DOE/ER/25048-40

**Linear Stationary Second-Degree Methods
for the Solution of Large Linear Systems**

BY

David M. Young
David R. Kincaid

July 1990

CNA-244

RECEIVED

NOV 20 1998

OSTI

CENTER FOR NUMERICAL ANALYSIS

THE UNIVERSITY OF TEXAS AT AUSTIN

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

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.

Linear Stationary Second-Degree Methods for the Solution of Large Linear Systems *

David M. Young
David R. Kincaid

July 9, 1990

Abstract

The optimum linear stationary second-degree iterative method for solving linear systems of equations is not as good in general as the optimum semi-iterative method. However, for a suitable choice of parameters, the rate of convergence of the stationary method is very nearly as good as that of the semi-iterative method. We present a straightforward determination of these optimum parameter values and the asymptotic rate of convergence.

1 Introduction

In this paper, we consider a class of linear stationary second-degree methods for solving the linear system

$$Au = b, \quad (1)$$

where A is a given real nonsingular $N \times N$ matrix and b is a given vector or $N \times 1$ (column) matrix. The second-degree methods that we shall consider are related to a *basic* linear stationary method of first degree of the form

$$u^{(n+1)} = Gu^{(n)} + k, \quad \text{depends on } u^{(n)} \quad (2)$$

---

*This work was supported, in part, by the National Science Foundation under Grant ASC-8917592, by the Department of Energy under Grant DE-FG05-87ER25048, and by Cray Research, Inc., under Grant LTR DTD with The University of Texas at Austin.

where G is a real $N \times N$ matrix and k is a real vector such that $I - G$ is nonsingular and $k = (I - G)A^{-1}b$. We assume that all eigenvalues μ of G are real and lie in the interval $\alpha \leq \mu \leq \beta < 1$, for some constants α and β . Evidently, the basic method (2) is *completely consistent* with (1) in the sense of Young [12] [13], since the unique solution $\bar{u} = A^{-1}b$ of system (1) is also the unique solution of the *related equation*

$$u = Gu + k, \quad (3)$$

corresponding to method (2). We do *not* require that the basic method be convergent in the sense that the sequence $u^{(1)}, u^{(2)}, \dots$ must converge to $\bar{u}$, given any starting vector $u^{(0)}$. Recall that the **spectral radius** of G is defined to be

$$S(G) = \max_{\mu \in \Lambda(G)} |\mu|,$$

where the maximum is taken over the set $\Lambda(G)$ of all the eigenvalues of G . We remark that a necessary and sufficient condition for such convergence is that $S(G) < 1$.

We consider a linear stationary second-degree method of the form

$$\begin{cases} u^{(1)} = \rho^* Gu^{(0)} + \delta^* u^{(0)} + v^*, \\ u^{(n+1)} = \rho Gu^{(n)} + \delta u^{(n)} + \tau u^{(n-1)} + v, \end{cases} \quad \begin{matrix} \text{depends on } u^{(0)}, u^{(n-1)} \\ (4) \end{matrix}$$

where $\rho^*, \delta^*, \rho, \delta$, and τ are constants, and where v^* and v are vectors. We require that the constants $\rho^*, \delta^*, \rho, \delta$, and τ be chosen so that v^* and v can be *conveniently* computed in terms of k . We also require that if $u^{(0)} = \bar{u}$, then $u^{(1)} = u^{(2)} = \dots = u^{(n)} = \dots = \bar{u}$. This condition corresponds to the *consistency* condition for linear stationary methods of first degree (see Young [12], Chapters 3, 9, and 16, and Young [13]). Since $\bar{u}$ satisfies (3), we have

$$v^* = (1 - \rho^* - \delta^*)\bar{u} + \rho^*k.$$

So that v^* can be conveniently computed in terms of k , we require that

$$\rho^* + \delta^* = 1. \quad (5)$$

Similarly, we obtain

$$v = (1 - \rho - \delta - \tau)\bar{u} + \rho k,$$

linear stationary second-degree method

linear: linear function of G

stationary: coefficients are constant

second degree: depends on 2 previous values $u^{(n)}, u^{(n-1)}$

and in order that v can be conveniently computed in terms of k , we require that

$$\rho + \delta + \tau = 1. \quad (6)$$

From (4)-(6), we have

$$\begin{cases} u^{(1)} = \rho^*(\overbrace{Gu^{(0)} + k}^{\tilde{u}^{(1)}}) + (1 - \rho^*)u^{(0)}, \\ u^{(n+1)} = \rho(Gu^{(n)} + k) + \delta u^{(n)} + \tau u^{(n-1)}. \end{cases} \quad \begin{aligned} u^{(1)} &= L(\tilde{u}^{(1)}, u^{(0)}) \\ u^{(n+1)} &= L(\tilde{u}^{(n+1)}, u^{(n)}, u^{(n-1)}) \end{aligned}$$

Now letting $\omega_1 = \rho^*$, $e = \rho$, and $d = -\tau$, we can rewrite this method in the form

$$\begin{cases} u^{(1)} = \omega_1(Gu^{(0)} + k) + (1 - \omega_1)u^{(0)}, \\ u^{(n+1)} = \underbrace{u^{(n)} + e(Gu^{(n)} + k - u^{(n)})}_{\tilde{u}^{(n+1)} - u^{(n)}} + d \underbrace{(u^{(n)} - u^{(n-1)})}_{u^{(n)} - u^{(n-1)}}. \end{cases} \quad (7)$$

Let us define ω , γ , and ω^* by

$$\omega = 1 + d, \quad \gamma = e/(1 + d), \quad \omega^* = \omega_1/\gamma.$$

Evidently, we have

$$d = \omega - 1, \quad e = \omega\gamma.$$

Since $\rho^* = \omega^*\gamma$, $\delta^* = 1 - \omega^*\gamma$, $\rho = \omega\gamma$, $\delta = \omega(1 - \gamma)$, and $\tau = (1 - \omega)$, we obtain

$$\begin{cases} u^{(1)} = \omega^*\{\gamma(Gu^{(0)} + k) + (1 - \gamma)u^{(0)}\} + (1 - \omega^*)u^{(0)}, \\ u^{(n+1)} = \omega\{\gamma(Gu^{(n)} + k) + (1 - \gamma)u^{(n)}\} + (1 - \omega)u^{(n-1)}. \end{cases} \quad (8)$$

