Lecture 4 Linear Algebra

We now give a third proof of the spectral theorem for Hermitian matrices. This proof appeared in Mathematics Monthly in the early 1980's.

We will consider real symmetric matrices. The Hermitian case is similar and left as an Exercise. So let A be a given $n \times n$ real symmetric matrix and consider the map

$$f : \mathcal{O} \equiv \{ n \times n \text{ orthogonal matrices} \} \to \mathbb{R}$$
$$\mathcal{O} \ni U \to f(U) = \sum_{i \neq j} \left((U A U^T)_{ij} \right)^2$$

As $\mathcal{O}$ is compact (why?) and as f is continuous, f attains its minimum at some $U_A \in \mathcal{O}$. We will show that $f(U_A) = 0$, i.e. $D \equiv U_A A U_A^T$ is diagonal, which proves the spectral theorem $A = U_A^T D U_A$.

Suppose by contradiction that $f(U_A) > 0$ and set $\tilde{A} \equiv U_A A U_A^T$. As $f(U_A) > 0$, $\tilde{A}_{ij} \neq 0$ for some

i, j. For convenience suppose $\tilde{A}_{12} \neq 0$, but the same proof works for any $i \neq j$. Let

$$R = \begin{pmatrix} c & s \\ -s & c \end{pmatrix}, \qquad c = \cos\theta,\ s = \sin\theta$$

R is orthogonal in $\mathbb{R}^2$. Set

$$V = \left(\begin{array}{c|c} R & 0 \\ \hline 0 & I \end{array}\right)$$

where I is the identity matrix matrix in $\mathbb{R}^{n-2}$. As R is orthogonal in $\mathbb{R}^2$, V is orthogonal in $\mathbb{R}^n$. Write $\tilde{A}$ in block form,

$$\tilde{A} = \begin{pmatrix} \alpha & \beta \\ \beta^T & \gamma \end{pmatrix}$$

where α is 2×2, β is $2\times(n-2)$ and γ is $(n-2)\times(n-2)$.

Now $$V\tilde{A}V^T = \begin{pmatrix} R & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} \alpha & \beta \\ \beta^T & \gamma \end{pmatrix}\begin{pmatrix} R^T & 0 \\ 0 & I \end{pmatrix}$$

$$= \begin{pmatrix} R\alpha R^T & R\beta \\ \beta^T R^T & \gamma \end{pmatrix}$$

Now $\|R\beta\|^2 = (R\beta, R\beta) = (\beta, R^TR\beta) = (\beta, \beta) = \|\beta\|^2$

It follows that $\varphi(VU_n) = \sum_{i \neq j} (V\tilde{A}V^T)_{ij}^2$

$$= \sum_{i \neq j} (\tilde{A}_{ij})^2 + 2(R\alpha R^T)_{12}^2 - 2\tilde{A}_{12}^2$$

Now we can choose θ, i.e. R, such that $R\alpha R^T$ is diagonal — this is just the spectral theorem for 2×2 matrices, which we can prove directly as follows:

$$R\alpha R^T = \begin{pmatrix} c & s \\ -s & c \end{pmatrix} \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} c & -s \\ s & c \end{pmatrix}, \quad \alpha_{12} = \alpha_{21}$$

$$= \begin{pmatrix} c\alpha_{11} + s\alpha_{21} & c\alpha_{12} + s\alpha_{22} \\ -s\alpha_{11} + c\alpha_{21} & -s\alpha_{12} + c\alpha_{22} \end{pmatrix} \begin{pmatrix} c & -s \\ s & c \end{pmatrix}$$

$$= \begin{pmatrix} x & -sc\alpha_{11} - s^2\alpha_{21} + c^2\alpha_{12} + cs\,\alpha_{22} \\ x & x \end{pmatrix}$$

$$= \begin{pmatrix} x & \alpha_{12}(c^2 - s^2) + cs(\alpha_{22} - \alpha_{11}) \\ x & x \end{pmatrix}$$

Now $g(\theta) = (\cos^2\theta - \sin^2(\theta))\alpha_{12} + (\cos\theta \sin\theta)(\alpha_{22} - \alpha_{11})$

has the property
$g(0) = \alpha_{12}$
$g(\frac{\pi}{2}) = -\alpha_{12}$

It follows by continuity that $g(\theta) = 0$ for some θ', $0 \leq \theta' \leq \pi/2$. Thus $R(\theta')\alpha R(\theta')^T$ is diagonal.

