CONF-8810515--2

DE90 013963

# The Search for Omega

by

David M. Young and Tsun-Zee Mai

August 1988 CNA-223

An invited talk based on this paper will be presented at the Conference on Iterative Methods for Large Linear Systems to be held at The University of Texas, Austin, Texas, on October 19–21. The paper will appear in the Conference Proceedings.

---

# The Search for Omega[1]

by

David M. Young[2] and Tsun-Zee Mai[3]

## Abstract

For the effective use of iterative algorithms for solving large sparse linear systems it is often necessary to select certain iteration parameters. Examples of iteration parameters are the relaxation factor, omega, for the SOR and SSOR methods and the largest and smallest eigenvalues of the matrix for a basic iterative method when Chebyshev acceleration is used to speed up the convergence. For many iterative algorithms the performance is extremely sensitive to the choice of iteration parameters. Moreover, uncertainty as to how to choose iteration parameters has often, in the past, tended to discourage the use of iterative methods, as opposed to direct methods, for certain classes of problems.

The purpose of this paper is to review the development of procedures for choosing iteration parameters, with special emphasis on methods applicable to linear systems arising from the numerical solution of partial differential equations. The discussion will include a priori procedures including analytic techniques, spectral methods, and methods based on related differential equations. Automatic, or "adaptive", procedures, wherein the iteration parameters are improved as the computation proceeds, will also be discussed. Some of these procedures have been incorporated into the ITPACK software packages for solving large sparse linear systems. Numerical experiments indicate that the amount of overhead needed to determine satisfactory parameters is usually not excessive. Methods for choosing iteration parameters for nonsymmetric systems will also be considered.

[1] The research was supported in part by the National Science Foundation under Grant DCR-8518722, by the Department of Energy under Grant DE-FG05-87ER25048, and by the U.S. Air Force Office of Scientific Research and Development under Grant AFSOR-85-0052 with The University of Texas at Austin.

[2] Center for Numerical Analysis, The University of Texas, Austin, Texas.

[3] Mathematics Department, The University of Alabama, Tuscaloosa, Alabama.

## 1. Introduction

For many iterative algorithms for solving large sparse linear systems the rate of convergence is very sensitive to the choice of the iteration parameters. It seems fair to say that uncertainty as to how to choose iteration parameters has been an important stumbling block which has tended to discourage the use of iterative methods, as opposed to direct methods, for solving certain classes of problems. In this paper we review some of the procedures which have been developed over the years for choosing iteration parameters.

A prime example of an iteration parameter is the relaxation factor, omega, which is used for the SOR method and for the SSOR method. Much of our discussion will be devoted to the "search for omega" for these methods. Another example of an iterative algorithm involving iteration parameters arises when Chebyshev acceleration is used to speed up the convergence of a basic iterative method with an iteration matrix $G$. Here the iteration parameters which are needed are the smallest and largest eigenvalues of $G$.

Other examples of iteration parameters include: the shift parameter for the shifted incomplete Cholesky method proposed by Manteuffel [1980]; the parameters for the strongly implicit method, see e.g. Stone [1968] and Dupont *et al* [1968]; and the parameters for the alternating direction implicit method of Peaceman and Rachford [1955].

Before discussing procedures for choosing iteration parameters we give a review, in Section 2, of various types of iterative algorithms where iteration parameters arise. In Section 3 we describe *a priori* procedures for choosing iteration parameters. Here one attempts to estimate the parameters in advance. The techniques used include analytic techniques, spectral techniques, and techniques involving the use of related partial differential equations. While such techniques are often useful, in many cases they do not yield sufficiently accurate parameter estimates. In Section 4 we describe adaptive techniques where improved estimates are determined automatically as the computational process proceeds. In many cases these procedures yield satisfactory parameters without an excessive amount of additional computational effort. Some of the adaptive procedures have been incorporated into the ITPACK software packages. The discussion includes methods used in searching for omega for the SOR method as well as adaptive methods for finding acceleration parameters for Chebyshev acceleration. Adaptive SSOR with Chebyshev and conjugate gradient acceleration and variational-based adaptive methods are also considered.

For most of our discussion it is assumed that the matrix $A$ of the linear system is symmetric and positive definite (SPD). The case where the matrix $A$ is nonsymmetric is much more complicated. Various approaches for choosing iteration parameters for the SOR method and for Chebyshev acceleration are considered in Section 5. A composite adaptive procedure, based on the use of a generalized conjugate gradient method, is described for finding parameters associated with certain basic iterative methods.

We wish to emphasize that our review is by no means complete. There are no doubt

many excellent schemes of which we are not aware. We would be grateful to have these called to our attention.

## 2. Iterative Algorithms and Iteration Parameters

In this section we describe some iterative algorithms for solving the linear system

$$Au = b \tag{2.1}$$

where $A$ is given square matrix and $b$ is a given column vector. Usually, as in the case of a linear system derived from the discretization of a partial differential equation, the matrix $A$ is very large and very sparse.

A typical iterative algorithm for solving the linear system (2.1) consists of a basic iterative method together with an acceleration procedure. A basic iterative method is a one-step procedure of the form

$$u^{(n+1)} = Gu^{(n)} + k \tag{2.2}$$

where

$$\begin{cases} G = I - Q^{-1}A \\ k = Q^{-1}b. \end{cases} \tag{2.3}$$

Here $Q$ is the "splitting" matrix corresponding to the basic iterative method and the matrix $G$ is the iteration matrix. The matrix $Q$ is usually chosen to be a simple easily-inverted matrix such as a diagonal, tridiagonal, upper triangular or lower triangular matrix or as a product of such matrices.

Examples of frequently-used basic iterative methods include:

| | | |
|---|---|---|
| Richardson's method | : | $Q = I$ |
| Jacobi method | : | $Q = D$ |
| SOR method | : | $Q = \frac{1}{\omega}D - C_L$ |
| SSOR method | : | $Q = \frac{\omega}{2-\omega}(\frac{1}{\omega}D - C_L)D^{-1}(\frac{1}{\omega}D - C_U)$ |

Here $A = D - C_L - C_U$ where $D$ is the diagonal matrix with the same diagonal elements as $A$ and where $C_L$ and $C_U$ are strictly lower and strictly upper triangular matrices, respectively. The parameter $\omega$ is known as the "relaxation factor."

