Density of diagonalizable square matrices

Student: Daniel Cervone; Mentor: Saravanan Thiyagarajan
University of Chicago VIGRE REU, Summer 2007.

For this entire paper, we will refer to V as a vector space over $\mathbb{C}$ and L(V) as the set of linear operators $\{A \mid A : V \to V\}$. Recall the following definition: if A is a linear operator on a vector space V, and $\exists \mathbf{v} \neq 0 \in V$ and $\lambda \in \mathbb{C}$ st $A\mathbf{v} = \lambda\mathbf{v}$, then $\mathbf{v}$ and λ are an *eigenvector* and *eigenvalue* of A, respectively.

Theorem 1: A matrix is called *diagonalizable* if it is similar to some diagonal matrix. If $A \in L(V)$ has distinct eigenvalues then A is diagonalizable.

Proof: Let $\mathbf{w}_1 \dots \mathbf{w}_n$ (assuming dimV = n) be the eigenvectors that correspond to each eigenvalue. Let W be the matrix that has $\mathbf{w}_1 \dots \mathbf{w}_n$ for each of its columns. A quick calculation will verify that:

$$\begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{pmatrix} \begin{pmatrix} \mathbf{w}_1 & \mathbf{w}_2 \dots & \mathbf{w}_n \end{pmatrix} = \begin{pmatrix} \mathbf{w}_1 & \mathbf{w}_2 \dots & \mathbf{w}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

$$\text{LHS} = \begin{pmatrix} & \cdots & \\ A\mathbf{w}_1 & A\mathbf{w}_2 \dots & A\mathbf{w}_n \\ & \cdots & \end{pmatrix} \text{ and RHS} = \begin{pmatrix} \lambda_1\mathbf{w}_1 & \lambda_2\mathbf{w}_2 \dots & \lambda_n\mathbf{w}_n \end{pmatrix} \text{ and clearly } A\mathbf{w}_i = \lambda_i\mathbf{w}_i.$$

And we know that W is invertible since the fact that the eigenvalues of A are distinct implies that $\mathbf{w}_1 \dots \mathbf{w}_n$ are linearly independent. Thus:

$$\begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{pmatrix} = \begin{pmatrix} \mathbf{w}_1 & \mathbf{w}_2 \dots & \mathbf{w}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} \mathbf{w}_1 & \mathbf{w}_2 \dots & \mathbf{w}_n \end{pmatrix}^{-1}$$

This proves the theorem. □

Theorem 2: Suppose $T \in L(V)$ with nondistinct eigenvalues. Let $\lambda_1 \dots \lambda_m$ be the distinct eigenvalues of T, thus m < dim(V). Then ∃ a basis of V with respect to which T has the form:

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_n \end{pmatrix} \text{ where each } A_j \text{ is an upper triangular matrix of the form: } \begin{pmatrix} \lambda_j & * & * \\ & \ddots & * \\ 0 & & \lambda_j \end{pmatrix}$$

Proof : $\forall 1 \leq j \leq m$ let U_j be the subspace of generalized eigenvectors of T corresponding to λ_j :

$\forall j,\ U_j = \{\mathbf{v} \in V : (T - \lambda_j I)^k \mathbf{v} = 0 \text{ for some } k \in \mathbb{N}\}$.

It follows from this immediately that $\forall j,\ U_j = \text{null}(T - \lambda_j I)^k$, and that $(T - \lambda_j I)|_{U_j}$ is nilpotent.

Note that if A is any linear operator on V, then null(A) is a subspace of V since it contains 0 and clearly satisfies closure under addition and scalar multiplication (these follow from A being linear).

Before continuing, we need a crucial lemma:

Lemma 1: If N is a nilpotent linear operator on a vector space X, then $\exists$ a basis of X with respect to which N has the form: $\begin{pmatrix} 0 & * & * \\ & \ddots & * \\ 0 & & 0 \end{pmatrix}$ (i.e. N has 0's on and below the diagonal).

