Slide 1

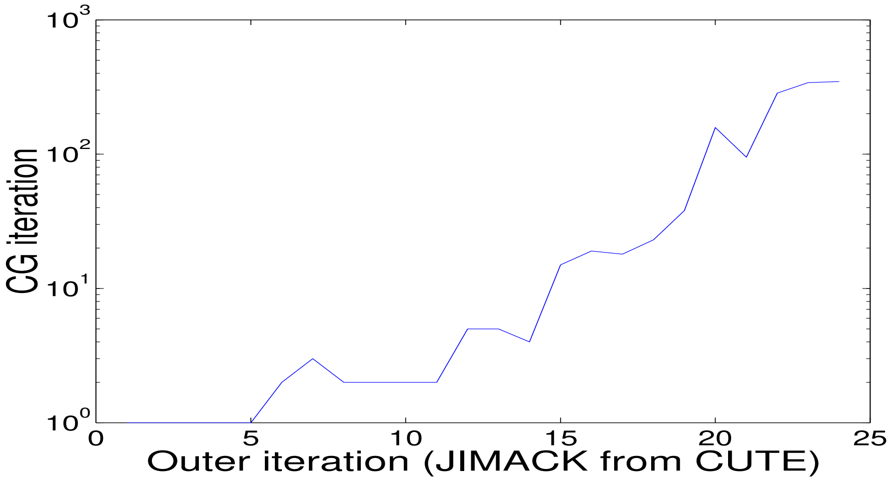
Fourth Kalamazoo Symposium on Matrix Analysis & Applications, Kalamazoo, Oct. 24-25, 1997

Incomplete Cholesky Factorizations with Limited Memory

Chih-Jen Lin and Jorge J. Moré

Slide 2

Motivation : from Large Scale Optimization



- In each iteration of optimization algorithm, a subproblem is solved by preconditioned conjugate gradient method.

Slide 3

Preconditioned Conjugate Gradient Method

- The linear system $Ax = b$, A is the Hessian or its approximation
- A may be indefinite, still find a decent direction and continue the optimization algorithm
- No general structure for matrix A
- Matrices (P.D. or indefinite) close to singular $\Rightarrow$ more difficult
- Need good preconditioners

Slide 4

Incomplete Cholesky Factorization (ICF)

Solve $Ax = b, LL^T$ as the preconditioner

$$A = LL^T + R$$

where

$$l_{i,j} = 0, \quad (i,j) \notin \mathcal{S}, \quad r_{i,j} = 0, \quad (i,j) \in \mathcal{S}.$$

Strategies

- Fixed fill : e.g. the same sparsity pattern of A
- Drop tolerance
- Variable fill

Questions

Is the sparsity pattern independent of numerical entries?
Are the storage requirements predictable?

Slide 5

ICF by selecting largest elements

Saad [1994], Jones and Plassmann [1995]

Let m_j be the number of nonzero off-diagonal elements in the j-th column of the lower triangular part of A.

For $j = 1, 2, \ldots,$

Store the jth column off-diagonal elements to w

Update w by using 1st to $j - 1$ st column of L

Select the largest $m_j + p$ nonzeros in w and store back to L

$n \times p$ more storage is required, known before running ICF

Slide 6

Shifting for Negative Pivots

- Incomplete factorization may fail. Facing negative pivots during factorization
- Two approaches : modify negative pivots to a positive number (local shift, $A + E$), or restart to factor $A + \alpha I$ (global shift)
- Local shift may generate a huge E. Consider the matrix

$$A = \begin{bmatrix} 1 & \gamma_1 & 0 \\ \gamma_1 & 1 & \gamma_2 \\ 0 & \gamma_2 & 1 \end{bmatrix}, \qquad \gamma_1, \gamma_2 \in (0,1), \gamma_1 \to 1.$$

Slide 7

- After two steps of factorization :

$$\begin{bmatrix} 1 & & \\ \gamma_1 & \delta_1 & \\ 0 & \delta_2 & 1-\delta_2^2 \end{bmatrix}, \qquad \delta_1 = \sqrt{1-\gamma_1^2}, \quad \delta_2 = \frac{\gamma_2}{\delta_1} \to \infty.$$

- Small αI is enough for global shift
- as γ_1 approaches 1, the matrix A approaches $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & \gamma_2 \\ 0 & \gamma_2 & 1 \end{bmatrix}$
- for $\gamma_2 \in [0, 0.1]$, the negative eigenvalue is near $-\frac{1}{2}\gamma_2^2$. Thus the Cholesky factorization of $A + \alpha I$, where $\alpha > \gamma_2^2$, succeeds.
- Shift incomplete factorization, Manteaffel [1980]

Slide 8

Shifting for Negative Pivots (Cont.)

