MATH 325 Q1: LINEAR ALGEBRA III, PART 2

The invariant subspace generated by a vector.

We start from the following result.

Lemma. *Let $\alpha : V \longrightarrow V$ be an F-linear map and $v \neq 0$ a vector in V. The subspace $W(\alpha, v)$ of V spanned by the set of vectors*

$$\{v, \alpha(v), \alpha^2(v), \alpha^3(v), \ \ldots\}$$

is α-invariant.

Proof. By construction, W is equal to the set of all finite linear combinations

$$\sum_{i=0}^{m} a_i \cdot \alpha^i(v)\,,\ a_i \in F\,,\ m \in \mathbb{N}\,,$$

where we have set $\alpha^0 = \mathrm{id}_V$. Since

$$\alpha\Big(\sum_{i=0}^{m} a_i \cdot \alpha^i(v)\Big) = \sum_{i=0}^{m} a_i \cdot \alpha^{i+1}(v)$$

the subspace $W(\alpha, v)$ is α-invariant. □

Remark. *Note that $W(\alpha, v)$ is the smallest α-invariant subspace of V which contains the vector v. Clearly, it is of dimension ≥ 1 since $v \neq 0$.*

Since $W(\alpha, v)$ is contained in V it has finite dimension. Hence for a big enough integer $l \geq 0$ the vectors

$$v = \alpha^0(v), \alpha(v), \ldots, \alpha^l(v)$$

are linearly dependent, i.e. there are scalars $a_0, \ldots, a_l \in F$, which are not all equal zero, such that

$$\sum_{i=0}^{l} a_i \cdot \alpha^i(v) = 0\,.$$

Lemma. *Let l be minimal with this property. Then $v, \alpha(v), \ldots, \alpha^{l-1}(v)$ is a basis of $W(\alpha, v)$.*

Proof. By the assumption on l this set of vectors is linear independent and so we have only to show that they generate (or span) the subspace $W(\alpha, v)$. For this it is enough to show that any vector of the generating system $\{\alpha^i(v) \mid i \in \mathbb{N}\}$ is a linear combination these vectors.

This is clear for $\alpha^i(v)$ if $0 \leq i \leq l-1$ and we show it for $\alpha^{(l-1)+i}(v)$ by induction on $i \geq 0$.

The case $i = 1$ is clear. Indeed, the minimality of l implies that the last coefficient a_l in the above equality

$$a_0 v + a_1 \alpha(v) + a_2 \alpha^2(v) + \cdots + a_l \alpha^l(v) = 0$$

is not equal to 0. Therefore,

$$\alpha^l(v) = -\frac{1}{a_l}(a_0 v + \alpha(v) + \cdots + a_{l-1}\alpha^{l-1}(v)) \in W(\alpha, v).$$

Abusing notation we will denote below $\frac{a_0}{a_l}$ by $a_0, \ldots,$ $\frac{a_{l-1}}{a_l}$ by a_{l-1}. Thus the above equality is of the form

$$\alpha^l(v) = -a_0 v - a_1 \alpha(v) + \cdots - a_{l-1}\alpha^{l-1}(v). \tag{1}$$

Next, by the induction assumption we may assume that the vectors

$$v, \alpha(v), \ldots, \alpha^{(l-1)+i-1}(v)$$

are in the span of $v, \alpha(v), \ldots, \alpha^{l-1}(v)$. Using (1) we then get

$$\alpha^{l-1+i}(v) = \sum_{j=0}^{l-1} -a_j \cdot \alpha^{i-1+j}(v)$$

is a linear combination of the vectors $v, \alpha(v), \alpha^2(v), \ldots, \alpha^{l-1}(v)$. $\square$

The linear map α defines by restriction an F-linear map from $W(\alpha, v)$ to itself. Using (1) we see that this linear map has with respect to the basis $v, \alpha(v), \ldots, \alpha^{l-1}(v)$ the following $l \times l$-matrix:

$$B = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & 0 & \ldots & 0 & -a_2 \\ & \vdots & \ddots & & & \vdots \\ & \vdots & & \ddots & & \vdots \\ 0 & & \ldots & & 1 & -a_{l-1} \end{pmatrix}. \tag{2}$$

The Cayley-Hamilton Theorem

The characteristic polynomial. Let

$$A = \begin{pmatrix} a_{11} & \dots & \dots & a_{1n} \\ a_{21} & \dots & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{n1} & \dots & \dots & a_{nn} \end{pmatrix}$$

be a $n \times n$-matrix over F.

Definition. *The characteristic polynomial $P_A(T)$ of the matrix A is defined as the polynomial*

$$\det(T\,\mathrm{I}_n - A) = \det \begin{pmatrix} T - a_{11} & \dots & \dots & -a_{1n} \\ -a_{21} & T - a_{22} & \dots & -a_{2n} \\ \vdots & & \ddots & \vdots \\ -a_{n1} & \dots & \dots & T - a_{nn} \end{pmatrix}$$

in the variable T, where I_n denotes the $n \times n$-identity matrix.

