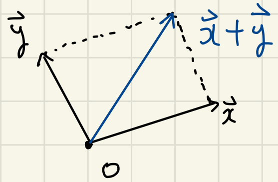
MAT224

Linear Algebra

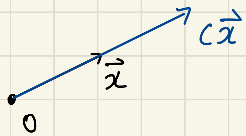
1.1 Vector Spaces

- **vector** – a directed line segment or "arrow"
    - describes quantities w/ a direction and a magnitude
    - magnitude – length of directed line
    - direction – direction the arrow is pointing
- Take two vectors $\vec{x}, \vec{y}$, we can define their **vector sum** $\vec{x}+\vec{y}$ to be the vector whose head is at the fourth corner of the parallelogram w/ sides $\vec{x}$ and $\vec{y}$

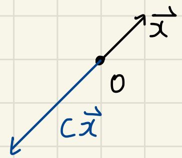


- Take a vector $\vec{x}$ and a positive real number c (called a **scalar**), we can define the **product** of the vector $\vec{x}$ and the scalar c to be the vector in the same direction as $\vec{x}$ but w/ a magnitude that is equal to c times the magnitude of $\vec{x}$

- In the case where $c < 0$, the vector $c\vec{x}$ will point along the same line through the origin as $\vec{x}$ but in the opposite direction from $\vec{x}$

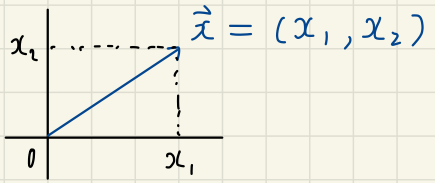


$c > 1$



$c < 0$

- vectors may be described as ordered pairs of real numbers



- if $\vec{x} = (x_1, x_2)$, $\vec{y} = (y_1, y_2)$, then $\vec{x} + \vec{y} = (x_1 + y_1, x_2 + y_2)$
- if c is a scalar, $\vec{x} = (x_1, x_2)$, then $c\vec{x} = (cx_1, cx_2)$

1. the vector sum is associative: $\forall \vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^2$,
$(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$

2. the vector sum is commutative: $\forall \vec{x}, \vec{y} \in \mathbb{R}^2$,

$\vec{x}+\vec{y}=\vec{y}+\vec{x}$

3. there is an additive identity element $\vec{0}=(0,0) \in \mathbb{R}^2$ w/ the property that $\forall \vec{x} \in \mathbb{R}^2$, $\vec{x}+\vec{0}=\vec{x}$

4. for each $\vec{x}=(x_1, x_2) \in \mathbb{R}^2$ there is an additive inverse $-\vec{x}=(-x_1,-x_2)$ w/ the property that $\vec{x}+(-\vec{x})=0$

5. multiplication by a scalar is distributive over vector sums: $\forall c \in \mathbb{R}$, $\vec{x},\vec{y} \in \mathbb{R}^2$, $c(\vec{x}+\vec{y})=c\vec{x}+c\vec{y}$

6. multiplication of a vector by a sum of scalars is distributive: $\forall c,d \in \mathbb{R}$, $\vec{x} \in \mathbb{R}^2$, $(c+d)\vec{x}=c\vec{x}+d\vec{x}$

7. $\forall c,d \in \mathbb{R}$, $\vec{x} \in \mathbb{R}^2$, $c(d\vec{x})=cd(\vec{x})$

8. $\forall \vec{x} \in \mathbb{R}^2$, $1\vec{x}=\vec{x}$

- consider the set of all functions $f: \mathbb{R} \to \mathbb{R}$ and denote the set by $F(\mathbb{R})$
  - if $f,g \in F(\mathbb{R})$, we can produce a new function denoted $f+g \in F(\mathbb{R})$ by defining

$(f+g)(x) = f(x)+g(x)$ for all $x \in \mathbb{R}$

- if $c \in \mathbb{R}$, we may multiply any $f \in F(\mathbb{R})$ by c to produce a new function denoted $cf \in F(\mathbb{R})$ and defined by

  $(cf)(x) = cf(x)$ for all $x \in \mathbb{R}$

1. the sum operation on functions is associative:

   $\forall f,g,h \in F(\mathbb{R}),\ (f+g)+h = f+(g+h)$

2. the sum operation on functions is commutative:

   $\forall f,g \in F(\mathbb{R}),\ f+g = g+f$

3. there is an additive identity element in $F(\mathbb{R})$:

   the constant function $z(x)=0$.

   $\forall f \in F(\mathbb{R}),\ f+z=0$

4. for each function f there is an additive inverse $-f \in F(\mathbb{R})$ with the property that

   $f+(-f)=z$ (the zero function)

5. $\forall f,g \in F(\mathbb{R}),\ c \in \mathbb{R},\ c(f+g) = cf+cg$

6. $\forall f \in F(\mathbb{R}),\ c,d \in \mathbb{R},\ (c+d)f = cf+df$

7. $\forall f \in F(\mathbb{R}),\ c,d \in \mathbb{R},\ (cd)f = c(df)$

8. $\forall f \in F(\mathbb{R})$, $1f = f$

- w.r.t the sum and scalar multiplication operations we have defined, the elements of our two sets $\mathbb{R}^2$, $F(\mathbb{R})$ behave the same way
- $\mathbb{R}^2$ and $F(\mathbb{R})$ are examples of a vector space
- (1.1.1) Definition. A (real) vector space is a set V (whose elements are called vectors) together w/
  - a) an operation called vector addition, which for each pair of vectors $\vec{x}, \vec{y} \in V$ produces another vector in V denoted $\vec{x}+\vec{y}$, and
  - b) an operation called multiplication by a scalar (a real number), which for each vector $\vec{x} \in V$, and each scalar $c \in \mathbb{R}$ produces another vector in V denoted $c\vec{x}$

Furthermore, the two operations must satisfy the following axioms:

1. For all vectors $\vec{x}, \vec{y}, \vec{z} \in V$,
   $(\vec{x}+\vec{y})+\vec{z} = \vec{x}+(\vec{y}+\vec{z})$

2. For all vectors $\vec{x}, \vec{y} \in V$, $\vec{x}+\vec{y} = \vec{y}+\vec{x}$

3. There exists a vector $\vec{0} \in V$ w/ the property that $\vec{x}+\vec{0} = \vec{x}$ for all vectors $\vec{x} \in V$

4. For each vector $\vec{x} \in V$, there exists a vector denoted $-\vec{x}$ w/ the property that $\vec{x} + (-\vec{x}) = 0$

5. For all vectors $\vec{x}, \vec{y} \in V$ and all scalars $c \in \mathbb{R}$, $c(\vec{x}+\vec{y}) = c\vec{x}+c\vec{y}$

6. For all vectors $\vec{x} \in V$, and all scalars $c, d \in \mathbb{R}$, $(c+d)\vec{x} = c\vec{x}+d\vec{x}$

7. For all vectors $\vec{x} \in V$, and all scalars $c, d \in \mathbb{R}$, $(cd)\vec{x} = c(d\vec{x})$

8. For all vectors $\vec{x} \in V$, $1\vec{x} = \vec{x}$

- Ex. In $V = \mathbb{R}^2$, if we defined

$(x_1, x_2) +' (y_1 + y_2) = (x_1 y_1, x_2 y_2)$

and $c(x_1, x_2) = (cx_1, cx_2)$

  - Axioms 1, 2 pass

- Axiom 3 passes w/ vector $(1,1)$
- Axiom 4 fails
  - no inverse for vector $(0,0)$
- And so we do not obtain a vector space

- Ex. Consider the addition operation $f +' g$ defined by $(f+'g)(x) = f(x) + 3g(x)$ and the usual scalar multiplication
  - Axiom 1 fails
- Ex. Consider the subset $V' = \mathbb{R}^+ = \{r \in \mathbb{R} \mid r > 0\}$ and define our vector sum and scalar multiplication operations to be $x +' y = x \cdot y$, $c \cdot' x = x^c$ for all $x, y \in V'$ and all $c \in \mathbb{R}$
  - Since multiplication of real numbers is associative and commutative, axioms 1 and 2 are satisfied
  - The element $1 \in V'$ is an identity element for the operation $+'$ (axiom 3 satisfied)
  - Each $x \in V'$ has an inverse $1/x \in V'$

under the operation $+'$ (axiom 4 satisfied)

- $c \cdot' (x +' y) = (x \cdot y)^c = x^c \cdot y^c = x^c +' y^c$

  $= (c \cdot' x) +' (c \cdot' y)$ (axiom 5 satisfied)

- $(c+d) \cdot' x = x^{c+d} = x^c \cdot x^d$

  $= (c \cdot' x) +' (d \cdot' x)$ (axiom 6 satisfied)

- $(cd) \cdot' x = x^{cd} = (x^d)^c = c \cdot' (d \cdot' x)$

  (axiom 7 satisfied)

- $1 \cdot' x = x^1 = x$ for all $x \in V'$

  (axiom 8 satisfied)

- In $\mathbb{R}^n$ the only one addictive identity is the zero vector $(0, 0, \ldots, 0) \in \mathbb{R}^n$, and each vector has only 1 additive inverse

- (1.1.6) Proposition. Let V be a vector space. Then

  a) The zero vector $\vec{0}$ is unique

  b) For all $\vec{x} \in V$, $0\vec{x} = \vec{0}$

  c) For each $\vec{x} \in V$, the additive inverse $-x$ is unique

  d) For all $\vec{x} \in V$, and all $c \in \mathbb{R}$,

$$(-c)\vec{x} = -(c\vec{x})$$

1.2 Subspaces

- (1.2.6) Definition. Let V be a vector space and let $W \subseteq V$ be a subset. Then W is a (vector) subspace of V if W is a vector space itself under the operations of vector sum and scalar multiplication from V.
  - Let $C(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous}\}$
  - $C(\mathbb{R})$ is a subspace of $F(\mathbb{R})$
- while in a subspace, we must use the same operations as in the larger vector space
  - $\mathbb{R}^+$ is not a subspace of $\mathbb{R}$ because vector sum and scalar multiplication operations in $\mathbb{R}^+$ are different from those in $\mathbb{R}$
- in every vector space, the subsets V and $\{0\}$ are subspaces
- if $W \subseteq V$ is to be a vector space, the

following properties must be true:

- $\forall \vec{x}, \vec{y} \in W$, $\vec{x} + \vec{y} \in W$
    - W is closed under addition
- $\forall \vec{x} \in W$, $\forall c \in \mathbb{R}$, $c\vec{x} \in W$
    - W is closed under scalar multiplication
- the zero vector of V must be contained in W

- (1.2.8) Theorem. Let V be a vector space, let W be a nonempty subset of V. Then W is a subspace of V if and only if for all $\vec{x}, \vec{y} \in W$, and all $c \in \mathbb{R}$, we have $c\vec{x} + \vec{y} \in W$
    - since a vector space contains an additive identity element, W has to be nonempty
- (1.2.13) Theorem. Let V be a subspace. Then the intersection of any collection of subspaces of V is a subspace of V
- (1.2.14) Corollary. Let a_{ij} $(1 \le i \le m, 1 \le j \le n)$

be any real numbers and let

$$W = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid a_{i1}x_1 + \cdots + a_{in}x_n = 0$$

for all i, $1 \leq i \leq m\}$. Then W is a subspace of $\mathbb{R}^n$

1.3 Linear Combinations

- (1.3.1) Definitions. Let S be a subset of a vector space V
  - a) a linear combination of vectors S is any sum $a_1\vec{x}_1 + \cdots + a_n\vec{x}_n$, where $a_i \in \mathbb{R}$, $x_i \in S$
  - b) if $S \neq \Phi$ (the empty subset of V), the set of all linear combinations of vectors in S is the (linear) span of S, denoted $\text{Span}(S)$.
    - if $S = \Phi$, we define
    $$\text{Span}(S) = \{0\}$$
  - c) if $W = \text{Span}(S)$, we say S spans (or generates) W
- the span of a set S is the set of all vectors that can be "built up" from the vectors in S by forming linear combinations

- Ex. In $V = \mathbb{R}^3$, let $S = \{(1,0,0), (0,1,0)\}$.

A typical linear combination of the vectors in S is a vector $a_1(1,0,0) + a_2(0,1,0) = (a_1, a_2, 0)$

- The span of S is the set of all such vectors
- $\text{Span}(S) = \{(a_1, a_2, 0) \in \mathbb{R}^3 \mid a_1, a_2 \in \mathbb{R}\}$
- $\text{Span}(S)$ is the x_1-x_2-plane in $\mathbb{R}^3$
- By Corollary (1.2.14), $\text{Span}(S)$ is a subspace in $\mathbb{R}^3$

- Ex. In $V = C(\mathbb{R})$ (set of all cont functions), let $S = \{1, x, x^2, \dots, x^n\}$

- we have $\text{Span}(S) = \{f \in C(\mathbb{R}) \mid f(x) = a_0 + a_1 x + \dots + a_n x^n$ for some $a_0, \dots, a_n \in \mathbb{R}\}$
- $\text{Span}(S)$ is the subspace $P_n(\mathbb{R})$ (set of all n^{th} degree polynomials) $\subset C(\mathbb{R})$

- The span of a set of vectors is a subspace of the vector space from which the vectors are chosen

- (1.3.4) Theorem. Let V be a vector space and let S be any subset of V. Then $\text{Span}(S)$ is a subspace of V

- (1.3.5) Definition. Let W_1 and W_2 be subspaces of a vector space V. The sum of W_1 and W_2 is the set $W_1 + W_2 = \{\vec{x} \in V \mid \vec{x} = \vec{x}_1 + \vec{x}_2$ for some $\vec{x}_1 \in W_1$ and $\vec{x}_2 \in W_2\}$

    - $W_1 + W_2$ is the set of vectors that can be "built up" from the vectors in W_1 and W_2 by linear combinations

    - Vectors in $W_1 + W_2$ are the vectors in V that can be "broken down" into the sum of a vector in W_1 and a vector in W_2

- Ex. Let $W_1 = \{(a_1, a_2) \in \mathbb{R}^2 \mid a_2 = 0\}$ and $W_2 = \{(a_1, a_2) \in \mathbb{R}^2 \mid a_1 = 0\}$

    - $W_1 + W_2 = \mathbb{R}^2$

    - every vector in $\mathbb{R}^2$ can be written as the sum of a vector in W_1 and a vectoin W_2

- Ex. In $V = C(\mathbb{R})$, let $W_1 = \text{Span}(\{1, e^x\})$ and $W_2 = \text{Span}(\{\sin x, \cos x\})$
  - $W_1 + W_2 = \text{Span}(\{1, e^x, \sin x, \cos x\})$
    - can be broken up into
      $a_1 + a_2 e^x + a_3 \sin x + a_4 \cos x$
  - every function in $\text{Span}(\{1, e^x, \sin x, \cos x\})$ is the sum of a function in W_1 and one in W_2
- if $W_1 = \text{Span}(S_1)$ and $W_2 = \text{Span}(S_2)$, then $W_1 + W_2$ is the span of the union of the spanning sets S_1 and S_2
- (1.3.8) Proposition. Let $W_1 = \text{Span}(S_1)$ and $W_2 = \text{Span}(S_2)$ be subspaces of a vector space V. Then $W_1 + W_2 = \text{Span}(S_1 \cup S_2)$
- (1.3.9) Theorem. Let W_1 and W_2 be subspaces of a vector space V. Then $W_1 + W_2$ is also a subspace of V

- If W_1 and W_2 are subspaces of V, then $W_1 \cup W_2$ is unlikely to be a subspace of V

$\Big\}\ W_1 \cup W_2$

$(1,0) + (0,1)$
$= (1,1) \notin W_1 \cup W_2$

- (1.3.11) Proposition. Let W_1 and W_2 be subspaces of a vector space V, let W be a subspace of V s.t. $W \supset (W_1 \cup W_2)$. Then $W \supset W_1 + W_2$
    - $W_1 + W_2$ is the smallest subspace of V containing $W_1 \cup W_2$
    - any other subspace of V that contains $W_1 \cup W_2$ contains $W_1 + W_2$

