J. Soc. Indust. Appl. Math.
Vol. 13, No. 4, December, 1965
Printed in U.S.A.

ON THE STATIONARY VALUES OF A SECOND-DEGREE POLYNOMIAL ON THE UNIT SPHERE*

GEORGE E. FORSYTHE AND GENE H. GOLUB†

1. The problem. Let A be a Hermitian square matrix of complex elements and order n. Let b be a known n-vector of complex numbers. For each complex n-vector x, the nonhomogeneous quadratic expression

$$\Phi(x) = (x - b)^H A(x - b) \tag{1.1}$$

(H denotes complex conjugate transpose) is a real number. C. R. Rao of the Indian Statistical Institute, Calcutta, suggested to us the problem of maximizing (or minimizing) $\Phi(x)$ for complex x on the unit sphere $S = \{x: x^H x = 1\}$. Since Φ is a continuous function on the compact set S, such maxima and minima always exist. We here extend the problem to include finding all stationary values of Φ.

In summary, our problem is:

$$\text{find all } x \text{ which make } \Phi(x) \text{ stationary for } x^H x = 1. \tag{1.2}$$

The purpose of this note is to reduce the problem (1.2) to the determination of a certain finite real point set which we shall call the *spectrum* of the system (A, b) (defined at end of §1) and show that a unique number λ in the spectrum determines the one or more $x = x^\lambda$ which maximize $\Phi(x)$ for given b. Theorem (4.1) is the main result. The development is an extension to general b of the familiar theory for the homogeneous case when $b = \theta$, the zero vector. No consideration to a practical computer algorithm is given here.

In §7 we show that determining the least squares solution of an arbitrary system of linear equations $Cy = f$, subject to the quadratic constraint $y^H y = 1$, is a special case of problem (1.2).

As an abstraction from optimal control theory, Balakrishnan [1] studies the minimization of $\| Cy - f \|^2$, subject to the quadratic inequality constraint $y^H y \leq 1$, when C is a linear operator from one Hilbert space to another. In connection with approximation theory, Davis [2] in 1952 had considered a similar problem in Hilbert space. Morrison [5] and Marquardt [4] study the same minimization problem in n-dimensional Euclidean space, as a local approximation to a nonlinear parameter-estimation problem. Phillips [6] and Twomey [7] begin the actual numerical solution of

---

* Received by the editors March 19, 1965.

† Computer Science Department, Stanford University, Stanford, California.
This paper was supported in part by Office of Naval Research Contract Nonr-225(37) (NR-044-211) at Stanford University.

certain particular integral equations by approximating them with algebraic problems very closely related to the minimum problem (1.2).

Our present contribution differs from those cited above in three ways.

(a) Our constraint restricts x to a nonconvex set $x^H x = \text{const.}$, rather than to a convex set $x^H x \leqq \text{const.}$ Here, as for many programming problems, this seems to be an important difference.

(b) We consider the general problem of rendering $\Phi(x)$ stationary, instead of just a minimum. Moreover, we deal with general Hermitian matrices A instead of only semidefinite ones.

(c) We are concerned here with the structure of the problem as an extension of the classical theory of Hermitian forms, rather than with finding an algorithm.

Let $\lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_n$ be the (necessarily real) eigenvalues of A, and let $\{u_1, \cdots, u_n\}$ be a corresponding real orthonormal set of eigenvectors, with $Au_i = \lambda_i u_i$, $i = 1, \cdots, n$.

Let the given b be written

$$b = \sum_{i=1}^{n} b_i u_i . \tag{1.3}$$

(1.4) THEOREM. *If x is any vector in S at which $\Phi(x)$ is stationary with respect to S, then there exists a real number $\lambda = \lambda(x)$ such that*

$$A(x - b) = \lambda x, \tag{1.5}$$

$$x^H x = 1. \tag{1.6}$$

Conversely, if any real λ and vector x satisfy (1.5)–(1.6), *then x renders $\Phi(x)$ stationary with respect to S.*

Proof. Let x_0 be a point of S. Now, as shown in (8.7), $\Phi(x)$ is stationary at x_0 with respect to x in S if and only if there exists a *real* Lagrange multiplier λ such that $\Psi(x) = (x - b)^H A(x - b) - \lambda x^H x$ is stationary at x_0 with respect to all neighboring *complex* vectors x. Since

$$0 = \tfrac{1}{2} \operatorname{grad} \Psi(x_0) = A(x_0 - b) - \lambda x_0 ,$$

the theorem is proved.

To see what conditions are satisfied by the λ of (1.4), we note that the system (1.5)–(1.6) is equivalent to the system

$$(A - \lambda I)x = Ab, \tag{1.7}$$

$$x^H x = 1. \tag{1.8}$$

Let $x = \sum_{i=1}^{n} x_i u_i$. Then (1.7) is equivalent to

$$\sum_{i=1}^{n} (\lambda_i - \lambda) x_i u_i = \sum_{i=1}^{n} \lambda_i b_i u_i . \tag{1.9}$$

DEFINITION. By the *spectrum* of the pair (A, b) we mean the set of all real λ for which there exists an x such that (1.7) and (1.8) are satisfied, i.e., such that $\Phi(x)$ is stationary at x with respect to S.

Given any λ, x satisfying (1.7) and (1.8), we shall say that x *belongs to* λ, and frequently write x in the form x^λ.

Note that the spectrum of (A, θ) is the ordinary spectrum $\{\lambda_i\}$ of A.

2. Special case: no $\lambda_i b_i = 0$. Assume for the present section that $\lambda_i b_i \neq 0$, all i. This implies that all $\lambda_i \neq 0$, i.e., that A is nonsingular. If λ is in the spectrum of (A, b), (1.9) implies that $\lambda \neq \lambda_i$ for all i, and also that

$$x_i = \frac{\lambda_i b_i}{\lambda_i - \lambda}, \qquad i = 1, \cdots, n. \tag{2.1}$$

Then the requirement that

$$x^H x = \sum_{i=1}^{n} | x_i |^2 = 1 \tag{2.2}$$

is equivalent to the condition

$$g(\lambda) = \sum_{i=1}^{n} \frac{\lambda_i^2 | b_i |^2}{| \lambda_i - \lambda |^2} = 1. \tag{2.3}$$

Although all λ corresponding to stationary values of $\Phi(x)$ are known by (1.4) to be real, it is useful to define $g(\lambda)$ by (2.3) for all complex λ not in $\{\lambda_i\}$.

