MATH 325 Q1: LINEAR ALGEBRA III, PART 3

The Jordan normal form

We start by recalling some facts which we already know. Let A be a matrix of size $n\times n$-matrix with coefficients in F. It defines an F-linear map

$$\alpha_A \,:\, F^n \longrightarrow F^n\,,\; v \longmapsto A\cdot v\,.$$

As we mentioned before A is precisely the matrix of α_A in the standard basis $e_1,\ldots,e_n$.

If $v_1,\ldots,v_n$ is another basis of F^n we can express $A\cdot v_j$ in terms of this basis:

$$A\cdot v_j \;=\; \sum_{i=1}^{n} b_{ij}\cdot v_i.$$

The matrix $B=(b_{ij})$ is the matrix of α_A with respect to the basis $v_1,\ldots,v_n$ and we have the conjugacy property

$$B \;=\; S^{-1}\cdot A\cdot S\,,$$

where $S=(s_{ij})$ is the matrix whose columns are the vectors $v_1,\ldots,v_n$:

$$v_j \;=\; \sum_{i=1}^{n} s_{ij}\cdot e_i \;=\; \begin{pmatrix} s_{1j} \\ s_{2j} \\ \vdots \\ s_{nj} \end{pmatrix}.$$

Going in the reverse direction, let T be an invertible matrix and $C=T^{-1}\cdot A\cdot T$. Since T is invertible the column vectors

$$w_1 = \begin{pmatrix} t_{11} \\ t_{21} \\ \vdots \\ t_{n1} \end{pmatrix},\; w_2 = \begin{pmatrix} t_{12} \\ t_{22} \\ \vdots \\ t_{n2} \end{pmatrix},\;\ldots\;,\; w_n = \begin{pmatrix} t_{1n} \\ t_{2n} \\ \vdots \\ t_{nn} \end{pmatrix}$$

form another basis of F^n and $C=(c_{ij})$ is the matrix of α_A with respect to this basis. As we mentioned before our aim is for a given matrix A to find a conjugate matrix $T^{-1}AT$ which is as simple as possible (equivalently, to find another basis of F^n in which the matrix of the map α_A is simple).

To understand what we can expect we first consider matrices of small size.

The case of 2×2-matrices.

Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

be a complex 2×2-matrix. The corresponding linear map $\alpha_A : \mathbb{C}^2 \longrightarrow \mathbb{C}^2$, $x \mapsto A \cdot x$, is given by the formula

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \longrightarrow \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix}.$$

Since any polynomial in $\mathbb{C}$ has a root we have

$$P_A(T) = (T - \lambda_1) \cdot (T - \lambda_2),$$

where λ_1, λ_2 are the eigenvalues of A, which may be equal.

Case (i): $\lambda_1 \neq \lambda_2$. Let v_1 and v_2 be eigenvectors for λ_1 and λ_2, respectively. Since $\lambda_1 \neq \lambda_2$ the vectors v_1, v_2 are linear independent and so they form a basis of $\mathbb{C}^2$. With respect to this basis the matrix of α_A is $\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$. (In this case we say that A is diagonalizable.)

Case (ii): $\lambda_1 = \lambda_2$. We denote this complex number by λ. In this case the matrix A can be diagonalizable or not. If there are two linear independent eigenvectors v, w for λ then the matrix of α_A with respect to the basis v, w is equal $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$, and so A is diagonalizable.

If A is not diagonalizable then the dimension of the eigenspace E_λ for λ is 1. Let v be an eigenvector for λ and $w' \in \mathbb{C}^2$ any other vector, such that v, w' is a basis of $\mathbb{C}^2$. Then $\alpha_A(v) = A \cdot v = \lambda \cdot v$ and $\alpha_A(w') = e \cdot v + f \cdot w'$ for some complex numbers e, f, and so the matrix of α_A with respect to this basis is $\begin{pmatrix} \lambda & e \\ 0 & f \end{pmatrix}$. Since this matrix is conjugate to A it has the same characteristic polynomial $P_A(T) = (T - \lambda)^2$. This implies $f = \lambda$. Hence A is conjugate to

$$B' := \begin{pmatrix} \lambda & e \\ 0 & \lambda \end{pmatrix}.$$

Note that the complex number e can not be zero since then w' would be an eigenvector and so the eigenspace of A for the eigenvalue λ would have dimension 2 (because it would then be generated by the linear independent vectors v and w').