1.4 Linear Dependence and Linear Independence

- Ex. In $V = \mathbb{R}^3$, consider the subspace W spanned by the set $S = \{(1,2,1),(0,-1,3),(1,0,7)\}$
    - By definition (1.3.1), we have
      $W = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid (x_1, x_2, x_3) = a_1(1,2,1) + a_2(0,-1,3) + a_3(1,0,7)$ for some $a_i \in \mathbb{R}\}$
    - We can see that $(1,2,1) + 2(0,-1,3) = (1,0,7)$
    - there is a linear combination of the vectors in S w/ nonzero coefficients that adds up to the zero vector in $\mathbb{R}^3$
    - every vector in $\mathrm{Span}(S)$ is also in the span of the subset $S' = \{(1,2,1),(0,-1,3)\}$
    - $\mathrm{Span}(S) = \mathrm{Span}(S')$
- If we have any vector space V and a set of vectors $S \subset V$, and there is a relation

$a_1\vec{x}_1 + \cdots + a_n\vec{x}_n = \vec{0}$ where $a_i \in \mathbb{R}$ and $\vec{x}_i \in S$,

and at least one of the $a_i \neq 0$ (say a_n), we can solve for $\vec{x}_n$ so that

$$\vec{x}_n = -\frac{a_1}{a_n}\vec{x}_1 - \cdots - \frac{a_{n-1}}{a_n}\vec{x}_n$$

- hence $\vec{x}_n$ is a linear combination of the vectors $\vec{x}_1, \ldots, \vec{x}_{n-1}$
- in any linear combination involving $\vec{x}_n$, we could replace $\vec{x}_n$ by that linear combination of $\vec{x}_1, \ldots, \vec{x}_{n-1}$
- a smaller set spans the same space

- (1.4.2) Definitions. Let V be a vector space, let S be a subset of V
  a) a linear dependence among the vectors of S is an equation $a_1\vec{x}_1 + \cdots + a_n\vec{x}_n = \vec{0}$ where $\vec{x}_i \in S$, and $a_i \in \mathbb{R}$ are not all zero
  b) the set S is linearly dependent if there exists a linear dependence among

vectors in S

- Ex. $S = \{\vec{0}\}$ is linearly dependent in any vector space, since the equation $a\vec{0} = \vec{0}$ is linear dependence
  - if S is any set containing the vector $\vec{0}$ then S is linearly dependent

- Ex. In $V = C'(\mathbb{R})$ (set of continuously differentiable functions), consider the set $S = \{\sin x, \cos x\}$. Is S linearly dependent?
  - Do there exist scalars a_1, a_2, at least one of which is nonzero, such that

    $a_1 \sin x + a_2 \cos x = 0$ in $C'(\mathbb{R})$
  - If $x = 0$, we obtain $a_1 \sin 0 + a_2 \cos 0 = 0$. This implies $a_2 = 0$
  - If $x = \frac{\pi}{2}$, we obtain $a_1 \sin\frac{\pi}{2} + a_2 \cos\frac{\pi}{2} = 0$. This implies $a_1 = 0$
  - There are no linear dependences among the functions in S, and S is not linearly

dependent

- The set of vectors not linearly dependent is called a linearly independent set
- (1.4.4) Definition. A subset S of a vector space V is linearly independent if whenever we have $a_i \in \mathbb{R}$ and $\vec{x}_i \in S$ s.t. $a_1\vec{x}_1 + \cdots + a_n\vec{x}_n = \vec{0}$, then $a_i = 0$ for all i
- Ex. Take $S = \{\vec{e}_1, \ldots, \vec{e}_n\}$ where by definition, $\vec{e}_i$ is the vector w/ a 1 in the ith component and 0 in every other component.
  - S is linearly independent in $\mathbb{R}^n$
  - S also spans $\mathbb{R}^n$
  - S is known as the standard basis of $\mathbb{R}^n$
- In any vector space the empty subset $\emptyset$ is linearly independent
  - satisfies " if there are any linear combinations among the vectors in the set that add up to

zero, then all the coefficients must be zero"

- (1.4.7) Proposition.

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

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

1.5 Solving Systems of Linear Equations

- Problem 1: Given a subspace W of a vector space V, defined by giving some conditions on the vectors, find a set of vectors S such that $W = \text{span}(S)$.
- The subspaces we deal with are usually subspaces of $\mathbb{R}^n$ defined by systems of equations
- Problem 2: Given a set of vectors S and a vector $\vec{x} \in V$, determine if $\vec{x} \in \text{span}(S)$
- If $V = \mathbb{R}^m$, S is a set of n vectors in V, say, $S = \{(a_{11}, \dots, a_{m1}), \dots, (a_{1n}, \dots, a_{mn})\}$, $\vec{b} = (b_1, \dots, b_m) \in \mathbb{R}^m$, then $\vec{b} \in \text{span}(S)$ iff there are scalars $x_1, \dots, x_n$ s.t.

$$(b_1, \dots, b_m) = x_1(a_{11}, \dots, a_{m1}) + \dots + x_n(a_{1n}, \dots, a_{mn})$$

$$= (a_{11}x_1 + \dots + a_{1n}x_n, \dots, a_{m1}x_1 + \dots + a_{mn}x_n)$$

Equating the components in these two vectors,

$$a_{11}x_1 + \cdots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + \cdots + a_{2n}x_n = b_2$$

$$\vdots$$

$$a_{m1}x_1 + \cdots + a_{mn}x_n = b_m$$

- Problem 3: Given a set of vectors, S, determine if S is linearly dependent or linearly independent.

- Let $V = \mathbb{R}^m$, let $S = \{(a_{11}, \dots, a_{m1}), \dots, (a_{1n}, \dots, a_{mn})\}$.

To determine if S is linearly dependent,

consider linear combinations

$$x_1(a_{11}, \dots, a_{m1}) + \cdots + x_n(a_{1n}, \dots, a_{mn}) = (0, \dots, 0)$$

or

$$(a_{11}x_1 + \cdots + a_{1n}x_n, \dots, a_{m1}x_1 + \cdots + a_{mn}x_n) = (0, \dots, 0)$$

By equating components, we obtain

$$a_{11}x_1 + \cdots + a_{1n}x_n = 0$$

$$a_{21}x_1 + \cdots + a_{2n}x_n = 0$$

$$\vdots$$

$$a_{m1}x_1 + \cdots + a_{mn}x_n = 0$$

- If there are solutions other than $(x_1, \dots, x_n) = (0, \dots, 0)$, then those n-tuples of scalars yield linear dependencies among the vectors in S, and consequently S is linearly dependent

- If $(0, \dots, 0)$ is the only solution of the system, then S is linearly independent

- (1.5.1) Definitions. A system of m equations in n unknowns $x_1, \dots, x_n$ of the form

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

where $a_{ij}, b_i \in \mathbb{R}$, is called a system of linear equations.

  - The a_{ij} are called coefficients of the system
  - The system is homogeneous if all the $b_i = 0$ and inhomogeneous otherwise
  - A solution (vector) of the above system is a vector $(x_1, \dots, x_n) \in \mathbb{R}^n$ whose components solve all the equations in

the system

- A homogeneous system always has the trivial solution $(0, \dots, 0)$

- (1.5.2) Definition Two systems of linear equations are equivalent if their sets of solutions are the same

- (1.5.3) Proposition. Consider the above system of linear equations:

  a) The system obtained by adding any multiple of any one equation to any second equation, while leaving the other equations unchanged, is an equivalent system.

  b) The system obtained by multiplying any one equation by a nonzero scalar and leaving the other equations unchanged is an equivalent system

  c) The system obtained by interchanging any two equations is an equivalent system

- The above are elementary operations

- (1.5.6) Definition. A system of linear equations as in (1.5.1) is in echelon form if:

  1) In each equation, all the coefficients are 0, or the first nonzero coefficient counting from the left of the equation is a 1. The corresponding term in the equation is called the leading term.

     - In the ith equation, we call the subscript of the variable in the leading term $j(i)$

  2) For each i, the coefficient of $x_{j(i)}$ is zero in every equation other than the ith

  3) For each i (for which i and $i+1$ have some nonzero coefficients), $j(i+1) > j(i)$

- (1.5.8) Theorem. Every system of linear equations is equivalent to a system in echelon form. This echelon form may be found by applying a sequence of elementary operations to the original system

- The echelon form system whose existence is guaranteed by this theorem is unique

- The proof to the above theorem may be elaborated into an algorithm for producing the echelon form the system

1. Begin w/ the entire system of m equations
2. Do the following steps for each i between 1 and m
   a) Among the equations numbered i through m in the "updated" system of equations, pick one equation containing the variable w/ the smallest index among the variables appearing w/ nonzero coefficients in those equations
   b) If necessary, apply a row swap to interchange the equation chosen in step 2a and the ith equation
   c) If necessary, multiply by a constant to make

the leading coefficient in the new (ith) equation equal to 1

d) If necessary, add by another equation to eliminate all other occurrances of the leading variable in the ith equation

3. The process terminates when we complete step 2d w/ $i = m$

- The above process is the elimination algorithm
- Echelon form allows us to write down the set of solutions immediately
- If a system contains an equation of the form $0 = c$ for some $c \neq 0$ (i.e. $0 = 1$), then there are no solutions, and the system is inconsistent
  - If there are no equations of this form, then the system is consistent, and solutions do exist
- (1.5.11) Definitions.

  a) In an echelon form system the variables appearing in leading terms of equations are called basic variables of the system

b) All the other variables are called free variables

- We can express each of the basic variables in terms of the free variables

- Ex. Consider the echelon form system

$$
\begin{aligned}
x_1 \quad + x_3 \qquad\qquad &= 0 \\
x_2 - x_3 \qquad + x_5 &= 0 \\
x_4 + x_5 &= 0
\end{aligned}
$$

- Basic variables: x_1, x_2, x_4

- Free variables: x_3, x_5

- We have $\begin{cases} x_1 = x_3 \\ x_2 = x_3 - x_5 \\ x_4 = \quad - x_5 \end{cases}$

- Set $x_3 = 1$, $x_5 = 0$, we obtain one solution $(-1, 1, 1, 0, 0)$; set $x_3 = 0$, $x_5 = 1$, we obtain a second solution $(0, -1, 0, -1, 1)$

- The general solution of the system, obtained by setting $x_3 = t_1$, $x_5 = t_2$ in $\mathbb{R}$, is a general linear combination of these two vectors

- Set of all solutions: $W = \{t_1(-1,1,1,0,0) + t_2(0,-1,0,-1,1) \mid t_1, t_2 \in \mathbb{R}\}$
    - This is a parametrization of the set of solutions
    - We defined a mapping $F: \mathbb{R}^2 \to \mathbb{R}^5$ whose image is the set of solutions
        - $F(t_1, t_2) = t_1(-1,1,1,0,0) + t_2(0,-1,0,-1,1) = (-t_1, t_1 - t_2, t_1, -t_2, t_2)$
- For any echelon form system, we can obtain a parametrization of the set of solutions
    - If k is the number of free variables, the set of solutions is the image of a mapping from $\mathbb{R}^k \to \mathbb{R}^n$ (where n is the total number of variables in the original system)
- (1.5.13) Corollary. If $m < n$, every homogeneous

system of m linear equations in n unknowns

has a nontrivial solution

1.6 Bases and Dimension

- (1.6.1) Definition. A subset S of a vector space V is called a basis of V if $V = \text{span}(S)$ and S is linearly independent
- The standard basis $S = \{\vec{e}_1, \dots, \vec{e}_n\} \in \mathbb{R}^n$ is a basis for $\mathbb{R}^n$
    - It is linearly independent
    - Every vector $(a_1, \dots, a_n) \in \mathbb{R}^n$ may be written as the linear combination

      $$(a_1, \dots, a_n) = a_1\vec{e}_1 + \dots + a_n\vec{e}_n$$
- The vector space $\mathbb{R}^2$ has many other bases
- Let $V = P_n(\mathbb{R})$ and $S = \{1, x, x^2, \dots, x^n\}$
    - S is a linearly independent set of functions
    - S spans V

  Hence S is a basis for $P_n(\mathbb{R})$
- The empty subset $\emptyset$ is a basis of the vector space consisting only of a zero vector $\{\vec{0}\}$

- Given a basis S of V and a vector $x \in V$, there is only one way to produce $\vec{x}$ as a linear combination of the vectors in S
- (1.6.3) Theorem. Let V be a vector space, and let S be a nonempty subset of V. Then S is a basis of V iff every vector $x \in V$ may be written uniquely as a linear combination of the vectors in S
  - The scalars appearing in the linear combination may be thought as the coordinates of vector $\vec{x}$ w.r.t. the basis S
- Consider the basis $S' = \{(1,2), (1,-1)\}$ for $\mathbb{R}^2$. If we have any vector $(b_1, b_2) \in \mathbb{R}^2$, then $(b_1, b_2) = ([1/3 b_1 + 1/3 b_2](1,2) + [2/3 b_1 - 1/3 b_2](1,-1))$
  - W.r.t. basis S', our vector is described by the vector of coordinates $(1/3 b_1 + 1/3 b_2, 2/3 b_1 - 1/3 b_2)$
- $\begin{cases} \text{finite set: } P_n(\mathbb{R}) = \{1, x, x^2, \dots, x^n\} \\ \text{infinite set: } P(\mathbb{R}) = \{1, x, x^2, \dots, x^n, \dots\} \end{cases}$

- (1.6.6) Theorem. Let V be a vector space that has a finite spanning set, and let S be a linearly independent subset of V. Then there exists a basis S' of V, with $S \subseteq S'$
    - i.e. every linearly independent set may be extended to a basis
- (1.6.8) Lemma. Let S be a linearly independent subset of V and let $\vec{x} \in V$, $\vec{x} \notin S$. Then $S \cup \{\vec{x}\}$ is linearly independent iff $\vec{x} \notin \text{span}(S)$
- Consider the subspace $W \subset \mathbb{R}^5$. Let $S = \{(-1,1,1,0,0), (0,-1,0,-1,1)\}$ span W.
    - $\mathbb{R}^5$ has a basis w/ 5 vectors
    - W is not the entirety of $\mathbb{R}^5$ since it has a basis w/ 2 vectors
    - Hence $W \subseteq \mathbb{R}^5$, $W \neq \mathbb{R}^5$
- Every basis of a given vector space has the same number of elements

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

- (1.6.11) Corollary. Let V be a vector space, let S and S' be two bases of V, w/ m and m' elements, respectively. Then $m = m'$.