Proof of Lemma: First note that N nilpotent on X $\Rightarrow$ $\exists \text{p} \in \mathbb{N}$ st $\text{N}^{\text{p}} = [\mathbf{0}] \Rightarrow \text{X} = \text{null}(\text{N}^{\text{p}})$

Next, choose a basis $\{b_1, \ldots, b_{k_1}\}$ of null(N) and extend this to a basis $\{b_1, \ldots, b_{k_2}\}$ of $\text{null}(\text{N}^2)$, where $k_1 \le k_2 \le p$.

We can do this because if $\mathbf{v} \in \text{null(N)}$, then $\text{N}\mathbf{v} = [\mathbf{0}]$, so clearly $\text{N}(\text{N}\mathbf{v}) = [\mathbf{0}]$. Thus $\text{null(N)} \subset \text{null}(\text{N}^2)$.

And since $b_1, \ldots, b_{k_1}$ are linearly independent vectors that span null(N), we can span $\text{null}(\text{N}^2)$ by $b_1, \ldots, b_{k_1}$ and 1 or more linearly independent vectors $b_{k_1+1}, \ldots, b_{k_2}$ in $\text{null}(\text{N}^2)$ that do not depend on $b_1, \ldots, b_{k_1}$.

We can keep extending the basis of $\text{null}(\text{N}^2)$ to a basis of $\text{null}(\text{N}^3)$ and eventually $\text{null}(\text{N}^{\text{p}})$. In doing so, we establish a basis B = $\{b_1, \ldots, b_p\}$ of X, since B is a basis of $\text{null}(\text{N}^{\text{p}}) = \text{X}$.

Now let us consider N with respect to this basis. We know that by changing the basis of N, we can write N with respect to B as the matrix: $[\text{N}b_1 | \text{N}b_2 | \text{N}b_3 \ldots | \text{N}b_p]$ where each column is the $(p \times 1)$ vector $\text{N}b_i$.

Since $b_1 \in$ null(N), the first column will be entirely 0. This is in fact true for each column through k_1.

The next column is $\text{N}b_{k_1+1}$, where $\text{N}b_{k_1+1} \in \text{null(N)}$ since $\text{N}^2 b_{k_1+1} = 0$ (recall that $b_{k_1+1} \in \text{null}(\text{N}^2)$).

$\text{N}b_{k_1+1} \in \text{null(N)} \Rightarrow$ $\text{N}b_{k_1+1}$ is a linear combination of $b_1 \ldots b_{k_1}$ $\Rightarrow$ all nonzero entries in the $k_1 + 1$ column lie above the diagonal. This is in fact true for all columns from $k_1 \ldots k_2$ where $b_1 \ldots b_{k_2}$ span $\text{null}(\text{N}^2)$.

Similarly, we can take the next column, $\text{N}b_{k_2+1}$, which is in $\text{null}(\text{N}^2)$ since b_{k_2+1} is a basis vector of $\text{null}(\text{N}^3)$. Thus $\text{N}b_{k_2+1}$ depends on $b_1 \ldots b_{k_2}$ and any nonzero entries in the $k_2 + 1$ column lie above the diagonal. We can continue this process through column p, thus confirming that N with respect to the basis B is of the form: $\begin{pmatrix} 0 & * & * \\ & \ddots & * \\ 0 & & 0 \end{pmatrix}$. This proves the lemma.

