Sparse Matrix Orderings for Factorized Inverse Preconditioners

Title:

Author(s):
Michele Benzi
Miroslav Tuma

Submitted to:
Proc. Copper Mountain Conf. on Iterative Methods 1998
Copper Mountain, Colorado
March 30–April 3, 1998

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. The Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.
DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.
DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.
Sparse Matrix Orderings for Factorized Inverse Preconditioners

Michele Benzi* and Miroslav Tůma†

Abstract. The effect of reorderings on the performance of factorized sparse approximate inverse preconditioners is considered. It is shown that certain reorderings can be very beneficial both in the preconditioner construction phase and in terms of the rate of convergence of the preconditioned iteration.

Key Words. Sparse linear systems, sparse matrices, preconditioned Krylov subspace methods, graph theory, orderings, factorized sparse approximate inverses, incomplete biconjugation.

1. Introduction. We consider the solution of sparse linear systems $Ax = b$ by preconditioned iterative methods, where the preconditioner is a sparse approximate inverse of $A$. Such preconditioners have particular interest from the point of view of parallel computation, since their application at each step of an iterative method only requires sparse matrix-vector products, which are relatively easy to parallelize. A comprehensive survey of sparse approximate inverse preconditioners, together with the results of extensive numerical tests aimed at assessing the performance of the various methods, can be found in [5]. One of the conclusions of that study was that factorized forms, in which the approximate inverse is the product of two sparse triangular matrices, tend to perform better than nonfactorized ones, in the sense that they tend to deliver better convergence rates for the same amount of nonzeros in the preconditioner. Factorized approximate inverses are also less expensive to compute than other forms, at least in a sequential environment. As mentioned in [5], another potential advantage of the factorized approach is the fact that such preconditioners are sensitive to the ordering of the unknowns. Indeed, the amount of fill-in in the inverse triangular factors of a sparse matrix $A$ depends very strongly on the ordering. In contrast, the inverse $A^{-1}$ is usually full, regardless of the ordering chosen. In this note we focus on the factorized sparse approximate inverse preconditioner $A\text{INV}$ based on incomplete (bi)conjugation, developed in [2],[4]. As already noted in [4], this algorithm can benefit from the Minimum Degree ordering, which tends to reduce the amount of fill in the inverse factors without negatively impacting the rate of convergence. Here we present results of experiments with several orderings in addition to Minimum Degree. A partial explanation of the results in terms of graph theory is also given. We note that while the effect of

* Scientific Computing Group (CIC-19), MS B256, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA (benzi@lanl.gov). This work was supported in part by the Department of Energy through grant W-7405-ENG-36 with Los Alamos National Laboratory.

† Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic (tuma@iuvt.cas.cz). This work was supported by the Grant Agency of the Czech Academy of Sciences by grants No. 2030706 and 205/96/0921.
ordering on ILU-type preconditioners has been studied for many years [3], interest in the same problem for factorized approximate inverses has emerged only very recently. A recent study of the effect of orderings on the performance of AINV preconditioning can be found in [6]. Our main conclusions are similar to those reached in [6]; however, in [6] some new orderings specifically tailored to the AINV algorithm are also considered, including some weighted graph-based heuristics designed for dealing with anisotropic problems.

2. Structural considerations. The AINV algorithm [2],[4],[5] constructs a factorized sparse approximate inverse of the form

\[ M = ZD^{-1}W^T \approx A^{-1} \]

where \( Z, W \) are unit upper triangular matrices and \( D \) is diagonal. The approximate inverse factors \( Z \) and \( W \) are sparse approximations of the inverses of the \( L \) and \( U \) factors in the LDU decomposition of \( A \). The AINV algorithm computes \( Z, W \) and \( D \) directly from \( A \) by means of an incomplete biconjugation process, in which small elements are dropped to preserve sparsity.

Let \( A \) be an unsymmetric \( n \times n \) matrix that has a factorization \( A = LU \) without pivoting. Let \( G(A) = (V(A), E(A)) \) be the directed graph of the matrix \( A \), where \( V(A) = \{1, \ldots, n\} \) is the vertex set and \( E(A) \) is the set of edges \((i, j)\) with \( i, j \) such that \( a_{ij} \neq 0 \). Let \( x \in V(A) \). The closure \( \text{cl}_{G(A)}(x) \) of \( x \) in \( G(A) \) is the set of vertices of \( G(A) \) from which there are paths to \( x \). The structure of a vector \( v \) is defined as \( \text{Struct}(v) = \{i|v_i \neq 0\} \). In the following we will state results for the factor \( L \). Similar results hold, of course, also for the factor \( U \). These results were given in [6]. The usual no cancellation assumption is made throughout. We denote by \( L(*, i) \) the \( i \)-th column of \( L \).

