Transactions of the Illinois State Academy of Science
(1983), Volume 76, 1 and 2, pp. 153-160

A NEW LOOK AT THE EIGENVALUE PROBLEM FOR REAL SYMMETRIC MATRICES

by

John Gregory and Don Redmond

Section 1. Introduction. In this paper, we will take a new look at the eigenvalue problem for real symmetric matrices. Although the basic results are well-known, we will present ideas and methods that are not usually associated with this subject nor with each other. Our goal is to present these ideas using elementary techniques.

We begin with some preliminary notation and results. We assume that the reader is familiar with the definition of $H = \mathbb{R}^n$ as a real vector space. We have the usual inner product on H: if $x = (a_1, \ldots, a_n)$ and $y = (b_1, \ldots, b_n)$, then

$$(x,y) = \sum_{j=1}^{n} a_j b_j = y^T \cdot x,$$

where y^T is the transpose of the vector y. This induces the norm on H

$$\|x\| = \sqrt{(x,x)}.$$

A is a real symmetric matrix if $a_{ij} = a_{ji}$. In this case the associated quadratic form is $Q(x) = (Ax,x) = x^T Ax$, for $x \in H$. The associated bilinear form is $Q(x,y) = (Ax,y) = y^T Ax$. It is immediate $Q(x,y) = Q(y,x)$ for x,y in H and that there is a one to one, onto mapping (isomorphism) between the collection of symmetric matrices and the collection of quadratic forms.

The principal result we will need in what follows in the so-called principal axiom theorem. Recall that an orthogonal matrix P is one satisfying $P^{-1} = P^T$. Theorem 1 may be found in [2, p. 266].

THEOREM 1. If A is a real symmetric matrix, then there exists an orthogonal matrix P such that $P^T AP = D$, where D is a real diagonal matrix.

Note that if $D = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$, then $\lambda_2, \ldots, \lambda_n$ are the eigenvalues of A. In terms of the associated quadratic form this gives, if $A = (a_{ij})$

$$\sum_{i,j=1}^{n} a_{ij}x_i x_j = Q(x) = (Ax,x) = x^T Ax = x^T PDP^T x = (P^T s)^T D(P^T x) = y^T dy = \sum_{i=1}^{n} \lambda_i y_i^2.$$

Thus, by a change of basis, which in this case amounts to a rotation of the axes, we can obtain diagonal quadratic forms, providing we know the eigenvalues of the matrix A.

Section 2. Rayleigh-Ritz Method. One of the standard techniques for obtaining the eigenvalues of a real symmetric matrix A, or the corresponding quadratic form Q(x), is the Rayleight-Ritz procedure. The Rayleigh quotient of A is defined to be

$$R(x) = \frac{(Ax,x)}{(x,x)} = \frac{Q(x)}{\|x\|^2} \quad \text{for } x \neq 0.$$

The eigenvalues of A (or Q(x)), then turn out to be maximums (or as we shall see minimum) of R(x) taken over an appropriate subspace of $\mathbb{R}^n - \{0\}$. Indeed we have the following result. (See [1, pg. 78]).

THEOREM 2. If A is a symmetric matrix, there exists an orthonormal set of eigenvectors $\{x_1, x_2, \ldots, x_n\}$ of A such that the corresponding eigenvalue $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ satisfy $\lambda_k = R(x_k)$ and $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. The vector x_1 minimizes R(x) on $\mathbb{R}^n - \{0\}$ and the vector x_n maximizes R(x) on $\mathbb{R}^n - \{0\}$. For each k such that $1 < k < n$, the vector x_k minimizes R(x) on the set