We replace now the basis v, w' by v and $w := e^{-1} \cdot w'$. Then we have $\alpha_A(v) = \lambda \cdot v$ and

$$\begin{aligned} \alpha_A(w) &= \alpha_A(e^{-1} \cdot w') \\ &= e^{-1} \cdot \alpha_A(w') \\ &= e^{-1} \cdot (\lambda \cdot w' + e \cdot v) \\ &= \lambda \cdot w + v \,, \end{aligned}$$

and so the matrix of α_A with respect to this basis is

$$B = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} .$$

(We could reach the same if we replace v, w' by $e \cdot v$ and w.)

Summary: if A is not diagonalizable A is similar to the so-called 2×2-*Jordan block*, *i.e.* there exists a basis of $\mathbb{C}^2$, such that the linear map $\alpha_A : \mathbb{C}^2 \longrightarrow \mathbb{C}^2$, $v \mapsto A \cdot v$, has matrix $\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$, where λ is the only eigenvalue of A, with respect to this basis. Such a basis is called a *Jordan basis* for A.

Remark. *The precise definition of Jordan- block and Jordan basis will be given later.*

We illustrate the above consideration with an example of a not diagonalizable matrix.

Example. Let $A = \begin{pmatrix} 1 & 2 \\ -2 & 5 \end{pmatrix} \in \mathrm{M}_{2\times 2}(\mathbb{C})$. Then we have

$$P_A(T) = (T-1) \cdot (T-5) + 4 = T^2 - 6T + 9 = (T-3)^2 \,,$$

and so $\lambda = 3$ is the only eigenvalue of A. One can check that $v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector for the eigenvalue $\lambda = 3$. We extend this to a basis of $\mathbb{C}^2$ by $w' = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

We compute $\alpha_A(v) = 3 \cdot v$ and

$$\alpha_A(w') = A \cdot w' = \begin{pmatrix} 2 \\ 5 \end{pmatrix} = 2 \cdot v + 3 \cdot w' \,.$$

Hence in above notation we have $e = 2$. We replace now the basis v, w' by v and $w = \frac{1}{2} \cdot w'$, *i.e.* by

$$v = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix},$$

and compute $\alpha_A(v) = 3 \cdot v$ and $\alpha_A(w) = v + 3 \cdot w$, *i.e.* α_A has the matrix $B = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$ with respect to this basis. This matrix is conjugate to A and we have

$$B = S^{-1} \cdot A \cdot S$$

for $S = \begin{pmatrix} 1 & 0 \\ 1 & \frac{1}{2} \end{pmatrix}$.

Generalized eigenvalues.

Let V be a finite dimensional complex vector space and $\alpha : V \longrightarrow V$ a $\mathbb{C}$-linear map.

Definition. *A complex number λ is called a generalized eigenvalue of α if there is an integer $l \geq 1$, and a nonzero vector $v \in V$, such that*

$$(\alpha - \lambda \cdot \mathrm{id}_V)^l(v) = 0\,.$$

The vector v is then called a generalized eigenvector for λ.

Notation. We denote the union of the zero vector and all generalized eigenvectors for the generalized eigenvalue λ by K_λ, or more precisely by $K_\lambda(\alpha)$. Thus,

$$K_\lambda(\alpha) = \{v \in V \;\mid\; (\alpha - \lambda \cdot \mathrm{id}_V)^l(v) = 0 \quad \text{for some integer} \quad l\}.$$

Note that we proved before that $K_\lambda(\alpha)$ is a vector subspace in V.

Definition. *The vector subspace $K_\lambda(\alpha)$ is called the generalized eigenspace for the eigenvalue λ.*

Remark. *Note that the eigenspace*

$$E_\lambda(\alpha) = \{v \in V \;\mid\; \alpha(v) = \lambda \cdot v\}$$

for the eigenvalue λ of α is contained in $K_\lambda(\alpha)$.