**Lemma 2.1.** \( \text{Struct}(L(*, i)) = \text{cl}_{G(L)}(i) \).

Let \( G^o(L) \) denote the transitive reduction of the directed acyclic graph (dag) \( G(L) \). This is a graph with a minimal number of edges which satisfies the following condition: \( G^o(L) \) has a directed path from \( i \) to \( j \) if and only if \( G(L) \) has a directed path from \( i \) to \( j \). Then the following result holds [9].

**Lemma 2.2.** \( \text{Struct}(L(*, i)) = \text{cl}_{G^o(L)}(i) \).

We mention two simple consequences of this relation.

**Lemma 2.3.** If \( \text{cl}_{G^o(L)}(i) \cap \text{cl}_{G^o(L)}(j) = \emptyset \), then \( \text{Struct}(L^{-1}(*, i)) \cap \text{Struct}(L^{-1}(*, j)) = \emptyset \).

**Lemma 2.4.** Let \( K = \text{cl}_{G^o(L)}(i) \cap \text{cl}_{G^o(L)}(j) \). If \( K \neq \emptyset \), then \( L^{-1}(*, K) \) is a dense matrix.

The construction of factorized approximate inverse preconditioners is naturally influenced by the inverse fill (i.e., by the fill in \( L^{-1} \) and \( U^{-1} \)). Orderings that cause relatively low inverse fill are likely to result in faster computation of the preconditioner. On the other hand, it is difficult to predict the impact that reorderings obtained using unweighted graph information only will have on the rate of convergence of the preconditioned iteration. However, if the factorized approximate
inverse is computed by means of a drop tolerance rather than on the basis of position only, then the rate of convergence should not be affected too much. Indeed, the numerical experiments in the following section indicate that in some cases, orderings that cause low inverse fill result in dramatically improved convergence rates; see also the results in [6].

From the above considerations it follows that it is important to keep the fill-in in $L^{-1}$ small, i.e., to try to minimize the sum

$$
\sum_{i<j} |c_{G^*(L)}(i) \cap c_{G^*(L)}(j)|.
$$

The optimization problem (2.1) has close links to that of finding elimination tree height-minimizing orderings used for parallel elimination. As is well-known, the problem of minimum height elimination tree orderings for general graphs is NP-hard; see [15].

To illustrate the effect of ordering on the inverse fill-in, consider a matrix (symmetric, for simplicity) before and after the Reverse Cuthill-McKee (RCM) reordering, which naturally tends to keep the sum (2.1) rather big. Figure 1.1 shows the patterns of the matrix and the inverse of its factor $L$. Figure 1.2 shows the patterns of the matrix and the inverse of $L$ after RCM reordering. The inverse is much more dense in the second case.

**Figure 1.1:** Patterns of the matrix $A$ (left) and of $L^{-1}$, the inverse of its lower triangular factor (right). Stars denote matrix nonzeros, $f$ is used to denote filled positions in $L$.

**Figure 1.2:** Patterns of the matrix $A$ (left) and of $L^{-1}$, the inverse of its lower triangular factor (right) after the RCM reordering.
Hence, we can expect that reorderings aimed at reducing the envelope or the band will tend to make the inverses of the triangular factors rather dense. Therefore, the RCM-like orderings do not seem to be advantageous as reordering options before the factorized approximate inverse construction. Minimum Degree and Nested Dissection orderings are in principle more acceptable as matrix reorderings for construction of factorized matrix inverses. They provide more bushy elimination dags resulting, typically, in less fill-in in the inverse factors.

One natural choice of an ordering which keeps the elimination tree rather low is Nested Dissection. It is known that for every graph there exists a Nested Dissection ordering with minimal separators which produces an elimination tree of minimum height; see [14]. However, finding a Nested Dissection ordering with minimal separators that produces a minimum height elimination tree is also NP-hard. Next we state, without proof, a result for the 5-point formula discretization of Poisson's equation with Dirichlet boundary conditions on a 2D regular \( k \times k \) grid.

**Theorem 2.5.** Consider the matrix \( A \) from a \( k \times k \) regular grid problem with the Nested Dissection ordering. The number of nonzeros in the inverse factor \( L^{-1} \) is bounded above by \( O(k^3) \). The number of nonzeros in the inverse factor with the lexicographic and RCM orderings is \( O(k^4) \).

