

Preconditioners for model order reduction by interpolation and random sketching of operators

Oleg Balabanov* and Anthony Nouy†

Abstract

The performance of projection-based model order reduction methods for solving parameter-dependent systems of equations highly depends on the properties of the operator, which can be improved by preconditioning. In this paper we present strategies to construct a parameter-dependent preconditioner by an interpolation of operator’s inverse. The interpolation is obtained by minimizing a discrepancy between the (preconditioned) operator and the matrix defining the metric of interest. The discrepancy measure is chosen such that its minimization can be efficiently performed online for each parameter value by the solution of a small least-squares problem. Furthermore, we show how to tune the discrepancy measure for improving the quality of Petrov-Galerkin projection or residual-based error estimation. This paper also addresses preconditioning for the randomized model order reduction methods from [Balabanov and Nouy 2019, Part I]. Our methodology can be readily used for efficient and stable solution of ill-conditioned parametric systems and an effective error estimation/certification without the need to estimate expensive stability constants.

The proposed approach involves heavy computations in both offline and online stages that are circumvented by random sketching. The norms of high-dimensional matrices and vectors are estimated by ℓ_2 -norms of their low-dimensional images, called sketches, through random embeddings. For this we extend the framework from [Balabanov and Nouy 2019, Part I] to random embeddings of operators.

Key words— model order reduction, preconditioner, reduced basis, proper orthogonal decomposition, error estimation, random sketching, subspace embedding.

1 Introduction

We consider a large-scale parameter-dependent system of equations

$$\mathbf{A}(\mu)\mathbf{u}(\mu) = \mathbf{b}(\mu), \quad \mu \in \mathcal{P}, \quad (1)$$

where \mathcal{P} is the parameter set. Such a system may result, for instance, from the discretization of a parameter-dependent partial differential equation (PDE). We assume that the solution manifold

*Inria Paris, Sorbonne Université, Université de Paris, LJLL, UMR CNRS 7598, France.

†Centrale Nantes, LMJL, UMR CNRS 6629, France.

$\mathcal{M} := \{\mathbf{u}(\mu) : \mu \in \mathcal{P}\}$ can be well approximated by a projection onto a space of low or moderate dimension. The linear system of equations (1) can then be approximately solved using projection-based model order reduction (MOR) methods such as Reduced Basis (RB) method, Proper Orthogonal Decomposition (POD) and (recycling) Krylov methods (see [5, 6, 17, 22, 23] and the references therein). The performance of projection-based methods highly depends on the properties of the matrix $\mathbf{A}(\mu)$, which can be improved by preconditioning.

Let the solution space be characterized by a weighted Euclidean (or Hermitian) inner product $\langle \cdot, \cdot \rangle_U := \langle \mathbf{R}_U \cdot, \cdot \rangle_2$, where \mathbf{R}_U is some self-adjoint positive definite matrix. More details regarding the problem's setting can be found in Section 2. Let the preconditioner $\mathbf{P}(\mu)$ be an approximate inverse of $\mathbf{A}(\mu)$. Then the (approximate) solution of (1) can be obtained from

$$\mathbf{B}(\mu)\mathbf{u}(\mu) = \mathbf{f}(\mu), \quad \mu \in \mathcal{P}, \quad (2)$$

where $\mathbf{B}(\mu) := \mathbf{R}_U \mathbf{P}(\mu) \mathbf{A}(\mu)$ and $\mathbf{f}(\mu) := \mathbf{R}_U \mathbf{P}(\mu) \mathbf{b}(\mu)$. If $\mathbf{P}(\mu) \mathbf{A}(\mu)$ is close to the identity matrix, then $\mathbf{B}(\mu)$ should have better properties than the original operator $\mathbf{A}(\mu)$, which implies better performance of projection-based methods. In particular, if $\mathbf{P}(\mu) = \mathbf{A}(\mu)^{-1}$ then (2) is perfectly conditioned relatively to the metric induced by \mathbf{R}_U . It is important to note that in the context of projection-based MOR, the invertibility of $\mathbf{P}(\mu)$ is not required for obtaining an approximate solution to (1). Since we operate only on a subset of vectors it is sufficient to ensure that $\mathbf{P}(\mu) \mathbf{A}(\mu)$ is close to the identity on this subset. Note also that the computation of the explicit form of $\mathbf{B}(\mu)$ can be extremely expensive and has to be avoided. Instead, this matrix should be operated as an implicit map outputting products with vectors.

Here we provide efficiently computable estimators of the quality of $\mathbf{B}(\mu)$ for the solution of (2) with projection-based methods or for residual-based error estimation. Each estimator basically measures a discrepancy between $\mathbf{B}(\mu)$ and \mathbf{R}_U with respect to a certain semi-norm, and is seen as an error indicator on $\mathbf{P}(\mu)$ as an approximation of the inverse of $\mathbf{A}(\mu)$. The proposed error indicators can be readily employed to efficiently construct $\mathbf{P}(\mu)$ by interpolation of the inverse of $\mathbf{A}(\mu)$ or to estimate the stability constants associated with the given parameter-dependent preconditioner. Unlike the minimization of the condition number of $\mathbf{B}(\mu)$ or the classical stability constants, the minimization of each error indicator over a low-dimensional space of matrices is a small least-squares problem, which can be efficiently solved online. The heavy offline computations are here circumvented with randomized linear algebra. More specifically, a drastic reduction of the computational cost is attained by the usage of the framework from [3] and its extension to the context of approximation of inner products between matrices. The ℓ_2 -embeddings are no longer seen as matrices, but rather as linear maps from a space of matrices to a low-dimensional Euclidean (or Hermitian) space. In Section 5 we propose a probabilistic way for the construction of such ℓ_2 -embeddings with the precise conditions on the size of the random sketch to guarantee a user-specified accuracy of approximation and probability of success.

The construction of an efficient parameter-dependent preconditioner was addressed in [11, 13, 21, 24, 31]. In particular, in [31] the authors proposed to use randomized linear algebra for the efficient construction of a preconditioner by interpolation of matrix inverse. This principle is taken as the starting point for the present paper. The Galerkin methods with an improved stability for ill-conditioned and non-coercive parametric equations were developed in [1, 28]. The effective/numerically stable, yet efficient error estimation/certification was addressed in [10, 15, 27] with deterministic approaches, and in [3, 9,

14, 25, 26, 29] with statistical learning/randomized approaches.

The randomized linear algebra recently became a popular approach for the enhancement of MOR methods [3, 4, 8, 9, 18, 25, 26, 31]. It was first used as a simple sample-based approach for error estimation [9, 18]. These ideas then underwent a more significant development in [3, 4, 25, 26]. It has to be noted that the probabilistic error estimator from [26] has a close relation to the multi-purpose preconditioner-based error estimator proposed in the present paper (see Section 1.2 for details). In [2] the low-rank approximation algorithm from [16] was used for enhancing the efficiency of computation of the POD basis. A POD method, based on the simultaneous low-rank approximation of the matrix of snapshots and the estimation of the reduced order model with random sketching was proposed in [3]. In [4], the authors employed for MOR the ideas from compressed sensing [30], based on random sketching.

1.1 Construction of a preconditioner

A parameter-dependent preconditioner $\mathbf{P}(\mu)$ is here obtained by a projection of $\mathbf{A}(\mu)^{-1}$ onto a linear span of some basis matrices $\{\mathbf{Y}_i : 1 \leq i \leq p\}$, i.e.,

$$\mathbf{P}(\mu) = \sum_{i=1}^p \lambda_i(\mu) \mathbf{Y}_i, \quad (3)$$

with coefficients $\lambda_i(\mu)$ computed online by minimizing some discrepancy measure between $\mathbf{P}(\mu)$ and $\mathbf{A}(\mu)^{-1}$.

Throughout the paper, the basis matrices \mathbf{Y}_i are considered to be $\mathbf{A}(\mu^i)^{-1}$ at some interpolation points $\mu^i \in \mathcal{P}$, but they could also be chosen as approximate inverses. The set of interpolation points can be simply obtained by random sampling in \mathcal{P} . Another way is an iterative greedy selection, at each iteration enriching the set of interpolation points by a parameter value where the error indicator is the largest. For methods where the approximation space U_r is constructed from snapshots $\mathbf{u}(\hat{\mu}^j)$ as RB or POD methods, the interpolation points can be selected among the parameters $\hat{\mu}^j$, providing recycling of the computations, since each snapshot (typically) requires computation of the implicit inverse (e.g., factorization) of the operator. Finally, for the reduced basis methods where U_r is constructed with a greedy algorithm based on Petrov-Galerkin projection, it can be useful to consider the same interpolation points for the construction of U_r and $\{\mathbf{Y}_i : 1 \leq i \leq p\}$. In this case the error indicator for the greedy selection of an interpolation point should be defined as a (weighted) average of the error indicator characterizing the quality of U_r (e.g., an upper bound of the error of the Galerkin projection) and the error indicator characterizing the quality of $\{\mathbf{Y}_i : 1 \leq i \leq p\}$ (e.g., one of the error indicators presented below). Other strategies for finding the parameter values μ^i can be found in [31].

1.2 Contributions

We here consider preconditioners in several contexts. Besides the multi-purpose context as in [31], where one is interested in minimization of the condition number, we also consider Galerkin projection onto a fixed approximation space, and residual-based error certification. We also cover the case of preconditioners for the randomized methods from [3]. A detailed presentation of the major contributions is given below.

Preconditioner for multi-purpose context

This work presents a generalization and improvement of the methodology introduced in [31] using random sketching of operators. First of all, the quality of the preconditioner is characterized with respect to a general norm represented by a self-adjoint positive definite matrix instead of the ℓ_2 -norm. This is important, for instance, in the context of numerical methods for PDEs to control the quality of an approximation regardless the used discretization. Secondly, the theoretical bounds from [31] for the size of sketching matrices are considerably improved. For instance our bound for Subsampled Randomized Hadamard Transform (SRHT) matrix is linear in the dimension p of the low-dimensional space of operators, and not quadratic as in [31]. Furthermore, thanks to the (extended) framework presented in [3], we here obtain a great improvement of the efficiency (both offline and online) and numerical stability of the algorithms. More specifically, if $\mathbf{A}(\mu)$ is a $n \times n$ sparse matrix and admits an affine expansion with m_A terms¹, if $\mathbf{P}(\mu)$ is a linear combination of p basis matrices, each requiring $\mathcal{O}(nk_P)$ (for some small $k_P > 1$) complexity and amount of storage for multiplication by a vector, and if k is the number of rows of the sketching matrices (typically $k = \mathcal{O}(p)$), then the precomputation of our error indicator, using SRHT, takes only $\mathcal{O}(nkp(m_A \log(k) + k_P))$ flops and $\mathcal{O}(nk_P)$ bytes of memory, while the approach from [31] can require up to $\mathcal{O}(nkp m_A^2(kp + k_P))$ flops and $\mathcal{O}(n(kp + k_P))$ bytes of memory. Moreover, we also improve the efficiency and numerical stability of the online stage. The online assembling of the reduced matrix for the computation (or minimization) of the indicator in [31] takes $\mathcal{O}(m_A^2 p^2)$ flops, while our approach essentially consumes only $\mathcal{O}(m_A p^2)$ flops. Our approach is also less sensitive to round-off errors since we proceed with direct solution of the least-squares problem without the need to appeal to the normal equation. We also derive a quasi-optimality result for the preconditioned Galerkin projection and error estimation with the proposed error indicator.

Preconditioner for Galerkin projection

The estimation of the operator norm by a Hilbert-Schmidt norm (as used in the multi-purpose context) can be very ineffective. In general a very high overestimation is possible. For numerical methods for PDEs, this may result in a high sensitivity to discretization. We show how to overcome this issue, if the preconditioner is used for a Galerkin projection onto a small or moderately large approximation space. In such a case effective error indicators can be obtained by ignoring the component of the residual which is orthogonal to the approximation space.

Preconditioner for error certification

Our methodology can be used for efficient and effective error estimation/certification without the need to estimate expensive stability constants. The error $\|\mathbf{u}(\mu) - \mathbf{u}_r(\mu)\|_U$ of an approximation $\mathbf{u}_r(\mu)$ of the solution $\mathbf{u}(\mu)$ can be estimated by a sketched norm [3] of the preconditioned residual $\mathbf{f}(\mu) - \mathbf{B}(\mu)\mathbf{u}_r(\mu)$. This approach is related to the one from [26], which consists in approximating the error by projections of the (unpreconditioned) residual onto approximate solutions of the dual problems $\mathbf{A}(\mu)^H \mathbf{y}_i(\mu) = \mathbf{z}_i$ with random right-hand sides. The difference is that in [26] the authors proposed to tackle the random dual

¹A parameter-dependent quantity $\mathbf{v}(\mu)$ with values in vector space V over a field \mathbb{K} is said to admit an affine representation if $\mathbf{v}(\mu) = \sum_{i=1}^d \mathbf{v}_i \lambda_i(\mu)$ with $\lambda_i(\mu) \in \mathbb{K}$ and $\mathbf{v}_i \in V$.

problems separately with RB methods, while we here consider a monolithic approach, approximating solutions by $\mathbf{y}_i(\mu) \approx \mathbf{P}(\mu)^H \mathbf{z}_i$, where $\mathbf{P}(\mu)$ is a preconditioner constructed by an interpolation of the operator's inverse based on minimization of an error indicator. Our method has several important advantages over the one in [26]. First, our efficient error certification procedure with the multi-purpose error indicator does not rely on the assumption that the error of the solution(s) of the random dual problem(s) is uniformly small on \mathcal{P} as in [26]. Furthermore, in contrast to [26] our methodology yields guarantees of success not only for finite parameter sets \mathcal{P} , which can be of particular interest for adaptive algorithms. Finally, we propose an additional, more robust approach for error estimation without requiring $\mathbf{A}(\mu)^{-1}$ (or $\mathbf{y}_i(\mu)$) to be well-approximated (uniformly over \mathcal{P}) in a low-dimensional space. For this we restrict the operator to a linear subspace that approximates well the solution manifold. Remarkably, the complexity of online computations in our approach does not depend on the dimension of the subspace in contrast to the existing methods (such as [15]) typically requiring online computation of a projection onto this subspace. This allows to choose a subspace of rather high dimension without deteriorating online efficiency.

Preconditioner for MOR with random sketching

A particular interest is given to construction of preconditioners for large-scale parametric problems tackled with randomized methods from [3] for a drastic reduction of the offline cost. We show how to tune preconditioner's error indicator to improving the quasi-optimality constants of the randomized, sketched Galerkin projection and error estimator from [3], with the precise guarantees of the quality of the preconditioner. The resulting method preserves the efficiency of the randomized methods so that it can be effectively executed with any computational architecture.

1.3 Outline

The paper is organized as follows. Section 2 introduces some notations. In Section 3 we provide indicators that quantify the quality of the preconditioner for different objectives such as multi-purpose context, Galerkin projection and residual-based error estimation. The measures of the quality of the preconditioner in the context of the probabilistic MOR methods from [3] are discussed in Section 4. The indicators from Sections 3 and 4 are efficiently estimated by using random sketching in Section 5. For this we first extend the framework from [3] to random embeddings of operators and then particularize the results to each indicator individually. The numerical validation of the methodology will be presented in the nearest future.

2 Preliminaries

Let $\mathbb{K} = \mathbb{R}$ or \mathbb{C} . The solution space is identified with $U := \mathbb{K}^n$. This space is equipped with inner product

$$\langle \cdot, \cdot \rangle_U := \langle \mathbf{R}_U \cdot, \cdot \rangle_2,$$

where $\langle \cdot, \cdot \rangle_2$ is the ℓ_2 -inner product on \mathbb{K}^n and $\mathbf{R}_U \in \mathbb{K}^{n \times n}$ is some self-adjoint (symmetric if $\mathbb{K} = \mathbb{R}$ and Hermitian if $\mathbb{K} = \mathbb{C}$) positive definite matrix. The dual of U is identified with $U' := \mathbb{K}^n$ and is endowed

with the canonical (dual) norm

$$\|\cdot\|_{U'} = \max_{\mathbf{w} \in U} \frac{|\langle \cdot, \mathbf{w} \rangle_2|}{\|\mathbf{w}\|_U}.$$

This norm is associated with the inner product $\langle \cdot, \cdot \rangle_{U'} := \langle \cdot, \mathbf{R}_U^{-1} \cdot \rangle_2$. The solution vector $\mathbf{u}(\mu)$ is seen as an element from U , the matrices $\mathbf{A}(\mu)$ and \mathbf{R}_U are seen as operators from U to U' , and $\mathbf{b}(\mu)$ is seen as an element from U' . The parameter set \mathcal{P} can be a subset of \mathbb{K}^l or a subset of an infinite dimensional space such as a function space. See [3] for more details on the meaning of this semi-discrete setting for numerical methods for PDEs. For problems described simply by algebraic equations the notions of solution spaces and dual spaces can be disregarded and matrix \mathbf{R}_U can be taken as identity.

For finite-dimensional (Hilbert) spaces V and W identified with an Euclidean or a Hermitian space, we denote by $HS(V, W)$ the space of matrices representing operators from V to W . Assuming that V and W are equipped with inner products $\langle \cdot, \cdot \rangle_V$ and $\langle \cdot, \cdot \rangle_W$, respectively, we endow $HS(V, W)$ with the Hilbert-Schmidt inner product

$$\langle \mathbf{X}, \mathbf{Y} \rangle_{HS(V, W)} := \sum_{i=1}^{\dim V} \langle \mathbf{X} \mathbf{v}_i, \mathbf{Y} \mathbf{v}_i \rangle_W,$$

where $\mathbf{X}, \mathbf{Y} : V \rightarrow W$ and $\{\mathbf{v}_i : 1 \leq i \leq \dim V\}$ is some orthonormal basis of V . Below we particularize the above setting to specific choices of V and W .

For $V = \ell_2(\mathbb{K}^r)$ and $W = \ell_2(\mathbb{K}^k)$, $HS(V, W)$ is identified with the space of matrices $\mathbb{K}^{k \times r}$ equipped with the Frobenius inner product $\langle \cdot, \cdot \rangle_{HS(\ell_2, \ell_2)} = \langle \cdot, \cdot \rangle_F$.

For $V = \ell_2(\mathbb{K}^r)$ and $W = U$ or $W = U'$, $HS(V, W)$ is identified with the space of matrices $\mathbb{K}^{n \times r}$ equipped with the inner products

$$\langle \cdot, \cdot \rangle_{HS(\ell_2, U)} = \langle \mathbf{R}_U \cdot, \cdot \rangle_F, \text{ or } \langle \cdot, \cdot \rangle_{HS(\ell_2, U')} = \langle \cdot, \mathbf{R}_U^{-1} \cdot \rangle_F,$$

respectively.

