NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2011; **00**:1–0 *Prepared using* nlaauth.cls *[Version: 2002/09/18 v1.02]*

Effectiveness and robustness revisited for an SIF preconditioning technique

Zixing Xin[1], Jianlin Xia*[1†], Stephen Cauley[2], and Venkataramanan Balakrishnan[3]

[1] *Department of Mathematics, Purdue University, West Lafayette, IN 47907, U.S.A.*
[2] *Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard University, Charlestown, MA 02129, U.S.A.*
[3] *Case School of Engineering, Case Western Reserve University, Cleveland, OH 44106, U.S.A.*

SUMMARY

In this work, we provide new analysis for a preconditioning technique called structured incomplete factorization (SIF) for symmetric positive definite (SPD) matrices. In this technique, a scaling-and-compression strategy is applied to construct SIF preconditioners, where off-diagonal blocks of the original matrix are first scaled and then approximated by low-rank forms. Some spectral behaviors after applying the preconditioner are shown. The effectiveness is confirmed with the aid of a type of 2D and 3D discretized model problems. We further show that previous studies on the robustness are too conservative. In fact, the practical multilevel version of the preconditioner has a robustness enhancement effect, and is unconditionally robust (or breakdown free) for the model problems regardless of the compression accuracy for the scaled off-diagonal blocks. The studies give new insights into the SIF preconditioning technique and confirm that it is an effective and reliable way for designing structured preconditioners. The studies also provide useful tools for analyzing other structured preconditioners. Various spectral analysis results can be used to characterize other structured algorithms and study more general problems. Copyright © 2011 John Wiley & Sons, Ltd.

KEY WORDS: SIF preconditioning; scaling-and-compression strategy; effectiveness; robustness; spectral analysis

1. INTRODUCTION

Designing effective and robust preconditioners is typically the key issue in iterative solutions of large symmetric positive definite (SPD) linear systems. An effective preconditioner can significantly reduce the number of iterations and thus the computational cost. In the meantime, it is often preferable to have robust preconditioners that remain positive definite. Various types of robust preconditioners have been designed [2, 3, 4, 11, 18, 19], where some stabilization strategies are often used.

In recent years, low-rank compression methods have often been used to design effective preconditioners, and are typically based on the low-rank approximation of certain dense off-diagonal

*Correspondence to: Jianlin Xia, Department of Mathematics, Purdue University, West Lafayette, IN 47907, U.S.A.
†Email: xiaj@math.purdue.edu

The research of Jianlin Xia was supported in part by an NSF grant DMS-1819166.

blocks. The resulting structured approximations are used as preconditioners. Such structured preconditioners can be quickly applied and it is convenient to control the accuracy of how they approximate the original matrix. On the other hand, it is usually nontrivial to analyze the effectiveness.

Recently, a robust preconditioning technique called structured incomplete factorization (SIF) is proposed in [24] for SPD matrices. The technique relies on a scaling-and-compression strategy reformulated from an earlier paper [23]. In the strategy, off-diagonal blocks are not directly compressed. Instead, they are first scaled by the inverses of the Cholesky factors of relevant diagonal blocks and the scaled off-diagonal blocks are then approximated by low-rank forms. It is shown in [23, 24] that the resulting SIF preconditioners have some attractive features. For example, some effectiveness results can be conveniently shown for the preconditioners and under certain assumptions, the scaled off-diagonal blocks can be aggressively compressed so as to yield fast and effective multilevel SIF preconditioners. Practical numerical tests have shown superior convergence results [24]. The scaling-and-compression strategy is later also followed by a series of other work [1, 7, 8, 25] for designing structured preconditioners for both dense and sparse matrices. Related ideas also appear in some work for preconditioning sparse matrices [12, 13, 14, 20].

The analysis in [24] aims at general SPD matrices and ignores specific properties and backgrounds. Thus, some of the results are very conservative. For example, the analysis in [24] for a multilevel SIF preconditioner is based on some restrictive robustness requirements. Specifically, the effectiveness and robustness analysis requires that the matrix is not too ill conditioned, the off-diagonal compression accuracy is not too low, or the number of levels is not too large. These requirements are needed in order to guarantee the positive definiteness of the preconditioner. However, the requirements either limit the applicability of the preconditioner or make the preconditioner too expensive. On the other hand, many practical tests have shown nice performance even though such requirements are not met.

In addition, there are two types of SIF preconditioners in [24], one based on Cholesky factorizations and another based on a so-called ULV factorization [6, 22]. The analysis is done for Cholesky SIF preconditioning while the implementation is for ULV SIF preconditioning since the latter has better scalability and stability. The effectiveness of ULV SIF preconditioning is not clear.

In this work, we revisit the analysis for the SIF preconditioning technique and give new insights into the effectiveness and robustness. Our aim is to provide better understanding of the performance in terms of both general spectral analysis and studies of some model problems and show that it is possible to relax the robustness requirements in [24]. The main contributions are as follows.

1. We provide more intuitive studies on the effectiveness of SIF preconditioning, especially some spectral analysis for ULV SIF preconditioners and show that they are as effective as the Cholesky SIF preconditioners. This confirms that ULV SIF preconditioners are the better choice in practice due to the nice stability and scalability.
2. We give concrete illustrations of the effectiveness of SIF preconditioning in terms of a type of 2D and 3D discretized model problems that has often been used to study some similar preconditioners in other work [12, 13, 14]. Singular values of the scaled off-diagonal blocks are derived and are used to show the condition number and eigenvalue distribution after preconditioning. Explicit forms of the preconditioners are also derived so as to understand the behaviors of the scaling-and-compression strategy in multilevel SIF preconditioning.
3. Furthermore, our studies indicate that multilevel SIF preconditioning has an implicit Schur complement compensation effect [10, 23], which can help enhance the robustness of the

resulting preconditioners. In fact, for the model problems, we can show that the requirements in [24] needed to guarantee positive definiteness are too conservative and may be relaxed. Actually, the multilevel SIF preconditioners are unconditionally robust or breakdown free for those problems. That is, the SIF preconditioners remain positive definite regardless of the off-diagonal compression accuracy and the number of levels. More specifically, in the multilevel scaling-and-compression strategy, the leading singular values of the scaled off-diagonal blocks remain unchanged. Such studies give a good indication that SIF preconditioners likely have much better robustness than predicted in [24].

Our studies give new perspectives for the SIF preconditioning technique and the scaling-and-compression strategy, and confirm that the technique is an effective and reliable way to design new structured preconditioners with guaranteed performance. That is, it is beneficial to combine scaling with off-diagonal compression in the design of structured preconditioners. Our work suggests that it is feasible to obtain even stronger analysis results for SIF preconditioning applied to specific applications. It also provides useful tools for analyzing and understanding other structured preconditioners. Various spectral analysis results can be used to characterize other structured algorithms and study more general problems. The work also suggests new directions for improving SIF preconditioning.

Our discussions include three parts. We provide some spectral analysis to illustrate the effectiveness of SIF preconditioning in Section 2. The effectiveness of SIF preconditioning is further demonstrated in terms of the model problems in Section 3. Section 4 discusses the robustness of multilevel SIF schemes. The analysis is also aided by some numerical evidences. To facilitate the discussions, we list commonly used notation as follows.

- $\lambda(A)$ denotes an eigenvalue of a symmetric matrix A and $\lambda_j(A)$ denotes the jth *smallest* eigenvalue of A.
- $\sigma_j(C)$ denotes the jth *largest* singular value of a matrix C.
- $\kappa(A)$ is the 2-norm condition number of A.
- $\mathrm{diag}(\cdot)$ denotes a (block) diagonal matrix with the given (block) diagonals.
- I_r is the $r \times r$ identity matrix.
- When an $n \times n$ matrix A is partitioned, the partitioning is denoted by the splitting of its index set $\{1:n\}$. For example, $\{1:n\} = \{1:n_1\} \cup \{n_1+1:n\}$ denotes a block 2×2 partitioning of A with the $(1,1)$ and $(2,2)$ diagonal blocks corresponding to the index sets $\{1:n_1\}$ and $\{n_1+1:n\}$, respectively.

2. SPECTRAL ANALYSIS FOR SIF PRECONDITIONING

In this section, we first give a quick review of SIF preconditioning for SPD matrices and then revisit the effectiveness in terms of some spectral analysis.

2.1. Review of SIF preconditioning for SPD matrices

The SIF preconditioning strategy is built upon a scaling-and-compression strategy [23, 24]. In this strategy, off-diagonal blocks are first scaled and then compressed so as to justify the effectiveness and to better control of the performance. The basic idea from [24] is as follows.

Consider a block 2×2 SPD matrix

$$A \equiv \begin{pmatrix} A_{\mathbf{11}} & A_{\mathbf{12}} \\ A_{\mathbf{12}}^T & A_{\mathbf{22}} \end{pmatrix}, \tag{1}$$

where the two diagonal blocks are assumed to have Cholesky factorizations

$$A_{\mathbf{kk}} = L_{\mathbf{k}} L_{\mathbf{k}}^T, \quad \mathbf{k} = \mathbf{1}, \mathbf{2}. \tag{2}$$

(Here, bold fonts are used for the subscripts in order to be consistent with later notation.) Then A can be factorized as

$$A = \begin{pmatrix} L_{\mathbf{1}} & \\ & L_{\mathbf{2}} \end{pmatrix} \begin{pmatrix} I & C \\ C^T & I \end{pmatrix} \begin{pmatrix} L_{\mathbf{1}}^T & \\ & L_{\mathbf{2}}^T \end{pmatrix}, \quad \text{with} \quad C = L_{\mathbf{1}}^{-1} A_{\mathbf{12}} L_{\mathbf{2}}^{-T}. \tag{3}$$

Suppose the full SVD of C and its rank-r truncation look like

$$C = \begin{pmatrix} \tilde{U}_{\mathbf{1}} & \hat{U}_{\mathbf{1}} \end{pmatrix} \begin{pmatrix} \tilde{\Sigma}_{\mathbf{1}} & \\ & \hat{\Sigma}_{\mathbf{1}} \end{pmatrix} \begin{pmatrix} \tilde{U}_{\mathbf{2}}^T \\ \hat{U}_{\mathbf{2}}^T \end{pmatrix} \approx \tilde{U}_{\mathbf{1}} \tilde{\Sigma}_{\mathbf{1}} \tilde{U}_{\mathbf{2}}^T, \tag{4}$$

where $\tilde{\Sigma}_{\mathbf{1}}$ is a diagonal matrix for the r singular values of C that are supposed to be greater than or equal to a tolerance τ: $\sigma_1(C) \geq \ldots \geq \sigma_r(C) \geq \tau$. That is, $\sigma_{r+1}(C) \leq \tau$ is the largest dropped singular value of C. With the truncated SVD, A can be approximated by

$$\tilde{A} \equiv \begin{pmatrix} L_{\mathbf{1}} & \\ & L_{\mathbf{2}} \end{pmatrix} \begin{pmatrix} I & \tilde{U}_{\mathbf{1}} \tilde{\Sigma}_{\mathbf{1}} \tilde{U}_{\mathbf{2}}^T \\ \tilde{U}_{\mathbf{2}} \tilde{\Sigma}_{\mathbf{1}} \tilde{U}_{\mathbf{1}}^T & I \end{pmatrix} \begin{pmatrix} L_{\mathbf{1}}^T & \\ & L_{\mathbf{2}}^T \end{pmatrix}.$$

Then we get a prototype SIF preconditioner

$$\tilde{A} = \tilde{\mathbf{L}} \tilde{\mathbf{L}}^T, \tag{5}$$

where

$$\tilde{\mathbf{L}} \equiv \begin{pmatrix} L_{\mathbf{1}} & \\ L_{\mathbf{2}} \tilde{U}_{\mathbf{2}} \tilde{\Sigma}_{\mathbf{1}} \tilde{U}_{\mathbf{1}}^T & L_{\mathbf{2}} \tilde{D}_{\mathbf{2}} \end{pmatrix}, \quad \text{with} \quad \tilde{D}_{\mathbf{2}} \tilde{D}_{\mathbf{2}}^T = I - \tilde{U}_{\mathbf{2}} \tilde{\Sigma}_{\mathbf{1}}^2 \tilde{U}_{\mathbf{2}}^T. \tag{6}$$

In practice, a ULV-type factorization [6, 22] is used to enhance the scalability since it avoids the sequential computation of Schur complements and uses a hierarchical scheme where local factorizations at each level can be done simultaneously. Let $Q_{\mathbf{1}}$ be an orthogonal matrix constructed from $\tilde{U}_{\mathbf{1}}$ with $\tilde{U}_{\mathbf{1}}$ as the first r columns and $Q_{\mathbf{2}}$ be constructed similarly from $\tilde{U}_{\mathbf{2}}$. Then (5) becomes a ULV factorization with $\tilde{\mathbf{L}}$ a ULV factor

$$\tilde{\mathbf{L}} = \begin{pmatrix} L_{\mathbf{1}} & \\ & L_{\mathbf{2}} \end{pmatrix} \begin{pmatrix} Q_{\mathbf{1}} & \\ & Q_{\mathbf{2}} \end{pmatrix} \Pi \begin{pmatrix} H & \\ & I \end{pmatrix}, \tag{7}$$

where H is the lower triangular Cholesky factor of $\begin{pmatrix} I & \tilde{\Sigma}_{\mathbf{1}} \\ \tilde{\Sigma}_{\mathbf{1}} & I \end{pmatrix}$ and Π is a permutation matrix used to assemble $\begin{pmatrix} I & \tilde{\Sigma}_{\mathbf{1}} \\ \tilde{\Sigma}_{\mathbf{1}} & I \end{pmatrix}$ like

$$\Pi^T \begin{pmatrix} I & \mathrm{diag}(\tilde{\Sigma}_{\mathbf{1}}, 0) \\ \mathrm{diag}(\tilde{\Sigma}_{\mathbf{1}}, 0) & I \end{pmatrix} \Pi = \mathrm{diag}\left(\begin{pmatrix} I & \tilde{\Sigma}_{\mathbf{1}} \\ \tilde{\Sigma}_{\mathbf{1}} & I \end{pmatrix}, I \right). \tag{8}$$

Generalization of the prototype preconditioner to practical multilevel schemes is also made in [24]. The procedure above with 1-level partitioning of A is called a 1-*level* (or *prototype*) *SIF scheme*. For convenience, we call the preconditioner (5) with the factor in (6) a 1-level Cholesky

SIF preconditioner and (5) with the factor in (7) a 1-level ULV SIF preconditioner. The same idea may be applied to $A_{\mathbf{11}}$ and $A_{\mathbf{22}}$ to yield approximate factors $\tilde{L}_{\mathbf{1}} \approx L_{\mathbf{1}}$ and $\tilde{L}_{\mathbf{2}} \approx L_{\mathbf{2}}$, respectively. If $\tilde{L}_{\mathbf{1}}$ and $\tilde{L}_{\mathbf{2}}$ are used to replace $L_{\mathbf{1}}$ and $L_{\mathbf{2}}$ respectively in the procedure above, then the procedure is a 2-*level SIF scheme*. Similarly, a general l-level SIF scheme can be obtained.

