A Survey of Incomplete Factorization Preconditioners

Edmond Chow

Center for Applied Scientific Computing
Lawrence Livermore National Laboratory
Incomplete LU Factorizations

\[ A = LU - R \]

- Classical algorithms for ILU
  - ILU for General Matrices
  - ILU for Difference Operators
  - Dropping by position
  - Dropping by numerical size
- Existence problem and breakdown-free variants
- Stability problem and remedies
- Effect of ordering
- Some implementation considerations
ILU for General Matrices

Denote

\[ A_{k-1} = \begin{pmatrix} b_k & f_k^T \\ e_k & C_k \end{pmatrix} \]

starting with \( A_0 = A \), and consider step \( k \) of the outer-product form of Gaussian elimination

\[ A_{k-1} = \begin{pmatrix} I & 0 \\ e_k b_k^{-1} & I \end{pmatrix} \begin{pmatrix} b_k & f_k^T \\ 0 & A_k \end{pmatrix} \]

where \( A_k = C_k - e_k b_k^{-1} f_k^T \).

To make the factorization *incomplete*, entries are dropped in \( A_k \), i.e., the factorization proceeds with \( \tilde{A}_k = A_k + R_k \).
ILU for General Matrices

- The dropped entries form \(-R\) in \(A = LU - R\), that is, \(R_{ij} = 0\) if no dropping in position \((i, j)\)

- How to select which entries to drop?
  - By position or by numerical size

- Does the factorization exist? Remain positive?

- Actual computation is row-wise (or column-wise) for \(L\) and \(U\)

Modified ILU (MILU)

- \(LUe = Ae\) and \((LU)^{-1} Ae = e\)

- The entries dropped from \(A_k\) are added back to its diagonal

- A further diagonal perturbation of size \(O(h^2)\) is often used
ILU for Difference Operators

(A)  (L)  (U)  (LU)  (A–LU)
ILU for Difference Operators

- Make $LU$ and $A$ match on the nonzeros of $A$
- Make the rowsums of $LU$ and $A$ match
- Factorization can be written as $(D + L_A)D(D + U_A)$
ILU for Difference Operators

Increasingly larger stencils for $L$ (Gustafsson, 1978)
Convergence rate for 5-point Poisson problem

<table>
<thead>
<tr>
<th>Grid</th>
<th>num. equations</th>
<th>IC(0)-PCG</th>
<th>MIC(0)-PCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>32 × 32</td>
<td>1024</td>
<td>34</td>
<td>24</td>
</tr>
<tr>
<td>64 × 64</td>
<td>4096</td>
<td>66</td>
<td>35</td>
</tr>
<tr>
<td>128 × 128</td>
<td>16384</td>
<td>123</td>
<td>51</td>
</tr>
<tr>
<td>256 × 256</td>
<td>65536</td>
<td>246</td>
<td>74</td>
</tr>
</tbody>
</table>

\( \kappa = O(h^{-2}) \)

\( O(h^{-1}) \) steps

\( O(h^{-1/2}) \) steps
Convergence rate for 5-point Poisson problem
Earlier History

ILU for Difference Operators
- Stone (1968), Dupont, Kendall, and Rachford (1968)

ILU for General Matrices
- Meijerink and Van der Vorst (1977)
- Gustafsson (1978)
- Kershaw (1978)

Dropping Strategies for General Matrices
- Based on numerical size (Munksgaard, 1980, Zlatev, 1982)
- Based on position (Watts, 1981)
Dropping by position or “level”

\[ A_0 = \begin{pmatrix} b & f^T \\ e & C \end{pmatrix}, \quad A_1 = C - ef^T / b \]

Let \( A_0 \) have diagonal elements of size \( O(\varepsilon^0) \) and off-diagonal elements of size \( O(\varepsilon^1) \), with \( \varepsilon < 1 \), represented by

\[
A_0 = \begin{pmatrix}
1 & \varepsilon & \varepsilon & \varepsilon \\
\varepsilon & 1 & \varepsilon & \varepsilon \\
\varepsilon & \varepsilon & 1 & \varepsilon \\
\varepsilon & \varepsilon & \varepsilon & 1
\end{pmatrix}, \quad A_1 = \begin{pmatrix}
1 - \varepsilon^2 & \varepsilon - \varepsilon^2 & -\varepsilon^2 \\
\varepsilon - \varepsilon^2 & 1 - \varepsilon^2 & \varepsilon - \varepsilon^2 \\
-\varepsilon^2 & \varepsilon - \varepsilon^2 & 1 - \varepsilon^2
\end{pmatrix}
\]
Dropping by position or “level”

- Initial level-of-fill

\[
\text{level}^{(0)}_{ij} = \begin{cases} 
0 & \text{if } a_{ij} \neq 0 \\
\infty & \text{otherwise}
\end{cases}
\]