Furthermore, $HS(U, U')$ and $HS(U', U)$ are identified with $\mathbb{K}^{n \times n}$. These spaces are seen as spaces of linear operators from U to U' and from U' to U , respectively, and are endowed with inner products

$$\langle \cdot, \cdot \rangle_{HS(U, U')} := \langle \cdot, \mathbf{R}_U^{-1} \cdot \mathbf{R}_U^{-1} \rangle_F \text{ and } \langle \cdot, \cdot \rangle_{HS(U', U)} := \langle \mathbf{R}_U \cdot \mathbf{R}_U, \cdot \rangle_F. \quad (4)$$

We also let $\|\cdot\|_{HS(\ell_2, U)}$, $\|\cdot\|_{HS(\ell_2, U')}$, $\|\cdot\|_{HS(U, U')}$ and $\|\cdot\|_{HS(U', U)}$ be the associated norms.

Finally, the minimal and the maximal singular values (or inf-sup constant and operator norm) of a matrix $\mathbf{C} \in HS(U, U')$, seen as an operator from U to U' , are defined as

$$\alpha(\mathbf{C}) := \min_{\mathbf{v} \in U \setminus \{0\}} \frac{\|\mathbf{C} \mathbf{v}\|_{U'}}{\|\mathbf{v}\|_U}, \quad (5a)$$

$$\beta(\mathbf{C}) := \max_{\mathbf{v} \in U \setminus \{0\}} \frac{\|\mathbf{C} \mathbf{v}\|_{U'}}{\|\mathbf{v}\|_U} =: \|\mathbf{C}\|_{U, U'}, \quad (5b)$$

and the condition number $\kappa(\mathbf{C}) := \frac{\beta(\mathbf{C})}{\alpha(\mathbf{C})}$. Supposing that $V, W \subseteq U$, the maximal singular value of the operator \mathbf{C} , restricted to V and used as an operator from V to $W' := \{\mathbf{R}_U \mathbf{w} : \mathbf{w} \in W\}$, is given by

$$\|\mathbf{C}\|_{V, W'} := \max_{\mathbf{v} \in V \setminus \{0\}} \frac{\|\mathbf{C} \mathbf{v}\|_{W'}}{\|\mathbf{v}\|_U}, \quad (6)$$

where

$$\|\cdot\|_{W'} := \max_{\mathbf{w} \in W \setminus \{\mathbf{0}\}} \frac{|\langle \cdot, \mathbf{w} \rangle_2|}{\|\mathbf{w}\|_U}.$$

Note that choosing $V = W = U$ in (6) is consistent with the definition (5b).

3 Measures of quality of a preconditioner for classical projection-based methods

In this section we derive computable error indicators that measure the quality of a preconditioner for classical projection-based methods. They essentially represent a discrepancy between $\mathbf{P}(\mu)$ and $\mathbf{A}(\mu)^{-1}$. Different error indicators (or discrepancy measures) shall be considered depending on the objectives.

Further, all considerations are for a fixed parameter value $\mu \in \mathcal{P}$, unless specified otherwise. To simplify the presentation, the dependencies on μ are omitted. We let $\mathbf{E} := \mathbf{R}_U(\mathbf{I} - \mathbf{P}\mathbf{A}) = \mathbf{R}_U - \mathbf{B}$ be the error matrix.

3.1 Multi-purpose context

Here we consider the preconditioned system of equations (2) and provide an error indicator that characterizes the performance of the preconditioner for projection-based methods such as (possibly adaptive or randomized) Galerkin methods, Krylov methods (with or without recycling), RB methods, etc.

The matrix \mathbf{B} in (2) can be seen as a linear operator from U to U' . The performance of a projection-based method and a residual-based error estimator usually depends on the condition number $\kappa(\mathbf{B})$. A smaller condition number yields better stability constants. The value of $\kappa(\mathbf{B})$ can be characterized by the distance between \mathbf{B} and \mathbf{R}_U measured with the operator norm, i.e., by $\|\mathbf{E}\|_{U,U'}$. More specifically, it directly follows from the definitions (5) of the minimal and maximal singular values that

$$1 - \|\mathbf{E}\|_{U,U'} \leq \alpha(\mathbf{B}) \leq \beta(\mathbf{B}) \leq 1 + \|\mathbf{E}\|_{U,U'}. \quad (7)$$

The minimization of $\|\mathbf{E}\|_{U,U'}$ for multiple operators \mathbf{B} may be an unfeasible task. Therefore the condition number of \mathbf{B} shall be approximated with a computable upper bound of $\|\mathbf{E}\|_{U,U'}$.

Proposition 3.1. *For an operator $\mathbf{C} : U \rightarrow U'$ and a vector $\mathbf{v} \in U$, it holds*

$$\|\mathbf{C}\mathbf{v}\|_{U'} \leq \|\mathbf{C}\|_{HS(U,U')} \|\mathbf{v}\|_U. \quad (8)$$

Proof. See appendix. □

From Proposition 3.1 it follows that $\|\mathbf{E}\|_{HS(U,U')}$ is an upper bound of $\|\mathbf{E}\|_{U,U'}$, which implies the first main result of this paper.

Define the following error indicator

$$\boxed{\Delta_{U,U'} = \|\mathbf{E}\|_{HS(U,U')}}. \quad (9)$$

If $\Delta_{U,U'} < 1$, then

$$\kappa(\mathbf{B}) \leq \frac{1 + \Delta_{U,U'}}{1 - \Delta_{U,U'}}. \quad (10)$$

Therefore a good performance of a projection-based method can be guaranteed if $\Delta_{U,U'}$ is sufficiently small.

In practice, the condition $\Delta_{U,U'} < 1$, which is required for the bound (10) to hold, is hard to reach. Our empirical studies, however, suggest that the operators which come from real applications have a small condition number also when $\Delta_{U,U'}$ is small but larger than one.

In general, a good effectivity of $\|\cdot\|_{HS(U,U')}$ as an estimator of the operator norm $\|\cdot\|_{U,U'}$ may not be guaranteed. In some situations, a large overestimation (up to a factor of $n^{1/2}$) happens. This issue can be particularly dramatic for numerical methods for PDEs, where each discrete operator \mathbf{C} (e.g., $\mathbf{C} = \mathbf{E}$) represents a finite-dimensional approximation of some partial differential operator C . The operator norm of C is an upper bound of $\|\mathbf{C}\|_{U,U'}$ regardless of the chosen discretization. The norm $\|\mathbf{C}\|_{HS(U,U')}$ is an approximation of the Hilbert-Schmidt norm of C , which can be infinite (if C is not a Hilbert-Schmidt operator). Therefore, even if C has a small operator norm (implying that $\|\mathbf{C}\|_{U,U'}$ is also small), $\|\mathbf{C}\|_{HS(U,U')}$ can be highly sensitive to the discretization and tend to infinity with n . This implies a possible failure of $\Delta_{U,U'}$ for characterizing the quality of the preconditioned operator. This problem can be circumvented for the projection-based MOR context, where the solution is approximated in a moderately large space, or for the residual-based error estimation.

3.2 Galerkin projection

Further, we consider the projection-based MOR context where the solution \mathbf{u} in (2) is approximated by the Galerkin projection \mathbf{u}_r onto a subspace $U_r \subseteq U$. Let $\mathbf{U}_r \in \mathbb{K}^{n \times r}$ be a matrix whose columns form a basis for U_r . The subspace U_r can be constructed with a greedy algorithm for RB method or low-rank approximation of the matrix of solution samples, called snapshots, for POD. The basis vectors for U_r can also be chosen a priori by exploiting the structure of the problem. In the context of numerical methods for PDEs, such basis vectors can be obtained by computing the coordinates of the basis functions (associated, for instance, with an approximation on a coarse grid) on the space of functions identified with U .

For given $W \subseteq U$, let $\mathbf{\Pi}_W : U \rightarrow W$ denote the orthogonal projection on W with respect to $\|\cdot\|_U$, such that

$$\forall \mathbf{x} \in U, \mathbf{\Pi}_W \mathbf{x} = \arg \min_{\mathbf{w} \in W} \|\mathbf{x} - \mathbf{w}\|_U. \quad (11)$$

The Galerkin orthogonality condition can be stated as follows

$$\langle \mathbf{B}(\mathbf{u} - \mathbf{u}_r), \mathbf{w} \rangle_2 = 0, \quad \forall \mathbf{w} \in U_r, \quad (12)$$

or, equivalently (see [3]),

$$\|\mathbf{B}(\mathbf{u} - \mathbf{u}_r)\|_{U'_r} = 0. \quad (13)$$

Next we use the following lemma to provide conditions for controlling the accuracy of \mathbf{u}_r in Proposition 3.3.

Lemma 3.2. *Let \mathbf{u}_r satisfy (13). Then*

$$\|\mathbf{u}_r - \Pi_{U_r} \mathbf{u}\|_U \leq \|\mathbf{E}(\mathbf{u} - \Pi_{U_r} \mathbf{u})\|_{U'_r} + \|\mathbf{E}(\mathbf{u}_r - \Pi_{U_r} \mathbf{u})\|_{U'_r}. \quad (14)$$

Proof. See appendix. \square

Proposition 3.3. *If $\|\mathbf{E}\|_{U_r, U'_r} < 1$, then the solution \mathbf{u}_r to (13) is unique and*

$$\|\mathbf{u} - \mathbf{u}_r\|_U \leq \left(1 + \frac{\|\mathbf{E}\|_{U, U'_r}}{1 - \|\mathbf{E}\|_{U_r, U'_r}}\right) \|\mathbf{u} - \Pi_{U_r} \mathbf{u}\|_U. \quad (15)$$

Proof. See appendix. \square

According to Proposition 3.3, the quasi-optimality of \mathbf{u}_r can be guaranteed by making sure that the semi-norms $\|\mathbf{E}\|_{U_r, U'_r}$ and $\|\mathbf{E}\|_{U, U'_r}$ are small enough. We observe that

$$\|\mathbf{E}\|_{U_r, U'_r} \leq \|\mathbf{E}\|_{U, U'_r} \leq \|\mathbf{E}\|_{U, U'} \leq \Delta_{U, U'}. \quad (16)$$

Moreover, for $U_r = U$ we clearly have $\|\mathbf{E}\|_{U_r, U'_r} = \|\mathbf{E}\|_{U, U'_r} = \|\mathbf{E}\|_{U, U'}$. The relation (16) and Proposition 3.3 yield an error bound for the Galerkin projection using the multi-purpose indicator $\Delta_{U, U'}$:

$$\|\mathbf{u} - \mathbf{u}_r\|_U \leq \left(1 + \frac{\Delta_{U, U'}}{1 - \Delta_{U, U'}}\right) \|\mathbf{u} - \Pi_{U_r} \mathbf{u}\|_U. \quad (17)$$

However, the quality of the preconditioner can be better characterized by taking into account the fact that the error norms $\|\mathbf{E}\|_{U_r, U'_r}$ and $\|\mathbf{E}\|_{U, U'_r}$ represent a discrepancy between \mathbf{B} and \mathbf{R}_U with the solution space and/or the test space being restricted to U_r (see Proposition 3.4).

Proposition 3.4. *Define*

$$\boxed{\Delta_{U_r, U'_r} := \|\mathbf{U}_r^H \mathbf{E} \mathbf{U}_r\|_F} \quad \text{and} \quad \boxed{\Delta_{U, U'_r} := \|\mathbf{E}^H \mathbf{U}_r\|_{HS(\ell_2, U')}}. \quad (18)$$

The following relations hold:

$$\frac{1}{\sigma_1^2 \sqrt{r}} \Delta_{U_r, U'_r} \leq \|\mathbf{E}\|_{U_r, U'_r} \leq \frac{1}{\sigma_r^2} \Delta_{U_r, U'_r} \quad \text{and} \quad \frac{1}{\sigma_1 \sqrt{r}} \Delta_{U, U'_r} \leq \|\mathbf{E}\|_{U, U'_r} \leq \frac{1}{\sigma_r} \Delta_{U, U'_r}, \quad (19)$$

where

$$\sigma_r := \min_{\mathbf{a} \in \mathbb{K}^r / \{\mathbf{0}\}} \frac{\|\mathbf{U}_r \mathbf{a}\|_U}{\|\mathbf{a}\|_2} \quad \text{and} \quad \sigma_1 := \max_{\mathbf{a} \in \mathbb{K}^r / \{\mathbf{0}\}} \frac{\|\mathbf{U}_r \mathbf{a}\|_U}{\|\mathbf{a}\|_2}$$

are the minimal and the maximal singular values of \mathbf{U}_r with respect to $\|\cdot\|_U$ -norm.

Proof. See appendix. \square

Clearly, the bounds in Proposition 3.4 are tighter when the columns of \mathbf{U}_r are orthonormal with respect to $\langle \cdot, \cdot \rangle_U$.

Corollary 3.5. Let $\Delta_{U_r, U_r'}$ and $\Delta_{U, U_r'}$ be the error indicators from Proposition 3.4. If the columns of \mathbf{U}_r are orthonormal vectors with respect to $\langle \cdot, \cdot \rangle_U$, then

$$\frac{1}{\sqrt{r}}\Delta_{U_r, U_r'} \leq \|\mathbf{E}\|_{U_r, U_r'} \leq \Delta_{U_r, U_r'} \quad \text{and} \quad \frac{1}{\sqrt{r}}\Delta_{U, U_r'} \leq \|\mathbf{E}\|_{U, U_r'} \leq \Delta_{U, U_r'}. \quad (20)$$

Furthermore, it is easy to see that if \mathbf{U}_r has orthonormal columns with respect to $\langle \cdot, \cdot \rangle_U$, then

$$\Delta_{U_r, U_r'} \leq \Delta_{U, U_r'} \leq \Delta_{U, U'},$$

which implies that the quasi-optimality constants obtained with $\Delta_{U_r, U_r'}$ and $\Delta_{U, U_r'}$ shall always be better than the ones obtained with $\Delta_{U, U'}$. Note that if $U_r = U$ then $\Delta_{U_r, U_r'} = \Delta_{U, U_r'} = \Delta_{U, U'}$. Unlike the multi-purpose context, here the effectiveness of $\Delta_{U_r, U_r'}$ and $\Delta_{U, U_r'}$ as estimators of $\|\mathbf{E}\|_{U_r, U_r'}$ and $\|\mathbf{E}\|_{U, U_r'}$ is guaranteed. For PDEs this implies a robust characterization of the quality of the preconditioned operator regardless the discretization.

Note that U_r can have a relatively high dimension making the orthogonalization of the basis with respect to $\langle \cdot, \cdot \rangle_U$ very expensive. In this case one should use \mathbf{U}_r with columns that are only approximately orthogonal. For PDEs such a matrix can be obtained with domain decomposition (i.e., by local orthogonalization).

Let us now summarize the results of Propositions 3.3 and 3.4 in a practical form.

Consider error indicators $\Delta_{U_r, U_r'}$, $\Delta_{U, U_r'}$ defined in (18) and $\Delta_{U, U'}$ defined in (9).

If $\min \{ \Delta_{U, U'}, \sigma_r^{-1} \Delta_{U, U_r'}, \sigma_r^{-2} \Delta_{U_r, U_r'} \} < 1$, then the solution \mathbf{u}_r to (13) is such that

$$\|\mathbf{u} - \mathbf{u}_r\|_U \leq \left(1 + \frac{\min \{ \Delta_{U, U'}, \sigma_r^{-1} \Delta_{U, U_r'} \}}{1 - \min \{ \Delta_{U, U'}, \sigma_r^{-1} \Delta_{U, U_r'}, \sigma_r^{-2} \Delta_{U_r, U_r'} \}} \right) \|\mathbf{u} - \mathbf{\Pi}_{U_r} \mathbf{u}\|_U, \quad (21)$$

where σ_r is the minimal singular value of \mathbf{U}_r with respect to $\|\cdot\|_U$ -norm. Therefore to ensure high quality of the Galerkin projection one can seek a preconditioner that minimizes

$$\Delta_{U, U'}, \quad \text{or} \quad \Delta_{U, U_r'}, \quad \text{or} \quad \gamma \Delta_{U, U_r'}^2 + \Delta_{U_r, U_r'}^2, \quad (22)$$

with a weight γ selected depending on the problem.

Remark 3.6. The minimization of $\Delta_{U, U'}$ shall be the right choice when the subspace U_r is unknown, and $\Delta_{U, U_r'}$ and $\Delta_{U_r, U_r'}$ can not be computed, or when \mathbf{U}_r has a high condition number and may not be approximately orthogonalized. On the other hand, if \mathbf{U}_r has a moderately high condition number, then the minimization of $\Delta_{U, U_r'}$ would provide a better quasi-optimality constant than the minimization of $\Delta_{U, U'}$.

If \mathbf{U}_r has a condition number sufficiently close to 1, then the quasi-optimality constant in (21) is $K = 1 + \frac{\sigma_r^{-1} \Delta_{U, U_r'}}{1 - \sigma_r^{-2} \Delta_{U_r, U_r'}}$. In this case, a good Galerkin projection should be obtained by minimizing a weighted sum $\gamma \Delta_{U, U_r'}^2 + \Delta_{U_r, U_r'}^2$. Next we propose a way to choose the weight γ . Let us consider a situation where one seeks a preconditioner yielding a quasi-optimality constant $K \leq K^*$, with a user-specified parameter K^* . Moreover, it is assumed that $K^* \geq 3/2$, since having the quasi-optimality constant of $3/2$ is sufficient for most applications. We notice that the indicator $\Delta_{U_r, U_r'}$ can have a high impact on K when

$\Delta_{U_r, U_r'} \geq 1/3\sigma_r^2$, but if $\Delta_{U_r, U_r'} \leq 1/3\sigma_r^2$, the quasi-optimality constant can be bounded by $1 + 3/2\sigma_r^{-1}\Delta_{U_r, U_r'}$ and there will be almost no effect of further minimization of $\Delta_{U_r, U_r'}$. From this consideration, one can derive the value $\gamma = 1/((\frac{2(K^*-1)}{\sigma_r})^2 - 1)$. It can be shown that for such a value of γ , the target quasi-optimality, $K \leq K^*$, is attained if $\frac{1}{\sqrt{1+\gamma}}(\gamma\Delta_{U_r, U_r'}^2 + \Delta_{U_r, U_r'}^2)^{1/2} \leq 1/3\sigma_r^2$.

