EFFICIENT STRUCTURED MULTIFRONTAL FACTORIZATION FOR GENERAL LARGE SPARSE MATRICES

JIANLIN XIA*

Abstract. Rank structures provide an opportunity to develop new efficient numerical methods for practical problems, when the off-diagonal blocks of certain dense intermediate matrices have small (numerical) ranks. In this work, we present a framework of structured direct factorizations for general sparse matrices, including discretized PDEs on general meshes, based on the multifrontal method and hierarchically semiseparable (HSS) matrices. We prove an idea of replacing certain complex structured operations by fast simple ones performed on compact reduced matrix forms. Such forms result from the hierarchical factorization of a tree-structured HSS matrix in a ULV scheme, so that the tree structure is reduced into a single node, the root of the original tree. This idea is shown to work in the partial ULV factorization of an HSS matrix (for quickly computing Schur complements, similar to the classical LU factorization), and also in the solution stage by substitutions. These techniques are then built into the multifrontal method for sparse factorizations after nested dissection, so as to convert the intermediate dense factorizations into fast structured ones. This method keeps certain Schur complements dense so as to avoid complicated data assembly, and is much simpler and more general than some existing methods. In particular, if the matrix arises from the discretization of certain PDEs, the factorization costs roughly $O(n)$ flops in 2D, and roughly $O(n^{5/3})$ flops or less in 3D. The solution cost and memory are nearly $O(n)$ in both cases. These counts are obtained with an idea of rank relaxation, so that this method is more generally applicable to problems where the intermediate off-diagonal ranks are not small. We demonstrate the performance of the method with 2D and 3D discretized equations, as well as various examples from a sparse matrix collection. The ideas here are also useful in future developments of fast structured solvers.

Key words. structured multifrontal method, general sparse matrix, HSS matrix, ULV factorization, reduced matrix, sparse rank relaxation

AMS subject classifications. 15A23, 65F05, 65F30, 65F50

1. Introduction. In scientific computations and engineering simulations, the major computational work is often to solve large sparse linear systems. Consider a linear system

\[
Ax = b, \ A : n \times n.
\]

Classical direct solvers for (1.1) are robust and are efficient for multiple right-hand sides. However, they are often expensive in the costs and memory, due to the creation of fill-in in the factorization. In fact, if $A$ arises from the discretization of an $N \times N$ 2D mesh, it needs at least $O\left(n^{3/2}\right)$ flops to factor $A$ [21]. For 3D, this cost is $O\left(n^2\right)$, and even the solution costs $O\left(n^{4/3}\right)$. Iterative methods can take good advantage of the sparsity with matrix-vector multiplications. But without good preconditioners, iterative methods may converge very slowly or may not even converge.

A lot of recent developments focus on structured approximate factorizations. It has been observed that some discretized PDEs and integral equations have a low-rank property. That is, the fill-in in their direct factorization has low-rank (or low-numerical-rank) off-diagonal blocks [1, 2, 9, 17, 18, 23, 24, 25, 34]. Based on this property, the fill-in can be approximated by rank structured matrices such as quasiseparable, semiseparable, or hierarchical ($H$, $H^2$) matrices (see, e.g., [3, 4, 5, 8, 13, 19, 31]

*Department of Mathematics, Purdue University, West Lafayette, IN 47907 (xiaj@math.purdue.edu). The research of Jianlin Xia was supported in part by NSF grants DMS-1115572 and CHE-0957024.
for some introductions and surveys). Examples of such methods include $\mathcal{H}$-LU methods [17, 18] and structured multifrontal methods [34]. These methods often significantly reduce classical lower complexity bounds for direct factorizations.

The structured multifrontal method in [34] employs the multifrontal method [12, 22] and hierarchically semiseparable (HSS) matrix representations [6, 7, 35], after the nested dissection ordering of $A$ [15]. For 2D elliptic equations, the method computes structured approximate factorizations with nearly linear complexity and linear storage. The major intermediate operations are performed in HSS forms, including dense factorization, block permutation, splitting, merging, etc. This thus makes the algorithm difficult to implement. The existing implementation as in [34] concentrates on symmetric positive definite (SPD) problems on 2D regular meshes. A variation of the method is given in [29]. However, this variation still mainly focuses on 2D meshes where the orientation of the mesh points roughly follow regular meshes. Also, it uses inversions for the intermediate structured matrices, followed by matrix-vector multiplications. These are often slower than ULV-type factorizations in [34].

This paper proposes a new structured multifrontal method which is much simpler and more general. It can be applied to discretized matrices in both 2D and 3D, as well as general sparse matrices. The work includes general sparse matrix reordering with graph partitioning, new improved dense structured factorizations, flexible factorization framework, and relaxed rank requirements.

First, our nested dissection reordering is performed with the aid of graph partitioning tools. Separators [15] in nested dissection are allowed to have general shapes and orientations. We also accommodate general connectivity of the separators. These make the method more generally applicable than those in [29, 34].

Second, HSS matrices are used to approximate dense intermediate fill-in, and then factored in our modified factorization scheme with improved efficiency.

Third, we organize the overall factorization in a simplified scheme. The multifrontal method converts the sparse factorization into that of a sequence of local dense ones, called frontal matrices. Partial factorizations of the frontal matrices provide certain columns of the factors, and the intermediate Schur complements are called update matrices. Here, unlike the method in [34], we approximate the frontal matrices by HSS forms, but keep the update matrices dense. This enables us to avoid complicated data assembly (called extend-add operation [22]) for the HSS form update matrices. The HSS extend-add operation in [34] requires regular meshes. Our new scheme uses the classical extend-add operation so as to accommodate more general problems.

Fourth, we show an important concept of reduced (HSS) matrices, which is not involved in [34]. That is, the ULV factorization of an HSS matrix $F$ generally results in smaller intermediate matrices (called reduced matrices) after orthogonal transformations. The original HSS matrix corresponds to a tree structure called HSS tree, which is also reduced accordingly. The overall ULV factorization leads to a final reduced matrix which is generally much smaller and corresponds to one single node of the HSS tree, its root. We prove that various complex operations involving $F$ can be replaced by simple ones performed on this final reduced matrix. The related HSS operation costs can then be reduced from, say, $O(r^2N)$ to $O(r^3)$, where $N$ is the order of $F$ and $r$ is its maximum off-diagonal rank. Such an idea is used for the fast computation of update matrices and the fast solution.

Finally, we systematically relax the classical rank requirement for structured sparse solutions. Traditional HSS operations sometimes require the related off-diagonal (numerical) ranks to be bounded in order to achieve high efficiency. In [37], it is shown
that the same or similar complexity can be obtained for HSS operations without this requirement. That is, the ranks are actually allowed to increase along the block sizes. Such rank phenomenon is indeed observed for certain sparse discretized PDEs. Thus, we generalize the dense rank relaxation idea in [37] to sparse factorizations. This brings more flexibility and broader opportunities for using structured factorizations.

With certain rank conditions which are much more flexible than the one in [34], the total factorization complexity of our new algorithm is similar to that in [34]. The costs for some 2D problems are roughly $O(n)$ flops (see Theorem 4.3). We also show that, for some 3D problems, the complexity is about $O(n^{4/3})$ or less. The storage requirement in both 2D and 3D is nearly $O(n)$. The cost of solving systems with the structured factors is also nearly $O(n)$. This type of structured approximate solvers are then very attractive for real applications such as Helmholtz equations in seismic imaging, where linear systems with a large number of right-hand sides are solved, but only modest accuracy is desired.

The remaining sections are organized as follows. HSS structures are reviewed in Section 2, followed by some improved HSS algorithms. Section 3 presents our new structured factorization and solution schemes. The algorithms are summarized and analyzed in Section 4, which shows the idea of sparse rank relaxation. Section 5 provides some numerical examples for 2D and 3D Helmholtz equations and more general examples from a classical sparse matrix collection. We draw some concluding remarks in Section 6. The following notation is used:

- $F_{i \times j}$ is the submatrix of $F$ specified by the row index set $i$ and column index set $j$.
- $F_{i \times j}$ is the submatrix formed by the columns of $F$ specified by the index set $j$.
- $b_{i \times j}$ is a vector formed by the entries of the vector $b$ from the row index set $i$.
- $\text{diag}(A_1, \ldots, A_k)$ is a block diagonal matrix with the diagonal blocks $A_1, \ldots, A_k$.
- $\text{root}(T)$ is the root of a binary tree $T$, and $\text{par}(j)$ and $\text{sib}(j)$ denote the parent and sibling nodes of node $j$ in $T$, respectively, with the nodes of $T$ labeled by $j = 1, 2, \ldots$.
- The operation $F_1 \leftrightarrow F_2$ is called an extend-add operation [12, 22], which permutes $F_1$ and $F_2$ following certain consistent index set (Section 3.2), patches zero rows and columns if necessary, and then adds the resulting matrices.

2. HSS structures and algorithms.

2.1. Review of HSS structures. HSS structures are very useful in handling dense matrices with the low-rank property. An HSS representation is generally defined recursively [6, 7]. The definition for a postordered HSS matrix is as follows [35].

Assume $F$ is an $N \times N$ dense matrix, and $\mathcal{I} = \{1 : N\} \equiv \{1, 2, \ldots, N\}$. Let $T$ be a binary tree with $k$ nodes, denoted by $j = 1, 2, \ldots, k \equiv \text{root}(T)$, and $t_j \subset \mathcal{I}$ be a set of contiguous indices associated with each node $j$ of $T$. We say $F$ is in an HSS form with the corresponding postordered HSS tree $T$ if:

1. $T$ is a full binary tree in its postordering, or, each node $j$ is either a leaf or a non-leaf node with two children $c_1$ and $c_2$ which satisfy $c_1 < c_2 < j$;
2. The index sets satisfy $t_{c_1} \cup t_{c_2} = t_j$ and $t_{c_1} \cap t_{c_2} = \emptyset$ for each non-leaf node $j$, with $t_k \equiv \mathcal{I}$;
3. For each node $j$, there exist matrices $D_j, U_j, V_j, R_j, W_j, B_j$ (called HSS gen-
ulators), which satisfy the following recursions for each non-leaf node $j$:

$$D_j \equiv F|_{t_j \times t_j} = \begin{pmatrix} D_{c_1} & U_{c_2}B_{c_2}V_{c_2}^T \\ U_{c_2}B_{c_2}V_{c_2}^T & D_{c_2} \end{pmatrix}, \quad U_j = \begin{pmatrix} U_{c_1}R_{c_1} \\ U_{c_2}R_{c_2} \end{pmatrix}, \quad V_j = \begin{pmatrix} V_{c_1}W_{c_1} \\ V_{c_2}W_{c_2} \end{pmatrix},$$

where $U_k, V_k, R_k, W_k,$ and $B_k$ are not needed (since $D_k \equiv F$ is the entire diagonal block without a corresponding off-diagonal one).