- When an element is updated, update its level-of-fill

\[
\text{level}^{(k)}_{ij} = \min(\text{level}^{(k-1)}_{ik} + \text{level}^{(k-1)}_{kj} + 1, \text{level}^{(k-1)}_{ij})
\]

- ILU\((k)\): Retain the nonzeros with level \(\leq k\)

- In practice, the best \(k\) are 0, 1, and 2 for 2-D and 0 and 1 for 3-D
Graph interpretation of “level-of-fill”

- Numbers indicate order of elimination
- Nonzero created at (4,6) from eliminating 1 and 2, since the path (4, 2, 1, 6) exists
- Level of fill-in is one less than the length of the shortest path between 4 and 6 through 1 and 2; in this case, level = 2
- Multilevel dropping strategies?
Dropping by numerical size (Threshold ILU)

Do not know beforehand which nonzeros to keep

Define a drop tolerance $\tau$; Two places to drop nonzeros:
  – small pivots, and small entries in $L$ and $U$

To control the maximum size of $L$ and $U$, restrict the maximum number of nonzeros per row: ILUT (Saad, 1994)
Existence

**Definition.** A is an *M-matrix* if *A* is nonsingular, $a_{ij} \leq 0$ for $i \neq j$, and $A^{-1} \geq 0$.

- The ILU factorization exists for an *M*-matrix, using any sparsity pattern including the diagonal (Meijerink and Van der Vorst, 1977)


- Note: the ILU factorization may break down or become indefinite for a positive matrix; the IC factorization may not exist for a SPD matrix
Shifted factorization

- Replace negative or zero pivots with small positive values (Kershaw, 1978)
- Shifted factorization: Factor $A + \alpha \text{diag}(A)$. An $\alpha$ exists such that this factorization exists (Manteuffel, 1980)
Ajiz-Jennings factorization

If \( d \) is to be dropped, \( s > 0 \), the submatrix is modified by adding

\[
\begin{pmatrix}
\vdots & \vdots & \vdots \\
0 & \frac{1}{s} |d| & 0 \\
0 & s|d| & -d \\
-2d & \frac{1}{s} |d| & \ddots \\
\end{pmatrix}
\]

which is positive semidefinite. The modified matrix remains positive definite and factorization cannot break down.

Ajiz and Jennings, 1984

Cf. diagonally compensated reduction (Axelsson and Kolotilina, 1994)
Tismenetsky’s factorization

\[ A = \begin{pmatrix} b & f^T \\ e & C \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ e/b & I \end{pmatrix} \begin{pmatrix} b & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} 1 & f^T/b \\ 0 & I \end{pmatrix} \]

where \( S = C - ef^T/b \). Now define \( p_e \) and \( p_f^T \) as \( e/b \) and \( f^T/b \) after dropping. Tismenetsky’s factorization uses

\[ \tilde{S} = ( -p_e \quad I ) A \begin{pmatrix} -p_f^T & I \end{pmatrix}^T \]

\[ = C + bp_e p_f^T - ep_f^T - p_e f^T \]


- \( \tilde{S} \) is SPD when \( A \) is SPD
- Need to keep track of \( (p_e - e/b) \) and \( (p_f - f^T/b) \)
- Very effective, but high intermediate storage costs
Factorization via $A$-orthogonalization

Use $A$-orthogonalization to produce $Z^T AZ = D$, with $Z$ upper-triangular. The root-free Cholesky factor is $L = AZD^{-1}$.

---

Benzi and Tůma, 2002

- Make incomplete by dropping in $Z$ (and $L$)
- Breakdowns can be avoided
- Needs intermediate storage, but not as much as Tismenetsky’s
Stability

- When an ILU factorization fails to help convergence, inaccuracy is often blamed.
- For nonsymmetric and indefinite matrices, instability of the LU factors is a common problem, i.e., $||L^{-1}||$ and $||U^{-1}||$ are very large.
- Note: $R = LU - A$ and $L^{-1}AU^{-1} = I + L^{-1}RU^{-1}$.
- This problem is rare in complete factorizations.
Unstable triangular factor

\[
\begin{pmatrix}
1 & & & \\
-2 & \ddots & & \\
& \ddots & \ddots & \\
& & -2 & 1
\end{pmatrix}
\begin{pmatrix}
x_1 \\
\vdots \\
x_i \\
\vdots
\end{pmatrix}
=
\begin{pmatrix}
b_1 \\
\vdots \\
b_i \\
\vdots
\end{pmatrix}
\]

Triangular solve recurrence:

\[x_i = 2x_{i-1} + b_i\]
Unstable triangular solves

Measure $\log_{10} \| (LU)^{-1} e \|_\infty$ (Chow and Saad, 1997)
Another difficulty: Very small pivots

- Lead to unstable factorizations, i.e., $\|L\|$ and $\|U\|$ are large
- Which lead to numerically zero pivots (via swamping)
- The small pivots might have been caused initially by inaccuracy due to dropping
Possible effect of small pivots

- Originally symmetric structure
- Large, erroneous, off-diagonal entries are propagated
### Assessing a factorization