We remark that the optimum semi-iterative method based on (2) can be written in the form

$$\begin{cases} u^{(1)} = \bar{\gamma}(Gu^{(0)} + k) + (1 - \bar{\gamma})u^{(0)}, \\ u^{(n+1)} = \omega_{n+1}\{\bar{\gamma}(Gu^{(n)} + k) + (1 - \bar{\gamma})u^{(n)}\} + (1 - \omega_{n+1})u^{(n-1)}, \end{cases} \quad (9)$$

coefficients change

where

$$\begin{cases} \omega_1 = 1, \\ \omega_2 = (1 - 2(\sigma/2))^{-1}, \\ \omega_{n+1} = (1 - (\sigma/2)^2)^{-1}, \end{cases}$$

and

$$\begin{cases} \bar{\gamma} = 2/(2 - (\beta + \alpha)), \\ \bar{\sigma} = (\beta - \alpha)/(2 - (\beta + \alpha)). \end{cases}$$

(See, for instance, Young [14].) Thus, if we replace ω_{n+1} in (9) by ω , we get (8) with $\omega^* = 1$. Clearly, if we let $\gamma = \bar{\gamma}$, then the optimum linear stationary second-degree method (8) will not in general be as good as the optimum semi-iterative method (9). However, as shown by Young [12],[14] (see also Golub and Varga [5]), for suitable choice of ω , the rate of convergence of the stationary method is very nearly as good as that of the semi-iterative method (9).

In this paper, we present a straightforward determination of the optimum value of γ and ω , and we show that, asymptotically, the convergence rate is independent of ω^* .

2 Convergence Analysis

Let us now assume that $d \neq -1$ and that $e \neq 0$. (The case $d = -1$ and/or $e = 0$ will be treated later in Section 4.) To simplify the discussion, we introduce the following notation. For fixed $\gamma > 0$, we let

$$\begin{cases} \hat{G} = \gamma G + (1 - \gamma)I, \\ \hat{\mu} = \gamma \mu + (1 - \gamma), \\ \hat{\alpha} = \gamma \alpha + (1 - \gamma), \\ \hat{\beta} = \gamma \beta + (1 - \gamma). \end{cases}$$

Thus, we can replace (8) by

$$\begin{cases} u^{(1)} = \omega^*(\hat{G}u^{(0)} + \gamma k) + (1 - \omega^*)u^{(0)}, \\ u^{(n+1)} = \omega(\hat{G}u^{(n)} + \gamma k) + (1 - \omega)u^{(n-1)}. \end{cases}$$

Let us define the polynomial $P_0(G), P_1(G), \dots$ by

$$\begin{cases} P_0(G) = I, \\ P_1(G) = \omega^*\hat{G} + (1 - \omega^*)I, \\ P_{n+1}(G) = (\omega\hat{G})P_n(G) + (1 - \omega)P_{n-1}(G). \end{cases} \quad (10)$$

Evidently, if we let $\epsilon^{(n)} = u^{(n)} - \bar{u}$, we have

$$\epsilon^{(n)} = P_n(G)\epsilon^{(0)}.$$

As a measure of convergence, let us define the **virtual spectral radius** of $P_n(G)$ by

$$\bar{S}(P_n(G)) = \max_{\alpha \leq \mu \leq \beta} |P_n(\mu)|,$$

and the **virtual average spectral radius** by

$$\bar{S}_n(P_n(G)) = \sqrt[n]{\bar{S}(P_n(G))}.$$

Finally, we define the **virtual asymptotic average spectral radius** by

$$\bar{S}(P_n(G)) = \limsup_{n \rightarrow \infty} \bar{S}_n(P_n(G)).$$

We seek to show that unless G has only one distinct eigenvalue then, for any choice of ω^* , we have

$$\bar{S}(P_n(G)) = R(\omega, \gamma),$$

where

$$R(\omega, \gamma) = \max_{\mu} \rho(\omega, \gamma, \mu),$$

and where $\rho(\omega, \gamma, \mu)$ is the root radius ¹ of the quadratic equation

$$\lambda^2 - (\omega\hat{\mu})\lambda + (\omega - 1) = 0.$$

We remark that if G has only one distinct eigenvalue, say η , then we can choose γ by

$$\gamma = 1/(1 - \eta).$$

In this case, the spectral radius of the matrix $\hat{G}$, corresponding to the iterative method

$$\begin{aligned} u^{(n+1)} &= \hat{G}u^{(n)} + \gamma k, \\ &= \gamma(Gu^{(n)} + k) + (1 - \gamma)u^{(n)}, \end{aligned}$$

---

¹The root radius of a quadratic equation is the maximum of the moduli of its radius.

vanishes. This method is often known as an *extrapolated method* based on (2). Because the spectral radius of $\hat{G}$ vanishes there is no need to consider acceleration by a second-degree method. If we let

$$Q_n(\hat{G}) = P_n \left(\frac{1}{\gamma} [\hat{G} - (1 - \gamma)I] \right),$$

then (10) becomes

$$\begin{cases} Q_0(\hat{G}) = I, \\ Q_1(\hat{G}) = \omega^* \hat{G} + (1 - \omega^*)I, \\ Q_{n+1}(\hat{G}) = (\omega \hat{G}) Q_n(\hat{G}) + (1 - \omega) Q_{n-1}(\hat{G}). \end{cases}$$

The eigenvalues $\hat{\mu}$ of $\hat{G}$ lie in the range $\hat{\alpha} \leq \hat{\mu} \leq \hat{\beta} < 1$. Moreover,

$$\overline{S}(P_n(G)) = \max_{\hat{\alpha} \leq \hat{\mu} \leq \hat{\beta}} |Q_n(\hat{\mu})|.$$

