Low-Rank Factorizations in Data Sparse Hierarchical Algorithms for Preconditioning Symmetric Positive Definite Matrices

Emmanuel Agullo, Eric Darve, Luc Giraud, Yuval Harness

Low-rank Factorizations in Data Sparse Hierarchical Algorithms for Preconditioning Symmetric Positive Definite Matrices

EMMANUEL AGULLO*, ERIC DARVE†, LUC GIRAUD†, AND YUVAL HARNESS†‡

Abstract. We consider the problem of choosing low-rank factorizations in data sparse matrix approximations for preconditioning large scale symmetric positive definite matrices. These approximations are memory efficient schemes that rely on hierarchical matrix partitioning and compression of certain sub-blocks of the matrix. Typically, these matrix approximations can be constructed very fast, and their matrix product can be applied rapidly as well. The common practice is to express the compressed sub-blocks by low-rank factorizations, and the main contribution of this work is the numerical and spectral analysis of SPD preconditioning schemes represented by 2×2 block matrices, whose off-diagonal sub-blocks are low-rank approximations of the original matrix off-diagonal sub-blocks. We propose an optimal choice of low-rank approximations which minimizes the condition number of the preconditioned system, and demonstrate that the analysis can be applied to the class of hierarchically off-diagonal low-rank matrix approximations. Spectral estimates that take into account the error propagation through levels of the hierarchy which quantify the impact of the choice of low-rank compression on the global condition number are provided. The numerical results indicate that the properties of the preconditioning scheme using proper low-rank compression are superior to employing standard choices for low-rank compression. A major goal of this work is to provide an insight into how proper reweighted prior to low-rank compression influences the condition number for a simple case, which would lead to an extended analysis for more general and more efficient hierarchical matrix approximation techniques.

Key words. Preconditioning, Symmetric Positive Definite, Data Sparse, Hierarchical Algorithms, Low-rank Factorization, Minimal Condition Number

AMS subject classifications. 15A16, 15B99, 65F08, 65F30, 65F35, 65F50

1. Introduction.

In this paper we consider preconditioning for iterative solution of large scale linear systems

$$Ax = b\,, \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$ is a *symmetric positive definite* (SPD) matrix. Such systems arise in a wide range of engineering applications, as means to model and understand physical phenomena. Typical example is the result of a finite element discretization of underlying differential equations of a boundary value problem. In many practical applications the matrix A becomes ill-conditioned and, thus, challenging for iterative techniques. In that case the use of preconditioned iterative methods, such as the *preconditioned conjugate gradient* (PCG) [19, 25] technique, becomes an imperative. The choice of a suitable preconditioning scheme can, often, drastically improve the convergence behavior of the iterative method and, generally, plays a vital role in the success of solving the system.

A preconditioning scheme for linear systems is, essentially, composed of linear operations or matrices that approximate A^{-1} (1), but with considerable less computational effort than explicitly inverting A. Transforming the system (1) with such a scheme is called the preconditioned system. The major concern when setting up a preconditioning scheme is to ensure that the preconditioned system has a bounded condition number, and that the number of iterations to convergence in an iterative method remains small while maintaining low associated complexity and reduced mem-

*Inria Bordeaux, France (emmanuel.agullo@inria.fr,luc.giraud@inria.fr,yuval.harness@inria.fr).
†Stanford University, Stanford, CA (eric.darve@stanford.edu).
‡Current address: Department of Applied Mathematics, Tel Aviv University, Israel.

ory cost. The literature on preconditioning techniques is vast, and many robust and efficient methods have been introduced in the last 50 years. These include, among others, incomplete factorization schemes such as ILU and Incomplete Cholesky, sparse matrix approximations, polynomial techniques, domain decomposition methods, as well as multigrid and algebraic multilevel iterations schemes. For a recent comprehensive review on this topic see [26].

The main contribution of this work is the numerical and spectral analysis of SPD preconditioning schemes represented by 2×2 block matrices, whose off-diagonal sub-blocks are low-rank approximations of the original matrix off-diagonal sub-blocks. We re-examine the way low-rank factorizations are obtained, by considering reweighting of the sub-blocks prior to the low-rank compression. Reweighting can be done in many ways, e.g., diagonal scaling, and the fundamental question that we attempt to answer is: which reweighting is optimal with respect to the condition number of the preconditioned system?

The mathematical theory for 2×2 matrices is derived in section 2. We present an optimal 1-level preconditioning scheme using proper reweighting prior to low-rank compression, which minimizes the spectral condition number of the preconditioned system. Thus, a preconditioning scheme employing such low-rank factorizations is expected to attain the same condition number with less computational resources and associated complexity, compared to employing other standard techniques for the low-rank factorizations. Spectral analysis shows that the scheme maps both small and large eigenvalues of the original system exactly to 1. This 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.

In section 3 we propose an application of the 1-level theory for *hierarchically off-diagonal low-rank* (HODLR) matrix structure, as means to demonstrate the applicability of the 1-level theory to the hierarchical multilevel case. We also provide spectral estimates that take into account the error propagation through levels of the hierarchy. This leads to quantification of the impact of the reweighting on the global condition number of the preconditioned system. In essence, weighted HODLR locally minimizes the condition number at each level of the hierarchy by approximately filtering the smallest and largest eigenvalues. Since this approach is employed hierarchically, it effectively creates a strong effect of global spectrum clustering.

The HODLR structure is a member of a wide class of hierarchical data sparse approximations. These approximations rely on the fact that the matrix can be subdivided into a hierarchy of smaller block matrices, and certain sub-blocks can be efficiently approximated as low-rank matrices by low-rank factorizations. The low-rank compressions of sufficient sub-blocks leads to a dramatic reduction of the complexity and computational cost. The best known example for such schemes is the class of hierarchical matrix ($\mathcal{H}$-matrix) approximations [15, 17, 18, 5] which has gained growing attention in recent years.

To the contrary of the more general strong hierarchical matrix structure which allows further decomposition of the off-diagonal blocks into low-rank and full-rank blocks, HODLR is a weak hierarchical matrix structure, which relies on a single low-rank compression for the off-diagonal blocks. Closely related to HODLR is the *hierarchically semi-separable* (HSS) [8, 28] structure , which is, in fact, a HODLR matrix format possessing a nested off-diagonal low-rank structure.

Essentially, weak hierarchical methods, i.e., HODLR and HSS, are not considered competitive with the more general strong hierarchical matrix methods, when the underlying problem is of very large scale. However, the proposed study provides a novel theoretical basis for optimality conditions of hierarchical preconditioning schemes. Thus, the presented analysis can serve as starting point for a more general theory on optimal $\mathcal{H}$-matrix preconditioning which is deferred to future work.

The weighted HODLR scheme is similar in nature to the methods proposed in [29, 30] which propose practical HSS schemes that rely on similar ideas of reweighting prior to compression, but without the complete numerical and spectral analysis of this study. The costs to apply these multilevel preconditioners are about $\mathcal{O}(n)$, where n is the matrix size.

The experimental part of this work, whose goal is to demonstrate the effectiveness of properly chosen reweighting for the low-rank approximations, is given in section 4. The section contains a detailed comparative study of HODLR preconditioning using different methods for the low-rank compressions. As alternatives to the proper reweighting strategy, we consider the conventional low-rank approximation in the 2-norm and the low-rank approximation with constraints [6]. The latter employs low-rank approximations that also preserve constraints, forcing sub-blocks of the preconditioning scheme to be identical to the corresponding sub-blocks of the input matrix on predetermined subspaces. Employing the method for preconditioning SPD matrices of discretized elliptic PDEs has been demonstrated in [7], and a similar approach for non-symmetric sparse matrices has been recently suggested in [31].

The numerical results indicate, that employing proper reweighting prior to low-rank compression, leads to a HODLR preconditioning scheme that requires far less computational resources for the same quality of convergence performance compared to using other low-rank compression techniques. The experiments also show, that the HODLR preconditioning scheme with proper reweighting retains the SPD property of the system when other standard techniques fail, and remains efficient and robust even if low accuracy compression is employed with ranks of $\mathcal{O}(1)$ for the low-rank approximations of the sub-blocks. Summary and plans for future work follow in section 5.

2. The Optimal One-level Preconditioning Scheme.

In this section we introduce the optimal 1-level scheme for the preconditioning of SPD matrices. We consider an input $n \times n$ SPD matrix A with a 2×2 block structure and a corresponding 1-level approximation K,

$$A = \begin{bmatrix} A_1 & M \\ M^T & A_2 \end{bmatrix}, \quad K = \begin{bmatrix} A_1 & U_1 S V_2^T \\ V_2 S U_1^T & A_2 \end{bmatrix}, \quad A_i \in \mathbb{R}^{n_i \times n_i}, \tag{2}$$

where $n = n_1 + n_2$, and the off-diagonal blocks of K are low-rank factorizations satisfying

$$U_1 \in \mathbb{R}^{n_1 \times r}, \quad S \in \mathbb{R}^{r \times r}, \quad V_2 \in \mathbb{R}^{n_2 \times r}, \tag{3}$$

with a, typically, small rank r. The matrix U_1 is the *interpolation operator*, the matrix V_2 is the *anterpolation operator*, and the matrix S whose rank is r is known as the *interaction operator*. In some cases S is omitted, i.e., equivalently represented by an $r \times r$ identity matrix.

We present an explicit formula for a 1-level approximation, K (2), which minimizes the spectral condition number

$$\operatorname{cond}_2\left(R^{-T} A R^{-1}\right) = \left\|R^{-T} A R^{-1}\right\|_2 \cdot \left\|R A^{-1} R^T\right\|_2, \tag{4}$$

of the preconditioned system,

$$R^{-T}AR^{-1}x = R^{-T}y\,, \tag{5}$$

for any given rank $r = 0, 1, \ldots, \min\{n_1, n_2\}$, where $\|\cdot\|_2$ is the 2-norm, and R denotes any square root (not necessarily principal) of K in the sense that

$$K = R^T R \in \mathbb{R}^{n\times n}\,. \tag{6}$$

The key idea is to reweight the off-diagonal blocks prior the low-rank factorization. Proper choice of reweighting leads to a minimum spectral condition number of the preconditioned system as well as clustering of the spectrum of the preconditioned system around 1.

We begin in subsection 2.1 by introducing the method for obtaining the minimum condition number low-rank approximation. In subsection 2.2 we provide the theorem on the minimum condition number property, including a detailed description of the spectral properties of the preconditioned system. A rigorous and detailed proof of the theorem is given in subsection 2.3.