Thus $f(VU_A) = f(U_A) - 2\tilde{A}_{12}^2 < f(U_A)$

as $\tilde{A}_{12} \neq 0$. This is a contradiction and the spectral th^m follows. ☐

We now make some remarks on <u>topology</u>. Recall that a topology τ on a set X is a collection of subsets of X with the following properties

- the empty set and X are in τ
- any union of elements in τ is an element of τ
- any intersection of a finite elements of τ is an element of τ

The element of τ are called <u>open sets</u> in X. We refer to the pair (X, τ) as a <u>topological space</u>

Recall that a norm $\|\cdot\|$ on a vector space V is a map $\|\cdot\| : V \to \mathbb{R}_+$ with the following properties:
$v \mapsto \|v\|$

(1) $\|v\| \geq 0$ and $\|v\| = 0 \iff v = 0$

(2) $\|v + u\| \leq \|v\| + \|u\| \quad \forall\, u, v \in V$

(3) $\|\lambda v\| = |\lambda|\, \|v\| \quad \forall\, v \in V$ and scalar λ

A norm $\|\cdot\|$ on V induces a topology $\tau_{\|\cdot\|}$ on V with basis sets

$$O_{v,\varepsilon} = \{v' \in V : \|v' - v\| < \varepsilon\}, \quad v \in V, \varepsilon > 0.$$

A subset $B \subset V$ is in $\tau_{\|\cdot\|}$ if and only if it can be expressed as a union of basis sets.

It follows that $v_n \to v$ in $(V, \tau_{\|\cdot\|})$, iff for each $\varepsilon > 0$ $\exists N$ such that $n \geq N \Rightarrow \|v - v_n\| < \varepsilon$ i.e. $v_n \in O_{v,\varepsilon}$.

If V is finite dimensional, then all norms on V are equivalent i.e. if $\|\cdot\|$, $|||\cdot|||$ are 2 norms, then $\exists$ $c_1 > c_2 > 0$ such that

(50.1) $$c_2 |||u||| \leq \|u\| \leq c_1 |||u||| \qquad \forall u \in V$$

It follows that for any $v \in V$, $\varepsilon > 0$

$$\{v' : |||v'-v||| < \varepsilon/c_2\} \subset \{v' : \|v'-v\| < \varepsilon\} \subset \{v' : |||v'-v||| < \varepsilon/c_1\}$$

and hence $\|\cdot\|$ and $|||\cdot|||$ induce the same topology on V.

To prove (50.1), fix a basis $u_1, \ldots, u_n$ for V, $\dim V = n < \infty$. For $u = \sum_{i=1}^{n} a_i u_i$, set

$$\|u\|_0 = \sum_{i=1}^{n} |a_i|$$

As the a_i's are unique, $\|u\|_0$ is well-defined

It is easy to see (exercise) that $\|\cdot\|_0$ is a norm on V. Now to prove (50.1) it is clearly enough to show that given any norm $\|\cdot\|$, $\exists\, c_1 > c_2 > 0$ such that $c_1 \|u\|_0 \leq \|u\| \leq c_2 \|u\|_0$. Of course c_1, c_2 depend on $\|\cdot\|$.

As $\|\cdot\|$ is a norm, by properties (2), 3

$$\|u\| = \|\Sigma a_i u_i\| \leq \sum_i |a_i| \, \|u_i\| \leq \left(\max_i \|u_i\|\right) \sum_i |a_i| = c_1 \|u\|_0$$

where $c_1 = \max_i \|u_i\|$. Now consider the map

$$f : B \to \mathbb{R}_+$$

where $B \equiv \{(a_1, \ldots, a_n) : \sum_i |a_i| = 1\}$ is compact and $f(a_1, \ldots, a_n) \equiv \|\sum_i a_i u_i\|$

Now f is a continuous function on a compact set and hence f attains its minimum at some point $(a_1^0, \ldots, a_n^0)$ in B : $f(a_1^0, \ldots, a_n^0) \leq f(a_1, \ldots, a_n)$ for all $(a_1, \ldots, a_n) \in B$. Suppose $f(a_1^0, \ldots, a_n^0) = \|\sum_i a_i^0 u_i\| = 0$ Then by property (1), $\sum_i a_i^0 u_i = 0$, but then $a_i^0 = 0$, $i = 1, \ldots, n$, as as $\{u_i\}$ is a basis. But $\sum_i |a_i^0| = 1$, so this is a contradiction. Thus $f(a_i, \ldots, a_n) \geq f(a_i^0, \ldots, a_n^0) = c_2 > 0$

ie $\|\sum_i a_i u_i\| \geq c_2 \quad \forall \sum |a_i| = 1$.

