then it can be shown that for the $\hat{\rho}_i$ of Equation 1 under the assumption that each H_i provides a new row of $\hat{A}$,

$$\sum_i \rho_i^2 \leq \sum_i \hat{\rho}_i^2.$$

Thus, in a sense, it is not possible to do better than the reflection coefficients ρ_i. Although this measure of optimality is not the most appropriate for the problem, the canonical set of reflection coefficients, ρ_i, can provide insights into the conditioning of transformations used in the generalized Schur algorithm.

As a short digression, it is worth noting that the decomposition given in Equation (2) applies when the matrix to be eliminated by H, B, is rank one and A is triangular, yielding a novel way of looking at Cholesky downdating. Suppose that for some upper triangular C and some vector x, $C^TC - xx^T$ is positive definite. Then the downdating problem is to find upper triangular $\hat{C}$ such that $\hat{C}^T\hat{C} = C^TC - xx^T$. The two standard algorithms for this are the Linpack algorithm and Chambers' algorithm. Chambers algorithm computes hyberbolic transformations and applies them in factored form to compute a downdate. The Linpack algorithm solves a triangular set of equations to obtain a sequence of plane rotations which would solve the updating problem $\hat{C}^T\hat{C} + xx^T$, thus solving the downdating problem indirectly.

If $y^TC = x^T$ and $y^TQ = \|y\|e_1^T$ then $e_1^TQ^TC = x^T/\|y\|$. It is possible to construct Q from plane rotations so that Q^TC will be upper Hessenberg. Thus, it is possible to transform C to a Hessenberg matrix for which the downdating is trivial. This is essentially a special case of Equation (2). Although this yields a downdating algorithm which is distinct from the two more conventional algorithms, it doesn't appear to have any advantages. When the plane rotations needed to restore the downdated H to triangular structure are taken into account it is clearly less efficient than either of the other two algorithms.

The main direction of this work is an attempt to understand the stability properties of the Schur and generalized Schur algorithms. We have looked at the remarkable robustness of the stability properties of the algorithm for Toeplitz matrices, its implications for semi-definite Toeplitz matrices and we have attempted to explain why stability is more delicate for other types of structured matrices.

ERIC DE STURLER
Swiss Center for Scientific Computing SCSC-ETHZ
Swiss Federal Institute of Technology Zurich
ETH Zentrum, RZ F-11
CH-8092 Zurich, Switzerland
+41-1-632 5566
sturler@scsc.ethz.ch

Inner-outer Methods with Deflation for Linear Systems with Multiple Right Hand Sides

ERIC DE STURLER

We will discuss inner-outer iterative methods with deflation for the solution of non-Hermitian complex linear systems with multiple right hand sides, and an application to problems that arise in lattice QCD [2]. The outer method is formed by GCR, which computes the optimal approximation over a set of (outer) search vectors. The inner method computes new search vectors by approximately solving the residual equation with a deflated operator. The deflation is computed using the approximate inverse from the outer GCR iteration [3, 4]. If we use a few steps (say five) of GMRES [7] as inner method the resulting scheme converges almost as fast as full GMRES, however at much lower cost. Other inner methods, like BiCGSTAB [8], also perform well, combining an optimal outer iteration with a cheap, short recurrence, inner method. Moreover, for each new right hand side we can reuse the approximation to the inverse computed for previous right hand sides and also improve the deflation at each iteration. This leads to a large reduction in the total number of iterations, while still solving for only one right hand side at a time.

We will briefly outline the method for a single right hand side [3]. The extension to multiple right hand sides is then straightforward. We solve the system $Ax = b$.

After k outer-iteration steps without truncation we have the following relations:

$$\begin{array}{ll} C_k^H C_k = I_k, & AU_k = C_k, \\ r_k = (I - C_k C_k^H) b, & x_k = U_k C_k^H b. \end{array}$$

If we now make m (non-fixed) inner GMRES steps with $(I - C_k C_k^H)A$ followed by an outer step we have

$$\begin{array}{rcl} AV_m &=& C_k C_k^H A V_m + V_{m+1} \bar{H}_m, \quad \text{where } C_k^H V_m = 0, \\ c_{k+1} &=& V_{m+1} \bar{H}_m y / \|V_{m+1} \bar{H}_m y\|, \\ u_{k+1} &=& (V_m y - U_k C_k^H A V_m y) / \|V_{m+1} \bar{H}_m y\|, \\ r_{k+1} &=& r_k - c_{k+1} c_{k+1}^H r_k \quad = \quad r_k - V_{m+1} \bar{H}_m y, \\ x_{k+1} &=& x_k + u_{k+1} c_{k+1}^H r_k, \end{array}$$

where y is chosen to minimize the (inner GMRES) residual. It is easy to show that this leads to the optimal approximation to $A^{-1}b$ over $range(U_k) \oplus range(V_m)$ in minimum residual sense [4]. Instead of GMRES we can also use other methods for the inner iteration, e.g. BiCGSTAB, where we also use the approximate inverse from the outer GCR for deflation to improve the convergence. After some number of inner iterations an outer step is computed as follows:

$$\begin{array}{rcl} c_{k+1} &=& r_k^{outer} - r^{inner} / \|r_k^{outer} - r^{inner}\|, \\ u_{k+1} &=& x^{inner} / \|r_k^{outer} - r^{inner}\|, \\ r_{k+1} &=& r_k - c_{k+1} c_{k+1}^H r_k, \\ x_{k+1} &=& x_k + u_{k+1} c_{k+1}^H r_k. \end{array}$$

Notice that the update of the outer residual and approximate solution with c_{k+1} and u_{k+1} are the same whether we use GMRES or some other method for the inner iteration. In fact one may also consider the outer iteration combined with the inner deflation as a way to improve the convergence of another iterative scheme. The variant with GMRES (typically for a small number of steps, say five or ten) as inner method generally leads to a convergence very close to that of full GMRES but at a much lower cost. Apart from updating the matrices U_k and C_k as indicated above, we can also add additional vectors to the set of outer vectors to improve the convergence further, e.g. eigenvectors corresponding to a suitable part of the spectrum.