The work [24] provides some analysis results on the effectiveness the 1-level Cholesky SIF preconditioner. It is shown that the preconditioned matrix has a form

$$\tilde{\mathbf{L}}^{-1} A \tilde{\mathbf{L}}^{-T} = \begin{pmatrix} I & \hat{C} \\ \hat{C} & I \end{pmatrix}, \tag{9}$$

where

$$\hat{C} = \hat{U}_{\mathbf{1}} \hat{\Sigma}_{\mathbf{1}} \hat{U}_{\mathbf{2}}^T \tilde{D}_{\mathbf{2}}^{-T}, \quad \|\hat{C}\|_2 = \sigma_{r+1}(C). \tag{10}$$

Thus, the 2-norm condition number of the preconditioned matrix is

$$\kappa(\tilde{\mathbf{L}}^{-1} A \tilde{\mathbf{L}}^{-T}) = \frac{1 + \sigma_{r+1}(C)}{1 - \sigma_{r+1}(C)}. \tag{11}$$

2.2. Spectral analysis for Cholesky and ULV SIF preconditioning

The effectiveness analysis in [24] is done only for the Cholesky SIF scheme in Section 2.1. On the other hand, the actual implementation is based on the ULV SIF schemes which have better scalability and stability. Here, we show that both types of schemes are similarly effective and also give a more intuitive explanation of the effectiveness by extending (10).

Theorem 2.1. *Suppose the smaller of the row and column sizes of C in (3) is k. For the Cholesky SIF factor $\tilde{\mathbf{L}}$ in (6), the equation (9) holds with the nonzero singular values of $\hat{C}$ in* (10) *given by*

$$\sigma_j(\hat{C}) = \sigma_{r+j}(C) \le \tau, \quad j = 1, 2, \ldots, k - r. \tag{12}$$

For the ULV SIF factor $\tilde{\mathbf{L}}$ in (7), we have

$$\tilde{\mathbf{L}}^{-1} A \tilde{\mathbf{L}}^{-T} = \operatorname{diag}\left(I, \begin{pmatrix} I & \bar{C} \\ \bar{C} & I \end{pmatrix} \right), \tag{13}$$

where $\bar{C}$ is an $(k-r) \times (k-r)$ matrix with singular values

$$\sigma_j(\bar{C}) = \sigma_{r+j}(C) \le \tau, \quad j = 1, 2, \ldots, k - r. \tag{14}$$

Accordingly, for $\tilde{\mathbf{L}}$ in either (6) or (7), the eigenvalues of $\tilde{\mathbf{L}}^{-1} A \tilde{\mathbf{L}}^{-T}$ are

$$1, 1 \pm \sigma_{r+1}(C), \ldots, 1 \pm \sigma_k(C), \tag{15}$$

where the eigenvalue 1 *has multiplicity $n - 2(k-r)$ with n the order of A.*

Proof. For $\tilde{\mathbf{L}}$ in (6), the proof for (10) [24, Theorem 2.5] already implies (12). That is, any eigenvalue $\lambda(\hat{C}\hat{C}^T)$ of $\hat{C}\hat{C}^T$ satisfies

$$\begin{aligned} \lambda(\hat{C}^T \hat{C}) &= \lambda(\tilde{D}_{\mathbf{2}}^{-1} \hat{U}_{\mathbf{2}} \hat{\Sigma}_{\mathbf{1}}^T \hat{\Sigma}_{\mathbf{1}} \hat{U}_{\mathbf{2}}^T \tilde{D}_{\mathbf{2}}^{-T}) = \lambda(\tilde{D}_{\mathbf{2}}^{-T} \tilde{D}_{\mathbf{2}}^{-1} \hat{U}_{\mathbf{2}} \hat{\Sigma}_{\mathbf{1}}^T \hat{\Sigma}_{\mathbf{1}} \hat{U}_{\mathbf{2}}^T) \\ &= \lambda\left((I - \tilde{U}_{\mathbf{2}} \tilde{\Sigma}_{\mathbf{1}}^2 \tilde{U}_{\mathbf{2}}^T)^{-1} \hat{U}_{\mathbf{2}} \hat{\Sigma}_{\mathbf{1}}^T \hat{\Sigma}_{\mathbf{1}} \hat{U}_{\mathbf{2}}^T \right) = \lambda(\hat{U}_{\mathbf{2}} \hat{\Sigma}_{\mathbf{1}}^T \hat{\Sigma}_{\mathbf{1}} \hat{U}_{\mathbf{2}}^T) \\ &\in \left\{ \sigma_{r+1}^2(C), \ldots, \sigma_k^2(C), 0, \ldots, 0 \right\}, \end{aligned}$$

where the equality in the second line is due to the Sherman-Morrison-Woodbury formula and the result $\tilde{U}_{\mathbf{2}}^T \hat{U}_{\mathbf{2}} = 0$.

Thus, we just focus on the ULV SIF factor $\tilde{\mathbf{L}}$ in (7). From (3), we have

$$\begin{aligned}\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T} &= \begin{pmatrix} H^{-1} & \\ & I \end{pmatrix}\Pi^T\begin{pmatrix} Q_{\mathbf{1}}^T & \\ & Q_{\mathbf{2}}^T \end{pmatrix}\begin{pmatrix} I & C \\ C^T & I \end{pmatrix}\begin{pmatrix} Q_{\mathbf{1}} & \\ & Q_{\mathbf{2}} \end{pmatrix}\Pi\begin{pmatrix} H^{-T} & \\ & I \end{pmatrix} \\ &= \begin{pmatrix} H^{-1} & \\ & I \end{pmatrix}\Pi^T\begin{pmatrix} I & Q_{\mathbf{1}}^TCQ_{\mathbf{2}} \\ Q_{\mathbf{2}}^TC^TQ_{\mathbf{1}} & I \end{pmatrix}\Pi\begin{pmatrix} H^{-T} & \\ & I \end{pmatrix}.\end{aligned}$$

Based on the construction of $Q_{\mathbf{1}}$ and $Q_{\mathbf{2}}$, we can let $Q_{\mathbf{1}} = \begin{pmatrix} \tilde{U}_{\mathbf{1}} & \hat{Q}_{\mathbf{1}} \end{pmatrix}$, $Q_{\mathbf{2}} = \begin{pmatrix} \tilde{U}_{\mathbf{2}} & \hat{Q}_{\mathbf{2}} \end{pmatrix}$. Then

$$\begin{aligned}Q_{\mathbf{1}}^TCQ_{\mathbf{2}} &= \begin{pmatrix} \tilde{U}_{\mathbf{1}}^T \\ \hat{Q}_{\mathbf{1}}^T \end{pmatrix}\begin{pmatrix} \tilde{U}_{\mathbf{1}} & \hat{U}_{\mathbf{1}} \end{pmatrix}\begin{pmatrix} \tilde{\Sigma}_{\mathbf{1}} & \\ & \hat{\Sigma}_{\mathbf{1}} \end{pmatrix}\begin{pmatrix} \tilde{U}_{\mathbf{2}}^T \\ \hat{U}_{\mathbf{2}}^T \end{pmatrix}\begin{pmatrix} \tilde{U}_{\mathbf{2}} & \hat{Q}_{\mathbf{2}} \end{pmatrix} \\ &= \begin{pmatrix} I & \\ & \hat{Q}_{\mathbf{1}}^T\hat{U}_{\mathbf{1}} \end{pmatrix}\begin{pmatrix} \tilde{\Sigma}_{\mathbf{1}} & \\ & \hat{\Sigma}_{\mathbf{1}} \end{pmatrix}\begin{pmatrix} I & \\ & \hat{U}_{\mathbf{2}}^T\hat{Q}_{\mathbf{2}} \end{pmatrix} \\ &= \mathrm{diag}(\tilde{\Sigma}_{\mathbf{1}}, \bar{C}),\end{aligned}$$

where

$$\bar{C} = \hat{Q}_{\mathbf{1}}^T\hat{U}_{\mathbf{1}}\hat{\Sigma}_{\mathbf{1}}\hat{U}_{\mathbf{2}}^T\hat{Q}_{\mathbf{2}}. \tag{16}$$

Thus,

$$\begin{aligned}\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T} &= \begin{pmatrix} H^{-1} & \\ & I \end{pmatrix}\Pi^T\begin{pmatrix} I & \mathrm{diag}(\tilde{\Sigma}_{\mathbf{1}}, \bar{C}) \\ \mathrm{diag}(\tilde{\Sigma}_{\mathbf{1}}, \bar{C}^T) & I \end{pmatrix}\Pi\begin{pmatrix} H^{-T} & \\ & I \end{pmatrix} \\ &= \begin{pmatrix} H^{-1} & \\ & I \end{pmatrix}\mathrm{diag}\left(\begin{pmatrix} I & \tilde{\Sigma}_{\mathbf{1}} \\ \tilde{\Sigma}_{\mathbf{1}} & I \end{pmatrix}, \begin{pmatrix} I & \bar{C} \\ \bar{C}^T & I \end{pmatrix}\right)\begin{pmatrix} H^{-T} & \\ & I \end{pmatrix} \\ &= \mathrm{diag}\left(H^{-1}\begin{pmatrix} I & \tilde{\Sigma}_{\mathbf{1}} \\ \tilde{\Sigma}_{\mathbf{1}} & I \end{pmatrix}H^{-T}, \begin{pmatrix} I & \bar{C} \\ \bar{C}^T & I \end{pmatrix}\right) \\ &= \mathrm{diag}\left(I, \begin{pmatrix} I & \bar{C} \\ \bar{C}^T & I \end{pmatrix}\right),\end{aligned}$$

where the second step follows from the definition of Π as in (8).

Then we show (14) for $\bar{C}$ in (16). Note that $\hat{Q}_{\mathbf{1}}^T\hat{U}_{\mathbf{1}}$ and $\hat{Q}_{\mathbf{2}}^T\hat{U}_{\mathbf{2}}$ are orthogonal matrices since, say,

$$\hat{Q}_{\mathbf{1}}^T\hat{U}_{\mathbf{1}}(\hat{Q}_{\mathbf{1}}^T\hat{U}_{\mathbf{1}})^T = \hat{Q}_{\mathbf{1}}^T\hat{U}_{\mathbf{1}}\hat{U}_{\mathbf{1}}^T\hat{Q}_{\mathbf{1}} = \hat{Q}_{\mathbf{1}}^T(I - \tilde{U}_{\mathbf{1}}\tilde{U}_{\mathbf{1}}^T)\hat{Q}_{\mathbf{1}} = \hat{Q}_{\mathbf{1}}^T\hat{Q}_{\mathbf{1}} = I,$$

where we used $\hat{Q}_{\mathbf{1}}^T\tilde{U}_{\mathbf{1}} = 0$. Thus, the singular values of $\bar{C}$ are the diagonal entries of $\hat{\Sigma}_{\mathbf{1}}$ and are $\sigma_{r+1}(C), \ldots, \sigma_k(C)$. (14) then holds.

