EIGENVALUES IN LINEAR ALGEBRA*

P. F. Leung
National University of Singapore

§1. General discussion

In this talk, we shall present an elementary study of eigenvalues in linear algebra. Very often in various branches of mathematics we may come across the problem of finding certain powers say A^k of a square matrix A. This comes up for example in the theory of Markov Chains, in the solution of a system of ordinary differential equations with constant coefficients, etc. In general, for small k we can calculate A^k directly by multiplication, but this becomes complicated for large k or for a general value k. To find another approach, we first observe that A^k can be calculated easily when A is a diagonal matrix. For example, if

$$A = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

then

$$A^k = \begin{pmatrix} \lambda_1^k & & & 0 \\ & \lambda_2^k & & \\ & & \ddots & \\ 0 & & & \lambda_n^k \end{pmatrix}$$

for any positive integer k.

---

*Text of a talk given at the Workshop on Linear Algebra and its Teaching from 9 - 10 September 1985 at the National University of Singapore.

However, since very often the matrix we come across need not be diagonal so this observation doesn't help us directly. Still if we think about it more carefully, we come to realise that even when A itself is not diagonal, but if it is *similar* to a diagonal matrix D in the sense that $A = MDM^{-1}$ for some non- singular matrix M, then we have

$$A^k = (MDM^{-1})^k$$

$$= \underbrace{(MDM^{-1})(MDM^{-1})\dots(MDM^{-1})}_{k \text{ terms}}$$

$$= MD^k M^{-1}$$

and hence A^k can again be calculated easily. Now let us spend a little time and find out something about the equation $A = MDM^{-1}$. We will represent vectors in column form and write

$$M = (m_1, m_2, \dots, m_n)$$

where m_i is the i-th column of the matrix M.

Then from $A = MDM^{-1}$, we have $AM = MD$ and so

$$(Am_1, Am_2, \dots, Am_n) = \left(M\begin{pmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, M\begin{pmatrix} 0 \\ \lambda_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, M\begin{pmatrix} 0 \\ \vdots \\ 0 \\ \lambda_n \end{pmatrix} \right)$$

and therefore we have

$$Am_i = \lambda_i m_i ,$$

since M is non-singular, therefore $m_i \neq 0$, and so we are led to study the equation

$$Ax = \lambda x$$

for $x \neq 0$.

Definition. Let A be an $n \times n$ matrix over the field of real numbers $\mathbb{R}$. A real number λ is called an *eigenvalue* of A if there is a vector x (in column form) in $\mathbb{R}^n$, $x \neq 0$, and such that $Ax = \lambda x$. In this case x is called an *eigenvector* of A corresponding to the eigenvalue λ.

From $Ax = \lambda x$, we have $A(cx) = cAx = c\lambda x = \lambda(cx)$ for any other real number c and hence we can see that any non-zero multiple of an eigenvector is again an eigenvector corresponding to the same eigenvalue and the linear subspace generated by x is mapped into itself under A. In general a subspace X is called an *invariant subspace* of A if $AX \subseteq X$.

Let us now return to the question of finding a diagonal matrix D and a nonsingular one M such that $A = MDM^{-1}$. Suppose we can find n eigenvalues $\lambda_1, \ldots, \lambda_n$ and the corresponding n eigenvectors $m_1, \ldots, m_n$ are linearly independent, then $M = (m_1, \ldots, m_n)$ is non-singular and letting

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

we see from (1) that $AM = MD$ and so $A = MDM^{-1}$.

In §2 we shall see how to find the eigenvalues and eigenvectors. In §3 we shall study one very important particular case when A is symmetric and we shall find out that in this case D and M can always be found (at least in theory) and furthermore

M can be assumed to be orthogonal (i.e. $M^{-1} = M^t$ the transpose of M) as well.

§2. Finding eigenvalues and eigenvectors

A is our $n \times n$ matrix and we use I to denote the $n \times n$ identity matrix. We have

$$Ax = \lambda x$$

$$\Leftrightarrow \quad Ax - \lambda Ix = 0$$

$$\Leftrightarrow \quad (A - \lambda I)x = 0. \qquad (2)$$

Now (2) represents a system of linear equations with coefficient matrix $A - \lambda I$ and the requirement $x \neq 0$ means that (2) has a non-trivial solution and hence we must have

$$\det(A - \lambda I) = 0 ,$$

where det denotes the determinant function. Therefore λ is a solution of the *characteristic equation*

$$\det(A - zI) = 0 . \qquad (*)$$

Conversely, if λ satisfies (*), then (2) will have a non-trivial solution x and hence $Ax = \lambda x$, $x \neq 0$. Hence finding eigenvalues of A is equivalent to solving its characteristic equation (*). We see that (*) is a polynomial equation (in z) of degree n and so A has at most n eigenvalues (counting multiplicity). Once an eigenvalue is found, an eigenvector corresponding to it can be found by solving for x in (2) using standard methods from linear algebra. Let us illustrate this procedure in some detail in an example.

