Interpolation of inverse operators for preconditioning parameter-dependent equations

Olivier Zahm, Anthony Nouy

Interpolation of inverse operators for preconditioning parameter-dependent equations *

Olivier ZAHM† and Anthony NOUY†‡

January 28, 2016

Abstract

We propose a method for the construction of preconditioners of parameter-dependent matrices for the solution of large systems of parameter-dependent equations. The proposed method is an interpolation of the matrix inverse based on a projection of the identity matrix with respect to the Frobenius norm. Approximations of the Frobenius norm using random matrices are introduced in order to handle large matrices. The resulting statistical estimators of the Frobenius norm yield quasi-optimal projections that are controlled with high probability. Strategies for the adaptive selection of interpolation points are then proposed for different objectives in the context of projection-based model order reduction methods: the improvement of residual-based error estimators, the improvement of the projection on a given reduced approximation space, or the re-use of computations for sampling based model order reduction methods.

1 Introduction

This paper is concerned with the solution of large systems of parameter-dependent equations of the form

$$A(\xi)u(\xi) = b(\xi), \tag{1}$$

where ξ takes values in some parameter set Ξ. Such problems occur in several contexts such as parametric analyses, optimization, control or uncertainty quantification, where ξ are random variables that parametrize model or data uncertainties. The efficient solution of equation (1) generally requires the construction of preconditioners for the operator $A(\xi)$,

†Ecole Centrale de Nantes, GeM, UMR CNRS 6183, France.

‡Corresponding author (anthony.nouy@ec-nantes.fr).

*This work was supported by the French National Research Agency (Grant ANR CHORUS MONU-0005)

either for improving the performance of iterative solvers or for improving the quality of residual-based projection methods.

A basic preconditioner can be defined as the inverse (or any preconditioner) of the matrix $A(\bar{\xi})$ at some nominal parameter value $\bar{\xi} \in \Xi$ or as the inverse (or any preconditioner) of a mean value of $A(\xi)$ over Ξ (see e.g. [21, 20]). When the operator only slightly varies over the parameter set Ξ, these parameter-independent preconditioners behave relatively well. However, for large variabilities, they are not able to provide a good preconditioning over the whole parameter set Ξ. A first attempt to construct a parameter-dependent preconditioner can be found in [17], where the authors compute through quadrature a polynomial expansion of the parameter-dependent factors of a LU factorization of $A(\xi)$. More recently, a linear Lagrangian interpolation of the matrix inverse has been proposed in [11]. The generalization to any standard multivariate interpolation method is straightforward. However, standard approximation or interpolation methods require the evaluation of matrix inverses (or factorizations) for many instances of ξ on a prescribed structured grid (quadrature or interpolation), that becomes prohibitive for large matrices and high dimensional parametric problems.

In this paper, we propose an interpolation method for the inverse of matrix $A(\xi)$. The interpolation is obtained by a projection of the inverse matrix on a linear span of samples of $A(\xi)^{-1}$ and takes the form

$$P_m(\xi) = \sum_{i=1}^{m} \lambda_i(\xi) A(\xi_i)^{-1},$$

where $\xi_1, \ldots, \xi_m$ are m arbitrary interpolation points in Ξ. A natural interpolation could be obtained by minimizing the condition number of $P_m(\xi)A(\xi)$ over the $\lambda_i(\xi)$, which is a Clarke regular strongly pseudoconvex optimization problem [30]. However, the solution of this non standard optimization problem for many instances of ξ is intractable and proposing an efficient solution method in a multi-query context remains a challenging issue. Here, the projection is defined as the minimizer of the Frobenius norm of $I - P_m(\xi)A(\xi)$, that is a quadratic optimization problem. Approximations of the Frobenius norm using random matrices are introduced in order to handle large matrices. These statistical estimations of the Frobenius norm allow to obtain quasi-optimal projections that are controlled with high probability. Since we are interested in large matrices, $A(\xi_i)^{-1}$ are here considered as implicit matrices for which only efficient matrix-vector multiplications are available. Typically, a factorization (e.g. LU) of $A(\xi_i)$ is computed and stored. Note that when the storage of factorizations of several samples of the operator is unaffordable or when efficient preconditioners are readily available, one could similarly consider projections of the inverse operator on the linear span of preconditioners of samples of the operator. However, the resulting parameter-dependent preconditioner would be no more an interpolation of preconditioners.

This straightforward extension of the proposed method is not analyzed in the present paper.

The paper then presents several contributions in the context of projection-based model order reduction methods (e.g. Reduced Basis, Proper Orthogonal Decomposition (POD), Proper Generalized Decompositon) that rely on the projection of the solution $u(\xi)$ of (1) on a low-dimensional approximation space. We first show how the proposed preconditioner can be used to define a Galerkin projection-based on the preconditioned residual, which can be interpreted as a Petrov-Galerkin projection of the solution with a parameter-dependent test space. Then, we propose adaptive construction of the preconditioner, based on an adaptive selection of interpolation points, for different objectives: (i) the improvement of error estimators based on preconditioned residuals, (ii) the improvement of the quality of projections on a given low-dimensional approximation space, or (iii) the re-use of computations for sample-based model order reduction methods. Starting from a m-point interpolation, these adaptive strategies consist in choosing a new interpolation point based on different criteria. In (i), the new point is selected for minimizing the distance between the identity and the preconditioned operator. In (ii), it is selected for improving the quasi-optimality constant of Petrov-Galerkin projections which measures how far the projection is from the best approximation on the reduced approximation space. In (iii), the new interpolation point is selected as a new sample determined for the approximation of the solution and not of the operator. The interest of the latter approach is that when direct solvers are used to solve equation (1) at some sample points, the corresponding factorizations of the matrix can be stored and the preconditioner can be computed with a negligible additional cost.

The paper is organized as follows. In Section 2 we present the method for the interpolation of the inverse of a parameter-dependent matrix. In Section 3, we show how the preconditioner can be used for the definition of a Petrov-Galerkin projection of the solution of (1) on a given reduced approximation space, and we provide an analysis of the quasi-optimality constant of this projection. Then, different strategies for the selection of interpolation points for the preconditioner are proposed in Section 4. Finally, in Section 5, numerical experiments will illustrate the efficiency of the proposed preconditioning strategies for different projection-based model order reduction methods.

Note that the proposed preconditioner could be also used (a) for improving the quality of Galerkin projection methods where a projection of the solution $u(\xi)$ is searched on a subspace of functions of the parameters (e.g. polynomial or piecewise polynomial spaces) [16, 31, 33], or (b) for preconditioning iterative solvers for (1), in particular solvers based on low-rank truncations that require a low-rank structure of the preconditioner [28, 32, 22, 23]. These two potential applications are not considered here.

2 Interpolation of the inverse of a parameter-dependent matrix using Frobenius norm projection

In this section, we propose a construction of an interpolation of the matrix-valued function $\xi \mapsto A(\xi)^{-1} \in \mathbb{R}^{n\times n}$ for given interpolation points $\xi_1, \ldots, \xi_m$ in Ξ. We let $P_i = A(\xi_i)^{-1}$, $1 \leq i \leq m$. For large matrices, the explicit computation of P_i is usually not affordable. Therefore, P_i is here considered as an implicit matrix and we assume that the product of P_i with a vector can be computed efficiently. In practice, factorizations of matrices $A(\xi_i)$ are stored.

2.1 Projection using Frobenius norm

We introduce the subspace $Y_m = \text{span}\{P_1, \ldots, P_m\}$ of $\mathbb{R}^{n\times n}$. An approximation $P_m(\xi)$ of $A(\xi)^{-1}$ in Y_m is then defined by

$$P_m(\xi) = \underset{P\in Y_m}{\text{argmin}} \ \|I - PA(\xi)\|_F, \tag{2}$$

where I denotes the identity matrix of size n, and $\|\cdot\|_F$ is the Frobenius norm such that $\|B\|_F^2 = \langle B, B\rangle_F$ with $\langle B, C\rangle_F = \text{trace}(B^T C)$. Since $A(\xi_i)^{-1} \in Y_m$, we have the interpolation property $P_m(\xi_i) = A(\xi_i)^{-1}$, $1 \leq i \leq m$. The minimization of $\|I - PA\|_F$ has been first proposed in [25] for the construction of a preconditioner P in a subspace of matrices with given sparsity pattern (SPAI method). The following proposition gives some properties of the operator $P_m(\xi)A(\xi)$ (see Lemma 2.6 and Theorem 3.2 in [24]).

Proposition 2.1 *Let $P_m(\xi)$ be defined by (2). We have*

$$(1 - \alpha_m(\xi))^2 \leq \|I - P_m(\xi)A(\xi)\|_F^2 \leq n(1 - \alpha_m^2(\xi)), \tag{3}$$

where $\alpha_m(\xi)$ is the lowest singular value of $P_m(\xi)A(\xi)$ verifying $0 \leq \alpha_m(\xi) \leq 1$, with $P_m(\xi)A(\xi) = I$ if and only if $\alpha_m(\xi) = 1$. Also, the following bound holds for the condition number of $P_m(\xi)A(\xi)$:

$$\kappa(P_m(\xi)A(\xi)) \leq \frac{\sqrt{n - (n-1)\alpha_m^2(\xi)}}{\alpha_m(\xi)}. \tag{4}$$

Under the condition $\|I - P_m(\xi)A(\xi)\|_F < 1$, equations (3) and (4) imply that

$$\kappa(P_m(\xi)A(\xi)) \leq \frac{\sqrt{n - (n-1)(1 - \|I - P_m(\xi)A(\xi)\|_F)^2}}{1 - \|I - P_m(\xi)A(\xi)\|_F}.$$

For all $\lambda \in \mathbb{R}^m$, we have

$$\|I - \sum_{i=1}^{m} \lambda_i P_i A(\xi)\|_F^2 = n - 2\lambda^T S(\xi) + \lambda^T M(\xi)\lambda,$$

where the matrix $M(\xi) \in \mathbb{R}^{m\times m}$ and the vector $S(\xi) \in \mathbb{R}^m$ are given by

$$M_{i,j}(\xi) = \text{trace}(A^T(\xi)P_i^T P_j A(\xi)) \quad \text{and} \quad S_i(\xi) = \text{trace}(P_i A(\xi)).$$

Therefore, the solution of problem (2) is $P_m(\xi) = \sum_{i=1}^m \lambda_i(\xi)P_i$ with $\lambda(\xi)$ the solution of $M(\xi)\lambda(\xi) = S(\xi)$. When considering a small number m of interpolation points, the computation time for solving this system of equations is negligible. However, the computation of $M(\xi)$ and $S(\xi)$ requires the evaluation of traces of matrices $A^T(\xi)P_i^T P_j A(\xi)$ and $P_i A(\xi)$ for all $1 \le i, j \le m$. Since the P_i are implicit matrices, the computation of such products of matrices is not affordable for large matrices. Of course, since $\text{trace}(B) = \sum_{i=1}^n e_i^T B e_i$, the trace of an implicit matrix B could be obtained by computing the product of B with the canonical vectors $e_1, \ldots, e_n$, but this approach is clearly not affordable for large n.

Hereafter, we propose an approximation of the above construction using an approximation of the Frobenius norm which requires less computational efforts.

2.2 Projection using a Frobenius semi-norm

Here, we define an approximation $P_m(\xi)$ of $A(\xi)^{-1}$ in Y_m by

$$P_m(\xi) = \underset{P\in Y_m}{\text{argmin}} \ \|(I - PA(\xi))V\|_F, \tag{5}$$

where $V \in \mathbb{R}^{n\times K}$, with $K \le n$. $B \mapsto \|BV\|_F$ defines a semi-norm on $\mathbb{R}^{n\times n}$. Here, we assume that the linear map $P \mapsto PA(\xi)V$ is injective on Y_m so that the solution of (5) is unique. This requires $K \ge m$ and is satisfied when $\text{rank}(V) \ge m$ and Y_m is the linear span of linearly independent invertible matrices. Then, the solution $P_m(\xi) = \sum_{i=1}^m \lambda_i(\xi)P_i$ of (5) is such that the vector $\lambda(\xi) \in \mathbb{R}^m$ satisfies $M^V(\xi)\lambda(\xi) = S^V(\xi)$, with

$$M^V_{i,j}(\xi) = \text{trace}(V^T A^T(\xi)P_i^T P_j A(\xi)V) \quad \text{and} \quad S_i^V(\xi) = \text{trace}(V^T P_i A(\xi)V). \tag{6}$$

The procedure for the computation of $M^V(\xi)$ and $S^V(\xi)$ is given in Algorithm 1. Note that only mK matrix-vector products involving the implicit matrices P_i are required.

Now the question is to choose a matrix V such that $\|(I - PA(\xi))V\|_F$ provides a good approximation of $\|I - PA(\xi)\|_F$ for any $P \in Y_m$ and $\xi \in \Xi$.

Algorithm 1 Computation of $M^V(\xi)$ and $S^V(\xi)$

Require: $A(\xi)$, $\{P_1, \ldots, P_m\}$ and $V = (v_1, \ldots, v_K)$

Ensure: $M^V(\xi)$ and $S^V(\xi)$

1: Compute the vectors $w_{i,k} = P_i A(\xi) v_k \in \mathbb{R}^n$, for $1 \leq k \leq K$ and $1 \leq i \leq m$

2: Set $W_i = (w_{i,1}, \ldots, w_{i,K}) \in \mathbb{R}^{n \times K}$, $1 \leq i \leq m$

3: Compute $M^V_{i,j}(\xi) = \mathrm{trace}(W_i^T W_j)$ for $1 \leq i, j \leq m$

4: Compute $S^V_i(\xi) = \mathrm{trace}(V^T W_i)$ for $1 \leq i \leq m$

2.2.1 Hadamard matrices for the estimation of the Frobenius norm of an implicit matrix

Let B an implicit n-by-n matrix (consider $B = I - PA(\xi)$, with $P \in Y_m$ and $\xi \in \Xi$). Following [3], we show how Hadamard matrices can be used for the estimation of the Frobenius norm of an implicit matrix. The goal is to find a matrix V such that $\|BV\|_F$ is a good approximation of $\|B\|_F$. The relation $\|BV\|_F^2 = \mathrm{trace}(B^T B V V^T)$ suggests that V should be such that VV^T is as close as possible to the identity matrix. For example, we would like V to minimize

$$err(V)^2 = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} (VV^T)_{i,j}^2 = \frac{\|I - VV^T\|_F^2}{n(n-1)},$$

which is the mean square magnitude of the off-diagonal entries of VV^T. The bound $err(V) \geq \sqrt{(n-K)/((n-1)K)}$ is known to hold for any $V \in \mathbb{R}^{n \times K}$ whose rows have unit norm [39]. Hadamard matrices can be used to construct matrices V such that the corresponding error $err(V)$ is close to the bound, see [3].

A Hadamard matrix H_s is a s-by-s matrix whose entries are ± 1, and which satisfies $H_s H_s^T = sI$ where I is the identity matrix of size s. For example,

$$H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

is a Hadamard matrix of size $s = 2$. The Kronecker product (denoted by $\otimes$) of two Hadamard matrices is again a Hadamard matrix. Then it is possible to build a Hadamard matrix whose size s is a power of 2 using a recursive procedure: $H_{2^{k+1}} = H_2 \otimes H_{2^k}$. The (i,j)-entry of this matrix is $(-1)^{a^T b}$, where a and b are the binary vectors such that $i = \sum_{k \geq 0} 2^k a_k$ and $j = \sum_{k \geq 0} 2^k b_k$. For a sufficiently large $s = 2^k \geq \max(n, K)$, we define the *rescaled partial Hadamard matrix* $V \in \mathbb{R}^{n \times K}$ as the first n rows and the first K columns of $H_s/\sqrt{K}$.

2.2.2 Statistical estimation of the Frobenius norm of an implicit matrix

For the computation of the Frobenius norm of B, we can also use a statistical estimator as first proposed in [26]. The idea is to define a random matrix $V \in \mathbb{R}^{n\times K}$ with a suitable distribution law $\mathcal{D}$ such that $\|BV\|_F$ provides a controlled approximation of $\|B\|_F$ with high probability.

Definition 2.2 *A distribution $\mathcal{D}$ over $\mathbb{R}^{n\times K}$ satisfies the (ε,δ)-concentration property if for all $B \in \mathbb{R}^{n\times n}$,*

$$\mathbb{P}(|\|BV\|_F^2 - \|B\|_F^2| \geq \varepsilon\|B\|_F^2) \leq \delta. \tag{7}$$

Two distributions $\mathcal{D}$ will be considered here.

