ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

Adaptive Sparse Grid Model Order Reduction for Fast Bayesian Estimation and Inversion

P. Chen and Ch. Schwab

Research Report No. 2015-08
February 2015

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

Funding ERC: 247277 STAHDPDE

Adaptive Sparse Grid Model Order Reduction for fast Bayesian Estimation and Inversion *

Peng Chen and Christoph Schwab

Abstract We present new sparse-grid based algorithms for fast Bayesian estimation and inversion of parametric operator equations. We propose Reduced Basis (RB) acceleration of numerical integration based on Smolyak sparse grid quadrature. To tackle the curse-of-dimensionality in high-dimensional Bayesian inversion, we exploit sparsity of the parametric forward solution map as well as of the Bayesian posterior density with respect to the random parameters. We employ an dimension adaptive Sparse Grid method (aSG) for both, offline-training the reduced basis as well as for deterministic quadrature of the conditional expectations which arise in Bayesian estimates. For the forward problem with nonaffine dependence on the random variables, we perform further affine approximation based on the Empirical Interpolation Method (EIM) proposed in [1]. A combined algorithm to adaptively refine the sparse grid quadrature, reduced basis approximation and empirical interpolation is proposed and its computational efficiency is demonstrated in numerical experiments for nonaffine-parametric, stationary, elliptic diffusion problems, in two spacial and in parameter space dimensions up to 1024.

1 Introduction

Bayesian estimation, ie. the "most likely" prediction of responses of ordinary and partial differential differential equations (ODEs and PDEs, for short), subject to uncertain input data, given noisy observation data, is a key topic in computational

* Work supported in part by the European Research Council (ERC) under AdG427277

Peng Chen
HG G 56.1, Seminar for Applied Mathematics, ETH Zürich, Rämistrasse 101, CH-8092 Zürich, Switzerland. e-mail: peng.chen@sam.math.ethz.ch

Christoph Schwab
HG G 57.1, Seminar for Applied Mathematics, ETH Zürich, Rämistrasse 101, CH-8092 Zürich, Switzerland. e-mail: christoph.schwab@sam.math.ethz.ch

statistics and in computational science. We refer to [28, 16] and the references there for a survey.

As a rule, Monte-Carlo (MC) based sampling methods are used. Then, the methods are computationally intensive due to the slow convergence of MC methods (implying a rather large number of samples) and due to the high cost of forward solves per "sample" (which, in PDE models, amounts to one numerical PDE solve per sample). Distributed uncertain inputs such as, for example, uncertain diffusion coefficients, require forward solves of *infinite-dimensional, parametrized PDEs*.

Recent mathematical results [14, 20, 18, 12, 19, 13] indicate that the parameter-to-solution maps of these parametric operator equations exhibit *sparsity* in the sense that their n-widths are small, independent of the number of parameters which are activated in the approximation. This has lead to the proposal of *sparse, dimension-adaptive interpolation schemes* for the exploitation of this sparsity in the solution map, see, e.g. [11], and to deterministic *sparse, adaptive quadrature methods in Bayesian inversion [25]*. Small n-widths of parametric PDE solution maps can, in principle, be exploited by greedy approximation strategies, due to recent mathematical results [2, 29]. This observation was used in recent works by one of the authors [5, 6, 3, 4, 8] to accelerate forward solves of PDEs. One way to accelerate MCMC methods is via gpc-based surrogates of the parametric forward maps analyzed in [21].

Other acceleration via model order reduction approaches [23, 17, 15] have also been investigated. Here, we consider acceleration via *reduced basis methods* (RB) relying on adaptive sparse grid (aSG) scheme for both interpolation and integration. At each sparse grid node generated in the adaptive algorithm, we evaluate the RB surrogate density of the Bayesian posterior or some other Quantity of Interest (QoI for short) for inference. If the surrogate is not determined to sufficient accuracy, as certified by some reliable and efficient goal-oriented error estimator, then a full forward problem is solved numerically to evaluate the posterior density and the RB approximation is refined by the full solution at this grid node. The efficient online evaluation of the RB surrogate and the error estimator introduced in [9] depends on an affine-parametric structure of the underlying PDEs that enables efficient offline-online decomposition. For more general nonaffine problems, we propose in the present paper to compute their affine approximations by empirical interpolation methods [1, 7] in combination with aSG. To this end, a combined adaptive sparse grid, empirical interpolation and reduced basis methods (aSG-EIM-RB) are developed to reliably and efficiently accelerate the Bayesian estimation and inversion. In this paper, we present the first detailed description of the corresponding algorithm, and the first numerical experiments performed with the proposed numerical inversion strategy. Dimension-independent convergence rates are demonstrated with parameter space dimensions up to 1024.

This paper is organized as follows: the formulation of parametric Bayesian inversion in function space is presented in section 2. The admissible forward maps are countably-parametric operator equations. Section 3 presents a constructive algorithm based on adaptive sparse grid interpolation for gpc approximation of the parametric solution, of the posterior density and of the related QoI. Section 4 is devoted to the development of reduced basis acceleration for the Bayesian inversion, where a com-

bined aSG-EIM-RB algorithm is presented. Numerical experiments in demonstrating the dimension-independent convergence rates and computational efficiency of the proposed algorithm are provided in section 5, followed by conclusions in section 6.

2 Bayesian Inversion

We review the mathematical setting of Bayesian inversion of partial differential equations with distributed uncertainty, in the mathematical framework of [16].

By *uncertainty parametrization* together with a Bayesian prior on the (generally infinite-dimensional) parameter space, the problem of Bayesian estimation is converted to a problem of quadrature of the parametric, deterministic posterior density.

Sparsity of the parametric forward solution map and of the Bayesian posterior density as well as its integration with respect to the infinite-dimensional parameter sequence will be presented. Dimension-independent convergence rates of sparsity-exploiting Smolyak quadratures are stated and verified in numerical experiments.

2.1 Formulation of Bayesian Inversion

We consider Bayesian inversion problems that consist of: given observation data subject to additive, centered gaussian observation noise of some system output, and given a prior distribution of the uncertain, distributed system input, Bayes' theorem yields the posterior distribution and an estimate of the "most likely" system response for some quantity of interest (QoI), that depends on the system state [16]. More precisely, we consider a system with state variable q belonging to a separable Hilbert space $\mathscr{X}$ (with (anti-)dual $\mathscr{X}'$), which is determined through the system forward map $G : X \to \mathscr{X}$ by an uncertain system input variable u that takes values in a separable Banach space X. The observation data $\delta \in Y = \mathbb{R}^K$, $K \in \mathbb{N}$, is determined by the system output $\mathscr{O}(\cdot) = (o_1, \ldots, o_K) \in (\mathscr{X}')^K$ which is assumed to be corrupted with additive Gaussian noise $\eta \sim \mathscr{N}(0, \Gamma)$ with symmetric positive definite correlation matrix $\Gamma \in \mathbb{R}^{K \times K}$, i.e.

$$\delta = \mathscr{O}(G(u)) + \eta \ . \tag{1}$$

We assume that a prior probability distribution $\pi_0 : X \to \mathbb{R}$ is prescribed for the uncertain system input u with $\pi_0(X) = 1$. Under appropriate continuity conditions on the uncertainty-to-observation map $\mathscr{G} := \mathscr{O} \circ G(\cdot) : X \to Y$, Bayes' rule guarantees the existence of a posterior distribution π^δ that is absolutely continuous with density $\Theta : X \to \mathbb{R}$ with respect to the prior distribution. Various concrete, sufficient conditions for absolute posterior continuity are provided in the surveys [28, 16]. We parametrize the uncertain input u taking values in the separable Banach space X by postulating a countable basis $(\phi_j)_{j \in \mathbb{J}} \in X$, where $\mathbb{J} = \{1, \ldots, J\}$ with $J \in \mathbb{N}$ or $\mathbb{J} = \mathbb{N}$. We assume that the uncertain input u can be represented by these bases: there is an

affine parametrization, i.e. $u = u(\mathbf{y})$ with the parameter (after possible rescaling and shifting) $\mathbf{y} = (y_j)_{j\in\mathbb{J}} \in U$, being $U = [-1,1]^{\mathbb{J}}$ a reference parameter space. Under appropriate re-scaling of the basis ϕ_j, we may reparametrize the distributed, uncertain input data u as

$$u(\mathbf{y}) = \bar{u} + \sum_{j\in\mathbb{J}} y_j \phi_j, \quad \mathbf{y} \in U \ , \tag{2}$$

where $\bar{u} \in X$ is the "nominal" value of the uncertain data u and $u - \bar{u}$ entails possible fluctuations of u through $\mathbf{y} \in U$. An example of the expression (2) is Karhunen–Loève expansion of a random field u with mean field $\bar{u}$ and (rescaled) eigenfunctions $(\phi_j)_{j\in\mathbb{J}}$. In practice, the parametrization may also be a nonlinear transformation of an affine input, i.e. the parameters can not be separated from the bases. For instance in describing a positive permeability field κ, we may assume that $\kappa = e^u$ with u defined in (2), so that κ is positive at each parameter value $\mathbf{y} \in U$ but nonaffine with respect to $\mathbf{y}$. An example for a "non-affine" uncertainty parameterization is a Karhunen–Loève expansion of $\log(\kappa)$, which typically arises in log-gaussian models for u, in which case the Bayesian prior is a Gaussian measure on X. Under the parametrization, for prescribing a prior distribution of the uncertain input data u we only need to prescribe a prior measure for the parameters, i.e.

$$\pi_0 = \bigotimes_{j\in\mathbb{J}} \frac{1}{2}\lambda^1 \text{ or } d\pi_0(\mathbf{y}) = \bigotimes_{j\in\mathbb{J}} \frac{1}{2} dy_j \ , \tag{3}$$

where λ^1 denotes the Lebesgue measure on $[-1,1]$. By Bayes' theorem, there exists a posterior measure which is absolutely continuous with respect to the prior. For the corresponding Radon–Nikodym derivative holds

$$\frac{d\pi^\delta}{d\pi_0}(\mathbf{y}) = \frac{1}{Z}\Theta(\mathbf{y}) \ , \tag{4}$$

where the (rescaled) posterior density Θ is given by

$$\Theta(\mathbf{y}) = \exp\left(-\frac{1}{2}(\delta - \mathscr{O}(G(u(\mathbf{y}))))^\top \Gamma^{-1}(\delta - \mathscr{O}(G(u(\mathbf{y}))))\right) \ , \tag{5}$$

and the renormalization constant Z is defined as

$$Z := \mathbb{E}^{\pi_0}[\Theta] = \int_U \Theta(\mathbf{y}) d\pi_0(\mathbf{y}) \ . \tag{6}$$