3 Basic Theorem

Theorem 1. Let $\hat{\alpha}$ and $\hat{\beta}$ be numbers such that $\hat{\alpha} < \hat{\beta} < 1$ and, for given ω and ω^* , let $Q_0(\hat{\mu}), Q_1(\hat{\mu}), \dots$ be the family of polynomials of degree zero, one, etc., such that

$$\begin{cases} Q_0(\hat{\mu}) = 1, \\ Q_1(\hat{\mu}) = \omega^* \hat{\mu} + (1 - \omega^*), \\ Q_{n+1}(\hat{\mu}) = (\omega \hat{\mu}) Q_n(\hat{\mu}) + (1 - \omega) Q_{n-1}(\hat{\mu}). \end{cases}$$

Then, for all ω^* , we have

$$\limsup_{n \rightarrow \infty} \sqrt[n]{\max_{\hat{\alpha} < \hat{\mu} < \hat{\beta}} |Q_n(\hat{\mu})|} = R(\omega),$$

where

$$R(\omega) = \max_{\hat{\alpha} < \hat{\mu} < \hat{\beta}} \rho(\omega, \hat{\mu}), \quad (11)$$

and where $\rho(\omega, \hat{\mu})$ is the root radius of the quadratic equation

$$\lambda^2 - (\omega \hat{\mu})\lambda + (\omega - 1) = 0. \quad (12)$$

Proof. It is easy to verify that if the roots λ_1 and λ_2 of (12) are distinct then

$$Q_n(\hat{\mu}) = \left[\frac{-\lambda_2 + Q_1}{\lambda_1 - \lambda_2} \right] \lambda_1^n + \left[\frac{\lambda_1 - Q_1}{\lambda_1 - \lambda_2} \right] \lambda_2^n, \quad (13)$$

while if $\lambda_1 = \lambda_2$ then

$$Q_n(\hat{\mu}) = \lambda_1^n + [-\lambda_1 + Q_1]n\lambda_1^{n-1}. \quad (14)$$

We now prove

Lemma 1. Under the hypothesis of Theorem 1, we have

$$\limsup_{n \rightarrow \infty} \sqrt[n]{\max_{\hat{\alpha} \leq \hat{\mu} \leq \hat{\beta}} |Q_n(\hat{\mu})|} \leq R(\omega). \quad (15)$$

Proof. If $\lambda_1(\hat{\mu}) \neq \lambda_2(\hat{\mu})$ we have, for $n \geq 2$,

$$Q_n(\hat{\mu}) = Q_1 \left(\frac{\lambda_1^n - \lambda_2^n}{\lambda_1 - \lambda_2} \right) - \lambda_1 \lambda_2 \left(\frac{\lambda_1^{n-1} - \lambda_2^{n-1}}{\lambda_1 - \lambda_2} \right), \quad (16)$$

$$\begin{aligned} |Q_n(\hat{\mu})| &\leq |Q_1| |\lambda_1^{n-1} + \lambda_1^{n-2} \lambda_2 + \cdots + \lambda_2^{n-1}| \\ &\quad + |\lambda_1 \lambda_2| |\lambda_1^{n-2} + \lambda_1^{n-3} \lambda_2 + \cdots + \lambda_2^{n-2}| \\ &\leq |Q_1| n \rho(\omega, \hat{\mu})^{n-1} + |\omega - 1| (n-1) \rho(\omega, \hat{\mu})^{n-2} \\ &\leq \left\{ |Q_1| \rho(\omega, \hat{\mu}) + \frac{n-1}{n} |\omega - 1| \right\} n \rho(\omega, \hat{\mu})^{n-2} \\ &\leq (\gamma \rho(\omega, \hat{\mu}) + |\omega - 1|) n \rho(\omega, \hat{\mu})^{n-2} \\ &\leq (\gamma R(\omega) + |\omega - 1|) n R(\omega)^{n-2}, \end{aligned} \quad (17)$$

where

$$\gamma = \max_{\hat{\alpha} \leq \hat{\mu} \leq \hat{\beta}} |Q_1(\hat{\mu})|.$$

Consequently, (15) holds for all $\hat{\mu}$ such that $\lambda_1(\hat{\mu}) \neq \lambda_2(\hat{\mu})$.

If $\lambda_1(\hat{\mu}) = \lambda_2(\hat{\mu})$, then by (14) we have, for $n \geq 2$,

$$|Q_n(\hat{\mu})| \leq \rho(\omega, \hat{\mu})^n + \delta n \rho(\omega, \hat{\mu})^{n-1} \leq \left[\frac{1}{n} R(\omega) + \delta \right] n R(\omega)^{n-1}, \quad (18)$$

where

$$\delta = \gamma + R(\omega).$$

By (17) and (18), it follows that for some constant K , independent of $\hat{\mu}$ and n , we have, for $n \geq 2$,

$$|Q_n(\hat{\mu})| \leq K n R(\omega)^n.$$

Therefore, (15) follows.

Lemma 2. *Under the hypothesis of Theorem 1, we have*

$$\limsup_{n \rightarrow \infty} \sqrt[n]{\max_{\hat{\alpha} \leq \hat{\mu} \leq \hat{\beta}} |Q_n(\hat{\mu})|} \geq R(\omega). \quad (19)$$

Proof. We seek to show that for any $\epsilon > 0$, there exists $\hat{\mu} = \hat{\mu}(\epsilon)$ such that $\hat{\alpha} \leq \hat{\mu}(\epsilon) \leq \hat{\beta}$ and such that

$$\limsup_{n \rightarrow \infty} \sqrt[n]{|Q_n(\hat{\mu}(\epsilon))|} \geq R(\omega) - \epsilon.$$

From this, (19) will follow.

For fixed ω , the root radius $\rho(\omega, \hat{\mu})$ is maximized over $\hat{\mu}$ in the interval $[\hat{\alpha}, \hat{\beta}]$ when $\hat{\mu} = \hat{\mu}_0$, where

$$\hat{\mu}_0 = \begin{cases} \hat{\beta} & \text{if } |\hat{\alpha}| \leq |\hat{\beta}|, \\ \hat{\alpha} & \text{if } |\hat{\alpha}| > |\hat{\beta}|. \end{cases} \quad (20)$$

(See, for instance, Young [12], page 186, or Young [11], Lemma 3.6.) Thus, we have

$$R(\omega) = \rho(\omega, \hat{\mu}_0).$$

In our subsequent discussion, let $\lambda_1(\hat{\mu}_0)$ and $\lambda_2(\hat{\mu}_0)$ be the roots of

$$\lambda^2 - (\omega \hat{\mu}_0) \lambda + (\omega - 1) = 0. \quad (21)$$

We label the roots so that

$$|\lambda_1(\hat{\mu}_0)| \geq |\lambda_2(\hat{\mu}_0)|.$$

We now consider various cases.