Example. Our matrix A is

$$\begin{pmatrix} 2 & -3 \\ -2 & 1 \end{pmatrix}$$

From (*) the characteristic equation is

$$\begin{vmatrix} 2-z & -3 \\ -2 & 1-z \end{vmatrix} = 0 .$$

This simplifies to $z^2 - 3z - 4 = 0$,

$$\text{i.e.} \quad (z-4)(z+1) = 0$$

and hence we see that there are two eigenvalues $\lambda_1 = -1$ and $\lambda_2 = 4$.

To find an eigenvector corresponding to $\lambda_1 = -1$, we look at equation (2) for this case which is

$$\begin{pmatrix} 3 & -3 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and solve for the vector $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. This equation reduces to $x_1 = x_2$ and so

$$m_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

is an eigenvector corresponding to $\lambda_1 = -1$.

Next we do the same for $\lambda_2 = 4$ and solve

$$\begin{pmatrix} -2 & -3 \\ -2 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

which reduces to $2x_1 = -3x_2$ and hence

$$m_2 = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$

is an eigenvector corresponding to $\lambda_2 = 4$. The matrix M will be given by

$$M = \begin{pmatrix} 1 & 3 \\ 1 & -2 \end{pmatrix} ,$$

and using the formula for the inverse of a 2×2 matrix, we see that

$$M^{-1} = \begin{pmatrix} \frac{2}{5} & \frac{3}{5} \\ \frac{1}{5} & -\frac{1}{5} \end{pmatrix}$$

and we finally conclude from §1 that

$$\begin{pmatrix} 2 & -3 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} \frac{2}{5} & \frac{3}{5} \\ \frac{1}{5} & -\frac{1}{5} \end{pmatrix}$$

and hence

$$\begin{pmatrix} 2 & -3 \\ -2 & 1 \end{pmatrix}^n = \begin{pmatrix} 1 & 3 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} (-1)^n & 0 \\ 0 & 4^n \end{pmatrix} \begin{pmatrix} \frac{1}{5} & \frac{3}{5} \\ \frac{1}{5} & -\frac{1}{5} \end{pmatrix}$$

$$= \frac{1}{5} \begin{bmatrix} 2(-1)^n + 3(4^n) & 3(-1)^n - 3(4^n) \\ 2(-1)^n - 2(4^n) & 3(-1)^n + 2(4^n) \end{bmatrix}$$

§3. The spectral theorem for symmetric matrices

We shall consider the n-dimensional Euclidean space $\mathbb{R}^n$ with its usual inner product (i.e. dot product) denoted by <,> so that

for any two vectors $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ in $\mathbb{R}^n$, we have

$$\langle x,y\rangle = x_1y_1 + \dots + x_ny_n.$$

A $n \times n$ matrix $A = (a_{ij})$ is called a *symmetric matrix* if we have $\langle Ax,y\rangle = \langle x,Ay\rangle$ for any $x,y \in \mathbb{R}^n$. Since in general, we always have $\langle Ax,y\rangle = \langle x,A^ty\rangle$, a matrix is symmetric if $A = A^t$, where A^t is the transpose of A, i.e. $a_{ij} = a_{ji}$ for all $i,j = 1,\dots,n$. For the rest of this section A will denote a symmetric matrix. A matrix M is called an *orthogonal matrix* if $M^{-1} = M^t$. For an orthogonal matrix M, we have $M^tM = I$ and since the (i,j) entry in M^tM is $\langle m_i,m_j\rangle$ where as in §1, m_i, m_j denote the i-th and j-th column of M respectively, we see that the columns $m_1,\dots,m_n$ form an orthonormal basis of $\mathbb{R}^n$. Conversely, if $m_1,\dots,m_n$ form an orthonormal basis of $\mathbb{R}^n$, then $M = (m_1,\dots,m_n)$ is an orthogonal matrix. Our main result in this section is the following.

Spectral Theorem. Let A be a symmetric matrix, then there exists an orthogonal matrix M and a diagonal matrix D such that

$$A = MDM^t.$$

As a consequence, we see that A has n real eigenvalues and that there is an orthonormal basis of $\mathbb{R}^n$ relative to which the transformation A is represented in diagonal form D.

