



# Preconditioning of linear least squares by RIF for implicitly held normal equations

J Scott, M Tuma

January 2016

Submitted for publication in SIAM Journal of Scientific Computing

RAL Library  
STFC Rutherford Appleton Laboratory  
R61  
Harwell Oxford  
Didcot  
OX11 0QX

Tel: +44(0)1235 445384  
Fax: +44(0)1235 446403  
email: [libraryral@stfc.ac.uk](mailto:libraryral@stfc.ac.uk)

Science and Technology Facilities Council preprints are available online  
at: <http://epubs.stfc.ac.uk>

**ISSN 1361- 4762**

Neither the Council nor the Laboratory accept any responsibility for loss or damage arising from the use of information contained in any of their reports or in any communication about their tests or investigations.

# Preconditioning of linear least squares by RIF for implicitly held normal equations

Jennifer Scott\* and Miroslav Tůma†

January 27, 2016

## Abstract

The efficient solution of the normal equations corresponding to a large sparse linear least squares problem can be extremely challenging. Robust incomplete factorization (RIF) preconditioners represent one approach that has the important feature of computing an incomplete  $LL^T$  factorization of the normal equations matrix without having to form the normal matrix itself. The right-looking implementation of Benzi and Tůma has been used in a number of studies but experience has shown that it can be computationally slow and its memory requirements are not known a priori. Here a new left-looking variant is presented that employs a symbolic preprocessing step to replace the potentially expensive searching through entries of the normal matrix. This involves a directed acyclic graph (dag) that is computed on-the-fly. An inexpensive but effective pruning algorithm is proposed to limit the number of edges in the dag. Problems arising from practical applications are used to compare the performance of the right-looking approach with a left-looking implementation that computes the normal matrix explicitly and our new implicit dag-based left-looking variant.

---

\*Scientific Computing Department, Rutherford Appleton Laboratory, Harwell Oxford, Oxfordshire, OX11 0QX, UK. Correspondence to: jennifer.scott@stfc.ac.uk Supported by EPSRC grants EP/I013067/1 and EP/M025179/1.

†Department of Numerical Mathematics, Faculty of Mathematics and Physics, Charles University and Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic, (mirektuma@karlin.mff.cuni.cz) Supported by the project P201/13-06684S of the Grant Agency of the Czech Republic and partially also by the ERC project MORE LL1202 financed by the MSMT of the Czech Republic.

# 1 Introduction

Linear least squares problems arise in a wide variety of practical applications. Let us consider the algebraic problem of linear least squares (LS) in the following form

$$\min_x \|b - Ax\|_2, \quad (1.1)$$

where  $A \in \mathbb{R}^{m \times n}$  ( $m \geq n$ ) is a large sparse matrix with full column rank and  $b \in \mathbb{R}^m$  is given. Solving (1.1) is mathematically equivalent to solving the  $n \times n$  *normal equations*

$$Cx = A^T b, \quad C = A^T A, \quad (1.2)$$

where, since  $A$  has full column rank, the *normal matrix*  $C$  is symmetric positive definite. To solve very large LS problems, an iterative method may be the method of choice because they can require much less storage and fewer operations than their direct counterparts. However, iterative methods do not offer the same level of reliability and their successful application often needs a good preconditioner to achieve acceptable convergence rates (or, indeed, to obtain convergence at all).

In recent years, a number of techniques for preconditioning LS problems have been proposed. In particular, significant attention has been devoted to the development of algorithms based on incomplete orthogonal factorizations of  $A$  [30, 37, 39, 40]. Most recently, there is the Multilevel Incomplete QR (MIQR) factorization of Li and Saad [33]. An alternative is the LU-based strategy that was first introduced as a direct method in 1961 in [32]; see its further development as the Peters-Wilkinson method [38] (and later in [11, 14]).

While these approaches are very useful in some circumstances, here we focus on a third and most traditional approach that is based on the normal equations and an incomplete factorization of the symmetric positive definite matrix  $C$ . A direct variant of this approach dates back to 1924 [4]. One problem connected to the normal equations is that they may not be very sparse, for example, if  $A$  contains some dense (or close to dense) rows. A possible way to overcome this is to treat the dense rows separately. An alternative approach is to avoid explicitly forming  $C$  (that is, to work only with  $A$  and  $A^T$ ) and to compute its factorization implicitly. Working with  $C$  implicitly is also important for very large problems for which computing  $C$  may be too costly (in terms of both time and memory). Moreover, forming the normal equations may lead to severe loss of information in highly ill-conditioned cases. The recent limited memory incomplete Cholesky factorization code HSL\_MI35 of Scott and Tůma [29] is designed for the normal equations. It offers an option to input either the matrix  $C$  or  $A$ . In the latter case, a single column of  $C$  is computed at each stage of the incomplete factorization process, thus avoiding the need to store  $C$  explicitly but not the work needed to form the product  $A^T A$  [10]. Recent results by Gould and Scott [26, 27] illustrate that HSL\_MI35 can perform well on a range of problems but that constructing the incomplete factorization can be unacceptably slow when  $A$  has dense rows.

An implicit factorization scheme that uses a (Schur complement-based) approach is the robust incomplete factorization (known as the RIF algorithm) of Benzi and Tůma [9]. RIF is based on  $C$ -orthogonalization and works entirely with  $A$  and  $A^T$  and can be derived as the “dual” of the SAINV (stabilized approximate inverse) preconditioner [9]. The preconditioner is guaranteed to be positive definite and, in exact arithmetic, the incomplete factorization process is breakdown free. A right-looking implementation of RIF for LS problems is available as part of the SPARSLAB software collection of Benzi and Tůma (see <http://www.karlin.mff.cuni.cz/~mirektuma>) and has been used in a number of studies on least squares preconditioners (including [2, 35]). Important weaknesses of this code are that it can be computationally slow (see the results of [26]) and the amount of memory needed is not known a priori (although for the experiments in [9] the total storage is estimated to be approximately 25% more than the storage needed for the final incomplete factor but this itself is not known and will depend on the dropping strategy). The aim of this paper is to present a new potentially more computationally efficient left-looking RIF algorithm that avoids the need to compute **any** entries of  $C$ . A key contribution is to propose and implement the use of a symbolic step to replace the searching through  $C$  that is needed in

previous implementations. Symbolic preprocessing is a standard tool in sparse direct methods for linear systems (see, for example, the early symbolic decomposition for symmetric and positive definite matrices in [20], symbolic evaluations in LU decomposition [24] and the overview of the theory of nonsymmetric elimination trees given in [19]). The proposed symbolic step involves a search using a directed acyclic graph that we construct on-the-fly. When used for RIF applied to LS problems, it has the potentially attractive property of offering the possibility of replacing the update step in the left-looking algorithm by a parallelizable sparse saxpy operation, although the design and development of parallel implementations lies beyond the scope of the current paper.

The rest of the paper is organised as follows. Section 2 recalls the RIF algorithm and then introduces our new left-looking RIF algorithm that employs symbolic preprocessing. The use of RIF for solving LS problems is described in Section 3. Numerical experiments in Section 4 demonstrate the efficiency and robustness of the new approach using a range of problems from real-world linear systems and least squares applications. Finally, Section 5 presents some concluding remarks.

**Notation:** we end this section by introducing the notation that is employed in the rest of the paper. We let the number of nonzero entries of a matrix  $A$  be  $nnz(A)$  and we let these entries be  $a_{i,j}$ . Furthermore, we denote the  $j$ th column of  $A$  by  $a_j$  and  $e_j$  denotes the  $j$ th unit basis vector. For any column  $a_j$ , we define its structure (sparsity pattern) to be the set  $Struct(a_j) = \{i \mid a_{i,j} \neq 0\}$ . Section notation is used to denote part of the matrix so that  $A_k \equiv A_{1:k,1:k}$  is the leading submatrix of order  $k$  and  $A_{1:k,j}$  is the first  $k$  entries of column  $j$ .

For a symmetric positive definite matrix  $C$ , we define the  $C$ -inner product to be

$$\langle x, y \rangle_C = y^T C x, \quad \forall x, y \in \mathbb{R}^n, \quad (1.3)$$

and associated  $C$ -norm

$$\|x\|_C = \sqrt{x^T C x}.$$

In the least squares case,

$$\langle x, y \rangle_C = (y^T A^T)(Ax), \quad \forall x, y \in \mathbb{R}^n.$$

We also recall some standard notation related to sparse matrices and their graphs, see, e.g., [23]. We define  $G(A) = (V, E)$  to be the directed graph of the (nonsymmetric) matrix  $A \in \mathbb{R}^{n \times n}$  with nonzero diagonal entries as follows: the vertex set is  $V = \{1, \dots, n\}$  and for  $1 \leq i \neq j \leq n$  there is an edge  $(i, j) \in E$  from  $i$  to  $j$  if and only if  $a_{i,j} \neq 0$ . There is a directed path from vertex  $i \in V$  to vertex  $k \in V$  if there exists a sequence of vertices  $i = i_0, i_1, i_2, \dots, i_{k-1}, i_k = k$  belonging to  $V$  such that each edge  $(i_{j-1}, i_j)$  is in  $E$ . This path is denoted by  $i \Rightarrow k$  and  $k$  is said to be reachable from  $i$ . The set of vertices that are reachable from  $i$  is denoted by  $Reach(i)$ .

We let  $L = \{l_{i,j}\}$  and  $Z = \{z_{i,j}\}$  denote a lower and upper triangular matrix, respectively. A nonzero off-diagonal entry  $l_{i,j}$  must have  $i > j$ , so any edge in the directed graph of  $L$  must satisfy  $i > j$ . Thus the directed graph of a triangular matrix has no loops and belongs to the class of directed acyclic graphs (dags). We denote the directed acyclic graph of  $L$  by  $dag(L)$ .

