DYNAMIC ACCELERATION OF NONLINEAR ITERATIONS

James M. Hyman

Center for Nonlinear Studies
Los Alamos National Laboratory
Los Alamos, NM 87545

Thomas A. Manteuffel

Computing Division
Los Alamos National Laboratory
Los Alamos, NM 87545

I. INTRODUCTION

Many large and complicated production codes require iteration to the solution of a system of nonlinear equations. For example, the solution to a time-dependent partial differential equation may require the solution of a nonlinear system at each time step. Unfortunately, these codes often implement suboptimal methods and convergence is very slow.

EMB

Due to unstructured programming and often political considerations, it is difficult to implement major changes in an existing applications computer code. In this paper we describe algorithms for accelerating the convergence iteration methods with a minimal impact on the original code. In most cases a single subroutine call and additional work space are sufficient.

Suppose we monitor an iteration by examining a sequence of iterates,

$$\cdots \underline{v}_{i-1}, \underline{v}_i, \underline{v}_{i+1} \cdots$$

If each iteration is a function of the previous iterate, then

$$\underline{v}_{i+1} = F(\underline{v}_i) \quad (1)$$

for some function F . The iteration has converged when a $\underline{v}$ is found such that

$$\underline{v} = F(\underline{v}) \quad (2)$$

The basis for this work is the assumption that the suboptimally implemented algorithm yields a sequence of iterates that can be well approximated by an iteration based on a linear system of equations when the convergence is very slow; that is,

$$\underline{v}_{i+1} \approx G_i \underline{v}_i + \underline{b}_i \quad (3)$$

where G_i , $\underline{b}_i$ vary slowly. This model allows us to obtain eigenvalue estimates of G at each step from a sequence of previous iterates. These eigenvalue estimates can be

used to find parameters for one-step, two-step, or other acceleration schemes.

Since the underlying iteration is nonlinear, the algorithm must be able to determine when acceleration will be helpful. The eigenvalue estimates can be used to predict how viable the assumption of linearity is. Also, trust regions are used to avoid catastrophic moves.

We will discuss one-step and two-step acceleration schemes and show how eigenvalue estimates may be found. These estimates can be used to choose the acceleration parameters. We will then briefly discuss the implications of nonlinearity. Finally we will present numerical results.

Due to the constraints on the length of this paper many details will be omitted. These may be found in a forthcoming Los Alamos Report.

II. ONE-STEP METHOD

If G_i and b_i are independent of i , then the convergence of the iteration

$$\underline{v}_{i+1} = G\underline{v}_i + \underline{b} \quad (4)$$

is determined by the spectral radius of G , $\sigma(G)$. Let

$$\underline{v} = G\underline{v} + \underline{b} \quad (5)$$

be the exact solution, and $\underline{e}_i = \underline{v} - \underline{v}_i$ be the error vector.

Subtracting (4) from (5) yields

$$\underline{e}_{i+1} = G \underline{e}_i = G^{i+1} \underline{e}_0 . \quad (6)$$

If the spectral radius, $\sigma(G) < 1$, then (4) will converge. One-step methods attempt to effectively reduce the spectral radius of G . Given $\underline{v}_i$, let

$$\underline{r}_i = (G\underline{v}_i + \underline{b}) - \underline{v}_i = \underline{v}_{i+1} - \underline{v}_i . \quad (7)$$

We see that (4) can be written as

$$\underline{v}_{i+1} = \underline{v}_i + \underline{r}_i . \quad (8)$$

If we let

$$\underline{v}_{i+1} = \underline{v}_i + \alpha \underline{r}_i \quad (9)$$

then (6) becomes

$$\underline{e}_{i+1} = ((1-\alpha)I + \alpha G) \underline{e}_i = ((1-\alpha)I + \alpha G)^{i+1} \underline{e}_0 . \quad (10)$$

If

$$\sigma[(1-\alpha)I + \alpha G] < \sigma(G) ,$$

then the new iteration will converge faster. The relation between the eigenvalues of G and of $(1-\alpha)I + \alpha G$ is simple, so the spectral radii are easily compared. Also note that the new iteration is consistent in that if $\underline{v}_i = \underline{v}$, then $\underline{v}_{i+1} = \underline{v}$.