Example. If $V = \mathbb{C}^n$ and $\alpha = \alpha_A : \mathbb{C}^n \longrightarrow \mathbb{C}^n$, $v \mapsto A \cdot v$, for some $n \times n$-matrix A then

$$(\alpha_A - \lambda\,\mathrm{id}_V)(v) = (A - \lambda \cdot \mathrm{I}_n) \cdot v = A \cdot v - \lambda \cdot v\,,$$

and so $(\alpha_A - \lambda \cdot \mathrm{id}_V)^l(v) = (A - \lambda \cdot \mathrm{I}_n)^l \cdot v$. In this case a generalized eigenvalue (respectively eigenvector) of α_A is also called a generalized eigenvalue (respectively eigenvector) of the matrix A.

We observe that an eigenvalue is also a generalized eigenvalue, and vice versa.

Lemma. *Every generalized eigenvalue is an eigenvalue.*

Proof. Let v be a generalized eigenvector for the generalized eigenvalue λ of α, say we have $(\alpha - \lambda \cdot \mathrm{id}_V)^l(v) = 0$ for some $l \geq 1$. We can assume that l is minimal with this property. Then for the nonzero vector $v' = (\alpha - \lambda \cdot \mathrm{id}_V)^{l-1}(v) \neq 0$ we have

$$(\alpha - \lambda \cdot \mathrm{id}_V)(v') = (\alpha - \lambda \cdot \mathrm{id}_V)^l(v) = 0.$$

The later equation implies $\alpha(v') = \lambda \cdot v'$ and so v' is an eigenvector for λ, *i.e.* λ is an eigenvalue. □

On the other hand eigenvectors for λ are also generalized eigenvectors, but generalized eigenvectors for λ have not to be eigenvectors of α. For instance consider the $\mathbb{C}$-linear map $\alpha_A : \mathbb{C}^2 \longrightarrow \mathbb{C}^2$, $v \mapsto A \cdot v$, where A is the 2×2-matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. The sole eigenvalue (hence also the only generalized eigenvalue) of A is $\lambda = 1$. We have $A \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and so $v = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is not an eigenvector of A. But

$$\mathrm{I}_n - A = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix},$$

and therefore $(\mathrm{I}_n - A)^2 = 0$, which has as a consequence that

$$(\mathrm{I}_n - A)^2 \cdot v = 0,$$

i.e. v is a generalized eigenvector of A.

Lemma. *The set $K_\lambda = K_\lambda(\alpha) \subset V$ is a linear subspace which is α-invariant.*

Proof. This was proved in class before.

□

Digression: Complex polynomials.

Before we continue our discussion of the *generalized eigenspaces* $K_\lambda(\alpha)$ of a linear map $\alpha : V \longrightarrow V$ we prove first two useful results about complex polynomials.

Recall for this that a constant polynomial $\neq 0$ has degree 0 but the zero polynomial has here by convention degree -1. (There are other conventions for the degree of the zero polynomial, some authors define its degree to be $-\infty$.)

Lemma A (division algorithm). *Let $f(T)$ and $g(T)$ be two complex polynomials with $g(T) \neq 0$. Then there exists unique polynomials $h(T)$ and $r(T)$ with $\deg r(T) < \deg g(T)$, such that*

$$f(T) \;=\; h(T)\cdot g(T) + r(T)\,.$$

Proof. We prove this by induction on $d := \deg f(T)$. If $\deg f(T) < e := \deg g(T)$ we set $h(T) = 0$ and $r(T) = f(T)$. So let now $\deg f(T) \geq \deg g(T)$ and assume (by induction) that we have proven the lemma for all polynomials of degree smaller than $\deg f(T)$.

