Field Theory and Canonical Forms of Matrices

Lecturer: Prof. S.K. Sardar (S.K.S.)
Notes taken by Sayan Das

6th Semester, 2026

Contents

Everything comes in time to him who knows how to wait.

—Leo Tolstoy, [Tol69]

Theme 1

Canonical forms of matrices

1.1 19 January

1.1.1 Matrix representations of linear transformations

Linear transformations between two finite-dimensional vector spaces is determined by a matrix and vice-versa.

Example 1.1 (Examples of finite dimensional vector spaces). For any field $\mathbb{F}$, and any positive integer n, $\mathbb{F}^n$ is a vector space over $\mathbb{F}$ with $\dim_{\mathbb{F}}(\mathbb{F}^n) = n$.

Let V be a vector space with $\dim V = n$ and $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis of V. For any $v \in V$, $\exists!\ (b_1, \ldots, b_n) \in \mathbb{C}^n$ s.t. $v = b_1 v_1 + \cdots + b_n v_n$. Then $[v]_{\mathcal{B}} = (b_1, \ldots, b_n)$ is the coordinate of v w.r.t. the basis $\mathcal{B}$. The choice of a basis $\mathcal{B}$ gives an isomorphism $V \cong \mathbb{C}^n$, defined by the projection map

$$\begin{aligned} \pi_{\mathcal{B}} : V &\longrightarrow \mathbb{C}^n \\ v &\longmapsto [v]_{\mathcal{B}} \end{aligned}$$

Theorem 1.2. Let V be a vector space with $\dim V = n$. Then TFAE (the following are equivalent):

- $\mathcal{B} = \{v_1, \ldots, v_n\}$ is a basis[a] of V.
- $\mathcal{B}$ is a minimal spanning subset of V.
- $\mathcal{B}$ is a maximal linearly independent subset of V.
- For any $v \in V$, $\exists!\ (b_1, \ldots, b_n) \in \mathbb{C}^n$ s.t. $v = b_1 v_1 + \cdots + b_n v_n$.

[a] linearly independent spanning set

Let V, W be vector spaces with $\dim V = n$ and $\dim W = m$. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ and $\mathcal{B}' = \{w_1, \ldots, w_m\}$ be bases for V and W respectively. Then any $\mathbb{C}$-linear map $T : V \to W$ is determined by the images of the basis vectors,

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i \qquad \text{for } j = 1, \ldots, n \text{ and } a_{ij} \in \mathbb{C}.$$

Thus, we can associate to T the matrix $[T]_{\mathcal{B}}^{\mathcal{B}'} = (a_{ij})_{m \times n} \in \mathrm{Mat}_{m \times n}(\mathbb{C})$. This gives an isomorphism

$$\begin{aligned} \pi_{\mathcal{B}}^{\mathcal{B}'} : \mathrm{Hom}_{\mathbb{C}}(V, W) &\longrightarrow \mathrm{Mat}_{m \times n}(\mathbb{C}) \\ T &\longmapsto [T]_{\mathcal{B}}^{\mathcal{B}'} \end{aligned}$$

where $\mathrm{Hom}_{\mathbb{C}}(V,W)$ is the space of $\mathbb{C}$-linear maps from V to W. In particular, the space of endomorphisms is isomorphic to the space of square matrices of order n,

$$\mathrm{End}_{\mathbb{C}}(V) \coloneqq \mathrm{Hom}_{\mathbb{C}}(V,V) \cong \mathrm{Mat}_{n\times n}(\mathbb{C}).$$

The space of invertible linear operators on V is called the general linear group of V, $GL(V)$,

$$\mathrm{GL}(V) \coloneqq \{T \in \mathrm{End}_{\mathbb{C}}(V) : \ T \text{ is invertible}\} \cong \mathrm{GL}_n(\mathbb{C}) \coloneqq \mathrm{GL}(\mathbb{C}^n).$$

1.2 22 January

Theorem 1.3. Let U, V, W be vector spaces with $\dim U = p$, $\dim V = n$ and $\dim W = m$. Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be bases for U, V and W respectively and $U \xrightarrow{T_1} V \xrightarrow{T_2} W$. Then $T_2 \circ T_1 : U \to W$ is a linear map, and we have

$$[T_2 \circ T_1]^{\mathcal{C}}_{\mathcal{A}} = [T_2]^{\mathcal{C}}_{\mathcal{B}}[T_1]^{\mathcal{B}}_{\mathcal{A}}. \tag{1.1}$$

Proof. Let $T : V \to W$ be a linear map and $\mathcal{B}$ and $\mathcal{C}$ be bases for V and W respectively. Suppose $\mathcal{B} = \{v_1, \dots, v_n\}$ and $\mathcal{C} = \{w_1, \dots, w_m\}$. Then

$$[T(v)]_{\mathcal{C}} = [T]^{\mathcal{C}}_{\mathcal{B}}[v]_{\mathcal{B}}.$$

If $v = b_1v_1 + \cdots + b_nv_n$, then $T(v) = b_1T(v_1) + \cdots + b_nT(v_n)$. Since $T(v_j) = \sum_{i=1}^{m} a_{ij}w_i$, we have

$$T(v) = \sum_{j=1}^{n} b_jT(v_j) = \sum_{j=1}^{n} b_j \sum_{i=1}^{m} a_{ij}w_i = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij}b_j \right) w_i.$$

So, $[T(v)]_{\mathcal{C}}$ is the vector whose i-th component is $\sum_{j=1}^{n} a_{ij}b_j$, which is precisely the i-th component of $[T]^{\mathcal{C}}_{\mathcal{B}}[v]_{\mathcal{B}}$,

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_j \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} \vdots \\ \sum_{j=1}^{n} a_{ij}b_j \\ \vdots \end{pmatrix}.$$

Returning to the composition of linear maps (eq. (1.1)), we have for any $u \in U$,

$$(T_2 \circ T_1)(u) = T_2(T_1(u)) \implies [(T_2 \circ T_1)(u)]_{\mathcal{C}} = [T_2]^{\mathcal{C}}_{\mathcal{B}}[T_1(u)]_{\mathcal{B}} = [T_2]^{\mathcal{C}}_{\mathcal{B}}[T_1]^{\mathcal{B}}_{\mathcal{A}}[u]_{\mathcal{A}}.$$

Hence,

$$[T_2 \circ T_1]^{\mathcal{C}}_{\mathcal{A}}[u]_{\mathcal{A}} = [T_2]^{\mathcal{C}}_{\mathcal{B}}[T_1]^{\mathcal{B}}_{\mathcal{A}}[u]_{\mathcal{A}} \forall u \in U.$$

But $Av = Bv$ for all $v \in V$ implies that $A = B$, so we conclude that $[T_2 \circ T_1]^{\mathcal{C}}_{\mathcal{A}} = [T_2]^{\mathcal{C}}_{\mathcal{B}}[T_1]^{\mathcal{B}}_{\mathcal{A}}$. ∎

Corollary 1.4. Let $T \in GL(V)$ and $\mathcal{B}$ be a basis for V. Then $[T^{-1}]_{\mathcal{B}} = [T]^{-1}_{\mathcal{B}}$.

1.2.1 Change of basis matrix

Let $T \in \mathrm{End}_{\mathbb{C}}(V)$ and $\mathcal{B}$ and $\mathcal{B}'$ be two bases for V ($\dim V = n$). On changing the basis from $\mathcal{B}$ to $\mathcal{B}'$, the matrix representation of T changes from $[T]_{\mathcal{B}}$ to $[T]_{\mathcal{B}'}$. We want to find how these two matrices are related.

Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ and $\mathcal{B}' = \{v_1', \ldots, v_n'\}$. Since $\mathcal{B}$ is a basis, we can write

$$v_j' = c_{1j}v_1 + \cdots + c_{nj}v_n = \sum_{i=1}^{n} c_{ij}v_i \qquad \text{for } c_{ij} \in \mathbb{C}.$$

Let $P_{\mathcal{B}\leftarrow\mathcal{B}'} = (c_{ij})_{n\times n} \in \mathrm{Mat}_{n\times n}(\mathbb{C})$ be the change of basis matrix from $\mathcal{B}'$ to $\mathcal{B}$. Then we have

$$P_{\mathcal{B}\leftarrow\mathcal{B}'} = \begin{pmatrix} [v_1']_{\mathcal{B}} & [v_2']_{\mathcal{B}} & \cdots & [v_n']_{\mathcal{B}} \end{pmatrix}.$$

For $v \in V$, we have $v = b_1v_1' + \cdots + b_nv_n'$. So

$$v = \sum_{j=1}^{n} b_j v_j' = \sum_{j=1}^{n} b_j \sum_{i=1}^{n} c_{ij}v_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} c_{ij}b_j \right) v_i.$$

This means

$$P_{\mathcal{B}\leftarrow\mathcal{B}'}[v]_{\mathcal{B}'} = [v]_{\mathcal{B}}. \tag{1.2}$$

Similarly,

$$P_{\mathcal{B}'\leftarrow\mathcal{B}}[v]_{\mathcal{B}} = [v]_{\mathcal{B}'}. \tag{1.3}$$

We want to compare $[T]_{\mathcal{B}}$ and $[T]_{\mathcal{B}'}$. We will compute the coordinate vectors of the image w.r.t. both bases.

$$\begin{aligned} [T(v)]_{\mathcal{B}'} &= [T]_{\mathcal{B}'}[v]_{\mathcal{B}'} \\ [T(v)]_{\mathcal{B}'} &= P_{\mathcal{B}'\leftarrow\mathcal{B}}[T(v)]_{\mathcal{B}} \\ \implies [T]_{\mathcal{B}'}[v]_{\mathcal{B}'} &= P_{\mathcal{B}'\leftarrow\mathcal{B}}[T(v)]_{\mathcal{B}} \\ &= P_{\mathcal{B}'\leftarrow\mathcal{B}}[T]_{\mathcal{B}}[v]_{\mathcal{B}}. \end{aligned}$$

Now, using eq. (1.2) for $[v]_{\mathcal{B}}$, we have

$$[T]_{\mathcal{B}'}[v]_{\mathcal{B}'} = P_{\mathcal{B}'\leftarrow\mathcal{B}}[T]_{\mathcal{B}}P_{\mathcal{B}\leftarrow\mathcal{B}'}[v]_{\mathcal{B}'}. \tag{1.4}$$

Since eq. (1.4) holds for all $v \in V$, we conclude that

$$[T]_{\mathcal{B}'} = P_{\mathcal{B}'\leftarrow\mathcal{B}}[T]_{\mathcal{B}}P_{\mathcal{B}\leftarrow\mathcal{B}'}. \tag{1.5}$$

Now, using eq. (1.2) and eq. (1.3), we have

$$P_{\mathcal{B}'\leftarrow\mathcal{B}} = P_{\mathcal{B}\leftarrow\mathcal{B}'}^{-1}. \tag{1.6}$$

Hence, eq. (1.5) can be rewritten as

$$[T]_{\mathcal{B}'} = P^{-1}[T]_{\mathcal{B}}P, \tag{1.7}$$

where $P = P_{\mathcal{B}\leftarrow\mathcal{B}'}$. So $[T]_{\mathcal{B}}$ and $[T]_{\mathcal{B}'}$ are similar matrices.

This allows us to define the trace and determinant of a linear operator $T \in \mathrm{End}_{\mathbb{C}}(V)$ as the trace and determinant of any matrix representation of T. Since similar matrices have the same trace and determinant, this is well-defined. So, we have $\mathrm{tr}(T) = \mathrm{tr}([T]_{\mathcal{B}})$ and $\det(T) = \det([T]_{\mathcal{B}})$, both independent of the choice of basis.

1.3 2 February

1.3.1 Direct sum decompositions

As internal and external direct sums are isomorphic, we will only talk about internal direct sums.

Theorem 1.5. Let V be a vector space and $U, W \subseteq V$ be subspaces of V. TFAE:

- For any $v \in V$, $\exists! u \in U$ and $w \in W$ s.t. $v = u + w$. We then say that V is the internal direct sum of U and W, denoted by $V = U \oplus W$.
- $V = U + W$ and $U \cap W = \{0\}$.
- $V = U + W$ and $\dim V = \dim U + \dim W$.
- External direct sum is isomorphic to internal direct sum:

$$\begin{aligned} \phi : U \times W &\xrightarrow{\sim} V \\ (u, w) &\mapsto u + w \end{aligned}$$

Theorem 1.6. $\mathcal{B}_1$ is a basis of U and $\mathcal{B}_2$ is a basis of $W \implies \mathcal{B}_1 \sqcup \mathcal{B}_2$ is a basis of $U \oplus W$.

1.3.2 Inner product spaces

Definition 1.7 (Inner product space)**.** An inner product on a $\mathbb{C}$-vector space V is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ such that

- $\langle \lambda v_1 + \mu v_2, w \rangle = \lambda \langle v_1, w \rangle + \mu \langle v_2, w \rangle$ (linearity in the first argument)
- $\langle v, w \rangle = \overline{\langle w, v \rangle}$ (conjugate symmetry)
- $\langle v, v \rangle \geq 0$ with equality iff $v = 0$ (positive-definiteness)

for all $v, v_1, v_2, w \in V$ and $\lambda, \mu \in \mathbb{C}$. $(V, \langle \cdot, \cdot \rangle)$ is called an inner product space.

There are some natural examples of inner product spaces coming from the standard inner product on $\mathbb{C}^n$.

Example 1.8. On $\mathbb{C}^n$, we have the inner product defined by $\langle a, b \rangle = \sum_{k=1}^{n} a_k \bar{b}_k$ where $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$. This is called the *standard inner product* on $\mathbb{C}^n$.

We will be working with a slightly different space, which is the free vector space on a finite set. Start with a finite set X. Consider all maps from X to $\mathbb{C}$, denoted by $\mathbb{C}^X$. Then $\mathbb{C}^X$ is a vector space under pointwise addition and scalar multiplication. It is easy to see that under these operations, $\mathbb{C}^X$ is a vector space. For any $x \in X$, let $\delta_x : X \to \mathbb{C}$ be the function defined by

$$\delta_x(y) = \begin{cases} 1 & \text{if } y = x \\ 0 & \text{if } y \neq x \end{cases}$$

Then $\mathcal{B}_x = \{\delta_x : x \in X\}$ forms a basis for $\mathbb{C}^X$. So, $\dim \mathbb{C}^X = |X|$.

Example 1.9. We can define an inner product on $\mathbb{C}^X$ by

$$\langle f, g \rangle = \sum_{x \in X} f(x) \overline{g(x)}.$$

This is called the *standard inner product* on $\mathbb{C}^X$. With this inner product, $\mathcal{B}_x$ is an orthonormal basis for $\mathbb{C}^X$.

Recall that *orthonormal* = *orthogonal* + unit vector.

- $u, v \in V$ are *orthogonal* if $\langle u, v \rangle = 0$.

- $S \subseteq V$ is an *orthogonal* if any two distinct vectors in S are orthogonal.