Under the posterior distribution, we can evaluate some quantity of interest (QoI) Ψ that depends on the system input u, e.g. the system input itself $\Psi(u) = u$ or the system response $\Psi(u) = G(u)$, as well as some statistics of the QoI, e.g. the expectation

$$\mathbb{E}^{\pi^\delta}[\Psi(u)] = \mathbb{E}^{\pi_0}[\Psi(u)\Theta] = \int_U \Psi(u(\mathbf{y}))\Theta(\mathbf{y}) d\pi_0(\mathbf{y}) \ . \tag{7}$$

2.2 Parametric Operator Equations

Under the above parametrization of the system input, we consider a class of parametric operator equations for the modelling of the system, which read as: for any parameter value $\mathbf{y} \in U$, find the solution $q(\mathbf{y}) \in \mathscr{X}$ such that

$$A(\mathbf{y})q(\mathbf{y}) = f(\mathbf{y}) \quad \text{in } \mathscr{Y}' , \tag{8}$$

where $\mathscr{Y}$ is a separable Hilbert space with anti-dual $\mathscr{Y}'$, $A(\mathbf{y})$ is a parametric operator and $f(\mathbf{y})$ is a parametric right hand side, both depending on the parameter $\mathbf{y}$ through the uncertain system input $u(\mathbf{y})$. In particular, we consider linear systems modelled by countably-parametric, linear operator families $A(\mathbf{y}) \in \mathscr{L}(\mathscr{X}, \mathscr{Y}')$. We associate the parametric operator $A(\mathbf{y})$ and $f(\mathbf{y})$ with sesquilinear and antilinear forms, respectively, in the Hilbert spaces $\mathscr{X}$ and $\mathscr{Y}$ over $\mathbb{C}$ as

$$a(\mathbf{y}; w, v) =_{\mathscr{Y}} \langle v, A(\mathbf{y})w\rangle_{\mathscr{Y}'} \text{ and } f(\mathbf{y}; v) :=_{\mathscr{Y}} \langle v, f(\mathbf{y})\rangle_{\mathscr{Y}'} \quad \forall w \in \mathscr{X}, v \in \mathscr{Y} . \tag{9}$$

The weak formulation of the parametric operator equation (8) reads: for any parameter value $\mathbf{y} \in U$, find the solution $q(\mathbf{y}) \in \mathscr{X}$ such that

$$a(\mathbf{y}; q(\mathbf{y}), v) = f(\mathbf{y}; v) \quad \forall v \in \mathscr{Y} . \tag{10}$$

For the well-posedness of problem (10) and for the approximation of its solution, the Bayesian posterior density, and some QoI, we make the following assumptions.

Assumption 1 ***A1*** *For $\varepsilon > 0$ and $0 < p < 1$, there exists a positive sequence $(b_j)_{j\in\mathbb{J}} \in \ell^p(\mathbb{J})$ such that for any sequence $\boldsymbol{\rho} := (\rho_j)_{j\in\mathbb{J}}$ with $\rho_j > 1$ for all $j \in \mathbb{J}$ and*

$$\sum_{j\in\mathbb{J}} (\rho_j - 1) b_j \le \varepsilon, \tag{11}$$

the parametric maps a and f in (10) *admit holomorphic extensions to certain cylindrical sets $\mathscr{O}_{\boldsymbol{\rho}} = \otimes_{j\in\mathbb{J}} \mathscr{O}_{\rho_j}$, where $\mathscr{O}_{\rho_j} \subset \mathbb{C}$ is an open set containing the Bernstein ellipse $\mathscr{E}_{\rho_j}$ with semi axes of length $(\rho_j + \rho_j^{-1})/2$ and $(\rho_j - \rho_j^{-1})/2 > 1$.*

A2 *There exist constants $0 < \beta < \gamma < \infty$ and $\theta > 0$ such that these extensions satisfy for all $\mathbf{z} \in \mathscr{O}_{\boldsymbol{\rho}}$ the uniform continuity conditions*

$$\sup_{v\in\mathscr{Y}} \frac{f(\mathbf{z}; v)}{||v||_{\mathscr{Y}}} \le \theta \text{ and } \sup_{w\in\mathscr{X}} \sup_{v\in\mathscr{Y}} \frac{a(\mathbf{z}; w, v)}{||w||_{\mathscr{X}} ||v||_{\mathscr{Y}}} \le \gamma \tag{12}$$

and the uniform inf-sup conditions

$$\inf_{0\neq w\in\mathscr{X}} \sup_{0\neq v\in\mathscr{Y}} \frac{|a(\mathbf{z}; w, v)|}{||w||_{\mathscr{X}} ||v||_{\mathscr{Y}}} \ge \beta \text{ and } \inf_{0\neq v\in\mathscr{Y}} \sup_{0\neq w\in\mathscr{X}} \frac{|a(\mathbf{z}; w, v)|}{||w||_{\mathscr{X}} ||v||_{\mathscr{Y}}} \ge \beta . \tag{13}$$

The following results are key to dimension-robust convergence rate of the model order reduction methods; we refer to [12], [27] and [25] for proofs.

Theorem 1. *Under Assumption 1, there exists a positive constant $C < \infty$ depending on $\theta, \gamma, \beta, p, \varepsilon$ and ρ, such that the operator equation* (8) *admits a unique uniformly bounded solution satisfying a generalized polynomial chaos expansion (gPC)*

$$\sup_{\mathbf{z}\in\mathcal{O}_\rho} ||q(\mathbf{z})||_{\mathscr{X}} \le C \text{ and } q(\mathbf{y}) = \sum_{\nu\in\mathscr{F}} q_\nu P_\nu(\mathbf{y}) \tag{14}$$

where $P_\nu(\mathbf{y}) := \prod_{j\in\mathbb{J}} P_{\nu_j}(y_j)$, with P_n denoting the univariate Legendre polynomial of degree n for the interval $[-1,1]$ normalized such that $||P_n||_{L^\infty([-1,1])} = 1$ and $\mathscr{F}$ denotes the countable set of all finitely supported sequences $\nu \in \mathbb{N}_0^{\mathbb{J}}$. Moreover, there exists a downward closed index set $\Lambda_M \subset \mathscr{F}$ *("dc set", for short)*[2] *with at most M indices such that the dimension-independent convergence rate holds*

$$\sup_{\mathbf{y}\in U} \|q(\mathbf{y}) - \sum_{\nu\in\Lambda_M} q_\nu P_\nu(\mathbf{y})\|_{\mathscr{X}} \le C_q M^{-s}, \quad s = \frac{1}{p} - 1\,. \tag{15}$$

Here the constant C_q neither depends on M nor on the number of active coordinates, ie. $\max\{\#\{j\in\mathbb{N} : \nu_j \neq 0\} : \nu \in \Lambda_M\}$. *The same convergence rate* (15) *also holds for the approximation of the posterior density $\Theta(\mathbf{y})$ as well as for the QoI $\Psi(\mathbf{y})$.*

3 Adaptive Sparse Grid Approximation

Theorem 1 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 [26], 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 first 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 $-1 \le y_1 < \cdots < y_m \le 1$, we define the interpolation operator $\mathscr{I} : C(U;\mathscr{Z}) \to \mathscr{P}_{m-1}(U)\otimes\mathscr{Z}$ as

$$\mathscr{I}g(y) = \sum_{k=1}^{m} g(y_k)\ell_k(y), \tag{16}$$

[2] A subset $\Lambda \subset \mathscr{F}$ is a dc set if for every $\nu \in \Lambda_M$ also $\mu \in \Lambda_M$ for any $\mu \preceq \nu$ ($\mu_j \le \nu_j$ for all $j \in \mathbb{J}$)

where the function $g \in C(U;\mathscr{Z})$, representing e.g. the parametric forward solution map q with $\mathscr{Z}=\mathscr{X}$ or the posterior density Θ with $\mathscr{Z}=\mathbb{R}$; $\ell_k(y)$, $1\le k\le m$, are the associated Lagrange polynomials in $\mathscr{P}_{m-1}(U)$, the space of polynomials of degree at most $m-1$. To define the sparse collocation, as usual the interpolation operator defined in (16) is recast as telescopic sum, ie.,

$$\mathscr{I}_L g(y) = \sum_{l=1}^{L} \triangle^l g(y)\,, \tag{17}$$

where L represents the level of interpolation grid; $\triangle^l := \mathscr{I}_l - \mathscr{I}_{l-1}$ with $\mathscr{I}_0 g \equiv 0$. Let Ξ^l denote the set of all interpolation nodes in the grid of level l, such that the grid is nested, i.e. $\Xi^l \subset \Xi^{l+1}$, $l=0,\dots,L-1$, with $\Xi^0=\emptyset$ and $\Xi^L=\{y_1,\dots,y_m\}$. As $\mathscr{I}_{l-1}g(y)=g(y)$ for any $y\in\Xi^{l-1}$, we have $\mathscr{I}_{l-1}=\mathscr{I}_l\circ\mathscr{I}_{l-1}$ and, with the notation $\Xi^l_\triangle = \Xi^l \setminus \Xi^{l-1}$, the interpolation operator (17) can be written in the form

$$\mathscr{I}_L g(y) = \sum_{l=1}^{L}\sum_{y_k^l\in\Xi^l_\triangle} (\mathscr{I}_l - \mathscr{I}_l\circ\mathscr{I}_{l-1})g(y) = \sum_{l=1}^{L}\sum_{y_k^l\in\Xi^l_\triangle} \underbrace{(g(y_k^l) - \mathscr{I}_{l-1}g(y_k^l))}_{s_k^l}\ell_k^l(y)\,, \tag{18}$$

where s_k^l represents the interpolation error of $\mathscr{I}_{l-l}g$ evaluated at the node $y_k^l\in\Xi^l_\triangle$, $k=1,\dots,|\Xi^l_\triangle|$, so that we can use it as a posteriori error estimator for adaptive construction of the interpolation (18). More precisely, we start from the root level $L=1$ with the root interpolation node $y=0$, whenever the interpolation error estimator

$$\mathscr{E}_i := \max_{y_k^L\in\Xi^L_\triangle} |s_k^L| \tag{19}$$

is larger than a given tolerance, we refine the interpolation to the next level $L+1$ by taking new interpolation node, for instance one Leja node

$$y_1^{L+1} = \operatorname*{argmax}_{y\in U} \prod_{l=1}^{L} |y-y^l|\,, \tag{20}$$

or Clenshaw–Curtis nodes

$$y_k^{L+1} = \cos\left(\frac{k}{2^{L-1}}\pi\right), \quad k=0,1 \text{ for } L=1; k=1,3,\dots,2^{L-1}-1 \text{ for } L\ge 2\,. \tag{21}$$