We now continue the proof of the theorem. Recall that $\forall 1 \le \text{j} \le \text{m}$ $(\text{T} - \lambda_j \text{I})|_{\text{U}_\text{j}}$ is nilpotent. Thus, by the lemma we just proved, $\forall$j $\exists$ a basis B_j of U_j st with respect to B_j:

$$(\text{T} - \lambda_j \text{I})|_{\text{U}_\text{j}} = \begin{pmatrix} 0 & * & * \\ & \ddots & * \\ 0 & & 0 \end{pmatrix}, \text{ and therefore } \text{T}|_{\text{U}_\text{j}} = \begin{pmatrix} \lambda_j & * & * \\ & \ddots & * \\ 0 & & \lambda_j \end{pmatrix}.$$

Moreover, if B is the basis $\{B_1 ... B_m\}$ of U where $U = U_1 \oplus U_2 \oplus \ldots \oplus U_m$ then $T|_U$ with respect to B is in the form:

$$\begin{pmatrix} T_1 & & 0 \\ & \ddots & \\ 0 & & T_m \end{pmatrix}$$ where each T_j is an upper triangular matrix of the form: $$\begin{pmatrix} \lambda_j & * & * \\ & \ddots & * \\ 0 & & \lambda_j \end{pmatrix}$$

Note that this is the desired form corresponding to our theorem. However, we still need to show that this form is possible for T with respect to a basis of V. It suffices to show that V = U, then clearly a basis of U is a basis of V.

To do this, consider the linear operator $S \in L(V)$ where $S = \prod_{j=1}^{m} (T - \lambda_j I)^{\dim V}$. Our claim is that $S|_U = 0$.

To verify this, consider that $\text{null}(T - \lambda_j I)^1 \subseteq \text{null}(T - \lambda_j I)^2 \subseteq \ldots \text{null}(T - \lambda_j I)^n$ for any n. We want to strengthen this statement into the following lemma:

Lemma 2: $\text{null}(T - \lambda_j I)^1 \subseteq \text{null}(T - \lambda_j I)^2 \subseteq \ldots \text{null}(T - \lambda_j I)^{\dim V} = \text{null}(T - \lambda_j I)^{\dim V+1} \ldots = \text{null}(T - \lambda_j I)^{\dim V+n}$

Proof: Suppose $\exists k$ st $\text{null}(T - \lambda_j I)^k = \text{null}(T - \lambda_j I)^{k+1}$. If $\mathbf{x} \in \text{null}(T - \lambda_j I)^{k+n+1}$ for $n \in \mathbb{N}$ then $(T - \lambda_j I)^{k+n+1}\mathbf{x} = 0$.

$\Rightarrow (T - \lambda_j I)^{k+1}(T - \lambda_j I)^n \mathbf{x} = 0 \Rightarrow (T - \lambda_j I)^n \mathbf{x} \in \text{null}(T - \lambda_j I)^{k+1} = (T - \lambda_j I)^k$.

Thus $\text{null}(T - \lambda_j I)^{k+n+1} \subseteq \text{null}(T - \lambda_j I)^{k+n} \Rightarrow \text{null}(T - \lambda_j I)^{k+n+1} = \text{null}(T - \lambda_j I)^{k+n}$.

So $\text{null}(T - \lambda_j I)^k = \text{null}(T - \lambda_j I)^{k+1} \Rightarrow$

$\text{null}(T - \lambda_j I)^1 \subseteq \text{null}(T - \lambda_j I)^2 \subseteq \ldots \text{null}(T - \lambda_j I)^k = \text{null}(T - \lambda_j I)^{k+1} \ldots = \text{null}(T - \lambda_j I)^{k+n}$.

Now we want to show that $\text{null}(T - \lambda_j I)^{\dim V} = \text{null}(T - \lambda_j I)^{\dim V+1}$. To prove this, assume the contrary, i.e.:

