1 Vector spaces

1.1 Definition (Vector space)

Let V be a set with a binary operation $+$, F a field, and $(c, v) \mapsto cv$ be a mapping from $F \times V$ into V. Then V is called a *vector space over* F (or a *linear space over* F) if

(i) $u + v = v + u$	for all $u, v \in V$
(ii) $u + (v + w) = (u + v) + w$	for all $u, v, w \in V$
(iii) $\exists 0 \in V : v + 0 = v$	for all $v \in V$
(iv) $\forall u \in V \; \exists - u \in V : \; u + (-u) = 0$	
(v) $(ab)v = a(bv)$	for all $a, b \in F$ and $v \in V$
(vi) $(a + b)v = av + bv$	for all $a, b \in F$ and $v \in V$
(vii) $a(u + v) = au + av$	for all $a \in F$ and $u, v \in V$
(viii) $1v = v$	for all $v \in V$

Note that (i)-(iv) mean that $(V, +)$ is an abelian group. The mapping $(c, v) \mapsto cv$ is called *scalar multiplication*, the elements of F *scalars*, the elements of V *vectors*. The vector $0 \in V$ is called *zero vector* (do not confuse it with 0 in F).

In this course we practically study two cases: $F = \mathbb{R}$ and $F = \mathbb{C}$. We call them *real vector spaces* and *complex vector spaces*, respectively.

1.2 Examples of vector spaces

a) F field, $F^n = \{(x_1, \ldots, x_n) : x_i \in F\}$ n-tuple space. The operations are defined by

$$\begin{aligned} (x_1, \ldots, x_n) + (y_1, \ldots, y_n) &= (x_1 + y_1, \ldots, x_n + y_n) \\ c(x_1, \ldots, x_n) &= (cx_1, \ldots, cx_n) \end{aligned}$$

Note that $\mathbb{R}^2$ and $\mathbb{R}^3$ can be interpreted as ordinary plane and space, respectively, in rectangular coordinate systems.

b) F field, S set, $V = \{f : S \to F\}$ (arbitrary functions from S to F). The operations are defined by

$$\begin{aligned} (f + g)(s) &= f(s) + g(s) \qquad \text{for } \; f, g \in V, \; s \in S \\ (cf)(s) &= cf(s) \qquad\quad \text{for } \; c \in F, \; f \in V, \; s \in S \end{aligned}$$

c) $C[0, 1]$ - the space of continuous functions on $[0, 1]$; $C^r[0, 1]$ - the space of functions on $[0, 1]$ that have (at least) r continuous derivatives. The operations are defined as in (b). Here the closed interval $[0, 1]$ may be replaced by any open or closed interval in $\mathbb{R}$ or the entire real line.

d) $P_n(\mathbb{R})$ - the space of all real polynomials of degree $\leq n$ with real coefficients. Also, $P(\mathbb{R})$ the space of all real polynomials (of any degree) with real coefficients.

e) F field, $F^{m\times n}$ the space of $m \times n$ matrices with components in F. For $A \in F^{m\times n}$ denote by A_{ij} the component of the matrix A in the ith row and the jth column, $i \in [1, m]$, $j \in [1, n]$. Then the operations are defined by

$$\begin{aligned} (A+B)_{ij} &= A_{ij} + B_{ij} \\ (cA)_{ij} &= cA_{ij} \end{aligned}$$

1.3 Basic properties of vector spaces

The following properties follow easily from the definition:

$$\begin{aligned} c0 &= 0 \\ 0v &= 0 \\ (-1)v &= -v \\ cv = 0 &\Rightarrow c = 0 \text{ or } v = 0 \end{aligned}$$

1.4 Definition (Linear combination)

Let $v_1, \ldots, v_n \in V$ and $c_1, \ldots, c_n \in F$. Then the vector $c_1v_1 + \cdots + c_nv_n$ is called a *linear combination* of $v_1, \ldots, v_n$.

1.5 Definition (Subspace)

A subset of a vector space V which is itself a vector space with respect to the operations in V is called a *subspace* of V.

1.6 Theorem (Subspace criterion)

A non-empty subset W of a vector space V is a subspace of V if and only if

$$au + bv \in W \quad \text{for all } a, b \in F,\ u, v \in W$$

1.7 Examples of subspaces

a) $C^r[0, 1]$ is a subspace of $C[0, 1]$ for all $r \geq 1$.

b) $P_n(\mathbb{R})$ is a subspace of $P(\mathbb{R})$.

c) the space of symmetric $n \times n$ matrices (i.e. such that $A_{ij} = A_{ji}$) is a subspace of $F^{n\times n}$, the space of all $n \times n$ matrices.

d) the space of antisymmetric $n \times n$ matrices (i.e. such that $A)ij = -A_{ji}$) is a subspace of $F^{n\times n}$, the space of all $n \times n$ matrices. Note that for any antisymmetric matrix $A_{ii} = 0$, i.e. the main diagonal is zero.

e) $W = \{0\}$ is a trivial subspace in any vector space (it only contains the zero vector).

f) If C is a collection of subspaces of V, then their intersection

$$\cap_{W\in C} W$$

is a subspace of V. This is not true for unions.

1.8 Definition (Span)

Let A be a subset of a vector space V. Denote by C the collection of all subspaces of V that contain A. Then the *span* of A is the subspace of V defined by

$$\operatorname{span} A = \cap_{W \in C} W$$

1.9 Theorem

$\operatorname{span} A$ is the set of all linear combinations of elements of A.

1.10 Example

If $W_1, W_2 \subset V$ are subspaces, then

$$\operatorname{span}(W_1 \cup W_2) = W_1 + W_2$$

where

$$W_1 + W_2 \stackrel{\text{def}}{=} \{u + v : u \in W_1,\, v \in W_2\}$$

1.11 Linear independence, linear dependence

A finite set of vectors $v_1, \ldots, v_n$ is said to be *linearly independent* (loosely, one can say that the vectors $v_1, \ldots, v_n$ are linearly independent), if $c_1 v_1 + \cdots + c_n v_n = 0$ implies $c_1 = \cdots = c_n = 0$. In other words, all nontrivial linear combinations are different from zero. Otherwise, the set $\{v_1, \ldots, v_n\}$ is said to be linearly dependent.

An infinite set of vectors is said to be linearly independent if every finite subset of it is linearly independent.

1.12 Examples

a) in the space $\mathbb{R}^n$, the vectors

$$\begin{aligned} e_1 &= (1, 0, 0, \ldots, 0) \\ e_2 &= (0, 1, 0, \ldots, 0) \\ &\ldots \\ e_n &= (0, 0, 0, \ldots, 1) \end{aligned}$$

are linearly independent.

b) In $P_n(\mathbb{R})$, the polynomials $1, x, x^2, \ldots, x^n$ are linearly independent. In $P(\mathbb{R})$, the infinite collection of polynomials $\{x^i\}$, $0 \le i < \infty$, is linearly independent.

c) If the set of vectors contains a zero vector, it is linearly dependent.

d) In $\mathbb{R}^2$ (=the ordinary xy plane), two vectors are linearly dependent iff they are collinear (parallel). Any three or more vectors in $\mathbb{R}^2$ are linearly dependent.

e) In $\mathbb{R}^3$ (=the ordinary space) three vectors are linearly dependent iff they are coplanar (belong in one plane). Any four or more vectors are linearly dependent.

1.13 Theorem

(a) If a set S of vectors is linearly dependent, then some vector $v \in S$ is a linear combination of other vectors $v_1, \ldots, v_m \in S$, i.e. $v = c_1 v_1 + \cdots + c_m v_m$ and $v \neq v_i$ for $1 \leq i \leq m$. In that case $\operatorname{span} S = \operatorname{span} S \setminus \{v\}$.

(b) Let $V \neq \{0\}$. If a finite subset $S \subset V$ of vectors spans V, then there is a linearly independent subset $S' \subset S$ that also spans V.

1.14 Basis

A *basis* of V is a linearly independent set $B \subset V$ which spans the entire V, i.e. $\operatorname{span} B = V$. The space V is said to be finite dimensional if it has a finite basis.

1.15 Examples

a) The vectors $\{e_1, \ldots, e_n\}$ make a basis is $\mathbb{R}^n$.

b) The polynomials $1, x, \ldots, x^n$ make a basis in $P_n(\mathbb{R})$. The infinite collection of polynomials $\{x^i\}_{i=0}^{\infty}$ make a basis in $P(\mathbb{R})$.

1.16 Theorem

Let V be spanned by $u_1, \ldots, u_m$, and let $v_1, \ldots, v_n$ be linearly independent in V. Then $n \leq m$.

1.17 Corollary

If $\{u_1, \ldots, u_m\}$ and $\{v_1, \ldots, v_n\}$ are two bases of V, then $m = n$.

1.18 Definition (Dimension)

The dimension of a finite-dimensional vector space V is the number of vectors in any basis of V. It is denoted by $\dim V$.

The trivial vector space $\{0\}$, the one consisting of a single zero vector, has no bases, and we define $\dim\{0\} = 0$.

1.19 Examples

a) $\dim F^n = n$

b) $\dim P_n(\mathbb{R}) = n + 1$

c) $\dim F^{m \times n} = mn$

d) $P(\mathbb{R})$ is not a finite-dimensional space

e) Let $A \in \mathbb{R}^{m \times n}$, $S = \{X \in \mathbb{R}^{n \times 1} : \ AX = 0\}$ (the solution space of $AX = 0$), and R is the row echelon matrix equivalent to A. Then $\dim S = n - r$, where r is the number of non-zero rows of R.

1.20 Theorem

Let S be a linearly independent subset of V. Suppose that $u \in V$ is not contained in $\operatorname{span} S$. Then the set $S \cup \{u\}$ is linearly independent.

1.21 Theorem

Let V be finite dimensional, $W \subset V$ a subspace, and S a linearly independent subset of W. Then S can be extended to a finite basis of W.

1.22 Corollary

Let V be a finite dimensional vector space.

a) If W is a subspace of V, then $\dim W \leq \dim V$, in particular W is finite dimensional.

b) If W is a proper subspace of V, then $\dim W < \dim V$.

c) Every nonempty linearly independent subset of V is a part of a basis of V.

1.23 Theorem

Let W_1 and W_2 be finite dimensional subspaces of V. Then

$$\dim W_1 + \dim W_2 = \dim (W_1 \cap W_2) + \dim (W_1 + W_2)$$

In particular, $W_1 + W_2$ is finite dimensional.

1.24 Lemma

If $\{u_1, \ldots, u_n\}$ is a basis for V, then for any $v \in V$ there exist unique scalars $c_1, \ldots, c_n$ such that

$$v = c_1 u_1 + \cdots + c_n u_n$$

1.25 Definition (Coordinates)

Let $B = \{u_1, \ldots, u_n\}$ be an ordered basis of V. If $v = c_1 u_1 + \cdots + c_n u_n$, then $(c_1, \ldots, c_n)$ are the *coordinates* of the vector v with respect to the basis B. Notation:

$$(v)_B = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$$

is the coordinate (row) vector of v relative B.

1.26 Examples

a) The canonical coordinates of a vector $v = (x_1, \ldots, x_n) \in F^n$ in the standard (canonical) basis $\{e_1, \ldots, e_n\}$ are its components $x_1, \ldots, x_n$, since $v = x_1 e_1 + \cdots + x_n e_n$.

b) The coordinates of a polynomial $p \in P_n(\mathbb{R})$ given by

$$p = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

in the basis $\{1, x, \ldots, x^n\}$ are its coefficients $a_0, a_1, \ldots, a_n$.

1.27 Coordinate mapping

Let $\dim V = n$ and $B = \{u_1, \ldots, u_n\}$ be an ordered basis of V. Then the mapping

$$M_B : \ V \to F^n \quad \text{given by} \quad v \mapsto (v)_B$$

is a bijection. For any $u, v \in V$ and $c \in F$ we have

$$\begin{aligned} M_B(u+v) &= M_B(u) + M_B(v) \\ M_B(cv) &= cM_B(v) \end{aligned}$$

Note: The mapping $M_B : V \to F^n$ not only is a bijection, but also preserves the vector operations. Since there is nothing else defined in V, we have a complete identity of V and F^n. Any property of V can be proven by first substituting F^n for V and then using the mapping M_B.

1.28 Theorem (Change of coordinates)

Let $B = \{u_1, \ldots, u_n\}$ and $B' = \{u'_1, \ldots, u'_n\}$ be two ordered bases of V. Denote by $P_{B',B}$ the $n \times n$ matrix with jth column given by

$$(P_{B',B})_j = (u'_j)_B$$

for $j = 1, \ldots, n$. Then the matrix $P_{B',B}$ is invertible with $P_{B',B}^{-1} = P_{B,B'}$ and

$$(v)_B = P_{B',B}(v)_{B'}$$

for every $v \in V$.

1.29 Definition (Row space, row rank)

Let $A \in F^{m \times n}$. Denote by $v_i = (A_{i1}, \ldots, A_{in}) \in F^n$ the ith row vector of A. Then the subspace $\text{span}\{v_1, \ldots, v_m\}$ of F^n is called the *row space* of A, and $\dim(\text{span}\{v_1, \ldots, v_m\})$ is called the *row rank* of A.

1.30 Theorem

Row equivalent matrices have the same row space.

Note: Theorem 1.30 has a converse: if two matrices have the same row space, then they are row equivalent (a proof may be found in textbooks). We will not need this fact.

1.31 Theorem

If R is a row echelon matrix, then the non-zero row vectors of R make a basis of the row space of R.

Note: Every matrix is row equivalent to a row echelon matrix. Thus, Theorems 1.30 and 1.31 show how to find a basis of $\text{span}\{v_1, \ldots, v_m\}$ for a given set of vectors $v_1, \ldots, v_m \in F^n$. In particular, this gives the dimension of that subspace.

1.32 Definition (Independent subspaces, direct sums)

Let $W_1, \ldots, W_k$ be subspaces of a vector space V. We say that they are independent if, whenever $w_1 \in W_1, \ldots, w_k \in W_k$ with $w_1 + \cdots + w_k = 0$, then $w_1 = \cdots = w_k = 0$.