For any $(b_1, \ldots, b_n) \neq 0$, set $a_i = b_i / \sum_j |b_j|$, $i = 1, \ldots, n$.

Then $\sum_i |a_i| = 1$ and so $\left\| \sum_i \frac{b_i}{\sum_j |b_j|} u_i \right\| \geq c_2$

But then by property (i), $\|\sum b_i u_i\| \geq c_2 \sum |b_i| = c_2 \|b\|_0$

$\forall (b_1, \ldots, b_n)$. This completes the proof of (50.V

There are many examples of norms. Let $u_1, \ldots, u_n$ be a basis for V and express $u = \sum_i^n a_i u_i$, $u \in V$.

(52.1) $\|u\|_\infty = \max_i |a_i|$ called l_∞ norm

(52.2) $\|u\|_1 = \sum_i |a_i|$ called l_1 norm

(52.3) $\|u\|_2 = \left(\sum_i |a_i|^2\right)^{\frac{1}{2}}$ called l_2 norm

(52.4) $\|u\|_p = \left(\sum_i |a_i|^p\right)^{\frac{1}{p}}$, $1 \leq p < \infty$ called l_p norm,

Exercise (i) prove that all of these are bona fide norms

(ii) draw the balls $\|u\|_p \leq 1$, $1 \leq p \leq \infty$ in $\mathbb{R}^2$

Note: The norms (52.1)–(52.4) all depend on the choice of basis $u_1, \ldots, u_n$ for V.

Now as the space of $n \times n$ matrices, $M(n, \mathbb{R})$ or $M(n, \mathbb{C})$ is just $\mathbb{R}^{n^2}$ or $\mathbb{C}^{n^2}$ with the usual addition, it follows that $M(n, \mathbb{R})$ and $M(n, \mathbb{C})$ carry many (equivalent) norms, and hence many (equivalent) topologies

e.g. for $A = (a_{ij})$

$$\|A\| = \max_{i,j} |a_{ij}|$$

$$\|A\|_2 = \left(\sum_{i,j} |a_{ij}|^2 \right)^{\frac{1}{2}}$$ (already encountered earlier – see Toda Flow)

etc.

There is one norm, however, that we will use which is particularly useful and which, henceforth in these lectures, we will write simply as $\|A\|$.

This norm, called the operator norm, is defined as follows:

(53.1) $$\|A\| = \sup_{u \neq 0} \frac{\|Au\|}{\|u\|}$$

Here $\|u\| = \left(\sum_{i=1}^{n} |u_i|^2\right)^{\frac{1}{2}}$ for $u = (u_1, \dots, u_n)^T \in \mathbb{R}^n$ or $\mathbb{C}^n$.

It is easy to see (exercise) that (53.1) is indeed a norm on $M(n, \mathbb{R})$ or $M(n, \mathbb{C})$

Observe that for $u = (u_1, \dots, u_n)^T$, $(Au)_i = \sum_{j=1}^{n} a_{ij} u_j$

(54.1) $$\|Au\|^2 = \sum_i \left|\sum_{j=1}^{n} a_{ij} u_j\right|^2 \leq \sum_i \left(\sum_{j=1}^{n} |a_{ij}|^2\right) \sum_{j=1}^{n} |u_j|^2$$

are Cauchy-Schwarz inequality.

$$\leq n^2 \|A\|_\infty^2 \|u\|^2$$

Thus

(54.2) $$\|A\| \leq n \|A\|_\infty$$

On the other hand for $e_j = (0 \dots 0, \underset{j}{1}, \dots 0)^T$, $1 \leq j \leq n$, $\|e_j\| = 1$,

$$Ae_j = (a_{1j}, \dots, a_{nj})^T \quad \text{and so} \quad \|Ae_j\|^2 = \sum_i |a_{ij}|^2$$

Thus $\|A\| \geq \frac{\|Ae_j\|}{\|e_j\|} \geq |a_{ij}| \quad \forall i, j \in (1, \dots, n$
and so

(54.3) $$\|A\|_\infty \leq \|A\| \leq n \|A_\infty\|$$

Also from (54.1), $\|Au\|^2 \leq \left(\sum_{i,j} |a_{ij}|^2\right) \|u\|^2 = \|A\|_2^2 \|u\|^2$
and so

$$\|A\| \leq \|A\|_2$$

On the other hand from (54.3) $\|A\|_2 \leq n \|A\|$ and so

(55.1) $$\|A\| \le \|A\|_2 \le n\|A\|$$

Important Remark: note from (54.3) and (55.1), in particular, the relationship between the norms blows up as $n \to \infty$; in particular in ∞ dimensions, the norms can no longer be expected to be equivalent.

