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 $\tilde{A}_k = A_k + R_k$.

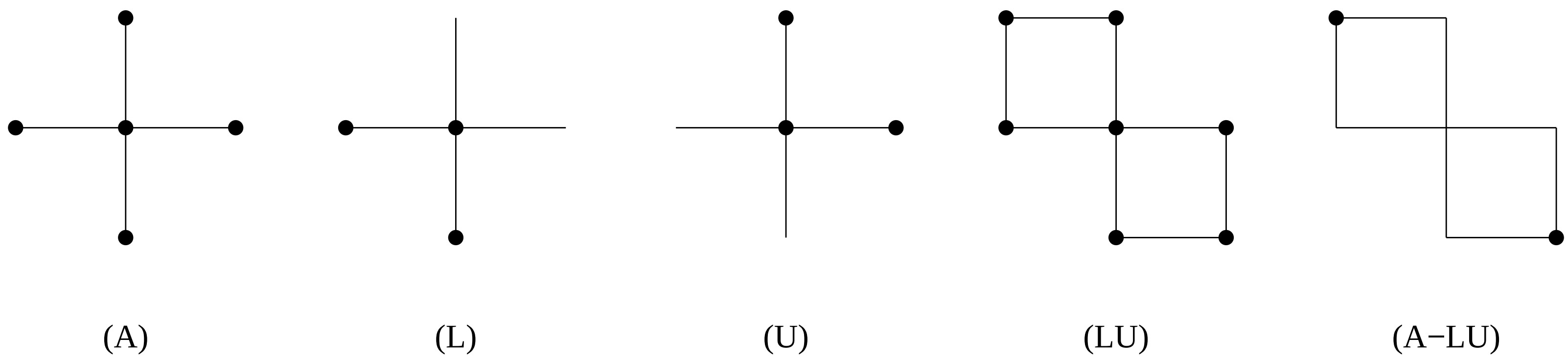
ILU for General Matrices

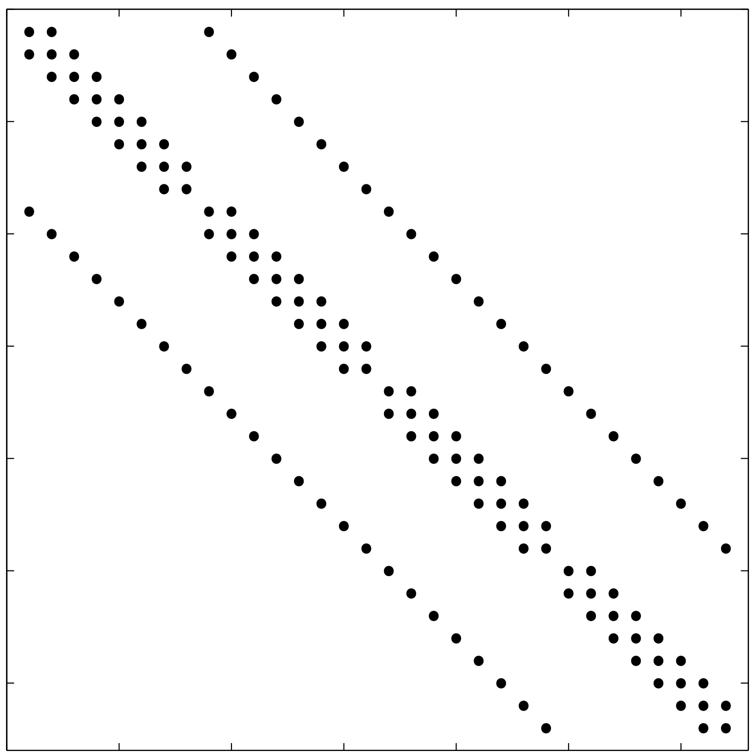
- The dropped entries form $-R$ in $A = LU - R$, that is, $R_{ij} = 0$ if no dropping in position (i, j)
- How to select which entries to drop?
  – By *position* or by *numerical size*
- Does the factorization exist? Remain positive?
- Actual computation is row-wise (or column-wise) for L and U

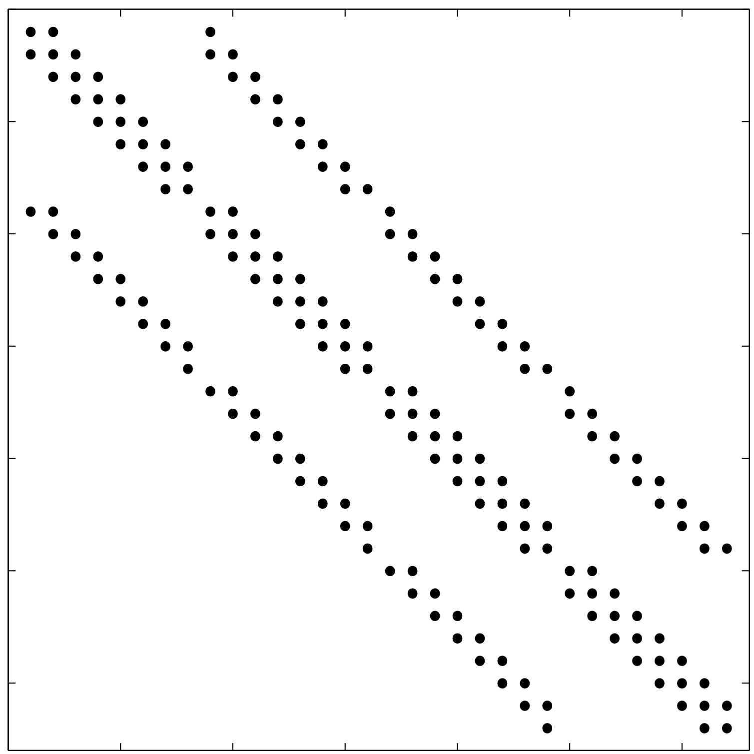
Modified ILU (MILU)

