MAT224 Linear Algebra II
Lecture Notes

Yuchen Wang

April 4, 2019

Contents

1 Vector Spaces

1.1 Vector Spaces

Definition 1.1.1 A (real) vector space is a set V (whose elements are called vectors) together with

1. an operation called vector addition, which for each pair of vectors $\mathbf{x}, \mathbf{y} \in V$ produced another vector in V denoted $\mathbf{x} + \mathbf{y}$, and
2. an operation called multiplication by a scalar (a real number), which for each vector $\mathbf{x} \in V$, and each scalar $c \in \mathbb{R}$ produced another vector in V denoted $c\mathbf{x}$

Furthermore, the two operations must satisfy the following axioms: $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in V, \forall c, d \in \mathbb{R}$,

1. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$
2. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$
3. $\exists \mathbf{0} \in V$ s.t. $\mathbf{x} + \mathbf{0} = \mathbf{x}$ (additive identity)
4. $\exists -\mathbf{x} \in V$ s.t. $\mathbf{x} + -\mathbf{x} = \mathbf{0}$ (additive inverse)
5. $c(\mathbf{x} + \mathbf{y}) = c\mathbf{x} + c\mathbf{y}$
6. $(c + d)\mathbf{x} = c\mathbf{x} + d\mathbf{x}$
7. $(cd)\mathbf{x} = c(d\mathbf{x})$
8. $1\mathbf{x} = \mathbf{x}$

Smooth functions C^∞
Most functions are not smooth.

1.2 Subspaces

Example $C^\infty(\mathbb{R}) < C^k(\mathbb{R}) <$ Differentiable functions $< C(\mathbb{R}) < F(\mathbb{R})$

Definition Let V be a vector space and Let $W \subseteq V$ be a subset. Then W is a (vector) subspace of V if W is a vector space itself under the operations of vector sum and scalar multiplication from V.

Theorem 1.2.8 Let V be a vector space and Let $W \subseteq V$ be a nonempty subset of V. Then W is a subspace of V iff $\forall \mathbf{x}, \mathbf{y} \in W$, and all $c \in \mathbb{R}$, we have $c\mathbf{x} + \mathbf{y} \in W$.
proof: $\rightarrow$: If W is a subspace of V, then $\forall \mathbf{x}, \mathbf{y} \in W$ and $c \in \mathbb{R}, c\mathbf{x} + \mathbf{y} \in W$ holds since W itself is a real vector space.
$\leftarrow$: If $\forall \mathbf{x}, \mathbf{y} \in W$, and all $c \in \mathbb{R}$, we have $c\mathbf{x} + \mathbf{y} \in W$
Can have $c = 1$, so $\mathbf{x} + \mathbf{y} \in W$ (close under addition)
$c = -1$ and $\mathbf{y} = \mathbf{x}$, so $-\mathbf{x} + \mathbf{x} = \mathbf{0} \in W$ (additive identity)
$\mathbf{y} = \mathbf{0}$, so $c\mathbf{x} \in W$ (close under scalar multiplication)
These implies all the axioms. ■

Examples

1. $W = \{f \in C(\mathbb{R}) | f(\pi) = 0\}$. W subspace of $C(\mathbb{R})$? -Yes
2. $W = \{f \in C(\mathbb{R}) | f(e) = e\}$. W subspace of $C(\mathbb{R})$? -No, not close under addition
3. $W = \{(x_1, ..., x_n) | x_i \geq 0 \forall i\}$. W subspace of $C(\mathbb{R})$? -No, there is no additive inverse for each item in W.

Theorem 1.2.13 Let V be a vector space. Then the intersection of any collection of subspaces of V is a subspace of V.
proof: Consider any collection of subspace of V. Note that the intersection of the subspaces is not empty since at least the zero vector from V is in it. Now let $\mathbf{x}, \mathbf{y}$ be any two vectors in the intersection, so they are in every single subspace in the collection. Therefore $c\mathbf{x} + \mathbf{y}$ is also in every single subspace in the collection, so that it is in the intersection as well. Hence the intersection is a subspace of V. ■

Application The set of all solutions of any simultaneous system of equations is a subspace of $\mathbb{R}^n$.

Corollary 1.2.14 Let $a_{ij} (1 \leq i \leq m, 1 \leq j \leq n)$ be any real numbers and let $W = \{(x_1, ..., x_n) \in \mathbb{R}^n | a_{i1}x_1 + \ldots + a_{in}x_n = 0 \text{ for all } i, 1 \leq i \leq m\}$. Then W is a subspace of $\mathbb{R}^n$.

1.3 Linear Combinations

Definition 1.3.1 Let S be a subset of a vector space V.

1. A *linear combination* of vectors in S is any sum $a_1\mathbf{x}_1 + \ldots + a_n\mathbf{x}_n$ where the $a_i \in \mathbb{R}$, and the $\mathbf{x}_i \in S$

2. If $S \neq \emptyset$, the set of all linear combinations of vectors in S is called the *span* of S, and denoted Span(S). If $S = \emptyset$, we define Span(S) $= \{\mathbf{0}\}$. (Remark: It is a mathematician convention)

3. If W = Span(S), we say S spans (or generates) W.

Theorem 1.3.4 Let V be a vector space and let S be any subset of V. Then Span(S) is a subspace of V.

proof: Span(S) is non-empty by definition. Let $\mathbf{x}, \mathbf{y} \in Span(S)$, then they are linear combinations of vectors in S. Check that $c\mathbf{x} + \mathbf{y}$ is also a linear combination of vectors in S, so $c\mathbf{x} + \mathbf{y} \in Span(S)$. Hence Span(S) is a subspace of V. ■

Definition Let W_1 and W_2 be subspaces of a vector space V. The *sum* of W_1 and W_2 is the set

$$W_1 + W_2 = \{\mathbf{x} \in V | \mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2, \text{ for some } \mathbf{x}_1 \in W_1 \text{ and } \mathbf{x}_2 \in W_2\}$$

Proposition 1.3.8 The basis of sum is the union of two bases Let $W_1 = Span(S_1)$ and $W_2 = Span(S_2)$ be subspaces of a vector space V. Then $W_1 + W_2 = Span(S_1 \cup S_2)$

Theorem 1.3.9 Let W_1 and W_2 be subspaces of a vector space V. Then $W_1 + W_2$ is also a subspace of V.

Proposition 1.3.11 $W_1 + W_2$ is the smallest subspace containing $W_1 \cup W_2$: Let W_1 and W_2 be subspaces of a vector space V and let W be a subspace of V such that $W_1 \cup W_2 \subseteq W$. Then $W_1 + W_2 \subseteq W$

Remark $W_1 \cup W_2$ is a subspace of V iff one is contained in another.

1.4 Linear Dependence and Linear Independence

Definitions 1.4.2 Let V be a vector space, and let S be a subset of V.

1. A *linear dependence* among the vectors of S is an equation

$$a_1\mathbf{x}_1 + \ldots + a_n\mathbf{x}_n = \mathbf{0}$$

   where the $x_i \in S$, and the $a_i \in \mathbb{R}$ are not all zero (i.e., at least one of the $a_i \neq \mathbf{0}$

2. the set S is said to be *linearly dependent* if there exists a linear dependence among the vectors in S.

Fact Let S be a set. If $\mathbf{0} \in S$, then S is dependent.

Definition 1.4.4 A subset S of a vector space V is *linearly independent* if whenever we have $a_i \in \mathbb{R}$ and $\mathbf{x}_i \in S$ such that $a_1\mathbf{x}_1 + \ldots + a_n\mathbf{x}_n = \mathbf{0}$, then $a_i = 0$ for all i.

Example In any vector space the empty subset $\emptyset$ is linearly independent.

Proposition 1.4.7

1. Let S be a linearly independent subset of a vector space V, and let S' be another subset of V that contains S. Then S' is also linearly dependent.

2. Let S be linearly independent subset of a vector space V and let S' be another subset of V that is contained in S. Then S' is also linearly independent.

1.5 Interlude on Solving Systems of Linear Equations (MAT223)

1.6 Bases And Dimension (Jan 17)

Definition A subset S of vector space V is called a *basis* of V if V = Span(S) and S is linearly independent.

Remark A basis is the maximal set of linearly independent vectors / minimal set of spanning vectors.

Examples

1. the standard basis S = $\{e_1,...,e_n\}$ in $\mathbb{R}^n$, since every vector $(a_1, ..., a_n) \in \mathbb{R}^n$ may be written as the linear combination $(a_1, ..., a_n) = a_1e_1 + ... + a_ne_n$

2. The vector space $\mathbb{R}^n$ has many other bases as well. e.g., in $\mathbb{R}^2$, consider the set $S = \{(1, 2), (1, -1)\}$, which is l.i.

3. Let $V = P_n(\mathbb{R})$ and consider $S = \{1, x, x^2, ..., x^n\}$, which is a basis of V.
   proof: It is clear that S spans V. For independence, consider

   $$a_0 + a_1x + a_2x^2 + ... + a_{n-1}x^{n-1} + a_nx^n = \mathbf{0}$$

   Take the derivative of both sides,

   $$\frac{d^n}{dx^n}(a_0 + a_1x + a_2x^2 + ... + a_{n-1}x^{n-1} + a_nx^n) = \frac{d^n}{dx^n}(0)$$

   $$n!a_n = 0 \implies a_n = 0$$

   Similarly, we have $a_i = 0$ for all i, as wanted.

4. The empty subset, $\emptyset$, is a basis of the vector space consisting only of a zero vector, $\{\mathbf{0}\}$.

Theorem 1.6.3 Let V be a vector space, and let S be a nonempty subset of V. Then S is a basis of V iff every vector $\mathbf{x} \in V$ may be written uniquely as a linear combination of the vectors in S.
Proof: $\rightarrow$: Assume S is a basis of V, then given $\mathbf{x} \in V$, there are scalars $a_i \in \mathbb{R}$ and vectors $x_i \in S$ s.t. $\mathbf{x} = a_1x_1 + ... + a_nx_n$. To show this linear combination is unique, consider a possible second linear combination of vectors in S which also adds up to $\mathbf{x}$: $x = b_1x_1 + ... + b_nx_n$. Subtracting these two expressions for $\mathbf{x}$, we find that

$$\mathbf{0} = a_1x_1 + ... + a_nx_n - (b_1x_1 + ... + b_nx_n)$$

$$= (a_1 - b_1)x_1 + ... + (a_n - b_n)x_n$$

Since S is linearly independent, the equation implies that $a_i = b_i$ for all i.

$\leftarrow$: Assume every vector $\mathbf{x} \in V$ may be written uniquely as a linear combination of the vectors in S. This implies $Span(S) = V$. We must show that S is l.i. Consider an equation

$$a_1x_1 + ... + a_nx_n = \mathbf{0}$$

Note that it is also the case that

$$0\mathbf{x} = 0(x_1 + ... + x_n) = \mathbf{0}$$

Since we assumed that every $\mathbf{x}$ has a unique representation in S, then it must be true that $a_i = 0$ for all i. Hence S is l.i.

Theorem 1.6.6 Let V be a vector space that has a finite spanning set, and let S be a linearly independent subset of V. Then there exists a basis S' of V, with $S \subset S'$

Lemma 1.6.8 Let S be a linearly independent subset of V and let $x \in V$, but $x \notin S$. Then $S \cup \{\mathbf{x}\}$ is l.i. iff $\mathbf{x} \notin Span(S)$.