- (1.6.12) Definitions.

  a) If V is a vector space w/ some finite basis (possibly empty), we say V is finite-dimensional

  b) Let V be a finite-dimensional vector space. The dimension of V, denoted $\dim(V)$, is the number of vectors in any basis of V.

  c) If $V = \{\vec{0}\}$, we define $\dim(V) = 0$

- For each n, $\dim(\mathbb{R}^n) = n$ since the standard basis $\{\vec{e}_1, \dots, \vec{e}_n\}$ contains n vectors

- $\dim(P_n(\mathbb{R})) = n+1$, since $\{1, x, x^2, \dots, x^n\}$ contains $n+1$ functions

- The vector spaces $P(\mathbb{R})$, $C'(\mathbb{R})$, $C(\mathbb{R})$ are not finite-dimensional
  - They are infinite-dimensional
- (1.6.14) Corollary. Let W be a subspace of a finite-dimensional vector space V. Then $\dim(W) \leq \dim(V)$. Furthermore, $\dim(W) = \dim(V)$ iff $W = V$
- (1.6.15) Corollary. Let W be a subspace of $\mathbb{R}^n$ defined by a system of homogeneous linear equations. Then $\dim(W)$ is equal to the number of free variables in the corresponding echelon form system.
- Ex. In $\mathbb{R}^5$, consider the subspace defined by the system

$$\begin{aligned} x_1 + 2x_2 + x_4 &= 0 \\ x_1 + x_3 + x_5 &= 0 \\ x_2 + x_3 + x_4 - x_5 &= 0 \end{aligned}$$

After reducing to row-echelon form,

$$\begin{aligned} x_1 - \tfrac{1}{3}x_4 + \tfrac{4}{3}x_5 &= 0 \\ x_2 + \tfrac{2}{3}x_4 - \tfrac{2}{3}x_5 &= 0 \\ x_3 + \tfrac{1}{3}x_4 - \tfrac{1}{3}x_5 &= 0 \end{aligned}$$

The free variables are x_4 and x_5, hence $\dim(W)=2$, and a basis for W may be obtained by setting $x_4=1$, $x_5=0$ to obtain $(1/3, -2/3, -1/3, 1, 0)$, then $x_4=0$, $x_5=1$ to obtain $(-4/3, 2/3, 1/3, 0, 1)$

- $\dim(V_1+V_2) \leq \dim(V_1)+\dim(V_2)$
  - the overlap decreases the dimension of the sum
- (1.6.18) Theorem. Let W_1, W_2 be finite-dimensional subspaces of a vector space V. Then
$\dim(W_1+W_2) = \dim(W_1)+\dim(W_2)-\dim(W_1 \cap W_2)$

2.1 Linear Transformations

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

  i) $T(\vec{u}+\vec{v}) = T(\vec{u}) + T(\vec{v})$ for all $\vec{u}, \vec{v} \in V$

  ii) $T(\alpha\vec{v}) = \alpha T(\vec{v})$ for all $\alpha \in \mathbb{R}$, $\vec{v} \in V$

  V is called the domain of T and W is called the target of T

  - Two different operations of addition are used in condition (i)
    - $\vec{u} + \vec{v}$ takes place in vector space V
    - $T(\vec{u}) + T(\vec{v})$ takes place in vector space W
  - Two different operations of scalar multiplication
  - A linear mapping always takes the zero vector in the domain vector space to the zero vector in the target vector space
    - i.e. $T(\vec{0}_V) = \vec{0}_W$

- (2.1.2) Proposition. A function $T: V \to W$ is a linear transformation iff for all $a, b \in \mathbb{R}$ and all $\vec{u}, \vec{v} \in V$, $T(a\vec{u} + b\vec{v}) = aT(\vec{u}) + bT(\vec{v})$
- (2.1.3) Corollary. A function $T: V \to W$ is a linear transformation iff for all $a_1, \dots, a_k \in \mathbb{R}$ and for all $\vec{v}_1, \dots, \vec{v}_k \in V$,

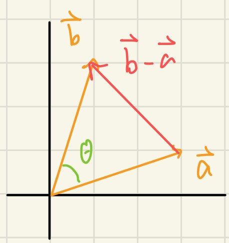
$$T\left(\sum_{i=1}^{k} a_i(\vec{v}_i)\right) = \sum_{i=1}^{k} a_i T(\vec{v}_i)$$

- Let V be any vector space. Let $W = V$. The identity transformation $I: V \to V$ is defined by $I(\vec{v}) = \vec{v}$ for all $\vec{v} \in V$
  - $I(a\vec{u} + b\vec{v}) = a\vec{u} + b\vec{v} = aI(\vec{u}) + bI(\vec{v})$
- Let V, W be any vector spaces. Let $T: V \to W$ be a mapping that takes every vector in V to the zero vector in W, i.e. $T(\vec{v}) = \vec{0}_W$
  - T is called the zero transformation
  - $aT(\vec{u}) + bT(\vec{v}) = a\vec{0}_W + b\vec{0}_W = \vec{0}_W$
- Consider the operation of differentiation. Let V be the

vector space $C^\infty(\mathbb{R})$ (of functions $f: \mathbb{R} \to \mathbb{R}$ w/ derivatives of all orders). Let $D: C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ be the mapping that takes each function $f \in C^\infty(\mathbb{R})$ to its derivative function: $D(f) = f' \in C^\infty(\mathbb{R})$

- We can show that D is a linear transformation

- Definite integration may also be viewed as linear mapping
  - Let V denote a vector space $C[a,b]$ of continuous functions on the closed interval $[a,b] \subset \mathbb{R}$, and let $W = \mathbb{R}$. Then we can define the integration mapping $Int: V \to W$ by the rule $Int(f) = \int_a^b f(x)\,dx \in \mathbb{R}$

- Geometry / trigonometry review
  - 

    $$\begin{cases} \vec{a} = (a_1, a_2) \\ \vec{b} = (b_1, b_2) \end{cases}$$

  - $\|\vec{a}\| = \sqrt{a_1^2 + a_2^2}$
  - $\|\vec{b}\| = \sqrt{b_1^2 + b_2^2}$
  - $\|\vec{b} - \vec{a}\| = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}$

- By law of cosines:

$$\|\vec{b}-\vec{a}\|^2 = \|\vec{a}\|^2 + \|\vec{b}\|^2 - 2\|\vec{a}\| \cdot \|\vec{b}\| \cdot \cos\theta$$

Rewriting the above equation,

$$a_1 b_1 + a_2 b_2 = \|\vec{a}\| \cdot \|\vec{b}\| \cdot \cos\theta$$

- The quantity $a_1 b_1 + a_2 b_2$ is called the inner product (or dot product) of the vectors $\vec{a}$ and $\vec{b}$ and is denoted $\langle \vec{a}, \vec{b} \rangle$

- (2.1.9) Proposition. If $\vec{a}$, $\vec{b}$ are nonzero vectors in $\mathbb{R}^2$, the angle θ between $\vec{a}$ and $\vec{b}$ satisfies

$$\cos\theta = \frac{\langle a, b \rangle}{\|\vec{a}\| \cdot \|\vec{b}\|}$$

- (2.1.10) Corollary. If $\vec{a} \neq \vec{0}$, $\vec{b} \neq \vec{0}$ are vectors in $\mathbb{R}^2$, the angle θ between them is a right angle iff $\langle \vec{a}, \vec{b} \rangle = 0$

- The length of a vector can be expressed in terms of the inner product by $\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v} \rangle}$

- Every vector $\vec{v}$ in the plane may be expressed in terms of the angle φ it makes w/ the first coordinate axis.

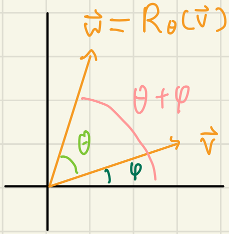
Since $\cos(\varphi) = v_1 / \|\vec{v}\|$ and $\sin(\varphi) = v_2/\|\vec{v}\|$ by trig identity, we have $\vec{v} = \|\vec{v}\| (\cos(\varphi), \sin(\varphi))$

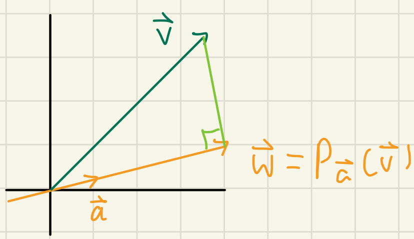
- Let $V = W = \mathbb{R}^2$. Let θ be a fixed real number that represents an angle (in radians). Define a function $R_\theta : V \to V$ by $R_\theta(\vec{v}) =$ the vector obtained by rotating the vector $\vec{v}$ through an angle θ while preserving its length
  - If $\vec{w} = R_\theta(\vec{v})$, then

    $\vec{w} = \|\vec{v}\| (\cos(\varphi + \theta), \sin(\varphi + \theta))$ where φ is the angle $\vec{v}$ makes w/ the first coordinate axis
  - Using the trig formulas,

    $$\vec{w} = \|\vec{v}\| (\cos(\varphi) \cdot \cos(\theta) - \sin(\varphi) \cdot \sin\theta, \\ \cos(\varphi) \cdot \sin(\theta) + \sin(\varphi) \cdot \cos(\theta))$$

    $$= (v_1 \cos\theta - v_2 \sin\theta, v_1 \sin\theta + v_2 \cos\theta)$$
  - We can easily check that R_θ is a linear transformation

- 

- Let $V = W = \mathbb{R}^2$. Let $\vec{a} = (a_1, a_2)$ be a nonzero vector in V. Let L denote the line spanned by the vector $\vec{a}$. The projection to L, denoted as $P_{\vec{a}}$ is defined as follows: given $\vec{v} \in \mathbb{R}^2$, construct the line perpendicular to L passing through $\vec{v}$. This line intersects L in a point $\vec{w}$. $P_{\vec{a}}(\vec{v})$ is defined to be this vector $\vec{w}$.
  - $P_{\vec{a}}(\vec{v}) = \vec{w}$ is a multiple of the vector $\vec{a}$ since $\vec{w} \in L$
  - $\vec{w} = c\vec{a}$ for some scalar c
  - $\vec{v} - P_{\vec{a}}(\vec{v}) = \vec{v} - c\vec{a}$ must be perpendicular to the vector $\vec{a}$
  - 

- $P_{\vec{a}}(\vec{v})$ must satisfy $\langle \vec{a}, (v - c\vec{a}) \rangle = 0$
- $\langle \vec{a}, \vec{v} \rangle = c \cdot \langle \vec{a}, \vec{a} \rangle$, and so

  $c = \langle \vec{a}, \vec{v} \rangle / \langle \vec{a}, \vec{a} \rangle$
  - Since $\vec{a} \neq \vec{0}$, $\langle \vec{a}, \vec{a} \rangle \neq 0$
- $$P_{\vec{a}}(\vec{v}) = c\vec{a} = \frac{\langle \vec{a}, \vec{v} \rangle}{\langle \vec{a}, \vec{a} \rangle} \vec{a}$$
- $P_{\vec{a}}$ is a linear transformation

- Let $T: V \to W$ be a linear transformation. If V is finite-dimensional, then we may choose a basis $\{\vec{v}_1, \ldots, \vec{v}_k\}$ for V. For each $\vec{v} \in V$, there is a unique choice of scalars $a_1, \ldots, a_k$ so that $\vec{v} = a_1\vec{v}_1 + \cdots + a_k\vec{v}_k$. Then

$$\begin{aligned} T(\vec{v}) &= T(a_1\vec{v}_1 + \cdots + a_k\vec{v}_k) \\ &= T(a_1\vec{v}_1) + \cdots + T(a_k\vec{v}_k) \\ &= a_1 T(\vec{v}_1) + \cdots + a_k T(\vec{v}_k) \end{aligned}$$

Therefore, if the values of T on the members of a basis for V are known, then all the values of T are known

- (2.1.14) Proposition. If $T: V \to W$ is a linear transformation and V is a finite-dimensional, then T is uniquely determined by its values on the members of a basis V.
- If $\{\vec{v}_1, \dots, \vec{v}_k\}$ is a basis for V and $\{\vec{w}_1, \dots, \vec{w}_k\}$ are any k vectors for W, then we can define a linear transformation $T: V \to W$ by insisting that $T(\vec{v}_i) = \vec{w}_i$, $i = 1, \dots, k$ and that T satisfies the linearity condition.
    - For an arbitrary $\vec{v} \in V$ expressed as
      $\vec{v} = a_1 \vec{v}_1 + \cdots + a_k \vec{v}_k$,
      $T(\vec{v}) = a_1 T(\vec{v}_1) + \cdots + a_k T(\vec{v}_k)$
      $= a_1 \vec{w}_1 + \cdots + a_k \vec{w}_k$ defines T
- Let $V = W = P_2(\mathbb{R})$. A basis for V is given by the polynomials $1, 1+x, 1+x+x^2$. Define T on this basis by $T(1) = x$, $T(1+x) = x^2$, $T(1+x+x^2) = 1$. If we insist that T be linear, this defines a linear transformation. If $p(x) = a_2 x^2 + a_1 x + a_0$, then

the equation

$$p(x) = (a_0 - a_1)1 + (a_1 - a_2)(1+x) + a_2(1+x+x^2)$$

expresses $p(x)$ in terms of the basis. Calculating,

$$T(p(x)) = (a_1 - a_2)x^2 + (a_0 - a_1)x + a_2$$

2.2 Finite-Dimensional Vector Spaces

- In a finite-dimensional vector space, every vector can be expressed uniquely as a linear combination of the vectors in a finite set

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

- Ex. Let $V = W = \mathbb{R}^2$. Define T by $T(\vec{e}_1) = \vec{e}_1 + \vec{e}_2$ and $T(\vec{e}_2) = 2\vec{e}_1 - 2\vec{e}_2$.
  - The four scalars $a_{11} = 1$, $a_{21} = 1$, $a_{12} = 2$, $a_{22} = -2$ determine T.

- Ex. Let $V = W = P_2(\mathbb{R})$. Choose the basis $p_0(x) = 1$, $p_1(x) = x$, $p_2(x) = x^2$ for both V, W. Define T by

$T(p(x)) = x \frac{d}{dx} p(x)$. Then $T(p_0(x)) = 0$, $T(p_1(x)) = p_1(x)$, $T(p_2(x)) = 2p_2(x)$. Thus, T is determined by the scalars a_{ij}, $0 \leq i,j \leq 2$, w/ $a_{ij} = 0$ if $i \neq j$ and $a_{00} = 0$, $a_{11} = 1$, $a_{22} = 2$

- (2.2.5) Definition. Let a_{ij}, $1 \leq i \leq l$, $1 \leq j \leq k$, be $l \cdot k$ scalars. The matrix whose entries are the scalars a_{ij} is the rectangular array of l rows and k columns:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1k} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2k} \\ & \vdots & & & \\ a_{l1} & a_{l2} & a_{l3} & \cdots & a_{lk} \end{bmatrix}$$

  - The scalar a_{ij} is the entry in the ith row and the jth column of the array
  - A matrix w/ l rows and k columns is an $l \times k$ matrix ("l by k matrix")

- A transformation $T: V \to W$ is determined by the choice of bases in V, W and a set of $l \cdot k$

scalars, where $k = \dim(V)$, $\ell = \dim(W)$

- We can write those scalars in the form of a matrix

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

  - If $T(\vec{v}_j) = a_{1j}\vec{w}_1 + \cdots + a_{\ell j}\vec{w}_\ell$, then the coefficients expressing $T(\vec{v}_j)$ in terms of the $\vec{w}_1, \dots, \vec{w}_\ell$ form the jth column of $[T]_\alpha^\beta$

  - In forming the matrix of a linear transformation, we always assume that the basis vectors in the domain and target spaces are written in some particular order

- Let $T: V \to V$ be the identity transformaton of a finite-dimensional vector space to itself, $T = I$, then w.r.t. any choice of basis α for V, the matrix of I is the $k \times k$ matrix w/ 1 in each diagonal position and 0 in each off-diagonal position, i.e.

$$\begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix}$$

  - We can omit the 0s for convenience

- Ex (the matrix of rotation) Let $V = W = \mathbb{R}^2$. Take both bases α, β to be the standard basis. Let $T = R_\theta$ be rotation through an angle θ in the plane. Then for an arbitrary vector $\vec{v} = (v_1, v_2)$,

$R_\theta(\vec{v}) = (v_1 \cos\theta - v_2 \sin\theta, v_1 \sin\theta + v_2 \cos\theta)$

  - $R_\theta(\vec{e}_1) = (\cos\theta, \sin\theta) = \cos\theta \cdot \vec{e}_1 + \sin\theta \cdot \vec{e}_2$

- $R_\theta(\vec{e}_2) = -\sin\theta \cdot \vec{e}_1 + \cos\theta \cdot \vec{e}_2$

- The matrix of R_θ is

$$[R_\theta]_\alpha^\alpha = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{where}$$

α is the standard basis in $\mathbb{R}^2$

- Ex. (the matrix of projection in the plane) Let $V = W = \mathbb{R}^2$ and both bases α, β are taken to be the standard basis. Let $\vec{a} = (a_1, a_2)$ be a fixed nonzero vector. Let $T = P_{\vec{a}}$ be the projection to the line spanned by the vector $\vec{a}$. Then for any arbitrary vector $\vec{v} = (v_1, v_2)$,

$$P_{\vec{a}}(\vec{v}) = \|\vec{a}\|^{-2} (a_1^2 v_1 + a_1 a_2 v_2 , a_1 a_2 v_1 + a_2^2 v_2)$$

