Linear Algebra

Math 405

R. Ebrahimian
University of Maryland, College Park

January 24, 2024

Notations

- $\in$ belongs to.
- $\forall$ for all.
- $\exists$ there exists or for some.
- $\operatorname{Im} f$, the image of function f.
- $\mathbb{N}$, the set of nonnegative integers.
- $\mathbb{Z}^+$, the set of positive integers.
- $\mathbb{Q}$, the set of rational numbers.
- $\mathbb{R}$, the set of real numbers.
- $A \subseteq B$, set A is a subset of set B.
- $A \subsetneqq B$, set A is a proper subset of set B.
- $A \cup B$, the union of sets A and B.
- $A \cap B$, the intersection of sets A and B.
- $A \backslash B$, the difference set of B from A.
- $\bigcup_{i=1}^{n} A_i$, the union of sets $A_1, A_2, \ldots, A_n$.
- $\bigcap_{i=1}^{n} A_i$, the intersection of sets $A_1, A_2, \ldots, A_n$.
- $A_1 \times A_2 \times \cdots \times A_n$, the Cartesian product of sets $A_1, A_2, \ldots, A_n$.
- $\emptyset$, the empty set.
- $f^{-1}(T)$, the inverse image or pre-image of set T under function f.
- $f(S)$, the image of set S under function f.
- $\mathbb{P}_n$, the vector space of all polynomials of degree not exceeding n and coefficients in $\mathbb{F}$.
- $\mathbb{P}$, the vector space of all polynomials with coefficients in $\mathbb{F}$.
- $C[a, b]$, the vector space of all continuous functions from $[a, b]$ to $\mathbb{R}$.
- $\operatorname{span} \mathcal{S}$, the subspace spanned by the set $\mathcal{S}$.
- $\dim V$, the dimension of vector space V.
- A^{-1}, the inverse of a matrix or a function.

- $\det A$, the determinant of a square matrix A.
- S_n, the set consisting of all permutations of $\{1, 2, \ldots, n\}$.
- $\overline{A}$, the complex conjugate of A.
- A^T, the transpose of a matrix A.
- A^*, the adjoint of a matrix or linear transformation.

Contents

These note may contain occasional typos or errors. Feel free to email me at ebrahimi@umd.edu if you notice a typo or an error.

Week 1

1.1 Vector Spaces

Throughout these notes $\mathbb{F}$ denoted either the field of real numbers or complex numbers.

Throughout these notes **list** refers to an *unordered* collection of objects, where repetition is allowed. So, the three lists $1, 1, 2$ and $1, 2, 1$ and $2, 1, 1$ are all the same, but they are all different from both $1, 2, 2$ and $1, 2$.

Definition 1.1. Let V be a set of elements, called **vectors**. Suppose $+$, called **vector addition**, and $\cdot$, called **scalar multiplication**, are two functions for which $+$ assigns a vector $\mathbf{v} + \mathbf{w}$ to every two vectors $\mathbf{v}, \mathbf{w} \in V$ and $\cdot$ assign a vector $c \cdot \mathbf{v}$ to every $c \in \mathbb{F}$ and every $\mathbf{v} \in V$. We say V is a vector space over $\mathbb{F}$ iff the following properties are satisfied:

(I) (Closure) For every two vectors $\mathbf{x}, \mathbf{y} \in V$, and every $c \in \mathbb{F}$, both $\mathbf{x} + \mathbf{y}$ and $c \cdot \mathbf{x}$ are in V.

(II) (Associativity) For every $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, and every $a, b \in \mathbb{F}$, we have $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$, and $a \cdot (b \cdot \mathbf{x}) = (ab) \cdot \mathbf{x}$.

(III) (Commutativity) For every $\mathbf{x}, \mathbf{y} \in V$, we have $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.

(IV) (Additive Identity) There is a vector $\mathbf{e} \in V$ for which, for every $\mathbf{x} \in V$ we have $\mathbf{x} + \mathbf{e} = \mathbf{x}$.

(V) (Additive Inverse) For every $\mathbf{x} \in V$, there is an element $\mathbf{y} \in V$ for which $\mathbf{x} + \mathbf{y} = \mathbf{e}$.

(VI) (Distributivity) For every $a, b \in \mathbb{F}$ and every $\mathbf{x}, \mathbf{y} \in V$, we have $(a + b) \cdot \mathbf{x} = a \cdot \mathbf{x} + b \cdot \mathbf{x}$, and $a \cdot (\mathbf{x} + \mathbf{y}) = a \cdot \mathbf{x} + a \cdot \mathbf{y}$.

(VII) (Multiplicative Identity) For every $\mathbf{x} \in V$ we have $1 \cdot \mathbf{x} = \mathbf{x}$.

When $\mathbb{F} = \mathbb{R}$, we say V is a **real vector space** and when $\mathbb{F} = \mathbb{C}$, we say V is a **complex vector space**. For simplicity $c \cdot \mathbf{x}$ is often denoted by $c\mathbf{x}$. Elements of $\mathbb{F}$ are called **scalars**.

Example 1.1. The following are examples of vector spaces:

(a) $\mathbb{F}^n$, the set of n-tuples with entries in $\mathbb{F}$ along with the standard componentwise vector addition and scalar multiplication.

(b) $\mathbb{P}_n = \{a_0 + a_1 t + \cdots + a_n t^n \mid a_0, a_1, \ldots, a_n \in \mathbb{F}\}$, the set of polynomials of degree not exceeding n, with the usual polynomial addition and scalar multiplication:

$$(a_0 + a_1 t + \cdots + a_n t^n) + (b_0 + b_1 t + \cdots + b_n t^n) = (a_0 + b_0) + (a_1 + b_1)t + \cdots + (a_n + b_n)t^n$$

$$c(a_0 + a_1 t + \cdots + a_n t^n) = ca_1 + ca_1 t + \cdots + ca_n t^n.$$

To emphasize the field $\mathbb{F}$ we often denoted $\mathbb{P}_n$ by $\mathbb{P}_n(\mathbb{F})$.

(c) $\mathbb{P} = \bigcup\limits_{n=1}^{\infty} \mathbb{P}_n$, the set of all polynomials with coefficients in $\mathbb{F}$ along with the usual polynomial addition and scalar multiplication. To emphasize the field $\mathbb{F}$ we often denoted $\mathbb{P}$ by $\mathbb{P}(\mathbb{F})$.

(d) $M_{m\times n}(\mathbb{F})$, the set of $m \times n$ matrices with entries in $\mathbb{F}$ along with the standard entrywise matrix addition and scalar multiplication.

(e) $C[D]$, the set of continuous functions $f : D \to \mathbb{R}$, where D is a given nonempty subset of $\mathbb{R}$, along with the standard addition and scalar multiplication is a real vector space.

Theorem 1.1. *Let V be a vector space. Then,*

(a) the additive identity is unique.

(b) the additive inverse of every vector is unique.

Remark 1.1. The additive identity of a vector space V is denoted by $\mathbf{0}_V$, or $\mathbf{0}$ if there is no ambiguity. The additive inverse of every vector $\mathbf{v}$ is denoted by $-\mathbf{v}$. The vector $\mathbf{u} - \mathbf{v}$ denotes the sum of vectors $\mathbf{u}$ and $-\mathbf{v}$ is called the **difference** of $\mathbf{v}$ from $\mathbf{u}$.

1.2 Linear Independence, Generating, and Bases

Definition 1.2. Let V be a vector space. A vector $\mathbf{w} \in V$ is said to be a **linear combination** of a list of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$, iff there are scalar $c_1, \ldots, c_n$ for which

$$\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n.$$

The only linear combination of a list of no vectors is the zero vector. The linear combination above is said to be **trivial** iff $c_1 = \cdots = c_n = 0$.

Definition 1.3. A list of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in a vector space V is called a **basis** for V iff every vector $\mathbf{w} \in V$ can uniquely be written as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

Definition 1.4. A list of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in V is said to be **generating** or **spanning** iff every vector in V can be represented as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

Definition 1.5. A list of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is said to be **linearly independent** iff their only linear combination that is $\mathbf{0}$ is the trivial one. Otherwise, we say they are linearly dependent.

Theorem 1.2. *A list of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is linearly dependent if and only if one of the v_j's is a linear combination of the others.*

Theorem 1.3. *A list of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ in a vector space V form a basis if and only if they are linearly independent and generating.*

Theorem 1.4. *Any finite[1] generating list of vectors in a vector space contains a basis.*

1.3 Linear Transformations

Definition 1.6. Let V, W be two vector spaces over the same field $\mathbb{F}$. A function $T : V \to W$ is said to be **linear** iff both of the following are satisfied:

(a) (Additivity) For every $\mathbf{u}, \mathbf{v} \in V$, we have $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$, and

(b) (Homogenenity) For every $\mathbf{v} \in V$ and every $c \in \mathbb{F}$, we have $T(c\mathbf{v}) = cT(\mathbf{v})$.

Note that this definition requires V and W to be vector spaces over the *same* field $\mathbb{F}$.

Theorem 1.5. *Suppose V, W are vector spaces over the same field $\mathbb{F}$, and $T : V \to W$ is a function. The following are equivalent:*

(a) T is linear.

(b) For every $a, b \in \mathbb{F}$ and every $\mathbf{u}, \mathbf{v} \in V$, we have $T(a\mathbf{u} + b\mathbf{v}) = aT(\mathbf{u}) + bT(\mathbf{v})$.

(c) For every $a \in \mathbb{F}$ and every $\mathbf{u}, \mathbf{v} \in V$, we have $T(\mathbf{u} + a\mathbf{v}) = T(\mathbf{u}) + aT(\mathbf{v})$.

Example 1.2. The following are linear transformations.

(a) $D : \mathbb{P} \to \mathbb{P}$ given by $D(f(t)) = f'(t)$.

(b) $T : \mathbb{P}_n(\mathbb{R}) \to \mathbb{R}$ given by $T(f(t)) = \displaystyle\int_0^1 f(t)\ dt$.

(c) Rotation about the origin in $\mathbb{R}^2$.

Theorem 1.6. *Let V, W be vector spaces over the same field $\mathbb{F}$. Consider $\mathcal{L}(V, W)$, the set of all linear transformations $T : V \to W$. Define the following addition and scalar multiplication for every $T, S \in \mathcal{L}(V, W)$ and every $c \in \mathbb{F}$.*

$$(T + S)(\mathbf{v}) = T(\mathbf{v}) + S(\mathbf{v}),\ \text{and}\ (cT)(\mathbf{v}) = cT(\mathbf{v}),\ \text{for all}\ \mathbf{v} \in V.$$

$\mathcal{L}(V, W)$ equipped with these two operations is a vector space over $\mathbb{F}$.

[1] The finiteness can be dropped if we assume the Axiom of Choice.

1.4 Matrix Multiplication

Theorem 1.7. *Suppose $T : V \to W$ and $S : W \to U$ are linear transformations. Then, $S \circ T : V \to U$ is linear.*

There is a close relation between linear transformations and matrices.

Definition 1.7. A **matrix** is an arrangement of objects in a rectangular array with a finite number of columns and rows. A matrix with m rows and n columns is said to be an $m \times n$ matrix. The set of $m \times n$ matrices with entries from a field $\mathbb{F}$ is denoted by $M_{m\times n}(\mathbb{F})$. The entry in the j-th row and k-th column of a matrix A is often denoted by a_{jk}, and is called the (j, k) entry of A. This is often written as $A = (a_{jk})_{m\times n}$, or $A = (a_{jk})$, to indicate A is an $m \times n$ matrix whose (j, k) entry is a_{jk}. The **transpose** of an $m \times n$ matrix A, denoted by A^T, is an $n \times m$ matrix whose (j, k) entry is a_{kj} for all $j = 1, \ldots, n$ and $k = 1, \ldots, m$. When the number of rows and the number of columns of a matrix are the same, we say the matrix is a **square matrix**. A matrix A is called **symmetric** iff $A^T = A$. It is called **antisymmetric** or **skew symmetric** iff $A^T = -A$.

Note that symmetric and antisymmetric matrices are square matrices.

Definition 1.8. Let $A \in M_{m\times n}(\mathbb{F})$ and $\mathbf{v} \in \mathbb{F}^n$ be a column vector. The product $A\mathbf{v}$ is an $m \times 1$ vector that is obtain by taking a linear combination of columns of A with scalars from entries of $\mathbf{v}$. In other words, if $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are columns of A from left to right, and $c_1, \ldots, c_n$ are entries of $\mathbf{v}$ from top to bottom, then $A\mathbf{v} = c_1\mathbf{a}_1 + \cdots c_n\mathbf{a}_n$.

Theorem 1.8. *Let $L : \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation. Then, there is a unique $A \in M_{m\times n}(\mathbb{F})$ for which $L(\mathbf{v}) = A\mathbf{v}$ for all $\mathbf{v} \in \mathbb{F}^n$. Furthermore, for every $A \in M_{m\times n}(\mathbb{F})$, the function $L : \mathbb{F}^n \to \mathbb{F}^m$ defined by $L(\mathbf{v}) = A\mathbf{v}$ is linear.*

Definition 1.9. Let $A \in M_{m\times n}(\mathbb{F})$ and $B \in M_{n\times k}(\mathbb{F})$. The product AB is an $m \times k$ matrix whose j-th column $A\mathbf{b}_j$, where $\mathbf{b}_j$ is the j-th column of B.

Remark 1.2. Let $A = (a_{jk}) \in M_{m\times n}(\mathbb{F})$ and $B = (b_{jk}) \in M_{n\times p}(\mathbb{F})$. If $AB = (c_{jk})$, then $c_{jk} = \sum\limits_{\ell=1}^{n} a_{j\ell}b_{\ell k}$.

Remark 1.3. Note that in general $AB \neq BA$. For example if A is a 2×3 matrix and B is a 3×3 matrix, then AB is well-defined, but BA is not. If $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, then

$$AB = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \neq BA = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Theorem 1.9 (Properties of Matrix Multiplication)**.** *For every three matrices A, B, C:*

(a) (Associativity) $(AB)C = A(BC)$ and $\alpha(AB) = (\alpha A)B = A(\alpha B)$.

(b) (Distributivity) $A(B + C) = AB + AC$ and $(B + C)A = BA + CA$.

(c) $(AB)^T = B^T A^T$.

In each case, we assume the sizes of matrices are so that one side of the equality is well-defined. Then, the other side will automatically be well-defined and the equality holds.

Definition 1.10. The trace of an $n \times n$ matrix $A = (a_{jj})$, denote by $\mathrm{tr}\, A$ or $\mathrm{tr}\,(A)$, is defined as $\mathrm{tr}\, A = \sum_{j=1}^{n} a_{jj}$.

Example 1.3. $\mathrm{tr}\, : M_n(\mathbb{F}) \to \mathbb{F}$ is a linear transformation.

Example 1.4. Suppose $A \in M_{m\times n}(\mathbb{F})$ and $B \in M_{n\times m}(\mathbb{F})$. Prove that $\mathrm{tr}\,(AB) = \mathrm{tr}\,(BA)$.

Definition 1.11. The **diagonal** of an $n \times n$ matrix $A = (a_{ij})_{n\times n}$ is the list of all entries of the form a_{kk} with $k = 1, \ldots, n$. An entry of the form a_{jk} with $j \neq k$ is called an **off-diagonal entry**. A is called **diagonal** iff all of its off-diagonal entries are zero. A is called **upper triangular** (resp., **lower triangular**) iff $a_{jk} = 0$ for all $1 \le k < j \le n$ (resp., $1 \le j < k \le n$).

Theorem 1.10 (Block Multiplication)**.** *Suppose matrices A and B are given as block matrices below*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \; B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22}. \end{pmatrix},$$

where A_{jk}, B_{jk} are themselves matrices. Assuming all appropriate multiplications and additions are defined we have

$$AB = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}$$

Similar result holds for block matrices with a larger number of blocks.

1.5 Examples

Example 1.5 ("Zero" does not mean *zero*)**.** Consider the set $\mathbb{F}^2$ along with the following two operations:

$$(x, y) \oplus (z, t) = (x + z - 1, y + t - 2), \text{ and } c \cdot (x, y) = (cx - c + 1, cy - 2c + 2), \text{ for all } x, y \in \mathbb{F}$$

Prove that $\mathbb{F}^2$ along with the above vector addition and scalar multiplication is a vector space. What is the zero of this vector space? What is the additive inverse of (x, y)?

Solution. The given vector addition can be written as $\mathbf{u} \oplus \mathbf{v} = \mathbf{u} + \mathbf{v} - \mathbf{e}$, for every $\mathbf{u}, \mathbf{v} \in \mathbb{F}^2$, where $\mathbf{e} = (1, 2)$. Here, $+$ and $-$ are the usual vector addition and subtraction of $\mathbb{F}^2$. We also see that $c \cdot \mathbf{u} = c\mathbf{u} - c\mathbf{e} + \mathbf{e}$ for every $\mathbf{u} \in \mathbb{F}^2$. We will now show $\oplus$ and $\cdot$ satisfy all properties of a vector space.

Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{F}^2$ and $a, b \in \mathbb{F}$.

Closure: The vectors $\mathbf{u} + \mathbf{v} - \mathbf{e}$ and $a\mathbf{u} - a\mathbf{e} + \mathbf{e}$ are in $\mathbb{F}^2$, since $\mathbb{F}^2$ is closed under the standard vector addition and scalar multiplication.

Associativity:

$$\begin{aligned}(\mathbf{u}\oplus\mathbf{v})\oplus\mathbf{w} &= (\mathbf{u}+\mathbf{v}-\mathbf{e})\oplus\mathbf{w}\\ &= (\mathbf{u}+\mathbf{v}-\mathbf{e})+\mathbf{w}-\mathbf{e}\\ &= \mathbf{u}+\mathbf{v}+\mathbf{w}-2\mathbf{e}\end{aligned}$$

On the other hand,

$$\begin{aligned}\mathbf{u}\oplus(\mathbf{v}\oplus\mathbf{w}) &= \mathbf{u}\oplus(\mathbf{v}+\mathbf{w}-\mathbf{e})\\ &= \mathbf{u}+(\mathbf{v}+\mathbf{w}-\mathbf{e})-\mathbf{e}\\ &= \mathbf{u}+\mathbf{v}+\mathbf{w}-2\mathbf{e}\end{aligned}$$

Therefore, $(\mathbf{u}\oplus\mathbf{v})\oplus\mathbf{w}=\mathbf{u}\oplus(\mathbf{v}\oplus\mathbf{w})$.

We also see:

$$(ab)\cdot\mathbf{u}=(ab)\mathbf{u}-(ab)\mathbf{e}+\mathbf{e}, \text{and}$$

$$a\cdot(b\cdot\mathbf{u})=a\cdot(b\mathbf{u}-b\mathbf{e}+\mathbf{e})=a(b\mathbf{u}-b\mathbf{e}+\mathbf{e})-a\mathbf{e}+\mathbf{e}=a(b\mathbf{u})-a(b\mathbf{e})+a\mathbf{e}-a\mathbf{e}+\mathbf{e}=(ab)\mathbf{u}-(ab)\mathbf{e}+\mathbf{e}.$$

Thus, $(ab)\cdot\mathbf{u}=a\cdot(b\cdot\mathbf{u})$.

Commutativity: $\mathbf{u}\oplus\mathbf{v}=\mathbf{u}+\mathbf{v}-\mathbf{e}$ and $\mathbf{v}\oplus\mathbf{u}=\mathbf{v}+\mathbf{u}-\mathbf{e}$. Since the standard vector addition is commutative $\mathbf{u}\oplus\mathbf{v}=\mathbf{v}\oplus\mathbf{u}$.

Additive Identity: We need to find a vector $\mathbf{x}$ for which $\mathbf{x}\oplus\mathbf{u}=\mathbf{u}$ for every vector $\mathbf{u}$. This is equivalent to $\mathbf{x}+\mathbf{u}-\mathbf{e}=\mathbf{u}$, which is equivalent to $\mathbf{x}=\mathbf{e}$. Thus, $\mathbf{e}$ is the additive identity of $\mathbb{F}^2$.

Additive Inverse: $\mathbf{v}$ is an additive inverse of $\mathbf{u}$ if and only if $\mathbf{u}\oplus\mathbf{v}=\mathbf{e}$. This is equivalent to $\mathbf{u}+\mathbf{v}-\mathbf{e}=\mathbf{e}$, i.e. $\mathbf{v}=-\mathbf{u}+2\mathbf{e}$. Thus, the additive inverse of $\mathbf{u}$ is $-\mathbf{u}+2\mathbf{e}$.

Distributivity: $(a+b)\cdot\mathbf{u}=(a+b)\mathbf{u}-(a+b)\mathbf{e}+\mathbf{e}=a\mathbf{u}+b\mathbf{u}-a\mathbf{e}-b\mathbf{e}+\mathbf{e}$.

$$\begin{aligned}a\cdot\mathbf{u}\oplus b\cdot\mathbf{u} &= a\cdot\mathbf{u}+b\cdot\mathbf{u}-\mathbf{e}\\ &= a\mathbf{u}-a\mathbf{e}+\mathbf{e}+b\mathbf{u}-b\mathbf{e}+\mathbf{e}-\mathbf{e}\\ &= a\mathbf{u}+b\mathbf{u}-a\mathbf{e}-b\mathbf{e}+\mathbf{e}\end{aligned}$$

Therefore, $(a+b)\cdot\mathbf{u}=a\cdot\mathbf{u}\oplus b\cdot\mathbf{u}$.

$a\cdot(\mathbf{u}\oplus\mathbf{v})=a(\mathbf{u}\oplus\mathbf{v})-a\mathbf{e}+\mathbf{e}=a(\mathbf{u}+\mathbf{v}-\mathbf{e})-a\mathbf{e}+\mathbf{e}=a\mathbf{u}+a\mathbf{v}-2a\mathbf{e}+\mathbf{e}$

$$\begin{aligned}a\cdot\mathbf{u}\oplus a\cdot\mathbf{v} &= a\cdot\mathbf{u}+a\cdot\mathbf{v}-\mathbf{e}\\ &= a\mathbf{u}-a\mathbf{e}+\mathbf{e}+a\mathbf{v}-a\mathbf{e}+\mathbf{e}-\mathbf{e}\\ &= a\mathbf{u}+a\mathbf{v}-2a\mathbf{e}+\mathbf{e}\end{aligned}$$

Therefore, $a\cdot(\mathbf{u}\oplus\mathbf{v})=a\cdot\mathbf{u}\oplus a\cdot\mathbf{v}$.

Multiplicative Identity: $1 \cdot \mathbf{u} = 1\mathbf{u} - 1\mathbf{e} + \mathbf{e} = \mathbf{u} - \mathbf{e} + \mathbf{e} = \mathbf{u}$. Therefore, $\mathbb{F}^2$ along with $\oplus$ and $\cdot$ is a vector space.

Note that the additive identity of $\mathbb{F}^2$ (i.e. its "zero") is $\mathbf{e} = (1, 2)$. The additive inverse of (x, y) is $-(x, y) + 2(1, 2) = (-x + 2, -y + 4)$. $\square$

Example 1.6. Prove that for every scalar c we have $c\mathbf{0} = \mathbf{0}$.

Solution. Since $\mathbf{0}$ is the additive identity we have $\mathbf{0} + \mathbf{0} = \mathbf{0}$. Using the distributive property we obtain $c\mathbf{0} + c\mathbf{0} = c\mathbf{0}$. Adding the additive inverse of $c\mathbf{0}$ to both sides we conclude $c\mathbf{0} + \mathbf{0} = \mathbf{0}$. Therefore, $c\mathbf{0} = \mathbf{0}$, as desired. $\square$

Example 1.7. Prove that if for a scalar c and a vector $\mathbf{v}$ we have $c\mathbf{v} = \mathbf{0}$, then $c = 0$ or $\mathbf{v} = \mathbf{0}$.

Solution. Suppose $c\mathbf{v} = \mathbf{0}$, but $c \neq 0$. Then $\frac{1}{c}(c\mathbf{v}) = \frac{1}{c}\mathbf{0} = \mathbf{0}$. By associativity, and multiplicative identity we have $\frac{1}{c}(c\mathbf{v}) = (\frac{1}{c}c)\mathbf{v} = 1\mathbf{v} = \mathbf{v}$. Therefore, $\mathbf{v} = \mathbf{0}$, as desired. $\square$

Example 1.8. Prove part (a) of Theorem 1.1.

Solution. Suppose $\mathbf{0}$ and $\mathbf{0}'$ are additive identities of a vector space V. We have the following:

$$\left.\begin{array}{ll} \mathbf{0} + \mathbf{0}' = \mathbf{0} & \text{Since } \mathbf{0}' \text{ is an additive identity} \\ \mathbf{0} + \mathbf{0}' = \mathbf{0}' & \text{Since } \mathbf{0} \text{ is an additive identity} \end{array}\right\} \Rightarrow \mathbf{0} = \mathbf{0}'$$

This completes the proof. $\square$

Example 1.9. Let V be a vector space. Prove that for every vector $\mathbf{v} \in V$, we have $(-1)\mathbf{v} = -\mathbf{v}$.

Solution. We note that

$$\begin{aligned} \mathbf{v} + (-1)\mathbf{v} &= 1\mathbf{v} + (-1)\mathbf{v} && \text{Multiplicative identity} \\ &= (1 + (-1))\mathbf{v} && \text{Distributive property} \\ &= 0\mathbf{v} = \mathbf{0} && \text{An exercise} \end{aligned}$$

Therefore, by adding $-\mathbf{v}$ to both sides of $\mathbf{v} + (-1)\mathbf{v} = \mathbf{0}$ we conclude $(-1)\mathbf{v} = -\mathbf{v}$, as desired. $\square$

Definition 1.12. For every nonempty set S, the vector space of all functions $f : S \to \mathbb{F}$ is denoted by $\mathcal{F}(S, \mathbb{F})$.

Example 1.10. Find a basis for $\mathcal{F}(\{1, 2, \ldots, n\}, \mathbb{F})$.

Scratch: We see that each function $f : \{1, 2, \ldots, n\} \to \mathbb{F}$ is determined by $f(1), f(2), \ldots, f(n)$. We need to find n functions that generate all functions. We can do that by choosing functions that are 1 at one value and zero for all other values.

Solution. Let $V = \mathcal{F}(\{1, 2, \ldots, n\}, \mathbb{F})$. For every i, define $f_i : \{1, 2, \ldots, n\} \to \mathbb{F}$ by $f_i(i) = 1$, and $f_i(j) = 0$ for all $j \neq i$.

Linear independence: Suppose $c_1 f_1 + c_2 f_2 + \cdots + c_n f_n = 0$ for some $c_1, c_2, \ldots, c_n \in \mathbb{R}$. Evaluating both sides at i we obtain $c_i f_i(i) = 0$, and thus $c_i = 0$, which completes the proof.

Generating: Let $f \in V$ and let $g = f(1)f_1 + f(2)f_2 + \cdots + f(n)f_n$. We see that $g(i) = f(i)f_i(i) = f(i)$ for all i. Therefore, $f = g$. This means f is a linear combination of $f_1, f_2, \ldots, f_n$. □

Example 1.11. Prove that every set of vectors that contains the vector $\mathbf{0}$ is linearly dependent.

Solution. Let $\mathcal{S}$ be a set of vectors containing $\mathbf{0}$. We see that $1\mathbf{0} = \mathbf{0}$ and the coefficient 1 is nonzero. Therefore the set $\mathcal{S}$ is linearly dependent. □

Example 1.12. Let $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{F}^n$ be linearly independent. Consider arbitrary vectors $\mathbf{w}_1, \ldots, \mathbf{w}_m \in \mathbb{F}^k$ and form the vectors $\mathbf{x}_1 = (\mathbf{v}_1, \mathbf{w}_1), \ldots, \mathbf{x}_m = (\mathbf{v}_m, \mathbf{w}_m) \in \mathbb{F}^{n+k}$ formed by placing each $\mathbf{w}_j$ next to $\mathbf{v}_j$. Prove that $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are linearly independent.

Solution. Let $c_1, \ldots, c_m \in \mathbb{F}$ be scalars for which

$$c_1\mathbf{x}_1 + \cdots + c_m\mathbf{x}_m = \mathbf{0}.$$

Using the way $\mathbf{x}_j$'s are created we have

$$c_1(\mathbf{v}_1, \mathbf{w}_1) + \cdots + c_m(\mathbf{v}_m, \mathbf{w}_m) = \mathbf{0} \Rightarrow (c_1\mathbf{x}_1 + \cdots + c_m\mathbf{v}_m, c_1\mathbf{w}_1 + \cdots + c_m\mathbf{w}_m) = \mathbf{0} \Rightarrow c_1\mathbf{x}_1 + \cdots + c_m\mathbf{v}_m = \mathbf{0}.$$

Since $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly independent we obtain $c_1 = \cdots = c_m = 0$, and hence $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are linearly independent. □

Example 1.13. Let V, W be vector spaces over $\mathbb{F}$. Assume $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form a basis for V, and let $\mathbf{w}_1, \ldots, \mathbf{w}_n \in W$. Prove that $T : V \to W$ defined by

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n) = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_n\mathbf{w}_n, \text{ for all } c_1, c_2, \ldots, c_n \in \mathbb{F}$$

is a linear transformation.

Solution. Suppose $\mathbf{x}, \mathbf{y} \in V$, and $c \in \mathbb{F}$. Since $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form a basis for V, there are scalars $a_j, b_j \in \mathbb{F}$ for which $\mathbf{x} = \sum_{j=1}^{n} a_j\mathbf{v}_j$ and $\mathbf{y} = \sum_{j=1}^{n} b_j\mathbf{v}_j$. Since $\mathbf{x} + c\mathbf{y} = \sum_{j=1}^{n}(a_j + cb_j)\mathbf{v}_j$, we have

$$T(\mathbf{x} + c\mathbf{y}) = \sum_{j=1}^{n}(a_j + cb_j)\mathbf{w}_j = \sum_{j=1}^{n} a_j\mathbf{w}_j + \sum_{j=1}^{n} cb_j\mathbf{w}_j = T(\mathbf{x}) + cT(\mathbf{y}).$$

Therefore, T is linear. □

Example 1.14. Let V, W be vector spaces over $\mathbb{F}$, and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be a basis for V. Assume $S, T : V \to W$ are linear transformations. Prove that $S = T$ if and only if $S(\mathbf{v}_j) = T(\mathbf{v}_j)$ for $j = 1, \ldots, n$.

Solution. $\Rightarrow$: If $S = T$, then $S(\mathbf{v}_j) = T(\mathbf{v}_j)$, as desired.

$\Leftarrow$: Suppose $S(\mathbf{v}_j) = T(\mathbf{v}_j)$ for $j = 1, \ldots, n$. Let $\mathbf{v} \in V$. Since $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a basis for V, there are scalars $c_1, c_2, \ldots, c_n$ for which $\mathbf{v} = \sum_{j=1}^{n} c_j \mathbf{v}_j$. By linearity of S and T, and the fact that $S(\mathbf{v}_j) = T(\mathbf{v}_j)$ we have

$$S(\mathbf{v}) = S(\sum_{j=1}^{n} c_j \mathbf{v}_j) = \sum_{j=1}^{n} c_j S(\mathbf{v}_j) = \sum_{j=1}^{n} c_j T(\mathbf{v}_j) = T(\sum_{j=1}^{n} c_j \mathbf{v}_j) = T(\mathbf{v}).$$

Therefore, $S = T$, as desired. $\square$

Example 1.15. Suppose $T : \mathbb{F}^2 \to \mathbb{F}^3$ is a linear transformation for which $T(1, 2) = (1, 0, 1)$ and $T(2, 1) = (1, 1, 0)$. Find the matrix A for which $T(\mathbf{v}) = A\mathbf{v}$ for all $\mathbf{v} \in \mathbb{F}^2$.

Solution. We need to find $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$. We see

$$(1, 0) = \frac{2}{3}(2, 1) - \frac{1}{3}(1, 2), \text{ and } (0, 1) = \frac{2}{3}(1, 2) - \frac{1}{3}(2, 1).$$

By linearity of T we have

$$T(\mathbf{e}_1) = \frac{2}{3}T(2, 1) - \frac{1}{3}T(1, 2) = \frac{2}{3}(1, 1, 0) - \frac{1}{3}(1, 0, 1) = (1/3, 2/3, -1/3),$$

and

$$T(\mathbf{e}_2) = \frac{2}{3}T(1, 2) - \frac{1}{3}T(2, 1) = \frac{2}{3}(1, 0, 1) - \frac{1}{3}(1, 1, 0) = (1/3, -1/3, 2/3).$$

Therefore, by a theorem the matrix A is given by

$$A = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{2}{3} \end{pmatrix}$$

$\square$

Example 1.16. Let $T : V \to W$ be a linear transformation of vector spaces over $\mathbb{F}$. Prove that for every $c_1, \ldots, c_n \in \mathbb{F}$ and $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ we have

$$T(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) = c_1 T(\mathbf{v}_1) + \cdots + c_n T(\mathbf{v}_n).$$

Solution. We prove this by induction on n.

Basis step. The equality $T(c_1\mathbf{v}_1) = c_1 T(\mathbf{v}_1)$ follows from homogeneity.

Inductive Step. Suppose the given equality holds for a positive integer n and let $c_1, \ldots, c_{n+1} \in \mathbb{F}$, $\mathbf{v}_1, \ldots, \mathbf{v}_{n+1} \in V$. We have the following:

$$\begin{aligned} T(\sum_{i=1}^{n+1} c_i\mathbf{v}_i) &= T(\sum_{i=1}^{n} c_i\mathbf{v}_i) + T(c_{n+1}\mathbf{v}_{n+1}) \text{ by additivity} \\ &= \sum_{i=1}^{n} c_i T(\mathbf{v}_i) + T(c_{n+1}\mathbf{v}_{n+1}) \text{ by inductive hypothesis} \\ &= \sum_{i=1}^{n} c_i T(\mathbf{v}_i) + c_{n+1}T(\mathbf{v}_{n+1}) \text{ by homogeneity} \\ &= \sum_{i=1}^{n+1} c_i T(\mathbf{v}_i) \end{aligned}$$

This completes the proof. □

Example 1.17. Prove that for every linear transformation $T : V \to W$ we have $T(\mathbf{0}) = \mathbf{0}$.

Solution. We see the following:

$$\begin{aligned} T(\mathbf{0}) &= T(\mathbf{0} + \mathbf{0}) \text{ Since } \mathbf{0} \text{ is the additive identity} \\ &= T(\mathbf{0}) + T(\mathbf{0}) \text{ By additivity of } T \end{aligned}$$

Adding $-T(\mathbf{0})$ to both sides, we conclude $T(\mathbf{0}) = \mathbf{0}$, as desired. □

1.6 Exercises

Exercise 1.1. *Prove part (b) of Theorem 1.1.*

Exercise 1.2. *Prove that if* $\mathbf{v}$ *is a vector and* c *is a scalar, then* $0\mathbf{v} = \mathbf{0}$ *and* $c\mathbf{0} = \mathbf{0}$.

Exercise 1.3. *Let* V *be a vector space over* $\mathbb{F}$. *Prove that for every* $a, b \in \mathbb{F}$ *and every* $\mathbf{u}, \mathbf{v} \in V$ *we have:*

(a) $-(-\mathbf{u}) = \mathbf{u}$.

(b) $a(\mathbf{u} - \mathbf{v}) = a\mathbf{u} - a\mathbf{v}$.

(c) $-(\mathbf{u} + \mathbf{v}) = -\mathbf{u} - \mathbf{v}$.

(d) $(a - b)\mathbf{u} = a\mathbf{u} - b\mathbf{v}$.

Exercise 1.4. *Prove that if for a vector* $\mathbf{v}$ *and a scalar* c *we have* $c\mathbf{v} = \mathbf{0}$, *then* $c = 0$ *or* $\mathbf{v} = \mathbf{0}$.

Exercise 1.5. *Let* S *be a nonempty set. Consider the set* $V = \mathcal{F}(S, \mathbb{F})$ *consisting of all functions* $f : S \to \mathbb{F}$ *equipped with the addition and scalar multiplication defined below for all* $f, g \in V$ *and* $c \in \mathbb{F}$:

$$(f + g)(x) = f(x) + g(x), \text{ and } (cf)(x) = cf(x).$$

Using the definition, prove that V *is a vector space.*

Exercise 1.6. *Given a vector* $\mathbf{v}$ *in a vector space* V, *define a vector addition and a scalar multiplication that turns* V *into a vector space, where* $\mathbf{v}$ *is the zero of this new vector space.*

Hint: See Example 1.5.

Exercise 1.7. *Using the fact that* $(XY)_{ij} = \sum X_{i\ell}Y_{\ell j}$, *prove that for every three matrices* A, B, C *we have* $(AB)C = A(BC)$. *Assume all the products are well-defined.*

Exercise 1.8. *Write down a basis for* $M_n(\mathbb{F})$.

Exercise 1.9. *Suppose a list of vectors* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *is linearly independent but not generating. Prove that if* $\mathbf{v}_{n+1}$ *is a vector that cannot be written as a linear combination of* $\mathbf{v}_1, \ldots, \mathbf{v}_n$, *then* $\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1}$ *are linearly independent.*

Exercise 1.10. *Suppose* $\mathbf{u}, \mathbf{v}, \mathbf{w}$ *are linearly dependent vectors. Prove that* $\mathbf{u}+\mathbf{v}$, $\mathbf{v}+\mathbf{w}$ *and* $\mathbf{w}+\mathbf{u}$ *are also linearly dependent.*

Exercise 1.11. *Construct a nonzero matrix* A *for which* $A^2 = 0$.

Exercise 1.12. *Determine if each of the following is a linear transformation:*

(a) $T : \mathbb{F}^2 \to \mathbb{F}^3$ *defined by* $T(x, y, z) = (2x + y, z - y + x)$.

(b) $L : M_{m\times n}(\mathbb{F}) \to M_{n\times m}(\mathbb{F})$ *defined by* $L(A) = A^T$.

Exercise 1.13. *Find a basis for* $\mathbb{C}$, *once as a real vector space and once as a complex vector space.*

Exercise 1.14. *Let* $T : \mathbb{C} \to \mathbb{C}$ *be defined by* $T(a + bi) = 2a + (3b - a)i$ *for every* $a, b \in \mathbb{R}$. *Is* T *a linear transformation when* $\mathbb{C}$ *is considered a complex vector space? How about when* $\mathbb{C}$ *is considered a real vector space?*

Exercise 1.15. *Prove that* $L : M_{m\times n}(\mathbb{F}) \to M_{n\times m}(\mathbb{F})$ *defined by* $L(A) = A^T$ *is a linear transformation.*

Exercise 1.16. *Suppose* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *are linearly independent vectors in a vector space* V. *Let* $\mathbf{w} \in V$. *Prove that the vectors* $\mathbf{v}_1 - \mathbf{w}, \mathbf{v}_2 - \mathbf{w}, \ldots, \mathbf{v}_n - \mathbf{w}$ *are linearly dependent if and only if* $\mathbf{w} = \sum_{j=1}^{n} c_j \mathbf{v}_j$, *where,* $\sum_{j=1}^{n} c_j = 1$.

Exercise 1.17. *Using the definition, determine if each of the following is a vector space.*

(a) The set consisting of all polynomials on variable t *with degree 3, along with the usual polynomial addition and scalar multiplication.*

(b) The set consisting of all 2×2 *matrices* $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *that satisfy* $a + b = c + d$.

(c) The set of all functions $f : \mathbb{R} \to \mathbb{R}$ *that are continuous but not differentiable along with the standard function addition and scalar multiplication.*

Exercise 1.18. *Suppose vectors* $\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{w}$ *in a vector space are linearly independent. Prove that for every scalar* c, *the vectors* $c\mathbf{w} + \mathbf{v}_1, c\mathbf{w} + \mathbf{v}_2, \ldots, c\mathbf{w} + \mathbf{v}_n$ *are linearly independent as well.*

Exercise 1.19. *Determine if each statement is true or false.*

(a) Closure of V under addition states: $\mathbf{x}, \mathbf{y} \in V$ if and only if $\mathbf{x} + \mathbf{y} \in V$.

(b) By definition of a vector space, $0\mathbf{v} = \mathbf{0}$.

(c) The set of all quadratic polynomials of the form $at^2 + bt + c$, with $a, b, c \in \mathbb{R}$ and $a \neq 0$, is a real vector space.

(d) The set of functions $f : \mathbb{R} \to \mathbb{R}$ of the form $f(x) = c_1 x + c_2 \sin x$ is a real vector space.

(e) Every vector space is nonempty.

Exercise 1.20. *Suppose $A \in M_{m\times n}(\mathbb{C})$ satisfies $A\mathbf{x} \in \mathbb{R}^m$ for every column vector $\mathbf{x} \in \mathbb{R}^n$. Prove $A \in M_{m\times n}(\mathbb{R})$.*

Exercise 1.21. *Suppose $A \in M_{m\times n}(\mathbb{C})$ satisfies $A\mathbf{x} \in \mathbb{R}^m$ for every column vector $\mathbf{x} \in \mathbb{C}^n$. Prove $A = 0$.*

Exercise 1.22. *Prove vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent if and only if $\mathbf{v}_1 \neq \mathbf{0}$ and for every j, $1 \leq j < n$, the vector $\mathbf{v}_j$ is not a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}$.*

Exercise 1.23. *Suppose $L : V \to W$ is a linear transformation. Suppose $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ are such that $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ are linearly independent. Prove $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ are linearly independent. Is the converse true?*

1.7 Challenge Problems

Exercise 1.24. *Prove that if an $n \times n$ matrix A commutes with every $n \times n$ matrix, then A is a diagonal matrix all of whose diagonal entries are the same:*

$$A = \begin{pmatrix} c & 0 & \cdots & 0 & 0 \\ 0 & c & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & c & 0 \\ 0 & 0 & \cdots & 0 & c \end{pmatrix}.$$

Exercise 1.25. *Suppose V is a nonempty set of elements, called vectors. Assume V is equipped with a vector addition and a scalar multiplication, and that V is closed under both operations. The following list includes all properties of a vector space.*

(a) Associativity of Addition

(b) Scalar multiplication is associative

(c) Commutativity of Addition

(d) Additive Identity

(e) Additive Inverse

(f) Additive Inverse

(g) Distributivity of Scalar multiplication over scalar addition

(h) Multiplicative Identity

(i) Distributivity of Scalar multiplication over vector addition

For each of these properties, either provide an example of a set that satisfies all of the other properties but that specific property, or show that property follows form the rest of the properties.