For some basic iterative methods, often referred to as "approximate factorization methods," $Q$ has the form $LU$ where $L$ and $U$ are lower and upper triangular matrices, respectively. (Often $L$ and $U$ have the same "sparsity" as $A$). Examples of such methods are the incomplete Cholesky (IC) method considered by Varga [1960] and by Meijerink and Van der Vorst [1977], the shifted incomplete Cholesky (SIC) method of Manteuffel [1980], and the modified incomplete Cholesky (MIC) method of Gustafsson [1979]. Another

example is the strongly implicit (SIP) method of Stone [1968]. See also Dupont *et al* [1968]. Except for the IC method, each of the above methods involves a parameter. We refer to a parameter associated with a basic iterative method as a "splitting parameter."

In many problems arising from the numerical solution of partial differential equations by discretization methods the rate of convergence of the SOR method with the optimum value of $\omega$ is faster by an order-of-magnitude than that of the Jacobi method. However, the rate of convergence is very sensitive to $\omega$. Similarly, the rates of convergence of many other basic iterative methods are sensitive to the choice of the splitting parameter.

A basic iterative method (2.2) is *symmetrizable* if the matrix $I - G$ is similar to an SPD matrix or, equivalently, if $Z(I - G)$ is SPD for some SPD matrix $Z$. Most of the standard iterative methods, involving Richardson's method, the Jacobi method, the SSOR method, and the IC method are symmetrizable if $A$ is SPD. More generally, a basic iterative method is symmetrizable if $A$ and $Q$ are SPD. It should be noted, however, that the SOR method is not symmetrizable.

If a basic iterative method (2.2) is symmetrizable, then the eigenvalues of $G$ are real and less than unity. The convergence of a symmetrizable basic iterative method can be speeded up by the use of a polynominal acceleration procedure. Examples of polynominal acceleration procedures are Chebyshev acceleration, see e.g. Varga [1957, 1962], Golub and Varga [1961], and Hageman and Young [1981], and conjugate gradient (CG) acceleration.

The formulas for Chebyshev acceleration are given by

$$u^{(n+1)} = \rho_{n+1}\{u^{(n)} + \gamma_{n+1}(Gu^{(n)} + k - u^{(n)})\} + (1 - \rho_{n+1})u^{(n-1)}. \tag{2.4}$$

Here the parameters $\{\rho_{n+1}\}$ and $\{\gamma_{n+1}\}$ are given by

$$\gamma_1 = \gamma_2 = \cdots = \frac{2}{2 - M(G) - m(G)} \tag{2.5}$$

$$\rho_{n+1} = \begin{cases} 1 & n = 0 \\ (1 - \frac{\sigma^2}{2})^{-1} & n = 1 \\ (1 - \frac{\sigma^2}{4}\rho_n)^{-1} & n \geq 2 \end{cases}$$

where

$$\sigma = \frac{M(G) - m(G)}{2 - M(G) - m(G)} \tag{2.6}$$

where $m(G)$ and $M(G)$ are the smallest and largest eigenvalues of $G$, respectively. In a given case one normally uses estimates $m_E$ and $M_E$ for $m(G)$ and $M(G)$ respectively. We refer to the parameters $m_E$ and $M_E$ as "acceleration parameters."

For a symmetrizable basic iterative method the rate of convergence of the Chebyshev acceleration procedure, with exact estimates $m_E$ and $M_E$, is an order-of-magnitude faster

than that of the unaccelerated method. However, the rate of convergence is often very sensitive to the choice of $m_E$ and $M_E$. Moreover in the case of Chebyshev acceleration applied to a basic iterative method with a splitting parameter, such as the SSOR method, one must choose the splitting parameter as well as the acceleration parameters, $m_E$ and $M_E$. In the case of the SSOR method, however, it can be shown that $m(G) \geq 0$; hence one can choose $m_E = 0$.

For conjugate gradient acceleration of a symmetrizable basic iterative method one must choose an auxiliary matrix $Z$ such that $Z$ and $Z(I-G)$ are SPD. In the case where $A$ and $Q$ are SPD, typical choices of $Z$ are $Z = A$ and $Z = Q$. There are several equivalent forms of CG acceleration; see e. g. the discussion of Young *et al* [1988]. One such form, which was considered by Concus *et al* [1976], see also Hageman and Young [1981], is given by (2.4) where

$$\gamma_{n+1} = \frac{(Z\delta^{(n)}, \delta^{(n)})}{(Z\delta^{(n)}, (I-G)\delta^{(n)})} \tag{2.7}$$

$$\rho_{n+1} = \left[1 - \frac{\gamma_{n+1}}{\gamma_n} \frac{(Z\delta^{(n)}, \delta^{(n)})}{(Z\delta^{(n-1)}, \delta^{(n-1)})} \frac{1}{\rho_n}\right]^{-1}, \quad n \geq 1, \quad (\rho_1 = 1).$$

Here the "pseudo-residual" vector $\delta^{(n)}$ is given by

$$\delta^{(n)} = Gu^{(n)} + k - u^{(n)}. \tag{2.8}$$

It should be noted that there are no acceleration parameters involved. Moreover, it can be shown that, for a given basic iteration method, CG acceleration converges at least as fast (in a certain norm) as any polynomial acceleration procedure including Chebyshev acceleration, applied to the basic method. On the other hand, CG acceleration requires the computation of inner products after each iteration and thus may require more time per iteration than Chebyshev iteration. Also, for some parallel computers the communication time needed to compute the inner products may be costly.

It should be noted that, given an accurate estimate $M_E$ of $M(G)$, one can obtain a good bound on the norm of the error $\|u^{(n)} - \bar{u}\|$, where $\bar{u} = A^{-1}b$ is the true solution of (2.1), in terms of the norm of the pseudo-residual vector $\delta^{(n)}$. Thus one can estimate $\|u^{(n)} - \bar{u}\|$ by

$$\|u^{(n)} - \bar{u}\| \approx \frac{1}{1 - M_E} \|\delta^{(n)}\| \tag{2.9}$$

(For a suitable choice of norm one can replace "$\approx$" by "$\leq$".) Thus an accurate estimate of $M(G)$ is useful in determining when an iterative process should be terminated. We remark that for the CG method an accurate estimate of $M(G)$ can be obtained by solving an eigenvalue problem for a tridiagonal matrix involving the $\{\gamma_i\}$ and the $\{\rho_i\}$ of (2.7); details are given in Hageman and Young [1981]; see also Mai [1986].

