)\|\delta\|_Y\psi(N) \tag{30}$$

where $\psi(N) \to 0$ *as* $N \to \infty$.

By $\boldsymbol{\pi}_N^\delta$ we denote the Bayesian posterior, given data $\delta \in Y$, with respect to Φ_Γ^N.

Proposition 2.4. *Under Assumption 3, and the assumption that for $\tilde{X} \subseteq X$ and for some bounded $B \subset X$ we have $\boldsymbol{\pi}_0(\tilde{X} \cap B) > 0$ and*

$$X \ni u \mapsto \exp(M_1(\|u\|_X))(M_2(\|u\|_X))^2 \in L^1_{\boldsymbol{\pi}_0}(X;\mathbb{R}) \,,$$

there holds, for every QoI $\phi : X \to \mathcal{Z}$ such that $\phi \in L^2_{\boldsymbol{\pi}^\delta}(X;\mathcal{Z}) \cap L^2_{\boldsymbol{\pi}^\delta_N}(X;\mathcal{Z})$ uniformly w.r.t. N, that $Z > 0$ in (24) *and*

$$\|\mathbb{E}^{\boldsymbol{\pi}^\delta}[\phi] - \mathbb{E}^{\boldsymbol{\pi}^\delta_N}[\phi]\|_{\mathcal{Z}} \leq C(\Gamma, r)\|\delta\|_Y \psi(N) \,. \tag{31}$$

For a proof of Proposition 2.4, we refer to [22, Thm. 4.7, Rem. 4.8].

Below, we shall present concrete choices for the convergence rate function $\psi(N)$ in estimates (30), (31) in terms of i) *"dimension truncation"* of the uncertainty parametrization (9), i.e. to a finite number of $s \geq 1$ terms in (9), and ii) *Petrov-Galerkin Discretization* of the dimensionally truncated problem, iii) *generalized polynomial chaos (gpc) approximation* of the dimensionally truncated problem for particular classes of forward problems. The verification of the consistency condition (30) in either of these cases will be based on (cf. [24])

Proposition 2.5. *Assume we are given a sequence $\{q^N\}_{N\geq 1}$ of approximations to the parametric forward response $X \ni u \mapsto q(u) \in \mathcal{X}$ such that, with the parametrization* (9)*,*

$$\sup_{\boldsymbol{y}\in U} \|(q - q^N)(\boldsymbol{y})\|_{\mathcal{X}} \leq \psi(N) \tag{32}$$

with a consistency error bound $\psi \downarrow 0$ as $N \to \infty$ monotonically and uniformly w.r.t. $u \in \tilde{X}$ (resp. w.r.t. $\boldsymbol{y} \in U$). By G^N we denote the corresponding (Galerkin) approximations of the parametric forward maps. Then the approximate Bayesian potential

$$\Phi^N(u;\delta) = \frac{1}{2}(\delta - \mathcal{G}^N(u))^\top \Gamma^{-1}(\delta - \mathcal{G}^N(u)) : X \times Y \mapsto \mathbb{R} \,, \tag{33}$$

where $\mathcal{G}^N := \mathcal{O} \circ G^N$, satisfies (30)*.*

2.6. Forward and Posterior Sparsity

A central role in convergence rate estimates of the approximations presented in the following sections are sparsity results on the parametric forward solution manifold $\{q(\boldsymbol{y}) : \boldsymbol{y} \in U\} \subset \mathcal{X}$ in (16) as well as on the corresponding Bayesian posterior manifold $\{\Theta(\boldsymbol{y}) : \boldsymbol{y} \in U\}$. As was shown in [18, 44, 40, 41, 42], the sparsity results take the form of summability conditions on coefficient sequences in gpc expansions of the parametric responses $q(\boldsymbol{y})$ and of the parametric posterior densities $\Theta(\boldsymbol{y})$ and $\Psi(\boldsymbol{y})$ in (23) and in (25), respectively. These results imply convergence rates of N-term gpc approximations of these countably-parametric quantities which are independent of number of active parameters in the approximations. The results obtained in [18, 40, 41, 42] require *holomorphic, parametric dependence* of the parametric forward solution map $U \ni \boldsymbol{y} \mapsto q(\boldsymbol{y})$ which we review next. *Throughout the rest of this section, all Banach spaces are understood as spaces over the complex coefficient field*, without notationally indicating this. *Holomorphic parameter dependence* with respect to the parameters $z_j = y_j + iw_j \in \mathbb{C}$ is made precise, with a view towards (Legendre series) polynomial approximation: we introduce for $s > 1$ the Bernstein ellipse in the complex plane

$$\mathcal{E}_s := \left\{ \frac{w + w^{-1}}{2} : w \in \mathbb{C} \,,\ \ 1 \leq |w| \leq s \right\}, \tag{34}$$

which has semi axes of length $(s + s^{-1})/2$ and $(s - s^{-1})/2$. For sequences $\boldsymbol{\rho} := (\rho_j)_{j\geq 1}$ with $\rho_j > 1$, with (34) we introduce tensorized poly-ellipses

$$\mathcal{E}_{\boldsymbol{\rho}} := \bigotimes_{j\geq 1} \mathcal{E}_{\rho_j} \,. \tag{35}$$

Definition 2.1. *For $\varepsilon > 0$ and $0 < p < 1$ and for a positive sequence $\boldsymbol{b} = (b_j)_{j\geq 1} \in \ell^p(\mathbb{N})$, the parametric residual map $\mathcal{R}$ in* (13) *satisfies the* $(\boldsymbol{b}, p, \varepsilon)$-holomorphy assumption *if and only if*

1. *For each $\boldsymbol{y} \in U$, there exists a unique solution $q(\boldsymbol{y}) \in \mathcal{X}$ of the problem* (1)*. The parametric solution map $\boldsymbol{y} \mapsto q(\boldsymbol{y})$ from U to $\mathcal{X}$ is uniformly bounded, i.e.*

$$\sup_{\boldsymbol{y}\in U} \|q(\boldsymbol{y})\|_{\mathcal{X}} \leq B_0 \,, \tag{36}$$

for some finite constant $B_0 > 0$.

2. *There exists a constant* $B_\varepsilon > 0$ *such that for any sequence* $\boldsymbol{\rho} := (\rho_j)_{j\geq 1}$ *of poly-radii* $\rho_j > 1$ *that satisfies*

$$\sum_{j=1}^{\infty}(\rho_j - 1)b_j \leq \varepsilon, \tag{37}$$

the map q *admits a complex extension* $\boldsymbol{z} \mapsto q(\boldsymbol{z})$ *that is holomorphic with respect to each coordinate* z_j *of the parameter sequence* $\boldsymbol{z} = (z_j)_{j\geq 1}$ *in a cylindrical set* $\mathcal{O}_{\boldsymbol{\rho}} := \bigotimes_{j\geq 1}\mathcal{O}_{\rho_j}$. *Here,* $\mathcal{O}_{\rho_j} \subset \mathbb{C}$ *is an open set containing the Bernstein ellipse* $\mathcal{E}_{\rho_j}$. *This extension is bounded on* $\mathcal{E}_{\boldsymbol{\rho}} := \bigotimes_{j\geq 1}\mathcal{E}_{\rho_j}$, *according to*

$$\sup_{\boldsymbol{z}\in\mathcal{E}_{\boldsymbol{\rho}}} \|q(\boldsymbol{z})\|_{\mathcal{X}} \leq B_\varepsilon \ . \tag{38}$$

The relevance of $(\boldsymbol{b}, p, \varepsilon)$-holomorphy of countably-parametric maps such as $\boldsymbol{z} \to q(\boldsymbol{z})$ is clear from the following approximation result, which is, for example, [18, Thm. 2.2]. To state it, we denote by $\Lambda = \{\boldsymbol{\nu} \in \mathbb{N}_0^{\mathbb{N}} : |\boldsymbol{\nu}| < \infty\}$ the countable set of sequences of finitely supported multi-indices.

Theorem 2.6. *Assume that* $\mathcal{X}$ *is a Hilbert space and that the map* $U \ni \boldsymbol{y} \mapsto q(\boldsymbol{y}) \in \mathcal{X}$ *is* $(\boldsymbol{b}, p, \varepsilon)$*-holomorphic. Then the parametric solution map admits the (unconditionally convergent) Legendre series expansion*

$$\forall \boldsymbol{y} \in U: \quad q(\boldsymbol{y}) = \sum_{\boldsymbol{\nu}\in\Lambda} q_{\boldsymbol{\nu}} L_{\boldsymbol{\nu}}(\boldsymbol{y}) \ , \tag{39}$$

where, for $\boldsymbol{\nu} \in \Lambda$, *we defined* $L_{\boldsymbol{\nu}}(\boldsymbol{y}) := \prod_{j\geq} L_{\nu_j}(\boldsymbol{y})$ *with* L_n *the Legendre polynomial of degree* n *on* $(-1, 1)$ *with normalization* $\|L_n\|_{L^\infty(-1,1)} = 1$, *and where the Legendre coefficients* $q_{\boldsymbol{\nu}} = (q, L_{\boldsymbol{\nu}}) \in \mathcal{X}$ *with* $(\cdot,\cdot)$ *denoting the* $L^2(U;\mu)$*-innerproduct with respect to the uniform probability measure* $\mu = \bigotimes_{j\geq 1} dy_j/2$ *on* U.

The sequence $\boldsymbol{q} := (\|q_{\boldsymbol{\nu}}\|_{\mathcal{X}})_{\boldsymbol{\nu}\in\Lambda} \in \ell^p_m(\Lambda)$, *which denotes the set of sequences indexed by* Λ *whose decreasing rearrangement is* p*-summable.*

Moreover, there exists a nested sequence $\{\Lambda_M\}_{M\in\mathbb{N}}$ *of index sets* $\Lambda_M \subset \Lambda$ *of cardinality at most* M *and an associated sequence* $\{\Gamma_M\}_{M\in\mathbb{N}}$ *of sparse grids which are unisolvent for* $\mathbb{P}_{\Lambda_M} = \operatorname{span}\{\boldsymbol{y}^{\boldsymbol{\nu}} : \boldsymbol{\nu} \in \Lambda_M\}$ *such that the corresponding sparse grid interpolation operators* $\mathcal{S}_{\Lambda_M}$ *satisfy*

$$\|q - \mathcal{S}_{\Lambda_M} q\|_{L^\infty(U;\mathcal{X})} \leq C_i M^{-s} \ , \quad s = 1/p - 1 \ . \tag{40}$$

Moreover, the sparse grid integration satisfies

$$||\mathbb{E}^{\boldsymbol{\pi}_0}[q] - \mathbb{E}^{\boldsymbol{\pi}_0}[\mathcal{S}_{\Lambda_M} q]||_{\mathcal{X}} \leq C_e M^{-s}, \quad s = 1/p - 1 \ . \tag{41}$$

Here, the constant $C_i > 0$ *and* C_e *are independent of* Λ_M *and of* Γ_M.

For a proof of Theorem 2.6, we refer to [18, Thms. 2.2, 4.1] for interpolation and to [41, Thm. 4.7] for integration.

By Theorem 2.6, $(\boldsymbol{b}, p, \varepsilon)$-holomorphic, countably-parametric mappings afford dimension-independent polynomial approximation order $s = 1/p - 1$. It is therefore of interest to identify problems whose parametric solution maps are $(\boldsymbol{b}, p, \varepsilon)$-holomorphic.

Theorem 2.7. *Assume (i) the uncertain input* $u \in X$ *satisfies* (9) *with basis fulfilling* (11), *(ii) problem* (1) *is well-posed for all* u *in the sense of Assumption 1, (iii) the isomorphism property* (4) *holds, (iv) the map* $(q, u) \mapsto \mathcal{R}(q; u)$ *is complex continuously differentiable from* $\mathcal{X} \times X$ *into* $\mathcal{Y}'$. *Then,*

1. *the parametric solution family* $\{q(\boldsymbol{y}) : \boldsymbol{y} \in U\}$ *of the nonlinear operator equation* (1) *with residual map* $\mathcal{R}$ *in* (13) *admits an extension to the complex domain which is* $(\boldsymbol{b}, p, \varepsilon)$*-holomorphic with the same sequence* $\boldsymbol{b}$ *and the same* p *defined in* (10), *see [18, Thm 4.3];*
2. *the parametric Bayesian posterior density* $\Theta(\boldsymbol{y})$ *in* (23) *and the density* $\Psi(\boldsymbol{y})$ *in* (25) *are likewise* $(\boldsymbol{b}, p, \varepsilon)$*-holomorphic, with the same sequence* $\boldsymbol{b}$ *and the same* p. *The constants* B_ε *in* (38) *for* Θ *and* Ψ *depend on the observation noise covariance* $\Gamma > 0$ *in* (18) *as* $B_\varepsilon(\Gamma) \leq \bar{B}_\varepsilon \exp(c/\Gamma)$ *for some* $\bar{B}_\varepsilon > 0$ *and* $c > 0$ *which are independent of* Γ, *see [44, 41, 42].*

Theorem 2.7 provides with Theorem 2.6 the basis for sparse grid approximations in the parametric space with dimension-independent convergence rate $s = 1/p - 1$ which is only limited by the sparsity p of the basis expansion (9), but is independent of the dimension of the parameter space resp. of the dimension of the interpolation space.

3. Adaptive Sparse Grid Algorithms

Theorem 2.6 in the last section guarantees the existence of sparse generalized polynomial approximations of the forward solution map and of the posterior density which approximate these quantities with dimension-independent convergence rate. We exploit this sparsity in two ways: first, in the choice of sparse parameter samples during the offline-training phase of model order reductions, and, as already proposed in [41, 42], for adaptive, Smolyak-based numerical integration for the evaluation of the Bayesian estimate. Both are based on constructive algorithms for the computation of such sparse polynomial approximations. To this end, we follow [16] and introduce adaptive univariate interpolation and integration, and then present the corresponding adaptive sparse grid approximation.

3.1. Adaptive Univariate Approximation

In the univariate case $U = [-1, 1]$, given a set of interpolation nodes $-1 \leq y_1 < \cdots < y_m \leq 1$, the interpolation operator $\mathcal{I} : C(U; \mathcal{Z}) \to \mathcal{P}_{m-1}(U) \otimes \mathcal{Z}$ reads

$$\mathcal{I}g(y) = \sum_{k=1}^{m} g(y_k) l_k(y), \tag{42}$$

where the function $g \in C(U; \mathcal{Z})$, representing e.g. the parametric forward solution map q with $\mathcal{Z} = \mathcal{X}$ or the posterior density Θ with $\mathcal{Z} = \mathbb{R}$; $l_k(y)$, $1 \leq k \leq m$, are the associated Lagrange polynomials in $\mathcal{P}_{m-1}(U)$, the space of polynomials of degree at most $m - 1$. To define the sparse collocation, as usual the interpolation operator defined in (42) is recast as telescopic sum, ie.,

$$\mathcal{I}_L g(y) = \sum_{l=1}^{L} \triangle^l g(y) \ , \tag{43}$$

where L represents the level of interpolation grid; $\triangle^l := \mathcal{I}_l - \mathcal{I}_{l-1}$ with $\mathcal{I}_0 g \equiv 0$. Let Ξ^l denote the set of all interpolation nodes in the grid of level l, such that the grid is nested, i.e. $\Xi^l \subset \Xi^{l+1}$, $l = 0, \ldots, L-1$, with $\Xi^0 = \emptyset$ and $\Xi^L = \{y_1, \ldots, y_m\}$. As $\mathcal{I}_{l-1} g(y) = g(y)$ for any $y \in \Xi^{l-1}$, we have $\mathcal{I}_{l-1} = \mathcal{I}_l \circ \mathcal{I}_{l-1}$ and, with the notation $\Xi^l_\triangle = \Xi^l \setminus \Xi^{l-1}$, the interpolation operator (43) can be written in the form

$$\mathcal{I}_L g(y) = \sum_{l=1}^{L} \sum_{y_k^l \in \Xi^l_\triangle} (\mathcal{I}_l - \mathcal{I}_l \circ \mathcal{I}_{l-1}) g(y) = \sum_{l=1}^{L} \sum_{y_k^l \in \Xi^l_\triangle} \underbrace{\left(g(y_k^l) - \mathcal{I}_{l-1} g(y_k^l)\right)}_{s_k^l} l_k^l(y) \ , \tag{44}$$

where s_k^l represents the interpolation error of $\mathcal{I}_{l-l} g$ evaluated at the node $y_k^l \in \Xi^l_\triangle$, $k = 1, \ldots, |\Xi^l_\triangle|$, so that we can use it as a posteriori error estimator for adaptive construction of the interpolation (44). More precisely, we start from the root level $L = 1$ with the root interpolation node $y = 0$; then, whenever the interpolation error estimator

$$\mathcal{E}_i := \max_{y_k^L \in \Xi^L_\triangle} |s_k^L| \tag{45}$$

is larger than a given tolerance, we refine the interpolation to level $L + 1$ by adding a new interpolation node. One possible choice is adding a Leja node

$$y_1^{L+1} = \operatorname*{argmax}_{y \in U} \prod_{l=1}^{L} |y - y^l| \ , \tag{46}$$

or Clenshaw–Curtis (more than one) nodes

$$y_k^{L+1} = \cos\left(\frac{k}{2^{L-1}}\pi\right), \quad k = 0, 1 \text{ for } L = 1; k = 1, 3, \ldots, 2^{L-1} - 1 \text{ for } L \geq 2 \ . \tag{47}$$

Based on the adaptive interpolation, an associated quadrature formula is given by

$$\mathbb{E}[g] \approx \mathbb{E}[\mathcal{I}_L g] = \sum_{l=1}^{L} \sum_{y_k^l \in \Xi^l_\triangle} s_k^l w_k^l, \quad \text{being } w_k^l = \mathbb{E}[l_k^l] \ , \tag{48}$$

for which the integration error estimator can be taken as

$$\mathcal{E}_e := \left| \sum_{y_k^L \in \Xi_\triangle^L} s_k^L w_k^L \right| . \tag{49}$$

3.2. Adaptive Sparse Grid Approximation

In multiple dimensions $\boldsymbol{y} \in U = [-1,1]^J$, an adaptive sparse grid (SG) interpolation is obtained by tensorization of the univariate interpolation formula (43)

$$\mathcal{S}_{\Lambda_M} g(\boldsymbol{y}) = \sum_{\boldsymbol{\nu} \in \Lambda_M} (\triangle_1^{\nu_1} \otimes \cdots \otimes \triangle_J^{\nu_J})\, g(\boldsymbol{y}) \,, \tag{50}$$

where Λ_M is the downward closed index set defined in Theorem 2.6. As $\Lambda_1 \subset \cdots \subset \Lambda_M$ and as the interpolation nodes are nested, the SG formula (50) can be rewritten as

$$\mathcal{S}_{\Lambda_M} g(\boldsymbol{y}) = \sum_{m=1}^{M} \sum_{\boldsymbol{y}_{\boldsymbol{k}}^{\boldsymbol{\nu}^m} \in \Xi_\triangle^{\boldsymbol{\nu}^m}} \underbrace{\left(g(\boldsymbol{y}_{\boldsymbol{k}}^{\boldsymbol{\nu}^m}) - \mathcal{S}_{\Lambda_{m-1}} g(\boldsymbol{y}_{\boldsymbol{k}}^{\boldsymbol{\nu}^m}) \right)}_{\mathbf{s}_{\boldsymbol{k}}^{\boldsymbol{\nu}^m}} \boldsymbol{l}_{\boldsymbol{k}}^{\boldsymbol{\nu}^m}(\boldsymbol{y}) \,, \tag{51}$$

where $\Xi_\triangle^{\boldsymbol{\nu}^m}$ is the set of added nodes corresponding to the index $\boldsymbol{\nu}^m = (\nu_1^m, \ldots, \nu_J^m) = \Lambda_m \setminus \Lambda_{m-1}$; $\boldsymbol{l}_{\boldsymbol{k}}^{\boldsymbol{\nu}^m}(\boldsymbol{y}) = l_{k_1}^{\nu_1}(y_1) \otimes \cdots \otimes l_{k_J}^{\nu_J}(y_J)$, is the multidimensional Lagrange polynomial; $\mathbf{s}_{\boldsymbol{k}}^{\boldsymbol{\nu}^m}$ denotes the interpolation error of $\mathcal{S}_{\Lambda_{m-1}} g$ evaluated at $\boldsymbol{y}_{\boldsymbol{k}}^{\boldsymbol{\nu}^m}$, which can be used as an interpolation error estimator for the construction of the SG.