Insight the number of vectors in a basis is, in a rough sense, a measure of "how big" the space is.

Theorem 1.6.10 (Basis Theorem) Let V be a vector space and let S be a spanning set for V, which has m elements. Then no linearly independent set in V can have more than m elements.

proof: It suffices to show that every set in V with more than m elements is linearly dependent. Write $S = y_1, ..., y_m$ and suppose $S' = x_1, ..., x_n$ is a subset of V with $n > m$ vectors. Consider an equation

$$(1)a_1x_1 + ... + a_nx_n = \mathbf{0}$$

Our goal is to show that a_i not all 0. Since S spans V, there are scalars b_{ij} s.t. for each i,

$$x_i = b_{i1}y_1 + ... + b_{im}y_m$$

Substituting these equations into (1), we get

$$a_1(b_{11}y_1 + ... + b_{1m}y_m) + ... + a_n(b_{n1}y_1 + ... + b_{nm}y_m) = \mathbf{0}$$

Collecting terms and rearranging,

$$(a_1b_{11} + ... + a_nb_{n1})y_1 + ... + (a_1b_{1m} + ... + a_nb_{nm})y_m = \mathbf{0}$$

Since S is l.i., this is equivalent to solving the system

$$b_{11}a_1 + ... + b_{n1}a_n = 0$$

$$\vdots$$

$$b_{1m}a_1 + ... + b_{nm}a_n = 0$$

But this is a system with n unknowns and m equations and $n > m$, so there must exist a non-trivial solution $\{a_1, ..., a_n\}$, which is what we wanted to show. ■

Corollary 1.6.11 Let V be a vector space and let S and S' be two bases of V, with m and m' elements, respectively. Then $m = m'$.

proof:
Since S is a spanning set of V and S' is l.i., we have $m' \leq m$. Since S' is a spanning set of V and S is l.i.m we have $m \leq m'$. Hence $m = m'$. ■

Definitions 1.6.12

1. If V is a vector space with some finite basis(possibly empty), we say V is *finite-dimentional*.
2. Let V be a finite-dimensional vector space. The dimension of V, denoted dim(V), is the number of vectors in a (hence any) basis of V.
3. If $V = \{\mathbf{0}\}$, we define dim(V) = 0.

Examples

1. For each n, $\dim(\mathbb{R}^n) = n$, since the standard basis contains n vectors.
2. $\dim(P_n(\mathbb{R})) = n + 1$, since a basis for $P_n(\mathbb{R})$ contains n + 1 functions.
3. The vector spaces $P(\mathbb{R}), C^1(\mathbb{R})$ and $C(\mathbb{R})$ are not finite-dimensional. We say that such spaces are *infinite-dimentional*.
4. $\dim(Span\{(1, 2, 3), (4, 5, 6), (7, 8, 9)\}) = 2$

Corollary 1.6.14 Let W be a subspace of a finite-dimensional vector space V. Then $\dim(W) \leq \dim(V)$. Furthermore, $\dim(W) = \dim(V)$ iff $W = V$.

Corollary 1.6.15 Let W be a subspace of $\mathbb{R}^n$ defined by a system of homogeneous linear equations. Then dim(W) is equal to the number of free variables in the corresponding echelon form system.

Theorem 1.6.18 Let W_1 and W_2 be finite-dimensional subspaces of a vector space V. Then

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

Remark Analogous to the Principle of Inclusion-Exclusion

proof: Result obvious if either W_1 or W_2 is $\{\mathbf{0}\}$.
Therefore, we assume that neither W_1 nor W_2 is $\{\mathbf{0}\}$. Starting from a basis S of $W_1 \cap W_2$. We can always find sets T_1 and T_2 (disjoint from S) such that $S \cup T_1$ is a basis for W_1 and $S \cup T_2$ is a basis for W_2. We claim that $U = S \cup T_1 \cup T_2$ is a basis for $W_1 + W_2$, since

$$U = S \cup T_1 \cup T_2 = (S \cup T_1) \cup (S \cup T_2)$$

$$Span(U) = Span((S \cup T_1) \cup (S \cup T_2)) = W_1 + W_2$$

Next, prove that U is linearly independent. Any potential linear dependence among the vectors in U must have the form

$$\mathbf{v} + \mathbf{w}_1 + \mathbf{w}_2 = \mathbf{0}$$

where $\mathbf{v} \in Span(S) = W_1 \cap W_2, \mathbf{w}_1 \in Span(T_1) \subset W_1, \mathbf{w}_2 \in Span(T_2) \subset W_2$. (slice the linear combination into the sum of the vectors from 3 vector spaces). It suffices to prove that in any such potential linear dependence, we must have $\mathbf{v} = \mathbf{w}_1 = \mathbf{w}_2 = \mathbf{0}$ (each vector is a lin comb, and equals $\mathbf{0}$).
Consider $\mathbf{w}_2 = -\mathbf{v} - \mathbf{w}_1$. Since $-\mathbf{v} - \mathbf{w}_1 \in W_1, \mathbf{w}_2 \in W_2$, we must have $\mathbf{w}_2 \in W_1 \cap W_2$. By definition, $\mathbf{w}_2 \in Span(T_2)$ But $S \cap T_2 = \emptyset$, hence $Span(S) \cap Span(T_2) = \{\mathbf{0}\}$. Therefore we must have $\mathbf{w}_2 = \mathbf{0}$. So then $-\mathbf{v} = \mathbf{w}_1 \in W_1 \cap W_2$. Since $S \cap T_1 = \emptyset$, $Span(S) \cap Span(T_1) = \{\mathbf{0}\}$ and we have $\mathbf{w}_1 = \mathbf{0}$, so $\mathbf{v} = \mathbf{0}$ as well. So U is independent.

$$\begin{aligned}|U| &= |S| + |T_1| + |T_2| \\ &= \dim W_1 \cap W_2 + (\dim W_1 - \dim W_1 \cap W_2) + (\dim W_2 - \dim W_1 \cap W_2) \\ &= \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2)\end{aligned}$$

Exercises for 1.4 1.(k), 7

Exercises for 1.6 1.(d)(e)(f), 3, 4, 16

2 Linear Transformations

2.1 Linear Tranformations

A function T from V to W is denoted by $T : V \to W$. The vector $\mathbf{w} = T(\mathbf{v})$ in W is called the *image* of $\mathbf{v}$ under the function T. Loosely speaking, we want our functions to turn the algebraic operations of addition and scalar multiplication in V into addition and scalar multiplication in W.

Definition 2.1.1 A function $T : V \to W$ is called a *linear mapping* or a *linear transformation* if it satisfies

1. $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}$ and $\mathbf{v} \in$ V
2. $T(a\mathbf{v}) = aT(\mathbf{v})$ for all $a \in \mathbb{R}$ and $\mathbf{v} \in V$

V is called the *domain* of T and W is called the *target* of T.
We say that a linear transformation preserves the operations of addition and scalar multiplication.

Property A linear mapping always takes the zero vector in the domain vector space to the zero vector in the target vector space.

Proposition 2.1.2 A function $T : V \to W$ is a linear transformation if and only if for all a and $b \in \mathbb{R}$ and all $\mathbf{u}$ and $\mathbf{v} \in V$

$$T(a\mathbf{u} + b\mathbf{v}) = aT(\mathbf{u}) + bT(\mathbf{v})$$

Corollary 2.1.3 A function $T : V \to W$ is a linear transformation if and only if for all $a_1, .., a_k \in \mathbb{R}$ and for all $\mathbf{v}_1, ..., \mathbf{v}_k \in V$:

$$T(\sum_{i=1}^{k} a_i\mathbf{v}_i) = \sum_{i=1}^{k} a_i T(\mathbf{v}_i)$$

Remark prove this by induction!

Examples

1. Let V be any vector space, and let W = V. The <u>*identity transformation*</u> $I : V \to V$ is defined by I($\mathbf{v}$) = $\mathbf{v}$ for all $\mathbf{v} \in$ V.

2. Let V and W be any vector spaces, and let $T : V \to W$ be the mapping that takes every vector in V to the zero vector in W:

$$T(\mathbf{v}) = \mathbf{0}_W$$

for all $\mathbf{v} \in$ V. T is called *zero transformation*.

3. *$T(\boldsymbol{x}) = a_1x_1 + ... + a_nx_n$*

4. *Differentiation, definite integration*

Remark The inner product plays a crucial role in linear algebra in that it provides a bridge between algebra and geometry, which is the heart of the more advanced material that appears later in the text.

Proposition 2.1.14 If $T : V \to W$ is a linear transformation and V is finite-dimensional, then T is uniquely determined by its values on the members of a basis of V.
proof: Show that if S and T are linear transformations that take the same values on each member of a basis for V, then in fact S = T.

$$\begin{aligned} T(v) &= T(a_1v_1 + ... + a_kv_k) \\ &= a_1T(v_1) + ... + a_kT(v_k) \\ &= a_1S(v_1) + ... + a_kS(v_k) \\ &= S(a_1v_1 + ... + a_kv_k) \\ &= S(v) \end{aligned}$$

Therefore, S and T are equal as mappings from V to W. ■

2.2 Linear Transformations Between Finite-Dimensional Vector Spaces

Proposition 2.2.1 Let $T : V \to W$ be a linear transformation between the finite-dimensional vector spaces V and W. If $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ is a basis for V and $\{\mathbf{w}_1, ..., \mathbf{w}_l\}$ is a basis for W, then $T : V \to W$ is uniquely determined by the $l \cdot k$ scalars used to express $T(\mathbf{v}_j), j = 1, ..., k$, in terms of $\mathbf{w}_1, ..., \mathbf{w}_l$.

Definition 2.2.6 Let $T : V \to W$ be a linear transformation between the finite-dimensional vector spaces V and W, and let $\alpha = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ and $\beta = \{\mathbf{w}_1, ..., \mathbf{w}_l\}$, respectively, be any bases for V and W. Let $a_{ij}, 1 \le i \le l$ and $1 \le j \le k$ be the $l \cdot k$ scalars that determine T with respect to the bases α and β. The matrix whose entries are the scalars a_{ij}, $1 \le i \le l$ and $1 \le j \le k$, is called the *matrix of the linear transformation T with respect to the bases α for V and β for W.* This matrix is denoted by $[T]_\alpha^\beta$.

Remark The basis vectors in the domain and target spaces are written in some particular order.

