27-28-81
Sub.
240115
SAND80-2371
Unlimited Release

UC-32

OPTIMAL PARAMETERS FOR LINEAR SECOND-DEGREE
STATIONARY ITERATIVE METHODS

MASTER

Thomas A. Manteuffel

Prepared by Sandia Laboratories, Albuquerque, New Mexico 87185
and Livermore, California 94550 for the United States Department of
Energy under Contract DE-AC04-76DP00789.

Printed November 1980

Sandia National Laboratories

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Issued by Sandia Laboratories, operated for the United States
Department of Energy by Sandia Laboratories.

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

PAGES 1 to 2
WERE INTENTIONALLY
LEFT BLANK

UC-32
SAND80-2371
Unlimited Release

Optimal Parameters for Linear Second-Degree
Stationary Iterative Methods

T. A. Manteuffel
Sandia National Laboratories - 5642
Albuquerque, New Mexico 87185

Abstract

In this paper we show that the optimal parameters for linear second-degree stationary iterative methods applied to nonsymmetric linear systems can be found by solving the same minimax problem used to find optimal parameters for the Tchebychev iteration. In fact, the Tchebychev iteration is asymptotically equivalent to a linear second-degree stationary method. The method of finding optimal parameters for the Tchebychev iteration given in Manteuffel [5] can be used to find optimal parameters for the stationary method as well.

Printed in the United States of America

Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, Virginia 22161
Price: \$4.50 , Microfiche \$3.00

DISCLAIMER

This book was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

1. Introduction

Given the possibly nonsymmetric linear system

$$(1.1) \quad A \underline{x} = \underline{b},$$

linear second-degree stationary iterative methods produce a sequence of approximate solutions $\{\underline{x}_i\}$ by a formula of the type

$$(1.2) \quad \underline{x}_{n+1} = \underline{x}_n + \alpha \underline{r}_n + \beta (\underline{x}_n - \underline{x}_{n-1}),$$

where

$$(1.3) \quad \underline{r}_n = \underline{b} - A \underline{x}_n$$

is the residual vector and $\underline{x}_0, \underline{x}_1$ are arbitrary (cf. Young [11], p. 487). In this paper we will show that the parameters α, β can be expressed as functions of the parameters d, c where d represents the center of a family of ellipses with foci $d \pm c$. The optimal choice of d, c results in the same minimax problem whose solution is given in Manteuffel [5] in terms of the convex hull of the spectrum of A . There, the parameters d, c define the nonstationary Tchebychev iteration:

$$(1.4) \quad \underline{x}_{n+1} = \underline{x}_n + \alpha_n \underline{r}_n + \beta_n (\underline{x}_n - \underline{x}_{n-1}),$$

where

$$(1.5) \quad \begin{aligned} \alpha_0 &= 1/d, \beta_0 = 0, \\ \alpha_1 &= \frac{2d}{2d^2 - c^2}, \beta_1 = d\alpha_0 - 1, \\ \alpha_n &= [d - (c/2)^2 \alpha_{n-1}]^{-1}, \beta_n = d\alpha_n - 1. \end{aligned}$$

We will show that a choice of d, c results in the same asymptotic convergence factor for either (1.2) or (1.4) and, in fact, α, β are the asymptotic values of α_n, β_n .

The above results are well known in the case in which the eigenvalues of A lie on the positive real axis (Golub and Varga [2]). Here we extend those results to general nonsymmetric A . Unless otherwise indicated, the following applies to A with complex elements. It will be shown that in order for convergence to be possible for either (1.2) or (1.4), the spectrum of A must lie in some open half-plane in the complex plane.

Work by Neithammer [7], [8], Faddeev and Faddeeva [1], and Kincaid [4] can be used to find optimal parameters for (1.2) when the spectrum of A lies in an ellipse that does not include the origin. Of course, the spectrum of A may be enclosed by many such ellipses. In Manteuffel [5] an algorithm is presented that finds the best iteration parameters for (1.4) by finding the best ellipse enclosing the spectrum of A . The purpose of this paper is to show that because of the asymptotic relationship between (1.2) and (1.4) mentioned above the algorithm in [5] will find the best parameters for (1.2) as well as (1.4).