In the following, we assume some basic knowledge of the concept of the elimination tree and its role in sparse factorization that is well described, for example, in the survey paper [34].

## 2 Robust incomplete factorization and symbolic decomposition

We start by considering general sparse symmetric and positive definite matrices, that are not necessarily normal matrices.

### 2.1 Left- and right-looking approaches

Consider the factorization of a general sparse symmetric and positive definite matrix  $C$  into the product of two triangular factors of the form

$$C = LL^T. \quad (2.1)$$

This factorization is unique but there are a number of approaches to computing it. For example, one can consider different computational variants of the Cholesky decomposition or the Gram-Schmidt process with the  $C$ -inner product. While the former is well-known, our focus is on the latter and, in particular, the RIF algorithm [8, 9]. Given  $n$  linear independent vectors, the Gram-Schmidt process builds a  $C$ -orthogonal set of vectors  $z_1, z_2, \dots, z_n$ . This can be written as

$$Z^T C Z = I, \quad I = L^T Z. \quad (2.2)$$

where  $Z = [z_1, z_2, \dots, z_n]$  is upper triangular with positive diagonal entries. In exact arithmetic,  $L^T$  is the transposed Cholesky factor of  $C$  and  $Z$  is its inverse. The left-looking Gram-Schmidt process is outlined as Algorithm 2.1; the right-looking process is outlined in Algorithm 2.2. Three computational options are included that correspond to the classical, modified and mixed variants (denoted by CGS, MGS and AINV, respectively); for details, see [31]. The relationship between  $L$  and  $Z$  can be found, for example, in the 1952 seminal paper on the conjugate gradient method by Hestenes and Stiefel [28].

**Algorithm 2.1.** *Left-looking Gram-Schmidt process (classical CGS, modified MGS and mixed AINV variants).*

**Input:** *Symmetric positive definite matrix  $C \in R^{n \times n}$ .*

**Output:** *Factors  $Z$  and  $L$  satisfying (2.2).*

1. **for**  $k = 1 : n$  **do**
2.   Set  $z_k^{(0)} = e_k$
3.   **for**  $j = 1 : k - 1$  **do**
4.     **if** MGS
5.       Set  $l_{k,j} = \langle z_k^{(j-1)}, z_j \rangle_C$
6.     **else if** CGS
7.       Set  $l_{k,j} = \langle z_k^{(0)}, z_j \rangle_C$
8.     **else if** AINV
9.       Set  $l_{k,j} = \langle z_k^{(j-1)}, e_j \rangle_C$
10.    **end if**
11.    Set  $z_k^{(j)} = z_k^{(j-1)} - l_{k,j} z_j$
12.    **end do**
13.    Set  $l_{k,k} = \|z_k^{(k-1)}\|_C$
14.    Set  $z_k = z_k^{(k-1)} / l_{k,k}$
15. **end do**

---

**Algorithm 2.2.** *Right-looking Gram-Schmidt process (classical CGS, modified MGS and mixed AINV variants).*

**Input:** Symmetric positive definite matrix  $C \in R^{n \times n}$ .

**Output:** Factors  $Z$  and  $L$  satisfying (2.2).

1. **for**  $k = 1 : n$  **do**
  2.   Set  $z_k^{(0)} = e_k$
  3. **end do**
  4. **for**  $k = 1 : n$  **do**
  5.   Set  $l_{k,k} = \|z_k^{(k-1)}\|_C$
  6.   Set  $z_k = z_k^{(k-1)} / l_{k,k}$
  7.   **for**  $j = k + 1 : n$  **do**
  8.     **if** MGS
  9.       Set  $l_{k,j} = \langle z_j^{(k-1)}, z_k \rangle_C$
  10.    **else if** CGS
  11.      Set  $l_{k,j} = \langle z_j^{(0)}, z_k \rangle_C$
  12.    **else if** AINV
  13.      Set  $l_{k,j} = \langle z_j^{(k-1)}, e_k \rangle_C / l_{k,k}$
  14.    **end if**
  15.    Set  $z_j^{(j)} = z_j^{(j-1)} - l_{k,j} z_k$
  16.    **end do**
  17. **end do**
- 

For large sparse matrices, we need to consider incomplete (approximate) factorizations. Here and elsewhere, we let  $\tilde{L} = \{\tilde{l}_{i,j}\}$  and  $\tilde{Z} = \{\tilde{z}_{i,j}\}$  denote incomplete factors such that

$$\tilde{Z}^T C \tilde{Z} \approx I, \quad I \approx \tilde{L}^T \tilde{Z}. \quad (2.3)$$

These incomplete factors may be used as preconditioners for the conjugate gradient (CG) method. Two different types of preconditioner can be obtained by carrying out the  $C$ -orthogonalization process incompletely. The first approach drops small entries from the computed vectors as the  $C$ -orthogonalization proceeds, that is, after line 11 of Algorithm 2.1 entries that are smaller in absolute value than some chosen drop tolerance are discarded. Alternatively, a relative drop tolerance can be used. Whatever dropping strategy is used, the result is an incomplete inverse factorization of the form

$$C^{-1} \approx \tilde{Z} \tilde{Z}^T.$$

This is a factored sparse approximate inverse and is known as the stabilized approximate inverse (SAINV) preconditioner. The diagonal entries are positive and so the preconditioner is guaranteed to be positive definite.

The second approach (the RIF preconditioner) is obtained by saving the multipliers  $\tilde{l}_{i,j}$  and discarding the computed sparsified column  $\tilde{z}_i$  of  $\tilde{Z}$  as soon as it has been used to form the corresponding parts of the incomplete factor  $\tilde{L}$  of  $C$ , that is, after line 14 of Algorithm 2.1. This gives an algorithm for computing an incomplete Cholesky factorization

$$C \approx \tilde{L} \tilde{L}^T.$$

Again, the preconditioner is guaranteed to be positive definite. Benzi and Tũma [9] report that, for LS problems, the RIF preconditioner is generally more effective at reducing the number of CG iterations than the SAINV preconditioner and thus it is the one we consider.

## 2.2 Implementing left- and right-looking approaches

We now consider the basic data structures and techniques used to implement the left- and right-looking approaches. The left-looking algorithm (Algorithm 2.1) computes the (sparse) columns of  $Z$  one-by-one. This requires knowledge of which indices  $j$  give nonzero  $C$ -inner products  $l_{k,j}$ ; this is discussed in Sections 2.3 and 2.4 below. The columnwise computation allows an approximation  $\tilde{Z}$  to be computed by limiting the number of nonzeros that are retained in each column or by discarding entries with respect to a drop tolerance. To compute the  $l_{k,j}$ ,  $C$  is accessed by columns. Note that the AINV variant requires a single column at a time. The actual computation of  $l_{k,j}$  is straightforward and is based on standard sparse matrix-vector products and sparse dot products. These multipliers form the RIF factor stored by rows.

For the right-looking algorithm (Algorithm 2.2), determining the nonzero  $l_{k,j}$  in the loop starting at line 7 is straightforward. Even in the most involved MGS variant, it reduces to a sparse product of the vector  $z_k$  with the matrix  $C$ . The result is then compared with the matrix whose columns comprise the vectors  $z_j^{(k-1)}$  for  $j = k + 1$  to  $n$ . For efficiency, this matrix needs to be held as a sparse matrix. The amount of fill-in is typically modest and the sparse rank-one updates can often be fast. But a data structure that allows fast dynamic operations is needed. The Benzi and Tůma right-looking implementation [8, 9] employs the data structure described in [36, 41] (and which is used in the early sparse direct solvers MA28 [17] and Y12M [41]). While developments in sparse matrix technologies (see, for example, [18, 24]), as well as the advent of block methods, means that such dynamic schemes have long been superseded in sparse direct solver software, the same is not true for incomplete factorization algorithms. This is at least partly because their significantly smaller memory demands keep right-looking schemes viable. At the start of the computation, three arrays are allocated. The compressed row indices, column indices and matrix values in individual columns are held as sections within in these arrays with some additional space. This is gradually filled and, at each step  $j$  of the factorization, for each array the memory used is recorded to be the first unused entry in the array plus the size of the  $j$  computed columns of the lower triangular factor. At some step, the size an array may be found to be insufficient. In this case, it is reallocated to be larger and the data in the old array that is still required (which may have become fragmented) is copied to the front of the new array and the memory pointer reset; this is reflected by a sharp drop in the reported memory. This reallocation process may need to be repeated several times as the memory required is not known a priori. Note there are no drops in the memory usage as the factorization proceeds if the originally array is sufficiently large but because of fragmentation, using such a large array will result in more memory being used than is necessary.

## 2.3 Existing strategies for exploiting sparsity

The practical success of solvers based on factorizing large sparse matrices crucially depends on exploiting sparsity. Over the last forty or more years, many sophisticated techniques have been developed for sparse direct solvers. For incomplete factorizations, far less has been done. This is at least partially because of the relative simplicity of many incomplete factorization schemes. However, the RIF approach is not straightforward to implement and the computational schemes in the original papers [5, 6, 8] do not discuss the exploitation of sparsity in depth.

To take advantage of sparsity in the left-looking approach, we first observe that the inner products at line 5, 7 and 9 of Algorithm 2.1 involve matrix-vector products and, if  $C$  is sparse, most of these inner products are zero and so the corresponding update operation at line 11 of Algorithm 2.1 can be skipped. The most crucial step from the point of view of exploiting sparsity is thus the determination of which inner products are nonzero. Once known, it is straightforward to use sparsity in the other steps of the algorithm.

In an early left-looking implementation of Algorithm 2.1, determining the nonzero inner products was interleaved with the actual numerical updates and was based on an efficient search of columns of  $C$ . For least squares problems,  $C$  may not be available explicitly and, in this case, such searching is unacceptably slow. Consequently, a right-looking approach was used by Benzi and Tůma [9] but, as already observed,



this has the major disadvantage of not being memory-limited. Thus we want to develop a left-looking limited memory approach that is able to exploit sparsity when  $C$  is stored implicitly.