- SLAP (Greenbaum and Seager), "if a zero pivot is found it is set to be 1.0 and the factorization proceeds", "This eventuality hardly ever occurs in practice."
- Running BCSSTK10 problem from Matrix Market(P.D., cond $\approx 2 \times 10^8$, size = 1000)

```
33   -1410728.0881009
36   -44435433003.158
37   -3.6208110524319D+19
38   -1.4219949825513D+36
41   -3.2931075508642D+60
42   -9.0673552689541+116
43   -1.6451457346854+233
46   -Infinity
```

- Shift ICF, $D^{-1/2}AD^{-1/2} + 0.008I$ succeeds, is usually more robust

Slide 9

Computing the ICF

- Reduce α of $A + \alpha I$ by scaling $D^{-1/2}AD^{-1/2}$.
- Let A be a symmetric matrix.

  Compute $\hat{A} = D^{-1/2}AD^{-1/2}$ where $D = \text{diag}(\|Ae_i\|_2)$.
  Let μ be a small positive number.
  Set $\alpha_0 = 0$ if $\min(a_{i,i}) > 0$; otherwise $\alpha_0 = -\min(a_{i,i}) + \mu$.
  For $k = 0, 1, \ldots$,
  Compute the ICF of $\hat{A}_k = \hat{A} + \alpha_k I$; exit if successful.
  Set $\alpha_{k+1} = \max(2\alpha_k, \mu)$

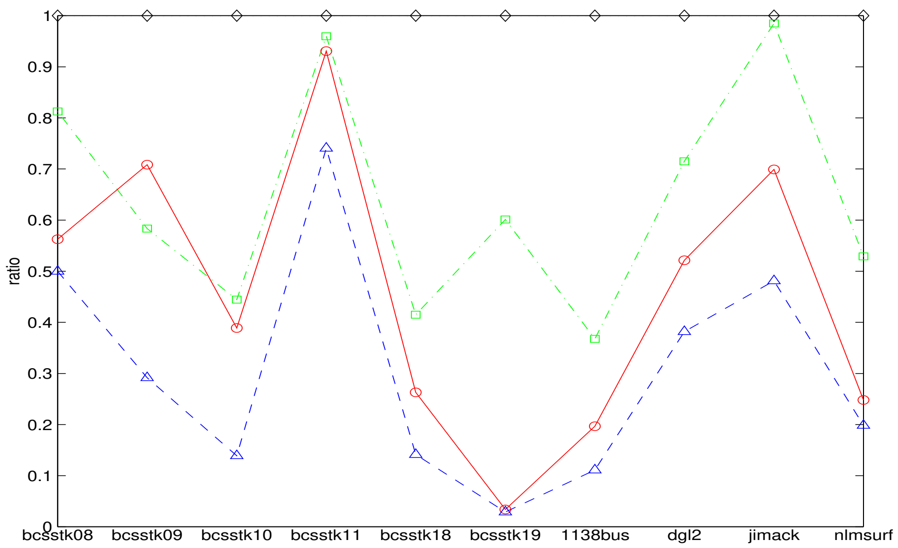
Slide 10

Scaling and Shifting

- **Choices of** $D = \text{diag}(d_i)$: improve the condition and reduce α of $\hat{A} + \alpha I$
  - $d_i = \max\{|a_{i,i}|, \epsilon\}$
  - $d_i = \max\{a_{i,i}, \epsilon\}$
  - $d_i = \|Ae_i\|_2$
- $\hat{A} + \sqrt{\beta} I$: H-matrix and ICF succeeds, β : max. # of nonzeros in any row of A
- l_r norm vs. l_s norm : If $r \leq s$ and $\widehat{A}_s + \alpha I$ is an H-matrix with positive diagonal elements, then $\widehat{A}_r + \alpha I$ is also an H-matrix with positive diagonal elements.

