Spectrum Transformation for Divergent Iterations

Murli M. Gupta
The George Washington University
Washington, D.C.

and Institute for Computational Mechanics in Propulsion
Lewis Research Center
Cleveland, Ohio

March 1991
SPECTRUM TRANSFORMATIONS FOR DIVERGENT ITERATIONS

Murli M. Gupta
Department of Mathematics
The George Washington University
Washington, D.C. 20052

and

Institute for Computational Mechanics in Propulsion
Lewis Research Center
Cleveland, Ohio 44135

SUMMARY

In this paper we describe certain spectrum transformation techniques that can be used to transform a diverging iteration into a converging one. We consider two techniques called spectrum scaling and spectrum enveloping and discuss how to obtain the optimum values of the transformation parameters. Numerical examples are given to show how this technique can be used to transform diverging iterations into converging ones; this technique can also be used to accelerate the convergence of otherwise convergent iterations.

1. INTRODUCTION

Consider the linear system

\[ Ax = b \] (1)

which may be rewritten as

\[ x = Tx = c \] (2)

and solved using the iteration

\[ x_{n+1} = T x_n + c \quad (n \geq 0) \] (3)

This iteration, called the basic iteration, converges to the unique solution of equation (2) if and only if the spectral radius of the iteration matrix \( T \) is smaller than 1, i.e., \( \rho(T) < 1 \).

When the basic iteration defined by equation (3) is convergent, a number of methods can be used to accelerate the rate of convergence. Such acceleration schemes are known as Chebyshev acceleration (ref. 7), semi-iterative methods (SIM) (refs. 4 and 10), and hybrid SIM (ref. 3). Several important optimality results have been established in the cited papers.

Now consider the case when the basic iteration described by equation (3) is divergent, i.e., when some eigenvalues of the iteration matrix \( T \) lie outside the unit circle \( (\rho(T) > 1) \). In such cases, only a few techniques are known that allow for the solution of equation (2) to be obtained iteratively. In particular, when the real parts of all eigenvalues of the basic iteration matrix \( T \) lie in the interval \((-1,1)\), it is possible to use the spectrum enveloping technique (refs. 2 and 8) to transform the basic iteration in such a way that the modified iteration is guaranteed to converge. When the real part of any eigenvalue of \( T \) lies outside the interval
(-1,1) most of the acceleration schemes fail, except possibly for the hybrid SIM which has been shown to apply in certain special situations (ref. 3).

We consider the situation where the eigenvalues of the iteration matrix $T$ may lie anywhere in the left half of the complex plane, in fact, real ($\xi$) $<$ 1, where $\xi = \xi + i\eta \in \sigma(T)$, $\sigma(T) =$ set of all eigenvalues of $T$. In this situation, a simple linear transformation can be used to map $\sigma(T)$ onto $S = \{\xi = \xi + i\eta: -1 < \text{real}(\xi) = \xi < 1\}$. Once the transformed eigenvalue spectrum lies in $S$, an enveloping ellipse can be defined to obtain guaranteed convergence. The composite procedure is equivalent to a polynomial mapping of $\sigma(T)$ the unit circle.

In this paper, we examine the conditions under which the spectrum scaling and spectrum enveloping transformations are successful. We determine the optimum values of the transformation parameters and show that the parameters of spectrum scaling may be chosen so as to obtain convergence. The spectrum scaling may also be used to accelerate the convergence when the basic iteration given by equation (3) is convergent. The spectrum enveloping transformation is used to obtain even faster convergence. We present examples to exhibit the theoretical and computational values of the transformation parameters and corresponding values of asymptotic convergence factors.

2. THE SPECTRUM SCALING

Consider the basic iteration:

$$x_{n+1} = Tx_n + c, \quad (n \geq 0), \quad x_0 \text{ arbitrary}$$

Assume that the eigenvalue spectrum of the iteration matrix $T$, denoted by $\sigma(T)$, is a compact set in the complex plane $C$. Let $\zeta = \xi + i\eta$ be an eigenvalue of $T$. The iteration (4) is convergent if $\rho(t) < 1$, i.e., $|\zeta|^2 = |\xi|^2 + |\eta|^2 < 1$ for all $\zeta \in \sigma(T)$. The iteration (4) diverges if $|\zeta| > 1$ for any $\zeta \in \sigma(T)$.

