SIAM J. MATRIX ANAL. APPL.
Vol. 9, No. 4, October 1988

003

ON MINIMIZING THE SPECTRAL RADIUS OF A NONSYMMETRIC MATRIX FUNCTION: OPTIMALITY CONDITIONS AND DUALITY THEORY*

MICHAEL L. OVERTON† AND ROBERT S. WOMERSLEY‡

Abstract. Let $A(x)$ be a nonsymmetric real matrix affine function of a real parameter vector $x \in \mathscr{R}^m$, and let $\rho(x)$ be the spectral radius of $A(x)$. The article addresses the following question: Given $x_0 \in \mathscr{R}^m$, is $\rho(x)$ minimized locally at x_0, and, if not, is it possible to find a descent direction for $\rho(x)$ from x_0? If any of the eigenvalues of $A(x_0)$ that achieve the maximum modulus $\rho(x_0)$ are multiple, this question is not trivial to answer, since the eigenvalues are not differentiable at points where they coalesce. In the symmetric case, $A(x) = A(x)^T$ for all x, $\rho(x)$ is convex, and the question was resolved recently by Overton following work by Fletcher and using Rockafellar's theory of subgradients. In the nonsymmetric case $\rho(x)$ is neither convex nor Lipschitz, and neither the theory of subgradients nor Clarke's theory of generalized gradients is applicable. A new necessary and sufficient condition is given for $\rho(x)$ to have a first-order local minimum at x_0, assuming that all multiple eigenvalues of $A(x_0)$ that achieve the maximum modulus are nondefective. The optimality condition is computationally verifiable and involves computing "dual matrices." If the condition does not hold, the dual matrices provide information that leads to the generation of a descent direction. The result can be extended to the case where $\rho(x)$ is replaced by the maximum real part of the eigenvalues of $A(x)$. The authors use the eigenvalue perturbation theory of Rellich and Kato, which provides expressions for directional derivatives of $\rho(x)$. They also derive formulas for the codimension of manifolds on which certain eigenvalue structures of $A(x)$ are maintained; these are due to Von Neumann and Wigner and to Arnold. Finally, they discuss the much more difficult question of resolving optimality when $A(x_0)$ has a defective multiple eigenvalue achieving the maximum modulus $\rho(x_0)$.

Key words. nonsmooth optimization, nondifferentiable optimization, eigenvalue minimization, minimum spectral radius, nonconvex optimization

AMS(MOS) subject classifications. 65F99, 65K10, 90C25

1. Introduction.

Let $A(x)$ be a real $n \times n$ matrix affine function of $x = (\xi_1, \cdots, \xi_m)^T \in \mathscr{R}^m$, i.e.,

$$A(x) = A_0 + \sum_{k=1}^{m} \xi_k A_k, \tag{1.1}$$

where $\{A_k\}$ are given real $n \times n$ matrices. Define $\rho(x)$ to be the spectral radius of $A(x)$, i.e.,

$$\rho(x) = \max_{1 \leq i \leq n} |\lambda_i(x)|, \tag{1.2}$$

where $\lambda_i(x)$, $i = 1, \cdots, n$, are the (not necessarily distinct) eigenvalues of $A(x)$. Because $A(x)$ is real, the eigenvalues $\{\lambda_i(x)\}$ are either real or occur in complex conjugate pairs. In this paper we address the following question: Given $x_0 \in \mathscr{R}^m$, is $\rho(x)$ minimized locally by $x = x_0$, and if not, can we find a descent direction for ρ from x_0, that is, a direction $d \in \mathscr{R}^m$ such that $\rho(x_0 + \alpha d) < \rho(x_0)$ for sufficiently small $\alpha > 0$? There are several cases of increasing level of difficulty.

* Received by the editors August 20, 1987; accepted for publication (in revised form) February 1, 1988.

† Computer Science Department, Courant Institute of Mathematical Sciences, New York University, New York, New York 10012. This author's research was supported in part by National Science Foundation grant DCR-85-02014, and took place while he was on sabbatical leave at the Centre for Mathematical Analysis and Mathematical Sciences Research Institute, Australian National University, Canberra ACT 2601, Australia.

‡ School of Mathematics, University of New South Wales, Kensington, New South Wales 2033, Australia.

If $A(x_0)$ has real eigenvalues of distinct modulus, $\rho(x)$ is differentiable, indeed analytic, at x_0 (see Kato (1984, p. 64)). The question is therefore answered by examining $\nabla\rho(x_0)$ and $\nabla^2\rho(x_0)$. The same is true when complex conjugate pairs of eigenvalues, each pair having different modulus, are permitted. For example, let $n = 2$, $m = 1$, and define

$$A(x) = \begin{bmatrix} 1+\xi_1 & 1 \\ 1 & 1-\xi_1 \end{bmatrix}.$$

Then $\lambda_{1,2}(x) = 1 \pm \sqrt{1+\xi_1^2}$, and $\rho(x)$ is minimized at $x_0 = [0]$, where $\nabla\rho = 0$ and $\nabla^2\rho$ is positive.

If several eigenvalues, not complex conjugates of each other, achieve the maximum modulus at x_0 but each eigenvalue is distinct, then $\rho(x)$ is simply the pointwise maximum function of several differentiable functions, and may be analyzed by standard min-max theory (see, e.g., Fletcher (1981, p. 175)).

Example 1.1. Let $n = 2$, $m = 1$, and define

$$A(x) = \begin{bmatrix} 1+\xi_1 & 1 \\ -\xi_1 & 1+\xi_1 \end{bmatrix}.$$

The eigenvalues are

$$\lambda_{1,2}(x) = 1 + \xi_1 \pm \sqrt{-\xi_1}$$

and the spectral radius is given by

$$\rho(x) = \begin{cases} -1-\xi_1+\sqrt{-\xi_1} & \text{if } \xi_1 \leqq -1, \\ 1+\xi_1+\sqrt{-\xi_1} & \text{if } -1 \leqq \xi_1 \leqq 0, \\ \sqrt{\xi_1^2+3\xi_1+1} & \text{if } \xi_1 \geqq 0 \end{cases}$$

(see Fig. 1.1). We see that at $x = [\xi_1] = -1$, the eigenvalues $\lambda_1(x)$ and $\lambda_2(x)$ have the same modulus, although they are distinct. The function $\rho(x)$ is a standard "max function" here; in particular, it is Lipschitz. On the other hand, at $x = 0$, the eigenvalues $\lambda_1(x)$ and $\lambda_2(x)$ coalesce and $\rho(x)$ has a completely different, non-Lipschitz, character. In fact, A is defective, i.e., not diagonalizable, at $x = 0$, and we say that $\lambda_1(x) = \lambda_2(x)$ is a defective eigenvalue. In general, even if $A(x)$ is nondefective at $x = x_0$, $\rho(x)$ is not differentiable at x_0 if $A(x_0)$ has multiple eigenvalues, and cannot be analyzed by standard min-max theory.

Besides showing the very different character of the two local minima, Fig. 1.1 also shows that, as typical with nonconvex problems, several local minima may occur and finding a global minimum would be very difficult in general. The example also shows that it is possible for $\rho(x)$ to have a smooth local maximum, so that the condition $\nabla\rho(x_0) = 0$ is not sufficient for f to have a local minimum at x_0.

An example with $m > 1$ gives additional insight.

Example 1.2. Let $n = 2$, $m = 2$, and define

$$A(x) = \begin{bmatrix} 2+\xi_1 & \xi_2 \\ 2\xi_1 & 2+\xi_1+\xi_2 \end{bmatrix}.$$

Figure 1.2 shows a contour plot of $\rho(x)$. There is no unconstrained local minimum of ρ. At the origin $x = [0, 0]^T$, $A(x)$ has a nondefective eigenvalue of multiplicity 2. Along the two lines $\xi_2 = 0$ and $\xi_2 = -8\xi_1$, except at the origin, $A(x)$ has a defective eigenvalue of multiplicity 2. These two lines divide the (ξ_1, ξ_2) plane into four quadrants; the ei-

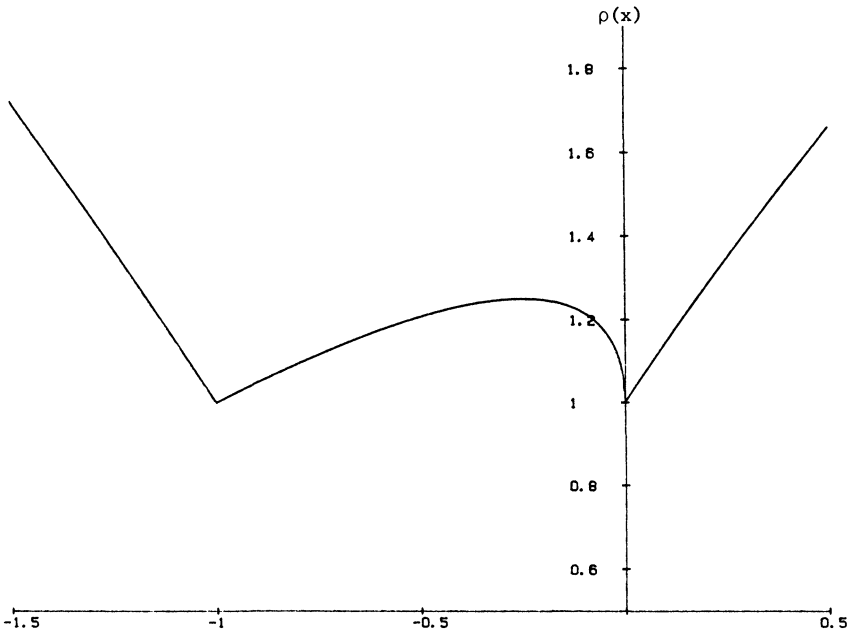


FIG. 1.1. *Plot of Example* 1.1.

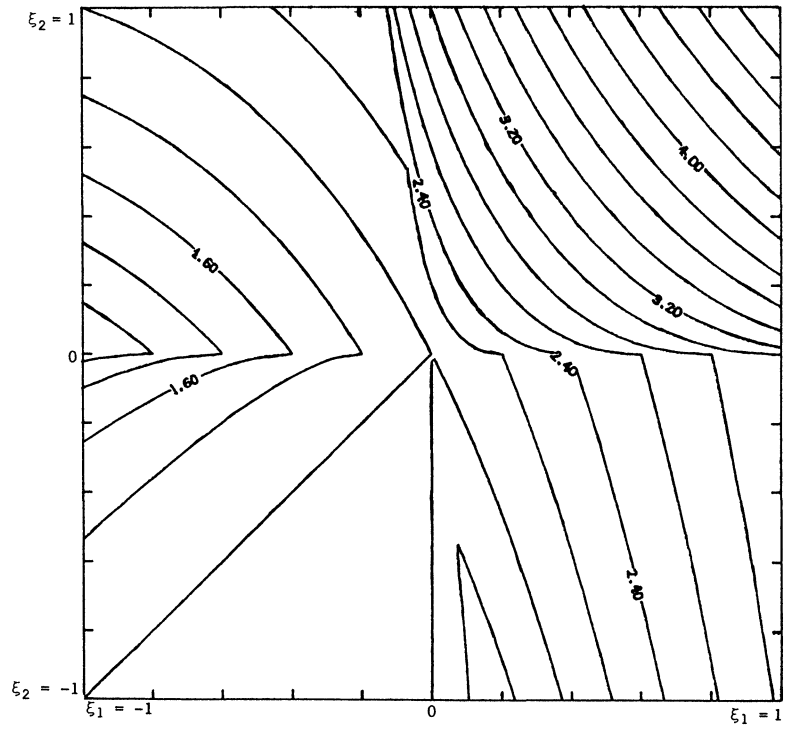


FIG. 1.2. *Contours of Example* 1.2.

genvalues are real and distinct in the top right and bottom left quadrants, and a complex conjugate pair in the other two quadrants. Note how the contours of ρ change sharply as they cross the defective manifold. This is because on the real side of the defective manifold, one of the eigenvalues is sharply increased by $O(\sqrt{\varepsilon})$ as the point x moves a distance ε away from the manifold, while on the complex side it is the imaginary part of the eigenvalue that is perturbed by $O(\sqrt{\varepsilon})$, which has only an $O(\varepsilon)$ effect on ρ. (The same effect is observed in Fig. 1.1 at $x = 0$.) Along lines passing through the origin, the function ρ is Lipschitz, but it is not Lipschitz along any other line in the (ξ_1, ξ_2) plane. Note that even if two vectors d_1 and d_2 are descent directions from the origin, a convex combination of d_1 and d_2 may be an ascent direction. We shall return to this phenomenon later.

