HESSIAN-BASED SAMPLING FOR HIGH-DIMENSIONAL MODEL REDUCTION

Peng Chen\textsuperscript{1,*} & Omar Ghattas\textsuperscript{1,2}

\textsuperscript{1}Institute for Computational Engineering & Sciences, The University of Texas at Austin, Austin, TX 78712.
\textsuperscript{2}Department of Mechanical Engineering, and Department of Geological Sciences, The University of Texas at Austin, Austin, TX 78712

*Address all correspondence to: Peng Chen, E-mail: peng@ices.utexas.edu

Original Manuscript Submitted: mm/dd/yyyy; Final Draft Received: mm/dd/yyyy

In this work we develop a Hessian-based sampling method for the construction of goal-oriented reduced order models with high-dimensional parameter inputs. Model reduction is known very challenging for high-dimensional parametric problems whose solutions also live in high-dimensional manifolds. However, the manifold of some quantity of interest (QoI) depending on the parametric solutions may be low-dimensional. We use the Hessian of the QoI with respect to the parameter to detect this low-dimensionality, and draw training samples by projecting the high-dimensional parameter to a low-dimensional subspace spanned by the eigenvectors of the Hessian corresponding to its dominating eigenvalues. Instead of forming the full Hessian, which is computationally intractable for a high-dimensional parameter, we employ a randomized algorithm to efficiently compute the dominating eigenpairs of the Hessian whose cost does not depend on the nominal dimension of the parameter but only on the intrinsic dimension of the QoI. We demonstrate that the Hessian-based sampling leads to much smaller errors of the reduced basis approximation for the QoI compared to a random sampling for a diffusion equation with random input obeying either uniform or Gaussian distributions.

KEY WORDS: goal-oriented model reduction, reduced basis method, Hessian-based sampling, randomized SVD, high-dimensional approximation, uncertainty quantification

1. INTRODUCTION

Partial differential equations (PDEs) with stochastic or parametric inputs can be found in many different contexts such as uncertainty quantification, inverse problems, control and optimization, sensitivity and risk analysis. In the case that the dimension of the parameter is very high or even infinite, approximation of the parametric PDEs is computationally very challenging because of the curse of dimensionality—the computational complexity grows exponentially with respect to the dimension of the parameter. Recently, different approximation methods have been developed to deal with the high-dimensional approximation, such as Monte Carlo approximation and its variants—multilevel, quasi, high-order quasi Monte Carlo\cite{37,38,44}, sparse polynomial approximation with Galerkin projection or collocation\cite{5,34,65}, low-rank approximation\cite{49,57}, Taylor approximation or perturbation analysis\cite{3,12}, and reduced basis approximation\cite{11,13,29,30,60}.

The reduced basis approximation, or more generally model reduction\cite{8}, seek the PDE solution by a Galerkin projection in a reduced basis space that is constructed from some ‘snapshots’—PDE solutions at properly selected parameter samples. The dimension of the reduced basis space is expected to be much smaller than the dimension of a high-fidelity approximation space such as the finite element space, so that solving the reduced basis system is much faster than solving the high-fidelity system. Therefore, the way to construct the reduced basis space becomes crucial for the accuracy and efficiency of the reduced basis approximation, which depends on two factors—the training samples and the construction method. For the latter, proper orthogonal decomposition (POD) or the related singular value decomposition (SVD) of the snapshot matrix, and greedy algorithms with a posteriori error estimates have been
developed as the two most successful methods. For the former, the mostly used training samples are random samples drawn from the parameter space [47,59]. Quasi Monte Carlo samples [59], structured collocation or quadrature points [27], sparse grid points [25,41], have also been used.

When it comes to problems with high-dimensional parameters, the PDE solutions may also live in high-dimensional manifolds. Therefore, a large number of reduced basis functions have to be used in order to achieve certain required accuracy of the reduced basis approximation, which makes the reduced order model less efficient or useful. However, in many cases the goal of our computation is some QoI depending on the PDE solution, e.g., the average of the solution at a certain location, which may live in a low-dimensional manifold even it depends on the high-dimensional parameter through the PDE solution. To detect this low-dimensionality structure, we use the Hessian information of the QoI with respect to the parameter, which describes its local curvature, or the extent of its variation with respect to the parameter in different directions. More specifically, one expects that the QoI varies the most along the eigenvectors corresponding to the dominating eigenvalues of the Hessian. Hence, instead of sampling in the whole parameter space, we draw samples by projecting the parameter to the subspace spanned by these eigenvectors, namely a Hessian-based sampling, which is supposed to capture the most variation of the QoI. When the dimension of the parameter is very high, the Hessian matrix becomes very large, to form which one needs to solve a large number of PDEs that is computational intractable. To address this difficulty, we employ a randomized SVD algorithm to compute the dominating eigenpairs of the Hessian, which requires only a limited number of PDE solves. To demonstrate the accuracy of the Hessian-based sampling, we perform numerical experiments based on a diffusion model with parametric diffusion coefficient. We consider both a uniform distribution and a Gaussian distribution for the parameter. In the former case, the coefficient is a piecewise random variable in each subdomain of the physical domain; for the latter, the coefficient is a log-normal random field. We construct the reduced order model by both a POD/SVD algorithm and a greedy algorithm with random training samples, as well as by the POD/SVD algorithm with Hessian-based training samples. From the comparison of the error decay of the reduced basis approximation for both the solution and the QoI, we show that the Hessian-based sampling leads to more accurate approximation for the QoI than the random sampling, not necessarily for the PDE solution. We mention that a Hessian-based model reduction with initial-condition inputs was developed in [7], which does not involve any parameter and the Hessian has different meaning from the second order variation of the QoI in our context. In [36,51], the parameter and state are simultaneously projected to their subspaces constructed in a greedy manner in the context of model reduction for inverse problems, and the Hessian of the likelihood function is employed in [36] to seek the parameter subspace for inverse problems.

The following of the paper is organized as follows: in Section 2 we present the basic elements for model reduction, including the reduced basis approximation, offline-online decomposition, two methods for the construction of reduced basis spaces, and a short survey of sampling methods for the construction. Section 3 is devoted to the development of the Hessian-based sampling method, the randomized SVD algorithm for the computation of the eigenpairs of the Hessian, and the way to compute the Hessian action in certain given direction. Numerical experiments are presented in Section 4 for the demonstration of the efficiency and the accuracy of the Hessian-based sampling method, for both a uniform distributed parameter of 256 dimensions and a Gaussian distributed parameter of 16,641 dimensions. At last, conclusions and perspectives are provided in Section 5.

2. MODEL REDUCTION

In this section, we briefly present the main ingredients of model reduction for a linear parametric partial differential equation (PDE) by a reduced basis method, which include a high-fidelity approximation and a reduced basis approximation for the PDE and QoI, offline-online decomposition of the reduced basis approximation, the construction algorithms (POD/SVD and greedy) of the reduced basis space, and a short survey of sampling methods for the construction.

2.1 Parametric partial differential equations

Let $V$ denote a Hilbert space on $\mathbb{R}$ with its dual space $V'$. Let $P \subset \mathbb{R}^K$ denote a $K$-dimensional parameter space, where $K \in \mathbb{N}$. We consider an abstract weak form of a linear parametric PDE: at any given parameter
\( \mathbf{p} = (p_1, \ldots, p_K) \in P \), find \( u \in V \) such that

\[
a(u, v; \mathbf{p}) = f(v; \mathbf{p}) \quad \forall v \in V,
\]

where \( a(\cdot, \cdot; \mathbf{p}): V \times V \to \mathbb{R} \) is a bilinear form and \( f(\cdot; \mathbf{p}) \in V' \) is a linear functional for any given \( \mathbf{p} \). By \( s(u) \in \mathbb{R} \) we denote a QoI that depends on the solution \( u \), which is our goal of computation.