Case $\lambda_1(\hat{\mu}_0)$ and $\lambda_2(\hat{\mu}_0)$ real and $|\lambda_1(\hat{\mu}_0)| > |\lambda_2(\hat{\mu}_0)|$.

In this case if

$$-\lambda_2(\hat{\mu}_0) + Q_1(\hat{\mu}_0) \neq 0, \quad (22)$$

then we have, by (13),

$$Q_n(\hat{\mu}_0) = \left[\frac{-\lambda_2(\hat{\mu}_0) + Q_1(\hat{\mu}_0)}{\lambda_1(\hat{\mu}_0) - \lambda_2(\hat{\mu}_0)} \right] \lambda_1(\hat{\mu}_0)^n \left\{ 1 + \left(\frac{\lambda_2(\hat{\mu}_0)}{\lambda_1(\hat{\mu}_0)} \right)^n \left(\frac{\lambda_1(\hat{\mu}_0) - Q_1(\hat{\mu}_0)}{-\lambda_2(\hat{\mu}_0) + Q_1(\hat{\mu}_0)} \right) \right\}.$$

Hence

$$\limsup_{n \rightarrow \infty} \sqrt[n]{|Q_n(\hat{\mu}_0)|} = \rho(\omega, \hat{\mu}_0) = R(\omega).$$

Next, suppose that the roots $\lambda_1(\hat{\mu}_0)$ and $\lambda_2(\hat{\mu}_0)$ of (21) are real and that $|\lambda_1(\hat{\mu}_0)| > |\lambda_2(\hat{\mu}_0)|$, but

$$-\lambda^2(\hat{\mu}_0) - Q_1(\hat{\mu}_0) = 0.$$

Given any $\epsilon > 0$, by the continuity of the roots of the quadratic equation

$$\lambda^2 - (\omega\hat{\mu})\lambda + (\omega - 1) = 0,$$

as a function of its coefficients, and hence as a function of $\hat{\mu}$, we can find $\hat{\mu}_1$, close to $\hat{\mu}_0$, such that the following conditions hold:

$$\hat{\mu}_1 \in [\hat{\alpha}, \hat{\beta}],$$

$$|\lambda_1(\hat{\mu}_1)| > |\lambda_2(\hat{\mu}_1)|,$$

$$|\lambda_1(\hat{\mu}_1)| \geq R(\omega) - \epsilon,$$

$$-\lambda_2(\hat{\mu}_1) + \omega^*\hat{\mu}_1 + (1 - \omega^*) \neq 0.$$

The last condition is possible since the condition

$$-\lambda_2(\hat{\mu}) + \omega^*\hat{\mu} + (1 - \omega^*) = 0,$$

implies that

$$\omega^*(\omega^* - \omega)\hat{\mu}^2 + (2\omega - \omega)(1 - \omega^*)\hat{\mu} + (\omega + (\omega^*)^2 - 2\omega^*) = 0. \quad (23)$$

Since (23) has at most two distinct real roots, it follows that if (23) is satisfied for $\hat{\mu}_0$, then, for $\hat{\mu}$ sufficiently close to $\hat{\mu}_0$, but different from $\hat{\mu}_0$, the condition will not be satisfied.

As in the case where (22) holds, we can show that

$$\limsup_{n \rightarrow \infty} \sqrt[n]{|Q_n(\hat{\mu}_1)|} \geq R(\omega) - \epsilon.$$

Since ϵ can be arbitrarily small the result (19) follows for this case.

Case $\lambda_1(\hat{\mu}_0)$ and $\lambda_2(\hat{\mu}_0)$ are real and $\lambda_1(\hat{\mu}_0) = -\lambda_2(\hat{\mu}_0)$.

By (13) we have, if n is even,

$$Q_n(\hat{\mu}_0) = \lambda_1^n,$$

and hence

$$\limsup_{n \rightarrow \infty} \sqrt[n]{|Q_n(\hat{\mu}_0)|} \geq R(\omega).$$

Case $\lambda_1(\hat{\mu}_0) = \lambda_2(\hat{\mu}_0)$.

If

$$\omega^*\hat{\mu}_0 + (1 - \omega^*) - \lambda_1(\hat{\mu}_0) \neq 0,$$

then by (14) we have

$$Q_n(\hat{\mu}_0) = n[\lambda_1(\hat{\mu}_0)]^{n-1} \left\{ 1 + \frac{\lambda_1(\hat{\mu}_0)}{n[Q_1(\hat{\mu}_0) - \lambda_1(\hat{\mu}_0)]} \right\},$$

and hence

$$\limsup_{n \rightarrow \infty} \sqrt[n]{|Q_n(\hat{\mu}_0)|} = R(\omega). \quad (24)$$

On the other hand, if

$$Q_1(\hat{\mu}_0) - \lambda_1(\hat{\mu}_0) = 0,$$

then

$$Q_n(\hat{\mu}_0) = \lambda_1^n,$$

and (24) also holds. The result (19) holds for this case.

Case $\lambda_1(\hat{\mu}_0)$ and $\lambda_2(\hat{\mu}_0)$ are complex.