(derived from $\dfrac{\langle \vec{a}, \vec{v} \rangle}{\langle \vec{a}, \vec{a} \rangle} \vec{a}$)

- $P_{\vec{a}}(\vec{e}_1) = \|\vec{a}\|^{-2} a_1^2 \vec{e}_1 + \|\vec{a}\|^{-2} a_1 a_2 \vec{e}_2$

- $P_{\vec{a}}(\vec{e}_2) = \|\vec{a}\|^{-2} a_1 a_2 \vec{e}_1 + \|\vec{a}\|^{-2} a_2^2 \vec{e}_2$

- The matrix of $P_{\vec{a}}$ w.r.t. the standard basis is

$$[P_{\vec{a}}]_\alpha^\alpha = \|\vec{a}\|^{-2} \begin{bmatrix} a_1^2 & a_1 a_2 \\ a_1 a_2 & a_2^2 \end{bmatrix}$$

- The scalar $\|\vec{a}\|^{-2}$ multiplies each entry of the matrix

- If $\vec{v} = a_1\vec{v}_1 + \cdots + a_k\vec{v}_k$ and $\vec{w} = b_1\vec{w}_1 + \cdots + b_\ell\vec{w}_\ell$, we can express $\vec{v}, \vec{w}$ in coordinates, respectively, as a $k \times 1$ matrix and as an $\ell \times 1$ matrix, i.e. as column vectors
  - We call these coordinate vectors of the given vectors w.r.t. the chosen bases
  - These coordinate vectors are denoted by $[\vec{v}]_\alpha$, $[\vec{w}]_\beta$, respectively
  - $$[\vec{v}]_\alpha = \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}, \quad [\vec{w}]_\beta = \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix}$$

- (2.2.10) Definition. Let A be an $\ell \times k$ matrix, let $\vec{x}$ be a column vector w/ k entries, then the product of the vector $\vec{x}$ by the matrix A is defined to be the column vector w/ ℓ entries:

$$\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1k}x_k \\ \vdots \\ a_{\ell 1}x_1 + a_{\ell 2}x_2 + \cdots + a_{\ell k}x_k \end{bmatrix}$$ and

is denoted $A\vec{x}$

- If we write out the entire matrix A and vector $\vec{x}$, this becomes

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1k} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2k} \\ & & \vdots & & \\ a_{\ell 1} & a_{\ell 2} & a_{\ell 3} & \cdots & a_{\ell k} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}$$

- The ith entry of the product $A\vec{x}$, $a_{i1}x_1 + \cdots + a_{ik}x_k$, is the product of the ith row of A, considered as a $1 \times k$ matrix, w/ the column vector $\vec{x}$

- The product of a $1 \times k$ matrix (i.e. a row vector) and a column vector generalizes the notion of the dot product in the plane

- i.e. $[x_1, x_2]\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = x_1y_1 + x_2y_2$

$= \langle \vec{x}, \vec{y} \rangle$

- If the number of columns of the matrix A is not equal to the number of entries in the column vector $\vec{x}$, matrix multiplication $A\vec{x}$ is not defined.

- To evaluate $T(\vec{v})$:

  1. Compute the matrix of T w.r.t. given bases for V and W

  2. Express the given vector $\vec{v}$ in terms of the basis for V

  3. Multiply the coordinate vector of $\vec{v}$ by the matrix of T to obtain a second coordinate vector

- (2.2.15) Proposition. Let $T: V \to W$ be a linear transformation between vector spaces V of dimension k and W of dimension l. Let $\alpha = \{\vec{v}_1, \ldots, \vec{v}_k\}$ be a basis for V and $\beta = \{\vec{w}_1, \dots, \vec{w}_\ell\}$ be a basis for W. Then for each $\vec{v} \in V$,

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

- If $\vec{v}_j$ is the jth member of the basis α of V, the coordinate vector $\vec{v}_j$ w.r.t. the basis α is $(0, \dots, 1, \dots, 0)$, i.e. 1 in the jth position and 0 in the remaining positions
- The coordinate vector of $T(\vec{v}_j)$, i.e. $[T(\vec{v}_j)]_\beta$, is $[T]_\alpha^\beta [\vec{v}_j]_\alpha =$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1k} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2k} \\ & & \vdots & & \\ a_{\ell 1} & a_{\ell 2} & a_{\ell 3} & \cdots & a_{\ell k} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{\ell j} \end{bmatrix}$$

which is the jth column of the matrix $[T]_\alpha^\beta$

- (2.2.18) Proposition. Let A be an $\ell \times k$ matrix and $\vec{u}, \vec{v}$ be column vectors w/ k entries. Then for every pair of real numbers a, b,

$$A(a\vec{u} + b\vec{v}) = aA\vec{u} + bA\vec{v}$$

- We can use a matrix to define a linear transformation
- Let A be a fixed $\ell \times k$ matrix w/ entries a_{ij}, $1 \le i \le \ell$, $1 \le j \le k$. Let $\alpha = \{\vec{v}_1, \ldots, \vec{v}_k\}$ be a basis for V and $\beta = \{\vec{w}_1, \ldots, \vec{w}_k\}$ be a basis for W. We can define a function $T: V \to W$ as follows:
  - For each vector $\vec{v} = x_1\vec{v}_1 + \cdots + x_k\vec{v}_k$, define $T(\vec{v})$ to be the vector in W whose coordinate vector in the β coordinates is $A[\vec{v}]_\alpha$, i.e. $[T(\vec{v})]_\beta = A[\vec{v}]_\alpha$ so that

$$T(\vec{v}) = \sum_{i=1}^{\ell} \left(\sum_{j=1}^{k} x_j a_{ij} \right) \vec{w}_i$$

- (2.2.19) Proposition. Let $\alpha = \{\vec{v}_1, \ldots, \vec{v}_k\}$ be a basis for V and $\beta = \{\vec{w}_1, \ldots, \vec{w}_\ell\}$ be a basis for W, and let $\vec{v} = x_1\vec{v}_1 + \cdots + x_k\vec{v}_k \in V$.
  i) If A is an $\ell \times k$ matrix, then the function $T(\vec{v}) = \vec{w}$ where $[\vec{w}]_\beta = A[\vec{v}]_\alpha$ is a linear transformation.

ii) If $A = [S]_\alpha^\beta$ is the matrix of a transformation $S: V \to W$, then the transformation T constructed from $[S]_\alpha^\beta$ is equal to S.

iii) If T is the transformation of i) constructed from A, then $[T]_\alpha^\beta = A$

- (2.2.20) Proposition. Let V, W be finite-dimensional vector spaces. Let α be a basis for V and β a basis for W. Then the assignment of a matrix to a linear transformation from V to W given by T goes to $[T]_\alpha^\beta$ is injective and surjective.
    - This assignment of a matrix to a transformation depends on the choice of the bases α and β.
        - i.e. $[T]_\alpha^\beta \neq [T]_{\alpha'}^{\beta'}$

2.3 Kernel and Image

- (2.3.1) Definition. The kernel of T, denoted $\mathrm{Ker}(T)$, is the subset of V consisting of all vectors $\vec{v} \in V$ s.t. $T(\vec{v}) = \vec{0}$
- (2.3.2) Proposition. Let $T: V \to W$ be a linear transformation. $\mathrm{Ker}(T)$ is a subspace of V.
- In general, let $V = \mathbb{R}^n$ and $W = \mathbb{R}$, and let $\vec{a} = (a_1, \dots, a_n)$ be a fixed vector in $\mathbb{R}^n$. If T is defined by $T(\vec{x}) = a_1x_1 + \cdots + a_nx_n$, then $\mathrm{Ker}(T)$ is equal to the subspace of $V = \mathbb{R}^n$ defined by the single linear equation $a_1x_1 + \cdots + a_nx_n = 0$
- Ex. Let $V = P_3(\mathbb{R})$. Define $T: V \to V$ by $T(p(x)) = \frac{d}{dx}p(x)$.
  - The only polynomials w/ zero derivatives are the constant polynomials
  - $\mathrm{Ker}(T)$ is the subspace of V

consisting of the constant polynomials

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

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

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

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

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

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

- (2.3.12) Proposition. If $\{\vec{v}_1, \dots, \vec{v}_m\}$ is any set that spans V (e.g. a basis of V), then $\{T(\vec{v}_1), \dots, T(\vec{v}_m)\}$ spans $Im(T)$.

- (2.3.13) Corollary. If $\alpha = \{\vec{v}_1, \dots, \vec{v}_k\}$ is a basis for V and $\beta = \{\vec{w}_1, \dots, \vec{w}_\ell\}$ is a basis for W, then the vectors in W whose coordinate vectors (in terms of β) are the columns of $[T]_\alpha^\beta$ span $Im(T)$.
  - These vectors are the vectors $T(\vec{v}_j)$, $1 \leq j \leq k$

- We can construct a basis for $Im(T)$

- Procedure 1: Those columns of $[T]_\alpha^\beta$ that correspond to basic variables of the system of equations $[T]_\alpha^\beta [\vec{x}]_\alpha = \vec{0}$ are linearly independent
- Procedure 2: Choose a basis $\{\vec{v}_1, \dots, \vec{v}_k\}$ for V so that $\vec{v}_1, \dots, \vec{v}_q$ (w/ $q \leq k$) is a basis of $Ker(T)$. Then $T(\vec{v}_1) = \cdots = T(\vec{v}_q) = \vec{0}$. We claim that $\{T(\vec{v}_{q+1}), \dots, T(\vec{v}_k)\}$ is a basis for $Im(T)$.
- The maximum number of linearly independent columns of a matrix A is the rank of A.
  - The rank of $[T]_\alpha^\beta$ is the dimension of the image of T
- (2.3.17) Theorem. If V is a finite-dimensional vector space and $T: V \to W$ is a linear transformation, then $\dim(Ker(T)) + \dim(Im(T)) = \dim(V)$

2.4 Applications of the Dimension Theorem

- A function between sets $f: S_1 \to S_2$ is injective if whenever $f(p_1) = f(p_2)$ for $p_1, p_2 \in S$, we have $p_1 = p_2$
- The function f is surjective if for each $q \in S_2$ there is some $p \in S_1$ w/ $f(p) = q$
- Let $V = W = \mathbb{R}^2$. Let $T = R_\theta$, rotation through an angle θ
  - T is injective.

    If $\vec{w} = R_\theta(\vec{u}) = R_\theta(\vec{v})$, then $\vec{w}$ was obtained by rotating both $\vec{u}$ and $\vec{v}$ through an angle θ.

    Therefore, rotation of $\vec{w}$ through an angle $-\theta$ yields both $\vec{u}$ and $\vec{v}$. Thus $\vec{u} = \vec{v}$.
  - If $\vec{v}$ is obtained from rotating $\vec{w}$ by $-\theta$, then $R(\vec{v}) = \vec{w}$. This holds for all $\vec{w}$, so T is surjective.
  - To prove that R_θ is injective, show that there is at most 1 solution to the

system of equations $[R_\theta]_\alpha^\alpha [\vec{u}]_\alpha = [\vec{w}]_\alpha$ for every choice of $\vec{w}$, where α is the standard basis for $\mathbb{R}^2$

- To prove that R_θ is surjective, show that there is at least 1 solution to the same system of equations for every choice of $\vec{w}$

- (2.4.2) Proposition. A linear transformation $T: V \to W$ is injective iff $\dim(\ker(T)) = 0$
    - Injective - only 1 vector mapped to $\vec{0}$
- (2.4.3) Corollary. A linear mapping $T: V \to W$ on a finite-dimensional vector space V is injective iff $\dim(\operatorname{Im}(T)) = \dim(V)$
- (2.4.4) Corollary. If $\dim(W) < \dim(V)$ and $T: V \to W$ is a linear mapping, then T is not injective.
- (2.4.5) Corollary. If V and W are finite-dimensional, then a linear mapping $T: V \to W$ can be injective only if $\dim(W) \geq \dim(V)$
- (2.4.7) Proposition. If W is finite-dimensional, then a

linear mapping $T: V \to W$ is surjective iff $\dim(\mathrm{Im}(T)) = \dim(W)$

- (2.4.8) Corollary. If V and W are finite-dimensional, w/ $\dim(V) < \dim(W)$, then there is no surjective linear mapping $T: V \to W$
- (2.4.9) Corollary. A linear mapping $T: V \to W$ can be surjective only if $\dim(V) \geq \dim(W)$
- (2.4.10) Proposition. Let $\dim(V) = \dim(W)$. A linear transformation $T: V \to W$ is injective iff it is surjective
- If $\dim(V) > \dim(W)$, a linear transformation $T: V \to W$ may be surjective, but cannot be injective.
- If $\dim(V) < \dim(W)$, a linear transformation $T: V \to W$ may be injective, but cannot be surjective.
- We can describe the set of all vectors $\vec{v}$ w/ $T(\vec{v}) = \vec{w}$ by the set $T^{-1}(\{\vec{w}\})$
    - $T^{-1}(\{\vec{w}\})$ is the inverse image of $\vec{w}$ under the transformation T

- We are not claiming that T has an inverse function
- $T^{-1}(\{\vec{w}\}) = \{\vec{v} \in V \mid T(\vec{v}) = \vec{w}\}$
- If $\vec{w} = \vec{0}$, then $T^{-1}(\{\vec{w}\}) = \ker(T)$
- If $\vec{w} \notin \operatorname{Im}(T)$, then $T^{-1}(\{\vec{w}\}) = \emptyset$

- If $\vec{v}_1, \vec{v}_2 \in T^{-1}(\{\vec{w}\})$, then $T(\vec{v}_1) = T(\vec{v}_2) = \vec{w}$, and so $T(\vec{v}_1) - T(\vec{v}_2) = \vec{0}$ and $T(\vec{v}_1 - \vec{v}_2) = \vec{0}$.

  Thus, $\vec{v}_1 - \vec{v}_2 \in \ker(T)$
- If $\vec{u}_1 \in T^{-1}(\{\vec{w}\})$, $\vec{u}_2 \in \ker(T)$, then $T(\vec{u}_1) = \vec{w}$, $T(\vec{u}_2) = \vec{0}$, and so $T(\vec{u}_1) + T(\vec{u}_2) = T(\vec{u}_1 + \vec{u}_2) = \vec{w}$.

  Thus $\vec{u}_1 + \vec{u}_2 \in T^{-1}(\{\vec{w}\})$
- (2.4.11) Proposition. Let $T: V \to W$ be a linear transformation, and let $\vec{w} \in \operatorname{Im}(T)$. Let $\vec{v}_1$ be any fixed vector w/ $T(\vec{v}_1) = \vec{w}$. Then every vector $\vec{v}_2 \in T^{-1}(\{\vec{w}\})$ can be written uniquely as $\vec{v}_2 = \vec{v}_1 + \vec{u}$, where $\vec{u} \in \ker(T)$.
  - If a different $\vec{v}_1$ is used, the corresponding $\vec{u}$

would be different too.

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

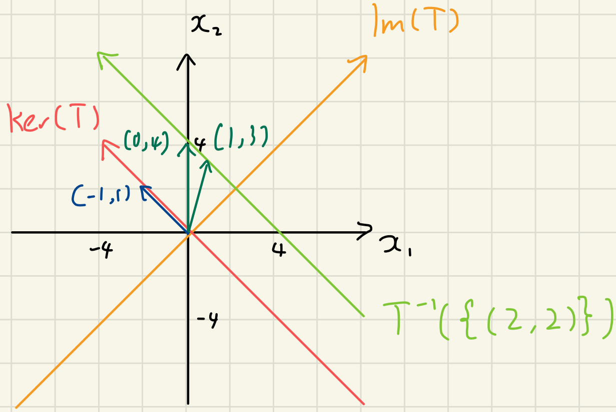
- Let $V = W = \mathbb{R}^2$, let $T = P_{\vec{a}}$ for $\vec{a} = (1,1)$.

Let $\vec{w} = (2,2)$.