Recall that conjugate (similar) matrices have the same characteristic polynomial: If S is an invertible matrix then

$$\begin{aligned} P_{S\cdot A\cdot S^{-1}}(T) &= \det(T \cdot \mathrm{I}_n - S \cdot A \cdot S^{-1}) \\ &= \det(T \cdot (S \cdot S^{-1} - S \cdot A \cdot S^{-1}) \\ &= \det\left(S \cdot (T\,\mathrm{I}_n - A) \cdot S^{-1}\right) \\ &= \det(S) \cdot \det(T\,\mathrm{I}_n - A) \cdot \det(S^{-1}) = \det(T\,\mathrm{I}_n - A) \end{aligned}$$

(the last equation since $\det(S) \cdot \det(S^{-1}) = \det(S \cdot S^{-1}) = \det(\mathrm{I}_n) = 1$).

Using this we can define the *characteristic polynomial* of an F-linear map $\alpha : V \longrightarrow V$ of a finite dimensional F-vector space into itself. If $v_1, \dots, v_n$ is a basis of V and A the matrix of α with respect to this basis we set

$$P_\alpha(T) := P_A(T).$$

Lemma. *This definition does not depend on the choice of the basis.*

Proof. Let $v'_1, \dots, v'_n$ be another basis and A' the matrix of α with respect to the new basis $v'_1, \dots, v'_n$. Then, as we proved before, we have $A' = S^{-1} \cdot A \cdot S$, where $S = (s_{ij})$ is the base change matrix, *i.e.* $v'_i = \sum_{h=1}^{n} s_{hi} v_h$. Hence by the above remarks $P_A(T) = P_{A'}(T)$ and so $P_\alpha(T)$ does not depend on the choice of a basis of V. $\square$

Remark. *Let A be a $n \times n$-matrix and $\alpha_A : F^n \longrightarrow F^n$ be the F-linear map $v \mapsto A \cdot v$. Then A is the matrix of α_A with respect to the standard basis, and so $P_{\alpha_A}(T) = P_A(T)$, i.e. the characteristic polynomial of the linear map defined by a matrix A is equal to the characteristic polynomial of the matrix A.*

Example. Let $B \in \mathrm{M}_{l \times l}(F)$, $C \in \mathrm{M}_{l \times (n-l)}(F)$, and $D \in \mathrm{M}_{(n-l)\times(n-l)}(F)$. Then $A = \begin{pmatrix} B & C \\ \mathrm{O}_{(n-l)\times l} & D \end{pmatrix}$ is a $n \times n$-matrix, and we have

$$P_A(T) \;=\; \det\left(T \cdot \mathrm{I}_n - A\right) \;=\; \det \begin{pmatrix} T \cdot \mathrm{I}_l - B & C \\ \mathrm{O}_{(n-l)\times l} & T \cdot \mathrm{I}_{n-l} - D \end{pmatrix}.$$

It follows that

$$P_A(T) \;=\; P_B(T) \cdot P_D(T)\,.$$

The characteristic polynomial and the invariant subspace generated by a vector. Let $\alpha : V \longrightarrow V$ be a F-linear map and $v \neq 0$ a vector in V. Recall that there is an $l \geq 1$, such that

$$\{v, \alpha(v), \ldots, \alpha^{l-1}(v)\}$$

is a basis of the α-invariant subspace $W(\alpha, v)$ of V spanned by $\alpha^i v$ where $i = 0, 1, 2, \ldots$. Hence there are in particular $a_0, \ldots, a_{l-1} \in F$, such that

$$\alpha^l(v) + \sum_{i=0}^{l-1} a_i \cdot \alpha^i(v) \;= 0\,.$$

The restriction of α to $W(\alpha, v)$ has then the matrix

$$B \;=\; \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & 0 & \ldots & 0 & -a_2 \\ \vdots & \ddots & & & & \vdots \\ \vdots & & & \ddots & & \vdots \\ 0 & \ldots & & & 1 & -a_{l-1} \end{pmatrix}$$

with respect to this basis. Let us compute the characteristic polynomial $P_B(T)$. This is the determinant of

$$B = \begin{pmatrix} T & 0 & 0 & \dots & 0 & a_0 \\ -1 & T & 0 & \dots & 0 & a_1 \\ 0 & -1 & T & \dots & 0 & a_2 \\ & \vdots & \ddots & & & \vdots \\ & \vdots & & \ddots & & \vdots \\ 0 & & \dots & & -1 & T + a_{l-1} \end{pmatrix}.$$

We claim that

$$P_B(T) = T^l + a_{l-1}T^{l-1} + \dots + a_1 T + a_0 .$$

This is easily checked for $l = 1$ or 2 and follows by induction using expansion by minors along the first row.