Let G be the set of complex numbers λ such that $g(\lambda) = 1$. For small enough $\sum_{i=1}^{n} | b_i |$, the set G is the union of n simple closed curves in the complex plane, the kth of which surrounds λ_k. As the $| b_i |$ grow, adjacent curves first coalesce in double points, and then merge into single curves. For very large values of all $| b_i |$, the set G is one simple closed curve including the set $\{\lambda_i\}$ in its interior. The family of sets G resembles the family of lemniscates $\prod_{i=1}^{n} | \lambda - \lambda_i | = \text{const.}$

Note, moreover, that $g(\lambda) > 1$ for λ inside any component curve G_j of G, while $g(\lambda) < 1$ in the exterior of all components G_j of G.

Now we shall show for the special case of §2 that each λ in G determines a unique x^λ which satisfies (1.7)–(1.8). For that x^λ,

$$\Phi(x^\lambda) = f(\lambda), \tag{2.4}$$

where we define f by

$$f(\lambda) = | \lambda |^2 \sum_{i=1}^{n} \frac{\lambda_i | b_i |^2}{| \lambda_i - \lambda |^2}. \tag{2.5}$$

Fix λ, and drop the superscript λ on x. To prove (2.4), note that (1.7) says $(\lambda_i - \lambda)x_i = \lambda_i b_i$. Thus

$$(\lambda_i - \lambda)(x_i - b_i) = \lambda_i b_i - b_i(\lambda_i - \lambda) = \lambda b_i .$$

Hence

$$x_i - b_i = \frac{\lambda b_i}{\lambda_i - \lambda},$$

and

$$\Phi(x) = \sum_{i=1}^{n} \lambda_i |x_i - b_i|^2 = |\lambda|^2 \sum_{i=1}^{n} \frac{\lambda_i |b_i|^2}{|\lambda_i - \lambda|^2} = f(\lambda),$$

proving (2.4).

Since the Lagrange multipliers λ must be real, the spectrum of (A, b) is the intersection of G with the real axis. This consists of from 2 to $2n$ distinct real numbers. How many numbers are actually in the spectrum depends on b; this will be discussed in §5 for $n = 2$.

We wish to determine which λ in the spectrum corresponds to the maximum [minimum] value of $f(\lambda)$. Let G_j be any component curve of the set G.

(2.6) THEOREM. *The maximum and minimum real parts of* λ, *for* λ *in any one* G_j, *both occur for* λ *on the real axis.*

Proof. Let $\lambda = \sigma + i\tau$, with σ, τ real. Then

$$g(\lambda) = g_1(\sigma, \tau) = \sum_{i=1}^{n} \frac{\lambda_i^2 |b_i|^2}{(\sigma - \lambda_i)^2 + \tau^2}.$$

Hence, for $\tau > 0$ and fixed σ, $g(\lambda)$ strictly decreases as τ increases. Then in the upper halfplane $\tau > 0$, any line $\sigma = $ const. intersects G_j in exactly one point. The theorem follows from this.

DEFINITION. Let Λ_R $[\Lambda_L]$ denote the unique real value of λ of maximum [minimum] real part in the set G.

(2.7) THEOREM. *Under the assumptions that A is regular (i.e.,* $\lambda_i \neq 0$ *for all* i) *and* $b_i \neq 0$, $i = 1, 2, \cdots, n$, *for all* λ, λ' *in* G, *we have*

$$\operatorname{Re}(\lambda) < \operatorname{Re}(\lambda') \textit{ implies } f(\lambda) < f(\lambda').$$

In particular, for λ *in* G *with* $\lambda \neq \Lambda_L$, $\lambda \neq \Lambda_R$, *we have*

$$f(\Lambda_L) < f(\lambda) < f(\Lambda_R).$$

Proof. Let $a_i = \lambda_i^2 |b_i|^2$, $i = 1, \cdots, n$. Introduce two independent complex variables λ, μ, where μ will later be set equal to $\bar{\lambda}$. In order to study the gradients of the functions g, f, and h(defined below) for complex λ, we shall use the tools of §8. This requires extending these functions into the space of λ and μ.

Let $\lambda = \sigma + i\tau$ where σ, τ are real. For all complex $\lambda \neq \lambda_i$, define the functions g_1 and g_2 by

$$g(\lambda) = g_1(\sigma, \tau) = g_2(\lambda, \bar{\lambda}),$$

where

$$g_2(\lambda, \mu) = \sum_{i=1}^{n} \frac{a_i}{(\lambda_i - \lambda)(\lambda_i - \mu)}.$$

(This definition is consistent with (2.3).) Then by (8.1),

$$\begin{aligned} \frac{1}{2}\left(\frac{\partial g_1}{\partial \sigma} + i\frac{\partial g_1}{\partial \tau}\right) &= \left[\frac{\partial g_2}{\partial \mu}\right]_{\mu=\bar{\lambda}} = \sum_{i=1}^{n} \frac{a_i}{(\lambda_i - \lambda)(\lambda_i - \bar{\lambda})^2} \\ &= \sum_{i=1}^{n} \frac{a_i}{|\lambda_i - \lambda|^2 \cdot (\lambda_i - \bar{\lambda})}. \end{aligned} \tag{2.8}$$

For all complex $\lambda \neq \lambda_i$, define $f(\lambda)$ by (2.5). We then define the functions f_1 and f_2 by

$$f(\lambda) = f_1(\sigma, \tau) = f_2(\lambda, \bar{\lambda}),$$

where

$$f_2(\lambda, \mu) = \lambda\mu \sum_{i=1}^{n} \frac{a_i}{\lambda_i(\lambda_i - \lambda)(\lambda_i - \mu)}.$$

Then by (8.1) and (2.8),

$$\begin{aligned} \frac{1}{2}\left(\frac{\partial f_1}{\partial \sigma} + i\frac{\partial f_1}{\partial \tau}\right) &= \left[\frac{\partial f_2}{\partial \mu}\right]_{\mu=\bar{\lambda}} \\ &= \lambda \sum_{i=1}^{n} \frac{a_i}{\lambda_i |\lambda_i - \lambda|^2} + \lambda\bar{\lambda} \sum_{i=1}^{n} \frac{a_i}{\lambda_i(\lambda_i - \lambda)(\lambda_i - \bar{\lambda})^2} \\ &= \lambda \sum_{i=1}^{n} \frac{a_i}{\lambda_i |\lambda_i - \lambda|^2}\left[1 + \frac{\bar{\lambda}}{\lambda_i - \bar{\lambda}}\right] \\ &= \lambda \sum_{i=1}^{n} \frac{a_i}{|\lambda_i - \lambda|^2(\lambda_i - \bar{\lambda})} = \lambda \left[\frac{\partial g_2}{\partial \mu}\right]_{\mu=\bar{\lambda}}, \end{aligned}$$

i.e.,

$$\left[\frac{\partial f_2}{\partial \mu}\right]_{\mu=\bar{\lambda}} = \lambda \cdot \left[\frac{\partial g_2}{\partial \mu}\right]_{\mu=\bar{\lambda}}. \tag{2.9}$$