## 3. A Priori Techniques

In this section and in the next section we review some of the techniques which can be used to estimate optimum iteration parameters. We consider two types of procedures, namely "a priori" techniques, which are usually carried out before the iteration process begins, and "adaptive techniques" which are carried out along with the iterative process. The a priori techniques described in this section include analytic techniques, spectral methods and procedures involving the use of related differential equations.

### Analytic Techniques

In some cases it is possible to relate the eigenvalues of the iteration matrix $G$ corresponding to a relatively complicated iterative method, such as the SOR method, to those of a relatively simple method, such as the Jacobi method. Thus, for example, if $A$ is a consistently ordered matrix then the eigenvalues $\lambda$ of the SOR matrix and the eigenvalues $\mu$ of the iteration matrix $\mathcal{B}$ of the Jacobi method are related by

$$\lambda + \omega - 1 = \omega\mu\sqrt{\lambda} \tag{3.1}$$

see Young [1971] for details. If $A$ is an SPD matrix the optimum value of $\omega$ for the SOR method is

$$\omega_b = \frac{2}{1 + \sqrt{1 - S(\mathcal{B})^2}} \tag{3.2}$$

where $S(\mathcal{B})$ is the spectral radius of $\mathcal{B}$. Thus for the consistently ordered case the determination of the optimum value of $\omega$ can be reduced to the determination of $S(\mathcal{B})$. Moreover for some linear systems derived from problems involving elliptic partial differential equations with constant coefficients one can find $S(\mathcal{B})$ using spectral methods as described below.

Even if $A$ is not consistently ordered, but is an "$L$-matrix" (i.e. if $a_{i,i} > 0$ for all $i$ and $a_{i,j} \leq 0$ for $i \neq j$) the use of (3.2) gives a "good" value of $\omega$. This follows from work of Kahan [1958]; see also Wachspress [1966] and Young [1971], Chapter 12.

Another example of the use of an analytic technique involves the SSOR method for the case where $A$ is SPD.* Good estimates for the optimum value of $\omega$ and the corresponding value of $S(\mathcal{S}_\omega)$, where $\mathcal{S}_\omega$ is the matrix of the SSOR method, can be obtained; see for instance Habetler and Wachspress [1961], Ehrlich [1964], and Young [1971, 1972, 1977]. The values of $\omega$ and $S(\mathcal{S}_\omega)$ are given in terms of the largest eigenvalue $M(\mathcal{B})$ of the matrix $\mathcal{B}$ corresponding to the Jacobi method and $S(LU)$ where $L$ and $U$ are strictly lower triangular

* Ehrlich [1964] obtained estimates for the optimum $\omega$ for block SSOR for certain separable problems using spectral techniques.

and strictly upper triangular matrices, respectively, and $\mathcal{B} = L + U$. Thus it is shown in Young [1977] that a good value, $\omega_1$, of $\omega$ is given by

$$\omega_1 = \frac{2}{1 + \sqrt{1 - 2M(\mathcal{B}) + 4\tilde{\beta}}} \tag{3.3}$$

where

$$\tilde{\beta} = \max\{\frac{1}{4}, S(LU)\} \tag{3.4}$$

The corresponding value of $S(\mathcal{S}_{\omega_1})$ satisfies

$$S(\mathcal{S}_{\omega_1}) \leq \frac{1 - \frac{1-M(\mathcal{B})}{\sqrt{1-2M(\mathcal{B})+4\tilde{\beta}}}}{1 + \frac{1-M(\mathcal{B})}{\sqrt{1-2M(\mathcal{B})+4\tilde{\beta}}}} \tag{3.5}$$

It is not difficult to see that for the SSOR method to be effective $S(LU)$ must not exceed $\frac{1}{4} + \epsilon$ for some small $\epsilon$. This condition is satisfied if one uses the natural ordering of the grid points and if the coefficients of the differential equation are sufficiently smooth; see Young [1977].

## Spectral Methods

For linear systems corresponding to separable partial differential equations with constant coefficients over rectangular regions, the eigenvalues of $A$ can often be found using "spectral methods," i.e. methods based on the use of finite Fourier series.* Thus, consider the "model problem" based on the Dirichlet problem in the unit square. Using the standard 5-point difference equation with a square grid of size $h = M^{-1}$ with $M$ an integer, we have the difference equation

$$u(x+h,y) + u(x-h,y) + u(x,y+h) + u(x,y-h) - 4u(x,y) = 0 \tag{3.6}$$

for all interior grid points $(x, y)$. Multiplying the equations by $-1$ and moving the terms involving the (known) boundary values to the right hand side, we obtain a system of the form (2.1) where $A$ has 4's on the main diagonal, $-1$'s in certain off-diagonal elements and 0's elsewhere. It can be verified directly that the eigenvectors of the difference operator corresponding to $A$ are

$$\nu_{p,q}(x,y) = \sin p\pi x \sin q\pi y \tag{3.7}$$

* For one of the earliest papers to use finite Fourier series to analyze difference equations see Phillips and Wiener [1923].

and the corresponding eigenvalues are

$$\mu_{p,q} = \frac{1}{2}[\cos p\pi h + \cos q\pi h] \tag{3.8}$$

where $p, q = 1, 2, \cdots, M-1$. Thus the eigenvalues $\mu$ of the Jacobi method lie in the interval

$$-\cos \pi h \leq \mu \leq \cos \pi h. \tag{3.9}$$

Hence Chebyshev acceleration can be used for the Jacobi method with

$$-m_E = M_E = \cos \pi h. \tag{3.10}$$

A similar technique can be used for rectangular regions with an elliptic differential equation of the form

$$(Au_x)_x + (Cu_y)_y + Fu = G \tag{3.11}$$

where $A, C$ and $F$ are constants; see e.g. Young [1971], Chapter 6. Again $M(\mathcal{B})$ and $m(\mathcal{B})$ can be computed. Moreover, bounds on $M(\mathcal{B})$ and $m(\mathcal{B})$ can be computed even for non-rectangular regions and for variable coefficients. The bounds involve upper and lower bounds for $A, C$, and $F$ in the region. For non-rectangular regions a monotonicity theorem is used which shows that $S(\mathcal{B})$ is an increasing function of the region involved.