Now consider the case when $|\zeta| > 1$. In particular, let

$$-\infty < a \leq \text{Real} (\zeta) = \xi \leq A < 1$$

$$-\infty < -b \leq \text{Imag} (\zeta) = \eta \leq b < \infty$$

i.e., $\sigma(T)$ is contained in the rectangle $[a,A] \times [-b,b]$ in the complex plane.

We define a two-step iteration

$$z_{n+1} = Tx_n + c$$

$$x_{n+1} = p \cdot z_{n+1} + (1 - p) x_n$$

The combined form of equation (6) is given by

$$x_{n+1} = T' x_n + pc$$
where

\[ T' = pT + (1 - p)I \]  \hspace{1cm} (8)

The eigenvalues \( \zeta' (= \xi' + i\eta') \) of \( T' \) are related to the eigenvalues \( \zeta \) of \( T \) through the following relations:

\[
\begin{align*}
\zeta' &= p\zeta + (1 - p) \\
\xi' &= p\xi + (1 - p), \hspace{0.5cm} \eta' = p\eta
\end{align*}
\]

The scaling parameter \( p \) must be chosen such that \( |\xi'| < 1 \) for all \( \zeta \in \sigma(T) \) and for some \( \gamma \in (0, 1) \):

\[-1 < -\gamma < \xi' = p(\xi - 1) + 1 < \gamma < 1 \]

or,

\[
\frac{1 - \gamma}{1 - \xi} < p < \frac{1 - \gamma}{1 - \xi} \]  \hspace{1cm} (9)

As equation (9) must hold for all \( \zeta = \xi + i\eta \), the value of the scaling parameter \( p \) must satisfy

\[
0 < \frac{1 - \gamma}{1 - A} \leq p \leq \frac{1 + \gamma}{1 - A} \]  \hspace{1cm} (10)

This condition requires that

\[
0 < \gamma_0 = \frac{A - a}{2 - A - a} \leq \gamma < 1 \]  \hspace{1cm} (11)

With conditions (10) and (11), the eigenvalue spectrum \( \sigma(T') \) of \( T' \) lies in \( S \):

\[
\begin{align*}
\zeta' &= \xi' + i\eta' \in \sigma(T') \\
-1 < -\gamma < \xi' &\leq \gamma < 1 \\
-\infty < -b \frac{1 - \gamma}{1 - A} &\leq \eta' \leq b \frac{1 + \gamma}{1 - A} < \infty
\end{align*}
\]  \hspace{1cm} (12)

The spectral radius \( \rho(T') \) of the transformed matrix \( T' \) is given by \( \rho^2(T') = \gamma^2 + p^2b^2 \), which is the smallest when \( p \) takes its minimum permissible value \( (1 - \gamma)/(1 - A) \). Thus,

\[
\rho^2(T) = \gamma^2 + b^2 \frac{(1 - \gamma)^2}{(1 - A)^2} \]  \hspace{1cm} (13)

**Theorem 1.** - The value of \( \rho(T') \) given by equation (13) is minimum when \( \gamma \) is chosen such that \( \gamma = \max \left( \gamma_0, \gamma_1 \right) \), where \( \gamma_0 = (A - a)/(2 - A - a) \) and \( \gamma_1 = k/(1 + k) \), \( k = b^2/(1 - A)^2 \).
Proof. - From equation (13), \( p^2(T') = \gamma^2 + k(1 - \gamma)^2 \), where \( k = b^2/(1 - A)^2 > 0 \). This expression takes a local minimum when \( \gamma = \gamma_1 = k/(1 + k) \). Consider two cases: if \( \gamma_0 < \gamma_1 < 1 \), the minimum value of \( p(T') \) is clearly obtained when \( \gamma = \gamma_1 \). In this case, \( p(T') = \sqrt{T_1} = \sqrt{[k/(1 + k)]} < 1 \). If \( \gamma_1 < \gamma_0 < 1 \), the value of \( p(T') \) is an increasing function of \( \gamma \) in \((\gamma_0, 1)\) and the minimum value of \( p(T') \) is obtained when \( \gamma = \gamma_0 \). In either case, the minimum value of \( p(T') \) is obtained when \( \gamma = \max(\gamma_0, \gamma_1) \).