The eigenvalues of $\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T}$ can then be immediately obtained based on (9) or (13). ∎

The theorem has two implications. One is in the understanding of the effectiveness of SIF preconditioning. If A is preconditioned with just the block diagonal preconditioner $\mathrm{diag}(A_{\mathbf{11}}, A_{\mathbf{22}})$, it is known that the preconditioned matrix has eigenvalues $1 \pm \sigma_1(C), \ldots, 1 \pm \sigma_k(C)$ and a repeated eigenvalue 1 of multiplicity $n-k$. The condition number after preconditioning is $\frac{1+\sigma_1(C)}{1-\sigma_1(C)}$. By keeping the r largest singular values of C in the off-diagonal compression, the r largest (smallest) eigenvalues $1 \pm \sigma_1(C), \ldots, 1 \pm \sigma_r(C)$ are mapped to 1. If $\sigma_j(C)$ has reasonable decay, then the eigenvalues of the preconditioned matrix cluster reasonably close to 1. In fact, the scaling of $A_{\mathbf{12}}$ may help enhance the singular value decay (an example is the model problem that will be considered later; see Remark 2). Also, the condition number of the preconditioned matrix becomes (11). As pointed out in [24], the significance of (11) lies in a decay magnification effect. That is, if the singular values $\sigma_i(C)$ slightly

decays, then $\kappa(\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T})$ decays much faster, so that an aggressive truncation of $\sigma_j(C)$ leads to a reasonable condition number. (11) and (15) indicate that SIF preconditioning can improve both the condition number and the eigenvalue distribution. They also suggest that, to design new effective structured preconditioners, it is desirable to accelerate the decay of the singular values of the scaled off-diagonal blocks.

Another implication of the theorem is in terms of the choice between Cholesky and ULV SIF preconditioners. The eigenvalue distribution is the same after either the Cholesky SIF preconditioning or the ULV SIF preconditioning. The ULV SIF scheme avoids the computation of explicit Schur complements and has better scalability. It also mainly uses orthogonal rotations (like $Q_{\mathbf{1}}, Q_{\mathbf{2}}$) in the intermediate factorizations and has better stability. Thus in practical implementations, ULV SIF preconditioners are preferred. On the other hand, Cholesky SIF preconditioners are often more convenient for the analysis purpose. Thus in our later analysis, we then only use the Cholesky SIF scheme.

For general l-level SIF schemes, results similar to (11) can be obtained [24], provided that the SIF preconditioners remain positive definite. This positive definiteness requirement will be investigated in Section 4.

3. EFFECTIVENESS OF SIF PRECONDITIONING FOR 2D AND 3D DISCRETIZED PROBLEMS

We then illustrate the effectiveness of SIF preconditioning via a type of model problems so as to obtain more concrete estimates. Consider the finite difference discretization of $-\Delta$ on 2D or 3D grids with Dirichlet boundary conditions. 5-point and 7-point stencils are used for the 2D and 3D cases, respectively. We would like to analyze the performance of SIF preconditioning when it is applied to the discretized matrix A and give specific effectiveness and robustness estimates.

Analysis for such a model problem is important due to multiple reasons.

1. As already mentioned in [24], multilevel SIF preconditioners can be directly applied to sparse matrices due to some attractive features. For example, the fast sparse matrix-vector products can be used to quickly compress scaled off-diagonal blocks like in (3) based on randomized SVDs [15]. This can help significantly reduce the cost to construct a multilevel SIF preconditioner from about $O(n^2)$ flops in [24] to about $O(n)$ flops. Such randomized structured approximation is similar to the procedures used in [9, 16, 17, 21]. (Some similar preconditioners not based on randomization have also been applied to sparse matrices [12, 13, 14].)

2. This model problem is a useful representative discretized problem and can help us gain better insights into the performance of SIF preconditioning for sparse discretized problems. In fact, the model problem has often been used to analyze and understand low-rank compression based multilevel preconditioners in other work [12, 13, 14].

3. This model problem is actually a somewhat challenging problem for standard rank structured preconditioners based on direct off-diagonal compression, since the off-diagonal blocks of A involve a negative identity submatrix that is not compressible in the usual sense. On the other hand, off-diagonal scaling like in (3) leads to reasonable decay in the singular values of the scaled off-diagonal blocks, as can be seen later. Our numerical tests later also show significant advantages over preconditioners based on straightforward structured approximations.

4. Our results here for the model problem can further serve as tools for studying other similar structured preconditioners. In fact, even for the multilevel SIF scheme, it is still feasible to perform various analysis such as spectral analysis for the scaled off-diagonal blocks. Indeed, some strong claims can be made, as shown in the remaining discussions.

The discretized matrix A from the model problem has a block tridiagonal form. Without loss of generality, we assume that the discretization mesh is $N \times M$ in two dimensions and $N \times N \times M$ in three dimensions with the outermost ordering of the mesh points in the last direction, so that A has M diagonal blocks T. In the 2D case, each diagonal block T corresponds to a 1D slice of the mesh and has size $\mathcal{N} \equiv N$. In the 3D case, T corresponds to a 2D slice of the mesh and has size $\mathcal{N} \equiv N^2$. Also assume any partitioning of A does not split the T blocks in the analysis later. A has the index set $\{1 : M\mathcal{N}\}$. Later, when we refer to *the model problem*, we assume this setup is used.

REMARK 1. Note that, like various other related model problem studies in [5, 12, 13, 14], *the focus here is not on how to solve such "easy" discretized problems.* Rather, we use the model problems to gain useful insights into the behaviors of the techniques under consideration. Here, we use the model problems to better understand the potential of SIF preconditioning. As shown in our numerical tests later, multilevel structured preconditioners based on straightforward off-diagonal compression have difficulties to handle the model problems. On the other hand, SIF preconditioning works significantly better. Even for such standard model problems, the analysis for SIF preconditioning is nontrivial. We anticipate that the analysis here can serve as a starting point for studying and designing SIF preconditioners for more practical discretized problems. Readers who are interested in numerical evidences for practical sparse problems are referred to [7, 8, 13, 14, 24].

3.1. Singular values of the scaled off-diagonal blocks

A key point in the analysis of the effectiveness and robustness of SIF preconditioning for the model problem is to derive the singular values of the scaled off-diagonal blocks. In this subsection, we focus on the scaled off-diagonal block C in (3) from the 1-level SIF scheme. Results on the multilevel SIF scheme will be given in Section 4.1. Suppose the partitioning in (1) follows the index set

$$\{1 : M\mathcal{N}\} = \{1 : m_{\mathbf{1}}\mathcal{N}\} \cup \{m_{\mathbf{1}}\mathcal{N} + 1 : M\mathcal{N}\}, \tag{17}$$

such that $A_{\mathbf{11}}$ corresponds to the leading $m_{\mathbf{1}}$ diagonal blocks T in A, $A_{\mathbf{22}}$ corresponds to the remaining $m_{\mathbf{2}} = M - m_{\mathbf{1}}$ diagonal blocks T in A, and

$$A_{\mathbf{12}} = \begin{pmatrix} 0 & 0 \\ -I_{\mathcal{N}} & 0 \end{pmatrix}. \tag{18}$$

Suppose $A_{\mathbf{11}}$ and $A_{\mathbf{22}}$ have Cholesky factorizations as in (2). We would like to derive the singular values of $C = L_{\mathbf{1}}^{-1} A_{\mathbf{12}} L_{\mathbf{2}}^{-T}$.

The specific forms of $L_{\mathbf{1}}$ and $L_{\mathbf{2}}$ can be conveniently written down as follows. Let

$$S_1 = T, \quad S_i = T - S_{i-1}^{-1}, \quad i = 2, 3, \ldots \tag{19}$$

Suppose the Cholesky factorization of S_i is

$$S_i = K_i K_i^T. \tag{20}$$

Then $L_{\mathbf{k}}$ for $\mathbf{k} = \mathbf{1}, \mathbf{2}$ has the following form:

$$L_{\mathbf{k}} \equiv \begin{pmatrix} K_1 & & & \\ -K_1^{-T} & \ddots & & \\ & \ddots & \ddots & \\ & & -K_{m_{\mathbf{k}}-1}^{-T} & K_{m_{\mathbf{k}}} \end{pmatrix}. \tag{21}$$

(Here, the subscripts in black fonts are associated with the partitioning of A as in (1), and the subscripts in regular fonts are for the original block tridiagonal partitioning of A due to the discretization.)

Let the eigenvalue decomposition of T be

$$T = Q\Lambda_1 Q^T, \quad \text{with} \quad \Lambda_1 = \operatorname{diag}\left(\lambda_1(T), \lambda_2(T), \ldots, \lambda_{\mathcal{N}}(T)\right), \tag{22}$$

where the eigenvalues are ordered as $\lambda_1(T) < \lambda_2(T) < \cdots < \lambda_{\mathcal{N}}(T)$. Then all S_i and S_i^{-1} have the same eigenvector matrices Q because of (19). Corresponding to (19), let the eigenvalue decomposition of S_i be

$$S_i = Q\Lambda_i Q^T, \quad \text{with} \quad \Lambda_i = \Lambda_1 - \Lambda_{i-1}^{-1}, \quad i = 2, 3, \ldots \tag{23}$$

Again, $\Lambda_i = \operatorname{diag}(\lambda_1(S_i), \lambda_2(S_i), \ldots, \lambda_{\mathcal{N}}(S_i))$ with the eigenvalues $\lambda_1(S_i) < \lambda_2(S_i) < \cdots < \lambda_{\mathcal{N}}(S_i)$.

The following lemma will be used frequently later.

Lemma 3.1. *Let $d_1 > 2$ and $d_i = d_1 - d_{i-1}^{-1}$ for $i = 2, 3, \ldots$ Then*

$$1 < d_i < d_{i-1}, \tag{24}$$

$$d_1^{-1} + d_1^{-1}d_2^{-1}d_1^{-1} + \cdots + d_1^{-1}\cdots d_{i-1}^{-1}d_i^{-1}d_{i-1}^{-1}\cdots d_1^{-1} = d_i^{-1}. \tag{25}$$

Accordingly, Λ_i in (23) satisfies

$$\Lambda_1^{-1} + \Lambda_1^{-1}\Lambda_2^{-1}\Lambda_1^{-1} + \cdots + \Lambda_1^{-1}\cdots\Lambda_{i-1}^{-1}\Lambda_i^{-1}\Lambda_{i-1}^{-1}\cdots\Lambda_1^{-1} = \Lambda_i^{-1}. \tag{26}$$

Proof. We prove this by induction. (24)–(25) are obviously true for $i = 1, 2$. Suppose they hold for $i-1$ with $i > 3$. We show they also hold for i. Let $w_i = d_1^{-1} + d_1^{-1}d_2^{-1}d_1^{-1} + \cdots + d_1^{-1}\cdots d_{i-1}^{-1}d_i^{-1}d_{i-1}^{-1}\cdots d_1^{-1}$. Since $d_1 > 2$, $d_{i-1} > 1$, we have $d_i = d_1 - d_{i-1}^{-1} > 1$. Then

$$\begin{aligned} d_i - d_{i-1} &= d_1 - d_{i-1}^{-1} - (d_1 - d_{i-2}^{-1}) = d_{i-2}^{-1} - d_{i-1}^{-1} = w_{i-2} - w_{i-1} \\ &= -d_1^{-1}\cdots d_{i-2}^{-1}d_{i-1}^{-1}d_{i-2}^{-1}\cdots d_1^{-1} < 0, \end{aligned}$$

and (24) holds. Also,

$$\begin{aligned} w_i &= w_{i-1} + d_1^{-1}\cdots d_{i-1}^{-1}d_i^{-1}d_{i-1}^{-1}\cdots d_1^{-1} = d_{i-1}^{-1} + d_1^{-1}\cdots d_{i-1}^{-1}d_i^{-1}d_{i-1}^{-1}\cdots d_1^{-1} \\ &= d_i^{-1}d_{i-1}^{-1}(d_i + d_1^{-1}\cdots d_{i-2}^{-1}d_{i-1}^{-1}d_{i-2}^{-1}\cdots d_1^{-1}) \\ &= d_i^{-1}d_{i-1}^{-1}(d_1 - d_{i-1}^{-1} + d_1^{-1}\cdots d_{i-2}^{-1}d_{i-1}^{-1}d_{i-2}^{-1}\cdots d_1^{-1}) \\ &= d_i^{-1}d_{i-1}^{-1}(d_1 - w_{i-1} + d_1^{-1}\cdots d_{i-2}^{-1}d_{i-1}^{-1}d_{i-2}^{-1}\cdots d_1^{-1}) = d_i^{-1}d_{i-1}^{-1}(d_1 - w_{i-2}) \\ &= d_i^{-1}d_{i-1}^{-1}(d_1 - d_{i-2}^{-1}) = d_i^{-1}d_{i-1}^{-1}d_{i-1} = d_i^{-1}. \end{aligned}$$

It is known that, for the 2D or 3D model problem, all the eigenvalues of T are greater than 2. Then (23) and (25) yield (26). ∎

We are now ready to present the following theorem.