### 2.2 High-fidelity approximation

To solve problem (1), we introduce an approximation space \( V_h \subset V \) with dimension \( N_h = \dim(V_h) \), e.g., a finite element space, where \( h \) stands for the mesh size. In the following, we call \( V_h \) a high-fidelity approximation space and \( u_h \) a high-fidelity solution, which solves the high-fidelity approximation problem: at any given \( \mathbf{p} \in P \), find \( u_h \in V_h \) such that

\[
a(u_h, v_h; \mathbf{p}) = f(v_h; \mathbf{p}) \quad \forall v_h \in V_h.
\]

Let \( \{ \zeta^n_h \}_{n=1}^{N_h} \) denote the basis functions in \( V_h \), i.e., \( V_h = \text{span}\{ \zeta^n_h, n = 1, \ldots, N_h \} \), so that the high-fidelity solution \( u_h \) can be represented as

\[
u_h = \sum_{n=1}^{N_h} u^h_n \zeta^n_h,
\]

where \( u_h = (u^h_1, \ldots, u^h_{N_h})^\top \in \mathbb{R}^{N_h} \) is the coefficient vector. Then the algebraic formulation of problem (2) can be written as: find \( u_h \in \mathbb{R}^{N_h} \) such that

\[
A_h(\mathbf{p})u_h = f_h(\mathbf{p}),
\]

where the high-fidelity matrix \( A_h(\mathbf{p}) \in \mathbb{R}^{N_h \times N_h} \) and vector \( f_h(\mathbf{p}) \in \mathbb{R}^{N_h} \) at \( \mathbf{p} \) are given by

\[
(A_h(\mathbf{p}))_{mn} = a(\zeta^n_h, \zeta^m_h; \mathbf{p}) \quad \text{and} \quad (f_h(\mathbf{p}))_m = f(\zeta^m_h; \mathbf{p}), \quad m, n = 1, \ldots, N_h.
\]

As a result, the QoI \( s(u) \) can be approximated by

\[
s(u_h) = s_h^\top u_h, \quad \text{where} \quad (s_h)_n = s(\zeta^n_h), \quad n = 1, \ldots, N_h,
\]

where we assume that the QoI is linear with respect to the solution for simplicity.

### 2.3 Reduced-basis approximation

As \( N_h \) is typically very big if high accuracy of the solution/QoI is required, solving the large-scale system (4) at each \( \mathbf{p} \in P \) is computational expensive and only a limited number of solves can be afforded. To reduce the computational cost, we introduce a reduced basis approximation: for any given \( \mathbf{p} \in P \), find \( u_N \in V_N \) such that

\[
a(u_N, v_N; \mathbf{p}) = f(v_N; \mathbf{p}) \quad \forall v_N \in V_N,
\]

where \( V_N \subset V_h \) is called the reduced basis space with dimension \( N \). Let \( \{ \zeta^n_N \}_{n=1}^{N} \) denote the basis functions of \( V_N \), i.e., \( V_N = \text{span}\{ \zeta^n_N, n = 1, \ldots, N \} \), then the reduced basis solution can be represented as

\[
u_N = \sum_{n=1}^{N} u^N_n \zeta^n_N,
\]

with coefficient vector \( u_N = (u^N_1, \ldots, u^N_N)^\top \in \mathbb{R}^N \). Consequently, the algebraic formulation of the reduced basis approximation problem (7) can be obtained as

\[
A_N(\mathbf{p})u_N = f_N(\mathbf{p}),
\]

where the reduced basis matrix \( A_N(\mathbf{p}) \) and vector \( f_N(\mathbf{p}) \) at \( \mathbf{p} \) are given by

\[
(A_N(\mathbf{p}))_{mn} = a(\zeta^n_N, \zeta^m_N; \mathbf{p}) \quad \text{and} \quad (f_N(\mathbf{p}))_m = f(\zeta^m_N; \mathbf{p}), \quad m, n = 1, \ldots, N.
\]

Moreover, the reduced basis approximation of the QoI can be evaluated as

\[
s(u_N) = s_N^\top u_N \quad \text{where} \quad (s_N)_n = s(\zeta^n_N), \quad n = 1, \ldots, N.
\]
2.4 Offline-online decomposition

Assume that the bilinear form $a$ and the linear functional $f$ allows the following affine representations with $Q_a$ and $Q_f$ terms

$$a(w, v; p) = \sum_{q=1}^{Q_a} \theta^q_a(p) a^q(w, v) \quad \text{and} \quad f(v; p) = \sum_{q=1}^{Q_f} \theta^q_f(p) f^q(v),$$

i.e., $a(\cdot; p)$ and $f(\cdot; p)$ depend on the parameter $p$ through the coefficients $\theta^q_a(p) \in \mathbb{R}$ and $\theta^q_f(p) \in \mathbb{R}$. Then the reduced basis algebraic system can be written as

$$\left( \sum_{q=1}^{Q_a} \theta^q_a(p) A^q_N \right) u_N = \sum_{q=1}^{Q_f} \theta^q_f(p) F^q_N,$$

where the reduced basis matrices $A^q_N \in \mathbb{R}^{N \times N}$, $q = 1, \ldots, Q_a$ and vectors $F^q_N \in \mathbb{R}^N$, $q = 1, \ldots, Q_f$, are given by

$$(A^q_N)_{mn} = a^q(\zeta^m_N, \zeta^n_N) \quad \text{and} \quad (F^q_N)_{mn} = f^q(\zeta^m_N), \quad m, n = 1, \ldots, N.$$  

Therefore, once the reduced basis matrices and vectors are computed and stored in the offline stage, solution of the reduced basis system (13) in the online stage takes $O(Q_a N^2 + Q_f N)$ operations for assembling and $O(N^3)$ operations for solving, evaluation of the reduced basis approximation of the QoI takes $O(N)$ and the linear functional

Assume that the bilinear form

2.5 Construction of the reduced basis space

Both the accuracy of the reduced basis approximation and the performance of the computational reduction critically depend on the reduced basis space $V_N$. Here we present two common algorithms for the construction of $V_N$: POD/SVD and a goal-oriented greedy algorithm.

2.5.1 The POD/SVD algorithm

For the construction by proper orthogonal decomposition (POD), one first takes a training sample set

$$\Xi_t = \{p^n, n = 1, \ldots, N_t \}$$

with $N_t$ samples. Then the high-fidelity solution vector $u_h(p)$ is computed by solving the high-fidelity problem (4) at each of the training sample $p \in \Xi_t$. By forming the coefficient matrix $U = (u_h(p^1), \ldots, u_h(p^{N_t})) \in \mathbb{R}^{N_h \times N_t}$, one then compute its singular value decomposition (SVD)

$$U = \mathbb{V} \Sigma \mathbb{W}^T,$$

where $\mathbb{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_{N_h}) \in \mathbb{R}^{N_h \times N_h}$ and $\mathbb{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_{N_t}) \in \mathbb{R}^{N_t \times N_t}$ are orthonormal matrices, and $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \in \mathbb{R}^{N_h \times N_t}$ is the diagonal matrix of positive singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ with $r \leq \min(N_h, N_t)$, denoting the rank of $U$. Then the reduced basis space $V_N$ is constructed with the first $N$ singular vectors of $\mathbb{V}$ as the coefficient vectors for its basis functions, for $N$ such that

$$N = \arg\min_{n \leq r} \frac{\sum_{i=1}^{n} \sigma_i^2}{\sum_{i=1}^{r} \sigma_i} \geq 1 - \varepsilon$$

with a given tolerance $\varepsilon > 0$ representing the information/energy loss. Note that the singular vectors are orthonormal in the discrete $\ell_2$-norm. To construct basis functions orthogonal with respect to the norm $X$, e.g., $L_2$-norm or energy $V$-norm, we only need to perform SVD on $BU$ where $X = BU^T B$, e.g., a Cholesky factorization of $X \in \mathbb{R}^{N_h \times N_h}$ where $X_{mn} = (\zeta_n^m, \zeta_n^m)_X$, and construct $V_N$ with $B^{-1} \zeta_n$ as the coefficient vector of its $n$-th basis function $\zeta_N^n$, so that $(\zeta_N^n, \zeta_n^m)_X = \zeta_n^T B^{-T} XB^{-1} \zeta_m = \delta_{mn}$.