The iteration (9) is called a stationary one-step method. We can also implement a nonstationary one-step method by letting α also depend upon i :

$$\underline{v}_{i+1} = \underline{v}_i + \alpha_i \underline{r}_i . \quad (11)$$

The error equation becomes

$$e_{i+1} = ((1-\alpha_i)I + \alpha_i G) e_i = \prod_{j=0}^i ((1-\alpha_j)I + \alpha_j G) e_0 . \quad (12)$$

It is clear that there is some choice for $\{\alpha_j\}_{j=0}^i$ such that

$$\sigma \left[\prod_{j=0}^i ((1-\alpha_j)I + \alpha_j G) \right] \leq \sigma(G^{i+1}) , \quad (13)$$

because $\alpha_j = 1$ yields equality.

To correctly choose α or $\{\alpha_i\}$ requires a knowledge of the spectrum of G . Consider the sequence $\{\underline{v}_i\}$ as generated by (8). The corresponding $\{\underline{r}_i\}$ has the property that

$$\underline{r}_i = G \underline{r}_{i-1} . \quad (14)$$

For sufficiently large i , the vector $\underline{r}_i$ will be dominated by a linear combination of the eigenvectors of G associated with the eigenvalues of largest modulus. Let $p_s(\lambda)$ be the orthogonal polynomial of degree s with respect to the Krylov sequence based upon G and $\underline{r}_i$ (see Householder [1,p. 24]). The roots of $p_s(\lambda)$ are approximations to the eigenvalues of G . The polynomial

$$p_s(\lambda) = \lambda^s + \rho_{s-1}^{(s)} \lambda^{s-1} + \dots + \rho_0^{(s)}$$

satisfies the least squares problem

$$\begin{pmatrix} | & & | \\ \hline \underline{r}_{i+s} & \underline{r}_{i+s-1} & \cdots & \underline{r}_i \\ | & & | \end{pmatrix} \begin{pmatrix} | \\ \hline 1 \\ \rho_{s-1}^{(s)} \\ \cdot \\ \cdot \\ \cdot \\ \rho_0^{(s)} \\ \hline | \end{pmatrix} \approx \begin{pmatrix} | \\ \hline 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ \hline | \end{pmatrix} \quad (15)$$

The residuals associated with the stationary one-step method also form a Krylov sequence. From Eq. (10) we see that

$$\underline{r}_{i+1} = ((1-\alpha)I + \alpha G) \underline{r}_i. \quad (16)$$

Thus, eigenvalue estimates of the eigenvalues of $(1-\alpha)I + \alpha G$ with largest modulus can be found. From these, eigenvalues of G can be estimated.

The residuals formed by a nonstationary one-step method can also be used to find the orthogonal polynomials with respect to the Krylov sequence formed by G and $\underline{r}_i$. As in (12) we have

$$\underline{r}_{i+1} = ((1-\alpha_i)I + \alpha_i G) \underline{r}_i. \quad (17)$$

Thus,

$$\underline{r}_{i+s} = \prod_{j=0}^{s-1} ((1-\alpha_{i+j})I + \alpha_{i+j} G) \underline{r}_i. \quad (18)$$

This polynomial relationship can be used to construct the Krylov sequence. More importantly, the orthogonal polynomial, $p_s(\lambda)$, can be constructed directly from the sequence of residuals.

Eigenvalue estimates can be used to choose new parameters. The convex hull of the eigenvalue estimates can be used to find the optimal stationary one-step parameter. If the residual is almost a linear combination of a few eigenvectors, then the parameters may be chosen so that the polynomial in (18) has roots at the corresponding eigenvalues. We will refer to this as annihilation.