Theorem 3.1. *For the discretized matrix A from the 2D or 3D model problem partitioned as in (1) following* (17)*, the nonzero singular values of $L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T}$ are*

$$\sigma_j(L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T}) = \sqrt{\lambda_j(S_{m_{\mathbf{1}}}^{-1})\lambda_j(S_{m_{\mathbf{2}}}^{-1})}, \quad j = 1, 2, \ldots, \mathcal{N},$$

where the S_i matrices are given in (19)*.*

Proof. According to (21), for $\mathbf{k} = \mathbf{1}, \mathbf{2}$,

$$\begin{aligned} L_{\mathbf{k}}^{-1} &= \left(\begin{pmatrix} I & & & \\ -S_1^{-1} & \ddots & & \\ & \ddots & \ddots & \\ & & -S_{m_{\mathbf{k}}-1}^{-1} & I \end{pmatrix} \begin{pmatrix} K_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & K_{m_{\mathbf{k}}} \end{pmatrix} \right)^{-1} \\ &= \begin{pmatrix} K_1^{-1} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & K_{m_{\mathbf{k}}}^{-1} \end{pmatrix} \begin{pmatrix} I & & & \\ S_1^{-1} & & I & \\ \vdots & & \vdots & \ddots \\ S_{m_{\mathbf{k}}-1}^{-1}\cdots S_1^{-1} & S_{m_{\mathbf{k}}-1}^{-1}\cdots S_2^{-1} & \cdots & I \end{pmatrix} \\ &= \begin{pmatrix} K_1^{-1} & & & \\ K_2^{-1}S_1^{-1} & & K_2^{-1} & \\ \vdots & & \vdots & \ddots \\ K_{m_{\mathbf{k}}}^{-1}S_{m_{\mathbf{k}}-1}^{-1}\cdots S_1^{-1} & K_{m_{\mathbf{k}}}^{-1}S_{m_{\mathbf{k}}-1}^{-1}\cdots S_2^{-1} & \cdots & K_{m_{\mathbf{k}}}^{-1} \end{pmatrix}. \end{aligned} \tag{27}$$

With $A_{\mathbf{12}}$ in (18), we have

$$L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T} = \begin{pmatrix} 0 \\ -K_{m_{\mathbf{1}}}^{-1}Z \end{pmatrix}, \tag{28}$$

where Z is the first row of $L_{\mathbf{2}}^{-T}$ as follows:

$$Z = \begin{pmatrix} K_1^{-T} & S_1^{-1}K_2^{-T} & \ldots & S_1^{-1}\cdots S_{m_{\mathbf{2}}-1}^{-1}K_{m_{\mathbf{2}}}^{-T} \end{pmatrix}. \tag{29}$$

Thus, the nonzero singular values of $L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T}$ are given by

$$\sigma_j(L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T}) = \sqrt{\lambda_{\mathcal{N}-j}(K_{m_{\mathbf{1}}}^{-1}ZZ^TK_{m_{\mathbf{1}}}^{-T})} = \sqrt{\lambda_{\mathcal{N}-j}(ZZ^TS_{m_{\mathbf{1}}}^{-1})},\ j = 1, 2, \ldots, \mathcal{N}. \tag{30}$$

(Notice that σ_j's are ordered from the largest to the smallest, and λ_j's are ordered from the smallest to the largest.) Accordingly to (19) and (23),

$$\begin{aligned} ZZ^T &= S_1^{-1} + S_1^{-1}S_2^{-1}S_1^{-1} + \cdots + S_1^{-1}\cdots S_{m_{\mathbf{2}}-1}^{-1}S_{m_{\mathbf{2}}}^{-1}S_{m_{\mathbf{2}}-1}^{-1}\cdots S_1^{-1} \\ &= Q\left(\Lambda_1^{-1} + \Lambda_1^{-1}\Lambda_2^{-1}\Lambda_1^{-1} + \cdots + \Lambda_1^{-1}\cdots\Lambda_{m_{\mathbf{2}}-1}^{-1}\Lambda_{m_{\mathbf{2}}}^{-1}\Lambda_{m_{\mathbf{2}}-1}^{-1}\cdots\Lambda_1^{-1}\right)Q^T \\ &= Q\Lambda_{m_{\mathbf{2}}}^{-1}Q^T, \end{aligned} \tag{31}$$

where the last equality is due to Lemma 3.1. Thus,

$$ZZ^TS_{m_{\mathbf{1}}}^{-1} = Q\Lambda_{m_{\mathbf{2}}}^{-1}\Lambda_{m_{\mathbf{1}}}^{-1}Q^T.$$

The result then follows from (30). ∎

In fact, we can further write the explicit SVD of $L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T}$ as follows, which will be useful later.

Corollary 3.1. *With Q and Λ_i in (23), let the full SVD of K_i in (20) be*

$$K_i = Q\Lambda_i^{\frac{1}{2}} V_i^T. \tag{32}$$

Then the SVD of $L_{\mathbf{1}}^{-1} A_{\mathbf{12}} L_{\mathbf{2}}^{-T}$ is

$$L_{\mathbf{1}}^{-1} A_{\mathbf{12}} L_{\mathbf{2}}^{-T} = U_{\mathbf{1}} \Sigma_{\mathbf{1}} U_{\mathbf{2}}^T, \quad \textit{with} \tag{33}$$
$$U_{\mathbf{1}} = \begin{pmatrix} 0 \\ V_{m_{\mathbf{1}}} \end{pmatrix}, \quad \Sigma_{\mathbf{1}} = \Lambda_{m_{\mathbf{1}}}^{-\frac{1}{2}} \Lambda_{m_{\mathbf{2}}}^{-\frac{1}{2}},$$
$$U_{\mathbf{2}}^T = -\Lambda_{m_{\mathbf{2}}}^{\frac{1}{2}} \begin{pmatrix} \Lambda_1^{-\frac{1}{2}} V_1^T & \Lambda_1^{-1}\Lambda_2^{-\frac{1}{2}} V_2^T & \cdots & \Lambda_1^{-1}\cdots\Lambda_{m_{\mathbf{2}}-1}^{-1}\Lambda_{m_{\mathbf{2}}}^{-\frac{1}{2}} V_{m_2}^T \end{pmatrix}.$$

Proof. With (29), (19), (23), and (32), we have

$$\begin{aligned} -K_{m_{\mathbf{1}}}^{-1} Z &= -K_{m_{\mathbf{1}}}^{-1} \begin{pmatrix} K_1^{-T} & S_1^{-1} K_2^{-T} & \cdots & S_1^{-1}\cdots S_{m_{\mathbf{2}}-1}^{-1} K_{m_{\mathbf{2}}}^{-T} \end{pmatrix} \\ &= -V_{m_{\mathbf{1}}} \Lambda_{m_{\mathbf{1}}}^{-\frac{1}{2}} Q^T \begin{pmatrix} Q\Lambda_1^{-\frac{1}{2}} V_1^T & S_1^{-1} Q \Lambda_2^{\frac{1}{2}} V_2^T & \cdots & S_1^{-1}\cdots S_{m_{\mathbf{2}}-1}^{-1} Q \Lambda_{m_{\mathbf{2}}}^{-\frac{1}{2}} V_{m_2}^T \end{pmatrix} \\ &= -V_{m_{\mathbf{1}}} \Lambda_{m_{\mathbf{1}}}^{-\frac{1}{2}} \begin{pmatrix} \Lambda_1^{-\frac{1}{2}} V_1^T & \Lambda_1^{-1}\Lambda_2^{-\frac{1}{2}} V_2^T & \cdots & \Lambda_1^{-1}\cdots\Lambda_{m_{\mathbf{2}}-1}^{-1}\Lambda_{m_{\mathbf{2}}}^{-\frac{1}{2}} V_{m_2}^T \end{pmatrix} \\ &= V_{m_{\mathbf{1}}} (\Lambda_{m_{\mathbf{1}}}^{-\frac{1}{2}} \Lambda_{m_{\mathbf{2}}}^{-\frac{1}{2}}) \left[-\Lambda_{m_{\mathbf{2}}}^{\frac{1}{2}} \begin{pmatrix} \Lambda_1^{-\frac{1}{2}} V_1^T & \Lambda_1^{-1}\Lambda_2^{-\frac{1}{2}} V_2^T & \cdots & \Lambda_1^{-1}\cdots\Lambda_{m_{\mathbf{2}}-1}^{-1}\Lambda_{m_{\mathbf{2}}}^{-\frac{1}{2}} V_{m_2}^T \end{pmatrix} \right]. \end{aligned}$$

Then (28) yields the SVD in (33), as long as $U_{\mathbf{2}}$ has orthonormal columns. In fact, according to Lemma 3.1,

$$U_{\mathbf{2}}^T U_{\mathbf{2}} = \Lambda_{m_{\mathbf{2}}} (\Lambda_1^{-1} + \Lambda_1^{-1}\Lambda_2^{-1}\Lambda_1^{-1} + \cdots + \Lambda_1^{-1}\cdots\Lambda_{m_{\mathbf{2}}-1}^{-1}\Lambda_{m_{\mathbf{2}}}^{-1}\Lambda_{m_{\mathbf{2}}-1}^{-1}\cdots\Lambda_1^{-1}) = I.$$

∎

Based on Theorem 3.1, we can obtain specific expressions of $\sigma_j(L_{\mathbf{1}}^{-1} A_{\mathbf{12}} L_{\mathbf{2}}^{-T})$ for the model problem in two or three dimensions. For example, the 2D case (where $\mathcal{N} = N$) looks as follows.

Corollary 3.2. *Suppose the same conditions as Theorem 3.1 hold and A is further from the 2D model problem. Let*

$$\theta_j = \eta_j + \sqrt{\eta_j^2 - 1}, \quad \textit{with} \quad \eta_j = 1 + 2\sin^2 \frac{j\pi}{2(N+1)}, \quad j = 1, 2, \ldots, N. \tag{34}$$

Then

$$\sigma_j(L_{\mathbf{1}}^{-1} A_{\mathbf{12}} L_{\mathbf{2}}^{-T}) = \sqrt{\gamma_{m_{\mathbf{1}},j}\gamma_{m_{\mathbf{2}},j}}, \quad j = 1, 2, \ldots, N,$$

where

$$\gamma_{m,j} = \frac{\theta_j^m - \theta_j^{-m}}{\theta_j^{m+1} - \theta_j^{-m-1}}, \quad m = m_{\mathbf{1}}, m_{\mathbf{2}}. \tag{35}$$

Proof. η_j's in (34) are the eigenvalues of $\frac{1}{2}T$. It is known that the eigenvalues of S_m^{-1} are (see, e.g., [14])

$$\lambda_j(S_m^{-1}) = \gamma_{m,j} = \frac{\sinh(m \cosh^{-1}(\eta_j))}{\sinh((m+1)\cosh^{-1}(\eta_j))}, \quad j = 1, 2, \ldots, N.$$

Since $e^{\cosh^{-1}(\eta_j)} = \theta_j$, we have

$$\sinh(m \cosh^{-1}(\eta_j)) = \frac{\theta_j^m - \theta_j^{-m}}{2}.$$

This yields (35). The results then follow from Theorem 3.1. ∎

REMARK 2. The studies in this subsection indicates that, although $A_{\mathbf{12}}$ in (18) has a negative identity block that is not compressible in the usual sense, after the scaling, $L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T}$ has decaying singular values and becomes reasonably compressible. This then further fits the effectiveness results in Section 2.2. It confirms that the scaling-and-compression framework can serve as a useful guideline for designing effective structured preconditioners. Instead of straightforward off-diagonal low-rank compression, it is beneficial to integrate diagonal information into off-diagonal blocks before they are compressed.

3.2. Effectiveness of 1-level SIF preconditioning

With the studies in the previous subsection, we can give concrete effectiveness estimates for the 1-level SIF preconditioner.

Corollary 3.3. *Suppose the same conditions as in Theorem 3.1 hold. Let $\tilde{\mathbf{L}}$ be the 1-level Cholesky SIF factor obtained with rank-r truncated SVD in (4). Then the eigenvalues of $\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T}$ are*

$$1, 1 \pm \sqrt{\lambda_{r+1}(S_{m_{\mathbf{1}}}^{-1})\lambda_{r+1}(S_{m_{\mathbf{2}}}^{-1})}, \ldots, 1 \pm \sqrt{\lambda_{\mathcal{N}}(S_{m_{\mathbf{1}}}^{-1})\lambda_{\mathcal{N}}(S_{m_{\mathbf{2}}}^{-1})}, \tag{36}$$

where the eigenvalue 1 has multiplicity $n-\mathcal{N}+r$ with n the order of A. If A is further from the 2D model problem, then with the same notation as in Corollary 3.2,

$$\|I - \tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T}\|_2 = \sqrt{\gamma_{m_{\mathbf{1}},r+1}\gamma_{m_{\mathbf{2}},r+1}} < \eta_{r+1} - \sqrt{\eta_{r+1}^2 - 1}, \tag{37}$$

$$\kappa(\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T}) = \frac{1+\sqrt{\gamma_{m_{\mathbf{1}},r+1}\gamma_{m_{\mathbf{2}},r+1}}}{1-\sqrt{\gamma_{m_{\mathbf{1}},r+1}\gamma_{m_{\mathbf{2}},r+1}}} < \sqrt{\frac{\eta_{r+1}+1}{\eta_{r+1}-1}}.$$

Proof. Theorems 2.1 and 3.1 yield (36). For the 2D case, since $\theta_j > 1$, we have

$$\gamma_{m,j} < \frac{\theta_j^m}{\theta_j^{m+1} + \theta_j^{-m} - \theta_j^{-m-1}} < \frac{\theta_j^m}{\theta_j^{m+1}} = \frac{1}{\theta_j}.$$

Thus,

$$\sigma_j(L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T}) = \sqrt{\gamma_{m_{\mathbf{1}},r+1}\gamma_{m_{\mathbf{2}},r+1}} < \frac{1}{\theta_j} = \eta_j - \sqrt{\eta_j^2 - 1}, \quad j = 1, 2, \ldots, N.$$

This leads to in (37). In addition,

$$\kappa(\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T}) < \frac{1+\eta_{r+1}-\sqrt{\eta_{r+1}^2-1}}{1-\eta_{r+1}+\sqrt{\eta_{r+1}^2-1}} = \sqrt{\frac{\eta_{r+1}+1}{\eta_{r+1}-1}}.$$

∎

To get more specific estimates, we suppose $m_{\mathbf{1}}$ and $m_{\mathbf{2}}$ are large enough. Then

$$\sigma_j(L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T}) \approx \frac{1}{\theta_j}, \quad \kappa(\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T}) \approx \sqrt{1 + \sin^{-2}\frac{(r+1)\pi}{2(N+1)}},$$

For sufficiently large N and small r, we have

$$\kappa(\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T}) \approx \sqrt{\frac{2(N+1)}{(r+1)\pi}}. \tag{38}$$

Thus, if $r = O(1)$, then $\kappa(\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T}) = O(\sqrt{N})$. If r is a small fraction of N, then $\kappa(\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T}) = O(1)$.