There are two basic graph-based strategies to determine the nonzero inner products in Algorithm 2.1 (with no dropping). The most straightforward (which we will call Strategy I) is based on the sparsity pattern of  $Z$ , which is known theoretically. The following result was shown in [7, 12] (and is a consequence of [22]).

**Lemma 2.1.** *Assume there is no cancellation in the factorization process. Then  $z_{k,j} \neq 0$  if and only if  $j$  is an ancestor of  $k$  in the elimination tree of  $C$ .*

Consequently, for each  $k$ , the sparsity pattern of column  $z_k$  can be found using the elimination tree. Once the pattern is known, which inner products should be evaluated can be determined; that is, the required inner products are flagged by a search through a submatrix of  $C$  that is determined by the symbolic structure of  $z_k$ . A bound on the complexity of the corresponding symbolic procedure for the mixed Gram-Schmidt method (AINV), for example, is given by the following straightforward result.

**Lemma 2.2.** *For Strategy I and a given  $k \geq 2$ , the complexity to determine all indices  $j$ ,  $j < k$  of the AINV nonzero inner products at line 9 of Algorithm 2.1 is bounded by  $O(\text{nnz}(\sum_{\{j|j \in \text{anc}(k)\}} C_{1:k-1,j}))$ , where  $j \in \text{anc}(k)$  denotes  $j$  is an ancestor of  $k$  in the elimination tree of  $C$ .*

However, Strategy I has a serious flaw: it typically results in many more inner products being evaluated than is necessary, even in exact arithmetic. This is illustrated by Bridson and Tang [13]. The reason for this over determination is that the structure of  $z_k$  is just the final structure of  $z_k^{(k-1)}$  and this is generally a superset of the patterns that are considered in the minor steps of the algorithm. Therefore, many of the inner products determined using Strategy I will still be zero.

A significant enhancement can be achieved by using the relation between  $L$  and  $Z$  given in [28] and which has been considered in [13]; we will refer to it here as Strategy II. It is based on the fact that the inner products used to update column  $z_k^{(j-1)}$  are the entries of the  $k$ th row of  $L$ . Assuming no dropping, we can formulate this observation in terms of the elimination tree and its row subtrees [34], which can be cheaply computed on-the-fly.

**Lemma 2.3.** *The inner product  $\langle z_k^{(j-1)}, z_j \rangle_C \equiv \langle z_k^{(j-1)}, e_j \rangle_C$  is nonzero if and only if  $j$  belongs to the  $k$ th row subtree  $T_r(k)$  of the elimination tree of  $L$ .*

Based on this characterization, we have the following complexity result for the AINV Gram-Schmidt process (but note that, in the case of exact arithmetic, the MGS, CGS and AINV variants are equivalent).

**Lemma 2.4.** *Suppose there is no cancellation in the factorization process. Then for a given  $k \geq 2$ , the complexity to determine all indices  $j < k$  of the AINV nonzero inner products at line 9 of Algorithm 2.1 is bounded by  $O(\text{nnz}(\sum_{\{j|j \in T_r(k)\}} C_{1:k-1,j}))$ .*

## 2.4 Exploiting sparsity for implicit $C$ and incomplete factorizations

If dropping is used (incomplete factorization), the row subtrees are no longer useful and we need an alternative approach. Consider the following approximate bordering scheme for a symmetric positive definite matrix  $C_{k-1}$  that is extended symmetrically by appending one row and column  $c_k$ :

$$\begin{pmatrix} C_{k-1} & c_k \\ c_k^T & \gamma_k \end{pmatrix} \approx \begin{pmatrix} \tilde{L}_{k-1} & \\ \tilde{l}_k^T & \tilde{\lambda}_k \end{pmatrix} \begin{pmatrix} \tilde{L}_{k-1}^T & \tilde{l}_k \\ & \tilde{\lambda}_k \end{pmatrix}. \quad (2.4)$$

The off diagonal entries  $\tilde{l}_k$  of the new column of  $\tilde{L}^T$  satisfy  $\tilde{l}_k = \tilde{L}_{k-1}^{-1} c_k$ . To obtain the sparsity pattern of  $\tilde{l}_k$ , recall the following lemma of Gilbert [22] that holds for any triangular matrix (and thus in particular for the incomplete factor  $\tilde{L}_{k-1}$ ).

**Lemma 2.5.** *The sparsity structure of  $\tilde{l}_k$  is equal to the subset of vertices of the directed acyclic graph of  $\tilde{L}_{k-1}$  that is reachable by directed paths from the nonzeros of  $c_k$  (that is,  $\text{Reach}_{k-1}(\text{Struct}(c_k))$ ).*

Employing this result in Algorithm 2.1, the sparsity pattern of row  $k$  of  $\tilde{L}$  determines the nonzero inner products that should be evaluated by the Gram-Schmidt process and used to obtain the numerical values of the entries in the row. There are two main problems connected with this but, as we now show, they can both be overcome. First, we do not have the directed acyclic graph  $dag(\tilde{L})$  readily available since Algorithm 2.1 computes  $\tilde{L}$  by rows but the efficient computation of the pattern of  $\tilde{L}$  needs access by columns. To deal with this,  $dag(\tilde{L})$  is computed on-the-fly. Note that in the case of a complete factorization this is unnecessary since the elimination tree is available and no other graph structure is needed. Second,  $dag(\tilde{L})$  may have more edges than are needed to generate the row structure. The graph with the smallest number of entries needed to generate the row structure is the transitive reduction [1] and for a dag it is unique. In general, finding the transitive reduction is expensive; instead, a cheap preprocessing called pruning is used. Pruning aims to remove edges that are both cheap to find and are redundant in preserving the set of paths in  $dag(\tilde{L})$ . We note that a similar mechanism for pruning based on structural symmetry in LU factorizations was proposed by Eisenstat and Liu [21].

The first simple pruning approach that we consider is based on the following lemma.

**Lemma 2.6.** *Let  $\tilde{L}$  be a lower triangular matrix with graph  $dag(\tilde{L})$ . Assume that for some  $i < j < k$ ,  $\tilde{l}_{k,i} \neq 0$  and  $\tilde{l}_{k,j} \neq 0$ ,  $\tilde{l}_{j,i} \neq 0$ . Then the set of paths in  $dag(\tilde{L})$  stays the same if  $\tilde{l}_{k,i}$  is set to 0.*

*Proof.* Because of the edges  $(k, j)$  and  $(j, i)$ , we have the path  $k \Rightarrow i$  and so the edge  $(k, i)$  can be pruned.  $\square$

Figure 2.1 gives an illustration of Lemma 2.6. Its assumptions imply the edge  $(i, k)$  can be pruned from the dag.

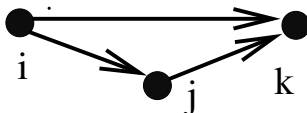


Figure 2.1: Simple illustration of Lemma 2.6. The edge  $(i, k)$  can be pruned.

Algorithm 2.3 applies the test in Lemma 2.6 successively to vertices of  $dag(\tilde{L})$  to get the pruned  $dag_p(\tilde{L})$ . The algorithm is illustrated using the example given in Figure 2.2.

---

**Algorithm 2.3.** *Simple pruning algorithm.*

**Input:** Lower triangular matrix  $\tilde{L} \in R^{n \times n}$ .

**Output:** Pruned graph  $dag_p(\tilde{L})$  with vertex set  $V$  and edge set  $E$ .

1. Set  $V = \{1\}$ ,  $E = \emptyset$
2. Set  $PREV(1 : n) = i$
3. **for**  $k = 2 : n$
4.     Set  $V = V \cup \{k\}$
5.     Obtain the sparsity structure  $J_k = \{j_1, \dots, j_p\}$  of  $\tilde{L}_{k,1:k-1}$  with  $1 \leq j_1 < \dots < j_p \leq k-1$
6.     Set  $E' = \{(k, j_1)\} \cup \dots \cup \{(k, j_p)\}$
7.     **for**  $i = 1 : p$
8.         **if**  $PREV(j_i) = i$  **or**  $(k, PREV(j_i)) \notin E'$  **then**
9.             Set  $E = E \cup \{(k, j_i)\}$
10.             $PREV(j_i) = k$
11.         **end if**
12.     **end for**
13. **end for**



```

19.     end for
20.     if found = false
21.         Set  $E = E \cup \{(k, j_i)\}$ 
22.     end if
23. end for
24. end for
25. end for

```

---

It is straightforward to see that after applying Algorithm 2.4 to the lower triangular matrix given Figure 2.3,  $\text{dag}_p(\tilde{L})$  has  $O(n)$  nonzero entries but that it is still not transitively reduced since the worst case complexity of this reduction based on the appropriate reachability sets is in general  $O(n^2)$ . The following result describes the complexity of Algorithm 2.4.

**Lemma 2.7.** *Assume that the number of nonzero entries in each column of  $\tilde{L}$  or in each row of  $\tilde{L}$  is bounded by  $lsize$ . Then the number of comparisons in Algorithm 2.4 is bounded by  $O(n \cdot lsize^2)$ .*

*Proof.* The comparisons in the algorithm imply that for each nonzero entry  $\tilde{l}_{i,j}$ , all other row indices  $i_k$  such that  $\tilde{l}_{i_k,j} \neq 0$  and  $j < i_k < i$  should be tested. If each column of  $\tilde{L}$  has at most  $lsize$  entries then  $\text{nnz}(\tilde{L}) \leq n \cdot lsize$  and for each entry at most  $lsize$  comparisons are performed. If each row of  $\tilde{L}$  has at most  $lsize$  entries then for each of its  $n$  rows we need to compare only entries in those rows that correspond to its nonzero entries. The maximum number of such rows is  $lsize$  and the result follows.  $\square$

Note that the above bound is asymptotically the same as the number of operations needed by the exact Cholesky factorization with the same upper bound for the number of entries in its columns. Our case of the incomplete factorization based on inner products and different data structures is not exactly comparable but it is clear that the complexity of pruning specified in Lemma 2.7 is affordable. Further additional dropping in  $\tilde{L}$  based on the magnitudes of the computed inner products can be applied, typically for each  $k$  once the structure of column  $\tilde{l}_k^T$  has been computed.