More explicitly, we start from the initial index $\boldsymbol{\nu} = \mathbf{1} = (1, \ldots, 1)$, thus $\Lambda_1 = \{\mathbf{1}\}$, with root node $\boldsymbol{y} = \mathbf{0} = (0, \ldots, 0)$. We then look for the *active index set* Λ_M^a such that $\Lambda_M \cup \{\boldsymbol{\nu}\}$ remains downward closed for any $\boldsymbol{\nu} \in \Lambda_M^a$. E.g. for $\Lambda_M = \{\mathbf{1}\}$ when $M = 1$, we have $\Lambda_M^a = \{\mathbf{1} + \mathbf{e}_j, j = 1, \ldots, J\}$, being $\mathbf{e}_j = (0, \ldots, j, \ldots, 0)$ whose j-th entry equals one while all other entries are zero. For each $\boldsymbol{\nu} \in \Lambda_M^a$, we evaluate the errors of the interpolation $\mathcal{S}_{\Lambda_M} g$ at the nodes $\Xi_\triangle^{\boldsymbol{\nu}}$, and enrich the index set $\Lambda_{M+1} = \Lambda_M \cup \{\boldsymbol{\nu}^{M+1}\}$ with the new index

$$\boldsymbol{\nu}^{M+1} := \operatorname*{argmax}_{\boldsymbol{\nu} \in \Lambda_M^a} \max_{\boldsymbol{y}_{\boldsymbol{k}}^{\boldsymbol{\nu}} \in \Xi_\triangle^{\boldsymbol{\nu}}} \frac{1}{|\Xi_\triangle^{\boldsymbol{\nu}}|} |\mathbf{s}_{\boldsymbol{k}}^{\boldsymbol{\nu}}| \,, \tag{52}$$

where the error is balanced by the work measured in terms of the number of new nodes $|\Xi_\triangle^{\boldsymbol{\nu}}|$. An adaptive sparse grid quadrature can be constructed similar to (50) as

$$\mathbb{E}[g] \approx \mathbb{E}[\mathcal{S}_{\Lambda_M} g] = \sum_{m=1}^{M} \sum_{\boldsymbol{y}_{\boldsymbol{k}}^{\boldsymbol{\nu}^m} \in \Xi_\triangle^{\boldsymbol{\nu}^m}} \mathbf{s}_{\boldsymbol{k}}^{\boldsymbol{\nu}^m} \mathbf{w}_{\boldsymbol{k}}^{\boldsymbol{\nu}^m}, \quad \text{where} \mathbf{w}_{\boldsymbol{k}}^{\boldsymbol{\nu}^m} = \mathbb{E}[\boldsymbol{l}_{\boldsymbol{k}}^{\boldsymbol{\nu}^m}] \,, \tag{53}$$

for which can enrich the index set with the new index

$$\boldsymbol{\nu}^{M+1} := \operatorname*{argmax}_{\boldsymbol{\nu} \in \Lambda_M^a} \frac{1}{|\Xi_\triangle^{\boldsymbol{\nu}}|} \left| \sum_{\boldsymbol{y}_{\boldsymbol{k}}^{\boldsymbol{\nu}} \in \Xi_\triangle^{\boldsymbol{\nu}}} \mathbf{s}_{\boldsymbol{k}}^{\boldsymbol{\nu}} \mathbf{w}_{\boldsymbol{k}}^{\boldsymbol{\nu}} \right| . \tag{54}$$

To terminate the SG algorithm for either interpolation or quadrature, we monitor the following heuristic error estimators compared to some prescribed tolerances, respectively:

$$\mathcal{E}_i := \max_{\boldsymbol{\nu} \in \Lambda_M^a} \max_{\boldsymbol{y}_{\boldsymbol{k}}^{\boldsymbol{\nu}} \in \Xi_\triangle^{\boldsymbol{\nu}}} |\mathbf{s}_{\boldsymbol{k}}^{\boldsymbol{\nu}}| \quad \text{and} \quad \mathcal{E}_e := \left| \sum_{\boldsymbol{\nu} \in \Lambda_M^a} \sum_{\boldsymbol{y}_{\boldsymbol{k}}^{\boldsymbol{\nu}} \in \Xi_\triangle^{\boldsymbol{\nu}}} \mathbf{s}_{\boldsymbol{k}}^{\boldsymbol{\nu}} \mathbf{w}_{\boldsymbol{k}}^{\boldsymbol{\nu}} \right| . \tag{55}$$

Remark 3.1. *The interpolation and integration error estimators* (55) *are not rigorous compared to the true interpolation and integration errors. In fact, $\mathcal{E}_i$ may underestimate the true worst-case scenario interpolation error as it is only the maximum interpolation error over a finite number of SG nodes instead of the whole parameter domain: $\mathcal{E}_e$ only measures the contribution of integration forward neighboring SG nodes.*

4. High-Fidelity Petrov-Galerkin Approximation

For the high-fidelity ("HiFi" for short) numerical solution of the parametric, nonlinear problem (1) at any given $\boldsymbol{y} \in U$, we consider the Petrov-Galerkin (PG) discretization in the one-parameter family of pairs of *subspaces* $\mathcal{X}_h \subset \mathcal{X}$ and $\mathcal{Y}_h \subset \mathcal{Y}$ where h represents a discretization parameter, for instance the meshwidth of a PG Finite Element discretization (when $\mathcal{X}_h = \mathcal{Y}_h$, the PG discretization becomes the classical Galerkin discretization). We assume that the dimensions of the two subspaces are equal, i.e. $N_h = \dim(\mathcal{X}_h) = \dim(\mathcal{Y}_h) < \infty$, $\forall h > 0$. To ensure the convergence of the HiFi PG solution $q_h \in \mathcal{X}_h$ to the exact solution $q \in \mathcal{X}$ as $h \to 0$, we assume the subspace families $\mathcal{X}_h$ and $\mathcal{Y}_h$ to be dense in $\mathcal{X}$ and $\mathcal{Y}$ as the discretization parameter (being, for example, a meshwidth or an inverse spectral order) $h \to 0$, i.e.

$$\forall w \in \mathcal{X}: \quad \lim_{h\to 0} \inf_{w_h \in \mathcal{X}_h} ||w - w_h||_{\mathcal{X}} = 0, \quad \text{and} \quad \forall v \in \mathcal{Y}: \quad \lim_{h\to 0} \inf_{v_h \in \mathcal{Y}_h} ||v - v_h||_{\mathcal{Y}} = 0 \,. \tag{56}$$

Moreover, to quantify the convergence rate of the discrete approximation, we introduce *scales of smoothness spaces* $\mathcal{X}^s \subset \mathcal{X} = \mathcal{X}^0$ and $\mathcal{Y}^s \subset \mathcal{Y} = \mathcal{Y}^0$ indexed by the smoothness parameter $s > 0$. Here, we have in mind for example spaces of functions with s extra derivatives in Sobolev or Besov spaces. Then, for appropriate choices of the subspaces $\mathcal{X}_h$ and $\mathcal{Y}_h$ hold the approximation properties: there exist constants $C_s > 0$ such that for all $0 < h \leq 1$ holds

$$\forall w \in \mathcal{X}^s: \inf_{w_h \in \mathcal{X}_h} ||w - w_h||_{\mathcal{X}} \leq C_s h^s ||w||_{\mathcal{X}^s} \quad \text{and} \quad \forall v \in \mathcal{Y}^s: \inf_{v_h \in \mathcal{Y}_h} ||v - v_h||_{\mathcal{Y}} \leq C_s h^s ||v||_{\mathcal{Y}^s} \,. \tag{57}$$

Here, the constant C_s is assumed independent of the discretization parameter h but may depend on the smoothness parameter s. For small values of h and/or if s is large, the PG discretization produces high-fidelity (HiFi) approximations $q_h \in \mathcal{X}_h$ of the true solution $q \in \mathcal{X}$ by solving

$$\text{given } \boldsymbol{y} \in U, \;\; \text{find } q_h(\boldsymbol{y}) \in \mathcal{X}_h: \quad {}_{\mathcal{Y}'}\langle \mathcal{R}(q_h(\boldsymbol{y}); \boldsymbol{y}), v_h \rangle_{\mathcal{Y}} = 0 \quad \forall v_h \in \mathcal{Y}_h \,. \tag{58}$$

To solve the nonlinear, parametric HiFi-PG approximation problem (58) numerically, we use a Newton iteration based on the parametric tangent operator of the nonlinear residual: for any given $\boldsymbol{y} \in U$, choose the initial guess of the solution $q_h^{(1)}(\boldsymbol{y}) \in \mathcal{X}_h$. Then, for $k = 1, 2, \ldots$, we search $\delta q_h^{(k)}(\boldsymbol{y}) \in \mathcal{X}_h$ such that

$${}_{\mathcal{Y}'}\langle D_q \mathcal{R}(q_h^{(k)}(\boldsymbol{y}); \boldsymbol{y})(\delta q_h^{(k)}(\boldsymbol{y})), v_h \rangle_{\mathcal{Y}} = -{}_{\mathcal{Y}'}\langle \mathcal{R}(q_h^{(k)}(\boldsymbol{y}); \boldsymbol{y}), v_h \rangle_{\mathcal{Y}} \quad \forall v_h \in \mathcal{Y}_h \,, \tag{59}$$

and update the solution by

$$q_h^{(k+1)}(\boldsymbol{y}) = q_h^{(k)}(\boldsymbol{y}) + \delta q_h^{(k)}(\boldsymbol{y}) \,. \tag{60}$$

We terminate the Newton Iteration with $q_h(\boldsymbol{y}) = q_h^{(k+1)}(\boldsymbol{y})$ once the following stopping-criterion is satisfied:

$$||\delta q_h^{(k)}(\boldsymbol{y})||_{\mathcal{X}} \leq \varepsilon_{tol} \quad \text{or} \quad ||\mathcal{R}(q_h^{(k)}(\boldsymbol{y}); \boldsymbol{y})||_{\mathcal{Y}'} \leq \varepsilon_{tol} \,. \tag{61}$$

Remark 4.1. *The Newton iteration* (59), (60) *converges only locally, and globalization techniques in* (60) *are required: for example, a rescaled increment* $\alpha^{(k)} \delta q_h^{(k)}(\boldsymbol{y})$*, where the step length* $\alpha^{(k)}$ *is determined by a line search and Wolfe or Goldstein conditions. We refer to [46, Sec. 4, 5] for details (see also [23]).*

In the HiFi-PG discretization, we denote the bases for $\mathcal{X}_h$ and $\mathcal{Y}_h$, respectively, by $\{w_h^n\}_{n=1}^{N_h}$ and $\{v_h^n\}_{n=1}^{N_h}$. In these bases, the PG approximation $q_h^{(k)}(\boldsymbol{y})$ and $\delta q_h^{(k)}(\boldsymbol{y})$ takes the form

$$q_h^{(k)}(\boldsymbol{y}) = \sum_{n=1}^{N_h} q_{h,n}^{(k)}(\boldsymbol{y}) w_h^n \quad \text{and} \quad \delta q_h^{(k)}(\boldsymbol{y}) = \sum_{n=1}^{N_h} \delta q_{h,n}^{(k)}(\boldsymbol{y}) w_h^n \,, \tag{62}$$

and the algebraic formulation of the parametric HiFi-PG problem (59) reads: find $\boldsymbol{\delta q}_h^{(k)}(\boldsymbol{y}) \in \mathbb{R}^{N_h}$ such that

$$J_R^{(k)}(\boldsymbol{y}) \boldsymbol{\delta q}_h^{(k)}(\boldsymbol{y}) = \boldsymbol{r}_h^{(k)}(\boldsymbol{y}) \,, \tag{63}$$

where $\boldsymbol{\delta q}_h^{(k)}(\boldsymbol{y}) = (q_{h,1}^{(k)}(\boldsymbol{y}), \ldots, q_{h,N_h}^{(k)}(\boldsymbol{y}))^\top$ and where the tangent stiffness matrix $J_R^{(k)}(\boldsymbol{y}) \in \mathbb{R}^{N_h \times N_h}$ is given by

$$\left(J_R^{(k)}(\boldsymbol{y})\right)_{nn'} = {}_{\mathcal{Y}'}\left\langle D_q \mathcal{R}(q_h^{(k)}(\boldsymbol{y}); \boldsymbol{y})(w_h^{n'}), v_h^n \right\rangle_{\mathcal{Y}}, \quad n, n' = 1, \ldots, N_h \,, \tag{64}$$

and the load vector $\boldsymbol{r}_h^{(k)}(\boldsymbol{y}) \in \mathbb{R}^{N_h}$ takes the form

$$\left(\boldsymbol{r}_h^{(k)}(\boldsymbol{y})\right)_n = -{}_{\mathcal{Y}'}\left\langle \mathcal{R}\left(q_h^{(k)}(\boldsymbol{y});\boldsymbol{y}\right), v_h^n\right\rangle_{\mathcal{Y}}, \quad n = 1,\dots,N_h \ . \tag{65}$$

With the notation $\boldsymbol{O}_h = (\mathcal{O}_h(w_h^1),\dots,\mathcal{O}_h(w_h^{N_h}))^\top$ and $\boldsymbol{q}_h(\boldsymbol{y}) = (q_{h,1}(\boldsymbol{y}),\dots,q_{h,N_h}(\boldsymbol{y}))^\top$, we have

$$\Theta_h(\boldsymbol{y}) = \exp\left(-\left(\delta - \boldsymbol{O}_h^\top \boldsymbol{q}_h(\boldsymbol{y})\right)^\top \Gamma^{-1}\left(\delta - \boldsymbol{O}_h^\top \boldsymbol{q}_h(\boldsymbol{y})\right)\right) \ . \tag{66}$$

When the number N_h of HiFi degrees of freedom is large, the system (63) may become too costly to solve. This renders the HiFi-PG solution $q_h(\boldsymbol{y})$ at a large number of parameter values $\boldsymbol{y} \in U$ computationally infeasible.

5. Reduced Basis Compression

In order to reduce the computational cost for the solution of the parametric HiFi-PG problem (58), we propose a *projection-based model order reduction*. Precisely, we project the parametric solution on a few problem-adapted, reduced bases rather than onto a large number N_h of HiFi (finite element) basis functions. In doing this, the reduction of computational complexity critically depends on *affine separability of the uncertain parameters from the physical variables [28, 7, 6]*. To construct an affine representation/approximation of the HiFi-PG problem (58), we rely on the *empirical interpolation method* applied to the HiFi-PG discretization.

5.1. Reduced Basis Compression

Instead of projecting the solution into the HiFi space $\mathcal{X}_h \subset \mathcal{X}$, we project it into a RB space $\mathcal{X}_N \subset \mathcal{X}_h$ of reduced dimension $N = \dim(\mathcal{X}_N) \leq N_h$. We also choose the RB test space such that $\mathcal{Y}_N \subset \mathcal{Y}_h$ with $\dim(\mathcal{Y}_N) = N$, which will be constructed to guarantee the stability of the RB-PG compression problem

$$\text{given } \boldsymbol{y} \in U, \quad \text{find } q_N(\boldsymbol{y}) \in \mathcal{X}_N : \quad {}_{\mathcal{Y}'}\langle \mathcal{R}(q_N(\boldsymbol{y});\boldsymbol{y}), v_N\rangle_{\mathcal{Y}} = 0 \quad \forall v_N \in \mathcal{Y}_N \ . \tag{67}$$

The RB-PG problem (67) can be solved by a Newton iteration: for given $\boldsymbol{y} \in U$, initialize the iteration with $q_N^{(1)}(\boldsymbol{y}) \in \mathcal{X}_N$, e.g. a simple case $q_N^{(1)}(\boldsymbol{y}) = 0$. For $k = 1, 2, \dots$, find $\delta q_N^{(k)}(\boldsymbol{y}) \in \mathcal{X}_N$ such that

$${}_{\mathcal{Y}'}\langle D_q\mathcal{R}(q_N^{(k)}(\boldsymbol{y});\boldsymbol{y})(\delta q_N^{(k)}(\boldsymbol{y})), v_N\rangle_{\mathcal{Y}} = -{}_{\mathcal{Y}'}\langle \mathcal{R}(q_N^{(k)}(\boldsymbol{y});\boldsymbol{y}), v_N\rangle_{\mathcal{Y}} \quad \forall v_N \in \mathcal{Y}_N \ . \tag{68}$$

Next, update the RB solution as

$$q_N^{(k+1)}(\boldsymbol{y}) = q_N^{(k)}(\boldsymbol{y}) + \delta q_N^{(k)}(\boldsymbol{y}) \ . \tag{69}$$

We termine the iteration (68) with $q_N(\boldsymbol{y}) = q_N^{(k+1)}(\boldsymbol{y})$, once the following criterion is satisfied for a prescribed tolerance ε_{tol}:

$$||\delta q_N^{(k)}(\boldsymbol{y})||_{\mathcal{X}} \leq \varepsilon_{tol} \quad \text{or} \quad ||\mathcal{R}(q_N^{(k)}(\boldsymbol{y});\boldsymbol{y})||_{\mathcal{Y}'} \leq \varepsilon_{tol} \ . \tag{70}$$

As pointed out in Remark 4.1, in order to ensure convergence of the Newton scheme, we update with rescaled increment $\alpha^{(k)}\delta q_N^{(k)}$ corresponding to step length $\alpha^{(k)}$ determined by a line search (e.g. [23]).

The accuracy of the RB-PG compression crucially depends on the RB space $\mathcal{X}_N$. The space $\mathcal{X}_N$ should consist of N basis functions which are "most representative" among all HiFi-PG approximations of the parametric solutions $q_h(\boldsymbol{y}) \in \mathcal{X}_h$, *uniformly with respect to* $\boldsymbol{y} \in U$. The basis functions can be constructed either as a set of HiFi solutions selected by a greedy algorithm or as their compression obtained via proper orthogonal decomposition. As the latter require a large number of HiFi solutions prior to compression, especially for high-dimensional parametric problems, we use the former approach. Formally, the RB space $\mathcal{X}_N$ is given by the span of "snapshots", ie., of HiFi-PG solutions at N selected parameter samples $\boldsymbol{y}^n$, $1 \leq n \leq N$. We express this formally as

$$\mathcal{X}_N = \text{span}\{q_h(\boldsymbol{y}^1),\dots,q_h(\boldsymbol{y}^N)\} \ . \tag{71}$$

For the selection of the parameter samples, we apply a greedy algorithm which is steered by a computable, goal-oriented error estimator $\triangle_N(\boldsymbol{y})$ which depends on the QoI. For example, $\triangle_N^\Theta(\boldsymbol{y})$ for the approximation of the density of the Bayesian posterior, ie., as estimator of $|\Theta_h(\boldsymbol{y}) - \Theta_N(\boldsymbol{y})|$. To ensure computational efficiency, we require that the estimator is computable in complexity that does not depend on the number N_h of HiFi-PG degrees of freedom. We defer the design of computable error estimators to the next section and assume them available for now. The greedy algorithm for the selection of the RB samples reads: pick the first parameter sample $\boldsymbol{y}^1$ randomly or as the barycenter of the (assumed convex) parameter space U, and define the first RB

space $\mathcal{X}_1 = \text{span}\{q_h(\boldsymbol{y}^1)\}$; then, for $N = 1, 2, \ldots$, until a stopping criterion is satisfied (e.g. maximum number of basis functions or prescribed error tolerance), we pick the next sample as

$$\boldsymbol{y}^{N+1} = \underset{\boldsymbol{y}\in U}{\text{argmax}}\, \triangle_N(\boldsymbol{y}) \,, \tag{72}$$

where the parameter space U is replaced by a finite training sample set $\Xi_t \subset U$ as, e.g., the set of adaptive sparse grid nodes or random samples. After solving the HiFi-PG problem (58) at $\boldsymbol{y}^{N+1}$, we enrich the RB space as $\mathcal{X}_{N+1} = \mathcal{X}_N \oplus \text{span}\{q_h(\boldsymbol{y}^{N+1})\}$. In order to obtain a well-conditioned algebraic RB system (90), we perform Gram-Schmidt orthogonalization procedure on the RB basis functions with respect to a suitable inner project, e.g. the one associated with the norm of $\mathcal{X}$, yielding a set of orthonormal basis w_N^n with $(w_N^n, w_N^{n'})_{\mathcal{X}} = \delta_{nn'}$, $1 \le n, n' \le N$.

To ensure the stability of the RB-PG approximations (90), we construct the RB test space $\mathcal{Y}_N$ via a "supremizer" approach [39], where the operator $T_{q_N^{(k)}(\boldsymbol{y})} : \mathcal{X}_h \to \mathcal{Y}_h$ is defined in our setting as

$$(T_{q_N^{(k)}(\boldsymbol{y})} w_h, v_h)_{\mathcal{Y}} = \ {}_{\mathcal{Y}'}\langle D_q \mathcal{R}(q_N^{(k)}(\boldsymbol{y}); \boldsymbol{y})(w_h), v_h\rangle_{\mathcal{Y}} \quad \forall v_h \in \mathcal{Y}_h, \quad 1 \le n \le N \,, \tag{73}$$

where $(\cdot,\cdot)_{\mathcal{Y}}$ is the inner-product in the Hilbert space $\mathcal{Y}$. By definition, we have that $T_{q_N^{(k)}(\boldsymbol{y})} w_h$ is the supremizer (maximizer by continuity and compactness) for the element $w_h \in \mathcal{X}_h$ with respect to the functional ${}_{\mathcal{Y}'}\langle D_q \mathcal{R}(q_N^{(k)}(\boldsymbol{y}); \boldsymbol{y})(w_h), \cdot\,\rangle_{\mathcal{Y}} : \mathcal{Y}_h \to \mathbb{R}$, i.e.

$$T_{q_N^{(k)}(\boldsymbol{y})} w_h = \underset{v_h\in\mathcal{Y}_h}{\text{argsup}}\ {}_{\mathcal{Y}'}\langle D_q \mathcal{R}(q_N^{(k)}(\boldsymbol{y}); \boldsymbol{y})(w_h), v_h\rangle_{\mathcal{Y}} \,. \tag{74}$$

Then, at every intermediate solution $q_N^{(k)}(\boldsymbol{y})$, we define the RB test space $\mathcal{Y}_N$ as

$$\mathcal{Y}_N \equiv \mathcal{Y}_{q_N^{(k)}(\boldsymbol{y})} := \text{span}\left\{T_{q_N^{(k)}(\boldsymbol{y})} w_N^1, \ldots, T_{q_N^{(k)}(\boldsymbol{y})} w_N^N\right\} \,, \tag{75}$$

where w_N^n, $1 \le n \le N$, are the basis functions of $\mathcal{X}_N$. Under this choice of the RB test space, we have

$$\begin{aligned}
\beta_N &:= \inf_{w_N\in\mathcal{X}_N} \sup_{v_N\in\mathcal{Y}_N} \frac{{}_{\mathcal{Y}'}\langle D_q \mathcal{R}(q_N^{(k)}(\boldsymbol{y}); \boldsymbol{y})(w_N), v_N\rangle_{\mathcal{Y}}}{||w_N||_{\mathcal{X}} ||v_N||_{\mathcal{Y}}} \\
(\text{by (74)}) &= \inf_{w_N\in\mathcal{X}_N} \frac{{}_{\mathcal{Y}'}\langle D_q \mathcal{R}(q_N^{(k)}(\boldsymbol{y}); \boldsymbol{y})(w_N), T_{q_N^{(k)}(\boldsymbol{y})} w_N\rangle_{\mathcal{Y}}}{||w_N||_{\mathcal{X}} ||T_{q_N^{(k)}(\boldsymbol{y})} w_N||_{\mathcal{Y}}} \\
(\text{by (73)}) &= \inf_{w_N\in\mathcal{X}_N} \frac{\left(T_{q_N^{(k)}(\boldsymbol{y})} w_N, T_{q_N^{(k)}(\boldsymbol{y})} w_N\right)_{\mathcal{Y}}}{||w_N||_{\mathcal{X}} ||T_{q_N^{(k)}(\boldsymbol{y})} w_N||_{\mathcal{Y}}} \\
((\cdot,\cdot)_{\mathcal{Y}} = ||\cdot||_{\mathcal{Y}}^2) &= \inf_{w_N\in\mathcal{X}_N} \frac{||T_{q_N^{(k)}(\boldsymbol{y})} w_N||_{\mathcal{Y}}}{||w_N||_{\mathcal{X}}} \\
(\mathcal{X}_N \subset \mathcal{X}_h) &\ge \inf_{w_h\in\mathcal{X}_h} \frac{||T_{q_N^{(k)}(\boldsymbol{y})} w_h||_{\mathcal{Y}}}{||w_h||_{\mathcal{X}}} \\
(\text{by (73) and (74)}) &= \inf_{w_h\in\mathcal{X}_h} \sup_{v_h\in\mathcal{Y}_h} \frac{{}_{\mathcal{Y}'}\langle D_q \mathcal{R}(q_N^{(k)}(\boldsymbol{y}); \boldsymbol{y})(w_h), v_h\rangle_{\mathcal{Y}}}{||w_h||_{\mathcal{X}} ||v_h||_{\mathcal{Y}}} =: \beta_h^{k,N}(\boldsymbol{y}) > 0 \,,
\end{aligned} \tag{76}$$

which implies that the RB-PG problem (90) is well-posed in the RB spaces $\mathcal{X}_N$ and $\mathcal{Y}_N$ as long as the HiFi-PG problem (59) is well-posed at $q_N^{(k)}(\boldsymbol{y})$, *with quantitative preservation of the discrete stability* $\beta_h^{k,N}(\boldsymbol{y})$.

