A COMPARISON OF THE SUCCESSIVE OVERRELAXATION
METHOD AND SEMI-ITERATIVE METHODS USING
CHEBYSHEV POLYNOMIALS*

RICHARD S. VARGA

1. Introduction. It is the main purpose of this paper to compare the
mean rates of convergence of two well-known schemes for solving self-
adjoint partial difference equations of elliptic type: the Young-Frankel
[6, 1] successive overrelaxation method, and the semi-iterative Chebyshev
polynomial method as described by Lanczos [2, p. 42], Stiefel [3], and
others. More generally, the analysis is applicable to any matrix equation
of the form

\[ Au = k, \]

provided the matrix \( A = \| a_{i,j} \| \) is symmetric and positive definite, and, in
the sense of Young [6, p. 93], satisfies property (A).

For semi-iterative methods, one considers iterates \( u_i \), where

\[ u_{i+1} = Mu_i + k \quad (i = 0, 1, 2, \ldots), \]

\( M \) being a specific matrix. Then, one forms from the sequence of vectors
\( u_i \) a new sequence of vectors

\[ t_n = \sum_{j=0}^{n} \nu_j(n) u_j \quad (n = 0, 1, 2, \ldots), \]

the constants \( \nu_j(n) \) being real numbers. Such a procedure is called a semi-
iterative method with respect to the matrix \( M \).

While it is known [8, p. 293] that the successive overrelaxation method
converges at least twice as fast as any semi-iterative method with respect
to the Jacobi method, we shall give a different proof of this result, which
generalizes to semi-iterative methods with respect to the Gauss-Seidel
method. For the Gauss-Seidel method, the result is that the successive
overrelaxation method converges at least as fast as any semi-iterative
method with respect to the Gauss-Seidel method.

When solution by high-speed computing machines is involved, it should
be noted that the successive overrelaxation method has the further advan-
tage of requiring no auxiliary storage of extra iterates \( u_m \), whereas
semi-iterative methods require that a few iterates \( u_m \) be stored, along with
suitable coefficients.

In a final section, an analogous discussion is given for semi-iterative

* Received by the editors October 3, 1956, and in revised form, February 28,
1957.
methods applied to matrices whose (complex) eigenvalues are known to be confined to a circle in the complex plane. The result is that the best semi-iterative method with respect to the successive overrelaxation method, assuming only that all eigenvalues \( \lambda \) of the successive overrelaxation method satisfy \( |\lambda| \leq \rho < 1 \), is simply the basic method repeated \( n \) times, where \( n \) is the order of the semi-iterative method.

2. Description of methods. Equation (1) is rewritten in the form

\[
u = Bu + f
\]

where the \( N \times N \) matrix \( B = \| b_{i,j} \| \) and the vector \( f \) are given by\(^1\)

\[
b_{i,j} = \begin{cases} -a_{i,i}/a_{i,i} & , \quad i \neq j \\ 0 & , \quad i = j \end{cases}, \quad f_i = k_i/a_{i,i},
\]

for \( i = 1, 2, \ldots, N \). For the successive overrelaxation method, one forms the sequence of vectors defined by

\[
u_i^{(m+1)} = \omega \left( \sum_{j=1}^{i-1} b_{i,j} u_j^{(m+1)} + \sum_{j=i+1}^{N} b_{i,j} u_j^{(m)} + f_i \right) + (1 - \omega) u_i^{(m)}
\]

where the "overrelaxation factor" \( \omega \) is a parameter which is fixed throughout the course of iteration. The equation above may be written symbolically as

\[
u_{m+1} = L_{\omega} \nu_m + g,
\]

where \( g \) is a fixed vector, and \( L_{\omega} \) denotes a linear operator.