In this case, the sum $W_1 + \cdots + W_k$ is called a *direct sum* and denoted by $W_1 \oplus \cdots \oplus W_k$.

If $k = 2$ and $W_1 \oplus W_2 = V$, then the subspaces W_1 and W_2 are called *complementary subspaces* of V.

Note: For every $w \in W_1 \oplus \cdots \oplus W_k$ there exists a unique collection of vectors $w_i \in W_i$, $1 \leq i \leq k$, such that $w = w_1 + \cdots + w_k$.

1.33 Theorem

Let V be finite dimensional. Let $W_1, \ldots, W_k$ be subspaces of V. Then the following are equivalent:

(i) $W_1, \ldots, W_k$ are independent.

(ii) $W_i \cap (W_1 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_k) = \{0\}$ for every $i = 1, \ldots, k$.

(iii) If B_i is a basis of W_i, then $B_1 \cup \cdots \cup B_k$ is a basis of $W_1 + \cdots + W_k$.

Note: If $k = 2$, then W_1 and W_2 are independent if and only if $W_1 \cap W_2 = \{0\}$.

1.34 Examples

a) Three lines in the ordinary space are independent if they do not belong in one plane. A plane and a line are independent if the line does not lie in the plane. Two planes are always dependent.

b) Let $V = C[0, 1]$. Let $W_1 = \{f \in V : \int_0^1 f(x)\, dx = 0\}$ and $W_2 = \{f \in V : f \equiv \text{const}\}$. Then $V = W_1 \oplus W_2$.

2 Linear transformation

2.1 Definition (Linear transformation)

Let V and W be vector spaces over a field F. A *linear transformation* from V to W is a mapping $T : V \to W$ such that

(i) $T(u+v) = Tu + Tv$ for all $u, v \in V$;

(ii) $T(cu) = cTu$ for all $u \in V$, $c \in F$.

Note: if $V = W$ then T is often called a *linear operator* on V.

2.2 Elementary properties

$T(0) = 0$, $T(-u) = -Tu$, $T(\sum c_i u_i) = \sum c_i T u_i$,

$\operatorname{Im} T = \{Tu : u \in V\}$ is a subspace of W (also denoted by $R(T)$),

$\operatorname{Ker} T = \{u \in V : Tu = 0\}$ is a subspace of V (also denoted by $N(T)$),

T is injective (one-to-one) if and only if $\operatorname{Ker} T = \{0\}$.

2.3 Examples

(a) A matrix $A \in F^{m\times n}$ defines a linear transformation $T : F^n \to F^m$ by $Tu = Au$ (multiplication). We denote this transformation by T_A.

Conversely, one has: if $T : F^n \to F^m$ is a linear transformation, then there is a matrix $A \in F^{m\times n}$ such that $T = T_A$.

(b) $T : C^1(0,1) \to C(0,1)$ defined by $Tf = f'$, where f' is the derivative of the function f.

(c) $T : P_n(\mathbb{R}) \to P_{n-1}(\mathbb{R})$ defined by $Tf = f'$, as above.

(d) $T : C[0,1] \to \mathbb{R}$ defined by $Tf = \int_0^1 f(x)\,dx$.

2.4 Theorem

Let V and W be vector spaces, where V is finite dimensional with basis $\{u_1, \ldots, u_n\}$. Let $\{v_1, \ldots, v_n\}$ be a subset of W. Then there exists a unique linear transformation $T : V \to W$ such that $Tu_i = v_i$ for all $1 \le i \le n$.

2.5 Definition (Rank, nullity)

Let $T : V \to W$ be a linear transformation. Then

$$\begin{aligned} \operatorname{rank}(T) : &= \dim \operatorname{Im}(T) \quad \text{if } \operatorname{Im}(T) \text{ is finite dimensional} \\ \operatorname{nullity}(T) : &= \dim \operatorname{Ker}(T) \quad \text{if } \operatorname{Ker}(T) \text{ is finite dimensional} \end{aligned}$$

2.6 Theorem

Let $T : V \to W$ be a linear transformation, V finite dimensional. Then

$$\operatorname{rank}(T) + \operatorname{nullity}(T) = \dim V$$

2.7 Definition (Column rank)

Let $A \in F^{m\times n}$, then the *column rank* of A is the dimension of the column space of A as a subspace of F^m.

2.8 Theorem + Definition (Rank of a matrix)

Let $A \in F^{m\times n}$. Then

$$\text{row rank}(A) = \text{column rank}(A) =: \text{rank}(A)$$

2.9 Notation

Let V and W be vector spaces over F. The set of all linear transformations from V to W is denoted by $L(V, W)$.

$L(V, W)$ is a vector space (over F), with ordinary addition and multiplication by scalars defined for functions (from V to W).

2.10 Theorem

If $\dim V = n$ and $\dim W = m$, then $\dim L(V, W) = mn$, in particular, $L(V, W)$ is finite dimensional.

2.11 Theorem

Let V, W, Z be vector spaces over F. For any $T \in L(V, W)$ and $U \in L(W, Z)$ the composition $U \circ T$, also denoted by UT, is a linear transformation from V to Z, i.e. $UT \in L(V, Z)$.

2.12 Example

Let $A \in F^{m\times n}$, $B \in F^{k\times m}$ and T_A, T_B be defined as in example 2.3(a). Then the composition $T_B T_A$ is a linear transformation from F^n to F^k given by the product matrix BA, i.e. $T_B T_A = T_{BA}$.

2.13 Definition (Isomorphism)

A transformation $T \in L(V, W)$ is called an *isomorphism* if T is bijective. If an isomorphism $T \in L(V, W)$ exists, the vector spaces V and W are called *isomorphic*.

Note: $T \in L(V, W)$ is an isomorphism if and only if there is a $U \in L(W, V)$ such that $UT = I_V$ (the identity on V) and $TU = I_W$ (the identity on W). In other words, T is invertible and $U = T^{-1}$, the inverse of T. Note that T^{-1} is also an isomorphism.

2.14 Theorem

Let V and W be finite dimensional, and $T \in L(V, W)$.

(i) T is injective (one-to-one) if and only if whenever $\{u_1, \ldots, u_k\}$ is linearly independent, then $\{Tu_1, \ldots, Tu_k\}$ is also linearly independent.

(ii) T is surjective (onto, i.e. $\text{Im}(T) = W$) if and only if $\text{rank}(T) = \dim W$

(iii) T is an isomorphism if and only if whenever $\{u_1, \ldots, u_n\}$ is a basis in V, then $\{Tu_1, \ldots, Tu_n\}$ is a basis in W.

(iv) If T is an isomorphism, then $\dim V = \dim W$.

2.15 Theorem

Let V and W be finite dimensional, $\dim V = \dim W$, and $T \in L(V, W)$. Then the following are equivalent:

(i) T is an isomorphism

(ii) T is injective (one-to-one)

(iii) T is surjective (onto)

2.16 Definition (Automorphism, $GL(V)$)

An isomorphisms $T \in L(V, V)$ is called an *automorphism* of V. The set of all automorphisms of V is denoted by $GL(V)$, which stands for *general linear group*. In the special case $V = F^n$ one sometimes writes $GL(V) = GL(n, F)$.

Note: $GL(V)$ is not a subspace of $L(V, V)$ (it does not contain zero, for example), but it is a group with respect to composition. This group is not abelian, i.e. $TU \neq UT$ for some $T, U \in GL(V)$, unless $\dim V = 1$.

2.17 Example

Let $T : P_n(\mathbb{R}) \to P_n(\mathbb{R})$ defined by $Tf(x) = f(x+a)$, where $a \in \mathbb{R}$ is a fixed number. Then T is an isomorphism.

2.18 Theorem

If $\dim V = n$, then V is isomorphic to F^n.

2.19 Definition (Matrix representation)

Let $\dim V = n$ and $\dim W = m$. Let $B = \{u_1, \ldots, u_n\}$ be a basis in V and $C = \{v_1, \ldots, v_m\}$ be a basis in W. Let $T \in L(V, W)$. The unique matrix A defined by

$$Tu_j = \sum_{i=1}^{m} A_{ij} v_i$$

is called the *matrix of T relative to B, C* and denoted by $[T]_{B,C}$.

2.20 Theorem

In the notation of 2.18, for each vector $u \in V$

$$(Tu)_C = [T]_{B,C} (u)_B$$

2.21 Theorem

the mapping $T \mapsto [T]_{B,C}$ defines an isomorphisms of $L(V,W)$ and $F^{m\times n}$.

2.22 Theorem

Let B, C, D be bases in vector spaces V, W, Z, respectively. For any $T \in L(V,W)$ and $U \in L(W,Z)$ we have

$$[ST]_{B,D} = [S]_{C,D}\,[T]_{B,C}$$

2.23 Corollary

$T \in L(V,W)$ is an isomorphism if and only if $[T]_{B,C}$ is an invertible matrix.

2.24 Corollary

Let $B = \{u_1, \ldots, u_n\}$ and $B' = \{u'_1, \ldots, u'_n\}$ be two bases in V. Then

$$[I]_{B',B} = P_{B',B}$$

where $P_{B',B}$ is the transition matrix defined in 1.28.

2.25 Theorem

Let V and W be finite dimensional, and $T \in L(V,W)$. Let B and B' be bases in V, and C and C' be bases in W. Then

$$[T]_{B',C'} = Q\,[T]_{B,C}\,P$$

where

$$P = [I]_{B',B} \qquad \text{and} \qquad Q = [I]_{C,C'}$$

Note: in the special case $V = W$, $B = C$ and $B' = C'$ one has

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

where $P = [I]_{B',B}$.

2.26 Definition (Similar matrices)

Two matrices $A, A' \in F^{n\times n}$ are said to be *similar* if there is an invertible matrix $P \in F^{n\times n}$ such that

$$A' = P^{-1} A\,P$$

The similarity is denoted by $A \sim A'$.

Note: similarity is an equivalence relation.

2.27 Theorem

Two matrices A and A' are similar if and only if they are matrix representations of one linear operator.

2.28 Theorem

Let V and W be finite dimensional and $T \in L(V, W)$. Let B be a basis in V and C a basis in W. Then

$$\operatorname{rank}(T) \;=\; \operatorname{rank}[T]_{B,C}$$

2.29 Corollary

If A and A' are similar, then $\operatorname{rank}(A) = \operatorname{rank}(A')$.

2.30 Definition (Linear functional, Dual space)

Let V be a vector space over F. Then $V^* := L(V, F)$ is called the *dual space* of V. The elements $f \in V^*$ are called *linear functionals* on V (they are linear transformations from V to F).

2.31 Corollary

V^* is a vector space. If V is finite dimensional, then $\dim V^* = \dim V$.

2.32 Examples

(a) see Example 2.3(d): $T : C[0,1] \to \mathbb{R}$ defined by $Tf = \int_0^1 f(x)\,dx$

(b) The trace of a square matrix $A \in F^{n\times n}$ is defined by $\operatorname{tr} A = \sum_i A_{ii}$. Then $\operatorname{tr}\colon F^{n\times n} \to F$ is a linear functional

(c) Every linear functional f on F^n is on the form

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} a_i x_i \qquad \text{for some } a_i \in F$$

2.33 Theorem + Definition

Let V be finite dimensional and $B = \{u_1, \ldots, u_n\}$ be a basis of V. Then there is a unique basis $B^* = \{f_1, \ldots, f_n\}$ of V^* such that

$$f_i(u_j) = \delta_{ij}$$

where $\delta_{ij} = 1$ if $i = j$ and 0 if $i \neq j$ ('Kronecker delta symbol').

The basis B^* is called the *dual basis* of B. It has the property that

$$f = \sum_{i=1}^{n} f(u_i) f_i$$

for all $f \in V^*$.

Note: In Example 2.32(c) the dual basis of the standard basis $B = \{e_1, \ldots, e_n\}$ is given by $f_i(x_1, \ldots, x_n) = x_i$.

3 Determinants

3.1 Definition (Permutation, Transposition, Sign)

A *permutation of a set* S is a bijective mapping $\sigma : S \to S$. If $S = \{1, \ldots, n\}$ we call σ a *permutation of degree* n (or *permutations of* n *letters*).

Notation: $\sigma = (\sigma(1), \ldots, \sigma(n))$.

S_n is the set of all permutations of degree n.

A permutation which interchanges two numbers and leaves all the others fixed is called a *transposition*. Notation: $\tau(j, k)$.

If $\sigma \in S_n$, then the *sign* of σ is defined as follows: $\mathrm{sg}(\sigma) = 1$ if σ can be written as a product of an even number of transpositions, and $\mathrm{sg}(\sigma) = -1$ if σ can be written as a product of an odd number of transpositions.

Note: The sign of permutation is well-defined, see MA634.

3.2 Definition (Determinant)

Let $A \in F^{n\times n}$. Then the *determinant* of A is

$$\det(A) = \sum_{\sigma\in S_n} \mathrm{sg}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

Note: every term in this sum contains exactly one element from each row and exactly one element from each column.

3.3 Examples

(1) If $A \in F^{2\times 2}$, then $\det(A) = a_{11}a_{22} - a_{12}a_{21}$

(2) Let $A = \mathrm{diag}(a_{11}, \ldots, a_{nn})$ be a diagonal matrix (i.e., A is the $n \times n$-matrix with diagonal elements $a_{11}, \ldots, a_{nn}$ and zeros off the diagonal). Then $\det(A) = a_{11} \cdots a_{nn}$.

3.4 Theorem

Let $A, B, C \in F^{n\times n}$.

a) If B is obtained from A by multiplying one row of A by k,then

$$\det(B) = k \det(A)$$

b) If A,B and C are identical, except for the i-th row, where

$$c_{ij} = a_{ij} + b_{ij} \quad 1 \le j \le n$$

then

$$\det(C) = \det(A) + \det(B)$$

(Note that in general $\det(A + B) \neq \det(A) + \det(B)$.)

c) If B is obtained from A by interchange of two rows, then

$$\det(B) = -\det(A)$$

d) If A has two equal rows, then $\det(A) = 0$.

e) If B is obtained from A by adding a multiple of one row to another row, then

$$\det(B) = \det(A)$$

3.5 Algorithm

Let $A \in F^{n\times n}$. By using two elementary row operations – row interchange and adding a multiple of one row to another row – we can transform A into a row echelon matrix R, whose leading entries in non-zero rows are non-zero numbers (not necessarily ones). By the above theorem, $\det(A) = (-1)^p \det(R)$, where p is the number of row interchanges used.