Theorem 1.10. Let $S \subseteq V$, $0 \notin S$ and S is orthogonal. Then S is linearly independent.

Using Gram-Schmidt process, we can always find an orthonormal basis for a finite-dim. inner product space.

Theorem 1.11. Every finite-dimensional inner product space has an orthonormal basis.

The advantage of working with an orthonormal basis is that it is easy to write down the coordinates w.r.t. the orthonormal basis. Let $\mathcal{B} = \{e_1, \ldots, e_n\}$ be an orthonormal basis for V. Then for any $v \in V$, we have

$$v = \sum_{k=1}^{n} \langle v, e_k \rangle e_k, \qquad [v]_{\mathcal{B}} = \begin{pmatrix} \langle v, e_1 \rangle & \ldots & \langle v, e_n \rangle \end{pmatrix}.$$

Definition 1.12 (Norm). The *norm* of a vector $v \in V$ is defined by $\|v\| = \sqrt{\langle v, v \rangle}$.

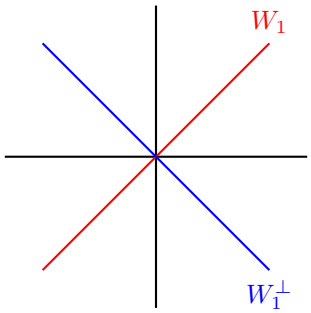
Theorem 1.13 (Triangle inequality). For any $v, w \in V$, we have

$$\|v + w\| \leq \|v\| + \|w\|. \tag{1.8}$$

Theorem 1.14 (Cauchy-Schwarz inequality). For any $v, w \in V$, we have

$$|\langle v, w \rangle| \leq \|v\| \, \|w\|. \tag{1.9}$$

What is the advantage of working in an inner product space ? For example, let's say we're in a 2-dimensional space V. Given a subspace W_1 of V, there will be many complementary subspaces of W_1. We say W_2 is a complement of W_1 if $V = W_1 \oplus W_2$.



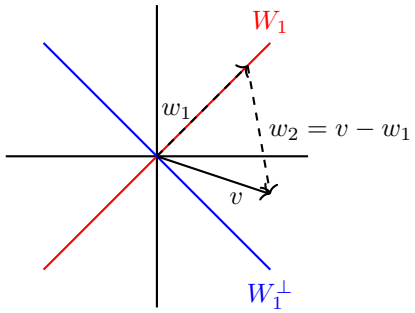
If we have an inner product on V, then we can take the orthogonal complement of W_1, denoted by $W_1^\perp$, as a complement of W_1. So, we have a canonical choice of complement for any subspace of V. In the absence of an inner product, we have no canonical choice of complement for a given subspace of V.

Theorem 1.15. Let V be an inner product space and $W_1 \subseteq V$ be a subspace of V and

$$W_1^\perp = \{v \in V : \langle v, w \rangle = 0 \text{ for all } w \in W_1\}.$$

Then $W_2 = W_1^\perp$ is a complement of W_1, i.e. $V = W_1 \oplus W_1^\perp$.

Proof. We need to check that $W_1 \cap W_1^\perp = \{0\}$ and $W_1 + W_1^\perp = V$ which is easy, because if $v \in W_1 \cap W_1^\perp$, then $\langle v, v \rangle = 0 \implies v = 0$. So, $W_1 \cap W_1^\perp = \{0\}$. For any $v \in V$, we want to write v as $v = w_1 + w_2$ where $w_1 \in W_1$ and $w_2 \in W_1^\perp$. To make this choice we can take w_1 to be the *orthogonal projection* of v onto W_1,



We can write w_1 as a linear combination of the orthonormal basis vectors of W_1, say $\mathcal{B} = \{e_1, \dots, e_k\}$, so that

$$w_1 = \hat{v} = \sum_{i=1}^{k} \langle v, e_i \rangle e_i.$$

Now, $\langle w_2, e_i \rangle = \langle v, e_i \rangle - \langle w_1, e_i \rangle = \langle v, e_i \rangle - \langle \hat{v}, e_i \rangle = \langle v, e_i \rangle - \langle v, e_i \rangle = 0$. So $w_2 = v - w_1 \in W_1^\perp$ as required. ■

1.3.3 Unitary operators

Definition 1.16 (Unitary operator)**.** Let V be an inner product space. A linear operator $T : V \to V$ is called *unitary* if $\langle T(v), T(w) \rangle = \langle v, w \rangle$ for all $v, w \in V$.

Theorem 1.17. If $T \in \mathrm{End}_{\mathbb{C}}(V)$ is unitary, then $T \in \mathrm{GL}(V)$.

Proof. If T is unitary and $v \in \ker T$, then

$$0 = \langle T(v), T(v) \rangle = \langle v, v \rangle \implies v = 0 \implies \ker T = \{0\}.$$

Hence, $T \in \mathrm{GL}(V)$. ■

Theorem 1.18. The set $U(V)$ of all unitary operators on V is a subgroup of $\mathrm{GL}(V)$.

Definition 1.19 (Transpose and conjugate transpose)**.** Let $A = (a_{ij}) \in \mathrm{Mat}_n(\mathbb{C})$. Then $A^T = (a_{ji})$ is called the *transpose* of A and $A^* = \overline{A^T} = (\overline{a}_{ji})$ is called the *conjugate transpose* of A.

Theorem 1.20.

1. Let $A, B \in \mathrm{Mat}_n(\mathbb{C})$, then $(AB)^* = B^* A^*$.
2. For $v, w \in \mathbb{C}^n$, $\langle Av, w \rangle = \langle v, A^* w \rangle$ is the standard inner product on $\mathbb{C}^n$.
3. $A \in \mathrm{GL}_n(\mathbb{C})$ is unitary iff $A^{-1} = A^*$.
   We then say the matrix A is unitary and denote by $U_n(\mathbb{C}) = U(\mathbb{C}^n)$ the group of all $n \times n$ unitary matrices.

Definition 1.21 (Self-adjoint and symmetric matrices)**.** A matrix $A \in \mathrm{Mat}_n(\mathbb{C})$ is *self-adjoint* if $A = A^*$ and *symmetric* if $A = A^T$.

Clearly $A \in \mathrm{Mat}_n(\mathbb{R})$ is symmetric iff A is self-adjoint.

1.3.4 Adjoint of a linear operator

Let V be a finite-dim. $\mathbb{C}$-inner product space. Given any $T \in \mathrm{End}_{\mathbb{C}}(V)$, there exists a unique $T^* \in \mathrm{End}_{\mathbb{C}}(V)$ such that $\langle T(v), w\rangle = \langle v, T^*(w)\rangle$ for all $v, w \in V$. We call T^* the *adjoint* of T. Then T is self-adjoint iff $T = T^*$.

Theorem 1.22. Let $T \in \mathrm{End}_{\mathbb{C}}(V)$ and $\mathcal{B}$ be an orthonormal basis for V. Then $[T^*]_{\mathcal{B}} = [T]^*_{\mathcal{B}}$.

Question. Given $A \in \mathrm{Mat}_n(\mathbb{C})$ how can we compute $\det(A)$, A^k for $k \geq 1$, $\exp(A)$?

A acts on $\mathbb{C}^n$ via the linear transformation $x \xmapsto{T_A} Ax$. Let $\mathcal{B}_{\text{std}} = \{e_1, \ldots, e_n\}$ be the standard basis of $\mathbb{C}^n$. Then

$$\boxed{[T_A]_{\mathcal{B}_{\text{std}}} = A}.$$

If we change the basis $\mathcal{B}_{\text{std}} \rightsquigarrow \mathcal{B}$, then by eq. (1.7), we have $[T_A]_{\mathcal{B}} \sim A$, i.e. $\exists P \in \mathrm{GL}_n(\mathbb{C})$ such that

$$\boxed{P[T_A]_{\mathcal{B}}P^{-1} = A},$$

where P is the change of basis matrix from $\mathcal{B}$ to $\mathcal{B}_{\text{std}}$. Now,

- $\det A = \det [T_A]_{\mathcal{B}}$.
- $A^k = P[T_A]^k_{\mathcal{B}}P^{-1}$, for all $k \geq 1$.
- $\exp(A) = \displaystyle\sum_{k=0}^{\infty} \frac{A^k}{k!}$.

The simplest class of matrices are the diagonal matrices. If A is diagonal, then $\det A$, A^k and $\exp(A)$ are easy to compute. So, if we can find a basis $\mathcal{B}$ such that $[T_A]_{\mathcal{B}}$ is diagonal, then we can easily compute $\det A$, A^k and $\exp(A)$. This motivates us to study the problem of diagonalisation of linear operators.

1.4 9 February

We want nice representatives for the similarity class of matrices.

- Diagonal matrix
- Block diagonal matrices
  - elementary Jordan matrices
  - companion matrices

1.4.1 Elementary Jordan matrix

Let $\mathbb{F}$ be a field and $\lambda \in \mathbb{F}$. Then the following matrices

$$(\lambda)\,,\ \begin{pmatrix}\lambda & 1\\ 0 & \lambda\end{pmatrix},\ \begin{pmatrix}\lambda & 1 & 0\\ 0 & \lambda & 1\\ 0 & 0 & \lambda\end{pmatrix},\ \begin{pmatrix}\lambda & 1 & 0 & 0\\ 0 & \lambda & 1 & 0\\ 0 & 0 & \lambda & 1\\ 0 & 0 & 0 & \lambda\end{pmatrix}\ \cdots \tag{1.10}$$

with diagonal λ and superdiagonal 1, or

$$(\lambda)\,,\ \begin{pmatrix}\lambda & 0\\ 1 & \lambda\end{pmatrix},\ \begin{pmatrix}\lambda & 0 & 0\\ 1 & \lambda & 0\\ 0 & 1 & \lambda\end{pmatrix},\ \begin{pmatrix}\lambda & 0 & 0 & 0\\ 1 & \lambda & 0 & 0\\ 0 & 1 & \lambda & 0\\ 0 & 0 & 1 & \lambda\end{pmatrix}\ \cdots \tag{1.11}$$

are called the *elementary Jordan matrices* corresponding to λ of order $1,\ 2,\ 3,\ \ldots$ respectively.

Example 1.23.

$$\begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

is an upper triangular matrix.

1.4.2 Companion matrix

Let $\mathbb{F}$ be a field and $f(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n \in \mathbb{F}[x]$ be a monic polynomial. The $n \times n$ matrix

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & & & & \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix}_{n\times n}, \text{ or } \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & a_{n-2} & -a_{n-1} \end{pmatrix}_{n\times n}$$

is called the *companion matrix* corresponding to $f(x)$.

1.4.3 Diagonalisation

Not all matrices are diagonalisable, but one can approximate any given matrix by diagonalisable matrices.

Theorem 1.24. The set of diagonalisable matrices is dense in $\mathrm{Mat}_n(\mathbb{C})$.

So there is a sequence of diagonalisable matrices $\{A_n\}_{n\in\mathbb{N}}$ such that $A_n \to A$.

Then $\det A_n \to \det A$, $A_n^k \to A^k$ and $\exp(A_n) \to \exp(A)$ as $n \to \infty$. The red sequences are easy to compute as the A_n's are diagonalisable, so we can approximate the limits by the respective sequences.

Definition 1.25 (Diagonalisable). $T \in \mathrm{End}_{\mathbb{C}}(V)$ is *diagonalisable* if there exists a basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}$ is diagonal.

Theorem 1.26. Let $A \in \mathrm{Mat}_n(\mathbb{C})$. Then $\exists P \in \mathrm{GL}_n(\mathbb{C})$ such that PAP^{-1} is upper triangular.

So the best we can say about any matrix is that it is similar to an upper triangular matrix. But this is not really much better for computational purposes. A matrix has n^2 entries, while an upper triangular matrix has

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2} = \mathrm{O}(n^2)$$

which means that for all n sufficiently large computing with upper triangular matrices is no better than computing with general matrices. So we need something simpler than upper triangular matrices when we can't get a diagonal matrix. This is the motivation for the Jordan canonical form.

1.4.4 Jordan canonical form

Definition 1.27 (Jordan block). A Jordan block of size k with eigenvalue λ is the $k \times k$ matrix

$$\mathcal{J}_k(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix}. \tag{1.12}$$

Remark. We could also have defined the Jordan blocks as lower triangular instead of upper triangular matrices, it doesn't matter. For a Jordan block of size n the number of entries is $n+(n-1)=2n-1=\mathrm{O}(n)$. This grows linearly with n, which is much better than the $\mathrm{O}(n^2)$ growth of upper triangular matrices. So, Jordan blocks are much better for computing.

How do we even expect to achieve something as nice as Jordan blocks ? Why does there exist a basis $\mathcal{B}$ such that PAP^{-1} "looks nice" ? All we are looking for is a coordinate system in which the matrix of the given linear transformation T looks simpler, and to achieve that coordinate system we decompose the given vector space as T-invariant subspaces.

Definition 1.28. [Invariant subspace and indecomposable subspace] A subspace $W \subseteq V$ is called T-*invariant* if $T(W) \subseteq W$.

W is said to be *indecomposable* if $W = W_1 \oplus W_2$ for some T-invariant subspaces $W_1, W_2 \subseteq W$, then $W_1 = \{0\}$ or $W_2 = \{0\}$.

Theorem 1.29. If $T \in \mathrm{End}_{\mathbb{C}}(V)$ for a finite-dim. vector space V, then there V is a direct sum of indecomposable T-invariant subspaces,

$$\underset{\mathcal{B}}{V} = \underset{\mathcal{B}_1}{V_1} \oplus \cdots \oplus \underset{\mathcal{B}_k}{V_k}, \qquad \text{where } V_i \text{ are indecomposable.} \tag{1.13}$$

Proof. If $\dim V = 1$ then V is indecomposable and we're done. So let $\dim V > 1$. If V is indecomposable then there's nothing to prove. Otherwise, $V = W_1 \oplus W_2$ for some T-invariant subspaces $W_1, W_2 \subsetneq V$ and $W_1, W_2 \neq 0$. Hence, $1 \leq \dim W_i < \dim V$ for $i = 1, 2$. We can restrict T to W_i, i.e. $T\big|_{W_i} : W_i \to W_i$, for $i = 1, 2$. By induction on the dimension of the vector space, we can write W_i as a direct sum of indecomposable T-invariant subspaces of W_i, hence T-invariant subspaces of V, for $i = 1, 2$. Putting these together, we get a decomposition of $V = W_1 \oplus W_2$ into indecomposable T-invariant subspaces as required. ∎

$$[T]_{\mathcal{B}} = \begin{pmatrix} \boxed{[T\big|_{V_1}]_{\mathcal{B}_1}} & 0 & \cdots & 0 \\ 0 & \boxed{[T\big|_{V_2}]_{\mathcal{B}_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boxed{[T\big|_{V_k}]_{\mathcal{B}_k}} \end{pmatrix}$$

is a block decomposition of $[T]_{\mathcal{B}}$. In order to achieve the decomposition our first objective is to find T-invariant subspaces of V.

