Received: 22 August 2019 | Revised: 15 January 2020 | Accepted: 4 February 2020

DOI: 10.1002/nla.2294

RESEARCH ARTICLE

WILEY

Effectiveness and robustness revisited for a preconditioning technique based on structured incomplete factorization

Zixing Xin[1] | **Jianlin Xia**[1] | **Stephen Cauley**[2] | **Venkataramanan Balakrishnan**[3]

[1]Department of Mathematics, Purdue University, West Lafayette, Indiana

[2]Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard University, Charlestown, Massachusetts

[3]Case School of Engineering, Case Western Reserve University, Cleveland, Ohio,

Correspondence
Jianlin Xia, Department of Mathematics, Purdue University, West Lafayette, IN 47907.
Email: xiaj@math.purdue.edu

Funding information
National Science Foundation, Grant/Award Number: DMS-1819166

Summary

In this work, we provide new analysis for a preconditioning technique called structured incomplete factorization (SIF) for symmetric positive definite 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 two-dimensional and three-dimensional 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.

KEYWORDS

effectiveness, robustness, scaling and compression strategy, SIF preconditioning, 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. In the meantime, it is often preferable to have robust preconditioners that remain positive definite. Various types of robust preconditioners have been designed,[1–6] where some stabilization strategies are often used. A commonly used stabilization strategy is diagonal compensation for preserving positive definiteness, where positive numbers are added to diagonal entries according to the off-diagonal entries that are dropped during incomplete factorizations.

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 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 Reference 7 for SPD matrices. The technique relies on a scaling and compression strategy reformulated from an earlier article.[8] 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 References 7,8 that the resulting SIF preconditioners have some attractive features. For example, some effectiveness results can be conveniently shown for the preconditioners. The effectiveness means that the scaled off-diagonal blocks can be aggressively compressed so as to yield effective multilevel SIF preconditioners that can improve the condition numbers and eigenvalue distributions. (Here, the effectiveness estimate or quantification is generally done in terms of the condition numbers and eigenvalue distributions.) In other words, by using low-rank approximations (with very small ranks) to the scaled off-diagonal blocks, the resulting preconditioners can greatly accelerate the convergence of iterative solution. The low-rank forms make it fast to apply the preconditioners. Practical numerical tests have shown superior convergence results as compared with standard structured preconditioning based on direct off-diagonal compression.[7] The scaling and compression strategy is later also followed by a series of other work[9–12] for designing structured preconditioners for both dense and sparse matrices. Related ideas also appear in some work for preconditioning sparse matrices.[13–16]

The analysis in Reference 7 aims at general SPD matrices and ignores specific properties and backgrounds. Thus, some of the results are very conservative. For example, the analysis in Reference 7 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 Reference 7, one based on Cholesky factorizations and another based on a so-called ULV factorization.[17,18] 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 Reference 7. 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 two-dimensional (2D) and three-dimensional (3D) discretized model problems that has often been used to study some similar preconditioners in other work.[13–15] 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,[8,19] which can help enhance the robustness of the resulting preconditioners. In fact, for the model problems, we can show that the requirements in Reference 7 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 Reference 7.

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.
- 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.[7,8] 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 Reference 7 is as follows.

Consider an order-n 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

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

(Here, bold fonts are used for the subscripts in order to be consistent with later notation.) Suppose $A_{\mathbf{12}}$ is $k_{\mathbf{1}} \times k_{\mathbf{2}}$. Then A can be factorized as

$$A = \begin{pmatrix} L_{\mathbf{1}} & \\ & L_{\mathbf{2}} \end{pmatrix} \begin{pmatrix} I_{k_1} & C \\ C^T & I_{k_2} \end{pmatrix} \begin{pmatrix} L_{\mathbf{1}}^T & \\ & L_{\mathbf{2}}^T \end{pmatrix}, \quad \text{with} \quad C = L_{\mathbf{1}}^{-1} A_{\mathbf{12}} L_{\mathbf{2}}^{-T}. \tag{3}$$

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

$$\tilde{A} \equiv \begin{pmatrix} L_1 & \\ & L_2 \end{pmatrix} \begin{pmatrix} I_{k_1} & \tilde{U}_1\tilde{\Sigma}_1\tilde{U}_2^T \\ \tilde{U}_2\tilde{\Sigma}_1\tilde{U}_1^T & I_{k_2} \end{pmatrix} \begin{pmatrix} L_1^T & \\ & L_2^T \end{pmatrix}.$$

Then we get a prototype SIF preconditioner

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

where

$$\tilde{\mathbf{L}} \equiv \begin{pmatrix} L_1 & \\ L_2\tilde{U}_2\tilde{\Sigma}_1\tilde{U}_1^T & L_2\tilde{D}_2 \end{pmatrix}, \quad \text{with} \quad \tilde{D}_2\tilde{D}_2^T = I_{k_2} - \tilde{U}_2\tilde{\Sigma}_1^2\tilde{U}_2^T. \tag{6}$$

In practice, a ULV-type factorization[17,18] is used to enhance the scalability since it avoids the sequential computation of large Schur complements and uses a hierarchical scheme where local factorizations at each level can be done simultaneously. Let Q_1 be an orthogonal matrix constructed from $\tilde{U}_1$ with $\tilde{U}_1$ as the first r columns and Q_2 be constructed similarly from $\tilde{U}_2$. Then Equation (5) becomes a ULV factorization with $\tilde{\mathbf{L}}$ a ULV factor

$$\tilde{\mathbf{L}} = \begin{pmatrix} L_1 & \\ & L_2 \end{pmatrix} \begin{pmatrix} Q_1 & \\ & Q_2 \end{pmatrix} \Pi \begin{pmatrix} H & \\ & I_{n-2r} \end{pmatrix}, \tag{7}$$

where H is the lower triangular Cholesky factor of a $2r \times 2r$ matrix $\begin{pmatrix} I_r & \tilde{\Sigma}_1 \\ \tilde{\Sigma}_1 & I_r \end{pmatrix}$ and Π is a permutation matrix used to assemble $\begin{pmatrix} I_r & \tilde{\Sigma}_1 \\ \tilde{\Sigma}_1 & I_r \end{pmatrix}$ like

$$\Pi^T \begin{pmatrix} I_{k_1} & \mathrm{diag}(\tilde{\Sigma}_1, 0) \\ \mathrm{diag}(\tilde{\Sigma}_1, 0) & I_{k_2} \end{pmatrix} \Pi = \mathrm{diag}\left(\begin{pmatrix} I_r & \tilde{\Sigma}_1 \\ \tilde{\Sigma}_1 & I_r \end{pmatrix}, I_{n-2r} \right). \tag{8}$$