Only the $D_j, U_j, V_j$ generators associated with each leaf node $j$ of $T$ are stored. Here, $U_j$ and $V_j$ are called basis matrices, since the columns of $U_j$ and the rows of $V_j^T$ form bases of the following HSS (off-diagonal) blocks:

$$F_j^- \equiv F|_{t_j \times (T \setminus t_j)} \quad \text{and} \quad F_j^j \equiv F|_{(T \setminus t_j) \times t_j},$$

respectively. Also, we call the block row of $F$ formed by $D_j$ and $F_j^-$ the $j$-th block row of $F$. Similarly define the $j$-th block column. Clearly, a low-rank off-diagonal block used in HSS representations is an entire block row/column without the diagonal block. On the other hand, $\mathcal{H}$-matrices involve low-rank blocks more general than (2.1) and can be considered as generalizations of HSS matrices.

For example, for an $N \times N$ matrix $F$ as shown in Figure 2.1(i), we partition the set $t_7 = \{1 : N\}$ hierarchically into two levels, and assign the index sets to a binary tree (Figure 2.1(ii)) into two levels following a hierarchical partition of $A$ parent index set is the union of the children index sets. Then we can define an HSS form and an HSS tree $T$ as in Figure 2.2.

Moreover, the HSS tree can help quickly identify any off-diagonal block of the matrix [7, 35]. For example, the block corresponding to nodes 1 and 4 in Figure 2.2 can be decided by visiting the path connecting these two nodes: $1 \to 3 \to 6 \to 4$. That is

$$F|_{t_1 \times t_4} = U_1R_1B_4W_4^TV_4^T.$$

An HSS form for $F$ can be constructed with recursive compression of the HSS blocks $F_j^-$ and $F_j^j$ [7, 35]. For example, in a scalable scheme [32, 33], $F_j^-$ is compressed as follows to get the $U, R$ generators. If $j$ is a leaf of the HSS tree $T$, compute a QR factorization

$$F_j^- = U_jG_j.$$
(Note: if rank-revealing factorizations are used for compression, then the HSS form approximates $F_i$.) If $j$ is a non-leaf node with children $c_1$ and $c_2$, stack the columns of $G_{c_1}$ and $G_{c_2}$ that correspond to $F_j^{-}$, and compute a QR factorization

$$
\begin{pmatrix}
G_{c_1}|_{\times (T\setminus j)} \\
G_{c_2}|_{\times (T\setminus j)}
\end{pmatrix} = \begin{pmatrix}
R_{c_1} \\
R_{c_2}
\end{pmatrix} G_j.
$$

Then by recursion, we can verify that $F_j^{-} = U_j F|_{j, \times (T\setminus j)}$. Similarly, we compress $F_j^l$ to get the $V, W$ generators, and then extract the $B$ generators [33].

If $F$ is symmetric, only $F_j^{-}$ needs to be compressed. Then $D_j = D_j^T$, and we can set [35]

$$
V_j = U_j, \quad B_i = B_j^T,
$$

where $i = \text{sib}(j)$.

### 2.2. An improved ULV factorization and reduced HSS matrices.

An HSS linear system can be quickly solved with ULV-type factorizations and solutions [7, 35, 33]. We briefly review the original version in [7] and then show a modified one. For convenience, assume the $D_j$ blocks corresponding to all leaves $j$ of the HSS tree $T$ have sizes $m$, and all HSS blocks have ranks $r$. The scheme in [7] includes the following major steps:

1. For a leaf $j$, introduce zeros into $F_j^{-}$ by multiplying $Q_j^T$ to $F_j^{-}$, where $Q_j$ is obtained from a full QL factorization

$$
U_j = Q_j \begin{pmatrix}
0 \\
\tilde{U}_j
\end{pmatrix} \begin{pmatrix}
m-r \\
r
\end{pmatrix}.
$$

Also update the diagonal block $D_j$ to

$$
\begin{align*}
D_j = Q_j^T D_j = \begin{pmatrix}
\tilde{D}_{j;1,1} & \tilde{D}_{j;1,2} \\
\tilde{D}_{j;2,1} & \tilde{D}_{j;2,2}
\end{pmatrix} \begin{pmatrix}
m-r \\
r
\end{pmatrix}.
\end{align*}
$$

2. Compute an LQ factorization

$$
\begin{pmatrix}
\tilde{D}_{j;1,1} & \tilde{D}_{j;1,2}
\end{pmatrix} = \begin{pmatrix}
L_j & 0
\end{pmatrix} P_j.
$$
Multiply \( P_j^T \) to \( \tilde{D}_j \) and \( F_j \) on the right by updating \( \tilde{D}_j \) and \( V_j \), respectively:

\[
\tilde{D}_j P_j^T \equiv \begin{pmatrix} m-r & r \\ D_{j;2,1} & \tilde{D}_j \end{pmatrix}^{m-r} \ , \quad P_j V_j \equiv \begin{pmatrix} \tilde{V}_j \\ \hat{V}_j \end{pmatrix}^{m-r} .
\]

3. Eliminate \( L_j \) and remove node \( j \) from the tree \( T \).
4. If \( j \) is a non-leaf node with children \( c_1 \) and \( c_2 \) which has been eliminated in the previous steps, merge blocks to obtain new generators:

\[
\tilde{D}_j = \begin{pmatrix} \tilde{D}_{c_1} & \hat{U}_{c_1} B_{c_1} V_{c_1}^T \\ \hat{U}_{c_2} B_{c_2} V_{c_2}^T & \tilde{D}_{c_2} \end{pmatrix} , \quad \tilde{U}_j = \begin{pmatrix} \hat{U}_{c_1} & R_{c_1} \\ \hat{U}_{c_2} & R_{c_2} \end{pmatrix} , \quad \tilde{V}_j = \begin{pmatrix} \hat{V}_{c_1} & W_{c_1} \\ \hat{V}_{c_2} & W_{c_2} \end{pmatrix} .
\]

Then \( j \) becomes a leaf corresponding to generators \( \tilde{D}_j, \tilde{U}_j, \tilde{V}_j, B_j, \) and \( F \) is reduced to a smaller HSS matrix, called a reduced matrix.

**Definition 2.1.** In the ULV factorization of an HSS matrix \( F \), a new HSS matrix with generators in (2.3) resulting from the elimination of the children of a node \( j \) of the HSS tree is called a reduced (HSS) matrix.

The above process then repeats for the reduced matrix.

This procedure can be modified to improve the efficiency, especially to reduce the costs for dense block multiplications such as the one in (2.2). We convert each diagonal block to an identity matrix and then preserve it. A similar method has been proposed for SPD matrices [33]. Here, we handle nonsymmetric ones. The details are presented since they are needed in Theorems 3.1 and 3.3 later.

For a leaf \( j \) of \( T \), compute an LU factorization

\[
D_j = L_j T_j .
\]

Then multiply \( L_j^{-1} \) to the \( j \)-th block row of \( F \) on the left, and \( T_j^{-1} \) to the \( j \)-th block column of \( F \) on the right, so as to convert the diagonal block to an identity matrix. This is done via the update of the generators:

\[
\tilde{D}_j = I , \quad \tilde{U}_j = L_j^{-1} U_j , \quad \tilde{V}_j = T_j^{-1} V_j .
\]

See Figure 2.3(i). Next, compute a full QL factorization

\[
\tilde{U}_j = Q_j \begin{pmatrix} 0 \\ \hat{U}_j \end{pmatrix} ,
\]

and multiply \( Q_j^T \) to \( \tilde{U}_j \) on the left, so that the first \( m-r \) rows of the updated off-diagonal block row \( Q_j^T L_j^{-1} F_j \) are zeros. Also, update \( \tilde{V}_j \) to

\[
Q_j^T \tilde{V}_j \equiv \begin{pmatrix} \tilde{V}_j \\ \hat{V}_j \end{pmatrix}^{m-r} .
\]

The diagonal block becomes an identity matrix again. See Figure 2.3(ii)–(iii). After this, the diagonal identity matrix can be partially eliminated, as in Figure 2.3(iii)–(iv). Then remove node \( j \) from \( T \), which is reduced to a smaller HSS tree for a reduced HSS matrix. See Figure 2.4.

If \( j \) is a non-leaf node with its children \( c_1 \) and \( c_2 \) partially eliminated as above, we merge appropriate blocks just like (2.3), except \( \tilde{D}_{c_1} = I \) and \( \tilde{D}_{c_2} = I \) in (2.3).
(i) LU factorization of diagonal blocks

(ii) Introducing zeros and recovering identity

(iii) After introducing zeros

(iv) Reduced matrix after merging blocks

Fig. 2.3. An improved ULV factorization scheme for a nonsymmetric HSS matrix.

(i) An HSS tree with root $k$

(ii) After eliminating lower levels

(iii) Only the root left

Fig. 2.4. Elimination of the nodes of the HSS tree in the ULV HSS factorization schemes.

(Figure 2.3(iv)). This is useful in improving the efficiency of upper-level eliminations. For example, the upper level LU factorization of $D_j$ is

$$
D_j = L_j T_j \equiv \begin{pmatrix}
I & \tilde{L}_j \\
\tilde{U}_{c_1} B_{c_2} V_{c_2}^T & \tilde{T}_j
\end{pmatrix}
\begin{pmatrix}
I & \tilde{U}_{c_1} B_{c_1} \tilde{V}_{c_2}^T \\
\tilde{U}_{c_1} B_{c_1} \tilde{V}_{c_2}^T & \tilde{T}_j
\end{pmatrix},
$$

where $\tilde{L}_j, \tilde{T}_j$ is the LU factorization of the Schur complement $I - \tilde{U}_{c_2} B_{c_2} V_{c_2}^T \tilde{U}_{c_1} B_{c_1} \tilde{V}_{c_2}^T$. Similarly, (2.5) can also be accelerated due to the special form of $L_j$ and $T_j$. The elimination proceeds until the root $k$ is reached. Then we compute a direct LU factorization as in (2.4) for $j = k$. 

\[Q^T \rightarrow \]

\[L_6 \]

\[U_6 \]

\[D_6 \]

\[\sim \]

\[U_6 \sim \]

\[B_6 \]

\[\sim \]

\[V_3 \]

\[T \sim \]

\[T \sim \]
Similar to the idea in [35], it can be verified that the ULV factorization has a form
\[ H = LU, \]
where \( L \) and \( U \) are given by a sequence of block orthogonal and triangular matrices.

2.3. ULV HSS solution. The ULV HSS solution scheme follows the idea in [33], but is presented with more details here for later use.

In a forward substitution stage, we traverse the HSS tree in a bottom-up order to solve the following system:
\[
Ly = b. \tag{2.9}
\]
In the process, there is a piece of \( b \) and a piece of \( y \) associated with each node \( j \) of \( T \), denoted \( b_j \) and \( y_j \), respectively. Initially, \( b_j \equiv b_{|e_j}| \) for all leaves \( j \).

