Matrix Theory, Math6304
Lecture Notes from Sept 11, 2012

taken by Tristan Whalen

1 Further Review continued

Warm-up

Let $A, B \in M_n$ and suppose $\det(A) \neq 0$. Define a matrix valued function as follows:

$$F(t) = (A + tB)^{-1}$$

The determinant of $F(t)$ is just a polynomial involving the matrix entries. So in particular the determinant is continuous, so if it is not 0 at $t = 0$, then it also is not 0 for for t sufficiently small. Thus, $A + tB$ is invertible for sufficiently small t so our function F at least makes sense for such t values. Using an explicit formula for the inverse of a matrix in terms of its determinant and the cofactors, we see that $F(t)$ is actually differentiable for sufficiently small t.

Now, we wish to compute $F'(0)$. Inspired by the blissful days of Calculus, we begin with

$$(F(t))(F(t))^{-1} = (A + tB)^{-1}(A + tB) = I$$

and then take the derivative of both sides with respect to t:

$$\frac{d}{dt}\big((F(t))(F(t))^{-1}\big) = \frac{d}{dt}(I)$$

The right hand side is 0. The left hand side is:

$$\begin{aligned}
\frac{d}{dt}\big((F(t))(F(t))^{-1}\big) &= F'(t)(F(t))^{-1} + F(t)\frac{d}{dt}\big((F(t))^{-1}\big) \\
&= F'(t)(A + tB) + F(t)\frac{d}{dt}(A + tB) \\
&= F'(t)(A + tB) + F(t)B
\end{aligned}$$

Together, then, we have that $F'(t)(A + tB) + F(t)B = 0$, and evaluating at $t = 0$ gives $F'(0)(A) + F(0)(B) = F'(0)(A) + (A^{-1})(B) = 0$. Right multiplication by A^{-1} gives $F'(0) + A^{-1}BA^{-1} = 0$, and we have the answer:

$$F'(0) = -A^{-1}BA^{-1}\,.$$

1.5 Simultaneous Diagonalization (continued)

1.5.34 Definition. A family $F \subset M_n$ is a *commuting family* if for any $A, B \in F$ we have $AB = BA$.

1.5.35 Definition. Let $A \in M_n$. A subspace $W \subset \mathbb{C}^n$ is *A-invariant* if $Aw \in W$ for each $w \in W$. If $F \subset M_n$ is any family of matrices, then W is called *F-invariant* if it is A-invariant for every $A \in F$.

So, if $W \subset \mathbb{C}^n$ is A-invariant, A maps W into itself and does not take any vectors "out of" W.

1.5.36 Remark. If $W \subset \mathbb{C}^n$ is A-invariant for some $A \in M_n$ and $\dim(W) \geq 1$, then there exists an $x \in W \setminus \{0\}$ such that $Ax = \lambda x$ for some $\lambda \in \mathbb{C}$ (i.e., x is an eigenvector for A). Why? If we restrict A to W, then $A|_W : W \to W$ is a linear map on a vector space, so it has an eigenvalue (since our field is $\mathbb{C}$ of course).

The notion of invariant subspace gives, in some sense, a "weaker" quality to observe than a basis of eigenvectors (i.e., diagonalizability). Even if we are not able to find a basis of eigenvectors, we (may?) be able to break things down into proper invariant subspaces:

1.5.37 Example. Consider the matrix $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ as a map $A : \mathbb{C}^3 \to \mathbb{C}^3$. Then $W = \{[x, y, 0]^* \mid x, y \in \mathbb{C}\}$ is an invariant subspace. Two eigenvectors of A are $v_1 = [1, 0, 0]^*$ and $v_2 = [0, 0, 1]^*$ with corresponding eigenvalues $\lambda_1 = 0$ and $\lambda_2 = 1$. Note that there is only a one-dimensional space of eigenvectors in W, but W is two-dimensional. We also have that $V = \{[0, 0, z]^* \mid z \in \mathbb{C}\} = \mathbb{C}$ is an invariant subspace. So we can write $\mathbb{C}^3$ as a direct sum of proper A-invariant subspaces $W \oplus V$ but not as the direct sum of eigenspaces, because the eigenvectors of A do not span $\mathbb{C}^3$.