Definition of coordinate vectors If $\mathbf{v} = a_1\mathbf{v}_1 + ... + a_k\mathbf{v}_k$ and $\mathbf{w} = b_1\mathbf{w}_1 + ... + b_l\mathbf{w}_l$, we can express $\mathbf{v}$ and $\mathbf{w}$ in coordinates, respectively, as a $k \times 1$ matrix and as an $l \times 1$ matrix, with respect to the chosen bases. These coordinate vectors will be denoted by $[\mathbf{v}]_\alpha$ and $[\mathbf{w}]_\beta$, respectively. Thus $[\mathbf{v}]_\alpha = \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}$ and $[\mathbf{w}]_\beta = \begin{bmatrix} b_1 \\ \vdots \\ b_l \end{bmatrix}$

Proposition 2.2.15 Let $T : V \to W$ be a linear transformation between vector spaces V of dimension k and W of dimension l. Let $\alpha = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ be a basis for V and $\beta = \{\mathbf{w}_1, ..., \mathbf{w}_l\}$ be a basis for W. Then for each $\mathbf{v} \in V$,

$$[T(\mathbf{v})]_\beta = [T]_\alpha^\beta [\mathbf{v}]_\alpha$$

<u>*proof:*</u> Let $\mathbf{v} = x_1\mathbf{v}_1 + ... + x_k\mathbf{v}_k \in V$. Then if $T(\mathbf{v}_j) = a_{1j}\mathbf{w}_1 + ... + a_{lj}\mathbf{w}_l$

$$\begin{aligned} T(\mathbf{v}) &= \sum_{j=1}^{k} x_j T(\mathbf{v}_j) \\ &= \sum_{j=1}^{k} x_j (\sum_{i=1}^{l} a_{ij}\mathbf{w}_i) \\ &= \sum_{i=1}^{l} (\sum_{j=1}^{k} x_j a_{ij})\mathbf{w}_i \end{aligned}$$

Thus, the ith coefficient of T($\mathbf{v}$) in terms of β is $\sum_{j=1}^{k} x_j a_{ij}$ and $[T(\mathbf{v})]_\beta = \begin{bmatrix} \sum_{j=1}^{k} x_j a_{1j} \\ \vdots \\ \sum_{j=1}^{k} x_j a_{lj} \end{bmatrix}$ which is precisely $[T(\mathbf{v})]_\beta = [T]_\alpha^\beta [\mathbf{v}]_\alpha$. ∎

Proposition 2.2.19 Let $\alpha = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ be a basis for V and $\beta = \{\mathbf{w}_1, ..., \mathbf{w}_l\}$ be a basis for W, and let $\mathbf{v} = x_1\mathbf{v}_1 + ... + x_k\mathbf{v}_k \in V$

1. If A is an $l \times k$ matrix, then the function
$$T(\mathbf{v}) = \mathbf{w}$$
where $[\mathbf{w}]_\beta = A[\mathbf{v}]_\alpha$ is a linear transformation.
2. If $A = [S]_\alpha^\beta$ is the matrix of a transformation $S : V \to W$, then the transformation T constructed from $[S]_\alpha^\beta$ is equal to S.
3. If T is the transformation of (1) constructed from A, then $[T]_\alpha^\beta = A$

Proposition 2.2.20 Let V and W be finite-dimensional vector spaces. Let α be a basis for V and β a basis for W. Then the assignment of a matrix to a linear transformation from V to W given by T goes to $[T]_\alpha^\beta$ is injective and surjective.

Notes

1. When proving a function T is not a linear transformation, can consider $T(\mathbf{0}) \neq \mathbf{0}$.

2.3 Kernel and Image

Definition 2.3.1 The *kernel* of T, denoted $Ker(T)$, is the subset of V consisting of all vectors $\mathbf{v} \in V$ such that $T(\mathbf{v}) = 0$.

Remark Kernel is different from null spaces. A null space is about a matrix, and it is something in $\mathbb{R}^n$.

Proposition 2.3.2 Let $T : V \to W$ be a linear transformation. Ker(T) is a subspace of V.

Examples

1. Let $V = P_3(\mathbb{R})$. Define $T : V \to V$ by $T(p(x)) = \frac{d}{dx}p(x)$. Ker(T) only consists constant polynomials.
2. Let $V = W = \mathbb{R}^2$. Let T be a rotation R_θ. Then $Ker(T) = \{\mathbf{0}\}$.

Proposition 2.3.7 Let $T : V \to W$ be a linear transformation of finite-dimensional vector spaces, and let α and β be bases for V and W, respectively. Then $\mathbf{x} \in Ker(T)$ ife lf the coordinate vector of $\mathbf{x}$, $[\mathbf{x}]_\alpha$, satisfies the system of equations

$$\begin{aligned} a_{11}x_1 + ... + a_{1k}x_k &= 0 \\ &\vdots \\ a_{l1}x_1 + ... + a_{lk}x_k &= 0 \end{aligned}$$

where the coefficients a_{ij} are the entries of the matrix $[T]_\alpha^\beta$.

Remark This says

$$x \in \ker(T) \iff [x]_\alpha \in Nul[T]_\alpha^\beta$$

Proposition 2.3.8 Independence is Basis-Independent Let V be a finite-dimensional vector space, and let $\alpha = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ be a basis for V. Then the vectors $\mathbf{x}_1, ..., \mathbf{x}_m \in V$ are linearly independent iff their corresponding coordinate vectors $[\mathbf{x}_1]_\alpha, ..., [\mathbf{x}_m]_\alpha$ are linearly independent.

Definition 2.3.10 The subset of W consisting of all vectors $\mathbf{w} \in W$ for which there exists a $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{w}$ is called the *image* of T and is denoted by $Im(T)$.

Proposition 2.3.11 Let $T : V \to W$ be a linear transformation. The image of T is a subspace of W.

Proposition 2.3.12 If $\{\mathbf{v}_1, ..., \mathbf{v}_m\}$is any set that spans V (in particular, it could be a basis of V), then $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_m)\}$ spans $Im(T)$.

Corollary 2.3.13 If $\alpha = \{\mathbf{v}_1, ..., \mathbf{v}_k\}$ is a basis for V and $\beta = \{\mathbf{w}_1, ..., \mathbf{w}_l\}$ is a basis for W, then the vectors in W, whose coordinate vectors (in terms of β) are the columns of $[T]_\alpha^\beta$, span $Im(T)$.

Rank-Nullity Theorem 2.3.17 If V is finite-dimensional vector space and $T : V \to W$ is a linear transformation, then

$$\dim(Ker(T)) + \dim(Im(T)) = \dim(V)$$

Equivalently,

$$Nullity(T) + Rank(T) = \dim(V)$$

2.4 Applications of the Dimension Theorem

Proposition 2.4.2 A linear transformation $T : V \to W$ is injective iff $\dim(Ker(T)) = 0$, or $\dim(Im(T)) = \dim(V)$.

Remark Analogously, in MAT223 we said that a matrix is one-to-one if all the columns are l.i..

Corollary 2.4.3 A linear mapping $T : V \to W$ on a finite-dimensional vector space V is injective iff $\dim(Im(T)) = \dim(V)$.

Corollary 2.4.4 If $\dim(W) < \dim(V)$ and $T : V \to W$ is a linear mapping, then T is not injective.
proof:

$$\dim(Im(T)) \leq \dim(W) < \dim(V)$$

$$\implies dim(Ker(T)) > 0$$

Proposition 2.4.7 If W is finite-dimensional, then a linear mapping $T : V \to W$ is surjective iff $\dim(Im(T)) = \dim(W)$

Remark Analogously, in MAT223 we said that a matrix $A \in M_{m\times n}(\mathbb{R})$ is onto if there is a pivot in every row, or the columns of A spans $\mathbb{R}^m$.

Corollary 2.4.8 If V and W are finite-dimensional, with $\dim(V) < \dim(W)$, then there is no surjective linear mapping $T : V \to W$
proof: $\dim(Im(T)) \leq \dim(V) < \dim(W) \implies T$ is not surjective

Corollary 2.4.9 A linear mapping $T : V \to W$ can be surjective iff

$$\dim(V) \geq \dim(W)$$

Proposition 2.4.10 Let $\dim(V) = \dim(W)$. A linear transformation $T : V \to W$ is injective iff it is surjective.

Proposition 2.4.11 Let $T : V \to W$ be a linear transformation, and let $w \in Im(T)$. Let v_1 be any fixed vector with $T(v_1) = w$. Then every vector $v_2 \in T^{-1}(\{w\})$ can be written uniquely as $v_2 = v_1 + u$, where $u \in Ker(T)$

Remark In this situation $T^{-1}(\{w\})$ is a subspace of V iff $w = 0$.

Corollary 2.4.15 Let $T : V \to W$ be a linear transformation of finite-dimensional vector spaces, and let $w \in W$. Then there is a unique vector $v \in V$ such that $T(v) = w$ iff

1. $w \in Im(T)$, and
2. dim(Ker(T)) = 0

Proposition 2.4.16 With notation as before

1. The set of solutions of the system of linear equations $A\mathbf{x} = \mathbf{b}$ is the subset $T^{-1}(\{\mathbf{b}\})$ of $V = \mathbb{R}^n$
2. The set of solutions of the system of linear equations $A\mathbf{x} = \mathbf{b}$ is a subspace of V iff the system is homogeneous, in which case the set of solutions is $Ker(T)$.

Corollary 2.4.17

1. The number of free variables in the homogeneous system $A\mathbf{x} = \mathbf{0}$ (or its echelon form equivalent) is equal to $\dim(Ker(T))$
2. The number of basic variables of the system is equal to $\dim(Im(T))$

Definition 2.4.18 Given an inhomogeneous system of equations, $A\mathbf{x} = \mathbf{b}$, any single vector $\mathbf{x}$ satisfying the system (necessarily $\mathbf{x} \neq \mathbf{0}$) is called a particular solution of the system of equations.

Proposition 2.4.19 Let $\mathbf{x}_p$ be a particular solution of the system $A\mathbf{x} = \mathbf{b}$. Then every other solution to $A\mathbf{x} = \mathbf{b}$ is of the form $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$, where $\mathbf{x}_h$ is a solution of the corresponding homogeneous system of equations $A\mathbf{x} = \mathbf{0}$. Furthermore, given $\mathbf{x}$ and $\mathbf{x}_p$, there is a unique $\mathbf{x}_h$ such that $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$.

Corollary 2.4.20 The system $A\mathbf{x} = \mathbf{b}$ has a unique solution iff $\mathbf{b} \in Im(T)$ and the only solution to $A\mathbf{x} = \mathbf{0}$ is the zero vector.

2.5 Composition of Linear Transformations

Definition Let U, V, and W be vector spaces, and let $S : U \to V$ and $T : V \to W$ be linear transformations. The *composition* of S and T is denoted $TS : U \to W$ and is defined by

$$TS(\mathbf{v}) = T(S(\mathbf{v}))$$

Notice that this is well defined since the image of S is contained in V, which is the domain of T.

Proposition 2.5.1 Let $S : U \to V$ and $T : V \to W$ be linear transformations, then TS is a linear transformation.

Remark In general, ST is not equal to TS. We emphasize that the composition is well defined only if the image of the first transformation is contained in the domain of the second.