$$B_k = \{x \in \mathbb{R}^n : (x,x_i) = 0,\ 1 \leq i \leq k-1\} - \{0\}$$

$$= \text{span}\{x_k, \ldots, x_n\} - \{0\}.$$

Similarly x_k maximizes R(x) on the set

$$C_k = \{x \in \mathbb{R}^n : (x,x_i) = 0,\ k+1 \leq i \leq n\} - \{0\}$$

$$= \text{span}\{x_1, \ldots, x_{k-1}\} - \{0\}.$$

Finally, if D^k denotes a subspace of dimension k without $x = 0$ and $\mathcal{D}^k$ denotes the collection of all such spaces, then λ_k satisfies the following (min-max or max-min) principal

$$(1) \qquad \lambda_k = \min_{\mathcal{D}^k}\{\max_{x \in D^k} R(x)\} = \max_{\mathcal{D}^{n-k+1}}\{\min_{x \in D^{n-k+1}} R(x)\}.$$

This last equality, (1), seems particularly formidable. Since B_k is one of the sets D^{n-k+1}, we have, by an earlier part of the theorem

$$\min_{x \in D^{n-k+1}} R(x) \leq \min_{x \in B_k} R(x) = \lambda_k,$$

but since λ_k is "obtained" we may maximize both sides to obtain the max-min equality in (1). The first equality in (1) follows similarly or one may follow the suggestion of Hestenes [1] and apply this argument to the eigenvalues of $-A$, which are $-\lambda_n \leq \cdots \leq -\lambda_1$.

We can give an easy heurestic proof of Theorem 2 as follows. By Theorem 1 we may diagonalize A, that is, we may find an orthogonal matrix P such that

$$\Lambda = P^T A P = \mathrm{diag}(\lambda_1, \ldots, \lambda_n),$$

where we have assumed $\lambda_1 \leq \ldots \leq \lambda_n$. If $e_1, \ldots, e_n$ are the standard unit vectors, then they are the eigenvectors of Λ and we hence $Q(e_k) = \lambda_k$. If $x = (c_1, \ldots, c_n)^T$ is such that

$$\|x\|^2 = \sum_{i=1}^{n} c_i^2 = 1,$$

then $R(x) = \sum_{i=1}^{n} \lambda_i c_i^2$ has the smallest value, when $c_1 = 1$ and $c_k = 0$ for $k = 2, \ldots, n$. To see this, note that $R(x)$ is a convex combination of the $\{\lambda_i\}$ and

$$S = \{(\lambda_1 d_1, \ldots, \lambda_n d_n) \ : \ d_i \geq 0, \ \sum_{i=1}^{n} d_i = 1\}$$

is the face or the intersection of the $n - 1$ dimensional hyperplane determined by $\{\lambda_i\}$ and the positive "octant" in n space. In the above $c_i^2 = d_i$, $1 \leq i \leq n$. In fact $R(x)$ is the sum of the coordinates on the S above. In linear programming terminology we wish to minimize $f(t) = t_1 + \ldots + t_n$, where $t = (t_1, \ldots, t_n)^T$ is in S. Since S is a convex set and f is linear, the minimum value of f on S exists and occurs at "corner" points of S. An immediate calculation gives the desired result since the extreme points are on the coordinate axes.

Figure 1 illustrates these ideas with $n = 3$ and $\lambda_1 > 0$, since the geometric ideas hold under the translation $s_i = |\lambda_n| + t_i$. In this case, $\min\{R(x) \ : x \in S\} = \lambda_1$ as stated. Similarly (in this picture) S_1 in the set $\{(0, a_1, a_2)\}$, where (a_1, a_2) lie on the line segment connecting λ_2 and λ_3. More generally the set S above becomes an $n - 2$ dimensional "face"

$$S_1 = \{x \in S : d_1 = 0\}$$

for the "λ_2-problem." Once again $\lambda_2 = \min\{R(x) \ : x \in S_1\}$. The similar argument holds for λ_k, $2 \leq k \leq n$, by constructing a collection $S, S_1, S_2, \ldots, S_{n-1}$ where each S_{i+1} is the "positive" edge of S_i. In Figure 1, S_1 is the positive edge of the face S.