In particular, let $A \in F^{n\times n}$ be invertible. Then R above will be upper triangular, i.e. $R_{ii} \neq 0$ for all $1 \leq i \leq n$, and $R_{ij} = 0$ for all $i > j$. By performing some more row operations one can trandform R into a diagonal matrix $D = \operatorname{diag}(d_1, \ldots, d_n)$ with non-zero diagonal entries $d_i \neq 0$.

3.6 Theorem

$A \in F^{n\times n}$ is invertible if and only if $\det(A) \neq 0$.

3.7 Lemma

Let $A, B \in F^{n\times n}$ and A invertible. Then

$$\det(AB) = \det(A)\ \det(B)$$

3.8 Theorem

Let $A, B \in F^{n\times n}$, then

$$\det(AB) = \det(A)\ \det(B)$$

3.9 Corollary

(i) If $A \in F^{n\times n}$ is invertible, then $\det A^{-1} = 1/\det(A)$.

(ii) If A and B are similar, then $\det(A) = \det(B)$.

3.10 Theorem + Definition (Transpose)

The *transpose* A^t of $A \in F^{n\times n}$ is defined by $A^t_{ij} := A_{ji}$. Then

$$\det(A) = \det(A^t)$$

Note: This implies that all parts of Theorem 3.4 are valid with "row" replaced by "column".

3.11 Theorem

If $A_{11} \in F^{r\times r}$, $A_{21} \in F^{(n-r)\times r}$, and $A_{22} \in F^{(n-r)\times(n-r)}$, then

$$\det \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix} = \det(A_{11})\ \det(A_{22})$$

Note: By induction it follows that

$$\det \begin{pmatrix} A_{11} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ A_{m1} & \cdots & \cdots & A_{mm} \end{pmatrix} = \det(A_{11}) \cdots \det(A_{mm})$$

In particular, for a *lower triangular* matrix one has

$$\det \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{n1} & \cdots & \cdots & a_{nn} \end{pmatrix} = a_{11} \cdots a_{nn}$$

By Theorem 3.10 this also holds for *upper triangular* matrices.

3.12 Definition (Cofactors)
Let $A \in F^{n\times n}$ and $i, j \in \{1, \ldots, n\}$. Denote by $A(i|j) \in F^{(n-1)\times(n-1)}$ the matrix obtained by deleting the i-th row and j-th column from A. Then

$$c_{ij} := (-1)^{i+j} \det(A(i|j))$$

is called the i, j-*cofactor* of A.

3.13 Theorem (Cofactor expansion)
(a) For every $i \in \{1, \ldots, n\}$ one has the i-th *row cofactor expansion*

$$\det(A) = \sum_{j=1}^{n} a_{ij} c_{ij}$$

(b) For every $j \in \{1, \ldots, n\}$ one has the j-th *column cofactor expansion*

$$\det(A) = \sum_{i=1}^{n} a_{ij} c_{ij}$$

3.14 Theorem
If A is invertible, then

$$A^{-1} = \frac{1}{\det(A)} \mathrm{adj}(A)$$

where $\mathrm{adj}(A) \in F^{n\times n}$ is the *adjoint matrix* of A defined by

$$(\mathrm{adj}(A))_{ij} = c_{ji}$$

3.15 Theorem (Cramer's rule)

If $A \in F^{n\times n}$ is invertible and $b \in F^n$, then the unique solution $x \in F^n$ of the equation $Ax = b$ is given by

$$x_j = \frac{\det(B_j)}{\det(A)} \qquad j = 1, \ldots, n$$

where B_j is the matrix obtained by replacing the j-th column of A by the vector b.

Note: Using $\mathrm{adj}(A)$ to compute A^{-1} or Cramer's rule to solve the equation $Ax = b$ is numerically impractical, since the computation of determinants is "too expensive" compared to other methods. However, 3.12 and 3.13 have theoretical value: they show the continuous dependence of entries of A^{-1} on entries of A and, respectively, of x on the entries of A and b.

3.16 Definition

A matrix $A \in F^{n\times n}$ is said to be *upper triangular* if $A_{ij} = 0$ for all $i > j$. If, in addition, $A_{ii} = 1$ for all $1 \leq i \leq n$, the matrix A is said to be *unit upper triangular*. Similarly, lower triangular matrices and unit lower triangular matrices are defined.

3.17 Theorem

The above four classes of matrices are closed under multiplication. For example, if $A, B \in F^{n\times n}$ are unit lower triangular matrices, then so is AB.

3.18 Theorem

The above four classes of matrices are closed under taking inverse. For example, if $A \in F^{n\times n}$ is unit lower triangular matrices, then so is A^{-1}.

4 The LU Decomposition Method

4.1 Algorithm (Gauss Elimination)

Let $A = (a_{ij})$ be an $n \times n$ matrix with $a_{11} \neq 0$. Denote $A^{(1)} = A$ and $a_{ij}^{(1)} = a_{ij}$. We define multipliers

$$m_{i1} = a_{i1}^{(1)}/a_{11}^{(1)} \quad \text{for} \;\; i = 2, \ldots, n$$

and replace the i-th row R_i of the matrix A with $R_i - m_{i1}R_1$ for all $i = 2, \ldots, n$. This creates zeros in the first column of $A^{(1)}$, which then takes the form

$$A^{(2)} = \begin{pmatrix} a_{11}^{(1)} & \cdots & & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2}^{(2)} & \cdots & a_{nn}^{(2)} \end{pmatrix}$$

where

$$a_{ij}^{(2)} = a_{ij}^{(1)} - m_{i1}a_{1j}^{(1)} \quad \text{for} \;\; 2 \leq i, j \leq n$$

Next, assume that $a_{22}^{(2)} \neq 0$. Then one can continue this process and define multipliers

$$m_{i2} = a_{i2}^{(2)}/a_{22}^{(2)} \quad \text{for} \;\; i = 3, \ldots, n$$

and replace the i-th row R_i of the matrix $A^{(2)}$ with $R_i - m_{i2}R_2$ for all $i = 3, \ldots, n$. This creates a matrix, $A^{(3)}$, with zeros in the second column below the main diagonal, and so on. If all the numbers $a_{ii}^{(i)}$, $1 \leq i \leq n-1$, are not zero, then one ultimately obtains an upper triangular matrix

$$A^{(n)} = U = \begin{pmatrix} a_{11} & \cdots & & & a_{1n} \\ 0 & a_{22}^{(2)} & \cdots & & a_{2n}^{(2)} \\ 0 & 0 & a_{33}^{(3)} & \cdots & a_{3n}^{(3)} \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_{nn}^{(n)} \end{pmatrix}$$

The elements $a_{ii}^{(i)}$, $1 \leq i \leq n$, are called *pivots*; the method fails if (and only if) one of the pivots $a_{ii}^{(i)}$, $1 \leq i \leq n-1$, is zero.

4.2 Definition (Principal minors)

Let $A \in F^{n \times n}$. For $1 \leq k \leq n$, the k-th principal minor of A is the matrix $k \times k$ formed by intersecting the first k rows and the first k columns of A (i.e., the top left $k \times k$ block of A). We denote the k-th principal minor of A by A_k.

4.3 Criterion of failure in the Gauss elimination process

The method of Gauss elimination fails if and only if $\det A_k = 0$ for some $k = 1, \ldots, n-1$. This follows because for each $k = 1, \ldots, n$

$$\det A_k = a_{11}^{(1)} \cdots a_{kk}^{(k)}$$

4.4 Gauss matrices

Assume that $A \in F^{n \times n}$ has non-singular principal minors up to the order $n-1$, so that the Gauss elimination works. For each $j = 1, \ldots, n-1$ the Gauss matrix G_j is defined by

$$G_j = \begin{pmatrix} 1 & 0 & \cdots & & & & 0 \\ 0 & 1 & & & & & \\ \vdots & 0 & \ddots & & & & \\ & \vdots & & 1 & & & \\ & & & -m_{j+1,j} & \ddots & & \vdots \\ & & & \vdots & & 1 & 0 \\ 0 & 0 & & -m_{nj} & & 0 & 1 \end{pmatrix}$$

Note that $G_j = I - m^{(j)} e_j^t$ where

$$m^{(j)} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ m_{j+1,j} \\ \vdots \\ m_{nj} \end{pmatrix}$$

4.5 Lemma

For each $j = 1, \ldots, n-1$ we have $G_j A^{(j)} = A^{(j+1)}$.

4.6 Corollary

We have

$$U = A^{(n)} = G_{n-1} \cdots G_2 G_1 A$$

4.7 Lemma

For each $j = 1, \ldots, n-1$ we have

$$L_j := G_j^{-1} = I + m^{(j)} e_j^t$$

so that

$$L_j = \begin{pmatrix} 1 & 0 & \cdots & & & 0 \\ 0 & 1 & & & & \\ \vdots & 0 & \ddots & & & \\ & \vdots & & 1 & & \\ & & & m_{j+1,j} & \ddots & \vdots \\ & & & \vdots & 1 & 0 \\ 0 & 0 & & m_{nj} & 0 & 1 \end{pmatrix}$$

4.8 Lemma

We have

$$L := (I + m^{(1)} e_1^t)(I + m^{(2)} e_2^t) \cdots (I + m^{(n-1)} e_{n-1}^t) = I + \sum_{k=1}^{n-1} m^{(k)} e_k^t$$

so that

$$L = \begin{pmatrix} 1 & 0 & \cdots & & 0 \\ m_{21} & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & \\ & & & 1 & 0 \\ m_{n1} & m_{n2} & \cdots & m_{n,n-1} & 1 \end{pmatrix},$$

is a unit lower triangular matrix.

4.9 Corollary

Assume that $A \in F^{n \times n}$ has non-singular principal minors up to the order $n-1$, so that the Gauss elimination works. Then

$$A = LU$$

i.e. A is the product of a unit lower triangular matrix and an upper triangular matrix. In that case

$$\det A = \det U = a_{11}^{(1)} \cdots a_{nn}^{(n)}$$

4.10 Theorem (LU Decomposition)

Assume that $A \in F^{n \times n}$ has non-singular principal minors up to the order $n-1$, so that the Gauss elimination works. Then there exists a factorization for A of the form $A = LU$, where U is upper triangular and L is unit lower triangular. If in addition, A is non-singular, then such a factorization is unique.

Proof. It remains to prove uniqueness. By way of contradiction, let $A = \tilde{L}\tilde{U} = LU$. As A is non-singular, it follows that $\tilde{U}$ is also non-singular, and hence $L^{-1}\tilde{L} = U\tilde{U}^{-1}$. Now, L^{-1} is unit lower triangular, so that $L^{-1}\tilde{L}$ is also unit lower triangular. On the other hand, $U\tilde{U}^{-1}$ is upper triangular. The only matrix with this property is the identity

matrix I, and thus $\tilde{L} = L$ and $\tilde{U} = U$.

4.11 Algorithm (Forward and backward substitution)

Assume that $A \in F^{n\times n}$ is nonsingular and is decomposed as $A = LU$, where L is lower triangular and U upper triangular. To solve a system $Ax = b$, one writes it as $LUx = b$, calls $Ux = y$, then solves the lower triangular system $Ly = b$ for y using "forward substitution" (finding $y_1, \ldots, y_n$ subsequently), then solves the system $Ux = y$ for x via a "back substitution" (finding $x_n, \ldots, x_1$ subsequently).

4.12 Cost of computation

The cost of computation is usually measured in "flops", where a flop (*fl*oating point *op*eration) is a multiplication or division together with one addition or subtraction. (There is also a different definition of "flop" in the literature, but we use this one.) Let us estimate the cost of the LU decomposition. The cost of arithmetic in computation of $A^{(2)}$ is $n-1$ divisions to compute the multipliers and $n(n-1)$ flops ($n-1$ rows with n flops per row) to make the zeros in the first column, i.e. total of approximately n^2 flops. The computation of $A^{(3)}$ then takes $(n-1)^2$ flops, and so on. Thus the total computational cost for the LU factorization is

$$n^2 + (n-1)^2 + \cdots + 2^2 = \frac{n(n+1)(2n+1)}{6} \approx \frac{n^3}{3}$$

flops.

If one solves a system $Ax = b$, then the LU decomposition is followed by solving two triangular systems. The cost to solve one triangular system is about $n^2/2$ flops. Hence, the total cost is still $\approx n^3/3$.

Note that most of the computational expense is incurred in finding the factors L and U. Thus, this method is particularly well suited to situations in which one is solving systems $Ax = b$ for more that one vector b: each additional b will require $\approx n^2$ flops.

4.13 Algorithm for finding A^{-1}

Assume that $A \in F^{n\times n}$ is non-singular and has non-singular principal minors up to the order $n-1$, so that the Gauss elimination works. One can find the matrix $X = A^{-1}$ by solving the system $AX = I$ for $X \in F^{n\times n}$, where I is the identity matrix. This amounts to solving n systems of linear equations $Ax_k = e_k$, for $k = 1, \ldots, n$, where x_k stands for the k-th column of the matrix X. The computational cost of this procedure is $n^3/3 + n \times n^2 = 4n^3/3$.

Note that, in principle, one can solve the system $Ax = b$ by $x = A^{-1}b$, finding A^{-1} as above. However, this is *very* inefficient, it costs approximately four times more flops than the the LU decomposition followed by forward and backward substitutions.

4.14 Definition (Diagonally dominant matrix)

An $n \times n$ matrix $A = (a_{ij})$ for which we have

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|$$

for all i is said to be *strictly row diagonally dominant*. If

$$|a_{jj}| > \sum_{i \neq j} |a_{ij}|$$

for all j the matrix is said to be *strictly column diagonally dominant*.

4.15 Theorem

If a matrix A is strictly row (or column) diagonally dominant, then $\det A_k \neq 0$ for all $1 \leq k \leq n$. Hence, no zero pivots will be encountered during Gaussian elimination.

Note that if A is strictly column diagonally dominant, then all the multipliers (i.e., the elements of L_j) have absolute value less than one.