**3. Numerical experiments.** Here we report on some preliminary tests with symmetrically structured, unsymmetric problems arising from the discretization of convection-diffusion equations with finite differences. First we consider the following partial differential equation in \( \Omega = (0,1) \times (0,1) \)

\[
-\varepsilon \Delta u + \frac{\partial e^\varepsilon y u}{\partial x} + \frac{\partial e^{-\varepsilon y} u}{\partial y} = g
\]

with Dirichlet boundary conditions. The problem is discretized using centered differences for both the second order and first order derivatives with grid size \( h = 1/33 \), leading to a block tridiagonal linear system of order \( n = 1024 \) with \( nz = 4992 \) nonzero coefficients. The right-hand side is chosen so that the solution \( u \) to the discrete system is the vector \((1,2,\ldots,n)\). The parameter \( \varepsilon > 0 \) controls the difficulty of the problem—the smaller is \( \varepsilon \), the harder it is to solve the discrete problem by iterative methods (see also [3]). For our experiments, we generated ten linear systems of increasing difficulty, corresponding to \( \varepsilon^{-1} = 100, 200, \ldots, 1000 \). The coefficient matrix \( A \) becomes increasingly nonsymmetric and far from diagonally dominant as \( \varepsilon \) gets smaller. Moreover, Green function arguments show that the rate of decay in the inverse of the coefficient matrix becomes slower with decreasing \( \varepsilon \). In Table 1 we give the number of Bi-CGSTAB iterations required to reduce the initial residual by at least four orders of magnitude when the preconditioner \( AINV \) with drop tolerance \( Tol = 0.2 \) is used. In parenthesis, we give the number of nonzeros in the approximate inverse (in thousands). The different orderings considered are the lexicographic, or natural ordering (denoted no in the tables), Cuthill-McKee (cm), Reverse Cuthill-McKee (rcm), Multiple Minimum Degree (mmd), Nested Dissection (nd), and Red-Black (rb). A \( \dagger \) means that convergence was not attained within 500 iterations.
Orderings for Factorized Inverse Preconditioners

Table 1

Number of Bi-CGSTAB iterations and fill-in for AINV(0.2) preconditioning.

<table>
<thead>
<tr>
<th>$\varepsilon^{-1}$</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>600</th>
<th>700</th>
<th>800</th>
<th>900</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>14 (8.6)</td>
<td>14 (11)</td>
<td>24 (14)</td>
<td>26 (17)</td>
<td>58 (19)</td>
<td>73 (22)</td>
<td>700</td>
<td>800</td>
<td>900</td>
<td>1000</td>
</tr>
<tr>
<td>cm</td>
<td>12 (9.8)</td>
<td>19 (14)</td>
<td>43 (18)</td>
<td>26 (22)</td>
<td>27 (27)</td>
<td>33 (31)</td>
<td>62 (36)</td>
<td>63 (39)</td>
<td>15 (19)</td>
<td>21 (18)</td>
</tr>
<tr>
<td>rcm</td>
<td>12 (10)</td>
<td>16 (14)</td>
<td>29 (18)</td>
<td>19 (23)</td>
<td>25 (28)</td>
<td>29 (32)</td>
<td>63 (39)</td>
<td>63 (39)</td>
<td>15 (19)</td>
<td>21 (18)</td>
</tr>
<tr>
<td>mmd</td>
<td>8 (7.8)</td>
<td>8 (9.6)</td>
<td>9 (12)</td>
<td>10 (14)</td>
<td>13 (16)</td>
<td>13 (17)</td>
<td>15 (17)</td>
<td>18 (20)</td>
<td>23 (20)</td>
<td>19 (26)</td>
</tr>
<tr>
<td>nd</td>
<td>9 (7.7)</td>
<td>9 (9.7)</td>
<td>10 (12)</td>
<td>11 (14)</td>
<td>16 (16)</td>
<td>15 (17)</td>
<td>20 (21)</td>
<td>23 (22)</td>
<td>29 (29)</td>
<td>29 (32)</td>
</tr>
<tr>
<td>rb</td>
<td>8 (9.2)</td>
<td>9 (11)</td>
<td>11 (14)</td>
<td>12 (17)</td>
<td>15 (19)</td>
<td>19 (26)</td>
<td>25 (23)</td>
<td>24 (29)</td>
<td>25 (31)</td>
<td>25 (31)</td>
</tr>
</tbody>
</table>

The best results are obtained with the Minimum Degree heuristic. Nested Dissection is a close second. The natural and Cuthill-McKee-type orderings do poorly, particularly for small $\varepsilon$. The amount of inverse fill grows quickly and convergence is eventually lost. Thus, these orderings are not robust. The Red-Black ordering is much better, but not as good as Minimum Degree or Nested Dissection.

