Matrix Theory
Lecture Notes from September 2, 2025

taken by Joel Sleeba

Warm Up

Let $A, B \in M_n(\mathbb{C})$, and A be an invertible matrix. Consider the function $F : \mathbb{R} \to M_n(\mathbb{C}) := t \to A + tB$. We are interested in

Clearly F is continuous. Observe that $F(0) = A$ is invertible. We claim that $F(t)$ is invertible for all t in some neighborhood $B_\varepsilon(0)$ of 0. Notice that $F(t) = A + tB = A(I_n + tA^{-1}B)$. Since A is invertible it is enough if we show that $I_n + tA^{-1}B$ is invertible for some $t \in B_\varepsilon(0)$.

Let $\|\cdot\|$ be the operator norm on $M_n(\mathbb{C})$. Let $|t| < \frac{1}{\|A^{-1}B\|}$. We claim that

$$(I_n + tA^{-1}B)^{-1} = \sum_{i=0}^{\infty} (-tA^{-1}B)^i$$

if the infinite sum is well defined. By the triangle inequality we see that

$$\Big\| \sum_{i=1}^{\infty} (-tA^{-1}B)^n \Big\| \leq \sum_{i=1}^{\infty} \|tA^{-1}B\|^n = \sum_{i=1}^{\infty} (|t|\|A^{-1}B\|)^n < \infty$$

where the last inequality is because of the convergence of the geometric series. Hence the infinite sum makes sense. Moreover

$$(I_n + tA^{-1}B)\Big(\sum_{i=1}^{n} (-tA^{-1}B)^i\Big) = I_n + (-tA^{-1}B)^{n+1}$$

and as $n \to \infty$, the tail $(-tA^{-1}B)^{n+1}$ converges to $\mathbf{0}$ by the convergence of the geometric series above. Thus we get that $F(t)$ is invertible in the interval of radius $\frac{1}{\|A^{-1}B\|}$ of 0.

Let $G : B_\varepsilon(0) \to M_n(\mathbb{C}) := t \to F(t)^{-1}$, where $\varepsilon = \frac{1}{\|A^{-1}B\|}$. We are interested in the differentiability of G in $B_\varepsilon(0)$. Then we can linearily approximate $G(t)$ about $t = 0$ as $G(0) + tG'(0)$. Clearly $G(0) = A^{-1}$. Now to show that G is differentiable, observe that if X, Y are invertible, then $X^{-1} - Y^{-1} = -Y^{-1}(X - Y)X^{-1}$. Then

$$\frac{G(t+h) - G(t)}{h} = \frac{-G(t)[F(t+h) - F(t)]G(t+h)}{h} = \frac{-G(t)hBG(t+h)}{h} = -G(t)BG(t+h)$$

and thus

$$G'(t) = \lim_{h \to 0} \frac{G(t+h) - G(t)}{h} = -G(t)BG(t)$$

is well defined in $B_\varepsilon(0)$. Hence G is differentiable everywhere in $B_\varepsilon(0)$ and $G'(0) = -A^{-1}BA^{-1}$ gives that

$$H(t) = A - tA^{-1}BA^{-1}$$

is a linear approximation for $G(t) = F(t)^{-1}$.

1.7 Conditions for Diagonalizability

Now we look for some more conditions for diagonalizability.

1.7.23 Theorem. *Let $A \in M_n(\mathbb{C})$, with its characteristic polynomial $p_A(t) = \prod_{j=1}^n (t-\lambda_j)$, and $\lambda_i \neq \lambda_j$ for $j \neq k$, then A is diagonalizable.*

Proof. We'll show that there's a linearly independent set of n eigenvectors. Then by what we've proved in the last lecture, we'll be done. Let $\{x_1, x_2, \ldots, x_n\}$ be such that $x_j \in \mathbb{C}^n$ with $Ax_j = \lambda_j x_j$. If $\{x_1, x_2, \ldots, x_n\}$ were linearly dependent, then there is a linear combination

$$\alpha_1 x_{j_1} + \alpha_2 x_{j_2} + \ldots + \alpha_s x_{j_s} = 0$$

with $s \leq n$, and all $\alpha_j \neq 0$. Let r be smallest such $s \leq n$, and assume with possible renumbering that $j_i = i$. Then applying A to the linear combination gives us

$$A(\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n) = \alpha_1 \lambda_1 x_1 + \alpha_2 \lambda_2 x_2 + \ldots + \alpha_n \lambda_n x_n = 0$$

multiplying the previous equation with λ_r and then subtracting gives us

$$\begin{aligned} 0 &= (\alpha_1\lambda_1 x_1 + \alpha_2\lambda_2 x_2 + \ldots + \alpha_n\lambda_n x_n) - (\alpha_1\lambda_r x_1 + \alpha_2\lambda_r x_2 + \ldots + \alpha_r\lambda_r x_r) \\ &= \alpha_1(\lambda_1 - \lambda_r)x_1 + \alpha_2(\lambda_2 - \lambda_r)x_2 + \ldots + \alpha_{r-1}(\lambda_{r-1} - \lambda_r)x_{r-1} + \alpha_r(\lambda_r - \lambda_r)x_r \\ &= \alpha_1(\lambda_1 - \lambda_r)x_1 + \alpha_2(\lambda_2 - \lambda_r)x_2 + \ldots + \alpha_{r-1}(\lambda_{r-1} - \lambda_r)x_{r-1} \end{aligned}$$

which contradicts the minimality of r. □

Unfortunately this is just a sufficient condition, as in the next example.