1.4.5 T-invariant subspaces

Recall the definition of T-invariant subspace from definition 1.28.

Theorem 1.30. Fix $T \in \mathrm{End}_{\mathbb{C}}(V)$ where $\dim V < \infty$. If $S \in \mathrm{End}_{\mathbb{C}}(V)$ such that $ST = TS$, then

- $\ker S$ is T-invariant.
- $V_S = \{v \in V : S^r v = 0 \text{ for some } r \in \mathbb{N}_0\}$ is T-invariant.

Proof.

- If $v \in \ker S$, then $(ST)(v) = (TS)(v) = 0 \implies S(T(v)) = 0$. So, $T(v) \in \ker S$.

- If $v \in V_S$, then $S^r(v) = 0$ for some $r \in \mathbb{N}_0$. Then, as $ST = TS$,

$$S^r(T(v)) = T(S^r(v)) = 0 \implies T(v) \in V_S.$$

■

How do we get an operator that commutes with a given operator T ? We can take $T, T^2, \ldots$, i.e. the algebra generated by T inside $\mathrm{End}_{\mathbb{C}}(V)$, which is the set $\mathbb{C}[T]$ of polynomials in T and forms a subalgebra of $\mathrm{End}_{\mathbb{C}}(V)$. Any operator in $\mathbb{C}[T]$ commutes with T.

The simplest polynomial is the linear polynomial $S = T - \lambda I$. Then $\ker S$ is T-invariant. We are not interested in the case $\ker S = \{0\}$. This motivates us to look for $\lambda \in \mathbb{C}$ such that $\ker T - \lambda I \neq \{0\}$, i.e. $T - \lambda I$ is not injective, i.e. $T - \lambda I$ is not invertible, i.e. $\det(T - \lambda I) = 0$. So, we want to find $\lambda \in \mathbb{C}$ such that $\det(T - \lambda I) = 0$. This leads us to define the characteristic polynomial of T.

Definition 1.31 (Characteristic polynomial, eigenvalues, eigenspaces)**.** The *characteristic polynomial* of $T \in \mathrm{End}_{\mathbb{C}}(V)$ is defined by $\chi_T(x) = \det(xI - T)$. The roots of $\chi_T(x)$ are called the *eigenvalues* of T. If λ is an eigenvalue of T, then the *eigenspace* of T corresponding to λ is defined by

$$E_\lambda = \ker(T - \lambda I) = \{v \in V : (T - \lambda I)(v) = 0\},$$

which is the space of all *eigenvectors* of T corresponding to λ together with the zero vector.

We can generalise this notion of eigenspaces.

Definition 1.32 (Generalised eigenspace)**.** Let $T \in \mathrm{End}_{\mathbb{C}}(V)$ and λ be an eigenvalue of T. The *generalised eigenspace* of T corresponding to λ is defined by

$$V_\lambda = \{v \in V : (T - \lambda I)^r(v) = 0 \text{ for some } r \in \mathbb{N}_0\}.$$

Theorem 1.33. $V_\lambda = \ker (T - \lambda I)^{m_\lambda}$, where m_λ is the multiplicity of λ as a root of $\chi_T(x)$.

We have the $\mathbb{C}$-algebra homomorphism

$$\begin{aligned} \pi_T : \mathbb{C}[x] &\longrightarrow \mathrm{End}_{\mathbb{C}}(V) \\ x &\longmapsto T \\ 1 &\longmapsto I \end{aligned}$$

where the image is $\mathrm{im}\, \pi_T = \mathbb{C}[T]$. Also $\ker \pi_T \trianglelefteq \mathbb{C}[x]$, and $\ker \pi_T$ generates the unique monic polynomial $m_T(x)$ called the *minimal polynomial* of T.

Theorem 1.34. $T \in \mathrm{End}_{\mathbb{C}}(V)$ is diagonalisable/semisimple iff the minimal polynomial $m_T(x)$ has no repeated roots.

1.5 23 March

1.5.1 Jordan-Chevalley decomposition

Definition 1.35 (Semisimple and nilpotent operators)**.** Let $T \in \mathrm{End}_{\mathbb{C}}(V)$ where $\dim V < \infty$. Then T is

- *semisimple* if T is diagonalisable.
- *nilpotent* if $T^r = 0$ for some $r \in \mathbb{N}$.

Theorem 1.36 (Jordan-Chevalley decomposition). Let $T \in \mathrm{End}_{\mathbb{C}}(V)$ where $\dim V < \infty$. Then

1. $\exists!\ T_s, T_n \in \mathrm{End}_{\mathbb{C}}(V)$ s.t.
$$T = T_s + T_n,$$
where T_s is semisimple, T_n is nilpotent, and T_s and T_n commute.
2. $\exists\ p(x), q(x) \in \mathbb{C}[x]$ s.t. $T_s = p(T)$, $T_n = q(T)$ (both T_s and T_n commute with T).

Proof. Let the minimal polynomial of T be

$$m_T(x) = \prod_{i=1}^{k} (x - \lambda_i)^{m_i}$$

where m_i is the multiplicity of λ_i in $m_T(x)$. Then for each λ_i we have the corresponding generalised eigenspaces

$$V_i = V_{\lambda_i} = \ker (T - \lambda_i I)^{m_i}.$$

Claim (*Primary decomposition theorem*): V is a direct sum of these generalised eigenspaces,

$$V = V_1 \oplus \cdots \oplus V_k. \tag{1.14}$$

This uses the fact that the λ_i's are distinct - in particular, that $\dfrac{m_T(x)}{(x-\lambda_i)^{m_i}}$ and $(x - \lambda_i)^{m_i}$ are relatively prime.

Now, we apply the Chinese remainder theorem to locate the polynomial $p(x) \in \mathbb{C}[x]$ such that

$$p(x) \equiv \lambda_i \mod (x - \lambda_i)^{m_i} \qquad \text{for } i = 1, 2, \ldots, k$$

which implies that $p(x) \equiv 0 \mod x$. Set $q(x) = x - p(x)$, then $x = p(x) + q(x)$. Then $T_s = p(T)$ is semisimple and $T_n = q(T)$ is nilpotent. Hence

$$T = T_s + T_n \qquad \text{where } T, T_s, T_n \text{ commute with each other.}$$

First, observe that $p(x) - \lambda_i =$ some multiple of $(x - \lambda_i)^{m_i}$, i.e.

$$p(T) - \lambda_i I = g(T)(T - \lambda_i I)^{m_i}$$

where the RHS is 0 on $V_i = \ker (T - \lambda_i)^{m_i}$. Hence,

$$T_s\Big|_{V_i} = \lambda_i I.$$

T_s is semisimple as $V = V_1 \oplus \cdots \oplus V_k$ by eq. (1.14). Hence,

$$[T_s]_{\mathcal{B}} = \begin{pmatrix} \boxed{\begin{matrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_1 \end{matrix}} & \cdots & 0 & \cdots & 0 \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \boxed{\begin{matrix} \lambda_k & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_k \end{matrix}} \end{pmatrix}$$

Now, $T_n\Big|_{V_i} = T\Big|_{V_i} - p(T)\Big|_{V_i} = T\Big|_{V_i} - \lambda_i I\Big|_{V_i}$, as $p(T) \equiv \lambda_i I$ on V_i, so this implies that

$$T_n^{m_i} = 0 \text{ on } V_i.$$

Hence, T_n is nilpotent. Now, let S be semisimple and N be nilpotent such that

$$T = S + N = T_s + T_n \implies T_s - S = N - T_n$$

where the LHS is semisimple and the RHS is nilpotent. Now, $[S, N] = 0 \implies [S, T] = 0$ where $[a, b] = ab - ba$ is the commutator. Hence, T_s and S commute, T_n and N commute. ■

1.5.2 Primary decomposition theorem

Theorem 1.37 (Primary decomposition theorem). Let $T \in \mathrm{End}_{\mathbb{C}}(V)$, $\dim_{\mathbb{C}}(V) < \infty$ and

$$m_T(x) = \prod_{i=1}^{k} (x - \lambda_i)^{m_i}, \qquad V_i = \ker (T - \lambda_i I)^{m_i}$$

then

$$V = V_1 \oplus \cdots \oplus V_k. \tag{1.15}$$

Theorem 1.38. T is diagonalisable iff $m_T(x)$ has distinct roots.

Proof. $\Longrightarrow$: T is diagonalisable $\implies \exists \mathcal{B} \subseteq V$ s.t.

$$[T]_{\mathcal{B}} = \begin{pmatrix} \mu_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mu_n \end{pmatrix}$$

and the eigenvalue λ_i appears with multiplicity ℓ_i for $i = 1, \ldots, k$. Then

$$\chi_T(x) = \prod_{i=1}^{k} (x - \lambda_i)^{\ell_i}, \; m_T(x) = \prod_{i=1}^{k} (x - \lambda_i)$$

$$[T]_{\mathcal{B}'} = \begin{pmatrix} \boxed{\begin{matrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_1 \end{matrix}}_{\ell_1 \times \ell_1} & & \\ & \ddots & \\ & & \boxed{\begin{matrix} \lambda_k & & \\ & \ddots & \\ & & \lambda_k \end{matrix}}_{\ell_k \times \ell_k} \end{pmatrix}$$

$\Longleftarrow$: Given the minimal polynomial

$$m_T(x) = (x - \lambda_1) \cdots (x - \lambda_k),$$

λ_i's are distinct with corresponding (generalised) eigenspaces $V_i = \ker(x - \lambda_i I)$. Then

$$\underset{\mathcal{B}}{V} = \underset{\mathcal{B}_1}{V_1} \oplus \cdots \oplus \underset{\mathcal{B}_k}{V_k}, \qquad \boxed{T\big|_{V_i} = \lambda_i I}.$$

Hence, the matrix

$$[T]_{\mathcal{B}} = \begin{pmatrix} \boxed{\lambda_1 I_{V_1}} & & \\ & \ddots & \\ & & \boxed{\lambda_k I_{V_k}} \end{pmatrix}$$

is diagonal. ■

Theorem 1.39. If $T \in \mathrm{End}_{\mathbb{C}}(V)$ s.t. $T^N = I_V$ for some N, then T is diagonalisable.

Theorem 1.40. If $\gcd(f'(x), f(x)) = 1$, then f has no repeated roots.
($f(x) = x^N - 1 \implies f'(x) = Nx^{N-1}$.)

Question. Given $T \in \mathrm{End}_{\mathbb{C}}(V)$, does there exist a basis $\mathcal{B}$ of V s.t. $[T]_{\mathcal{B}}$ looks "somewhat simpler" ? In other words, we want to find a representative for the equivalence class of similar matrices

$$\boxed{\{[T]_{\mathcal{B}} : \mathcal{B} \text{ is a basis of } V\}}$$

1.5.3 Cyclic decomposition theorem

Let $T \in \mathrm{End}_{\mathbb{C}}(V)$, $S \subseteq V$, and let the T-invariant subspace generated by S be

$$\langle S \rangle = \bigcap_{S \subseteq W \subseteq V,\ T-\text{invariant}} W.$$

For $v \in V$, $\langle v \rangle = W$. Then $T(v) \in W, T^2(v) \in W, \ldots\ T^k(v) \in W$ for $k \geq 1$. Now, we have the T-invariant subspace

$$\mathrm{span}_{\mathbb{C}} \left\{ T^k(v),\ k \geq 1 \right\} \subseteq W.$$

Definition 1.41 (Cyclic subspace). Let $v \in V$. Then the cyclic subspace of V generated by v is

$$Z_T(v) = \mathrm{span}_{\mathbb{C}} \left\{ T^k(v),\ k \geq 1 \right\}.$$

Theorem 1.42 (Cyclic decomposition theorem). Given $T \in \mathrm{End}_{\mathbb{C}}(V)$, $\exists$ nonzero vectors $v_1, \ldots, v_k \in V \setminus \{0\}$ such that

$$V = Z_T(v_1) \oplus \cdots \oplus Z_T(v_k).$$

1.6 13 April

1.6.1 Modules

Let R be a ring (with identity) and let M be an abelian group $(M, +)$. We say M is a (left) R-*module* if there is a scalar multiplication

$$R \times M \longrightarrow M, \qquad (\lambda, v) \longmapsto \lambda v,$$

satisfying, for all $v, v_1, v_2 \in M$ and $\lambda, \mu \in R$,

$$\lambda(v_1 + v_2) = \lambda v_1 + \lambda v_2, \tag{1.16}$$
$$(\lambda + \mu)v = \lambda v + \mu v, \tag{1.17}$$
$$(\lambda\mu)v_1 = \lambda(\mu v_1), \tag{1.18}$$
$$1 \cdot v = v. \tag{1.19}$$

Remark 1.43. A module is precisely the generalisation obtained by replacing the field $\mathbb{F}$ in the definition of a vector space by an arbitrary ring R. This is exactly what makes the difference when $\mathbb{F}$ is replaced by a ring R: a vector space is an $\mathbb{F}$-module where $\mathbb{F}$ is a field, while modules over general rings need not have bases, dimensions, etc.

1.6.2 The $\mathbb{F}[x]$-module structure induced by an operator

Suppose V is a vector space over a field $\mathbb{F}$ and T is a linear operator on V. Define

$$\mathbb{F}[x] \times V \longrightarrow V, \qquad \big(f(x), v\big) \longmapsto f(x) \cdot v := f(T)(v).$$

Proposition 1.44. The above operation makes V into an $\mathbb{F}[x]$-module.

(The verification of the module axioms is a direct, routine check using the fact that $T \mapsto f(T)$ respects sums and products of polynomials.)

We say that V is an $\mathbb{F}[x]$-*module via* T.

In particular, for a scalar $\lambda \in \mathbb{F} \subset \mathbb{F}[x]$ (viewed as a constant polynomial $f(x) = \lambda$), we have

$$\lambda v = T(v).$$

Indeed, if $f(x) = \lambda$ then $f(T) = \lambda I$, so

$$\lambda v = f(x) \cdot v = f(T)(v) = \lambda I(v) = \lambda v,$$

which is consistent: the scalar multiplication of the $\mathbb{F}$-vector-space structure on V is recovered as the restriction of the $\mathbb{F}[x]$-module structure to constant polynomials, and the action of x itself is exactly T:

$$x \cdot v = T(v).$$

So V now carries *two* compatible structures:

$$\mathbb{F} \times V \to V,\ (\lambda, v) \mapsto \lambda v \quad \text{(vector space structure)},$$
$$\mathbb{F}[x] \times V \to V,\ (f(x), v) \mapsto f(x)v := f(T)(v) \quad \text{(module structure)}.$$

1.6.3 What are the submodules?

Given the dictionary $xv = T(v)$, $\lambda v = f(T)(v)$ where $f(x) = \lambda$, the natural question is: *what do submodules of the $\mathbb{F}[x]$-module V (via T) look like, in vector-space language?* The following dictionary answers this.