Week 2

2.1 Review of Inverse Functions

All sets in this section are arbitrary. In other words, we do not assume sets are vector spaces or functions are linear transformations of vector spaces.

Definition 2.1. Let A be an arbitrary nonempty set. The function $I_A : A \to A$ defined by $I_A(x) = x$ for all $x \in A$ is called the **identity** function of A. Given a function $f : A \to B$, we say $g : B \to A$ is a **left inverse** for f iff $g \circ f = I_A$. If such a function g exists we say f is **left invertible**. We say $h : B \to A$ is a right inverse for f iff $f \circ h = I_B$. If such a function h exists we say f is **right invertible**. We say f is **invertible** iff it has a left and a right inverse.

Theorem 2.1. *Let $f : A \to B$ be a function between arbitrary sets. Then,*

(a) f is left invertible if and only if f is one-to-one.

(b) f is right invertible if and only if f is onto.

(c) f is invertible if and only if for every $b \in B$, the equation $f(x) = b$ has a unique solution $x \in A$.

Theorem 2.2. *Suppose $f : A \to B$ is right and left invertible (i.e. invertible). Then, f has a unique left inverse, a unique right inverse and these two inverses are equal.*

Remark 2.1. The left and right inverse of an invertible function f is denoted by f^{-1}.

Theorem 2.3. *Suppose $f : A \to B$ and $g : B \to C$ are invertible functions of arbitrary sets. Then, $g \circ f$ is invertible and $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. Furthermore, $f^{-1} : B \to A$ is invertible and $(f^{-1})^{-1} = f$.*

2.2 Inverse Functions and Linear Transformations

Definition 2.2. An $n \times n$ matrix is called the **identity matrix**, denoted by I_n or I, iff all of its main diagonal entries are 1 and the rest of its entries are all zero.

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

Theorem 2.4. *Suppose $L : V \to W$ is a linear transformation of vector spaces. If L is invertible, then L^{-1} is linear.*

Definition 2.3. An invertible linear transformation of vector spaces is called an **isomorphism**. If there is an isomorphism $L : X \to Y$, the we say X and Y are **isomorphic**. This is written as $X \cong Y$.

Example 2.1. The linear transformation $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ defined by rotation about the origin with angle θ is invertible with $R_\theta^{-1} = R_{-\theta}$.

Recall that all linear transformations $L : \mathbb{F}^n \to \mathbb{F}^m$ are given by $L(\mathbf{v}) = A\mathbf{v}$, where $A \in M_{m\times n}(\mathbb{F})$. We say L is the linear transformation corresponding to the matrix A.

Definition 2.4. We say a matrix is **left invertible** (resp. **right invertible**, **invertible**) iff its corresponding linear transformation is left invertible (resp. **right invertible**, **invertible**).

Definition 2.5. Given a matrix $A \in M_{m\times n}(\mathbb{F})$, we say $B \in M_{n\times m}(\mathbb{F})$ is a **left inverse** (resp. **right inverse**) of A iff $BA = I_m$ (resp. $AB = I_n$).

Theorem 2.5. *A matrix is left invertible (resp. right invertible) if and only if it has a left inverse (resp. right inverse). If a matrix is invertible, then the left and right inverse are equal and unique.*

Remark 2.2. The inverse of a matrix A, if it exists, is denoted by A^{-1}.

Example 2.2. Find all left and right inverses of the matrix $(1\ 2\ -1)$

Theorem 2.6. *Suppose A, B are invertible matrices for which AB is well-defined. Then,*

(a) AB is invertible, and $(AB)^{-1} = B^{-1}A^{-1}$.

(b) A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.

(c) A^{-1} is invertible and $(A^{-1})^{-1} = A$.

Theorem 2.7. *Suppose $L : V \to W$ is a linear transformtion. Then, the following are equivalent.*

(a) L is an isomorphism.

(b) If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *form a basis for* V*, then* $L(\mathbf{v}_1), \ldots, L(\mathbf{v}_n)$ *form a basis for* W*.*

(c) $L(\mathbf{v}_1), \ldots, L(\mathbf{v}_n)$ *form a basis for* W *for some basis* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *of* V*.*

Example 2.3. $L : \mathbb{F}^{n+1} \to \mathbb{P}_n$ defined by $L(a_0, a_1, \ldots, a_n) = a_0 + a_1 t + \cdots + a_n t^n$ is an isomorphism.

Theorem 2.8. *An* $m \times n$ *matrix* A *is invertible if and only if its columns form a basis for* $\mathbb{F}^m$*.*

2.3 Subspaces

Definition 2.6. Let V be a vector space over $\mathbb{F}$. A subset W of V is called a **subspace** of W iff W along with the vector addition and scalar multiplication of V is itself a vector space.

Theorem 2.9 (Subspace Criterion)**.** *A subset* W *of a vector space* V *is a subspace of* V *if and only if* W *satisfies the following:*

(a) $\mathbf{0}_V \in W$*.*

(b) For every $\mathbf{x}, \mathbf{y} \in W$ *and every* $c \in \mathbb{F}$*, we have* $\mathbf{x} + \mathbf{y}, c\mathbf{x} \in W$*. [We say* W *is closed under vector addition and scalar multiplication.]*

Corollary 2.1. If W is a subspace of V, then $\mathbf{0}_W = \mathbf{0}_V$.

Example 2.4. The following are examples of subspaces:

(a) For every vector space V, the sets $\{\mathbf{0}\}$ and V are subspaces of V.

(b) For every subset $\mathcal{A}$ of a vector space V, the set "$\operatorname{span} \mathcal{A}$" consisting of all vectors that are linear combinations of some vectors of $\mathcal{A}$, is a subspace of V.

(c) Given a linear transformation $L : V \to W$ of vector spaces, $\operatorname{Ker} L = \{\mathbf{v} \in V \mid L(\mathbf{v}) = \mathbf{0}\}$ is a subspace of V.

(d) The image of every linear transformation $L : V \to W$ is a subspace of W.

(e) Given a positive integer n and an open interval I, the set $C^n[I]$ consisting of all n times differentiable functions $f : I \to \mathbb{R}$ whose n-th derivative $f^{(n)} : I \to \mathbb{R}$ is continuous is a subspace of $C[I]$.

(f) Given an open interval I, the set $C^\infty[I]$ consisting of all infinitely differentiable functions $f : I \to \mathbb{R}$ is a subspace of $C^n[I]$ for every positive integer n.

Remark 2.3. The subspace $\{\mathbf{0}\}$, containing only the zero vector, is called the **trivial subspace** of V.

2.4 Systems of Linear Equations

Consider a system of linear equations

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{cases}$$

Here, a_{ij}'s and b_j's are known scalars in $\mathbb{F}$, and $x_1, \ldots, x_n$ are unknown variables. This system can be written in two other ways that are often useful:

$$\underbrace{\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}}_{\text{Coefficient matrix}} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

or

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b},$$

where $\mathbf{a}_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$.

Keeping all the known scalars in one place we can just work with the **augmented matrix** seen below:

$$\left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right)$$

2.5 Examples

Example 2.5. Let S and T be two subsets of a vector space V. Prove that $\operatorname{span} S = \operatorname{span} T$ if and only if $S \subseteq \operatorname{span} T$ and $T \subseteq \operatorname{span} S$.

Solution. Suppose $\operatorname{span} S = \operatorname{span} T$. By definition of span, $S \subseteq \operatorname{span} S = \operatorname{span} T$. Similarly $T \subseteq \operatorname{span} T = \operatorname{span} S$, as desired.

Now, suppose $S \subseteq \operatorname{span} T$, and $T \subseteq \operatorname{span} S$.

Every element $\mathbf{v} \in \operatorname{span} T$ is of the form $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$ for some $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in T$. Since $T \subseteq \operatorname{span} S$ and $\operatorname{span} S$ is a subspace, $\mathbf{v} \in \operatorname{span} S$. Therefore, $\operatorname{span} T \subseteq \operatorname{span} S$. Similarly $\operatorname{span} S \subseteq \operatorname{span} T$. This implies $\operatorname{span} S = \operatorname{span} T$, as desired. $\square$

Example 2.6. Let V be a vector space. Suppose $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent vectors of V and $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$ are also linearly independent vectors of V. Prove that $\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w}_1, \dots, \mathbf{w}_m$ are linearly independent if and only if

$$\operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_n\} \cap \operatorname{span}\{\mathbf{w}_1, \dots, \mathbf{w}_m\} = \{\mathbf{0}\}.$$

Solution. For simplicity, let $U = \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, and $W = \operatorname{span}\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$.

$\Rightarrow$: Suppose $\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w}_1, \dots, \mathbf{w}_m$ are linearly independent and $\mathbf{x} \in U \cap W$. Thus $\mathbf{x} = \sum_{i=1}^{n} a_i \mathbf{v}_i = \sum_{j=1}^{m} b_j \mathbf{w}_j$, for some $a_i, b_j \in \mathbb{R}$. Therefore, $\sum_{i=1}^{n} a_i \mathbf{v}_i - \sum_{j=1}^{m} b_j \mathbf{w}_j = \mathbf{0}$. Since $\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{w}_1, \dots, \mathbf{w}_m$ are linearly independent, we must have $a_i = b_j = 0$ and thus $\mathbf{x} = \mathbf{0}$. On the other hand $\mathbf{0}$ is in any subspace. Therefore, $U \cap W = \{\mathbf{0}\}$.

$\Leftarrow$: Now assume $U \cap W = \{\mathbf{0}\}$. Suppose $\sum_{i=1}^{n} a_i \mathbf{v}_i + \sum_{j=1}^{m} b_j \mathbf{w}_j = \mathbf{0}$. This implies $\sum_{i=1}^{n} a_i \mathbf{v}_i = -\sum_{j=1}^{m} b_j \mathbf{w}_j \in U \cap W$, which implies $\sum_{i=1}^{n} a_i \mathbf{v}_i = -\sum_{j=1}^{m} b_j \mathbf{w}_j = \mathbf{0}$. Since $\mathbf{v}_1, \dots, \mathbf{v}_n$ and $\mathbf{w}_1, \dots, \mathbf{w}_m$ are linear independent we must have $a_i = b_j = 0$ for all i, j. This completes the proof. □

Example 2.7. Consider the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ given by $T(x, y) = (2x + y, 0, x - y)$. Find all linear transformations that are left or right inverse of T.

Solution. $S : \mathbb{R}^3 \to \mathbb{R}^2$ is a left inverse of T iff $S \circ T(x, y) = (x, y)$. This is equivalent to $S(2x+y, 0, x-y) = (x, y)$. Setting $a = 2x + y, b = x - y$ and solving for x, y we obtain $x = (a + b)/3, y = (a - 2b)/3$. Therefore, $S(a, 0, b) = (\frac{a+b}{3}, \frac{a-2b}{3})$. Since S is linear, we will have

$$S(a, c, b) = cS(0, 1, 0) + (\frac{a+b}{3}, \frac{a-2b}{3}).$$

Now, if we let $\mathbf{v} \in \mathbb{R}^2$ and define $S : \mathbb{R}^3 \to \mathbb{R}^2$ by

$$S(a, c, b) = c\mathbf{v} + (\frac{a+b}{3}, \frac{a-2b}{3}) \qquad (*)$$

Then, S is linear (why?) and is a left inverse of T. So, all left inverses of T are of the form $(*)$.

Note that $(0, 1, 0)$ is not in the range of T. Thus, T is not onto. By Theorem 2.1, T does not have a right inverse. □

Example 2.8. Give an example of two invertible matrices A, B for which $A + B$ is not invertible.

Solution. Let $A = I$ be the identity matrix, and $B = -I$. We know $A^2 = B^2 = I$, but $A + B = 0$ does not have an inverse. □

Example 2.9. Give an example of non-invertible matrices A, B, for which $A + B$ is invertible.

Solution. $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ is one such example. (Prove these work!) $\square$

Example 2.10. Suppose X is a subspace of a vector space V. Let $\mathbf{x} \in X$ and $\mathbf{y}$ be a vector in V that is not in X. Prove that for every nonzero scalar $c \in \mathbb{F}$, the vector $\mathbf{x} + c\mathbf{y}$ does not belong to X.

Solution. On the contrary assume $\mathbf{x} + c\mathbf{y} \in X$. Since X is closed under addition and scalar multiplication, $(\mathbf{x} + c\mathbf{y}) - \mathbf{x} \in X$. Therefore, $c\mathbf{y} \in X$. Since X is closed under scalar multiplication, $\frac{1}{c}c\mathbf{y} \in X$. Thus, $\mathbf{y} \in X$, which is a contradiction. $\square$

Example 2.11. Suppose A, B are matrices for which A and AB are both invertible. Prove that B is also invertible.

Solution. We have $B = IB = (A^{-1}A)B = A^{-1}(AB)$. Note that by assumption, both A^{-1} and AB are invertible. Thus, by Theorem2.6, their product B is also invertible. $\square$

Example 2.12. Suppose for two nonzero matrices A, B we have $AB = 0$. Can A have a right inverse? How about a left inverse?

Solution. A could have a right inverse. For example letting $A = (1\ 0)$ and $B = (0\ 1)^T$ we see that $AB = (0)$, and that $AA^T = (1)$. So, A^T is a right inverse of A.

A cannot have a left inverse. Suppose on the contrary $CA = I$ for some matrix C. Then $CAB = C0 = 0$. On the other hand $CAB = IB = B$. Thus, $B = 0$, which is a contradiction. $\square$

Example 2.13. Let X and Y are subspaces of a vector space V. Prove that $X + Y$ defined below is a subpace of V.

$$X + Y = \{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in X, \text{ and } \mathbf{y} \in Y\}.$$

Solution. We wil use the Subspace Criterion.

Since X and Y are subspaces, we have $\mathbf{0} \in X$ and $\mathbf{0} \in Y$. Thus $\mathbf{0} = \mathbf{0} + \mathbf{0} \in X + Y$.

Suppose $\mathbf{u}, \mathbf{v} \in X + Y$ and $c \in \mathbb{F}$. By definition, $\mathbf{u} = \mathbf{x}_1 + \mathbf{y}_1$ and $\mathbf{v} = \mathbf{x}_2 + \mathbf{y}_2$ for some vectors $\mathbf{x}_1, \mathbf{x}_2 \in X$ and $\mathbf{y}_1, \mathbf{y}_2 \in Y$. We have

$$\mathbf{u} + \mathbf{v} = (\mathbf{x}_1 + \mathbf{x}_2) + (\mathbf{y}_1 + \mathbf{y}_2), \text{ and } c\mathbf{u} = c\mathbf{x}_1 + c\mathbf{y}_1.$$

Since X and Y are subspaces, they are closed under vector addition and scalar multiplication. Thus, $\mathbf{x}_1 + \mathbf{x}_2, c\mathbf{x}_1 \in X$ and $\mathbf{y}_1 + \mathbf{y}_2, c\mathbf{y}_1 \in Y$. Thus, by definition, $\mathbf{u} + \mathbf{v}, c\mathbf{u} \in X + Y$ $\square$

Example 2.14. Prove that the second condition in the Subspace Criterion can be replaced by

$$\text{“For every } \mathbf{x}, \mathbf{y} \in W \text{ and } c \in \mathbb{F}, \text{ we have } \mathbf{x} + c\mathbf{y} \in W.\text{”}$$

Solution. Suppose W satisfies the first condition of the Subspace Criterion and the given condition above. We will show W satisfies the Subspace Criterion.

First, note that by assumption $\mathbf{0} \in W$. Let $\mathbf{x}, \mathbf{y} \in W$ and $c \in \mathbb{F}$. Since $\mathbf{0} \in W$, by given assumption, $\mathbf{0} + c\mathbf{x} \in W$. Thus, $c\mathbf{x} \in W$. Therefore, W is closed under scalar multiplication. By assumption, we can also see that $\mathbf{x} + \mathbf{y} = \mathbf{x} + 1\mathbf{y} \in W$. Thus, W is closed under vector addition. Therefore, W is a subspace of V.

On the other hand if W is a subspace of V, it clearly satisfies the given condition above, by definition of a vector space. □

2.6 Exercises

Exercise 2.1. *Prove that the set of $n \times n$ symmetric matrices is a subspace of $M_n(\mathbb{F})$. Prove that the set of antisymmetric matrices is a subspace of $M_n(\mathbb{F})$.*

Exercise 2.2. *Prove Theorem 2.7.*

Exercise 2.3. *Suppose for two matrices A, B, the product AB is invertible. Prove that A is right invertible and B is left invertible. By an example show that A and B may not be invertible.*

Exercise 2.4. *Is there a 2×2 matrix A for which $A + B$ is invertible for every 2×2 matrix B?*

Exercise 2.5. *Prove that the inverse of an invertible symetric matrix is also symmetric.*

Exercise 2.6. *Find all right and left inverses of the 3×1 matrix $(3\ 2\ -1)^T$.*

Exercise 2.7. *Let $T : V \to W$ be a linear transformation between vector spaces V and W. Suppose X is a subspace of V. Prove that the image of X under T, defined below, is a subspace of W:*

$$T(X) = \{T(\mathbf{x}) \mid \mathbf{x} \in X\}.$$

Exercise 2.8. *Let $T : V \to W$ be a linear transformation between vector spaces V and W. Suppose Y is a subspace of W. Prove that the pre-image of Y under T, defined below, is a subspace of V:*

$$T^{-1}(Y) = \{\mathbf{v} \in V \mid T(\mathbf{v}) \in Y\}.$$

Exercise 2.9. *What is the smallest subspace of $M_3(\mathbb{F})$ matrices that contains all upper triangular and all symmetric matrices? What is the largest subspace that is contained in both?*

Exercise 2.10. *In the Subspace Criterion we assume* $\mathbf{0} \in W$*. This condition seems to follow from the fact that* W *is closed under scalar multiplication, since* $0\mathbf{x} = \mathbf{0} \in W$*. Given that, can we drop the condition that* $\mathbf{0} \in W$*?*

Exercise 2.11. *Give an example of each of the following:*

(a) A vector space V*, a subset* W *of* V *that contains the zero vector and is closed under vector addition, but* W *is not a subspace of* V*.*

(b) A vector space V*, a subset* W *of* V *that contains the zero vector and is closed under scalar multiplication, but* W *is not a subspace of* V*.*

Exercise 2.12. *Prove that the relation "*$\cong$*" between vector spaces is an equivalence relation. In other words, prove that for every three vector spacex* X, Y, Z *over a field* $\mathbb{F}$ *we have the following:*

(a) $X \cong X$*.*

(b) If $X \cong Y$*, then* $Y \cong X$*.*

(c) If $X \cong Y$ *and* $Y \cong Z$*, then* $X \cong Z$*.*

Exercise 2.13. *Suppose* $f : A \to B$ *is a function between arbitrary sets. Prove that if* f *has a unique left inverse, or a unique right inverse, then it is invertible. Deduce that if* $A \in M_{m\times n}(\mathbb{F})$ *has a unique left inverse or a unique right inverse, then* A *is invertible.*

Exercise 2.14. *Prove that a linear transformation* $L : V \to W$ *is one-to-one if and only if* $L^{-1}(\mathbf{0}) = \{\mathbf{0}\}$*.*

Exercise 2.15. *Determine if each statement is true or false.*

(a) Suppose W *is a subspace of a vector space* V*. If for two vectors* $\mathbf{u}, \mathbf{v} \in V$ *we have* $\mathbf{u} + \mathbf{v} \in W$*, then* $\mathbf{u}, \mathbf{v} \in W$*.*

(b) Any system of linear equations with more variables than equations has a nontrivial solution.

(c) The solution sets to $A\mathbf{x} = \mathbf{b}$ *and* $A\mathbf{x} = \mathbf{c}$ *are either identical or disjoint.*

Exercise 2.16. *Prove every square matrix* A *can uniquely be written as* $A = B + C$*, where* B *is a symmetric matrix and* C *is an antisymmetric matrix.*

Exercise 2.17. *Determine if each of the following implies* W *is a subspace of* V*.*

(a) $\mathbf{0} \in W$ *and* $c\mathbf{x} + (1 - c)\mathbf{y} \in W$ *for every* $\mathbf{x}, \mathbf{y} \in W$ *and every* $\mathbf{c} \in \mathbb{F}$*.*

(b) If $\mathbf{x}, \mathbf{y} \in W$ *and* $c \in \mathbb{F}$*, then* $\mathbf{x} + c\mathbf{y} \in W$*.*

(c) $\mathbf{0} \in W$ *and* $c\mathbf{x} - c\mathbf{y} \in W$ *for every* $\mathbf{x}, \mathbf{y} \in W$ *and every* $c \in \mathbb{F}$*,*

Exercise 2.18. *Suppose* V *is a subset of* $\mathbb{R}^n$ *that is also a subspace of the complex vector space* $\mathbb{C}^n$*. Prove* $S = \{\mathbf{0}\}$*.*

Exercise 2.19. *Suppose W is a subspace of V satisfying the following:*

"For every $\mathbf{x}, \mathbf{y} \in V$, if $\mathbf{x} + \mathbf{y} \in W$, then $\mathbf{x} \in W$ or $\mathbf{y} \in W$."

Prove $W = V$.

Exercise 2.20. *Let V be the subset of $\mathbb{P}$ consisting of all even polynomials, i.e. polynomials of the form $a_0 + a_1t^2 + a_1t^4 + \cdots + a_nt^{2n}$, where n is a positive integer and $a_0, a_1, \ldots, a_n \in \mathbb{F}$ are scalars. Prove V is a proper subspace of $\mathbb{P}$ that is isomorphic to $\mathbb{P}$.*

Exercise 2.21. *Suppose $W_1 \subseteq W_2 \subseteq W_3 \subseteq \cdots$ is an infinite sequence of subspaces of a vector space V. Prove $\bigcup_{n=1}^{\infty} W_n$ is a subspace of V.*

2.7 Challenge Problems

Exercise 2.22. *Prove that if a vector space V is a union of n of its subspaces $W_1, \ldots W_n$, then $W_j = V$ for some j.*

Week 3

3.1 Echelon Form and Reduced Echelon Form

In order to solve a system of linear equations $A\mathbf{x} = \mathbf{b}$, we form the augmented matrix $(A|\mathbf{b})$. We then apply the following three row operations to obtain a matrix is echelon form.

- **Row Addition:** Adding a scalar multiple of a row to another row.
- **Row Interchange:** Interchanging two rows.
- **Row Multiplication:** Multiplying a row by a nonzero number.

Definition 3.1. A matrix is in **echelon form** iff it satisfies all of the following:

- All zero rows are at the bottom.
- The entries below the first nonzero entry of each row are all zero.
- The leading nonzero entry of each row is to the left of the leading nonzero entry of all rows below it.

Every leading nonzero entry of a row of a matrix in echelon form is called a **pivot entry**. The column of every pivot entry is called a **pivot column**. Every variable corresponding to a nonpivot column is called a **free variable**.

If in addition to the above, we also have the following two conditions:

- the first nonzero entry of each row is 1, and
- these 1's are the only nonzero entry of their column.

Then, we say the matrix is in **reduced (row) echelon form**.

Assume we apply these row operations to a matrix A. Each row operation can be seen as multiplying A by an invertible matrix from the left. Let $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ be the standard basis for $\mathbb{R}^n$. In other words, $\mathbf{e}_j$ is the vector whose j-th component is 1, and all of whose other components are zero. Suppose rows of A are

$\mathbf{a}_1, \ldots, \mathbf{a}_n$ from top to bottom. Then, the matrix obtained by adding $c\mathbf{a}_k$ to $\mathbf{a}_j$ is given by EA, where E is the matrix obtain from the identity matrix, by adding c times its k-th row to its j-th row:

$$E = \begin{pmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_j + c\mathbf{e}_k \\ \vdots \\ \mathbf{e}_n \end{pmatrix} \begin{array}{l} \\ \\ \leftarrow j\text{-th row} \\ \\ \\ \end{array}$$

The matrix obtained by interchanging $\mathbf{a}_j$ and $\mathbf{a}_k$ is EA, where E is obtained from the identity matrix by interchanging its j-th and k-th row.

$$E = \begin{pmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_j \\ \vdots \\ \mathbf{e}_k \\ \vdots \\ \mathbf{e}_n \end{pmatrix} \begin{array}{l} \\ \\ \leftarrow k\text{-th row} \\ \\ \leftarrow j\text{-th row} \\ \\ \\ \end{array}$$

The matrix obtained by scaling j-th row of A by a nonzero scalar c is EA, where E is the matrix obtained by scaling the j-th row of the identity matrix by the nonzero scalar c.

$$E = \begin{pmatrix} \mathbf{e}_1 \\ \vdots \\ c\ \mathbf{e}_j \\ \vdots \\ \mathbf{e}_n \end{pmatrix} \begin{array}{l} \\ \\ \leftarrow j\text{-th row} \\ \\ \\ \end{array}$$

Definition 3.2. Each of the above matrices E is called an **elementary matrix**. An elementary matrix is called a row interchange elementary matrix (resp. a row multiplication elementary matrix; a row replacement elementary matrix) if it corresponds to the appropriate row operation.

Remark 3.1. They are all invertible. Therefore, the equations $A\mathbf{x} = \mathbf{b}$ and $EA\mathbf{x} = E\mathbf{b}$ have the same solution set.

Theorem 3.1. *Every matrix can be turned into a matrix in reduced echelon form. This matrix in reduced echelon form is unique.*

Definition 3.3. Given a matrix $A \in M_{m\times n}(\mathbb{F})$ and a column vector $\mathbf{b} \in \mathbb{F}^m$, an equation $A\mathbf{x} = \mathbf{b}$ is called **inconsistent** iff it has no solution $\mathbf{x} \in \mathbb{F}^n$.

Theorem 3.2. *Consider the linear system* $A\mathbf{x} = \mathbf{b}$.

(a) This system is inconsistent if and only if the last column is a pivot column.

(b) The solution to $A\mathbf{x} = \mathbf{0}$ is unique if and only if there are no free variables.

Theorem 3.3. *Consider column vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{F}^n$ and the matrix $A = (\mathbf{v}_1 \cdots \mathbf{v}_m) \in M_{n\times m}(\mathbb{F})$.*

(a) $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly independent if and only if every column of A is a pivot column.

(b) $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are generating if and only if there is a pivot entry in every row of A.

(c) $\mathbf{v}_1, \ldots, \mathbf{v}_m$ form a basis for $\mathbb{F}^n$ if and only if there is a pivot entry in every row and every column of A.

Theorem 3.4. *(a) Any linear independent set of vectors of $\mathbb{F}^n$ contains at most n vectors.*

(b) Any generating set of vectors of $\mathbb{F}^n$ contains at least n vectors.

(c) Every basis of $\mathbb{F}^n$ contains precisely n vectors.

Theorem 3.5. *Any two bases for a vector space V have the same number of elements.*

Theorem 3.6. *A matrix is invertible iff there is a pivot entry in every row and every column of its echelon form. Consequently every invertible matrix must be a square matrix.*

Theorem 3.7. *Suppose a square matrix has a right or left inverse. Then, it is invertible.*

3.2 Evaluating the Inverse of a Matrix

By Theorem 3.6 a matrix $A \in M_n(\mathbb{F})$ is invertible iff there is a pivot entry in every row and every column of an echelon form A_e of A. Since the number of rows and columns of A are the same, the entries on the main diagonal of A are all pivot entries. Reducing this further to obtain the reduced echelon form of A, we obtain the identity matrix iff A is invertible. In other words, A is invertible iff its reduced echelon form is the identity matrix.

Assume A is invertible. Since each row operation corresponds to a multiplication by an elementary matrix from the left, we obtain the following for some elementary matrices $E_1, \ldots, E_m$:

$$E_m \cdots E_1 A = I \Rightarrow A^{-1} = E_m \cdots E_1 \text{ and } A = E_1^{-1} \cdots E_m^{-1}$$

If we start with the augmented matrix $(A|I)$ and apply row operations we obtain the following:

$$(A|I) \to (E_1 A|E_1) \to (E_2 E_1 A|E_2 E_1) \to \cdots (E_m \cdots E_1 A|E_m \cdots E_1) = (I|A^{-1})$$

In other words, to obtain the inverse of a square matrix A:

- Form the augmented matrix $(A|I)$.

- Apply row operations until you obtain $(I|B)$. (If one of the columns of A is not a pivot column, then A is not invertible.)
- Conclude that $B = A^{-1}$.

Example 3.1. Evaluate the inverse of A.

$$A = \begin{pmatrix} 1 & 3 & 3 \\ 1 & 4 & 3 \\ 1 & 3 & 4 \end{pmatrix}$$

Theorem 3.8. *Suppose V is a vector space with a basis of size n. Then:*

(a) Any linearly independent set in V has at most n elements.

(b) Any generating set in V has at least n elements.

(c) Any basis of V has precisely n elements.

Definition 3.4. For a vector space V, we define its **dimension**, denoted by $\dim V$, to be the size of a basis of V. If V has no finite basis we write $\dim V = \infty$.

Even though many of the topics discussed in this class can be generalized to infinite-dimensional vector spaces, we will assume every vector space is finite-dimensional, unless otherwise stated.

Theorem 3.9. *Let V be a (finite-dimensional) vector space. Any linearly independent set of vectors can be completed to a basis.*

Theorem 3.10. *Suppose V is a vector space with $\dim V = n < \infty$. Then the dimension of every subspace W of V does not exceed n. Furthermore, if $\dim V = \dim W$, then $V = W$.*

3.3 General Solutions to Linear Systems

Given $A \in M_{m\times n}(\mathbb{F})$ and a column vector $\mathbf{b} \in \mathbb{F}^m$ we are interested in solutions $\mathbf{x} \in \mathbb{F}^n$ to $A\mathbf{x} = \mathbf{b}$. When $\mathbf{b} = \mathbf{0}$, we say the system $A\mathbf{x} = \mathbf{0}$ is **homogeneous**.

Theorem 3.11. *Let H be the set of solutions to the homogeneous system $A\mathbf{x} = \mathbf{0}$, and $\mathbf{x}_p$ be a particular solution to $A\mathbf{x} = \mathbf{b}$ (i.e. $A\mathbf{x}_p = \mathbf{b}$). Then, the set of solutions to $A\mathbf{x} = \mathbf{b}$ is given by*

$$\{\mathbf{x}_p + \mathbf{x}_h \mid \mathbf{x}_h \in H\}.$$

Example 3.2. Find a system of linear equations $A\mathbf{x} = \mathbf{b}$ whose solution set is given by

$$\mathbf{x} = \begin{pmatrix} 1+2s \\ -1-s \\ 2+s \end{pmatrix}, \text{ with } s \in \mathbb{F}.$$

3.4 Examples

Example 3.3. Prove that if V is a vector space of dimension $n \geq 2$, then V can be written as the union of its proper subspaces.

Solution. Let W be the union of all proper subspaces of V. Since every element of W is also in V, we have $W \subseteq V$. Let $\mathbf{v} \in V$. The dimension of the subspace $X = \operatorname{span}\{\mathbf{v}\}$ is at most one and since $\dim V \geq 2$, X is a proper subspace of V. Therefore, $\mathbf{v} \in W$. Thus, $V \subseteq W$, which implies $V = W$, as desired. $\square$

Example 3.4. Find a system of linear equations on three variables x_1, x_2, x_3 whose general solution is given as

$$\begin{cases} x_1 = 3t + 2 \\ x_2 = -t + 1 \\ x_3 = t - 2 \end{cases}$$

Here, t is a free variable.

Solution. Since the general solution to the corresponding homogeneous system is $x_1 = 3t, x_2 = -t, x_3 = t$ which has one free variable, the reduced echelon form must have two pivot column. So, we may assume the reduce echelon form is of the following form:

$$\begin{pmatrix} 1 & 0 & a \\ 0 & 1 & b \end{pmatrix} \qquad (*)$$

Multiplying this matrix with $(3 \ -1 \ 1)^T$ and setting that equal to the zero vector we obtain $3 + a = 0$ and $-1 + b = 0$. Thus, the corresponding homogeneous system is

$$\begin{cases} x_1 - 3x_3 = 0 \\ x_2 + x_3 = 0 \end{cases}$$

Now, we will need to make sure $x_1 = 2, x_2 = 1, x_3 = -2$ is a solution to the system. Multiplying the above matrix $(*)$ with $(2 \ 1 \ -2)^T$ we obtain $(8 \ -1 \ 0)$. Thus, one such system is given as

$$\boxed{\begin{cases} x_1 - 3x_3 = 8 \\ x_2 + x_3 = -1 \end{cases}}$$

$\square$

Example 3.5. Prove the inverse of every elementary matrix is also an elementary matrix.

Solution. We will prove this for all three different types of elementary matrices.

Assume E is a row addition elementary matrix. Suppose E adds c times the k-th row to the j-th row. This means the j-th row of E is $\mathbf{e}_j + c\mathbf{e}_k$, and its ℓ-th row is $\mathbf{e}_\ell$ for every $\ell \neq j$. Consider the elementary matrix D that adds $-c$ times the k-th row to the j-th row. When multiplying DE, the j-th row becomes $(\mathbf{e}_j + c\mathbf{e}_k) - c\mathbf{e}_k = \mathbf{e}_j$, and the rest of the rows remain unchanged. Thus $DE = I$. So, E has a left inverse, and since it is a square matrix, by a theorem, it is invertible. Thus $D = E^{-1}$.

Assume E is a row exchange elementary matrix. Suppose E swaps the k-th and j-th rows. This means the k-th row of E is $\mathbf{e}_j$, the j-th row of E is $\mathbf{e}_k$ and its ℓ-th row is $\mathbf{e}_\ell$ for every $\ell \neq j.k$. The matrix EE is obtained by swapping the k-th and j-th rows of E. Thus, $EE = I$, which means E is its own inverse.

Assume E is a row scaling elementary matrix. Assume E multiplies the j-th row by a nonzero scalar c. The j-th row of E is $c\mathbf{e}_j$ and its k-th row is $\mathbf{e}_k$ for every $k \neq j$. Let D be the elementary matrix that multiplies the j-th row by a factor of c^{-1}. We see that $DE = I$. Therefore, by the argument made above, E is invertible, and thus $D = E^{-1}$. □

Example 3.6. Using row operations find the inverse of the following matrix

$$\begin{pmatrix} 1 & 3 & 3 \\ 1 & 4 & 3 \\ 1 & 3 & 4 \end{pmatrix}.$$

Solution. Let A be the given matrix. We will row reduce the matrix $(A|I)$.

$$\left(\begin{array}{ccc|ccc} 1 & 3 & 3 & 1 & 0 & 0 \\ 1 & 4 & 3 & 0 & 1 & 0 \\ 1 & 3 & 4 & 0 & 0 & 1 \end{array}\right) \xrightarrow{R_2-R_1, R_3-R_1} \left(\begin{array}{ccc|ccc} 1 & 3 & 3 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{R_1-3R_2} \left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 4 & -3 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{array}\right) \xrightarrow{R_1-3R_3} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 7 & -3 & -3 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{array}\right)$$

Therefore, the inverse of the given matrix is

$$\begin{pmatrix} 7 & -3 & -3 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

□

Example 3.7. Suppose A, B are two matrices for which AB and BA are both invertible. Prove that A and B must be of the same size.

Solution. Suppose sizes of A and B are $m \times n$ and $n \times m$, respectively. On the contrary assume $m \neq n$. WLOG we may assume $n < m$. Since columns of B are in $\mathbb{F}^n$ and B has m columns, and $m > n$, by a theorem, columns of B are linearly dependent. Therefore, there is a nonzero vector $\mathbf{v} \in \mathbb{F}^m$ for which $B\mathbf{v} = \mathbf{0}$. Multiplying both sides by A from the left, we obtain $AB\mathbf{v} = \mathbf{0}$. Multiplying by $(AB)^{-1}$ we obtain $\mathbf{v} = \mathbf{0}$, which is a contradiction. Therefore, $m = n$, as desired. □

Example 3.8. Determine (with full justification) if each statement is true or false for all matrices A, B with A being an invertible matrix. Assume the product AB is defined.

(a) If all columns of a matrix B are linearly independent, then all columns of AB are also linearly independent.

(b) If all rows of a matrix B are linearly independent, then all rows of AB are also linearly independent.

Solution. Note that in order to show the columns of a matrix X are linearly independent, we need to show if for a vector $\mathbf{v}$ we have $X\mathbf{v} = \mathbf{0}$, then $\mathbf{v} = \mathbf{0}$.

(a) This is true. Suppose $AB\mathbf{v} = \mathbf{0}$ for a vector $\mathbf{v}$. Since A is invertible, $A^{-1}AB\mathbf{v} = A^{-1}\mathbf{0} = \mathbf{0}$. Therefore, $B\mathbf{v} = \mathbf{0}$. Since columns of B are linearly independent, we must have $\mathbf{v} = \mathbf{0}$.

(b) This is true. We will work with the transpose of AB and use its columns, instead. Suppose $(AB)^T\mathbf{v} = \mathbf{0}$. We know $(AB)^T = B^T A^T$. Thus, $B^T A^T \mathbf{v} = \mathbf{0}$. Since columns of B^T (which are rows of B) are linearly independent, we must have $A^T\mathbf{v} = \mathbf{0}$. Since A is invertible, so is A^T, and thus $\mathbf{v} = \mathbf{0}$, as desired. □

Example 3.9. Prove that if a system of linear equations $A\mathbf{x} = \mathbf{b}$ has more than one solution, then it has infinitely many solutions.

Solution. Let $\mathbf{x}_p$ be a solution to $A\mathbf{x} = \mathbf{b}$. By Theorem 3.11, the solution set to this system is given by

$$A = \{\mathbf{x}_p + \mathbf{x}_h \mid \mathbf{x}_h \in H\},$$

where H is the solution set to the homogeneous system $A\mathbf{x} = \mathbf{0}$. Since A has at least two elements, H contains a nontrivial element $\mathbf{y}$. Since $A\mathbf{y} = \mathbf{0}$, for every $c \in \mathbb{F}$ we have $A(c\mathbf{y}) = cA\mathbf{y} = \mathbf{0}$. Therefore, $c\mathbf{y} \in H$. Since $\mathbf{y} \neq \mathbf{0}$, for every two distinct scalars a, b we have $a\mathbf{y} \neq b\mathbf{y}$. Thus, $\mathbf{x}_p + a\mathbf{y} \neq \mathbf{x}_p + b\mathbf{y}$. Therefore, there are infinitely many vectors of the form $\mathbf{x}_p + c\mathbf{y}$ in A, which means A is an infinite set. □

Example 3.10. Can 4 vectors in $\mathbb{P}_4$ be linearly independent? Can they be generating?

Solution. First, note that $1, t, t^2, t^3$ are four linearly independent polynomials in $\mathbb{P}_4$. So, the answer to the first question is yes!

By an example, $\{1, t, t^2, t^3, t^4\}$ is a basis for $\mathbb{P}_4$. Therefore, $\dim \mathbb{P}_4 = 5$. Thus, no four polynomials can be generating by Theorem 3.8(b). □

3.5 Exercises

Exercise 3.1. *Determine which of the following vectors form a basis for the appropriate* $\mathbb{F}^n$.

(a) $(1,0,1),(1,1,2),(-1,-2,-3)$.

(b) $(1,0),(2,3),(1,1)$.

(c) $(1,0,0),(0,1,1),(0,1,2)$.

Exercise 3.2. *Show that a matrix* A *has a left inverse if and only if* $A\mathbf{x}=\mathbf{0}$ *has a unique solution for* $\mathbf{x}$.

Exercise 3.3. *Find the inverse of each matrix or show the matrix is not invertible.*

$$\begin{pmatrix} 4 & 3 & -1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 3 \\ 2 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & n-1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & n \end{pmatrix}.$$

Exercise 3.4. *Determine (with full justification) if each statement is true or false for all matrices* A,B *with* B *being an invertible matrix. Assume the product* AB *is defined.*

(a) If all columns of a matrix A *span* $\mathbb{F}^n$, *then all columns of* AB *also span* $\mathbb{F}^n$.

(b) If all rows of a matrix A *are linearly independent, then all rows of* AB *are also linearly independent.*

Exercise 3.5. *Suppose for a square matrix* A, *we know* A^2 *is invertible. Prove that* A *is also invertible.*

Exercise 3.6. *Find a system of linear equations whose solution is given by*

$$x_1 = 2+3t+s, x_2 = 5-t-s, x_3 = t, \text{ with } r,s \in \mathbb{R}.$$

Exercise 3.7. *Let* n *be a positive integer. Suppose* $f_j \in \mathbb{P}_n$ *is a nonzero polynomial of degree* j *for* $j = 0,1,\ldots,n$. *Prove that* $\{f_0,f_1,\ldots,f_n\}$ *is a basis for* $\mathbb{P}_n$.