Again, the condition $\Delta_{U_r, U_r'} < \sigma_r^2$ (or $\Delta_{U_r, U_r'} < \sigma_r$) can require too expensive preconditioners and may not be attained for some problems. Without this condition we do not have any a priori guaranty of quality of the Galerkin projection. On the other hand, our experimental observations revealed that, in practice, minimizing $\Delta_{U_r, U_r'}$ and $\Delta_{U, U'}$ yields good preconditioners even when $\Delta_{U_r, U_r'} \geq \sigma_r^2$.

3.3 Error estimation

Let $\mathbf{u}_r \in U_r$ be an approximation of \mathbf{u} in the subspace U_r . Here we address the question of estimating and bounding the error $\|\mathbf{u} - \mathbf{u}_r\|_U$. The standard way is the certification of the error with the residual norm:

$$\|\mathbf{u} - \mathbf{u}_r\|_U \leq \frac{1}{\eta} \|\mathbf{r}(\mathbf{u}_r)\|_{U'}, \quad (23)$$

where $\mathbf{r}(\mathbf{u}_r) = \mathbf{b} - \mathbf{A}\mathbf{u}_r$ and η is a computable lower bound of the smallest singular value of \mathbf{A} (the operator norm of \mathbf{A}^{-1}). For ill-conditioned operators \mathbf{A} , the accuracy of such error estimator is very poor. A straightforward approach to overcome this issue is to replace \mathbf{A} in (23) by the preconditioned operator \mathbf{B} and use a preconditioned residual norm:

$$\|\mathbf{u} - \mathbf{u}_r\|_U \leq \frac{1}{\eta^*} \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'}, \quad (24)$$

with $\mathbf{r}^*(\mathbf{u}_r) = \mathbf{f} - \mathbf{B}\mathbf{u}_r$ and η^* a lower bound of the smallest singular value of \mathbf{B} . The coefficients η and η^* can be obtained theoretically in some particular problems or with the Successive Constraint Method [12, 19, 20]. The above approach can be intractable, since the computation of η and (especially) η^* with classical procedures can be very expensive. A more efficient certification of the error can be obtained with a multi-purpose error indicator $\Delta_{U, U'}$, as proposed in Proposition 3.7.

Proposition 3.7. *If $\Delta_{U, U'} < 1$, then*

$$\frac{1}{1 + \Delta_{U, U'}} \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'} \leq \|\mathbf{u} - \mathbf{u}_r\|_U \leq \frac{1}{1 - \Delta_{U, U'}} \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'}. \quad (25)$$

Proof. See appendix. □

The advantage of certification with $\Delta_{U, U'}$ is that it does not require the use of expensive methods to estimate the operator's minimal singular value. However, the effectivenesses of the certification with $\Delta_{U, U'}$ can still be poor (for instance, for PDEs context with non Hilbert-Schmidt operators). Furthermore, both certifications with (24) and (25) require \mathbf{B} to have a relatively high minimal singular value, which can be in general hard to attain. These drawbacks can be (again) circumvented by restriction of the test and/or solution spaces as is discussed below.

Suppose that we are given an m -dimensional approximation subspace $U_m \subseteq U$, with $r \leq m \ll n$, that is known to well approximate the solution \mathbf{u} . Assume that $\mathbf{u}_r \in U_m$, and

$$\|\mathbf{u} - \Pi_{U_m} \mathbf{u}\|_U \leq \tau \|\mathbf{u} - \mathbf{u}_r\|_U, \quad (26)$$

for some $\tau < 1$. We require τ to be sufficiently small, say $\tau \leq 1/2$. Note that such a condition in general may not be efficiently certified in the online stage, and therefore it has to be guaranteed a priori. In practice, this can be done by considering U_m with high dimension, possibly much larger than r (but still much lower than n). Our methodology allows this since it does not involve any online computations depending on m . The subspace U_m can be taken as a span of snapshots (plus U_r). For PDEs, this space can also be chosen to be associated with an approximation on a coarse grid.

Clearly, when $\mathbf{B} \approx \mathbf{R}_U$, the (mapped) residual $\mathbf{R}_U^{-1} \mathbf{r}^*(\mathbf{u}_r)$ can serve as an estimator of $\mathbf{u} - \mathbf{u}_r$. We observe that for the error estimation we may consider instead of the full vector $\mathbf{R}_U^{-1} \mathbf{r}^*(\mathbf{u}_r)$ its projection onto U_m . This consideration leads to the following error estimator

$$\|\mathbf{u} - \mathbf{u}_r\|_U \approx \|\Pi_{U_m} \mathbf{R}_U^{-1} \mathbf{r}^*(\mathbf{u}_r)\|_U = \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'_m}.$$

Its effectivity can be characterized by the distance between operators \mathbf{B} and \mathbf{R}_U restricted to the space U_m , as shown in Proposition 3.8.

Proposition 3.8. *If $a := \|\mathbf{E}\|_{U_m, U'_m} + \tau \|\mathbf{E}\|_{U, U'_m} < \sqrt{1 - \tau^2}$, then*

$$\frac{1}{1 + a} \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'_m} \leq \|\mathbf{u} - \mathbf{u}_r\|_U \leq \frac{1}{\sqrt{1 - \tau^2} - a} \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'_m}. \quad (27)$$

Proof. See appendix. □

By choosing $r := m$ in Proposition 3.4 one may define the error indicators Δ_{U_m, U'_m} and Δ_{U, U'_m} and use them for bounding $\|\mathbf{E}\|_{U_m, U'_m}$ and $\|\mathbf{E}\|_{U, U'_m}$. Similarly to the Galerkin projection, here the effectiveness of $\|\mathbf{r}^*(\mathbf{u}_r)\|_{U'_m}$ as an error estimator can be attained by minimizing Δ_{U, U'_m} or $\Delta_{U_m, U'_m}^2 + \tau^2 \Delta_{U, U'_m}^2$.

A great advantage of certification of the error with Proposition 3.8 and Proposition 3.4 is that such a certification no longer requires \mathbf{B} to have a moderate minimal singular value as in (24) and (25). The only requirement is that the preconditioner is such that \mathbf{B} is close to \mathbf{R}_U with the solution space and/or the test space being restricted to the subspace U_m .

4 Measures of quality of a preconditioner for randomized model order reduction methods

In [3], the authors proposed a methodology, based on random sketching, to drastically improve the efficiency of classical model reduction methods. Here we introduce error indicators (or discrepancy measures) for estimating the quality of a preconditioner in this context.

4.1 Preliminaries on random sketching

Let us first present the basic ingredients of random sketching from [3]. In this work, the original inner products $\langle \cdot, \cdot \rangle_U$ and $\langle \cdot, \cdot \rangle_{U'}$ are estimated by

$$\langle \cdot, \cdot \rangle_U^\Theta := \langle \Theta \cdot, \Theta \cdot \rangle, \text{ and } \langle \cdot, \cdot \rangle_{U'}^\Theta := \langle \Theta \mathbf{R}_U^{-1} \cdot, \Theta \mathbf{R}_U^{-1} \cdot \rangle, \quad (28)$$

where $\Theta \in \mathbb{K}^{k \times n}$, with $k \ll n$, is seen as an operator from U to ℓ_2 , and is called a $U \rightarrow \ell_2$ subspace embedding. Let $\| \cdot \|_U^\Theta$ and $\| \cdot \|_{U'}^\Theta$ denote the associated semi-norms.

Here, Θ is taken as a realization of a random distribution of matrices satisfying the (ε, δ, d) oblivious $U \rightarrow \ell_2$ subspace embedding property (see below the definition).

Definition 4.1. Θ is called a (ε, δ, d) oblivious $U \rightarrow \ell_2$ subspace embedding if for any d -dimensional subspace V of U , the relation

$$\forall \mathbf{x}, \mathbf{y} \in V, \quad |\langle \mathbf{x}, \mathbf{y} \rangle_U - \langle \mathbf{x}, \mathbf{y} \rangle_U^\Theta| \leq \varepsilon \|\mathbf{x}\|_U \|\mathbf{y}\|_U \quad (29)$$

holds with probability at least $1 - \delta$.

Furthermore, Θ is chosen such that it can be efficiently multiplied by a vector.

A random matrix which is a (ε, δ, d) oblivious $U \rightarrow \ell_2$ subspace embedding, with $U = \mathbb{K}^n$ and $\langle \cdot, \cdot \rangle_U = \langle \cdot, \cdot \rangle_2$, is referred to as a (ε, δ, d) oblivious $\ell_2 \rightarrow \ell_2$ subspace embedding. Some distributions of matrices are known to be (ε, δ, d) oblivious $\ell_2 \rightarrow \ell_2$ subspace embeddings, such as the rescaled Gaussian matrices, the rescaled Rademacher matrices, the Subsampled Randomized Hadamard Transform (SRHT), the Subsampled Randomized Fourier Transform (SRFT) and others. In this work, from the oblivious $\ell_2 \rightarrow \ell_2$ subspace embeddings, we shall only use the rescaled Gaussian and the SRHT distributions. A $k \times n$ rescaled Gaussian matrix has i.i.d. entries with mean 0 and variance k^{-1} . Assuming that n is the power of 2, a $k \times n$ SRHT matrix is defined as $k^{-1/2}(\mathbf{R}\mathbf{H}_n\mathbf{D}) \in \mathbb{R}^{k \times n}$, where $\mathbf{R} \in \mathbb{R}^{k \times n}$ are the first k rows of a uniform random permutation of rows of the identity matrix, $\mathbf{H}_n \in \mathbb{R}^{n \times n}$ is a Walsh-Hadamard matrix and $\mathbf{D} \in \mathbb{R}^{n \times n}$ is a random diagonal matrix with random entries such that $\mathbb{P}([\mathbf{D}]_{i,i} = \pm 1) = 1/2$. The partial-SRHT (P-SRHT) is used when n is not necessarily a power of 2, and is defined as the first n columns of a SRHT matrix of size s , where s is a power of 2 and $n \leq s < 2n$.

From [3] it follows that the rescaled Gaussian distribution with

$$k \geq 7.87\varepsilon^{-2}(D6.9d + \log(1/\delta)), \quad (30a)$$

where $D = 1$ for $\mathbb{K} = \mathbb{R}$ or $D = 2$ for $\mathbb{K} = \mathbb{C}$, and the P-SRHT distribution with

$$k \geq 2(\varepsilon^2 - \varepsilon^3/3)^{-1} \left[\sqrt{d} + \sqrt{8 \log(6n/\delta)} \right]^2 \log(3d/\delta), \quad (30b)$$

respectively, are (ε, δ, d) oblivious $\ell_2 \rightarrow \ell_2$ subspace embeddings. The logarithmic dependence or no dependence of bounds (30) on n and δ explain the interest of random sketching methods. It is important to note that despite having higher theoretical bounds for the required number of rows, SRHT matrices typically provide similar accuracy of approximation as Gaussian matrices of the same size.

The $\ell_2 \rightarrow \ell_2$ embeddings can be used for the construction of $U \rightarrow \ell_2$ embeddings. Let \mathbf{Q} be a matrix such that $\mathbf{Q}^H \mathbf{Q} = \mathbf{R}_U$.² This matrix can be obtained with a Cholesky factorization or a more efficient approach proposed in [3, Remark 2.7] based on domain decomposition.

Proposition 4.2 (Proposition 3.11 in [3]). *Let Ω be a (ε, δ, d) oblivious $\ell_2 \rightarrow \ell_2$ subspace embedding. The random matrix*

$$\Theta := \Omega \mathbf{Q}$$

is a (ε, δ, d) oblivious $U \rightarrow \ell_2$ subspace embedding of subspaces of U .

For a subspace $W \subseteq U$, we let $\Pi_W^\Theta : U \rightarrow W$ denote the orthogonal projection on W with respect to $\|\cdot\|_U^\Theta$, such that

$$\forall \mathbf{x} \in U, \Pi_W^\Theta \mathbf{x} = \arg \min_{\mathbf{w} \in W} \|\mathbf{x} - \mathbf{w}\|_U^\Theta. \quad (31)$$

The semi-norm $\|\cdot\|_{W'}^\Theta$ can be estimated by [3]

$$\|\cdot\|_{W'}^\Theta := \|\Pi_W^\Theta \mathbf{R}_U^{-1} \cdot\|_U^\Theta = \max_{\mathbf{w} \in W \setminus \{\mathbf{0}\}} \frac{|\langle \mathbf{R}_U^{-1} \cdot, \mathbf{w} \rangle_U^\Theta|}{\|\mathbf{w}\|_U^\Theta}. \quad (32)$$

Finally, the maximal singular value, $\|\mathbf{C}\|_{V, W'}$, of an operator $\mathbf{C} : U \rightarrow U'$ used as an operator from V to W' , can be estimated by

$$\|\mathbf{C}\|_{V, W'}^\Theta := \max_{\mathbf{v} \in V \setminus \{\mathbf{0}\}} \frac{\|\mathbf{C}\mathbf{v}\|_{W'}^\Theta}{\|\mathbf{v}\|_U}. \quad (33)$$

4.2 Sketched Galerkin projection

Following [3], the sketched Galerkin orthogonality condition can be stated as

$$\|\mathbf{B}(\mathbf{u} - \mathbf{u}_r)\|_{U_r'}^\Theta = 0. \quad (34)$$

First of all, it follows directly from [3, Propositions 4.1 and 4.2] that $\mathbf{u}_r \in U_r$ satisfying (34) is an almost optimal minimizer of the error over U_r , if Θ is an $(\varepsilon, \delta, r + 1)$ oblivious embedding and \mathbf{B} is a well-conditioned operator. Therefore, one way to ensure a good quality of \mathbf{u}_r in (34) is to choose a preconditioner that minimizes the multi-purpose error indicator $\Delta_{U, U'}$.

A better way to measure the quality of the sketched Galerkin projection can be to use error indicators with solution and/or test spaces restricted to low-dimensional spaces, as is discussed below. Our analysis will be based on the assumption that Θ satisfies (29) with $V = U_r$. This condition can be satisfied with high probability by taking Θ as (ε, δ, r) oblivious $\ell_2 \rightarrow \ell_2$ subspace embedding, and serves as the basis for the randomized methods in [3].

We have the following result characterizing the quasi-optimality of \mathbf{u}_r in (34).

Proposition 4.3. *Let Θ satisfy (29) with $V = U_r$. Let \mathbf{u}_r satisfy (34). If $\|\mathbf{E}\|_{U_r, U_r'}^\Theta < \sqrt{1 - \varepsilon}$, then*

$$\|\mathbf{u} - \mathbf{u}_r\|_U \leq \left(1 + \frac{\|\mathbf{E}\|_{U, U_r'}^\Theta}{\sqrt{1 - \varepsilon} - \|\mathbf{E}\|_{U_r, U_r'}^\Theta} \right) \|\mathbf{u} - \Pi_{U_r}^\Theta \mathbf{u}\|_U. \quad (35)$$

²Matrix \mathbf{Q} (respectively \mathbf{Q}^H) is interpreted as a map from U to ℓ_2 (respectively from ℓ_2 to U').

Proof. See appendix. \square

Let $\mathcal{M} := \{\mathbf{u}(\mu) : \mu \in \mathcal{P}\}$ denote the solution manifold. Proposition 4.3 implies the quasi-optimality of the sketched Galerkin projection if the semi-norms $\|\mathbf{E}\|_{U, U_r}^\Theta$ and $\|\mathbf{E}\|_{U, U_r}^\Theta$ are small enough, with the condition that $\|\mathbf{u} - \Pi_{U_r}^\Theta \mathbf{u}\|_U$ is approximately equal to $\|\mathbf{u} - \Pi_{U_r} \mathbf{u}\|_U$. The latter condition can be ensured with high probability, if Θ is an oblivious subspace embedding of small size. For instance, by the first part of Proposition 4.4, the inequalities

$$\begin{aligned} \|\mathbf{u} - \Pi_{U_r} \mathbf{u}\|_U &\leq \|\mathbf{u} - \Pi_{U_r}^\Theta \mathbf{u}\|_U \leq \frac{1}{\sqrt{1-\varepsilon}} \|\mathbf{u} - \Pi_{U_r}^\Theta \mathbf{u}\|_U^\Theta \leq \frac{1}{\sqrt{1-\varepsilon}} \|\mathbf{u} - \Pi_{U_r} \mathbf{u}\|_U^\Theta \\ &\leq \sqrt{\frac{1+\varepsilon}{1-\varepsilon}} \|\mathbf{u} - \Pi_{U_r} \mathbf{u}\|_U, \end{aligned}$$

hold with probability at least $1 - \delta$ for a single $\mathbf{u} \in \mathcal{M}$, if Θ is a $(\varepsilon, \delta, r + 1)$ oblivious $U \rightarrow \ell_2$ subspace embedding. The guaranty for the whole solution manifold, if it is finite, can then be obtained using a union bound argument. On the other hand, such a guaranty for infinite \mathcal{M} requires knowing how well \mathcal{M} can be approximated with a low-dimensional space, which can be quantified with the Kolmogorov m -width

$$d_m(\mathcal{M}) := \min_{\dim(W)=m} \sup_{\mathbf{u} \in \mathcal{M}} \min_{\mathbf{w} \in W} \|\mathbf{u} - \mathbf{w}\|_U. \quad (36)$$

Proposition 4.4. *We have the following results.*

- If Θ is a $(\varepsilon, \delta, r + 1)$ oblivious $U \rightarrow \ell_2$ subspace embedding, then

$$\sqrt{1-\varepsilon} \|\mathbf{u} - \mathbf{w}\|_U \leq \|\mathbf{u} - \mathbf{w}\|_U^\Theta \leq \sqrt{1+\varepsilon} \|\mathbf{u} - \mathbf{w}\|_U$$

holds with probability at least $1 - \delta$ for all $\mathbf{w} \in U_r$ and a single $\mathbf{u} \in \mathcal{M}$.

- If Θ is a $(\varepsilon, \delta, r + m)$ and $(\varepsilon, \delta(\frac{n}{m} + 1)^{-1}, m)$ oblivious $U \rightarrow \ell_2$ subspace embedding, then

$$\sqrt{1-\varepsilon} \|\mathbf{u} - \mathbf{w}\|_U - D d_m(\mathcal{M}) \leq \|\mathbf{u} - \mathbf{w}\|_U^\Theta \leq \sqrt{1+\varepsilon} \|\mathbf{u} - \mathbf{w}\|_U + D d_m(\mathcal{M}),$$

with $D = \sqrt{1+\varepsilon}(\sqrt{n/m} + 1 + 1)$, hold with probability at least $1 - 2\delta$ for all $\mathbf{w} \in U_r$ and all $\mathbf{u} \in \mathcal{M}$, simultaneously.