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

$$\mathbb{E}[g] \approx \mathbb{E}[\mathscr{I}_L g] = \sum_{l=1}^{L}\sum_{y_k^l\in\Xi^l_\triangle} s_k^l w_k^l, \quad \text{being } w_k^l = \mathbb{E}[\ell_k^l]\,, \tag{22}$$

for which the integration error estimator can be taken as

$$\mathscr{E}_e := \left| \sum_{y_k^L \in \Xi_\triangle^L} s_k^L w_k^L \right| . \tag{23}$$

3.2 Adaptive Sparse Grid Approximation

In multiple dimensions $\mathbf{y} \in U = [-1,1]^J$, we construct an adaptive sparse grid (aSG) interpolation by tensorizing the univariate interpolation formula (17)

$$\mathscr{S}_{\Lambda_M} g(\mathbf{y}) = \sum_{\nu \in \Lambda_M} \left(\triangle_1^{\nu_1} \otimes \cdots \otimes \triangle_J^{\nu_J} \right) g(\mathbf{y}) , \tag{24}$$

where Λ_M is a downward closed index set defined in Theorem 1. As $\Lambda_1 \subset \cdots \subset \Lambda_M$ and the interpolation nodes are nested, the aSG formula (24) can be rewritten as

$$\mathscr{S}_{\Lambda_M} g(\mathbf{y}) = \sum_{m=1}^{M} \sum_{\mathbf{y}_\mathbf{k}^{\nu^m} \in \Xi_\triangle^{\nu^m}} \underbrace{\left(g(\mathbf{y}_\mathbf{k}^{\nu^m}) - \mathscr{S}_{\Lambda_{m-1}} g(\mathbf{y}_\mathbf{k}^{\nu^m}) \right)}_{\mathbf{s}_\mathbf{k}^{\nu^m}} \ell_\mathbf{k}^{\nu^m}(\mathbf{y}) , \tag{25}$$

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

More explicitly, we start from the initial index $\nu = \mathbf{1} = (1, \ldots, 1)$, thus $\Lambda_1 = \{\mathbf{1}\}$, with root node $\mathbf{y} = \mathbf{0} = (0, \ldots, 0)$. We then look for the maximal *active index set* Λ_M^a such that $\Lambda_M \cup \{\nu\}$ remains downward closed for any $\nu \in \Lambda_M^a$, 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 is one and all other entries are zeros. For each $\nu \in \Lambda_M^a$, we evaluate the errors of the interpolation $\mathscr{S}_{\Lambda_M} g$ at the nodes $\Xi_\triangle^\nu$, and enrich the index set $\Lambda_{M+1} = \Lambda_M \cup \{\nu^{M+1}\}$ with the new index

$$\nu^{M+1} := \operatorname*{argmax}_{\nu \in \Lambda_M^a} \max_{\mathbf{y}_\mathbf{k}^\nu \in \Xi_\triangle^\nu} \frac{1}{|\Xi_\triangle^\nu|} |\mathbf{s}_\mathbf{k}^\nu| , \tag{26}$$

where the error is balanced by the work measured in the number of new nodes $|\Xi_\triangle^\nu|$. An adaptive sparse grid quadrature can be constructed similar to (24) as

$$\mathbb{E}[g] \approx \mathbb{E}[\mathscr{S}_{\Lambda_M} g] = \sum_{m=1}^{M} \sum_{\mathbf{y}_\mathbf{k}^{\nu^m} \in \Xi_\triangle^{\nu^m}} \mathbf{s}_\mathbf{k}^{\nu^m} \mathbf{w}_\mathbf{k}^{\nu^m}, \quad \text{being } \mathbf{w}_\mathbf{k}^{\nu^m} = \mathbb{E}[\ell_\mathbf{k}^{\nu^m}] , \tag{27}$$

for which can enrich the index set with the new index

$$\boldsymbol{\nu}^{M+1} := \underset{\boldsymbol{\nu}\in\Lambda_M^a}{\operatorname{argmax}} \frac{1}{|\Xi_{\triangle}^{\boldsymbol{\nu}}|} \left| \sum_{\mathbf{y}_{\mathbf{k}}^{\boldsymbol{\nu}}\in\Xi_{\triangle}^{\boldsymbol{\nu}}} \mathbf{s}_{\mathbf{k}}^{\boldsymbol{\nu}} \mathbf{w}_{\mathbf{k}}^{\boldsymbol{\nu}} \right| . \tag{28}$$

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

$$\mathscr{E}_i := \max_{\boldsymbol{\nu}\in\Lambda_M^a} \max_{\mathbf{y}_{\mathbf{k}}^{\boldsymbol{\nu}}\in\Xi_{\triangle}^{\boldsymbol{\nu}}} |\mathbf{s}_{\mathbf{k}}^{\boldsymbol{\nu}}| \quad \text{and} \quad \mathscr{E}_e := \left| \sum_{\boldsymbol{\nu}\in\Lambda_M^a} \sum_{\mathbf{y}_{\mathbf{k}}^{\boldsymbol{\nu}}\in\Xi_{\triangle}^{\boldsymbol{\nu}}} \mathbf{s}_{\mathbf{k}}^{\boldsymbol{\nu}} \mathbf{w}_{\mathbf{k}}^{\boldsymbol{\nu}} \right| . \tag{29}$$

The following convergence results can be obtained for the aSG interpolation and integration errors based on that for gPC approximation in Theorem 1, see [12, 25].

Theorem 2. *Under Assumption 1, there exists a downward closed set Λ_M such that the interpolation error*

$$\sup_{\mathbf{y}\in U} ||q(\mathbf{y}) - \mathscr{S}_{\Lambda_M} q(\mathbf{y})||_{\mathscr{X}} \le C_i M^{-s}, \quad s = \frac{1}{p} - 1 , \tag{30}$$

where C_i is independent of M. Analogously, there exists a dc set Λ_M such that the integration error

$$||\mathbb{E}^{\pi_0}[q] - \mathbb{E}^{\pi_0}[\mathscr{S}_{\Lambda_M} q]||_{\mathscr{X}} \le C_e M^{-s}, \quad s = \frac{1}{p} - 1; , \tag{31}$$

where C_e is independent of M. The same convergence rate holds also for the aSG interpolation and integration errors of the posterior density Θ and the QoI Ψ.

4 Model Order Reduction

The evaluation of the posterior density Θ, the renormalization constant Z, the QoI Ψ as well as its statistics, e.g. $\mathbb{E}^{\pi^\delta}[\Psi]$, requires the solutions of the forward parametric equation (10) at many interpolation or integration nodes $\mathbf{y} \in U$ by the aSG algorithms. This section is devoted to the development of model order reduction techniques to effectively reduce the computational cost for the forward solutions.

4.1 High-Fidelity Petrov-Galerkin Approximation

For the solution of the forward parametric problem (10) at any given parameter $\mathbf{y}$, we introduce the finite-dimensional trial space $\mathscr{X}_h$ and test space $\mathscr{Y}_h$, with $\dim(\mathscr{X}_h) = \dim(\mathscr{Y}_h) = \mathscr{N}$, $\mathscr{N} \in \mathbb{N}$. Here h denotes a discretization parameter, such as the mesh width of finite element discretization or the reciprocal of polynomial degree

for spectral discretization. The Petrov–Galerkin (PG), high-fidelity (HiFi for short) approximation of problem (10) reads: for any $\mathbf{y} \in U$, find $q_h(\mathbf{y}) \in \mathscr{X}_h$ such that

$$a(\mathbf{y}; q_h(\mathbf{y}), v_h) = f(\mathbf{y}; v_h) \quad \forall v_h \in \mathscr{Y}_h \tag{32}$$

We proceed under the hypothesis that Assumption 1 holds also in the finite-dimensional spaces $\mathscr{X}_h$ and $\mathscr{Y}_h$, in particular the inf-sup condition (13) is satisfied with constant $\beta_h > 0$ uniformly w.r. to $\mathbf{y}$. The parametric Bayesian posterior density $\Theta(\mathbf{y})$ in (5) can then be approximated by

$$\Theta_h(\mathbf{y}) = \exp\left(-\frac{1}{2}(\delta - \mathscr{O}_h(q_h(\mathbf{y})))^\top \Gamma^{-1}(\delta - \mathscr{O}_h(q_h(\mathbf{y})))\right), \tag{33}$$

where $\mathscr{O}_h$ represents the finite-dimensional approximation of the observation functional $\mathscr{O}$. Similarly, the QoI Ψ can be approximated by the corresponding quantity Ψ_h. Under Assumption 1 in $\mathscr{X}_h$ and $\mathscr{Y}_h$, the well-posedness and gPC as well as the aSG approximation properties in Theorem 1 and Theorem 2 hold with the same convergence rates. In order to compute an approximation subject to a prescribed error tolerances for the quantities q, Θ and Ψ, the dimension $\mathscr{N}$ of the finite-dimensional spaces used in the PG approximation problem (32) is, typically, large. Thus, the numerical solution of the HiFi problem (32) is generally expensive, rendering the aSG approximation that requires one solution at each of many interpolation/integration nodes computationally unfeasible in many cases. To reduce it, we propose a model order reduction technique based on reduced basis (RB) approximations constructed by a greedy algorithm with goal-oriented a-posteriori error estimation and Offline-Online decomposition [22, 24, 3, 9].