- $LUe = Ae$ 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

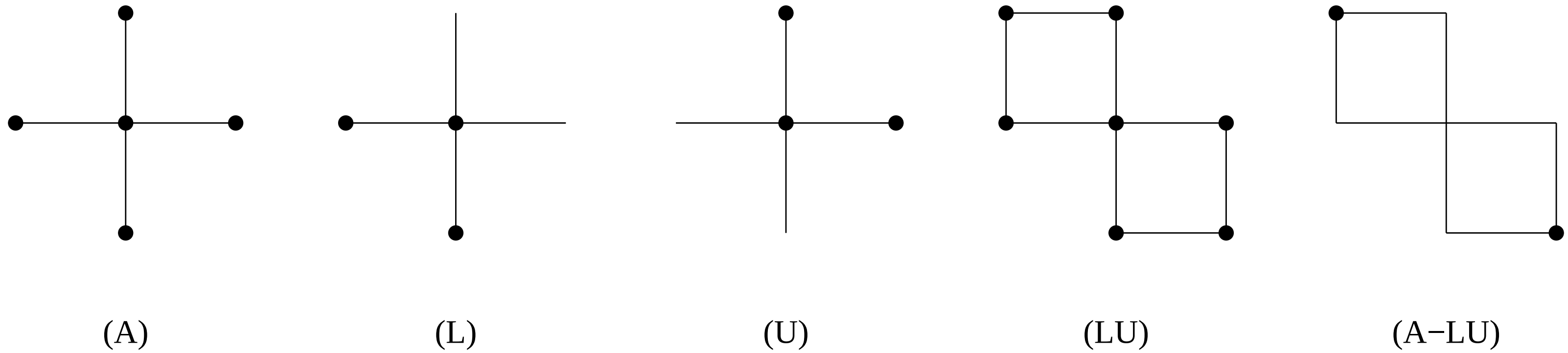
ILU for Difference Operators





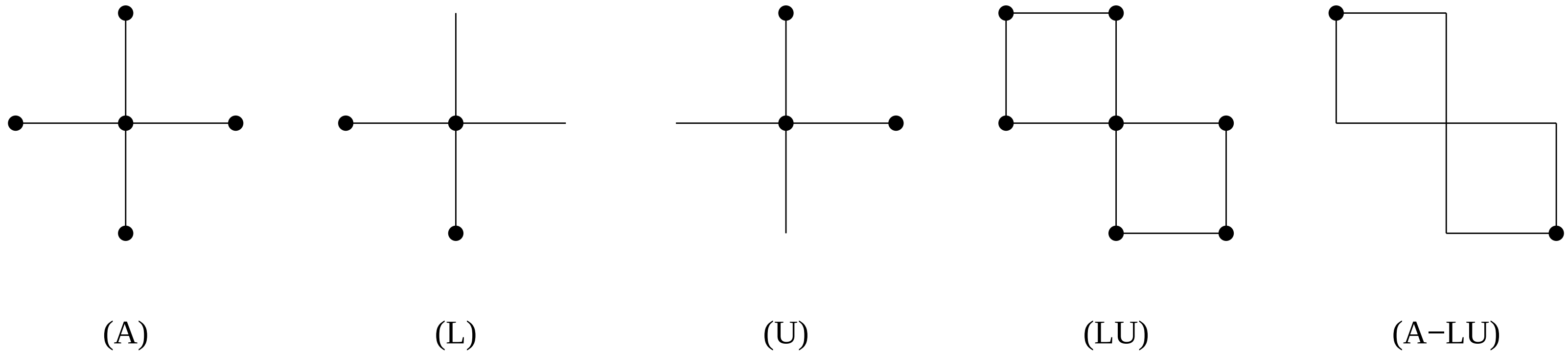


ILU for Difference Operators

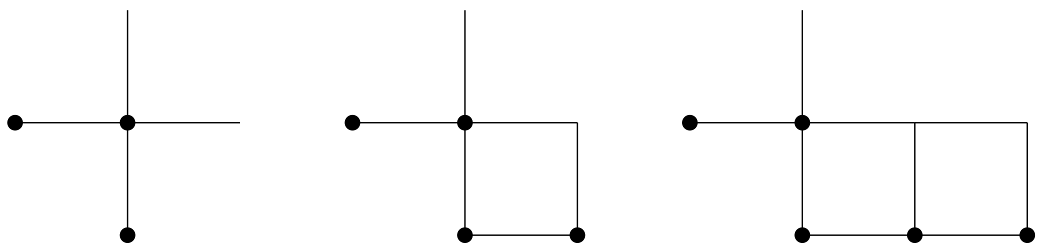


- Make LU and A match on the nonzeros of A
- Make the rowsums of LU and A match
- Factorization can be written as $(D + L_A)D(D + U_A)$