2.1. Explicit Formula of the Optimal One-level Scheme. The construction of the minimum condition number 1-level preconditioner K (2) subject to

$$\operatorname{rk}\left(U_1 S V_2^T\right) \leq r\,, \tag{7}$$

is based on the following two-step method ensuring that the preconditioned matrix $R^{-T}AR^{-1}$ also inherits the SPD property of A:

1. Apply a two-sided block Jacobi transformation,

$$\begin{bmatrix} R_1^{-T} & 0 \\ 0 & R_2^{-T} \end{bmatrix} \cdot A \cdot \begin{bmatrix} R_1^{-1} & 0 \\ 0 & R_2^{-1} \end{bmatrix} = \begin{bmatrix} I_1 & R_1^{-T} M R_2^{-1} \\ R_2^{-T} M^T R_1^{-1} & I_2 \end{bmatrix}, \tag{8}$$

   where I_i denotes the $n_i \times n_i$ identity matrix, and $R_i \in \mathbb{R}^{n_i\times n_i}$ denotes a square root of A_i i.e., $R_i^T R_i = A_i$.
2. Extract the off-diagonal triple products (3) by setting,

$$U_1 = R_1^T \mathcal{U}_r\,, \quad S = \Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)\,, \quad V_2 = R_2^T \mathcal{V}_r\,, \tag{9}$$

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

$$R_1^{-T} M R_2^{-1} = \mathcal{U}\Sigma\mathcal{V}^T\,, \quad \Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_{\min\{n_1,n_2\}})\,.$$

The theory presented in this study implies, that in practice for a given rank bound $r \geq 0$, any low-rank factorization, $U_1 S V_2^T$, satisfying

$$\left\|R_2^{-T} M^T R_1^{-1} - R_2^{-T} U_1 S V_2^T R_1^{-1}\right\|_2 \leq \left\|R_2^{-T} M^T R_1^{-1} - \mathcal{U}_r \Sigma_r \mathcal{V}_r^T\right\|_2\,,$$

would achieve the same minimal spectral condition number. However, the truncated SVD of the reweighted off-diagonal block, $\mathcal{U}_r\Sigma_r\mathcal{V}_r^T$, also ensures that the spectrum of the preconditioned system is optimally clustered. This observation is discussed and explained in the next subsection.

2.2. Minimal Condition Number and Spectral Analysis. Let us now focus on the spectral properties of the preconditioned system, $R^{-T}AR^{-1}$, where R is a square root of K (2) whose off-diagonal low-rank blocks are given by (9). First, let us consider the degenerate case $r = 0$. In this case $U_1SV_2^T = 0$ and the square root of K reduces to the following block diagonal form,

$$R(r = 0) = \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} . \tag{10}$$

The preconditioning scheme (5) with $r = 0$ is, in fact, the two-sided block Jacobi (8). There is a known result [12] showing that the two-sided block Jacobi preconditioner (8) is optimal, in the sense that

$$\mathrm{cond}_2 \left(\begin{bmatrix} R_1^{-T} & 0 \\ 0 & R_2^{-T} \end{bmatrix} A \begin{bmatrix} R_1^{-1} & 0 \\ 0 & R_2^{-1} \end{bmatrix} \right) \leq \mathrm{cond}_2 \left(\begin{bmatrix} B_1^T & 0 \\ 0 & B_2^T \end{bmatrix} A \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix} \right) ,$$

for any non-singular $\left[\begin{smallmatrix} B_1 & 0 \\ 0 & B_2 \end{smallmatrix}\right]$ with the same dimensions and partition as $R(r = 0)$ (10). The analysis we present, thus, naturally extends this classic result.

The main results for the general case $r \geq 0$ are presented in Theorem 1, whose principal component is the spectral analysis of the preconditioned system. Our proof shows that the spectrum of the two-sided block Jacobi preconditioned system (8) contains (or equals to)

$$1 + \sigma_1, \ldots, 1 + \sigma_{min\{n_1,n_2\}}, 1 - \sigma_{min\{n_1,n_2\}}, \ldots, 1 - \sigma_1 ,$$

where $1 - \sigma_1$ and $1 + \sigma_1$ are the smallest and largest, respectively, eigenvalues of the preconditioned system. Thus, the two-sided block Jacobi redistributes the spectrum of the matrix symmetrically around 1. We show that the optimal 1-level preconditioning scheme 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 (8) exactly to 1. Hence, the spectral condition number (4) as a function of r is

$$\mathrm{cond}_2 \left(R^{-T}AR^{-1} \right) = \frac{1 + \sigma_{r+1}}{1 - \sigma_{r+1}} .$$

An illustration of the spectral clustering done by the optimal 1-level preconditioning scheme is displayed in Figure 1.

THEOREM 1. *Let*

$$A = \begin{bmatrix} A_1 & M \\ M^T & A_2 \end{bmatrix} , \quad K = \begin{bmatrix} A_1 & U_1SV_2^T \\ V_2SU_1^T & A_2 \end{bmatrix} ,$$

have the same dimensions and partition where A is SPD, and let R_i denote a square root of A_i, i.e., $A_i = R_i^TR_i$.

If the off-diagonal triple product approximation $U_1SV_2^T$ satisfies

$$U_1 = R_1^T\mathcal{U}_r , \quad S = \Sigma_r , \quad V_2 = R_2^T\mathcal{V}_r ,$$

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

$$R_1^{-T}MR_2^{-1} = \mathcal{U}\Sigma\mathcal{V}^T , \quad \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_{\min\{n_1,n_2\}}) , \tag{11}$$

then:

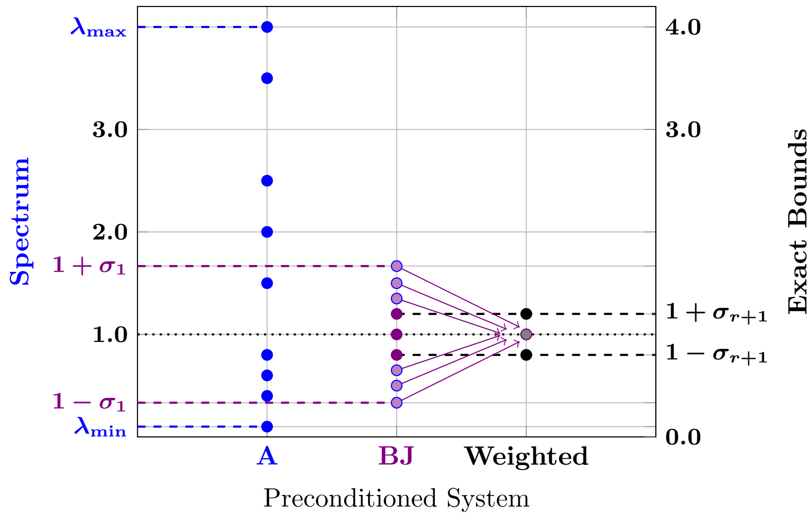

Fig. 1: **Spectrum Clustering of the Optimal 1-level Preconditioning Scheme.** The spectrum of some SPD matrix A and the transformations it goes after preconditioning by block Jacobi (BJ) and the optimal 1-level preconditioning scheme are displayed. The spectra are ordered from the left to the right starting from A, followed by BJ and end up with the optimal scheme.

1. *The matrix K is SPD and possesses a square root, R, i.e., $K = R^T R$.*
2. *For any $r < \min\{n_1, n_2\}$ and any square root R, the spectrum of the preconditioned system is contained in $]0, 2[$ and equal to*

$$\left\{1+\sigma_{r+1}, \ldots, 1+\sigma_{\min\{n_1,n_2\}}, 1, 1-\sigma_{\min\{n_1,n_2\}}, \ldots, 1-\sigma_{r+1}\right\} .$$

3. *Any inverse square root of K, R^{-1}, is a minimizer of the spectral condition number (4) in the sense that*

$$\operatorname{cond}_2\left(R^{-T} A R^{-1}\right) \leq \operatorname{cond}_2\left(\widehat{R}^{-T} A \widehat{R}^{-1}\right), \quad \widehat{R}^T \widehat{R} = \widehat{K},$$

for any SPD matrix with the same dimensions and partition as K of the form,

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

whose off-diagonal blocks satisfy $\operatorname{rk}(\widehat{M}) \leq r$.

2.3. Proof of Theorem 1. The proof of Theorem 1 regarding the spectral properties relies on Lemma 1, while the minimum condition number property is based on the *Cauchy (eigenvalue) interlacing theorem* [23, p. 202]. The latter 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 $i = 1, \dots, r$,

$$\lambda_i(H) \geq \lambda_i(E) \geq \lambda_{i+n-r}(H)\,, \tag{12}$$

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

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

LEMMA 1. *Let* $H = \left[\begin{smallmatrix} \delta I_1 & \mathcal{M} \\ \mathcal{M}^T & \delta I_2 \end{smallmatrix}\right] \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$ *where* I_i *is the* $n_i \times n_i$ *identity matrix and* $\delta \in \mathbb{R}$*, and let* $\sigma_1, \dots, \sigma_{\min\{n_1,n_2\}}$ *denote the singular values of* $\mathcal{M}$*.*

1. *If* $n_1 = n_2$ *then*

$$\operatorname{spec}(H) = \{\delta - \sigma_1, \dots, \delta - \sigma_{n_1}, \delta + \sigma_{n_1}, \dots, \delta + \sigma_1\}\,.$$

2. *If* $n_1 \neq n_2$ *then*

$$\operatorname{spec}(H) = \{\delta - \sigma_1, \dots, \delta - \sigma_{\min\{n_1,n_2\}}, \delta + \sigma_{\min\{n_1,n_2\}}, \dots, \delta + \sigma_1\} \cup \{\delta\}\,,$$

where the multiplicity of the eigenvalue δ *is at least* $|n_1 - n_2|$*.*

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

$$\mathcal{W} = \frac{1}{\sqrt{2}}\begin{bmatrix} \widetilde{\mathcal{U}} & \mathcal{U} \\ \widetilde{\mathcal{V}} & -\mathcal{V} \end{bmatrix} \in \mathbb{R}^{n\times n}\,, \tag{13}$$

whose blocks are given by

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

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

$$\mathcal{W}^T \begin{bmatrix} 0 & \mathcal{M} \\ \mathcal{M}^T & 0 \end{bmatrix} \mathcal{W} = \begin{bmatrix} \mathcal{S}_1 & 0 \\ 0 & -\mathcal{S}_2 \end{bmatrix}\,,$$

where $\mathcal{S}_i = \operatorname{diag}\left[\sigma_1, \dots, \sigma_{\min\{n_1,n_2\}}, 0, \dots, 0\right] \in \mathbb{R}^{n_i\times n_i}$. Thus, by the orthogonality of $\mathcal{W}$ we obtain

$$\mathcal{W}^T \begin{bmatrix} \delta I_1 & \mathcal{M} \\ \mathcal{M}^T & \delta I_2 \end{bmatrix} \mathcal{W} = \begin{bmatrix} \delta I_1 + \mathcal{S}_1 & 0 \\ 0 & \delta I_2 - \mathcal{S}_2 \end{bmatrix}\,.$$

Hence, the spectrum of H is given by

$$\operatorname{spec}(H) = \{\delta - \sigma_1, \dots, \delta - \sigma_{\min\{n_1,n_2\}}, \delta + \sigma_{\min\{n_1,n_2\}}, \dots, \delta + \sigma_1\} \cup \{\delta\}\,,$$

where the multiplicity of δ is at least $n_1 - n_2$. Note that in case $n_2 > n_1$, we can simply interchange the principal blocks of H by reordering the columns and rows of H, and repeat the proof. ∎

Proof. of Theorem 1.
Let $\widehat{K}$ be a partitioned SPD matrix with the same dimensions and partition as K (2) whose off-diagonal blocks rank is bounded by r,

$$\widehat{K} = \begin{bmatrix} A_1 & U_1 S V_2^T \\ V_2 S U_1^T & A_2 \end{bmatrix}, \quad \mathrm{rk}(U_1 S V_2^T) \leq r\,.$$

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

$$\widehat{R}^{-T} A \widehat{R}^{-1}\,, \quad \widehat{K} = \widehat{R}^T \widehat{R}\,, \tag{14}$$

then by employing the change of variables, $\zeta = \widehat{R}\xi$, we obtain

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

Since $\widehat{R}^{-1}\widehat{R}^{-T} = \widehat{K}^{-1}$, we conclude that regardless to the particular choice of square root, $\widehat{R}$, the spectrum of the preconditioned system (14) remains unchanged.