III. TWO-STEP METHOD

The one-step method only uses information from the previous step. The two-step method uses the two previous iterates. Instead of (9) we have

$$\underline{v}_{i+1} = \underline{v}_i + \alpha \underline{r}_i + \beta(\underline{v}_i - \underline{v}_{i-1}) \quad (19)$$

The error equation becomes

$$\underline{e}_{i+1} = ((1+\beta-\alpha)I + \alpha G)\underline{e}_i - \beta \underline{e}_{i-1} \quad (20)$$

This error equation can also be written as a system of dimension $2N$,

$$\begin{pmatrix} \underline{e}_{i+1} \\ \underline{e}_i \end{pmatrix} = \begin{pmatrix} (1+\beta-\alpha)I + \alpha G & -\beta I \\ I & 0 \end{pmatrix} \begin{pmatrix} \underline{e}_i \\ \underline{e}_{i-1} \end{pmatrix} \quad (21)$$

To simplify the notation we rewrite this as

$$\tilde{\underline{e}}_{i+1} = \tilde{G} \tilde{\underline{e}}_i = \tilde{G}^{i+1} \tilde{\underline{e}}_0 \quad (22)$$

We would like to choose α, β to make $\sigma(\tilde{G})$ as small as possible. The solution to this can be well approximated

by fitting the best ellipse around the spectrum of G and choosing the parameters to minimize the iteration polynomial over this ellipse. We use the algorithm derived in Manteuffel [2,3].

Notice that if both sides of (20) are multiplied by $I-G$ the $\underline{e}$'s are replaced by $\underline{r}$'s. Thus, as in (22), we have

$$\tilde{\underline{r}}_{i+1} = \tilde{G} \tilde{\underline{r}}_i . \quad (23)$$

The residuals may be used to form a Krylov sequence in the operator $\tilde{G}$. Using the techniques described above we can find eigenvalue estimates of the operator $\tilde{G}$. Examination of (21) shows that the eigenvalues of $\tilde{G}$, say μ , are related to the eigenvalues of G , say λ , by the formula

$$\mu^2 - ((1+\beta-\alpha) + \alpha\lambda)\mu + \beta = 0 . \quad (24)$$

IV. DETERMINING LINEARITY

Since the analysis above is based upon a linear model it is important to determine when this model is viable. If convergence is swift, then this algorithm has no purpose. It is when each iterate differs little from its predecessor that some improvement may be found. From (7) we see that a small change from $\underline{v}_i$ to $\underline{v}_{i+1}$ implies that $\underline{r}_i$ is relatively small. This, of course, does not imply that $\underline{e}_i$ is small. When the change in $\underline{r}_i$ is linear, then the acceleration methods in Secs. II and III may be used to enhance the convergence.

Linearity can be ascertained by monitoring a moving window of past residuals. Consider the matrix in (15). If the previous residuals are linearly dependent then this matrix has a null space. If this null space differs little from step to step we may assume that the iteration is behaving linearly and the error is in the direction of only a few eigenvectors of G . A sequence of acceleration parameters may be chosen to annihilate these errors.

Even if these previous residuals are not linearly dependent, linear behavior may be detected by monitoring the eigenvalue estimates obtained as described above. If the eigenvalue estimates differ little from step to step we may assume that the iteration is behaving linearly and use the eigenvalue estimates to choose stationary one- or two-step parameters. This is done by accumulating a pile of previous eigenvalue estimates and forming the convex hull of the pile. Because the underlying iteration is nonlinear, these estimates must be aged and discarded after an appropriate number of steps.

Note that the eigenvalue estimates can be used to determine the relationship between $\underline{r}_i$ and $\underline{e}_i$. Thus, the convergence criteria can be greatly improved.

