Nearly optimal fast preconditioning of symmetric positive definite matrices

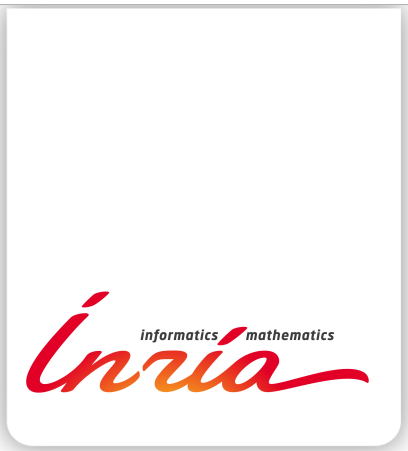
Emmanuel Agullo, Eric Darve, Luc Giraud, Yuval Harness

Nearly optimal fast preconditioning of symmetric positive definite matrices

E. Agullo, Eric Darve, L. Giraud, Y. Harness

RESEARCH REPORT
N° 8984
November 2016
Project-Teams HiePACS

ISSN 0249-6399 ISRN INRIA/RR--8984--FR+ENG



Nearly optimal fast preconditioning of symmetric positive definite matrices

E. Agullo*, Eric Darve†, L. Giraud*, Y. Harness*

Project-Teams HiePACS

Research Report n° 8984 — November 2016 — 34 pages

Abstract: We consider the hierarchical off-diagonal low-rank preconditioning of symmetric positive definite matrices arising from second order elliptic boundary value problems. When the scale of such problems becomes large combined with possibly complex geometry or unstable of boundary conditions, the representing matrix is large and typically ill-conditioned. Multilevel methods such as the hierarchical matrix approximation are often a necessity to obtain an efficient solution. We propose a novel hierarchical preconditioner that attempts to minimize the condition number of the preconditioned system. The method is based on approximating the low-rank off-diagonal blocks in a norm adapted to the hierarchical structure. Our analysis shows that the new preconditioner effectively maps both small and large eigenvalues of the system approximately to 1. Finally through numerical experiments, we illustrate the effectiveness of the new designed scheme which outperforms more classical techniques based on regular SVD to approximate the off-diagonal blocks and SVD with filtering.

Key-words: Fast direct solvers, preconditioning, hierarchical matrices, sparse matrices, low-rank matrices, hierarchical compression

* Inria, France
† Stanford University, USA

RESEARCH CENTRE
BORDEAUX – SUD-OUEST

200 avenue de la Vielle Tour
33405 Talence Cedex

Préconditionneur rapide quasi-optimal pour des systèmes linéaires symétriques définis positifs

Résumé : Nous étudions un préconditionnement "data sparse" (HODLR) pour matrices symétriques définies positives provenant de problèmes aux limites elliptiques du second ordre. Pour des problèmes de grandes tailles, des conditions aux limites quasi-singulières ou des géomètries complexes, les matrices de discrétisation associées sont très mal conditionnées. Le recours à des méthodes multi-niveaux sont souvent une nécessité pour obtenir une solution efficace. Nous proposons un nouveau préconditionneur hiérarchique qui, dans le cas deux niveaux, est mimise le conditionnement du système préconditionné. Dans le cas multi-niveau le preconditionneur tente de conserver cette propriété qui n'est plus prouvée; en revanche nous établissons que les valeurs propres extrémales sont clusterisées dans un intervalle autour de 1. Finalement, à travers des expérimentations numériques, nous illustrons l'efficacité du nouveau schéma proposé qui surpasse les techniques plus classiques basées sur une SVD régulière pour approximer les blocs hors-diagonaux ou une SVD filtrée.

Mots-clés : Solveurs directs rapides, préconditionnement, matrices hiérarchiques, matrices creuses, matrice de rang faible, compression hiérarchique

1 Introduction

The solution of boundary value problems is of immense importance in predictive science and in a wide range of engineering applications, as means to model and understand physical phenomena. Real life complex physical models can only be solved approximately by discretized models which are typically represented by large-scale algebraic systems. In this work we consider such linear systems of the form,

$$Ax = b\,, \tag{1.1}$$

where $A \in \mathbb{R}^{n\times n}$ is a *symmetric positive definite* (SPD) matrix arising from a finite element of finite difference discretization of an elliptic *partial differential equation* (PDE). In many practical applications the matrix A becomes ill-conditioned and, thus, challenging for iterative methods. This is especially true when the number of unknowns is very large and additional complexities such as irregular geometries, unstable boundary or interface conditions, and possibly non-smooth, high-contrast coefficients are considered.

In this work we focus on large-scale SPD linear systems representing discretized elliptic PDEs. For their solution, we consider preconditioned iterative methods, such as the *preconditioned conjugate gradient* (PCG) method [14, 22]. The main contribution of this work is the introduction of a nearly optimal hierarchically low-rank structured preconditioner. The optimality is expressed by two features: (a) the new preconditioner locally minimizes the spectral condition number for any given distribution of low ranks, (b) our method effectively maps both small and large eigenvalues of the original system approximately to 1. The last feature is of great importance to Krylov subspace methods, since it is equivalent to the minimization of the effective degree of the minimal polynomial of A that defines the maximal dimension of the search space. An added value of the method is that given certain conditions are met, the spectral bounds of principal preconditioned subsystems can be analytically estimated. Hence, we have some control over the potential growth of the condition number in the hierarchy.

Hierarchical or multilevel methods are, often, a necessity for obtaining solutions of very large-scale linear systems, due to their capabilities to efficiently process, often in parallel, large partitioned matrices and extract important features from sub-partitions. The class of hierarchical matrix approximations [11, 18], which has gained growing attention in recent years, offers unique advantages over other traditional multilevel methods, e.g., multigrid or algebraic domain decomposition. These include efficient parallelization schemes with provable convergence rates to some of the most challenging complex problems.

Hierarchical matrices are, essentially, data-sparse approximations of a class of dense matrices that often arise in elliptic boundary value problems. These approximations rely on the fact that the matrices can be sub-divided into a hierarchy of smaller block matrices, and certain sub-blocks can be efficiently approximated as low-rank matrices. If $A_{t\times s}$ denotes a sub-block of A associated with a subset of row indices, t, and a subset of column indices, s, then generally it can be approximated by a low-rank matrix if t and s represent well separated coordinates with respect to a given admissibility criterion. Typically, the low-rank property can be rigorously proven for some of the most complex boundary value problems, which leads to provable convergence rates.

A major concern when setting up a preconditioner is to ensure that the preconditioned system has a bounded condition number, and that the number of iterations in an iterative scheme remains small. Essentially, a hierarchical matrix approximation, $\widehat{A}$, is a perturbed version of A. Hence, we are required, in principle, to verify that each sub-block of the input matrix, $A_{t\times s}$, which is replaced by a low-rank block, $\widehat{A}_{t\times s}$, in $\widehat{A}$ satisfies

$$\left\| A_{s\times t} - \widehat{A}_{s\times t} \right\| < \lambda_{\min}\left(A\right)\,, \tag{1.2}$$

with respect to some norm $\|\cdot\|$ where $\lambda_{\min}(A)$ denotes the minimal eigenvalue of the input matrix matrix, A. See [15] for further details. Thus, when the given problem is large and ill-conditioned, conforming with eq. (1.2) can produce a hierarchical preconditioner, $\widehat{A}$, of degraded quality, whose complexity and memory usage are high.

The common practice in hierarchical matrix approximation is to set a distribution of positive tolerances, $\{\tau_{s\times t}\}$, corresponding to the sub-blocks $A_{t\times s}$ that are replaced by low-rank approximations $\widehat{A}_{t\times s}$ such that

$$\left\| A_{s\times t} - \widehat{A}_{s\times t} \right\|_2 < \tau_{s\times t} \cdot \|A_{s\times t}\|_2 \ ,$$

where $\|\cdot\|_2$ denotes the 2-norm. The question of choosing a distribution of tolerances for hierarchical matrix approximations that ensures a bounded number of iterations to convergence using PCG, which is also independent of the number of unknowns has been explored in [5]. The authors consider an elliptic PDE discretization on quasi-uniform finite element mesh with typical mesh size h, and show that for obtaining a uniformly h-independent bounded number of iterations with PCG it is necessary to set $\tau_{t\times s} \sim h^2$ on smaller blocks of the partition and $\tau_{t\times s} \sim h$ on the larger blocks. This is, in fact, an expression for the limitation eq. (1.2). The authors suggest a modified approach which preserves piecewise constant vectors to overcome eq. (1.2). In a recent paper [6] the authors extend and generalize the theory of the method. A similar approach for non-symmetric sparse matrices has been recently suggested in [25].

In the present work we take a different approach, whose key idea is to obtain the low-rank approximations in a properly chosen weighted norm that overcomes the limitation eq. (1.2). The method is algebraic and can be used in a black box fashion without taking into considerations the specific features of the underlined PDE. In fact, our method can be combined with the method suggested in [6] to, potentially, achieve even better results. The last point, however, is not explored in this work. For its simplicity we employ in this work the *hierarchical off-diagonal low-rank* (HODLR) structure, which is the simplest of all hierarchical matrix formats. We exploit the simplicity to analyze spectral properties and derive the optimal preconditioner introduced in Section 3. We also note that the simplicity of HODLR facilitates a highly parallelizable matrix factorization scheme [1].

The paper is organized as follows. In Section 2 we review the HODLR matrix structure for the symmetric case and present some notational conventions. Our theoretical results based on a 2-level analysis are detailed in Section 3. The implementation of the theoretical part along with practical considerations for constructing a nearly optimal multilevel preconditioner are given in Section 4. Our experimental study is given in Section 5 and includes a comparative study of our method with the conventional low-rank SVD approximation and the method suggested in [6] adapted to HODLR. The conclusions and plans for future work follow in Section 6.

2 The HODLR Structure

In this section we review the HODLR matrix structure, which will be employed throughout this article. We focus on the symmetric case, since this work is concerned with the preconditioning of SPD matrices. We also detail the various notations that are used later in the manuscript.

2.1 The Symmetric HODLR Matrix

Consider a symmetric non-singular matrix $A \in \mathbb{R}^{N\times N}$. Its multilevel HODLR approximation, $\widehat{A} \in \mathbb{R}^{N\times N}$, can be described in the following recursive manner,

$$\widehat{A} = \widehat{A}_1^{(0)}\,, \quad \widehat{A}_k^{(\ell)} = \begin{bmatrix} \widehat{A}_{2k-1}^{(\ell+1)} & \widehat{M}_k^{(\ell)} \\ \left(\widehat{M}_k^{(\ell)}\right)^T & \widehat{A}_{2k}^{(\ell+1)} \end{bmatrix} \in \mathbb{R}^{n_k^{(\ell)}\times n_k^{(\ell)}}\,, \quad k = 1,2,\ldots,2^\ell\,,$$

where $\ell = 0,1,\ldots,L-1$ denotes the level of $\widehat{A}_k^{(\ell)}$ in the hierarchy. The submatrices in the lowest level of the hierarchy, $A_k^{(L)}$, are full-rank and the off-diagonal blocks are low-rank matrices of the following form

$$\widehat{M}_k^{(\ell)} = \widehat{U}_k^{(\ell)} \left(\widehat{W}_k^{(\ell)}\right)^T\,, \quad \widehat{U}_k^{(\ell)} \in \mathbb{R}^{n_{2k-1}^{(\ell+1)}\times r_k^{(\ell)}}\,, \quad \widehat{W}_k^{(\ell)} \in \mathbb{R}^{n_{2k}^{(\ell+1)}\times r_k^{(\ell)}}\,, \tag{2.1}$$

where $r_k^{(\ell)}$ is the rank of $\widehat{M}_k^{(\ell)}$. The structure is typically obtained after proper reordering of the matrix rows and columns, which ensures the low-rank assumption,

$$r_k^{(\ell)} \ll n_{2k-1}^{(\ell)}, n_{2k}^{(\ell)}\,. \tag{2.2}$$

For further details on this issue, see [18]. An illustration of the hierarchical structure is displayed in fig. 1.

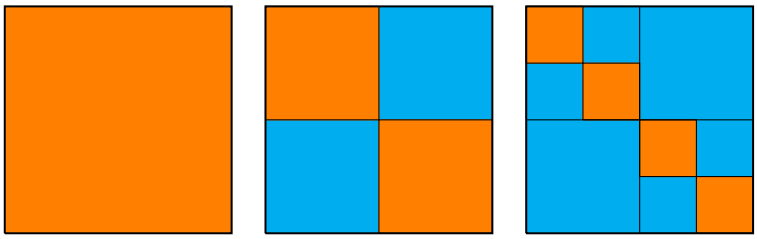


Figure 1: **The HODLR Structure.** The first 3 levels, $\ell = 0,1,2$, of the HODLR structure are illustrated: at each level the blue color blocks are the low rank blocks and the orange blocks are the HODLR principal submatrices of the level, $A_k^{(\ell)}$.

The common practice is to obtain the low rank off-diagonal blocks such that

$$\left\|M_k^{(\ell)} - \widehat{M}_k^{(\ell)}\right\|_2 \le \tau_k^{(\ell)} \cdot \left\|M_k^{(\ell)}\right\|_2\,, \tag{2.3}$$

in the 2-norm where $\tau_k^{(\ell)} > 0$ denotes a chosen tolerance and $M_k^{(\ell)}$ denotes the corresponding off-diagonal block in the properly reordered input matrix, denoted by $A_1^{(0)}$, i.e., $A = A_1^{(0)}$ up to a reordering of rows and columns and $A_1^{(0)}$ has the same hierarchical partitioning as $\widehat{A}_1^{(0)}$,

$$A_k^{(\ell)} = \begin{bmatrix} A_{2k-1}^{(\ell+1)} & M_k^{(\ell)} \\ \left(M_k^{(\ell)}\right)^T & A_{2k}^{(\ell+1)} \end{bmatrix} \in \mathbb{R}^{n_k^{(\ell)}\times n_k^{(\ell)}}\,, \quad k = 1,2,\ldots,2^\ell\,,$$

where $\ell = 0,1,\ldots,L-1$. The submatrices in the lowest level of the hierarchy satisfy $\widehat{A}_k^{(L)} = A_k^{(L)}$ for all $k = 0,1,\ldots,2^L$.

For obtaining the approximation $\widehat{M}_k^{(\ell)}$ satisfying eq. (2.3) the low-rank *singular value decomposition* (SVD) [10] which originated from [7] is, generally, considered the best choice, since it is optimal with respect to any unitarily invariant norm (2-norm, Frobenius). However, the computational cost to construct an SVD is relatively expensive requiring an $\mathcal{O}\left(m^3\right)$ operations, where $m = n_k^{(\ell)}/2$. For this reason a variety of fast approximation algorithms attempting to efficiently obtain a low-rank approximation close enough to the low-rank SVD have been proposed. These include, among others, the rank revealing LU [19], rank revealing QR [12], randomized algorithms [9, 16, 24], adaptive cross approximation [21] and boundary distance low-rank [2]. For more details see a review on this topic in [2]. Note that in this work we are not concerned with the specific method used to obtain the low-rank approximation, $\widehat{M}_k^{(\ell)}$. Rather we wish to explore which norm or what modification are needed for the low-rank approximation to ensure efficient preconditioning that overcomes the limitation eq. (1.2).

2.2 Notations

In this work we consistently use the following conventions and guidelines to denote various quantities:

- The $\widehat{\ }$ accent denotes a data-sparse (either HODLR or low-rank) approximation of the matrix or submatrix of the same notation, e.g.,
  - $\widehat{A}$ is an HODLR approximation of A.
  - $\widehat{A}_i$ is an HODLR approximation of A_i.
  - $\widehat{M}$ is a low-rank approximation of the off-diagonal block M, i.e., by eq. (2.1) and eq. (2.2) satisfies
$$\widehat{M} = \widehat{U}\widehat{W}^T \in \mathbb{R}^{n_1 \times n_2}\,, \quad \widehat{U} \in \mathbb{R}^{n_1 \times r}\,, \quad \widehat{W} \in \mathbb{R}^{n_2 \times r}\,, \quad r \ll n_1, n_2\,.$$
- To ease the reading when considering a principal subsystem in the hierarchy we will often drop superscripts and subscripts and consider the simplified representations
$$\begin{bmatrix} A_1 & M \\ M^T & A_2 \end{bmatrix} = \begin{bmatrix} A_{2k-1}^{(\ell+1)} & M_k^{(\ell)} \\ \left(M_k^{(\ell)}\right)^T & A_{2k}^{(\ell+1)} \end{bmatrix}\,,$$
$$\begin{bmatrix} \widehat{A}_1 & \widehat{M} \\ \widehat{M}^T & \widehat{A}_2 \end{bmatrix} = \begin{bmatrix} \widehat{A}_{2k-1}^{(\ell+1)} & \widehat{M}_k^{(\ell)} \\ \left(\widehat{M}_k^{(\ell)}\right)^T & \widehat{A}_{2k}^{(\ell+1)} \end{bmatrix}\,,$$
  where n_1, n_2, m denote the dimensionality of the partition according to
$$A_1, \widehat{A}_1 \in \mathbb{R}^{n_1 \times n_1}\,, \quad A_2, \widehat{A}_2 \in \mathbb{R}^{n_2 \times n_2}\,, \quad m = \min\{n_1, n_2\}\,.$$
- B denotes an inverse (possibly non-principal) square root of A in the sense that
$$B^T A B = I\,.$$
  Similarly we will denote:
  - B_i as the inverse square root of each principal submatrix A_i,
$$B_i^T A_i B_i = I_i\,, \tag{2.4}$$
    where I_i is the $n_i \times n_i$ identity matrix.

