JOURNAL OF COMPUTER AND SYSTEM SCIENCES 13, 213–222 (1976)

Some Convergence Results on the Method of Gradients for $Ax = \lambda Bx$

S. F. McCormick*

Department of Mathematics, Colorado State University, Fort Collins, Colorado 80521

Received December 5, 1973; revised November 1, 1975

The method of gradients is a recently developed iterative technique for the solution of the generalized eigenvalue problem $Ax = \lambda Bx$. This paper essentially determines the size of the neighborhoods about the unit vectors in each eigenspace in which choices of initial estimates guarantee convergence of the iteration to the appropriate eigenvector.

1. Introduction

Lately, a great deal of attention has been directed toward methods for solution of the generalized eigenvalue problem

$$Ax = \lambda Bx \tag{1}$$

for square matrices A and B, which includes the classical eigenvalue problem

$$Ax = \lambda x \tag{2}$$

as a special case. Solutions of (1) will henceforth be referred to as eigenvectors of (A, B). Of the various gradient-like methods treated (cf. [1–9]), what we will call the method of gradients [1, 3, 6, 9] is especially important when A and B are large, sparse, and unsymmetric. This method is the subject of this paper.

The procedure is a simple gradient projection algorithm which is given formally by

$$\begin{aligned} y_n &= x_n - s(F(x_n)\,\nabla F(x_n)/\|\nabla F(x_n)\|^2), \\ x_{n+1} &= y_n/\|y_n\|, \end{aligned} \tag{3}$$

where the scalar $0 < s < 4$, $\|y\|$ represents the Euclidean norm of y, and the functional F is defined by

$$F(x) = \langle Ax, Ax\rangle\,\langle Bx, Bx\rangle - \langle Ax, Bx\rangle^2. \tag{4}$$

* This work was supported in part by NSF Grant number GJ34737.

Here, $\langle x, y\rangle$ denotes the Euclidean inner-product of the vectors x and y. Note that the gradient of F is then given by

$$\nabla F(x) = 2[\langle Ax, Ax\rangle B^*Bx + \langle Bx, Bx\rangle A^*Ax - \langle Ax, Bx\rangle (A^*Bx + B^*Ax)], \tag{5}$$

where A^* and B^* denote the adjoints of A and B, respectively. The fact that (3) is in the form of a gradient projection algorithm follows from the equivalence of problem (1) to the constrained minimization problem of finding a unit vector x such that

$$F(x) = \min\{F(u): u \text{ in } H, \| u \| = 1\}. \tag{6}$$

The extension of iteration (3) to general optimization problems is realized in this way and has been exploited in an article by the author [7].

Rodrigue [9] chose to treat the technique with the projection phase ommitted so that the method has the form

$$x_{n+1} = x_n - s\,(F(x_n)\,\nabla F(x_n)/\|\nabla F(x_n)\|^2). \tag{7}$$

The difficulty with this approach is that convergence to the null vector often occurs and, consequently, the theoretical results are somewhat unsettling. Note that convergence of the two procedures is otherwise the same up to differing by a normalizing constant. Moreover, for a given choice of x_0, the limit of the sequence in either (3) or (7), if it exists, must be a zero of and, hence, an eigenvector of (A, B). This follows from taking limits of both sides of the defining equations and noting

$$F(x) = \tfrac{1}{4}\langle \nabla F(x), x\rangle. \tag{8}$$

It is therefore important to determine which unit eigenvectors of (A, B) are indeed limit points of sequences (3) produced by different choices of the initial guess x_0 and to also have some idea of the nature of the initial estimates that provide convergence to a particular unit eigenvector. More precisely, it is desirable to be able to specify neighborhood balls about each of the eigenvectors of (A, B) such that convergence of iteration (3) to a particular eigenvector is guaranteed for any normalized choice of x_0 in its surrounding neighborhood. Such is the aim of this paper.

The central result is that each nondegenerate eigenvector of (A, B) exhibits a neighborhood ball with the above properties and, quite surprisingly, are all of equal size and depend only on s when $B^{-1}A$ is symmetric. Moreover, in the limiting case $s = 0$, the balls are as large as possible without intersecting. Although the general situation for problem (1) is somewhat more complex, the size of the balls are nevertheless independent of the eigenvalues, a property which is in marked contrast to other techniques like the power method, inverse iteration, and the angular gradient technique [8]. This nature of the method of gradients is one of its most significant characteristics and suggests advantages over other iterative techniques when intermediate eigenvalues and their eigenvector are sought. This remains true even though one should keep in mind that convergence rates do depend on the relative positions of the eigenvalues of (A, B).