4.2 Reduced Basis Approximation

4.2.1 Reduced Basis Construction

Analogous to the HiFi -PG approximation, we look for a RB trial space $\mathscr{X}_N \subset \mathscr{X}_h$ and a RB test space $\mathscr{Y}_N \subset \mathscr{Y}_h$ with $\dim(\mathscr{X}_N) = \dim(\mathscr{Y}_N) = N$, $N \in \mathbb{N}$, $N << \mathscr{N}$.

Then we approximate the forward solution map by solving a PG-RB problem: for any $\mathbf{y} \in U$, find $q_N(\mathbf{y}) \in \mathscr{X}_N$ such that

$$a(\mathbf{y}; q_N(\mathbf{y}), v_N) = f(\mathbf{y}; v_N) \quad \forall v_N \in \mathscr{Y}_N\,. \tag{34}$$

For accurate and efficient approximation of the solution manifold $\mathscr{M}_h = \{q_h(\mathbf{y}), \mathbf{y} \in U\}$, RB takes the HiFi solutions $q_h(\mathbf{y})$ at N carefully chosen parameter values $\mathbf{y} = \mathbf{y}^n$, $1 \le n \le N$, called *snapshots*, as the basis functions of the trial space, i.e.

$$\mathscr{X}_N = \text{span}\{q_h(\mathbf{y}^n), 1 \le n \le N\}\,. \tag{35}$$

In order to select "most representative snapshots" for the approximation of the posterior density Θ (or the QoI Ψ, which can be approximated in the same way),

$$\Theta_N(\mathbf{y}) = \exp\left(-\frac{1}{2}(\delta - \mathscr{O}_h(q_N(\mathbf{y})))^\top \Gamma^{-1}(\delta - \mathscr{O}_h(q_N(\mathbf{y})))\right) , \tag{36}$$

which is nonlinear with respect to the solution q_N, we propose a greedy algorithm based on a goal-oriented a-posteriori error estimator $\triangle_N^\Theta(\mathbf{y})$ for the RB approximation error of the posterior density, $|\Theta_h(\mathbf{y}) - \Theta_N(\mathbf{y})|$ for any $\mathbf{y} \in U$. We start with the first parameter value $\mathbf{y}^1$, e.g. the center of U or a random sample, and construct the initial RB trial space as $\mathscr{X}_1 = \text{span}\{q_h(\mathbf{y}^1)\}$. Then, for $N = 1, 2, \ldots$, we pick the next parameter value by

$$\mathbf{y}^{N+1} := \underset{\mathbf{y}\in U}{\text{argmax}}\, \triangle_N^\Theta(\mathbf{y}) , \tag{37}$$

and enrich the RB space as $\mathscr{X}_{N+1} = \mathscr{X}_N \oplus \text{span}\{q_h(\mathbf{y}^{N+1})\}$. In practice, instead of solving a high-dimensional optimization problem (37), we can replace the parameter domain U by a suitable training set Ξ_{train}, e.g. the sparse grid nodes. The basis functions for the test space $\mathscr{Y}_N$ are chosen such that the PG-RB approximation is stable. In the case that the bilinear form $a(\mathbf{y};\cdot,\cdot)$ is coercive in $\mathscr{X}_h \times \mathscr{Y}_h$ for $\mathscr{Y}_h = \mathscr{X}_h$, the choice $\mathscr{Y}_N = \mathscr{X}_N$ satisfies the stability condition for the PG-RB approximation. For noncoercive problems, we construct the RB test space through the *supremizer operator* $T^\mathbf{y} : \mathscr{X}_h \to \mathscr{Y}_h$ defined as

$$(T^\mathbf{y} w_h, v_h)_\mathscr{Y} = a(\mathbf{y}; w_h, v_h) \quad \forall v_h \in \mathscr{Y}_h . \tag{38}$$

Then $T^\mathbf{y} w_h \in \mathscr{Y}_h$ is the supremizer for the element $w_h \in \mathscr{X}_h$ with respect to the functional $a(\mathbf{y}; w_h, \cdot) : \mathscr{Y}_h \to \mathbb{R}$, i.e.

$$T^\mathbf{y} w_h = \underset{v_h\in\mathscr{Y}_h}{\text{argsup}} \frac{a(\mathbf{y}; w_h, v_h)}{||w_h||_\mathscr{X} ||v_h||_\mathscr{Y}} . \tag{39}$$

For any $\mathbf{y} \in U$, the $\mathbf{y}$-dependent RB test space $\mathscr{Y}_N^\mathbf{y}$ is defined as

$$\mathscr{Y}_N^\mathbf{y} := \text{span}\{T^\mathbf{y} q_h(\mathbf{y}^n), 1 \le n \le N\} . \tag{40}$$

It can be shown (see [9]) that

$$\beta_N(\mathbf{y}) := \inf_{0\neq w_N\in\mathscr{X}_N} \sup_{v_N\in\mathscr{Y}_N^\mathbf{y}} \frac{|a(\mathbf{y}; w_N, v_N)|}{||w_N||_\mathscr{X} ||v_N||_\mathscr{Y}} \ge \beta_h > 0 , \tag{41}$$

ie., the PG-RB approximation problem (34) is uniformly well-posed w.r.to $\mathbf{y} \in U$.

4.2.2 A-Posteriori Error Estimator

The goal-oriented a-posteriori error estimator $\triangle_N^{\Theta}$ plays a key role in constructing the RB spaces, which should be reliable and efficient, i.e. there exist two constants $0 < c_{\triangle} \leq C_{\triangle} < \infty$ such that

$$c_{\triangle}|\Theta_h(\mathbf{y}) - \Theta_N(\mathbf{y})| \leq \triangle_N^{\Theta}(\mathbf{y}) \leq C_{\triangle}|\Theta_h(\mathbf{y}) - \Theta_N(\mathbf{y})| \,. \tag{42}$$

As we can view the function $\Theta_h : U \to \mathbb{R}$ as a functional $\Theta_h(\cdot) : \mathscr{X}_h \to \mathbb{R}$ through $\Theta(\mathbf{y}) = \Theta_h(q_h(\mathbf{y}))$, following the derivation in [9], smooth dependence of the posterior on the parameters in the forward map implies a formal Taylor expansion of $\Theta_h(q_h(\mathbf{y}))$ about $q_N(\mathbf{y})$:

$$\Theta_h(q_h(\mathbf{y})) = \Theta_h(q_N(\mathbf{y})) + \frac{\partial \Theta_h}{\partial q_h}\Big|_{q_N(\mathbf{y})}(q_h(\mathbf{y}) - q_N(\mathbf{y})) + O(||q_h(\mathbf{y}) - q_N(\mathbf{y})||^2_{\mathscr{X}}) \,, \tag{43}$$

where the second term of the right hand side is the Fréchet derivative of Θ_h at $q_N(\mathbf{y})$ with respect to q_h, evaluated at the error $e_N^h(\mathbf{y}) = q_h(\mathbf{y}) - q_N(\mathbf{y})$. As the first term $\Theta_h(q_N(\mathbf{y})) = \Theta_N(\mathbf{y})$, as long as the last term is dominated by the second term, we can define the error estimator for $|\Theta_h(\mathbf{y}) - \Theta_N(\mathbf{y})|$ as the second term in (43),i.e.

$$\triangle_{N,h}^{\Theta}(\mathbf{y}) := \frac{\partial \Theta_h}{\partial q_h}\Big|_{q_N(\mathbf{y})}(e_N^h(\mathbf{y})) \,. \tag{44}$$

In order to evaluate $\triangle_{N,h}^{\Theta}(\mathbf{y})$ more efficiently, we propose a dual HiFi PG approximation [22, 9]: for any $\mathbf{y} \in U$, find the dual solution $\varphi_h(\mathbf{y}) \in \mathscr{Y}_h$ such that

$$a(\mathbf{y}; w_h, \varphi_h(\mathbf{y})) = \frac{\partial \Theta_h}{\partial q_h}\Big|_{q_N(\mathbf{y})}(w_h) \quad \forall w_h \in \mathscr{X}_h \,. \tag{45}$$

Then, with the definition of the residual for the primal HiFi problem (32) evaluated at the RB solution of (34), i.e.

$$r(\mathbf{y}; v_h) := f(\mathbf{y}; v_h) - a(\mathbf{y}; q_N(\mathbf{y}), v_n) \quad \forall v_h \in \mathscr{Y}_h \,, \tag{46}$$

we obtain, as the primal HiFi equation (32) holds for $\varphi_h \in \mathscr{Y}_h$,

$$r(\mathbf{y}; \varphi_h(\mathbf{y})) = f(\mathbf{y}; \varphi_h(\mathbf{y})) - a(\mathbf{y}; q_N(\mathbf{y}), \varphi_h(\mathbf{y})) = a(\mathbf{y}; e_N^h(\mathbf{y}), \varphi_h(\mathbf{y})) \,, \tag{47}$$

which, together with definition (44) and (45), imply

$$\triangle_{N,h}^{\Theta}(\mathbf{y}) = r(\mathbf{y}; \varphi_h(\mathbf{y})) \,. \tag{48}$$

As it is computationally expensive to obtain the solution $\varphi_h(\mathbf{y})$, we propose to use RB approximation for the HiFi -PG approximation of the dual problem (45) following the same development as for the primal HiFi problem in the last section. With the dual RB solution $\varphi_N(\mathbf{y})$ (where number N of degrees of freedom of the dual problem

could be different from N which was used in the RB-PG approximation of the primal problem), we define the a-posteriori error estimator for the error $|\Theta_h(\mathbf{y}) - \Theta_N(\mathbf{y})|$ as

$$\triangle_N^{\Theta}(\mathbf{y}) = r(\mathbf{y}; \varphi_N(\mathbf{y})) \,, \tag{49}$$

whose difference from $\triangle_h^{\Theta}(\mathbf{y})$ can be bounded by

$$|\triangle_h^{\Theta}(\mathbf{y}) - \triangle_N^{\Theta}(\mathbf{y})| = r(\mathbf{y}; \varepsilon_N^h(\mathbf{y})) = a(\mathbf{y}; e_N^h(\mathbf{y}), \varepsilon_N^h(\mathbf{y})) \leq \gamma ||e_N^h(\mathbf{y})||_{\mathscr{X}} ||\varepsilon_N^h(\mathbf{y})||_{\mathscr{Y}} \,, \tag{50}$$

where $\varepsilon_N^h(\mathbf{y}) = \varphi_h(\mathbf{y}) - \varphi_N(\mathbf{y})$ and γ represents the continuity constant of the bilinear form a. In general, the primal and dual RB errors $e_N^h(\mathbf{y})$ and $\varepsilon_N^h(\mathbf{y})$ tend to zero so that, asymptotically, (50) and the second order term in (43) are both dominated by the first order term of (43), we can expect to obtain a reliable and efficient, computable a-posteriori error estimator $\triangle_N^{\Theta}(\mathbf{y})$ for the error $|\Theta_h(\mathbf{y}) - \Theta_N(\mathbf{y})|$, with the corresponding constants $c_{\triangle}$ and $C_{\triangle}$ in (42) close to one uniformly w.r. to $\mathbf{y}$.

