A flexible, truncation scheme for FOM and FGMRES

Andreas Stathopoulos

Department of Computer Science
College of William and Mary
Williamsburg, VA 23187-8795

February 1, 2002

Krylov iterative methods for linear systems that maintain explicitly an orthonormal basis of the Krylov vectors have memory requirements and execution time that increase with the number of iterations. Therefore, methods like GMRES, FGMRES, or FOM have traditionally been used in various restarted or truncated variants.

Restarted methods are stopped after an allowed number of iterations, m, at which time the best approximation from the subspace is extracted, x_m. This projection step is either a Galerkin (in FOM) or a Petrov-Galerkin method (in FGMRES). The approximation x_m is then used to restart the method anew. Restarting may cause two well documented problems: stagnation, or/and significant convergence deterioration. The reason is that the approximation x_m does not include certain directions from the Krylov space that method keeps rebuilding at every outer loop. Morgan has shown that approximate eigenvectors often capture this lost information well; at least when eigenvalues close to zero are the source of the problem and when these can be computed.

What is disconcerting with restarting techniques is that they present the same problems even for symmetric matrices, where the whole Krylov space can be build with a three-term recurrence. Although FOM(m) or FGMRES(m) are rarely used for symmetric matrices, they are often used for nearly symmetric ones. Yet, they still present the same limitations.

For nearly symmetric matrices, incomplete orthogonalization methods have been described, such as DIOM(s) and DQGMRES(s), that maintain only a local orthogonality among the Krylov vectors V, i.e., each $v_i^T v_j = 0$ only for $|i-j| \leq s$. If the matrix is symmetric and $s \geq 2$, this local orthogonality is equivalent to a global one and these methods are optimal. For non symmetric matrices, the Faber-Manteuffel theorem states that such methods cannot be optimal, at least for reasonably small s. In practice, truncation methods are relatively effective when $\Delta = \|A - A^T\|/\|A\|$ is small, but not with larger Δ. Interestingly, DIOM(s) is still a Galerkin projection, albeit an oblique one against a space with deteriorating condition number. Similarly, the loss of orthogonality in V results in the well known relationship between the residuals for DQGMRES(s) (r^Q) and GMRES (r^G):

$$\|r_m^Q\| < \kappa(V_{m+1}) \|r_m^G\|.$$

To improve the conditioning of V, truncated methods are usually restarted, as well. However, restarting destroys the local information that the method keeps over the last s iterations, again hurting the efficiency in nearly symmetric methods.

Based on the fact that optimal projections (e.g., residual minimization with Petrov Galerkin) are eventually orthogonality constraints, de Sturler noted that a truncation technique should maintain those directions that new vectors tend to lose orthogonality against. Although future vectors are not known, loss of orthogonality is extrapolated from the current vectors in the basis. Coupled with GCRO this technique has worked surprisingly well. Additional research is needed to determine the choice of parameters, as well as to identify its relation to Morgan's augmented Krylov methods.

Ideally, the space V should be truncated to x_m and an orthogonal subspace Q of V, so that an application of the projection (Galerkin) step over $\{Q, x_m, r_m\}$ would produce the same x_{m+1} approximation of the unrestarted method. Interestingly, even for symmetric matrices, the choice of the last s vectors in V is not optimal, but the appropriate space involves $\{x_{m-1}, x_m, r_m\}$. This suggests an extension to truncation methods which is similar in spirit to de Sturler's method.

We propose an FOM-like method, that allows the use of arbitrary vectors in the space, much like FGMRES, and thus facilitates the following truncation/restarting strategy. After m vectors have been generated in V, we compute the "best" set of directions from

$$\{X, x_m\} = \{x_1, x_2, \ldots, x_{m-1}, x_m\},$$

which is a different basis for the same space. Instead of keeping a truncation window of the last s vectors $x_{m-s}, \ldots, x_{m-1}$, we examine two more general ways of obtaining the "best" directions. These methods apply both to our new FOM method and to FGMRES.

In the first method, we pick the vectors that might describe x_{m+1} well, by extrapolating from how well subsets of X vectors can describe $x_m(1:m-1)$. Specifically, we compute the coefficients c, such that $x_m(1:m-1) = Xc$. Then we pick Q as the s vectors x_i with the largest normalized coefficients: $|c_i| \|x_i\|$. If we denote as $L = X - Q$ the subspace of all vectors that are left out, this choice guarantees that the distance of the $x_m(1:m-1)$ vector from L is bounded by:

$$dist(x_m(1:m-1), X - Q) < \sum_{x_i \in L} |c_i| \|x_i\|,$$

which is a minimum bound over all choices of x_i.

In the second method, we obtain the singular value decomposition of X and pick Q as the singular vectors with the largest singular values. This guarantees that the most important directions in recent approximation vectors are kept.

The described techniques have the advantage of being optimal for matrices where there is a s-term recurrence. Also, the x_i previous approximation vectors provide a basis more suitable for truncation, and the techniques are not limited to the last s recurrence vectors. We have to emphasize that all the above operations can be performed inexpensively in the $m \times m$ coordinate space, with negligible cost to the main iteration. Finally, the flexibility of the proposed method (as well as that of FGMRES) allows also the incorporation of additional information into the truncation window. Specifically, we can augment the truncated space by eigenvectors as computed by Morgan's method. The resulting method is more robust, because it does not rely solely in the computation of eigenvectors or the identification of the appropriate window.

Our preliminary experiments have shown that our truncation method improves convergence significantly when compared with FOM(m) and FGMRES(m). We have also found that combining our trunction with Morgan's method always improves convergence, sometimes matching that of the unrestarted methods. The combination method seems to perform similarly and sometimes even improve de Sturler's method, but only in one experiment we could reproduce from his paper. More experimentation is needed toward that direction.