– $\widehat{B}$ and $\widehat{B}_i$ as inverse square roots of $\widehat{A}$ and $\widehat{A}_i$, respectively,

$$\widehat{B}^T \widehat{A} \widehat{B} = I_i \,, \quad \widehat{B}_i^T \widehat{A}_i \widehat{B}_i = I_i \,. \tag{2.5}$$

- To denote spectral properties of various matrices we will use the letter λ:
  - $\lambda_{\min}(H)$ denotes the minimum eigenvalue of a symmetric matrix H.
  - $\lambda_{\max}(H)$ denotes the maximum eigenvalue of a symmetric matrix H.

  the sepctrum of a given symmetric matrix H, i.e., the set of all eigenvalues, will be denoted as $\mathrm{spec}(H)$.

- The Greek letters α and β denote spectral bounds of the preconditioned system,

$$\alpha \leq \lambda_{\min}\left(\widehat{B}^T A \widehat{B}\right) \leq \lambda_{\max}\left(\widehat{B}^T A \widehat{B}\right) \leq \beta \,.$$

  Similarly α_i and β_i denote spectral bounds of the preconditioned principal submatrices,

$$\alpha_i \leq \lambda_{\min}\left(\widehat{B}_i^T A_i \widehat{B}_i\right) \leq \lambda_{\max}\left(\widehat{B}_i^T A_i \widehat{B}_i\right) \leq \beta_i \,.$$

- For the SVD of an $n_1 \times n_2$ matrix we assume the thin representation

$$\mathcal{U} \Sigma \mathcal{V}^T \,, \quad \Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_m) \,,$$

  where:
  - $\sigma_1 \geq \sigma_1 \geq \cdots \geq \sigma_m \geq 0$ are the singular values and $m = \min\{n_1, n_2\}$.
  - $\mathcal{U} \in \mathbb{R}^{n_1 \times m}$ is the matrix of left-singular vectors.
  - $\mathcal{V} \in \mathbb{R}^{n_2 \times m}$ is the matrix of right-singular vectors.

- The $\widetilde{\ }$ accent denotes an alternative data-sparse or low-rank approximation of the matrix or submatrix of the same notation, e.g.,
  - $\widetilde{A}$ is a data-sparse approximation of A.
  - $\widetilde{A}_i$ is a data-sparse approximation of A_i.
  - $\widetilde{M}$ is a low-rank approximation of the off-diagonal block M, i.e., by eq. (2.1) and eq. (2.2) satisfies

$$\widetilde{M} = \widetilde{U} \widetilde{W}^T \in \mathbb{R}^{n_1 \times n_2} \,, \quad \widetilde{U} \in \mathbb{R}^{n_1 \times r} \,, \quad \widetilde{W} \in \mathbb{R}^{n_2 \times r} \,, \quad r \ll n_1, n_2 \,.$$

  This notation is used to express optimality of a specific choice accented by $\widehat{\ }$.

3 Two-level Analysis

In this section we adopt a 2-level view and consider a simplified problem. The matrix and its HODLR approximation are given by,

$$A = \begin{bmatrix} A_1 & M \\ M^T & A_2 \end{bmatrix} \,, \quad \widehat{A} = \begin{bmatrix} \widehat{A}_1 & \widehat{M} \\ \widehat{M}^T & \widehat{A}_2 \end{bmatrix} \,, \quad A_i, \widehat{A}_i \in \mathbb{R}^{n_i \times n_i} \,, \tag{3.1}$$

respectively, where A and the principal blocks $\widehat{A}_i$ $(i = 1, 2)$ are assumed to be SPD.

The goal of this section is to determine for any given rank $r = 1, \ldots, m-1$, which off-diagonal block with rank bounded by r, $\widehat{M}$, ensures that $\widehat{A}$ is SPD and minimizes the spectral condition number of the preconditioned system,

$$\mathrm{cond}_2\left(\widehat{B}^T A \widehat{B}\right) = \left\|\widehat{B}^T A \widehat{B}\right\|_2 \cdot \left\|\widehat{B}^{-1} A^{-1} \widehat{B}^{-T}\right\|_2 , \tag{3.2}$$

where $\| \cdot \|_2$ is the 2-norm and $\widehat{B}$ is an inverse square root of $\widehat{A}$ in the sense of eq. (2.5).

The theory is developed in two steps:

1. In Section 3.1 we assume that $A_i = \widehat{A}_i$ and present a specific choice which we prove to be a (non-unique) minimizer of eq. (3.2). We also show that this specific choice clusters the spectrum by mapping the largest r eigenvalues and smallest r eigenvalues of $\widehat{B}^T A \widehat{B}$ to 1.

2. In Section 3.2 we extend our choice from Section 3.1 to the more general case, $A_i \neq \widehat{A}_i$ $(i = 1, 2)$. We provide necessary and sufficient conditions ensuring $\widehat{A}$ is SPD and obtain estimates of the spectral bounds of $\widehat{B}^T A \widehat{B}$, given the spectral bounds of the preconditioned principal systems $\widehat{B}_i^T A_i \widehat{B}_i$ $(i = 1, 2)$.

The choice made at Section 3.2 is a heuristic. However, when $A_i \approx \widehat{A}_i$ the results of Section 3.2 imply it is nearly optimal. A more optimal choice would, generally, be much more costly to implement and will not necessarily produce a far better preconditioner in terms of convergence. Our experiments in Section 5 indicate that the resulting preconditioner is, indeed, very robust and, in fact, outperforms standard and state-of-the-art alternative methods.

3.1 An Optimal Two-level Preconditioner

Let us assume that $\widehat{A}_i = A_i$, and let us first consider the degenerate case $r = 0$. In this case $\widehat{M} = 0$ and the preconditioner reduces to block Jacobi,

$$\widehat{B} = \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix} , \tag{3.3}$$

where $B_i \in \mathbb{R}^{n_i \times n_i}$ is an inverse square root of A_i as defined in eq. (2.4). There is a known result [8] that shows that the two-sided block Jacobi preconditioner eq. (3.3) is optimal in the sense that

$$\mathrm{cond}_2\left(\widehat{B}^T A \widehat{B}\right) \leq \mathrm{cond}_2\left(\widetilde{B}^T A \widetilde{B}\right) ,$$

for any other non-singular block diagonal matrix with the same partition as $\widehat{B}$,

$$\widetilde{B} = \begin{bmatrix} \widetilde{B}_1 & 0 \\ 0 & \widetilde{B}_2 \end{bmatrix} .$$

The analysis we present, thus, naturally extends this classic result.

The major result given in theorem 1 is that minimizing eq. (3.2) is equivalent to minimizing the condition number of the two-sided block Jacobi preconditioned exact system,

$$\begin{bmatrix} B_1^T & 0 \\ 0 & B_2^T \end{bmatrix} A \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix} = \begin{bmatrix} I_1 & B_1^T M B_2 \\ B_2^T M^T B_1 & I_2 \end{bmatrix} , \tag{3.4}$$

by the two-sided block Jacobi preconditioned approximate system,

$$\begin{bmatrix} B_1^T & 0 \\ 0 & B_2^T \end{bmatrix} \widehat{A} \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix} = \begin{bmatrix} I_1 & B_1^T \widehat{M} B_2 \\ B_2^T \widehat{M}^T B_1 & I_2 \end{bmatrix} .$$

This is obtained by setting the r-rank off-diagonal block $\widehat{M}$ in eq. (3.1) such that

$$B_1^T \widehat{M} B_2 = \mathcal{U}_r \Sigma_r \mathcal{V}_r^T , \quad \mathcal{U}_r \in \mathbb{R}^{n_1 \times r} , \quad \mathcal{V}_r \in \mathbb{R}^{n_2 \times r} ,$$

where $\mathcal{U}_r$ and $\mathcal{V}_r$ are composed of the first r left and right, respectively, singular vectors of the SVD,

$$B_1^T M B_2 = \mathcal{U} \Sigma \mathcal{V}^T , \quad \Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_m) , \quad m = \min\{n_1, n_2\} ,$$

and $\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$.

The second consequence of theorem 1 regards the spectrum of the preconditioned system. Our proof shows that the spectrum of the two-sided block Jacobi preconditioned system eq. (3.4) contains (or equals to)

$$1 + \sigma_1, \ldots, 1 + \sigma_m, 1 - \sigma_m, \ldots, 1 - \sigma_1 ,$$

where $1-\sigma_1$ and $1+\sigma_1$ are the smallest and largest, respectively, eigenvalues of the preconditioned system. Thus, block Jacobi, essentially, redistributes the spectrum of the matrix evenly around 1. The preconditioner we suggest, does the same but also maps the largest r eigenvalues $(1 + \sigma_1, \ldots, 1 + \sigma_r)$ and the smallest r eigenvalues $(1 - \sigma_r, \ldots, 1 - \sigma_1)$ of eq. (3.4) exactly to 1. Thus, the condition number eq. (3.2) as a function of r is

$$\operatorname{cond}_2 \left(\widehat{B}^T A \widehat{B} \right) = \frac{1 + \sigma_{r+1}}{1 - \sigma_{r+1}} ,$$

where $\sigma_1, \ldots, \sigma_m$ are the singular values of $B_1^T M B_2$.

Theorem 1. *Let*

$$A = \begin{bmatrix} A_1 & M \\ M^T & A_2 \end{bmatrix} , \quad \widehat{A} = \begin{bmatrix} A_1 & \widehat{M} \\ \widehat{M}^T & A_2 \end{bmatrix} ,$$

have the same partition where A is SPD, and let B_i denote an inverse square root of A_i in the sense that $B_i^T A_i B_i = I_i$ where I_i is the $n_i \times n_i$ identity matrix.

If the off-diagonal block $\widehat{M}$ satisfies

$$B_1^T \widehat{M} B_2 = \mathcal{U}_r \Sigma_r \mathcal{V}_r^T ,$$

where $\mathcal{U}_r$ and $\mathcal{V}_r$ are composed of the first r left and right, respectively, singular vectors of the SVD,

$$B_1^T M B_2 = \mathcal{U} \Sigma \mathcal{V}^T , \quad \Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_m) , \quad m = \min\{n_1, n_2\} , \tag{3.5}$$

then:

1. *The matrix $\widehat{A}$ is SPD and possesses an inverse square root, $\widehat{B}$,*

$$\widehat{B}^T \widehat{A} \widehat{B} = I .$$

2. *For any $r \geq 1$ and any inverse square root $\widehat{B} \in \mathbb{R}^{n \times n}$, the spectrum of the preconditioned system is contained in $]0, 2[$ and equal to*

$$\{1 + \sigma_{r+1}, \ldots, 1 + \sigma_m, 1, 1 - \sigma_m, \ldots, 1 - \sigma_{r+1}\} .$$

3. The inverse square root $\widehat{B}$ is a minimizer of the spectral condition number in the sense that

$$\mathrm{cond}_2\left(\widehat{B}^T A \widehat{B}\right) \leq \mathrm{cond}_2\left(\widetilde{B}^T A \widetilde{B}\right),$$

for any $\widetilde{B}$ satisfying $\widetilde{B}^T \widetilde{A} \widetilde{B} = I$, where $\widetilde{A}$ is a partitioned SPD matrix of the same partition as $\widehat{A}$ eq. (3.1)*, whose off-diagonal blocks rank is bounded by r*

$$\widetilde{A} = \begin{bmatrix} A_1 & \widetilde{M} \\ \widetilde{M}^T & A_2 \end{bmatrix}, \quad \mathrm{rk}(\widetilde{M}) \leq r\,.$$

Remark 1. *If we denote*

$$\boldsymbol{U} = B_1^{-T}\mathcal{U}\,, \quad \boldsymbol{V} = B_2^{-T}\mathcal{V}\,,$$

we obtain the decompositions

$$M = \boldsymbol{U}\Sigma\boldsymbol{V}^T\,, \quad \widehat{M} = \boldsymbol{U}_r\Sigma\boldsymbol{V}_r^T\,, \tag{3.6}$$

where $\boldsymbol{U}_r$ and $\boldsymbol{V}_r$ are composed of the first r columns of $\boldsymbol{U}$ and $\boldsymbol{V}$, respectively. Hence, the columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ form orthonormal bases with respect to the inner product induced by A_1^{-1} and A_2^{-1}, respectively:

$$\boldsymbol{U}^T A_1^{-1}\boldsymbol{U} = I_1\,, \quad \boldsymbol{V}^T A_2^{-1}\boldsymbol{V} = I_2\,.$$

The decomposition eq. (3.6) *is known as the weighted version of the* generalized singular value decomposition *[17] of M with respect to A_1^{-1} and A_2^{-1}.*

The distinct property of the preconditioner suggested in theorem 1 is that the bases representing the low-rank off-diagonal blocks are orthogonal with respect to a weighted inner product. Thus, we will denote the new preconditioner as the *weighted singular value decomposition* (WSVD) preconditioner. The usage of the weighted inner product ensures the minimality of the spectral condition number and an optimal spectral clustering, as defined in theorem 1. An illustration of the spectral clustering done by the WSVD preconditioner is displayed in fig. 2.

The proof of theorem 1 is based on the *Cauchy (eigenvalue) interlacing theorem* [20, p. 202] which asserts that the eigenvalues of any principal submatrix of a symmetric matrix interlace those of the symmetric matrix. To be precise, if $H \in \mathbb{R}^{n\times n}$ is a partitioned symmetric matrix of the following form

$$H = \begin{bmatrix} E & F \\ F^T & G \end{bmatrix},$$

in which E is a $r \times r$ principal submatrix, then for each $i = 1, \ldots, r$,

$$\lambda_i(H) \geq \lambda_i(E) \geq \lambda_{i+n-r}(H)\,, \tag{3.7}$$

where the eigenvalues of the symmetric matrix H are assumed to be arranged in a decreasing order:

$$\lambda_1(H) \geq \lambda_2(H) \geq \cdots \geq \lambda_n(H)\,.$$

The proof itself is verified with the aid of the next lemma.

Lemma 1. *Let $\mathcal{C} = \left[\begin{smallmatrix} \delta I_1 & \mathcal{M} \\ \mathcal{M}^T & \delta I_2 \end{smallmatrix}\right] \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$ where I_i denotes the $n_i \times n_i$ identity matrix and $\delta \in \mathbb{R}$. If $n_1 = n_2$ then*

$$\mathrm{spec}(\mathcal{C}) = \{\delta - \sigma_1, \ldots, \delta - \sigma_m, \delta + \sigma_m, \ldots, \delta + \sigma_1\}\,,$$

where $m = n_1 = n_2$. Otherwise

$$\mathrm{spec}(\mathcal{C}) = \{\delta - \sigma_1, \ldots, \delta - \sigma_m, \delta + \sigma_m, \ldots, \delta + \sigma_1\} \cup \{\delta\}\,,$$

where $m = \min\{n_1, n_2\}$ and the multiplicity of the eigenvalue δ is at least $|n_1 - n_2|$.