Proposition 2.5.4

1. Let $R : U \to V, S : V \to W$ and $T : W \to X$ be linear transformations of the vector space U,V,W and X as indicated. Then

$$T(SR) = (TS)R \text{ (associativity)}$$

2. Let $R : U \to V, S : V \to W$ and $T : W \to X$ be linear transformations of the vector space U,V,W and X as indicated. Then

$$T(R + S) = TR + TS \text{ (distributivity)}$$

3. Let $R : U \to V, S : V \to W$ and $T : W \to X$ be linear transformations of the vector space U,V,W and X as indicated. Then

$$(T + S)R = TR + SR \text{ (distributivity)}$$

Proposition 2.5.6 Let $S : U \to V$ and $T : V \to W$ be linear transformations. Then

1. $Ker(S) \subset Ker(TS)$

2. $Im(TS) \subset Im(T)$

proof:

1. If $\mathbf{u} \in Ker(S), S(\mathbf{u}) = \mathbf{0}$. Then $TS(\mathbf{u}) = T(\mathbf{0}) = \mathbf{0}$. Therefore $\mathbf{u} \in Ker(TS)$.
2. If $\mathbf{x} \in Im(TS)$, then $\exists \mathbf{u} \in U$ s.t. $TS(\mathbf{u}) = T(S(\mathbf{u})) = \mathbf{x}$, then $\exists \mathbf{v} = S(\mathbf{u}) \in V$ s.t. $T(\mathbf{v}) = \mathbf{x}$. Therefore $\mathbf{x} \in Im(T)$ ■

Corollary 2.5.7 Let $S : U \to V$ and $T : V \to W$ be linear transformations of finite-dimensional vector spaces. Then

1. $\dim(Ker(S)) \leq \dim(Ker(TS))$
2. $\dim(Im(TS)) \leq \dim(Im(T))$

Proposition 2.5.9 If $[S]_\alpha^\beta$ has entries a_{ij}, $i = 1, \ldots, n$ and $j = 1, \ldots, m$ and $[T]_\beta^\gamma$ has entries b_{kl}, $k = 1, \ldots, p$ and $l = 1, \ldots, n$, then the entries of $[TS]_\alpha^\gamma$ are $\sum_{l=1}^n b_{kl}a_{lj}$

Definition 2.5.10 Let A be an $n \times m$ matrix and B a $p \times n$ matrix, then the *matrix product* BA is defined to be the $p \times m$ matrix whose entries are $\sum_{l=1}^n b_{kl}a_{lj}$ for $k = 1, \ldots, p$ and $j = 1, \ldots, m$.

Proposition 2.5.13 Let $S : U \to V$ and $T : V \to W$ be linear transformations between finite-dimensional vector spaces. Let α, β and γ be bases for U,V and W, respectively. Then

$$[TS]_\alpha^\gamma = [T]_\beta^\gamma [S]_\alpha^\beta$$

In words, the matrix of the composition of two linear transformations is the product of the matrices of the transformations

Proposition 2.5.14

1. Let A,B and C be $m \times n, n \times p$ and $p \times r$ matrices, then
$$(AB)C = A(BC) \text{ (associativity)}$$
2. Let A,B and C be $m \times n, n \times p$ and $p \times r$ matrices, then
$$A(B + C) = AB + AC \text{ (distributivity)}$$
3. Let A,B and C be $m \times n, n \times p$ and $p \times r$ matrices, then
$$(A + B)C = AC + BC \text{ (distributivity)}$$

2.6 The Inverse of a Linear Transformation

Definition If $f : S_1 \to S_2$ is a function from one set to another, we say that g is the *inverse function of* f if for every $x \in S_1, g(f(x)) = x$ and for every $y \in S_2, f(g(y)) = y$. If such a g exists, f must be both injective and surjective(bijective).
To see this, notice that if $f(x_1) = f(x_2)$, then

$$x_1 = g(f(x_1)) = g(f(x_2)) = x_2$$

So that f is injective. If $y \in S_2$, then for $x = g(y), f(x) = f(g(y)) = y$ so that f is surjective.
Converse is true: bijective $\implies$ exists an inverse

Proposition 2.6.1 If $T : V \to W$ is injective and surjective, then the inverse function $S : W \to V$ is a linear transformation.

proof: Let $\mathbf{w}_1$ and $\mathbf{w}_2 \in W$ and a and $b \in \mathbb{R}$. By definition, $S(\mathbf{w}_1) = \mathbf{v}_1$ and $S(\mathbf{w}_2) = \mathbf{v}_2$ are the unique vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ satisfying $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$. By definition, $S(a\mathbf{w}_1 + b\mathbf{w}_2)$ is the unique vector $\mathbf{v}$ with $T(\mathbf{v}) = a\mathbf{w}_1 + b\mathbf{w}_2$ but $\mathbf{v} = a\mathbf{v}_1 + b\mathbf{v}_2$ satisfies $T(a\mathbf{v}_1 + b\mathbf{v}_2) = aT(\mathbf{v}_1) + bT(\mathbf{v}_2) = a\mathbf{w}_1 + b\mathbf{w}_2$. Thus, $S(a\mathbf{w}_1 + b\mathbf{w}_2) = a\mathbf{v}_1 + b\mathbf{v}_2 = aS(\mathbf{w}_1) + bS(\mathbf{w}_2)$ as we desired. ■

Proposition 2.6.2 A linear transformation $T : V \to W$ has an inverse linear transformation S if and only if T is injective and surjective.

Definition 2.6.3 If $T : V \to W$ is a linear transformation that has an inverse transformation $S : W \to V$, we say that T is invertible, and we denote the inverse of T by T^{-1}.

Definition 2.6.4 If $T : V \to W$ is an invertible transformation, T is called an isomorphism, and we say V and W are isomorphic vector spaces.

Notes $T^{-1}T(\mathbf{v})$ is the identity linear transformation of V, $T^{-1}T = I_V$, and TT^{-1} is the identity linear transformation of W, $TT^{-1} = I_W$. If S is a linear transformation that is a candidate for the inverse, we need only verify that $ST = I_V$ and $TS = I_W$.

Proposition 2.6.7 If V and W are finite-dimensional vector spaces, then there is an isomorphism $T : V \to W$ if and only if $\dim(V) = \dim(W)$.

Definition 2.6.10 An $n \times n$ matrix A is called *invertible* if there exists an $n \times n$ matrix B so that $AB = BA = I$. B is called the *inverse* of A and is denoted by A^{-1}.

Proposition 2.6.11 Let $T : V \to W$ be an isomorphism of finite-dimensional vector spaces. Then for any choice of bases α for V and β for W

$$[T^{-1}]_{\beta}^{\alpha} = [T]_{\alpha}^{\beta\,-1}$$

2.7 Change of Basis

Proposition 2.7.3 Let V be a finite-dimensional vector space, and let α and α' be bases for V. Let $\mathbf{v} \in V$. Then the coordinate vector $[\mathbf{v}]_{\alpha'}$ of $\mathbf{v}$ in the basis α' is related to the coordinate vector $[\mathbf{v}]_{\alpha}$ of $\mathbf{v}$ in the basis α by

$$[I]_{\alpha}^{\alpha'}[\mathbf{v}]_{\alpha} = [\mathbf{v}]_{\alpha'}$$

Theorem 2.7.5 Let $T : V \to W$ be a linear transformation between finite-dimensional vector spaces V and W. Let $I_V : V \to V$ and $I_W : W \to W$ be the respective identity transformations of V and W. Let α and α' be two bases for V, and let β and β' be two bases for W. Then

$$[T]_{\alpha'}^{\beta'} = [I_W]_{\beta}^{\beta'} \cdot [T]_{\alpha}^{\beta} \cdot [I_V]_{\alpha'}^{\alpha}$$

Definition 2.7.6 Let A,B be $n \times n$ matrices. A and B are said to be *similar* if there is an invertible $n \times n$ matrix Q such that

$$B = Q^{-1}AQ$$

3 The Determinant Function

3.1 The Determinant as Area

Corollary 3.1.2 Let $V = \mathbb{R}^2$. $T : V \to V$ is an isomorphism if and only if the area of the parallelogram constructed previously is nonzero.

Proposition 3.1.3 The function $Area(\mathbf{a}_1, \mathbf{a}_2)$ has the following properties for $\mathbf{a}_1, \mathbf{a}_2, , \mathbf{a}_1'$, and $\mathbf{a}_2' \in \mathbb{R}^2$

1. $Area(b\mathbf{a}_1 + c\mathbf{a}_1', \mathbf{a}_2) = bArea(\mathbf{a}_1, \mathbf{a}_2) + cArea(\mathbf{a}_1', \mathbf{a}_2)$ for $b, c \in \mathbb{R}$

2. $Area(\mathbf{a}_1, b\mathbf{a}_2 + c\mathbf{a}_2') = bArea(\mathbf{a}_1, \mathbf{a}_2) + cArea(\mathbf{a}_1, \mathbf{a}_2')$ for $b, c \in \mathbb{R}$
3. $Area(\mathbf{a}_1, \mathbf{a}_2) = -Area(\mathbf{a}_2, \mathbf{a}_1)$
4. $Area((1,0),(0,1)) = 1$

Proposition 3.1.4 If $B(\mathbf{a}_1, \mathbf{a}_2)$ is any real-valued function of $\mathbf{a}_1$ and $\mathbf{a}_2 \in \mathbb{R}^2$ that satisfies Properties (i),(ii),(iii) of Proposition (3.1.3), then B is equal to the area function.

Definition 3.1.5 The determinant of a 2×2 matrix A, denoted by $\det(A)$ or $\det(\mathbf{a}_1, \mathbf{a}_2)$, is the unique function of the rows of A satisfying

1. $\det(\mathbf{a}_1, b\mathbf{a}_2 + c\mathbf{a}_2') = b\det(\mathbf{a}_1, \mathbf{a}_2) + c\det(\mathbf{a}_1, \mathbf{a}_2')$ for $b, c \in \mathbb{R}$
2. $\det(\mathbf{a}_1, \mathbf{a}_2) = -\det(\mathbf{a}_2, \mathbf{a}_1)$
3. $\det(\mathbf{e}_1, \mathbf{e}_2) = 1$

As a consequence of (3.1.4), $\det(A)$ is given explicitly by

$$\det(A) = a_{11}a_{22} - a_{12}a_{21}$$

We can rephrase the work of this section as follows

Proposition 3.1.6

1. A 2×2 matrix A is invertible if and only if $\det(A) \neq 0$
2. If $T : V \to V$ is a linear transformation of a two-dimensional vector space V, then T is an isomorphism if and only if $\det([T]_\alpha^\alpha) \neq 0$

3.2 The Determinant of an $n \times n$ Matrix

Definition 3.2.1 A function f of the rows of a matrix A is called multilinear if f is a linear function of each of its rows when the remaining rows are held fixed. That is, f is multilinear if for all b and $b' \in \mathbb{R}$,

$$f(\mathbf{a}_1, \ldots, b\mathbf{a}_l + b'\mathbf{a}_l', \ldots, \mathbf{a}_n) = bf(\mathbf{a}_1, \ldots, \mathbf{a}_l, \ldots, \mathbf{a}_n) + b'f(\mathbf{a}_1, \ldots, \mathbf{a}_l', \ldots, \mathbf{a}_n)$$

Definition 3.2.2 A function f of the rows of a matrix A is said to be alternating if whenever any two rows of A are interchanged f changes sign. That is, for all$i \neq j, 1 \leq i, j \leq n$, we have

$$f(\mathbf{a}_1, \ldots, \mathbf{a}_l, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_n) = -f(\mathbf{a}_1, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_n)$$

Lemma 3.2.3 If f is an alternating real-valued function of the rows of an $n \times n$ matrix and two rows of the matrix A are identical, then $f(A) = 0$

Definition 3.2.4 Let A be an $n \times n$ matrix with entries $a_{ij}, i, j = 1, \ldots, n$. The ijth minor of A is defined to be the $(n-1) \times (n-1)$ matrix obtained by deleting the ith row and jth column of A. The ijth minor is denoted by A_{ij}.