Note that if a is a nonzero real number, then

$$R(ax) = Q(ax)/\|ax\|^2 = \frac{a^2 Q(x)}{a^2 \|x\|^2} = R(x).$$

Thus R is homogeneous of degree 0. Pick D^k in $\overline{D}^k$. We may search for optimal values of the component function on S (by the homogeneity of R). The value max $R(x)$, x in D^k, is taken on at the intersection point $\underline{P}$ of an edge S and D^k, since, as above, extremal values of a linear function cannot occur in the interior of S_1. By the above argument, if $\underline{P}$ is not a corner or extreme point, it is not optimal. The minimum of such values occurs when D^k = space $\{e_1, \ldots, e_n\}$.

Finally, the problems of finding optimal values of A and Λ are

are equivalent, since P is one to one and preserves lengths. For example, $e_1 = (1, 0, \ldots, 0)^T$ satisfies

$$\lambda_1 = (Ae_1, e_1) = \max_{x\neq 0} \frac{(Ax,x)}{(x,x)} = \max_{x\neq 0} \frac{(P^TAx,Px)}{(x,x)} = \max_{x\neq 0} \frac{(APx,Px)}{(x,x)}$$

$$= \max_{y\neq 0} \frac{Ay,y}{(P^Ty,P^Ty)} = \max_{y\neq 0} \frac{(Ay,y)}{(y,y)}$$

since $(P^Ty,P^Ty) = (y, PP^Ty) = (y,y)$. Thus Theorem 2 is a "rotation", or more correctly an isometry (distance preserving mapping) of an apparent geometrical picture.

The remainder of this section is partly out of order developmentally with Theorem 2. That is, we might have placed this material before Theorem 2. We have included this material to show interesting calculations and concepts for quadratic forms. The following theorem is stated in Hestenes [1, p. 73]. Of interest are both the results and the computations involved in obtaining these results.

THEOREM 3. A vector x is an eigenvector of A if and only if it is a critical point of R(x). The eigenvalues of A are the corresponding critical values.

Let x_0 be a critical point of R(x) with $\lambda = R(x_0) = Q(x_0)/\|x_0\|^2 = (Ax_0,x_0)/(x_0,x_0)$ the critical value. Then for any vector and $\varepsilon > 0$ and small we have

$$\frac{1}{\varepsilon}\left[\frac{Q(x_0 + \varepsilon y)}{\|x_0 + \varepsilon y\|^2} - \frac{Q(x_0)}{\|x_0\|^2}\right] =$$

$$= \frac{1}{\varepsilon} \; \frac{\|x_0\|^2[Q(x_0) + 2\varepsilon Q(x_0,y) + Q(y)]-Q(x_0)\;[\|x_0\|^2 + 2\varepsilon(x_0,y) + \|y\|^2]}{\|x_0 + \varepsilon y\|^2\|x_0\|^2}$$

$$=2\;\frac{\|x_0\|^2Q(x_0,y) - Q(x_0)\;(x_0,y)}{\|x_0 + \varepsilon y\|^2\|x_0\|^2} + \varepsilon\;\frac{\|x_0\|Q(y) - Q(x_0)\|y\|^2}{\|x_0 + \varepsilon y\|^2\|x_0\|^2}$$

Taking the limit as $\varepsilon \to 0$ and noting that $Q(x_0) = \lambda\|x_0\|^2$ we have

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\,[R(x_0 + \varepsilon y) - R(x_0)\,] = \frac{2}{\|x_0\|^2}\,[Q(x_0,y) - \lambda(x_0,y)\,].$$

This limit is zero for all y if and only if $Q(x_0,y) - (x_0,y) = 0$ or $(Ax_0 - \lambda x_0, y) = 0$ for all y, i.e., $Ax_0 = \lambda x_0$. Thus our result follows.

