ADVANCED LINEAR ALGEBRA

MATH 4378

These Notes are meant as a supplement of the required Text *Linear Algebra* by **K. Hoffman** and **R. Kunze.** The first seven section cover the geometric/algebraic theory of the structure of one linear map; section eight deals with the computational aspect of the theory. In section nine we present an elementary proof of the spectral theorem for symmetric maps on Euclidean spaces.

THE STRUCTURE OF ONE LINEAR MAP

1. The Minimal Polynomial. Let $\mathbb{V}$ be a vector space over the field $\mathbb{F}$ and let $T: \mathbb{V} \longrightarrow \mathbb{V}$ be linear. We assume that $\mathbb{V}$ is of finite dimension, $\dim(\mathbb{V})=n$. Let $\alpha \in \mathbb{V}$ where $\alpha \neq 0$. Then

$$\alpha,\ T(\alpha),\ T^2(\alpha),\ldots.,\ T^n(\alpha)$$

are n+1-many vectors in $\mathbb{V}$, so they must be linearly dependent. Hence, there is some k, $0 < k \leq n$, such that

$$\alpha,\ T(\alpha),\ldots.,\ T^{k-1}(\alpha) \text{ are linearly independent}$$

while $T^k(\alpha)$ is a unique linear combination of α, $T(\alpha)$, ..., $T^{k-1}(\alpha)$:

$$T^k(\alpha) = -a_0\alpha - a_1T(\alpha) - \ldots. - a_{k-1}T^{k-1}(\alpha)$$

That is: $\quad a_0\alpha + a_1T(\alpha) + \ldots + a_{k-1}T(\alpha) + T^k(\alpha) = 0.$

Let: $\quad p_\alpha(x) = a_0 + a_1x \ldots +a_{k-1}x^{k-1} + x^k$

Then: $\quad p_\alpha(T)(\alpha) = 0$

$p_\alpha(x)$ is called the *minimal polynomial* of T for the *vector* $\alpha \in \mathbb{V}$.

For the zero vector $\alpha = 0$ one has that $T^0(\alpha) = id_{\mathbb{V}}(\alpha) = 0$. Hence we may define for the zero vector, $p_0(x) = 1$. If $T = 0$ then $p_\alpha(x) = x$ for every $\alpha \neq 0$. If $T = id_V$ is the identity map on $\mathbb{V}$, then $p_\alpha(x) = 1-x$ for every $\alpha \neq 0$.

Assume one has for a polynomial $p(x) = b_0 + b_1x \ldots. +b_{m-1}x^{m-1} + x^m$, $m\geq 0$, that $p(T)(\alpha) = 0$. Such a polynomial is called a non-trivial *annihilating polynomial* for α. Then:

$$T^m(\alpha) = -b_0\alpha - b_1T(\alpha) - \ldots. - b_{m-1}T^{m-1}(\alpha)$$

Thus: $\quad \alpha,\ T(\alpha),\ \ldots.,\ T^m(\alpha)$ are linearly dependent

Because α, $T(\alpha)$, ..., $T^{k-1}(\alpha)$ are linearly independent we must have $m \geq k$ for any non-trivial annihilating polynomial.

If p(x) is any non-trivial annihilating polynomial then

$$\deg(p(x)) \geq \deg(p_\alpha(x)).$$

Assume that $p(x)$ is annihilating for α. We divide $p(x)$ by $p_\alpha(x)$ with remainder:

$$p(x) = q(x)p_\alpha(x) + r(x), \quad \deg(r(x)) < \deg(p_\alpha(x)) \text{ or } r(x) = 0$$

We then have

$$p(T) = q(T)\circ p_\alpha(T) + r(T), \text{ and, in particular,}$$

$$p(T)(\alpha) = q(T)(p_\alpha(T)(\alpha))+r(T)(\alpha) = 0$$

because $p(x)$ is annihilating for α. But also $p_\alpha(T)(\alpha) = 0$, so $r(T)(\alpha) = 0$. Hence, $r(x)$ is also annihilating for α. This is possible only if $r(x) = 0$. Hence: *Every annihilating polynomial for α is divisible by the minimal polynomial for α.* Recall the following fact about polynomials: If $p_1(x)$ and $p_2(x)$ divide each other then $p_1(x) \sim p_2(x)$, i.e., $p_1(x)$ and $p_2(x)$ are either both zero or $p_1(x) = cp_2(x)$ with a unique $c \in \mathbb{F}$ where $c \neq 0$. In particular, two monic polynomials $p_1(x)$ and $p_2(x)$ which divide each other are equal.

Proposition 1. The minimal polynomial of the vector α for the linear map T on the finite dimensional vector space $\mathbb{V}$ is the unique polynomial $p_\alpha(x)$ with the following properties.

(a) $p_\alpha(x)$ is annihilating, i.e., $p_\alpha(T)(\alpha) = 0$.

(b) If $p(x)$ is annihilating then $p_\alpha(x) | p(x)$.

(c) $p_\alpha(x)$ is monic, i.e., the highest coefficient of $p_\alpha(x)$ is one. □

Now let $\alpha_1,\ldots,\alpha_n$ be any base for $\mathbb{V}$ and let $p_{\alpha_1}(x),\ldots,p_{\alpha_n}(x)$ be the minimal polynomials for $\alpha_1,\ldots,\alpha_n$, respectively. Then define:

$$p_{\mathbb{V}}(x) = \text{l.c.m.}(p_{\alpha_1}(x),\ldots,p_{\alpha_n}(x))$$

That is, $p_{\alpha_\nu}(x) | p_{\mathbb{V}}(x)$ for $\nu = 1,\ldots,n$, and if $p_{\alpha_\nu}(x) | p(x)$ holds for $\nu = 1,\ldots,n$, then $p_{\mathbb{V}}(x) | p(x)$.

Proposition 2. The minimal polynomial $p_{\mathbb{V}}(x)$ for the linear map T on the finite dimensional vector space $\mathbb{V}$ is the unique polynomial with the following properties.

(a) $p_{\mathbb{V}}(x)$ annihilates $\mathbb{V}$, i.e., $p_{\mathbb{V}}(T)(\alpha) = 0$ for every $\alpha \in \mathbb{V}$.

(b) If $p(x)$ is annihilating $\mathbb{V}$ then $p_{\mathbb{V}}(x) | p(x)$.

(c) $p_{\mathbb{V}}(x)$ is monic.

Proof. Let $\alpha = a_1\alpha_1 + a_2\alpha_2 + \ldots + a_n\alpha_n$. Then:

$$p_{\mathbb{V}}(T)(\alpha) = a_1 p_{\mathbb{V}}(T)(\alpha_1) + \ldots + a_n p_{\mathbb{V}}(T)(\alpha_n)$$

But $p_{\mathbb{V}}(x) = q_\nu(x)p_{\alpha_\nu}(x)$ and therefore $p_{\mathbb{V}}(T)(\alpha_\nu) = 0$ for $\nu = 1,\ldots,n$. This proves (a). Now, if $p(T)(\alpha) = 0$ holds for every $\alpha \in \mathbb{V}$, then , in particular, $p(T)(\alpha_\nu) = 0$. Hence, $p_{\alpha_\nu}(x) \mid p(x)$ for $\nu = 1,\ldots,n$. But then $p_{\mathbb{V}}(x) \mid p(x)$ because $p_{\mathbb{V}}(x)$ is the lowest common multiple of the $p_{\alpha_\nu}(x)$. This proves (b).□

Thus $P_{\mathbb{V}}(x)$ is independent of the chosen basis $\alpha_1,\ldots,\alpha_n$. It is the unique annihilating polynomial which divides any other annihilating polynomial and which has highest coefficient one. Note:

If α *has minimal polynomial* $p_\alpha(x)$ *then* $p_\alpha(x) \mid p_{\mathbb{V}}(x)$.

Recall that the ring $\mathbb{F}[x]$ of polynomials over a field $\mathbb{F}$ is a *principal ideal domain.* That is, any ideal I of $\mathbb{F}[x]$ is principal, $I = \{q(x)p(x) \mid q(x) \in \mathbb{F}[x]\}$. The generator $p(x)$ is either zero or can be chosen as a unique monic polynomial in I.

If $\mathbb{V}$ is any vector space over $\mathbb{F}$ and T any linear map on $\mathbb{V}$, then for any subset S of $\mathbb{V}$

$$I = \{p(x) \mid p(T)(\alpha) = 0 \text{ for every } \alpha \in S\} = \mathrm{Ann}(S)$$

is an ideal of $\mathbb{F}[x]$. If $I \neq 0$, e.g., if $\mathbb{V}$ is of finite dimension, then the unique monic generator of Ann(S) is called the minimal polynomial $p_S(x)$ for S. If $\mathbb{V}$ is of infinite dimension an annihilator ideal might very well be zero, e.g., $\mathbb{V} = \mathbb{F}[x]$ and $T = D$, D being the differential operator.

Homework problems:

Page 198, Exercises 7, 8, 10

2. **The Primary Decomposition Theorem.** Let $\mathbb{U}_1$ and $\mathbb{U}_2$ be subspaces of the vector space $\mathbb{V}$. Then $\mathbb{V}$ is the *direct sum* of $\mathbb{U}_1$ and $\mathbb{U}_2$ if (a) $\mathbb{U}_1 \cap \mathbb{U}_2 = \mathbb{O}$ and (b) $\mathbb{U}_1 + \mathbb{U}_2 = \{\alpha_1 + \alpha_2 | \alpha_1 \in \mathbb{U}_1,\ \alpha_2 \in U_2\} = \mathbb{V}$ hold. It is easy to see that (a) and (b) together are equivalent with (c) For every vector $\alpha \in \mathbb{V}$ one has $\alpha = \alpha_1 + \alpha_2$ with unique $\alpha_i \in \mathbb{U}_i$, i=1,2. We say that $\mathbb{V}$ is the direct sum of subspaces $\mathbb{V}_1,\ldots,\mathbb{V}_n$ if every vector $\alpha \in \mathbb{V}$ is the unique sum of vectors $\alpha_i \in \mathbb{V}_i$. If this is the case then one writes $\mathbb{V} = \mathbb{V}_1 \oplus \ldots \oplus \mathbb{V}_n$.

If $T\colon \mathbb{V} \longrightarrow \mathbb{V}$ is linear then a subspace $\mathbb{U}$ of $\mathbb{V}$ is called *invariant* if

$$T(\alpha) \in \mathbb{U} \text{ holds for all } \alpha \in \mathbb{U}$$

Examples. 1. $\ker(T) = \{\alpha | T(\alpha) = 0\}$ and $\operatorname{im}(T) = \{\beta | \beta = T(\alpha) \text{ for some } \alpha \in \mathbb{V}\}$ are invariant subspaces of $\mathbb{V}$.

2. Let $p(x)$ be any polynomial and $T\colon \mathbb{V} \longrightarrow \mathbb{V}$ be any linear map. Then $\mathbb{U} = \{\alpha | p(T)(\alpha) = 0\} = \ker(p(T))$ is an invariant subspace: If $p(T)(\alpha) = 0$, then $p(T)(T(\alpha)) = (p(T)\circ T)(\alpha) = (T\circ p(T))(\alpha) = T(p(T)(\alpha)) = T(O) = 0$. Hence, if $\alpha \in U$ then $T(\alpha) \in \mathbb{U}$.

3. Let $\mathbb{U}$ be an invariant subspace of $\mathbb{V}$. Then T can be restricted to $\mathbb{U}$ and we may safely talk about the minimal polynomial $p_{\mathbb{U}}(x)$ of $\mathbb{U}$ for T.

4. The intersection of invariant subspaces is invariant. Hence there is for every subset S of $\mathbb{V}$ the invariant span ,$\langle S\rangle_T$ of S, i.e., the smallest invariant subspace which contains S.

Lemma. Let $T\colon \mathbb{V} \longrightarrow \mathbb{V}$ be a linear map on the finite dimensional vector space $\mathbb{V}$. Assume that

$$p_{\mathbb{V}}(x) = p_1(x)p_2(x)$$

where $p_1(x)$ and $p_2(x)$ are monic polynomials which are relatively prime. Then $\mathbb{V}$ is a direct sum of invariant subspaces whose minimal polynomials are $p_1(x)$ and $p_2(x)$, respectively:

$\mathbb{V} = \mathbb{V}_1 \oplus \mathbb{V}_2$; $\mathbb{V}_i$ are invariant, i.e., $T(\mathbb{V}_i) \subseteq \mathbb{V}_i$ and $p_{\mathbb{V}_i}(x) = p_i(x)$.

Proof. We define $\mathbb{V}_1 = \{\alpha | p_1(T)(\alpha) = 0\}$ and $\mathbb{V}_2 = \{\alpha | p_2(T)(\alpha) = 0\}$. We already know that $\mathbb{V}_1$ and $\mathbb{V}_2$ are invariant subspaces of $\mathbb{V}$. Because $p_1(x)$ and $p_2(x)$ are relatively prime one has

$$1 = q_1(x)p_1(x) + q_2(x)p_2(x)$$

Thus: $$\mathrm{id}_{\mathbb{V}} = I = q_1(T)\circ p_1(T) + q_2(T)\circ p_2(T)$$

Let $\alpha \in \mathbb{V}$. Then: $\alpha = (q_1(T)\circ p_1(T))(\alpha) + (q_2(T)\circ p_2(T))(\alpha)$

We claim : $\alpha_1 = (q_2(T)\circ p_2(T))(\alpha) \in \mathbb{V}_1$, $\alpha_2 = (q_1(T)\circ p_1(T))(\alpha) \in \mathbb{V}_2$

Notice: $p_1(T)(\alpha_1) = (p_1(T)\circ q_2(T)\circ p_2(T))(\alpha) = (q_2(T)\circ p_{\mathbb{V}}(T))(\alpha) = 0$

Hence $\alpha_1 \in V_1$, and a similar argument shows $\alpha_2 \in \mathbb{V}_2$.