2. Main Results

For simplification we shall henceforth assume that A and B are real n by n matrices and that B is nonsingular.

Remark. Let u be a real normalized n vector. Then the following are equivalent:

a. $F(u) = 0$,

b. $F(u) = \min\{F(x)\colon x \text{ in } R^n, \|Bx\| = 1\}$,

c. $\nabla F(u) = 0$,

d. $Au = \lambda Bu$ for some real λ.

The equivalence of a to b follows directly from applying the Schwarz inequality to the definition of $F(x)$ in (4). The equality condition implies the equivalence of a to c. Finally, substitution shows that d implies a and that a implies d is a consequence of (8).

Our first theorem summarizes results presently known about the method of gradients [3, 7, 9], so the proof is ommitted. Not that the conclusions indicate qualitative local convergence only.

THEOREM 1. *Let* $0 < s < 4$ *and suppose that* λ *is a nondegenerate eigenvalue of* (A, B) *in the sense that the minimal polynomial of* $B^{-1}A$ *has only an elementary divisor associated with* λ. *Let* $E_\lambda = \{u \text{ in } R^n\colon \|u\| = 1, Au = Bu\}$ *and for each* x *in* R^n *where possible define* $\nabla x = x - u_x$ *letting* u_x *represent the closest point in* E_λ *to* x. *Then for some* $\delta > 0$ *the sequence* x_n *in* (3) *is well defined and linearly convergent to* u_{x_0} *for any* x_0 *in the set*

$$\mathscr{B}_\delta = \{x \text{ in } R^n\colon \|x\| = 1, \|\Delta x\| < \delta\}.$$

Theorem 1 establishes the existence of a "neighborhood ball" about E_λ in which convergence is guaranteed for the method of gradients. To have some idea of the usefulness of the technique it now becomes necessary to determine allowable sizes of such neighborhoods, which is what the next two theorems attempt to do. We first consider the special case in which $B = I$, the identity matrix, for which it is necessary to have some estimate of the quantity

$$d_\lambda = \max\{|\langle (A - \lambda I)h, u\rangle| (\|(A - \lambda I)h\|^2 - \langle (A - \lambda I)h, u\rangle^2)^{-1/2}\colon u \text{ in } E_\lambda, h \text{ in } E_\lambda^\perp\}. \quad (9)$$

Here $E_\lambda^\perp$ denotes the set of unit vectors in R^n that are orthogonal to E_λ. Note that d_λ

can be thought of as a measurement of the nondegeneracy of λ and that an alternate definition is given by

$$d_\lambda = \max\{|a/b| : (A - \lambda I)h = au + bv,\ h \text{ and } v \text{ in } E_\lambda^\perp\}. \tag{10}$$

Observe that $d_\lambda < \infty$ for λ nondegenerate and that $d_\lambda = 0$ if each u in E_λ is also an eigenvector of A^*.

THEOREM 2. *Suppose $B = I$ and $0 < s < 4$ and let λ represent a nondegenerate eigenvalue of (A, B). Choose α_0 so that*

$$1 > \alpha_0 \geqslant [4/(8 - s)]^{1/2}, \tag{11}$$

$$d_\lambda^2 \leqslant (\alpha_0^2(8 - s) - 4)^2/[\alpha_0^2(8 - s)^2(1 - \alpha_0^2)], \tag{12}$$

and

$$d_\lambda^2 \leqslant \alpha_0^2/(1 - \alpha_0^2). \tag{13}$$

Then the conclusions of Theorem 1 are valid for the choice $\delta = 2(1 - \alpha_0)$. Note that, for this choice,

$$\mathscr{B}_\delta = \{\alpha u + \beta h : u \text{ in } E_\lambda,\ h \text{ in } E_\lambda^\perp,\ \alpha \text{ and } \beta \text{ reals, } \alpha > \alpha_0\}.$$

Before we present the proof, we first remark that condition (11) is quite a loose restriction on the size of $\mathscr{B}_\delta$, as we shall see in the corollary; condition (13) simply ensures that no other eigenvectors of (A, B) lie in $\mathscr{B}_\delta$; and condition (12) is restrictive only when d_λ is much different from zero. The possibility that the quantities $\langle Ah, u\rangle$ are nonzero obscures conception of the size of $\mathscr{B}_\delta$, since otherwise (12) and (13) are empty restrictions. The corollary immediately following the proof should provide a clearer picture.