We are considering matrices over $\mathbb{C}$, so the reader may replace $\mathbb{R}$ with $\mathbb{C}$ and the claims still hold.

1.5.38 Lemma. *If $F \subset M_n$ is a commuting family, then there is an $x \in \mathbb{C}^n$ such that $Ax = \lambda x$ for* each $A \in F$.

By the previous remark, each $A \in F$ has an eigenvector. This lemma says all $A \in F$ share a common eigenvector.

Proof. Choose an F-invariant subspace $W \in \mathbb{C}^n$. (There is such a space, since $\mathbb{C}^n$ itself is F-invariant.) Furthermore, choose W so that it has a minimal non-zero dimension. (We are able to do this by the Well-Ordering Principle for the natural numbers. Observe that $\{k : k = dim(W)$ and W is an F-invariant subspace$\}$ is a non-empty subset of $\mathbb{N}$, since it contains n, so it has a minimum.)

Next, we show that any $x \in W \setminus \{0\}$ is an eigenvector for each $A \in F$ by means of contradiction. So, for the purpose of contradiction, suppose this is not the case. That is, there is an $A \in F$ and a $y \in W$, $y \neq 0$, such that $Ay \notin \mathbb{C}y$ (this is just saying y is not an eigenvector of A). But, from the remark above, we know A has an eigenvector in W, so choose $x \in W \setminus \{0\}$ such that $Ax = \lambda x$.

Now consider the set

$$W_0 := \{z \in W : Az = \lambda z\}$$

Note that $W_0 \subsetneq W$ since $y \in W$ but $y \notin W_0$. Also note that W_0 is a subspace of W (which is straightforward to check; note that λ is fixed).

However, if $B \in F$, then for $z \in W_0$ we at least have that $Bz \in W$ by F-invariance of W. Then we have that for $z \in W_0$,

$$A(Bz) = ABz = BAz = B(\lambda z) = \lambda Bz$$

where we have used the fact that A and B commute by hypothesis and the fact that $Az = \lambda z$ by definition of W_0. But this shows that $Bz \in W_0$ by definition of W_0. So B maps W_0 into W_0. Since $B \in F$ was arbitrary, this implies W_0 is F-invariant. But now we have a contradiction. We chose W to be F-invariant of minimum dimension and have produced a proper F-invariant subspace of W.

Thus, it must be that for each $y \in W$, $Ay \in \mathbb{C}y$ for all $A \in F$. $\square$

1.5.39 Remark. Again, this lemma says that when you have a family of commuting matrices, this family has at least one common eigenvector.

Now we up the ante and define the following:

1.5.40 Definition. A *simultaneously diagonalizable* family $F \subset M_n$ is a family such that there exists an invertible $S \in M_n$ and $S^{-1}AS$ is diagonal for each $A \in F$.

Notice that in the definition the same S "works" for every member of the family.

1.5.41 Theorem. *Let $F \subset M_n$ be a family of diagonalizable matrices. Then it is a commuting family if and only if it is simultaneously diagonalizable.*

Proof. First suppose that F is simultaneously diagonalizable (via S). Then for $A, B \in M_n$,

$$AB = S^{-1}D_1SS^{-1}D_2S = S^{-1}D_1D_2S = S^{-1}D_2D_1S = S^{-1}D_2SS^{-1}D_1S = BA$$

So F is a commuting family.

Conversely suppose that F is a commuting family of diagonalizable matrices. We proceed by induction on n via a common trick in matrix theory known as "deflation". The base case, $n = 1$, has nothing to prove ($M_1 = \mathbb{C}$ and everything's awesome). Now we'll suppose the theorem holds for matrices of size $n-1$ or less. We'll take on the size n case by "deflating" it to some $n-1$ size cases.

Inductive hypothesis: given a family of diagonalizable matrices $F \in M_k$ for $k \leq n-1$, if it is commuting then it is simultaneously diagonalizable.