Some comments are in order. We have $R(x + \varepsilon y) = R(x) + R'(x,\varepsilon y) + \frac{1}{2}R''(x,\varepsilon y) + \ldots$, where $R'(x,h) = \nabla R(x)h$ and $R''(x,h) = h^TAh$. R' is linear in its second argument and equal to $\nabla R\cdot(\varepsilon y)$ so that subtracting

$R(x)$ from both sides, dividing by ε and letting $\varepsilon \to 0$ we have $R'(X_0,y) = \frac{2}{\|x_0\|^2} [Q(x_0,y) - \lambda(x_0,y)]$ or the value of the gradient is $\nabla R(x_0) = \frac{2}{\|x_0\|^2} [Ax_0 - R(x_0)x_0]$. Locally at $x = x_0$, $R(x + \varepsilon y) - R(x)$ is linear in y if x_0 is not an eigenvector of A. If x_0 is an eigenvector of A, then this expression is locally quadratic in y.

It is illustrative to use elementary calculus to obtain the (first and) second directional derivatives of the Rayleigh quotient $R(x)$ and the Taylor series expansion as in (3). This will enable us to independently derive a "stronger" result than in Theorems 2 and 3. The critical point x_k is an (ℓ,m) saddle point of $R(x)$ if there exists subspaces S_1 and S_2 of $\mathbb{R}^n$ of dimension ℓ and m respectively such that $y_1 \neq 0$ in S_1 and $y_2 \neq 0$ in S_2 imply that there exists $\delta > 0$ such that $|\varepsilon| < \delta$ implies $R(x_k + \varepsilon y_1) < R(x_k) < R(x_k + \varepsilon y_2)$. The above means that locally we may decrease the critical value $\lambda_k = R(x_k)$ by "moving" from $x = x_k$ in an ℓ dimensional direction and increase $R(x_k)$ by "moving" from $x = x_k$ in an m dimensional direction In Theorem 4 we show that we may choose $S_1 = \text{span}\{x_1, \ldots, x_\ell\}$ and $S_2 = \text{span}\{x_{n-m+1}, \ldots, x_n\}$.

THEOREM 4. Let $\lambda_1 \leq \lambda_2 \leq \ldots < \lambda_n$ be the eigenvalues of A with corresponding eigenvectors $x_1, x_2, \ldots, x_n$ respectively. If for some k $(1 < k < n)$ we have $\lambda_1 < \lambda_k < \lambda_n$, then the critical point x_k is a saddle point of the Rayleigh quotient (neither a local maximum or minimum). More precisely, if $\lambda_\ell < \lambda_k < \lambda_{n-m+1}$ then x_k is an (ℓ,m) saddle point. Finally λ_1 and λ_n are respectively the absolute minimum and maximum of $R(x)$ on $\mathbb{R}^n - \{0\}$.

Let $h(\varepsilon) = Q(x + \varepsilon y)$, $g(\varepsilon) = \|x + \varepsilon y\|^2$, and $R(x + \varepsilon y) = f(\varepsilon) = h(\varepsilon)/g(\varepsilon)$. Now $f(\varepsilon) = f(0) + \varepsilon f'(0) + \frac{1}{2}\varepsilon^2 f''(0) + \ldots$ where $f'(\varepsilon) = [g(\varepsilon)h'(\varepsilon) - h(\varepsilon)g'(\varepsilon)]/g^2(\varepsilon)$ and

$$f''(\varepsilon) = \frac{g^2(\varepsilon)[g'(\varepsilon)h'(\varepsilon) + g(\varepsilon)h''(\varepsilon) - h'(\varepsilon)g'(\varepsilon) - h(\varepsilon)g''(\varepsilon)] - \{\ \}}{g^4(\varepsilon)}.$$

We have not bothered to determine $\{\ \}$ since it is zero when $\varepsilon = 0$ (a critical point). Thus $f'(0) = [g(0)h'(0) - h(0)g'(0)]/g^2(0)$ and $f''(0) = [g(0)h''(0) - h(0)g''(0)]/g^2(0)$.