Remark 5.1. *The construction of the RB test space presented above is equivalent to the least-squares PG approximation of the nonlinear reduced basis problem* (67)*, which is more expensive to assemble compared to the matrix of the Galerkin approximation or of PG with a fixed test basis.*

5.2. Generalized Empirical Interpolation Method (GEIM)

The resulting computable solution of problem (68) generally depends on the number N_h of HiFi degrees of freedom (unless the residual operator $\mathcal{R}$ is affine-parametric with respect to $\boldsymbol{y}$ and linear with respect to the solution q_N, which case was considered in detail in [16]). For nonaffine and nonlinear problems under

consideration here, we ensure efficiency by *greedy compression of the nonaffine parametric, nonlinear residual* $\mathcal{R}(\cdot;\boldsymbol{y}) \in \mathcal{Y}'$, through the so-called *generalized empirical interpolation method (GEIM) [32].* As explained, for example, in [32, Section 1], we require that the range of the parametric operator residual map $\mathcal{F} := \{\mathcal{R}(\cdot;\boldsymbol{y}) : \boldsymbol{y} \in U\} \subset \mathcal{Y}'$, is compact. This is, in engineering applications, a consequence of some (minimal) PDE regularity and of Rellich's theorem. Since the residual $\mathcal{R}(\cdot;\boldsymbol{y})$ is evaluated on the HiFi space $\mathcal{X}_h$, this compactness will also follow from the continuity $\mathcal{R}(\cdot;\boldsymbol{y}) : \mathcal{X} \to \mathcal{Y}'$ and from the fact that $N_h := \dim(\mathcal{X}_h) < \infty$ (albeit not uniformly w.r. to the HiFi discretization parameter h). The GEIM assumes at hand a *dictionary* $\Sigma \subset \mathcal{Y} \simeq \mathcal{L}(\mathcal{Y}')$ which dictionary satisfies the *GEIM properties:* $\Sigma 1$: $\forall \sigma \in \Sigma : \|\sigma\|_{\mathcal{Y}} = 1$, and $\Sigma 2$ (unisolvency): for every $\varphi \in \mathcal{F}$ holds: $\sigma(\varphi) = 0$ for every $\sigma \in \Sigma$ implies $\varphi = 0$. The GEIM builds recursively a sequence $\{\mathcal{F}_M\}_{M\geq 1}$ of M-dimensional subspaces $\mathcal{F}_M = \text{span}\{\varphi_1, ..., \varphi_M\} \subset \mathcal{F}$ and associated subsets $\Sigma_M = \{\sigma_1, ..., \sigma_M\} \subset \Sigma$ of linearly independent functionals such that for every $\varphi \in \mathcal{F}$ exists a unique GEIM interpolant $\mathcal{J}_M[\varphi]$, given by the affine expression

$$\mathcal{J}_M[\varphi] = \sum_{m=1}^{M} \lambda_m[\varphi] r_m \ , \tag{77}$$

such that $\sigma_i(\mathcal{J}_M[\varphi]) = \sigma_i(\varphi)$ for $i = 1, ..., M$. By GEIM property $\Sigma 2$, given collections $\{r_1, ..., r_M\}$ of reduced bases with $\text{span}\{r_1, ..., r_M\} = \text{span}\{\varphi_1, ..., \varphi_M\} = \mathcal{F}_M \subset \mathcal{F}$, the GEIM interpoland $\mathcal{J}_M[\varphi]$ is well-defined, and the coefficients $\lambda_m[\varphi]$ in (77) can be obtained from the solution of a (dense!) linear system of equations of (hopefully small) order M (see [32, Section 1]). We emphasize that the *linear system becomes diagonal* if the reduced bases $\text{span}\{r_1, ..., r_M\}$ are biorthogonal to the sets Σ_M of observation functionals, ie. when $\sigma_i(r_j) = \delta_{ij}$.

If $\mathcal{F} \subset C^0(\overline{D})$, it is possible to choose Σ as a collection of Dirac (nodal) functionals; the GEIM reduces in this case to its *nodal version*, the (original) empirical interpolation method (EIM for short) [1]. In the numerical experiments ahead, we build affine approximations of the nonaffine and nonlinear residual $\mathcal{R}$ by interpolating at a finite number of HiFi finite element nodes as follows: let $\mathcal{T}$ denote a shape regular, simplicial triangulation of the physical domain D with N_e elements $\{K_i\}_{i=1}^{N_e}$ and the finite element nodes $\{x_n\}_{n=1}^{N_h}$. We say that the finite element basis functions $\{w_h^n\}_{n=1}^{N_h}$ are a locally supported nodal basis of the HiFi space $\mathcal{X}_h$ if

$$w_h^n(x_m) = \delta_{mn}, \quad x_m \in K^{(n)} \quad \text{and} \quad w_h^n(x) = 0, \quad \forall x \notin K^{(n)}, \quad \forall m, n = 1, \ldots, N_h \ , \tag{78}$$

where δ_{mn} denotes the Kronecker symbol, i.e. $\delta_{mn} = 1$ if $m = n$ and $\delta_{mn} = 0$ otherwise, and $K^{(n)}$ denotes the set of elements surrounding the node x_n. Nodal bases are available, for instance, for any Lagrangian Finite Element space in the sense of Ciarlet, cp. e.g. [3, Section 3.1]. We develop this now for the particular case of *continuous, Lagrangian Finite Elements on regular, simplicial triangulation $\mathcal{T}$ of D, where the GEIM dictionary Σ becomes simply Dirac masses σ_n concentrated* on the HiFi Finite Element nodes x_n in the element $K_i \in \mathcal{T}$, $i = 1, \ldots, N_e$. Then $\Sigma = \{\sigma_n, 1 \leq n \leq N_h\}$, and for any $\varphi \in C^0(\overline{D})$ holds

$$\sigma_m(\varphi) = \varphi(x_m), \quad \forall m = 1, \ldots, N_h \ , \tag{79}$$

Then $\mathcal{J}_M\mathcal{R}(q_N^{(k)}(\boldsymbol{y});\boldsymbol{y})$, the empirical interpolation (EI) of the parametric residual map $\mathcal{R}(q_N^{(k)}(\boldsymbol{y});\boldsymbol{y})$ in (68), is given by

$$\mathcal{J}_M\mathcal{R}(q_N^{(k)}(\boldsymbol{y});\boldsymbol{y}) := \sum_{m=1}^{M} \lambda_m^{(k)}(\boldsymbol{y}) r_m(x) \ , \tag{80}$$

where the terms $\lambda_m^{(k)}(\boldsymbol{y})$, $1 \leq m \leq M$, $k = 1, 2, \ldots$, depend only on the parameter $\boldsymbol{y}$, and the reduced bases $r_m(\cdot)$, $1 \leq m \leq M$, to be constructed later, depend only on the physical variable $x \in D$. The coefficient vector $\boldsymbol{\lambda}^{(k)}(\boldsymbol{y}) = (\lambda_1^{(k)}(\boldsymbol{y}), \ldots, \lambda_M^{(k)}(\boldsymbol{y}))^\top$ is obtained by solving a nodal interpolation problem

$$\sigma_{m'}\left(\sum_{m=1}^{M} \lambda_m^{(k)}(\boldsymbol{y}) r_m(x)\right) = \sigma_{m'}\left(\mathcal{R}(q_N^{(k)}(\boldsymbol{y});\boldsymbol{y})\right), \quad m' = 1, \ldots, M \ , \tag{81}$$

whose computational cost depends on (a) the cost for evaluating the right hand side, and (b) the cost for the numerical solution of the dense (in fact, lower triangular) $M \times M$ linear system. For partial differential equation models, evaluation of the right hand side involves $q_N^{(k)}(\boldsymbol{y})$ only at the nodes in the set of elements $K^{(m')}$ surrounding the node $x_{m'}$ for locally supported HiFi basis functions. With the approximation of the residual

in (80), the Jacobian of the residual can be evaluated at any $w_h \in \mathcal{X}_h$, $1 \le n \le N_h$ by

$$D_q(\mathcal{J}_M\mathcal{R}(q_N^{(k)}(\boldsymbol{y});\boldsymbol{y}))(w_h) = \sum_{m=1}^{M} D_q(\lambda_m^{(k)}(\boldsymbol{y}))(w_h) r_m \equiv \boldsymbol{r} D_q(\boldsymbol{\lambda}^{(k)}(\boldsymbol{y}))(w_h) \;, \tag{82}$$

where we have denoted the vector of basis functions $\boldsymbol{r} = (r_1, \ldots, r_M)$; the coefficient $\boldsymbol{\lambda}^{(k)}(\boldsymbol{y})$ implicitly depends on the state variable $q_N^{(k)}(\boldsymbol{y})$ through the solution of the interpolation problem (81). To make this dependence explicit, we denote by R the matrix with entries $R_{mm'} = r_{m'}(x_m)$, $1 \le m, m' \le M$, and by $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_M)^\top$ a vector of the nodal functionals. Then, upon inverting (81), we obtain

$$\boldsymbol{\lambda}^{(k)}(\boldsymbol{y}) = R^{-1}\boldsymbol{\sigma}\left(\mathcal{R}\left(q_h^{(k)}(\boldsymbol{y});\boldsymbol{y}\right)\right) \;, \tag{83}$$

where $R^{-1} \in \mathbb{R}^{M\times M}$ (or a factorization of R) needs to be computed and stored only once. The cost for evaluation of (83) depends only on M: it is $O(M)$ for the functional evaluation and $O(M^2)$ for the matrix-vector product. The accuracy of the affine approximation (80) depends on the number M, the basis functions r_m, $1 \le m \le M$, as well as the functionals σ_m, $1 \le m \le M$. In order to obtain an accurate and efficient approximation (80), we adopt a greedy algorithm following that for empirical interpolation [1] to construct the basis functions r_m and the functionals σ_m, $1 \le m \le M$, which we describe next. Suppose we are given a realization of the residual (to be specified later), denoted (with slight abuse of notation) as $\mathcal{R} : D \to \mathbb{R}$ with pointwise definition, we define the first functional σ_1 and the first basis function r_1 as

$$\sigma_1 = \operatorname*{argmax}_{\sigma\in\Sigma} |\sigma(\mathcal{R})| \quad \text{and} \quad r_1 := \frac{\mathcal{R}}{\sigma_1(\mathcal{R})} \;. \tag{84}$$

Then, for $M = 1, 2, \ldots$, given a new realization of the residual (provided according to a posteriori error estimator in Sec. 5.3), and still denoted as $\mathcal{R}$, we define the next functional and the next basis function as

$$\sigma_{M+1} = \operatorname*{argmax}_{\sigma\in\Sigma} |\sigma(\mathcal{R} - \mathcal{J}_M\mathcal{R})| \quad \text{and} \quad r_{M+1} = \frac{\mathcal{R} - \mathcal{J}_M\mathcal{R}}{\sigma_{M+1}(\mathcal{R} - \mathcal{J}_M\mathcal{R})} \;, \tag{85}$$

where $\mathcal{J}_M\mathcal{R}$ is the EI defined in (80). The construction of EI is terminated until the interpolation error (or some computable estimator of it such as the one presented in Sec. 5.3) is below a prescribed tolerance.

Remark 5.2. *Instead of the Lagrange interpolation* (81)*, the coefficient can also be obtained by solving minimum residual (or least squares) problems when the number of functionals is larger than the number of EI basis functions, as employed in [6] using gappy POD reconstruction for affine approximation.*

Remark 5.3. *The EI can be applied selectively to interpolate only the nonlinear term rather than the complete residual term: when the residual operator can be written more explicitly as*

$$\mathcal{R}(q(\boldsymbol{y});\boldsymbol{y}) := A(\boldsymbol{y})q(\boldsymbol{y}) + B(q(\boldsymbol{y});\boldsymbol{y}) - F(\boldsymbol{y}) \;, \tag{86}$$

where A and B are linear and nonlinear parametric operators w.r.t. q, respectively, the EI may be applied only for an affine approximation of the nonlinear term $B(q(\boldsymbol{y});\boldsymbol{y})$ if A and F are affine w.r.t. $\boldsymbol{y}$ and still denote the approximation as (with slight abuse of notation)

$$\mathcal{J}_M\mathcal{R}(q(\boldsymbol{y});\boldsymbol{y}) = A(\boldsymbol{y})q(\boldsymbol{y}) + \mathcal{J}_M B(q(\boldsymbol{y});\boldsymbol{y}) - F(\boldsymbol{y}) \;. \tag{87}$$

If A and/or F depends in a nonaffine fashion on the parameter sequence $\boldsymbol{y}$, besides direct interpolation of the whole residual $\mathcal{R}$, one can also apply the EI for the nonlinear term B and for the nonaffine function in A and/or F, leading to separate empirical interpolation of each nonlinear and/or nonaffine term, i.e.

$$\mathcal{J}_M\mathcal{R}(q(\boldsymbol{y});\boldsymbol{y}) = \mathcal{J}^A_{M_A}A(\boldsymbol{y})q(\boldsymbol{y}) + \mathcal{J}^B_{M_B}B(q(\boldsymbol{y});\boldsymbol{y}) - \mathcal{J}^F_{M_F}F(\boldsymbol{y}) \;, \tag{88}$$

where $M = (M_A, M_B, M_F)$; $\mathcal{J}^A_{M_A}$, $\mathcal{J}^B_{M_B}$, and $\mathcal{J}^F_{M_F}$ are the EI interpolants with M_A, M_B, and M_F terms.

Remark 5.4. *The EIM [1] and its variants, including the* discrete empirical interpolation method (DEIM) *[7], the* empirical operator interpolation method (EOIM) *[25], as well as the nodal version of GEIM proposed here, employ the same greedy construction algorithm in different contexts. DEIM was proposed to directly approximates the assembled finite element quantities [7], e.g. $-{}_{\mathcal{Y}'}\langle\mathcal{R}(u_h^{(k)}(\boldsymbol{y});\boldsymbol{y}), v_h\rangle_{\mathcal{Y}}$; EOIM approximates the nonlinear*

operator and was applied in the finite volume context [25], while the nodal EIM approximates the residual $\mathcal{R}$ with nodal representation of the solution in the context of Petrov-Galerkin finite element approximation.

Let $q_{N,M}(\boldsymbol{y}) \in \mathcal{X}_N$ denote the RB-PG solution of (67) with the residual operator $\mathcal{R}$ approximated by its empirical interpolation $\mathcal{J}_M\mathcal{R}$, then we have the RB-EI problem:

$$\text{given } \boldsymbol{y} \in U, \quad \text{find } q_{N,M}(\boldsymbol{y}) \in \mathcal{X}_N : \quad {}_{\mathcal{Y}'}\langle \mathcal{J}_M\mathcal{R}(q_{N,M}(\boldsymbol{y}); \boldsymbol{y}), v_N\rangle_{\mathcal{Y}} = 0 \quad \forall v_N \in \mathcal{Y}_N \ . \tag{89}$$

Problem (89) can again be solved by a Newton iteration: find $\delta q^{(k)}_{N,M}(\boldsymbol{y}) \in \mathcal{X}_N$ such that

$${}_{\mathcal{Y}'}\langle D_q \left(\mathcal{J}_M\mathcal{R}(q^{(k)}_{N,M}(\boldsymbol{y}); \boldsymbol{y})\right)(\delta q^{(k)}_{N,M}(\boldsymbol{y})), v_N\rangle_{\mathcal{Y}} = -{}_{\mathcal{Y}'}\langle \mathcal{J}_M\mathcal{R}(q^{(k)}_{N,M}(\boldsymbol{y}); \boldsymbol{y}), v_N\rangle_{\mathcal{Y}} \quad \forall v_N \in \mathcal{Y}_N \ . \tag{90}$$

Given $\{w^n_N\}_{n=1}^N$ a basis of $\mathcal{X}_N$, we express the parametric, RB solution $q^{(k)}_{N,M}(\boldsymbol{y})$ and the increment $\delta q^{(k)}_{N,M}(\boldsymbol{y})$ as

$$q^{(k)}_{N,M}(\boldsymbol{y}) = \sum_{n=1}^N q^{(k,n)}_{N,M}(\boldsymbol{y}) w^n_N \quad \text{and} \quad \delta q^{(k)}_{N,M}(\boldsymbol{y}) = \sum_{n=1}^N \delta q^{(k,n)}_{N,M}(\boldsymbol{y}) w^n_N \ . \tag{91}$$

The algebraic formulation of problem (90) can be written as: find $\boldsymbol{\delta q}^{(k)}_{N,M}(\boldsymbol{y}) \in \mathbb{R}^N$ such that

$$J^{(k)}_{N,M}(\boldsymbol{y})\boldsymbol{\delta q}^{(k)}_{N,M}(\boldsymbol{y}) = \boldsymbol{r}^{(k)}_{N,M}(\boldsymbol{y}) \ . \tag{92}$$

Here, $\boldsymbol{\delta q}^{(k)}_{N,M}(\boldsymbol{y}) = (\delta q^{(k,1)}_{N,M}(\boldsymbol{y}), \ldots, \delta q^{(k,N)}_{N,M}(\boldsymbol{y}))^\top$ and the Jacobian matrix $J^{(k)}_{N,M}(\boldsymbol{y}) \in \mathbb{R}^{N\times N}$ is given by

$$\left(J^{(k)}_{N,M}(\boldsymbol{y})\right)_{nn'} = \sum_{m=1}^M \left(D_q(\boldsymbol{\lambda}^{(k)}(\boldsymbol{y}))(w^{n'}_N)\right)_m \ {}_{\mathcal{Y}'}\langle r_m, v^n_N\rangle_{\mathcal{Y}}, \quad 1 \le n, n' \le N \ . \tag{93}$$

where the vector $D_q(\boldsymbol{\lambda}^{(k)}(\boldsymbol{y}))(w^{n'}_N) \in \mathbb{R}^M$ is generated in $O(M^2)$ operations as in (82). The quantities ${}_{\mathcal{Y}'}\langle r_m, v^n_N\rangle_{\mathcal{Y}}$ (with $\{v^n_N\}_{n=1}^N$ denoting the basis of $\mathcal{Y}_N$), $1 \le m \le M$, $1 \le n \le N$, can be computed and stored for only once when $\{v^n_N\}_{n=1}^N$ are $\boldsymbol{y}$-independent. The right hand side of (92) is given by

$$\left(\boldsymbol{r}^{(k)}_{N,M}(\boldsymbol{y})\right)_n = -\sum_{m=1}^M \lambda^{(k)}_m(\boldsymbol{y}) \ {}_{\mathcal{Y}'}\langle r_m, v^n_N\rangle_{\mathcal{Y}}, \quad 1 \le n \le N \ , \tag{94}$$

where $\lambda^{(k)}_m(\boldsymbol{y})$ is computed as in 83 with $O(M^2)$ operations. In summary, the solution of the algebraic problem (92) takes $O(M^2N^2)$ operations for assembling and $O(N^3)$ operations for solving. These numbers are independent of the number N_h of HiFi degrees of freedom. Upon termination of the Newton iteration, the RB-EI surrogate $q_{N,M}(\boldsymbol{y})$ in (90) of the parametric forward map is available and *for any instance $\delta \in Y$ of data $\delta \in Y$*, a *parsimonious RB-EI compression of the Bayesian posterior density* $\Theta_{N,M}(\boldsymbol{y})$ can be obtained in $O(KN)$ operations (K is the number of observations)

$$\Theta_{N,M}(\boldsymbol{y}) = \exp\left(-\left(\delta - \boldsymbol{O}^\top_N \boldsymbol{q}_{N,M}(\boldsymbol{y})\right)^\top \Gamma^{-1}\left(\delta - \boldsymbol{O}^\top_N \boldsymbol{q}_{N,M}(\boldsymbol{y})\right)\right) \ , \tag{95}$$

where $\boldsymbol{O}_N = (\mathcal{O}(w^1_N), \ldots, \mathcal{O}(w^N_N))^\top$ and $\boldsymbol{q}_{N,M}(\boldsymbol{y}) = (q^{(K,1)}_{N,M}(\boldsymbol{y}), \ldots, q^{(K,N)}_{N,M}(\boldsymbol{y}))^\top$.

As the approximate Jacobian (82) allows an affine-parametric representation, for the definition of the parametric RB test basis functions $\{v^n_N(\boldsymbol{y})\}_{n=1}^N$, we can write the supremizing operator more explicitly as (see also [15])

$$T_{q^{(k)}_{N,M}(\boldsymbol{y})} w_h = \sum_{m=1}^M D_q(\lambda^{(k)}_m(\boldsymbol{y}))(w_h) T_m \ , \tag{96}$$

where the basis function T_m, is given as the solution of the following problem: find $T_m \in \mathcal{Y}_h$ such that

$$(T_m, v_h)_{\mathcal{Y}} = {}_{\mathcal{Y}'}\langle r_m, v_h\rangle_{\mathcal{Y}} \quad \forall v_h \in \mathcal{Y}_h, \forall m = 1, \ldots, M \ . \tag{97}$$

Therefore, the (n, n'), $1 \leq n, n' \leq N$, entry of the Jacobian (93) is given by

$$\left(J^{(k)}_{N,M}(\boldsymbol{y})\right)_{nn'} = \sum_{m=1}^{M}\sum_{m'=1}^{M} \left(D_q(\boldsymbol{\lambda}^{(k)}(\boldsymbol{y}))(w_N^{n'})\right)_m \left(D_q(\boldsymbol{\lambda}^{(k)}(\boldsymbol{y}))(w_N^{n})\right)_{m'} (T_m, T_{m'})_{\mathcal{Y}} \ , \tag{98}$$

where the HiFi quantities $(T_m, T_{m'})_{\mathcal{Y}}$, $1 \leq m, m' \leq M$ need to be computed and stored only once, while at each Newton iteration, the Jacobian matrix is assembled with $O(M^2N^2)$ operations.