Let \( u_0 \) be a "trial solution" of (1), and let the sequences of vectors \( u_i \) and \( t_n \) be defined, respectively, by (2) and (3). If \( u_0 \) is the unique solution of (1), then each \( t_n \) is also a solution of (1) if and only if\(^2\)

\[
\sum_{j=0}^{n} v_j(n) = 1 \quad \quad (n = 0, 1, 2, \ldots).
\]

If \( v \) denotes the unique solution of (1), and the \( n \)th error vector \( t_n - v \) is denoted by \( e_n^* \), then

\[
e_n^* = p_n(M)e_0, \quad p_n(x) = \sum_{j=0}^{n} v_j(n)x^j,
\]

and \( e_0 = u_0 - v \) is the error vector associated with the trial solution \( u_0 \). As

\(^1\) Since \( A \) is symmetric and positive definite, (1) has a unique solution, and \( a_{i,i} > 0 \) for all \( i \).

\(^2\) For polynomials similarly normalized, see [2, p. 41], and [3, p. 63].
a consequence of (8), we shall henceforth assume $p_n(1) = 1$. For comparison, we note that in the successive overrelaxation method

(10) \[ \varepsilon_n = L^n_{\omega, \omega}[e_0], \]

which corresponds to the choice $p_n(x) = x^n$, and $M = L_{\omega, \omega}$.

If the matrix $M$ has real eigenvalues $\lambda_k$, and the interval $a \leq x \leq b$ is the smallest interval containing the $\lambda_k$'s, we define

\[ \mu[p_n(M)] = \max_{a \leq x \leq b} \left\{ |p_n(x)|^{1/n} \right\}, \]

(11) \[ R[p_n(M)] = -\log \mu[p_n(M)]. \]

The quantities $\mu[p_n(M)]$ and $R[p_n(M)]$ are respectively the average spectral norm and the average rate of convergence\(^3\) at the $n^{th}$ step of the semi-iterative method with respect to the matrix $M$. For $p_n(x) = x^n$, we see that $\mu[p_n(M)]$ and $R(p_n(M))$ are independent of $n$, and are, for matrices $M$ with real eigenvalues, respectively the usual definition of the spectral norm $\bar{\mu}$ and rate of convergence $R$ of $M$ [6, p. 96].

3. Basis of Comparisons. In this section, we shall compare the rate of convergence of the successive overrelaxation method with that of the semi-iterative method with respect to the Jacobi method. We choose the matrix $M$ of (2) to be the matrix $B$ defined by (5). Under our initial assumptions, $A$ is symmetric and positive definite, and satisfies property (A). For this case, it is known [6] that all the eigenvalues $\lambda$ of $B$ are real, and lie in the symmetric interval $-\bar{\mu} \leq \lambda \leq + \bar{\mu} < 1$, where $\bar{\mu}$ is the spectral norm of $B$. Without loss of generality, we may assume that $A$ is consistently ordered [6, p. 93]. The best choice [6] of $\omega$ is given by

(12) \[ \omega_b = 1 + \left[ \frac{\bar{\mu}}{1 + [1 - \bar{\mu}^2]^{1/2}} \right]^2, \]

and

(13) \[ R[L_{\omega_b, \omega_b}] = - \log (\omega_b - 1). \]

To select the best semi-iterative method with respect to the matrix $B$, we choose the polynomial $p_n(x)$ such that

(14) \[ \max_{-\bar{\mu} \leq x \leq +\bar{\mu}} |p_n(x)| \]

\[^3\text{Specifically, if the largest degree of the elementary divisors [5, Ch. III] of the matrix } p_n(M), \text{ for } n \text{ fixed, is unity, then the reciprocal of } R[p_n(M)] \text{ is an estimate of the least positive integer } k \text{ for which} \]

\[ \| (p_n(M))^{k}\|_0 \leq e^{-n} \| e_0 \|, \]

where $\| \chi \|$ refers to the Euclidean norm of the vector $x$. 
is smallest. It is well-known that the solution of this problem is in Chebyshev polynomials, and we have explicitly

\[ p_n(x) = \frac{T_n(x/\bar{\mu})}{T_n(1/\bar{\mu})}, \]