The first and second derivatives are found by the Taylor series expansion. Thus $h(\varepsilon) = Q(x + \varepsilon y) = Q(x) + 2\varepsilon Q(x,y) + \varepsilon^2 Q(y) = h(0) + \varepsilon h'(0) + \frac{1}{2}\varepsilon^2 h''(0)$ so that $h'(0) = 2Q(x,y)$ and $h''(0) = 2Q(y)$; similarly (or replacing A by I) we have $g'(0) = 2(x,y)$ and $g'(0) = 2\|x\|^2$. Thus the critical points of $R(x)$ are when $f'(0) = 0$. Letting $x = x_0$ be a critical point with critical value $\lambda_0 = R(x_0)$ we obtain $0 = g(0)h'(0) - h(0)g'(0) = 2\|x_0\|^2 Q(x_0,y) - Q(x_0)(x_0,y) = 2\|x_0\|^2 [Q(x_0,y) - \lambda_0(x_0,y)]$. Since y is arbitrary we obtain, as above, in Theorem 3, that $f'(0) = 0$ or equivalently $Ax_0 = \lambda_0 x_0$ if and only if (λ_0,x_0) is a critical solution of the Rayleigh quotient.

At a critical solution we have $f(\varepsilon) = f(0) + \frac{1}{2}\varepsilon^2 f''(0) + \ldots$ so that (to second order)

$$R(x_0 + \varepsilon y) \cong R(x_0) + \frac{1}{2}\varepsilon^2 \left[\frac{\|x_0\|^2 (2Q(y)) - Q(x_0)(2\|y\|^2)}{\|x_0\|^4} \right].$$

Thus $R(x_0 + \varepsilon y) \cong R(x_0) + \varepsilon^2 (Q(y) - \lambda_0 \|y\|^2)/\|x_0\|^2 = R(x_0) + \varepsilon^2((A - \lambda_0) y,y)/\|x_0\|^2$. If (x_k, x_k) is an eigensolution of A with $\lambda_1 < \lambda_k < \lambda_n$ then $Q_1(y) = Q(y) - \lambda_k(y,y)$ satisfies $Q_1(x_1) = Q(x_1) - \lambda_k\|x_1\|^2\ R(x_1)\|x_1\|^2 - \lambda_k\|x_1\|^2 = (\lambda_1 - \lambda_k)\|x_1\|^2 < 0$ while $Q_1(x_1) = Q(x_n) - \lambda_k\|x_n\|^2 = (\lambda_n - \lambda_k)\|x_n\|^2 > 0$. Hence for ε small, $(R(x_0 + \varepsilon x_1) < R(x_0) < R(x_0 + \varepsilon x_n)$. The next to last sentence of the theorem about (ℓ,m) saddle points follows by direct computation and the fact that $Q(x_p,x_q) = 0$ if $p \neq q$. Thus if $y = \sum_{m=1}^{\ell} a_m x_m$ then

$$Q_1(y) = Q(y) - \lambda_k(y,y) = \sum_{m=1}^{\ell} [a_m^2 Q(x_m) - \lambda_k a_m^2 \|x_m\|^2]$$

$$= \sum_{m=1}^{\ell} a_m^2 (\lambda_m - \lambda_k)\|x_m\|^2 < 0.$$

The last statement of the Theorem 4 follows by advanced calculus ideas. The minimum value of $R(x)$ on the unit disc $C = \{x \mid \|x\| = 1\}$ in $\mathbb{R}^n$ is obtained since C is compact and $R(x)$ is continuous. The unit eigenvectors are the only critical points on C so that λ_1 is the minimum value of $R(x)$ on C. Similarly λ_n is the maximum value of $R(x)$ on C. The last sentence of the theorem now follows as $R(x)$ is homogeneous of degree zero.