Spectral methods can sometimes be used to determine the eigenvalues of the iteration matrix $\mathcal{L}_\omega$, for the SOR method even when no relation, such as (3.1), is available between the eigenvalues of $\mathcal{L}_\omega$ and the eigenvalues of $\mathcal{B}$. Thus, if one uses the standard 9-point difference equation instead of the 5-point difference equation for the model problem, the matrix $A$ is not consistently ordered and (3.1) does not hold. (Of course such a relation would hold if line SOR were used.) Nevertheless the difference equation is separable and the eigenvalues of the SOR method can be determined, implicitly, as described by Adams, LeVeque and Young [1986]; (see also Van de Vooren and Vliegenthart [1967]), by the following polynominal equation

$$\begin{aligned} &\alpha^4 - [\frac{4}{5}\omega c_q + \frac{2}{25}\omega^2 c_p^2 c_q]\alpha^3 - [2(1-\omega) - \frac{4}{25}\omega^2 c_q^2 + \frac{1}{25}\omega^2 c_p^2(c_q^2+4)]\alpha^2 \\ &+ [\frac{4}{5}\omega(1-\omega)c_q - \frac{2}{25}\omega^2 c_p^2 c_q]\alpha + (1-\omega)^2 = 0 \end{aligned} \tag{3.12}$$

Here $\lambda = \alpha^2$ is an eigenvalue of $\mathcal{L}_\omega$ and

$$\begin{cases} c_p = \cos p\pi h \\ c_q = \cos q\pi h \end{cases} \tag{3.13}$$

To find $S(\mathcal{L}_\omega)$ one must solve the above quartic equation numerically for $p, q = 1, \cdots, M-1$. One can then determine $S(\mathcal{L}_\omega)$. Next, one can find the value of $\omega$ which minimizes $S(\mathcal{L}_\omega)$ by a direct search procedure.

Using asymptotic analysis it can be shown that for small $h$

$$S(\mathcal{L}_\omega) \approx 1 - 1.79\pi h \tag{3.14}$$

We remark that an approximate asymptotic result, namely,

$$S(\mathcal{L}_\omega) \approx 1 - 2.35\pi h \tag{3.15}$$

can be obtained using the much less laborious analysis of Garabedian [1956], which is based on the use of a related hyperbolic differential equation.

## Use of Differential Equations

A technique which has sometimes been used for analyzing the behavior of an iterative method for solving an elliptic problem is to consider a related time-dependent problem. Essentially the iteration number, $n$, is replaced by the time variable $t$. Thus, suppose one is trying to solve the model problem using the "extrapolated Jacobi method" defined by

$$u^{(n+1)} = \gamma(\mathcal{B}u^{(n)} + c) + (1-\gamma)u^{(n)} \tag{3.16}$$

where

$$\begin{cases} \mathcal{B} = I - D^{-1}A \\ c = D^{-1}b \end{cases} \tag{3.17}$$

Here $D$ is defined in Section 2. The number $\gamma$ is the extrapolation factor. The extrapolated Jacobi method corresponds to the forward difference method for solving the time-dependent equation

$$u_t = u_{xx} + u_{yy}. \tag{3.18}$$

The extrapolation factor $\gamma$ depends on the time step. If one uses variable time steps, this correspond to the use of variable extrapolation with the Jacobi method, as defined by

$$u^{(n+1} = \gamma_{n+1}(\mathcal{B}u^{(n)} + c) + (1-\gamma_{n+1})u^{(n)} \tag{3.19}$$

Here again the $\{\gamma_{n+1}\}$ are determined by the time steps. We remark that with a suitable choice of the $\{\gamma_{n+1}\}$ the convergence of the Jacobi method can be speeded up by an amount comparable to the speedup achieved using the Chebyshev acceleration. The use of variable extrapolation with the Jacobi method was studied by Huang [1987] in his PhD thesis.

Another example of the use of a related differential equation to analyze the convergence of an iterative method for solving a linear system is the analysis of Garabedian [1956] for the SOR method to which we have previously referred.

## 4. Adaptive Techniques

In many cases it is difficult or impractical to determine the iteration parameters in advance to sufficient accuracy. For example suppose one is using the SOR method and that the coefficient matrix $A$ is consistently ordered. If the true value of $S(\mathcal{B})$ is .9999 but if one uses an estimate of .99 to determine omega, the rate of convergence may be reduced by a factor of nearly 20. This illustrates the fact that it is seldom possible to get a sufficiently good estimate of $S(\mathcal{B})$ using a priori methods. In order to get $S(\mathcal{B})$ to sufficient accuracy it is often necessary to use adaptive techniques.

The basic idea of an adaptive procedure is as follows: One starts the iterative process with initial estimates of the optimum parameters. Periodically the convergence of the procedure is tested and compared with the expected convergence rate (which one would obtain if the exact optimum parameters were used). If the convergence rate is satisfactory the process is continued; otherwise new parameter estimates are determined according to some procedures. The development of such procedures is not simple in general, since one must develop procedures to tell how rapidly the method, with the given choice of parameters, is converging and, if the method is found to be converging too slowly, how the parameters should be changed to speed up the convergence.

In this section we discuss the following: the search for omega for the SOR method; adaptive procedures for Chebyshev acceleration; and the choice of omega for the SSOR method with Chebyshev and CG acceleration. We also consider variational-based adaptive procedures, including a composite adaptive procedure developed by Mai [1986].

### The Search for Omega for the SOR Method

Early work on the search for omega for the SOR method was based on the use of (3.1) and (3.2) as well as other related formulas. One such approach was considered by Young [1955]. Use was made of the fact that if $A$ is consistently ordered then the eigenvalues of the matrix $\mathcal{L}$ for the Gauss-Seidel method are the square of the eigenvalues of the matrix $\mathcal{B}$ for the Jacobi method. Thus by (3.2) the optimum value of $\omega$ is given by

$$\omega_b = \frac{2}{1 + \sqrt{1 - S(\mathcal{L})}} \tag{4.1}$$

The eigenvalues of $\mathcal{L}$ are real, nonnegative and less than one. Hence one can try to determine $S(\mathcal{L})$ using the ordinary power method. Then, having estimated $S(\mathcal{L})$, one can determine $\omega_b$ by (4.1).