These studies give concrete estimates of effectiveness for a given truncation rank r. In another word, they show how to choose r to achieve a desired condition number $\kappa(\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T})$. For the 3D case, a bound on $\kappa(\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T})$ can be similarly derived and is omitted.

3.3. Explicit form of the 1-level SIF preconditioner

We can understand the behavior of SIF preconditioning from another perspective by looking at the actual forms of the preconditioners for the model problem.

Theorem 3.2. *Suppose the same conditions as in Theorem 3.1 hold. Let r be the truncation rank for the SVD truncation step in (4) and let $\tilde{Q}$ be matrix given by the first r columns of Q in* (22)*. Then the 1-level SIF preconditioner is*

$$\tilde{\mathbf{L}}\tilde{\mathbf{L}}^T \equiv \begin{pmatrix} A_{\mathbf{11}} & \tilde{A}_{\mathbf{12}} \\ \tilde{A}_{\mathbf{12}}^T & A_{\mathbf{22}} \end{pmatrix}, \quad \textit{with} \quad \tilde{A}_{\mathbf{12}} = \begin{pmatrix} 0 & 0 \\ -\tilde{Q}\tilde{Q}^T & 0 \end{pmatrix}. \tag{39}$$

Proof. For the lower-triangular Cholesky factor $L_{\mathbf{k}}$ of $A_{\mathbf{kk}}$ for $\mathbf{k} = \mathbf{1}, \mathbf{2}$ in (21), $L_{\mathbf{k}}^{-1}$ has the form (27). Also, the SVD of $L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T}$ is given in Corollary 3.1. Note that in the SVD of K_i in (32), V_i is orthogonal and the singular values in $\Lambda_i^{\frac{1}{2}}$ are ordered from the smallest to the largest. In the SIF scheme, we truncate the SVD of $L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T}$ in (33) by keeping the r largest singular values in $\Sigma_{\mathbf{1}}$. That is, the r smallest singular values in $\Lambda_{m_{\mathbf{1}}}^{\frac{1}{2}}$ and $\Lambda_{m_{\mathbf{2}}}^{\frac{1}{2}}$ are kept. Use $\tilde{\Lambda}_i^{\frac{1}{2}}$ to denote the leading $r \times r$ principal submatrix of $\Lambda_i^{\frac{1}{2}}$ and use $\tilde{V}_i$ to denote the singular vectors in V_i in (32) that correspond to $\tilde{\Lambda}_i^{\frac{1}{2}}$. Then in the SIF scheme, $L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T}$ is approximated by a rank-r truncated SVD as follows:

$$L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T} = U_{\mathbf{1}}\Sigma_{\mathbf{1}}U_{\mathbf{2}}^T \approx \tilde{U}_{\mathbf{1}}\tilde{\Sigma}_{\mathbf{1}}\tilde{U}_{\mathbf{2}}^T, \tag{40}$$

where

$$\tilde{U}_{\mathbf{1}} = \begin{pmatrix} 0 \\ \tilde{V}_{m_{\mathbf{1}}} \end{pmatrix}, \quad \tilde{\Sigma}_{\mathbf{1}} = \tilde{\Lambda}_{m_{\mathbf{1}}}^{-\frac{1}{2}}\tilde{\Lambda}_{m_{\mathbf{2}}}^{-\frac{1}{2}},$$

$$\tilde{U}_{\mathbf{2}}^T = -\tilde{\Lambda}_{m_{\mathbf{2}}}^{\frac{1}{2}} \begin{pmatrix} \tilde{\Lambda}_1^{-\frac{1}{2}}\tilde{V}_1^T & \tilde{\Lambda}_1^{-1}\tilde{\Lambda}_2^{-\frac{1}{2}}\tilde{V}_2^T & \cdots & \tilde{\Lambda}_1^{-1}\cdots\tilde{\Lambda}_{m_{\mathbf{2}}-1}^{-1}\tilde{\Lambda}_{m_{\mathbf{2}}}^{-\frac{1}{2}}\tilde{V}_{m_2}^T \end{pmatrix}. \tag{41}$$

Accordingly, in the SIF preconditioner, $A_{\mathbf{12}} = L_{\mathbf{1}}(L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T})L_{\mathbf{2}}$ is approximated by

$$\tilde{A}_{\mathbf{12}} \equiv L_{\mathbf{1}}\tilde{U}_{\mathbf{1}}\tilde{\Sigma}_{\mathbf{1}}\tilde{U}_{\mathbf{2}}^T L_{\mathbf{2}}^T = \begin{pmatrix} 0 \\ K_{m_{\mathbf{1}}}\tilde{V}_{m_{\mathbf{1}}}\tilde{\Sigma}_{\mathbf{1}}\tilde{U}_{\mathbf{2}}^T L_{\mathbf{2}}^T \end{pmatrix}. \tag{42}$$

From (21), (32), and (41),

$$L_{\mathbf{2}}\tilde{U}_{\mathbf{2}} = -\begin{pmatrix} K_1 & & & \\ -K_1^{-T} & \ddots & & \\ & \ddots & \ddots & \\ & & -K_{m_{\mathbf{2}}-1}^{-T} & K_{m_{\mathbf{2}}} \end{pmatrix} \begin{pmatrix} \tilde{V}_1\tilde{\Lambda}_1^{-\frac{1}{2}} \\ \tilde{V}_2\tilde{\Lambda}_2^{-\frac{1}{2}}\tilde{\Lambda}_1^{-1} \\ \vdots \\ \tilde{V}_{m_{\mathbf{2}}}\tilde{\Lambda}_{m_{\mathbf{2}}}^{-\frac{1}{2}}\tilde{\Lambda}_{m_{\mathbf{2}}-1}^{-1}\cdots\tilde{\Lambda}_1^{-1} \end{pmatrix} \tilde{\Lambda}_{m_{\mathbf{2}}}^{\frac{1}{2}}.$$

Notice that for $i = 1, \ldots, m_{\mathbf{2}}$,

$$K_i\tilde{V}_i = Q\Lambda_i^{\frac{1}{2}}V_i^T\tilde{V}_i = Q\begin{pmatrix} \tilde{\Lambda}_i^{\frac{1}{2}} \\ 0 \end{pmatrix}, \quad K_1^{-T}\tilde{V}_i = Q\Lambda_i^{-\frac{1}{2}}V_i^T\tilde{V}_i = Q\begin{pmatrix} \tilde{\Lambda}_i^{-\frac{1}{2}} \\ 0 \end{pmatrix}.$$

Then for $i = 2, \ldots, m_{\mathbf{2}}$,

$$\begin{aligned} &- K_{i-1}^{-T}\tilde{V}_{i-1}\tilde{\Lambda}_{i-1}^{-\frac{1}{2}}\tilde{\Lambda}_{i-2}^{-1}\cdots\tilde{\Lambda}_1^{-1} + K_i\tilde{V}_i\tilde{\Lambda}_i^{-\frac{1}{2}}\tilde{\Lambda}_{i-1}^{-1}\cdots\tilde{\Lambda}_1^{-1} \\ =& - Q\begin{pmatrix} I \\ 0 \end{pmatrix}\tilde{\Lambda}_{i-1}^{-1}\cdots\tilde{\Lambda}_1^{-1} + Q\begin{pmatrix} I \\ 0 \end{pmatrix}\tilde{\Lambda}_{i-1}^{-1}\cdots\tilde{\Lambda}_1^{-1} = 0. \end{aligned}$$

Thus,

$$L_{\mathbf{2}}\tilde{U}_{\mathbf{2}} = \begin{pmatrix} K_1\tilde{V}_1\tilde{\Lambda}_1^{-\frac{1}{2}} \\ 0 \\ \vdots \\ 0 \end{pmatrix}\tilde{\Lambda}_{m_{\mathbf{2}}}^{\frac{1}{2}} = -\begin{pmatrix} Q\begin{pmatrix} \tilde{\Lambda}_i^{\frac{1}{2}} \\ 0 \end{pmatrix}\tilde{\Lambda}_i^{-\frac{1}{2}} \\ 0 \\ \vdots \\ 0 \end{pmatrix}\tilde{\Lambda}_{m_{\mathbf{2}}}^{\frac{1}{2}} = -\begin{pmatrix} \tilde{Q} \\ 0 \end{pmatrix}\tilde{\Lambda}_{m_{\mathbf{2}}}^{\frac{1}{2}}.$$

Therefore, from (42),

$$\begin{aligned} \tilde{A}_{\mathbf{12}} &= \begin{pmatrix} 0 \\ -K_{m_{\mathbf{1}}}\tilde{V}_{m_{\mathbf{1}}}^T\tilde{\Sigma}_{\mathbf{1}}\tilde{\Lambda}_{m_{\mathbf{2}}}^{\frac{1}{2}}\begin{pmatrix} \tilde{Q}^T & 0 \end{pmatrix} \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ -Q\begin{pmatrix} \tilde{\Lambda}_{m_{\mathbf{1}}}^{\frac{1}{2}} \\ 0 \end{pmatrix}(\tilde{\Lambda}_{m_{\mathbf{1}}}^{-\frac{1}{2}}\tilde{\Lambda}_{m_{\mathbf{2}}}^{-\frac{1}{2}})\tilde{\Lambda}_{m_{\mathbf{2}}}^{\frac{1}{2}}\begin{pmatrix} \tilde{Q}^T & 0 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ -\tilde{Q}\tilde{Q}^T & 0 \end{pmatrix}. \end{aligned}$$

∎

Note that originally $A_{\mathbf{12}}$ has a negative identity subblock so it is not clear what rank-r truncated SVD should be used in standard off-diagonal compression. Theorem 3.2 indicates that the SIF preconditioning technique chooses $-\tilde{Q}\tilde{Q}^T$ as the truncated SVD, where $\tilde{Q}$ corresponds to the eigenspace associated with the r *smallest* eigenvalues of S_i. Such a truncation leads to the quantification of the effectiveness as in the previous subsection.

3.4. Effectiveness of multilevel SIF preconditioning

We now look at the effectiveness of multilevel SIF preconditioning for the model problem. Suppose the discretized matrix A is hierarchically partitioned, with the finest level partitioning following the index set splitting

$$\begin{aligned} \{1 : M\mathcal{N}\} = \{1 : m_{\mathbf{1}}\mathcal{N}\} &\cup \{m_{\mathbf{1}}\mathcal{N} + 1 : (m_{\mathbf{1}} + m_{\mathbf{2}})\mathcal{N}\} \cup \cdots \\ &\cup \{(m_{\mathbf{1}} + \cdots + m_{\mathbf{s}-1})\mathcal{N} + 1 : M\mathcal{N}\}, \end{aligned} \tag{43}$$

where $m_{\mathbf{1}} + m_{\mathbf{2}} + \cdots + m_{\mathbf{s}} = M$.

We can use induction and explicit computations similar to the proof of Theorem 3.2 to show the following result. The details are omitted.

Corollary 3.4. *Suppose the multilevel SIF scheme is applied to the discretized matrix A from the 2D or 3D model problem, where A is hierarchically partitioned with the finest level partitioning following the index splitting (43). Then the resulting SIF preconditioner $\tilde{A} \equiv \tilde{\mathbf{L}}\tilde{\mathbf{L}}^T$ that is the same as A except*

$$\tilde{A}_{m_{\mathbf{k}}, m_{\mathbf{k}+1}} = \tilde{A}_{m_{\mathbf{k}+1}, m_{\mathbf{k}}} = -\tilde{Q}\tilde{Q}^T, \quad \mathbf{k} = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{s}. \tag{44}$$

Thus in the multilevel SIF scheme, the compression of any scaled off-diagonal block replace the corresponding $-I$ subblock in A by $-\tilde{Q}\tilde{Q}^T$.

We can also illustrate the effectiveness of the l-level SIF preconditioner $\tilde{\mathbf{L}}\tilde{\mathbf{L}}^T$ for the model problems based on the results in [24]. The spectral analysis is much more sophisticated, since it depends on how the singular values and singular vectors of the approximately scaled off-diagonal blocks approximate those of the exact ones. Here, we just numerically illustrate how the condition number varies when l increases. In Table I, we use the 2D model problem discretized on a 64×64 mesh and the 3D model problem discretized on a $32 \times 32 \times 32$ mesh to test $\kappa(\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T})$, and the original condition numbers $\kappa(A)$ are 1.71×10^3 and 440.69, respectively. Clearly, after l-level SIF preconditioning, all the condition numbers remain reasonably small when l increases, even for r as small as 2. In comparison, if we use standard off-diagonal compression (here, by just keeping r diagonal entries in the off-diagonal $-I$ blocks), the resulting approximation fails to be positive definite for the multilevel cases.

