1212: Linear Algebra II

Dr. Vladimir Dotsenko (Vlad)

Lecture 23

Proof of Cayley–Hamilton theorem

The Cayley–Hamilton theorem stated yesterday claims that "every matrix is a root of its own characteristic polynomial", that is that if we consider the characteristic polynomial $\chi_A(t) = \det(A-tI) = a_0+a_1t+\cdots+a_nt^n$ for the given $n \times n$-matrix A, then we have

$$\chi_A(A) = a_0I + a_1A + \cdots + a_nA^n = 0.$$

Today we shall discuss two proofs of this result.

Proof 1. Let $\lambda_1, \ldots, \lambda_k$ be all different complex eigenvalues of A. Then of course there exist positive integers $m_1, \ldots, m_k$ such that

$$\chi_A(t) = \det(A - tI) = a_n(t - \lambda_1)^{m_1} \cdots (t - \lambda_k)^{m_k},$$

and hence

$$\chi_A(A) = a_n(A - \lambda_1 I)^{m_1} \cdots (A - \lambda_k I)^{m_k}.$$

At the same time, we know that for some positive integers $n_1, \ldots, n_k$ we have

$$V = \mathrm{Ker}(A - \lambda_1 I)^{n_1} \oplus \cdots \oplus \mathrm{Ker}(A - \lambda_k I)^{n_k}.$$

In this decomposition, all eigenvalues of A on $\mathrm{Ker}(A - \lambda_i I)^{n_i}$ are equal to λ_i, so the total multiplicity of that eigenvalue, that is m_i, is equal to the sum of lengths of the threads we obtain from that subspace. The number n_i, that is the exponent which annihilates the linear transformation $A - \lambda_i I$, is equal to the maximum of all lengths of threads, since for a thread of length s, the power $(A-\lambda_i I)^s$ annihilates all vectors of that thread, and the power $(A - \lambda_i I)^{s-1}$ does not. This shows that $m_i \geqslant n_i$ (the first of them is sum of lengths of threads, the second is the maximum of lengths of threads). Therefore, the linear transformation

$$\chi_A(A) = a_n(A - \lambda_1 I)^{m_1} \cdots (A - \lambda_k I)^{m_k}$$

annihilates each of the subspaces $\mathrm{Ker}(A - \lambda_i I)^{n_i}$, therefore annihilates their direct sum, that is V, therefore vanishes, as required. □

The second proof uses a bit of analysis that you would learn in due course in other modules.

Proof 2. Let us first assume that A is diagonalisable, that is has a basis of eigenvectors $v_1, \ldots, v_n$, with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then

$$\chi_A(t) = \det(A - tI) = a_n(t - \lambda_1) \cdots (t - \lambda_n),$$

and hence

$$\chi_A(A) = a_n(A - \lambda_1 I) \cdots (A - \lambda_n I).$$

In this product (of commuting factors), there is a factor to annihilate each eigenvector v_i, since $(A-\lambda_i I)v_i = 0$. Therefore, each element of the basis is annihilated by $\chi_A(A)$, therefore every vector is annihilated by that transformation, therefore $\chi_A(A) = 0$.

To handle an arbitrary linear transformation, note that every matrix is a limit of diagonalisable matrices (e.g. one can take the Jordan normal form and change the diagonal entries a little bit so that they are all distinct), and the expression $\chi_A(A)$ is a continuous function of A, so if it vanishes on all diagonalisable matrices, it must vanish everywhere. □

Examples of computations with Jordan normal forms

Let us consider an example of how Jordan decompositions can be used in particular computations. One instance where it is useful to compute powers of matrices is when dealing with recurrent sequences. An example of that sort was considered before in the first semester when we dealt with Fibonacci numbers. We shall now consider a similar question where however Jordan decomposition will be important.

Let us consider the sequence defined as follows: $x_0 = 7$, $x_1 = 3$, $x_{n+2} = -10x_{n+1} - 25x_n$ for $n \geqslant 0$. In order to find a closed formula for x_n (that is, a formula that expresses it in terms of n, without the need to compute all the previous terms of the sequence), we consider the vectors $v_n = \begin{pmatrix} x_n \\ x_{n+1} \end{pmatrix}$, for which we have

$$v_{n+1} = \begin{pmatrix} x_{n+1} \\ x_{n+2} \end{pmatrix} = \begin{pmatrix} x_{n+1} \\ -10x_{n+1} - 25x_n \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -25 & -10 \end{pmatrix} \begin{pmatrix} x_n \\ x_{n+1} \end{pmatrix} = Av_n,$$

where $A = \begin{pmatrix} 0 & 1 \\ -25 & -10 \end{pmatrix}$. Thus, $v_n = A^n v_0$, so in order to compute x_n, it is enough to find a formula for A^n.

We have $\chi_A(t) = \det(A - tI) = t^2 + 10t + 25 = (t+5)^2$, so -5 is the only eigenvalue. The kernel of $A + 5I$ is spanned by the vector $\begin{pmatrix} 1 \\ -5 \end{pmatrix}$. We take the vector $v = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ outside the kernel of $A + 5I$ that would compensate for the missing pivot; we have $(A + 5I)v = \begin{pmatrix} 1 \\ -5 \end{pmatrix}$, so the columns of $C = \begin{pmatrix} 0 & 1 \\ 1 & -5 \end{pmatrix}$ form a Jordan basis. Thus, $C^{-1}AC = \begin{pmatrix} -5 & 0 \\ 1 & -5 \end{pmatrix}$, and

$$A^n = C \begin{pmatrix} -5 & 0 \\ 1 & -5 \end{pmatrix}^n C^{-1} = \begin{pmatrix} (-5)^n - n(-5)^n & n(-5)^{n-1} \\ -n(-5)^{n+1} & (-5)^n + n(-5)^n \end{pmatrix}.$$

Finally, $v_n = A^n \begin{pmatrix} 7 \\ 3 \end{pmatrix} = \begin{pmatrix} (-5)^{n-1}(38n - 35) \\ (-5)^n(38n + 3) \end{pmatrix}$, so $x_n = (-5)^{n-1}(38n - 35)$.

Another example where computing powers of matrices is important is suggested by probabilistic models known as Markov chains. Let us mention a simple example. Suppose that a particle can be in two states, that we call 1 and 2. Suppose that if it is in the state 1, then with probability p_{11} it remains in that state in one second, and with probability p_{12} changes to the state 2, and similarly, if it is in the state 2, then with probability p_{21} it changes to the state 1, and with probability p_{22} remains in the same state (of course, $p_{11} + p12 = 1$ and $p_{21} + p_{22} = 1$). Then, if in the beginning we only know that the particle is in the state 1 with probability p and in the state 2 with probability $q = 1 - p$, then in one second the probabilities change to $p' = pp_{11} + qp_{21}$ and $q' = pp12 + qp22$, in other words,

$$\begin{pmatrix} p' \\ q' \end{pmatrix} = \begin{pmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix},$$

and the probabilities after n seconds are computed using the n-th power of the "transfer matrix" $\begin{pmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{pmatrix}$.