We shall present two different proofs of this very important theorem based on two different ideas both of which are of independent interest. In Part I of the following, we shall take a geometric approach to obtain a characterisation of the eigenvalues of A in terms of a maximum-minimum property and this coupled with a little analysis and linear algebra will give us an inductive proof of the spectral theorem. In Part II an idea due to Jacobi in the numerical computation of eigenvalues will be studied and as a consequence we shall obtain a very simple and elegant proof of the spectral theorem. We need two important facts from analysis :

Fact 1 : A subset of the Euclidean space is compact if and only if it is closed and bounded.

Fact 2 : A continuous real-valued function on a compact set must attain a minimum at some point in this set.

Part I

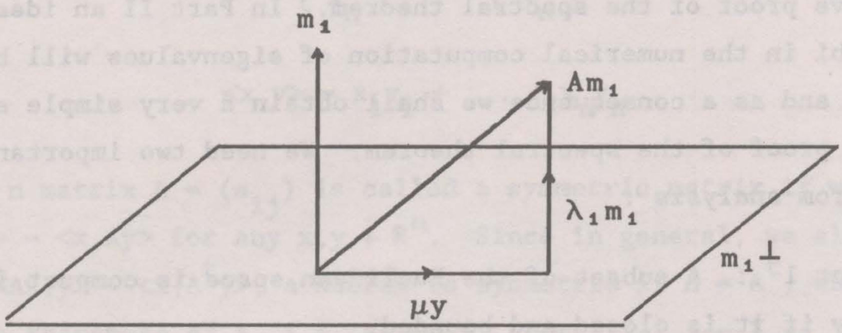
Let S^{n-1} denote the unit sphere in $\mathbb{R}^n$, i.e. $S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$ where $\|x\|^2 = \langle x,x\rangle$. Clearly S^{n-1} is closed and bounded and hence compact. Now consider the real-valued function f on S^{n-1} defined by $f(x) = \langle Ax,x\rangle$, this attains a minimum at a point m_1 say in S^{n-1} and this minimum value is $\lambda_1 = \langle Am_1, m_1\rangle$.

Claim. m_1 is an eigenvector corresponding to the eigenvalue λ_1, i.e. $Am_1 = \lambda_1 m_1$.

Proof. By using orthogonal decomposition, we can let

$$Am_1 = \lambda_1 m_1 + \mu y$$

where y is a unit vector in the orthogonal complement $m_1^{\perp}$ (see figure).



Consider now the unit circle

$$\gamma(\theta) = m_1 \cos\theta + y\sin\theta, \quad -\pi \le \theta \le \pi.$$

The restriction of f on this circle $f(\gamma(\theta))$ as a function of one variable θ attains a minimum when $\theta = 0$ and so we have

$$\frac{d}{d\theta} f(\gamma(\theta)) \Big|_{\theta=0} = 0. \tag{3}$$

Now

$$f(\gamma(\theta)) = \langle A(\gamma(\theta)), \gamma(\theta)\rangle$$

$$= \langle Am_1 \cos\theta + Ay\sin\theta,\ m_1\cos\theta + y\sin\theta\rangle$$

$$= \lambda_1 \cos^2\theta + \langle Am_1, y\rangle \sin 2\theta + \langle Ay, y\rangle \sin^2\theta\ .$$

Therefore

$$\frac{d}{d\theta} f(\gamma(\theta))\Big|_{\theta=0} = 2<Am_1,y> = 2\mu .$$

Hence from (3) we have $\mu = 0$ and therefore

$$Am_1 = \lambda_1 m_1 .$$

Now we apply a similar reasoning to the unit sphere S^{n-2} in the (n-1)-dimensional subspace $m_1^\perp$ and obtain a second eigenvalue λ_2 and a corresponding unit eigenvector m_2. Apply this again to the unit sphere in $\{m_1,m_2\}^\perp$ we obtain λ_3 and m_3 and so on by induction we then obtain all n eigenvalues $\lambda_1,\ldots,\lambda_n$ and corresponding unit eigenvectors $m_1,\ldots,m_n$ which are mutually orthogonal. So $M = (m_1,\ldots,m_n)$ is an orthogonal matrix and we have

$$A = MDM^t$$

where $D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$ which proves the spectral theorem.

The above reasoning gives us the following characterization of the eigenvalues $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$.

