Math 504: Advanced Linear Algebra

Hugo Woerdeman, with edits by Darij Grinberg*

October 1, 2021 (unfinished!)

Contents

Math 504 Lecture 6

1. Schur triangularization (cont'd)

1.1. Application: Cayley–Hamilton theorem

Let us recall some properties of the characteristic polynomial of an $n \times n$-matrix A:

Definition 1.1.1. Let $\mathbb{F}$ be a field. Let $A \in \mathbb{F}^{n \times n}$ be an $n \times n$-matrix over $\mathbb{F}$.
The **characteristic polynomial** p_A of A is defined to be the polynomial

$$\det(tI_n - A) \in \underbrace{\mathbb{F}[t]}_{\substack{\text{ring of all polynomials} \\ \text{in the indeterminate } t \\ \text{with coefficients in } \mathbb{F}}}.$$

*Drexel University, Korman Center, 15 S 33rd Street, Philadelphia PA, 19104, USA

Example 1.1.2. Let $n = 2$ and $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then,

$$\begin{aligned} tI_n - A = tI_2 - A &= t\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \begin{pmatrix} t & 0 \\ 0 & t \end{pmatrix} - \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} t-a & -b \\ -c & t-d \end{pmatrix}, \end{aligned}$$

so that

$$\begin{aligned} p_A &= \det(tI_n - A) = \det\begin{pmatrix} t-a & -b \\ -c & t-d \end{pmatrix} = (t-a)(t-d) - (-b)(-c) \\ &= t^2 - (a+d)\,t + (ad - bc). \end{aligned}$$

Example 1.1.3. Let $n = 3$ and $A = \begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix}$. Then,

$$tI_n - A = tI_3 - A = \begin{pmatrix} t-a & -b & -c \\ -a' & t-b' & -c' \\ -a'' & -b'' & t-c'' \end{pmatrix},$$

so that

$$\begin{aligned} p_A &= \det\begin{pmatrix} t-a & -b & -c \\ -a' & t-b' & -c' \\ -a'' & -b'' & t-c'' \end{pmatrix} \\ &= t^3 - \left(a + b' + c''\right) t^2 + \left(ab' - ba' + ac'' - ca'' + b'c'' - b''c'\right) t \\ &\qquad - \left(ab'c'' - ab''c' - ba'c'' + ba''c' + ca'b'' - ca''b'\right). \end{aligned}$$

By the way, some authors define p_A to be $\det(A - tI_n)$ instead of $\det(tI_n - A)$. This differs from our definition only by a factor of $(-1)^n$, so the difference is insignificant.

Proposition 1.1.4 (properties of the char. poly.)**.** Let $\mathbb{F}$ be a field. Let $A \in \mathbb{F}^{n\times n}$ be an $n \times n$-matrix over $\mathbb{F}$.

(a) The characteristic polynomial p_A is a monic polynomial in t of degree n. (That is, its leading term is t^n.)

(b) The constant term of p_A is $(-1)^n \det A$.

(c) The t^{n-1}-coefficient of p_A is $-\operatorname{Tr} A$. (Recall that $\operatorname{Tr} A$ is defined to be the sum of all diagonal entries of A; this is known as the **trace** of A.)

Proof. All of this should be more or less clear from the examples. Part **(b)** follows from observing that the constant term of p_A is $p_A(0) = \det(0I_n - A) = \det(-A) = (-1)^n \det A$.

For details, I'll give references in the notes. □

Theorem 1.1.5 (Cayley–Hamilton theorem)**.** Let $\mathbb{F}$ be a field. Let $A \in \mathbb{F}^{n\times n}$ be an $n \times n$-matrix. Then,

$$p_A(A) = 0.$$

(The "0" on the RHS is the zero matrix.)

Example 1.1.6. Let $n = 2$ and $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then, as we know,

$$p_A = t^2 - (a+d)\,t + (ad - bc).$$

Thus,

$$\begin{aligned} p_A(A) &= A^2 - (a+d)\,A + (ad-bc)\,I_2 \\ &= \begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 - (a+d)\begin{pmatrix} a & b \\ c & d \end{pmatrix} + (ad-bc)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0. \end{aligned}$$

Remark 1.1.7. You cannot argue that $p_A(A) = \det(AI_n - A)$ "by substituting A for t into $p_A = \det(tI_n - A)$". Indeed, $tI_n - A$ is a matrix whose entries are polynomials in t. If you substitute A for t into it, it will become a matrix whose entries are matrices. First of all, it is not quite clear how to take the determinant of such a matrix; second, this matrix is not $AI_n - A$. For example, for $n = 2$, plugging A for t in $tI_n - A$ gives

$$\begin{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} - a & -b \\ -c & \begin{pmatrix} a & b \\ c & d \end{pmatrix} - d \end{pmatrix},$$

which doesn't quite look like $AI_n - A$ (which is the zero matrix). There **is** a correct proof of the Cayley–Hamilton theorem along the lines of "substituting A for t", but it requires a lot of work.