Let $R_i \in \mathbb{R}^{n_i \times n_i}$ denote a, generally, non-symmetric square root of A_i. By direct calculations we obtain

$$\begin{bmatrix} R_1^{-T} & 0 \\ 0 & R_2^{-T} \end{bmatrix} A \begin{bmatrix} R_1^{-1} & 0 \\ 0 & R_2^{-1} \end{bmatrix} = \begin{bmatrix} I_1 & R_1^{-T} M R_2^{-1} \\ R_2^{-T} M^T R_1^{-1} & I_2 \end{bmatrix}, \tag{15}$$

and by Lemma 1, the spectrum of (15) is contained in (or equal to)

$$\{1+\sigma_1, \ldots, 1+\sigma_{\min\{n_1,n_2\}}, 1, 1-\sigma_{\min\{n_1,n_2\}}, \ldots, 1-\sigma_1\}\,,$$

where σ_i are the singular values of $R_1^{-T} M R_2^{-1}$. Since R_i are non-singular, the preconditioned matrix (15) is SPD. Hence, we have

$$1 - \sigma_1 > 0 \quad \Rightarrow \quad \mathrm{spec}\left(\begin{bmatrix} I_1 & R_1^{-T} M R_2^{-1} \\ R_2^{-T} M^T R_1^{-1} & I_2 \end{bmatrix} \right) \subset]0, 2[\,.$$

Consider the specific choice of inverse square root of $\widehat{K}$,

$$\widehat{R}^{-1} = \begin{bmatrix} R_1^{-1} & 0 \\ 0 & R_2^{-1} \end{bmatrix} \widehat{\mathcal{W}} \widehat{\mathcal{D}}^{-1/2} \widehat{\mathcal{W}}^T, \quad \widehat{\mathcal{D}} = \begin{bmatrix} I_1 + \widehat{\mathcal{S}}_{1,r} & 0 \\ 0 & I_2 - \widehat{\mathcal{S}}_{2,r} \end{bmatrix}.$$

The matrix $\widehat{\mathcal{W}}$ is orthogonal of the same form as (13) in the proof of Lemma 1 with respect to (15), and $\widehat{\mathcal{S}}_{i,r} = \mathrm{diag}\,(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \in \mathbb{R}^{n_i \times n_i}$ where σ_i are the singular values of $R_1^{-T} U_1 S V_2^T R_2^{-1}$. Setting the choice $\widehat{R}^{-1}$ into (4) and employing the fact that the 2-norm is invariant under unitary transformations, we obtain

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

where H is an SPD matrix given by

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

The matrices $\mathcal{W}$ (13) and $\mathcal{S}_{i,m}$ are defined and constructed in the proof of Lemma 1. Note that like $\widehat{\mathcal{W}}$, the matrix $\mathcal{W}$ is orthogonal. Hence, the product $\mathcal{W}^T \widehat{\mathcal{W}}$ is also an orthogonal matrix.

Our definitions so far indicate that the following diagonal matrices,

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

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

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

in the sense that $\left(\widehat{\mathcal{D}} - \underline{\mathcal{D}}\right)$ and $\left(\overline{\mathcal{D}} - \widehat{\mathcal{D}}\right)$ are non-negative definite. Thus, applying the change of variables $\xi = \widetilde{\mathcal{D}}^{-1/2}x$ and exploiting the properties of the Rayleigh quotient, we can write

$$\begin{aligned} \left\|\widehat{\mathcal{D}}^{-1/2}H\widehat{\mathcal{D}}^{-1/2}\right\|_2 &= \max_{x\neq 0} \frac{x^T\widehat{\mathcal{D}}^{-1/2}H\widehat{\mathcal{D}}^{-1/2}x}{x^Tx} = \max_{\xi\neq 0} \frac{\xi^T H\xi}{\xi^T\widehat{\mathcal{D}}\xi} \\ &\geq \max_{\xi\neq 0} \frac{\xi^T H\xi}{\xi^T\overline{\mathcal{D}}\xi} = \max_{y\neq 0} \frac{y^T\overline{\mathcal{D}}^{-1/2}H\overline{\mathcal{D}}^{-1/2}y}{y^Ty} = \left\|\overline{\mathcal{D}}^{-1/2}H\overline{\mathcal{D}}^{-1/2}\right\|_2, \end{aligned} \tag{16}$$

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

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

Let $Z = \text{span}\{e_{r+1}, \ldots, e_n\}$ where e_i denotes the i-th canonical basis vector, and let P_Z denote the orthogonal projection matrix on Z. The structure of $\overline{\mathcal{D}}$ implies

$$\left\|\overline{\mathcal{D}}^{-1/2}H\overline{\mathcal{D}}^{-1/2}\right\|_2 \geq \max_{P_Zy\neq 0} \frac{(P_Zy)^T\overline{\mathcal{D}}^{-1/2}H\overline{\mathcal{D}}^{-1/2}P_Zy}{(P_Zy)^TP_Zy} = \max_{P_Zy\neq 0} \frac{(P_Zy)^THP_Zy}{(P_Zy)^TP_Zy}.$$

Essentially, $P_Z^THP_Z$ represents an $(n-r)\times(n-r)$ principal block of a unitarily equivalent matrix of H whose eigenvalues are identical to H. Thus, by the Cauchy interlacing theorem (12),

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

Applying similar arguments it can also be shown that

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

which leads to the following lower bound on the spectral condition number,

$$\text{cond}_2(R^{-T}AR^{-1}) \geq \frac{1+\sigma_{r+1}}{1-\sigma_{r+1}}.$$

Finally, let us consider the specific choice $U_1SV_2^T = R_1^T\mathcal{U}_r\Sigma\mathcal{V}_r^TR_2$ 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 $R_1^{-T}MR_2^{-1}$. Consequently, we have $\sigma_i = \sigma_i$, $i = 1, \ldots, r$. Thus, setting accordingly $\widehat{\mathcal{W}} = \mathcal{W}$ and $\widehat{R} = R$ we obtain by direct calculations:

$$\text{cond}_2(R^{-T}AR^{-1}) = \left\|R^{-T}AR^{-1}\right\|_2\left\|RA^{-1}R^T\right\|_2 = \frac{1+\sigma_{r+1}}{1-\sigma_{r+1}},$$

and the proof is complete. ∎

3. The Multilevel Weighted HODLR Preconditioning Scheme.

In this section we introduce the multilevel HODLR preconditioning scheme for SPD matrices. The method is based on the theory presented in section 2 and relaxation of the original problem. The motivation is twofold. First we demonstrate that the 1-level analysis can be extended to a multilevel preconditioning scheme. Second, we provide spectral bounds on the eigenvalues of the preconditioned system which give account for error propagation through the levels of the hierarchy.

In subsection 3.1 we give a brief review of the HODLR matrix structure which will be employed throughout the remainder of this paper. We focus on the symmetric case, since this work is concerned with the preconditioning of SPD matrices. In subsection 3.2 we introduce the preconditioning scheme, which is based on hierarchical construction and fast application of the inverse square roots, R^{-1} and R^{-T}. In subsection 3.3 we briefly consider the associated memory and the computational costs of constructing and applying the scheme. An in-depth spectral analysis is presented in subsection 3.4. Our analysis provides estimates of the spectral bounds of the preconditioned system at each level, that take into account the approximation error at the lower levels of the hierarchy. These bounds are, however, of qualitative nature as they reflect a possible worst case scenario which is not likely to occur in practice. A rigorous and detailed proof of the theory is given in subsection 3.5.

3.1. Symmetric HODLR Matrix Structure.

A symmetric HODLR matrix, $K \in \mathbb{R}^{n\times n}$, can be described in the following recursive manner,

$$K = K_1^{(0)}\,, \quad K_k^{(\ell)} = \begin{bmatrix} K_{2k-1}^{(\ell+1)} & U_{2k-1}^{(\ell+1)} S_k^{(\ell)} {V_{2k}^{(\ell+1)}}^T \\ V_{2k}^{(\ell+1)} S_k^{(\ell)} {U_{2k-1}^{(\ell+1)}}^T & K_{2k}^{(\ell+1)} \end{bmatrix} \in \mathbb{R}^{n_k^{(\ell)}\times n_k^{(\ell)}}\,, \tag{17}$$

for $\ell = 0, 1, \ldots, L-1$ and $k = 1, 2, \ldots, 2^\ell$, where ℓ is the level of $K_k^{(\ell)}$ in the hierarchy. The triple products, $U_{2k-1}^{(\ell+1)} S_k^{(\ell)} {V_{2k}^{(\ell+1)}}^T$, represent low-rank blocks in the sense that

$$U_{2k-1}^{(\ell+1)} \in \mathbb{R}^{n_{2k-1}^{(\ell+1)}\times r_k^{(\ell)}}\,, \quad V_{2k}^{(\ell+1)} \in \mathbb{R}^{n_{2k}^{(\ell+1)}\times r_k^{(\ell)}}\,, \quad S_k^{(\ell)} \in \mathbb{R}^{r_k^{(\ell)}\times r_k^{(\ell)}}\,, \tag{18}$$

where $r_k^{(\ell)}$ is the rank of the corresponding off-diagonal block of K. Typically, $r_k^{(\ell)} \ll n_{2k-1}^{(\ell+1)}, n_{2k}^{(\ell+1)}$. An illustration of the hierarchical structure of K is displayed in Figure 2.

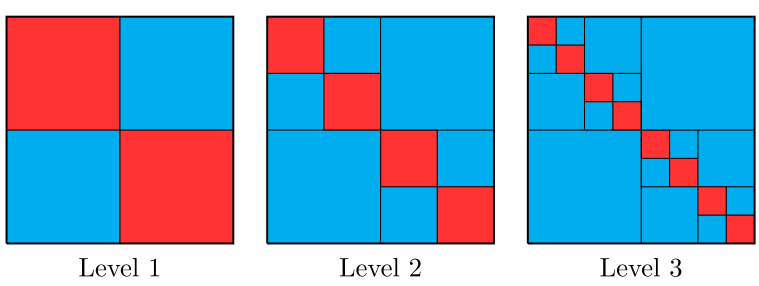

Fig. 2: **The HODLR Structure.** The first 3 levels, $\ell = 1, 2, 3$, of the HODLR structure are illustrated: at each level the blue color blocks are the low-rank off-diagonal blocks and the red blocks are the HODLR principal submatrices.