Now let $F \in M_n$. If each $A \in M_n$ is of the form λI, then there is no work to do (take $S = I$). So assume A is diagonalizable with eigenvalues $\{\lambda_1, \ldots, \lambda_r\}$ where $r \geq 2$ and $AB = BA$ for each $B \in F$. A is similar to a diagonal matrix, so without loss of generality assume A is diagonal. Since each B commutes with the diagonal matrix A, each $B \in F$ is a block diagonal matrix (see the theorem about commuting and simultaneously diagonalizable).

Since A had at least two distinct entries, each block of each B has size $n-1$ or less.

By the block-wise commutation property (if we take all the blocks of size $k \leq n-1$ we have a commuting family), together with the inductive hypothesis, all the blocks (of size k) in all $B \in F$ are simultaneously diagonalizable. Thus there exist fixed matrices $T_1, T_2, \ldots, T_r$ such that conjugating each $B \in F$ with $T = \begin{bmatrix} T_1 & & & \\ & T_2 & & \\ & & \ddots & \\ & & & T_r \end{bmatrix}$ gives a block matrix in which the blocks are diagonal, which is a diagonal matrix. That is, for any $B \in F$,

$$\begin{aligned} T^{-1}BT &= \begin{bmatrix} T_1^{-1}B_1T_1 & & & \\ & T_2^{-1}B_2T_2 & & \\ & & \ddots & \\ & & & T_r^{-1}B_rT_r \end{bmatrix} \\ &= \begin{bmatrix} D_1 & & & \\ & D_2 & & \\ & & \ddots & \\ & & & D_r \end{bmatrix} \end{aligned}$$

□

1.6 Hermitian, Normal, Unitary

1.6.42 Definition. Let $A \in M_{n,m}$. The *adjoint* of A, denoted A^*, is an $m \times n$ matrix such that $< Ax, y >=< x, A^*y >$ for all $x \in \mathbb{C}^m$ and $y \in \mathbb{C}^n$. If $A \in M_n$ and $A = A^*$, we say A is *Hermitian* or *self-adjoint*. If $A \in M_n$ satisfies $A^* = -A$, then it is *skew-Hermitian*.

Unlike the determinant of a matrix in $\mathbb{C}^n$, the adjoint of a matrix in $\mathbb{C}^n$ can be computed easily:

1.6.43 Proposition (The Adjoint is the Conjugate Transpose)**.** *If $A \in \mathbb{C}^n$ where $A = [a_{j,k}]$, then $A^* = [\overline{a_{k,j}}]$. That is, A^* is the matrix obtained by taking the transpose of A and replacing each entry with its complex conjugate.*

Proof. Recall that for complex vectors $x = [x_1, ..., x_n]$ and $y = [y_1, ..., y_n]$, then the dot product is defined as $\langle x, y\rangle = \Sigma_{k=1}^n x_k \overline{y_k}$.

Now, let $A = [a_{j,k}]$ and $A^* = [b_{j,k}]$. By definition of matrix multiplication and of the dot product,

$$\langle Ax, y\rangle = \langle [\Sigma_{k=1}^n a_{j,k}x_k]_{j=1}^n, [y]_{j=1}^n\rangle = \Sigma_{j=1}^n \Sigma_{k=1}^n a_{j,k}x_k\overline{y_j}$$

On the other hand,

$$\begin{aligned} \langle x, A^*y\rangle &= \langle [x_j]_{j=1}^n, [\Sigma_{k=1}^n b_{j,k}y_k]_{j=1}^n\rangle \\ &= \Sigma_{j=1}^n \left[x_j \overline{(\Sigma_{k=1}^n b_{j,k}y_k)} \right] \\ &= \Sigma_{j=1}^n \left[x_j \Sigma_{k=1}^n \overline{b_{j,k}}\,\overline{y_k} \right] \\ &= \Sigma_{j=1}^n \Sigma_{k=1}^n \overline{b_{j,k}} x_j \overline{y_k} \end{aligned}$$

where we have used basic facts about taking the conjugate (i.e., $\overline{ab+c} = \overline{a}\overline{b} + \overline{c}$). Substituting $b_{j,k} = \overline{a_{k,j}}$, the above equals

$$\Sigma_{j=1}^{n}\Sigma_{k=1}^{n} a_{k,j} x_j \overline{y_k}$$

and we have shown that $A^* = [\overline{a_{k,j}}]$ (simply swap the roles of the indices j and k and compare to the first equation in this proof). That this is the only such A^* follows since this must be true for all $x, y \in \mathbb{C}^n$, or by appealing to advanced linear algebra and the Riesz Representation Theorem. □