where \( T_n(x) = \cos [n \cos^{-1} x] \) is the Chebyshev polynomial of degree \( n \). We shall call this particular method the Chebyshev semi-iterative method with respect to the Jacobi method. By definition, we have

\[ \max_{-\bar{\mu} \leq x \leq \bar{\mu}} |p_n(x)| = \frac{\max_{-1 \leq x \leq 1} |T_n(x)|}{|T_n(1/\bar{\mu})|} = \frac{1}{T_n(1/\bar{\mu})}, \]

since \( T_n(\alpha) > 1 \) for \( \alpha > 1 \). Forming the ratio of the average rates of convergence of the successive overrelaxation method and this Chebyshev semi-iterative method, we have

\[ R[L_{\omega_k}, R[p_n(B)]] = \left[ \frac{2 \log \left[ 1 + \left( 1 - \frac{1}{\bar{\mu}} \right)^2 \right]}{n} \right]^2 \left( \log T_n(1/\bar{\mu}) \right). \]

Since, for \( \alpha > 1 \), \( \log T_n(\alpha)/n \) increases monotonically to \( \cosh^{-1}(\alpha) \), and since \( \cosh^{-1}(\alpha) = \log (\alpha + \sqrt{\alpha^2 - 1}) \) for \( \alpha > 1 \), we have, replacing \( \alpha \) by \( 1/\bar{\mu} \), \( R[L_{\omega_{\omega_k}}, R[p_n(B)]] \geq 2 \). This proves in a different manner the following theorem of Young [8, p. 293].

**Theorem 1.** The successive overrelaxation method with the optimum overrelaxation factor converges at least twice as fast as the Chebyshev semi-iterative method with respect to the Jacobi method, and therefore at least twice as fast as any semi-iterative method with respect to the Jacobi method. Furthermore, as the number of iterations tends to infinity, the successive overrelaxation method becomes exactly twice as fast as this Chebyshev semi-iterative method.

We finally remark that for values of \( \omega \) near the optimum \( \omega_*, \omega_k \), the successive overrelaxation method still has a faster rate of convergence than the Chebyshev semi-iterative method with respect to the Jacobi method, and therefore a faster rate of convergence than any semi-iterative method with respect to the Jacobi method.

4. **Extension to polynomials** in \( L_{\omega,1} \). In this section, we merely replace the Jacobi matrix \( B \) in the previous section by the Gauss-Seidel or Liebmann operator \( L_{\omega,1} \) [1, 6]. Assuming \( A \) to be consistently ordered, it is then known [6, p. 100] that the eigenvalues \( \lambda_k \) of \( L_{\omega,1} \) satisfy \( 0 \leq \lambda_k \leq \mu^2 < 1 \), where \( \mu \) is the spectral norm of the matrix \( B \). As before, the polynomial \( p_n(x) \) of degree \( n \), normalized so that \( p_n(1) = 1 \), which has the property

---

*See also [4, Chapter VI].*
that max$_{0 \leq x \leq \bar{\mu}}$ | $p_n(x)$ | is smaller than all other such normalized polynomials, will be

$$(16) \quad p_n(x) = T_n \left( \frac{2x}{\bar{\mu}^2} - 1 \right) / T_n \left( \frac{2}{\bar{\mu}^2} - 1 \right).$$

The semi-iterative method based on these polynomials will be called the Chebyshev semi-iterative method with respect to the operator $L_{\alpha,1}$. The average rate of convergence of this iterative scheme at the $n$th step is

$$(17) \quad R[p_n(L_{\alpha,1})] = \log T_n \left( \frac{2}{\bar{\mu}^2} - 1 \right) / n.$$