(a) The *rescaled Rademacher distribution*. Here the entries of $V \in \mathbb{R}^{n\times K}$ are independent and identically distributed with $V_{i,j} = \pm K^{-1/2}$ with probability $1/2$. According to Theorem 13 in [2], the rescaled Rademacher distribution satisfies the (ε,δ)-concentration property for

$$K \geq 6\varepsilon^{-2}\ln(2n/\delta). \tag{8}$$

(b) The *subsampled Randomized Hadamard Transform distribution* (SRHT), first introduced in [1]. Here we assume that n is a power of 2. It is defined by $V = K^{-1/2}(RH_nD)^T \in \mathbb{R}^{n\times K}$ where

- $D \in \mathbb{R}^{n\times n}$ is a diagonal random matrix where $D_{i,i}$ are independent Rademacher random variables (i.e. $D_{i,i} = \pm 1$ with probability $1/2$),
- $H_n \in \mathbb{R}^{n\times n}$ is a Hadamard matrix of size n (see Section 2.2.1),
- $R \in \mathbb{R}^{K\times n}$ is a subset of K rows from the identity matrix of size n chosen uniformly at random and without replacement.

In other words, we randomly select K rows of H_n without replacement, and we multiply the columns by $\pm K^{-1/2}$. We can find in [37, 7] an analysis of the SRHT matrix properties. In the case where n is not a power of 2, we define the partial SRHT (P-SRHT) matrix $V \in \mathbb{R}^{n\times K}$ as the first n rows of a SRHT matrix of size $s \times K$, where $s = 2^{\lceil \log_2(n)\rceil}$ is the smallest power of 2 such that $n \leq s < 2n$. The following proposition shows that the (P-SRHT) distribution satisfies the (ε,δ)-concentration property.

Proposition 2.3 *The (P-SRHT) distribution satisfies the (ε,δ)-concentration property for*

$$K \geq 2(\varepsilon^2 - \varepsilon^3/3)^{-1}\ln(4/\delta)(1 + \sqrt{8\ln(4n/\delta)})^2. \tag{9}$$

Proof: Let $B \in \mathbb{R}^{n\times n}$. We define the square matrix $\widetilde{B}$ of size $s = 2^{\lceil \log_2(n)\rceil}$, whose first $n \times n$ diagonal block is B, and 0 elsewhere. Then we have $\|\widetilde{B}\|_F = \|B\|_F$. The rest of the proof is similar to the one of Lemma 4.10 in [7]. We consider the events $A = \{(1-\varepsilon)\|\widetilde{B}\|_F^2 \leq \|\widetilde{B}V\|_F^2 \leq (1+\varepsilon)\|\widetilde{B}\|_F^2\}$ and $E = \{\max_i \|\widetilde{B}DH_s^T e_i\|_2^2 \leq (1+\sqrt{8\ln(2s/\delta)})^2\|\widetilde{B}\|_F^2\}$, where e_i is the i-th canonical vector of $\mathbb{R}^s$. The relation $\mathbb{P}(A^c) \leq \mathbb{P}(A^c|E) + \mathbb{P}(E^c)$ holds. Thanks to Lemma 4.6 in [7] (with $t = \sqrt{8\ln(2s/\delta)}$) we have $\mathbb{P}(E^c) \leq \delta/2$. Now, using the scalar Chernoff bound (Theorem 2.2 in [37] with $k = 1$) we have

$$\begin{aligned}
\mathbb{P}(A^c|E) &= \mathbb{P}(\|\widetilde{B}V\|_F^2 \leq (1-\varepsilon)\|\widetilde{B}\|_F^2|E) + \mathbb{P}(\|\widetilde{B}V\|_F^2 \geq (1+\varepsilon)\|\widetilde{B}\|_F^2|E) \\
&\leq (e^{-\varepsilon}(1-\varepsilon)^{-1+\varepsilon})^{K(1+\sqrt{8\ln(2s/\delta)})^{-2}} + (e^{\varepsilon}(1+\varepsilon)^{-1-\varepsilon})^{K(1+\sqrt{8\ln(2s/\delta)})^{-2}} \\
&\leq 2(e^{\varepsilon}(1+\varepsilon)^{-1-\varepsilon})^{K(1+\sqrt{8\ln(2s/\delta)})^{-2}} \leq 2e^{K(-\varepsilon^2/2+\varepsilon^3/6)(1+\sqrt{8\ln(2s/\delta)})^{-2}}.
\end{aligned}$$

The condition (9) implies $\mathbb{P}(A^c|E) \leq \delta/2$, and then $\mathbb{P}(A^c) \leq \delta/2 + \delta/2 = \delta$, which ends the proof. ∎

Such statistical estimators are particularly interesting for that they provide approximations of the Frobenius norm of large matrices, with a number of columns K for V which scales as the logarithm of n, see (8) and (9). However, the concentration property (7) holds only for a given matrix B. The following proposition 2.4 extends these concentration results for any matrix B in a given subspace. The proof is inspired from the one of Theorem 6 in [15]. The essential ingredient is the existence of an ε-net for the unit ball of a finite dimensional space (see [6]).

Proposition 2.4 *Let $V \in \mathbb{R}^{n\times K}$ be a random matrix whose distribution $\mathcal{D}$ satisfies the (ε,δ)-concentration property, with $\varepsilon \leq 1$. Then, for any L-dimensional subspace of matrices $M_L \subset \mathbb{R}^{n\times n}$ and for any $C > 1$, we have*

$$\mathbb{P}\big(|\|BV\|_F^2 - \|B\|_F^2| \geq \varepsilon(C+1)/(C-1)\|B\|_F^2, \forall B \in M_L\big) \leq (9C/\varepsilon)^L\delta. \tag{10}$$

Proof: We consider the unit ball $\mathcal{B}_L = \{B \in M_L : \|B\|_F \leq 1\}$ of the subspace M_L. It is shown in [6] that for any $\widetilde{\varepsilon} > 0$, there exists a net $\mathcal{N}_L^{\widetilde{\varepsilon}} \subset \mathcal{B}_L$ of cardinality lower than $(3/\widetilde{\varepsilon})^L$ such that

$$\min_{B_{\widetilde{\varepsilon}} \in \mathcal{N}_L^{\widetilde{\varepsilon}}} \|B - B_{\widetilde{\varepsilon}}\|_F \leq \widetilde{\varepsilon}, \quad \forall B \in \mathcal{B}_L.$$

In other words, any element of the unit ball $\mathcal{B}_L$ can be approximated by an element of $\mathcal{N}_L^{\widetilde{\varepsilon}}$ with an error less than $\widetilde{\varepsilon}$. Using the (ε,δ)-concentration property and a union bound, we obtain

$$|\|B_{\widetilde{\varepsilon}}V\|_F^2 - \|B_{\widetilde{\varepsilon}}\|_F^2| \leq \varepsilon\|B_{\widetilde{\varepsilon}}\|_F^2, \quad \forall B_{\widetilde{\varepsilon}} \in \mathcal{N}_L^{\widetilde{\varepsilon}}, \tag{11}$$

with a probability at least $1-\delta(3/\widetilde{\varepsilon})^L$. We now impose the relation $\widetilde{\varepsilon}=\varepsilon/(3C)$, where $C>1$. To prove (10), it remains to show that equation (11) implies

$$|\|BV\|_F^2-\|B\|_F^2|\leq \varepsilon(C+1)/(C-1)\|B\|_F^2, \quad \forall B\in M_L. \tag{12}$$

We define $B^*\in \arg\max_{B\in\mathcal{B}_L}|\|BV\|_F^2-\|B\|_F^2|$. Let $B_{\widetilde{\varepsilon}}\in\mathcal{N}_L^{\widetilde{\varepsilon}}$ be such that $\|B^*-B_{\widetilde{\varepsilon}}\|_F\leq\widetilde{\varepsilon}$, and $B_{\widetilde{\varepsilon}}^*=\arg\min_{B\in\mathrm{span}(B_{\widetilde{\varepsilon}})}\|B^*-B\|_F$. Then we have $\|B^*-B_{\widetilde{\varepsilon}}^*\|_F^2=\|B^*\|_F^2-\|B_{\widetilde{\varepsilon}}^*\|_F^2\leq\widetilde{\varepsilon}^2$ and $\langle B^*-B_{\widetilde{\varepsilon}}^*,B_{\widetilde{\varepsilon}}^*\rangle=0$, where $\langle\cdot,\cdot\rangle$ is the inner product associated to the Frobenius norm $\|\cdot\|_F$. We have

$$\begin{aligned}
\eta &:= |\|B^*V\|_F^2-\|B^*\|_F^2| = |\|(B^*-B_{\widetilde{\varepsilon}}^*)V+B_{\widetilde{\varepsilon}}^*V\|_F^2-\|B^*-B_{\widetilde{\varepsilon}}^*+B_{\widetilde{\varepsilon}}^*\|_F^2| \\
&= |\|(B^*-B_{\widetilde{\varepsilon}}^*)V\|_F^2+2\langle(B^*-B_{\widetilde{\varepsilon}}^*)V,B_{\widetilde{\varepsilon}}^*V\rangle+\|B_{\widetilde{\varepsilon}}^*V\|_F^2-\|B^*-B_{\widetilde{\varepsilon}}^*\|_F^2-\|B_{\widetilde{\varepsilon}}^*\|_F^2| \\
&\leq |\|(B^*-B_{\widetilde{\varepsilon}}^*)V\|_F^2-\|B^*-B_{\widetilde{\varepsilon}}^*\|_F^2|+|\|B_{\widetilde{\varepsilon}}^*V\|_F^2-\|B_{\widetilde{\varepsilon}}^*\|_F^2|+2\|(B^*-B_{\widetilde{\varepsilon}}^*)V\|_F\|B_{\widetilde{\varepsilon}}^*V\|_F.
\end{aligned}$$

We now have to bound the three terms in the previous expression. Firstly, since $(B^*-B_{\widetilde{\varepsilon}}^*)/\|B^*-B_{\widetilde{\varepsilon}}^*\|_F\in\mathcal{B}_L$, the relation $|\|(B^*-B_{\widetilde{\varepsilon}}^*)V\|_F^2-\|B^*-B_{\widetilde{\varepsilon}}^*\|_F^2|\leq\|B^*-B_{\widetilde{\varepsilon}}^*\|_F^2\eta\leq\widetilde{\varepsilon}^2\eta$ holds. Secondly, (11) gives $|\|B_{\widetilde{\varepsilon}}^*V\|_F^2-\|B_{\widetilde{\varepsilon}}^*\|_F^2|\leq\varepsilon\|B_{\widetilde{\varepsilon}}^*\|_F^2\leq\varepsilon$. Thirdly, by definition of η, we can write $\|(B^*-B_{\widetilde{\varepsilon}}^*)V\|_F^2\leq(1+\eta)\|B^*-B_{\widetilde{\varepsilon}}^*\|_F^2\leq\widetilde{\varepsilon}^2(1+\eta)$ and $\|B_{\widetilde{\varepsilon}}^*V\|_F^2\leq(1+\varepsilon)\|B_{\widetilde{\varepsilon}}^*\|_F^2\leq 1+\varepsilon$, so that we obtain $2\|(B^*-B_{\widetilde{\varepsilon}}^*)V\|_F\|B_{\widetilde{\varepsilon}}^*V\|_F\leq 2\widetilde{\varepsilon}\sqrt{1+\varepsilon}\sqrt{1+\eta}$. Finally, from (11), we obtain

$$\eta\leq\widetilde{\varepsilon}^2\eta+\varepsilon+2\widetilde{\varepsilon}\sqrt{1+\varepsilon}\sqrt{1+\eta} \tag{13}$$

Since $\varepsilon\leq 1$, we have $\widetilde{\varepsilon}=\varepsilon/(3C)<1/3$. Then $\widetilde{\varepsilon}^2\leq\widetilde{\varepsilon}$ and $\sqrt{1+\varepsilon}\leq 3/2$, so that (13) implies

$$\eta\leq\widetilde{\varepsilon}\eta+\varepsilon+3\widetilde{\varepsilon}\sqrt{1+\eta}\leq\widetilde{\varepsilon}\eta+\varepsilon+3\widetilde{\varepsilon}(1+\eta/2)\leq 3\widetilde{\varepsilon}\eta+\varepsilon+3\widetilde{\varepsilon},$$

and then $\eta\leq(\varepsilon+3\widetilde{\varepsilon})/(1-3\widetilde{\varepsilon})\leq\varepsilon(C+1)/(C-1)$. By definition of η, we can write $|\|BV\|_F^2-\|B\|_F^2|\leq\varepsilon(C+1)/(C-1)$ for any $B\in\mathcal{B}_L$, that implies (12). ∎

Proposition 2.5 *Let $\xi\in\Xi$, and let $P_m(\xi)\in Y_m$ be defined by (5) where $V\in\mathbb{R}^{n\times K}$ is a realization of a rescaled Rademacher matrix with*

$$K\geq 6\varepsilon^{-2}\ln(2n(9C/\varepsilon)^{m+1}/\delta), \tag{14}$$

or a realization of a P-SRHT matrix with

$$K\geq 2(\varepsilon^2-\varepsilon^3/3)^{-1}\ln(4(9C/\varepsilon)^{m+1}/\delta)(1+\sqrt{8\ln(4n(9C/\varepsilon)^{m+1}/\delta)})^2 \tag{15}$$

for some $\delta>0$, $\varepsilon\leq 1$ and $C>1$. Assuming $\varepsilon'=\varepsilon(C+1)/(C-1)<1$,

$$\|I-P_m(\xi)A(\xi)\|_F\leq\sqrt{\frac{1+\varepsilon'}{1-\varepsilon'}}\min_{P\in Y_m}\|I-PA(\xi)\|_F \tag{16}$$

holds with a probability higher than $1-\delta$.

Proof: Let us introduce the subspace $M_{m+1} = Y_m A(\xi) + \text{span}(I)$ of dimension less than $m+1$, such that $\{I - PA(\xi) : P \in Y_m\} \subset M_{m+1}$. Then, we note that with the conditions (14) or (15), the distribution law $\mathcal{D}$ of the random matrix V satisfies the $(\varepsilon, \delta(\varepsilon/(9C))^{m+1})$-concentration property. Thanks to Proposition 2.4, the probability that

$$|\|(I - PA(\xi))V\|_F^2 - \|I - PA(\xi)\|_F^2| \leq \varepsilon' \|I - PA(\xi)\|_F^2$$

holds for any $P \in Y_m$ is higher than $1 - \delta$. Then, by definition of $P_m(\xi)$ (5), we have with a probability at least $1 - \delta$ that for any $P \in Y_m$, it holds

$$\begin{aligned}\|I - P_m(\xi)A(\xi)\|_F &\leq \frac{1}{\sqrt{1-\varepsilon'}} \|(I - P_m(\xi)A(\xi))V\|_F, \\ &\leq \frac{1}{\sqrt{1-\varepsilon'}} \|(I - PA(\xi))V\|_F \leq \frac{\sqrt{1+\varepsilon'}}{\sqrt{1-\varepsilon'}} \|I - PA(\xi)\|_F.\end{aligned}$$

Then, taking the minimum over $P \in Y_m$, we obtain (16). ∎

Similarly to Proposition 2.1, we obtain the following properties for $P_m(\xi)A(\xi)$, with $P_m(\xi)$ the solution of (5).