Minimum Principle

$$\lambda = \min_{\|x\|=1} <Ax,x>$$

and $$\lambda_k = \min_{\substack{x\perp\{m_1,\ldots,m_{k-1}\} \\ \|x\|=1}} <Ax,x>, \qquad 2 \le k \le n.$$

As a simple application of the minimum principle, we can derive a very easy upper bound for the smallest eigenvalue λ_1. We just look for the smallest element in the diagonal of A, say this is a_{ii}, then we have $\lambda_1 \le a_{ii}$. This may sometimes be useful when the size of A is large so that a direct computation of its eigenvalues is not practical, and we only need some numerical estimate on its smallest eigenvalue. For example if A is a 100×100 symmetric matrix with a -1 in a diagonal entry, then we know without any work that A cannot have all eigenvalues positive.

Supplement to Part one

For practical purposes, the minimum principle is not very useful because of the dependence of the characterization of later eigenvalues on the eigenvectors of the preceeding ones. We shall now use the minimum principle to derive a different characterization.

The Mini-Max Principle

(1) $$\lambda_1 = \min_{\|x\|=1} \langle Ax, x\rangle$$

(2) For $2 \le k \le n$ and any collection of k-1 vectors $v_1, \ldots, v_{k-1}$ we define a number α depending on $v_1, \ldots, v_{k-1}$ by

$$\alpha(v_1, \ldots, v_{k-1}) = \min_{\substack{x \perp (v_1, \ldots, v_{k-1}) \\ \|x\|=1}} \langle Ax, x\rangle$$

Then $\lambda_k = \max_{\{v_1,\ldots,v_{k-1}\}} \alpha(v_1,\ldots,v_{k-1})$ where the maximum is taken over all choices of $v_1,\ldots,v_{k-1}$.

Proof. Let $\mu_k = \max_{\{v_1,\ldots,v_{k-1}\}} \alpha(v_1,\ldots,v_{k-1})$. By minimum principle, we have

$$\lambda_k = \alpha(m_1,\ldots,m_{k-1})$$

and so

$$\lambda_k \le \mu_k.$$

Let us now take a fixed arbitrary choice of $v_1,\ldots,v_{k-1}$. We consider the vector

$$x = a_1 m_1 + \ldots + a_k m_k ,$$

where $a_1,\ldots,a_k$ are chosen such that $\langle x,v_1\rangle = \ldots = \langle x,v_{k-1}\rangle = 0$ and $a_1^2 + \ldots + a_k^2 = 1$. This is possible because $\langle x,v_1\rangle = 0,\ldots,$ $\langle x,v_{k-1}\rangle = 0$ is a system of (k-1) linear equations in k unknowns $(a_1,\ldots,a_k)$ and so a non-trivial solution can be found to satisfy $a_1^2 + \ldots + a_k^2 = 1$. Now that the conditions $x \perp \{v_1 \ldots, v_{k-1}\}$ and $\|x\| = 1$ are satisfied, we check that

$$\begin{aligned}\langle Ax,x\rangle &= \langle a_1\lambda_1 m_1 + \ldots + a_k\lambda_k m_k,\ a_1 m_1 + \ldots + a_k m_k\rangle \\ &= \lambda_1 a_1^2 + \lambda_2 a_2^2 + \ldots + \lambda_k a_k^2 \\ &\le \lambda_k a_1^2 + \lambda_k a_2^2 + \ldots + \lambda_k a_k^2 = \lambda_k .\end{aligned}$$

Hence $\alpha(v_1,\ldots,v_{k-1}) \le \lambda_k$.

Finally, since $v_1, \ldots, v_{k-1}$ is any arbitrary choice, we therefore have

$$\mu_k \leq \lambda_k ,$$

and hence $\lambda_k = \mu_k$.

Part II

We shall give another proof of the spectral theorem based on a numerical method due to Jacobi. We begin with a simple fact:

Lemma. The set O(n) of n × n orthogonal matrices is compact.

Proof. We shall realize O(n) as a closed and bounded set in $\mathbb{R}^{n^2}$ and hence its compactness follows. First we observe that the space of all n × n matrices can be identified with $\mathbb{R}^{n^2}$ because there are n^2 entries for such a matrix. Therefore we can use the norm in $\mathbb{R}^{n^2}$ to induce a norm on any n × n matrix $B = (b_{ij})$ be defining $\|B\|^2 = \Sigma b_{ij}^2$. A simple calculation shows that this is equivalent to

$$\|B\|^2 = \text{trace}(BB)^t.$$

Therefore for any $M \in O(n)$, we have

$$\|M\|^2 = \text{trace}(MM^t) = \text{trace}(I) = n$$

and so $O(n)$ is a subset of the hypersphere of radius $\sqrt{n}$ in $\mathbb{R}^{n^2}$. and hence $O(n)$ is bounded. To see that $O(n)$ is closed we take a sequence $M_i \in O(n)$ such that $M_i \to M$, then we observe that $M_i^t \to M$ and hence

$$M_i M_i^t \to MM^t$$

but $M_i M_i^t = I$ and so $MM^t = I$, i.e., $M \in O(n)$.