$T((0,4)) = (2,2)$

$T((1,3)) = (2,2)$

$(0,4) - (1,3)$
$= (-1,1) \in \ker(T)$



- (2.4.15) Corollary. Let $T: V \to W$ be a linear transformation of finite-dimensional vector spaces. Let $\vec{w} \in W$. Then there is a unique vector $\vec{v} \in V$ s.t. $T(\vec{v}) = \vec{w}$ iff
  i. $\vec{w} \in \mathrm{Im}(T)$ and
  ii. $\dim(\ker(T)) = 0$

- Solving the system $A\vec{x} = \vec{b}$ is equivalent to finding the set of all vectors $\vec{x} \in T^{-1}(\{\vec{b}\})$

- (2.4.16) Proposition. With notation as before.

i. The set of solutions of the system of linear equations $A\vec{x}=\vec{b}$ is the subset $T^{-1}(\{\vec{b}\})$ of $V=\mathbb{R}^n$.

ii. The set of solutions of the system of linear equations $A\vec{x}=\vec{b}$ is a subspace of V iff the system is homogeneous, in which case the set of solutions is $\ker(T)$

- (2.4.17) Corollary.

  i. The number of free variables in the homogeneous system $A\vec{x}=\vec{0}$ (or its echelon form equivalent) is equal to $\dim(\ker(T))$

  ii. The number of basic variables of the system is equal to $\dim(\mathrm{Im}(T))$

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

- (2.4.19) Proposition. Let $\vec{x}_p$ be a particular solution of the system $A\vec{x} = \vec{b}$. Then every other solution to $A\vec{x} = \vec{b}$ is of the form $\vec{x} = \vec{x}_p + \vec{x}_n$, where $\vec{x}_n$ is a solution of the corresponding homogeneous system of equations $A\vec{x} = \vec{0}$.

Furthermore, given $\vec{x}_p$ and $\vec{x}$, there is a unique $\vec{x}_n$ s.t. $\vec{x} = \vec{x}_p + \vec{x}_n$

- When $\vec{b} \in \operatorname{Im}(T)$, the set of solutions to the system $A\vec{x} = \vec{b}$ consists of all vectors in $\ker(T)$ (a subspace of $V = \mathbb{R}^n$) translated by (any) particular solution $\vec{x}_p$
- Translation of a subspace $U \subset V$ by a vector $\vec{a}$ means the set

  $$\{\vec{v} \in V \mid \exists \vec{u} \in U \text{ s.t. } \vec{v} = \vec{a} + \vec{u}\} \subset V$$

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

2.5 Composition of Linear Transformations

- Let $S: U \to V$ and $T: V \to W$ be linear transformations. The composition of S and T is denoted $TS: U \to W$ defined by $TS(\vec{v}) = T(S(\vec{v}))$
  - Well defined since the image of S is contained in V, which is the domain of T
- (2.5.1) Proposition. Let $S: U \to V$ and $T: V \to W$ be linear transformations, then TS is a linear transformation.
  - Usually $ST \neq TS$
  - The composition is well-defined only if the image of the first transformation is contained in the domain of the second.
- (2.5.4) Proposition.
  i. Let $R: U \to V$, $S: V \to W$, $T: V \to X$ be linear transformations of the vector spaces U, V, W, X as indicated.

Then $T(SR) = (TS)R$ (associativity)

- We may write TSR and compute the composition in any order

ii. Let $R: U \to V$, $S: U \to V$, $T: V \to W$ be linear transformations of the vector spaces U, V, W as indicated.

Then $T(R+S) = TR + TS$ (distributivity)

iii. Let $R: U \to V$, $S: V \to W$, $T: V \to W$ be linear transformations of the vector spaces U, V, W as indicated.

Then $(T+S)R = TR + SR$ (distributivity)

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

i. $\mathrm{Ker}(S) \subset \mathrm{Ker}(TS)$

ii. $\mathrm{Im}(TS) \subset \mathrm{Im}(T)$

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

i. $\dim(\mathrm{Ker}(S)) \leq \dim(\mathrm{Ker}(TS))$

ii. $\dim(\mathrm{Im}(TS)) \leq \dim(\mathrm{Im}(T))$

- Let $U = V = W = \mathbb{R}^2$. Let $S = R_\theta$ and $T = P_{\vec{a}}$ for $\vec{a} \neq \vec{0}$.
  - $\dim(\mathrm{Ker}(R_\theta)) = 0$, $\dim(\mathrm{Ker}(P_{\vec{a}} R_\theta)) = 1$
  - $\dim(\mathrm{Im}(P_{\vec{a}})) = 1$, $\dim(\mathrm{Im}(P_{\vec{a}} R_\theta)) = 1$
  - $\dim(\mathrm{Ker}(P_{\vec{a}})) = 1$, $\dim(\mathrm{Ker}(R_\theta P_{\vec{a}})) = 1$
  - $\dim(\mathrm{Im}(R_\theta)) = 2$, $\dim(\mathrm{Im}(R_\theta P_{\vec{a}})) = 1$
- (2.5.9) Proposition. If $[S]_\alpha^\beta$ has entries a_{ij}, $i = 1, \dots, n$ and $j = 1, \dots, m$ and $[T]_\beta^\gamma$ has entries $b_{k\ell}$, $k = 1, \dots, p$ and $\ell = 1, \dots, n$.

  Then the entries of $[TS]_\alpha^\gamma$ are $\sum_{\ell=1}^{n} b_{k\ell} a_{\ell j}$.
- (2.5.10) Definition. Let A be an $n \times m$ matrix and B be a $p \times n$ matrix. Then the matrix product BA is defined to be the $p \times m$ whose entries are $\sum_{\ell=1}^{n} b_{k\ell} a_{\ell j}$ for $k = 1, \dots, p$ and $j = 1, \dots, m$.
  - The entry in the kth row and jth column of BA is the product of the kth row of B

w/ the jth column of A

$$\begin{array}{cc} & \\ \text{kth row} & \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{k1} & \cdots & b_{kn} \\ \vdots & & \vdots \\ b_{p1} & \cdots & b_{pn} \end{bmatrix} \end{array} \begin{array}{c} \text{jth column} \\ \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1m} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nm} \end{bmatrix} \end{array}$$

$$= \begin{array}{cc} \begin{bmatrix} & \vdots & \\ \cdots & \sum_{l=1}^{n} b_{kl} a_{lj} & \cdots \\ & \vdots & \end{bmatrix} & \text{kth row} \\ \text{jth column} & \end{array}$$

- The multiplication of matrices BA is not defined if the number of columns of B is not equal to the number of rows of A

- (2.5.13) Proposition. Let $S: U \to V$ and $T: V \to W$ be linear transformations between finite-dimensional vector spaces. Let α, β, γ be bases for U, V, W, respectively. Then $[TS]_\alpha^\gamma = [T]_\beta^\gamma [S]_\alpha^\beta$.
  - The matrix of the composition of two linear transformations is the product of

the matrices of the transformations.

- (2.5.14) Proposition.

    i. Let A, B, C be $m \times n$, $n \times p$, $p \times r$ matrices, respectively. Then $(AB)C = A(BC)$ (associativity)

    ii. Let A be an $m \times n$ matrix and B and C, $n \times p$ matrices. Then $A(B+C) = AB + AC$ (distributivity)

    iii. Let A and B be $m \times n$ matrices and let C be an $n \times p$ matrix. Then $(A+B)C = AC + BC$ (distributivity)

2.6 The Inverse of a Linear Transformation

- If $f: S_1 \to S_2$ is a function from one set to another, we say that $g: S_2 \to S_1$ is the inverse function of f if for every $x \in S_1$, $g(f(x)) = x$ and for every $y \in S_2$, $f(g(y)) = y$
  - If such a g exists, f must be both injective and surjective
- (2.6.1) Proposition. If $T: V \to W$ is injective and surjective, then the inverse function $S: W \to V$ is a linear transformation
- (2.6.2) Proposition. A linear transformation $T: V \to W$ has an inverse linear transformation S iff T is injective and surjective
- (2.6.3) Definition. If $T: V \to W$ is a linear transformation that has an inverse transformation $S: W \to V$, we say that T is invertible, and we denote the inverse of T by T^{-1}

- (2.6.4) Definition. If $T: V \to W$ is an invertible linear transformation, T is called an isomorphism, and we say V and W are isomorphic vector spaces
  - Isomorphism - "same form"
- $\forall \vec{v} \in V$, $T^{-1}T(\vec{v}) = \vec{v}$
  - $T^{-1}T = I_V$ is the identity linear transformation of V
- $\forall \vec{w} \in W$, $TT^{-1}(\vec{w}) = \vec{w}$
  - $TT^{-1} = I_W$ is the identity linear transformation of W
- If S is a linear transformation that is a candidate for the inverse, we need to verify that $ST = I_V$ and $TS = I_W$
- Let V be any vector space, and let $I: V \to V$ be the identity transformation. $I(\vec{v}) = \vec{v}$ for all $\vec{v} \in V$.
  - I is an isomorphism and the inverse of I is I itself
    - $I(I(\vec{v})) = \vec{v}$ for all $\vec{v} \in V$

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

- If $T: V \to W$ is an isomorphism of finite-dimensional vector spaces, and we choose bases

  $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$ for V and $\beta = \{\vec{w}_1, \dots, \vec{w}_n\}$ for W,

  T is represented by an $n \times n$ matrix $[T]_\alpha^\beta$.

  T^{-1} is reprented by a matrix $[T^{-1}]_\beta^\alpha$

- We can use the Gauss-Jordan method to construct $[T^{-1}]_\beta^\alpha$ if we already know $[T]_\alpha^\beta$.

  Let A denote $[T]_\alpha^\beta$ w/ entries a_{ij}

  Let B denote $[T^{-1}]_\beta^\alpha$ w/ entries b_{ij}

$$[A \mid I] = \left[\begin{array}{ccc|ccc} a_{11} & \cdots & a_{1n} & 1 & & \\ \vdots & \ddots & \vdots & & \ddots & \\ a_{n1} & \cdots & a_{nn} & & & 1 \end{array}\right]$$

Perform row operations on A and I simultaneously until A is in echelon form, and the resulting matrix on the right will be $B = [T^{-1}]_\beta^\alpha$

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

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

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

- Let $T: V \to W$ be an isomorphism. To solve for $T(\vec{v}) = \vec{w}$, apply T^{-1} on both sides, and the solution is $\vec{v} = T^{-1}(\vec{w})$

  - Likewise, $A\vec{x} = \vec{b}$ is solved by $\vec{x} = A^{-1}\vec{b}$

2.7 Change of Basis

- (2.7.3) Proposition. Let V be a finite-dimensional vector space, and let α and α' be bases for V. Let $\vec{v} \in V$. Then the coordinate vector $[\vec{v}]_{\alpha'}$ of $\vec{v}$ in the basis α' is related to the coordinate vector $[\vec{v}]_{\alpha}$ of $\vec{v}$ in the basis α by $[I]_{\alpha}^{\alpha'}[\vec{v}]_{\alpha} = [\vec{v}]_{\alpha'}$
  - The matrix $[I]_{\alpha}^{\alpha'}$ is the change of basis matrix which changes from α coordinates to α' coordinates
  - $([I]_{\alpha}^{\alpha'})^{-1} = [I]_{\alpha'}^{\alpha}$
- (2.7.5) Theorem. Let $T: V \to W$ be a linear transformation between finite-dimensional vector spaces V and W. Let $I_V: V \to V$ and $I_W: W \to W$ be the respective identity transformations of V and W. Let α and α' be two bases for V, and let β and β' be two bases for W. Then

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

- Rewriting, $[T]_{\alpha'}^{\beta'} = ([I_W]_{\beta'}^{\beta})^{-1} \cdot [T]_{\alpha}^{\beta} \cdot [I_V]_{\alpha'}^{\alpha}$
- When $V = W$ and $\alpha = \beta$ and $\alpha' = \beta'$,

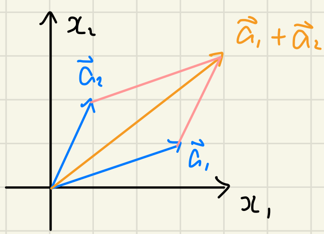
  $[T]_{\alpha'}^{\alpha'} = ([I_V]_{\alpha'}^{\alpha})^{-1} \cdot [T]_{\alpha}^{\alpha} \cdot [I_V]_{\alpha'}^{\alpha}$

- (2.7.6) Definition. Let A, B be $n \times n$ matrices. A and B are similar if there is an invertible $n \times n$ matrix Q such that $B = Q^{-1}AQ$
- If $T: V \to V$ is a linear transformation and α and α' are two bases for V, then $A = [T]_{\alpha}^{\alpha}$ is similar to $B = [T]_{\alpha'}^{\alpha'}$ and the invertible matrix Q in the definition is the matrix $Q = [I_V]_{\alpha'}^{\alpha}$
- Ex. Find the change-of-basis matrix from $\alpha = \{(1,2),(3,4)\}$ to $\alpha' = \{(0,3),(1,1)\}$ in $\mathbb{R}^2$
  - Use the standard basis $\beta = \{(1,0),(0,1)\}$ as an intermediate role
  - $[I]_{\alpha}^{\beta} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$, $[I]_{\alpha'}^{\beta} = \begin{bmatrix} 0 & 1 \\ 3 & 1 \end{bmatrix}$
  - $[I]_{\alpha}^{\alpha'} = [I]_{\beta}^{\alpha'}[I]_{\alpha}^{\beta} = ([I]_{\alpha'}^{\beta})^{-1}[I]_{\alpha}^{\beta}$

3.1 The Determinant as Area

- Let $V = \mathbb{R}^2$. Let α be the standard basis.

  $[T]_\alpha^\alpha$ is a 2×2 matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$

  - Let $\vec{a}_1 = (a_{11}, a_{12})$, $\vec{a}_2 = (a_{21}, a_{22})$ denote the rows of A.
  - To determine if T is an isomorphism, we need to check if the rows of A are linearly independent
  - 

    If $\vec{a}_1$ and $\vec{a}_2$ are linearly independent, then $\vec{a}_1$ is not a scalar multiple of $\vec{a}_2$, thus the sides of the parallelogram are not collinear and the area of the parallelogram is not zero

- (3.1.1) Proposition.

    i) The area of the parallelogram with vertices $\vec{0}$, $\vec{a}_1$, $\vec{a}_2$, $\vec{a}_1+\vec{a}_2$ is $\pm(a_{11}a_{22} - a_{12}a_{21})$

    ii) The area is not zero iff the vectors $\vec{a}_1$ and $\vec{a}_2$ are linearly independent.

- (3.1.2) Corollary. Let $V=\mathbb{R}^2$, $T: V \to V$ is an isomorphism iff the area of the parallelogram constructed previously is nonzero.

