Lecture VII: Hermitian Matrices and the Diagonalization of Hermitian and Normal Operators

Definition: An Hermitian inner product on a vector space V over $\mathbb{C}$ is a function $\langle\cdot,\cdot\cdot\rangle : V \times V \to \mathbb{C}$ such that

(1) $\langle\cdot,\cdot\rangle$ is $\mathbb{C}$-linear in the first slot $(\cdot)$

(2) $\langle v,w\rangle = \overline{\langle w,v\rangle}$ where $\overline{}$ denotes complex conjugation

(3) the (necessarily real) number $\langle v,v\rangle \geq 0$ for all $v \in V$ and $\langle v,v\rangle = 0$ if and only if $v = \vec{0}$

Basic example $V = \mathbb{C}^n$

$$\langle (z_1,\dots,z_n),(w_1\dots,w_n)\rangle = \sum_{j=1}^{n} z_j \overline{w}_j$$

Definition: An <u>orthonormal</u> <u>basis</u> (for a finite-dimensional vector space V over $\mathbb{C}$ with an Hermitian inner product $\langle\, ,\,\rangle$) is a basis $v_1,\dots,v_n$ such that $\langle v_i, v_j\rangle = 0$ if $i \neq j$ and $\langle v_i, v_i\rangle = 1$, for each $i = 1, 2 \dots n$.

Theorem: A finite-dimensional vector space V over $\mathbb{C}$ with dimension n has an orthonormal basis $v_1,\dots,v_n$.

Proof. Choose a basis $w_1,\dots,w_n$ for V over $\mathbb{C}$. Make it orthonormal by the Gram Schmidt process (which works the same way as in the real case). $\square$

Corollary (of the proof): If $W \subset V$, V as in the theorem, then $\exists$ an orthonormal basis $v_1, \ldots, v_n$ such that

(1) $v_1, \ldots, v_\ell$ is an orthonormal basis for W ($\ell = \dim_{\mathbb{C}} W$)

and

(2) $W^\perp =$ the $\mathbb{C}$-linear span of $v_{\ell+1} \ldots v_n$

where $W^\perp = \{v \in V : \langle v, w \rangle = 0 \ \forall w \in W\}$.

Proof: Choose $w_1, \ldots, w_\ell$ a basis for W. Let $w_{\ell+1} \ldots w_n$ be such that $w_1, \ldots, w_\ell, w_{\ell+1} \ldots w_n$ is a basis for W. Then Gram Schmidt produces $v_1, \ldots, v_n$ orthonormal basis for V such that $v_1, \ldots, v_\ell$ is an orthonormal basis for W. In this basis, it is clear that $w \in W^\perp$ ($w = \sum_{i=1}^{n} \alpha_i v_i$) if and only if $\alpha_1 = \ldots = \alpha_\ell = 0$. $\square$

Definition: With V, $\mathbb{C}$-vector space with Hermitian inner product as before, a linear transformation $T: V \to V$ is <u>Hermitian</u> if and only if $\langle Tv, w \rangle = \langle v, Tw \rangle$ for all $v, w \in V$.

Lemma: If V has an orthonormal basis $v_1, \ldots, v_n$ (V finite dimensional) then

T is Hermitian if and only if, with A = the matrix of T relative to $v_1, \ldots, v_n$ (at both ends),

$$A = \overline{A^t}$$

(here $\overline{}$ denote complex conjugation as before).

Proof: Calculate using $\langle \sum_{i=1}^n \alpha_i v_i, \sum_{j=1}^n \beta_j v_j \rangle$

$$= \sum_{i=1}^n \alpha_i \overline{\beta_i} \ . \ \square$$

Notation: Hereafter $A^* = \overline{A^t}$.

Theorem: If $T: V \to V$ is Hermitian (V $\mathbb{C}$-vector space, finite-dimensional, with Hermitian inner product) then $\exists$ an orthonormal basis $v_1, \ldots, v_n$ of V, $n = \dim_{\mathbb{C}} V$, and real numbers $\lambda_1, \ldots, \lambda_n$ such that

$$T v_j = \lambda_j v_j \qquad \text{for each } j \in \{1, \ldots n\}$$

Proof: First note that if $Tv = \lambda v$, $v \neq \vec{0}$ $\lambda \in \mathbb{C}$ then in fact λ is real because $\langle Tv, v \rangle = \langle \lambda v, v \rangle = \lambda \langle v, v \rangle$ while $\langle Tv, v \rangle = \langle v, Tv \rangle = \langle v, \lambda v \rangle = \overline{\lambda} \langle v, v \rangle$ so $v \neq \vec{0} \Rightarrow \lambda = \overline{\lambda}$.

Now consider the (polynomial) equation in λ

$$\det(A - \lambda I_n) = 0$$

where A is the matrix of A relative an arbitrary orthonormal basis (which need not be the same one as the $v_1 \ldots v_n$ we are going to produce eventually). This has a solution λ_1 (since it is a polynomial equation of degree $n \geq 1$ over $\mathbb{C}$).

Then

$$\ker(T - \lambda_1 I) \neq \{0\}$$

Choose $v_1 \in \ker(T - \lambda_1 I)$ with $\langle v_1, v_1 \rangle = 1$. Note that as before λ_1 is actually necessarily a real number.

Set $W = (\mathrm{span}(v_1))^{\perp}$.

Then

$$T(W) \subset W$$

(W is "T invariant") because $0 = \langle w, v \rangle$

$$\Rightarrow \langle Tw, v \rangle = \langle w, Tv \rangle = \langle w, \lambda_1 v \rangle$$
$$= \bar{\lambda}_1 \langle w, v \rangle = 0.$$

(that $\bar{\lambda}_1 = \lambda_1$ plays no role in this argument).