For a leaf \( j \), let
\[
\hat{b}_j = Q_j^T L_j^{-1} b_j \equiv \begin{pmatrix} \hat{b}_{j,1} \\ \hat{b}_{j,2} \end{pmatrix} m - r, \quad y_j \equiv \hat{b}_{j,1}, \quad z_j = \tilde{V}_j y_j,
\]
where \( \tilde{V}_j \) is given in (2.7). For a non-leaf node \( j \) with children \( c_1 \) and \( c_2 \), set
\[
b_j = \begin{pmatrix} \hat{b}_{c_1,2} \\ \hat{b}_{c_2,2} \end{pmatrix} - \begin{pmatrix} \tilde{U}_{c_1} B_{c_1} z_{c_1} \\ \tilde{U}_{c_2} B_{c_2} z_{c_2} \end{pmatrix}.
\]
Then we similarly compute \( y_j \) as in (2.10), except to set
\[
z_j = W_{c_1} z_{c_1} + W_{c_2} z_{c_2}.
\]
These operations can be applied recursively, until the root node \( k \) is reached, where we compute
\[
y_k = L_k^{-1} b_k. \tag{2.11}
\]
See Figure 2.5(i). Then we merge all \( y_j \) pieces with appropriate permutations to form \( y \). The details are shown in the proof of Theorem 3.3 (as in (3.25)).

In a backward substitution stage, we traverse the HSS tree in a top-down order to solve the following system:
\[
Ux = y. \tag{2.12}
\]
Initially, let
\[
x_k = T_k^{-1} y_k.
\]
For each non-leaf node \( j \) with children \( c_1 \) and \( c_2 \), partition \( y_j \) as
\[
y_j = \begin{pmatrix} y_{j,1} \\ y_{j,2} \end{pmatrix} r \]
and compute
\[
x_{c_1} = T_{c_1}^{-1} Q_{c_1} \begin{pmatrix} \tilde{y}_{c_1} \\ y_{j,1} \end{pmatrix}, \quad x_{c_2} = T_{c_2}^{-1} Q_{c_2} \begin{pmatrix} \tilde{y}_{c_2} \\ y_{j,2} \end{pmatrix}.
\]
When all the non-leaf nodes are visited, stack the pieces \( x_j \) (by setting \( x_{|e_j} \equiv x_j \)) for all the leaves \( j \) to form \( x \). See Figure 2.5(ii).
3. New structured multifrontal method. In this section, we present our new structured sparse factorization method. The HSS algorithms in the previous sections are applied to the dense intermediate matrices in a multifrontal factorization framework. For simplicity, we assume $A$ is SPD in the discussions. For nonsymmetric ones, the idea of using reduced matrices is also shown (Corollaries 3.2 and 3.5), and the algorithm can be similarly derived. Note that we do not consider pivoting issues since they are not our focus. In fact, we may use static pivoting in a preprocessing step similar to that in SuperLU [11], followed by iterative refinements in a postprocessing step.

3.1. Nested dissection and separator partitioning for general adjacency graphs. For a sparse symmetric matrix $A$, the adjacency graph has a vertex corresponding to each row/column of $A$, so that there is an edge $(i, j)$ connecting vertices $i$ and $j$ if $a_{ij} = a_{ji} \neq 0$. For a discretized matrix, the mesh can often serve as the adjacency graph. In nested dissection [15], the graph is recursively divided with separators (small sets of vertices). A top level separator divides the graph into two subregions, which are further divided recursively. See Figure 3.1. Lower level separators are ordered and eliminated before upper level ones. The elimination of a separator mutually connects its (upper level) neighbors which creates fill-in [27, 30]. (When we say the neighbors of a separator, we mean those separators which are ordered after this separator and are connected to it due to the elimination of lower level separators.) Nested dissection is very useful in reducing the fill-in. In fact, it can help the factorization of a discretized matrix in 2D or 3D achieve the theoretical lower complexity bounds tightly, in general [21].

Here in the context of our structured solution method, we allow the flexibility of using an adjacency graph or mesh which is irregular. Since the graph partition is not the major focus of this work, we only point out the following aspects:

1. Unlike [34], we can handle separators (and its neighbors) with general orientation and connectivity. We use graph partitioning tools such as METIS [26] to construct a nested dissection ordering. The ordering needs only the matrix $A$, although additional mesh information (if any) can help improve the quality of the ordering.

2. Each separator is partitioned into multiple pieces, which correspond to the partitions of appropriate HSS matrices. The partition information can be accumulated from lower levels, so as to preserve the graph connectivity.
3.2. Review of the multifrontal method. As one of the most important factorization methods, the multifrontal method [12, 22] converts a sparse factorization into a sequence of factorizations of smaller intermediate dense matrices. A tree structured called elimination tree or assembly tree is used so that these local factorizations can be done in parallel at each level. For convenience, assume the sparse matrix $A$ is SPD with the Cholesky factorization $A = LL^T$. The elimination tree $T$ has $n$ nodes corresponding to the $n$ rows/columns of $A$, and the parent of a node $i$ of $T$ is defined as

$$\text{par}(i) = \min\{j > i : L|_{j \times i} \neq 0\},$$

Let $T[i]$ be the subtree of $T$ with root $i$, nodes $c_1, c_2, \ldots, c_q$ be the children of $i$, and $\mathcal{N}_i = \{j_1, j_2, \ldots, j_d\}$ be the set of nonzero row indices in $L|_{(1:n) \times i}$. The $i$-th frontal matrix is defined to be

$$F_i = \begin{pmatrix} A|_{i \times i} & (A|_{\mathcal{N}_i \times i})^T \\ A|_{\mathcal{N}_i \times i} & 0 \end{pmatrix} - \sum_{j \in T[i]\setminus i} L|_{(i \cup \mathcal{N}_i) \times j} (L|_{(i \cup \mathcal{N}_i) \times j})^T \equiv F^0_i \oplus \mathcal{U}_{c_1} \oplus \mathcal{U}_{c_2} \oplus \cdots \oplus \mathcal{U}_{c_q},$$

where each $\mathcal{U}_{c_j}$ is called the $c_j$-th update matrix obtained from $F_{c_j}$ by recursion as in (3.2) below, and $\oplus$ denotes an extend-add operation which aligns indices and adds matrix entries. For example, assume that $\mathcal{U}_{c_1}$ and $\mathcal{U}_{c_2}$ correspond to index sets $\{3, 1\}$ and $\{3, 2\}$, respectively, and

$$\mathcal{U}_{c_1} = \begin{pmatrix} u_{11}^{(1)} & u_{12}^{(1)} \\ u_{21}^{(1)} & u_{22}^{(1)} \end{pmatrix}, \quad \mathcal{U}_{c_2} = \begin{pmatrix} u_{11}^{(2)} & u_{12}^{(2)} \\ u_{21}^{(2)} & u_{22}^{(2)} \end{pmatrix}. $$
Then
\[
U_{c_1} \leftrightarrow U_{c_2} = \begin{pmatrix} u^{(1)}_{22} & 0 & u^{(1)}_{21} \\ 0 & 0 & 0 \\ u^{(1)}_{12} & 0 & u^{(1)}_{11} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ u^{(2)}_{22} & u^{(2)}_{21} & 0 \\ 0 & u^{(2)}_{12} & u^{(2)}_{11} \end{pmatrix} = \begin{pmatrix} u^{(1)}_{22} & 0 & u^{(1)}_{21} \\ 0 & u^{(2)}_{22} & u^{(2)}_{21} \\ u^{(1)}_{12} & u^{(1)}_{11} + u^{(2)}_{12} \end{pmatrix}.
\]

Factorizing the leading entry in (3.1) yields one column of \(L\):
\[
F_i = \begin{pmatrix} \mathcal{L}_{i \times i} \\ \mathcal{L}_{N_i \times i} \end{pmatrix} \begin{pmatrix} I \\ U_i \end{pmatrix} \begin{pmatrix} (\mathcal{L}_{i \times i})^T & (\mathcal{L}_{N_i \times i})^T \\ I \end{pmatrix},
\]
where \(U_i\) is the Schur complement (the \(i\)-th update matrix). Following (3.1)–(3.2), the factorization process repeats for all the nodes of \(T\).

3.3. Structured multifrontal method. We derive a scalable structured sparse solver based on a supernodal version of the multifrontal method. That is, nested dissection in Section 3.1 is used to reorder the adjacency graph of \(A\), so that a binary assembly tree \(T\) is formed, where the separators are denoted by \(i = 1, 2, \ldots\) Each separator \(i\) is treated as a node in \(T\) and replaces the index \(i\) in Section 3.2. See Figure 3.2. This is similar to the method in [34] for 2D regular meshes, but is more general.

Let \(N_i \equiv \{j_1, j_2, \ldots, j_d\}\) be the set of neighbor separators of separator \(i\). For example, for separator \(i = 1\) in Figure 3.2, \(N_i = \{3, 7, 15\}\). Assume separator \(j\) corresponds to the index set \(t_j\) for \(A\), and \(t_j\) is the subset of \(t_i\) that is connected to separator \(i\) (due to nonzeros in \(A\) or fill-in created by earlier eliminations [20]). If \(c_1\) and \(c_2\) are the children of \(i\) in \(T\), then the frontal matrix \(F_i\) is formed by the block form extend-add operation
\[
(3.3) \quad F_i = F^0_i \leftrightarrow U_{c_1} \leftrightarrow U_{c_2}, \quad \text{with} \quad F^0_i \equiv \begin{pmatrix} A|_{t_i \times t_i} & (A|_{(\cup_{j=1}^d t_j) \times t_i})^T \\ (A|_{(\cup_{j=1}^d t_j) \times t_i})^T & 0 \end{pmatrix}.
\]

Partition \(F_i\) conformably as
\[
(3.4) \quad F_i = \begin{pmatrix} F_{i,1} & F^T_{N_i,1} \\ F_{N_i,1} & F_{N_i, N_i} \end{pmatrix}.
\]

Our structured multifrontal method has the following major steps:
1. Approximate $F_1$ by an HSS matrix.
2. Partially factor $F_1$ with a ULV factorization scheme.
3. Compute the Schur complement or update matrix $U_i$ with a low-rank update.
4. Perform the extend-add operation as in the standard multifrontal method for (3.3). The details of this can be found in [12] and are skipped here.

The first three steps are elaborated as follows. Firstly, we construction an HSS approximation to (3.4) as mentioned in Section 2. (For convenience, we assume the HSS forms are exact instead of approximate, so as to avoid using additional notation.) An HSS tree $T_i$ with $k + 2$ nodes as in Figure 3.3(i) is used, and root ($T_i$) has children $k$ and $k + 1$. That is, we assume the HSS form of $F_i$ is

\[
F_i = \begin{pmatrix}
H & U_k B_k U_k^T \\
U_{k+1} B_k^T U_k & D_{k+1}
\end{pmatrix},
\]

where $H$ is an HSS representation for $F_i$ that corresponds to the subtree $T_i[k]$, and $D_{k+1} \equiv F_{N_i N_i}$. See Figure 3.4(i). For convenience, assume the HSS rank of each $F_i$ is $r$.