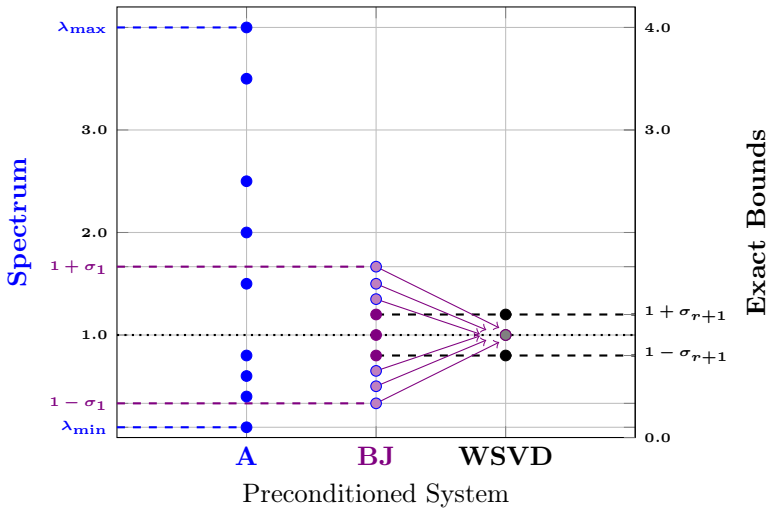


Figure 2: **Spectrum Clustering for $\widehat{\boldsymbol{A}}_{\boldsymbol{i}} = \boldsymbol{A}_{\boldsymbol{i}}$.** The spectrum of some SPD matrix A and the transformations it goes after preconditioning by block Jacobi (BJ) and the WSVD preconditioner are displayed. The spectra are ordered from the left to the right starting from A, followed by BJ and ends up with WSVD.

Proof. of lemma 1.
Let us assume without the loss of generality that $n_1 \geq n_2 = m$ and let

$$\mathcal{M} = \mathcal{U}\Sigma\mathcal{V}^T\,, \quad \mathcal{U} \in \mathbb{R}^{n_1 \times n_2}\,, \quad \mathcal{V} \in \mathbb{R}^{n_2 \times n_2}\,,$$

denote the SVD of $\mathcal{M}$. Let

$$\mathcal{W} = \frac{1}{\sqrt{2}} \begin{bmatrix} \widetilde{\mathcal{U}} & \mathcal{U} \\ \widetilde{\mathcal{V}} & -\mathcal{V} \end{bmatrix} \in \mathbb{R}^{n \times n}\,, \tag{3.8}$$

whose blocks are given by

$$\widetilde{\mathcal{U}} = \begin{cases} \mathcal{U} & \text{if} \quad n_1 = n_2 \\ \begin{bmatrix} \mathcal{U} & \sqrt{2}\mathcal{U}^{\perp} \end{bmatrix} & \text{if} \quad n_1 > n_2 \end{cases}\,, \quad \widetilde{\mathcal{V}} = \begin{cases} \mathcal{V} & \text{if} \quad n_1 = n_2 \\ \begin{bmatrix} \mathcal{V} & 0 \end{bmatrix} & \text{if} \quad n_1 > n_2 \end{cases}\,,$$

where $\mathcal{U}^{\perp}$ is an $n_1 \times (n_1 - n_2)$ matrix with orthonormal columns, whose range is orthogonal to the range of $\mathcal{U}$,

$$\mathcal{U}^T\mathcal{U}^{\perp} = 0 \in \mathbb{R}^{n_2 \times (n_1 - n_2)}\,.$$

Direct calculations show that $\mathcal{W}$ is an orthonormal matrix satisfying

$$\mathcal{W}^T \begin{bmatrix} 0 & \mathcal{M} \\ \mathcal{M}^T & 0 \end{bmatrix} \mathcal{W} = \begin{bmatrix} \mathcal{S}_{1,m} & 0 \\ 0 & -\mathcal{S}_{2,m} \end{bmatrix}\,,$$

where $\mathcal{S}_{i,m} = \operatorname{diag}\big[\sigma_1, \ldots, \sigma_m, 0, \ldots, 0\big] \in \mathbb{R}^{n_i \times n_i}$. Thus, by the orthogonality of $\mathcal{W}$ we obtain

$$\mathcal{W}^T \begin{bmatrix} \delta I_1 & \mathcal{M} \\ \mathcal{M}^T & \delta I_2 \end{bmatrix} \mathcal{W} = \begin{bmatrix} \delta I_1 + \mathcal{S}_{1,m} & 0 \\ 0 & \delta I_2 - \mathcal{S}_{2,m} \end{bmatrix}\,.$$

Hence, the spectrum of $\mathcal{C}$ is given by

$$\operatorname{spec}(\mathcal{C}) = \{\delta - \sigma_1, \ldots, \delta - \sigma_m, \delta + \sigma_m, \ldots, \delta + \sigma_1\} \cup \{\delta\}\,,$$

where the multiplicity of δ is at least $n_1 - n_2$. □

Proof. of theorem 1.
Let $\widetilde{A}$ be a partitioned SPD matrix with the same partition as $\widehat{A}$ eq. (3.1) whose off-diagonal blocks rank is bounded by r,

$$\widetilde{A} = \begin{bmatrix} A_1 & \widetilde{M} \\ \widetilde{M}^T & A_2 \end{bmatrix}, \quad \mathrm{rk}(\widetilde{M}) \leq r\,.$$

If $(\lambda, \zeta) \in \mathbb{R} \times \mathbb{R}^n$ is an eigenpair of the preconditioned system

$$\widetilde{B}^T A \widetilde{B}\,, \tag{3.9}$$

where $\widetilde{B}$ is an inverse square root of $\widetilde{A}$ in the sense of eq. (2.5), then by employing the change of variables, $\zeta = \widetilde{B}^{-1}\xi$, we obtain

$$\widetilde{B}^T A \widetilde{B}\zeta = \lambda\zeta \;\Leftrightarrow\; \widetilde{B}^T A\xi = \lambda \widetilde{B}^{-1}\xi \;\Leftrightarrow\; \widetilde{B}\widetilde{B}^T A\xi = \lambda\xi\,.$$

Since $\widetilde{B}\widetilde{B}^T = \widetilde{A}^{-1}$, we conclude that regardless to the specific choice of inverse square root, $\widetilde{B}$, the spectrum of the preconditioned system eq. (3.9) remains unchanged.

Let $B_i \in \mathbb{R}^{n_i \times n_i}$ denote a, generally, non-symmetric inverse square root of A_i in the sense of eq. (3.9). By direct calculations we obtain

$$\begin{bmatrix} B_1^T & 0 \\ 0 & B_2^T \end{bmatrix} \widetilde{A} \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix} = \begin{bmatrix} I_1 & \mathcal{M} \\ \mathcal{M}^T & I_2 \end{bmatrix} = \widetilde{\mathcal{A}}\,,$$

and by lemma 1, the spectrum of $\widetilde{\mathcal{A}}$ is contained in (or equal to)

$$\{1 + \widetilde{\sigma}_1, \ldots, 1 + \widetilde{\sigma}_r, 1, 1 - \widetilde{\sigma}_r, \ldots, 1 - \widetilde{\sigma}_1\}\,,$$

where $\widetilde{\sigma}_i$ are the singular values of $\widetilde{\mathcal{M}} = B_1^T \widetilde{M} B_2$. Since $\widetilde{A}$ is SPD and B_i are non-singular, $\widetilde{\mathcal{A}}$ is SPD as well, hence

$$\lambda_{\min}\left(\widetilde{\mathcal{A}}\right) = 1 - \widetilde{\sigma}_1 > 0 \quad \Rightarrow \quad \mathrm{spec}(\widetilde{\mathcal{A}}) \subset]0, 2[\,.$$

Consider the specific choice of inverse square root

$$\widetilde{B} = \begin{bmatrix} B_1^T & 0 \\ 0 & B_2^T \end{bmatrix} \widetilde{\mathcal{W}} \widetilde{\mathcal{D}}^{-1/2} \widetilde{\mathcal{W}}^T, \quad \widetilde{\mathcal{D}} = \begin{bmatrix} I_1 + \widetilde{\mathcal{S}}_{1,r} & 0 \\ 0 & I_2 - \widetilde{\mathcal{S}}_{2,r} \end{bmatrix},$$

where $\widetilde{\mathcal{W}}^T$ is an orthogonal matrix of the same form as eq. (3.8) as defined in the proof of lemma 1, and $\widetilde{\mathcal{S}}_{i,r} = \mathrm{diag}\big[\widetilde{\sigma}_1, \ldots, \widetilde{\sigma}_r, 0, \ldots, 0\big] \in \mathbb{R}^{n_i \times n_i}$. Using once more lemma 1 and the fact that the 2-norm is unitarily invariant, we obtain

$$\mathrm{cond}_2(\widetilde{B}^T A \widetilde{B}) = \left\| \widetilde{\mathcal{D}}^{-1/2} H \widetilde{\mathcal{D}}^{-1/2} \right\|_2 \left\| \widetilde{\mathcal{D}}^{1/2} H^{-1} \widetilde{\mathcal{D}}^{1/2} \right\|_2,$$

where H is an SPD matrix given by

$$H = \widetilde{\mathcal{W}}^T \mathcal{W} \mathcal{D} \mathcal{W}^T \widetilde{\mathcal{W}}, \quad \mathcal{D} = \begin{bmatrix} I_1 + \mathcal{S}_{1,m} & 0 \\ 0 & I_2 - \mathcal{S}_{2,m} \end{bmatrix}.$$

The matrices $\mathcal{W}$ eq. (3.8) and $\mathcal{S}_{i,m}$ are defined and constructed in the proof of lemma 1. Note that like $\widetilde{\mathcal{W}}$, the matrix $\mathcal{W}$ is orthogonal. Hence, the product $\mathcal{W}^T\widetilde{\mathcal{W}}$ is also an orthogonal matrix.

Our definitions so far indicate that the following diagonal matrices,

$$\underline{\mathcal{D}} = \begin{bmatrix} I_1 & 0 \\ 0 & (I_2 - \widetilde{\mathcal{S}}_{2,r}) \end{bmatrix}, \quad \overline{\mathcal{D}} = \begin{bmatrix} (I_1 + \widetilde{\mathcal{S}}_{1,r}) & 0 \\ 0 & I_2 \end{bmatrix}.$$

bound the diagonal matrix $\widetilde{\mathcal{D}}$

$$\underline{\mathcal{D}} \leq \widetilde{\mathcal{D}} \leq \overline{\mathcal{D}}$$

in the sense that $\left(\widetilde{\mathcal{D}} - \underline{\mathcal{D}}\right)$ and $\left(\overline{\mathcal{D}} - \widetilde{\mathcal{D}}\right)$ are non-negative definite. Thus, applying the change of variables $\xi = \widetilde{\mathcal{D}}^{-1}x$ we can write

$$\begin{aligned} \left\|\widetilde{\mathcal{D}}^{-1/2} H \widetilde{\mathcal{D}}^{-1/2}\right\|_2 &= \max_{x\neq 0} \frac{x^T \widetilde{\mathcal{D}}^{-1/2} H \widetilde{\mathcal{D}}^{-1/2} x}{x^T x} = \max_{\xi \neq 0} \frac{\xi^T H \xi}{\xi^T \widetilde{\mathcal{D}} \xi} \\ &\geq \max_{\xi\neq 0} \frac{\xi^T H \xi}{\xi^T \overline{\mathcal{D}} \xi} = \max_{y\neq 0} \frac{y^T \overline{\mathcal{D}}^{-1/2} H \overline{\mathcal{D}}^{-1/2} y}{y^T y} = \left\|\overline{\mathcal{D}}^{-1/2} H \overline{\mathcal{D}}^{-1/2}\right\|_2, \end{aligned} \tag{3.10}$$

where $y = \overline{\mathcal{D}}^{1/2}\xi$. Using the same arguments it can also be shown that

$$\left\|\widetilde{\mathcal{D}}^{1/2} H^{-1} \widetilde{\mathcal{D}}^{1/2}\right\|_2 \geq \left\|\underline{\mathcal{D}}^{1/2} H^{-1} \underline{\mathcal{D}}^{1/2}\right\|_2.$$

Let $\overline{Z} = \mathrm{span}\{e_{r+1}, \ldots, e_n\}$ where e_i denotes the i-th canonical basis vector. By the Cauchy interlacing theorem eq. (3.7),

$$\begin{aligned} \left\|\overline{\mathcal{D}}^{-1/2} H \overline{\mathcal{D}}^{-1/2}\right\|_2 &= \max_{y\neq 0} \frac{y^T \overline{\mathcal{D}}^{-1/2} H \overline{\mathcal{D}}^{-1/2} y}{y^T y} \\ &\geq \max_{P_Z y \neq 0} \frac{(P_Z y)^T \overline{\mathcal{D}}^{-1/2} H \overline{\mathcal{D}}^{-1/2} P_Z y}{(P_Z y)^T P_Z y} = \max_{P_Z y \neq 0} \frac{(P_Z y)^T H P_Z y}{(P_Z y)^T P_Z y} \\ &\geq \lambda_{r+1}(H) = 1 + \sigma_{r+1}, \end{aligned} \tag{3.11}$$

where P_Z is the orthogonal projection on the subspace Z. Applying similar arguments it can also be shown that

$$\left\|\underline{\mathcal{D}}^{1/2} H^{-1} \underline{\mathcal{D}}^{1/2}\right\|_2 \geq \lambda_{r+1}(H^{-1}) = \frac{1}{1-\sigma_{r+1}}.$$

Thus, we obtain the lower bound

$$\mathrm{cond}_2(\widetilde{B}^T A \widetilde{B}) \geq \frac{1+\sigma_{r+1}}{1-\sigma_{r+1}},$$

Finally, setting $\widetilde{\mathcal{M}} = \mathcal{U}_r \Sigma \mathcal{V}_r^T$ where $\mathcal{U}_r$ and $\mathcal{V}_r$ are composed of the first r columns of $\mathcal{U}$ and $\mathcal{V}$, respectively, in the SVD of $\mathcal{M} = B_1^T M B_2$,

We have $\widetilde{\sigma}_i = \widehat{\sigma}_i = \sigma_i$, $i = 1, \ldots, r$. Thus, setting accordingly $\widetilde{\mathcal{W}} = \widehat{\mathcal{W}} = \mathcal{W}$ we obtain by direct calculations:

$$\mathrm{cond}_2(\widetilde{B}^T A \widetilde{B}) = \mathrm{cond}_2(\widehat{B}^T A \widehat{B}) = \left\|\widehat{B}^T A \widehat{B}\right\|_2 \left\|\widehat{B}^{-1} A^{-1} \widehat{B}^{-T}\right\|_2 = \frac{1+\sigma_{r+1}}{1-\sigma_{r+1}},$$

and the proof is complete. □

3.2 The Two-level Preconditioner in Practice

In this subsection we consider the case $\widehat{A}_i \neq A_i$, which represents a more practical view. Indeed, if the matrices eq. (3.1) are submatrices of a multilevel SPD HODLR approximation for some ℓ and k,

$$\begin{bmatrix} A_1 & M \\ M^T & A_2 \end{bmatrix} = \begin{bmatrix} A_{2k-1}^{(\ell+1)} & M_k^{(\ell)} \\ \left(M_k^{(\ell)}\right)^T & A_{2k}^{(\ell+1)} \end{bmatrix}, \quad \begin{bmatrix} \widehat{A}_1 & \widehat{M} \\ \widehat{M}^T & \widehat{A}_2 \end{bmatrix} = \begin{bmatrix} \widehat{A}_{2k-1}^{(\ell+1)} & \widehat{M}_k^{(\ell)} \\ \left(\widehat{M}_k^{(\ell)}\right)^T & \widehat{A}_{2k}^{(\ell+1)} \end{bmatrix},$$

unless $\ell = L-1$ we can not assume that $\widehat{A}_i = A_i$. Clearly the important case is $\ell = 0$, since we are ultimately interested in preconditioning the global matrix, $A_1^{(0)}$.

We begin with the definition of the WSVD preconditioner in the case $\widehat{A}_i \neq A_i$, which is motivated by the optimality result of theorem 1. We proceed with lemma 2 which gives conditions ensuring that the WSVD preconditioner is SPD assuming the spectral bounds of the lower level preconditioned principal submatrices,

$$\alpha_i \leq \lambda_{\min}\left(\widehat{B}_i^T A_i \widehat{B}_i\right) \leq \lambda_{\max}\left(\widehat{B}_i^T A_i \widehat{B}_i\right) \leq \beta_i\,,$$

are known. theorem 2 completes the picture in the spirit of theorem 1 stating that the WSVD preconditioner, essentially, maps both the r largest and the r smallest eigenvalues to a closed segment containing 1. When this segment is small, the preconditioner, essentially, retains optimality or near optimality. The results of theorem 2, also indicate that the sensitivity of the spectral bounds to the inaccuracies $\widehat{A}_i \neq A_i$ is governed by the *Cauchy-Bunyakowski-Schwarz* (CBS) constant.