Vector space language	**Module language**
$\mathbb{F} \times V \to V,\ (\lambda, v) \mapsto \lambda v$	$\mathbb{F}[x] \times V \to V,\ (f(x), v) \mapsto f(x)v := f(T)(v)$
(i) Invariant subspace	(i) Submodule
(ii) T-cyclic subspace generated by v_0	(ii) Cyclic submodule generated by $v_0 \in V$: $\{f(x)v_0 : f(x) \in \mathbb{F}[x]\} = \langle v_0 \rangle$ the *smallest* submodule containing v_0.
(iii) Order/Annihilator of v_0	(iii) $\mathrm{Ann}_{\mathbb{F}[x]}(v_0)$, the annihilator of v_0 in $\mathbb{F}[x]$

Definition 1.45 (Invariant subspace vs. submodule)**.** Let $W \leq V$ be a subspace.

- W is a *T-invariant subspace* if $T(w) \in W$ for all $w \in W$. Equivalently,
$$xw \in W \quad \text{for all } w \in W.$$
- W is a *submodule* of the $\mathbb{F}[x]$-module V if
$$f(x)w \in W \quad \text{for all } w \in W,\ f(x) \in \mathbb{F}[x].$$

Remark 1.46. The two notions above coincide: a subspace is T-invariant if and only if it is a submodule of V viewed as an $\mathbb{F}[x]$-module via T. Indeed, T-invariance gives $xw \in W$ for $w \in W$; iterating shows $x^k w \in W$ for every $k \geq 0$, and taking linear combinations shows $f(x)w \in W$ for every polynomial $f(x) \in \mathbb{F}[x]$.

1.6.4 $\mathbb{Z}$-modules and abelian groups

Let $(G, +)$ be an abelian group. Define

$$\mathbb{Z} \times G \longrightarrow G, \qquad (n, a) \longmapsto na := \begin{cases} \underbrace{a + \cdots + a}_{n \text{ times}}, & n > 0, \\ \underbrace{(-a) + \cdots + (-a)}_{|n| \text{ times}}, & n < 0, \\ 0, & n = 0. \end{cases}$$

This makes G into a $\mathbb{Z}$-module, and

$$\mathrm{Ann}_{\mathbb{Z}}(a) = \{n \in \mathbb{Z} : na = 0\} = \langle m \rangle$$

for some $m \geq 0$ (a left ideal of $\mathbb{Z}$, hence principal).

Remark 1.47. A module is a simultaneous generalisation of a vector space and of an abelian group: every abelian group is automatically a $\mathbb{Z}$-module via the action above, and every vector space over $\mathbb{F}$ is an $\mathbb{F}$-module.

Example 1.48. $\mathbb{Z}_n$ is a $\mathbb{Z}$-module. In the $\mathbb{Z}$-module $\mathbb{Z}_6$ we have $6v = 0$ for every $v \in \mathbb{Z}_6$, yet $6 \neq 0$ (in $\mathbb{Z}$) and possibly $v \neq 0$ (in $\mathbb{Z}_6$). Such an element v is said to be a *torsion element.*

1.6.5 Torsion elements

Definition 1.49. Let M be a left R-module and let $m \in M$. We say m is a *torsion element* if there exists $r\ (\neq 0) \in R$ such that $rm = 0$.

For a fixed $m \in M$, the set
$$\{r \in R : rm = 0\}$$
(which is nonempty, since $0 \cdot m = 0$) is called the *annihilator of m in R*, denoted $\mathrm{Ann}_R(m)$.

Exercise 1.50. $\mathrm{Ann}_R(m) \trianglelefteq R$, i.e. $\mathrm{Ann}_R(m)$ is a left ideal of R.

Definition 1.51. Let $v_0 \in V$. The *T-annihilator of v_0* (also called *the T-annihilator of v_0*, cf.[HK70]) is

$$\mathrm{Ann}_{\mathbb{F}[x]}(v_0) = \{f(x) \in \mathbb{F}[x] : f(x)v_0 = 0\} = \{f(x) \in \mathbb{F}[x] : f(T)(v_0) = 0\}.$$

This is a (two-sided) ideal of $\mathbb{F}[x]$. Any generator of this ideal is called *an order* of v_0. The ideal has many generators, but it has one and only one *monic* generator, which is called *the order* of v_0.

Remark 1.52. If $m(x)$ is the minimal polynomial of T, then

$$m(x)v = m(T)(v) = 0 \qquad \text{for every } v \in V.$$

1.6.6 Torsion Modules

Definition 1.53. An R-module M is said to be a *torsion module* if every element of M is a torsion element; i.e. for every $m \in M$ there exists $r\ (\neq 0) \in R$ (depending on m) such that $rm = 0$.

Example 1.54.

(i) The $\mathbb{Z}$-module $\mathbb{Z}_n$ is a torsion module.

(ii) The $\mathbb{Z}$-module G, where $(G, +)$ is an abelian group in which every element has finite order, is a torsion module.

(iii) If V is a finite-dimensional vector space over $\mathbb{F}$ and $T \in \mathcal{L}(V)$, then, as an $\mathbb{F}[x]$-module (via T), V is a torsion module: for each $v \in V$ the minimal polynomial $m(x)$ of T satisfies $m(x)v = m(T)(v) = 0$, and $m(x) \neq 0$.

1.6.7 Finitely Generated Modules

Definition 1.55. Suppose M is an R-module and $S \subseteq M$. The submodule of M generated by S, written $\langle S \rangle$, is defined as the smallest submodule of M containing S. If $\langle S \rangle = M$, then S is said to be a *generator* of M. If M has a finite generating set, then M is said to be *finitely generated* (f.g.).

Exercise 1.56 (Homework). If V is a finite-dimensional vector space over a field $\mathbb{F}$ and $T \in \mathcal{L}(V)$, then V, as an $\mathbb{F}[x]$-module (via T), is finitely generated.

1.6.8 A finitely generated torsion module over a PID

Since V is finite-dimensional, let $\beta = \{v_1, \ldots, v_n\}$ be a basis of V. For any $v \in V$,

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

for some $\alpha_i \in \mathbb{F} \subset \mathbb{F}[x]$. Hence $\langle \beta \rangle = V$, i.e. the $\mathbb{F}[x]$-module V is finitely generated.

Combined with example 1.54 (Example above, part (iii)), this shows:

Theorem 1.57. If V is a finite-dimensional vector space over $\mathbb{F}$ and $T \in \mathcal{L}(V)$, then V, as an $\mathbb{F}[x]$-module via T, is a *finitely generated torsion module* over the principal ideal domain $\mathbb{F}[x]$.

This single fact — V is a finitely generated torsion $\mathbb{F}[x]$-module — is the starting point for the entire structure theory of canonical forms developed below.

1.6.9 Cyclic submodules

Definition 1.58. The *cyclic submodule of V generated by $v_0 \in V$* is

$$\langle v_0 \rangle \coloneqq \{f(x)v_0 : f(x) \in \mathbb{F}[x]\},$$

equivalently written $\mathcal{Z}(v_0; T)$: the smallest T-invariant subspace of V containing v_0.

1.6.10 Unitary cyclic modules

Definition 1.59. A (left, unitary — i.e. $1 \cdot m = m$ for all m) R-module M is *cyclic* if $M = \langle m_0 \rangle$ for some $m_0 \in M$.

Theorem 1.60. Suppose M is a unitary cyclic (left) R-module generated by m_0, i.e. $M = \langle m_0 \rangle$. Then

$$M \cong R/I$$

as R-modules, for some left ideal I of R.

Here, for a left ideal $I \trianglelefteq R$, the quotient R/I is made into an R-module via

$$R \times R/I \longrightarrow R/I, \qquad (r,\, s+I) \longmapsto r(s+I) \coloneqq rs + I.$$

Remark 1.61 (Well-definedness). If $s + I = t + I$, i.e. $s - t \in I$, then $r(s-t) \in I$ for every $r \in R$ (since I is a left ideal), so $rs + I = rt + I$ for every $r \in R$. Thus the action is well defined. (Recall also that the submodules of the R-module R are exactly the left ideals of R.)

Proof of theorem 1.60. Define

$$\Phi : R \longrightarrow M = \langle m_0 \rangle, \qquad \Phi(r) \coloneqq rm_0.$$

Clearly Φ is surjective (every element of $\langle m_0 \rangle$ has the form rm_0).

We show Φ is a module homomorphism. Let $r, s \in R$. Then

$$\Phi(r+s) = (r+s)m_0 = rm_0 + sm_0 = \Phi(r) + \Phi(s),$$

$$\Phi(rs) = (rs)m_0 = r(sm_0) = r\Phi(s).$$

Hence Φ is an R-module homomorphism. By the First Isomorphism Theorem for modules, as left R-modules,

$$M \cong R/\ker \Phi.$$

Now $\ker \Phi$ is a submodule of the left R-module R, hence $\ker \Phi$ is a left ideal of R. This completes the proof. ∎

Remark 1.62 (Observation).

$$\ker \Phi = \{r \in R : \Phi(r) = 0\} = \{r \in R : rm_0 = 0\} = \operatorname{Ann}_R(m_0).$$

So in fact $M \cong R/\operatorname{Ann}_R(m_0)$.

1.6.11 The case R a PID

Proposition 1.63. In the discussion above, if R is a PID, then $\mathrm{Ann}_R(m_0)$ is a principal ideal of R. Any generator of $\mathrm{Ann}_R(m_0)$ is called *an order* of m_0; it is unique up to multiplication by a unit of R.

Specialising to $R = \mathbb{F}[x]$ and $M = \langle v_1 \rangle = \{f(x)v_1 : f(x) \in \mathbb{F}[x]\}$:

$$\mathrm{Ann}_{\mathbb{F}[x]}(v_1) = \{f(x) \in \mathbb{F}[x] : f(x)v_1 = 0\} = \langle g(x) \rangle = \langle u\, g(x) \rangle \quad \text{for every } u \in \mathbb{F} \setminus \{0\}.$$

This ideal has a *monic* generator, and it is unique; denote it by $f_1(x)$. Then

$$\langle v_1 \rangle = \{f(x)v_1 : f(x) \in \mathbb{F}[x]\} \;\cong\; \mathbb{F}[x]/\,\mathrm{Ann}_{\mathbb{F}[x]}(v_1) \;=\; \mathbb{F}[x]/\big(f_1(x)\big).$$

Theorem 1.64 (Structure theorem for finite abelian groups — recalled)**.** Suppose G is a finite abelian group. Then

$$G \cong \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_k}, \qquad n_1 \mid n_2 \mid \cdots \mid n_k.$$

The integers $n_1, \ldots, n_k$ are called the *invariant factors* of G.

By theorem 1.57, the $\mathbb{F}[x]$-module V (via T) is a finitely generated torsion module over the PID $\mathbb{F}[x]$, exactly as a finite abelian group is a finitely generated torsion $\mathbb{Z}$-module. The structure theory runs in complete parallel.

1.6.12 The Cyclic Decomposition Theorem

Theorem 1.65 (Cyclic Decomposition Theorem)**.** Suppose V is a finite-dimensional vector space over a field $\mathbb{F}$ and T is a linear operator on V. Then there exist vectors $v_1, \ldots, v_k \in V$ such that

$$V \cong \langle v_1 \rangle \oplus \cdots \oplus \langle v_k \rangle \qquad \text{as } \mathbb{F}[x]\text{-modules},$$

equivalently, as vector spaces over $\mathbb{F}$,

$$V = \mathcal{Z}(v_1; T) \oplus \cdots \oplus \mathcal{Z}(v_k; T),$$

where $\mathcal{Z}(v_i; T) = \langle v_i \rangle$ is the smallest T-invariant subspace of V containing v_i (equivalently, the cyclic submodule of the $\mathbb{F}[x]$-module V generated by v_i). Moreover, writing $f_i(x)$ for the monic generator of $\mathrm{Ann}_{\mathbb{F}[x]}(v_i)$,

$$V \;\cong\; \mathbb{F}[x]/\big(f_1(x)\big) \oplus \cdots \oplus \mathbb{F}[x]/\big(f_k(x)\big),$$

where

$$f_1(x) \mid f_2(x) \mid \cdots \mid f_k(x).$$

Definition 1.66. The polynomials $f_1(x) \mid f_2(x) \mid \cdots \mid f_k(x)$ in theorem 1.65 are called the *invariant factors of* T. The largest invariant factor $f_k(x)$ is the minimal polynomial of T, and the product $f_1(x)f_2(x)\cdots f_k(x)$ is the characteristic polynomial of T.

1.6.13 Companion matrices and the Rational Canonical Form

Writing $T_i = T|_{\mathcal{Z}(v_i;T)}$, the set

$$\big\{v_i,\; T_i v_i,\; \ldots,\; T_i^{n_i-1} v_i\big\}, \qquad n_i = \deg f_i,$$

is a basis β_i of $\mathcal{Z}(v_i;T)$, and the matrix of T_i with respect to β_i is the *companion matrix* C_{f_i} of $f_i(x)$. Taking $\beta = \beta_1 \cup \cdots \cup \beta_k$ as an (ordered) basis of V, we obtain a block-diagonal representation

$$[T]_\beta = \begin{pmatrix} C_{f_1} & & & \\ & C_{f_2} & & \\ & & \ddots & \\ & & & C_{f_k} \end{pmatrix}, \qquad f_1 \mid f_2 \mid \cdots \mid f_k.$$

This is the *Rational Canonical Form* (RCF) of T.

Remark 1.67 (Three levels of "nice" representation, and what is actually guaranteed)**.** Given an operator T, one looks for a nice matrix representation:

(1) a *diagonal* matrix — **not** guaranteed (only when T is diagonalisable);

(2) a block-diagonal matrix with *elementary Jordan blocks* — **not** guaranteed in general (only when the characteristic polynomial splits over $\mathbb{F}$);

(3) a block-diagonal matrix with companion matrices as blocks (with the divisibility chain $f_1 \mid f_2 \mid \cdots \mid f_k$) — **always** guaranteed, by the Cyclic Decomposition Theorem.

1.6.14 Definition of invariant factors, restated

Definition 1.68.

(1) Let $A \in M_n(\mathbb{F})$. Suppose A is similar to a block-diagonal matrix whose blocks are companion matrices $C_{f_1}, \ldots, C_{f_R}$ with $f_1 \mid f_2 \mid \cdots \mid f_R$. Then $f_1, \ldots, f_R$ are called the *invariant factors of* A; they are unique in the sense that any other such chain $g_1 \mid g_2 \mid \cdots \mid g_R$ obtained from a similar block-diagonal matrix of companion matrices must satisfy $g_i = f_i$ for all i.

(2) Suppose V is a finite-dimensional vector space and $T \in \mathcal{L}(V)$. Realise T by a matrix $A = [T]_\beta$ with respect to some (any) ordered basis β of V (e.g. for $A \in M_n(\mathbb{F})$, $V = \mathbb{F}^n$ and $T : \mathbb{F}^n \to \mathbb{F}^n$, $x \mapsto Ax$). The *invariant factors of* T are, by definition, the invariant factors of A.