Secondly, apply a symmetric ULV HSS factorization method to $H$ (a symmetric version of the original one in [7, 35] or the improved one in Section 2.2). This yields (see (3.23) below)

\[
H = L_i L_i^T.
\]

Then

\[
F_i = \begin{pmatrix}
L_i \\
U_{k+1} B_k^T U_k L_i^{-T} & I
\end{pmatrix} \begin{pmatrix} I \\ U_i \end{pmatrix} \begin{pmatrix} L_i^T & L_i^{-1} U_k B_k U_k^T \\
U_k & I
\end{pmatrix},
\]

where $U_i$ is the update matrix:

\[
U_i = D_{k+1} - (U_{k+1} B_k^T U_k) H^{-1} (U_k B_k U_k^T)
= D_{k+1} - ((L_i^{-1} U_k) B_k U_k^T)^T ((L_i^{-1} U_k) B_k U_k^T).
\]

$U_i$ is formed quickly in the next step, not directly with (3.8). After the ULV factorization of $H$, its HSS tree $T_i[k]$ is reduced to a single node $k$ with the associated generators $\tilde{D}_k$ and $\tilde{U}_k$ as in Figures 2.4(iii) and 3.3(ii). That is, the HSS form of $F_i$ is converted into a reduced matrix:

\[
\begin{pmatrix}
\tilde{D}_k & \tilde{U}_k B_k U_k^T \\
U_{k+1} B_k^T U_k \tilde{U}_k & D_{k+1}
\end{pmatrix}.
\]

The procedure is illustrated in 3.4.

Finally, consider the fast computation of $U_i$. We emphasize that $L_i$ is not explicitly available. We do not need to directly apply $L_i^{-1}$ to $U_k B_k U_k^T$ as in (3.7), neither do we compute $U_i$ by fully forming $H^{-1}$ as in (3.8). The first issue is actually addressed in the solution stage in Section 3.4 and Theorem 3.3 below, and the second issue is elaborated as follows. That is, although the ULV factorization $H = L_i L_i^T$ is not a classical triangular factorization, we can still compute its Schur complement $U_i$ conveniently. Using the notation in Section 2.2, we have the following theorem.

**Theorem 3.1.** Assume a symmetric ULV HSS factorization (3.6) for $H$ is computed, and $\tilde{D}_k$ and $\tilde{U}_k$ are the generators in the reduced matrix (3.9). Then

\[
U_k^T H^{-1} U_k = U_k^T \tilde{U}_k \tilde{D}_k^{-1} \tilde{U}_k.
\]
Therefore,

\[(3.11) \quad U_1 = D_{k+1} - \Theta_k^T \Theta_k, \quad \text{with} \quad \Theta_k = \left( L_k^{-1} \tilde{U}_k \right) B_k U_{k+1}^T,\]

where \(\tilde{D}_k = L_k L_k^T\) is the Cholesky factorization of \(\tilde{D}_k\).

Proof. We consider the improved ULV factorization in Section 2.2 only. Assume \(T_i[k]\) (the HSS tree of \(H\)) has \(l_{\max}\) levels with the root \(k\) at level 0 and the leaves at level \(l_{\max}\). For \(l = l_{\max}, l_{\max} - 1, \ldots, 1, 0\), let \(H^{(l)}\) be the reduced matrix at level \(l\) in the ULV factorization. That is, let \(H^{(l_{\max})} \equiv H\), and \(H^{(l)}\) be the reduced matrix obtained from \(H^{(l+1)}\) after the elimination of all nodes at level \(l+1\). This is formulated as follows. For convenience, we denote the children of a node \(j\) of \(T_i[k]\) by \(c_{j,1}\) and \(c_{j,2}\).

Assume the notation in Section 2.2 is specifically used for \(H\). Noticing (2.4) and
(2.6), we define

\[ U^{(l)} = \text{diag}(\tilde{U}_{j_1}, \ldots, \tilde{U}_{j_\alpha}), \quad l = l_{\max}, l_{\max} - 1, \ldots, 1, \quad U^{(0)} = \tilde{U}_k, \]

\[ R^{(l)} = \text{diag} \left( \begin{pmatrix} R_{e_{j_1,1}} & \cdots & R_{e_{j_\alpha,1}} \\ R_{e_{j_1,2}} & \cdots & R_{e_{j_\alpha,2}} \end{pmatrix} \right), \quad l = l_{\max} - 1, l_{\max} - 2, \ldots, 1, \quad R^{(0)} \equiv I, \]

\[ X^{(l)} = \text{diag} \left( \begin{pmatrix} Q^T_{e_{j_1,1}} L_{e_{j_1,1}}^{-1} & \cdots & Q^T_{e_{j_1,2}} L_{e_{j_1,2}}^{-1} \\ Q^T_{e_{j_\alpha,1}} L_{e_{j_\alpha,1}}^{-1} & \cdots & Q^T_{e_{j_\alpha,2}} L_{e_{j_\alpha,2}}^{-1} \end{pmatrix} \right), \]

\[ l = l_{\max} - 1, l_{\max} - 2, \ldots, 0, \]

where \( j_1, j_2, \ldots, j_\alpha \) are the nodes at level \( l \) of \( T_l[k] \). Clearly, the hierarchical structure of the HSS form of \( H \) means

\[ (3.12) \quad U_k = U^{(l_{\max})} \left( R^{(l_{\max} - 1)} R^{(l_{\max} - 2)} \cdots R^{(0)} \right) \equiv U^{(l_{\max})} \prod_{l=l_{\max} - 1}^{0} R^{(l)}. \]

Also let \( \Omega_l \) be a permutation matrix during the elimination at level \( l+1 \) which performs all the merging steps on \( H^{(l+1)} \) to form \( H^{(l)} \). Then the ULV factorization process can be recursively represented by

\[ (3.13) \quad \begin{pmatrix} 0 \\ U^{(l)} \end{pmatrix} = \Omega^{(l)} X^{(l)} U^{(l+1)} R^{(l)}, \quad \text{(Figure 2.3(i)–(ii))} \]

\[ (3.14) \quad \begin{pmatrix} I \\ H^{(l)} \end{pmatrix} = \Omega^{(l)} X^{(l)} H^{(l+1)} (\Omega^{(l)} X^{(l)})^T, \quad \text{(Figure 2.3(iii)–(iv))} \]

\[ l = l_{\max} - 1, l_{\max} - 2, \ldots, 1, 0. \]

For \( l = l_{\max} - 1, l_{\max} - 2, \ldots, 0 \), we have the following recursive relationship:

\[ (R^{(l)})^T \left[ \left( U^{(l+1)} \right)^T \left( H^{(l+1)} \right)^{-1} U^{(l+1)} \right] R^{(l)} \quad \text{(equation (3.14))} \]

\[ = (U^{(l+1)} R^{(l)})^T \left[ \left( \Omega^{(l)} X^{(l)} \right)^{-1} \left( \begin{pmatrix} I \\ H^{(l)} \end{pmatrix} \right) \left( \Omega^{(l)} X^{(l)} \right)^{-T} \right]^{-1} (U^{(l+1)} R^{(l)}) \]

\[ = \left( \Omega^{(l)} X^{(l)} U^{(l+1)} R^{(l)} \right)^T \left( \begin{pmatrix} I \\ H^{(l)} \end{pmatrix}^{-1} \right) \left( \Omega^{(l)} X^{(l)} U^{(l+1)} R^{(l)} \right) \quad \text{(equation (3.13))} \]

\[ = \begin{pmatrix} 0 \\ (U^{(l)})^T \end{pmatrix} \left( \begin{pmatrix} I \\ H^{(l)} \end{pmatrix}^{-1} \right) \begin{pmatrix} 0 \\ U^{(l)} \end{pmatrix} \]

\[ = (U^{(l)})^T (H^{(l)})^{-1} U^{(l)}. \]
Thus,
\[
\tilde{U}_k^T \tilde{D}_k^{-1} \tilde{U}_k = \left( U^{(0)} \right)^T \left( H^{(0)} \right)^{-1} U^{(0)}
\]
\[
= (R^{(0)})^T \left[ \left( U^{(1)} \right)^T \left( H^{(1)} \right)^{-1} U^{(1)} \right] R^{(0)}
\]
\[
= \ldots
\]
\[
= \left( \prod_{l=l_{max}-1}^{0} R^{(l)} \right)^T \left[ \left( U^{(l_{max})} \right)^T \left( H^{(l_{max})} \right)^{-1} U^{(l_{max})} \right] \left( \prod_{l=l_{max}-1}^{0} R^{(l)} \right)
\]
\[
= \left( \prod_{l=l_{max}-1}^{0} R^{(l)} \right)^T H^{-1} \left( U^{(l_{max})} \right) \left( \prod_{l=l_{max}-1}^{0} R^{(l)} \right) \quad \text{(equation (3.12))}
\]
\[
= U_k^T H^{-1} U_k.
\]

Then (3.10) holds. (3.11) follows from (3.8) and (3.10).

**Remark 3.1.** Note that \( \tilde{D}_k \) is the final reduced matrix after the ULV factorization of \( H \). This theorem indicates that, in the computation of \( U_i \), the roles of \( H \) and \( U_k \) can be replaced by those of \( \tilde{D}_k \) and \( \tilde{U}_k \), respectively. Thus, \( U_i \) can be computed quickly with a low-rank update in (3.11), since \( \tilde{D}_k \) is a much smaller matrix with size equal to the HSS rank of \( H \). In fact, if \( H \) has size \( N \) and HSS rank \( r \), then the computation of \( L_i^{-1} U_k \) in (3.8) costs \( O(r^2 N) \) flops, while the computation of \( L_k^{-1} \tilde{U}_k \) in (3.11) costs only \( O(r^3) \).

For the nonsymmetric case, we can similarly prove the following result.

**Corollary 3.2.** Assume \( H \) and its ULV HSS factorization in Theorem 3.1 are nonsymmetric. Then
\[
V_k^T H^{-1} U_k = \tilde{V}_k^T \tilde{D}_k^{-1} \tilde{U}_k.
\]
*(See (2.3) with \( j \) set to be \( k \)).* Therefore,
\[
U_i = D_{k+1}-\Theta_k^T \Phi_k, \quad \Theta_k = \left( T_k^{-T} \tilde{V}_k \right) B_k U_{k+1}^T, \quad \Phi_k = \left( L_k^{-1} \tilde{U}_k \right) B_k V_{k+1}^T.
\]

where \( \tilde{D}_k = L_k T_k \) is the LU factorization of \( \tilde{D}_k \).

After the computation of \( U_i \), it participates in the standard extend-add operation to form upper level frontal matrices. This process then proceeds along the elimination tree. When the process finishes, we have a structured sparse factorization
\[
A = L \mathcal{L}^T.
\]

If rank-revealing factorizations are used in the intermediate HSS constructions, \( \mathcal{L} \mathcal{L}^T \) approximates \( A \).

**Remark 3.2.** The structure of \( \mathcal{L} \) from the above partially structured scheme (with dense \( U_i \)) is the same as that produced by a fully structured version (with HSS \( U_i \)). According to the discussions in Section 4, the costs of these two schemes are also similar. However, by keeping \( U_i \) dense, we significantly reduce the complication of the extend-add operation. The idea of reduced matrices in Theorem 3.1 and Corollary 3.2 is also very useful in future developments of fully structured versions, where the computation of \( L_i^{-1} U_k \) in (3.8) dominates the cost.
3.4. Structured multifrontal solution. After the factorization (3.16), we solve two structured systems

$$ \mathcal{L}y = b, $$
$$ \mathcal{L}^T x = y. $$

For convenience, partition the vectors conformably into $b_i, y_i,$ and $x_i$ pieces according to the sizes of the separators in nest dissection, so that $b_i, y_i,$ and $x_i$ correspond to the variables associated with separator $i$ (node $i$ in the assembly tree $T$).