As an example let $A = \text{diag}\{-1, 1, 2, 2\}$ with eigenvalues $\lambda_1 = -1$, $\lambda_2 = 1$, $\lambda_3 = 2$ and $\lambda_4 = 2$. Let $\{e_i\}$ be the associated standard eigenvectors where $e_1 = (1, 0, 0, 0)^T$ etc. Then e_2 is a (1,2) saddle point since $\lambda_1 < \lambda_2 < \lambda_3$ and $n - k + 1 = 4 - 2 + 1 = 3$. If $S_1 = \text{span}\{e_1\}$ and $S_2 = \text{span}\{e_3,e_4\}$, then $y_1 \neq 0$ in S_1 and $y_2 \neq 0$ in S_2 implies $R(e_2 + \varepsilon(ae_1)) \cong R(e_2) + \varepsilon^2 a^2 R(e_1) = 1 - \varepsilon^2 a^2$ and $R(e_2 + \varepsilon(be_1 + ce_2)) = R(e_2) + \varepsilon(b^2 + c^2)2$ so that $R(e_2 + \varepsilon(ae_1)) < 1 = R(e_2) < R(e_2 + \varepsilon(be_1 + ce_2))$. As above, if the reader believes our example with a diagonal matrix is too special, since $P^T AP = \Lambda$ or $A = P\Lambda P^T$, the reader may make up his own matrix A with "diagonal form" Λ. That is, for any orthogonal matrix 0, form $A = 0\Lambda 0^T$ and $x_k = 0e_k$, the k^{th} column of 0.

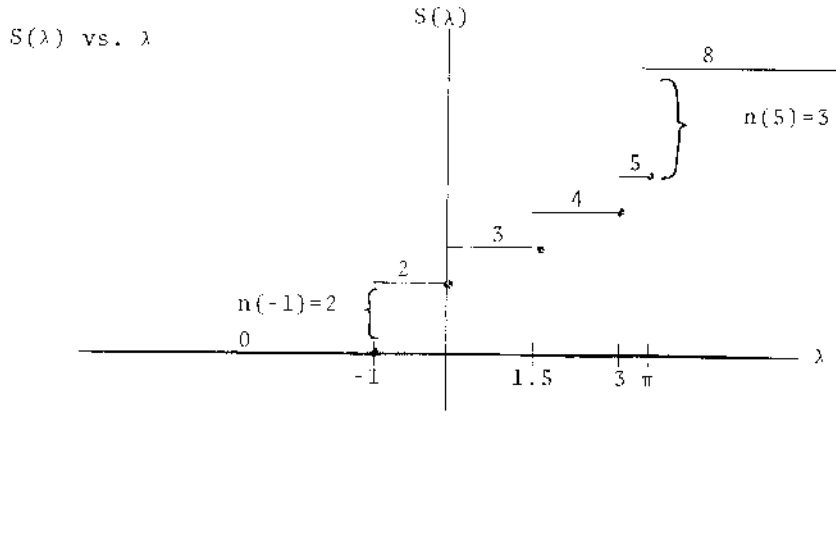
Section 3. Signature. The purpose of this section is to give an alternate definition of eigenvalues of a real symmetric matrix A or quadratic form $Q(x) = (Ax,x)$. For many problems this definition is more practical than the usual definition of $Ax = \lambda x$. Thus, for example, it is often more practical to think of a zero of a continuous real valued function $f(t)$ as a value such that $f(t_0+)\ f(t_0-)$ is negative. In this example this is not an equivalent definition, but in our case the definitions are equivalent. It also contains the Rayleigh quotient ideas but is easier to apply. Finally this definition involves the signature idea contained above.