1.6.15 Similar Linear Operators

Definition 1.69. Given two linear operators $S, T : V \to V$, we say S and T are *similar* if there exists a bijective linear operator $U : V \to V$ such that

$$T = U^{-1}SU.$$

Remark 1.70. Corresponding $\mathbb{F}[x]$-modules (i.e. V viewed via S and V viewed via T) are isomorphic precisely when S and T are similar. This is the module-theoretic reformulation of similarity, and it is the conceptual reason invariant factors — being isomorphism invariants of the module V — are a complete similarity invariant.

1.6.16 Worked Examples

Theorem 1.71. If A is the companion matrix of a monic polynomial $f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + x^n$, that is,

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix},$$

then $f(x)$ is *both* the characteristic polynomial and the minimal polynomial of A.

Proof. Let $e_1 = (1, 0, \ldots, 0)$. Then

$$Ae_1 = (0, 1, 0, \ldots, 0) = e_2, \qquad A^2e_1 = (0, 0, 1, \ldots, 0) = e_3, \qquad \ldots, \qquad A^{n-1}e_1 = (0, \ldots, 0, 1) = e_n.$$

Hence

$$\mathcal{Z}(e_1; A) = \operatorname{span}\{e_1, \ldots, e_n\} = V,$$

so e_1 is a *cyclic vector* for A. (As a general fact, if a linear operator on a finite-dimensional space has a cyclic vector, its characteristic and minimal polynomials coincide.)

To see this directly: write $p(A) = a_0I + a_1A + \cdots + a_{n-1}A^{n-1} + A^n$. A direct computation using $Ae_1 = e_2, \ldots, A^{n-1}e_1 = e_n$ and the last column of A shows

$$p(A)e_1 = 0,$$

and then, using $Ap(A) = p(A)A$,

$$p(A)e_2 = p(A)Ae_1 = Ap(A)e_1 = A \cdot 0 = 0, \qquad \ldots, \qquad p(A)e_n = 0.$$

Since $\{e_1, \ldots, e_n\}$ is a basis of V, $p(A) = 0$. As e_1 is a cyclic vector, the minimal polynomial of A has degree exactly $\dim V = n = \deg p$. Hence p is the minimal polynomial, and since p is monic of degree n and annihilates A while having the right degree, p is also the characteristic polynomial of A. ∎

Example 1.72. The companion matrix of $f(x) = x^2 + 1$ is

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in M_2(\mathbb{R}).$$

Example 1.73 (A 5×5 matrix with $\chi_A = m_A$)**.** Give an example of a 5×5 real matrix whose characteristic polynomial and minimal polynomial are the same.

Take the diagonal matrix

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \in M_5(\mathbb{R}).$$

Then $\chi_A(x) = (x-1)(x-2)(x+1)(x-3)(x-4)$. Since χ_A and m_A always have the same roots, and here all five roots are simple and distinct,

$$m_A(x) = (x-1)(x-2)(x+1)(x-3)(x-4) = \chi_A(x).$$

Example 1.74 (A non-diagonalisable 5×5 matrix with $\chi_A = m_A$)**.** Give an example of a 5×5 real matrix (not necessarily diagonalisable) whose characteristic and minimal polynomials are the same.

Take the single elementary Jordan block of size 5:

$$A = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

An elementary Jordan block always has $\chi_A = m_A = (x-\lambda)^n$, because the standard basis vector e_1 is a cyclic vector for A: T has a cyclic vector if and only if the minimal and characteristic polynomials are identical.

More generally, any *cyclic* operator (one possessing a cyclic vector) can be put, via $\{v, Av, A^2v, A^3v, A^4v\}$, into the companion-matrix form of theorem 1.71: if $f(x) = x^5 + ax^4 + bx^3 + cx^2 + dx + e$ is the desired common characteristic/minimal polynomial, the companion matrix

$$\begin{pmatrix} 0 & 0 & 0 & 0 & -e \\ 1 & 0 & 0 & 0 & -d \\ 0 & 1 & 0 & 0 & -c \\ 0 & 0 & 1 & 0 & -b \\ 0 & 0 & 0 & 1 & -a \end{pmatrix}$$

works equally well.

Example 1.75 (The all-ones 100×100 matrix)**.** Let $A = (a_{ij}) \in M_{100}(\mathbb{R})$ with $a_{ij} = 1$ for all i, j. Determine the minimal and characteristic polynomials of A.

Solution. If a matrix B has constant row sum $n\lambda$, then $n\lambda$ is an eigenvalue, since

$$\begin{pmatrix} \lambda & \lambda & \cdots & \lambda \\ \vdots & & & \vdots \\ \lambda & \lambda & \cdots & \lambda \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = n\lambda \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

(For constant column sum, apply this to B^T; recall $\chi_B = \chi_{B^T}$.)

Here every row sum of A is 100, so 100 is an eigenvalue of A. Also $\operatorname{rank} A = 1$, so

$$\operatorname{Null}(A) = 100 - 1 = 99, \qquad \dim \ker(A - 0 \cdot I_{100}) = 99,$$

i.e. 0 is an eigenvalue with geometric multiplicity 99. (Recall: for any matrix B and scalar λ, λ is an eigenvalue of B iff $\ker(B - \lambda I) \neq 0$, and the geometric multiplicity is $\dim \ker(B - \lambda I)$.) Hence the algebraic multiplicity of 0 is at least 99. Since A is 100×100 and 100 is also an eigenvalue, the algebraic multiplicity of 0 is exactly 99, and

$$\chi_A(x) = x^{99}(x - 100).$$

The algebraic and geometric multiplicities of *every* eigenvalue of A coincide ($99 = 99$ for 0, and $1 = 1$ for 100), so A is diagonalisable; hence m_A is a product of distinct linear factors, and since m_A and χ_A have the same roots,

$$m_A(x) = x(x - 100).$$

In particular A is similar to $\operatorname{diag}(100, 0, \ldots, 0)$ (99 zeros).

Example 1.76 (Operators of rank one)**.** Suppose T is a linear operator on a finite-dimensional vector space with $\operatorname{rank} T = 1$. What can be said about T?

Solution. Exactly one of the following holds (never both):

- T is diagonalisable: then T has eigenvalue 0 with multiplicity $n-1$ and a single nonzero eigenvalue λ, and $m_T(x) = x(x-\lambda)$;
- T is nilpotent (but not zero): then $T^2 = 0$, $m_T(x) = x^2$, and the Jordan form of T consists of Jordan blocks of order 1 (i.e. zero blocks) except for exactly one block of order 2 — this is forced because, for a nilpotent operator with blocks of size at most 2, $\operatorname{rank} T$ equals the number of blocks of size ≥ 2.

Example 1.77 (Possible invariant factors from χ and m)**.** Suppose $\chi_T(x) = (x-2)^5(x-3)^4$ and $m_T(x) = (x-2)^2(x-3)^2$. What are the possible invariant factors of T?

Method. If $f_1 \mid f_2 \mid \cdots \mid f_k$ are the invariant factors, each f_i is of the form $(x-2)^{a_i}(x-3)^{b_i}$ with

$$0 \leq a_1 \leq a_2 \leq \cdots \leq a_k, \qquad 0 \leq b_1 \leq b_2 \leq \cdots \leq b_k,$$

$$a_1 + \cdots + a_k = 5, \qquad b_1 + \cdots + b_k = 4, \qquad a_k = b_k = 2$$

(the last condition coming from $f_k = m_T$). So one writes 5 as an ordered sum (a partition, read in increasing order) of parts each ≤ 2 with largest part $= 2$, and likewise 4, and pairs them up using a common number of parts (padding the shorter list with leading 0's). The partitions of 5 into parts ≤ 2 with largest part 2 are

$$5 = 1+2+2 = 1+1+1+2,$$

and the partitions of 4 into parts ≤ 2 with largest part 2 are

$$4 = 2+2 = 1+1+2.$$

Pairing lists of equal length gives two valid invariant factor chains:

$$(x-2)(x-3) \mid (x-2)^2(x-3) \mid (x-2)^2(x-3)^2,$$

$$(x-2) \mid (x-2)(x-3) \mid (x-2)(x-3) \mid (x-2)^2(x-3)^2.$$

(A pairing such as $(x-2)^3(x-3)^2 \mid (x-2)^2(x-3)^2$ fails, because the exponents of $(x-3)$, namely 0 and 2, sum to $2 \neq 4$; similarly for other mismatched pairings.) Each valid chain corresponds to a possible *primary rational canonical form* for T (rational canonical form with respect to elementary divisors), and — when the field is such that $x-2, x-3$ split off completely — each also corresponds to a Jordan canonical form, obtained by further decomposing each $(x-2)^{a_i}$, $(x-3)^{b_i}$ factor into its own Jordan block. This is the precise correspondence between the Rational Canonical Form and the Jordan Canonical Form: they encode the same data (the elementary divisors), merely packaged differently.

1.7 5 May

1.7.1 Bilinear Forms

A *linear form* (or *linear functional*) on V is a linear map

$$f : V \longrightarrow \mathbb{F}, \qquad f(\alpha v + \beta w) = \alpha f(v) + \beta f(w).$$

A *bilinear form* combines two "slots", each behaving linearly.

Definition 1.78. Suppose V is a vector space over a field $\mathbb{F}$. A mapping

$$f : V \times V \longrightarrow \mathbb{F}$$

is said to be a *bilinear form* if

$$f(\alpha v_1 + \beta v_2,\ v_3) = \alpha f(v_1, v_3) + \beta f(v_2, v_3), \tag{1.20}$$
$$f(v_1,\ \alpha v_2 + \beta v_3) = \alpha f(v_1, v_2) + \beta f(v_1, v_3), \tag{1.21}$$

for all $v_1, v_2, v_3 \in V$ and all $\alpha, \beta \in \mathbb{F}$ (i.e. f is linear in each variable separately, the other held fixed). If instead $f : V \times V \to V$ satisfies the analogous conditions, f is called a *bilinear mapping*.

Definition 1.79. A *sesquilinear form*, or simply a *form*, on a real or complex vector space V is a function $f : V \times V \to F$ ($F = \mathbb{R}$ or $\mathbb{C}$) such that

- $f(\alpha u + v, w) = \alpha f(u, w) + f(v, w)$
- $f(u, \alpha v + w) = \overline{\alpha} f(u, v) + f(u, w)$

$\forall u, v, w \in V$ and $\forall \alpha \in F$.

1.7.2 The matrix of a bilinear form

Suppose $\dim V = n$ is finite and $\beta = \{v_1, \dots, v_n\}$ is an ordered basis of V. The $n \times n$ matrix

$$A = (a_{ij}) \in M_n(\mathbb{F}), \qquad a_{ij} \coloneqq f(v_i, v_j) \quad (1 \le i, j \le n),$$

$$A = \begin{pmatrix} f(v_1, v_1) & f(v_1, v_2) & \cdots & f(v_1, v_n) \\ f(v_2, v_1) & f(v_2, v_2) & \cdots & f(v_2, v_n) \\ \vdots & & & \vdots \\ f(v_n, v_1) & f(v_n, v_2) & \cdots & f(v_n, v_n) \end{pmatrix}$$

is called the *matrix of f with respect to β*, denoted $[f]_\beta$.

Theorem 1.80. Suppose V is a finite-dimensional vector space over $\mathbb{F}$ and f is a bilinear form on V. Let $\beta = \{v_1, \dots, v_n\}$ be an ordered basis of V, and let $A = [f]_\beta$. Then for all $v, w \in V$,

$$f(v, w) = [v]_\beta^T \, A \, [w]_\beta.$$

Proof. Write $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ and $w = \gamma_1 v_1 + \cdots + \gamma_n v_n$, so that

$$[v]_\beta = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}, \qquad [w]_\beta = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix}.$$

Using bilinearity of f repeatedly,

$$f(v, w) = f\Big(\sum_{i=1}^n \alpha_i v_i,\ w\Big) = \sum_{i=1}^n \alpha_i f(v_i, w) = \sum_{i=1}^n \alpha_i f\Big(v_i, \sum_{j=1}^n \gamma_j v_j\Big)$$
$$= \sum_{i=1}^n \alpha_i \sum_{j=1}^n \gamma_j f(v_i, v_j) = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \gamma_j \, a_{ij}.$$

On the other hand,

$$[v]_\beta^T A[w]_\beta = (\alpha_1, \dots, \alpha_n)\, A \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix} = \sum_{i=1}^{n} \alpha_i \Big(\sum_{j=1}^{n} a_{ij}\gamma_j \Big) = \sum_{i,j} \alpha_i \gamma_j a_{ij}.$$

The two expressions agree, proving the claim. (One may also verify this column by column: $[v]_\beta^T A[w]_\beta$ applied with $v = v_i, w = v_j$ standard basis vectors of $\mathbb{F}^n$ recovers $a_{ij} = f(v_i, v_j)$ directly, and the general case follows by bilinearity in each argument separately, as carried out above.) ■

1.7.3 The converse: every matrix gives a bilinear form

Theorem 1.81 (Generating bilinear forms from matrices)**.** Suppose V is an n-dimensional vector space over a field $\mathbb{F}$ and $A = (a_{ij})$ is an $n \times n$ matrix over $\mathbb{F}$. Let $\beta = \{v_1, \dots, v_n\}$ be an ordered basis of V. Define $f : V \times V \to \mathbb{F}$ by

$$f(v, w) = [v]_\beta^T \, A \, [w]_\beta.$$

Then f is a bilinear form on V with $[f]_\beta = A$.

Proof. First, $[f]_\beta = A$: taking $v = v_1, w = v_1$,

$$f(v_1, v_1) = (1, 0, \dots, 0)\, A\, (1, 0, \dots, 0)^T = a_{11},$$

and likewise $f(v_1, v_2) = (1, 0, \dots, 0)A(0, 1, 0, \dots, 0)^T = a_{12}$, and in general $f(v_i, v_j) = a_{ij}$.

Next, bilinearity. Let $u, v, w \in V$ and $\alpha, \beta \in \mathbb{F}$ (reusing β generically here for the scalar, with apologies for notation clash with the basis). Then $[\alpha u + \beta v] = \alpha[u] + \beta[v]$ (coordinates are linear in the vector), so writing $[u] = (x_1, \dots, x_n)^T$, $[v] = (y_1, \dots, y_n)^T$, $[w] = (z_1, \dots, z_n)^T$,

$$\begin{aligned} f(\alpha u + \beta v,\, w) &= (\alpha x_1 + \beta y_1, \dots, \alpha x_n + \beta y_n)\, A\, (z_1, \dots, z_n)^T \\ &= \alpha\, (x_1, \dots, x_n) A (z_1, \dots, z_n)^T \;+\; \beta\, (y_1, \dots, y_n) A (z_1, \dots, z_n)^T \\ &= \alpha f(u, w) + \beta f(v, w), \end{aligned}$$

using the distributivity of matrix multiplication over addition. This proves linearity in the first slot. Linearity in the second slot is proved the same way, since $A[\alpha w_1 + \beta w_2] = \alpha A[w_1] + \beta A[w_2]$. Hence f is bilinear. ■

Example 1.82. Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \in M_2(\mathbb{R})$, $V = \mathbb{R}^2$, $\beta = \{e_1, e_2\} = \{(1, 0), (0, 1)\}$. For $(x_1, x_2), (y_1, y_2) \in \mathbb{R}^2$,

$$(x_1, x_2) \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = x_1 y_1 + 2x_1 y_2 + 3x_2 y_1 + 4x_2 y_2.$$

This is a homogeneous polynomial in x_1, x_2, y_1, y_2 (of degree one in the x's and degree one in the y's).