5.3. Goal-oriented A-posteriori Error Estimates

The construction of both the RB compression and the EI relies on the greedy algorithm. For its computational realization an inexpensive and reliable a-posteriori error estimator is crucial. We develop now such an estimator to compute the posterior density and related QoI in Bayesian inversion. The estimator is based on the HiFi posterior density Θ_h as example to present a goal-oriented a-posteriori error estimator based on the dual-weighted-residual. Firstly, we formally expand $\Theta_h(\boldsymbol{y}) = \Theta(q_h(\boldsymbol{y}))$ at its approximate $\Theta_{N,M}(\boldsymbol{y}) = \Theta(u_{N,M}(\boldsymbol{y}))$ as

$$\Theta(q_h(\boldsymbol{y})) - \Theta(q_{N,M}(\boldsymbol{y})) = D_q\Theta(q_{N,M}(\boldsymbol{y}))(q_h(\boldsymbol{y}) - q_{N,M}(\boldsymbol{y})) + O(||q_h(\boldsymbol{y}) - q_{N,M}(\boldsymbol{y})||^2_{\mathcal{X}}) \ , \tag{99}$$

where $D_q\Theta(q_{N,M}(\boldsymbol{y}))$ denotes the Fréchet derivative of Θ with respect to q evaluated at $q_{N,M}(\boldsymbol{y})$, whose existence can be proved as in [15]. The dual problem associated with the first term reads: given $q_{N,M}(\boldsymbol{y}) \in \mathcal{X}_N$, find the HiFi solution $\psi_h(\boldsymbol{y}) \in \mathcal{Y}_h$ such that

$${}_{\mathcal{Y}'}\langle D_q\mathcal{R}(q_{N,M}(\boldsymbol{y}); \boldsymbol{y})(w_h), \psi_h(\boldsymbol{y})\rangle_{\mathcal{Y}} = D_q\Theta(q_{N,M}(\boldsymbol{y}))(w_h) \quad \forall w_h \in \mathcal{X}_h \ . \tag{100}$$

In order to reduce the computational cost for solving the HiFi linear dual problem (100), we apply the RB compression, where the RB trial space $\mathcal{Y}_N^{du}$ is constructed as

$$\mathcal{Y}_N^{du} = \text{span}\{\psi_h(\boldsymbol{y}^1), \ldots, \psi_h(\boldsymbol{y}^N)\} \equiv \text{span}\{\psi_N^1, \ldots, \psi_N^N\} \ , \tag{101}$$

where ψ_N^n, $1 \leq n \leq N$, are the orthonormal basis of $\mathcal{Y}_N^{du}$. The RB test space $\mathcal{X}_N^{du}$ with basis $\{\phi_N^n\}_{n=1}^N$ can be constructed in the same way using the supremizer operator as in the last section. Then the RB-PG compression of the dual problem (100) reads: given $q_{N,M}(\boldsymbol{y}) \in \mathcal{X}_N$, find $\psi_N(\boldsymbol{y}) \in \mathcal{Y}_N^{du}$ such that

$${}_{\mathcal{Y}'}\langle D_q\mathcal{R}(q_{N,M}(\boldsymbol{y}); \boldsymbol{y})(\phi_N^n), \psi_N(\boldsymbol{y})\rangle_{\mathcal{Y}} = D_q\Theta(q_{N,M}(\boldsymbol{y}))(\phi_N^n) \quad \forall \phi_N^n \in \mathcal{X}_N^{du} \ . \tag{102}$$

Moreover, upon application the EI of the residual $\mathcal{R}(q_{N,M}(\boldsymbol{y}); \boldsymbol{y})$, we obtain the RB-EI compression of the dual problem (100): given $q_{N,M}(\boldsymbol{y}) \in \mathcal{X}_N$, find $\psi_{N,M}(\boldsymbol{y}) \in \mathcal{Y}_N^{du}$ such that

$${}_{\mathcal{Y}'}\langle D_q\left(\mathcal{J}_M\left(\mathcal{R}(q_{N,M}(\boldsymbol{y}); \boldsymbol{y})\right)\right)(\phi_N^n), \psi_{N,M}(\boldsymbol{y})\rangle_{\mathcal{Y}} = D_q\Theta(q_{N,M}(\boldsymbol{y}))(\phi_N^n) \quad \forall \phi_N^n \in \mathcal{X}_N^{du} \ . \tag{103}$$

Let $\boldsymbol{\psi}_{N,M}(\boldsymbol{y}) = (\psi^{(1)}_{N,M}(\boldsymbol{y}), \ldots, \psi^{(N)}_{N,M})^\top$ denote the coefficient vector of the RB-EI dual solution $\psi_{N,M}(\boldsymbol{y})$ on the basis $\{\psi_N^n\}_{n=1}^N$, then the algebraic formulation of problem (103) can be written as

$$J_{N,M}(\boldsymbol{y})\boldsymbol{\psi}_{N,M}(\boldsymbol{y}) = D_q\boldsymbol{\Theta}(q_{N,M}(\boldsymbol{y})) \ , \tag{104}$$

where the Jacobian matrix $J_{N,M}(\boldsymbol{y})$ is defined as

$$(J_{N,M}(\boldsymbol{y}))_{nn'} = \sum_{m=1}^{M} D_q(\lambda_m(\boldsymbol{y}))(\phi_N^n)_{\mathcal{Y}'}\langle r_m, \psi_N^{n'}\rangle_{\mathcal{Y}}, \quad 1 \leq n, n' \leq N \ . \tag{105}$$

As ${}_{\mathcal{Y}'}\langle r_m, \psi_N^{n'}\rangle_{\mathcal{Y}}$ is parameter-independent (or $\psi_N^{n'}$ can be written as an affine expansion on parameter-independent supremizers as done in (98)), they can be computed and stored once and for all. Assembling of $J_{N,M}(\boldsymbol{y})$ takes $O(M^2N^2)$ operations for any $\boldsymbol{y} \in U$. The right hand side $D_q\boldsymbol{\Theta}(q_{N,M}(\boldsymbol{y}))$ is defined as

$$(D_q\boldsymbol{\Theta}(q_{N,M}(\boldsymbol{y})))_n = D_q\Theta(q_{N,M}(\boldsymbol{y}))(\phi_N^n) = 2\Theta(q_{N,M}(\boldsymbol{y}))\left(-\delta + \boldsymbol{O}_N^\top \boldsymbol{q}_{N,M}(\boldsymbol{y})\right)^\top \Gamma^{-1}\mathcal{O}(\phi_N^n) \ , \tag{106}$$

where $\mathcal{O}(\phi_N^n)$ can be computed and stored once and for all. Evaluation of $D\boldsymbol{\Theta}(q_{N,M}(\boldsymbol{y}))$ takes $O(KN^2)$ operations. Therefore, it takes $O((M^2 + K)N^2)$ operations to assembe (104) and $O(N^3)$ operations to solve it.

To this end, we compute the first term of the expansion (99) by the definition of the dual problem (100) as

$$\begin{aligned} D_q\Theta(q_{N,M}(\boldsymbol{y}))(q_h(\boldsymbol{y})-q_{N,M}(\boldsymbol{y})) =&\ {}_{\mathcal{Y}'}\langle D_q\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})(q_h(\boldsymbol{y})-q_{N,M}(\boldsymbol{y})),\psi_h(\boldsymbol{y})\rangle_{\mathcal{Y}} \\ =&\ {}_{\mathcal{Y}'}\langle D_q\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})(q_h(\boldsymbol{y})-q_{N,M}(\boldsymbol{y})),\psi_{N,M}(\boldsymbol{y})\rangle_{\mathcal{Y}} \\ &+O(||q_h(\boldsymbol{y})-q_{N,M}(\boldsymbol{y})||_{\mathcal{X}}||\psi_h(\boldsymbol{y})-\psi_{N,M}(\boldsymbol{y})||_{\mathcal{Y}}) \ . \end{aligned} \tag{107}$$

Moreover, by the (formal) Taylor expansion of $\mathcal{R}(q_h(\boldsymbol{y});\boldsymbol{y})$ at $q_{N,M}(\boldsymbol{y})$, we have

$$D_q\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})(q_h(\boldsymbol{y})-q_{N,M}(\boldsymbol{y})) = \mathcal{R}(q_h(\boldsymbol{y});\boldsymbol{y})-\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})+O(||q_h(\boldsymbol{y})-q_{N,M}(\boldsymbol{y})||_{\mathcal{X}}^2) \ . \tag{108}$$

By the definition of the HiFi-PG problem (58) and $\psi_{N,M}(\boldsymbol{y})\in\mathcal{Y}_N^{du}\subset\mathcal{Y}_h$, we have

$${}_{\mathcal{Y}'}\langle\mathcal{R}(q_h(\boldsymbol{y});\boldsymbol{y}),\psi_{N,M}(\boldsymbol{y})\rangle_{\mathcal{Y}}=0 \ ; \tag{109}$$

moreover, by the EI compression of $\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})$, we have

$$\begin{aligned} -{}_{\mathcal{Y}'}\langle\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y}),\psi_{N,M}(\boldsymbol{y})\rangle_{\mathcal{Y}} =&\ -{}_{\mathcal{Y}'}\langle\mathcal{J}_M\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y}),\psi_{N,M}(\boldsymbol{y})\rangle_{\mathcal{Y}} \\ &-{}_{\mathcal{Y}'}\langle(\mathcal{J}_{M+M_{est}}-\mathcal{J}_M)\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y}),\psi_{N,M}(\boldsymbol{y})\rangle_{\mathcal{Y}} \\ &-{}_{\mathcal{Y}'}\langle(\mathcal{J}-\mathcal{J}_{M+M_{est}})\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y}),\psi_{N,M}(\boldsymbol{y})\rangle_{\mathcal{Y}} \ , \end{aligned} \tag{110}$$

where we use a (presumably more accurate) EI with M_{est} more terms to obtain a computable a-posteriori error estimator for the EI error with M terms. In doing this, we assume implicitly that the third term is dominated by the second term when M_{est} is large. A combination of (107), (108), (109), and (110) leads to the following *computable a-posteriori error estimate*

$$D_q\Theta(q_{N,M}(\boldsymbol{y}))(q_h(\boldsymbol{y})-q_{N,M}(\boldsymbol{y}))\approx\triangle_{N,M}^{\Theta}(\boldsymbol{y})=\triangle_{N,M}^{\Theta,RB}(\boldsymbol{y})+\triangle_{N,M}^{\Theta,EI}(\boldsymbol{y}) \tag{111}$$

where the first term, defined as

$$\triangle_{N,M}^{\Theta,RB}(\boldsymbol{y}):=-{}_{\mathcal{Y}'}\langle\mathcal{J}_M\left(\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})\right),\psi_{N,M}(\boldsymbol{y})\rangle_{\mathcal{Y}} \tag{112}$$

provides an error estimate of the RB compression error, and the second term given by

$$\triangle_{N,M}^{\Theta,EI}(\boldsymbol{y}):=-{}_{\mathcal{Y}'}\langle(\mathcal{J}_{M+M_{est}}-\mathcal{J}_M)\left(\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})\right),\psi_{N,M}(\boldsymbol{y})\rangle_{\mathcal{Y}} \ , \tag{113}$$

provides an error estimate of the EI compression error. Given $q_{N,M}(\boldsymbol{y})$ and $\psi_{N,M}(\boldsymbol{y})$, evaluation of $\triangle_{N,M}^{\Theta}(\boldsymbol{y})$ takes $O((M+M_{est})^2N^2)$ operations in total, which is independent of the number N_h of HiFi degrees of freedom. As $\triangle_{N,M}^{\Theta}(\boldsymbol{y})$ is an approximation of the first order Taylor expansion of $\Theta(q_h(\boldsymbol{y}))$ at $q_{N,M}(\boldsymbol{y})$, we may correct the RB-EI compression $\Theta_{N,M}(\boldsymbol{y})=\Theta(q_{N,M}(\boldsymbol{y}))$ by

$$\Theta_{N,M}^c(\boldsymbol{y})=\Theta_{N,M}(\boldsymbol{y})+\triangle_{N,M}^{\Theta}(\boldsymbol{y}) \ , \tag{114}$$

which is supposed to be more accurate than $\Theta_{N,M}(\boldsymbol{y})$ for the approximation of $\Theta(q_h(\boldsymbol{y}))$, as long as $\Theta_{N,M}(\boldsymbol{y})$ provides a good approximation of $D_q\Theta(q_{N,M}(\boldsymbol{y}))(q_h(\boldsymbol{y})-q_{N,M}(\boldsymbol{y}))$ for any given $\boldsymbol{y}\in U$.

Remark 5.5. *When the residual operator is explicitly given as* (86) *with affine-parametric A and F, we have*

$$\triangle_{N,M}^{\Theta,RB}(\boldsymbol{y}):={}_{\mathcal{Y}'}\langle F(\boldsymbol{y})-A(\boldsymbol{y})q_{N,M}(\boldsymbol{y})-\mathcal{J}_M(B(q_{N,M}(\boldsymbol{y});\boldsymbol{y})),\psi_{N,M}(\boldsymbol{y})\rangle_{\mathcal{Y}} \ , \tag{115}$$

and

$$\triangle_{N,M}^{\Theta,EI}(\boldsymbol{y}):=-{}_{\mathcal{Y}'}\langle(\mathcal{J}_{M+M_{est}}-\mathcal{J}_M)\left(B(q_{N,M}(\boldsymbol{y});\boldsymbol{y})\right),\psi_{N,M}(\boldsymbol{y})\rangle_{\mathcal{Y}} \ . \tag{116}$$

When the residual operator is given as (88) *with nonaffine terms A and F, we have*

$$\triangle_{N,M}^{\Theta,RB}(\boldsymbol{y}):={}_{\mathcal{Y}'}\langle\mathcal{J}_{M_F}^F F(\boldsymbol{y})-\mathcal{J}_{M_A}^A A(\boldsymbol{y})q_{N,M}(\boldsymbol{y})-\mathcal{J}_{M_B}^B(B(q_{N,M}(\boldsymbol{y});\boldsymbol{y})),\psi_{N,M}(\boldsymbol{y})\rangle_{\mathcal{Y}} \ , \tag{117}$$

and

$$\begin{aligned} \triangle_{N,M}^{\Theta,EI}(\boldsymbol{y}):=&{}_{\mathcal{Y}'}\langle(\mathcal{J}_{M_F+M_{est}}^F-\mathcal{J}_{M_F}^F)F(\boldsymbol{y}) \\ &-(\mathcal{J}_{M_A+M_{est}}^A-\mathcal{J}_{M_A}^A)A(\boldsymbol{y})q_{N,M}(\boldsymbol{y})-(\mathcal{J}_{M_B+M_{est}}^B-\mathcal{J}_{M_B}^B)B(q_{N,M}(\boldsymbol{y});\boldsymbol{y}),\psi_{N,M}(\boldsymbol{y})\rangle_{\mathcal{Y}} \ . \end{aligned} \tag{118}$$

5.4. Adaptive greedy algorithm

Algorithm 1 Adaptive greedy algorithm for SG-RB-EI construction

procedure INITIALIZATION:

Specify the tolerance ε_{ei}^{rb} for the RB-EI compression of Θ; specify the tolerances ε_{sg}, ε_{ei} and ε_{rb}, and/or the maximum number of nodes/bases M_{sg}^{max}, M_{ei}^{max} and N_{max} for the construction of SG, RB and EI, respectively. Set $\mathcal{E}_{sg} = 2\varepsilon_{sg}$; choose $M_{est} \in \mathbb{N}$ (e.g. as 10);

Initialize the SG, RB and EI compression with $M_{sg} = N = 1$ and $M_{ei} = M_{est}$;

a. solve the primal and dual HiFi problems (58) and (100) at the root node $\boldsymbol{y}^1 = \mathbf{0} \in U$;

b. initialize the sparse grid index set $\Lambda_1 = \{\mathbf{1}\}$, and construct the SG approximation, either the interpolation as $\mathcal{S}_{\Lambda_1}\Theta_h(\boldsymbol{y}) = \Theta_h(\boldsymbol{y}^1)$ or the integration as $\mathbb{E}[\mathcal{S}_{\Lambda_1}\Theta_h] = \Theta_h(\boldsymbol{y}^1)$;

c. initialize the EI compression (80) by (84) and (85) based on a set of $\mathcal{R}(q_h^{(k)}(\boldsymbol{y}); \boldsymbol{y})$ at some random samples $\boldsymbol{y} \in U$, yielding basis functions r_m and nodes x_m, $1 \le m \le M_{est}$;

d. initialize the RB-EI compression with the primal trial space $\mathcal{X}_1 = \text{span}\{q_h(\boldsymbol{y}^1)\}$ and the dual trial space $\mathcal{Y}_1 = \text{span}\{\psi_h(\boldsymbol{y}^1)\}$, compute all the offline quantities for (92) and (103).

end procedure

procedure CONSTRUCTION:

  while $M_{sg} < M_{sg}^{max}$ and $\mathcal{E}_{sg} > \varepsilon_{sg}$ **do**

    compute the active index set $\Lambda_{M_{sg}}^a$ for the SG approximation;

    for each $\nu \in \Lambda_{M_{sg}}^a$ **do**

      compute the set of added nodes $\Xi_{\triangle}^{\nu}$ associated to ν;

      for each $\boldsymbol{y} \in \Xi_{\triangle}^{\nu}$ **do**

        solve the primal and dual RB-EI problems (89) and (103);

        compute the error estimator $\triangle_{N,M}^{\Theta}(\boldsymbol{y})$, $\triangle_{N,M}^{\Theta,RB}(\boldsymbol{y})$ and $\triangle_{N,M}^{\Theta,EI}(\boldsymbol{y})$ in (111);

        if $\triangle_{N,M}^{\Theta}(\boldsymbol{y}) \ge \varepsilon_{ei}^{rb}$ **then**

          solve the primal and dual HiFi problems (58) and (100);

          if $\triangle_{N,M}^{\Theta,EI} \ge \varepsilon_{ei}$ **then**

            refine the EI according to (85) based on the collections $\mathcal{R}(q_h^{(k)}(\boldsymbol{y}))$;

          end if

          if $\triangle_{N,M}^{\Theta,RB} \ge \varepsilon_{rb}$ **then**

            enrich the primal and dual RB spaces with $q_h(\boldsymbol{y})$ and $\psi_h(\boldsymbol{y})$, respectively;

          end if

          compute the offline quantities for RB-EI compressions and error estimates;

        end if

      end for

    end for

    compute the SG error estimate $\mathcal{E}_a$ with the RB-EI approximate $\Theta_{N,M}^{\Theta}$;

    enrich $\Lambda_{M_{sg}}$ by the index $\nu^{M_{sg}+1}$ corresponding to the largest SG error estimate;

    set $M_{sg} = M_{sg} + 1$ and go to next step;

  end while

end procedure

We construct the EI and RB compressions by a greedy search algorithm based on the a-posteriori error estimator (111). To illustrate this, based on the parametric posterior density $\Theta(\boldsymbol{y})$ and present a combined and goal-oriented greedy algorithm for the simultaneous construction of SG, EI and RB in Algorithm 1. During initialization, we specify the tolerances for the each of the SG, EI and RB compression errors as well as for the RB-EI compression of the posterior density Θ at every $\boldsymbol{y} \in U$. We may also specify the maximum number of the nodes/bases for each approximation in order to terminate the greedy search. To initialize the SG, either for interpolation or integration, we solve the primal and dual HiFi problems at the root node of the sparse grid and construct a constant (w.r. to $\boldsymbol{y}$) approximation based on the solution. The EI is initialized with M_{est} terms by performing a greedy algorithm outlined as in (84) and (85) based on a set of the residuals $\mathcal{R}(q_h^{(k)}(\boldsymbol{y}); \boldsymbol{y})$, $1 \le k \le K$, at a sequence of random samples $\boldsymbol{y} \in U$. More explicitly, the following algorithms are feasible:

(i) take the residual (e.g. $\mathcal{R}(q_h^{(k^*)}(\boldsymbol{y}^1); \boldsymbol{y}^1)$ at $k^* := \text{argmax}_{k \in \{1,\dots,K(\boldsymbol{y}^1)\}} ||\mathcal{R}(q_h^{(k)}(\boldsymbol{y}^1); \boldsymbol{y}^1)||_{\mathcal{Y}'}$, here $\boldsymbol{y}^1$ is the first sample) as the first basis function to construct the EI with $M = 1$. Then for each new sample $\boldsymbol{y}$,

take the residual $\mathcal{R}(q_h^{(k^*)}(\boldsymbol{y});\boldsymbol{y})$ at $k^* = \text{argmax}_{k\in\{1,\dots,K(\boldsymbol{y})\}} ||\mathcal{R}(q_h^{(k)}(\boldsymbol{y});\boldsymbol{y}) - \mathcal{J}_M\mathcal{R}(q_h^{(k)}(\boldsymbol{y});\boldsymbol{y})||_{\mathcal{Y}'}$ as the next basis function to update the EI, and set $M := M + 1$;

(ii) perform a proper orthogonal decomposition (POD) for $\mathcal{R}(q_h^{(k)}(\boldsymbol{y});\boldsymbol{y})$, $1 \le k \le K$, at each sample $\boldsymbol{y} \in U$, and take the first POD mode (or the first few POD modes) as the new basis function to construct and update the EI, governed by the EI error estimate as in (i).

(iii) in the case the residual is approximated as in (87), we may adopt the same scheme as in (i); an alternative scheme is to take $B(q_h^{K(\boldsymbol{y})}(\boldsymbol{y});\boldsymbol{y})$ as the basis function to update the EI for the nonlinear term B, which may provide accurate approximation of B at the solution of (86).

(iv) in the case the residual is approximated as in (88), we can update the EI compression for each term, using the bases $A(\boldsymbol{y})$, $B(q_h(\boldsymbol{y});\boldsymbol{y})$ and $F(\boldsymbol{y})$ at the selected realization $\boldsymbol{y} \in U$, respectively.

It is also of practical importance to allow for construction of more than one basis functions for the EI compression at each $\boldsymbol{y}$. The RB-PG compressions for both the primal and dual problems are initialized by the solution at the root node. In the procedure of construction for the SG-RB-EI compression, we replace $\Theta(q_h(\boldsymbol{y}))$ for all $\boldsymbol{y}$ in the sparse grids by their RB-EI compression $\Theta_{N,M}(\boldsymbol{y})$ (or the corrected value $\Theta^c_{N,M}(\boldsymbol{y})$); whenever the approximation does not meet the prescribed tolerance at some $\boldsymbol{y} \in U$, we refine the EIM and/or the RB compressions if their a-posteriori error estimates are larger than their error tolerances.

6. Numerical Experiments

We present two numerical examples, one focusing on a nonlinear operator with affine parametrization, the second addressing a simultaneously nonlinear operator and nonaffine parametrization arising from the transformation of shape uncertainty from a given realization of a random domain to a fixed, "nominal" domain.

6.1. Affine-parametric, nonlinear operator equation

We consider a semilinear, parametric elliptic problem in the physical domain $D = (0,1)^2$: given $\boldsymbol{y} \in U$, find $q(\boldsymbol{y}) : D \to \mathbb{R}$ such that

$$-\operatorname{div}(u(\boldsymbol{y})\nabla q(\boldsymbol{y})) + q^3(\boldsymbol{y}) = f \quad \text{in } D, \;\; q(\boldsymbol{y}) = 0 \quad \text{on } \partial D, \tag{119}$$

where homogeneous Dirichlet boundary data are prescribed on the whole boundary only for simplicity (mixed or inhomogeneous boundary data can be handled by the present techniques without essential changes). This problem fits into the abstract setting (1) with $\mathcal{X} = \mathcal{Y} = H^1_0(D)$.

The parametric diffusion coefficient $u(\boldsymbol{y})$ is defined as in (9), where for some truncation dimension $J \in \mathbb{N}_+$ we specify

$$\langle u \rangle = 2 \quad \text{and} \quad \psi_j = \frac{1}{j^\alpha}\sin(j_1\pi x_1)\sin(j_2\pi x_2), \quad j = 1,\dots,J\,, \tag{120}$$

where $x = (x_1, x_2) \in D$, $j_1 = j - [j/\sqrt{J}] \times \sqrt{J}$, and $j_2 = (j - j_1)/\sqrt{J}$, and we set the scaling parameter as $\alpha = 2$. Note that we set the random coefficient as affine w.r.t. the parameter $\boldsymbol{y}$ in order to focus on the approximation for the nonlinear term. See [16] for the treatment of nonaffine coefficients such as e^u. We assume that the $K = 9$ observation data are given by Gaussian convolutions (signifying, for example, sensors such as transducers) of the forward solution at randomly sampled parameters $\boldsymbol{y} \in U$,

$$o_k(q(\boldsymbol{y})) = \text{Gauss}(q(\boldsymbol{y}); x^{(k)}, d_k) := \int_D \frac{1}{\sqrt{2\pi}d_k}\exp\left(-\frac{(x - x^{(k)})^2}{2d_k^2}\right) q(\boldsymbol{y})dx, \tag{121}$$

where the locations $x^{(1)},\dots,x^{(K)}$ are uniformly distributed inside the domain D and the width $d_1 = \cdots = d_K$ is such that an d_k ball-about $x^{(k)}$ is contained in D. The covariance operator of the gaussian observation noise η is chosen as $\Gamma = \sigma^2 I$, being I the $K \times K$ identity matrix. We set $d_1 = 0.1$ and $\sigma = 0.1$. For the high-fidelity approximation, we use the Finite Element Method with continuous, piecewise polynomials on a uniform mesh of size $h = 2^{-n}$, $n \in \mathbb{N}_+$. As indicated in Remark 5.3, we apply the EI only for the nonlinear term $q^3(\boldsymbol{y})$ of problem (119), whose construction is as in section 5.

