MA1112: Linear Algebra II

Dr. Vladimir Dotsenko (Vlad)

Lecture 11

Uniqueness of the normal form

If the matrix of a linear transformation can be made diagonal, then the diagonal entries are eigenvalues. In case of Jordan normal forms, we have blocks of varying sizes corresponding to eigenvalues. It is natural to ask whether the sizes of those blocks are uniquely defined, or different choices of bases can lead to different block structures.

Let us first focus on the case of a nilpotent transformation with $\varphi^k = 0$. We denote by m_d the number of threads of length d, where $1 \leqslant d \leqslant k$. In that case, we have

$$\begin{aligned} m_k &= \dim N_k - \dim N_{k-1}, \\ m_{k-1} + m_k &= \dim N_{k-1} - \dim N_{k-2}, \\ &\dots \\ m_2 + \ldots + m_k &= \dim N_2 - \dim N_1, \\ m_1 + m_2 + \ldots + m_k &= \dim N_1, \end{aligned}$$

so the numbers of threads of various lengths are uniquely determined by dimensions of kernels of powers, which are intrinsic characteristics of the linear transformation φ that do not depend on any choices.

Next, we consider an arbitrary linear transformation. Suppose that λ_1 is one of its eigenvalues, and let $B_{\lambda_1} = \varphi - \lambda_1 I$. In course of the proof, we established that $V = \mathrm{Ker}(B_{\lambda_1}^{k_1}) \oplus \mathrm{Im}(B_{\lambda_1}^{k_1})$, where k_1 is the first place where the kernels of powers of B_{λ_1} stabilise, so that $\mathrm{Ker}(B_{\lambda_1}^{k_1}) = \mathrm{Ker}(B_{\lambda_1}^{k_1+1}) = \ldots$. This number k_1, as well as the subspaces $\mathrm{Ker}(B_{\lambda_1}^{k_1})$ and $\mathrm{Im}(B_{\lambda_1}^{k_1})$ are intrinsic characteristics of the linear transformation φ. Thus, sizes of blocks with λ_1 on the diagonal are intrinsically determined by the previous observation, and the others are intrinsically determined by induction.

Overall, we conclude that up to the order of blocks, the Jordan normal form is an invariant of a linear transformation; moreover, it is the only invariant, since we already established that every linear transformation has a Jordan normal form.

This statement can also be formalised in a slightly different way. Two $n \times n$ matrices A and B are said to be *similar* if $A = C^{-1}BC$ for some invertible matrix C. Geometrically, this means that A and B represent the same linear transformations with respect to different bases (and C is the transition matrix between two bases). We established that A and B are similar if and only if their Jordan forms are the same.

Cayley–Hamilton theorem

The celebrated Cayley–Hamilton theorem states that "every matrix is a root of its own characteristic polynomial", that is that if we consider the characteristic polynomial $\chi_A(t) = \det(A - tI) = a_0 + a_1 t + \cdots + a_n t^n$ for the given matrix A, then we have

$$\chi_A(A) = a_0 I + a_1 A + \cdots + a_n A^n = 0.$$

Of course, it is tempting to say that this theorem is obvious, because $\chi_A(A) = \det(A - A \cdot I) = \det(0) = 0$. However, $A - tI$ is a matrix whose entries depend on t, and we cannot simply substitute $t = A$ in those entries! Another way to see the problem is to note that our "proof" would be equally applicable to $\operatorname{tr}(A - tI) = \operatorname{tr}(A) - t \operatorname{tr}(I) = \operatorname{tr}(A) - nt$, but substituting $t = A$ in that polynomial yields $\operatorname{tr}(A)I - nA$, which only is equal to zero for matrices A proportional to I. Today we shall discuss two mathematically sound 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

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

In this decomposition, all eigenvalues of A on $\operatorname{Ker}(A - \lambda_i I)^{n_i}$ 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 $\operatorname{Ker}(A - \lambda_i I)^{n_i}$, 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. □