Remarks. - The spectrum scaling defined in this section is applicable whenever \( \xi = \text{Real}(\zeta) < 1 \) for all eigenvalues \( \zeta \) of the iteration matrix \( T \). Similar scaling can be defined, and corresponding values of optimal parameters obtained, when \( \xi > 1 \) for all eigenvalues of \( T \).

3. THE SPECTRUM ENVELOPING

If the eigenvalue spectrum of the iteration matrix \( T \) lies in the infinite set \( S \{ \zeta : -1 < \text{real}(\zeta) < 1 \} \) in the complex plane, irrespective of whether \( T \) defines a convergent or a divergent iteration, a spectrum enveloping transformation can be defined to obtain accelerated convergence (refs. 2 and 8). In this section, we consider the problem of obtaining the optimum values of the transformation parameters.

Let the eigenvalues \( \zeta = \xi + i\eta \) of the basic iteration matrix \( T \) satisfy the following conditions:

\[
-1 < \Re(\zeta) = \xi < \gamma < 1
\]
\[
-\infty < -\Im(\zeta) = \eta < \beta < \infty
\]

Define an ellipse, lying entirely within \( S \), with semi-axis of length \( M \) lying on the imaginary line and semi-axis of length \( m \) lying on the real line, \( 0 < m < 1 \). The spectrum enveloping iteration for the basic iteration (3) may be written as follows:

\[
z_{n+1} = Ty_n + c
\]
\[
y_{n+1} = (1 + \lambda^2)z_{n+1} - \lambda^2 y_{n-1}
\]

where \( \lambda = (m - M)/(m + M) \), and \( \mu \) is the unique root in \((0, 1)\) of \((m - M)^2 - 2\mu + (m + M) = 0 \).

It is known (refs. 2 and 8) that the sequence \( \{y\} \) defined by equation (15) is convergent whenever \( \lambda \) is nonzero; the asymptotic convergence factor is given by

\[
\mu = \frac{M + m}{1 + \sqrt{[M^2 - m^2 + 1]}}
\]
\[
\phi_n = \frac{M^2 + m^2}{M + 1}, \quad \text{as } m < 1
\]
\[
= \frac{1 - \frac{m}{1 + M}}{1 + M} < 1
\]
The semi-axes \( m, M \) of the enveloping ellipse must be chosen so as to contain all elements of the set \( \sigma(T) \); i.e., \( m, M \) must be chosen such that
\[ \gamma^2/m^2 + \beta^2/M^2 \leq 1. \] This requires that \( M \geq m\beta/\sqrt{[m^2 - \gamma^2]} \). As \( \mu \) is an increasing function of \( M \), we use the smallest value of \( M \) to define the enveloping ellipse:
\[ M = \frac{m\beta}{\sqrt{[m^2 - \gamma^2]}} \] (16)

We now have,
\[ \mu < \mu_0 = 1 - f(m) \]
where
\[ f(m) = \frac{1 - m}{1 + \frac{m\beta}{\sqrt{[m^2 - \gamma^2]}}} \] (17)

In order to obtain the optimum convergence we minimize the value of \( \mu_0 \); this is equivalent to the problem of maximizing the value of \( f(m) \).

**Theorem 2.** \( f(m) \) given by equation (17) is maximum when \( m \) lies in the interval \( (\gamma, \gamma^{2/3}) \).

**Proof.** Note that \( f(\gamma) = f(1) = 0 \) and \( f(m) > 0 \) for \( m \in (\gamma, 1) \). The derivative of \( f(m) \) is given by
\[ f'(m) = \frac{-1 + \frac{\beta(y^2 - m^3)}{(m^2 - \gamma^2)^{3/2}}}{\Delta^2} \] (18)
where \( \Delta \) is the denominator in equation (17). As \( f'(\gamma) > 0 \), and \( f'(\gamma^{2/3}) < 0 \), the peak in the graph of \( f(m) \) is obtained somewhere between \( m = \gamma \) and \( m = \gamma^{2/3} \).

The next result describes an optimum value of \( m \).