Proposition 2.6 *Under the assumptions of Proposition 2.5, the inequalities*

$$(1 - \alpha_m(\xi))^2 (1 - \varepsilon')^{-1} \leq \|(I - P_m(\xi)A(\xi))V\|_F^2 \leq n\left(1 - (1 - \varepsilon')\alpha_m^2(\xi)\right) \tag{17}$$

and

$$\kappa(P_m(\xi)A(\xi)) \leq \alpha_m(\xi)^{-1} \sqrt{n(1-\varepsilon')^{-1} - (n-1)\alpha_m^2(\xi)} \tag{18}$$

hold with probability $1 - \delta$*, where* $\alpha_m(\xi)$ *is the lowest singular value of* $P_m(\xi)A(\xi)$*.*

Proof: The optimality condition for $P_m(\xi)$ yields $\|(I - P_m(\xi)A(\xi))V\|_F^2 = \|V\|_F^2 - \|P_m(\xi)A(\xi)V\|_F^2$. Since $P_m(\xi)A(\xi) \in M_{m+1}$ (where M_{m+1} is the subspace introduced in the proof of Proposition (2.5)), we have

$$\|P_m(\xi)A(\xi)V\|_F^2 \geq (1 - \varepsilon')\|P_m(\xi)A(\xi)\|_F^2 \tag{19}$$

with a probability higher than $1 - \delta$. Using $\|V\|_F^2 = n$ (which is satisfies for any realization of the rescaled Rademacher or the P-SRHT distribution), we obtain $\|(I - P_m(\xi)A(\xi))V\|_F^2 \leq n - (1 - \varepsilon')\|P_m(\xi)A(\xi)\|_F^2$ with a probability higher than $1 - \delta$. Then, $\|P_m(\xi)A(\xi)\|_F^2 \geq n\alpha_m(\xi)^2$ yields the right inequality of (17). Following the proof of Lemma 2.6 in [24], we have $(1 - \alpha_m(\xi)^2) \leq \|I - P_m(\xi)A(\xi)\|_F^2$. Together with (19), it yields the left inequality of (17). Furthermore, with probability $1 - \delta$, we have $n - (1 - \varepsilon')\|P_m(\xi)A(\xi)\|_F^2 \geq 0$. Since

the square of the Frobenius norm of matrix $P_m(\xi)A(\xi)$ is the sum of squares of its singular values, we deduce

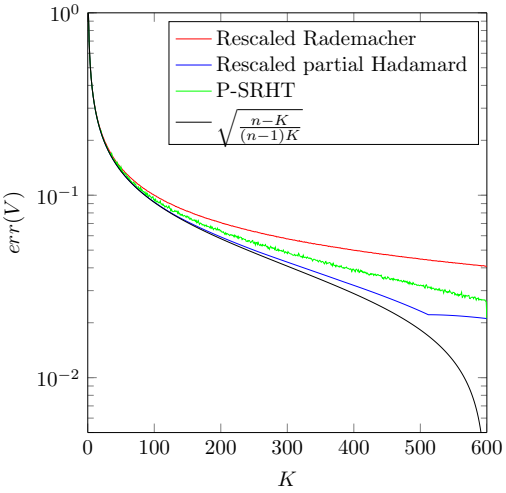
$$(n-1)\alpha_m(\xi)^2 + \beta_m(\xi)^2 \leq \|P_m(\xi)A(\xi)\|_F^2 \leq n(1-\varepsilon')^{-1}$$

with a probability higher than $1-\delta$, where $\beta_m(\xi)$ is the largest singular value of $P_m(\xi)A(\xi)$. Then (18) follows from the definition of $\kappa(P_m(\xi)A(\xi)) = \beta_m(\xi)/\alpha_m(\xi)$. ∎

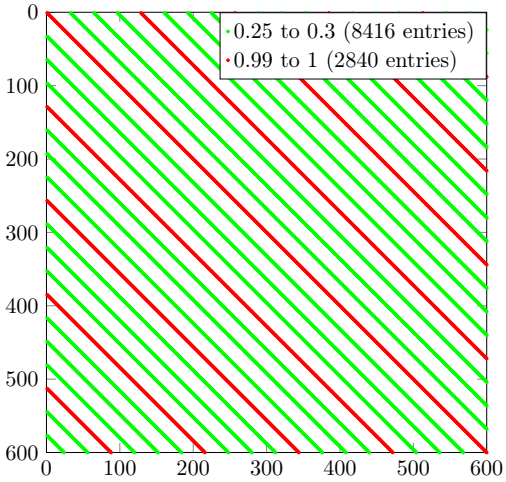
2.2.3 Comparison and comments

We have presented different possibilities for the definition of V. The rescaled partial Hadamard matrices introduced in section 2.2.1 have the advantage that the error $err(V)$ is close to the theoretical bound $\sqrt{(n-K)/((n-1)K)}$, see Figure 1(a) (note that the rows of V have unit norm). Furthermore, an interesting property is that VV^T has a structured pattern (see Figure 1(b)). As noticed in [3], when $K = 2^q$ the matrix VV^T have non-zero entries only on the 2^{qk}-th upper and lower diagonals, with $k \geq 0$. As a consequence, the error on the estimation of $\|B\|_F$ will be induced only by the non-zero off-diagonal entries of B that occupy the 2^{qk}-th upper and lower diagonals, with $k \geq 1$. If the entries of B vanish away from the diagonal, the Frobenius norm is expected to be accurately estimated. Note that the P-SRHT matrices can be interpreted as a "randomized version" of the rescaled partial Hadamard matrices, and Figure 1(a) shows that the error $err(V)$ associated to the P-SRHT matrix behaves almost like the rescaled partial Hadamard matrix. Also, P-SRHT matrices yield a structured pattern for VV^T, see Figure 1(c). The rescaled Rademacher matrices give higher errors $err(V)$ and yield matrices VV^T with no specific patterns, see Figure 1(d).

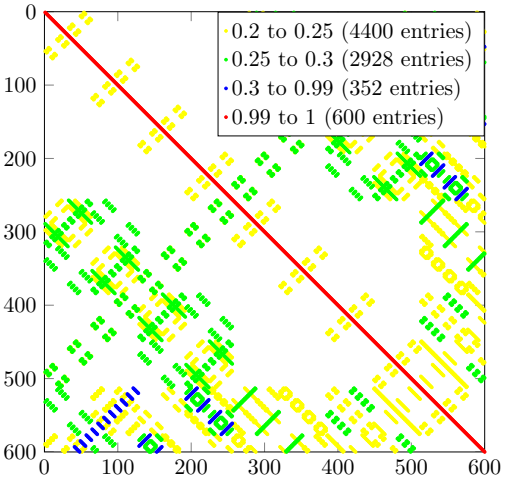
The advantage of using rescaled Rademacher matrices or P-SRHT matrices is that the quality of the resulting projection $P_m(\xi)$ can be controlled with high probability, provided that V has a sufficiently large number of rows K (see Proposition 2.5). Table 1 shows the theoretical value for K in order to obtain the quasi-optimality result (16) with $\sqrt{(1+\varepsilon')/(1-\varepsilon')} = 10$ and $\delta = 0.1\%$. It can be observed that K grows very slowly with the matrix size n. Also, K depends on m linearly for the rescaled Rademacher matrices and quadratically for the P-SRHT matrices (see equations (14) and (15)). However, these theoretical bounds for K are very pessimistic, especially for the P-SRHT matrices. In practice, it can be observed that a very small value for K may provide very good results (see Section 5). Also, it is worth mentioning that our numerical experiments do not reveal significant differences between the rescaled partial Hadamard, the rescaled Rademacher and the P-SRHT matrices.



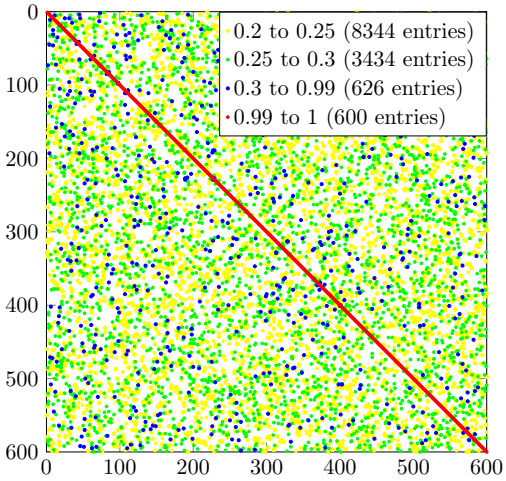
(a) $err(V)$ as function of K.



(b) Distribution of the entries of VV^T (in absolute value) where V is the rescaled partial Hadamard matrix with $K = 100$.



(c) Distribution of the entries of VV^T (in absolute value) where V is a sample of the P-SRHT matrix with $K = 100$.



(d) Distribution of the entries of VV^T (in absolute value) where V is a sample of the rescaled Rademacher matrix with $K = 100$.

Figure 1: Comparison between the rescaled partial Hadamard, the rescaled Rademacher and the P-SRHT matrix for the definition of matrix V, with $n = 600$.

2.3 Ensuring the invertibility of the preconditioner for positive definite matrix

Here, we propose a modification of the interpolation which ensures that $P_m(\xi)$ is invertible when $A(\xi)$ is positive definite.

Since $A(\xi_i)$ is positive definite, $P_i = A(\xi_i)^{-1}$ is positive definite. We introduce the vectors

(a) Rescaled Rademacher distribution.

	$m = 2$	$m = 5$	$m = 10$	$m = 20$	$m = 50$
$n = 10^4$	239	363	567	972	2 185
$n = 10^6$	270	395	599	1 005	2 219
$n = 10^8$	301	427	632	1 038	2 253

(b) P-SRHT distribution.

	$m = 2$	$m = 5$	$m = 10$	$m = 20$	$m = 50$
$n = 10^4$	27 059	63 298	155 129	455 851	2 286 645
$n = 10^6$	30 597	69 129	164 750	473 011	2 326 301
$n = 10^8$	34 112	74 929	174 333	490 126	2 365 914

Table 1: Theoretical number of columns K for the random matrix V in order to ensure (16), with $\sqrt{(1+\varepsilon')/(1-\varepsilon')} = 10$ and $\delta = 0.1\%$. The constant C has been chosen in order to minimize K.

$\gamma^- \in \mathbb{R}^m$ and $\gamma^+ \in \mathbb{R}^m$ whose components

$$\gamma_i^- = \inf_{w \in \mathbb{R}^n} \frac{\langle P_i w, w \rangle}{\|w\|^2} > 0 \quad \text{and} \quad \gamma_i^+ = \sup_{w \in \mathbb{R}^n} \frac{\langle P_i w, w \rangle}{\|w\|^2} < \infty$$

correspond respectively to the lowest and highest eigenvalues of the symmetric part of P_i. Then, for any $P = \sum_{i=1}^m \lambda_i P_i \in Y_m$,

$$\inf_{w \in \mathbb{R}^n} \frac{\langle Pw, w \rangle}{\|w\|^2} \geq \langle \lambda^+, \gamma^- \rangle - \langle \lambda^-, \gamma^+ \rangle, \tag{20}$$

where $\lambda^+ \geq 0$ and $\lambda^- \geq 0$ are respectively the positive and negative parts of $\lambda = \lambda^+ - \lambda^- \in \mathbb{R}^m$. As a consequence, if the right hand side of (20) is strictly positive, then P is invertible. Furthermore, we have $\|P\| \leq \langle \lambda^+ + \lambda^-, C \rangle$, where $C \in \mathbb{R}^m$ is the vector of component $C_i = \|P_i\|$, where $\|P_i\|$ denotes the operator norm of P_i. If we assume that $\langle \lambda^+, \gamma^- \rangle - \langle \lambda^-, \gamma^+ \rangle > 0$, the condition number of P satisfies

$$\kappa(P) = \|P\| \, \|P^{-1}\| \leq \|P\| \left(\inf_{w \in \mathbb{R}^n} \frac{\langle Pw, w \rangle}{\|w\|^2} \right)^{-1} \leq \frac{\langle \lambda^+ + \lambda^-, C \rangle}{\langle \lambda^+, \gamma^- \rangle - \langle \lambda^-, \gamma^+ \rangle}.$$

It is then possible to bound $\kappa(P)$ by $\bar{\kappa}$ by imposing

$$\langle \lambda^+ + \lambda^-, C \rangle \leq \bar{\kappa} (\langle \lambda^+, \gamma^- \rangle - \langle \lambda^-, \gamma^+ \rangle),$$

which is a linear inequality constraint on λ^+ and λ^-. We introduce two convex subsets of

Y_m defined by

$$Y_m^{\bar\kappa} = \left\{ \sum_{i=1}^m \lambda_i^+ P_i - \sum_{i=1}^m \lambda_i^- P_i \ : \ \begin{array}{c} \lambda_i^+ \geq 0,\ \lambda_i^- \geq 0 \\ \langle \lambda^+, \gamma^- \rangle - \langle \lambda^-, \gamma^+ \rangle \geq 0 \\ \langle \lambda^+, \bar\kappa\gamma^- - C \rangle - \langle \lambda^-, \bar\kappa\gamma^+ + C \rangle \geq 0 \end{array} \right\},$$

$$Y_m^+ = \left\{ \sum_{i=1}^m \lambda_i P_i \ : \ \lambda_i \geq 0 \right\}.$$

From (20), we have that any nonzero element of Y_m^+ is invertible, while any nonzero element of $Y_m^{\bar\kappa}$ is invertible and has a condition number lower than $\bar\kappa$. Under the condition $\bar\kappa \geq \max_i C_i/\gamma_i^-$, we have

$$Y_m^+ \subset Y_m^{\bar\kappa} \subset Y_m. \tag{22}$$

Then definitions (2) and (5) for the approximation $P_m(\xi)$ can be replaced respectively by

$$P_m(\xi) = \underset{P \in Y_m^+ \text{ or } Y_m^{\bar\kappa}}{\operatorname{argmin}} \|I - PA(\xi)\|_F, \tag{23a}$$

$$P_m(\xi) = \underset{P \in Y_m^+ \text{ or } Y_m^{\bar\kappa}}{\operatorname{argmin}} \|(I - PA(\xi))V\|_F, \tag{23b}$$

which are quadratic optimization problems with linear inequality constraints. Furthermore, since $P_i \in Y_m^+$ for all i, all the resulting projections $P_m(\xi)$ interpolate $A(\xi)^{-1}$ at the points $\xi_1, \ldots, \xi_m$.

The following proposition shows that properties (3) and (4) still hold for the preconditioned operator.

Proposition 2.7 *The solution $P_m(\xi)$ of* (23a) *is such that $P_m(\xi)A(\xi)$ satisfies* (3) *and* (4)*. Also, under the assumptions of Proposition 2.5, the solution $P_m(\xi)$ of* (23b) *is such that $P_m(\xi)A(\xi)$ satisfies* (17) *and* (18) *with a probability higher than $1 - \delta$.*

Proof: Since Y_m^+ (or $Y_m^{\bar\kappa}$) is a closed and convex positive cone, the solution $P_m(\xi)$ of (23a) is such that $trace((I - P_m(\xi)A(\xi))^T(P_m(\xi) - P)A(\xi)) \geq 0$ for all $P \in Y_m^+$ (or $Y_m^{\bar\kappa}$). Taking $P = 2P_m(\xi)$ and $P = 0$, we obtain that $\text{trace}((I - P_m(\xi)A(\xi))^T P_m(\xi)A(\xi)) = 0$, which implies $\|P_m(\xi)A(\xi)\|_F^2 = \text{trace}(P_m(\xi)A(\xi))$. We refer to the proof of Lemma 2.6 and Theorem 3.2 in [24] to deduce (3) and (4). Using the same arguments, we prove that the solution $P_m(\xi)$ of (23b) satisfies $\|P_m(\xi)A(\xi)V\|_F^2 = \text{trace}(V^T P_m(\xi)A(\xi)V)$, and then that (17) and (18) hold with a probability higher than $1 - \delta$. ∎

2.4 Practical computation of the projection

Here, we detail how to efficiently compute $M^V(\xi)$ and $S^V(\xi)$ given in equation (6) in a multi-query context, i.e. for several different values of ξ. The same methodology can be applied for computing $M(\xi)$ and $S(\xi)$. We assume that the operator $A(\xi)$ has an affine expansion of the form

$$A(\xi) = \sum_{k=1}^{m_A} \Phi_k(\xi) A_k, \tag{24}$$

where the A_k are matrices in $\mathbb{R}^{n\times n}$ and the $\Phi_k : \Xi \to \mathbb{R}$ are real-valued functions. Then $M^V(\xi)$ and $S^V(\xi)$ also have the affine expansions

$$M^V_{i,j}(\xi) = \sum_{k=1}^{m_A}\sum_{l=1}^{m_A} \Phi_k(\xi)\Phi_l(\xi)\ \mathrm{trace}(V^T A_k^T P_i^T P_j A_l V), \tag{25a}$$

$$S^V_i(\xi) = \sum_{k=1}^{m_A} \Phi_k(\xi)\ \mathrm{trace}(V^T P_i A_k V), \tag{25b}$$

respectively. Computing the multiple terms of these expansions would require many computations of traces of implicit matrices and also, it would require the computation of the affine expansion of $A(\xi)$. Here, we use the methodology introduced in [9] for obtaining affine decompositions with a lower number of terms. These decompositions only require the knowledge of functions Φ_k in the affine decomposition (24), and evaluations of $M^V_{i,j}(\xi)$ and $S^V_i(\xi)$ (that means evaluations of $A(\xi)$) at some selected points. We briefly recall this methodology.

Suppose that $g : \Xi \to X$, with X a vector space, has an affine decomposition $g(\xi) = \sum_{k=1}^{m} \zeta_k(\xi) g_k$, with $\zeta_k : \Xi \to \mathbb{R}$ and $g_k \in X$. We first compute an interpolation of $\zeta(\xi) = (\zeta_1(\xi), \ldots, \zeta_m(\xi))$ under the form $\zeta(\xi) = \sum_{k=1}^{m_g} \Psi_k(\xi)\zeta(\xi^*_k)$, with $m_g \leq m$, where $\xi^*_1, \ldots, \xi^*_{m_g}$ are interpolation points and $\Psi_1(\xi), \ldots, \Psi_{m_g}(\xi)$ the associated interpolation functions. Such an interpolation can be computed with the Empirical Interpolation Method [29] described in Algorithm 2. Then, we obtain an affine decomposition $g(\xi) = \sum_{k=1}^{m_g} \Psi_k(\xi) g(\xi^*_k)$ which can be computed from evaluations of g at interpolation points ξ^*_k.

Applying the above procedure to both $M^V(\xi)$ and $S^V(\xi)$, we obtain

$$M^V(\xi) \approx \sum_{k=1}^{m_M} \Psi_k(\xi)\ M^V(\xi^*_k), \quad S^V(\xi) \approx \sum_{k=1}^{m_S} \widetilde{\Psi}_k(\xi)\ S^V(\widetilde{\xi}^*_k). \tag{26}$$

The first (so-called *offline*) step consists in computing the interpolation functions $\Psi_k(\xi)$ and $\widetilde{\Psi}_k(\xi)$ and associated interpolation points ξ^*_k and $\widetilde{\xi}^*_k$ using Algorithm 2 with input $\{\Phi_i\Phi_j\}_{1\leq i,j\leq m_A}$ and $\{\Phi_i\}_{1\leq i\leq m_A}$ respectively, and then in computing matrices $M^V(\xi^*_k)$ and vectors $S^V(\widetilde{\xi}^*_k)$ using Algorithm 1. The second (so-called *online*) step simply consists in computing the matrix $M^V(\xi)$ and the vector $S^V(\xi)$ for a given value of ξ using (26).