Clearly, we are ready for an induction on n, just as we were in the situation of diagonalization of real symmetric matrices. When $n=1$, the result of the Theorem is obvious. If we assume the result for $n-1$, $n \geq 2$, then as above we can "split" V as

$\mathrm{span}(v_1) \oplus (\mathrm{span}(v_1))^\perp$, $v_1 \ni Tv_1 = \lambda_1 v_1$, an orthogonal decomposition. Applying inductively the Theorem to $T|(\mathrm{span}(v_1))^\perp$, which has $\mathbb{C}$-dimension $n-1$, we get $v_2, \ldots, v_n$ o.n. basis for

$(\mathrm{span}(v_1))^\perp$ and $\lambda_2, \ldots, \lambda_n$ with $T(v_i) = \lambda_i v_i$ $i \geq 2$. Then $\lambda_1, \ldots \lambda_n$ and $v_1, \ldots v_n$ are as required for $T: V \to V$, $\dim_{\mathbb{C}} V = n$. $\square$

Definition: An $n \times n$ matrix with $\mathbb{C}$ entries is <u>normal</u> if and only if $AA^* = A^*A$. — A commutes with its "Hermitian conjugate" as people like to call $A^* = \overline{A^t}$.

Theorem: If $T: \mathbb{C}^n \to \mathbb{C}^n$ is the linear transformation with matrix A and if A is normal, then there are an orthonormal basis $v_1, \ldots, v_n$ of $\mathbb{C}^n$ and complex numbers $\lambda_1, \ldots, \lambda_n$ such that
$$Tv_j = \lambda_j v_j \quad \text{for each } j \in \{1, \ldots n\}.$$

"Normal operators are diagonalizable".

Proof: Every $n \times n$ matrix A can be written as $B + iC$ where $B^* = B$ and $C^* = C$: namely

$$B = \frac{1}{2}(A + A^*) \qquad C = \frac{1}{2i}(A - A^*)$$

will work (Exercise: Check this and check that B, C with $B^* = B$, $C^* = C$ and $B + iC = A$ are uniquely determined, given A).

When is $AA^* = A^*A$?

This happen exactly when (since if $A = B + iC$ $A^* = B^* - iC^*$)

$$(B + iC)(B^* - iC^*) = (B^* - iC^*)(B + iC).$$

But since $B^* = B$ & $C^* = C$, this happens (expanding and cancelling) exactly when

$$(B + iC)(B - iC) = (B - iC)(B + iC)$$

or

$$2CB = 2BC$$

or

$$CB = BC.$$

In other words, A is normal if and only if B and C commute, when $A = B + iC$, with $B^* = B$, $C^* = C$.

Now we shall apply the following fundamental lemma, which is important in other contexts too:

Lemma on commuting linear transformation: If $T: V \to V$ and $S: V \to V$ are linear transformation of a vector space V over a field F, then for each $\lambda \in F$,

$$S(\ker(T - \lambda I)) \subset \ker(T - \lambda I)$$

where $I : V \to V$ is the identity map.

"The eigenspaces of T are S-invariant if S and T commute".

Proof of the Lemma: If $v \in \ker(T - \lambda I)$, then

$$T(Sv) = S(Tv) = S(\lambda v) = \lambda S(v)$$

so $Sv \in \ker(T - \lambda I)$. $\square$

Returning now to A normal, $A = B + iC$, B, C Hermitian and hence $BC = CB$, we choose an orthonormal basis $v_1, \ldots, v_n$ for $\mathbb{C}^n$ such that B is diagonal for $v_1 \ldots v_n$ i.e. $B v_j = \lambda_j v_j$ for some λ_j's $\in \mathbb{C}$

(actually, in $\mathbb{R}$). Now not all the λ_j's need be distinct. Let $\mu_1, \ldots, \mu_l$ $l \leq n$ be the distinct λ-values that occur. Each μ_j, $j \in \{1, \ldots, l\}$ has some v's in the orthonormal basis associated to it. Set V_j = span of these v's (the λ's of which $= \mu_j$).

Then V is an orthogonal direct sum of $V_1 \oplus \ldots \oplus V_l$ and T preserves this direct sum decomposition with

$B|V_j$ being multiplication by μ_j. Footnote*

By the Lemma (since clearly $V_j = \ker(B - \mu_j I_n)$, we get

$CS(V_j) \subset V_j$. So $\exists$ an orthonormal basis, say $\hat{v}_j^1, \ldots \hat{v}_j^{d_j}$ where

d_j = dimension of V_j such that, for some $\lambda_j^1 \ldots \lambda_j^{d_j}$, $C(\hat{v}_j^k) = \lambda_j^k \hat{v}_j^k$.

Taking all such $\hat{v}_j^k$ $j \in \{1, \ldots l\}$ $k \in \{1, \ldots d_j\}$ together gives a diagonalizing basis for $B + iC$ $\Box$

(Note that the $\hat{v}_j^k$ are probably different from the v's that originally generated V_j here)

*Here we use the usual orthogonal direct sum decomposition idea: $V = V_1 \oplus \ldots \oplus V_l$ means every vector $v \in V$ has the form $w_1 + \ldots + w_l$, $w_j \in V_j$ with ws uniquely determined & pairwise $\perp$: $\langle w_{j_1}, w_{j_2} \rangle = 0$ if $j_1 \neq j_2$.