4.2.3 Offline-Online Computation

To this end, we make a crucial assumption that the HiFi PG discretization of the parametric problem, (32) is affine, i.e. $\forall w_h \in \mathscr{X}_h, v_h \in \mathscr{Y}_h$, the bilinear and linear forms can be written as

$$a(\mathbf{y}; w_h, v_h) = \sum_{m=1}^{M_a} \lambda_m^a(\mathbf{y}) a_m(w_h, v_h) \quad \text{and} \quad f(\mathbf{y}; v_h) = \sum_{m=1}^{M_f} \lambda_m^f(\mathbf{y}) f_m(v_h) \,. \tag{51}$$

For instance, for a diffusion problem with affine-parametric diffusion coefficient (2), we have $\lambda_m^a(\mathbf{y}) = y_m$ and $a_m(w_h, v_h) = (\phi_m \nabla w_h, \nabla v_h)$, $1 \leq m \leq M_a = J$. We defer the discussion of linearization in parameter space, ie., the approximation of the non-affine parametric problem by an affine parametric model in (51).

For the sake of algebraic stability of the PG-RB approximation (34), we compute the orthonormal bases $(w_N^n)_{n=1}^N$ of $\mathscr{X}_N$ obtained by Gram–Schmidt orthonormalization algorithm for the bases $(q_h(\mathbf{y}^n))_{n=1}^N$. Then the RB solution of problem (34) at any $\mathbf{y} \in U$ can be represented by

$$q_N(\mathbf{y}) = \sum_{n=1}^{N} q_N^n(\mathbf{y}) w_N^n \,, \tag{52}$$

where $\mathbf{q}_N(\mathbf{y}) = (\mathbf{q}_N^1(\mathbf{y}), \ldots, \mathbf{q}_N^N(\mathbf{y}))^{\top} \in \mathbb{R}^N$, denoting the coefficient of $q_N(\mathbf{y})$. In the coercive case where $\mathscr{Y}_N = \mathscr{X}_N$ with basis $v_N^n = w_N^n$, $1 \leq n \leq N$, the algebraic system of the PG-RB problem (34) becomes

$$\left(\sum_{m=1}^{M_a} \lambda_m^a(\mathbf{y}) A_m \right) \mathbf{q}_N(\mathbf{y}) = \sum_{m=1}^{M_f} \lambda_m^f(\mathbf{y}) \mathbf{f}_m \,, \tag{53}$$

where the RB matrix A_m, $1 \le m \le M_a$, and the RB vector $\mathbf{f}_m$, $1 \le m \le M_f$, are given respectively by

$$(A_m)_{n',n} = a_m(w_N^n, v_N^{n'}) \quad \text{and} \quad (\mathbf{f}_m)_n = f_m(v_N^n) \quad n,n' = 1,\dots,N\,, \tag{54}$$

which do not depend on the parameter $\mathbf{y} \in U$ and can therefore be assembled and stored once and for all in the Offline stage. Given any $\mathbf{y} \in U$, the algebraic system (53) can be assembled and solved Online with $O(M_a N^2 + M_f N)$ and $O(N^3)$ operations, respectively, which do not depend on the number $\mathscr{N}$ of high-fidelity degrees of freedom. In the noncoercive case, for any $\mathbf{y} \in U$, the test basis v_N^n, $1 \le n \le N$, is given by

$$v_N^n = T^{\mathbf{y}} w_N^n = \sum_{m=1}^{M_a} \lambda_m^a(\mathbf{y}) T_m w_N^n\,, \tag{55}$$

where $T_m w_N^n$, $1 \le m \le M_a$, $1 \le n \le N$, is the solution of

$$(T_m w_N^n, v_h)_{\mathscr{Y}} = a_m(w_N^n, v_h) \quad \forall v_h \in \mathscr{Y}_h\,, \tag{56}$$

which does not depend on $\mathbf{y} \in U$ and which can be computed and stored once and for all during the Offline stage. The corresponding algebraic system of the PG-RB problem (34) is given by

$$\left(\sum_{m}^{M_a} \sum_{m'}^{M_a} \lambda_m^a(\mathbf{y}) \lambda_{m'}^a(\mathbf{y}) A_{m,m'} \right) \mathbf{q}_N(\mathbf{y}) = \sum_{m=1}^{M_f} \lambda_m^f(\mathbf{y}) \mathbf{f}_m\,, \tag{57}$$

where the (densely populated) RB matrix $A_{m,m'}$, $1 \le m,m' \le M_a$, is given by

$$(A_{m,m'})_{n',n} = a_m(w_N^n, T_{m'} w_N^{n'}) \quad 1 \le n,n' \le N\,. \tag{58}$$

This matrix does not depend on $\mathbf{y}$ and can be computed and stored once and for all during the Office stage. Given any $\mathbf{y} \in U$, the algebraic system (57) can be assembled and solved Online in $O(M_a^2 N^2 + M_f N)$ and $O(N^3)$ operations.

The dual RB solution $\varphi_N(\mathbf{y})$ can be computed by the same Offline-Online procedure. The a-posteriori error estimator (49) takes the explicit form

$$\begin{aligned} \triangle_N^{\Theta}(\mathbf{y}) = & \sum_{m=1}^{M_f} \sum_{n=1}^{N} \lambda_m^f(\mathbf{y}) f_m(v_N^{n,du}) \varphi_N^n(\mathbf{y}) \\ & - \sum_{m=1}^{M_a} \sum_{n=1}^{N} \sum_{n'=1}^{N} \lambda_m^a(\mathbf{y}) q_N^n(\mathbf{y}) a_m(w_N^n, v_N^{n,du}) \varphi_N^{n'}(\mathbf{y})\,, \end{aligned} \tag{59}$$

where $\varphi_N^{n'}(\mathbf{y})$ is the coefficient of the dual RB solution $\varphi_N(\mathbf{y})$ on the trial RB basis $v_N^{n',du} \in \mathscr{Y}_N^{du}$, $1 \le n' \le N$. As $f_m(v_N^{n,du})$, $1 \le m \le M_f$ and $a_m(w_N^n, v_N^{n,du})$, $1 \le m \le M_a$, $1 \le n,n' \le N$, are independent of $\mathbf{y}$, they can be computed and stored once during

the Offline stage and the error estimator (59) can be assembled during the Online stage for any given $\mathbf{y} \in U$ with $O(M_f N + M_a N^2)$ operations.

Finally, the RB posterior density $\Theta_N(\mathbf{y})$ can be computed by

$$\Theta_N(\mathbf{y}) = \exp\left(-\frac{1}{2}\left(\delta - \mathscr{O}_K^N \mathbf{q}_N(\mathbf{y})\right)^\top \Gamma^{-1}\left(\delta - \mathscr{O}_K^N \mathbf{q}_N(\mathbf{y})\right)\right), \tag{60}$$

where the observation matrix $\mathscr{O}_K^N \in \mathbb{R}^{K\times N}$ with elements $(\mathscr{O}_K^N)_{k,n} = o_k(w_N^n)$, $1 \le k \le K$, $1 \le n \le N$, is computed and stored for once during Offline stage and $\Theta_N(\mathbf{y})$ is assembled for any $\mathbf{y} \in U$ during the Online stage in $O(NK^2)$ operations.

As the error estimator $\triangle_N^\Theta(\mathbf{y})$ is an approximation of the second term in the Taylor expansion (43) for $\Theta_h(\mathbf{y})$, we correct the RB approximation $\Theta_N(\mathbf{y})$ by

$$\Theta_N^\triangle(\mathbf{y}) = \Theta_N(\mathbf{y}) + \triangle_N^\Theta(\mathbf{y}), \tag{61}$$

which is generally more accurate than $\Theta_N(\mathbf{y})$.

Theorem 3. *([9]) Under Assumption 1, the RB error for the posterior density satisfies*

$$\sup_{\mathbf{y}\in U} |\Theta_h(\mathbf{y}) - \Theta_N^\triangle(\mathbf{y})| \le C_\Theta^\triangle N^{-2s}, \quad s = \frac{1}{p} - 1, \tag{62}$$

where the constant $C_\Theta^\triangle$ is independent of the number of RB bases N and the active dimension J. The same convergence rate holds for RB approximation of the QoI Ψ.

4.3 Empirical Interpolation Method (EIM)

As the computational reduction due to the $\mathscr{N}$-independent Online RB evaluation crucially depends on the assumption (51), which is however not necessarily valid in practice: we mention only diffusion problems with lognormal diffusion coefficient given by $\kappa = e^u$. We outline the Empirical Interpolation Method (EIM) for affine-parametric approximation of problems with nonaffine parameter dependence. More precisely, suppose $\mathscr{X}_h$ is defined in the domain $D \subset \mathbb{R}^d$, $d \in \mathbb{N}$, with the finite set of discretization nodes $D_h \in D$, we seek to approximate an arbitrary, non-affine function $g : D_h \times U \to \mathbb{R}$ in the bilinear and linear forms by

$$g(x,\mathbf{y}) \approx \mathscr{I}_M[g](x,\mathbf{y}) = \sum_m^M \lambda_m(\mathbf{y}) g_m(x), \tag{63}$$

which results in an approximation of the problem (32) with affine representation (51). For instance, when g is the diffusion coefficient of a diffusion problem, we obtain (51) with $\lambda_m^a(\mathbf{y}) = \lambda_m(\mathbf{y})$ and $a_m(w_h, v_h) = (g_m \nabla w_h, \nabla v_h)$, $1 \le m \le M_a = M$.