Generalization of the prototype preconditioner to practical multilevel schemes is also made in Reference 7. The procedure above with 1-level partitioning of A is called a 1-level (or *prototype*) *SIF scheme*. For convenience, we call the preconditioner Equation (5) with the factor in Equation (6) a 1-level Cholesky SIF preconditioner and Equation (5) with the factor in Equation (7) a 1-level ULV SIF preconditioner. The same idea may be applied to A_{11} and A_{22} to yield approximate factors $\tilde{L}_1 \approx L_1$ and $\tilde{L}_2 \approx L_2$, respectively. If $\tilde{L}_1$ and $\tilde{L}_2$ are used to replace L_1 and L_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[7] provides some analysis results on the effectiveness the 1-level Cholesky SIF preconditioner. It is shown that the preconditioned matrix has a form

$$\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T} = \begin{pmatrix} I_{k_1} & \hat{C} \\ \hat{C}^T & I_{k_2} \end{pmatrix}, \tag{9}$$

where

$$\hat{C} = \hat{U}_1\hat{\Sigma}_1\hat{U}_2^T\tilde{D}_2^{-T}, \quad \|\hat{C}\|_2 = \sigma_{r+1}(C). \tag{10}$$

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 Reference 7 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 Equation (10).

Theorem 1. *Suppose the smaller of the row and column sizes of C in Equation (3) is k. For the Cholesky SIF factor* $\tilde{\mathbf{L}}$ *in Equation (6), the Equation (9) holds with the nonzero singular values of* $\hat{C}$ *in Equation (10) given by*

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

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

$$\tilde{\mathbf{L}}^{-1} A \tilde{\mathbf{L}}^{-T} = \operatorname{diag}\left(I_{n-2(k-r)}, \begin{pmatrix} I_{k-r} & \overline{C} \\ \overline{C}^T & I_{k-r} \end{pmatrix} \right), \tag{13}$$

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

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

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

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

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

Proof. For $\tilde{\mathbf{L}}$ in Equation (6), the proof for Equation (10) Theorem 2.5 of Reference 7, already implies Equation (12). That is, any eigenvalue $\lambda(\hat{C}\hat{C}^T)$ of $\hat{C}\hat{C}^T$ satisfies

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

where the last equality in the second line is due to the Sherman-Morrison-Woodbury formula and the result $\tilde{U}_{\mathbf{2}}^T\hat{U}_{\mathbf{2}} = 0$. The eigenvalues of $\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T}$ can then be immediately obtained based on Equation (9).

For the ULV SIF factor $\tilde{\mathbf{L}}$ in Equation (7), since the matrix $\tilde{A}$ in Equation (5) remains the same as the Cholesky SIF case, the same result as above holds for the eigenvalues of $\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T}$. On the other hand, the preconditioned matrix $\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T}$ differs. It can be shown that Equation (13) holds with

$$\overline{C} = (\hat{Q}_{\mathbf{1}}^T\hat{U}_{\mathbf{1}})\hat{\Sigma}_{\mathbf{1}}(\hat{Q}_{\mathbf{2}}^T\hat{U}_{\mathbf{2}})^T,$$

where we let $Q_{\mathbf{1}} = (\tilde{U}_{\mathbf{1}}\ \hat{Q}_{\mathbf{1}})$, $Q_{\mathbf{2}} = (\tilde{U}_{\mathbf{2}}\ \hat{Q}_{\mathbf{2}})$ based on the construction of $Q_{\mathbf{1}}$ and $Q_{\mathbf{2}}$. It can be verified that $\hat{Q}_{\mathbf{1}}^T\hat{U}_{\mathbf{1}}$ and $\hat{Q}_{\mathbf{2}}^T\hat{U}_{\mathbf{2}}$ are orthogonal matrices. The singular values of $\overline{C}$ are then obvious. ∎

This theorem means that the Cholesky and ULV SIF preconditioners produce similar preconditioned matrices. The eigenvalues of the preconditioned matrices are the same. Note that if A is preconditioned with just the block diagonal preconditioner $\operatorname{diag}(A_{\mathbf{11}}, A_{\mathbf{22}})$, it is known that the preconditioned matrix has eigenvalues $1\pm\sigma_1(C), \dots, 1\pm\sigma_k(C)$ 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 as in the Cholesky or ULV SIF preconditioning, the r largest (smallest) eigenvalues $1\pm\sigma_1(C), \dots, 1\pm\sigma_r(C)$ are mapped to 1. The condition number of the preconditioned matrix becomes Equation (11). As pointed out in Reference 7, the result Equation (11) also leads to a decay magnification effect. That is, if the singular values $\sigma_i(C)$ decays slightly, 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.

In practical implementations, ULV SIF preconditioners are preferred since they avoid the computation of explicit Schur complements and have better scalability. They also mainly use orthogonal rotations (like $Q_{\mathbf{1}}, Q_{\mathbf{2}}$) in the intermediate

factorizations and have better stability. On the other hand, Cholesky SIF preconditioners are often more convenient for analysis. Thus in our later analysis, we just use the Cholesky SIF scheme.

For general l-level SIF schemes, results similar to Equation (11) can be obtained,[7] 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. Five-point and seven-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 Reference 7, 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 Equation (3) based on randomized SVDs.[20] This can help significantly reduce the cost to construct a multilevel SIF preconditioner from about $O(n^2)$ flops in Reference 7 to roughly $O(n)$ flops (with $r = O(1)$). See Remark 2 below for more details. Such randomized structured approximation is similar to the procedures used in References 21–24. (Some similar preconditioners not based on randomization have also been applied to sparse matrices.[13–15])
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.[13–15]
3. This model problem is actually a somewhat challenging problem for standard structured preconditioners like HSS ones that are 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 Equation (3) leads to reasonable decay in the singular values of the scaled off-diagonal blocks, as can be seen later.
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 suppose that the discretization uses the five-point stencil on an $N \times M$ mesh in two dimensions and the seven-point stencil on an $N \times N \times M$ mesh in three dimensions. Also suppose that the outermost ordering of the mesh points is in the last direction, so that A has M diagonal blocks T. In the 2D case, each diagonal block T corresponds to a one-dimensional 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 $\mathcal{N} \equiv N^2$. It is known that the diagonal block T in the 2D case has eigenvalues $2\eta_j$, where

$$\eta_j = 1 + 2\sin^2\frac{j\pi}{2(N+1)}, \quad j = 1, 2, \ldots, N. \tag{16}$$

The eigenvalues of the diagonal block T in the 3D case are $2 + 2(\eta_j + \eta_k), j, k = 1, 2, \ldots, N$. Based on these results, it is convenient to write out the eigenvalues of A (which won't be used in this article). 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 References 13–15,25, *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 References 7,10,11,14,15.