If we form the ratio of $R[L_{\alpha\omega}]$ to $R[p_n(L_{\alpha,1})]$, then using the monotone property of $\log T_n(\alpha)/n$ and the previously used identity for $\cosh^{-1}(\alpha)$ for $\alpha > 1$, we obtain

$$R[L_{\alpha\omega}] / R[p_n(L_{\alpha,1})] \geq 2 \log \left( \frac{1}{\bar{\mu} + \left( \frac{1}{\bar{\mu}^2} - 1 \right)^{\frac{1}{2}}} \right) / \log \left( \frac{2}{\bar{\mu}^2} - 1 + \left( \left( \frac{2}{\bar{\mu}^2} - 1 \right)^{\frac{1}{2}} - 1 \right)^{\frac{1}{2}} \right).$$

But the right hand side of the inequality above reduces identically to unity for $\bar{\mu} < 1$. This proves

**Theorem 2.** The successive overrelaxation method with the optimum overrelaxation factor converges at least as fast as the Chebyshev semi-iterative method with respect to the operator $L_{\alpha,1}$, and therefore at least as fast as any semi-iterative method with respect to the operator $L_{\alpha,1}$. Furthermore, as the number of iterations tends to infinity, the successive overrelaxation method becomes exactly as fast as the Chebyshev semi-iterative method with respect to the operator $L_{\alpha,1}$.

5. Extensions to polynomials in the operator $L_{\alpha\omega}$. - If we have, as before, that the eigenvalues of $B$ are real and lie in $-\bar{\mu} \leq x \leq \bar{\mu}$, then we can formulate the problem of finding the best polynomial of degree $n$, normalized so that $p_n(1) = 1$, having the smallest absolute value on the interval $-\bar{\mu} \leq x \leq \bar{\mu}$, and we are naturally led to Chebyshev polynomials. With these polynomials, we then defined the Chebyshev semi-iterative method with respect to the matrix $B$, which was, in some sense, the optimum semi-iterative method with respect to the matrix $B$. The same is true if we consider, rather than the matrix $B$, the linear operator $L_{\alpha,1}$ whose eigenvalues $\lambda_k$ are also real and satisfy $0 \leq \lambda_k \leq \bar{\mu}^2$, and optimize the selection of a sequence of normalized polynomials whose absolute value on the interval $0 \leq x \leq \bar{\mu}^2$ is smallest. The resulting semi-iterative method defined by this sequence of polynomials was called the Chebyshev semi-iterative method with respect to the operator $L_{\alpha,1}$. As we pass to the case where
1 < \omega < 2$, the operator $L_{\omega, \omega}$ does not have all real eigenvalues [6, p. 101], and the selection of a sequence of normalized polynomials to define a semi-iterative method with respect to the operator $L_{\sigma, \omega}$ is not immediate. As before, we have

\[ \varepsilon_n^* = \sum_{j=0}^n \nu_j(n) L^{j}_{\sigma, \omega} \varepsilon_0 = p_n(L_{\sigma, \omega}) \varepsilon_0, \]

where $p_n(1) = 1$. We now assume that the eigenvalues $\lambda_k$ of $L_{\sigma, \omega}$ satisfy $|\lambda_k| \leq \rho < 1$. If $g_n(z)$ is any complex polynomial of degree $n$, let $M_{g_n}(r)$ denote the maximum modulus function of $g_n(z)$, i.e.,

\[ M_{g_n}(r) = \max_{|z|^2 \leq r} |g_n(z)|. \]

Completely analogous to the previous sections, let $S_n$ be the set of all polynomials $g_n(z)$ of degree $n$ for which $g_n(1) = 1$, and consider

\[ \min_{g_n \in S_n} \{ M_{g_n}(r) \}. \]

The following theorem, due to E. H. Zarantonello\(^5\), seems to be of interest by itself.

**Theorem 3.** For all $r$ such that $0 \leq r \leq 1$, $\min_{g_n \in S_n} \{ M_{g_n}(r) \} = r^n$ for all positive integers $n$.