The common practice is to set the HODLR matrix as an approximation of a given matrix, $A \in \mathbb{R}^{n\times n}$. The low-rank off-diagonal blocks of K satisfy

$$\left\| M_k^{(\ell)} - U_{2k-1}^{(\ell+1)} S_k^{(\ell)} V_{2k}^{(\ell+1)^T} \right\|_2 \le \tau_k^{(\ell)} \cdot \left\| M_k^{(\ell)} \right\|_2 , \tag{19}$$

where $M_k^{(\ell)}$ denotes the corresponding off-diagonal block of A and $\tau_k^{(\ell)} > 0$ is a chosen tolerance. Typically, a prior reordering of the matrix rows and columns is needed to confirm that $r_k^{(\ell)}$ are, indeed, low.

For obtaining low-rank approximations satisfying (19), the low-rank *singular value decomposition* (SVD) [14] which originated in [11] is, generally, considered the best choice, since it is optimal with respect to any unitarily invariant norm (2-norm, Frobenius). The computational cost required to obtain the SVD of $M_k^{(\ell)}$ is relatively expensive necessitating 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 [22], rank revealing QR [16], randomized algorithms [13, 20, 27], adaptive cross approximation [24] and boundary distance low-rank [2]. For more details see a review on this topic in [2].

3.2. Recursive Formula of the Multilevel Preconditioning Scheme. By our definitions (17) the principal blocks of K and the corresponding blocks of A are described by

$$K_k^{(\ell)} = \begin{bmatrix} K_{2k-1}^{(\ell+1)} & U_{2k-1}^{(\ell+1)} S_k^{(\ell)} V_{2k}^{(\ell+1)^T} \\ V_{2k}^{(\ell+1)} S_k^{(\ell)} U_{2k-1}^{(\ell+1)^T} & K_{2k}^{(\ell+1)} \end{bmatrix}, A_k^{(\ell)} = \begin{bmatrix} A_{2k-1}^{(\ell+1)} & M_k^{(\ell)} \\ M_k^{(\ell)^T} & A_{2k}^{(\ell+1)} \end{bmatrix}, \tag{20}$$

where $K_k^{(\ell)}, A_k^{(\ell)} \in \mathbb{R}^{n_k^{(\ell)}\times n_k^{(\ell)}}$, and the rank of each off-diagonal triple product approximation satisfies

$$\operatorname{rk}\left(U_{2k-1}^{(\ell+1)} S_k^{(\ell)} V_{2k}^{(\ell+1)^T}\right) \le r_k^{(\ell)} . \tag{21}$$

The key idea we propose is to construct each $K_k^{(\ell)}$ as an optimal 1-level preconditioning scheme of the matrix

$$B_k^{(\ell)} = \begin{bmatrix} K_{2k-1}^{(\ell+1)} & M_k^{(\ell)} \\ M_k^{(\ell)^T} & K_{2k}^{(\ell+1)} \end{bmatrix} \in \mathbb{R}^{n_k^{(\ell)}\times n_k^{(\ell)}} ,$$

which is obtained by replacing the principal blocks of $A_k^{(\ell)}$ with the corresponding principal blocks of $K_k^{(\ell)}$. This is a relaxation of the original problem, which facilitates a fast construction method. The resulting preconditioned global system condition number is no longer ensured to be minimal. However, the numerical results in section 4 indicate that the proposed approach is highly robust and, in general, attains superior condition number compared to HODLR approximations using other low-rank approximation schemes.

Before proceeding we introduce some necessary notations. Let $R_k^{(\ell)} \in \mathbb{R}^{n_k^{(\ell)}\times n_k^{(\ell)}}$ denote the square root of $K_k^{(\ell)}$ in the sense that $K_k^{(\ell)} = R_k^{(\ell)^T} R_k^{(\ell)}$. Let $\mathcal{U}_{k,r}^{(\ell)} \in$

$\mathbb{R}^{n_{2k-1}^{(\ell+1)} \times r_k^{(\ell)}}$ and $\mathcal{V}_{k,r}^{(\ell)} \in \mathbb{R}^{n_{2k}^{(\ell+1)} \times r_k^{(\ell)}}$ be two thin matrices with orthogonal columns composed of the first $r_k^{(\ell)}$ left and right, respectively, singular vectors of the SVD,

$$R_{2k-1}^{(\ell+1)^{-T}} M_k^{(\ell)} R_{2k}^{(\ell+1)^{-1}} = \mathcal{U}_k^{(\ell)} \Sigma_k^{(\ell)} \mathcal{V}_k^{(\ell)^T} , \tag{22}$$

where $\Sigma_{k,r}^{(\ell)} \in \mathbb{R}^{r_k^{(\ell)} \times r_k^{(\ell)}}$ is the principal submatrix of

$$\Sigma_k^{(\ell)} = \operatorname{diag}(\sigma_{k,1}^{(\ell)}, \ldots, \sigma_{k,\min\{n_{2k-1}^{(\ell+1)}, n_{2k}^{(\ell+1)}\}}^{(\ell)}) .$$

For brevity and clarity we will abuse the notation and employ $\mathcal{U} = \mathcal{U}_{k,r}^{(\ell)}$ and $\mathcal{V} = \mathcal{V}_{k,r}^{(\ell)}$. The proof of Theorem 1 implies that the following recursive formulas for a fast application of the inverse square roots hold,

$$R_k^{(\ell)^{-T}} = \left(I + \frac{1}{2} \begin{bmatrix} \mathcal{U}_r & \mathcal{U}_r \\ \mathcal{V}_r & -\mathcal{V}_r \end{bmatrix} \begin{bmatrix} \mathcal{S}_r^+ - I & 0 \\ 0 & \mathcal{S}_r^- - I \end{bmatrix} \begin{bmatrix} \mathcal{U}_r^T & \mathcal{V}_r^T \\ \mathcal{U}_r^T & -\mathcal{V}_r^T \end{bmatrix} \right) \begin{bmatrix} R_{2k-1}^{(\ell+1)} & 0 \\ 0 & R_{2k}^{(\ell+1)} \end{bmatrix}^{-T} ,$$

and

$$R_k^{(\ell)^{-1}} = \begin{bmatrix} R_{2k-1}^{(\ell+1)} & 0 \\ 0 & R_{2k}^{(\ell+1)} \end{bmatrix}^{-1} \left(I + \frac{1}{2} \begin{bmatrix} \mathcal{U}_r & \mathcal{U}_r \\ \mathcal{V}_r & -\mathcal{V}_r \end{bmatrix} \begin{bmatrix} \mathcal{S}_r^+ - I & 0 \\ 0 & \mathcal{S}_r^- - I \end{bmatrix} \begin{bmatrix} \mathcal{U}_r^T & \mathcal{V}_r^T \\ \mathcal{U}_r^T & -\mathcal{V}_r^T \end{bmatrix} \right) ,$$

where $\mathcal{S}_r^{\pm} = (I \pm \Sigma_{k,r}^{(\ell)})^{-\frac{1}{2}}$. These formulas can be verified by writing the product $R_k^{(\ell)^{-T}} K_k^{(\ell)} R_k^{(\ell)^{-1}}$ which, indeed, equals to the identity matrix, assuming $K_k^{(\ell)}$ is SPD.

3.3. Utilization and Construction of the Preconditioning Scheme. The recursive representations of $R_k^{(\ell)^{-1}}$ and $R_k^{(\ell)^{-T}}$ indicate that both matrices, essentially, posses HODLR structure. Thus, $R_k^{(\ell)^{-1}}$ and $R_k^{(\ell)^{-T}}$ can be applied relatively fast in matrix product operations. In the case that a constant average rank, $r^{(\ell)} = \mathcal{O}(r)$, is taken for the off-diagonal blocks, the recursive representations implies the following relation

$$\mathcal{C}^{(\ell)}(m,n,r) = 2\mathcal{C}^{(\ell+1)}(m,n,r) + \mathcal{O}\left(m \cdot \frac{n}{2^\ell} \cdot r \right) , \tag{23}$$

where $\mathcal{C}^{(\ell)}(m,n,r)$ denotes the computational cost of the operation $R_k^{(\ell)^{-T}} \cdot F_k^{(\ell)}$ at level ℓ, $F_k^{(\ell)} \in \mathbb{R}^{n_k^{(\ell)} \times m}$, and $n_k^{(\ell)} = n^{(\ell)} = n/2^\ell$ is assumed. The first contribution in (23) stems from the recursive calls of the inverse square roots of the diagonal blocks. The second contribution is associated with the cost of the matrix product operations. Expanding (23) into a sum yields the total computational cost estimate,

$$\mathcal{C}^{(0)}(m,n,r) = \sum_{\ell=0}^{\log\left(\frac{n}{r}\right)} \mathcal{O}\left(m \cdot rn\right) = \mathcal{O}\left(m \cdot rn \log\left(n\right)\right) ,$$

where the depth of the hierarchy is set by $n = 2^L r$. Similarly, the cost of storing $R_k^{(0)^{-1}}$ and $R_k^{(0)^{-T}}$ is equal to $\mathcal{O}(rn \log n)$ in the case where the average rank per level, $r^{(\ell)}$, is of $\mathcal{O}(r)$. See [2] for further details. As noted in the introduction if the HODLR scheme is generalized to HSS the costs to apply the preconditioner reduce to about $\mathcal{O}(n)$, see [30] for further details.

Constructing the preconditioner is accomplished by performing a single pass over the hierarchy from bottom to top. At each level ℓ we compute the low-rank factorizations of the triple products

$$R_{2k-1}^{(\ell+1)^{-T}} M_k^{(\ell)} R_{2k}^{(\ell+1)^{-1}} \tag{24}$$

where $k = 1, 2, \ldots, 2^\ell$. Obtaining the low-rank factorization is performed by the following procedure:

- Capture the range of (24) in a matrix $Q_L^{(k,\ell)}$ whose columns are orthonormal,

$$\left(I - Q_L^{(k,\ell)} Q_L^{(k,\ell)^T}\right) \cdot R_{2k-1}^{(\ell+1)^{-T}} M_k^{(\ell)} R_{2k}^{(\ell+1)^{-1}} \approx \mathbf{0}\,, \quad Q_L^{(k,\ell)^T} Q_L^{(k,\ell)} = I\,.$$

- Capture the range of the transpose of (24) in a matrix $Q_R^{(k,\ell)}$ whose columns are orthonormal,

$$\left(I - Q_R^{(k,\ell)} Q_R^{(k,\ell)^T}\right) \cdot R_{2k}^{(\ell+1)^{-T}} M_k^{(\ell)^T} R_{2k-1}^{(\ell+1)^{-1}} \approx \mathbf{0}\,, \quad Q_R^{(k,\ell)^T} Q_R^{(k,\ell)} = I\,.$$

- Compute the rank $r^{(k,\ell)}$ truncated SVD of the reduced matrix,

$$Q_L^{(k,\ell)^T} \cdot R_{2k-1}^{(\ell+1)^{-T}} M_k^{(\ell)} R_{2k}^{(\ell+1)^{-1}} \cdot Q_R^{(k,\ell)} \approx U_{k,r}^{(\ell)} \Sigma_{k,r}^{(\ell)} V_{k,r}^{(\ell)}$$

- Reconstruct the WSVD left and right singular vectors matrices,

$$\mathcal{U}_{k,r}^{(\ell)} = Q_L \cdot U_{k,r}\,, \quad \mathcal{V}_{k,r}^{(\ell)} = Q_R \cdot V_{k,r}\,.$$

If the effective rank of (24) is small, e.g. $\mathcal{O}(1)$, we can capture the range matrices, $Q_L^{(k,\ell)}$ and $Q_R^{(k,\ell)}$, quickly by applying (24) and its traspose on a small set of randomized column vectors. See [21] for more details. However, if the effective rank of (24) is not small, the procedure can become costly.