While it is possible to use (2.9) to study the behavior of $f(\lambda)$ on the set G where $g(\lambda) - 1 = 0$, it is more convenient here and in §3 to introduce a new function $h(\lambda)$, which agrees with $f(\lambda)$ on G. For all complex $\lambda \neq \lambda_i$, define

$$h(\lambda) = f(\lambda) + \frac{\lambda + \bar{\lambda}}{2}[1 - g(\lambda)], \tag{2.10}$$

and note that

$$h(\lambda) = f(\lambda) \qquad \text{for } \lambda \in G. \tag{2.11}$$

As with f and g, we introduce functions h_1 and h_2 so that

$$h(\lambda) = h_1(\sigma, \tau) = h_2(\lambda, \mu),$$

where

$$h_2(\lambda, \mu) = f_2(\lambda, \mu) + \frac{\lambda + \mu}{2}[1 - g_2(\lambda, \mu)].$$

Then by (2.9),

$$\begin{aligned} \frac{\partial h_2}{\partial \mu} &= \frac{\partial f_2}{\partial \mu} + \frac{1}{2}[1 - g_2(\lambda, \mu)] - \frac{\lambda + \mu}{2}\frac{\partial g_2}{\partial \mu} \\ &= \frac{\lambda - \mu}{2}\frac{\partial g_2}{\partial \mu} + \frac{1}{2}[1 - g_2(\lambda, \mu)]. \end{aligned} \tag{2.12}$$

Hence by (8.1),

$$\begin{aligned} \frac{1}{2}\left(\frac{\partial h_1}{\partial \sigma} + i\frac{\partial h_1}{\partial \tau}\right) = \left[\frac{\partial h_2}{\partial \mu}\right]_{\mu=\bar\lambda} &= \frac{\lambda - \bar\lambda}{2}\left[\frac{\partial g_2}{\partial \mu}\right]_{\mu=\bar\lambda} + \frac{1}{2}[1 - g_2(\lambda, \bar\lambda)] \\ &= \frac{\tau i}{2}\left(\frac{\partial g_1}{\partial \sigma} + i\frac{\partial g_1}{\partial \tau}\right) + \frac{1}{2}[1 - g(\lambda)]. \end{aligned} \tag{2.13}$$

Now any component G_j of the set G where $g(\lambda) = 1$ encloses a region where $g(\lambda) > 1$. On G the gradient vector of g,

$$\frac{\partial g_1}{\partial \sigma} + i\frac{\partial g_1}{\partial \tau},$$

is nonzero, is normal to G_j, and points to the interior of G_j. Then, by (2.12), the gradient vector of h on G_j, namely

$$\frac{\partial h_1}{\partial \sigma} + i\frac{\partial h_1}{\partial \tau} = i\tau\left(\frac{\partial g_1}{\partial \sigma} + i\frac{\partial g_1}{\partial \tau}\right),$$

is nonzero for $\tau \neq 0$ and points along the tangent to G_j in the direction of increasing σ. Hence

(2.14) $h(\lambda)$ *is strictly increasing, as* λ *traces* G_j *in the direction of increasing* σ.

From (2.14) it follows that $h(\lambda)$ assumes its maximum value, for each separate component curve G_j of G, at the point β_j on G_j of maximum real part. By (2.6), β_j is on the axis of real λ.

Note that setting $\mu = \bar\lambda = \lambda$ in (2.12) yields the result that

$$h'(\lambda) = 1 - g(\lambda), \qquad \text{for real } \lambda. \tag{2.15}$$

To complete the proof of the present theorem, we must show that $f(\lambda)$ is larger at the point α_j of least real part on the component G_j of G than it is at the rightmost point β_{j-1} of the component G_{j-1} of G immediately to the left of G_j.

Note that g is continuous for $\lambda \in [\beta_{j-1}, \alpha_j]$, and that $g(\beta_{j-1}) = g(\alpha_j) = 1$, but $g(\lambda) < 1$ for $\beta_{j-1} < \lambda < \alpha_j$. Then by (2.15),

$$\begin{aligned} h(\alpha_j) &= h(\beta_{j-1}) + \int_{\beta_{j-1}}^{\alpha_j} h'(\lambda)\, d\lambda \\ &= h(\beta_{j-1}) + \int_{\beta_{j-1}}^{\alpha_j} [1 - g(\lambda)]\, d\lambda > h(\beta_{j-1}). \end{aligned}$$

Thus

$$h(\beta_{j-1}) < h(\alpha_j), \tag{2.16}$$

as was to be proved.

We conclude that $h(\lambda)$ increases as λ increases along the real axis between adjacent components of G. Since $h(\lambda) = f(\lambda)$ on G, we see that (2.7) follows from (2.14) and (2.16). In particular,

$$\max_{\lambda \in G} f(\lambda) = f(\Lambda_R),$$

$$\min_{\lambda \in G} f(\lambda) = f(\Lambda_L).$$

It follows trivially from (2.7) that, if the *real* numbers in G (i.e., the spectrum of (A, b) in this case) consist of the set $\{\Lambda_1, \Lambda_2, \cdots, \Lambda_p\}$, with

$$\Lambda_L = \Lambda_1 < \Lambda_2 < \cdots < \Lambda_p = \Lambda_R,$$

then

$$f(\Lambda_1) < f(\Lambda_2) < \cdots < f(\Lambda_p).$$

By (1.9), our condition that no $\lambda_i b_i = 0$ implies that $\lambda \neq \lambda_i$ for all i and for all λ in G. Hence no Λ_k is an eigenvalue of A, and so $A - \Lambda_k I$ is nonsingular for all k. Therefore we can solve (1.7) uniquely for a vector x^k belonging to Λ_k:

$$x^k = x^{\Lambda_k} = (A - \Lambda_k I)^{-1} A b.$$

In particular,

$$x_{\max} = x^{\Lambda_R} = (A - \Lambda_R I)^{-1} A b,$$

$$x_{\min} = x^{\Lambda_L} = (A - \Lambda_L I)^{-1} A b.$$

These equations give unique solutions to the problem of minimizing and maximizing $\Phi(x) = (x - b)^H A(x - b)$, for nonsingular A and b such that no $b_i = 0$.

It would be desirable to be able to prove[1] that $f(\alpha_j) < f(\beta_j)$, in the notation of (2.7), without analyzing $f(\lambda)$ and $g(\lambda)$ for complex values of λ.