Proof. See appendix. \square

It follows from the second part of Proposition 4.4 (and the bounds (30)) that if

$$d_m(\mathcal{M}) \leq \mathcal{O}(\sqrt{m/n}) \|\mathbf{u} - \Pi_{U_r} \mathbf{u}\|_U, \quad (37)$$

then we can use Θ with $k = \mathcal{O}(\varepsilon^{-2}m)$ rows to ensure that $\|\mathbf{u} - \Pi_{U_r}^\Theta \mathbf{u}\|_U = (1 + \mathcal{O}(\varepsilon)) \|\mathbf{u} - \Pi_{U_r} \mathbf{u}\|_U$. In its turn, the property (37) holds for $m = \mathcal{O}(r + \log n)$ if \mathcal{M} has a Kolmogorov m -width decaying exponentially, i.e., $d_m(\mathcal{M}) \leq E e^{-Fm}$, for some constants $E, F > 0$.

Let $\mathbf{U}_r : \mathbb{K}^r \rightarrow U$ be a matrix whose columns form a basis for U_r . Define the sketched matrices $\mathbf{U}_r^\Theta := \Theta \mathbf{U}_r$, $\mathbf{V}_r^\Theta := \Theta \mathbf{R}_U^{-1} \mathbf{B} \mathbf{U}_r$ as in [3], and $\mathbf{E}^\Theta := \Theta \mathbf{R}_U^{-1} \mathbf{E}$. In Proposition 4.5 we propose estimators for $\|\mathbf{E}\|_{U, U_r}^\Theta$ and $\|\mathbf{E}\|_{U, U_r}^\Theta$ with a guaranty of their effectivenesses.

Proposition 4.5. *Define*

$$\boxed{\Delta_{U_r, U_r'}^\Theta := \|\mathbf{U}_r^H \Theta^H \Theta \mathbf{R}_U^{-1} \mathbf{E} \mathbf{U}_r\|_F} = \|\mathbf{I} - (\mathbf{U}_r^\Theta)^H \mathbf{V}_r^\Theta\|_F \quad (38a)$$

and

$$\boxed{\Delta_{U, U'}^\Theta := \|\mathbf{E}^H \mathbf{R}_U^{-1} \Theta^H \Theta \mathbf{U}_r\|_{HS(\ell_2, U')}} = \|(\mathbf{E}^\Theta)^H \mathbf{U}_r^\Theta\|_{HS(\ell_2, U')}. \quad (38b)$$

Let Θ satisfy (29) with $V = U_r$. Assume that \mathbf{U}_r is orthonormal with respect to $\langle \cdot, \cdot \rangle_U^\Theta$. The following relations hold:

$$\sqrt{\frac{1-\varepsilon}{r}} \Delta_{U_r, U_r'}^\Theta \leq \|\mathbf{E}\|_{U_r, U_r'}^\Theta \leq \sqrt{1+\varepsilon} \Delta_{U_r, U_r'}^\Theta \quad \text{and} \quad \frac{1}{\sqrt{r}} \Delta_{U, U'}^\Theta \leq \|\mathbf{E}\|_{U, U'}^\Theta \leq \Delta_{U, U'}^\Theta. \quad (39)$$

Proof. See appendix. □

Remark 4.6. *The orthogonality of \mathbf{U}_r with respect to $\langle \cdot, \cdot \rangle_U^\Theta$ is equivalent to the orthogonality of \mathbf{U}_r^Θ with respect to $\langle \cdot, \cdot \rangle_2$. This condition can be fulfilled by finding (e.g., with QR factorization) a square matrix \mathbf{T} such that $\mathbf{U}_r^\Theta \mathbf{T}$ is ℓ_2 -orthonormal, and then replace the basis \mathbf{U}_r by $\mathbf{U}_r \mathbf{T}$. Note that the possibly expensive computation of $\mathbf{U}_r \mathbf{T}$ is not needed for the error indicators from Proposition 4.5. One only needs to compute the products $\mathbf{U}_r^\Theta \mathbf{T}$ and $\mathbf{V}_r^\Theta \mathbf{T}$.*

Further we provide a guaranty of almost preservation of the quasi-optimality constants of the classical Galerkin projection by the sketched Galerkin projection. This can be done under assumption that the operators $\mathbf{B}(\mu)$ belong to a low-dimensional space.

Proposition 4.7. *Let $Y \subseteq HS(U, U')$ be a low-dimensional subspace of matrices seen as operators from U to U' . Let $\Delta_{U_r, U_r'}$ and $\Delta_{U, U'}$ be the error indicators from (18) associated with \mathbf{U}_r that has orthonormal columns with respect to $\langle \cdot, \cdot \rangle_U$, and let $\Delta_{U, U'}$ be the multi-purpose indicator from (9). Furthermore, define*

$$\Delta_{U_r, U'} := \|\mathbf{E} \mathbf{U}_r\|_{HS(\ell_2, U')} \leq \Delta_{U, U'}.$$

- *If Θ is a $(\varepsilon, \delta/r, r + \dim Y + 1)$ oblivious $U \rightarrow \ell_2$ subspace embedding, then with probability at least $1 - \delta$,*

$$(\Delta_{U_r, U_r'}^\Theta)^2 \leq \frac{2}{1-\varepsilon} (\Delta_{U_r, U_r'}^2 + \varepsilon^2 \Delta_{U_r, U'}^2),$$

holds for all $\mathbf{B} \in Y$.

- *If Θ is a $(\varepsilon, \delta/n, r + \dim Y + 1)$ oblivious $U \rightarrow \ell_2$ subspace embedding, then with probability at least $1 - \delta$,*

$$(\Delta_{U, U'}^\Theta)^2 \leq \frac{2}{1-\varepsilon} (\Delta_{U, U'}^2 + \varepsilon^2 \Delta_{U, U'}^2),$$

holds for all $\mathbf{B} \in Y$.

Proof. See appendix. □

By combining the statements of Proposition 4.7 with the results of Propositions 3.4 and 4.5 we deduce an almost preservation of the quasi-optimality constants, if Θ is an oblivious embedding of sufficiently large size. Note that the upper bound for Δ_{U,U_r}^Θ is proportional to $\varepsilon\Delta_{U,U_r}$ (for large Δ_{U,U_r}) that can be as big as $n^{1/2}\varepsilon\|\mathbf{E}\|_{U,U_r}$ (e.g., for PDEs with non Hilbert-Schmidt operators). In such a case one may be forced to use a very small value for ε to have the preservation of the quasi-optimality constants. This implies that for some problems the classical preconditioned Galerkin projection characterized with $\Delta_{U_r,U_r'}$ and $\Delta_{U,U_r'}$ can be a more suitable choice than the sketched one.

4.3 Sketched error estimation

As seen in Section 3.3, the effectiveness of the (preconditioned) residual $\|\mathbf{r}^*(\mathbf{u}_r)\|_{U'}$ as an estimator of the error $\|\mathbf{u} - \mathbf{u}_r\|_U$ can be certified by using a multi-purpose indicator $\Delta_{U,U'}$. In its turn, $\|\mathbf{r}^*(\mathbf{u}_r)\|_{U'}$ can be efficiently estimated by $\|\mathbf{r}^*(\mathbf{u}_r)\|_{U'}^\Theta$. The accuracy of such an estimation can be guaranteed if Θ is an oblivious embedding of small size (for more details, see [3] taking $\mathbf{A} := \mathbf{B}$ and $\mathbf{b} := \mathbf{f}$).

The quality of an error estimation can be improved by a restriction of the solution and test spaces to a low-dimensional space. As in Section 3.3, we assume to be given a subspace U_m that contains \mathbf{u}_r and approximates well \mathbf{u} . It is assumed that

$$\begin{aligned} \|\mathbf{u} - \Pi_{U_m} \mathbf{u}\|_U &\leq \|\mathbf{u} - \Pi_{U_m}^\Theta \mathbf{u}\|_U \leq \tau^* \|\mathbf{u} - \mathbf{u}_r\|_U, \text{ and} \\ \|\mathbf{u} - \Pi_{U_m}^\Theta \mathbf{u}\|_U^\Theta &\leq \|\mathbf{u} - \Pi_{U_m} \mathbf{u}\|_U^\Theta \leq \tau^* \|\mathbf{u} - \mathbf{u}_r\|_U, \end{aligned}$$

for some $\tau^* < 1$. This condition can be ensured by using a sufficiently large space U_m and sufficiently large sketching matrix Θ (see Proposition 4.4 taking $r := m$). Then we propose the following error estimator:

$$\|\mathbf{u} - \mathbf{u}_r\|_U \approx \|\mathbf{r}^*(\mathbf{u}_r)\|_{U_m'}^\Theta.$$

The effectiveness of such an error estimator can be characterized by the following proposition under the condition that $\|\mathbf{u} - \mathbf{u}_r\|_U = (1 + \mathcal{O}(\varepsilon))\|\mathbf{u} - \mathbf{u}_r\|_U^\Theta$, which can be ensured by using Proposition 4.4.

Proposition 4.8. *Let Θ be such that*

$$\sqrt{1 - \varepsilon} \|\mathbf{u} - \mathbf{u}_r\|_U \leq \|\mathbf{u} - \mathbf{u}_r\|_U^\Theta \leq \sqrt{1 + \varepsilon} \|\mathbf{u} - \mathbf{u}_r\|_U.$$

If $a^ := \|\mathbf{E}\|_{U_m,U_m'}^\Theta + \tau^* \|\mathbf{E}\|_{U,U_m'}^\Theta < \sqrt{1 - \varepsilon - \tau^{*2}}$, then*

$$\frac{1}{\sqrt{1 + \varepsilon} + a^*} \|\mathbf{r}^*(\mathbf{u}_r)\|_{U_m'}^\Theta \leq \|\mathbf{u} - \mathbf{u}_r\|_U \leq \frac{1}{\sqrt{1 - \varepsilon - \tau^{*2}} - a^*} \|\mathbf{r}^*(\mathbf{u}_r)\|_{U_m'}^\Theta. \quad (40)$$

Proof. See appendix. □

The semi-norms $\|\mathbf{E}\|_{U_m,U_m'}^\Theta$ and $\|\mathbf{E}\|_{U,U_m'}^\Theta$ can be estimated by $\Delta_{U_m,U_m'}^\Theta$ and $\Delta_{U,U_m'}^\Theta$ defined in Proposition 4.5 with $r := m$. The minimization of $\Delta_{U,U_m'}^\Theta$ or a weighted sum $(\Delta_{U_m,U_m'}^\Theta)^2 + \tau^{*2}(\Delta_{U,U_m'}^\Theta)^2$ can ensure the effectiveness of $\|\mathbf{r}^*(\mathbf{u}_r)\|_{U_m'}^\Theta$ as an error estimator.

5 Random sketching for operators and construction of preconditioners

We consider the construction of a preconditioner of the form (3), i.e,

$$\mathbf{P}(\mu) = \sum_{i=1}^p \lambda_i(\mu) \mathbf{Y}_i, \quad (41)$$

with the coefficients $\lambda_i(\mu)$ obtained by minimizing one of the error indicators from Sections 3 and 4 (or their quadratic weighted average). For each parameter value the minimization of an error indicator consists in solving a least-squares problem. Its solution can be classically obtained with the normal equation, which is a $p \times p$ linear system of equations. For each parameter value the normal system can be formed from its affine decomposition precomputed in the offline stage, which allows to have the online complexity independent of n . This approach, however, can involve heavy, and in some cases even unfeasible, offline computations such as evaluations of Frobenius inner products of large, dense, possibly implicit matrices. Furthermore, this approach also can suffer from numerical instabilities and can require expensive online computations.

The aforementioned issues can be circumvented by extending the probabilistic methods from [3] to efficiently estimate inner products between operators from low-dimensional spaces. Then, by exploiting the fact that $\mathbf{P}(\mu)$ belongs to a low-dimensional space of operators, the error indicators from Sections 3 and 4, given as norms of certain matrices, can be estimated by ℓ_2 -norms of the images (so-called sketches) of the matrices through a carefully chosen random linear map to a low-dimensional ℓ_2 -space. Such random maps are here constructed using random sketching matrices that are $\ell_2 \rightarrow \ell_2$ oblivious subspace embeddings, as is discussed next.

5.1 Oblivious ℓ_2 -embeddings for matrices

Let X be a space of matrices equipped with an inner product $\langle \cdot, \cdot \rangle_X$. We will consider several cases: the space of matrices $X = HS(\ell_2, \ell_2)$ equipped with $\langle \cdot, \cdot \rangle_X = \langle \cdot, \cdot \rangle_F$, the space of matrices $X = HS(\ell_2, U)$ equipped with $\langle \cdot, \cdot \rangle_X = \langle \cdot, \cdot \rangle_{HS(\ell_2, U)} = \langle \mathbf{R}_U \cdot, \cdot \rangle_F$, or the space of matrices $X = HS(U', U)$ equipped with $\langle \cdot, \cdot \rangle_X = \langle \cdot, \cdot \rangle_{HS(U', U)} = \langle \mathbf{R}_U \cdot \mathbf{R}_U, \cdot \rangle_F$.

Definition 5.1. *A random linear map $\Upsilon(\cdot)$ from X to $\ell_2(\mathbb{K}^k)$ with $k \leq \dim(X)$, is called a (ε, δ, d) oblivious $X \rightarrow \ell_2$ subspace embedding if for any d -dimensional subspace V of X it holds*

$$\mathbb{P}(\forall \mathbf{X}, \mathbf{Y} \in V, |\langle \mathbf{X}, \mathbf{Y} \rangle_X - \langle \Upsilon(\mathbf{X}), \Upsilon(\mathbf{Y}) \rangle_2| \leq \varepsilon \|\mathbf{X}\|_X \|\mathbf{Y}\|_X) \geq 1 - \delta. \quad (42)$$

Next we propose ways to construct (ε, δ, d) oblivious $X \rightarrow \ell_2$ embeddings from the classical oblivious $\ell_2 \rightarrow \ell_2$ embeddings such as Gaussian matrices or P-SRHT. Let \mathbf{Q} be a $s \times n$ matrix such that $\mathbf{Q}^H \mathbf{Q} = \mathbf{R}_U$. The dimension s can be larger than n but is assumed to satisfy $s = \mathcal{O}(n)$. Note that \mathbf{Q} and \mathbf{Q}^H are seen as operators from U to ℓ_2 and from ℓ_2 to U' , respectively. Let us also define the operation $\text{vec}(\cdot)$ that reshapes (say, column-wise) a matrix to a vector, and defines a linear isometry between $HS(\ell_2, \ell_2)$ and ℓ_2 .

Proposition 5.2. *The random map $\Lambda(\cdot)$ defined by*

$$\Lambda(\mathbf{X}) := \Gamma \text{vec}(\Omega \mathbf{X} \Sigma^H), \quad \mathbf{X} : \mathbb{K}^r \rightarrow \mathbb{K}^r,$$

where Γ , Ω and Σ are $(\varepsilon_\Gamma, \delta_\Gamma, d)$, $(\varepsilon_\Omega, \delta_\Omega, d)$ and $(\varepsilon_\Sigma, \delta_\Sigma, d)$ oblivious $\ell_2 \rightarrow \ell_2$ subspace embeddings, is a (ε, δ, d) oblivious $HS(\ell_2, \ell_2) \rightarrow \ell_2$ subspace embedding with $\varepsilon = (1 + \varepsilon_\Gamma)(1 + \varepsilon_\Sigma)(1 + \varepsilon_\Omega) - 1$ and $\delta = \min(k_\Sigma \delta_\Omega + r \delta_\Sigma, k_\Omega \delta_\Sigma + r \delta_\Omega) + \delta_\Gamma$, where k_Ω and k_Σ are the numbers of rows of Ω and Σ , respectively.

Proof. See appendix. □

Proposition 5.3. *The random map $\Xi(\cdot)$ defined by*

$$\Xi(\mathbf{X}) := \Gamma \text{vec}(\Omega \mathbf{Q} \mathbf{X} \Sigma^H), \quad \mathbf{X} : \mathbb{K}^r \rightarrow U,$$

where Γ , Ω and Σ are $(\varepsilon_\Gamma, \delta_\Gamma, d)$, $(\varepsilon_\Omega, \delta_\Omega, d)$ and $(\varepsilon_\Sigma, \delta_\Sigma, d)$ oblivious $\ell_2 \rightarrow \ell_2$ subspace embeddings, is a (ε, δ, d) oblivious $HS(\ell_2, U) \rightarrow \ell_2$ subspace embedding of subspaces of matrices with r columns (representing r vectors in U) with $\varepsilon = (1 + \varepsilon_\Gamma)(1 + \varepsilon_\Sigma)(1 + \varepsilon_\Omega) - 1$ and $\delta = \min(k_\Sigma \delta_\Omega + s \delta_\Sigma, k_\Omega \delta_\Sigma + r \delta_\Omega) + \delta_\Gamma$, where k_Ω and k_Σ are the numbers of rows of Ω and Σ , respectively.

Proof. See appendix. □

Proposition 5.4. *The random map $\Psi(\cdot)$ defined by*

$$\Psi(\mathbf{X}) := \Gamma \text{vec}(\Omega \mathbf{Q} \mathbf{X} \mathbf{Q}^H \Sigma^H), \quad \mathbf{X} : U' \rightarrow U,$$

where Γ , Ω and Σ are $(\varepsilon_\Gamma, \delta_\Gamma, d)$, $(\varepsilon_\Omega, \delta_\Omega, d)$ and $(\varepsilon_\Sigma, \delta_\Sigma, d)$ oblivious $\ell_2 \rightarrow \ell_2$ subspace embeddings, is a (ε, δ, d) oblivious $HS(U', U) \rightarrow \ell_2$ subspace embedding of matrices representing operators from U' to U with $\varepsilon = (1 + \varepsilon_\Gamma)(1 + \varepsilon_\Sigma)(1 + \varepsilon_\Omega) - 1$ and $\delta = \min(k_\Sigma \delta_\Omega + s \delta_\Sigma, k_\Omega \delta_\Sigma + s \delta_\Omega) + \delta_\Gamma$, where k_Ω and k_Σ are the numbers of rows of Ω and Σ , respectively.

Proof. See appendix. □

It follows from (30) that to satisfy the (ε, δ, d) oblivious $X \rightarrow \ell_2$ subspace embedding property, the random maps $\Lambda(\cdot)$, $\Xi(\cdot)$ and $\Psi(\cdot)$ in Propositions 5.2 to 5.4 can be constructed with small Gaussian matrices Γ , Ω and Σ with $\mathcal{O}(\varepsilon^{-2}(d + \log(n) + \log(1/\delta)))$ rows. For P-SRHT this bound is larger by a factor of $\mathcal{O}(\log(n) + \log(1/\delta))$, but still remains sufficiently small. Moreover, despite having worse theoretical guarantees, in practice P-SRHT matrices provide similar results to Gaussian matrices of the same size.