Example 1.2 is not generic in the sense that a two-parameter family of matrices cannot be expected to have a nondefective multiple eigenvalue; this is explained in § 2. However, the example can be extended to three variables without changing its essential character by adding a term $\xi_3 A_3$ to $A(x)$. In that case the defective manifold becomes a cone instead of a pair of lines (see Arnold (1971, p. 40)). The eigenvalues of $A(x)$ are complex in the two disconnected "interior" parts of the cone and real elsewhere.

If $A(x) = A(x)^T$ for all x, $\rho(x)$ is a convex function and Rockafellar's theory of subgradients applies. In a recent paper, Overton (1988), following Fletcher (1985), has given verifiable optimality conditions for the symmetric case and shown how, if not optimal, a descent direction may always be obtained, even if this requires splitting a multiple eigenvalue. (There are exceptions in degenerate cases.) Both the optimality conditions and the method for obtaining descent directions involve an interesting duality theory. The same paper provides a practical, accurate algorithm for minimizing $\rho(x)$ in the symmetric case.

In the nonsymmetric case $\rho(x)$ is generally not convex and the problem is much more difficult. The main contribution of the present paper concerns the case where the (multiple) eigenvalues achieving the maximum modulus at x_0 are all nondefective. Even in this case, $\rho(x)$ is generally not Lipschitz at x_0, and hence the usual definition of the generalized gradient of Clarke (1975) is not applicable. However, the function ρ is Lipschitz at x_0 if its argument is restricted to the line $\{x_0 + \alpha d | \alpha \in \mathscr{R}\}$, for any $d \in \mathscr{R}^m$, and indeed the usual directional derivative of ρ (in the direction d) always exists. By considering this we are able to give a new necessary and sufficient condition for x_0 to be a local first-order minimizer of $\rho(x)$, excluding degenerate cases. The condition is computationally verifiable and involves computing "dual matrices." If the condition is found not to hold, the dual matrices are used to provide information that produces a descent direction, even if this requires splitting a multiple eigenvalue or making a multiple eigenvalue defective.

The paper is organized as follows. In the next section we derive formulas for the codimensions of manifolds defined by maintaining a given Jordan structure for $A(x)$. In the most general case, these formulas are due to Arnold (1971), (1983). In § 3 we characterize the directional derivative of $\rho(x)$. This derivation relies on the classic work of Kato and Rellich (see Kato (1984)). In § 4 we begin by summarizing the known optimality conditions for the symmetric case; we then derive new optimality conditions in the nonsymmetric case when only one multiple eigenvalue, which is nondefective at x_0, achieves the maximum modulus at x_0. In § 5 we extend this result to cover the case of several nondefective multiple eigenvalues achieving the same maximum modulus at x_0. In § 6 we briefly discuss the situation where a multiple eigenvalue achieving the maximum modulus is defective. The question of optimality seems very difficult to resolve in this case.

This paper is motivated by many applications. Perhaps the major source of applications is control engineering, where, for example, an optimal spectral radius value below

1 would represent system stability while a value greater than 1 would represent instability. See, for example, Mäkilä and Toivonen (1987) and Miller, Cochran, and Howze (1978) for applications where $A(x)$ is nonsymmetric; see Boyd (1988) and Kamenetskii and Pyatnitskii (1987) for applications where $A(x)$ is symmetric. Another source of applications is the design of iterative methods for solving linear systems of equations, where certain parameters must be chosen to minimize the spectral radius of the iteration matrix (see, for example, Young (1971)). The most well-known example is the SOR method, which depends on a single parameter ω whose optimal value is well known. More generally, we might consider a general preconditioner design problem. Since the latter application class generally involves nonlinear parameter dependence, the results of this paper cannot be applied directly. However, the results reported here will be an essential starting point for the analysis of problems where $A(x)$ is a nonlinear function. Other applications may involve constraints on the variables; it should be possible to extend the results given here to handling such constraints using standard Lagrange multiplier techniques.

As mentioned earlier, a practical algorithm is already available to minimize $\rho(x)$ in the symmetric case. We believe the results in this paper are an important first step towards the long-term goal of obtaining an efficient algorithm for the nonsymmetric case. There are many difficulties to be overcome before such a goal can be achieved. For example, even computing the Jordan form of $A(x)$ at a single point x is known to be a hard problem numerically, although there has been substantial progress in this direction in recent years (see Golub and Van Loan (1983) and Demmel (1983)).

It is important to note that the techniques used in this paper are also relevant to other functions of the eigenvalues $\lambda(x)$ besides the spectral radius. In fact, they could be used to analyze any real convex function of the eigenvalue function $\lambda(x) \in \mathscr{C}^m$. In our analysis of $\rho(x)$, we note that minimizing $\rho(x)$ is equivalent to minimizing

$$f(x) = \tfrac{1}{2}\rho(x)^2 = \tfrac{1}{2} \max_{1 \le i \le n} \lambda_i(x)\bar{\lambda}_i(x), \tag{1.3}$$

where $\bar{z}$ denotes the complex conjugate of $z \in \mathscr{C}$. Most of the analysis is then concerned with the nondifferentiable nature of $\lambda_i(x)$. Similarly, we can also consider minimizing another function that frequently arises in applications:

$$g(x) = \max_{1 \le i \le n} \operatorname{Re} \lambda_i(x) = \tfrac{1}{2} \max_{1 \le i \le n} (\lambda_i(x) + \bar{\lambda}_i(x)). \tag{1.4}$$

Of course, $\rho(x)$ and $g(x)$ are related to each other by exponential transformation of the matrix $A(x)$, but this is to be avoided numerically (Golub and Van Loan (1983)). In control engineering, for example, the form $g(x)$ arises when we consider stability of initial value problems; the form $\rho(x)$ arises when we consider discrete-time systems.

There is a large literature on extremal eigenvalue problems (see, for example, Nowasad (1968), Friedland (1978), and references therein). However, most of this work seems to be concerned with special problems that arise in infinite-dimensional spaces. The questions raised here do not seem to have been considered in detail previously.

2. The codimension of manifolds. An eigenvalue of multiplicity t is said to be nondefective (or semisimple) if the corresponding part of the Jordan form of the matrix is diagonal. Let x_0 be given, with $A(x_0)$ having a nondefective eigenvalue of multiplicity t, say $\lambda_1(x_0) = \cdots = \lambda_t(x_0)$. What, generically, is the codimension of the manifold containing x_0 on which $A(x)$ has a nondefective multiple eigenvalue $\lambda_1(x) = \cdots = \lambda_t(x)$? This question was answered in the symmetric case by von Neumann and Wigner (1929) and, in the context of requiring a matrix to have a given rank, by Ledermann (1937), although the answer does not seem to be widely known. More recently, the

symmetric case was discussed by Friedland, Nocedal, and Overton (1987) in the context of inverse eigenvalue problems. Arnold (1971), (1983) answers the question in the general complex nonsymmetric case, including the defective case when nontrivial Jordan blocks must be considered. In this section we motivate and summarize these results, which are essential for a complete understanding of the later sections. We do not give a rigorous derivation, for which the reader is referred to Arnold's work.

First assume that $\lambda_1(x_0) = \cdots = \lambda_t(x_0)$ is real and that the other eigenvalues of $A(x_0)$ are real and distinct. For x to lie in the desired manifold, we require

$$A(x)Q = Q\Lambda, \qquad \Lambda = \begin{bmatrix} \lambda I_t & \\ & \Lambda_2 \end{bmatrix}, \tag{2.1}$$

where Q is a nonsingular real matrix, I_t is the identity matrix of order t, and Λ_2 is a real diagonal matrix of order $n - t$. (None of the eigenvalues can become complex near enough to x_0 since the only multiple eigenvalue is being preserved.) We may view (2.1) as $h_1 = n^2$ equations that restrict x; but we have introduced additional variables Q and Λ. These variables are correctly counted as follows. There are $h_2 = n - t + 1$ variables in Λ. The matrix Q has $h_3 = n^2$ components, but not all n^2 degrees of freedom are useful in satisfying (2.1). Let $Q = [Q_1, Q_2]$, where the columns of Q_1 correspond to $\lambda_1(x) = \cdots = \lambda_t(x)$. We may postmultiply Q_1 by any nonsingular $t \times t$ matrix, and postmultiply Q_2 by any nonsingular diagonal matrix, without affecting (2.1). Let $h_4 = t^2$ and $h_5 = n - t$; therefore, we see that the total number of introduced variables useful in solving (2.1) is

$$h_2 + h_3 - h_4 - h_5 = n^2 - t^2 + 1.$$

The codimension of the desired manifold is obtained by subtracting this from h_1, the number of equations in (2.1), giving

$$c_N(t) = h_1 - h_2 - h_3 + h_4 + h_5 = t^2 - 1. \tag{2.2}$$

Since this manifold is embedded in $\mathscr{R}^m$, and the codimension describes the number of degrees of freedom restricted by requiring x to be in the manifold, the dimension of the manifold is $m + 1 - t^2$. For example, if $t = 2$ and $m = 3$, the dimension of the manifold is zero, i.e., a three-parameter matrix family $A(x)$ generically has only a single point x_0, where $A(x_0)$ has a nondefective multiple eigenvalue. Of course, this argument is generic and there are exceptions in degenerate cases.

A similar argument for the symmetric case ($A(x) = A(x)^T$ for all x) gives the Von Neumann–Wigner result $h_1 = n(n + 1)/2$, $h_2 = n - t + 1$, $h_3 = n(n - 1)/2$ (since Q is orthogonal), $h_4 = t(t - 1)/2$, $h_5 = 0$ (since Q is already restricted to being orthogonal by h_3), so

$$c_S(t) = \frac{t(t+1)}{2} - 1. \tag{2.3}$$

Von Neumann and Wigner also derived the codimension for the case that $A(x)$ is complex but Hermitian for all x, where we continue to view $A(x)$ as a function of *real* variables; thus (2.1) is n^2 real equations, namely $n(n - 1)/2$ complex off-diagonal equations and n real diagonal equations. We then obtain the same formula as (2.2).

Returning to the real nonsymmetric case, if $\lambda_1(x_0) = \cdots = \lambda_t(x_0)$ is real but we allow the other eigenvalues to be complex, the codimension (2.2) does not change. This is because Λ_2 and Q_2, although complex, consist of complex conjugate pairs.

If the multiple eigenvalue $\lambda_1(x_0) = \cdots = \lambda_t(x_0)$ is one of a complex conjugate pair, we require

$$A(x)Q = Q\Lambda, \qquad \Lambda = \begin{bmatrix} \lambda_1 I_t & & \\ & \bar{\lambda}_1 I_t & \\ & & \Lambda_2 \end{bmatrix},$$

where $Q = [Q_1, \bar{Q}_1, Q_2]$, and Λ_2 is a diagonal matrix of order $n - 2t$. We obtain $h_1 = n^2$, $h_2 = n - 2t + 2$, $h_3 = n^2$, $h_4 = 2t^2$, and $h_5 = n - 2t$, i.e.,

$$c_C(t) = 2t^2 - 2. \tag{2.4}$$

Thus the codimension is the same as if two real multiple eigenvalues, each of multiplicity t, were to be preserved separately.