Since the Lagrange multipliers are real, we see from (1.7) and (1.8) that Λ_R and Λ_L must be roots of the equation

$$\Delta(\lambda) = \Gamma(\lambda) - 1 = 0, \tag{2.17}$$

where

$$\begin{aligned}\Gamma(\lambda) &= [(A - \lambda I)^{-1}Ab]^H[(A - \lambda I)^{-1}Ab] \\ &= b^H A(A - \lambda I)^{-2}Ab = \sum_{i=1}^{n} \frac{\lambda_i^2 |b_i|^2}{(\lambda_i - \lambda)^2}.\end{aligned} \tag{2.18}$$

Since $\Delta(\lambda)$ has a pole at λ_n and $\Delta(\lambda) \to -1$ as $\lambda \to \infty$, we see that $\Lambda_R > \lambda_n$. Similarly, $\Lambda_L < \lambda_1$.

We now use the identity

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A) \cdot \det(D - CA^{-1}B), \tag{2.19}$$

valid for any partitioned matrix with A and D square and $\det(A) \neq 0$. From (2.19) and (2.18) we may replace (2.17) with the equation

$$\begin{aligned}1 - \Gamma(\lambda) &= 1 - b^H A(A - \lambda I)^{-2}Ab \\ &= \det \begin{bmatrix} (A - \lambda I)^2 & Ab \\ b^H A & 1 \end{bmatrix} = 0.\end{aligned} \tag{2.20}$$

It follows from (2.20) that, when λ is either Λ_L or Λ_R, there exist a column vector r and a number s, both depending on λ, such that

$$(A - \lambda I)^2 r + Abs = \theta, \qquad b^H Ar + s = 0.$$

A simple elimination shows that Λ_R and Λ_L must satisfy the determinantal equation

$$\det(\lambda^2 I - 2\lambda A + A(I - bb^H)A) = 0. \tag{2.21}$$

Equation (2.21) may well lead to an effective computational algorithm for obtaining Λ_R and Λ_L.

Let us assume, for the remainder of §2 only, that A is positive definite; such problems may certainly occur, as is demonstrated in the least squares application discussed in §7. We shall now give bounds for Λ_R and Λ_L which depend upon $\| b \|^2 = b^H b$.

[1] Professor W. Kahan of the University of Toronto has just shown us such a proof.

Case 1. $\| b \| < 1$. For $\| b \| < 1$, $A(I - bb^H)A$ is positive definite, since $(I - bb^H)$ is positive definite. Consequently, if

$$[\lambda^2 I - 2\lambda A + A(I - bb^H)A]r = \theta \tag{2.22}$$

with $\| r \| = 1$, the roots of (2.21) satisfy

$$\lambda^2 - 2\lambda a_1(r) + b_1(r) = 0, \tag{2.23}$$

where

$$a_1(r) = r^H A r, \qquad b_1(r) = r^H A(I - bb^H)Ar.$$

Thus, all the real roots of (2.21) are positive, and, hence, by Descartes' rule of signs and the fact that the sum of the two roots of (2.23) is $2a_1(r) \leqq 2\lambda_n$, we see that

$$0 < \Lambda_L < \lambda_1 \quad \text{and} \quad \lambda_n < \Lambda_R < 2\lambda_n .$$

Case 2. $\| b \| = 1$. Here $A(I - bb^H)A$ is semidefinite and for any vector r, $a_1(r) > 0$, $b_1(r) \geqq 0$. If $A(I - bb^H)Aw = \theta$, w satisfies (2.22) with $\lambda = 0$. Thus,

$$\Lambda_L = 0, \qquad \lambda_n < \Lambda_R \leqq 2\lambda_n .$$

Case 3. $\| b \| > 1$. Since $\Delta(0) > 0$ and $\Delta(x) \to -1$, as $\lambda \to \infty$, we see that

$$\Lambda_L < 0 < \lambda_1 .$$

3. General case: some $\lambda_i b_i = 0$. We now study the general case where one or more $\lambda_i b_i = 0$. To be explicit, let $\mathcal{L}$ be the set of all real numbers μ for which there exists an eigenvalue λ_k of A with $\lambda_k = \mu$ and $\lambda_k b_k = 0$. We wish to find the spectrum of (A, b).

If $\lambda \notin \mathcal{L}$, the condition for λ to be in the spectrum is precisely what it was in §2 above—namely, that

$$\sum_{i=1}^{n} \frac{\lambda_i^2 |b_i|^2}{|\lambda_i - \lambda|^2} = 1. \tag{3.1}$$

If certain of the $\lambda_i^2 | b_i |^2$ are 0, then, if $\lambda \notin \mathcal{L}$ is in the spectrum, the corresponding x_i^λ are 0, by (1.9).

Now we fix attention on an arbitrary λ in $\mathcal{L}$, and consider whether it is in the spectrum of (A, b). Let m be the multiplicity of λ as an eigenvalue of A. Define the set of integers $\mathcal{I} = \mathcal{I}_\lambda = \{i: \lambda_i = \lambda\}$. Then card $(\mathcal{I}) = m$. If $\lambda_i b_i \neq 0$ for some $i \in \mathcal{I}$, then it would be impossible to satisfy the condition

$$(\lambda_i - \lambda)x_i = \lambda_i b_i \tag{1.9}$$

for that i, and so λ could not be in the spectrum of (A, b). (By the same argument, no eigenvalue λ_k not in $\mathcal{L}$ could be in the spectrum.)

Now suppose that $\lambda_i b_i = 0$ for all m indices i in $\mathcal{I}$. Then when is λ in the spectrum? As stated in §1, λ is in the spectrum if and only if (1.8) and (1.9) are satisfied. The condition (1.9) determines a unique x_i^λ for each $i \notin \mathcal{I}$, but it places no restriction on x_i^λ for $i \in \mathcal{I}$. Form the sum

$$\sum_{i \notin \mathcal{I}} | x_i^\lambda |^2 = \sum_{i \notin \mathcal{I}} \frac{\lambda_i^2 | b_i |^2}{| \lambda_i - \lambda |^2}.$$

If we make the convention that $0/0 = 0$, the above sum can be written as

$$g(\lambda) = \sum_{i=1}^{n} \frac{\lambda_i^2 | b_i |^2}{| \lambda_i - \lambda |^2}. \tag{3.2}$$

If $g(\lambda) > 1$, then it would be impossible to define x_i^λ for $i \in \mathcal{I}$ so that $\sum_{i=1}^n | x_i^\lambda |^2 = 1$, and so λ cannot be in the spectrum of (A, b), by (1.8).

If $g(\lambda) = 1$, then we can make $\sum_{i=1}^n | x_i^\lambda |^2 = 1$ if and only if we set $x_i^\lambda = 0$ for all i in $\mathcal{I}$. Thus λ is in the spectrum, and the corresponding x^λ is unique.