This treatment includes as a special case the situation in which A can be expressed as $A = I - B$ where B is two-cyclic. The spectrum of B must lie in the infinite vertical strip $\{z: -1 < \operatorname{Re}(z) < 1\}$. In this case the optimal value of d is known to be unity. The choice of parameters becomes a one parameter problem. Several authors have addressed this problem for either (1.2) or (1.4) (cf. Hageman [3], di Pillis [9]). If B is not two-cyclic, however, forcing $d=1$ may yield slower convergence.

Finally, the results of this work can easily be applied to preconditioned systems. If C is a preconditioning matrix, then A may be replaced by $C^{-1}A$ throughout this paper.

2. Linear Second-Degree Iterative Methods

Let $\underline{e}_n = \underline{x} - \underline{x}_n$ be the error at step n . We can rewrite (1.2) as

$$(2.1) \quad \underline{e}_{n+1} = \underline{e}_n - \alpha A \underline{e}_n + \theta (\underline{e}_n - \underline{e}_{n-1}) .$$

If the dimension of A is N , then we can form a system of dimension $2N$:

$$(2.2) \quad \begin{pmatrix} \underline{e}_n \\ \underline{e}_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & I \\ -\theta I & (1+\theta)I - \alpha A \end{pmatrix} \begin{pmatrix} \underline{e}_{n-1} \\ \underline{e}_n \end{pmatrix} ,$$

which we will denote

$$(2.4) \quad \hat{\underline{e}}_{n+1} = \hat{G} \hat{\underline{e}}_n .$$

The eigenvalues, μ , of $\hat{G}$ can be found by taking determinants. We have

$$(2.5) \quad \delta(\mu I - \hat{G}) = \delta(\mu^2 I - ((1+\theta)I - \alpha A) \mu + \theta I) = 0 .$$

If λ is an eigenvalue of A , we have the relation

$$(2.6) \quad \mu^2 - ((1+\theta) - \alpha\lambda) \mu + \theta = 0 .$$

With the judicious change of parameters,

$$(2.7) \quad \alpha = \frac{1}{d} (\theta + 1) ,$$

this becomes

$$(2.8) \quad \mu^2 - (1+\theta)(1 - \frac{\lambda}{d}) \mu + \theta = 0 .$$

Solving for λ yields

$$(2.9) \quad \lambda = d(1 - \frac{1}{1+\theta} (\mu + \frac{\theta}{\mu})) .$$

Letting $\mu = \rho e^{i\theta}$, $\rho > 0$, $\theta = |\theta| e^{i\gamma}$, we have

$$(2.10) \quad \begin{aligned} \lambda &= d(1 - \frac{1}{1+\theta} (\rho e^{i\theta} + \frac{\theta}{\rho} e^{-i\theta})) \\ &= d(1 - \frac{e^{i\gamma/2}}{1+\theta} ((\rho + \frac{|\theta|}{\rho}) \sin(\theta - \gamma/2) + i(\rho - \frac{|\theta|}{\rho}) \cos(\theta - \gamma/2))) . \end{aligned}$$

Thus, circles centered at the origin in the μ -plane correspond to ellipses in the λ -plane centered at d with foci at $d \pm c$ where

$$(2.11) \quad c^2 = \frac{4\theta d^2}{(1+\theta)^2} .$$

Note that $\rho^2 = |\delta|$ yields the degenerate ellipse, the straight line connecting the foci.