ILU for Difference Operators



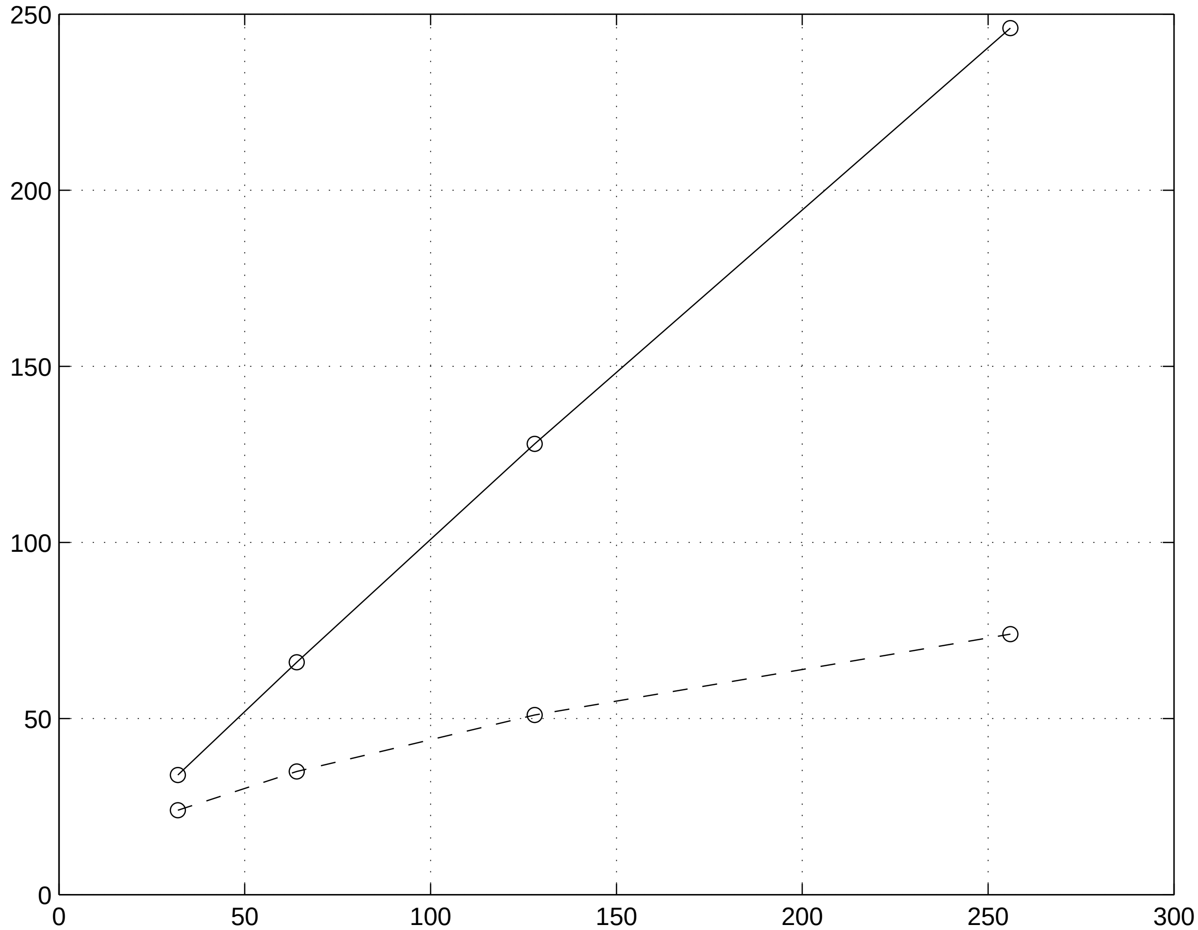
Increasingly larger stencils for L (Gustafsson, 1978)



Convergence rate for 5-point Poisson problem

Grid	num. equations	IC(0)-PCG	MIC(0)-PCG
32×32	1024	34	24
64×64	4096	66	35
128×128	16384	123	51
256×256	65536	246	74
$\kappa = O(h^{-2})$		$\kappa = O(h^{-2})$ $O(h^{-1})$ steps	$\kappa = O(h^{-1})$ $O(h^{-1/2})$ steps

Convergence rate for 5-point Poisson problem



Earlier History

ILU for Difference Operators

- Buleev (1960), Oliphant (1961), Varga (1961)
- 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"

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

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

$$A_0 = \left(\begin{array}{c|ccc} 1 & \varepsilon & \varepsilon & \varepsilon \\ \hline \varepsilon & 1 & \varepsilon & \\ \varepsilon & \varepsilon & 1 & \varepsilon \\ \varepsilon & & \varepsilon & 1 \end{array}\right), \qquad A_1 = \begin{pmatrix} (1-\varepsilon^2) & (\varepsilon-\varepsilon^2) & (-\varepsilon^2) \\ (\varepsilon-\varepsilon^2) & (1-\varepsilon^2) & (\varepsilon-\varepsilon^2) \\ (-\varepsilon^2) & (\varepsilon-\varepsilon^2) & (1-\varepsilon^2) \end{pmatrix}$$

Dropping by position or "level"

- Initial level-of-fill

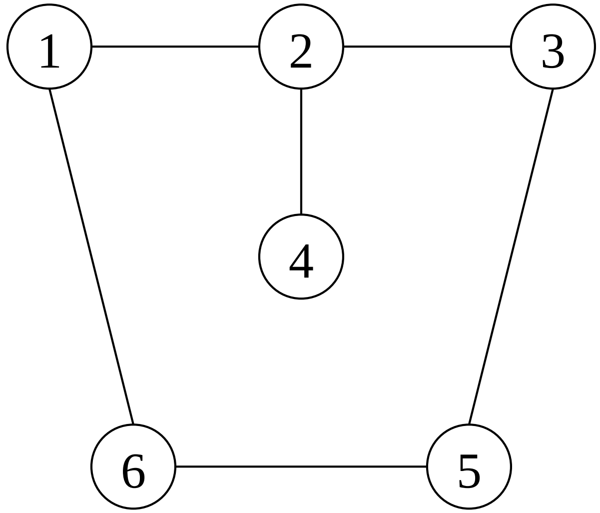
$$\text{level}_{ij}^{(0)} = \begin{cases} 0 & \text{if } a_{ij} \neq 0 \\ \infty & \text{otherwise} \end{cases}$$