If $g(\lambda) < 1$, then we can define x_i^λ, arbitrarily, for $i \in \mathcal{I}$, subject to the condition that

$$\sum_{i \in \mathcal{I}} | x_i^\lambda |^2 = 1 - g(\lambda), \tag{3.3}$$

and we will have $\sum_{i=1}^n | x_i^\lambda |^2 = 1$ in any case. The set of vectors x^λ determined by (1.9) for $i \notin \mathcal{I}$ and by (3.3) for $i \in \mathcal{I}$ forms an $(m - 1)$-dimensional sphere $\mathcal{V}$. For, if an x^λ in $\mathcal{V}$ has components x_k, $k = 1, \cdots, n$, then any y with components y_k, $k = 1, \cdots, n$, is also in $\mathcal{V}$, if

$$y_k = \begin{cases} x_k & \text{for } k \notin \mathcal{I} \\ x_k e^{i\theta_k} \quad (\theta_k \text{ real}) & \text{for } k \in \mathcal{I}, \end{cases}$$

since $\sum_{k \in \mathcal{I}} | y_k |^2 = 1 - g(\lambda)$ for all y. Thus, when $g(\lambda) < 1$, uniqueness of x^λ is lost. The sphere $\mathcal{V}$ is analogous to (in fact, is a generalization of) the sphere of unit eigenvectors of a Hermitian matrix A belonging to an eigenvalue of multiplicity m.

Note that the inequality $g(\lambda) < 1$ states that λ is in the exterior of the graph

$$G = \left\{ \mu : \sum_{i \notin \mathcal{I}} \frac{\lambda_i^2 | b_i |^2}{| \lambda_i - \mu |^2} = 1 \right\},$$

i.e., λ can be joined to ∞ by an arc not cutting G. Thus, in brief, the spectrum of (A, b) consists of the union of all real numbers in the set

$$G = \left\{ \mu : \sum_{i=1}^{n} \frac{\lambda_i^2 | b_i |^2}{| \lambda_i - \mu |^2} = 1 \right\}, \tag{3.4}$$

where we interpret 0/0 as 0, together with those eigenvalues λ_k of A which are exterior to the graph G. (If G is the null set, then $b = \theta$ and the spectrum of (A, θ) consists of all eigenvalues λ_k.)

We must now examine $\Phi(x^\lambda)$ for λ in the spectrum of (A, b). The study of $\Phi(x^\lambda)$, for real $\lambda \in G$ in (3.4), is the same as in §2, and yields the same results (2.4) and (2.5). First, for $\lambda \in G$,

$$\Phi(x^\lambda) = f(\lambda) = |\lambda|^2 \sum_{i=1}^{n} \frac{\lambda_i |b_i|^2}{|\lambda_i - \lambda|^2},$$

where $0/0 = 0$. Second, let μ_R, μ_L be the rightmost [respectively, leftmost] points of G. Then $f(\mu_R)$ maximizes [respectively, $f(\mu_L)$ minimizes] $f(\lambda)$ for $\lambda \in G$. It remains to consider $\Phi(x^{\lambda_k})$, for eigenvalues λ_k outside G.

(3.5) THEOREM. *For any* λ *in the spectrum of* (A, b) *we have*

$$\Phi(x^\lambda) = h(\lambda) = f(\lambda) + \lambda[1 - g(\lambda)], \tag{3.6}$$

where $f(\lambda)$ *is given by* (2.5) *with* 0/0 *interpreted as* 0.

Proof. Take any λ in the spectrum of (A, b).

If $\lambda \neq \lambda_k$, $k = 1, \cdots, n$, then $\lambda \in G$, and everything proceeds as in the proof of (2.4), showing that $\Phi(x^\lambda) = f(\lambda)$. Since $g(\lambda) = 1$, we have proved (3.6) when $\lambda \neq \lambda_k$.

If $\lambda = \lambda_k$, an eigenvalue of A, let x_i denote the ith coordinate of any corresponding vector x^{λ_k}. Since λ_k is in the spectrum of (A, b), we have $\lambda_i b_i = 0$ for all $i \in \mathscr{I}$, where $\mathscr{I}$ is defined above after (3.1), and hence $\lambda_i |x_i - b_i|^2 = \lambda_i |x_i|^2 = \lambda_k |x_i|^2$, for all $i \in \mathscr{I}$. Then, by (3.3),

$$\sum_{i \in \mathscr{I}} \lambda_i |x_i - b_i|^2 = \lambda_k \sum_{i \in \mathscr{I}} |x_i|^2 = \lambda_k[1 - g(\lambda_k)]. \tag{3.7}$$

Moreover, like (2.4) we can prove (where $0/0 = 0$)

$$\sum_{i \notin \mathscr{I}} \lambda_i |x_i - b_i|^2 = \lambda_k^2 \sum_{i \notin \mathscr{I}} \frac{\lambda_i |b_i|^2}{(\lambda_i - \lambda_k)^2} = \lambda_k^2 \sum_{i=1}^{n} \frac{\lambda_i |b_i|^2}{(\lambda_i - \lambda_k)^2} = f(\lambda_k). \tag{3.8}$$

Adding (3.7) to (3.8), we get

$$\begin{aligned} \phi(x^{\lambda_k}) &= \sum_{i=1}^{n} \lambda_i |x_i - b_i|^2 = \sum_{i \notin \mathscr{I}} \lambda_i |x_i - b_i|^2 + \sum_{i \in \mathscr{I}} \lambda_i |x_i - b_i|^2 \\ &= f(\lambda_k) + \lambda_k[1 - g(\lambda_k)] = h(\lambda_k). \end{aligned} \tag{3.9}$$

This proves (3.6) when $\lambda = \lambda_k$.

It is property (3.6) of h which motivated our use of h in §2.

Remark. It is easily shown from (3.6) or (3.9) that, for all λ in the spectrum of (A, b),

$$h(\lambda) = \lambda + \lambda \sum_{i=1}^{n} \frac{\lambda_i |b_i|^2}{\lambda - \lambda_i}, \tag{3.10}$$

where $0/0 = 0$. If λ is in the spectrum of (A, b), but is not an eigenvalue of A, we can derive (3.10) as follows. Let x belong to λ. Then using (1.5), (1.6), and (1.7), we obtain

$$\Phi(x) = (x - b)^H A(x - b) = (x - b)^H \lambda x = \lambda x^H x - \lambda b^H x = \lambda - \lambda b^H x$$

$$= \lambda - \lambda b^H (A - \lambda I)^{-1} A b = \lambda - \lambda \sum_{i=1}^{n} \frac{\lambda_i |b_i|^2}{\lambda_i - \lambda}.$$

We shall not make use of (3.10) here.

We now use (3.6) to extend the domain of h to all real λ where $g(\lambda) < \infty$, i. e., to all λ except where, for some i, $\lambda = \lambda_i$ and $\lambda_i b_i \neq 0$.