The parameters d, c describe a family of confocal ellipses, $\mathcal{F}(d, c)$, in the λ -plane. Each member of this family, $F(d, c)$, corresponds to two circles in the μ -plane. One has radius ρ and the other has radius $\rho' = \frac{|\delta|}{\rho}$. (If $\delta = 0$, the iteration reduces to $x_{n+1} = x_n + \alpha r_n$, Richardson's method.) Let ρ be such that $\rho \geq \sqrt{|\delta|} \geq \rho'$. If $F_1(d, c)$, $F_2(d, c) \in \mathcal{F}(d, c)$ are such that the ellipse $F_2(d, c)$ is inside $F_1(d, c)$, the corresponding circles have radii ρ_1 and ρ_2 such that

$$(2.12) \quad \rho_1 > \rho_2 \geq \sqrt{|\delta|} \geq \rho_2' > \rho_1'.$$

(See Figure 2.1.) The degenerate ellipse is mapped onto the circle of radius $\sqrt{|\delta|}$.

Figure 2.1

Let $F_0(d, c) \in \mathcal{F}(d, c)$ be the ellipse passing through the origin in the λ -plane. We see from (2.8) that the corresponding values of μ are $\mu = 1$ and $\mu = \beta$. Thus, if the iteration is to converge, the eigenvalues of A must lie in the interior of $F_0(d, c)$. Clearly, there exists some d, c such that the spectral radius of $\hat{G}$ is less than one if and only if the spectrum of A lies in some open half complex plane. Further, if λ_1 is an eigenvalue of A , then (2.12) shows that $\hat{G}$ has an eigenvalue μ_1 such that $|\mu_1| \geq \sqrt{|\beta|}$. We must have $|\beta| < 1$ for convergence to be possible.

Now let us solve (2.8) for μ . We have

$$\begin{aligned} \mu &= \frac{(1+\beta)(1 - \frac{\lambda}{d}) \pm ((1+\beta)^2(1 - \frac{\lambda}{d})^2 - 4\beta)^{\frac{1}{2}}}{2} \\ &= \frac{(d - \lambda) \pm \left((d - \lambda)^2 - \frac{4d^2\beta}{(1+\beta)^2} \right)^{\frac{1}{2}}}{\frac{2d}{1+\beta}}. \end{aligned}$$

Using (2.11) we note that if $|\beta| < 1$, then

$$d + (d^2 - c^2)^{\frac{1}{2}} = d + (d^2 - \frac{4d^2\beta}{(1+\beta)^2})^{\frac{1}{2}} = d + \frac{d(1-\beta)}{1+\beta} = \frac{2d}{1+\beta}.$$

Thus, we may write

$$(2.13) \quad \mu = \frac{(d-\lambda) \pm ((d-\lambda)^2 - c^2)^{\frac{1}{2}}}{d + (d^2 - c^2)^{\frac{1}{2}}}.$$

To minimize the spectral radius of $\hat{G}$, and thus the asymptotic convergence factor associated with iteration (1.2), we must choose d, c to satisfy

$$(2.14) \quad \min_{d,c} \max_{\lambda_i} \left| \frac{(d-\lambda_i) \pm ((d-\lambda_i)^2 - c^2)^{\frac{1}{2}}}{d + (d^2 - c^2)^{\frac{1}{2}}} \right|,$$

where the max is taken over all eigenvalues of A .

This is the same minimax problem described in Manteuffel [5]. There it is shown that the solution can be found in terms of the eigenvalues appearing as vertices of the convex hull of the spectrum of A . For A real valued, the solution is found by comparing a finite number of ellipses whose parameters are either given explicitly or in terms of a root of a certain 5th degree polynomial whose coefficients depend upon the spectrum of A . Further, if A is real valued, then $\mathcal{F}(d,c)$ is symmetric with respect to the real axis; that is, $d > 0$ ($d < 0$), $c^2 < d^2$.

The following lemma gives a visual image of the relation between the family of ellipses, $\mathcal{F}(d,c)$, and the spectrum of $\hat{G}$.