Exercise 3.8. *Suppose* A,B *are square matrices for which* AB *is invertible. Prove that both* A *and* B *are invertible.*

Definition 3.5. A **hyperplane** in $\mathbb{F}^n$ is a subspace of dimension $n-1$.

Exercise 3.9. *Suppose* $(a_1,\ldots,a_n) \in \mathbb{F}^n$ *is a given nonzero vector. Prove the set of all* $(x_1,\ldots,x_n) \in \mathbb{F}^n$ *given by* $a_1x_1+\cdots+a_nx_n = 0$ *is an* $(n-1)$*-dimensional subspace of* $\mathbb{F}^n$.

Exercise 3.10. *Suppose* X,Y *are subspaces of a vector space* V *for which* $\dim X + \dim Y > \dim V$. *Prove* $X \cap Y \neq \{\mathbf{0}\}$.

Exercise 3.11. *Suppose* $A \in M_{m\times n}(\mathbb{R})$. *Since* $\mathbb{R}$ *is a subset of* $\mathbb{C}$, *we can also consider* A *as a matrix in* $M_{m\times n}(\mathbb{C})$. *Prove that* $\dim \operatorname{Ker} A$ *is the same whether* A *is assumed to be a matrix in* $M_{m\times n}(\mathbb{R})$ *or a matrix in* $M_{m\times n}(\mathbb{C})$. *Do the same for* $\dim \operatorname{Col} A$.

Exercise 3.12. *Let* V *and* W *be finite dimensional vector spaces of dimensions* m *and* n, *respectively. Prove that the dimension of* $\mathcal{L}(V, W)$ *is* mn.

Exercise 3.13. *Suppose* V *and* W *are vector spaces over the same field* $\mathbb{F}$. *Prove that their Cartesian product* $V \times W$ *along with vector addition and scalar multiplication defined below is also a vector space:*

$$(\mathbf{v}_1, \mathbf{w}_1) + (\mathbf{v}_2, \mathbf{w}_2) = (\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}_1 + \mathbf{w}_2), \text{ and } c(\mathbf{v}_1, \mathbf{w}_1) = (c\mathbf{v}_1, c\mathbf{w}_1) \text{ for all } \mathbf{v}_1, \mathbf{v}_2 \in V, \mathbf{w}_1, \mathbf{w}_2 \in W.$$

Prove that if both V *and* W *are finite dimensional, then* $\dim(V \times W) = \dim V = \dim W$.

Exercise 3.14. *Suppose* V *is an* n*-dimensional complex vector space. Prove that if we restrict the set of scalars to the real numbers,* V *with the same vector addition and scalar multiplication is a real vector space with dimension* $2n$.

3.6 Challenge Problems

Exercise 3.15. *Prove that the subspace of* $C[\mathbb{R}]$ *generated by* $\sin t, \sin(2t), \sin(3t), \ldots$ *is infinite dimensional.*

Exercise 3.16. *If we allow only rational numbers to be scalars we can turn* $\mathbb{R}$ *into a vector space over* $\mathbb{Q}$, *i.e. the vectors are real numbers and scalars are rational numbers. Prove that* $\mathbb{R}$ *is an infinite dimensional vector space over* $\mathbb{Q}$.

Week 4

4.1 Fundamental Subspaces of a Matrix

Definition 4.1. Given a linear transformation $L : V \to W$, we define its **kernel** or **null** space by:

$$\mathrm{Ker}\, L = \mathrm{Null}\, L = \{\mathbf{v} \in V \mid \}.$$

The **range** or **image** of L is defined as

$$RanL = L(V) = \{L(\mathbf{v}) \mid \mathbf{v} \in V\}.$$

Given a matrix A, the kernel or null space of A, denoted by $\mathrm{Ker}\,(A) = \mathrm{Null}\,(A)$ is defined to be the kernel of the linear transformation associated to A. The range or image of A is the range of the linear transformation associated with A.

Note that $Ran(A)$ is sometimes denoted by $\mathrm{Col}\,(A)$. The range of A^T is called the **row space** of A and is denoted by $\mathrm{Row}\,(A)$.

The four vector spaces $\mathrm{Ker}\,(A), \mathrm{Ker}\,(A^T), \mathrm{Col}\,(A)$, and $\mathrm{Row}\,(A)$ are called four fundamental subspaces corresponding to a matrix A.

Theorem 4.1. *Let A be a matrix, and A_e be a matrix in echelon form obtained from A by performing row operations. Then,*

(a) The nonzero rows of A_e form a basis for the $\mathrm{Row}\,(A)$.

(b) The pivot columns of A form a basis for $\mathrm{Col}\,(A)$.

Definition 4.2. The **rank** of a matrix is the dimension of its column space.

Theorem 4.2. *For every matrix A, we have* $\mathrm{rank}\, A = \mathrm{rank}\, A^T$.

Theorem 4.3 (Rank-Nullity Theorem)**.** *Let $L : V \to W$ be a linear transformation of finite-dimensional vector spaces. Then,*

$$\dim \mathrm{Ker}\, L + \dim L(V) = \dim V.$$

Furthermore, if $A \in M_{m\times n}(\mathbb{F})$, *then*

$$\dim \operatorname{Ker} A + \operatorname{rank} A = n.$$

Definition 4.3. The **nullity** of a matrix is the dimension of its null space.

Theorem 4.4. *Suppose* $A \in M_{m\times n}(\mathbb{F})$. *Then, the equation* $A\mathbf{x} = \mathbf{b}$ *has a solution for every* $\mathbf{b} \in \mathbb{F}^m$ *if and only if the equation* $A^T\mathbf{x} = \mathbf{0}$ *has a unique solution.*

Theorem 3.9 indicates that every set of linear independent vectors can be completed to a basis, however its proof did not give us a clear algorithm as to how we can do that. What we discussed above gives us a clear algorithm for how that can be done. First, we will consider the case where the vector space is $\mathbb{F}^n$. In order to complete linear independent vectors $\mathbf{v}_1, \dots, \mathbf{v}_m \in \mathbb{F}^n$ to a basis, we create a matrix whose rows are $\mathbf{v}_1, \dots, \mathbf{v}_m$. We row reduce this matrix and insert appropriate elements of the standard basis to obtain a square matrix in standard form. Since rows form a basis for $\mathbb{F}^n$, the original vectors $\mathbf{v}_1, \dots, \mathbf{v}_m$ along with the newly added vectors from the standard basis form a basis for $\mathbb{F}^n$.
To replicate this for a (finite-dimensional) vector space V, first write down an isomorphism between V and $\mathbb{F}^n$, where $n = \dim V$. Then, repeat the process above and use the fact that under isomorphisms every basis is mapped to a basis.

4.2 Representation of Transformations in Arbitrary Bases

Definition 4.4. An **ordered basis** for a vector space is a basis with a specified order. In other words, an ordered basis for a vector space V is an n-tuple $\mathcal{B} = (\mathbf{b}_1, \dots, \mathbf{b}_n)$, where $\mathbf{b}_1, \dots, \mathbf{b}_n$ form a basis for V.

Definition 4.5. Suppose V is a vector space with an ordered basis $\mathcal{B} = (\mathbf{b}_1, \dots, \mathbf{b}_n)$. We know every vector $\mathbf{v} \in V$ has a unique representation as $\mathbf{v} = \sum_{j=1}^{n} c_j \mathbf{b}_j$. The coefficients $c_1, \dots, c_n$ are called **coordinates** of $\mathbf{v}$ in basis $\mathcal{B}$. This is written as

$$[\mathbf{v}]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}.$$

Theorem 4.5. *Let* $\mathcal{B}$ *be an ordered basis for an* n*-dimensional vector space* V. *Then,* $L : V \to \mathbb{F}^n$ *defined by* $L(\mathbf{v}) = [\mathbf{v}]_{\mathcal{B}}$ *is an isomorphism.*

Theorem 4.6. *Let* $\mathcal{A} = (\mathbf{a}_1, \dots, \mathbf{a}_n)$ *and* $\mathcal{B}$ *be ordered bases for vector spaces* V *and* W, *respectively. Suppose* $T : V \to W$ *be a linear transformation. Then, there is a unique matrix* A *for which* $[T(\mathbf{v})]_{\mathcal{B}} = A[\mathbf{v}]_{\mathcal{A}}$. *Furthermore, this matrix is given by*

$$A = ([T(\mathbf{a}_1)]_{\mathcal{B}} \cdots T(\mathbf{a}_n)]_{\mathcal{B}})$$

Definition 4.6. The unique matrix A in the theorem above is called the **matrix of** T **relative to ordered bases** $\mathcal{A}$ **and** $\mathcal{B}$. This matrix is denoted by $[T]_{\mathcal{BA}}$.

Theorem 4.7. *Suppose $T : V \to W$ and $S : W \to U$ are linear transformations between vector spaces on the same field $\mathbb{F}$. Let $\mathcal{A}, \mathcal{B}$, and $\mathcal{C}$ be ordered bases for vector spaces V, W, and U, respectively. Then,*

$$[S \circ T]_{\mathcal{C}\mathcal{A}} = [S]_{\mathcal{C}\mathcal{B}}[T]_{\mathcal{B}\mathcal{A}}$$

Corollary 4.1. Suppose $T : V \to W$ is an isomorphism between vector spaces. Let $\mathcal{A}$ and $\mathcal{B}$ be ordered bases for V and W, respectively. Then,

$$[T^{-1}]_{\mathcal{A}\mathcal{B}} = [T]^{-1}_{\mathcal{B}\mathcal{A}}.$$

Definition 4.7. Given a vector space V and two ordered bases $\mathcal{A}$ and $\mathcal{B}$ of V, the matrix $[I_V]_{\mathcal{B}\mathcal{A}}$ is called the **change of coordinate matrix** from $\mathcal{A}$ to $\mathcal{B}$.

Example 4.1. Find the change of basis matrix from the basis $\mathcal{A} = (1+t, t-1)$ to the basis $\mathcal{B} = (2-3t, 1-3t)$ for $\mathbb{P}_1$. Assume $\mathcal{A}$ and $\mathcal{B}$ are bases for $\mathbb{P}_1$.

Definition 4.8. Two square matrices A and B of the same size are said to be **similar** iff there is an invertible matrix P for which $A = PBP^{-1}$.

Theorem 4.8. *Suppose $\mathcal{A}$ and $\mathcal{B}$ are two ordered bases for a vector space V and $T : V \to V$ is a linear transformation. Then $[T]_{\mathcal{B}\mathcal{B}}$ and $[T]_{\mathcal{A}\mathcal{A}}$ are similar matrices.*

4.3 Examples

Example 4.2. Suppose the dimension of the kernel of a 31×17 matrix A is 12. Find the dimension of all of its four fundamental subspaces.

Solution. By the Rank-Nullity Theorem, we have $\dim \operatorname{Ker} A + \operatorname{rank} A = 17$. Therefore, $\operatorname{rank} A = 5$. Therefore, $\dim \operatorname{Col}(A) = \dim \operatorname{Row}(A) = 5$. Applying the Rank-Nullity Theorem to A^T we obtain $\dim \operatorname{Ker} A^T + \operatorname{rank} A^T = 31$. Since $\operatorname{rank} A^T = 5$, we have $\dim \operatorname{Ker} A^T = 26$. $\square$

Example 4.3. Show the following vectors are linearly independent and complete them to a basis of $\mathbb{F}^4$

$$(1, 2, -1, 3)^T, (1, 2, 1, 0)^T, (0, 1, 2, 0)^T$$

Solution. We will form a matrix whose rows are the given vectors.

$$\begin{pmatrix} 1 & 2 & -1 & 3 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 0 \end{pmatrix} \xrightarrow{R_2 - R_1} \begin{pmatrix} 1 & 2 & -1 & 3 \\ 0 & 0 & 2 & -3 \\ 0 & 1 & 2 & 0 \end{pmatrix} \xrightarrow{R_2 \leftrightarrow R_3} \begin{pmatrix} 1 & 2 & -1 & 3 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 2 & -3 \end{pmatrix}$$

This matrix is in echelon form with 3 pivot entries. Thus, the three given vectors are linearly independent. Furthermore, placing $\mathbf{e}_4$ in the last row creates a matrix in echelon form with four pivot entries. Therefore, the following vectors form a basis for $\mathbb{F}^4$.

$$(1, 2, -1, 3)^T, (1, 2, 1, 0)^T, (0, 1, 2, 0)^T, \mathbf{e}_4$$

□

Example 4.4. Consider $n-1$ linearly independent vectors $\mathbf{v}_1, \dots, \mathbf{v}_{n-1} \in \mathbb{F}^n$. Suppose the only vector in the standard basis that completes these $n-1$ vectors to a basis for $\mathbb{F}^n$ is $\mathbf{e}_n$. Prove that the n-th entry of every $\mathbf{v}_j$, $j = 1, \dots, n-1$ is zero.

Solution. For simplicity let $W = \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_{n-1}\}$. Note that since $\mathbf{v}_j$'s are linearly independent, $\dim W = n-1$.

First, we claim that for every j with $1 \le j < n$, we have $\mathbf{e}_j \in W$. We will prove that by contradiction. Suppose $\mathbf{e}_j \notin W$. By Exercise 1.9 the set $\mathcal{A} = \{\mathbf{v}_1, \dots, \mathbf{v}_{n-1}, \mathbf{e}_j\}$ is linearly independent. Since $\dim \mathbb{F}^n = n$, by Theorem 3.8, the set $\mathcal{A}$ is a basis for $\mathbb{F}^n$, which is a contradiction. Therefore, for every j with $1 \le j < n$, we have $\mathbf{e}_j \in W$. Therefore, $\mathbf{e}_1, \dots, \mathbf{e}_{n-1}$. On the other hand since $\dim W = n-1$, by Theorem 3.8, $\mathbf{e}_1, \dots, \mathbf{e}_{n-1}$ is a basis for W. Therefore, $W = \operatorname{span}\{\mathbf{e}_1, \dots, \mathbf{e}_{n-1}\}$. However the n-th coordinate of every vector that is in the span of $\mathbf{e}_1, \dots, \mathbf{e}_{n-1}$ is zero. Thus, the n-th coordinates of $\mathbf{v}_j$'s are all zero. □

Example 4.5. Consider the ordered bases $\mathcal{A} = (1+t, 2+3t)$ and $\mathcal{B} = (2-t, 1+2t)$ of $\mathbb{P}_1$. Find the change of coordinate matrix from $\mathcal{B}$ to $\mathcal{A}$.

Solution. Consider the standard ordered basis $\mathcal{S} = (1, t)$ of $\mathbb{P}_1$. We know

$$[I]_{\mathcal{S}\mathcal{A}} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}, \text{ and } [I]_{\mathcal{S}\mathcal{B}} = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$$

Therefore,

$$[I]_{\mathcal{A}\mathcal{B}} = [I]_{\mathcal{A}\mathcal{S}}[I]_{\mathcal{S}\mathcal{B}} = [I]_{\mathcal{S}\mathcal{A}}^{-1}[I]_{\mathcal{S}\mathcal{B}} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$$

The answer is $\begin{pmatrix} 8 & -1 \\ -3 & 1 \end{pmatrix}$. □

Example 4.6. Write down the coordinate vector of $3t^2 - t + 1$ with respect to each given basis.

(a) $\mathcal{A} = (1, t, t^2)$.

(b) $\mathcal{B} = (t, 1, t^2)$.

(c) $\mathcal{C} = (1+t, 1-t^2, t-t^2)$.

Solution. (a) By definition the answer is $(1 \quad -1 \quad 3)^T$.

(b) By definition the answer is $(-1 \;\; 1 \;\; 3)^T$.

(c) For simplicity let $p(t) = 1 - t + 3t^2$. We know $[p(t)]_{\mathcal{A}} = (1 \quad -1 \quad 3)^T$. In order to find $[p(t)]_{\mathcal{C}}$ we will find $[I]_{\mathcal{CA}}$. Then use the fact that $[p(t)]_{\mathcal{C}} = [I]_{\mathcal{CA}}[p(t)]_{\mathcal{A}}$. By Theorem **??** we have

$$[I]_{\mathcal{AC}} = \big([1+t]_{\mathcal{A}} \ [1-t^2]_{\mathcal{A}} \ [t-t^2]_{\mathcal{A}}\big) = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & -1 \end{pmatrix}.$$

Therefore, by a Theorem we have the following:

$$[I]_{\mathcal{CA}} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 \\ -1/2 & 1/2 & -1/2 \end{pmatrix}$$

The final answer is obtained by evaluating $[I]_{\mathcal{CA}}[p(t)]_{\mathcal{A}}$. The answer is $(3/2 \ -1/2 \ -5/2)^T$. □

Example 4.7. Prove that if A and B are similar matrices, then $\operatorname{tr} A = \operatorname{tr} B$. Is it true that if $\operatorname{tr} A = \operatorname{tr} B$, then A and B must be similar?

Solution. Since A and B are similar, $B = PAP^{-1}$ for some invertible matrix P. By Example 1.4, we have $\operatorname{tr}(PAP^{-1}) = \operatorname{tr}(P^{-1}PA) = \operatorname{tr} A$. Therefore, $\operatorname{tr} A = \operatorname{tr} B$, as desired. □

Example 4.8. Let $T : \mathbb{R}^2 \to \mathbb{P}_1$ be the linear transformation given by $T(a,b) = (a+b) + (b-a)t$. Find $[T]_{\mathcal{BA}}$, where $\mathcal{A} = (\mathbf{e}_2, \mathbf{e}_1)$ and $\mathcal{B} = (1,t)$ are ordered bases of $\mathbb{R}^2$ and $\mathbb{P}_1$. You may assume T is linear and $\mathcal{A}, \mathcal{B}$ are bases for $\mathbb{R}^2$ and $\mathbb{P}_1$.

Solution. By a theorem, we have

$$[T]_{\mathcal{BA}} = ([T(0,1)]_{\mathcal{B}}[T(1,0)]_{\mathcal{B}}) = ([1+t]_{\mathcal{B}}[1-t]_{\mathcal{B}}) = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

□

Example 4.9. Consider the linear transformation $T : \mathbb{P}_2 \to \mathbb{P}_2$ given by $T(at^2+bt+c) = (a+b)t^2 + (b+c)$. Let $\mathcal{A} = (1, 2+3t, 1-t^2)$ and $\mathcal{B} = (1+t, 1-t, t+t^2)$ be ordered bases for $\mathbb{P}_1$. Find $[T]_{\mathcal{BA}}$. You may assume $\mathcal{A}$ and $\mathcal{B}$ are bases and T is linear.

Solution. Let $S = (1, t, t^2)$ be the standard ordered basis for $\mathbb{P}_2$. By Theorem 4.7, we know

$$[T]_{\mathcal{BA}} = [I]_{\mathcal{BS}}[T]_{\mathcal{SA}}.$$

We will now evaluate each one of the three matrices above.

$$[I]_{\mathcal{BS}} = [I]_{\mathcal{SB}}^{-1} = \big([1+t]_{\mathcal{S}} \ [1-t]_{\mathcal{S}} \ [t+t^2]_{\mathcal{S}}\big) = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 0 & 1 \end{pmatrix}^{-1}$$

Furthermore,

$$[T]_{\mathcal{S}\mathcal{A}} = \big([T(1)]_{\mathcal{S}}\ [T(2+3t)]_{\mathcal{S}}\ [T(1-t^2)]_{\mathcal{S}}\big) = \big([1]_{\mathcal{S}}\ [5+3t^2]_{\mathcal{S}}\ [1-t^2]_{\mathcal{S}}\big) = \begin{pmatrix} 1 & 5 & 1 \\ 0 & 0 & 0 \\ 0 & 3 & -1 \end{pmatrix}$$

Therefore, the answer can be evaluated by performing the following operations:

$$[T]_{\mathcal{B}\mathcal{A}} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 5 & 1 \\ 0 & 0 & 0 \\ 0 & 3 & -1 \end{pmatrix}$$

(For a complete solution, this calculation must be done.) □

Example 4.10. Prove for every two matrices A, B of the same size $\operatorname{rank}(A+B) \le \operatorname{rank}(A) + \operatorname{rank}(B)$.

Solution. Let $\mathbf{v}_1, \ldots, \mathbf{v}_r$ and $\mathbf{w}_1, \ldots, \mathbf{w}_s$ be bases for $\operatorname{Col}(A)$ and $\operatorname{Col}(B)$, respectively.

Suppose columns of A are $\mathbf{a}_1, \ldots, \mathbf{a}_n$ in that order and columns of B are $\mathbf{b}_1, \ldots, \mathbf{b}_n$ in that order. Thus, columns of $A+B$ are $\mathbf{a}_1 + \mathbf{b}_1, \ldots, \mathbf{a}_n + \mathbf{b}_n$. Every vector $\mathbf{x}$ in $\operatorname{Col}(A+B)$ can be written as

$$\mathbf{x} = \sum_{j=1}^{n} c_j(\mathbf{a}_j + \mathbf{b}_j) = \sum_{j=1}^{n} c_j \mathbf{a}_j + \sum_{j=1}^{n} c_j \mathbf{b}_j$$

Since $\sum\limits_{j=1}^{n} c_j \mathbf{a}_j \in \operatorname{Col}(A)$, the vector $\sum\limits_{j=1}^{n} c_j \mathbf{a}_j$ can be written as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_r$. Similarly $\sum\limits_{j=1}^{n} c_j \mathbf{b}_j$ is a linear combination of $\mathbf{w}_1, \ldots, \mathbf{w}_s$. Therefore, $\mathbf{x}$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{w}_1, \ldots, \mathbf{w}_s$. This means $\operatorname{Col}(A+B)$ has a generating set of size $r+s$. Therefore, by Theorem 3.8, the dimension of $\operatorname{Col}(A)$ is at most $r+s$. Thus, $\operatorname{rank}(A+B) \le r+s = \operatorname{rank} A + \operatorname{rank} B$. □

Example 4.11. Prove that a matrix $A \in M_{m\times n}(\mathbb{F})$ has rank 1 iff $A = \mathbf{u}\mathbf{v}^T$ for some nonzero column vectors $\mathbf{u} \in \mathbb{F}^m$ and $\mathbf{v} \in \mathbb{F}^n$.

Solution. A is of rank 1 iff the dimension of the column space of A is 1. This is equivalent to $\operatorname{Col}(A) = \operatorname{span}\{\mathbf{u}\}$, for some nonzero column vector $\mathbf{u} \in \mathbb{F}^m$. This is equivalent to all columns of A being scalar multiples of $\mathbf{u}$ and one of these columns must be nonzero. Let the columns of A be $c_1\mathbf{u}, \ldots, c_n\mathbf{u}$. We conclude $\operatorname{rank} A = 1$ if and only if

$$A = (c_1\mathbf{u} \cdots c_n\mathbf{u}) = \mathbf{u}\begin{pmatrix} c_1 & \cdots & c_n \end{pmatrix},$$

and at least one of c_j's is nonzero. □

Example 4.12. Suppose for a matrix A we have $\operatorname{Col}(A) \subseteq \operatorname{Row}(A)$. Prove that $\operatorname{Row}(A) = \operatorname{Col}(A)$.

Solution. By Theorem 4.2, the dimensions of $\operatorname{Row}(A)$ and $\operatorname{Col}(A)$ are the same. By Theorem 3.10, since $\operatorname{Col}(A) \subseteq \operatorname{Row}(A)$, we must have $\operatorname{Col}(A) = \operatorname{Row}(A)$. □

4.4 Exercises

Exercise 4.1. *Determine if each statement is true or false.*

(a) The nullity of a matrix A is at least the number of zero rows of A.

(b) The rank of a matrix A is the same as the number of nonzero columns of A.

(c) The nullity of a matrix is the same as the nullity of its transpose.

(d) If for a matrix A we have $\mathrm{Row}\,(A) \subseteq \mathrm{Col}\,(A)$*, then A is a square matrix.*

(e) If A is an invertible matrix, then $\mathrm{Row}\,(A) = \mathrm{Col}\,(A)$.

Exercise 4.2. *Suppose $A \in M_{m\times n}(\mathbb{F})$ and $B \in M_{n\times k}(\mathbb{F})$.*

(a) Prove that $\mathrm{Col}\,(AB) \subseteq \mathrm{Col}\,(A)$.

(b) Prove that $\mathrm{Row}\,(AB) \subseteq \mathrm{Row}\,(B)$.

(c) Deduce that $rank(AB) \leq \min(\mathrm{rank}\, A, \mathrm{rank}\, B)$.

(d) Prove that if A is left invertible, then $\mathrm{Row}\,(AB) = \mathrm{Row}\,(B)$*, and if B is right invertible, then* $\mathrm{Col}\,(AB) = \mathrm{Col}\,(A)$.

Exercise 4.3. *Suppose $A \in M_{m\times n}(\mathbb{F})$ and $B \in M_{n\times k}(\mathbb{F})$ such that* $\mathrm{Col}\,(AB) = \mathrm{Col}\,(A)$*. Prove that there is a matrix $C \in M_{k\times n}$ for which $ABC = A$.*

Exercise 4.4. *For each of the following matrices:*

1. *Find its rank and nullity.*

2. *Find all of its four fundamental subspaces.*

$$\begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & 1 & 3 \end{pmatrix}; \begin{pmatrix} -1 & 0 & 2 \\ 1 & 0 & 2 \\ -1 & 0 & -2 \\ 2 & 1 & -8 \end{pmatrix}; \begin{pmatrix} i & 1 & 0 \\ 1 & 1-i & 1 \\ 2 & -i & i \end{pmatrix}.$$

Exercise 4.5. *A 25×64 matrix has rank 14. What are the dimensions of its four fundamental subspaces?*

Exercise 4.6. *Prove that if the nullity of A and A^T are the same, then A is a square matrix.*

Exercise 4.7. *In each case provide an example of a matrix satisfying the given conditions or show no such matrix exists.*

(a) A has 16 columns, $\mathrm{rank}\, A = 15$ *and* $\mathrm{Null}\,(A)$ *contained two linear independent vectors.*

(b) $A \in M_{15\times 31}(\mathbb{F})$, $\dim \mathrm{Ker}\, A = 10$ *and A has no entries of zero.*

(c) $A \in M_{4\times 6}(\mathbb{F})$, *and* $\operatorname{rank} A = 7$.

(d) $A \in M_{10\times 9}(\mathbb{F})$, $\operatorname{rank} A = 9$ *and* A *has precisely 82 entries that are zero.*

Exercise 4.8. *Two matrices* A, B *satisfy* $\operatorname{Row}(A) = \operatorname{Row}(B)$ *and* $\dim \operatorname{Null}(A) = \dim \operatorname{Null}(B)$. *Do* A *and* B *have to be the same matrices? Do they have to have the same size?*

Exercise 4.9. *How many matrices* $A \in M_4(\mathbb{F})$ *satisfy all of the following?*

(a) $\operatorname{rank} A = 3$;

(b) All entries of A *are from the set* $\{0, 1, 2, \ldots, 7\}$; *and*

(c) A *has precisely 3 nonzero entries.*

Exercise 4.10. *Two matrices* A, B *have the same corresponding four fundamental subspaces. Can we conclude* $A = B$*?*

Exercise 4.11. *Prove that for a matrix* $A \in M_n(\mathbb{R})$ *it is not possible to have* $\operatorname{Ker} A = \operatorname{Col}(A^T)$. *By an example show this can happen when* A *is allowed to have nonreal entries.*

Exercise 4.12. *Suppose* $A, B \in M_n(\mathbb{F})$ *are similar matrices. Prove that*

(a) $\operatorname{rank} A = \operatorname{rank} B$.

(b) A *and* B *have the same nullity, but their null spaces may be different.*

Exercise 4.13. *Let* $T : V \to W$ *be a linear transformation, and* $\mathcal{A}$ *and* $\mathcal{B}$ *be ordered bases for* V *and* W, *respectively. Prove that* T *is invertible if and only if* $[T]_{\mathcal{B}\mathcal{A}}$ *is an invertible matrix.*

Exercise 4.14. *Write down the coordinate vector of* $t^3 - 2t + 1$ *with respect to the given bases of* $\mathbb{P}_3$.

(a) $\mathcal{A} = (1, t^3, t, t^2)$.

(b) $\mathcal{B} = (1 + t, 1 - t^2, t^3 + t, -t^2)$.

You may assume $\mathcal{A}, \mathcal{B}$ *are bases for* $\mathbb{P}_3$.

Exercise 4.15. *Suppose* $A, B \in M_n(\mathbb{F})$ *are similar matrices. Prove that there is an ordered basis for* $\mathbb{F}^n$ *for which* $[T]_{\mathcal{A}\mathcal{A}} = B$, *where* $T : \mathbb{F}^n \to \mathbb{F}^n$ *is the linear transformation associated with* A, *i.e.* $T(\mathbf{v}) = A\mathbf{v}$.

Exercise 4.16. *Prove that "being similar" is an equivalence relation in* $M_n(\mathbb{F})$. *In other words, prove the following:*

(a) Every matrix is similar to itself. (Reflexive)

(b) If A *is similar to* B, *then* B *is similar to* A. *(Symmetric)*

(c) If A *is similar to* B *and* B *is similar to* C, *then* A *is similar to* C. *(Transitive)*

Exercise 4.17. *Suppose every entry of $A \in M_n(\mathbb{F})$ is an integer. Suppose every entry of $B \in M_n(\mathbb{R})$ except for its $(1,1)$ entry is an integer, and $(1,1)$ entry of B is not an integer. Prove that A and B cannot be similar.*

Exercise 4.18. *Prove that if A and B are similar matrices, then so are A^T and B^T.*

Exercise 4.19. *Find the change of coordinate matrix from $\mathcal{A}$ to $\mathcal{B}$, where $\mathcal{A} = (1+5t, t-5)$ and $\mathcal{B} = (1+t, 2-t)$ are bases for $\mathbb{P}_1$. You may assume $\mathcal{A}, \mathcal{B}$ are bases.*

Exercise 4.20. *Find the matrix of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x,y) = (x-2y, 3x+y)$ once in the standard ordered basis $(\mathbf{e}_1, \mathbf{e}_2)$ and once in the ordered basis $((1,2),(2,1))$.*

Exercise 4.21. *Determine if each statement is true or false.*

(a) $\operatorname{rank}(AB) = \operatorname{rank}(BA)$ *for every $A, B \in M_n(\mathbb{F})$ and every n.*

(b) $\operatorname{Col}(A) = \operatorname{Row}(A^T)$ *for every matrix A.*

(c) The rank of a matrix is the number of nonzero columns of its reduced echelon form.

(d) The nullity of a matrix is the number of zero rows of its reduced echelon form.

Exercise 4.22. *Let $T, S : V \to W$ be linear transformations of vector spaces over $\mathbb{F}$ and $c \in \mathbb{F}$ be a scalar. Suppose $\mathcal{A}$ and $\mathcal{B}$ are bases for V and W, respectively. Prove that:*

$$[T + cS]_{\mathcal{BA}} = [T]_{\mathcal{BA}} + c[S]_{\mathcal{BA}}.$$

Suppose $\dim V = m$ *and* $\dim W = n$. *Use the above to prove the function $\varphi : \mathcal{L}(V,W) \to M_{n\times m}(\mathbb{F})$ given by $\varphi(T) = [T]_{\mathcal{BA}}$ is an isomorphism.*

Exercise 4.23. *Suppose V and W are finite-dimensional vector spaces of dimensions m and n, respectively. Let $\mathcal{A}$ and $\mathcal{B}$ be bases for V and W, respectively. Prove that for every $A \in M_{n\times m}(\mathbb{F})$, there is $T \in \mathcal{L}(V,W)$ for which $[T]_{\mathcal{BA}} = A$.*

Exercise 4.24. *Suppose $A \in M_n(\mathbb{F})$ satisfies $A^2 = 0$. Prove that* $\operatorname{rank} A \leq \dfrac{n}{2}$.

Exercise 4.25. *Consider the following block diagonal matrix, where $A_1, \ldots, A_r$ are matrices.*

$$A = \begin{pmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_r \end{pmatrix}$$

Prove $\operatorname{rank} A = \sum_{j=1}^{r} \operatorname{rank} A_j$.

Exercise 4.26. *Suppose $a_1, \ldots, a_n$ are scalars and $b_1, \ldots, b_n$ is a permutation of $a_1, \ldots, a_n$. Prove the diagonal matrices* $\operatorname{diag}(a_1, \ldots, a_n)$ *and* $\operatorname{diag}(b_1, \ldots, b_n)$ *are similar.*

Exercise 4.27 (A Generalization of the Previous Exercise). *Let $A_1, \ldots, A_r$ be square matrices. Suppose σ is a permutation of $\{1, 2, \ldots, n\}$. Prove the two block matrices below are similar:*

$$\begin{pmatrix} A_1 & & & \mathit{0} \\ & A_2 & & \\ & & \ddots & \\ \mathit{0} & & & A_r \end{pmatrix} \text{ and } \begin{pmatrix} A_{\sigma(1)} & & & \mathit{0} \\ & A_{\sigma(2)} & & \\ & & \ddots & \\ \mathit{0} & & & A_{\sigma(r)} \end{pmatrix}.$$

Hint: Try $r = 2$ first.

Exercise 4.28. *Let $A \in M_{m\times n}(\mathbb{F})$. Define $T : M_{n\times k}(\mathbb{F}) \to M_{m\times k}(\mathbb{F})$ by $T(X) = AX$.*

(a) Prove T is a linear transformation.

(b) Suppose $\operatorname{rank} A = r$. *Prove* $\dim \operatorname{Ker} T = (n - r)^k$.

Exercise 4.29. *Prove the following 3×3 matrices are similar.*

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \text{ and } \begin{pmatrix} a_{33} & a_{32} & a_{31} \\ a_{23} & a_{22} & a_{21} \\ a_{13} & a_{12} & a_{11} \end{pmatrix}.$$

Exercise 4.30. *Generalize the previous exercise for an $n \times n$ matrix.*

4.5 Challenge Problems

Exercise 4.31. *Find all permutations σ of $1, 2, \ldots, n^2$ for which the following matrices are similar, for every choice of scalars $a_1, a_2, \ldots, a_{n^2}$.*

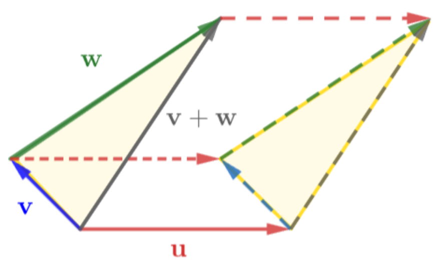
$$\begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ a_{n+1} & a_{n+2} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{(n-2)n+1} & a_{(n-2)n+2} & \cdots & a_{n(n-1)} \\ a_{(n-1)n+1} & a_{(n-1)n+2} & \cdots & a_{n^2} \end{pmatrix} \text{ and } \begin{pmatrix} a_{\sigma(1)} & a_{\sigma(2)} & \cdots & a_{\sigma(n)} \\ a_{\sigma(n+1)} & a_{\sigma(n+2)} & \cdots & a_{\sigma(2n)} \\ \vdots & \vdots & \vdots & \vdots \\ a_{\sigma((n-2)n+1)} & a_{\sigma((n-2)n+2)} & \cdots & a_{\sigma(n(n-1))} \\ a_{\sigma((n-1)n+1)} & a_{\sigma((n-1)n+2)} & \cdots & a_{\sigma(n^2)} \end{pmatrix}.$$

Week 5

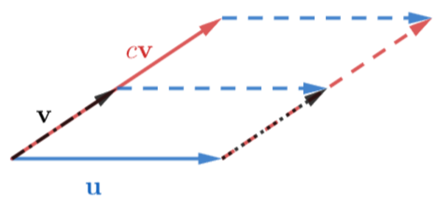
5.1 Determinants

In this section we would like to define the determinant of a square matrix. One interpretation of determinant is “volume”. Given n vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{F}^n$, we want the $n \times n$ determinant corresponding to $\mathbf{v}_1, \ldots, \mathbf{v}_n$ to determine the volume of the parallelepiped determined by these n vectors. We expect any reasonable volume to satisfy the following properties:

$\mathrm{Area}(\mathbf{u}, \mathbf{u}) = 0$



$\mathrm{Area}(\mathbf{u}, \mathbf{v} + \mathbf{w}) = \mathrm{Area}(\mathbf{u}, \mathbf{v}) + \mathrm{Area}(\mathbf{u}, \mathbf{w})$



$\mathrm{Area}(\mathbf{u}, c\mathbf{v}) = c\mathrm{Area}(\mathbf{u}, \mathbf{v})$

Definition 5.1. Let $D : M_n(\mathbb{F}) \to \mathbb{F}$ be a function.

(a) We say D is **multi-linear** iff D is linear with respect to each row. In other words, for every j, $1 \le j \le n$, we have

$$D\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ a\mathbf{v}_j + b\mathbf{w} \\ \vdots \\ \mathbf{v}_n \end{pmatrix} = aD\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_j \\ \vdots \\ \mathbf{v}_n \end{pmatrix} + bD\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{w} \\ \vdots \\ \mathbf{v}_n \end{pmatrix} \leftarrow j-\text{th row.}$$

(b) We say D is **alternating** iff $D\begin{pmatrix}\mathbf{v}_1\\ \vdots\\ \mathbf{v}_n\end{pmatrix} = 0$ when $\mathbf{v}_j = \mathbf{v}_k$ for some $j \neq k$.

To keep the notations more compact, instead of writing $D\begin{pmatrix}\mathbf{v}_1\\ \vdots\\ \mathbf{v}_n\end{pmatrix}$ we write $D(\mathbf{v}_1, \dots, \mathbf{v}_n)$; inserting commas to indicate $\mathbf{v}_1, \dots, \mathbf{v}_n$ are rows and not columns.

Theorem 5.1. *Let $D : M_n(\mathbb{F}) \to \mathbb{F}$ be alternating and multi-linear, then it satisfies the following properties.*

(a) Swapping two rows, negates D. In other words,

$$D(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_k, \dots, \mathbf{v}_n) = -D(\mathbf{v}_1, \dots, \mathbf{v}_k, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n)$$

(b) Scaling a row by c scales D by c. In other words,

$$D(\mathbf{v}_1, \dots, c\mathbf{v}_j, \dots, \mathbf{v}_n) = cD(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n).$$

(c) Adding a multiple of one row to another does not change D. In other words,

$$D(\mathbf{v}_1, \dots, \mathbf{v}_j + c\mathbf{v}_k, \dots, \mathbf{v}_n) = D(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n) \text{ if } j \neq k.$$

(d) If $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly dependent, then $D(\mathbf{v}_1, \dots, \mathbf{v}_n) = 0$.

Clearly the first three operations are very familiar. These are precisely the row operations that we explored when solving systems of linear equations.

Example 5.1. Find all alternating, multi-linear function $D : M_2(\mathbb{F}) \to \mathbb{F}$.

Theorem 5.2. *For every positive integer n, there is a unique multi-linear, alternating function $D : M_n(\mathbb{F}) \to \mathbb{F}$ satisfying $D(I) = 1$.*

Definition 5.2. Let n be a positive integer. The **determinant** is the unique multi-linear, alternating function $D : M_n(\mathbb{F}) \to \mathbb{F}$ for which $D(I_n) = 1$. Determinant of a matrix A is denoted by $\det A$ or $\det(A)$.

Corollary 5.1 (Leibniz formula for determinants)**.** For every matrix $A = (a_{jk}) \in M_n(\mathbb{F})$, we have

$$\det A = \sum_{\sigma \in S_n} \epsilon_\sigma a_{1\sigma(1)} \cdots a_{n\sigma(n)},$$

where S_n is the set of all permutations $\sigma : \{1, 2, \dots, n\} \to \{1, 2, \dots, n\}$ and $\epsilon_\sigma = \pm 1$ only depends on σ,

Example 5.2. Evaluate

$$\det\begin{pmatrix} 1 & 2 & -1 \\ 2 & 0 & 1 \\ 3 & 2 & 1 \end{pmatrix}.$$

5.2 Row Operations and Matrix Multiplication

Theorem 5.3. *Let A and B be two $n \times n$ matrices, then* $\det(AB) = (\det A)(\det B)$*.*

Corollary 5.2. If A and B are square similar matrices, then they have the same determinant.

Proof. By definition, $A = PBP^{-1}$ for some invertible matrix P. By Theorem 5.3, we have $\det A = (\det P)(\det B)(\det P^{-1})$. Note that since $PP^{-1} = I$, we have $(\det P)(\det P^{-1}) = \det I = 1$. Therefore, $\det A = \det B$. □

Definition 5.3. Determinant of a linear transformation $T : V \to V$ is defined as the determinant of $[T]_{\mathcal{A}\mathcal{A}}$, where $\mathcal{A}$ is some ordered basis for V. Note that since the matrices of T in different bases are similar, their determinants are the same. Therefore, this is well-defined.

Determinants can be evaluated using co-factor expansions. Here is an example.

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11} \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{12} \det \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} + a_{13} \det \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}.$$

In other words, we can write the determinant of a 3×3 matrix A as follows:

$$\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13},$$

where A_{ij} is obtained by removing the i-th row and the j-th row of A.

Theorem 5.4. *(Cofactor Expansion Along a Row or a Column) Let $A \in M_n(\mathbb{F})$, with a_{jk} as its (j,k) entry. Then, for every j with $1 \leq j \leq n$, we have*

$$\det A = \sum_{k=1}^{n} (-1)^{j+k} a_{jk} \det A_{jk},$$

where A_{jk} is obtained by removing the j-th row and the k-th column of A. Similarly, for every k with $1 \leq k \leq n$, we have

$$\det A = \sum_{j=1}^{n} (-1)^{j+k} a_{jk} \det A_{jk},$$

Theorem 5.5. *For a square matrix A the following are equivalent:*

(a) A is invertible.