Suppose we require $r + s$ nondefective multiple eigenvalues to have the same modulus, where r of them are real with respective multiplicities, $t_1, \cdots, t_r$, and s of them are complex with positive imaginary part and respective multiplicity $t_{r+1}, \cdots, t_{r+s}$. (Note $r \leqq 2$.) Then the codimension of the manifold along which multiplicities are preserved and all eigenvalues have the same modulus is

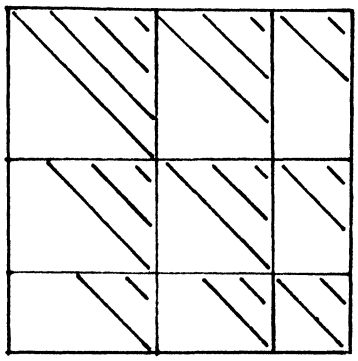
$$\begin{aligned} c_G(t_1, \cdots, t_r; t_{r+1}, \cdots, t_s) &= \sum_{j=1}^{r} (t_j^2 - 1) + \sum_{j=r+1}^{r+s} (2t_j^2 - 2) + (r + s - 1) \\ &= \sum_{j=1}^{r} t_j^2 + 2 \sum_{j=r+1}^{r+s} t_j^2 - s - 1, \end{aligned} \tag{2.5}$$

reflecting the fact that $(r + s - 1)$ additional restrictions are being placed on the moduli.

Now let us drop the assumption that $\lambda_1(x_0)$ is nondefective. Assume $A(x_0)$ has a real multiple eigenvalue $\lambda_1(x_0) = \cdots = \lambda_t(x_0)$, corresponding to Jordan blocks of size $u_1 \geqq u_2 \geqq \cdots \geqq u_p$, $1 \leqq p \leqq t$. We are interested in the dimension of the manifold passing through x_0 along which the same Jordan structure is maintained. For x to lie in the manifold, we require that

$$A(x)Q = QJ, \qquad J = \begin{bmatrix} J_1 & \\ & \Lambda_2 \end{bmatrix},$$

where $Q = [Q_1, Q_2]$ is any nonsingular matrix, Λ_2 is diagonal of order $n - t$, and J_1 is the desired Jordan form. We have $h_1 = n^2$, $h_2 = n - t + 1$, $h_3 = n^2$, and $h_5 = n - t$ as before. To determine h_4 we need to answer the following question: What class of matrices commute with J_1? If J_1 equals $\lambda_1 I$, the answer is all $t \times t$ matrices; if J_1 is a single Jordan block, the answer is all $t \times t$ upper triangular Toeplitz matrices. In general, the answer is given by Arnold (1971, p. 34), namely matrices of the following form:



Here the block partitioning conforms to the Jordan block partitioning of J_1, and each block is an upper triangular (rectangular) Toeplitz matrix. The example shown here corresponds to $u_1 = 4$, $u_2 = 3$, $u_3 = 2$. The number of variables in such a matrix is

$$h_4 = u_1 + 3u_2 + 5u_3 + \cdots + (2p+1)u_p.$$

We therefore obtain the codimension

$$c_D(u_1, \cdots, u_p) = u_1 + 3u_2 + 5u_3 + \cdots + (2p+1)u_p - 1. \tag{2.6}$$

Note that, as before, the codimension is independent of n. We have

$$c_D(1, 1, \cdots, 1) = t^2 - 1 = c_N(t)$$

and the codimension for a single Jordan block is

$$c_D(t) = t - 1. \tag{2.7}$$

The arguments given here are not rigorous; in particular we have not attempted to prove independence of the various restricting equations. For a full derivation, see Arnold (1971).

3. Directional derivatives. Let x_0 be given with $A(x_0)$ having a nondefective multiple eigenvalue $\lambda_1(x_0) = \cdots = \lambda_t(x_0)$. In general the eigenvalues $\lambda_i(x)$, $i = 1, \cdots, t$, are not Lipschitz functions even at x_0. For example, let

$$A(x) = \begin{bmatrix} 1 & \xi_1 \\ \xi_2 & 1 \end{bmatrix}$$

so that

$$\lambda_{1,2}(x) = 1 \pm \sqrt{\xi_1 \xi_2}.$$

Given any ball of radius $\varepsilon > 0$ around $x_0 = [0, 0]^T$, let $x_1 = [\varepsilon/\sqrt{2}, 0]^T$ and $x_2 = [\varepsilon/\sqrt{2}, \delta]^T$, where $\delta > 0$. Both x_1 and x_2 lie in the given ball if $\delta \leqq \varepsilon/\sqrt{2}$, but

$$|\lambda_i(x_1) - \lambda_i(x_2)|$$

cannot be bounded by $K\delta$ for any constant K independent of δ. This contradicts the definition of a Lipschitz function (which may be found in, e.g., Clarke (1983)). Of course, the eigenvalues $\lambda_i(x)$ are always continuous functions, regardless of x_0, provided a consistent ordering is used.

Although the eigenvalues $\lambda_i(x)$, $i = 1, \cdots, t$, are not Lipschitz with respect to several variables, they may be ordered so that they are locally continuously differentiable along any line passing through x_0. This follows from the classical eigenvalue perturbation theory of Rellich and Kato. Before stating the result let us introduce some notation. Let Q_1 be an $n \times t$ matrix whose columns are independent right eigenvectors of $A(x_0)$ corresponding to $\lambda_1(x_0) = \cdots = \lambda_t(x_0)$ and let P_1^T be a $t \times n$ matrix whose rows are corresponding independent left eigenvectors. We may normalize P_1 so that

$$P_1^T Q_1 = I_t \tag{3.1}$$

and we then have

$$P_1^T A(x_0) Q_1 = \lambda_1(x_0) I_t. \tag{3.2}$$

The quantity $Q_1 P_1^T$ is called the eigenprojection for $\lambda_1(x_0)$. Define the $t \times t$ matrices

$$B_k = P_1^T A_k Q_1, \qquad k = 1, \cdots, m, \tag{3.3}$$

where A_k is given by (1.1). Note that if $\lambda_1(x_0)$ is real, all of P_1, Q_1, and B_k, $k = 1, \cdots, m$, are also real, but if $\lambda_1(x_0)$ is complex, all these matrices generally will also be complex. If $\lambda_1(x_0)$ is complex, it has an associated complex conjugate multiple eigenvalue, with corresponding eigenvector matrices $\bar{P}_1$, $\bar{Q}_1$, and

$$\bar{B}_k = \bar{P}_1^T A_k \bar{Q}_1, \qquad k = 1, \cdots, m$$

corresponding to (3.3).

Now define the directional derivative of $\lambda_i(x)$ in the direction $d = [\delta_1, \cdots, \delta_m]^T \in \mathscr{R}^m$ by

$$\lambda_i'(x_0; d) = \lim_{\alpha \to 0^+} \frac{\lambda_i(x_0 + \alpha d) - \lambda_i(x_0)}{\alpha}. \tag{3.4}$$

LEMMA 3.1. *We have*

$$\lambda_i'(x_0; d) = \mu_i, \qquad i = 1, \cdots, t, \tag{3.5}$$

where $\{\mu_i\}$ *are the eigenvalues of*

$$B(d) = \sum_{k=1}^{m} \delta_k B_k. \tag{3.6}$$

Proof. See Kato (1984, p. 81) and preceding pages for the proof. Note that, although we assume that $\lambda_1(x_0) = \cdots = \lambda_t(x_0)$ is nondefective, we do not assume that $B(d)$ is nondefective.

Remark. It is useful to motivate the result as follows. Suppose for simplicity that $B(d)$ is nondefective, and let its eigensystem be

$$B(d) = Z D Y^T, \tag{3.7}$$

where Y, Z are nonsingular $t \times t$ matrices, $Y^T Z = I_t$, and D is diagonal with entries $\{\mu_i\}$. We have

$$Y^T P_1^T A(x_0 + \alpha d) Q_1 Z = \lambda_1(x_0) + \alpha D. \tag{3.8}$$

If $t = n$, this proves the lemma, since (3.8) gives the eigensystem of $A(x_0 + \alpha d)$, with linear eigenvalues $\lambda_1(x_0) + \alpha\mu_i$. On the other hand, if $t = 1$, the lemma is trivial since μ_1 is the inner product of d with the gradient of the differentiable function $\lambda_1(x)$, namely $[p_1^T A_1 q_1, \cdots, p_1^T A_m q_1]^T$. More generally, suppose that $1 < t < n$. Then (3.8) represents a generalized Rayleigh quotient, the key point being that the right-hand side is diagonal. Thus the diagonal entries approximate the first t eigenvalues of $A(x_0 + \alpha d)$, and the columns of $Q_1 Z$ (respectively, the rows of $Y^T P_1^T$) are the particular right (respectively, left) eigenvectors of $A(x_0)$ to which the right (left) eigenvectors of $A(x_0 + \alpha d)$ generally converge as $\alpha \to 0$. (If the $\{\mu_i\}$ are not distinct, the corresponding eigenvectors need not converge.)

Now let us turn to the functions $f(x)$ and $g(x)$ defined by (1.3) and (1.4); it is easier to work with $f(x) = \frac{1}{2}\rho(x)^2$ than directly with $\rho(x)$. Note that as long as $f'(x_0; d)$ exists with $f(x_0) \neq 0$, the quantity $\rho'(x_0; d)$ exists and is related by

$$\rho'(x_0; d) = \frac{f'(x_0; d)}{\rho(x_0)}.$$

LEMMA 3.2. *Suppose that* $\lambda_1(x_0) = \cdots = \lambda_t(x_0)$ *is a real nondefective eigenvalue, and that all other eigenvalues of* $A(x_0)$ *have smaller modulus than* $|\lambda_1(x_0)|$. *Then for any* $d \in \mathcal{R}^m$

$$f'(x_0; d) = \lambda_1(x_0) \max_{1 \leq i \leq t} \operatorname{Re} \mu_i,$$

where, as before, $\{\mu_i\}$ *are the eigenvalues of* $B(d)$.

Proof. It is clear that

$$f'(x_0; d) = \max_{1 \leq i \leq t} f_i'(x_0; d),$$

where

$$f_i(x_0; d) = \tfrac{1}{2}\lambda_i(x)\bar{\lambda}_i(x).$$

Now

$$f_i'(x_0; d) = \tfrac{1}{2}(\lambda_i(x_0)\bar{\lambda}_i'(x_0; d) + \bar{\lambda}_i(x_0)\lambda_i'(x_0; d)),$$

so the result follows from Lemma 3.1, since $\lambda_1(x_0)$ is real. □

LEMMA 3.3. *Suppose that* $\lambda_1(x_0) = \cdots = \lambda_t(x_0)$ *is a nondefective eigenvalue, and that all other eigenvalues of* $A(x_0)$ *have smaller real part. Then*

$$g'(x_0; d) = \max_{1 \leq i \leq t} \operatorname{Re} \mu_i,$$

where again $\{\mu_i\}$ *are the eigenvalues of* $B(d)$.

Proof. The proof is straightforward.

More generally, consider the case where several different eigenvalues achieve the maximum modulus or the maximum real part, respectively. It is convenient to change notation as follows. Let $\lambda_{jl}(x)$ denote the eigenvalues of $A(x)$ with the following properties:

(i) $\lambda_{j1}(x_0) = \cdots = \lambda_{jt_j}(x_0)$, for $j = 1, \cdots, r$, is a real nondefective multiple eigenvalue of $A(x_0)$ with multiplicity t_j.

(ii) $\lambda_{j1}(x_0) = \cdots = \lambda_{jt_j}(x_0)$, for $j = r+1, \cdots, r+s$, is a complex nondefective multiple eigenvalue of $A(x_0)$ with multiplicity t_j and positive imaginary part.