We can extend the basis $v, \alpha(v), \dots, \alpha^{l-1}(v)$ of the subspace $W(\alpha, v) \subset V$ to a basis of V:

$$v, \alpha(v), \dots, \alpha^{l-1}(v), v_{l+1}, \dots, v_n .$$

With respect to this basis α is represented by the matrix

$$A = \begin{pmatrix} B & C \\ \mathrm{O}_{(n-l)\times l} & D \end{pmatrix},$$

where C is in $\mathrm{M}_{l\times(n-l)}(F)$ and $D \in \mathrm{M}_{(n-l)\times(n-l)}(F)$. Hence by Example above we have

$$P_\alpha(T) = P_A(T) = P_B(T) \cdot P_D(T) = \big(T^l + \sum_{i=0}^{l-1} a_i T^i\big) \cdot P_D(T) . \quad (3)$$

Matrix polynomials. Let $f(T) = \sum\limits_{i=0}^{n} a_i T^i$ be a polynomial in one variable T over the field F, and A an $n \times n$-matrix over F. The matrix $f(A)$ is then defined by evaluating the polynomial $f(T)$ at A, *i.e.*

$$f(A) := \sum_{i=0}^{n} a_i \cdot A^i = \mathrm{I}_n + a_1 \cdot A + a_2 \cdot A^2 + \dots + a_n \cdot A^n$$

(by A^0 is the $n \times n$-identity matrix I_n understood). We will need the following facts.

Lemma. *Let* $f(T) = \sum_{i=0}^{m} a_i T^i$ *and* $g(T) = \sum_{i=0}^{n} b_i T^i$ *be two polynomials and* A *a* $n \times n$*-matrix. Then:*

(i) *Matrices* $f(A)$ *and* $g(A)$ *commute with each other, i.e.*

$$f(A) \cdot g(A) \;=\; g(A) \cdot f(A)\,.$$

(ii) *If* $W \subset F^n$ *is* A*-invariant then it is also* $f(A)$*-invariant.*

Proof. Exercise. □

Similarly we can define the polynomial of a linear map $\alpha : V \longrightarrow V$. If $f(T) = \sum_{i=0}^{n} a_i T^i$ then

$$f(\alpha) \;=\; \sum_{i=0}^{n} a_i \alpha^i\,,$$

where $\alpha^0 = \mathrm{id}_V$ is understood. For instance, if $f(T) = 3T^3 + 10T^2 - 1$ then $f(\alpha)$ is the F-linear map

$$v \longmapsto 3 \cdot \alpha(\alpha(\alpha(v))) + 10 \cdot \alpha(\alpha(v)) - v\,.$$

We have then also the analog of the lemma above:

$$f(\alpha) \circ g(\alpha) \;=\; g(\alpha) \circ f(\alpha) \tag{4}$$

for all polynomials $f(T), g(T)$, and if $W \subseteq V$ is an α-invariant subspace then it is also an $f(\alpha)$-invariant subspace.

The Cayley-Hamilton Theorem.

If $n = 2$ and $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ then

$$P_A(T) \;=\; (T - a_{11}) \cdot (T - a_{22}) - a_{21} \cdot a_{12} \;=\; T^2 - (a_{11} + a_{22})T + (a_{11} \cdot a_{22} - a_{12} \cdot a_{21})$$

is a quadratic polynomial.

Exercise. Show that that $P_A(A) = 0$.

We now prove that this is true for all $n \times n$-matrices:

Theorem (Cayley-Hamilton). *Let* $\alpha : V \longrightarrow V$ *be a linear map. Then*

$$P_\alpha(\alpha) \;=\; 0\,.$$

In particular, we have $P_A(A) = 0$ *for all* $n \times n$*-matrices* A*.*

Proof. Note that $P_\alpha(\alpha)$ is a map from V to V and we need to show that this is zero map, i.e. if $v \in V$ then $P_\alpha(\alpha)(v) = 0$.

This is clear for $v = 0$, so let $v \neq 0$. Let $n = \dim V$ and $1 \leq l \leq n$ be the dimension of the α-invariant subspace $W(\alpha, v)$ generated by α. By (3) we have

$$P_\alpha(T) \;=\; \big(T^l + \sum_{i=0}^{l-1} a_i T^i\big) \cdot P_D(T)\,,$$

where the $a_i \in F$ are such that

$$\alpha^l(v) + \sum_{i=0}^{l-1} a_i \alpha^i(v) \;= 0\,,$$

and D is a matrix in $\mathrm{M}_{n-l}(F)$.

Hence $P_\alpha(\alpha) = \big(\alpha^l + \sum_{i=0}^{l-1} a_i \alpha^i\big) \cdot P_D(\alpha)$, and so by (4) we have

$$P_\alpha(\alpha) = P_D(\alpha) \cdot \big(\alpha^l + \sum_{i=0}^{l-1} a_i \alpha^i\big)\,.$$

Hence we compute

$$P_\alpha(\alpha)(v) \;= P_D(\alpha)\Big[\,\big(\alpha^l(v) + \sum_{i=0}^{l-1} a_i \alpha^i(v)\big)\Big] \;=\; P_D(\alpha)(0) \;=\; 0$$

as claimed. $\square$