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

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

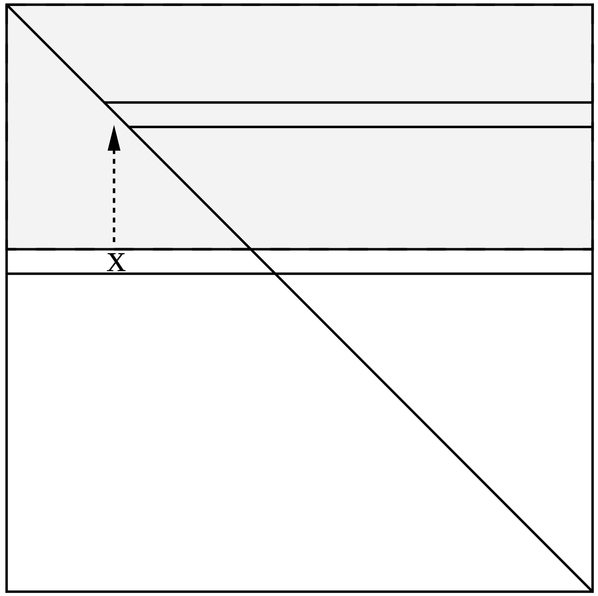
- ILU(k): Retain the nonzeros with level $\leq k$
- In practice, the best k are 0, 1, and 2 for 2-D and 0 and 1 for 3-D

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 τ; 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)
- Same result for H-matrices (Varga, Saff, and Mehrman, 1980, Manteuffel, 1980, Robert, 1982)
- 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 $A + \alpha \mathrm{diag}(A)$. An α 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

$$\begin{pmatrix} \ddots & & & & \\ & s|d| & & -d & \\ & & \ddots & & \\ & -d & & \frac{1}{s}|d| & \\ & & & & \ddots \end{pmatrix}$$

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

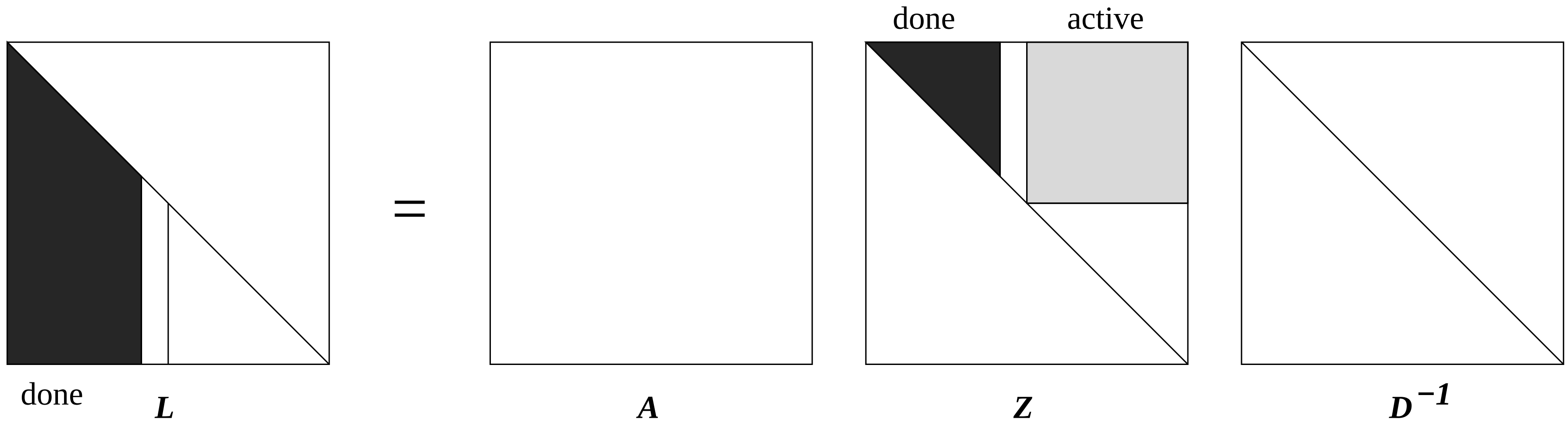
$$\begin{aligned} \widetilde{S} &= \begin{pmatrix} -p_e & I \end{pmatrix} A \begin{pmatrix} -p_f^T & I \end{pmatrix}^T \\ &= C + bp_e p_f^T - ep_f^T - p_e f^T \end{aligned}$$

Tismenetsky, 1991, Kaporin, 1998

- $\widetilde{S}$ 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 $L^{-1}AU^{-1} = I + L^{-1}RU^{-1}$
- Van der Vorst (1981), Elman (1986), Chow and Saad (1997)
- This problem is rare in *complete* factorizations