The above theorem give some conditions under which the Gauss elimination method works. It is easy to find matrices to which the method does not apply. For example, if $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, then the method fails instantly. (Furthermore, for this A there is no decomposition $A = LU$ with L lower triangular and U upper triangular.) In practice, however, due to the effects of round-off error, one rarely encounters a pivot that is exactly zero. Does this mean that the Gauss elimination is practically perfect? By no means. In practice, one can easily encounter a pivot that is very small, and this is equally disastrous. You can read [**?**, §2.7, p. 123] carefully to appreciate the phenomenon of *swamping*. One tries to avoid this type of problem by using an appropriate pivoting strategy.

4.16 Algorithm of partial pivoting – general

The idea is to avoid small (in absolute value) pivots by interchanging rows, if necessary. At any stage of Gauss elimination, one looks for the largest (in absolute value) element in the pivot column (at or below the main diagonal). For a non-singular matrix it cannot happen that all of these numbers are zero. Then the row containing that element is interchanged with the current row. Now the largest element is on the main diagonal. After that the usual elimination step is performed.

It is important to note that partial pivoting implies that all the multipliers (i.e., the elements of L_j) have absolute value less than or equal to one.

4.17 Theorem

If a matrix A is strictly column diagonally dominant, then the Gaussian elimination with no pivoting is equivalent to Gaussian elimination with partial pivoting (i.e., no row

interchanges are necessary).

4.18 Definition (Permutation matrix)
Let $1 \leq i, j \leq n$. The *permutation matrix* $E(i, j)$ is obtained from the identity matrix I by interchanging the i-th and j-th rows.

4.19 Lemma
Let $A \in F^{n\times n}$. Then the matrix $E(i, j)A$ is obtained from A by interchanging the rows i and j. Similarly, the matrix $AE(i, j)$ is obtained from A by interchanging the columns i and j. Also, $E(i, j)^{-1} = E(i, j)$.

4.20 Algorithm of partial pivoting – formalism
Let us analyze the partial pivoting procedure. At every stage $j \in [1, n-1]$, we find the row, r_j, containing the largest element in the j-th (pivot) column at or below the main diagonal. Then we interchange the rows j and r_j, note that $r_j \geq j$. This will be better described by the permutation matrix $E(j, r_j)$, which we denote for brevity by E_j. Hence, we have

$$A^{(j+1)} = G_j E_j A^{(j)}$$

(Note that if we do not interchange rows then simply $r_j = j$ and $E_j = I$, and we recover our old Gauss elimination procedure.) Therefore,

$$U = A^{(n)} = G_{n-1}E_{n-1} \cdots G_1 E_1 A$$

By construction, U is an upper triangular matrix. Now, taking inverses gives

$$A = E_1 L_1 \cdots E_{n-1} L_{n-1} U$$

The next question is, what can we say about the matrix $E_1 L_1 \cdots E_{n-1} L_{n-1}$?

4.21 Theorem
Let $A \in F^{n\times n}$ be non-singular. Then we can apply the partial pivoting algorithm. Also, let

$$P = E_{n-1} \cdots E_1$$

Then the matrix

$$\tilde{L} := P^{-1} E_1 L_1 \cdots E_{n-1} L_{n-1}$$

is unit lower triangular, and so we have

$$PA = \tilde{L}U$$

Proof. First, observe that for any $m > j$ the matrix $E_m L_j E_m$ is obtained from the (old) Gauss matrix L_j by interchanging the m-th and r_m-th elements in the j-th column below the main diagonal. So, the matrix

$$\tilde{L}_j = E_{n-1} \cdots E_{j+1} L_j E_{j+1} \cdots E_{n-1}$$

is obtained from the (old) Gauss matrix L_j by some permutation of its elements in the j-th (pivot) column below the main diagonal. In particular, $\tilde{L}_j$ is also unit lower triangular.

Now, in the expression

$$E_1 L_1 \cdots E_{n-1} L_{n-1}$$

we fill in as many E's as necessary (remembering that $E_i^2 = I$) to convert all L_j's into $\tilde{L}_j$'s, and obtain

$$PA = \tilde{L}_1 \cdots \tilde{L}_{n-2} L_{n-1} U$$

where

$$P = E_{n-1} \cdots E_1$$

is called a permutation matrix. The matrix

$$\tilde{L} = \tilde{L}_1 \cdots \tilde{L}_{n-2} L_{n-1}$$

is unit lower triangular, and we get

$$PA = \tilde{L}U$$

4.22 Algorithm of partial pivoting – implementation

In practice, one wants to solve a system of equations $Ax = b$ with a nonsingular A. Applying the algorithm of partial pivoting gives $PA = \tilde{L}U$. Therefore, $\tilde{L}Ux = Pb =: c$. Then x can be found by the standard forward and backward substitution, one only needs to obtain c from b. In computer implementation, one does not actually interchange values in memory locations for the elements of A or b, but the row interchanges are recorded in a special permutation vector $S = (s_1, \ldots, s_n)$. It holds initially the values $s_1 = 1, \cdots, s_n = n$. The interchange of rows 2 and 5, for example, is recorded by interchanging the integers 2 and 5 in this vector. Thus if one finds the permutation vector S (which is returned when the LU factorization is completed) then the components of the vector $c = Pb$ can be retrieved from b via the vector S, which is used as a 'pointer'. (see [**?**, p. 64]). The method then proceeds in the usual manner.

4.23 Complete pivoting – a sketch

The method of *complete pivoting* involves both row and column interchanges to make use of the largest pivot available. By a similar argument to that used above this is equivalent to effecting the factorization

$$PAQ = LU,$$

where P and Q are permutation matrices. This method, which provides additional insurance against round-off error buildup is useful if the method of partial pivoting proves to be unstable. The main disadvantage of the method is that it is expensive in that an extra $n^3/3$ comparisons are required, compared to a cost on the order of n^2 comparisons for the partial pivoting algorithm.

5 Diagonalization

Throughout this section, we use the following notation: V is a finite dimensional vector space, $\dim V = n$, and $T \in L(V, V)$. Also, B is a basis in V and $[T]_B := [T]_{B,B} \in F^{n\times n}$ is a matrix representing T.

Our goal in this and the following sections is to find a basis in which the matrix $[T]_B$ is "as simple as possible". In this section we study conditions under which the matrix $[T]_B$ can be made diagonal.

5.1 Definition (Eigenvalue, eigenvector)

A scalar $\lambda \in F$ is called an *eigenvalue* of T if there is a non-zero vector $v \in V$ such that

$$Tv = \lambda v$$

In this case, v is called an *eigenvector* of T corresponding to the eigenvalue λ. (Sometimes these are called *characteristic value, characteristic vector.*)

5.2 Theorem + Definition (Eigenspace)

For every $\lambda \in F$ the set

$$E_\lambda = \{v \in V : Tv = \lambda v\} = \operatorname{Ker}(\lambda I - T)$$

is a subspace of V. If λ is an eigenvalue, then $E_\lambda \neq \{0\}$, and it is called the *eigenspace* corresponding to λ.

Note that E_λ always contains 0, even though 0 is never an eigenvector. At the same time, the zero scalar $0 \in F$ may be an eigenvalue.

5.3 Remark

0 is an eigenvalue $\Leftrightarrow \operatorname{Ker} T \neq \{0\} \Leftrightarrow T$ is not invertible $\Leftrightarrow [T]_B$ is singular $\Leftrightarrow \det[T]_B = 0$.

5.4 Definition (Eigenvalue of a matrix)

$\lambda \in F$ is called an *eigenvalue* of a matrix $A \in F^{n\times n}$ if there is a nonzero $v \in F^n$ such that

$$Av = \lambda v$$

Eigenvectors and *eigenspaces* of matrices are defined accordingly.

5.5 Simple properties

(a) λ is an eigenvalue of $A \in F^{n\times n} \Leftrightarrow \lambda$ is an eigenvalue of $T_A \in L(F^n, F^n)$.

(b) λ is an eigenvalue of $T \Leftrightarrow \lambda$ is an eigenvalue of $[T]_B$ for any basis B.

(c) if $A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, then $\lambda_1, \ldots, \lambda_n$ are eigenvalues of A with eigenvectors $e_1, \ldots, e_n$.

(d) $\dim E_\lambda = \text{nullity}(\lambda I - T) = n - \text{rank}(\lambda I - T)$.

5.6 Lemma

The function $p(x) := \det(xI - [T]_B)$ is a polynomial of degree n. This function is independent of the basis B.

5.7 Definition (Characteristic polynomial)

The function

$$C_T(x) := \det(xI - [T]_B)$$

is called the *characteristic polynomial* of T.

For a matrix $A \in F^{n\times n}$, the function

$$C_A(x) := \det(xI - A)$$

is called the *characteristic polynomial* of A.

Note that $C_T(x) = C_{[T]_B}(x)$ for any basis B.

5.8 Lemma

λ is an eigenvalue of T if and only if $C_T(\lambda) = 0$.

5.9 Corollary

T has at most n eigenvalues. If V is complex ($F = \mathbb{C}$), then T has at least one eigenvalue. In this case, according to the Fundamental Theorem of Algebra, $C_T(x)$ is decomposed into linear factors.

5.10 Corollary

If $A \sim B$ are similar matrices, then $C_A(x) \equiv C_B(x)$, hence A and B have the same set of eigenvalues.

5.11 Examples

(a) $A = \begin{pmatrix} 3 & 2 \\ -1 & 0 \end{pmatrix}$. Then $C_A(x) = x^2 - 3x + 2 = (x-1)(x-2)$, so that $\lambda = 1, 2$. Also, $E_1 = \text{span}(1, -1)$ and $E_2 = \text{span}(2, -1)$, so that $E_1 \oplus E_2 = \mathbb{R}^2$.

(b) $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Then $C_A(x) = (x-1)^2$, so that $\lambda = 1$ is the only eigenvalue, $E_1 = \text{span}(1, 0)$.

(c) $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Then $C_A(x) = x^2 + 1$, so that there are no eigenvalues in the real case and two eigenvalues (i and $-i$) in the complex case.

(d) $A = \begin{pmatrix} 3 & 3 & 2 \\ 1 & 2 & 2 \\ -1 & -1 & 0 \end{pmatrix}$. Then $C_A(x) = (x-1)(x-2)^2$, so that $\lambda = 1, 2$. $E_1 = \text{span}(-1, 1, 0)$ and $E_2 = \text{span}(2, 0, -1)$. Here $E_1 \oplus E_2 \neq \mathbb{R}^3$ ("not enough eigenvectors").

(e) $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. Then $C_A(x) = (x-1)(x-2)^2$, so that $\lambda = 1, 2$. $E_1 = \text{span}\{e_1\}$ and $E_2 = \text{span}\{e_2, e_3\}$. Now $E_1 \oplus E_2 = \mathbb{R}^3$.

5.12 Definition (Diagonalizability)

T is said to be *diagonalizable* if there is a basis B such that $[T]_B$ is a diagonal matrix.

A matrix $A \in F^{n\times n}$ is said to be *diagonalizable* if there is a similar matrix $D \sim A$ which is diagonal.

5.13 Lemma

(i) If $v_1, \ldots, v_k$ are eigenvalues of T corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, then the set $\{v_1, \ldots, v_n\}$ is linearly independent.

(ii) If $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of T, then $E_{\lambda_1} + \cdots + E_{\lambda_n} = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_n}$.

Proof of (i) goes by induction on k.

5.14 Corollary

If T has n distinct eigenvalues, then T is diagonalizable. (The converse is not true.)

5.15 Theorem

T is diagonalizable if and only if there is a basis B consisting entirely of eigenvectors of T. (In this case we say that T has a complete set of eigenvectors.)

Not all matrices are diagonalizable, even in the complex case, see Example 5.11(b).

5.16 Definition (Invariant subspace)

A subspace W is said to be *invariant* under T if $TW \subset W$, i.e. $Tw \in W$ for all $w \in W$.

The *restriction* of T to a T-invariant subspace W is denoted by $T|_W$. It is a linear transformation of W into itself.

5.17 Examples

(a) Any eigenspace E_λ is T-invariant. Note that any basis in E_λ consists of eigenvectors.

(b) Let $A = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Then the subspaces $W_1 = \text{span}\{e_1, e_2\}$ and $W_2 =$

span$\{e_3\}$ are T_A-invariant. The restriction $T_A|_{W_1}$ is represented by the matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, and the restriction $T_A|_{W_2}$ is the identity.

5.18 Lemma

Let $V = W_1 \oplus W_2$, where W_1 and W_2 are T-invariant subspaces. Let B_1 and B_2 be bases in W_1, W_2, respectively. Denote $[T|_{W_1}]_{B_1} = A_1$ and $[T|_{W_2}]_{B_2} = A_2$. Then

$$[T]_{B_1 \cup B_2} = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$$

Matrices like this are said to be *block-diagonal*. Note: by induction, this generalizes to $V = W_1 \oplus \cdots \oplus W_k$.

5.19 Definition (Algebraic multiplicity, geometric multiplicity)

Let λ be an eigenvalue of T. The *algebraic multiplicity* of λ is its multiplicity as a root of $C_T(x)$, i.e. the highest power of $x - \lambda$ that divides $C_T(x)$. The *geometric multiplicity* of λ is the dimension of the eigenspace E_λ.

The same definition goes for eigenvalues of matrices.

Both algebraic and geometric multiplicities are at least one.

5.20 Theorem

T is diagonalizable if and only if the sum of geometric multiplicities of its eigenvalues equals n. In this case, if $\lambda_1, \ldots, \lambda_s$ are all distinct eigenvalues, then

$$V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_s}$$

Furthermore, if $B_1, \ldots, B_s$ are arbitrary bases in $E_{\lambda_1}, \ldots, E_{\lambda_s}$ and $B = B_1 \cup \cdots \cup B_s$, then

$$[T]_B = \operatorname{diag}(\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots, \lambda_s, \ldots, \lambda_s)$$

where each eigenvalue λ_i appears $m_i = \dim E_{\lambda_i}$ times.

5.21 Corollary

Assume that

$$C_T(x) = (x - \lambda_1) \cdots (x - \lambda_n)$$

where λ_i's are not necessarily distinct.