2.5.2 The Greedy algorithm

Different from the POD/SVD algorithm, the greedy algorithm seeks to construct $V_N$ iteratively in the parameter training set $\Xi_t$. At the initial step, one often picks the first sample $p^1$ from $\Xi_t$, solve the high-fidelity problem (4) at $p^1$, and construct $V_1 = \text{span}\{u_h(p^1)\}$. Then, for $N = 1, 2, \ldots$, one chooses the next sample as

$$p^{N+1} = \arg\max_{p \in \Xi_t} \Delta_N(p), \quad (18)$$

solve the high-fidelity problem at $p^{N+1}$, and enrich $V_{N+1} = V_N \oplus \text{span}\{u_h(p^{N+1})\}$, which is often orthogonalized by Gram–Schmidt process. Here $\Delta_N(p)$ is an a-posteriori error indicator of the solution error $||u_h(p) - u_N(p)||_V$ or the goal-oriented error $|s(u_h(p)) - s(u_N(p))|$. As our goal is the computation of the QoI, we consider a goal-oriented (dual-weighted residual) error indicator for the latter, which is defined as

$$\Delta_N(p) := |r(\psi_N; p)| = |f(\psi_N; p) - a(u_N, \psi_N; p)|, \quad (19)$$

where $\psi_N$ is the solution of the dual problem: given $p \in P$, find $\psi_N \in W_N$ such that

$$a(w_N, \psi_N; p) = s(w_N) \quad \forall w_N \in W_N, \quad (20)$$

where the reduced basis space $W_N$ can be constructed as $W_N = \text{span}\{\psi_h(p^n), n = 1, \ldots, N\}$ with $\psi_h(p)$ the high-fidelity solution of the dual problem (20) in $V_h$. Under the affine assumption (12), we can evaluate the weighted residual by

$$r(\psi_N; p) = \sum_{q=1}^{Q_f} \theta_q^f(p) (f_q^f)^T \psi_N - \sum_{q=1}^{Q_a} \theta_q^a(p) u_N^T h_N^q \psi_N, \quad (21)$$

with $O(Q_f N + Q_a N^2)$ operations, independent of $N_h$, where

$$(\bar{f}_N^f)_m = f_q(\eta_N^m) \quad \text{and} \quad (\bar{h}_N^a)_m = a_q(\zeta_N^m, \eta_N^m), \quad m, n = 1, \ldots, N, \quad (22)$$

are computed and stored for once. Here by $\{\eta_N^m\}_{N=1}^N$ we denote the basis functions of $W_N$, which are obtained by Gram–Schmidt orthogonalization from $\{\psi_h(p^n)\}_{n=1}^N$.

2.6 A short survey of sampling methods

Both the POD/SVD construction and the greedy construction algorithms require a training sample set $\Xi_t$, which plays a crucial role in the approximation property of the reduced basis space $V_N$, especially in the case of high-dimensional parameter. On the one hand, $\Xi_t$ should be rich enough such that the main information of the manifold of the solution or the QoI can be captured by the snapshots in the training set. On the other hand, the size of $\Xi_t$ should not be redundantly large as one has to solve the expensive high-fidelity problem at each of the training sample by the POD/SVD algorithm, or compute the error indicator $\Delta_N(p)$ at each of the training sample for each $N = 1, 2, \ldots$, by the greedy algorithm.

One of the most widely used sampling method is random sampling from the probability distribution of the parameter [57, 60]. It is rather straightforward and does not take the property of the computational QoI into account. A variant is the quasi-random sampling using low-discrepancy sequences [55, 59], such as Halton or Sobol sequence, which tends to provide more equidistributed samples in the parameter space. For different probability distributions
of the parameter, weighted reduced basis/POD methods \[26,62,64\] were developed by sampling from the probability distribution with a weighted a-posteriori error estimator for the construction of the reduced basis space. Structured sampling methods using quadrature/collocation points such as Chebyshev points and Gauss Legendre/Hermite points have also been investigated \[27\] in comparison with the random sampling methods. In \[40\], an “hp” adaptive sampling method was proposed, where the parameter domain is decomposed into smaller subdomains and in each subdomain a random sampling is used. An adaptive greedy sampling algorithm was proposed in \[46\] by adaptively cleaning and enriching the training sample set with random samples. A goal-oriented sampling method was developed in \[24\] in the context of failure probability computation, where the samples are adapted to the critical limit state surface. In high-dimensional parameter space, a greedy sampling method in combination with the isotropic sparse grid and dimension-adaptive sparse grid has been developed in \[41\] and \[25\]. Sampling from a subspace of the parameter space constructed using Karhunen–Loève expansion or gradient information were investigated in \[22,45,63\].

3. HESSIAN-BASED SAMPLING

In this section, we develop a new sampling method particularly suited for high-dimensional parametric problems based on the Hessian of the QoI with respect to the parameter. The rationale is that even the intrinsic dimension of the solution manifold is high, that of the QoI manifold could still be low, which can be captured by the low rank structure or fast spectral decay of the Hessian of the QoI with respect to the parameter. In fact, the low rank or fast spectral decay property of the Hessian has been proven for some specific problems and observed numerically for many others \[1–3,7,14,16–20,32,35,43,48,53,58\]. We can therefore draw samples by projecting the high-dimensional parameter into a low-dimensional subspace spanned by the eigenvectors corresponding to the largest (absolute) eigenvalues.

3.1 Hessian

Our computational goal is the QoI \(s(u(p))\), which depends the parameter \(p \in P \subset \mathbb{R}^{K \times K}\) through the PDE solution \(u(p)\). In the following, we simply denote it as \(s(p)\). Hessian is the square matrix \(\mathbb{H} \in \mathbb{R}^{K \times K}\) of the second-order partial derivatives of \(s\) with respect to \(p\), i.e.,

\[
\mathbb{H}_{kl} = \frac{\partial^2 s}{\partial p_k \partial p_l}, \quad k, l \in 1, \ldots, K. \tag{23}
\]

It describes the local curvature of \(s\) at \(p\) in the parameter space \(P\), and has been widely used in large-scale optimization \[16,33,56\], Bayesian inversion \[9,32,53\], and data assimilation \[7,50\]. The eigenvectors corresponding to the dominating eigenvalues of the Hessian are the directions along which the QoI changes the most in the parameter space, which is illustrated by a simple example in Fig. 1. We can see that the QoI varies only along the first eigenvector and does not change along the second. Thus, sampling in the subspace spanned by the eigenvectors corresponding to the dominating eigenvalues will presumably capture the most variation of the QoI.

3.2 Hessian-based sampling

Without loss of generality, suppose the parameter \(p\) obeys probability distribution \(\mu\) and has mean \(\bar{p}\) and covariance \(\mathbb{C}\). For instance, \(\mu = \mathcal{U}([-\sqrt{3}, \sqrt{3}]),\) a uniform distribution in the box \([-\sqrt{3}, \sqrt{3}]^K\) with mean \(\bar{p} = 0\) and covariance \(\mathbb{C} = \mathbb{I}\), or \(\mu = \mathcal{N}(\bar{p}, \mathbb{C})\), a Gaussian distribution with mean \(\bar{p}\) and covariance \(\mathbb{C}\). Note that a parameter \(p \in P\) in the compact support \(P \in \mathbb{R}^K\) without any prescribed probability distribution can be regarded as a random variable uniformly distributed in the parameter space \(P\). To proceed with the Hessian-based construction of the subspace for the parameter projection, we first consider the quadratic approximation of \(s\) given by

\[
s(p) \approx s_{\text{quad}}(p) = s(\bar{p}) + g_p^T (p - \bar{p}) + \frac{1}{2} (p - \bar{p})^T \mathbb{H}_{\bar{p}} (p - \bar{p}), \tag{24}
\]
where $g_p$ and $H_p$ represent the gradient and the Hessian of $s$ at $\bar{p}$. The expectation of $s$ can thus be approximated by

$$E[s] \approx E[s_{\text{quad}}],$$

which has the analytic expression (see the proof in Appendix A)

$$E[s_{\text{quad}}] = s(\bar{p}) + \frac{1}{2} \text{tr}(\tilde{H}_p),$$

(25)

where the second term is the trace of the covariance preconditioned Hessian $\tilde{H}_p = CH_p$ at the mean $\bar{p}$. It is equivalent to the sum of all its eigenvalues, i.e.,

$$\text{tr}(\tilde{H}_p) = \sum_{k=1}^{K} \lambda_k(\tilde{H}_p).$$

(26)

In many problems, e.g., [1–3,7,14,16–20,32,35,43,48,53,58], it can be proven or numerically demonstrated that the (absolute) eigenvalues are dominated by a only a few of them $L \ll K$. Moreover, $L$ typically does not change even $K$ becomes bigger, e.g., as the mesh is refined for a random field parameter. Therefore, the variation of $s$ can be captured by the dominating eigenvalues, which implies that the parameter in the subspace spanned by the corresponding eigenvectors contribute to the most variation of the QoI in the parameter space.