Algorithm 2 Empirical Interpolation Method (EIM).

Require: $(\zeta_1(\cdot), \ldots, \zeta_m(\cdot))$

Ensure: $\Psi_1(\cdot), \ldots, \Psi_k(\cdot)$ and $\xi_1^*, \ldots, \xi_k^*$

1: Define $R_1(i, \xi) = \zeta_i(\xi)$ for all i, ξ
2: Initialize $e = 1$, $k = 0$
3: **while** $e \geq tolerance$ (in practice the machine precision) **do**
4: $\quad k = k + 1$
5: $\quad$ Find $(i_k^*, \xi_k^*) \in \underset{i,\xi}{\operatorname{argmax}} |R_k(i, \xi)|$
6: $\quad$ Set the error to $e = |R_k(i_k^*, \xi_k^*)|$
7: $\quad$ Actualize $R_{k+1}(i, \xi) = R_k(i, \xi) - R_k(i, \xi_k^*) R_k(i_k^*, \xi) / R_k(i_k^*, \xi_k^*)$ for all i, ξ
8: **end while**
9: Fill in the k-by-k matrix Q : $Q_{i,j} = \zeta_{i_i^*}(\xi_j^*)$ for all $1 \leq i, j \leq k$
10: Compute $\Psi_i(\xi) = \sum_{j=1}^k (Q^{-1})_{i,j} \zeta_{i_j^*}(\xi)$ for all ξ and $1 \leq i \leq k$

3 Preconditioners for projection-based model reduction

We consider a parameter-dependent linear equation

$$A(\xi)u(\xi) = b(\xi), \tag{27}$$

with $A(\xi) \in \mathbb{R}^{n \times n}$ and $b(\xi) = \mathbb{R}^n$. Projection-based model reduction consists in projecting the solution $u(\xi)$ onto a well chosen approximation space $X_r \subset X := \mathbb{R}^n$ of low dimension $r \ll n$. Such projections are usually defined by imposing the residual of (27) to be orthogonal to a so-called test space of dimension r. The quality of the projection on X_r depends on the choice of the test space. The latter can be defined as the approximation space itself X_r, thus yielding the classical Galerkin projection. However when the operator $A(\xi)$ is ill-conditioned (for example when $A(\xi)$ corresponds to the discretization of non coercive or weakly coercive operators), this choice may lead to projections that are far from optimal. Choosing the test space as $\{R_X^{-1} A(\xi) v_r : v_r \in X_r\}$, where $R_X^{-1} A(\xi)$ is called the "supremizer operator" (see e.g. [35]), corresponds to a minimal residual approach, which may also results in projections that are far from optimal. In this section, we show how the preconditioner $P_m(\xi)$ can be used for the definition of the test space. We also show how it can improve the quality of residual-based error estimates, which is a key ingredient for the construction of suitable approximation space X_r in the context of the Reduced Basis method.

X is endowed with the norm $\|\cdot\|_X$ defined by $\|\cdot\|_X^2 = \langle R_X \cdot, \cdot \rangle$, where R_X is a symmetric positive definite matrix and $\langle \cdot, \cdot \rangle$ is the canonical inner product of $\mathbb{R}^n$. We also introduce the

dual norm $\|\cdot\|_{X'} = \|R_X^{-1}\cdot\|_X$ such that for any $v, w \in X$ we have $|\langle v, w\rangle| \leq \|v\|_X \, \|w\|_{X'}$.

3.1 Projection of the solution on a given reduced subspace

Here, we suppose that the approximation space X_r has been computed by some model order reduction method. The best approximation of $u(\xi)$ on X_r is $u_r^*(\xi) = \arg\min_{v\in X_r} \|u(\xi) - v\|_X$ and is characterized by the orthogonality condition

$$\langle u_r^*(\xi) - u(\xi), R_X v_r\rangle = 0, \quad \forall v_r \in X_r, \tag{28}$$

or equivalently by the Petrov-Galerkin orthogonality condition

$$\langle A(\xi)u_r^*(\xi) - b(\xi), A^{-T}(\xi)R_X v_r\rangle = 0, \quad \forall v_r \in X_r. \tag{29}$$

Obviously the computation of test functions $A^{-T}(\xi)R_X v_r$ for basis functions v_r of X_r is prohibitive. By replacing $A(\xi)^{-1}$ by $P_m(\xi)$, we obtain the feasible Petrov-Galerkin formulation

$$\langle A(\xi)u_r(\xi) - b(\xi), P_m^T(\xi)R_X v_r\rangle = 0, \quad \forall v_r \in X_r. \tag{30}$$

Denoting by $U \in \mathbb{R}^{n\times r}$ a matrix whose range is X_r, the solution of (30) is $u_r(\xi) = Ua(\xi)$ where the vector $a(\xi) \in \mathbb{R}^r$ is the solution of

$$\big(U^T R_X P_m(\xi) A(\xi) U\big)a(\xi) = \big(U^T R_X P_m(\xi) b(\xi)\big).$$

Note that (30) corresponds to the standard Galerkin projection when replacing $P_m(\xi)$ by R_X^{-1}. Indeed, the orthogonality condition (30) becomes $\langle A(\xi)u_r(\xi) - b(\xi), v_r\rangle = 0$ for all $v_r \in X_r$.

Remark 3.1 *Here, the preconditioner $P_m(\xi)$ is used for the definition of the parameter-dependent test space $\{P_m^T(\xi)R_X v_r : v_r \in X_r\}$ which defines the Petrov-Galerkin projection* (30)*. However, $P_m(\xi)$ could also be used to construct preconditioners for the solution of the linear system $\big(U^T A(\xi)U\big)a(\xi) = \big(U^T b(\xi)\big)$ corresponding to the Galerkin projection on X_r. Following the idea proposed in [19], such preconditoner can take the form $(U^T P_m(\xi)U)$, thus yielding the preconditioned reduced linear system*

$$\big(U^T P_m(\xi)U\big)\big(U^T A(\xi)U\big)a(\xi) = \big(U^T P_m(\xi)U\big)\big(U^T b(\xi)\big).$$

Such preconditioning strategy can be used to accelerate the solution of the reduced system of equations when using iterative methods. However, and contrarily to (30)*, this strategy does not change the definition of $u_r(\xi)$, which is the standard Galerkin projection.*

We give now a quasi-optimality result for the approximation $u_r(\xi)$. This analysis relies on the notion of δ-proximality introduced in [13].

Proposition 3.2 *Let $\delta_{r,m}(\xi) \in [0,1]$ be defined by*

$$\delta_{r,m}(\xi) = \max_{v_r \in X_r} \min_{w_r \in X_r} \frac{\|v_r - R_X^{-1}(P_m(\xi)A(\xi))^T R_X w_r\|_X}{\|v_r\|_X}. \tag{31}$$

The solutions $u_r^*(\xi) \in X_r$ and $u_r(\xi) \in X_r$ of (28) and (30) satisfy

$$\|u_r^*(\xi) - u_r(\xi)\|_X \leq \delta_{r,m}(\xi)\|u(\xi) - u_r(\xi)\|_X. \tag{32}$$

Moreover, if $\delta_{r,m}(\xi) < 1$ holds, then

$$\|u(\xi) - u_r(\xi)\|_X \leq (1 - \delta_{r,m}(\xi)^2)^{-1/2}\|u(\xi) - u_r^*(\xi)\|_X. \tag{33}$$

Proof: The orthogonality condition (28) yields

$$\langle u_r^*(\xi) - u_r(\xi), R_X v_r\rangle = \langle u(\xi) - u_r(\xi), R_X v_r\rangle = \langle b(\xi) - A(\xi)u_r(\xi), A^{-T}(\xi)R_X v_r\rangle$$

for all $v_r \in X_r$. Using (30), we have that for any $w_r \in X_r$,

$$\begin{aligned}
\langle u_r^*(\xi) - u_r(\xi), R_X v_r\rangle &= \langle b(\xi) - A(\xi)u_r(\xi), A^{-T}(\xi)R_X v_r - P_m(\xi)^T R_X w_r\rangle, \\
&= \langle u(\xi) - u_r(\xi), R_X v_r - (P_m(\xi)A(\xi))^T R_X w_r\rangle, \\
&\leq \|u(\xi) - u_r(\xi)\|_X \ \|R_X v_r - (P_m(\xi)A(\xi))^T R_X w_r\|_{X'} \\
&= \|u(\xi) - u_r(\xi)\|_X \ \|v_r - R_X^{-1}(P_m(\xi)A(\xi))^T R_X w_r\|_X.
\end{aligned}$$

Taking the infimum over $w_r \in X_r$ and by the definition of $\delta_{r,m}(\xi)$, we obtain

$$\langle u_r^*(\xi) - u_r(\xi), R_X v_r\rangle \leq \delta_{r,m}(\xi)\|u(\xi) - u_r(\xi)\|_X \ \|v_r\|_X.$$

Then, noting that $u_r^*(\xi) - u_r(\xi) \in X_r$, we obtain

$$\|u_r^*(\xi) - u_r(\xi)\|_X = \sup_{v_r \in X_r} \frac{\langle u_r^*(\xi) - u_r(\xi), R_X v_r\rangle}{\|v_r\|_X} \leq \delta_{r,m}(\xi)\|u(\xi) - u_r(\xi)\|_X,$$

that is (32). Finally, using orthogonality condition (28), we have that

$$\begin{aligned}
\|u(\xi) - u_r(\xi)\|_X^2 &= \|u(\xi) - u_r^*(\xi)\|_X^2 + \|u_r^*(\xi) - u_r(\xi)\|_X^2, \\
&\leq \|u(\xi) - u_r^*(\xi)\|_X^2 + \delta_{r,m}(\xi)^2\|u(\xi) - u_r(\xi)\|_X^2,
\end{aligned}$$

from which we deduce (33) when $\delta_{r,m}(\xi) < 1$. ∎

An immediate consequence of Proposition 3.2 is that when $\delta_{r,m}(\xi) = 0$, the Petrov-Galerkin projection $u_r(\xi)$ coincides with the orthogonal projection $u_r^*(\xi)$. Following [14], we show in the following proposition that $\delta_{r,m}(\xi)$ can be computed by solving an eigenvalue problem of size r.

Proposition 3.3 *We have $\delta_{r,m}(\xi) = \sqrt{1-\gamma}$, where γ is the lowest eigenvalue of the generalized eigenvalue problem $Cx = \gamma Dx$, with*

$$C = U^T B(B^T R_X^{-1} B)^{-1} B^T U \in \mathbb{R}^{r\times r},$$
$$D = U^T R_X U \in \mathbb{R}^{r\times r},$$

where $B = (P_m(\xi)A(\xi))^T R_X U \in \mathbb{R}^{n\times r}$ and where $U \in \mathbb{R}^{n\times r}$ is a matrix whose range is X_r.

Proof: Since the range of U is X_r, we have

$$\delta_{r,m}(\xi)^2 = \max_{a\in\mathbb{R}^r} \min_{b\in\mathbb{R}^r} \frac{\|Ua - R_X^{-1}Bb\|_X^2}{\|Ua\|_X^2}.$$

For any $a \in \mathbb{R}^r$, the minimizer b^* of $\|Ua - R_X^{-1}Bb\|_X^2$ over $b \in \mathbb{R}^r$ is given by $b^* = (B^T R_X^{-1} B)^{-1} B^T Ua$. Therefore, we have $\|Ua - R_X^{-1}Bb^*\|_X^2 = \|Ua\|_X^2 - \langle Ua, Bb^*\rangle$, and

$$\delta_{r,m}^2(\xi) = 1 - \inf_{a\in\mathbb{R}^r} \frac{\langle U^T B(B^T R_X^{-1} B)^{-1} B^T Ua, a\rangle}{\langle U^T R_X Ua, a\rangle},$$

which concludes the proof. ∎

3.2 Greedy construction of the solution reduced subspace

Following the idea of the Reduced Basis method [36, 38], a sequence of nested approximation spaces $\{X_r\}_{r\geq 1}$ in X can be constructed by a greedy algorithm such that $X_{r+1} = X_r + \text{span}(u(\xi_{r+1}^{RB}))$, where ξ_{r+1}^{RB} is a point where the error of approximation of $u(\xi)$ in X_r is maximal. An ideal greedy algorithm using the best approximation in X_r and an exact evaluation of the projection error is such that

$$u_r^*(\xi) \text{ is the orthogonal projection of } u(\xi) \text{ on } X_r \text{ defined by (28)}, \tag{34a}$$
$$\xi_{r+1}^{RB} \in \operatorname*{argmax}_{\xi\in\Xi} \|u(\xi) - u_r^*(\xi)\|_X. \tag{34b}$$

This ideal greedy algorithm is not feasible in practice since $u(\xi)$ is not known. Therefore, we rather rely on a feasible weak greedy algorithm such that

$$u_r(\xi) \text{ is the Petrov-Galerkin projection of } u(\xi) \text{ on } X_r \text{ defined by (30)}, \tag{35a}$$
$$\xi_{r+1}^{RB} \in \operatorname*{argmax}_{\xi\in\Xi} \|P_m(\xi)(A(\xi)u_r(\xi) - b(\xi))\|_X. \tag{35b}$$

Assume that

$$\underline{\alpha}_m \|u(\xi) - u_r(\xi)\|_X \leq \|P_m(\xi)(A(\xi)u_r(\xi) - b(\xi))\|_X \leq \bar{\beta}_m \|u(\xi) - u_r(\xi)\|_X$$

holds with $\underline{\alpha}_m = \inf_{\xi\in\Xi} \alpha_m(\xi) > 0$ and $\bar{\beta}_m = \sup_{\xi\in\Xi} \beta_m(\xi) < \infty$, where $\alpha_m(\xi)$ and $\beta_m(\xi)$ are respectively the lowest and largest singular values of $P_m(\xi)A(\xi)$ with respect to the norm $\|\cdot\|_X$, respectively defined by the infimum and supremum of $\|P_m(\xi)A(\xi)v\|_X$ over $v \in X$ such that $\|v\|_X = 1$. Then, we easily prove that algorithm (35) is such that

$$\|u(\xi_{r+1}^{RB}) - u_r(\xi_{r+1}^{RB})\|_X \geq \gamma_m \max_{\xi\in\Xi} \|u(\xi) - u_r(\xi)\|_X, \tag{36}$$

where $\gamma_m = \underline{\alpha}_m / \bar{\beta}_m \leq 1$ measures how far the selection of the new point is from the ideal greedy selection. Under condition (36), convergence results for this weak greedy algorithm can be found in [5, 18].

We give now sharper bounds for the preconditoned residual norm that exploits the fact that the approximation $u_r(\xi)$ is the Petrov-Galerkin projection.

Proposition 3.4 *Let $u_r(\xi)$ be the Petrov-Galerkin projection of $u(\xi)$ on X_r defined by* (29)*. Then we have*

$$\alpha_{r,m}(\xi)\ \|u(\xi) - u_r(\xi)\|_X \leq \|P_m(\xi)(A(\xi)u_r(\xi) - b(\xi))\|_X \leq \beta_{r,m}(\xi)\ \|u(\xi) - u_r(\xi)\|_X,$$

with

$$\alpha_{r,m}(\xi) = \inf_{v\in X} \sup_{w_r\in X_r} \frac{\|(P_m(\xi)A(\xi))^T R_X v\|_{X'}}{\|v - w_r\|_X},$$
$$\beta_{r,m}(\xi) = \sup_{v\in X} \inf_{w_r\in X_r} \frac{\|(P_m(\xi)A(\xi))^T R_X (v - w_r)\|_{X'}}{\|v\|_X}.$$

Proof: For any $v \in X$ and $w_r \in X_r$ and according to (30), we have

$$\begin{aligned}
\langle u(\xi) - u_r(\xi), R_X v\rangle &= \langle b(\xi) - A(\xi)u_r(\xi), A^{-T}(\xi)R_X v - P_m^T(\xi)R_X w_r\rangle \\
&= \langle P_m(\xi)(b(\xi) - A(\xi)u_r(\xi)), (P_m(\xi)A(\xi))^{-T}R_X v - R_X w_r\rangle \\
&\leq \|R\|_X\ \|(P_m(\xi)A(\xi))^{-T}R_X v - R_X w_r\|_{X'},
\end{aligned}$$

where $R(\xi) := P_m(\xi)(b(\xi) - A(\xi)u_r(\xi))$. Taking the infimum over $w_r \in X_r$, dividing by $\|v\|_X$ and taking the supremum over $v \in X$, we obtain

$$\begin{aligned}
\|u(\xi) - u_r(\xi)\|_X &\leq \|R(\xi)\|_X \sup_{v\in X} \inf_{w_r\in X_r} \frac{\|(P_m(\xi)A(\xi))^{-T}R_X v - R_X w_r\|_{X'}}{\|v\|_X}, \\
&= \|R(\xi)\|_X \sup_{v\in X} \inf_{w_r\in X_r} \frac{\|v - w_r\|_X}{\|(P_m(\xi)A(\xi))^T R_X v\|_{X'}}, \\
&= \|R(\xi)\|_X \left(\inf_{v\in X} \sup_{w_r\in X_r} \frac{\|(P_m(\xi)A(\xi))^T R_X v\|_{X'}}{\|v - w_r\|_X}\right)^{-1},
\end{aligned}$$

which proves the first inequality. Furthermore, for any $v \in X$ and $w_r \in X_r$, we have

$$\begin{aligned}\langle P_m(\xi)(b(\xi) - A(\xi)u_r(\xi)), R_X v\rangle &= \langle b(\xi) - A(\xi)u_r(\xi), P_m^T(\xi)R_X(v - w_r)\rangle \\ &\leq \|u(\xi) - u_r(\xi)\|_X \ \|(P_m(\xi)A(\xi))^T R_X(v - w_r)\|_{X'}.\end{aligned}$$

Taking the infimum over $w_r \in X_r$, dividing by $\|v\|_X$ and taking the supremum over $v \in X$, we obtain the second inequality. ∎

Since $X_r \subset X_{r+1}$, we have $\alpha_{r+1,m}(\xi) \geq \alpha_{r,m}(\xi) \geq \alpha_m(\xi)$ and $\beta_{r+1,m}(\xi) \leq \beta_{r,m}(\xi) \leq \beta_m(\xi)$. Equation (36) holds with γ_m replaced by the parameter $\gamma_{r,m} = \underline{\alpha}_{r,m}/\ \bar{\beta}_{r,m}$. Since $\gamma_{r,m}$ increases with r, a reasonable expectation is that the convergence properties of the weak greedy algorithm will improve when r increases.