Unstable triangular factor

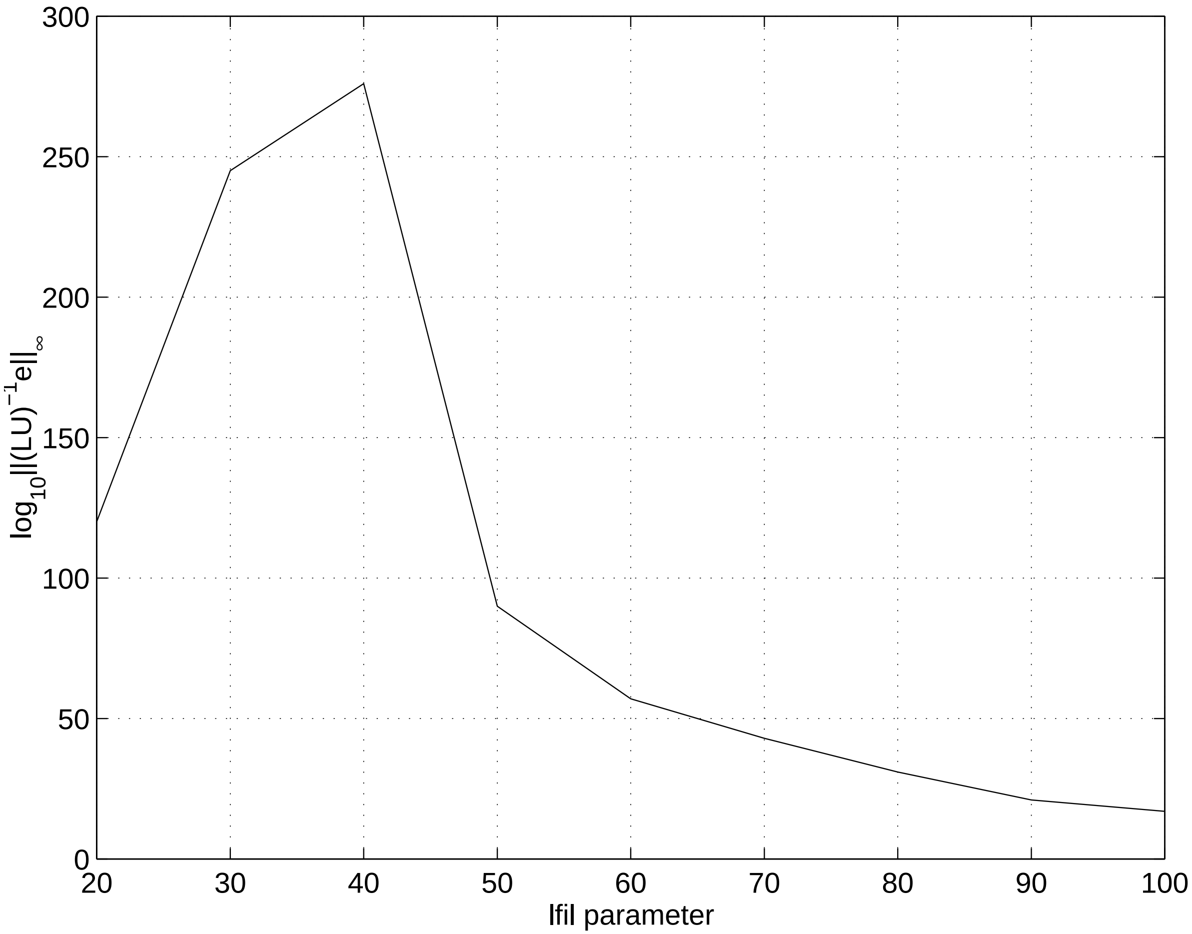
$$\begin{pmatrix} 1 & & & & \\ -2 & \ddots & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & -2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ \vdots \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_i \\ \vdots \\ \vdots \end{pmatrix}$$

Triangular solve recurrence:

$$x_i = 2x_{i-1} + b_i$$

Unstable triangular solves

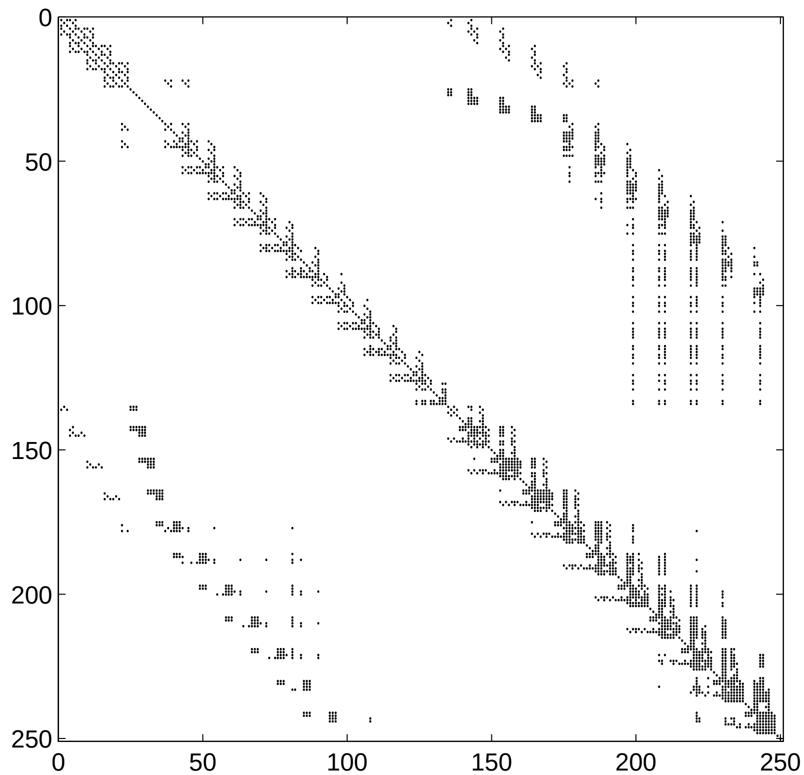
Measure $\log_{10} \|(LU)^{-1}e\|_\infty$ (Chow and Saad, 1997)



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

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

1/pivot \ **condest**	small	large
small	*Inaccuracy due to dropping*	*unstable triangular solves*
large		*very small pivots*