As stated before (3.5), we know that the largest value of $\Phi(x^\lambda) = h(\lambda)$ for λ in G occurs at the rightmost point μ_R of G. It remains to see whether $h(\lambda_k)$ may be still larger for any λ_k in the spectrum of (A, b), if $\mu_R < \lambda_k$.

The answer is furnished by (2.15), which is valid for the general case of §3 with the understanding that $0/0 = 0$. Thus h is increasing on all segments of the real axis between or exterior to components of the curve G. It follows that $h(\lambda)$ takes its maximum at the rightmost point Λ_R of the spectrum of (A, b) and its minimum value at the leftmost point Λ_L of the spectrum of (A, b), whether or not these are eigenvalues of A.

From the considerations following (3.3), we see that the maximizing x is unique if $\Lambda_R \in G$. If, however, Λ_R is not in G and is an eigenvalue of A of multiplicity m, then the maximizing x include all points of an $(m - 1)$-sphere of nonzero radius, whose center is not at θ when $b \neq \theta$.

Analogous statements can be made about the uniqueness of the vectors x belonging to any λ which is in the spectrum of (A, b).

The above result about Λ_R and Λ_L for the case where some $\lambda_i b_i = 0$ can be obtained by continuity from the case where no $\lambda_i b_i = 0$. It is not clear that we could use continuity to deduce the nature of the maximizing and minimizing vectors, for multiple roots.

4. The main result. In §2 and §3 we have proved our main result:

(4.1) THEOREM. *Given A, a Hermitian matrix of order n with eigenvalues $\{\lambda_i\}$, and b, an arbitrary complex n-vector, define $\{b_i\}$ as in* (1.3). *Then the spectrum of (A, b) consists of all real λ such that*

$$g(\lambda) = \sum_{i=1}^{n} \frac{\lambda_i^2 |b_i|^2}{|\lambda_i - \lambda|^2} = 1 \qquad (0/0 = 0, 1/0 = \infty),$$

together with each eigenvalue λ_k of A for which $g(\lambda_k) < 1$.

For each λ in the spectrum with $g(\lambda) = 1$, a unique x^λ is found by solving (1.7)–(1.8). *For each λ in the spectrum with $g(\lambda) < 1$, there exists an $(m - 1)$-sphere of x^λ satisfying* (1.7)–(1.8), *where* $m = \text{card}\{j : \lambda_j = \lambda_k\}$.

Each x^λ so found renders $\Phi(x)$ stationary on S. Let the spectrum of (A, b) consist of the numbers $\{\Lambda_1, \cdots, \Lambda_p\}$, with

$$\Lambda_L = \Lambda_1 < \Lambda_2 < \cdots < \Lambda_p = \Lambda_R .$$

For $i = 1, 2, \cdots, p$, let x^i be any x^{λ_i}. Then

$$\Phi(x^1) < \Phi(x^2) < \cdots < \Phi(x^p).$$

In particular, $\Phi(x^1)$, $\Phi(x^p)$ are, respectively, the minimum and maximum values of $\Phi(x)$ on S.

5. The number of points in the spectrum. As we noted in §2, if A is of order n, then the spectrum of (A, b) contains anywhere from 2 to $2n$ real points. When does it have the full number $2n$? If any $\lambda_i b_i = 0$, then the discussion of §3 shows that the spectrum necessarily has fewer than $2n$ points. So we are limited to the case where all $\lambda_i b_i \neq 0$. But then, as shown in §2, we know that the spectrum is the intersection of the graph of

$$\mu = \sum_{i=1}^{n} \frac{\lambda_i^2 \mid b_i \mid^2}{\mid \lambda_i - \lambda \mid^2} \tag{5.1}$$

for real λ with the line $\mu = 1$.

The graph of (5.1) for real λ consists of $n + 1$ branches between the n vertical asymptotes $\lambda = \lambda_i$, $i = 1, \cdots, n$. Since $\mu > 0$ for all λ, and $\mu \to 0$ as $\lambda \to \infty$ and $\lambda \to -\infty$, the rightmost and leftmost branches necessarily cut $\mu = 1$. The spectrum has the full number $2n$ of points if and only if each of the $n - 2$ interior branches of the curve reaches its minimum with $\mu < 1$. For general n a condition for this is probably too complicated to derive. For $n = 2$, however, we can answer the question, as follows.

(5.2) THEOREM.

(5.3) *Let $n = 2$, and assume A is in diagonal form with $\lambda_1 < \lambda_2$. If the spectrum of (A, b) consists of 4 distinct numbers, then*

$$0 < \mid b_1\lambda_1 \mid \quad \textit{and} \quad 0 < \mid b_2\lambda_2 \mid, \tag{5.4}$$

and also

$$\mid b_1\lambda_1 \mid^{2/3} + \mid b_2\lambda_2 \mid^{2/3} < (\lambda_2 - \lambda_1)^{2/3}. \tag{5.5}$$

(5.6) *Conversely, if* (5.4) *and* (5.5) *hold, then the spectrum of (A, b) consists of 4 distinct numbers.*

Proof of (5.3). Let $a_i = \mid \lambda_i b_i \mid^2$, $i = 1, 2$. If either a_1 or a_2 were zero, then the development in §3 shows that the spectrum would consist of at most 3 points. Hence $a_1 > 0$ and $a_2 > 0$, i.e., (5.4) holds. Let $M = (a_2/a_1)^{1/3}$. Now the development in §2 shows that the spectrum of (A, b) consists precisely of the real roots λ of the equation

$$g(\lambda) = \frac{a_1}{(\lambda - \lambda_1)^2} + \frac{a_2}{(\lambda - \lambda_2)^2} = 1. \tag{5.7}$$

Since (5.7) has 4 real roots, we know that two roots must lie in the interval (λ_1, λ_2). Now let μ be the unique real root of

$$g'(\lambda) = \frac{-2a_1}{(\lambda - \lambda_1)^3} - \frac{2a_2}{(\lambda - \lambda_2)^3} = 0.$$

Then, because there are two roots of (5.7) in (λ_1, λ_2),

$$g(\mu) < 1. \tag{5.8}$$

We now show that (5.8) implies (5.4).