One choice for the approximation (63) is by the aSG interpolation based on some structured interpolation nodes, e.g. Leja nodes or Clenshaw-Curtis nodes, presented

in Section 3. As the work for each Online RB evaluation is proportional to the number M of affine terms, it is important to keep M as small as possible. To this end, we propose an adaptive construction of a sparse interpolation set by the following greedy algorithm. We start by searching for the first parameter value $\mathbf{y}^1 \in U$ and the first discretization node $x^1 \in D_h$ such that

$$\mathbf{y}^1 = \operatorname*{argsup}_{\mathbf{y}\in U} \max_{x\in D_h} |g(x,\mathbf{y})| \text{ and } x^1 = \operatorname*{argmax}_{x\in D_h} |g(x,\mathbf{y}^1)| . \tag{64}$$

The first basis g_1 is taken as $g_1(x) = g(x,\mathbf{y}^1)/g(x^1,\mathbf{y}^1)$, $x \in D_h$. We define the EIM node set $S_1 = \{x^1\}$. For $M = 1,2,\ldots$, for any $\mathbf{y} \in U$, the coefficient $\lambda_m(\mathbf{y})$, $1 \le m \le M$, of the interpolation (63) is obtained by Lagrange interpolation at the selected discretization nodes, ie.

$$\sum_{m=1}^{M} \lambda_m(\mathbf{y}) g_m(x) = g(x,\mathbf{y}) \quad \forall x \in S_M . \tag{65}$$

Then we define the *empirical interpolation residual* as

$$r_{M+1}(x,\mathbf{y}) = g(x,\mathbf{y}) - \sum_{m=1}^{M} \lambda_m(\mathbf{y}) g_m(x) . \tag{66}$$

The next parameter sample $\mathbf{y}^{M+1}$ and discretization node x^{M+1} are chosen as

$$\mathbf{y}^{M+1} = \operatorname*{argsup}_{\mathbf{y}\in U} \max_{x\in D_h} |r_{M+1}(\mathbf{y})| \quad \text{and} \quad x^{M+1} = \operatorname*{argmax}_{x\in D_h} |r_{M+1}(x,\mathbf{y}^{M+1})| . \tag{67}$$

We define $S_{M+1} = S_M \cup \{x^{M+1}\}$ and take the next basis function g_{M+1} as

$$g_{M+1}(x) = \frac{r_{M+1}(x,\mathbf{y}^{M+1})}{r_{M+1}(x^{M+1},\mathbf{y}^{M+1})} \quad x \in D_h . \tag{68}$$

We remark that in practice the parameter domain U is replaced with a finite training set Ξ_{train} to avoid directly solving a maximization problem (67), which is very expensive in a high-dimensional parameter domain. Details and error bounds will be available in [10].

4.4 Adaptive aSG-EIM-RB Algorithm

In this section, we propose an adaptive algorithm for the evaluations of the posterior density Θ as well as its expectation Z for Bayesian inversion with nonaffine forward map by incorporation of approximations of aSG, EIM and RB in order to reduce the total computational cost. The same algorithm applies for the evaluation of the QoI Ψ and its statistical moments as well. The basic idea is that at each step of the construction of aSG with new interpolation or integration nodes, we refine the EIM

approximation of the nonaffine parametric function and refine the RB approximation of Θ when their approximation errors are larger than prescribed tolerances at the new nodes. In the end, instead of solving a large number of HiFi problems for the evaluation of $\Theta_h(\mathbf{y})$ at all aSG nodes, we approximate $\Theta_h(\mathbf{y})$ by the $\Theta_N(\mathbf{y})$ resp. by $\Theta_N^{\triangle}(\mathbf{y})$, which only requires inexpensive RB solutions. The main procedure of simultaneous aSG-EIM-RB construction and evaluation is provided in Algorithm 1.

Algorithm 1: Adaptive aSG-EIM-RB Algorithm

1. Specify the tolerances ε_{aSG}, ε_{EIM} and ε_{RB} and the maximum numbers of nodes M_{aSG}^{max}, M_{EIM}^{max} for aSG, EIM and bases N_{max} for RB approximations, respectively, set $\mathscr{E}_{aSG} = 2\varepsilon_{aSG}$;
2. Initialize the aSG, EIM and RB approximation with $M_{aSG} = M_{EIM} = N = 1$:
   a. solve the primal and dual HiFi problems (32) and (45) at the root node $\mathbf{y}^1 = \mathbf{0} \in U$;
   b. initialize the index set $\Lambda_1 = \{\mathbf{1}\}$, and construct the aSG approximation, either the interpolation as $\mathscr{S}_{\Lambda_1}\Theta_h(\mathbf{y}) = \Theta_h(\mathbf{y}^1)$ or the integration as $\mathbb{E}[\mathscr{S}_{\Lambda_1}\Theta_h] = \Theta_h(\mathbf{y}^1)$;
   c. set the first EIM basis as $\mathscr{I}_1[g](\mathbf{y}) = g(\mathbf{y}^1)$, set $x^1 \in \operatorname{argmax}_{x\in D_h} |g(x,\mathbf{y}^1)|$;
   d. construct the first RB primal trial space $\mathscr{X}_1 = \operatorname{span}\{q_h(\mathbf{y}^1)\}$ and dual trial space $\mathscr{Y}_1^{du} = \operatorname{span}\{\varphi_h(\mathbf{y}^1)\}$, compute and store all quantities in Offline stage.
3. **While** $M_{aSG} < M_{aSG}^{max}$ and $\mathscr{E}_{aSG} > \varepsilon_{aSG}$
   a. compute the active index set $\Lambda_{M_{aSG}}^a$ for the aSG approximation;
   b. **For** each $\nu \in \Lambda_{M_{aSG}}^a$
      i. compute the set of added nodes $\Xi_{\triangle}^{\nu}$ associated to ν;
      ii. **For** each $\mathbf{y} \in \Xi_{\triangle}^{\nu}$
         A. compute EIM interpolation of g at $\mathbf{y}$ and the interpolation error $\mathscr{E}_{EIM}(\mathbf{y})$;
         B. **If** $M_{EIM} < M_{EIM}^{max}$ and $\mathscr{E}_{EIM}(\mathbf{y}) > \varepsilon_{EIM}$
            - refine the EIM interpolation with the new basis $g(\mathbf{y})$, select $x^{M_{EIM}+1}$;
            - set $M_{EIM} = M_{EIM} + 1$;

            EndIf
         C. compute the RB solution and $\Theta_N^{\triangle}(\mathbf{y})$ and the error estimator $\mathscr{E}_{RB}(\mathbf{y}) = \triangle_N^{\Theta}(\mathbf{y})$;
         D. **If** $N < N_{max}$ and $\mathscr{E}_{RB}(\mathbf{y}) > \varepsilon_{RB}$
            - enrich the RB trial spaces $\mathscr{X}_N$ with $q_h(\mathbf{y})$ and $\mathscr{Y}_N^{du}$ with $\varphi_h(\mathbf{y})$;
            - compute and save the all Offline quantities;
            - set $N = N+1$;

            EndIf
      EndFor
   EndFor
   c. compute the aSG error estimator $\mathscr{E}_{aSG}$ as one of (29) with the RB approximation $\Theta_N^{\triangle}$;
   d. enrich $\Lambda_{M_{aSG}}$ by $\nu^{M_{aSG}+1}$ according to (26) for interpolation or (28) for integration;
   e. set $M_{aSG} = M_{aSG} + 1$;

   EndWhile

In the adaptive refinement of EIM interpolation, we may replace the set of discretization nodes D_h in (67), which depends on the HiFi degree of freedom $\mathscr{N}$, by (i) a smaller number of randomly selected discretization nodes in $D_h \setminus S_M$; or (ii) the last s (e.g. $s = 1,2,\dots$) selected nodes $\{x^M, x^{M-1}, x^{M-s+1}\}$ and use the first $M-s$ EIM bases to evaluate the error estimator $\mathscr{E}_{EIM}(\mathbf{y})$.

5 Numerical Experiments

We consider a diffusion problem in the physical domain $D = (0,1)^2$: for $\mathbf{y} \in U$, find the solution $q(\mathbf{y}) \in H_0^1(D)$ such that

$$-div(\kappa(\mathbf{y})\nabla q(\mathbf{y})) = f, \tag{69}$$

where we set $f = 1$ and prescribe homogeneous Dirichlet boundary condition $q(\mathbf{y}) = 0$ on ∂D; the diffusion coefficient is given by

$$\kappa(\mathbf{y}) = e^{u(\mathbf{y})} \text{ with } u(\mathbf{y}) = 1 + \sum_{j=1}^{J} y_j \frac{1}{j^3} \sin((j_1+1)\pi x_1) \sin((j_2+1)\pi x_2)\,, \tag{70}$$

where $j_1, j_2 = 1, \ldots, \sqrt{J}$ such that $j = j_1 + j_2\sqrt{J}$ for a square J; $x = (x_1, x_2) \in D$. Note that $u(\mathbf{y})$ is nonaffine with respect to $\mathbf{y}$. We perform interpolation for an affine decomposition of $\kappa(\mathbf{y})$ by applying both aSG and EIM. We first investigate the convergence of the aSG interpolation error with respect to the number of dimensions J. For simplicity, we only consider the interpolation for the function $\kappa(\mathbf{y})$ at a sample node $x = (0.3, 0.6)$ (interpolation at any other node (or set of nodes) or the worst case scenario measured in $L^\infty(D)$-norm can be performed in the same way, but with much more computational cost for the latter case). We test the cases of $J = 16, 64, 256$, and 1024, and construct the aSG using Clenshaw–Curtis nodes defined in (21) with the maximum number of interpolation nodes set to 10^5. Fig. 1 displays the convergence of the interpolation error estimator defined in (29) with respect to the number of interpolation nodes. We can see that the convergence rate converges to the one close to M^{-2} when the number of dimensions increases from 16 to 1024, which demonstrates the theoretical prediction of the error convergence rate in Theorem 2 for high-(infinite-)dimensional sparse interpolation.