$\text{null}(T - \lambda_j I)^1 \subsetneq \text{null}(T - \lambda_j I)^2 \ldots \subsetneq \text{null}(T - \lambda_j I)^{\dim V} \subsetneq \text{null}(T - \lambda_j I)^{\dim V+1}$. Since each $\text{null}(T - \lambda_j I)^i$ is a subspace of V, $\text{null}(T - \lambda_j I)^i \subsetneq \text{null}(T - \lambda_j I)^{i+1} \Rightarrow \dim(\text{null}(T - \lambda_j I)^i) + 1 \leq \dim(\text{null}(T - \lambda_j I)^{i+1})$

since the term left of "$\subsetneq$" has a lower dim than the one to the right. But then $\dim(\text{null}(T - \lambda_j I)^{\dim V+1}) > \dim V$, which is a contradiction since $\text{null}(T - \lambda_j I)^{\dim V+1}$ is a subspace of V. Therefore the following is true:

$\text{null}(T - \lambda_j I)^1 \subseteq \text{null}(T - \lambda_j I)^2 \subseteq \ldots \text{null}(T - \lambda_j I)^{\dim V} = \text{null}(T - \lambda_j I)^{\dim V+1} \ldots = \text{null}(T - \lambda_j I)^{\dim V+n}$. This completes the proof of lemma 2.

We again return to verifying that $S|_U = 0$. Now consider $S\mathbf{u}$ for some $\mathbf{u} \in U$.

$\mathbf{u} \in U \Rightarrow \mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2 \ldots \mathbf{u}_m$ for $\mathbf{u}_i \in U_i$. Since matrix multiplication is distributive, $S\mathbf{u} = S\mathbf{u}_1 + S\mathbf{u}_2 \ldots S\mathbf{u}_m$.

Moreover, we know that $\forall i,j \leq m$, $(T - \lambda_i I)^{\dim V}$ and $(T - \lambda_j I)^{\dim V}$ are commutable (this is because their product in either direction consists of terms of T of some order and terms of TI or IT of some order. And clearly T commutes with T and I commutes with any matrix).

So $S\mathbf{u}_i = (T - \lambda_1 I)^{\dim V}(T - \lambda_2 I)^{\dim V} \cdots (T - \lambda_{i-1} I)^{\dim V}(T - \lambda_{i+1} I)^{\dim V} \cdots (T - \lambda_i I)^{\dim V}\mathbf{u}_i$. Of course, $(T - \lambda_i I)^{\dim V}\mathbf{u}_i = 0$ since $\forall 1 \leq i \leq m$ $(T - \lambda_i I)^{\dim V}\mathbf{u}_i$. Thus $S\mathbf{u} = 0$ and we have proven our claim that $S|_U = 0$, which gives $U \subseteq \text{null}(S)$.

Yet suppose $\mathbf{u} \in \text{null}(S)$. Then $(\prod_{j=1}^{m}(T-\lambda_j I)^{\dim V})(\mathbf{u})=0$. Therefore for some $i \le m$, $(T-\lambda_j I)^{\dim V})(\mathbf{u})=0$.

$\Rightarrow \mathbf{u} \in U_i \Rightarrow \mathbf{u} \in U \Rightarrow \text{null}(S) \subseteq U \Rightarrow \text{null}(S) = U$.

Now, we have shown that $\forall i,j \le m$, $(T-\lambda_i I)^{\dim V}$ and $(T-\lambda_j I)^{\dim V}$ are commutable. From this it follows that S and T are commutable. For a vector $\mathbf{v} \in V$, of course $S(T\mathbf{v}) \in \text{Img}(S)$. Yet since T and S commute, $T(S\mathbf{v}) \in \text{Img}(S)$ (i.e. Img(S) is invariant over T). Let us assume that $\text{Img}(S) \ne 0$. Img(S) invariant over T $\Rightarrow \exists \mathbf{w} \in \text{Img}(S)$ where $\mathbf{w}$ is an eigenvector for T. Moreover, $\mathbf{w} \in \text{Img}(S) \Rightarrow \exists \mathbf{x} \in V$ st $S\mathbf{x}$ is an eigenvector of T. By definition, $S\mathbf{x} \ne 0$, thus $\mathbf{x} \notin \text{null}(S)$. But $S\mathbf{x}$ is an eigenvector of T, so clearly $SS\mathbf{x} = 0$. Thus, $\text{null}(S) \subsetneq \text{null}(S^2)$.