There are various proofs of the Cayley–Hamilton theorem (I'll give references in the notes). We will here only prove it for $\mathbb{F} = \mathbb{C}$:

Proof of the Cayley–Hamilton theorem for $\mathbb{F} = \mathbb{C}$. Assume that $\mathbb{F} = \mathbb{C}$. The Schur triangularization theorem shows that A is unitarily similar to an upper-triangular matrix. Hence, A is similar to an upper-triangular matrix (because unitarily similar matrices always are similar). In other words, there exists an invertible matrix U and an upper-triangular matrix T such that $A = UTU^{-1}$. Consider these U and T.

Now, let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the diagonal entries of T. Then, by Proposition 2.3.4, these diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A (with algebraic multiplicities). Hence,

$$p_A = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n)$$

(since p_A is monic, and the roots of p_A are precisely the eigenvalues of A with algebraic multiplicities).

Now, substituting A for t in the polynomial identity $p_A = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n)$, we obtain

$$p_A(A) = (A - \lambda_1 I_n)(A - \lambda_2 I_n) \cdots (A - \lambda_n I_n).$$

For each $i \in [n]$, we have

$$\underbrace{A}_{=UTU^{-1}} - \lambda_i \underbrace{I_n}_{=UU^{-1}} = UTU^{-1} - \lambda_i UU^{-1} = U(T - \lambda_i I_n) U^{-1}.$$

Hence, the above equality becomes

$$\begin{aligned} p_A(A) &= (A - \lambda_1 I_n)(A - \lambda_2 I_n) \cdots (A - \lambda_n I_n) \\ &= U(T - \lambda_1 I_n) \underbrace{U^{-1}U}_{=I_n} (T - \lambda_2 I_n) U^{-1} \cdots U(T - \lambda_n I_n) U^{-1} \\ &= U(T - \lambda_1 I_n)(T - \lambda_2 I_n) \cdots (T - \lambda_n I_n) U^{-1}. \end{aligned}$$

Thus, it suffices to show that

$$(T - \lambda_1 I_n)(T - \lambda_2 I_n) \cdots (T - \lambda_n I_n) = 0.$$

Let us show this on an example for $n = 3$:

$$\begin{aligned} & T = \begin{pmatrix} \lambda_1 & * & * \\ 0 & \lambda_2 & * \\ 0 & 0 & \lambda_3 \end{pmatrix} \\ \Longrightarrow \quad & T - \lambda_1 I_n = \begin{pmatrix} 0 & * & * \\ 0 & \lambda_2 - \lambda_1 & * \\ 0 & 0 & \lambda_3 - \lambda_1 \end{pmatrix} \\ \Longrightarrow \quad & (T - \lambda_1 I_n)(T - \lambda_2 I_n) = \begin{pmatrix} 0 & * & * \\ 0 & \lambda_2 - \lambda_1 & * \\ 0 & 0 & \lambda_3 - \lambda_1 \end{pmatrix} \begin{pmatrix} \lambda_1 - \lambda_2 & * & * \\ 0 & 0 & * \\ 0 & 0 & \lambda_3 - \lambda_1 \end{pmatrix} \\ & \qquad = \begin{pmatrix} 0 & 0 & * \\ 0 & 0 & * \\ 0 & 0 & * \end{pmatrix} \\ \Longrightarrow \quad & (T - \lambda_1 I_n)(T - \lambda_2 I_n)(T - \lambda_3 I_n) = \begin{pmatrix} 0 & 0 & * \\ 0 & 0 & * \\ 0 & 0 & * \end{pmatrix} \begin{pmatrix} \lambda_1 - \lambda_3 & * & * \\ 0 & \lambda_2 - \lambda_3 & * \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

The general proof follows the same pattern: Every time you add a new factor, one more column of your matrix becomes 0. Formally speaking, this means that you are proving the following fact by induction on j:

For each $j \in \{0, 1, \ldots, n\}$, the first j columns of the matrix

$$(T - \lambda_1 I_n)(T - \lambda_2 I_n) \cdots (T - \lambda_j I_n)$$

are 0.

Once this is proved, we can apply this to $j = n$, and conclude that the first n columns of the matrix

$$(T - \lambda_1 I_n)(T - \lambda_2 I_n) \cdots (T - \lambda_n I_n)$$

are 0. But this means that the whole matrix is 0, qed. □

1.2. Sylvester's equation

Definition 1.2.1. Let $A \in \mathbb{C}^{n \times n}$. Then, the **spectrum** of A is defined to be the set of all eigenvalues of A. This spectrum is denoted by $\sigma(A)$ (or by spec A).

Theorem 1.2.2. Let A be an $n \times n$-matrix, and let B be an $m \times m$-matrix (both with complex entries). Let C be an $n \times m$-matrix. Then, the following statements are equivalent:

- $\mathcal{U}$: There is a **unique** matrix $X \in \mathbb{C}^{n \times m}$ such that $AX - XB = C$.
- $\mathcal{V}$: We have $\sigma(A) \cap \sigma(B) = \varnothing$.