(iii) $\{\lambda_{j1}(x_0)\}$, $j = 1, \cdots, r+s$, are distinct quantities with, in the case of minimizing the spectral radius, the same modulus $\rho(x_0) = \sqrt{2f(x_0)}$, or, in the case of minimizing the maximum real part, the same real part $g(x_0)$. These eigenvalues are said to be *active*. The complex conjugates $\bar{\lambda}_{j1}(x_0)$, $j = r+1, \cdots, r+s$, are also active, so there are a total of $r + 2s$ distinct active eigenvalues. All other eigenvalues of $A(x_0)$ are inactive, i.e., they have smaller modulus or smaller real part, respectively. (Note that $r \leqq 2$.)

Now, for $j = 1, \cdots, r+s$, define Q_j, P_j^T as matrices whose columns (respectively, rows) are independent right (respectively, left) eigenvectors of $A(x_0)$ corresponding to $\lambda_{j1}(x_0) = \cdots = \lambda_{jt_j}(x_0)$, with $P_j^T Q_j = I_{t_j}$. Define the $t_j \times t_j$ matrix

$$B_k^{(j)} = P_j^T A_k Q_j, \quad k = 1, \cdots, m, \quad j = 1, \cdots, r+s. \tag{3.9}$$

LEMMA 3.4. *Let* $A(x_0)$ *have nondefective active eigenvalues with respect to the function* $f(x)$. *For any* $d = [\delta_1, \cdots, \delta_m]^T \in \mathcal{R}^m$,

$$f'(x_0; d) = \max_{1 \leq j \leq r+s} \max_{1 \leq l \leq t_j} \operatorname{Re}(\bar{\lambda}_{j1}(x_0)\mu_{jl}),$$

where μ_{jl}, $l = 1, \cdots, t_j$ *are the eigenvalues of*

$$B^{(j)}(d) = \sum_{k=1}^{m} \delta_k B_k^{(j)}. \tag{3.10}$$

Proof. It is clear that

$$f'(x_0;d) = \max_{1 \le j \le r+s} \max_{1 \le l \le t_j} f'_{jl}(x_0;d),$$

where

$$f_{jl}(x) = \tfrac{1}{2}\lambda_{jl}(x)\bar{\lambda}_{jl}(x).$$

Since

$$f'_{jl}(x_0;d) = \tfrac{1}{2}(\lambda_{jl}(x_0)\bar{\lambda}'_{jl}(x_0;d) + \bar{\lambda}_{jl}(x_0)\lambda'_{jl}(x_0;d)) \tag{3.11}$$

the result follows from Lemma 3.1. □

LEMMA 3.5. *Let $A(x_0)$ have nondefective active eigenvalues with respect to the function $g(x)$. For any $d \in \mathscr{R}^m$,*

$$g'(x_0;d) = \max_{1 \le j \le r+s} \max_{1 \le l \le t_j} \operatorname{Re} \mu_{jl},$$

where μ_{jl}, $l = 1, \cdots, t_j$, are the eigenvalues of $B^{(j)}(d)$ defined by (3.10).

Proof. The proof is straightforward.

We complete this section with the definition of a matrix inner product that will be needed in § 4. Following Fletcher (1985), define

$$A:B = \operatorname{tr} A^T B \tag{3.12}$$

for any real rectangular matrices A and B with the same dimension.

LEMMA 3.6. $XAY^T:B = A:X^TBY$.

Proof. The proof is straightforward.

4. Optimality conditions in the case of one active nondefective multiple eigenvalue. Assume that $A(x_0)$ has one active multiple eigenvalue that is real, nonzero, and nondefective, and that we denote by $\lambda_1(x_0) = \cdots = \lambda_t(x_0)$, reverting to our original notation. Let us define $d \in \mathscr{R}^m$ to be a descent direction for f from x_0 if $f'(x_0; d) < 0$. If no such direction exists, f is said to have a *first-order local minimum* at x_0. We wish to give a procedure for determining whether f has a first-order local minimum at x_0 and, if it does not, for obtaining a descent direction.

It is useful to first consider the symmetric case.

(1) Symmetric case ($A(x) = A(x)^T$ for all x).

In this case the eigenvalues $\lambda_i(x)$ are always real, the eigenvector matrix Q is orthogonal, $P_1 = Q_1$, and $B_k = Q_1^T A_k Q_1$. Furthermore, $f(x)$ and $\rho(x)$ are convex; this follows from Fletcher (1985, p. 510).

THEOREM 4.1. *Define the set*

$$\Omega = \{ v = [\nu_1, \cdots, \nu_m]^T \in \mathscr{R}^m \mid \text{there exists a symmetric positive semidefinite } t \times t \text{ matrix } U \text{ satisfying } \operatorname{tr} U = 1, \lambda_1(x_0)U:B_k = \nu_k, k = 1, \cdots, m \}.$$

(The matrix inner product operator ":" was defined by (3.12).) A necessary and sufficient condition for x_0 to minimize f is that $0 \in \Omega$.

Proof. Let $v \in \Omega$, let $d = [\delta_1, \cdots, \delta_m]^T \in \mathscr{R}^m$, and let the eigenvalue decomposition of the $t \times t$ symmetric matrix $B(d) = \sum_{k=1}^m \delta_k B_k$ be given by $B = ZMZ^T$, where Z is orthogonal and $M = \operatorname{diag}(\mu_i)$. We have

$$v^T d = \lambda_1(x_0) \sum_{k=1}^m \delta_k U:B_k$$

$$= \lambda_1(x_0) U:ZMZ^T.$$

Therefore

$$\sup_{v \in \Omega} v^T d = \lambda_1(x_0) \sup_U U : Z M Z^T, \tag{4.1}$$

where the second "sup" is taken over all $t \times t$ symmetric positive semidefinite matrices U with tr $U = 1$. Since Z is orthogonal and U is symmetric, without loss of generality we may write (4.1) as

$$\lambda_1(x_0) \sup_U U : M = \lambda_1(x_0) \sup_U \sum_{i=1}^{m} U_{ii} \mu_i \tag{4.2}$$

(see Lemma 3.6). Now U cannot have negative diagonal elements, and it has trace equal to one, so we see from (4.1), (4.2) that

$$\begin{aligned} \sup_{v \in \Omega} v^T d &= \lambda_1(x_0) \max_{1 \leq i \leq t} \mu_i \\ &= f'(x_0; d) \end{aligned} \tag{4.3}$$

by Lemma 3.2. It follows that if $0 \in \Omega$,

$$f'(x_0; d) \geqq 0 \quad \forall d \in \mathscr{R}^m,$$

i.e., x_0 minimizes f. On the other hand, if $0 \notin \Omega$, then by the separating hyperplane theorem and the convexity of Ω, there exists d with $v^T d < 0$ for all $v \in \Omega$, i.e., d is a descent direction by (4.3). (For a statement of the separating hyperplane theorem, see Rockafellar (1970, p. 95).)

Remark. This theorem was proved in Overton (1988). The proof here is more direct, since it does not use Rockafellar's theory of subgradients, but only the separating hyperplane theorem. Nonetheless, the proof technique is similar to those used in the theory of subgradients, and it is doubtful whether the theorem would have been obtained without the motivation of that theory (and also the paper of Fletcher (1985)).

COROLLARY. $\Omega = \partial f(x_0)$, *the subdifferential of the convex function f as defined by Rockafellar* (1970).

Proof. The proof follows from (4.3).

Remark. Because f is convex, there is no distinction between "first-order local minimum" and "minimum."

Remark. The matrix U is called the dual matrix (or Lagrange matrix), and it plays the role of Lagrange multipliers familiar from constrained optimization.

We now discuss the generation of descent directions if x_0 is not optimal. There are three cases.

(1A) Symmetric case, assuming $I_t \in \text{Span}\{B_1, \cdots, B_m\}$.

In this case we simply solve

$$\lambda_1(x_0) \sum_{k=1}^{m} \delta_k B_k = -I_t. \tag{4.4}$$

By Lemma 3.2, $d = [\delta_1, \cdots, \delta_m]^T$ is a descent direction for f. Furthermore, all the eigenvalues $\lambda_1(x), \cdots, \lambda_t(x)$ decrease at the same rate along d; that is, the eigenvalue does not split to first order. This case holds generically if $m \geqq t(t+1)/2$, i.e., $m > c_S(t)$, i.e., the generic dimension of the manifold defined by

$$\lambda_1(x) = \cdots = \lambda_t(x) \tag{4.5}$$

is greater than zero.

(1B) Symmetric case, assuming (1A) does not apply and the set $\{I_t, B_1, \cdots, B_m\}$ has full rank $t(t+1)/2$.

This case holds generically when $m = c_S(t) = t(t+1)/2 - 1$, i.e., the manifold defined by (4.5) is the single point x_0. It also holds if $m > c_S(t)$, but f is minimized on the manifold (4.5) at x_0. To make further progress we must split the multiple eigenvalue λ_1. Solve for the dual matrix $U = U^T$ in the linear system

$$\text{tr } U = 1, \quad \lambda_1(x_0) U : B_k = 0, \quad k = 1, \cdots, m. \tag{4.6}$$

This is a system of $m+1$ equations in $t(t+1)/2$ unknowns. Although it is possible that the $\{B_k\}$ are not independent, (4.6) has a unique solution U in view of the homogeneity of all equations except the trace equation (which is equivalent to $I_t : U = 1$). If $0 \notin \Omega$, i.e., x_0 is not optimal, it follows that U is not positive semidefinite.

THEOREM 4.2. *Assume* $0 \notin \Omega$, *so that* U *has an eigenvalue* $\theta < 0$. *Let* $z \in \mathcal{R}^t$ *be a corresponding normalized eigenvector of* U. *Solve for* $[\delta_0, \delta_1, \cdots, \delta_m]^T \in \mathcal{R}^{m+1}$ *in*

$$\delta_0 I_t + \lambda_1(x_0) \sum_{k=1}^{m} \delta_k B_k = -zz^T. \tag{4.7}$$

Then $d = [\delta_1, \cdots, \delta_m]^T$ *is a descent direction*.

Proof. The linear system (4.7) is solvable by assumption, although if $\{B_k\}$ are not independent, d is not unique. Taking an inner product of U with (4.7) we obtain

$$\delta_0 \text{ tr } U + \lambda_1(x_0) \sum_{k=1}^{m} \delta_k U : B_k = -U : zz^T,$$

i.e.,

$$\delta_0 = -\theta > 0$$

by (4.6). From (4.7) and Lemma 3.2, $f'(x_0; d)$ is the maximum eigenvalue of the symmetric matrix $-zz^T - \delta_0 I_t$. The eigenvalues of this matrix are $(-1 + \theta, \theta, \cdots, \theta)$, so $f'(x_0; d) < 0$.

Remark. This theorem was given by Overton (1988). The proof here is slightly different. The theorem shows that we can progress by splitting the multiple eigenvalue while maintaining multiplicity $t - 1$ (to first order). This is analogous to moving off a single active constraint in the context of constrained optimization. Note that it is the dual matrix U that provides information leading to a descent direction, just as negative Lagrange multipliers provide similar information in constrained optimization. Note in particular that the coefficient matrix of the left-hand side of the linear system (4.6), which defines the dual matrix, is the transpose of the coefficient matrix of the linear system (4.7), which gives the descent direction.

(1C) Symmetric case, where neither (1A) nor (1B) applies.

Although this applies generally if $m < c_S(t)$, such cases are degenerate in the sense that, generically, a point x_0 satisfying (4.5) will not exist. In such degenerate situations, verifying optimality or finding a descent direction is very difficult, just as it is in the much simpler case of linear programming. We may be able to solve (4.6), but the dual matrix U is not uniquely defined and generally (4.7) will not be solvable. Theorem 4.1 still applies, so x_0 is optimal if and only if there exists a dual matrix U with the required properties. However, because the solution to (4.6) is not unique, finding such a matrix U may be very difficult.

We now turn to the nonsymmetric problem. We first dispose of the trivial case.

(2A) Nonsymmetric case, assuming $I_t \in \text{Span}\{B_1, \cdots, B_m\}$.

A descent direction is obtained by solving (4.4). The eigenvalue is not split (to first order). This case holds generically if $m > c_N(t) = t^2 - 1$.