Proposition 3.2.5 Let A be a 3×3 matrix, and let f be an alternating multilinear function. Then

$$f(A) = [a_{11} \det(A_{11}) - a_{12} \det(A_{12}) + a_{13} \det(A_{13})] f(I)$$

Corollary 3.2.6 There exists exactly one multilinear alternating function f of the rows of a 3×3 matrix such that $f(I) = 1$

Definition 3.2.7 The determinant function of a 3×3 matrix is the unique alternating multilinear function f with $f(I) = 1$. This function will be denoted by $\det(A)$.

Theorem 3.2.8 There exists exactly one alternating multilinear function $f : M_{n \times n}(\mathbb{R}) \to \mathbb{R}$ satisfying $f(I) = 1$, which is called the determinant function $f(A) = \det(A)$. Further, any alternating multilinear function f satisfies $f(A) = \det(A) f(I)$

Proposition 3.2.10 If an $n \times n$ matrix A is not invertible, then $\det(A) = 0$.

Proposition 3.2.11

$$\det(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \det(\mathbf{a}_1, \ldots, \mathbf{a}_i + b\mathbf{a}_j, \ldots, \mathbf{a}_n)$$

Lemma 3.2.12 If A is an $n \times n$ diagonal matrix, then $\det(A) = a_{11} a_{22} \ldots a_{nn}$

Proposition 3.2.13 If A is invertible, then $\det(A) \neq 0$

Theorem 3.2.14 Let A be an $n \times n$ matrix. A is invertible if and only if $\det(A) \neq 0$

3.3 Further Properties of the Determinant

Let A' be the matrix whose entries a'_{ij} are the scalars $(-1)^{i+j}\det(A_{ji})$. The quantity a'_{ij} is called the jith cofactor of A.

Proposition 3.3.1

$$AA' = det(A)I$$

Corollary 3.3.2 If A is an invertible $n \times n$ matrix, then A^{-1} is the matrix whose ijth entry is $(-1)^{i+j}\det(A_{ji})/\det(A)$

Proposition 3.3.4 For any fixed j, $1 \leq j \leq n$,

$$\det(A) = \sum_{i=1}^{n}(-1)^{i+j}a_{ij}\det(A_{ij})$$

Remark 3.3.5 In general, if $\mathbf{b}$ is a vector in $\mathbb{R}^n$, $A'\mathbf{b}$ is a vector whose ith entry is $\sum_{j=1}^{n} a'_{ij}b_j = \sum_{j=1}^{n} b_j(-1)^{i+j}\det(A_{ji})$. This is the determinant of the matrix whose columns are $\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{b}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_n$, where $\mathbf{a}_j, 1 \leq j \leq n$, is the jth column of A. The determinant is expanded along the ith column. This fact will be used in the discussion of Cramer's rule, which appears later in this section.

Proposition 3.3.7 If A and B are $n \times n$ matrices, then

1. $\det(AB) = \det(A)\det(B)$
2. If A is invertible, then $\det(A^{-1}) = 1/\det(A)$

Corollary 3.3.8 If $T : V \to V$ is a linear transformation, $\dim(V) = n$, then

$$\det([T]_{\alpha}^{\alpha}) = \det([T]_{\beta}^{\beta})$$

for all choices of bases α and β for V.

Definition 3.3.9 The determinant of a linear transformation $T : V \to V$ of a finite-dimensional vector space is the determinant of $[T]_{\alpha}^{\alpha}$ for any choice of α. We denote this by $\det(T)$.

Proposition 3.3.11 A linear transformation $T : V \to V$ of a finite-dimensional vector space is an isomorphism if and only if $\det(T) \neq 0$

Proposition 3.3.12 Let $S : V \to V$ and $T : V \to V$ be linear transformations of a finite-dimensional vector space, then

1. $\det(ST) = \det(S)\det(T)$ and
2. if T is an isomorphism, $\det(T^{-1}) = \det(T)^{-1}$

Proposition 3.3.13 (Cramer's rule) Let A be an invertible $n \times n$ matrix. The solution $\mathbf{x}$ to the system of equations $A\mathbf{x} = \mathbf{b}$ is the vector whose jth entry is the quotient

$$\det(B_j)/\det(A)$$

where B_j is the matrix obtained from A by replacing the jth column of A by the vector $\mathbf{b}$.

4 Eigenvalues, Eigenvectors, Diagonalization, and the Spectral Theorem in $\mathbb{R}^n$

4.1 Eigenvalues and Eigenvectors

Definition 4.1.2 Let $T : V \to V$ be a linear mapping

1. A vector $\mathbf{x} \in V$ is called an eigenvector of T if $\mathbf{x} \neq \mathbf{0}$ and there exists a scalar $\lambda \in \mathbb{R}$ such that $T(\mathbf{x}) = \lambda\mathbf{x}$
2. If $\mathbf{x}$ is an eigenvector of T and $T(\mathbf{x}) = \lambda\mathbf{x}$, the scalar λ is called the *eigenvalue* of T corresponding to $\mathbf{x}$.

Proposition 4.1.15 A vector $\mathbf{x}$ is an eigenvector of T with eigenvalue λ if and only if $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{x} \in Ker(T - \lambda I)$.

Definition 4.1.16 Let $T : V \to V$ be a linear mapping, and let $\lambda \in \mathbb{R}$. The λ-eigenspace of T, denoted E_λ, is the set

$$E_\lambda = \{\mathbf{x} \in V | T(\mathbf{x}) = \lambda\mathbf{x}\}$$

That is, E_λ is the set containing all the eigenvectors of T with eigenvalue λ, together with the vector $\mathbf{0}$. If λ is not an eigenvalue of T, then we have $E_\lambda = \{\mathbf{0}\}$.

Proposition 4.1.7 E_λ is a subspace of V for all λ.

Proposition 4.1.9 Let $A \in M_{n\times n}(\mathbb{R})$. Then $\lambda \in \mathbb{R}$ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$.

Definition 4.1.11 Let $A \in M_{n\times n}(\mathbb{R})$. The polynomial $\det(A - \lambda I)$ is called the *characteristic polynomial* of A.

Remark The characteristic polynomial should only depends on the linear mapping defined by the matrix A and not on the matrix itself. (if change to another basis, the characteristic polynomial should be the same.)

Proposition 4.1.12 Similar matrices have equal characteristic polynomials. *proof:* Suppose A and B are two similar matrices, so that $B = Q^{-1}AQ$ for some invertible matrix Q. Then we have

$$\begin{aligned}
\det(B - \lambda I) &= \det(Q^{-1}AQ - \lambda I) \\
&= \det(Q^{-1}AQ - Q^{-1}\lambda I Q) \\
&= \det(Q^{-1}(A - \lambda I)Q) \\
&= \det(Q^{-1})\det(A - \lambda I)\det(Q) \\
&= \frac{1}{\det(Q)}\det(A - \lambda I)\det(Q) \\
&= \det(A - \lambda I)
\end{aligned}$$

■

Examples 4.1.13

1. For a general 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ we have

$$\det(A - \lambda I) = \lambda^2 - Tr(A)\lambda + \det(A)$$

2. If we substitute A into its own characteristic polynomial, we get $p(A) = 0$. We find that A satisfies its own polynomial equation.

3. For a general 3×3 matrix $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ we have

$$\det(A-\lambda I) = -\lambda^3 + Tr(A)\lambda^2 - ((ae-bd)+(ai-cg)+(ei-fh))\lambda + \det(A)$$

4. For any $n \times n$ matrix A, the characteristic polynomial has the form

$$(-1)^n\lambda^n + (-1)^{n-1}Tr(A)\lambda^{n-1} + c_{n-1}\lambda^{n-2} + \ldots + c_1\lambda + \det(A)$$

where the c_i are other polynomial expressions in the entries of the matrix A.

Corollary 4.1.14 Let $A \in M_{n\times n}(\mathbb{R})$. Then A has no more than n distinct eigenvalues. In addition, if $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of A and λ_i is an m_i-fold root of the characteristic polynomial, then $m_1+\ldots+m_k \leq n$

Theorem 4.1.18 Let $A \in M_{n\times n}(\mathbb{R})$, and let $p(t) = \det(A - tI)$ be its characteristic polynomial. Then $p(A) = 0$ (the $n \times n$ zero matrix).

4.2 Diagonalizability

Definition 4.2.1 Let V be a finite-dimensional vector space, and let $T : V \to V$ be a linear mapping. T is said to be diagonalizable if there exists a basis of V, all of whose vectors are eigenvectors of T.

Proposition 4.2.2 $T : V \to V$ is diagonalizable if and only if, for any basis α of V, the matrix $[T]_\alpha^\alpha$ is similar to a diagonal matrix.

Proposition 4.2.4 Let $\mathbf{x}_i (1 \leq i \leq k)$ be eigenvectors of a linear mapping $T : V \to V$ corresponding to distinct eigenvalues λ_i. Then $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ is a linearly independent subset of V.

Corollary 4.2.5 For each $i(1 \leq i \leq k)$, let $\{\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{l,n_l}\}$ be a linearly independent set of eigenvectors of T all with eigenvalue λ_i and suppose the λ_i are distinct. Then $S = \{\mathbf{x}_{1,1}, \ldots, \mathbf{x}_{1,n_1}\} \cup \ldots \cup \{\mathbf{x}_{k,1}, \ldots, \mathbf{x}_{k,n_k}\}$ is linearly independent.

Proposition 4.2.6 Let V be finite-dimensional, and let $T : V \to V$ be linear. Let λ be an eigenvalue of T, and assume that λ is an m-fold root of the characteristic polynomial of T. Then we have

$$1 \leq \dim(E_\lambda) \leq m$$

Theorem 4.2.7 Let $T : V \to V$ be a linear mapping on a finite-dimensional vector space V, and let $\lambda_1, \ldots, \lambda_k$ be its distinct eigenvalues. Let m_l be the multiplicity of λ_i as a root of the characteristic polynomial of T. Then T is diagonalizable if and only if

1. $m_1 + \ldots + m_k = n = \dim(V)$, and
2. for each $i, \dim(E_{\lambda_i}) = m_i$

Corollary 4.2.8 Let $T : V \to V$ be a linear mapping on a finite-dimensional space V, and assume that T has $n = \dim(V)$ distinct real eigenvalues. Then T is diagonalizable.

Corollary 4.2.9 A linear mapping $T : V \to V$ on a finite-dimensional space V is diagonalizable if and only if the sum of the multiplicities of the real eigenvalues is $n = \dim(V)$, and either

1. We have $\sum_{i=1}^{k} \dim(E_{\lambda_i}) = n$, where the λ_i are the distinct eigenvalues of T, or
2. We have $\sum_{i=1}^{k}(n - \dim(Im(T - \lambda_i I))) = n$, where again λ_i are the distinct eigenvalues.

Remark In order for a linear mapping or a matrix to be diagonalizable, it must have enough linearly independent eigenvectors to form a basis of V.

4.3 Geometry in $\mathbb{R}^n$

Example

$$f \cdot g = \int_a^b f(x)g(x)\, dx$$

defines an inner product on $[a, b]$.