Let

$$\lambda_1(\hat{\mu}_0) = R e^{i\theta}, \quad \lambda_2(\hat{\mu}_0) = R e^{-i\theta},$$

where θ is not a multiple of π and where $R > 0$. By (16), we have, for $n \geq 2$, since $\lambda_1 \lambda_2 = R^2$,

$$\begin{aligned} Q_n(\hat{\mu}_0) &= (R^n / \sin \theta) \{ (Q_1/R) \sin n\theta - \sin(n-1)\theta \} \\ &= (R^n / \sin \theta) \{ [(Q_1/R) - \cos \theta] \sin n\theta + \sin \theta \cos n\theta \} \\ &= (R^n \Gamma / \sin \theta) \{ \cos \alpha \sin n\theta + \sin \alpha \cos n\theta \} \\ &= (R^n \Gamma / \sin \theta) \sin(n\theta + \alpha), \end{aligned} \tag{25}$$

where

$$\begin{cases} \Gamma^2 &= [(Q_1/R) - \cos \theta]^2 + \sin^2 \theta, \\ \cos \alpha &= [(Q_1/R) - \cos \theta] \Gamma^{-1}, \\ \sin \alpha &= (\sin \theta) \Gamma^{-1}. \end{cases}$$

Here $Q_1 = Q_1(\hat{\mu}_0)$. We claim that for some $\eta > 0$, given any positive integer N , there exists $n > N$ such that

$$|\sin(n\theta + \alpha)| \geq \eta.$$

If this is true then, by (25), it follows that

$$\limsup_{n \rightarrow \infty} \sqrt[n]{Q_n(\hat{\mu}_0)} \geq R.$$

On the other hand, if the claim is not true, then

$$\lim_{n \rightarrow \infty} \sin(n\theta - \alpha) = 0. \tag{26}$$

But the following lemma shows that (26) cannot hold and, hence, that Lemma 1 is true.

Lemma 3. *We have*

$$\lim_{n \rightarrow \infty} \sin(n\theta - \alpha) = 0, \tag{27}$$

if and only if θ and α are each multiples of π .

Proof. If α and θ are each multiples of π then $\sin(n\theta - \alpha) = 0$ for all n and hence (27) follows. Suppose now that α is a multiple of π and that (27) holds. Then (27) is equivalent to

$$\lim_{n \rightarrow \infty} \sin(n\theta) = 0. \quad (28)$$

If θ is not a multiple of π then

$$\theta/\pi = m + f,$$

where m is an integer and $0 < f < 1$. Let N be any integer and let n be an integer greater than N . Then

$$\sin(n\theta) = \sin(nm\pi + nf\pi) = (-1)^{nm} \sin(nf\pi).$$

If nf is not an integer, let $n' = n$. If nf is an integer then

$$(n+1)f = nf + f$$

is not an integer and we let $n' = n + 1$. Let k be the closest integer of $n'f$. Then

$$0 < |n'f - k| \leq 1/2,$$

and

$$0 < |n'f\pi - k\pi| < \pi/2.$$

If $|n'f\pi - k\pi| \geq \pi/4$ then

$$|\sin(nf\pi)| \geq \sqrt{2}/2.$$

Otherwise, let s be the smallest integer such that

$$|sn'f\pi - sk\pi| \geq \pi/4.$$

Evidently

$$|sn'f\pi - sk\pi| \leq \pi/2,$$

and hence

$$|\sin(sn'f\pi)| \geq \sqrt{2}/2.$$

Thus, for any N , we have an integer, say t , greater than N such that

$$|\sin(t\theta)| \geq \sqrt{2}/2,$$

which contradicts (28). Hence θ must be a multiple of π .

Suppose now that α is not a multiple of π . If (28) holds, then given any $\epsilon > 0$ there exists N such that, for all $n > N$ we have $|\sin(n\theta - \alpha)| < \epsilon$. For any such n , let $k = k(n)$ be an integer that minimizes $|n\theta - \alpha - k\pi|$. Then

$$|\sin(n\theta - \alpha)| = |\sin(n\theta - \alpha - k\pi)| < \epsilon,$$

where

$$|n\theta - \alpha - k\pi| \leq \pi/2.$$

We note that if $|x| \leq \pi/2$ then

$$|x| \leq \pi/2 |\sin x|. \quad (29)$$

This follows since for $0 \leq x \leq \pi/2$ we have

$$\frac{d}{dx} \left(\frac{\sin x}{x} \right) = \left(\frac{x - \tan x}{x^2} \right) \cos x \leq 0.$$

Since $\sin x/x$ is a decreasing function of x for $0 \leq x \leq \pi/2$ it follows that

$$\left(\frac{\sin x}{x} \right) \geq \left(\frac{\sin(\pi/2)}{\pi/2} \right) = 2/\pi,$$

and hence

$$x \leq (\pi/2) \sin x.$$

Therefore, (29) follows.

Evidently, we have, for all $n > N$,

$$|n\theta - \alpha - \pi k| \leq (\pi/2)\epsilon.$$

Moreover, for some ϵ_1 such that $|\epsilon_1| \leq \epsilon$,

$$\begin{cases} n\theta = \alpha + \pi k \pm (\pi/2)\epsilon_1, \\ 2n\theta - \alpha = 2\pi k + \alpha \pm \pi\epsilon_1. \end{cases}$$

Therefore

$$\sin(2n\theta - \alpha) = \sin(\alpha \pm \pi\epsilon_1).$$

But for some ϕ we have

$$|\sin(\alpha \pm \pi\epsilon_1) - \sin \alpha| \leq \pi|\epsilon_1| |\cos \phi| \leq \pi\epsilon,$$

and hence

$$\sin(2n\theta - \alpha) = \sin \alpha + \tau,$$

where

$$|\tau| \leq \pi\epsilon.$$

Thus, if we let $\epsilon < |\sin \alpha|/(2\pi)$ we have

$$|\sin(2n\theta - \alpha)| \geq |\sin \alpha|/2.$$

Thus given any N , there exists an integer, say t , such that $t > N$ and

$$|\sin(t\theta - \alpha)| \geq |\sin \alpha|/2.$$

This contradicts (28), and the proof of Lemma 3 is complete.

For Lemma 3, it follows that (26) cannot hold since θ is not a multiple of π . This completes the proof of Lemma 2. The proof of Theorem 1 is complete.

4 Determination of γ and ω

We have shown that no matter what value of ω^* is used the virtual asymptotic average spectral radius $\overline{S}(P_n(G))$ is given by $R(\omega, \gamma)$ where $R(\omega, \gamma)$ is given by (11). We now seek to determine γ and ω to minimize $R(\omega, \gamma)$.