In the operator topology we have for matrices A, B

$$\|AB\| = \sup_{u \neq 0} \frac{\|ABu\|}{\|u\|} \le \sup_{\|u\| \neq 0} \frac{\|A\|\,\|Bu\|}{\|u\|}$$

$$\le \|A\|\,\|B\|$$

ie $$\|AB\| \le \|A\|\,\|B\|$$

Thus $M(n, \mathbb{R})$ or $M(n, \mathbb{C})$ is a Banach algebra in the operator norm. We saw earlier (see Toda flow) that $\|A\|_2$ also gives rise to a Banach algebra for $M(n,\mathbb{R})$ and $M(n, \mathbb{C})$.

The following result is basic.

Theorem

The matrices with distinct eigenvalues are dense in the space of all matrices. Also the invertible matrices are dense.

Proof: Let A be an arbitrary $n \times n$ matrix. Then by the Schur Theorem, $\exists$ a unitary matrix U and an upper triangular matrix D such that

$$A = U D U^*$$

Write

$$D = \begin{pmatrix} \alpha_1 & & * \\ & \ddots & \\ 0 & & \alpha_n \end{pmatrix}$$

Now clearly it is possible for each m to find

$$\varepsilon_1^m, \dots, \varepsilon_n^m, \quad \varepsilon_i^m \to 0 \quad \text{as } m \to \infty$$

such that

$$\alpha_i^m = \alpha_i + \varepsilon_i^m, \qquad 1 \le i \le m$$

forms a set of distinct numbers for each m. Set

$$D^m = \begin{pmatrix} \alpha_1^m & & * \\ & \ddots & \\ 0 & & \alpha_n^m \end{pmatrix}$$

where the strict upper parts $*$ of D and D^m are the same.

Set

$$A^m = U D^m U^*$$, which is clearly invertible.

Then

(57.0) $$\|A^m - A\| = \| U(D^m - D)U^* \| \leq \|U\| \, \|D^m - D\| \, \|U^*\|$$

Now

(57.1) $$\|U\| = 1$$

Indeed for any $x \in \mathbb{R}^n$ or $\mathbb{C}^n$.

$$\|Ux\|^2 = (Ux, Ux) = (x, U^*Ux) = (x, x) = \|x\|^2$$

Hence

$$\|U\| = \sup_{x \neq 0} \frac{\|Ux\|}{\|x\|} = 1.$$

Also $D^m - D = \operatorname{diag}(\alpha_1^m - \alpha_1, \ldots, \alpha_n^m - \alpha_n)$

Thus $\|D^m - D\| \to 0$ as $m \to \infty$. Thus $\|A^m - A\| \to 0$ as $m \to \infty$, and so the invertible matrices are dense.

Finally $\exists\, \varepsilon_m \to 0$ s.t. $\alpha_i^m = \alpha_i + \varepsilon_m$, $1 \leq i \leq n$, are non-zero. Then $A^m = U \begin{pmatrix} \alpha_1 + \varepsilon_m & & 0 \\ & \ddots & \\ 0 & & \alpha_n + \varepsilon_m \end{pmatrix} U^*$ is invertible

and $A^m \to A$ as $n \to \infty$. $\square$

The following result is also basic.

For a linear map $A : V \to V$ recall that spectrum of A, denoted by $\sigma(A)$, is the set of all λ such $A - \lambda$ is a bijection i.e. the inverse of $A - \lambda$, $(A-\lambda)^{-1}$ exists.

Let

(58.0) $$\rho(A) = \mathbb{C}^n / \sigma(A) \quad = \text{resolvent set of } A$$

If V is finite dimensional, $\sigma(A) =$

$$\sigma(A) = \{\lambda : \lambda \text{ is an eigenvalue of } A\}$$

and is clearly a finite set.

Theorem (58.1)

For a matrix A, $\rho(A)$ is an open set and $z \to (A-z)^{-1}$ is an analytic map on $\rho(A)$.