In the numerical convergence study of the empirical interpolation error, we consider the $J = 64$ dimensional case for uniform, triangular meshes with mesh widths $h = 1/16, 1/32, 1/64$, and $1/128$. The tolerance is chosen as 10^{-8} and the same 1000 random samples as the training samples for the construction of EIM are selected. $M = 161, 179, 179$, and 179 EIM bases are constructed for $h = 1/16, 1/32, 1/64$ and $1/128$, respectively. This shows that at a given level of accuracy, the number of EIM bases is independent of HiFi mesh width, provided it is sufficiently fine. We use $h = 1/32$, i.e. with Finite Element nodes $x = (i_1/32, i_2/32)$, $i_1, i_2 = 0, \ldots, 32$, and 1000 random training samples to evaluate the convergence of EIM error with respect to the number of dimensions $J = 16, 64, 256$, and 1024, which is shown in Fig. 2. We observe that as J increases, the convergence rate tends to M^{-2}, as could be expected from the results in the affine-parametric setting in [9]. However, as the number of EIM bases increases beyond the dimension J of the set of active parameters, the convergence for EIM error exceeds the rate M^{-2} and becomes much faster (in fact, exponential) than the aSG error that still converges with a rate close to M^{-2}. This is further demonstrated in the case $J = 64$ in Fig. 3, where the aSG is constructed for interpolation only at the sample node $x = (0.3, 0.6)$ and at the Finite Element nodes

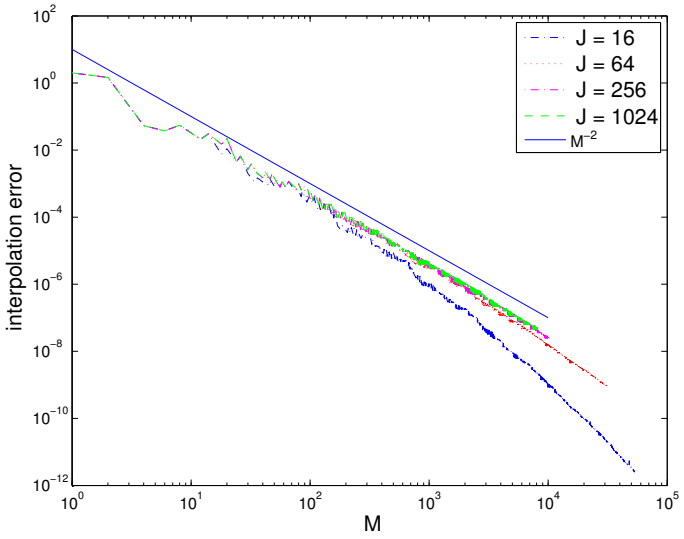

Fig. 1 Decay of the aSG interpolation error with respect to the number of interpolation nodes M for $\kappa(\mathbf{y})$ in $J = 16, 64, 256$ and 1024 dimensions at a sample node $x = (0.3, 0.6)$.

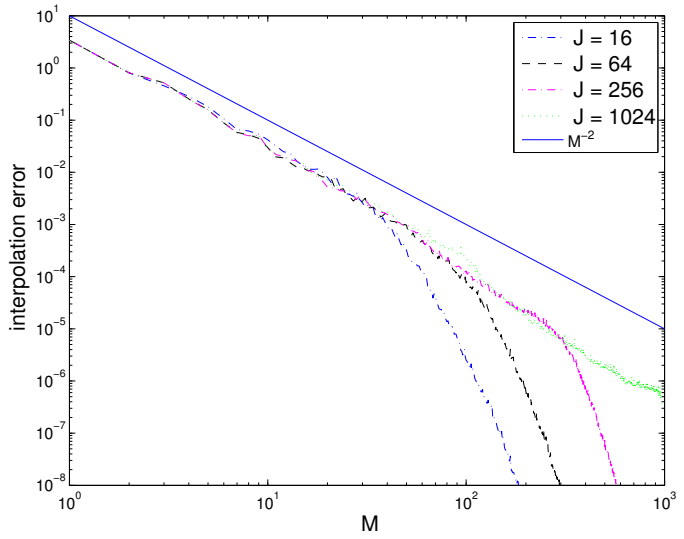

Fig. 2 Decay of the EIM interpolation error with respect to the number of interpolation nodes M for $\kappa(\mathbf{y})$ in $J = 16, 64, 256$ and 1024 dimensions uniformly at all Finite Elment nodes.

$x = (i_1/32, i_2/32)$, $i_1, i_2 = 0, \ldots, 32$, respectively. The EIM bases are constructed with all previously computed aSG nodes (5×10^4) as training samples.

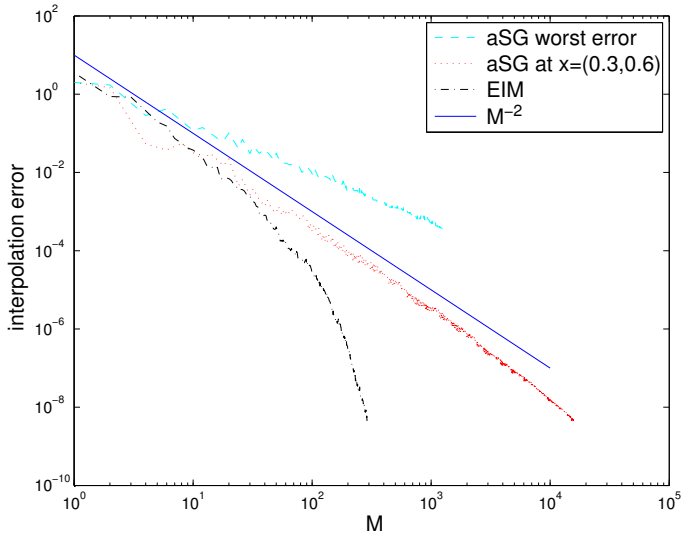

Fig. 3 Decay of interpolation error with respect to the number of interpolation nodes M for $\kappa(\mathbf{y})$ in $J = 64$ dimensions by aSG at the sample node $x = (0.3, 0.6)$ (aSG at x=(0.3,0.6)) and at the Finite Element nodes (aSG worst error), and by EIM at the Finite Element nodes.

From Fig. 3 we see that the worst aSG interpolation error over all mesh nodes decays at a lower rate (with rate about $M^{-1.2}$) than the theoretical prediction M^{-2} in Theorem 2 and that of aSG at only one sample node. This indicates that the aSG constructed to minimize the maximal interpolation error over all mesh nodes can produce approximations which do not converge at the rate afforded by the N-approximation results.

We also see that, in order to achieve the same interpolation accuracy, a much smaller number of EIM bases is needed compared to that of aSG nodes. For example, only 50 EIM bases are needed in order to achieve the same accuracy 3.7×10^{-4} as for the worst case scenario aSG that requires 1248 interpolation nodes, while 289 EIM bases are needed to attain the interpolation accuracy 4.5×10^{-9}, for which about 1.3×10^7 interpolation nodes are expected (according to the estimated convergence rate $M^{-1.2}$) for the worst case scenario aSG, even only 15748 nodes are needed for aSG interpolation at a single mesh sample point $x = (0.3, 0.6)$. Therefore, in the affine approximation of the nonaffine function $\kappa(\mathbf{y})$ for this example with $J = 64$ parameters, EIM is much more efficient than aSG. For the higher dimensional case, e.g. for $J = 1024$, the same conclusion can be drawn as the worst aSG interpolation error converges at a lower rate (about $M^{-1.2}$) than EIM, which converges at a rate of about M^{-2} when the number of EIM bases is smaller than J and much faster than M^{-2} when the number of EIM bases becomes larger than J.

To study the convergence of the RB errors and the error estimator as well as its effectivity for the approximation of the posterior density Θ in different dimensions,

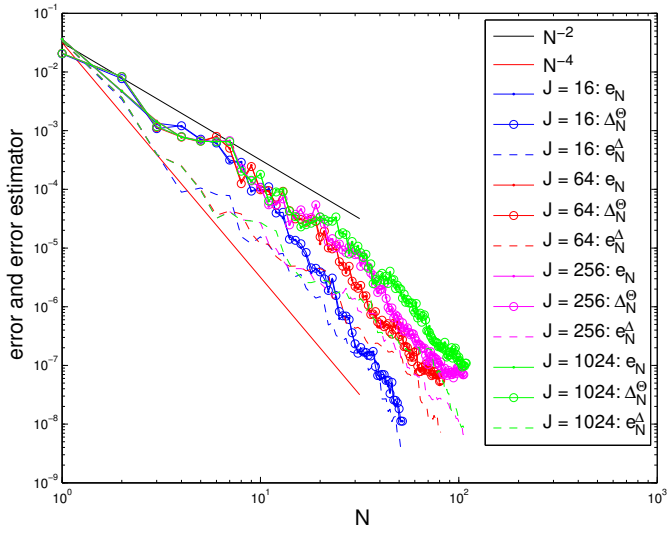

Fig. 4 Decay of the RB approximation errors $e_N(\mathbf{y}_{max})$, $e_N^{\Delta}(\mathbf{y}_{max})$, and the RB error estimator $\triangle_N^{\Theta}(\mathbf{y}_{max})$ with respect to the number of RB bases N in $J = 16, 64, 256$ and 1024 dimensions. The training set consists of 1000 random samples for the construction of RB approximation with EIM and RB tolerances set as 10^{-8}. The test set Ξ_{test} consists of another 100 random samples.

here in particular $J = 16, 64, 256$ and 1024, we first construct the EIM approximation of the nonaffine random field using 1000 random samples with tolerance 10^{-8} (selected so small that EIM interpolation error is dominated by the RB error). We next construct the RB approximation for the posterior density using the same 1000 samples with tolerance 10^{-8}. Then, the RB approximation errors of the posterior density, defined as $e_N = |\Theta_h(\mathbf{y}_{max}) - \Theta_N(\mathbf{y}_{max})|$, where $\mathbf{y}_{max} = \text{argmax}_{\mathbf{y}\in\Xi_{test}} |\Theta_h(\mathbf{y}) - \Theta_N(\mathbf{y})|$, $e_N^{\Delta} = |\Theta_h(\mathbf{y}_{max}) - \Theta_N^{\triangle}(\mathbf{y}_{max})|$, and the RB error estimator $\triangle_N^{\Theta}(\mathbf{y}_{max})$ defined in (49), are computed in a test set Ξ_{test} with 100 random samples that are independent of the 1000 training samples. Fig. 4 displays the convergence of the RB errors and the error estimator with respect to the number of RB bases in different dimensions. We can see that the RB error e_N can hardly be distinguished from the error estimator $\triangle_N^{\Theta}$, which implies that the error estimator is very effective. As parameter space dimension J increases, the approximation error becomes larger. The corrected density $\Theta_N^{\triangle}$ is more accurate than Θ_N in all cases, especially when N is relatively small. In fact, a convergence rate N^{-2} can be observed for e_N compared to N^{-4} for e_N^{Δ} when N is small. When N and J become larger, both errors converge with *a dimension-independent, asymptotic convergence rate* N^{-4}, which is in complete agreement with Theorem 3.