Remark 3.5 *When replacing $P_m(\xi)$ by R_X^{-1}, the preconditioned residual norm $\|P_m(\xi)(A(\xi)u_r(\xi) - b(\xi))\|_X$ turns out to be the residual norm $\|A(\xi)u_r(\xi) - b(\xi)\|_{X'}$, which is a standard choice in the Reduced Basis method for the greedy selection of points (with R_X being associated with the natural norm on X or with a norm associated with the operator at some nominal parameter value). This can be interpreted as a basic preconditioning method with a parameter-independent preconditioner.*

4 Selection of the interpolation points

In this section, we propose strategies for the adaptive selection of the interpolation points. For a given set of interpolation points $\xi_1, \ldots, \xi_m$, three different methods are proposed for the selection of a new interpolation point ξ_{m+1}. The first method aims at reducing uniformly the error between the inverse operator and its interpolation. The resulting interpolation of the inverse is pertinent for preconditioning iterative solvers or estimating errors based on preconditioned residuals. The second method aims at improving Petrov-Galerkin projections of the solution of a parameter-dependent equation on a given approximation space. The third method aims at reducing the cost for the computation of the preconditioner by reusing operators computed when solving samples of a parameter-dependent equation.

4.1 Greedy approximation of the inverse of a parameter-dependent matrix

A natural idea is to select a new interpolation point where the preconditioner $P_m(\xi)$ is not a good approximation of $A(\xi)^{-1}$. Obviously, an ideal strategy for preconditioning would be

to choose ξ_{m+1} where the condition number of $P_m(\xi)A(\xi)$ is maximal. The computation of the condition number for many values of ξ being computationaly expensive, one could use upper bounds of this condition number, e.g. computed using SCM [27].

Here, we propose the following selection method: given an approximation $P_m(\xi)$ associated with interpolation points $\xi_1, \ldots, \xi_m$, a new point ξ_{m+1} is selected such that

$$\xi_{m+1} \in \operatorname*{argmax}_{\xi \in \Xi} \|(I - P_m(\xi)A(\xi))V\|_F, \tag{38}$$

where the matrix V is either the random rescaled Rademacher matrix, or the P-SRHT matrix (see Section 2.2). This adaptive selection of the interpolation points yields the construction of an increasingsequence of subspaces $Y_{m+1} = Y_m + \operatorname{span}(A(\xi_{m+1})^{-1})$ in $Y = \mathbb{R}^{n\times n}$. This algorithm is detailed below.

Algorithm 3 Greedy selection of interpolation points.

Require: $A(\xi), V, M$.

Ensure: Interpolation points $\xi_1, \ldots, \xi_M$ and interpolation $P_M(\xi)$.

1: Initialize $P_0(\xi) = I$
2: **for** $m = 0$ to $M - 1$ **do**
3: Compute the new point ξ_{m+1} according to (38)
4: Compute a factorization of $A(\xi_{m+1})$
5: Define $A(\xi_{m+1})^{-1}$ as an implicit operator
6: Update the space $Y_{m+1} = Y_m + \operatorname{span}(A(\xi_{m+1})^{-1})$
7: Compute $P_{m+1}(\xi) = \arg\min_{P\in Y_{m+1}} \|(I - PA(\xi))V\|_F$
8: **end for**

The following lemma interprets the above construction as a weak greedy algorithm.

Lemma 4.1 *Assume that $A(\xi)$ satisfies $\underline{\alpha}_0\|\cdot\| \le \|A(\xi)\cdot\| \le \bar{\beta}_0\|\cdot\|$ for all $\xi \in \Xi$, and let $P_m(\xi)$ be defined by (5). Under the assumption that there exists $\varepsilon \in [0, 1[$ such that*

$$|\|(I - PA(\xi))V\|_F^2 - \|I - PA(\xi)\|_F^2| \le \epsilon\|I - PA(\xi)\|_F^2 \tag{39}$$

holds for all $\xi \in \Xi$ and $P \in Y_m$, we have

$$\|P_m(\xi_{m+1}) - A(\xi_{m+1})^{-1}\|_F \ge \gamma_\varepsilon \max_{\xi\in\Xi} \min_{P\in Y_m} \|P - A(\xi)^{-1}\|_F, \tag{40}$$

with $\gamma_\varepsilon = \underline{\alpha}_0\sqrt{1-\varepsilon}/(\bar{\beta}_0\sqrt{1+\varepsilon})$, and with ξ_{m+1} defined by (38).

Proof: Since $\|BC\|_F \le \|B\|_F\|C\|$ holds for any matrices B and C, with $\|C\|$ the operator norm of C, we have for all $P \in Y$,

$$\|A(\xi)^{-1} - P\|_F \le \|I - PA(\xi)\|_F\|A(\xi)^{-1}\| \le \underline{\alpha}_0^{-1}\|I - PA(\xi)\|_F,$$
$$\|I - PA(\xi)\|_F \le \|A(\xi)^{-1} - P\|_F\|A(\xi)\| \le \bar{\beta}_0\|A(\xi)^{-1} - P\|_F.$$

Then, thanks to (39) we have

$$\begin{aligned} & \|A(\xi)^{-1} - P\|_F \leq (\underline{\alpha}_0\sqrt{1-\varepsilon})^{-1}\|(I - PA(\xi))V\|_F \\ \text{and} \quad & \|(I - PA(\xi))V\|_F \leq \bar{\beta}_0\sqrt{1+\varepsilon}\|A(\xi)^{-1} - P\|_F, \end{aligned}$$

which implies

$$\frac{1}{\bar{\beta}_0\sqrt{1+\varepsilon}}\|(I - PA(\xi))V\|_F \leq \|A(\xi)^{-1} - P\|_F \leq \frac{1}{\underline{\alpha}_0\sqrt{1-\varepsilon}}\|(I - PA(\xi))V\|_F.$$

We easily deduce that ξ_{m+1} is such that (40) holds. ∎

Remark 4.2 *The assumption* (39) *of Lemma 4.1 can be proved to hold with high probability in two cases. A first case is when Ξ is a training set of finite cardinality, where the results of Proposition 2.5 can be extended to any $\xi \in \Xi$ by using a union bound. We then obtain that* (39) *holds with a probability higher than $1 - \delta(\#\Xi)$. A second case is when $A(\xi)$ admits an affine decomposition* (24) *with m_A terms. Then the space $M_L = span\{I - PA(\xi) : \xi \in \Xi, P \in Y_m\}$ is of dimension $L \leq 1 + m_A m$ and Proposition 2.4 allows to prove that assumption* (39) *holds with high probability.*

The quality of the resulting spaces Y_m have to be compared with the Kolmogorov m-width of the set $A^{-1}(\Xi) := \{A(\xi)^{-1} : \xi \in \Xi\} \subset Y$, defined by

$$d_m(A^{-1}(\Xi))_Y = \min_{\substack{Y_m \subset Y \\ \dim(Y_m) = m}} \sup_{\xi \in \Xi} \min_{P \in Y_m} \|A(\xi)^{-1} - P\|_F, \tag{41}$$

which evaluates how well the elements of $A^{-1}(\Xi)$ can be approximated on a m-dimensional subspace of matrices. (40) implies that the following results holds (see Corollary 3.3 in [18]):

$$\|A(\xi)^{-1} - P_m(\xi)\|_F = \begin{cases} \mathcal{O}(m^{-a}) & \text{if } d_m(A^{-1}(\Xi))_Y = \mathcal{O}(m^{-a}) \\ \mathcal{O}(e^{-\tilde{c}m^b}) & \text{if } d_m(A^{-1}(\Xi))_Y = \mathcal{O}(e^{-cm^b}) \end{cases},$$

where $\tilde{c} > 0$ is a constant which depends on c and b. That means that if the Kolmogorov m-width has an algebraic or exponential convergence rate, then the weak greedy algorithm yields an error $\|P_m(\xi) - A(\xi)^{-1}\|_F$ which has the same type of convergence. Therefore, the proposed interpolation method will present good convergence properties when $d_m(A^{-1}(\Xi))_Y$ rapidly decreases with m.

Remark 4.3 *When the parameter set Ξ is $[-1, 1]^d$ (or a product of compact intervals), an exponential decay can be obtained when $A(\xi)^{-1}$ admits an holomorphic extension to a domain in $\mathbb{C}^d$ containing Ξ (see [10]).*

Remark 4.4 *Note that here, there is no constraint on the minimization problem over Y_m (either optimal subspaces or subspaces constructed by the greedy procedure), so that we have no guaranty that the resulting approximations Y_m are invertible (see Section 2.3).*

4.2 Selection of points for improving the projection on a reduced space

We here suppose that we want to find an approximation of the solution $u(\xi)$ of a parameter-dependent equation (27) onto a low-dimensional approximation space X_r, using a Petrov-Galerkin orthogonality condition given by (30). The best approximation is considered as the orthogonal projection defined by (28). The quantity $\delta_{r,m}(\xi)$ defined by (31) controls the quality of the Petrov-Galerkin projection on X_r (see Proposition 3.2). As indicated in Proposition 3.3, $\delta_{r,m}(\xi)$ can be efficiently computed. Thus, we propose the following selection strategy which aims at improving the quality of the Petrov-Galerkin projection: given a preconditioner $P_m(\xi)$ associated with interpolation points $\xi_1, \ldots, \xi_m$, the next point ξ_{m+1} is selected such that

$$\xi_{m+1} \in \operatorname*{argmax}_{\xi \in \Xi} \delta_{r,m}(\xi). \tag{42}$$

The resulting construction is described by Algorithm 3 with the above selection of ξ_{m+1}. Note that this strategy is closely related with [14], where the authors propose a greedy construction of a parameter-independent test space for Petrov-Galerkin projection, with a selection of basis functions based on an error indicator similar to $\delta_{r,m}(\xi)$.

4.3 Re-use of factorizations of operator's evaluations - Application to reduced basis method

When using a sample-based approach for solving a parameter-dependent equation (27), the linear system is solved for many values of the parameter ξ. When using a direct solver for solving a linear system for a given ξ, a factorization of the operator is usually available and can be used for improving a preconditioner for the solution of subsequent linear systems.

We here describe this idea in the particular context of greedy algorithms for Reduced Basis method, where the interpolation points $\xi_1, \ldots, \xi_r$ for the interpolation of the inverse $A(\xi)^{-1}$ are taken as the evaluation points $\xi_1^{RB}, \ldots, \xi_r^{RB}$ for the solution. At iteration r, having a preconditioner $P_r(\xi)$ and an approximation $u_r(\xi)$, a new interpolation point is defined such that

$$\xi_{r+1}^{RB} \in \operatorname*{argmax}_{\xi \in \Xi} \|P_r(\xi)(A(\xi)u_r(\xi) - b(\xi))\|_X.$$

Algorithm 4 describes this strategy.

Algorithm 4 Reduced Basis method with re-use of operator's factorizations.

Require: $A(\xi), b(\xi)$, V, and R.

Ensure: Approximation $u_R(\xi)$.

1: Initialize $u_0(\xi) = 0$, $P_0(\xi) = I$
2: **for** $r = 0$ to $R - 1$ **do**
3: Find $\xi_{r+1}^{RB} \in \arg\max_{\xi \in \Xi} \|P_r(\xi)(A(\xi)u_r(\xi) - b(\xi))\|_X$
4: Compute a factorization of $A(\xi_{r+1}^{RB})$
5: Solve the linear system $v_{r+1} = A(\xi_{r+1}^{RB})^{-1} b(\xi_{r+1}^{RB})$
6: Update the approximation subspace $X_{r+1} = X_r + \text{span}(v_{r+1})$
7: Define the implicit operator $P_{r+1} = A(\xi_{r+1}^{RB})^{-1}$
8: Update the space Y_{r+1} (or Y_{r+1}^+)
9: Compute the preconditioner : $P_{r+1}(\xi) = \underset{P \in Y_{r+1} (\text{or } Y_{r+1}^+)}{\text{argmin}} \|(I - PA(\xi))V\|_F$
10: Compute the Petrov-Galerkin approximation $u_{r+1}(\xi)$ of $u(\xi)$ on X_{r+1} using equation (30)
11: **end for**

5 Numerical results

5.1 Illustration on a one parameter-dependent model

In this section we compare the different interpolation methods on the following one parameter-dependent advection-diffusion-reaction equation:

$$-\Delta u + v(\xi) \cdot \nabla u + u = f \tag{43}$$

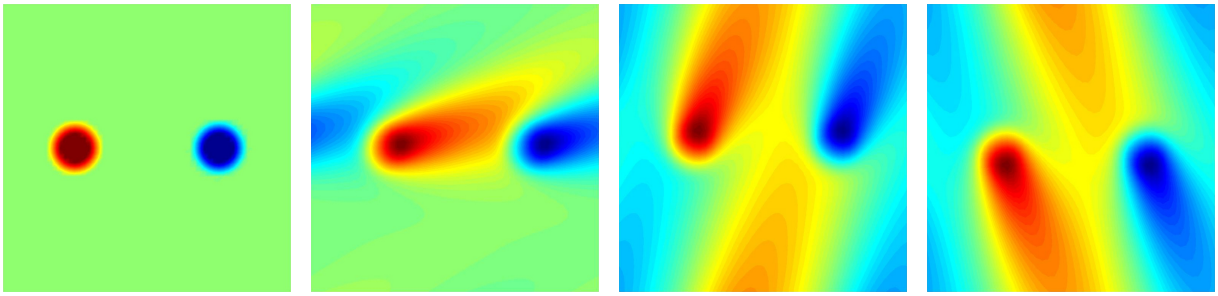
defined over a square domain $\Omega = [0, 1]^2$ with periodic boundary conditions. The advection vector field $v(\xi)$ is spatially constant and depends on the parameter ξ that takes values in $[0, 1]$: $v(\xi) = D\cos(2\pi\xi)e_1 + D\sin(2\pi\xi)e_2$, with $D = 50$ and (e_1, e_2) the canonical basis of $\mathbb{R}^2$. Ξ denotes a uniform grid of 250 points on $[0, 1]$. The source term f is represented in Figure 2(a). We introduce a finite element approximation space of dimension $n = 1600$ with piecewise linear approximations on a regular mesh of Ω. The mesh Péclet number takes moderate values (lower than one), so that a standard Galerkin projection without stabilization is here sufficient. The Galerkin projection yields the linear system of equations $A(\xi)u(\xi) = b$, with

$$A(\xi) = A_0 + \cos(2\pi\xi)A_1 + \sin(2\pi\xi)A_2,$$

where the matrices A_0, A_1, A_2 and the vector b are given by

$$(A_0)_{i,j} = \int_\Omega \nabla\phi_i \cdot \nabla\phi_j + \phi_i\phi_j \ , \ (A_1)_{i,j} = \int_\Omega \phi_i(e_1 \cdot \nabla\phi_j)$$
$$(A_2)_{i,j} = \int_\Omega \phi_i(e_2 \cdot \nabla\phi_j) \ , \ (b)_i = \int_\Omega \phi_i f,$$

where $\{\phi_i\}_{i=1}^n$ is the basis of the finite element space. Figures 2(b), 2(c) and 2(d) show three samples of the solution.



(a) f (b) $u(0.05)$ (c) $u(0.2)$ (d) $u(0.8)$

Figure 2: Plot of the source term f (a) and 3 samples of the solution corresponding to parameter values $\xi = 0.05$ (b), $\xi = 0.2$ (c) and $\xi = 0.8$ (d) respectively.