Assume $\alpha \in \mathbb{V}_1 \cap \mathbb{V}_2$. Then

$$p_1(T)(\alpha) = 0,\ p_2(T)(\alpha) = 0$$

and

$$\mathrm{id}_{\mathbb{V}} = I = q_1(T)\circ p_1(T) + q_2(T)\circ p_2(T)$$

yields

$$\alpha = (q_1(T)\circ p_1(T))(\alpha) + (q_2(T)\circ p_2(T))(\alpha) = 0$$

Hence:

$$\mathbb{V} = \mathbb{V}_1 \oplus \mathbb{V}_2$$

By the very definition $p_1(x)$ is annihilating for $\mathbb{V}_1 = \{\alpha \mid p_1(T)(\alpha) = 0\}$ and the same holds for $\mathbb{V}_2$. Hence:

$$p_{\mathbb{V}_1}(x) \mid p_1(x) \text{ and } p_{\mathbb{V}_2}(x) \mid p_2(x)$$

We are going to show that $p(x) = p_{\mathbb{V}_1}(x)\cdot p_{\mathbb{V}_2}(x)$ is annihilating for $\mathbb{V}$. Let $\alpha \in \mathbb{V}$. Then $\alpha = \alpha_1 + \alpha_2$ where $\alpha_i \in \mathbb{V}_i$. But then

$$p(T)(\alpha) = p_{\mathbb{V}_2}(T)\circ p_{\mathbb{V}_1}(T)(\alpha_1) + p_{\mathbb{V}_1}(T)\cdot p_{\mathbb{V}_2}(T)(\alpha_2) = 0 + 0 = 0$$

This shows that

$$p_{\mathbb{V}}(x) \mid p_{\mathbb{V}_1}(x)\cdot p_{\mathbb{V}_2}(x)$$

But

$p_{\mathbb{V}_1}(x) \mid p_1(x)$ and $p_{\mathbb{V}_2}(x) \mid p_2(x) \Rightarrow p_{\mathbb{V}_1}(x)\cdot p_{\mathbb{V}_2}(x) \mid p(x) = p_1(x)\cdot p_2(x) = p_{\mathbb{V}}(x)$

Hence:

$$p_{\mathbb{V}_1}(x)\cdot p_{\mathbb{V}_2}(x) = p_{\mathbb{V}}(x)$$

We have $\deg(p_{\mathbb{V}_1}(x)) \leq \deg(p_1(x))$ and $\deg(p_{\mathbb{V}_2}(x)) \leq \deg(p_2(x))$, but also

$$\deg(p_{\mathbb{V}_1} p_{\mathbb{V}_2}) = \deg(p_{\mathbb{V}_1}) + \deg(p_{\mathbb{V}_2}) = \deg(p_{\mathbb{V}}) = \deg(p_1) + \deg(p_2)$$

Hence: $\deg(p_{\mathbb{V}_1}) = \deg(p_1)$ and $\deg(p_{\mathbb{V}_2}) = \deg(p_2)$

Therefore $p_{\mathbb{V}_1}(x) = p_1(x)$ and $p_{\mathbb{V}_2}(x) = p_2(x)$. □

Definition: Let $T: \mathbb{V} \longrightarrow \mathbb{V}$ be a linear map on the finite dimensional vector space $\mathbb{V}$. Let $p_{\mathbb{V}}(x) = p_1(x)^{r_1}\ldots\ldots p_k{}^{r_k}(x)$ be the prime factorization of the minimal polynomial for T. Then the invariant subspaces

$$\mathbb{V}_i = \{\alpha \mid p_i(T)^{r_i}(\alpha) = 0\} = \ker(p_i(T)^{r_i}(\alpha) \ ,\ i = 1,\ldots,k$$

are called the *primary components* of T.

If we put $q_1(x) = p_1(x)^{r_1}$ and $q_2(x) = p_2(x)^{r_1}.....p_k{}^{r_k}(x)$ then it follows from the lemma:

$$\mathbb{V} = \mathbb{V}_1 \oplus \mathbb{U}\ ,\ p_{\mathbb{V}_1}(x) = p_1(x)^{r_1} \text{ and } p_{\mathbb{U}}(x) = p_2(x)^{r_2}.....p_k{}^{r_k}(x)$$

Assume $p_2(T)^{r_2}(\alpha) = 0$. We wish to show that $\alpha \in \mathbb{U}$. Now $\alpha = \alpha_1 + \alpha_2$ where $\alpha_1 \in \mathbb{V}_1$ and $\alpha_2 \in \mathbb{U}$. Thus $p_2(T)^{r_2}(\alpha) = p_2(T)^{r_2}(\alpha_1) + p_2(T)^{r_2}(\alpha_2) = 0$. So $p_2(T)^{r_2}(\alpha_1) = 0$ and $p_2(T)^{r_2}(\alpha_2) = 0$. Because the minimal polynomial of $\mathbb{V}_1$ is $p_1(x)^{r_1}$, the minimal polynomial of α_1 is of the form $p_1(x)^s$ where $0 \leq s \leq r_1$. Now $p_1(x)^s | p_2(T)^{r_2}(\alpha_1)$ because of $p_2(T)^{r_2}(\alpha_1) = 0$. But this is only possible in case that $s = 0$, i.e., $\alpha_1 = 0$. Hence

$$\mathbb{V}_2 = \{\alpha | p_2(T)^{r_2}(\alpha) = 0\} = \{\alpha | \alpha \in \mathbb{U},\ p_2(T)^{r_2}(\alpha) = 0\}$$

and

$$p_{\mathbb{V}_2}(x) = p_2(x)^{r_2}$$

This proves

Theorem 3 (Primary Decomposition Theorem) Let T: $\mathbb{V} \longrightarrow \mathbb{V}$ be a linear map on the finite dimensional vector space $\mathbb{V}$. Then $\mathbb{V}$ is the direct sum of its primary components $\mathbb{V}_i = \{\alpha | p_i(T)^{r_i}(\alpha) = 0\}$ where $p_{\mathbb{V}}(x) = p_1(x)^{r_1}.....p_k{}^{r_k}(x)$ is the prime factorization of the minimal polynomial of $\mathbb{V}$ for T. Moreover, $p_i(x)^{r_i}$ is the minimal polynomial for $\mathbb{V}_i$. □

Homework problems:

Page 213, Exercise 9

Page 219, Exercise 4

Page 225, Exercises 3, 10, 15

3. Quotient Spaces. Let $\mathbb{V}$ be a vector space over the field $\mathbb{F}$ and let $\mathbb{U}$ be a subspace of $\mathbb{V}$. Then

$$\alpha_1 \equiv_{\mathbb{U}} \alpha_2 \quad \text{iff} \quad \alpha_1 - \alpha_2 \in \mathbb{U}$$

is an equivalence relation on $\mathbb{V}$. It is also easily proved that $\equiv_{\mathbb{U}}$ is a *congruence* on the vector space $\mathbb{V}$. That is,

$$\alpha_1 \equiv_{\mathbb{U}} \alpha_2 \text{ and } \beta_1 \equiv_{\mathbb{U}} \beta_2 \quad \text{implies that } \alpha_1 + \beta_1 \equiv \alpha_2 + \beta_2$$

$$\alpha_1 \equiv_{\mathbb{U}} \alpha_2 \quad \text{implies that } c.\alpha_1 \equiv c.\alpha_2 \text{ holds for every } c \in \mathbb{F}$$

We may define on the set $\mathbb{V}/\mathbb{U}$ of all equivalence classes a vector space structure according to the following rules:

$$[\alpha] + [\beta] = [\alpha + \beta] \quad \text{and} \quad c.[\alpha] = [c.\alpha]$$

The map $$q_{\mathbb{U}}: \mathbb{V} \twoheadrightarrow \mathbb{V}/\mathbb{U} \ , \ \alpha \longmapsto [\alpha]$$

is obviously linear and

$$\ker(q_{\mathbb{U}}) = \mathbb{U}$$

Hence, by the equation: $\text{rank}(q_{\mathbb{U}}) + \text{nullity}(q_{\mathbb{U}}) = \dim(\mathbb{V})$ we see that

$$\dim(\mathbb{V}/\mathbb{U}) + \dim(\mathbb{U}) = \dim(\mathbb{V})$$

The proof of this equation is based on the fact that

If $\alpha_1, \dots , \alpha_d$ *is a basis of* $\mathbb{U}$ *and* $[\alpha_{d+1}]. \dots ,[\alpha_k]$ *a basis of* $\mathbb{V}/\mathbb{U}$ *then* $\alpha_1, \dots , \alpha_d, \alpha_{d+1}, \dots , \alpha_k$ *is a basis of* $\mathbb{V}$.

Lemma: $[\alpha_1], \dots ,[\alpha_m]$ are linearly independent in $\mathbb{V}/\mathbb{U}$ iff

$$a_1\alpha_1 + \dots + a_m\alpha_m \in \mathbb{U} \quad \text{iff} \quad a_1 = \dots = a_m = 0$$

Proof. This is obvious because of

$$a_1\alpha_1 + \dots + a_m\alpha_m \in \mathbb{U} \text{ iff } [a_1\alpha_1 + \dots + a_m\alpha_m] = \mathbb{U} \quad \text{iff}$$
$$a_1[\alpha_1] + \dots + a_m[\alpha_m] = [0] \ . \ \square$$

Now assume that $T: \mathbb{V} \longrightarrow \mathbb{V}$ is linear and $\mathbb{U}$ is invariant under T. Then T induces a linear map between the factor spaces:

$$\overline{T}: \mathbb{V}/\mathbb{U} \longrightarrow \mathbb{V}/\mathbb{U} \ , \ [\alpha] \longmapsto [T(\alpha)]$$

The map $\overline{T}$ is well defined: If $\alpha_1 \equiv \alpha_2$ then $T(\alpha_1) \equiv T(\alpha_2)$. Here we use that $\mathbb{U}$ is invariant.

Lemma: The minimal polynomial $p_{[\alpha]}(x)$ of $[\alpha]$ for the linear map $\overline{T}$ on $\mathbb{V}/\mathbb{U}$ divides the minimal polynomial $p_\alpha(x)$ of α for the linear map T on $\mathbb{V}$.

Proof. Let $p_\alpha(x) = a_0 + a_1x + \dots + x^k$. Then:

$$p_\alpha(\overline{T})([\alpha]) = (a_0 + a_1\overline{T} + \dots + (\overline{T})^k)([\alpha]) = [a_0\alpha + a_1T(\alpha) + \dots + T^k(\alpha)] = \mathbb{U}$$

Thus $p_\alpha(x)$ is an annihilating polynomial of $[\alpha]$. This is

$$p_{[\alpha]}(x) \mid p_\alpha(x) \text{ . } \square$$

Similarly:
$$p_{\mathbb{V}/\mathbb{U}}(x) \mid p_{\mathbb{V}}(x) \text{ .}$$

Let $S: \mathbb{U} \twoheadrightarrow \mathbb{V}$ be a surjective linear map between vector spaces. Then one has the *homomorphism theorem for linear maps:*

$$\mathbb{U}/\ker(S) \cong \mathbb{V} \text{ , under the map } \dot{S} : [\alpha] \longmapsto S(\alpha)$$

If we are considering vector spaces with designated linear maps, say $(\mathbb{V}_1,T_1)$ and $(\mathbb{V}_2,T_2)$, then a linear map is *admissible* if it also obeys T. That is,

$$S(T_1(\alpha)) = T_2(S(\alpha)) \text{ holds for all } \alpha \in \mathbb{V}_1$$

Exercises: Show that for an admissible map S, ker(S) is T_1-invariant in $\mathbb{V}_1$ and $\mathrm{im}(S_1)$ is T_2-invariant in $\mathbb{V}_2$. Formulate and prove the generalization of the homomorphism theorem for admissible maps between vector spaces with designated linear maps.

4. Cyclic Subspaces. An invariant subspace $\mathbb{U}$ of $\mathbb{V}$ is called *cyclic* if there is some vector α such that $\mathbb{U}$ is the smallest T-invariant subspace which contains $\mathbb{U}$:

$$\mathbb{U} = \langle\{\alpha\}\rangle_T$$

If $p_\alpha(x) = a_0 + a_1x + \dots + x^k$ is the minimal polynomial of α, then

$\alpha_0 = \alpha$, $\alpha_1 = T(\alpha)$, ..., $\alpha_{k-1} = T^{k-1}(\alpha)$ are linearly independent and

$$T^k(\alpha) = -a_0\alpha - \dots - a_{k-1}\alpha_{k-1}$$

Hence, if $\beta = c_0\alpha_0 + \dots + c_{k-1}\alpha_{k-1}$ then $T(\beta) = c_0\alpha_1 + \dots + c_{k-1}\alpha_k$. It follows that $\langle\alpha, T(\alpha), \dots, T^{k-1}(\alpha)\rangle$ is invariant and therefore,

$$\mathbb{U} = \langle\{\alpha\}\rangle_T = \langle\alpha, T(\alpha), \dots, T^{k-1}(\alpha)\rangle$$

Theorem 4. If $\mathbb{U} = \langle\{\alpha\}\rangle_T$ is a cyclic subspace then $\dim(\mathbb{V}) = k$, where k is the degree of the minimal polynomial of α for T. □

Let $\beta \in \mathbb{U}$. Then $\beta = c_0\alpha_0 + \dots + c_{k-1}\alpha_{k-1}$ with unique $c_i \in \mathbb{F}$. If we put $r(x) = c_0 + c_1x + \dots + c_{k-1}x^{k-1}$ then we see that there is for every β a unique $r(x)$ such that

$$r(T)(\alpha) = \beta \text{ , where } r(x) = 0 \text{ or } \deg(r(x)) < k$$

If $p(x) = q(x)p_\alpha(x) + r(x)$, then $p(T)(\alpha) = r(T)(\alpha)$.