(i) If all the eigenvalues have the same algebraic and geometric multiplicities, then T is diagonalizable.

(ii) If all the eigenvalues are distinct, then T is diagonalizable.

5.22 Corollary

Let T be diagonalizable, and D_1, D_2 are two diagonal matrices representing T (in different bases). Then D_1 and D_2 have the same the diagonal elements, up to a permutation.

5.23 Examples (continued from 5.11)

(a) The matrix is diagonalizable, its diagonal form is $D = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$.

(b) The matrix is not diagonalizable.

(c) The matrix is not diagonalizable in the real case, but is diagonalizable in the complex case. Its diagonal form is $D = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$.

(d) The matrix is not diagonalizable.

6 Generalized eigenvectors, Jordan decomposition

Throughout this and the next sections, we use the following notation: V is a finite dimensional *complex* vector space, $\dim V = n$, and $T \in L(V, V)$.

In this and the next sections we focuse on nondiagonalizable transformations. Those, as we have seen by examples, do not have 'enough eigenvectors'.

6.1 Definition (Generalized eigenvector, Generalized eigenspace)

Let λ be an eigenvalue of T. A vector $v \neq 0$ is called a *generalized eigenvector* of T corresponding to λ if

$$(T - \lambda I)^k v = 0$$

for some positive integer k.

The *generalized eigenspace* corresponding to λ is the set of all generalized eigenvectors corresponding to λ plus the zero vector. We denote that space by U_λ.

6.2 Example

Let $A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$ for some $\lambda \in \mathbb{C}$. Then λ is the (only) eigenvalue of A, and $E_\lambda = \operatorname{span}\{e_1\}$. Since $(A - \lambda I)^2$ is the zero matrix, $U_\lambda = \mathbb{C}^2$.

6.3 Notation

For each $k \geq 1$, denote $U_\lambda^{(k)} = \operatorname{Ker}(T - \lambda I)^k$. Clearly, $U_\lambda^{(1)} = E_\lambda$ and $U_\lambda^{(k)} \subset U_\lambda^{(k+1)}$ (i.e., $\{U_\lambda^{(k)}\}$ is an *increasing* sequence of subspaces of V). Note that if $U_\lambda^{(k)} \neq U_\lambda^{(k+1)}$, then $\dim U_\lambda^{(k)} < \dim U_\lambda^{(k+1)}$.

Observe that $U_\lambda = \cup_{k=1}^{\infty} U_\lambda^{(k)}$. Since each $U_\lambda^{(k)}$ is a subspace of V, their union U_λ is a subspace, too.

6.4 Lemma

There is an $m = m_\lambda$ such that $U_\lambda^{(k)} \neq U_\lambda^{(k+1)}$ for all $1 \leq k \leq m-1$, and

$$U_\lambda^{(m)} = U_\lambda^{(m+1)} = U_\lambda^{(m+2)} = \cdots$$

In other words, the sequence of subspaces $U_\lambda^{(k)}$ strictly increases up to $k = m$ and stabilizes after $k \geq m$. In particular, $U_\lambda = U_\lambda^{(m)}$.

Proof. By way of contradiction, assume that

$$U_\lambda^{(k)} = U_\lambda^{(k+1)} \neq U_\lambda^{(k+2)}$$

Pick a vector $v \in U_\lambda^{(k+2)} \setminus U_\lambda^{(k+1)}$ and put $u := (T - \lambda I)v$. Then $(T - \lambda I)^{k+2} v = 0$, so $(T - \lambda I)^{k+1} u = 0$, hence $u \in U_\lambda^{k+1}$. But then $u \in U_\lambda^k$, and so $(T - \lambda I)^k u = 0$, hence $(T - \lambda I)^{k+1} v = 0$, a contradiction. $\square$

6.5 Corollary

$m_\lambda \le n$, where $n = \dim V$. In particular, $U_\lambda = U_\lambda^{(n)}$.

6.6 Definition (Polynomials in T)

By T^k we denote $T \circ \cdots \circ T$ (k times), i.e. inductively $T^k v = T(T^{k-1}v)$. A polynomial in T is

$$a_k T^k + \cdots + a_1 T + a_0 I$$

where $a_0, \ldots, a_k \in \mathbb{C}$. Example: $(T - \lambda I)^k$ is a polynomial in T for any $k \ge 1$.

Note: $T^k \circ T^m = T^m \circ T^k$. For arbitrary polynomials p, q we have $p(T)q(T) = q(T)p(T)$.

6.7 Lemma.

The generalized eigenvectors of T span V.

Proof goes by induction on $n = \dim V$. For $n = 1$ the lemma follows from 5.9. Assume that the lemma holds for all vector spaces of dimension $< n$. Let λ be an eigenvalue of T. We claim that $V = V_1 \oplus V_2$, where $V_1 := \operatorname{Ker}(T - \lambda I)^n = U_\lambda$ and $V_2 := \operatorname{Im}(T - \lambda I)^n$.

Proof of the claim:

(i) Show that $V_1 \cap V_2 = 0$. Let $v \in V_1 \cap V_2$. Then $(T - \lambda I)^n v = 0$ and $(T - \lambda I)^n u = v$ for some $u \in V$. Hence $(T - \lambda I)^{2n} u = (T - \lambda I)^n v = 0$, i.e. $u \in U_\lambda$. Now 6.5 implies that $0 = (T - \lambda I)^n u = v$.

(ii) Show that $V = V_1 + V_2$. In view of 2.6 we have $\dim V_1 + \dim V_2 = \dim V$. Since V_1 and V_2 are independent by (i), we have $V_1 + V_2 = V$.

The claim is proven.

Since, $V_1 = U_\lambda$, it is already spanned by generalized eigenvectors. Next, V_2 is T-invariant, because for any $v \in V_2$ we have $v = (T - \lambda I)^n u$ for some $u \in V$, so $Tv = T(T - \lambda I)^n U = (T - \lambda I)^n T u \in V_2$. Since $\dim V_2 < n$ (remember that $\dim V_1 \ge 1$), by the inductive assumption V_2 is spanned by generalized eigenvectors of $T|_{V_2}$, which are, of course, generalized eigenvectors for T. This proves the lemma.□

6.8 Lemma

Generalized eigenvectors corresponding to distinct eigenvalues of T are linearly independent.

Proof. Let $v_1, \ldots, v_m$ be generalized eigenvectors corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_m$. We need to show that if $v_1 + \cdots + v_m = 0$, then $v_1 = \cdots = v_m = 0$. It is enough to show that $v_1 = 0$.

Let k be the smallest positive integer such that $(T - \lambda_1 I)^k v_1 = 0$. We now apply the transformation

$$R := (T - \lambda_1 I)^{k-1} (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n$$

to the vector $v_1 + \cdots + v_m$ (if $k = 1$, then the first factor in R is missing). Since all the factors in R commute (as polynomials in T), R kills all the vectors $v_2, \ldots, v_m$, and we

get $Rv_1 = 0$. Next, we replace $T - \lambda_i I$ by $(T - \lambda_1 I) + (\lambda_1 - \lambda_i)I$ for all $i = 2, \ldots, m$ in the product formula for R. We then expand this formula by Binominal Theorem. All the terms but one will contain $(T - \lambda_1 I)^r$ with some $r \geq k$, which kills v_1, due to our choice of k. The equation $Rv_1 = 0$ is then equivalent to

$$(\lambda_1 - \lambda_2)^n \cdots (\lambda_1 - \lambda_m)^n (T - \lambda_1 I)^{k-1} v_1 = 0$$

This contradicts our choice of k (remember that the eigenvalues are distinct, so that $\lambda_i \neq \lambda_1$). □

6.9 Definition (Nilpotent transformations)

The transformation $T : V \to V$ is said to be *nilpotent*, if $T^k = 0$ for some positive integer k. The same definition goes for matrices.

6.10 Example

If $A \in \mathbb{C}^{n \times n}$ is an upper triangular matrix whose diagonal entries are zero, then it is nilpotent.

6.11 Lemma

T is nilpotent if and only if 0 is its only eigenvalue.

Proof. Let T be nilpotent and λ be its eigenvalue with eigenvector $v \neq 0$. Then $T^k v = \lambda^k v$ for all k, and then $T^k v = 0$ implies $\lambda^k = 0$, so $\lambda = 0$.

Conversely, if 0 is the only eigenvalue, then by 6.7, $V = U_0 = \operatorname{Ker} T^n$, so $T^n = 0$. □

6.12 Theorem (Structure Theorem)

Let $\lambda_1, \ldots, \lambda_m$ be all distinct eigenvalues of $T : V \to V$ with corresponding generalized eigenspaces $U_1, \ldots, U_m$. Then
(i) $V = U_1 \oplus \cdots \oplus U_m$
(ii) Each U_j is T-invariant
(iii) $(T - \lambda_j I)|_{U_j}$ is nilpotent for each j
(iv) Each $T|_{U_j}$ has exactly one eigenvalue, λ_j
(v) $\dim U_j$ equals the algebraic multiplicity of λ_j

Proof. (i) Follows from 6.7 and 6.8.
(ii) Recall that $U_j = \operatorname{Ker}(T - \lambda_j I)^n$ by 6.5. Hence, for any $v \in U_j$ we have

$$(T - \lambda_j I)^n Tv \overset{\text{by } 6.6}{=} T(T - \lambda_j I)^n v = 0$$

so that $Tv \in U_j$.
(iii) Follows from 6.5.
(iv) From 6.11 and (iii), $(T - \lambda_j I)|_{U_j}$ has exactly one eigenvalue, 0. Therefore, T_{U_j} has exactly one eigenvalue, λ_j. (A general fact: λ is an eigenvalue of T if and only if $\lambda - \mu$

is an eigenvalue of $T - \mu I$.)
(v) Pick a basis B_j in each U_j, then $B := \cup_{j=1}^{m} B_j$ is a basis of V by (i). Due to 5.18 and (i), the matrix $[T]_B$ is block-diagonal, whose diagonal blocks are $[T|_{U_j}]_{B_j}$, $1 \le j \le m$. Then

$$\begin{aligned} C_T(x) &= \det(xI - [T]_B) \\ &= C_{[T|_{U_1}]_{B_1}}(x) \cdots C_{[T|_{U_m}]_{B_m}}(x) \\ &= C_{T|_{U_1}}(x) \cdots C_{T|_{U_m}}(x) \end{aligned}$$

due to 3.11. Since $T|_{U_j}$ has the only eigenvalue λ_j, we have $C_{T|_{U_j}}(x) = (x - \lambda_j)^{\dim U_j}$, hence

$$C_T(x) = (x - \lambda_1)^{\dim U_1} \cdots (x - \lambda_m)^{\dim U_m}$$

Theorem is proven. □

6.13 Corollary
Since $E_\lambda \subset U_\lambda$, the geometric multiplicity of λ never exceeds its algebraic multiplicity.

6.14 Definition (Jordan block matrix)
An $m \times m$ matrix J is called a *Jordan block matrix* for the eigenvalue λ if

$$J = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & 0 & \ddots & 0 \\ 0 & 0 & \lambda & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 & \lambda & 1 \\ 0 & \cdots & \cdots & 0 & 0 & \lambda \end{pmatrix}$$

6.15 Properties of Jordan block matrices
(i) $C_J(x) = (x - \lambda)^m$, hence λ is the only eigenvalue of J, its algebraic multiplicity is m
(ii) $(J - \lambda I)^m = 0$, i.e. the matrix $J - \lambda I$ nilpotent
(iii) $\operatorname{rank}(J - \lambda I) = m - 1$, hence $\operatorname{nullity}(J - \lambda I) = 1$, so the geometric multiplicity of λ is 1.
(iv) $Je_1 = \lambda e_1$, hence $(J - \lambda I)e_1 = 0$, i.e. $E_\lambda = \operatorname{span}\{e_1\}$
(v) $Je_k = \lambda e_k + e_{k-1}$, hence $(J - \lambda I)e_k = e_{k-1}$ for $2 \le k \le m$
(vi) $(J - \lambda I)^k e_k = 0$ for all $1 \le k \le m$.
Note that the map $J - \lambda I$ takes

$$e_m \mapsto e_{m-1} \mapsto \cdots \mapsto e_1 \mapsto 0$$

6.16 Definition (Jordan chain)
Let λ be an eigenvalue of T. A *Jordan chain* is a sequence of non-zero vectors $\{v_1, \ldots, v_m\}$ such that $(T - \lambda I)v_1 = 0$ and $(T - \lambda I)v_k = v_{k-1}$ for $k = 2, \ldots, m$. The

length of the Jordan chain is m.

Note that a Jordan chain contains exactly one eigenvector (v_1), and all the other vectors in the chain are generalized eigenvectors.

6.17 Lemma

Let λ be an eigenvalue of T. Suppose we have $s \geq 1$ Jordan chains corresponding to λ, call them $\{v_{11}, \ldots, v_{1m_1}\}$, …, $\{v_{s1}, \ldots, v_{sm_s}\}$. Assume that the vectors $\{v_{11}, v_{21}, \ldots, v_{s1}\}$ (the eigenvectors in these chains) are linearly independent. Then all the vectors $\{v_{ij} : i = 1, \ldots, s,\ j = 1, \ldots, m_j\}$ are linearly independent.

Proof. Let $M = \max\{m_1, \ldots, m_s\}$ be the maximum length of the chains. The proof goes by induction on M. For $M = 1$ the claim is trivial. Assume the lemma is proved for chains of lengths $\leq M-1$. Without loss of generality, assume that $m_1 \geq m_2 \geq \cdots \geq m_s$, i.e. the lengths are decreasing, so $M = m_1$.

By way of contradiction, let

$$\sum_{i,j} c_{ij} v_{ij} = 0$$

Applying $(T - \lambda I)^{m_1 - 1}$ to the vector $\sum c_{ij} v_{ij}$ kills all the terms except the last vectors v_{im_1} in the chains of maximum length (of length m_1). Those vectors will be transformed to $(T - \lambda I)^{m_1 - 1} v_{im_1} = v_{i1}$, so we get

$$\sum_{i=1}^{p} c_{im_1} v_{i1} = 0$$

where $p \leq s$ is the number of chains of length m_1. Since the vectors $\{v_{i1}\}$ are linearly independent, we conclude that $c_{im_1} = 0$ for all i. That reduces the problem to the case of chains of lengths $\leq M - 1$. □

6.18 Corollary

Let $B = \{v_1, \ldots, v_m\}$ be a Jordan chain corresponding to an eigenvalue λ. Then B is linearly independent, i.e. it is a basis in the subspace $W := \operatorname{span}\{v_1, \ldots, v_m\}$. Note that W is T-invariant, and the matrix $[T|_W]_B$ is exactly a Jordan block matrix.