Proof. Let $x = \alpha u + \beta h$ denote an element of $\mathscr{B}_\delta$ where we suppress the subscript of u_x and where $h \in E_\lambda^\perp$ and, therefore, $\alpha^2 + \beta^2 = 1$. That $\mathscr{B}_\delta$ is given according to the theorem follows from noting

$$\begin{aligned} \| \Delta x \|^2 &= \|(\alpha - 1)u + \beta h\|^2 \\ &= (\alpha^2 - 2\alpha + 1) + \beta^2 \\ &= 2(1 - \alpha), \end{aligned}$$

so that the requirement $\| \Delta x \| < \delta$ is equivalent to $\alpha > \alpha_0$.

To prove the remainder of the theorem, we first formally define

$$q(x) = x - (sF(x)\, \nabla F(x)/\| \nabla F(x) \|^2).$$

Now temporarily suppose we could prove for x in $\mathscr{B}_\delta - E_\lambda$ that

a. $\| \nabla F(x) \| \neq 0$ and $\| q(x) \| \neq 0$,

b. $\langle q(x), u \rangle \geqslant 0$,

and

c. $\langle x, u \rangle (8 - s) F(x) \geqslant 2 \langle \nabla F(x), u \rangle$.

Then a would imply that $q(x)$ and $q(x)/\| q(x) \|$ are well defined. From c it would then follow that

$$\left(\langle x, u \rangle - \frac{sF(x)\langle \nabla F(x), u \rangle}{\| \nabla F(x) \|^2} \right)^2 > \langle x, u \rangle^2 \left(1 - \frac{s(8 - s) F^2(x)}{\| \nabla F(x) \|^2} \right)$$

which, coupled with b, would imply that

$$\langle q(x), u \rangle > \langle x, u \rangle \| q(x) \|$$

and, hence

$$\begin{aligned} \|(q(x)/\| q(x) \|) - E_\lambda \|^2 &= \|(q(x)/\| q(x) \| - u_x \|^2 \\ &= 2(1 - (\langle q(x), u \rangle / \| q(x) \|) \\ &< 2(1 - \langle x, u \rangle) \\ &= \| \Delta x \|^2. \end{aligned}$$

Then, choosing $\delta' > 0$ small enough to ensure the validity of the conclusions of Theorem 1 and letting

$$k = \max\{\|(q(x)/\| q(x) \|) - E_\lambda \| / \| \Delta x \| : x \text{ in } \mathscr{B}_\delta - \mathscr{B}_{\delta'}\},$$

it would then follow from the finite dimensionality and boundedness of $\mathscr{B}_\delta - \mathscr{B}_{\delta'}$ that $k < 1$. Linear convergence would therefore be guaranteed in all of $\mathscr{B}_\delta$ and the theorem would be proved. We therefore complete the proof by establishing a, b, and c.

First note that if A is replaced by $A - \lambda I$, then neither $F(x)$ nor, therefore, $\nabla F(x)$ is altered. We may for the remainder of the proof, then, assume that $\lambda = 0$.

To prove (a) we first assume $\| \nabla F(x) \| = 0$ and, thus, that $Ax = \mu x$ for some $\mu \neq 0$. Since $Ax = \beta Ah$ it follows that $Ah = k\alpha u + k\beta h$ for some scalar $k \neq 0$. From (10), therefore,

$$| k\alpha / k\beta | \leqslant d\lambda.$$

Thus,

$$d_\lambda^2 \geqslant \alpha^2/(1 - \alpha^2),$$

which contradicts (13) and establishes that $\| \nabla F(x) \| \neq 0$. (a) now follows from the observation

$$\begin{aligned} \| q(x) \| &\geqslant \langle q(x), x \rangle \\ &= 1 - (s \langle \nabla F(x), x \rangle^2 / 4 \| \nabla F(x) \|^2) \\ &\geqslant 1 - (s/4) \\ &> 0. \end{aligned}$$