Definition 1. *Let*

$$A = \begin{bmatrix} A_1 & M \\ M^T & A_2 \end{bmatrix}, \quad \widehat{A} = \begin{bmatrix} \widehat{A}_1 & \widehat{M} \\ \widehat{M}^T & \widehat{A}_2 \end{bmatrix},$$

have the same partition where A and $\widehat{A}_i \in \mathbb{R}^{n_i \times n_i}$ $(i = 1, 2)$ are SPD, and let $\widehat{B}_i$ denote an inverse square root of $\widehat{A}_i$ in the sense that $\widehat{B}_i^T \widehat{A}_i \widehat{B}_i = I_i$ where I_i is the $n_i \times n_i$ identity matrix.

If the off-diagonal block $\widehat{M}$ satisfies

$$\widehat{B}_1^T \widehat{M} \widehat{B}_2 = \widehat{\mathcal{U}}_r \widehat{\Sigma}_r \widehat{\mathcal{V}}_r^T\,, \tag{3.12}$$

where $\widehat{\mathcal{U}}_r$ and $\widehat{\mathcal{V}}_r$ are composed of the first r left and right, respectively, singular vectors of the SVD,

$$\widehat{B}_1^T M \widehat{B}_2 = \widehat{\mathcal{U}} \widehat{\Sigma} \widehat{\mathcal{V}}^T\,, \quad \widehat{\Sigma} = \operatorname{diag}(\widehat{\sigma}_1, \ldots, \widehat{\sigma}_m)\,, \quad m = \min\{n_1, n_2\}\,, \tag{3.13}$$

we say $\widehat{A}$ is the weighted SVD (WSVD) HODLR preconditioner of A.

Clearly, when $\widehat{A}_i = A_i$ eq. (3.12) coincides with eq. (3.5). Note that by our fundamental assumptions the submatrices $\widehat{A}_i$ $(i = 1, 2)$ are SPD, hence the corresponding inverse square roots $\widehat{B}_i \in \mathbb{R}^{n_i \times n_i}$ $(i = 1, 2)$ exist. Now, assuming that $\widehat{A}$ is SPD as well, there exists an inverse square root, $\widehat{B}$, satisfying $\widehat{B}^T \widehat{A} \widehat{B} = I$ whose corresponding preconditioned system is

$$\widehat{B}^T A \widehat{B}\,. \tag{3.14}$$

Validating that $\widehat{A}$ is SPD is essential and can be established with the following lemma.

Lemma 2. *Let*

$$A = \begin{bmatrix} A_1 & M \\ M^T & A_2 \end{bmatrix}, \quad \widehat{A} = \begin{bmatrix} \widehat{A}_1 & \widehat{M} \\ \widehat{M}^T & \widehat{A}_2 \end{bmatrix},$$

have the same partition where A and $\widehat{A}_i \in \mathbb{R}^{n_i \times n_i}$ $(i = 1, 2)$ are SPD, and let $\widehat{B}_i$ denote an inverse square root of $\widehat{A}_i$ in the sense that $\widehat{B}_i^T \widehat{A}_i \widehat{B}_i = I_i$ where I_i is the $n_i \times n_i$ identity matrix.

Let us assume that we are given real positive constants,

$$0 < \alpha_1, \alpha_2 \leq 1 \leq \beta_1, \beta_2\,, \tag{3.15}$$

such that

$$0 < \alpha_i x_i^T \widehat{A}_i x_i \leq x_i^T A_i x_i \leq \beta_i x_i^T \widehat{A}_i x_i \quad \forall\, x_i \in \mathbb{R}^{n_i}$$

and let

$$\underline{A} = \begin{bmatrix} \alpha_1 \widehat{A}_1 & M \\ M^T & \alpha_2 \widehat{A}_2 \end{bmatrix}, \quad \overline{A} = \begin{bmatrix} \beta_1 \widehat{A}_1 & M \\ M^T & \beta_2 \widehat{A}_2 \end{bmatrix}. \tag{3.16}$$

Assuming $\widehat{B}_1^T \widehat{M} \widehat{B}_2 = \widehat{\mathcal{U}}_r \widehat{\Sigma}_r \widehat{\mathcal{V}}_r^T$ for some r as defined in eq. (3.12)*, we have:*

1. *The matrices $\underline{A}$, $\widehat{A}$, $\overline{A}$ are SPD iff*

$$\widehat{\sigma}_1 < \sqrt{\alpha_1 \alpha_2}\,, \quad \widehat{\sigma}_1 < 1\,, \quad \widehat{\sigma}_1 < \sqrt{\beta_1 \beta_2}\,,$$

   respectively, where $\widehat{\sigma}_1$ is the largest singular value in eq. (3.13)*.*

2. *If $\underline{A}$, $\widehat{A}$, $\overline{A}$ are SPD, there exist two positive constants, α and β, such that*

$$\alpha = \min_{x \neq 0} \frac{x^T \underline{A} x}{x^T \widehat{A} x} \leq \lambda_{\min}(\widehat{B}^T A \widehat{B}) \leq \lambda_{\max}(\widehat{B}^T A \widehat{B}) \leq \max_{x \neq 0} \frac{x^T \overline{A} x}{x^T \widehat{A} x} = \beta\,. \tag{3.17}$$

Remark 2. *The justification for assumption eq.* (3.15) *will be given in Section 4.*

The following theorem completes the results of lemma 2 and, in fact, constitutes an extension of theorem 1 to the case $\widehat{A}_i \neq A_i$ $(i = 1, 2)$.

Theorem 2. *Let*

$$A = \begin{bmatrix} A_1 & M \\ M^T & A_2 \end{bmatrix}, \quad \widehat{A} = \begin{bmatrix} \widehat{A}_1 & \widehat{M} \\ \widehat{M}^T & \widehat{A}_2 \end{bmatrix},$$

have the same partition where A and $\widehat{A}_i \in \mathbb{R}^{n_i \times n_i}$ $(i = 1, 2)$ are SPD, and let $\widehat{B}_i$ denote an inverse square root of $\widehat{A}_i$ in the sense that $\widehat{B}_i^T \widehat{A}_i \widehat{B}_i = I_i$ where I_i is the $n_i \times n_i$ identity matrix.

Let us assume that we are given real positive constants,

$$0 < \alpha_1, \alpha_2 \leq 1 \leq \beta_1, \beta_2\,,$$

such that

$$0 < \alpha_i x_i^T \widehat{A}_i x_i \leq x_i^T A_i x_i \leq \beta_i x_i^T \widehat{A}_i x_i \quad \forall\, x_i \in \mathbb{R}^{n_i}\,,$$

and let $\widehat{B}_1^T \widehat{M} \widehat{B}_2 = \widehat{\mathcal{U}}_r \widehat{\Sigma}_r \widehat{\mathcal{V}}_r^T$ for some $r = 1, \ldots, m$ as defined in eq. (3.12) *such that*

$$\widehat{\sigma}_1 < \sqrt{\alpha_1 \alpha_2}\,, \tag{3.18}$$

and $\widehat{\sigma}_{r+1} > 0$. *Then the spectral bounds eq.* (3.17) *are given by*

$$\alpha = \min\left\{ \alpha_{1,2}^{\text{avg}} - \sqrt{\widehat{\sigma}_{r+1}^2 + \left(\alpha_{1,2}^{\text{dif}}\right)^2} \quad , \quad \frac{\alpha_{1,2}^{\text{avg}} + \widehat{\sigma}_1}{1+\widehat{\sigma}_1} - \delta_{\alpha}^{1,2} \right\}, \tag{3.19}$$

$$\beta = \max\left\{ \beta_{1,2}^{\text{avg}} + \sqrt{\widehat{\sigma}_{r+1}^2 + \left(\beta_{1,2}^{\text{dif}}\right)^2} \quad , \quad \frac{\beta_{1,2}^{\text{avg}} - \widehat{\sigma}_1}{1-\widehat{\sigma}_1} + \delta_{\beta}^{1,2} \right\}, \tag{3.20}$$

where

$$\gamma_{1,2}^{\text{avg}} = \frac{\gamma_1 + \gamma_2}{2}, \quad \gamma_{1,2}^{\text{dif}} = \frac{\gamma_1 - \gamma_2}{2}, \quad (\gamma = \alpha \ \ or \ \ \beta),$$

and

$$\delta_{\gamma}^{1,2} = \sqrt{\frac{1}{4}\left|\frac{\gamma_{1,2}^{\text{avg}} - \widehat{\sigma}_1}{1-\widehat{\sigma}_1} - \frac{\gamma_{1,2}^{\text{avg}} + \widehat{\sigma}_1}{1+\widehat{\sigma}_1}\right|^2 + \frac{\left(\gamma_{1,2}^{\text{dif}}\right)^2}{1-\widehat{\sigma}_1^2}} - \frac{1}{2}\left|\frac{\gamma_{1,2}^{\text{avg}} - \widehat{\sigma}_1}{1-\widehat{\sigma}_1} - \frac{\gamma_{1,2}^{\text{avg}} + \widehat{\sigma}_1}{1+\widehat{\sigma}_1}\right|.$$

From theorem 2 we observe that each estimated bound, α or β, is a minimum or a maximum, respectively, of two competing terms: the first depends on the largest singular value, $\widehat{\sigma}_1$, and the second is a function of the truncation error, $\widehat{\sigma}_{r+1}$. In fact, when the truncation error becomes sufficiently small it does not affect the values of the bounds, which are governed solely by the terms depending on the largest singular value. Thus in this case, improving the approximation by increasing the rank r does not improve the corresponding condition number estimate, β/α. An illustration of this observation is given in fig. 3.

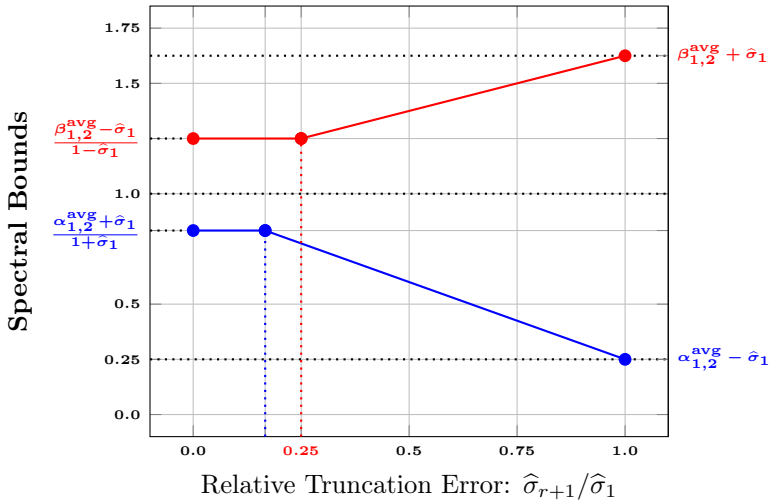


Figure 3: **Spectral Bounds.** A typical behavior of the spectral bounds displayed for the case $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$. The lower bound α eq. (3.19) vs. $\widehat{\sigma}_{r+1}/\widehat{\sigma}_1$ is plotted in blue, and the upper bound β eq. (3.20) vs. $\widehat{\sigma}_{r+1}/\widehat{\sigma}_1$ is plotted in red.

The last observation as displayed in fig. 3 indicates that the value of $\widehat{\sigma}_1$ is central to the estimation of the spectral bounds, and effectively dominates the condition number of the preconditioned system. In particular when $\widehat{\sigma}_1 \to 1$ the bounds α eq. (3.19) and β eq. (3.20) can become very small and very large, respectively. In this sense $\widehat{\sigma}_1$ reflects the sensitivity of the condition number of the preconditioned system eq. (3.14) to the lower level inaccuracies $\widehat{A}_i \neq A_i$ $(i = 1, 2)$.

It is important to note that σ_1 eq. (3.5) and $\widehat{\sigma}_1$ eq. (3.13) are, in fact, the the so-called *Cauchy-Bunyakowski-Schwarz* (CBS) constants of the matrices A and $\widehat{A}$, respectively. The CBS

constant originated from the theory of *Algebraic Multilevel Iterations Methods* [3, 4], its formal definition (with respect to the matrix A) is

$$\gamma = \sup_{x_1,x_2\neq 0} \frac{x_1^T M x_2}{\sqrt{x_1^T A_1 x_1}\sqrt{x_2^T A_2 x_2}} \geq 0\,, \tag{3.21}$$

and a similar definition follows for the matrix $\widehat{A}$. Definition eq. (3.21) coincides with the *principal angle* (cosine of the smallest angle) between the column space of $\begin{bmatrix} I_1 & 0 \end{bmatrix}^T$ and the column space of $\begin{bmatrix} 0 & I_2 \end{bmatrix}^T$ with respect to the inner product $\langle x, y\rangle_A = y^T A x$. Thus, γ represents the local contribution of the upper level to the overall condition number. Now, combining eq. (3.21) with the assumptions of lemma 2 leads to the following relation

$$\frac{1}{\sqrt{\beta_1\beta_2}} \leq \frac{\widehat{\sigma}_1}{\sigma_1} \leq \frac{1}{\sqrt{\alpha_1\alpha_2}}\,,$$

where α_i and β_i are the bounds eq. (3.15). The important conclusion here is that σ_1 and $\widehat{\sigma}_1$ are correlated where σ_1 is intrinsicly predetermined by A and the chosen partition. If $\widehat{A}$ is close to A then we can expect $\widehat{\sigma}_1$ to be close to σ_1, and in this case we have little influence over its value.

Regarding the spectrum of the preconditioned system, the interpretation of theorem 2 is similar to the interpretation of theorem 1. From the proof it can be inferred that two-sided block Jacobi (i.e., the case $r = 0$) effectively maps the spectra of the bounding preconditioned systems to two segments centered around $\alpha_{1,2}^{\text{avg}}$ and $\beta_{1,2}^{\text{avg}}$,

$$\operatorname{spec}\left(\widehat{B}_{\text{BJ}}^T \underline{A} \widehat{B}_{\text{BJ}}\right) \subset \left[\alpha_{1,2}^{\text{avg}} - \sqrt{\widehat{\sigma}_1^2 + \left(\alpha_{1,2}^{\text{dif}}\right)^2}\ ,\ \alpha_{1,2}^{\text{avg}} + \sqrt{\widehat{\sigma}_1^2 + \left(\alpha_{1,2}^{\text{dif}}\right)^2}\right],$$

$$\operatorname{spec}\left(\widehat{B}_{\text{BJ}}^T \overline{A} \widehat{B}_{\text{BJ}}\right) \subset \left[\beta_{1,2}^{\text{avg}} - \sqrt{\widehat{\sigma}_1^2 + \left(\beta_{1,2}^{\text{dif}}\right)^2}\ ,\ \beta_{1,2}^{\text{avg}} + \sqrt{\widehat{\sigma}_1^2 + \left(\beta_{1,2}^{\text{dif}}\right)^2}\right],$$

where $\operatorname{spec}(H)$ denotes the spectrum of the symmetric matrix H, and $\widehat{B}_{\text{BJ}} = \widehat{B}(r = 0)$. The WSVD preconditioner $\widehat{A}$ as described in definition 1 does the same but also maps the r largest and r smallest eigenvalues of $\overline{A}$ eigenvalues of $\underline{A}$ to the segments

$$\left[\frac{\widehat{\sigma}_1 + \alpha_{1,2}^{\text{avg}}}{1 + \widehat{\sigma}_1} - \delta_\alpha^{1,2}\ ,\ \frac{\widehat{\sigma}_1 - \alpha_{1,2}^{\text{avg}}}{1 - \widehat{\sigma}_1} + \delta_\alpha^{1,2}\right], \tag{3.22}$$

$$\left[\frac{\beta_{1,2}^{\text{avg}} + \widehat{\sigma}_1}{1 + \widehat{\sigma}_1} - \delta_\beta^{1,2}\ ,\ \frac{\beta_{1,2}^{\text{avg}} - \widehat{\sigma}_1}{1 - \widehat{\sigma}_1} + \delta_\beta^{1,2}\right], \tag{3.23}$$

respectively. Thus, assuming the segments eqs. (3.22) and (3.23) are small, a significant improvement in the condition number as well as the clustering of the spectrum of the original preconditioned system eq. (3.14) is expected. An illustration is given in fig. 4.