6.19 Definition (Jordan basis)

A basis B of V is called a *Jordan basis* for T if it is a union of some Jordan chains.

6.20 Remark

If B is a Jordan basis of V, then $[T]_B$ is a block diagonal matrix, whose diagonal blocks are Jordan block matrices.

6.21 Definition (Jordan matrix)

A matrix Q is called a *Jordan matrix corresponding to an eigenvalue* λ if

$$Q = \begin{pmatrix} J_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & J_s \end{pmatrix}$$

where $J_1, \ldots, J_s$ are Jordan block matrices corresponding to λ, and their lengths decrease: $|J_1| \geq \cdots \geq |J_s|$.

A matrix A is called a *Jordan matrix* if

$$A = \begin{pmatrix} Q_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Q_r \end{pmatrix}$$

where $Q_1, \ldots, Q_r$ are Jordan matrices corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_r$.

6.22 Example

Let $T : \mathbb{C}^4 \to \mathbb{C}^4$ have only one eigenvalue, λ. We can find all possible Jordan matrices for T; there are 5 distinct such matrices.

6.23 Theorem (Jordan decomposition)

Let V be a finite dimensional complex vector space.
(i) For any $T : V \to V$ there is a Jordan basis B of V, so that the matrix $[T]_B$ is a Jordan matrix. The latter is unique, up to a permutation of eigenvalues.
(ii) Every matrix $A \in \mathbb{C}^{n \times n}$ is similar to a Jordan matrix. The latter is unique, up to a permutation of eigenvalues (i.e., Q_j's in 6.21).

Note: the uniqueness of the matrices Q_j is achieved by the requirement $|J_1| \geq \cdots \geq |J_s|$ in 6.21.

Note: the Jordan basis B is not unique, not even after fixing the order of eigenvalues.

Proof of Theorem 6.23.
(ii) follows from (i).
It is enough to prove (i) assuming that T has just one eigenvalue λ, and then use 6.12. We can even assume that the eigenvalue of T is zero, by switching from T to $T - \lambda I$. Hence, we assume that T is nilpotent.

The proof of existence goes by induction on $n = \dim V$. The case $n = 1$ is trivial. Assume the theorem is proved for all spaces of dimension $< n$. Consider $W := \operatorname{Im} V$. If $\dim W = 0$, then $T = 0$, so the matrix $[T]_B$ is diagonal (actually, zero) in any basis. So, assume that $m := \dim W \geq 1$.

Note that W is T-invariant, and $m < n$, because $n - m = \operatorname{nullity} T \neq 0$. So, by the inductive assumption there is a basis B in W that is the union of Jordan chains. Let k be the number of those chains.

The last vector in each Jordan chain has a pre-image under T (because it belongs in $W = \operatorname{Im} T$). So, we can extend each of those k Jordan chains by one more vector. Now we get k Jordan chains for the transformation $T : V \to V$. By 6.17 all the vectors in those chains are linearly independent, so they span a subspace of dimension $m + k$.

Next, consider the space $K := \operatorname{Ker} T$. Note that $K_0 := \operatorname{Ker} T|_W$ is a subspace of K. The first vectors in our Jordan chains make a basis in K_0, so $k = \dim K_0 = \operatorname{nullity} T|_W$. This basis can be extended to a basis in K, and thus we can get $r := \dim K - \dim K_0$ new vectors. Note that $\dim K - \dim K_0 = n - m - k$. Hence, the total number of vectors we found is n. The last r vectors are eigenvectors, so they make r Jordan chains of length 1.

Finally, note that all our vectors are independent, therefore they make a basis in V.

To prove the uniqueness, note that for any $k \geq 1$ the number of Jordan chains in the Jordan basis that have length $\geq k$ equals $\operatorname{rank} T^{k-1} - \operatorname{rank} T^k$, so it is independent of the basis. This easily proves the uniqueness. □

6.24 Strategy to find the Jordan matrix

Given a matrix $A \in \mathbb{C}^{n \times n}$ you can find a similar Jordan matrix $J \sim A$ as follows. First, find all the eigenvalues. Then, for every eigenvalue λ and $k \geq 1$ find the number $r_k = \operatorname{rank} (T - \lambda I)^k$. You will get a sequence $n > r_1 > r_2 > \cdots > r_p = r_{p+1} = \cdots$ (in view of 6.4). Then $r_{k-1} - r_k$ is the number of Jordan blocks of length $\geq k$ corresponding to λ.

Example: $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$. Here $\lambda = 1$ is the only eigenvalue, and $r_1 = \operatorname{rank} (A - I) = 1$. Then $(A - I)^2$ is a zero matrix, so $r_2 = 0$. Therefore, $J = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

In addition, we can find a Jordan basis in this example, i.e. a basis in which the transformation is represented by a Jordan matrix. The strategy: pick a vector $v_1 \in \operatorname{Im} (A - I)$, then take one of its preimages $v_2 \in (A - I)^{-1} v_1$, and an arbitrary eigenvector $v_3 \in \operatorname{Ker} (A - I)$ independent of v_1. Then $\{v_1, v_2\}$ and $\{v_3\}$ are two Jordan chains making a basis.

7 Minimal Polynomial and Cayley-Hamilton Theorem

We continue using the notation V, n, T of Section 6. The statements 7.1–7.4 below hold for any field F, but in the rest of the section 7 we work in the complex field.

7.1 Lemma

For any $T : V \to V$ there exists a nonzero polynomial $f \in P(F)$ such that $f(T) = 0$. If we require that the leading coefficient equal 1 and the degree of the polynomial be minimal, then such a polynomial is unique.

Proof: Recall that $L(V, V)$ is a vector space of a finite dimension, $\dim L(V, V) = n^2$. The transformations $I, T, T^2, \ldots$ belong in $L(V, V)$, so there is a $k \geq 1$ such that $I, T, \ldots, T^k$ are linearly dependent, i.e. $\exists\, c_0, \ldots, c_k \in F$:

$$c_0 I + c_1 T + \cdots + c_k T^k = 0$$

and we can assume that $c_k \neq 0$. Dividing through by c_k gives $c_k = 1$. Lastly, if there are two distinct polynomials of degree k and with $c_k = 1$ satisfying the above equation, then their difference is a polynomial g of degree $< k$ such that $g(T) = 0$. This proves the uniqueness. □

7.2 Definition (Minimal polynomial)

The unique polynomial found in Lemma 7.1 is called the *minimal polynomial* $M_T(x)$ of $T : V \to V$. If A is an $n \times n$-matrix then the minimal polynomial of T_A is denoted by $M_A(x)$.

7.3 Examples

(a) $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Here $A^2 = 0$, so that the minimal polynomial is $M_A(x) = x^2$. (It is easy to check that $cI + A \neq 0$ for any c, so the degree one is not enough.) Compare this to the characteristic polynomial $C_A(x) = x^2$.

(b) It is an easy exercise to prove $C_A(A) = 0$ for any 2×2 matrix A (over any field F). Hence $M_A(x)$ either coincides with $C_A(x)$ or is of degree one. In the latter case $A - cI = 0$ for some $c \in F$, hence $A = cI$, in which case $C_A(x) = (x - c)^2$ and $M_A(x) = x - c$.

(c) $A = \begin{pmatrix} 2 & 3 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. Here $A - 2I$ is an upper triangular matrix with zero main diagonal, so it is nilpotent, see 6.10. It is easy to check that $(A - 2I)^2 = 0$, hence the minimal polynomial is $M_A(x) = (x-2)^2 = x^2 - 4x + 4$. Compare this to the characteristic polynomial $C_A(x) = (x - 2)^3$. Note that $C_A(x)$ is a multiple of $M_A(x)$.

7.4 Corollary
(i) Let B be a basis of V, then $M_T(x) = M_{[T]_B}(x)$.
(ii) Similar $n\times n$-matrices have the same minimal polynomial.

From now on, $F = \mathbb{C}$ again.

7.5 Theorem
Let $\lambda_1, \ldots, \lambda_r$ be all distinct eigenvalues of T and U_j the corresponding generalized eigenspaces. Let m_j is the maximum size of Jordan blocks corresponding to the eigenvalue λ_j. Consider the polynomial

$$p(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_r)^{m_r}$$

Then:
(i) $\deg p \leq \dim V$
(ii) if $f(x)$ is a nonzero polynomial such that $f(T) = 0$, then f is a multiple of p
(iii) $p(x) = M_T(x)$

Note: m_j is the length of the longest Jordan chain in U_j. Also, m_j is the smallest positive integer s.t. $(T - \lambda_j I)^{m_j} v = 0$ for all $v \in U_j$.

Proof:
(i) Follows from 6.23.
(ii) Let $f(T) = 0$. We will show that $f(x)$ is a multiple of $(x - \alpha_j)_j^m$ for all j. Fix a j and write $f(x) = c(x - c_1)^{t_1} \cdots (x - c_s)^{t_s}(x - \lambda_j)^t$ where $c_1, \ldots c_s$ denote the roots of f other than λ_j (if λ_j is not a root of f, we simply put $t = 0$). If $t < m_j$, then there is a vector $v \in U_j$ such that $u := (T - \lambda_j I)^t v \neq 0$. Recall that U_j is T-invariant, so it is $(T - cI)$-invariant for any c. Hence, $u \in U_j$. Furthermore, each transformation $T - c_i I$ leaves U_j invariant, and is a bijection of U_j because $c_i \neq \lambda_j$. Therefore, the transformation $c(T - c_1 I)^{t_1} \cdots (T - c_s I)^{t_s}$ is a bijection of U_j, so it takes u to a nonzero vector w. Thus, $w = f(T)v \neq 0$, hence $f(T) \neq 0$, a contradiction. This proves (ii).
(iii) By (ii), $M_T(x)$ is a multiple of $p(x)$. It remains to prove that $p(T) = 0$, then use 7.1. To prove that $p(T) = 0$, recall that $V = U_1 \oplus \cdots \oplus U_r$, and for every $v \in U_j$ we have $(T - \lambda_j I)^{m_j} v = 0$. $\square$

7.6. Example
Let J be an $m\times m$ Jordan block matrix for eigenvalue λ, see 6.14. Then $U_\lambda = \mathbb{C}^m$ and $(T - \lambda I)^m = 0$ (and m is the minimal such power). So, $M_J(x) = (x - \lambda)^m$. Note that $C_J(x) = M_J(x)$. In general, though, $C_A(x) \neq M_A(x)$.

7.7 Theorem (Cayley-Hamilton)
The characteristic polynomial $C_T(x)$ is a multiple of the minimal polynomial $M_T(x)$. In particular, $C_T(T) = 0$, i.e. any linear operator satisfies its own characteristic equation.

Proof: Let $\lambda_1, \ldots, \lambda_r$ be all distinct eigenvalues of T, and $p_1, \ldots, p_r$ their algebraic multiplicities. Then $C_T(x) = (x-\lambda_1)^{p_1} \cdots (x-\lambda_r)^{p_r}$. Note that $p_j = \dim U_j \geq m_j$, where m_j is the maximum size of Jordan blocks corresponding to λ_j. So, by 7.5 the minimal polynomial $M_T(x) = (x-\lambda_1)^{m_1} \cdots (x-\lambda_r)^{m_r}$ divides $C_T(x)$. $\square$

7.8 Corollary
(i) $C_T(x)$ and $M_T(x)$ have the same linear factors.
(ii) T is diagonalizable if and only if $M_T(x) = (x-\lambda_1)\cdots(x-\lambda_r)$, i.e. $M_T(x)$ has no multiple roots.

7.9 Examples
(a) $A = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 1 \\ -1 & -1 & -2 \end{pmatrix}$. Here $C_A(x) = (x+1)^2(x-1)$. Therefore, $M_T(x)$ may be either $(x+1)^2(x-1)$ or $(x+1)(x-1)$. To find it, it is enough to check if $(A+I)(A-I) = 0$, i.e. if $A^2 = I$. This is not true, so $M_T(x) = (x+1)^2(x-1)$. The Jordan form of the matrix is $J = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

(b) Assume that $C_A(x) = (x-2)^4(x-3)^2$ and $M_A(x) = (x-2)^2(x-3)$. Find all possible Jordan forms of A. Answer: there are two Jordan blocks of length one for $\lambda = 3$ and two or three Jordan blocks of lengths 2+2 or 2+1+1 for $\lambda = 2$.

8 Norms and Inner products

This section is devoted to vector spaces with an additional structure - inner product. The underlining field here is either $F = \mathbb{R}$ or $F = \mathbb{C}$.

8.1 Definition (Norm, distance)

A *norm* on a vector space V is a real valued function $||\cdot||$ satisfying

1. $||v|| \geq 0$ for all $v \in V$ and $||v|| = 0$ if and only if $v = 0$.
2. $||cv|| = |c|\,||v||$ for all $c \in F$ and $v \in V$.
3. $||u + v|| \leq ||u|| + ||v||$ for all $u, v \in V$ (*triangle inequality*).

A vector space $V = (V, ||\cdot||)$ together with a norm is called a *normed space*.

In normed spaces, we define the *distance* between two vectors u, v by $d(u, v) = ||u - v||$.

Note that property 3 has a useful implication: $||u|| - ||v|| \leq ||u - v||$ for all $u, v \in V$.