**Theorem 3.** An optimum value of the semi-axis \( m \) is given by the positive root of the equation
\[ (1 + \beta)m^3 - 1.5 \gamma^2 m - \beta \gamma^2 = 0 \] (19)

**Proof.** The maximum value of the function \( f(m) \) is obtained at the point where \( f'(m) = 0 \). From equation (18), \( f'(m) = 0 \) when
\[ \beta(y^2 - m^3) = (m^2 - \gamma^2)^{3/2} \]
which gives, as a first approximation,
\[ \beta(y^2 - m^3) = m^3 \left(1 - \frac{1.5 \gamma^2}{m^2}\right) \]
which yields the cubic equation (19).
From Theorem 3, the positive root of equation (19) gives the optimum value of $m$:

$$m = \left( \frac{0.5\gamma^3\beta}{1 + \beta} + D \right)^{1/3} + \left( \frac{0.5\gamma^3\beta}{1 + \beta} - D \right)^{1/3}$$  \hspace{1cm} (20)$$

when $D = (\gamma/2)^4[2\beta^2(1 + \beta) - \gamma^2]/(1 + \beta)^3 \geq 0$. The above root of equation (19) is valid when $\beta$ is large enough so that $D \geq 0$. In this case, the value of $m$ may often be approximated by

$$m = m_1 = 2\left( \frac{0.5\gamma^3\beta}{1 + \beta} \right)^{1/3}$$  \hspace{1cm} (21)$$

When $\beta$ is small so that $\gamma^2 > [2\beta^2(1 + \beta)]$, the three (real) roots of equation (19) are given by

$$m = \gamma\sqrt{\frac{2}{(1 + \beta)}} \cos \frac{\phi}{3}, \quad \gamma\sqrt{\frac{2}{(1 + \beta)}} \cos \frac{\phi + \pi}{3}, \quad \gamma\sqrt{\frac{2}{(1 + \beta)}} \cos \frac{\phi - \pi}{3}$$  \hspace{1cm} (22)$$

where $\cos(\phi) = \gamma^{-2}\beta\sqrt{2 + 2\beta}$. In this case, the optimum value of $m$ is given by the root in equation (22) that lies in the interval $(\gamma, \gamma^{2/3})$. If no such value of $m$ is available from equation (22), the length of the semi-axis $m$ is taken as $\gamma^{2/3}$. In either case, the value of $m$ given by equations (20) and (22) provides a near optimum value of the asymptotic convergence factor $\mu < \mu_0 = 1 - f(m)$ and the asymptotic rate of convergence is given by $-\log \mu > -\log \mu_0$.

4. THE COMPOSITE TRANSFORMATION

The composite transformation is obtained by first applying the spectrum scaling to the iteration given in equation (3), and then applying the spectrum enveloping to the scaled iteration given by equation (7). The composite iteration is described as follows:

$$y_{n+1} = p(1 + \lambda\mu^2)(T_y y_{n+1} + c) + (1 - p)(1 + \lambda\mu^2)y_n - \lambda\mu^2 y_{n-1}$$  \hspace{1cm} (23)$$

This is a stationary two-step iteration. As all optimized semi-implicit methods for equation (2) degenerate into stationary two-step methods on computers with finite word length (ref. 7), we need consider only two-step iterations. This iteration involves the parameters $p$, $\gamma$, $m$, and $M$ defined by

$$0 < \frac{1 - \gamma}{1 - A} < p < \frac{1 + \gamma}{1 - A}, \quad 0 < \gamma_0 = \frac{A - a}{2 - A - a} < \gamma < 1,$$

$$\lambda = \frac{m - M}{m + M}, \quad 0 < \gamma < m < 1, \quad M = \frac{mpb}{\sqrt{(m^2 - \gamma^2)}}$$

$$\quad m < 1.$$
The optimum values of these parameters are given by:

\[ \gamma = \max(\gamma_0, \gamma_1) \]  

(24a)

where

\[ \gamma_0 = (A - a)/(2 - A - a) \]
\[ \gamma_1 = k/(1 + k), \quad k = b^2/(1 - A)^2 \]

(24b)

\[ \lambda = \frac{m - M}{m + M} \]  

(24c)

\[ M = \frac{mpb}{\sqrt{m^2 - \gamma^2}} \]  

(24d)

where \( m \) is given by Theorem 3. The asymptotic convergence factor \( \mu \) for iteration (23) is given by

\[ \mu = \frac{M + m}{1 + M^2 - m^2 + 1} < 1 \]

The asymptotic rate of convergence is \(-\log \mu > -\log \beta(T)\).