Remark 1.83. A bilinear form over $\mathbb{F}^n$ is nothing but (the formula for) a homogeneous polynomial expression that is linear in each of two blocks of n variables. For example, with $A = \begin{pmatrix} 1 & 2 & 0 \\ 3 & 4 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in M_3(\mathbb{R})$,

$$(x_1, x_2, x_3)A(y_1, y_2, y_3)^T = x_1 y_1 + 2x_1 y_2 + 3x_2 y_1 + 4x_2 y_2 + 0,$$

which happens to have no x_3, y_3 terms; whether one regards this as a bilinear form over $\mathbb{R}^2$ or over $\mathbb{R}^3$ depends entirely on which space V was declared to be. If $\dim V$ is not specified, the question is simply *incomplete information.*

1.7.4 Matrices and linear transformations: an aside

For comparison, recall the dictionary between matrices and linear maps:

$$A \in M_{m\times n}(\mathbb{F}) \;\rightsquigarrow\; L_A : \mathbb{F}^n \to \mathbb{F}^m,\; x \mapsto Ax,$$

and, more generally, for $\dim V = n$ with basis β_1 and $\dim W = m$ with basis β_2, every linear map $T : V \to W$ is represented by a matrix $[T]_{\beta_1}^{\beta_2} \in M_{m\times n}(\mathbb{F})$, giving an isomorphism of vector spaces

$$\mathcal{L}(V, W) \;\cong\; M_{m\times n}(\mathbb{F}).$$

This is the model for the analogous statement about bilinear forms below.

1.7.5 The Vector Space of Bilinear Forms

Definition 1.84. Let $\mathrm{Bil}(V)$ denote the set of all bilinear forms on V. Define, for $f, g \in \mathrm{Bil}(V)$ and $\lambda \in \mathbb{F}$,

$$(f+g)(v,w) \coloneqq f(v,w) + g(v,w), \qquad (\lambda f)(v,w) \coloneqq \lambda f(v,w) \qquad \forall\, v, w \in V.$$

Proposition 1.85. With the operations above, $\mathrm{Bil}(V)$ is a vector space over $\mathbb{F}$.

Theorem 1.86. If $\dim V = n$, then $\mathrm{Bil}(V) \cong M_n(\mathbb{F})$, via $f \mapsto [f]_\beta$ for any fixed ordered basis β of V.

1.7.6 Inner Products

Definition 1.87. Let V be a vector space over $\mathbb{F}$ (where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$). A mapping

$$\langle \cdot, \cdot \rangle : V \times V \longrightarrow \mathbb{F}$$

is called an *inner product* on V if it satisfies, for all $u, v, w \in V$ and $\lambda \in \mathbb{F}$:

(i) $\overline{\langle v, w\rangle} = \langle w, v\rangle$ (conjugate symmetry)

(ii) $\langle \lambda v, w\rangle = \lambda \langle v, w\rangle$ (linearity in the first slot, scalars)

(iii) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ (linearity in the first slot, vectors)

(iv) $\langle v, v\rangle > 0$ for all $v \neq 0 \in V$ (positivity)

Here $\overline{\cdot}$ denotes complex conjugation. Note that if $\mathbb{F} = \mathbb{R}$, condition (i) simply reads $\langle v, w\rangle = \langle w, v\rangle$ for all $v, w \in V$.

1.7.7 Basic consequences

Proposition 1.88. Let $\langle \cdot, \cdot \rangle$ be an inner product on V. Then:

(I) Conditions (i) and (ii) together imply

$$\langle v, \lambda w \rangle = \overline{\lambda}\, \langle v, w \rangle \qquad \forall\, v, w \in V,\ \lambda \in \mathbb{F}.$$

(II) Condition (i) implies $\langle v, v \rangle$ is real for every $v \in V$.

(III) Conditions (ii) and (iii), used separately, give

$$\langle 0, 0 \rangle = \langle 0 \cdot v,\ 0 \rangle = 0 \cdot \langle v, 0 \rangle = 0$$

(the first 0 a scalar, the second a vector).

(IV)

$$\langle v,\ u + w \rangle = \overline{\langle u + w, v \rangle} = \overline{\langle u, v \rangle + \langle w, v \rangle} = \overline{\langle u, v \rangle} + \overline{\langle w, v \rangle} = \langle v, u \rangle + \langle v, w \rangle.$$

Proof. (I): $\overline{\langle w, \lambda v \rangle} \overset{\text{(i)}}{=} \langle \lambda v, w \rangle \overset{\text{(ii)}}{=} \lambda \langle v, w \rangle$. Also $\overline{\langle w, \lambda v \rangle} = \overline{\overline{\lambda} \langle w, v \rangle}$ once we know the result for the first slot w... more directly: apply (i) and (ii) as in the displayed computation in part (IV) above with the roles of the arguments adjusted; the stated identity follows by conjugating $\langle \lambda w, v \rangle = \lambda \langle w, v \rangle$ and using (i) twice. (II): by (i), $\overline{\langle v, v \rangle} = \langle v, v \rangle$, so $\langle v, v \rangle \in \mathbb{R}$. (III) and (IV) are the displayed computations. ■

Remark 1.89.

(i) An inner product on a complex vector space is *conjugate-linear* in the second slot (by Proposition 1.88(I)).

(ii) An inner product on a real vector space *is* a bilinear form (no conjugation needed, since $\mathbb{F} = \mathbb{R}$).

1.7.8 The matrix of an inner product

Definition 1.90. Let $\langle \cdot, \cdot \rangle$ be an inner product on a finite-dimensional vector space V over $\mathbb{F}$ ($= \mathbb{R}$ or $\mathbb{C}$), and let $\beta = \{v_1, \ldots, v_n\}$ be an ordered basis of V. The $n \times n$ matrix $A = (a_{ij}) \in M_n(\mathbb{F})$ is said to be *the matrix of the inner product* $\langle \cdot, \cdot \rangle$ *with respect to* β if

$$a_{ij} \coloneqq \langle v_j, v_i \rangle \qquad \text{for } 1 \le i, j \le n.$$

Theorem 1.91. With the notation above, for $v, w \in V$:

(i) $\langle v, w \rangle = [w]_\beta^T\, A\, [v]_\beta$.

(ii) $A^* = A$, where $A^* \coloneqq \left(\overline{A}\right)^T$ (the conjugate transpose).

(iii) For every $v \neq 0 \in V$, $[v]_\beta^T\, A\, [v]_\beta > 0$.

Proof. (i) follows exactly as in theorem 1.80, with the convention $a_{ij} = \langle v_j, v_i \rangle$ producing the transposed argument order. (ii): since $a_{ij} = \langle v_j, v_i \rangle$,

$$\overline{a_{ij}} = \overline{\langle v_j, v_i \rangle} = \langle v_i, v_j \rangle = a_{ji},$$

which is exactly the statement $(\overline{A})^T = A$, i.e. $A^* = A$. (iii) is condition (iv) (positivity) of the definition of inner product, rewritten using part (i) with $v = w$. ■

Definition 1.92. A matrix $A \in M_n(\mathbb{F})$ satisfying $A^* = A$ (*Hermitian*, or *symmetric* when $\mathbb{F} = \mathbb{R}$) and $x^T Ax > 0$ for all $x \neq 0$ (suitably interpreted with a conjugate where $\mathbb{F} = \mathbb{C}$) is called a *positive definite matrix*.

1.7.9 Building bilinear/inner-product forms from linear forms

Example 1.93 (Combining two linear forms)**.** Given linear forms $f, g : V \to \mathbb{F}$ (or, more generally, two real-valued functionals), one obtains a bilinear form on $V \times V$ via

$$(v, w) \longmapsto f(v)\, g(w).$$

For instance, on $\mathbb{R}^2 \times \mathbb{R}^2$, writing $v_1 = (x_1, y_1)$, $v_2 = (x_2, y_2)$, the assignment

$$\langle v_1, v_2 \rangle = x_1 x_2 + 2 y_1 y_2$$

is built this way (and is, in fact, an inner product on $\mathbb{R}^2$).

Exercise 1.94 ([HK70])**.** Can you define an inner product on $\mathbb{R}^2$ such that $\langle (1,0), (0,1) \rangle \neq 0$? How many such inner products are there?

Example 1.95 (The standard inner product on $\mathbb{R}^2$)**.** With $\beta = \{e_1, e_2\} = \{(1,0), (0,1)\}$, the standard inner product has matrix

$$\begin{pmatrix} \langle e_1, e_1 \rangle & \langle e_2, e_1 \rangle \\ \langle e_1, e_2 \rangle & \langle e_2, e_2 \rangle \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2,$$

and for $x, y \in \mathbb{R}^2$, $\langle x, y \rangle = y^T I x = y^T x$.

1.7.10 Change of Basis

Suppose f is a bilinear form on a finite-dimensional space V, and $\beta_1 = \{v_1, \ldots, v_n\}$, $\beta_2 = \{w_1, \ldots, w_n\}$ are two ordered bases of V. What is the relationship between $[f]_{\beta_1}$ and $[f]_{\beta_2}$? (One may ask the analogous question for an inner product.)

Let $P = (p_{ij})_{n \times n}$ be the change-of-basis matrix expressing β_2 in terms of β_1:

$$w_j = p_{1j} v_1 + p_{2j} v_2 + \cdots + p_{nj} v_n, \qquad j = 1, \ldots, n.$$

Equivalently, $P = [\mathrm{Id}_V]_{\beta_2}^{\beta_1}$; or, letting $T : V \to V$ be the (necessarily bijective) linear operator with $T(v_i) = w_i$, $P = [T]_{\beta_1}$. Either description shows that P is invertible (its inverse is $[\mathrm{Id}_V]_{\beta_1}^{\beta_2}$, respectively $[T^{-1}]_{\beta_1}$).

Theorem 1.96 (Congruence transformation)**.** With the notation above,

$$[f]_{\beta_2} = P^T [f]_{\beta_1} P,$$

and, if $\langle \cdot, \cdot \rangle$ is an inner product on V,

$$\big[\langle \cdot, \cdot \rangle\big]_{\beta_2} = P^* \big[\langle \cdot, \cdot \rangle\big]_{\beta_1} P.$$

(This is left as an exercise: combine theorem 1.80 (resp. theorem 1.91) with the change-of-basis formula $[v]_{\beta_1} = P[v]_{\beta_2}$.)

Remark 1.97. theorem 1.96 is the bilinear-form/inner-product analogue of the similarity transformation $[T]_{\beta_2} = P^{-1}[T]_{\beta_1}P$ for linear operators (§6). The key structural difference is that matrices of bilinear forms transform by *congruence* ($P^T(\cdot)P$), while matrices of linear operators transform by *similarity* ($P^{-1}(\cdot)P$); for inner products the relevant transformation is congruence by P^*, reflecting the conjugate-symmetric nature of the form.

Theme 2

Field and Galois theory

2.1 20 April

The key idea of Galois theory is to relate a field extension $K \subset F$ to the group of all automorphisms of F that fix K elementwise (the Galois group of the extension). The fundamental theorem of Galois theory states that there is a one-to-one correspondence between the intermediate fields of a (finite dimensional) Galois field extension and the subgroups of the Galois group of the extension.

2.1.1 Characteristic of a field

We already know that a field is a commutative division ring.

Definition 2.1. The *characteristic* of a field K is the smallest positive integer n such that $n \cdot 1_K = 0$, where 1_K is the multiplicative identity in K. If no such positive integer exists, we say that the characteristic of K is zero. We denote the characteristic of K by $\text{char}(K) = n$.

Example 2.2. For any prime p, the finite field $\mathbb{Z}_p$ has characteristic p. The field of rational numbers $\mathbb{Q}$ has characteristic zero. The field of real numbers $\mathbb{R}$ and the field of complex numbers $\mathbb{C}$ also have characteristic zero.

Theorem 2.3. The characteristic of a field is either zero or a prime number.

Proof. Let K be a field and let $n = \text{char}(K)$. Then $n \cdot 1_K = 0$. Assume for contradiction that n is composite, say $n = ab$ with $1 < a, b < n$. Then we have

$$(ab) \cdot 1_K = 0 \implies (a \cdot 1_K)(b \cdot 1_K) = 0 \implies a \cdot 1_K = 0 \text{ or } b \cdot 1_K = 0,$$

since K is a field. This contradicts the minimality of n. Therefore, n must be prime or zero. ■

Definition 2.4. In a commutative ring R with identity,

1. for $a, b \in R$, $a \mid b$ means $\exists c \in R$ such that $b = ac$.
2. $u \in R$ is said to be *invertible* or *unit* if $\exists v \in R$ such that $uv = 1$.
3. $a \in R$ is said to be *irreducible* if a is not a unit and $a = bc \implies b$ is a unit or c is a unit.
4. a nonzero nonunit $a \in R$ is said to be *prime* if whenever $a \mid bc \implies a \mid b$ or $a \mid c$.

Definition 2.5. Let K be a field. A polynomial $f(x) \in K[x]$ is *irreducible* if $\deg f \geq 1$ and if we cannot write $f(x)$ as the product

$$f(x) = g(x)h(x)$$

with $g, h \in K[x]$, $g, h \notin K$ and $\deg g, \deg h < \deg f$.

In other words, an irreducible polynomial is a non-constant polynomial that cannot be factored into the product of two non-constant polynomials. If R is a ring, then $f \in R[x]$ is irreducible if f is irreducible over the field of fractions of R.

Example 2.6. $f(x) = 2(x+2) \in \mathbb{Z}[x]$ is not irreducible element but it is an irreducible polynomial.

For a field K, the polynomial ring $K[x]$ is a principal ideal domain (PID) and hence a unique factorization domain (UFD). Therefore, in $K[x]$, the notions of irreducible and prime elements coincide. In particular, an irreducible polynomial over a field is also an irreducible, and hence prime, element.

A polynomial of degree 2 or 3 over K is irreducible if and only if it has no roots in K.

Example 2.7. The degree 4 polynomial $x^4 + x^2 + 1 = (x^2 + x + 1)(x^2 - x + 1)$ is reducible.

Now as $K[x]$ is PID, $f \in K[x]$ is irreducible if and only if the ideal (f) is a prime ideal, and in a PID every prime ideal is maximal. So f is irreducible if and only if (f) is a maximal ideal. Now, an ideal M in a commutative ring R is maximal if and only if the quotient ring R/M is a field. Therefore, f is irreducible if and only if the quotient ring $K[x]/(f)$ is a field.