Definition 4.3.5 The angle, θ, between two nonzero vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is defined to be

$$\theta = \cos^{-1}\left(\frac{<\mathbf{x}, \mathbf{y}>}{||x|| \cdot ||y||}\right)$$

Definition of Orthogonal and Orthonormal Sets

1. S is an <u>orthogonal set</u> if $\forall \mathbf{x}, \mathbf{y} \in S, \mathbf{x} \neq \mathbf{y} \implies \mathbf{x} \cdot \mathbf{y} = 0$.
2. S is an <u>orthonormal set</u> if it is orthogonal and all elements are unit vectors.

Proposition 4.3.10 If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are orthogonal, nonzero vectors, then $\{\mathbf{x}, \mathbf{y}\}$ is linearly independent.

Theorem Orthogonal sets of nonzero vectors are independent.
<u>*proof:*</u>

$$S = \{x_1, \ldots, x_n\}$$

Suppose $c_1 x_1 + c_2 x_2 + \ldots + c_n x_n = \mathbf{0}$

$$\begin{aligned} 0 &= x_i \cdot \mathbf{0} \\ &= x_i \cdot \sum_{j=1}^{n} c_j x_j \\ &= \sum_{j=1}^{n} c_j (x_i \cdot x_j) \\ &= \sum_{j=1}^{n} c_j (0 \text{ if } i \neq j) \\ &= c_i ||x_i||^2 \end{aligned}$$

Since x_i nonzero, then $c_i = 0 \forall i$ ■

Definition of Bilinearity A mapping $B : V \times V \to \mathbb{R}$ is said to be <u>bilinear</u> if B is linear in each variable, or more precisely if

1. $B(c\mathbf{x} + \mathbf{y}, \mathbf{z}) = cB(\mathbf{x}, \mathbf{z}) + B(\mathbf{y}, \mathbf{z})$ and
2. $B(\mathbf{x}, c\mathbf{y} + \mathbf{z}) = cB(\mathbf{x}, \mathbf{y}) + B(\mathbf{x}, \mathbf{z})$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and all $c \in \mathbb{R}$

4.4 Orthogonal Projections and the Gram-Schmidt Process

Definition 4.4.1 The *orthogonal complement* of W, denoted $W^\perp$, is the set $W^\perp = \{\mathbf{v} \in \mathbb{R}^n | <\mathbf{v}, \mathbf{w}>= 0 \text{ for all } \mathbf{w} \in W\}$

Remark If we choose a basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ for W, then $\mathbf{v} \in W^\perp$ iff $\mathbf{v}$ is orthogonal to every vector in the basis.

Examples

1. $W = \{\mathbf{0}\}$, then $W^\perp = \mathbb{R}^n$
2. $u_1, u_2 \in \mathbb{R}^3$
   $Span\{u_1, u_2\}^\perp = \{x | x \cdot u_1 = 0\} \cap \{x | x \cdot u_2 = 0\} = Ker \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$

Proposition 4.4.3

1. For every subspace W of $\mathbb{R}^n$, $W^\perp$ is also a subspace of $\mathbb{R}^n$
2. We have $\dim(W) + \dim(W^\perp) = \dim(\mathbb{R}^n) = n$
3. For all subspaces W of $\mathbb{R}^n, W \cap W^\perp = \{\mathbf{0}\}$
4. Given a subspace W of $\mathbb{R}^n$, every vector $\mathbf{x} \in \mathbb{R}^n$ can be written uniquely as $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$, where $\mathbf{x}_1 \in W$ and $\mathbf{x}_2 \in W^\perp$. In other words, $\mathbb{R}^n = W \bigoplus W^\perp$

Definition of Orthogonal Projection Every vector $\mathbf{x} \in \mathbb{R}^n$ can be written uniquely as $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$, where $\mathbf{x}_1 \in W$ and $\mathbf{x}_2 \in W^\perp$. Define $P_W : \mathbb{R}^n \to \mathbb{R}^n$ by $P_W(\mathbf{x}) = \mathbf{x}_1$.

Proposition 4.4.5

1. P_W is a linear mapping
2. $Im(P_W) = W$, and if $\mathbf{w} \in W$, then $P_W(\mathbf{w}) = \mathbf{w}$ (Identity transformation)
3. $Ker(P_W) = W^\perp$

Proposition 4.4.6 Let $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ be an orthonormal basis for the subspace $W \subset \mathbb{R}^n$

1. For each $\mathbf{w} \in W$, we have

$$\mathbf{w} = \sum_{i=1}^{k} < \mathbf{w}, \mathbf{w}_i > \mathbf{w}$$

2. For all $\mathbf{x} \in \mathbb{R}^n$, we have

$$P_W(\mathbf{x}) = \sum_{i=1}^{k} < \mathbf{x}, \mathbf{w}_i > \mathbf{w}_i$$

Remarks The real meaning of the statement is that we can use the inner product to compute the scalars needed to express the relevant vector in W as a linear combination of the basis vectors $\mathbf{w}_i$
proof: see textbook p194

Gram-Schmidt Orthogonalization Process Suppose we are given vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ that are linearly independent but not necessarily orthogonal, and we want to construct an orthogonal set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ with the property that $Span(\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}) = Span(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\})$.

$$\begin{aligned}
&\mathbf{v}_1 = \mathbf{u}_1 \\
&W_1 = Span\{\mathbf{u}_1\} \\
&\mathbf{v}_2 = \mathbf{u}_2 - P_{W_1}(\mathbf{u}_2) \\
&W_2 = Span\{\mathbf{v}_1, \mathbf{v}_2\} \\
&\mathbf{v}_3 = \mathbf{u}_3 - P_{W_2}(\mathbf{u}_3) \\
&\cdots \\
&W_k = Span\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}\} \\
&\mathbf{v}_k = \mathbf{u}_k - P_{W_k}(\mathbf{u}_k)
\end{aligned}$$

By proposition 4.4.6, we see that

$$P_{W_j}(\mathbf{v}) = \sum_{i=1}^{j} \frac{< \mathbf{v}_i, \mathbf{v} >}{< \mathbf{v}_i, \mathbf{v}_i >} \mathbf{v}_i$$

Therefore,

$$\begin{aligned}
&\mathbf{v}_1 = \mathbf{u}_1 \\
&\mathbf{v}_2 = \mathbf{u}_2 - \frac{< \mathbf{v}_1, \mathbf{u}_2 >}{< \mathbf{v}_1, \mathbf{v}_1 >} \mathbf{v}_2 \\
&\vdots \\
&\mathbf{v}_{j+1} = \mathbf{u}_{j+1} - \sum_{i=1}^{j} \frac{< \mathbf{v}_i, \mathbf{u}_{j+1} >}{< \mathbf{v}_i, \mathbf{v}_i >} \mathbf{v}_i
\end{aligned}$$

Remark The real meaning of the statement is that we can use the inner product to compute the scalars needed to express $\mathbf{w}$ and $P_W(\mathbf{x})$ as a linear combination of the basis vectors $\mathbf{w}_i$.

Theorem 4.4.9 Let W be a subspace of $\mathbb{R}^n$. Then there exists an orthonormal basis of W.

4.5 Symmetric Matrices

Definition 4.5.1 A square matrix A is said to be *symmetric* if $A = A^T$, where A^T denotes the transpose of A.

Proposition 4.5.2a Let $A \in M_{n\times n}(\mathbb{R})$.

1. For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n, < A\mathbf{x}, \mathbf{y} >=< \mathbf{x}, A^T\mathbf{y} >$
2. A is symmetric if and only if $< A\mathbf{x}, \mathbf{y} >=< \mathbf{x}, A\mathbf{y} >$ for all vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

proof:

1. $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n, < A\mathbf{x}, \mathbf{y} >= (A\mathbf{x})^T\mathbf{y} = \mathbf{x}^T A^T \mathbf{y} =< \mathbf{x}, A^T\mathbf{y} >$
2. Obvious

■

Corollary 4.5.2b Let V be any subspace of $\mathbb{R}^n$, let $T : V \to V$ be any linear mapping, and let $\alpha = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ be any orthonormal basis of V. Then $[T]_\alpha^\alpha$ is a symmetric matrix if and only if $< T(\mathbf{x}), \mathbf{y} >=< \mathbf{x}, T(\mathbf{y}) >$ for all vectors $\mathbf{x}, \mathbf{y} \in V$.

Definition 4.5.3 Let V be a subspace of $\mathbb{R}^n$. A linear mapping $T : V \to V$ is said to be symmetric if $< T(\mathbf{x}), \mathbf{y} >=< \mathbf{x}, T(\mathbf{y}) >$ for all vectors $\mathbf{x}, \mathbf{y} \in V$.

Example An important class of symmetric mappings is orthogonal projections see textbook p202

Remark Write out the matrix of orthogonal projection transformation with an orthonormal basis, we see a direct proof that orthogonal projections are diagonalizable.

Fact 4.5.6 For any symmetric matrix:

1. All the roots of the characteristic polynomial are real.
2. eigenvectors corresponding to distinct eigenvalues are orthogonal.

proof:
WLOG assume $\lambda_1 \neq 0$

$$\begin{aligned}< v_1, v_2 > &= < \frac{T(v_1)}{\lambda_1}, v_2 > \\ &= \frac{1}{\lambda_1} < T(v_1), v_2 > \\ &= \frac{1}{\lambda_1} < v_1, T(v_2) > \\ &= \frac{1}{\lambda_1} < v_1, \lambda_2 v_2 > \\ &= \frac{\lambda_2}{\lambda_1} < v_1, v_2 >\end{aligned}$$

Since $\lambda_1 \neq \lambda_2$, we have $< v_1, v_2 >= 0$ ■

Theorem 4.5.7 Let $A \in M_{n\times n}(\mathbb{R})$ be a symmetric matrix, let $\mathbf{x}_1$ be an eigenvector of A with eigenvalue λ_1, and let $\mathbf{x}_2$ be an eigenvector of A with eigenvalue λ_2, where $\lambda_1 \neq \lambda_2$. Then $\mathbf{x}_1$ and $\mathbf{x}_2$ are orthogonal vectors in $\mathbb{R}^n$.

4.6 The Spectral Theorem

Theorem 4.6.1 - The Spectral Theorem Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a symmetric linear mapping. Then there is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of T. In particular, T is diagonalizable.
proof: By induction

Base Case
If $n = 1$, then every linear mapping is symmetric and diagonalizable.

Inductive Step
Assume the theorem is true for mappings from $\mathbb{R}^k$ to $\mathbb{R}^k$ and consider $T : \mathbb{R}^{k+1} \to \mathbb{R}^{k+1}$.
Let λ be any one of the eigenvalues, and let $\mathbf{x}_1$ be any unit eigenvector with eigenvalue λ.
Let $W = Span(\{\mathbf{x}_1\})$. Note that $W^\perp$ is a k-dimensional subspace of $\mathbb{R}^{k+1}$, so $W^\perp$ is isomorphic to $\mathbb{R}^k$, and we can apply I.H. to $T|_{W^\perp}$

1. To see that T takes vectors in $W^\perp$ to vectors in $W^\perp$, note that if

$\mathbf{y} \in W^{\perp}$, then

$$\begin{aligned} < \mathbf{x}_1, T(\mathbf{y}) > &= < T(\mathbf{x}_1), \mathbf{y} > \\ &= < \lambda \mathbf{x}_1, \mathbf{y} > \\ &= 0 && \text{(Since } \mathbf{y} \in W^{\perp}) \end{aligned}$$

Hence $T(\mathbf{y}) \in W^{\perp}$

2. To see that the restriction of T to $W^{\perp}$ is still symmetric, note that if $\mathbf{y}_1, \mathbf{y}_2 \in W^{\perp}$, then $< T(\mathbf{y}_1), \mathbf{y}_2 >=< \mathbf{y}_1, T(\mathbf{y}_2) >$, since this holds more generally for all vectors in $\mathbb{R}^{k+1}$.