Remark 2. The computational complexity of the SIF scheme for sparse matrices can be understood as follows. In an l-level SIF scheme, the factorizations of the diagonal blocks in Equation (2) are done via an $(l-1)$-level scheme. A key operation is to compute a truncated SVD of C in Equation (3). This can be done conveniently via matrix-vector multiplications and randomized SVDs[20] in the following way. Let

$$Y = CX = L_{\mathbf{1}}^{-1}(A_{\mathbf{12}}(L_{\mathbf{2}}^{-T}X)),$$

where X is a skinny Gaussian random matrix with column size equal to the desired truncation rank r for C plus a small constant. Since $L_{\mathbf{1}}$ and $L_{\mathbf{2}}$ are structured and $A_{\mathbf{12}}$ is sparse, Y can be quickly computed via structured solutions and sparse matrix-vector multiplications. Y can be used to extract an approximate column basis matrix $\tilde{U}_{\mathbf{1}}$ for C. We then have a low-rank approximation $C \approx \tilde{U}_{\mathbf{1}}(\tilde{U}_{\mathbf{1}}^T C)$. Thus, the costs at level l of an l-level ULV SIF scheme includes the formation of Y, the compression of Y, and the computation of the Q, H matrices like in Equation (7). The total costs of constructing and applying the l-level SIF preconditioner can be counted by recursion similarly to the counts in many other hierarchical structured methods.[26] Larger l means more hierarchical levels in the SIF structured approximation. This can potentially improve the efficiency if it is expensive to directly factorize the intermediate diagonal blocks. However, larger l means more approximation levels which may reduce effectiveness of preconditioning. On the other hand, smaller l may be used if there are fast ways to factorize the intermediate diagonal blocks at a certain hierarchical level. This would help avoid too many levels of approximations in the SIF scheme. If $l = O(\log n)$, it can be shown that it costs $O(r^2 n \log^2 n)$ flops to construct the preconditioner, where n is the size of A. The cost to apply the preconditioner to a vector is $O(rn \log n)$. A larger r means better effectiveness in preconditioning but higher costs. In practice, the parameters l, r need to be carefully tuned for specific problems so as to reach the optimal total iterative solution cost. The technical details of the complexity are omitted since the algorithm design or implementation is not our focus here.

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 Equation (3) from the 1-level SIF scheme. Results on the multilevel SIF scheme will be given in Section 4.1. Suppose the partitioning in Equation (1) follows the index set

$$\{1 : M\mathcal{N}\} = \{1 : m_{\mathbf{1}}\mathcal{N}\} \cup \{m_{\mathbf{1}}\mathcal{N} + 1 : (m_{\mathbf{1}} + m_{\mathbf{2}})\mathcal{N}\}, \quad (M = m_{\mathbf{1}} + m_{\mathbf{2}}), \tag{17}$$

such that $A_{\mathbf{11}}$ corresponds to the leading $m_{\mathbf{1}}$ diagonal blocks T in A, $A_{\mathbf{22}}$ corresponds to the remaining $m_{\mathbf{2}}$ 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 Equation (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 bold fonts are associated with the partitioning of A as in Equation (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

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

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

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

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

The following lemma will be used frequently later, and shows that the eigenvalues of S_i decreases with increasing i.

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

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

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

Accordingly, Λ_i in Equation (23) satisfies

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

Proof. We prove this by induction. Equations (24) to (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 $w_i = d_1^{-1} + d_1^{-1}d_2^{-1}d_1^{-1} + \dots + d_1^{-1}\cdots d_{i-1}^{-1}d_i^{-1}d_{i-1}^{-1}\cdots d_1^{-1}$. Since $d_1 > 2$, $d_{i-1} > 1$, we have $d_i = d_1 - d_{i-1}^{-1} > 1$. Then

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

and Equation (24) holds. Also,

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

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

We are now ready to present the following theorem.

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

$$\sigma_j(L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T}) = \sqrt{\lambda_j(S_{m_1}^{-1})\lambda_j(S_{m_2}^{-1})}, \quad j = 1, 2, \dots, \mathcal{N},$$

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

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

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

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

$$L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T} = \begin{pmatrix} 0 \\ -K_{m_1}^{-1}Z \end{pmatrix}, \tag{28}$$

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

$$Z = \begin{pmatrix} K_1^{-T} & S_1^{-1}K_2^{-T} & \cdots & S_1^{-1}\cdots S_{m_2-1}^{-1}K_{m_2}^{-T} \end{pmatrix}. \tag{29}$$

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

$$\sigma_j(L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T}) = \sqrt{\lambda_{\mathcal{N}-j}(K_{m_1}^{-1}ZZ^TK_{m_1}^{-T})} = \sqrt{\lambda_{\mathcal{N}-j}(ZZ^TS_{m_1}^{-1})}, \quad j = 1, 2, \dots, \mathcal{N}. \tag{30}$$

(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.) According to Equations (19) and (23),

$$\begin{aligned} ZZ^T &= S_1^{-1} + S_1^{-1}S_2^{-1}S_1^{-1} + \cdots + S_1^{-1}\cdots S_{m_2-1}^{-1}S_{m_2}^{-1}S_{m_2-1}^{-1}\cdots S_1^{-1} \\ &= Q\left(\Lambda_1^{-1} + \Lambda_1^{-1}\Lambda_2^{-1}\Lambda_1^{-1} + \cdots + \Lambda_1^{-1}\cdots\Lambda_{m_2-1}^{-1}\Lambda_{m_2}^{-1}\Lambda_{m_2-1}^{-1}\cdots\Lambda_1^{-1}\right)Q^T \\ &= Q\Lambda_{m_2}^{-1}Q^T, \end{aligned} \tag{31}$$

where the last equality is due to Lemma 1. Thus,

$$ZZ^TS_{m_1}^{-1} = Q\Lambda_{m_2}^{-1}\Lambda_{m_1}^{-1}Q^T.$$

The result then follows from Equation (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 1. *With Q and Λ_i in Equation (23), let the full SVD of K_i in Equation (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

$$\begin{aligned} L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T} &= U_{\mathbf{1}}\Sigma_{\mathbf{1}}U_{\mathbf{2}}^T, \quad \text{with} \\ U_{\mathbf{1}} &= \begin{pmatrix} 0 \\ V_{m_1} \end{pmatrix}, \quad \Sigma_{\mathbf{1}} = \Lambda_{m_1}^{-\frac{1}{2}}\Lambda_{m_2}^{-\frac{1}{2}}, \\ U_{\mathbf{2}}^T &= -\Lambda_{m_2}^{\frac{1}{2}} \begin{pmatrix} \Lambda_1^{-\frac{1}{2}}V_1^T & \Lambda_1^{-1}\Lambda_2^{-\frac{1}{2}}V_2^T & \cdots & \Lambda_1^{-1}\cdots\Lambda_{m_2-1}^{-1}\Lambda_{m_2}^{-\frac{1}{2}}V_{m_2}^T \end{pmatrix}. \end{aligned} \tag{33}$$