Let A be a real symmetric matrix and $Q(x) = (Ax,x)$ be the associated quadratic form. Let $s(\lambda)$ denote the signature of the quadratic form $J(x;\lambda) = Q(x) - \lambda\|x\|^2$. That is, $s(\lambda)$ is the dimension of a maximal subspace C of $\mathbb{R}^n$ such that $x \neq 0$ in C implies $J(x;\lambda) < 0$. Note that $\lambda_1 < \lambda_2$ implies that $J(x;\lambda_2) - J(x;\lambda_1) = Q(x) - \lambda_2\|x\|^2 - (Q(x)-\lambda_1\|x\|^2) = (\lambda_1 - \lambda_2)\|x\|^2 < 0$ if $x \neq 0$. Thus $J(x;\lambda_2) \leq J(x;\lambda_1)$ and so $x \neq 0$, $J(x;\lambda_1) < 0$, implies $J(x;\lambda_2) < 0$. Thus $s(\lambda_2) \geq s(\lambda_1)$, i.e., $s(\lambda)$ is a nondecreasing, nonnegative, integer valued function. Assume as above that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ and $x_1, x_2, \ldots, x_n$ are respectively the n eigenvalues and eigenvectors of A or $Q(x) = (Ax,x)$. If $\lambda^* < \lambda_1 = \min R(x)$ for $x \neq 0$, then $Q(x)/\|x\|^2 > \lambda_1$ or $J(x;\lambda^*) = Q(x) - \lambda^*\|x\|^2 > (\lambda_1 - \lambda^*)\|x\|^2 > 0$. That is $J(x;\lambda^*)$ is positive definite and $s(\lambda^*) = 0$. Similarly if $\hat{\lambda} > \lambda_n = \max R(x)$ for $x \neq 0$, then $Q(x)/\|x\|^2 < \lambda_n$ or $J(x;\hat{\lambda}) = Q(x) - \hat{\lambda}\|x\|^2 < (\lambda_n - \hat{\lambda})\|x\|^2 < 0$. Thus $J(x;\hat{\lambda})$ is negative definite and $s(\hat{\lambda}) = n$. The reader may verify as an exercise that the intermediate eigenvalues between λ_1 and λ_n behave as we expect. Thus

THEOREM 5. $s(\hat{\lambda})$ is a nonnegative, nondecreasing, integer value function of λ. It is continuous from the left, i.e., $s(\lambda_0 - 0) = s(\lambda_0)$ and its jump at $\lambda = \lambda_0$ equal to the number of eigenvalues equal to λ_0, i.e., $s(\lambda + 0) = s(\lambda_0) + n(\lambda_0)$ where $n(\lambda_0)$ is the number of eigenvalues equal to λ_0. Finally $s(\lambda_0) = \sum_{\lambda<\lambda_0} n(\lambda)$.

In future work we will refer to $n(\lambda_0)$ as the nullity of $J(x;\lambda_0)$. In this case it is the dimension of the null space $N(\lambda) = \{x \in \mathbb{R}^n \mid J(x,y;\lambda_0) = 0$ for all y in $\mathbb{R}^n\}$. Note that $N(\lambda) \neq \{0\}$ for exactly n values of λ counting multiplicity. By Theorem 2, this space is the span of the eigenvectors corresponding to the eigenvalues equal to λ_0 since x in $N(\lambda)$ implies $0 = (Ax,y) - \lambda_0(x,y) = ((A-\lambda_0 I)x,y)$ for all y in $\mathbb{R}^n$. For future work we note that $m(\lambda) = s(\lambda) + n(\lambda)$ is the dimension of a maximal subspace E of $\mathbb{R}^n$ for which $J(x;\lambda) \leq 0$ on E. This nonincreasing integer valued function is continuous from the right with $m(\lambda_0) - m(\lambda_0 - 0) = n(\lambda_0)$.

We also have an interesting comparison result. Note for $x \neq 0$, $J(x;\lambda) = Q(x) - \lambda\|x\|^2 < 0$ implies $Q(x) < \lambda\|x\|^2$. Thus $s(\lambda)$ gives the dimension of the subspace for which $Q(x)$ is less than $\lambda\|x\|^2$. This concept can be generalized if we replace $\|x\|^2$ by $K(x) = (Bx,x)$ where B is symmetric.