6.1.1. Sparse grid approximation error

In the first experiment, we test the convergence of the adaptive sparse grid interpolation and integration errors w.r.t. the number of indices and w.r.t. the number of PDE solves in $J = 64$ parameter dimensions. We construct the adaptive sparse grid using the interpolation error indicator (52) and the integration error indicator (54), respectively. We specify the mesh size as $h = 2^{-5}$. Hierarchical Clenshaw-Curtis nodes (47) are used for both the interpolation and the integration.

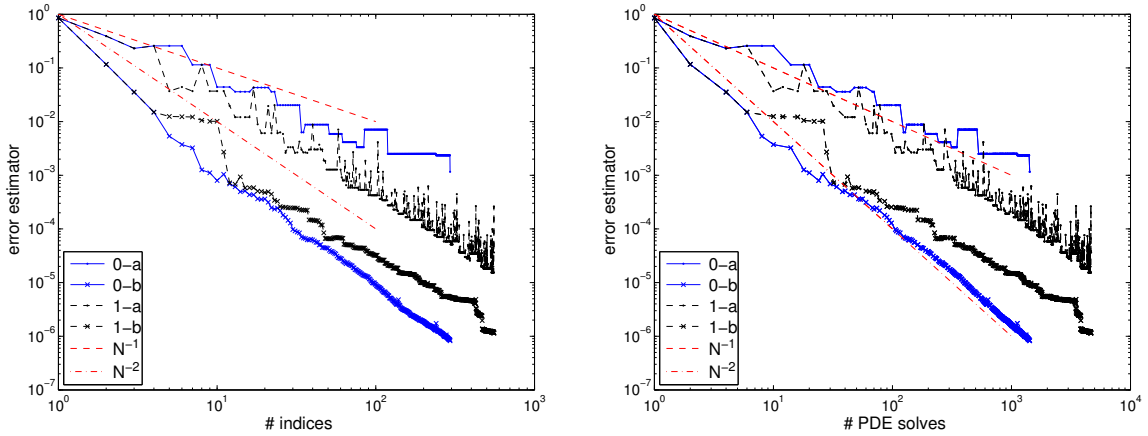

Figure 1: Decay of the interpolation (a) and the integration (b) error estimators (55) w.r.t. the number of indices (left) and PDE solves (right), for which the adaptive sparse grid is constructed by the integration error indicator (0) (54) and the interpolation error indicator (1) (52).

The convergence results of the interpolation and the integration of the parametric posterior density $\Theta(\boldsymbol{y})$ are reported in Figure 1. From these results we can draw the following observations: (i) the interpolation error indicator leads to much smaller interpolation error compared to that by the integration error indicator; it is also evident that the integration error indicator results in much smaller integration error compared to that by interpolation error indicator, which suggests that for different approximation purpose, choosing an appropriate error indicator is very important. (ii) the convergence rate of the interpolation w.r.t. the number of indices is larger than M^{-s}, with $s = 1/p - 1 = 2 - 1 = 1$ for the choice of basis in (120), as predicted in Theorem 2.6, which is due to that more than one node is used for each index for the choice of the Clenshaw-Curtis nodes; this effect can be demonstrated by the fact that the convergence rate is in good agreement with the prediction M^{-1} when we plot the error against the number of nodes (PDE solves). (iii) the observed convergence rate of the integration error estimator w.r.t. the number of PDE solves, evidently M^{-2}, is of higher order than the first order M^{-1} predicted in Theorem 2.6. We conjecture that this is due to that the integration error is measured in $L^1(U)$-norm (in average) while the interpolation error is measured in $L^\infty(U)$-norm (the worst case scenario).

6.1.2. High-fidelity approximation error

In this experiment, we study the convergence of the high-fidelity approximation error w.r.t. the mesh size with different polynomial degrees as well as the dependence of the number of basis functions for EI and RB compressions on the mesh size. We first pick a random sample $\bar{\boldsymbol{y}} \in U$ of dimension $J = 64$, at which we solve the high-fidelity problem (58) on a uniform mesh of size $h = 2^{-n}$, where we choose $n = 4, 5, 6, 7, 8$ and set $\bar{h} = 2^{-8}$ as the reference value to compute the accurate approximation of the observation functional $\mathcal{O}_{\bar{h}}(q_{\bar{h}}(\bar{\boldsymbol{y}})) \in \mathbb{R}^K$. We compute the finite element error at h by

$$\text{finite element error } (h) = \left|\mathcal{O}_h(q_h(\bar{\boldsymbol{y}})) - \mathcal{O}_{\bar{h}}(q_{\bar{h}}(\bar{\boldsymbol{y}}))\right| . \tag{122}$$

We use both linear and quadratic continuous, piecewise polynomial functions as the finite element basis functions to test the convergence rate of the finite element error w.r.t. the mesh size. The convergence result is displayed in the left part of Figure 2, where the finite element error is plotted against the reciprocal of the mesh size $1/h$. From this figure we can observe an empirical asymptotic convergence rate of $O(h^2)$ for P1 Lagrange Finite Elements and of $O(h^4)$ for P2 Lagrange Finite Elements, indicating the sharpness of the asymptotic convergence estimates in Theorem 8.2 ahead. These rates are, in fact, consistent with the Aubin-Nitsche duality argument mentioned in Remark 8.1 ahead: based on this argument, we expect a rate of $h^{k+k'}$ with $k = k' = 1$ for P1 Lagrange Finite Elements and a rate of $k = k' = 2$ for P2 Lagrange Finite Elements. In order to investigate the

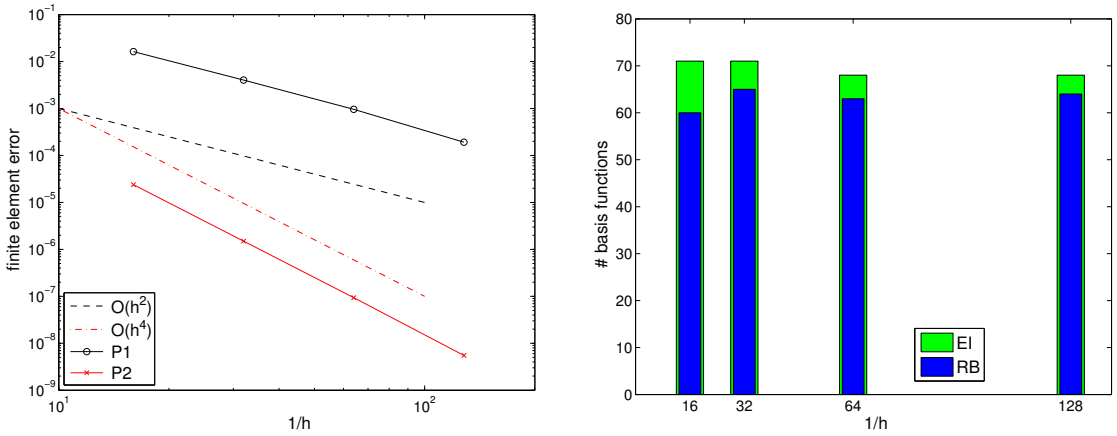

Figure 2: Left: decay of finite element error $|\mathcal{O}_h(q_h(\bar{\boldsymbol{y}})) - \mathcal{O}_{\bar{h}}(q_{\bar{h}}(\bar{\boldsymbol{y}}))|$ w.r.t. the reciprocal of the mesh size $(1/h)$; right: the number of basis functions at different $1/h$ constructed for EI and RB compressions by Algorithm 1, where the EI and RB tolerances $\varepsilon_{ei} = 10^{-7}$ and $\varepsilon_{rb} = 10^{-5}$.

influence of the mesh size (P1 element is used) to the number of the basis functions in EI and RB, we construct the SG-RB-EI interpolation (corresponding to the SG interpolation shown in Figure 1) by the adaptive greedy algorithm 1 with the maximum number of sparse grid nodes set as $M_a^{max} = 10000$ (in these experiments, we also set the maximum number of EI basis functions $M_n^{max} = 10000$ and RB basis functions $N_{max} = 10000$) and the tolerance of the EI as $\varepsilon_{ei} = 10^{-7}$ and $\varepsilon_{rb} = 10^{-5}$ (at the same level of SG interpolation error, see Figure 1), while the tolerance for SG is set as 10^{-16} to guarantee that the maximum number of sparse grid nodes is reached for all the cases of mesh size. The number of basis functions is shown for different cases in the right part of Figure 2. It is evident from the numerical results that for sufficiently refined mesh in the HiFi PG solution (in the considered example already at $h = 2^{-5}$), the number of RB functions becomes independent of the meshwidth h; specificially, there are 65 RB's for $h = 2^{-5}$, 63 RB's for $h = 2^{-6}$ and 64 RB's for $h = 2^{-7}$. This implies that the parametric solution manifold (more precisely, solutions at the 10000 sparse grid nodes in the parameter domain) is well represented by the small number of RB basis functions, regardless of the Finite Element mesh width employed in the HiFi discretization, *provided it is "small enough", meaning that the HiFi PG discretization resolves all reduced bases contributing to the solution.* The same observation holds for EI compression, where the number of EI basis functions is independent of h for $h = 2^{-4}$ and $h = 2^{-5}$ and for $h = 2^{-6}$ and $h = 2^{-7}$, which is again much smaller than the number of sparse grid nodes 10000. This result demonstrates that the EI and RB approximations in the parametric forward model can allow for a significant reduction of computational work in Bayesian inversion; in our numerical experiments, we solved the HiFi problem only at a small portion ($< 1\%$) of the number of SG interpolation nodes, and solved the RB surrogate at all remaining SG nodes.

6.1.3. Reduced basis compression errors

In this experiment, we first investigate the convergence of the RB compression errors, including RB error and EI error, w.r.t. the number of basis functions, and demonstrate the effectivity of the a-posteriori error estimators. We run the greedy algorithm (lines 10 to 23 in the adaptive greedy algorithm 1) for the construction of the RB-EI compression with 1000 random samples. The tolerances for EI and RB are set as 10^{-8} and 10^{-5}, respectively. We record all the errors and error estimators during the greedy construction, which can be regarded as the test errors and error estimators with the same 1000 random parameter samples, and which we denote as a test set Ξ_{test}.

Figure 3 depicts the errors and error estimators under consideration. On the left, we show the actual RB-EI compression error, denoted Error, the error estimator $\triangle_{dwr}$ and the RB error estimator $\triangle_{rb}$. Both estimators are very close to each other, implying that the error estimators are effective. All three quantities decay with a rate of about $N^{-1.5}$ when the number of basis functions is relatively small, and with a rate of approximately N^{-3} for a larger number of basis functions. The error of the corrected posterior density $\Theta^c_{N,M}$ is smaller than that without correction and decays asymptotically with a rate N^{-3}. Note that this convergence rate is larger than that of the adaptive sparse grid interpolation error N^{-1}. On the right side of Figure 3, the error and error estimator for the EI compression are shown besides the RB-EI compression error. In this example, the EI error and its error estimator are very close, indicating that the error estimator is rather effective in this problem. Moreover, the EI error asymptotically converges with a rate N^{-3}, which is larger than the SG convergence

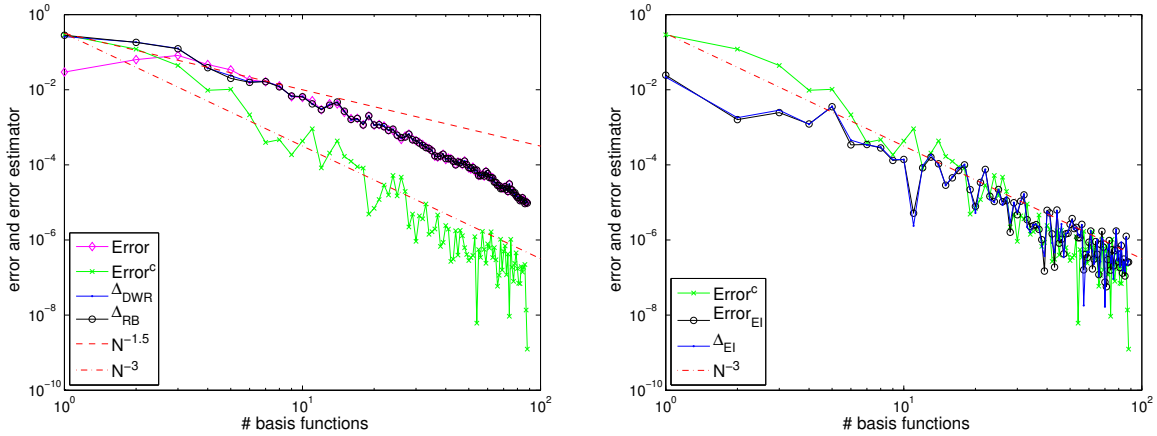

Figure 3: Decay of RB compression errors and error estimators (left) w.r.t. the number of RB basis functions, where $\triangle_{DWR} = \triangle^{\Theta}_{N,M}(\boldsymbol{y}^*)$ is the dual-weighted residual error estimator at $\boldsymbol{y}^* = \text{argmax}_{\boldsymbol{y}\in\Xi_{test}} \triangle^{\Theta}_{N,M}(\boldsymbol{y})$ in (111), Error $= |\Theta_h(\boldsymbol{y}^*) - \Theta_{N,M}(\boldsymbol{y}^*)|$, Error$^c = |\Theta_h(\boldsymbol{y}^*) - \Theta^c_{N,M}(\boldsymbol{y}^*)|$, $\triangle_{RB} = \triangle^{\Theta,RB}_{N,M}(\boldsymbol{y}^*)$ is the RB error estimator in (111); (right) w.r.t. the number of EI basis functions, where Error$_{EI} = |\Theta_N(\boldsymbol{y}^*) - \Theta_{N,M}(\boldsymbol{y}^*)|$, where $\Theta_N(\boldsymbol{y}^*)$ is the RB compression without performing EI evaluated at $\boldsymbol{y}^*$, $\triangle_{EI} = \triangle^{\Theta,EI}_{N,M}(\boldsymbol{y}^*)$ is the EI error estimator in (111).

rate N^{-1} as predicted by Theorem 8.5. Moreover, we observe that the corrected RB-EI error and the EI error are quite close. Therefore, in order to achieve a more accurate, corrected RB-EI compression, the EI should be constructed such that its error is smaller than that of the RB error. This can be realized by choosing the tolerance for EI substantially smaller than that for RB. However, it is difficult to choose the optimal tolerances for RB-EI, which depend on the number of basis functions and the quantity of interest. In the numerical experiments reported here, it was essential to choose the tolerance for EI two to three orders of magnitude smaller than that for RB.

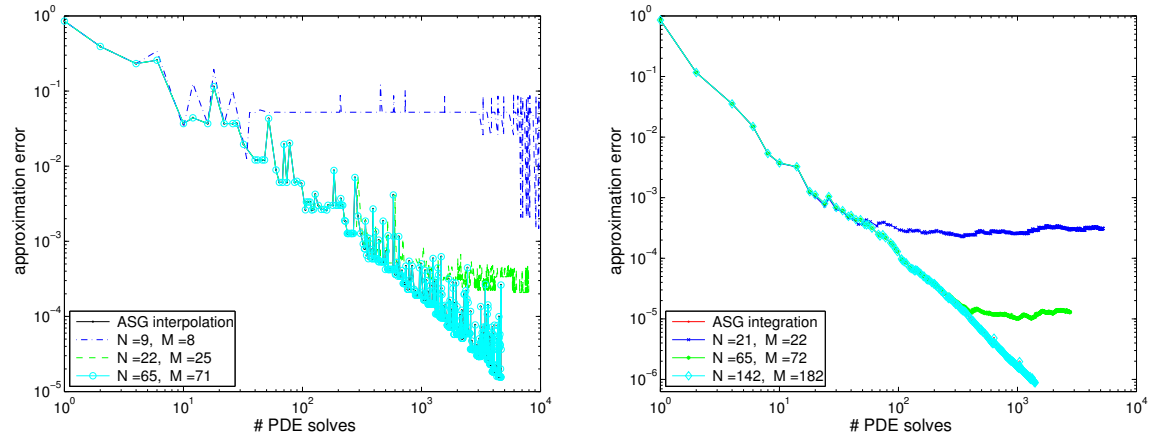

Figure 4: Decay of interpolation (left) and integration (right) error estimators by the adaptive SG and the adaptive SG-RB-EI at different tolerances for the construction of EI and RB by Algorithm 1

In order to see how the tolerances for the EI-RB construction influence the accuracy of the SG interpolation and in particular SG integration, we run the adaptive greedy algorithm 1 with $(\varepsilon_{rb}, \varepsilon_{ei}) = (10^{-2}, 10^{-4})$, $(10^{-3}, 10^{-5})$ and $(10^{-5}, 10^{-7})$ for SG interpolation and $(\varepsilon_{rb}, \varepsilon_{ei}) = (10^{-3}, 10^{-5})$, $(10^{-5}, 10^{-7})$ and $(10^{-7}, 10^{-9})$ for SG integration according to the error in Figure 1. The results are shown in Figure 4. For SG interpolation, only 65 RB basis functions and 71 EI basis functions are sufficient to produce the same interpolation accuracy as that without RB-EI compression, so that only 71 HiFi problems are solved, compared to 10^4. For SG integration, more basis functions are required for both EI (182) and RB (142) in order to guarantee that the same SG integration accuracy around 10^{-8} is maintained. Note that the number of RB and EI basis functions are still dramatically less that the number of HiFi PDE solves required by SG without RB-EI compression.

In the last experiment, we investigate the convergence of SG and RB-EI for parametric problems for various

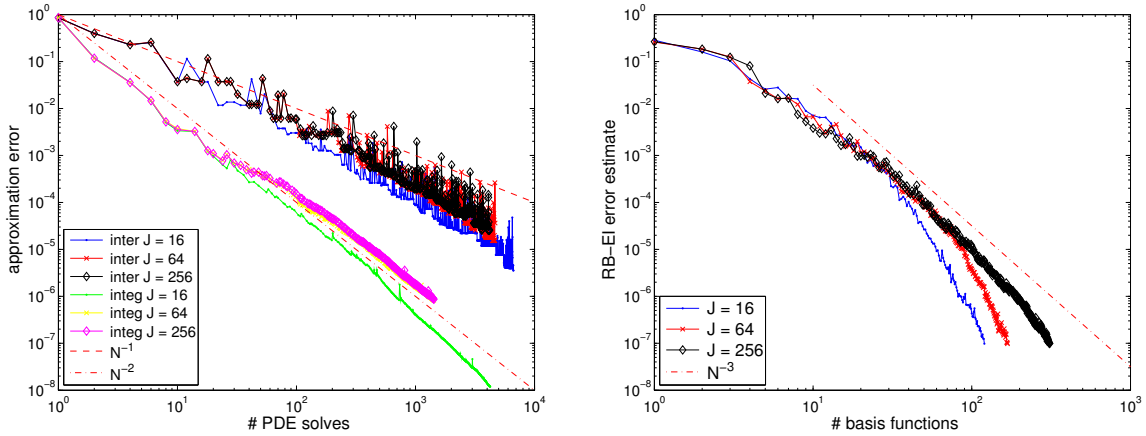

Figure 5: Decay of error estimators of the adaptive SG (left) and the RB-EI compression (right) at different dimensions.

parameter dimensions. Specifically, we consider parameter dimensions $J = 16, 64$ and 256 and run the SG construction for both interpolation and integration as well as the greedy strategy (lines 10 to 24 in Algorithm 1) for RB-EI construction with 1000 random samples. We report the results in Figure 5. The convergence rate of the SG interpolation and integration only slightly increases to the rate N^{-1} and N^{-2}, respectively, as the dimension grows from 16 to 256 confirming that the convergence rate is independent of the dimension of the parameter space. The same observation holds for the convergence rate of the RB-EI compression error. The convergence rate increases to $O(N^{-3})$ as the dimension of the parameter space increases; the order 3 is larger than the convergence rate for either SG interpolation and SG based Smolyak integration. Moreover, when the number of RB basis functions surpasses the number of parameter dimensions, the convergence rate becomes larger than 3, empirically displaying an exponential error decay in terms of N.

6.2. Nonaffine, nonlinear problem

We consider a nonlinear operator equation in a random domain [18], which is non-affine w.r. to the parameter sequence $\boldsymbol{y}$ and nonlinear w.r. to the 'state' (being the domain): given $\boldsymbol{y} \in U$, find $q(\boldsymbol{y}) : D_{u(\boldsymbol{y})} \to \mathbb{R}$ such that

$$-\triangle q(\boldsymbol{y}) + q^3(\boldsymbol{y}) = f \quad \text{in } D_{u(\boldsymbol{y})}, \quad q(\boldsymbol{y}) = 0 \quad \text{on } \partial D_{u(\boldsymbol{y})}, \tag{123}$$

where the random domain $D_{u(\boldsymbol{y})}$ is homeomorphic to the unit disc, and explicitly given by

$$D_{u(\boldsymbol{y})} := \{x = (r\cos(\theta), r\sin(\theta)) : 0 \le r < u(\boldsymbol{y}), 0 \le \theta < 2\pi\}\,. \tag{124}$$

Here, the random radius $u(\boldsymbol{y})$, as defined in (9), is given explicitly by

$$u(0) = \langle u \rangle = 1 \text{ and } \psi_j = \frac{0.5}{j^\alpha}\sin(j\theta) \quad j \ge 1, \text{ where } \alpha > 2\,. \tag{125}$$

Let F_u denote a transformation map from the reference domain $D_{\langle u \rangle}$, the unit disk of $\mathbb{R}^2$ centered at the origin, to the parametric domain D_u, given by $F_u(r\cos(\theta), r\sin(\theta)) := (u(\boldsymbol{y})r\cos(\theta), u(\boldsymbol{y})r\sin(\theta))$. Then the nonlinear operator equation (123) becomes: given $\boldsymbol{y} \in U$, find $q(\boldsymbol{y}) : D_{\langle u \rangle} \to \mathbb{R}$ such that

$$\begin{cases} -\mathrm{div}(M(\boldsymbol{y})\nabla q(\boldsymbol{y})) + q^3(\boldsymbol{y})d(\boldsymbol{y}) &= f d(\boldsymbol{y}) \quad \text{in } D_{\langle u \rangle}, \\ q(\boldsymbol{y}) &= 0 \quad \text{on } \partial D_{\langle u \rangle}, \end{cases} \tag{126}$$

where $d(\boldsymbol{y})$ denotes the determinant of the Jacobian dF_u of the map F_u, given as $d(\boldsymbol{y}) = (u(\boldsymbol{y}))^2$;

$$M(\boldsymbol{y}) := d(\boldsymbol{y}) dF_u^{-1} dF_u^{-\top} = \begin{pmatrix} 1 + (b(\boldsymbol{y}))^2 & -b(\boldsymbol{y}) \\ -b(\boldsymbol{y}) & 1 \end{pmatrix} \quad \text{where } b(\boldsymbol{y}) := \frac{\partial_\theta u(\boldsymbol{y})}{u(\boldsymbol{y})}\,. \tag{127}$$

This (highly non-affine parametric) problem also fits into the abstract setting (1) with the choices $\mathcal{X} = \mathcal{Y} = H_0^1(D_{\langle u \rangle})$. We assume that the $K = 9$ observation data are given as in (121) with $x^{(k)} = (0.5\cos(2\pi k/K), 0.5\sin(2\pi k/K))$ and with $d_k = 0.1$, $k = 1, \ldots, K$. The observation noise is as in Section 6.1.