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

$$\begin{aligned} -K_{m_1}^{-1}Z &= -K_{m_1}^{-1}\begin{pmatrix} K_1^{-T} & S_1^{-1}K_2^{-T} & \dots & S_1^{-1}\dots S_{m_2-1}^{-1}K_{m_2}^{-T} \end{pmatrix} \\ &= -V_{m_1}\Lambda_{m_1}^{-\frac{1}{2}}Q^T\begin{pmatrix} Q\Lambda_1^{-\frac{1}{2}}V_1^T & S_1^{-1}Q\Lambda_2^{\frac{1}{2}}V_2^T & \dots & S_1^{-1}\dots S_{m_2-1}^{-1}Q\Lambda_{m_2}^{-\frac{1}{2}}V_{m_2}^T \end{pmatrix} \\ &= -V_{m_1}\Lambda_{m_1}^{-\frac{1}{2}}\begin{pmatrix} \Lambda_1^{-\frac{1}{2}}V_1^T & \Lambda_1^{-1}\Lambda_2^{-\frac{1}{2}}V_2^T & \dots & \Lambda_1^{-1}\dots\Lambda_{m_2-1}^{-1}\Lambda_{m_2}^{-\frac{1}{2}}V_{m_2}^T \end{pmatrix} \\ &= V_{m_1}(\Lambda_{m_1}^{-\frac{1}{2}}\Lambda_{m_2}^{-\frac{1}{2}})\left[-\Lambda_{m_2}^{\frac{1}{2}}\begin{pmatrix} \Lambda_1^{-\frac{1}{2}}V_1^T & \Lambda_1^{-1}\Lambda_2^{-\frac{1}{2}}V_2^T & \dots & \Lambda_1^{-1}\dots\Lambda_{m_2-1}^{-1}\Lambda_{m_2}^{-\frac{1}{2}}V_{m_2}^T \end{pmatrix}\right]. \end{aligned}$$

Then Equation (28) yields the SVD in Equation (33), as long as $U_{\mathbf{2}}$ has orthonormal columns. In fact, according to Lemma 1,

$$U_{\mathbf{2}}^T U_{\mathbf{2}} = \Lambda_{m_2}(\Lambda_1^{-1} + \Lambda_1^{-1}\Lambda_2^{-1}\Lambda_1^{-1} + \cdots + \Lambda_1^{-1}\cdots\Lambda_{m_2-1}^{-1}\Lambda_{m_2}^{-1}\Lambda_{m_2-1}^{-1}\cdots\Lambda_1^{-1}) = I.$$

■

Based on Theorem 2, 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 2. *Suppose the same conditions as Theorem 2 hold, and furthermore, A is from the 2D model problem. Let $\theta_j = \eta_j + \sqrt{\eta_j^2 - 1}$, $j = 1, 2, \dots, N$, where η_j is given in Equation (16). Then*

$$\sigma_j(L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T}) = \sqrt{\gamma_{m_1,j}\gamma_{m_2,j}}, \quad j = 1, 2, \dots, N,$$

where

$$\gamma_{m,j} = \frac{\theta_j^m - \theta_j^{-m}}{\theta_j^{m+1} - \theta_j^{-m-1}}, \quad m = m_{\mathbf{1}}, m_{\mathbf{2}}. \tag{34}$$

Proof. η_j's in Equation (16) are the eigenvalues of $\frac{1}{2}T$. It is known that the eigenvalues of S_m^{-1} are (see, eg, Reference 15)

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

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 Equation (34). The results then follow from Theorem 2. ■

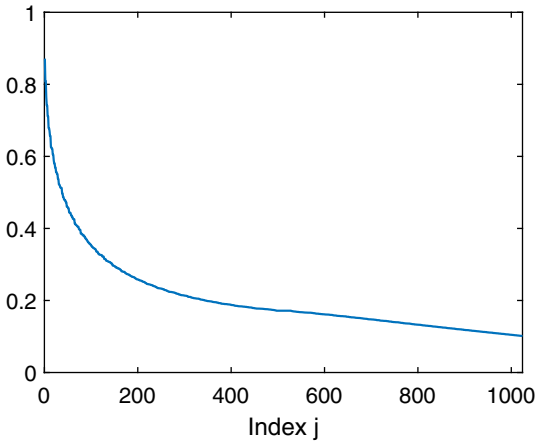


FIGURE 1 Nonzero singular values $\sigma_j(L_1^{-1}A_{12}L_2^{-T})$ for the three-dimensional model problem discretized on a $32 \times 32 \times 32$ mesh

For the 3D model problem, we just show $\sigma_j(L_1^{-1}A_{12}L_2^{-T})$ numerically with the discretization on a $32 \times 32 \times 32$ mesh. See Figure 1. We can observe reasonable decay in the nonzero singular values.

Remark 3. The studies in this subsection indicate that, although A_{12} in Equation (18) has a negative identity block that is not compressible in the usual sense, after the scaling, $L_1^{-1}A_{12}L_2^{-T}$ 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. *Suppose the same conditions as in Theorem 2 hold. Let $\tilde{\mathbf{L}}$ be the 1-level Cholesky SIF factor obtained with rank-r truncated SVD in Equation (4). Then the eigenvalues of $\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T}$ are*

$$1, 1 \pm \sqrt{\lambda_{r+1}(S_{m_1}^{-1})\lambda_{r+1}(S_{m_2}^{-1})}, \dots, 1 \pm \sqrt{\lambda_{\mathcal{N}}(S_{m_1}^{-1})\lambda_{\mathcal{N}}(S_{m_2}^{-1})}, \tag{35}$$

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

$$\begin{aligned} \|I - \tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T}\|_2 &= \sqrt{\gamma_{m_1,r+1}\gamma_{m_2,r+1}} < \eta_{r+1} - \sqrt{\eta_{r+1}^2 - 1}, \\ \kappa(\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T}) &= \frac{1 + \sqrt{\gamma_{m_1,r+1}\gamma_{m_2,r+1}}}{1 - \sqrt{\gamma_{m_1,r+1}\gamma_{m_2,r+1}}} < \sqrt{\frac{\eta_{r+1}+1}{\eta_{r+1}-1}}. \end{aligned} \tag{36}$$

Proof. Theorems 1 and 2 yield Equation (35). 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,

$$\sigma_j(L_1^{-1}A_{12}L_2^{-T}) = \sqrt{\gamma_{m_1,r+1}\gamma_{m_2,r+1}} < \frac{1}{\theta_j} = \eta_j - \sqrt{\eta_j^2 - 1}, \quad j = 1, 2, \dots, N.$$