Possible Remedies for Instability and Small Pivots

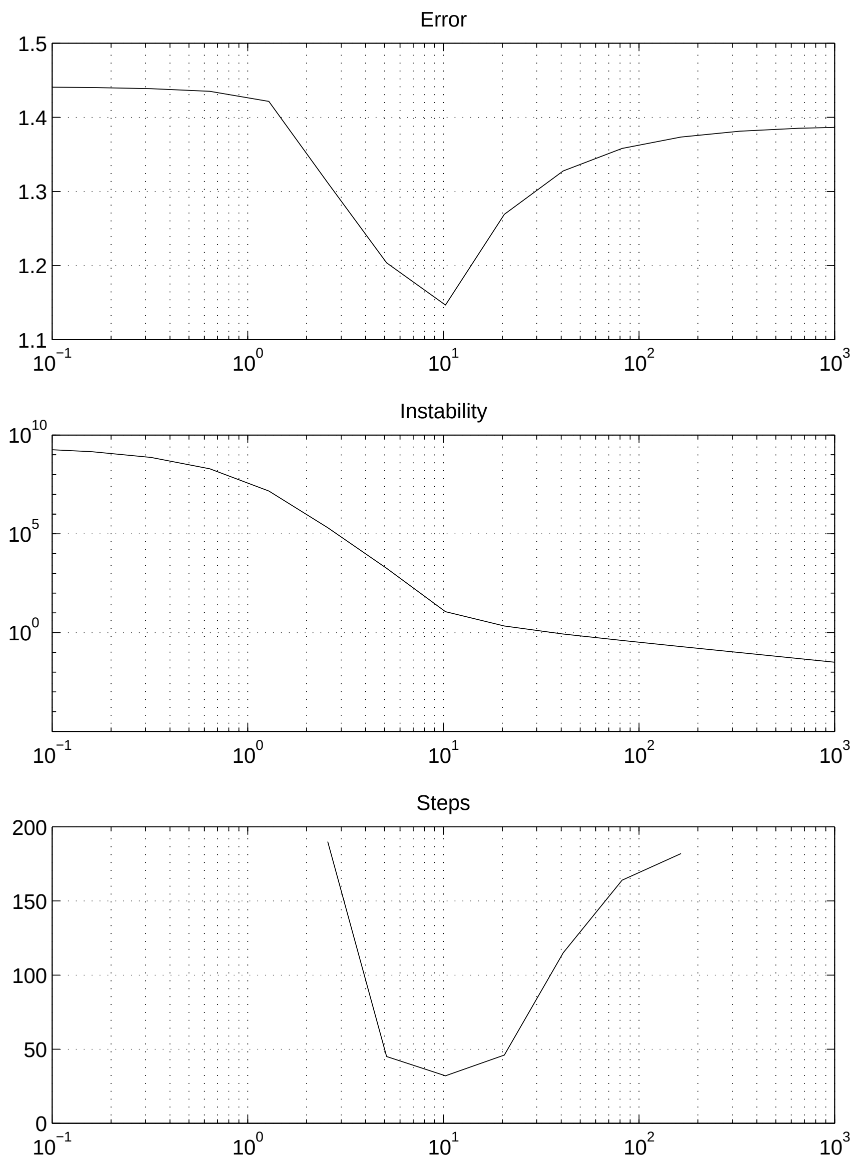
Stabilization

- Shifted factorization: $A + \alpha \text{diag}(A)$, best α is larger than the one that makes factorization exist (Manteuffel, 1980)
- Modify diagonals of L and U to make the factors diagonally dominant (Van der Vorst, 1981, Munksgaard, 1980, Elman, 1989)
- Replace small pivots: sign of the pivot matters