- If the angle from $\vec{a}_1$ to $\vec{a}_2$ (which is $\leq \pi$) is traced counterclockwise (clockwise) from $\vec{a}_1$ to $\vec{a}_2$, then the area is positive (negative)

- (3.1.3) Proposition. The function $\text{Area}(\vec{a}_1, \vec{a}_2)$ has the following properties for $\vec{a}_1, \vec{a}_2, \vec{a}_1', \vec{a}_2' \in \mathbb{R}^2$:

    i) $\text{Area}(b\vec{a}_1 + c\vec{a}_1', \vec{a}_2)$

    $= b\,\text{Area}(\vec{a}_1, \vec{a}_2) + c\,\text{Area}(\vec{a}_1', \vec{a}_2)$ for $b, c \in \mathbb{R}$

    i') $\text{Area}(\vec{a}_1, b\vec{a}_2 + c\vec{a}_2')$

    $= b\,\text{Area}(\vec{a}_1, \vec{a}_2) + c\,\text{Area}(\vec{a}_1, \vec{a}_2')$ for $b, c \in \mathbb{R}$

ii) $\text{Area}(\vec{a}_1, \vec{a}_2) = -\text{Area}(\vec{a}_2, \vec{a}_1)$

iii) $\text{Area}((1,0),(0,1)) = 1$

- Properties i) and i') tells us that Area is a linear function of each variable when the other is held fixed
  - This property is multilinearity
- Property ii) is the alternating property
- Property iii) is the normalization property

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

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

  i) $\det(b\vec{a}_1 + c\vec{a}_1', \vec{a}_2)$

  $= b\det(\vec{a}_1, \vec{a}_2) + c\det(\vec{a}_1', \vec{a}_2)$ for $b, c \in \mathbb{R}$

  ii) $\det(\vec{a}_1, \vec{a}_2) = -\det(\vec{a}_2, \vec{a}_1)$

iii) $\det(\vec{e}_1, \vec{e}_2) = 1$

- $\det A$ is given explicitly by

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

- (3.1.6) Proposition.

  i) A 2×2 matrix A is invertible iff $\det A \neq 0$

  ii) If $T: V \to V$ is a linear transformation of a two-dimensional vector space V, then T is an isomorphism iff $\det([T]_\alpha^\alpha) \neq 0$

- $$A^{-1} = \frac{1}{\det A}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

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

- For a 3×3 matrix A w/ rows $\vec{a}_1, \vec{a}_2, \vec{a}_3$ we could define the determinant by using the signed volume of the parallelopiped in $\mathbb{R}^3$ w/ vertices $\vec{0}, \vec{a}_1, \vec{a}_2, \vec{a}_3, \vec{a}_1 + \vec{a}_2, \vec{a}_1 + \vec{a}_3, \vec{a}_2 + \vec{a}_3, \vec{a}_1 + \vec{a}_2 + \vec{a}_3$
    - The set of vectors $\{\vec{v} \mid \vec{v} = x_1\vec{a}_1 + x_2\vec{a}_2 + x_3\vec{a}_3, 0 \leq x_i \leq 1\}$
- (3.2.1) Definition. A function f of the rows of a matrix A is multilinear if f is a linear function of each of its rows when the remaining rows are held fixed. That is, f is multilinear if for all $b, b' \in \mathbb{R}$, $f(\vec{a}_1, \dots, b\vec{a}_i + b'\vec{a}_i', \dots, \vec{a}_n)$
  $= bf(\vec{a}_1, \dots, \vec{a}_i, \dots, \vec{a}_n) + b'f(\vec{a}_1, \dots, \vec{a}_i', \dots, \vec{a}_n)$
- (3.2.2) Definition. A function f of the rows of a matrix A is alternating if whenever any two rows of A are interchanged f changes sign. That is, for all $i \neq j$, $1 \leq (i, j) \leq n$, we have

$f(\vec{a}_1, \ldots, \vec{a}_i, \ldots, \vec{a}_j, \ldots, \vec{a}_n)$

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

- (3.2.3) Lemma. If f is an alternating real-valued function of the rows of an $n \times n$ matrix and two rows of the matrix A are identical, then $f(A) = 0$.
- (3.2.4) Definition. Let A be an $n \times n$ matrix w/ entries a_{ij}, $(i,j) = 1, \ldots, n$. The ijth minor of A is defined to be the $(n-1) \times (n-1)$ matrix obtained by deleting the ith row and the jth column of A. The ijth minor is denoted by A_{ij}

$$A_{ij} = \begin{bmatrix} a_{11} & \cdots & a_{1,j-1} & a_{1,j+1} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{i-1,1} & \cdots & a_{i-1,j-1} & a_{i-1,j+1} & \cdots & a_{i-1,n} \\ a_{i+1,1} & \cdots & a_{i+1,j-1} & a_{i+1,j+1} & \cdots & a_{i+1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{n,j-1} & a_{n,j+1} & \cdots & a_{nn} \end{bmatrix}$$

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

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

- (3.2.6) Corollary. There exists exactly 1 multilinear alternating function f of the rows of a 3×3 matrix such that $f(I) = 1$
- (3.2.7) Definition. The determinant function of a 3×3 matrix is the unique alternating multilinear function f w/ $f(I) = 1$. This function will be denoted by $\det(A)$
- From Corollary (3.2.6), $\det(A) =$

$$(-1)^{1+1} a_{11} \det(A_{11}) + (-1)^{1+2} a_{12} \det(A_{12}) + (-1)^{1+3} a_{13} \det(A_{13})$$

  - However, we can choose any other row instead of the first row
  - $\det(A) = \sum_{j=1}^{3} (-1)^{i+j} a_{ij} \det(A_{ij})$ for $i = 1, 2, 3$

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

$f(A) = \det(A) f(I)$

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

- (3.2.11) Proposition. $\det(\vec{a}_1, \dots, \vec{a}_n) =$

  $\det(\vec{a}_1, \dots, \underbrace{\vec{a}_i + b\vec{a}_j}_{\text{ith position}}, \dots, \vec{a}_n)$

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

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

- (3.2.14) Theorem. Let A be an $n \times n$ matrix. A is invertible iff $\det A \neq 0$

3.3 Further Properties of the Determinant

- Consider the matrix $B_{k\ell}$ whose rows are

  $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_k, \ldots, \underbrace{\vec{a}_k}_{\ell\text{th position}}, \ldots, \vec{a}_n$

  - $B_{k\ell}$ is equal to A except that we have replaced the ℓth row of A by the kth row of A
  - Determinant is 0 if $k \neq \ell$
    - $0 = \det(B_{k\ell}) = \sum_{j=1}^{n} (-1)^{j+\ell} a_{kj} \det(A_{\ell j})$
  - Determinant is $\det A$ if $k = \ell$
- Let A' be the matrix whose entries a_{ij}' are the scalars $(-1)^{i+j} \det(A_{ji})$. The quantity a_{ij}' is the jith cofactor of A
- (3.3.1) Proposition. $AA' = \det(A) I$
  - $k\ell$th entry is $\sum_{j=1}^{n} a_{kj} a_{j\ell}'$

    $= \sum_{j=1}^{n} a_{kj} (-1)^{j+\ell} \det(A_{\ell j})$
    - 0 if $k \neq \ell$, otherwise $\det A$

- (3.3.2) Corollary. If A is an invertible $n \times n$ matrix, then A^{-1} is the matrix whose ijth entry is $(-1)^{i+j} \det(A_{ji}) / \det A$
    - Divide by $\det A$ in Proposition (3.3.1)
- (3.3.4) Proposition. For any fixed j, $1 \leq j \leq n$

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

- (3.3.7) Proposition. If A and B are $n \times n$ matrices, then
    a) $\det(AB) = \det A \det B$
    b) If A is invertible, then $\det(A^{-1}) = 1/\det A$
- (3.3.8) Corollary. If $T: V \to V$ is a linear transformation, $\dim V = n$, then $\det([T]_\alpha^\alpha) = \det([T]_\beta^\beta)$ for all choices of bases α and β of V
- (3.3.9) Definition. The determinant of a linear transformation $T: V \to V$ of a finite-dimensional vector space is the determinant of $[T]_\alpha^\alpha$ for any choice of α. We denote this by $\det(T)$.
- (3.3.11) Proposition. A linear transformation $T: V \to V$ of a finite-dimensional vector space is an

isomorphism iff $\det T \neq 0$

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

  i) $\det(ST) = \det S \det T$

  ii) if T is an isomorphism, then

  $\det(T^{-1}) = \det(T)^{-1}$

- (3.3.13) Proposition. (Cramer's rule) Let A be an invertible $n \times n$ matrix. The solution $\vec{x}$ to the system of equations $A\vec{x} = \vec{b}$ is the vector whose jth entry is the quotient $\det(B_j)/\det A$ where B_j is the matrix obtained from A by replacing the jth column of A by the vector $\vec{b}$

  - If $\vec{b} \in \mathbb{R}^n$, $A'\vec{b}$ is a vector whose ith entry is $\sum_{j=1}^{n} a'_{ij} b_j$

    $= \sum_{j=1}^{n} b_j (-1)^{i+j} \det(A_{ji})$

    - This is the determinant of the matrix whose columns are

$\vec{a}_1, \dots, \vec{a}_{i-1}, \vec{b}, \vec{a}_{i+1}, \dots, \vec{a}_n$ where

$\vec{a}_j$, $1 \leq j \leq n$, is the jth column of A

- $A'\vec{b} / \det A = A^{-1}\vec{b}$
- $A'\vec{b} = \det(B_j)$

4.1 Eigenvalues and Eigenvectors

- (4.1.2) Definitions. Let $T: V \to V$ be a linear mapping.
  - a) A vector $\vec{x} \in V$ is called an eigenvector of T if $\vec{x} \neq \vec{0}$ and there exists a scalar $\lambda \in \mathbb{R}$ such that $T(\vec{x}) = \lambda \vec{x}$
  - b) If $\vec{x}$ is an eigenvector of T and $T\vec{x} = \lambda \vec{x}$, the scalar λ is called the eigenvalue of T corresponding to $\vec{x}$
  - Those terms are also known as characteristic value and characteristic vector
- Ex. Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be the orthogonal projection on the x_1-axis.
  - Every vector of the form $\begin{bmatrix} a \\ 0 \end{bmatrix}$, $a \neq 0$ is an eigenvector of T w/ eigenvalue of $\lambda = 1$
  - Vectors of the form $\begin{bmatrix} 0 \\ b \end{bmatrix}$, $b \neq 0$ are eigenvectors of T w/ eigenvalue $\lambda = 0$
- Ex. Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be a rotation through an angle

$\theta \neq 0, \pm\pi, \pm 2\pi$, etc.

- T has no eigenvectors or eigenvalues
- $T(\vec{x})$ never lies on the same line through the origin as $\vec{x}$ if $\vec{x} \neq 0$

- (4.1.5) Proposition. A vector $\vec{x}$ is an eigenvector of T w/ eigenvalue λ iff $\vec{x} \neq \vec{0}$ and $\vec{x} \in \ker(T - \lambda I)$
- (4.1.6) Definition. Let $T: V \to V$ be a linear mapping, and let $\lambda \in \mathbb{R}$. The λ-eigenspace of T, denoted E_λ, is the set

$$E_\lambda = \{\vec{x} \in V \mid T(\vec{x}) = \lambda \vec{x}\}$$

  - E_λ is the set containing all the eigenvectors of T w/ eigenvalue λ, together w/ $\vec{0}$
  - If λ is not an eigenvalue of T, then $E_\lambda = \{\vec{0}\}$
- (4.1.7) Proposition. E_λ is a subspace of V for all λ
  - $E_\lambda = \ker(T - \lambda I)$, which is a subspace
- (4.1.9) Proposition. Let $A \in M_{n \times n}(\mathbb{R})$. Then $\lambda \in \mathbb{R}$ is an eigenvalue of A iff $\det(A - \lambda I) = 0$
  - I denotes the $n \times n$ identity matrix

- Ex. Let $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$

  - $\det A = \det \begin{bmatrix} -\lambda & 1 \\ -1 & -\lambda \end{bmatrix} = \lambda^2 + 1$

  - There are no real roots of the equation $\lambda^2 + 1 = 0$, so A has no real eigenvalues

- When we expand out a determinant of the form $\det(A - \lambda I)$, we obtain a polynomial in λ whose coefficients depend on the entries in the matrix A

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

- The characteristic polynomial depends on the linear mapping defined by the matrix A, not on the matrix itself

- (4.1.12) Proposition. Similar matrices have equal characteristic polynomials

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

$$\det \begin{bmatrix} a-\lambda & b \\ c & d-\lambda \end{bmatrix} = \lambda^2 - (a+d)\lambda + (ad-bc)$$

- The constant term is $\det A$
- The coefficient of λ is called the trace of A, denoted by $\mathrm{Tr}(A)$
    - The trace is the sum of the diagonal entries
- $\det(A - \lambda I) = \lambda^2 - \mathrm{Tr}(A)\lambda + \det(A)$

- In general, a characteristic polynomial has the form

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

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

- (4 1. 14) Corollary. Let $A \in M_{n \times n}(\mathbb{R})$. Then A has no more than n distinct eigenvalues. In addition, if $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of A and λ_i is an m_i-fold root of the characteristic polynomial, then $m_1 + \cdots + m_k \leq n$

- Once the eigenvalues of a given matrix have been found, the corresponding eigenvectors may be found by solving the systems $(A - \lambda_i I)\vec{x} = \vec{0}$ for each eigenvalue λ_i

- Ex. $A = \begin{bmatrix} 2 & 5 \\ 0 & 2 \end{bmatrix}$
  - Characteristic polynomial is $\det(A - \lambda I) = (2-\lambda)^2$
  - The only eigenvalue is $\lambda = 2$
  - Eigenspace E_2 is the set of solutions to $(A - 2I)\vec{x} = \vec{0}$
  - $E_2 = \text{span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\}$
- (4.1.18) Theorem (Cayley-Hamilton Theorem).

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

4.2 Diagonalizability

- (4.2.1) Definition. Let V be a finite-dimensional vector space, and let $T: V \to V$ be a linear mapping. T is **diagonalizable** if there exists a basis of V, all of whose vectors are eigenvectors of T.
  - If $\beta = \{\vec{x}_1, \dots, \vec{x}_n\}$ is such a basis, then for each $\vec{x}_i$, we have $T(\vec{x}_i) = \lambda_i \vec{x}_i$ for some eigenvalue λ_i. Hence,

$$[T]_\beta^\beta = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} = D$$

- (4.2.2) Proposition. $T: V \to V$ is diagonalizable iff for any basis α of V, the matrix $[T]_\alpha^\alpha$ is similar to a diagonal matrix.
- In order for a linear mapping or a matrix to be diagonalizable, it must have enough linearly independent eigenvectors to form a basis of V
- (4.2.4) Proposition. Let $\vec{x}_i$ $(1 \le i \le k)$ be eigenvectors of

a linear mapping $T: V \to V$ corresponding to distinct eigenvalues λ_i. Then $\{\vec{x}_1, \dots, \vec{x}_k\}$ is a linearly independent subset of V.

- (4.2.5) Corollary. For each i $(1 \le i \le k)$, let $\{\vec{x}_{i,1}, \dots, \vec{x}_{i,n_i}\}$ be a linearly independent set of eigenvectors of T all w/ eigenvalue λ_i and suppose the λ_i are distinct. Then
$$S = \{\vec{x}_{1,1}, \dots, x_{1,n_1}\} \cup \{\vec{x}_{k,1}, \dots, \vec{x}_{k,n_k}\}$$ is linearly independent.

- (4.2.6) Proposition. Let V be finite-dimensional, and let $T: V \to V$ be linear. Let λ be an eigenvalue of T, and assume that λ is an m-fold root of the characteristic polynomial of T. Then we have
$$1 \le \dim(E_\lambda) \le m$$

- (4.2.7) Theorem. Let $T: V \to V$ be a linear mapping on a finite-dimensional vector space V, and let $\lambda_1, \dots, \lambda_k$ be its distinct eigenvalues. Let m_i be the multiplicity of λ_i as a root of the characteristic polynomial of T.