This leads to Equation (36). 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_1 and m_2 in Equation (17) 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

$$\kappa(\tilde{\mathbf{L}}^{-1}A\tilde{\mathbf{L}}^{-T}) \approx \sqrt{\frac{2(N+1)}{(r+1)\pi}}. \tag{37}$$

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 other words, 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. *Suppose the same conditions as in Theorem 2 hold. Let r be the truncation rank for the SVD truncation step in Equation (4) and let $\tilde{Q}$ be the matrix given by the first r columns of Q in Equation (22). Then the 1-level SIF preconditioner is*

$$\tilde{\mathbf{L}}\tilde{\mathbf{L}}^T \equiv \begin{pmatrix} A_{\mathbf{11}} & \tilde{A}_{\mathbf{12}} \\ \tilde{A}_{\mathbf{12}}^T & A_{\mathbf{22}} \end{pmatrix}, \quad \text{with} \quad \tilde{A}_{\mathbf{12}} = \begin{pmatrix} 0 & 0 \\ -\tilde{Q}\tilde{Q}^T & 0 \end{pmatrix}. \tag{38}$$

Proof. For the lower triangular Cholesky factor $L_{\mathbf{k}}$ of $A_{\mathbf{kk}}$ for $\mathbf{k} = \mathbf{1}, \mathbf{2}$ in Equation (21), $L_{\mathbf{k}}^{-1}$ has the form Equation (27). Also, the SVD of $L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T}$ is given in Corollary 1. Note that in the SVD of K_i in Equation (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 Equation (33) by keeping the r largest singular values in $\Sigma_{\mathbf{1}}$. That is, the r smallest singular values in $\Lambda_{m_1}^{\frac{1}{2}}$ and $\Lambda_{m_2}^{\frac{1}{2}}$ 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 Equation (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:

$$L_{\mathbf{1}}^{-1}A_{\mathbf{12}}L_{\mathbf{2}}^{-T} = U_{\mathbf{1}}\Sigma_{\mathbf{1}}U_{\mathbf{2}}^T \approx \tilde{U}_{\mathbf{1}}\tilde{\Sigma}_{\mathbf{1}}\tilde{U}_{\mathbf{2}}^T, \tag{39}$$

where

$$\begin{aligned} \tilde{U}_{\mathbf{1}} &= \begin{pmatrix} 0 \\ \tilde{V}_{m_1} \end{pmatrix}, \quad \tilde{\Sigma}_{\mathbf{1}} = \tilde{\Lambda}_{m_1}^{-\frac{1}{2}}\tilde{\Lambda}_{m_2}^{-\frac{1}{2}}, \\ \tilde{U}_{\mathbf{2}}^T &= -\tilde{\Lambda}_{m_2}^{\frac{1}{2}}\begin{pmatrix} \tilde{\Lambda}_1^{-\frac{1}{2}}\tilde{V}_1^T & \tilde{\Lambda}_1^{-1}\tilde{\Lambda}_2^{-\frac{1}{2}}\tilde{V}_2^T & \dots & \tilde{\Lambda}_1^{-1} \dots \tilde{\Lambda}_{m_2-1}^{-1}\tilde{\Lambda}_{m_2}^{-\frac{1}{2}}\tilde{V}_{m_2}^T \end{pmatrix}. \end{aligned} \tag{40}$$

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

$$\tilde{A}_{12} \equiv L_1 \tilde{U}_1 \tilde{\Sigma}_1 \tilde{U}_2^T L_2^T = \begin{pmatrix} 0 \\ K_{m_1} \tilde{V}_{m_1} \tilde{\Sigma}_1 \tilde{U}_2^T L_2^T \end{pmatrix}. \tag{41}$$

From Equations (21), (32), and (40),

$$L_2 \tilde{U}_2 = -\begin{pmatrix} K_1 & & & \\ -K_1^{-T} & \ddots & & \\ & \ddots & \ddots & \\ & & -K_{m_2-1}^{-T} & K_{m_2} \end{pmatrix} \begin{pmatrix} \tilde{V}_1 \tilde{\Lambda}_1^{-\frac{1}{2}} \\ \tilde{V}_2 \tilde{\Lambda}_2^{-\frac{1}{2}} \tilde{\Lambda}_1^{-1} \\ \vdots \\ \tilde{V}_{m_2} \tilde{\Lambda}_{m_2}^{-\frac{1}{2}} \tilde{\Lambda}_{m_2-1}^{-1} \dots \tilde{\Lambda}_1^{-1} \end{pmatrix} \tilde{\Lambda}_{m_2}^{\frac{1}{2}}.$$

Notice that for $i = 1, \dots, m_2$,

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

Then for $i = 2, \dots, m_2$,

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

Thus,

$$L_2 \tilde{U}_2 = \begin{pmatrix} K_1 \tilde{V}_1 \tilde{\Lambda}_1^{-\frac{1}{2}} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tilde{\Lambda}_{m_2}^{\frac{1}{2}} = -\begin{pmatrix} Q \begin{pmatrix} \tilde{\Lambda}_i^{\frac{1}{2}} \\ 0 \end{pmatrix} \tilde{\Lambda}_i^{-\frac{1}{2}} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tilde{\Lambda}_{m_2}^{\frac{1}{2}} = -\begin{pmatrix} \tilde{Q} \\ 0 \end{pmatrix} \tilde{\Lambda}_{m_2}^{\frac{1}{2}}.$$

Therefore, from Equation (41),

$$\begin{aligned} \tilde{A}_{12} &= \begin{pmatrix} 0 \\ -K_{m_1} \tilde{V}_{m_1}^T \tilde{\Sigma}_1 \tilde{\Lambda}_{m_2}^{\frac{1}{2}} \begin{pmatrix} \tilde{Q}^T & 0 \end{pmatrix} \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ -Q \begin{pmatrix} \tilde{\Lambda}_{m_1}^{\frac{1}{2}} \\ 0 \end{pmatrix} (\tilde{\Lambda}_{m_1}^{-\frac{1}{2}} \tilde{\Lambda}_{m_2}^{-\frac{1}{2}}) \tilde{\Lambda}_{m_2}^{\frac{1}{2}} \begin{pmatrix} \tilde{Q}^T & 0 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ -\tilde{Q}\tilde{Q}^T & 0 \end{pmatrix}. \end{aligned}$$

■

Note that originally A_{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 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

$$\{1 : M\mathcal{N}\} = \{1 : m_1\mathcal{N}\} \cup \{m_1\mathcal{N} + 1 : (m_1 + m_2)\mathcal{N}\} \cup \cdots \cup \{(m_1 + \cdots + m_{s-1})\mathcal{N} + 1 : M\mathcal{N}\}, \tag{42}$$

where $m_1 + m_2 + \ldots + m_s = M$.