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>condest</td>
<td>$| (LU)^{-1}e |_\infty$, $e = (1, \ldots, 1)^T$</td>
</tr>
<tr>
<td>1/pivot</td>
<td>size of reciprocal of the smallest pivot</td>
</tr>
<tr>
<td>max(L+U)</td>
<td>size of largest element in $L$ and $U$</td>
</tr>
</tbody>
</table>

#### condest

<table>
<thead>
<tr>
<th>small</th>
<th>large</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Inaccuracy due to dropping</em></td>
<td><em>unstable triangular solves</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>small</th>
<th>large</th>
</tr>
</thead>
<tbody>
<tr>
<td>very small pivots</td>
<td></td>
</tr>
</tbody>
</table>
Possible Remedies for Instability and Small Pivots

Stabilization

- Shifted factorization: $A + \alpha \text{diag}(A)$, best $\alpha$ is larger than the one that makes factorization exist (Manteuffel, 1980)
- Replace small pivots: sign of the pivot matters

Other Techniques

- Preserving symmetric structure
- Pivoting
- Reordering
- Blocking
Shifted factorization, nonsymmetric problem
Static, structure-based orderings

Natural  Reverse Cuthill-McKee  Minimum degree
Effect of ordering

Symmetric positive definite problems (Duff and Meurant, 1989)

- Natural and RCM orderings work well
- Minimum degree is better only with large amounts of fill-in

Nonsymmetric problems (Dutto, 1993, Benzi et al., 1997)

- RCM ordering is generally best
- Natural ordering generally worst
Coefficient-dependent orderings

Very unstructured problems
- ILUT with pivoting, called ILUTP (Saad, 1988)
- Maximum product transversals (Duff and Koster, 1999)
Anisotropy: complete $U$ factor, two orderings

Ordering along weak directions is better. This is counter-intuitive.
Dynamic, coefficient-dependent ordering

Recall

\[ A_{k-1} = \begin{pmatrix} b_k & f_k^T \\ e_k & C_k \end{pmatrix} \]

and

\[ A_k = C_k - e_k b_k^{-1} f_k^T, \quad \tilde{A}_k = A_k + R_k \]

Anisotropic problems

- Given a sparsity pattern for the factorization, dynamically choose an ordering for \( A_{k-1} \) that will reduce some norm of \( R_k \) (D’Azevedo, Forsyth, and Tang, 1991)
Implementation considerations for Threshold ILU

- Nonzeros in $L$ part must be eliminated in topological order
Crout version of ILU

Li, Saad, and Chow, 2002

- Avoids the topological sort
- Can produce a factorization with symmetric structure
- Dropping based on $L^{-1}$ and $U^{-1}$ can be implemented
Let $A_{k+1}$ be the $(k+1)$-st leading principal submatrix of $A$ and assume we have the decomposition $A_k = L_k D_k U_k$. Compute the factorization of $A_{k+1}$ via

$$
\begin{pmatrix}
A_k & v_k \\
w_k & \alpha_{k+1}
\end{pmatrix} =
\begin{pmatrix}
L_k & 0 \\
y_k & 1
\end{pmatrix}
\begin{pmatrix}
D_k & 0 \\
0 & d_{k+1}
\end{pmatrix}
\begin{pmatrix}
U_k & z_k \\
0 & 1
\end{pmatrix}
$$
Skyline version of ILU

\[
\begin{pmatrix}
A_k & v_k \\
w_k & \alpha_{k+1}
\end{pmatrix}
= 
\begin{pmatrix}
L_k & 0 \\
y_k & 1
\end{pmatrix}
\begin{pmatrix}
D_k & 0 \\
0 & d_{k+1}
\end{pmatrix}
\begin{pmatrix}
U_k & z_k \\
0 & 1
\end{pmatrix}
\]

Compute:

\[
\begin{align*}
z_k &= D_k^{-1} L_k^{-1} v_k \\
y_k &= w_k U_k^{-1} D_k^{-1} \\
d_{k+1} &= \alpha_{k+1} - y_k D_k z_k.
\end{align*}
\]

Chow and Saad, 1997

- Need sparse approximate solves
- May need a companion structure for \( L \) and \( U \)
- A running condition estimate \((L_k U_k)^{-1}\) is available
What we didn’t cover

- Block variants
  - BPKIT Software: Chow and Heroux (1998)

- Multilevel versions
  - Relation of block variants to multigrid methods
What we didn’t cover (cont’d)

- Parallel ILU for General Matrices
  - Multicoloring: Jones and Plassmann (1995)

- Perturbed MILU
  - Beuwens, Notay, Magolu, Eijkhout, and others
What we covered

- Classical algorithms for ILU
  - ILU for General Matrices
  - ILU for Difference Operators
  - Dropping by position
  - Dropping by numerical size
- Existence problem and breakdown-free variants
- Stability problem and remedies
- Effect of ordering
- Some implementation considerations
Acknowledgment

This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.