Proof. of lemma 2.
To show the first part of the lemma we consider a general matrix of the form

$$C = \begin{bmatrix} \delta_1 \widehat{A}_1 & M \\ M^T & \delta_2 \widehat{A}_2 \end{bmatrix}, \quad \delta_1, \delta_2 > 0\,.$$

Now applying the following change of basis

$$x_1 = \frac{1}{\sqrt{\delta_1}} \widehat{B}_1 \xi_1\,, \quad x_2 = \frac{1}{\sqrt{\delta_2}} \widehat{B}_2 \xi_2\,,$$

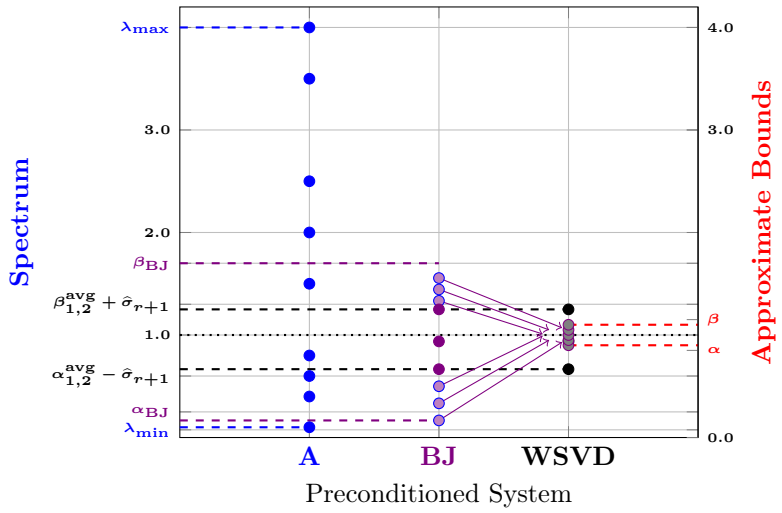


Figure 4: **Spectrum Clustering for $\widehat{A}_i \neq A_i$.** The spectrum of some SPD matrix A and the transformation it goes after preconditioning by block Jacobi (BJ) and the WSVD preconditioner is displayed. The spectra are ordered from the left to the right starting from A, followed by BJ and ends up with WSVD. The spectral bounds α eq. (3.19) and β eq. (3.20) are marked in red on the right y-axis, while the spectral bounds for the block Jacobi case $\alpha_{\mathrm{BJ}} = \alpha(r = 0)$ and $\beta_{\mathrm{BJ}} = \beta(r = 0)$ are marked in purple on the left y-axis. The main difference compared to fig. 2, is that the WSVD preconditioner now maps the extreme eigenvalues to an interval and not exactly to 1.

where $x^T = (x_1^T, x_2^T)$ corresponds to the partition of C, yields

$$\mathcal{C} = \begin{bmatrix} \frac{1}{\sqrt{\delta_1}}\widehat{B}_1^T & 0 \\ 0 & \frac{1}{\sqrt{\delta_2}}\widehat{B}_2^T \end{bmatrix} C \begin{bmatrix} \frac{1}{\sqrt{\delta_1}}\widehat{B}_1 & 0 \\ 0 & \frac{1}{\sqrt{\delta_2}}\widehat{B}_2 \end{bmatrix} = \begin{bmatrix} I & \frac{1}{\sqrt{\delta_1\delta_2}}\mathcal{M} \\ \frac{1}{\sqrt{\delta_1\delta_2}}\mathcal{M}^T & I \end{bmatrix},$$

where $\mathcal{M} = \widehat{B}_1^T M \widehat{B}_2$. The matrix C is SPD iff $\mathcal{C}$ is SPD as well. Thus, by lemma 1 the matrix C is SPD iff $1 - \sigma_1/\sqrt{\delta_1\delta_2} > 0$, and the conditions ensuring $\underline{A}$, $\widehat{A}$, and $\overline{A}$ are SPD immediately follow.

For the second part of the lemma it is sufficient to assume that $\underline{A}$ SPD, which, by the first part, ensures that $\widehat{A}$ and $\overline{A}$ are SPD as well. Accordingly, we obtain the following inequalities

$$\frac{x^T\widehat{A}x}{x^T\underline{A}x} \leq \frac{x^T\widehat{A}x}{x^TAx} \leq \frac{x^T\widehat{A}x}{x^T\overline{A}x} \quad \forall\, x \neq 0\,.$$

The Lagrangian stationary points of each generalized Rayleigh quotient in the inequalities above constitute the spectrum of each preconditioned system. See appendix A for further details. Thus, the proof is complete. □

The following lemma is a key ingredient in the proof of theorem 2.

Lemma 3. *Let* $\mathcal{C} = \left[\begin{smallmatrix} \mathcal{D}^{(1)} & \mathcal{D}^{(2)} \\ \mathcal{D}^{(2)} & \mathcal{D}^{(3)} \end{smallmatrix}\right] \in \mathbb{R}^{2m\times 2m}$, *where* $\mathcal{D}^{(i)}$ $(i = 1, 2, 3)$ *are diagonal matrices,*

$$\mathcal{D}^{(i)} = \mathrm{diag}(d_1^{(i)}, \ldots, d_m^{(i)})\,.$$

If $d_j^{(2)} \neq 0$ *for all* $j = 1, 2, \ldots, m$, *then*

$$\mathrm{spec}(\mathcal{C}) = \left\{\lambda_j^-\right\}_{j=1}^m \cup \left\{\lambda_j^+\right\}_{j=1}^m , \quad \lambda_j^{\pm} = \frac{d_j^{(1)} + d_j^{(2)}}{2} \pm \sqrt{\left(\frac{d_j^{(1)} - d_j^{(2)}}{2}\right)^2 + \left(d_j^{(2)}\right)^2},$$

where $\mathrm{spec}(\mathcal{C})$ *denotes the spectrum of the symmetric matrix* $\mathcal{C}$.

Proof. of lemma 3.
From the given structure of $\mathcal{C}$ it is clear that $\lambda \in \mathbb{R}$ is an eigenvalue of $\mathcal{C}$ iff for some $j = 1, 2, \ldots, m$ the vectors $(d_j^{(1)} - \lambda, d_j^{(2)})$ and $(d_j^{(2)}, d_j^{(3)} - \lambda)$ are linearly dependent. Since we have assumed $d_j^{(2)} \neq 0$, we have that $(d_j^{(1)} - \lambda, d_j^{(2)})$ and $(d_j^{(2)}, d_j^{(3)} - \lambda)$ are linearly dependent iff

$$\frac{d_j^{(1)} - \lambda}{d_j^{(2)}} = \frac{d_j^{(2)}}{d_j^{(3)} - \lambda} \quad \Leftrightarrow \quad (d_j^{(1)} - \lambda)(d_j^{(3)} - \lambda) - (d_j^{(2)})^2 = 0\,.$$

The solution to the quadratic equation above is

$$\lambda = \lambda_j^{\pm} = \frac{d_j^{(1)} + d_j^{(2)}}{2} \pm \sqrt{\left(\frac{d_j^{(1)} - d_j^{(2)}}{2}\right)^2 + \left(d_j^{(2)}\right)^2},$$

and the proof is complete. □

Proof. of theorem 2.
The conditions given in theorem 2, ensure by lemma 2 the existence of the spectral bounds, α and β, satisfying

$$0 < \alpha = \min_{x \neq 0} \frac{x^T \underline{A} x}{x^T \widehat{A} x} \leq \lambda_{\min}(\widehat{B}^T A \widehat{B}) \leq \lambda_{\max}(\widehat{B}^T A \widehat{B}) \leq \max_{x \neq 0} \frac{x^T \overline{A} x}{x^T \widehat{A} x} = \beta\,.$$

where $\underline{A}$ and $\overline{A}$ are defined in lemma 2.

To find the exact values of α and β, we consider a generalized Rayleigh quotient

$$R(x) = \frac{x^T \widetilde{A} x}{x^T \widehat{A} x}, \quad \widetilde{A} = \begin{bmatrix} \delta_1 \widehat{A}_1 & M \\ M & \delta_2 \widehat{A}_2 \end{bmatrix},$$

where $\delta_1, \delta_2 > 0$ such that $\widetilde{A}$ is SPD. Applying the change of variables, $x = \left[\begin{smallmatrix} \widehat{B}_1 & 0 \\ 0 & \widehat{B}_2 \end{smallmatrix}\right] \xi$, where $\widehat{B}_i^T \widehat{A}_i \widehat{B}_i = I_i$ and I_i is the $n_i \times n_i$ identity matrix, yields

$$R(x) = \frac{\xi^T \widetilde{\mathcal{A}} \xi}{\xi^T \widehat{\mathcal{A}} \xi}, \quad \widetilde{\mathcal{A}} = \begin{bmatrix} \delta_1 I_1 & \mathcal{M} \\ \mathcal{M}^T & \delta_2 I_2 \end{bmatrix}, \quad \widehat{\mathcal{A}} = \begin{bmatrix} I_1 & \mathcal{M}_r \\ \mathcal{M}_r^T & I_2 \end{bmatrix},$$

where $\mathcal{M} = \widehat{B}_1^T M \widehat{B}_2$ and $\mathcal{M}_r$ is the r-rank weighted SVD approximation of $\mathcal{M}$.

Let w_i denote the i-th column of the orthogonal matrix $\mathcal{W}$ eq. (3.8) as defined in lemma 1. Then we have:

1. $\widehat{\mathcal{A}} w_i = (1 + \widehat{\sigma}_1) w_i$, $i = 1, 2, \ldots, r$.
2. $\widehat{\mathcal{A}} w_{n_1+i} = (1 - \widehat{\sigma}_1) w_{n_1+i}$, $i = 1, 2, \ldots, r$.

3. $\widehat{\mathcal{A}}w_j = w_j$, $j \neq 1, \ldots, r, n_1+1, \ldots, n_1+r$.

and similarly for $\widetilde{\mathcal{A}}$:

1. $\widetilde{\mathcal{A}}w_i = (\delta_{1,2}^{\text{avg}} + \widehat{\sigma}_1)w_i + \delta_{1,2}^{\text{dif}} w_{n_1+i}$, $i = 1, 2, \ldots, m$.

2. $\widetilde{\mathcal{A}}w_{n_1+i} = (\delta_{1,2}^{\text{avg}} - \widehat{\sigma}_1)w_{n_1+i} + \delta_{1,2}^{\text{dif}} w_i$, $i = 1, 2, \ldots, m$.

3. $\widetilde{\mathcal{A}}w_j = w_j$, $j \neq 1, \ldots, m, n_1+1, \ldots, n_1+m$.

where $\delta_{1,2}^{\text{avg}} = (\delta_1 + \delta_2)/2$ and $\delta_{1,2}^{\text{dif}} = (\delta_1 - \delta_2)/2$. Clearly, both $\widehat{\mathcal{A}}$ and $\widetilde{\mathcal{A}}$ are invariant over the subspaces $Z = \text{span}\{w_1, \ldots, w_r, w_{n_1+1}, \ldots, w_{n_1+r}\}$ and its orthogonal complement, $Z^\perp$. Hence, by the properties of the generalized Rayleigh quotient we have:

$$\max_{x \neq 0} R(x) = \max \left\{ \max_{\xi \in Z \setminus \{0\}} R(x) \, , \max_{\xi \in Z^\perp \setminus \{0\}} R(x) \right\} ,$$

and

$$\min_{x \neq 0} R(x) = \min \left\{ \min_{\xi \in Z \setminus \{0\}} R(x) \, , \min_{\xi \in Z^\perp \setminus \{0\}} R(x) \right\} .$$

By our results so far, if $x = \xi \in Z^\perp$ then $R(x) = \xi^T \widetilde{\mathcal{A}} \xi / \xi^T \xi$. Let us apply the change of variables of the form $\xi = C\zeta \in Z^\perp$, given explicitly by

$$\xi = \zeta_1 w_{r+1} + \ldots + \zeta_{n_1-r} w_{n_1} + \zeta_{n_1-r+1} w_{n_1+r+1} + \ldots + \zeta_{n_1+n_2-2r} w_{n_1+n_2} \, ,$$

where ζ_i is the i-th coordinate of ζ and as before w_i denotes the i-th column in the orthogonal matrix $\mathcal{W}$ eq. (3.8). Then, for any $\xi \in Z^\perp$ we obtain

$$R(x) = \frac{\zeta^T \mathcal{C}_{Z^\perp} \zeta}{\zeta^T \zeta} \, , \quad \mathcal{C}_{Z^\perp} = \begin{bmatrix} \mathcal{D}_{Z^\perp}^{(1)} & \mathcal{D}_{Z^\perp}^{(2)} \\ (\mathcal{D}_{Z^\perp}^{(2)})^T & \mathcal{D}_{Z^\perp}^{(3)} \end{bmatrix} ,$$

where $\mathcal{D}_{Z^\perp}^{(2)} = \delta_{1,2}^{\text{dif}} I_{n_1,n_2}$ and

$$\mathcal{D}_{Z^\perp}^{(1)} = \begin{cases} \text{diag}\left(\delta_{1,2}^{\text{avg}} + \widehat{\sigma}_{r+1}, \ldots, \delta_{1,2}^{\text{avg}} + \widehat{\sigma}_{n_1}\right) & \text{if} \quad n_1 \leq n_2 \quad , \\ \text{diag}\left(\delta_{1,2}^{\text{avg}} + \widehat{\sigma}_{r+1}, \ldots, \delta_{1,2}^{\text{avg}} + \widehat{\sigma}_{n_1}, \delta_1, \ldots, \delta_1\right) & \text{if} \quad n_1 > n_2 \quad . \end{cases}$$

$$\mathcal{D}_{Z^\perp}^{(3)} = \begin{cases} \text{diag}\left(\delta_{1,2}^{\text{avg}} - \widehat{\sigma}_{r+1}, \ldots, \delta_{1,2}^{\text{avg}} - \widehat{\sigma}_{n_2}\right) & \text{if} \quad n_2 \leq n_1 \quad , \\ \text{diag}\left(\delta_{1,2}^{\text{avg}} - \widehat{\sigma}_{r+1}, \ldots, \delta_{1,2}^{\text{avg}} - \widehat{\sigma}_{n_2}, \delta_2, \ldots, \delta_2\right) & \text{if} \quad n_2 > n_1 \quad . \end{cases}$$

Now,by lemma 3, we obtain that the spectrum of $\mathcal{C}_{Z^\perp}$ contains the sets

$$\left\{ \delta_{1,2}^{\text{avg}} + \sqrt{\widehat{\sigma}_{r+1}^2 + \left(\delta_{1,2}^{\text{dif}}\right)^2}, \ldots, \delta_{1,2}^{\text{avg}} + \sqrt{\widehat{\sigma}_m^2 + \left(\delta_{1,2}^{\text{dif}}\right)^2}, \right\} ,$$

$$\left\{ \delta_{1,2}^{\text{avg}} - \sqrt{\widehat{\sigma}_m^2 + \left(\delta_{1,2}^{\text{dif}}\right)^2}, \ldots, \delta_{1,2}^{\text{avg}} - \sqrt{\widehat{\sigma}_{r+1}^2 + \left(\delta_{1,2}^{\text{dif}}\right)^2} \right\} .$$

Hence, we conclude that

$$\min_{\xi \in Z^\perp \setminus \{0\}} R(x) = \delta_{1,2}^{\text{avg}} - \sqrt{\widehat{\sigma}_{r+1}^2 + \left(\delta_{1,2}^{\text{dif}}\right)^2} ,$$

$$\max_{\xi\in Z^{\perp}\setminus\{0\}} R(x) = \delta_{1,2}^{\mathrm{avg}} + \sqrt{\widehat{\sigma}_{r+1}^2 + \left(\delta_{1,2}^{\mathrm{dif}}\right)^2}\,.$$