We have $f(T) = a_d T^d + a_{d-1}T^{d-1} + \ldots + a_0$ and $g(T) = b_e T^e + b_{e-1}T^{e-1} + \ldots + b_0$ with $a_d \neq 0$ and $b_e \neq 0$. The polynomial

$$f_1(T) := f(T) - \frac{a_d}{b_e}\cdot T^{d-e}\cdot g(T) \;=\; (a_{d-1} - \frac{a_d b_{e-1}}{b_e})T^{d-1} + \ldots + a_0 - \frac{a_d b_0}{b_e}$$

has degree at most $d-1 < d = \deg f(T)$ and so by induction there is $h_1(T)$ and $r(T)$ with $\deg r(T) < \deg g(T)$, such that $f_1(T) = h_1(T)\cdot g(T) + r(T)$, and so we have

$$f(T) = (h_1(T) + \frac{a_d}{b_e}\cdot T^{d-e})\cdot g(T) + r(T)\,.$$

Setting $h(T) := h_1(T) + \frac{a_d}{b_e}\cdot T^{d-e}$ finishes the proof. □

Lemma B. *Let $f_1(T), \ldots, f_m(T)$ be complex polynomials, which are not all zero. Then there exists complex polynomials $h_1(T), \ldots, h_m(T)$, such that the polynomial*

$$l(T) = h_1(T)\cdot f_1(T) + \ldots + h_m(T)\cdot f_m(T)$$

divides all polynomials $f_i(T)$, i.e. $f_i(T) = g_i(T)\cdot l(T)$ for some $g_i(T) \in \mathbb{C}[T]$ for all $1 \leq i \leq m$.

Proof. Let

$$U \;:=\; \Big\{\, \sum_{i=1}^{m} a_i(T)\cdot f_i(T) \;\Big|\; a_i(T) \in \mathbb{C}[T]\,,\; i = 1,\ldots,m \,\Big\}\,.$$

Let $l(T)$ be a polynomial of minimal but ≥ 0 degree in the set U, say $l(T) = \sum_{i=1}^{m} c_i(T)\cdot f_i(T)$ for some $c_i(T) \in \mathbb{C}[T]$. Note that $l(T) \neq 0$ since $\deg l(T) \geq 0$ by assumption.

By Lemma A above we find $g_i(T)$ and $r_i(T)$ with $\deg r_i(T) < \deg l(T)$, such that

$$f_i(T) = g_i(T) \cdot l(T) + r_i(T) \tag{1}$$

for all $1 \leq i \leq m$. This is equivalent to

$$r_i(T) = (1 - c_i(T) \cdot g_i(T)) \cdot f_i(T) \; - \sum_{j \neq i} (g_i(T) \cdot c_j(T)) \cdot f_j(T) \,,$$

and so $r_i(T)$ is in the set U for all $1 \leq i \leq m$. Since $l(T)$ is of degree strictly bigger than $r_i(T)$ and also has the smallest degree ≥ 0 in the set U we conclude that $r_i(T) = 0$ for all $1 \leq i \leq m$, and so $f_i(T) = g_i(T) \cdot l(T)$ as desired. □

This has the following consequence.

Corollary. *Let $f_1(T), \ldots, f_m(T)$, $m \geq 2$, be complex polynomials without common root, i.e. there does not exists $\lambda \in \mathbb{C}$, such that $f_i(\lambda) = 0$ for all $1 \leq i \leq m$. Then there exists $k_1(T), \ldots, k_m(T) \in \mathbb{C}[T]$, such that*

$$1 = k_1(T) \cdot f_1(T) + \, \ldots \, + k_m(T) \cdot f_m(T) \,.$$

Proof. Let $l(T) = \sum_{i=1}^{m} h_i(T) \cdot f_i(T)$ be as in Lemma B. If $\deg l(T) \geq 1$ then there exists a complex number λ, such that $l(\lambda) = 0$. But since $f_i(T) = g_i(T) \cdot l(T)$ this implies $f_i(\lambda) = g_i(\lambda) \cdot l(\lambda) = 0$ for all $1 \leq i \leq m$, and so the polynomials $f_i(T)$ would have a common root. Hence $l(t) = c \neq 0$ is constant, and so with $k_i(T) = c^{-1} \cdot h_i(T)$ we get $1 = \sum_{i=1}^{m} k_i(T) \cdot f_i(T)$. □