To compute the dominating eigenvalues $(\lambda_k)_{k=1}^L$ of $\tilde{H}_p = CH_p$ for some $L \leq K$, which are the same as the dominating generalized eigenvalues of $(H_p, C^{-1})$, we solve the generalized eigenvalue problem

$$H_p \phi_k = \lambda_k C^{-1} \phi_k, \text{ such that } \phi_k^T C^{-1} \phi_{k'} = \delta_{kk'}, \quad k, k' = 1, \ldots, L.$$  

(27)

We remark that $C^{-1}$ is used in the computation as it is often readily available, e.g., when the covariance is given by the discretization of an inverse of a fractional elliptic operator as shown later in Section 4.2. The parameter dimension reduction is then obtained by projecting the parameter $p - \bar{p}$ to the $L$-dimensional subspace spanned by the eigenvectors $\Phi_L = \text{span}\{\phi_l, l = 1, \ldots, L\}$, with properly chosen $L \leq K$, i.e.,

$$P_L(p - \bar{p}) = \sum_{l=1}^{L} \phi_l \phi_l^T C^{-1}(p - \bar{p}).$$

(28)

Then projected parameter (sample from the subspace $\Phi_L$), denoted as $p_L$, is given by

$$p_L = \bar{p} + P_L(p - \bar{p}).$$

(29)
In the case of Gaussian distribution \( \mu = \mathcal{N}(\bar{p}, \mathbb{C}) \), the parameter \( p \) can be expressed by the Karhunen–Loève expansion as
\[
p = \bar{p} + \sum_{k=1}^{K} \sqrt{\rho_k} \phi_k \xi_k, \quad \text{i.i.d.} \quad \xi_k \sim \mathcal{N}(0, 1),
\] (30)
where \( (\rho_k, \phi_k)_{k=1}^{K} \) are the eigenpairs of the covariance \( \mathbb{C} \). Then the projection (29) becomes
\[
p_L = \bar{p} + \sum_{l=1}^{L} \varphi_l \sum_{k=1}^{K} \sqrt{\rho_k} \phi_l^T \mathbb{C}^{-1} \phi_k \xi_k.
\] (31)
Since a linear combination of Gaussian random variations is still a Gaussian random variable denoted as \( \sqrt{\beta_l} \omega_l, \omega_l \in \mathcal{N}(0, 1) \), with the variance given by
\[
\beta_l = \sum_{k=1}^{K} \rho_k (\varphi_l^T \mathbb{C}^{-1} \phi_k)^2 = \varphi_l^T \left( \sum_{k=1}^{K} \rho_k^{-1} \phi_k \phi_k^T \right) \varphi_l = \varphi_l^T \mathbb{C}^{-1} \varphi_l = 1,
\] (32)
where we used (27) in the last equality, so that we can sample \( p_L \) simply as
\[
p_L = \bar{p} + \sum_{l=1}^{L} \varphi_l \omega_l.
\] (33)
Note that for high-dimension parameters with \( K \gg L \), \( \omega_l \) can be taken as i.i.d. random variables.

We remark that the Hessian \( \mathbb{H}_p \) is local, evaluated at the mean \( \bar{p} \), which may fail to characterize the variation of the QoI globally in the parameter space. To deal with this issue, we propose two schemes for the computation of a global Hessian—namely, an averaged Hessian and a combined Hessian—to account for the variation of the QoI globally in the parameter space, as presented in APPENDIX B. Extension of the Hessian-based sampling for multiple quantities of interest, or a vector-valued output is presented in APPENDIX C.

3.3 Randomized SVD for generalized eigenvalue problems

To solve the generalized eigenvalue problem (27), it is prohibitive to form the full Hessian matrix when the parameter dimension is high. Instead, we apply a randomized SVD algorithm to compute the dominating generalized eigenpairs which only requires Hessian action in some random parameter directions. This is presented in Algorithm 1; see [4, 61] for more details on randomized SVD algorithms.

**Algorithm 1: Randomized SVD for the generalized eigenvalue problem (27)**

**Input:** matrices \( \mathbb{H}_p, \mathbb{C} \), the number of eigenpairs \( k \), an oversampling factor \( c \leq 10 \).

**Output:** \( (\Lambda_L, \Phi_L) \) with \( \Lambda_L = \text{diag}(\lambda_1, \ldots, \lambda_L) \) and \( \Phi_L = (\varphi_1, \ldots, \varphi_L) \).

1. Draw a Gaussian random matrix \( \Omega \in \mathbb{R}^{n \times (L+c)} \).
2. Compute \( Y = \mathbb{C} (\mathbb{H}_p \Omega) \).
3. Compute QR-factorization \( Y = QR \) such that \( Q^T \mathbb{C}^{-1} Q = I_{L+c} \).
4. Form \( T = Q^T \mathbb{H}_p Q \) and compute eigendecomposition \( T = SAS^T \).
5. Extract \( \Lambda_L = \Lambda(1: L, 1: L) \) and \( \Phi_L = QS_L \) with \( S_L = S(:, 1: L) \).

We remark that the computation is dominated by the Hessian actions \( \mathbb{H}_p \Omega \) and \( \mathbb{H}_p Q \), which are presented in the next section. The advantages of Algorithm 1 are: (i) the error of the eigenvalues \( \lambda_j, j = 1, \ldots, L \), are bounded by the remaining ones \( \lambda_j, j > L \), which is small if they decay fast; (ii) the computational cost is dominated by \( 2(L + c) \) Hessian actions (the application of \( \mathbb{C} \) on a vector is inexpensive, e.g., it only takes \( O(n) \) operations by a multigrid solver for \( \mathbb{C} \) discretized from a differential operator); (iii) it is tractable as \( L \) typically does not change when \( K \) becomes bigger; (iv) computing the Hessian actions \( \mathbb{H}_p \Omega \) and \( \mathbb{H}_p Q \) can be asynchronously parallelized.
3.4 Hessian action in a given direction

To compute the Hessian action in a certain given parameter direction, we employ a Lagrange multiplier method. We take the Hessian $\mathcal{H}_p$ at any given $p \in P$ as an example to compute its action in a given direction $\hat{p} \in P$, which readily applies to the local, the averaged, and the combined Hessian actions. We first form the Lagrangian as

$$\mathcal{L}(u, v; p) = s(u) + f(v; p) - a(u, v; p),$$  \hspace{1cm} (34)

where $v$ is the adjoint variable or the Lagrange multiplier. Then by setting the first variation of $\mathcal{L}$ with respect to the adjoint and the state variables to be zero we obtain the state and adjoint variables at $p$ as the solutions of the state problem (1) and the adjoint problem: find $v \in V$ such that

$$a(w, v; p) = s(w) \quad \forall w \in V.$$  \hspace{1cm} (35)

Then we can evaluate the gradient of $s$ with respect to $p$ as

$$g_p = \partial_p \mathcal{L}(u, v; p) = \partial_p f(v; p) - \partial_p a(u, v; p).$$  \hspace{1cm} (36)