Lemma Let $d > 0$, $c^2 < d^2$. Let λ_i be an eigenvalue of A and let $F_i(d,c)$ be the member of $\mathcal{F}(d,c)$ passing through λ_i , and let $F_0(d,c)$ be the member of $\mathcal{F}(d,c)$ passing through $\lambda = 0$. Let M_i, m_i be the semimajor and semiminor axes of $F_i(d,c)$, and let M_0, m_0 be the semimajor and semiminor axes of $F_0(d,c)$. If μ_i, μ_i' are the corresponding eigenvalues of $\hat{G}$ (see 2.13), then

$$|\mu_i| = \frac{M_i + m_i}{M_0 + m_0},$$

$$|\mu_i'| = \frac{M_i - m_i}{M_0 + m_0}.$$

Proof Since all members of the ellipse $F_i(d, c)$ are mapped onto the same two circles in the μ -plane, let us choose a convenient λ_0 , namely the intersection of $F_i(d, c)$ with the real axis closest to the origin. We have

$$|\mu_i| = \frac{(d - \lambda_0) + ((d - \lambda_0)^2 - c^2)^{\frac{1}{2}}}{d + (d^2 - c^2)^{\frac{1}{2}}},$$

$$|\mu_i'| = \frac{|(d - \lambda_0) - ((d - \lambda_0)^2 - c^2)^{\frac{1}{2}}|}{d + (d^2 - c^2)^{\frac{1}{2}}}.$$

Notice that $(d - \lambda_0)$ is the real semiaxis of $F_i(d, c)$ and d is the real semiaxis of $F_0(d, c)$. The result follows from the geometry of ellipses. ■

A similar result holds for arbitrary d, c .

Note that if the spectrum of A lies on the positive real axis, then it can be exactly covered by a degenerate ellipse; that is, if $a \leq \lambda \leq b$, then $d = \frac{b+a}{2}$, $c = \frac{b-a}{2}$. This is, in fact, the optimal solution. The geometrical picture provided by the above lemma yields an easy proof.

Having chosen d, c we may use (2.7) and (2.11) to rewrite (1.2) as

$$(2.15) \quad x_{n+1} = x_n + \alpha r_n + \beta(x_n - x_{n-1}),$$

where

$$(2.16) \quad \alpha = \frac{2}{d \pm (d^2 - c^2)^{\frac{1}{2}}}, \quad \beta = d\alpha - 1,$$

and the sign in α is chosen so that $|\beta| < 1$. If $d > 0$, $c^2 < d^2$, this yields

$$(2.17) \quad \alpha = \frac{2}{d + (d^2 - c^2)^{\frac{1}{2}}}, \quad \beta = d\alpha - 1.$$

3. Asymptotic Form of the Tchebychev Iteration

Again let $\underline{e}_n = \underline{x} - \underline{x}_n$ be the error vector. The iteration (1.4) yields the relation

$$(3.1) \quad \underline{e}_n = P_n(A) \underline{e}_0$$

where $P_n(\lambda)$ is the n th scaled and translated Tchebychev polynomial (cf. Manteuffel [5], p. 316); that is,

$$(3.2) \quad P_n(\lambda) = \frac{T_n\left(\frac{d-\lambda}{c}\right)}{T_n\left(\frac{d}{c}\right)},$$

where

$$(3.3) \quad T_n(x) = \cosh(n \cosh^{-1}(x)).$$

One can show that for $\frac{d-\lambda}{c} \notin [-1,1]$, $\frac{d}{c} \notin [-1,1]$, we have

$$(3.4) \quad P_n(\lambda) \doteq (S(\lambda))^n = \left[\frac{(d-\lambda) + ((d-\lambda)^2 - c^2)^{\frac{1}{2}}}{d + (d^2 - c^2)^{\frac{1}{2}}} \right]^n$$

for large n (Manteuffel [6], p. 186). Let $S(A)$ be the operator equivalent of $S(\lambda)$ in (3.4). The eigenvalues, σ_i , of $S(A)$ are related to the eigenvalues λ_i of A by the expression

$$(3.5) \quad \sigma_1 = \frac{(d-\lambda_1) + ((d-\lambda_1)^2 - c^2)^{\frac{1}{2}}}{d + (d^2 - c^2)^{\frac{1}{2}}}.$$

Notice that (3.5) corresponds to the branch of (2.13) having larger modulus. (Care must be used in choosing the branch of the square root.)