The solution of (3.17) with forward substitution involves a forward (or postordering) traversal of the assembly tree. We show the intermediate structured solution steps associated with a node $i$. According to (3.7), we need to solve intermediate systems that look like

$$ (L_i U_{k+1} B_{k+1} U_k^T L_i - T_i) (y_i \tilde{b}_{N_i}) = (b_i b_{N_i}), $$

where $b_{N_i}$ corresponds to union of $b_j$ for $j \in N_i$ (the set of neighbors of $i$). Here, $b_i$ may have been updated in the previous solution steps associated with the lower levels of $T$. See (3.19). (We still use $b_i$ for notational convenience.)

We first solve $L_i y_i = b_i$ with an ULV-type forward substitution method in Section 2.3. Then $U_{k+1} B_{k+1} U_k^T L_i - T_i y_i$ is the contribution of separator $i$ to its neighbors. That is, we update $b_{N_i}$ by

$$ b_{N_i} \leftarrow \tilde{b}_{N_i} = b_{N_i} - U_{k+1} B_{k+1} U_k^T L_i - T_i y_i. $$

Notice that the computation cost of (3.19) can be significantly reduced based on the idea of reduced matrices similar to Theorem 3.1.

**Theorem 3.3.** For node $i$ in the assembly tree $T$, denote $y_j$ in (2.10) by $y_{i,j}$, and $y_k$ in (2.11) by $y_{i,k}$, which are the solution pieces obtained by the forward substitution procedure in Section 2.3 applied to $L_i y_i = b_i$. Assume the same condition as in Theorem 3.1 holds. Then

$$ U_k^T L_i - T_i y_i = U_k^T L_k - T_k y_{i,k}. $$

Therefore, (3.19) can be computed as

$$ \tilde{b}_{N_i} = b_{N_i} - \Theta_k^T y_{i,k}, $$

where $\Theta_k$ is available from (3.11).

**Proof.** Consider $H^{(0)}$ in the proof of Theorem 3.1. Assume $H^{(0)} = L^{(0)} (L^{(0)})^T$ is the Cholesky factorization of $H^{(0)} \equiv D_k$. Then according to (3.14), we have

$$ H^{(l)} = L^{(l)} (L^{(l)})^T, $$

where $L^{(l)}$ is recursively defined as:

$$ L^{(l+1)} = (X^{(l)})^{-1} (\Omega^{(l)})^T \left( \begin{array}{c} I \\ L^{(l)} \end{array} \right) \Omega^{(l)}, \ l = l_{\text{max}} - 1, l_{\text{max}} - 2, \ldots, 1, 0. $$

This gives the actual form of the ULV factorization:

$$ H = L_i L_i^T, \ L_i \equiv L^{(l_{\text{max}})}. $$
Also, assume the solution pieces \( y_{ij} \equiv y_{i,j} \) in (2.10) for all nodes \( j \) at each level \( l \) of \( T[k] \) form a vector \( \bar{y}^{(l)} \) (Figure 2.5). That is, let

\[
\begin{align*}
(3.24) \quad &\bar{y}^{(l)} = (y_{j_1}^T \cdots y_{j_{i_0}}^T)^T, \quad (j_1, \ldots, j_{i_0} : \text{all nodes at level } l \text{ of } T[k]) \\
(3.25) \quad &y^{(0)} = y_k, \quad y^{(l+1)} = (\Omega^{(l)})^T \left( \begin{array}{c} \bar{y}^{(l)} \\ y^{(l)} \end{array} \right), \quad l = 0, 1, \ldots, l_{\max} - 1.
\end{align*}
\]

Then it can be verified that

\[
(3.26) \quad y_i \equiv y^{(l_{\max})}.
\]

According to (3.22) and (3.25), we have the following recursive relationship for \( l = l_{\max} - 1, l_{\max} - 2, \ldots, 0: \)

\[
\begin{align*}
(R^{(l)})^T[(U^{(l+1)})^T(L^{(l+1)})^{-T}y^{(l+1)}] \\
= (R^{(l)})^T \left[(U^{(l+1)})^T(\Omega^{(l)})^{-1}(\Omega^{(l)})^T \left( \begin{array}{c} I \\ L^{(l)} \end{array} \right) \Omega^{(l)} \right]^{T} \left(\Omega^{(l)})^T \left( \begin{array}{c} \bar{y}^{(l)} \\ y^{(l)} \end{array} \right) \right] \\
= (\Omega^{(l)}X^{(l)}U^{(l+1)}R^{(l)})^T \left( \begin{array}{c} I \\ (L^{(l)})^{-T} \end{array} \right) \left( \begin{array}{c} \bar{y}^{(l)} \\ y^{(l)} \end{array} \right) \\
= \left( \begin{array}{c} 0 \\ (U^{(l)})^T \end{array} \right) \left( \begin{array}{c} I \\ (L^{(l)})^{-T} \end{array} \right) \left( \begin{array}{c} \bar{y}^{(l)} \\ y^{(l)} \end{array} \right) \quad \text{(equation (3.13))} \\
= (U^{(l)})^T(L^{(l)})^{-T}y^{(l)}.
\end{align*}
\]

Therefore,

\[
\begin{align*}
(3.24) \quad &\bar{y}^{(l)} = (y_{j_1}^T \cdots y_{j_{i_0}}^T)^T, \quad (j_1, \ldots, j_{i_0} : \text{all nodes at level } l \text{ of } T[k]) \\
(3.25) \quad &y^{(0)} = y_k, \quad y^{(l+1)} = (\Omega^{(l)})^T \left( \begin{array}{c} \bar{y}^{(l)} \\ y^{(l)} \end{array} \right), \quad l = 0, 1, \ldots, l_{\max} - 1.
\end{align*}
\]

Then it can be verified that

\[
(3.26) \quad y_i \equiv y^{(l_{\max})}.
\]

According to (3.22) and (3.25), we have the following recursive relationship for \( l = l_{\max} - 1, l_{\max} - 2, \ldots, 0: \)

\[
\begin{align*}
(R^{(l)})^T[(U^{(l+1)})^T(L^{(l+1)})^{-T}y^{(l+1)}] \\
= (R^{(l)})^T \left[(U^{(l+1)})^T(\Omega^{(l)})^{-1}(\Omega^{(l)})^T \left( \begin{array}{c} I \\ L^{(l)} \end{array} \right) \Omega^{(l)} \right]^{T} \left(\Omega^{(l)})^T \left( \begin{array}{c} \bar{y}^{(l)} \\ y^{(l)} \end{array} \right) \right] \\
= (\Omega^{(l)}X^{(l)}U^{(l+1)}R^{(l)})^T \left( \begin{array}{c} I \\ (L^{(l)})^{-T} \end{array} \right) \left( \begin{array}{c} \bar{y}^{(l)} \\ y^{(l)} \end{array} \right) \\
= \left( \begin{array}{c} 0 \\ (U^{(l)})^T \end{array} \right) \left( \begin{array}{c} I \\ (L^{(l)})^{-T} \end{array} \right) \left( \begin{array}{c} \bar{y}^{(l)} \\ y^{(l)} \end{array} \right) \quad \text{(equation (3.13))} \\
= (U^{(l)})^T(L^{(l)})^{-T}y^{(l)}.
\end{align*}
\]

where equations (3.12), (3.23), and (3.26) are used. Then, (3.19) and (3.20) lead to (3.21).

Again, since \( L_k \) is generally a much smaller triangular matrix, the computation with (3.21) is much faster than with (3.19). A direct solution of \( L_i^{-T}y_i \) needs \( O(rN) \) flops if \( H \) has size \( N \) and HSS rank \( r \), while \( L_k^{-T}y_k \) only costs \( O(r^2) \). Such a saving is even more significant in a fully structured version (Remark 3.2).

In the backward substitution stage, we solve intermediate systems of the following form for \( x_i \):

\[
\begin{pmatrix}
L_i^T & L_i^{-1}U_k B_k U_{k+1}^T \\
I & \\
\end{pmatrix}
\begin{pmatrix}
x_i \\
\end{pmatrix} = \begin{pmatrix}
y_i \\
\end{pmatrix},
\]
where $x_{N_i}$ is already available from solution steps associated with upper level separators. That is, we actually solve

$$L_i^T x_i = \tilde{y}_i, \quad \text{with} \quad \tilde{y}_i = y_i - L_i^{-1} U_k B_k U_{k+1}^T x_{N_i}. $$

This involves a ULV-type backward substitution in Section 2.3. Similar to Theorem 3.3, we can prove that $\tilde{y}_i$ can be computed quickly as follows.

**Theorem 3.4.** Let $y_{i,j}$ and $y_{i,k}$ be given in Theorem 3.3. The computation of $\tilde{y}_i = y_i - L_i^{-1} U_k B_k U_{k+1}^T x_i$ can be done by updating only the piece $y_k$ by

$$y_k \leftarrow y_k - \Theta_k x_{N_i}, \quad \Theta_k \text{ available from (3.11)}.$$  

For nonsymmetric problems, similar results can be shown.

**Corollary 3.5.** Assume the conditions in Corollary 3.2 hold, and (3.16) is replaced by a nonsymmetric structured factorization $A = LU$. Then in the solutions of $Ly = b$ and $Ux = y$, (3.21) and (3.27) are replaced by

$$\tilde{b}_{N_i} = b_{N_i} - \Theta_k^T y_{i,k}, \quad \text{and} \quad y_k \leftarrow y_k - \Phi_k x_{N_i},$$

respectively, where $\Theta_k$ and $\Phi_k$ are available from (3.15).

**4. Algorithms and performance analysis with sparse rank relaxation.**

In practice, the structured factorization only starts from a certain switching level $l_s$ of the assembly tree $T$, not only to avoid small block operations, but also to achieve nearly optimal complexity. This is similar to the method in [34], except that the optimization criteria vary. See Theorem 4.1 below. We summarize the structured factorization and solution algorithms as follows. For convenience, $A$ is assumed to be SPD. The nonsymmetric case can be similarly considered.