3.4. Spectral Estimates and Error Propagation. Let us now focus on the spectral properties of the preconditioned submatrices, $R_k^{(\ell)^{-T}} A_k^{(\ell)} R_k^{(\ell)^{-1}}$, where $R_k^{(\ell)}$ is the square root of the principal submatrix $K_k^{(\ell)}$. The submatrix $A_k^{(\ell)}$ is given in (20). Clearly the important case is $\ell = 0$, since we are ultimately interested in preconditioning the input matrix, $A = A_1^{(0)}$.

For brevity and clarity we will abuse the notation and employ the following representations in the spirit of section 2,

$$K_k^{(\ell)} = \begin{bmatrix} K_1 & U_1 S V_2^T \\ V_2 S U_1^T & K_2 \end{bmatrix}, \quad A_k^{(\ell)} = \begin{bmatrix} A_1 & M \\ M^T & A_2 \end{bmatrix}, \quad R_i^T R_i = K_i \ (i = 1, 2)\,.$$

Note that R_i represents an approximate square root of A_i, as opposed to the exact square root that was assumed in section 2. We make the fundamental assumption that we have at our disposal spectral bounds estimates,

$$\alpha_i \le \lambda_{\min}\left(R_i^{-T} A_i R_i^{-1}\right) \le \lambda_{\max}\left(R_i^{-T} A_i R_i^{-1}\right) \le \beta_i \quad (i = 1, 2)\,. \tag{25}$$

The lower-level bounds, α_i and β_i $(i = 1, 2)$, can be obtained numerically, or possibly estimated analytically by the theory presented in this subsection. Note that in the case $\ell = L - 1$ we have $\alpha_i = 1 = \beta_i$, and in the case $\ell = L - 2$ we have from section 2 the exact bounds

$$\alpha_i = \lambda_{\min}\left(R_i^{-T} A_i R_i^{-1}\right) = 1 - \sigma_{2k-2+i,r}^{(L-1)} \in (0, 1]\,,$$

$$\beta_i = \lambda_{\max}\left(R_i^{-T} A_i R_i^{-1}\right) = 1 + \sigma_{2k-2+i,r}^{(L-1)} \in [1,2)\,.$$

The main result of the current subsection is presented in Theorem 2. The theorem provides a description of the behavior of the current-level spectral bounds,

$$\alpha \le \lambda_{\min}\left(R_k^{(\ell)^{-T}} A_k^{(\ell)} R_k^{(\ell)^{-1}}\right) \le \lambda_{\max}\left(R_k^{(\ell)^{-T}} A_k^{(\ell)} R_k^{(\ell)^{-1}}\right) \le \beta\,, \tag{26}$$

as a function of the lower-level bounds (25) and the rank of the off-diagonal blocks, $r = r_k^{(\ell)}$. The definition of the bounds α and β (26) is based on variational formulation and provided in Lemma 2. The analysis requires sufficient (but not necessary) conditions on the given lower-level bounds, α_i and β_i $(i = 1, 2)$. We show that the proposed HODLR preconditioning scheme, essentially, maps both the r largest and the r smallest eigenvalues to a closed segment containing 1. When this segment is small, the preconditioner retains optimality or near optimality. We also show that the sensitivity of the spectral bounds to the inaccuracies $K_i \ne A_i$ $(i = 1, 2)$ is governed by the *Cauchy-Bunyakowski-Schwarz* (CBS) constant [3, 4].

THEOREM 2. *Let*

$$A = \begin{bmatrix} A_1 & M \\ M^T & A_2 \end{bmatrix}, \quad K = \begin{bmatrix} K_1 & U_1 S V_2^T \\ V_2 S U_1^T & K_2 \end{bmatrix},$$

be symmetric matrices of the same dimensions and partition where A is SPD. Assume that the off-diagonal triple product approximation $U_1 S V_2^T$ satisfy

$$U_1 = R_1^T \mathcal{U}_r\,, \quad S = \Sigma_r\,, \quad V_2 = R_2^T \mathcal{V}_r\,,$$

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

$$R_1^{-T} M R_2^{-1} = \mathcal{U}\Sigma\mathcal{V}^T\,, \quad \Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_{\min\{n_1,n_2\}})\,,$$

and $R_i^T R_i = K_i$ $(i = 1, 2)$.

Assuming there exist real positive constants,

$$0 < \alpha_1, \alpha_2 \le 1 \le \beta_1, \beta_2\,, \tag{27}$$

such that

$$0 < \alpha_i x_i^T K_i x_i \le x_i^T A_i x_i \le \beta_i x_i^T K_i x_i \quad \forall x_i \in \mathbb{R}^{n_i}\,, \tag{28}$$

we have the following spectral estimates:

1. *If $\sigma_1 < \sqrt{\alpha_1 \alpha_2}$ then*

$$\alpha = \min\left\{\alpha_{1,2}^{\text{avg}} - \sqrt{\sigma_{r+1}^2 + \left(\alpha_{1,2}^{\text{dif}}\right)^2}\quad,\quad \frac{\alpha_{1,2}^{\text{avg}} + \sigma_1}{1+\sigma_1} - \delta_\alpha^{1,2}\right\}, \tag{29}$$

where

$$\alpha_{1,2}^{\text{avg}} = \frac{\alpha_1 + \alpha_2}{2}\,, \quad \alpha_{1,2}^{\text{dif}} = \frac{\alpha_1 - \alpha_2}{2}\,, \quad (\alpha = \alpha \text{ or } \beta)\,,$$

and

$$\delta_\alpha^{1,2} = \sqrt{\frac{1}{4}\left|\frac{\alpha_{1,2}^{\text{avg}} - \sigma_1}{1-\sigma_1} - \frac{\alpha_{1,2}^{\text{avg}} + \sigma_1}{1+\sigma_1}\right|^2 + \frac{\left(\alpha_{1,2}^{\text{dif}}\right)^2}{1-\sigma_1^2}} - \frac{1}{2}\left|\frac{\alpha_{1,2}^{\text{avg}} - \sigma_1}{1-\sigma_1} - \frac{\alpha_{1,2}^{\text{avg}} + \sigma_1}{1+\sigma_1}\right|,$$

is a positive lower spectral bound of the preconditioned system,

$$0 < \alpha \le \lambda\left(R^{-T} A R^{-1}\right), \quad K = R^T R\,.$$

2. *If* $\sigma_1 < 1$ *then*

$$(30) \qquad \beta = \max\left\{ \beta_{1,2}^{\text{avg}} + \sqrt{\sigma_{r+1}^2 + \left(\beta_{1,2}^{\text{dif}}\right)^2} \quad , \quad \frac{\beta_{1,2}^{\text{avg}} - \sigma_1}{1-\sigma_1} + \delta_\beta^{1,2} \right\},$$

where

$$\beta_{1,2}^{\text{avg}} = \frac{\beta_1 + \beta_2}{2}, \quad \beta_{1,2}^{\text{dif}} = \frac{\beta_1 - \beta_2}{2}, \quad (\beta = \alpha \ \text{ or } \ \beta),$$

and

$$\delta_\beta^{1,2} = \sqrt{\frac{1}{4}\left|\frac{\beta_{1,2}^{\text{avg}} - \sigma_1}{1-\sigma_1} - \frac{\beta_{1,2}^{\text{avg}} + \sigma_1}{1+\sigma_1}\right|^2 + \frac{\left(\beta_{1,2}^{\text{dif}}\right)^2}{1-\sigma_1^2}} - \frac{1}{2}\left|\frac{\beta_{1,2}^{\text{avg}} - \sigma_1}{1-\sigma_1} - \frac{\beta_{1,2}^{\text{avg}} + \sigma_1}{1+\sigma_1}\right|.$$

is an upper spectral bound of the preconditioned system,

$$0 < \lambda\left(R^{-T} A R^{-1}\right) \leq \beta, \quad K = R^T R.$$

Remarks. *The justification for (27) is a consequence of Theorem 2, which shows that* α *(29) and* β *(30) are monotonically non-increasing and non-decreasing as functions of the level, respectively. This observation is also supported by numerical evidence in section 4. If* $\sqrt{\alpha_1\alpha_2} \leq \sigma_1 < 1$*, then the preconditioned system remains SPD. However, the theory presented here can not predict the positive value of the lower spectral bound,* α *(29).*

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, σ_1, and the second is a function of the truncation error, σ_{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 Figure 3.

The last observation as displayed in Figure 3 indicates that the value of σ_1 is central to the estimation of the spectral bounds, and effectively dominates the condition number of the preconditioned system. In this sense σ_1 reflects the sensitivity of the condition number of $R_k^{(\ell)^{-T}} A_k^{(\ell)} R_k^{(\ell)^{-1}}$ to the lower level inaccuracies. It can be shown that σ_1 is the so-called *Cauchy-Bunyakowski-Schwarz* (CBS) constants of the matrix K, which is defined by

$$(31) \qquad \sigma_1 = \sup_{x_1, x_2 \neq 0} \frac{x_1^T M x_2}{\sqrt{x_1^T K_1 x_1}\sqrt{x_2^T K_2 x_2}} \geq 0.$$

Definition (31) originated from the theory of *Algebraic Multilevel Iterations Methods* [3, 4], and 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, σ_1 represents the local contribution of the upper level to the overall condition number. Using (31) with assumption (27) leads to the following relation

$$\frac{1}{\sqrt{\beta_1\beta_2}} \leq \frac{\sigma_1}{\sigma_1^{\text{exact}}} \leq \frac{1}{\sqrt{\alpha_1\alpha_2}},$$

where σ_1^{exact} is the corresponding CBS constant of A. The important conclusion here is that σ_1 and σ_1^{exact} are correlated where σ_1^{exact} is intrinsically predetermined by the

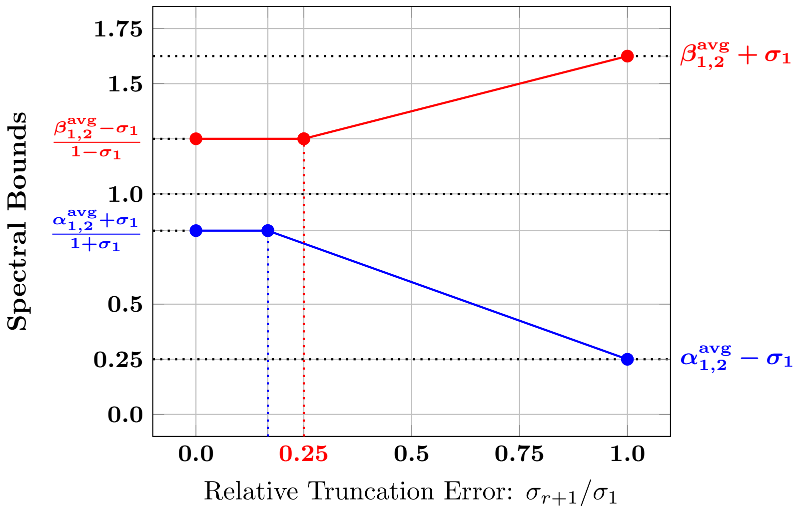

Fig. 3: **Spectral Bounds.** A typical behavior of the spectral bounds displayed for the case $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$. The lower bound α (29) vs. σ_{r+1}/σ_1 is plotted in blue, and the upper bound β (30) vs. σ_{r+1}/σ_1 is plotted in red.