5. THE COMPUTATIONAL ALGORITHM

When the basic iteration defined by equation (3) is divergent under the conditions given by equation (5), it is often a simple matter to estimate the dominant eigenvalue of the iteration matrix \( T \) using the power method (ref. 12). This eigenvalue may provide a good estimate for \( a \) and \( b \); the value of \( A \) may be estimated through the use of the power method on the matrix \((T - aI)\). Of course, more sophisticated methods are available in the literature for estimating the extreme eigenvalues of \( T \) (see, e.g., (refs. 6 and 11)).

Once the values of \( a, A, \) and \( b \) are available, the scaling parameter \( p \), the scaled matrix \( T' \), and the optimum scaling interval \((-\gamma, \gamma)\) may be obtained from Section 2. If the eigenvalues of \( T \) satisfy the conditions (14) of Section 3, or if it is desired to apply the spectrum enveloping to the scaled iteration defined by equation (7), the parameters \( m, M \) of the enveloping ellipse may be obtained from Section 3 and convergence follows.

The composite iteration described by equation (23) may be carried out in three phases:

Phase 0: Compute using equation (23) with \( p = 1, \lambda = 0 \) to estimate \( a, A, b \).

Phase 1: Obtain scaling parameters \( p \) and \( \lambda \) from Theorem 1 and compute using spectrum scaling with equation (7) (or eq. (23) with \( \lambda = 0 \)).
Phase 2: Obtain enveloping parameters $m$, $M$, $\lambda$, and $\mu$ from Theorems 2 and 3 and compute using spectrum enveloping with equation (15) (or eq. (23) with $p = 1$).

Phase 3: Obtain scaling and enveloping parameters $p$, $\gamma$, $\lambda$, $\mu$, $m$, and $M$ from equation (24) and compute using the composite iteration (23).

6. NUMERICAL EXAMPLES

We now consider a few examples.

Example 1: Consider the following convection diffusion equation representing convection dominated flows:

$$-\varepsilon \left( \frac{\partial u}{\partial x^2} + \frac{\partial u}{\partial y^2} \right) + \frac{\partial u}{\partial x} = 0, \quad 0 < x < 1, \quad 0 < y < 1$$

with appropriate boundary conditions (refs. 5 and 8).

This differential equation is discretized on a uniform $(N + 1)$ by $(N + 1)$ grid for $\varepsilon = 0.005$ using central difference approximations. When the resulting linear system of order $(N - 1)^2$ is solved using Jacobi iteration, the iterations diverge for $N \leq 32$ (ref. 8). The eigenvalue spectrum of the Jacobi iteration matrix lies in the set $S: -1 < \text{real} (\zeta) = \xi + i\eta < 1$ for all eigenvalues $\zeta$; an enveloping ellipse may be defined to obtain convergence. As an example, when $N = 8$, the eigenvalues $\zeta$ of the iteration matrix $T$ satisfy $-\gamma \leq \zeta \leq \gamma$, $-\beta \leq \eta \leq \beta$ with $\gamma = 0.46194$, $\beta = 5.75574$, and $\rho(T) = 5.7742$ (fig. 1). The parameters for the enveloping ellipse in phase 2 may be obtained from equation (21) as $m = 0.5665$ and $M = 9.943266$ and the asymptotic convergence factor is given by $\mu = 0.957404$.

The eigenvalue spectrum of the Gauss-Seidel iteration matrix lies outside the set $S$; the conditions (5) are satisfied with $\alpha = -33.1385$, $\lambda = 0.2134$, $b = 5.3176$, $\rho(T) = 33.3419$ (fig. 2). Optimum values of the scaling parameters $p$, $\gamma$ are obtained from Theorem 1: $\gamma = \max(\gamma_0, \gamma_1) = \max(0.954955, 0.978587) = 0.978587$ and $p = (1 - \gamma)/(1 - A) = 0.027222$. The scaled matrix $T'$ of phase 1 satisfies $\rho(T') = \sqrt{\gamma^2 + p^2 b^2} = 0.989235 < 1$ (fig. 3). The use of spectrum enveloping of phase 2 provides the parameters for the enveloping ellipse: $m = 0.988568$, $M = 1.208813$ which provides even faster convergence with the asymptotic convergence factor $\mu = 0.988280$.