Example 1.2.3. Let us take $n = 1$ and $m = 1$. In this case, A, B and C are 1×1-matrices, so we can view them as scalars. Let us therefore write a, b and c for them. Then, the theorem says that the following statements are equivalent:

- $\mathcal{U}$: There is a **unique** complex number x such that $ax - xb = c$.
- $\mathcal{V}$: We have $\{a\} \cap \{b\} = \varnothing$ (that is, $a \neq b$).

This is not surprising, because the equation $ax - xb = c$ has a unique solution (namely, $x = \dfrac{c}{a-b}$) when $a \neq b$, and otherwise has either none or infinitely many solution.

The equation $AX - XB = C$ in the Theorem is known as **Sylvester's equation**. Because the X is on different sides in AX and in XB, it cannot be factored out (matrices do not generally commute).

Proof of the $\mathcal{V} \Longrightarrow \mathcal{U}$ part of the theorem. First, observe that the matrix space $\mathbb{C}^{n\times m}$ is itself a $\mathbb{C}$-vector space of dimension nm.

Consider the map

$$\begin{aligned} L : \mathbb{C}^{n\times m} &\to \mathbb{C}^{n\times m}, \\ X &\mapsto AX - XB. \end{aligned}$$

This map L is linear, because

$$\begin{aligned} L(\alpha X + \beta Y) &= A(\alpha X + \beta Y) - (\alpha X + \beta Y) B \\ &= \alpha AX + \beta AY - \alpha XB - \beta YB \\ &= \alpha (AX - XB) + \beta (AY - YB) = \alpha L(X) + \beta L(Y). \end{aligned}$$

Thus, L is a linear map between two vector spaces that have the same (finite) dimension. Hence, we have the following equivalence:

$$\begin{aligned} & (L \text{ is surjective (= onto)}) \\ \Longleftrightarrow\ & (L \text{ is injective (= one-to-one)}) \\ \Longleftrightarrow\ & (L \text{ is bijective (= invertible)}). \end{aligned}$$

Now, statement $\mathcal{U}$ is saying that the matrix C has a **unique** preimage under L (that is, there exists a unique $X \in \mathbb{C}^{n\times m}$ such that $L(X) = C$). As we know from general properties of linear maps, this is true whenever L is bijective, and false otherwise. So statement $\mathcal{U}$ is equivalent to L being bijective.

Now, let us prove that $\mathcal{V} \Longrightarrow \mathcal{U}$. To wit, we will show that L is **injective**. This will imply that L is bijective (by the above equivalence), and therefore statement $\mathcal{U}$ will follow.

In order to prove that a linear map is injective, it suffices to show that its kernel (= nullspace) is 0. So let $X \in \operatorname{Ker} L$; we will show that $X = 0$.

From $X \in \operatorname{Ker} L$, we get $L(X) = 0$. Since $L(X) = AX - XB$, this means that $AX - XB = 0$. In other words, $AX = XB$. Hence,

$$A^2X = A \underbrace{AX}_{=XB} = \underbrace{AX}_{=XB} B = XBB = XB^2.$$

Similarly,

$$A^3X = XB^3, \qquad A^4X = XB^4, \qquad A^5X = XB^5, \qquad \ldots.$$

That is,

$$A^kX = XB^k \qquad \text{for each } k \in \mathbb{N}.$$

(Strictly speaking, this is proved by induction on k.)

Therefore, I claim that

$$f(A)\,X = Xf(B) \qquad \text{for any polynomial } f \in \mathbb{C}[t].$$

(Indeed, if we write the polynomial f as $f = \sum_{k=0}^{m} f_k t^k$ with $f_k \in \mathbb{C}$, then

$$f(A)X = \sum_{k=0}^{m} f_k \underbrace{A^k X}_{=XB^k} = \sum_{k=0}^{m} f_k X B^k = X \underbrace{\sum_{k=0}^{m} f_k B^k}_{=f(B)} = Xf(B),$$

as desired.)

Apply this claim to $f = p_A$. We obtain

$$p_A(A)X = Xp_A(B) = X(B - \lambda_1 I_n)(B - \lambda_2 I_n) \cdots (B - \lambda_n I_n),$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A (with algebraic multiplicities), because

$$p_A = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n).$$

Thus,

$$X(B - \lambda_1 I_n)(B - \lambda_2 I_n) \cdots (B - \lambda_n I_n) = \underbrace{p_A(A)}_{\substack{=0 \\ \text{(by Cayley–Hamilton)}}} X = 0.$$

We want to prove that $X = 0$. This would follow from this equation if we knew that the factors

$$B - \lambda_1 I_n, \ B - \lambda_2 I_n, \ \ldots, \ B - \lambda_n I_n$$

are invertible (because then we can cancel these factors). However, they are indeed invertible, because each λ_i is an eigenvalue of A and therefore **not** an eigenvalue of B (since $\sigma(A) \cap \sigma(B) = \varnothing$). This completes the proof of $\mathcal{V} \Longrightarrow \mathcal{U}$. $\square$

Maybe $\mathcal{U} \Longrightarrow \mathcal{V}$ will be homework. Also a nice exercise(?):

$$\sigma(L) = \sigma(A) - \sigma(B) = \{\lambda - \mu \ \mid \ \lambda \in \sigma(A) \text{ and } \mu \in \sigma(B)\}.$$