### 3 Implicit left-looking RIF for LS problems

While the previous section looked at exploiting sparsity in the left-looking RIF algorithm for general symmetric positive matrices, here we focus on its application to the solution of LS problems. In Algorithm 3.1, we present an outline of our implicit left-looking RIF algorithm for LS problems that avoids computing the normal matrix  $C = A^T A$  explicitly. Algorithm 3.1 includes scaling and the optional use of a nonzero shift  $\alpha$  so that the RIF factorization of  $S(C + \alpha I)S$  is computed where the diagonal matrix  $S$  is an  $n \times n$  column scaling matrix.

---

**Algorithm 3.1.** *Left-looking RIF algorithm for LS with  $C = A^T A$  held implicitly.*

**Input:**  $A \in R^{m \times n}$  with full column rank, a shift  $\alpha \geq 0$ , and drop tolerance  $\tau > 0$ .

**Output:** Incomplete RIF factor  $\tilde{L}$  (stored by rows).

1. Compute a column scaling  $S$  and scale:  $A \leftarrow AS$
2. Set  $\tilde{L}_{1,:} = (1 + \alpha)\sqrt{a_1^T a_1}$
3. Set  $z_1 = e_1$
4. **for**  $k = 2 : n$  **do**
5.     Set  $z_k^{(0)} = e_k$
6.     Let  $c_k = A^T A_{1:k-1,k}$
7.     Compute the sparsity structure  $J_k$  of  $\tilde{L}_{k-1}^{-1} c_k$  as  $\text{Reach}_{k-1}(\text{Struct}(c_k))$ .
8.     Prune  $J_k$  using Algorithm 2.3 or 2.4 to get  $J_k = \{j_1 < \dots < j_p\}$  with  $p \leq k - 1$ ; set  $j_0 = 0$

```

9.   for  $s = 1 : p$  do
10.     $j = j_s$ 
11.    if MGS
12.        Set  $v_j = z_j$  and  $u_k = z_k^{(j_{s-1})}$ 
13.    else if CGS
14.        Set  $v_j = e_j$  and  $u_k = z_k^{(0)}$ 
15.    else if AINV
16.        Set  $v_j = e_j$  and  $u_k = z_k^{(j_{s-1})}$ 
17.    end if
18.    Compute  $p_j = Av_j$ ,  $q_k = Au_k$  and, if  $\alpha > 0$ ,  $\beta_k = v_j^T S^2 u_k$ 
19.    if  $p_j^T q_k + \alpha \beta_k > \tau$  do
20.        Set  $\tilde{l}_{k,j} = p_j^T q_k + \alpha \beta_k$ 
21.        Set  $z_k^{(j)} = z_k^{(j_{s-1})} - \tilde{l}_{k,j} z_j$ 
22.        Discard all components of  $z_k^{(j)}$  less than  $\tau$  in absolute value
23.    end if
24.    end do
25.    Set  $\tilde{l}_{k,k} = \sqrt{(Az_k^{(j_p)})^T (Az_k^{(j_p)})}$ 
26.    Set  $z_k = z_k^{(j_p)} / \tilde{l}_{k,k}$ 
27. end do

```

Observe that since the algorithm treats  $C = A^T A$  implicitly, the shift is performed within the main loop and the scaling is applied to the shifted entries. In some applications in optimization, such as the Levenberg-Marquardt method for solving nonlinear least squares, a non-zero shift is used (see, e.g., [16]). However, if  $\alpha = 0$ , the algorithm to compute the RIF factorization can be significantly accelerated, as we state in the following proposition.

**Proposition 3.1.** *If the shift  $\alpha$  is equal to zero then the computed  $v_j$  do not need to be stored. Instead, if  $v_j = z_j$  (MGS variant,) the matrix-vector products  $p_j = Av_j$  can be precomputed and stored (that is, once  $z_k$  is computed at line 23,  $Az_k$  may be computed and stored). Note that the case  $v_j = e_j$  (CGS and AINV variants) trivially uses a column of  $A$ , that is,  $p_j = a_j$ .*

## 4 Numerical experiments

### 4.1 Test environment

Most of the test problems used in our experiments are taken from the University of Florida Sparse Matrix Collection [15]. The exceptions are a Laplacian test example, problem IPROB (which is part of the CUTEst linear programme set [25]), and problem PIGS\_large1 from a pig breeding application (see [3] for details). In our experiments on symmetric positive definite linear systems  $Cx = b$ , the right-hand side vector  $b$  is computed so that the exact solution is  $x = 1$ , and the stopping criteria used for preconditioned conjugate gradients (PCG) is

$$\|Cx - b\|_2 \leq 10^{-6} \|b\|_2. \quad (4.5)$$

We define the *density ratio* of the computed incomplete factor  $\tilde{L}$  to be

$$\rho = \text{nnz}(\tilde{L}) / \text{nnz}(C).$$

For the least squares tests, we use preconditioned conjugate gradients for the normal equations (CGNE). We employ the following stopping rules that are taken from [26]:

C1: Stop if  $\|r\|_2 < \delta_1$

C2: Stop if

$$\frac{\|A^T r\|_2}{\|r\|_2} < \frac{\|A^T r_0\|_2}{\|r_0\|_2} * \delta_2,$$

where  $r = Ax - b$  is the residual,  $r_0$  is the initial residual and  $\delta_1$  and  $\delta_2$  are convergence tolerances that we set to  $10^{-8}$  and  $10^{-6}$ , respectively. In all our experiments, we take the initial solution guess to be  $x_0 = 0$  and in this case C2 reduces to

$$\frac{\|A^T r\|_2}{\|r\|_2} < \frac{\|A^T b\|_2}{\|b\|_2} * \delta_2.$$

We define the density ratio of the computed incomplete factor  $\tilde{L}$  to be

$$\rho = nnz(\tilde{L})/nnz(A).$$

All the reported experiments use with the MGS variant of the left- and right-looking RIF algorithms. The experiments were performed on an Intel(R) Core(TM) i5-4590 CPU running at 3.30 GHz with 12 GB of internal memory. The codes were written in Fortran and the Visual Fortran Intel(R) 64 XE compiler (version 14.0.3.202) was used.

## 4.2 The case for pruning

Our first experiment, which we report on in Figure 4.4, demonstrates the need for pruning. Results are given for two examples: a two dimensional Laplacian matrix of dimension 25,00 and problem Pothen/bodyy4 ( $n = 17,546$ ,  $nnz(C) = 121,550$ ). In each case, we repeatedly increase the density ratio  $\rho$  by decreasing the drop tolerance  $\tau$  used in the incomplete factorization (30 different tolerances are used for the Laplacian and 11 for Pothen/bodyy4). We see that Algorithm 2.3 is highly effective in limiting the growth in the number of edges in  $dag_p(\tilde{L})$  but that some further reductions are possible using the more sophisticated Algorithm 2.4, particularly as the number of entries in  $\tilde{L}$  increases. Note that we get exactly the same preconditioner (that is, the same non zero entries) with pruning as without pruning.

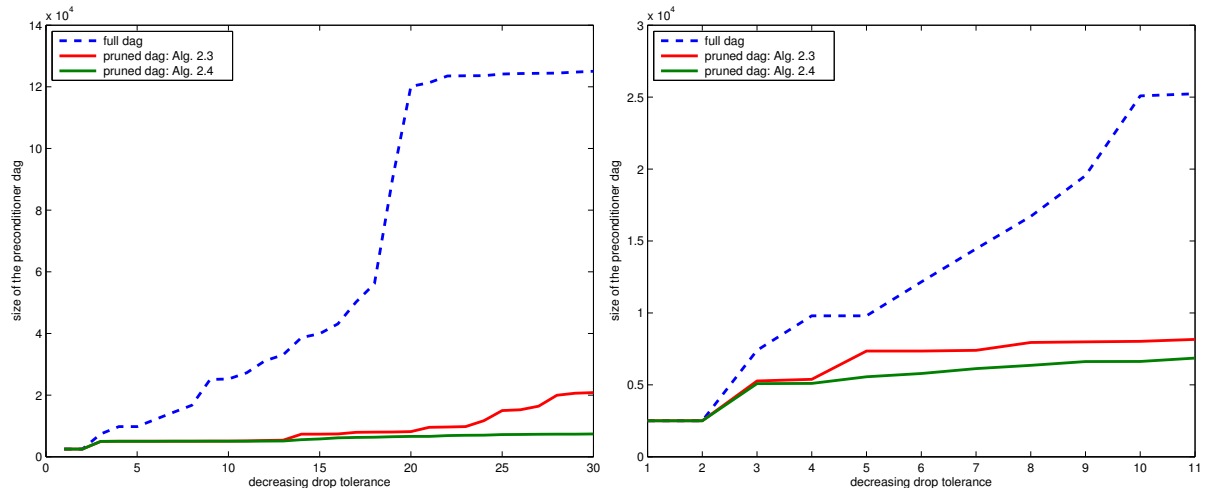


Figure 4.4: Effect of the two pruning approaches (Algorithms 2.3 and 2.4) on the two dimensional Laplacian matrix (left) and on matrix Pothen/bodyy4 (right). We depict here the dependence of the number of edges in the auxiliary dag on the drop tolerance.