The polynomial ring $\mathbb{F}[x]$ is a vector space over $\mathbb{F}$ and multiplication by x may be considered as a linear map: $f(x) \longmapsto x \cdot f(x)$. Let I be a set of polynomials which is a subspace of $\mathbb{F}[x]$. It is quite obvious that I is invariant if and only if I is an ideal of $\mathbb{F}[x]$.

Theorem 5. Let $\mathbb{U} = \langle\{\alpha\}\rangle$ be cyclic. Then the surjection

$$\varepsilon_T\colon \mathbb{F}[x] \twoheadrightarrow \mathbb{U} \text{ , } p(x) \longmapsto p(T)(\alpha)$$

is a an admissible linear map where

$$\ker(\varepsilon_T) = (p_\alpha(x))$$

is the principal ideal generated by the minimal polynomial $\mathbb{V}$ under T. Hence:

$$\mathbb{F}[x]/(p_{\mathbb{U}}(x)) \cong \mathbb{U}$$

as vector spaces with designated linear maps. □

The last theorem tells us that *cyclic spaces are isomorphic as spaces with designated maps if and only if they have the same minimal polynomial.*

Exercise. Let $p(x) \in \mathbb{F}[x]$. Then the x-invariant vector space $\mathbb{F}[x]/(p(x))$ is cyclic and its minimal polynomial is $p(x)$. Hence, for *any polynomial $p(x)$ there is a cyclic vector space $\mathbb{U}$ with designated map T such that $p_{\mathbb{U}} = p(x)$.*

5. Cyclic Decomposition. Let T be a linear map on the finite dimensional vector space $\mathbb{V}$. An invariant subspace $\mathbb{U}$ of $\mathbb{V}$ is *directly indecomposable* if $\mathbb{U}$ is not the direct sum of non-zero invariant subspaces. That is, if $\mathbb{U} = \mathbb{U}_1 \oplus \mathbb{U}_2$, then either $\mathbb{U}_1 = \mathbb{O}$, or $\mathbb{U}_2 = \mathbb{O}$. Of course, the zero space $\mathbb{O}$ is directly indecomposable. The following is a rather trivial observation.

Theorem 6. Let T be a linear map on the finite dimensional vector space $\mathbb{V}$. Then $\mathbb{V}$ is a direct sum of directly indecomposable subspaces.

Proof. $\mathbb{O}$ is the empty sum of directly indecomposable spaces. Now, given $(\mathbb{V},T)$ then either $\mathbb{V}$ is directly indecomposable, and we are done, or $\mathbb{V} = \mathbb{V}_1 \oplus \mathbb{V}_2$ where the $\mathbb{V}_i$ are invariant under T and different from the zero space. Hence, $\dim(\mathbb{V}_i) < \dim(\mathbb{V})$. The claim follows now by induction on $\dim(\mathbb{V})$. □

We are now facing the important problem to characterize the directly indecomposable subspaces of $\mathbb{V}$. We need some preliminary lemmas.

Lemma 1: Assume that the minimal polynomials $p_1(x)$ and $p_2(x)$ of α_1 and α_2, respectively, are relatively prime. Then the minimal polynomial of $\alpha = \alpha_1 + \alpha_2$ is $p(x) = p_1(x)\cdot p_2(x)$.

Proof. We have that $p(T)(\alpha_1 + \alpha_2) = p_2(T)p_1(T)(\alpha_1) + p_1(T)p_2(T)(\alpha_2) = 0$. Hence, $p(x)$ is annihilating α. This is $p_\alpha(x) \mid p(x)$. Now, let $q(x)$ be any annihilating polynomial for α. Then $p_2(T)q(T)(\alpha_1) = p_2(T)q(T)(\alpha - \alpha_2) = 0$. Hence, $p_1(x) \mid p_2(x)q(x)$, and, by the same token, $p_2(x) \mid p_1(x)q(x)$. Because $p_1(x)$ and $p_2(x)$ are relatively prime, one has that $p_1(x) \mid q(x)$ and that $p_2(x) \mid q(x)$. Again, because the $p_i(x)$ are relatively prime, one has that $p_1(x)\cdot p_2(x) \mid q(x)$. □

Lemma 2: Assume that the minimal polynomial for $\mathbb{V}$ is a power of an irreducible polynomial. Then there is a vector α such that $p_\alpha(x) = p_{\mathbb{V}}(x)$.

Proof. Let $\alpha_1, \dots, \alpha_n$ be any basis of $\mathbb{V}$. Then $p_{\alpha_i}(x) \mid p_{\mathbb{V}}(x)$ where $p_{\mathbb{V}}(x) = p(x)^r$ and where $p(x)$ is irreducible. Hence, each $p_{\alpha_i}(x)$ is a power of $p(x)$. That is, $p_{\alpha_i}(x) = p(x)^{r_i}$ and $p_{\mathbb{V}}(x)$ is as the lowest common multiple of the $p_{\alpha_i}(x)$ equal to some $p_{\alpha_k}(x)^{r_k}$ where $r = r_k$. □

Theorem 7. For any linear map $T: \mathbb{V} \longrightarrow \mathbb{V}$ on the finite dimensional vector space $\mathbb{V}$ there is some $\alpha \in \mathbb{V}$ such that $p_\alpha(x) = p_{\mathbb{V}}(x)$.

Proof. We first apply Lemma 1 to each primary component and then Lemma 2. □

Corollary: $\deg(p_{\mathbb{V}}(x)) \leq \dim(\mathbb{V})$. □

Lemma 3: Let $T: \mathbb{V} \longrightarrow \mathbb{V}$ be a linear map on the finite dimensional vector space $\mathbb{V}$. Let α_1 be any vector in $\mathbb{V}$ such that $p_{\alpha_1}(x) = p_{\mathbb{V}}(x)$. Assume that we have further vectors $\alpha_2, \ldots, \alpha_k$ such that $p_{\alpha_i} = p_{\mathbb{V}/\mathbb{U}_{i-1}}$ where $\mathbb{U}_0 = \mathbb{O}$ and where $\mathbb{U}_i = \langle \alpha_1 \rangle_T \oplus \ldots \oplus \langle \alpha_i \rangle_T$ for $0 < i \leq k$. Then, in case that $\mathbb{U}_k \subset \mathbb{V}$, one can find a vector α_{k+1} such that $p_{\alpha_{k+1}} = p_{\mathbb{V}/\mathbb{U}_k}$ and where the sum $\mathbb{U}_{k+1} = \mathbb{U}_k + \langle \alpha_{k+1} \rangle_T$ is direct.

Proof. We assume $\mathbb{U}_k \subset \mathbb{V}$, i.e., $\mathbb{V}/\mathbb{U}_k \neq \mathbb{O}$. In order to simplify notation we set $p_i = p_{\alpha_i}$, $i = 1, \ldots, k$ and $p_{k+1} = p_{\mathbb{V}/\mathbb{U}_k}$. The map T induces the various factor maps $\mathbb{V}/\mathbb{U}_i \longrightarrow \mathbb{V}/\mathbb{U}_i$, $[\alpha]_{\mathbb{U}_i} \longmapsto [T(\alpha)]_{\mathbb{U}_i}$. All these maps are denoted as $\overline{T}$.

Because $p_1(T)(\alpha) = 0$, we have that $p_1(T)(\alpha) \in \mathbb{U}_1$ is true for every α. Hence, $p_1(x)$ is an annihilating polynomial for $\overline{T}$ where $\overline{T}$ is the induced linear map on $\mathbb{V}/\mathbb{U}_1$, $[\alpha] \longmapsto [T(\alpha)]$. Because $p_2(x)$ is the minimal polynomial for $\mathbb{V}/\mathbb{U}_1$, we conclude $p_2(x) \mid p_1(x)$. Similarly, because $p_2(\overline{T})([\alpha]) = \mathbb{U}_1$ holds for every class $[\alpha] \in \mathbb{V}/\mathbb{U}_1$, we have that $p_2(T)(\alpha) \in \mathbb{U}_1 \subseteq \mathbb{U}_2$. Hence, $p_2(x)$ is an annihilating polynomial for $\overline{T}$, where $\overline{T}$ is the induced map $\mathbb{V}/\mathbb{U}_2 \longmapsto \mathbb{V}/\mathbb{U}_2$. Because $p_3(x)$ is the minimal polynomial for $\overline{T}$ on $\mathbb{V}/\mathbb{U}_2$, we conclude $p_3(x) \mid p_2(x)$. Finally, because $p_k(x)$ is the minimal polynomial for $\overline{T}$ on $\mathbb{V}/\mathbb{U}_{k-1}$ we have that $p_k(\overline{T})[\alpha] = \mathbb{U}_{k-1}$ holds for every class $[\alpha] \in \mathbb{V}/\mathbb{U}_{k-1}$. Hence, $p_k(T)(\alpha) \in \mathbb{U}_{k-1} \subseteq \mathbb{U}_k$. Hence, $p_k(x)$ is an annihilating polynomial for $\overline{T}$ on $\mathbb{V}/\mathbb{U}_k$. Because $p_{k+1}(x)$ is the minimal polynomial for $\overline{T}$ on $\mathbb{V}/\mathbb{U}_k$, we conclude $p_{k+1}(x) \mid p_k(x)$. Hence, we have a divisor chain of polynomials:

$$p_{k+1}(x) \mid p_k(x) \ldots \mid p_2(x) \mid p_1(x)$$

In order to find some α_{k+1} such that $p_{\alpha_{k+1}}(x) = p_{k+1}(x)$, we first pick any class $[\beta]$ in $\mathbb{V}/\mathbb{U}_k$ such that

$$p_{[\beta]}(x) = p_{k+1}(x)$$

We have for any $\alpha \in [\beta]$ that $p_{k+1}(x) | p_\alpha(x)$. Indeed,

$$p_\alpha(\overline{T})([\beta]) = p_\alpha(\overline{T})([\alpha]) = [p_\alpha(T)(\alpha)] = [0].$$

Hence, $p_{[\beta]}(x) | p_\alpha(x)$. But $p_{[\beta]}(x) = p_{k+1}(x)$. We are going to show that we can find in $[\beta]$ some α such that $p_\alpha(x) = p_{k+1}(x)$. Any such α will serve as an α_{k+1} for the lemma.

Because $p_{k+1}(x)$ divides all other polynomials $p_i(x)$ we have $q_i(x)$ such that

$$p_1(x) = q_1(x) \cdot p_{k+1}(x),\ p_2(x) = q_2(x) \cdot p_{k+1}(x),\ \ldots,\ p_k(x) = q_k(x) \cdot p_{k+1}(x)$$

We also note that

$$p_{k+1}(T)(\beta) \in \mathbb{U}_k = \langle \alpha_1 \rangle_T \oplus \ldots \oplus \langle \alpha_k \rangle_T$$

Hence: $$p_{k+1}(T)(\beta) = g_1(T)\alpha_1 + \ldots + g_k(T)\alpha_k$$

for certain polynomials $g_i(x)$. We are going to show that all these polynomials are divisible by $p_{k+1}(x)$:

$$q_1(T) \circ p_{k+1}(T)(\beta) = p_1(T)(\beta) = 0 = q_1(T) \circ g_1(T)\alpha_1 + \ldots + q_1(T) \circ g_k(T)\alpha_k$$

Because $\mathbb{U}_k$ is a direct sum of the $\langle \alpha_i \rangle_T$, we must have that

$$q_1(T) \circ g_1(T)\alpha_1 = 0,\ \ldots.,\ q_1(T) \circ g_k(T)\alpha_k = 0$$

Because $p_1(x)$ is the minimal polynomial of α_1,

$p_1(x) | q_1(x)g_1(x)$, i.e., $q_1(x)p_{k+1}(x) | q_1(x)g_1(x)$ or $p_{k+1}(x) | g_1(x)$. Hence:

$$g_1(x) = p_{k+1}(x) \cdot h_1(x)$$

Similarly,

$$q_2(T) \circ p_{k+1}(T)(\beta) = p_2(T)(\beta) = q_2(T)g_1(T)\alpha_1 + \ldots + q_2(T)g_k(T)\alpha_k \in \mathbb{U}_1 = \langle \alpha_1 \rangle$$

Again, because $\mathbb{U}_k$ is a direct sum of the $\langle \alpha_i \rangle_T$, we must have that

$$q_2(T) \circ g_2(T)(\alpha_2) = 0,\ \ldots\ ,\ q_2(T) \circ g_k(T)(\alpha_k) = 0$$

Because $p_2(x)$ is the minimal polynomial of α_2,

$p_2(x) | q_2(x)g_2(x)$, i.e., $q_2(x)p_{k+1}(x) | q_2(x)g_2(x)$ or $p_{k+1}(x) | g_2(x)$. Hence:

$$g_2(x) = p_{k+1}(x) \cdot h_2(x)$$

Finally,

$$q_k(T)\circ p_{k+1}(T)(\beta) = p_k(T)(\beta) = q_k(T)g_1(T)\alpha_1 + \dots + q_k(T)g_k(T)\alpha_k \in \mathbb{U}_{k-1}$$