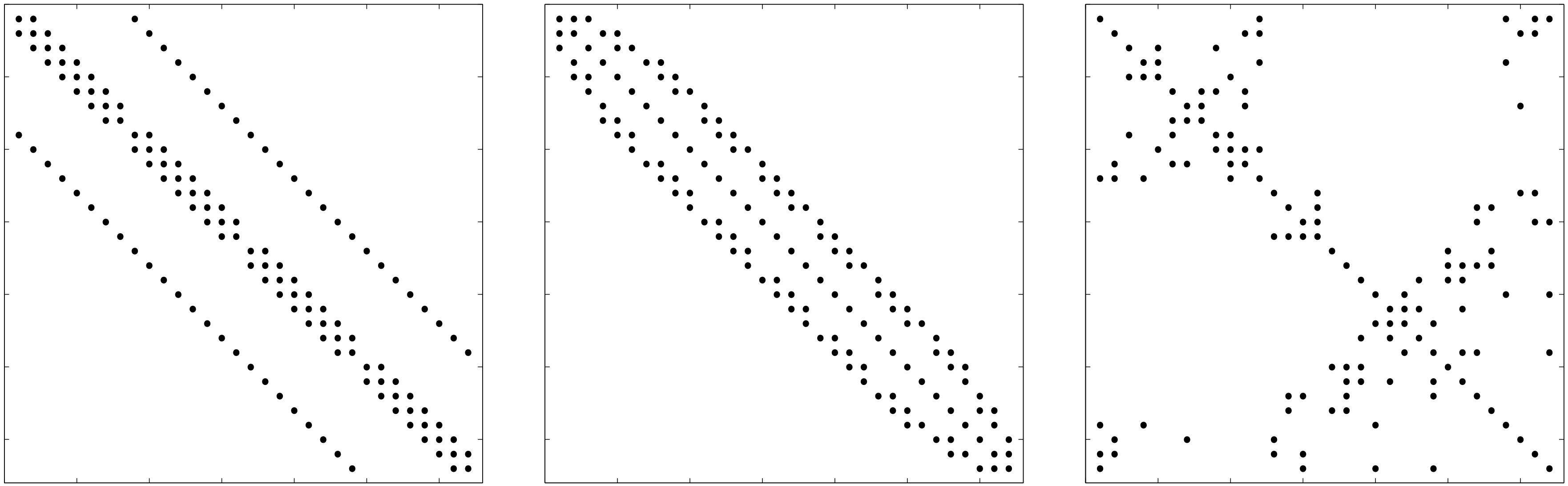
Other Techniques

- Preserving symmetric structure
- Pivoting
- Reordering
- Blocking

Shifted factorization, nonsymmetric problem



Static, structure-based orderings



Natural Reverse Cuthill-McKee Minimum degree

Effect of ordering

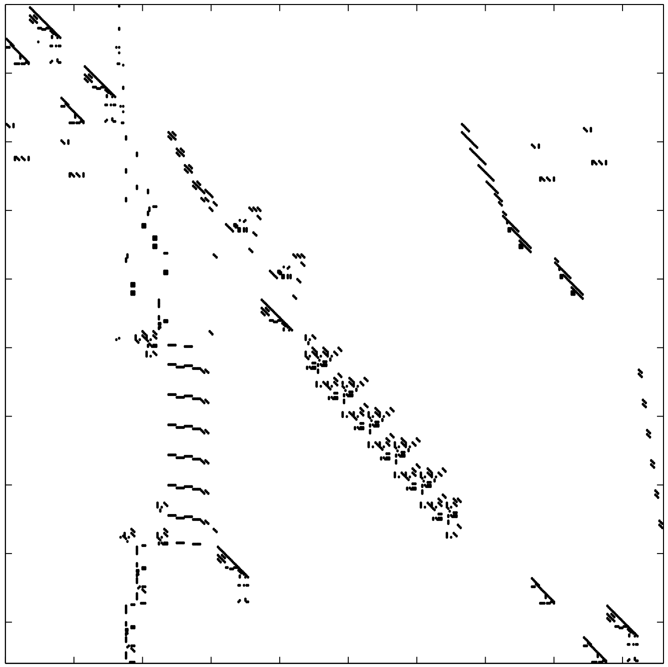
Symmetric positive definite problems (Duff and Meurant, 1989)