Proof: ~~As the complement of a finite set in $\mathbb{C}$,~~ ~~$\rho(A)$ is clearly open.~~ For $\lambda, \mu \in \rho(A)$ we have the second resolvent identity (exercise)

(58.2) $$\frac{1}{A-\lambda} = \frac{1}{A-\mu} + (\lambda-\mu)\frac{1}{(A-\lambda)(A-\mu)}$$

$$= \frac{1}{A-\mu} + (\lambda-\mu)\frac{1}{(A-\mu)}\frac{1}{(A-\lambda)}$$

Thus

$$\frac{1}{A-\lambda}\left(I - (\lambda-\mu)\frac{1}{A-\mu}\right) = \frac{1}{A-\mu}.$$

Now suppose we know $\mu \in \rho(A)$. Then consider the formula

(59.1) $$f(A) = \frac{1}{A-\mu}\left(I - (\lambda-\mu)\frac{1}{A-\mu}\right)^{-1}$$

Now suppose B is any $n \times n$ matrix with $\|B\| < 1$.

Then $I - B$ is invertible with inverse given by

(59.2) $$(I-B)^{-1} = \sum_{k=0}^{\infty} B^k \quad = \text{Neumann series}$$

As $\|B\| < 1$, the Neuman series converges: indeed

for $m > n$, $$\left\| \sum_{k=n}^{m} B^k \right\| \leq \sum_{k=n}^{m} \|B^k\|$$
$$\leq \sum_{k=n}^{m} \|B\|^k.$$

$\to 0$ as $n, m \to \infty$, as $\|B\| < 1$.

Thus $\left\{ \sum_{k=0}^{n} B^k \right\}$ is a Cauchy sequence, and so converges

as $n \to \infty$. But
$$(I-B)\frac{1}{I-B} = (I-B)\sum_{k=0}^{\infty} B^k = \sum_{k=0}^{\infty} B^k - \sum_{k=1}^{\infty} B^k$$
$$= I$$

and similarly $\frac{1}{I-B}(I-B) = I$, and so (59.2) indeed gives an inverse for $(I-B)$.

Returning to (59.1), we note that
$$\left\|(\lambda-\mu)\frac{1}{A-\mu}\right\| = |\lambda-\mu| \left\|\frac{1}{A-\mu}\right\| < 1$$

for λ sufficiently close to μ, and for such λ,

$(I - (\lambda-\mu)(A-\mu)^{-1})^{-1}$ and hence $f(A)$ exists.

A simple calculation then shows that $f(A) = \frac{1}{A-\lambda}$

so that $A-\lambda$ is invertible for all λ close enough

to μ. Then for λ close enough to μ, we can use

(58.2) to find

$$\frac{(A-\lambda)^{-1} - (A-\mu)^{-1}}{\lambda-\mu} = \frac{1}{(A-\lambda)} \frac{1}{A-\mu}.$$

and letting $\lambda \to \mu$ we see that $\frac{d}{d\lambda}(A-\lambda)^{-1}$ $\exists$ an

(60.1) $$\frac{d}{d\lambda}(A-\lambda)^{-1}\Big|_{\lambda=\mu} = \frac{1}{(A-\mu)^2}$$

This proves the theorem. $\square$

<u>Remarks</u> In the above proof, we are dealing with

matrices and the derivatives are pointwise ie

$$\frac{d}{d\lambda}((A-\lambda)^{-1})_{ij}\Big|_{\lambda=\mu} = \left(\frac{1}{(A-\mu)^2}\right)_{ij}, \quad 1 \le i,j \le n.$$

But the calculation goes through for bounded linear operators in an (infinite dimensional) <u>Banach space</u>

(see eg. Reed-Simon Vol I, Methods of Math. Physics).

In the matrix case for $A = (a_{ij})$

(61.1) $$(A-\lambda)^{-1}_{ij} = \frac{\alpha_{ij}(\lambda)}{\det(A-\lambda)} \qquad 1 \le i,j \le n$$

where the cofactor $\alpha_{ij}(\lambda)$ is clearly a polynomial in λ, and hence analytic. So the theorem follows from (61.1) for $\lambda \notin \operatorname{spec}(A) = \{\lambda_i : \lambda_i \text{ eigenvalue of } A\}$

As $\operatorname{spec}(A)$ is finite, $\rho(A) = \mathbb{C} \setminus \sigma(A)$ is clearly open.

Rather than invoking (61.1), we have given an operator theoretic proof, because that is the one that goes through for infinite dimensional spaces: Thus Thm (58.1) goes through for bounded operators A in a Banach space V with the understanding that

$$\rho(A) \equiv \{\lambda : (A-\lambda)^{-1} \text{ exists}\}$$

and $f : \mathbb{C} \to V$ is analytic if $\lim_{h \to 0} \frac{f(z+h) - f(z)}{h} = f'(z)$ exists. (Note that for general A, $\sigma(A) = \{\lambda : (A-\lambda)^{-1} \nexists\}$ is not necessarily finite, so it is not clear a priori that $\rho(A)$ is open.)