Because $\mathbb{U}_k$ is a direct sum of the $\langle\alpha_i\rangle_T$, we must have that

$$q_k(T)\circ g_k(T)(\alpha_k) = 0$$

Because $p_k(x)$ is the minimal polynomial of α_k,

$p_k(x) \mid q_k(x)g_k(x)$, i.e., $q_k(x)p_{k+1}(x) \mid q_k(x)g_k(x)$ or $p_{k+1}(x) \mid g_k(x)$. Hence:

$$g_k(x) = p_{k+1}(x)\cdot h_k(x)$$

We now define:

$$\alpha_{k+1} = \beta - h_1(T)\alpha_1 - \dots - h_k(T)\alpha_k$$

We have $\alpha_{k+1} - \beta \in \mathbb{U}_k$, hence $[\alpha_{k+1}] = [\beta] \bmod(\mathbb{U}_k)$. Now,

$$p_{k+1}(T)(\alpha_{k+1}) = p_{k+1}(T)(\beta) - p_{k+1}(T)h_1(T)\alpha_1 - \dots - p_{k+1}(T)h_k(T)\alpha_k =$$
$$p_{k+1}(T)(\beta) - g_1(T)\alpha_1 - \dots - g_k(T)\alpha_k = 0$$

Hence, $p_{\alpha_{k+1}}(x) \mid p_{k+1}(x)$. But $p_{k+1}(x) \mid p_{\alpha_{k+1}}(x)$ as seen before. This is $p_{k+1}(x) = p_{\alpha_{k+1}}(x)$

Finally, we wish to show that the sum of $\mathbb{U}_k$ and $\langle\alpha_{k+1}\rangle_T$ is direct. Assume that $g(T)(\alpha_{k+1}) \in \mathbb{U}_k$. This is the same as saying that $g(\overline{T})[\alpha_{k+1}] = \mathbb{U}_k$ holds in $\mathbb{V}/\mathbb{U}_k$. But $[\alpha_{k+1}] = [\beta]$. Therefore, $p_{[\alpha_{k+1}]} = p_{[\beta]} = p_{k+1}$. Thus $p_{k+1}(x)$, which is the minimal polynomial of $[\alpha_{k+1}]$, must divide the annihilating polynomial $g(x)$. Hence, $g(x) = h(x)p_{k+1}(x)$. Thus, $g(T)(\alpha_{k+1}) = 0$. □

Theorem 8 (Cyclic Decomposition Theorem). Let $T\colon \mathbb{V} \longrightarrow \mathbb{V}$ be a linear map on the finite dimensional vector space $\mathbb{V}$. Then $\mathbb{V}$ is a direct sum of non-zero cyclic subspaces:

$$\mathbb{V} = \mathbb{V}_1 \oplus \mathbb{V}_2 \oplus \dots \oplus \mathbb{V}_r, \quad \mathbb{V}_i = \langle\alpha_i\rangle_T$$

such that $p_{\mathbb{V}_1}(x) = p_{\mathbb{V}}(x)$ and $p_{\mathbb{V}_i}(x) \mid p_{\mathbb{V}_{i-1}}(x)$ for $i = 1, 2, \dots, r$. □

Corollary 1. An invariant subspace $\mathbb{U}$ of $\mathbb{V}$ is cyclic if and only if

$$\deg(p_{\mathbb{U}}(x)) = \dim(\mathbb{U})$$

Proof. In Theorem 8 we have $\dim(\mathbb{V}) = \deg(p_{\mathbb{V}})$ if and only if $\mathbb{V} = \mathbb{V}_1$, hence, if and only if $\mathbb{V}$ is cyclic. □

Lemma 4. (a) Assume that $\mathbb{V}_1 = \langle\alpha_1\rangle_T$ and $\mathbb{V}_2 = \langle\alpha_2\rangle_T$ are cyclic and that $p_1(x) = p_{\alpha_1}(x)$ and $p_2(x) = p_{\alpha_2}(x)$ are relatively prime. Then $\mathbb{V} = \mathbb{V}_1 + \mathbb{V}_2$ is cyclic, the sum is direct and $\mathbb{V} = \mathbb{V}_1 \oplus \mathbb{V}_2 = \langle\alpha_1 + \alpha_2\rangle_T$.

(b) Assume that $\mathbb{V} = \mathbb{V}_1 \oplus \mathbb{V}_2$ where $\mathbb{V}$ is cyclic. Then $p_{\mathbb{V}_1}(x)$ and $p_{\mathbb{V}_2}(x)$ are relatively prime.

Proof. (a) If $\alpha \in \langle\alpha_1\rangle \cap \langle\alpha_2\rangle$, then $p_{\alpha_1}(T)(\alpha) = p_{\alpha_2}(T)(\alpha) = 0$ and, therefore, $p_\alpha | p_{\alpha_1}$ and $p_\alpha | p_{\alpha_2}$. Because $p_{\alpha_1}(x)$ and $p_{\alpha_2}(x)$ are relatively prime, this is only possible in case that $p_\alpha(x) = 1$. Hence, $\alpha = 0$ and the sum is direct. By Lemma 1, the minimal polynomial of $\alpha = \alpha_1 + \alpha_2$ is $p_\alpha(x) = p_1(x) \cdot p_2(x)$. Hence:

$\dim(\langle\alpha\rangle_T) = \deg(p_\alpha) = \deg(p_1) + \deg(p_2) = \dim(\mathbb{V}_1) + \dim(\mathbb{V}_2) =$

$\dim(\mathbb{V}_1 + \mathbb{V}_2) - \dim(\mathbb{V}_1 \cap \mathbb{V}_2) = \dim(\mathbb{V}_1 + \mathbb{V}_2)$

Therefore, $\langle\alpha\rangle_T = \mathbb{V}_1 + \mathbb{V}_2$.

(b) Let $\deg(p_{\mathbb{V}}(x)) = m$, $\deg(p_{\mathbb{V}_1}(x)) = m_1$, $\deg(p_{\mathbb{V}_2}(x)) = m_2$ and

$\dim(\mathbb{V}) = n$, $\dim(\mathbb{V}_1) = n_1$, $\dim(\mathbb{V}_2) = n_2$

Then, $n = n_1 + n_2$, and $m \le m_1 + m_2$ because $\mathbb{V} = \mathbb{V}_1 \oplus \mathbb{V}_2$ and $p_{\mathbb{V}_1}(x) \cdot p_{\mathbb{V}_2}(x)$ is annihilating $\mathbb{V}$. Because $\mathbb{V}$ is cyclic, $m = n$. Also, the degree of the minimal polynomial is $\le$ than the dimension, hence $m_1 \le n_1$, $m_2 \le n_2$. Therefore:

$$m \le m_1 + m_2 \le n_1 + n_2 = n$$

We conclude $m = m_1 + m_2$, i.e.,

$p_{\mathbb{V}}(x) = p_{\mathbb{V}_1}(x) \cdot p_{\mathbb{V}_2}(x) = \text{l.c.m.}(p_{\mathbb{V}_1}(x), p_{\mathbb{V}_2}(x))$, i.e., $p_{\mathbb{V}_1}(x)$ and $p_{\mathbb{V}_2}(x)$ are relatively prime. □

Corollary 2. $\mathbb{V}$ is directly indecomposable if and only if
(a) $\mathbb{V}$ is cyclic. (b) $p_{\mathbb{V}}(x) = (p(x))^r$ where $p_{\mathbb{V}}(x)$ is irreducible.

Proof. Assume that (a) and (b) hold. Then, if $\mathbb{V} = \mathbb{V}_1 \oplus \mathbb{V}_2$, $p_{\mathbb{V}_1}(x)$ and $p_{\mathbb{V}_2}(x)$ must divide $p_{\mathbb{V}}(x)$ because $p_{\mathbb{V}}(x)$ certainly annihilates $\mathbb{V}_1$ and $\mathbb{V}_2$. Hence, $p_{\mathbb{V}_1}(x)$ and $p_{\mathbb{V}_2}(x)$ are powers of $p(x)$ and, according to the last lemma, also relatively prime. This is only possible if one has that $r_i = 0$, i.e., $\mathbb{V}_i = \mathbb{O}$, for i = 1 or i = 2.

On the other hand, if $\mathbb{V}$ is indecomposable then, because of the cyclic decomposition theorem, $\mathbb{V}$ has to be cyclic. The minimal polynomial of $\mathbb{V}$ has to be a prime power because of the primary decomposition theorem. □

Theorem 9. Let $T: \mathbb{V} \longrightarrow \mathbb{V}$ b a linear map on the finite dimensional vector space $\mathbb{V}$. Then $\mathbb{V}$ decomposes into a direct sum of cyclic spaces $\mathbb{E}_i$ where the minimal polynomials of the $\mathbb{E}_i$'s are powers of irreducible polynomials. The spaces $\mathbb{E}_i$ cannot be decomposed any further into direct sums of invariant subspaces.

Proof. This is an immediate consequence of Theorem 6 and Corollary 2. □

In order to obtain this decomposition, we may start with the decomposition into the primary components:

$\mathbb{V} = \mathbb{V}_1 \oplus \dots \oplus \mathbb{V}_k$ where $p_{\mathbb{V}}(x) = p_1(x)^{r_1} \cdot \dots \cdot p_k(x)^{r_k}$, $\mathbb{V}_i = \{\alpha \mid p_i(T)^{r_i}(\alpha)=0\}$
and then apply the cyclic decomposition theorem to each $\mathbb{V}_i$:

$\mathbb{V}_i = \mathbb{E}_{i1} \oplus \dots \oplus \mathbb{E}_{is_i}$ where $p_{\mathbb{E}_{ij}}(x) = p_i(x)^{r_{ij}}$ and $0 < r_{is_i} < \dots < r_{i1} = r_i$

Homework problems:

Page 190, Exercises 6, 7
Page 205, Exercises 5, 6, 7, 8
Page 213, Exercises 1, 2
Page 231, Exercises 7, 8

6. Invariant Factors, Elementary Divisors. Let T be a linear map on the finite dimensional vector space $\mathbb{V}$. Assume that:

$$\mathbb{V} = \mathbb{V}_1 \oplus \dots \oplus \mathbb{V}_r \quad \text{where } \mathbb{V}_i = \langle \alpha_i \rangle_T \text{ and } p_{\mathbb{V}_i}(x) \mid p_{\mathbb{V}_{i-1}}(x) \quad i = 2,\dots, r$$

The polynomial $p_1 = p_{\mathbb{V}_1}$ annihilates $\mathbb{V}_1$ but also $\mathbb{V}_2,\dots,\mathbb{V}_r$ because the minimal polynomials of these summands form a divisor chain. Hence, $p_1(x)$ annihilates the whole space $\mathbb{V}$. On the other hand, there is a vector whose minimal polynomial is $p_1(x)$, namely the generator α_1 of the cyclic space $\mathbb{V}_1$. Hence, $p_1(x) = p_{\mathbb{V}}(x)$. That is, given any other such decomposition of $\mathbb{V}$:

$$\mathbb{V} = \mathbb{U}_1 \oplus \dots \oplus \mathbb{U}_s \quad \text{where } \mathbb{U}_i = \langle \beta_i \rangle_T \text{ and } p_{\mathbb{U}_i}(x) \mid p_{\mathbb{U}_{i-1}}(x) \quad i = 2,\dots, s$$

we certainly have that $p_1(x) = p_{\mathbb{V}_1}(x) = p_{\mathbb{U}_1}(x) = q_1(x)$. We are going to show that $r = s$ and $p_i(x) = q_i(x)$. But we need to know more about cyclic spaces.

Lemma 1. Let α be any vector in $\mathbb{V}$ and $q(x)$ be any polynomial. Let $d(x)$ be the greatest common divisor of $p_\alpha(x)$ and $q(x)$. Then the minimal polynomial of the vector $\beta = q(T)(\alpha)$ is $p(x) = \dfrac{p_\alpha(x)}{d(x)}$. Moreover, $\langle\beta\rangle_T = \langle d(T)(\alpha)\rangle_T$

Proof. We have $p_\alpha(x) = d(x)\cdot p(x)$, $q(x) = d(x)\cdot r(x)$ and where $r(x)$ and $p(x)$ are relatively prime. (If $e \mid r$, $e \mid p$, then $p_\alpha = d\cdot e\cdot p'$, $q = d\cdot e\cdot r'$. Hence, $d\cdot e$ divides p_α as well as q. This is $d\cdot e \mid d$ because d is the g.c.d.(p_α, q). But this is possible only if $e = 1$.)

$p(x)$ annihilates β: $p(T)\beta = p(T)\circ q(T)\alpha = p(T)\circ d(T)\circ r(T)\alpha = r(T)\circ p_\alpha(T)\alpha = 0$.

Now assume that $s(x)$ is any polynomial which annihilates β. We are going to show that it is divisible by $p(x)$:

$0 = s(T)\beta = s(T)\circ q(T)\alpha = s(T)\circ d(T)\circ r(T)\alpha$, i.e., $p_\alpha(x) \mid s(x)\cdot d(x)\cdot r(x)$, $d(x)\cdot p(x) \mid s(x)\cdot d(x)\cdot r(x)$ hence, $p(x) \mid s(x)\cdot r(x)$. Because $p(x)$ and $r(x)$ are relatively prime, we conclude that $p(x) \mid s(x)$. Hence, $p(x)$ is the minimal polynomial of β.

Let $\beta' = d(T)\alpha$. Because $d(x)$ is a divisor of $q(x)$, it is clear that β belongs to the cyclic subspace generated by β'. On the other hand, because $d(x)$ is the g.c.d of $p_\alpha(x)$ and $q(x)$ we have polynomials $q_1(x)$ and $q_2(x)$ such that: $d(x) = q_1(x)p_\alpha(x) + q_2(x)q(x)$, i.e., $d(T)(\alpha) = q_2(T)q(T)(\alpha) = q_2(T)(\beta)$ and it follows that β' belongs to $\langle\beta\rangle_T$. □

Lemma 2. Any invariant subspace of a cyclic space is cyclic.