(b) $\det A \neq 0$.

(c) Columns of A are linearly independent.

(d) Rows of A are linearly independent.

Theorem 5.6 (Cramer's Rule)**.** *Let $A = (\mathbf{a}_1 \cdots \mathbf{a}_n)$ be an invertible matrix. Then for every column vector $\mathbf{b}$, the only solution to $A\mathbf{x} = \mathbf{b}$ is*

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

where $x_j = \dfrac{\det(\mathbf{a}_1 \cdots \mathbf{a}_{j-1}\ \mathbf{b}\ \mathbf{a}_{j+1} \cdots \mathbf{a}_n)}{\det(A)}$.

Example 5.3. Solve the system of equations using Cramer's Rule:

$$\begin{cases} x + y - 2z = 1 \\ y + 2z = 1 \\ x - z = 3 \end{cases}$$

Theorem 5.7. *Let A be an invertible matrix. Then the (j,k) entry of A^{-1} equals $\dfrac{(-1)^{j+k}\det(A_{kj})}{\det A}$, where A_{kj} is the matrix obtained from A by removing the k-th row and the j-th column of A.*

Example 5.4. Prove the inverse of any invertible 2×2 matrix is given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Example 5.5. Suppose $A \in M_n(\mathbb{R})$ is an invertible matrix whose entries are all rational. Prove that for every $\mathbf{b} \in \mathbb{Q}^n$, the solution $\mathbf{x}$ to $A\mathbf{x} = \mathbf{b}$ is in $\mathbb{Q}^n$.

5.3 Minors, Rank and Determinant

We know determinant of a square matrix is nonzero if and only if it is invertible. We would like to find a relation between rank of a matrix and some determinant.

Definition 5.4. Let A be a matrix. A **submatrix** of A is a matrix obtained by selecting some arbitrary columns and some arbitrary rows of A and looking at the entries that lie at the intersections of these rows and columns. In other words, the submatrix corresponding to rows numbered $i_1 < i_2 < \cdots < i_k$ and columns numbered $j_1 < j_2 < \cdots < j_\ell$, is a matrix whose (r,s) entry is $A_{i_r j_s}$, where A_{ij} is the (i,j) entry of A. A **minor** of order k of a matrix A is the determinant of a $k \times k$ submatrix of A.

Theorem 5.8. *For a nonzero matrix A, the rank of A equals to the maximum integer k for which there exists a nonzero minor of order k.*

5.4 Examples

Example 5.6. For scalars $a_1, \ldots, a_n$ let $A = \operatorname{diag}(a_1, \ldots, a_n)$ be the $n \times n$ matrix whose diagonal entries are $a_1, \ldots, a_n$ in that order. Prove that $\det A = a_1 \cdots a_n$ in two ways:

(a) Using induction along with co-factor expansion.

(b) Using row operations

Solution. (a) We will prove this by induction on n.

Basis step. For $n = 1$, $A = (a_1)$, and we have $\det(a_1) = a_1$.

Inductive step. Expanding $\det A$ along the last row we obtain $\det A = (-1)^{n+n} a_n \det(\mathrm{diag}(a_1, \dots, a_{n-1}))$ $(*)$, since the rest of the terms in the expansion are zero. By inductive hypothesis $\det(\mathrm{diag}(a_1, \dots, a_{n-1})) = a_1 \cdots a_{n-1}$. Combining this with $(*)$ we obtain the result.

(b) Note that rows of the given matrix are $a_1\mathbf{e}_1, \dots, a_n\mathbf{e}_n$. By the rescaling row operation with a factor of a_1 and with respect to the first row we obtain the following:

$$\det \begin{pmatrix} a_1\mathbf{e}_1 \\ a_2\mathbf{e}_2 \\ \vdots \\ a_n\mathbf{e}_n \end{pmatrix} = a_1 \det \begin{pmatrix} \mathbf{e}_1 \\ a_2\mathbf{e}_2 \\ \vdots \\ a_n\mathbf{e}_n \end{pmatrix}$$

Repeating this we conclude that

$$\det(\mathrm{diag}(a_1, \dots, a_n)) = a_1 \cdots a_n \det I = a_1 \cdots a_n,$$

as desired. □

Example 5.7. Suppose A is a square matrix such that A and A^{-1} both only have integer entries. Prove that $\det A = \pm 1$.

Solution. First note that if B is a square matrix with integer entries, then $\det B$ is also an integer. This can be shown by induction on the size of B and cofactor expansion. (Show this!) Therefore, $\det A$ and $\det A^{-1}$ are both integer. Since $\det(AA^{-1}) = \det I = 1$, we must have $(\det A)(\det A^{-1}) = 1$. Since both $\det A$ and $\det A^{-1}$ are integers, we must have $\det A = \pm 1$. □

Example 5.8. Prove the converse of the previous example: Suppose A is a matrix with integer entries for which $\det A = \pm 1$. Prove that all entries of A^{-1} are integers.

Solution. We know the (j,k) entry of A^{-1} is $\dfrac{(-1)^{j+k}\det(A_{kj})}{\det A} = \pm\det(A_{kj})$, since $\det A = \pm 1$. Note that since all entries of A are integers, by what we saw in the previous example $\det(A_{kj})$ is an integer. Therefore, every entry of A^{-1} is an integer. □

Example 5.9. Let a, b, c be three real numbers. Evaluate the following determinant:

$$\det \begin{pmatrix} a & a^2 & a^3 \\ b & b^2 & b^3 \\ c & c^2 & c^3 \end{pmatrix}$$

Solution. We will use properties of determinant.

$$\det \begin{pmatrix} a & a^2 & a^3 \\ b & b^2 & b^3 \\ c & c^2 & c^3 \end{pmatrix} = abc \det \begin{pmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{pmatrix}$$

Use row operations $R_2 - R_1$ and $R_3 - R_1$ we obtain the following:

$$abc \det \begin{pmatrix} 1 & a & a^2 \\ 0 & b-a & b^2 - a^2 \\ 0 & c-a & c^2 - a^2 \end{pmatrix} = abc(b-a)(c-a) \det \begin{pmatrix} 1 & a & a^2 \\ 0 & 1 & b+a \\ 0 & 1 & c+a \end{pmatrix},$$

which is obtained by taking out scalars $b - a$ and $c - a$ from the second and third rows of the matrix. Using the row operation $R_3 - R_2$ we obtain the following:

$$abc(b-a)(c-a) \det \begin{pmatrix} 1 & a & a^2 \\ 0 & 1 & b+a \\ 0 & 0 & c-b \end{pmatrix} = abc(b-a)(c-a)(c-b) \det \begin{pmatrix} 1 & a & a^2 \\ 0 & 1 & b+a \\ 0 & 0 & 1 \end{pmatrix}$$

Expanding this along the first column and the fist column again we obtain $\boxed{abc(b-a)(c-a)(c-b)}$. □

Example 5.10. Let $A \in M_n(\mathbb{F})$ and $c \in \mathbb{F}$. Prove $\det(cA) = c^n \det A$.

Solution. Suppose rows of A are $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Then rows of cA are $c\,\mathbf{a}_1, \ldots, c\,\mathbf{a}_n$. By properties of determinant, we have the following:

$$\det(cA) = \det \begin{pmatrix} c\,\mathbf{a}_1 \\ \vdots \\ c\,\mathbf{a}_n \end{pmatrix} = c \det \begin{pmatrix} \mathbf{a}_1 \\ c\,\mathbf{a}_2 \\ \vdots \\ c\,\mathbf{a}_n \end{pmatrix} = c^2 \det \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ c\,\mathbf{a}_3 \\ \vdots \\ c\,\mathbf{a}_n \end{pmatrix} = \cdots = c^n \det \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{pmatrix} = c^n \det A.$$

□

Example 5.11. A matrix A is called skew-symmetric if $A^T = -A$. Prove that if an $n \times n$ matrix A is skew-symmetric and n is odd, then A is not invertible.

Solution. By a theorem we know $\det(A^T) = \det A$. At the same time, we know $\det(-A) = (-1)^n \det A = -\det A$, since n is odd. Therefore, $\det A = -\det A$ and hence $\det A = 0$. This implies that A is not invertible. □

Example 5.12. Suppose A is an upper triangular $n \times n$ matrix whose diagonal entries are $\lambda_1, \dots, \lambda_n$. Prove $\det A = \lambda_1 \cdots \lambda_n$.

Solution. We will prove this by induction on n.

Basis step. For $n = 1$, $A = (\lambda_1)$, and $\det A = \lambda_1 \det(1) = \lambda_1$.

Inductive step. Suppose A is an $n \times n$ upper triangular matrix with λ_1 in its $(1, 1)$ position. Using cofactor expansion along the first column, we conclude $\det A = (-1)^{1+1}\lambda_1 \det B$, where B is an upper triangular matrix whose diagonal entries are $\lambda_2, \dots, \lambda_n$. By inductive hypothesis, $\det B = \lambda_2 \cdots \lambda_n$. This implies $\det A = \lambda_1 \cdots \lambda_n$, as desired. □

Example 5.13. Let A be the $n \times n$ matrix, whose entries above or on the main diagonal are all 1's and whose entries below the main diagonal are all a variable t. Find $\det A$ in terms of n and t.

Solution. The matrix A is as follows:

$$A = \begin{pmatrix} 1 & 1 & \dots & 1 & 1 \\ t & 1 & \dots & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ t & t & \dots & 1 & 1 \\ t & t & \dots & t & 1 \end{pmatrix}_{n \times n}$$

We will apply the row operations $R_2 - tR_1, R_3 - tR_1, \dots, R_n - tR_1$ to obtain the following:

$$\det A = \det \begin{pmatrix} 1 & 1 & \dots & 1 & 1 \\ 0 & 1-t & \dots & 1-t & 1-t \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1-t & 1-t \\ 0 & 0 & \dots & 0 & 1-t \end{pmatrix}_{n \times n}$$

This is an upper triangular matrix and thus $\det A = (1 - t)^{n-1}$. □

Example 5.14. Evaluate the determinant of an $n \times n$ matrix whose off-diagonal entries are all 1 and whose diagonal entries are all a variable t.

Solution. Let

$$E = \begin{pmatrix} t & 1 & \dots & 1 & 1 \\ 1 & t & \dots & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \dots & t & 1 \\ 1 & 1 & \dots & 1 & t \end{pmatrix}_{n \times n}$$

Subtracting the first row from all other rows does not change the determinant and we obtain the following determinant:

$$\det E = \det \begin{pmatrix} t & 1 & \dots & 1 & 1 \\ 1-t & t-1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1-t & 0 & \dots & t-1 & 0 \\ 1-t & 0 & \dots & 0 & t-1 \end{pmatrix}$$

Adding all columns to the first we obtain the following:

$$\det E = \det \begin{pmatrix} t+n-1 & 1 & \dots & 1 & 1 \\ 0 & t-1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & t-1 & 0 \\ 0 & 0 & \dots & 0 & t-1 \end{pmatrix}$$

This is an upper triangular matrix. Thus, its determinant is the product of its diagonal entries. Therefore, $\det E = (t-1)^{n-1}(t+n-1)$. □

Example 5.15. Let $A(t)$ be an $m \times n$ matrix whose entries are all polynomials on variable t with complex coefficients. Prove that rank of $A(t)$ is constant for every $t \in \mathbb{C}$, except possibly finitely many complex numbers t, where the rank is smaller.

Solution. Consider all minors of $A(t)$. Since all entries of $A(t)$ are polynomials, all minors are also polynomials. If all of these minors are identically zero, then $A(t) = 0$, since minors of order 1 are the entries of $A(t)$. Otherwise, assume k is the largest integer for which there is a nonzero minor of order k. Since this minor is a polynomial, it has finitely many roots. Let S be the set of roots of this minor, which is a finite set. For every complex number not in S, this minor is nonzero. Thus, by Theorem 5.8 the rank of $A(c)$ is k for every $c \in \mathbb{C} \backslash S$. By the choice of k the rank of $A(c)$ never exceeds k for any $c \in \mathbb{C}$, as desired. □

5.5 Exercises

Exercise 5.1. *Consider a function $D : M_n(\mathbb{F}) \to \mathbb{F}$ and $c \in \mathbb{F}$ is a scalar.*

(a) Prove that if D is alternating, then so is cD, defined by $(cD)(A) = cD(A)$.

(b) Prove that if D is multi-linear, the so is cD.

Exercise 5.2. *Suppose $D : M_n(\mathbb{F}) \to \mathbb{F}$ is a multi-linear, alternating function. Prove that there is a scalar c for which $D(A) = c\ \det A$ for all $A \in M_n(\mathbb{F})$.*

Exercise 5.3. *To turn a matrix into one in reduced echelon form, we use three row operations: Row Addition, Row Interchange, and Row Scaling.*

(a) Prove that the Row Interchange operation is not needed. In other words, show that Row Interchange can be obtained from Row Addition and Row Scaling.

(b) Prove that Both Row Addition and Row Scaling are necessary to turn a matrix into one in reduced Echelon form.

Exercise 5.4. *Suppose $A, B \in M_n(\mathbb{F})$, and assume A is invertible. Prove there are infinitely many $r \in \mathbb{F}$ for which $A + rB$ is also invertible.*

Exercise 5.5. *Let A be a square matrix. Prove that all of the following matrices*

$$\begin{pmatrix} A & * \\ 0 & I \end{pmatrix}, \begin{pmatrix} I & * \\ 0 & A \end{pmatrix}, \begin{pmatrix} A & 0 \\ * & I \end{pmatrix}, \begin{pmatrix} I & 0 \\ * & A \end{pmatrix}$$

have determinant equal to $\det A$. *In each case $*$ is an arbitrary matrix that makes the given matrix a square matrix.*

Exercise 5.6. *Find the determinant of an $n \times n$ matrix whose minor diagonal entries are $a_1, \dots, a_n$ and all of whose entries below the minor diagonal are zero. In other words, find the determinant of the matrix:*

$$\begin{pmatrix} * & * & \cdots & a_1 \\ * & \cdots & a_2 & 0 \\ \vdots & \iddots & & \vdots \\ a_n & 0 & \cdots & 0 \end{pmatrix}.$$

Exercise 5.7. *Let D_n be the determinant of the $n \times n$ matrix–shown below–whose main diagonal entries are all 1's, the entries immediately above the main diagonal (if any exists) are all -1's and the entries immediately below the main diagonal (if any exists) are all 1's, and whose all other entries (if any exists) are all 0's.*

$$\begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & -1 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 1 \end{pmatrix}$$

(a) Evaluate D_1, D_2, D_3, D_4 and D_5.

(b) Conjecture a recursion for D_n.

(c) *Prove your claim in part (b). (Hint: Expand along the first column.)*

Exercise 5.8. *Let D_n be the determinant of the $n \times n$ matrix–shown below–whose main diagonal entries are all 3's, the entries immediately above the main diagonal (if any exists) are all 2's and the entries immediately below the main diagonal (if any exists) are all 1's, and whose all other entries (if any exists) are all 0's.*

$$\begin{pmatrix} 3 & 2 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & 3 & 2 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & 3 & 2 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 3 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 3 & 2 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 3 & 2 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 3 \end{pmatrix}$$

(a) *Evaluate D_1, D_2, D_3 and D_4.*

(b) *Conjecture a formula for D_n, for every n.*

(c) *Prove your claim in part (b) using induction.*

For the next exercise you will need the following familiar theorem:

Theorem 5.9. *Suppose $p(t) = A_0 + A_1 t + \cdots + A_n t^n$ is a polynomial with complex coefficients $A_0, A_1, \ldots, A_n$. Suppose $p(t)$ has n distinct roots $r_1, \ldots, r_n \in \mathbb{C}$. Then*

$$p(t) = A_n(t - r_1) \cdots (t - r_n).$$

Exercise 5.9 (Vandermonde Determinant)**.** *In this exercise you will prove the Vandermonde Determinant using induction:*

$$\det \begin{pmatrix} 1 & c_0 & c_0^2 & \cdots & c_0^n \\ 1 & c_1 & c_1^2 & \cdots & c_1^n \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & c_n & c_n^2 & \cdots & c_n^n \end{pmatrix} = \prod_{0 \le j < k \le n} (c_k - c_j) \qquad (*)$$

(a) *Prove $(*)$ for $n = 1$.*

(b) *Prove $(*)$ holds if $c_j = c_k$ for some $j \neq k$. For the rest of the problem assume c_j's are distinct.*

(c) *Instead of c_n in the last row use a variable t. Using cofactor expansion along the last row show that this determinant can be written as $A_0 + A_1 t + \cdots + A_n t^n$, where A_j's are constants depending on $c_0, \ldots, c_{n-1}$.*

(d) *Prove that the polynomial $p(t) = A_0 + A_1 t + \cdots + A_n t^n$ has n roots $t = c_0, c_1, \ldots, c_{n-1}$. Use this to show $p(t) = A_n(t - c_0) \cdots (t - c_{n-1})$. (Hint: Use Theorem 5.9.)*

(e) Assuming $(*)$ is true for $n-1$, find A_n. Use that to obtain a proof of the Vandermonde determinant using induction.

Exercise 5.10. *Is there a subspace of $M_2(\mathbb{R})$ of dimension larger than 1 whose only noninvertible matrix is the zero matrix? How about $M_3(\mathbb{R})$? How about $M_n(\mathbb{R})$ for other positive integers n? Discuss this for $M_n(\mathbb{C})$ and a positive integer n.*

Exercise 5.11. *Prove that if A, B are square matrices of the same size and that $AB = cI$ for some nonzero scalar c, then $BA = cI$.*

Exercise 5.12. *Suppose $A, B \in M_n(\mathbb{F})$. Prove $\det(A^2 + 2AB + B^2) = (\det(A + B))^2$.*

Exercise 5.13. *Determine if each statement is true or false.*

(a) $\det(A + BC) = \det(A + CB)$ *for every $A, B, C \in M_n(\mathbb{F})$ and all n.*

(b) $\det(AB^T) = \det(A^T B)$ *for every $A, B \in M_n(\mathbb{F})$ and all n.*

(c) $\det(AB) = \det(A^T B)$ *for every $A, B \in M_n(\mathbb{F})$ and all n.*

Exercise 5.14. *Prove that for every $A, B, C \in M_n(\mathbb{F})$ we have*

$$\det\begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = (\det A)(\det C), \ \text{ and } \det\begin{pmatrix} 0 & A \\ -B & C \end{pmatrix} = (\det A)(\det B).$$

Definition 5.5. A square matrix P is called a permutation matrix iff it is obtained by applying row interchange operations to the identity matrix. In other words, a permutation matrix is a matrix with precisely one entry of 1 in every row and all other entries 0.

Exercise 5.15. *Let P be an $n \times n$ permutation matrix.*

(a) Describe the inverse of P.

(b) Find the number of $n \times n$ permutation matrices.

(c) Prove that $P^N = I$ for some positive integer N.

(d) Suppose rows of P, in order from top to bottom, are $e_{\sigma(1)}, \dots, e_{\sigma(n)}$. Describe PA for a matrix $A \in M_n(\mathbb{F})$.

(e) What are columns of P, the permutation matrix provided in part (c)? Use that to describe AP for a matrix $A \in M_n(\mathbb{F})$.

Exercise 5.16. *Suppose $A, B \in M_n(\mathbb{R})$ are matrices that are similar in $M_n(\mathbb{C})$. Is it true that A and B must be similar as matrices of $M_n(\mathbb{R})$?*

Exercise 5.17. *Prove that for every square matrix A, we have $\det \overline{A} = \overline{\det A}$.*

Exercise 5.18. *Let $A \in M_n(\mathbb{F})$ be an invertible matrix. Prove the following:*

(a) $\det A = 1$ *if and only if* $A = E_1 \cdots E_k$ *for some row replacement elementary matrices* $E_1, \ldots, E_k$*.*

(b) $\det A = -1$ *if and only if* $A = E_1 \cdots E_k E_{k+1}$ *for some row replacement elementary matrices* $E_1, \ldots, E_k$ *and a row interchange elementary matrix* E_{k+1}*.*

(c) $A = E_1 \cdots E_k E_{k+1}$ *for some row replacement elementary matrices* $E_1, \ldots, E_k$ *and a row multiplication elementary matrix* E_{k+1}*.*

5.6 Challenge Problems

Exercise 5.19. *Consider a square matrix* A *whose entries in the* j*-th row from left to right form an arithmetic sequence with common difference* d_j *and first term* x_j*. Find* $\det A$ *in terms of* x_j*'s and* d_j*'s.*

Exercise 5.20. *Find the determinant of the* $n \times n$ *matrix whose entries from left to right and from top to bottom are* $\cos 1, \cos 2, \ldots, \cos(n^2)$*, where all angles are measured in radians.*

Exercise 5.21. *Suppose* $A, B \in M_n(\mathbb{R})$ *satisfy* $AB = BA$*. Prove that* $\det(A^2 + B^2) \geq 0$*.*

Week 6

6.1 Eigenvalues and Eigenvectors

Definition 6.1. Consider a linear transformation $L : V \to V$. We say a scalar $\lambda \in \mathbb{F}$ is an **eigenvalue** (or e-value, for short) of L iff $L(\mathbf{v}) = \lambda \mathbf{v}$ for some nonzero vector $\mathbf{v} \in V$. This vector $\mathbf{v}$ is called an **eigenvector** (or e-vector, for short) and the pair $(\lambda, \mathbf{v})$ is called an **eigenpair** (or e-pair, for short) of L. For a matrix $A \in M_n(\mathbb{F})$, we define the same notions, replacing $L(\mathbf{v})$ by $A\mathbf{v}$. The set of all eigenvalues of a matrix A (resp. a linear transformation L) is called the **spectrum** and is denoted by $\sigma(A)$ (resp. $\sigma(L)$).

Theorem 6.1. *Let $L : V \to V$ be a linear transformation, and $\mathcal{A}$ be a basis for V. A scalar $\lambda \in \mathbb{F}$ and a nonzero vector $\mathbf{v} \in V$ form an eigenpair $(\lambda, \mathbf{v})$ for L if and only if $\mathbf{v} \in \operatorname{Ker}(T - \lambda I_V)$. Similarly, for a matrix $A \in M_n(\mathbb{F})$, a scalar $\lambda \in \mathbb{F}$ and a nonzero vector $\mathbf{v} \in \mathbb{F}^n$, the pair $(\lambda, \mathbf{v})$ is an eigenpair for A if and only if $\mathbf{v} \in \operatorname{Ker}(A - \lambda I)$.*

Theorem 6.2. *Let $T : V \to V$ be a linear transformation of a finite dimensional vector space V, and $\mathcal{A}$ be an ordered basis for V. A scalar $\lambda \in \mathbb{F}$ is an eigenvalue of T if and only if $\det([T]_{\mathcal{A}\mathcal{A}} - \lambda I) = 0$.*

Theorem 6.3. *If A and B are similar matrices, then the polynomials $\det(A - tI)$ and $\det(B - tI)$ are the same.*

Since the matrices $[T]_{\mathcal{A}\mathcal{A}}$ and $[T]_{\mathcal{B}\mathcal{B}}$ of a linear transformation $T : V \to V$ in two ordered bases $\mathcal{A}$ and $\mathcal{B}$ are similar, the polynomials $\det([T]_{\mathcal{A}\mathcal{A}} - tI)$ and $\det([T]_{\mathcal{B}\mathcal{B}} - tI)$ are the same. This brings us to the following definition.

Definition 6.2. Let $T : V \to V$ be a linear transformation of a finite dimensional vector space V and $\mathcal{A}$ be an ordered basis for V. The polynomial $\det([T]_{\mathcal{A}\mathcal{A}} - tI)$ is called the **characteristic polynomial** of T. Similarly, for a square matrix A, the polynomial $\det(A - tI)$ is called the characteristic polynomial of A.

Remark 6.1. Suppose V is a finite dimensional vector space with an ordered basis $\mathcal{A}$. Let $T : V \to V$ be a linear transformation. Since $[\cdot]_{\mathcal{A}}$ is an isomorphism (See Theorem 4.5), the equality $T(\mathbf{v}) = \lambda \mathbf{v}$ is equivalent to $[T(\mathbf{v})]_{\mathcal{A}} = [\lambda \mathbf{v}]_{\mathcal{A}} = \lambda [\mathbf{v}]_{\mathcal{A}}$. By Theorem 4.6 this is equivalent to $[T]_{\mathcal{A}\mathcal{A}}[\mathbf{v}]_{\mathcal{A}} = \lambda [\mathbf{v}]_{\mathcal{A}}$. So, finding eigenpairs of T is equivalent to finding eigenpairs of its matrix $[T]_{\mathcal{A}\mathcal{A}}$. Because of this, we will mainly focus on understanding eigenpairs of square matrices.

Example 6.1. Consider the matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Find all eigenvalues of A, once as a matrix in $M_2(\mathbb{R})$ and once as a matrix in $M_2(\mathbb{C})$.

Definition 6.3. We say a root c of a polynomial $p(t)$ has **multiplicity** m iff $p(t) = (t-c)^m q(t)$, where m is a positive integer, and $q(t)$ is a polynomial with $q(c) \neq 0$.

Definition 6.4. Let λ be an eigenvalue of a matrix A, and let $p(t) = \det(A - tI)$ be the characteristic polynomial of A. The multiplicity of λ as a root of $p(t)$ is called the **algebraic multiplicity** of this eigenvalue λ. The dimension of $\operatorname{Ker}(A - \lambda I)$ is called the **geometric multiplicity** of λ.

Theorem 6.4. *For every eigenvalue λ of a matrix A, the geometric multiplicity of λ does not exceed its algebraic multiplicity.*

Example 6.2. Consider the matrix

$$A = \begin{pmatrix} -2 & 1 & 7 \\ -4 & 2 & 8 \\ 0 & 0 & 6 \end{pmatrix}$$

Find the spectrum of A, the algebraic and geometric multiplicity of each eigenvalue of A.

Theorem 6.5. *Let $\lambda_1, \ldots, \lambda_n$ be all eigenvalues of a matrix $A \in M_n(\mathbb{C})$. Then,*

(a) $\operatorname{tr} A = \lambda_1 + \cdots + \lambda_n$.

(b) $\det A = \lambda_1 \cdots \lambda_n$.

6.2 Examples

Example 6.3. Suppose $(\lambda, \mathbf{v})$ is an eigenpair for a square matrix A with real entries. Prove that $(\overline{\lambda}, \overline{\mathbf{v}})$ is also an eigenpair for A.

Solution. By assumption $A\mathbf{v} = \lambda\mathbf{v}$. Note that since for every two complex numbers z, w we have $\overline{zw} = \overline{z}\,\overline{w}$ and $\overline{z+w} = \overline{z} + \overline{w}$, we will obtain the following:

$$\overline{A\mathbf{v}} = \overline{\lambda\mathbf{v}} \Rightarrow \overline{A}\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}} \Rightarrow A\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}.$$

Above we use the fact that all entries of A are real and thus $\overline{A} = A$. Since $\mathbf{v}$ is nonzero, $\overline{\mathbf{v}}$ is also nonzero and thus $(\overline{\lambda}, \overline{\mathbf{v}})$ is an eigenpair of A, as desired. □

Example 6.4. Prove that the set of all eigenvectors of a linear transformation $T : V \to V$ corresponding to a fixed eigenvalue λ along with the zero vector, is a subspace of V. Prove a similar result for a square matrix.

Solution. Let W be the set of all eigenvectors of T corresponding to λ along with the zero vector. We see that $\mathbf{x} \in W$ if and only if $T(\mathbf{x}) = \lambda\mathbf{x}$ or $\mathbf{x} = \mathbf{0}$, however $T(\mathbf{0}) = \mathbf{0} = \lambda\mathbf{0}$. Therefore, $\mathbf{x} \in W$ if and only if $T(\mathbf{x}) = \lambda\mathbf{x}$, which is equivalent to $(T - \lambda I)(\mathbf{x}) = \mathbf{0}$, which is equivalent to $\mathbf{x} \in \operatorname{Ker}(T - \lambda I)$. Therefore, $W = \operatorname{Ker}(T - \lambda I)$ and hence it is a subspace of V. A similar argument works for a square matrix. □

Example 6.5. Find all scalars c for which $\lambda = 1$ is an eigenvalue of the matrix

$$A = \begin{pmatrix} 1 & c & -1 \\ c & 1 & 0 \\ 2 & 3 & -1 \end{pmatrix}$$

Solution. $\lambda = 1$ is an eigenvalue of A if and only if $\det(A - I) = 0$. This is equivalent to

$$\det \begin{pmatrix} 1-1 & c & -1 \\ c & 1-1 & 0 \\ 2 & 3 & -1-1 \end{pmatrix} = 0$$

Expanding along the second row we obtain

$$-c \det \begin{pmatrix} c & -1 \\ 3 & -2 \end{pmatrix} = 0 \Rightarrow c(-2c+3) = 0.$$

Therefore, the answer is $c = 0, 3/2$. □

Example 6.6. Show that the characteristic polynomial of an $n \times n$ matrix has degree n and its leading coefficient is $(-1)^n$.

Solution. Suppose $A = (a_{jk}) \in M_n(\mathbb{F})$. Set $A - zI = (b_{jk})$. We will use Leibniz formula for determinant.

$$\det(A - zI) = \det(b_{jk}) = \sum_{\sigma \in S_n} \pm b_{1\sigma(1)} b_{2\sigma(2)} \cdots b_{n\sigma(n)}.$$

We know $b_{jk} = a_{jk}$ iff $j \neq k$, which means those terms that $\sigma(j) \neq j$ for some j yield a polynomial of degree less than n. The only term in the above sum with $\sigma(j) = j$ for all j is

$$b_{11} b_{22} \cdots b_{nn} (a_{11} - z)(a_{22} - z) \cdots (a_{nn} - z).$$

Distributing the above product, we note that all terms have degree less than n except the term obtained by mutiplying $-z$ with itself n times. Therefore, the term with the highest degree in the polynomial $\det(A - zI)$ is $(-z)^n = (-1)^n z^n$, as desired. □

6.3 Exercises

Exercise 6.1. *Determine if each statement is true or false.*

(a) The rank of a square matrix A is equal to the number of nonzero eigenvalues of A, counting multiplicity.

(b) A square matrix A is invertible if and only if $0 \notin \sigma(A)$.

(c) Every matrix in $M_n(\mathbb{R})$ has n real eigenvalues counting (algebraic) multiplicities.

(d) Similar matrices have the same eigenvalues.

(e) For every $A, B \in M_n(\mathbb{F})$, if $\lambda \in \sigma(A)$, then $\lambda \in \sigma(AB)$.

Exercise 6.2. *For each of the following matrices:*

1. *Find $\sigma(A)$.*

2. *Find the geometric and algebraic multiplicity of each eigenvalue.*

$$\begin{pmatrix} 1 & 7 \\ -1 & 4 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 1 \\ 1 & 0 & 2 \\ 1 & 1 & 0 \end{pmatrix}.$$

Exercise 6.3. *Find all eigenvalues of the $n \times n$ matrix all of whose entries are 1.*

Exercise 6.4. *Prove that the eigenvalues of an upper triangular matrix is its diagonal entries.*

Definition 6.5. A matrix $A \in M_n(\mathbb{F})$ is called **nilpotent** iff $A^k = 0$ for some positive integer k.

Exercise 6.5. *Prove that if A is a nilpotent matrix, then $\sigma(A) = \{0\}$.*

Exercise 6.6. *Suppose $(\lambda, \mathbf{v})$ is an eigenpair for a linear transformation $T : V \to V$, and $\mathcal{A}, \mathcal{B}$ are ordered bases for V. Prove that $\det([T]_{\mathcal{B}\mathcal{A}} - \lambda[I_V]_{\mathcal{B}\mathcal{A}}) = 0$. By an example show that the matrix $[T]_{\mathcal{B}\mathcal{A}} - \lambda I$ may be invertible. (Compare this to Theorem 6.2.)*

Exercise 6.7. *Suppose $A \in M_n(\mathbb{R})$, where n is an odd integer. Prove that A has a real eigenvalue.*

Exercise 6.8. *For any list of complex numbers $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ create an $n \times n$ matrix whose list of eigenvalues is the given list $\lambda_1, \ldots, \lambda_n$.*

Exercise 6.9. *Suppose A, B are $m \times n$ and $n \times m$ matrices, respectively, where $n \leq m$. Let $p(t), q(t)$ be the characteristic polynomials of AB and BA, respectively. Prove that $p(t) = t^{m-n}q(t)$.*

Exercise 6.10. *Prove that if A is a nilpotent matrix, the matrix $I + A$ is invertible.*

6.4 Challenge Problems

Exercise 6.11. *Prove that if for a square matrix A we know $\operatorname{tr}(A^k) = 0$ for all positive integers k, then A is nilpotent.*

Exercise 6.12. *Let $a_1 = a, a_2 = a + d, \dots, a_{n^2} = a + (n^2 - 1)d$ be terms of an arithmetic sequence of length n^2. Place these numbers in entries of an $n \times n$ matrix A from top left to bottom right as follows.*

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_{n-1} & a_n \\ a_{n+1} & a_{n+2} & \cdots & a_{2n-1} & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \\ a_{n^2-2n+1} & a_{n^2-2n+2} & \cdots & a_{n^2-n-1} & a_{n^2-n} \\ a_{n^2-n+1} & a_{n^2-n+2} & \cdots & a_{n^2-1} & a_{n^2} \end{pmatrix}$$

Find the characteristic polynomial of A. Use that to find all eigenvalues of A.

Week 7

7.1 Diagonalization

Matrix operations for diagonal matrices is very easy. For example:

$$\begin{pmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_n \end{pmatrix}\begin{pmatrix} b_1 & & 0 \\ & \ddots & \\ 0 & & b_n \end{pmatrix} = \begin{pmatrix} a_1b_1 & & 0 \\ & \ddots & \\ 0 & & a_nb_n \end{pmatrix}$$

Definition 7.1. A square matrix A is called **diagonalizable** iff $A = SDS^{-1}$ for a diagonal matrix D and an invertible matrix S. The representation $A = SDS^{-1}$, where D is diagonal and S is invertible is called a **diagonalization** of A.

Theorem 7.1. *A matrix $A \in M_n(\mathbb{F})$ is diagonalizable iff there is a basis $\mathcal{B}$ of $\mathbb{F}^n$ for which all elements of $\mathcal{B}$ are eigenvectors of A. Furthermore, if $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ is a basis for $\mathbb{F}^n$, where $(\lambda_j, \mathbf{b}_j)$ is an eigenpair of A for $j = 1, \dots, n$, then $A = SDS^{-1}$, where $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, and $S = (\mathbf{b}_1 \cdots \mathbf{b}_n)$. Conversely, if $A = SDS^{-1}$ is diagonalization of A, then each diagonal entry of D is an eigenvalue of A and its corresponding column in S is a corresponding eigenvector of A.*

Example 7.1. Consider the matrices

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \text{ and } B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Prove that A is not diagonalizable over $\mathbb{R}$, but it is diagonalizable over $\mathbb{C}$. Show that B is not diagonalizable.

For every diagonal matrix $D = \operatorname{diag}(a_1, \dots, a_n)$ and every positive integer m, we have $D^m = \operatorname{diag}(a_1^m, \dots, a_n^m)$. If we define e^D using the Taylor series for e^x we get

$$e^D = \sum_{m=0}^{\infty} \frac{D^m}{m!} = \sum_{m=0}^{\infty} \operatorname{diag}\left(\frac{\lambda_1^m}{m!}, \dots, \frac{\lambda_n^m}{m!}\right) = \operatorname{diag}\left(\sum_{m=0}^{\infty} \frac{\lambda_1^m}{m!}, \dots, \sum_{m=0}^{\infty} \frac{\lambda_n^m}{m!}\right) = \operatorname{diag}\left(e^{\lambda_1}, \dots, e^{\lambda_n}\right).$$

Given $A = SDS^{-1}$, we have $A^m = SD^mS^{-1}$, so if A is diagonalizable, it is easy to evaluate its powers. We could also define e^A for every diagonalizable matrix A using Taylor series of e^x as:

$$\sum_{m=0}^{\infty} \frac{A^m}{m!} = \sum_{m=0}^{\infty} \frac{SD^mS^{-1}}{m!} = S\left(\sum_{m=0}^{\infty} \frac{D^m}{m!}\right)S^{-1} = Se^DS^{-1} = S\operatorname{diag}\left(e^{\lambda_1}, \dots, e^{\lambda_n}\right)S^{-1}.$$

Theorem 7.2. *Suppose $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are eigenvectors associated to distinct eigenvalues of a matrix A. Then $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are linearly independent.*

Corollary 7.1. If $A \in M_n(\mathbb{F})$ has n distinct eigenvalues, then A is diagonalizable.

Definition 7.2. Suppose $V_1, \ldots, V_r$ are subspaces of a vector space V.

(a) We say $V_1, \ldots, V_r$ form a basis for V iff every vector $\mathbf{v} \in V$ has a unique representation $\mathbf{v} = \sum_{j=1}^{r} \mathbf{v}_j$ with $\mathbf{v}_j \in V_j$ for $j = 1, \ldots, r$.

(b) We say $V_1, \ldots, V_r$ are linearly independent iff the only solution to $\mathbf{0} = \sum_{j=1}^{r} \mathbf{v}_j$ with $\mathbf{v}_j \in V_j$ for $j = 1, \ldots, r$ is $\mathbf{v}_j = \mathbf{0}$ for $j = 1, \ldots, r$.

Corollary 7.2. Suppose $\lambda_1, \ldots, \lambda_r$ are distinct eigenvalues of a matrix $A \in M_n(\mathbb{F})$. Then subspaces $E_j = \operatorname{Ker}(A - \lambda_j I)$ with $j = 1, \ldots, r$ are linearly independent.

Theorem 7.3. *Suppose $V_1, \ldots, V_r$ are linearly independent subspaces of a vector space V. Let $\mathcal{B}_j$ be a basis for V_j for $j = 1, \ldots, r$. Then $\bigcup_{j=1}^{k} \mathcal{B}_j$ is linearly independent. Furthermore, if $V_1, \ldots, V_r$ are a basis for V, then $\bigcup_{j=1}^{k} \mathcal{B}_j$ is a basis for V.*

Theorem 7.4. *A matrix $A \in M_n(\mathbb{F})$ is diagonalizable if and only if the characteristic polynomial of A has n (not necessarily distinct) roots and for every eigenvalue λ, the algebraic multiplicity of λ is the same as its geometric multiplicity.*

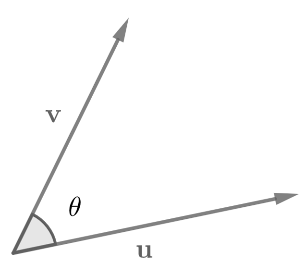
Corollary 7.3. If a matrix $A \in M_n(\mathbb{R})$ has n real eigenvalues and it can be diagonalized over $\mathbb{C}$, then it can be diagonalized over $\mathbb{R}$.

7.2 Inner Product Spaces

To better understand the geometry of vector spaces, we would like to define the notion of "angle" between vectors.

Example 7.2. Consider the vectors $\mathbf{u} = (x_1, y_1)$ and $\mathbf{v} = (x_2, y_2)$ in $\mathbb{R}^2$. Let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$. Using the law of cosines, prove that

$$x_1 x_2 + y_1 y_2 = \sqrt{x_1^2 + y_1^2}\,\sqrt{x_2^2 + y_2^2}\cos\theta.$$



Definition 7.3. An **inner product** (or **scalar product**) on a vector space V is a function that assigns a scalar $\langle \mathbf{x}, \mathbf{y} \rangle$ to every pair of vectors $\mathbf{x}, \mathbf{y} \in V$ that satisfies the following for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and all $a, b \in \mathbb{F}$:

(a) $\langle \mathbf{x}, \mathbf{x} \rangle > 0$ if $\mathbf{x} \neq \mathbf{0}$ (Positivity),

(b) $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$ (Conjugate Symmetry),

(c) $\langle a\mathbf{x} + b\mathbf{y}, \mathbf{z} \rangle = a\langle \mathbf{x}, \mathbf{z} \rangle + b\langle \mathbf{y}, \mathbf{z} \rangle$ (Linearity).

Any vector space equipped with an inner product is called an **inner product vector space** or simply an **inner product space**.

Example 7.3. The following are examples of inner product spaces:

(a) $\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle = \sum\limits_{j=1}^{n} x_j \overline{y_j}$ over $\mathbb{F}^n$.

(b) $\langle f, g \rangle = \displaystyle\int_0^1 f(t)\overline{g(t)}\, dt$ for every $f, g \in \mathbb{P}_n$.

(c) $\langle A, B \rangle = \operatorname{tr}(B^* A)$ for every $A, B \in M_{m \times n}(\mathbb{F})$.

(d) $\langle f, g \rangle = \displaystyle\int_a^b f(x)g(x)\, dx$ for every $f, g \in C[a, b]$, where $a < b$ are real numbers.

Note that by conjugate symmetry and linearity we can prove conjugate linearity with respect to the second vector:

$$\langle \mathbf{x}, a\mathbf{y} + b\mathbf{z} \rangle = \overline{a}\langle \mathbf{x}.\mathbf{y} \rangle + \overline{b}\langle \mathbf{x}, \mathbf{z} \rangle.$$

Definition 7.4. In an inner product space, the **length** of a vector $\mathbf{v}$, denoted by $||\mathbf{v}||$, is defined by:

$$||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}.$$

Definition 7.5. We say vectors $\mathbf{v}$ and $\mathbf{w}$ in an inner product space V are **orthogonal**, written as $\mathbf{v} \perp \mathbf{w}$, iff $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. We say vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ are **orthogonal** iff each pair of them are orthogonal, i.e. $\langle \mathbf{v}_j, \mathbf{v}_k \rangle = 0$ for every $j \neq k$. If in addition $||\mathbf{v}_j|| = 1$ for every j, then we say $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are **orthonormal**.