A drawback of the above scheme is that many iterations of the Guass-Seidel method are required to determine $S(\mathcal{L})$ and that the iterations are essentially wasted. A much more

sophisticated scheme was used by Hageman and Kellogg [1968] where the power method is speeded up by the use of Chebyshev polynomials. Other procedures are described by Varga [1957] and by Bilodeau *et al* [1957].

Ideally, one would like to avoid carrying out "wasted iterations." Instead it would be desirable if every iteration could be useful, to some extent, even if perhaps not as useful as though the optimum value of the $\omega$ is used. Several schemes have been proposed for doing this including schemes of Kulsrud [1961], Carré [1961], Rigler [1965] and Reid [1966]. Each scheme is based on the following ideas. If one knows the spectral radius of the SOR matrix, $\mathcal{L}_\omega$, corresponding to a given $\omega$, one could, using (3.1) obtain a good estimate of $S(\mathcal{B})$ by

$$S(\mathcal{B}) = \frac{S(\mathcal{L}_\omega) + \omega - 1}{\omega\sqrt{S(\mathcal{L}_\omega)}} \tag{4.2}$$

Having found $S(\mathcal{B})$ one could get an improved value of $\omega$ using (3.2).

The basic procedure is as follows. Choose a value of $\omega$ not greater than $\omega_b$. (For example let $\omega = 1.$) Iterate for several iterations and estimate $S(\mathcal{L}_\omega)$. Then compute $S(\mathcal{B})$ from (4.2) and then a new value of $\omega$ by (3.2).

The problem of adaptively determining the optimum value of $\omega$ for the SOR method for the case where the coefficient matrix is SPD and consistently ordered is complicated by the fact that the matrix $\mathcal{L}_\omega$ of the SOR method may have principal vectors of grade 2. This leads to slow the convergence and, together with the fact that the eigenvalues of $\mathcal{L}_\omega$ are often complex, makes the convergence erratic and also makes it hard to estimate $S(\mathcal{L}_\omega)$. A rather elaborate program, which is based on the ideas given above and which takes into account a great many of the potential difficulties, has been developed by L. Hageman. The program is described in Chapter 9 of Hageman and Young [1981]. The program is also included in the ITPACK 2C software package; see Kincaid *et al* [1982].

## Adaptive Chebyshev Acceleration

For Chebyshev acceleration of a basic iterative method two acceleration parameters, $m_E$ and $M_E$, are normally required in addition to the splitting parameters, if any. However, in many cases a good lower bound, say $\underline{m}$, is available for $m(G)$. An example is the SSOR method where it is known that $0 \leq m(G)$. In such cases one can let $m_E$ be fixed and equal to $\underline{m}$. There remains one acceleration parameter, $M_E$, to be determined. For a given value of $M_E$ one can determine how fast the method should be converging if $M_E$ were equal to $M(G)$. Moreover one can measure the actual rate of convergence. Also procedures are available to improve $M_E$ whenever it is found that the convergence is too slow. An adaptive procedure based on these components is described in Hageman and Young [1981]. An important property of the procedure is that the values of $M_E$ are nondecreasing and are never greater than $M(G)$.

Extensive numerical experiments over a wide range of problems indicate that the "overhead" associated with the adaptive process is seldom more than 25%-30%; see Hageman and Young [1981] and Jea and Young [1988]

Several approaches have been used for the case where $m_E$ and $M_E$ must both be determined adaptively, see e.g. Diamond [1971], Hageman and Young [1981], and Mai [1986]. Recently Mai and Young [1988] have considered the use of a "dual adaptive procedure," based on a suggestion of Dr. Tom Manteuffel. Numerical experiments based on this procedure indicate that it is somewhat more effective in many cases then the Hageman and Young procedure and the Mai procedure.

An alternative approach for finding $m(G)$ and $M(G)$ based on the use of modified moments has been used by Golub and Kent [1987].

Adaptive SSOR With Chebyshev and CG Acceleration *

Since the SSOR method is symmetrizable it is usually applied in conjunction with an acceleration procedure – either Chebyshev acceleration or CG acceleration. In either case the relaxation factor $\omega$ must be estimated. If Chebyshev acceleration is used $M(\mathcal{S}_\omega)$ must also be determined. In Hayes and Young [1977] and Grimes *et al* [1979] an adaptive procedure is described which is focused on estimations of $S(\mathcal{B})$ and $S(LU)$. The procedure is based on analytic formulas given in Section 3 relating the good value of $\omega$ and the corresponding value of $M(\mathcal{S}_\omega)$ to $S(\mathcal{B})$ and $S(LU)$. The procedure is used in the ITPACK 2C software package; see Kincaid *et al* [1982]. Another adaptive procedure for finding $\omega$ was given by Benokraitis [1974]. An alternative procedure based on a composite adaptive procedure corresponding to a variational method is described below.

Variational-Based Adaptive Methods: The Composite Adaptive Procedures

Let us again consider the linear system (2.1) where $A$ is SPD. It is easy to show that the problem of solving (2.1) is equivalent to that of minimizing the quadratic form

$$F(u) = \frac{1}{2}(u, Au) - (b, u) \tag{4.3}$$

Suppose that we have a basic iterative method (2.2) for solving (2.1) where the splitting matrix $Q$ depends on a (splitting) parameter, say $\omega$. Given $u^{(n)}$ it would seem reasonable to choose $\omega$ to minimize $F(u^{(n+1)})$. In the case of the SOR method, for example, it can be shown that, (assuming $D = I, L = C_L$, and $U = C_U$)

$$F(u^{(n+1)}) - F(u^{(n)}) = -\frac{1}{2}\left[\frac{2-\omega}{\omega}\right](\delta^{(n)}, \delta^{(n)}) \tag{4.4}$$

* Evans and Forrington [1963] considered an iteration procedure for finding the optimum value of SSOR using a procedure not related to an acceleration procedure.

where $\delta^{(n)} = \omega(I - \omega L)^{-1} r^{(n)}$ and $r^{(n)} = b - Au^{(n)}$. One could determine $\omega$ to minimize $F(u^{(n+1)})$ at each step or periodically. This might involve a direct search procedure. However, in some cases, for example for the case where $A$ is red-black the relation $(I - \omega L)^{-1} = I + \omega L$ holds and the problem of minimizing $F(u^{(n+1)})$ is equivalent to that of minimizing a quartic polynomial in $\omega$.