given matrix, A, and the chosen partition. If K is close to A then we can expect σ_1 to be close to σ_1^{exact}, 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 $\beta_{1,2}^{\text{avg}}$,

$$\left[\alpha_{1,2}^{\text{avg}} - \sqrt{\sigma_1^2 + \left(\alpha_{1,2}^{\text{dif}}\right)^2} \ , \ \alpha_{1,2}^{\text{avg}} + \sqrt{\sigma_1^2 + \left(\alpha_{1,2}^{\text{dif}}\right)^2}\right] ,$$

$$\left[\beta_{1,2}^{\text{avg}} - \sqrt{\sigma_1^2 + \left(\beta_{1,2}^{\text{dif}}\right)^2} \ , \ \beta_{1,2}^{\text{avg}} + \sqrt{\sigma_1^2 + \left(\beta_{1,2}^{\text{dif}}\right)^2}\right] .$$

The multilevel Weighted HODLR preconditioning scheme does the same, but also maps the largest and smallest eigenvalues to the segments

$$\left[\frac{\sigma_1 + \alpha_{1,2}^{\text{avg}}}{1+\sigma_1} - \delta_\alpha^{1,2} \ , \ \frac{\sigma_1 - \alpha_{1,2}^{\text{avg}}}{1-\sigma_1} + \delta_\alpha^{1,2}\right] , \tag{32}$$

$$\left[\frac{\beta_{1,2}^{\text{avg}} + \sigma_1}{1+\sigma_1} - \delta_\beta^{1,2} \ , \ \frac{\beta_{1,2}^{\text{avg}} - \sigma_1}{1-\sigma_1} + \delta_\beta^{1,2}\right] , \tag{33}$$

respectively. Thus, assuming the segments (32) and (33) are small, a significant improvement in the condition number as well as the clustering of the spectrum of the original preconditioned system is expected. An illustration is given in Figure 4. The figure is similar to Figure 1 where the main difference is that the weighted HODLR preconditioning scheme now maps the extreme eigenvalues to an interval containing 1 and not exactly to 1.

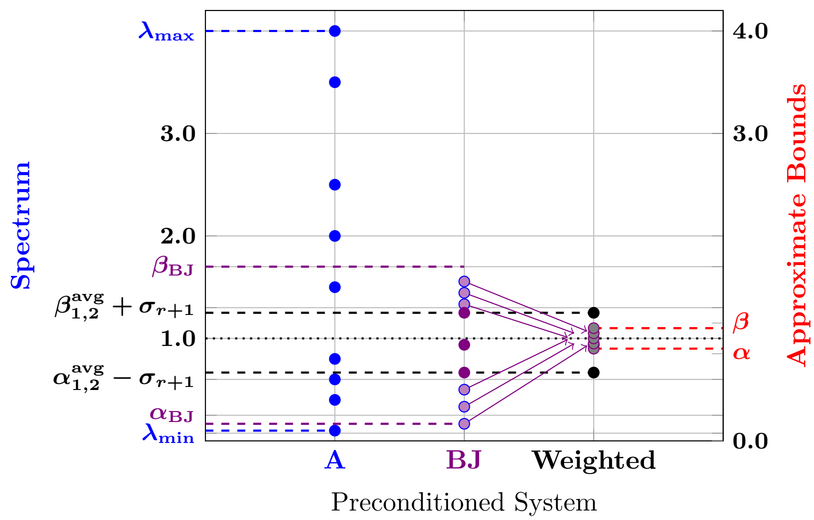

Fig. 4: **Spectrum Clustering for the Multilevel Weighted HODLR Preconditioning Scheme.** The spectrum of some SPD matrix A and the transformation it goes after preconditioning by block Jacobi (BJ) and the multilevel weighted HODLR preconditioning scheme are displayed. The spectra are ordered from the left to the right starting from A, followed by Block Jacobi (BJ) and ends up with the multilevel weighted HODLR scheme. The spectral bounds α (29) and β (30) are marked 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 on the left y-axis.

3.5. Proof of Theorem 2. The proof of Theorem 2 is based on Lemma 2 which provides the definition of the bounds α (29) and β (30), and on the technical result presented in Lemma 3.

LEMMA 2. *Let*

$$A = \begin{bmatrix} A_1 & M \\ M^T & A_2 \end{bmatrix}, \quad K = \begin{bmatrix} K_1 & U_1 S V_2^T \\ V_2 S U_1^T & K_2 \end{bmatrix},$$

and the associated lower-level bounds

$$0 < \alpha_1, \alpha_2 \leq 1 \leq \beta_1, \beta_2 , \tag{34}$$

satisfy the assumptions of Theorem 2.

Let us define

$$\underline{K} = \begin{bmatrix} \alpha_1 K_1 & M \\ M^T & \alpha_2 K_2 \end{bmatrix}, \quad \overline{K} = \begin{bmatrix} \beta_1 K_1 & M \\ M^T & \beta_2 K_2 \end{bmatrix}. \tag{35}$$

Then we have:

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

$$\sigma_1 < \sqrt{\alpha_1 \alpha_2}, \quad \sigma_1 < \sqrt{\beta_1 \beta_2},$$

respectively, where σ_1 is the largest singular of $R_1^{-T} M R_2^{-1}$.

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

$$\alpha = \min_{x\neq 0} \frac{x^T \underline{K} x}{x^T K x} \leq \lambda_{\min}(\widehat{B}^T A \widehat{B}) \leq \lambda_{\max}(\widehat{B}^T A \widehat{B}) \leq \max_{x\neq 0} \frac{x^T \overline{K} x}{x^T K x} = \beta \,. \tag{36}$$

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

$$H = \begin{bmatrix} \delta_1 K_1 & M \\ M^T & \delta_2 K_2 \end{bmatrix}, \quad \delta_1, \delta_2 > 0 \,.$$

Let us apply the following two-sided transformations

$$\widehat{H} = \begin{bmatrix} \frac{1}{\sqrt{\delta_1}} R_1^{-T} & 0 \\ 0 & \frac{1}{\sqrt{\delta_2}} R_2^{-T} \end{bmatrix} H \begin{bmatrix} \frac{1}{\sqrt{\delta_1}} R_1^{-1} & 0 \\ 0 & \frac{1}{\sqrt{\delta_2}} R_2^{-1} \end{bmatrix} = \begin{bmatrix} I & \frac{1}{\sqrt{\delta_1 \delta_2}} \mathcal{M} \\ \frac{1}{\sqrt{\delta_1 \delta_2}} \mathcal{M}^T & I \end{bmatrix},$$

where $\mathcal{M} = R_1^{-T} M R_2^{-1}$. The matrix $\widehat{H}$ is SPD iff H is SPD as well. Thus, by Lemma 1 the matrix H is SPD iff $1 - \sigma_1/\sqrt{\delta_1 \delta_2} > 0$, and the conditions ensuring $\underline{K}$, K, and $\overline{K}$ are SPD immediately follow.

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

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

The Lagrangian stationary points of each generalized Rayleigh quotient in the inequalities above constitute the spectrum of each preconditioned system. Thus, the proof is complete. ∎

LEMMA 3. *Let $H = \begin{bmatrix} \mathcal{D}^{(1)} & \mathcal{D}^{(2)} \\ \mathcal{D}^{(2)} & \mathcal{D}^{(3)} \end{bmatrix} \in \mathbb{R}^{2r \times 2r}$, where $\mathcal{D}^{(i)}$ $(i = 1, 2, 3)$ are diagonal matrices,*

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

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

$$\mathrm{spec}(H) = \left\{\lambda_j^-\right\}_{j=1}^m \cup \left\{\lambda_j^+\right\}_{j=1}^m, \quad \lambda_j^{\pm} = \frac{d_j^{(1)} + d_j^{(3)}}{2} \pm \sqrt{\left(\frac{d_j^{(1)} - d_j^{(3)}}{2}\right)^2 + \left(d_j^{(2)}\right)^2},$$

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

Proof. of Lemma 3.
From the given structure of H it is clear that $\lambda \in \mathbb{R}$ is an eigenvalue of H 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

$$\lambda = \lambda_j^{\pm} = \frac{d_j^{(1)} + d_j^{(3)}}{2} \pm \sqrt{\left(\frac{d_j^{(1)} - d_j^{(3)}}{2}\right)^2 + \left(d_j^{(2)}\right)^2},$$

and the proof is complete. ∎

Proof. of Theorem 2.
Considering the conditions of Theorem 2 we have by Lemma 2 that the spectral bounds, α and β, satisfying

$$0 < \alpha = \min_{x\neq 0} \frac{x^T \underline{K} x}{x^T K x} \leq \lambda_{\min}(R^{-T}AR^{-1}) \leq \lambda_{\max}(R^{-T}AR^{-1}) \leq \max_{x\neq 0} \frac{x^T \overline{K} x}{x^T K x} = \beta\,,$$

exist where $\underline{K}$ and $\overline{K}$ are defined in Lemma 2.

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

$$Q(x) = \frac{x^T H x}{x^T K x}\,, \quad H = \begin{bmatrix} \delta_1 K_1 & M \\ M^T & \delta_2 K_2 \end{bmatrix}\,,$$

whose range is strictly positive. Thus, $Q(x)$ represents either $x^T\underline{K}x/x^TKx$ or $x^T\overline{K}x$. We apply the change of variables, $x = \left[\begin{smallmatrix} R_1^{-T} & 0 \\ 0 & R_2^{-1} \end{smallmatrix}\right]\xi$, and obtain the following equivalent representation

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

where $\mathcal{M} = R_1^{-T} M R_2^{-1}$ and $\mathcal{M}_r = \mathcal{U}_r \Sigma_r \mathcal{V}_r^T$ is the r-rank weighted SVD approximation of $\mathcal{M}$.