(2B) Nonsymmetric case, assuming (2A) does not hold and the set $\{I_t, B_1, \cdots, B_m\}$ has full rank t^2.

This case holds generically when $m = c_N(t)$, i.e., the manifold defined by maintaining the nondefective multiple eigenvalue is the single point x_0. It also holds if $m > c_N(t)$, but f is minimized on the manifold at x_0. To make further progress we must either split the multiple eigenvalue λ_1 or make it defective.

Our initial work on this problem involved the following set, intended to generalize the subdifferential Ω to the nonconvex case. Define the (nonconvex) set Ψ by

$$\Psi = \{v = [v_1, \cdots, v_m]^T \in \mathscr{R}^m \mid \text{there exists a real } t \times t \text{ diagonalizable matrix } U \text{ with real nonnegative eigenvalues satisfying tr } U = 1, \lambda_1(x_0)U : B_k = v_k, k = 1, \cdots, m\}.$$

However, it is not the case that (4.3) holds when we substitute Ψ for Ω on the left-hand side. On the contrary,

$$\sup_{v \in \Psi} v^T d = \infty.$$

The point where the proof of Theorem 4.1 breaks down in the nonsymmetric case is that U can have negative diagonal elements, even though it is similar to a nonnegative diagonal matrix with trace equal to 1.

Nonetheless, it is true that $0 \in \Psi$ is a necessary condition for x_0 to minimize f. A weaker result which is easier to show, following the lines of Theorem 4.1, is that $0 \in \text{Conv } \Psi$ is a necessary condition for optimality, but this is of no interest since it turns out that $\text{Conv } \Psi = \mathscr{R}^m$. We note that if we were to apply the usual definition of Clarke's generalized gradient (Clarke (1983, p. 10)) to f, ignoring the fact that f is not Lipschitz, we would obtain $\partial f(x_0) = \Re^m$. Rockafellar has extended the definition of the generalized gradient to the non-Lipschitz case, but this apparently still gives $\partial f(x_0) = \Re^m$ for our function f (Rockafellar (1985), Burke (1987)).

It may be worth noting at this point that there cannot exist any set $\tilde{\Psi}$, convex or not, such that

$$\sup_{v \in \tilde{\Psi}} v^T d = f'(x_0; d) \quad \text{for all } d \in \mathscr{R}^m.$$

The existence of such a set would contradict the possibility of the existence of descent directions whose convex combination is an ascent direction, which was noted in Example 1.2.

To show that $0 \in \Psi$ is a necessary condition for x_0 to minimize f, first observe that, as in case (1B), the linear system (4.6) is solvable, although since the matrices are nonsymmetric, it is now a system of $m + 1$ equations in t^2 unknowns, namely the elements of the dual matrix U. If U has a negative real eigenvalue, we can obtain a descent direction by solving (4.7), replacing the right-hand side by yz^T, where z and y^T are, respectively, right and left eigenvectors for the negative eigenvalue of U. If U has complex eigenvalues or is defective, we can also find a descent direction by appropriate choice of the right-hand side of (4.7). In view of the subsequent remarks, there is no need to elaborate on this further.

We now show that the set Ψ is too large to be useful and that a necessary and sufficient optimality condition can be obtained from using a smaller set. Define

$$\Phi = \{ v = [v_1, \cdots, v_m]^T \in \mathscr{R}^m \mid U = \frac{1}{t} I_t, \lambda_1(x_0) U : B_k = v_k, k = 1, \cdots, m \},$$

i.e., Φ consists of the single point $v = (\lambda_1(x_0)/t)[\operatorname{tr} B_1, \cdots, \operatorname{tr} B_m]^T$.

THEOREM 4.3. *A necessary and sufficient condition for f to have a first-order local minimum at x_0 is that $0 \in \Phi$.*

Remark. The theorem does not require the assumption that $\lambda_1(x_0) \neq 0$. However, it is convenient to assume throughout that $\lambda_1(x_0) \neq 0$, as stated at the beginning of the section, so that (4.6) remains solvable. With this assumption, $0 \in \Phi \Leftrightarrow \operatorname{tr} B_k = 0$, $k = 1, \cdots, m$.

Proof. Define U by solving (4.6). The theorem states that $U = (1/t)I_t$ if and only if f has a first-order local minimum at x_0. First suppose that $U = (1/t)I_t$, and suppose also that x_0 is not a first-order local minimizer, i.e., there exists a descent direction $d \in \mathscr{R}^m$. By Lemma 3.2, this implies that

$$\lambda_1(x_0) \operatorname{Re} \mu_i < 0, \qquad i = 1, \cdots, t,$$

where μ_i are the eigenvalues of $B(d)$. Because $\lambda_1(x_0)$ is real, $B(d)$ is also real, so this implies $\operatorname{tr} \lambda_1(x_0) B(d) < 0$. However, this is a contradiction, since $U = (1/t)I_t$ implies $\lambda_1(x_0) \operatorname{tr} B_k = 0$, $k = 1, \cdots, m$.

Now suppose that f has a first-order local minimum at x_0, but that $U \neq (1/t)I_t$. The latter assumption implies that there exists a $t \times t$ real matrix E with zero eigenvalues such that $U : E \neq 0$, namely, one of the following $t^2 - 1$ linearly independent defective matrices:

$$e_p e_q^T, \quad p, q = 1, \cdots, t, \quad p \neq q$$

or

$$e_p e_p^T - e_{p+1} e_{p+1}^T + e_p e_{p+1}^T - e_{p+1} e_p^T, \qquad p = 1, \cdots, t-1.$$

Here e_p denotes the pth column of I_t. Now solve the following linear system for $[\delta_0, \delta_1, \cdots, \delta_m]^T = [\delta_0, d^T]^T \in \mathscr{R}^{m+1}$:

$$\delta_0 I_t + \lambda_1(x_0) \sum_{k=1}^{m} \delta_k B_k = E. \tag{4.8}$$

(This system is a nonsymmetric version of (4.7), and therefore the coefficient matrix of the left-hand side is the transpose of that in the nonsymmetric version of (4.6).) Taking an inner product of U with (4.8), we get

$$\delta_0 = U : E \neq 0. \tag{4.9}$$

But by (4.8) and Lemma 3.2, $f'(x_0; d)$ is the largest real part of the eigenvalues of $E - \delta_0 I_t$, i.e., $-\delta_0$. This contradicts the assumption that a descent direction does not exist, since if $\delta_0 < 0$ we may replace $[\delta_0, d^T]^T$ by $-[\delta_0, d^T]^T$. □

Any direction d that preserves the multiple eigenvalue $\lambda_1 = \cdots = \lambda_t$ (to first order) by making it defective (to first order) has the property that $f'(x_0; -d) = -f'(x_0; d)$, since all the active eigenvalues have the same first-order charge. It follows that either d or $-d$ is a descent direction unless the first-order charge is zero; Theorem 4.3 states that this happens for all such "defective" directions if and only if $U = (1/t)I_t$. An example of this is the following.

Example 4.1. Let $n = 2$, $m = 3$, and define

$$A(x) = \begin{bmatrix} 1+\xi_3 & \xi_1 \\ \xi_2 & 1-\xi_3 \end{bmatrix}.$$

The eigenvalues are

$$\lambda_{1,2} = 1 \pm \sqrt{\xi_3^2 + \xi_1\xi_2}.$$

At the origin, $\lambda_1 = \lambda_2$ is nondefective (with value 1) and we may take $P_1 = Q_1 = I$. Thus $B_k = A_k$, $k = 1, 2, 3$, and $\operatorname{tr} B_k = \operatorname{tr} A_k = 0$, so U, defined by (4.6), is $\frac{1}{2}I$. The spectral radius $\rho(x)$ is one at every point on the manifold where $\lambda_1 = \lambda_2$ is defective. Figure 4.1 shows a contour plot of $\rho(x)$ restricted to the (ξ_1, ξ_2) plane, where the defective manifold reduces to the coordinate axes.

COROLLARY. *There is always a direction d satisfying $f'(x_0; d) \leqq 0$, i.e., f never has a strongly unique local minimum at x_0.*

Proof. The proof is straightforward.

From both a practical and a theoretical point of view, obtaining a descent direction by making the active eigenvalue defective to first order is far from satisfactory. Because defective eigenvalues are very ill-conditioned, roundoff error may be overwhelming. Even in exact arithmetic, it is possible that a very small stepsize α may be required to make $f(x_0 + \alpha d) < f(x_0)$. In any case, finding the next descent direction to further reduce f may be very difficult, as explained in § 6. The following theorem greatly improves the situation.

THEOREM 4.4. *Suppose that $0 \notin \Phi$, i.e., f does not have a first-order local minimum at x_0 and therefore U, defined by (4.6), is not equal to $(1/t)I_t$. Then there exists a descent direction d along which $\lambda_1 = \cdots = \lambda_t$ is split into several nondefective eigenvalues. All eigenvalues maintain a common real part to first order, but they may have several different imaginary parts.*

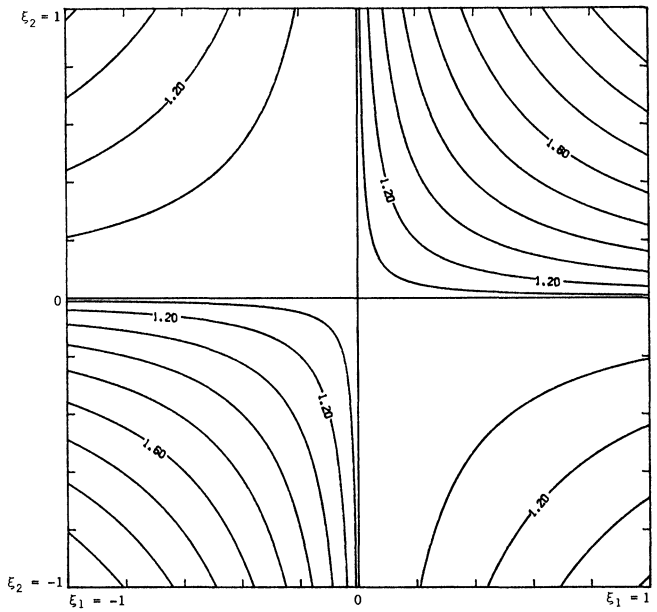


FIG. 4.1. *Contours of Example* 4.1 *in* $\xi_3 = 0$ *plane.*

Proof. Since $U \neq (1/t)I_t$, there exists a matrix E with *imaginary* eigenvalues such that $U:E \neq 0$, namely one of the following $t^2 - 1$ linearly independent matrices:

$$2e_pe_q^T - e_qe_p^T, \quad p,q = 1, \cdots, t, \quad p \neq q$$

or

$$e_pe_p^T - e_{p+1}e_{p+1}^T + 2e_pe_{p+1}^T - e_{p+1}e_p^T, \qquad p = 1, \cdots, t-1.$$

Now solve (4.8) for $[\delta_0, d^T]^T$, using the new right-hand side matrix E. As before, we obtain (4.9). Also as before, $f'(x_0; d)$ is the largest real part of the eigenvalues of $E - \delta_0 I_t$, i.e., $-\delta_0$, since E has imaginary eigenvalues. Thus a descent direction is obtained with the required property, since d may be replaced by $-d$ if $\delta_0 < 0$. Note that, to first order, multiplicity $t - 2$ is maintained along d, the common value being reduced by δ_0, while the other two eigenvalues split into a complex conjugate pair. It may be possible to split λ_1 further, with several eigenvalues taking on several different imaginary parts to first order, by choosing a less elementary matrix E with several different imaginary eigenvalues for the right-hand side of (4.8). □

The following question might arise: Can we obtain a descent direction along which $\lambda_1 = \cdots = \lambda_t$ is split into several distinct real eigenvalues? Obtaining such a descent direction d is much more difficult, since it is not true that $f'(x_0; -d) = -f'(x_0; d)$. If U has a negative real eigenvalue, such a descent direction may be obtained by using yz^T on the right-hand side of (4.8), where y^T, z are the left and right eigenvectors corresponding to the negative eigenvalue of U, as already explained. However, we have observed examples where there exists such a descent direction even if U has no negative eigenvalue.