If we think of the adjoint as analogous to the complex conjugate, then Hermitian matrices correspond to real numbers and skew-Hermitian matrices correspond to imaginary numbers.

1.6.44 Facts. $A = A^*$ if and only if $(iA) = -(iA)^*$. For any $A \in M_n$, there exist unique self-adjoint $B, C \in M_n$ with $A = B + iC$. Note that

$$B = \frac{A + A^*}{2} \text{ and } C = \frac{A - A^*}{2i}$$

B is called the *Hermitian part of* A, and iC is called the *skew-Hermitian part of* A.

An upcoming proposition gives more credence to the notion of thinking of Hermitian matrices as analogous to real numbers. But to prove it we make use of the Polarization Identity. This identity works for general sesquilinear forms, but we only need it for the inner product on $\mathbb{C}^n$.

1.6.45 Lemma (The Polarization Identity)**.** *Given $x, y \in \mathbb{C}^n$ and a matrix $A \in M_n$, we have that*

$$\langle Ax, y\rangle = \frac{1}{4}\left[\langle A(x+y), x+y\rangle - \langle A(x-y), x-y\rangle + i\langle A(x+iy), x+iy\rangle - i\langle A(x-iy), x-iy\rangle\right]$$

Proof. This can be proven directly by simplifying the right-hand side, using the properties of the inner product (linearity in first coordinate, conjugate-linearity in second coordinate). □

1.6.46 Remark. Taking $A = I$ above we have the usual presentation of the polarization identity. Notice that on the right-hand side all the terms are of the form $\langle Az, z\rangle$, so we can conclude something about $\langle Ax, y\rangle$ when we only know $\langle Ax, x\rangle$ for all x.

1.6.47 Proposition. *A matrix $A \in M_n$ is Hermitian if and only if for all $x \in \mathbb{C}^n$ we have that $\langle Ax, x\rangle \in \mathbb{R}$.*

Proof. If A is Hermitian, then

$$\langle Ax, x\rangle = \langle x, A^*x\rangle = \langle x, Ax\rangle = \overline{\langle Ax, x\rangle}$$

which implies $\langle Ax, x\rangle$ is a real number. Conversely, suppose $A \in M_N$ satisfies, for all $x \in \mathbb{C}^n$, $\langle Ax, x\rangle \in \mathbb{R}$. We write A in terms of its Hermitian and skew-Hermitian parts: $A = B + iC$, where $B = B^*$ and $C = C^*$. Then (using linearity of the inner product in the first slot), we have

$$\langle Ax, x\rangle = \langle Bx, x\rangle + i\langle Cx, x\rangle$$

Since the left hand side is real, the right hand side must be real as well. So it must be that $\langle Cx, x\rangle = 0$ for all $x \in C^n$. We show that $C = 0$. By the Polarization Identity (see lemma above), we have that for any $x, y \in \mathbb{C}^n$,

$$\begin{aligned}\langle Cx, y\rangle = \frac{1}{4}[&\langle C(x+y), x+y\rangle \\ &- \langle C(x-y), x-y\rangle \\ &+ i\langle C(x+iy), x+iy\rangle \\ &- i\langle C(x-iy), x-iy\rangle] \\ &= 0\end{aligned}$$

since all the terms on the right equal 0, since $\langle Cx, x\rangle = 0$ for all $x \in \mathbb{C}^n$. Since $\langle Cx, y\rangle = 0$ for all $x, y \in \mathbb{C}^n$, we conclude that $C = 0$. This implies that $A = B$, so $A^* = B^* = B = A$, so A is Hermitian. $\square$