Gary Howell

Applied Mathematics
Florida Institute of Technology
Melbourne, Florida 32901–6975

Internet: howell@zach.fit.edu

Efficient Reduction of General Matrices to Small Band Form

Gary Howell

Department of Applied Mathematics
Florida Institute of Technology
Melbourne, FL 32901
e-mail howell@zach.fit.edu

Bidiagonalization for SVD

The first step in determining singular values of a general matrix is in bidiagonalizing by alternately eliminating rows and columns by Householder transformations. If only the singular values are to be determined, this is the only $O(n^3)$ part of the algorithm.

- The current LAPACK implementation performs left and right matrix vector multiplies for each paired row-column elimination.

- The matrix vector multiplies are accomplished by two calls to the BLAS-2 algorithm GEMV.

- For each call to GEMV the entire active part of the matrix must be called from main memory through cache to the CPU.

A More Efficient SVD

- A more efficient approach is to combine the two matrix vector multiplies in one call to GEMVT.

- GEMVT will allow half as many reads from memory to cache.

- Potentially, overall execution time can be halved.

- GEMVT will be part of the new BLAS standard.

Computation of eigenvalues by reduction to small-band form by the BHESS algorithm

- Requires 8/3 n^3 double precision precision flops + flops required for iterations.

- Reduction is stable (or unstable) in a similar sense to EISPACK ELMHES (standard 1980's routine for reduction to similar Hessenberg form).

- The bulk of the flops are in the reduction to small band form and run at a good Mflop rate.

- Potentially the best throughput (problems/second).

- Reduction code available from howell@zach.fit.edu

Compare to Reduction to Hessenberg form followed by QR iteration

- Requires 10-20 n^3 double precision flops.

- The bulk of the flops are in QR iteration and run at a low Mflop rate.

- Best accuracy. Known examples of non-convergence are rare and can be easily rectified by allowing rare exceptional shifts and/or exceptional balancing.

- Quality code available from LAPACK and others.

Stability (Instability) of Direct Reduction BHESS

- Let A by an $n \times n$ matrix and

$$H = fl(X^{-1}AX)$$

be the Hessenberg matrix produced by the BHESS algorithm. Then

$$H = \widetilde{X}^{-1}(A_0 + E)\widetilde{X} \qquad (1)$$

where

$$\|E\|_2 \leq 12\, n\, cond_2(X)\, |A|\, u \quad \sum_{k=1}^{n-2} [\|\widetilde{m}_{k+1}\|_2^2] + O(u^2)$$

where u the machine precision, $A_0 = PAP^T$ is A with rows and columns permuted so that the reduction to banded Hessenberg form proceeds without permutations.

The Direct Reduction Algorithm BHESS

- If both a row and column can be eliminated, the pivot is chosen to minimize the maximal multiplier for the paired elimination.

- A row-column pair is eligible to be eliminated if $sec\theta/j < tol$, where j is row length and θ is the angle between row and column.

- tol approximately bounds maximal row multiplier if the maximal column multiplier is one. The maximal multiplier is more typically $tol^{1/2}$.

- For $k = 1 \ldots n-2$, the column is eliminated below the subdiagonal.

- For the kth step, eliminate the first eligible row with index less than k.

- If no row is eligible for elimination, use maximal column pivoting in the column elimination.

Comparison To Look-Ahead Lanczos

- For direct reduction each row-column elimination is allowed to have condition number at worst user-specified tol, e.g. 35 – In most look-ahead Lanczos methods pivots are accepted if condition is no worse than $\sqrt{u}$, e.g. 1.e8 .

- Multipliers used for direct reduction produce a globally well-conditioned similarity transformation – The left and right vectors for look-ahead Lanczos are locally biorthogonal, but not globally well-conditioned.

- Failures do not occur in direct reduction – Failures sometimes occur in look-ahead Lanczos.

- N steps of direct reduction typically preserve all eigenvalues – N steps of look-ahead Lanczos result in N correct eigenvalues only when the matrix size is small.

Computing Eigenvalues of Small-Band Hessenberg Matrices

- Efficient methods are Rayleigh quotient iteration and BR iteration.

- We can often compute eigenvalues of an unsymmetric matrix in about 1/5 Lapack time. For example, with $tol = 35$, and matrix size 1000, LAPACK took 300 seconds, BHESS-BR 60.

- On the result cited above BHESS was running at about 50 MFLOPs, compared to 90 MFLOPS for the LAPACK reduction to Hessenberg form.

The new BLAS standard will enable further speed-up

- Using GEMVT in BHESS can bring the BHESS Mflop rate up to that of LAPACK dgehrd. For the example of the graph BHESS-BR will compute eigenvalues eight times faster than LAPACK.

- Eight times faster gives scope for improving accuracy and reliability by iterative refinement and by using high precision arithmetic.

- High precision operations and GEMVT will be available in the new BLAS (Basic Linear Algebra Subroutines) standard.