- 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

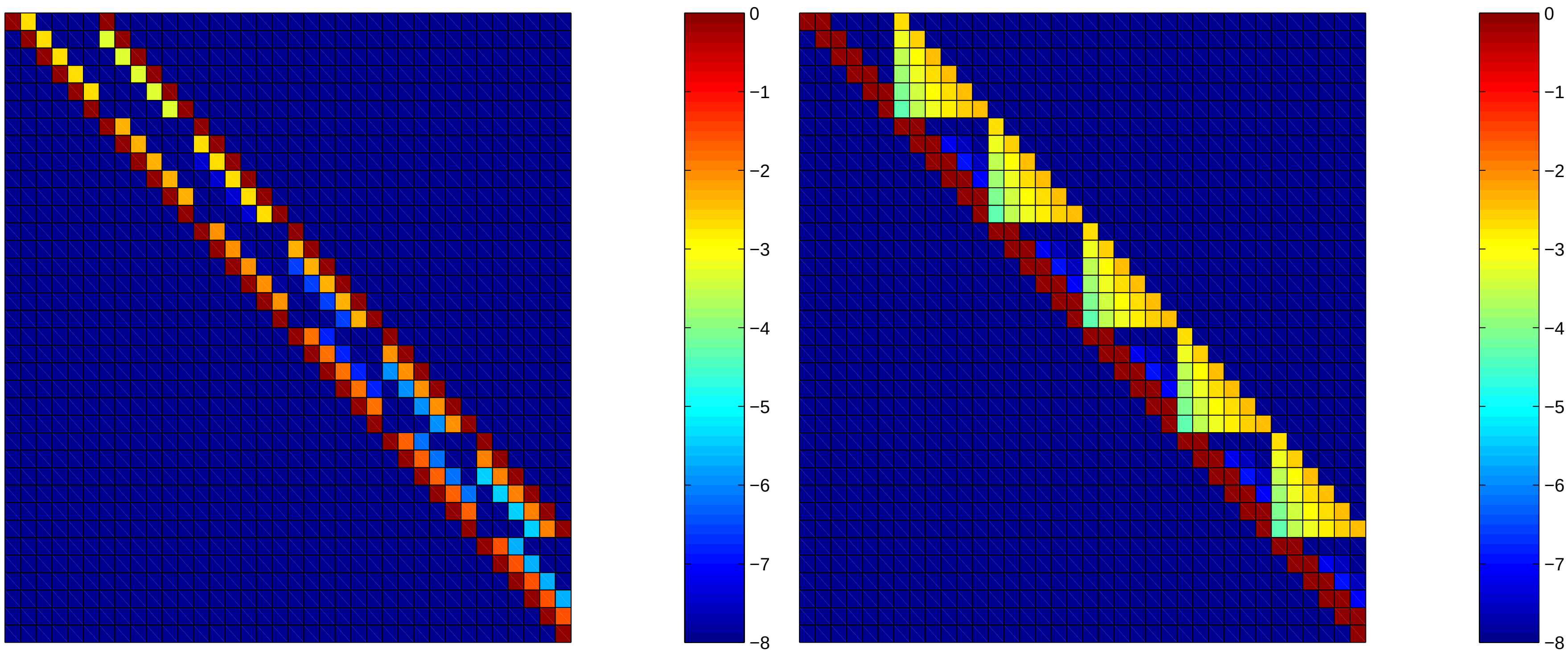
Coefficient-dependent orderings



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

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

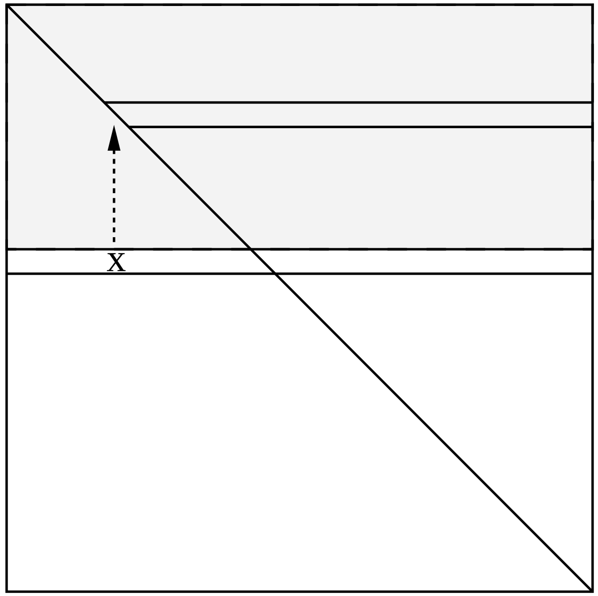
and

$$A_k = C_k - e_k b_k^{-1} f_k^T, \qquad \tilde{A}_k = A_k + R_k$$

Anisotropic problems

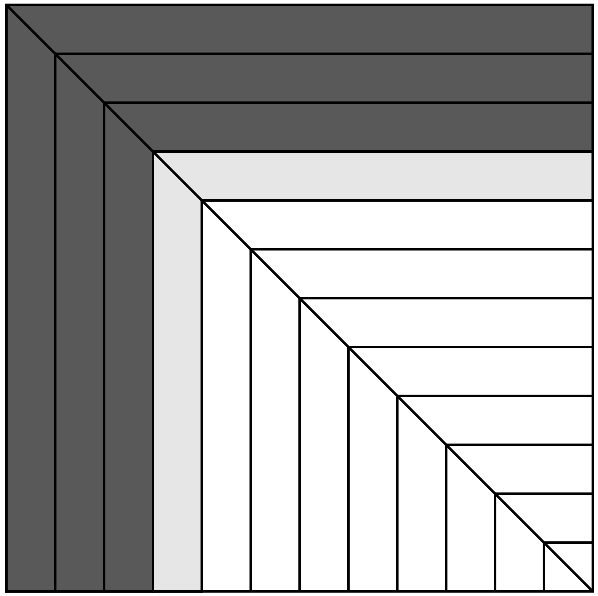
- Given a sparsity pattern for the factorization, dynamically choose an ordering for A_{k-1} that will reduce some norm of R_k (D'Azevedo, Forsyth, and Tang, 1991)

Implementation considerations for Threshold ILU



- Nonzeros in L part must be eliminated in topological order

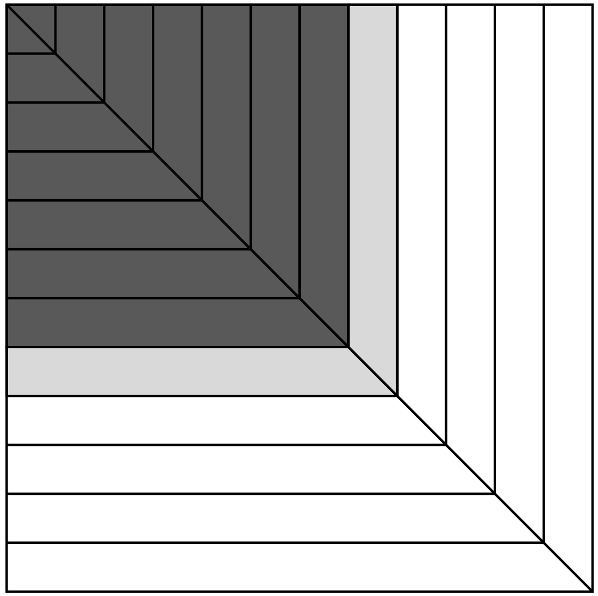
Crout version of ILU



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
- Cholesky and IC versions: Eisenstat, Schultz, and Sherman (1981), Jones and Plassmann (1995)

Skyline version of ILU



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:

$$\begin{aligned} z_k &= D_k^{-1} L_k^{-1} v_k \\ y_k &= w_k U_k^{-1} D_k^{-1} \\ d_{k+1} &= \alpha_{k+1} - y_k D_k z_k. \end{aligned}$$

Chow and Saad, 1997

- Need sparse approximate solves
- May need a *companion structure* for L and U
- A running condition estimate $\|(L_k U_k)^{-1}\|_\infty$ is available

What we didn't cover

- Block variants
  - Block tridiagonal: Axelsson, Brinkkemper, and Il'in (1984), Concus, Golub, and Meurant (1985), Kolotilina and Yeremin (1986)
  - Dense blocks: Fan, Forsyth, McMacken, and Tang (1996), Ng, Peyton, and Raghavan (1999)
  - BPKIT Software: Chow and Heroux (1998)
- Multilevel versions
  - Brand and Heinemann (1989), Saad (1996), Botta, van der Ploeg, and Wubs (1996), Saad and Zhang (1999), Saad, Sosonkina, and Suchomel (2000)
  - Relation of block variants to multigrid methods

What we didn't cover (cont'd)

- Parallel ILU for General Matrices
  - Multicoloring: Jones and Plassmann (1995)
  - Domain Decomposition: Saad and others (1994), Karypis and Kumar (1996), Hysom and Pothen (1998)
- Perturbed MILU
  - Beauwens, 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.