Example 7.4. Prove that 1 and t are orthogonal vectors in $\mathbb{P}_2$ under the inner product $\langle f, g \rangle = \displaystyle\int_{-1}^1 f(t)\overline{g(t)}\, dt$.

Solution. $\langle 1, t \rangle = \displaystyle\int_{-1}^1 t\, dt = \frac{t^2}{2}\Big]_{-1}^1 = 0$. Therefore, 1 and t are orthogonal. □

Theorem 7.5 (Generalized Pythagorean Theorem)**.** *Suppose* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *are orthogonal vectors in an inner product space. Then*

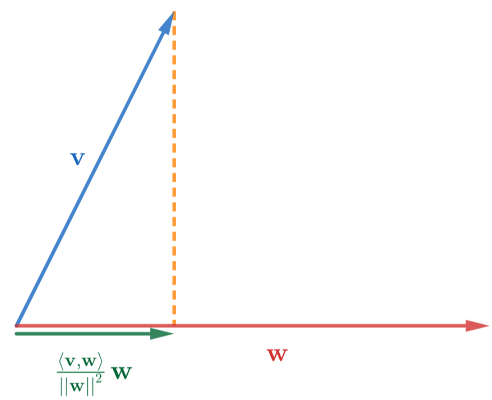
$$||\mathbf{v}_1 + \cdots + \mathbf{v}_n||^2 = ||\mathbf{v}_1||^2 + \cdots + ||\mathbf{v}_n||^2.$$

Corollary 7.4. Any set of nonzero orthogonal vectors are linearly independent.

Theorem 7.6. *Given two vectors* $\mathbf{v}, \mathbf{w}$ *in an inner product space* V *with* $\mathbf{w} \neq \mathbf{0}$*, the vector* $\dfrac{\langle \mathbf{v}, \mathbf{w} \rangle}{||\mathbf{w}||^2}\mathbf{w}$ *is the unique vector* $\mathbf{x}$ *that satisfies both of the following*

1. $\mathbf{x}$ *is a scalar multiple of* $\mathbf{w}$*, and*

2. $\mathbf{v}-\mathbf{x} \perp \mathbf{w}$.



Definition 7.6. Given two vectors $\mathbf{v}, \mathbf{w}$ in an inner product space V with $\mathbf{w} \neq \mathbf{0}$, the vector $\dfrac{\langle \mathbf{v}, \mathbf{w}\rangle}{||\mathbf{w}||^2}\mathbf{w}$ is called the (orthogonal) **projection of v onto w** and is denoted by $P_{\mathbf{w}}\mathbf{v}$.

Theorem 7.7 (Cauchy-Schwarz Inequality)**.** *For every two vectors* $\mathbf{v}, \mathbf{w}$ *in an inner product space we have*

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

Proof. First, assume $\mathbf{w} \neq \mathbf{0}$. Set $\mathbf{x} = P_{\mathbf{w}}\mathbf{v}$. Since $\mathbf{v}-\mathbf{x} \perp \mathbf{w}$ and $\mathbf{x}$ is a scalar multiple of $\mathbf{w}$, we have $\mathbf{v}-\mathbf{x} \perp \mathbf{x}$.
By the Pythagorean Theorem, $||\mathbf{v}||^2 = ||\mathbf{v}-\mathbf{x}||^2 + ||\mathbf{x}||^2 \geq ||\mathbf{x}||^2$. Note that for every scalar c we have $||c\mathbf{w}||^2 = \langle c\mathbf{w}, c\mathbf{w}\rangle = c\bar{c}\langle \mathbf{w}, \mathbf{w}\rangle = |c|^2||\mathbf{w}||^2$. Therefore, we obtain the following:

$$||\mathbf{v}||^2 \geq ||\mathbf{x}||^2 = \frac{|\langle \mathbf{v}, \mathbf{w}\rangle|^2}{|\langle \mathbf{w}, \mathbf{w}\rangle|^2}||\mathbf{w}||^2 = \frac{|\langle \mathbf{v}, \mathbf{w}\rangle|^2}{||\mathbf{w}||^4}||\mathbf{w}||^2 = \frac{|\langle \mathbf{v}, \mathbf{w}\rangle|^2}{||\mathbf{w}||^2} \Rightarrow ||\mathbf{v}|| \geq \frac{|\langle \mathbf{v}, \mathbf{w}\rangle|}{||\mathbf{w}||} \Rightarrow ||\mathbf{v}||\ ||\mathbf{w}|| \geq |\langle \mathbf{v}, \mathbf{w}\rangle|.$$

The case where $\mathbf{w} = \mathbf{0}$ is left as an exercise. □

Definition 7.7. The angle between two vectors $\mathbf{v}, \mathbf{w}$ in a real inner product space is given by

$$\theta = \cos^{-1}\left(\frac{\langle \mathbf{v}, \mathbf{w}\rangle}{||\mathbf{v}||\ ||\mathbf{w}||}\right)$$

7.3 Normed Spaces

Definition 7.8. Let V be a vector space. A function, denoted by $||\cdot||$, that assigns to every vector $\mathbf{v} \in V$ a real number $||\mathbf{v}||$ satisfying the following properties is called a **norm** on V.

(a) $||\mathbf{v}|| > 0$ for every nonzero $\mathbf{v} \in V$. (Positivity)

(b) $||c\ \mathbf{v}|| = |c|\ ||\mathbf{v}||$ for every $\mathbf{v} \in V$ and every $c \in \mathbb{F}$. (Homogeneity)

(c) $||\mathbf{v}+\mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||$ for every $\mathbf{v}, \mathbf{w} \in V$. (Triangle Inequality)

Any vector space equipped with a norm is called a **normed space** or a **normed vector space**.

Theorem 7.8. *Let V be an inner product space. Then, $||\mathbf{v}||$ defined as $\sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$ is a norm on V.*

Theorem 7.9. *A norm on a vector space V is produced from an inner product if and only if the following identity (called the parallelogram identity) holds for every $\mathbf{v}, \mathbf{w} \in V$:*

$$||\mathbf{v}+\mathbf{w}||^2 + ||\mathbf{v}-\mathbf{w}||^2 = 2||\mathbf{v}||^2 + 2||\mathbf{w}||^2.$$

7.4 Examples

Example 7.5. Diagonalize each matrix or show the matrix is not diagonalizable.

$$A = \begin{pmatrix} -1 & -2 & 2 \\ -2 & -1 & 2 \\ -2 & -2 & 3 \end{pmatrix}, B = \begin{pmatrix} -2 & -4 & 5 \\ -2 & 0 & 1 \\ -3 & -3 & 5 \end{pmatrix}, C = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Solution. $\det(A - \lambda I) = -\lambda^3 + \lambda^2 + \lambda - 1$. We guess $\lambda = 1$ as a root. After performing long division we can factor this polynomial as $(\lambda - 1)(-\lambda^2 + 1)$. Therefore, the eigenvalues of A are $1, 1, -1$.

For $\lambda = 1$ we can find the eigenvectors by solving the following:

$$\begin{pmatrix} -2 & -2 & 2 \\ -2 & -2 & 2 \\ -2 & -2 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{0} \Rightarrow -2x - 2y + 2z = 0 \Rightarrow z = x + y.$$

This yields, two linearly independent eigenvectors for $\lambda = 1$: $(1\ 0\ 1)^T$ and $(0\ 1\ 1)^T$.

For $\lambda = -1$ we can find the eigenvectors by solving the following:

$$\begin{pmatrix} 0 & -2 & 2 \\ -2 & 0 & 2 \\ -2 & -2 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{0} \Rightarrow \begin{cases} -2y + 2z = 0 \\ -2x + 2z = 0 \\ -2x - 2y + 4z = 0 \end{cases} \Rightarrow z = x = y.$$

This yields an eigenvector $(1\ 1\ 1)^T$ for $\lambda = -1$. Therefore, $A = PDP^{-1}$, where $D = \mathrm{diag}(1, 1, -1)$ and $P = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$.

$\det(B - \lambda I) = -\lambda^3 + 3\lambda^2 - 4$. By inspection a root of this polynomial can be obtained as $\lambda = -1$. After performing long division we obtain $-\lambda^3 + 3\lambda^2 - 4 = (\lambda + 1)(-\lambda^2 + 4\lambda - 4) = -(\lambda + 1)(\lambda - 2)^2$. The eigenvalues are $\lambda = -1, 2, 2$. Following the same process as before we can find an eigenvector for $\lambda = -1$. For $\lambda = 2$ we need to solve the following:

$$\begin{pmatrix} -4 & -4 & 5 \\ -2 & -2 & 1 \\ -3 & -3 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{0} \Rightarrow \begin{cases} -4x - 4y + 5z = 0 \\ -2x - 2y + z = 0 \\ -3x - 3y + 3z = 0 \end{cases}$$

After solving we obtain $z = 0$ and $y = -x$. Therefore, the eigenspace corresponding to $\lambda = 2$ is one-dimensional. This means we cannot find three linearly independent eigenvectors, which implies B is not diagonalizable.

$\det(C - \lambda I) = \lambda^2 - 4\lambda + 3$. The eigenvalues, thus, are $\lambda = 1, 3$, which are distinct and thus C is diagonalizable. After finding the eigenvectors we will get the following diagonalization of C:

$$C = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}^{-1}.$$

$\square$

Example 7.6. Suppose $(\lambda, \mathbf{v})$ is an eigenpair for a square matrix A with real entries. Prove that $(\overline{\lambda}, \overline{\mathbf{v}})$ is also an eigenpair for A.

Solution. By assumption $A\mathbf{v} = \lambda\mathbf{v}$. Note that since for every two complex numbers z, w we have $\overline{zw} = \overline{z}\,\overline{w}$ and $\overline{z + w} = \overline{z} + \overline{w}$, we will obtain the following:

$$\overline{A\mathbf{v}} = \overline{\lambda\mathbf{v}} \Rightarrow \overline{A}\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}} \Rightarrow A\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}.$$

Above we use the fact that all entries of A are real and thus $\overline{A} = A$. Since $\mathbf{v}$ is nonzero, $\overline{\mathbf{v}}$ is also nonzero and thus $(\overline{\lambda}, \overline{\mathbf{v}})$ is an eigenpair of A, as desired. $\square$

Example 7.7. Prove that a 2×2 matrix with complex entries is not diagonalizable if and only if it is similar to a matrix of the form

$$\begin{pmatrix} a & b \\ 0 & a \end{pmatrix},$$

where $a, b \in \mathbb{C}$ and $b \neq 0$.

Solution. ($\Rightarrow$) Assume A is a 2×2 matrix that is not diagonalizable. By Corollary 7.1 the two eigenvalues of A must be identical. Assume a is the only eigenvalue of A and let $\mathbf{v}$ be an eigenvector corresponding to a. Let $\mathbf{w} \in \mathbb{C}^2$ be a vector that is not a scalar multiple of $\mathbf{v}$. Since $A\mathbf{v} = a\mathbf{v}$, the matrix A in the basis $(\mathbf{v}, \mathbf{w})$ is of the form

$$\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}.$$

Since this matrix is similar to A, its only eigenvalue must be a. Therefore, $c = a$. On the other hand b cannot be zero, for otherwise A would be diagnozalizable.

($\Leftarrow$) Assume A is similar to a matrix of the form

$$\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \qquad (*)$$

where $a, b \in \mathbb{C}$ and $b \neq 0$. On the contrary assume A is diagonalizable. Since the eigenvalues of A are both a, for some invertible matrix S we must have:

$$A = S \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} S^{-1} = SaIS^{-1} = aI.$$

Therefore, aI is similar to the matrix $(*)$, and thus,

$$aI = P \begin{pmatrix} a & b \\ 0 & a \end{pmatrix} P^{-1} \Rightarrow P^{-1}aIP = \begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \Rightarrow aI = \begin{pmatrix} a & b \\ 0 & a \end{pmatrix}.$$

This implies $b = 0$, which is a contradiction. Therefore, A is not diagonalizable. □

Example 7.8. Find all scalars c for which the matrix A given below is not diagonalizable.

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

Solution. The characteristic polynomial is $(1-t)(-1-t) - 2c = t^2 - 1 - 2c$. The eigenvalues of A are then $t = \pm\sqrt{1+2c}$. If the eigenvalues are distinct, then by Corollary 7.1 the matrix A is diagonalizable. Suppose the two eigenvalues are identical. This means $1+2c = 0$, which implies $c = -1/2$. In this case the eigenvalues are both zero. If A were diagonalizable, then $A = P0P^{-1} = 0$, which is a contradiction, because A is not the zero matrix. Therefore, the answer is $c = -1/2$. □

Example 7.9. Determine which of the following matrices are similar.

$$\begin{pmatrix} 2 & -2 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 \\ 1 & 3 \end{pmatrix}, \begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 3 & 2 \end{pmatrix}$$

Solution. Let's call these matrices A, B, C, D, E in order. We note that

$$\det A = \det C = 6, \quad \det B = \det D = \det E = 1.$$

Therefore, A and C may be similar and B, D, E may be similar. We notice $\operatorname{tr} A = 4$ and $\operatorname{tr} C = 5$ are not the same. Therefore, A and C are also not similar. Thus, A and C are not similar to any of the above matrices. Note that $\operatorname{tr} B = \operatorname{tr} D = \operatorname{tr} E = 4$, so these three may be similar. The characteristic equations of B, D and E are all $z^2 - 4z + 6 = 0$ which has 2 distinct roots r, s. Thus, all matrices B, D, E are similar to the diagonal matrix $\operatorname{diag}(r, s)$. This means B, D, E are all similar. □

Example 7.10. Suppose a diagonalizable matrix A is also nilpotent. Prove that $A = 0$.

Solution. Since A is diagonalizable $A = PDP^{-1}$ for some diagonal matrix $D = \operatorname{diag}(c_1, \ldots, c_n)$. By assumption, $A^k = 0$ for some positive integer k. Therefore,

$$(PDP^{-1})^k = 0 \Rightarrow PD^kP^{-1} = 0 \Rightarrow D^k = 0.$$

This implies

$$\operatorname{diag}(c_1^k, \ldots, c_n^k) = 0 \Rightarrow c_1^k = \cdots = c_n^k = 0 \Rightarrow c_1 = \cdots = c_n = 0 \Rightarrow D = 0.$$

Therefore, $A = P0P^{-1} = 0$, as desired. $\square$

Example 7.11. Prove that if $||\cdot||$ is a norm relative to an inner product in a vector space V and $\mathbf{v}, \mathbf{w} \in V$, then

$$||\mathbf{v} + \mathbf{w}||^2 + ||\mathbf{v} - \mathbf{w}||^2 = 2(||\mathbf{v}||^2 + ||\mathbf{w}||^2).$$

Solution. By definition we have $||\mathbf{v} \pm \mathbf{w}||^2 = \langle \mathbf{v} \pm \mathbf{w}, \mathbf{v} \pm \mathbf{w} \rangle$. By linearity and conjugate symmetry this simplifies to

$$\langle \mathbf{v} \pm \mathbf{w}, \mathbf{v} \pm \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{v} \rangle \pm \langle \mathbf{v}, \mathbf{w} \rangle \pm \langle \mathbf{w}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle.$$

Summing the two together and using the fact that $\langle \mathbf{v}, \mathbf{v} \rangle = ||\mathbf{v}||^2$ and $\langle \mathbf{w}, \mathbf{w} \rangle = ||\mathbf{w}||^2$ we obtain the result. $\square$

Example 7.12. Prove that if $||\cdot||$ is a norm on a vector space V, then $||\mathbf{0}|| = 0$.

Solution. By homogeneity $||0\mathbf{0}|| = |0|\ ||\mathbf{0}|| = 0||\mathbf{0}|| = 0$. Since $0\mathbf{0} = \mathbf{0}$, we obtain $||\mathbf{0}|| = 0$, as desired. $\square$

Example 7.13. Prove that if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are vectors in normed space V, then

$$||\mathbf{v}_1 + \cdots + \mathbf{v}_n|| \leq ||\mathbf{v}_1|| + \cdots + ||\mathbf{v}_n||.$$

Solution. We will prove this by induction on n.
Basis step: For $n = 1$ both sides of the inequality are $||\mathbf{v}_1||$. This proves the basis step.
Inductive Step: Let $\mathbf{v}_1, \ldots, \mathbf{v}_{n+1}$ be vectors in V. Suppose

$$||\mathbf{v}_1 + \cdots + \mathbf{v}_n|| \leq ||\mathbf{v}_1|| + \cdots + ||\mathbf{v}_n|| \qquad (*)$$

By the Triangle Inequality we obtain:

$$||\mathbf{v}_1 + \cdots + \mathbf{v}_{n+1}|| \leq ||\mathbf{v}_1 + \cdots + \mathbf{v}_n|| + ||\mathbf{v}_{n+1}||.$$

Combining this with $(*)$ completes the inductive step. $\square$

Example 7.14. Suppose $c_1, \ldots, c_n \in \mathbb{F}$ are constants. Define a function $\langle , \rangle$ by

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle = \sum_{j=1}^{n} c_j x_j \overline{y}_j, \text{ for all } x_1, \ldots, x_n, y_1, \ldots, y_n \in \mathbb{F}.$$

(a) Show $\langle , \rangle$ satisfies linearity with respect to the first vector.

(b) Prove $\langle , \rangle$ satisfies conjugate symmetry if and only if $c_1, \ldots, c_n$ are all real.

(c) Show that $\langle , \rangle$ is an inner product if and only if $c_1, \ldots, c_n$ are all positive real numbers.

Solution. Let $\mathbf{u} = (x_1, \ldots, x_n), \mathbf{v} = (y_1, \ldots, y_n), \mathbf{w} = (z_1, \ldots, z_n)$ be in $\mathbb{F}^n$ and $\alpha, \beta \in \mathbb{F}$.

(a) We have

$$\begin{aligned} \langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w} \rangle &= \sum_{j=1}^{n} c_j(\alpha x_j + \beta y_j)\overline{z_j} \\ &= \alpha \sum_{j=1}^{n} c_j x_j \overline{z_j} + \beta \sum_{j=1}^{n} c_j y_j \overline{z_j} \\ &= \alpha\langle \mathbf{u}, \mathbf{w} \rangle + \beta\langle \mathbf{v}, \mathbf{w} \rangle \end{aligned}$$

This proves the linearity of $\langle , \rangle$ with respect to the first vector.

(b) By definition, we have

$$\begin{aligned} \overline{\langle \mathbf{u}, \mathbf{v} \rangle} &= \overline{\sum_{j=1}^{n} c_j x_j \overline{y_j}} \\ &= \sum_{j=1}^{n} \overline{c_j}, \overline{x}_j, \overline{\overline{y_j}} \qquad\qquad (*) \\ &= \sum_{j=1}^{n} \overline{c}_j y_j \overline{x}_j \\ \langle \mathbf{v}, \mathbf{u} \rangle &= \sum_{j=1}^{n} c_j y_j \overline{x}_j \end{aligned}$$

$(\Rightarrow)$ Assume $\overline{\langle \mathbf{u}, \mathbf{v} \rangle} = \langle \mathbf{v}, \mathbf{u} \rangle$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{F}^n$. Setting $\mathbf{u} = \mathbf{v} = \mathbf{e}_j$ in $(*)$, we conclude $\overline{c}_j = c_j$ and hence $c_j \in \mathbb{R}$ for $j = 1, \ldots, n$.

$(\Leftarrow)$ Assume c_j's are all real. This means $c_j = \overline{c}_j$. Hence, $(*)$ shows $\overline{\langle \mathbf{u}, \mathbf{v} \rangle} = \langle \mathbf{v}, \mathbf{u} \rangle$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{F}^n$. This completes the proof.

(c) $(\Rightarrow)$ Assume $\langle , \rangle$ defined an inner product. By the Positivity axiom of an inner product, we must have $\langle \mathbf{e}_j, \mathbf{e}_j \rangle$ is a positive real number. By definition, we obtain $c_j > 0$ for $j = 1, \ldots, n$, as desired.

$(\Leftarrow)$ Suppose $c_1, \ldots, c_n$ are all positive. By definition, $\langle \mathbf{u}, \mathbf{u} \rangle = \sum_{j=1}^{n} c_j x_j \overline{x}_j = \sum_{j=1}^{n} c_j |x_j|^2 \geq 0$, since $|x_j|^2 \geq 0$ for all j and $c_j > 0$. Each term $c_j|x_j|^2$ is positive unless $x_j = 0$. Therefore, if $\sum_{j=1}^{n} c_j|x_j|^2 = 0$, then every x_j is zero. This proves the Positivity. We have already shown $\langle , \rangle$ is linear with respect to the first vector and it satisfies conjugate symmetry. Therefore, $\langle , \rangle$ is an inner product. $\square$

Example 7.15. Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for a vector space V. For every two vectors

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n, \text{ and } \mathbf{w} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_n\mathbf{v}_n \text{ in } V,$$

define $\langle \mathbf{v}, \mathbf{w} \rangle = a_1\overline{b}_1 + a_2\overline{b}_2 + \cdots + a_n\overline{b}_n$. Prove that this defines an inner product on V.

Solution. $\langle \mathbf{v}, \mathbf{w}\rangle = a_1\bar{b}_1 + \cdots + a_n\bar{b}_n = \overline{b_1\bar{a}_1 + \cdots + b_n\bar{a}_n} = \overline{\langle \mathbf{w}, \mathbf{v}\rangle}$, which proves conjugate symmetry.

Suppose $\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, and $a, b \in \mathbb{F}$.

$$\langle a\mathbf{v} + b\mathbf{w}, \mathbf{u}\rangle = \langle \sum_{j=1}^{n}(aa_j + bb_j)\mathbf{v}_j, \sum_{j=1}^{n} c_j\mathbf{v}_j\rangle = \sum_{j=1}^{n}(aa_j + bb_j)\bar{c}_j = a\sum_{j=1}^{n} a_j\bar{c}_j + b\sum_{j=1}^{n} b_j\bar{c}_j = a\langle \mathbf{v}, \mathbf{u}\rangle + b\langle \mathbf{w}, \mathbf{u}\rangle.$$

This proves linearity.

Suppose $\mathbf{v} \neq \mathbf{0}$. Therefore, at least one a_j is nonzero, and thus $\langle \mathbf{v}, \mathbf{v}\rangle = |a_1|^2 + \cdots + |a_n|^2 > 0$, which proves the Positivity.

Therefore, $\langle , \rangle$ is an inner product. □

Example 7.16. Find the angle between 1 and t as vectors in $\mathbb{P}_1(\mathbb{R})$ under the inner product $\langle f, g\rangle = \int_0^1 f(t)g(t)\ dt$.

Solution. We see that:

$$\begin{aligned} \langle 1, t\rangle &= \int_0^1 t\ dt = \frac{1}{2} \\ \langle 1, 1\rangle &= \int_0^1 1\ dt = 1 \\ \langle t, t\rangle &= \int_0^1 t^2\ dt = 1/3 \end{aligned}$$

Therefore, $\dfrac{\langle 1, t\rangle}{||1||\ ||t||} = \dfrac{1/2}{\sqrt{1/3}} = \dfrac{\sqrt{3}}{2}$. Thus, the angle is $\dfrac{\pi}{6}$. □

Example 7.17. Prove that if $\mathbf{u}, \mathbf{v}$ are vectors in an inner product space, then

$$||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2 + 2\mathrm{Re}\,(\langle \mathbf{u}, \mathbf{v}\rangle).$$

Here, $\mathrm{Re}\, z$ is the real part of the complex number z.

Solution.

$$\begin{aligned} ||\mathbf{u} + \mathbf{v}||^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v}\rangle \\ &= \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{u}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle \\ &= ||\mathbf{u}||^2 + \langle \mathbf{u}, \mathbf{v}\rangle + \overline{\langle \mathbf{u}, \mathbf{v}\rangle} + ||\mathbf{v}||^2 \\ &= ||\mathbf{u}||^2 + ||\mathbf{v}||^2 + 2\mathrm{Re}\,(\langle \mathbf{u}, \mathbf{v}\rangle) \end{aligned}$$

The last equality is obtained from the fact that for every complex number $z = a + bi$, with $a, b \in \mathbb{R}$ we have $z + \bar{z} = a + bi + a - bi = 2a = 2\mathrm{Re}\, z$. □

7.5 Exercises

Exercise 7.1. *Consider the matrix* $A = \begin{pmatrix} 1 & 2 \\ 4 & -1 \end{pmatrix}$.

(a) Diagonalize A*, i.e. find an invertible matrix* S *and a diagonal matrix* D *for which* $A = SDS^{-1}$.

(b) Use part (a) to evaluate A^n *for every positive integer* n*. You can leave your answer as a product of three matrices, but evaluate any possible inverses in your product.*

Exercise 7.2. *Prove that if a matrix* A *is diagonalizable so is its transpose.*

Exercise 7.3. *Suppose* $A = SDS^{-1}$ *is a diagonalization of* A *and* $\lambda \in \sigma(A)$*. Prove that columns of* S *that are eigenvectors corresponding to* λ *form a basis for* $\operatorname{Ker}(A - \lambda I)$.

Exercise 7.4. *Suppose* V *is a 1-dimensional normed space with basis* $\{\mathbf{e}\}$.

(a) Prove that there is a positive real constant α *for which* $||c\,\mathbf{e}|| = \alpha|c|$ *for all* $c \in \mathbb{F}$.

(b) Prove that given a positive real constant α*, the function given by* $||c\,\mathbf{e}|| = \alpha|c|$ *defines a norm on* V.

(c) Deduce that every norm of V *is obtained from an inner product.*

Exercise 7.5 (Equality Case of the Cauchy-Schwarz Inequality)**.** *Prove that if in an inner product space for two vectors* $\mathbf{v}$ *and* $\mathbf{w}$ *we have* $|\langle \mathbf{v}, \mathbf{w}\rangle| = ||\mathbf{v}||\;||\mathbf{w}||$*, then* $\mathbf{v} = c\,\mathbf{w}$ *for some* $c \in \mathbb{F}$ *or* $\mathbf{w} = \mathbf{0}$.

Hint: Follow the proof of the Cauchy-Schwarz Inequality.

Exercise 7.6 (Equality Case of the Triangle Inequality)**.** *Suppose* $||\cdot||$ *is a norm obtained from an inner product of a vector space* V*. Prove that* $||\mathbf{v}+\mathbf{w}|| = ||\mathbf{v}||+||\mathbf{w}||$ *for two vector* $\mathbf{v}, \mathbf{w} \in V$*, if and only if* $\mathbf{v} = c\mathbf{w}$ *for some positive real number* c *or* $\mathbf{w} = \mathbf{0}$

Exercise 7.7. *Suppose* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *is a spanning set in an inner product space* V *and* $\mathbf{x} \in V$*. Prove that if* $\mathbf{x} \perp \mathbf{v}_j$ *for* $j = 1, \ldots, n$*, then* $\mathbf{x} = \mathbf{0}$.

Exercise 7.8. *Suppose* $A \in M_{m\times n}(\mathbb{R})$*. Show* $\operatorname{Ker} A$ *is the set of all vectors that are orthogonal to all rows of* A *under the usual dot product of* $\mathbb{R}^n$.

Exercise 7.9. *Suppose* $n > 1$ *is an integer. Prove the following defines a norm on* $\mathbb{C}^n$ *that cannot be obtained from an inner product:*

$$||(z_1, \ldots, z_n)|| = \max(|z_1|, \ldots, |z_n|).$$

Hint: To show this norm cannot be obtained from an inner product, show it does not satisfy the parallelogram identity.

Exercise 7.10. *Suppose* $T : V \to W$ *is an isomorphism. Assume* V *is an inner product space. Prove* W *can be turned into an inner product space under the inner product defined by:*

$$\langle \mathbf{x}, \mathbf{y}\rangle = \langle T^{-1}(\mathbf{x}), T^{-1}(\mathbf{y})\rangle.$$

Exercise 7.11. *Suppose $\langle , \rangle_1$ and $\langle , \rangle_2$ are inner products for a vector space V. Let c be a positive scalar. Prove the following define inner products on V:*

(a) $\langle \mathbf{x}, \mathbf{y} \rangle = c\langle \mathbf{x}, \mathbf{y} \rangle_1$.

(b) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle_1 + \langle \mathbf{x}, \mathbf{y} \rangle_2$.

Exercise 7.12 (Reverse Triangle Inequlity)**.** *Prove in every normed space for every two vectors* $\mathbf{v}, \mathbf{w}$*:*

$$||\mathbf{v} - \mathbf{w}|| \geq |\, ||\mathbf{v}|| - ||\mathbf{w}||\, |.$$

Exercise 7.13 (Polarization Identities)**.** *Assume* $\mathbf{x}, \mathbf{y}$ *are vectors in an inner product space. Prove the following:*

(a) $\langle \mathbf{x}, \mathbf{y} \rangle = \dfrac{1}{4}\left(||\mathbf{x}+\mathbf{y}||^2 - ||\mathbf{x}-\mathbf{y}||^2\right)$, *if V is a real vector space.*

(b) $\langle \mathbf{x}, \mathbf{y} \rangle = \dfrac{1}{4}\sum_{c=\pm 1, \pm i} c\, ||\mathbf{x}+c\mathbf{y}||^2$, *if V is a complex vector space.*

Exercise 7.14. *Consider the matrix*

$$A = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}.$$

Prove there are precisely four distinct matrices B for which $B^2 = A$.

Exercise 7.15. *Suppose $A \in M_n(\mathbb{C})$ has n distinct eigenvalues. Prove there are either precisely 2^{n-1} or precisely 2^n distinct matrices $B \in M_n(\mathbb{C})$ for which $B^2 = A$.*

Exercise 7.16. *Prove vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ in an inner product space are linearly independent if and only if the $n \times n$ matrix whose (j,k) entry, for all $1 \leq j,k \leq n$, is $\langle \mathbf{v}_j, \mathbf{v}_k \rangle$ is invertible.*

Exercise 7.17. *Suppose $n > 1$ is an integer. Prove the following defines a norm on $\mathbb{C}^n$ that cannot be obtained from an inner product:*

$$||(z_1, \dots, z_n)|| = |z_1| + \cdots + |z_n|.$$

Exercise 7.18. *Consider the vector space $M_2(\mathbb{C})$ equipped with the inner product given by $\langle A, B \rangle = \operatorname{tr}(B^* A)$. Find the orthogonal projection of* $A = \begin{pmatrix} 1 & 2 \\ -i & 0 \end{pmatrix}$ *onto* $B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

Week 8

8.1 Orthogonal Projections and Orthogonal Bases

Definition 8.1. We say a vector $\mathbf{v}$ is **orthogonal** to a subspace W of an inner product space V iff $\mathbf{v} \perp \mathbf{w}$, for every $\mathbf{w} \in W$. We say two subspaces W_1 and W_2 of V are orthogonal iff $\mathbf{w}_1 \perp \mathbf{w}_2$ for every $\mathbf{w}_1 \in W_1$ and every $\mathbf{w}_2 \in W_2$.

Theorem 8.1. *Suppose W is a subspace of an inner product space V. Suppose $\mathbf{w}_1, \ldots, \mathbf{w}_n$ form an orthogonal basis for W. Then, for every $\mathbf{v} \in V$, there is a unique vector $\mathbf{w} \in W$ for which $\mathbf{v} - \mathbf{w} \perp W$. Furthermore, this vector $\mathbf{w}$ is given by*

$$\mathbf{w} = \sum_{j=1}^{n} \frac{\langle \mathbf{v}, \mathbf{w}_j \rangle}{||\mathbf{w}_j||^2} \mathbf{w}_j.$$

We would like to call the vector $\mathbf{w}$ in the above theorem the orthogonal projection of $\mathbf{v}$ onto W. Before we can do so, we need to prove such a vector $\mathbf{w}$ exists. In Theorem 7.6 we discussed the projection of a vector $\mathbf{v}$ onto a vector $\mathbf{w}$, i.e. $P_{\mathbf{w}}\mathbf{v}$. In order to prove the existence of the projection of a vector $\mathbf{v}$ onto an arbitrary subspace W using Theorem 8.1, we need to prove W has an orthogonal basis. This is achieved in the following theorem.

Theorem 8.2 (Gram-Schmidt Orthogonalization Process)**.** *Suppose $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form a basis for an inner product space V. Define vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ recursively as follows:*

$$\begin{aligned}
\mathbf{w}_1 &= \mathbf{v}_1 \\
\mathbf{w}_2 &= \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 \\
\\
\mathbf{w}_3 &= \mathbf{v}_3 - \frac{\langle \mathbf{v}_3, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \frac{\langle \mathbf{v}_3, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 \\
&\vdots \\
\mathbf{w}_n &= \mathbf{v}_n - \frac{\langle \mathbf{v}_n, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \frac{\langle \mathbf{v}_n, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 - \cdots - \frac{\langle \mathbf{v}_n, \mathbf{w}_{n-1} \rangle}{\langle \mathbf{w}_{n-1}, \mathbf{w}_{n-1} \rangle}\mathbf{w}_{n-1}
\end{aligned}$$

Then $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ form an orthogonal basis for V.

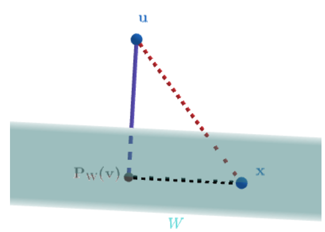
Definition 8.2. A basis of orthogonal vectors for an inner product space is called an **orthogonal basis**. Similarly, a basis of orthonormal vectors for an inner product space is called an **orthonormal basis**.

Corollary 8.1. Every finite dimensional inner product space has an orthonormal basis.

Corollary 8.2. Let W be a finite dimensional subspace of an inner product space V. Then, for every $\mathbf{v} \in V$ there is a unique vector $\mathbf{w} \in W$ for which $\mathbf{v} - \mathbf{w} \perp W$.

Definition 8.3. The **orthogonal projection** of a vector $\mathbf{v} \in V$ onto a finite dimensional subspace W of V, denoted by $P_W(\mathbf{v})$, is the unique vector $\mathbf{w} \in W$ for which $\mathbf{v} - \mathbf{w} \perp W$.

Theorem 8.3. *Suppose W is a finite dimensional subspace of an inner product space V, then for every $\mathbf{v} \in V$, we have $||P_W(\mathbf{v}) - \mathbf{v}|| \leq ||\mathbf{x} - \mathbf{v}||$ for every $\mathbf{x} \in W$. Furthermore, equality occurs if and only if $\mathbf{x} = P_W(\mathbf{v})$.*



Definition 8.4. The **orthogonal complement** of a subspace E of a vector space V, denoted by $E^\perp$, is the set consisting of all vectors $\mathbf{v} \in V$ for which $\mathbf{v} \perp E$.

Theorem 8.4. *Suppose E is a subspace of a finite dimensional inner product space V. Then,*

(a) $E^\perp$ *is a subspace of* V.

(b) $\dim E + \dim E^\perp = \dim V$.

(c) $(E^\perp)^\perp = E$.

Theorem 8.5. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form an orthogonal basis for a vector space V, then every $\mathbf{v} \in V$ can be written as*

$$\mathbf{v} = \sum_{j=1}^{n} \frac{\langle \mathbf{v}, \mathbf{v}_j \rangle}{||\mathbf{v}_j||^2} \mathbf{v}_j.$$

8.2 Least Square Solution

We want to find the best solution $\mathbf{x}$ to the equation $A\mathbf{x} = \mathbf{b}$. In other words, we would like to find $\mathbf{x}$ for which $||A\mathbf{x} - \mathbf{b}||$ is minimized. This means we are looking for $\mathbf{u} \in \mathrm{Col}\,(A)$ that is closest to $\mathbf{b}$. This is precisely the projection of $\mathbf{b}$ onto $\mathrm{Col}\,(A)$. This means we need $A^*(A\mathbf{x} - \mathbf{b}) = \mathbf{0}$. This gives us the **normal equation** $A^*A\mathbf{x} = A^*\mathbf{b}$.

Theorem 8.6. $\mathrm{Ker}\,(A^*A) = \mathrm{Ker}\,A$.

Corollary 8.3. Let $A \in M_{m\times n}(\mathbb{F})$. The matrix A^*A is invertible if and only if $\mathrm{rank}\,A = n$.

Example 8.1 (Line of best fit)**.** Given a data set $(x_1.y_1), (x_2, y_2), \ldots, (x_n, y_n)$ with $n \geq 2$ and distinct x_j's, we want to find a line $y = a + bx$ that best describes this data set. In other words, we would like to minimize $||(y_1 - a - bx_1, \ldots, y_n - a - bx_n)||$. This is the same as finding the least square solution to the system:

$$\begin{pmatrix} a + bx_1 \\ a + bx_2 \\ \vdots \\ a + bx_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \text{ i.e. } \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

Example 8.2. We want to estimate the data set $(x_1.y_1), (x_2, y_2), \ldots, (x_n, y_n)$ with a parabola $a + bx + cx^2$. We need to find the least square solution to

$$\begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & & \\ 1 & x_n & x_n^2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

Definition 8.5. The **least square solution** to an equation $A\mathbf{x} = \mathbf{b}$, where $A \in M_{m \times n}(\mathbb{F})$ and $\mathbf{b} \in \mathbb{F}^m$, is a vector $\mathbf{x} \in \mathbb{F}^n$ for which $||A\mathbf{x} - \mathbf{b}||$ is miminized.

8.3 Adjoint of a Linear Transformation

Definition 8.6. The **conjugate** of a matrix $A \in M_{m \times n}(\mathbb{F})$, denoted by $\overline{A}$, is an $m \times n$ matrix whose (j, k) entry is the complex conjugate of the (j, k) entry of A for all $j = 1, \ldots, m$ and $k = 1, \ldots, n$. The **adjoint** or **conjugate transpose** or **Hermitian** of A, denotes by A^*, is the matrix $\overline{A}^T$.

Remark 8.1. The standard inner product $\mathbf{x} \cdot \mathbf{y}$ of two column vectors $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$ is given by $\mathbf{y}^*\mathbf{x}$, where $\mathbf{y}^*$ is the transpose conjugate of $\mathbf{y}$.

Theorem 8.7. *Let $A \in M_{m \times n}(\mathbb{F})$. A matrix $B \in M_{m \times n}(\mathbb{F})$ is the adjoint of A if and only if $\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, B\mathbf{y} \rangle$.*

Theorem 8.8. *Let $T : V \to W$ be a linear transformation between inner product spaces. Then, there is a unique linear transformation $S : W \to V$ satisfying:*

$$\langle T(\mathbf{x}), \mathbf{y} \rangle = \langle \mathbf{x}, S(\mathbf{y}) \rangle, \ \textit{for all } \mathbf{x} \in V, \mathbf{y} \in W.$$

As seen in the proof of the above theorem, if $\mathcal{A}$ and $\mathcal{B}$ are orthonormal bases for V and W, respectively, then $[T^*]_{\mathcal{AB}} = ([T]_{\mathcal{BA}})^*$.

Definition 8.7. Suppose $T : V \to W$ be a linear transformation between inner product spaces. Then, the unique linear transformation $S : W \to V$ satisfying:

$$\langle T(\mathbf{x}), \mathbf{y} \rangle = \langle \mathbf{x}, S(\mathbf{y}) \rangle, \ \text{for all } \mathbf{x} \in V, \mathbf{y} \in W,$$

is called the **adjoint of** T and is denoted by T^*.

Theorem 8.9. *Given two matrices A, B and a scalar c we have*

(a) $(A+B)^* = A^* + B^*$

(b) $(AB)^* = B^* A^*$

(c) $(cA)^* = \overline{c}\, A^*$

(d) $(A^*)^* = A$

as long as the appropriate operation is defined. Similar properties are true for linear transformations between inner product spaces.

Theorem 8.10. *Given a matrix A we have*

(a) $(\mathrm{Ker}\, A)^{\perp} = \mathrm{Col}\,(A^*)$

(b) $(\mathrm{Col}\, A)^{\perp} = \mathrm{Ker}\,(A^*)$

8.4 Isometries

Definition 8.8. A linear transformation $T : V \to W$ between inner product spaces is called an **isometry** iff it preserves the norm, i.e. $||U(\mathbf{x})|| = ||\mathbf{x}||$ for all $\mathbf{x} \in V$.

Theorem 8.11. *Suppose $T : V \to W$ is a linear transformation between inner product spaces. T is an isometry if and only if T preserves the inner product, i.e.*

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle T(\mathbf{x}), T(\mathbf{y}) \rangle, \qquad \forall \mathbf{x}, \mathbf{y} \in V.$$

Theorem 8.12. *Suppose $T : V \to W$ is an isometry. If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are orthogonal (resp., orthonormal) vectors in V, then $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ are orthogonal (resp., orthonormal) vectors in W.*

Theorem 8.13. *A linear transformation $T : V \to W$ between inner product spaces is an isometry if and only if $T^* \circ T = I_V$.*

Definition 8.9. An isometry is called **unitary** iff it is invertible.

Theorem 8.14. *An isometry $T : V \to W$ is unitary iff $\dim V = \dim W$.*

Corollary 8.4. A linear transformation $T : V \to W$ between inner product spaces is unitary if and only if $T^{-1} = T^*$. A similar result hold for square matrices.

Definition 8.10. A square matrix U is called **unitary** iff $U^*U = I$, i.e. $U^{-1} = U^*$. A unitary matrix with real entries is called an **orthogonal** matrix.