Proof. Let $\mathbb{U}$ be a subspace of the cyclic space $\langle\alpha\rangle$. Then let $I = \{q(x) \mid q(T)\alpha \in \mathbb{U}\}$. It is easy to see that I is an ideal of $\mathbb{F}[x]$. Hence, $I = (p(x))$ for some polynomial $p(x)$. Let $\beta = p(T)\alpha$. Then $\langle\beta\rangle_T \subseteq \mathbb{U}$. Let γ be any vector in $\mathbb{U}$. Then $\gamma = q(T)\alpha$ and, therefore, $q(x) \in I$. But then $q(x) = d(x)p(x)$ and it follows that $\gamma = d(T)p(T)\alpha = d(T)\beta \in \langle\beta\rangle_T$. □

Theorem 10. Let $T\colon \mathbb{V} \longrightarrow \mathbb{V}$ be a linear map on the finite dimensional vector space $\mathbb{V}$. Then $q(x) \longmapsto \langle q(T)\alpha\rangle_T$ defines a bijective correspondence between the divisors of $p_\alpha(x)$ and the invariant subspaces of $\langle\alpha\rangle_T$.

Proof. Let $\mathbb{U}$ be any invariant subspace of $\langle\alpha\rangle_T$. Then, by Lemma 2, $\mathbb{U}$ is cyclic. That is, $\mathbb{U} = \langle\beta\rangle_T$ where $\beta \in \langle\alpha\rangle_T$. But then $\beta = q(T)\alpha$ for some $q(x)$. According to Lemma 1, we may assume that $\mathbb{U} = \langle d(T)(\alpha)\rangle$ where $d(x)$ divides $p_\alpha(x)$. Now, if we have $\langle d_1(T)\alpha\rangle = \langle d_2(T)\alpha\rangle = \mathbb{U}$, where $d_1(x)$ and $d_2(x)$ both divide $p_\alpha(x)$, then, by Lemma 1, $p_{\mathbb{U}}(x) = \dfrac{p_\alpha(x)}{d_1(x)} = \dfrac{p_\alpha(x)}{d_2(x)}$. It follows that $d_1(x) = d_2(x)$. □

We certainly have that $\langle d_1(T)\alpha\rangle \subseteq \langle d_2(T)\alpha\rangle$, in case that $d_2(x) \mid d_1(x)$. On the other hand, if we assume that $\langle d_1(T)\alpha\rangle \subseteq \langle d_2(T)\alpha\rangle$ for divisors $d_1(x)$ and $d_2(x)$ of $p_\alpha(x)$, we first conclude $\langle d_1(T)\alpha\rangle = \langle e_2(T)d_2(T)\alpha\rangle$ where $e_2(x)$ is a divisor of the minimal polynomial of $d_2(T)\alpha$. This is because of Lemma 1, applied to the subspace $\langle d_1(T)\alpha\rangle$ of $\langle d_2(T)\alpha\rangle$. Hence, $e_2(x)$ divides $\dfrac{p_\alpha(x)}{d_2(x)}$, and we conclude that $e_2(x)d_2(x)$ is a divisor of $p_\alpha(x)$. By the last theorem, we have that $d_1(x) = e_2(x)d_2(x)$, i.e., $d_2(x) \mid d_1(x)$.

The last theorem establishes an *order reversing isomorphism between the monic divisors of $p_\alpha(x)$ and the invariant subspaces of $\langle\alpha\rangle_T$*.

Equipped with this complete knowledge about cyclic spaces and their invariant subspaces, we pick any vector in α and any polynomial $q(x)$. The image $q(T)\langle\alpha\rangle_T$ is an invariant subspace of $\langle\alpha\rangle_T$. It is cyclic and trivially generated by $q(T)(\alpha)$. Hence,

Lemma 3. Let α be any vector in $\mathbb{V}$ and let $q(x)$ be any polynomial. Then $q(T)(\langle\alpha\rangle_T) = \langle q(T)(\alpha)\rangle_T$ and $\dim(q(T)\langle\alpha\rangle_T) = \deg(p(x))$ where $p(x) = \dfrac{p_\alpha(x)}{d(x)}$ and $d(x) = \text{g.c.d.}(p_\alpha(x), q(x))$. □

Assume: $\mathbb{V} = \langle\alpha_1\rangle_T \oplus \ldots \oplus \langle\alpha_r\rangle_T = \langle\beta_1\rangle_T \oplus \ldots \oplus \langle\beta_s\rangle_T$, $p_r|\ldots|p_1$, $q_s|\ldots|q_1$

We already know that $p_1 = q_1 = p_{\mathbb{V}}$. We apply the map $p_2(T)$ to both decompositions:

$$p_2(T)(\mathbb{V}) = p_2(T)(\langle\alpha_1\rangle_T) = p_2(T)(\langle\beta_1\rangle_T) \oplus \ldots \oplus p_2(T)(\langle\beta_s\rangle_T)$$

This is true because $p_2(T)$ annihilates α_2 and all the other α_i. But notice:

$$\dim(p_2(T)\langle\alpha_1\rangle_T) = \dim(p_2(T)\langle\beta_1\rangle_T) = \deg\frac{p_{\mathbb{V}}(x)}{p_2(x)}$$

This follows from Lemma 3, α_1 and β_1 have the same minimal polynomial, namely $p_{\mathbb{V}}$, and the g.c.d. of p_2 and p_1 is p_2, because of $p_2|p_1$. Hence all the other summands of the second decomposition of $p_2(T)\mathbb{V}$ must also be zero. In particular:

$$p_2(T)(\beta_2) = 0, \text{ i.e., } p_2(x)|q_2(x).$$

Of course, by symmetry, we also must have $q_2(x)|p_2(x)$, i.e.,

$$p_2(x) = q_2(x)$$

We can repeat this argument:

$$p_3(T)\mathbb{V} = p_3(T)\langle\alpha_1\rangle_T \oplus p_3(T)\langle\alpha_2\rangle_T = p_3(T)\langle\beta_1\rangle_T \oplus p_3(T)\langle\beta_2\rangle_T \ldots \oplus p_3(T)\langle\beta_s\rangle_T$$

Again, the dimensions of the first two summands of the second decomposition are equal to the dimensions of the two summands of the first decomposition. Hence, all other summands of the second decomposition must be zero. In particular,

$$p_3(T)\langle\beta_3\rangle = 0, \text{ i.e., } p_3(x)|q_3(x)$$

Of course, we then must also have that $q_3(x)|p_3(x)$, i.e.,

$$p_3(x) = q_3(x)$$

We conclude, that $r = s$ and that $p_i(x) = q_i(x)$.

Theorem 11. Let T be a linear map on the finite dimensional vector space $\mathbb{V}$. Then $\mathbb{V}$ admits essentially only one cyclic decomposition subject to the divisor chain condition for the minimal polynomials.

Proof. $\langle\alpha_i\rangle_T \cong \langle\beta_i\rangle_T \cong \mathbb{F}[x]/(p_i(x))$. □

Definition. The divisor chain $p_r | \ldots | p_1$ of polynomials in the Cyclic Decomposition Theorem is called the list of *invariant factors.*

Let p(x) be irreducible in $\mathbb{F}[x]$. $\mathbb{U} = \mathbb{U}_{p(x)}$ is called a *generalized eigenspace* if $\mathbb{U}$ is different from the zero-space and if $\mathbb{U} = \{\alpha | p(T)^r(\alpha) = 0$ for some $r \geq 0\}$. If p(x) is linear, i.e., $p(x) = x-c$, then $\mathbb{U}$ contains an eigenvector for c and the eigenspace $\mathbb{E}_c = \{\alpha | T(\alpha) = c.\alpha\}$ is an invariant subspace of $\mathbb{U}_{x-c}$. The subspace $\mathbb{E}_c$ is in general not a direct summand of $\mathbb{U}$, e.g., if $\mathbb{U}$ is cyclic and $p_{\mathbb{U}}$ is $p(x)^r$ for some $r > 0$. According to the Primary Decomposition Theorem, $\mathbb{V}$ is the direct sum of its generalized eigenspaces. Now, every generalized eigenspaces is according to the Cyclic Decomposition Theorem, an essentially unique sum of cyclic spaces whose minimal polynomials are powers of p(x). This decomposes $\mathbb{V}$ into a direct sum of indecomposable spaces. On the other hand, if $\mathbb{V}$ is in any way decomposed into a direct sum of indecomposable spaces, we may collect all those summands $\langle\alpha\rangle_T$, where p_α is $p(x)^r$ for a fixed irreducible factor p(x) of the minimal polynomial. These spaces add up to $\mathbb{U}_{p(x)}$ and can be arranged in such a way that the minimal polynomials form a divisor chain. Hence they are essentially the same indecomposable spaces as before.

Theorem 12. Let $T: \mathbb{V} \longrightarrow \mathbb{V}$ be a linear map on the finite dimensional vector space $\mathbb{V}$. Then $\mathbb{V}$ admits essentially only one decomposition into non-zero directly indecomposable subspaces. □

Definition. The minimal polynomials of the directly indecomposable subspaces of $\mathbb{V}$ are called the *elementary divisors* of T.

Let α be any non-zero vector in $\mathbb{V}$. Then $\deg(p_\alpha(x)) = k > 0$ and α, $T(\alpha)$, $T^{k-1}(\alpha)$ is a basis for $\langle\alpha\rangle_T$. If $p_\alpha(x) = c_0 + c_1x + \ldots + x^n$ is the minimal polynomial for α, then the matrix A_α for T with respect to this basis looks like:

$$\begin{bmatrix} 0 & 0 & 0 & \cdot & \cdot & 0 & -c_0 \\ 1 & 0 & 0 & \cdot & \cdot & 0 & -c_1 \\ 0 & 1 & 0 & \cdot & \cdot & 0 & -c_2 \\ \cdot & \cdot & \cdot & & & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 1 & -c_{k-1} \end{bmatrix}$$

If $\mathbb{V} = \langle\alpha_1\rangle_T \oplus \dots \oplus \langle\alpha_k\rangle_T$ then the base vectors $\alpha_1, \dots, T^{n_1-1}(T)(\alpha), \dots, \alpha_k, \dots, T^{n_k-1}(\alpha_k)$ of the cyclic subspaces form a basis of $\mathbb{V}$. The matrix of T is with respect to this basis a *direct sum* of matrices $A_i = A_{\alpha_i}$:

$$\begin{bmatrix} A_1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & A_2 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & A_3 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ 0 & 0 & 0 & & & & A_k \end{bmatrix}$$

That is, for every linear map T on the finite dimensional vector space $\mathbb{V}$ one can finsd a basis such that the matrix A of T is in *rational* form.

Example: Let $\mathbb{F} = \mathbb{R}$ and $\mathbb{V} = \mathbb{R}^2$. Then for any linear map there is a basis such that the matrix of T looks like one of the following:

$$\begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix}, \quad \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix}, \quad \begin{bmatrix} 0 & -c^2 \\ 1 & +2c \end{bmatrix}, \quad \begin{bmatrix} 0 & -a^2-b^2 \\ 1 & +2a \end{bmatrix}$$

This corresponds to the possibilities for the elementary divisors:

$$(x-c),(x-c); \quad (x-c_1),(x-c_2); \quad (x-c)^2; \quad x^2 - 2ax + (a^2 + b^2)$$

Exercise. Do the same analysis for $\mathbb{V} = R^3$. Notice, the degree of the product of the elementary divisors must be three.

7. Cayley–Hamilton Theorem, Jordan Normal Form. Let T: $\mathbb{V} \longrightarrow \mathbb{V}$ be a linear map on the finite dimensional vector space $\mathbb{V}$. The *characteristic polynomial* is defined by $c_T(x) = \det(xI - A)$ where A is the matrix of T with respect to any basis. It is easy to see that $c_T(x)$ is independent of the chosen basis. Here we perceive the determinant as a function with polynomial entries, i.e., as an n-linear, alternating function with entries from the commutative ring $\mathbb{K} = \mathbb{F}[x]$.

Theorem 13. Let T: $\mathbb{V} \longrightarrow \mathbb{V}$ be a linear map on the finite dimensional vector space $\mathbb{V}$. Let $\mathbb{U} = \langle\alpha\rangle_T$ be a non-zero cyclic subspace of $\mathbb{V}$. Then the characteristic polynomial of the restriction of T to $\mathbb{U}$ is equal to the minimal polynomial of $\mathbb{U}$.

Proof. We may choose as a basis for $\mathbb{U}$ the vectors α, $T(\alpha)$, ... ,$T^{k-1}(\alpha)$ where k is the degree of $p_\alpha(x)$. Now the claim follows by means of simple determinant manipulations. □

Recall that the determinant of a direct sum of matrices is just the product of the determinants, i.e.,

$$\det \begin{bmatrix} A_1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & A_2 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & A_3 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ 0 & 0 & 0 & & & & A_k \end{bmatrix} = \det(A_1)\cdot\det(A_2)\cdot\ldots\cdot\det(A_k)$$

If $\mathbb{V} = \mathbb{V}_1 \oplus \mathbb{V}_2 \oplus \ldots \oplus \mathbb{V}_k$ is a decomposition of $\mathbb{V}$ into cyclic subspaces, where $T|\mathbb{V}_i$ has matrix A_i, we get

$$c_T(x) = \det(xI-A_1)\cdot\det(xI-A_2)\cdot\ldots\det(xI-A_k) = p_{\mathbb{V}_1}(x)\cdot p_{\mathbb{V}_2}(x)\cdot\ldots\cdot p_{\mathbb{V}_k}(x)$$

We may apply this observation to the cyclic decomposition according to the invariant factors or according to elementary divisors and arrive at

Theorem 14. Let T: $\mathbb{V} \longrightarrow \mathbb{V}$ be a linear map on the finite dimensional vector space $\mathbb{V}$. Then the characteristic polynomial $c_T(x)$ is equal to the product of the invariant factors and also equal to the product of all elementary divisors. □

In particular, $p_{\mathbb{V}}(x) \mid c_T(x)$. That is the

Corollary (Cayley–Hamilton Theorem). $c_T(T) = 0$. □

The characteristic polynomial has the same irreducible factors as the minimal polynomial but, in general, they occur with higher multiplicities: Assume that T is diagonalizable, i.e., $p_{\mathbb{V}}(x) = (x-c_1)\cdot\ldots\cdot(x-c_k)$. We then have for the characteristic polynomial $c(x) = (x-c_1)^{n_1} \cdot\ldots\cdot (x-c_k)^{n_k}$ where $n_i = \dim(\mathbb{E}_{c_i})$.