### 4.3 Memory management for left- and right-looking approaches

Our second experiment is designed to demonstrate the principal differences in the memory management of the left- and right-looking approaches. Figure 4.5 depicts the reported memory for three preconditioners of different densities for problem Nasa/nasa1824 ( $n = 1,824$ ,  $nnz(C) = 39,208$ ). The density ratios are  $\rho = 1.2, 1.6$  and  $2.8$ , respectively. The PCG iteration counts are 26, 18 and 8, respectively. For the right-looking code, the reported memory is as described in Section 2.2. For the left-looking code, it is the size of the approximate inverse factor  $\tilde{Z}$  plus that of the preconditioner  $\tilde{L}$ . We can see that as the density increases, the influence of the memory to store  $\tilde{Z}$  becomes more significant. But, in general, in our experience this occurs when the preconditioner is too large to be practical.

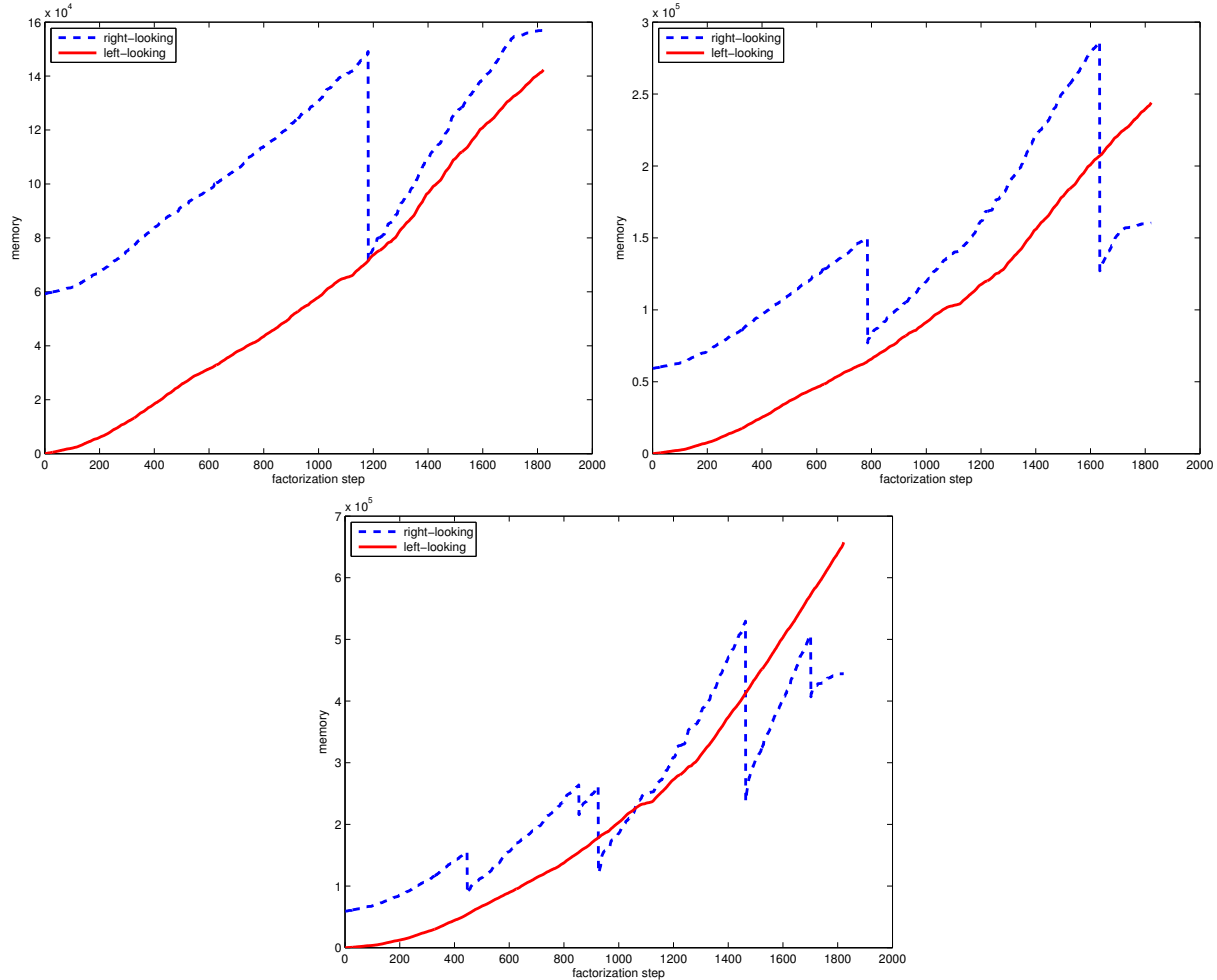


Figure 4.5: Memory comparison for the left- and right-looking RIF approaches for Nasa/nasa1824 for three different density ratios.

Similar results for problems Nasa/nassarb and GHS\_psdef/hood are given in Figures 4.6 and 4.7, respectively. For Nasa/nassarb,  $\rho$  increases from approximately 5 to nearly 50 and the PCG iteration count decreases from 194 to 43. For GHS\_psdef/hood, the densities are 0.5 and 1.5 and the corresponding PCG iteration counts are 224 and 51. GHS\_psdef/hood illustrates large but not atypical differences between the right- and left-looking memory demands.

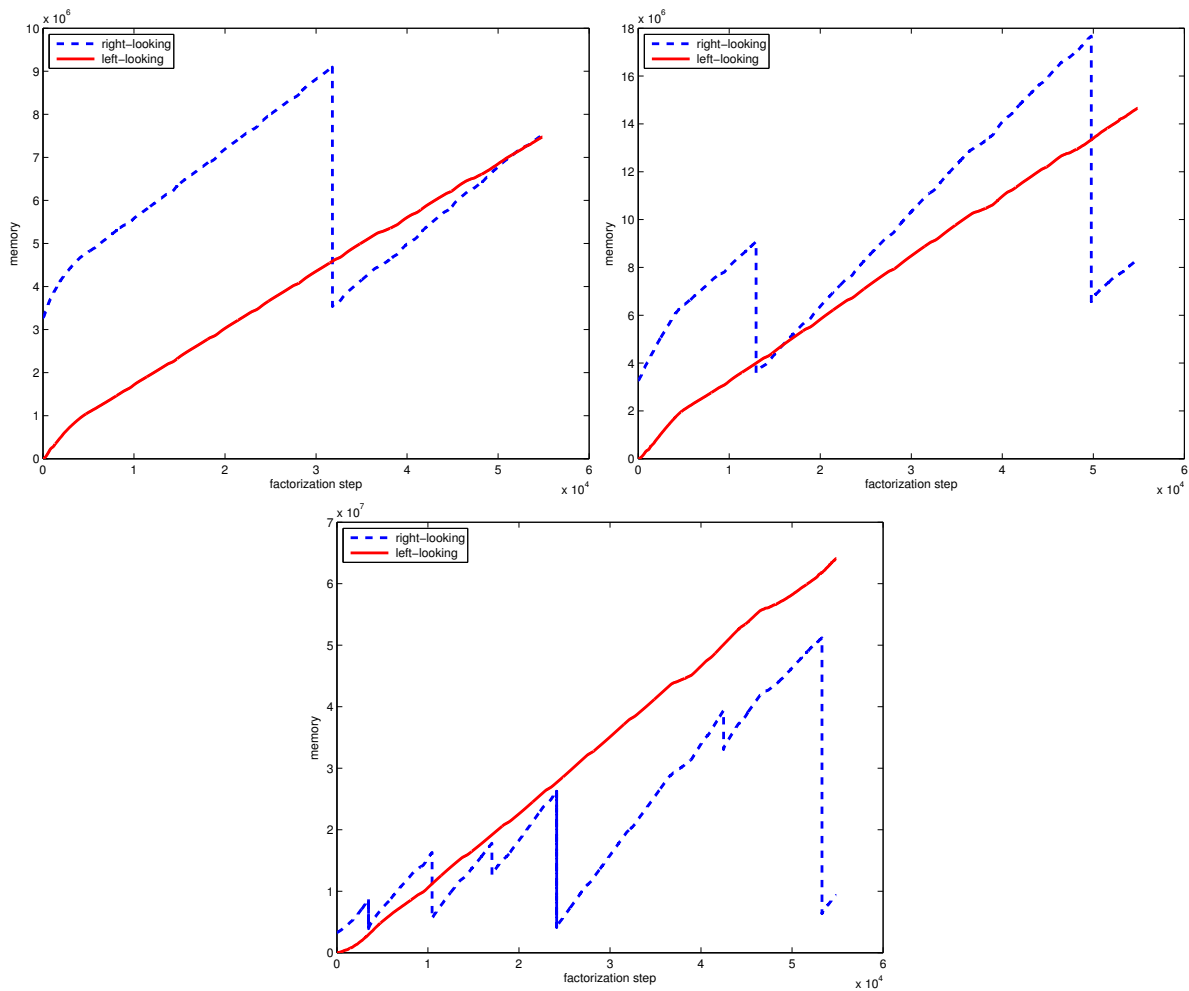


Figure 4.6: Memory comparison for the left- and right-looking RIF approaches for problem Nasa/nasarb for three different density ratios.