Let w_i denote the i-th column of the orthogonal matrix $\mathcal{W}$ (13) as defined in Lemma 1. Then we have:

1. $\widehat{K} w_i = (1+\sigma_1) w_i$, $i = 1, 2, \ldots, r$.
2. $\widehat{K} w_{n_1+i} = (1-\sigma_1) w_{n_1+i}$, $i = 1, 2, \ldots, r$.
3. $\widehat{K} w_j = w_j$, $j \neq 1, \ldots, r, n_1+1, \ldots, n_1+r$.

and similarly for $\widehat{H}$, it can be verified that:

1. $\widehat{H} w_i = (\delta_{1,2}^{\text{avg}} + \sigma_1) w_i + \delta_{1,2}^{\text{dif}} w_{n_1+i}$, $i = 1, 2, \ldots, \min\{n_1, n_2\}$.
2. $\widehat{H} w_{n_1+i} = (\delta_{1,2}^{\text{avg}} - \sigma_1) w_{n_1+i} + \delta_{1,2}^{\text{dif}} w_i$, $i = 1, 2, \ldots, \min\{n_1, n_2\}$.
3. $\widehat{H} w_j = w_j$, $j \neq 1, \ldots, \min\{n_1, n_2\}, n_1+1, \ldots, n_1 + \min\{n_1, n_2\}$.

where $\delta_{1,2}^{\text{avg}} = (\delta_1+\delta_2)/2$ and $\delta_{1,2}^{\text{dif}} = (\delta_1-\delta_2)/2$. Clearly, both $\widehat{K}$ and $\widehat{H}$ 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:

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

and

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

By our results so far, if $x = \xi \in Z^\perp$ then $Q(x) = \xi^T \widehat{H} \xi / \xi^T \xi$. 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 $\mathcal{W}$ (13). Then, for any $\xi \in Z^\perp$ we obtain

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

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

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

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

Now,by Lemma 3, we obtain that the spectrum of $\widehat{H}_{Z^\perp}$ contains the sets

$$\left\{\delta_{1,2}^{\text{avg}} + \sqrt{\sigma_{r+1}^2 + \left(\delta_{1,2}^{\text{dif}}\right)^2}, \dots, \delta_{1,2}^{\text{avg}} + \sqrt{\sigma_{\min\{n_1,n_2\}}^2 + \left(\delta_{1,2}^{\text{dif}}\right)^2}, \right\},$$

$$\left\{\delta_{1,2}^{\text{avg}} - \sqrt{\sigma_{\min\{n_1,n_2\}}^2 + \left(\delta_{1,2}^{\text{dif}}\right)^2}, \dots, \delta_{1,2}^{\text{avg}} - \sqrt{\sigma_{r+1}^2 + \left(\delta_{1,2}^{\text{dif}}\right)^2}\right\}.$$

Hence, we conclude that

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

$$\max_{\xi \in Z^\perp \setminus \{0\}} Q(x) = \delta_{1,2}^{\text{avg}} + \sqrt{\sigma_{r+1}^2 + \left(\delta_{1,2}^{\text{dif}}\right)^2}.$$

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 $\mathcal{W}$ (13). Then, for any $\xi \in Z$ we obtain

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

$$\mathcal{D}_Z^{(1)} = \text{diag}\left(\frac{\delta_{1,2}^{\text{avg}} + \sigma_1}{1+\sigma_1}, \dots, \frac{\delta_{1,2}^{\text{avg}} + \sigma_r}{1+\sigma_r}\right),$$

$$\mathcal{D}_Z^{(2)} = \text{diag}\left(\frac{\delta_{1,2}^{\text{dif}}}{\sqrt{1-\sigma_1^2}}, \dots, \frac{\delta_{1,2}^{\text{dif}}}{\sqrt{1-\sigma_r^2}}\right),$$

$$\mathcal{D}_Z^{(3)} = \text{diag}\left(\frac{\delta_{1,2}^{\text{avg}} - \sigma_1}{1-\sigma_1}, \dots, \frac{\delta_{1,2}^{\text{avg}} - \sigma_r}{1-\sigma_r}\right).$$

Applying once more the outcome of Lemma 3 we have that the spectrum of $\widehat{H}_Z$ is composed of the following values

$$\frac{1}{2}\left(\frac{\delta_{1,2}^{\text{avg}} + \sigma_i}{1+\sigma_i} + \frac{\delta_{1,2}^{\text{avg}} - \sigma_i}{1-\sigma_i}\right) \pm \sqrt{\frac{1}{4}\left(\frac{\delta_{1,2}^{\text{avg}} + \sigma_i}{1+\sigma_i} + \frac{\delta_{1,2}^{\text{avg}} - \sigma_i}{1-\sigma_i}\right)^2 + \frac{\left(\delta_{1,2}^{\text{dif}}\right)^2}{1-\sigma_i^2}},$$

where $i = 1, 2, \dots, r$ and the proof is complete. ◻

4. Numerical Study.

This section contains the experimental part of this work. The main goal is to demonstrate the effect of different low-rank approximations (18) for the off-diagonal blocks on the preconditioned system using HODLR. We perform a comparative study and consider the following low-rank techniques:

- **R-HODLR:** the off-diagonal low-rank factorizations are obtained in the standard or regular approach using truncated SVD.
- **C-HODLR:** the off-diagonal low-rank factorizations are obtained using truncated SVD with additional imposed constraints as described in [6].
- **W-HODLR:** the off-diagonal low-rank factorizations are obtained using the weighted HODLR preconditioning scheme for the multi-level case. The construction and application of the scheme follows the outlined procedure in subsection 3.3.

Employing SVD is done for convenience and uniformity of the comparison, and can be replaced, in practice, by other more efficient low-rank approximation techniques.

Subsection 4.1 describes the computational setting, and presents a pair of severely ill-conditioned sparse systems which have been used in the numerical simulations. Another simplified numerical example along a more detailed analysis can be found in [1]. In subsection 4.2 we describe the numerical results of the PCG approximation using HODLR preconditioning. The results indicate that the weighted low-rank factorization scheme proves to be superior to other standard techniques.

4.1. Sparse Matrices and Computational Setting.

In the presented numerical study we have explored and analyzed the PCG solution for the following sparse matrices, which have been picked from the SuiteSparse matrix collection [10].

- **bcsstk16:** $4,884 \times 4,884$, SPD , spectral condition number $\approx 4.94 \cdot 10^9$.
- **bcsstk15:** $3,948 \times 3,948$, SPD , spectral condition number $\approx 6.53 \cdot 10^9$.

For constructing the HODLR preconditioning schemes we interpret each matrix as a discrete graph and apply a balanced partitioning using successive bisections. We employ SCOTCH [9] for each bisection dividing a given vertex set into two distinct sets of approximately equal size whose cut is minimal, i.e., the number of edges running between the separated subsets is as small as possible. The process starts with the entire set of vertices, and then applied recursively on each separated subset until reaching the predetermined bottom level of the hierarchy, L.

Construction of the preconditioning schemes and the iterative solution of the preconditioned system has been implemented with a Fortran90 code. In all the simulations we have employed the following selections:

- $L = \lceil \log_2 (n/100) \rceil$ for an $n \times n$ matrix as the lowest level of the hierarchy, which forces the size of the smallest blocks in the partition under 100.
- Constant off-diagonal block ranks over all levels of the hierarchy with the following $\mathcal{O}(1)$ values:

$$r_k^{(\ell)} = 0, 5, 10, 15, 20, 25\,, \quad \ell = 0, 1, \ldots, L\,, \quad k = 1, \ldots, 2^\ell\,. \tag{37}$$

Note that $r_k^{(\ell)} \equiv 0$ reduces the preconditioning scheme to *block Jacobi* (BJ), regardless of the specific low-rank factorization technique.

The construction of the low-rank factorizations (24) follows the path outlined in subsection 3.3. We have produced fast low-rank factorizations by first removing all the zero rows and columns of the sparse block $M_k^{(\ell)}$ (20), and then computing the low-rank factorization (24) on the reduced block. For the sparse case, this procedure is, typically, equivalent in terms of complexity to the randomized technique [21].

4.2. Numerical Results and Analysis. This subsection contains the numerical results of PCG solution for the chosen sparse systems, using R-HODLR, C-HODLR and W-HODLR preconditioning schemes. For both matrices, bcsstk16 and bcsstk15, we have set the right-hand side to $b = 1$, and the iterative approximation was stopped 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 $x_{(i)}$ denotes the approximate solution at step i. The results indicate that in all the test cases, the W-HODLR scheme outperforms the other techniques, and retain good properties even when low accuracy for the off-diagonal blocks approximations is employed.

Figure 5 contains plots of the PCG convergence history profiles for bcsstk16. All plots in this case indicate that increasing the constant rank (37), improves the approximation quality, and achieves faster convergence rate. It is also evident that R-HODLR and C-HODLR achieve similar convergence with the same memory resources, while W-HODLR converges faster with the same memory resources. Figure 6 contains plots of the PCG convergence history profiles for bcsstk15. The results show that R-HODLR and C-HODLR fail to converge in 1000 PCG iterations. In fact, setting constant rank 0, i.e., Block Jacobi, performs better then using these schemes with a constant rank greater than zero. This occurs because the use of naive approximations for the off-diagonal blocks makes the problem even more ill-conditioned. The W-HODLR scheme, however, converges with excellent convergence rates, where the convergence rate improves when the constant rank (37) is increased.

5. Summary and Future Work. In this work we have addressed the problem of choosing low-rank factorizations in fast hierarchical algorithms for preconditioning SPD matrices.

We have presented a mathematical analysis for obtaining low-rank factorizations, that minimize the spectral condition number of the preconditioned system for the 1-level (2×2) case. The key idea was to properly reweight the blocks prior the low-rank factorization, which leads to a minimum spectral condition number.

The presented theory has been extended to HODLR preconditioning schemes, including analysis of the spectral properties and bounds that take into account the error propagation through the levels of the hierarchy.

The numerical experiments indicate, that employing proper reweighting for the off-diagonal blocks prior to low-rank compression, leads to a HODLR preconditioning scheme that requires far less computational resources for the same quality of performance of convergence than using the other low-rank compression techniques.

As noted in the introduction a major goal of this work is to provide an analysis of optimal choice of low-rank approximations for a simple case; i.e., HODLR, which could lead to an extended analysis for the strong hierarchical case. This point will be explored and pursued in a future study.

Acknowledgments. We would like to thank the anonymous referees for their valuable remarks, questions and comments that enabled us to substantially improve this paper.