In light of (3.1) and (3.4) we would like to choose d, c so that the spectral radius of $S(A)$ is as small as possible. This yields the minimax problem (2.14). Given d, c the spectral radius of $S(A)$ will equal the spectral radius of $\hat{G}$. In fact, the spectrum of $S(A)$ corresponds to the half of the spectrum of $\hat{G}$ having modulus greater than or equal to $\sqrt{|B|}$.

Finally, notice that if we let $\gamma = (\frac{c}{2d})^2$, then from (1.5) we can write

$$(3.6) \quad d\alpha_n = \frac{1}{1 - \gamma d\alpha_{n-1}}.$$

Thus, $d\alpha_n$ can be written as the continued fraction

$$(3.7) \quad d\alpha_n = \frac{1}{1 - \frac{\gamma}{1 - \frac{\gamma}{1 - \dots}}}.$$

If $\gamma \notin (1/4, \infty)$, that is, if the degenerate ellipse connecting $d+c$ and $d-c$ does not include the origin, then α_n converges to

$$(3.8) \quad \alpha = \frac{2}{d \pm (d^2 - c^2)^{\frac{1}{2}}}$$

where the sign is chosen to minimize $|\alpha|$ (cf. Wall [10], p. 39). Notice that β_n must converge to

$$(3.9) \quad \beta = d\alpha - 1 = \begin{cases} \frac{d - (d^2 - c^2)^{\frac{1}{2}}}{d + (d^2 - c^2)^{\frac{1}{2}}} \\ \frac{d + (d^2 - c^2)^{\frac{1}{2}}}{d - (d^2 - c^2)^{\frac{1}{2}}} \end{cases}.$$

The sign choice for α insures that $|\beta| < 1$. Comparing (3.8) and (3.9) to (2.16) we see that (1.4) is asymptotically equivalent to (1.2).

- [1] Faddeev, D. K. and Faddeeva, V. N., Computational Methods of Linear Algebra, W. H. Freedman, San Francisco, 1963.
- [2] Golub, G. H. and Varga, R. S., "Chebyshev Semi-iterative Methods, Successive Over Relaxation Iterative Methods, and Second Order Richardson Iterative Methods," Numer. Math., 3 (1961), p. 147.
- [3] Hageman, L. A., "The Estimation of Acceleration Parameters for the Chebyshev Polynomial and The Successive Over Relaxation Iteration Methods," AEC Research and Development Report WAPD TM-1038, June, 1972.
- [4] Kincaid, D. R., "On Complex Second-Degree Iterative Methods," SIAM J. Numer. Anal., Vol. II, No. 2 (1974), 211-218.
- [5] Manteuffel, T. A. "The Tchebychev Iterations for Nonsymmetric Linear Systems," Numer. Math., 28 (1977), 307-327.
- [6] Manteuffel, T. A. "Adaptive Procedure for Estimations Parameters for the Nonsymmetric Tchebychev Iteration" Numer. Math., 31 (1978), 183-208.
- [7] Neithammer, W. "Iterationverfahren and allgemeine Euler-verfahren," Math. Zeitschr., 102 (1967), 288-317.
- [8] Neithammer, W., "Konvergenzbeschleunigung bei einstufigen iterationsverfahren durch summierungsmethoden," Internat. Ser. on Numer. Math., 15 (1970), 235-240.
- [9] di Pillis, J., "How to Embrace your Spectrum for Faster Iterative Results," to appear.
- [10] Wall, H. S., Analytic Theory of Continued Fractions, D. Van Nostrand, New York, 1948.
- [11] Young, D., Iterative Solution of Large Linear Systems, Academic Press, New York and London, 1971.