Theorem 8.15. *Suppose $T : V \to W$ is a linear transformation between inner product spaces. If T is unitary, then so is T^{-1}. Similarly, if a matrix is unitary, then, so is its inverse.*

Proof. By assumption $T^{-1} = T^*$. In order to show T^{-1} is unitary, we need to show $(T^{-1})^{-1} = (T^{-1})^*$. We know $(T^{-1})^{-1} = T$. We also see that $(T^{-1})^* = (T^*)^* = T$. Therefore, $(T^{-1})^{-1} = (T^{-1})^*$, as desired. $\square$

Theorem 8.16. *Suppose $T : V \to V$ is a unitary transformation, and $\lambda \in \sigma(T)$. Then,*

(a) $|\lambda| = 1$.

(b) $|\det T| = 1$.

A similar result holds for square matrices.

Proof. (a) By definition of an eigenvalue, $T(\mathbf{v}) = \lambda\mathbf{v}$ for some nonzero vector $\mathbf{v}$. Since T is unitary, it is an isometry and hence, $||T(\mathbf{v})|| = ||\mathbf{v}||$. Therefore,

$$||\mathbf{v}|| = ||\lambda\mathbf{v}|| = |\lambda|\,||\mathbf{v}|| \Rightarrow |\lambda| = 1, \text{ since } \mathbf{v} \neq \mathbf{0}.$$

(b) Let A be a matrix of T in some basis of V. By Theorem 6.5, $\det T = \lambda_1 \cdots \lambda_n$, where $\lambda_1, \ldots, \lambda_n$ are all eigenvalues of A (including multiplicities.) We now take the absolute value of both sides and use part (a) to obtain:

$$|\det T| = |\lambda_1| \cdots |\lambda_n| = 1.$$

$\square$

8.5 Examples

Example 8.3. Find the orthogonal projection of t^3 onto the subspace of $\mathbb{P}_3$ spanned by 1 and t^2. Use the inner product $\langle p, q\rangle = \int_0^1 p(t)\overline{q(t)}\ dt$

Solution. First, we need to find an orthogonal basis for $E = span\{1, t^2\}$. For that we will use the Gram-Schmidt process.

$$\mathbf{v}_1 = 1, \mathbf{v}_2 = t^2 - \frac{\langle t^2, 1\rangle}{\langle 1, 1\rangle}1.$$

We see $\langle t^2, 1\rangle = \int_0^1 t^2\ dt = \frac{1}{3}$, and $\langle 1, 1\rangle = \int_0^1 1\ dt = 1$. This yields $\mathbf{v}_2 = t^2 - \frac{1}{3}$. The projection of t^3 onto E is

$$\frac{\langle t^3, 1\rangle}{\langle 1, 1\rangle}1 + \frac{\langle t^3, (t^2 - 1/3)\rangle}{\langle (t^2 - 1/3), (t^2 - 1/3)\rangle}(t^2 - 1/3).$$

$$\langle t^3, 1\rangle = \int_0^1 t^3\ dt = \frac{1}{4}.$$

$$\langle t^3, (t^2 - 1/3)\rangle = \int_0^1 t^5 - \frac{t^3}{3}\ dt = \frac{1}{6} - \frac{1}{12} = \frac{1}{12}.$$

$$\langle (t^2 - 1/3), (t^2 - 1/3)\rangle = \int_0^1 (t^2 - 1/3)^2\ dt = \int_0^1 t^4 + 1/9 - 2t^2/3\ dt = \frac{1}{5} + \frac{1}{9} - \frac{2}{9} = \frac{4}{45}.$$

Therefore,

$$P_E(t^3) = \frac{1}{4} + \frac{45}{48}(t^2 - 1/3)$$

$\square$

Example 8.4. Prove that $P_E \circ P_E = P_E$ for every subspace E of an inner product space.

Solution. Let $\mathbf{v} \in V$, the vector $P_E(\mathbf{v})$ is in E and $P_E(\mathbf{v}) - P_E(\mathbf{v})$ is orthogonal to E. Thus, $P_E(P_E(\mathbf{v})) = P_E(\mathbf{v})$. Therefore, $P_E \circ P_E = P_E$, as desired. $\square$

Example 8.5. Let E_1 and E_2 be two subspaces of an inner product space V for which the projection of every $\mathbf{v} \in V$ onto E_1 and E_2 are the same. Prove $E_1 = E_2$.

Solution. Let $\mathbf{v} \in E_1$. By definition $P_{E_1}(\mathbf{v}) = \mathbf{v}$. Since $P_{E_1} = P_{E_2}$, we have $P_{E_2}(\mathbf{v}) = \mathbf{v}$. By definition, $P_{E_2}(\mathbf{v}) \in E_2$. Thus, $\mathbf{v} \in E_2$. Therefore, $E_1 \subseteq E_2$. Similarly, $E_2 \subseteq E_1$. Therefore, $E_1 = E_2$. $\square$

Example 8.6. Suppose E is a subspace of an inner product vector space V. Let P, Q be projections onto E and $E^\perp$, respectively.

(a) Prove that $P \circ Q(\mathbf{v}) = \mathbf{0}$, for every vector $\mathbf{v} \in V$.

(b) Prove that $P + Q$ is the identity transformation.

(c) Prove $(P - Q)^{-1} = P - Q$.

Solution. (a) $Q(\mathbf{v})$ is a vector in $E^\perp$. Thus, $Q(\mathbf{v}) - \mathbf{0}$ is orthogonal to E. Since $\mathbf{0} \in E$, we have $P(Q(\mathbf{v})) = \mathbf{0}$.
(b) By definition, $\mathbf{v} - P(\mathbf{v})$ is orthogonal to E. Thus, $\mathbf{v} - P(\mathbf{v}) \in E^\perp$. On the other hand $\mathbf{v} - (\mathbf{v} - P(\mathbf{v})) = P(\mathbf{v})$ is orthogonal to $E^\perp$. Therefore, $\mathbf{v} - P(\mathbf{v}) = Q(\mathbf{v})$. Thus, $P(\mathbf{v}) + Q(\mathbf{v}) = \mathbf{v}$, which implies $P + Q$ is the identity transformation.
(c) By linearity, $(P - Q) \circ (P - Q) = P \circ P + Q \circ Q + P \circ Q + Q \circ P = P + Q$. Here we used the fact that $P \circ P = P$, $P \circ Q = Q \circ P = 0$. Therefore, $P - Q$ is its own inverse. $\square$

Example 8.7. Consider the subspace V of $\mathbb{F}^4$ spanned by $\mathbf{v} = (1, 2, 0, 1)$ and $\mathbf{w} = (1, -1, 1, 2)$. Find a basis for the orthogonal complement of V relative to the standard inner product.

Solution. Note that since $\mathbf{v}$ and $\mathbf{w}$ are not multiples of each other, $\dim V = 2$. By Theorem 8.4, we have $\dim V^\perp = 4 - 2 = 2$.

We will find a basis for $\mathbb{F}^4$ containing $\mathbf{v}$ and $\mathbf{w}$. To do that, we will place these vectors in rows of a matrix, and row reduce the matrix as below:

$$\begin{pmatrix} 1 & 2 & 0 & 1 \\ 1 & -1 & 1 & 2 \end{pmatrix} \xrightarrow{R_2 - R_1} \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & -3 & 1 & 1 \end{pmatrix}$$

Therefore, by adding $\mathbf{e}_3$ and $\mathbf{e}_4$ to the rows of this matrix, we obtain a matrix in echelon form. Thus, $\mathbf{v}, \mathbf{w}, \mathbf{e}_3, \mathbf{e}_4$ form a basis for $\mathbb{F}^4$. Now, we will apply the Gram-Schmidt process.

$$\begin{aligned}
\mathbf{w}_1 &= \mathbf{v} \\
\mathbf{w}_2 &= \mathbf{w} - \frac{\langle \mathbf{w}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v} \\
\mathbf{w}_3 &= \mathbf{e}_3 - \frac{\langle \mathbf{e}_3, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \frac{\langle \mathbf{e}_3, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 \\
\mathbf{w}_4 &= \mathbf{e}_4 - \frac{\langle \mathbf{e}_4, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \frac{\langle \mathbf{e}_4, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 - \frac{\langle \mathbf{e}_4, \mathbf{w}_3 \rangle}{\langle \mathbf{w}_3, \mathbf{w}_3 \rangle}\mathbf{w}_3
\end{aligned}$$

The vectors $\mathbf{w}_3, \mathbf{w}_4$ are linearly independent and are in $V^\perp$. Since $\dim V^\perp = 2$, the two vectors $\mathbf{w}_3$ and $\mathbf{w}_4$ form a basis for $V^\perp$. (The calculation must be done!) □

Example 8.8. Suppose E, F are subspaces of an inner product space V for which $P_E \circ P_F = 0$, the zero function. Prove $E \perp F$. Conversely, prove that if $E \perp F$, then $P_E \circ P_F = 0$.

Solution. ($\Rightarrow$) Let $\mathbf{x} \in F$. Since $\mathbf{x} \in F$ and $\mathbf{x} - \mathbf{x} \perp F$, we have $P_F(\mathbf{x}) = \mathbf{x}$. By assumption $P_E(\mathbf{x}) = \mathbf{0}$. Therefore, by definition of projection $\mathbf{x} - \mathbf{0} \perp E$. Thus, every element of F is orthogonal to E, which means $E \perp F$.

($\Leftarrow$) Suppose $E \perp F$. Let $\mathbf{x} \in V$. We know $P_F(\mathbf{x}) \in F$, by definition of P_F. Therefore, $P_F(\mathbf{x}) - \mathbf{0} = P_F(\mathbf{x}) \perp E$, which implies $P_E(P_F(\mathbf{x})) = \mathbf{0}$, as desired. □

Example 8.9. Suppose $A \in M_n(\mathbb{F})$ is self adjoint. Prove that for every $S \in M_n(\mathbb{F})$ the matrix S^*AS is self adjoint.

Solution. By properties of adjoint and the fact that A is self adjoint, we have $(S^*AS)^* = S^*A^*(S^*)^* = S^*A^*S = S^*AS$. Therefore, S^*AS is self adjoint. □

8.6 Exercises

Exercise 8.1. *Determine if each statement is true or false.*

(a) If a square matrix has determinant 1, then it is unitary.

(b) Any matrix corresponding to an isometry transformation has a left inverse.

(c) Any matrix corresponding to a unitary transformation is invertible.

(d) If $T : V \to W$ is a linear transformation and T sends some orthonormal basis of V to an orthonormal basis of W, then T is unitary.

(e) If $T : V \to V$ is a linear transformation of an inner product space V and $||T(\mathbf{v}_j)|| = ||\mathbf{v}_j||$ for some orthogonal basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V, then T is unitary.

Exercise 8.2. *Find the shortest distance from a point* $(x_0, y_0, z_0) \in \mathbb{R}^3$ *to the plane* $ax_1 + bx_2 + cx_3 = 0$*, where* $(a, b, c) \in \mathbb{R}^3$ *is a fixed nonzero vector.*

Hint: Show the given plane is the orthogonal complement of the span of (a, b, c). Use that to find the projection of (x_0, y_0, z_0) onto the plane.

Exercise 8.3. *Prove that every projection is its own adjoint.*

Hint: Consider $\langle P_E(\mathbf{x}) - \mathbf{x}, P_E(\mathbf{y})\rangle$.

Exercise 8.4. *Find the least square solution to the system*

$$\begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 2 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

Exercise 8.5. *Find the equation of the line of best fit for the data set* $(x_1, y_1), \ldots, (x_n, y_n)$*, where* $n \geq 2$ *and* x_j*'s are distinct.*

Exercise 8.6. *Suppose a linear transformation* $T : V \to V$ *is its own adjoint (aka self adjoint) and* $T \circ T = I_V$*. Prove* T *is an orthogonal projection.*

Exercise 8.7. *Find an orthonormal basis for* $M_{m\times n}(\mathbb{F})$ *under the inner product* $\langle A, B\rangle = \mathrm{tr}\,(B^*A)$*.*

Exercise 8.8. *Prove that the product of two unitary matrices of the same size is unitary.*

Exercise 8.9. *Prove that if* $U \in M_2(\mathbb{R})$ *is orthogonal with* $\det U = 1$*, then* U *is a rotation matrix.*

Exercise 8.10. *Prove there is no square matrix* A *for which* $A^* = A + I$*.*

Exercise 8.11. *Consider* $\mathbb{P}_n(\mathbb{R})$ *equipped with the inner product* $\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)\ dt$*. Prove that the orthogonal complement of the subspace of even polynomials, (i.e. the subspace spanned by* $1, t^2, t^4, \ldots, t^{2\lfloor n/2\rfloor}$*) is the subspace of odd polynomials (i.e. the subspace spanned by* $t, t^3, \ldots, t^{2\lceil n/2\rceil-1}$*).*

Exercise 8.12. *Prove that the adjoint of an elementary matrix is an elementary matrix of the same type.*

Exercise 8.13. *Consider the vector space* $M_2(\mathbb{C})$ *with the inner product given by* $\langle A, B\rangle = \mathrm{tr}\,(B^*A)$*. Find the orthogonal projection of* $A = \begin{pmatrix} 1 & i \\ 1 & 0 \end{pmatrix}$ *onto the subspace of symmetric matrices.*

Exercise 8.14. *Suppose* X*,* Y *are inner product spaces of the same dimension. Prove there is an isomorphism* $T : X \to Y$ *for which*

$$\langle T(\mathbf{v}), T(\mathbf{w})\rangle = \langle \mathbf{v}, \mathbf{w}\rangle \ \forall\ \mathbf{v}, \mathbf{w} \in X.$$

Exercise 8.15. *Suppose* X, Y *are normed spaces of the same dimension. Is it true that there must be an isomorphism* $T : X \to Y$ *for which*

$$||T(\mathbf{v})|| = ||\mathbf{v}|| \ \forall\ \mathbf{v} \in X.$$

Exercise 8.16. *Consider the vector space $C[-\pi,\pi]$ equipped with the inner product $\langle f,g\rangle = \int_{-\pi}^{\pi} f(x)g(x)\,dx$. Prove the following list is an orthonormal list of functions in $C[-\pi,\pi]$:*

$$\frac{\sin(x)}{\sqrt{\pi}}, \frac{\cos(x)}{\sqrt{\pi}}, \frac{\sin(2x)}{\sqrt{\pi}}, \frac{\cos(2x)}{\sqrt{\pi}}, \frac{\sin(3x)}{\sqrt{\pi}}, \frac{\cos(3x)}{\sqrt{\pi}}, \ldots.$$

Exercise 8.17. *Let W be the subspace of $C[-1,1]$ spanned by the functions*

$$\sin(x), \cos(x), \sin(2x), \cos(2x), \sin(3x), \cos(3x), \ldots.$$

Prove the function $f \in C[-1,1]$ defined by $f(x) = x$ does not have an orthogonal projection onto W.

Week 9

9.1 Unitary Equivalent Matrices

Definition 9.1. Two matrices A, B are said to be **unitarily equivalent** iff $A = UBU^*$, where U is a unitary matrix. Similarly, two linear transformations $S, T : X \to X$ of an inner product space X are said to be unitarily equivalent iff there is a unitary transformation U for which $S = U \circ T \circ U^*$.

Remark 9.1. Note that two unitarily equivalent matrices must be square of the same size. Furthermore, since $U^* = U^{-1}$, two unitarily equivalent matrices are similar.

Definition 9.2. A square matrix is said to be **unitarily diagonalizable** iff it is unitarily equivalent to a diagonal matrix. Similarly, a linear transformation $L : V \to V$ of an inner product space is called unitarily diagonalizable iff the matrix of T relative to an orthonormal basis is diagonal.

Theorem 9.1. *A matrix $A \in M_n(\mathbb{F})$ is unitarily diagonalizable if and only if $\mathbb{F}^n$ has an orthogonal basis of eigenvectors. Furthermore, if A is unitarily diagonalizable, then $\mathbb{F}^n$ has an orthonormal basis of eigenvector.*

Example 9.1. Determine if the matrix is unitarily diagonalizable:

$$\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

Theorem 9.2. *If a matrix $A \in M_n(\mathbb{F})$ is unitarily diagonalizable, and $\mathbf{u}, \mathbf{v}$ are eigenvectors of A corresponding to distinct eigenvalues, then $\mathbf{u} \perp \mathbf{v}$.*

9.2 Rigid Motions

Definition 9.3. A **rigid motion** of an inner product space V is a distance preserving function $f : V \to V$, i.e.

$$||f(\mathbf{x}) - f(\mathbf{y})|| = ||\mathbf{x} - \mathbf{y}|| \ \forall \mathbf{x}, \mathbf{y} \in V.$$

Note that a rigid motion is not assumed to be linear.

Example 9.2. Suppose $U : V \to V$ is a unitary transformation and $\mathbf{v} \in V$. Then, $f : V \to V$ defined by $f(\mathbf{x}) = U(\mathbf{x}) + \mathbf{v}$ is a rigid motion.

Theorem 9.3. *Suppose $f : V \to V$ is a rigid motion of a real inner product space V. Then, the function $T : V \to V$ defined by $T(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{0})$ is orthogonal.*

9.3 Schur's Theorem

Theorem 9.4. *Suppose a matrix $A \in M_n(\mathbb{F})$ has n eigenvalues. Then, there is an upper triangular matrix T and a unitary matrix U for which $A = UTU^*$. Consequently, for every $A \in M_n(\mathbb{C})$, there is an upper triangular matrix T and a unitary matrix U for which $A = UTU^*$. Similarly, if $L : V \to V$ is a linear transformation on an inner product n-dimensional space V has n eigenvalues. Then, there is an ordered orthonormal basis $\mathcal{A}$ of V for which the matrix $[L]_{\mathcal{A}\mathcal{A}}$ is upper triangular.*

9.4 Self-Adjoint and Normal Linear Transformations

Definition 9.4. A square matrix is called **self-adjoint** iff it is its own adjoint. Similarly, a linear transformation $L : V \to V$ on an inner product space V is called self-adjoint iff $T = T^*$.

Theorem 9.5. *Suppose $L : V \to V$ is a self-adjoint linear transformation. Then, all eigenvalues of L are real. Similarly, all eigenvalues of a self-adjoint matrix is real.*

Question. When is a matrix (or linear transformation) unitarily diagonalizable?

Definition 9.5. A square matrix N is called **normal** iff $NN^* = N^*N$. Similarly, a linear transformation $N : V \to V$ on an inner product space V is called normal iff $NN^* = N^*N$.

Theorem 9.6. *If $A \in M_n(\mathbb{F})$ has n eigenvalues, then it is normal if and only if it is unitarily diagonalizable. Similarly, if a linear transformation $N : V \to V$ on an inner product n-dimensional space V has n eigenvalues. Then, N is normal if and only if it is unitarily diagonalizable.*

Corollary 9.1. Every self adjoint matrix is unitarily equivalent to a diagonal matrix with real entries.

Theorem 9.7. *A matrix $N \in M_n(\mathbb{F})$ is normal if and only if*

$$||N\mathbf{x}|| = ||N^*\mathbf{x}|| \qquad \forall \mathbf{x} \in \mathbb{F}^n.$$

Similarly, a linear transformation $N : V \to V$ on an inner product space V is normal if and only if

$$||N(\mathbf{x})|| = ||N^*(\mathbf{x})|| \qquad \forall \mathbf{x} \in V.$$

9.5 Examples

Example 9.3. Prove that a normal matrix A is self-adjoint if and only if all of its eigenvalues are real.

Solution. $\Rightarrow$: Since A is self-adjoint, by a theorem all of its eigenvalues are real.

$\Leftarrow$: Since A is normal, there is a unitary matrix U and a diagonal matrix D for which $A = UDU^*$. Since A and D are similar, they have the same eigenvalues. Since D is diagonal, its diagonal entries are its eigenvalues. Therefore, diagonal entries of D are all real, and hence $D^* = D$. We now see that

$$A^* = (UDU^*)^* = (U^*)^*D^*U^* = UD^*U^* = UDU^* = A.$$

Therefore, A is self-adjoint. $\square$

Example 9.4. Consider the complex vector space $\mathbb{C}$. By an example show that there is a rigid motion $f : \mathbb{C} \to \mathbb{C}$ for which $T : \mathbb{C} \to \mathbb{C}$ given by $T(z) = f(z) - f(0)$ is not linear.

Solution. Let $f : \mathbb{C} \to \mathbb{C}$ be given by $f(z) = \overline{z}$ for all $z \in \mathbb{C}$. We have

$$||f(z) - f(w)|| = ||\overline{z} - \overline{w}|| = ||\overline{z - w}|| = ||z - w||.$$

Thus, f is a rigid motion. Furthermore, $f(0) = 0$. However, f is not linear, since $f(i) = -i \neq if(1) = i$. $\square$

Example 9.5. Suppose A is a square matrix for which $AA^* = \lambda I$ for some $\lambda \in \mathbb{F}$. Prove that A is normal.

Solution. First, note that if $\lambda \neq 0$, then $(\det A)(\det A^*) = \lambda^n$ is not zero and thus A is invertible. This implies

$$A^* = A^{-1}\lambda I = \lambda I A^{-1} \Rightarrow A^*A = \lambda I \Rightarrow AA^* = A^*A$$

Therefore, A is normal.

Now, assume $\lambda = 0$. Let $\mathbf{a}$ be the first row of A. The first column of A^* is $\mathbf{a}^*$. The $(1,1)$ entry of AA^* is $\mathbf{a}\mathbf{a}^*$ which is the same as $\mathbf{a} \cdot \mathbf{a} = ||\mathbf{a}||^2$. Since $\lambda = 0$, we must have $AA^* = 0$ and thus $||\mathbf{a}||^2 = 0$, which implies the first row of A is zero. A similar argument shows all rows of A are zero. Thus, $A = 0$, which implies $AA^* = 0 = A^*A$. Hence, A is normal. $\square$

Example 9.6. Suppose $A, B \in M_n(\mathbb{F})$ are self-adjoint. Prove AB is self-adjoint if and only if $AB = BA$.

Solution. AB is self-adjoint if and only if $(AB)^* = AB$. By properties of adjoint, $(AB)^* = B^*A^* = BA$, since A, B are self-adjoint. Therefore, $(AB)^* = AB$ if and only if $BA = AB$, as desired. $\square$

9.6 Exercises

Exercise 9.1. *Suppose V is a vector space. Prove that if $T \in \mathcal{L}(V, V)$, then, there is an ordered basis $\mathcal{A}$ of V for which $[T]_{\mathcal{A}\mathcal{A}}$ is upper triangular. (Do not assume V is an inner product space.)*

Exercise 9.2. *Suppose T is an upper triangular normal matrix. Prove T is diagonal.*

Exercise 9.3. *Use the fact that every matrix $A \in M_n(\mathbb{C})$ is unitarily equivalent to an upper triangular matrix to prove,*

(a) $\det A$ *is the product of eigenvalues of* A*.*

(b) $\operatorname{tr} A$ *is the sum of eigenvalues of* A*.*

Exercise 9.4. *Consider the matrix*

$$A = \begin{pmatrix} 2 & 1+i \\ 1-i & 1 \end{pmatrix}.$$

(a) Without evaluating the eigenvalues of A*. Prove* A *is unitarily diagonalizable.*

(b) Unitarily diagonalize A*.*

Exercise 9.5. *Prove that a normal matrix* A *is unitary if and only if the absolute value of each of its eigenvalues is 1.*

Exercise 9.6. *Prove that a normal matrix* A *is self-adjoint if and only if all of its eigenvalues are real.*

Exercise 9.7. *Determine if each statement is true or false:*

(a) Any symmetrix matrix in $M_n(\mathbb{C})$ *is diagonalizable.*

(b) Any symmetric matrix in $M_n(\mathbb{C})$ *is unitarily diagonalizable.*

(c) Any diagonalizable matrix in $M_n(\mathbb{C})$ *is unitarily diagonalizable.*

(d) The product of every two normal matrices is normal.

Exercise 9.8. *Suppose an invertible square matrix* A *satisfies* $A^* = A^2$*. Prove* $A^3 = I$*.*

Exercise 9.9. *Determine if each of the following is true or false.*

(a) If A *is a self-adjoint matrix, then* $A + iI$ *is invertible.*

(b) If A *is a unitary matrix, then* $A + 2I$ *is invertible.*

(c) If A *is a normal matrix, then* $A + 3I$ *is invertible.*

(d) If $A \in M_n(\mathbb{R})$*, then* $A + iI$ *is invertible.*

(e) If $A \in M_n(\mathbb{R})$ *and* $A + (2-i)I$ *is invertible, then* $A + (2+i)I$ *is invertible.*

Week 10

10.1 Positive Definite and Square Roots

Definition 10.1. A self adjoint linear transformation $T : V \to V$ on an inner product spaces V is called **positive definite**, written as $T > 0$, iff

$$\langle T(\mathbf{x}), \mathbf{x} \rangle > 0 \quad \forall \mathbf{x} \in V \backslash \{\mathbf{0}\}.$$

Similarly, we say T is **positive semidefinite**, written $A \geq 0$, iff

$$\langle T(\mathbf{x}), \mathbf{x} \rangle \geq 0 \quad \forall \mathbf{x} \in V.$$

Similarly, a self adjoint matrix $A \in M_n(\mathbb{F})$ is said to be positive definite, written $A > 0$, iff

$$\mathbf{x}^* A \mathbf{x} > 0 \quad \forall \mathbf{x} \in \mathbb{F}^n \backslash \{\mathbf{0}\}.$$

A is called positive semidefinite, written $A \geq 0$, iff

$$\mathbf{x}^* A \mathbf{x} \geq 0 \quad \forall \mathbf{x} \in \mathbb{F}^n.$$

Similarly, we define **negative definite** and **negative semidefinite** linear transformations and matrices. If a self adjoint linear transformation (resp. a self adjoint matrix) is neither positive semidefinite nor negative semidefinite, we say it is **indefinite**. Determining the **definiteness** of a linear transformation or a matrix means determining if it is positive definite, positive semidefinite, negative definite, negative semidefinite or indefinite.

Theorem 10.1. *Let A be a self adjoint matrix. Then,*

(a) $A > 0$ if and only if all eigenvalues of A are positive.

(b) $A \geq 0$ if and only if all eigenvalues of A are nonnegative.

Theorem 10.2. *Let A be a positive semidefinite matrix. Then, there exists a unique positive semidefinite matrix B for which $B^2 = A$.*

Definition 10.2. Given a positive semidefinite matrix A, the unique matrix B satisfying $B^2 = A$ is called the **square root** of A and is denoted by $\sqrt{A}$.

Definition 10.3. The **modulus** of a matrix A, denoted by $|A|$, is defined as $\sqrt{A^*A}$.

Theorem 10.3. *For any matrix A we have $\|\,|A|\mathbf{x}\,\| = \|\,A\mathbf{x}\,\|$ for all $\mathbf{x} \in \mathbb{F}^n$.*

Corollary 10.1. For any matrix A we have $\operatorname{Ker} A = \operatorname{Ker} |A|$.

10.2 Polar, Singular Value and Schmidt Decompositions

Any real number x can be written as $x = \epsilon|x|$, where $\epsilon = \pm 1$. Similarly, every complex number z can be written as $z = (\cos\theta + i\sin\theta)|z|$ (See Appendix). Note that $|\cos\theta + i\sin\theta| = 1$. In other words, $|z|$ tells us how "large" z is and $\cos\theta + i\sin\theta$ tells us the angle that we need to rotate $|z|$ to get to z. We will do something similar for matrices.

Theorem 10.4 (Polar Decomposition)**.** *Every square matrix A can be represented as $A = U|A|$, where U is a unitary matrix.*

We know we cannot unitarily diagonalize all matrices, but can we find unitary matrices W, V and a diagonal matrix D for which $A = WDV^*$? Singular Value Decomposition answers this question.

Definition 10.4. Given a matrix A, every eigenvalue of $|A|$ is called a **singular value** of A. In other words, if $\lambda_1, \dots, \lambda_n$ are eigenvalues of A^*A, then $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$ are singular values of A.

Theorem 10.5. *Suppose $\sigma_1, \dots, \sigma_k$ is the list of all nonzero singular values of A, and $\mathbf{v}_1, \dots, \mathbf{v}_k$ is an orthonormal basis of eigenvectors of A^*A, where $(\sigma_j^2, \mathbf{v}_j)$ is an eigenpair for A^*A. Then, $\mathbf{w}_1, \dots, \mathbf{w}_k$ defined below are orthonormal:*

$$\mathbf{w}_j = \frac{1}{\sigma_j} A\mathbf{v}_j.$$

Theorem 10.6. *Given a matrix $A \in M_{m\times n}(\mathbb{F})$, there are unitary matrices $W \in M_{m\times m}(\mathbb{F})$, $V \in M_{n\times n}(\mathbb{F})$ and a matrix $\Sigma \in M_{m\times n}(\mathbb{F})$ for which the following are satisfied:*

(a) $A = W\Sigma V^*$.

(b) Every (j,k) entry of Σ, where $j \neq k$, is zero.

(c) Every (j,j) entry of Σ is a nonnegative real number.

(d) All zero rows of Σ are at the bottom.

Definition 10.5. Any decomposition $A = W\Sigma V^*$ that satisfies the properties stated in the previous theorem is called the **singular value decomposition** (SVD) of A.

Theorem 10.7. *Every $A \in M_{m\times n}(\mathbb{F})$ can be written as*

$$A = \sum_{j=1}^{k} \sigma_j \mathbf{w}_j \mathbf{v}_j^* \qquad (*),$$

where $\sigma_j > 0$ for $j = 1, \dots, k$ and both $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{F}^m$ and $\mathbf{w}_1, \dots, \mathbf{w}_k \in \mathbb{F}^m$ are orthonormal.

Definition 10.6. Any decomposition of the form $(*)$ satisfying the conditions stated in the previous theorem is called a **Schmidt decomposition** of A.

10.3 Examples

Example 10.1. Suppose A is a normal matrix whose eigenvalues, including multiplicity, are $\lambda_1, \ldots, \lambda_n$. Prove that the singular values of A are $|\lambda_1|, \ldots, |\lambda_n|$

Solution. Since A is normal, by Theorem 9.6, $A = UDU^*$ for some unitary matrix U and a diagonal matrix $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. We have

$$A^*A = (UDU^*)^*(UDU^*) = UD^*U^*UDU^* = UD^*DU^* = U\mathrm{diag}(|\lambda_1|^2, \ldots, |\lambda_n|^2)U^*.$$

The eigenvalues of A^*A are $|\lambda_1|^2, \ldots, |\lambda_n|^2$. Therefore, the singular values of A are $|\lambda_1|, \ldots, |\lambda_n|$, as desired. □

Example 10.2. Let $A \in M_{m\times n}(\mathbb{F})$. Prove that all non-zero eigenvalues of AA^* and A^*A, counting multiplicities, are the same.

Solution. Assume $A = UDV^*$ is a singular value decomposition of A. Note that $U \in M_m(\mathbb{F})$ and $V \in M_n(\mathbb{F})$ are unitary, and D is an $m \times n$ "diagonal" matrix, i.e. $D_{ij} = 0$ for all $i \neq j$. Also, all entries of D are nonnegative. We have

$$AA^* = UDV^*VD^*U^* = UDD^*U^*, \text{ and } A^*A = VD^*U^*UDV^* = VD^*DV^*.$$

Therefore, AA^* is similar to DD^* and A^*A is similar to D^*D. If the (j,j) entry of D is λ_j, then the (j,j) entry of DD^* is λ_j^2. Therefore, the diagonal entries of DD^* are λ_j^2's and zero, if any diagonal entries are left. Similar is true for D^*D.

The number of zero eigenvalues of DD^* and D^*D are the same if and only if $m = n$. □

Example 10.3. Suppose a matrix $A \in M_{m\times n}(\mathbb{F})$ is written as

$$A = \sum_{j=1}^{k} \sigma_j \mathbf{w}_j \mathbf{v}_j^*,$$

where $\sigma_j > 0$ for $j = 1, \ldots, k$ and both $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{F}^m$ and $\mathbf{w}_1, \ldots, \mathbf{w}_k \in \mathbb{F}^m$ are orthonormal. Prove that

(a) $\sigma_1, \ldots, \sigma_k$ are all nonzero singular values of A.

(b) $\mathbf{w}_j = \dfrac{1}{\sigma_j} A\mathbf{v}_j$ for $j = 1, \ldots, k$.

Solution. (a) Suppose $\mathbf{v}_1, \ldots, \mathbf{v}_k, \ldots, \mathbf{v}_n \in \mathbb{F}^n$ is an orthonormal basis for $\mathbb{F}^n$, which exists by extending $\mathbf{v}_1, \ldots, \mathbf{v}_k$ to a basis of $\mathbb{F}^n$ (See Theorem 3.9) and the Gram-Schmidt Orthogonalization Process. Using properties of adjoint we obtain $A^* = \sum_{j=1}^{k} \sigma_j \mathbf{v}_j \mathbf{w}_j^*$. Therefore,

$$A^*A = \sum_{j=1}^{k} \sigma_j \mathbf{v}_j \mathbf{w}_j^* \sum_{j=1}^{k} \sigma_j \mathbf{w}_j \mathbf{v}_j^* = \sum_{j=1}^{k}\sum_{r=1}^{k} \sigma_j \sigma_r \mathbf{v}_j \mathbf{w}_j^* \mathbf{w}_r \mathbf{v}_r^* = \sum_{j=1}^{k}\sum_{r=1}^{k} \sigma_j \sigma_r \mathbf{v}_j (\mathbf{w}_r \cdot \mathbf{w}_j) \mathbf{v}_r^*.$$

Since $\mathbf{w}_1, \ldots, \mathbf{w}_k$ are orthonormal we obtain:

$$\mathbf{w}_r \cdot \mathbf{w}_j = \begin{cases} 1 & \text{if } r = j \\ 0 & \text{if } r \neq j \end{cases}$$

Therefore,

$$A^*A = \sum_{j=1}^{k} \sigma_j^2 \mathbf{v}_j \mathbf{v}_j^*.$$

Since

$$\mathbf{v}_j^* \mathbf{v}_r = \mathbf{v}_r \cdot \mathbf{v}_j = \begin{cases} 1 & \text{if } j = r \\ 0 & \text{if } j \neq r \end{cases}$$

we obtain $A^*A\mathbf{v}_r = \sum_{j=1}^{k} \sigma_j^2 \mathbf{v}_j \mathbf{v}_j^* \mathbf{v}_r = \sigma_r^2 \mathbf{v}_r$ if $r \leq k$, and $A^*A\mathbf{v}_r = \mathbf{0}$ if $r > k$. Thus, $(\sigma_r^2, \mathbf{v}_r)$ is an eigenpair of A^*A for $r = 1, \ldots, k$ and $(0, \mathbf{v}_r)$ is an eigenpair for $r = k+1, \ldots, n$. Therefore, $\sigma_1, \ldots, \sigma_k$ are all nonzero singular values of A.

(b) For every $1 \leq r \leq k$ we have:

$$A\mathbf{v}_r = \sum_{j=1}^{k} \sigma_j \mathbf{w}_j \mathbf{v}_j^* \mathbf{v}_r = \sum_{j=1}^{k} \sigma_j \mathbf{w}_j (\mathbf{v}_r \cdot \mathbf{v}_j) = \sigma_r \mathbf{w}_r \Rightarrow \frac{1}{\sigma_r} A\mathbf{v}_r = \mathbf{w}_r.$$

This completes the proof. □

Example 10.4. Prove the number of nonzero singular values of a matrix is the same as the rank of a matrix.

Solution. Suppose $A = W\Sigma V^*$ is a singular value decomposition of A. By Exercise 4.2(d), we have $\mathrm{Col}\,(W\Sigma V^*) = \mathrm{Col}\,(W\Sigma)$. Therefore, $\mathrm{rank}\, A = \mathrm{rank}\,(W\Sigma)$. Similarly, by Exercise 4.2(d), we have $\mathrm{Row}\,(W\Sigma) = \mathrm{Row}\,(\Sigma)$ and thus $\mathrm{rank}\,(W\Sigma) = \mathrm{rank}\,\Sigma$. Therefore, $\mathrm{rank}\, A = \mathrm{rank}\,\Sigma$. Note that Σ is in echelon form and its rank is the same as the number of nonzero singular values of A, as desired. □

Example 10.5. Suppose $A = W\Sigma V^*$ is a singular value decomposition of a matrix A. Prove that the nonzero entries of Σ are all nonzero singular values of A.

Solution. We have the following:

$$A^*A = V\Sigma^* W^* W \Sigma V^* = V\Sigma^T \Sigma V^*.$$

Therefore, A^*A and $\Sigma^T\Sigma$ are similar. Since Σ is "diagonal" the diagonal entries of $\Sigma^T\Sigma$ are $\sigma_1^2, \dots, \sigma_k^2, 0, \dots, 0$, where $\sigma_1, \dots, \sigma_k$ are nonzero enrties of the diagonal of Σ. Therefore, all nonzero eigenvalues of A^*A are $\sigma_1^2, \dots, \sigma_k^2$. Thus, by definition of singular values, $\sigma_1, \dots, \sigma_k$ are all nonzero singular values of A. $\square$

Example 10.6. Prove the following matrix is positive definite and find its square root.

$$A = \begin{pmatrix} 2 & \sqrt{3} \\ \sqrt{3} & 4 \end{pmatrix}.$$

Solution. The characteristic polynomial of A is $z^2 - 6z + 5 = (z-1)(z-5)$. So, $\sigma(A) = \{1, 5\}$. Since both eigenvalues of A are positive, A is positive definite. An eigenvector corresponding to the eigenvalue of 1 is $(\sqrt{3}\ \ -1)^T$ and an eigenvector corresponding to the eigenvalue of 5 is $(1\ \ \sqrt{3})^T$. Therefore,

$$A = \begin{pmatrix} \sqrt{3} & 1 \\ -1 & \sqrt{3} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} \sqrt{3} & 1 \\ -1 & \sqrt{3} \end{pmatrix}^{-1}.$$

Therefore, from the proof of Theorem 10.2,

$$\sqrt{A} = \begin{pmatrix} \sqrt{3} & 1 \\ -1 & \sqrt{3} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{5} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 1 \\ -1 & \sqrt{3} \end{pmatrix}^{-1} = \frac{1}{4} \begin{pmatrix} 3+\sqrt{5} & -\sqrt{3}+\sqrt{15} \\ -\sqrt{3}+\sqrt{15} & 1+\sqrt{15} \end{pmatrix}.$$

$\square$

Example 10.7. Prove a square matrix A is positive semidefinite if and only if $A = |A|$.

Solution. ($\Rightarrow$) Since A is positive semidefinite, it is self adjoint. Thus, $A^*A = A^2$. Therefore, A is the positive semidefinite matrix whose square is A^*A and thus $A = \sqrt{A^*A}$. Therefore, $A = |A|$.

($\Leftarrow$) Suppose $A = |A|$. Since $|A|$ is positive semidefinite, so is A, as desired. $\square$

10.4 Exercises

Exercise 10.1. *We know for every matrix A, the matrix A^*A is positive semidefinite. Prove the converse: If $A \geq 0$, then there is a matrix B for which $A = B^*B$.*

Exercise 10.2. *Find a Schmidt decomposition and a singular value decomposition of* $\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 1 \end{pmatrix}$.

Exercise 10.3. *Show that for every square matrix A we have* $\det|A| = |\det A|$.

Exercise 10.4. *Let σ be the largest singular value of a matrix A. Prove that if $\lambda \in \sigma(A)$, then $|\lambda| \leq \sigma$.*

Exercise 10.5. *Consider the matrix* $A = \begin{pmatrix} -2 & -2 \\ 1 & -2 \end{pmatrix}$.

(a) Find a singular value decomposition for A.

(b) Use part (a) to find the maximum and minimum of $||A\mathbf{x}||$, where $\mathbf{x}$ ranges over all unit vectors in $\mathbb{R}^2$.

Exercise 10.6. *Prove a positive semidefinite matrix A is positive definite if and only if $\det A \neq 0$.*

Exercise 10.7. *Suppose A is a normal matrix whose list of eigenvalues is $\lambda_1, \ldots, \lambda_n$. Prove that the list of singular values of A is $|\lambda_1|, \ldots, |\lambda_n|$.*

Exercise 10.8. *Suppose $A = W\Sigma V^*$ is a singular value decomposition for a square matrix A. Prove $|A| = V\Sigma V^*$ and if we set $U = WV^*$, the decomposition $A = U|A|$ is a polar decomposition of A.*

Exercise 10.9. *Let $A = \begin{pmatrix} 7 & 1 \\ 0 & 0 \\ 5 & 5 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}$. Find*

(a) $|A|$ and $|B|$.

(b) Two singular value decompositions of A.

(c) A polar decomposition for B.

10.5 Summary

- To find a singular value and a Schmidt decomposition of a matrix $A \in M_{m\times n}(\mathbb{F})$:
  - Evaluate A^*A.
  - Find the list $\lambda_1, \ldots, \lambda_n$ of all eigenvalues of A^*A.
  - Find an orthonormal basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathbb{F}^n$, where $(\lambda_j, \mathbf{v}_j)$ is an eigenpair of A^*A.
  - Find singular values of A by evaluating square roots of λ_j's. Assume the list of singular values of A is $\sigma_1, \ldots, \sigma_k, \underbrace{0, \ldots, 0}_{n-k \text{ times}}$. Suppose $\sigma_1, \ldots, \sigma_k$ are positive.
  - Evaluate $\mathbf{w}_j = \frac{1}{\sigma_j} A\mathbf{v}_j$ for $j = 1, \ldots, k$.
  - If $k < m$, extend $\mathbf{w}_j$'s to an orthonormal basis $\mathbf{w}_1, \ldots, \mathbf{w}_m$ of $\mathbb{F}^m$. You may need to use the Gram-Schmidt Orthogonalization Process.
  - Set $V = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_n)$ and $W = (\mathbf{w}_1 \ \ldots \ \mathbf{w}_m)$.
  - Create an $m \times n$ matrix Σ whose (j, j) entry is σ_j for $j = 1, \ldots, k$ and all of whose other entries are zero.
  - $A = W\Sigma V^*$ is a singular value decomposition of A.
  - $A = \sum_{j=1}^{k} \sigma_j \mathbf{w}_j \mathbf{v}_j^*$ is a Schmidt decomposition of A.