V. TRUST REGIONS

It is assumed that the underlying nonlinear iteration will converge, albeit slowly, if left alone. From (8) we see that $\underline{r}_i$ is a safe step. From (9) and (19) it is clear how this safe step is being altered. If the new step is shorter, then no problems should arise.

However, if the step is lengthened substantially then care must be taken. The eigenvalue estimates from which the parameters are chosen also yield a predicted reduction in the residual. If the actual reduction is substantially different then the step length will have to be shortened.

VI. NUMERICAL RESULTS

To test the effectiveness of the acceleration methods, we studied the convergence of a second-order Runge-Kutta method to the approximate steady-state solution of the scalar three-dimensional Burger's equation

$$u_t + \frac{1}{2} \nabla u^2 = \varepsilon \Delta u \quad (25a)$$

on the unit cube. The small parameter $\varepsilon = 0.02$ and Dirichlet boundary conditions

$$\begin{aligned} u(0, y, z) &= u(x, 0, z) = u(x, y, 0) = 0 \\ u(1, y, z) &= u(x, 1, z) = u(x, y, 1) = 1 \end{aligned} \quad (25b)$$

were chosen so there was a thin boundary layer. The equation was differenced in divergence form with second order centered finite differences on a uniform grid with N grid points. The time step was chosen to be

$$\Delta t = \frac{1}{2} \Delta x / \|u\|_{\max} = \frac{1}{2} \Delta x.$$

The results for $N = 8,000$ are shown in Fig. 1. Similar results, not described here, are obtained for $N = 27,000$ and $N = 64,000$. We compared three methods:

Fig. 1. The residual errors in solving the steady-state solution to the 3-D scalar Burgers equation (25).

Method 1 - No acceleration.

Method 2 - After four time steps, eigenvalue estimates of G are obtained and used to generate stationary two-step acceleration parameters. From then on, up to four eigenvalue estimates are obtained on every iteration using the past five residuals. If the residual increases on an iteration, as happened on iterations 34, 35, and 36, or if twelve iterations have been made since new acceleration parameters were calculated, as happened on iterations 16, 28, and 48, then new parameters are calculated. These are marked by dots in Fig. 1. The new parameters are generated based on the best ellipse enclosing the past twenty five eigenvalue estimates. The maximum accelerated step-length in this run was limited to less than five times the original step-length.

Method 3 - This method is included because of its simplicity and effectiveness. If a single real eigenvalue estimate is obtained on every time step using only the previous two residuals,

$$\lambda_i = \lambda_{i-1} + (1-\lambda_{i-1}) \langle \underline{r}^i, \underline{r}^{i-1} \rangle / \langle \underline{r}^{i-1}, \underline{r}^{i-1} \rangle, \quad (26)$$

then the annihilation one-step acceleration parameter is

$$\alpha_i = 1/(1-\lambda_i) \quad . \quad (27)$$

Again we limited the maximum step size to less than five times the original step size ($\alpha_i \leq 5$).

Its clear that both acceleration methods substantially improve the convergence to the steady-state solution. Method 3 works extremely well on this problem, but we mention that it is much more prone to instabilities, especially when the real part of the some of the complex eigenvalues of G are close to one.

Without the trust region model the iterations soon diverge. If the underlying equation is linear (as in some examples not presented here including a linear Burger's equation and an SOR iterative solution of Laplaces equation) the trust region is less critical and methods can recover.

REFERENCES

1. Householder, A. S., The Theory of Matrices in Numerical Analysis, Blaisdell Publishing Co., New York, 1964.
2. Manteuffel, T. A., "Adaptive Procedure for Estimations Parameters for the Nonsymmetric Tchebychev Iteration," Numer. Math. 31 (1978), 183-208.
3. Manteuffel, T. A., "Optimal Iteration Parameters for Linear Second-Degree Stationary Iterative Methods," SIAM J. Numer. Anal. V. 19, No. 4, Aug. 1982.