TABLE 1 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
	$l = 2$	15.76	8.61	4.75	11.46	7.61	5.56
	$l = 3$	24.12	10.89	5.03	18.98	11.52	7.64
	$l = 4$	44.32	18.01	6.76	32.65	21.23	13.32
	$l = 5$	86.64	34.05	11.59	58.86	40.07	25.58
Standard	$l = 1$	37.95	37.88	37.51	14.55	14.55	14.55
	$l \geq 2$	Breakdown (approximation not SPD)					

Abbreviations: 2D, two-dimensional; 3D, three-dimensional; SIF, structured incomplete factorization; SPD, symmetric positive definite.

We can use induction and explicit computations similar to the proof of Theorem 3 to show the following result. The details are omitted.

Corollary 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 Equation (42). Then the resulting SIF preconditioner* $\tilde{A} \equiv \tilde{\mathbf{L}}\tilde{\mathbf{L}}^T$ *that is the same as A except*

$$\tilde{A}_{m_{\mathbf{k}},m_{\mathbf{k}+1}} = \tilde{A}_{m_{\mathbf{k}+1},m_{\mathbf{k}}} = -\tilde{Q}\tilde{Q}^T, \quad \mathbf{k} = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{s}. \tag{43}$$

Thus in the multilevel SIF scheme, the compression of any scaled off-diagonal block replaces 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 Reference 7. 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 1, 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.

In the test, we can also refine the meshes and increase the problem sizes. For example, for the 2D model problem, we test $N \times N$ meshes with $N = 64, 128, 256, 512$. The right-hand side vector is obtained with the exact solution given by a vector of all ones. The preconditioner form in Corollary 4 of this section is used to avoid the impact of different implementations of the algorithm. If we let $r = 4$ and $l = 4$, the preconditioned conjugate gradient (PCG) method takes 24, 33, 44, 60 steps, respectively, to reach relative residuals smaller than 10^{-6}. If we let r be a small portion of N (ie, $N/16$) and let l increase with N, then PCG needs 24, 29, 30, 32 steps, respectively. These multilevel test results roughly follow similar patterns in the estimates for the prototype case after Corollary 3. Numerical tests for more practical problems can be found in References 7,10,11,14,15.

Note that the effectiveness results in Reference 7 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 following 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 Equation (42). 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

$$A_{\mathbf{pp}} = \begin{pmatrix} A_{\mathbf{ii}} & A_{\mathbf{ij}} \\ A_{\mathbf{ij}}^T & A_{\mathbf{jj}} \end{pmatrix}. \tag{44}$$

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 Equation (44). Similarly to Equation (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 Equation (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 Reference 7, 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*, that is, 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 2. *Consider S_i in Equations (19) and (23). For $k > 1$, let*

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

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

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

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 Equation (22), we have

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

where

$$\begin{aligned} \tilde{\Lambda}_{k+1} &= \operatorname{diag}\left(\lambda_j(T) - \lambda_j(\tilde{S}_k^{-1}), j = 1, \dots, r\right), \\ \overline{\Lambda}_{k+1} &= \operatorname{diag}\left(\lambda_j(T) - \lambda_j(\tilde{S}_k^{-1}), j = r+1, \dots, \mathcal{N}\right). \end{aligned}$$

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

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

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

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. *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 two-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*

$$\sigma_j(\tilde{L}_{\mathbf{i}}^{-1} A_{\mathbf{ij}} \tilde{L}_{\mathbf{j}}^{-T}) = \sigma_j(L_{\mathbf{i}}^{-1} A_{\mathbf{ij}} L_{\mathbf{j}}^{-T}) < 1, \quad j = 1, 2, \dots, r. \tag{45}$$

Proof. To facilitate the proof, suppose $\mathcal{T}$ has the form in Figure 2. The matrix A corresponds to the root $\mathbf{p}$, and the first-level partitioning of A looks like Equation (44). 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 Equation (42) 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 Equation (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.

According to Theorem 3,

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

where $\tilde{A}_{\mathbf{12}}$ looks like Equation (38). The Cholesky factorization of $\tilde{A}_{\mathbf{ii}}$ has the form

$$\tilde{A}_{\mathbf{ii}} = \tilde{L}_{\mathbf{i}} \tilde{L}_{\mathbf{i}}^T, \quad \text{with} \quad \tilde{L}_{\mathbf{i}} = \begin{pmatrix} L_{\mathbf{1}} & \\ \tilde{L}_{\mathbf{i}}^{(21)} & \tilde{L}_{\mathbf{i}}^{(22)} \end{pmatrix}, \tag{46}$$

where

$$\tilde{L}_{\mathbf{i}}^{(21)} = \begin{pmatrix} 0 & -\tilde{Q}\tilde{Q}^T K_{m_{\mathbf{1}}}^{-T} \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}, \quad \tilde{L}_{\mathbf{i}}^{(22)} = \begin{pmatrix} \tilde{K}_{m_{\mathbf{1}}+1} & & & \\ -\tilde{K}_{m_{\mathbf{1}}+1}^{-T} & \ddots & & \\ & \ddots & \ddots & \\ & & -\tilde{K}_{m_{\mathbf{1}}+m_{\mathbf{2}}-1}^{-T} & \tilde{K}_{m_{\mathbf{1}}+m_{\mathbf{2}}} \end{pmatrix},$$

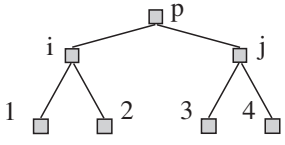


FIGURE 2 A two-level tree $\mathcal{T}$ for organizing the partitioning of A

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

$$\tilde{S}_{m_1+1} \equiv T - \tilde{Q}\tilde{Q}^T K_{m_1}^{-T} K_{m_1}^{-1} \tilde{Q}\tilde{Q}^T, \quad \tilde{S}_{m_1+i} \equiv T - \tilde{S}_{m_1+i-1}^{-1}, \quad i = 2, 3, \dots, \boldsymbol{m_2}. \tag{47}$$

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

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

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 $\tilde{S}_{m_1+1}$:

$$\begin{aligned} \tilde{S}_{m_1+1} &= Q\Lambda_1 Q^T - \tilde{Q}\tilde{Q}^T Q \Lambda_{m_1}^{-1} Q^T \tilde{Q}\tilde{Q}^T \\ &= Q\Lambda_1 Q^T - (\,\tilde{Q}\ 0\,)\Lambda_{m_1}^{-1} \begin{pmatrix} \tilde{Q}^T \\ 0 \end{pmatrix} = Q\Lambda_1 Q^T - Q \begin{pmatrix} \tilde{\Lambda}_{m_1}^{-1} & \\ & 0 \end{pmatrix} Q^T, \end{aligned} \tag{48}$$