Week 11

11.1 Structure of Orthogonal Matrices

Theorem 11.1. *Suppose A is an orthogonal matrix. Then, there is an orthogonal matrix U for which*

$$A = U \begin{pmatrix} R_{\varphi_1} & & & & & & & \\ & R_{\varphi_2} & & & & \mathit{0} & & \\ & & \ddots & & & & & \\ & & & R_{\varphi_k} & & & & \\ & & & & \epsilon_1 & & & \\ & \mathit{0} & & & & \ddots & & \\ & & & & & & \epsilon_m \end{pmatrix} U^T,$$

where $R_{\varphi_j} = \begin{pmatrix} \cos\varphi_j & -\sin\varphi_j \\ \sin\varphi_j & \cos\varphi_j \end{pmatrix}$ *is a 2×2 rotation matrix for $j = 1, \ldots, k$ and $\epsilon_1, \ldots, \epsilon_m \in \{\pm 1\}$.*

11.2 Bilinear Forms

Definition 11.1. A **bilinear form** on a vector space V is a function $L : V \times V \to \mathbb{F}$ that is linear with respect to both of its components. In other words for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and all $a, b \in \mathbb{F}$,

$$L(a\mathbf{x} + b\mathbf{y}, \mathbf{z}) = aL(\mathbf{x}, \mathbf{z}) + bL(\mathbf{y}, \mathbf{z}), \text{ and } L(\mathbf{z}, a\mathbf{x} + b\mathbf{y}) = aL(\mathbf{z}, \mathbf{x}) + bf(\mathbf{z}, \mathbf{y}).$$

Theorem 11.2. *A function $L : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ is bilinear if and only if $L(\mathbf{x}, \mathbf{y}) = \mathbf{y}^T A\mathbf{x}$ for a fixed matrix $A \in M_n(\mathbb{F})$ and all column vectors $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$. Furthermore, for a given bilinear form L, this matrix A is unique.*

11.3 Quadratic Forms

Suppose $\langle \mathbf{x}, \mathbf{y} \rangle$ is an inner product on $\mathbb{F}^n$. We can write $\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{e}_j$ and $\mathbf{y} = \sum_{j=1}^{n} y_j \mathbf{e}_j$, which yields:

$$\begin{aligned}
\langle \mathbf{x}, \mathbf{y} \rangle &= \langle \sum_{j=1}^{n} x_j \mathbf{e}_j, \sum_{j=1}^{n} y_j \mathbf{e}_j \rangle \\
&= \sum_{j=1}^{n} x_j \sum_{k=1}^{n} \overline{y}_k \langle \mathbf{e}_j, \mathbf{e}_k \rangle \text{ by linearity and conjugate symmetry} \\
&= \begin{pmatrix} \sum_{k=1}^{n} \overline{y}_k \langle \mathbf{e}_1, \mathbf{e}_k \rangle & \sum_{k=2}^{n} \overline{y}_k \langle \mathbf{e}_2, \mathbf{e}_k \rangle & \cdots & \sum_{k=1}^{n} \overline{y}_k \langle \mathbf{e}_n, \mathbf{e}_k \rangle \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}. \\
&= (\overline{y}_1 \ \cdots \ \overline{y}_n) \begin{pmatrix} \langle \mathbf{e}_1, \mathbf{e}_1 \rangle & \cdots & \langle \mathbf{e}_n, \mathbf{e}_1 \rangle \\ \vdots & \vdots & \vdots \\ \langle \mathbf{e}_1, \mathbf{e}_n \rangle & \cdots & \langle \mathbf{e}_n, \mathbf{e}_n \rangle \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\
&= \mathbf{y}^* A \mathbf{x}
\end{aligned}$$

Note that the matrix A above is independent of $\mathbf{x}$ and $\mathbf{y}$ and only depends on the inner product. Furthermore, since $\langle , \rangle$ satisfies conjugate symmetry, A is self adjoint. We will now investigate for what matrices A, the expression $\mathbf{y}^* A \mathbf{x}$ defines an inner product on $\mathbb{F}^n$. One of the conditions we need to prove is Positivity. For that reason, we state the following definition.

Definition 11.2. A **quadratic form** $Q : \mathbb{F}^n \to \mathbb{F}$ is a function defined by $Q(\mathbf{x}) = \mathbf{x}^* A \mathbf{x}$, where $A \in M_n(\mathbb{F})$ is a fixed matrix. When $\mathbb{F} = \mathbb{R}$ we say Q is a **quadratic form on** $\mathbb{R}^n$ or a **real quadratic form**. When $\mathbb{F} = \mathbb{C}$ we say Q is a **quadratic form on** $\mathbb{C}^n$ or a **complex quadratic form**.

Example 11.1. Write down the real quadratic form $Q(x_1, x_2) = x_1^2 - 2x_2^2 + 3x_1 x_2$ in the form $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$. How many different such matrices A can you find?

Theorem 11.3. *Let Q be a quadratic form on $\mathbb{R}^n$. Then, there is a unique symmetric matrix A for which $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$.*

Definition 11.3. For a real quadratic form Q, the unique symmetric matrix A for which $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$ is called the **matrix associated with** Q. The quadratic form Q is called the **quadratic form defined by** A.

We cannot guarantee that the matrix of a quadratic form is symmetric or self adjoint. In fact distinct matrices yield distinct complex quadratic forms.

Theorem 11.4. *Suppose Q is a complex quadratic form. The matrix $A \in M_n(\mathbb{C})$ for which $Q(\mathbf{x}) = \mathbf{x}^* A \mathbf{x}$ is unique.*

Theorem 11.5. *Let $A \in M_n(\mathbb{F})$. Then, $\mathbf{x}^* A \mathbf{y} = \overline{\mathbf{y}^* A \mathbf{x}}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$ if and only if A is self adjoint.*

Theorem 11.6. *Suppose $A \in M_n(\mathbb{C})$ is a fixed matrix. $\mathbf{x}^* A\mathbf{x}$ is real for all $\mathbf{x} \in \mathbb{C}^n$, if and only if A is self adjoint.*

Theorem 11.7. *Given a matrix $A \in M_n(\mathbb{F})$, the function $\langle , \rangle : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ defined by $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{y}^* A\mathbf{x}$ is an inner product if and only if A is positive definite. (Recall that any positive definite matrix is, by definition, self adjoint.)*

11.4 Diagonalization of Quadratic Forms

Consider a quadratic form $Q(x) = \mathbf{x}^* A\mathbf{x}$, with $A = A^*$. Since A is self adjoint, all eigenvalues of A are real, and by Corollary 9.1, A is unitarily diagonalizable. Therefore, we can write $A = UDU^*$ for some unitary matrix U and a diagonal matrix $D \in M_n(\mathbb{R})$. Setting $\mathbf{y} = U^*\mathbf{x}$ we can write

$$Q(\mathbf{x}) = \mathbf{x}^* A\mathbf{x} = \mathbf{x}^* UDU^*\mathbf{x} = (U^*\mathbf{x})^* D(U^*\mathbf{x}) = \mathbf{y}^* D\mathbf{y} = \sum_{j=1}^{n} \lambda_j |y_j|^2,$$

where $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $\mathbf{y} = (y_1 \ \cdots \ y_n)^T$.

Example 11.2. Unitarily diagonalize the matrix of the quadratic form $Q(x, y) = x^2 + y^2 + 4xy$. Use that to write this quadratic form as a sum or difference of squares of linear combinations of x and y.

Unitary diagonalization of a matrix requires solving polynomial equations, which is not always easy or even possible. For that reason we would seek non-unitary diagonalizations.

Theorem 11.8. *Suppose $Q(\mathbf{x}) = \mathbf{x}^* A\mathbf{x}$ is a quadratic form on $\mathbb{F}^n$ with $A^* = A$. Assume $D = S^* AS$, where $D = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$ and $S \in M_n(\mathbb{F})$ is invertible. (Note that S is not assumed to be unitary.) Let $\mathbf{y} = S^{-1}\mathbf{x} = (y_1 \ \cdots \ y_n)^T$. Then, $Q(\mathbf{x}) = \sum_{j=1}^{n} \alpha_j |y_j|^2$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$.*

Definition 11.4. Suppose $Q(\mathbf{x}) = \mathbf{x}^* A\mathbf{x}$ is a quadratic form with $A^* = A$. A **diagonalization** for Q is a way of writing Q as

$$Q(\mathbf{x}) = \sum_{j=1}^{n} \alpha_j |y_j|^2,$$

where $\alpha_j \in \mathbb{R}$ is constant for $j = 1, \ldots, n$ and

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = S^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

for some invertible matrix $S \in M_n(\mathbb{F})$. We say this diagonalization is **unitary** (or **orthogonal** when $\mathbb{F} = \mathbb{R}$) iff the $n \times n$ matrix S is a unitary matrix.

Remark 11.1. In order to diagonalize Q we will find an invertible matrix S and a diagonal matrix D for which $D = S^* AS$. We then apply Theorem 11.8.

Example 11.3. Diagonalize $Q(x,y,z) = x^2 + 2y^2 + 4xy - 2xz + 2yz$.

Example 11.4. Write down the quadratic form associated with the following matrix and diagonalize it:

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

Determine the definiteness of this quadratic form.

11.5 Examples

Example 11.5. Consider the (real) quadratic form

$$Q(x_1,x_2,x_3) = -x_1^2 - x_2^2 + 8x_3^2 - 4x_1x_2 + 4x_1x_3 + 2x_2x_3.$$

(a) Find the matrix of this quadratic form.

(b) Diagonalize Q.

Solution. (a) The matrix of Q is

$$A = \begin{pmatrix} -1 & -2 & 2 \\ -2 & -1 & 1 \\ 2 & 1 & 8 \end{pmatrix}$$

(b) Applying $R_2 - 2R_1$ and $R_3 + 2R_1$ to The agumented matrix $(A|I)$ we obtain the following:

$$\left(\begin{array}{ccc|ccc} -1 & -2 & 2 & 1 & 0 & 0 \\ 0 & 3 & -3 & -2 & 1 & 0 \\ 0 & -3 & 12 & 2 & 0 & 1 \end{array}\right)$$

Applying the corresponding column operations in the same order (i.e. $C_2 - 2C_1$ followed by $C_3 + 2C_1$) we obtain the following:

$$\left(\begin{array}{ccc|ccc} -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 3 & -3 & -2 & 1 & 0 \\ 0 & -3 & 12 & 2 & 0 & 1 \end{array}\right)$$

Applying $R_3 + R_2$ to the augmented matrix we obtain:

$$\left(\begin{array}{ccc|ccc} -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 3 & -3 & -2 & 1 & 0 \\ 0 & 0 & 9 & 0 & 1 & 1 \end{array}\right)$$

Applying $C_3 + C_2$ to the above matrix we obtain:

$$\left(\begin{array}{ccc|ccc} -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 3 & 0 & -2 & 1 & 0 \\ 0 & 0 & 9 & 0 & 1 & 1 \end{array}\right)$$

Therefore, $S^TAS = D$, where,

$$S^T = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \text{ and } D = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 9 \end{pmatrix}$$

We now find S^{-1} by row reducing the augmented matrix $(S|I)$:

$$S^{-1} = \begin{pmatrix} 1 & 2 & -2 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}$$

Therefore, $S^{-1}\mathbf{x} = (x_1 + 2x_2 - 2x_3 \;\; x_2 - x_3 \;\; x_3)^T$, which yields the following diagonalization of Q:

$$Q(x_1, x_2, x_3) = -(x_1 + 2x_2 - 2x_3)^2 + 3(x_2 - x_3)^2 + 9x_3^2$$

□

Example 11.6. Prove that a square matrix $A \in M_n(\mathbb{F})$ is symmetric if and only if there is an invertible matrix S for which S^TAS is diagonal.

Solution. $\Leftarrow$: Since S^TAS is diagonal, it is symmetric. Therefore, $(S^TAS)^T = S^TAS$, which implies $S^TA^TS = S^TAS$. Since S is invertible, S^T is invertible. Multiplying both sides of $S^TA^TS = S^TAS$ by S^{-1} from the right and $(S^T)^{-1}$ from the left, we obtain $A^T = A$, as desired.

$\Rightarrow$: Suppose $A \in M_n(\mathbb{F})$ is symmetric. We will prove the statement by induction on n.

Basis step. For $n = 1$, the matrix A is itself diagonal. Thus $S = (1)$ works.

Inductive step. First assume the $(1, 1)$ entry of A is non-zero. By applying row addition operations, if needed, i.e. $R_2 + \alpha_2 R_1, R_3 + \alpha_3 R_1, \ldots, R_n + \alpha_n R_1$ we can turn A into a matrix whose first column is a multiple of $\mathbf{e}_1$. We know applying row operations is the same as multiplying by elementary matrices from the left. Thus, there is an invertible matrix P for which the first column of PA is a multiple of $\mathbf{e}_1$. Since P^T corresponds to column additions $C_2 + \alpha_2 C_1, C_3 + \alpha_3 C_1, \ldots, C_n + \alpha_n C_1$, the first column of PAP^T is the same as the first column of PA and thus, it is a multiple of $\mathbf{e}_1$. Since A is symmetric, so is PAP^T. Therefore, there is an $(n-1) \times (n-1)$ matrix B for which

$$PAP^T = \left(\begin{array}{c|c} c & \mathbf{0} \\ \hline \mathbf{0} & B \end{array}\right).$$

Since PAP^T is symmetric, so is B. By inductive hypothesis, there is an $(n-1) \times (n-1)$ invertible matrix K for which K^TBK is diagonal. Therefore the following matrix is diagonal.

$$\left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & K^T \end{array}\right) PAP^T \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & K \end{array}\right) = \left(\begin{array}{c|c} c & \mathbf{0} \\ \hline \mathbf{0} & K^TBK \end{array}\right)$$

Thus, the matrix S^TAS is diagonal, where

$$S = P^T \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & K \end{array} \right).$$

Now, assume the $(1,1)$ entry of A is zero. If the first column of A is the zero vector, then the first row of A must also be the zero vector, so the above proof with $c = 0$ and $P = I$ would give us the result. If $A_{11} = 0$ but some other entry of the first column of A, say A_{21}, is non-zero, then we apply the row operation $R_1 + R_2$ to turn A_{11} into A_{21}. We then apply $C_1 + C_2$, we change the the first entry of the first column into $A_{21} + A_{12} = 2A_{21}$ which is non-zero. Therefore, there is an invertible (in fact elementary) matrix Q for which the $(1,1)$ entry of QAQ^T is non-zero. By what we showed above, there is an invertible matrix S for which S^TQAQ^TS is diagonal. Therefore, $(Q^TS)^TA(Q^TS)$ is diagonal. □

Example 11.7. Suppose $A = SDS^*$ for an invertible matrix S and a diagonal matrix D. Is it true that diagonal entries of D must be eigenvalues of A?

Solution. No! Choose D to be the identity matrix. Then $A = SIS^*$ does not have to have any eigenvalues of 1. Here is an example:

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

Characteristic polynomial of A is $(2-t)(1-t) - 1 = t^2 - 3t + 1$. Therefore, the eigenvalues of A are $(3 \pm \sqrt{5})/2$. □

Example 11.8. Is it possible to write a square matrix A as $A = SDS^*$ for infinitely many diagonal matrices D and invertible matrices S?

Solution. Yes. In the example above, for every real number r we can write

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1/r & 1/r \\ 0 & 1/r \end{pmatrix} \begin{pmatrix} r^2 & 0 \\ 0 & r^2 \end{pmatrix} \begin{pmatrix} 1/r & 0 \\ 1/r & 1/r \end{pmatrix}.$$

□

Example 11.9. Suppose $D = S^*AS$ for an invertible matrix S, a square matrix A and a diagonal matrix D. Prove if all entries of D are real, then A is self adjoint. Does the conclusion remain true without the condition that S is invertible?

Solution. Since S is invertible, so is S^* and $(S^*)^{-1} = (S^{-1})^*$. Therefore, we have $A = (S^*)^{-1}DS^{-1} = (S^{-1})^*DS^{-1}$. We have:

$$A^* = ((S^{-1})^*DS^{-1})^* = (S^{-1})^*D^*S^{-1} = (S^{-1})^*DS^{-1} = A.$$

Therefore, A is self adjoint.

This is not true if we drop the assumption that S is invertible. For example we can take $D = S = 0$ and $A = \mathrm{diag}(i, 1)$. $\square$

Example 11.10. Diagonalize the quadratic form and determine its definiteness:

$$Q(x_1, x_2) = |x_1|^2 - i\overline{x}_1 x_2 + 3|x_2|^2 + ix_1\overline{x}_2.$$

Solution. The matrix associated with this quadratic form is $A = \begin{pmatrix} 1 & -i \\ i & 3 \end{pmatrix}$, which is self adjoint. We will proceed by diagonalizing this matrix using row and column operations.

$$\left(\begin{array}{cc|cc} 1 & -i & 1 & 0 \\ i & 3 & 0 & 1 \end{array}\right) \xrightarrow{R_2 - iR_1} \left(\begin{array}{cc|cc} 1 & -i & 1 & 0 \\ 0 & 2 & -i & 1 \end{array}\right) \xrightarrow{C_2 + iC_1} \left(\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 2 & -i & 1 \end{array}\right).$$

Therefore, $D = S^*AS$, where

$$D = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \text{ and } S^* = \begin{pmatrix} 1 & 0 \\ -i & 1 \end{pmatrix}.$$

This gives us $S = \begin{pmatrix} 1 & -i \\ 0 & 1 \end{pmatrix}$. Evaluating S^{-1} we obtain

$$S^{-1} = \begin{pmatrix} 1 & i \\ 0 & 1 \end{pmatrix}.$$

Therefore, $S^{-1}(x_1 \;\; x_2)^T = (x_1 + ix_2 \;\; x_2)^T$ which yields the following diagonalization of Q:

$$Q(x_1, x_2) = |x_1 + ix_2|^2 + 2|x_2|^2.$$

Therefore, $Q(x_1, x_2) \geq 0$ for all $x_1, x_2 \in \mathbb{C}$. Furthermore, $Q(x_1, x_2) = 0$ implies $x_1 + ix_2 = 0$ and $x_2 = 0$, which implies $x_1 = x_2 = 0$. Thus, Q is positive definite. $\square$

11.6 Exercises

Exercise 11.1. *Prove Theorem 11.4.*

Exercise 11.2. *Prove Theorem 11.6.*

Exercise 11.3. *Find the matrix associated with the bilinear form*

$$L(\mathbf{x}, \mathbf{y}) = 2x_1y_1 - x_1y_2 + x_2y_1 - x_2y_2 + x_1y_3 + x_3y_3.$$

Here, $\mathbf{x} = (x_1 \; x_2 \; x_3)^T$ *and* $\mathbf{y} = (y_1 \; y_2 \; y_3)^T$.

Exercise 11.4. *Suppose $A \in M_n(\mathbb{C})$ is a matrix for which $\mathbf{x}^* A\mathbf{x}$ is real for all $\mathbf{x} \in \mathbb{C}^n$. Prove that A is self adjoint.*

Exercise 11.5. *Consider the following (real) quadratic form:*

$$Q(x_1, x_2, x_3) = 2x_1^2 + x_2^2 + 2x_3^2 + 4x_1x_2 + 2x_1x_3 + 4x_2x_3$$

(a) Write down the matrix associated with Q.

(b) Write down a unitary diagonalization of Q.

(c) Write down a diagonalization of Q using row and column operations.

Exercise 11.6. *Consider the following matrix:*

$$A = \begin{pmatrix} 1 & 1-i \\ 1+i & 0 \end{pmatrix}.$$

Let $Q(\mathbf{x}) = \mathbf{x}^* A\mathbf{x}$ be the (complex) qudaratic form associated with A.

(a) Write down an expression for $Q(x_1, x_2)$.

(b) Write down a diagonalization of Q using row and column operations.

(c) Use part (b) to determine the type of this quadratic form. (i.e. positive definite, positive semidefinite, etc.)

Exercise 11.7. *Define $L : \mathbb{F}^2 \times \mathbb{F}^2 \to \mathbb{F}$ by $L(\mathbf{x}, \mathbf{y}) = \det(\mathbf{x} \;\; \mathbf{y})$, where $\mathbf{x}, \mathbf{y} \in \mathbb{F}^2$ are column vectors.*

(a) Prove L is bilinear.

(b) Find the matrix associated with L.

Exercise 11.8. *Determine every inner product space V for which the function $L : V \times V \to \mathbb{F}$ defined by $L(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle$ is bilinear.*

Exercise 11.9. *Suppose $A \in M_n(\mathbb{C})$. Prove $\mathbf{x}^T A\mathbf{x} \in \mathbb{R}$ for every column vector $\mathbf{x} \in \mathbb{R}^n$ if and only if $A + A^T \in M_n(\mathbb{R})$.*

Exercise 11.10. *Suppose $\langle , \rangle$ is an inner product of $\mathbb{F}^n$. Prove, there is a positive definite matrix A for which $\langle \mathbf{x}, \mathbf{y} \rangle = (A\mathbf{x}) \cdot (A\mathbf{y})$. (This shows every inner product of $\mathbb{F}^n$ is "essentially" the same as the dot product.)*

Exercise 11.11. *Find an orthogonal diagonalization for the quadratic form*

$$Q(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - 2x_1x_2 - 2x_1x_3.$$

Find another one using row and column operations.

Week 12

12.1 Sylvester's Law of Inertia and Positivity Criterion

Definition 12.1. Let $A \in M_n(\mathbb{F})$ be self adjoint. We say a subspace E of $\mathbb{F}^n$ is A-positive (resp. A-negative) iff $\mathbf{x}^* A\mathbf{x} > 0$ (resp. $\mathbf{x}^* A\mathbf{x} < 0$) for all nonzero $\mathbf{x} \in E$.

Theorem 12.1 (Sylvester's Law of Intertia)**.** *Suppose A is a self adjoint matrix, S is an invertible matrix for which $D = S^* AS$ is diagonal. Then, the number of positive entries (resp. negative entries) of D is the same as the maximum dimension of an A-positive (resp. A-negative) subspace of $\mathbb{F}^n$. Consequently, the number of positive entries (resp. negative entries) of D only depends on A, and not the choice of S.*

Corollary 12.1. Suppose A is a self adjoint matrix, S is an invertible matrix for which $D = S^* AS$ is diagonal.

(a) $A > 0$ if and only if all diagonal entries of D are positive.

(b) $A < 0$ if and only if all diagonal entries of D are negative.

(c) $A \geq 0$ if and only if all diagonal entries of D are nonnegative.

(d) $A \leq 0$ if and only if all diagonal entries of D are nonpositive.

(e) A is indefinite if and only if D has at least one positive and one negative entry.

Theorem 12.2 (Minmax Characterization of Eigenvalues)**.** *Let $A \in M_n(\mathbb{F})$ be self adjoint and $\lambda_1 \geq \cdots \geq \lambda_n$ be the list of all of its eigenvalues. Then, for every $k, 1 \leq k \leq n$,*

$$\lambda_k = \max_{E, \dim E = k} \left(\min_{||\mathbf{x}||=1, \mathbf{x} \in E} (\mathbf{x}^* A\mathbf{x}) \right) = \min_{F, \dim F = n+1-k} \left(\max_{||\mathbf{x}||=1, \mathbf{x} \in F} (\mathbf{x}^* A\mathbf{x}) \right).$$

Theorem 12.3 (Intertwining of Eigenvalues)**.** *Suppose $A \in M_n(\mathbb{F})$ is self adjoint, and let B be the submatrix of A formed by the first $n-1$ rows and the first $n-1$ columns of A. Let $\lambda_1 \geq \cdots \geq \lambda_n$ and $\mu_1 \geq \cdots \geq \mu_{n-1}$ be all eigenvalues of A and B, respectively. Then,*

$$\lambda_1 \geq \mu_1 \geq \lambda_1 \geq \mu_2 \geq \cdots \geq \lambda_{n-1} \geq \mu_{n-1} \geq \lambda_n.$$

Definition 12.2. The k-th **principal minor** of an $n \times n$ matrix A is the determinant of the square submatrix of A formed by the first k columns and the first k rows of A for $k = 1, \dots, n$. We denote the k-th principal minor of A by $\Delta_k(A)$.

Theorem 12.4 (Sylvester's Criterion for Positivity)**.** *Let A be a self adjoint $n \times n$ matrix.*

(a) A is positive definite if and only if $\Delta_k(A) > 0$ for $k = 1, \dots, n$.

(b) A is negative definite if and only if $(-1)^k \Delta_k(A) > 0$ for $k = 1, \dots, n$.

(c) Suppose A is invertible. A is indefinite if and only if neither (a) nor (b) occurs.

12.2 Advanced Spectral Theory

Definition 12.3. Given a square matrix A and a polynomial $p(t) = c_0 + c_1 t + \cdots + c_n t^n$ with $c_j \in \mathbb{F}$ for $j = 1, \dots, n$, $p(A)$ is defined as

$$p(A) = c_0 I + c_1 A + \cdots + c_n A^n.$$

Remark 12.1. Suppose $A \in M_n(\mathbb{F})$, and $p(t), q(t), f(t) \in \mathbb{P}$ for which $p(t) + q(t) = f(t)$. Then, $p(A) + q(A) = f(A)$. Similarly, if $p(t)q(t) = f(t)$, then $p(A)q(A) = f(A)$. Furthermore, by block multiplication of matrices, if A has the following block form,

$$A = \begin{pmatrix} B & * \\ 0 & C \end{pmatrix}.$$

Then,

$$p(A) = \begin{pmatrix} p(B) & * \\ 0 & p(C) \end{pmatrix}.$$

Furthermore, in any polynomial identity that is valid over $\mathbb{R}$ we can substitute the variables by square matrices of the same size as long as the matrices commute. For example, $(A - B)(A + B) = A^2 - B^2$ for every two matrices $A, B \in M_n(\mathbb{F})$ for which $AB = BA$.

Theorem 12.5. *Suppose $\lambda_1, \dots, \lambda_n$ is the list of all eigenvalues of a matrix $A \in M_n(\mathbb{C})$ and $p(t)$ is a polynomial with complex coefficients. Then, the list of all eigenvalues of $p(A)$ is*

$$p(\lambda_1), p(\lambda_2), \dots, p(\lambda_n).$$

Example 12.1. Suppose $(\lambda, \mathbf{v})$ is an eigenpair of a square matrix A and $p(t)$ is a polynomial with coefficients in $\mathbb{F}$. Prove that $p(A)\mathbf{v} = p(\lambda)\mathbf{v}$.

Theorem 12.6 (Cayley-Hamilton)**.** *Let $A \in M_n(\mathbb{F})$ and $p(t) = \det(A - tI)$ be the characteristic polynomial of A. Then, $p(A) = 0$.*

12.3 Jordan Canonical Form

We have shown that every matrix is similar to an upper triangular matrix, but can we write these upper triangular matrices in a more specific form? Let's look at an example.

Example 12.2. Consider the matrix $A = \begin{pmatrix} 2 & 0 & 1 & 0 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 2 \end{pmatrix}$.

(a) Find all eigenpairs of A and show A is not diagonalizable.

(b) Find a basis for $\operatorname{Ker}(A-2I)^n$ for all $n \in \mathbb{Z}^+$.

(c) Use that to write a matrix similar to A in an "almost diagonal" form.

Theorem 12.7. *Let λ be an eigenvalue of an $n \times n$ matrix A. Then, there is an integer $k \leq n$ for which*

$$\operatorname{Ker}(A-\lambda I) \subsetneqq \operatorname{Ker}(A-\lambda I)^2 \subsetneqq \cdots \subsetneqq \operatorname{Ker}(A-\lambda I)^k = \operatorname{Ker}(A-\lambda I)^{k+1} = \cdots .$$

Definition 12.4. For an eigenvalue λ of an $n \times n$ matrix A, every nonzero vector in $\operatorname{Ker}(A-\lambda I)^n$ is called a **generalized eigenvector** of A corresponding to eigenvalue λ. The vector space $\operatorname{Ker}(A-\lambda I)^n$ is called the **generalized eigenspace** of A corresponding to eigenvalue λ. The vector space $\operatorname{Ker}(A-\lambda I)^n$ is generally denoted by E_λ.

Definition 12.5. Suppose $T : V \to V$ is a linear transformation. A subspace E of V is said to be T-invariant iff $T(E) \subseteq E$. Similarly, for a matrix $A \in M_n(\mathbb{F})$, a subspace E of $\mathbb{F}^n$ is called A-invariant iff $A\mathbf{x} \in E$ for all $\mathbf{x} \in E$.

Theorem 12.8. *Generalized eigenspaces corresponding to a linear transformation T are T-invariant.*

Theorem 12.9. *Let $\lambda_1, \ldots, \lambda_k$ be all distinct eigenvalues of a matrix $A \in M_n(\mathbb{C})$. Assume $E_{\lambda_1}, \ldots, E_{\lambda_k}$ are all generalized eigenspaces of A. Then, $E_{\lambda_1}, \ldots, E_{\lambda_k}$ are linearly independent.*

Theorem 12.10. *The dimension of the generalized eigenspace corresponding to the eigenvalue λ for a square matrix A is the same as the algebraic multiplicity of λ as an eigenvalue of A.*

The above theorem implies that given an $n \times n$ matrix A we can find a basis for $\mathbb{F}^n$ consisting of generalized eigenvectors of A by finding a basis for each generalized eigenspace and putting all of these bases together.

Definition 12.6. A matrix is said to be in **Jordan canonical form** (or Jordan form for short) if it is a block matrix of the form

$$\begin{pmatrix} B_1 & 0 & \cdots & 0 & 0 \\ 0 & B_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & B_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & B_n \end{pmatrix},$$

where each B_j, called a **Jordan block**, is a matrix with an eigenvalue λ_j on its main diagonal, 1's immediately above the main diagonal, and zeros everywhere else. In other words:

$$B_j = \begin{pmatrix} \lambda_j & 1 & 0 & \cdots & 0 \\ 0 & \lambda_j & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_j & 1 \\ 0 & 0 & \cdots & 0 & \lambda_j \end{pmatrix}.$$

12.4 Examples

Example 12.3. Suppose A is a self adjoint matrix and S is an invertible matrix for which $S^*AS = D$ is diagonal. Prove the number of zeros on the diagonal of D is the same as $\dim \operatorname{Ker} A$.

Solution. Since A is self adjoint, there is a unitary matrix U and a diagonal matrix D_0 for which $A = UD_0U^*$, (and hence $U^*AU = D_0$) where the diagonal entries of D_0 are all eigenvalues of A. By Theorem 12.1, the number of positive entries of D is the same as the number of positive entries of D_0 and the number of negative entries of D is the same as the number of negative entries of D_0. Therefore, the number of zeros on the diagonals of D and D_0 are the same integer k. Since A and D are similar, they have the same characteristic polynomial and thus the algebraic multiplicity of 0 as an eigenvalue of A is k. Since A is diagonalizable, by Theorem 7.4, $k = \dim \operatorname{Ker} A$, as desired. □

Example 12.4. Give an example of a self adjoint matrix $A = (a_{jk}) \in M_3(\mathbb{R})$ for which $a_{11} > 0$ and $\det A_k \geq 0$ for all k, but A is not positive semidefinite.

Solution. Consider the diagonal matrix $A = \operatorname{diag}(1, 0, -1)$. Since A has a negative eigenvalue of -1, it is not positive semidefinite. Also, $a_{11} = 1$ is positive and $\det A_2 = \det A_3 = 0$. □

Example 12.5. Prove that if $A = (a_{jk}) \in M_2(\mathbb{C})$ is a self adjoint matrix satisfying $a_{11} > 0$ and $\det A \geq 0$, then A is positive semidefinite.

Solution. Suppose $A = \begin{pmatrix} a & b \\ \bar{b} & c \end{pmatrix}$, where $a, c \in \mathbb{R}$ and $b \in \mathbb{C}$. By assumption, $ac - |b|^2 \geq 0$ and $a > 0$. Since $ac \geq |b|^2 \geq 0$ and $a > 0$, we have $c \geq 0$. Since A is self-adjoint, the eigenvalues of A are real. Let λ_1, λ_2 be eigenvalues of A. By an exercise, $\lambda_1 + \lambda_2 = \operatorname{tr}(A) = a + c > 0$. Also, $\lambda_1\lambda_2 = \det A \geq 0$. Thus, either both λ_1 and λ_2 are nonnegative or they are both nonpositive. Since their sum is positive, they must both be nonnegative. Therefore, A is positive semidefinite. □

Example 12.6. Suppose a 2×2 matrix A satisfies $\operatorname{tr} A = 0$. Prove that $A^2 = cI$ for some scalar c.

Solution. Since A is 2×2 and $\operatorname{tr} A = 0$, the matrix A must be of the form

$$A = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}.$$

The characteristic polynomial of A is $p(z) = (a - z)(-a - z) - bc = z^2 - a^2 - bc$. By the Cayley-Hamilton Theorem, we must have $p(A) = 0$ and thus $A^2 = (a^2 + bc)I$, as desired. □

Example 12.7. Find a matrix in Jordan form that is similar to the following matrix:

$$A = \begin{pmatrix} -6 & 5 & -3 & 9 \\ -1 & 2 & 0 & 1 \\ 4 & -4 & 4 & -4 \\ -5 & 3 & -2 & 8 \end{pmatrix}.$$

Solution. The characteristic polynomial of A is $\det(A - zI) = z^4 - 8z^3 + 23z^2 - 28z + 12$. By inspection, we see that $z = 1$ is a root. After performing long division we obtain the following:

$$z^4 - 8z^3 + 23z^2 - 28z + 12 = (z - 1)(z^3 - 7z^2 + 16z - 12).$$

By inspection, we find $z = 2$ as a root of $z^3 - 7z^2 + 16z - 12 = 0$. Repeating this process we find out that the four eigenvalues of A are $1, 2, 2, 3$. For the eigenvalue $z = 2$, the eigenspace $\operatorname{Ker}(A - 2I)$ is one-dimensional (and is generated by $(1\ 1\ 2\ 1)^T$). Thus, A is not diagonalizable and thus the Jordan block corresponding to eigenvalue 2 must be 2×2. Therefore, the matrix in Jordan form that is similar to A is

$$\begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

□

Example 12.8. Find an explicit formula for A^n, where n is a positive integer, and

$$A = \begin{pmatrix} 1 & 1 & -1 \\ 3 & -3 & 7 \\ 2 & -3 & 6 \end{pmatrix}.$$

Solution. The characteristic polynomial of this matrix is $\det(A - zI) = -z^3 + 4z^2 - 5z + 2$. By inspection we can find a root of this polynomial to be $z = 1$. Dividing by $z - 1$ and factoring we obtain $(z - 1)^2(2 - z)$. For $z = 2$ we find $\mathbf{v}_1 = (1\ 2\ 1)^T$ as an eigenvector. For $z = 1$ we see that $\operatorname{Ker}(A - I)$ is one-dimensional and is generated by $(-1\ 1\ 1)^T$. We also obtain

$$(A - I)^2 = \begin{pmatrix} 0 & 1 & -1 \\ 3 & -4 & 7 \\ 2 & -3 & 5 \end{pmatrix}^2 = \begin{pmatrix} 1 & -1 & 2 \\ 2 & -2 & 4 \\ 1 & -1 & 2 \end{pmatrix}.$$

The vector (x, y, z) is in $\operatorname{Ker}(A-I)^2$ if and only if

$$x - y + 2z = 0, \text{ and } 2x - 2y + 4z = 0.$$

Solving for x we obtain $x = y - 2z$. Thus, elements of $\operatorname{Ker}(A-I)^2$ are of the form

$$\begin{pmatrix} y-2z \\ y \\ z \end{pmatrix} = y \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}.$$

Now, we will choose a vector in $\operatorname{Ker}(A-I)^2$ that does not belong to $\operatorname{Ker}(A-I)$. We set $\mathbf{v}_3 = (1\ 1\ 0)^T$, and

$$\mathbf{v}_2 = (A-I)\mathbf{v}_3 = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}.$$

Since $\mathbf{v}_2 \in \operatorname{Ker}(A-I)$, we have $A\mathbf{v}_2 = \mathbf{v}_2$. We also know $A\mathbf{v}_1 = 2\mathbf{v}_1$, and $A\mathbf{v}_3 = \mathbf{v}_2 + \mathbf{v}_3$. Note also that $\mathbf{v}_2, \mathbf{v}_3$ are generalized eigenvectors correponding to eigenvalue of 1 and they are linearly independent. We also know $\mathbf{v}_1$ is an eigenvector corresponding to the eigenvalue of 2. Thus, $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are linearly independent by Theorem 12.9. Therefore, we obtain the following decomposition:

$$A = PJP^{-1}, \text{ where } P = \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & 1 \\ 1 & -1 & 0 \end{pmatrix}, \text{ and } J = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

We know $A^n = PJ^nP^{-1}$. By block multiplication of matrices $J^n = \begin{pmatrix} 2^n & 0 \\ 0 & B^n \end{pmatrix}$, where $B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Note that $B = I + E$, where I is the 2×2 identity matrix and $E^2 = 0$. By the binomial theorem we have $B^n = I + nE + \binom{n}{2}E^2 + \cdots = I + nE$. Therefore,

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

□

12.5 Exercises

Exercise 12.1. *Determine if each of the following is true or false. Assume every appropriate operation is defined.*

(a) A self adjoint matrix A is positive definite if and only if $-A$ is negative definite.

(b) If $A > 0$ and $B > 0$, then $A + B > 0$.

(c) If $A > 0$ then A is invertible.

(d) If $A > 0$, then $A^3 > 0$.

(e) If $A < 0$, then $A^4 < 0$.

(f) If $A < 0$, then $A^4 > 0$.

(g) If $A > 0$ and B is indefinite, then $A + B > 0$.

(h) If $A \geq 0$ and $A \leq 0$, then $A = 0$.

Exercise 12.2. *Suppose $A \in M_n(\mathbb{F})$ is self adjoint. Prove that if the set consisting of the zero vector and all vectors $\mathbf{x} \in \mathbb{F}^n$ for which $\mathbf{x}^* A\mathbf{x} > 0$ is a subspace of $\mathbb{F}^n$, then A is positive definite.*

Exercise 12.3. *Determine the definiteness of each real quadratic form:*

(a) $Q(x, y, z) = x^2 + 3y^2 + 4z^2 - xy + 6yz - xz$.

(b) $Q(x, y) = x^2 + 2y^2 + 2axy$, where $a \in \mathbb{R}$ is a constant.

(c) $Q(x, y) = 2x^2 + 3y^2 - xy$.

(d) $Q(x_1, \ldots, x_n) = \sum_{j=1}^{n} x_j^2 + x_1 x_2$.

(e) $Q(x_1, \ldots, x_n) = \sum_{j=1}^{n} x_j^2 - \sum_{j=1}^{n-1} x_j x_{j+1}$.

Exercise 12.4. *Consider the (real) quadratic form*

$$Q(x, y, z) = 2(x + y)^2 + 2(2x + y - z)^2 - (x - z)^2.$$

(a) Using algebra show $Q(x, y, z) = (3x + 2y - z)^2$.

(b) Determine the definiteness of Q.

(c) Q seems to have two diagonalizations, one with positive and negative coefficients, and one with only nonnegative coefficients. How do you reconcile this with Theorem 12.1 and Corollary 12.1?

Exercise 12.5. *Suppose $A \in M_n(\mathbb{F})$ is self adjoint such that $\Delta_k(A) > 0$ for $k = 1, \ldots, n - 1$ and $\det A \geq 0$. Prove A is positive semidefinite.*

Exercise 12.6. *For each of the following quadratic forms*

1. $Q(x, y, z) = x^2 + y^2 + z^2 + 2xy - xz + 4yz$

2. $Q(x_1, x_2) = |x_1|^2 - i\overline{x}_1 x_2 + 3|x_2|^2 + ix_1\overline{x}_2$

do the following:

(a) Diagonalize the quadratic form.

(b) Determine the definiteness of Q using diagonalization found in part (a).

(c) Determine the definiteness of Q using Sylvester's Criterion.

Exercise 12.7. *Suppose $A \in M_n(\mathbb{F})$ is a nilpotent matrix. Prove that $A^n = 0$.*

Hint: Recall that $\sigma(A) = \{0\}$. Then use the Cayley-Hamilton Theorem.

Exercise 12.8. *Let $A = \begin{pmatrix} 0 & c \\ 0 & 0 \end{pmatrix}$, where $c \in \mathbb{F}$ is nonzero. Prove there is no matrix B for which $B^2 = A$.*

Exercise 12.9. *Let A be a square matrix.*

(a) Consider the matrices $I, A, A^2, \ldots$. Using the fact that $M_n(\mathbb{F})$ is a finite dimensional vector space over $\mathbb{F}$, prove that there are constants c_i, not all zero, for which $c_0 I + \cdots + c_{n^2} A^{n^2} = 0$. Deduce that there is a nonzero polynomial $f(t)$ for which $f(A) = 0$. (Do not use the Cayley-Hamilton Theorem for this part.)

(b) Prove that if $f(A) = 0$ for some $f(t) \in \mathbb{P}$ and λ is an eigenvalue of A, then $f(\lambda) = 0$.