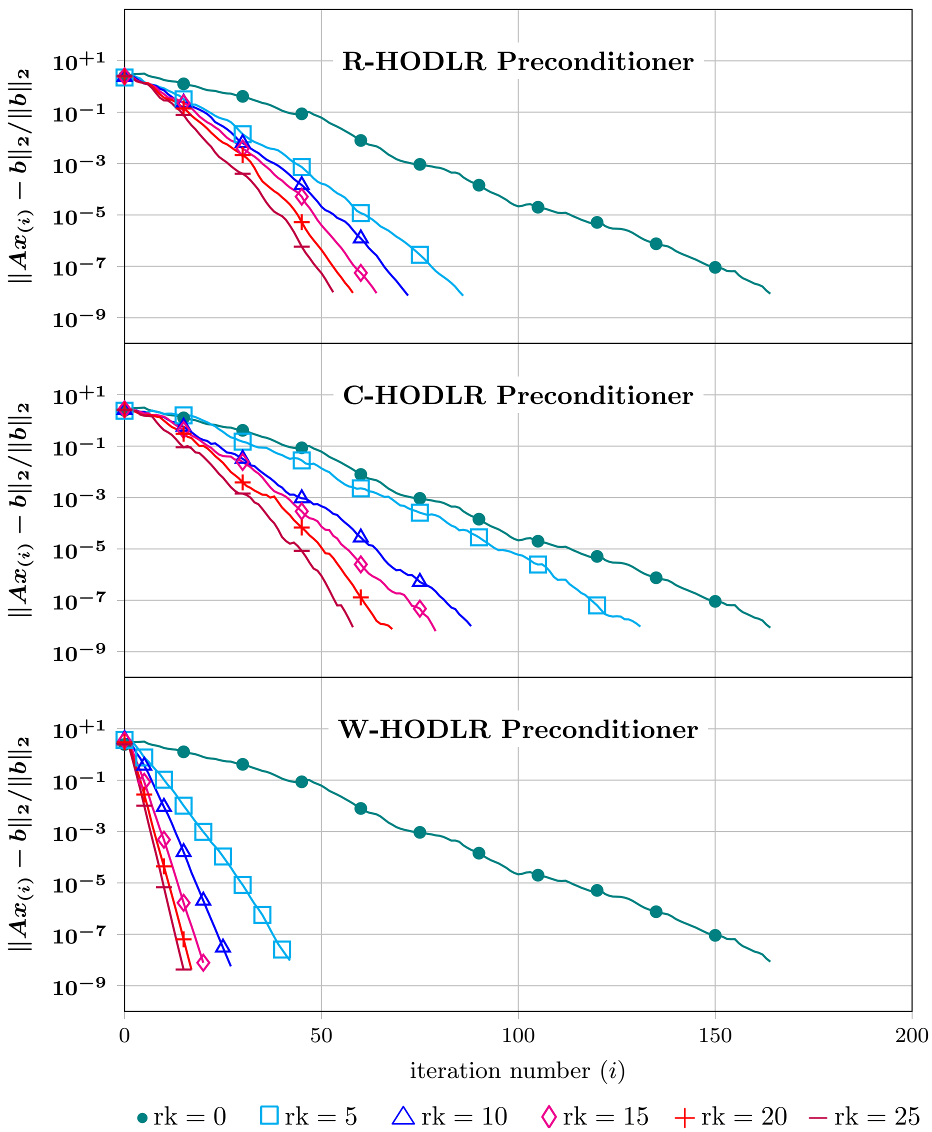

Fig. 5: **PCG Convergence History: 'bcsstk16'.** Three plots showing PCG convergence history profiles, i.e., the values $\|Ax_{(i)} - b\|_2/\|b\|_2$ as a function of the iteration number i, for each preconditioning scheme. Each plot displays various profiles, where each profile corresponds to a different constant rank value (37) of the approximations for the off-diagonals blocks by low-rank factorizations.

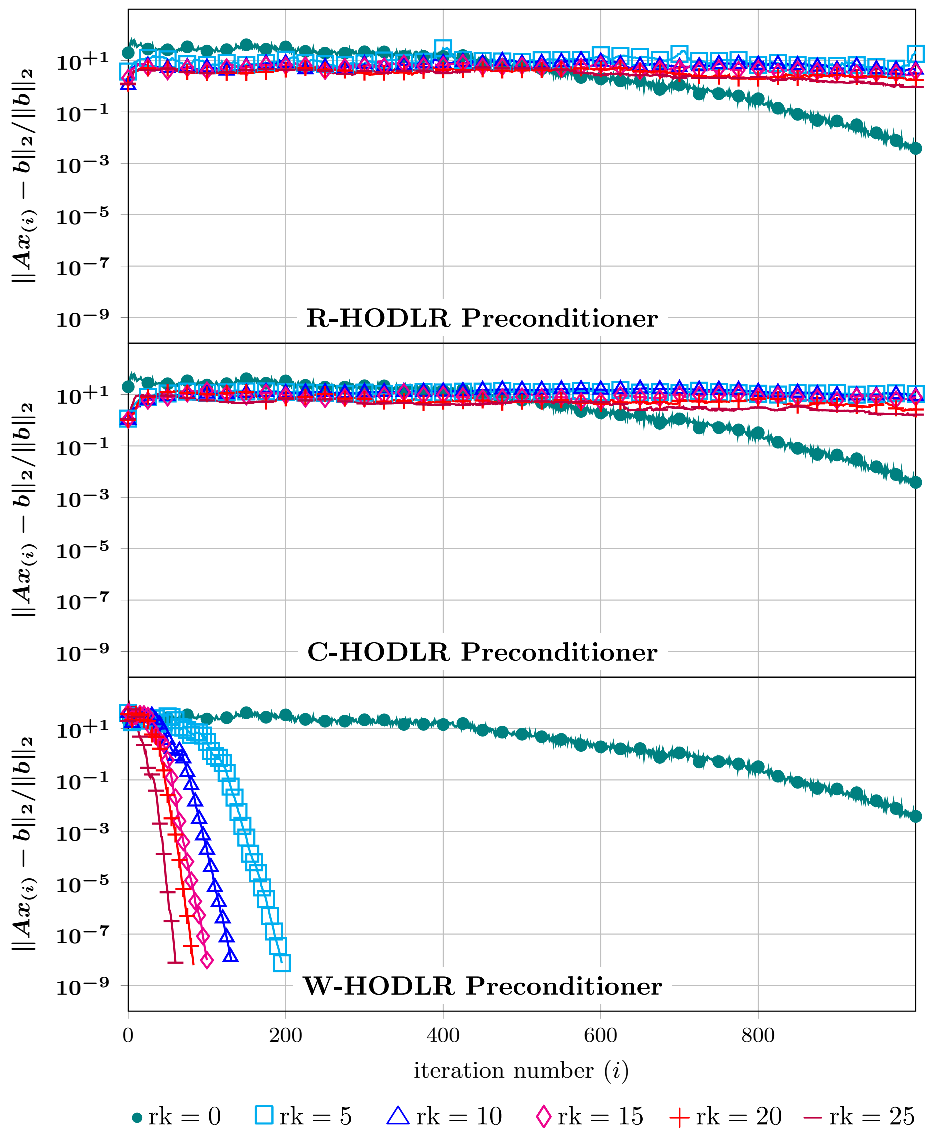

Fig. 6: **PCG Convergence History: 'bcsstk15'.** Three plots showing PCG convergence history profiles, i.e., the values $\|Ax_{(i)} - b\|_2/\|b\|_2$ as a function of the iteration number i, for each preconditioning scheme. Each plot displays various profiles, where each profile corresponds to a different constant rank value (37) of the approximations for the off-diagonals blocks by low-rank factorizations.

REFERENCES

[1] E. Agullo, E. Darve, L. Giraud, and Y. Harness, *Nearly optimal fast preconditioning of symmetric positive definite matrices*, Research Report RR-8984, Inria Bordeaux Sud-Ouest, Nov. 2016.

[2] A. Aminfar, S. Ambikasaran, and E. Darve, *A fast block low-rank dense solver with applications to finite-element matrices*, Journal of Computational Physics, 304 (2016), pp. 170 – 188.

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

[4] O. Axelsson and P. S. Vassilevski, *Algebraic Multilevel Preconditioning Methods, II*, SIAM Journal on Numerical Analysis, 27 (1990), pp. 1569–1590.

[5] M. Bebendorf, *Hierarchical Matrices: A Means to Efficiently Solve Elliptic Boundary Value Problems*, Springer Berlin Heidelberg, 2008.

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

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

[8] S. Chandrasekaran, M. Gu, and T. Pals, *A Fast ULV Decomposition Solver for Hierarchically Semiseparable Representations*, SIAM Journal on Matrix Analysis and Applications, 28 (2006), pp. 603–622.

[9] C. Chevalier and F. Pellegrini, *PT-Scotch: A tool for efficient parallel graph ordering*, Parallel Computing, 34 (2008), pp. 318 – 331. Parallel Matrix Algorithms and Applications.

[10] T. A. Davis and Y. Hu, *The University of Florida Sparse Matrix Collection*, ACM Trans. Math. Softw., 38 (2011), pp. 1:1–1:25.

[11] C. Eckart and G. Young, *The approximation of one matrix by another of lower rank*, Psychometrika, 1 (1936), pp. 211–218.

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

[13] A. Frieze, R. Kannan, and S. Vempala, *Fast Monte-carlo Algorithms for Finding Low-rank Approximations*, J. ACM, 51 (2004), pp. 1025–1041.

[14] G. H. Golub and C. F. V. Loan, *Matrix computations, Forth Edition*, Johns Hopkins University Press, 2013.

[15] L. Grasedyck and W. Hackbusch, *Construction and Arithmetics of H-Matrices*, Springer Journal of Computing, 70 (2003), pp. 295–334.

[16] M. Gu and S. C. Eisenstat, *Efficient Algorithms for Computing a Strong Rank-Revealing QR Factorization*, SIAM Journal on Scientific Computing, 17 (1996), pp. 848–869.

[17] W. Hackbusch, *A Sparse Matrix Arithmetic Based on $\mathcal{H}$-Matrices. Part I: Introduction to $\mathcal{H}$-Matrices*, Computing, 62 (1999), pp. 89–108.

[18] W. Hackbusch and B. N. Khoromskij, *A Sparse $\mathcal{H}$-Matrix Arithmetic. Part II: Application to Multi-Dimensional Problems*, Computing, 64 (2000), pp. 21–47.

[19] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, Journal of Research of the National Bureau of Standards, 49 (1952).

[20] E. Liberty, F. Woolfe, P.-G. Martinsson, V. Rokhlin, and M. Tygert, *Randomized algorithms for the low-rank approximation of matrices*, Proceedings of the National Academy of Sciences, 104 (2007), pp. 20167–20172.

[21] P. G. Martinsson, *A Fast Randomized Algorithm for Computing a Hierarchically Semiseparable Representation of a Matrix*, SIAM Journal on Matrix Analysis and Applications, 32 (2011), pp. 1251–1274.

[22] L. Miranian and M. Gu, *Strong rank revealing LU factorizations*, Linear Algebra and its Applications, 367 (2003), pp. 1 – 16.

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

[24] S. Rjasanow, *Adaptive cross approximation of dense matrices*, in Int. Association Boundary Element Methods Conf., IABEM, 2002, pp. 28–30.

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

[26] A. J. Wathen, *Preconditioning*, Acta Numerica, 24 (2015), pp. 329–376.

[27] F. Woolfe, E. Liberty, V. Rokhlin, and M. Tygert, *A fast randomized algorithm for the approximation of matrices*, Applied and Computational Harmonic Analysis, 25 (2008), pp. 335 – 366.

[28] J. Xia, S. Chandrasekaran, M. Gu, and X. S. Li, *Fast algorithms for hierarchically semisep-*

arable matrices, Numerical Linear Algebra with Applications, 17 (2010), pp. 953–976.

[29] J. Xia and M. Gu, *Robust Approximate Cholesky Factorization of Rank-Structured Symmetric Positive Definite Matrices*, SIAM Journal on Matrix Analysis and Applications, 31 (2010), pp. 2899–2920.

[30] J. Xia and Z. Xin, *Effective and Robust Preconditioning of General SPD Matrices via Structured Incomplete Factorization*, SIAM Journal on Matrix Analysis and Applications, 38 (2017), pp. 1298–1322.

[31] K. Yang, H. Pouransari, and E. Darve, *Sparse Hierarchical Solvers with Guaranteed Convergence*, ArXiv e-prints, (2016), arXiv:1611.03189.