To compute the Hessian action in direction $\hat{p}$, we form another Lagrangian for the first order variation constraints as

$$\mathcal{L}^H(u, v, p, \hat{u}, \hat{v}; \hat{p}) = a(u, \hat{v}; p) + a(\hat{u}, v; p) + (\partial_p f(v; p) - \partial_p a(u, v; p), \hat{p}),$$  \hspace{1cm} (37)

where $\hat{v}$ and $\hat{u}$ are adjoint variables. They can be obtained by setting variation of (37) with respect to $u$ and $v$ as zero, which leads to the incremental adjoint problem: find $\hat{v} \in V$ such that

$$a(\hat{u}, \hat{v}; p) = (\partial_p a(\hat{u}, v; p), \hat{p}) \quad \forall \hat{u} \in V,$$  \hspace{1cm} (38)

and the incremental state problem: find $\hat{u} \in V$ such that

$$a(\hat{u}, \hat{v}; p) = (\partial_p a(u, \hat{v}; p) - \partial_p f(\hat{v}; p), \hat{p}) \quad \forall \hat{v} \in V.$$  \hspace{1cm} (39)

We remark that the adjoint problem and the two incremental problems are linear and have the same linear operator (or its adjoint), even when the state problem is nonlinear. Then the Hessian action in direction $\hat{p}$ is given by

$$\mathcal{H}_p \hat{p} = \partial_p \mathcal{L}^H \hat{p} = (\partial_p a(u, \hat{v}; p) + \partial_p a(\hat{u}, v; p) + \partial_{pp} f(v; p) - \partial_{pp} a(u, v; p), \hat{p}).$$  \hspace{1cm} (40)

Therefore, once the solutions of the state and adjoint problems at $p$ are obtained, the Hessian action $\mathcal{H}_p \hat{p}$ only involves solving two linear incremental problems for each $\hat{p}$, which implies that the computational cost of the randomized SVD Algorithm I is dominated by $4(L+c)$ linear PDE solves with the same linear operator (or its adjoint).

4. NUMERICAL EXPERIMENTS

To demonstrate the performance of the Hessian-based sampling algorithm for goal-oriented model reduction with high-dimensional parameter, we consider the diffusion problem

$$-\nabla(\kappa(p)\nabla u) = g, \quad \text{in } D,$$  \hspace{1cm} (41)

in a physical domain $D = [0, 1]^2$, with parametric coefficient $\kappa(p)$ and suitable boundary conditions on $\partial D$. We consider the following QoI

$$s(p) = \frac{1}{|D_s|} \int_{D_s} u(p) dx,$$  \hspace{1cm} (42)

where we set $D_s = [0, 0.1]^2$ with volume $|D_s| = 0.1^2$. As for the parameter $p \in P \subset \mathbb{R}^K$, we consider the cases of a uniform distribution with dimension $K = 16^2$ and a Gaussian distribution with dimension $K = 129^2$. 
4.1 Uniform distribution

In this example, we consider the coefficient $\kappa(p)$ as a piecewise random variable with uniform distribution given by

$$\kappa(p) = \kappa_0 + \sum_{k=1}^{K} k^{-\beta} \chi_{D_k} p_k,$$

where $\chi_{D_k}$ is a characteristic function taking value one in $D_k$ and zero elsewhere, $p \sim U([-\sqrt{3}, \sqrt{3}]^K)$ with mean $\bar{p} = 0$ and covariance $\Sigma = I$, $\beta$ is a scaling parameter. Here $D = \bigcup_k D_k = [0, 1]^2$, where the $k$-th subdomain $D_k = [i * h, (j + 1) * h] \times [(i + 1) * h, (j + 1) * h]$ with $h = 1/\sqrt{K}$, $j = \text{mod}(k, \sqrt{K})$ and $i = (k - j)/\sqrt{K}$. We take $g = 0$, and set the Dirichlet boundary conditions $u = 1$ on $x = [0, 1] \times 0$, and $u = 0$ on $x = [0, 1] \times 1$, while setting zero Neumann boundary conditions elsewhere.

The weak form of problem (41) can be expressed as (1) with affine representation (12), where we have $\theta^1_a(p) = \kappa_0, a^1(w, v) = \int_{\hat{D}} \nabla w \cdot \nabla v dx$ and

$$\theta^{k+1}_a(p) = k^{-\beta} \chi_{D_k} p_k, \text{ and } a^{k+1}(w, v) = \int_{D_k} \nabla w \cdot \nabla v dx, \quad k = 1, \ldots, K,$$

and $\theta^1_f(p) = \kappa_0, f^1(v) = \int_{\hat{D}} \nabla u_D \cdot \nabla v dx$, being $u_D$ the Dirichlet boundary condition on $\Gamma_D = [0, 1] \times \{0, 1\}$, and

$$\theta^{k+1}_f(p) = k^{-\beta} \chi_{D_k} p_k, \text{ and } f^{k+1}(v) = \int_{\partial D_k \cap \Gamma_D} \nabla u_D \cdot \nabla v dx, \quad k = 1, \ldots, K.$$

In the numerical test, we use piecewise linear finite element in a uniform mesh of size $65 \times 65$ for the discretization of the problem. We consider a relatively high dimension $K = 256$. We set the parameter $\kappa_0 = \sqrt{3} + 0.01$ to guarantee that the coefficient is positive, and set $\beta = 1$ so that the solution manifold is relatively high-dimensional yet its reduced basis approximation error still shows evident decay with respect to the number of reduced basis functions.

At first, we compute the Hessian $\mathbb{H}_p$ of the QoI $s$ at the mean $\bar{p} = 0$. Note that here $\mathbb{H}_p \in \mathbb{R}^{K \times K}$, which can be formed via (40) by solving the incremental adjoint and state problems (38) and (39) with $\bar{p} = e_k$, whose $k$-th element is one and all the other elements are zero, $k = 1, \ldots, K$. Computing the full Hessian is used in this test of uniform distribution for $K = 256$, which becomes very expensive if $K$ is much larger as in the next test of Gaussian distribution where we use the randomized SVD Algorithm (7) instead of computing the full Hessian. The eigenvalues of the Hessian are computed as the solution of problem (27) with $C = I$, which are shown in Fig. (2). We can observe that the eigenvalues decay very fast in the first few dimensions, with four orders of magnitude of difference in the first 20 dimensions.

In the construction of the reduced basis space, we compare three different schemes: POD with random training samples, (goal-oriented) greedy with random training samples, and POD with Hessian-based training samples. We draw $N_t = 1000$ training samples for each scheme and construct the corresponding reduced basis space by the POD/greedy algorithms presented in Sec. 2.5 with $N = 200$ basis functions. Then we compute the relative error of the reduced basis approximation for the solution and the QoI as

$$\mathcal{E}_N = \frac{1}{|\Xi_{test}|} \sum_{p \in \Xi_{test}} \frac{||u_h(p) - u_N(p)||_V}{||u_h(p)||_V}; \mathcal{E}_N^s = \frac{1}{|\Xi_{test}|} \sum_{p \in \Xi_{test}} \frac{|s_h(p) - s_N(p)|}{|s_h(p)|},$$

where $\Xi_{test}$ is a test sample set with $|\Xi_{test}| = 10$ samples randomly drawn from the uniform distribution. The decay of the relative error of the RB approximation for the PDE solution is shown in the left of Fig. (3) from which we can observe that the RB space obtained by POD construction with random training samples leads to the smallest approximation error, smaller than that by the greedy construction with the same random training samples. The errors get stagnated for the POD construction with the Hessian-based training samples. Moreover, the larger the number of the modes $L$ used in the projection (29), the smaller the errors become. This observation indicates that for the
Hessian-based sampling for high-dimensional model reduction

FIG. 2: The decay of the generalized eigenvalues of the Hessian $H_{\bar{p}}$ at the mean $\bar{p} = 0 \in \mathbb{R}^{256}$, as the solution of problem (27). Positive and negative eigenvalues are shown with red dot and blue cross, respectively.