In the last experiment, we consider the influence of the tolerance for RB training to the accuracy of the RB approximation of the posterior density Θ and its integration Z using the aSG-EIM-RB Algorithm 1, where we set the maximum number of the

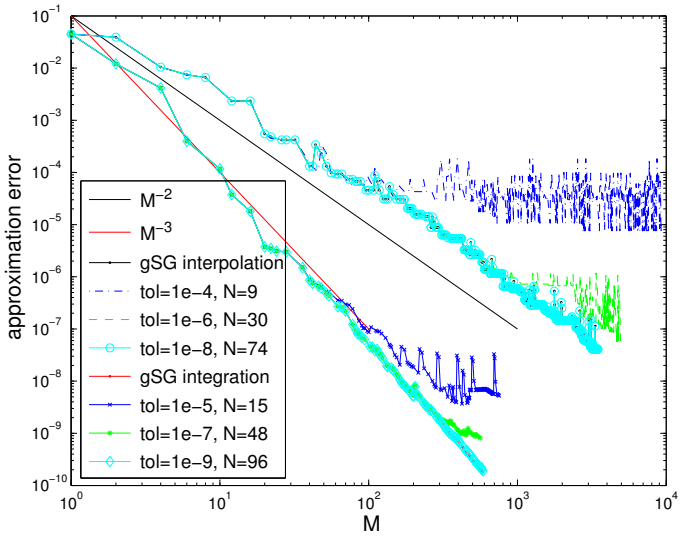

Fig. 5 Decay of aSG interpolation and integration errors with respect to the number of aSG nodes in $J = 64$ dimensions; RB is trained adaptively at the aSG nodes with different tolerances.

sparse grid nodes as 10^4 and 2×10^3 for the interpolation of Θ and the integration of Z, respectively, and set the tolerance for RB training as 10^{-4}, 10^{-6} and 10^{-8} for the construction of aSG interpolation, and 10^{-5}, 10^{-7} and 10^{-9} for the construction of aSG integration. Fig. 5 shows the convergence rates of the aSG interpolation and integration error estimators defined in (29), which are close to M^{-2} (the same as theoretical prediction in Theorem 2) and M^{-3} (faster than the theoretical prediction M^{-2}). Fig. 5 also displays the number of RB and its approximation accuracy with different tolerances. We see that in order to achieve the same approximation accuracy for Θ, the number of RB required is considerably smaller than the number of aSG nodes, e.g. 74 RB bases compared to 3476 aSG nodes. This entails the need to solve a smaller number of high-fidelity problems by RB. The same observation holds for the approximation of Z, where 96 RB bases are constructed out of 2×10^3 aSG nodes, which preserves the same integration accuracy as aSG with 584 nodes. Note that in this test, we set the tolerance of EIM as 10^{-9} for the interpolation of Θ and as 10^{-10} for the integration of Z, both of which are negligible compared to the accuracy/tolerance of RB and aSG. When the tolerances for the EIM were selected smaller, the number of EIM bases, whose cost of construction depends linearly on $\mathcal{N}$, are relatively large. In order to balance the errors from aSG, EIM and RB to reach a prescribed numerical accuracy at minimum computational cost, an algorithm will be presented in [10].

6 Conclusion

We investigated acceleration of computational Bayesian inversion for PDEs with distributed parameter uncertainty. Upon reformulation, forward models which are given in terms of PDEs with random input data take the form of countably-parametric, deterministic operator equations. Sparsity of the parameter-to-solution maps is exploited computationally by the reduced basis approach. Sparse grids enter the proposed numerical methods in several ways: first, sparse dimension-adaptive quadratures are used to evaluate conditional expectations in Bayesian estimates and second, sparse grids are used in the offline stage of the reduced basis algorithms (in particular, the empirical interpolation method) to "train" the greedy algorithms and to facilitate the greedy searches over the high-dimensional parameter spaces. For a model diffusion problem, we present detailed numerical experiments of the proposed algorithms, indicating their essentially dimension-independent performance and convergence rates which are only limited by the sparsity in the data-to-solution map.

In the present paper, we considered only a model problem with uniform Bayesian prior on the parameter space U. The proposed approach is, however, also applicable directly to priors with separable densities w.r. to uniform priors. Generalizations to nonseparable prior densities will be provided in the forthcoming work [10], where nonlinear parametric problems will also be addressed.

References

1. M. Barrault, Y. Maday, N. Nguyen, and A. 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 of Mathematical Analysis*, 43(3):1457–1472, 2011.
3. 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.
4. P. Chen and A. Quarteroni. A new algorithm for high-dimensional uncertainty problems based on dimension-adaptive and reduced basis methods. *EPFL, MATHICSE Report 09, submitted*, 2014.
5. 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.
6. P. Chen, A. Quarteroni, and G. Rozza. Comparison of reduced basis and stochastic collocation methods for elliptic problems. *Journal of Scientific Computing*, 59:187–216, 2014.
7. P. Chen, A. Quarteroni, and G. Rozza. A weighted empirical interpolation method: a priori convergence analysis and applications. *ESAIM: Mathematical Modelling and Numerical Analysis*, 48:943–953, 7 2014.
8. P. Chen, A. Quarteroni, and G. Rozza. Reduced order methods for uncertainty quantification problems. *ETH Report 03, Submitted*, 2015.
9. P. Chen and C. Schwab. Sparse grid, reduced basis Bayesian inversion. *ETH Zurich, SAM Report 36, submitted*, 2014.

10. P. Chen and C. Schwab. Sparse grid, reduced basis bayesian inversion II: nonaffine parametric, nonlinear operator equations. *in preparation*, 2015.
11. A. Chkifa, A. Cohen, R. DeVore, and C. Schwab. Adaptive algorithms for sparse polynomial approximation of parametric and stochastic elliptic PDEs. *M2AN Math. Mod. and Num. Anal.*, 47(1):253–280, 2013.
12. A. Chkifa, A. Cohen, and C. Schwab. Breaking the curse of dimensionality in sparse polynomial approximation of parametric pdes. *Journal de Mathématiques Pures et Appliquées*, 2014.
13. A. Cohen and R. DeVore. Kolmogorov widths under holomorphic mappings. *manuscript*, 2014.
14. A. Cohen, R. Devore, and C. Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's. *Analysis and Applications*, 9(01):11–47, 2011.
15. T. Cui, Y. Marzouk, and K. Willcox. Data-driven model reduction for the Bayesian solution of inverse problems. *arXiv preprint arXiv:1403.4290*, 2014.
16. M. Dashti and A. Stuart. The bayesian approach to inverse problems. 2014.
17. D. Galbally, K. Fidkowski, K. 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.
18. M. Hansen and C. Schwab. Sparse adaptive approximation of high dimensional parametric initial value problems. *Vietnam Journal of Mathematics*, 41(2):181–215, 2013.
19. V. Hoang and C. Schwab. Analytic regularity and polynomial approximation of stochastic, parametric elliptic multiscale pdes. *Analysis and Applications (Singapore)*, 11(01), 2013.
20. V. Hoang and C. Schwab. Sparse tensor galerkin discretizations for parametric and random parabolic pdes - analytic regularity and gpc approximation. *SIAM J. Mathematical Analysis*, 45(5):3050–3083, 2013.
21. V. Hoang, C. Schwab, and A. Stuart. Complexity analysis of accelerated mcmc methods for bayesian inversion. *Inverse Problems*, 29(8), 2013.
22. Y. Maday, A. Patera, and D. Rovas. A blackbox reduced-basis output bound method for noncoercive linear problems. *Studies in Mathematics and its Applications*, 31:533–569, 2002.
23. N. Nguyen, G. Rozza, D. Huynh, and A. 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.
24. G. Rozza, D. Huynh, and A. 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.
25. C. Schillings and C. Schwab. Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. *Inverse Problems*, 29(6), 2013.
26. C. Schillings and C. Schwab. Sparsity in Bayesian inversion of parametric operator equations. *Inverse Problems*, 30(6), 2014.
27. C. Schwab and A. Stuart. Sparse deterministic approximation of bayesian inverse problems. *Inverse Problems*, 28(4):045003, 2012.
28. A. Stuart. Inverse problems: a Bayesian perspective. *Acta Numerica*, 19(1):451–559, 2010.
29. G. Turinici, C. Prud'Homme, A. Patera, Y. Maday, and A. Buffa. A priori convergence of the greedy algorithm for the parametrized reduced basis method. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46(3):595, 2012.

Recent Research Reports

Nr.	Authors/Title

2014-38 D. Schoetzau and Ch. Schwab
Exponential Convergence for hp-Version and Spectral Finite Element Methods for Elliptic Problems in Polyhedra

2014-39 P. Grohs and M. Sprecher
Total Variation Regularization by Iteratively Reweighted Least Squares on Hadamard Spaces and the Sphere

2014-40 R. Casagrande and R. Hiptmair
An A Priori Error Estimate for Interior Penalty Discretizations of the Curl-Curl Operator on Non-Conforming Meshes

2015-01 X. Claeys and R. Hiptmair
Integral Equations for Electromagnetic Scattering at Multi-Screens

2015-02 R. Hiptmair and S. Sargheini
Scatterers on the substrate: Far field formulas

2015-03 P. Chen and A. Quarteroni and G. Rozza
Reduced order methods for uncertainty quantification problems

2015-04 S. Larsson and Ch. Schwab
Compressive Space-Time Galerkin Discretizations of Parabolic Partial

Differential Equations

2015-05 S. May
New spacetime discontinuous Galerkin methods for solving convection-diffusion systems

2015-06 H. Heumann and R. Hiptmair and C. Pagliantini
Stabilized Galerkin for Transient Advection of Differential Forms