Table I. Condition number $\kappa(\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T})$ with A from the 2D and 3D model problems when the preconditioner $\tilde{\mathbf{L}}\tilde{\mathbf{L}}^T$ is generated with the l-level SIF scheme.

Model problem		2D			3D		
r		2	4	8	2	4	8
SIF	$l=1$	13.84	8.36	4.74	9.44	6.74	5.22
	$l=2$	15.14	8.53	4.75	11.01	7.33	5.45
	$l=3$	17.63	9.50	4.89	12.93	8.57	6.25
	$l=4$	20.19	11.15	5.54	14.60	9.97	7.37
	$l=5$	21.95	12.73	6.56	15.71	11.07	8.41
Standard	$l=1$	37.95	37.88	37.51	1.55	14.55	14.55
	$l \geq 2$	Breakdown (approximation not SPD)					

In the test, we can also refine the meshes and increase the problem sizes. The condition numbers slowly increase. For the 2D case, the increase is similar to that in (38). Numerical tests for more practical problems can be found in [7, 8, 13, 14, 24].

Note that the effectiveness results in [24] has some strict robustness requirements in order to guarantee that the multilevel SIF scheme does not break down or the approximation to A remains positive definite. In the next section, we show that such requirements are too conservative and may be relaxed.

4. ROBUSTNESS OF MULTILEVEL SIF PRECONDITIONING FOR 2D AND 3D DISCRETIZED PROBLEMS

The 1-level SIF scheme always produces a positive definite approximation to any SPD matrix A. This is not the case for multilevel SIF schemes. The generalization to multiple levels is done through recursive applications of the 1-level scheme to the diagonal blocks in the hierarchical partition of A. For convenience, we organize the partitioning procedure with a binary tree $\mathcal{T}$. The matrix A is partitioned hierarchically according to the nodes at each level of the tree. The leaf nodes correspond

to the individual index sets at the bottom level partitioning in (43). The index set associated with a parent node is the union of the child index sets. Thus, if a node $\mathbf{p}$ of $\mathcal{T}$ has two children $\mathbf{i}$ and $\mathbf{j}$, the corresponding diagonal block $A_{\mathbf{pp}}$ is then partitioned as

$$A_{\mathbf{pp}} = \begin{pmatrix} A_{\mathbf{ii}} & A_{\mathbf{ij}} \\ A_{\mathbf{ij}}^T & A_{\mathbf{jj}} \end{pmatrix}. \tag{45}$$

In the following, we study the robustness of the l-level SIF scheme applied to A from the 2D and 3D model problems in Section 3. The 1-level SIF scheme is applied to all diagonal blocks of A like $A_{\mathbf{pp}}$ in (45). Similarly to (2), let $L_{\mathbf{i}}$ and $L_{\mathbf{j}}$ be the lower-triangular Cholesky factors of $A_{\mathbf{ii}}$ and $A_{\mathbf{jj}}$, respectively. In the multilevel scheme, $L_{\mathbf{i}}$ and $L_{\mathbf{j}}$ are further approximated by $\tilde{L}_{\mathbf{i}}$ and $\tilde{L}_{\mathbf{j}}$, respectively, which are obtained via the recursive application of the 1-level SIF scheme. Like in (3), the condition for the multilevel preconditioner to exist is that $\begin{pmatrix} I & \tilde{L}_{\mathbf{i}}^{-1} A_{\mathbf{ij}} \tilde{L}_{\mathbf{j}}^{-T} \\ \tilde{L}_{\mathbf{j}}^{-1} A_{\mathbf{ij}}^T \tilde{L}_{\mathbf{i}}^{-T} & I \end{pmatrix}$ is SPD for any pair of siblings $\mathbf{i}, \mathbf{j}$. This needs $\|\tilde{L}_{\mathbf{i}}^{-1} A_{\mathbf{ij}} \tilde{L}_{\mathbf{j}}^{-T}\|_2 < 1$, which may not hold for a general SPD matrix A. In [24], a condition $[(1+\tau)^l - 1]\kappa(A) < 1$ is used to guarantee the existence of the l-level SIF preconditioner. This condition essentially means that the condition number of A cannot be too large, the truncation rank r cannot be too small, or the number of levels l cannot be too large. Here, we would like to use the model problems to show that these requirements are too conservative.

Indeed, when A is from the 2D or 3D model problems, we show that the multilevel SIF scheme is *unconditionally robust*, i.e., it never breaks down and always produces SPD approximations to A. In fact, as a stronger result, it can be shown that $\tilde{L}_{\mathbf{i}}^{-1} A_{\mathbf{ij}} \tilde{L}_{\mathbf{j}}^{-T}$ always preserves the leading r singular values of $L_{\mathbf{i}}^{-1} A_{\mathbf{ij}} L_{\mathbf{j}}^{-T}$ when a fixed numerical rank r is used in the compression of the scaled off-diagonal blocks at all the hierarchical levels of $\mathcal{T}$. The details are as follows.

4.1. Singular values of scaled off-diagonal blocks within multilevel SIF schemes

Here, we study $\sigma_j(\tilde{L}_{\mathbf{i}}^{-1} A_{\mathbf{ij}} \tilde{L}_{\mathbf{j}}^{-T})$ in detail. The following lemma will be used.

Lemma 4.1. *Consider S_i in (19) and (23). For $k > 1$, let*

$$\tilde{S}_k = Q \begin{pmatrix} \tilde{\Lambda}_k & \\ & \bar{\Lambda}_k \end{pmatrix} Q^T,$$

where $\tilde{\Lambda}_k$ is an $r \times r$ diagonal matrix with the r smallest eigenvalues of S_k and $\bar{\Lambda}_k$ is any $(\mathcal{N} - r) \times (\mathcal{N} - r)$ diagonal matrix with diagonal entries greater than those in $\tilde{\Lambda}_k$. Also, let

$$\tilde{S}_i = T - \tilde{S}_{i-1}^{-1}, \quad i = k+1, k+2, \ldots$$

Then for $i \geq k$, the smallest r eigenvalues of $\tilde{S}_i$ are the same as those of S_i.

Proof. Clearly, the columns of Q are also the eigenvectors of each $\tilde{S}_i$ for $i \geq k$. From (22), we have

$$\tilde{S}_{k+1} = T - \tilde{S}_k^{-1} = Q\Lambda_1 Q^T - Q \begin{pmatrix} \tilde{\Lambda}_k^{-1} & \\ & \bar{\Lambda}_k^{-1} \end{pmatrix} Q^T \equiv Q \begin{pmatrix} \tilde{\Lambda}_{k+1} & \\ & \bar{\Lambda}_{k+1} \end{pmatrix} Q^T,$$

where

$$\tilde{\Lambda}_{k+1} = \operatorname{diag}\left(\lambda_j(T) - \lambda_j(\tilde{S}_k^{-1}), j = 1, \ldots, r\right),$$
$$\bar{\Lambda}_{k+1} = \operatorname{diag}\left(\lambda_j(T) - \lambda_j(\tilde{S}_k^{-1}), j = r+1, \ldots, \mathcal{N}\right).$$

Since $\lambda_j(S_k) = \lambda_j(\tilde{S}_k)$ for $j = 1, \ldots, r$, according to (19) and (23), $\lambda_j(S_{k+1}) = \lambda_j(\tilde{S}_{k+1})$ for $j = 1, \ldots, r$.

Also, the diagonal entries of $\bar{\Lambda}_{k+1}$ are greater than those of $\tilde{\Lambda}_{k+1}$ since for $j = r+1, \ldots, \mathcal{N}$,

$$\lambda_j(\tilde{S}_{k+1}) = \lambda_j(T) - \lambda_j(\tilde{S}_k^{-1}) > \lambda_r(T) - \lambda_r(\tilde{S}_k^{-1}) = \lambda_r(\tilde{S}_{k+1}).$$

Therefore, the r smallest eigenvalues of $\tilde{S}_{k+1}$ (the diagonal entries of $\tilde{\Lambda}_{k+1}$) are the same as those of S_{k+1}.

Similarly, we can extend this proof to show the result for any $i > k+1$. ∎

Now we are ready to study the singular values of the scaled off-diagonal blocks in the multilevel SIF scheme. The essential idea can be illustrated in terms of the 2-level SIF scheme.

Theorem 4.1. *Suppose the 2-level SIF scheme is applied to A from the 2D or 3D model problem, where the tree $\mathcal{T}$ for organizing the partitioning of A is a 2-level tree with $\mathbf{p}$ the root node and $\mathbf{i}$ and $\mathbf{j}$ the two children of $\mathbf{p}$. Suppose $L_{\mathbf{i}}$ and $L_{\mathbf{j}}$ are approximated by 1-level SIF factors $\tilde{L}_{\mathbf{i}}$ and $\tilde{L}_{\mathbf{j}}$, respectively. Also suppose r is the truncation rank at every SVD truncation step. Then*

$$\sigma_j(\tilde{L}_{\mathbf{i}}^{-1} A_{\mathbf{ij}} \tilde{L}_{\mathbf{j}}^{-T}) = \sigma_j(L_{\mathbf{i}}^{-1} A_{\mathbf{ij}} L_{\mathbf{j}}^{-T}) < 1, \quad j = 1, 2, \ldots, r. \tag{46}$$

Proof. To facilitate the proof, suppose $\mathcal{T}$ has the form in Figure 1. The matrix A corresponds to the root $\mathbf{p}$, and the first-level partitioning of A looks like (45). Similarly, $A_{\mathbf{ii}}$ and $A_{\mathbf{jj}}$ are further partitioned following the child nodes $\mathbf{1}, \mathbf{2}$ and $\mathbf{3}, \mathbf{4}$, respectively:

$$A_{\mathbf{ii}} \equiv \begin{pmatrix} A_{\mathbf{11}} & A_{\mathbf{12}} \\ A_{\mathbf{12}}^T & A_{\mathbf{22}} \end{pmatrix}, \quad A_{\mathbf{jj}} \equiv \begin{pmatrix} A_{\mathbf{33}} & A_{\mathbf{34}} \\ A_{\mathbf{43}}^T & A_{\mathbf{44}} \end{pmatrix}.$$

The corresponding finest level partitioning of the index set (43) for A looks like

$$\begin{aligned} \{1 : M\mathcal{N}\} = &\{1 : m_{\mathbf{1}}\mathcal{N}\} \cup \{m_{\mathbf{1}}\mathcal{N} + 1 : (m_{\mathbf{1}} + m_{\mathbf{2}})\mathcal{N}\} \\ &\cup \{(m_{\mathbf{1}} + m_{\mathbf{2}})\mathcal{N} + 1 : (m_{\mathbf{1}} + m_{\mathbf{2}} + m_{\mathbf{3}})\mathcal{N}\} \\ &\cup \{(m_{\mathbf{1}} + m_{\mathbf{2}} + m_{\mathbf{3}})\mathcal{N} + 1 : M\mathcal{N}\}, \end{aligned}$$

The off-diagonal blocks $A_{\mathbf{ij}}$, $A_{\mathbf{12}}$, and $A_{\mathbf{34}}$ have forms like in (18). In the following, we derive an analytical form for $\tilde{L}_{\mathbf{i}}^{-1} A_{\mathbf{ij}} \tilde{L}_{\mathbf{j}}^{-T}$ when $\tilde{L}_{\mathbf{i}}$ and $\tilde{L}_{\mathbf{j}}$ are 1-level SIF factors.

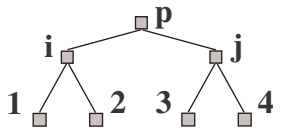


Figure 1. A two-level tree $\mathcal{T}$ for organizing the partitioning of A.