**Algorithm 1.** *(Structured sparse factorization)*

```plaintext```
for node/separator $i$ from 1 to root ($T$)
1. if $i$ is a leaf, form $F_i \equiv F_i^0$ as in (3.3)
2. if $i$ is at level $l > l_s$ of $T$
   (a) Compute the (exact) Cholesky factorization $F_{i,i} = L_i L_i^T$
   (a) Compute $L_{N_i,i} = F_{N_i,i} L_i^T$
   (b) Compute (dense) $U_i = F_{N_i,N_i} - L_{N_i,i} L_{N_i,i}^T$
else
   (a) Compute an HSS approximation $H$ to $F_i$
   (b) Compute an ULV HSS factorization $F_{i,i} = L_i L_i^T$
   (c) Compute (dense) $U_i = D_{k+1} - \Theta_k^T \Theta_k$, with $\Theta_k = L_k^{-1} \tilde{U}_k B_k U_{k+1}^T$
3. if $i$ is a left node, push $U_i$ onto the update matrix stack
else
   (a) Pop $U_j$ from the update matrix stack for $j = \text{sib } (i)$
   (b) Compute $F_p = F_p^0 + U_j \bigoplus U_i$ for $p = \text{par } (i)$
```

**Algorithm 2.** *(Structured sparse solution)*

1. Partition $b$ into $b_i$ pieces according to the sizes of leaf separators
2. for node/separator $i$ from 1 to root ($T$)
   if $i$ is at level $l > l_s$ of $T$
   (a) Solve a lower-triangular system $L_i y_i = b_i$
(b) Update $b_{\tilde{N}_i}$ to $\tilde{b}_{\tilde{N}_i} = b_{\tilde{N}_i} - L_{\tilde{N}_i,i}y_{l,k}$

else

(a) Solve $L_{\tilde{N}_i}y_{i} = b_{i}$ with the ULV forward substitution in Section 2.3

(b) Update $b_{\tilde{N}_i}$ to $\tilde{b}_{\tilde{N}_i}$ in (3.21)

3. for $i$ from the root of $T$ to node (separator) 1
   if $i$ is at level $1 > l_s$ of $T$
   (a) Solve an upper-triangular system $L_i^T x_i = y_i$
   (b) for all nodes $j$ such that $i \in N_j$, update $y_j$ to $\tilde{y}_j = y_j - L_{j,i}y_i$
   else
   (a) Solve $L_i^T y_i = b_i$ with the ULV forward substitution in Section 2.3
   (b) for all nodes $j$ such that $i \in N_j$, update the piece $y_k$ of $y_j$ to $\tilde{y}_k$ in (3.27)

Both the multifrontal method and the HSS algorithms have good scalability. In a parallel implementation, we traverse $T$ and also the intermediate HSS trees $T_i$ levelwise. The algorithm can be applied to general sparse SPD matrices and can be generalized to nonsymmetric ones. For simplicity, we only consider its complexity in terms of sparse matrices arising from the discretizations of 2D and 3D PDEs.

4.1. Classical complexity analysis. We first consider the classical case where the HSS ranks of all the frontal matrices are bounded by $r$, and then relax this requirement to get more flexible results.

**Theorem 4.1.** (Complexity optimization strategies) Assume Algorithms 1 and 2 are applied to a sparse matrix $A$, and the HSS ranks of all the frontal matrices $F_i$ in Algorithm 1 are bounded by $r$. Denote the structured multifrontal factorization cost, solution cost, and memory size by $\xi_{\text{fact}}$, $\xi_{\text{sol}}$, and $\sigma_{\text{mem}}$, respectively. Let root ($T$) be at level 0 and the leaves of $T$ be at level $l_{\text{max}}$. The switching level $l_s$ ($0 \leq l_s \leq l_{\text{max}}$) is chosen to obtain optimal complexity as follows:

- If $A$ is obtained from a 2D $N \times N$ mesh and $n = N^2$, the counts are given in row 2 of Table 4.1. The switching level $l_s$ satisfies $l_{\text{max}} - l_s = O(\log N)$, so that the factorization costs before and after the switching level are the same.
- If $A$ is obtained from a 3D $N \times N \times N$ mesh and $n = N^3$, the results are given in row 3 of Table 4.1. The switching level $l_s$ satisfies $l_{\text{max}} - l_s = O(\log N)$, so that the solution costs before and after the switching level are the same.

<table>
<thead>
<tr>
<th></th>
<th>$\xi_{\text{fact}}$</th>
<th>$\xi_{\text{sol}}$</th>
<th>$\sigma_{\text{mem}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D</td>
<td>$O(rn \log n)$</td>
<td>$O(n \log r + O(n \log \log n))$</td>
<td>$O(n \log r + O(n \log \log n))$</td>
</tr>
<tr>
<td>3D</td>
<td>$O(rn^{4/3})$</td>
<td>$O(r^{1/2}n)$</td>
<td>$O(r^{1/2}n)$</td>
</tr>
</tbody>
</table>

*Table 4.1*

Factorization cost $\xi_{\text{fact}}$, solution cost $\xi_{\text{sol}}$, and storage $\sigma_{\text{mem}}$ of the structured multifrontal method applied to a discretized matrix $A$ of order $n$ on a regular mesh.

**Proof.** We briefly sketch the proof. For the 2D case, with an process similar to the proof in [34, Theorem 4.2], we can obtain the total cost for the factorization algorithm:

\[
(4.1) \quad \xi_{\text{fact}} = \sum_{l=0}^{l_{\text{max}}} 4^l O\left(\frac{N}{2^l}\right)^3 + \sum_{l=0}^{l_s} 4^l O\left(\frac{N}{2^l}\right)^2 = O\left(\frac{N^3}{2^{l_s}}\right) + O(rN^2l_s).
\]
Assume $N$ and $l_{\text{max}}$ are sufficiently large. $\xi_{\text{fact}}$ is optimized when the costs before and after the switch level $l_s$ are equal. That is,

$$
O\left(\frac{N^3}{2^{l_s}}\right) = O(r N^2 l_s), \quad \text{or} \quad 1_s = l_{\text{max}} - O(\log r) - O(\log l_s) = O(\log N).
$$

Then $\xi_{\text{fact}} = O(r n \log n)$. In such a situation, the solution cost is

$$
\xi_{\text{sol}} = \sum_{l=1}^{l_{\text{max}}} 4^4 O\left(\frac{N^2}{2^l}\right) + \sum_{l=0}^{l_s} 4^4 O\left(\frac{N^2}{2^l}\right) = O(n (l_{\text{max}} - l_s)) + O(r 2^{l_s} N)
$$

Then we get the results in row 3 of Table 4.1.

(4.3) $2^{l_s} = O(N/r^{1/2}).$

According to this theorem, when $r$ is small, our factorization method has performance close to the fully structured version (Remark 3.2). The solution costs and memory requirements of the two versions are in about the same orders in both 2D and 3D cases.

For 2D discrete elliptic equations, it is known that $r = O(1)$ [9]. Thus, our structured algorithms have nearly linear complexity. This similarly holds for 2D Helmholtz equations where the rank bound is $r = O(\log n)$ under certain assumptions [14]. For 3D, the bound is $r = O(n^{1/3})$ [9], and Theorem 4.1 indicates $\xi_{\text{fact}} = O(n^{5/3})$, $\xi_{\text{sol}} = O(n^{7/6})$. However, these bounds can be significantly improved as follows.

### 4.2. Rank relaxation.

The estimates above are useful when the HSS rank bound $r$ is small. When $r$ is large or depends on the HSS matrix size, the analysis can highly overestimate the actual costs. In fact, the numerical ranks of the individual HSS blocks at different levels of the HSS tree can be allowed to increase along the level, so that even if $r$ is large, we can still achieve satisfactory complexity. This is explained as follows.

**Lemma 4.2.** [37] (Dense rank relaxation) Suppose $F$ is an $N \times N$ dense matrix, and $T$ is a perfect binary tree with $l_{\text{max}} = O(\log N)$ levels. Partition $F$ into $O(\log N)$ levels of HSS blocks following $T$ so that the HSS block rows corresponding to the nodes at level $l$ of $T$ have row dimensions $N_l \equiv O(N/2^l)$ and maximum numerical rank $r_l$. Then for an $r_l$ value, the costs for the construction of an HSS form for $F$, its ULV factorization, and the ULV solution are $\tilde{\xi}_{\text{constr}}$, $\tilde{\xi}_{\text{fact}}$, and $\tilde{\xi}_{\text{sol}}$ flops, respectively, and the memory size is $\tilde{\sigma}_{\text{mem}}$, as shown in Table 4.2.
The factorization of a 3D discretized Helmholtz equation. We observe that, for each dense example, Figure 4.1 shows in various practical problems, although an analytical proof is not yet available. As order $n$, the memory size be $O(\frac{\sigma}{\alpha^3})$, where $\sigma$ is obtained from a 2D frontal matrix and $\alpha < \alpha < 2$. Suppose each order $N$ frontal matrix $F$ satisfies the condition of $F$ in Lemma 4.2. Let the costs of Algorithms 1–2 applied to $A$ be $O(\xi_{fact})$ and $O(\xi_{sol})$ respectively, and the memory size be $O(\sigma_{mem})$. Then if $r_l$ satisfies the patterns as in Lemma 4.2, we have the following results:

- If $A$ is obtained from a 2D $N \times N$ mesh and $n = N^2$, the results are given in Table 4.3. The switching level $l_s$ satisfies $l_{max} - l_s = O(\log N)$, so that the factorization costs at the levels before and after $l_s$ are the same.
- If $A$ is obtained from a 3D $N \times N \times N$ mesh and $n = N^3$, the results are given in Table 4.4. The switching level $l_s$ satisfies $l_{max} - l_s = O(\log N)$, so that the solution costs at the levels before and after $l_s$ are the same.

We call each $r_l$ in Table 4.2 a rank pattern. Such rank patterns have been observed in various practical problems, although an analytical proof is not yet available. As an example, Figure 4.1 shows $r_l$ for two dense frontal matrices in the multifrontal factorization of a 3D discretized Helmholtz equation. We observe that, for each dense frontal matrix, $r_l$ is roughly $O(N_l^{1/2})$.

![Numerical rank patterns $r_l$ of the HSS blocks at level $l$ for two dense frontal matrices of sizes $N = 2 \times 10^4$ and $4 \times 10^4$, respectively, in the multifrontal factorization of a 3D discretized Helmholtz equation.]

Based on this lemma, we have the following result, whose earlier variations can be found in the report [36].