For the case $\xi \in Z$ let us apply the change of variables of the form $\xi = C\psi \in Z$, given explicitly by

$$\xi = \psi_1 w_1 + \ldots + \psi_r w_r + \psi_{r+1} w_{n_1+1} + \ldots + \psi_{2r} w_{n_1+r}\,,$$

where ψ_i is the i-th coordinate of ψ and as before w_i denotes the i-th column in the orthogonal matrix $\mathcal{W}$ eq. (3.8). Then, for any $\xi \in Z$ we obtain

$$R(x) = \frac{\psi^T \widetilde{\mathcal{C}}_Z \psi}{\psi^T \psi}\,, \quad \mathcal{C}_Z = \begin{bmatrix} \mathcal{D}_Z^{(1)} & \mathcal{D}_Z^{(2)} \\ \mathcal{D}_Z^{(2)} & \mathcal{D}_Z^{(3)} \end{bmatrix}\,,$$

$$\mathcal{D}_Z^{(1)} = \mathrm{diag}\left(\frac{\delta_{1,2}^{\mathrm{avg}} + \widehat{\sigma}_1}{1+\widehat{\sigma}_1}, \ldots, \frac{\delta_{1,2}^{\mathrm{avg}} + \widehat{\sigma}_r}{1+\widehat{\sigma}_r}\right)\,,$$

$$\mathcal{D}_Z^{(2)} = \mathrm{diag}\left(\frac{\delta_{1,2}^{\mathrm{dif}}}{\sqrt{1-\widehat{\sigma}_1^2}}, \ldots, \frac{\delta_{1,2}^{\mathrm{dif}}}{\sqrt{1-\widehat{\sigma}_r^2}}\right)\,,$$

$$\mathcal{D}_Z^{(3)} = \mathrm{diag}\left(\frac{\delta_{1,2}^{\mathrm{avg}} - \widehat{\sigma}_1}{1-\widehat{\sigma}_1}, \ldots, \frac{\delta_{1,2}^{\mathrm{avg}} - \widehat{\sigma}_r}{1-\widehat{\sigma}_r}\right)\,.$$

Applying once more the outcome of lemma 3 we have that the spectrum of $\mathcal{C}_Z$ is composed of the following values

$$\frac{1}{2}\left(\frac{\delta_{1,2}^{\mathrm{avg}} + \widehat{\sigma}_i}{1+\widehat{\sigma}_i} + \frac{\delta_{1,2}^{\mathrm{avg}} - \widehat{\sigma}_i}{1-\widehat{\sigma}_i}\right) \pm \sqrt{\frac{1}{4}\left(\frac{\delta_{1,2}^{\mathrm{avg}} + \widehat{\sigma}_i}{1+\widehat{\sigma}_i} + \frac{\delta_{1,2}^{\mathrm{avg}} - \widehat{\sigma}_i}{1-\widehat{\sigma}_i}\right)^2 + \frac{\left(\delta_{1,2}^{\mathrm{dif}}\right)^2}{1-\widehat{\sigma}_i^2}}\,,$$

where $i = 1, 2, \ldots, r$ and the proof is complete. □

4 Practical Multilevel Implementation

In this section we present the implementation of the theory we have developed in Section 3 for constructing a WSVD version of the multilevel HODLR preconditioner as described in Section 2. We begin with the description of the procedure, which we interpret as a greedy approach attempting to construct a global minimizer of the global condition number. We end with a discussion on the behavior of the spectral bounds and the condition number as a function of the level, and, essentially, justify assumption eq. (3.15).

For an efficient construction of the preconditioner, $\widehat{A}_1^{(0)}$, we perform a single-pass over the hierarchy from bottom (level $\ell = L$) to top (level $\ell = 0$). At level $\ell = L$ we set $\widehat{A}_k^{(L)} = A_k^{(L)}$ for all $k = 1, 2, \ldots, 2^L$. At each level $\ell < L$ we consider the matrices

$$\begin{bmatrix} A_1 & M \\ M^T & A_2 \end{bmatrix} = \begin{bmatrix} A_{2k-1}^{(\ell+1)} & M_k^{(\ell)} \\ \left(M_k^{(\ell)}\right)^T & A_{2k}^{(\ell+1)} \end{bmatrix}\,, \quad \begin{bmatrix} \widehat{A}_1 & \widehat{M} \\ \widehat{M}^T & \widehat{A}_2 \end{bmatrix} = \begin{bmatrix} \widehat{A}_{2k-1}^{(\ell+1)} & \widehat{M}_k^{(\ell)} \\ \left(\widehat{M}_k^{(\ell)}\right)^T & \widehat{A}_{2k}^{(\ell+1)} \end{bmatrix}\,,$$

and compute the preconditioner by approximating each off-diagonal block according to definition 1,

$$\widehat{B}_1^T \widehat{M} \widehat{B}_2 = \widehat{\mathcal{U}}_r \widehat{\Sigma}_r \widehat{\mathcal{V}}_r^T\,, \quad r = r_k^{(\ell)}\,,$$

where $\widehat{\mathcal{U}}_r$ and $\widehat{\mathcal{V}}_r$ are composed of the first r left and right, respectively, singular vectors of the SVD,

$$\widehat{B}_1^T M \widehat{B}_2 = \widehat{\mathcal{U}} \widehat{\Sigma} \widehat{\mathcal{V}}^T\,, \quad \widehat{\Sigma} = \mathrm{diag}(\widehat{\sigma}_1, \ldots, \widehat{\sigma}_m)\,, \quad m = \min\{n_1, n_2\}\,.$$

The process ends at level $\ell = 0$, which yields the global preconditioner $\widehat{A}_1^{(0)}$. Clearly, this procedure assumes that we can accurately and efficiently apply the inverse square roots at each level, $\widehat{B} = \widehat{B}_k^{(\ell)}$. This can be achieved, in principle, by employing the hierarchical Cholesky factorization [5].

It is important to note that the suggested procedure attempts to greedily obtain an optimal global preconditioner that minimizes the global spectral condition number,

$$\mathrm{cond}_2(\widehat{B}^T A_1^{(0)} \widehat{B}) \,, \quad \widehat{B} = \widehat{B}_1^{(0)} \,,$$

given a distribution of ranks $\left\{ r_k^{(\ell)} \right\}_{\ell,k=0,1}^{L-1,2^\ell}$. Indeed, if $\widehat{A}_i$ $(i = 1, 2)$ are optimal minimizers of

$$\mathrm{cond}_2(\widehat{B}_i^T A_i \widehat{B}_i) \,, \quad A_i = A_{2k-1+i}^{(\ell+1)} \,,$$

then the matrix

$$\begin{bmatrix} \widehat{A}_1 & M \\ M^T & \widehat{A}_2 \end{bmatrix} = \begin{bmatrix} \widehat{A}_{2k-1}^{(\ell+1)} & M_k^{(\ell)} \\ \left(M_k^{(\ell)}\right)^T & \widehat{A}_{2k}^{(\ell+1)} \end{bmatrix} \,, \tag{4.1}$$

is an optimal choice for approximating $A = A_k^{(\ell)}$. By theorem 1 an optimal choice for approximating eq. (4.1) by replacing its off-diagonal blocks with low-rank blocks is given by $\widehat{M}$ satisfying

$$\widehat{B}_1^T \widehat{M} \widehat{B}_2 = \widehat{\mathcal{U}}_r \widehat{\Sigma}_r \widehat{\mathcal{V}}_r^T \,, \quad r = r_k^{(\ell)} \,,$$

with respect to the SVD of $\widehat{B}_1^T M \widehat{B}_2$ as described in definition 1.

To determine the rank of each off-diagonal low-rank block, $\widehat{M} = \widehat{M}_k^{(\ell)}$, we apply

$$r_k^{(\ell)} = \underset{0 \le r \le m}{\mathrm{argmin}} \left\{ \widehat{\sigma}_{r+1} \le \tau_k^{(\ell)} \cdot \widehat{\sigma}_1 \right\} \,, \quad k = 1, 2, \ldots, 2^\ell \,, \tag{4.2}$$

where the distribution of positive tolerances, $\{\tau_k^{(\ell)}\}$, is predetermined. Note, that at the bottom of the hierarchy at level $\ell = L - 1$ we have $\widehat{A}_i = A_i$. Thus, by theorem 1 the spectral bounds,

$$\alpha = 1 - \sigma_{r+1} \,, \quad \beta = 1 + \sigma_{r+1} \,, \quad r = r_k^{(L-1)} \,,$$

are sharp $\left(\alpha = \lambda_{\min}\left(\widehat{B}^T A B\right) \le \lambda_{\max}\left(\widehat{B}^T A B\right) = \beta \right)$, and satisfy assumption eq. (3.15) with respect to the level above $(\ell = L - 2)$,

$$0 < \alpha \le 1 \le \beta \,.$$

In particular, the condition number at level $L - 1$ is the maximal ratio $(1 + \sigma_{r+1})/(1 - \sigma_{r+1})$ of all the preconditioned submatrices, $\widehat{B}^T A_k^{(L-1)} \widehat{B}$, where $\widehat{B} = \widehat{B}_k^{(L-1)}$.

From the proofs of theorem 2 it is clear that α is monotonically non-increasing as a function of the level, while β is monotonically non-decreasing as a function of the level. Hence, the spectral bounds at each level $\alpha = \alpha_k^{(\ell)}, \beta = \beta_k^{(\ell)}$ are expected to satisfy assumption eq. (3.15). This observation is also supported by numerical evidence in Section 5. Note, that the monotonic behaviour of the spectral bounds implies similar monotonic behaviour of the condition number per level. Thus, the condition number at level $\ell = L-1$ serves as a lower bound for any condition number at higher levels.

5 Numerical Results

This section describes the experimental part of this work. We begin with the description of the model problem and its properties in Section 5.1. In Section 5.2 we describe two additional low-rank approximation methods for preconditioning. Section 5.3 contains the numerical results for the given model problem, including a detailed comparison with the methods described in the previous subsection.

5.1 The Model Problem

Consider the 2D Poisson problem

$$u_{xx} + u_{yy} = f(x,y)\,, \quad (x,y) \in (0,1) \times (0,1)\,,$$

with the following Robin boundary condition,

$$\epsilon u + \partial_\nu u|_{\partial\Omega} = 0\,, \tag{5.1}$$

where $\epsilon > 0$ and $\partial_\nu u$ denotes the normal derivative. When $\epsilon \to 0^+$ the problem becomes increasingly unstable since the limit case is the ill posed Poisson problem with Neumann boundary conditions.

We discretize the problem by setting a uniform grid of $(N+1)^2$ equally spaced gridpoints,

$$x_i = \frac{i}{(2N+2)}\,, \quad y_j = \frac{j}{(2N+2)}\,, \quad i,j = 1,2,\ldots,N+1\,,$$

and apply the five-point *finite difference* discretization rule. Thus, we end up with a symmetric linear system of $(N+1)^2$ equations in $(N+1)^2$ unknowns, $u_{i,j} \approx u(x_i, y_j)$, where $i,j = 1,2,\ldots,N+1$. It can be shown, that the spectrum of the discrete operator, i.e., the matrix, is bounded. Hence, the matrix is SPD. To avoid numerical instability we also scale the system and multiply it by $(N+1)^2$.

To reduce the dimensionality and conditioning of the given problem, we separate the effect of the unstable boundary from the interior part in the following manner. First we define the gridpoints that interact with the boundary,

$$(x_i, y_j) \quad : \quad i = 1 \text{ or } j = N+1 \text{ or } i = N+1 \text{ or } j = 1\,, \tag{5.2}$$

as the 'border' set. Now, computing the Schur complement associated with the border set eq. (5.2) we obtain a new dense matrix whose dimensionality and condition number are significantly reduced. Indeed, from a sparse system of $(N+1)^2$ unknowns the Schur complement is only a $4N \times 4N$ matrix. Using the *Cauchy interlacing theorem* eq. (3.7) it can be shown [26, p. 47] that the eigenvalues of the Schur complement are interlacing with those of the original finite difference matrix. Thus, the condition number of the Schur is ensured to be lower than the condition number of the original finite difference system. Also, note that it is well known that the solution of the original system can be easily obtained from the solution of the Schur system.

Obtaining the Schur system involves the factorization of the submatrix corresponding to the set of non-border gridpoints. In our model problem this submatrix is the 5-point Laplacian with Dirichlet boundary conditions, whose condition number is better than the original matrix. Using a sparse direct solver software, e.g., PaStiX [13], which efficiently stores the factorization. Thus, we can implicitly apply the Schur complement without the need to explicitly form it.

For constructing the WSVD HODLR preconditioner we apply a balanced geometric partitioning on the border set eq. (5.2). We start by bisecting the set at the lower-left and upper-right

corners, next we bisect each subset at the lower-right and upper-left corners, respectively. The following bisections are applied to each subset of the partition, which is separated at the middle into two equally sized subsets. We continue the process until reaching the predetermined bottom level, L. An illustration of the process is displayed in fig. 5.

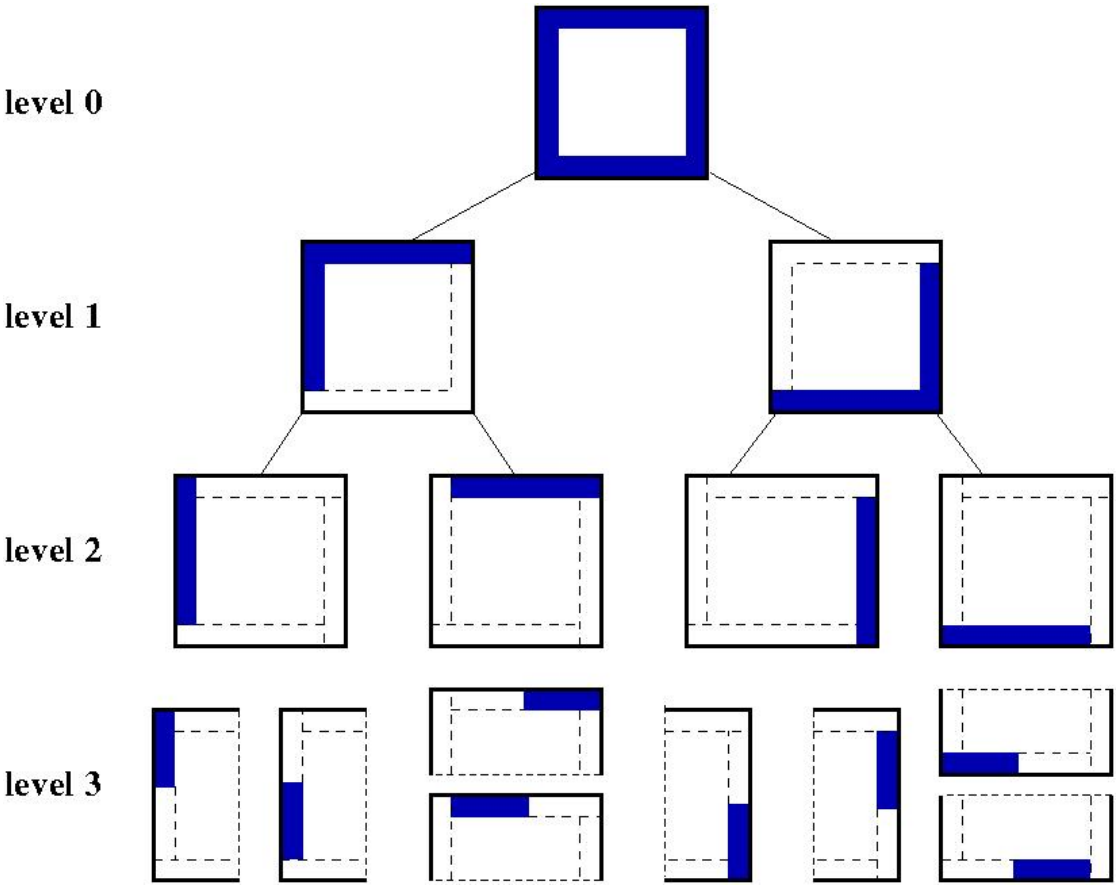


Figure 5: **Border Set Hierarchical Partitioning.**

At this point we can explore numerically the properties of the partitioned and accordingly reordered Schur complement, denoted by $A_1^{(0)}$. fig. 6 displays a plot of the average condition number at each level,

$$\mathrm{cond}_2^{\mathrm{avg}}\left(A^{(\ell)}\right) = \frac{1}{2^\ell}\sum_{k=1}^{2^\ell}\mathrm{cond}_2\left(A_k^{(\ell)}\right),$$

as a function of the level, $\ell = 0, 1, \ldots, 6$. From the plot it is clear, that as we go up the hierarchy from the bottom ($\ell = L$) to the top ($\ell = 0$), the condition numbers of the principle blocks increase. This fact can be attributed to the nature of the partition. The largest grid captures the eigenmode corresponding to the smallest eigenvalue, which is not visible on the lower level grids in the partition. Similarly as we descend the hierarchy, additional small eigenvalues are effectively removed, leading to the decrease of the average condition number.