Other examples have led us to the following conjecture that might be of interest.

CONJECTURE. Assume $n = 2$, $m = 3$, $\lambda_1(x_0) = \lambda_2(x_0)$ is nondefective, and $\{I_t, B_1, B_2, B_3\}$ has full rank. Then U has real eigenvalues if and only if there exist descent directions in *both* of the disconnected regions where $\lambda_{1,2}$ splits into a complex conjugate pair.

Remark. When there are descent directions in both of these disconnected regions, a convex combination of descent directions can give an ascent direction, namely in the region where $\lambda_{1,2}$ splits into a distinct real pair.

In the case $n = 2$, $t = 2$, $m = 3$, it is usually easy to find a descent direction by random search, since we need only that $\lambda_1(x_0)$ max Re $\mu_i < 0$, $i = 1, 2$. However, for larger t, the chance of finding a descent direction rapidly diminishes. In some randomly generated tests, we found that it was usually possible to obtain a descent direction with less than 500 random attempts for $n = t = 6$, $m = 35$, but this was not usually possible for $n = t = 8$, $m = 63$. Presumably the chance of success decreases exponentially with t.

We have now completed the discussion of case (2B). The degenerate case remains.

(2C) Nonsymmetric case, where neither (2A) nor (2B) applies.

This case generally applies if $m < c_N(t) = t^2 - 1$. As in case (1C), such situations are degenerate. Unlike the symmetric case, the nonsymmetric case no longer has an applicable optimality condition.

5. Optimality conditions in the case of several active nondefective multiple eigenvalues. Assume that $A(x_0)$ has several distinct active eigenvalues, all nondefective and with nonzero common modulus. Denote those that are real by λ_{jl}, $j = 1, \cdots, r$, and those that have positive imaginary parts by λ_{jl}, $j = r + 1, \cdots, s$, as described in the latter part of § 3. Recall that $\lambda_{j1} = \cdots = \lambda_{jt_j}$ is a multiple eigenvalue of multiplicity t_j, and recall the definition of $B_k^{(j)}$ given by (3.9). We now wish to generalize the results of the previous section.

(1) Symmetric case. This is easily generalized, since $r \leqq 2$ and $s = 0$. Details may be found in Overton (1988).

(2) Nonsymmetric case. To avoid confusing notation we entitle the three cases (A), (B), (C) somewhat differently than in § 4.

(2A) Nonsymmetric case, where we can obtain a descent direction without splitting a multiple eigenvalue or making it defective, or separating moduli.

For this case to apply, assume that the following linear system is solvable for $[\delta_1, \cdots, \delta_m, \varepsilon_{r+1}, \cdots, \varepsilon_{r+s}]^T \in \mathscr{R}^{m+s}$:

$$\sum_{k=1}^{m} \delta_k \lambda_{j1}(x_0) B_k^{(j)} = -I_{t_j}, \qquad j = 1, \cdots, r, \tag{5.1}$$

$$\sum_{k=1}^{m} \delta_k \operatorname{Re}(\bar{\lambda}_{j1}(x_0) B_k^{(j)}) = -I_{t_j}, \qquad j = r+1, \cdots, r+s, \tag{5.2}$$

$$\varepsilon_j I_{t_j} + \sum_{k=1}^{m} \delta_k \operatorname{Im}(\bar{\lambda}_{j1}(x_0) B_k^{(j)}) = 0, \qquad j = r+1, \cdots, r+s. \tag{5.3}$$

The system is generically solvable if $m > c_G(t_1, \cdots, t_{r+s})$, which is given by (2.5). Since we do not use the index i in this section, let $i = \sqrt{-1}$. Adding (5.2) to i times (5.3) we get

$$\sum_{k=1}^{m} \delta_k \bar{\lambda}_{j1}(x_0) B_k^{(j)} = -(1 + \varepsilon_j i) I_{t_j}, \qquad j = r+1, \cdots, r+s.$$

From Lemma 3.1, the first-order changes in the eigenvalue λ_{jl}, $l = 1, \cdots, t_j$, along the direction $d = \{\delta_1, \cdots, \delta_m\}^T$, are thus all the same quantity $-(1 + \varepsilon_j i)/\bar{\lambda}_{j1}(x_0)$, for each $j = r+1, \cdots, r+s$. Similarly, by (5.1), the first-order charges in λ_{jl}, $l = 1, \cdots, t_j$, are all $-1/\lambda_{j1}(x_0)$, for each $j = 1, \cdots, r$. Thus all multiple eigenvalues are preserved. Furthermore, by Lemma 3.4, or more specifically (3.11), the first-order change in $f_{jl} = \frac{1}{2}|\lambda_{jl}|^2$ is -1 for all $l = 1, \cdots, t_j$, $j = 1, \cdots, r+s$, i.e., all moduli are reduced along d and remain equal to first order.

(2B) Nonsymmetric case, where we can obtain a descent direction by splitting a multiple eigenvalue or making it defective or separating moduli, or else demonstrate optimality.

For this case to apply, assume that the coefficient matrix of the left-hand side of the following linear system has full column rank, and that the system is solvable. This case applies generically if $m = c_G(t_1, \cdots, t_{r+s})$. It also applies if $m > c_G(t_1, \cdots, t_{r+s})$, but f is minimized at x_0 on the manifold that preserves the nondefective multiplicities and the equal moduli. The linear system defines square dual matrices, $U_1, \cdots, U_{r+s}$, $V_{r+1}, \cdots, V_{r+s}$, of dimension $t_1, \cdots, t_{r+s}, t_{r+1}, \cdots, t_{r+s}$, respectively, by

$$\sum_{j=1}^{r} U_j : \lambda_{j1}(x_0) B_k^{(j)} + \sum_{j=r+1}^{r+s} U_j : \operatorname{Re}(\bar{\lambda}_{j1}(x_0) B_k^{(j)}) + \sum_{j=r+1}^{r+s} V_j : \operatorname{Im}(\bar{\lambda}_{j1}(x_0) B_k^{(j)}) = 0, \tag{5.4}$$

$$k = 1, \cdots, m,$$

$$\sum_{j=1}^{r+s} \operatorname{tr} U_j = 1, \tag{5.5}$$

$$\operatorname{tr} V_j = 0, \qquad j = r+1, \cdots, r+s. \tag{5.6}$$

The system (5.4)–(5.6) consists of $m + s + 1$ equations in $t_1^2 + \cdots + t_r^2 + 2t_{r+1}^2 + \cdots + 2t_{r+s}^2$ unknowns, so that it is square if $m = c_G(t_1, \cdots, t_{r+s})$.

THEOREM 5.1. *Define the dual matrices by* (5.4)–(5.6). *Then f has a first-order local minimum at x_0 if and only if $U_j = \kappa_j I_{t_j}$, where κ_j is a nonnegative real number, $j = 1, \cdots, r + s$, and $V_j = 0, j = r + 1, \cdots, r + s$.*

Proof. First suppose that f does not have a first-order local minimum at x_0 and assume the given condition on the dual matrices holds. Let d be a descent direction for f from x_0. Then by Lemma 3.4,

$$\operatorname{Re}(\bar{\lambda}_{j1}(x_0)\mu_{jl}) < 0, \quad l = 1, \cdots, t_j, \quad j = 1, \cdots, r+s,$$

where $\{\mu_{jl}\}$ are the eigenvalues of $B^{(j)}(d)$, defined by (3.10). It follows that

$$\sum_{j=1}^{r+s} \kappa_j \operatorname{Re}(\operatorname{tr}(\bar{\lambda}_{j1}(x_0)B^{(j)}(d))) < 0.$$

(Note that $\sum_{j=0}^{r+s} \kappa_j t_j = 1$ by (5.5), so not all the $\{\kappa_j\}$ are zero.) Therefore, since the trace is the sum of diagonal elements,

$$\sum_{j=1}^{r+s} \kappa_j \operatorname{tr}(\operatorname{Re}(\bar{\lambda}_{j1}(x_0)B^{(j)}(d))) < 0.$$

But from (5.4), using the facts that $U_j = \kappa_j I_{t_j}$ and $V_j = 0$, and that $\bar{\lambda}_{j1}(x_0)B_k^{(j)}$ is real for $j = 1, \cdots, r$, we have

$$\sum_{j=1}^{r+s} \kappa_j \operatorname{tr}(\operatorname{Re}(\bar{\lambda}_{j1}(x_0)B_k^{(j)})) = 0, \qquad k = 1, \cdots, m.$$

By (3.10), this is a contradiction.

Now suppose that the given condition on the dual matrices does not hold. We wish to show that there exists a descent direction. Solve the following linear system in $[\delta_0, \delta_1, \cdots, \delta_m, \varepsilon_{r+1}, \cdots, \varepsilon_{r+s}] \in \mathscr{R}^{m+s+1}$:

$$\delta_0 I_{t_j} + \sum_{k=1}^{m} \delta_k \lambda_{j1}(x_0)B_k^{(j)} = E_j, \qquad j = 1, \cdots, r, \tag{5.7}$$

$$\delta_0 I_{t_j} + \sum_{k=1}^{m} \delta_k \operatorname{Re}(\bar{\lambda}_{j1}(x_0)B_k^{(j)}) = E_j, \qquad j = r+1, \cdots, r+s, \tag{5.8}$$

$$\varepsilon_j I_{t_j} + \sum_{k=1}^{m} \delta_k \operatorname{Im}(\bar{\lambda}_{j1}(x_0)B_k^{(j)}) = F_j, \qquad j = r+1, \cdots, r+s, \tag{5.9}$$

where the right-hand sides $\{E_j, F_j\}$ will now be defined. First note that the coefficient matrix of the left-hand side has full row rank, since it is the transpose of the coefficient matrix of the system (5.4)–(5.6), which defines the dual matrices. Now define all right-hand side matrices $\{E_j, F_j\}$ to be zero except one, namely E_h or F_h, which is to be defined by the first applicable case from the following list. At least one case must apply by assumption.

(i) Set $E_h = e_p e_q^T$ if there is a dual matrix U_h with a nonzero element in the (p, q) position, with $p \neq q$. Here e_p denotes the pth column of I_{t_h}.

(ii) Set $F_h = e_p e_q^T$ if there is a dual matrix V_h with a nonzero element in the (p, q) position, with $p \neq q$.

(iii) Set E_h to

$$e_p e_p^T - e_{p+1} e_{p+1}^T + e_p e_{p+1}^T - e_{p+1} e_p^T \tag{5.10}$$

if U_h is diagonal but has different pth and $(p+1)$th diagonal entries.

(iv) Set F_h to (5.10) if V_h is diagonal but has different pth and $(p+1)$th diagonal entries.

(v) The only other possibility is that $U_h = \kappa_h I$ where $\kappa_h < 0$ for some h, since we know tr $V_j = 0, j = r+1, \cdots, r+s$ by (5.6). Set $E_h = -I_{t_h}$.