**Theorem 4.3.** (Sparse rank relaxation) Assume $A$ is a discretized matrix of order $n$. Suppose each order $N$ frontal matrix $F$ satisfy the condition of $F$ in Lemma 4.2. Let the costs of Algorithms 1–2 applied to $A$ be $O(\xi_{fact})$ and $O(\xi_{sol})$, respectively, and the memory size be $O(\sigma_{mem})$. Then if $r_l$ satisfies the patterns as in Lemma 4.2, we have the following results:

<table>
<thead>
<tr>
<th>$r_l$</th>
<th>$r = \max r_l$</th>
<th>$\xi_{constr}$</th>
<th>$\xi_{fact}$</th>
<th>$\xi_{sol}$</th>
<th>$\sigma_{mem}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O(1)$</td>
<td>$O(1)$</td>
<td>$O(N^2)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
</tr>
<tr>
<td>$O((\log N)^p)$</td>
<td>$O((\log N)^p)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
</tr>
<tr>
<td>$O(N^{1/p})$</td>
<td>$O(N^{1/p})$</td>
<td>$O(N^2)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
</tr>
<tr>
<td>$O(N^{1/p})$</td>
<td>$O(N^{1/p})$</td>
<td>$O(N^2)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
</tr>
<tr>
<td>$\alpha &lt; \sqrt[2]{2}$</td>
<td>$O(N^{1/2})$</td>
<td>$O(N^2)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
</tr>
<tr>
<td>$\alpha = 2^{1/3}$</td>
<td>$O(N^{1/3})$</td>
<td>$O(N^2)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
</tr>
<tr>
<td>$2^{1/3} &lt; \alpha &lt; 2^{1/2}$</td>
<td>$O(N^{1/2})$</td>
<td>$O(N^2)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
</tr>
<tr>
<td>$\alpha = 2^{1/2}$</td>
<td>$O(N^{1/2})$</td>
<td>$O(N^2)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
<td>$O(N)$</td>
</tr>
</tbody>
</table>

Table 4.2

**Costs and storage of HSS construction, ULV factorization, and ULV solution algorithms with rank relaxation, where $p$ is a positive integer, $\alpha > 0$, and $r = \max r_l$ is the HSS rank of $H$.**

Fig. 4.1. Numerical rank patterns $r_l$ of the HSS blocks at level $l$ for two dense frontal matrices of sizes $N = 2 \times 10^4$ and $4 \times 10^4$, respectively, in the multifrontal factorization of a 3D discretized Helmholtz equation.
the 2D case, (4.1) is replaced by

\[ \text{tolerance in the compression. This saves the costs of the HSS factorizations above} \]

\[ \text{the following:} \]

\[ \text{algorithms with some examples. In addition to the notation in Section 4.2, we also use} \]

\[ \text{Lemma 4.2 are used to replace some operations. For example, when} \]

\[ \text{method applied to a discretized matrix} \]

\[ \text{method applied to a discretized matrix} \]

\[ \text{We demonstrate the performance of our algo-} \]

\[ \text{Proof. The proof is similar to that of Theorem 4.1, except that the results in} \]

\[ \text{Lemma 4.2 are used to replace some operations. For example, when} \]

\[ \text{in the 2D case, (4.1) is replaced by} \]

\[ \xi_{\text{fact}} = \sum_{l=1}^{l_{\text{max}}} 4^l O\left(\frac{n^{1/2}}{2^l}\right)^3 + \sum_{l=0}^{l_{\text{max}}} 4^l O\left(\frac{n^{1/2}}{2^l} \log\left(\frac{n^{1/2}}{2^l}\right)\right) \]

\[ = \sum_{l=1}^{l_{\text{max}}} 4^l O\left(\frac{n^{1/2}}{2^l}\right)^3 + n \sum_{l=0}^{l_{\text{max}}} \frac{1}{2} \log n - l = n^{3/2} + n l_{\text{max}} (\log n - 1) \]

The minimum is \( O(n \log^2 n) \), when \( O(n^{1/2} / [l_{\text{max}} (\log n - 1)]) \) = \( 2^l \).

**Remark 4.1.** According to this theorem, the discretized 3D Helmholtz equation mentioned in Figure 4.1 can be factored in about \( O(n^{4/3}) \) flops, even though \( r \) depends on \( n \). Moreover, the solution costs and the storage are nearly linear in \( n \).

**Remark 4.2.** If the solver is used as a preconditioner, we can set a relatively large tolerance in the compression. This saves the costs of the HSS factorizations above the switching level \( l_{\text{max}} \). According to the complexity optimization strategies, this also indicates that the number of HSS factorization levels in \( T \) can be increased, so as to further improve the efficiency of the approximate factorization.

**5. Numerical experiments.** We demonstrate the performance of our algorithms with some examples. In addition to the notation in Section 4.2, we also use the following:

<table>
<thead>
<tr>
<th>( r_1 )</th>
<th>( r = \max r_1 )</th>
<th>( \xi_{\text{fact}} )</th>
<th>( \xi_{\text{sol}} )</th>
<th>( \sigma_{\text{mem}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( O(1) )</td>
<td>( O(1) )</td>
<td>( O(n \log n) )</td>
<td>( O(n \log n) )</td>
<td>( O(n \log n) )</td>
</tr>
<tr>
<td>( O((\log N)^p) )</td>
<td>( O((\log N)^p) )</td>
<td>( O(n \log n) )</td>
<td>( O(n \log n) )</td>
<td>( O(n \log n) )</td>
</tr>
<tr>
<td>( O(N^{l/p}) )</td>
<td>( p \geq 3 )</td>
<td>( O(N^{l/p}) )</td>
<td>( O(n \log^2 n) )</td>
<td>( O(n \log n) )</td>
</tr>
<tr>
<td>( O(N^{l/p}) )</td>
<td>( p = 2 )</td>
<td>( O(N^{l/2}) )</td>
<td>( O(n \log^2 n) )</td>
<td>( O(n \log n) )</td>
</tr>
<tr>
<td>( O(\alpha^{l_{\text{max}}-l_{\text{max}}}^{-l}) )</td>
<td>( \alpha \leq 2^{1/3} )</td>
<td>( O(N^{l/3}) )</td>
<td>( O(n \log n) )</td>
<td>( O(n \log n) )</td>
</tr>
<tr>
<td>( O(\alpha^{l_{\text{max}}-l_{\text{max}}}^{-l}) )</td>
<td>( 2^{1/3} &lt; \alpha \leq 2^{1/2} )</td>
<td>( O(N^{l/2}) )</td>
<td>( O(n \log^2 n) )</td>
<td>( O(n \log n) )</td>
</tr>
</tbody>
</table>

**Table 4.3.** Factorization cost \( \xi_{\text{fact}} \), solution cost \( \xi_{\text{sol}} \), and storage \( \sigma_{\text{mem}} \) of the structured multifrontal method applied to a discretized matrix \( A \) of order \( n \) on a 2D \( N \times N \) mesh, where \( p \in \mathbb{N} \) and \( \alpha > 0 \).

<table>
<thead>
<tr>
<th>( r_1 )</th>
<th>( r = \max r_1 )</th>
<th>( \xi_{\text{fact}} )</th>
<th>( \xi_{\text{sol}} )</th>
<th>( \sigma_{\text{mem}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( O(1) )</td>
<td>( O(1) )</td>
<td>( O(n^{4/3}) )</td>
<td>( O(n) )</td>
<td>( O(n) )</td>
</tr>
<tr>
<td>( O((\log_2 N)^p) )</td>
<td>( O((\log_2 N)^p) )</td>
<td>( O(n \log^{1/2} n) )</td>
<td>( O(n \log^{1/2} n) )</td>
<td>( O(n \log^{1/2} n) )</td>
</tr>
<tr>
<td>( O(N^{l/p}) )</td>
<td>( p \geq 3 )</td>
<td>( O(N^{l/3}) )</td>
<td>( O(n \log^{1/2} n) )</td>
<td>( O(n \log^{1/2} n) )</td>
</tr>
<tr>
<td>( O(N^{l/p}) )</td>
<td>( p = 2 )</td>
<td>( O(N^{l/2}) )</td>
<td>( O(n \log^{1/2} n) )</td>
<td>( O(n \log n) )</td>
</tr>
<tr>
<td>( O(\alpha^{l_{\text{max}}-l_{\text{max}}}^{-l}) )</td>
<td>( \alpha \leq 2^{1/3} )</td>
<td>( O(N^{l/3}) )</td>
<td>( O(n \log^{1/2} n) )</td>
<td>( O(n \log^{1/2} n) )</td>
</tr>
<tr>
<td>( O(\alpha^{l_{\text{max}}-l_{\text{max}}}^{-l}) )</td>
<td>( 2^{1/3} &lt; \alpha \leq 2^{1/2} )</td>
<td>( O(N^{l/2}) )</td>
<td>( O(n \log^{3/2} n) )</td>
<td>( O(n \log n) )</td>
</tr>
</tbody>
</table>

**Table 4.4.** Factorization cost \( \xi_{\text{fact}} \), solution cost \( \xi_{\text{sol}} \), and storage \( \sigma_{\text{mem}} \) of the structured multifrontal method applied to a discretized matrix \( A \) of order \( n \) on a 3D \( N \times N \times N \) mesh, where \( p \in \mathbb{N} \) and \( \alpha > 0 \).
NEW: The new structured multifrontal factorization and solution.

MF: The exact multifrontal method.

e_2 = \frac{\|x - \hat{x}\|_2}{\|x\|_2}, \quad \gamma_2 = \frac{\|Ax - b\|_2}{\|b\|_2}, \quad \text{where} \quad x \text{ and } \hat{x} \text{ are the exact and approximation solutions of } Ax = b, \text{ respectively.}

Example 1. We consider a Helmholtz equation:

\[(5.1) \quad [-\Delta + \omega^2 c(x)^{-2}]u = f\]

where \(\omega\) the angular frequency, \(c(x)\) the velocity field, and \(f\) is the forcing term.

First, consider the 2D case. The Helmholtz operator is discretized on \(N \times N\) meshes. We solve linear systems \(Ax = b\) with the discretized matrix \(A\) of order \(n = N^2\). A sequential code in Fortran 90 for NEW is tested on a 2.33GHz Intel E5410 processor. The flop counts and timings for the factorization method is shown in Figure 5.1, where \(\omega = 5\text{Hz}\). For \(N = 2^8, 2^9, \ldots, 2^{12} = 4096\), we use \(l_{\text{max}}\) as in Table 5.1. A relative tolerance \(\tau = 10^{-6}\) and \(l_{\text{max}} - l_s = 9\) are used in NEW. Such choice of \(l_s\) roughly gives the minimum factorization cost for each \(N\). This is also consistent with Theorem 4.3, which indicates that \(l_{\text{max}} - l_s\) only changes very slowly for different \(N\).

![Flop counts and timings comparison](image)

(i) Flop counts \(\xi_{\text{fact}}\)  
(ii) Timing in parallel

The memory sizes \(\sigma_{\text{mem}}\) of the algorithms are also reported in Table 5.1. \(\sigma_{\text{mem}}\) of both NEW and MF in the table scales roughly linearly, but NEW performs better. That is, when \(N\) doubles and \(n\) quadruples, \(\sigma_{\text{mem}}\) also roughly quadruples.

<table>
<thead>
<tr>
<th>(N) ((n = N^2))</th>
<th>256</th>
<th>512</th>
<th>1024</th>
<th>2048</th>
<th>4096</th>
<th>8192</th>
</tr>
</thead>
<tbody>
<tr>
<td>(l_{\text{max}})</td>
<td>13</td>
<td>15</td>
<td>17</td>
<td>19</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>MF</td>
<td>(5.36E6)</td>
<td>(2.54E7)</td>
<td>(1.18E8)</td>
<td>(5.34E8)</td>
<td>(2.40E9)</td>
<td>(1.06E10)</td>
</tr>
<tr>
<td>NEW</td>
<td>(4.88E6)</td>
<td>(2.06E7)</td>
<td>(8.59E7)</td>
<td>(3.53E8)</td>
<td>(1.44E9)</td>
<td>(5.82E9)</td>
</tr>
</tbody>
</table>

Table 5.1

Example 1: Memory size \(\sigma_{\text{mem}}\) (number of nonzero entries in the factors) of NEW and MF for (5.1) discretized on \(N \times N\) meshes in 2D.