According to Theorem 3.2,

$$A_{\mathbf{ii}} \approx \tilde{A}_{\mathbf{ii}} = \begin{pmatrix} A_{\mathbf{11}} & \tilde{A}_{\mathbf{12}} \\ \tilde{A}_{\mathbf{12}}^T & A_{\mathbf{22}} \end{pmatrix},$$

where $\tilde{A}_{\mathbf{12}}$ looks like (39). The Cholesky factorization of $\tilde{A}_{\mathbf{ii}}$ has the form

$$\tilde{A}_{\mathbf{ii}} = \tilde{L}_{\mathbf{i}} \tilde{L}_{\mathbf{i}}^T, \quad \text{with} \quad \tilde{L}_{\mathbf{i}} = \begin{pmatrix} L_{\mathbf{1}} & \\ \tilde{L}_{\mathbf{i}}^{(21)} & \tilde{L}_{\mathbf{i}}^{(22)} \end{pmatrix}, \tag{47}$$

where

$$\tilde{L}_{\mathbf{i}}^{(21)} = \begin{pmatrix} 0 & -\tilde{Q}\tilde{Q}^T K_{m_1}^{-T} \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}, \ \tilde{L}_{\mathbf{i}}^{(22)} = \begin{pmatrix} \tilde{K}_{m_1+1} & & & \\ -\tilde{K}_{m_1+1}^{-T} & \ddots & & \\ & \ddots & \ddots & \\ & & -\tilde{K}_{m_1+m_2-1}^{-T} & \tilde{K}_{m_1+m_2} \end{pmatrix},$$

and $\tilde{K}_{m_1+i}$ is the Cholesky factor of $\tilde{S}_{m_1+i}$ defined by

$$\tilde{S}_{m_1+1} \equiv T - \tilde{Q}\tilde{Q}^T K_{m_1}^{-T} K_{m_1}^{-1} \tilde{Q}\tilde{Q}^T, \quad \tilde{S}_{m_1+i} \equiv T - \tilde{S}_{m_1+i-1}^{-1}, \quad i = 2, 3, \ldots, m_2. \tag{48}$$

Note that, in comparison, the computation of the exact Cholesky factor of $A_{\mathbf{ii}}$ involves

$$S_{m_1+1} = T - K_{m_1}^{-T} K_{m_1}^{-1}, \quad S_{m_1+i} \equiv T - S_{m_1+i-1}^{-1} = K_{m_1+i} K_{m_1+i}^T, \quad i = 2, 3, \ldots, m_2.$$

We can show that $\tilde{S}_{m_1+i}$ preserves the r smallest eigenvalues of S_{m_1+i} for $i \geq 1$. We first verify this for $\tilde{S}_{m_1+1}$:

$$\begin{aligned} \tilde{S}_{m_1+1} &= Q\Lambda_1 Q^T - \tilde{Q}\tilde{Q}^T Q \Lambda_{m_1}^{-1} Q^T \tilde{Q}\tilde{Q}^T \\ &= Q\Lambda_1 Q^T - \begin{pmatrix} \tilde{Q} & 0 \end{pmatrix} \Lambda_{m_1}^{-1} \begin{pmatrix} \tilde{Q}^T \\ 0 \end{pmatrix} = Q\Lambda_1 Q^T - Q \begin{pmatrix} \tilde{\Lambda}_{m_1}^{-1} & \\ & 0 \end{pmatrix} Q^T, \end{aligned} \tag{49}$$

where $\tilde{\Lambda}_{m_1}^{-1}$ contains the r largest diagonal entries of $\Lambda_{m_1}^{-1}$. Since the exact matrix S_{m_1+1} satisfies $S_{m_1+1} = Q\Lambda_1 Q^T - Q\Lambda_{m_1}^{-1} Q^T$, we can see that the r smallest eigenvalues of S_{m_1+1} are the same as those of $\tilde{S}_{m_1+1}$. Then by applying Lemma 4.1 (together with Lemma 3.1) to (48), we get that the r smallest eigenvalues of $\tilde{S}_{m_1+i}$ are the same as those of S_{m_1+i}.

Note that a form similar to (47) can be derived for $\tilde{L}_{\mathbf{j}}$, which uses matrices $\tilde{S}_{m_3+i}, i = 1, 2, \ldots, m_4$ similar to those in (48). With the same reasoning, we can get that $\tilde{S}_{m_3+i}$ preserves the r smallest eigenvalues of S_{m_3+i}. We can further obtain the forms of $\tilde{L}_{\mathbf{i}}^{-1}$ and $\tilde{L}_{\mathbf{j}}^{-1}$ similar to (27).

We are then ready to derive the singular values of $\tilde{L}_{\mathbf{i}}^{-1} A_{\mathbf{ij}} \tilde{L}_{\mathbf{j}}^{-T}$. Similarly to (28), we have

$$\tilde{L}_{\mathbf{i}}^{-1} A_{\mathbf{ji}}^T \tilde{L}_{\mathbf{j}}^{-T} = \begin{pmatrix} 0 \\ -\tilde{K}_{m_1+m_2}^{-1} \tilde{Z} \end{pmatrix},$$

where $\tilde{K}_{m_1+m_2}^{-1}$ is the lower-right block of $\tilde{L}_{\mathbf{i}}^{-1}$ and $\tilde{Z}$ is the first block row of $\tilde{L}_{\mathbf{j}}^{-T}$ with a form similar to (29):

$$\begin{aligned} \tilde{Z} = \Big(& K_1^{-T} \quad S_1^{-1} K_2^{-T} \quad \cdots \quad S_1^{-1} \cdots S_{m_3-1}^{-1} K_{m_3}^{-T} \quad \Big| \\ & S_1^{-1} \cdots S_{m_3}^{-1} \tilde{Q}\tilde{Q}^T \tilde{K}_{m_3+1}^{-T} \quad S_1^{-1} \cdots S_{m_3}^{-1} \tilde{Q}\tilde{Q}^T \tilde{S}_{m_3+1}^{-1} \tilde{K}_{m_3+2}^{-T} \\ & \cdots \quad S_1^{-1} \cdots S_{m_3}^{-1} \tilde{Q}\tilde{Q}^T \tilde{S}_{m_3+1}^{-1} \cdots \tilde{S}_{m_3+m_4-1}^{-1} \tilde{K}_{m_3+m_4}^{-T} \Big). \end{aligned}$$

Then we have

$$\begin{aligned} \sigma_j(\tilde{L}_{\mathbf{i}}^{-1} A_{\mathbf{ij}} \tilde{L}_{\mathbf{j}}^{-T}) &= \sqrt{\lambda_j(\tilde{K}_{m_1+m_2}^{-1} \tilde{Z}\tilde{Z}^T \tilde{K}_{m_1+m_2}^{-T})} \\ &= \sqrt{\lambda_j(\tilde{Z}\tilde{Z}^T \tilde{S}_{m_1+m_2}^{-1})}, \quad j = 1, 2, \ldots, \mathcal{N}. \end{aligned} \tag{50}$$

Here,

$$\begin{aligned}\tilde{Z}\tilde{Z}^T &= S_1^{-1} + S_1^{-1}S_2^{-1}S_1^{-1} + \cdots + S_1^{-1}\cdots S_{m_3-1}^{-1}S_{m_3}^{-1}S_{m_3-1}^{-1}\cdots S_1^{-1} \\ &\quad + S_1^{-1}\cdots S_{m_3}^{-1}\tilde{Q}\tilde{Q}^T(\tilde{S}_{m_3+1}^{-1} + \tilde{S}_{m_3+1}^{-1}\tilde{S}_{m_3+2}^{-1}\tilde{S}_{m_3+1}^{-1} + \cdots \\ &\quad + \tilde{S}_{m_3+1}^{-1}\cdots\tilde{S}_{m_3+m_4}^{-1}\cdots\tilde{S}_{m_3+1}^{-1})\tilde{Q}\tilde{Q}^T S_{m_3}^{-1}\cdots S_1^{-1} \\ &= Q\Big(\Lambda_1^{-1} + \Lambda_1^{-1}\Lambda_2^{-1}\Lambda_1^{-1} + \cdots + \Lambda_1^{-1}\cdots\Lambda_{m_3-1}^{-1}\Lambda_{m_3}^{-1}\Lambda_{m_3-1}^{-1}\cdots\Lambda_1^{-1} \\ &\quad + \Lambda_1^{-1}\cdots\Lambda_{m_3}^{-1}\,\mathrm{diag}(I_r,0)(\check{\Lambda}_{m_3+1}^{-1} + \check{\Lambda}_{m_3+1}^{-1}\check{\Lambda}_{m_3+2}^{-1}\check{\Lambda}_{m_3+1}^{-1} + \cdots \\ &\quad + \check{\Lambda}_{m_3+1}^{-1}\cdots\check{\Lambda}_{m_3+m_4}^{-1}\cdots\check{\Lambda}_{m_3+1}^{-1})\,\mathrm{diag}(I_r,0)\Lambda_{m_3}^{-1}\cdots\Lambda_1^{-1}\Big)Q^T,\end{aligned} \tag{51}$$

where $\check{\Lambda}_{m_3+i}$ is a diagonal matrix for the eigenvalues of $\tilde{S}_{m_3+i}$ with the eigenvalues ordered from the smallest to the largest on the diagonal. According to the discussions above, the smallest r eigenvalues of $\tilde{S}_{m_3+i}$ and S_{m_3+i} are the same for $i = 1, 2, \ldots, m_4$. According to Lemma 3.1, we get

$$\begin{aligned}\tilde{Z}\tilde{Z}^T &= Q\begin{pmatrix} \tilde{\Lambda}_{m_3+m_4}^{-1} & \\ & \mathrm{diag}(\lambda_{r+1}(S_{m_3}^{-1}),\ldots,\lambda_{\mathcal{N}}(S_{m_3}^{-1})) \end{pmatrix}Q^T \\ &= Q\,\mathrm{diag}(\lambda_1(S_{m_3+m_4}^{-1}),\ldots,\lambda_r(S_{m_3+m_4}^{-1}),\lambda_{r+1}(S_{m_3}^{-1}),\ldots,\lambda_{\mathcal{N}}(S_{m_3}^{-1}))Q^T.\end{aligned} \tag{52}$$

Note

$$\lambda_1(S_{m_3+m_4}) < \cdots < \lambda_r(S_{m_3+m_4}) < \lambda_{r+1}(S_{m_3+m_4}), \quad \lambda_{r+1}(S_{m_3}) < \cdots < \lambda_{\mathcal{N}}(S_{m_3}).$$

Also, Lemma 3.1 means $\lambda_{r+1}(S_{m_3+m_4}) < \lambda_{r+1}(S_{m_3})$. Thus, the eigenvalues on the right-hand side of (52) are ordered from the largest to the smallest, and the r largest eigenvalues of $\tilde{Z}\tilde{Z}^T$ are the same as those of $S_{m_3+m_4}^{-1}$. As discussed above, the r largest eigenvalues of $\tilde{S}_{m_1+m_2}^{-1}$ are also the same as those of $S_{m_1+m_2}^{-1}$.

Since $\lambda_j(\tilde{Z}\tilde{Z}^T\tilde{S}_{m_1+m_2}^{-1}) = \lambda_j(\tilde{Z}\tilde{Z}^T)\lambda_j(\tilde{S}_{m_1+m_2}^{-1})$, we see that the r largest eigenvalues of $\tilde{Z}\tilde{Z}^T\tilde{S}_{m_1+m_2}^{-1}$ are the same as those of $S_{m_3+m_4}^{-1}S_{m_1+m_2}^{-1}$. Therefore, we get (46) from Theorem 3.1 and (50). ∎

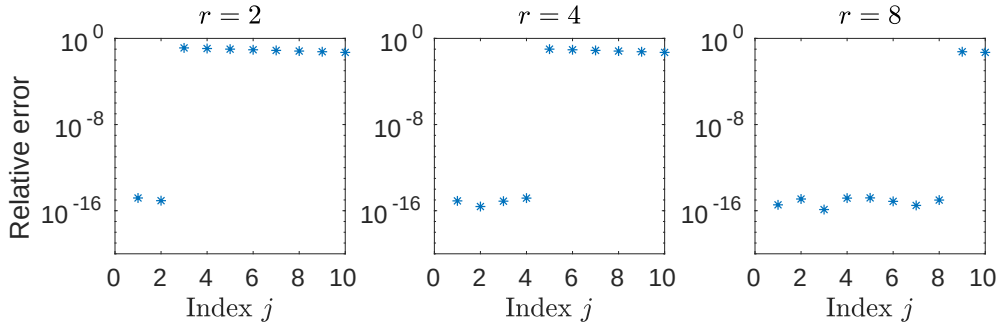
Based on Corollary 3.4, a procedure similar to the proof of Theorem 4.1 can be used to show the following result.

Corollary 4.1. *The result of Theorem 4.1 still holds if a multilevel SIF scheme is used. That is, in the l-level SIF scheme ($l > 1$) where $\tilde{L}_{\mathbf{i}}$ and $\tilde{L}_{\mathbf{j}}$ in Theorem 4.1 are $(l-1)$-level SIF factors,* (46) *is still true.*

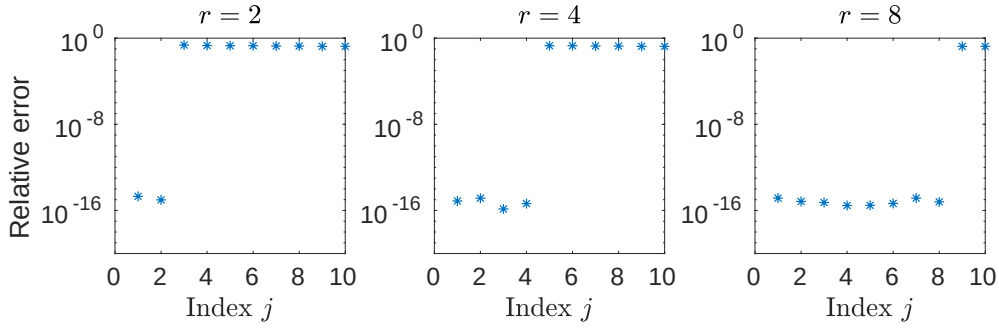
To illustrate the studies, we apply the l-level Cholesky SIF scheme to the model problems in two dimensions with a 64×64 mesh and three dimensions with a $32 \times 32 \times 32$ mesh. $l = 5$ is used. In Figure 2, we plot $\frac{|\sigma_j(\tilde{L}_{\mathbf{i}}^{-1}A_{\mathbf{ij}}\tilde{L}_{\mathbf{j}}^{-T}) - \sigma_j(L_{\mathbf{i}}^{-1}A_{\mathbf{ij}}L_{\mathbf{j}}^{-T})|}{|\sigma_j(L_{\mathbf{i}}^{-1}A_{\mathbf{ij}}L_{\mathbf{j}}^{-T})|}$ for the top level scaled off-diagonal block (where $\mathbf{i}$ and $\mathbf{j}$ are the children of the root node of $\mathcal{T}$). It can be seen that for a given truncation rank r, this errors for $j = 1, 2, \ldots, r$ are nearly in the machine precision.