(c) Suppose $g(t) \in \mathbb{P}$ is a monic polynomial with the smallest degree for which $g(A) = 0$. (Such a polynomial exists by part (a).) Let $h(t)$ be a polynomial for which $h(A) = 0$. Prove that $h(t)$ is divisible by $g(t)$. Use that to prove such a polynomial $g(t)$ is unique. (Hint: Using long division write $h(t) = g(t)q(t) + r(t)$, where $q(t)$ and $r(t)$ are the quotient and remainder when $h(t)$ is divided by $g(t)$.)

(d) Show that every root of the polynomial $g(t)$ in part (c) is an eigenvalue of A.

Remark 12.2. The polynomial $g(t)$ in the above exercise is called the **minimal polynomial** of A.

Exercise 12.10. *Write the following matrix as $A = PJP^{-1}$, where P is invertible and J is a matrix in Jordan form:*

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

Week 13

13.1 Jordan Canonical Form, Continued

Example 13.1. Write down the following matrix in the form SJS^{-1}, where J is a Jordan matrix and S is invertible:

$$\begin{pmatrix} 2 & -1 & 0 \\ -1 & 5 & -1 \\ -4 & 13 & -2 \end{pmatrix}$$

Theorem 13.1. *Every matrix in $M_n(\mathbb{C})$ is similar to a matrix in Jordan form.*

Theorem 13.2. *Let A be a matrix in $M_n(\mathbb{C})$ and J be a matrix in Jordan form that is similar to A. Then, for every positive integer k, the number of Jordan blocks of J with size at least $k \times k$ corresponding to an eigenvalue λ is given by*

$$\dim \operatorname{Ker} (A - \lambda I)^k - \dim \operatorname{Ker} (A - \lambda I)^{k-1}.$$

Here $(A - \lambda I)^0 = I$ and thus its kernel has dimension zero. Consequently, the matrix in Jordan form similar to A is unique up to a permutation of Jordan blocks.

Using the above theorem, we can find the Jordan form of any matrix rather easily, however finding the matrix P in $A = PJP^{-1}$ is more difficult.

Example 13.2. How many 5×5 nonsimilar matrices in Jordan form are there all of whose eigenvalues are 0?

Example 13.3. Find the number of nonsimilar 6×6 matrices in Jordan form whose eigenvalues are $1, 2, 2, 3, 3, 3$.

Example 13.4. The characteristic polynomial of a matrix A is $p(t) = t^6(t-1)^4$. Suppose

$$\dim \operatorname{Ker} A = 1, \text{ and } \dim \operatorname{Ker} (A - I) = 3.$$

Find a matrix in Jordan form that is similar to A.

To find P in $A = PJP^{-1}$, where J is in Jordan form, start with a vector $\mathbf{v}_k$ in $\operatorname{Ker}(A - \lambda I)^k$ that does not lie in $\operatorname{Ker}(A - \lambda I)^{k-1}$. Evaluate vectors $\mathbf{v}_{k-1} = (A - \lambda I)\mathbf{v}_k$, $\mathbf{v}_{k-2} = (A - \lambda I)\mathbf{v}_{k-1}, \ldots, \mathbf{v}_1 = (A - \lambda I)\mathbf{v}_2$. We need to repeat this process and make sure we obtain a basis for each generalized eigenspace. This process is not easy for large matrices and we skip the details. These $\mathbf{v}_j$'s give us columns of matrix P.

13.2 Applications of Jordan Canonical Form

Theorem 13.3. *For a matrix $A \in M_n(\mathbb{C})$, there are matrices $D, N \in M_n(\mathbb{C})$ for which all of the following hold:*

- *D is diagonalizable and N is nilpotent;*
- *The eigenvalues of D and A are the same;*
- *$DN = ND$; and*
- *$A = D + N$.*

Previously we defined e^A for any square matrix A by

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!},$$

but we never proved this sum in fact converges. Here we will prove that and we will also define $f(A)$ for a class of functions called analytic functions.

Definition 13.1. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be **analytic** iff there is a sequence a_n of real numbers for which

$$f(x) = \sum_{n=0}^{\infty} a_n x^n, \text{ for all } x \in \mathbb{R}.$$

Theorem 13.4. *Suppose $f : \mathbb{R} \to \mathbb{R}$ is analytic, i.e. $f(x) = \sum\limits_{n=0}^{\infty} a_n x^n$ for some sequence $a_n \in \mathbb{R}$. Then, $a_n = \dfrac{f^{(n)}(0)}{n!}$ and $\sum\limits_{n=0}^{\infty} a_n z^n$ converges, for all $z \in \mathbb{C}$.*

Definition 13.2. Suppose $f : \mathbb{R} \to \mathbb{R}$ is analytic, i.e. $f(x) = \sum\limits_{n=0}^{\infty} a_n x^n$ for some sequence $a_n \in \mathbb{R}$. The function $F : \mathbb{C} \to \mathbb{C}$ defined by $F(z) = \sum\limits_{n=0}^{\infty} a_n z^n$ is called the **analytic continuation** of f. Eventhough F and f have different domains, for simplicity, we denote both function by "f".

Definition 13.3. Suppose $f(x) = \sum\limits_{n=0}^{\infty} a_n x^n$ is an analytic function. Given a square matrix A we define $f(A) = \lim\limits_{m \to \infty} \sum\limits_{n=0}^{m} a_n A^n$, if the limit exists.

Definition 13.4. Given a sequence of matrices $A_n = (a_{ij,n})$, we define the matrix $A = \lim\limits_{n \to \infty} A_n$ to be the matrix whose (i, j) entry is the limit of the sequence of (i, j) entries of A_n. In other words

$$\lim_{n \to \infty} (a_{ij,n}) = (\lim_{n \to \infty} a_{ij,n}).$$

When $A_n(t)$ is a sequence of matrices whose entries are functions of t, then their limit $A(t)$ is defined the same way for every real number t.

The following theorem can be easily proved using the above definition and properties of limit.

Theorem 13.5. *Let j, k, ℓ be three positive integers and let A_n, B_n be two sequences of $j \times k$ matrices, and C_n be a sequence of $k \times \ell$ matrices. Let $a_n \in \mathbb{F}$ be a sequence of scalars. Suppose*

$$\lim_{n\to\infty} A_n = A, \lim_{n\to\infty} B_n = B, \lim_{n\to\infty} C_n = C, \text{ and } \lim_{n\to\infty} a_n = a.$$

Then,

- $\lim_{n\to\infty}(A_n + B_n) = A + B.$
- $\lim_{n\to\infty}(A_n C_n) = AC.$
- $\lim_{n\to\infty}(a_n A_n) = aA.$

Now, suppose $f(t)$ is an analytic function. By Theorem 13.3, we can write $A = D + N$, where D is diagonalizable and N is nilpotent and $ND = DN$. We will define $f(A) = \sum_{k=0}^{\infty} a_k A^k$, where $\sum_{k=0}^{\infty} a_k t^k$ is the Taylor series for $f(t)$. We need to show $\sum_{k=0}^{\infty} a_k A^k$ converges for every square matrix A.

Let $p_m(t) = \sum_{k=0}^{m} a_k t^k$ be the m-th partial sum of the Taylor series for $f(t)$.

Since D is diagonalizable we can write $D = S \operatorname{diag}(\lambda_1, \ldots, \lambda_n)\, S^{-1}$. Since, p_m is a polynomial, we have:

$$p_m(D) = S \operatorname{diag}(p_m(\lambda_1), \ldots, p_m(\lambda_n))\, S^{-1}$$

As $m \to \infty$, we have $p_m(\lambda_j) \to f(\lambda_j)$ for every j. Therefore,

$$f(D) = \lim_{m\to\infty} p_m(D) = S \operatorname{diag}(f(\lambda_1), \ldots, f(\lambda_n))\, S^{-1} \qquad (*)$$

From calculus, we know the Taylor polynomial of p_m centered at x_0 is given by

$$p_m(x_0 + h) = \sum_{k=0}^{\infty} \frac{p_m^{(k)}(x_0)}{k!} h^k.$$

Since the k-th derivative of p_m when $k > \deg p_m$ is zero, the above sum is a finite sum, and it will terminate at $k = \deg p_m$. Since N and D commute, we can substitute $x_0 = D$ and $h = N$ to obtain

$$p_m(D + N) = \sum_{k=0}^{\infty} \frac{p_m^{(k)}(D)}{k!} N^k.$$

Since N is nilpotent, $N^\ell = 0$ for some positive integer ℓ. Therefore,

$$p_m(A) = \sum_{k=0}^{\ell-1} \frac{p_m^{(k)}(D)}{k!} N^k,$$

since $N^\ell = N^{\ell+1} = \cdots = 0$.

Applying the term by term differentiation of power series to $f(x) = \sum_{n=0}^{\infty} a_n x^n$ we obtain $f^{(k)}(x) = \sum_{n=0}^{\infty} \dfrac{d^k(a_n x^n)}{dx^k}$, which means the polynomials $p_m^{(k)}(x)$ are partial sums of $f^{(k)}(x)$. By $(*)$ we have $\lim_{m\to\infty} p_m^{(k)}(D) = f^{(k)}(D)$ exists. Therefore, we have:

$$f(A) = \lim_{m\to\infty} p_m(A) = \sum_{k=0}^{\ell-1} \frac{f^{(k)}(D)}{k!} N^k.$$

This means the $p_m(A)$ converges and thus the power series for $f(A)$ is convergent.

If we substitute $f(x) = e^x$, we obtain: $e^A = e^D \sum_{k=0}^{n-1} \dfrac{N^k}{k!}$.

We summarize this in the following two theorems.

Theorem 13.6. *Suppose $f : \mathbb{R} \to \mathbb{R}$ is an analytic function, $D = P\text{diag}(\lambda_1, \ldots, \lambda_n)P^{-1}$ is a diagonalizable matrix. Then,*

$$f(D) = P\text{diag}(f(\lambda_1), \ldots, f(\lambda_n))P^{-1}.$$

Theorem 13.7. *Suppose $f : \mathbb{R} \to \mathbb{R}$ is an analytic function, with its Taylor series given as $f(t) = \sum_{k=0}^{\infty} a_k t^k$. Then, for every matrix $A \in M_n(\mathbb{C})$, the series $f(A) = \sum_{k=0}^{\infty} a_k A^k$ converges. Furthermore, if $A = D + N$ with $DN = ND$, D diagonalizable, and N nilpotent, then $f(A) = \sum_{k=0}^{n-1} \dfrac{f^{(k)}(D)}{k!} N^k$.*

(Note: We define $A^0 = I$ for every square matrix A.)

Example 13.5. Evaluate $\sin A$ and e^A, where $A = \begin{pmatrix} 3 & -1 \\ 1 & 1 \end{pmatrix}$.

13.3 Examples

Example 13.6. Suppose $A = PBP^{-1}$ for three square matrices A, B, P. Prove that $e^A = Pe^B P^{-1}$.

Solution. By definition $e^A = \lim_{m\to\infty} p_m(A)$, where $p_m(z) = \sum_{k=0}^{m} \dfrac{z^k}{k!}$. Substituting $A = PBP^{-1}$ we obtain the following:

$$p_m(A) = \sum_{k=0}^{m} \frac{(PBP^{-1})^k}{k!} = \sum_{k=0}^{m} \frac{PB^kP^{-1}}{k!} = P\left(\sum_{k=0}^{m} \frac{B^k}{k!}\right)P^{-1} = P\ p_m(B)\ P^{-1}.$$

By properties of limit we have the following:

$$e^A = \lim_{m\to\infty} p_m(A) = \lim_{m\to\infty} P\ p_m(B)\ P^{-1} = P\left(\lim_{m\to\infty} p_m(B)\right)P^{-1} = Pe^BP^{-1}.$$

Therefore, $e^A = Pe^BP^{-1}$. □

Example 13.7. Find all $A \in M_n(\mathbb{C})$ satisfying $A^2 = A$.

Solution. First, note that if λ is an eigenvalue of A, then $A\mathbf{v} = \lambda\mathbf{v}$ for some nonzero $\mathbf{v}$ and thus $A^2\mathbf{v} = \lambda^2\mathbf{v}$, which implies $\lambda^2 = \lambda$, since $A = A^2$. Therefore, $\lambda = 0, 1$. We will now write A in Jordan form: $A = PJP^{-1}$.

$$A^2 = A \iff PJ^2P^{-1} = PJP^{-1} \iff J^2 = J.$$

By block multiplication of matrices $B^2 = B$ for every Jordan block of J. If B is of size more than 1×1, then the $(1, 2)$ entry of B^2 is 2λ, while the $(1, 2)$ entry of B is 1. Therefore, $\lambda = 1/2$, which is a contradiction. Therefore, all Jordan blocks of J are 1×1, and thus A is diagonalizable. This means $A^2 = A$ if and only if $A = PDP^{-1}$, where D is a diagonal matrix whose diagonal entries are 0 and 1. $\square$

Example 13.8. Evaluate $\sin(A)$, where $A = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$.

Solution. Let $A = -I + N$, where $N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. By Theorem 13.7 and Theorem 13.6, we have,

$$\sin(A) = \sin(-I) + \cos(-I)N = \begin{pmatrix} \sin(-1) & 0 \\ 0 & \sin(-1) \end{pmatrix} + \begin{pmatrix} \cos(-1) & 0 \\ 0 & \cos(-1) \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Therefore,

$$\sin(A) = \begin{pmatrix} -\sin 1 & \cos 1 \\ 0 & -\sin 1 \end{pmatrix}.$$

$\square$

13.4 Exercises

Definition 13.5. For every positive integer n, the number of sequences of positive integers $a_1 \leq a_2 \leq a_2 \leq \cdots \leq a_k$ satisfying $a_1 + a_2 + \cdots + a_k = n$ is denoted by $p(n)$.

The answer to the next problem could be in terms of the function $p(n)$ defined above.

Exercise 13.1. *How many $n \times n$ matrices J in Jordan form are there that have a single given eigenvalue λ? How about if J were to have two distinct given eigenvalues λ_1 and λ_2?*

Exercise 13.2. *Two $(n+1) \times (n+1)$ matrices A and B with complex entries are given. Assume the list of eigenvalues of both A and B is*

$$1, 2, \ldots, n-2, n-1, n, n.$$

Suppose further that neither A nor B is diagonalizable. Prove that A and B are similar matrices.

Hint: Find Jordan forms of A and B.

Exercise 13.3. *Find $\cos(A)$ and e^A if $A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$.*

Exercise 13.4. *Suppose B is a Jordan block with eigenvalue λ, i.e. a square matrix of the form:*

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

(a) Prove that B is similar to B^T.

(b) Using part (a) and Jordan canonical form, prove that every matrix in $M_n(\mathbb{C})$ is similar to its transpose.

Exercise 13.5. *Evaluate e^B, where B is the $n \times n$ Jordan block with λ on its main diagonal.*

Exercise 13.6. *Show the number of Jordan blocks of size $k \times k$ corresponding to an eigenvalue λ of a square matrix A is given by the following formula:*

$$2 \dim \operatorname{Ker} (A - \lambda I)^k - \dim \operatorname{Ker} (A - \lambda I)^{k+1} - \dim \operatorname{Ker} (A - \lambda I)^{k-1}.$$

Exercise 13.7. *Suppose $A \in M_n(\mathbb{C})$. Assume $\operatorname{rank} A^m = \operatorname{rank} A^{m+1}$ for some positive integer m. Prove $\operatorname{rank} A^m = \operatorname{rank} A^k$ for every integer $k \geq m$.*

Exercise 13.8. *Prove a matrix $A \in M_n(\mathbb{C})$ is diagonalizable if and only if $\operatorname{rank} (A - \lambda I) = \operatorname{rank} (A - \lambda I)^2$ for every $\lambda \in \sigma(A)$.*

Exercise 13.9. *Given nonzero numbers $a_1, \ldots, a_n$ find the matrix in Jordan form that is similar to the $(n+1) \times (n+1)$ matrix (shown below) whose entries immediately below the main diagonal are $a_1, \ldots, a_n$, and all of its other entries are zero.*

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ a_1 & 0 & \cdots & 0 & 0 \\ 0 & a_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a_n & 0 \end{pmatrix}.$$

Exercise 13.10. *Show that a matrix is diagonalizable if and only if its Jordan form J is diagonal.*

Exercise 13.11. *Write the matrix A in the form PJP^{-1}, where J is in Jordan form:*

$$A = \begin{pmatrix} 3 & -1 & 0 & 0 \\ 9 & -3 & 0 & 0 \\ 0 & 0 & 5 & -2 \\ 0 & 0 & 12 & -5 \end{pmatrix}.$$

Use this to find e^A. You could leave your answers as products of matrices.

Exercise 13.12. *Let A be an $n \times n$ matrix. Recall that A can be written as $A = D + N$, where D is diagonalizable, N is nilpotent and $ND = DN$. Also, recall that eigenvalues of A and D are the same.*

(a) Using the fact that N is nilpotent, prove that for any positive integer m, we have $A^m = \sum_{k=0}^{n-1} \binom{m}{k} D^{m-k} N^k$.

(b) Suppose every eigenvalue λ of A satisfies $|\lambda| < 1$. Prove that D^m approaches the zero matrix as $m \to \infty$. Deduce that A^m approaches the zero matrix as $m \to \infty$.

Hint: You may use the fact that exponential decay is faster than polynomial growth.

Exercise 13.13. *Suppose A is a square matrix satisfying $A^4 = A$. Prove that A is diagonalizable.*

Exercise 13.14. *Find $\cos(A)$ and e^A if A is each of the following matrices.*

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

Exercise 13.15. *Suppose λ is an eigenvalue of an $n \times n$ matrix A for which $\operatorname{Ker}(A - \lambda I)^{n-1} \neq \operatorname{Ker}(A - \lambda I)^n$. Prove that A is similar to a single Jordan block.*

Exercise 13.16. *Prove that for every analytic function f, two square matrices A, P of the same size, with P being invertible, we have $f(PAP^{-1}) = Pf(A)P^{-1}$.*

Exercise 13.17. *Prove that for every block matrix*

$$A = \begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix},$$

where B, C are square matrices, and every analytic function $f : \mathbb{R} \to \mathbb{R}$ we have

$$f(A) = \begin{pmatrix} f(B) & 0 \\ 0 & f(C) \end{pmatrix}.$$

Exercise 13.18. *Consider the matrices*

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \text{ and } \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}.$$

Evaluate e^A, e^B and e^{A+B}. Show that $e^A e^B \neq e^{A+B}$.

Week 14

14.1 Quotient Spaces

Definition 14.1. Given a subspace W of a vector space V and a vector $\mathbf{x} \in V$, the **coset** of W in V corresponding to $\mathbf{x}$ is the set denoted by $\mathbf{x} + W$ and defined by:

$$\mathbf{x} + W = \{\mathbf{x} + \mathbf{w} \mid \mathbf{w} \in W\}.$$

Example 14.1. Let W be a 1-dimensional subspace of $\mathbb{R}^2$. Then cosets of W in $\mathbb{R}^2$ are precisely all lines identical or parallel to W.

Theorem 14.1. *Suppose W is a subspace of a vector space V and let $\mathbf{x}, \mathbf{y} \in V$. Then,*

(a) Either $\mathbf{x} + W = \mathbf{y} + W$ or $(\mathbf{x} + W) \cap (\mathbf{y} + W) = \emptyset$ (but not both!)

(b) $\mathbf{x} + W = \mathbf{y} + W$ if and only if $\mathbf{x} - \mathbf{y} \in W$.

(c) $\mathbf{x} + W = W$ if and only if $\mathbf{x} \in W$.

Remark 14.1. Note that a coset $\mathbf{x} + W$ typically has many representations. In other words $\mathbf{x} + W$ may be the same as $\mathbf{y} + W$ even though $\mathbf{x}$ and $\mathbf{y}$ are not the same. Because of that when a definition depends on the choice of a representative $\mathbf{x}$ we must make sure the definition is valid. This is phrased as the definition is **well-defined**.

Definition 14.2. Given a subspace W of a vector space V, the set V/W (or $\dfrac{V}{W}$) is the set of all cosets of W in V. In other words,

$$\frac{V}{W} = \{\mathbf{x} + W \mid \mathbf{x} \in V\}.$$

Theorem 14.2. *Given a subspace W of a vector space V, the set V/W along with the following addition and scalar multiplication*

$$(\mathbf{x} + W) + (\mathbf{y} + W) = (\mathbf{x} + \mathbf{y}) + W, \text{ and } c(\mathbf{x} + W) = c\mathbf{x} + W, \forall \mathbf{x}, \mathbf{y} \in V, \forall c \in \mathbb{F}$$

is a vector space.

Definition 14.3. The vector space $\dfrac{V}{W}$ is called the **quotient space** of V by W.

Theorem 14.3. *Let W be a subspace of a finite dimensional vector space V. Then,*

$$\dim(V/W) = \dim V - \dim W.$$

Furthermore, if $\mathbf{v}_1, \ldots, \mathbf{v}_k$ form a basis for W and $\mathbf{v}_{k+1}+W, \ldots, \mathbf{v}_n+W$ form a basis for $\dfrac{V}{W}$, then $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form a basis for V.

Theorem 14.4 (First Isomorphism Theorem)**.** *Let $L : V \to U$ be a linear transformation of (not necessarily finite dimensional) vector spaces. Then, $\overline{L} : \dfrac{V}{\operatorname{Ker} L} \to L(V)$ given by $\overline{L}(\mathbf{v}+\operatorname{Ker} L) = L(\mathbf{v})$ is an isomorphism.*

Example 14.2. Let $c \in \mathbb{F}$. Find a linear transformation from $\mathbb{P}_n$ to $\mathbb{F}$ whose kernel is

$$V = \{f(x) \in \mathbb{P}_n \mid f(c) = 0\}.$$

Use that to show V is a subspace of $\mathbb{P}_n$ of dimension n.

Example 14.3. For every vector space V, the vector space $V/\{\mathbf{0}\}$ is isomorphic to V.

Definition 14.4. For every two subspaces W and U of a vector space V, we define

$$W + U = \{\mathbf{w} + \mathbf{u} \mid \mathbf{w} \in W, \text{ and } \mathbf{u} \in U\}.$$

Theorem 14.5 (Second Isomorphism Theorem)**.** *Let W, U be two subspaces of a (not necessarily finite dimensional) vector space V. Then, $W + U$ is a subspace of V and*

$$L : \frac{W+U}{U} \to \frac{W}{W \cap U}, \text{ given by } L(\mathbf{w} + U) = \mathbf{w} + (W \cap U), \forall \mathbf{w} \in W,$$

is an isomorphism. Furthermore, if $W + U$ is finite dimensional, then

$$\dim(W + U) = \dim W + \dim U - \dim(W \cap U).$$

Theorem 14.6 (Third Isomorphism Theorem)**.** *Suppose $W \subseteq U$ are subspaces of a (not necessarily finite dimensional) vector space V. Then, $\dfrac{U}{W}$ is a subspace of $\dfrac{V}{W}$, and*

$$L : \frac{V/W}{U/W} \to \frac{V}{U}, \text{ given by } L\left((\mathbf{v} + W) + \frac{U}{W}\right) = \mathbf{v} + U, \forall \mathbf{v} \in V,$$

is an isomorphism.

Quotient spaces can be used to eliminate "unwanted" portions of a vector space. What follows illustrates an example.

Definition 14.5. A **seminorm** on a vector space V is a function $p : V \to \mathbb{R}$ satisfying all of the following:

(a) (Homogeneity) $p(c\mathbf{x}) = |c|p(\mathbf{x})$ for all $c \in \mathbb{F}$ and all $\mathbf{x} \in V$,

(b) (Triangle Inequality) $p(\mathbf{x} + \mathbf{y}) \leq p(\mathbf{x}) + p(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in V$, and

(c) (Nonnegativity) $p(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in V$.

Note that in this definition we do not require $p(\mathbf{x})$ to be nonzero for every nonzero vector $\mathbf{x}$.

Example 14.4. Consider the vector space V consisting of all integrable function $f : [0,1] \to \mathbb{R}$. The following is a seminorm on V, which is not a norm:

$$p(f) = \left(\int_0^1 (f(x))^2 \, dx\right)^{1/2}.$$

Theorem 14.7. *Suppose V is a vector space equipped with a seminorm p. Let W be the subset of V consisting of every vector in V whose seminorm is zero. In other words*

$$W = \{\mathbf{x} \in V \mid p(\mathbf{x}) = 0\}.$$

Then, the following is a norm on V/W:

$$||\mathbf{x} + W|| = p(\mathbf{x}), \ \forall \mathbf{x} \in V.$$

14.2 Examples

Example 14.5. Let V be the subspace of $\mathbb{F}^3$ spanned by $(1, -1, 3)$. Find all vectors $\mathbf{v} \in \mathbb{F}^3$ for which $\mathbf{v} + V = (1, 2, 0) + V$.

Solution. Let $\mathbf{v} = (x, y, z)$. By Theorem 14.1, $(x, y, z) + V = (1, 2, 0) + V$ if and only if $(x-1, y-2, z) \in V$. This is equivalent to $(x - 1, y - 2, z) = (c, -c, 3c)$, which implies $x = c + 1, y = -c + 2, z = 3c$. Thus, $\mathbf{v} = (c+1, -c+2, 3c)$, for some $c \in \mathbb{F}$. □

Example 14.6. Let W be a subspace of a vector space V. Prove that every subspace of V/W is of the form X/W, where X is a subspace of V containing W.

Solution. Let A be a subspace of V/W, and let X be given by

$$X = \{\mathbf{x} \in V \mid \mathbf{x} + W \in A\}.$$

We will show X is a subspace of V containing W and that $A = X/W$.

First, note that if $\mathbf{w} \in W$, then $\mathbf{w} + W = \mathbf{0} + W \in A$, and thus $\mathbf{w} \in X$. Therefore, X contains W.

Now, assume $\mathbf{x}, \mathbf{y} \in X$ and $c \in \mathbb{F}$. By definition, $\mathbf{x} + W$ and $\mathbf{y} + W$ are both in A. Since A is a subspace of V/W, we have $(\mathbf{x} + W) + (\mathbf{y} + W), c(\mathbf{x} + W) \in A$. Therefore, $(\mathbf{x} + \mathbf{y}) + W$ and $(c\mathbf{x}) + W$ are both in A. Thus, $\mathbf{x} + \mathbf{y}$ and $c\mathbf{x}$ are in X. Thus, by the subspace criterion, X is a subspace of V.

We will now show $A = X/W$. If $\mathbf{a} + W \in A$, then by definition of X, we have $\mathbf{a} \in X$ and thus $\mathbf{a} + W \in X/W$. Suppose $\mathbf{x} + W \in X/W$. By definition of X we have $\mathbf{x} + W \in A$. Thus, $A = X/W$, as desired. □

In the next problem we will use the fact that for every two vector spaces X and Y, their Cartesian product, $X \times Y$ is also a vector space of dimension $\dim X + \dim Y$. See Exercise 3.13.

Example 14.7. Suppose V is a finite dimensional vector space. The **diagonal** of the vector space $V \times V$ is the set given by

$$D = \{(\mathbf{v}, \mathbf{v}) \mid \mathbf{v} \in V\}.$$

Using the First Isomorphism Theorem, prove that D is a subspace of $V \times V$ whose dimension is the same as $\dim V$.

Solution. Define $L : V \times V \to V$ by $L(\mathbf{x}, \mathbf{y}) = \mathbf{x} - \mathbf{y}$. We will show L is linear. Let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_1, \mathbf{y}_2 \in V$ and $c \in \mathbb{F}$.

$$L((\mathbf{x}_1, \mathbf{y}_1) + c(\mathbf{x}_2, \mathbf{y}_2)) = L(\mathbf{x}_1 + c\mathbf{x}_2, \mathbf{y}_1 + c\mathbf{y}_2) = \mathbf{x}_1 + c\mathbf{x}_2 - \mathbf{y}_1 - c\mathbf{y}_2 = L(\mathbf{x}_1, \mathbf{y}_1) + cL(\mathbf{x}_2, \mathbf{y}_2).$$

Therefore, L is linear. The kernel of L is precisely D. Note also that $L(\mathbf{x}, \mathbf{0}) = \mathbf{x}$, which implies L is onto. Therefore by the First Isomorphism Theorem, $(V \times V)/D$ is isomorphic to V and thus $\dim(V \times V) - \dim D = \dim V$. Since $\dim(V \times V) = \dim V + \dim V$, we conclude $\dim D = \dim V$. $\square$

Example 14.8. Let V be a vector space and E be a subspace of V. Define $\phi : \dfrac{V}{E} \to E^\perp$ by $\phi(\mathbf{x} + E) = P_{E^\perp}(\mathbf{x})$. Show that ϕ is an isomorphism. Solve this problem once directly using the definition of an isomorphism, and once using the First Isomorphism Theorem.

Solution 1. First, we will show ϕ is well-defined. Suppose $\mathbf{x} + E = \mathbf{y} + E$ for two vectors $\mathbf{x}, \mathbf{y} \in V$. By Theorem 14.1, $\mathbf{x} - \mathbf{y} \in E$. Since $P_{E^\perp}(\mathbf{x}) - \mathbf{x} \in E$, we have $P_{E^\perp}(\mathbf{x}) - \mathbf{y} \in E$. On the other hand $P_{E^\perp}(\mathbf{x})$ is in $E^\perp$. Therefore, $P_{E^\perp}(\mathbf{y}) = P_{E^\perp}(\mathbf{x})$.

First, we will show ϕ is linear. Let $\mathbf{x}, \mathbf{y} \in V$ and $c \in \mathbb{F}$.

$$\phi((\mathbf{x} + E) + c(\mathbf{y} + E)) = \phi((\mathbf{x} + c\mathbf{y}) + E) = P_{E^\perp}(\mathbf{x} + c\mathbf{y}) = P_{E^\perp}(\mathbf{x}) + cP_{E^\perp}(\mathbf{y}) = \phi(\mathbf{x} + E) + c\phi(\mathbf{y} + E).$$

Thus, ϕ is linear. Now, we will show ϕ is one-to-one. Suppose $\phi(\mathbf{x} + E) = \phi(\mathbf{y} + E)$. By definition of ϕ we have $P_{E^\perp}(\mathbf{x}) = P_{E^\perp}(\mathbf{y})$. Therefore, $P_{E^\perp}(\mathbf{x} - \mathbf{y}) = \mathbf{0}$. This implies $\mathbf{x} - \mathbf{y} - \mathbf{0}$ is orthogonal to $E^\perp$, and thus $\mathbf{x} - \mathbf{y} \in E$. By Theorem 14.1, we have $\mathbf{x} + E = \mathbf{y} + E$. Thus, ϕ is one-to-one.

Now, we will show ϕ is onto. Let $\mathbf{x} \in E^\perp$. Note that $\phi(\mathbf{x} + E) = P_{E^\perp}(\mathbf{x}) = \mathbf{x}$, since $\mathbf{x} \in E^\perp$. This implies ϕ is an isomorphism. $\square$

Solution 2. We apply the First Isomorphism Theorem: We know $P_{E^\perp} : V \to E^\perp$ is a linear transformation. Also, $P_{E^\perp}$ is onto since if $\mathbf{x} \in E^\perp$, then $P_{E^\perp}(\mathbf{x}) = \mathbf{x}$. Thus, by the First Isomorphism Theorem, ϕ defined above is an isomorphism. $\square$

14.3 Exercises

Exercise 14.1. *Consider $\mathbb{P}_2$ as a subspace of $\mathbb{P}_4$. Find all polynomials $f(t) \in \mathbb{P}_4$ for which $f(t) + \mathbb{P}_2 = (1 + t^4) + \mathbb{P}_2$.*

Exercise 14.2. *Suppose W is a subspace of a vector space V that has only finitely many cosets. Prove that $W = V$.*

Exercise 14.3. *We know the intersection of every two subspaces is another subspace. Prove that the intersection of every two cosets is either empty or it is another coset. In other words, show if U and W are subspaces of a vector space V and $\mathbf{x}, \mathbf{y} \in V$, then $(\mathbf{x} + U) \cap (\mathbf{y} + W)$ is either empty or it is a coset of $U \cap W$.*

Exercise 14.4. *Suppose X, Y are vector spaces over the same field. Using the First Isomorphism Theorem, prove that $\dfrac{X \times Y}{X \times \{\mathbf{0}\}}$ is isomorphic to Y. Deduce that $\dim(X \times Y) = \dim X + \dim Y$.*

Exercise 14.5. *Suppose $L : X \to Y$ is a linear transformation of vector spaces. Using the First Isomorphism Theorem, prove that the set given by*

$$A = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{y} = L(\mathbf{x})\}.$$

is a subspace of $X \times Y$ and that its dimension is the same as the dimension of X.

Exercise 14.6. *Let E be the subspace of $\mathbb{R}^4$ generated by $(1, -1, 2, 1)^T$, $(2, 0, 1, -1)^T$ and $(4, -2, 5, 1)^T$. Find a basis for $\dfrac{\mathbb{R}^4}{E}$.*

Exercise 14.7. *Suppose W is a subspace of a finite dimensional vector space V. Prove that if $\mathbf{v}_1, \ldots, \mathbf{v}_k$ form a basis for W and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form a basis for V, then $\mathbf{v}_{k+1} + W, \ldots, \mathbf{v}_n + W$ form a basis for $\dfrac{V}{W}$.*

Exercise 14.8. *Determine if the vector space $C[\mathbb{R}]/C^1[\mathbb{R}]$ is finite dimensional. How about $C^n[\mathbb{R}]/C^{n+1}[\mathbb{R}]$, for a given positive integer n?*

Exercise 14.9. *Write an onto linear transformation from $\mathbb{P}$ to $\mathbb{P}$ whose kernel is all constant polynomials. Use the First Isomorphism Theorem to deduce $\dfrac{\mathbb{P}}{\mathbb{F}} \cong \mathbb{P}$. Using a similar method show $\dfrac{C^\infty[\mathbb{R}]}{A} \cong C^\infty[\mathbb{R}]$, where A is the vector space of all constant functions from $\mathbb{R}$ to $\mathbb{R}$.*

Basics of Complex Numbers

A.1 Definition and Basic Operations

Definition A.1. The set of **complex numbers**, denoted by $\mathbb{C}$, is defined as

$$\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\},$$

where i is a solution of the equation $i^2 = -1$. The form $a + bi$ for a complex number is called its **standard form**. Two complex numbers $a+bi$ and $c+di$ with $a, b, c, d \in \mathbb{R}$ are said to be equal if and only if $a = c$ and $b = d$. We say a and b are the **real** and the **imaginary** parts of the complex number $z = a+bi$, respectively. We denote these by $\mathrm{Re}\,(z)$ and $\mathrm{Im}\,(z)$, respectively.

The set $\mathbb{C}$ is equipped with two binary operations $+$ and $.$ as follows:

- $\forall a, b, c, d \in \mathbb{R}\ \ (a + bi) + (c + di) = (a + c) + (b + d)i.$
- $\forall a, b, c, d \in \mathbb{R}\ \ (a + bi) \cdot (c + di) = (ac - bd) + (ad + bc)i.$ (Or $(a + bi)(c + di)$, without the dot.)

Note: Both real and imaginary parts of a complex number are *real*.

Definition A.2. For a complex number $z = a + bi$, where a and b are real numbers, we define its **complex conjugate** as $\overline{z} = a - bi$ and its **absolute value** (or **norm**) as $|z| = \sqrt{a^2 + b^2}$.

Theorem A.1 (Field properties of $\mathbb{C}$)**.** *For every $x, y, z \in \mathbb{C}$*

- *(Commutativity) $x + y = y + x$, and $xy = yx$.*
- *(Associativity) $(x + y) + z = x + (y + z)$ and $(xy)z = x(yz)$.*
- *(Additive Identity) $x + 0 = x$. (Here zero of $\mathbb{C}$ is given by $0 = 0 + 0i$.)*
- *(Additive Inverse) There is $t \in \mathbb{C}$ for which $x + t = 0$. (When $x = a + bi$, we have $t = -a + (-b)i$.)*
- *(Multiplicative Inverse) If $x \neq 0$, there is some $t \in \mathbb{C}$ for which $xt = 1$. (t is denoted by x^{-1} or $1/x$.)*
- *(Distributivity) $x(y + z) = xy + xz$.*

Theorem A.2 (Properties of complex conjugate and norm)**.** *For every two complex numbers z and w, we have*

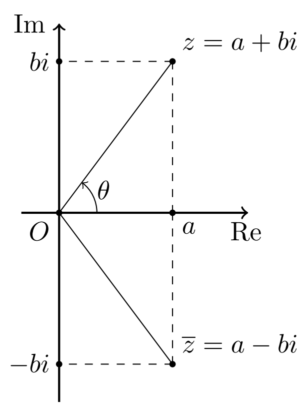
- $\overline{zw} = \overline{z}\,\overline{w}$.
- $|zw| = |z|\,|w|$.
- $|z|^2 = z\,\overline{z}$.
- *If $z \neq 0$, then* $z^{-1} = \dfrac{\overline{z}}{|z|^2}$.
- $|z+w| \leq |z| + |w|$. *(Triangle Inequality.)*

Example A.1. Find the additive and multiplicative inverse of $3+2i$.

Solution. Its additive inverse is $-3-2i$. Its multiplicative inverse is $\dfrac{3-2i}{9+4} = \dfrac{3}{13} - \dfrac{2}{13}i$. □

A.2 Geometry of $\mathbb{C}$

Each complex number $z = a+bi$ can be plotted on a plane called the **complex plane**. The horizontal axis consists of all real numbers and the vertical axis consists of all complex numbers with zero real part. If θ is the angle between $0z$ and the positive real axis, then $z = |z|(\cos\theta + i\sin\theta)$.



Recall the power series of $\cos\theta$, $\sin\theta$ and e^x are as follows:

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + - \cdots$$

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + - \cdots$$

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots$$

Multiplying $\sin\theta$ by i and adding it to $\cos\theta$ we obtain the series for e^x, when x is replaced by $i\theta$. This yields the following theorem:

Theorem A.3. *For every real number θ, we have $e^{i\theta} = \cos\theta + i\sin\theta$.*

Theorem A.4. *Let x, y be two real numbers and n be an integer. Then,*

(a) $e^{ix}e^{iy} = e^{i(x+y)}$.

(b) (De Moivre's Formula) $(e^{ix})^n = e^{inx}$.

A.3 Examples

Example A.2. Let $z = 2+i, w = 1-3i$. Write down the complex numbers $z+w, z-w, zw$, and z/w in standard form.

Solution. $z+w = 3-2i$, $z-w = 1+4i$, $zw = 2+i-6i-3i^2 = 5-5i$. $z/w = z\overline{w}/|w|^2 = (2+i+6i+3i^2)/(1+9) = -0.1+0.7i$. □

Example A.3. Find all real numbers a, b for which $a^2 + bi + 2i = (7+3i)(1-i)$.

Solution. Writing the left hand side in standard form and multiplying the right hand side we obtain:

$$a^2 + (b+2)i = 7 - 7i + 3i + 3 = 10 - 4i \Rightarrow a^2 = 10, \text{ and } b+2 = -4.$$

The answer is $a = \pm\sqrt{10}$, and $b = -6$. □

Example A.4. Evaluate $(1+i)^{1000}$

Solution. Since we are finding large exponents of a complex number, De Moiver's formula would be helpful. So, we will first write down this complex number in polar form. $|1+i| = \sqrt{2}$. The angle between the segment from 0 to $1+i$ and the positive real axis is $\pi/4$. This means $1+i = \sqrt{2}e^{i\pi/4}$. Therefore, $(1+i)^{1000} = 2^{500}e^{i250\pi} = 2^{500}(\cos(250\pi) + i\sin(250\pi)) = 2^{500}$. □

Example A.5. Given a positive integer n, solve the equation $z^n = 1$ over complex numbers.

Solution. By taking the absolute value of both sides we obtain $|z|^n = 1$. Since $|z|$ is a nonnegative real number, we must have $|z| = 1$. Therefore, using the polar form we obtain $z = e^{i\theta}$ for some angle $\theta \in [0, 2\pi)$. This means, we must solve $e^{in\theta} = 1$. This holds if and only if $\cos(n\theta) = 1$ and $\sin(n\theta) = 0$. This is equivalent to $n\theta = 2k\pi$ for some integer k. Since $\theta \in [0, 2\pi)$, we have $k = 0, 1, \ldots, n-1$. Therefore, all roots of $z^n = 1$ are $z = e^{2ik\pi/n}$ with $k = 0, 1, \ldots, n-1$. □

Example A.6. Prove that a complex number z satisfies $|z| = 1$ if and only if $z = e^{i\theta}$ for some real number θ.

Solution. ($\Rightarrow$) Suppose $|z| = 1$. By the polar form of z we know $z = |z|e^{i\theta} = e^{i\theta}$ for some $\theta \in [0, 2\pi)$.
($\Leftarrow$) Suppose $z = e^{i\theta}$ for some real number θ. Then, $z = \cos\theta + i\sin\theta$. Therefore, $|z| = \sqrt{\cos^2\theta + \sin^2\theta} = 1$. $\square$

A.4 Exercises

Exercise A.1. *Find all real numbers a, b for which $a + 3bi + a^2b = 2ab + ai + 2bi$.*

Exercise A.2. *Using the method of Mathematical Induction, prove the De Moivre's formula: $(e^{ix})^n = e^{inx}$ for every real number x and every integer n. Note that the cases where n is negative or positive should be dealt with separately.*

Exercise A.3. *Suppose z, w are complex numbers for which both $z + w$ and $i(z - w)$ are real. Prove that $z = \overline{w}$.*