Example 2: Consider the biharmonic equation:

$$\frac{\partial^4 u}{\partial x^4} + \frac{2\partial^4 u}{\partial x^2 \partial y^2} + \frac{\partial^4 u}{\partial y^4} = 0, \quad 0 < x < 1, \quad 0 < y < 1$$

with prescribed boundary conditions. This equation is discretized on a uniform $(N + 1)$ by $(N + 1)$ grid using the standard 13-point finite difference approximation (refs. 1 and 9). The resulting linear system of order $(N - 1)^2$ may be solved iteratively.

When Jacobi iteration is used to solve the resulting linear system, the iterations are found to be divergent for all values of $N$ and the spectral radius of the iteration matrix is found to be close to 2. As an example, for $N = 8$,
a = -1.9699, A = 0.9454, b = 0.02927, and \( \rho(T) = 1.9699 \) (fig. 4). The spectrum scaling is applicable with the optimum values of scaling parameters given by Theorem 1: \( \gamma = 0.9639 \) and \( p = 0.6613 \). The resulting iteration of phase 1 is convergent with \( \rho(T') = 0.9641 \) (fig. 5). When the spectrum enveloping of phase 2 is applied to the transformed iteration given by \( T' \), the optimum values of the enveloping ellipse are given by \( m = 0.98205 \) and \( M = 0.101171 \) and the composite iteration is convergent with the asymptotic convergence factor \( \mu = 0.892256 \).

When Gauss-Seidel iteration is used to solve the linear system corresponding to discrete biharmonic equation, the iterations are convergent, e.g., for \( N = 8 \), \( \rho(T) = 0.901248 \). Spectrum scaling of phase 1 may still be carried out: For \( N = 8 \), \( a = -0.001816 \), \( A = 0.898199 \), \( b = 0.074083 \) (fig. 6). The spectrum scaling is applicable with the scaling parameters \( \gamma = 0.815514 \), \( p = 1.81222 \), and \( \rho(T') = 0.826491 \). The composite use of spectrum scaling and spectrum enveloping in phase 3 gives \( \beta = pb = 0.134255 \), \( m = 0.872879 \), and \( M = 0.376549 \) and the composite iteration defined by equation (23) converges with the asymptotic convergence factor \( \mu = 0.7730 \).

CONCLUSIONS

We have described two spectrum transformation techniques, viz., spectrum scaling and spectrum enveloping, that can be used to transform a diverging iteration into a converging one. We have also discussed the method of obtaining the optimum values of the transformation parameters. Numerical examples show that, in addition to their effectiveness on divergent iterations, these techniques are also useful for accelerating the convergence of an already converging iteration.

REFERENCES


Figure 1: Jacobi eigenvalues $\sigma(T)$ of CDS matrix ($N=8$)
Figure 2: Gauss-Seidel eigenvalues $\sigma(T)$ of CDS matrix $(N=8)$
Figure 3: Transformed Gauss-Seidel eigenvalues $\sigma(T')$ of CDS matrix ($N=8$)
Figure 4: Jacobi eigenvalues $\sigma(T)$ of biharmonic matrix $(N=8)$
Figure 5: Transformed Jacobi eigenvalues $\sigma(T')$ of biharmonic matrix $(N=8)$
Figure 6: Gauss-Seidel eigenvalues $\sigma(T)$ of biharmonic matrix $(N=8)$
In this paper we describe certain spectrum transformation techniques that can be used to transform a diverging iteration into a converging one. We consider two techniques called spectrum scaling and spectrum enveloping and discuss how to obtain the optimum values of the transformation parameters. Numerical examples are given to show how this technique can be used to transform diverging iterations into converging ones; this technique can also be used to accelerate the convergence of otherwise convergent iterations.