6.2.1. Sparse grid approximation error

In this test, we generate a regular, triangle mesh with 14060 vertices and use continuous, piecewise linear P2 elements on this mesh for the HiFi discretization. Scaling parameters $\alpha = 3, 4$ are used in the adaptive construction of the sparse grid approximation (interpolation and integration) of the HiFi density. We truncate the parameter domain at dimension $J = 64$. The convergence results are shown in Figure 6. As the convergence rates for the interpolation error is approximately $\alpha - 2$ (M^{-1} when $\alpha = 3$ and M^{-2} when $\alpha = 4$) and for the integration error is approximately $\alpha - 1$ (M^{-2} when $\alpha = 3$ and M^{-3} when $\alpha = 4$), which is one order lower compared to the rates $\alpha - 1$ for interpolation and α for integration in the last section. This is mainly due to the first order derivative with respect to θ being involved in the matrix $M(\boldsymbol{y})$ in (127) through $b(\boldsymbol{y})$; this leads to a less sparse problem as the series of $\partial_\theta u(\boldsymbol{y})$, more explicitly $\partial_\theta \psi_j(\theta) = \cos(j\theta)/2j^{\alpha-1}$, such that $\{||\partial_\theta \psi_j(\theta)||_{L^\infty((0,2\pi))}\}_{j\geq 1} \in \ell^p$ for $p > 1/(\alpha-1)$ instead of $p > 1/\alpha$ as for the example in the last section, which further leads to the dimension-independent convergence rate M^{-s} in Theorem 2.6, for $s = 1/p - 1 < \alpha - 2$. Moreover, it is also shown in Figure 6 that for interpolation (respectively integration), the interpolation (resp. integration) error estimator gives rise to more accurate approximation than the integration (resp. interpolation) error estimator.

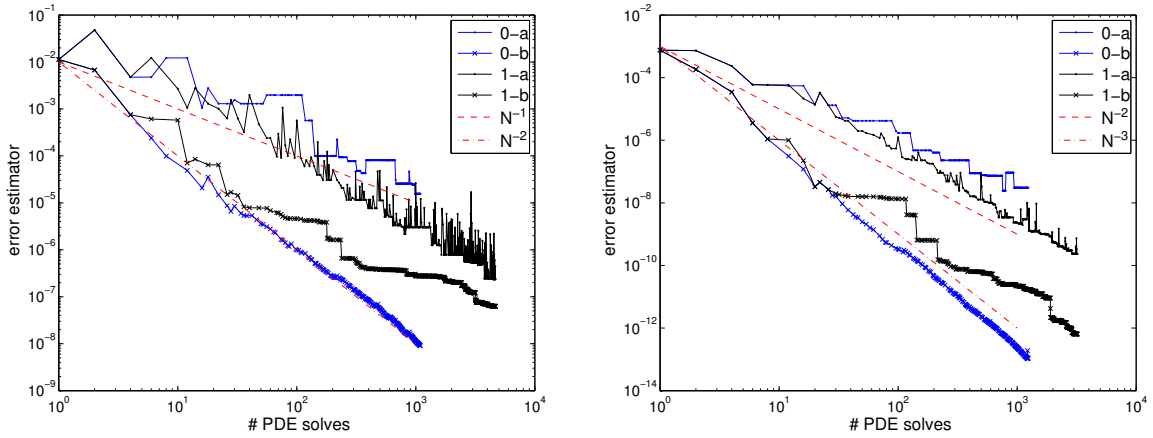

Figure 6: Interpolation (a) and integration (b) error estimates w.r.t. the number of PDE solves (the same as the number of sparse grid nodes) with the sparse grid constructed by integration (0) and interpolation (1) error indicators; left: $\alpha = 3$; right: $\alpha = 4$.

6.2.2. Reduced basis compression error

In this experiment, we test the compression error from RB and EI, respectively, and demonstrate the SG-RB-EI algorithm for the high dimensional integration in Bayesian inversion. At first, we run the greedy algorithm for the construction of RB space at $J = 64$ dimensions with the goal-oriented a posteriori error estimator $\triangle_N^\Theta$, defined as

$$\triangle_N^\Theta(\boldsymbol{y}) := -_{\mathcal{Y}'}\langle (\mathcal{R}(q_N(\boldsymbol{y}); \boldsymbol{y})), \psi_N(\boldsymbol{y})\rangle_{\mathcal{Y}} , \tag{128}$$

where $q_N(\boldsymbol{y})$ and $\psi_N(\boldsymbol{y})$ are the solutions of the primal and dual RB-PG compression problems. We use a training set Ξ_{train} of 1000 random samples for the construction. The RB compression error and error estimators are shown in Figure 7, where we can see that the error estimator and the error (compared to HiFi Θ_h) can not be distinguished, which implies that the goal-oriented RB error estimator is very effective. Both the error and the error estimator converge with rate $N^{-(\alpha-1)}$ (N^{-2} for $\alpha = 3$ on the left part of Figure 7 and N^{-3} for $\alpha = 4$ on the right part of Figure 7), which is as that for the SG integration error as shown in Figure 6. Moreover, we can observe that the RB density with correction, i.e. $\Theta_N^c(\boldsymbol{y}) = \Theta_N(\boldsymbol{y}) + \triangle_N^\Theta(\boldsymbol{y})$, achieves much smaller error compared to the RB density without correction $\Theta_N(\boldsymbol{y}) = \Theta(q_N(\boldsymbol{y}))$, and displays asymptotically a convergence rate of $N^{-2(\alpha-1)}$. We remark that this much faster convergence attributes to that the error estimator $\triangle_N^\Theta(\boldsymbol{y})$ is very effective as a surrogate of the RB error $|\Theta_h(\boldsymbol{y}) - \Theta_N(\boldsymbol{y})|$.

In order to see the interpolatory property of the EI for the residual $\mathcal{R}$, as well as for the linear term A, the nonlinear term B and the forcing term F, we first solve the HiFi-PG problem (58) and collect the residual $\mathcal{R}(q_h^{(k)}(\boldsymbol{y}); \boldsymbol{y})$ ($1 \leq k \leq K(\boldsymbol{y})$), respectively $A(\boldsymbol{y})$, $B(q_h(\boldsymbol{y}); \boldsymbol{y})$ (only collected at the final solution $q_h(\boldsymbol{y}) = q_h^{K(\boldsymbol{y})}(\boldsymbol{y})$) and $F(\boldsymbol{y})$, at a training set Ξ_{train} of 1000 random samples. Then we perform the greedy algorithm (steps (84) and (85)) for the construction of the EI to approximate the collected quantities. The interpolation errors are shown in Figure 8, which displays a convergence rate of $M^{-(\alpha-1)}$ (M^{-2} when $\alpha = 3$ in

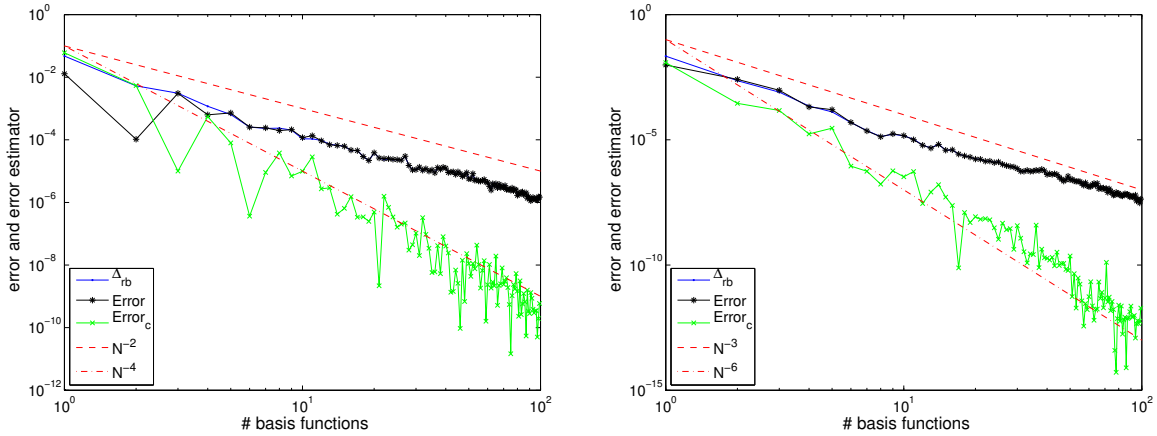

Figure 7: Convergence of the RB error $\text{Error} = \max_{\boldsymbol{y}\in\Xi_{train}} |\Theta_h(\boldsymbol{y}) - \Theta_N(\boldsymbol{y})|$, RB error estimator $\triangle_{RB} = \max_{\boldsymbol{y}\in\Xi_{train}} \triangle_N^{\Theta}(\boldsymbol{y})$, and the RB error for the corrected RB density $\text{Error}_c = \max_{\boldsymbol{y}\in\Xi_{train}} |\Theta_h(\boldsymbol{y}) - \Theta_N^c(\boldsymbol{y})|$. Left: $\alpha = 3$; right: $\alpha = 4$.

the left part and M^{-3} for $\alpha = 4$ in the right part of Figure 8) for the residual term $\mathcal{R}$, as well as for A, B and F. Note that as $A(\boldsymbol{y})$ and $F(\boldsymbol{y})$ do not depend on the solution $q_h(\boldsymbol{y})$, a much faster decay is observed for the interpolation errors of $A(\boldsymbol{y})$ after 128 terms and of $F(\boldsymbol{y})$ after 64 terms, which are twice as and equal to the number of dimensions $J = 64$. The interpolation errors of the residual $\mathcal{R}(q_h^{(k)})$ and in particular the nonlinear term $B(q_h(\boldsymbol{y}); \boldsymbol{y})$, which depend on the solutions, decay with a consistent convergence rate. We remark that the convergence rate of the EI interpolation error is identical to the rate for SG integration and larger than the SG interpolation, see Figure 6. We remark further that the EI error is measured as the maximum of the approximation of the residual for all the random samples at all finite element nodes, namely pointwise worst-case scenario error.

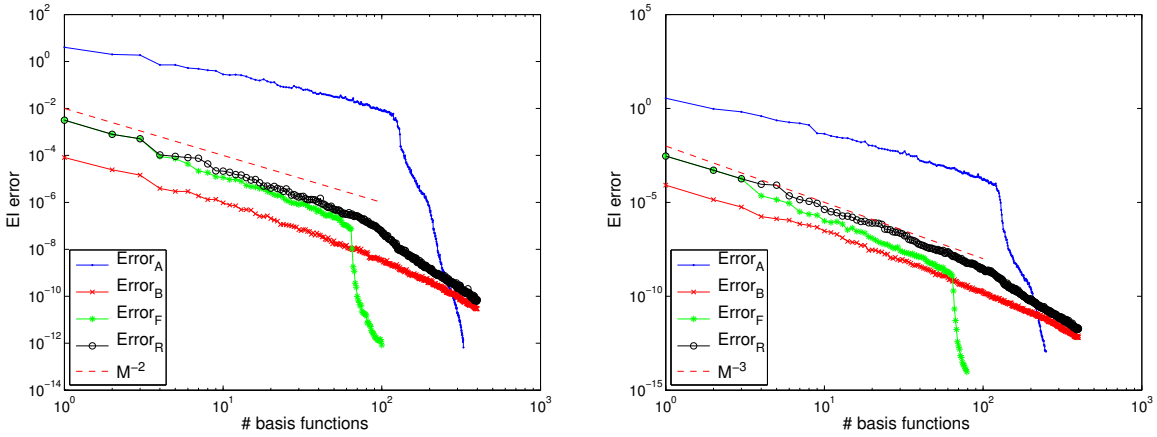

Figure 8: Convergence of the EI error $\text{Error}_A = \max_{\boldsymbol{y}\in\Xi_{train},\sigma\in\sigma_{set}} |\sigma(A(\boldsymbol{y}) - \mathcal{J}_{M_A}^A A(\boldsymbol{y}))|$, the same definition for Error_B and Error_F, and $\text{Error}_R = \max_{k,\boldsymbol{y}\in\Xi_{train},\sigma\in\sigma_{set}} |\sigma(\mathcal{R}(q_h^{(k)}(\boldsymbol{y}); \boldsymbol{y}) - \mathcal{J}_M \mathcal{R}(q_h^{(k)}(\boldsymbol{y}); \boldsymbol{y}))|$. Left: $\alpha = 3$; right: $\alpha = 4$.

Next, we run the greedy algorithm 1 for the construction of RB-EI compression based on a training set Ξ_{train} of 1000 random samples and using the goal-oriented a posteriori error estimator (111). Two types of the EI are applied, the first is to interpolation the whole residual term $\mathcal{R}$ and the second is to interpolate the linear term A, nonlinear term B and constant term F separately as in (88). The comparison of the convergence of the compression errors and error estimators is shown in Figure 9. It is evident that the uniform interpolation leads to rather ineffective a posteriori error estimate, displaying large oscillation. On the other hand, the error estimator $\triangle_{DWR} = \triangle_{N,M}^{\Theta}$ with separate interpolation of each term A, B, and F provides a good estimate for the RB-EI compression error, even if it is not as effective as the estimator $\triangle_N^{\Theta}$ without the EI compression of the residual shown in Figure 7. The same observation holds for the RB error and for its error estimator in the right part of Figure 9 as well as for the EI error and its estimator, in the left part of Figure 10. However,

when we construct a very fine EI at the initial step by using 200 training samples, the error estimator $\triangle_{DWR}$ provides a rather effective estimate of the RB-EI compression error, as shown in the right part of Figure 10, so that we can correct the RB-EI posterior density with the error estimator, leading to smaller compression error Error^c as displayed in Figure 10. The RB and the EI errors and their estimators decay asymptotically with a rate $N^{-(\alpha-1)}$ for $\alpha = 3$, the same as that of SG integration error.

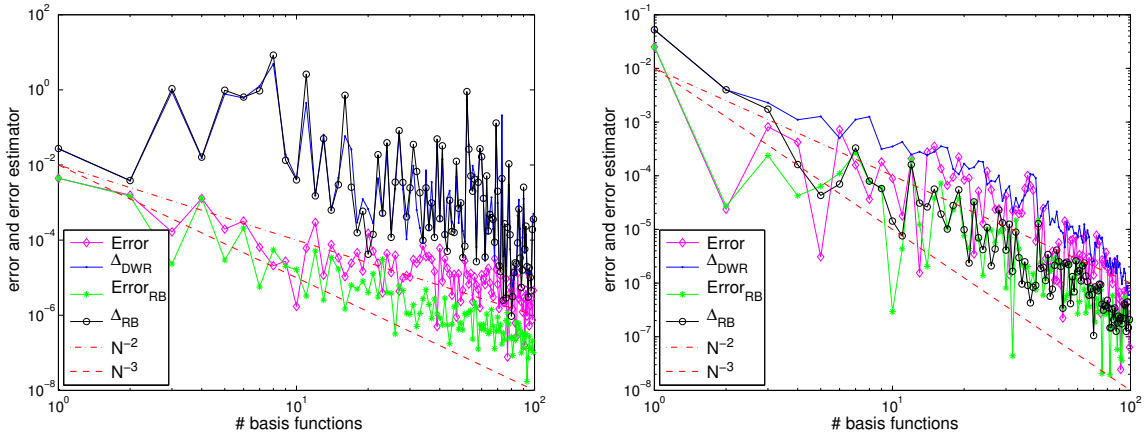

Figure 9: Convergence of the RB compression error $\text{Error} = \max_{\boldsymbol{y}\in\Xi_{train}} |\Theta_h(\boldsymbol{y}) - \Theta_{N,M}(\boldsymbol{y})|$, the error estimator $\triangle_{DWR} = \max_{\boldsymbol{y}\in\Xi_{train}} \triangle^{\Theta}_{N,M}(\boldsymbol{y})$; the RB error $\text{Error}_{RB} = \max_{\boldsymbol{y}\in\Xi_{train}} |\Theta_h(\boldsymbol{y}) - \Theta_N(\boldsymbol{y})|$ and error estimator $\triangle_{EI} = \max_{\boldsymbol{y}\in\Xi_{train}} \triangle^{\Theta,RB}_{N,M}(\boldsymbol{y})$. Left: using EI directly for the residual $\mathcal{R}$; right: using EI separately for A, B, and F.

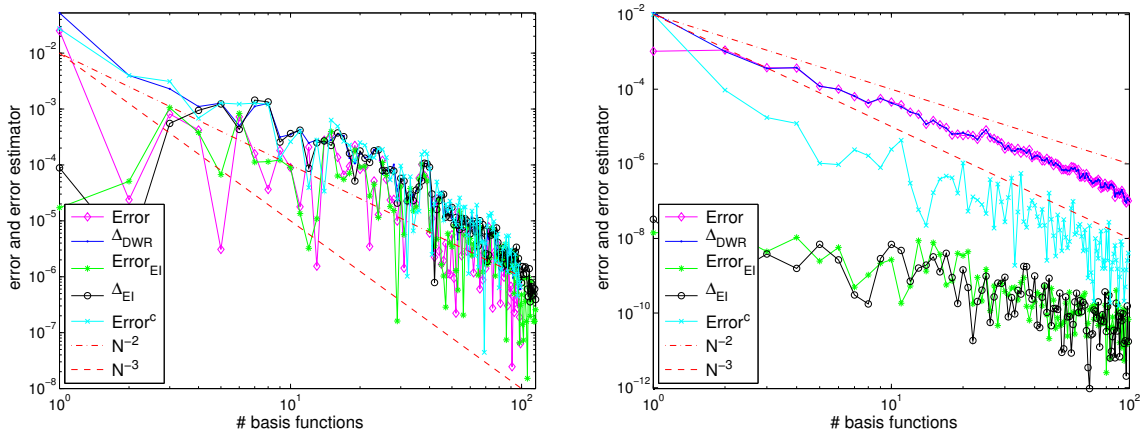

Figure 10: Convergence of the RB-EI compression error $\text{Error} = \max_{\boldsymbol{y}\in\Xi_{train}} |\Theta_h(\boldsymbol{y}) - \Theta_{N,M}(\boldsymbol{y})|$, the error estimator $\triangle_{DWR} = \max_{\boldsymbol{y}\in\Xi_{train}} \triangle^{\Theta}_{N,M}(\boldsymbol{y})$; the EI error $\text{Error}_{EI} = \max_{\boldsymbol{y}\in\Xi_{train}} |\Theta_N(\boldsymbol{y}) - \Theta_{N,M}(\boldsymbol{y})|$ and error estimator $\triangle_{EI} = \max_{\boldsymbol{y}\in\Xi_{train}} \triangle^{\Theta,EI}_{N,M}(\boldsymbol{y})$. Left: initial training size for EI interpolation $M_0 = 20$; right: $M_0 = 200$.

In the last test, we run the full greedy algorithm for the SG-RB-EI construction and the evaluation of the integration of the density Θ. We set the number of maximal SG nodes as 10000 and the SG tolerance as 10^{-16}. Different tolerances, 10^{-4}, 10^{-6}, 10^{-8} are set in the termination criteria for the relative error of the RB-EI construction with separate EI for the interpolation of A, B and F. The corresponding convergence of the integration errors are shown in Figure 11. The decay is locked earlier for higher tolerance, which indicates that as more SG nodes are used for the integration, a lower tolerance should be used for the RB-EI construction. At tolerance 10^{-8}, the RB and the EI error estimators at each of the 10000 SG nodes are displayed in the left and right parts of Figure 11, respectively. We observe that in this example, the two error estimators decay asymptotically with a rate of $M^{-(\alpha-1)}$. As for the EI, the decay rate is superior to this rate when the number of EI functions exceeds 128, which is about twice the truncation dimension $J = 64$ as also observed in Figure 8 for the linear and constant terms. The total numbers of RB functions are 45, 256, 708 at the three different tolerances, which is approximately the same as the number of SG nodes for integration with the same level

of accuracy. The corresponding numbers of EI functions are $(M_A, M_B, M_F) = (82, 45, 37)$, $(194, 149, 70)$, and $(303, 314, 86)$ for the three tolerances.

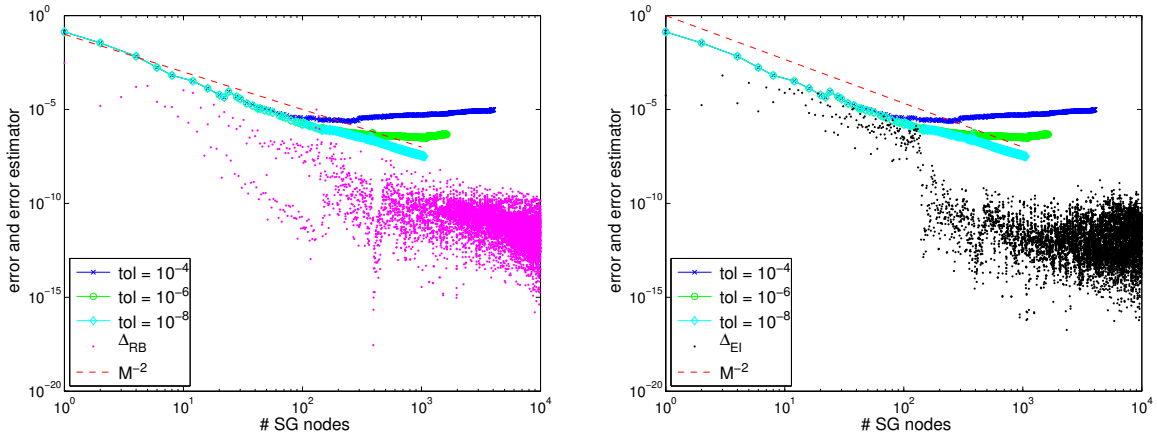

Figure 11: Decay of the error for the integration of the density Θ, by the SG-RB-EI approximation with different tolerances (for relative error) for RB-EI construction, tol $= 10^{-4}, 10^{-6}$ and 10^{-8}. The RB error estimator $\triangle_{N,M}^{\Theta,RB}$ (left) and the EI error estimator $\triangle_{N,M}^{\Theta,EI}$ (right) at the 10000 SG nodes are shown for the case tol $= 10^{-8}$.