5.1.1 Comparison of the interpolation strategies

We first choose arbitrarily 3 interpolation points ($\xi_1 = 0.05$, $\xi_2 = 0.2$ and $\xi_3 = 0.8$) and show the benefits of using the Frobenius norm projection for the definition of the preconditioner. For the comparison, we consider the Shepard and the nearest neighbor interpolation strategies. Let $\|\cdot\|_\Xi$ denote a norm on the parameter set Ξ. The Shepard interpolation method is an inverse weighted distance interpolation:

$$\lambda_i(\xi) = \begin{cases} \dfrac{\|\xi - \xi_i\|_\Xi^{-s}}{\sum_{j=1}^m \|\xi - \xi_j\|_\Xi^{-s}} & \text{if } \xi \neq \xi_i \\ 1 & \text{if } \xi = \xi_i \end{cases},$$

where $s > 0$ is a parameter. Here we take $s = 2$. The nearest neighbor interpolation method consists in choosing the value taken by the nearest interpolation point, that means $\lambda_i(\xi) = 1$ for some $i \in \arg\min_j \|\xi - \xi_j\|_\Xi$, and $\lambda_j(\xi) = 0$ for all $j \neq i$.

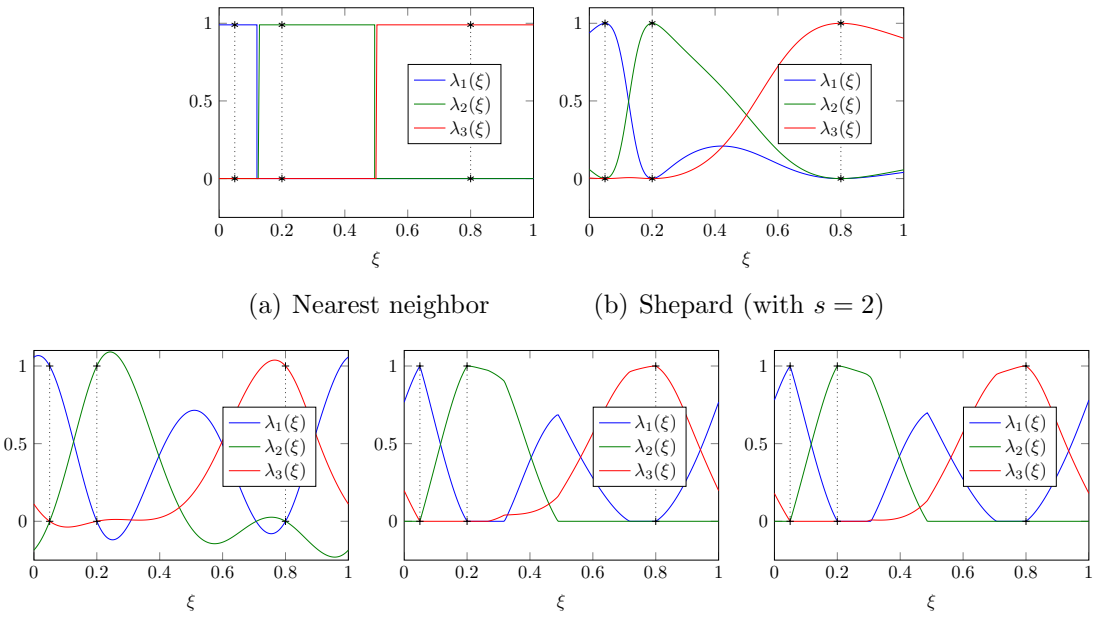
Concerning the Frobenius norm projection on Y_m (or Y_m^+), we first construct the affine decomposition of $M(\xi)$ and $S(\xi)$ as explained in Section 2.4. The interpolation points ξ_k^* (resp. $\widetilde{\xi}_k^*$) given by the EIM procedure for $M(\xi)$ (resp. $S(\xi)$) are $\{0.0; 0.25; 0.37; 0.56; 0.80\}$

(resp. $\{0.0; 0.25; 0.62\}$). The number of terms $m_M = 5$ in the resulting affine decomposition of $M(\xi)$ (see equation (26)) is less than the expected number $m_A^2 = 9$ (see equation (25a)). Considering the functions $\Phi_1(\xi) = 1$, $\Phi_2(\xi) = \cos(2\pi\xi)$, $\Phi_3(\xi) = \sin(2\pi\xi)$, and thanks to relation $\cos^2 = 1 - \sin^2$, the space

$$\mathrm{span}_{i,j}\{\Phi_i\Phi_j\} = \mathrm{span}\{1, \cos, \sin, \cos\sin, \cos^2, \sin^2\} = \mathrm{span}\{1, \cos, \sin, \cos\sin, \cos^2\}$$

is of dimension $m_M = 5$. The EIM procedure automatically detects the redundancy in the set of functions and reduces the number of terms in the decomposition (26). Then, since the dimension n of the discretization space is reasonable, we compute the matrices $M(\xi_k^*)$ and the vectors $S(\widetilde{\xi_k^*})$ using equation (26).

The functions $\lambda_i(\xi)$ are plotted on Figure 3 for the proposed interpolation strategies. It is important to note that contrary to the Shepard or the nearest neighbor method, the Frobenius norm projection (on Y_m or Y_m^+) leads to periodic interpolation functions, *i.e.* $\lambda_i(\xi = 0) = \lambda_i(\xi = 1)$. This is consistent with the fact that the application $\xi \mapsto A(\xi)$ is 1-periodic. The Frobenius norm projection automatically detects such a feature.



(a) Nearest neighbor (b) Shepard (with $s = 2$)

(c) Projection on Y_m (d) Projection on Y_m^+ (e) Projection on $Y_m^{\bar{\kappa}}$

Figure 3: Interpolation functions $\lambda_i(\xi)$ for different interpolation methods.

Figure 4 shows the condition number $\kappa_m(\xi)$ of $P_m(\xi)A(\xi)$ with respect to ξ. We first note that for the constant preconditioner $P_1(\xi) = A(\xi_2)^{-1}$, the resulting condition number is higher than the one of the non preconditioned matrix $A(\xi)$ for $\xi \in [0.55; 0.95]$. We also note that the interpolation strategies based on the Frobenius norm projection lead to

better preconditioners than the Shepard and nearest neighbor interpolation strategies. When considering the projection on Y_m^+ and $Y_m^{\bar{\kappa}}$ (with $\bar{\kappa} = 5\times10^4$ such that (22) holds), the resulting condition number is roughly the same, so as the interpolation functions of Figures 3(d) and 3(e). Since the projection on $Y_m^{\bar{\kappa}}$ requires the expensive computation of the constants γ^+, γ^- and C (see Section 2.3), we prefer to simply use the projection on Y_m^+ in order to ensure the preconditioner to be invertible. Finally, for this example, it is not necessary to impose any constraint since the projection on Y_m leads to the best preconditioner and this preconditioner appears to be invertible for any $\xi \in \Xi$.

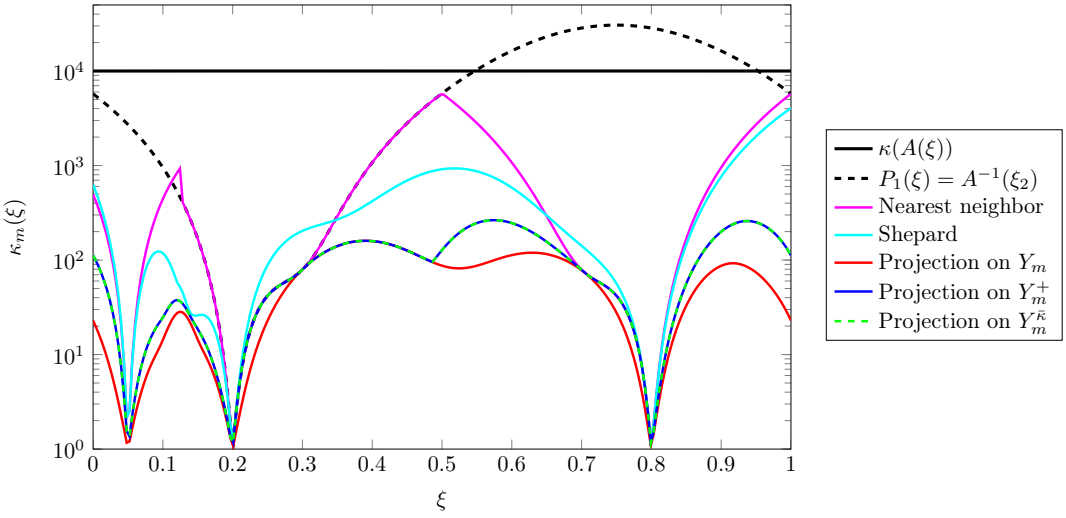


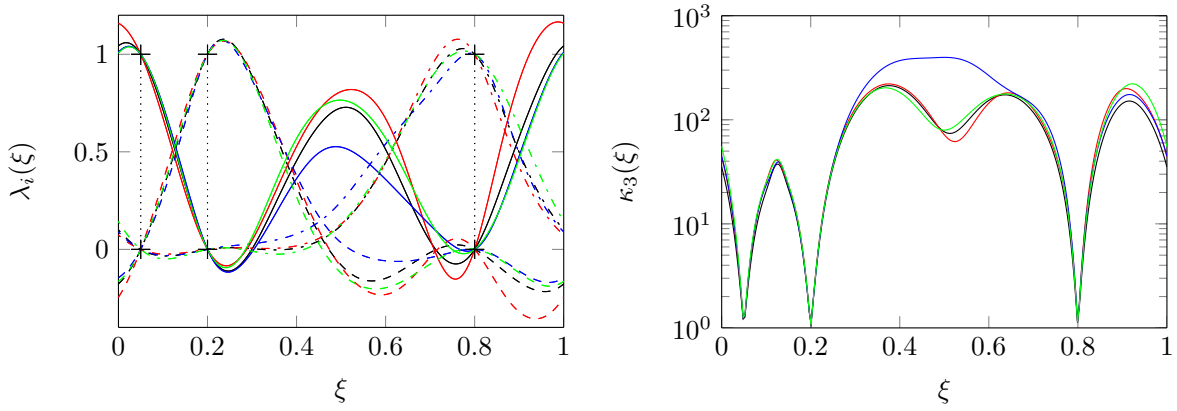
Figure 4: Condition number of $P_m(\xi)A(\xi)$ for different interpolation strategies. The condition number of $A(\xi)$ is given as a reference.

5.1.2 Using the Frobenius semi-norm

We analyze now the interpolation method defined by the Frobenius semi-norm projection on Y_m (5) for the different definitions of $V \in \mathbb{R}^{n\times K}$ proposed in sections 2.2.2 and 2.2.1. According to Table 2, the error on the interpolation functions decreases slowly with K (roughly as $\mathcal{O}(K^{-1/2})$), and the use of the P-SRHT matrix leads to a slightly lower error. The interpolation functions are plotted on Figure 5(a) in the case where $K = 8$. Even if we have an error of 36% to 101% on the interpolation functions, the condition number given on Figure 5(b) remains close to the one computed with the Frobenius norm. Also, an important remark is that with $K = 8$ the computational effort for computing $M^V(\xi_k^*)$ and $S^V(\widetilde{\xi}_k^*)$ is negligible compared to the one for $M(\xi_k^*)$ and $S(\widetilde{\xi}_k^*)$.

K	8	16	32	64	128	256	512
Rescaled partial Hadamard	0.4131	0.3918	0.3221	0.1010	0.0573	0.0181	0.0255
Rescaled Rademacher (1)	0.5518	0.0973	0.2031	0.1046	0.1224	0.1111	0.0596
Rescaled Rademacher (2)	1.0120	0.6480	0.1683	0.1239	0.0597	0.0989	0.0514
Rescaled Rademacher (3)	0.7193	0.2014	0.1241	0.1051	0.1235	0.1369	0.0519
P-SRHT (1)	0.4343	0.2081	0.2297	0.0741	0.0723	0.0669	0.0114
P-SRHT (2)	0.3624	0.2753	0.0931	0.1285	0.0622	0.0619	0.0249
P-SRHT (3)	0.8133	0.4227	0.1138	0.0741	0.0824	0.0469	0.0197

Table 2: Relative error $\sup_\xi \|\lambda(\xi) - \lambda^V(\xi)\|_{\mathbb{R}^3} / \sup_\xi \|\lambda(\xi)\|_{\mathbb{R}^3}$: $\lambda^V(\xi)$ (resp. $\lambda(\xi)$) are the interpolation functions associated to the Frobenius semi-norm projection (resp. the Frobenius norm projection) on Y_m, with V either the rescaled partial Hadamard matrix, the random rescaled Rademacher matrix or the P-SRHT matrix (3 different samples for random matrices).



(a) Interpolation functions.

(b) Condition number of $P_3(\xi)A(\xi)$.

Figure 5: Comparison between the Frobenius norm projection on Y_3 (black lines) and the Frobenius semi-norm projection on Y_3, using for V either a sample of the rescaled Rademacher matrix (blue lines), the rescaled partial Hadamard matrix (red lines) or a sample of the P-SRHT matrix (green lines) with $K = 8$.

5.1.3 Greedy selection of the interpolation points

We now consider the greedy selection of the interpolation points presented in Section 4. We start with an initial point $\xi_1 = 0$ and the next points are defined by (38), where matrix V is a realization of the P-SRHT matrix with $K = 128$ columns. $P_m(\xi)$ is the projection on Y_m using the Frobenius semi-norm defined by (5). The first 3 steps of the algorithm are illustrated on Figure 6. We observe that at each iteration, the new interpolation point ξ_{m+1} is close to the point where the condition number of $P_m(\xi)A(\xi)$ is maximal. Table 3 presents

the maximal value over $\xi \in \Xi$ of the residual, and of the condition number of $P_m(\xi)A(\xi)$. Both quantities are rapidly decreasing with m. This shows that this algorithm, initially designed to minimize $\|(I - P_m(\xi)A(\xi))V\|_F$, seems to be also efficient for the construction of preconditioners, in the sense that the condition number decreases rapidly.

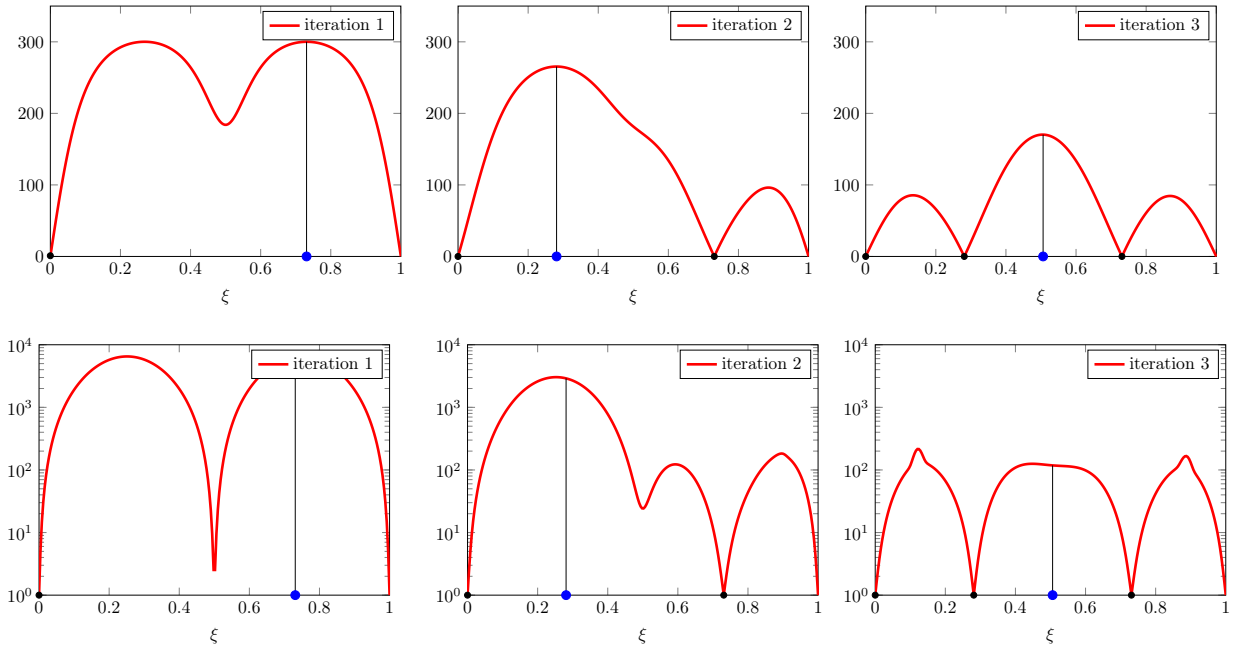


Figure 6: Greedy selection of the interpolation points: the first row is the residual $\|(I - P_k(\xi)A(\xi))V\|_F$ (the blue points correspond to the maximum of the residual) with V a realization of the P-SRHT matrix with $K = 128$ columns, and the second row is the condition number of $P_m(\xi)A(\xi)$.

iteration m	0	1	2	5	10	20	30
$\sup_\xi \kappa(P_m(\xi)A(\xi))$	10001	6501	3037	165,7	51,6	16,7	7,3
$\sup_\xi \|(I - P_m(\xi)A(\xi))V\|_F$	-	300	265	80,5	35,4	10,0	7,6

Table 3: Convergence of the greedy algorithm: supremum over $\xi \in \Xi$ of the condition number (first row) and of the Frobenius semi-norm residual (second row).