Distribution:

Richard Bartels
Department of Mathematical Sciences
Johns Hopkins University
Baltimore, MD 21218

Jim Bunch
Dept. of Computer Science
University of California, San Diego
P.O. Box 109
La Jolla, CA 92037

B. L. Buzbee
C-Division
Los Alamos Scientific Laboratory
Box 1663
Los Alamos, NM 87544

Tony Chan
Computer Science Department
Yale University
Box 2158 Yale Station
New Haven, CT 06520

Ray Chin
Mail Stop L-71
Lawrence Livermore National Laboratory
P.O. Box 808
Livermore, CA 94550

Paul Concus
University of California
Lawrence Berkeley Laboratory
5DA-2129
Berkeley, CA 94720

Jack Dongarra
Applied Mathematics
Argonne National Labs
9700 S. Cass Avenue
Argonne, IL 60439

Iain S. Duff
Harwell Computer Science Division
Building 8.9
AERE Harwell, Oxfordshire
England

Louis W. Ehrlich
The Johns Hopkins University
Applied Physics Laboratory
Johns Hopkins Road
Laurel, MD 20810

Stan Eisenstat
Department of Computer Science
Yale University
New Haven, CT 06520

Gene H. Golub
Department of Computer Science
Stanford University
Stanford, CA 94305

Louis Hageman
Advisory Mathematician
Bettis Atomic Power Laboratory
Box 79
West Mifflin, PA 15122

I. Karasallo
Lawrence Berkeley Lab
Room 2113 Bldg. 50A
Berkeley, CA 94720

Davis R. Kincaid
Center for Numerical Analysis
University of Texas at Austin
Austin, TX 78712

John Lewis
Boeing Computing Center
565 Andover Park West
Mail Stop 9C01
Tukwilla, WA 98188

Cleve B. Moler
Department of Mathematics
University of New Mexico
Albuquerque, NM 87106

J. diPilllis
Dept. of Math
University of California, Riverside
Riverside, CA 92521

John K. Reid
Computer Science System Division
Building 8.9 AERE
Harwell, Didcot
England OX11 0RA

Garry H. Rodrigue
M.S. 74
Lawrence Livermore National Laboratory
P.O. Box 808
Livermore, CA 94550

Paul Saylor
Department of Computer Science
University of Illinois
Urbana, IL 61820

Martin Schultz
Department of Computer Science
Yale University
New Haven, CT 06520

Andy Sherman
Center for Numerical Analysis
Computer Science Dept.
University of Texas
Austin, TX 78712

Richard S. Varga
Department of Mathematics
Kent State University
Kent, OH 44242

H. van der Vorst
Academic Computer Center Utrecht
Budapestlaan 6
De Uithof-Utrecht
The Netherlands
030 531436

David Young
Center for Numerical Analysis
University of Texas
Austin, TX 78712

2620 R. J. Detry
2646 M. R. Scott
2646 K. Haskell
2646 H. Watts
3141 T. L. Werner (5)
3151 W. L. Garner (3)
For DOE/TIC (Unlim. Release)
3154-3 R. P. Campbell (25) (For DOE/TIC)
5600 D. B. Shuster
Attn: 5610 A. A. Lieber
5620 M. M. Newsom
5630 R. C. Maydew
5640 G. J. Simmons
5641 R. J. Thompson
5642 L. F. Snampine
5642 R. J. Hanson
5642 K. L. Hiebert
5642 T. A. Manteuffel (25)
8000 T. B. Cook
Attn: 8100 W. J. Spencer
8200 A. N. Blackwell
8300 B. F. Murphey
8400 L. Gutierrez
8330 G. W. Anderson
8331 R. J. Kee
8331 M. S. Smooke
8331 L. R. Petzold
8332 R. E. Huddleston
8332 R. E. Chang
8332 D. L. Crawford
8332 T. H. Jefferson