**Proof.** As is well known, we have

\[ M_{g_n}(r) = \lim_{p \to \infty} \left( \int_0^{2\pi} |g_n(re^{i\theta})|^2 r^2 d\theta \right)^{1/2p}. \]

Clearly, we have for any positive integer $p$,

\[ \min_{g_n \in S_n} \left( \int_0^{2\pi} |g_n(re^{i\theta})|^2 r^2 d\theta \right)^{1/p} \geq \min_{Q \in S_{np}} \left[ \left( \int_0^{2\pi} |Q(re^{i\theta})|^2 r^2 d\theta \right)^{1/p} \right]^{1/p}, \]

since if $g_n \in S_n$, then

\[ Q(z) = [g_n(z)]^p \in S_{np}. \]

If

\[ Q(z) = \sum_{k=0}^{np} a_k z^k, \]

then

\[ \left( \int_0^{2\pi} |Q(re^{i\theta})|^2 r^2 d\theta \right)^{1/p} = \sqrt{2\pi} \left( \sum_{k=0}^{np} |a_k|^2 r^{2k} \right)^{1/p}. \]

---

\(^5\) Personal communication.
OVERRELAXATION AND SEMI-ITERATIVE METHODS

But

\[
\sqrt{2\pi} \left( \sum_{k=0}^{n^p} |a_k|^2 r^{2k} \right)^{\frac{1}{p}} \geq \sqrt{2\pi} r^{n^p} \left( \sum_{k=0}^{n^p} |a_k|^2 \right)^{\frac{1}{p}},
\]

since \(0 \leq r \leq 1\). Since \(Q \in S_{n^p}\), then \(\sum_{k=0}^{n^p} a_k = 1\). Therefore, using Schwarz' inequality, we have

\[
1 = \left| \sum_{k=0}^{n^p} a_k \right| \leq \sum_{k=0}^{n^p} |a_k| \leq (n^p + 1)^{\frac{1}{p}} \left( \sum_{k=0}^{n^p} |a_k|^2 \right)^{\frac{1}{p}},
\]

and

\[
\left\{ \int_0^{2\pi} |Q(e^{i\theta})|^2 \, d\theta \right\}^{\frac{1}{p}} \geq \sqrt{2\pi} r^{n^p} \left( \sum_{k=0}^{n^p} |a_k|^2 \right)^{\frac{1}{p}} \geq \left( \frac{2}{n^p + 1} \right)^{\frac{1}{p}} r^{n^p}.
\]

Thus,

\[
\min_{Q \in S_{n^p}} \left[ \left( \int_0^{2\pi} |Q(e^{i\theta})|^2 \, d\theta \right)^{\frac{1}{p}} \right]^{\frac{1}{p}} \geq \left( \frac{2}{n^p + 1} \right)^{\frac{1}{2p}} r^{n^p}.
\]

Letting \(p \to \infty\), we have

\[
\min_{Q \in S_n} M_n(r) \geq r^n.
\]

Since \(z^n \in S_n\), and \(M_{z^n}(r) = r^n\), then

\[
\min_{Q \in S_n} M_n(r) = r^n \quad \text{for all } 0 \leq r \leq 1,
\]

for all positive integers \(n\).

In view of this theorem, we have

**Theorem 4.** The best semi-iterative method with respect to the successive overrelaxation method, \(\omega > 1\), which can be obtained, assuming only that eigenvalues \(\lambda_i\) of the successive overrelaxation method satisfy \(|\lambda| \leq \rho < 1\), is simply the basic method repeated \(n\) times, where \(n\) is the order of the semi-iterative method.

We remark that, for \(\omega > 1\), the error vector associated with the best semi-iterative method with respect to \(L_{\omega,\omega}\) satisfies

\[
\varepsilon_n^* = L_{\omega,\omega}^n \varepsilon_0 = \varepsilon_n.
\]

This particularly simple form of a semi-iterative method with respect to \(L_{\omega,\omega}\) incidentally has been used repeatedly to solve multigroup diffusion problems in two or more (space) dimensions [9, 10]. Apparently, the choice for the iteration method is not related to the theorem above, but rather to the inherent simplicity of the iteration method of (20).
REFERENCES


Westinghouse Electric Corporation
Atomic Power Division