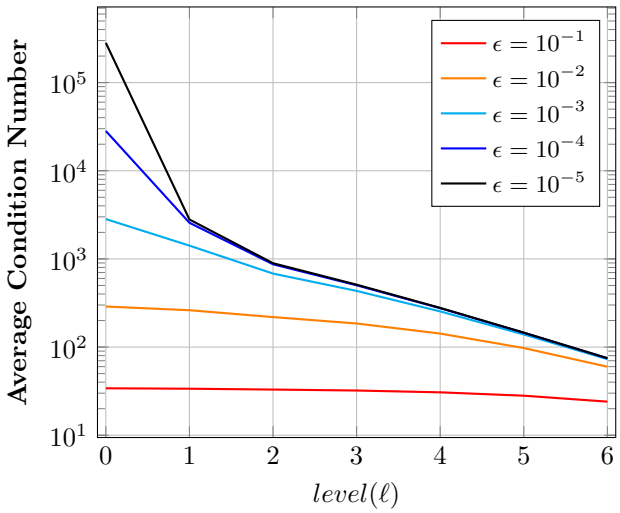


Figure 6: **Conditioning vs. Level of $\boldsymbol{A_1^{(0)}}$.**

5.2 Low-rank Approximation Schemes

For the comparative study of the weighted SPD HODLR preconditioner we consider two other techniques for obtaining low-rank approximations of the off-diagonal blocks.

The first option is to employ *regular singular value decomposition* (RSVD), which is, in fact, the common approach. Each off-diagonal block $M_k^{(\ell)}$ is approximated by its r-rank SVD approximation

$$\widehat{M}_k^{(\ell)} = U_{r,k}^{(\ell)} S_k^{(\ell)} \left(V_{r,k}^{(\ell)}\right)^T, \quad S_{r,k}^{(\ell)} = \mathrm{diag}(s_{1,k}^{(\ell)}, \ldots, s_{r,k}^{(\ell)}), \tag{5.3}$$

where the matrices $U_{r,k}^{(\ell)}$ and $\widehat{\mathcal{V}}_{r,k}^{(\ell)}$ are composed of the first r left singular vectors and the first r right singular vectors, respectively, of the SVD of $M_k^{(\ell)}$.

The other method is the *filtered singular value decomposition* (FSVD), that has been suggested in [6]. The method is considered state-of-the art for preconditioning by hierarchical matrices. The idea is to choose an approximation of the form

$$\widehat{M}_k^{(\ell)} = U_{r,k}^{(\ell)} S_k^{(\ell)} \left(V_{r,k}^{(\ell)}\right)^T + \left[P_{r,2k-1}^{(\ell)}\right]^T E_{r,k}^{(\ell)} \left[P_{r,2k}^{(\ell)}\right], \tag{5.4}$$

where $P_{r,j}^{(\ell)}$ $(j = 2k-1, 2k)$ is a projection matrix onto a low-dimensional subspace, and $E_{r,k}^{(\ell)}$ is the error of the regular r-rank SVD approximation eq. (5.3),

$$E_{r,k}^{(\ell)} = M_k^{(\ell)} - U_{r,k}^{(\ell)} S_k^{(\ell)} \left(V_{r,k}^{(\ell)}\right)^T, \quad S_{r,k}^{(\ell)} = \mathrm{diag}(s_{1,k}^{(\ell)}, \ldots, s_{r,k}^{(\ell)}).$$

Choosing $P_{r,k}^{(\ell)}$ to be the projection onto the 1-dimensional subspace of constant vectors has exhibited superior results using weak hierarchical $\mathcal{H}$-matrix approximation compared to the RSVD approach. The intuition behind this choice, is that the global approximation, $\widehat{A}$, becomes nearly exact on a subspace of piecewise-constant functions. Thus, it can better capture the slow varying eigenmodes of elliptic problems than the standard RSVD approach.

5.3 Experimental Results

In this subsection we present the experimental study of the WSVD HODLR preconditioner for the iterative solution of the Schur complement system, A.

For the computational setting we have employed:

- PasTiX 5.22 [13] to evaluate the Schur complement matrix.
- Scilab 5.5.2 [23] to construct the SPD HODLR preconditioners, $\widehat{A}_1^{(0)}$, and perform spectrum evaluations and PCG simulations.

The inverse square roots were evaluated by inverse symmetric square roots,

$$\widehat{B}_k^{(\ell)} = \left(\widehat{A}_k^{(\ell)}\right)^{-1/2} .$$

In principle, it would be more efficient to use the hierarchical Cholesky factorization [5] to evaluate $\widehat{B}_k^{(\ell)}$. However, to avoid uncertainty which may be induced by an approximate factorization we preferred an exact and stable inversion. An implementation of an inverse square root factorization adapted to the HODLR structure is left for future work.

For the setting of the model problem we have chosen:

- Grid parameter $N = 10^3$, which gives a $4 \cdot 10^3 \times 4 \cdot 10^3$ Schur complement matrix, A. Note that the dimensions of the original sparse finite difference matrix are roughly $10^6 \times 10^6$.
- $L = 6$ as the lowest level of the hierarchy.
- $\epsilon = 10^{-3}$ in the Robin boundary condition eq. (5.1).

In all the simulations we employed the uniform approach for the truncation rule

$$\tau_k^{(\ell)} = \tau \quad \forall\, \ell = 0, 1, \ldots, L-1\,,\ k = 1, \ldots, 2^\ell\,,$$

with the uniform τ taking the following tolerance values:

$$\tau = 0.4, 0.2, 0.1, 0.05, 0.025, 0.005, 0.002, 0.001, 0.0001\,. \tag{5.5}$$

In the following paragraphs we describe the comparative study of the three different methods we have employed to construct the SPD HODLR preconditioner, namely RSVD, FSVD and WSVD. Since each method approximated the low rank off-diagonal blocks differently, each chosen τ eq. (5.5) produces a different distribution of ranks, $\{r_k^{(\ell)}\}$. To compare the methods properly we also examine their performance as a function of the total compression ratio, i.e., the ratio of the memory used in practice and the total available memory. The compression ratio of each method as a function of τ is given in table 1.

The FSVD preconditioner had been implemented by filtering the errors eq. (5.4) of the RSVD preconditioner with the same τ. Thus, we obtain for the same τ worse compression ratio than in the RSVD case, since the rank of each low-rank off-diagonal block larger by 2 in the FSVD case compared to the RSVD case. However, a better condition number and possibly a better clustering of the spectrum are expected when employing FSVD compared to RSVD with the same τ. Note, that the compression ratio of WSVD is, essentially, unrelated to RSVD and FSVD with the same τ, since the truncation is performed in a different (weighted) norm.

fig. 7 displays three plots showing the effect of each SPD HODLR method, namely RSVD, FSVD and WSVD, on the spectrum of the matrix A for the range of uniform tolerances given in

τ	RSVD	FSVD	WSVD
0.0001	0.11964606	0.13146138	0.12899986
0.001	0.10290766	0.11472300	0.11558552
0.002	0.09896922	0.11078454	0.10979994
0.005	0.09109232	0.10290766	0.10586150
0.01	0.08813848	0.09995382	0.10069228
0.025	0.08124620	0.09306154	0.09503076
0.05	0.07829236	0.09010770	0.09103128
0.1	0.07140008	0.08321542	0.08518464
0.2	0.07140008	0.08321542	0.08124620
0.4	0.06844626	0.08026160	0.07140008

Table 1: **Compression Ratio vs. τ.** Each column contains the compression ratios of each method, namely RSVD, FSVD and WSVD. Each row contains the compression ratio of each method for the same uniform tolerance value, τ.

table 1. The spectra of the preconditioned system by each method are ordered from the left to the right starting from the spectrum of A (in green) and then according to τ descending from the largest to the smallest value. As expected from table 1, the FSVD preconditioner compression ratio is higher than the RSVD preconditioner with the same τ. The main conclusions that are clear from the graphs when comparing spectrum clustering of the methods at the same compression ratio are:

- FSVD removes more small eigenvalues than RSVD with the same compression ratio.
- FSVD removes less large eigenvalues than RSVD with the same compression ratio, and in general does not cluster the spectrum better than RSVD (with the same compression ratio).
- WSVD removes large and small eigenvalues better than RSVD and FSVD with the same compression ratio, and generally clusters the spectrum substantially better than the other methods.

To complete the result displayed in fig. 7, we also plot the condition number achieved as a function of the compression ratio for each method in fig. 8. As before, the compression ratio is the ratio of the memory used in practice and the total available memory. The graph clearly shows that the WSVD achieves a better condition number with less memory resources than any other method.

Another important aspect is the number of iterations to convergence that is achieved in practice with the iterative scheme. A natural choice of iterative scheme for SPD matrices is the *preconditioned conjugate gradient* (PCG) method. We employ the PCG for the solution of eq. (1.1) using each preconditioner for every τ in table 1 with the a right-hand-side, $b = 1$. The stopping criterion is set according to the non-preconditioned system with a relative threshold of 10^{-8}, by stopping the iterations at the first occurrence of

$$\left\| Ax_{(i)} - b \right\|_2 \leq 10^{-8} \left\| b \right\|_2 ,$$

where $i = 1, 2, \ldots$ is the iteration step index and $x_{(i)}$ is the approximate solution of the non-preconditioned system at step i.

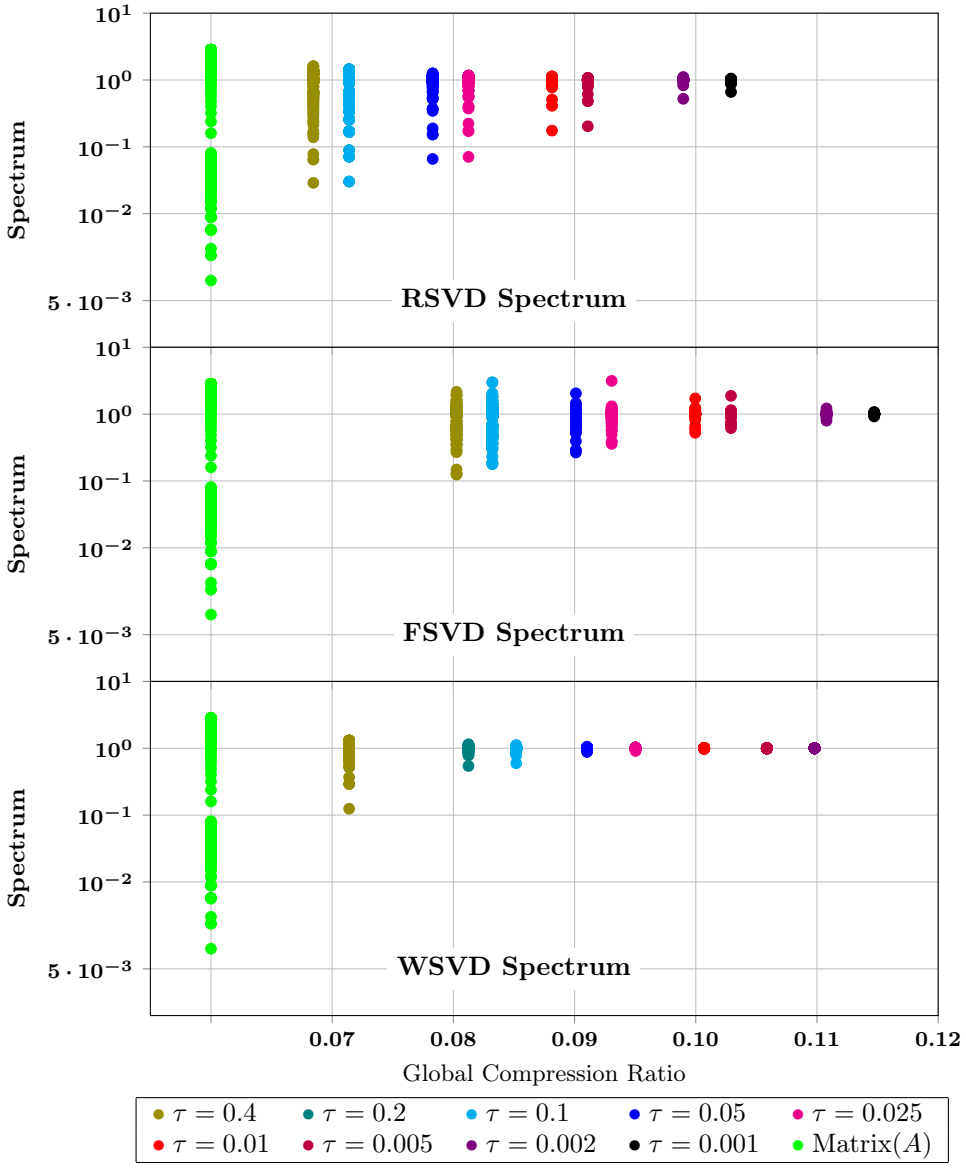


Figure 7: **Spectrum Clustering of the Preconditioned system.**

fig. 9 displays the number of iterations to convergence as a function of the compression ratio.The main conclusions from the graphs are the following:

- PCG with FSVD preconditioner converges faster than RSVD preconditioner for the same τ. However, for the same compression ratio the convergence rate in the RSVD case is slightly faster than in the FSVD case.
- PCG with WSVD preconditioner converges faster than any other preconditioner with same compression ratio.

To complete the results given in fig. 9, we also plot the PCG convergence history, i.e., the

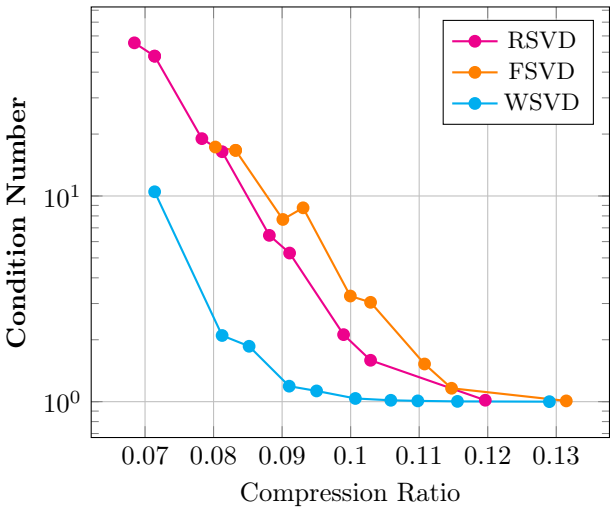


Figure 8: **Condition Number vs. Compression.**

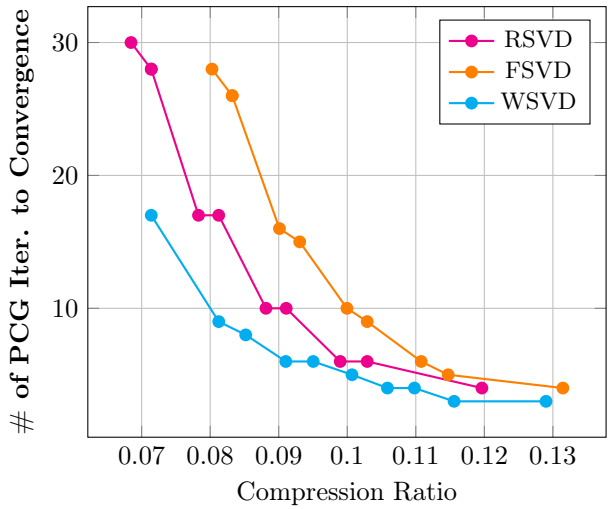


Figure 9: **Number of Iterations to Convergence vs. Compression.**

values $\|Ax_{(i)} - b\|_2/\|b\|_2$ as a function of the iteration number i, for each method and for every τ in fig. 10. All the graphs exhibit steady monotonic decrease in logarithmic scale. This implies that the methods are not sensitive to the specific choice of τ in the sense that regardless of the chosen τ the features of fig. 9 would not dramatically change.