Obviously, the approximate inverse from the outer GCR can be used to compute good initial approximations for successive right hand sides and to improve the convergence in the inner method for successive right hand sides. If the total number of iterations becomes very large, we 'truncate' the matrices U_k and C_k by making suitable combinations of the vectors, that is, we set $U_l = U_k W_l$, and $C_l = C_k W_l$ with $W_l^H W_l = I_l$, and $l < k$. We will discuss several strategies for choosing W_l. The overall scheme leads to an efficient method that can combine the information from the iterations for successive right hand sides, while still solving for one right hand side at a time (in contrast to block methods). Therefore, the same amount of memory is needed independent of the number of right hand sides. It also allows us to improve the convergence for right hand sides that are not known before a previous system has been solved.

rhs	GCRO(5)	BiCGγ_5
1	738	803
2	1097	1604
3	1379	2393
4	1681	3277
5	2002	4074
6	2353	4862
7	2702	5653
8	3066	6437
9	3427	7242
10	3789	8039
11	4154	8829
12	4514	9611

Table 1: Cumulative iteration counts (matrix-vector products) for GCRO(5) and BiCGγ_5 for twelve right hand sides

We apply these methods to problems arising in Lattice Quantum Chromo-Dynamics [2], where the matrix is of the form $A \equiv (I - \kappa M)$, and we solve for a small multiple of twelve right hand sides. In the 'simple' case, where κ is real, several methods with cheap iterations still perform well. One important example of such a method is BiCG where one exploits the fact that a Hermitian (but not positive) matrix P exists such that AP is Hermitian and $PP = I$, and hence $A^H = PAP$. So BiCG can be implemented very cheaply with an appropriate choice of the starting residuals [5, 6, 1]. If P would be positive definite, this BiCG would actually amount to CG with a special inner product. Although this is not the case, the convergence of the method is nearly optimal. In Table 1 we compare the convergence of this special BiCG (BiCGγ_5) with that of the method outlined above with 5 steps of GMRES as inner method (GCRO(5)) and a maximum of 100 outer vectors. We give the cumulative iteration counts to reach a relative reduction of the residual norm of $1.0e-10$ for twelve right hand sides. In the 'harder' case κ is complex and most of the methods with cheap iterations do not converge anymore (most notably BiCGγ_5). However, the nested methods still converge well. We will discuss different strategies for selecting vectors from the inner iteration to be kept in the outer iteration to improve the effect of the deflation, and better truncation schemes. Also the possibility to combine different methods at different stages will be discussed. For example, one could use GMRES as inner method for the first two right hand sides, to compute accurate eigenvector approximations to be added to the set of outer vectors, and then switch to use BiCGSTAB or BiCGSTAB(k) for the other right hand sides.

Bibliography

[1] A. Borici and P. De Forcrand Fast Krylov space methods for calculation of Quark Propagator. In Proceedings of Physics Computing '94, August 1994, Lugano, Switzerland.

[2] P. De Forcrand. Progress on lattice QCD algorithms. IPS Research Report No. 95-24, Interdisciplinary Project Center for Supercomputing, Swiss Federal Institute of Technology, Zurich, Switzerland, 1995. To appear in Nuclear Physics B (LAT95 Proceedings).

[3] E. De Sturler and D.R. Fokkema. Nested Krylov methods and preserving the orthogonality. In N. Duane Melson, T. A. Manteuffel, and S. F. McCormick, editors, *Sixth Copper Mountain Conference on Multigrid Methods*, NASA Conference Publication 3224, Part 1, pages 111–125, Hampton, VA, USA, 1993. NASA Langley Research Center.

[4] E. De Sturler. Nested Krylov methods based on GCR. Technical Report 93-50, Faculty of Technical Mathematics and Informatics, Delft University of Technology, Delft, The Netherlands, 1993. Accepted for publication in Journal of Comp. and Appl. Mathematics, North-Holland.

[5] R. W. Freund. Lanczos-type algorithms for structured non-Hermitian eigenvalue problems. In J. David Brown, Moody T. Chu, Donald C. Ellison, and Robert J. Plemmons, editors, *Proceedings of the Cornelius Lanczos International Centenary Conference*, SIAM Proccedings Series, pages 243–245, Philadelphia, PA 19103-5052, USA, 1994. SIAM Publications.

[6] A. Frommer et al. Preprint, available from hep-lat/9504020

[7] Y. Saad and M. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 7:856–869, 1986.

[8] H. A. Van der Vorst. BI-CGSTAB: A fast and smoothly converging variant of BI-CG for the solution of nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 13:631–644, 1992.