Additional experiments were performed with different convection-diffusion problems. The first one (denoted 3dcd) represents a 7-point centered differences discretization on a $20 \times 20 \times 20$ cubic grid of a three-dimensional analogue of equation (3.1) with $\varepsilon^{-1} = 100$ and with Dirichlet boundary conditions. Problems elman1 and elman2 were suggested by H. Elman and are described in [3]. They are two-dimensional, convection-dominated problems with constant coefficients discretized on locally refined meshes. These are stable, second-order discretizations leading to matrices of order $n = 5041$ for elman1 and $n = 7921$ for elman2, with 24921 and 39249 nonzeros, respectively. The last problem (zhang) is the most difficult of all for iterative solution. It was provided by J. Zhang and represents a recirculating flow problem. The dimension and number of nonzeros are $n = 1024$ and $nz = 4992$. The results of our experiments are reported in Table 2. For all problems, we used the AINV preconditioner with drop tolerance $Td1 = 0.1$. The stopping criterion was a reduction by at least six orders of magnitude for 3dcd, elman1 and elman2, and by at least four orders of magnitude for zhang. A * means that the maximum amount of fill allowed in the preconditioner, about 300,000 nonzeros, has been exceeded before the construction of the preconditioner was completed. It appears from these experiments that the Minimum Degree ordering is the most robust among those considered here, and also the most effective.

As with ILU, the ordering of grid points becomes increasingly important as convection becomes stronger. However, whereas for ILU preconditioning (with fill) the RCM-type orderings were found to be highly robust and effective [3], such reorderings are unsuitable for AINV. With AINV, the best...
results are obtained with Minimum Degree, which is inferior to RCM when used with ILU.

### Table 2

*Bi-CGSTAB iterations and fill-in for AINV(0.1) preconditioning, various problems.*

<table>
<thead>
<tr>
<th>Problem</th>
<th>t</th>
<th>Ordering</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no</td>
<td>cm</td>
</tr>
<tr>
<td>3dcd</td>
<td>15 (102)</td>
<td>14 (116)</td>
</tr>
<tr>
<td>elman1</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>elman2</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>zhang</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

## 4. Conclusions and future work.

In this contribution we have presented a few numerical experiments aimed at appraising the effect of different sparse matrix orderings on the factorized approximate inverse preconditioner AINV. These experiments appear to confirm the intuition that orderings which produce relatively sparse inverse factors, such as Minimum Degree and Nested Dissection, perform better than orderings (like RCM) that result in matrices with a narrow profile but full inverse factors. A good ordering for AINV results not only in greatly reduced effort in computing the approximate inverse factorization, but also in better convergence rates for the preconditioned iteration. This is probably due to the fact that with a good ordering, less entries in the inverse factors are discarded, resulting in a more accurate approximation to the exact inverse. In a way, the situation for factorized approximate inverse preconditioners is the opposite to that for ILU-type techniques.

The results presented here must be regarded as preliminary, and we do not expect the Minimum Degree ordering to be a good ordering for all problems. Weighted graph orderings, which take into account the numerical value of the nonzero entries in $A$ as well as their position, are likely to lead to more robust approaches. Some such heuristics have been presented in [6], and this is a topic that deserves further research.

For the near future, besides performing extensive experiments on a wide range of problems, we plan to investigate reorderings obtained from postprocessing the graph of an already existing sparse triangular factor. One strategy is based on tree rotations which approximately minimize the height of the elimination tree; see [12],[13]. Another strategy can be based on Hafsteinsson's chain reordering algorithm, again motivated by minimization of the height of the elimination tree. Other strategies will be considered as well. As the initial (pre)ordering, we will investigate combinations of “top-down” (ND-like) and “bottom-up” (MMD-like) ordering techniques; see, e.g., [1]. The idea behind this two-phase approach is to find orderings that attempt to reduce both the fill-in in the triangular factors of $A$, and the inverse fill. The same objective is also behind the Minimum Inverse Penalty (MIP) heuristic proposed in [6], which performed quite well on a number of problems.

Moreover, we expect that a combination of some of the symmetric permutations considered in
this paper with nonsymmetric permutations for permuting large entries to the main diagonal [7] will improve the performance of factorized approximate inverse preconditioners. Extensive numerical experiments will be performed to test this intuition.

In this paper we have restricted our attention to the AINV factorized approximate inverse preconditioner. We plan to include in future studies also other factorized forms, such as the FSAI algorithm [11] and inverse preconditioners based on bordering [5],[16].

REFERENCES