FIG. 3: Decay of the relative error $E_u$ (left) and $E_s$ (right), defined in (46), of the RB approximation of the solution and the QoI, respectively. $L = 5, 10, 20$ modes are used in the projection (29) for the Hessian-based samples.

RB approximation of the solution, the Hessian-based sampling does not capture the solution manifold as well as the random sampling, which is expected since the Hessian is for the QoI, not for the PDE solution.

As for the RB approximation of the QoI, from the right of Fig. 3 we can observe that the Hessian-based sampling (with $L = 20$ modes) leads to much smaller errors compared to both the POD and the greedy construction with random training samples. Moreover, the Hessian-based sampling with small number of modes ($L = 5$) yields smaller errors for small $N$ but larger errors for large $N$ than that with large number of modes ($L = 20$), which implies that the parameter in the subspace spanned by fewer eigen-modes can capture more representative reduced basis functions for small $N$, while more eigen-modes are needed if higher approximation accuracy is required. Therefore, an adaptive Hessian-based sampling by gradually increasing the number of modes could probably lead to a better construction...
method, which is subject to further investigation. Note that the greedy construction yields smaller errors than the POD construction with the same random training samples, due to the use of the goal-oriented a-posteriori error indicator that can be efficiently computed as in thanks to the affine representation.

4.2 Gaussian distribution

In the second example, we consider the coefficient as a log-normal random field, i.e.,

$$\kappa(p) = e^p,$$  \hfill (47)

where $p$ is a Gaussian random field with distribution $\mathcal{N}(\bar{p}, C)$. Here the covariance operator $C$ is taken as the inverse of a fractional elliptic operator

$$C = (-\delta \Delta + \gamma I)^{-\alpha}.$$  \hfill (48)

In this example, we take $\gamma = 0.5$, $\delta = 1$, $\alpha = 2$, such that $C$ is of trace class. By piecewise linear finite element discretization in a uniform mesh of size $129 \times 129$, we obtain a $129^2$-dimensional parameter $p$ as the coefficient for the random field $p$, which obeys the Gaussian distribution $p \sim \mathcal{N}(\bar{p}, C)$ with covariance given such that

$$C^{-1} = A M^{-1} A,$$  \hfill (49)

where $M$ is the mass matrix and $A$ is the stiffness matrix given by

$$M_{ij} = \int_D \psi_j \psi_i dx \text{ and } A_{ij} = \int_D (\delta \nabla \psi_j \cdot \nabla \psi_i + \gamma \psi_j \psi_i) dx, \quad i, j = 1, \ldots, 129^2,$$  \hfill (50)

where $\psi_i, i = 1, \ldots, 129^2$, are the finite element basis functions. For simplicity, we take the source term $g = 1$ and use homogeneous Dirichlet boundary conditions.

![Graph](image.png)

**FIG. 4:** The decay of the generalized eigenvalues of the Hessian $\nabla^2 p$ at the mean $\bar{p} = 0 \in \mathbb{R}^{16641}$, as the solution of problem (27). Positive and negative eigenvalues are shown with red dot and blue cross, respectively.

We solve the generalized eigenvalue problem by the randomized SVD algorithm presented in Section where the Hessian action in a random direction is evaluated as in Section. We compute the first 100 eigenpairs.
by using 110 random directions, for which the decay of the eigenvalues is shown in Fig. 4. We can observe that the eigenvalues decay very fast in the first few dimensions, with four orders of magnitude of difference in the first 15 dimensions, and in particular the first eigenvalue is dominating, which indicates that sampling in a rather low-dimensional subspace spanned by the eigenvectors could be sufficient to capture the major variation of the QoI in the parameter space.

To demonstrate the efficacy of the Hessian-based sampling, we project the $129^2$-dimensional parameter $p$ to low-dimensional subspaces as in (29) with $L = 1, 3, 7, 15$, where the sampling is performed efficiently as in (33). We construct the reduced basis space by the POD algorithm in Section 2.5.1 with both random training samples and the Hessian-based training samples of size 1000. Note that the problem is nonaffine due to the log-normal coefficient $\kappa(p) = e^p$, we do not use greedy algorithm which is very expansive without the offline-online decomposition for the evaluation of the dual-weighted residual (19). Affine approximation (e.g., by empirical interpolation) of the log-normal random field is not considered here and can be found in [31] for details. A Hessian-based sampling for empirical interpolation is out of scope of this paper and subject to further investigation. We compute the reduced basis approximation errors for the solution and the QoI defined in (46) with 10 test samples randomly drawn from the whole parameter space. The decay of the errors are shown in the left of Fig. 5 for the solution and in the right of Fig. 5 for the QoI. From the former figure we can see that with only 1 mode for the Hessian-based sampling, the RB error for the solution remains large, while with 3 modes, the RB error becomes much smaller, and with 7 and 15 modes, the RB errors are comparable to and become even smaller than that obtained by random samples at large number of RB basis functions. On the other hand, with 1 mode for the Hessian-based sampling, the RB error for the QoI is already close to that obtained by random samples as seen from the right of Fig. 5. Moreover, with 3, 7, and 15 modes, the RB errors become much smaller than that obtained by random samples, which demonstrate the efficiency of the Hessian-based sampling in capturing the QoI variation in high-dimensional parameter space.

FIG. 5: Decay of the relative error $E_u^N$ (left) and $E_s^N$ (right), defined in (46), of the RB approximation of the solution and the QoI, respectively. $L = 1, 3, 7, 15$ modes are used in the projection (29) for the Hessian-based samples.

5. CONCLUSION

We developed an efficient Hessian-based sampling method to construct goal-oriented reduced order models for high-dimensional parametric problems. Even the dimension of the solution manifold is high due to the high-dimensionality of the parameter space, the QoI related to the solution may live in a low-dimensional manifold. We detected this low-dimensionality by exploring the Hessian of the QoI with respect to the parameter and proposed to sample the parameter from a subspace spanned by the eigenvectors of the Hessian instead of the whole parameter space. For the computation of the eigenpairs of the Hessian, we proposed to use a randomized SVD algorithm, whose cost depends
only on the number of eigenpairs, not on the nominal dimension of the parameter. Based on a diffusion model, we demonstrated that this sampling method leads to much smaller errors of the reduced basis approximation of the QoI for parameters with both uniform distribution and Gaussian distribution.

Further investigation includes adaptive Hessian-based sampling on a systematic way to determine the number of eigenvectors and the required accuracy for the reduced basis approximation of the QoI. Another direction is on the development of the Hessian-based sampling method for nonaffine and nonlinear PDE models and quantities of interest. Moreover, study of the comparison of subspace based sampling methods for function approximation and model reduction using local and global Hessian, as well as gradient information is ongoing.

REFERENCES

1. Alen Alexanderian, Noemi Petra, Georg Stadler, and Omar Ghattas. A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized \( \ell_0 \)-sparsification. *SIAM Journal on Scientific Computing*, 36(5):A2122–A2148, 2014.

2. Alen Alexanderian, Noemi Petra, Georg Stadler, and Omar Ghattas. A fast and scalable method for A-optimal design of experiments for infinite-dimensional Bayesian nonlinear inverse problems. *SIAM Journal on Scientific Computing*, 38(1):A243–A272, 2016.

3. Alen Alexanderian, Noemi Petra, Georg Stadler, and Omar Ghattas. Mean-variance risk-averse optimal control of systems governed by PDEs with random parameter fields using quadratic approximations. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):1166–1192, 2017. arXiv preprint arXiv:1602.07592.

4. H. Avron and S. Toledo. Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix. *Journal of the ACM (JACM)*, 58(2):8, 2011.

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

6. 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.

7. O. Bashir, K. Willcox, O. Ghattas, B. van Bloemen Waanders, and J. Hill. Hessian-based model reduction for large-scale systems with initial condition inputs. *International Journal for Numerical Methods in Engineering*, 73:844–868, 2008.