Solving $g'(\mu) = 0$ shows that

$$\frac{\lambda_2 - \mu}{\lambda - \lambda_1} = M,$$

whence

$$\mu - \lambda_1 = \frac{1}{1 + M}(\lambda_2 - \lambda_1),$$

$$\lambda_2 - \mu = \frac{M}{1 + M}(\lambda_2 - \lambda_1).$$

For brevity, define

$$Q = a_1^{1/3} + a_2^{1/3}.$$

Hence

$$g(\mu) = \frac{a_1(1 + M)^2}{(\lambda_2 - \lambda_1)^2} + \frac{a_2(1 + M)^2}{M^2(\lambda_2 - \lambda_1)^2} = \frac{(1 + M)^2}{(\lambda_2 - \lambda_1)^2}\left[a_1 + \frac{a_2}{M^2}\right]$$

$$= \frac{(1 + M)^2 a_1^{2/3}}{(\lambda_2 - \lambda_1)^2} \cdot Q = \frac{Q^2}{(\lambda_2 - \lambda_1)^2} \cdot Q = \frac{Q^3}{(\lambda_2 - \lambda_1)^2}.$$

Thus $g(\mu) < 1$ implies

$$a_1^{1/3} + a_2^{1/3} < (\lambda_2 - \lambda_1)^{2/3}, \tag{5.9}$$

which implies (5.5). Thus (5.3) is proved.

Proof of (5.6). We have $a_1 > 0$, $a_2 > 0$, and (5.9). The above steps are reversible, and so $g(\mu) < 1$, whence there are 4 real roots of $g(\mu) = 1$.

Thus (5.2) is completely proved.

Condition (5.4) says that neither λ_1 nor λ_2 is 0, and that the point $b = (b_1, b_2)$ does not lie on an axis of the (x_1, x_2)-plane. Condition (5.5) requires that (b_1, b_2) be inside a curve Γ which depends only on the ratio λ_2/λ_1. If $\lambda_2/\lambda_1 = 2$, for example, the curve Γ is $| b_1 |^{2/3} + | 2b_2 |^{2/3} = 1$.

In Fig. 1 the number of points in the spectrum of (A, b) is indicated, for different b in the first quadrant, by integers in circles.

If $| \lambda_2/\lambda_1 | > 2$, the curve Γ includes values $| b_1 | > 1$. But $| \lambda_2/\lambda_1 | > 1$ implies that, on Γ, $| b_2 | < 1$.

6. Geometrical interpretation. The surfaces $\Phi(x) = k$ are similar conic surfaces with center b in the Euclidean n-space $\mathcal{E}_n$ of vectors x. The maximum problem (1.2) is to find the conic surface with maximum k which touches the constraint surface S, the unit sphere in $\mathcal{E}_n$. The rotation of A to diagonal form is a rotation of $\mathcal{E}_n$ (leaving S invariant, of course) which causes principal axes of the conic surface to coincide with the axes of $\mathcal{E}_n$.

The vector $Ax - b$ is half the gradient of $\Phi(x)$, and x is the radius vector. Condition (1.5) merely states that at a point where $\Phi(x)$ is stationary, for x on S, the surface $\Phi(x) = k$ is tangent to S.

Fix x at a solution of (1.5), and let t be real. If the constant λ of (1.5) is positive, the value of $\Phi(tx)$ increases as t increases from 1; if λ is negative, $\Phi(tx)$ decreases as t increases from 1.

The main result of §2 and §3 is that the maximum problem of §1 is solved for the largest value of λ satisfying (1.5), for x on S. The authors see no obvious geometrical reason why this should be so.

If all $b_i\lambda_i \neq 0$, then §2 shows that any vector $x = x^\lambda$ which makes $\Phi(x)$ stationary on S is uniquely determined by λ.

Fig. 2 shows, for $n = 2$ and $0 < \lambda_1 < \lambda_2$, a case where there are 4 distinct points of tangency of an ellipse with the unit circle. All ellipses have center b and common value of $\lambda_2/\lambda_1 > 2$. Since $\lambda_2/\lambda_1 > 2$, it was shown in §5 that 4 distinct tangencies are possible for certain b outside S.

Whenever some $b_k = 0$, then, provided that (3.2) holds with the inequality sign $<$, we get more than one x belonging to a given λ. That is illustrated in Fig. 3, where $n = 2$ and $k = 1$. What is not obvious to the authors is a geometrical reason why necessarily $\lambda = \lambda_k$ in this case.

7. A constrained least squares problem. Let C be an $m \times n$ matrix, $m \geqq n$, and f an m-vector, both over the complex field. We wish to study the set of complex n-vectors y of Euclidean length $\| y \| = (y^H y)^{1/2} = 1$ such that

$$\| Cy - f \|^2 = (Cy - f)^H(Cy - f) = \min. \tag{7.1}$$

The constraint is

$$\| y \|^2 = y^H y = 1. \tag{7.2}$$

Because Euclidean length is invariant under unitary transformations, it is useful to rotate coordinates in both the space of y and the space of f. To do this, let $r = \text{rank}\ (C)$, and write

$$C = U^H D V, \tag{7.3}$$

where U, V are unitary, and where the only nonzero elements of D are the first r elements of the leading diagonal, which we may arrange so that $d_1 \geqq d_2 \geqq \cdots \geqq d_r > 0$. Now let $Vy = x$ and $Uf = g$. Then

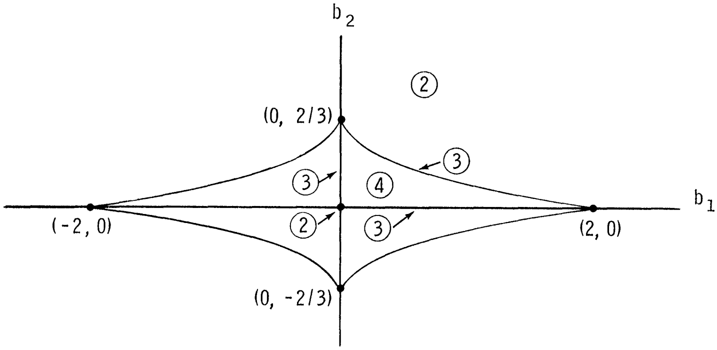

FIG. 1

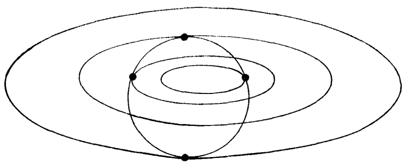

FIG. 2

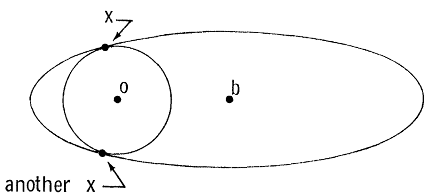

FIG. 3

$$Cy - f = U^H DVy - U^H Uf = U^H(Dx - g).$$

Hence

$$\| Cy - f \|^2 = \| Dx - g \|^2 = \sum_{i=1}^{r} | d_i x_i - g_i |^2 + \sum_{i=r+1}^{m} | g_i |^2. \tag{7.4}$$