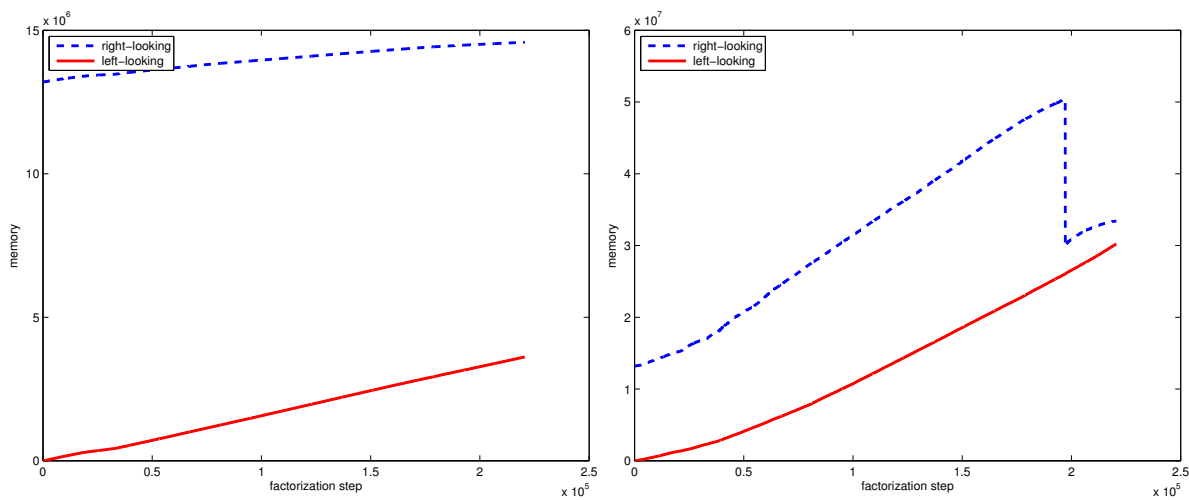


Figure 4.7: Memory comparison for the left- and right-looking RIF approaches for GHS\_psdef/hood for two different density ratios.



## 4.4 Least-squares problems

We now explore some differences between the left- and right-looking approaches when used for solving the normal equations (1.2). Our first example is a small square matrix IPROB of order  $n = 3001$ . For this example,  $A^T A$  is dense, with  $nnz(A^T A) \approx 9 \times 10^6$ . The RIF preconditioner is of size  $nnz(\tilde{L}) \approx 2.5 \times 10^6$  and  $nnz(\tilde{Z}) \approx 8.6 \times 10^4$ ; three CGNE iterations are required for convergence. The plot on the left in Figure 4.8 shows that both the left- and right-looking algorithms have similar memory requirements if the size of  $A$  is not taken into account. However, in the plot on the right we include the size of  $A^T A$  for the left-looking approach and of  $A$  and  $A^T$  for the right-looking approach ( $A$  must be held by both rows and columns) and we now see that the explicit computation of the normal matrix  $A^T A$  can result in an unacceptable overhead for the left-looking algorithm.

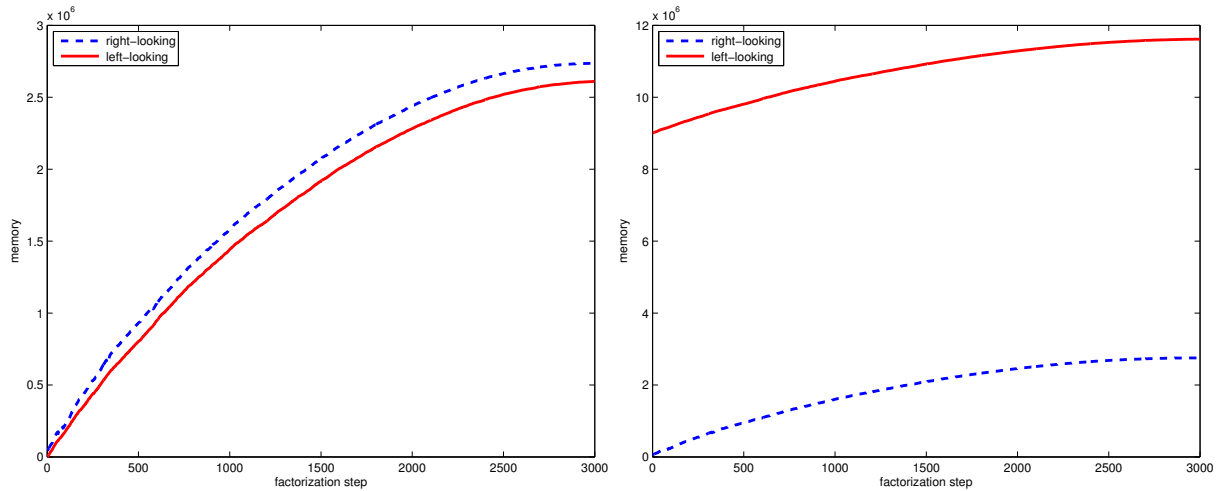


Figure 4.8: Memory comparison for the left- and right-looking RIF approaches for problem IPROB for which the normal equations  $A^T A$  are dense. The plot on the left does not take into account the memory for input matrix  $A$ . The plot on the right adds to the explicit left-looking approach the size of  $A^T A$  and adds  $nnz(A) + nnz(A^T)$  to the right-looking approach.

In many applications,  $m$  is significantly larger than  $n$  and consequently  $nnz(A^T A)$  can be smaller than  $nnz(A) + nnz(A^T)$ . For problem LPnetlib/lp\_osa\_30 ( $m = 104,374$ ,  $n = 4,350$  and  $nnz(A) = 604,488$ ), Figure 4.9 shows the RIF memory requirements for a sparse preconditioner (here the density ratio  $\rho$  is approximately 0.5). The corresponding comparison for an increased density ratio of 0.8 is given in Figure 4.10. Note the sudden drops in the memory for the right-looking approach indicates memory reallocations are now needed; on each reallocation the size of the array used is doubled. We also observe that for this example the factors fill in significantly towards the end of the factorization resulting in a sudden increase in the memory for the left-looking RIF where the memory is dominated by  $\tilde{L}$  (for the right-looking approach the total memory dominates that of  $\tilde{L}$  so the increase from the fill in is not seen). In Figure 4.11, we plot memory usage for problem Yoshiyasu/mesh\_deform ( $m = 234,023$ ,  $n = 9393$  and  $nnz(A) = 853,829$ ); the density ratio is approximately 4. Again, for this example with  $m \gg n$ , the left-looking approach is considerably more memory efficient than the right-looking one.

## 4.5 Explicit versus implicit left-looking approaches

We next consider the differences between the left-looking algorithm that explicitly forms the normal matrix  $C = A^T A$  and the new implicit dag-based variant that avoids forming  $C$ . Note that both result in preconditioners of a similar quality and lead to essentially the same CGNE iteration counts. In some practical applications it may not be possible to form  $C$ , in which case the only computational possibility

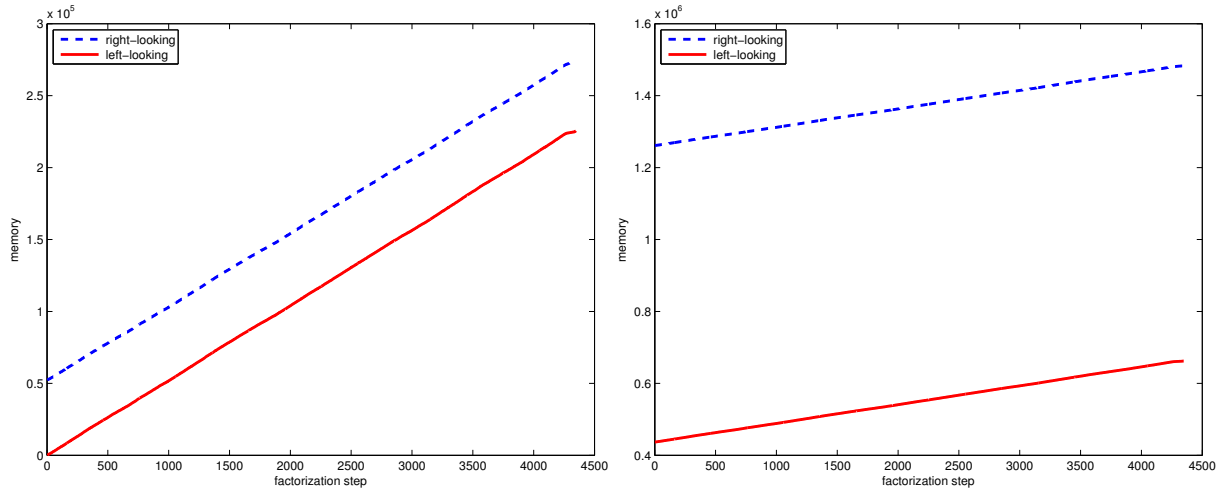


Figure 4.9: Memory comparison for the left- and right-looking RIF approaches applied to the normal equations for problem LPnetlib/lp\_osa\_30 with  $\rho = 0.5$ . The plot on the left does not take into account the input matrix  $A$ . The plot on the right adds to the explicit left-looking approach the size of  $A^T A$  and adds to the right-looking approach the sizes of  $A$  and  $A^T$ .