Note that the usage of random embeddings Σ and Ω in Proposition 5.2 and Σ in Proposition 5.3 is not pertinent for d not so small compared to r (say, $d \geq r/4$). In such cases these matrices should be taken as the identity matrix (noting that the identity matrix is a $(0, 0, d)$ oblivious $\ell_2 \rightarrow \ell_2$ subspace embedding).

5.2 Randomized error indicators

We consider a situation where the preconditioned operators $\mathbf{B}(\mu)$ lie in a space Y of operators from U to U' and Y has a low dimension l . This is the case when the preconditioners are of the form (41). Then for a fixed parameter value, $\mathbf{B}(\mu)$ belongs to $Y = \text{span}\{\mathbf{R}_U \mathbf{Y}_i \mathbf{A}(\mu) : 1 \leq i \leq p\}$ with dimension $l \leq p$. Furthermore, if $\mathbf{A}(\mu)$ has an affine form $\mathbf{A}(\mu) = \sum_{j=1}^{m_A} \phi_j(\mu) \mathbf{A}_j$, then $\mathbf{B}(\mu)$ belong to $Y = \text{span}\{\mathbf{R}_U \mathbf{Y}_i \mathbf{A}_j : 1 \leq i \leq p, 1 \leq j \leq m_A\}$, with $l \leq pm_A$, for all $\mu \in \mathcal{P}$.

Multi-purpose, Galerkin projection and error estimation contexts. The error indicator $\Delta_{U,U'}(\mu)$ defined by (9) can be approximated by

$$\boxed{[\Delta_{U,U'}]^\Psi(\mu) := \|\Psi(\mathbf{R}_U^{-1} \mathbf{E}(\mu) \mathbf{R}_U^{-1})\|_2}, \quad (43)$$

where $\Psi(\cdot)$ is an oblivious $HS(U', U) \rightarrow \ell_2$ subspace embedding.

The following result is a direct consequence of Definition 5.1 and the fact that $\|\mathbf{E}\|_{HS(U,U')} = \|\mathbf{R}_U^{-1} \mathbf{E} \mathbf{R}_U^{-1}\|_{HS(U',U)}$.

Proposition 5.5. *If $\Psi(\cdot)$ is an $(\varepsilon, \delta, l+1)$ oblivious $HS(U', U) \rightarrow \ell_2$ subspace embedding, then*

$$\mathbb{P}(\forall \mathbf{B} \in Y, |\Delta_{U,U'}^2 - ([\Delta_{U,U'}]^\Psi)^2| \leq \varepsilon \Delta_{U,U'}^2) \geq 1 - \delta. \quad (44)$$

Observe that (44) implies with high probability the quasi-optimality of the minimizer of $[\Delta_{U,U'}]^\Psi$ over Y (or a subspace of Y) compared to a minimizer of $\Delta_{U,U'}$. A random map $\Psi(\cdot)$, which is a $(\varepsilon, \delta, l+1)$ oblivious $HS(U', U) \rightarrow \ell_2$ subspace embedding, can be constructed using Proposition 5.4 with Gaussian matrices or P-SRHT as $\ell_2 \rightarrow \ell_2$ subspace embeddings. The conditions (30) can be used for a priori selection of the sizes of random sketching matrices. In particular, it follows that in Proposition 5.4, the random matrices $\mathbf{\Gamma}$, $\mathbf{\Omega}$ and $\mathbf{\Sigma}$ can be chosen as rescaled Gaussian matrices or (in practice) SRHT with $\mathcal{O}(l + \log(n) + \log(1/\delta))$ rows.

The estimators for the error indicators $\Delta_{U_r, U_r'}(\mu)$ and $\Delta_{U, U_r'}(\mu)$ in (18) are given by

$$\boxed{[\Delta_{U_r, U_r'}]^\Lambda(\mu) := \|\Lambda(\mathbf{U}_r^H \mathbf{E}(\mu) \mathbf{U}_r)\|_2} \quad (45a)$$

and

$$\boxed{[\Delta_{U, U_r'}]^\Xi(\mu) := \|\Xi(\mathbf{R}_U^{-1} \mathbf{E}(\mu)^H \mathbf{U}_r)\|_2}, \quad (45b)$$

where $\Lambda(\cdot)$ is an oblivious $HS(\ell_2, \ell_2) \rightarrow \ell_2$ embedding, and $\Xi(\cdot)$ is an oblivious $HS(\ell_2, U) \rightarrow \ell_2$ embedding.

The following result is a direct consequence of Definition 5.1 and the fact that $\|\mathbf{E}^H \mathbf{U}_r\|_{HS(\ell_2, U')} = \|\mathbf{R}_U^{-1} \mathbf{E}^H \mathbf{U}_r\|_{HS(\ell_2, U)}$.

Proposition 5.6.

- *If $\Lambda(\cdot)$ is a $(\varepsilon, \delta, l+1)$ oblivious $HS(\ell_2, \ell_2) \rightarrow \ell_2$ subspace embedding, then*

$$\mathbb{P}(\forall \mathbf{B} \in Y, |\Delta_{U_r, U_r'}^2 - ([\Delta_{U_r, U_r'}]^\Lambda)^2| \leq \varepsilon \Delta_{U_r, U_r'}^2) \geq 1 - \delta. \quad (46)$$

- If $\Xi(\cdot)$ is a $(\varepsilon, \delta, l+1)$ oblivious $HS(\ell_2, U) \rightarrow \ell_2$ subspace embedding, then

$$\mathbb{P}(\forall \mathbf{B} \in Y, |\Delta_{U, U_r'}^2 - ([\Delta_{U, U_r'}]^\Xi)^2| \leq \varepsilon \Delta_{U, U_r'}^2) \geq 1 - \delta. \quad (47)$$

The random maps $\Lambda(\cdot)$ and $\Xi(\cdot)$ can be constructed with Propositions 5.2 and 5.3. For each indicator, the oblivious $\ell_2 \rightarrow \ell_2$ subspace embeddings Γ , Ω and Σ can be readily taken as Gaussian or P-SRHT matrices with $\mathcal{O}(l + \log(r) + \log(1/\delta))$ rows (according to (30)).

Relations (44), (46) and (47) imply the quasi-optimality of the minimizer of

$$[\Delta_{U, U_r'}]^\Psi, \text{ or } [\Delta_{U, U_r'}]^\Xi, \text{ or } ([\Delta_{U, U_r'}]^\Xi)^2 + \gamma([\Delta_{U_r, U_r'}]^\Lambda)^2$$

as a minimizer of, respectively,

$$\Delta_{U, U_r'}, \text{ or } \Delta_{U, U_r'}, \text{ or } \Delta_{U, U_r'}^2 + \gamma \Delta_{U_r, U_r'}^2$$

over Y (or a subspace of Y) with high probability.

The indicators $\Delta_{U_m, U_m'}(\mu)$ and $\Delta_{U, U_m'}(\mu)$ for the error estimation context in Section 3.3 can be estimated by $[\Delta_{U_m, U_m'}]^\Lambda(\mu)$ and $[\Delta_{U, U_m'}]^\Xi(\mu)$ defined by (45) choosing $r := m$. The accuracy of such an estimation can be characterized in exactly the same manner as the accuracy of estimation of the indicators $\Delta_{U_r, U_r'}$ and $\Delta_{U, U_r'}$ for the Galerkin projection.

Sketched Galerkin projection and error estimation. The error indicators $\Delta_{U_r, U_r'}^\Theta(\mu)$ and $\Delta_{U, U_r'}^\Theta(\mu)$ defined by (38) can be estimated by

$$\boxed{[\Delta_{U_r, U_r'}^\Theta]^\Lambda(\mu) := \|\Lambda(\mathbf{U}_r^H \Theta^H \Theta \mathbf{R}_U^{-1} \mathbf{E}(\mu) \mathbf{U}_r)\|_2} \quad (48a)$$

and

$$\boxed{[\Delta_{U, U_r'}^\Theta]^\Xi(\mu) := \|\Xi(\mathbf{R}_U^{-1} \mathbf{E}(\mu)^H \mathbf{R}_U^{-1} \Theta^H \Theta \mathbf{U}_r)\|_2}, \quad (48b)$$

where $\Lambda(\cdot)$ (respectively $\Xi(\cdot)$) is an oblivious $HS(\ell_2, \ell_2) \rightarrow \ell_2$ (respectively $HS(\ell_2, U) \rightarrow \ell_2$) subspace embedding.

Proposition 5.7.

- If $\Lambda(\cdot)$ is a $(\varepsilon, \delta, l+1)$ oblivious $HS(\ell_2, \ell_2) \rightarrow \ell_2$ subspace embedding, then

$$\mathbb{P}(\forall \mathbf{B} \in Y, |(\Delta_{U_r, U_r'}^\Theta)^2 - ([\Delta_{U_r, U_r'}^\Theta]^\Lambda)^2| \leq \varepsilon (\Delta_{U_r, U_r'}^\Theta)^2) \geq 1 - \delta. \quad (49)$$

- If $\Xi(\cdot)$ is a $(\varepsilon, \delta, l+1)$ oblivious $HS(\ell_2, U) \rightarrow \ell_2$ subspace embedding, then

$$\mathbb{P}(\forall \mathbf{B} \in Y, |(\Delta_{U, U_r'}^\Theta)^2 - ([\Delta_{U, U_r'}^\Theta]^\Xi)^2| \leq \varepsilon (\Delta_{U, U_r'}^\Theta)^2) \geq 1 - \delta. \quad (50)$$

Propositions 5.2 and 5.3 can be employed for the construction of $\Lambda(\cdot)$ and $\Xi(\cdot)$. A quasi-optimality of the preconditioner can be guaranteed with probability at least $1 - \delta$ by choosing random matrices Γ , Ω and Σ as Gaussian or P-SRHT matrices with $\mathcal{O}(l + \log(r) + \log(1/\delta))$ rows.

The error indicators $\Delta_{U_m, U_m'}^\Theta(\mu)$ and $\Delta_{U, U_m'}^\Theta(\mu)$ for the error estimation context can be efficiently estimated by $[\Delta_{U_m, U_m'}^\Theta]^\Lambda(\mu)$ and $[\Delta_{U, U_m'}^\Theta]^\Xi(\mu)$ defined in (48) choosing $r := m$.

Probability of success for all parameter values. There are two ways to guaranty a success of the sketched estimation of an error indicator (or a weighted average of error indicators) with probability at least $1 - \delta^*$ for all parameter values in \mathcal{P} , simultaneously. If \mathcal{P} is finite then one can simply consider success for each parameter value, separately, and then use a union bound argument, therefore using $\delta = \delta^*/\#\mathcal{P}$ and $l = p$ for the selection of sizes of random matrices. A second way is to exploit the fact that the set $\{\mathbf{B}(\mu) : \mu \in \mathcal{P}\}$ is a subset of some low-dimensional space. For instance, if $\mathbf{A}(\mu)$ has affine expansion with m_A terms and $\mathbf{P}(\mu)$ is of the form (3) then there exists such a space with dimension $l \leq pm_A$. This space can be readily chosen as the space Y in the above considerations.

Let us underline that these guarantees would only hold if $\{\mathbf{Y}_i : 1 \leq i \leq p\}$ is (statistically) independent of $\mathbf{\Gamma}$, $\mathbf{\Omega}$ and $\mathbf{\Sigma}$, which is not the case in the greedy algorithm for the selection of the parameter values for the interpolation of matrix inverse in Section 1.1. For the adaptive algorithms, one has to consider all possible outcomes with another union bound for the probability of success. In particular, if the training set has cardinality M , then there can exist up to $\binom{M}{p}$ possible outcomes of the greedy selection of p basis matrices and a success has to be guaranteed for each of them. In practice, this implies an increase of the probability of failure by a factor of $\binom{M}{p}$. Luckily the required sizes of random matrices depend only logarithmically on the probability of failure, therefore the replacement of δ by $\delta/\binom{M}{p}$ shall not dramatically affect the computational costs.

5.3 Practical aspects

Here, we summarize the main results of this section from the practical point of view.

The error indicators $\Delta_{U,U'}(\mu)$, $\Delta_{U_r,U'_r}(\mu)$, $\Delta_{U,U'_r}(\mu)$, $\Delta_{U_m,U'_m}(\mu)$, $\Delta_{U,U'_m}(\mu)$ introduced in Section 3 for the classical MOR methods, and the error indicators $\Delta_{U_r,U'_r}^\ominus(\mu)$, $\Delta_{U,U'_r}^\ominus(\mu)$, $\Delta_{U_m,U'_m}^\ominus(\mu)$, $\Delta_{U,U'_m}^\ominus(\mu)$ introduced in Section 4 for the randomized MOR methods from [3], can be respectively estimated by efficiently computable random estimators $[\Delta_{U,U'}]^\Psi(\mu)$, $[\Delta_{U_r,U'_r}]^\Lambda(\mu)$, $[\Delta_{U,U'_r}]^\Xi(\mu)$, $[\Delta_{U_m,U'_m}]^\Lambda(\mu)$, $[\Delta_{U,U'_m}]^\Xi(\mu)$, and $[\Delta_{U_r,U'_r}^\ominus]^\Lambda(\mu)$, $[\Delta_{U,U'_r}^\ominus]^\Xi(\mu)$, $[\Delta_{U_m,U'_m}^\ominus]^\Lambda(\mu)$, $[\Delta_{U,U'_m}^\ominus]^\Xi(\mu)$ which are the ℓ_2 -norms of the images of the corresponding residual matrices through random linear maps to low-dimensional Euclidean (or Hermitian) spaces. For $\mathbf{P}(\mu)$ defined in (41), the sketched error indicators from above (or their quadratic weighted average) can be written in the following form

$$\|\mathbf{W}_p(\mu)\mathbf{a}_p(\mu) - \mathbf{h}(\mu)\|_2, \quad (51)$$

where $[\mathbf{a}_p(\mu)]_i = \lambda_i(\mu)$, $1 \leq i \leq p$. The $k \times p$ matrix $\mathbf{W}_p(\mu) = [\mathbf{w}_1(\mu), \dots, \mathbf{w}_p(\mu)]$ and the vector $\mathbf{h}(\mu)$ represent the sketches of the corresponding large matrices. For instance, for the multi-purpose context: $\mathbf{w}_i(\mu) = \mathbf{\Psi}(\mathbf{Y}_i\mathbf{A}(\mu)\mathbf{R}_U^{-1})$, $1 \leq i \leq p$, and $\mathbf{h}(\mu) = \mathbf{\Psi}(\mathbf{R}_U^{-1})$. The accuracy of estimation of an error indicator can be guaranteed with high probability, for all $\mu \in \mathcal{P}$, if $k = \mathcal{O}(p + \log(\#\mathcal{P}))$ or $k = \mathcal{O}(m_{Ap})$, where m_A is the number of terms in the affine decomposition of $\mathbf{A}(\mu)$. The minimization of (51) over $\mathbf{a}_p(\mu)$ can be efficiently and numerically stably performed online for each parameter value with a standard routine such as QR factorization or SVD. For this, the affine decompositions of $\mathbf{W}_p(\mu)$ and $\mathbf{h}(\mu)$ have to be precomputed in the offline stage and then used for the efficient solution of (51) for each μ , with a cost independent (or weakly dependent) of the full dimension n , and dimensions m and r . The affine decompositions can be obtained from (given) affine decomposition of $\mathbf{A}(\mu)$ or with empirical interpolation method.

The computational cost of the offline stage is dominated by two operations: the products of \mathbf{Y}_i (and \mathbf{R}_U^{-1} , \mathbf{Q}) with multiple vectors and the products with random matrices $\mathbf{\Gamma}$, $\mathbf{\Omega}$ and $\mathbf{\Sigma}$.

With a good choice of random matrices, the offline computational cost associated with multiplications of these matrices by (explicit) matrices and vectors should have only a minor impact on the overall cost. Indeed, due to their specific structure, SRHT matrices of dimension $k \times n$ can be multiplied by a vector with only $2n \log_2(k+1)$ flops, while Gaussian matrices are very efficient in terms of scalability of computations for parallel architectures. Moreover, the random matrices can be generated, maintained or transferred (for the computations on multiple computational devices) with a negligible computational cost by using a seeded random number generator. For more details please see [3, 7, 16].

Maintaining the basis matrices \mathbf{Y}_i in explicit form can be intractable. In general, one should maintain and operate with \mathbf{Y}_i (and \mathbf{R}_U^{-1}) in an efficient implicit form, e.g, obtained with LU factorization, which can be precomputed once and then used for computing products of \mathbf{Y}_i (and \mathbf{R}_U^{-1}) with multiple vectors. For the architectures that do not allow precomputation of (approximate) factorizations of $\mathbf{A}(\mu^i)$, the cost of multiplication of \mathbf{Y}_i by a vector shall be comparable to the cost of solving the full system for a single parameter value. Next we provide an analysis of the computational cost of precomputation of an affine decomposition of $\mathbf{W}_p(\mu)$ associated with the multipurpose error indicator. Notice that, if $\mathbf{A}(\mu)$ has an affine form $\sum_{j=1}^{m_A} \phi_j(\mu) \mathbf{A}_j$, then the columns $\mathbf{w}_i(\mu) = \mathbf{\Psi}(\mathbf{Y}_i \mathbf{A}(\mu) \mathbf{R}_U^{-1}) = \mathbf{\Gamma} \text{vec}(\mathbf{\Omega} \mathbf{Q} \mathbf{Y}_i \mathbf{A}(\mu) \mathbf{R}_U^{-1} \mathbf{Q}^H \mathbf{\Sigma}^H)$ of $\mathbf{W}_p(\mu)$ can be expressed as $\mathbf{w}_i(\mu) = \sum_{j=1}^{m_A} \phi_j(\mu) [\sum_{e=1}^k \mathbf{w}_{i,j}^{(e)}]$, where $\mathbf{w}_{i,j}^{(e)} = \mathbf{\Gamma} \mathbf{R}_e \text{vec}(\omega_e \mathbf{Q} \mathbf{Y}_i \mathbf{A}_j \mathbf{R}_U^{-1} \mathbf{Q}^H \mathbf{\Sigma}^H)$, with ω_e denoting the e -th row of $\mathbf{\Omega}$, and \mathbf{R}_e denoting the extension operator, which is a sparse boolean matrix, mapping the local vector to the global one. If $\mathbf{\Omega}$ is a Gaussian matrix with k rows, if $\mathbf{\Sigma}$, $\mathbf{\Gamma}$ are SRHT matrices with k rows, if \mathbf{Q} , \mathbf{R}_U^{-1} , \mathbf{Y}_i can be multiplied by a vector using $\mathcal{O}(nk_P)$ flops and bytes of memory, and if \mathbf{A}_j are sparse matrices, then each product $\omega_e \mathbf{Q} \mathbf{Y}_i \mathbf{A}_j \mathbf{R}_U^{-1} \mathbf{Q}^H \mathbf{\Sigma}^H$ can be computed from left to right using $\mathcal{O}(n(\log(k) + k_P))$ flops and $\mathcal{O}(nk_P)$ bytes of memory. This implies that the total computational cost for the precomputation of $\mathbf{W}_p(\mu)$ is $\mathcal{O}(nk_P m_A (\log(k) + k_P))$ flops and $\mathcal{O}(nk_P)$ bytes of memory. Moreover, if \mathbf{R}_U and \mathbf{Q} are identity matrices, then the complexity can be reduced to $\mathcal{O}(nk_P (m_A \log(k) + k_P))$ flops. The cost of precomputation of $\mathbf{W}_p(\mu)$ associated with the other error indicators can be analyzed similarly.