We now describe a somewhat more general approach which is related to the method of steepest descent.

Given $u^{(n)}$ choose $\lambda_n$ and $\omega$ to minimize

$$F(u^{(n+1)}) = F(u^{(n)}) - \lambda_n(\delta^{(n)}, r^{(n)}) + \frac{\lambda_n^2}{2}(\delta^{(n)}, A\delta^{(n)}). \tag{4.5}$$

Here $\delta^{(n)} = Q(\omega)^{-1} r^{(n)}$.

The minimization of $F(u^{(n+1)})$ with respect to $\lambda_n$ gives

$$\lambda_n = \frac{(\delta^{(n)}, r^{(n)})}{(\delta^{(n)}, A\delta^{(n)})} \tag{4.6}$$

and

$$F(u^{(n+1)}) - F(u^{(n)}) = -\frac{1}{2}\frac{(\delta^{(n)}, r^{(n)})^2}{(\delta^{(n)}, A\delta^{(n)})}. \tag{4.7}$$

One then can seek to choose $\omega$ to minimize $F(u^{(n+1)})$. Again this might be done after each iteration or periodically. A direct search might be required unless an analytic formula can be found.

We now consider a more general approach which we refer to as the "composite adaptive procedure." We assume that $Q(\omega)$ is SPD for each $\omega$ so that the basic iterative method is symmetrizable. It can be shown that if CG acceleration with $Z = Q$ is applied to (2.2) then, for each $n$, $F(u^{(n+1)})$ is minimized with respect to any polynomial acceleration procedure based on (2.2). Here we write the CG acceleration procedure in the two-term form

$$\begin{aligned} u^{(n+1)} &= u^{(n)} + \lambda_n p^{(n)} \\ p^{(n)} &= \delta^{(n)} + \alpha_{n,n-1} p^{(n-1)} \end{aligned} \tag{4.8}$$

where $\lambda_n$, and $\alpha_{n,n-1}$ are chosen to minimize $F(u^{(n+1)})$. For the composite adaptive procedure where $\omega$ varies we choose an integer $s \geq 0$ and let

$$p^{(n)} = \delta^{(n)} + \alpha_{n,n-1} p^{(n-1)} + \cdots + \alpha_{n,n-s} p^{(n-s)} \tag{4.9}$$

where again $\alpha_{n,n-1}, \alpha_{n,n-2}, \cdots, \alpha_{n,n-s}, \lambda_n$ and $\omega$ are chosen to minimize $F(u^{(n+1)})$.

The case $s = 0$ reduces to the procedure defined by (4.5) – (4.7) above. The case $s = 1$ has been analyzed by Mai [1986]. Numerical experiments have been carried out which show that the convergence rate is improved and that the optimum value of $\omega$ is obtained to reasonable accuracy. However, the amount of work required to find $\omega$ at each step is quite high. It is expected that with the use of parallel computers the search for the optimum $\omega$ can be greatly speeded up.

## 5. The Nonsymmetric Case

For many problems involving partial differential equations the matrix $A$ is not SPD. For example consider the convection-diffusion equation

$$u_{xx} + u_{yy} + Du_x = 0. \tag{5.1}$$

If the standard 5-point finite difference equation based on central difference formulas is used, the matrix $A$ is not symmetric and hence is not SPD. If $A$ is not SPD, then most of the standard basic iterative methods are not symmetrizable; i.e. $I - G$ is not similar to an SPD matrix. One consequence of this is that the eigenvalues of the iteration matrix $G$ are not necessarily real and less than one.

We now describe some of the approaches which can be used to solve linear systems where the matrix is not SPD. One approach is to use the ordinary SOR method. One can also use Chebyshev iteration. In each case the formulas are the same as in the symmetrizable case but the optimum parameters are determined in a different way. A number of generalized CG acceleration algorithms are also available; see e.g. Hageman and Young [1981]. As an alternative to generalized CG acceleration algorithms, one can use various Lanczos algorithms. These are also described in Hageman and Young [1981] and in Young *et al* [1988].

### The SOR Method

For the case where $A$ is consistently ordered the relation between the eigenvalues of the SOR matrix $\mathcal{L}_\omega$ and those of the Jacobi matrix $\mathcal{B}$ holds. For separable problems it is often possible to find regions in the complex plane which contain all eigenvalues of $\mathcal{B}$. Programs are available, see e.g. Young and Eidson [1970], Young and Huang [1983] and Huang [1983] for finding the optimum value of $\omega$ when all eigenvalues of $\mathcal{B}$ are known. Usually it is sufficient to have available certain key eigenvalues of $\mathcal{B}$.

For more general problems Ehrlich [1984] and others have developed an "ad hoc SOR" procedure. This procedure involves choosing values of $\omega$ at individual mesh points. The value of $\omega$ used at a given mesh point is that which one would obtain for a representative rectangle if the coefficients of the differential equations were constant throughout the rectangle. (In such a related problem the optimum value of $\omega$ can be obtained analytically). This ad hoc procedure has been found to work well in practice; however, a rigorous analysis of the procedure remains to be made.

## Chebyshev Acceleration

If the eigenvalues of the basic iterative method $G$ are complex but have real parts less than one, we can apply Chebyshev acceleration. The method is defined by (7). However the $\{\gamma_i\}$ and the $\{\rho_i\}$ are given in terms of two real numbers $d$ and $\sigma^2$ which relate to an "optimum ellipse". This ellipse contains all eigenvalues of $G$ and has center at $d$ and foci at $d \pm |\sigma|$ if $\sigma^2$ is positive and at $d \pm |\sigma| i$ if $\sigma^2$ is negative. Here $\sigma^2 = b^2 - a^2$ where $a$ and $b$ are the lengths of the semi-major axis and the semi-minor axis respectively of the ellipse. (See Hageman and Young [1981], Chapter 12, for more details.)