8.2 Examples

(i) Several standard norms in $\mathbb{R}^n$ and $\mathbb{C}^n$:

$$||x||_1 := \sum_{i=1}^{n} |x_i| \qquad (1 - \text{norm})$$

$$||x||_2 := \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2} \qquad (2 - \text{norm})$$

note that this is Euclidean norm,

$$||x||_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \qquad (p - \text{norm})$$

for any real $p \in [1, \infty)$,

$$||x||_\infty := \max_{1 \leq i \leq n} |x_i| \qquad (\infty - \text{norm})$$

(ii) Several standard norms in $C[a, b]$ for any $a < b$:

$$||f||_1 := \int_a^b |f(x)|\, dx$$

$$||f||_2 := \left(\int_a^b |f(x)|^2\, dx \right)^{1/2}$$

$$||f||_\infty := \max_{a \leq x \leq b} |f(x)|$$

8.3 Definition (Unit sphere, Unit ball)
Let $||\cdot||$ be a norm on V. The set

$$S_1 = \{v \in V : ||v|| = 1\}$$

is called a *unit sphere* in V, and the set

$$B_1 = \{v \in V : ||v|| \le 1\}$$

a *unit ball* (with respect to the norm $||\cdot||$). The vectors $v \in S_1$ are called *unit vectors*.

For any vector $v \neq 0$ the vector $u = v/||v||$ belongs in the unit sphere S_1, i.e. any nonzero vector is a multiple of a unit vector.

The following four theorems, 8.4–8.7, are given for the sake of completeness, they will not be used in the rest of the course. Their proofs involve some advanced material of real analysis. The students who are not familiar with it, may disregard the proofs.

8.4 Theorem
Any norm on $\mathbb{R}^n$ or $\mathbb{C}^n$ is a continuous function. Precisely: for any $\varepsilon > 0$ there is a $\delta > 0$ such that for any two vectors $u = (x_1, \dots, x_n)$ and $v = (y_1, \dots, y_n)$ satisfying $\max_i |x_i - y_i| < \delta$ we have $\Big|\,||u|| - ||v||\,\Big| < \varepsilon$.

Proof. Let $g = \max\{||e_1||, \dots, ||e_n||\}$. By the triangle inequality, for any vector

$$v = (x_1, \dots, x_n) = x_1 e_1 + \cdots + x_n e_n$$

we have

$$||v|| \le |x_1|\,||e_1|| + \cdots + |x_n|\,||e_n|| \le (|x_1| + \cdots + |x_n|) \cdot g$$

Hence if $\max\{|x_1|, \dots, |x_n|\} < \delta$, then $||v|| < ng\delta$.

Now let two vectors $v = (x_1, \dots, x_n)$ and $u = (y_1, \dots, y_n)$ be close, so that $\max_i |x_i - y_i| < \delta$. Then

$$\Big|\,||u|| - ||v||\,\Big| \le ||u - v|| < ng\delta$$

It is then enough to set $\delta = \varepsilon/ng$. This proves the continuity of the norm $||\cdot||$ as a function of v. $\square$

8.5 Corollary
Let $||\cdot||$ be a norm on $\mathbb{R}^n$ or $\mathbb{C}^n$. Denote by

$$S_1^e = \{(x_1, \dots, x_n) : |x_1|^2 + \cdots + |x_n|^2 = 1\}$$

the Euclidean unit sphere in $\mathbb{R}^n$ (or $\mathbb{C}^n$). Then the function $||\cdot||$ is bounded above and below on S_1^e:

$$0 < \min_{v \in S_1^e} ||v|| \le \max_{v \in S_1^e} ||v|| < \infty$$

Proof. Indeed, it is known in real analysis that S_1^e is a compact set. Note that $||\cdot||$ is a continuous function on S_1^e. It is known in real analysis that a continuous function on a compact set always takes its maximum and minimum values. In our case, the minimum value of $||\cdot||$ on S_1^e is strictly positive, because $0 \notin S_1^e$. This proves the corollary. □

8.6 Definition (Equivalent norms)

Two norms, $||\cdot||_a$ and $||\cdot||_b$, on V are said to be equivalent if there are constants $0 < C_1 < C_2$ such that $C_1 \leq ||u||_a/||u||_b \leq C_2 < \infty$ for all $u \neq 0$. This is an equivalence relation.

8.7 Theorem

In any finite dimensional space V, any two norms are equivalent.

Proof. Assume first that $V = \mathbb{R}^n$ or $V = \mathbb{C}^n$. It is enough to prove that any norm is equivalent to the 2-norm. Any vector $v = (x_1, \ldots, x_n) \in V$ is a multiple of a Euclidean unit vector $u \in S_1^e$, so it is enough to check the equivalence for vectors $u \in S_1^e$, which immediately follows from 8.5. So, the theorem is proved for $V = \mathbb{R}^n$ and $V = \mathbb{C}^n$. An arbitrary n-dimensional vector space over $F = \mathbb{R}$ or $F = \mathbb{C}$ is isomorphic to $\mathbb{R}^n$ or $\mathbb{C}^n$, respectively. □

8.8 Theorem + Definition (Matrix norm)

Let $||\cdot||$ be a norm on $\mathbb{R}^n$ or $\mathbb{C}^n$. Then

$$A := \sup_{||x||=1} ||Ax|| = \sup_{x \neq 0} \frac{||Ax||}{||x||}$$

defines a norm in the space of $n \times n$ matrices. It is called the *matrix norm induced by* $||\cdot||$.

Proof is a direct inspection.

Note that supremum can be replaced by maximum in Theorem 8.8. Indeed, one can obviously write $A = \sup_{||x||_2=1} ||Ax||/||x||$, then argue that the function $||Ax||$ is continuous on the Euclidean unit sphere S_1^e (as a composition of two continuous functions, Ax and $||\cdot||$), then argue that $||Ax||/||x||$ is a continuous function, as a ratio of two continuous functions, of which $||x|| \neq 0$, so that $||Ax||/||x||$ takes its maximum value on S_1^e.

Note that there are norms on $\mathbb{R}^{n\times n}$ that are not induced by any norm on $\mathbb{R}^n$, for example $||A|| := \max_{i,j} |a_{ij}|$ (Exercise, use 8.10(ii) below).

8.9 Theorem

(i) $||A||_1 = \max_{1\leq j\leq n} \sum_{i=1}^n |a_{ij}|$ (maximum column sum)
(ii) $||A||_\infty = \max_{1\leq i\leq n} \sum_{j=1}^n |a_{ij}|$ (maximum row sum)

Note: There is no explicit characterization of $||A||_2$ in terms of the a_{ij}.

8.10 Theorem

Let $|| \cdot ||$ be a norm on $\mathbb{R}^n$ or $\mathbb{C}^n$. Then
(i) $||Ax|| \leq ||A||\,||x||$ for all vectors x and matrices A.
(ii) $||AB|| \leq ||A||\,||B||$ for all matrices A, B.

8.11 Definition (Real Inner Product)

Let V be a real vector space. A *real inner product* on V is a real valued function on $V \times V$, denoted by $\langle \cdot, \cdot \rangle$, satisfying
1. $\langle u, v \rangle = \langle v, u \rangle$ for all $u, v \in V$
2. $\langle cu, v \rangle = c\langle u, v \rangle$ for all $c \in \mathbb{R}$ and $u, v \in V$
3. $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ for all $u, v, w \in V$
4. $\langle u, u \rangle \geq 0$ for all $u \in V$, and $\langle u, u \rangle = 0$ iff $u = 0$

Comments: 1 says that the inner product is symmetric, 2 and 3 say that it is linear in the first argument (the linearity in the second argument follows then from 1), and 4 says that the inner product is non-negative and non-degenerate (just like a norm). Note that $\langle 0, v \rangle = \langle u, 0 \rangle = 0$ for all $u, v \in V$.

A real vector space together with a real inner product is called a *real inner product space*, or sometimes a *Euclidean space*.

8.12 Examples

(i) $V = \mathbb{R}^n$: $\langle u, v \rangle = \sum_{i=1}^n u_i v_i$ (standard inner product). Note: $\langle u, v \rangle = u^t v = v^t u$.
(ii) $V = C([a, b])$ (real functions): $\langle f, g \rangle = \int_a^b f(x)g(x)\,dx$

8.13 Definition (Complex Inner Product)

Let V be a complex vector space. A *complex inner product* on V is a complex valued function on $V \times V$, denoted by $\langle \cdot, \cdot \rangle$, satisfying
1. $\langle u, v \rangle = \overline{\langle v, u \rangle}$ for all $u, v \in V$
2, 3, 4 as in 8.11.

A complex vector space together with a complex inner product is called a *complex inner product space*, or sometimes a *unitary space*.

8.14 Simple properties

(i) $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$
(ii) $\langle u, cv \rangle = \bar{c}\langle u, v \rangle$

The properties (i) and (ii) are called *conjugate linearity* in the second argument.

8.15 Examples

(i) $V = \mathbb{C}^n$: $\langle u, v\rangle = \sum_{i=1}^n u_i \bar{v}_i$ (standard inner product). Note: $\langle u, v\rangle = u^t \bar{v} = \bar{v}^t u$.

(ii) $V = C([a, b])$ (complex functions): $\langle f, g\rangle = \int_a^b f(x)\overline{g(x)}\, dx$

Note: the term "inner product space" from now on refers to either real or complex inner product space.

8.16 Theorem (Cauchy-Schwarz-Buniakowsky inequality)

Let V be an inner product space. Then

$$|\langle u, v\rangle| \leq \langle u, u\rangle^{1/2} \langle v, v\rangle^{1/2}$$

for all $u, v \in V$.

The equality holds if and only if $\{u, v\}$ is linearly dependent.

Proof. We do it in the complex case. Assume that $v \neq 0$. Consider the function

$$\begin{aligned} f(z) &= \langle u - zv, u - zv\rangle \\ &= \langle u, u\rangle - z\langle v, u\rangle - \bar{z}\langle u, v\rangle + |z|^2 \langle v, v\rangle \end{aligned}$$

of a complex variable z. Let $z = re^{i\theta}$ and $\langle u, v\rangle = se^{i\varphi}$ be the polar forms of the numbers z and $\langle u, v\rangle$. Set $\theta = \varphi$ and assume that r varies from $-\infty$ to ∞, then

$$0 \leq f(z) = \langle u, u\rangle - 2sr + r^2 \langle v, v\rangle$$

Since this holds for all $r \in \mathbb{R}$ (also for $r < 0$, because the coefficients are all nonnegative), the discriminant has to be ≤ 0, i.e. $s^2 - \langle u, u\rangle\langle v, v\rangle \leq 0$. This completes the proof in the complex case. It the real case it goes even easier, just assume $z \in \mathbb{R}$. The equality case in the theorem corresponds to the zero discriminant, hence the above polynomial assumes a zero value, and hence $u = zv$ for some $z \in \mathbb{C}$. (We left out the case $v = 0$, do it yourself as an exercise.) $\square$

8.17 Theorem + Definition (Induced norm)

If V is an inner product vector space, then $||v|| := \langle v, v\rangle^{1/2}$ defines a norm on V. It is called the *induced norm*.

To prove the triangle inequality, you will need 8.16.

8.18 Example

The inner products in Examples 8.12(i) and 8.15(i) induce the 2-norm on $\mathbb{R}^n$ and $\mathbb{C}^n$, respectively.

The inner products in Examples 8.12(ii) and 8.15(ii) induce the 2-norm on the spaces $C[a, b]$ of real and complex functions, respectively.

8.19 Theorem

Let V be a real vector space with norm $||\cdot||$. The norm $||\cdot||$ is induced by an inner product if and only if the function

$$\langle u, v\rangle := \frac{1}{4}\left(||u+v||^2 - ||u-v||^2\right) \qquad \text{(polarization identity)}$$

satisfies the definition of an inner product. In this case $||\cdot||$ is induced by the above inner product.

Note: A similar but more complicated polarization identity holds in complex inner product spaces.

9 Orthogonal vectors

In this section, V is always an inner product space (real or complex).

9.1 Definition (Orthogonal vectors)
Two vectors $u, v \in V$ are said to be *orthogonal* if $\langle u, v\rangle = 0$.

9.2 Example
(i) The canonical basis vectors $e_1, \ldots, e_n$ in $\mathbb{R}^n$ or $\mathbb{C}^n$ with the standard inner product are mutually (i.e., pairwise) orthogonal.
(ii) Any vectors $u = (u_1, \ldots, u_k, 0, \ldots, 0)$ and $v = (0, \ldots, 0, v_{k+1}, \ldots, v_n)$ are orthogonal in $\mathbb{R}^n$ or $\mathbb{C}^n$ with the standard inner product.
(iii) The zero vector 0 is orthogonal to any vector.

9.3 Theorem (Pythagoras)
If $\langle u, v\rangle = 0$, then $||u + v||^2 = ||u||^2 + ||v||^2$.

Inductively, it follows that if $u_1, \ldots, u_k$ are mutually orthogonal, then

$$||u_1 + \cdots + u_k||^2 = ||u_1||^2 + \cdots + ||u_k||^2$$

9.4 Theorem
If nonzero vectors $u_1, \ldots, u_k$ are mutually orthogonal, then they are linearly independent.

9.5 Definition (Orthogonal/Orthonormal basis)
A basis $\{u_i\}$ in V is said to be *orthogonal*, if all the basis vectors u_i are mutually orthogonal. If, in addition, all the basis vectors are unit (i.e., $||u_i|| = 1$ for all i), then the basis is said to be *orthonormal*, or an ONB.

9.6 Theorem (Fourier expansion)
If $B = \{u_1, \ldots, u_n\}$ is an ONB in a finite dimensional space V, then

$$v = \sum_{i=1}^{n} \langle v, u_i\rangle u_i$$

for every $v \in V$, i.e. $c_i = \langle v, u_i\rangle$ are the coordinates of the vector v in the basis B. One can also write this as $[v]_B^t = (\langle v, u_1\rangle, \ldots, \langle v, u_n\rangle)$.

Note: the numbers $\langle v, u_i\rangle$ are called the *Fourier coefficients* of v in the ONB $\{u_1, \ldots, u_n\}$.

For example, in $\mathbb{R}^n$ or $\mathbb{C}^n$ with the standard inner product, the coordinates of any vector $v = (v_1, \ldots, v_n)$ satisfy the equations $v_i = \langle v, e_i\rangle$.