The last point which is explored is the estimation of the spectral bounds α, β eq. (3.17) and the average spectral condition number of the preconditioned system at each level,

$$K_{\text{avg}}^{(\ell)} = \frac{1}{2^\ell} \sum_{k=1}^{2^\ell} \text{cond}_2 \left(\left(\widehat{B}_k^{(\ell)} \right)^T A_k^{(\ell)} \widehat{B}_k^{(\ell)} \right) ,$$

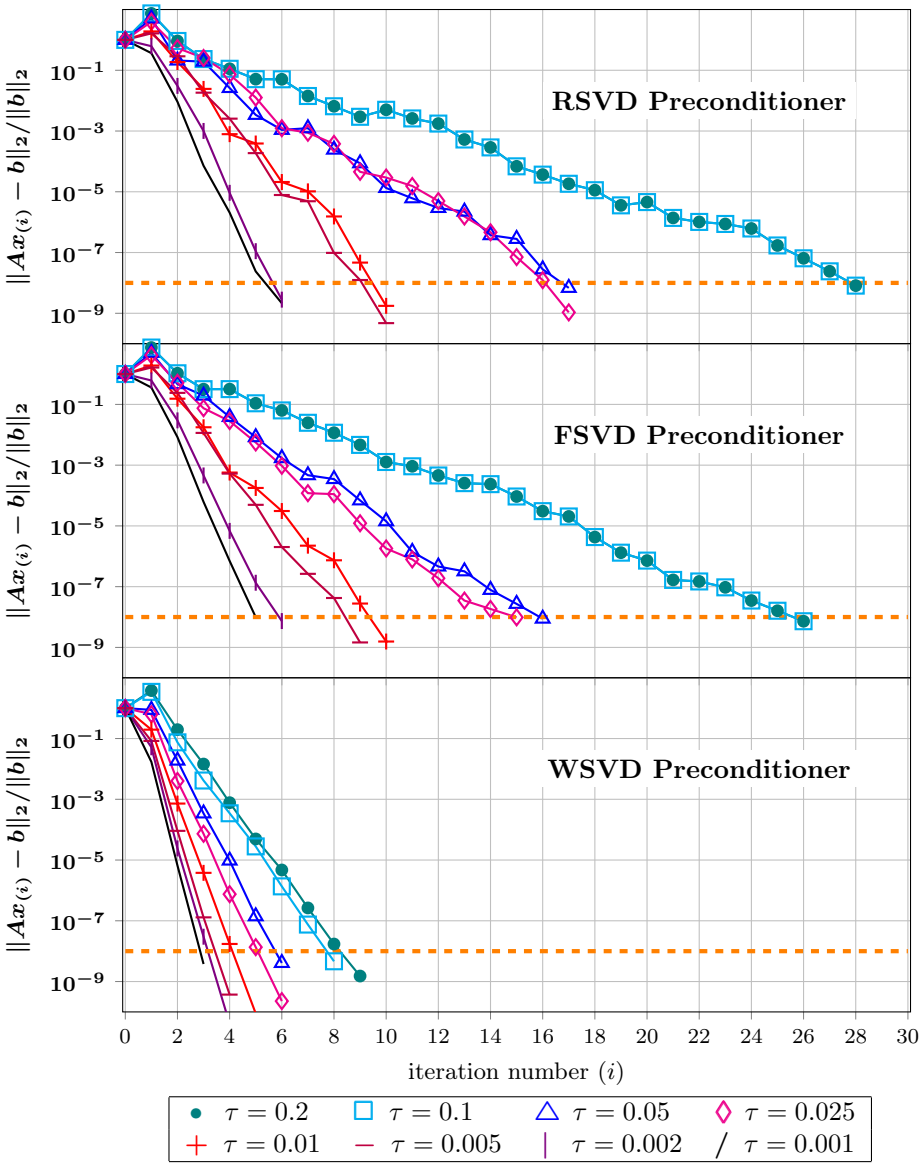


Figure 10: **Preconditioned CG Convergence History.**

using the analytic formulas eqs. (3.19) and (3.20). fig. 11 displays the average estimated spectral bounds and the average exact spectral bounds, i.e., smallest and largest eigenvalues, of the preconditioned matrix and principal submatrices at each level. We observe the following points:

- When the uniform tolerance, τ, is sufficiently small (roughly below 0.01) the prediction is quite good.
- As the uniform tolerance increases, the estimation becomes less accurate.
- For a sufficiently large uniform tolerance the lower spectral bound estimate fails, since the sufficient condition eq. (3.18) is not fulfilled.

The last figure, fig. 12, displays the errors of the average estimated condition number at each level. We examine two errors,

- Average Absolute Condition Number Error: $\left|K_{\text{avg}}^{(\ell)} - \beta/\alpha\right|$.
- Average Condition Number Amplification Error: $\left(K_{\text{avg}}^{(\ell)}\right)^{-1} \cdot \beta/\alpha$.

where β/α is the average ratio of the spectral bounds eqs. (3.19) and (3.20) of all the preconditioned submatrices $\left(\widehat{B}_k^{(\ell)}\right)^T A_k^{(\ell)} \widehat{B}_k^{(\ell)}$ at level ℓ, and the lower level bounds in eq. (3.15) are taken to be the exact extremal eigenvalues.

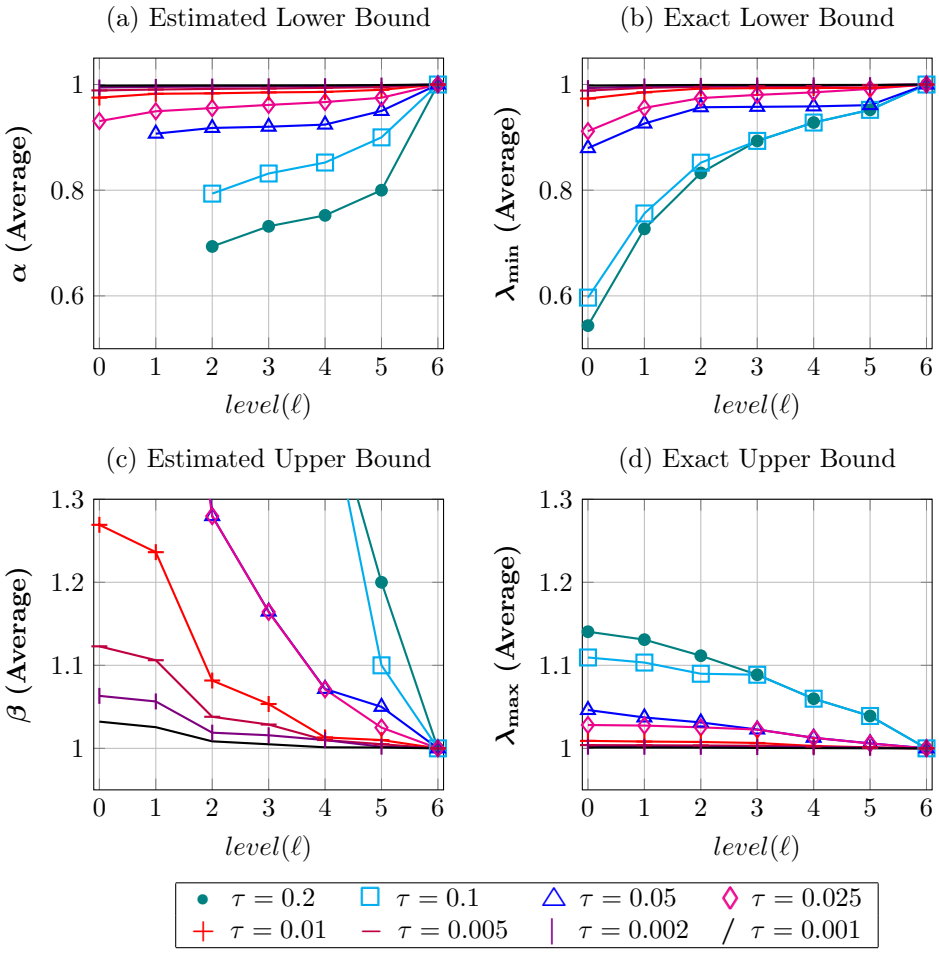


Figure 11: **Avg. Spectral Bounds vs. Level.**

6 Summary and Future Work

We have introduced a new weighted low-rank approximation scheme, which is nearly optimal for preconditioning SPD matrices that can be approximated by the SPD HODLR format. This class

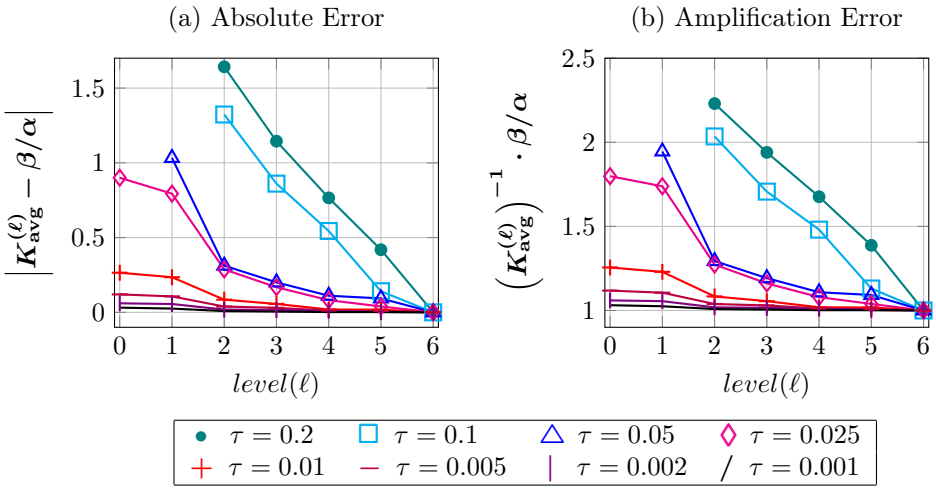


Figure 12: **Avg. Condition Number Error vs. Level.**

of matrices is typical, but not only, to finite element and finite difference matrices arising from the discretization of elliptic partial differential equations. Our theoretical study supported by numerical evidence shows that the preconditioner effectively removes small and large eigenvalues at each level in the hierarchy. Note, that our method is purely algebraic and can be implemented as a black box in a most general manner.

As we have shown the WSVD HODLR preconditioner is advantageous both in terms of achieving low condition number and reduced memory consumption compared to both the regular SVD (RSVD) low-rank approximation and the state-of-the-art filtered SVD (FSVD) low rank approximation. Furthermore, we have demonstrated that our method achieves a faster PCG convergence rate, and results in a significantly reduced number of iterations to convergence compared to the other methods.

The theoretical study of estimation of the spectral bounds of the preconditioned matrix and its principal submatrices works well as long as the tolerance used is small. When the tolerance becomes larger the theory fails, though the preconditioner remains robust and efficient in practice. The theory presented in this work implies that the estimated spectral bounds are attained in a worst case scenario, which is highly unlikely to occur. A future goal is to further improve these predictions by relying on the complete multilevel HODLR structure. This, we hope, will also lead to an effective adaptive approach for achieving low global condition number with minimized complexity and memory consumption.

Finally we note that the experimental example presented here (the 'border' problem) was of limited size and run on a single machine. Current implementation did not include a fast inverse square root algorithm, which would be a crucial ingredient for practical large-scale problems. To verify the efficiency of our method on much larger scale problems, a parallel implementation is necessary. This will be explored in a future study.

A Properties of the Generalized Rayleigh Quotient

Consider the basic problem of finding the extremal points of the generalized Rayleigh quotient

$$R(x) = \frac{x^T A x}{x^T B x}\,,$$

in the region $x \neq 0$, where A is symmetric and B is SPD. Since $R(cx) = R(x)$ for any constant $c \neq 0$ we can restrict the search to the spheroid

$$g(x) = x^T B x = 1\,,$$

on which the quotient reduces to $R(x) = f(x) = x^T A x$.

By the Lagrange multipliers theorem, a necessary condition for x to be an extremal point of $R(x)$ is to find some $\lambda \in \mathbb{R}$ such that

$$\nabla f = \lambda \nabla g\,.$$

Clearly by the symmetry of A,

$$\nabla \left(x^T A x\right) = x^T (A + A^T) = 2x^T A\,.$$

Thus, the necessary condition reduces to

$$Ax = \lambda B x\,.$$

That is, λ is an eigenvalue of $B^{-1}A$ and, equivalently, an eigenvalue of $B^{-1/2}AB^{-1/2}$, where $B^{-1/2}$ is the inverse square root of B.

References

[1] Sivaram Ambikasaran and Eric Darve. An $\mathcal{O}(N \log N)$ fast direct solver for partial hierarchically semi-separable matrices. *Springer Journal of Scientific Computing*, 57(3):477–501, 2013.

[2] AmirHossein Aminfar, Sivaram Ambikasaran, and Eric Darve. A fast block low-rank dense solver with applications to finite-element matrices. *Journal of Computational Physics*, 304:170 – 188, 2016.

[3] O. Axelsson and P. S. Vassilevski. Algebraic multilevel preconditioning methods. i. *Numerische Mathematik*, 56(2):157–177, 1989.

[4] O. Axelsson and P. S. Vassilevski. Algebraic multilevel preconditioning methods, ii. *SIAM Journal on Numerical Analysis*, 27(6):1569–1590, 1990.

[5] M. Bebendorf, M. Bollhöfer, and M. Bratsch. Hierarchical matrix approximation with blockwise constraints. *BIT Numerical Mathematics*, 53(2):311–339, 2013.

[6] M. Bebendorf, M. Bollhöfer, and M. Bratsch. On the spectral equivalence of hierarchical matrix preconditioners for elliptic problems. *Mathematics of Computation*, 85:2839–2861, 2016.

[7] Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.

[8] L. Elsner. A note on optimal block-scaling of matrices. *Numerische Mathematik*, 44(1):127–128, 1984.

[9] Alan Frieze, Ravi Kannan, and Santosh Vempala. Fast monte-carlo algorithms for finding low-rank approximations. *J. ACM*, 51(6):1025–1041, 2004.

[10] Gene H. Golub and Charles F. Van Loan. *Matrix computations, Forth Edition.* Johns Hopkins University Press, 2013.

[11] Lars Grasedyck and Wolfgang Hackbusch. Construction and arithmetics of h-matrices. *Springer Journal of Computing*, 70(4):295–334, 2003.

[12] Ming Gu and Stanley C. Eisenstat. Efficient algorithms for computing a strong rank-revealing qr factorization. *SIAM Journal on Scientific Computing*, 17(4):848–869, 1996.

[13] Pascal Hénon, Pierre Ramet, and Jean Roman. PaStiX: A High-Performance Parallel Direct Solver for Sparse Symmetric Definite Systems. *Parallel Computing*, 28(2):301–321, 2002.

[14] Magnus Rudolph Hestenes and Eduard Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6), 1952.

[15] Tosio Kato. *Perturbation theory for linear operators; 2nd ed.* Grundlehren Math. Wiss. Springer, Berlin, 1976.

[16] Edo Liberty, Franco Woolfe, Per-Gunnar Martinsson, Vladimir Rokhlin, and Mark Tygert. Randomized algorithms for the low-rank approximation of matrices. *Proceedings of the National Academy of Sciences*, 104(51):20167–20172, 2007.

[17] Charles F. Van Loan. Generalizing the singular value decomposition. *SIAM Journal on Numerical Analysis*, 13(1):76–83, 1976.

[18] Bebendorf Mario. *Hierarchical matrices: A Means to Efficiently Solve Elliptic Boundary Value Problems.* Springer Berlin Heidelberg, 2008.

[19] L Miranian and M Gu. Strong rank revealing lu factorizations. *Linear Algebra and its Applications*, 367:1 – 16, 2003.

[20] B. Parlett. *The Symmetric Eigenvalue Problem.* Society for Industrial and Applied Mathematics, 1998.

[21] Sergej Rjasanow. Adaptive cross approximation of dense matrices. In *Int. Association Boundary Element Methods Conf., IABEM*, pages 28–30, 2002.

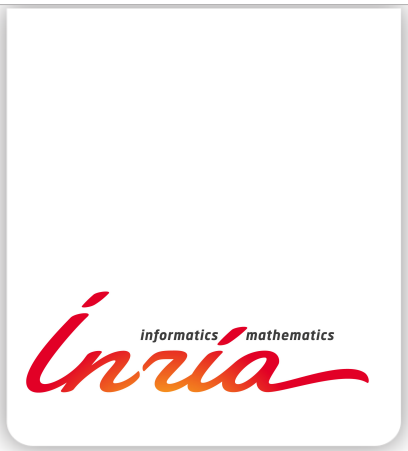
[22] Yousef Saad. *Iterative Methods for Sparse Linear Systems, Second Edition.* Siam, 2003.

[23] Scilab Enterprises. *Scilab: Free and Open Source software for numerical computation.* Scilab Enterprises, Orsay, France, 2012.

[24] Franco Woolfe, Edo Liberty, Vladimir Rokhlin, and Mark Tygert. A fast randomized algorithm for the approximation of matrices. *Applied and Computational Harmonic Analysis*, 25(3):335 – 366, 2008.

[25] K. Yang, H. Pouransari, and E. Darve. Sparse Hierarchical Solvers with Guaranteed Convergence. *ArXiv e-prints*, 2016.

[26] F. Zhang. *The Schur Complement and Its Applications.* Numerical Methods and Algorithms. Springer, 2005.



RESEARCH CENTRE
BORDEAUX – SUD-OUEST

200 avenue de la Vielle Tour
33405 Talence Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399