To prove (b), note that

$$\begin{aligned} \langle q(x), u \rangle &= \alpha - (s \langle \nabla F(x), x \rangle \langle \nabla F(x), u \rangle / 4 \| \nabla F(x) \|^2) \\ &\geqslant \alpha - (s/4). \end{aligned}$$

But since $\alpha^2 > 4/(8 - s)$ it suffices to show for $0 < s < 4$ that

$$4/(8 - s) \geqslant (s^2/4^2)$$

or, equivalently,

$$s^3 - 8s^2 + 64 \geqslant 0.$$

The only critical point of the polynomial $s^3 - 8s^2 + 64$ in the interval $[0, 4]$ is at $s = 0$. Examining the polynomial for a minimum at the endpoints $s = 0, 4$ shows that the minimum is zero, thus proving the validity of (b).

For (c), we first observe

$$F(x) = \beta^2(\langle Ah, Ah \rangle - \langle Ah, x \rangle^2)$$

and

$$\langle \nabla F(x), u \rangle = 2\beta^2(\langle Ah, Ah \rangle \alpha - \langle Ah, x \rangle \langle Ah, u \rangle).$$

We must therefore show that

$$\alpha(8 - s)(\langle Ah, Ah \rangle - \langle Ah, x \rangle^2) - 4(\langle Ah, Ah \rangle \alpha - \langle Ah, x \rangle \langle Ah, u \rangle) \geqslant 0. \quad (14)$$

Now let a, b, ϵ, and η be real numbers such that $\epsilon^2 + \eta^2 = 1$ and $Ah = au + b(\epsilon h + \eta v)$ for some unit vector v orthogonal to both u and h. Then a simple calculation shows that (14) is equivalent to the nonnegativity of the polynomial

$$P(\epsilon) = \alpha(8 - s)(a^2 + b^2 - (a\alpha + b\epsilon\beta)^2) - 4(\alpha b^2 - a\beta b\epsilon),$$

which is of degree two in ϵ. Since the coefficient of ϵ^2 is nonnegative, to prove (14) it suffices to show that $P(\pm 1) \geqslant 0$. Observe that

$$\alpha P(\pm 1) = \alpha^2(8 - s)(b\alpha \pm a\beta)^2 + 4\alpha b(\beta a \pm \alpha b)$$

and, dividing by $b^2\alpha^2$ and letting $d = | a\beta/b\alpha |$, we have

$$\begin{aligned}(P(\pm 1)/\alpha b^2) &= \alpha^2(8-s)(1 \pm d)^2 - 4(1 \pm d)\\ &\geqslant \alpha^2(8-s)(1-d)^2 - 4(1-d)\\ &= (1-d)(\alpha^2(8-s) - 4 - \alpha^2(8-s)d).\end{aligned}$$

Note that from (13) it follows that

$$\begin{aligned}| \alpha/\beta | &> \alpha_0(1-\alpha_0{}^2)^{-1/2}\\ &\geqslant d_\lambda\\ &\geqslant | a/b |.\end{aligned}$$

Clearly $d < 1$, and it suffices to prove that

$$(\alpha^2(8-s) - 4) - \alpha^2(8-s)d \geqslant 0.$$

This, of course, is (12) rewritten. We have thus proved (c) and, therefore, the theorem.

COROLLARY. *Suppose A is a symmetric matrix, $B = I$, and $0 < s < 4$. Define the set*

$$\mathscr{B} = \{\alpha u + \beta h\colon h \text{ in } E_\lambda^\perp, \alpha^2 > 4/(8-s)\}.$$

Then the sequence in (3) is well defined and converges linearly to u_{x_0} for any x_0 in $\mathscr{B}$.

Note that the "balls" $\mathscr{B}$ are equal for all values of λ and depend solely on s. Moreover, as s goes from four to zero they expand monotonically from the sets E_λ to "circular" surfaces on the unit sphere that are as large as possible. That is, in the limiting case $s = 0$, although the balls do not intersect, their closures do. This fact, coupled with the local properties of the method of gradients, suggests that numerical implementation of the algorithm should begin with small s and increase to the value $s = 2$ in later stages of the iterative procedure. No apparent advantage is gained by having $2 < s < 4$.