As an example to illustrate signature, if $A = P\Lambda P^t$ where Λ = diag{-1, -1, 0, 1.5, e, π, π, π} then the graph of $s(\lambda)$ is given in Figure 4. If $\lambda < -1$ then $J(x;\lambda)$ is positive definite.

Section 4. Lagrange Multipliers. We may also think of eigenvalues as Lagrange multipliers. It is of interest to show that the eigenvalues can be determined solely by a process that maximizes quadratic forms over the unit sphere in $\mathbb{R}^n$ and not on proper subsets of $\mathbb{R}^n$ such as constraints $L(x) = (x,x_1) = 0$ as in the previous work. Once again we assume that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the eigenvalues of A or Q(x) with corresponding orthonormal vectors $x_1, x_2, \ldots, x_n$, respectively. By translation we may assume that $\lambda_1 > 0$, since $A + (|\lambda_n| + 1)I$ has the same eigenvectors as A with eigenvalues translated by $|\lambda_n| + 1$. This assumption is not required in our proof by construction, but avoids some technical difficulties.

We know that $\lambda_n = \max Q(x) = Q(x_n)$ where the maximum is taken over $C = \{\|x\| = 1\}$ the unit ball in $\mathbb{R}^n$. Notice that $K(x,y) = (x,x_n)(y,x_n)$ is a bilinear form in x and y and $K(x) = K(x,x) = (x,x_n)^2$ is quadratic. We have $K(x + \varepsilon y) = (x + \varepsilon y,x_n)^2 = [(x,x_n) + \varepsilon(y,x_n)]^2 = (x,x_n)^2 + 2\varepsilon(x,x_n)(y,y_n) + \varepsilon^2(y,x_n)^2 = K(x) + 2\varepsilon K(x,y) + \varepsilon^2 K(y)$. Let $Q_1(x) = Q(x) - \lambda_n(x,x_n)^2$. The eigenvalues of $Q_1(x)$ are $\lambda_1 = 0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n-1}$ with corresponding eigenvectors $x_n, x_1, x_2, \ldots, x_{n-1}$. One way of seeing this is that $Q_1(x_i,x_j) = Q(x_i,x_j) - \lambda_n(x_i,x_n)(x_j,x_n) = (\lambda_i x_i, x_j) - \lambda_n(x_i,x_n)(x_j,x_n) = \lambda_i\delta_{in}\delta_{ij}$. This expression is zero if $i \neq j$ or if $i = j = n$. If $i = j \neq n$, then this expression is λ_i. If A were diagonal, i.e., $A = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. Then Q_1 would correspond to the real symmetric matrix $B = \text{diag}(\lambda_1, \lambda_2, \lambda_{n-1}, 0)$.

Now $\lambda_{n-1} = \max Q_1(x) = \max Q(x) - \lambda_n(x,x_n)^2 = Q_1(x_{n-1})$ where the maximum is over C and not some subset of C. Continuing in this way if $Q_2(x) = Q(x) - \lambda_{n-1}(x,x_{n-1})^2 - \lambda_n(x,x_n)^2$, then $\lambda_3 = \max Q_2(x) = Q_2(x_{n-2})$ where the maximum is over C. Finally note that we may decompose Q(x) into its finite "Fourier series"

$$Q(x,y) = \sum_{k=1}^{n} \lambda_k(x,x_k)(y,x_k).$$

Clearly at each step we have invoked a Lagrange multiplier type rule. This result is stronger than the min-max theory in that we maximize over all of C and do not restrict ourselves to certain subspaces.

References

1. M.R. Hestenes, Optimatization Theory, John Wiley and Sons, 1975.

2. K. Hoffman, R. Kunze, Linear Algebra, Prentice-Hall, 1961.

Department of Mathematics
Southern Illinois University
Carbondale, IL 62901