Now take inner products of $\{U_j\}$ with (5.7) and (5.8), respectively, and inner products of $\{V_j\}$ with (5.9), respectively. Summing the result and using (5.4) we obtain

$$\delta_0 \sum_{j=1}^{r+s} \text{tr } U_j + \sum_{j=r+1}^{r+s} \varepsilon_j \text{ tr } V_j = U_h : E_h + V_h : F_h.$$

Here one of the terms on the right-hand side is zero. The other is nonzero by construction. Using (5.5), (5.6) we therefore have $\delta_0 \neq 0$, and, as before, we may take $\delta_0 > 0$ by reversing the sign of the right-hand side and the solution of (5.7)–(5.9). Now add i times (5.9) to (5.8) to obtain

$$\sum_{k=1}^{m} \delta_k \bar{\lambda}_{j1}(x_0) B_k^{(j)} = E_j + F_j - (\delta_0 + \varepsilon_j i) I_{t_j}, \qquad j = r+1, \cdots, r+s. \tag{5.11}$$

In cases (i)–(iv) the eigenvalues of all $E_j, j = 1, \cdots, r$, and all $E_j + F_j, j = r+1, \cdots, r+s$, are zero, even for $j = h$. Therefore, by (5.7) and (5.11),

$$\text{Re } (\bar{\lambda}_{j1}(x_0) \mu_{jl}) = -\delta_0, \quad l = 1, \cdots, t_j, \quad j = 1, \cdots, r+s,$$

where $\{\mu_{jl}\}$ are the eigenvalues of

$$B^{(j)}(d) = \sum_{k=1}^{m} \delta_k B_k^{(j)}.$$

In case (v) the hth equation gives

$$\text{Re } (\bar{\lambda}_{h1}(x_0) \mu_{hl}) = -\delta_0 - 1$$

since $E_h = -I$. In both cases $f'(x_0; d) < 0$, where $d = [\delta_1, \cdots, \delta_m]^T$, by Lemma 3.4. □

Remark. In cases (i)–(iv), descent is obtained by maintaining all eigenvalue multiplicities but making $\lambda_{h1} = \cdots = \lambda_{ht_h}$ defective (to first order). We could just as well split $\lambda_{h1} = \cdots = \lambda_{ht_h}$ so that the change in all eigenvalues in the group has a common positive component in the direction (in the complex plane) $-\lambda_{h1}(x_0)$, and has different components in the orthogonal direction, i.e., tangent to the circle centered at the origin and passing through $\lambda_{h1}(x_0)$. This is what we did in Theorem 4.4, where the multiple eigenvalue is real. All we need do is set E_h or F_h, respectively, to a matrix with imaginary eigenvalues and nonzero inner product with U_h or V_h. In case (v), descent is obtained by preserving all nondefective eigenvalue multiplicities but reducing the modulus of λ_{hl} by more than the moduli of the other eigenvalues.

Remark. In the case $s = 0$, $t_j = 1$, $j = 1, \cdots, r$, Theorem 5.1 reduces to the standard min-max optimality condition where only case (v) applies. In the case $s = 0$, $r = 1$, the theorem reduces to Theorem 4.3. In the case $r = 0$, $s = 1$, the theorem reduces to a statement about splitting a multiple eigenvalue which is one of a single active complex conjugate pair.

We conclude this section with two examples.

Example 5.1. Reconsider Example 1.1. At $x_0 = [-1]$, we have $r = 2$, $t_1 = t_2 = 1$, $s = 0$. The codimension of the manifold defined by $|\lambda_1(x)| = |\lambda_2(x)|$ is $c_G(1, 1; 0) = 1$, so since $m = 1$, the dimension of the manifold is zero. The optimality condition is checked as follows. We have

$$\lambda_1(x_0) = 1, \qquad \lambda_2(x_0) = -1,$$

$$[Q_1\ Q_2] = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \qquad [P_1\ P_2] = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

$$B_1^{(1)} = \begin{bmatrix} \frac{1}{2} \end{bmatrix}, \qquad B_1^{(2)} = \begin{bmatrix} \frac{3}{2} \end{bmatrix}.$$

Equations (5.4)–(5.6), which define the dual matrices, in this case scalars, give

$$\frac{1}{2}U_1 - \frac{3}{2}U_2 = 0, \qquad U_1 + U_2 = 1.$$

The solution is $U_1 = \kappa_1 = \frac{3}{4}$, $U_2 = \kappa_2 = \frac{1}{4}$, so x_0 is indeed optimal.

Example 5.2. Let $n = 10$, let $x_0 = [0, \cdots, 0]^T$, and define $A_0 = A(x_0)$ by

$$A_0 = \begin{bmatrix} \sqrt{2} & & & & & & & & & \\ & \sqrt{2} & & & & & & & & \\ & & \sqrt{2} & & & & & & & \\ & & & \sqrt{2} & & & & & & \\ & & & & 1 & -1 & & & & \\ & & & & 1 & 1 & & & & \\ & & & & & & 1 & -1 & & \\ & & & & & & 1 & 1 & & \\ & & & & & & & & 1 & \\ & & & & & & & & & 0 \end{bmatrix}.$$

This matrix has one active quadruple real eigenvalue and one active double complex conjugate pair of eigenvalues, all with modulus $\sqrt{2}$. Thus $r = 1$, $s = 1$, $t_1 = 4$, $t_2 = 2$. In order for a generic family $A(x)$ to have x_0, and only x_0, as a point where $A(x)$ has a quadruple real eigenvalue and a complex conjugate pair with the same modulus, we require

$$m = c_G(4; 2) = 16 + 8 - 1 - 1 = 22.$$

The component matrices $\{A_k\}$, $k = 1, \cdots, 22$, are randomly generated by setting the elements, in the order $(A_1)_{11}, (A_1)_{1,2}, \cdots, (A_1)_{1,n}, \cdots, (A_1)_{n,n}, (A_2)_{1,1}, \cdots, (A_m)_{n,n}$, to the sequence ψ_ν, $\nu = 1, 2, \cdots$, defined by

$$\psi_\nu = \frac{\theta_\nu}{4095}, \qquad \theta_\nu = (445\theta_{\nu-1} + 1) \bmod 4096$$

and $\theta_0 = 1$.

We have $\lambda_{1,l} = \sqrt{2}$, $l = 1, \cdots, 4$, $\lambda_{2,l} = 1 + i$, $l = 1, 2$, and

$$P_1 = Q_1 = e_1e_1^T + e_2e_2^T + e_3e_3^T + e_4e_4^T,$$

$$P_2 = -\frac{i}{\sqrt{2}}(e_5e_1^T + e_7e_2^T) + \frac{1}{\sqrt{2}}(e_6e_1^T + e_8e_2^T),$$

$$Q_2 = \frac{i}{\sqrt{2}}(e_5e_1^T + e_7e_2^T) + \frac{1}{\sqrt{2}}(e_6e_1^T + e_8e_2^T).$$

Here e_p is the pth column of the identity matrix of the appropriate dimension, so that P_1, Q_1 are 10×4 and P_2, Q_2 are 10×2. Forming the system (5.4)–(5.6) and solving it, we obtain

$$U_1 = \begin{bmatrix} .455 & .040 & .039 & -.231 \\ -.110 & .094 & -.057 & -.187 \\ .007 & .017 & .335 & -.043 \\ -.227 & .092 & .002 & -.187 \end{bmatrix},$$

$$U_2 = \begin{bmatrix} .354 & -.017 \\ -.114 & -.050 \end{bmatrix}, \qquad V_2 = \begin{bmatrix} -.515 & -.078 \\ .338 & .515 \end{bmatrix}.$$

Thus there are many possible descent directions. For example, we have the following:

(i) Let $E_1 = -e_4 e_1^T$, $E_2 = 0$, $F_2 = 0$. Solving (5.7)–(5.9) we obtain $(\delta_0, d, \varepsilon_2)$ with $f'(0; d) = -\delta_0 = -.227$. Along this direction $\lambda_{1,1} = \cdots = \lambda_{1,4}$ does not split but becomes defective (to first order).

(ii) Let $E_1 = -e_2 e_1^T + e_1 e_2^T$, $E_2 = 0$, $F_2 = 0$. We get $f'(0; d) = -\delta_0 = -.150$. Because E_1 has imaginary eigenvalues, $\lambda_{1,1} = \cdots = \lambda_{1,4}$ splits into a complex conjugate pair and a double real eigenvalue (to first order).

(iii) Let $E_1 = 0$, $E_2 = -e_2 e_1^T$, $F_2 = 0$. We get $f'(0; d) = -\delta_0 = -.114$. This time it is the double complex conjugate pair of eigenvalues that becomes defective (to first order).

(iv) Let $E_1 = 0$, $E_2 = -e_2 e_1^T + e_1 e_2^T$, $F_2 = 0$. We get $f'(0; d) = -\delta_0 = -.097$. The double complex conjugate pair of eigenvalues splits in directions tangent to the circle in the complex plane centered at the origin with radius $\sqrt{2}$.

Finally, there is the degenerate case.

(2C) Nonsymmetric case, where neither (2A) nor (2B) applies.

This case generally applies if $m < c_G(t_1, \cdots, t_{r+s})$. As before such situations are degenerate, and the optimality condition does not apply.

6. The defective case. If $A(x_0)$ has a defective active eigenvalue, none of the previous results apply. In such cases it seems very hard to determine in general whether x_0 is a local minimizer of f, and, if not, to generate a descent direction. Indeed, it is well known that even determining the Jordan structure of $A(x_0)$ is difficult numerically.

Suppose there is one real active multiple eigenvalue $\lambda_1(x_0) = \cdots = \lambda_t(x_0)$, and suppose the orders of the corresponding Jordan blocks are $u_1 \geqq \cdots \geqq u_p$, $1 \leqq p \leqq t$. The codimension of the manifold on which the same Jordan structure is preserved is $c = c_D(u_1, \cdots, u_p)$, given by (2.6). If $m > c$, then generically the dimension of this manifold is at least one, and if x_0 does not minimize f on the manifold, it seems reasonable to suppose that a descent direction exists. This is not clear, however, since f is not Lipschitz along lines through x_0.

If $m = c$, then generically x_0 is the only point where $A(x)$ has the given Jordan structure. If $\lambda_1(x_0)$ is derogatory, i.e., there is more than one Jordan block corresponding to $\lambda_1(x_0)$, it may be possible to decrease $f(x)$ by making $\lambda_1(x)$ "more defective," i.e., moving to a point x where two of the Jordan blocks combine to form a larger block. Such points lie on a manifold with smaller codimension and hence larger dimension. If $m = c$ and $\lambda_1(x_0)$ is nonderogatory, i.e., $p = 1$, it will generally be necessary to split the multiple eigenvalue to obtain a reduction in f. It seems that the cases where x_0 is most likely to be a minimum are where $\lambda_1(x_0)$ is nonderogatory.

If $\lambda_1(x_0)$ is nonderogatory, an arbitrary perturbation of x with size ε will generally perturb the eigenvalues by $O(\varepsilon^{1/t})$. More specifically, the eigenvalues can be expanded

in Puiseux series; see Kato (1984, p. 65). The sum of the perturbed eigenvalues is analytic in ε (Kato (1984, p. 78)); accordingly, the $O(\varepsilon^{1/t})$ changes in the t eigenvalues are generally of equal magnitude and along directions in the complex plane separated by angles of $2\pi/t$. It follows that if $t > 2$, the spectral radius is increased by $O(\varepsilon^{1/t})$. If $t = 2$, the only case in which the spectral radius changes by $O(\varepsilon)$ is that in which the eigenvalues split into a complex conjugate pair; or more generally, if $\lambda_1(x_0)$ is complex, that in which the changes in the eigenvalues are tangent to the circle in the complex plane centred at the origin and passing through $\lambda_1(x_0)$. However, it is also true in the nondefective case that arbitrary perturbations to x generally increase the spectral radius; the question is whether a properly chosen perturbation can decrease f. It may be possible, even in the nonderogatory case, to perturb x so that the spectral radius is decreased. This would require that the first nonzero term in the Puiseux series be either an imaginary term of size $O(\varepsilon^{1/2})$ or a real term of size $O(\varepsilon)$. It might be achieved, for example, by splitting off a complex conjugate pair of eigenvalues and preserving multiplicity $t - 2$.