5.2 Multi-parameter-dependent equation

We introduce a benchmark proposed within the OPUS project (see http://www.opus-project.fr). Two electronic components Ω_{IC} (see Figure 7) submitted to a cooling air flow in the domain Ω_{Air} are fixed on a printed circuit board Ω_{PCB}. The temperature field defined over

$\Omega = \Omega_{IC} \cup \Omega_{PBC} \cup \Omega_{Air} \subset \mathbb{R}^2$ satisfies the advection-diffusion equation:

$$-\nabla \cdot (\kappa(\xi)\nabla u) + D(\xi) v \cdot \nabla u = f. \tag{44}$$

The diffusion coefficient $\kappa(\xi)$ is equal to κ_{PCB} on Ω_{PCB}, κ_{Air} on Ω_{Air} and κ_{IC} on Ω_{IC}. The right hand side f is equal to $Q = 10^6$ on Ω_{IC} and 0 elsewhere. The boundary conditions are $u = 0$ on Γ_d, $e_2 \cdot \nabla u = 0$ on Γ_u (e_1, e_2 are the canonical vectors of $\mathbb{R}^2$), and $u_{|\Gamma_l} = u_{|\Gamma_r}$ (periodic boundary condition). At the interface $\Gamma_C = \partial\Omega_{IC} \cap \partial\Omega_{PCB}$ there is a thermal contact conductance, meaning that the temperature field u admits a jump over Γ_C which satisfies

$$\kappa_{IC}(e_1 \cdot \nabla u_{|\Omega_{IC}}) = \kappa_{PCB}(e_1 \cdot \nabla u_{|\Omega_{PCB}}) = r(u_{|\Omega_{IC}} - u_{|\Omega_{PCB}}) \quad \text{ on } \Gamma_C.$$

The advection field v is given by $v(x, y) = e_2 g(x)$, where $g(x) = 0$ if $x \leq e_{PCB} + e_{IC}$ and

$$g(x) = \frac{3}{2(e - e_{IC})} \left(1 - \left(\frac{2x - (e + e_{IC} + 2e_{PCB})}{e - e_{IC}} \right)^2 \right)$$

otherwise. We have 4 parameters: the width $e := \xi_1$ of the domain Ω_{Air}, the thermal conductance parameter $r := \xi_2$, the diffusion coefficient $\kappa_{IC} := \xi_3$ of the components and the amplitude of the advection field $D := \xi_4$. Since the domain $\Omega = \Omega(e)$ depends on the parameter $\xi_1 \in [e_{min}, e_{max}]$, we introduce a geometric transformation $(x, y) = \phi_{\xi_1}(x_0, y_0)$ that maps a reference domain $\Omega_0 = \Omega(e_{max})$ to $\Omega(\xi_1)$:

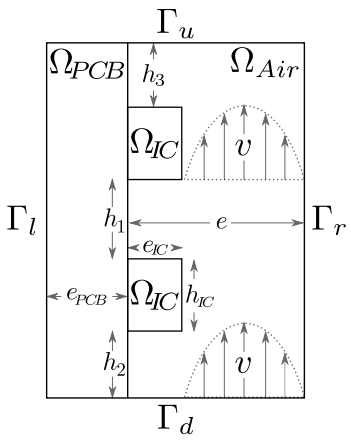
$$\phi_{\xi_1}(x_0, y_0) = \begin{pmatrix} \begin{cases} x_0 & \text{if } x_0 \leq e_0 \\ e_0 + (x_0 - e_0)\frac{\xi_1 - e_{IC}}{e_{max} - e_{IC}} & \text{otherwise.} \end{cases} \\ y_0 \end{pmatrix},$$

with $e_0 = e_{PCB} + e_{IC}$. This method is described in [36]: since the geometric transformation ϕ_{ξ_1} satisfies the so-called *Affine Geometry Precondition*, the operator of equation (44) formulated on the reference domain admits an affine representation.

For the spatial discretization we use a finite element approximation with $n = 2.8 \cdot 10^4$ degrees of freedom (piecewise linear approximation). We rely on a Galerkin method with SUPG stabilization (see [8]). Ξ is a set of 10^4 independent samples drawn according the loguniform probability laws of the parameters given on Figure 7.

5.2.1 Preconditioner for the projection on a given reduced space

We consider here a POD basis X_r of dimension $r = 50$ computed with 100 snapshots of the solution (a basis of X_r is obtained by the first 50 dominant singular vectors of a matrix



Geometry (m)	
$e = \xi_1$	$\sim \log\mathcal{U}(2.5 \times 10^{-3}, 5 \times 10^{-2})$
e_{PCB}	2×10^{-3}
e_{IC}	2×10^{-3}
h_{IC}	2×10^{-2}
h_1	3×10^{-2}
h_2	2×10^{-2}
h_3	4×10^{-2}
Thermal conductance ($Wm^{-2}K^{-1}$)	
$r = \xi_2$	$\sim \log\mathcal{U}(1 \times 10^{-1}, 1 \times 10^{2})$
Diffusion coefficients ($Wm^{-1}K^{-1}$)	
κ_{PCB}	2×10^{-1}
κ_{Air}	3×10^{-2}
$\kappa_{IC} = \xi_3$	$\sim \log\mathcal{U}(2 \times 10^{-1}, 1.5 \times 10^{2})$
Advection field ($Wm^{-2}K^{-1}$)	
$D = \xi_4$	$\sim \log\mathcal{U}(5 \times 10^{-4}, 1 \times 10^{-2})$

Figure 7: Geometry and parameters of the benchmark OPUS.

of 100 random snapshots of $u(\xi)$). Then we compute the Petrov-Galerkin projection as presented in Section 3.1. The efficiency of the preconditioner can be measured with the quantity $\delta_{r,m}(\xi)$: the associated quasi-optimality constant $(1 - \delta_{r,m}(\xi)^2)^{-1/2}$ should be as close to one as possible (see equation (33)). We introduce the quantile q_p of probability p associated to the quasi-optimality constant $(1 - \delta_{r,m}(\xi)^2)^{-1/2}$ defined as the smallest value $q_p \geq 1$ satisfying

$$\mathbb{P}\big(\{\xi \in \Xi : (1 - \delta_{r,m}(\xi)^2)^{-1/2} \leq q_p\}\big) \geq p,$$

where $\mathbb{P}(A) = \#A/\#\Xi$ for $A \subset \Xi$. Table 4 shows the evolution of the quantile with respect to the number of interpolation points for the preconditioner. Here the goal is to compare the different strategies for the selection of the interpolation points:

(a) the greedy selection (42) based on the quantity $\delta_{r,m}(\xi)$,

(b) the greedy selection (38) based on the Frobenius semi-norm residual, with V a P-SRHT matrix with $K = 256$ columns, and

(c) a random Latin Hypercube sample (LHS).

The projection on Y_m (or Y_m^+) is then defined with the Frobenius semi-norm using for V a P-SRHT matrix with $K = 330$ columns.

When considering a small number of interpolation points $m \leq 3$, the projection on Y_m^+ provides lower quantiles for the quasi-optimality constant compared to the projection on Y_m. The positivity constraint is useful for small m. But for high values of m (see $m = 15$) the positivity constraint is no longer necessary and the projection on Y_m provides lower quantiles.

Concerning the choice of the interpolation points, the strategy (a) shows the faster decay of the quantiles q_p, especially for $p = 50\%$. The strategy (b) shows also good results, but the

quantile q_p for $p = 100\%$ are still high compared to (a). These results show the benefits of the greedy selection based on the quasi-optimality constant. Finally the strategy (c) shows bad results (high values of the quantiles), especially for small m.

	Projection on Y_m								
	Greedy selection based on						(c) Latin Hypercube sampling		
	(a) $\delta_{r,m}(\xi)$			(b) Frob. residual					
	50%	90%	100%	50%	90%	100%	50%	90%	100%
$m = 0$	21.3	64.1	94.1	21.3	64.1	94.1	21.3	64.1	94.1
$m = 1$	18.3	74.1	286.7	10.2	36.1	161.6	18.3	104.1	231.8
$m = 2$	11.9	22.6	42.1	9.8	53.3	374.0	11.5	113.0	533.9
$m = 3$	11.1	49.2	200.4	7.8	31.2	60.2	18.3	138.7	738.5
$m = 5$	5.2	10.8	18.4	6.8	18.6	24.5	8.7	121.1	651.4
$m = 10$	3.1	9.0	13.2	5.3	22.3	62.1	4.0	21.6	345.7
$m = 15$	2.2	6.3	10.4	3.5	6.5	11.5	2.7	7.8	48.6

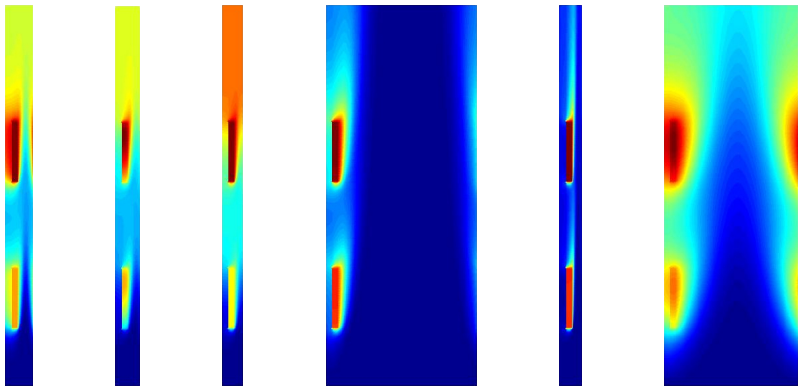
	Projection on Y_m^+								
	Greedy selection based on						(c) Latin Hypercube sampling		
	(a) $\delta_{r,m}(\xi)$			(b) Frob. residual					
	50%	90%	100%	50%	90%	100%	50%	90%	100%
$m = 0$	21.3	64.1	94.1	21.3	64.1	94.1	21.3	64.1	94.1
$m = 1$	18.3	74.1	286.7	10.2	36.1	161.6	18.3	104.1	231.8
$m = 2$	11.9	22.6	42.1	8.9	35.5	78.6	10.4	41.5	112.5
$m = 3$	9.7	24.4	48.0	7.9	27.7	57.9	12.1	48.8	114.1
$m = 5$	6.4	15.0	25.5	6.9	26.8	65.1	5.7	11.6	17.5
$m = 10$	4.6	9.5	16.8	7.3	18.9	38.0	4.3	10.0	18.5
$m = 15$	4.3	7.1	11.2	6.4	10.1	18.0	4.2	9.0	19.3

Table 4: Quantiles q_p of the quasi-optimality constant associated to the Petrov-Galerkin projection on the POD subspace X_r for $p = 50\%, 90\%$ and 100%. The row $m = 0$ corresponds to $P_0(\xi) = R_X^{-1}$, that is the standard Galerkin projection.

5.2.2 Preconditioner for Reduced Basis method

We now consider the preconditioned Reduced Basis method for the construction of the approximation space X_r, as presented in Section 3.2. Figures 9 and 10 show the convergence of the error with respect to the rank r of $u_r(\xi)$ for different constructions of the preconditioner $P_m(\xi)$. Two measures of the error are given: $\sup_{\xi\in\Xi} \|u(\xi) - u_r(\xi)\|_X / \|u(\xi)\|_X$, and the quantile of probability 0.97 for $\|u(\xi) - u_r(\xi)\|_X / \|u(\xi)\|_X$. The curve "Ideal greedy" corresponds to the algorithm defined by (34) which provides a reference for the ideally conditioned algorithm, *i.e.* with $\kappa_m(\xi) = 1$. Figure 8 shows the corresponding first interpolation points for the solution.

The greedy selection of the interpolation points based on (38) (see Figure 9) allows to almost recover the convergence curve of the ideal greedy algorithm when using the projection on Y_m with $m = 15$. For the strategy of re-using the operators factorizations, the approximation is rather bad for $r = m \leq 10$ meaning that the space Y_r (or Y_r^+) is not really adapted for the construction of a good preconditioner over the whole parametric domain. However, for higher values of r, the preconditioner is getting better and better. For $r \geq 20$, we almost reach the convergence of the ideal greedy algorithm. We conclude that this strategy of re-using the operator factorization, which has a computational cost comparable to the standard non preconditioned Reduced Basis greedy algorithm, allows obtaining asymptotically the performance of the ideal greedy algorithm. Note that the positivity constraint yields a better preconditioner for small values of r but is no longer necessary for large r.



(a) (b) (c) (d) (e) (f)

	(a) $r = 1$	(b) $r = 2$	(c) $r = 3$	(d) $r = 4$	(e) $r = 5$	(f) $r = 6$
e	$7.1 \cdot 10^{-3}$	$6.1 \cdot 10^{-3}$	$5.0 \cdot 10^{-3}$	$5.0 \cdot 10^{-2}$	$5.1 \cdot 10^{-3}$	$4.4 \cdot 10^{-2}$
r	$9.9 \cdot 10^{1}$	$2.3 \cdot 10^{0}$	$4.1 \cdot 10^{-1}$	$2.8 \cdot 10^{0}$	$8.6 \cdot 10^{-1}$	$4.8 \cdot 10^{1}$
κ_{IC}	$1.1 \cdot 10^{2}$	$2.1 \cdot 10^{-1}$	$3.2 \cdot 10^{1}$	$6.0 \cdot 10^{0}$	$1.1 \cdot 10^{2}$	$3.2 \cdot 10^{-1}$
D	$1.7 \cdot 10^{-3}$	$7.4 \cdot 10^{-4}$	$6.2 \cdot 10^{-4}$	$9.9 \cdot 10^{-3}$	$8.6 \cdot 10^{-3}$	$6.0 \cdot 10^{-4}$

Figure 8: First six interpolation points of the ideal reduced basis method and corresponding reduced basis functions.

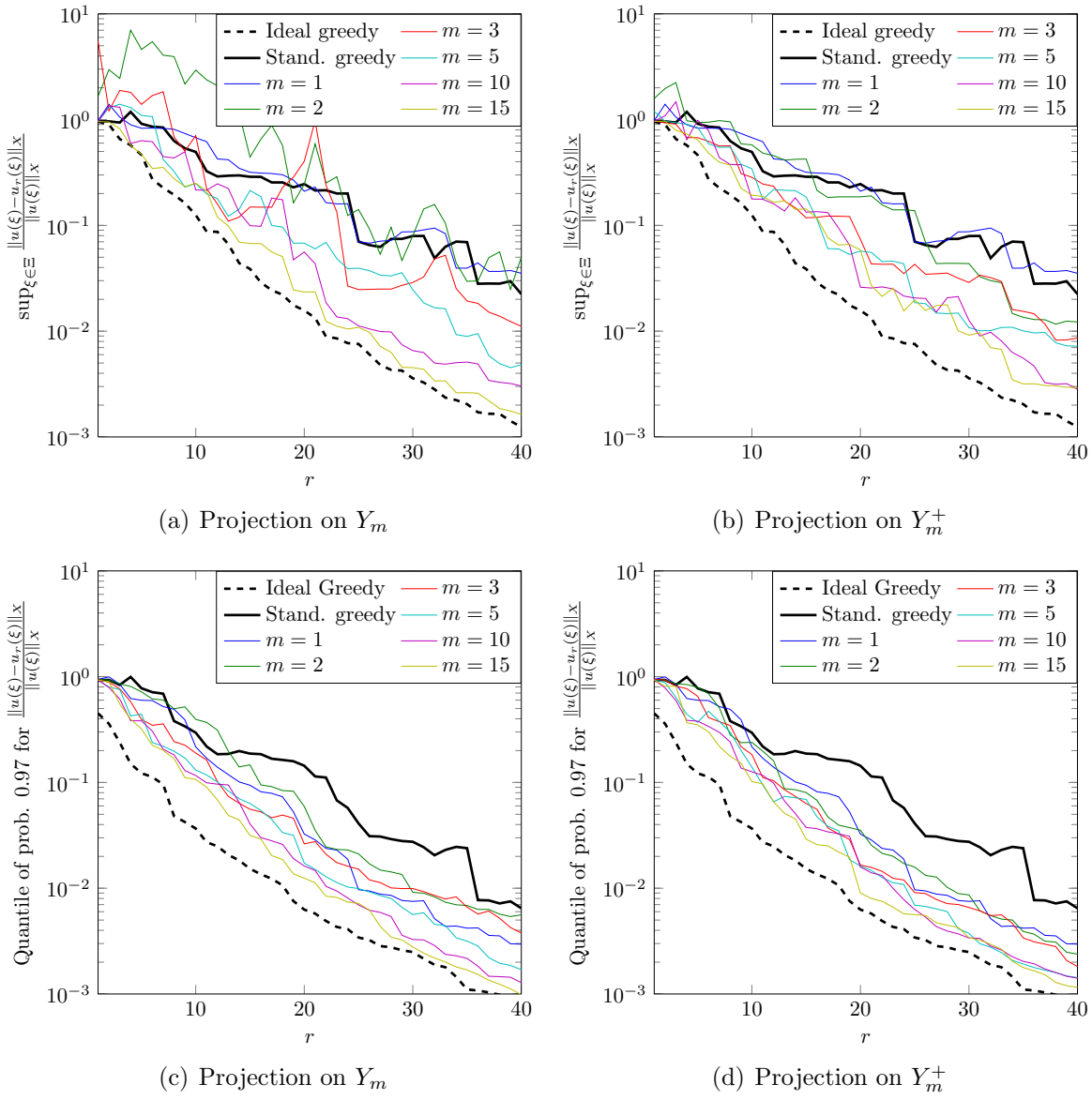
Let us finally consider the effectivity index

$$\eta_r(\xi) = \|P_r(\xi)(A(\xi)u_r(\xi) - b(\xi))\|_X / \|u(\xi) - u_r(\xi)\|_X,$$

which evaluates the quality of the preconditioned residual norm for error estimation. We introduce the confidence interval $I_r(p)$ defined as the smallest interval which satisfies

$$\mathbb{P}(\{\xi \in \Xi \ : \ \eta_r(\xi) \in I_r(p)\}) \geq p.$$

On Figure 11 we see that the confidence intervals are shrinking around 1 when r increases, meaning that the preconditioned residual norm becomes a better and better error estimator