As stated at the beginning of the proof of Lemma 2, $R(\omega, \gamma) = \rho(\omega, \gamma, \hat{\mu}_0)$, where $\hat{\mu}_0$ is given by (20). For given ω , we do not increase $\rho(\omega, \gamma, \hat{\mu}_0)$, if we choose γ to minimize

$$\max\{|\hat{\alpha}|, |\hat{\beta}|\} = \max\{|\gamma\alpha + (1 - \gamma)|, |\gamma\beta + (1 - \gamma)|\}.$$

The optimum value of γ is clearly

$$\overline{\gamma} = \frac{2}{2 - (\beta + \alpha)}.$$

For, if $\gamma' < \overline{\gamma}$ then

$$\begin{aligned} \gamma'\beta + (1 - \gamma') &= 1 - \gamma'(1 - \beta) > 1 - \overline{\gamma}(1 - \beta) = \frac{\beta - \alpha}{2 - (\beta + \alpha)} \\ &= \max\{|\overline{\gamma}\alpha + (1 - \overline{\gamma})|, |\overline{\gamma}\beta + (1 - \overline{\gamma})|\}. \end{aligned}$$

On the other hand, if $\gamma' > \bar{\gamma}$ then

$$-\gamma'\alpha - 1 + \gamma' > \bar{\gamma} - 1 - \bar{\gamma}\alpha = \frac{\beta - \alpha}{2 - (\beta + \alpha)}.$$

Evidently, $R(\omega, \bar{\gamma})$ is the root radius of

$$\lambda^2 - (\omega\bar{\sigma})\lambda + (\omega - 1) = 0, \quad (30)$$

where

$$\bar{\sigma} = \frac{\beta - \alpha}{2 - (\beta + \alpha)}. \quad (31)$$

We claim that the value of ω that minimizes the root radius of (30) is

$$\hat{\omega}_b = \frac{2}{1 + \sqrt{1 - \bar{\sigma}^2}}.$$

The corresponding root radius is

$$R(\hat{\omega}_b, \bar{\gamma}) = \sqrt{\hat{\omega}_b - 1}.$$

If $\omega > \hat{\omega}_b$ then the root radius is at least $\sqrt{\omega - 1} > \sqrt{\hat{\omega}_b - 1}$, since the product of the roots of (30) is $\omega - 1$. (We note that $2 > \omega_b > 1$ since $\bar{\sigma} < 1$.) Similarly, if $\omega \leq 0$, then $R(\omega, \bar{\gamma}) \geq 1 > \sqrt{\hat{\omega}_b - 1}$. If $0 < \omega < \hat{\omega}_b$, then the roots of (30) are real and the root radius is greater than $\sqrt{\hat{\omega}_b - 1}$, since

$$\frac{1}{2} \frac{d}{d\omega} \left[\omega\bar{\sigma} + \sqrt{(\omega\bar{\sigma})^2 - 4(\omega - 1)} \right] < 0,$$

for $0 < \omega < \hat{\omega}_b$. This follows since the above derivative is equal to

$$\frac{1}{2} \left\{ \frac{\bar{\sigma} + \sqrt{(\omega\bar{\sigma})^2 - 4(\omega - 1)} + (\omega\bar{\sigma}^2 - 2)}{\sqrt{(\omega\bar{\sigma})^2 - 4(\omega - 1)}} \right\},$$

and since

$$\left(\bar{\sigma} \sqrt{(\omega\bar{\sigma})^2 - 4(\omega - 1)} \right)^2 - (\omega\bar{\sigma}^2 - 2)^2 = 4(\bar{\sigma}^2 - 1) < 0.$$

To summarize, the optimum choices of γ and ω are given by

$$\begin{cases} \gamma = \bar{\gamma} = \frac{2}{2 - (\beta + \alpha)}, \\ \omega = \hat{\omega}_b = \frac{2}{1 + \sqrt{1 - \bar{\sigma}^2}}. \end{cases} \quad (32)$$

where

$$\bar{\sigma} = \frac{\beta - \alpha}{2 - (\beta + \alpha)}.$$

With these choices, we have

$$\bar{\mathcal{S}}(P_n(G)) = \sqrt{\hat{\omega}_b - 1}.$$

Let us now consider the special case $d = -1$ and also the case $e = 0$, which do not correspond to any choice of γ and ω_1 . For the case $e = 0$, we have, from (8),

$$\begin{cases} u^{(1)} = \omega_1(Gu^{(0)} + k) + (1 - \omega_1)u^{(0)}, \\ u^{(n+1)} = u^{(n)} + d(u^{(n)} - u^{(n-1)}). \end{cases}$$

The corresponding polynomials $P_0(G), P_1(G), \dots$ are given by

$$\begin{cases} P_0(G) = I, \\ P_1(G) = \omega_1 G + (1 - \omega_1)I, \\ P_n(G) = (1 + d)P_n(G) - dP_{n-1}(G). \end{cases}$$

By an analysis similar to that given in Section 3, we are led to consider the root radius of the quadratic equation

$$\lambda^2 - (1 + d)\lambda + d = 0,$$

whose roots are 1 and d . One can then show that

$$\bar{\mathcal{S}}(P_n(G)) \geq 1.$$

For the case $d = -1$, we have

$$\begin{cases} u^{(1)} = \omega_1(Gu^{(0)} + k) + (1 - \omega_1)u^{(0)}, \\ u^{(n+1)} = e(Gu^{(n)} + k - u^{(n)}) + u^{(n-1)}. \end{cases}$$

Here we are led to consider the root radius of the quadratic equation

$$\lambda^2 - e(\mu - 1)\lambda - 1 = 0,$$

which is not less than unity. One can then show that

$$\overline{S}(P_n(G)) \geq 1.$$

Thus, the two special cases that cannot be represented in terms of γ and ω are relatively unimportant.

5 Special Cases

We now consider the following two important special cases:

- (i) $\omega = \hat{\omega}_b, \quad \gamma = \bar{\gamma}, \quad \omega^* = 1,$
- (ii) $\omega = \hat{\omega}_b, \quad \gamma = \bar{\gamma}, \quad \omega^* = \hat{\omega}_b.$

Here $\hat{\omega}_b$ and $\bar{\gamma}$ are given by (32).