8. P. Benner, S. Gugercin, and K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM review*, 57(4):483–531, 2015.

9. L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, L. Tenorio, B. van Bloemen Waanders, K. Willcox, and Y. Marzouk. Large-scale inverse problems and quantification of uncertainty. John Wiley & Sons, Hoboken, NJ, 2011.

10. Lorenz T Biegler, Omar Ghattas, Matthias Heinkenschloss, and Bart Bloemen Waanders. Large-scale PDE-constrained optimization: an introduction. *Large-Scale PDE-Constrained Optimization*, pages 3–13, 2003.

11. 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.

12. Francesca Bonizzoni and Fabio Nobile. Perturbation analysis for the Darcy problem with log-normal permeability. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):223–244, 2014.

13. T. Bui-Thanh, K. Willcox, and O. Ghattas. Model reduction for large-scale systems with high-dimensional parametric input space. *SIAM Journal on Scientific Computing*, 30(6):3270–3288, 2008.

14. Tan Bui-Thanh, Carsten Burstedde, Omar Ghattas, James Martin, Georg Stadler, and Lucas C. Wilcox. Extreme-scale UQ for Bayesian inverse problems governed by PDEs. In *SC12: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 2012. Gordon Bell Prize finalist.

15. Tan Bui-Thanh, Murali Damodaran, and Karen E Willcox. Aerodynamic data reconstruction and inverse design using proper orthogonal decomposition. *AIAA journal*, 42(8):1505–1516, 2004.

16. Tan Bui-Thanh and Omar Ghattas. Analysis of the Hessian for inverse scattering problems. Part I: Inverse shape scattering of acoustic waves. *Inverse Problems*, 28(5):055001, 2012.
17. Tan Bui-Thanh and Omar Ghattas. Analysis of the Hessian for inverse scattering problems. Part II: Inverse medium scattering of acoustic waves. *Inverse Problems*, 28(5):055002, 2012.

18. Tan Bui-Thanh and Omar Ghattas. Analysis of the Hessian for inverse scattering problems. Part III: Inverse medium scattering of electromagnetic waves. *Inverse Problems and Imaging*, 7(4):1139–1155, 2013.

19. Tan Bui-Thanh and Omar Ghattas. A scalable MAP solver for Bayesian inverse problems with Besov priors. *Inverse Problems and Imaging*, 9(1):27–54, 2015.

20. Tan Bui-Thanh, Omar Ghattas, James Martin, and Georg Stadler. A computational framework for infinite-dimensional Bayesian inverse problems Part I: The linearized case, with application to global seismic inversion. *SIAM Journal on Scientific Computing*, 35(6):A2494–A2523, 2013.

21. 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.

22. K. Carlberg and C. Farhat. A low-cost, goal-oriented compact proper orthogonal decomposition basis for model reduction of static systems. *International Journal for Numerical Methods in Engineering*, 86(3):381–402, 2011.

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

24. 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.

25. P. Chen and A. Quarteroni. A new algorithm for high-dimensional uncertainty quantification based on dimension-adaptive sparse grid approximation and reduced basis methods. *Journal of Computational Physics*, 298:176–193, 2015.

26. 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.

27. 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.

28. 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.

29. P. Chen, A. Quarteroni, and G. Rozza. Reduced basis methods for uncertainty quantification. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):813–869, 2017.

30. P. Chen and C. Schwab. Model order reduction methods in computational uncertainty quantification. *Handbook of Uncertainty Quantification*, pages 1–53, 2016.

31. P. Chen and Ch. Schwab. Adaptive sparse grid model order reduction for fast bayesian estimation and inversion. In: Garcke J., Pfister D. (eds) Sparse Grids and Applications - Stuttgart 2014. Lecture Notes in Computational Science and Engineering, vol 109. Springer, Cham, 2016.

32. Tiangang Cui, Youssef Marzouk, and Karen Willcox. Scalable posterior approximations for large-scale Bayesian inverse problems. *Computer Methods in Applied Mechanics and Engineering*, 327:147–172, 2017.

33. P. Chen, U. Villa, and O. Ghattas. Taylor approximation and variance reduction for PDE-constrained optimal control problems under uncertainty. *arXiv:1804.04301*, 2018.

34. 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.

35. Benjamin Crestel, Alen Alexanderian, Georg Stadler, and Omar Ghattas. A-optimal encoding weights for nonlinear inverse problems, with application to the Helmholtz inverse problem. *Inverse Problems*, 33(7):074008, 2017.

36. Tiangang Cui, Youssef Marzouk, and Karen Willcox. Scalable posterior approximations for large-scale Bayesian inverse problems via likelihood-informed parameter and state reduction. *Journal of Computational Physics*, 315:363–387, 2016.

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

38. J. Dick, Q.T. Le Gia, and Ch. Schwab. Higher order quasi–Monte Carlo integration for holomorphic, parametric operator equations. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):48–79, 2016.
39. 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.

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

41. H. Elman and Q. Liao. Reduced basis collocation methods for partial differential equations with random coefficients. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1):192–217, 2013.

42. Richard Everson and Lawrence Sirovich. Karhunen–Loève procedure for gappy data. *JOSA A*, 12(8):1657–1664, 1995.

43. Pearl H. Flath, Lucas C. Wilcox, Volkan Akçelik, Judy Hill, Bart van Bloemen Waanders, and Omar Ghattas. Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial Hessian approximations. *SIAM Journal on Scientific Computing*, 33(1):407–432, 2011.

44. Michael B Giles. Multilevel monte carlo methods. *Acta Numerica*, 24:259, 2015.

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

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

47. J.S. Hesthaven, G. Rozza, and B. Stamm. Certified reduced basis methods for parametrized partial differential equations. *SpringerBriefs in Mathematics*, 2015.

48. Tobin Isaac, Noemi Petra, Georg Stadler, and Omar Ghattas. Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet. *Journal of Computational Physics*, 296:348–368, September 2015.

49. D. Kressner and C. Tobler. Low-rank tensor Krylov subspace methods for parametrized linear systems. *SIAM Journal on Matrix Analysis and Applications*, 32(4):1288–1316, 2011.

50. Kody Law, Andrew Stuart, and Kostas Zygalakis. *Data Assimilation*. Springer, 2015.

51. 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.

52. 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.

53. James Martin, Lucas C. Wilcox, Carsten Burstedde, and Omar 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.

54. NC Nguyen, AT Patera, and J Peraire. A 'best points' interpolation method for efficient approximation of parametrized functions. *International journal for numerical methods in engineering*, 73(4):521–543, 2008.

55. Harald Niederreiter. *Random number generation and quasi-Monte Carlo methods*. SIAM, 1992.

56. J. Nocedal and S. Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

57. Anthony Nouy. Low-rank methods for high-dimensional approximation and model order reduction. *arXiv preprint arXiv:1511.01554*, 2015.

58. Noemi Petra, James Martin, Georg Stadler, and Omar Ghattas. A computational framework for infinite-dimensional Bayesian inverse problems: Part II. Stochastic Newton MCMC with application to ice sheet inverse problems. *SIAM Journal on Scientific Computing*, 36(4):A1525–A1555, 2014.

59. A. Quarteroni, A. Manzoni, and F. Negri. *Reduced Basis Methods for Partial Differential Equations: An Introduction*, volume 92. Springer, 2015.

60. 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.

61. Arvind K Saibaba, Jonghyun Lee, and Peter K Kitanidis. Randomized algorithms for generalized hermitian eigenvalue problems with application to computing karhunen–loève expansion. *Numerical Linear Algebra with Applications*, 23(2):314–339, 2016.
Hessian-based sampling for high-dimensional model reduction

62. Christopher Spannring, Sebastian Ullmann, and Jens Lang. A weighted reduced basis method for parabolic PDEs with random data. *arXiv preprint arXiv:1712.07393*, 2017.

63. Marco Tezzele, Francesco Ballarin, and Gianluigi Rozza. Combined parameter and model reduction of cardiovascular problems by means of active subspaces and POD-Galerkin methods. *arXiv preprint arXiv:1711.10884*, 2017.