To find $d$ and $\sigma^2$ one needs to first find all eigenvalues of $G$ (or at least certain key eigenvalues) and, second, one needs to find $d$ and $\sigma^2$. Manteuffel [1977], [1978] developed an adaptive program which periodically determines key eigenvalues of $G$ and from these determines $d$ and $\sigma^2$. Another approach to finding the key eigenvalue of $G$ is the hybrid procedure of Elman *et al* [1984]. This procedure involves first carrying out several iterations using the GMRES procedure (see Saad and Schultz [1983]) and then using the coefficients thus obtained to estimate the key eigenvalues. From these key eigenvalues the values of $d$ and $\sigma^2$ are determined.

Huang [1983] developed an alternate procedure for finding the optimum values of $d$ and $\sigma^2$, given all of the eigenvalues of $G$.

## Generalized CG Methods and Lanczos Methods

Both the generalized CG acceleration methods (ORTHODIR, ORTHOMIN, and ORTHORES) and the Lanczos acceleration methods, (including the biconjugate gradient method of Fletcher [1975]) do not require any acceleration parameters. However, for the generalized CG acceleration procedures many of the terms must be discarded; otherwise the machine time and storage needed would be prohibitive. Thus for the generalized CG procedures one must chose a nonnegative integer $s$, and discard all but $s$ terms on each iteration. All of these methods may include splitting parameters. Mai [1986] has proposed a composite adaptive procedure for combining the adaptive determination of the splitting parameters with a generalized CG procedure. Numerical results indicate that the splitting parameter is determined fairly quickly and accurately; however, there is a need to improve the efficiency of the minimization procedure which must be carried out during each iteration.

## 6. References

Adams, L.M., LeVeque, R.J., and Young, D.M. [1986]. "Analysis of the SOR Method for the 9-point Laplacian," ICASE Report No. 86-81, Institute for Computer Applications in Science and Engineering, NASA, Langley Research Center, Hampton, VA (To appear in the *SIAM Jour. of Numerical Analysis*).

Benokraitis, V.J. [1974]. "On the Adaptive Acceleration of Symmetric Successive Overrelaxation," Doctoral Thesis, University of Texas, Austin, Texas.

Bilodeau, G.G., Cadwell, W.R., Dorsey, J.P., Fairey, J.M., and Varga, R.S. [1957]. "PDQ–An ABM-704 Code to Solve the Two-dimensional Few-group Neutron-diffu-sion Equations," Report WAPD-TM-70, Bettis Atomic Power Laboratory, Westinghouse Electric Corp., Pittsburg, Pennsylvania.

Concus, P., Golub, G.H., and O'Leary, D.P. [1976]. "A Generalized Conjugate Gradient Method for the Numerical Solution of Elliptic Partial Differential Equations," in *Sparse Computation* (Bunch and Rose eds), *309–332.* Academic Press, New York. Carré, B.A. [1961]. "The Determination of the Optimum Acceleration Factor for Successive Over-relaxation," *Comput. J., 4, 73–78.*

Diamond, M.A. [1971]. "An Economical Algorithm for the Solution of Finite Difference Equations," Doctoral thesis, Univ. of Illinois, Urbana, Ill.

Dupont, T., Kendall, R., and Rachford, H.H., Jr. [1968]. "An Approximate Factorization Procedure for Solving Self-adjoint Elliptic Difference Equations," *SIAM J. Numer. Anal., 5, 559–573.*

Ehrlich, L.W. [1964]. "The Block Symmetric Successive Overrelaxation Method," *J. Soc. Indust. Appl. Math., 12, 807–826.*

Ehrlich, L.W. [1984], "The Ad Hoc SOR Method: a Local Relaxation Scheme," *Elliptic Problem Solvers II,* edited by Garrett Birkhoff and Arthur Schoenstadt, Academic Press, Inc., Orlando, Florida.

Elman, H.C., Saad, Y., and Saylor, P.E. [1984]. "A Hybrid Chebyshev Krylov Subspace Algorithm for Solving Nonsymmetric Systems of Linear Equations," Report TR-301, Yale University, New Haven, CT.

Evans, D.J. and Forrington, C.V.D. [1963]. "An Iterative Process for Optimizing Successive Overrelaxation," *Comput. J., 6, 271–273.*

Fletcher, R. [1975]. "Conjugate Gradient Methods for Indefinite Systems," *Proc. of the Dundee Biennial Conference on Numerical Analysis,* (G.A. Watson, ed.) *73–89*, Springer-Verlag, Berlin and New York.

Garabedian, P.R. [1956]. "Estimation of the Relaxation Factor for Small Mesh Size," *Math. Tables Aids Comput., 10, 183–185; 228.*

Golub, G.H. and Kent, M.D. [1987]. "Estimates of Eigenvalues for Iterative Methods," Report NA-87-02, Numerical Analysis Project, Computer Science Department, Stanford University, Stanford, California.

Golub, G.H. and Varga, R.S. [1961]. "Chebyshev Semi-iterative methods, Successive Over-relaxation Iterative Methods, and Second- order Richardson Iterative Methods," *Numer. Math.,* Parts I and II 3, *147–168.*

Grimes, R.G., Kincaid, D.R., and Young, D.M. [1979]. "ITPACK 2.0 User's Guide," Report CNA-150, Center for Numerical Analysis, The University of Texas, Austin, Texas.

Gustafsson, I. [1979]. "Stability and Rate of Convergence of Modified Incomplete Cholesky Factorization Methods," Report 79.02R (Ph.D. Thesis), Department of Computer Sciences, Chalmers University of Technology, Goteborg, Sweden.

Habetler, G.J. and Wachspress, E.L. [1961]. "Symmetric Successive Overrelaxation in Solving Diffusion Difference Equations," *Math. Comp. 15, 356–362.*

Hageman, L.A. and Kellogg, R.B. [1968]. "Estimating Optimum Overrelaxation Parame-

ters," *Math. Comp., 22, 60-68.*

Hageman, L.A. and Young, David M. [1981]. *Applied Iterative Methods*, Academic Press, Inc., New York, N.Y.