Slide 11

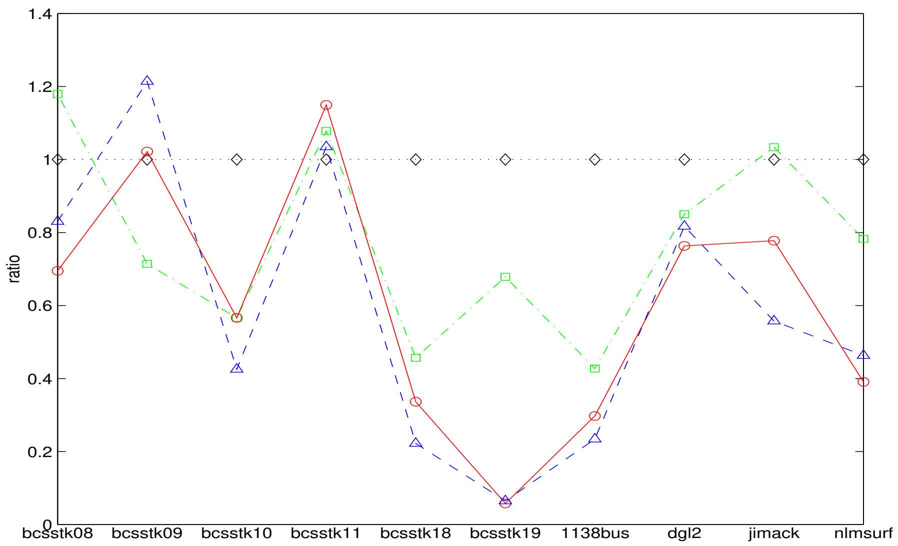
Number of Conjugate Gradient Iterations



($p = 0$ ◊, $p = 2$ □, $p = 5$ o, $p = 10$, △,Harwell-Boeing,Minpack-2,CUTE)

Slide 12

Computational Time and p



($p = 0$ ◊, $p = 2$ □, $p = 5$ o, $p = 10$, △)

Slide 13

CG Software Evaluation : CG iteration

	ICF	ma31 Munksgaard [1980]			MATLAB 5.0	
	$p = 5$	$(10^{-1}, 1)$	$(10^{-3}, 2)$	$(10^{-9}, 4)$	10^{-1}	10^{-3}
bcsstk08	9	36	5	1	52	14
bcsstk09	17	508	84	28	69	6
bcsstk10	14	1086*	11	1	1086*	8
bcsstk11	671	1473*	1473*	1	1473*	1262
bcsstk18	147	11948*	14	219	4893	54
bcsstk19	21	817*	31	2	817*	127
1138bus	23	258	4	1	148	31
dgl2	97	10000*	10000*	10000*	1760	215
jimack	93	1029*	126	2	1029*	128
nlmsurf	30	1024*	17	3	168	12

*: exceed maximal number of iterations

Slide 14

CG Software Evaluation : nnz(L)/nnz(A)

	ICF	ma31			MATLAB 5.0	
	$p = 5$	$(10^{-1}, 2)$	$(10^{-3}, 3)$	$(10^{-9}, 5)$	10^{-1}	10^{-3}
bcsstk08	1.73	1.31	2.31	4.31	0.36	2.88
bcsstk09	1.55	1.22	2.22	4.22	0.44	2.32
bcsstk10	1.46	1.19	2.19	4.19	0.46	1.50
bcsstk11	1.39	1.16	2.16	4.16	0.46	2.36
bcsstk19	2.02	1.43	2.43	4.43	0.43	1.71
bcsstk18	1.56	1.30	2.30	4.30	0.37	1.92
1138bus	2.19	1.88	2.88	4.88	0.95	2.95
dgl2	1.74	1.30	2.30	4.30	0.42	4.52
jimack	1.16	1.07	2.07	4.07	0.37	0.86
nlmsurf	2.00	1.42	2.42	4.42	0.69	3.06

Slide 15

Conclusions and Future Research

- Small amounts of additional storage lead to dramatic performance improvement in an ICF
- Drop tolerance strategies are unstable
- Experiments on other sets of matrices.
- The paper is available at
  `http://www.mcs.anl.gov/home/more/papers/icf.ps.gz`