Assume that the cyclic subspace $\langle\alpha\rangle_T$ has $(x-c)^r$ as its minimal polynomial. Then let $\alpha = \alpha_0$, $\alpha_1 = (T-c)\alpha$, $\alpha_2 = (T-c)^2(\alpha)$, $\ldots$, $\alpha_{r-1} = (T-c)^{r-1}(\alpha)$. These r many vectors are linearly independent in $\langle\alpha\rangle_T$. Otherwise we would have a polynomial $q(x) \neq 0$, $\deg(q(x)) < r$, such that $q(T)(\alpha) = 0$. Hence, $\alpha_0, \ldots, \alpha_{r-1}$ form a base of $\langle\alpha\rangle_T$. Notice:

$$(T-c)\alpha_0 = \alpha_1,\ \text{i.e.,}\ T(\alpha_0) = c\alpha_0 + \alpha_1$$

$$(T-c)\alpha_1 = \alpha_2,\ \text{i.e.,}\ T(\alpha_1) = c\alpha_1 + \alpha_2$$

$$(T-c)\alpha_2 = \alpha_3,\ \text{i.e.,}\ T(\alpha_2) = c\alpha_2 + \alpha_3$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$(T-c)\alpha_{r-1} = 0,\ \text{i.e.,}\ T(\alpha_{r-1}) = c\alpha_{r-1}$$

The matrix of T with respect to this basis is called a *Jordan Block*:

$$J = \begin{bmatrix} c & 0 & 0 & \cdot & \cdot & 0 & 0 \\ 1 & c & 0 & \cdot & \cdot & 0 & 0 \\ 0 & 1 & c & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & 1 & & & & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & & c & \cdot \\ 0 & 0 & 0 & & & 1 & c \end{bmatrix}$$

It is easy to see that an invariant subspace for which T has such a matrix representation must be cyclic with $(x-c)^r$ as minimal polynomial.

Theorem 15 (Jordan Normal Form) Let $T : \mathbb{V} \longrightarrow \mathbb{V}$ be a linear map on the finite dimensional vector space $\mathbb{V}$. Assume that all prime factors of the characteristic polynomial are linear. Then $\mathbb{V}$ admits a basis such that the matrix for T is a direct sum of Jordan matrices. □

Such a Jordan decomposition is always possible if $\mathbb{F}$ is algebraically closed, e.g., for $\mathbb{F} = \mathbb{C}$. The decomposition is essentially unique because it leads to the elementary divisors for T.

Assume $\mathbb{F} = \mathbb{C}$ or that $\mathbb{F}$ is algebraically closed. Then T has an eigenvector. This follows immediately from the Jordan Form. More generally, we have the following: *If* $S \circ T = T \circ S$ *then* T *and* S *have a common eigenvector.* Indeed, if $\mathbb{E}_C$ is an eigenspace for T then $\mathbb{E}_C$ is also invariant under S. Any eigenvector for S in $\mathbb{E}_C$ is a common eigenvector.

Within any eigenspace $\mathbb{E}_C$ any subspace is invariant and therefore has a direct complement. This fact can be generalized to arbitrary invariant subspaces of diagonalizable maps:

Theorem 16. Let $T: \mathbb{V} \longrightarrow \mathbb{V}$ be a linear map on the finite dimensional vector space $\mathbb{V}$ over the algebraically closed field $\mathbb{F}$. Then the following statements are equivalent:

(a) T is diagonalizable, i.e., $\mathbb{V}$ is a direct sum of eigenspaces $\mathbb{E}_{C_i}$.

(b) $p_{\mathbb{V}}(x) = (x-c_1)\cdot \ldots \cdot (x-c_k)$, c_i pairwise distinct.

(c) Every invariant subspace $\mathbb{U}$ has a direct invariant complement. That is, $\mathbb{V} = \mathbb{U} \oplus \mathbb{U}'$ for some invariant subspace $\mathbb{U}'$.

Proof. The equivalence of (a) and (b) is an immediate consequence of the Primary Decomposition Theorem and has been shown before.

Assume (b) and let $\mathbb{U}$ be any invariant subspace of $\mathbb{V}$. Because the minimal polynomial of $T|\mathbb{U}$ is a divisor of $p_{\mathbb{V}}(x)$, one has $\mathbb{U} = \mathbb{U}_1 \oplus \ldots \oplus \mathbb{U}_l$ with eigenspaces $\mathbb{U}_i \subseteq \mathbb{U}$. Now, each $\mathbb{U}_i$ is a subspace of some $\mathbb{E}_{C_j}$ and, without loss of generality, we may assume that $\mathbb{U}_i \subseteq \mathbb{E}_{C_i}$. But then $\mathbb{E}_{C_1} = \mathbb{U}_1 \oplus \mathbb{U}_1'$, ... $\mathbb{E}_{C_l} = \mathbb{U}_l \oplus \mathbb{U}_l'$. Hence, $\mathbb{V} = \mathbb{U}_1 \oplus \ldots \oplus \mathbb{U}_l \oplus (\mathbb{U}_1' \oplus \ldots \oplus \mathbb{U}_l' \oplus \mathbb{E}_{C_{l+1}} \oplus \ldots \; \mathbb{E}_{C_k})$, i.e., $\mathbb{V} = \mathbb{U} \oplus \mathbb{U}'$, for some invariant subspace $\mathbb{U}'$.

Now assume (c), i.e., every invariant subspace is a direct summand. Let α_1 be any eigenvector. Then $\langle\alpha_1\rangle$ is invariant and therefore $\mathbb{V} = \langle\alpha_1\rangle \oplus \mathbb{U}_2$ with some invariant subspace $\mathbb{U}_2$. If $\mathbb{U}_2 \neq \mathbb{O}$, pick any eigenvector α_2 of T within $\mathbb{U}_2$. The sum $\langle\alpha_1\rangle + \langle\alpha_2\rangle$ is direct and, obviously, an invariant subspace of $\mathbb{V}$. Hence, there is some invariant $\mathbb{U}_3$ such that $\mathbb{V} = \langle\alpha_1\rangle \oplus \langle\alpha_2\rangle \oplus \mathbb{U}_3$. If $\mathbb{U}_3$ is different from the zero-space, there is some eigenvector α_3 within $\mathbb{U}_3$ and the sum $\langle\alpha_1\rangle + \langle\alpha_2\rangle + \langle\alpha_3\rangle$ is direct etc. Hence, $\mathbb{V}$ is a direct sum of one dimensional spaces, each one generated by an eigenvector. That is, T is diagonalizable. □

Definition. A linear map T is called *semi-simple* if every invariant subspace is a direct summand.

We have shown that for finite dimensional vector spaces over algebraically closed fields the class of semi-simple linear maps coincides with the class of linear maps which admit eigenbases.

It follows from the Jordan Normal Form that *any* T *is a sum of a semi-simple map D and a nilpotent map* N: For each Jordan Block J we have that $T|\mathbb{E}$ is the sum of $D(\mathbb{E}) = c.I$ and a nilpotent map $N(\mathbb{E})$. Here $\mathbb{E}$ is the indecomposable subspace for which J is the matrix of $T|\mathbb{E}$. If we define $T(\mathbb{E})$ as the linear map which is $T|\mathbb{E}$ on $\mathbb{E}$ and zero on the other indecomposable spaces $\mathbb{E}'$, then it is clear that T is just the sum of the $T(\mathbb{E})$'s where $T(\mathbb{E}) = D(\mathbb{E}) + N(\mathbb{E})$. The sum of the $N(\mathbb{E})$'s is nilpotent because the sum of commuting nilpotent linear maps is nilpotent. Hence $T = D + N$ where D is the sum of the $D(\mathbb{E})$'s and N is the sum of all $N(\mathbb{E})$'s. It is also clear that we have $D \circ N = N \circ D$. It is quite remarkable that one can achieve such a decomposition for finite dimensional vector spaces over arbitrary fields.

Homework problems.

Page 189, Exercises 2, 3
Page 198, Exercise 2
Page 218, Exercise 2
Page 230, Exercises 1, 2
Page 241, Exercise 1
Page 250, Exercise 10
Page 261, Exercise 4

8. The Smith Normal Form. Computation of the Invariant Factors. Let M be an $m \times n$ matrix with entries in the polynomial ring $\mathbb{F}[x]$. M is said to be in *Smith normal form* if the only non-zero entries can be found on the main diagonal and if the non-zero entries form a divisor chain of monic polynomials. That is, M looks like

$$\begin{bmatrix}
1 &&&&&&&&&&& \\
& 1 &&&&&&&&&& \\
&& \ddots &&&&&&&&& \\
&&& 1 &&&&&&&& \\
&&&& p_1 &&&&&&& \\
&&&&& p_2 && 0 &&&& \\
& 0 &&&&& \ddots &&&&& \\
&&&&&&& p_k &&&& \\
&&&&&&&& 0 &&& \\
&&&&&&&&& 0 && \\
&&&&&&&&&& \ddots & \\
&&&&&&&&&&& 0 \cdots
\end{bmatrix}$$

Notice that 1 divides every polynomial and that 0 is divisible by every polynomial. Hence, all the entries on the main diagonal form a divisor chain. That is,

$$p_{ii} \mid p_{(i+1)(i+1)} \quad \text{for } i = 1,\ldots, l \quad \text{where } l = \min(m,n).$$

Assume that M is a scalar matrix, i.e., all entries are elements of $\mathbb{F}$. Then by means of elementary row and column operations, A can be transformed into

$$M = \begin{bmatrix}
1 &&&&&&&& \\
& 1 &&&&&&& \\
&& \ddots &&& 0 &&& \\
&&& 1 &&&&& \\
&&&& 0 &&&& \\
&&&&& 0 &&& \\
& 0 &&&&& \ddots && \\
&&&&&&& 0 \cdots
\end{bmatrix}$$

The number r of units on the main diagonal of M is the rank of A.

Examples. The following matrices are in Smith normal form.

$$\text{(a)}\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{(b)}\ \begin{bmatrix} 1 & 0 \\ 0 & x \\ 0 & 0 \end{bmatrix} \qquad \text{(c)}\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & x+1 & 0 & 0 \\ 0 & 0 & x^2-1 & 0 \end{bmatrix}$$

We are going to show that any matrix $A = A(x)$ with entries taken from $\mathbb{F}[x]$, i.e., $A \in \mathbb{F}[x]^{m \times n}$, can be transformed into Smith normal form by means of a succession of elementary row and column operations. As in the scalar case, we distinguish three types of elementary (row) operations:

R1. Multiply one row by some scalar $c \neq 0$.

R2. Replace the rth row A_r by $A_r + p(x)A_s$ where $r \neq s$ and $p(x)$ is any polynomial in $\mathbb{F}[x]$.

R3. Interchange two rows of A.

It's easy to see that **R3** can be obtained by a succession of operations of type 1 and type 2. The column operations C1 – C3 are defined similarly. We say that A and B are *row equivalent* if B can be obtained from A by means of finitely many row operations. Because any row operation is reversible by a similar row operation, it is clear that row equivalence is an equivalence relation. We need a preliminary result.

Lemma 1. Assume that the first column of the matrix $A(x)$ is different from zero. Let $p(x) = \text{g.c.d.}(p_{11}(x),\ldots, p_{m1}(x))$. Then $A(x)$ is row equivalent to a matrix $B(x)$ which has

$$\begin{bmatrix} p(x) \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

as its first column.

Proof. We need a preliminary remark. Let $P = (p_i(x))$, $i \in I$, be a family of polynomials. Then the greatest common divisor $d(x)$ of P is the generator of the principal ideal (P), which is the ideal generated by P. Recall, (P) is the set of all finite linear combinations:

$$(P) = \{f_{i_1}(x) \cdot p_{i_1}(x) + \dots + f_{i_k}(x) \cdot p_{i_k}(x) \mid i_\nu \in I,\ f_{i_\nu}(x) \in \mathbb{F}[x]\}$$

Let P' be obtained from P by replacing $p_i(x)$ by $c \cdot p_i(x)$, where c is a constant different from zero. Then $(P) = (P')$ and P and P' have the same greatest common divisor. Similarly, if we replace $p_i(x)$ by $p_i(x) + f(x)p_j(x)$ where $i \neq j$ and where $f(x) \in \mathbb{F}[x]$, then the g.c.d. remains the same. In particular, the elementary operations don't change the g.c.d. of a given matrix. Also, row operations don't change the g.c.d. of any given column and column operations don't change the g.c.d. of any given row.