Evidently for Case (i), if $|\hat{\mu}| < \bar{\sigma}$, where $\bar{\sigma}$ is given by (31), both roots of (12), with $\omega = \hat{\omega}_b$, are real and simple. By (16), if $\lambda_1 \neq \lambda_2$, we have

$$Q_n(\hat{\mu}) = \hat{\mu} \left(\frac{\lambda_1^n - \lambda_2^n}{\lambda_1 - \lambda_2} \right) - \lambda_1 \lambda_2 \left(\frac{\lambda_1^{n-1} - \lambda_2^{n-1}}{\lambda_1 - \lambda_2} \right),$$

since $\omega^* = 1$, or

$$Q_n(\hat{\mu}) = \hat{\mu} S_{n-1}(\hat{\mu}) + (1 - \hat{\omega}_b) S_{n-2}(\hat{\mu}), \quad (33)$$

where we let

$$S_n(\hat{\mu}) = \begin{cases} \frac{\lambda_1^{n+1} - \lambda_2^{n+1}}{\lambda_1 - \lambda_2} & \text{if } \lambda_1 \neq \lambda_2, \\ (n+1)\lambda_1^n & \text{if } \lambda_1 = \lambda_2. \end{cases} \quad (34)$$

If $|\hat{\mu}| = \sigma$, then $\lambda_1 = \lambda_2$ and, by (14), we have

$$\begin{aligned} Q_n(\hat{\mu}) &= \lambda_1^n + (\hat{\mu} - \lambda_1)n\lambda_1^{n-1} \\ &= \hat{\mu}n\lambda_1^{n-1} - (n-1)\lambda_1^n \\ &= \hat{\mu}S_{n-1}(\hat{\mu}) + (1 - \hat{\omega}_b)S_{n-2}(\hat{\mu}). \end{aligned}$$

since $\lambda_1^2 = \hat{\omega}_b - 1$. Hence (33) holds for $|\hat{\mu}| \leq \bar{\sigma}$.

It was shown by Golub [2] (see also Young [8, p. 316]) that

$$\begin{aligned} & \max_{-\bar{\sigma} \leq \hat{\mu} \leq \bar{\sigma}} |\hat{\mu} S_{n-1}(\hat{\mu}) + (1 - \hat{\omega}_b) S_{n-2}(\hat{\mu})| \\ &= |\bar{\sigma} S_{n-1}(\bar{\sigma}) + (1 - \hat{\omega}_b) S_{n-2}(\bar{\sigma})| \\ &= \left[\frac{2 + (1 - \hat{r})(n-1)}{1 + \hat{r}} \right] \hat{r}^{n/2}, \end{aligned}$$

where

$$\hat{r} = \hat{\omega}_b - 1.$$

Therefore, we have

$$\bar{S}(P_n(G)) = \left[\frac{2 + (1 - \hat{r})(n-1)}{(1 + \hat{r})} \right] \hat{r}^{n/2}. \quad (35)$$

In Case (ii), the polynomials $Q_0(\hat{\mu}), Q_1(\hat{\mu}), \dots$ are defined by

$$\begin{cases} Q_0(\hat{\mu}) &= 1, \\ Q_1(\hat{\mu}) &= \hat{\omega}_b \hat{\mu} + (1 - \hat{\omega}_b), \\ Q_{n+1}(\hat{\mu}) &= (\hat{\omega}_b \hat{\mu}) Q_n(\hat{\mu}) + (1 - \hat{\omega}_b) Q_{n-1}(\hat{\mu}). \end{cases}$$

By (16), we have for the case $\lambda_1(\hat{\mu}) \neq \lambda_2(\hat{\mu})$,

$$Q_n(\hat{\mu}) = Q_1(\hat{\mu}) \left(\frac{\lambda_1^n - \lambda_2^n}{\lambda_1 - \lambda_2} \right) - \lambda_1 \lambda_2 \left(\frac{\lambda_1^{n-1} - \lambda_2^{n-1}}{\lambda_1 - \lambda_2} \right).$$

But since λ_1 and λ_2 satisfy

$$\lambda^2 - \hat{\omega}_b \hat{\mu} \lambda + (\hat{\omega}_b - 1) = 0, \quad (36)$$

we have $\hat{\omega}_b - 1 = \lambda_1 \lambda_2$ and $\hat{\omega}_b \hat{\mu} = \lambda_1 + \lambda_2$. Therefore

$$\begin{aligned} Q_n(\hat{\mu}) &= \left(\frac{\lambda_1^{n+1} - \lambda_2^{n+1}}{\lambda_1 - \lambda_2} \right) - \lambda_1 \lambda_2 \left(\frac{\lambda_1^n - \lambda_2^n}{\lambda_1 - \lambda_2} \right) \\ &= S_n(\hat{\mu}) + (1 - \hat{\omega}_b) S_{n-1}(\hat{\mu}). \end{aligned}$$

As before $\lambda_1(\hat{\mu}) \neq \lambda_2(\hat{\mu})$, if $|\hat{\mu}| < \bar{\sigma}$. If $|\hat{\mu}| = \bar{\sigma}$, then $\lambda_1(\hat{\mu}) = \lambda_2(\hat{\mu}) = \hat{\omega}_b \hat{\mu}/2$. By (14) and (34), we have

$$\begin{aligned} Q_n(\hat{\mu}) &= \lambda_1^n + [-\lambda_1 + 2\lambda_1 + (1 - \hat{\omega}_b)] n \lambda_1^{n-1} \\ &= (n+1)\lambda_1^n + (1 - \hat{\omega}_b)n\lambda_1^{n-1} \\ &= S_n(\hat{\mu}) + (1 - \hat{\omega}_b)S_{n-1}(\hat{\mu}). \end{aligned}$$