For classical projection-based methods with preconditioning, the precomputation of the reduced system of equations associated with the Galerking projection, and the quantities needed for the error estimation, can require maintenance and operation with \mathbf{Y}_i (in an implicit form) that can be too expensive in some architectures. This can be avoided for the sketched Galerkin projection and error estimation contexts thanks to the methodology from [3]. As was indicated in [3], a reduced model can be accurately (with high probability) approximated from small random sketches of the approximation space and the associated residuals, represented by sketched matrices $\mathbf{U}_r^\Theta := \mathbf{\Theta} \mathbf{U}_r$, $\mathbf{V}_r^\Theta(\mu) := \mathbf{\Theta} \mathbf{R}_U^{-1} \mathbf{B}(\mu) \mathbf{U}_r$ and a vector $\mathbf{f}^\Theta(\mu) := \mathbf{\Theta} \mathbf{R}_U^{-1} \mathbf{f}(\mu)$. In their turn, the sketched matrices $\mathbf{V}_r^\Theta(\mu)$ and $\mathbf{f}^\Theta(\mu)$ can be obtained from $\mathbf{V}_r^{\Theta(i)}(\mu) = \mathbf{\Theta} \mathbf{Y}_i \mathbf{A}(\mu) \mathbf{U}_r$, $\mathbf{f}^{\Theta(i)}(\mu) = \mathbf{\Theta} \mathbf{Y}_i \mathbf{b}(\mu)$ and $\mathbf{a}_p(\mu)$ (obtained from $\mathbf{W}_p(\mu)$). Firstly, we see that, in the offline stage, rather than maintaining and operating with a basis matrix \mathbf{Y}_i , we can precompute its random sketch $\mathbf{\Theta} \mathbf{Y}_i$ (along with $\mathbf{w}_i(\mu)$) and operate with the sketch, which can be far more efficient in some architectures. Furthermore, for each $1 \leq i \leq p$ the precomputation of the affine decompositions of $\mathbf{V}_r^{\Theta(i)}(\mu)$, $\mathbf{f}^{\Theta(i)}(\mu)$ and $\mathbf{w}_i(\mu)$ requires operations only with \mathbf{Y}_i and no other basis matrices, which implies efficiency in terms of storage and distribution of computations.

Similarly as in [4], the online efficiency of the minimization of (51) for a finite test set $\mathcal{P}_{\text{test}}$ of parameter values can be improved by using an extra oblivious $\ell_2 \rightarrow \ell_2$ subspace embedding Φ statistically independent of $\mathcal{P}_{\text{test}}$, $\mathbf{W}_p(\mu)$ and $\mathbf{h}(\mu)$. The minimizer $\mathbf{a}_p(\mu)$ of (51) can be approximated by the minimizer of

$$\|\Phi \mathbf{W}_p(\mu) \mathbf{a}_p(\mu) - \Phi \mathbf{h}(\mu)\|_2. \quad (52)$$

If Φ is an $(\varepsilon, \delta(\#\mathcal{P}_{\text{test}})^{-1}, p+1)$ oblivious $\ell_2 \rightarrow \ell_2$ subspace embedding, then the minimizer $\mathbf{a}_p(\mu)$ of (52) is close to optimal with probability at least $1 - \delta$. The Gaussian matrices and P-SRHT (in practice) satisfy this property if they have $\mathcal{O}(p + \log(\#\mathcal{P}_{\text{test}}) + \log(1/\delta))$ rows. The number of rows of Φ should be in several times smaller than the size of $\mathbf{W}_p(\mu)$ required to guarantee the accuracy of the sketched error indicators for the whole parameter set \mathcal{P} or for adaptively chosen parameters in the algorithms for the construction of the preconditioner's basis. In the online stage, $\Phi \mathbf{W}_p(\mu)$ and $\Phi \mathbf{h}(\mu)$ can be evaluated from their affine decompositions, which can be efficiently precomputed beforehand (in the intermediate online stage) by applying the map Φ to the terms in affine decompositions of $\mathbf{W}_p(\mu)$ and $\mathbf{h}(\mu)$.

6 Conclusions and future work

In this work we developed a methodology for the construction of parameter-dependent preconditioners for large parameter-dependent matrices. A particular focus has been given to preconditioning techniques for the effective solution of ill-conditioned parametric systems of equations with projection-based model order reduction methods. We considered preconditioners with an affine form, with the parameter-dependent coefficients computed online by minimizing a discrepancy between the preconditioned matrix and the identity matrix measured with a semi-norm. Such minimization is a small least-squares problem, which can be efficiently solved online unlike the estimation of the classical stability constants. We provided several choices for measuring the discrepancy (or semi-norm) depending on the particular objective such as improving the quality of the Petrov-Galerkin projection or residual-based error estimation. The relations between the discrepancy measures and the stability constants have been derived, which directly yield the certification of the quality of the preconditioner in each of the considered contexts.

Besides the classical projection-based MOR methods, we also discussed preconditioning for randomized methods from [3]. Similarly to [3], the resulting methods do not require operations with large vectors and matrices, but only with their small random sketches, which implies offline efficiency with any computational architecture.

The heavy computational cost of construction of the preconditioner is bypassed by using randomized linear algebra. Instead of minimizing the Hilbert-Schmidt norm of a large matrix defining the discrepancy between the preconditioned operator and the identity, we minimize the ℓ_2 -norm of its image (called a sketch) to a low-dimensional ℓ_2 -space through a random linear map. The random map is chosen in such a way that the norm of the sketch estimates well the norm of the original matrix with high-probability. We described a procedure for the construction of the random maps and provided conditions on the dimensions of the sketching matrices required to attain the given accuracy of estimation with a user-specified probability of success.

The experimental validation of the methodology will be provided in the nearest future.

References

- [1] R. Abgrall, D. Amsallem, and R. Crisovan. “Robust model reduction by L_1 -norm minimization and approximation via dictionaries: application to nonlinear hyperbolic problems”. *Advanced Modeling and Simulation in Engineering Sciences* 3.1 (2016), pp. 1–16 (cit. on p. 2).
- [2] A. Alla and J. N. Kutz. “Randomized model order reduction”. *Advances in Computational Mathematics* 45.3 (2019), pp. 1251–1271 (cit. on p. 3).
- [3] O. Balabanov and A. Nouy. “Randomized linear algebra for model reduction. Part I: Galerkin methods and error estimation”. *Advances in Computational Mathematics* 45.5-6 (2019), pp. 2969–3019 (cit. on pp. 2, 3, 4, 5, 6, 8, 12, 13, 14, 15, 17, 18, 22, 23, 24, 31, 32).
- [4] O. Balabanov and A. Nouy. “Randomized linear algebra for model reduction. Part II: minimal residual methods and dictionary-based approximation”. *Advances in Computational Mathematics* 47.26 (2021) (cit. on pp. 3, 24).
- [5] P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, eds. *Model Reduction and Approximation: Theory and Algorithms*. SIAM, Philadelphia, PA, 2017 (cit. on p. 2).
- [6] P. Benner, S. Gugercin, and K. Willcox. “A survey of projection-based model reduction methods for parametric dynamical systems”. *SIAM review* 57.4 (2015), pp. 483–531 (cit. on p. 2).
- [7] C. Boutsidis and A. Gittens. “Improved matrix algorithms via the subsampled randomized Hadamard transform”. *SIAM Journal on Matrix Analysis and Applications* 34.3 (2013), pp. 1301–1340 (cit. on p. 23).
- [8] A. Buhr and K. Smetana. “Randomized local model order reduction”. *SIAM journal on scientific computing* 40.4 (2018), A2120–A2151 (cit. on p. 3).
- [9] Y. Cao and L. Petzold. “A posteriori error estimation and global error control for ordinary differential equations by the adjoint method”. *SIAM Journal on Scientific Computing* 26.2 (2004), pp. 359–374 (cit. on pp. 2, 3).
- [10] F. Casenave, A. Ern, and T. Lelièvre. “Accurate and online-efficient evaluation of the a posteriori error bound in the reduced basis method”. *ESAIM: Mathematical Modelling and Numerical Analysis* 48.1 (2014), pp. 207–229 (cit. on p. 2).
- [11] Y. Chen, S. Gottlieb, and Y. Maday. “Parametric analytical preconditioning and its applications to the reduced collocation methods”. *Comptes Rendus Mathématique* 352.7-8 (2014), pp. 661–666 (cit. on p. 2).
- [12] Y. Chen, J. S. Hesthaven, Y. Maday, and J. Rodríguez. “Improved successive constraint method based a posteriori error estimate for reduced basis approximation of 2D Maxwell’s problem”. *ESAIM: Mathematical Modelling and Numerical Analysis* 43.6 (2009), pp. 1099–1116 (cit. on p. 11).
- [13] C. Desceliers, R. Ghanem, and C. Soize. “Polynomial chaos representation of a stochastic preconditioner”. *International journal for numerical methods in engineering* 64.5 (2005), pp. 618–634 (cit. on p. 2).

- [14] M. Drohmann and K. Carlberg. “The ROMES method for statistical modeling of reduced-order-model error”. *SIAM/ASA Journal on Uncertainty Quantification* 3.1 (2015), pp. 116–145 (cit. on p. 3).
- [15] S. Hain, M. Ohlberger, M. Radic, and K. Urban. “A hierarchical a posteriori error estimator for the Reduced Basis Method”. *Advances in Computational Mathematics* (2019), pp. 1–24 (cit. on pp. 2, 5).
- [16] N. Halko, P.-G. Martinsson, and J. A. Tropp. “Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions”. *SIAM review* 53.2 (2011), pp. 217–288 (cit. on pp. 3, 23).
- [17] J. S. Hesthaven, G. Rozza, and B. Stamm. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. Springer Briefs in Mathematics. Switzerland: Springer, 2015, p. 135. ISBN: 978-3-319-22469-5 (cit. on p. 2).
- [18] C. Homescu, L. R. Petzold, and R. Serban. “Error estimation for reduced-order models of dynamical systems”. *SIAM Journal on Numerical Analysis* 43.4 (2005), pp. 1693–1714 (cit. on p. 3).
- [19] D. Huynh, D. Knezevic, Y. Chen, J. S. Hesthaven, and A. Patera. “A natural-norm successive constraint method for inf-sup lower bounds”. *Computer Methods in Applied Mechanics and Engineering* 199.29-32 (2010), pp. 1963–1975 (cit. on p. 11).
- [20] D. B. P. Huynh, G. Rozza, S. Sen, and A. T. Patera. “A successive constraint linear optimization method for lower bounds of parametric coercivity and inf-sup stability constants”. *Comptes Rendus Mathematique* 345.8 (2007), pp. 473–478 (cit. on p. 11).
- [21] D. Kressner, M. Plešinger, and C. Tobler. “A preconditioned low-rank CG method for parameter-dependent Lyapunov matrix equations”. *Numerical Linear Algebra with Applications* 21.5 (2014), pp. 666–684 (cit. on p. 2).
- [22] M. L. Parks, E. De Sturler, G. Mackey, D. D. Johnson, and S. Maiti. “Recycling Krylov subspaces for sequences of linear systems”. *SIAM Journal on Scientific Computing* 28.5 (2006), pp. 1651–1674 (cit. on p. 2).
- [23] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced basis methods for partial differential equations: an introduction*. Vol. 92. Springer, 2015 (cit. on p. 2).
- [24] N. D. Santo, S. Deparis, A. Manzoni, and A. Quarteroni. “Multi space reduced basis preconditioners for large-scale parametrized PDEs”. *SIAM Journal on Scientific Computing* 40.2 (2018), A954–A983 (cit. on p. 2).
- [25] K. Smetana and O. Zahm. “Randomized residual-based error estimators for the proper generalized decomposition approximation of parametrized problems”. *International Journal for Numerical Methods in Engineering* 121.23 (2020), pp. 5153–5177 (cit. on p. 3).
- [26] K. Smetana, O. Zahm, and A. T. Patera. “Randomized residual-based error estimators for parametrized equations”. *SIAM journal on scientific computing* 41.2 (2019), A900–A926 (cit. on pp. 3, 4, 5).

- [27] T. Taddei. “An offline/online procedure for dual norm calculations of parameterized functionals: empirical quadrature and empirical test spaces”. *Advances in Computational Mathematics* 45.5-6 (2019), pp. 2429–2462 (cit. on p. 2).
- [28] D. Torlo, F. Ballarin, and G. Rozza. “Stabilized weighted reduced basis methods for parametrized advection dominated problems with random inputs”. *SIAM/ASA Journal on Uncertainty Quantification* 6.4 (2018), pp. 1475–1502 (cit. on p. 2).
- [29] S. Trehan, K. T. Carlberg, and L. J. Durlofsky. “Error modeling for surrogates of dynamical systems using machine learning”. *International Journal for Numerical Methods in Engineering* 112.12 (2017), pp. 1801–1827 (cit. on p. 3).
- [30] J. A. Tropp and A. C. Gilbert. “Signal recovery from random measurements via orthogonal matching pursuit”. *IEEE Transactions on information theory* 53.12 (2007), pp. 4655–4666 (cit. on p. 3).
- [31] O. Zahm and A. Nouy. “Interpolation of inverse operators for preconditioning parameter-dependent equations”. *SIAM Journal on Scientific Computing* 38.2 (2016), A1044–A1074 (cit. on pp. 2, 3, 4).

7 Appendix

This section provides the proofs of propositions from the article.

Proof of Proposition 3.1. We have,

$$\|\mathbf{C}\mathbf{v}\|_{U'} = \|(\mathbf{R}_U^{-1/2}\mathbf{C}\mathbf{R}_U^{-1/2})(\mathbf{R}_U^{1/2}\mathbf{v})\|_2 \leq \|\mathbf{R}_U^{-1/2}\mathbf{C}\mathbf{R}_U^{-1/2}\|_F \|\mathbf{R}_U^{1/2}\mathbf{v}\|_2 = \|\mathbf{C}\|_{HS(U,U')} \|\mathbf{v}\|_U,$$

which ends the proof. □

Proof of Lemma 3.2. We have,

$$\begin{aligned} \|\mathbf{u}_r - \Pi_{U_r}\mathbf{u}\|_U &= \|\Pi_{U_r}(\mathbf{u}_r - \Pi_{U_r}\mathbf{u})\|_U \\ &\leq \|\Pi_{U_r}\mathbf{R}_U^{-1}\mathbf{B}(\mathbf{u}_r - \Pi_{U_r}\mathbf{u})\|_U + \|\Pi_{U_r}\mathbf{R}_U^{-1}\mathbf{E}(\mathbf{u}_r - \Pi_{U_r}\mathbf{u})\|_U \\ &= \|\mathbf{B}(\mathbf{u}_r - \Pi_{U_r}\mathbf{u})\|_{U'_r} + \|\mathbf{E}(\mathbf{u}_r - \Pi_{U_r}\mathbf{u})\|_{U'_r} \\ &\leq \|\mathbf{B}(\mathbf{u} - \Pi_{U_r}\mathbf{u})\|_{U'_r} + \|\mathbf{E}(\mathbf{u}_r - \Pi_{U_r}\mathbf{u})\|_{U'_r} \\ &\leq \|\mathbf{E}(\mathbf{u} - \Pi_{U_r}\mathbf{u})\|_{U'_r} + \|\mathbf{E}(\mathbf{u}_r - \Pi_{U_r}\mathbf{u})\|_{U'_r}, \end{aligned}$$

which completes the proof. □

Proof of Proposition 3.3. The relation (15) and the uniqueness of \mathbf{u}_r directly follow from Lemma 3.2 and the definitions of semi-norms $\|\mathbf{E}\|_{U_r,U'_r}$ and $\|\mathbf{E}\|_{U,U'_r}$. More specifically, Lemma 3.2 implies that

$$\begin{aligned} \|\mathbf{u}_r - \Pi_{U_r}\mathbf{u}\|_U &\leq \|\mathbf{E}(\mathbf{u} - \Pi_{U_r}\mathbf{u})\|_{U'_r} + \|\mathbf{E}(\mathbf{u}_r - \Pi_{U_r}\mathbf{u})\|_{U'_r} \\ &\leq \|\mathbf{E}\|_{U,U'_r} \|\mathbf{u} - \Pi_{U_r}\mathbf{u}\|_U + \|\mathbf{E}\|_{U_r,U'_r} \|\mathbf{u}_r - \Pi_{U_r}\mathbf{u}\|_U, \end{aligned}$$

which combined with the inequality

$$\|\mathbf{u}_r - \mathbf{\Pi}_{U_r} \mathbf{u}\|_U \geq \|\mathbf{u} - \mathbf{u}_r\|_U - \|\mathbf{u} - \mathbf{\Pi}_{U_r} \mathbf{u}\|_U$$

yields (15). The uniqueness of \mathbf{u}_r classically follows from the argument that if $\mathbf{u}_r \in U_r$ and $\mathbf{v}_r \in U_r$ satisfy the Galerkin orthogonality condition, then

$$\begin{aligned} 0 &= \|\mathbf{B}(\mathbf{u} - \mathbf{u}_r)\|_{U'_r} + \|\mathbf{B}(\mathbf{u} - \mathbf{v}_r)\|_{U'_r} \geq \|\mathbf{B}(\mathbf{v}_r - \mathbf{u}_r)\|_{U'_r} \\ &\geq \|\mathbf{R}_U(\mathbf{v}_r - \mathbf{u}_r)\|_{U'_r} - \|(\mathbf{R}_U - \mathbf{B})(\mathbf{v}_r - \mathbf{u}_r)\|_{U'_r} \\ &\geq (1 - \|\mathbf{E}\|_{U_r, U'_r}) \|\mathbf{v}_r - \mathbf{u}_r\|_U, \end{aligned}$$

which implies that $\mathbf{v}_r = \mathbf{u}_r$. □

Proof of Proposition 3.4. Let \mathbf{T} be a square matrix such that $\mathbf{U}_r^* := \mathbf{U}_r \mathbf{T}$ has orthonormal columns with respect to $\langle \cdot, \cdot \rangle_U$. The identity $\mathbf{\Pi}_{U_r} = \mathbf{U}_r^* \mathbf{U}_r^{*H} \mathbf{R}_U$ implies that

$$\|\cdot\|_{U'_r} = \|\mathbf{\Pi}_{U_r} \mathbf{R}_U^{-1} \cdot\|_U = \|\mathbf{U}_r^* \mathbf{U}_r^{*H} \cdot\|_U = \|\mathbf{U}_r^{*H} \cdot\|_2 = \|\mathbf{T}^H \mathbf{U}_r^H \cdot\|_2.$$