In order to prove the lemma, let $p_j(x)$ be a polynomial in the first column of lowest degree:

$$A(x) = \begin{bmatrix} p_1(x) & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & & \\ p_j(x) & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ p_m(x) & \cdot & \cdot \end{bmatrix}$$

We divide each $p_i(x)$, $i \neq j$, by $p_j(x)$ with remainder:

$$p_i(x) = q_i(x) \cdot p_j(x) + r_i(x),\ 0 \leq \deg(r_i(x)) < \deg(p_j(x)) \text{ or } r_i(x) = 0$$

The first entry of each of the rows $i \neq j$ can be made $r_i(x)$ by adding to the ith row $(-q_i(x))$-times the jth row. A multiplication of the jth row by a constant different from zero makes $p_j(x)$ to a monic polynomial $\tilde{p}_j(x$. If we interchange the first and jth row then $A(x)$ is row equivalent to:

$$\begin{bmatrix} \tilde{p}_j(x) & \cdot & \cdot \\ r_2(x) & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ r_1(x) & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ r_m(x) & \cdot & \cdot \end{bmatrix}$$

Any $r_i(x)$ is either zero or has a degree less than the degree of $\tilde{p}_j(x)$. We may repeat the process until we get the desired result. □

Lemma 2. Let $A(x)$ be an $m \times n$ matrix where the entries are polynomials in $F[x]$. Then $A(x)$ is equivalent to a matrix which is in Smith normal form.

Proof. There is nothing to prove for the zero matrix. If $A(x)$ is different from the zero matrix we show that $A(x)$ is equivalent to a matrix

$$B(x) = \begin{bmatrix} p_1(x) & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & R & & \cdot \\ \cdot & \cdot & & & & \cdot \\ 0 & \cdot & & & & \cdot \end{bmatrix}$$

where $p_1(x) = \text{g.c.d.}(A(x)) = \text{g.c.d.}(B(x))$. This will prove the claim.

Let $k = k(A(x))$ be the minimum of the degrees of the entries $p_{ij}(x)$ of $A(x)$. By interchanging columns we can move an entry of degree k to the first column. According to Lemma 1, $A(x)$ is equivalent to a matrix like

$$A_1 = \begin{bmatrix} q(x) & a(x) & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & & \\ \cdot & \cdot & & S & & \\ \cdot & \cdot & & & & \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

Of course, $\deg(q(x)) \leq k$. Now, if q divides each entry of the first row then by adding multiples of the first column to the other columns we get

$$A^{\sim} = \begin{bmatrix} q(x) & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & & \\ \cdot & \cdot & & S & & \\ \cdot & \cdot & & & & \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

Otherwise, an application of Lemma 1 to the first row of A_1 yields

$$A_2 = \begin{bmatrix} q_1(x) & 0 & \cdot & \cdot & \cdot & 0 \\ a'(x) & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & & \\ \cdot & \cdot & & S' & & \\ \cdot & \cdot & & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

where now $\deg(q_1(x)) < \deg(q(x)) \leq k$. (One has that $q_1 = \text{g.c.d.}(q,a,\ldots)$ and therefore $\deg(q_1) < \deg(q)$ in case that $q_1 \neq q$, i.e., q does not divide every entry of the first row of A_1.) Now, if q_1 divides each entry of the first column then by adding multiples of the first row to the other rows we get

$$\begin{bmatrix} q_1(x) & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & & \\ \cdot & \cdot & & S' & & \\ \cdot & \cdot & & & & \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

Otherwise, an application of Lemma 1 to the first column of A_2 yields

$$\begin{bmatrix} q_2(x) & a''(x) & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & & \\ \cdot & \cdot & & S'' & & \\ \cdot & \cdot & & & & \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

where $\deg(q_2) < \deg(q_1) < k$.

After finitely many steps, we must get a matrix of the form

$$A^{\sim} = \begin{bmatrix} q(x) & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & & \\ \cdot & \cdot & & S & & \\ \cdot & \cdot & & & & \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

where $\deg(q(x)) \leq k$.

If $q(x)$ divides every entry of S then, because

$$q(x) = \text{g.c.d.}(A^{\sim}) = \text{g.c.d.}(A)$$

we are done: $A^{\sim} = B$ and $S = R$.

Otherwise, there is a column in S which contains a polynomial $g(x)$ which is not divisible by $q(x)$. If we add this column to the first column of $A^{\sim}$ then we get a matrix like

$$\begin{bmatrix} q(x) & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & & \\ g & \cdot & & S & & \\ \cdot & \cdot & & & & \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

Now, $g(x) = s(x)q(x) + r(x)$ where $\deg(r) < \deg(q) \leq k$. Hence, this matrix is equivalent to

$$A' = \begin{bmatrix} q(x) & 0 & \cdots & 0 \\ 0 & \cdot & \cdots & \cdot \\ \cdot & \cdot & & \\ r & \cdot & S & \\ \cdot & \cdot & & \\ 0 & \cdot & \cdots & \cdot \end{bmatrix}$$

We apply the whole process now to A' and arrive at a matrix

$$(A')^{\sim} = \begin{bmatrix} q'(x) & 0 & \cdots & 0 \\ 0 & \cdot & \cdots & \cdot \\ \cdot & \cdot & & \\ \vdots & \vdots & S' & \\ \cdot & \cdot & & \\ 0 & \cdot & \cdots & \cdot \end{bmatrix}$$

where $\deg(q') \leq \deg(r) \leq k' < k$. After finitely many steps we must get the desired matrices $B(x)$ and $\mathring{R}(x)$. □

Definition. Let $M \in \mathbb{F}[x]^{m \times n}$. For any k, $1 \leq k \leq \min(m,n)$, we define $\delta_k(M)$ = g.c.d. of the determinants of all $k \times k$ submatrices.

Lemma 3. If M and N are equivalent matrices in $\mathbb{F}[x]$ then $\delta_k(M) = \delta_k(N)$.

Proof. For a fixed k, $1 \leq k \leq \min(m,n)$, and

$$I = (i_1, \ldots, i_k),\ 1 \leq i_1 < \ldots < i_k \leq m$$

$$J = (j_1, \ldots, j_k),\ 1 \leq j_1 < \ldots < j_k \leq n$$

let $D_{I,J}(M) = \det \begin{bmatrix} M_{i_1 j_1} & \cdots & M_{i_1 j_k} \\ M_{i_2 j_1} & \cdots & M_{i_2 j_k} \\ \vdots & \vdots & \vdots \\ M_{i_k j_1} & \cdots & M_{i_k j_k} \end{bmatrix}$

Then: $\delta_k(M) = \underset{I,J}{\text{g.c.d.}}\ D_{I,J}(M).$

If e is some elementary operation then e(M) is the matrix M after the operation e. We are going to show:

If $f(x)$ divides $D_{I,J}(M)$ for every I and J then $f(x)$ divides $D_{I,J}(e(M))$ for every I and J. Because elementary operations are invertible, the set of $D_{I,J}(M))$'s has the same set of divisors as the set of $D_{I,J}(e(M))$'s. Hence the g.c.d's are the same.

We have to go through the two types of essential row operations.

(a) Multiplication of the rth row by some $c \neq 0$:

$D_{I,J}(M) = D_{I,J}(e(M))$ if $r \notin \{i_1,\dots, i_k\}$

$D_{I,J}(e(M)) = D_{I,J}(M)$ if $r \in \{i_1,\dots, i_k\}$

Clearly, $f \mid D_{I,J}(M)$ iff $f \mid D_{I,J}(e(M))$.

(b) Replace row r by (row r plus $g(x) \times$ row s):

$D_{I,J}(M) = D_{I,J}(e(M))$ if $r \notin \{i_1,\dots, i_k\}$

If $r \in \{i_1,\dots, i_k\}$ then

$$D_{I,J}(e(M)) = \det\begin{bmatrix} M_{i_1j_1} & \cdots\cdot & M_{i_1j_k} \\ M_{i_2j_1} & \cdots\cdot & M_{i_2j_k} \\ \cdot & \cdots\cdot & \cdot \\ M_{r,j_1} + g(x)M_{s,j_1} & \cdots\cdot & M_{r,j_k} + g(x)M_{s,j_k} \\ \cdot & \cdots\cdot & \cdot \\ M_{i_kj_1} & \cdots\cdot & M_{i_kj_k} \end{bmatrix} =$$

$$D_{I,J}(M) + g(x)\det\begin{bmatrix} M_{i_1j_1} & \cdots\cdot & M_{i_1j_k} \\ M_{i_2j_1} & \cdots\cdot & M_{i_2j_k} \\ \cdot & \cdots\cdot & \cdot \\ M_{s,j_1} & \cdots\cdot & M_{s,j_k} \\ \cdot & \cdots\cdot & \cdot \\ M_{i_kj_1} & \cdots\cdot & M_{i_kj_k} \end{bmatrix} =$$

$D_{I,J} + 0$, in case that $s \in \{i_1,\dots, i_{r-1}, i_{r+1},\dots, i_k\}$ or is equal to

$D_{I,J} \pm D_{I',J}$ where I' is a permutation of $(i_1,\dots, i_{r-1}, s, i_{r+1},\dots, i_k)$.

At any rate, if some $f(x)$ divides every $D_{I,J}(M)$ then $f(x)$ divides every $D_{I,J}(e(M))$. □

Theorem 17. Every matrix $M \in \mathbb{F}[x]^{m \times n}$ is equivalent to exactly one matrix which is in Smith normal form.

Proof. Let N be in normal form:

$$N = \begin{bmatrix} f_1 & & & & \\ & f_2 & & 0 & \\ & & \ddots & & \\ & 0 & & f_t & \cdots \end{bmatrix} \quad \text{where } t = \min(m,n).$$

Of course, the first f_i's may be units and the last f_i's may be zeros. Let $k \leq t$. Assume that a $k \times k$ submatrix $N_{I,J}$ contains the ith row but not the ith column. Because the only term $N_{i,j}$ which is not necessarily zero is $N_{i,i}$, and this term is missing in $N_{I,J}$, we conclude that the determinant $D_{I,J}$ of $N_{I,J}$ is zero. Thus, in order to find the greatest common divisor of all determinants of $k \times k$ submatrices, we only have to consider submatrices where $I = J$. For $k = 1$, we get $\delta_1 = f_1$ because $f_1 | f_2 | \ldots | f_t$. For $k = 2$ we get $\delta_2 = f_1 \cdot f_2$ because $f_1 \cdot f_2 | f_i \cdot f_j$ for $i < j$. We have $1 \leq i$ and $2 \leq j$ and $f_1 | f_i$ and $f_2 | f_j$. In general,

$$\delta_k = f_1 \cdot f_2 \cdot \ldots \cdot f_k$$

Assume that M is equivalent to N and N' where N and N' are in normal form. Let s be the first k, if there is one, where $f_s = 0$. Then

$$\delta_s = f_1 \cdot \ldots \cdot f_s = 0 = f'_1 \cdot \ldots \cdot f'_s$$

Hence, $f'_k = 0$ for some $k \leq s$. This is, $s' \leq s$ where s' is the first k in N' where f'_k is zero. One concludes $s = s'$, by symmetry.

If $1 < s$ then $\delta_1 = f_1 = f'_1$. For any k, $1 < k < s$, and $f_i = f'_i$, $i < k$, one concludes

$$\delta_k = (f_1 \cdot \ldots \cdot f_{k-1}) \cdot f_k = (f'_1 \cdot \ldots \cdot f'_{k-1}) \cdot f'_k$$

Hence, $f_k = f'_k$. □

Example. Any m × n **scalar** matrix A is equivalent to exactly one diagonal matrix which has r-many units on the main diagonal, followed by (m–r) many zeros. Here r is the rank of A.

Let B be the matrix of the linear map T with respect to a basis $\beta_1,\ldots,\beta_n$. We are going to show that the Smith normal form of the **characteristic** matrix (xI –B) is

$$\begin{bmatrix} 1 & & & & & & & \\ & \cdot & & & & & & \\ & & \cdot & & & & & \\ & & & 1 & & & & \\ & & & & p_r & & & \\ & & & & & \cdot & & \\ & & & & & & \cdot & \\ & & & & & & & p_1 \end{bmatrix}$$

where $p_r|\ldots|p_1$ is the list of invariant factors.

We know that there is a basis $\alpha_1,\ldots,\alpha_n$ such that the matrix A for T is in rational form. But then A and B are similar, i.e.,

$$A = P^{-1}\circ B\circ P$$

for some invertible matrix P. But then

$$P^{-1}\circ(xI - B)\circ P = xI - P^{-1}\circ B\circ P = xI - A$$

where xI – A is a block of matrices and where each block looks like:

$$\begin{bmatrix} x & 0 & 0 & \cdot\ \cdot & 0 & +c_0 \\ -1 & x & 0 & \cdot\ \cdot & 0 & +c_1 \\ 0 & -1 & x & \cdot\ \cdot & 0 & +c_2 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \cdot\ \cdot & -1 & x+c_{k-1} \end{bmatrix}$$

If e is an elementary row operation and M an m × n matrix with polynomial entries then one has that e(M) = E∘M where E is the *elementary* matrix $e(I_m)$. I_m is the m × m unit matrix. If e^{-1} reverses the elementary operation e, then one has $e^{-1}(M) = E^{-1} \circ M$. If f is an elementary column operation then $f(M) = M \circ F$ where F is the matrix $f(I_n)$ and $f^{-1}(M) = F^{-1}(M)$. Hence, if N is the normal form for M then

$$N = E_u \circ \ldots \circ E_1 \circ M \circ F_1 \ldots \circ F_v$$

Assume that M is invertible. Then det(M) is a non-zero element $c \in \mathbb{F}$. According to Lemma 1, M is row equivalent to a matrix

$$M_1 = \begin{bmatrix} q(x) & a(x) & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & & \\ \cdot & \cdot & & S & & \\ \cdot & \cdot & & & & \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

Because M_1 is invertible, $\det(M_1) = q(x) \cdot \det(S)$ is a non-zero constant. That is, $q(x) = 1$. It follows that M is row equivalent to a matrix which has units on its main diagonal. But then M is row equivalent to the unit matrix.