In order to test the RB-EI compression for Bayesian inversion for new observation data, we randomly pick a new input $\boldsymbol{y} \in U$ and compute a noisy observation δ. We then run the SG-RB-EI algorithm for the evaluation of the integration of the posterior density, where the RB-EI functions are fixed as those constructed at tolerance 10^{-8} in the offline construction. The convergence of the SG integration error estimator, and of the RB and the EI error estimators are shown in Figure 12. We observe that the RB-EI errors remain very small over all SG nodes and the SG integration error estimator decays without saturation, which indicates that the RB-EI surrogates constructed offline are accurate for online evaluation. The advantage of the RB-EI compression is that, given new data, we only need to solve the RB-EI compression problem instead of the HiFi approximation problem in order to evaluate the quantities of interest by SG integration. This largely reduces the computational cost whenever the HiFi solution of the forward problem is computationally costly. We remark that the RB-EI minimizes the worst-case scenario error while the error for SG integration, being an integral quantity, is measured in the "average" sense.

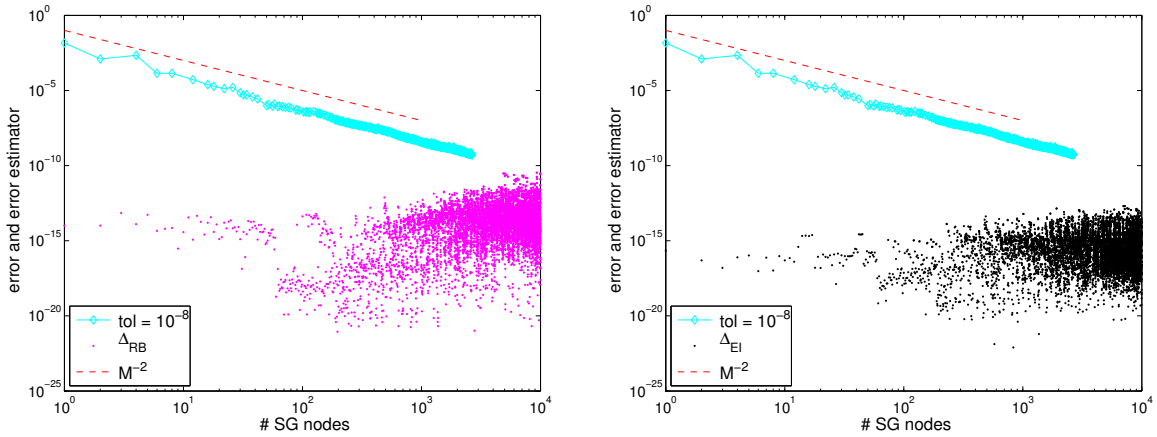

Figure 12: Decay of the error for the SG integration of the density Θ, with the density at each SG node evaluated by RB-EI approximation, which has been constructed with tolerance tol $= 10^{-8}$ offline. Shown are also the RB error estimator $\triangle_{N,M}^{\Theta,RB}$ (left) and the EI error estimator $\triangle_{N,M}^{\Theta,EI}$ (right) at the 10000 SG nodes.

7. Conclusions

We extended our work [15] on the RB acceleration of the deterministic quadrature approach from [41, 42] for computational Bayesian inversion of linear, affine-parametric operator equations with distributed, uncertain input data to nonlinear operator equations with possibly non-affine parametric inputs. The generalization is based on methods for RB acceleration for forward solves of the parametric problem with uncertain input data; specifically, a nodal version of EIM due to [1]. Based on sparsity results of the countably-parametric, deterministic Bayesian posterior densities, their N-widths are known to be small (cf. [41, 42]), being bounded by approximation errors of N-term truncated gpc expansions. To ease the presentation, we considered uniform prior $\boldsymbol{\pi}_0(d\boldsymbol{y})$. We add that the present results extend to "informed prior measures" $\boldsymbol{\pi}_0$ which admit a $(\boldsymbol{b}, p)$-holomorphic density $\rho(\boldsymbol{y})$ with respect to the uniform measure. In the present work we proposed algorithms for construction of parsimonious surrogate maps of the parametric forward solution, as well as for the parametric Bayesian posterior density. The construction of the RB surrogates of the parametric forward maps is effected by a greedy search which, as we showed, can be performed offline, before the actual Bayesian estimation step. In particular, before assimilating observation data for the quadrature approximation of the Bayesian expectation, conditional on observation data. In all numerical experiments, for nonlinear problems on uncertain domains of definition, convergence rates which are independent of the dimension of the set of active parameters and at least as good as the N-term approximations rates were achieved. The possibly large convergence rates of adaptive Smolyak quadrature approximation of the Bayesian estimates reported in [41, 42] were realized also here in all computed examples. The required online CPU time, however, was substantially reduced as compared to [41, 42]. We remark that the presently proposed approach of generating offline a parsimonious RB surrogate of the parametric forward map prior to the evaluation of the Bayesian estimate will also allow to accelerate other Bayesian estimation methods, such as MCMC methods; we refer to [29] for an error analysis of such acceleration methods.

8. Appendix: A-Priori Error estimates

We provide a-priori error estimates of each of the approximations in the proposed algorithms. We account in particular for dimension truncation of the parameter domains, HiFi-PG approximation, model order reduction including error contributions from the RB compression and the EI. A combined error estimate is also proved for the approximation of the Bayesian posterior density and the related quantities of interest.

8.1. Dimension truncation

For a *truncation dimension* $J \in \mathbb{N}$, denote the J-term truncation of parametric representation (9) of the uncertain datum u by $u_J \in X$. Dimension truncation is equivalent to setting $y_j = 0$ for $j > J$ in (9) and we denote by $q_J(\boldsymbol{y})$ the solution of the corresponding parametric weak problem (13). Unique solvability of (13) implies $q_J(\boldsymbol{y}) = q(\{y_1, y_2, ..., y_J, 0, ...\})$. For $\boldsymbol{y} \in U$, define $\boldsymbol{y}_{\{1:J\}} := (y_1, y_2, ..., y_J, 0, 0, ...)$. Proposition 2.1 holds when $u(\boldsymbol{y})$ is replaced by $u_J(\boldsymbol{y})$, with $\kappa > 0$ in (5) independent of J for sufficiently large J.

Our estimation of the *dimension truncation error* $q(\boldsymbol{y}) - q_J(\boldsymbol{y})$ relies on two assumptions.

Assumption 4. *(i) We assume the p-summability* (11) *of the sequence* $\boldsymbol{b}$ *given by* $b_j := \|\psi_j\|_X$ *in* (9)*. From the definition of the sequence* $\boldsymbol{b} = (b_j)_{j\geq 1}$ *in* (11)*, the condition is equivalent to* $\sum_{j\geq 1} b_j^p < \infty$*. (ii) the* b_j *in* (11) *are enumerated so that*

$$b_1 \geq b_2 \geq \cdots \geq b_j \geq \cdots . \tag{129}$$

Consider the J-term truncated problem: given $u_J \in \tilde{X}$,

$$\text{find } q_J \in \mathcal{X} : \quad {}_{\mathcal{Y}'}\langle \mathcal{R}(q_J; u_J), w\rangle_{\mathcal{Y}} = 0 \ \ \forall w \in \mathcal{Y} . \tag{130}$$

Proposition 8.1. *Under assumptions* (10), (11)*, for every* $F \in \mathcal{Y}'$*, for every* $\boldsymbol{y} \in U$ *and for every* $J \in \mathbb{N}$*, the parametric solution* $q_J(\boldsymbol{y})$ *of the dimensionally truncated, parametric weak problem* (13) *with* J*-term truncated parametric expansion* (9) *satisfies, with* b_j *as defined in* (11)*,*

$$\sup_{\boldsymbol{y}\in U} \|q(\boldsymbol{y}) - q_J(\boldsymbol{y})\|_{\mathcal{X}} \ \leq \ C(F, X) \sum_{j\geq J+1} b_j \tag{131}$$

for some constant $C > 0$ independent of J. The same bound (with different constants independent of J) holds for the approximation of the posterior density Θ and QoI Ψ as well as for their integration. In addition, if conditions (10), (11) *and* (129) *hold, then*

$$\sum_{j\geq J+1} b_j \;\leq\; \min\left(\frac{1}{1/p-1},1\right)\left(\sum_{j\geq 1} b_j^p\right)^{1/p} J^{-s}, \quad s=\frac{1}{p}-1\;. \tag{132}$$

8.2. High-fidelity PG Approximation

To establish the well-posedness of the HiFi-PG approximation problem (58) as well as the a-priori and a-posteriori error estimates for the approximate solution q_h, we impose classical assumptions from nonlinear Finite Element analysis, as e.g. in [37].

Assumption 5. *Let $a(\cdot,\cdot;\boldsymbol{y}) : \mathcal{X}\times\mathcal{Y}\to\mathbb{R}$ denote the parametric bilinear form for each $\boldsymbol{y}\in U$ associated with the Fréchet derivative of $\mathcal{R}$ at q, i.e.*

$$a(w,v;\boldsymbol{y}) :=_{\mathcal{Y}'} \langle D_q\mathcal{R}(q(\boldsymbol{y});\boldsymbol{y})(w),v\rangle_{\mathcal{Y}} \quad \forall w\in\mathcal{X},\forall v\in\mathcal{Y}\;. \tag{133}$$

We assume the following conditions hold

A1 stability: *the parametric bilinear form a satisfies the discrete PG inf-sup condition*

$$\forall \boldsymbol{y}\in U: \quad \inf_{0\neq w_h\in\mathcal{X}_h}\sup_{0\neq v_h\in\mathcal{Y}_h}\frac{a(w_h,v_h;\boldsymbol{y})}{||w_h||_{\mathcal{X}}||v_h||_{\mathcal{Y}}} =: \beta_h(\boldsymbol{y})\geq\beta_h>0\;, \tag{134}$$

where the inf-sup constant $\beta_h(\boldsymbol{y})$ depends on h and on $\boldsymbol{y}$ and may vanish $\beta_h(\boldsymbol{y})\to 0$ as $h\to 0$.

A2 consistency: *the best approximation satisfies the consistent approximation property*

$$\forall \boldsymbol{y}\in U: \quad \lim_{h\to 0}\frac{1}{\beta_h^2(\boldsymbol{y})}\inf_{w_h\in\mathcal{X}_h}\|q(\boldsymbol{y})-w_h\|_{\mathcal{X}}=0\;. \tag{135}$$

In view of the convergence rate in (57), (135) *amounts to require $h^s/\beta_h^2(\boldsymbol{y})\to 0$ as $h\to 0$.*

A3 local Lipschitz continuity: *there exists ϵ_0 and $L>0$ such that for all $w\in\mathcal{X}$ with $||q(\boldsymbol{y})-w||_{\mathcal{X}}\leq\epsilon_0$, there holds*

$$\forall \boldsymbol{y}\in U: \quad \|D_q\mathcal{R}(q(\boldsymbol{y});\boldsymbol{y})-D_q\mathcal{R}(w;\boldsymbol{y})\|_{\mathcal{L}(\mathcal{X},\mathcal{Y}')}\leq L||q(\boldsymbol{y})-w||_{\mathcal{X}}\;. \tag{136}$$

Assumption 5 is sufficient to guarantee the existence of a solution $q_h(\boldsymbol{y})\in\mathcal{X}_h$ of the HiFi-PG approximation problem (58) for any $\boldsymbol{y}\in U$, which is locally unique and satisfies a-priori error estimate. We present the results in the following theorem, whose proof follows that in [37].

Theorem 8.2. *Under Assumption 5, there exists $h_0>0$ and $\eta_0>0$ such that for $0<h\leq h_0$, there exists a solution $q_h(\boldsymbol{y})\in\mathcal{X}_h$ of the HiFi-PG approximation problem* (58)*, which is unique in $\mathcal{B}_{\mathcal{X}}(q(\boldsymbol{y});\eta_0\beta_h(\boldsymbol{y}))$. Moreover, for $0<h\leq h_0$, there holds the a-priori error estimate*

$$||q(\boldsymbol{y})-q_h(\boldsymbol{y})||_{\mathcal{X}}\leq 2\frac{||a(\boldsymbol{y})||}{\beta(\boldsymbol{y})}\left(1+\frac{||a(\boldsymbol{y})||}{\beta_h(\boldsymbol{y})}\right)\inf_{w_h\in\mathcal{X}_h}||q(\boldsymbol{y})-w_h||_{\mathcal{X}}\;, \tag{137}$$

where $||a(\boldsymbol{y})|| := ||D_q\mathcal{R}(q(\boldsymbol{y});\boldsymbol{y})||_{\mathcal{L}(\mathcal{X},\mathcal{Y}')}$. Depending on the smoothness parameter $s>0$ (see (57)*) and the polynomial degree $r\geq 1$ of the Finite Element space, we have*

$$\inf_{w_h\in\mathcal{X}_h}||q(\boldsymbol{y})-w_h||_{\mathcal{X}}\leq Ch^k||q(\boldsymbol{y})||_{\mathcal{X}^s},\quad k=\min\{s,r\}\;, \tag{138}$$

where C is independent of the mesh size h and uniformly bounded w.r.t. $\boldsymbol{y}$. Moreover, we have the a-posteriori error estimate

$$||q(\boldsymbol{y})-q_h(\boldsymbol{y})||_{\mathcal{X}}\leq\frac{4}{\beta(\boldsymbol{y})}||\mathcal{R}(q_h(\boldsymbol{y});\boldsymbol{y})||_{\mathcal{Y}'}\;. \tag{139}$$

Remark 8.1. *The same convergence rate as in* (137) *can be obtained for the HiFi approximation of the observation functional $\mathcal{O}_h(\cdot)$ evaluated at q_h (since it is linear and bounded) as well as for the posterior density Θ_h defined in* (66)*. In fact, by the Aubin-Nitsche duality argument, a larger convergence rate is expected for HiFi approximation of $\mathcal{O}_h(q_h)$, as demonstrated in Sec. 6.1.2.*

8.3. Reduced Basis Compression

We observe that the compression error for the RB-EI solution $q_{N,M}(\boldsymbol{y})$, at any $\boldsymbol{y} \in U$, can be split as

$$||q_h(\boldsymbol{y}) - q_{N,M}(\boldsymbol{y})||_{\mathcal{X}} \leq ||q_h(\boldsymbol{y}) - q_N(\boldsymbol{y})||_{\mathcal{X}} + ||q_N(\boldsymbol{y}) - q_{N,M}(\boldsymbol{y})||_{\mathcal{X}} \ , \tag{140}$$

where the first term is due to the RB compression and the second term is due to the EI compression of the residual. The same splitting holds for the compression error for the RB-EI dual solution $\psi_{N,M}(\boldsymbol{y})$. To bound the first term, we establish the optimal approximation property of the RB-PG compression problem (67).

Theorem 8.3. *Under Assumption 5, there exist $N_0 > 0$ and $\eta_0 > 0$ such that for $N \geq N_0$, there exists a solution $q_N(\boldsymbol{y}) \in \mathcal{X}_N$ of the RB-PG compression problem (67) which is unique in $\mathcal{B}(q_N(\boldsymbol{y}); \eta_0 \beta_N(\boldsymbol{y}))$. Moreover, for $N \geq N_0$, there holds the (uniform w.r. to $\boldsymbol{y}$) a-priori error estimate*

$$||q_h(\boldsymbol{y}) - q_N(\boldsymbol{y})||_{\mathcal{X}} \leq 2\frac{||a_N||}{\beta(\boldsymbol{y})} \left(1 + \frac{||a_N||}{\beta_N(\boldsymbol{y})}\right) \inf_{w_N \in \mathcal{X}_N} ||q_h(\boldsymbol{y}) - w_N||_{\mathcal{X}} \ , \tag{141}$$

where the bilinear form $a_N : \mathcal{X}_N \times \mathcal{Y}_N \to \mathbb{R}$ is defined at the RB-PG solution $q_N(\boldsymbol{y})$ as

$$a_N(w_N, v_N; \boldsymbol{y}) :=_{\mathcal{Y}'} \langle D_q \mathcal{R}(q_N(\boldsymbol{y}); \boldsymbol{y})(w_N), v_N \rangle_{\mathcal{Y}} \quad \forall w_N \in \mathcal{X}_N, \forall v_N \in \mathcal{Y}_N \ , \tag{142}$$

which is equipped with the norm $||a_N|| := ||D_q \mathcal{R}(q_N(\boldsymbol{y}); \boldsymbol{y})||_{\mathcal{L}(\mathcal{X}, \mathcal{Y}')}$. Moreover, there holds the a-posteriori error estimate

$$||q_h(\boldsymbol{y}) - q_N(\boldsymbol{y})||_{\mathcal{X}} \leq \frac{4}{\beta_h(\boldsymbol{y})} ||\mathcal{R}(q_N(\boldsymbol{y}); \boldsymbol{y})||_{\mathcal{Y}'} \ . \tag{143}$$

Proof We verify the assumptions 5 in the RB spaces $\mathcal{X}_N$ and $\mathcal{Y}_N$, in particular **A1 stability** and **A2 consistency** since **A3 Lipschitz continuity** is valid for $q_h(\boldsymbol{y})$ with constant $\epsilon_0/2$ by the triangle inequality with the fact that $||q(\boldsymbol{y}) - q_h(\boldsymbol{y})||_{\mathcal{X}} \leq \eta_0 \beta_h(\boldsymbol{y}) \leq \epsilon_0/2$. The rest of the proof follows the same as that for the a-priori error estimate of the HiFi-PG approximation error as in Theorem 8.2.

As the sequence of finite dimensional RB subspaces $\mathcal{X}_1 \subset \mathcal{X}_2 \subset \cdots \subset \mathcal{X}_N \subset \mathcal{X}_h$, and $\mathcal{X}_N = \mathcal{X}_h$ when $N = N_h$, moreover the solution manifold $\mathcal{M}_h := \{q_h(\boldsymbol{y}) \in \mathcal{X}_h : \boldsymbol{y} \in U\}$ is compact, we have that there exists N_0 such that for any $N \geq N_0$, there holds $||q_h(\boldsymbol{y}) - q_N(\boldsymbol{y})||_{\mathcal{X}} \leq \epsilon_0/2$, so that the **A3 Lipschitz continuity** in Assumption 5 holds at $q_h(\boldsymbol{y})$ for $w = q_N(\boldsymbol{y}) \in \mathcal{X}$. By the construction of the RB test functions with the supremizer approach in (75), we have the stability estimate as in (76)

$$\beta_N(\boldsymbol{y}) := \inf_{0 \neq w_N \in \mathcal{X}_N} \sup_{0 \neq v_N \in \mathcal{Y}_N} \frac{a_N(w_N, v_N; \boldsymbol{y})}{||w_N||_{\mathcal{X}} ||v_N||_{\mathcal{Y}}} \geq \inf_{0 \neq w_h \in \mathcal{X}_h} \sup_{0 \neq v_h \in \mathcal{Y}_h} \frac{_{\mathcal{Y}'}\langle D_q \mathcal{R}(q_N(\boldsymbol{y}); \boldsymbol{y})(w_h), v_h \rangle_{\mathcal{Y}}}{||w_h||_{\mathcal{X}} ||v_h||_{\mathcal{Y}}} =: \beta_h^N(\boldsymbol{y}) \ . \tag{144}$$

Moreover, by **A1 stability** in Assumption 5, for any $0 \neq w_h \in \mathcal{X}_h$ we have

$$\begin{aligned} \beta_h(\boldsymbol{y}) ||w_h||_{\mathcal{X}} &\leq \sup_{0 \neq v_h \in \mathcal{Y}_h} \frac{_{\mathcal{Y}'}\langle D_q \mathcal{R}(q_h(\boldsymbol{y}); \boldsymbol{y})(w_h), v_h \rangle_{\mathcal{Y}}}{||v_h||_{\mathcal{Y}}} \\ &= \sup_{0 \neq v_h \in \mathcal{Y}_h} \left(\frac{_{\mathcal{Y}'}\langle (D_q \mathcal{R}(q_h(\boldsymbol{y}); \boldsymbol{y}) - D_q \mathcal{R}(q_N(\boldsymbol{y}); \boldsymbol{y}))(w_h), v_h \rangle_{\mathcal{Y}}}{||v_h||_{\mathcal{Y}}} + \frac{_{\mathcal{Y}'}\langle D_q \mathcal{R}(q_N(\boldsymbol{y}); \boldsymbol{y})(w_h), v_h \rangle_{\mathcal{Y}}}{||v_h||_{\mathcal{Y}}} \right) \\ &\leq L ||q_h(\boldsymbol{y}) - q_N(\boldsymbol{y})||_{\mathcal{X}} ||w_h||_{\mathcal{X}} + \sup_{0 \neq v_h \in \mathcal{Y}_h} \frac{_{\mathcal{Y}'}\langle D_q \mathcal{R}(q_N(\boldsymbol{y}); \boldsymbol{y})(w_h), v_h \rangle_{\mathcal{Y}}}{||v_h||_{\mathcal{Y}}} \ , \end{aligned} \tag{145}$$

where the second equality is due to **A3 Lipschitz continuity** in Assumption 5 at $q_h(\boldsymbol{y})$ and $w = q_N(\boldsymbol{y})$. We choose N_0 such that when $N \geq N_0$, there holds $L||q_h(\boldsymbol{y}) - q_N(\boldsymbol{y})||_{\mathcal{X}} \leq \beta_h(\boldsymbol{y})/2$, then we obtain from (144) and (145) that $\beta_N(\boldsymbol{y}) \geq \beta_h(\boldsymbol{y})/2 > 0$, so that **A1 stability** in Assumption 5 holds in the RB trial and tes spaces $\mathcal{X}_N$ and $\mathcal{Y}_N$. As a consequence, we have (note that $||q_h(\boldsymbol{y}) - q_N(\boldsymbol{y})||_{\mathcal{X}} \to 0$ as $N \to N_h$)

$$\forall \boldsymbol{y} \in U : \quad \lim_{N \to N_h} \frac{1}{\beta_N^2(\boldsymbol{y})} \inf_{w_N \in \mathcal{X}_N} ||q_h(\boldsymbol{y}) - w_N||_{\mathcal{X}} = 0 \ , \tag{146}$$

i.e. **A2 consistency** holds in the RB space $\mathcal{X}_N$. □

The optimal approximation property of the RB-PG compression in Theorem 8.3 implies

Theorem 8.4. *Under Assumption 5 and the assumptions of Theorem 2.7, the compression error of the RB solution of the RB-PG problem* (67) *can be bounded by*

$$\sup_{\boldsymbol{y}\in U} ||q_h(\boldsymbol{y}) - q_N(\boldsymbol{y})||_{\mathcal{X}} \leq CN^{-s}, \quad s = \frac{1}{p} - 1 \ , \tag{147}$$

where the constants C does not depend on N. The same convergence rate holds for the compression error of the RB dual solution $\psi_N(\boldsymbol{y})$ of the RB-PG dual problem (102)*.*

Proof Assumption 5 and Theorem 8.3 implies that the assumptions in Theorem 2.7 hold in both the HiFi spaces $\mathcal{X}_h$ and $\mathcal{Y}_h$, and the RB spaces $\mathcal{X}_N$ and $\mathcal{Y}_N$, in particular (ii) the well-posedness and (iii) the isomorphism property due to the inf-sup condition in $\mathcal{X}_h \times \mathcal{Y}_h$ and $\mathcal{X}_N \times \mathcal{Y}_N$, which further implies that the HiFi solution q_h and the RB solution q_N admit extension to the complex domain which is $(\boldsymbol{b}, p, \varepsilon)$-holomorphic with the same sequence $\boldsymbol{b}$ and with the same p defined in (10). Consequently, there exist SG interpolants for both q_h and q_N, whose interpolation errors decay with dimension-independent convergence rate N^{-s}, $s = 1/p - 1$. Then the proof of (147) follows using the optimality of RB compression, by a comparison argument between the RB compression error and the SG interpolation error, which is bounded as in Theorem (2.6). We refer to [15] for more details of the proof in the linear and affine case. □

As for the EI compression error of the second term in (140), we have

Theorem 8.5. *Under Assumption 5 and the assumptions of Theorem 2.7, there exists an integer M_0 such that for every $M \geq M_0$ holds*

$$\sup_{\boldsymbol{y}\in U} ||q_N(\boldsymbol{y}) - q_{N,M}(\boldsymbol{y})||_{\mathcal{X}} \leq C_M M^{-s}, \quad s = \frac{1}{p} - 1 \ , \tag{148}$$

where the constant $C_M > 0$ depends on the Lebesgue constant L_M of the empirical interpolant.