(a) Projection on Y_m

(b) Projection on Y_m^+

(c) Projection on Y_m

(d) Projection on Y_m^+

Figure 9: Convergence of the preconditioned reduced basis method using the greedy selection of interpolation points for the preconditioner. Supremum over Ξ (top) and quantile of probability 97% (bottom) of the relative error $\|u(\xi) - u_r(\xi)\|_X / \|u(\xi)\|_X$ with respect to r. Comparison of preconditioned reduced basis algorithms with ideal and standard greedy algorithms.

when r increases. Again, the positivity constraint is needed for small values of r, but we obtain a better error estimation without imposing this constraint for $r \geq 20$. On the contrary, the standard residual norm leads to effectivity indices that spread from 10^{-1} to 10^{1} with no improvement as r increases, meaning that we can have a factor 10^{2} between the

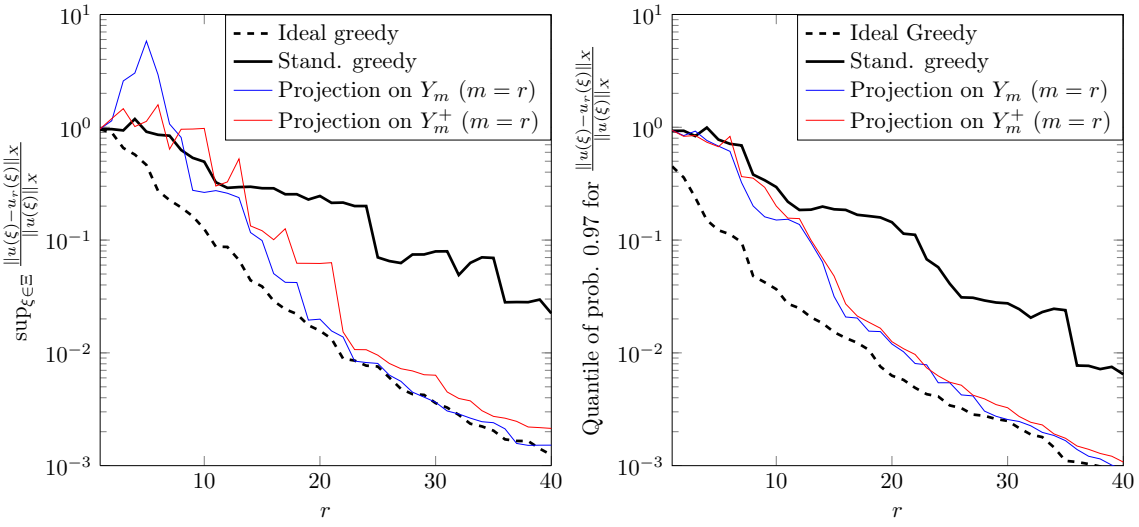
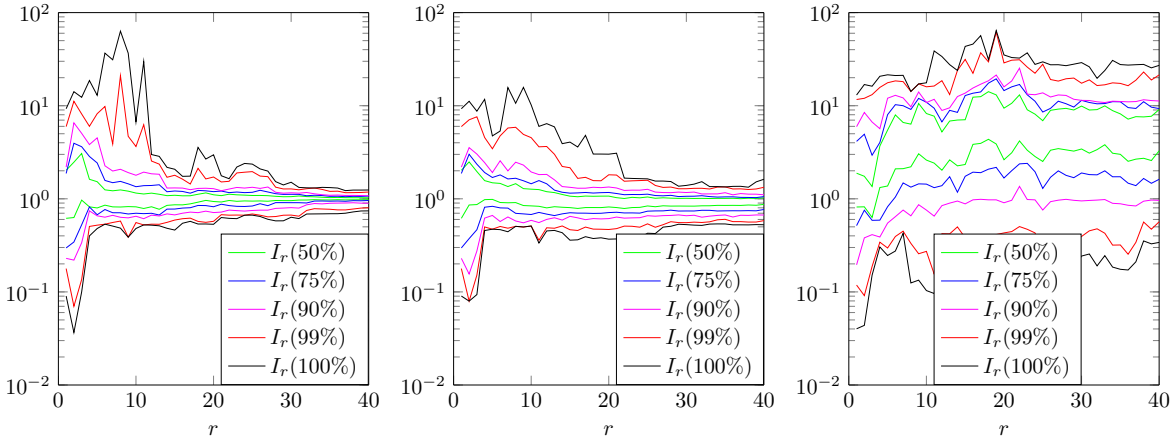


Figure 10: Preconditioned Reduced basis methods with re-use of operators. Supremum over Ξ (left) and quantile of probability 97% (right) of the relative error $\|u(\xi)-u_r(\xi)\|_X/\|u(\xi)\|_X$ with respect to r. Comparison of preconditioned reduced basis algorithms with ideal and standard greedy algorithms.

error estimator $\|A(\xi)u_r(\xi) - b(\xi)\|_{X'}$ and the true error $\|u_r(\xi) - u(\xi)\|_X$.



(a) Re-use, proj. on Y_m. (b) Re-use, proj. on Y_m^+. (c) Dual residual norm.

Figure 11: Confidence intervals of the effectivity index during the iterations of the Reduced Basis greedy construction. Comparison between preconditioned algorithms with re-use of operators factorizations (a,b) and the non preconditioned greedy algorithm (c).

6 Conclusion

We have proposed a method for the interpolation of the inverse of a parameter-dependent matrix. The interpolation is defined by the projection of the identity in the sense of the Frobenius norm. Approximations of the Frobenius norm have been introduced to make computationally feasible the projection in the case of large matrices. Then, we have proposed preconditioned versions of projection-based model reduction methods. The preconditioner can be used to define Petrov-Galerkin projections on a given approximation space with better quasi-optimality constants by introducing a parameter-dependent test space depending on the preconditioner. Also, the preconditioner can be used to improve residual-based error estimates that are used for assessing the quality of a given approximation, which is required in any adaptive approximation strategy. Different strategies have been proposed for the selection of interpolation points depending on the objective: (i) the construction of an optimal approximation of the inverse operator for preconditioning iterative solvers or for improving error estimators based on preconditioned residuals, (ii) the improvement of the quality of Petrov-Galerkin projections of the solution of a parameter-dependent equation on a given reduced approximation space, or (iii) the re-use of operators factorizations when solving a parameter-dependent equation with a sample-based approach. The performance of the obtained parameter-dependent preconditioners has been illustrated in the context of projection-based model reduction techniques such as the Proper Orthogonal Decomposition and the Reduced Basis method.

The proposed preconditioner has been used to define Petrov-Galerkin projections with better stability constants. For the solution of PDEs, the Petrov-Galerkin projection has been defined at the discrete (algebraic) level for obtaining a better approximation (in a reduced space) of the finite element Galerkin approximation of the PDE. Therefore, for convection-dominated problems, the proposed approach does not avoid using stabilized finite element formulations. Similar observations can be found in [34]. However, a Petrov-Galerkin method could be defined at the continuous level with a preconditioner being the interpolation of inverse partial differential operators. In this continuous framework, the preconditioner would improve the stability constant for the finite element Galerkin projection and may avoid the use of stabilized finite element formulations. Such Petrov-Galerkin methods have been proposed in [12, 13] for convection-dominated problems (as an alternative to standard stabilization methods), which can be interpreted as an implicit preconditioning method defined at the continuous level.

In the present paper, the parameter-dependent preconditioner is obtained by a projection onto the space generated by snapshots of the inverse operator. When the storage of many inverse operators (even as implicit matrices) is not feasible, a parameter-dependent precon-

ditioner could be obtained by a projection into the linear span of preconditioners, such as incomplete factorizations, sparse approximate inverses, H-matrices or other preconditoners that are readily available for a considered application. Also, we have restricted the presentation to the case of real matrices but the methodology can be naturally extended to the case of complex matrices.

References

[1] N. Ailon, and B. Chazelle, *The fast Johnson-Lindenstrauss transform and approximate nearest neighbors*, STOC (2006), pp. 557–563.

[2] H. Avron, and S. Toledo, *Randomized Algorithms for Estimating the Trace of an Implicit Symmetric Positive Semi-definite Matrix*, J. ACM, 58 (2011), pp. 8:1–8:34.

[3] C. Bekas, E. Kokiopoulou, and Y. Saad, *An estimator for the diagonal of a matrix*, Appl. Numer. Math., 57 (2007), pp. 1214–1229.

[4] M. Benzi, *Preconditioning Techniques for Large Linear Systems: A Survey*, J. Comput. Phys., 182 (2002), pp. 418–477.

[5] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova and P. Wojtaszczyk, *Convergence rates for greedy algorithms in reduced basis methods*, SIAM J. Math. Anal., 43 (2011), pp. 1457–1472.

[6] J. Bourgain, J. Lindenstrauss and V. Milman, *Approximation of zonoids by zonotopes*, Acta Mathematica, 162 (1989), pp. 73–141.

[7] C. Boutsidis and A. Gittens , *Improved Matrix Algorithms via the Subsampled Randomized Hadamard Transform*, SIAM J. Matrix Anal. A., 34 (2013), pp. 1301–1340.

[8] A. N. Brooks, and T. J. R. Hughes, *Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations*, Comput. Method. Appl. M., 32 (1982), pp. 199–259.

[9] F. Casenave, A. Ern, and T. Lelièvre, *A nonintrusive reduced basis method applied to aeroacoustic simulations*, Adv. Comput. Math. (2014), pp. 1–26.

[10] P. Chen, A. Quarteroni, and G. Rozza, *Comparison Between Reduced Basis and Stochastic Collocation Methods for Elliptic Problems*, Journal of Scientific Computing, 59 (2014), pp. 187–216.

[11] Y. Chen, S. Gottlieb and Y. Maday, *Parametric analytical preconditioning and its applications to the reduced collocation methods*, Comptes Rendus Mathematique, 352 (2014), pp. 661–666.

[12] A. Cohen, W. Dahmen, and G. Welper, *Adaptivity and variational stabilization for convection-diffusion equations*, ESAIM: M2AN, 46 (2012), pp. 1247–1273.

[13] W. Dahmen, C. Huang, C. Schwab, and G. Welper, *Adaptive Petrov–Galerkin Methods for First Order Transport Equations*, SIAM J. Numer. Anal., 50 (2012), pp. 2420–2445.

[14] W. Dahmen, C. Plesken, and G. Welper, *Double greedy algorithms: Reduced basis methods for transport dominated problems*, Esaim-Math. Model. Num., 48 (2013), pp. 623–663.

[15] A. Dasgupta, P. Drineas, B. Harb, R. Kumar and M.W. Mahoney, *Sampling Algorithms and Coresets for ℓ_p Regression*, SIAM J. Sci. Comput., 38 (2009), pp. 2060–2078.

[16] M. Deb, I. Babuska, and J. T. Oden, *Solution of stochastic partial differential equations using Galerkin finite element techniques*, Comput. Method. Appl. M., 190 (2001), pp. 6359–6372.

[17] C. Desceliers, R. Ghanem, and C. Soize, *Polynomial chaos representation of a stochastic preconditioner*, Int. J. Numer. Meth. Eng., 64 (2005), pp. 618–634.

[18] R. DeVore, G. Petrova, and P. Wojtaszczyk, *Greedy Algorithms for Reduced Bases in Banach Spaces*, Constr. Approx., 37 (2013), pp. 455–466.

[19] H. Elman and V. Forstall, *Preconditioning Techniques for Reduced Basis Methods for Parameterized Elliptic Partial Differential Equations*, SIAM J. Sci. Comput., 37 (2015), pp. S177–S194.

[20] O. G. Ernst, C. E. Powell, D. Silvester, and E. Ullmann, *Efficient Solvers for a Linear Stochastic Galerkin Mixed Formulation of Diffusion Problems with Random Data*, SIAM J. Sci. Comput., 31 (2009), pp. 1424–1447.

[21] R. Ghanem and R. M. Kruger, *Numerical solution of spectral stochastic finite element systems*, Comput. Method. Appl. M., 129 (1996), pp. 289–303.

[22] L. Giraldi, A. Litvinenko, D. Liu, H. Matthies, and A. Nouy, *To Be or Not to Be Intrusive? The Solution of Parametric and Stochastic Equations—the Plain Vanilla Galerkin Case*, SIAM J. Sci. Comput., 36 (2014), pp. A2720–A2744.

[23] L. Giraldi, A. Nouy, and G. Legrain., *Low-Rank Approximate Inverse for Preconditioning Tensor-Structured Linear Systems*, SIAM J. Sci. Comput., 36 (2014), pp. A1850–A1870.

[24] L. González, *Orthogonal Projections of the Identity: Spectral Analysis and Applications to Approximate Inverse Preconditioning*, SIAM Rev., 48 (2006), pp. 66–75.

[25] M. J. Grote, and T. Huckle, *Parallel Preconditioning with Sparse Approximate Inverses*, SIAM J. Sci. Comput., 18 (1997), pp. 838–853.

[26] M. F. Hutchinson, *A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines*, Commun. Stat. B-Simul., 19 (1990), pp. 433–450.

[27] 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 infsup stability constants*, Comptes Rendus Mathematique, 345 (2007), pp. 473–478.

[28] B. N. Khoromskij and C. Schwab, *Tensor-Structured Galerkin Approximation of Parametric and Stochastic Elliptic PDEs*, SIAM J. Sci. Comput., 33 (2011), pp. 364–385.

[29] Y. Maday, N. C. Nguyen, A. T. Patera, and G. S. H. Pau, *A general multipurpose interpolation procedure: the magic points*, CPAA, 8 (2009), pp. 383–404.

[30] P. Maréchal, and J. Ye, *Optimizing condition numbers*, SIAM J. Optim., 20 (2009), pp. 935–947.

[31] H. G. Matthies and A. Keese, *Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations*, Comput. Method. Appl. M., 194 (2005), pp. 1295–1331.

[32] H. G. Matthies and E. Zander, *Solving stochastic systems with low-rank tensor compression*, Linear Algebra Appl., 436 (2012), pp. 3819–3838.

[33] A. Nouy, *Recent Developments in Spectral Stochastic Methods forthe Numerical Solution of Stochastic Partial Differential Equations*, Arch. Comput. Method. E., 16 (2009), pp. 251–285.

[34] P. Pacciarini and G. Rozza, *Stabilized reduced basis method for parametrized advection-diffusion PDEs*. Comput. Methods Appl. Mech. Eng., 274 (2014), 1-18.

[35] G. Rozza and K. Veroy, *On the stability of the reduced basis method for Stokes equations in parametrized domains*, Comput. Methods Appl. Mech. Eng., 196-7 (2007), pp. 1244–1260.

[36] G. Rozza, D. B. P. Huynh, and A. T. Patera, *Reduced Basis Approximation and a Posteriori Error Estimation for Affinely Parametrized Elliptic Coercive Partial Differential Equations*, Arch. Comput. Method. E., 15 (2008), pp. 229–275.

[37] J. Tropp, *Improved analysis of the Subsampled Randomized Hadamard Transform*, AADA, 03 (2011), pp. 115–126.

[38] K. Veroy, C. Prud'homme, D.V. Rovas, and A.T. Patera, *A Posteriori Error Bounds for Reduced-Basis Approximation of Parametrized Noncoercive and Nonlinear Elliptic Partial Differential Equations*, Proceedings of the 16th AIAA Computational Fluid Dynamics Conference, 03 (2003), pp. 2003–3847.

[39] L. Welch, *Lower bounds on the maximum cross correlation of signals*, IEEE T. Inform. Theory., 20 (1974), pp. 397–399.