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 2 (together with Lemma 1) to Equation (47), 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 Equation (46) can be derived for $\tilde{L}_{\mathbf{j}}$, which uses matrices $\tilde{S}_{m_3+i}, i = 1, 2, \dots, \boldsymbol{m_4}$ similar to those in Equation (47). 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 Equation (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 Equation (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 Equation (29):

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

Then we have

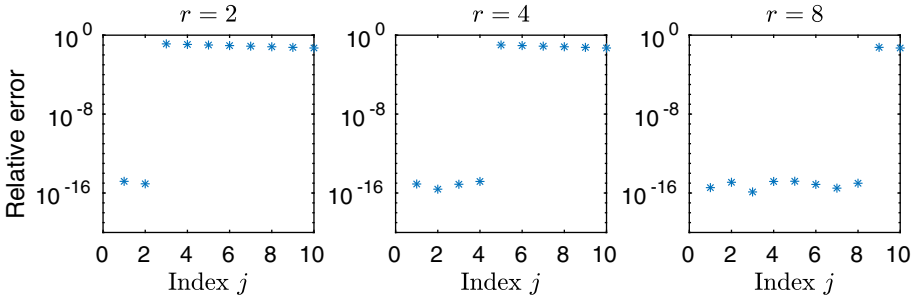
$$\begin{aligned} \sigma_j(\tilde{L}_{\mathbf{i}}^{-1} A_{\mathbf{ij}} \tilde{L}_{\mathbf{j}}^{-T}) &= \sqrt{\lambda_j(\tilde{K}_{m_1+m_2}^{-1} \tilde{Z}\tilde{Z}^T \tilde{K}_{m_1+m_2}^{-T})} \\ &= \sqrt{\lambda_j(\tilde{Z}\tilde{Z}^T \tilde{S}_{m_1+m_2}^{-1})}, \quad j = 1, 2, \dots, \mathcal{N}. \end{aligned} \tag{49}$$

Here,

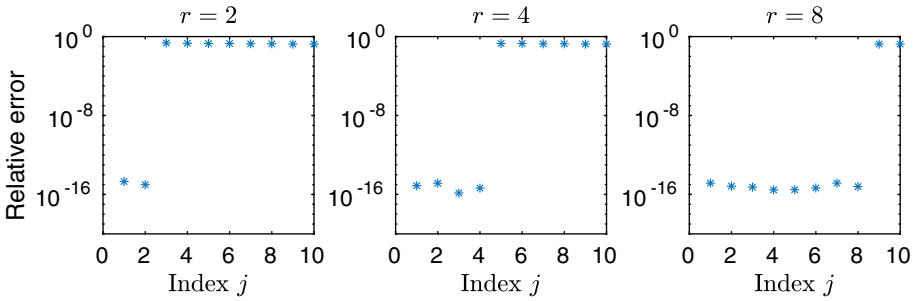
$$\begin{aligned} \tilde{Z}\tilde{Z}^T &= S_1^{-1} + S_1^{-1} S_2^{-1} S_1^{-1} + \cdots + S_1^{-1} \cdots S_{m_3-1}^{-1} S_{m_3}^{-1} S_{m_3-1}^{-1} \cdots S_1^{-1} + S_1^{-1} \cdots S_{m_3}^{-1} \tilde{Q}\tilde{Q}^T (\tilde{S}_{m_3+1}^{-1} + \tilde{S}_{m_3+1}^{-1} \tilde{S}_{m_3+2}^{-1} \tilde{S}_{m_3+1}^{-1} \\ &\quad + \cdots + \tilde{S}_{m_3+1}^{-1} \dots \tilde{S}_{m_3+m_4}^{-1} \cdots \tilde{S}_{m_3+1}^{-1}) \tilde{Q}\tilde{Q}^T S_{m_3}^{-1} \cdots S_1^{-1} \\ &= Q(\Lambda_1^{-1} + \Lambda_1^{-1}\Lambda_2^{-1}\Lambda_1^{-1} + \cdots + \Lambda_1^{-1} \cdots \Lambda_{m_3-1}^{-1} \Lambda_{m_3}^{-1} \Lambda_{m_3-1}^{-1} \cdots \Lambda_1^{-1} + \Lambda_1^{-1} \cdots \Lambda_{m_3}^{-1} \mathrm{diag}(I_r, 0)(\check{\Lambda}_{m_3+1}^{-1} + \check{\Lambda}_{m_3+1}^{-1} \check{\Lambda}_{m_3+2}^{-1} \check{\Lambda}_{m_3+1}^{-1} \\ &\quad + \cdots + \check{\Lambda}_{m_3+1}^{-1} \dots \check{\Lambda}_{m_3+m_4}^{-1} \cdots \check{\Lambda}_{m_3+1}^{-1}) \mathrm{diag}(I_r, 0) \Lambda_{m_3}^{-1} \cdots \Lambda_1^{-1}) Q^T, \end{aligned} \tag{50}$$

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 $i = 1, 2, \dots, \boldsymbol{m_4}$. According to Lemma 1, we get

$$\begin{aligned} \tilde{Z}\tilde{Z}^T &= Q \begin{pmatrix} \tilde{\Lambda}_{m_3+m_4}^{-1} & \\ & \mathrm{diag}(\lambda_{r+1}(S_{m_3}^{-1}), \dots, \lambda_{\mathcal{N}}(S_{m_3}^{-1})) \end{pmatrix} Q^T \\ &= Q\mathrm{diag}(\lambda_1(S_{m_3+m_4}^{-1}), \dots, \lambda_r(S_{m_3+m_4}^{-1}), \lambda_{r+1}(S_{m_3}^{-1}), \dots, \lambda_{\mathcal{N}}(S_{m_3}^{-1})) Q^T. \end{aligned} \tag{51}$$



(a) *2D model problem*



(b) *3D model problem*

FIGURE 3 $\frac{|\sigma_j(\tilde{L}_{\mathbf{i}}^{-1}A_{\mathbf{ij}}\tilde{L}_{\mathbf{j}}^{-T})-\sigma_j(L_{\mathbf{i}}^{-1}A_{\mathbf{ij}}L_{\mathbf{j}}^{-T})|}{|\sigma_j(L_{\mathbf{i}}^{-1}A_{\mathbf{ij}}L_{\mathbf{j}}^{-T})|}$: relative errors of the leading r singular values of the top-level scaled off-diagonal block in a multilevel structured incomplete factorization scheme, where A is from the model problem

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 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 Equation (51) 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 $S^{-1}_{m_3+m_4}$. As discussed above, the r largest eigenvalues of $\tilde{S}^{-1}_{m_1+m_2}$ are also the same as those of $S^{-1}_{m_1+m_2}$.