Theorem 18. $M \in \mathbb{F}[x]^{n \times n}$ is invertible iff M is a product of elementary matrices $E_u, \ldots, E_1$ iff M is a product of elementary matrices $F_v, \ldots, F_1$.

If we replace in the equation $P^{-1} \circ (xI - B) \circ P = xI - A$ the matrix P^{-1} by a product of elementary matrices E_i and P by a product of elementary matrices F_j then it becomes apparent that $(xI - B)$ is equivalent to $(xI - A)$.

It is quite easy to see that each block

$$xI_{n_i} - A_i = \begin{bmatrix} x & 0 & 0 & \cdot & \cdot & 0 & +c_0 \\ -1 & x & 0 & \cdot & \cdot & 0 & +c_1 \\ 0 & -1 & x & \cdot & \cdot & 0 & +c_2 \\ \cdot & \cdot & \cdot & & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & -1 & x+c_{k-1} \end{bmatrix}$$

of $(xI - A)$ is equivalent to the matrix

$$\begin{bmatrix} p(x) & & & & \\ & 1 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & 1 \end{bmatrix}$$

where $p(x) = c_0 + c_1 x \dots + x^k$.

Theorem 19. Let T: $\mathbb{V} \longrightarrow \mathbb{V}$ be a linear map on the finite dimensional vector space $\mathbb{V}$ and let B be the matrix of T with respect to any given basis. Then the Smith normal form of the characteristic matrix $(xI - B)$ is

$$\begin{bmatrix} 1 & & & & & & & \\ & \cdot & & & & & & \\ & & \cdot & & & & & \\ & & & \cdot & & & & \\ & & & & 1 & & & \\ & & & & & p_r & & \\ & & & & & & \ddots & \\ & & & & & & & p_1 \end{bmatrix}$$

where $p_r | \dots | p_1$ is the list of invariant factors. □

Homework problems.

Page 242, Exercises 7, 12, 13, 15, 19

Page 261, Exercises 1, 2, 3

9. Euclidean Spaces. Let $\mathbb{V}$ be a finite dimensional vector space over the field $\mathbb{R}$ of real numbers. A map

$$\langle\ |\ \rangle : \mathbb{V} \times \mathbb{V} \longrightarrow \mathbb{R}$$

is called an *inner* product if it has the following properties:

(a) $\langle c.\alpha|\beta\rangle = c\cdot\langle\alpha|\beta\rangle$, $\langle\alpha + \beta|\gamma\rangle = \langle\alpha|\beta\rangle + \langle\alpha|\gamma\rangle$

(b) $\langle\alpha|\beta\rangle = \langle\beta|\gamma\rangle$

(c) $\langle\alpha|\alpha\rangle \geq 0$ and $\langle\alpha|\alpha\rangle = 0$ iff $\alpha = 0$

Because of property (a), for every vector β, the map $f_\beta\colon \alpha \longmapsto \langle\alpha|\beta\rangle$ is linear. Because of (b), for every α the map $f_\alpha\colon \beta \longmapsto \langle\alpha|\beta\rangle$ is linear in β. That is, $\langle\ |\ \rangle$ is a *bilinear form*. In particular, $\langle\sigma|\beta\rangle = 0 = \langle\alpha|\sigma\rangle$ holds for every α and β. The property (c) says, that the form is *positive definite*. The mathematical structure $(\mathbb{V},\langle\ |\ \rangle$ is called a *Euclidean Space.*

For every α one has $\langle\alpha|\alpha\rangle \geq 0$. The number

$$\|\alpha\| = \sqrt{\langle\alpha|\alpha\rangle}$$

is called the *norm* of the vector α. For any $\alpha \neq \sigma$, the vector

$$\varepsilon_\alpha = \frac{1}{\|\alpha\|}\,\alpha$$

is a *unnit* vector, i.e., a vector of norm (or *length)* one.

The vectors α and β are called *perpendicular* to each other if $\langle\alpha|\beta\rangle = 0$. One writes for this $\alpha \perp \beta$. If α and β are perpendicular to each other then one has the

Phythagorean Theorem : $\|\alpha + \beta\|^2 = \|\alpha\|^2 + \|\beta\|^2$

If $\alpha \neq \sigma$ and β is any vector in $\mathbb{V}$, the vector

$$\mathrm{proj}_\alpha(\beta) = \langle\beta|\varepsilon_\alpha\rangle.\varepsilon_\alpha$$

is called the *projection* of β in the direction α. It is the only vector $\gamma = c.\gamma$ such that $(\beta - \gamma) \perp \alpha$.

Let $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ be a system of unit vectors which are pairwise perpendicular. If the vector α is a linear combination of the ε_i, that is, $\alpha = c_1.\varepsilon_1 + c_2.\varepsilon_2 + \dots + c_m.\varepsilon_m$, multiplication of both sides by ε_i shows that that $c_i = \langle \alpha | \varepsilon_i \rangle$. That is, the coefficients c_i are uniquely determined by α. In particular, any *orthonormal* system of vectors is linearly independent. If α is any vector in $\mathbb{V}$, the vector

$$\alpha' = \langle \alpha | \varepsilon_1 \rangle.\varepsilon_1 + \dots + \langle \alpha | \varepsilon_m \rangle.\varepsilon_m$$

is called the *Fourier expansion* of α or the *projection* of α along H, $\text{proj}_H(\alpha)$, where H is the subspace generated by $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$. It is easy to see that $\alpha - \text{proj}_H(\alpha)$ is perpendicular to every ε_i. Hence, $\alpha - \text{proj}_H(\alpha) \perp \beta$, where β is any vector in H. Let β be any vector in H. Then $(\text{proj}_H(\alpha) - \beta)$ belongs to H and in the equation

$$\alpha - \beta = (\alpha - \text{proj}_H(\alpha)) + (\text{proj}_H(\alpha) - \beta)$$

$(\alpha - \text{proj}_H(\alpha))$ is perpendicular to $(\text{proj}_H(\alpha) - \beta)$. Hence, by the Pythagorean Theorem, one has

$$\|\alpha - \beta\|^2 = \|\alpha - \text{proj}_H(\alpha)\|^2 + \|\text{proj}_H(\alpha) - \beta\|^2$$

We conclude, $\|\alpha - \beta\|^2 \geq \|\alpha - \text{proj}_H(\alpha)\|^2$ and equality holds exactly when β is equal to $\text{proj}_H(\alpha)$. That is, $\text{proj}_H(\alpha)$ is the unique vector β_0 in H for which the function $d(\beta) = \|\alpha - \beta\|$, $\beta \in H$, takes on its minimum.

Let $\alpha_1, \alpha_2, \dots, \alpha_m$ be a set of linearly independent vectors in $\mathbb{V}$. We define succesively a system of vectors $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ such that for every j, $\langle \alpha_1, \dots, \alpha_j \rangle = \langle \varepsilon_1, \dots, \varepsilon_j \rangle$ and where the ε_i are unit vectors which are pairwise perpendicular. This process is called the *Gram-Schmidt Orthogonalization:*

$\varepsilon_1 = \varepsilon_{\alpha_1}$;

$\varepsilon_2' = \alpha_2 - \langle \alpha_2 | \varepsilon_1 \rangle.\varepsilon_1$, $\varepsilon_2 = \dfrac{1}{\|\varepsilon_2'\|} \varepsilon_2'$;

$\varepsilon_3' = \alpha_3 - \langle \alpha_3, \varepsilon_1 \rangle.\varepsilon_1 - \langle \alpha_3, \varepsilon_2 \rangle.\varepsilon_2$, $\varepsilon_3 = \dfrac{1}{\|\varepsilon_3'\|} \varepsilon_3'$; etc

In particular, $\mathbb{V}$ admits a basis of unit vectors which are pairwise perpendicular to each other. Such an orthonormal basis is also called a *Cartesian* coordinate system. The unit vectors of $\mathbb{R}^n$ are an example of an orthonormal system.

Let $\alpha_1,\dots,\alpha_n$ and $\beta_1,\dots,\beta_n$ be two cartesian coordinate systems. Then:

$$\alpha_i = \sum_\nu c_{\nu i}\beta_\nu = \sum_\nu \langle\alpha_i \mid \beta_\nu\rangle\beta_\nu$$

and $\langle\alpha_i \mid \alpha_j\rangle = \delta^i_j = \langle\sum_\nu c_{\nu i}\beta_\nu \mid \sum_\nu c_{\mu j}\beta_\mu\rangle = \sum_\nu\sum_\mu c_{\nu i}c_{\mu j}\langle\beta_\nu,\beta_\mu\rangle = \sum_\nu\sum_\mu \nu_i c_{\mu j}\delta^\mu_\nu =$
$\sum_\nu c_{\nu i}c_{\nu j}$

That is, the columns of the matrix C form an orthonormal system. This is the same as saying that

$$C\circ C^* = I$$

where for a matrix A, the matrix A^* is the transposed of A. But then $C^* = C^{-1}$ (for matrices, left inverses are inverses) and $C^{-1}\circ C = I$ shows that

$$C^*\circ C = I$$

which now tells us that the rows of C form an orthonormal system.

Definition. An n×n-matrix C is called *orthogonal* if the rows form an orthonormal system. This is the same as saying that the columns form an orthonormal system. Equivalently, the inverse of C is equal to C^*.

We have shown that the coordinate transformation matrix between two cartesian systems is orthogonal. Let $T\colon \mathbb{V} \longrightarrow \mathbb{V}$ be linear. The map T is called *symmetric* if one has for all α, β:

$$\langle T(\alpha) \mid \beta\rangle = \langle\alpha, T(\beta)\rangle$$

Theorem 19. The following statements about a linear map on the finite dimensional Euclidean space are equivalent.

(a) T is symmetric.

(b) Mat(T) is symmetric for every cartesian coordinate system of $\mathbb{V}$.

Proof. Assume that T is symmetric and that $\alpha_1,\dots,\alpha_n$ is cartesian. Then,

$$\langle T(\alpha_i)|\alpha_j\rangle = a_{ji} = \langle\alpha_i|T(\alpha_j)\rangle = \langle T(\alpha_j)|\alpha_i\rangle = a_{ij}$$

shows that the matrix A for T is symmetric.
Now assume that A is symmetric. Then

$$\langle T(\alpha_i)|\alpha_j\rangle = a_{ji} = a_{ij} = \langle T(\alpha_j)|\alpha_i\rangle = \langle\alpha_i|T(\alpha_j)\rangle$$

Hence, $\langle T(\Sigma\, a_i\alpha_i)|\Sigma\, b_j\alpha_j\rangle = \sum_i\sum_j a_i b_j\ \langle T(\alpha_i)|\alpha_j\rangle = \sum_i\sum_j a_i b_j\langle\alpha_i|T(\alpha_j)\rangle =$ $\langle\Sigma a_i\alpha_i|T(\Sigma b_j\alpha_j\rangle$, i.e., T is symmetric. □

Theorem 20. Let T be a symmetric map on the finite dimensional Euclidean space $\mathbb{V}$. Then T admits a cartesian eigenbase.

Proof. We first need to show that the minimal polynomial p(x) for T has only linear factors of multiplicity one. Assume $x^2 - 2ax + a^2 + b^2 | p(x)$ where $b \neq 0$. Then there is a vector $\alpha \neq \sigma$ such that $(T^2 - 2aT + a^2).\alpha = -b^2.\alpha$. Hence,

$$\langle(T^2 - 2aT + a^2).\alpha|\alpha\rangle = -b^2\langle\alpha,\alpha\rangle,\ \langle(T-a)^2\alpha|\alpha\rangle = \langle(T-a)\alpha|(T-a)\alpha\rangle = -b^2\langle\alpha,\alpha\rangle.$$

Now, in $\langle(T-a)\alpha|(T-a)\alpha\rangle = -b^2\langle\alpha,\alpha\rangle$ the right-hand side is negative while the left-hand side is non-negative. This is a contradiction. Hence, all irreducible factors of p(x) have to be linear. Now assume that $(x-a)^2|p(x)$. Then there is a vector α such that $(T - a)^2.\alpha = \sigma$ but $(T - a).\alpha \neq \sigma$. But in

$$\langle(T - a)^2.\alpha|\alpha\rangle = \langle(T - a).\alpha|(T - a).\alpha\rangle$$

the left-hand side is zero while the right-hand side is positive. This is a contradiction.
Now we know that $\mathbb{V}$ admits a decomposition into eigenspaces $\mathbf{E}_c$. For each of these eigenspaces we can choose a cartesian base. We can combine these bases to a basis of $\mathbb{V}$. We are done if we can show that eigenvectors belonging to different eigenvalues are perpendicular. Assume that c and d are different eigenvalues and α and β eigenvectors for c and d, respectively. We then have $T(\alpha) = c.\alpha$, $T(\beta) = d.\beta$ and

$$\langle T(\alpha) | \beta\rangle = \langle c.\alpha | \beta\rangle = \langle \alpha | T(\beta)\rangle = \langle \alpha | d.\beta\rangle, \text{ i.e., } c\cdot\langle \alpha,\beta\rangle = d.\langle \alpha,\beta\rangle$$

Because of $c \neq d$ one concludes $\langle \alpha,\beta\rangle = 0$. □

Assume that A is a symmetric matrix with real entries. Let T be the linear map which has matrix A with respect to the unit vectors. There is a cartesian basis such that the matrix for T is a diagonal matrix D and A and D are conjugate via an orthogonal matrix C:

Corollary. Let $A \in \mathbb{R}^{n\times n}$ be symmetric. Then there is an orthogonal matrix C such that $D = C^* \circ A \circ C$ is a diagonal matrix. □