Hence by I.H. applied to $T|_{W^{\perp}}$, there exists an orthonormal basis $\{\mathbf{x}_2, \ldots, \mathbf{x}_{k+1}\}$ of $W^{\perp}$, consisting of eigenvectors of the restricted mapping. Union with $\mathbf{x}_1$, we have the conclusion. ■

Theorem 4.6.3 Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a symmetric linear mapping, and let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T. let P_i be the orthogonal projection of $\mathbb{R}^n$ onto the eigenspace E_{λ_i}. Then

1. $T = \lambda_1 P_1 + \ldots + \lambda_k P_k$, and
2. $I = P_1 + \ldots + P_k$

Remark Spectral Decomposition. This says that $\mathbf{x}$ can be recovered or built up from its projections on the various eigenspaces of T.

5 Complex Numbers and Complex Vector Spaces

5.1 Complex Numbers

Definition 5.1.1 The set of *complex numbers*, denoted $\mathbb{C}$, is the set of ordered pairs of real numbers (a, b) with the operations of addition and multiplication defined by

$$(a, b) + (c, d) = (a + c, b + d)$$

and the *product* of (a, b) and (c, d) is the complex number defined by

$$(a, b)(c, d) = (ac - bd, ad + cb)$$

Definition 5.1.2 Let $z = a + bi \in \mathbb{C}$, The *real part* of z, denoted $\mathrm{Re}(z)$, is the real number a. The *imaginary part* of z, denoted $\mathrm{Im}(z)$, is the real number b. z is called a *real number* if $\mathrm{Im}(z) = 0$, and purely imaginary if $\mathrm{Re}(z) = 0$.

Definition 5.1.4 A field is a set F with two operations, defined on ordered pairs of elements of F, called *addition* and *multiplication*. Addition assigns to the pair x and $y \in F$ there *sum*, which is denoted by $x + y$ and multiplication assigns to the pair x and $y \in F$ their *product*, which is denoted by $x \cdot y$ or xy. These two operations must satisfy the following properties for allx, y and $z \in F$:

1. Commutativity of addition: $x + y = y + x$
2. Associativity of addition: $(x + y) + z = x + (y + z)$
3. Existence of an additive identity: There is an element $0 \in F$, called zero, such that $x + 0 = x$
4. Existence of additive inverses: For each x there is an element $-x \in F$ such that $x + (-x) = 0$
5. Commutativity of multiplication: $xy = yx$
6. Associativity of multiplication: $(xy)z = x(yz)$
7. Existence of a multiplicative identity: There is an element $1 \in F$, called 1, such that $x \cdot 1 = x$
8. Existence of multiplicative inverses: If $x \neq 0$, then there is an element $x^{-1} \in F$ such that $xx^{-1} = 1$

Examples

1. $F = \mathbb{C}$
2. $F = \mathbb{R}$
3. $F = \mathbb{Q}$
4. $F = \mathbb{Z}/p\mathbb{Z}$, p prime
5. Algebraic numbers $= \{x|p(x) = 0$, for a polynomial p with integer coefficients$\}$

Counter-Example $P_n(\mathbb{R})$

Proposition 5.1.5 The set of complex numbers is a field with the operations of addition and scalar multiplication as defines previously.

More definitions about complex numbers $z = a + bi$
complex conjugate: $\bar{z} = a - bi$
$z^{-1} = \frac{\bar{z}}{z\bar{z}}$ since $z\bar{z} = a^2 + b^2$

Proposition 5.1.7

1. The additive identity in a field is unique
2. The additive inverse of an element of a field is unique
3. The multiplicative identity of a field is unique
4. The multiplicative inverse of a nonzero element of a field is unique

Definition 5.1.8 The absolute value of the complex number $z = a + bi$ is the nonnegative real number $\sqrt{a^2 + b^2}$ and is denoted by $|z|$ or $r = |z|$. The *argument* of the complex number z is the angle θ of the polar coordinate representation of z. Can write $z = |z|(\cos(\theta) + i\sin(\theta))$

Remark In general, if n is an integer,

$$z^n = r^n(\cos(n\theta) + i\sin(n\theta))$$

Definition 5.1.11 (de Moivre's formula) A field F is called algebraically closed if every polynomial $p(z) = a_n z^n + \ldots + a_1 z + a_0$ with coefficients in F, $a_i \in F$ for $i = 0, \ldots, n$, has n roots in F.

Theorem 5.1.12 $\mathbb{C}$ is algebraically closed and $\mathbb{C}$ is the smallest algebraically closed field containing $\mathbb{R}$

5.2 Vector Spaces Over a Field

Definition 5.2.1 A vector space over a field F is a set V (whose elements are called vectors) together with addition and multiplication and 8 axioms as in chapter 1.

Example $F^n = \{\mathbf{x} = (x_1, \ldots, x_n) | x_i \in F, \text{ for } i = 1, \ldots, n\}$

5.3 Geometry in a complex vector space

Definition 5.3.1 Let V be a complex vector space, A Hermitian inner product on V is a complex valued function on pairs of vectors in V, denoted by $< \mathbf{u}, \mathbf{v} > \in \mathbb{C}$ for $\mathbf{u}, \mathbf{v} \in V$, which satisfies the following properties:

1. For all $\mathbf{u}, \mathbf{v}$, and $\mathbf{w} \in V$ and $a, b \in \mathbb{C}$, $< a\mathbf{u} + b\mathbf{v}, \mathbf{w} > = a < \mathbf{u}, \mathbf{w} >$ $+ b < \mathbf{v}, \mathbf{w} >$
2. For all $\mathbf{u}, \mathbf{v}, \in V, < \mathbf{u}, \mathbf{v} > = \overline{< \mathbf{v}, \mathbf{u} >}$, and
3. For all $\mathbf{v} \in V, < \mathbf{v}, \mathbf{v} > \geq 0$ and $< \mathbf{v}, \mathbf{v} > = 0$ implies $\mathbf{v} = \mathbf{0}$

Example 5.3.2 Hermitian inner product on $\mathbb{C}^n$
For $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{C}^n$, we define their inner product by $< \mathbf{x}, \mathbf{y} > = x_1 \bar{y_1} + \ldots + x_n \bar{y_n}$, which satisfies the Hermitian inner product properties.

Definition 5.3.7 Let V be a finite dimensional Hermitian inner product space and let α be an orthonormal basis for V. The adjoint of the linear transformation $T : V \to V$ is the linear transformation T^* whose matrix with respect to the orthonormal basis α is the matrix $([\bar{T}]_\alpha^\alpha)^t$; that is, $[T^*]_\alpha^\alpha = ([\bar{T}]_\alpha^\alpha)^t$

Proposition 5.3.8 Let V be a finite dimensional Hermitian inner product space. The adjoint of $T : V \to V$ satisfies $< T(\mathbf{v}, \mathbf{w}) > = < \mathbf{v}, T^*(\mathbf{w}) >$ for all $\mathbf{v}$ and $\mathbf{w} \in V$.

Definition 5.3.9 $T : V \to V$ is called Hermitian or self-adjoint if $< T(\mathbf{u}), \mathbf{v} > = < \mathbf{u}, T(\mathbf{v}) >$ for all $\mathbf{u}$ and $\mathbf{v} \in V$. Equivalently, T is Hermitian or self-adjoint if $T = T^*$ or $[\bar{T}]_\alpha^{\alpha t} = [T]_\alpha^\alpha$ for an orthonormal basis α. An $n \times n$ complex matrix is called Hermitian or self-adjoint if $A = A^*$.

Proposition 5.3.10 If λ is an eigenvalue of the self-adjoint linear transformation T, then $\lambda \in \mathbb{R}$

Proposition 5.3.11 If $\mathbf{u}$ and $\mathbf{v}$ are eigenvectors, respectively, for distinct eigenvalues λ and μ of a self adjoint transformation $T : V \to V$, then $< \mathbf{u}, \mathbf{v} > = 0$, so $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

Theorem 5.3.12 Let $T : V \to V$ be a self-adjoint transformation of a complex vector space V with Hermitian inner product $<,>$. Then there is an orthonormal basis of V consisting of eigenvectors for T and, in particular, T is diagonalizable.

Theorem 5.3.13 Let $T : V \to V$ be a self-adjoint transformation of a complex vector space V with Hermitian inner product $<,>$. Let $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ be the distinct eigenvalues for T, and Let P_i be the orthogonal projections of V onto the eigenspaces E_{λ_i}, then

1. $T = \lambda_1 P_1 + \ldots + \lambda_k P_k$
2. $I = P_1 + \ldots + P_k$

6 Jordan Canonical Form

A next best form after a diagonal form for the matrices of linear mappings that are not necessarily diagonalizable

6.1 Triangular Form

Definition 6.1.2 Let $T : V \to V$ be a linear mapping. A subspace W $\subset V$ is said to be *invariant* (or *stable*) under T if $T(W) \subset W$.

Proposition 6.1.4 Let V be a vector space, let $T : V \to V$ be a linear mapping, and let $\beta = \{x_1, \ldots, x_n\}$ be a basis for V. Then $[T]_\beta^\beta$ is upper triangular if and only if each of the subspaces $W_i = Span(\{x_1, \ldots, x_i\})$ is invariant under T.
Note that the subspaces W_i in the proposition are related as follows:

$$\{\mathbf{0}\} \subset W_1 \subset W_2 \subset \ldots \subset W_{n-1} \subset W_n = V$$

The W_i form an *increasing sequence* of subspaces.

Definition 6.1.5 We say that a linear mapping $T : V \to V$ on a finite-dimensional vector space V is triangularizable if there exists a basis β such that $[T]_\beta^\beta$ is upper-triangular.

Proposition 6.1.6 Let $T : V \to V$, and let $W \subset V$ be an invariant subspace. Then the characteristic polynomial of $T|_W$ divides the characteristic polynomial of T.

Remark Every eigenvalue of $T|_W$ is also an eigenvalue of T (the set of eigenvalues of $T|_W$ is some subset of the eigenvalues of T on the whole space).

Theorem 6.1.8 Let V be a finite-dimensional vector space over a field F, and let $T : V \to V$ be a linear mapping. Then T is triangularizable if and only if the characteristic polynomial equation of $p(t)$ has $dim(V)$ roots (counted with multiplicities) in the field F.

Remark The theorem implies that every matrix $A \in M_{n\times n}(\mathbb{C})$ may be triangularized.