We now turn our attention to the general case for which we need the quantity

$$\begin{aligned}D_\lambda = \max\{&|\langle(A-\lambda B)h, Bu\rangle|(\|(A-\lambda B)h\|^2 \|Bu\|^2 - \langle(A-\lambda B)h, Bu\rangle^2)^{-1/2}\colon\\ &u \text{ in } E_\lambda, h \text{ in } E_\lambda^\perp\}.\end{aligned} \tag{15}$$

This quantity corresponds to d_λ with equality in the case $B = I$ and exhibits the same characteristics. As before, an alternate definition is given by

$$\begin{aligned}D_\lambda = \max\{| a/b |\colon (A-\lambda B)h = aBu + bBv,\\ \langle Bv, Bu\rangle = 0, \| Bu \| = \| Bv \| = 1\}.\end{aligned} \tag{16}$$

Note that $D_\lambda < \infty$ if λ is nondegenerate and a sufficient condition for $D_\lambda = 0$ is the symmetry of $B^{-1}A$.

Theorem 2 applies directly to the case when B is orthogonal. The possibility that B alters the orthogonality relationship between E_λ and $E_\lambda^\perp$, however, requires that we take this into account as we do by defining

$$\mathcal{O}_\lambda = \max\{(1/\| Bu \|^2)[(\| Bh \|^2 \| Bu \|^2 - \langle Bu, Bh\rangle^2 D_\lambda + |\langle Bu, Bh\rangle|]: u \text{ in } E_\lambda, h \text{ in } E_\lambda^\perp\}. \tag{17}$$

THEOREM 3. *Let* $0 < s < 4$ *and suppose* λ *is a nondegenerate eigenvalue of* (A, B). *Let* α_0 *be such that*

$$1 > \alpha_0 \geqslant (4/(8 - s))^{1/2}, \tag{18}$$

$$\mathcal{O}_\lambda^2 \leqslant (\alpha_0^2(8 - s) - 4)^2/[\alpha_0^2(8 - s)^2(1 - \alpha_0^2)], \tag{19}$$

and

$$\mathcal{O}_\lambda^2 \leqslant \alpha_0^2/(1 - \alpha_0^2). \tag{20}$$

Then the conclusions of Theorem 1 *are valid for* $\delta = 2(1 - \alpha_0)$.

Proof. Let $x = \alpha u + \beta h$ as in the proof of Theorem 2. With $q(x)$ as in the proof of Theorem 2, to prove Theorem 3 it suffices as before to establish (a), (b), and (c) listed in the proof of Theorem 2. For this purpose, we again assume without loss of generality and for convenience that $\lambda = 0$ noting that $F(x)$ and $\nabla F(x)$ are unchanged by the replacement of A by $A - \lambda B$.

The proof of (a) and (b) are just as before and we settle with the task of establishing (c). Note, as previously, that

$$F(x) = \beta^2(\| Ah \|^2 \| Bx \|^2 - \langle Ah, Bx\rangle^2)$$

and

$$\langle \nabla F(x), u\rangle = 2\beta^2(\| Ah \|^2 \langle Bx, Bu\rangle - \langle Ah, Bx\rangle\langle Ah, Bu\rangle.$$

We must therefore show that

$$\begin{aligned}&\alpha(8 - s)(\| Ah \|^2 \| Bx \|^2 - \langle Ah, Bx\rangle^2)\\ &- 4(\| Ah \|^2 \langle Bx, Bu\rangle - \langle Ah, Bx\rangle\langle Ah, Bu\rangle) \geqslant 0\end{aligned}$$

or, rewritten with the factor α, that

$$(\alpha^2(8 - s) - 4)(\| Ah \|^2 \| Bx \|^2 - \langle Ah, Bx\rangle^2) + 4\beta^2 F(h) - 4\alpha\beta\langle Ah, Bu\rangle\langle Ah, Bh\rangle \geqslant 0. \tag{22}$$

Let a, b, ϵ, and η be reals with $\epsilon^2 + \eta^2 = 1$ and $Ah = aBu + bB(\epsilon h' + \delta v)$. Here we require Bv to be orthogenal to both Bu and Bh and of length $\| Bv \| = \| Bu \|$ and we let

$$h' = (-\langle Ah, Bu\rangle u + \langle Bu, Bu\rangle h)(\| Bh \|^2 \| Bu \|^2 - \langle Bu, Bh\rangle^2)^{-1/2}.$$