It is easy to show that $|Q_n(\hat{\mu})|$ is maximized in the interval $-\bar{\sigma} \leq \hat{\mu} \leq \bar{\sigma}$ if $\hat{\mu} = \bar{\sigma}$. For we have, by (34),

$$|S_n(\hat{\mu})| \leq (n+1)\rho(\hat{\mu})^n \leq (n+1)\hat{r}^{n/2},$$

where $\rho(\hat{\mu})$ is the root radius of (36). Moreover,

$$S_n(\bar{\sigma}) = (n+1)\hat{r}^{n/2}.$$

But since

$$\begin{aligned} S_n(\bar{\sigma}) + (1 - \hat{\omega}_b)S_{n-1}(\bar{\sigma}) &= (n+1)\hat{r}^{n/2} - \hat{r}n\hat{r}^{(n-1)/2} \\ &= \hat{r}^{n/2} \{n+1 - n\hat{r}^{1/2}\}, \end{aligned}$$

it follows that $|Q_n(\hat{\mu})|$ is maximized for $\hat{\mu} = \bar{\sigma}$. Therefore, we have

$$\overline{S}(P_n(G)) = \hat{r}^{n/2} \{1 + n(1 - \hat{r}^{1/2})\}. \quad (37)$$

Clearly, by (35) and (37), $\overline{S}(P_n(G))$ for the case $\omega^* = \hat{\omega}_b$ is larger than $\overline{S}(P_n(G))$ for the case $\omega^* = 1$, even though $\mathcal{S}(P_n(G))$ is the same in both cases.

Let us now assume that A is positive definite and has the 2×2 block form

$$A = \begin{bmatrix} D_1 & H \\ K & D_2 \end{bmatrix},$$

where D_1 and D_2 are nonsingular square diagonal matrices. Let us also assume that G represents the matrix $B = I - D^{-1}A$ corresponding to the Jacobi method, where

$$D = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}.$$

It is easy to show that the eigenvalues of B are real and lie in the interval $-\bar{\mu} \leq \mu \leq \bar{\mu}$ for some $\bar{\mu} < 1$. The linear stationary second-degree method with $\gamma = 1$, $\omega = \omega_b$ and $\omega^* = \omega_b$ corresponds to the successive overrelaxation method provided that one uses *compression* (see Golub and Varga [3]). Similarly, if $\gamma = 1$, $\omega = \omega_b$ and $\omega^* = 1$, one obtains the modified Sheldon method (see for instance Young [10], Chapter 10). This is a variant of the successive overrelaxation method where one uses $\omega = 1$ for the first iteration for the equation of system (1) corresponding to D_1 and then uses ω_b thereafter. (The regular successive overrelaxation method involves the use of ω_b for all equations and for all iterations.) The fact that $\bar{S}(P_n(G))$ is larger for the case $\omega^* = \hat{\omega}_b$ is perhaps surprising in the view of the fact that the spectral radius of the modified Sheldon method itself is larger (see, for instance, Young et al. [7]). However, based on an analysis of certain matrix norms the modified Sheldon method is actually better than the successive overrelaxation method (see Golub [2], Golub and Varga [3], Young [8], and Young and Kincaid [11]).

References

- [1] Avdelas, G., Galanis, S., and Hadjidimos, A., "On the optimization of a class of second order iterative schemes," *BIT* **23** (1983) 50-64.
- [2] Avdelas, G., de Pillis, J., Hadjidimos, A., and Neumann, M., "A guide to the acceleration of iterative methods whose iteration matrix is nonnegative and convergent," *SIAM J. Matrix Anal. Appl.* **9** (1988) 329-342.
- [3] Frankel, S.P., "Convergence rates of iterative treatments of partial differential equations," *Math. Tables Aids Comput.* **4** (1950) 63-75.
- [4] Golub, G.H., *The Use of Chebyshev Matrix Polynomials in the Iterative Solution of Linear Systems Compared with the Methods of Successive Overrelaxation*, Ph.D. thesis, University of Illinois, 1959.
- [5] Golub, G.H., and R.S. Varga, "Chebyshev semi-iterative methods, successive overrelaxation iterative methods, and second-order Richardson iterative methods, Parts I and II," *Numer. Math.* **3** (1961) 147-168.

- [6] Kincaid, D.R., *An Analysis of a Class of Norms of Iterative Methods for Systems of Linear Equations*, Ph.D. thesis, University of Texas at Austin, 1971.
- [7] Kincaid, D.R., "Extrapolated, semi-iterative, and second-degree methods for solving systems of linear equations," Report CNA-23, Center for Numerical Analysis, University of Texas at Austin, June, 1971.
- [8] Kincaid, D.R., "On complex second-degree iterative methods," *SIAM J. Numer. Anal.* **11** (1974) 211-218. (Also, Report CNA-36, Center for Numerical Analysis, University of Texas at Austin, March 1972.)
- [9] Kincaid, D.R., "Numerical results of the application of complex second-degree and semi-iterative methods," Report CNA-90, Center for Numerical Analysis, University of Texas at Austin, October 1974.
- [10] Young, D.M., "Iterative methods for solving partial difference equations of elliptic type," *Trans. Amer. Math. Soc.* **76** (1954) 92-111.
- [11] Young, D.M., "Convergence properties of the symmetric and unsymmetric successive overrelaxation methods and related methods," Report TNN-96, Computation Center, University of Texas at Austin, 1969; revised version appeared in *Math. of Comp.* **24** (1970) 793-807.
- [12] Young, D.M., *Iterative Solution of Large Linear Systems*, Academic Press, New York, 1971.
- [13] Young, D.M., "On the consistency of linear stationary iterative methods," *SIAM J. Numer. Anal.* **9** (1972) 89-96.
- [14] Young, D.M., "Second-degree iterative methods for the solution of large linear systems," *J. Approx. Theory* **5** (1972) 137-148.
- [15] Young, D.M., and D.R. Kincaid, "Norms of the successive overrelaxation method and related methods," Report TNN-94, Computation Center, University of Texas at Austin, 1969.
- [16] Young, D.M., M.F. Wheeler, and J.A. Downing, "On the use of the modified successive overrelaxation method with several relaxation factors," *Proc. of IFIP 65* (W.A. Kalenich, ed.), Spartan Books, Inc., Washington, D.C., 177-182, 1965.