Example 2.8. The polynomial $x^2 + x + 1$ is irreducible over $\mathbb{Z}_2$ as it has no roots in $\mathbb{Z}_2$,

$$0^2 + 0 + 1 \equiv 1 \not\equiv 0 \pmod 2, \quad 1^2 + 1 + 1 \equiv 1 \not\equiv 0 \pmod 2$$

Therefore, the quotient ring $\mathbb{Z}_2[x]/(x^2 + x + 1)$ is a field.

Question 2.9. Explicitly describe the elements of $\mathbb{Z}_2[x]/(x^2 + x + 1)$.

Solution. Let $f(x) \in \mathbb{Z}_2[x]$ then by the division algorithm, we can write $f(x) = q(x)(x^2 + x + 1) + r(x)$ where $q(x), r(x) \in \mathbb{Z}_2[x]$, $r(x) = 0$ or $1 < \deg r(x) < \deg f(x)$. Then we have

$$\begin{aligned}
\mathbb{Z}_2[x]/(x^2 + x + 1) &= \left\{ f(x) + (x^2 + x + 1) : f(x) \in \mathbb{Z}_2[x] \right\} \\
&= \left\{ r(x) + (x^2 + x + 1) : r(x) = ax + b,\ a, b \in \mathbb{Z}_2 \right\} \\
&= \left\{ ax + b + (x^2 + x + 1) : a, b \in \mathbb{Z}_2 \right\} \\
&= \left\{ 0 + (x^2 + x + 1), 1 + (x^2 + x + 1), x + (x^2 + x + 1), x + 1 + (x^2 + x + 1) \right\} \\
&= \left\{ 0 + I, 1 + I, x + I, x + 1 + I \right\}.
\end{aligned}$$

Let $I = (x^2 + x + 1)$ be the ideal generated by $x^2 + x + 1$ in $\mathbb{Z}_2[x]$. We then have the following Cayley tables,

$+$	$0+I$	$1+I$	$x+I$	$x+1+I$
$0+I$	$0+I$	$1+I$	$x+I$	$x+1+I$
$1+I$	$1+I$	$0+I$	$x+1+I$	$x+I$
$x+I$	$x+I$	$x+1+I$	$0+I$	$1+I$
$x+1+I$	$x+1+I$	$x+I$	$1+I$	$0+I$

and

$\cdot$	$0+I$	$1+I$	$x+I$	$x+1+I$
$0+I$	$0+I$	$0+I$	$0+I$	$0+I$
$1+I$	$0+I$	$1+I$	$x+I$	$x+1+I$
$x+I$	$0+I$	$x+I$	$x+1+I$	$1+I$
$x+1+I$	$0+I$	$x+1+I$	$1+I$	$x+I$

where recall that $(a+I)(b+I) = ab+I$ so, for instance,

$$(x+I)(x+I) = x^2 + I = x + 1 + I,$$

$$(x+1+I)(x+I) = x^2 + x + I = 1 + I,$$

$$(x+1+I)(x+1+I) = (x+1)^2 + I$$

$$= x^2 + 2x + 1 + I = x^2 + 1 + I = x + I,$$

and so on, with the working principle[1] being to take $x^2 + x + 1$ as 0. From the Cayley tables, we can see that every non-zero element has its inverse

$$(1+I)^{-1} = 1 + I,$$
$$(x+I)^{-1} = x + 1 + I,$$
$$(x+1+I)^{-1} = x + I.$$

Hence, $\mathbb{Z}_2[x]/(x^2+x+1)$ is a field with 4 elements. ■

Theorem 2.10. Let p be a prime and $f(x) \in \mathbb{Z}_p[x]$ be an irreducible polynomial of degree n. Then we have

$$|\mathbb{Z}_p[x]/(f)| = p^n.$$

Proof. Let $g(x) \in \mathbb{Z}_p[x]$ then by the division algorithm, we can write $g(x) = q(x)(f)+r(x)$ where $q(x), r(x) \in \mathbb{Z}_p[x]$, $r(x) = 0$ or $1 < \deg r(x) < \deg f(x) = n$. Then we have $g(x) + (f) = r(x) + f$, so

$$\begin{aligned}\mathbb{Z}_p[x]/(f) &= \{g(x) + (f) : g(x) \in \mathbb{Z}_p[x]\} \\ &= \left\{r(x) + (f) : r(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1},\ a_i \in \mathbb{Z}_p\right\} \\ &= \left\{a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + (f) : a_i \in \mathbb{Z}_p\right\}\end{aligned}$$

and we have p choices for each a_i. Hence, $|\mathbb{Z}_p[x]/(f)| = p^n$. ■

Lemma 2.11. For a finite field F, $\text{char}(F) \neq 0$.

Proof. Let F be a finite field and let $\text{char}(F) = 0$. Then there is no positive integer n such that $n \cdot 1_F = 0$. Now, we can write

$$F = \{1_F, 1_F + 1_F, \dots\}$$

and as F is finite there exist positive integers $m < n$ such that $m \cdot 1_F = n \cdot 1_F$. Then $(n-m) \cdot 1_F = 0$. This contradicts the assumption that $\text{char}(F) = 0$. Hence, $\text{char}(F) \neq 0$. ■

Theorem 2.12. Every finite field is of order p^n for some prime p and positive integer n.

For proof, see section 2.1.3.

[1] In R/I, $a + I = b + I$ iff $a - b \in I$ iff $a \equiv b \pmod{I}$.

2.1.2 Prime subfield

Definition 2.13. The *prime subfield* of a field K is the smallest subfield of K, i.e. the intersection of all subfields of K, i.e. the subfield of K generated by 0_K and 1_K.

Theorem 2.14. The prime subfield of a field K is isomorphic to either $\mathbb{Q}$ (when $\text{char}(K) = 0$) or $\mathbb{Z}_p$ (when $\text{char}(K) = p$) for some prime p.

Proof. Define a map $\phi : \mathbb{Z} \to K$ by $\phi(n) = n \cdot 1_K$ for $n \in \mathbb{Z}$. Then ϕ is a ring homomorphism. If $\text{char}(K) = 0$, then $\ker \phi = \{0\}$ and hence ϕ is injective. Therefore, K contains an isomorphic copy of $\mathbb{Z}$ (by the first isomorphism theorem) and hence an isomorphic copy of $\mathbb{Q}$. If $\text{char}(K) = p$ for some prime p, then $\ker \phi = p\mathbb{Z}$ and hence by the first isomorphism theorem, $\mathbb{Z}/p\mathbb{Z} \cong \text{im}\, \phi$. So K contains an isomorphic copy of $\mathbb{Z}_p$. Therefore, the prime subfield of K is isomorphic to either $\mathbb{Q}$ or $\mathbb{Z}_p$. ■

Example 2.15. The prime subfield of both $\mathbb{Q}$ and $\mathbb{R}$ is $\mathbb{Q}$.

2.1.3 Field extension

Definition 2.16. A field F is said to be an *extension field* of a field K if K is a subfield of F or K is isomorphic to a subfield of F. We denote this by F/K. In particular, every field is an extension field of its prime subfield.

Every extension field F of K is a vector space over K with the scalar multiplication given by the field multiplication in F. The dimension of this vector space is called the *degree* of the extension and is denoted by $[F : K]$.

Example 2.17. $\mathbb{R}$ is an extension field of $\mathbb{Q}$ and $[\mathbb{R} : \mathbb{Q}] = \infty$. The field $\mathbb{C}$ is an extension field of $\mathbb{R}$ and $[\mathbb{C} : \mathbb{R}] = 2$.

Example 2.18. We saw $\mathbb{Z}_2[x]/(x^2 + x + 1) = \left\{ax + b + (x^2 + x + 1) : a, b \in \mathbb{Z}_2\right\}$, so $\mathbb{Z}_2[x]/(x^2 + x + 1)$ is an extension field of $\mathbb{Z}_2$ with degree $[\mathbb{Z}_2[x]/(x^2 + x + 1) : \mathbb{Z}_2] = 2$.

Proof of theorem 2.12. Let F be a finite field. Then $\text{char}(F) = p$ for some prime p. Then the prime subfield of F is isomorphic to $\mathbb{Z}_p$. Consider F as an extension of $\mathbb{Z}_p$ and let $[F : \mathbb{Z}_p] = n$ be the dimension of this vector space. Then $F \cong \mathbb{Z}_p \times \cdots \times \mathbb{Z}_p$ (n times), so $|F| = p^n$. ■

Theorem 2.19. Any ring homomorphism with domain a field is either a zero morphism or a monomorphism.

Proof. Let $\phi : F \to F'$ be a ring homomorphism with domain a field F. Then $\ker \phi$ is an ideal of F. Since F is a field, the only ideals of F are $\{0\}$ and F. If $\ker \phi = F$, then ϕ is the zero morphism. If $\ker \phi = \{0\}$, then ϕ is injective and hence a monomorphism. ■

Question 2.20. Make a field of order 343, two fields of order 8 and two fields of order 9.

Solution. The polynomial $x^3 - 2$ is irreducible in $\mathbb{Z}_7[x]$, so $\mathbb{Z}_7[x]/(x^3 - 2)$ is a field of order $7^3 = 343$.

Two fields of order 8 are $\mathbb{Z}_2[x]/(x^3 + x + 1)$ and $\mathbb{Z}_2[x]/(x^3 + x^2 + 1)$ as both $x^3 + x + 1$ and $x^3 + x^2 + 1$ are irreducible over $\mathbb{Z}_2$. Two fields of order 9 are $\mathbb{Z}_3[x]/(x^2 + 1)$ and $\mathbb{Z}_3[x]/(x^2 + x + 2)$ as both $x^2 + 1$ and $x^2 + x + 2$ are irreducible over $\mathbb{Z}_3$. ■

2.2 8 May

2.2.1 Algebraic and transcendental number

Definition 2.21. Suppose K/F is an extension field. Then $\alpha \in K$ is called *algebraic over* F if there exists a nonzero polynomial $f(x) \in F[x]$ such that $f(\alpha) = 0$. Otherwise, α is called *transcendental over* F.

Example 2.22. $\sqrt{2} \in \mathbb{R}$ is algebraic over $\mathbb{Q}$ since it is a root of $x^2 - 2 \in \mathbb{Q}[x]$. On the other hand, $\pi \in \mathbb{R}$ is transcendental over $\mathbb{Q}$ since it is not a root of any nonzero polynomial in $\mathbb{Q}[x]$. But π is algebraic over $\mathbb{R}$ since it is a root of $x - \pi \in \mathbb{R}[x]$.

Definition 2.23. The algebraic elements of $\mathbb{C}/\mathbb{Q}$ are called *algebraic numbers*, and the transcendental elements of $\mathbb{C}/\mathbb{Q}$ are called *transcendental numbers*.

$\alpha \in \mathbb{C}$ is called an *algebraic integer* if it is a root of a monic polynomial in $\mathbb{Z}[x]$.

Exercise 2.24. If $\alpha \in \mathbb{Q}$ is an algebraic integer, then show that $\alpha \in \mathbb{Z}$.

Solution. As $\alpha \in \mathbb{Q}$, we can write $\alpha = \frac{a}{b}$ for some $a, b \in \mathbb{Z}$ with $b \neq 0$. If α is an algebraic integer, then it is a root of a monic polynomial $f(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$ with coefficients in $\mathbb{Z}$. Substituting $\alpha = \frac{a}{b}$ into this equation and multiplying through by b^n gives

$$a^n + c_{n-1}a^{n-1}b + \cdots + c_1 ab^{n-1} + c_0 b^n = 0.$$

This implies that $a^n = -(c_{n-1}a^{n-1}b + \cdots + c_1 ab^{n-1} + c_0 b^n)$. Since the right-hand side is divisible by b, it follows that a^n is divisible by b. But since $\gcd(a, b) = 1$, we must have $b = 1$, so $\alpha = a \in \mathbb{Z}$. ■

2.2.2 Extension fields generated by subsets

Definition 2.25. Let K/F be an extension field and $S \subseteq K$. Then

- $F(S)$ is the smallest subfield of K containing $F \cup S$. It is called the *subfield generated by* S *over* F.
- $F[S]$ is the smallest subring of K containing $F \cup S$. It is called the *subring generated by* S *over* F.

Note that $F[S]$ is necessarily an integral domain. If S is finite, say $S = \{\alpha_1, \ldots, \alpha_n\}$, then we write $F[S] = F[\alpha_1, \ldots, \alpha_n]$ and $F(S) = F(\alpha_1, \ldots, \alpha_n)$.

Example 2.26. Consider the field extension $\mathbb{R}/\mathbb{Q}$ and let $S = \{\sqrt{2}\}$. Then the subring generated by S over $\mathbb{Q}$ is $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$. Also, if $\alpha \in \mathbb{Q}(\sqrt{2})$ then

$$\begin{aligned}
\alpha &= \frac{a + b\sqrt{2}}{c + d\sqrt{2}} \\
&= \frac{(a + b\sqrt{2})(c - d\sqrt{2})}{(c + d\sqrt{2})(c - d\sqrt{2})} \\
&= \frac{(ac - 2bd) + (bc - ad)\sqrt{2}}{c^2 - 2d^2} \\
&= \frac{ac - 2bd}{c^2 - 2d^2} + \frac{bc - ad}{c^2 - 2d^2}\sqrt{2}.
\end{aligned}$$

Thus, the subfield generated by S over $\mathbb{Q}$ is $\mathbb{Q}(\sqrt{2}) = \left\{a + b\sqrt{2} : a, b \in \mathbb{Q}\right\} = \mathbb{Q}[\sqrt{2}]$ in this case.

Theorem 2.27. Let F/K be an extension field, $X \subset F$ and $u, u_i \in F$.

1. the subring $K[u]$ is the set of all elements of the form $f(u)$ where $f \in K[x]$;
2. the subring $K[u_1, \ldots, u_m]$ is the set of all elements of the form $g(u_1, \ldots, u_m)$ where $g \in K[x_1, \ldots, x_m]$;
3. the subring $K[X]$ is the set of all elements of the form $h(u_1, \ldots, u_n)$ where each $u_i \in X$, n is a positive integer and $h \in K[x_1, \ldots, x_n]$;
4. the subfield $K(u)$ is the set of all elements of the form $\dfrac{f(u)}{g(u)} = f(u)g(u)^{-1}$ where $f, g \in K[x]$ and $g(u) \neq 0$;
5. the subfield $K(u_1, \ldots, u_m)$ is the set of all elements of the form
$$\frac{h(u_1, \ldots, u_m)}{k(u_1, \ldots, u_m)} = h(u_1, \ldots, u_m)k(u_1, \ldots, u_m)^{-1}$$
where $h, k \in K[x_1, \ldots, x_m]$ and $k(u_1, \ldots, u_m) \neq 0$;
6. the subfield $K(X)$ is the set of all elements of the form $\dfrac{f(u_1, \ldots, u_n)}{g(u_1, \ldots, u_n)} = f(u_1, \ldots, u_n)g(u_1, \ldots, u_n)^{-1}$ where each $u_i \in X$, n is a positive integer and $f, g \in K[x_1, \ldots, x_n]$ with $g(u_1, \ldots, u_n) \neq 0$.