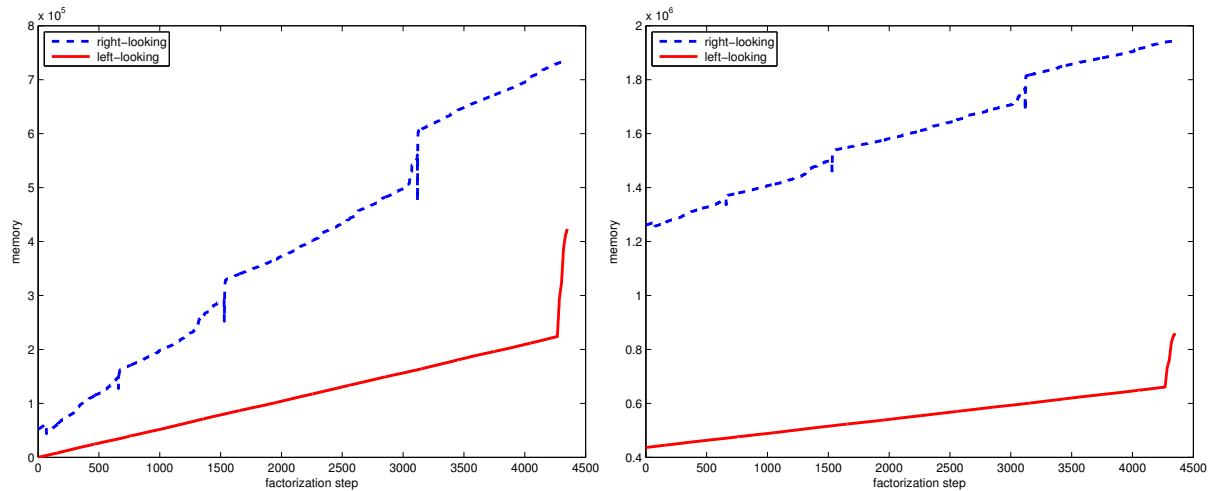


Figure 4.10: Memory comparison for the left- and right-looking RIF approaches applied to the normal equations for problem LPnetlib/lp\_osa\_30 with  $\rho = 0.8$ . The plot on the left does not take into account the input matrix  $A$ . The plot on the right adds to the explicit left-looking approach the size of  $A^T A$  and adds to the right-looking approach the sizes of  $A$  and  $A^T$ .

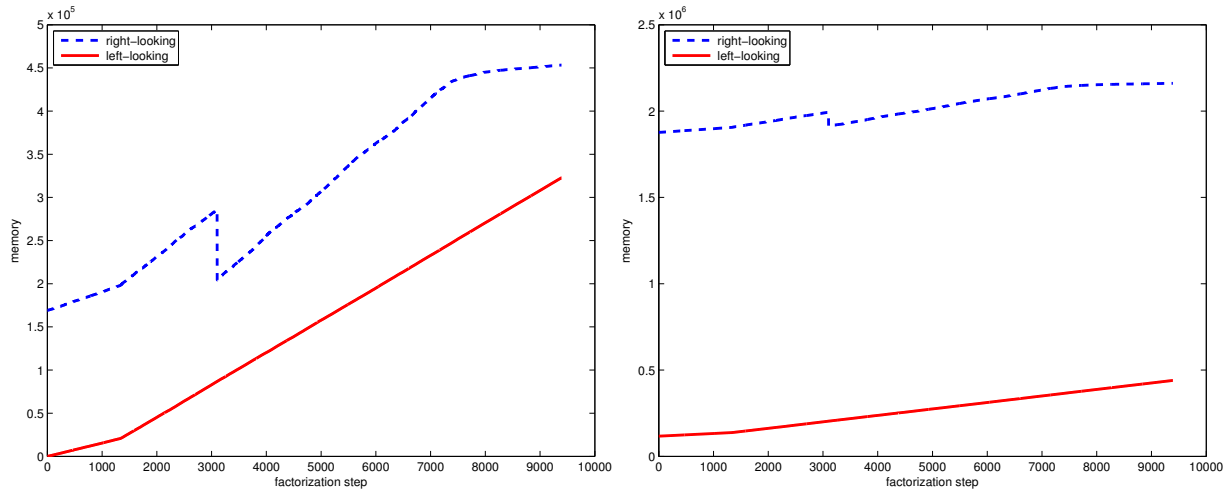


Figure 4.11: Memory comparison for the left- and right-looking RIF approaches applied to the normal equations for problem Yoshiyasu/mesh\_deform with  $\rho = 4$ . The plot on the left does not take into account the input matrix  $A$ . The plot on the right adds to the explicit left-looking approach the size of  $A^T A$  and adds to the right-looking approach the sizes of  $A$  and  $A^T$ .

with limited memory is the dag-based approach. But this is also the method of choice if  $nnz(A^T A) \gg nnz(A)$ , as is illustrated in Table 4.1 by our reported results for problems IPROB and Bydder/mri2. Note that for these examples, pruning (using Algorithm 2.4) leads to a significant reduction in the number of edges in the dag. However, if  $C$  can be formed and stored and is sufficiently sparse, the explicit algorithm can be much faster than the implicit one. This happens, in particular, if  $\tilde{L}$  is not very sparse. This is demonstrated by problems PIGS\_large1 and LPnetlib/lp\_ken.18. But note that the experiments for LPnetlib/lp\_osa30 show that even with relatively sparse  $A^T A$  the implicit algorithm is asymptotically better. If  $\tilde{L}$  is made denser by using a larger drop tolerance then the situation is more complicated. While the cost of the implicit algorithm increases because of the dag searches, the explicit version suffers from the dense  $\tilde{L}$  as well since it is involved in the searches for nonzero dot products. The final three problems in Table 4.1 are known to be challenging (see [26]) and for these relaxed stopping tolerances  $\delta_1 = 10^{-5}$  and  $\delta_2 = 10^{-3}$  are used together with a shift  $\alpha = 0.1 * \|(A^T A)\|_F$ .

Table 4.1: A comparison of the explicit and implicit left-looking RIF approaches when used to precondition CGNE. Time is the time (in seconds) to compute the preconditioner; Iters is the number of CGNE iterations; *size\_p* denotes the number of edges in the dag, and before and after indicate before and after pruning. \* denotes stopping tolerances  $\delta_1 = 10^{-5}$  and  $\delta_2 = 10^{-3}$  are used together with a shift  $\alpha = 0.1 * \|(A^T A)\|_F$ .

Identifier	$nnz(A)$	$nnz(A^T A)$	Explicit		Implicit			Iters
			Time	<i>size_p</i>	Time	<i>size_p</i> before	<i>size_p</i> after	
Pereyra/landmark	1151232	120,234	0.22	25,858	2.62	25,867	7,836	50
PIGS_large1	75,018	128,798	0.08	48,307	0.42	48,307	39,395	94
LPnetlib/lp_ken_18	358,171	609,271	3.02	399,026	1.46	410,372	297,614	103
LPnetlib/lp_osa30	604,488	436,738	11.5	220,537	14.8	220,537	104,601	155
JGD_Groebner/f855_mat9*	171,214	4,485,478	33.3	2,220,736	6.03	2,234,595	19,990	799
JGD_Groebner/c8_mat11*	2,462,970	16,899,062	277.	7,896,493	146.	8,137,794	46,427	961
Bydder/mri2*	569,160	31,335,808	569.	7,415,074	129.	10,318,767	164,171	671

While the results for example LPnetlib/lp\_osa30 in Table 4.1 illustrate that the implicit algorithm can be faster than the explicit one, this is dependent on the preconditioner size. This can be seen in Figure 4.12 (left-hand plot). As the density ratio  $\rho$  increases, the time for the implicit approach also increases and eventually it is slower than using the explicitly computed  $C$ . For problem Kemelmacher/Kemelmacher the findings are different (right-hand plot in Figure 4.12). We see here that as  $\rho$  increases, the explicitly computed  $C$  does not prevent the computational time from steadily increasing. Thus, summarizing our experience, we conclude that both the explicit and implicit algorithms can be useful and even when  $C$  is available, deciding which approach will be the computationally most efficient is not a clear choice. Construction of the preconditioner using the auxiliary dag can be considered as a way to a parallel implementation since the implicit algorithm enables the classical generalized Gram-Schmidt algorithm to be used to compute the factor  $\tilde{Z}$ . This offers the potential to significantly enhance the exploitation of parallelism, but a more detailed discussion lies outside the scope of the current study.

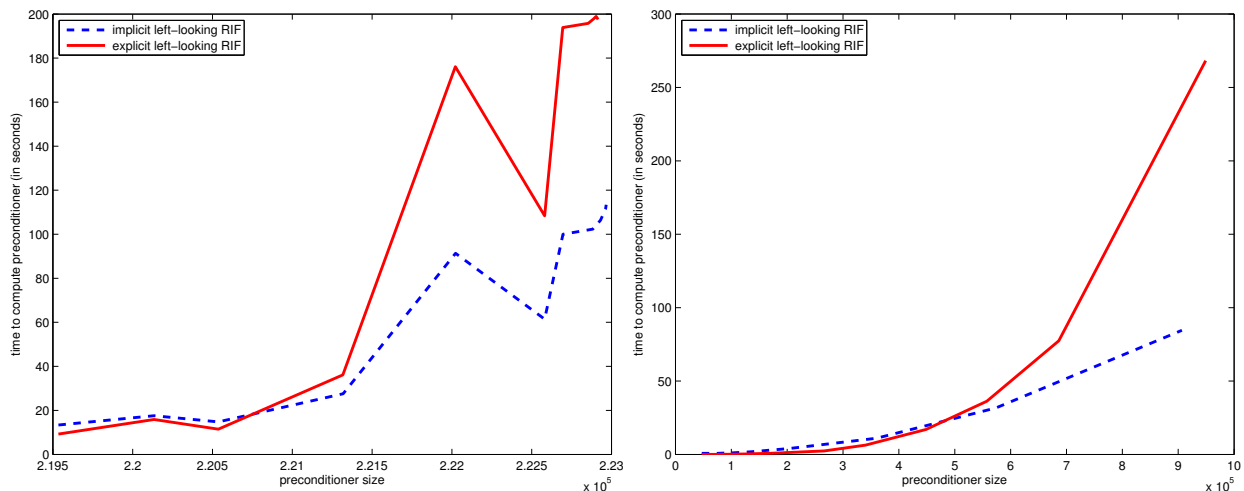


Figure 4.12: Dependence of the time to compute the explicit and implicit left-looking RIF on the size of the preconditioner  $\tilde{L}$  for problems LPnetlib/lp\_osa30 (left) and Kemelmacher/Kemelmacher (right).