Then T is diagonalizable iff

i) $m_1 + \cdots + m_k = n = \dim V$, and

- T must have n real eigenvalues

ii) for each i, $\dim(E_{\lambda_i}) = m_i$

- For each of the eigenvalues, the maximum possible dimension for the eigenspace E_{λ_i} must be attained

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

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

  i) We have $\sum_{i=1}^{k} \dim(E_{\lambda_i}) = n$, or

  ii) We have $\sum_{i=1}^{k} \left(n - \dim(\mathrm{im}(T - \lambda_i I))\right) = n$

  where the λ_i are the distinct eigenvalues of T

5.1 Complex Numbers

- Goal: to construct a set F such that $\mathbb{R} \subset F$, and possesses two operations on pairs of elements of F, addition and multiplication, satisfying the following rules:
  a) When restricted to the subset $\mathbb{R} \subset F$, these are the usual operations of addition and multiplication of real numbers
  b) The operations on F must satisfy the same properties as do addition and multiplication of real numbers
  c) Every polynomial $p(z) = a_n z^n + \cdots + a_1 z + a_0$, $a_i \in F$ for $i = 0, \ldots, n$, has n roots in F
     - A set satisfying this condition is algebraically closed
- (5.1.1) Definition. The set of complex numbers, denoted $\mathbb{C}$, is the set of ordered pairs of real numbers (a, b) w/ the operations of addition and multiplication defined by:

- For all $(a,b), (c,d) \in \mathbb{C}$:
    - The sum of (a,b) and (c,d) is the complex number defined by
      $(a,b)+(c,d) = (a+c, b+d)$
    - The product of (a,b) and (c,d) is the complex number defined by
      $(a,b)(c,d) = (ac-bd, ad+cb)$
- The subset $\{(a,0) \mid a \in \mathbb{R}\} \subset \mathbb{C}$ is the real numbers
    - $(a,0)+(c,0) = (a+c, 0)$
    - $(a,0)(c,0) = (ac-0, 0a+0c) = (ac, 0)$
- $(0,1)(0,1) = (0-1, 0+0) = (-1, 0)$
    - Square of $(0,1)$ is -1
- We denote the complex number $(1,0)$ by 1 and $(0,1)$ by i or $\sqrt{-1}$
    - We write $a+bi$ for $a(1,0)+b(0,1)$
- (5.1.2) Definition. Let $z = a+bi \in \mathbb{C}$.
  The real part of z, denoted $\mathrm{Re}(z)$, is the real number a. The imaginary part of z,

denoted $\mathrm{Im}(z)$, is the real number b.

z is called a real number if $\mathrm{Im}(z) = 0$ and purely imaginary if $\mathrm{Re}(z) = 0$

- Addition and multiplication of complex numbers:
  - $(a+bi) + (c+di) = (a+c) + (b+d)i$
  - $(a+bi)(c+di) = (ac-bd) + (ad+bc)i$
  - Examples:
    - $(1+i)(3-7i) = 10-4i$
    - $i^4 = 1$
    - $(-i)i = 1$

- (5.1.4) Definition. A field is a set F w/ 2 operations, defined on ordered pairs of elements of F, called addition and multiplication. Addition assigns to the pair $x, y \in F$ their sum, which is denoted by $x+y$ and multiplication assigns to the pair $x, y \in F$ their product, which is denoted by $x \cdot y$ or xy. These two operations must satisfy the following properties for all $x, y, z \in F$:

i) Commutativity of addition: $x+y=y+x$

ii) Associativity of addition: $(x+y)+z=x+(y+z)$

iii) Existence of an additive identity: There exists an element $0 \in F$, called zero, s.t. $x+0=x$

iv) Existence of additive inverses: For each x there is an element $-x \in F$ s.t. $x+(-x)=0$

v) Commutativity of multiplication: $xy=yx$

vi) Associativity of multiplication: $(xy)z=x(yz)$

vii) Distributivity:

$(x+y)z=xz+yz$ and $x(y+z)=xy+xz$

viii) Existence of multiplicative identity: There exists an element $1 \in F$, called 1, s.t. $x \cdot 1=x$

ix) Existence of multiplicative inverses: If $x \neq 0$, then there is an element $x^{-1} \in F$ s.t. $xx^{-1}=1$

- (5.1.5) Proposition. The set of complex numbers is a field w/ the operations of addition and multiplication as defined previously.

  - Additive identity is $0+0i$

- Additive inverse of $a+bi$ is $-a-bi$
- Multiplicative identity is $1+0i$
- Multiplicative inverse of $a+bi$ is $\frac{a-ib}{a^2+b^2}$

- The complex number $a-bi$ is the complex conjugate of $z=a+bi$ and is denoted by $\bar{z}$
  - $z^{-1} = \bar{z}/z\bar{z}$
    - $z\bar{z} = a^2+b^2$

- Examples
  - $(i)^{-1} = -i$
  - $\frac{3-i}{4+i} = \frac{1}{17}(11-7i)$

- (5.1.7) Proposition.

  i) The additive identity in a field is unique

  ii) The additive inverse of an element of a field is unique

  iii) The multiplicative identity of a field is unique

  iv) The multiplicative inverse of a nonzero element of a field is unique

- (5.1.8) Definition. The absolute value of the complex number $z = a + bi$ is the nonnegative real number $\sqrt{a^2 + b^2}$ and is denoted by $|z|$ or $r = |z|$.

  The argument of the complex number z is the angle θ of the polar coordinate representation of z

  - We can write $z = |z|(\cos\theta + i\sin\theta)$ for $z = a + bi$
  - If θ is replaced by $\theta \pm 2\pi k$, $k \in \mathbb{Z}$, we have defined the same complex number.

- Ex. Consider the equation $|x| = c$, $c \geq 0$
  - In $\mathbb{R}$, it has 2 solutions $x = c, -c$ when $c > 0$ and 1 solution $x = 0$ when $c = 0$
  - In $\mathbb{C}$, if $z = a + ib$ satisfies $|z| = c > 0$, then $\sqrt{a^2 + b^2} = c$, so that $a^2 + b^2 = c^2$
    - i.e. z lies on the circle of radius c centered at the origin
  - If $c = 0$, we have the unique solution $z = 0$
  - If z_0 is a fixed nonzero complex number,

the equation $|z-z_0| = c > 0$ defines a circle of radius c centered at $z_0 \in \mathbb{C}$; if $c=0$, $|z-z_0|=0$ is solved uniquely by $z=z_0$

- If $z_j = r_j(\cos\theta_j + i\sin\theta_j)$, then
  - $z_1 z_2 = r_1 r_2 (\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2))$
  - $z^2 = r^2(\cos(2\theta) + i\sin(2\theta))$
  - $z^{-1} = r^{-1}(\cos(-\theta) + i\sin(-\theta))$, if $z \neq 0$
- $z^n = r^n(\cos(n\theta) + i\sin(n\theta))$ (De Moivre's formula)
  - Used to find roots of complex numbers
  - Fix $z_0 = r_0(\cos\theta_0 + i\sin\theta_0) \in \mathbb{C}$.

    Let n be a pos. integer.
  - An nth root of z_0 is a complex number z s.t. $z^n = z_0$
  - From De Moivre's formula, $z = r(\cos\theta + i\sin\theta)$ must satisfy $r^n = r_0$, $n\theta = \theta_0 + 2\pi k$, $k \in \mathbb{Z}$
    - Same r, same angle
  - Since $r, r_0 > 0$, we know that $r = r_0^{1/n}$

defines r uniquely

- θ is of the form $\theta_0/n + 2\pi k/n$, $k \in \mathbb{Z}$
    - The values of $k = 0, 1, \ldots, n-1$ determine n distinct values for θ
    - Any other value of k (e.g. $k = n$) would yield one of the n values of θ we have already determined
- The n nth roots of the complex number $z_0 \neq 0$ are $r_0^{1/n}(\cos(\theta_0/n + 2\pi k/n) + i\sin(\theta_0/n + 2\pi k/n))$ for $k = 0, 1, \ldots, n-1$

- Ex. Consider $i = \cos(\pi/2) + i\sin(\pi/2)$
    - In polar coordinates, values of $\sqrt{i}$ are given by $r = 1$, $\theta = \pi/4$ and $r = 1$, $\theta = 5\pi/4$
        - $\sqrt{i} = \pm\frac{\sqrt{2}}{2}(1+i)$
    - $\sqrt[3]{i} = (\sqrt{3}+i)/2,\ (-\sqrt{3}+i)/2,\ -i$

- (5.1.11) Definition. A field F is algebraically closed if every polynomial $p(z) = a_n z^n + \cdots + a_1 z + a_0$ w/ $a_i \in F$, $a_n \neq 0$ has n roots counted w/ multiplicity in F.

- (5.1.12) Theorem (Fundamental Theorem of Algebra).

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

5.2 Vector Spaces Over a Field

- (5.2.1) Definition. A vector space over a field F is a set V (whose elements are called vectors) together w/

    a) An operation called vector addition, which for each pair of vectors $\vec{x}, \vec{y} \in V$ produces a vector denoted $\vec{x} + \vec{y} \in V$

    b) An operation called multiplication by a scalar (a field element), which for each vector $\vec{x} \in V$, and each scalar $c \in F$ produces a vector denoted $c\vec{x} \in V$

  Furthermore, the two operations must satisfy the following axioms:

    1) $\forall \vec{x}, \vec{y}, \vec{z} \in V, \ (\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$

    2) $\forall \vec{x}, \vec{y} \in V, \ \vec{x} + \vec{y} = \vec{y} + \vec{x}$

    3) $\exists \vec{0} \in V$ s.t. $\forall \vec{x} \in V, \ \vec{x} + \vec{0} = \vec{x}$

    4) $\forall \vec{x} \in V, \ \exists -\vec{x} \in V$ s.t. $\vec{x} + -\vec{x} = \vec{0}$

    5) $\forall \vec{x}, \vec{y} \in V, \ \forall c \in F, \ c(\vec{x} + \vec{y}) = c\vec{x} + c\vec{y}$

6) $\forall \vec{x} \in V$, $\forall c, d \in F$, $(c+d)\vec{x} = c\vec{x} + d\vec{x}$

7) $\forall \vec{x} \in V$, $\forall c, d \in F$, $(cd)\vec{x} = c(d\vec{x})$

8) $\forall \vec{x} \in V$, $1\vec{x} = \vec{x}$

- Ex. Consider F^n, the collection of n-tuples of elements of F, defined by

  $F^n = \{\vec{x} = (x_1, \dots, x_n) \mid x_i \in F \text{ for } i = 1, \dots, n\}$

  - We define addition and scalar multiplication coordinate by coordinate
    - If $\vec{x} = (x_1, \dots, x_n)$, $\vec{y} = (y_1, \dots, y_n) \in F^n$,

      $\vec{x} + \vec{y} = (x_1 + y_1, \dots, x_n + y_n)$
    - If $c \in F$, $c\vec{x} = (cx_1, \dots, cx_n)$
  - F^n is a vector space over F
- If V is a vector space over a field F, then the definition of a vector subspace still holds
  - $W \subset V$ is a vector subspace if W is a vector space in its own right using the operations inherited from V
- Linear dependence and independence are defined as before

- Let V, W be vector spaces over a field F.

  A linear transformation $T: V \to W$ is a function from V to W, which satisfies

  $T(a\vec{u} + b\vec{v}) = aT\vec{u} + bT\vec{v}$ for $\vec{u}, \vec{v} \in V$ and $a, b \in F$

- Ex. Consider $T: \mathbb{C}^3 \to \mathbb{C}^2$ defined by matrix multiplication by the matrix $\begin{bmatrix} 1+i & 0 & 3+i \\ 1 & -i & 2+i \end{bmatrix}$

  - Row-reduce it to obtain the kernal:

    $\ker T = \operatorname{span}\{(-2+i, 2, 1)\}$, $\dim \ker T = 1$

  - By Dimension Theorem $\dim \operatorname{im} T = 2$, and so

    $\operatorname{im} T = \mathbb{C}^2$ and T is surjective

- If A is any $n \times n$ matrix w/ real entries, we may view A as a matrix w/ complex entries, since $\mathbb{R} \subset \mathbb{C}$

  - Then we can use the FTA (Theorem 5.1.12)
  - We would only need to determine if the dimension of the eigenspace of the eigenvalue λ is equal to the multiplicity of λ for each eigenvalue λ of the matrix or transformation.

- $\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} = \begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix}$

- If $T: \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation, we can extend T to a complex linear transformation $T: \mathbb{C}^n \to \mathbb{C}^n$
    - We extend the basis α of real numbers to the basis α' of complex numbers

6.1 Triangular Form

- (6.1.2) Definition. Let $T: V \to V$ be a linear mapping. A subspace $W \subset V$ is invariant (or stable) under T if $T(W) \subset W$
    - $\{\vec{0}\}$ and V are invariant under all linear mappings $T: V \to V$
    - $\ker T$ and $\operatorname{im} T$ are invariant subspaces
    - If λ is an eigenvalue of T, then the eigenspace E_λ is invariant under T
        - i.e. for $\vec{v} \in E_\lambda$, $T(\vec{v}) = \lambda\vec{v} \in E_\lambda$
- (6.1.4) Proposition. Let V be a vector space, let $T: V \to V$ be a linear mapping, and let $\beta = \{\vec{x}_1, \dots, \vec{x}_n\}$ be a basis for V. Then $[T]_\beta^\beta$ is upper-triangular iff each of the subspaces $W_i = \operatorname{span}(\{\vec{x}_1, \dots, \vec{x}_i\})$ is invariant under T.
    - $\{\vec{0}\} \subset W_1 \subset W_2 \subset \cdots \subset W_{n-1} \subset W_n = V$
    - The W_i form an increasing sequence of subspaces

- (6.1.5) Definition. A linear mapping $T: V \to V$ on a finite-dimensional vector space V is triangularizable if there exists a basis β s.t. $[T]_\beta^\beta$ is upper-triangular
    - To show that a given mapping is triangularizable, we must produce the increasing sequence of invariant subspaces spanned by the subsets of the basis β and show that each W_i is invariant under T
- Let $T|_W : W \to W$ define the restriction of T to an invariant subspace $W \subset V$.
- (6.1.6) Proposition. Let $T: V \to V$, and let $W \subset V$ be an invariant subspace. Then the characteristic polynomial of $T|_W$ divides the characteristic polynomial of T
    - Every eigenvalue of $T|_W$ is also an eigenvalue of T
    - The set of eigenvalues of $T|_W$ is some

subset of the eigenvalues of T

- (6.1.8) Theorem. Let V be a finite-dimensional vector space over a field F, and let $T: V \to V$ be a linear mapping. Then T is triangularizable iff the characteristic polynomial equation of T has $\dim V$ roots (counted w/ multiplicities) in the field F
    - Satisfied automatically if $F = \mathbb{C}$ (by FTA)
- (6.1.10) Lemma. Let $T: V \to V$ be as in the theorem, and assume that the characteristic polynomial of T has $n = \dim V$ roots in F. If $W \subsetneq V$ is an invariant subspace under T, then there exists a vector $\vec{x} \neq \vec{0}$ in V s.t. $\vec{x} \notin W$ and $W + \text{span}(\{\vec{x}\})$ is also invariant under T
- (6.1.11) Corollary. If $T: V \to V$ is triangularizable, w/ eigenvalues λ_i w/ respective multiplicities m_i, then there exists a basis β for V s.t. $[T]_\beta^\beta$ is upper-triangular, and the diagonal entries of

$[T]_\beta^\beta$ are m_1 λ_1s, followed by m_2 λ_2s, and so on.

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

6.2 A Canonical Form for Nilpotent Mappings

- A matrix $A \in M_{n\times n}(\mathbb{R})$ is nilpotent if $A^k = 0$ for some integer $k \geq 1$
- $N: V \to V$ is nilpotent iff N has one eigenvalue $\lambda = 0$ w/ multiplicity $n = \dim V$
- Let $\dim V = n$, and let $N: V \to V$ be a nilpotent mapping
  - $N^n = 0$
  - Given any vector $\vec{v} \in V$, we will have $N^n(\vec{x}) = \vec{0}$
  - For a given $\vec{x}$, it may be true that $N^k(\vec{x}) = \vec{0}$ for some $k < n$
  - For each $\vec{x} \in V$, either $\vec{x} = \vec{0}$ or there is a unique integer k, $1 \leq k \leq n$, s.t. $N^k(\vec{x}) = \vec{0}$ but $N^{k-1}(\vec{x}) \neq 0$
    - If $\vec{x} \neq 0$, the set $\{N^{k-1}(\vec{x}), N^{k-2}(\vec{x}), \dots, N(\vec{x}), \vec{x}\}$ consists of distinct nonzero vectors

- (6.2.1) Definitions. Let N, $\vec{x} \neq 0$, k be as before.
  a) The set $\{N^{k-1}(\vec{x}), N^{k-2}(\vec{x}), \ldots, N(\vec{x}), \vec{x}\}$ is called the cycle generated by $\vec{x}$.
  $\vec{x}$ is the initial vector of the cycle
  b) The subspace $\text{span}(\{N^{k-1}(\vec{x}), N^{k-2}(\vec{x}), \ldots, \vec{x}\})$ is called the cyclic subspace generated by $\vec{x}$, and denoted $C(\vec{x})$
  c) The integer k is the length of the cycle
- (6.2.3) Proposition. With all notation as before:
  a) $N^{k-1}(\vec{x})$ is an eigenvector of N w/ eigenvalue $\lambda = 0$
  b) $C(\vec{x})$ is an invariant subspace of V under N
  c) The cycle generated by $\vec{x} \neq \vec{0}$ is a linearly independent set. Hence $\dim(C(\vec{x})) = k$, the length of the cycle
- Let N be an nilpotent mapping, let $C(\vec{x})$ be the cyclic subspace generated by some $\vec{x} \in V$, and let α be the cycle generated by $\vec{x}$, viewed as a

basis for $C(\vec{x})$. Since $N(N^i(\vec{x})) = N^{i+1}(\vec{x})$

for all i, we see that

$$[N|_{C(\vec{x})}]_\alpha^\alpha = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & 0 & \ddots & \\ & & & \ddots & 1 \\ & & & & 0 \end{bmatrix}$$

- (6.2.4) Proposition. Let $\alpha_i = \{N^{k_i-1}(\vec{x}_i), \ldots, \vec{x}_i\}$ $(1 \le i \le r)$ be cycles of length k_i, respectively. If the set of eigenvectors $\{N^{k_1-1}(\vec{x}_1), \ldots, N^{k_r-1}(\vec{x}_r)\}$ is linearly independent, then $\alpha_1 \cup \cdots \cup \alpha_r$ is linearly independent.