Consider Example 1.1. At $x_0 = 0$, $\lambda_1(x_0)$ is defective, with $n = t = 2$, $p = 1$. We have $c = m = 1$, and, indeed, x_0 is the only point where $\lambda_1(x)$ is defective. The point x_0 is a local minimizer of f. Now generalize the example to

$$A(x) = \begin{bmatrix} 1 + \gamma\xi_1 & 1 \\ -\xi_1 & 1 + \gamma\xi_1 \end{bmatrix}$$

with $x_0 = [0]$. Regardless of γ, the eigenvalues of $A(x)$ are real for $\xi_1 < 0$ and we may legitimately generalize the notion of directional derivative to say that $f'(0; -1) = +\infty$. For $\xi_1 > 0$, the eigenvalues are a complex conjugate pair, with

$$\lambda_{1,2}(\xi_1) = 1 + \gamma\xi_1 \pm i\sqrt{\xi_1}$$

so that

$$f'(0; +1) = \gamma + \tfrac{1}{2}.$$

Thus zero is a first-order local minimizer if and only if $\gamma \geqq -\frac{1}{2}$. In fact, we may without difficulty extend the definition of Clarke's generalized gradient to handle the case $m = 1$ regardless of whether $\lambda_1(x_0)$ is defective. In this particular case we obtain

$$\partial f(0) = [-\infty, \gamma + \tfrac{1}{2}]$$

so that, for any γ, f has a first-order local minimum at zero if and only if $0 \in \partial f(0)$.

The reason that duality theory, particularly the theorems in §§ 4 and 5, is so useful is that information computed only at x_0 defines dual variables, in our case matrices, that resolve the question of optimality and give information regarding descent directions. If $\lambda_1(x_0)$ is defective, however, it does not seem possible, even in the simple case just described, to resolve optimality directly from the information given by the Jordan form of $A(x_0)$ together with the component matrices $\{A_k\}$. It is possible, of course, to determine whether a given direction d is a descent direction by looking at the limit of the well defined quantities $f'(x_0 + \varepsilon d; d)$, where $\varepsilon > 0$ and $A(x_0 + \varepsilon d)$ has distinct eigenvalues, but this is of little use when $m > 1$.

Let us turn to Example 1.2 (see Fig. 1.2). We see that at, say, $x_0 = [1, 0]^T$, it is not trivial to determine which directions into the "complex region" are descent directions. In this case, the defective manifold shown in Fig. 1.2 is linear, so reducing f by keeping the eigenvalue defective poses no difficulty.

Finally, consider the following example.

Example 6.1. Let $n = 3$, $m = 2$ and define

$$A(x) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} + \xi_1 \begin{bmatrix} .5 & -.2 & -.4 \\ .7 & 1.2 & 1 \\ -2 & .8 & -.3 \end{bmatrix} + \xi_2 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Let $x_0 = [0, 0]^T$. At x_0, A has a nonderogatory triple eigenvalue. The codimension $c = 3 - 1 = 2$. Since $m = 2$, x_0 is the only point with this Jordan structure. Figure 6.1 gives a contour plot of $\rho(\xi_1, \xi_2)$. Figure 6.2 shows graphs of $\rho(\xi_1, \xi_2)$ along the lines $\xi_2 = 0.1$, $\xi_2 = 0$ and $\xi_2 = -0.1$, respectively.

There is a curve clearly visible in Fig. 6.1 across which $\rho(x)$ is not differentiable. Along the part of the curve above the point x_0, $A(x)$ is defective; more specifically, the triple eigenvalue splits into one defective double real eigenvalue and one single eigenvalue. On the part of the curve below x_0, $A(x)$ is not defective, and in fact it has distinct eigenvalues, one complex conjugate pair and one real eigenvalue. Along this part of the curve, $\rho(x)$ is a Lipschitz max function, with the complex conjugate pair and the real eigenvalue achieving the same modulus. Theorem 5.1 is trivially applicable at these points. It can be seen that ρ is Lipschitz along $\xi_2 = -0.1$ (Fig. 6.2(c)), that ρ is not Lipschitz along $\xi_2 = 0.1$ (Fig. 6.2(a)), and that ρ has even more rapid variation along $\xi_2 = 0$ (Fig. 6.2(b)); this is because a triple eigenvalue is being perturbed in the last case. There is another curve emanating up from x_0 along which the triple eigenvalue also splits into one defective double real eigenvalue and one single eigenvalue. This curve is not visible in the contour plot, since it is the *distinct* eigenvalue that has the maximum modulus. Thus the "defective manifold" has a cusp at x_0. This is consistent with the illustration given by Arnold (1971, p. 38); the manifold here corresponds to a cross-section of the one shown by Arnold.

We note that ρ is apparently locally but not globally minimized at x_0. There are lower values of ρ on the curve of discontinuity towards the bottom of Fig. 6.1.

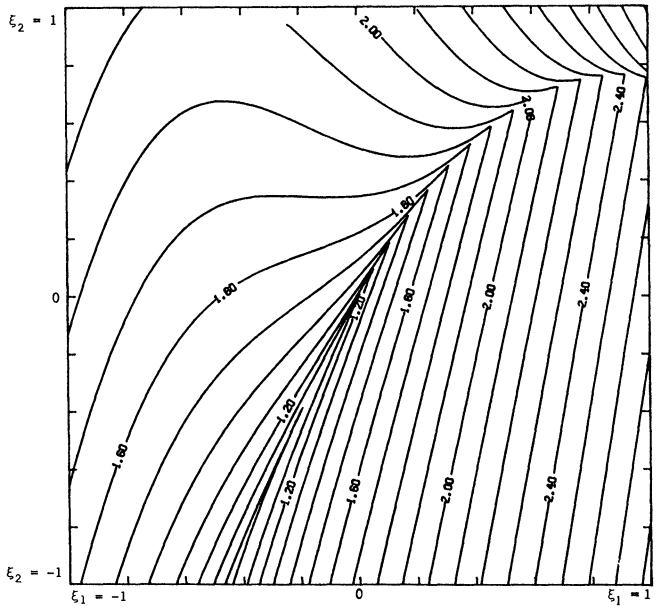


FIG. 6.1. *Contour plot of Example* 6.1.

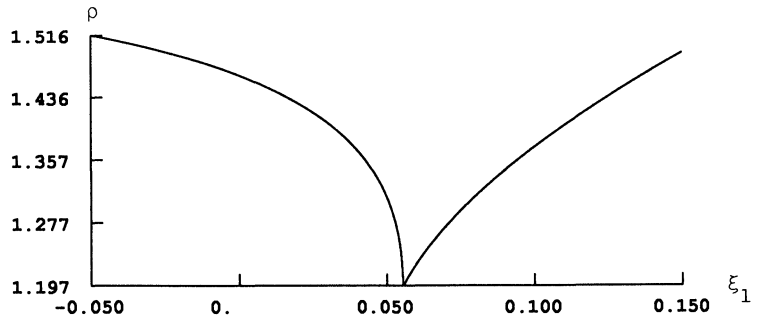


FIG. 6.2(a)

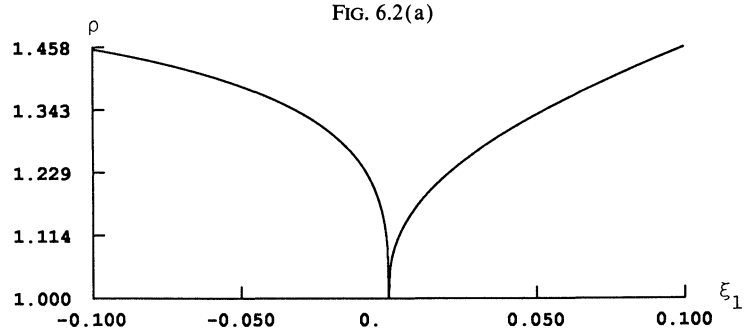


FIG. 6.2(b)

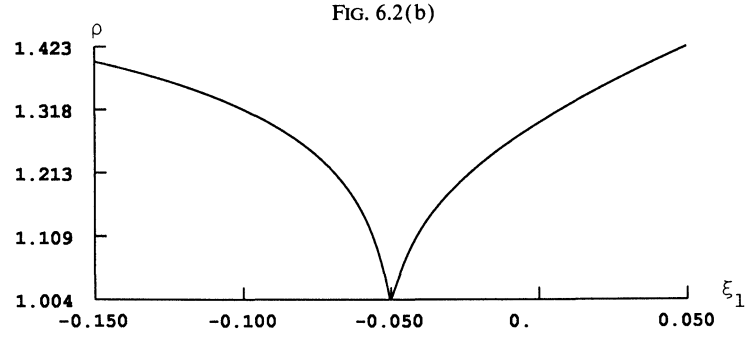


FIG. 6.2(c)

In summary, the question of optimality seems very hard to resolve in the defective case, and many interesting questions remain open.

Acknowledgments. The authors thank Dr. M. R. Osborne for many stimulating discussions, and Dr. Alastair Spence for bringing the work of Arnold to their attention. The first author thanks his hosts at the Australian National University, especially Mike Osborne, for their warm hospitality.

REFERENCES

V. I. ARNOLD (1971), *On matrices depending on parameters*, Russian Math. Surveys, 26, No. 2 pp. 29–43.

——— (1983), *Geometrical Methods in the Theory of Ordinary Differential Equations*, Springer-Verlag, Berlin, New York.

S. BOYD (1988), *Structured and simultaneous Lyapunov functions for system stability problems*, Information Systems Laboratory Report L-104-88-1, Stanford University, Stanford, CA.

J. V. BURKE (1987), private communication.

F. H. CLARKE (1975), *Generalized gradients and applications*, Trans. Amer. Math. Soc., 205, pp. 247–262.
——— (1983), *Optimization and Nonsmooth Analysis*, Wiley–Interscience, New York.
J. W. DEMMEL (1983), *A Numerical Analyst's Jordan Canonical Form*, Ph.D. thesis, Computer Science Dept., University of California, Berkeley, CA.
R. FLETCHER (1981), *Practical Methods of Optimization*, Vol. 2, John Wiley, Chichester, New York.
——— (1985), *Semi-definite matrix constraints in optimization*, SIAM J. Control Optim., 23, pp. 493–513.
S. FRIEDLAND (1978), *Extremal eigenvalue problems*, Bol. Soc. Brasil. Mat., 9, pp. 13–40.
S. FRIEDLAND, J. NOCEDAL, AND M. L. OVERTON (1987), *The formulation and analysis of numerical methods for inverse eigenvalue problems*, SIAM J. Numer. Anal., 24, pp. 634–667.
G. H. GOLUB AND C. VAN LOAN (1983), *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD.
V. A. KAMENETSKII AND E. S. PYATNITSKII (1987), *Gradient method of constructing Lyapunov functions in problems of absolute stability*, Automat. Remote Control, 48, pp. 1–8.
T. KATO (1984), *Perturbation Theory for Linear Operators*, 2nd edition, Springer-Verlag, Berlin, New York.
W. LEDERMANN (1937), *On the rank of the reduced correlational matrix in multiple-factor analysis*, Psychometrika, 2, pp. 85–93.
P. M. MÄKILÄ AND H. T. TOIVONEN (1987), *Computational methods for parametric LQ Problems—a survey*, IEEE Trans. Automat. Control, AC-32, pp. 658–671.
L. F. MILLER, R. G. COCHRAN, AND J. W. HOWZE (1978), *Output feedback stabilization by minimization of a spectral radius functional*, Internat. J. Control, 27, pp. 455–462.
P. NOWASAD (1968), *Isoperimetric eigenvalue problems in algebras*, Comm. Pure Appl. Math., 21, pp. 401–465.
M. L. OVERTON (1988), *On minimizing the maximum eigenvalue of a symmetric matrix*, SIAM J. Matrix Anal. Appl., 9, pp. 256–268.
R. T. ROCKAFELLAR (1970), *Convex Analysis*, Princeton University Press, Princeton, NJ.
——— (1981), *The Theory of Subgradients and Its Application to Problems of Optimization: Convex and Nonconvex Functions*, in Research and Education in Mathematics 1, Heldermann-Verlag, Berlin.
——— (1985), *Extensions of subgradient calculus with applications to optimization*, Nonlinear Anal. Theory Methods Appl., 9, pp. 665–698.
J. VON NEUMANN AND E. WIGNER (1929), *Über das Verhalten von Eigenwerten bei adiabatischen Prozessen*, Physik. Zeitschr., 30, pp. 467–470.
D. M. YOUNG (1971), *Iterative Solution of Large Linear Systems*, Academic Press, New York.