Hayes, Linda and Young, David M. [1977]. "The Accelerated SSOR Method for Solving Large Linear Systems," Report CNA-123, Center for Numerical Analysis, The University of Texas, Austin, Texas.

Huang, Chung-Yuan [1987]. "Optimization of Explicit Time - Stepping Algorithms and Stream-Function-Coordinate (SFC) Concept for Fluid Dynamic Problems," Ph.D. Thesis, The University of Texas, Austin, Texas.

Huang, R. [1983]. "On the Determination of Iterative Parameters for Complex SOR and Chebyshev Methods," Report CNA-187, Center for Numerical Analysis, The University of Texas, Austin, Texas. Also, M.A. Thesis.

Jea, Kang C. and Young, David M. [1988]. "On the Effectiveness of Adaptive Chebyshev Acceleration for Solving Systems of Linear Equations" Report CNA-218, Center for Numerical Analysis The University of Texas, Austin, Texas. To appear in a Special Issue of the *J. Comp. Appl. Math.*.

Kahan, W. [1958]. "Gauss-Seidel Methods of Solving Large Systems of Linear Equations," Doctoral thesis, Univ. of Toronto, Canada.

Kincaid, D.R., Respess, J.R., Grimes, R.G. and Young, D.M. [1982]. "ITPACK 2C: A FORTRAN Package for Solving Large Sparse Linear Systems by Adaptive Accelerated Iterative Methods," *ACM Transactions on Mathematical Software, 8, 302-322.*

Kulsrud, H.E. [1961]. "A Practical Technique for the Determination of the Optimum Relaxation Factor of the Successive Over-relaxation Method," *Comm. Assoc. Comput. Mach., 4, 184-187.*

Mai, Tsun-zee [1986]. "Adaptive Iterative Algorithms for Large Sparse Linear Systems," Report CNA-203, Center for Numerical Analysis, The University of Texas, Austin, Texas; also, Ph.D. Dissertation.

Mai, Tsun-zee and Young, David M. [1988]. "A Dual Adaptive Procedure for the Automatic Determination of Iteration Parameters for Chebyshev Acceleration" (to appear as a report of the Center for Numerical Analysis, The University of Texas, Austin, Texas).

Manteuffel, T.A. [1977]. "The Tchebyshev Iteration for Nonsymmetric Linear systems," *Numer. Math. 28, 307-327.*

Manteuffel, T.A. [1978]. "Adaptive Procedure for Estimating Parameters for the Nonsymmetric Tchebyshev Iteration," *Numer. Math. 31, 183-208.*

Manteuffel, T.A. [1980]. "An Incomplete Cholesky Factorization Technique for Positive Definite Linear Systems," *Math. Comp., 34, 473-497.*

Meijerink, J.A. and Van der Vorst, H.A. [1977]. "An Iterative Solution Method for Linear Systems of Which the Coefficient Matrix is a Symmetric M-Matrix," *Math. Comp., 31, 148-162.*

Peaceman, D.W. and Rachford, H.H. [1955]. "The Numerical Solution of Parabolic and Elliptic Differential Equations," *J. Soc. Indus. Appl. Math. 3, 28-41.*

Phillips, H.B. and Wiener, N. [1923]. "Nets and the Dirichlet Problem," *J. Math. and*

*Phys., 58, 105–124.*

Reid, J.K. [1966]. "A Method for Finding the Optimum Successive Over-relaxation Parameter," *Comput. J., 9, 200–204.*

Rigler, A.K. [1965]. "Estimation of the Successive Over-relaxation Factor," *Math. Comp., 19, 302–307.*

Saad, Y and Schultz, M.H. [1983]. "GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems," Technical Report 254, Department of Computer Science, Yale University, New Haven, Conn.

Stone, H.L. [1968], "Iterative Solutions of Implicit Approximations of Multi-Dimen-sional Partial Differential Equations," *SIAM J. Numer. Analy., 5, 530–538.*

Van de Vooren, A.I. and Vliegenthart, A.C. [1967]. "The 9-point Difference Formula for Laplace's Equation," *J. Eng. Math. 1, 187–202.*

Varga, R.S. [1957]. "A Comparison of the Successive Overrelaxation Method and Semi-Iterative methods Using Chebyshev Polynomials," *J. Soc. Indus. Appl. Math. 5, 39–46.*

Varga, R.S. [1960]. "Factorization and Normalized Iterative Methods," *Boundary Problems in Differential Equations* (R.E. Langer, ed.), *121–142,* University of Wisconsin Press, Madison, WI.

Varga, R.S. [1962], *Matrix Iterative Analysis,* Prentice-Hall, Englewood, Cliffs, New Jersey.

Wachspress, E.L. [1966]. *Iterative Solution of Elliptic Systems and Application to the Neutron Diffusion Equations of Reactor Physics,* Prentice-Hall, Englewood Cliffs, New Jersey.

Young, David M. [1955]. "Ordvac Solutions of the Dirichlet Problem," *J. Assoc. Comput. Math., 2, 137–161.*

Young, David M. [1971]. *Iterative Solution of Large Linear Systems,* Academic Press, New York, N.Y.

Young, David M. [1972]. "Second-degree Iterative Methods for the Solution of Large Linear Systems," *J. of Approx. Theory, 5, 137–148.*

Young, David M. [1977]. "On the Accelerated SSOR Method for Solving Large Linear Systems," *Adv. in Mathematics, 23, 215–271.*

Young, David M. and Eidson, H.D. [1970]. "On the Determination of the Optimum Relaxation Factor for the SOR Method When the Eigenvalues of the Jacobi Method are Complex," Report CNA-1, Center for Numerical Analysis , The University of Texas, Austin, Texas.

Young, David M. and Huang, R. [1983]. "Some Notes on Complex Successive Overrelaxation," Report CNA-185, Center for Numerical Analysis, The University of Texas, Austin, Texas.

Young, David M., Jea, K. C., and Mai, T. Z. [1988]. "Preconditioned Conjugate Gradient Algorithms and Software for Solving Large Sparse Linear Systems," in *Linear Algebra in Signals, Systems, and Control,* (Datta, Johnson, Kaashoek, Plemmons, and Sontag eds), *260–283.* SIAM, Philadelphia, PA.