- (6.2.5) Definition. The cycles $\alpha_i = \{N^{k_i-1}(\vec{x}_i), \ldots, \vec{x}_i\}$ are non-overlapping cycles if $\alpha_1 \cup \cdots \cup \alpha_r$ is linearly independent.

- Ex. Let $V = \mathbb{R}^6$, let $N: V \to V$ be a mapping defined by the matrix $A = \begin{bmatrix} 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

  - $\vec{e}_1, \vec{e}_4, \vec{e}_6$ are eigenvectors of N

- Cycles:
  - $\alpha_1 = \{\vec{e}_1, -\vec{e}_2, (-1/2)\vec{e}_2 + (-1/2)\vec{e}_3\}$
  - $\alpha_2 = \{\vec{e}_4, (1/4)\vec{e}_5\}$
  - $\alpha_3 = \{\vec{e}_6\}$
- Since $\{\vec{e}_1, \vec{e}_4, \vec{e}_6\}$ is linearly independent, $\alpha_1, \alpha_2, \alpha_3$ are non-overlapping cycles
- $\beta = \alpha_1 \cup \alpha_2 \cup \alpha_3$ contains $6 = \dim \mathbb{R}^6$ vectors, so β is a basis for $\mathbb{R}^6$
- $$[N]_\beta^\beta = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

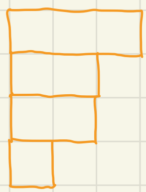
- If we can find a basis β for V that is the union of a collection of nonoverlapping cycles, then $[N]_\beta^\beta$ would only have 1s above the diagonal and 0s elsewhere
- (6.2.7) Definition. Let $N: V \to V$ be a nilpotent mapping on a finite-dimensional vector space V. We call a basis β a canonical basis (w.r.t. N) if β is the

union of a collection of nonoverlapping cycles for N

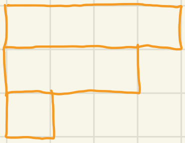
- (6.2.8) Theorem (Canonical form for nilpotent mappings)

  Let $N: V \to V$ be a nilpotent mapping on a finite-dimensional vector space. There exists a canonical basis β of V w.r.t. N.

- A cycle tableau is a pictorial representation of the vectors in cycles
  - Assume the cycles are arranged so that $k_1 \geq k_2 \geq \cdots \geq k_r$
  - The cycle tableau consists of r rows of boxes, where the boxes in the ith row represent the vectors in the ith cycle
    - k_i boxes in the ith row
  - Ex. if we have 4 cycles of length 3, 2, 2, 1, respectively, then the corresponding cycle tableau is



- The boxes in the left-most column represent the eigenvectors
- Applying the mapping N to a vector in any one of the cycles corresponds to shifting one box to the left on the corresponding row
    - Since $N(N^{k_i-1}(\vec{x}_i)) = \vec{0}$, the vectors in the left-most column get "pushed over the edge" and disappear
- (6.2.9) Lemma. Consider the cycle tableau corresponding to a canonical basis for a nilpotent mapping $N: V \to V$. Let r be the number of rows, and let k_i be the number of boxes in the ith row $(k_1 \geq k_2 \geq \cdots \geq k_r)$. For each j $(1 \leq j \leq k_1)$, the number of boxes in the jth column of the tableau is $\dim \ker(N^j) - \dim \ker(N^{j-1})$
- Ex. The cycle tableau corresponding to a canonical basis for N is

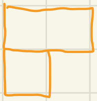


- $\dim \ker(N) = 3$
- $\dim \ker(N^2) = 5$
- $\dim \ker(N^3) = 7$
- $\dim \ker(N^4) = 8$

- (6.2.11) Corollary. The canonical form of a nilpotent mapping is unique (provided the cycles in the canonical basis are arranged so the lengths satisfy $k_1 \geq k_2 \geq \cdots \geq k_r$)
- Once we have the cycle tableau corresponding to a canonical basis, we do the following:
  1. Find an eigenvector of N that is an element of $\operatorname{Im}(N^{k-1})$ but not of $\operatorname{Im}(N^k)$
     - The final vector of the cycle of length k is in $\ker N \cap \operatorname{im}(N^{k-1})$
  2. The rest of the vectors can be found by solving systems of linear equations

- i.e. if $\vec{y}$ is the final vector of the cycle, solve for $N^{k-1}(\vec{x}) = y$

- Ex. Let $V = \mathbb{R}^3$, and consider the mapping $N : V \to V$ defined by $A = \begin{bmatrix} 1 & 1 & 1 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{bmatrix}$

  - $A^2 = 0$
  - $\dim \ker(N) = 2$, $\dim \ker(N^2) = 3$
  - 2 boxes in the first column, 1 box in the second
  - Cycle tableau for a canonical basis: 
  - $\ker N = \operatorname{span}\{(-1,1,0), (-1,0,1)\}$
  - $\operatorname{im} N = \operatorname{span}\{(1,-2,1)\}$
  - As the final vector of the cycle corresponding to the first row of the tableau, we take any nonzero vector in $\ker(N) \cap \operatorname{im}(N)$
    - i.e. $\vec{y} = (1,-2,1)$
  - The initial vector of the cycle of length 2 can be any vector that solves the system $A\vec{x} = \vec{y}$

- i.e. $\vec{x} = (1, 0, 0)$

- For the second cycle we find an eigenvector of N that together w/ vector $\vec{y} = (1, -2, 1)$ gives a linearly independent set
  - i.e. $(-1, 1, 0)$
- Hence, one canonical basis for N is $\beta = \{(1, -2, 1), (1, 0, 0), (-1, 1, 0)\}$

6.3 Jordan Canonical Form

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

  a) There exist subspaces $V_i' \subset V$ $(1 \le i \le k)$ s.t.

     i) Each V_i' is invariant under T

     ii) $T|_{V_i'}$ has exactly 1 distinct eigenvalue λ_i

     iii) $V = V_1' \oplus \cdots \oplus V_k'$

  b) There exists a basis β for V s.t.

     $[T]_\beta^\beta$ has a direct sum decomposition into upper-triangular blocks of the form

$$\begin{bmatrix} \lambda & & * \\ 0 & \lambda & \\ \vdots & \ddots & \ddots \\ 0 & \cdots \; 0 & \lambda \end{bmatrix}$$

(the entries above the diagonal are arbitrary and all entries in the matrix other than those in the diagonal blocks are 0)

- By Corollary (6.1.11), there is a basis α of V s.t. $[T]_\alpha^\alpha$ is upper-triangular of the form

$$\begin{bmatrix} \lambda_1 & & & & & & \\ & \ddots & & & * & & \\ & & \lambda_1 & & & & \\ & & & \lambda_2 & & & \\ & 0 & & & \ddots & & \\ & & & & & \lambda_2 & \\ & & & & & & \ddots \end{bmatrix}$$

  - We can block each eigenvalue and apply the same corollary (choosing another basis)

$$\begin{bmatrix} \boxed{\begin{matrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_1 \end{matrix}} & 0 & 0 \\ 0 & \boxed{\begin{matrix} \lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_2 \end{matrix}} & 0 \\ 0 & 0 & \ddots \end{bmatrix}$$

  - The first m_1 vectors in the basis, the second m_2 vectors, etc. all span subspaces that are invariant under T

- (6.3.2) Definitions. Let $T: V \to V$ be a linear mapping on a finite-dimensional vector space V. Let λ be an eigenvalue of T w/ multiplicity m.

a) The λ-generalized eigenspace, denoted by K_λ, is the kernal of the mapping $(T-\lambda I)^m$ on V

b) The non-zero elements of K_λ are called generalized eigenvectors of T

- A generalized eigenvector of T is any vector $\vec{x} \neq \vec{0}$ s.t. $(T-\lambda I)^m(\vec{x}) = \vec{0}$
    - Also works for $(T-\lambda I)^k(\vec{x}) = \vec{0}$ for some $k \leq m$
    - The eigenvectors w/ eigenvalue λ are also generalized eigenvectors
    - For each eigenvalue, $E_\lambda \subset K_\lambda$
- Ex. Let $T: \mathbb{C}^5 \to \mathbb{C}^5$ be the linear mapping whose matrix w.r.t. the standard basis is $\begin{bmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3-i & 1 \\ 0 & 0 & 0 & 0 & 3-i \end{bmatrix}$
    - Since $\lambda = 2$ is an eigenvalue of multiplicity $m=3$, the generalized eigenspace K_2 is $\ker((T-2I)^3)$
    - $K_2 = \text{span}\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$
    - Similarly, $K_{3-i} = \text{span}\{\vec{e}_4, \vec{e}_5\}$

- (6.3.4) Proposition.

  a) For each eigenvalue λ of T, K_λ is an invariant subspace of V

  b) If λ_i $(1 \leq i \leq k)$ are the distinct eigenvalues of T, then $V = K_{\lambda_1} \oplus \cdots \oplus K_{\lambda_k}$

  c) If λ is an eigenvalue of multiplicity m, then $\dim(K_\lambda) = m$

- Our canonical form for general linear mapping comes from combining the decomposition of V into the invariant subspaces K_{λ_i} for the different eigenvalues w/ our canonical form for nilpotent mappings (applied to the nilpotent mappings $N_i = (T - \lambda_i I)|_{K_{\lambda_i}}$). Since N_i is nilpotent, by Theorem (6.2.8) there exists a basis γ_i for K_{λ_i} s.t. $[N_i]_{\gamma_i}^{\gamma_i}$ has a direct sum decomposition into blocks of the

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

- $[T|_{K_{\lambda_i}}]_{\gamma_i}^{\gamma_i} = [N_i]_{\gamma_i}^{\gamma_i} + \lambda_i I$ has a direct sum decomposition into diagonal blocks of the form $\begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & & \vdots \\ \vdots & & \lambda_i & \ddots & 0 \\ & & & \ddots & 1 \\ 0 & \cdots & 0 & & \lambda_i \end{bmatrix}$

- (6.3.5) Definitions.

  a) A matrix of the above form is called a Jordan block matrix

  b) A matrix $A \in M_{n \times n}(F)$ is in Jordan canonical form if A is a direct sum of Jordan block matrices

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

  a) There exists a basis γ (called a canonical basis) of V s.t. $[T]_{\gamma}^{\gamma}$ has a direct sum decomposition into Jordan block matrices

b) In this decomposition the number of Jordan blocks and their sizes are uniquely determined by T

- The order in which the blocks appear in the matrix may be different for different canonical bases

6.4 Computing Jordan Form

- Computation of Jordan Canonical Form
  1. Find all the eigenvalues of T and their multiplicities by factoring the characteristic polynomial completely
  2. For each distinct eigenvalue λ_i, in turn, construct the cycle tableau for a canonical basis of K_{λ_i} w.r.t. the mapping $N_i = (T - \lambda_i I)|K_{\lambda_i}$ using the method of Lemma (6.2.9): for each j, the number of boxes in the jth column of the tableau for λ_i will be

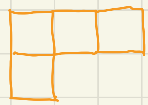
     $\dim \ker (T - \lambda_i I)^j - \dim \ker (T - \lambda_i I)^{j-1}$

     - $\ker (T - \lambda_i I)^j$ is contained in K_{λ_i} for all $j \geq 1$
  3. Form the corresponding Jordan blocks and assemble the matrix of T
- Ex. Assume we have a linear mapping $T: \mathbb{C}^8 \to \mathbb{C}^8$ w/ distinct eigenvalues $\lambda_1 = 2$, $\lambda_2 = i$, $\lambda_3 = -i$ having

multiplicities 4, 3, 1, respectively. Assume that

$$N_1 = (T-2I)|_{K_2} \text{ satisfies } \begin{cases} \dim \ker N_1 = 2 \\ \dim \ker N_1^2 = 3 \\ \dim \ker N_1^3 = 4 \end{cases}$$

$$N_2 = (T-iI)|_{K_i} \text{ satisfies } \begin{cases} \dim \ker N_2 = 1 \\ \dim \ker N_2^2 = 2 \\ \dim \ker N_2^3 = 3 \end{cases}$$

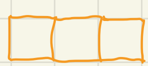
- Cycle tableau for N_1 is 

- Canonical form of the nilpotent mapping N_1 is

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- Hence, the canonical form for $T|_{K_2}$ is

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

- Cycle tableau for N_2 is . which corresponds to a canonical form containing

one Jordan block $\begin{bmatrix} i & 1 & 0 \\ 0 & i & 1 \\ 0 & 0 & i \end{bmatrix}$

- The one eigenvector for $\lambda = -i$ gives us a

  Jordan block of size 1

- $$C = \begin{bmatrix} 2 & 1 & & & & & & \\ & 2 & 1 & & & & 0 & \\ & & 2 & & & & & \\ & & & 2 & & & & \\ & & & & i & 1 & & \\ & & & & & i & 1 & \\ & 0 & & & & & i & \\ & & & & & & & -i \end{bmatrix}$$

- The union of the two nonoverlapping cycles for

  $\lambda = 2$, the cycle for $\lambda = i$, and the eigenvector

  for $\lambda = -i$ would form a canonical basis γ

  - $[T]_\gamma^\gamma = C$

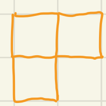
- Ex. Let $T: \mathbb{C}^5 \to \mathbb{C}^5$ be the linear mapping whose

  matrix w.r.t. the standard basis is

$$\begin{bmatrix} 1-i & 1 & 0 & 0 & 0 \\ 0 & 1-i & 0 & 0 & 0 \\ i & -1 & 1 & 0 & 0 \\ 0 & i & 1 & 1 & 1 \\ -i & 1 & 0 & 0 & 1 \end{bmatrix}$$

- $\det(T-\lambda I) = (1-\lambda)^3 [(1-i)-\lambda]^2$
    - $\lambda = 1$ is an eigenvalue w/ multiplicity 3
    - $\lambda = 1-i$ is an eigenvalue w/ multiplicity 2
- $\dim \ker (T-(1-i)I) = 1$

  $\dim \ker (T-(1-i)I)^2 = 2$
- Cycle tableau for eigenvalue $\lambda = 1-i$ is 
    - Jordan block in canonical form: $\begin{bmatrix} 1-i & 1 \\ 0 & 1-i \end{bmatrix}$
- $\dim \ker (T-I) = 2$

  $\dim \ker (T-I)^2 = 3$
- Cycle tableau for eigenvalue $\lambda = 1$ is 
    - Canonical form: $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
- Canonical form of T on $\mathbb{C}^5$:

$$\begin{bmatrix} 1-i & 1 & & & \\ & 1-i & & & \\ & & 1 & 1 & \\ & & & 1 & \\ & & & & 1 \end{bmatrix}$$