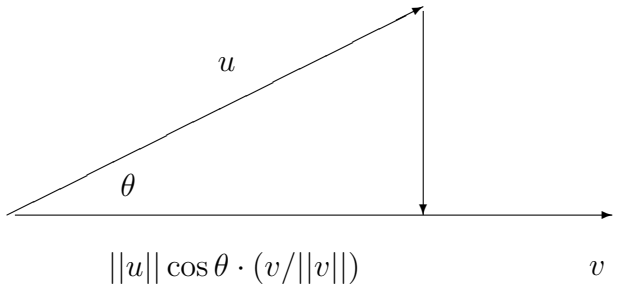
9.7 Definition (Orthogonal projection)

Let $u, v \in V$, and $v \neq 0$. The *orthogonal projection* of u onto v is

$$\mathrm{Pr}_v u = \frac{\langle u, v\rangle}{||v||^2} v$$

Note that the vector $w := u - \mathrm{Pr}_v u$ is orthogonal to v. Therefore, u is the sum of two vectors, $\mathrm{Pr}_v u$ parallel to v, and w orthogonal to v (see the diagram below).

Figure 1: Orthogonal projection of u to v



9.8 Definition (Angle)

In the real case, for any nonzero vectors $u, v \in V$ let

$$\cos\theta = \frac{\langle u, v\rangle}{||u||\,||v||}$$

By 8.16, we have $\cos\theta \in [-1, 1]$. Hence, there is a unique angle $\theta \in [0, \pi]$ with this value of cosine. It is called the *angle between* u and v.

Note that $\cos\theta = 0$ if and only if u and v are orthogonal. Also, $\cos\theta = \pm 1$ if and only if u, v are proportional, $v = cu$, then the sign of c coincides with the sign of $\cos\theta$.

9.9 Theorem

Let $B = \{u_1, \ldots, u_n\}$ be an ONB in a finite dimensional space V. Then

$$v = \sum_{i=1}^{n} \mathrm{Pr}_{u_i} v = \sum_{i=1}^{n} u_i \cos\theta_i$$

where θ_i is the angle between u_i and v.

9.10 Theorem (Gram-Schmidt)

Let nonzero vectors $w_1, \ldots, w_m$ be mutually orthogonal. For $v \in V$, set

$$w_{m+1} = v - \sum_{i=1}^{n} Pr_{w_i} v$$

Then the vectors $w_1, \ldots, w_{m+1}$ are mutually orthogonal, and

$$\text{span}\{w_1, \ldots, w_m, v\} = \text{span}\{w_1, \ldots, w_m, w_{m+1}\}$$

In particular, $w_{m+1} = 0$ if and only if $y \in \text{span}\{w_1, \ldots, w_m\}$.

9.11 Algorithm (Gram-Schmidt orthogonalization)

Let $\{v_1, \ldots, v_n\}$ be a basis in V. Define

$$w_1 = v_1$$

and then inductively, for $m \geq 1$,

$$\begin{aligned} w_{m+1} &= v_{m+1} - \sum_{i=1}^{m} \text{Pr}_{w_i} v_{m+1} \\ &= v_{m+1} - \sum_{i=1}^{m} \frac{\langle v_{m+1}, w_i \rangle}{||w_i||^2} w_i \end{aligned}$$

This gives an orthogonal basis $\{w_1, \ldots, w_n\}$, which 'agrees' with the basis $\{v_1, \ldots, v_n\}$ in the following sense:

$$\text{span}\{v_1, \ldots, v_m\} = \text{span}\{w_1, \ldots, w_m\}$$

for all $1 \leq m \leq n$.

The basis $\{w_1, \ldots, w_n\}$ can be normalized by $u_i = w_i/||w_i||$ to give an ONB $\{u_1, \ldots, u_n\}$.

Alternatively, an ONB $\{u_1, \ldots, u_n\}$ can be obtained directly by

$$w_1 = u_1 \qquad u_1 = w_1/||w_1||$$

and inductively for $m \geq 1$

$$w_{m+1} = v_{m+1} - \sum_{i=1}^{m} \langle v_{m+1}, u_i \rangle u_i \qquad u_{m+1} = w_{m+1}/||w_{m+1}||$$

9.12 Example

Let $V = P_n(\mathbb{R})$ with the inner product given by $\langle f, g \rangle = \int_0^1 f(x)g(x)\,dx$. Applying Gram-Schmidt orthogonalization to the basis $\{1, x, \ldots, x^n\}$ gives the first $n+1$ of the so called *Legendre polynomials*.

9.13 Corollary

Let $W \subset V$ be a finite dimensional subspace of an inner product space V, and $\dim W = k$. Then there is an ONB $\{u_1, \ldots, u_k\}$ in W. If, in addition, V is finite dimensional, then the basis $\{u_1, \ldots, u_k\}$ of W can be extended to an ONB $\{u_1, \ldots, u_n\}$ of V.

9.14 Definition (Orthogonal complement)

Let $S \subset V$ be a subset (not necessarily a subspace). Then

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

is called the *orthogonal complement* to S.

9.15 Theorem

$S^\perp$ is a subspace of V. If $W = \operatorname{span} S$, then $W^\perp = S^\perp$.

9.16 Example

If $S = \{(1,0,0)\}$ in $\mathbb{R}^3$, then $S^\perp = \operatorname{span}\{(0,1,0),(0,0,1)\}$.

Let $V = C[a,b]$ (real functions) with the inner product from 8.12(ii). Let $S = \{f \equiv \text{const}\}$ (the subspace of constant functions). Then $S^\perp = \{g : \int_a^b g(x)\,dx = 0\}$. Note that $V = S \oplus S^\perp$, see 1.34(b).

9.17 Theorem

If W is a finite dimensional subspace of V, then

$$V = W \oplus W^\perp$$

Proof. By 9.13, there is an ONB $\{u_1, \ldots, u_k\}$ of W. For any $v \in V$ the vector

$$v - \sum_{i=1}^{k} \langle v, u_i\rangle u_i$$

belongs in $W^\perp$. Hence, $V = W + W^\perp$. The linear independence of W and $W^\perp$ follows from 9.4. □

Note: the finite dimension of W is essential. Let $V = C[a,b]$ with the inner product from 8.12(ii) and $W \subset V$ be the set of real polynomials restricted to the interval $[a,b]$. Then $W^\perp = \emptyset$, and at the same time $V \neq W$.

9.18 Theorem (Parceval's identity)

Let $B = \{u_1, \ldots, u_n\}$ be an ONB in V. Then

$$\langle v, w\rangle = \sum_{i=1}^{n} \langle v, u_i\rangle \overline{\langle w, u_i\rangle} = [v]_B^t \overline{[w]_B} = \overline{[w]_B^t}[v]_B$$

for all $v, w \in V$. In particular,

$$||v||^2 = \sum_{i=1}^{n} |\langle v, u_i \rangle|^2 = [v]_B^t \overline{[v]_B}$$

Proof. Follows from 9.6. □

9.19 Theorem (Bessel's inequality)
Let $\{u_1, \ldots, u_n\}$ be an orthonormal subset of V. Then

$$||v||^2 \geq \sum_{i=1}^{n} |\langle v, u_i \rangle|^2$$

for all $v \in V$.

Proof. For any $v \in V$ the vector

$$w := v - \sum_{i=1}^{n} \langle v, u_i \rangle u_i$$

belongs in $\{u_1, \ldots, u_n\}^\perp$. Hence,

$$||v||^2 = ||w||^2 + \sum_{i=1}^{n} |\langle v, u_i \rangle|^2$$

9.20 Definition (Isometry)
Let V, W be two inner product spaces (both real or both complex). An *isomorphism* $T : V \to W$ is called an *isometry* if it preserves the inner product, i.e.

$$\langle Tv, Tw \rangle = \langle v, w \rangle$$

for all $v, w \in V$. In this case V and W are said to be *isometric*.

Note: it can be shown by polarization identity that $T : V \to W$ preserves inner product if and only if T preserves the induced norms, i.e. $||Tv|| = ||v||$ for all $v \in V$.

9.21 Theorem Let $\dim V < \infty$. A linear transformation $T : V \to W$ is an isometry if and only if whenever $\{u_1, \ldots, u_n\}$ is an ONB in V, then $\{Tu_1, \ldots, Tu_n\}$ is an ONB in W.

9.22 Corollary Finite dimensional inner product spaces V and W (over the same field) are isometric if and only if $\dim V = \dim W$.

9.23 Example
Let $V = \mathbb{R}^2$ with the standard inner product. Then the maps defined by matrices $A_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $A_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ are isometries of $\mathbb{R}^2$.

10 Orthogonal/Unitary Matrices

10.1 Definition

If V is a real inner product space, then isometries $V \to V$ are called *orthogonal operators*. If V is a complex inner product space, then isometries $V \to V$ are called an *unitary operators*.

Note: If $U_1, U_2 : V \to V$ are unitary operators, then so are U_1U_2 and U_1^{-1}, U_2^{-1}. Thus, unitary operators form a group, denoted by $U(V) \subset GL(V)$. Similarly, orthogonal operators on a real inner product space V make a group $O(V) \subset GL(V)$.

10.2 Definition.

A real $n \times n$ matrix Q is said to be *orthogonal* if $QQ^t = I$, i.e. Q is invertible and $Q^{-1} = Q^t$. A complex $n \times n$ matrix U is said to be *unitary* if $UU^H = I$, where $U^H = (\bar{U})^t$ is the *Hermitian transpose* of U.

Note: If Q is orthogonal, then also $Q^tQ = I$, and so Q^t is an orthogonal matrix as well. If U is unitary, then also $U^HU = I$, and so U^H is a unitary matrix, too. In the latter case, we also have $\bar{U}U^t = I$ and $U^t\bar{U} = I$.

10.3 Theorem

(i) The linear transformation in $\mathbb{R}^n$ defined by a matrix $Q \in \mathbb{R}^{n\times n}$ preserves the standard (Euclidean) inner product if and only if Q is orthogonal.

(ii) The linear transformation in $\mathbb{C}^n$ defined by a matrix $U \in \mathbb{C}^{n\times n}$ preserves the standard (Hermitian) inner product if and only if U is unitary.

Proof. In the complex case: $\langle Ux, Uy\rangle = (Ux)^t\overline{Uy} = x^tU^t\bar{U}\bar{y} = x^t\bar{y} = \langle x, y\rangle$. The real case is similar. □

10.4 Theorem

(i) An operator $T : V \to V$ of a finite dimensional complex space V is unitary iff the matrix $[T]_B$ is unitary in any ONB B.

(ii) An operator $T : V \to V$ of a finite dimensional real space V is orthogonal iff the matrix $[T]_B$ is orthogonal in any ONB B.

10.5 Corollary

Unitary $n \times n$ matrices make a group, denoted by $U(n)$. Orthogonal $n \times n$ matrices make a group, denoted by $O(n)$.

10.6 Theorem

(i) A matrix $U \in \mathbb{C}^{n\times n}$ is unitary iff its columns (resp., rows) make an ONB of $\mathbb{C}^n$.

(ii) A matrix $Q \in \mathbb{R}^{n\times n}$ is orthogonal iff its columns (resp., rows) make an ONB of $\mathbb{R}^n$.

10.7 Theorem

If Q is orthogonal, then $\det Q = \pm 1$. If U is unitary, then $|\det U| = 1$, i.e. $\det U = e^{i\theta}$ for some $\theta \in [0, 2\pi]$.

Proof. In the complex case: $1 = \det I = \det U^H U = \det \bar{U}^t \cdot \det U = |\det U|^2$. The real case is similar. □

Note: Orthogonal matrices have determinant 1 or -1. Orthogonal $n \times n$ matrices with determinant 1 make a subgroup of $O(n)$, denoted by $SO(n)$.

10.8 Example

The orthogonal matrix $Q = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ represents the counterclockwise rotation of $\mathbb{R}^2$ through the angle θ.

10.9 Theorem

If λ is an eigenvalue of an orthogonal or unitary matrix, then $|\lambda| = 1$. In the real case this means $\lambda = \pm 1$.

Proof. If $Ux = \lambda x$ for some $x \neq 0$, then $\langle x, x\rangle = \langle Ux, Ux\rangle = \langle \lambda x, \lambda x\rangle = |\lambda|^2 \langle x, x\rangle$, so that $|\lambda|^2 = 1$. □

10.10 Theorem

Let $T : V \to V$ be an isometry of a finite dimensional space V. If a subspace $W \subset V$ is invariant under T, then so is its orthogonal complement $W^\perp$.

10.11 Theorem

Any unitary operator T of a finite dimensional complex space is diagonalizable. Furthermore, there is an ONB consisting of eigenvectors of T. Any unitary matrix is diagonalizable.

Proof goes by induction on the dimension of the space, use 10.10. □

10.12 Theorem

Let $T : V \to V$ be an orthogonal operator on a finite dimensional real space V. Then $V = V_1 \oplus \cdots \oplus V_m$, where V_i are mutually orthogonal subspaces, each V_i is a T-invariant one- or two-dimensional subspace of V.

Proof. If T has an eigenvalue, we can use 10.10 and reduce the dimension of V by one. Assume now that T has no eigenvalues. The characteristic polynomial $C_T(x)$ has

real coefficients and no (real) roots, so its roots are pairs of conjugate complex numbers (because $C_T(z) = 0 \Leftrightarrow C_T(\bar{z}) = 0$). So, $C_T(x)$ is a product of quadratic polynomials with no real roots: $C_T(x) = P_1(x) \cdots P_k(x)$, where $\deg P_i(x) = 2$. By Cayley-Hamilton,

$$C_T(T) = P_1(T) \cdots P_k(T) = 0$$

Hence, at least one operator $P_i(T)$, $1 \leq i \leq k$, is not invertible (otherwise $C_T(T)$ would be invertible). Let $P_i(x) = x^2 + ax + b$. Then the operator $T^2 + aT + bI$ has a nontrivial kernel, i.e. $T^2v + aTv + bv = 0$ for some vectror $v \neq 0$. Let $w = Tv$ (note that $w \neq 0$, since the operator T is an isometry). We get $Tw + aw + bv = 0$, so that

$$Tv = w \qquad \text{and} \qquad Tw = -aw - bv$$

Hence, the subspace $\text{span}\{v, w\}$ is invariant under T. It is two-dimensional, since v and w are linearly independent (otherwise $Tv = w = \lambda v$, and λ would be an eigenvalue of T). Now we can use 10.10 and reduce the dimension of V by two. $\square$