The problem (7.1)–(7.2) is to minimize

$$\sum_{i=1}^{r} | d_i x_i - g_i |^2 = \sum_{i=1}^{r} d_i^2 \left| x_i - \frac{g_i}{d_i} \right|^2,$$

subject to the constraint

$$\sum_{i=1}^{n} | x_i |^2 = 1. \tag{7.5}$$

Now let

$$\lambda_i = \begin{cases} 0 & \text{if } i = 1, 2, \cdots, n-r, \\ d_{n+1-i}^2 & \text{if } i = n - r + 1, \cdots, n, \end{cases}$$

and let

$$b_i = \begin{cases} 0 & \text{if } i = 1, 2, \cdots, n-r, \\ \dfrac{g_{n+1-i}}{d_{n+1-i}} & \text{if } i = n - r + 1, \cdots, n. \end{cases}$$

We then have changed our problem to one of minimizing

$$\sum_{i=1}^{n} \lambda_i | x_i - b_i |^2, \tag{7.6}$$

subject to the constraint (7.5), where

$$0 = \lambda_1 = \cdots = \lambda_{n-r} < \lambda_{n-r+1} \leqq \cdots \leqq \lambda_n . \tag{7.7}$$

This is precisely the minimum problem (1.2) of §1. The special role of the $n - r$ zero eigenvalues of $C^H C$ is evident.

Thus the general problem of the least squares solution of $Cy = f$ with constraint (7.2) is a special case of our minimum problem (1.2).

8. Lemmas from complex function theory. In this final section we state and prove three lemmas relating partial derivatives of certain regular analytic functions of several complex variables to gradients of real-valued functions of vector variables. This technique is common in the study of second-order partial differential equations; for example, see [3]. We include the material mainly to keep our treatment self-contained, and partly to call explicit attention to the fact that the Lagrange multiplier λ must be real even though complex variables are used.

(8.1) LEMMA. *Let $\Phi(\lambda, \mu)$ be a regular analytic function of two complex variables λ, μ such that, for all real x, y,*

$$F(x, y) = \Phi(x + iy, x - iy) \tag{8.2}$$

is real-valued. Then

$$\frac{\partial F}{\partial x} + i \frac{\partial F}{\partial y} = 2 \left. \frac{\partial \Phi}{\partial \mu} \right|_{\substack{\lambda = x+iy \\ \mu = x - iy}} .$$

Proof. Differentiate (8.2):

$$\frac{\partial F}{\partial x} = \frac{\partial \Phi}{\partial \lambda} \cdot 1 + \frac{\partial \Phi}{\partial \mu} \cdot 1 \tag{8.3}$$

$$\frac{\partial F}{\partial y} = \frac{\partial \Phi}{\partial \lambda} \cdot i - \frac{\partial \Phi}{\partial \mu} \cdot i. \tag{8.4}$$

Add (8.3) to (8.4) multiplied by i, getting the desired result.

(8.5) LEMMA. *Let F and G be real-valued differentiable functions of real variables $x_1, y_1, \cdots, x_n, y_n$. For abbreviation, let $z_k = x_k + iy_k$, and let $z = (z_1, \cdots, z_n)$. Then, for $F(z)$ to be stationary at $z = a$ with respect to all neighboring z such that $G(z) = G(a)$, it is necessary and sufficient that there exist a* real *Lagrange constant λ such that*

$$\frac{\partial F}{\partial x_k} + i\frac{\partial F}{\partial y_k} - \lambda\left(\frac{\partial G}{\partial x_k} + i\frac{\partial G}{\partial y_k}\right) = 0 \tag{8.6}$$

for $z = a$ and $k = 1, \cdots, n$.

Proof. Condition (8.6) is nothing but the usual condition that the real gradient vector

$$\left(\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial y_1}, \frac{\partial F}{\partial x_2}, \frac{\partial F}{\partial y_2}, \cdots, \frac{\partial F}{\partial x_n}, \frac{\partial F}{\partial y_n}\right)$$

be parallel to the vector

$$\left(\frac{\partial G}{\partial x_1}, \frac{\partial G}{\partial y_1}, \frac{\partial G}{\partial x_2}, \frac{\partial G}{\partial y_2}, \cdots, \frac{\partial G}{\partial x_n}, \frac{\partial G}{\partial y_n}\right).$$

The use of the complex variables z_k is unessential.

Given any vector $z = (z_1, \cdots, z_n)$, we let $\bar{z}$ denote the vector of complex conjugates $(\bar{z}_1, \cdots, \bar{z}_n)$.

(8.7) LEMMA. *Let $\Phi(z, w)$* and *$\Psi(z, w)$ be regular analytic functions of the two complex vector variables $z = (z_1, \cdots, z_n)$ and $w = (w_1, \cdots, w_n)$ with the property that $\Phi(z, \bar{z})$ and $\Psi(z, \bar{z})$ are real. Then $\Phi(z, \bar{z})$ is stationary at $z = a$ with respect to all z such that $\Psi(z, \bar{z}) = \Psi(a, \bar{a})$ if and only if there exists a* real *Lagrange constant λ such that*

$$\frac{\partial \Phi}{\partial w_k} - \lambda\frac{\partial \Psi}{\partial w_k} = 0$$

for $z = a$ and $w = \bar{a}$ and $k = 1, 2, \cdots, n$.

Proof. Let $z = x + iy$. Then $\Phi(z, \bar{z}) = F(x, y)$, $\Psi(z, \bar{z}) = G(x, y)$. By (8.1) applied to each variable z_k,

$$\frac{\partial \Phi}{\partial w_k} = \frac{\partial F}{\partial x_k} + i\frac{\partial F}{\partial y_k} \quad \text{and} \quad \frac{\partial \Psi}{\partial w_k} = \frac{\partial G}{\partial x_k} + i\frac{\partial G}{\partial y_k}$$

for $z = a$, $w = \bar{a}$, and $k = 1, \cdots, n$. Then (8.7) follows from (8.5).

REFERENCES

[1] A. V. Balakrishnan, *An operator theoretic formulation of a class of control problems and a steepest descent method of solution*, J. SIAM Control Ser. A, 1 (1963), pp. 109–127.

[2] Philip Davis, *An application of doubly orthogonal functions to a problem of approximation in two regions*, Trans. Amer. Math. Soc., 72 (1952), pp. 104–137.

[3] P. R. Garabedian, *Partial Differential Equations*, John Wiley, New York, 1964, Chap. 3.

[4] Donald W. Marquardt, *An algorithm for least-squares estimation of nonlinear parameters*, this Journal, 11 (1963), pp. 431–441.

[5] David D. Morrison, *Methods for nonlinear least squares problems and convergence proofs*, Proc. Jet Propulsion Laboratory Seminar, 1960, pp. 1–9.

[6] David L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind*, J. Assoc. Comput. Mach., 9 (1962), pp. 84–97.

[7] S. Twomey, *On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature*, Ibid., 10 (1963), pp. 97–101.