## 5 Concluding remarks

In this paper, we have proposed a new left-looking variant of the RIF approach for computing an incomplete  $LL^T$  factorization of a sparse positive-definite matrix and, in particular, the normal equations matrix  $C = A^T A$ . The practical success of solvers for the solution of large sparse problems crucially depends on the efficient and effective exploitation of sparsity. In the case of direct methods, the importance of sparsity is well understood and, over many years, sophisticated techniques have been developed to take advantage of sparsity throughout the factorization. Much less has been done for incomplete factorizations. While this may partly be because of the relative simplicity of many such algorithms, this is definitely not the case for sophisticated schemes such as RIF that need complicated data structures and combine techniques from Gaussian elimination and orthogonalization. For the left-looking RIF algorithm, we have introduced a global symbolic preprocessing step that constructs a directed acyclic graph  $dag(\tilde{L})$  that determines the sparsity pattern of the preconditioner factor  $\tilde{L}$ , without the need to explicitly construct the matrix  $C$ . An efficient pruning algorithm has been proposed to limit the number of edges in  $dag(\tilde{L})$ . Numerical experiments have shown this pruning algorithm to be highly effective.

A fundamental difference between the left- and right-looking RIF approaches is their memory management. For the latter, the memory required is not known a priori and so it can be necessary

to increase the memory available during the factorization and this reallocation may have to be done more than once. Our results have illustrated that, for linear systems, which approach is most memory efficient is not only problem dependent but also depends on the density of the computed  $\tilde{L}$  (that is, on the choice of the dropping parameter). For least squares problems, we additionally need to take into account the number of entries in  $A$  and  $A^T$  or, in the case of the explicit left-looking algorithm, the number of entries in  $C$ . We have found that the performances of the different variants can be vary significantly but for a given problem, without some prior knowledge of the problem and its characteristics, it is not obvious which approach will be most efficient. Clearly, if a series of similar problems must be solved and it is possible to construct and store  $C$ , it may be worthwhile to try both the explicit and implicit approaches and select the most efficient. If it is not possible to explicitly compute  $C$ , then the new implicit variant offers a viable alternative.

While this study was carried out with prototype codes, in the future we plan to develop library-quality implementations that will be made available as part of the HSL software library [29]. We anticipate that these implementations will include options for ordering and scaling of the problem, which can potentially significantly enhance the performance.

## References

- [1] A. V. Aho, M. R. Garey, and J. D. Ullman. The transitive reduction of a directed graph. *SIAM J. on Computing*, 1:131–137, 1972.
- [2] M. Arioli and I. S. Duff. Preconditioning linear least-squares problems by identifying a basis matrix. *SIAM J. on Scientific Computing*, 37:S544–S561, 2015.
- [3] M. Arioli and I. S. Duff. Preconditioning of linear least-squares problems by identifying a basic matrix. *SIAM J. on Scientific Computing*, 37, 2015.
- [4] C. Benoit. Note sur une méthode de résolution des équations normales provenant de l’application de la méthode des moindres carrés a un système d’équations linéaires en nombre inférieur a celui des inconnues. application de la méthode a la résolution d’un système défini d’équations linéaires. *Bulletin Géodésique*, 2:5–77, 1924.
- [5] M. Benzi, J. K. Cullum, and M. Tũma. Robust approximate inverse preconditioning for the conjugate gradient method. *SIAM J. on Scientific Computing*, 22(4):1318–1332, 2000.
- [6] M. Benzi, C. D. Meyer, and M. Tũma. A sparse approximate inverse preconditioner for the conjugate gradient method. *SIAM J. on Scientific Computing*, 17(5):1135–1149, 1996.
- [7] M. Benzi and M. Tũma. Orderings for factorized sparse approximate inverse preconditioners. *SIAM J. on Scientific Computing*, 21(5):1851–1868, 2000.
- [8] M. Benzi and M. Tũma. A robust incomplete factorization preconditioner for positive definite matrices. *Numerical Linear Algebra with Applications*, 10(5-6):385–400, 2003.
- [9] M. Benzi and M. Tũma. A robust preconditioner with low memory requirements for large sparse least squares problems. *SIAM J. on Scientific Computing*, 25(2):499–512, 2003.
- [10] Å. Björck. *Numerical methods for Least Squares Problems*. SIAM, Philadelphia, 1996.
- [11] A. Björck and J. Y. Yuan. Preconditioners for least squares problems by  $LU$  factorization. *Electronic Transactions on Numerical Analysis*, 8:26–35, 1999.
- [12] R. Bridson and W.-P. Tang. Ordering, anisotropy, and factored sparse approximate inverses. *SIAM J. on Scientific Computing*, 21(3):867–882, 1999.
- [13] R. Bridson and W.-P. Tang. Refining an approximate inverse. *J. of Computational and Applied Mathematics*, 123:293–306, 2000.
- [14] R. Bru, J. Marín, J. Mas, and M. Tũma. Preconditioned iterative methods for solving linear least squares problems. *SIAM J. on Scientific Computing*, 36(4):A2002–A2022, 2014.
- [15] T. A. Davis and Y. Hu. The University of Florida sparse matrix collection. *ACM Transactions on Mathematical Software*, 38(1):1–28, 2011.

- [16] J. E. Dennis, Jr. and Robert B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*, volume 16 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996. Corrected reprint of the 1983 original.
- [17] I. S. Duff. MA28—a set of Fortran subroutines for sparse unsymmetric linear equations. Harwell Report, UK AERE-R.8730, Harwell Laboratories, 1980.
- [18] I. S. Duff and J. K. Reid. The multifrontal solution of indefinite sparse symmetric linear equations. *ACM Transactions on Mathematical Software*, 9(3):302–325, 1983.
- [19] S. Eisenstat and J.-W. H. Liu. The theory of elimination trees for sparse unsymmetric matrices. *SIAM J. on Matrix Analysis and Applications*, 26(3):686–705, 2005.
- [20] S. C. Eisenstat, M. C. Gursky, M. H. Schultz, and A. H. Sherman. The Yale Sparse Matrix Package (YSMP) – I : The symmetric codes. *International J. of Numerical Methods in Engineering*, 18:1145–1151, 1982.
- [21] S. C. Eisenstat and J.-W. H. Liu. Exploiting structural symmetry in unsymmetric sparse symbolic factorization. *SIAM J. Matrix Anal. Appl.*, 13(1):202–211, 1992.
- [22] J. R. Gilbert. Predicting structure in sparse matrix computations. *SIAM J. on Matrix Analysis and Applications*, 15:62–79, 1994.
- [23] J. R. Gilbert and J. W. H. Liu. Elimination structures for unsymmetric sparse lu factors. *SIAM J. on Matrix Analysis and Applications*, 14:334–352, 1993.
- [24] J. R. Gilbert and T. Peierls. Sparse partial pivoting in time proportional to arithmetic operations. *SIAM J. on Scientific and Statistical Computing*, 9:862–874, 1988.
- [25] N. I. M. Gould, D. Orban, and Ph. L. Toint. CUTEst: a constrained and unconstrained testing environment with safe threads for mathematical optimization. *Computational Optimization and Applications*, 60:545–557, 2015.
- [26] N. I. M. Gould and J. A. Scott. The state-of-the-art of preconditioners for sparse linear least squares problems. Technical Report RAL-P-2015-010, Rutherford Appleton Laboratory, 2015.
- [27] N. I. M. Gould and J. A. Scott. The state-of-the-art of preconditioners for sparse linear least squares problems: the complete results. Technical Report RAL-TR-2015-009, Rutherford Appleton Laboratory, 2015.
- [28] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. of Research of the National Bureau of Standards*, 49:409–435, 1952.
- [29] HSL. A collection of Fortran codes for large-scale scientific computation, 2013. <http://www.hsl.rl.ac.uk>.
- [30] A. Jennings and M. A. Ajiz. Incomplete methods for solving  $A^T Ax = b$ . *SIAM J. on Scientific and Statistical Computing*, 5(4):978–987, 1984.
- [31] J. Kopal, M. Rozložník, A. Smoktunowicz, and M. Tůma. Rounding error analysis of orthogonalization with a non-standard inner product. *BIT Numerical Mathematics*, 52:1035–1058, 2012.
- [32] P. Läuchli. Jordan-Elimination und Ausgleichung nach kleinsten Quadraten. *Numerische Mathematik*, 3:226–240, 1961.
- [33] N. Li and Y. Saad. MIQR: A multilevel incomplete QR preconditioner for large sparse least-squares problems. *SIAM J. on Matrix Analysis and Applications*, 28(2), 2006.
- [34] J. W. H. Liu. The role of elimination trees in sparse factorizations. *SIAM J. on Matrix Analysis and Applications*, 11(1):134–172, 1990.
- [35] K. Morikuni and K. Hayami. Inner-iteration Krylov subspace methods for least squares problems. *SIAM J. on Matrix Analysis and Applications*, 34(1):1–22, 2013.
- [36] O. Østerby and Z. Zlatev. *Direct methods for sparse matrices*, volume 157 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, 1983.
- [37] A. T. Papadopoulos, I. S. Duff, and A. J. Wathen. A class of incomplete orthogonal factorization methods. II: Implementation and results. *BIT Numerical Mathematics*, 45(1):159–179, 2005.
- [38] G. Peters and J. H. Wilkinson. The least squares problem and pseudo-inverse. *The Computer Journal*, 131:309–316, 1970.

- [39] Y. Saad. Preconditioning techniques for nonsymmetric and indefinite linear systems. *J. of Computational and Applied Mathematics*, 24(1-2):89–105, 1988.
- [40] X. Wang, K. A. Gallivan, and R. Bramley. CIMGS: an incomplete orthogonal factorization preconditioner. *SIAM J. on Scientific Computing*, 18(2):516–536, 1997.
- [41] Z. Zlatev. *Computational Methods for General Sparse Matrices*. Kluwer, Dordrecht, 1991.