1.7.24 Example. Consider the matrix

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Clearly A is diagonalizable. But the characteristic polynomial $p_A(x) = x^2(1-x)$ does not satisfy the conditions of the above theorem.

1.7.25 Definition. If for $A \in M_n(\mathbb{C})$, with characteristic polynomial

$$p_A(t) = (t-\lambda_1)^{m_1}(t-\lambda_2)^{m_2}\ldots(t-\lambda_r)^{m_r}$$

then we say that the eigenvalue λ_j has algebraic multiplicity m_j. We call $\mathrm{null}(\lambda_j I - A)$, the geometric multiplicity of λ_j

1.7.26 Lemma. *If $A \in M_n$ has eigenvalue λ, and characteristic polynomial $p_A(t) = (t-\lambda)^m q(t)$, with $q(\lambda) = 0$, then $r = nul(\lambda I - A)) \leq m$*

Proof. Choose a basis $\{x_1, x_2, \ldots, x_r\}$ of eigenvectors, spanning $E_\lambda = \{x \in \mathbb{C}^n \ : \ Ax = \lambda x\}$. Complete it to a basis $\{x_1, x_2, \ldots, x_n\}$ of $\mathbb{C}^n$. Let $S = [x_1, x_2, \ldots, x_n]$.

Then $AS = [\lambda x_1, \lambda x_2, \ldots \lambda x_r, y_{r+1}, \ldots y_n]$ with some vectors $y_{r+1}, \ldots, y_n$. Then

$$S^{-1}AS = \begin{bmatrix} \lambda I_r & 0 \\ 0 & C \end{bmatrix}$$

and we get

$$\begin{aligned} \det(tI - A) &= \det(tI - S^{-1}AS) \\ &= (t-\lambda)^r \det(t - C) \end{aligned}$$

Thus we conclude that algebraic multiplicity of λ is at least equal to r. □

1.7.27 Remark. By the definition the algebraic multiplicity of a matrix $A \in M_n(\mathbb{C})$ is the number of roots of its characteristic polynomial $p_A(x)$, counted upto multiplicities. But as a consequence of the fundamental theorem of algebra, every polynomial decomposes as linear factors in $\mathbb{C}$. Thus $p_A(x)$, being a polynomial of degree n, has n roots in $\mathbb{C}$, forcing its algebraic multiplicity equal to n.

1.7.28 Theorem. *The matrix $A \in M_n(\mathbb{C})$ is diagonalizable if and only if the algebraic and geometric multiplicities are equal for each eigenvalue.*

Proof. Let λ_i, λ_j be two distinct eigenvalues of A with their eigenspaces E_i, E_j respectively. We claim $E_i \cap E_j = \{\mathbf{0}\}$. If not, there exists $\mathbf{0} \neq x \in E_i \cap E_j$, and $Ax = \lambda_i x = \lambda_j x$. Then since $x \neq \mathbf{0}$, $\mathbf{0} = (\lambda_i - \lambda_j)x$ forces $\lambda_i = \lambda_j$ contradicting our assumption.

Next, we claim that if $\{v_1, v_2, \ldots, v_{r_1}\}$ and $\{u_1, u_2, \ldots, u_{r_2}\}$ form a basis for E_i and E_j respectively, then $\{v_1, v_2, \ldots, v_{r_1}, u_1, u_2, \ldots, u_{r_2}\}$ is linearly independent. If not there will be scalars $\alpha_1, \alpha_2, \ldots, \alpha_{r_1}$, $\beta_1, \beta_2, \ldots, \beta_{r_2}$, such that

$$\sum_{i=1}^{r_1} \alpha_i v_i + \sum_{j=1}^{r_2} \beta_j u_j = 0$$

Since we know that $E_j \cap E_j = \{\mathbf{0}\}$, this forces

$$\sum_{i=1}^{r_1} \alpha_i v_i = -\sum_{j=1}^{r_2} \beta_j u_j = \mathbf{0}$$

Linear independence of $u_1, u_2, \ldots, u_{r_2}$, $v_1, v_2, \ldots, v_{r_1}$ forces all $\alpha_i = 0 = \beta_j$ proving the linear independence of $\{v_1, v_2, \ldots, v_{r_1}, u_1, u_2, \ldots, u_{r_2}\}$.

Now by using induction over distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ of A, we get a basis for $E_1 + E_2 \ \ldots + E_k$ with dimension $r = \sum_{i=1}^{k} r_i$.

If algebraic and geometric multiplicities equal then $r = n$, and we have a basis of eigenvectors. Otherwise if $r < n$, then we do not have such a basis of eigenvectors. And since existence of a basis of eigenvectors characterizes diagonalizability (from previous lecture), our if and only if statement is proved. □

In the next lecture, we'll look when multiple matrices can be simultaneously diagonalizable with the same S matrix.