From this fact we obtain the following expressions for $\|\mathbf{E}\|_{U_r, U'_r}$ and $\|\mathbf{E}\|_{U, U'_r}$:

$$\|\mathbf{E}\|_{U_r, U'_r} = \max_{\mathbf{v} \in U_r \setminus \{\mathbf{0}\}} \frac{\|\mathbf{T}^H \mathbf{U}_r^H \mathbf{E} \mathbf{v}\|_2}{\|\mathbf{v}\|_U} = \|\mathbf{T}^H \mathbf{U}_r^H \mathbf{E} \mathbf{U}_r \mathbf{T}\|_2 \quad (53a)$$

and

$$\|\mathbf{E}\|_{U, U'_r} := \max_{\mathbf{v} \in U \setminus \{\mathbf{0}\}} \frac{\|\mathbf{T}^H \mathbf{U}_r^H \mathbf{E} \mathbf{v}\|_2}{\|\mathbf{v}\|_U} = \|\mathbf{T}^H \mathbf{U}_r^H \mathbf{E} \mathbf{R}_U^{-1/2}\|_2. \quad (53b)$$

It can be shown that the minimal and the maximal singular values of \mathbf{T} are equal to σ_1^{-1} and σ_r^{-1} , respectively. Then for any matrix \mathbf{X} and matrix $\mathbf{X}^* = \mathbf{X} \mathbf{T}$ or $\mathbf{T}^H \mathbf{X}$, it holds

$$\sigma_1^{-1} \|\mathbf{X}\|_2 \leq \|\mathbf{X}^*\|_2 \leq \sigma_r^{-1} \|\mathbf{X}\|_2. \quad (54)$$

By choosing in (54), first $\mathbf{X} = \mathbf{T}^H \mathbf{U}_r^H \mathbf{E} \mathbf{U}_r$ with $\mathbf{X}^* = \mathbf{X} \mathbf{T}$, and then $\mathbf{X} = \mathbf{U}_r^H \mathbf{E} \mathbf{U}_r$ with $\mathbf{X}^* = \mathbf{T}^H \mathbf{X}$, and using (53a) we obtain

$$\sigma_1^{-2} \|\mathbf{U}_r^H \mathbf{E} \mathbf{U}_r\|_2 \leq \|\mathbf{E}\|_{U_r, U'_r} \leq \sigma_r^{-2} \|\mathbf{U}_r^H \mathbf{E} \mathbf{U}_r\|_2.$$

At the same time, by choosing $\mathbf{X} = \mathbf{U}_r^H \mathbf{E} \mathbf{R}_U^{-1/2}$ with $\mathbf{X}^* = \mathbf{T}^H \mathbf{X}$ in (54), and using (53b) we get

$$\sigma_1^{-1} \|\mathbf{U}_r^H \mathbf{E} \mathbf{R}_U^{-1/2}\|_2 \leq \|\mathbf{E}\|_{U, U'_r} \leq \sigma_r^{-1} \|\mathbf{U}_r^H \mathbf{E} \mathbf{R}_U^{-1/2}\|_2.$$

These two relations combined with the fact that for a matrix \mathbf{X} with r rows,

$$r^{-1/2} \|\mathbf{X}\|_F \leq \|\mathbf{X}\|_2 \leq \|\mathbf{X}\|_F,$$

result in (19). □

Proof of Proposition 3.7. By definition of $\alpha(\mathbf{B})$ and $\beta(\mathbf{B})$, we have

$$\beta(\mathbf{B})^{-1} \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'} \leq \|\mathbf{u} - \mathbf{u}_r\| \leq \alpha(\mathbf{B})^{-1} \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'}. \quad (55)$$

Moreover, since $\Delta_{U,U'}$ is an upper bound of $\beta(\mathbf{E})$, the following inequalities hold

$$1 - \Delta_{U,U'} \leq 1 - \beta(\mathbf{E}) \leq \alpha(\mathbf{B}) \leq \beta(\mathbf{B}) \leq 1 + \beta(\mathbf{E}) \leq 1 + \Delta_{U,U'}. \quad (56)$$

The result of the proposition follows immediately from (55) and (56). \square

Proof of Proposition 3.8. We have,

$$\|\Pi_{U_m}(\mathbf{u} - \mathbf{u}_r)\|_U^2 = \|\mathbf{u} - \mathbf{u}_r\|_U^2 - \|\mathbf{u} - \Pi_{U_m} \mathbf{u}\|_U^2.$$

Therefore

$$\|\Pi_{U_m}(\mathbf{u} - \mathbf{u}_r)\|_U \leq \|\mathbf{u} - \mathbf{u}_r\|_U \leq \frac{1}{\sqrt{1 - \tau^2}} \|\Pi_{U_m}(\mathbf{u} - \mathbf{u}_r)\|_U. \quad (57)$$

On the other hand,

$$\begin{aligned} \|\Pi_{U_m}(\mathbf{u} - \mathbf{u}_r)\|_U &\leq \|\Pi_{U_m} \mathbf{R}_U^{-1} \mathbf{B}(\mathbf{u} - \mathbf{u}_r)\|_U + \|\Pi_{U_m} \mathbf{R}_U^{-1} \mathbf{E}(\mathbf{u} - \mathbf{u}_r)\|_U \\ &= \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'_m} + \|\Pi_{U_m} \mathbf{R}_U^{-1} \mathbf{E}(\mathbf{u} - \mathbf{u}_r)\|_U \\ &\leq \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'_m} + \|\Pi_{U_m} \mathbf{R}_U^{-1} \mathbf{E} \Pi_{U_m}(\mathbf{u} - \mathbf{u}_r)\|_U + \|\Pi_{U_m} \mathbf{R}_U^{-1} \mathbf{E}(\mathbf{u} - \Pi_{U_m} \mathbf{u})\|_U \\ &\leq \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'_m} + \|\mathbf{E}\|_{U_m, U'_m} \|\mathbf{u} - \mathbf{u}_r\|_U + \|\mathbf{E}\|_{U, U'_m} \|\mathbf{u} - \Pi_{U_m} \mathbf{u}\|_U \\ &\leq \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'_m} + (\|\mathbf{E}\|_{U_m, U'_m} + \tau \|\mathbf{E}\|_{U, U'_m}) \|\mathbf{u} - \mathbf{u}_r\|_U, \end{aligned} \quad (58)$$

and similarly

$$\|\Pi_{U_m}(\mathbf{u} - \mathbf{u}_r)\|_U \geq \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'_m} - (\|\mathbf{E}\|_{U_m, U'_m} + \tau \|\mathbf{E}\|_{U, U'_m}) \|\mathbf{u} - \mathbf{u}_r\|_U. \quad (59)$$

The statement of the proposition can be obtained by combining (57) to (59). \square

Proof of Proposition 4.3. If not unique, we choose $\Pi_{U_r}^\ominus$ such that $\Pi_{U_r}^\ominus \mathbf{x} = \mathbf{x}$ holds for all vectors $\mathbf{x} \in U_r$. The following result can be derived by following the proof of Lemma 3.2 substituting $\|\cdot\|_U$, Π_{U_r} , $\|\cdot\|_{U'_r}$ by $\|\cdot\|_{U^\ominus}$, $\Pi_{U_r}^\ominus$, $\|\cdot\|_{U'^\ominus_r}$:

$$\|\mathbf{u}_r - \Pi_{U_r}^\ominus \mathbf{u}\|_{U^\ominus} \leq \|\mathbf{E}(\mathbf{u} - \Pi_{U_r}^\ominus \mathbf{u})\|_{U'^\ominus_r} + \|\mathbf{E}(\mathbf{u}_r - \Pi_{U_r}^\ominus \mathbf{u})\|_{U'^\ominus_r}.$$

Then,

$$\sqrt{1 - \varepsilon} \|\mathbf{u}_r - \Pi_{U_r}^\ominus \mathbf{u}\|_{U^\ominus} \leq \|\mathbf{u}_r - \Pi_{U_r}^\ominus \mathbf{u}\|_{U^\ominus} \leq \|\mathbf{E}\|_{U, U'_r} \|\mathbf{u} - \Pi_{U_r}^\ominus \mathbf{u}\|_U + \|\mathbf{E}\|_{U_r, U'_r} \|\mathbf{u}_r - \Pi_{U_r}^\ominus \mathbf{u}\|_U.$$

This result and the inequality,

$$\|\mathbf{u}_r - \Pi_{U_r}^\ominus \mathbf{u}\|_U \geq \|\mathbf{u} - \mathbf{u}_r\|_U - \|\mathbf{u} - \Pi_{U_r}^\ominus \mathbf{u}\|_U,$$

yield (35). \square

Proof of Proposition 4.4. The first part of the proposition follows directly from the definition of Θ . More specifically, since Θ is a $(\varepsilon, \delta, r + 1)$ oblivious embedding, hence

$$\forall \mathbf{x}, \mathbf{y} \in V, \quad |\langle \mathbf{x}, \mathbf{y} \rangle_U - \langle \mathbf{x}, \mathbf{y} \rangle_U^\Theta| \leq \varepsilon \|\mathbf{x}\|_U \|\mathbf{y}\|_U, \quad (60)$$

holds with probability at least $1 - \delta$ for all $r + 1$ -dimensional subspaces $V \subset U$, and in particular, for $V = U_r + \text{span}(\mathbf{u})$. Now, let us provide the proof for the second part of the proposition.

Let $U_m^* := \arg \min_{\dim(W)=m} \sup_{\mathbf{u} \in \mathcal{M}} \|\mathbf{u} - \Pi_W \mathbf{u}\|_U$. Then, the Kolmogorov m -width is given by $d_m(\mathcal{M}) = \sup_{\mathbf{u} \in \mathcal{M}} \|\mathbf{u} - \Pi_{U_m^*} \mathbf{u}\|_U$. Since Θ is a $(\varepsilon, \delta, r + m)$ oblivious embedding, then (60) holds with probability at least $1 - \delta$ for $V = U_r + U_m^*$. This implies that

$$\begin{aligned} \|\mathbf{u} - \mathbf{w}\|_U^\Theta &\leq \|\Pi_{U_r+U_m^*} \mathbf{u} - \mathbf{w}\|_U^\Theta + \|\mathbf{u} - \Pi_{U_r+U_m^*} \mathbf{u}\|_U^\Theta \\ &\leq \sqrt{1 + \varepsilon} \|\Pi_{U_r+U_m^*} \mathbf{u} - \mathbf{w}\|_U + \|\mathbf{u} - \Pi_{U_r+U_m^*} \mathbf{u}\|_U^\Theta \\ &\leq \sqrt{1 + \varepsilon} (\|\mathbf{u} - \mathbf{w}\|_U + d_m(\mathcal{M})) + \|\mathbf{u} - \Pi_{U_r+U_m^*} \mathbf{u}\|_U^\Theta, \end{aligned} \quad (61)$$

and (similarly)

$$\|\mathbf{u} - \mathbf{w}\|_U^\Theta \geq \sqrt{1 - \varepsilon} (\|\mathbf{u} - \mathbf{w}\|_U - d_m(\mathcal{M})) - \|\mathbf{u} - \Pi_{U_r+U_m^*} \mathbf{u}\|_U^\Theta, \quad (62)$$

hold simultaneously for all $\mathbf{u} \in \mathcal{M}$ and $\mathbf{w} \in U_r$ with probability at least $1 - \delta$.

It remains to provide an upper bound for $\|\mathbf{u} - \Pi_{U_r+U_m^*} \mathbf{u}\|_U^\Theta$. Denote $\Delta \mathbf{u} := \mathbf{u} - \Pi_{U_r+U_m^*} \mathbf{u}$. Let $S = \{\mathbf{v}_i : 1 \leq i \leq n\}$ be some orthogonal basis for U . Let $\{S_j : 1 \leq j \leq \lceil \frac{n}{m} \rceil\}$, be a partition of S into subsets with $\#S_j \leq m$. Since Θ is a $(\varepsilon, \delta(\frac{n}{m} + 1)^{-1}, m)$ oblivious embedding, the relation (60) holds for every $V = \text{span}(S_j)$, $1 \leq j \leq \lceil \frac{n}{m} \rceil$, simultaneously, with probability at least $1 - \delta$. Therefore, for all $\mathbf{u} \in \mathcal{M}$,

$$\begin{aligned} \|\Delta \mathbf{u}\|_U^\Theta &= \left\| \sum_{j=1}^{\lceil \frac{n}{m} \rceil} \Pi_{S_j} \Delta \mathbf{u} \right\|_U^\Theta \leq \sum_{j=1}^{\lceil \frac{n}{m} \rceil} \|\Pi_{S_j} \Delta \mathbf{u}\|_U^\Theta \leq \sqrt{1 + \varepsilon} \sum_{j=1}^{\lceil \frac{n}{m} \rceil} \|\Pi_{S_j} \Delta \mathbf{u}\|_U \\ &\leq \sqrt{1 + \varepsilon} \sqrt{\lceil \frac{n}{m} \rceil} \sqrt{\sum_{j=1}^{\lceil \frac{n}{m} \rceil} \|\Delta \mathbf{u}\|_U^2} = \sqrt{1 + \varepsilon} \sqrt{\frac{n}{m} + 1} \|\Delta \mathbf{u}\|_U \leq D d_m(\mathcal{M}), \end{aligned} \quad (63)$$

holds with probability at least $1 - \delta$. The second part of the proposition then follows directly from relations (61) to (63) and an union bound. \square

Proof of Proposition 4.5. Let \mathbf{T} be a square matrix such that $\mathbf{U}_r \mathbf{T}$ has orthonormal columns with respect to $\langle \cdot, \cdot \rangle_U$. Since Θ satisfies (42) with $V = U_r$, we have

$$\forall \mathbf{x} \in \mathbb{K}^r, \quad \frac{1}{1 + \varepsilon} \|\mathbf{x}\|_2^2 \leq \|\mathbf{U}_r \mathbf{x}\|_U^2 = \|\mathbf{T}^{-1} \mathbf{x}\|_2^2 \leq \frac{1}{1 - \varepsilon} \|\mathbf{x}\|_2^2.$$

This implies that the singular values of \mathbf{T} lie inside $[\sqrt{1 - \varepsilon}, \sqrt{1 + \varepsilon}]$. The identity $\Pi_{U_r}^\Theta = \mathbf{U}_r \mathbf{U}_r^H \Theta^H \Theta$ implies that

$$\|\cdot\|_{U_r'}^\Theta = \|\Pi_{U_r}^\Theta \mathbf{R}_U^{-1} \cdot\|_U^\Theta = \|\mathbf{U}_r \mathbf{U}_r^H \Theta^H \Theta \mathbf{R}_U^{-1} \cdot\|_U^\Theta = \|\mathbf{U}_r^H \Theta^H \Theta \mathbf{R}_U^{-1} \cdot\|_2.$$

From this fact we obtain the following expressions for $\|\mathbf{E}\|_{U_r, U'_r}^\ominus$ and $\|\mathbf{E}\|_{U, U'_r}^\ominus$:

$$\|\mathbf{E}\|_{U_r, U'_r}^\ominus = \max_{\mathbf{v} \in U_r \setminus \{\mathbf{0}\}} \frac{\|\mathbf{U}_r^H \Theta^H \Theta \mathbf{R}_U^{-1} \mathbf{E} \mathbf{v}\|_2}{\|\mathbf{v}\|_U} = \|\mathbf{U}_r^H \Theta^H \Theta \mathbf{R}_U^{-1} \mathbf{E} \mathbf{U}_r \mathbf{T}\|_2 \quad (64a)$$

and

$$\|\mathbf{E}\|_{U, U'_r}^\ominus := \max_{\mathbf{v} \in U \setminus \{\mathbf{0}\}} \frac{\|\mathbf{U}_r^H \Theta^H \Theta \mathbf{E} \mathbf{v}\|_2}{\|\mathbf{v}\|_U} = \|\mathbf{U}_r^H \Theta^H \Theta \mathbf{R}_U^{-1} \mathbf{E} \mathbf{R}^{-1/2}\|_2. \quad (64b)$$

Then

$$\sqrt{1 - \varepsilon} \|\mathbf{U}_r^H \Theta^H \Theta \mathbf{R}_U^{-1} \mathbf{E} \mathbf{U}_r\|_2 \leq \|\mathbf{E}\|_{U_r, U'_r}^\ominus \leq \sqrt{1 + \varepsilon} \|\mathbf{U}_r^H \Theta^H \Theta \mathbf{R}_U^{-1} \mathbf{E} \mathbf{U}_r\|_2$$

The statement of the proposition then follows from the fact that for a matrix \mathbf{X} with r rows,

$$r^{-1/2} \|\mathbf{X}\|_F \leq \|\mathbf{X}\|_2 \leq \|\mathbf{X}\|_F.$$

□

Proof of Proposition 4.7. Assume that $\mathbf{U}_r = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r]$ is orthogonal with respect to $\langle \cdot, \cdot \rangle_U$. In addition let $\mathbf{U}_n = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n]$ be a matrix whose columns form an orthonormal basis of U .