When $L_M \leq CM^k$ for some $k > 0$, one has that C_M in (148) *does not depend on M.*

Proof By the RB-PG compression problem (67) and the RB-EI compression problem (89), we have

$${}_{\mathcal{Y}'}\langle \mathcal{R}(q_N(\boldsymbol{y}); \boldsymbol{y}), v_N\rangle_{\mathcal{Y}} = {}_{\mathcal{Y}'}\langle \mathcal{J}_M \mathcal{R}(q_{N,M}(\boldsymbol{y}); \boldsymbol{y}), v_N\rangle_{\mathcal{Y}} \quad \forall v_N \in \mathcal{Y}_N \ , \tag{149}$$

Subtracting $\mathcal{R}(q_{N,M}(\boldsymbol{y}); \boldsymbol{y})$ on both sides and changing the sign results in

$${}_{\mathcal{Y}'}\langle \mathcal{R}(q_{N,M}(\boldsymbol{y}); \boldsymbol{y}) - \mathcal{R}(q_N(\boldsymbol{y}); \boldsymbol{y}), v_N\rangle_{\mathcal{Y}} = {}_{\mathcal{Y}'}\langle \mathcal{R}(q_{N,M}(\boldsymbol{y}); \boldsymbol{y}) - \mathcal{J}_M \mathcal{R}(q_{N,M}(\boldsymbol{y}); \boldsymbol{y}), v_N\rangle_{\mathcal{Y}} \quad \forall v_N \in \mathcal{Y}_N \ . \tag{150}$$

By a formal Taylor expansion of $\mathcal{R}(q_{N,M}(\boldsymbol{y}); \boldsymbol{y})$ at $q_N(\boldsymbol{y})$, the left hand side can be written as

$${}_{\mathcal{Y}'}\langle \mathcal{R}(q_{N,M}(\boldsymbol{y}); \boldsymbol{y}) - \mathcal{R}(q_N(\boldsymbol{y}); \boldsymbol{y}), v_N\rangle_{\mathcal{Y}} = {}_{\mathcal{Y}'}\langle D_q \mathcal{R}(q_N(\boldsymbol{y}); \boldsymbol{y})(q_{N,M}(\boldsymbol{y}) - q_N(\boldsymbol{y})), v_N\rangle_{\mathcal{Y}} + O(||q_{N,M}(\boldsymbol{y}) - q_N(\boldsymbol{y})||^2_{\mathcal{X}}) \ , \tag{151}$$

By the inf-sup condition (144) with $\beta_N(\boldsymbol{y}) \geq \beta_h(\boldsymbol{y})/2$, we have

$$\sup_{0\neq v_N\in\mathcal{Y}_N} \frac{{}_{\mathcal{Y}'}\langle D_q \mathcal{R}(q_N(\boldsymbol{y}); \boldsymbol{y})(q_{N,M}(\boldsymbol{y}) - q_N(\boldsymbol{y})), v_N\rangle_{\mathcal{Y}}}{||v_N||_{\mathcal{Y}}} \geq \beta_N(\boldsymbol{y})||q_{N,M}(\boldsymbol{y}) - q_N(\boldsymbol{y})||_{\mathcal{X}} \ . \tag{152}$$

As $q_{N,M}(\boldsymbol{y}) \to q_N(\boldsymbol{y})$ when $M \to N_h$, i.e. the EI becomes more accurate with more EI basis functions, and $q_{N,M}(\boldsymbol{y}) = q_N(\boldsymbol{y})$ when $M = N_h$, there exists M_0 such that when $M \geq M_0$, the second term in (150) satisfies

$$O(||q_{N,M}(\boldsymbol{y}) - q_N(\boldsymbol{y})||^2_{\mathcal{X}}) \leq (\beta_N/2)||q_{N,M}(\boldsymbol{y}) - q_N(\boldsymbol{y})||_{\mathcal{X}} \ . \tag{153}$$

On the other hand, for the right hand side of (150) we have

$$\sup_{0\neq v_N\in\mathcal{Y}_N} \frac{{}_{\mathcal{Y}'}\langle \mathcal{R}(q_{N,M}(\boldsymbol{y}); \boldsymbol{y}) - \mathcal{J}_M \mathcal{R}(q_{N,M}(\boldsymbol{y}); \boldsymbol{y}), v_N\rangle_{\mathcal{Y}}}{||v_N||_{\mathcal{Y}}} \leq ||(\mathcal{I} - \mathcal{J}_M)\mathcal{R}(q_{N,M}(\boldsymbol{y}); \boldsymbol{y})||_{\mathcal{Y}'} \ , \tag{154}$$

where $\mathcal{I}$ denotes the identity operator in $\mathcal{Y}'$. As $q_{N,M}(\boldsymbol{y})$ is the solution of the RB-EI compression problem (89), which fulfils the assumptions in Theorem 2.7, in particular, (ii) problem (89) is well-posed by the supremizer approach; (iii) the isomorphism property holds due to the inf-sup stability condition in $\mathcal{X}_N$ and $\mathcal{Y}_N$; (iv) $\mathcal{J}_M \mathcal{R}(q, u) = \boldsymbol{r} R^{-1} \boldsymbol{\sigma}(\mathcal{R}(q, u))$ is complex continuously differentiable w.r.t. q and u as $\boldsymbol{r} R^{-1} \boldsymbol{\sigma}(\cdot)$ is linear and bounded. By Theorem 2.7, $q_{N,M}$ admits a continous extension to the complex domain which is $(\boldsymbol{b}, p, \varepsilon)$-holomorphic, and so does $\mathcal{R}(q_{N,M}(\boldsymbol{y}); \boldsymbol{y})$ due to the complex continuous differentiability. Therefore, as in

Theorem 2.6, there exists a SG interpolation $\mathcal{S}_M$ for $\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})$ such that

$$\sup_{\boldsymbol{y}\in U} \|(\mathcal{I}-\mathcal{S}_M)\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})\|_{\mathcal{Y}'} \leq CM^{-s}, \quad s=\frac{1}{p}-1 \tag{155}$$

with constant C independent of M. By the triangle inequality,

$$||(\mathcal{I}-\mathcal{J}_M)\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})||_{\mathcal{Y}'} \leq ||(\mathcal{I}-\mathcal{S}_M)\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})||_{\mathcal{Y}'} + ||(\mathcal{S}_M-\mathcal{J}_M)\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})||_{\mathcal{Y}'} \ , \tag{156}$$

where the second term can be rewritten due to $\mathcal{J}_M\mathcal{S}_M\mathcal{R}=\mathcal{S}_M\mathcal{R}$ (since $\mathcal{J}_M$ is exact for the subspace $Z_M=\text{span}\{\mathcal{R}(q_{N,M}(\boldsymbol{y}^m);\boldsymbol{y}^m): m=1,\dots,M\}$ due to the interpolation property, and due to $\mathcal{S}_M\mathcal{R}\in Z_M$) as

$$||(\mathcal{S}_M-\mathcal{J}_M)\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})||_{\mathcal{Y}'} = ||\mathcal{J}_M(\mathcal{S}_M-\mathcal{I})\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})||_{\mathcal{Y}'} \leq L_M||(\mathcal{I}-\mathcal{S}_M)\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})||_{\mathcal{Y}'} \ , \tag{157}$$

where L_M is the Lebesgue constant of $\mathcal{J}_M$. It is proved in [1] that $L_M\leq 2^M-1$. This bound, however, is likely too pessimistic: in practical experiments L_M is observed to grow linearly w.r.t. M [33]. Consequently, a combination of (155), (156) and (157) yields the estimate

$$||(\mathcal{I}-\mathcal{J}_M)\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})||_{\mathcal{Y}'} \leq C(1+L_M)M^{-s}, \quad s=\frac{1}{p}-1 \ . \tag{158}$$

In the case that the Lebesgue constant of EI grows algebraically w.r.t M, i.e. if there exists $C>0$ such that $L_M\leq CM^k$ for all $M\geq 1$ with some $k>0$, by the fact $\mathcal{J}_M\mathcal{S}_M\mathcal{R}=\mathcal{S}_M\mathcal{R}$, it can be shown as in [18, Thm. 3.1] that

$$||(\mathcal{I}-\mathcal{J}_M)\mathcal{R}(q_{N,M}(\boldsymbol{y});\boldsymbol{y})||_{\mathcal{Y}'} \leq CM^{-s}, \quad s=\frac{1}{p}-1 \ . \tag{159}$$

A combination of (150), (151), (152), (153), (154), together with (158) or (159) concludes with the constant $C_{EI}=2C_i(1+L_M)/\beta_N$ or $2C/\beta_N$ when $L_M\leq CM^k$ for some $k>0$. □

Note that the RB solution $q_N(\boldsymbol{y})$ and the RB-EI solution $q_{N,M}(\boldsymbol{y})$ are approximations of the HiFi solution $q_h(\boldsymbol{y})$, which is an approximation of the true solution $q(\boldsymbol{y})$. To emphasize this dependence, we denote the HiFi-RB solution as $q_{h,N}(\boldsymbol{y})$ and the HiFi-RB-EI solution as $q_{h,N,M}(\boldsymbol{y})$ for any $\boldsymbol{y}\in U$. A combination of the previous estimates on the HiFi, RB and EI errors leads to the following result.

Theorem 8.6. *Under Assumption 5 and the assumptions of Theorem 2.7, there holds the a priori error estimate for the HiFi-RB-EI compression of the solution*

$$\sup_{\boldsymbol{y}\in U} ||q(\boldsymbol{y})-q_{h,N,M}(\boldsymbol{y})||_{\mathcal{X}} \leq C_{HiFi}h^k + C_{RB}N^{-s} + C_{EI}M^{-s} \ , \tag{160}$$

where the constant C_{HiFi} does not depend on h, C_{RB} does not depend on N, C_{EI} does not depend on M when the EI Lebesgue constant $L_{M_{EI}}$ is bounded as $L_{M_{EI}}\leq CM_{EI}^k$ for some $k>0$, as in Theorem 8.5.

Proof The estimate is a result of the triangle inequality

$$||q(\boldsymbol{y})-q_{h,N,M}(\boldsymbol{y})||_{\mathcal{X}} \leq ||q(\boldsymbol{y})-q_h(\boldsymbol{y})||_{\mathcal{X}} + ||q_h(\boldsymbol{y})-q_{h,N}(\boldsymbol{y})||_{\mathcal{X}} + ||q_{h,N}(\boldsymbol{y})-q_{h,N,M}(\boldsymbol{y})||_{\mathcal{X}} \tag{161}$$

together with the estimates in Theorem 8.2, Theorem 8.4 and Theorem 8.5. Note that when taking the bound (161) the supremum over $\boldsymbol{y}\in U$, the first term can be bounded uniformly w.r.t. $\boldsymbol{y}$ as $||a(\boldsymbol{y})||$ and $||q(\boldsymbol{y})||_{\mathcal{X}^s}$ can be bounded from above, and $\beta(\boldsymbol{y})$ and $\beta_h(\boldsymbol{y})$ can be bounded from below, uniformly w.r.t. $\boldsymbol{y}$, by Proposition 2.1, and Assumption 5. □

Accounting for the SG integration error in the approximation of $\mathbb{E}^{\boldsymbol{\pi}_0}[\Theta]$ and $\mathbb{E}^{\boldsymbol{\pi}_0}[\Psi]$, we obtain

Theorem 8.7. *Under Assumption 2, the assumptions of Theorem 2.7, and Assumption 5, there holds the a priori error estimate*

$$|\mathbb{E}^{\boldsymbol{\pi}_0}[\Theta]-\mathbb{E}^{\boldsymbol{\pi}_0}[\mathcal{S}_{\Lambda_{M_{SG}}}\Theta_{h,N,M_{EI}}]| \leq C(\Gamma,r)||\delta||_Y(C_{HiFi}h^k + C_{RB}N^{-s} + C_{EI}M_{EI}^{-s}) + C_{SG}M_{SG}^{-s} \ , \tag{162}$$

where the constant C_{SG} does not depend on M_{SG}, and the remaining constants are as given in (31) *and Theorem 8.6. The same estimate holds also for any QoI Ψ defined in* (25).

Proof The proof is a result of the triangle inequality

$$|\mathbb{E}^{\boldsymbol{\pi}_0}[\Theta]-\mathbb{E}^{\boldsymbol{\pi}_0}[\mathcal{S}_{\Lambda_{M_{SG}}}\Theta_{h,N,M_{EI}}]| \leq |\mathbb{E}^{\boldsymbol{\pi}_0}[\Theta]-\mathbb{E}^{\boldsymbol{\pi}_0}[\Theta_{h,N,M_{EI}}]| + |\mathbb{E}^{\boldsymbol{\pi}_0}[\Theta_{h,N,M_{EI}}]-\mathbb{E}^{\boldsymbol{\pi}_0}[\mathcal{S}_{\Lambda_{M_{SG}}}\Theta_{h,N,M_{EI}}]| \ , \tag{163}$$

where the first term is bounded as in the first term of (162) by Proposition (31) and Theorem 8.6, and the second term can be bounded as

$$|\mathbb{E}^{\boldsymbol{\pi}_0}[\Theta_{h,N,M_{EI}}] - \mathbb{E}^{\boldsymbol{\pi}_0}[\mathcal{S}_{\Lambda_{M_{SG}}}\Theta_{h,N,M_{EI}}]| \leq \sup_{\boldsymbol{y}\in U} |\Theta_{h,N,M_{EI}} - \mathcal{S}_{\Lambda_{M_{SG}}}\Theta_{h,N,M_{EI}}|\,, \tag{164}$$

which is estimated using Theorem 2.6 and the fact that $\Theta_{h,N,M_{EI}}$ is $(\boldsymbol{b}, p, \varepsilon)$-holomorphic as $q_{h,N,M_{EI}}$ is so (see [44]), which is shown in the proof of Theorem 8.5. We denote this bound as $C_{SG}M_{SG}^{-s}$. □

[1] 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.

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

[3] S.C. Brenner and L.R. Scott. *The mathematical theory of finite element methods*. Springer Verlag, 2008.

[4] A. Buffa, Y. Maday, A.T. Patera, C. Prudhomme, and G. Turinici. A priori convergence of the greedy algorithm for the parametrized reduced basis method. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46(03):595–603, 2012.

[5] T. Bui-Thanh, O. Ghattas, J. Martin, and G. Stadler. A computational framework for infinite-dimensional Bayesian inverse problems Part I: The linearized case, with application to global seismic inversion. *SIAM J. Sci. Comput.*, 35(6):A2494–A2523, 2013.

[6] 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.

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

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

[9] 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.

[10] 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.

[11] P. Chen and A. Quarteroni. A new algorithm for high-dimensional uncertainty quantification based on dimension-adaptive sparse grid approximation and reduced basis methods. *EPFL, MATHICSE Report 09, 2014. Accepted in Journal of Computational Physics, in press.*, 2015.

[12] 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.

[13] P. Chen, A. Quarteroni, and G. Rozza. Multilevel and weighted reduced basis method for stochastic optimal control problems constrained by Stokes equations. *EPFL, MATHICSE Report 33, 2013, Accepted in Numerische Mathematik, in press*, 2015.

[14] P. Chen, A. Quarteroni, and G. Rozza. Reduced order methods for uncertainty quantification problems. *ETH Report 03, Submitted*, 2015.

[15] P. Chen and Ch. Schwab. Sparse-grid, reduced-basis Bayesian inversion. Technical Report 2014-36, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2014.

[16] P. Chen and Ch. Schwab. Adaptive sparse grid model order reduction for fast Bayesian estimation and inversion. Technical Report 2015-08, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2015.

[17] A. Chkifa, A. Cohen, R. DeVore, and Ch. Schwab. Adaptive algorithms for sparse polynomial approximation of parametric and stochastic elliptic PDEs. *M2AN Math. Mod. and Num. Anal.*, 47(1):253–280, 2013.

[18] A. Chkifa, A. Cohen, and Ch. Schwab. Breaking the curse of dimensionality in sparse polynomial approximation of parametric pdes. *Journal de Mathématiques Pures et Appliquées*, 2014.

[19] Z. Ciesielski and J. Domsta. Construction of an orthonormal basis in $C^m(I^d)$ and $W_p^m(I^d)$. *Studia Math.*, 41:211–224, 1972.

[20] S. L. Cotter, M. Dashti, and A. M. Stuart. Approximation of Bayesian inverse problems for PDEs. *SIAM J. Numer. Anal.*, 48(1):322–345, 2010.

[21] T. Cui, Y.M. Marzouk, and K.E. Willcox. Data-driven model reduction for the Bayesian solution of inverse problems. *arXiv preprint arXiv:1403.4290*, 2014.

[22] M. Dashti and A. M. Stuart. Inverse problems: a Bayesian perspective. Technical report.

[23] P. Deuflhard. *Newton methods for nonlinear problems: affine invariance and adaptive algorithms*, volume 35. Springer Science & Business Media, 2011.

[24] J. Dick, R. Gantner, Q.T. LeGia, and Ch. Schwab. Higher order quasi monte carlo integration for Bayesian inversion of holomorphic, parametric operator equations. Technical report, Seminar for Applied Mathematics, ETH Zürich, 2015.

[25] 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.

[26] L. Formaggia, A. Quarteroni, and A. Veneziani. *Cardiovascular Mathematics: Modeling and simulation of the circulatory system*, volume 1. Springer Science & Business Media, 2010.

[27] D. Galbally, K. Fidkowski, K.E. Willcox, and O. Ghattas. Non-linear model reduction for uncertainty quantification in large-scale inverse problems. *International journal for numerical methods in engineering*, 81(12):1581–1608, 2010.

[28] 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.

[29] V.H. Hoang, Ch. Schwab, and A. Stuart. Complexity analysis of accelerated MCMC methods for Bayesian inversion. *Inverse Problems*, 29(8), 2013.

[30] T. Lassila, A. Manzoni, A. Quarteroni, and G. Rozza. A reduced computational and geometrical framework for inverse problems in hemodynamics. *International journal for numerical methods in biomedical engineering*, 29(7):741–776, 2013.

[31] C. Lieberman, K. Willcox, and O. Ghattas. Parameter and state model reduction for large-scale statistical inverse problems. *SIAM Journal on Scientific Computing*, 32(5):2523–2542, 2010.

[32] Y. Maday, O. Mula, A. T. Patera, and M. Yano. The generalized empirical interpolation method: stability theory on Hilbert spaces with an application to the Stokes equation. *Comput. Methods Appl. Mech. Engrg.*, 287:310–334, 2015.

[33] Y. Maday, N.C. Nguyen, A.T. Patera, and G.S.H. Pau. A general, multipurpose interpolation procedure: the magic points. *Communications on Pure and Applied Analysis*, 8(1):383–404, 2009.

[34] J. Martin, L.C. Wilcox, C. Burstedde, and O. Ghattas. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing*, 34(3):A1460–A1487, 2012.

[35] Y.M. Marzouk, H.N. Najm, and L.A. Rahn. Stochastic spectral methods for efficient bayesian solution of inverse problems. *Journal of Computational Physics*, 224(2):560–586, 2007.

[36] 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.

[37] J Pousin and J Rappaz. Consistency, stability, a priori and a posteriori errors for Petrov-Galerkin methods applied to nonlinear problems. *Numerische Mathematik*, 69(2):213–231, 1994.

[38] 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.

[39] G. Rozza and K. Veroy. On the stability of the reduced basis method for stokes equations in parametrized domains. *Computer methods in applied mechanics and engineering*, 196(7):1244–1260, 2007.

[40] C. Schillings. A note on sparse, adaptive smolyak quadratures for bayesian inverse problems. *Oberwolfach Report*, 10(1):239–293, 2013.

[41] C. Schillings and Ch. Schwab. Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. *Inverse Problems*, 29(6), 2013.

[42] C. Schillings and Ch. Schwab. Sparsity in Bayesian inversion of parametric operator equations. *Inverse Problems*, 30(6), 2014.

[43] Ch. Schwab and C. Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numerica*, 20:291–467, 2011.

[44] Ch. Schwab and A. M. Stuart. Sparse deterministic approximation of Bayesian inverse problems. *Inverse Problems*, 28(4):045003, 32, 2012.

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

[46] P. Wolfe. Convergence conditions for ascent methods. *SIAM review*, 11(2):226–235, 1969.

Recent Research Reports

Nr.	Authors/Title

2015-11 R. Bourquin
Exhaustive search for higher-order Kronrod-Patterson Extensions

2015-12 R. Bourquin and V. Gradinaru
Numerical Steepest Descent for Overlap Integrals of Semiclassical Wavepackets

2015-13 A. Hiltebrand and S. Mishra
Entropy stability and well-balancedness of space-time DG for the shallow water equations with bottom topography

2015-14 B. Ayuso de Dios and R. Hiptmair and C. Pagliantini
Auxiliary Space Preconditioners for SIP-DG Discretizations of H(curl)-elliptic Problems with Discontinuous Coefficients

2015-15 A. Paganini and S. Sargheini and R. Hiptmair and C. Hafner
Shape Optimization of microlenses

2015-16 V. Kazeev and Ch. Schwab
Approximation of Singularities by Quantized-Tensor FEM

2015-17 P. Grohs and Z. Kereta
Continuous Parabolic Molecules

2015-18 R. Hiptmair
Maxwell’s Equations: Continuous and Discrete

2015-19 X. Claeys and R. Hiptmair and E. Spindler
Second-Kind Boundary Integral Equations for Scattering at Composite Partly Impenetrable Objects

2015-20 R. Hiptmair and A. Moiola and I. Perugia
A Survey of Trefftz Methods for the Helmholtz Equation