This contradicts lemma 2, since $\text{null}(S) \subsetneq \text{null}(S^2) \Rightarrow \dim(\text{null}(\prod_{j=1}^{m}(T-\lambda_j I))^{\dim V}) < \dim(\text{null}(\prod_{j=1}^{m}(T-\lambda_j I))^{2\dim V})$.

Therefore Img(S)=0. If we apply the rank-nullity theorem to $S:V \to V$, we get:

$\dim V = \dim(\text{null}(S)) + \dim(\text{Img}(S))$.

$\text{Img}(S)=0 \Rightarrow \dim(\text{Img}(S))=0$, so $\dim V = \dim(\text{null}(S))$. We showed earlier that $U = \text{null}(S)$, so $\dim U=\dim V$.

And U being a subspace of V and $\dim U=\dim V \Rightarrow U=V$.

Thus, a basis of U is also a basis of V. This proves the theorem. □

Theorem 3: $\exists A_n$ st $A_i \in L(V)$, A_i has distinct eigenvalues, and $A_n \to T$.

Proof: Theorem 1 (in light of the recent observation) shows that $\exists T'$ st $T \sim T'$ and T'

can be written: $\begin{pmatrix} \lambda_1 & & *_{T'} \\ & \ddots & \\ 0 & & \lambda_k \end{pmatrix}$ where λ_i are eigenvalues

but not necessarily distinct from one another ($i \ne j$ does not imply $\lambda_i \ne \lambda_j$).

Now let A_n be $\begin{pmatrix} \lambda_{1_n} & & *_{T'} \\ & \ddots & \\ 0 & & \lambda_{k_n} \end{pmatrix}$ where $\forall 1 \le i \le k$ $\lambda_{i_n} \to \lambda_i$ and $\forall i, j \le k$ $\lambda_{i_n} = \lambda_{j_n} \Leftrightarrow i = j$

(the eigenvalues of each A_n are distinct).

The fact that $\forall 1 \le i \le k$ $\lambda_{i_n} \to \lambda_i \Rightarrow A_n \to T'$ entrywise $\Rightarrow A_n \to T'$.

But $T \sim T' \Rightarrow \exists$ nonsingular matrix P st $T = P^{-1}T'P$. Now $A_n \to T' \Rightarrow P^{-1}A_nP \to T$ since matrix multiplication is continuous (this is fairly easy to verify: if $\mathbf{v}_n \to \mathbf{v}$ then surely $A\mathbf{v}_n \to \mathbf{v}$ entrywise $\Rightarrow A\mathbf{v}_n \to A\mathbf{v}$. And if a sequence of matrices $X_n \to X$, then clearly the column vectors converge, thus $AX_n \to AX$. And since $A_n \sim P^{-1}A_nP$, the eigenvalues of $P^{-1}A_nP$ are equal to thoseof A_n. Thus $P^{-1}A_nP$ is a sequence of matrices with distinct eigenvalues and $P^{-1}A_nP \to T$. This proves the theorem. □

Observation: If $T \in L(V)$ and $\exists$ a basis of V with respect to which T is triangular, this is equivalent to saying that $\exists T' \in L(V)$ st T' is triangular and $T \sim T'$ (T is similar to T′), i.e. $\exists$ nonsingular matrix P st $T = P^{-1}T'P$.

Corollary: Theorems 1, 2, and 3 imply that any square matrix is a limit point of a sequence of square matrices with distinct eigenvalues. By definition then, square matrices with distinct eigenvalues are *dense* in L(V). And Theorem 1 shows that any square matrix with distinct eigenvalues is diagonlizable, thus the diagonalizable matrices are also dense in L(V).

REFERENCES:

Axler, Sheldon. *Linear Algebra Done Right.* 1997.

http://planetmath.org