4.2. Positive definiteness of multilevel SIF preconditioners

Based on the previous studies, we can claim the positive definiteness of the approximation matrix $\tilde{A}$ produced by the multilevel SIF scheme applied to A from the model problem. This can be verified from two perspectives. One is based on the explicit form of $\tilde{A}$ as in Corollary 3.4, and another is based on the singular values of the scaled off-diagonal blocks as in Corollary 4.1.



(a) *2D model problem*



(b) *3D model problem*

Figure 2. $\frac{|\sigma_j(\tilde{L}_{\mathbf{i}}^{-1}A_{\mathbf{ij}}\tilde{L}_{\mathbf{j}}^{-T})-\sigma_j(L_{\mathbf{i}}^{-1}A_{\mathbf{ij}}L_{\mathbf{j}}^{-T})|}{|\sigma_j(L_{\mathbf{i}}^{-1}A_{\mathbf{ij}}L_{\mathbf{j}}^{-T})|}$: relative errors of the leading r singular values of the top-level scaled off-diagonal block in a multilevel SIF scheme, where A is from the model problem.

Theorem 4.2. *$\tilde{A}$ as in Corollary 3.4 produced by the multilevel SIF scheme is positive definite.*

Proof. There are two ways to prove this. One way is to use the explicit form of $\tilde{A}$ in Corollary 3.4. Let $\tilde{A}^{(\mathbf{1})}$ be obtained from A with only $A_{m_{\mathbf{1}},m_{\mathbf{1}+1}}$ and $A_{m_{\mathbf{1}+1},m_{\mathbf{1}}}$ replaced by $-\tilde{Q}\tilde{Q}^T$. Partition A and $\tilde{A}^{(\mathbf{1})}$ conformably as

$$A = \begin{pmatrix} A^{(\mathbf{0};1,1)} & A^{(\mathbf{0};1,2)} \\ A^{(\mathbf{0};2,1)} & A^{(\mathbf{0};2,2)} \end{pmatrix}, \quad \tilde{A}^{(\mathbf{1})} = \begin{pmatrix} A^{(\mathbf{0};1,1)} & \tilde{A}^{(\mathbf{1};1,2)} \\ \tilde{A}^{(\mathbf{1};2,1)} & A^{(\mathbf{0};2,2)} \end{pmatrix},$$

where the diagonal blocks of A and $\tilde{A}^{(\mathbf{1})}$ are the same, $A^{(\mathbf{0};1,1)}$ has $A_{m_{\mathbf{1}},m_{\mathbf{1}}}$ at its lower right corner, and $\tilde{A}^{(\mathbf{1};1,2)}$ has $-\tilde{Q}\tilde{Q}^T$ at its lower left corner like in (39).

Then the Schur complement of $A^{(\mathbf{0};1,1)}$ in $\tilde{A}^{(\mathbf{1})}$ can be obtained from $A^{(\mathbf{0};2,2)}$ with the leading diagonal block of $A^{(\mathbf{0};2,2)}$ modified to be $\tilde{S}_{m_{\mathbf{1}}+1}$ as in (48). According to (49) and the discussions following it, the r smallest eigenvalues of $\tilde{S}_{m_{\mathbf{k}}+1}$ are the same as those of $S_{m_{\mathbf{k}}+1}$. Furthermore, the remaining $\mathcal{N}-r$ eigenvalues of $\tilde{S}_{m_{\mathbf{k}}+1}$ are the same as those of T and are larger than the corresponding ones in $S_{m_{\mathbf{k}}+1}$ because of Lemma 3.1. In another word, $\tilde{S}_{m_{\mathbf{k}}+1}$ can be written as $S_{m_{\mathbf{k}}+1}$ plus a positive definite matrix. Accordingly, the Schur complement of $A^{(\mathbf{0};1,1)}$ in $\tilde{A}^{(\mathbf{1})}$ is SPD, and $\tilde{A}^{(\mathbf{1})}$ is SPD.

If we continue to replace the blocks $\tilde{A}^{(\mathbf{1})}_{m_{\mathbf{2}},m_{\mathbf{2}+1}}$ and $\tilde{A}^{(\mathbf{1})}_{m_{\mathbf{2}+1},m_{\mathbf{2}}}$ by $-\tilde{Q}\tilde{Q}^T$ to produce a new approximation matrix $\tilde{A}^{(\mathbf{2})}$, the same procedure as above shows that this modifies the exact Schur complement $S_{m_{\mathbf{2}}+1}$ by adding a positive definite matrix to it. Thus, $\tilde{A}^{(\mathbf{2})}$ is SPD. This process

then continues for all $\mathbf{k} = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{s}$ as in (44), and the final approximation matrix $\tilde{A}$ is the SIF preconditioner and remains SPD.

Another way to prove the positive definiteness is to use Theorem 4.1 and Corollary 4.1. In the 2-level SIF scheme, following the notation in Theorem 4.1, $\tilde{A}$ has the form

$$\tilde{A} = \begin{pmatrix} \tilde{L}_{\mathbf{i}} & \\ & \tilde{L}_{\mathbf{j}} \end{pmatrix} \begin{pmatrix} I & \tilde{U}_{\mathbf{i}}\tilde{\Sigma}_{\mathbf{i}}\tilde{U}_{\mathbf{j}}^T \\ \tilde{U}_{\mathbf{i}}\tilde{\Sigma}_{\mathbf{i}}\tilde{U}_{\mathbf{j}}^T & I \end{pmatrix} \begin{pmatrix} \tilde{L}_{\mathbf{i}}^T & \\ & \tilde{L}_{\mathbf{j}}^T \end{pmatrix}, \tag{53}$$

where $\mathbf{i}$ and $\mathbf{j}$ are the children of the root of $\mathcal{T}$ and $\tilde{U}_{\mathbf{i}}\tilde{\Sigma}_{\mathbf{i}}\tilde{U}_{\mathbf{j}}^T$ is the rank-r truncated SVD of $\tilde{L}_{\mathbf{i}}^{-1}A_{\mathbf{ij}}\tilde{L}_{\mathbf{j}}^{-T}$. According to Theorem 4.1, $\|\tilde{L}_{\mathbf{i}}^{-1}A_{\mathbf{ij}}\tilde{L}_{\mathbf{j}}^{-T}\|_2 < 1$. Thus, the matrix in the middle on the right-hand side of (53) is SPD. Accordingly, $\tilde{A}$ is SPD.

Similarly, if an l-level SIF scheme is used with $\tilde{L}_{\mathbf{i}}, \tilde{L}_{\mathbf{j}}$ in (53) being $(l-1)$-level SIF factors, we can show $\tilde{A}$ is SPD by induction using Corollary 4.1. ∎

Theorem 4.2 indicates that the multilevel SIF scheme for the model problem is unconditionally robust without the restrictions in [24]. The proof of the theorem indicates that the multilevel SIF scheme has an implicit *robustness enhancement (or Schur complement compensation) effect* [10, 23]. That is, whenever a scaled off-diagonal block is compressed, a positive (semi)definite matrix is implicitly added to the Schur complement. In 1-level SIF preconditioning, this guarantees the positive definiteness of $\tilde{A}$ for any SPD matrix A, as already shown in [24]. In multilevel SIF preconditioning, it still holds for specific applications like the model problem. Even for general A, this Schur complement compensation effect can help enhance the robustness of the resulting multilevel SIF preconditioner.

5. CONCLUSIONS

This work provides new insights into SIF preconditioning that is built on the scaling-and-compression strategy. We have shown how SIF preconditioning improves the spectral properties of SPD matrices and illustrated the specific effectiveness and robustness in terms of a type of model problems. In particular, for the model problem, we derived the singular values of scaled off-diagonal blocks as well as explicit forms of the preconditioners. The results are used to show that multilevel SIF preconditioning has a robustness enhancement effect, and the resulting preconditioner for the model problem remains positive definite regardless of the number of levels and the compression accuracy. The studies confirm that the scaling-and-compression strategy is a useful technique for designing effective structured preconditioners. Our results can also work as useful tools for studying various relevant structured preconditioners. The work also gives new hints for improving SIF preconditioning. For example, a plausible direction is to construct SIF preconditioners that could further accelerate the decay of the singular values of the scaled off-diagonal blocks, which will be explored in future work.

REFERENCES

1. Agullo E, Darve E, Giraud L, Harness Y. Low-rank factorizations in data sparse hierarchical algorithms for preconditioning symmetric positive definite matrices. SIAM J. Matrix Anal. Appl. 2018; 39:1701–1725.
2. Ajiz MA, Jennings A. A robust incomplete Choleski-conjugate gradient algorithm. Internat. J. Numer. Methods Engrg. 1984; 20:949–966.
3. Benzi M, Cullum JK, Tuma M. Robust approximate inverse preconditioning for the conjugate gradient method. SIAM J. Sci. Comput. 2000; 22:1318–1332.

4. Benzi M, Tuma M. A robust incomplete factorization preconditioner for positive definite matrices. Numer. Linear Algebra Appl. 2003; 10:385–400.
5. Chandrasekaran S, Dewilde P, Gu M, Somasunderam N. On the numerical rank of the off-diagonal blocks of Schur complements of discretized elliptic PDEs. SIAM J. Matrix Anal. Appl. 2010; 31:2261–2290.
6. Chandrasekaran S, Gu M, Pals T. A fast ULV decomposition solver for hierarchically semiseparable representations. SIAM J. Matrix Anal. Appl. 2006; 28:603–622.
7. Chen C, Cambier L, Boman EG, Rajamanickam, S, Tuminaro RS, Darve E. A robust hierarchical solver for ill-conditioned systems with applications to ice sheet modeling. arXiv preprint 2018; arXiv:1811.11248.
8. Feliu-Fabá J, Ho KL, Ying L. Recursively preconditioned hierarchical interpolative factorization for elliptic partial differential equations. arXiv preprint 2018; arXiv:1808:01364.
9. Gorman C, Chávez G, Ghysels P, Mary T, Rouet F-H, Li XS. Matrix-free construction of HSS representation using adaptive randomized sampling. arXiv preprint 2018; arXiv:1810.04125.
10. Gu M, Li XS, Vassilevski P. Direction-preserving and Schur-monotonic semiseparable approximations of symmetric positive definite matrices. SIAM J. Matrix Anal. Appl. 2010; 31:2650–2664.
11. Kaporin IE. High quality preconditioning of a general symmetric positive definite matrix based on its $U^TU + U^TR + R^TU$-decomposition. Numer. Linear Algebra Appl. 1998; 5:483–509.
12. Li R, Saad Y. Divide and conquer low-rank preconditioners for symmetric matrices. SIAM J. Sci. Comput. 2013; 35:A2069–A2095.
13. Li R, Saad Y. Low-rank correction methods for algebraic domain decomposition preconditioners. SIAM J. Matrix Anal. Appl. 2017; 38:807–828.
14. Li R, Xi Y, Saad Y. Schur complement based domain decomposition preconditioners with low-rank corrections. Numer. Linear Algebra Appl. 2016; 23:706–729.
15. Liberty E, Woolfe F, Martinsson PG, Rokhlin V, Tygert M. Randomized algorithms for the low-rank approximation of matrices. Proc. Natl. Acad. Sci. USA 2007; 104:20167–20172.
16. Lin L, Lu J, Ying L. Fast construction of hierarchical matrix representation from matrix-vector multiplication. J. Comput. Phys. 2011; 230:4071–4087.
17. Liu X, Xia J, de Hoop MV. Parallel randomized and matrix-free direct solvers for large structured dense linear systems. SIAM J. Sci. Comput. 2016; 38:S508–S538.
18. Manteuffel TA. An incomplete factorization technique for positive definite linear systems, Math. Comp. 1980; 34:473–497.
19. Meijerink JA, van der Vorst HA. An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix. Math. Comp. 1977; 31:148–162.
20. Xi Y, Li R, Saad Y. An algebraic multilevel preconditioner with low-rank corrections for sparse symmetric matrices. SIAM J. Matrix Anal. Appl. 2016; 37:235–259.
21. Xi Y, Xia J, Cauley S, Balakrishnan V.Superfast and stable structured solvers for Toeplitz least squares via randomized sampling. SIAM J. Matrix Anal. Appl. 2014; 35:44–72.
22. Xia J, Chandrasekaran S, Gu M, Li XS. Fast algorithms for hierarchically semiseparable matrices, Numer. Linear Algebra Appl. 2010;17:953–976.
23. Xia J, Gu M. Robust approximate Cholesky factorization of rank-structured symmetric positive definite matrices. SIAM J. Matrix Anal. Appl. 2010; 31:2899–2920.
24. Xia J, Xin Z. Effective and robust preconditioning of general SPD matrices via structured incomplete factorization. SIAM J. Matrix Anal. Appl. 2017; 38:1298–1322.
25. Xing X, Chow E. Preserving positive definiteness in hierarchically semiseparable matrix approximations. SIAM J. Matrix Anal. Appl. 2018; 39:829–855.