Proof of Lemma Let $\alpha = \{x_1, \dots, x_k\}$ be a basis for W and extend α by adjoining $\alpha' = \{x_{k+1}, \dots, x_n\}$ to form a basis $\beta = \{x_1, \dots, x_k, x_{k+1}, \dots, x_n\}$ for V. Let $W' = Span(W')$.
Define $P : V \to V$ by

$$P(a_1x_1 + \ldots + a_nx_n) = a_1x_1 + \ldots + a_kx_k$$

Notice that $Ker(P) = W', Im(P) = W, P^2 = P$. P is called the projection on W with kernel W'. Then $I - P$ is the projection on W' with kernel W. Then $I - P$ is the projection on W' with kernel W.
Let $S = (I - P)T$. Since $Im(I - P) = W'$, we see by prop2.5.6 that $Im(S) \subset Im(I - P) = W'$. Hence W' is an invariant subspace of S. Then the eigenvalues of $S|_{W'}$ is a subset of the set of eigenvalues of T. Since all the eigenvalues of T lie in the filed F, the same is true of all the eigenvalues of $S|_{W'}$. Hence there is some nonzero vector $\mathbf{x} \in W'$ and some $\lambda \in F$ such that $S(x) = \lambda\mathbf{x}$. So

$$(I - P)T(\mathbf{x}) = \lambda\mathbf{x}$$
$$\implies T(\mathbf{x}) - PT(\mathbf{x}) = \lambda\mathbf{x}$$
$$\implies T(\mathbf{x}) = \lambda\mathbf{x} + PT(\mathbf{x})$$

where $\lambda\mathbf{x} \in Span(\{\mathbf{x}\})$ and $PT(\mathbf{x}) \in W$. Therefore $W + Span(\{\mathbf{x}\})$ is also invariant under T and this finishes the proof. ■

Proof of Theorem 6.1.8 $\to$: If T is triangularizable, then there exists a basis β for V such that $[T]_\beta^\beta$ is upper-triangular. The eigenvalues of T are the diagonal entries of this matrix, so they are elements of the field F.
$\leftarrow$: If all the eigenvalues are in F:

Let λ be any eigenvalue of T, and let x_1 be an eigenvector of λ, let $W_1 = Span(\{x_1\})$. By definition W_1 is invariant under T. Now, assume by induction that we have constructed invariant subspaces $W_1 \subset W_2 \subset \ldots \subset W_k$ with $W_i = Span(\{x_1, \ldots, x_l\})$ for each i. By Lemma 6.1.10 there exists a vector $x_{k+1} \notin W_k$ such that the subspace $W_{k+1} = W_k + Span(\{x_{k+1}\})$ is also invariant under T. We continue this process until we have produced a basis for V. Hence, T is triangularizable. ■

Lemma 6.1.10 Let $T : V \to V$ be as in the theorem, and assume that the characteristic polynomial of T has $n = \dim(V)$ roots in F. If $W \subsetneqq V$ is an invariant subspace under T, then there exists a vector $\mathbf{x} \neq \mathbf{0}$ in V such that $\mathbf{x} \notin W$ and $W + Span(\{\mathbf{x}\})$ is also invariant under T.

Remark What this says is that we can make a T-invariant subspace 1-dimension bigger.

Corollary 6.1.11 If $T : V \to V$ is triangularizable, with eigenvalues λ_i with respective multiplicities m_i, then there exists a basis β for V such that $[T]_\beta^\beta$ is upper-triangular, and the diagonal entries of $[T]_\beta^\beta$ are $m_1 \lambda_1$'s, followed by $m_2 \lambda_2$'s, and so on.

Theorem 6.1.12 (Cayley-Hamilton) If $T : V \to V$ be a linear mapping on a finite-dimensional vector space V, and let $p(t) = \det(T - tI)$ be its characteristic polynomial. Assume that $p(t)$ has $\dim(V)$ roots in the field F over which V is defined. Then $p(T) = 0$

6.2 A Canonical Form For Nilpotent Mappings

Definition A linear mapping $N : V \to V$ is <u>nilpotent</u> if $N^k = 0$ for some integer $k \geq 1$.

Proposition $N : V \to V$ is nilpotent if and only if it has one eigenvalue $\lambda = 0$ with multiplicity $n = \dim(V)$.

Proposition 6.2.3 With all notations as before:

1. $N^{k-1}(\mathbf{x})$ is an eigenvector of N with eigenvalue $\lambda = 0$
2. $C(\mathbf{x})$ is an invariant subspace of V under N.

3. The cycle generated by $\mathbf{x} \neq 0$ is a linearly independent set. Hence $\dim(C(\mathbf{x})) = k$, the length of the cycle.

Proposition 6.2.4 Let $\alpha_1 = \{N^{k_i-1}(\mathbf{x}_i), \ldots, \mathbf{x}_i\}(1 \leq i \leq r)$ be cycles of lengths k_i, respectively. If the set of eigenvectors $\{N^{k_i-1}(\mathbf{x}_1), \ldots, N^{k_r-1}(\mathbf{x}_r)\}$ is linearly independent, then $\alpha_1 \cup \ldots \cup \alpha_r$ is linearly independent.

Remark For a given $\mathbf{x} \in V$, either $\mathbf{x} = \mathbf{0}$ or there is a unique integer k, $1 \leq k \leq n$, such that $N^k(\mathbf{x}) = \mathbf{0}$ but $N^{k-1}(\mathbf{x}) \neq \mathbf{0}$.

Definitions 6.2.1 Let $N, \mathbf{x} \neq \mathbf{0}$ and k be as before

1. The set $\{N^{k-1}(\mathbf{x}), N^{k-2}(\mathbf{x}), \ldots, \mathbf{x}\}$ is called the cycle generated by $\mathbf{x}$. $\mathbf{x}$ is called the initial vector of the cycle.
2. The subspace $Span(\{N^{k-1}(\mathbf{x}), N^{k-2}(\mathbf{x}), \ldots, \mathbf{x}\}$ is called the cyclic subspace generated by $\mathbf{x}$, and denoted $C(\mathbf{x})$
3. The integer k is called the length of the cycle

Definition 6.2.5 We say that the cycles $\alpha_i = \{N^{k_i-1}(\mathbf{x}_i), \ldots, x_i\}$ are non-overlapping cycles if $\alpha_1 \cup \ldots \cup \alpha_r$ is linearly independent.

Definition 6.2.7 Let $N : V \to V$ be a nilpotent mapping on a finite-dimensional vector space V. We call a basis β for V a canonical basis (with respect to N) if β is the union of a collection of nonoverlapping cycles for N.

Theorem 6.2.8 (Canonical form for nilpotent mappings) Let $N : V \to V$ be a nilpotent mapping on a finite-dimensional vector space. There exists a canonical basis β of V with respect to N.

Lemma 6.2.9 Consider the cycle tableau corresponding to a canonical basis for a nilpotent mapping $N : V \to V$. As before, let r be the number of rows, and let k_i be the number of boxes in the ith row $(k_1 \geq k_2 \geq \ldots \geq k_r)$. For each $j(1 \leq j \leq k_1)$, the number of boxes in the jth column of the tableau is $\dim(Ker(N^j)) - \dim(Ker(N^{j-1})$.

Corollary 6.2.11 The canonical form of a nilpotent mapping is unique (provided the cycles in the canonical basis are arranged so the lengths satisfy $k_1 \geq k_2 \geq \ldots \geq k_r$)

6.3 Jordan Canonical Form

Proposition 6.3.1 Let $T : V \to V$ be a linear mapping whose characteristic polynomial has $\dim(V)$ roots (λ_i with respective multiplicities $m_i, 1 \leq i \leq k$) in the field F over which V is defined.

(a) There exist subspaces $V_i' \subset V (1 \leq i \leq k)$ such that

1. Each V_i' is invariant under T
2. $T|_{V_i'}$ has exactly one distinct eigenvalue λ_i, and
3. $V = V_1' \bigoplus \ldots \bigoplus V_k'$

(b) There exists a basis β for V such that $[T]_\beta^\beta$ has a direct sum decomposition into upper-triangular blocks of the form $\begin{bmatrix} \lambda & 1 & 0 & \ldots & 0 \\ 0 & \lambda & 1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & & \ldots & 0 & \lambda \end{bmatrix}$

Definition 6.3.2 Let $T : V \to V$ be a linear mapping on a finite-dimensional vector space V. Let λ be an eigenvalue of T with multiplicity m.

1. The λ-generalized eigenspace, denoted by K_λ, is the kernel of the mapping $(T - \lambda I)^m$ on V.
2. The nonzero elements of K_λ are called generalized eigenvectors of T.

Definitions 6.3.5

1. A matrix of the form $\begin{bmatrix} \lambda_i & 1 & 0 & \ldots & 0 \\ 0 & \lambda_i & 1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & & \ldots & 0 & \lambda_i \end{bmatrix}$ is called a Jordan block matrix
2. A matrix $A \in M_{n\times n}(\mathbb{F})$ is said to be in Jordan canonical form if A is a direct sum of Jordan block matrices.

Theorem 6.3.6 (Jordan Canonical Form) Let $T : V \to V$ be a linear mapping on a finite-dimensional vector space V whose characteristic polynomial has $\dim(V)$ roots in the field $\mathbb{F}$ over which V is defined.

1. There exists a basis γ (called a canonical basis) of V such that $[T]_\gamma^\gamma$ has a direct sum decomposition into Jordan block matrices.
2. In this decomposition the number of Jordan blocks and their sizes are uniquely determined by T. (The order in which the blocks appear in the matrix may be different for different canonical bases, however).

6.4 Computing Jordan Form

Algorithm

1. Find all the eigenvalues of T and their multiplicities by factoring the characteristic polynomial completely (assume the field is algebraically closed)
2. For each distinct eigenvalue λ_i in turn, construct the cycle tableau for a canonical basis of K_{λ_i} with respect to the mapping $N_i = (T - \lambda_i I)|K_{\lambda_i}$ using the method: for each j, the number of boxes in the jth column of the tableau for λ_i will be

$$\dim(Ker(T - \lambda_i I)^j) - \dim(Ker(T - \lambda_i I)^{j-1}$$

3. Form the corresponding Jordan blocks and assemble the matrix of T.

7 Problem Notes

1 $S = \{\mathbf{a}\} \subseteq \mathbb{R}^2$, then we cannot determine whether S is dependent (when $\mathbf{a} = \mathbf{0}$) or independent (when $\mathbf{a} \neq \mathbf{0}$)

2 If a set in a vector space contains the zero vector, then it is linearly dependent.

3 The order of Jordan blocks does not matter: if you change the order of Jordan blocks, it is still equivalent to the original one.

8 Proof Clinic - JCF

Facts

1. $E_\lambda \subset K_\lambda$, and both are T-invariant
2. $\forall \mu \neq \lambda, (T - \mu I)|_{K_\lambda}$ is bijective
3. $K_\lambda = Ker((T - \lambda I)^{m_i})$
4. Bases β_i, β_j for K_{λ_i} and K_{λ_j}, respectively, are disjoint if $\lambda_i \neq \lambda_j$
5. $\bigcup\limits_\lambda \beta_{K_\lambda}$ is a basis for V if each β_{K_λ} is a basis for K_λ
6. T is diagonalizable $\iff K_\lambda = E_\lambda \, \forall \lambda$
7. $V = \bigoplus\limits_\lambda K_\lambda$
8. Similar matrices have the same JCF

Suppose $\beta = \bigcup\limits_\lambda \gamma_\lambda$ is a basis of V, where each γ_λ is a cycle of generalized eigenvectors of T. Then $Span(\gamma_\lambda)$ is T-invariant and $[T|_{Span(\gamma_\lambda)}]_{\gamma_\lambda}$ is a Jordan block and β is a Jordan canonical basis.