We have constructed h' as a linear combination of u and h so that $\langle Bh', Bu\rangle = 0$ and $\| Bh' \| = \| Bu \|$. Proving (21) reduces therefore to deriving the nonnegativity of the polynomial $P(\epsilon)$ in ϵ which is achieved by substituting the above representation of Ah into (21). As before it is sufficient to consider the case $\epsilon = \pm 1$, i.e., $Ah = aBu + bBh'$ where now

$$\begin{aligned} | a/b | &= |\langle Ah, Bu\rangle|(\| Ah \|^2 \| Bu \|^2 - \langle Ah, Bu\rangle^2)^{-1/2} \\ &\leqslant D_\lambda. \end{aligned}$$

Letting

$$a_0 = a - \langle Ah, Bu\rangle(\| Ah \|^2 \| Bu \|^2 - \langle Ah, Bu\rangle^2)^{-1/2}.$$

and

$$b_0 = \| Bu \|^2 (\| Ah \|^2 \| Bu \|^2 - \langle Ah, Bu\rangle^2)^{-1/2} b,$$

then we have $Ah = a_0 Bu + b_0 Bh$. Calculation now shows that, as before,

$$\alpha P(\pm 1) = \alpha^2(8 - s)(b_0\alpha \pm a_0\beta)^2 + 4\alpha b_0(\beta a_0 \pm \alpha b_0),$$

which implies

$$(P(\pm 1)/\alpha b_0^2) \geqslant \alpha^2(8 - s)(1 - d)^2 - 4(1 - d)$$

where now $d = | a_0\beta/b_0\alpha |$. If $d \geqslant 1$, we are done. Otherwise, it suffices to show that

$$\alpha^2(8 - s) - 4 \geqslant \alpha^2(8 - s)d$$

or

$$\frac{a_0^2}{b_0^2} \leqslant \frac{(\alpha^2(8 - s) - 4)^2}{\alpha^2(8 - s)^2(1 - \alpha^2)}.$$

(c) now follows from observing

$$\begin{aligned} \left| \frac{a_0}{b_0} \right| &= \left| \frac{(\| Ah \|^2 \| Bu \|^2 - \langle Ah, Bu\rangle^2)^{1/2}}{\| Bu \|^2} \frac{a}{b} - \frac{\langle Ah, Bu\rangle}{\| Bu \|^2} \right| \\ &\leqslant \mathscr{O}_\lambda \end{aligned}$$

and noting (19). The theorem is proved.

3. Concluding Remarks

Although conclusions in the general environment of problem (1) are complex at best, the situation is entirely expected from the nature of the parameters that affect the "regions of attraction" about E_λ. Specifically, we must take note of the effects of the choice of s, the exclusion of other eigenvalues, and, if you will, the nearness of $B^{-1}A$ to symmetry and the nearness of B to orthogonality. The size restrictions on $\mathscr{B}_\delta$ that correspond to these four aspects are, roughly speaking, contained in (18), (20), (19), and, again, (19), respectively.

The results presented here can be posed in greater generality by including the complex, rectangular, and Hilbert space cases. The expected effort, however, is prohibitive.

References

1. M. Altman, An iterative method for eigenvalues and eigenvectors of matrices, *Bull. Acad. Polon. Sci.* **9** (1961), 639–644.
2. E. K. Blum, The computation of eigenvalues and eigenvectors of a completely continuous self-adjoint operator, *J. Comput. System Sci.* **1** (1967), 362–370.
3. E. K. Blum and G. H. Rodrigue, Solution of eigenvalue problems and least squares problems in Hilbert space by a gradient method, *J. Comput. System Sci.* **8** (1974), 220–238.
4. M. R. Hestenes and W. Karush, "Solutions of $Ax = \lambda Bx$," *J. Res. Nat. Bur. Standards* **47** (1951), 471–478.
5. G. Hufford, "On the computation of selected eigenvalues," *J. Analyse Math.* **16** (1966), 423–451.
6. S. F. McCormick, A general approach to one-step iterative methods with application to eigenvalue problems, *J. Comput. System Sci.* **6** (1972), 354–372.
7. S. F. McCormick, An iterative procedure for the solution of constrained nonlinear equations with application to optimization problems, *Numer. Math.* **23** (1975), 371–385.
8. S. F. McCormick, Ph. D. Thesis, University of Southern California, 1971.
9. G. H. Rodrigue, A gradient method for the matrix eigenvalue problem $Ax = \lambda Bx$, *Numer. Math.* **22** (1973), 1–16.