We have,

$$\Delta_{U_r, U'_r}^2 = \|\Pi_{U_r} \mathbf{R}_U^{-1} \mathbf{E} \mathbf{U}_r\|_{HS(\ell_2, U)}^2 = \sum_{i=1}^r \|\mathbf{E} \mathbf{v}_i\|_{U'_r}^2, \quad \Delta_{U_r, U'}^2 = \|\mathbf{E} \mathbf{U}_r\|_{HS(\ell_2, U')}^2 = \sum_{i=1}^r \|\mathbf{E} \mathbf{v}_i\|_{U'}^2, \quad (65a)$$

$$\Delta_{U, U'_r}^2 = \|\Pi_{U_r} \mathbf{R}_U^{-1} \mathbf{E} \mathbf{U}_n\|_{HS(\ell_2, U)}^2 = \sum_{i=1}^n \|\mathbf{E} \mathbf{e}_i\|_{U'_r}^2, \quad \Delta_{U, U'}^2 = \|\mathbf{E} \mathbf{U}_n\|_{HS(\ell_2, U')}^2 = \sum_{i=1}^n \|\mathbf{E} \mathbf{e}_i\|_{U'}^2 \quad (65b)$$

and

$$(\Delta_{U_r, U'_r}^\ominus)^2 = (\|\Pi_{U_r} \mathbf{R}_U^{-1} \mathbf{E} \mathbf{U}_r\|_{HS(\ell_2, U)}^\ominus)^2 = \sum_{i=1}^r (\|\mathbf{E} \mathbf{v}_i\|_{U'_r}^\ominus)^2, \quad (65c)$$

$$(\Delta_{U, U'_r}^\ominus)^2 = (\|\Pi_{U_r} \mathbf{R}_U^{-1} \mathbf{E} \mathbf{U}_n\|_{HS(\ell_2, U)}^\ominus)^2 = \sum_{i=1}^n (\|\mathbf{E} \mathbf{e}_i\|_{U'_r}^\ominus)^2. \quad (65d)$$

By definition of Θ and a union bound argument, the relation

$$\forall \mathbf{x}, \mathbf{y} \in V_i, \quad |\langle \mathbf{x}, \mathbf{y} \rangle_U - \langle \mathbf{x}, \mathbf{y} \rangle_U^\ominus| \leq \varepsilon \|\mathbf{x}\|_U \|\mathbf{y}\|_U, \quad (66)$$

holds with probability at least $1 - \delta$ for all $V_i = U_r + \{\mathbf{R}_U^{-1} \mathbf{E} \mathbf{v}_i : \mathbf{B} \in Y\}$, $1 \leq i \leq r$, simultaneously. Then, by [3, Proposition 3.4],

$$\|\mathbf{E} \mathbf{v}_i\|_{U'_r}^\ominus \leq \frac{1}{\sqrt{1 - \varepsilon}} (\|\mathbf{E} \mathbf{v}_i\|_{U'_r} + \varepsilon \|\mathbf{E} \mathbf{v}_i\|_{U'}),$$

holds with probability at least $1 - \delta$ for all $1 \leq i \leq r$, simultaneously. This fact, combined with (65a) and (65c) results in the first part of the proposition.

Next is proved the second part of the proposition. By definition of Θ , the relation (66) holds with probability at least $1 - \delta$ for all $V_i = U_r + \{\mathbf{R}_U^{-1}\mathbf{E}\mathbf{e}_i : \mathbf{B} \in Y\}$, $1 \leq i \leq n$, simultaneously. Therefore, by [3, Proposition 3.4],

$$\|\mathbf{E}\mathbf{e}_i\|_{U'_r}^{\Theta} \leq \frac{1}{\sqrt{1-\varepsilon}}(\|\mathbf{E}\mathbf{e}_i\|_{U'_r} + \varepsilon\|\mathbf{E}\mathbf{e}_i\|_{U'}), \quad (67)$$

holds with probability at least $1 - \delta$ for all $1 \leq i \leq n$, simultaneously. The proof is completed by combining (67) with (65b) and (65d). \square

Proof of Proposition 4.8. If not unique, we choose $\Pi_{U_m}^{\Theta}$ such that $\Pi_{U_m}^{\Theta}\mathbf{x} = \mathbf{x}$ holds for all vectors $\mathbf{x} \in U_m$. We have,

$$(\|\Pi_{U_m}^{\Theta}(\mathbf{u} - \mathbf{u}_r)\|_U^{\Theta})^2 = (\|\mathbf{u} - \mathbf{u}_r\|_U^{\Theta})^2 - (\|\mathbf{u} - \Pi_{U_m}^{\Theta}\mathbf{u}\|_U^{\Theta})^2,$$

and

$$\frac{1}{\sqrt{1+\varepsilon}}\|\mathbf{u} - \mathbf{u}_r\|_U^{\Theta} \leq \|\mathbf{u} - \mathbf{u}_r\|_U \leq \frac{1}{\sqrt{1-\varepsilon}}\|\mathbf{u} - \mathbf{u}_r\|_U^{\Theta}.$$

Therefore

$$\frac{1}{\sqrt{1+\varepsilon}}\|\Pi_{U_m}^{\Theta}(\mathbf{u} - \mathbf{u}_r)\|_U^{\Theta} \leq \|\mathbf{u} - \mathbf{u}_r\|_U \leq \frac{1}{\sqrt{1-\varepsilon-\tau^{*2}}}\|\Pi_{U_m}^{\Theta}(\mathbf{u} - \mathbf{u}_r)\|_U^{\Theta}. \quad (68)$$

On the other hand,

$$\begin{aligned} \|\Pi_{U_m}^{\Theta}(\mathbf{u} - \mathbf{u}_r)\|_U^{\Theta} &\leq \|\Pi_{U_m}^{\Theta}\mathbf{R}_U^{-1}\mathbf{B}(\mathbf{u} - \mathbf{u}_r)\|_U^{\Theta} + \|\Pi_{U_m}^{\Theta}\mathbf{R}_U^{-1}\mathbf{E}(\mathbf{u} - \mathbf{u}_r)\|_U^{\Theta} \\ &= \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'_m}^{\Theta} + \|\Pi_{U_m}^{\Theta}\mathbf{R}_U^{-1}\mathbf{E}(\mathbf{u} - \mathbf{u}_r)\|_U^{\Theta} \\ &\leq \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'_m}^{\Theta} + \|\Pi_{U_m}^{\Theta}\mathbf{R}_U^{-1}\mathbf{E}\Pi_{U_m}(\mathbf{u} - \mathbf{u}_r)\|_U^{\Theta} + \|\Pi_{U_m}^{\Theta}\mathbf{R}_U^{-1}\mathbf{E}(\mathbf{u} - \Pi_{U_m}\mathbf{u})\|_U^{\Theta} \\ &\leq \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'_m}^{\Theta} + \|\mathbf{E}\|_{U_m, U'_m}^{\Theta}\|\mathbf{u} - \mathbf{u}_r\|_U + \|\mathbf{E}\|_{U, U'_m}^{\Theta}\|\mathbf{u} - \Pi_{U_m}\mathbf{u}\|_U \\ &\leq \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'_m}^{\Theta} + (\|\mathbf{E}\|_{U_m, U'_m}^{\Theta} + \tau^*\|\mathbf{E}\|_{U, U'_m}^{\Theta})\|\mathbf{u} - \mathbf{u}_r\|_U. \end{aligned} \quad (69)$$

Similarly, we have

$$\|\Pi_{U_m}^{\Theta}(\mathbf{u} - \mathbf{u}_r)\|_U^{\Theta} \geq \|\mathbf{r}^*(\mathbf{u}_r)\|_{U'_m}^{\Theta} - (\|\mathbf{E}\|_{U_m, U'_m}^{\Theta} + \tau^*\|\mathbf{E}\|_{U, U'_m}^{\Theta})\|\mathbf{u} - \mathbf{u}_r\|_U. \quad (70)$$

The statement of the proposition follows by combining (68) to (70). \square

Propositions 5.2 to 5.4 will follow from the following proposition.

Proposition 7.1. *The random map $\Upsilon(\cdot)$ defined by*

$$\Upsilon(\mathbf{X}) := \Gamma \text{vec}(\mathbf{\Omega}\mathbf{X}\mathbf{\Sigma}^H), \quad \mathbf{X} \in \mathbb{K}^{q \times p},$$

where Γ , $\mathbf{\Omega}$ and $\mathbf{\Sigma}$ are $(\varepsilon_{\Gamma}, \delta_{\Gamma}, d)$, $(\varepsilon_{\mathbf{\Omega}}, \delta_{\mathbf{\Omega}}, d)$ and $(\varepsilon_{\mathbf{\Sigma}}, \delta_{\mathbf{\Sigma}}, d)$ oblivious $\ell_2 \rightarrow \ell_2$ subspace embeddings, is a (ε, δ, d) oblivious HS(ℓ_2, ℓ_2) $\rightarrow \ell_2$ subspace embedding of matrices with $\varepsilon = (1 + \varepsilon_{\Gamma})(1 + \varepsilon_{\mathbf{\Sigma}})(1 + \varepsilon_{\mathbf{\Omega}}) - 1$ and $\delta = \min(k_{\mathbf{\Sigma}}\delta_{\mathbf{\Omega}} + q\delta_{\mathbf{\Sigma}}, k_{\mathbf{\Omega}}\delta_{\mathbf{\Sigma}} + p\delta_{\mathbf{\Omega}}) + \delta_{\Gamma}$, where $k_{\mathbf{\Omega}}$ and $k_{\mathbf{\Sigma}}$ are the numbers of rows of $\mathbf{\Omega}$ and $\mathbf{\Sigma}$, respectively.

Proof. Let us first assume that

$$k_{\Omega}\delta_{\Sigma} + p\delta_{\Omega} \leq k_{\Sigma}\delta_{\Omega} + q\delta_{\Sigma}$$

Let V be a d -dimensional space of matrices in $\mathbb{K}^{q \times p}$. Define

$$V^i = \{\mathbf{V}\mathbf{e}_i : \mathbf{V} \in V\} \subset \mathbb{K}^q,$$

where \mathbf{e}_i denotes the i -th column of the $p \times p$ identity matrix. By the definition of Ω and a union bound argument, we have that

$$\forall \mathbf{x} \in V^i, \quad \left| \|\mathbf{x}\|_2^2 - \|\Omega\mathbf{x}\|_2^2 \right| \leq \varepsilon_{\Omega} \|\mathbf{x}\|_2^2,$$

holds for all V^i , $1 \leq i \leq p$, with probability at least $1 - p\delta_{\Omega}$. This implies that

$$\left| \|\mathbf{V}\mathbf{e}_i\|_2^2 - \|\Omega[\mathbf{V}\mathbf{e}_i]\|_2^2 \right| \leq \varepsilon_{\Omega} \|\mathbf{V}\mathbf{e}_i\|_2^2 \quad (71)$$

holds with probability at least $1 - p\delta_{\Sigma}$ for all $\mathbf{V} \in V$ and $1 \leq i \leq p$. By (71) and the following identities

$$\|\mathbf{V}\|_F^2 = \sum_{i=1}^p \|\mathbf{V}\mathbf{e}_i\|_2^2 \quad \text{and} \quad \|\Omega\mathbf{V}\|_F^2 = \sum_{i=1}^p \|\Omega[\mathbf{V}\mathbf{e}_i]\|_2^2,$$

we deduce that

$$\mathbb{P}(\forall \mathbf{V} \in V, \left| \|\mathbf{V}\|_F^2 - \|\Omega\mathbf{V}\|_F^2 \right| \leq \varepsilon_{\Omega} \|\mathbf{V}\|_F^2) \geq 1 - q\delta_{\Omega}. \quad (72)$$

Furthermore, by taking $V := \{\mathbf{X}^H \Omega^H : \mathbf{X} \in V\}$ and replacing Ω by Σ in (72), we also deduce that

$$\mathbb{P}(\forall \mathbf{V} \in V, \left| \|\mathbf{V}^H \Omega^H\|_F^2 - \|\Sigma \mathbf{V}^H \Omega^H\|_F^2 \right| \leq \varepsilon_{\Sigma} \|\mathbf{V}^H \Omega^H\|_F^2) \geq 1 - k_{\Omega} \delta_{\Sigma}. \quad (73)$$

Finally, by definition of Γ and the identity $\|\Omega\mathbf{V}\Sigma^H\|_F^2 = \|\text{vec}(\Omega\mathbf{V}\Sigma^H)\|_2^2$, we have

$$\mathbb{P}(\forall \mathbf{V} \in V, \left| \|\Omega\mathbf{V}\Sigma^H\|_F^2 - \|\Gamma \text{vec}(\Omega\mathbf{V}\Sigma^H)\|_2^2 \right| \leq \varepsilon_{\Gamma} \|\Omega\mathbf{V}\Sigma^H\|_F^2) \geq 1 - \delta_{\Gamma}. \quad (74)$$

By combining (72) to (74) and using a union bound, we obtain that

$$\begin{aligned} & \left| \|\mathbf{V}\|_F^2 - \|\Gamma \text{vec}(\Omega\mathbf{V}\Sigma^H)\|_2^2 \right| \leq \left| \|\mathbf{V}\|_F^2 - \|\Omega\mathbf{V}\Sigma^H\|_F^2 \right| + \left| \|\Omega\mathbf{V}\Sigma^H\|_F^2 - \|\Gamma \text{vec}(\Omega\mathbf{V}\Sigma^H)\|_2^2 \right| \\ & \leq \left| \|\mathbf{V}\|_F^2 - \|\Omega\mathbf{V}\Sigma^H\|_F^2 \right| + \varepsilon_{\Gamma} \|\Omega\mathbf{V}\Sigma^H\|_F^2 \\ & \leq (1 + \varepsilon_{\Gamma}) \left| \|\mathbf{V}\|_F^2 - \|\Omega\mathbf{V}\Sigma^H\|_F^2 \right| + \varepsilon_{\Gamma} \|\mathbf{V}\|_F^2 \\ & \leq (1 + \varepsilon_{\Gamma}) \left(\left| \|\mathbf{V}\|_F^2 - \|\Omega\mathbf{V}\|_F^2 \right| + \left| \|\Omega\mathbf{V}\|_F^2 - \|\Omega\mathbf{V}\Sigma^H\|_F^2 \right| \right) + \varepsilon_{\Gamma} \|\mathbf{V}\|_F^2 \\ & \leq (1 + \varepsilon_{\Gamma}) \left(\left| \|\mathbf{V}\|_F^2 - \|\Omega\mathbf{V}\|_F^2 \right| + \varepsilon_{\Sigma} \|\Omega\mathbf{V}\|_F^2 \right) + \varepsilon_{\Gamma} \|\mathbf{V}\|_F^2 \\ & \leq (1 + \varepsilon_{\Gamma})(1 + \varepsilon_{\Sigma}) \left| \|\mathbf{V}\|_F^2 - \|\Omega\mathbf{V}\|_F^2 \right| + (\varepsilon_{\Gamma} + \varepsilon_{\Sigma}(1 + \varepsilon_{\Gamma})) \|\mathbf{V}\|_F^2 \\ & \leq (1 + \varepsilon_{\Gamma})(1 + \varepsilon_{\Sigma}) \varepsilon_{\Omega} \|\mathbf{V}\|_F^2 + (\varepsilon_{\Gamma} + \varepsilon_{\Sigma}(1 + \varepsilon_{\Gamma})) \|\mathbf{V}\|_F^2 \\ & = \varepsilon \|\mathbf{V}\|_F^2 \end{aligned}$$

holds with probability at least $1 - \delta$ for all $\mathbf{V} \in V$. This statement with the parallelogram identity imply that, with probability at least $1 - \delta$, for all $\mathbf{X}, \mathbf{Y} \in V$

$$|\langle \mathbf{X}, \mathbf{Y} \rangle_F - \langle \Upsilon(\mathbf{X}), \Upsilon(\mathbf{Y}) \rangle_F| \leq \varepsilon \|\mathbf{X}\|_F \|\mathbf{Y}\|_F,$$

which completes the proof for the case

$$k_{\Omega}\delta_{\Sigma} + p\delta_{\Omega} \leq k_{\Sigma}\delta_{\Omega} + q\delta_{\Sigma}.$$

For the alternative case, we can apply the proof of the first case by interchanging Ω with Σ , p and q , and considering a reshaping operator $\text{vec}^*(\cdot) := \text{vec}(\cdot^{\text{H}})$ instead of the operator $\text{vec}(\cdot)$ to show that the linear map

$$\mathbf{X} \in \mathbb{K}^{p \times q} \rightarrow \Gamma \text{vec}^*(\Sigma \mathbf{X} \Omega^{\text{H}})$$

is a (ε, δ, d) oblivious $HS(\ell_2, \ell_2) \rightarrow \ell_2$ subspace embedding. Since the Frobenius inner product of two matrices is equal to the Frobenius inner product of the (Hermitian-)transposed matrices, the linear map

$$\mathbf{X} \in \mathbb{K}^{p \times q} \rightarrow \Gamma \text{vec}^*(\Sigma \mathbf{X}^{\text{H}} \Omega^{\text{H}})$$

is also a (ε, δ, d) oblivious $HS(\ell_2, \ell_2) \rightarrow \ell_2$ subspace embedding. The proof is completed by noticing that

$$\Gamma \text{vec}^*(\Sigma \mathbf{X}^{\text{H}} \Omega^{\text{H}}) = \Gamma \text{vec}((\Sigma \mathbf{X}^{\text{H}} \Omega^{\text{H}})^{\text{H}}) = \Gamma \text{vec}(\Omega \mathbf{X} \Sigma^{\text{H}}).$$

□

Proof of Proposition 5.2. Proposition 5.2 directly follows from Proposition 7.1. □

Proof of Proposition 5.3. We have that $\Xi(\cdot) = \Upsilon(\mathbf{Q}\cdot)$, where $\Upsilon(\cdot)$ is a (ε, δ, d) oblivious $H(\ell_2, \ell_2) \rightarrow \ell_2$ embedding from Proposition 7.1. Since $\langle \cdot, \cdot \rangle_{HS(\ell_2, U)} = \langle \mathbf{Q}\cdot, \mathbf{Q}\cdot \rangle_F$, we have that $\Xi(\cdot)$ is a (ε, δ, d) oblivious $H(\ell_2, U) \rightarrow \ell_2$ embedding. □

Proof of Proposition 5.4. We have that $\Xi(\cdot) = \Upsilon(\mathbf{Q}\cdot\mathbf{Q}^{\text{H}})$, where $\Upsilon(\cdot)$ is a (ε, δ, d) oblivious $H(\ell_2, \ell_2) \rightarrow \ell_2$ embedding from Proposition 7.1. The proposition then follows from the fact that $\langle \cdot, \cdot \rangle_{HS(U', U)} = \langle \mathbf{Q}\cdot\mathbf{Q}^{\text{H}}, \mathbf{Q}\cdot\mathbf{Q}^{\text{H}} \rangle_F$. □