We shall now introduce the method to reduce a non-zero entry in the off-diagonal of a symmetric matrix A to zero, and of course, after all the off-diagonal entries are reduced to zero, we get a diagonal matrix. So let us now introduce a function to measure the deviation of any matrix from a diagonal one.

Definition. For any matrix $B = (b_{ij})$, define

$$\phi(B) = \sum_{i \neq j} b_{ij}^2 .$$

Clearly B is a diagonal matrix if and only if $\phi(B) = 0$.

Theorem (Jacobi).

Let A be a symmetric matrix. If $\phi(A) > 0$, then there exists an orthogonal matrix M such that

$$\phi(MAM^t) < \phi(A).$$

Proof. Since $\phi(A) > 0$, there is an entry, say $a_{ij} \neq 0$, with $i < j$. Let M be the $n \times n$ matrix given by

$$
M = \begin{array}{c}
\begin{array}{ccccccccc} & & & i & & & j & & \end{array} \\
\left[\begin{array}{ccccccccc}
1 & & & \vdots & & & \vdots & & \\
& \ddots & & \vdots & & & \vdots & & \\
& & 1 & \vdots & & & \vdots & & \\
\cdots & \cdots & \cdots & \cos\theta & \cdots & \cdots & \sin\theta & \cdots & \cdots \\
& & & \vdots & 1 & & \vdots & & \\
& & & \vdots & & \ddots & \vdots & & \\
& & & \vdots & & & 1 \;\vdots & & \\
\cdots & \cdots & \cdots & -\sin\theta & \cdots & \cdots & \cos\theta & \cdots & \cdots \\
& & & \vdots & & & \vdots & 1 & \\
& & & \vdots & & & \vdots & \ddots & \\
& & & \vdots & & & \vdots & & 1
\end{array}\right]
\end{array}
$$

Then it is easily checked that $MM^t = I$ and so M is orthogonal.

Now we just have to multiply out MAM^t. If we let $MAM^t = (b_{pq})$, then the result is

$$b_{pq} = b_{qp} = a_{pq} = a_{qp}\,, \qquad p \neq i,j,\; q \neq i,j$$

$$b_{pi} = a_{pi}\cos\theta + a_{pj}\sin\theta = b_{ip}\,, \qquad p \neq i,j$$

$$b_{pj} = -\,a_{pi}\sin\theta + a_{pj}\cos\theta = b_{jp}\,, \qquad p \neq i,j$$

$$b_{ii} = a_{ii}\cos^2\theta + a_{ij}\sin 2\theta + a_{jj}\sin^2\theta\,,$$

$$b_{jj} = a_{ii}\sin^2\theta + a_{ij}\sin 2\theta + a_{jj}\cos^2\theta\,,$$

$$b_{ij} = \frac{1}{2}\,(a_{jj} - a_{ii})\sin 2\theta + a_{ij}\cos 2\theta = b_{ji}\,.$$

Now we see that θ can be chosen so as to make $b_{ij} = 0$ (just solve this equation for θ) and with this choice of θ, a direct computation shows that

$$\phi(MAM^t) = \phi(A) - 2a_{ij}^2$$

and hence

$$\phi(MAM^t) < \phi(A) .$$

Proof of the spectral theorem.

A is our fixed symmetric matrix. We now consider the function $f : O(n) \to \mathbb{R}$ defined by $f(J) = \phi(J^t AJ)$. f is then a continuous function and since O(n) is compact, f attains a minimum, say at $M \in O(n)$.

Claim. $f(M) = \phi(M^t AM) = 0$.

Suppose $f(M) > 0$, let

$$D = M^t AM.$$

Then D is symmetric and $\phi(D) > 0$. Therefore by the theorem of Jacobi, there exists $N \in O(n)$ such that

$$\begin{aligned} f(M) = \phi(D) &> \phi(NDN^t) \\ &= \phi(NM^t AMN^t) \\ &= \phi((MN^t)^t A(MN^t)) \\ &= f(MN^t) \end{aligned}$$

But $M,N \in O(n)$, so $MN^t \in O(n)$ also, and this is a contradiction to the hypothesis that f(M) is a minimum.

Therefore $\phi(t) = 0$ so that D is diagonal. Since $D = M^t AM$, therefore $A = MDM^t$ and the spectral theorem is proved.