Proof. See [Hun74, Theorem 5.1.3, p. 232] ∎

In example 2.26, we saw that $\mathbb{Q}[\sqrt{2}] = \mathbb{Q}(\sqrt{2})$. But when is $F[\alpha] = F(\alpha)$ true?

Theorem 2.28. Let K/F be a field extension and $\alpha \in K$. Let $\psi : F[x] \to F[\alpha]$ be defined by $\psi(f(x)) = f(\alpha)$. Then ψ is a ring homomorphism.

Proof. Let $f, g \in F[x]$ and $c \in F$. Then
$$\begin{aligned} \psi(f+g) &= (f+g)(\alpha) = f(\alpha) + g(\alpha) = \psi(f) + \psi(g), \\ \psi(fg) &= (fg)(\alpha) = f(\alpha)g(\alpha) = \psi(f)\psi(g), \\ \psi(cf) &= (cf)(\alpha) = cf(\alpha) = c\psi(f). \end{aligned}$$
Thus, ψ is a ring homomorphism. ∎

Remark 2.29. Here,
$$\begin{aligned} \ker \psi &= \{f \in F[x] : \psi(f(x)) = 0\} \\ &= \{f \in F[x] : f(\alpha) = 0\} \\ &= \text{the set of all polynomials in } F[x] \text{ that have } \alpha \text{ as a root} \\ &= \text{the set of all polynomials in } F[x] \text{ that annihilate } \alpha. \end{aligned}$$
As $F[x]$ is PID, $\ker \psi$ is a principal ideal of $F[x]$. If $\ker \psi = \{0\}$, then ψ is injective and hence an isomorphism. Now,

- if α is transcendental over F, then $\ker \psi = \{0\}$,
- if α is algebraic over F, then $\ker \psi \neq \{0\}$.

So when α is transcendental, ψ is an isomorphism.

Theorem 2.30. Suppose α is algebraic over F. Then $F[\alpha]$ is a field and hence $F[\alpha] = F(\alpha)$. Further, $\ker \psi$ has a unique monic generator.

Proof. By the first isomorphism theorem,

$$F[x]/\ker \psi \cong F[\alpha].$$

Since $F[\alpha]$ is an integral domain, $F[x]/\ker \psi$ is also an integral domain. Since $F[x]$ is a commutative ring with identity 1, $\ker \psi$ is a prime ideal of $F[x]$. Since $F[x]$ is a PID, $\ker \psi$ is generated by a single element, say $f(x) \in F[x]$. Since in a PID, irreducible elements and primes coincide, $f(x)$ is an irreducible element of $F[x]$. Thus, $\ker \psi = (f(x))$ where $f(x) \in F[x]$. Now, since $F[x]$ is PID, $(f(x))$ is maximal, so $F[x]/\ker \psi = F[x]/(f(x))$ is a field. Hence, $F[\alpha]$ is a field. Now $F(\alpha) \subseteq F[\alpha]$, as $F(\alpha)$ is the smallest subfield of K containing α and $F[\alpha] \subseteq F(\alpha)$ by definition. Hence, $F[\alpha] = F(\alpha)$. ■

Definition 2.31. The unique monic generator of $\ker \psi$ in theorem 2.30 is called the *minimal polynomial of α over F*. It is an irreducible polynomial.

Example 2.32. $x^2 - 2$ is the minimal polynomial of $\sqrt{2}$ over $\mathbb{Q}$. $x - \sqrt{2}$ is the minimal polynomial of $\sqrt{2}$ over $\mathbb{R}$.

Theorem 2.33. Let F be a field and $f(x) \in F[x]$ be an irreducible polynomial. Then F has an extension field containing a root of $f(x)$, namely

$$K = F[x]/(f).$$

Theorem 2.34. Let K/F be a field extension and $\alpha \in K$ be algebraic over F. Then

$$\{1, \alpha, \ldots, \alpha^{n-1}\} \subseteq F(\alpha)$$

is a basis of $F(\alpha)$ over F and hence $[F(\alpha) : F] = n$ where n is the degree of the minimal polynomial of α over F.

Exercise 2.35. Compute:

1. $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}]$.
2. $[\mathbb{Q}(i) : \mathbb{Q}]$.
3. $[\mathbb{R}(i) : \mathbb{R}]$.
4. $[\mathbb{Z}_3(i) : \mathbb{Z}_3]$.
5. $[\mathbb{Z}_3[x]/(x^2+1) : \mathbb{Z}_3]$.

Solutions. $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = \deg\left(x^2 - 2\right) = 2$. Also, $[\mathbb{Q}(i) : \mathbb{Q}] = [\mathbb{R}(i) : \mathbb{R}] = [\mathbb{Z}_3(i) : \mathbb{Z}_3] = \deg\left(x^2 + 1\right) = 2$. Finally, we have

$$\mathbb{Z}_3[x]/(x^2+1) \cong \mathbb{Z}_3(i),$$

hence $[\mathbb{Z}_3[x]/(x^2+1) : \mathbb{Z}_3] = 2$ as well. ■

2.3 14 May

Recall that in a field extension K/F we defined the *evaluation map* $\psi : F[x] \to F[\alpha]$ by $f(x) \mapsto f(\alpha)$. We saw that $\ker \psi = \{0\}$ iff α is transcendental, and

$$\ker \psi = \{f(x) \in F[x] : f(\alpha) = 0\} \neq \{0\} \iff \alpha \text{ is algebraic over } F.$$

The unique monic generator of this ideal $\ker \psi$ is the *minimal polynomial* $m(x)$ of α over F. It is an irreducible polynomial. α algebraic implies $F(\alpha) = F[\alpha]$ and

$$F[x]/(m(x)) \cong F[\alpha].$$

If α, β are both algebraic and the root of the same minimal polynomial, then $F(\alpha) \cong F(\beta)$. If α, β are both transcendental, then also $F(\alpha) \cong F(\beta)$.

2.3.1 Finitely generated, simple extension

Definition 2.36. If S is finite, then we say $F(S)$ is a *finitely generated extension* of F. If S is a singleton, then $F(S)$ is a *simple extension.*

Theorem 2.37 (Tower law)**.** Let K/E and E/F be field extensions. Then

$$[K : F] = [K : E][E : F].$$

Furthermore, $[K : F]$ is finite iff $[K : E]$ and $[E : F]$ are finite.

Proof. See [Hun74, Theorem 5.1.2, p. 231]. ■

2.3.2 Algebraic extension

Definition 2.38. An extension field K of F is an *algebraic extension* of F if every element of K is algebraic over F. Similarly, a *transcendental extension* is one where every element is transcendental instead.

Exercise 2.39. Is x algebraic over F ?

Solution. No. If possible let x be algebraic over F. Then there is some nonzero polynomial $f \in F[x]$ such that

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n = 0 \implies a_0 = a_1 = \cdots = a_n = 0$$

so $f(x)$ is the zero polynomial, an absurdity. So, x must be transcendental over F. ■

Theorem 2.40. Every finite dimensional extension K/F is finitely generated and algebraic over F.

Proof. If $[K : F] = n$ and $u \in K$, then the set of $n + 1$ elements $\{1_F, u, u^2, \ldots, u^n\}$ must be linearly dependent. Hence there are $a_i \in F$, not all zero, such that

$$a_0 + a_1 u + a_2 u^2 + \cdots + a_n u^n = 0$$

which implies u is algebraic over F. As u was arbitrary, K is algebraic over F. If $\{v_1, \ldots, v_n\}$ is a basis of K over F, then it is easy to see that $K = F(v_1, \ldots, v_n)$. ■

Example 2.41. As x is not algebraic over F, $F(x)$ is not an algebraic extension of F. So $F(x)$ is not a finite extension of F, even though $F(x)$ is a *finitely generated* (even *simple*) extension of F.

Theorem 2.42. If F is an extension field of K and X is a subset of F such that $F = K(X)$ and every element of X is algebraic over K, then F is an algebraic extension of K. If X is a finite set, then F is finite dimensional over K.

Proof. If $v \in F$ then $v \in K(u_1, \ldots, u_n)$ for some $u_i \in X$ and there is a tower of subfields

$$K \subset K(u_1) \subset K(u_1, u_2) \subset \cdots \subset K(u_1, \ldots, u_{n-1}) \subset K(u_1, \ldots, u_n).$$

Since u_i is algebraic over K, it is necessarily algebraic over $K(u_1, \ldots, u_{i-1})$ for each $i \geq 2$, say of degree r_i. Since $K(u_1, \ldots, u_{i-1})(u_i) = K(u_1, \ldots, u_i)$, we have $[K(u_1, \ldots, u_i) : K(u_1, \ldots, u_{i-1})] = r_i$. Let r_1 be the degree of u_1 over K. Then repeated application of tower law yields $[K(u_1, \ldots, u_n) : K] = r_1 r_2 \cdots r_n$. Then by theorem 2.40, $K(u_1, \ldots, u_n)$, and hence v, is algebraic over K. As $v \in F$ was arbitrary, F is algebraic over K. If $X = \{u_1, \ldots, u_n\}$ is finite, the same proof with $F = K(u_1, \ldots, u_n)$ shows that $[F : K] = r_1 r_2 \cdots r_n$ is finite. ■

Theorem 2.43. Let F be an extension field of K and E the set of all elements of F which are algebraic over K. Then E is a subfield of F, which is, of course, algebraic over K.

Proof. If $u, v \in E$ then $K(u, v)$ is an algebraic extension field of K by theorem 2.42. Therefore, since $u - v$ and uv^{-1} $(v \neq 0)$ are in $K(u, v)$, $u - v$ and $uv^{-1} \in E$. This implies that E is a field. ■

Clearly the subfield E is the unique maximal algebraic extension of K contained in F.

Example 2.44. Not every algebraic extension is finite. For instance, the set of algebraic numbers $\mathbb{A}$ is an infinite-dimensional algebraic field extension of $\mathbb{Q}$.

2.3.3 Splitting field

Definition 2.45. Let $f(x) \in F[x]$. A *splitting field* of $f(x)$ over F is the smallest extension K of F containing all the roots of $f(x)$, i.e. if $f(x)$ 'splits' completely into linear factors in $K[x]$ and $f(x)$ does not split completely into linear factors over any proper subfield of K containing F.

Example 2.46. The splitting field of $x^2 + 1$ over $\mathbb{R}$ is $\mathbb{C}, \mathbb{R}(i), \mathbb{R}[x]/(x^2 + 1)$ all three being isomorphic (not unique). The splitting field of $x^2 - 2$ over $\mathbb{R}$ is $\mathbb{R}$ itself (over $\mathbb{Q}$ the splitting field is $\mathbb{Q}(\sqrt{2})$). The splitting field of $\mathbb{Q}$ is $\mathbb{A}$, for $\mathbb{R}$ is $\mathbb{C}$.

Splitting field is always an algebraic extension. Every field has a proper extension ($F[x]$), but not necessarily a proper *algebraic* extension ($\mathbb{C}$). Splitting field of all polynomials over F is the *algebraic closure* of F.

2.3.4 Algebraic closure

Definition 2.47. The field $\overline{F}$ is an *algebraic closure* of F if $\overline{F}$ is algebraic over F and if every polynomial $f(x) \in F[x]$ splits completely over $\overline{F}$, so that $\overline{F}$ can be said to contain all the elements algebraic over F.

Example 2.48. $\overline{\mathbb{Q}} = \mathbb{A}$, $\overline{\mathbb{R}} = \mathbb{C}$.

Definition 2.49. A field F is *algebraically closed* if every non-constant polynomial with coefficients in F has a root in F.

Example 2.50. $\mathbb{C}$ is algebraically closed.

Theorem 2.51. A finite field cannot be algebraically closed.

Proof. Let $|F| = m$, and $F = \{\alpha_1, \ldots, \alpha_m\}$. Consider the polynomial

$$f(x) = (x - \alpha_1) \cdots (x - \alpha_m) + \alpha \in F[x]$$

for some $\alpha \in F \setminus \{0\}$. Then for any $\alpha_i \in F$, $f(\alpha_i) = \alpha \neq 0$. Thus $f(x)$ has no roots in F, thus F cannot be algebraically closed. ■

2.3.5 Normal and separable extension

Definition 2.52. If K is an algebraic extension over F which is the splitting field of a collection of polynomials $f(x) \in F[x]$ over F, then K is called a *normal extension* of F.

Definition 2.53. An irreducible polynomial $f(x) \in F[x]$ is *separable* if in some splitting field of f over F every root of f is a simple root (multiplicity $= 1$). If K is an extension of F and $u \in K$ is algebraic over F, then u is said to be *separable* over F if its irreducible polynomial is separable. If every element of K is separable over F, then K is a *separable extension* of F.

Informally, a normal extension means if it has one root then there are other roots. A separable extension means all the elements are algebraic and the minimal polynomial is separable i.e. roots are distincts.

Example 2.54. Extension of any field of characteristic 0 is separable.

2.3.6 Galois extension and Galois group of extension field

Definition 2.55. The *Galois group* of K/F is the group of F-fixing automorphisms of K,

$$\mathrm{Aut}_F(K) = \{f \in \mathrm{Aut}(K) : f(\alpha) = \alpha \quad \forall \alpha \in F\}.$$

Definition 2.56. The *Galois group* of a polynomial $f(x) \in F[x]$ is defined to be the Galois group of K over F, where K is the splitting field of $f(x) \in F[x]$ over F.

Exercise 2.57. Find the Galois group of $\mathbb{Q}(\sqrt{2}, \sqrt{3})/\mathbb{Q}$.

Definition 2.58. A (finite) normal, separable extension is called a *Galois extension.*

Every automorphism of a field fixes the prime subfield.

Theorem 2.59 (The Fundamental Theorem of Galois Theory). If F is a Galois extension of K, then there is a one-to-one correspondence between the set of all intermediate fields of the extension and the set of all subgroups of the Galois group $\mathrm{Aut}_K(F)$ (given by $E \mapsto E' = \mathrm{Aut}_E(F)$) such that

1. the relative dimension of two intermediate fields is equal to the relative index of the corresponding subgroups; in particular, $\mathrm{Aut}_K(F)$ has order $[F : K]$;
2. F is Galois over every intermediate field E, but E is Galois over K if and only if the corresponding subgroup $E' = \mathrm{Aut}_E(F)$ is normal in $G = \mathrm{Aut}_K(F)$; in this case G/E' is (isomorphic to) the Galois group $Aut_K(E)$ of E over K.

Bibliography

[Tol69] Leo Tolstoy. *War and Peace*. Penguin Classics, 1869.

[HK70] Kenneth Hoffman and Raymond Kunze. *Linear Algebra*. Prentice Hall, 1970.

[Hun74] Thomas W. Hungerford. *Algebra*. Springer, 1974.

List of Definitions