Since $\lambda_j(\tilde{Z}\tilde{Z}^T\tilde{S}^{-1}_{m_1+m_2}) = \lambda_j(\tilde{Z}\tilde{Z}^T)\lambda_j(\tilde{S}^{-1}_{m_1+m_2})$, we see that the r largest eigenvalues of $\tilde{Z}\tilde{Z}^T\tilde{S}^{-1}_{m_1+m_2}$ are the same as those of $S^{-1}_{m_3+m_4}S^{-1}_{m_1+m_2}$. Therefore, we get Equation (45) from Theorem 2 and Equation (49). ∎

Based on Corollary 4, a procedure similar to the proof of Theorem 4 can be used to show the following result.

Corollary 5. *The result of Theorem* 4 *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 *are* $(l-1)$*-level SIF factors, Equation (*45*) 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 3, 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 **i** and **j** are the children of the root node of $\mathcal{T}$). It can be seen that for a given truncation rank r, these errors for $j = 1, 2, \ldots, r$ are near 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 4, and another is based on the singular values of the scaled off-diagonal blocks as in Corollary 5.

Theorem 5. *$\tilde{A}$ as in Corollary 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 4. Let $\tilde{A}^{(\mathbf{1})}$ be obtained from A with only $A_{m_1,m_{1+1}}$ and A_{m_{1+1},m_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_1,m_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 Equation (38).

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 $\tilde{S}_{m_1+1}$ as in Equation (47). According to Equation (48) 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 1. In other words, $\tilde{S}_{m_\mathbf{k}+1}$ 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 $\tilde{A}^{(\mathbf{1})}_{m_2,m_{2+1}}$ and $\tilde{A}^{(\mathbf{1})}_{m_{2+1},m_2}$ 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_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 Equation (43), 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 and Corollary 5. In the 2-level SIF scheme, following the notation in Theorem 4, $\tilde{A}$ has the form

$$\tilde{A} = \begin{pmatrix} \tilde{L}_\mathbf{i} & \\ & \tilde{L}_\mathbf{j} \end{pmatrix} \begin{pmatrix} I & \tilde{U}_\mathbf{i}\tilde{\Sigma}_\mathbf{i}\tilde{U}_\mathbf{j}^T \\ \tilde{U}_\mathbf{i}\tilde{\Sigma}_\mathbf{i}\tilde{U}_\mathbf{j}^T & I \end{pmatrix} \begin{pmatrix} \tilde{L}_\mathbf{i}^T & \\ & \tilde{L}_\mathbf{j}^T \end{pmatrix}, \tag{52}$$

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, $\|\tilde{L}_\mathbf{i}^{-1}A_{\mathbf{ij}}\tilde{L}_\mathbf{j}^{-T}\|_2 < 1$. Thus, the matrix in the middle on the right-hand side of (52) 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 (52) being $(l-1)$-level SIF factors, we can show $\tilde{A}$ is SPD by induction using Corollary 5. ∎

Theorem 5 indicates that the multilevel SIF scheme for the model problem is unconditionally robust without the restrictions in Reference 7. The proof of the theorem indicates that the multilevel SIF scheme has an implicit *robustness enhancement (or Schur complement compensation) effect*.[8,19] 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 Reference 7. 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.

ACKNOWLEDGEMENTS

Thank the two anonymous referees for the helpful suggestions. The research of J.X. was supported in part by an NSF grant DMS-1819166.

CONFLICTS OF INTEREST

This work does not have any conflicts of interest.

ORCID

Jianlin Xia https://orcid.org/0000-0002-9653-9312

REFERENCES

1. Ajiz MA, Jennings A. A robust incomplete Choleski-conjugate gradient algorithm. Int J Numer Methods Eng. 1984;20:949–966.
2. Benzi M, Cullum JK, Tuma M. Robust approximate inverse preconditioning for the conjugate gradient method. SIAM J Sci Comput. 2000;22:1318–1332.
3. Benzi M, Tuma M. A robust incomplete factorization preconditioner for positive definite matrices. Numer Linear Algebra Appl. 2003;10:385–400.
4. 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.
5. Manteuffel TA. An incomplete factorization technique for positive definite linear systems. Math Comput. 1980;34:473–497.
6. Meijerink JA, van der Vorst HA. An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix. Math Comput. 1977;31:148–162.
7. 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.
8. Xia J, Gu M. Robust approximate Cholesky factorization of rank-structured symmetric positive definite matrices. SIAM J Matrix Anal Appl. 2010;31:2899–2920.
9. 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.
10. 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; 2018. arXiv preprint: arXiv:1811.11248.
11. Feliu-Fabá J, Ho KL, Ying L. Recursively preconditioned hierarchical interpolative factorization for elliptic partial differential equations; 2018. arXiv preprint: arXiv:1808:01364.
12. Xing X, Chow E. Preserving positive definiteness in hierarchically semiseparable matrix approximations. SIAM J Matrix Anal Appl. 2018;39:829–855.
13. Li R, Saad Y. Divide and conquer low-rank preconditioners for symmetric matrices. SIAM J Sci Comput. 2013;35:A2069–A2095.
14. Li R, Saad Y. Low-rank correction methods for algebraic domain decomposition preconditioners. SIAM J Matrix Anal Appl. 2017;38:807–828.
15. Li R, Xi Y, Saad Y. Schur complement based domain decomposition preconditioners with low-rank corrections. Numer Linear Algebra Appl. 2016;23:706–729.
16. 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.
17. Chandrasekaran S, Gu M, Pals T. A fast ULV decomposition solver for hierarchically semiseparable representations. SIAM J Matrix Anal Appl. 2006;28:603–622.
18. Xia J, Chandrasekaran S, Gu M, Li XS. Fast algorithms for hierarchically semiseparable matrices. Numer Linear Algebra Appl. 2010;17:953–976.
19. 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.
20. 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.
21. Gorman C, Chávez G, Ghysels P, Mary T, Rouet F-H, Li XS. Matrix-free construction of HSS representation using adaptive randomized sampling; 2018. arXiv preprint: arXiv:1810.04125.
22. Lin L, Lu J, Ying L. Fast construction of hierarchical matrix representation from matrix-vector multiplication. J Comput Phys. 2011;230:4071–4087.
23. 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.

24. 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.
25. 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.
26. Xia J. On the complexity of some hierarchical structured matrix algorithms. SIAM J Matrix Anal Appl. 2012;33:388–410.

How to cite this article: Xin Z, Xia J, Cauley S, Balakrishnan V. Effectiveness and robustness revisited for a preconditioning technique based on structured incomplete factorization. *Numer Linear Algebra Appl*. 2020;e2294. https://doi.org/10.1002/nla.2294