64. Luca Venturi, Davide Torlo, Francesco Ballarin, and Gianluigi Rozza. Weighted reduced order methods for parametrized partial differential equations with random inputs. *arXiv preprint arXiv:1805.00828*, 2018.

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

**APPENDIX A.**

It is easy to see that \( \mathbb{E}[g^T_p (p - \bar{p})] = 0 \). We only to verify the quadratic term. Let \((\lambda_k, \Phi_k)_{k=1}^K\) be the eigenpairs of \(\mathbb{H}_p\) where \((\Phi_k)_{k=1}^K\) form a complete orthonormal basis in \(\mathbb{R}^K\) (with completion if the rank of \(\mathbb{H}_p\) is smaller than \(K\)). Let \(P_K : \mathbb{R}^K \rightarrow \mathbb{R}^K\) denote a projection operator defined as

\[
P_Kv = \sum_{k=1}^K \Phi_k \Phi_k^T v, \quad \forall v \in \mathbb{R}^K, \tag{A1}\]

As \((\Phi_k)_{k=1}^K\) form a complete orthonormal basis in \(\mathbb{R}^K\), we have \(P_Kv = v, \forall v \in \mathbb{R}^K\). Therefore, by replacing \(p - \bar{p} = P_K(p - \bar{p})\), we have

\[
\mathbb{E}[(p - \bar{p})^T \mathbb{H}_p (p - \bar{p})] = \sum_{k,l} \mathbb{E}[(p - \bar{p})^T \Phi_k \Phi_k^T \mathbb{H}_p \Phi_l \Phi_l^T (p - \bar{p})] \\
= \sum_{k} \lambda_k \mathbb{E}[(p - \bar{p})^T \Phi_k \Phi_k^T (p - \bar{p})] \\
= \sum_{k=1}^K \lambda_k \Phi_k^T \mathbb{E}[(p - \bar{p})(p - \bar{p})^T] \Phi_k \\
= \sum_{k=1}^K \Phi_k^T \mathbb{H}_p \Phi_k = \text{tr} (\mathbb{H}_p). \tag{A2}\]

**APPENDIX B. FROM LOCAL TO GLOBAL HESSIAN**

Note that the Hessian \(\mathbb{H}_p\) is local, evaluated at the mean \(\bar{p}\), which may fail to characterize the variation of the QoI globally in the parameter space. To deal with this issue, we propose two schemes for the computation of a global Hessian—namely, an averaged Hessian and a combined Hessian—to account for the variation of the QoI globally in the parameter space.

**APPENDIX B.1 Averaged Hessian**

As the Hessian at the mean describes the local curvature, which might not capture the important directions in the whole parameter spaces, we can replace the Hessian at the mean by an averaged Hessian defined as

\[
\mathbb{H} = \int_P \mathbb{H}_p d\mu(p) \approx \frac{1}{M} \sum_{m=1}^M \mathbb{H}_p^{m}, \tag{B1}\]

with \(p^m\) sampled according to its probability distribution \(\mu\) in the whole parameter space. Then as in the first case, we can compute the generalized eigenpairs of \((\mathbb{H}, \mathbb{C}^{-1})\) with averaged Hessian \(\mathbb{H}\) and project the parameter in the
subspace spanned by the first $L$ eigenvectors. We remark that the averaged Hessian was used in \[36\] for the projection of a parameter under posterior distribution into a subspace spanned by the dominating eigenvectors in the context of Bayesian inverse problems.

**APPENDIX B.2 Combined Hessian**

Another method to construct the subspace for the parameter projection is to combine all the eigenvectors of Hessian at different locations with suitable compression. Let $(\lambda_k^m, \varphi_k^m)_{k=1}^L$ denote the generalized eigenpairs of $(\mathbb{H}^m, \mathbb{C}^{-1})$ at the random sample $p^m$, $m = 1, \ldots, M$, i.e.,

$$\mathbb{H}^m \varphi_k^m = \lambda_k^m \mathbb{C}^{-1} \varphi_k^m,$$

such that $\varphi_k^m \mathbb{C}^{-1} \varphi_k^m = \delta_{kk'}$, $k, k' = 1, \ldots, L_m$. Then we form the matrix $\Phi \in \mathbb{R}^{K \times N}$ with

$$\Phi = (w_1^1 \varphi_1^1, \ldots, w_L^1 \varphi_L^1, \ldots, w_1^M \varphi_1^M, \ldots, w_L^M \varphi_L^M),$$

with suitable weight $w_k^m$ that reflects the importance of the eigenvector $\varphi_k^m$. A natural choice is $w_k^m = \sqrt{\lambda_k^m}$. Let $\mathbb{S}$ denote a matrix such that $\mathbb{C}^{-1} = \mathbb{S}^T \mathbb{S}$, e.g., $\mathbb{S} = \mathbb{C}^{-1/2}$ or $\mathbb{S}$ represents a Cholesky factorization of $\mathbb{C}^{-1}$. Then we compute the singular value decomposition of $\mathbb{S} \Phi$ for the largest $L$ singular values as

$$\mathbb{S} \Phi = \mathbb{P} \Sigma \mathbb{Q}^T,$$

where $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_L) \in \mathbb{R}^{L \times L}$ is a diagonal matrix with the $L$ largest non-negative singular values on the diagonal, $\mathbb{Q} = (q_1, \ldots, q_L) \in \mathbb{R}^{N \times L}$ are the $L$ right singular vectors. We construct the basis functions for the parameter projection as

$$\varphi_k = \sum_{j=1}^N \frac{1}{\sigma_k} q_{k,n} \varphi_n, \quad k = 1, \ldots, L,$$

where $q_{k,n}$ denotes the $n$-th element of $q_k$. One can verify that $\varphi_k^T \mathbb{C}^{-1} \varphi_k' = \delta_{kk'}$, $k, k' = 1, \ldots, L$. Fig. \[B.6\] displays the comparison of the three different Hessians for the construction of the reduced order model, from which we can see that there is almost no difference between using the local Hessian and the global Hessian for the examples in Section \[4\] Comparison of the differences of the local and global Hessian for both function approximation and model reduction is subject to further investigation.

**APPENDIX C. MULTIPLE QUANTITIES OF INTEREST**

In the case of multiple quantities of interest, $s_j$, $j = 1, \ldots, J$, instead of constructing different reduced order models with independent Hessian-based sampling for each QoI, we can construct a single reduced order model by designing Hessian with the following two approaches as used in Sec. \[3.2\]. Let $\mathbb{H}^j_p$ denote the Hessian for $s_j$ at $p$, $j = 1, \ldots, J$, then we can project the full parameter to the eigenvectors of the averaged Hessian at the mean

$$\mathbb{H}^j_p = \frac{1}{J} \sum_{j=1}^J \mathbb{H}^j_p,$$

or the double averaged Hessian

$$\mathbb{H} = \frac{1}{JM} \sum_{j=1}^J \sum_{m=1}^M \mathbb{H}^j_p^m.$$

Alternatively, we can compute the eigenpairs of the Hessian $\mathbb{H}^j_p$ at mean $\bar{p}$ separately, denoted as $(\lambda_l^j, \varphi_l^j)_{l=1}^L$, and combine them with weight $w_l^j$ (e.g. $w_l^j = \sqrt{\lambda_l^j}$) as

$$\Phi = (w_1^1 \varphi_1^1, \ldots, w_L^1 \varphi_L^1, \ldots, w_1^j \varphi_1^j, \ldots, w_L^j \varphi_L^j).$$
which we compress by SVD as in Appendix B.2 to obtain the dominating singular vectors for the projection of the parameter. Moreover, we may compute the eigenpairs of the sample averaged Hessian with $M$ samples, and combine the eigenvectors as

$$\Phi = (\Phi_1, \ldots, \Phi_J),$$

where $\Phi_j$ is the combined eigenvectors (B3) for each $j = 1, \ldots, J$. Then, similarly we perform SVD compression for $\Phi$ and project the full parameter to the subspace formed by the singular vectors corresponding to the dominating singular values.