The solution costs and the accuracies of the algorithms are given in Table 5.2, where the right-hand side \(b\) in \(Ax = b\) is obtained with a random \(x\). Again, the
solution cost of NEW scales linearly. NEW also produces modest accuracies. After few steps of iterative refinements, high accuracies are obtained, and are comparable to those of MF.

<table>
<thead>
<tr>
<th>$N (n = N^2)$</th>
<th>256</th>
<th>512</th>
<th>1024</th>
<th>2048</th>
<th>4096</th>
<th>8192</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l_{\text{max}}$</td>
<td>13</td>
<td>15</td>
<td>17</td>
<td>19</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>MF</td>
<td>$1.06E7$</td>
<td>$5.03E7$</td>
<td>$2.33E8$</td>
<td>$1.06E9$</td>
<td>$4.76E9$</td>
<td>$2.11E10$</td>
</tr>
<tr>
<td>NEW</td>
<td>$9.78E6$</td>
<td>$4.13E7$</td>
<td>$1.72E8$</td>
<td>$7.09E8$</td>
<td>$2.89E9$</td>
<td>$1.17E10$</td>
</tr>
</tbody>
</table>

(i) Solution flops $\xi_{\text{sol}}$ of NEW and MF

<table>
<thead>
<tr>
<th>$N (n = N^2)$</th>
<th>256</th>
<th>512</th>
<th>1024</th>
<th>2048</th>
</tr>
</thead>
<tbody>
<tr>
<td>MF</td>
<td>$e_2$</td>
<td>$2.13E-16$</td>
<td>$2.14E-16$</td>
<td>$2.93E-16$</td>
</tr>
<tr>
<td>NEW Original</td>
<td>$e_2$</td>
<td>$5.13E-6$</td>
<td>$1.17E-5$</td>
<td>$2.30E-5$</td>
</tr>
<tr>
<td>NEW After 2 steps of iterative refinement</td>
<td>$e_2$</td>
<td>$1.01E-14$</td>
<td>$1.90E-14$</td>
<td>$7.92E-14$</td>
</tr>
<tr>
<td></td>
<td>$\gamma_2$</td>
<td>$4.54E-8$</td>
<td>$6.59E-8$</td>
<td>$7.37E-8$</td>
</tr>
<tr>
<td></td>
<td>$\gamma_2$</td>
<td>$1.39E-16$</td>
<td>$1.56E-16$</td>
<td>$3.63E-16$</td>
</tr>
<tr>
<td></td>
<td>$\gamma_2$</td>
<td>$2.56E-9$</td>
<td>$2.12E-9$</td>
<td>$2.89E-9$</td>
</tr>
</tbody>
</table>

(ii) Accuracies of NEW and MF (relative error $e_2$ and relative residual $\gamma_2$)

Table 5.2

Example 1: Solution costs $\xi_{\text{sol}}$ and accuracies of the two algorithms discretized on $N \times N$ meshes in 2D.

Similarly, we test $A$ discretized on $N \times N \times N$ meshes in the 3D case. The results are shown in Table 5.3, with the values of $l_{\text{max}}$. A relative tolerance $\tau = 10^{-6}$ and $l_{\text{max}} - l_e = 10$ are used in NEW. Again, this choice of $l_e$ roughly gives the minimum factorization cost for each $N$. The accuracies are similar to those in the 2D case and are not shown.

<table>
<thead>
<tr>
<th>$N (n = N^3)$</th>
<th>48</th>
<th>96</th>
<th>192</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l_{\text{max}}$</td>
<td>12</td>
<td>15</td>
<td>18</td>
</tr>
<tr>
<td>$\xi_{\text{fact}}$</td>
<td>MF</td>
<td>$1.58E11$</td>
<td>$1.05E13$</td>
</tr>
<tr>
<td></td>
<td>NEW</td>
<td>$1.83E11$</td>
<td>$8.32E12$</td>
</tr>
<tr>
<td>$\sigma_{\text{mem}}$</td>
<td>MF</td>
<td>$1.06E08$</td>
<td>$1.87E09$</td>
</tr>
<tr>
<td></td>
<td>NEW</td>
<td>$9.82E07$</td>
<td>$1.24E09$</td>
</tr>
<tr>
<td>$\xi_{\text{sol}}$</td>
<td>MF</td>
<td>$2.13E08$</td>
<td>$3.74E09$</td>
</tr>
<tr>
<td></td>
<td>NEW</td>
<td>$2.01E08$</td>
<td>$2.56E09$</td>
</tr>
</tbody>
</table>

Table 5.3

Example 1: Costs (flops) and storage (number of nonzeros) of NEW and MF for (5.1) discretized on $N \times N \times N$ meshes in 3D.

**Example 2.** To show that the method is more generally applicable, we run it on some test matrices from the University of Florida Sparse Matrix Collection [10]. These matrices arise from various background.

See Table 5.4 for the performance. In NEW, we choose $\tau = 10^{-6}$, and $l_{\text{max}} - l_e = 7, 7, 8, 6, 6, 7$ for the matrices from the left to the right in Table 5.4, respectively. The matrices are relatively small. However, when they are factored, NEW is up to twice as fast as MF. NEW also requires less memory in the table.
more storage than the block diagonal one. faster, and the cost is much lower. The structured preconditioner needs about 80% accuracies are given in Table 5.6. Clearly, the convergence of the conjugate method with the new solver as the preconditioner (CG-NEW) is significantly faster, and the cost is much lower. The structured preconditioner needs about 80% more storage than the block diagonal one.

<table>
<thead>
<tr>
<th>Matrix A</th>
<th>apache2</th>
<th>ecology2</th>
<th>G3_circuit</th>
<th>parabo_fem</th>
<th>thermo</th>
<th>tmt_sym</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>715,176</td>
<td>999,999</td>
<td>1,585,478</td>
<td>525,825</td>
<td>204,316</td>
<td>726,713</td>
</tr>
<tr>
<td>nnz(A)</td>
<td>4,817,870</td>
<td>4,995,991</td>
<td>7,660,826</td>
<td>3,674,625</td>
<td>1,423,116</td>
<td>5,080,961</td>
</tr>
<tr>
<td>(l_{\max})</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>14</td>
<td>14</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 5.4

Example 2: Costs and storage (number of nonzeros) of NEW and MF for some test matrices in the University of Florida Sparse Matrix Collection [10], where nnz(A) is the number of nonzeros in A, and parabo_fem and thermo are abbreviations of the matrix names parabolic_fem and thermomech_dH in [10], respectively.

<table>
<thead>
<tr>
<th>Matrix A</th>
<th>apache2</th>
<th>ecology2</th>
<th>G3_circuit</th>
<th>parabo_fem</th>
<th>thermo</th>
<th>tmt_sym</th>
</tr>
</thead>
<tbody>
<tr>
<td>MF</td>
<td>(2.52E11)</td>
<td>(2.50E10)</td>
<td>(8.77E10)</td>
<td>(9.14E9)</td>
<td>(8.56E8)</td>
<td>(1.33E10)</td>
</tr>
<tr>
<td>NEW</td>
<td>(1.51E11)</td>
<td>(1.26E10)</td>
<td>(5.20E10)</td>
<td>(6.75E9)</td>
<td>(6.46E8)</td>
<td>(8.68E9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Matrix A</th>
<th>apache2</th>
<th>ecology2</th>
<th>G3_circuit</th>
<th>parabo_fem</th>
<th>thermo</th>
<th>tmt_sym</th>
</tr>
</thead>
<tbody>
<tr>
<td>MF</td>
<td>(1.77E8)</td>
<td>(4.88E7)</td>
<td>(1.14E8)</td>
<td>(2.50E07)</td>
<td>(6.51E6)</td>
<td>(3.31E7)</td>
</tr>
<tr>
<td>NEW</td>
<td>(9.19E7)</td>
<td>(2.77E7)</td>
<td>(6.40E7)</td>
<td>(1.66E07)</td>
<td>(4.79E6)</td>
<td>(1.93E7)</td>
</tr>
</tbody>
</table>

Table 5.5

Example 2: Accuracies of NEW and MF for the test matrices in Table 5.4.

Example 3. In this example, we test the effectiveness of the method as a preconditioner, when a larger relative tolerance \(\tau\) is used in the compression. The matrix tmt_sym in Example 2 is tested.

In Figure 5.2, we show the convergence behaviors of the conjugate method (CG), the conjugate method with a block diagonal preconditioner (CG-bdiag), and the conjugate method with the new solver as the preconditioner (CG-NEW). The costs and final accuracies are given in Table 5.6. Clearly, the convergence of CG-NEW is significantly faster, and the cost is much lower. The structured preconditioner needs about 80% more storage than the block diagonal one.
**Fig. 5.2.** Example 3: Convergence of the conjugate method (CG), CG with a block diagonal preconditioner (CG-bdiag), and CG with the new solver as the preconditioner (CG-NEW), where a diagonal block size 20 is used in CG-bdiag and $\tau = 10^{-3}$ is used in CG-NEW.

<table>
<thead>
<tr>
<th></th>
<th>Precomputation flops</th>
<th>Total flops</th>
<th>Number of iterations</th>
<th>$\gamma_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG</td>
<td>N/A</td>
<td>$1.18E11$</td>
<td>6,517</td>
<td>$2.93E-15$</td>
</tr>
<tr>
<td>CG-bdiag</td>
<td>$1.94E8$</td>
<td>$2.24E11$</td>
<td>4,822</td>
<td>$2.04E-15$</td>
</tr>
<tr>
<td>CG-NEW</td>
<td>$5.70E9$</td>
<td>$1.17E10$</td>
<td>23</td>
<td>$1.32E-16$</td>
</tr>
</tbody>
</table>

*Table 5.6* Example 3: Costs and accuracies for Figure 5.2, where the precomputation is to get the preconditioner (factorization).

**Remark 5.1.** These examples are intended to show the complexity and accuracies of our new algorithms. The parallel performance is not discussed here. Also, an efficient reordering scheme for irregular meshes is expected to be done so that we can solve larger problems. Ordering schemes and pivoting strategies for nonsymmetric matrices will be considered in future work.

**6. Conclusions.** We show a new structured multifrontal method which is simpler and more general than some similar ones. In a strategy of combing the multifrontal method with rank structured matrices, we show a concept of reduced matrices. This concept enables us to replace the operations on large structured matrices by those on simple compact ones. The performance is studied in terms of the complexity. The idea of a sparse rank relaxation also allows us to use our method to solve more problems where the related off-diagonal ranks are not bounded and follow certain rank patterns.

These ideas are very useful in developing other structured methods, especially those avoiding dense update matrices, which will appear in future work. The idea of reduced matrices helps overcome the major computational costs for computing structured update matrices in the factorization. Our algorithm here is also highly scalable. Its efficient parallel version is expected to be built as a toolbox for general sparse solutions.

**Acknowledgments.** The author would like to thank Shen Wang for helping with the tests in Example 1. The research of the author was supported in part by NSF grants DMS-1115572 and CHE-0957024.
REFERENCES


